AADE-15-NTCE-15

# Statistical Analysis of the Variation Inherent in the Measurement and Calculation of Drilling Fluid Properties

Anne Beckly, M-I SWACO, a Schlumberger company



## Abstract

During a drilling fluid assessment, mud-check numbers carry a great deal of importance. Verifying the drilling fluid specifications is necessary for day-to-day operations and in tender evaluations, where clients make decisions involving multi-million dollar contracts. It is important for all concerned parties to understand the errors involved and the likely variance (the amount the readings spread). Values quoted for the properties of a drilling fluid can never be absolute, because errors and variances associated with any measured and calculated values are possible.

This paper studies the statistical variance of readings taken on both water-based and oil-based drilling fluids. A group of competent people, using different sets of equipment, carried out the work. This exercise increased the variation between measurements and maximized the variance to provide a more realistic match to real-world applications. The variances seen in different mud-check properties are not the same. Some experiments have a lower inherent degree of error than others. This paper attempts to identify the important area of errors and variances within the mud-check numbers.

## Introduction

Most drilling fluid assessments involve few, if any, duplicate readings, due to time constraints and the allocation of resources. This can lead to misleading results. How can we be sure that the drilling fluid properties quoted are correct? There can be errors associated with the equipment used, with the samples selected and with the personnel carrying out the tests. These errors can be minimized, but not completely eliminated, by ensuring that equipment is calibrated to the required standard, all procedures are followed, and that the personnel carrying out the testing are competent. There can still be errors associated with calibrated equipment as they are only calibrated to a required tolerance. Some readings will have a greater degree of inherent errors, be it due to equipment or other factors. It can be assumed that these errors are likely to be consistent only for measurements taken under exactly the same conditions, for example by the same person, using the same equipment, on the same day.

For rheology measured using a Fann 35 viscometer the accepted tolerance for the plastic viscosity (PV) and the yield point (YP) is ±2 and for the 6rpm, 3rpm dial readings and gel strengths ±1. The manufacture's specification for this equipment is that the 600rpm reading is within ±3 and the 300rpm reading is within ±2. The equipment is calibrated monthly to these specifications and adjusted as necessary. However this means that as PV and YP values are calculated from the 600 and 300rpm readings, we may be working at, or slightly beyond, the limits of accuracy for the equipment to obtain the required precision.

This paper considers the variation seen in readings taken of properties of the same drilling fluid formulations tested in the same lab at different times, by different competent personnel, using different equipment. The data is from lab formulations not from field muds. The most commonly used statistical property quoted is the central value of a data set, this is usually the mean or average value, but can be the median or mode. The variance and standard deviation are both properties, which can be used to show the degree of spread within a data set. The standard deviation is the square root of the variance and can be used to calculate other statistical properties to assess the data. The symmetry around the central value, or skewness, and the peakedness, or kurtosis, are further characteristics used to describe a distribution

## Analysis

### *Fluid Loss Data*

Fluid loss data was obtained from fluid loss measurements on oil-based mud (OBM), in experiments using ceramic discs at 80°C, and 500 psi differential pressure for approximately 16 hours. A total of 305 fluid loss results were obtained from multiple batches of the same formulation of a robust low toxicity oil-based mud (LTOBM). Tests were carried out by at least two different competent people on different days using at least six different high-pressure high-temperature (HPHT) fluid loss cells. Fluid loss was measured to the nearest 0.5 mL. The results are shown graphically in **Figure 1**.

The mean for the data set, shown in **Figure 1**, is 6.9 mL with a standard deviation of 0.6 mL. The results would seem to be close to a normal distribution (the curve superimposed on the histogram in Figure 1). The large sample size, with a good fit to the normal distribution means that a normal distribution model could be appropriate to use to characterize the data. **Figure 2** shows the actual readings and those expected for a normal distribution of 305 values, with mean and standard deviation the same as that calculated for the actual data.

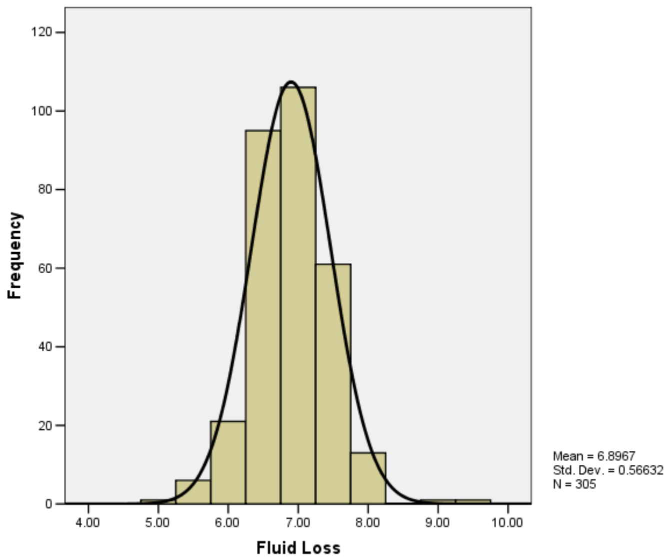


**Figure 1. Histogram of the results of multiple fluid loss experiments on a LTOBM.**

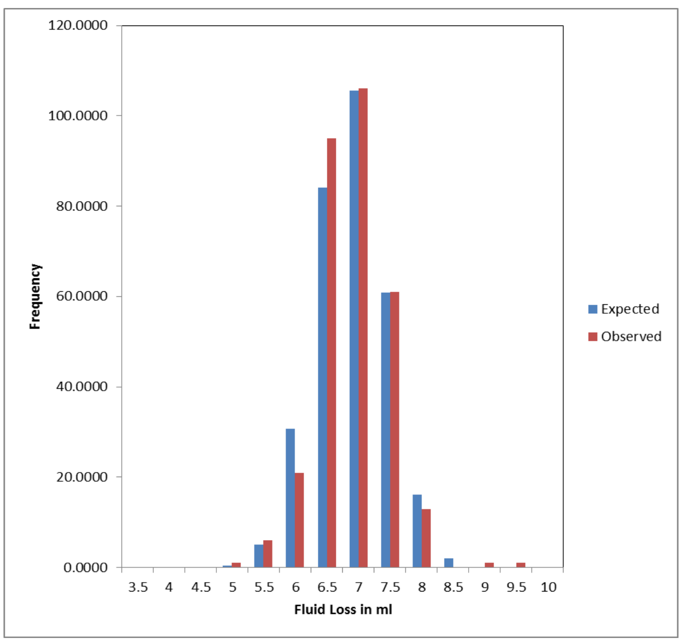


**Figure 2. Histogram of the 305 results of the fluid loss experiments on a LTOBM and 305 calculated from a normal distribution with mean and standard derivation the same as the observed readings.**

This fluid loss data has a relatively small range with the lowest reading of 5mL and the highest of 9.5 mL. However as **Figures 1 and 2** demonstrate, the lowest values are still within the expected range for a normal distribution. The probability of a value of 5 mL is 1.3% while for 7 mL the probability is 34.6%. The highest value, 9.5 mL, is slightly higher than would be expected assuming a normal distribution with a probability of 0.001%. For a normal distribution 95% of the data is expected to lie within ±2 standard deviations of the mean (i.e between 5.76 and 8.04 mL).

The distribution of the data is also slightly more peaked than might be expected as demonstrated by the Q-Q plot in **Figure 3**. There is slight deviation from the straight lines in the high and low values. The Q-Q plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other. In this case it is used to see if the normal distribution is a reasonable fit. The theoretical quantiles for a normal distribution are plotted against the actual sample quantiles. The asymmetry or skewness ($\sqrt{\beta_1}$) of the distribution has been calculated as 0.248 so $\beta_1$ is 0.161. A normal distribution is defined as having a skewness and $\beta_1$ value of zero. The low positive value of the skewness means that the distribution is slightly right skewed, but very close to a normal distribution. The other property to consider is the peakedness of the distribution. The quantity $\beta_2$ is a relative measure of kurtosis and for a normal distribution has a value of 3. The value calculated from the fluid loss data was 3.138. This means that the distribution is slightly more peaked than a normal distribution. There is some slight deviation from a straight line in the Q-Q plot and a slight difference in the values of $\beta_1$ and $\beta_2$, but overall a normal distribution would appear to be a good model for this data.

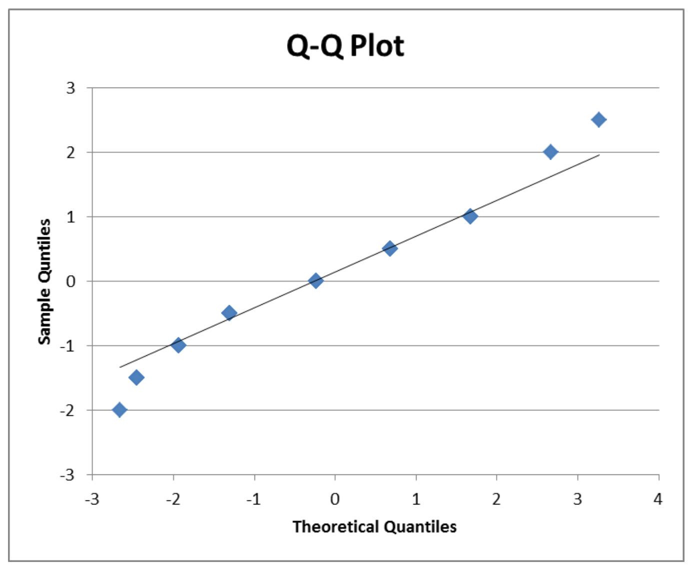


**Figure 3. Q-Q Plot for fluid loss data**

If only one reading was taken, then the statistically most likely value would be 7 mL, but it is probable that the reading would be within the range from 5 to 9 mL (assuming that measurements are rounded to the nearest 0.5 mL). In other words, a single reading will not necessarily give the most likely result.

### *Oil-Based Mud and Water-Based Mud Data*

Mud data was obtained as part of our competency procedure. The same, LTOBM formulation and a polymer-glycol water-based mud (WBM) were mixed by different people, at different times, over a period of a number of years, using different equipment and samples. There are insufficient samples to determine if the data collected can be modelled by a normal distribution, but for all the readings and calculated values, there is a spread in the data.

Descriptive statistics (i.e., the mean and standard deviation) determined for the measured Fann 35 dial readings and calculated PV and YP rheological data of OBM, before hot rolling (BHR) for 16 hours at 250°F, are shown in **Figure 4. Figure 5** shows the values for the data obtained after hot rolling (AHR). The apparent viscosity, PV and YP are all calculated from the 600 and 300rpm dial readings; therefore, the degree of variation of their values should be tied to the degree of variation seen in the 600 and 300rpm readings.

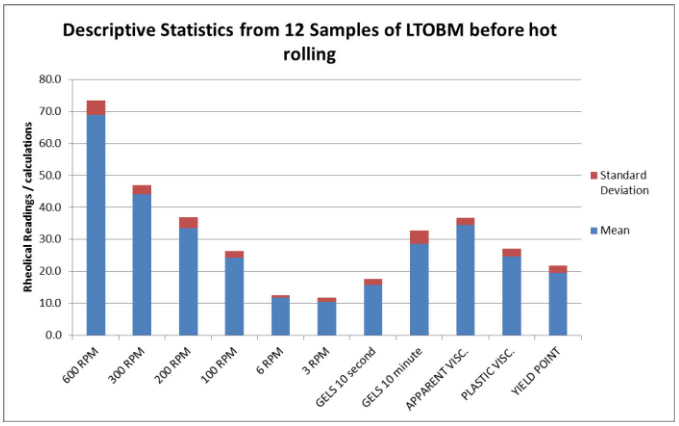


**Figure 4. Descriptive statistics from 12 samples of LTOBM before hot rolling.**

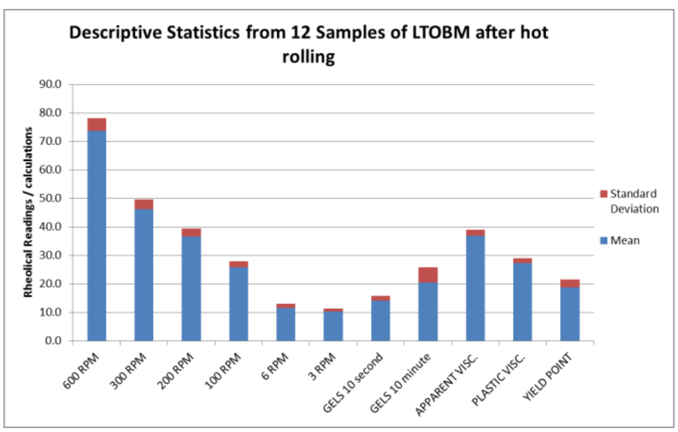


**Figure 5. Descriptive statistics from 12 samples of LTOBM after hot rolling.**

The standard deviation is the degree of variation or dispersion from the mean seen within the values used to calculate it and is the square root of the variance. The standard deviation and variance along with the difference between the maximum and minimum values will indicate how much variation there is within the data for a given measurement. As would be expected, the variations, given by the standard deviation, for the 6rpm, 3rpm dial readings and 10second gel strength are smaller than for the other readings; the actual readings in absolute numbers are lower. The standard deviation for the apparent viscosity is half the value for the 600rpm reading. This is to be expected as the apparent viscosity is calculated as half the 600rpm reading.

The standard deviation of the 600rpm reading tends to have the largest magnitude. The 6rpm and 3rpm readings tend to have the lowest standard deviation. The BHR sample readings have lower standard deviations than those for AHR samples. The calculated YP standard deviation for AHR samples is greater than that for the PV; however, both are lower than the standard deviations for the 600rpm and 300rpm measurements.

The descriptive statistics for the AHR measurements are shown in **Table 1**. The upper and lower limits are the values two standard deviation from the mean. The maximum and minimum values for the data lie between these upper and lower limits, apart from for the YP where the maximum value is slightly above the upper limit.

**Table 1. Descriptive Statistics for LTOBM After Hot Rolling.**

| Rheology measurement | Mean | SD | Var | Max | Min | Upper limit | Lower limit |
|---|---|---|---|---|---|---|---|
| 600rpm | 73.7 | 4.4 | 19.7 | 81 | 68 | 83 | 65 |
| 300rpm | 46.3 | 3.3 | 10.9 | 53 | 42 | 53 | 40 |
| 200rpm | 36.8 | 2.7 | 7.1 | 42 | 33 | 42 | 31 |
| 100rpm | 25.8 | 2.2 | 4.8 | 31 | 23 | 30 | 21 |
| 6rpm | 11.6 | 1.4 | 1.9 | 14 | 9 | 14 | 9 |
| 3rpm | 10.3 | 1.1 | 1.1 | 12 | 8 | 12 | 8 |
| Gels 10 sec | 14.2 | 1.7 | 3.1 | 17 | 12 | 18 | 11 |
| Gels 10 min | 20.6 | 5.1 | 26.4 | 31 | 15 | 31 | 10 |
| PV | 27.4 | 1.7 | 2.8 | 30 | 25 | 31 | 24 |
| YP | 18.8 | 2.8 | 7.8 | 25 | 16 | 24 | 13 |

Though the mean values have changed, the standard deviations are similar for the respective rheology values BHR and AHR. However, the addition of contamination either in the form of 35 lb/bbl HMP (a clay used to replicate drill solids) or 10% by volume seawater to the sample AHR fluid caused an increase in both the mean and standard deviation. For HMP contamination, the mean and standard deviations are approximately doubled. The increase in the mean and standard deviation are not as large for seawater contamination.

The mean and standard deviation for the rheology data of these batches of WBM, before hot rolling at 250 °F for 16 hours are shown in **Figure 6**. **Figure 7** shows the values for the readings taken and calculated after hot rolling.

.

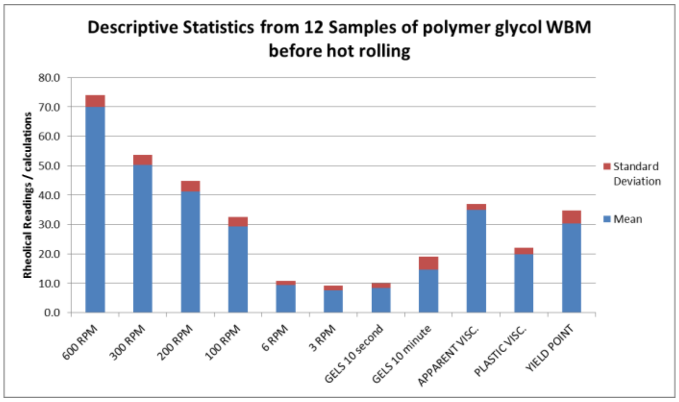


**Figure 6. Descriptive statistics from 12 samples of polymer-glycol WBM BHR.**

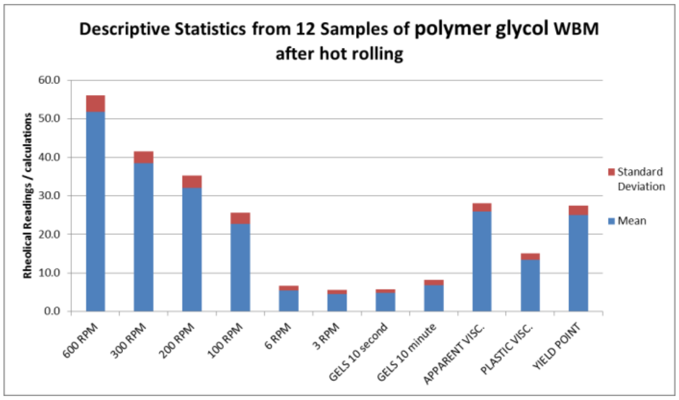


**Figure 7. Descriptive statistics from 12 samples of polymer-glycol WBM AHR.**

The analysis of the data for the WBM testing gives similar standard deviations as those for the respective tests on OBM. As would be expected the 600rpm readings gave the highest values of standard deviation with the 6rpm, 3rpm, and 10second gel readings having the lowest values

**Table 2. Descriptive Statistics - Polymer-Glycol WBM After Hot Rolling.**

| Rheology measurement | Mean | SD | Var | Max | Min | Upper limit | Lower limit |
|---|---|---|---|---|---|---|---|
| 600rpm | 51.8 | 4.4 | 19.3 | 59 | 46 | 61 | 43 |
| 300rpm | 38.4 | 3.1 | 9.4 | 43 | 33 | 45 | 32 |
| 200rpm | 32.1 | 3.1 | 9.7 | 36 | 27 | 38 | 26 |
| 100rpm | 22.8 | 2.9 | 8.2 | 28 | 18 | 28 | 17 |
| 6rpm | 5.5 | 1.2 | 1.4 | 8 | 4 | 8 | 3 |
| 3rpm | 4.5 | 1.2 | 1.4 | 7 | 3 | 7 | 2 |
| Gels 10 sec | 4.9 | 0.8 | 0.6 | 6 | 4 | 7 | 3 |
| Gels 10 min | 6.8 | 1.4 | 2.0 | 9 | 5 | 10 | 4 |
| PV | 13.3 | 1.8 | 3.3 | 16 | 11 | 17 | 10 |
| YP | 25.1 | 2.5 | 6.1 | 28 | 20 | 30 | 20 |

The descriptive statistics for after hot rolling measurements are shown in **Table 2**. All of the maximum and minimum values lie on or between the upper and lower limit. This suggests that even when there are not enough measurements to accurately predict a model to fit the data, it is possible to put expected limits to the variation of the data.

The addition of contamination, as described above for the OBM, to the AHR mixture has less effect on the means, but similar effects on the standard deviations. The means for 600, 300, 200, 100rpm dial readings and the 10second and 10minute gel strengths are slightly lower for the contaminated samples, while the means for the 6 and 3rpm readings are slightly higher. Statistics indicate the biggest disparity caused by the addition of contaminates is on the spread of results obtained. There is larger variation (indicated by the standard deviation) seen within the rheological numbers with the addition of contamination.

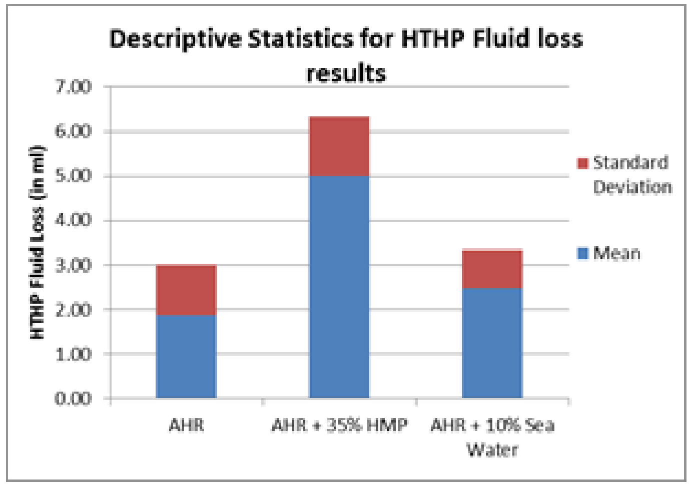


**Figure 8. Descriptive statistics for HPHT fluid loss results from different batches of LTOBM.**

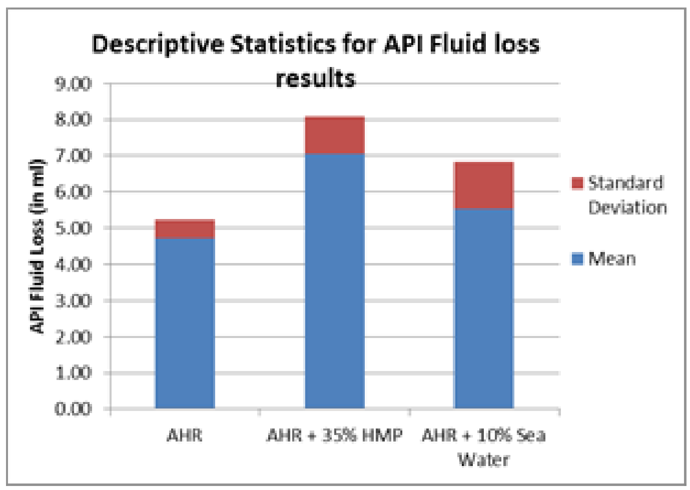


**Figure 9. Descriptive statistics for API fluid loss results from different batches of a polymer-glycol WBM.**

Studying the fluid loss results for these two types of muds, as seen in **Figures 8** and **9**, reveals again a difference in the spread of results obtained. The introduction of contamination increases the mean values and generally the standard deviation. This increase in the standard deviation indicates an increase in the spread of values obtained.

The 95% confidence interval for the means of HPHT and API fluid loss values obtained AHR are (1.25, 2.53) and (4.41, 5.01). This means that there is a 95% chance that the mean for these two distributions lie within these two values, assuming that the samples are from a normal distribution. Previously discussed results suggest that it is reasonable to assume that fluid loss readings can be modelled using a normal distribution. The 95% confidence interval reinforces the idea that even the mean is not a fixed parameter. There is still a 5% chance that the mean for the distribution could lie outside the interval.

### *Oil-Based Mud Data from Tender Formulations*

As part of a tender, two different formulations of a 10lb/gal LTOBM were considered containing different emulsifier packages; one designated "CBE" and the other is designated "FL & VB". The requirement was to have a fluid loss of less than 2 mL. **Figure 10** shows the range of values for this and other critical properties, displayed as box plots. Box plots show the range of values (the line) and the box showing the range between the upper and lower quartile readings. The solid line across the box indicates the median value. The dots indicate outliers.

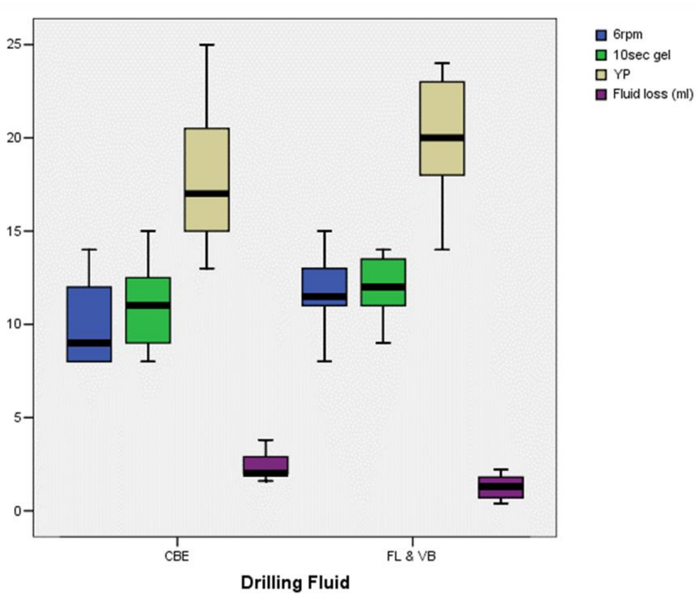


**Figure 10. Box plots for critical properties for the two formulations, CBE and FL & VB.**

The majority of the properties shown in **Figure 10** appear to have similar ranges and the means of the rheological properties are similar. This would suggest that both formulations had similarly acceptable rheological properties. However there appeared to be some difference in the fluid loss results. **Figure 11** shows the HPHT fluid loss results in more detail. It can be seen that although there is an overlap in the range of values the median values are different. This is shown more clearly in **Figure 12**.

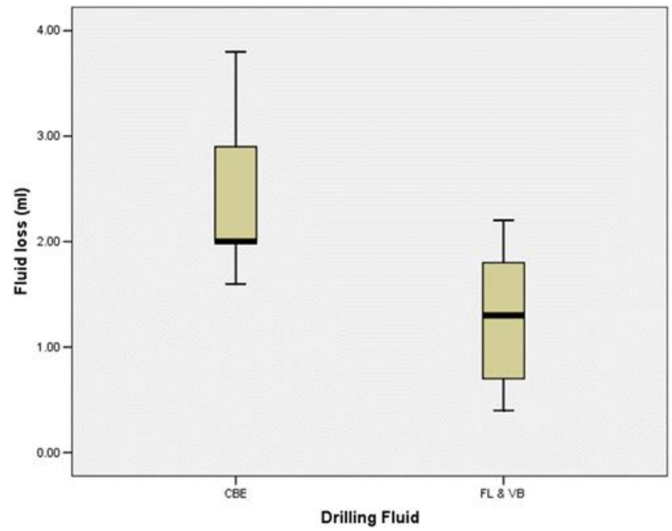


**Figure 11. Box plot showing HPHT fluid loss for two formulations, CBE and FL & VB.**

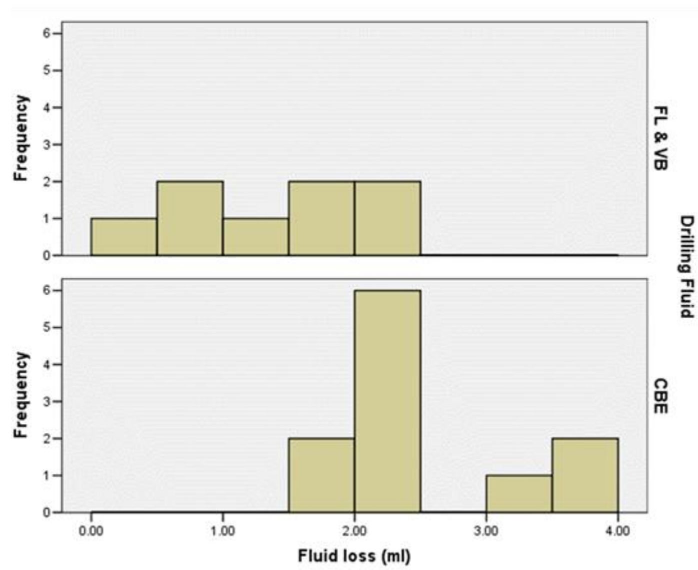


**Figure 12. Histograms of HPHT fluid loss results for the two formulations, CBE and FL & VB.**

The important question is whether these differences are significant. The mean for the LTOBM containing CBE is 2.5 mL, with a standard deviation of 0.8mL, based on 11 readings. The VB & FL fluid has a mean value of 1.3mL and standard deviation 0.7 mL, based on 8 readings. It was seen earlier in the paper that fluid loss may be modelled using a normal distribution. There is insufficient data from these experiments to determine that a normal model is correct; however, it is a reasonable assumption.

Assuming normal distribution allows us to apply a *t*-test to determine whether the mean values are different. The result is a probability of 0.004 for the null hypothesis of two means being identical. This would appear to indicate that there is a significant difference between the mean fluid loss values of the two formulations. This would suggest that the formulation containing FL & VB is the better formulation to use if a fluid loss of less than 2 mL is required. However, if only one reading was taken for each formulation, this might not have been clear, as the CBE formulation had values below 2 mL while the FL & VB formulation had reading above 2 mL. It is only with multiple testing and statistical analysis that a clear pattern can be seen.

## Conclusions

Variation, as described by the standard deviation and other statistical properties, within the data collected for different drilling fluid properties, as would be expected, is not constant. Different properties showed different amounts of variance or standard deviation.

Despite having between 12 and 16 results (many more than are commonly obtained) for the same formulation there was not enough data to be able to fully model the variation seen for most drilling fluid properties. While it cannot be demonstrated from these data that all measurements can be modeled by a normal distribution, the spread of results are in most cases consistent with it. The upper and lower limits (i.e., two standard deviations from the mean) will give the expected limits for the variation for approximately 95% of the data. There is still a 5% chance that the data could be outside these limits, assuming a normal distribution.

A normal distribution can be shown to apply to the HPHT fluid loss data for a particular LTOBM. It is possible that a normal distribution can be used to model all OBM fluid loss, but not enough data was available to confirm or deny this assumption.

The variance seen in the readings for 600rpm are far larger than those seen for 6rpm Fann 35 readings, as would be expected. However, if the standard deviation is considered as a percentage of the mean, then the value for 6rpm is much greater than for 600rpm. The standard deviation for the 600rpm measurement in both the WBM and OBM studied AHR was 4.4. As the Fann 35 viscometers are calibrated to ±3 dial reading at 600rpm, equipment error is not the only factor causing the range in the data.

The standard deviations for the calculated PV and YP are similar for the OBM and WBM, with PV: 1.7 and 1.8 cP and YP: 2.8 and 2.5 lb/100 $ft^2$ respectively. However, if the standard deviation is expressed as a percentage of the mean, there is a difference. For the PV the percentage changes from 6.1% for OBM to 13.7% for WBM. The difference is much less for the YP, with 14.8% for OBM and 13.7% for WBM.

The standard deviation and variance increases with the introduction of contamination (drill solids or seawater) into a drilling fluid formulation. This shows that the perturbation of the system by the contaminant is less well controlled than the measurements. As this study only looked at lab produced fluids, it is important to realize that the variance seen in these fluids could be different to those seen for field fluids.

Despite the limitations caused by the spread in measured data when testing a fluid's properties, it is still possible to assess the performance between different formulations of the same fluid.

If only one measurement is taken, it is not necessarily going to be the mean value. It could be anywhere within the spread of possible values. Multiple measurements are required to improve the confidence in the reliability of data.

## Acknowledgments

I would like to thank my colleagues at M-I SWACO Aberdeen Research and Technology Centre for their help in the collection of the data; and I thank the management of Schlumberger for their support and permission to use the information in this paper

## References

1. Daly, F., Hand, D. J., Jones, M., C., Lunn, A., D., and McConway, K. J.: "*Elements of Statistics*." Open University and Addison-Wesley (1999)
2. Hahn, G. J. and Shapiro, S. S.: "*Statistical Models in Engineering.*" John Wiley & Sons, Inc. (1994)