

UNITED STATES

Luminex Corporation
Austin, Texas
Tel: +1.512.219.8020
Fax: +1.512.219.5195

CANADA

Luminex Molecular Diagnostics
Toronto, Ontario
Tel: +1.416.593.4323
Fax: +1.416.593.1066

EUROPE

Luminex B.V.
Oosterhout, The Netherlands
Tel: +31.162.408333
Fax: +31.162.408337

www.luminexcorp.com

CHINA

Luminex Shanghai Trading Co.
Shanghai, China
Tel: +21.616.50809
Fax: +21.616.50811

info@luminexcorp.com

JAPAN

Luminex Japan Corporation, Ltd.
Tokyo, Japan
Tel: +81.3.5545.7440
Fax: +81.3.5545.0451

AUSTRALIA

BSD Robotics, A Luminex Company
Brisbane, Queensland
Tel: +61.7.3273.0273
Fax: +61.7.3273.0274

Luminex[®]



xPONENT[®]

xPONENT[®]
Logistic Curve Fitting
Technical Notes



Introduction

Immunoassays like all analytical methods give a fixed response to a given amount of analyte. The response units i.e. Median Fluorescent Intensity (MFI) must be converted into the proper units i.e. ng/ml. The conversion is achieved by the use of a standard curve. The standards that comprise the curve are a collection of samples of known and increasing concentration of the analyte in the matrix of interest. The standard curve is created with an equation relating assay response to sample concentration and is estimated from the responses of the standards.

Immunoassays are often characterized by non-linear relationships between the response and concentration of the analyte. Curve fitting is a technique used to express the relationship between quantitative response and explanatory variables for those non-linear relationships; for example the relationship between MFI and molecular concentration. This relationship can be modeled in the form of a mathematical function. The model can be approximated by a straight line (linear) or a "nth" order polynomial and other forms.

The actual response achieved for a given concentration of analyte may for many reasons vary with time, temperature, preparation, etc. The Luminex system allows for a very robust method to compensate for these variations by allowing the standardization process to be re-run every time data from a plate is acquired. This ensures that there is no significant effect on results due to the previously mentioned variables.

Luminex® xPONENT software is can utilize several different curve fitting formulas. The most popular are the Linear and Logistic regression methods.

Linear vs. Logistic Regression

Although linear regression requires fewer data points or standards (as few as 3) compared to logistic regression (5PL requires 6 data points); a more accurate fit is obtained by using at least 6 points for any of the regression types (Motulsky 1996). Linear regression is rarely accurate enough for most antibody binding processes. The nature of the binding process depending on the chemistry involved is often significantly non-linear.



Figure A. Example of Linear Regression for the same immunoassay data as regressed in Figure B

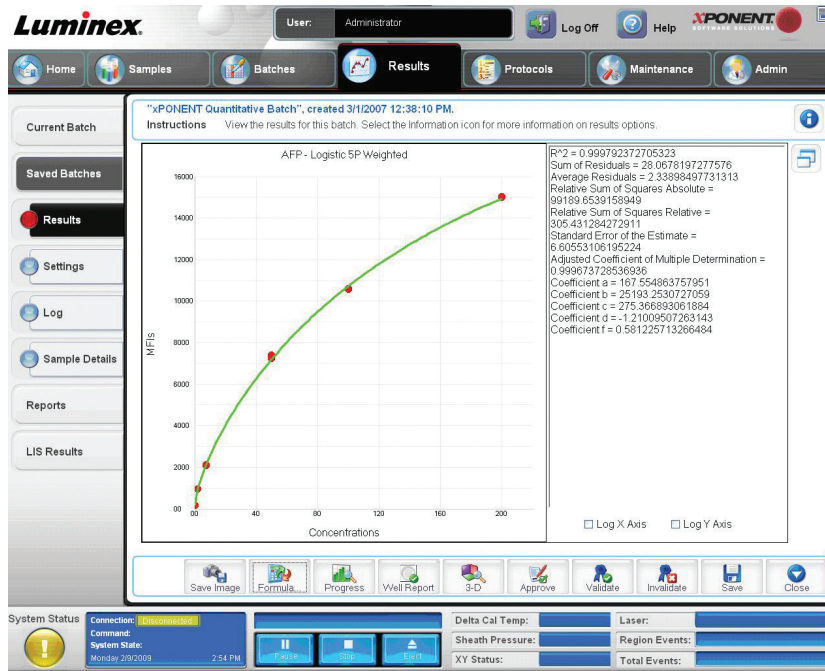


Figure B. Example of Logistic Regression for the same immunoassay data as regressed in Figure A

Figure A and B are respective examples of Linear and 5PL logistic curve fits for the same immunoassay data. It is obvious that the linear fit (Figure A) does not express the actual assay reaction properly and significant error in calculating actual concentration from the MFI response of an analyte results.

The goal of nonlinear regression or specifically Logistic Regression is to determine the best-fit parameters for a model by minimizing a chosen merit function. Where nonlinear regression differs from linear is that the model has a nonlinear dependence on the unknown parameters, and the process of merit function minimization is an iterative approach. The process starts with some initial estimates and then incorporates algorithms to improve the estimates iteratively. The new estimates then become a starting point for the next iteration. These iterations continue until the merit function effectively stops decreasing.

The nonlinear model to be fitted is represented by:

$$y = y(x; a)$$

The merit function minimized in the regression is the following:

$$\chi^2(a) = \sum_{i=1}^N \left[\frac{y_i - y(x_i; a)}{\sigma_i} \right]^2$$

Where σ_i is the measurement error or standard deviation of the i th data point. We are attempting to minimize the sum of the squares of the distances between the actual data points and the regression line.

Nonlinear regression iterations proceed as follows:

1. Obtain initial estimates for all of the variables being fitted for in the model.
2. Using the initial estimates, compute the merit function.
3. Use an algorithm to adjust the variables in order to improve the fit of the model to the data points. xPONENT utilizes the Levenberg-Marquardt method.
4. Again, compute the merit function and compare it to the previous iteration.
5. Repeat steps 3 and 4 until there is essentially no change in the merit function, then cease the iterations.
6. Calculate the goodness of fit statistics.

Why is this a better method than linear regression? To start with, it is a much more general procedure. There are a very limited number of models that can be expressed in linear form without transforming the data. Also, remember that transforming the data means that the fitting routine will be minimizing the merit function on the transformed data, not the actual data.

Goodness of Fit Statistics:

It is critical to be able to determine when a fit is good enough. Statistical metrics are produced by xPONENT to provide a detailed view into the goodness of the curve fit. xPONENT produces 7 key metrics as shown in Figure C.

xPONENT Logistic Regression Statistical Metrics

1. R² (R squared) or Coefficient of Multiple Determination:

R² measures the proportion of variation in the data points which is explained by the regression model.

$$R^2 = \frac{\text{explained variation}}{\text{total variation}} = \frac{\sum (Y_{est} - \bar{Y})^2}{\sum (Y - \bar{Y})^2}$$

Where Y are the observed values for the dependent variable; \bar{Y} is the average of the observed values and Y_{est} are predicted values for the dependent variable (the predicted values are calculated using the regression equation).

For example if R²= 0.97, then 97% of the variation in the dependent or response variable is explained by the regression model. In other words 97% of the variation of the concentration is explained by the variation in the response (MFI). A value of R²=1.0 means that the curve passes through every standard data point. A value of R² = 0.0 means that the

regression model does not describe the data any better than a horizontal line passing through the average of the data points. R² is not a good metric for distinguishing the goodness of a fit because even very poorly fitting curves can have a relatively high R². In Figure A, the R² of the linear regression is 0.94, but obviously the fit is quite poor since most of the standard points do not lie on the curve represented by a straight line.

2. Sum of Residuals: This is the total sum of the residuals for all data points. If the curve passed through each data point, the sum of residuals would be zero. A regression model can have large positive and negative residuals and still sum to a small number so this metric alone is not sufficient to determine the goodness of a fit.

3. Average of Residuals: This is just the average of the residuals.

4. RSSE: Residual or often called “Relative” sum of squares unlike R² measures how well the curve fits the individual data points.

If the data is weighted, the RSSE is X² distributed and this probability makes a good metric. xPONENT will display the RSSE in two forms as shown in the below with their equations.

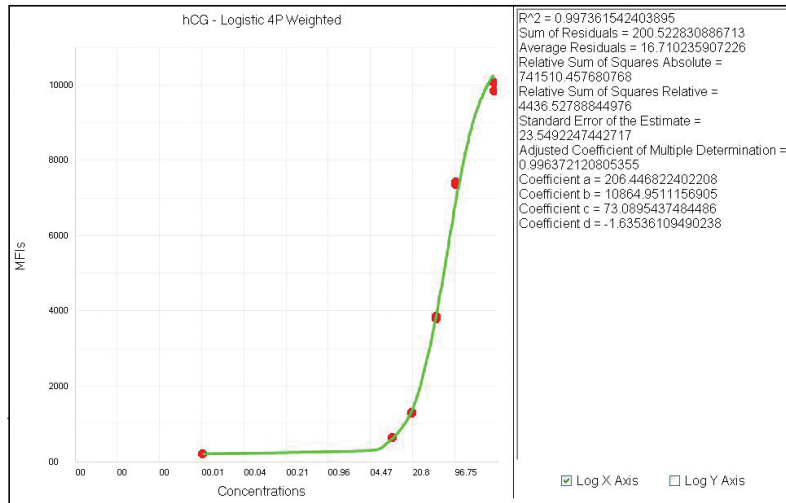


Figure C. Logistic Regression with 7 key metrics

Pay close attention to this number as it is a good indicator of fit accuracy. Two versions of this metric are calculated within xPONENT as described here below. Y_i is the actual measured MFI for a particular standard and \hat{Y}_i is the predicted standard MFI.

Residual or Error Sum of Squares (Absolute) =

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Residual or Error Sum of Squares (Relative) =

$$\sum_{i=1}^n [(Y_i - \hat{Y}_i)^2 * W_i]$$

Where: $W_i = \frac{1.0}{\sigma_i^2}$ normalized so that $\sum_{i=1}^n W_i = n$

σ_i = the standard deviation of the i th data point Y_i
 n is the number of data points or observation

a. Relative Sum of Squares Absolute: The Relative (or Residual) Sum of Squares (RSS), is the sum of the squares of the differences between the entered data and the curve generated from the fitted regression model. A perfect fit would yield a relative sum of squares of 0.0.

b. Relative Sum of Squares Relative: The Relative (or Residual) Sum of Squares (RSS), is the weighted sum of the squares of the differences between the entered data and the curve generated from the fitted regression model. A perfect fit would yield a relative sum of squares of 0.0. If the regression was not weighted the relative and absolute sum of squares will be equal.

5. **Standard Error of the Estimate:** The Standard Error of the Estimate is the standard deviation of the differences between the entered data and the curve generated from the fitted model. This gives you an idea about how scattered the residuals are around the average. As the standard error approaches 0.0, you can be more certain that the regression model accurately describes the data. A perfect fit would yield a standard error of 0.0.

6. **Adjusted Coefficient of Multiple Determination:** This is used to balance the cost of using a model with more parameters against the increase in R^2 .

4PL vs 5PL Logistic regression non-linear models:

The 4PL logistic method works best when the standard curve is symmetrical around its mid-point. When this requirement is not fulfilled the 4PL may give a poor fit to the data particularly around the asymptotes. The more general form of the logistic equation called a five parameter logistic (5PL) overcomes this problem.

$$\text{Logistic 4P: } y = a + b / (1 + (x/c)^d)$$

$$\text{Logistic 5P: } y = a + (b / (1 + (x/c)^d))^f$$

- a - Estimated response at zero concentration
- b - Estimated response at infinite concentration
- c - Mid-range concentration (EC50)
- d - Slope factor (Hill Slope)
- f - Asymmetry factor

The f parameter present in the 5PL model but not in the 4PL allows for asymmetry correction when the curve response is not similar around the mid-point of the concentration range.

Weighting the Curve:

Variance due to instrumentation variations, processing variations, and other unknown sources of influence are usually not evenly distributed through the standard curve. It is quite common in immunoassay and bioassay data for the error variance to be proportional to the magnitude of the response so that the errors are much larger where the responses (in our case MFI) are greater. In many cases this is at the top of the curve where the concentration measurements are the highest. Conversely the errors are often smaller where the response is small i.e. where the concentrations are closest to zero and the observed response is closer to zero.

An important reason that the variance of a standard is often a function of the response is because the kinetics associated with antibody binding are not linear so that the kinetic variations in the reaction change as the ratio of the analyte to tracer and binder changes (Wild 1994). This effect is known as heteroscedasticity and is evident in all fields of chemical and bioassays.

By adjusting the response curve to compensate for this effect, substantially more accurate results can be achieved. xPONENT provides the popular $1/y^2$ weighting schema as a user configurable option. Weights are assigned to each standard

data point in order to compensate for the non-linear reaction characteristics. Points on the lower part of the curve will then have a more significant effect on the curve shape since these are considered more accurate while those on the higher point have less of an effect since these points are more error prone.

Extrapolation:

Extrapolation is required when concentrations must be estimated for points that are within calculatable limits but are outside of the range of standards inside of our standard curve.

Specifically the case of calculated concentration being less than the minimum standard expected concentration or if the calculated concentration is greater than the maximum standard expected concentration. Extrapolation can be a poor estimate of actual response especially since as the curve flattens small changes in the MFI values measured can result in very large changes in concentration.

Figure D illustrates several examples of extrapolation and interpolation with specific references as to how these instances are described on the xPONENT data grid and standard curve chart.

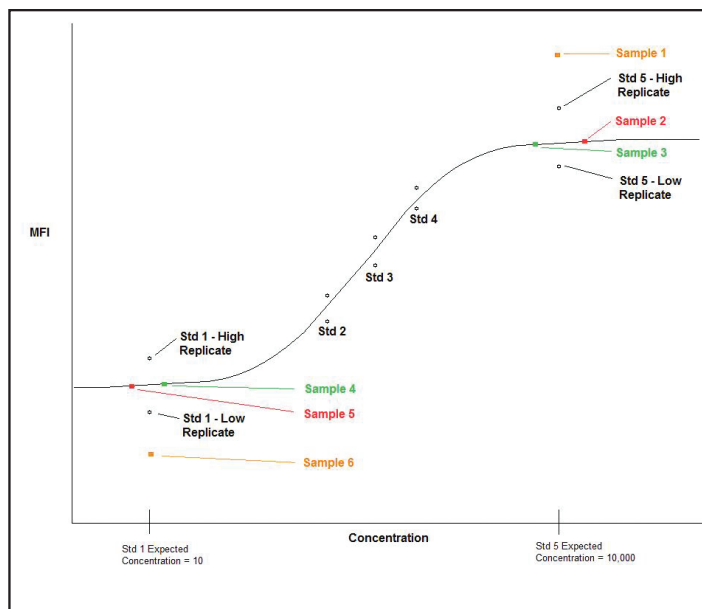


Figure D. Extrapolations and Interpolations

Extrapolation Interpolation Sample Explanation

As seen in Figure D

Sample 1 – The MFI does not intersect the curve (the MFI value is above the high asymptote of the curve); therefore, a numerical concentration result cannot be calculated. Instead, the result will be reported as “> (High Std Concentration)”, such as “>10,000”.

Sample 2 - The MFI intersects the curve; therefore, a numerical concentration result will be calculated, such as 10,707 pg/ml. However, because the MFI of the sample is greater than the lowest replicate of the high standard, the sample concentration result will be flagged as **extrapolated**. Also note that depending on where the intersection occurs, it is possible to obtain results that are numerically greater than the expected concentration of the high standard. For example, in this case a result of 10,707 was obtained, even though the expected concentration of the high standard is only 10,000. Also note, that for samples that do not intersect the curve, such as “Sample 1”, the result will be recorded as “>10,000”. Thus it is expected that some results may be reported with actual numbers greater than the expected concentration of the high standard, while other results may be reported simply as “> (High Std Concentration)”.

Sample 3 – The MFI intersects the curve; therefore, a numerical concentration result will be calculated, such as 9,599 pg/ml. However, because the MFI of the sample is greater than the lowest replicate of the high standard, the sample concentration result will be flagged as **extrapolated**.

Sample 4 - The MFI intersects the curve; therefore, a numerical concentration result will be calculated, such as 11.9 pg/ml. However, because the MFI of the sample is less than

the highest replicate of the low standard, the sample concentration result will be flagged as **extrapolated**.

Sample 5 – The MFI intersects the curve; therefore, a numerical concentration result will be calculated, such as 8.3 pg/ml. However, because the MFI of the sample is less than the highest replicate of the low standard, the sample concentration result will be flagged as **extrapolated**. Also note that depending on where the intersection occurs, it is possible to obtain results that are numerically less than the expected concentration of the low standard. For example, in this case a result of 8.3 was obtained, even though the expected concentration of the low standard is only 10. Also note, that for samples that do not intersect the curve, such as “Sample 6”, the result will be recorded as “<10”. Thus it is expected that some results may be reported with actual numbers less than the expected concentration of the low standard, while other results may be reported simply as “< (Low Std Concentration)”.

Sample 6 –The MFI does not intersect the curve (the MFI value is below the low asymptote of the curve); therefore, a numerical concentration result cannot be calculated. Instead, the result will be reported as “< (Low Std Concentration)” such as “<10 pg/ml”.

Note: For samples in which the bead count does not meet the specified “Minimum bead count for obtaining results” value found in the “Batch Option” section of the Admin page, the result will be recorded as “NaN”, or Not a Number.

References:

Motulsky, Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting, US, 2004

Wild, David (Eds), *The Immunoassay Handbook*, Stockton, New York, 1994.