A New Statistical Method for Determination of Wax Appearance Temperature

Reidar Barfod Schüller[1], Morten Tande[2], Trygve Almøy[1], Solve Sæbø[1], Rainer Hoffmann[2], Harald Kallevik[3], and Lene Amundsen[2]

[1] Norwegian University of Life Sciences, Ås, Norway
[2] StatoilHydro, Research Centre Porsgrunn, Norway
[3] StatoilHydro, Research Centre Trondheim, Norway

ABSTRACT

A new method for determination of wax appearance temperature (*WAT*) is presented in this paper. Viscosity data (or NIR absorption data) are recorded as the temperature is reduced, and a statistical description is made of the complete temperature data set from temperature T_1 to temperature T_n. An acceptance criterion, defined as the limits of 95% or 99% prediction intervals, is then made on whether the next data point at temperature T_{n+1} is below the *WAT*. The statistical model is described in detail, making it possible for the reader to apply it to all types of experimental data where the location of some type of break point shall be determined.

INTRODUCTION

The wax appearance temperature (*WAT*) for waxy oils and condensates can be determined by several different methods like microscopy, differential scanning calorimetry (DSC) and near infra-red (NIR) absorption. Viscosity data can also be used to determine the *WAT*. Determination of *WAT* from simple rheometer tests represents a convenient way of characterizing the fluid.

Relevant studies to mention from the literature are on wax precipitation[1], wax deposition[2], cloud point determination[3], characterization of wax fractions and precipitation[4], comparing methods of determining wax content and pour point[5] and on wax content and wax precipitation temperature[6].

The proposed method, when applied to rheological data, can assume an optional linear relationship between the logarithm of viscosity, $\ln(\eta)$, versus $1/T$ at temperatures (T) above the *WAT* according to a normal type Arrhenius behaviour. Viscosity data are measured as the temperature is reduced, and a linear regression model is estimated from the complete viscosity and temperature data set from temperature T_1 to temperature T_n. These data are considered as so-called "in-control" data using terminology known from Statistical Process Control (SPC)[7]. Based on the estimated model a 95% or 99% prediction interval for the next viscosity data point at temperature T_{n+1} is constructed. The upper interval limit is used to determine whether the next observed viscosity represents an "out-of-control" state. If an out-of control state is found, the temperature T_{n+1} is considered as the *WAT*.

We first describe the statistical method in general terms making it applicable for other types of experimental data where break points are to be identified. Then we present the application of our methodology to the determination of *WAT* for the viscosity data and to NIR absorption data.

THE MODEL

A linear regression model relating a dependent variable y to a single predictor variable x is assumed.

$$y_i = a + bx_i + \varepsilon_i \quad (1)$$

where the error terms ε_i's are assumed to be independent, normally distributed variables, each with expectation 0 and variance σ^2. The model is assumed to hold for x-values below (or above) an unknown limit x_{LIM}. Beyond this limit the relationship between y and x becomes non-linear. The purpose of the method described in the following is to identify the break point x_{LIM}.

Within the area of Statistical Process Control (SPC) methods have been developed for similar purposes[7]. SPC is used to monitor industrial processes in order to reveal any out-of-control situations in which case the process should be stopped. For the matter of identifying an out-of-control situation the properties of the process in-control must be learned from in-control observations.

Let (y_1,x_1) $\ldots(y_n, x_n)$ be a set of experimental data assumed to represent an in-control state. The linear model is fitted to the in-control data using Least Squares Estimation, and based on the estimated regression coefficients, a and b, a predictor for a new observation y^* given observed predictor value x^* will be:

$$\hat{y}^* = \hat{a} + \hat{b}x^* \quad (2)$$

In addition to the estimates of the regression coefficients an unbiased estimate of the error variance σ^2 is obtained by the Mean Sum Squares of the Error (MSE). The MSE is given by the standard formula:

$$\text{MSE} = \hat{\sigma}^2 = \frac{\sum_{i=1}^{N} e_i^2}{n-2} \quad (3)$$

Where e_i is the residual of observation i found as the deviation between the observed response value and the value predicted by the model:

$$e_i = y_i - \hat{y}_i \quad (4)$$

From the linear model fit a prediction interval for a new observation y^* can be constructed using the standard formulas from linear regression (e.g. Montgomery *et al.*[8]). A (1-α)100% prediction interval is an interval which with probability (1-α) will contain a future observation y^* for a given value x^*. The interval is defined by a lower limit LL and an upper limit UL. The upper limit is given by:

$$UL = \hat{y}^* + t_{\alpha/2,n-2}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \quad (5)$$

and the lower limit by:

$$LL = \hat{y}^* - t_{\alpha/2,n-2}\hat{\sigma}\sqrt{1+\frac{1}{n}+\frac{(x^*-\bar{x})^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}} \quad (6)$$

where $\bar{x}$ is the mean of the in-control x-observations. The statistic $t_{\alpha/2,n-2}$ is the upper $\alpha/2$ percentile of the student-t distribution with $n-2$ degrees of freedom. Percentiles of the t-distribution can be found in standard text-books on introductory statistics or by using the TINV-function in Excel. Typically the parameter α is chosen to be equal to 0.05 or 0.01, corresponding to 95% or 99% prediction intervals, respectively.

In order to identify a break point a linear prediction model is fitted based on the first n data points where we assume that the observations are ordered and that the last observation is closest to the break point. The next observation, for which the predictor value is x_{n+1}, is assumed to be even closer to the break point or being at the break point. Based on the fitted model the prediction interval is computed. If the observed value y_{n+1} lies below the LL or above the UL (an outlier) this is taken as an indication of a possible break point, since the probability of this to occur, given that the model is true, is only equal to α. In case the observed response value is found to be an inlier, the point (y_{n+1}, x_{n+1}) is added to the in-control

sample set and the prediction model is refitted using the extended data set. Then the procedure is repeated for the next data point. Since, in general, outliers will occur at a rate of α a single outlier is not a strong indication of a break-point, but repeated successive outliers are therefore a stronger indication.

For the special case of determining *WAT* from rheoogical data we can write:

$$x = \frac{1}{T} \tag{7}$$

$$y = \ln(\eta) \tag{8}$$

where T is absolute temperature in K and y is the natural logarithm of the viscosity.

The basic Arrhenius assumption is based on a linear relationship between $\ln(\eta)$ and *1/T*. As the temperature is decreased, then *1/T* increases, and $\ln(\eta)$ increases approximately linearly until wax precipitates bringing about a change in the basically linear Arrhenius behaviour. The present method is an attempt to determine the first significant data point indicating wax precipitation.

For the specific problem of detecting *WAT* it is always the upper limit, *UL*, which is of interest. If $y_{n+1} > UL$, we have found the *WAT* if successive points lie above the *UL*.

APPLICATION OF METHOD

It is possible to apply the method to any type of linear x, y data set, and it will here be shown examples with rheological data and for data from NIR absorption where a matrix (y = absorption, $x = T$) is used as input.

Case 1: A rheological case

A data set on a real North Sea condensate was recorded in rotation using a Physica MCR301 rheometer with a XL/PR400/TI high pressure cell and DG35.12/PR/TI measuring system. The viscosity was recorded every 0.1 °C as the temperature was reduced at a shear rate of 1000 s^{-1}. The raw data plot of viscosity versus temperature is shown in Fig. 1. In order to work on linear data above the *WAT* the data are transformed and an Arrhenius representation of the data is shown in Fig. 2 as logarithmic viscosity versus 1/*T*. At temperatures above the *WAT* the curve is approximately a straight line. It is not obvious to see where wax starts to precipitate from these plots. The 1st derivative of the raw data is also shown in Fig. 2. We observe that the slope is approximately constant in the region from 0.0030 to 0.0032 K^{-1}, which corresponds to a constant slope of a straight line. This documents that the Arrhenius plot is a straight line at temperatures above the *WAT*, and the data become unlinear at 1/*T*-values larger than 0.00325 K^{-1}.

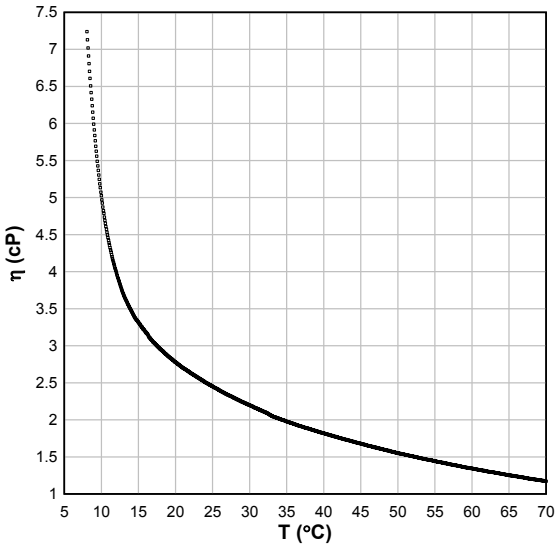

Figure 1. Plot of raw data series with data point interval of 0.1 °C.

If the data are analyzed according to the model presented above, Figs. 3, 4, and 5 show the values predicted from the model (Eq. 2) and the upper and lower boundaries for a 99% prediction interval; Eq. 5 and Eq. 6. It is possible to see where the data series cross the upper boundary, and the first data point that lies above the *UL* is taken as the first indication that the *WAT* has been reached at 32.9 °C.

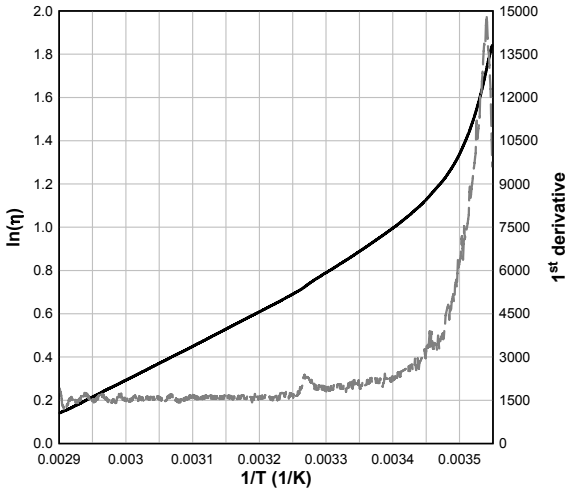

Figure 2. Raw data transformed to an Arrhenius plot of logarithmic viscosity versus $1/T$ (solid line) and the 1st derivative of the Arrhenius curve (dashed line). The derivative shows that the slope of the log viscosity curve is constant for values less than approximately 0.00325 K^{-1}.

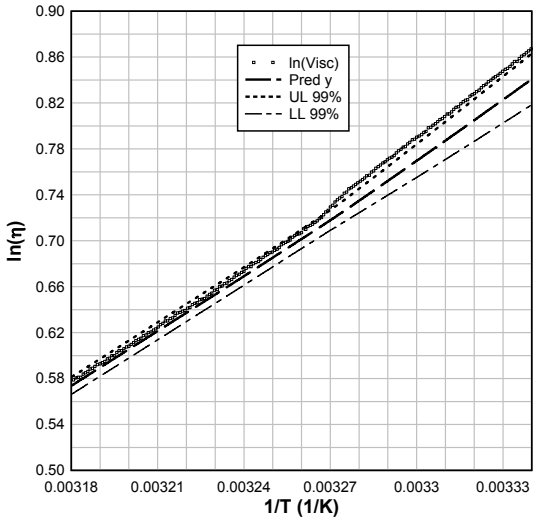

Figure 3. Arrhenius plot of raw data, predicted *y*, upper boundary (*UL*) and lower boundary (*LL*) for 99% prediction interval zoomed to region of the *WAT*.

If the Arrhenius linearization is skipped in the above analysis, and $\ln(\eta)$ versus T is used as the data set, then the predicted *WAT* also becomes 32.9 °C if a 99% prediction interval is chosen. It should be noted that the curve of $\ln(\eta)$ versus T above the *WAT* is not very unlinear ($\left|d^2 \ln(\eta)/dT^2\right| < 10^{-3}$.).

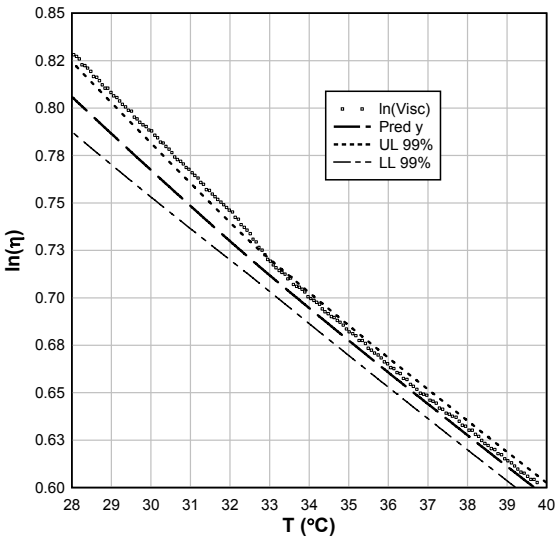

Figure 4. The data in Fig. 3 transformed to temperature on the *x*-axis.

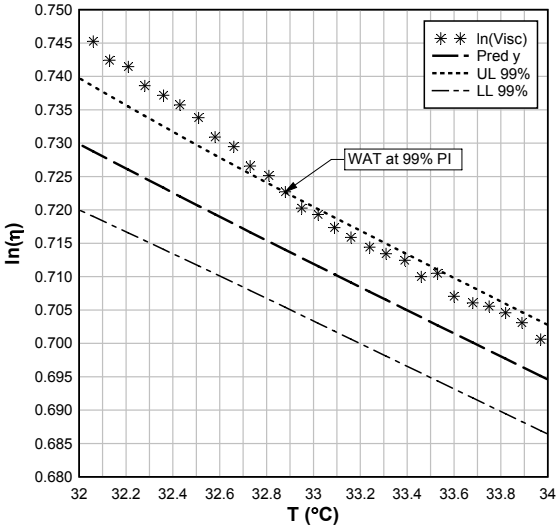

Figure 5. The determined *WAT* is located where the first data point, in a reducing temperature trend, lies above the *UL*. *WAT* = 32.9 °C.

Case 2: A NIR data case

The statistical method has also been applied to other data sets obtained from a NIR absorption system (Brimrose Luminar 2000 AOTF) with a fibre optic custom made high pressure transmission cell. Temperature measurements were recorded by a

Eurotherm/PT-100. The same fluid (Case 1) was tested at 1 bara and 80 bara.

The absorption versus temperature data were in this case used without any linearization above the *WAT*, and the resulting prediction models, with upper and lower boundaries for 99% prediction interval, are shown in Figs. 6 and 7, giving *WAT*-values of 33.4 °C and 31.1 °C, respectively.

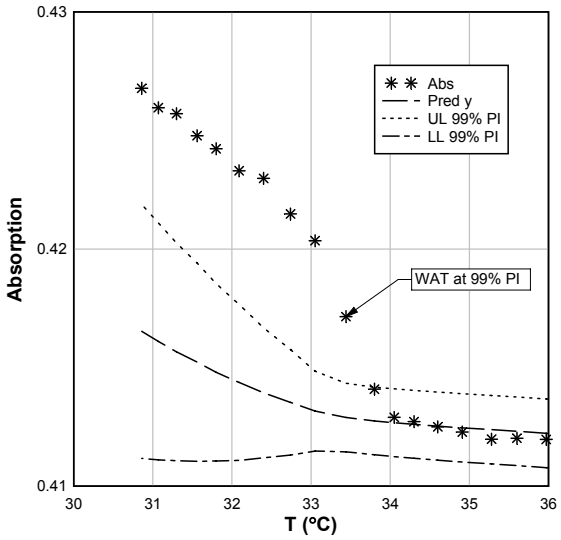

Figure 6. Example of determination of *WAT* from a NIR data set in Case 2, 1 bara data. *WAT* = 33.4 °C.

DISCUSSION

The quality of the data series to be analysed affects the resulting accuracy of the *WAT* determination. This is seen by examining Eq. 5 where the position of the upper limit boundary (*UL*) is a function of the MSE, Eq. 3.

If there is a large amount of noise in the data, this will result in an increase in the value of the MSE. The accuracy of the *WAT* determination is obviously also linked to the temperature difference between adjacent data points. In Case 1 the temperature step in the data is 0.1 K, and the accuracy of the basic *WAT* determination can therefore not be better than 0.1 K.

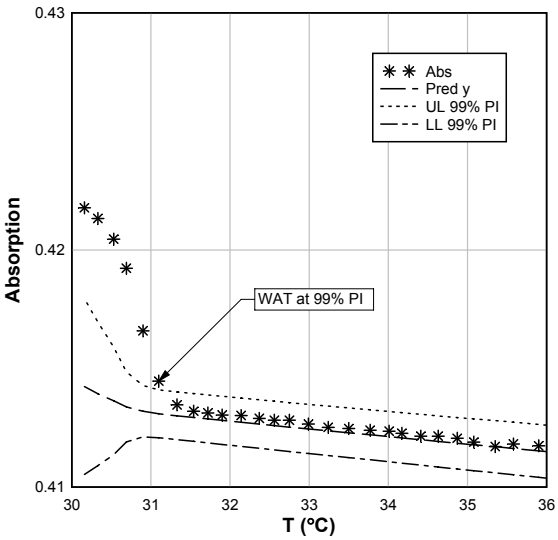

Figure 7. Example of determination of *WAT* from a NIR data set in Case 2, 80 bara data. *WAT* = 31.1 °C.

The new method for determination of wax appearance temperature has successfully been applied to both rheometer data and data from NIR absorption. The method is very simple to use since it does not involve any subjective interaction from the user that involves setting parameters that will influence on the result. Both the rheology data and the NIR data at 1 bara give *WAT* = 33 °C.

The squared prediction error graph, shown in Fig. 8, gives a representation of the region, between 32.5 °C and 33.5 °C, where the *WAT* is located, but the exact data point in this region that coincides with the 99% prediction interval criterion cannot be read directly from the graph.

A prediction interval of 99% has been used in this study, but this could have been chosen less strict. A 99% prediction interval ($\alpha = 0.01$) gives $t_{0.01/2,n-2} = 2.58$ for large values of n. A 95% prediction interval gives $t_{0.05/2,n-2} = 1.96$ for large values of n as shown in Fig. 9.

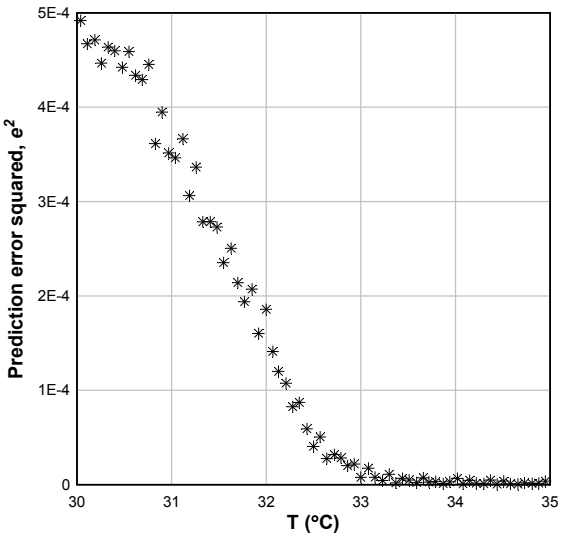

Figure 8. Squared error of prediction versus temperature for the individual data points.

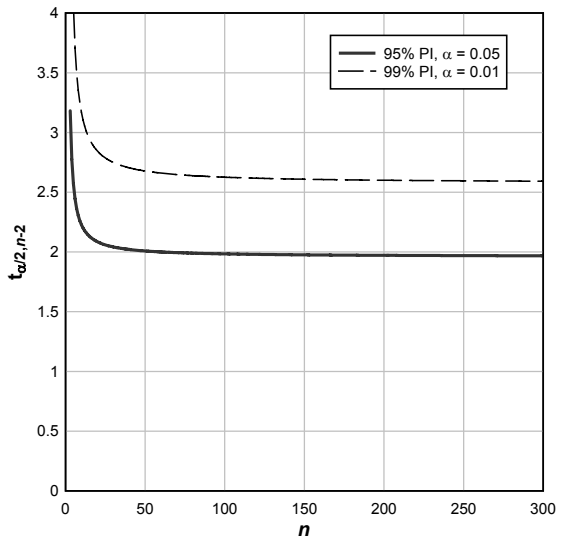

Figure 9. Values of $t_{\frac{\alpha}{2},n-2}$ as a function of n for $\alpha = 0.05$ and $\alpha = 0.01$.

It may be possible to apply some sort of data interpolation to determine the location where the data series crosses the *UL* line, but this approach has not been tested in the present paper, but it seems a feasible approach looking at the data in Fig. 6.

The sensitivity of the determined *WAT* to the chosen prediction interval is shown in Fig. 10 for Case 1.

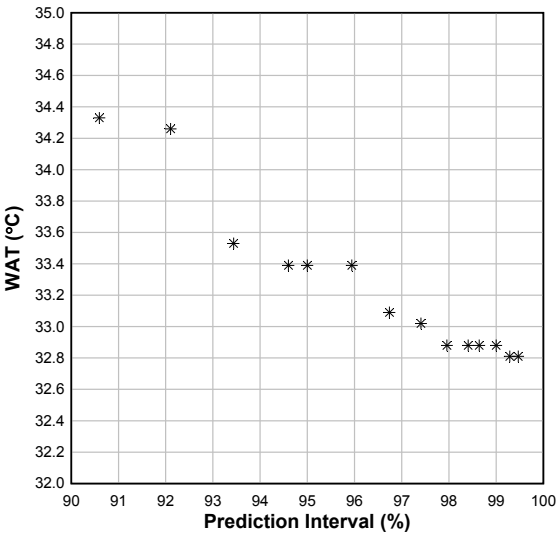

Figure 10. *WAT* as a function of chosen prediction interval for the rheological data points of Case 1.

We observe that the *WAT* changes by less than a degree when the prediction interval is changed from 95% to 99%. The *WAT* generally reduces with increasing prediction interval.

It is important to observe that the user does not need to make any decision on a tolerance limit for determination of the *WAT*. Although the chosen prediction interval will affect the value to some degree, the method still has an objective quality. The *WAT* at 95% prediction interval is 33.4 °C and that at 99% prediction interval is 32.9 °C (Fig. 10). The difference between the determined *WAT* with 95% and 99% prediction interval is therefore only 0.5 K for the data in Case 1.

If the model is applied to the raw viscosity data, without transformation to Arrhenius plot, the predicted *WAT* for Case 1 at 99% prediction interval is independent of the initial Arrhenius linearization. The method may therefore be applied directly to the raw data series if a high prediction interval is chosen. This means that the data do not have to be highly linear above the *WAT* for the method to work satisfactorily.

CONCLUSIONS

The conclusions from this work can be summarized as follows:

- The new method is able to determine the *WAT* from both rheological and NIR data.
- The method does not involve any subjective input from the user since the criterion for determination of *WAT* is defined for a specified prediction interval.
- The model is able to detect the first data point in the series that deviate from the accumulated prediction model.
- The data series must not necessarily exhibit a strong linear behaviour above the *WAT* as shown with the unlinearized rheology data and the NIR data.

REFERENCES

1. Coto, B., Martos, C., Peña, J.L., Espada, J.J., and Robustillo, M.D. (2008), "A new method for the determination of wax precipitation from non-diluted crude oils by fractional precipitation", *Fuel*, **87**, 2090-2094.

2. Elsharkawy, A.M., Al-Sahaf, T.A. and Fahim, M.A. (2000), "Wax deposition from Middle East crudes", *Fuel*, **79**, 1047-1055.

3. Coutinho, J.A.P. and J.-L. Daridon, (2005), "The limitations of the cloud point measurement techniques and the influence of the oil composition on its detection", *Petr. Sci. Technol.*, **23**, 1113-1128.

4. Martos, C., Coto, B., Espada, J.J. Robustillo, M.D., Gómez, S., and Peña, J.L. (2008), "Experimental determination and characterization of wax fractions precipitated as a function of temperature", *Energy & Fuels*, **22**(2), 708-714.

5. Kök, M.V., Letoffe, J.M., and Claudy, P. (2007), "Comparative methods in the determination of wax content and pour points of crude oils", *J. Therm. Anal. Cal.*, **90**(3), 827-831.

6. Davidsen, S. and Hamouda A.A. (1999), "Innovative method to determine the wax content and the wax precipitation temperature simultaneously for crude oils at pipeline pressures", *SPE International Symposium on Oilfield Chemistry*, Houston: Society of Petroleum Engineers, SPE50747, pp. 459-473.

7. Montgomery, D.C., (2005) "Introduction to Statistical Quality Control", 5 ed., Hoboken, New Jersey: Wiley & Sons. Inc.

8. Montgomery, D.C., E.A. Peck, and G.G. Vining, (2006) "Introduction to linear regression analysis", 4 ed, Wiley-Interscience. Hoboken, New Jersey: John Wiley & Sons, Inc. 612, 978-0-471-75495-4, 0-471-95495-1.