

STarStation curve fitting: An Introduction to standard curves



- **Optimise standard data**

Use one of four curve fitting models to optimise quantification of data via standard curves.

- **Reject outliers**

Outlying standard points can be excluded from the standard curves to discount erroneous data.

- **Weight standard points**

STarStation 2.0 introduces standard data weighting to assign more influence to data, further enhancing standard curve accuracy.

Introduction

When calculating quantitative results for an assay (analyte concentrations), a series of known concentrations of an analyte is used to construct a plot of intensity versus concentration. Mathematical modelling of this plot is then used to obtain an equation for the so called standard or calibration curve. This equation generated from the standard curve is then used to predict concentrations of unknown samples based on the response (fluorescence (MFI)) measured. STarStation software provides users with several models for calculating a standard curve.

The curve fitting model directly affects the interpretation of results generated by STarStation and users should ensure they are using the most suitable algorithm for the type assay data that is being analysed. The four standard curve fitting models and their uses in STarStation are: -

- **Spline Interpolation:** A simple, quick, method where subsequent data points are joined with a series of formula.
- **Log-Log regression:** A curve fit ideal when the data is linear when plotted on a logarithmic scale.
- **Four Parameter Logistic (4PL) equation:** This curve fit is used to model complex curves where the data doesn't show a linear trend (i.e. for sigmoid curves). Four parameters are taken to generate the curve.
- **Five Parameter Logistic (5PL) equation:** An extension of the 4PL curve fit which provides optimal generation of standard curves when using asymmetric data.

In general, users should choose the curve fitting method which best models the particular data generated by their assay. Guidelines for choosing the most appropriate curve fitting method is discussed in the STarStation curve fitting algorithms section.

STarStation 2.3 also includes a data point weighting system for 5PL curve fits, allowing users to increase or decrease the influence of certain data points on the standard curve shape. Outlier standard data points may also be removed from the data set used by the curve fitting algorithms; this affords the user additional control over the way in which results are calculated.

An overview of the various curve fitting formulae used by STarStation and guidelines for their use are presented below.

STarStation Standard Curve Fitting Methods

STarStation generates standard curve fits using any one of four methods: Spline interpolation, Log-log regression, Four Parameter and Five Parameter Logistic regression. Several Curve Fitting methods are included to allow users to choose the one which is most appropriate to the standard data. The curve fitting model selected for deriving the standard curve for a particular analyte is selected during multiplex assay creation in the STarStation Assay Manager and may be modified during the data review step of STarStation software workflow in the Reports View.

Spline Interpolation

Spline interpolation calculates curved lines between calibration points to produce a standard curve. The equation used for each segment is a cubic polynomial and each of the short curves can then be modified to compel it to join smoothly to the next (smoothing). Smoothing is a repetitive process as each segment is recalculated until the join is smooth. The composite mathematical function obtained is called a spline function.

STarStation uses cubic equations to join each subsequent data point. The equation will differ for each segment as different curves are calculated. The general form of the cubic equation is:

$$y = a + bx + cx^2 + dx^3$$

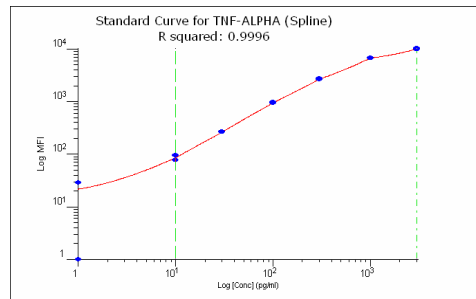


Figure 1: Example Standard Curve Plotted using the STarStation Spline Algorithm

Log Log regression

A log-log regression curve fit is calculated by transforming the data by taking logarithms and performing a linear regression on the transformed data to calculate a straight line of best fit. This curve fit should be used when the standard results follow a linear pattern on a log-log scale.

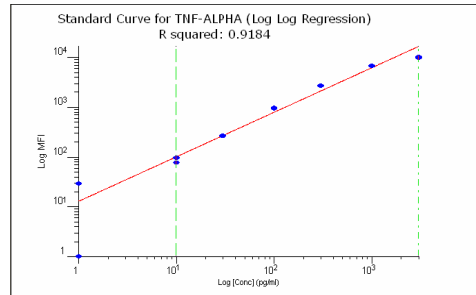


Figure 2: Example Standard Curve Plotted using the STarStation log-log Regression Algorithm

Nonlinear or Logistic Regression

Often the most applicable method for interpretation of immunoassay data is logistic regression. In this case the standard curve is defined by a non-linear equation that uses a number of parameters. The parameters are produced via calculating and re-calculating curves by altering the parameters (logistic regression) until the most accurate curve is generated. There are a large number of well-known nonlinear models used in different sciences. Two of the most commonly used nonlinear models used in immunoassay analysis are available in STarStation: the Four-parameter and the Five-parameter models.

Four-Parameter Logistic (4PL) Equation

A 4PL equation uses four parameters to define the standard curve. While it is a more complex method for calculating curves of best fit it is not ideal for asymmetric data. The form of the 4PL equation used in STarStation is detailed below:

$$y = B + \frac{A - B}{1 + \left(\frac{x}{C}\right)^D}$$

Where: *B* is the estimated response at infinite concentration
A is the estimated response at zero concentration
C is the mid-range concentration (*C*₅₀)
D is the slope (also known as Hill slope)

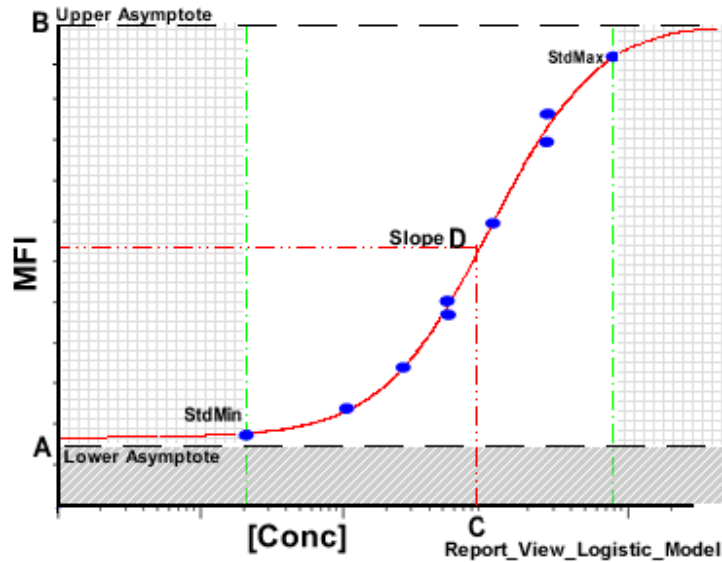


Figure 3: A Symmetric Standard Curve Plotted using the STarStation 4PL Curve Fit.

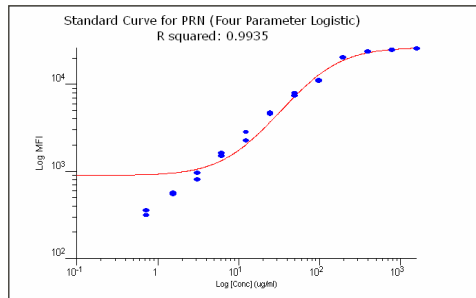


Figure 4: An Asymmetric Standard curve Plotted using the STarStation 4PL Curve Fit.

Five-parameter Logistic (5PL) Equation

The application of the 4PL logistic to asymmetric data can sometimes produce a poor curve fit, leading to erroneous results (see figure 4). In this case the 5PL equation, which has an asymmetry parameter, can provide a more optimal fit. STarStation implements the following version for 5PL curve fits:

$$y = B + \frac{A - B}{\left(1 + \left(\frac{x}{C}\right)^D\right)^E}$$

Where: *B* is the estimated response at infinite concentration
A is the estimated response at zero concentration
C is the mid-range concentration (*C*₅₀)
D is the slope (also known as Hill slope)
E is the asymmetry factor

When *E* = 1 the equation is identical to the 4PL equation.

PD 035-308 STarStation Standard Curve Fitting

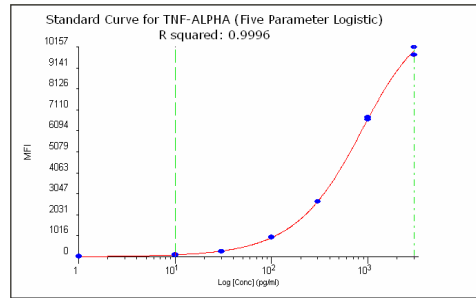


Figure 6: Standard Curve Plotted using the STarStation 5PL Curve Fit (symmetric data).

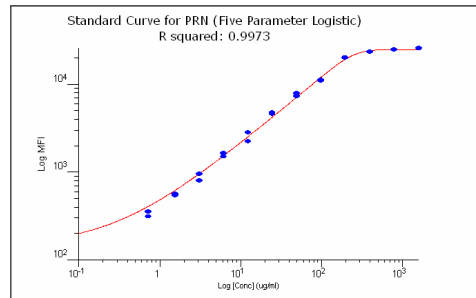


Figure 7: A Standard Curve Plotted using the 5 Parameter Logistic (5PL) Curve Fit (asymmetric data)

Weighted Five Parameter Logistic Curve Fits

The aim of performing logistic regression is to find the best parameter values to produce a best-fit curve. Most often nonlinear regression is done without weighting. The curve fitting minimises the sum-of-squares of the vertical distances of the data from the curve. This method gives equal weight to all points, as is appropriate when you expect experimental scatter to be the same in all parts of the curve. STarStation 2.0 allows data point weighting when a 5PL formula is used. The weighting factor applied to individual data points dictates how the data points influence the curve fitting process.

Ideally the smaller the variance (error) that a data points shows, the higher it should be weighted. This is because the data points with a lower error value are more reliable data and have a higher probability of lying on the true curve of best fit (the curve which STarStation is trying to estimate).

STarStation calculates which curve has the maximum likely hood estimate of the true curve by using the data points with the smallest sum of squared errors. The smaller the sum-of-squares errors in the curve, the more accurate the curve fit will be.

Three weighting models, along with a manual weighting function, are included in STarStation 2.0. Users are advised to apply different curve fitting methods to produce the best-fitting curve for the data in question.

Relative weighting (weighting by 1/y²)

This method attempts to minimise the sum-of-squares of the relative distances of the data from the curve.

This method is appropriate when you expect the average distance of the points from the curve to be higher when Y is higher, but the relative distance (distance divided by Y) to be a constant. In this common situation, minimising the sum-of-squares is inappropriate because points with high Y values will have a large influence on the sum-of-squares value while points with smaller Y values will have little influence. Minimising the sum of the square of the relative distances restores equal weighting to all points.

Reciprocal (Poisson) weighting (weighting by 1/y)

Weighting by 1/y is a compromise between minimising the actual distance squared and minimising the relative distance squared. 1/y weighting is appropriate when the y values follow a Poisson distribution. When this is the case, in order to find the best-fit curves from the standard data users can choose this weighting method.

Weighting by observed variability (reciprocal st. deviation squared (1/sd²))

It isn't always possible to determine how the data scatter varies relative to y values. When this is the case users can base the weighting of the data on the observed variation among replicates by choosing this weighting method.

This method assumes that the mean of replicates with a large standard deviation is less accurate than the mean of replicates with a small standard deviation. This assumption is not always true; when only a low number of replicate standards are used the standard deviation may vary considerably by chance (especially at lower concentrations).

Manual Weighting

Manual weighting is provided for users who wish to adjust their curve fit for if they aren't satisfied with the curve fit generated by StarStation. For example, occasionally the standard curve will not be plotted through some standard data points or will be plotted in negative y values. If the user wishes to force the curve through these points they can assign a higher weight to these points to influence the curve fit.

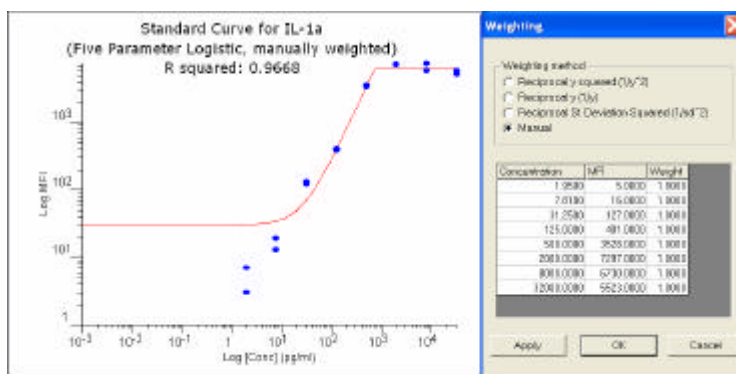


Figure 8: A 5PL standard curve plotted from non-weighted data points.

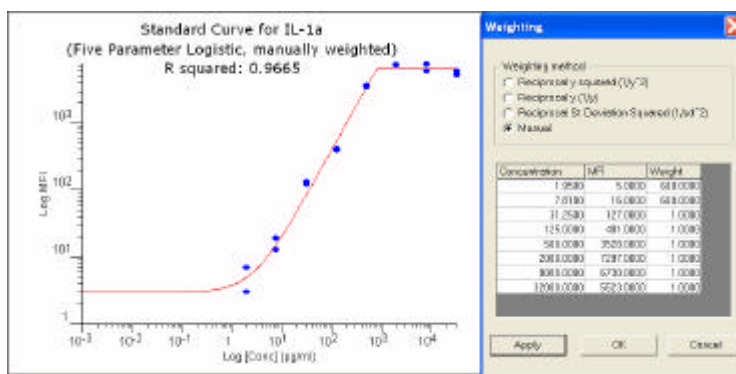


Figure 9: A 5PL standard curve plotted from manually weighted data points.

Outliers

Often one or more data points of a calibration curve are out of consensus with the other points. A pair of duplicate standards may produce significantly differing results (through pipetting error, or by a typing mistake, for example) and users may wish to exclude one (or more) erroneous data points to enhance the curve fit. This is done by de-selecting the data point using the data exclusion checkbox in the "sample ID" column of the Review Grid in the Reports View.

PD 035-308 StarStation Standard Curve Fitting

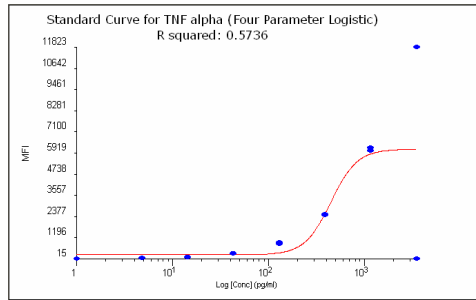


Figure 10: Standard curve with an erroneous upper standard data point

The standard Curve Fitting Model used for a particular analyte is modified in the STarStation Reports View via the *Net MFI-Curve Fitting* Toolbar or the *Curve Algorithm* option on the standard curve context menu.

The Standard Curve context menu is accessed by right mouse clicking the Standard Curve Window, selecting the *Edit Standard Curve* option from the *Edit* Menu or by pressing the Edit Standard Curve Icon on the *Sample Range Checking* Toolbar.

The Curve Fitting Method used to generate the standard curve can be modified via the Curve Fit Toolbar (1) or the Standard Curve Context Menu (3). The Standard Curve Context Menu is accessed by placing the mouse pointer in the standard curve window and clicking the right mouse button or via the Edit Standard Curve Icon on the Sample Range Checking Toolbar (2)

Summary	Mean [Conc]	CV% (n)
1 Four Parameter Logistic	29.00	#NUM NULL (0)
2 Five Parameter Logistic-not weighted	77.00	6.49 NULL (1)
3 Log Log Regression	263.00	30.74 1.09 (2)
4 Std03 (30.00 pg/ml)	961.00	105.98 NULL (1)
5 Std04 (100.00 pg/ml)	2691.00	294.25 NULL (1)
6 Std05 (300.00 pg/ml)	6701.00	1003.24 2.44 (2)

Review	SampleID	[Conc]	MFI	Bead CV%	Multiplex	LMD	Description	QC Comment
1	Std01 (0.00 pg/ml)		0.00	0.00	3-plex human 02 (2002-08-			User Rejected Std!
2	Std01 (0.00 pg/ml)	#NUM	29.00	493.54	3-plex human 10 (2002-08-			
3	Std02 (10.00 pg/ml)	6.49	77.00	454.24	3-plex human 03 (2002-08-			
4	Std02 (10.00 pg/ml)		95.00	50.06	3-plex human 11 (2002-08-			User Rejected Std!
5	Std03 (30.00 pg/ml)	30.50	261.00	142.59	3-plex human 04 (2002-08-			

Reports_View_Standard_Curve_Modifying_Standard_Curve

Reports_View_Curve_Algorithm_Menu

The calculated parameters for a particular standard curve can be displayed via the **Show Parameters** option on the Curve Algorithm menu.

Selecting the **Show Parameters...** option displays the parameters of the 4PL or 5PL equation for the standard curve.

Report_View_4PL_Parameters

PD 035-308 STArStation Standard Curve Fitting

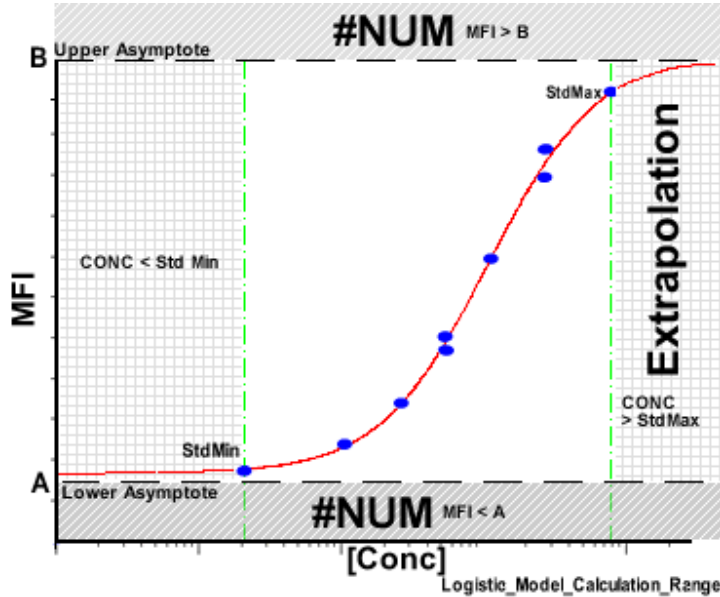
The values of the parameters A (lower asymptote) and B (upper asymptote) determine the calculation range of the equation (refer to the topic *The Calculation Ranges of the Four and Five Parameter Logistic Models* below).

The Calculation Ranges of the Four and Five Parameter Logistic Models.

The calculable ranges for the logistic equations in STArStation are defined by the upper and lower asymptotes.

Samples with MFI values below A have non-calculable concentrations and STArStation reports the concentration as #NUM (off-scale low).

Samples with MFI values above B have non-calculable concentrations and STArStation reports the concentration as #NUM (off-scale high).



Standard Curve for TNF-ALPHA (Four Parameter Logistic)
R squared: 0.9995

Summary	SampleID	Mean MFI	Mean [Conc]	CV% (n)
1	Std01 (0.00 pg/ml)	14.50	#NUM	NULL (0)
2	Std02 (10.00 pg/ml)	86.00	8.49	22.58 (
3	Std03 (30.00 pg/ml)	263.00	31.46	1.07 (2
4	Std04 (100.00 pg/ml)	947.00	105.65	1.96 (2
5	Std05 (300.00 pg/ml)	2676.00	294.41	0.85 (2
6	Std06 (1000.00 pg/ml)	6701.00	1006.21	2.43 (2
7	Std07 (3000.00 pg/ml)	9968.00	3011.06	12.36 (
8	Ctrl01 (1500.00 pg/ml)	5900.00	807.90	5.86 (2
9	Ctrl02 (1500.00 pg/ml)	5536.00	730.64	4.89 (2

Curve Fit Algorithm Parameters

A = 39, B = 1.2e+004, C = 8.5e+002, D = -1.2

OK

Review	SampleID	[Conc]	MFI	Read CV%	Description
1	Std01 (0.00 pg/ml)	#NUM	0.00		
2	Std01 (0.00 pg/ml)	#NUM	29.00		
3	Std02 (10.00 pg/ml)	7.14	77.00		
4	Std02 (10.00 pg/ml)	9.85	95.00		
5	Std03 (30.00 pg/ml)	31.22	261.00		
6	Std03 (30.00 pg/ml)	31.70	265.00		
7	Std04 (100.00 pg/ml)	107.12	961.00		

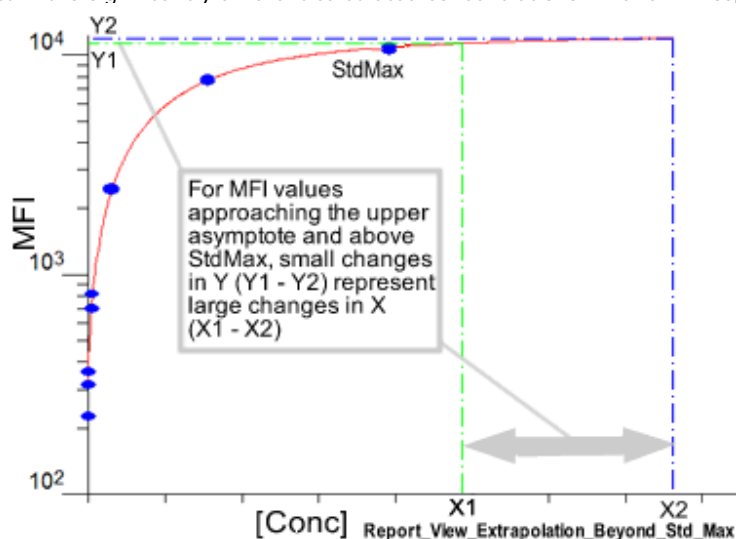
MFI values below the lower asymptote (A), are beyond the calculable range of the model and concentration values cannot be calculated. The Y value does not intercept the curve, STArStation reports the concentration as #NUM

For the example 4PL coefficients shown in the figure above, any MFI values below A (39) or above B (12000) are non-calculable. STArStation returns the value #NUM for any standards or samples with MFI values below A or above B.

A Spline curve will also report a calculated concentration as #NUM if the sample MFI value does not intercept the standard curve.

The 4-parameter, 5-parameter logistic and Spline models are also capable of extrapolating samples with MFI values above that of the highest included Standard (StdMax). The 4PL and 5PL models in STArStation are able to extrapolate beyond the highest standard concentration (StdMax), i.e. using the equation derived from the standard curve, samples that have an MFI value above StdMax but below the upper asymptote (coefficient B of the 4 and 5PL models) can be calculated.

However users should exercise caution when accepting samples with calculated concentrations above that of the back-calculated concentration of the highest standard. In the diagram below the reason for showing caution when calculating concentrations for MFI values above StdMax and which approach the upper asymptote is outlined. The two similar MFI values Y1 and Y2, which are both above StdMax, return the significantly different calculated concentrations X1 and X2 respectively.



Routinely this scenario occurs if the dynamic range of the standard curve is insufficient to quantify the range of anticipated sample analyte concentrations; this is typically corrected by extending the standard concentration range or by diluting the samples.

The Sample Range checking mechanism in STarStation can be used to indicate any samples that have been derived by extrapolation, by setting the Sample Range Checking Upper Limit to that of StdMax (the highest included standard). Any samples with the "More Than Expected" QC Comment have then been calculated by extrapolation. During the multiplex assay creation process when Sample Range checking is enabled the Upper Limit default setting is the concentration of the highest standard in the assay for the particular analyte).

Summary

Analyte concentrations in a Luminex assay are quantified from standard curves of known concentration vs. respective fluorescence. STarStation can use any one of four different formula to calculate standard curves; Spline interpolation, log-log regression, four parameter logistic and five parameter logistic curve fits.

Of the four curve fitting algorithms supplied by STarStation, when calculating the standard curve from a dynamic range of data the 4PL and 5PL curve fits are more applicable and the 5PL curve fit provides the most accuracy for asymmetric curve fits.

Standard data points may be given weighting in the 5PL algorithm to bias the curve fit in favour of specific data points or for data with normalised error.

STarStation also allows data to be excluded from calibration curves to account for "outlying" data. This allows users to ensure standard curves are only generated from reliable, reproducible data.

Further reading

Bevington, P.R. *Data Reduction and Error Analysis for the Physical Sciences*. New York: McGraw-Hill, 1969.

Lancaster, P. and Salkauskas, K. *Curve and Surface Fitting: An Introduction*. London: Academic Press, 1986.

Press, W. H.; et al. *Numerical Recipes in C*, 2nd ed.; Cambridge University Press: New York, 1992; 994 pp.; provides a detailed description of the Levenberg–Marquardt algorithm.