27th International North Sea Flow Measurement Workshop
20th– 23rd October 2009

Inferential Chemometric Allocation

Phillip Stockton, IMASS (formerly Smith Rea Energy Ltd)

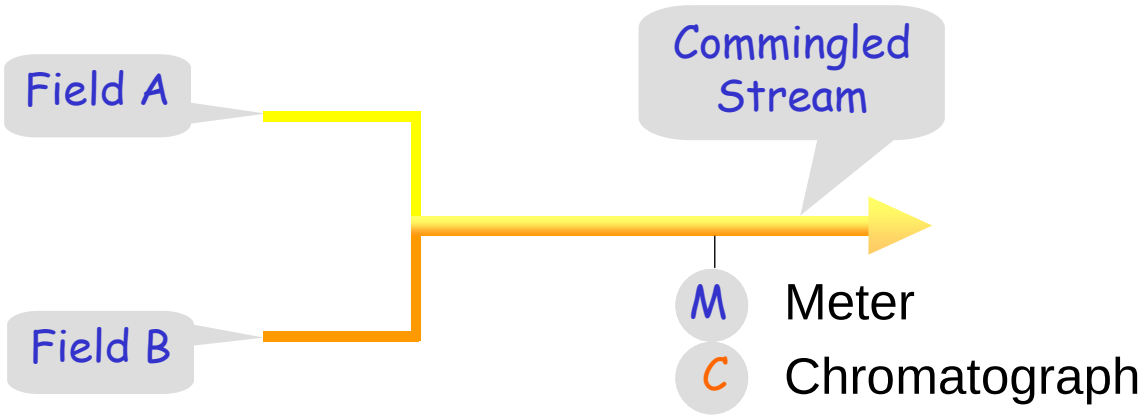
1 INTRODUCTION

What is "Inferential Chemometric Allocation"?

Chemometrics is the science of extracting information from chemical systems by data-driven means. The chemical data used in this approach to allocation are the compositions of the feed and product streams in a commingled system. These are used to infer an allocation of the product stream between two or more contributing streams without measuring the flow of the feed streams.

A simple example will illustrate the concept. Consider two streams: Field A and Field B, each of different compositions. The composition of each stream is known and remains constant. These streams are mixed together and the commingled product's composition is measured as shown in Figure 1.

Figure 1 – Two Streams of known Composition Commingled



The commingled stream's composition varies in accordance with the relative contributions of each of the feed streams.

For example if Field A is 70 wt% methane and Field B 60% wt methane, then if the measured composition has 65 wt% methane we can infer that half the feed can be attributed to Field A and half to Field B. Similarly, if the commingled composition was 67.5 wt% methane we would infer that three quarters of the flow is from Field A and one quarter from B. The product composition offers a means of allocating to the feed streams without the need for individual flow measurement of the contributing fields.

In simple terms this is the basis of inferential chemometric allocation, i.e. it utilises the composition of a commingled product stream to allocate between two or more

feed streams. This is a viable means of allocation if the feed streams' compositions are essentially constant or known and sufficiently dissimilar from one another.

However, a number of questions arise:

- In the example, if a component different to methane had been selected would the answer have been different?
- How accurate is the method when subject to the uncertainties in the measured compositions?
- How dissimilar do the compositions need to be?
- Is it possible to use this method to allocate between more than two fields?
- Would we know if one of the assumed constant compositions starts to drift?

Or, in summary:

- Is this a viable means of allocation in the real world?

This paper attempts to explore and answer these questions.

Section 2 presents various Inferential Chemometric Allocation methods and illustrates the concepts using simplified examples.

As this is a novel allocation methodology and is in the development stage it has not been implemented in any real systems. Hence in order to test its accuracy and practicability it has been necessary to construct theoretical but representative data. In addition, by utilising data from a real system in which the feed flows are known, it is possible to test the methods in what may be considered real world scenarios for both gas and liquid systems. Section 3 presents the results of this analysis and explores the questions of allocating to more than two fields and detection of compositional drift.

Having stated all this, there is a final question to be answered:

- Why would we want to do this?

Section 4 discusses possible applications and developments and Section 5 summarises the conclusions.

2 CONCEPTS

2.1 Simple Compositional Based Split

Consider Example 1, which comprises two fields, Field A and Field B, of differing compositions, being commingled in a 50:50 mixture. Imagine that the true composition and flow data are known and the commingled stream is a perfect mixture of the two fields. The true measurement data for this perfect system is presented in Table 1.

Table 1 – True Compositions and Flows of Example Perfect System

	Field A wt %	Field B wt %	Commingled wt%
C1	60%	70%	65.0%
C2	30%	25%	27.5%
C3	10%	5%	7.5%
Flow (te)	500	500	1,000

If the true compositions of all streams but only the flow of the commingled stream were known, we could still infer the flows of Field A and B by examining any of the component weight fractions in the commingled stream.

More formally, the equation for calculating the contribution of Field A, based on the weight fraction of component C1 in the three streams, is given by:

$$S_A = \frac{\left(C_{C1} - B_{C1}\right)}{\left(A_{C1} - B_{C1}\right)} \tag{1}$$

The derivation of (1) is presented in Section 6.1. A similar equation can be written for any of the components and these will give the same answer for S_A if the data is perfectly consistent.

However, in the real world the compositions and flows would not be known perfectly and the compositional analyses would be subject to measurement uncertainties. Applying randomly generated measurement errors to the compositions in Table 1, the actual measured data may typically look like that presented in Table 2:

Table 2 - Compositions of Measured Streams

	Field A wt %	Field B wt %	Commingled wt%	Relative Measurement Uncertainties wt%
C1	59.89%	69.88%	64.93%	± 0.6 %
C2	30.01%	25.17%	27.42%	± 1 %
C3	10.09%	4.96%	7.65%	± 3 %

Using the above data it is possible to infer the split of the two Fields using Equation (1). Based on the C1 weight fractions the contribution from (or split to) Field A is calculated to be 49.58% and therefore 50.42% is from Field B. These are reasonably close to the true split of 50%.

If we had chosen C2 to determine the Field A split, the answer would have been slightly different and if C3 had been chosen different again; the calculated splits based on each of the three components are presented in Table 3:

Table 3 – Calculated Field Splits

	Field A		Field B	
	Split	Error	Split	Error
Perfect	50.00%		50.00%	
C1	49.58%	-0.42%	50.42%	0.42%
C2	46.49%	-3.51%	53.51%	3.51%
C3	52.50%	2.50%	47.50%	-2.50%

Each component gives a slightly different answer and in a real system with 10+ components there will be many more choices. Which one is the best estimate?

C1 appears the obvious choice here, as it is the major component and its measurement uncertainty in relative terms is the lowest. However, what would happen if the C1 content of the two Fields was closer?

Consider Example 2, which is the similar to Example 1, except that Field B's true C1 content is changed to 61% and C3 to 14%. The inferred allocated quantities now become:

Table 4 – True and Measured Compositions
Field C1 Content Similar

True Values			
	Field A	Field B	Commingled
	Perfect	Perfect	Perfect
	wt%	wt%	wt%
C1	60.0%	61.0%	60.5%
C2	30.0%	25.0%	27.5%
C3	10.0%	14.0%	12.0%

Measured Values			
	Field A	Field B	Commingled
	wt%	wt%	wt%
C1	59.99%	61.36%	61.09%
C2	30.1%	25.0%	27.4%
C3	9.9%	13.7%	11.6%

Table 5 – Calculated Field Splits
Field C1 Content Similar

	Field A		Field B	
	Split	Error	Split	Error
Perfect	50.00%		50.00%	
C1	19.84%	-30.16%	80.16%	30.16%
C2	46.42%	-3.58%	53.58%	3.58%
C3	56.08%	6.08%	43.92%	-6.08%

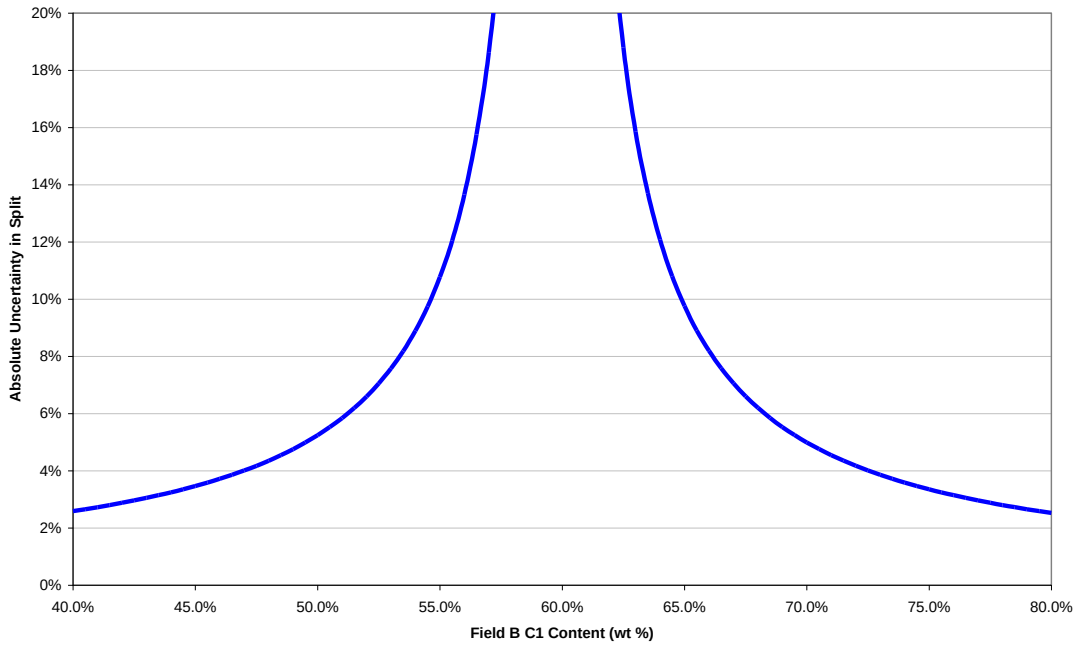
The inferred allocation based on C1 has produced a split to Field A that is grossly below the true value, clearly a poor result, though the splits based on the other components remain reasonable.

If the concentrations of a component are similar in the two Fields then the uncertainty in the results becomes large. The uncertainty in the calculated Field Split for component "i" is given by[1]:

$$U_{SA} = \frac{\sqrt{U_{A,i}^2 (C_i - B_i)^2 + U_{B,i}^2 (C_i - A_i)^2 + U_{C,i}^2 (A_i - B_i)^2}}{(A_i - B_i)^2} \qquad (2)$$

Inspection of (2), reveals that the uncertainty in Field A's split is inversely proportional to the square of the difference of the component's concentration in the two Field streams. Indeed as A_i approached B_i, the uncertainty tends to infinity. Field A's calculated split uncertainty based on C1 is plotted against the C1 content in Field B in Figure 2:

Figure 2 – Uncertainty in Calculated Field A Split as a Function of Field B's C1 Content



At 60%, both Fields' C1 contents are the same and the uncertainty becomes infinite.

2.2 Optimised Split

Is there a methodology, which utilises all the components and therefore maximises the use of the data, that is not be subject to the uncertainty issues encountered with single components, encountered in Example 2 above?

Consider a mass balance across component C1:

$$\delta_{C1} = S_A * A_{C1} + (1 - S_A) * B_{C1} - C_{C1} \qquad (3)$$

[1] (2) is derived in accordance with the Propagation of Uncertainties described in the GUM [3].

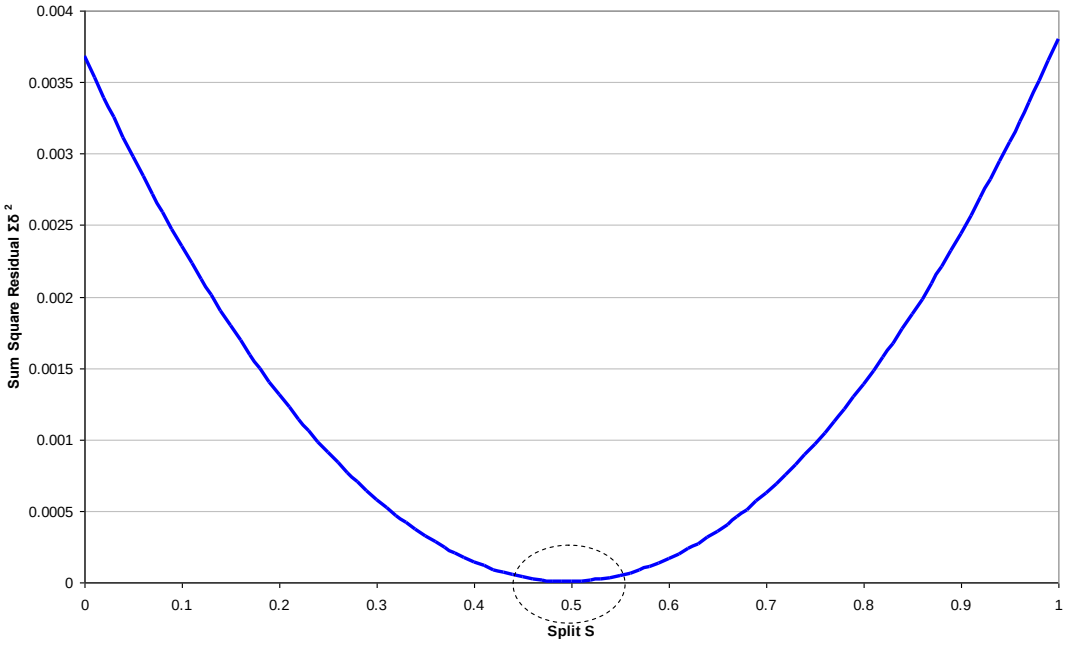
"δ" is termed the residual of the mass balance. Similar balances may be written for all components. As such, unless we have perfectly matching compositions, whatever value we use for the Field split there will always be a violation in the mass balances of some or all the components. Using the data in Example 1, the component mass balances residuals are:

Table 6 – Mass Balance Residuals

S_A	49.58%			
	Field A wt %	Field B wt %	Comm'ed wt%	δ wt%
C1	59.89%	69.88%	64.93%	0.0000%
C2	30.01%	25.17%	27.42%	0.1499%
C3	10.09%	4.96%	7.65%	-0.1499%

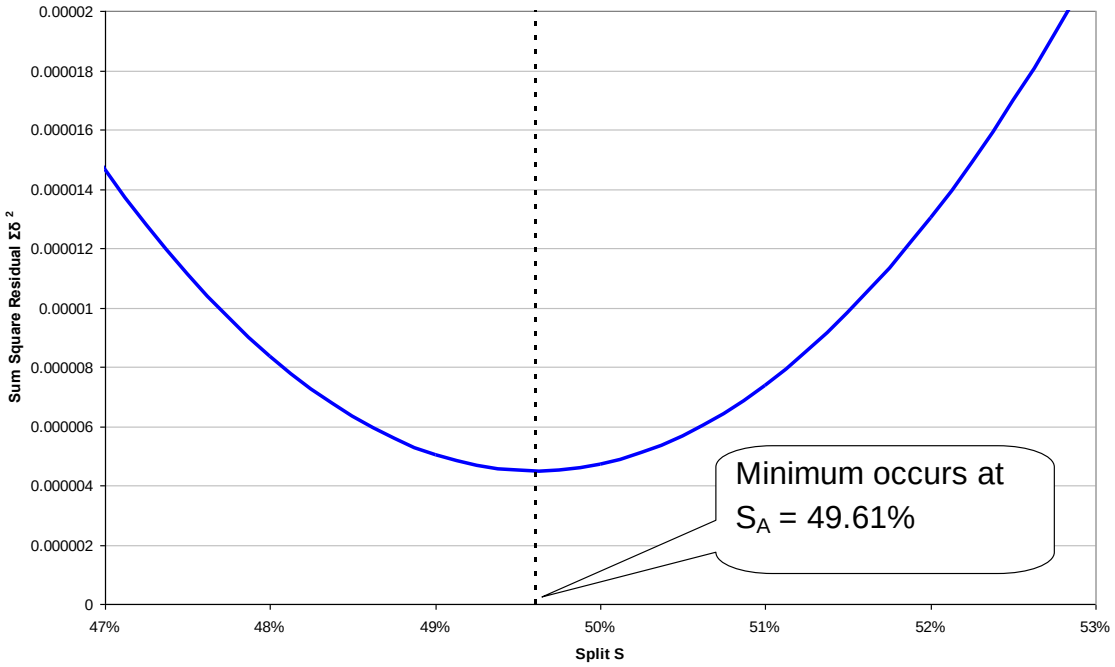
C1's mass balance is satisfied but C2 and C3's aren't. Now is there a value of S_A that would minimise these mass balance residuals (which can be both positive and negative)? The "Optimised Split" methodology, proposed in this paper, determines the value of S_A that produces the minimum of the sum of the squares of these residuals. Figure 3 plots the sum of square of the residuals as a function of Field A split.

Figure 3 – Plot of Sum of Square of Residuals versus Field A Split



The minimum value of the sum of square of the residuals occurs just below 0.5 and is more evidently visible in the amplified version of the plot around this area presented in Figure 4:

Figure 4 – Plot of Sum of Square of Residuals versus Field A Split (Magnified)



This minimum value of can be found analytically by evaluating the derivative of $\Sigma\delta^2$ (with respect to with S_A) when it is equal to zero and thereby the optimised value of S_A:

$$S_A = \frac{\sum_{i=1toN} \left(B_i \left(B_i - A_i - C_i \right) + A_i * C_i \right)}{\sum_{i=1toN} \left(A_i - B_i \right)^2} \qquad (4)$$

Equation (4) is derived in Section 6.2.

At first sight this may appear to be a formidable looking equation. However, it may be broken down into more tractable elements and is easily coded on a spreadsheet or in software code. So applying Equation (4) to Example 1:

Table 7 – Calculation of Optimised Field Split

	Field A	Field B	Commingled	Numerator Terms	Denominator Terms
Cpt (i)	Ai	Bi	Ci	Bi*(Bi -Ai - Ci) + Ai*Ci	$(Ai - Bi)^2$
C1	59.89%	69.88%	64.93%	0.4940%	0.9963%
C2	30.01%	25.17%	27.42%	0.1091%	0.2348%
C3	10.09%	4.96%	7.65%	0.1385%	0.2638%
Sum	100.00%	100.00%	100.00%	0.7416%	1.4949%
Field Split (SA) = 0.7416% / 1.4949% =				**49.61%**	

As can be observed, the inferred split is slightly improved over the individual component based splits. However, this is a single randomly generated snapshot of fictitious data and a more rigorous analysis is required to compare the relative accuracies of the various approaches and this is discussed with more meaningful data in Section 3.3.

2.3 Uncertainty Based Optimised Split

The Optimised Split takes advantage of all the data but it does not account for the relative variation in its accuracy (or uncertainty). We might believe the C1 mass balance residual has a lower relative uncertainty than the other components and therefore should carry more weight in the minimisation. An alternative formulation, based on the Optimised Split, has been developed, termed the "Uncertainty Based Optimised Split". In effect, this approach is a variation on uncertainty based allocation (see [4], [5] and [6]) and employs principles utilised in data reconciliation (see [1] and [2]).

Consider the uncertainty in each residual (δ in equation (3)), which may be calculated using the Propagation of Uncertainties [3] from:

$$U_{\delta,i} = \sqrt{\left(\frac{\partial \delta_i}{\partial A_i}\right)^2 * U_{A,i}^2 + \left(\frac{\partial \delta_i}{\partial B_i}\right)^2 * U_{B,i}^2 + \left(\frac{\partial \delta_i}{\partial C_i}\right)^2 * U_{C,i}^2} \quad (5)$$

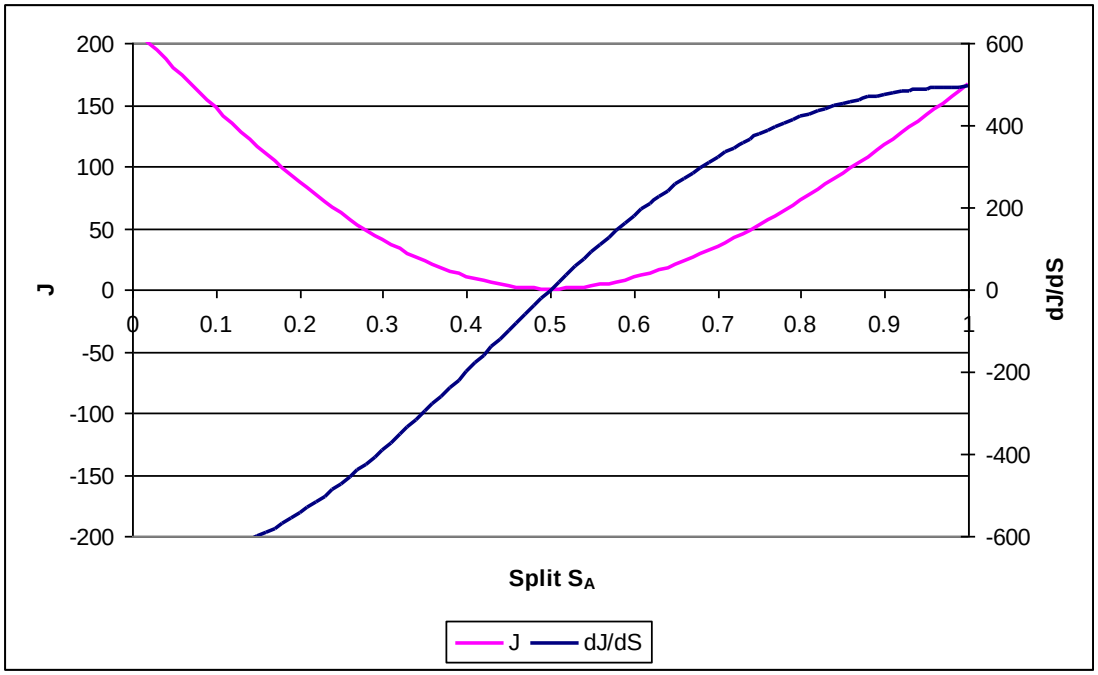
Residuals are then weighted according to their uncertainty. This is performed by dividing δ_i by its uncertainty calculated according to (5). S_A corresponding to the minimum value of the sum of squares of these weighted residuals is then calculated.

The weighted sum of squares of the residuals (J) is given by:

$$J = \sum_{i=1toN} \frac{\delta_i^2}{U_{\delta,i}^2} = \sum_{i=1toN} \frac{\left(S_A * A_i + (1 - S_A) * B_i - C_i\right)^2}{S_A^{\,2} * U_{A,i}^2 + \left(-S_A\right)^2 * U_{B,i}^2 + U_{C,i}^2} \quad (6)$$

The variation of J with S_A is presented in Figure 5:

Figure 5 – Plot of Weighted Sum of Square of Residuals versus Field A Split



Again the optimum value of S_A and can be calculated when the derivative of J with respect to S_A is zero. The equation that describes this is:

$$\frac{dJ}{dS_A} = \sum_{i=1 to N} \left(\frac{2*(S_A*A_i + (1-S_A)*B_i - C_i)*(A_i - B_i)}{S_A^2*U_{A,i}^2 + (-S_A)^2*U_{B,i}^2 + U_{C,i}^2} \right. \\ \left. \frac{(S_A*A_i + (1-S_A)*B_i - C_i)^2*(2*S_A*U_{A,i}^2 - 2*(-S_A)*U_{B,i}^2)}{(S_A^2*U_{A,i}^2 + (-S_A)^2*U_{B,i}^2 + U_{C,i}^2)^2} \right) \quad (7)$$

The derivative is also plotted on Figure 5 against the right hand axis and it equals zero at the minimum value of J. The full derivation of (7) is presented in Section 6.3.

S_A cannot be obtained directly from (7) and must be calculated using numerical techniques such as binary chop, Newton Raphson, secant, etc. The Uncertainty Based Optimised Split (S_A) for Example 1 occurs at 50.12%. This is the closest inferred split to the true value in this example. However, as mentioned Section 2.2 a more rigorous analysis is required to compare the relative accuracies of the various approaches and this is in the next section.

3 APPLICATION WITH REALISTIC DATA

3.1 Introduction

Having illustrated the concepts and the potential for these Chemometric Inferential approaches to allocation in Section 2, we have only examined theoretical models with a small number of components and only for a couple of cases – these were just random snapshots of what is, in any case, fictitious data. In order to determine if the

approaches are viable means of allocation and compare their performance, more meaningful data is required.

As this is a novel allocation methodology, in the development stage, it has not been implemented in any real systems. However, the use of real data from system in which the feed flows are also measured provides a test of how well the inferential methods perform.

First the methods are tested using data obtained from a gas plant in Section 3.2 and with some limited oil analysis data in Section 3.4.

In addition, to perform more exhaustive analysis in Section 3.3, the real data is also utilised to generate a data set with which theoretical tests can be performed that we might expect to encounter in further real systems hence the use of the term "Realistic" in the title.

3.2 Actual Gas Plant Data

All feed streams and product streams (except removed CO_2) are measured both in terms of flow and composition[2]. Using the compositional data alone it is possible to infer the split of the feed using the various inferential techniques described in Section 2 and compare these against the actual metered split of feed flows. Typical compositions of the flows and compositions used are presented in Table 8.

Table 8 – Typical Gas plant Feed and Product Streams

	Feed Streams		
	Field A wt%	Field B wt%	Products wt%
N2	1.13%	1.09%	1.12%
C1	66.81%	72.35%	70.19%
C2	13.54%	11.17%	12.15%
C3	9.85%	7.19%	8.27%
IC4	1.73%	1.38%	1.52%
NC4	3.42%	2.84%	3.00%
IC5	0.99%	0.90%	0.97%
NC5	1.09%	1.10%	1.06%
C6+	1.44%	1.99%	1.71%
Fraction of Feed	53%	47%	

The above data provides an indication of the difference in the feed stream compositions and the relative contribution from each Field. In reality the compositions fluctuate from day to day and the relative flow rates of the feed streams vary more considerably.

Daily data was available for a period exceeding 200 days. The daily compositions were used to infer the contribution from each field and this was compared against the field split calculated from the actual metered flows.

[2] The data had to be conditioned to exclude CO_2, since this was removed in the process and the discharge stream wasn't measured.

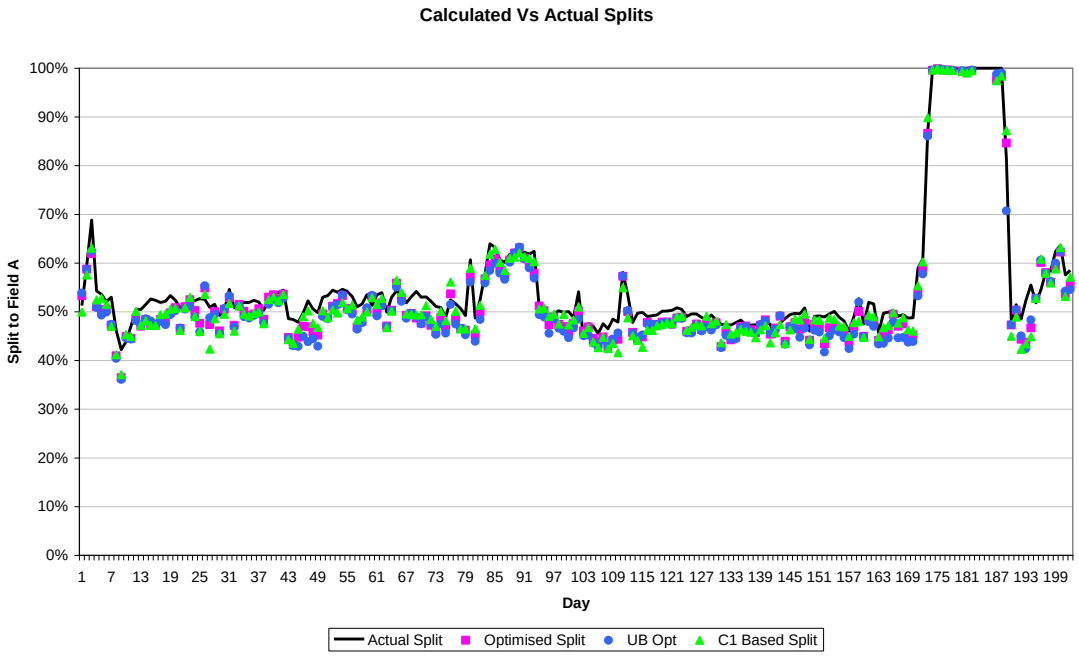
In order to calculate the Uncertainty Based Optimised Splits however, component measurement uncertainties are required. Rather than use published standard uncertainties, which may be appropriate for laboratory conditions, uncertainties based on the daily component mass balance residuals were calculated. By analysing the variance in the mass balance residuals for each component (calculated in accordance with Equation (3)), over the 200 days of data, the relative uncertainty associated with the measurement of that component could be estimated.[3] The values thus calculated are presented in Table 9:

Table 9 – Calculated Component Uncertainties

	Relative Uncertainty ± wt%
N2	2.4%
C1	0.5%
C2	1.4%
C3	3.6%
IC4	6.3%
NC4	4.3%
IC5	11.6%
NC5	9.5%
C6+	30.9%

For the 200 days of data, the inferred Field A splits, based on all the methods described in Section 2, were calculated. The results for the Optimised, Uncertainty Based Optimised and C1 Component based splits are compared with the actual Field A split in Figure 6:

Figure 6 – Inferred versus Actual Field A Splits – Real Gas Plant Data



[3] Using the propagation if uncertainties (as described in [3]), it is possible to infer an average relative uncertainty for a component that is consistent with observed variance in that components residual.

All three methods agree relatively well with the actual splits over the full 200 days. The results data is summarised in Table 10 and this also includes the remaining component based splits.

Table 10 – Inferred Field A Split

		Average Difference from Actual Split wt%	Uncertainty wt%
Optimised Split		-2.2%	4.5%
Uncertainty Based Optimised Split		-2.7%	4.8%
Component Based Splits	N2	1.5%	426.4%
	C1	-2.0%	4.9%
	C2	-3.5%	6.9%
	C3	-2.1%	11.2%
	IC4	3.7%	100.1%
	NC4	-5.7%	30.1%
	IC5	61.2%	174.2%
	NC5	-159.8%	4025.5%
	C6+	-0.1%	76.3%

The difference between the inferred split for each day was calculated for each day and the average of these differences is shown in the table. The uncertainties were calculated based on the standard deviations of the differences between the inferred and actual splits.

The two optimised splits and the C1 component based split performed the best. The minor components performed poorly, notably nC5, this is a result of both Fields having similar concentrations of this component (see Table 8) and is to be expected based on the discussion in Section 2.1.

Most of the more viable methods tend to under-predict the Field A Split by between 2 and 4 %. The uncertainties were calculated as double the standard deviation in the daily calculated differences. The consistent under-prediction possibly indicates a small bias in either: the actual compositional data or the metered flows.

In summary, the differences between the inferred and actual splits and the levels of uncertainties encountered, indicate that the inferential chemometric methods are viable options to allocate the feed contributions.

3.3 Realistic Gas Plant Data

In order to explore the accuracy and robustness of the various methods further, the gas plant data was conditioned to form a theoretical data set. This has the advantage that it allows us to test various scenarios using Monte Carlo methods to generate random uncertainties in the measurements. It also allows numerical uncertainties associated with the methods to be estimated.

In the first simulation the typical feed compositions in Table 8 were combined in a 50:50 mix to provide a perfectly balance product composition. In effect this is analogous to Example 1, but with realistic data. This approach provides a controlled

environment, in which the true measurement values are known (and systematic errors eliminated), that allows allocation uncertainties to be calculated whilst retaining an authentic data set.

The component measurement uncertainties presented in Table 9 were applied to each of the feed and product stream compositions to generate measurement errors randomly. The inferred Field splits were then calculated and compared with the known true 50:50 split. The above random generation of measurement errors was repeated over a number of trials and the results are summarised in Table 11.

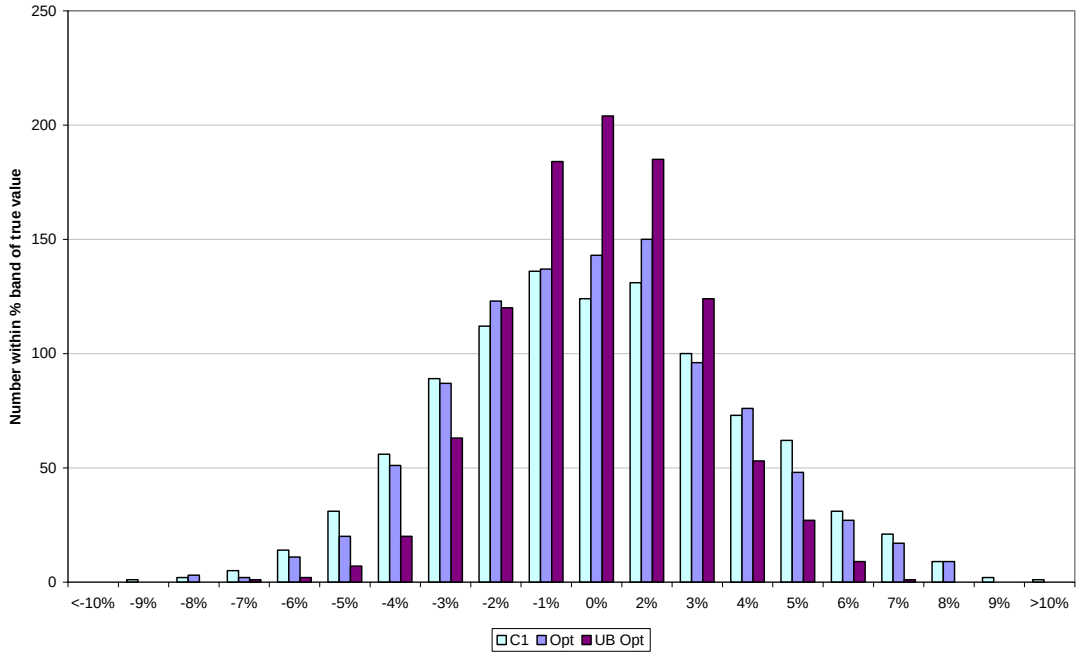
Table 11 – Uncertainties Inferred Allocation Methods, 50:50 Field Split

		Analytical Uncertainty ± wt%	Monte Carlo Uncertainty ± wt%
Optimised Split		4.3%	5.3%
Uncertainty Based Optimised Split			3.8%
Component Based Splits	N2	9.6%	9.9%
	C1	5.2%	5.8%
	C2	7.6%	7.3%
	C3	11.0%	10.6%
	IC4	18.4%	18.8%
	NC4	16.5%	16.9%
	IC5	54.0%	80.0%
	NC5	101.8%	23682.1%
	C6+	190.0%	7396.2%

Both analytical and numerical uncertainties (calculated based on the Monte Carlo simulation results) are provided. Equations for the analytical uncertainties are presented in Section 6.4. The differences between the analytical and numerical uncertainty values for some components is thought to be due to the analytical uncertainty equations becoming highly non-linear, strictly requiring the use of higher order derivatives in their calculation. The results agree reasonably with those encountered with real data presented in Table 10.

The distribution of the differences in the inferred splits from the true splits is presented in the bar chart below for the two optimised and C1 based component splits:

Figure 7 – Distribution of Errors of Inferred Allocation Methods



The above table and figure show that the Uncertainty Based Optimised method produces the most accurate estimates of the Field splits.

A similar analysis was performed with a true Field Split of 90:10 and the results presented in Table 12.

Table 12 – Uncertainties Inferred Allocation Methods 90:10 Field Split

		Analytical Uncertainty ± wt%	Monte Carlo Uncertainty ± wt%
Optimised Split		4.9%	5.3%
Uncertainty Based Optimised Split			2.5%
Component Based Splits	N2	9.2%	9.5%
	C1	6.0%	6.0%
	C2	7.6%	7.6%
	C3	9.9%	9.5%
	IC4	16.4%	16.7%
	NC4	15.5%	16.0%
	IC5	52.5%	7928.5%
	NC5	106.4%	547.9%
	C6+	227.3%	43899.8%

The results are similar to those for the 50:50 case.

To test the impact of the C1 content of the two fields being similar (analogous to Example 2) the 50:50 split was repeated but with Field B's C1 content reduced to 69% and the remaining components increased proportionately. The results are presented in Table 13.

Table 13 – Uncertainties Inferred Allocation Methods 50:50 Field Split, C1 Content Similar

		Analytical Uncertainty ± wt%	Monte Carlo Uncertainty ± wt%
Optimised Split		5.0%	4.7%
Uncertainty Based Optimised Split			2.7%
Component Based Splits	N2	9.6%	9.7%
	C1	86.3%	41608.4%
	C2	5.1%	4.6%
	C3	11.0%	10.0%
	IC4	18.4%	18.7%
	NC4	16.5%	16.4%
	IC5	54.0%	68.8%
	NC5	101.8%	9156.9%
	C6+	190.0%	9341.4%

In agreement with the analysis presented in Section 2, the C1 Component based split becomes very large but the optimised approaches remain robust.

3.4 Oil Samples

The data in examples considered so far has all pertained to gas systems. The techniques are equally applicable to liquid systems also. Consider the compositions of two liquid streams (from Fields X and Y) and the commingled blend of the two presented in Table 14.

Table 14 – Oil Stream Compositions

BP Fraction (°C)	Field X	Field Y	Commingled Product Z
45	4.1%	1.1%	7.3%
60	0.6%	0.5%	2.8%
75	4.2%	0.8%	7.6%
90	6.2%	0.8%	5.9%
105	12.3%	1.7%	13.3%
120	8.0%	1.7%	6.6%
135	10.3%	2.0%	9.4%
150	11.5%	1.6%	8.2%
165	7.9%	2.0%	4.7%
200	13.1%	4.8%	10.2%
250	12.7%	9.0%	7.9%
250+	9.2%	74.3%	16.1%
	100.0%	100.0%	100.0%

These are samples from a real system. The Optimised and component based inferential allocation methods were applied to this data and the results are presented in Table 15.

Table 15 – Inferred Field Split of Oil

Optimised Split		Field X Split 88.7%
Component Based Splits	BP Fraction (oC)	
	45	202%
	60	1359%
	75	200%
	90	95%
	105	110%
	120	77%
	135	90%
	150	67%
	165	45%
	200	65%
	250	-29%
	250+	89.3%

In this instance the true split was unknown, except that it was expected to be within 85% to 90% for Field X. As can be seen from the results the Optimised Split and the 250+ fraction based split both fall in this expected range.

The true split was unknown because of meter failure and this case serves as an example of how Inferential Allocation techniques could serve as secondary means of allocation – this is discussed further in Section 4.

3.5 Three Feed Streams

On of the questions posed in Section 1 was whether it was possible to infer the contribution of more than two Fields. The answer to this is yes and equations are presented below.

The component based split requires the use of two components to infer the split of three fields (in general M + 1 components are required to infer the contribution from M fields). The approach is similar to that developed for the two fields derived in Section 6.1. The final equation for three Fields (A, B, D) mixing to form a commingled stream (C), based on C1 and C2, is:

$$S_A = \frac{(C_{C2} - B_{C2}) * (D_{C1} - B_{C1}) - (C_{C1} - B_{C1}) * (D_{C2} - B_{C2})}{(A_{C2} - B_{C2}) * (D_{C1} - B_{C1}) - (A_{C1} - B_{C1}) * (D_{C2} - B_{C2})} \quad (8)$$

Similarly, the Optimised Split can be used to allocate between three fields and the analogous equation is:

$$S_A = \frac{\sum_{i=1toN}(B_i - C_i)*(D_i - B_i)*\sum_{i=1toN}(D_i - B_i)*(A_i - B_i) - \sum_{i=1toN}(B_i - C_i)*(A_i - B_i)*\sum_{i=1toN}(D_i - B_i)^2}{\sum_{i=1toN}(A_i - B_i)^2*\sum_{i=1toN}(D_i - B_i)^2 - \sum_{i=1toN}(D_i - B_i)*(A_i - B_i)^2} \quad (9)$$

The gas plant, from which the data presented in Section 3.2, also experienced periods when three fields were producing. Typical compositions and Field splits are presented in Table 8.

Table 16 – Typical Gas Plant Feed (3 Fields) and Product Streams

	Feed Streams			
	Field A wt%	Field B wt%	Field C wt%	Products wt%
N2	1.13%	1.09%	1.46%	1.12%
C1	66.81%	72.35%	81.86%	70.19%
C2	13.54%	11.17%	9.17%	12.15%
C3	9.85%	7.19%	3.18%	8.27%
IC4	1.73%	1.38%	0.97%	1.52%
NC4	3.42%	2.84%	1.18%	3.00%
IC5	0.99%	0.90%	0.60%	0.97%
NC5	1.09%	1.10%	0.44%	1.06%
C6+	1.44%	1.99%	1.15%	1.71%
Fraction of Feed	48%	46%	6%	

The figures below plot the inferred allocation to each of the three fields using the two methods described in Equations (8) and (9).

Figure 8 – Inferred versus Actual Field A Splits – Real Gas Plant Data for Three Fields

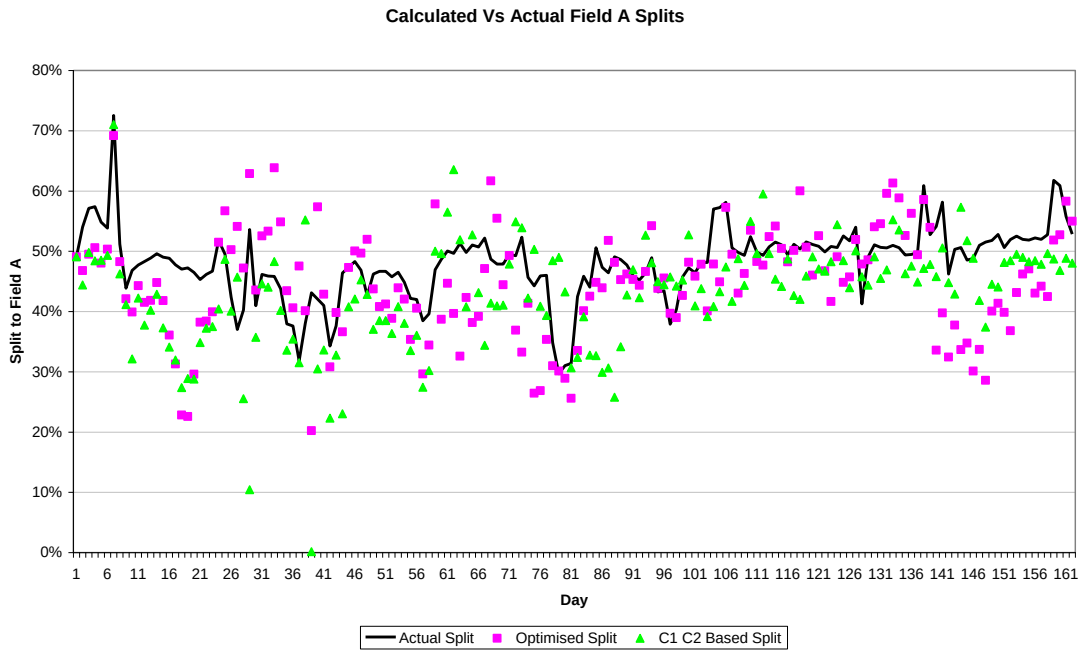


Figure 9 – Inferred versus Actual Field B Splits – Real Gas Plant Data for Three Fields

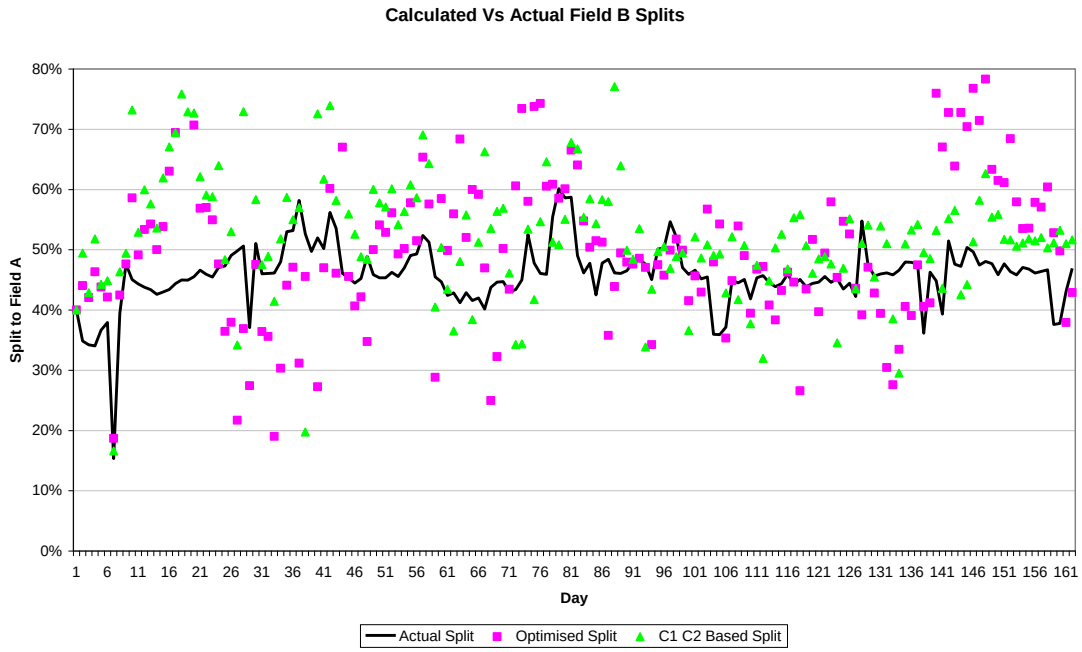
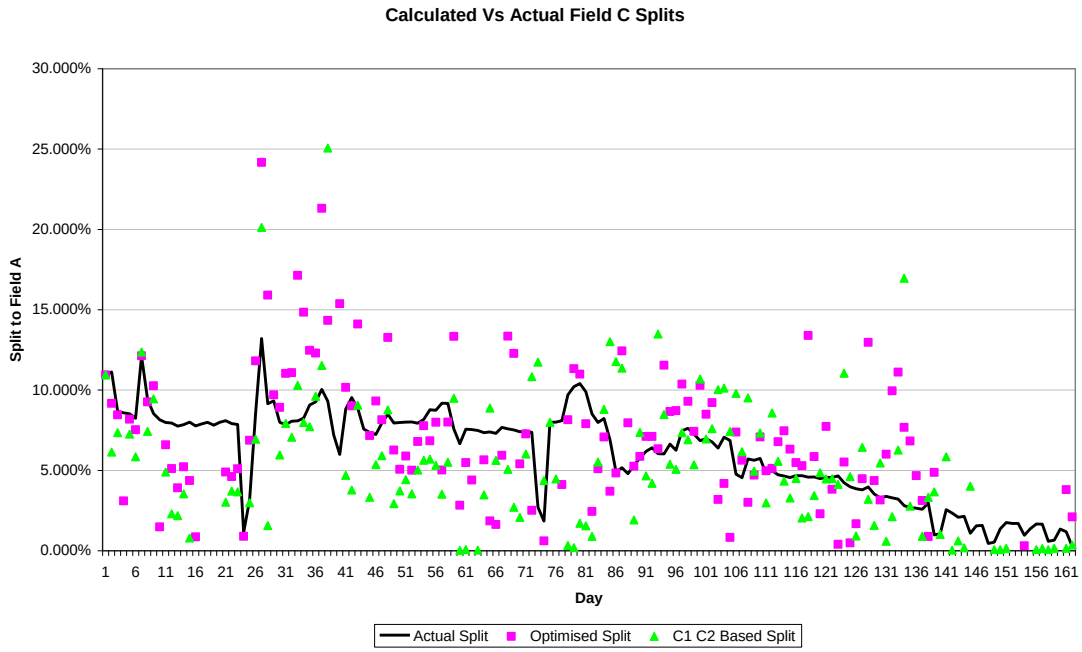


Figure 10 – Inferred versus Actual Field C Splits – Real Gas Plant Data for Three Fields



There is more scatter in the predicted splits and a reduction in accuracy compared to when two fields are present. The average deviation between the predicted and actual split and calculated uncertainty (based on twice the standard deviation) are presented in Table 17.

Table 17 – Summary of Inferred Field Splits vs Actual

	Optimsed Split		C1, C2 Based Split	
	Average Difference from Actual Split wt%	Uncertainty ± wt%	Average Difference from Actual Split wt%	Uncertainty ± wt%
Field A	-3.5%	16.6%	-5.2%	16.7%
Field B	4.4%	25.8%	7.3%	25.2%
Field C	-0.9%	9.5%	-2.1%	11.8%

Overall, the Optimised Split performs slightly better than the Component Based split.

3.6 Compositional Drift

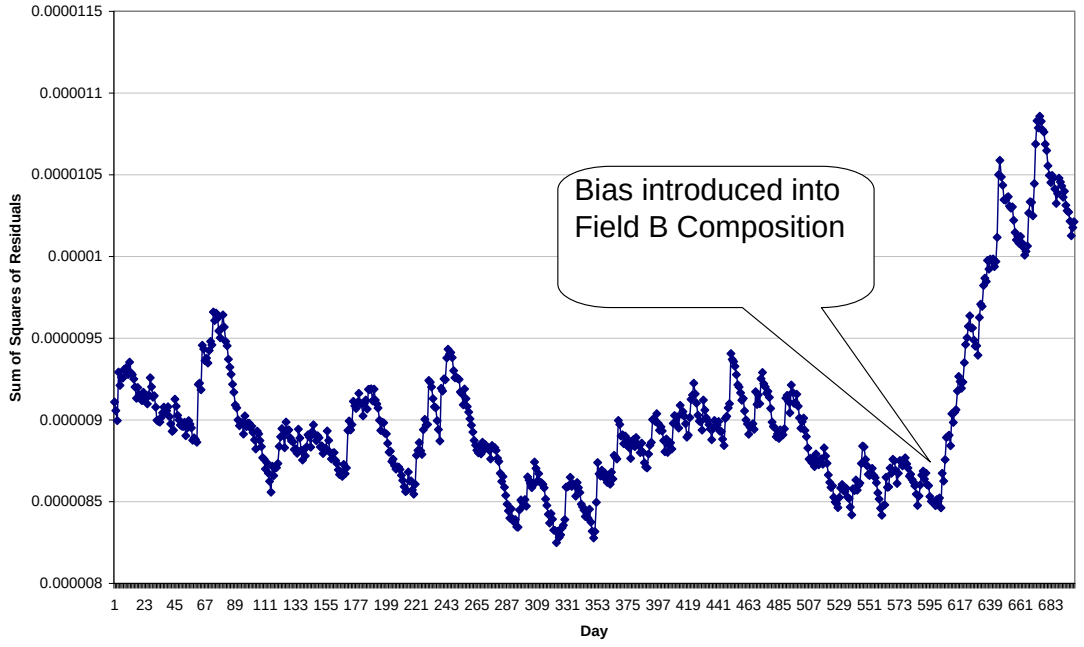
Another question posed in Section 1 was: "Would we know if one of the assumed constant compositions starts to drift?"

A feature of the optimised split and uncertainty based optimised split is that they provide a methodology to determine if the commingled mixture is genuinely a mix of our two assumed feed compositions.

A relatively easy metric to monitor to determine if there has been compositional drift is to plot the sum of the squares of the component mass balance residuals as given by equations (15) and (6). Plotting this metric each day provides a figure that may be monitored and if it increases then this would indicate compositional drift.

As an example, using the 50:50 split case described in Section 3.3, a bias was introduced part way through the simulation. A plot of the sum of squares of the residuals for the Optimised Split, filtered to smooth the data, is presented in Figure 11:

Figure 11 – Plot of Sum of Squares of Residuals 50:50 Split, Compositional Bias Introduced on Day 600



As can be seen the point at which the bias occurred is evident.

Though a useful metric to monitor from a practical viewpoint in some putative "imagined" allocation system, there are more rigorous approaches. The above approach does not provide any information on what is a significant level of the sum of squares number is. It simply looks for a rise in the value and infers that something has changed.

Alternatively, it is possible to determine what a significant value for the sum of squares figure is, above which systematic errors are indicated. This would provide a test statistic to determine, given the uncertainties in the compositions, whether the residuals could have arisen reasonably by chance or whether there is a systematic shift. Such techniques are used to detect gross errors in data reconciliation [1] and [2].

4 APPLICATIONS

Having shown the potential of these inferred allocation methodologies, the final question to be answered is: what use can they be for allocation?

For the case of the gas plant described above these techniques would be of limited use since all the streams are measured. This data was merely used to test the methodologies. However for the case of the oil sample example the contribution of the two streams was not known and this is where the inferential techniques could be usefully employed to allocate the product.

To be used as a primary method of allocation the confidence in the measured compositions would have to be high. Possible applications could include systems where the feed streams are known to have stable compositions.

The methods could be used as a secondary, back-up, means of allocation where there is a risk of the primary measurement failing. For example, in a system relying on a multi-phase flow meter, especially when located subsea, the inferential technique could be used as an alternative method of allocation should the meter fail.

Indeed even if the meter was working correctly, the secondary inferential allocation could be used as a quality control metric. These techniques provide alternative methods of allocation which can be used to monitor how well the primary measurement and allocation systems are performing.

5 CONCLUSIONS

This is paper has presented three methods of inferring allocated quantities from compositional data alone:

- Single component based split, using ratios of single components
- Optimised Split, calculated by minimising component mass balance residuals
- Uncertainty Based Optimised Split, calculated by minimising the uncertainty in component mass balance residuals

These have been demonstrated to be viable means of allocation, using both theoretical examples and real data.

Their accuracy is dependent on the relative flows of the streams and the dissimilarity of their compositions.

The optimised split approaches were generally found to be the most robust and accurate.

Though mainly analysed in terms of allocating between two Fields the methods have been demonstrated to work with three fields with diminished accuracy.

The mass balance residuals provide a metric with which to monitor the consistency of the compositional data and thereby detect systematic changes.

A number of possible applications have been suggested: as a primary allocation method, a secondary back-up method to mitigate the impact loss of metering on the allocation system and as a quality assurance check.

6 MATHEMATICAL DERIVATIONS OF EQUATIONS PRESENTED

6.1 Field Split Based on a Single Component

Consider a mass balance for component i:

$$F_A * A_i + F_B * B_i = F_C * C_i \tag{10}$$

Contribution or split of A is given by:

$$S_A = \left(\frac{F_A}{F_A + F_B} \right) \tag{11}$$

S_B can be similarly defined and noting that $F_A + F_B = F_C$, (10) becomes:

$$S_A * A_i + S_B * B_i = C_i \tag{12}$$

Also since the splits must sum to 1, $S_A + S_B = 1$, and rearranging to obtain S_A (13) becomes:

$$S_A = \frac{(C_i - B_i)}{(C_i - A_i)} \tag{13}$$

6.2 Optimised Field Split

Consider a mass balance for component i:

$$\delta_{C1} = S_A * A_i + (1 - S_A) * B_i - C_i \tag{14}$$

The sum of the squares these residuals, K, is calculated:

$$K = \sum_{i=1toN} \left(S_A * A_i + (1 - S_A) * B_i - C_i \right)^2 \tag{15}$$

K is differentiated with respect to S_A:

$$\frac{dK}{dS_A} = \sum_{i=1toN} 2 * \left(S_A * A_i + (1 - S_A) * B_i - C_i \right) * \left(A_i - B_i \right) \tag{16}$$

The minimum value of K is obtained by setting the derivative to equal zero and rearranging to obtain S_A:

$$S_A = \frac{\sum_{i=1toN} \left(B_i \left(B_i - A_i - C_i \right) + A_i * C_i \right)}{\sum_{i=1toN} \left(A_i - B_i \right)^2} \tag{17}$$

6.3 Uncertainty Based Optimised Field Split

Consider a mass balance for component i:

$$\delta_i = S_A * A_i + (1 - S_A) * B_i - C_i \tag{18}$$

The uncertainty in the residual δ_i for component i is calculated from the following:

$$U_{\delta,i} = \sqrt{\left(\frac{\partial \delta_i}{\partial A_i} \right)^2 * U_{A,i}^2 + \left(\frac{\partial \delta_i}{\partial B_i} \right)^2 * U_{B,i}^2 + \left(\frac{\partial \delta_i}{\partial C_i} \right)^2 * U_{C,i}^2} \tag{19}$$

The partial derivatives are:

$$\left(\frac{\partial \delta_i}{\partial A_i} \right) = S_A; \quad \left(\frac{\partial \delta_i}{\partial B_i} \right) = \left(1 - S_A \right); \quad \left(\frac{\partial \delta_i}{\partial C_i} \right) = -1 \tag{20}$$

The residuals are then weighted according to their uncertainty. This is performed by dividing δ_i by its uncertainty calculated according to (19).

$$\frac{\delta_i^2}{U_{\delta,i}^2} = \frac{\left(S_A * A_i + (1 - S_A) * B_i - C_i \right)^2}{S_A^2 * U_{A,i}^2 + (1 - S_A)^2 * U_{B,i}^2 + U_{C,i}^2} \tag{21}$$

The weighted sum of squares of the residuals (J) is given by:

$$J = \sum_{i=1toN} \frac{\delta_i^2}{U_{\delta,i}^2} = \sum_{i=1toN} \frac{(S_A * A_i + (1-S_A) * B_i - C_i)^2}{S_A^{\,2} * U_{A,i}^2 + (1-S_A)^2 * U_{B,i}^2 + U_{C,i}^2} \tag{22}$$

The optimum value of S_A corresponding to the minimum value of J can be calculated when the derivative of J with respect to S_A is zero. The equation that describes this is:

$$\frac{dJ}{dS_A} = \sum_{i=1toN} \left(\frac{2 * (S_A * A_i + (1-S_A) * B_i - C_i) * (A_i - B_i)}{S_A^{\,2} * U_{A,i}^2 + (1-S_A)^2 * U_{B,i}^2 + U_{C,i}^2} \right. \\ \left. \frac{(S_A * A_i + (1-S_A) * B_i - C_i)^2 * (2 * S_A * U_{A,i}^2 - 2 * (1-S_A) * U_{B,i}^2)}{(S_A^{\,2} * U_{A,i}^2 + (1-S_A)^2 * U_{B,i}^2 + U_{C,i}^2)^2} \right) \tag{23}$$

S_A cannot be obtained directly from (23) and must be calculated using numerical techniques such as binary chop, Newton Raphson, secant, etc.

6.4 Inferential Method Uncertainties

Single Component Based Split

From equation (13), the partial derivatives are calculated:

$$\left(\frac{\partial S_A}{\partial A_i}\right) = \frac{B_i - C_i}{(A_i - B_i)^2}; \quad \left(\frac{\partial S_A}{\partial B_i}\right) = \frac{C_i - A_i}{(A_i - B_i)^2}; \quad \left(\frac{\partial S_A}{\partial C_i}\right) = \frac{1}{A_i - B_i} \tag{24}$$

And the uncertainty is given by:

$$U_{SA} = \sqrt{\frac{(A_i - B_i)^2 * U_{A,i}^2 + (C_i - A_i)^2 * U_{B,i}^2 + (B_i - Z_i)^2 * U_{C,i}^2}{(A_i - B_i)^2}} \tag{25}$$

Optimised Split

Equation (17), has to be re-expressed in terms of independent variables and hence N-1 components. One component is not independent since the component mass fractions must sum to 1.

$$S_A = \frac{\sum_{i=1toN-1} B_i (B_i - A_i - C_i) + A_i * C_i + \left(1 - \sum_{i=1toN-1} B_i\right)\left(\sum_{i=1toN-1} A_i + \sum_{i=1toN-1} C_i - \sum_{i=1toN-1} B_i - 1\right) + \left(1 - \sum_{i=1toN-1} A_i\right) * \left(1 - \sum_{i=1toN-1} C_i\right)}{\sum_{i=1toN-1} (A_i - B_i)^2 + \left(\sum_{i=1toN-1} B_i - \sum_{i=1toN-1} A_i\right)^2} \tag{26}$$

Abbreviating (26) to:

$$S_A = \frac{Num}{Denom} \tag{27}$$

Defining,

$$A_d = 1 - \sum_{i=1toN-1} A_i; \tag{28}$$

And similar terms for B and C, the partial derivatives of S_A are:

$$\left(\frac{\partial S_A}{\partial A_i}\right) = \frac{C_i - B_i - C_d + B_d}{Denom} + \frac{2 * Num * \left(B_i - A_i - B_d + A_d\right)}{Denom^2} \tag{29}$$

$$\left(\frac{\partial S_A}{\partial B_i}\right) = \frac{2 * B_i - A_i - C_i - 2 * B_d + A_d + C_d}{Denom} - \frac{2 * Num * \left(B_i - A_i - B_d + A_d\right)}{Denom^2} \tag{30}$$

$$\left(\frac{\partial S_A}{\partial C_i}\right) = \frac{A_i - B_i - A_d + B_d}{Denom} \tag{31}$$

With these partial derivatives for each of N-1 and the associated absolute component uncertainties the uncertainty of S_A calculated using the propagation of uncertainties.

NOTATION

A	Mass fraction Field A
B	Mass fraction Field B
C	Mass fraction commingled stream
D	Mass fraction Field D
Denom	Denominator of Equation 26
F	Stream mass flow
J	Weighted sum of squares of residuals
K	Sum of squares of residuals
M	Number of meters
N	Number of components
Num	Numerator of Equation 26
S	Field split
U	Absolute uncertainty
δ	Mass balance residual

Subscripts

A	Field A
B	Field B
C	Commingled stream
C1	Component C1
C2	Component C2
d	Component as defined in Equation 28
D	Field D
i	component
SA	Field A split

7 REFERENCES

[1] Data Processing and Reconciliation for Chemical Process Operations, J Romagnoli and M Sanchez, Published by Academic Press (2000), ISBN 0-12-594460-8.

[2] Data Reconciliation and Gross Error Detection: An Intelligent Use of Process Data, S Narasimhan and C Jordache, Published by Gulf Publishing (2000), ISBN 0-88415-255-3

[3] Guide to the Expression of Uncertainty in Measurement, International Organisation for Standardisation, ISO/IEC Guide 98:1995.

[4] Use of Subsea Wet Gas Flowmeters in Allocation Measurement Systems, API RP 85, First Edition, August 2003.

[5] Determination of Measurement Uncertainty for the Purpose of Wet Gas Hydrocarbon Allocation Robert A. Webb, BP, Winsor Letton, Letton-Hall Group. Martin Basil, FLOW Ltd North Sea Flow Measurement Workshop, 22nd -25th October 2002.

[6] Experiences in the Use of Uncertainty Based Allocation in a North Sea Offshore Oil Allocation System, Phillip Stockton & Alan Spence, Smith Rea Energy Ltd, Production and Upstream Flow Measurement Workshop, Houston, 12-14 February 2008.