Gaussian Distribution

The objective of this case study is to investigate how signal intensity varies with retention time in a chromatographic experiment and to quantify that relationship using nonlinear curve fitting in **Isalos Analytics Platform**. In this case study, the **Gaussian Distribution** model is used to estimate the peak shape and location of a compound signal using a standard normal distribution framework.

The dataset used in this case study is derived from an LC–MS peak detection dataset available on Zenodo. It contains retention time values and the corresponding signal intensity measurements obtained from a chromatographic analysis. In this model, the independent variable (X) is the retention time, and the dependent variable (Y) is the signal intensity.

The Gaussian Distribution model describes a nonlinear relationship in which the response forms a symmetric peak around a central value. The equation used in this analysis is:

$$Y = \text{Amplitude} \cdot \exp\left(-\frac{1}{2}\left(\frac{X - \text{Mean}}{\text{SD}}\right)^2\right)$$

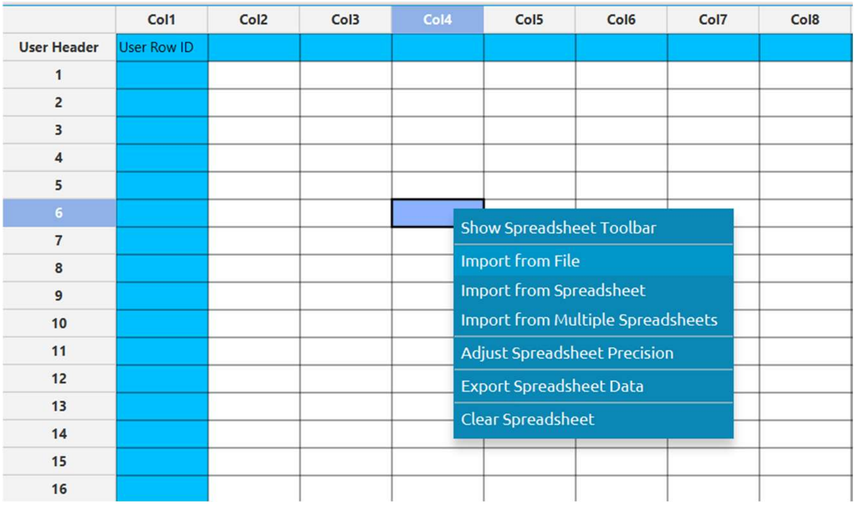
In this model, **Amplitude** represents the peak height of the distribution (maximum Y value), **Mean** represents the center of the peak (the retention time at which the maximum signal occurs), and **SD** represents the standard deviation, which determines the spread or width of the peak.

The purpose of this analysis is to determine whether the observed chromatographic data are consistent with a Gaussian distribution and to convert the measured signal values into meaningful quantitative outputs. The main results obtained from the fit are **Amplitude**, **Mean**, and **SD**, where Amplitude reflects the peak intensity, Mean reflects the retention time of the peak center, and SD reflects the dispersion of the signal around the center. These results are useful because they allow quantitative characterization of chromatographic peaks and support interpretation of compound identification and separation efficiency.

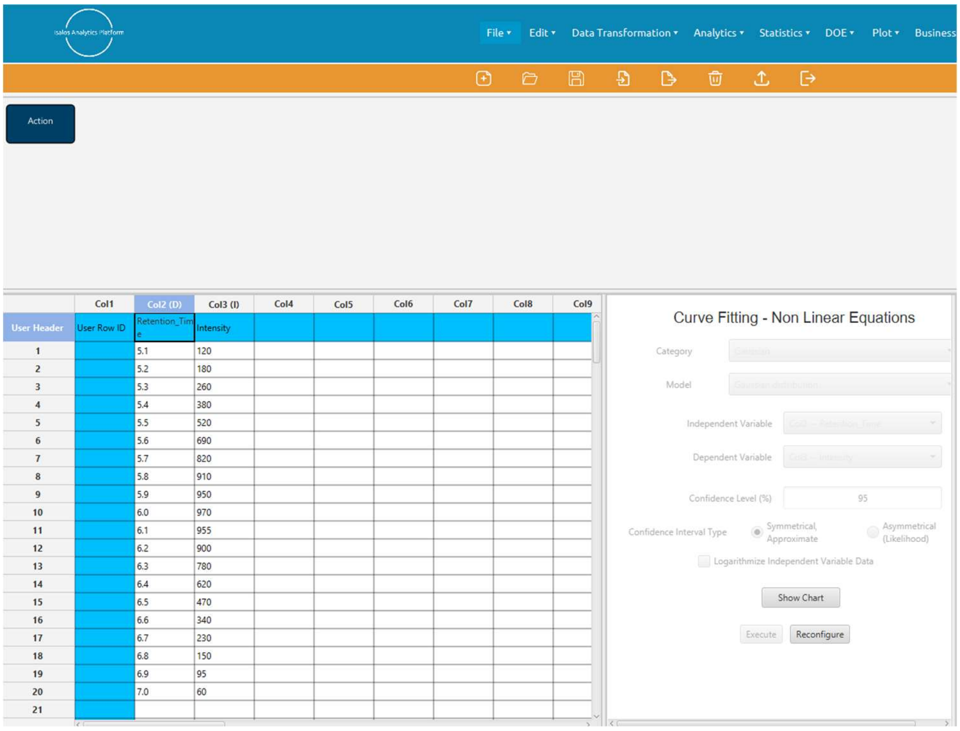
Isalos version used: 2.0.2
Scientific Dataset: https://zenodo.org/records/14984876

Step 1: Import data from file

Right-click on the input spreadsheet panel on the left and choose "**Import from File**". Then browse to the file containing the XY dataset for this case study and load the sheet in which each row corresponds to one retention time value and one signal intensity value.

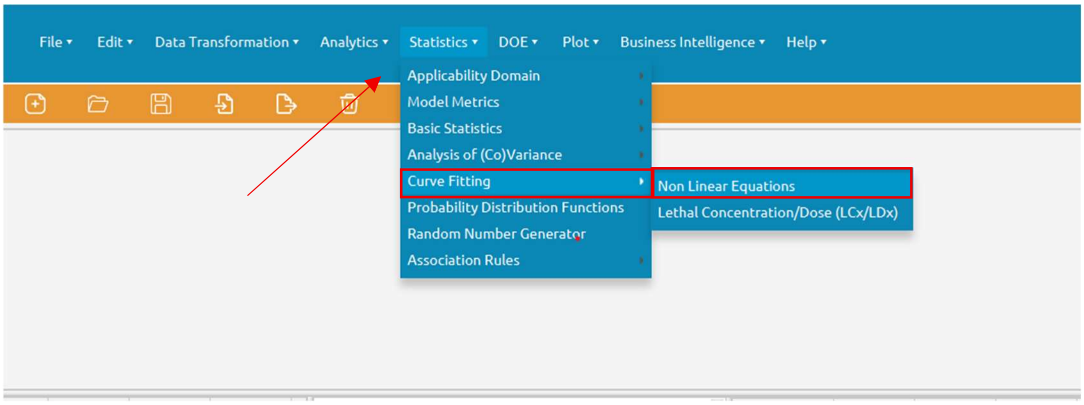


The data will appear on the left spreadsheet.

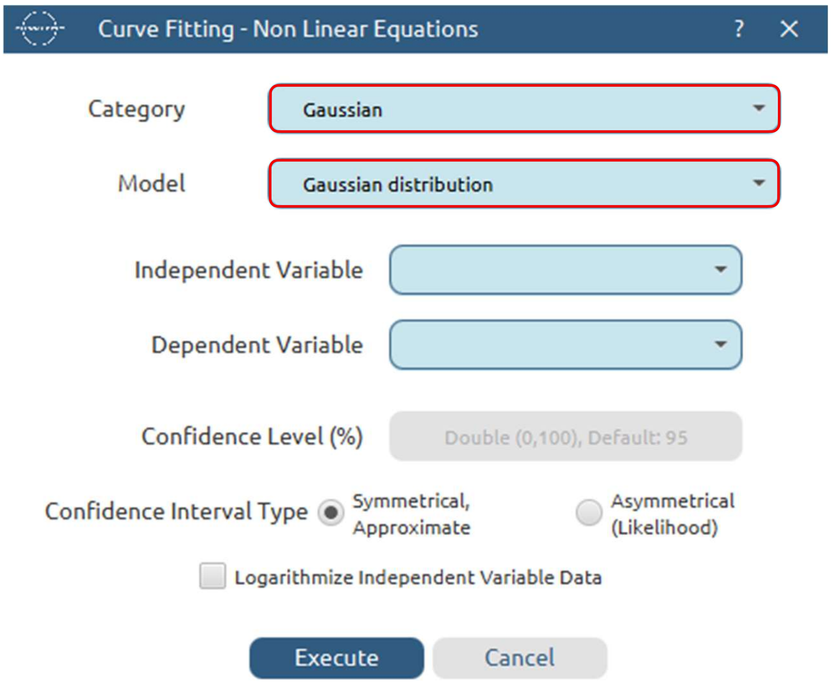


Step 2: Select the desired statistical analysis model

From the toolbar, open the **Statistics** drop-down list and navigate through: **Statistics > Curve Fitting > Non Linear Equations.**

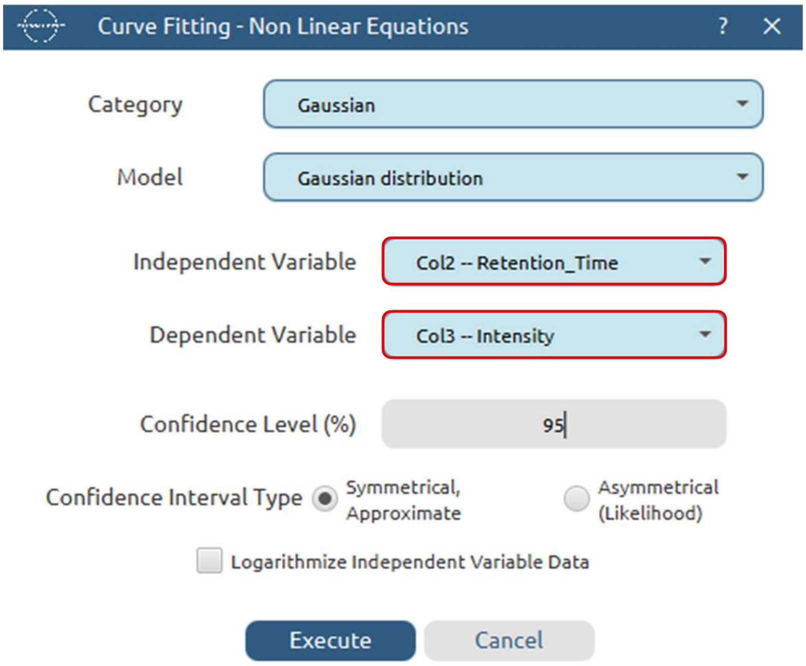


In the category list, select the category containing **Gaussian** models and then choose the **Gaussian Distribution model** from the model menu.



Step 3: Configure variables and confidence intervals

Set the column containing the **retention time** values as the **independent variable (X),** and set the column containing the **signal intensity** values as the **dependent variable (Y).** Set the confidence level to 95% and choose Symmetrical Approximate as the confidence interval type.

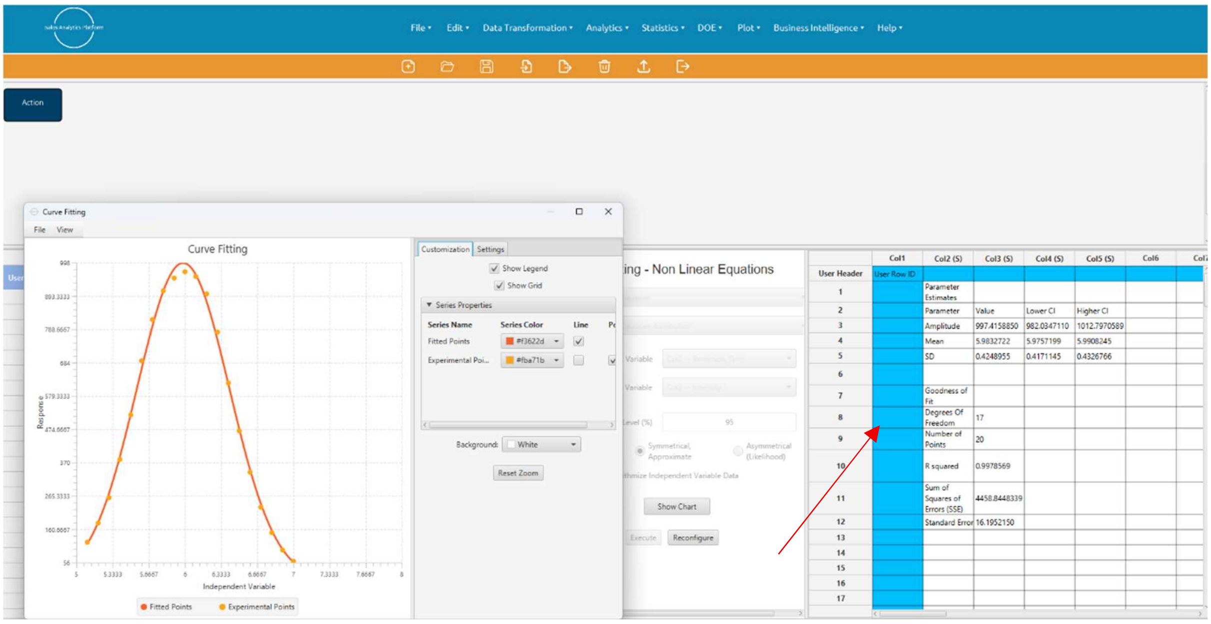


Step 4: Analyze the output and fitted curve

Once the analysis is completed, Isalos presents the fitted Gaussian curve together with the experimental data points, allowing direct visual assessment of how well the model captures the chromatographic peak shape. In this study, the graph shows a symmetric bell-shaped curve centered around the **Mean value**, supporting the use of a **Gaussian distribution model.**

The results page presents the estimated model parameters with their confidence limits, together with goodness-of-fit statistics and the corresponding fitted plot. In the **Gaussian Distribution model**, the main fitted parameters are **Amplitude**, **Mean**, and **SD**, and these values collectively determine the height, position, and width of the fitted curve.

The **Goodness of Fit** section of the table summarizes key fitting statistics such as the **number of data points used**, **degrees of freedom**, **residual sum of squares**, and the **standard error** of the regression. These outputs should be interpreted together with the fitted plot in order to evaluate how well the model describes the data and how reliable the estimated parameters are.



The fitted results indicate that the chromatographic signal data are well described by a **Gaussian Distribution model**, showing that the variation in signal intensity across retention time follows a symmetric peak rather than a linear or polynomial trend. The estimated **Amplitude** of **997.416** represents the maximum signal intensity of the peak, the estimated **Mean** of **5.983** represents the retention time at which the peak reaches its maximum, and the estimated **SD** of **0.4249** represents the spread of the peak, indicating how quickly the signal rises and falls around the center.

The confidence intervals for all three parameters are narrow and do not show excessive variation, which supports the stability and reliability of the fitted model. In particular, the confidence interval for **SD** does not approach zero, confirming a well-defined peak width, and the confidence interval for the **Mean** is tightly centered around the peak location, indicating precise estimation of the retention time.

Together with the high **R^2** value of **0.99786** and the relatively low **residual error** (SSE = 4458.84, Standard Error = 16.20), the results indicate that the **Gaussian Distribution model** provides a good description of the chromatographic data and that the fitted parameters are reliable for characterizing the peak shape and position in this study.

References:

(1) Zou, S., Cui, Q., Liu, J., Wu, Q., Zhu, L., Chen, D., Du, Y. and Wu, T., 2025. Local Asymmetric Gaussian Fitting Algorithm for Enhanced Peak Detection of Liquid Chromatography–High Resolution Mass Spectrometry Data. *Analytical Chemistry*, *97*(20), pp.10603-10610.