Comparison of Calibration Curves: an Application Example

Pier Giorgio Spazzini, Luca Callegaro & Francesca Pennecchi
INRIM – Italy
Strada delle Cacce, 91 – 10135 Torino (Italy)
Contact Author: P.G. Spazzini - +39 011 0906862/+39 011 0906859. p.spazzini@inrim.it

Bodo Mickan
PTB - Germany
Department 1.4 Gas - Bundesallee 100 - D-38116 Braunschweig (Germany)

Abstract: The present paper discusses an example of application of a method for the comparison of calibration curves. The method, presently in development, is expected to be an useful tool for the analysis of complex sets of data in several situations; two possible examples are the analysis of the drift of instrument with complex calibration curves or the analysis of data in Key Comparisons performed for quantities which require the use of measurement instruments as Transfer Standards.
The proposed example is based on data from a Key Comparison.
The paper discusses the merits of both the technique proposed for the determination of a reference calibration curve from the calibration curves pertaining to the various laboratories and the method for the computation of the distance between the single curves and the reference. Moreover, the advantages of the global evaluation of compatibility between calibration curves versus the point-to-point compatibility are discussed.

1. Introduction

The comparison of the outcomes of calibrations of instruments in different conditions (e.g. at various participating laboratories during International Comparisons, or in subsequent moments in time in order to analyze drift) is presently performed by comparing approximately equal calibration points obtained in the different conditions. Though, especially in the flow measurements field, this method presents some shortcomings as it is often difficult or even impossible to obtain that the reference stimulus is exactly the same in the various calibrations. Recently, our workgroup has developed a methodology to directly compare calibration curves obtained, usually, by fitting the calibration data [1,2,3].
In the present paper, after a brief recollection of the method for the comparison of calibration curves, a practical application will be demonstrated. The demonstration will be performed on actual data from the CCM.FF.KC5 Key Comparison, kindly provided by PTB, in order to show the usefulness of the method in real-life applications. Also, the data were already treated, therefore a comparison with other data analysis techniques will be possible.

2. Method Recapitulation

The method for the determination of a reference curve and its uncertainty from an arbitrary set of curves is described in full detail in [3], while [1,2] provide a complete discussion about the method for the computation of a distance between two arbitrary curves. We will provide here only a brief recall of the focal points of these methodologies.
The tool used for the computation of the distance between two curves is the L_p-norm, which is a very generic norm, which can be applied to many mathematical objects. In this context, however,

we restrict its definition to the case of bounded functions $f(x)$ defined on a closed and bounded subset of the real numbers, for example on the interval $[a, b]$, having length $\Delta = b - a$. Then,

$$\| f \|_p = \left(\frac{1}{\Delta} \int_a^b |f(x)|^p dx \right)^{1/p} \tag{1}$$

is the (normalized) L_p-norm of $f(x)$ on $[a, b]$ [7]. If $0 < p < 1$, expression (1) is no longer a norm, since it does not satisfy the triangular inequality.

Notice that the norm is defined over a support, and that its value is scaled to the extent of such support. Also notice that the computation of the norm includes a free parameter, namely the power p. The mathematical meaning of this parameter, and the consequences its choice implies on the computation of the distance between curves, was discussed in detail in [2]; in short, it can be said here that higher values of p tend to highlight the local peaks of curves, while low values tend to underline the overall distance from the axis; the utility and meaning of such features in practical applications will be discussed in the following.

If the L_p-norm computation is applied to the functions "difference" between a reference curve and the curves to be analyzed, it allows to determine an ordering between the various curves. Such ordering can be a function of p, as will be shown in the following.

The aforementioned reference curve can be determined from a set of given curves through an approach consistent with the application of the L_p-norm to the computation of the distance between curves; indeed, the reference curve can be built through a minimization method, i.e. as the curve such that the sum of its L_p-distances from the set of original curves is minimum. It is shown in [3] that this minimization problem is equivalent to the n problems of determining the points minimizing the distance from a set of points at the same abscissa (*slicing theorem*). In the following, this approach will be applied for the construction of the reference curves, using a value of $p=2$ in the minimization of the L_p-norm for most cases.

3. Test Case

In order to show the potential of the method, we present here an analysis performed on sets of data extracted from the CCM.FF.KC5 key comparison [8,9]. Data refer then to the calibration of the same instrument (gas flow meter) in different test rigs. First of all, data referring to the same nominal condition (flowing gas and flow pressure) will be discussed.
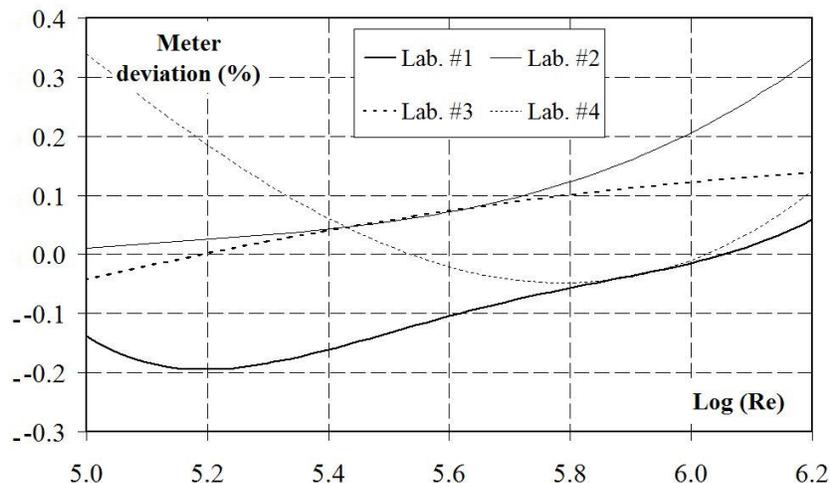

Fig. 1: Calibration Curves obtained from data measured at different Laboratories (case: flow pressure = 10 bar, working gas air).

2

The first analysis was performed on the data referring to the calibration of the transfer standard at a flow pressure of 10 bar, and using air as the working fluid. Such calibrations were performed at four laboratories. The corresponding data were extracted from the overall database, and the set of points from each laboratory was fitted using a standard Least Square Fit (LSF) method to provide a calibration curve. The chosen regression variables were the logarithm (base 10) of the flow Reynolds number as independent variable and the meter percent deviation as the dependent variable. The regression functions were polynomials of degree up to 4. The curves thus obtained are shown in Fig. 1.

After this, we faced the analysis of a more complex data, composed by data measured at four laboratories at different pressures. This allowed to increase the range of analysis, and to better highlight the advantages of the method proposed here

4. Results and Discussion

First of all, it is necessary to compute a reference curve. Application of the method discussed in [3] allows to determine various reference curves, depending on the choice of p.

When $p = 1$ is used, the pointwise estimator, calculated on each set of four points pertaining to the four curves at each abscissa point, is the median of such points, while when $p = 2$ it corresponds to their average. Based on this last observation, it is possible to perform a preliminary mathematical check. Actually, if the fitting functions all have the same mathematical form (e.g., they all are polynomials of the same degree), the reference curve generated with $p = 2$ should coincide with the curve obtained by averaging the coefficients of the original fitting curves. Therefore, in a first phase each of the four data sets discussed earlier was fitted with a polynomial of third degree; the method for reference curve generation was applied to these polynomials and the curve thus generated was compared to the one obtained by averaging the coefficients of the original fit; the curves coincided to within the numerical error of the approximations, confirming the reliability of the method for generating the reference.
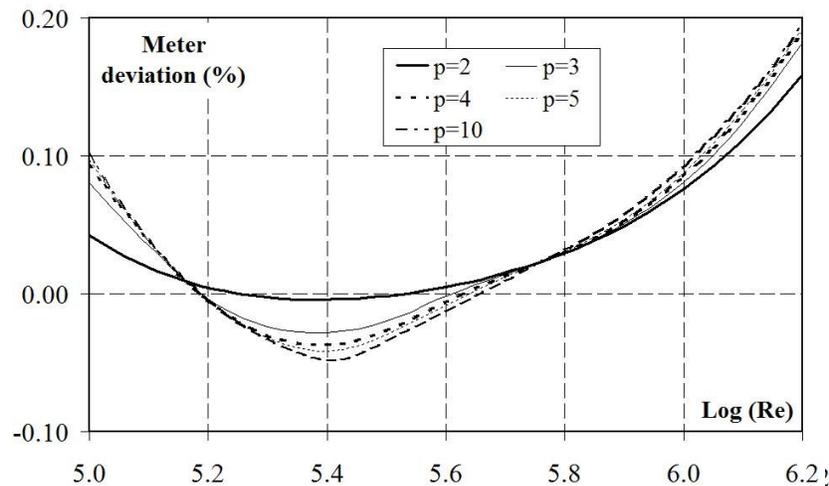

Fig. 2: Reference Curves generated for various values of the index p (case: flow pressure = 10 bar, working gas air).

Each data set was then fitted to a polynomial of optimal degree obtaining the curves plotted in Fig. 1; the degrees of the polynomials employed range from 2 to 4. Using these curves, it is possible to generate a reference for every value of p. Some of the curves thus generated are

3

plotted in Fig. 2. Notice that in the example only integer p values were used, but it would be possible to use non integer values.

It is possible to observe that the various reference curves differ slightly from a quantitative standpoint, but they have quite different behaviours. Notice that, while the curve corresponding to p=2 has a smooth trend, the more p increases, the more the curves tend to have high values at the limits of the considered range and a sharp peak at approximately Log (Re) = 5.4. This result is a consequence of the method used for the reference curves generation, and in turn of the features of the L_p-norm. Indeed, as discussed earlier, the norms with high values of p tend to highlight the large local differences, therefore a minimization based on those will generate a reference curve which tend to be the average of the curves with the maximum and minimum values, therefore not considering other curves. Considering now again the curves in Fig. 1, it can be observed that the curve referring to Lab. #1, which is always the lowest, is quite flat in the region around Log (Re) = 5.4, while the curve from Lab. #4, which in that region is the highest, is decreasing quickly. But, around that same point, there is a switch and Lab. #4 curve is surpassed by those from Labs. #2 and #3, which are increasing. This explains the low peak in Fig. 2. Similarly, in each region of the domain, a value of p lower than 2 will generate a reference curve influenced mainly by those which are the "central" curves in that region: the extreme case, $p = 1$, actually corresponds to the median, which is equivalent to say that only the intermediate curve (at every point) is considered. Now, when generating a reference curve, one would not want to give too much importance to the highest or lowest curves in a given region, because it is possible that the extrema are influenced by the presence of outliers (or, in limiting cases, the curve as a whole might be an outlier). On the other hand, all curves considered must contribute in the determination of the reference in order to keep into account all the information available.

These considerations lead to the conclusion that it is in general more appropriate to generate the reference curve using a low value of p, like $p = 2$ or not very much different from it. From now on, all reference curves considered in the present paper will be generated with $p = 2$.

Consider now Fig. 3, which displays the calibration curves pertaining to the various Labs together with the reference one determined from them, and 4, where the reference curve is plotted alongside with the original data points.
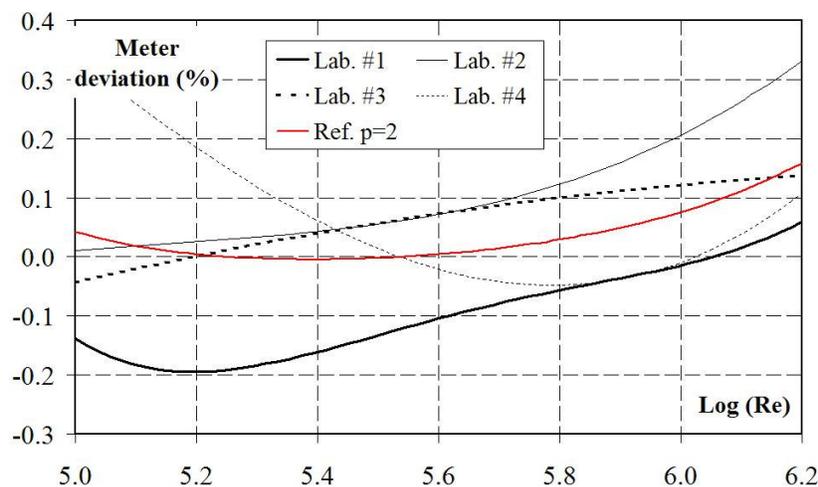

Fig. 3: Calibration Curves with the Reference generated by them for p = 2 (case: flow pressure = 10 bar, working gas air).

A qualitative evaluation allows to state that the Reference curve thus built can be considered satisfactory, as it clearly bears a relationship to all the originating curves. Moreover, it is relatively smooth and regular. Finally, it appears to be in good agreement with the original

experimental data. The trend of the curve also appears to be a good representation of the overall data trend.
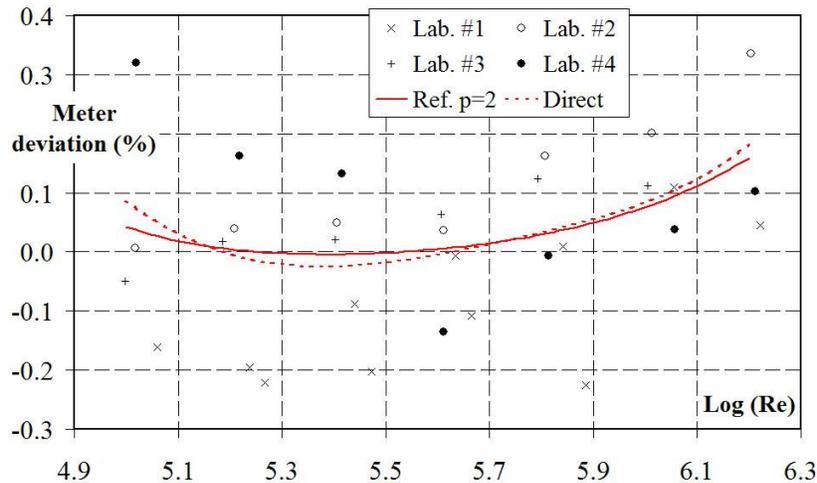

Fig. 4: Calibration Data Points with the References generated for p = 2 and by direct interpolation of the data points (case: flow pressure = 10 bar, working gas air).

Fig. 4 also reports a curve obtained by direct fitting of all the data points (dotted red curve). The procedure applied was again a fourth-degree polynomial LSF. This curve can be compared to the curve generated by the calibration curves obtained from all laboratories. It can be immediately observed that the two curves are very similar to each other. Though, the one obtained by direct interpolation appears to be slightly more sensitive to the high values at the extrema of the interpolation region and to the low values around 5.3. Despite these small differences, though, the two regression curves are very similar.

From these observations, it can be concluded that the procedure for the generation of the reference curve from single-Lab calibration curves is reliable and provides appropriate results.

Notice that this can be an advantage in situation where the original calibration data are not readily available, but only the calibration curves are provided.

Once the reference curve is determined, it is necessary to study the distance from it of the various single-Lab calibration curves. This task can be accomplished through the calculation of the L_p-norm of the difference of each curve to the reference. It can be observed at this point that, because of the normalization applied (see eq. (1)), the distance will have the same dimension as the quantity expressed by the curves analyzed, e.g. percent deviation in this specific case.

Fig. 5 reports the distances from the reference curve of the calibration curves from the single Laboratories. It can be observed that every plot tends to a maximum for increasing values of p; this is in agreement with the fact that, for very high values of p, the distance computed with the formula (1) tends to the maximum of the function under analysis (in this case, the difference between the reference and the calibration curve).

Another important observation is that the ordering of the distances is not independent of p. Indeed, while for low values of p the farthest curve is the one referring to Laboratory #1, when the parameter is increased the more distant curve is the one for Laboratory #4.

On the other hand, it can be observed that the result from Laboratory #3 has the minimum distance from the reference for any value of p and a very weak dependence on the value of p.

These results can be explained when considering the calibration curves reported in Fig. 3. Consider indeed the calibration curve referring to Laboratory #4. It can be observed that this curve has the maximum local distance from the reference, at the beginning of the range

5

considered (Log(Re) = 5). This difference is of approximately 0.3, which is exactly the value towards which the distance function tends. Also notice that, despite this high local difference, the curve is overall relatively close to the reference. This is reflected by the fact that, for low values of p, the difference is much lower; indeed, for this laboratory, the distance curve has a very steep gradient.
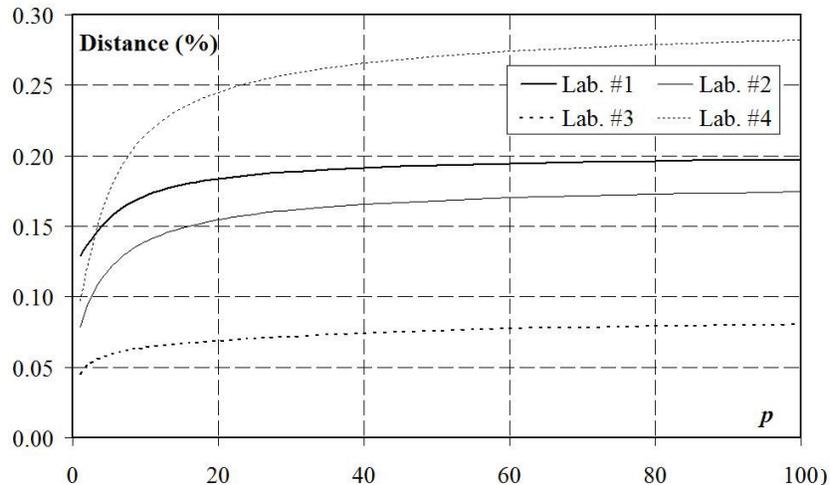

Fig. 5: Distances of calibration curves to the selected Reference for various values of p (case: flow pressure = 10 bar, working gas air).

On the other hand, the curve pertaining to Laboratory #1 is relatively far from the reference everywhere; therefore, its quantitative distance is not so strongly dependent on the parameter. Finally, the curve obtained by Laboratory #3 is very close to the reference throughout the examined range, therefore it has always the minimum distance and the distance is weakly dependent on p. It is interesting to observe that this curve can be considered as the best approximation to the reference although it clearly has a completely different mathematical expression (for instance, the two curves have opposite concavity). This result quantitatively expresses the qualitative evaluations that appear evident from the observation of Fig. 3, where clearly the curve from Laboratory #3 is "always close" to the reference, while the others are "evidently more distant".
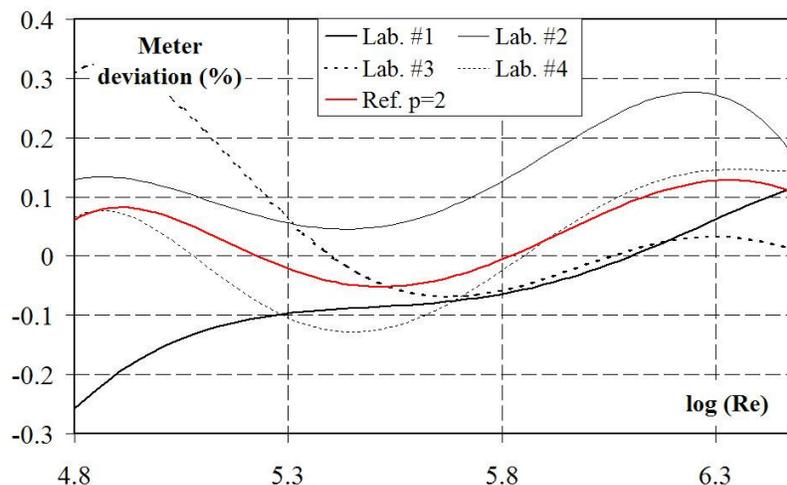

Fig. 6: Calibration Curves with the Reference generated by them for p = 2 (case: all flow pressures, working gas air).

6

The same analysis was then performed on a more complete data set, including data at various gas pressures in four laboratories, which allowed to extend the analysis range. Fig. 6 reports the calibration curves pertaining to the four laboratories and the reference generated from them in this case. It can be observed that the situation is more complex in this case; while it is still clear at a glance that the curve pertaining to Laboratory #4 is the closer to the reference, it is very difficult in this case to make a statement about the other ones.
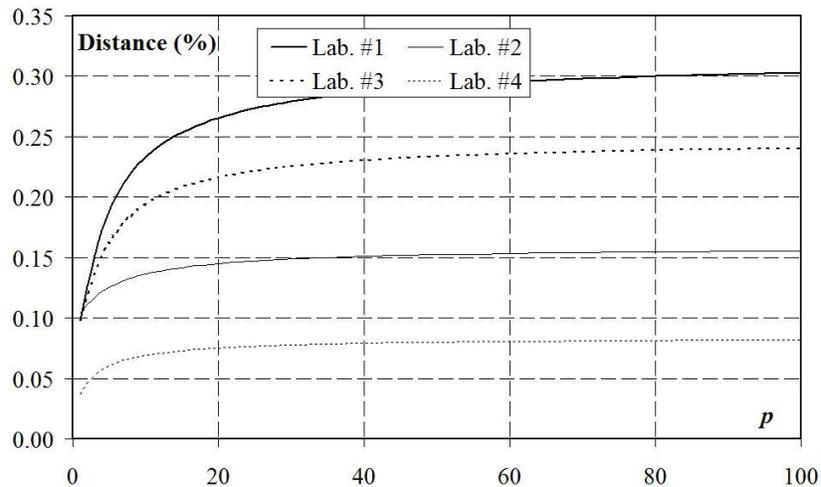

Fig. 7: Distances of calibration curves to the selected Reference for various values of p
(case: all flow pressures, working gas air).

The distance analysis, whose results are reported in Fig. 7, is in this case very useful. It can be observed that, as qualitatively expected, the result of Laboratory #4 is the closer to the reference for any p. For high values of p, there is a clear ordering between the other three laboratories. Though, it is interesting to observe that, despite the apparent large differences between them, the distance of the curves for Laboratories #1, #2 and #3 are practically equal at low p.

These observations can provide important information. The first conclusion that can be drawn is that these three Laboratories provided overall equivalent results; on the other hand, Laboratory #1 has higher local discrepancies from the reference result, while the local discrepancies of Laboratory #2 are much lower; etc…

5. Conclusions

In this paper an example of application of a method for the comparison of calibration curves was presented. A method for determining reference calibration curves was presented and commented, and a guideline for the selection of the appropriate value of the parameter p was indicated. Next, a reference curve thus obtained was used to analyse the relative distance of the various calibration curves from it and to classify them accordingly. A relatively simple case was presented for a first approach, showing that the method proposed actually allows to express quantitatively the "closeness" of two curves. A more complex case showed that the method, because of this ability, allows to draw important conclusions about the comparison of calibration curves performed at different laboratories (equivalently, it would be possible to draw similar conclusions about calibration performed in different moments at the same laboratory).

Of course, the methods discussed in the present paper are not yet complete. What is still missing is a method for the evaluation of the uncertainty of the reference curve and, more important, for using this uncertainty with the aim of defining a DOE (Degree of Equivalence) of the calibration curves from the various laboratories amongst them and with the reference curve. Clearly, such

techniques would complete the methodology and provide researchers in metrology with a powerful tool for the analysis of data sets. These challenges are currently being addressed by our workgroup, and will be the subject of future papers.

References

[1] Callegaro L., Pennecchi F. and Spazzini P. G. *Comparison of Calibration Curves Using the L_p Norm.* In "Proceedings of the 2008 IMEKO TC8 Workshop", Torino, Italy, 2008.

[2] Callegaro L., Pennecchi F. and Spazzini P. G. *Comparison of Calibration Curves Using the L_p Norm.* Accreditation and Quality Assurance 14 (11), pp. 587-592, 2009.

[3] Callegaro L., Pennecchi F. and Spazzini P. G. *Determination of Reference Calibration Curves.* In preparation.

[4] Gonin R. and Money A. H. *Nonlinear L_p - Norm Estimation.* Marcel Dekker, New York, 1989.

[5] Sposito, V.A. *Some properties of L_p - estimators.* In "Robust Regression, Analysis and Applications", K. D. Lawrence & J. L. Arthur, eds., pp. 23-58. Marcel Dekker, New York, 1990.

[6] Nyquist H., *The optimal L_p norm estimator in linear regression models.* Commun. Statist.– Theor. Meth., 12, 1983, pp. 2511–2524.

[7] Hwang S. Y., *On Monotonicity of L_p and l_p Norms.* IEEE Trans. Acoust., Speech, and Signal Processing, ASSP-f 3, 1975, pp. 593–594.

[8] Dopheide D., Mickan B., Kramer R., Hotze H. J., van der Beek M., Blom G., Vallet J. P., Gorieu O. *Final report on the CIPM Key Comparisons for Natural Gas at High-pressure Conducted in November / December 2004 CCM.FF-5.a.* 2005

[9] Mickan B., Kramer R., Dopheide D. *The Linking of the CIPM Key Comparisons CCM.FF-KC5a (for Natural Gas) and CCM.FF-KC5b (Compressed Air/Nitrogen) using Model Based Analysis of the data.* Proceedings of the 14th FLOMEKO Congress, Johannesburg, South Africa, 2007.