Main Index

North Sea
FLOW
Measurement Workshop

Paper ~~1~~: – 1.1

COMPARING PERFORMANCE OF MULTIPHASE METERS
«How to see the woods for the threes?»

Authors:
Chris J.M. Wolff, Shell International Exploration and Production b.v.
The Netherlands

Organiser:
Norwegian Society of Chartered Engineers
Norwegian Society for Oil and Gas Measurement
Co-organiser:
National Engineering Laboratory, UK

Comparing performance of multiphase meters
How to see the wood for the trees?

Chris J.M.Wolff

Shell International Exploration and Production b.v.
The Netherlands

SUMMARY

A method is presented to condense the results of multiphase meter performance evaluations into one or at the most a small number of characteristic values. Ranking of these allows easy performance comparison of different meters. One important characteristic is the fraction of the data points that satisfies a certain (in-)accuracy criterion, which is called the "yield". Another one is the width of the error distributions, for which the standard deviation appears to be unsuitable due to the large number of outliers. More workable numbers for the widths are obtained when at either side of the distribution 10% of the data points are discarded and the width between remaining extreme errors is selected (the 80% width). The approach also demonstrates that expressing the measurement errors in terms of relative oil and water flow rate errors produces results which give an unnecessary poor impression, due to the fact that without separation of the phases, the relative error of a quantity close to zero is bound to be large. Using the relative errors for gas and liquid and absolute errors for BSW gives more consistent and more useful results. Finally it is shown that the often heard statement: "the multiphase meters investigated, generally satisfied the 10% accuracy criterion" cannot be supported using the conventional 95% confidence limits.

INTRODUCTION

Various comparative tests of multiphase meters (Porsgrunn 1 & 2, Multiflow 1 & 2 at NEL, Texaco JIP at Humble) have invariably produced large quantities of data, which are usually presented in a large number of scatter graphs. The result is that the observer is inundated with information and it has become virtually impossible to assess or compare the performance of the different meters. For comparison of meter performance one would really need a means of ranking the meters. To allow ranking it is required that the performance is condensed in a single figure (a scalar performance indicator). In this article four different methods of calculating such a performance indicator will be introduced and their usefulness discussed. It will also be shown that statements of overall performance of meters are very dependent on which of these four performance indicators is actually used. Usually this specification is not included in the statements. It is needless to say that a performance indicator always applies to a certain operating envelope or test data set, hence it must not be used without that being defined.

Very often engineers will want to know why the performance of a meter is as it is and they want to analyse the test results. To facilitate this process some further key parameters are suggested. Their number should be as small as possible to allow easy comparison.

Finally, it is shown that the method of converting the three-dimensional measurement results in a scalar can be used to identify the area in the operating envelope where a meter has its worst performance. By excluding this area from the recommended operating envelope it should be possible to arrive at an improved performance indicator of the meter for the remaining area. It is hoped that meter suppliers will adopt this method to sharpen up the specification of their meters performance.

DIFFERENT PERFORMANCE DEFINITIONS

The first performance criterion to be considered is based on the accuracy requirement as formulated by the users of the information, which is usually:
"The relative uncertainty of oil, water and gas flow rates should be better than a certain percentage, for which often 10% is taken.".
Rather than the word "uncertainty", one often uses "inaccuracy" or even just "accuracy". In real life the distinction between these is not always adhered to. That may even apply to the rest of this article.

The above requirement means that for one measurement consisting of three components (oil, water and gas) none of the three relative errors must be larger than the 10% criterion. In mathematical formula:

$$MAX\left(ABS\left(\frac{\Delta O}{O}\right), ABS\left(\frac{\Delta W}{W}\right), ABS\left(\frac{\Delta G}{G}\right)\right) \leq Cri \qquad \text{(1)=M\sim OWG}$$

The criterion value (Cri) can be 10% or any other agreed number.

A different interpretation of the verbal requirement statement is also possible. And it would appear that in practice, people discussing the performance of multiphase meters rather use that different definition. This is usually not clearly defined, but it is stated here that in practice one seems to consider the relative errors of oil, water and gas separately.
In formula

$$\begin{aligned} &ABS\left(\frac{\Delta O}{O}\right) \leq Cri. \\ &ABS\left(\frac{\Delta W}{W}\right) \leq Cri. \qquad \text{(2)=S\sim OWG} \\ &ABS\left(\frac{\Delta G}{G}\right) \leq Cri. \end{aligned}$$

It will be shown that this difference in definition results in a completely different "yield" of data points satisfying the criterion.

A different formulation of the requirements was presented by Slijkerman et.al. [1] in 1995 on behalf of authors from Statoil, SAGA, Norsk Hydro, BP and Shell. It presents the following general requirements:

5% to 10% relative accuracy in total liquid flow rate (L);
5% to 10% relative accuracy in gas flow rate (G)
2 % absolute accuracy in water cut measurement (B)

In the mathematical formulation introduced above this would become either

$$MAX\left(ABS\left(\frac{\Delta L}{L}\right), ABS\left(\frac{\Delta G}{G}\right), p*ABS(\Delta B)\right) \le Cri \qquad \text{(3)=M~LGpB}$$

or

$$\begin{aligned} &ABS\left(\frac{\Delta L}{L}\right) \le Cri \\ &ABS\left(\frac{\Delta G}{G}\right) \le Cri \qquad \text{(4)=S~LGpB)} \\ &p*ABS(\Delta B) \le Cri \end{aligned}$$

The criterion value (Cri) can be taken 5%, 10% or any other value. The weight factor p has been introduced to allow a more stringent criterion to be applied to the water cut accuracy. When a relative accuracy of 5% is used for liquid and gas, and 2% absolute accuracy for water cut, then p is 2.5. In this article a default value for p of 2 will be used throughout.

A numerical example to illustrate the above has been calculated using the data set produced by one of the multiphase meter evaluation in the course of 1996. The data set comprised 120 measurement points at different flow rates, compositions and pressures. The yield of data points satisfying the criterion has been calculated and expressed as a percentage of the total number of data points. For definitions (1) and (3) this was straight forward. For definitions (2) and (4) the three errors constituting one measurement point were taken separately, as if the data set comprised 360 independent points. To obtain some insight in how the yield (of "good") points depended on the criterion value chosen, the latter was varied between 0 and 50%, resulting in a kind of cumulative error distribution. The result for one of the meters submitted to the test is shown in Fig.1.

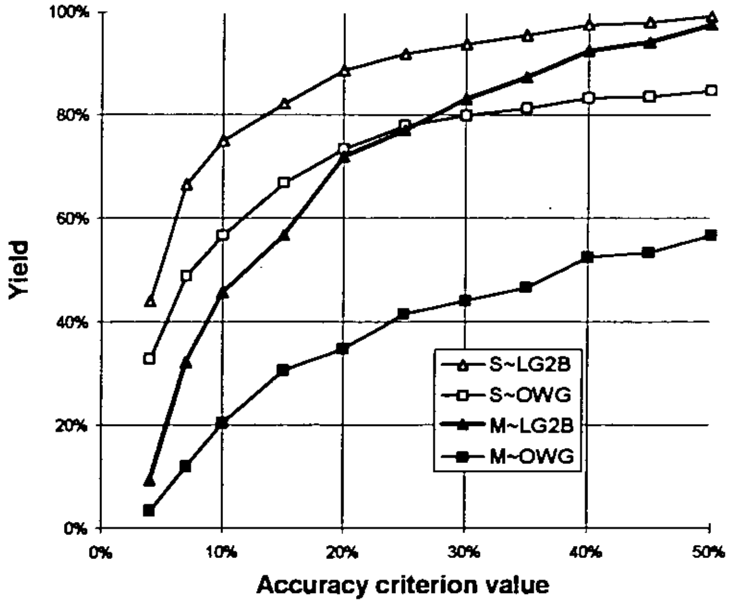

Fig.1 Typical cumulative error distribution curves (or yield curves) for an arbitrary meter submitted to a test and for four different accuracy definitions M~OWG, M~LGpB, S~OWG and S~LGpB as defined in the text.

Table 1 Yields of six meters (M1 - M6) during final test series at Porsgrunn 2

		M1	M2	M3	M4	M5	M6
Yields when Cri=10%; p=2							
	M~OWG	17%	7%	20%	15%	13%	10%
	M~LG2B	30%	21%	45%	42%	14%	24%
	S~OWG	52%	35%	57%	52%	50%	35%
	S~LG2B	69%	51%	75%	68%	64%	57%
Yields when Cri=20%; p=2							
	M~OWG	37%	25%	34%	34%	64%	29%
	M~LG2B	82%	54%	71%	68%	66%	48%
	S~OWG	75%	54%	73%	68%	82%	57%
	S~LG2B	93%	73%	88%	84%	89%	75%

Table 2a Ranking of Porsgrunn 1 meters on 10% criterion (1 is best, 5 is worst)

meter:	M1	M2	M3	M4	M5
M~OWG	5	1-2	1-2	3	4
M~LG2B	5	3	1	2	4
S~OWG	5	4	2	3	1
S~LG2B	5	4	2	3	1
AVERAGE	5	3.1	1.6	2.8	2.5

Table 2b Ranking of Porsgrunn 1 meters on 20% criterion (1 is best, 5 is worst)

meter:	M1	M2	M3	M4	M5
M~OWG	4	3	1	2	5
M~LG2B	5	3	1	2	4
S~OWG	5	4	2	3	1
S~LG2B	5	4	2	3	1
AVERAGE	4.8	3.5	1.5	2.5	2.7

Notice that at the accuracy criterion value of 10% the yields range from 20% to 75% depending on the accuracy definition used. E.g. only 20% of the data points satisfy the criterion if the worst (rel.) error of oil, water and gas is used (M~OWG). But, on the other extreme, if the liquid, gas and BSW are used as parameters and the errors considered separately (S~LG2B) 75% satisfy the 10% criterion value. This illustrates the drastic influence of the definitions used on the yields and hence on our conclusions about overall performance. The definitions acknowledging the three dimensional character of the measurement and using the worst accuracy of the three elements as the determining factor (M~OWG and M~LG2B) are clearly much more stringent than the definitions that treat all measurement errors separately (S~OWG and S~LG2B). In addition the oil, water gas definitions produce a much lower yield than the liquid, gas, BSW ones. This is also understandable since the relative error of a small fraction of any phase is bound to be large and this occurs more frequently when the pair of oil and water are considered than with the pair of liquid and BSW. This was actually the reason behind the formulation of the requirements presented by Slijkerman et.al. [1].

The author is of the opinion that for a proper characterisation of the performance of the multiphase meters the maximum type accuracy definitions should be used, i.e. either M~LGpB or M~OWG. Of these two, on practical grounds the liquid, gas, BSW definition (M~LGpB) is preferred, in line with Slijkerman et.al.

COMPARISON

Calculation of the yield curves for different meters submitted to the same test points, allows the performance of the meters to be compared. For ranking of the meters one only needs to calculate the yield figure at an agreed accuracy criterion value and do a ranking based on that single performance figure. According to Slijkerman et.al. the criterion values should be taken at 5% or 10% with p=2.5 resp. 5. However 5% produces such low yields in the data sets of the evaluations available (Porsgrunn 1995, Porsgrunn 1996 and NEL), that it was considered practical to standardise for the time being on 10%, with 20% as a back-up. The weighting factor p has been taken to be 2, as mentioned above.

As example, in Fig.2 the percentage data points satisfying a 10% accuracy criterion value (Yields) are given for the five meters of the Porsgrunn 1 test data set (340 data points) for each of the four different accuracy definitions as introduced above. In Fig.3 the same is done applying a 20% criterion value. In table 1 similar information is given in numerical form this time for the six meters tested in the Porsgrunn 2 trial. Whether graphical or tabular presentation should be used, is a matter of personal preference. For that reason an example of either form is presented in this paper.

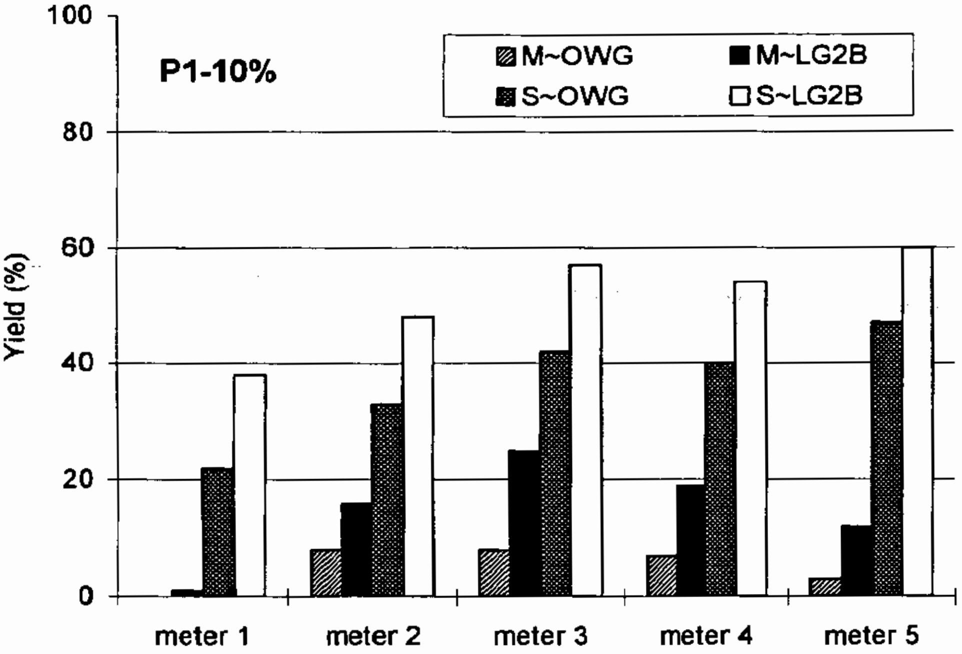

Fig.2 Percentage data points (Yields) for five meters used at Porsgrunn1 trial satisfying 10% accuracy criterion value

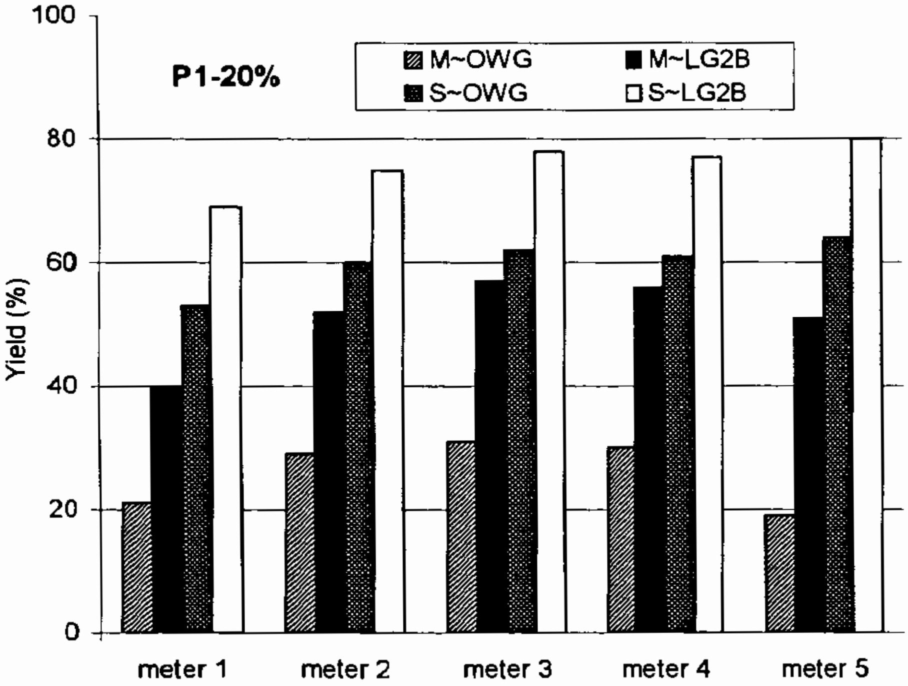

Fig.3 Percentage data points (Yields) for five meters used at Porsgrunn1 trial satisfying 20% accuracy criterion value

It is possible to do a performance ranking of meters submitted to the same test points. As an example the result of ranking the Porsgrunn 1 meters (yield data of Fig.2 and 3) is shown in table 2a and 2b. It appears that the ranking is hardly dependent on the criterion value used (10% or 20%)
Another rather surprising result from this exercise is that the ranking order is not too dependent on the choice of accuracy definition applied either, with one exception which is discussed below. Certain meters always rank high (or low), regardless the definition used. Hence the ranking method allows one to home in on the better performers. Meter 3 ranks clearly very high.
Note meter 5, which is the best if the three components of the measurement are taken separate, but is almost the worst if the three-dimensional character of the measurements is taken into account. With this meter apparently always one of the three components has a larger error.

However, a ranking does not tell us whether the better meters also satisfy our needs. For that we need a quantitative assessment. In the following paragraphs it will be shown that quantitative assessment is possible, but not as straight forward as it might seem to be at first.

In table 1, one can see that even for the best meter of the Porsgrunn 2 evaluation only 75% of all data points satisfy the S~LG2B<10% criterion and only 45% satisfy the M~LG2B<10% criterion. The better yields for the 20% criterion are 93% and 82% respectively.

Now, before trying to conclude on a typical accuracy level achieved, it has to be pointed out that the mathematical/statistical basis of this subject is not simple. Note that we are dealing here with both one-dimensional and three-dimensional error distributions. With single phase meters one would usually require that 95% of all data points satisfy the accuracy criterion, and assumes that the plus/minus 2-sigma range of the error distribution corresponds with a 95% confidence level. But for multi-dimensional normal distributions the relationship between the width of a distribution and the confidence levels is not the same as for one-dimensional normal distributions. As is shown in table 3 for a three dimensional distribution, the plus/minus 2 sigma envelope corresponds with only a 74% confidence level and not the 95% that we are used to. It would seem that for a fair assessment of the performance of multiphase meters one should take a 95% confidence level (and yield) target for the one dimensional error distributions (S~OWG and S~LG2B accuracy definitions), but use the 74% confidence level (and yield) target as an equivalent for the three-dimensional distributions (M~OWG and M~LG2B). The author would like to invite people better skilled in statistics to express their opinion on this. For the time being it suffices that this mathematical phenomenon is highlighted.

Table 3. Confidence levels for one-, two- and three-dimensional gaussian distributions with equal standard deviations in all directions

	One-dimensional (%)	Two-dimensional (%)	Three-dimensional (%)
+/- 1 σ	68	39	20
+/- 2 σ	95	86	74
+/- 3 σ	99.7	98.9	97.0

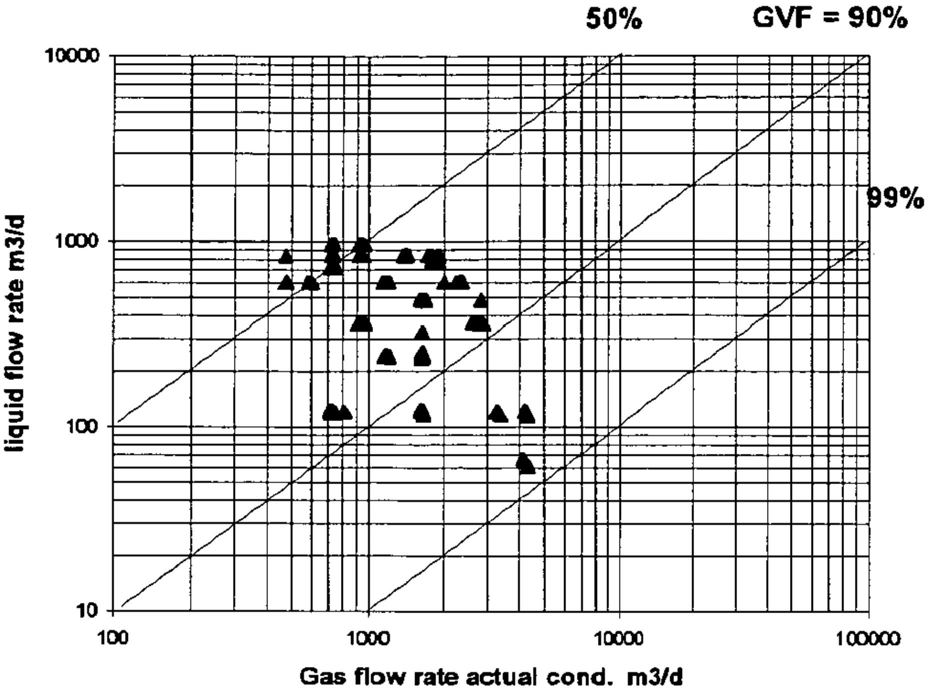

Fig.4a Example of graphical representation in the two-phase map of set of measurement points used

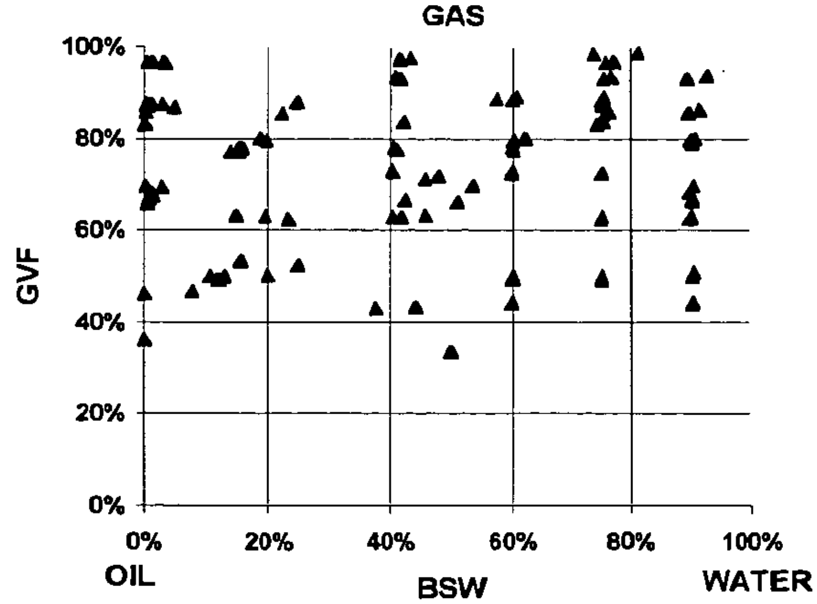

Fig.4b Example of graphical representation in the composition map of set of measurement points used

Assuming that the above mathematical reasoning is correct, then for an overall meter performance of 10% accuracy, the S (separate) criteria should produce a yield of more than 95% and the M (maximum) criteria more than 74% . This has not been achieved so far. (See Figs 2 and 3, and Table 1). A 20% accuracy (for the entire data sets) seems closer to the reality. For a reduced operating envelope better accuracies may apply. This will be addressed later.

The method proposed allows to compare performance of multiphase meters, however as mentioned earlier, this requires that the test data set to which the performance figures apply is known and well defined. Therefore this test set (the test envelope) must be presented together with the performance results. It is suggested that the reference measurement set is presented in graphical form, as shown in Fig.4 a&b, comprising of a standard two-phase map and a composition map (GVF vs. BSW). Tables would in principle give the information as well, but the quick insight one gets from a graphical representation is often preferred.

WIDTH AND MEDIAN OF DISTRIBUTIONS

Whilst the yield figures give a basic measure of the performance of a meter demonstrated during a test, they give little insight in the underlying strong points or shortcomings of the meter. Such insight can be achieved by looking at the error distributions of the individual measurement components. The components recommended for consideration are:

$\Delta O/O$, $\Delta W/W$,$\Delta G/G$, $\Delta L/L$, ΔB

Notice that the GVF is excluded. The reason for this in the nature of the quantity, which is strongly non-linear. E.g. a 1% error at GVF 50% has a different impact than the same error at e.g. GVF 98%. It would for that reason make sense to work with a gas-liquid ratio at line condition rather than the GVF. But this is not usual and will not be introduced in this article.

It has been investigated whether average values of the (rel.) errors would be a useful characteristic value, but it appeared that the averages are too strongly influenced by the many extreme outliers. The **median** of the error distribution works much better. The median is the data point with 50% of the data points having a smaller value and the other 50% having a larger value.

Similarly the standard deviation of the error distribution was adversely affected by the outliers as well. A more robust measure of the width of a distribution appeared to be the maximum and minimum errors when 10% worst error data points at either side had been discarded. The result has been called the 80% width. In a one-dimensional normal distribution this would correspond with the range from -1.28σ to $+1.28\sigma$. To allow direct comparison with the accuracy criterion value as used in the yield graphs (Fig. 1) rather **half the 80% width** will be used. (The "half 68%-width" would correspond with 1σ)

Initially the width between the upper and lower quartiles was considered, but this would leave only 50% of the data points covered in the range. This was felt to be too low.

Only the width of the gas, liquid and BSW error distribution appeared to be meaningful in general. The oil and water relative error distributions contained too many very extreme values. This is simply due to the fact that oil and water flows approach zero too frequently. The same reason as was behind the proposal to express meter performance in terms of gas, liquid and BSW.

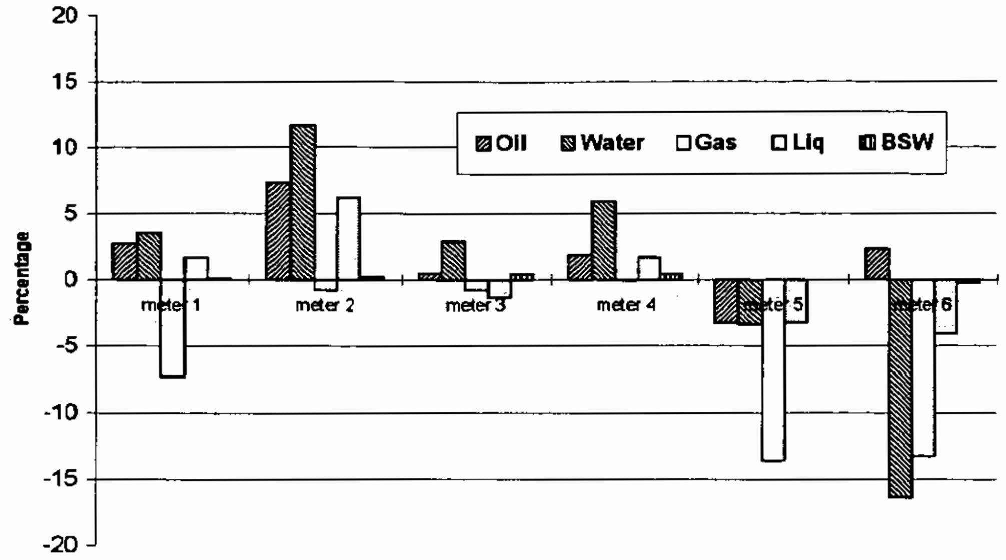

Fig.5 Medians of the error distributions of components for six different meters

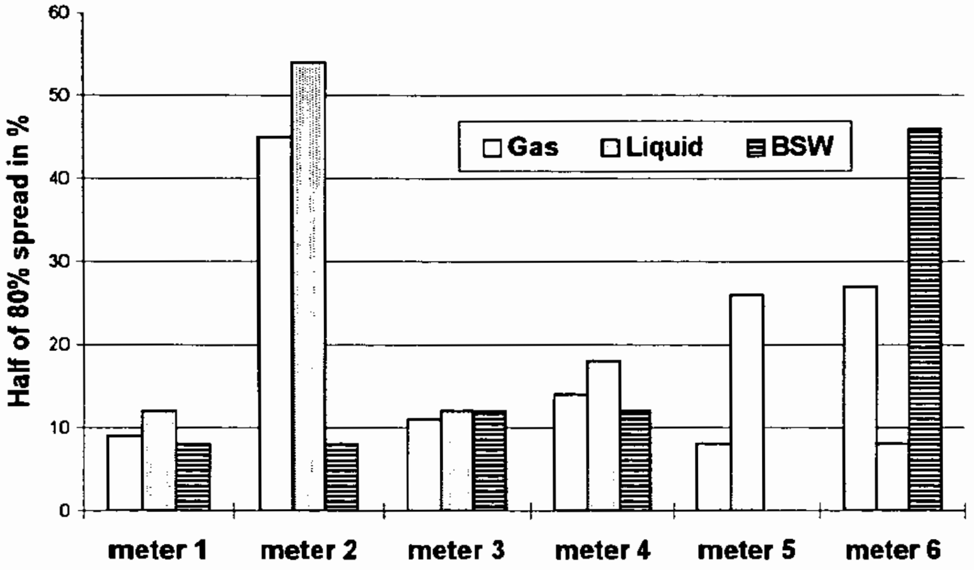

Fig.6 The width if the error distributions of compnents expressed as half the distance between the extreme error values when 10% of the data points at either end of the distribution have been discarded.

As an example the medians and half 80%-widths of the error distributions of the Porsgrunn 2 trial are depicted in Fig 5 and 6 respectively. The most remarkable features to be seen in the graphs are:
1) Meter 2 sticks out with half widths of the gas and liquid error distributions around 50%.
2) The half width of the BSW distribution of meter 6 is about 45%.
3) Meter 1 has a small width of the gas distribution combined with minus 8 % median. Hence the gas measurement is "well defined" but there seems to be a systematic error in the model used. Here seems to be scope for improvement through a better model for converting the raw measurements into reported quantities.

Using the width of a distribution and its number of points an estimate can be made of the uncertainty of the mean. With 120 points in the distribution and a width of plus or minus 10%, the reliability interval of the mean is approximately 1% (10% divided by square root 120). When the width of the distribution is plus or minus 40% the reliability interval of the mean is still only +/- 2%. The author has simply assumed that what applies to the mean also applies to the median.

In general it would appear that whilst the "yield" is a good measure for the observed performance of a meter, the "widths of the error distributions of the components" are a very useful indication for its potential, more so than the median.

SUPPLIER GIVEN SPECIFICATION

The yield method can also be used to check how many of the test data point actually satisfy the supplier given meter specifications. It is suggested that the suppliers do this themselves. The accuracy definitions used could be either MAX(O,W,G) or MAX(L,G,p*B) or both of them. It would give the clients a good measure of the level of confidence that can be attached to the specification.

STANDARD REPORT

The above analysis of the available test results suggest that the summary of the evaluation of a meter should consist of

Minimum:
1a) graphical representation of the test set (Two-phase map and composition map)
1b) Number of data points in the test
2) 10% (5%) yield figure according to M~LG2B
3) Yield percentage according to manufacturers own specification

To gain more insight this minimum could be expanded to an example characterisation as given in Table 4. As can be seen from the data for this particular meter, the gas measurement seems to be the weak spot for this meter (minus18% median), but the width of the gas distribution (only 8%) is good enough to enable improvement through better modelling.

SELECTING BEST OPERATING ENVELOPE

Suppliers could also improve their "score" by narrowing down the operating envelope of their

Table 4. Example of full set of numbers characterising the performance of a multiphase meter during a test. (Values are taken from a real meter evaluation comprising about 210 points. The corresponding data set is not presented)

Yields	**10% accuracy**	**20% accuracy**
M~OWG	0%	16%
M~LG2B	3%	30%
S~OWG	38%	69%
S~LG2B	55%	83%
	Median	**half 80% width**
Oil	2.9%	19%
Water	-0.4%	61%
Gas	-18.0%	8%
Liquid	0.3%	8%
BSW	-0.2%	6%

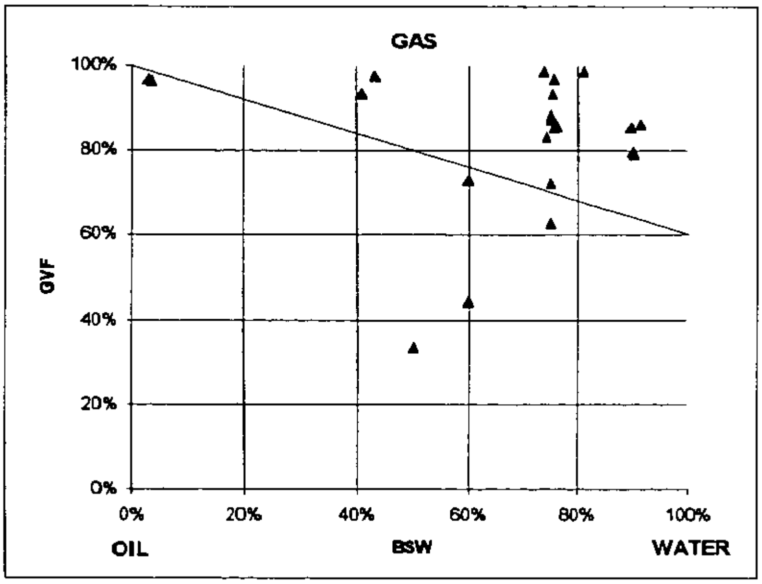

Fig.7 Twenty percent worst data points for one of the meters plotted in the composition map. It won't surprise that the points are concentrated in the comer of high BSW and high GVF.
Line is GVF = 1- 0.4 * BSW

meter to the area where the meter performs best. One way of determining this optimum area is by chopping the area(s) of worst performance. As an example this has been done for one of the meters out of the available data sets. For each data point the maximum of the liquid, gas and 2*BSW errors were calculated (M~LG2B). The points were ranked on this value and the 20% worst points were plotted in a composition map. The result is shown in Fig.7. One sees that as could be expected these "worst" performance points are concentrated around the high GVF, high BSW area. The next step was to reduce the specified operating envelope by excluding compositions with high GVF and high BSW. The split was made according to the line

GVF> 1 - 0.4*BSW

This was an arbitrary choice, just to demonstrate the method. For the remaining data points the yield figures were calculated again. The results are shown in table 5.

Table 5 Comparison of yields for one of the meters. Old is total data set and New is after removal of the high BSW, high GVF points.

	Old	New		Old	New
Cri	10%	10%		20%	20%
M~OWG	15%	19%		34%	44%
M~LG2B	42%	55%		68%	84%
S~OWG	52%	57%		68%	73%
S~LG2B	68%	77%		84%	91%

CONCLUSIONS

A method is presented that allows the results of multiphase meter evaluation tests to be condensed into no more than a small number of characteristic values. The prime parameter is the yield figure which is the percentage of data points satisfying a certain accuracy criterion.

Provided the evaluation is done on the same or at least a comparable reference data set, the performance of meters can be ranked based on these yield figures. The yield figures can also be used as a general qualification of meter performance if the reference measurement set is not explicitly defined but considered to be typically applicable for that meter.

Based on the test data available and the definitions introduced in this paper, it should be concluded that the overall accuracy of multiphase meters is closer to 20% relative for liquid and gas and 10% abs for BSW, than to 10% and 5% respectively.

The method can be used to demonstrate experimentally the confidence level one can attach to vendor supplied accuracy performance specifications of multiphase meters.

The method can also be used to identify the operating envelope corresponding with optimum performance of a certain meter.

REFERENCE

[1] Slijkerman, W.F.J. et.al. "Oil Companies Needs In Multiphase Flow Metering" NSFMW, Lillehammer 1995, paper 23.