Nonlinear LOB and LOD estimation

Cyril Galitzine
c.galitzine@neu.edu
Vitek Lab

Northeastern

Outline

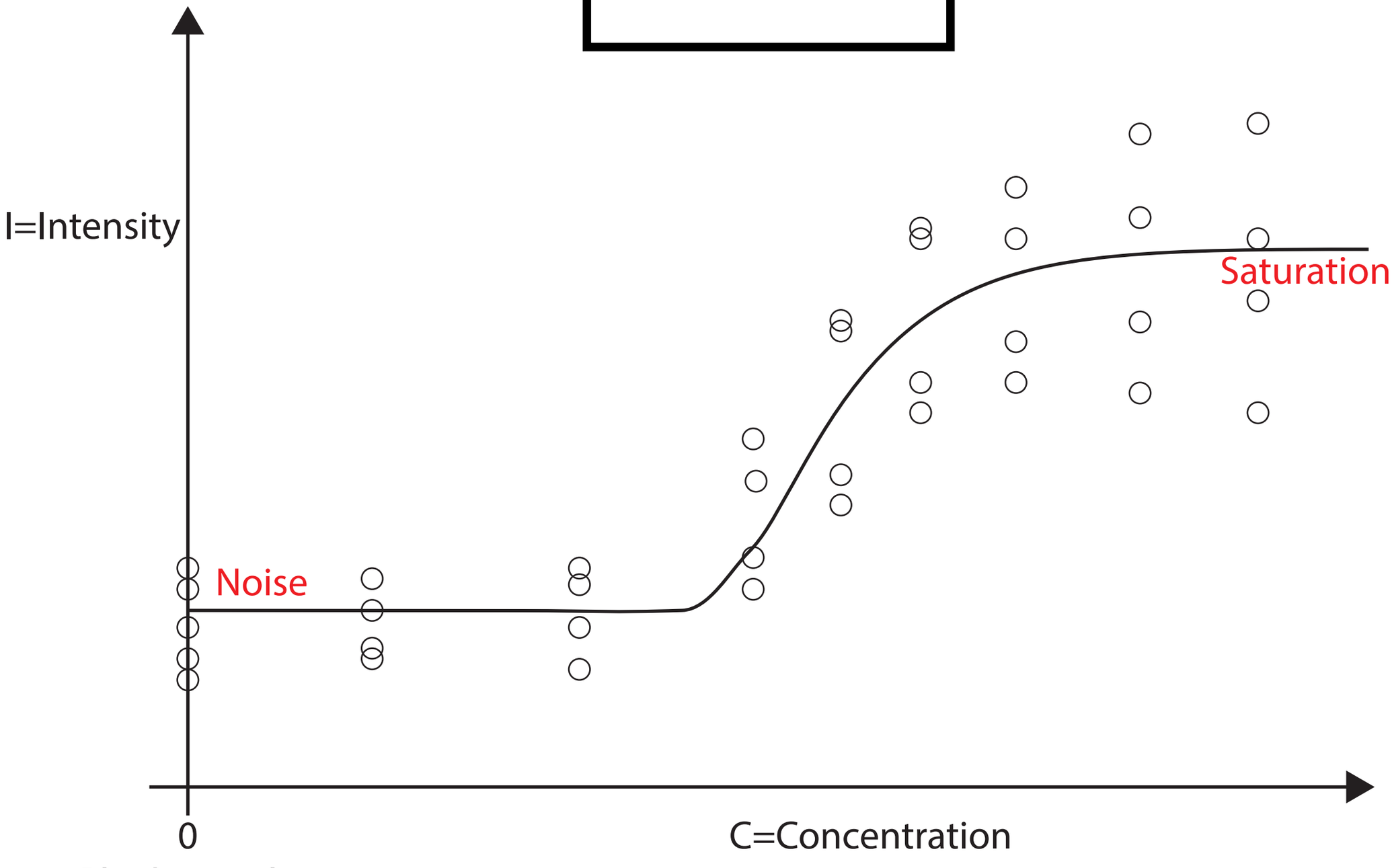
1. Basics of curve fit methods
2. An R package to calculate the LOB and LOD: MSstats

Response curve

<u>Input</u> (=known)

Protein concentration → LC-MS/MS → <u>Output</u>

Signal Intensity



LOB/LOD/LOQ definition

Will focus on curve fit methods

Clinical Chemistry 50:4
732–740 (2004)

General Clinical Chemistry

Partly Nonparametric Approach for Determining the Limit of Detection

KRISTIAN LINNET[1*] and MARINA KONDRATOVICH[2]

A STATISTICAL OVERVIEW ON UNIVARIATE CALIBRATION, INVERSE REGRESSION, AND DETECTION LIMITS: APPLICATION TO GAS CHROMATOGRAPHY/MASS SPECTROMETRY TECHNIQUE

Irma Lavagnini and Franco Magno*
Dipartimento di Scienze Chimiche, Università di Padova, via Marzolo 1, 35131 Padova, Italy

Received 8 March 2006; accepted 21 April 2006

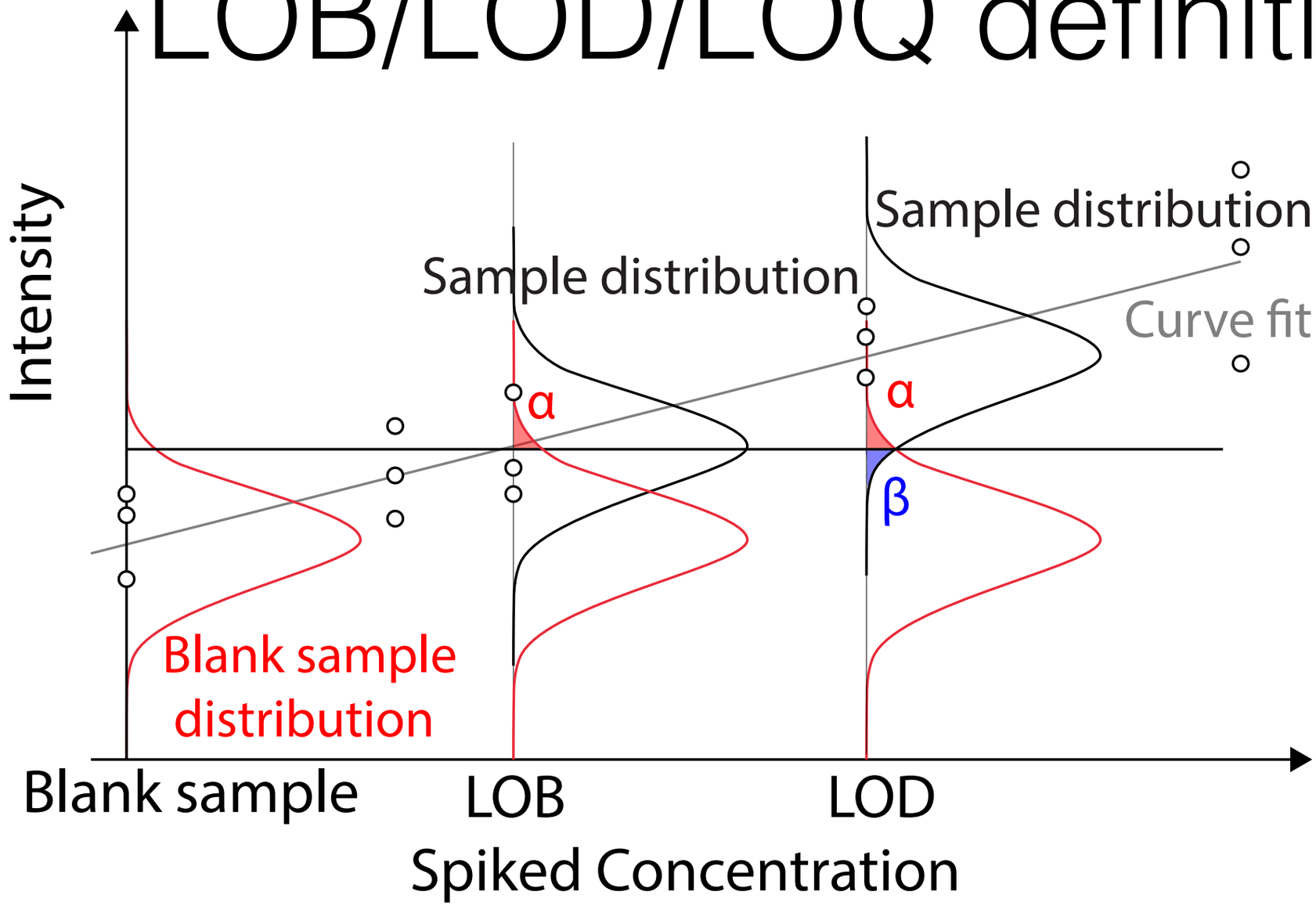
Anal. Chem. **1997,** *69,* 3069−3075

Weighted Least-Squares Approach To Calculating Limits of Detection and Quantification by Modeling Variability as a Function of Concentration

Michael E. Zorn,[†] Robert D. Gibbons,[‡] and William C. Sonzogni*,[†]

Water Chemistry Program, University of Wisconsin—Madison, Madison, Wisconsin 53706, and Department of Biostatistics, University of Illinois—Chicago, Chicago, Illinois 60612

LOB/LOD/LOQ definition



- LOB = Limit of Blank = Min. concentration above which false positive < α
- LOD = Limit of Detection = Min. concentration above which false negative < α and false positive < β

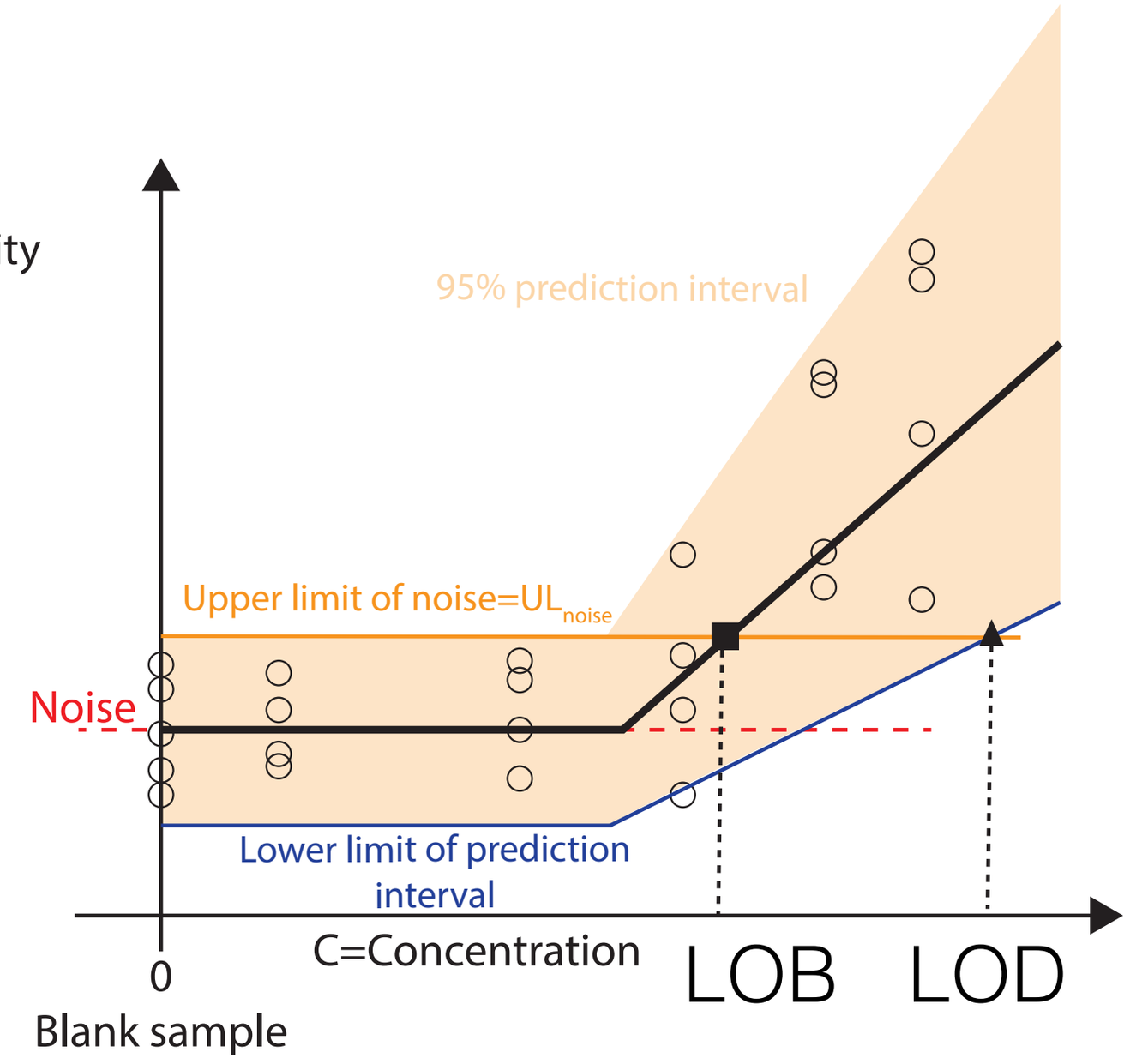
Usually α=β=5%

LOB/LOD/LOQ definition

- LOQ = Limit of Quantification = Min. concentration above at which the assay is able to provide quantitative results of a stated analytical quality
- No “universal” definition exists for the LOQ
  - ✳CV= mean/sd < 0.2
  - ✳Recovery factor < 0.2
  - ✳Methods based on blank sample

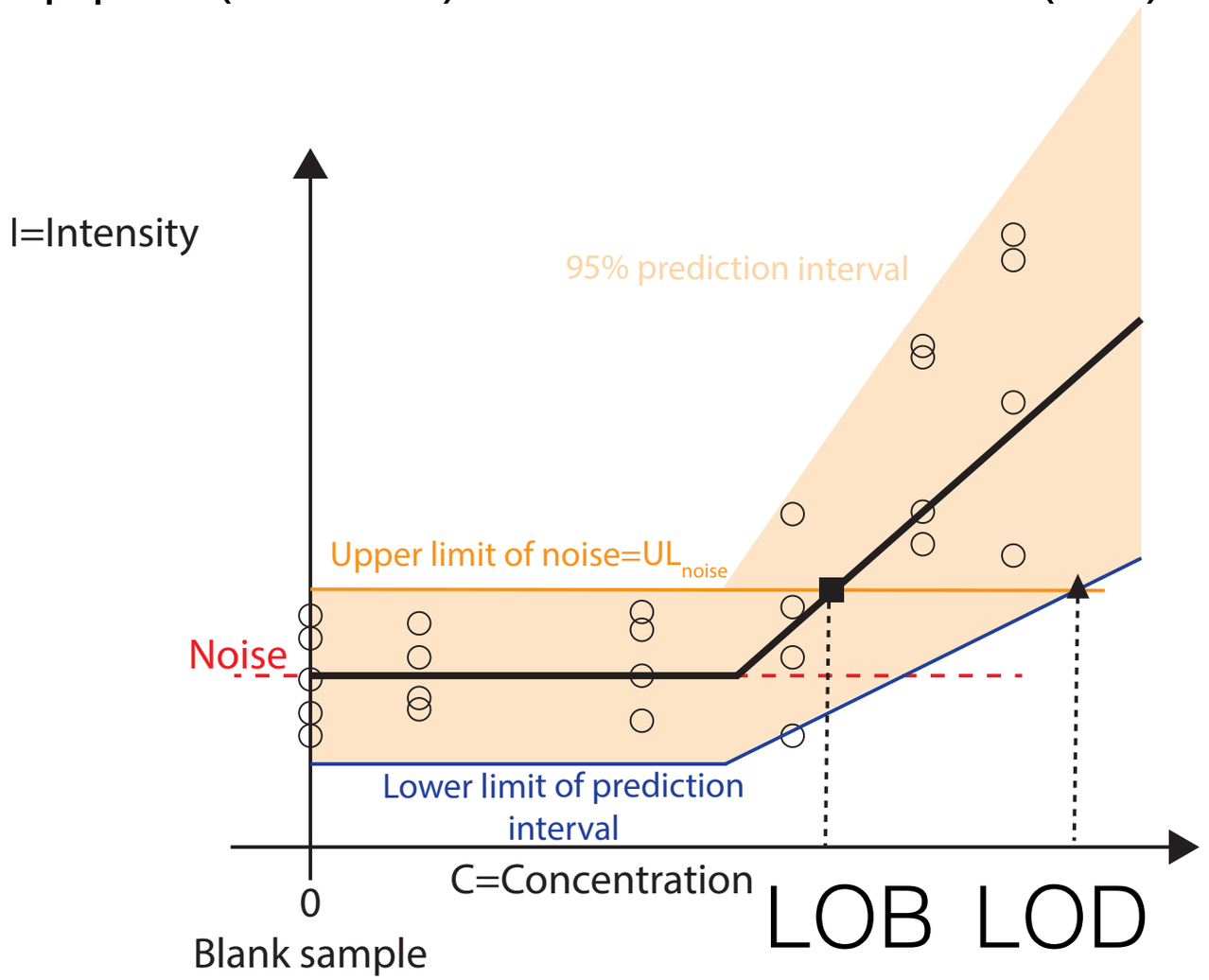
LOD/LOQ definition

- LOB = Concentration when Intensity > Noise “on average”
- LOD = Concentration when Intensity > Noise “always”



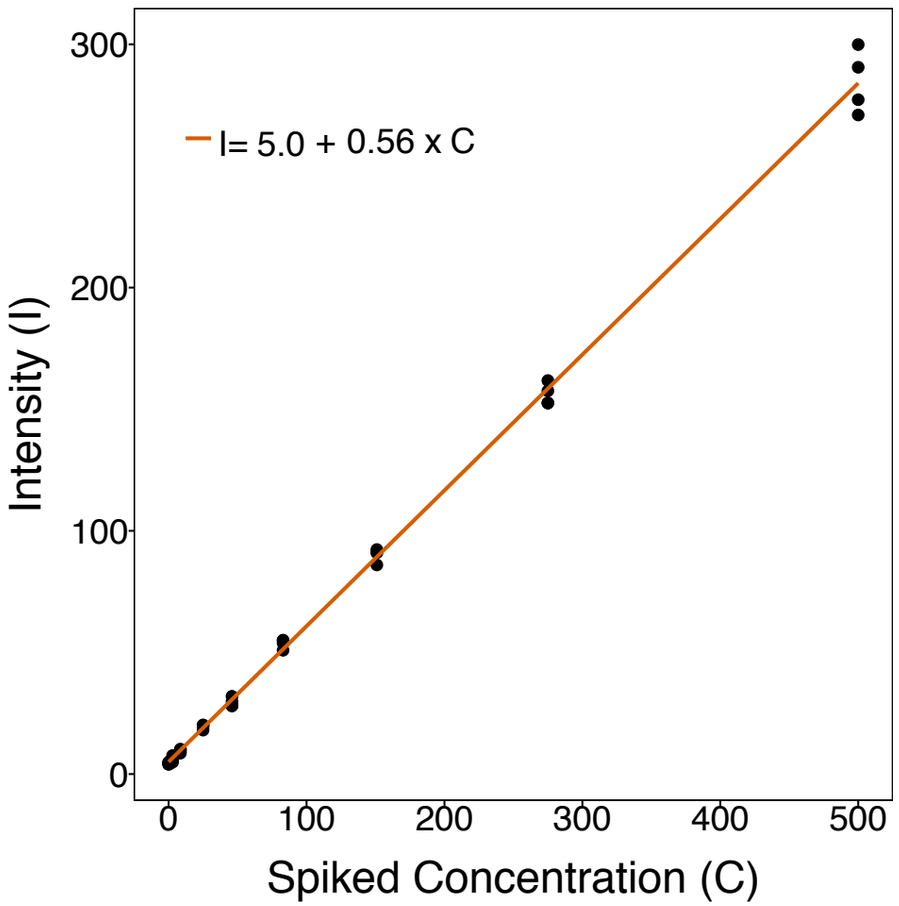
General LOB/LOD calculation procedure

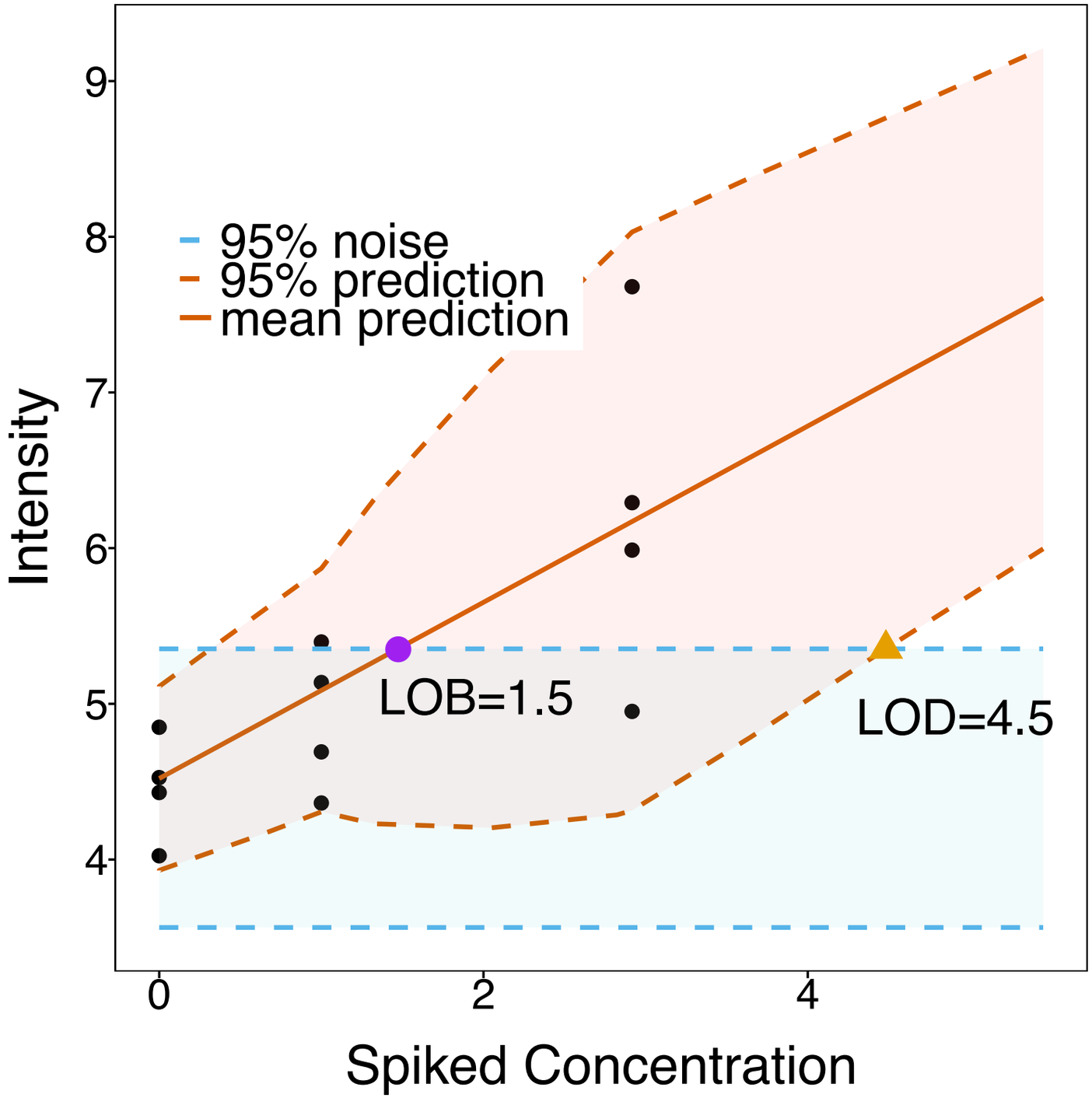
1. Normalize data (e.g. with heavy peptides)
2. Do a curve fit of Intensity ~ Concentration
3. Calculate prediction interval of fit
4. Estimate noise of blank samples
5. LOB = Fit ∩ Upper(Noise) LOD = Lower(Fit) ∩ Upper(Noise)



Linear LOD/LOQ calculation method

- Based on a weighted linear fit
- Prediction interval estimated by resampling data

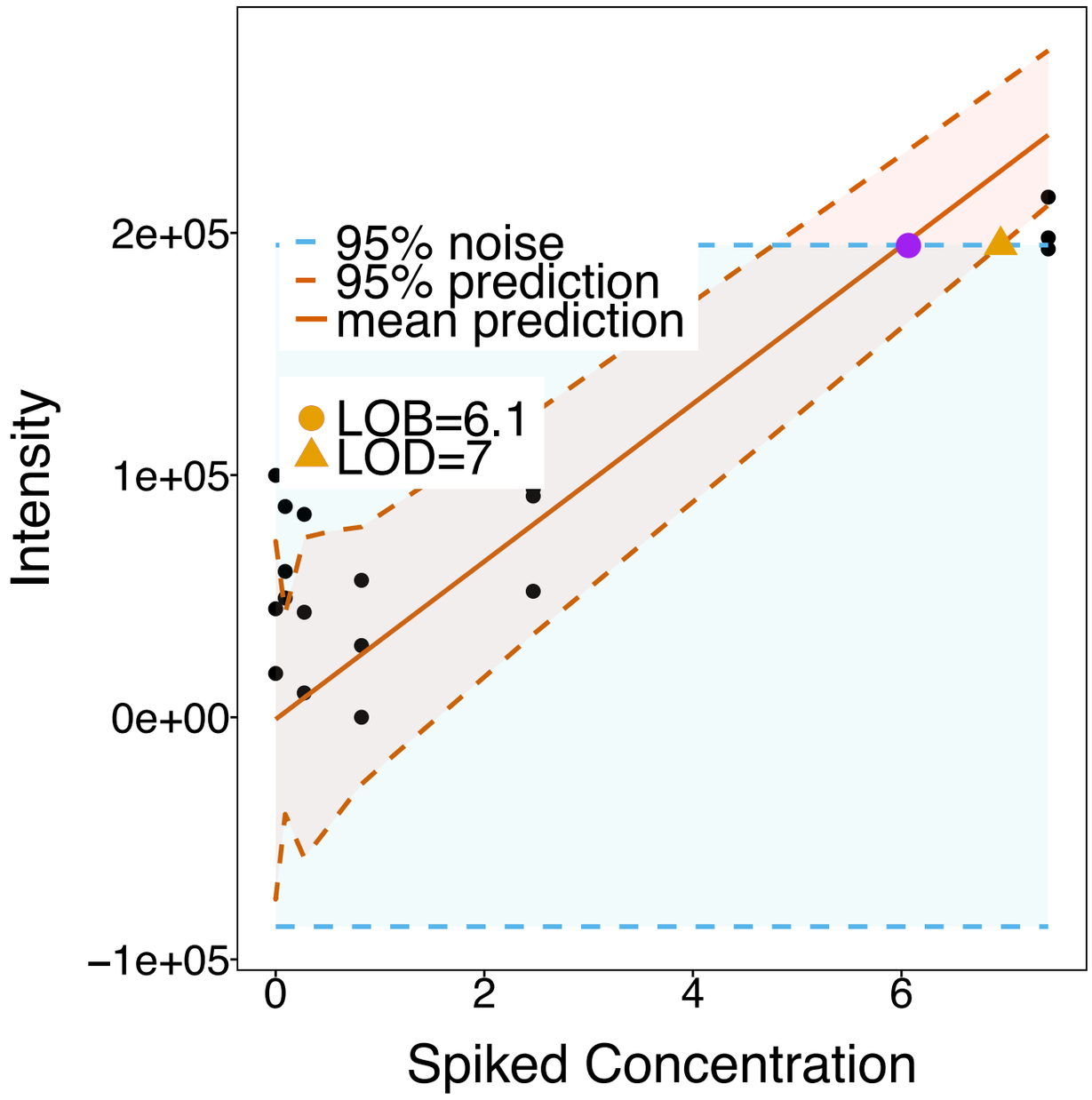




CPTAC 7, Peptide ESDTSYYVSLK, Site 86, Study IIIc

Linear LOB/LOD calculation method

- Does not capture threshold (=leveling off at low concentrations)
- Curve fit, LOB and LOD are clearly not accurate



Peptide HGGEDYVFSLLTGYCEPPTGVSLR of study CellLysate_5500QTRAP_directMRM

Nonlinear LOB/LOD calculation methods

- Multiple methods developed to remedy problems with linear fits

Drug Development and Industrial Pharmacy, 26(6), 661–669 (2000)

RESEARCH PAPER

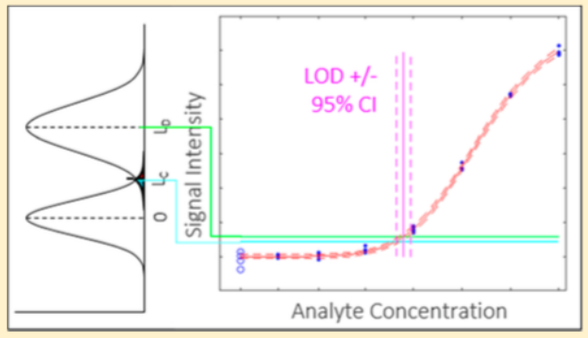
Statistical Method for Determining and Comparing Limits of Detection of Bioassays

Carly A. Holstein,*,† Maryclare Griffin,‡ Jing Hong,‡ and Paul D. Sampson‡

†Department of Bioengineering, University of Washington, Box 355061, Seattle, Washington 98195, United States
‡Department of Statistics, University of Washington, Box 354322, Seattle, Washington 98195, United States

Supporting Information

ABSTRACT: The current bioassay development literature lacks the use of statistically robust methods for calculating the limit of detection of a given assay. Instead, researchers often employ simple methods that provide a rough estimate of the limit of detection, often without a measure of the confidence in the estimate. This scarcity of robust methods is likely due to a realistic preference for simple and accessible methods and to a lack of such methods that have reduced the concepts of limit of detection theory to practice for the specific application of bioassays. Here, we have developed a method for determining limits of detection for bioassays that is statistically robust and reduced to practice in a clear and accessible manner geared at researchers, not statisticians. This method utilizes a four-



Calibration and LOD/LOQ Estimation of a Chemiluminescent Hybridization Assay for Residual DNA in Recombinant Protein Drugs Expressed in *E. coli* Using a Four-Parameter Logistic Model

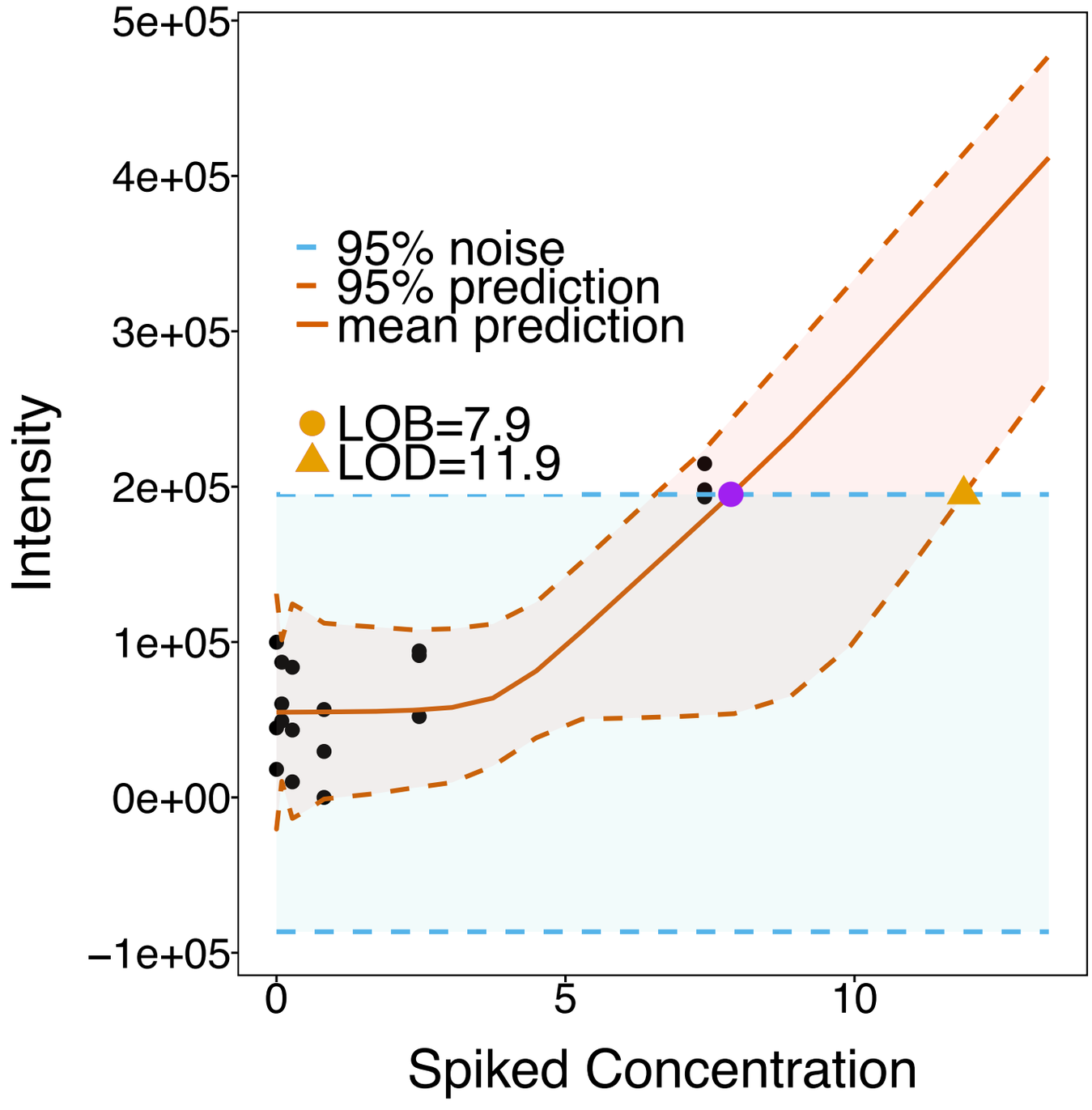
Kwan R. Lee,[1,*] Byron Dipaolo,[2] and Xiaoying Ji[2]

[1]*Statistical Sciences Department and* [2]*Bioanalytical Sciences Department, SmithKline Beecham Pharmaceuticals R&D, King of Prussia, PA 19406*

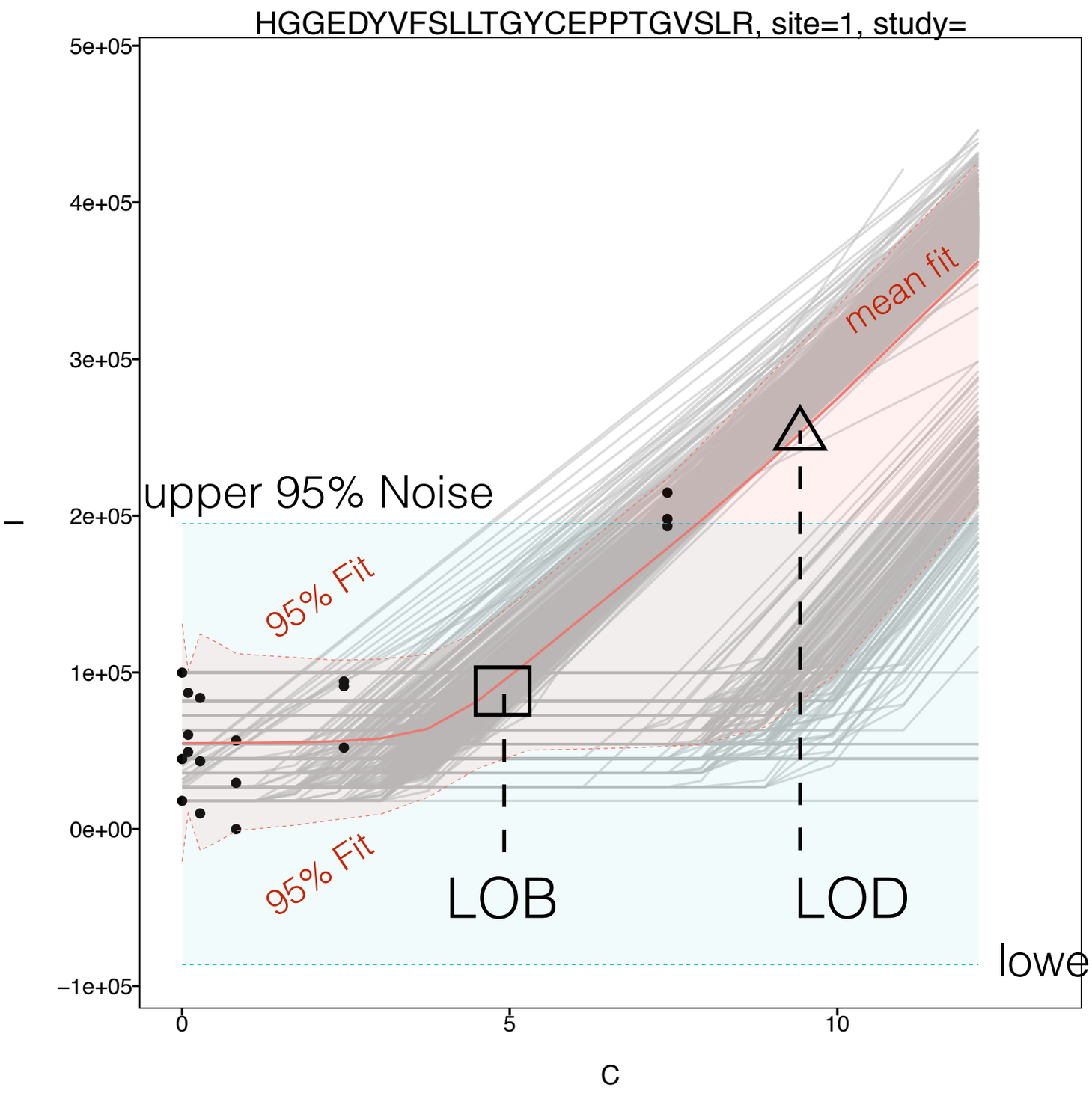
- Do not however take canonical shape of response into account
  - ✳Overfitting
  - ✳Numerical issues (convergence etc.)

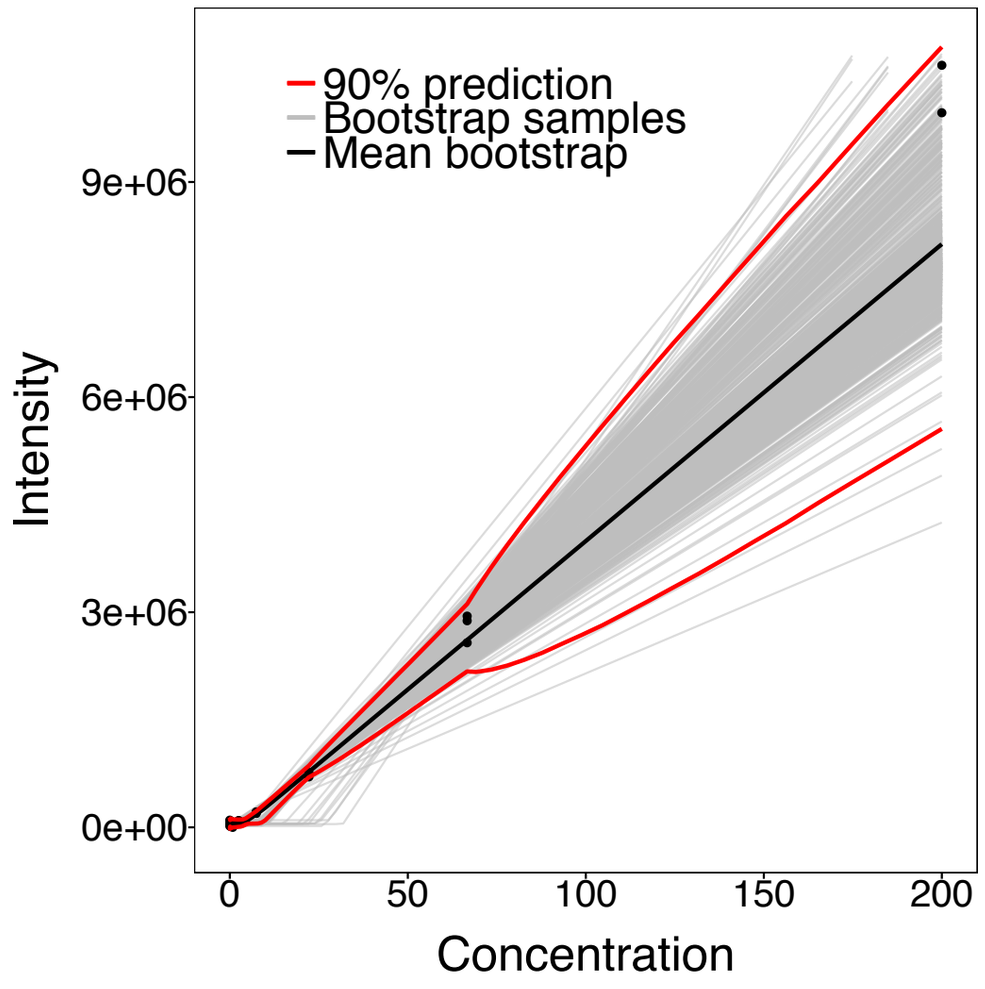
A new Nonlinear LOB/LOD calculation method

- More advanced method recently developed by the Vitek Lab
- Captures threshold at low concentrations but linear when no threshold is present
- Avoids problems with logistic regression



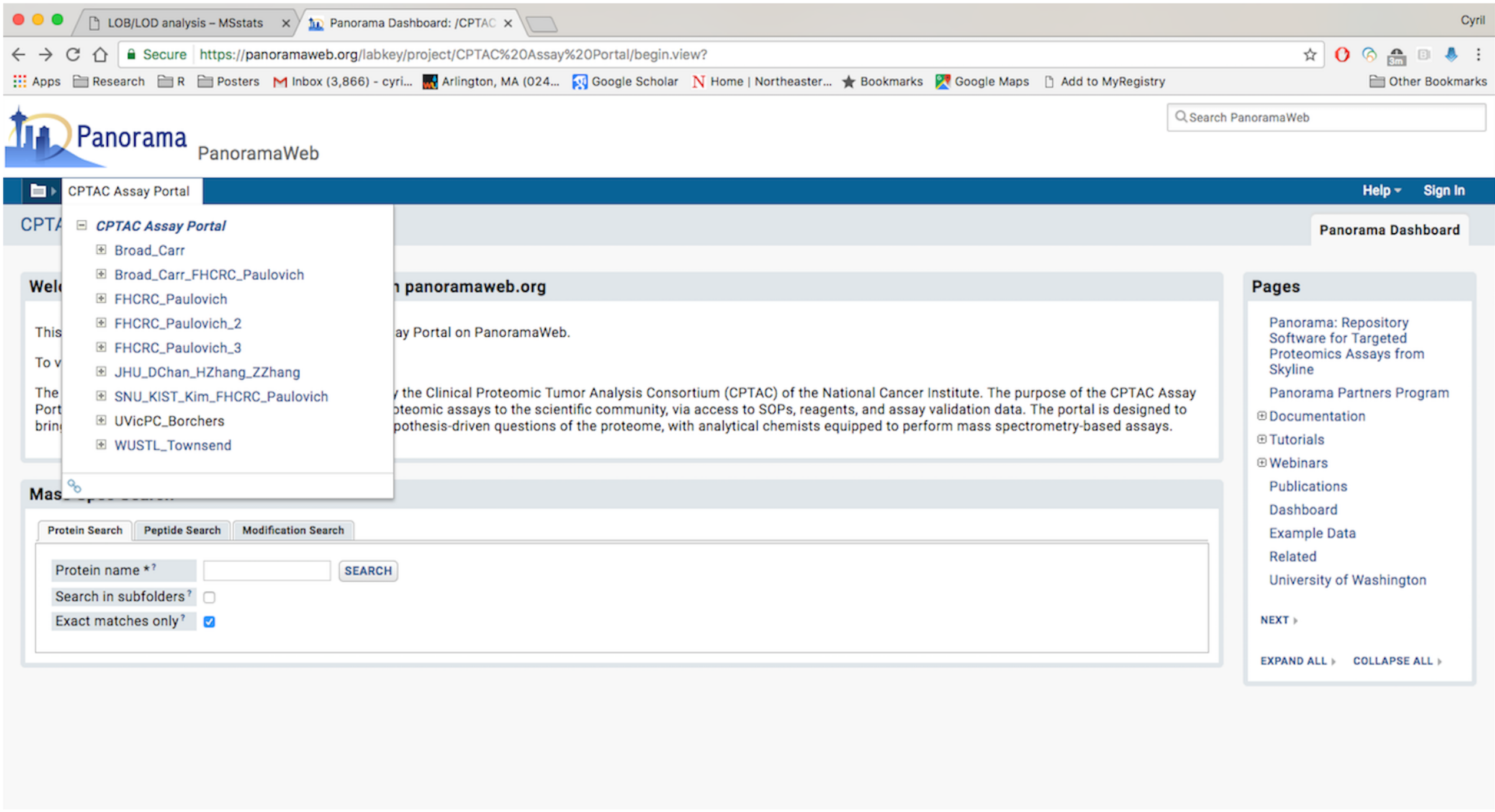
A new Nonlinear LOB/LOD calculation method





Blank sample

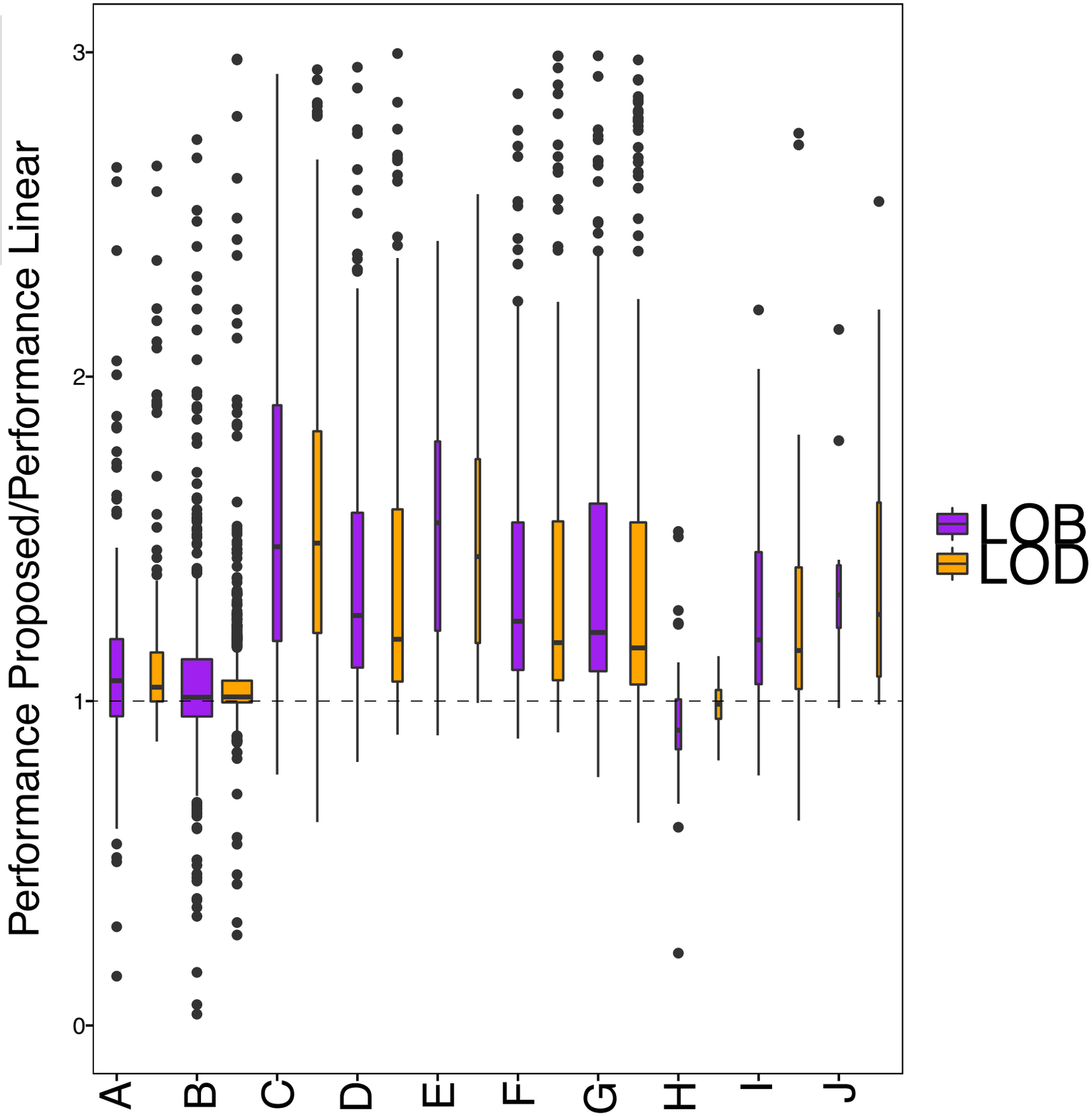
A new Nonlinear LOB/LOD calculation method: Results



A new Nonlinear LOB/LOD calculation method: Results

10 datasets from CPTAC assay portal

Nonlinear method corrects underestimation of LOD and LOB with linear method



A nonlinear LOB/LOD calculation method: R package

Part of MSstats R package: http://msstats.org

MSstats

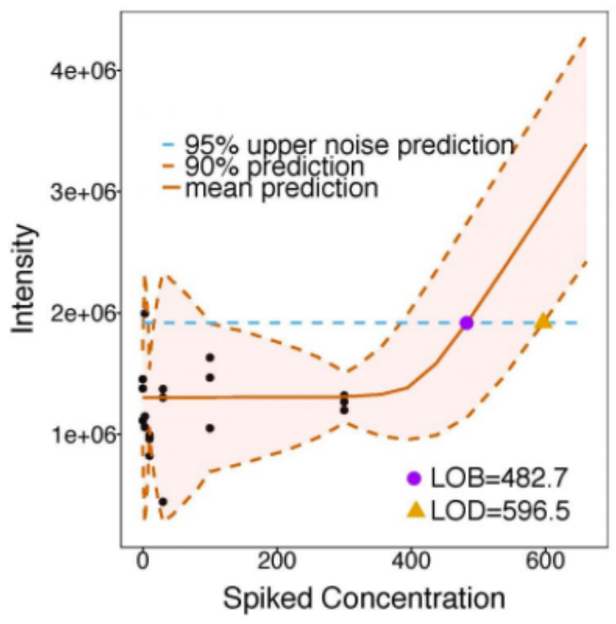
Statistical Tool For Quantitative Mass Spectrometry-Based Proteomics

HOME MSSTATS LOB/LOD MSSTATSQC DATASETS TRAINING CONTACT

LOB/LOD ANALYSIS

Assay characterization : estimation of limit of blanc (LoB) and limit of detection (LoD)

The need for assay characterization is ubiquitous in quantitative mass spectrometry-based proteomics. Although many assay characteristics exist, the limit of blank (LOB) and limit of detection (LOD) are particularly useful figures of merit. LOB and LOD are determined by repeatedly measuring the peak intensities of peptides in samples with known peptide concentrations, and deriving an intensity versus concentration response curve. Most commonly, a weighted linear regression is fit to the intensity-concentration response, and LOB and LOD are estimated from the fit. Linear methods, however, inaccurately characterize assays containing a noise threshold at low concentrations, which is a very common situation. We propose a new approach based on non-linear regression that correctly captures the noise threshold. In absence of a noise threshold, the estimates of LOB/LOD obtained with non-linear statistical modeling are identical to those of weighted linear regression. However, in presence of a noise threshold the non-linear model changed the estimates of LOB/LOD by up to 20-40%. It improved the accuracy of the results, and avoided the unduly optimistic estimation of these figures of merit. We implemented the non-linear regression approach in the open-source R-based software MSstats, and advocate its general use for mass spectrometric protein assay characterization.



MSstats R package: linear fit for LOB/LOD calculation

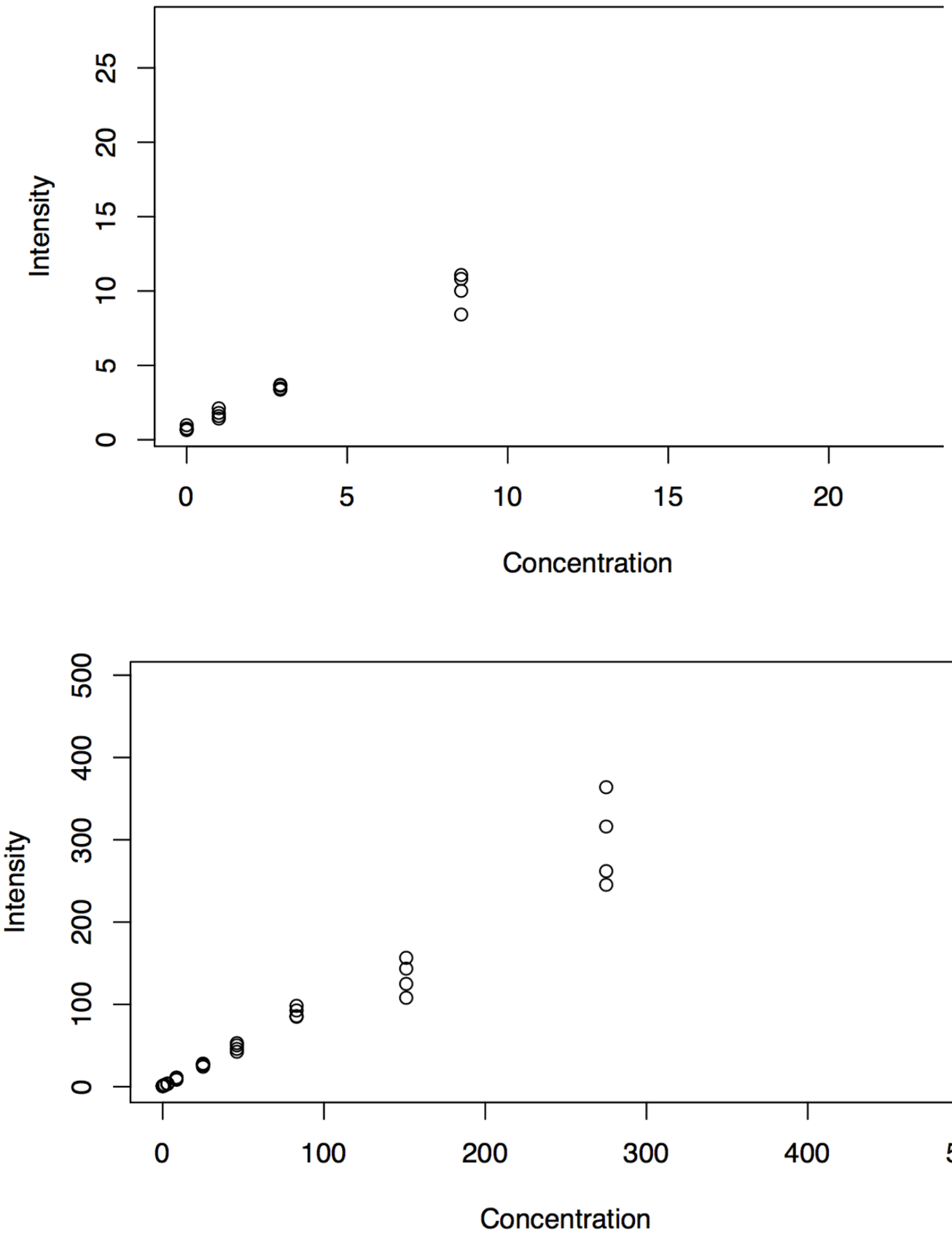
Input data set:

test.R | LOD_workflow.R | spikeindata

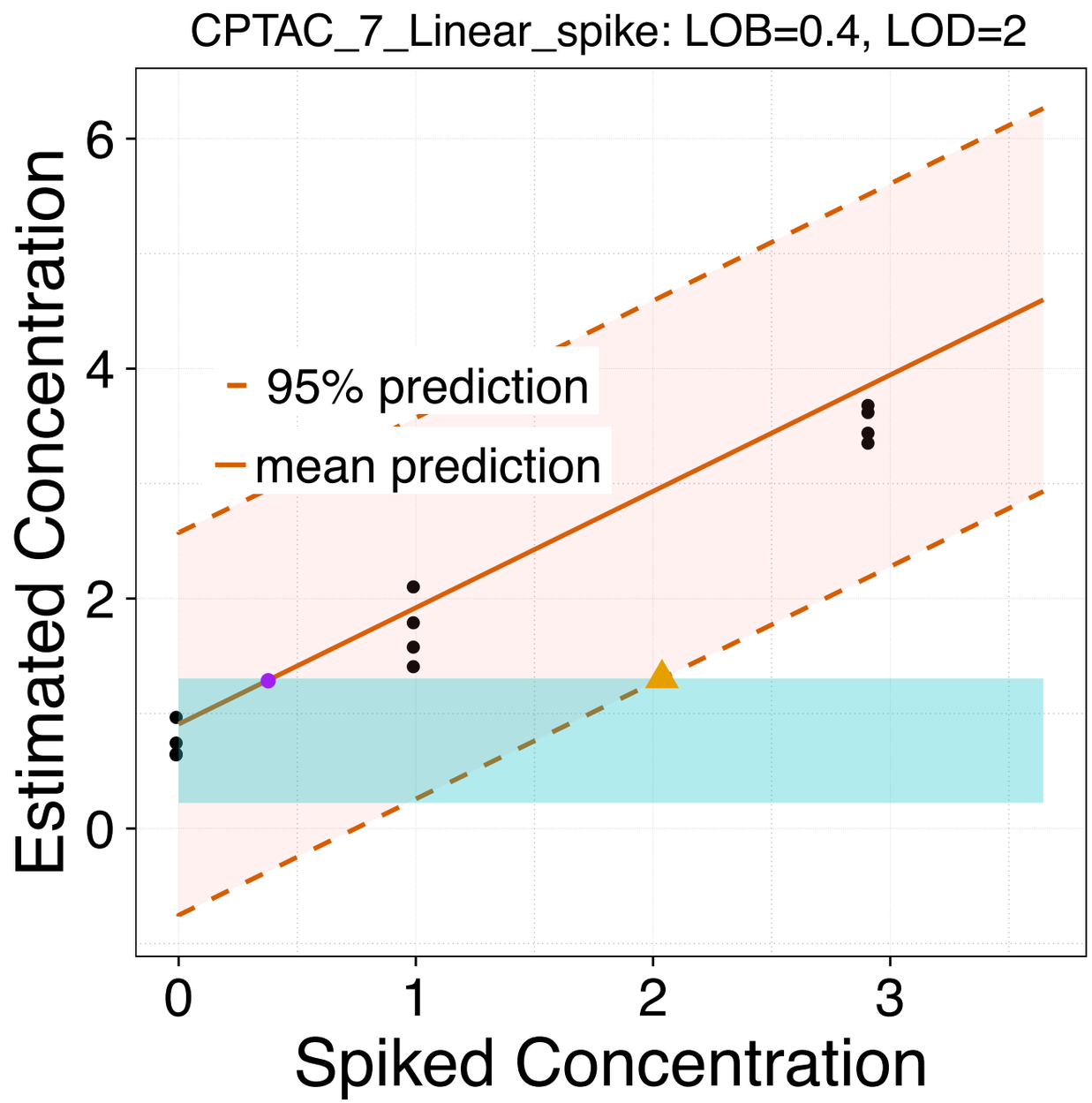
Filter

	INTENSITY	CONCENTRATION	NAME	REPLICATE
1	2.629110e+04	0.0576	LPPGLLANFTLLR	1
2	2.448406e+05	0.2880	LPPGLLANFTLLR	1
3	0.000000e+00	0.0000	LPPGLLANFTLLR	1
4	7.742742e+05	0.0000	LPPGLLANFTLLR	1
5	4.820082e+05	0.0000	LPPGLLANFTLLR	1
6	7.807922e+05	0.0000	LPPGLLANFTLLR	1
7	1.377281e+06	1.4400	LPPGLLANFTLLR	1
8	9.861708e+06	7.2000	LPPGLLANFTLLR	1
9	6.757440e+07	36.0000	LPPGLLANFTLLR	1
10	2.875355e+08	180.0000	LPPGLLANFTLLR	1
11	2.427150e+09	900.0000	LPPGLLANFTLLR	1
12	2.546919e+04	0.0576	LPPGLLANFTLLR	2
13	1.886122e+05	0.2880	LPPGLLANFTLLR	2
14	0.000000e+00	0.0000	LPPGLLANFTLLR	2
15	1.415216e+05	0.0000	LPPGLLANFTLLR	2
16	1.931595e+05	0.0000	LPPGLLANFTLLR	2

Showing 1 to 17 of 30 entries



MSstats R package: Linear fit for LOB/LOD calculation: linear_quantlim()

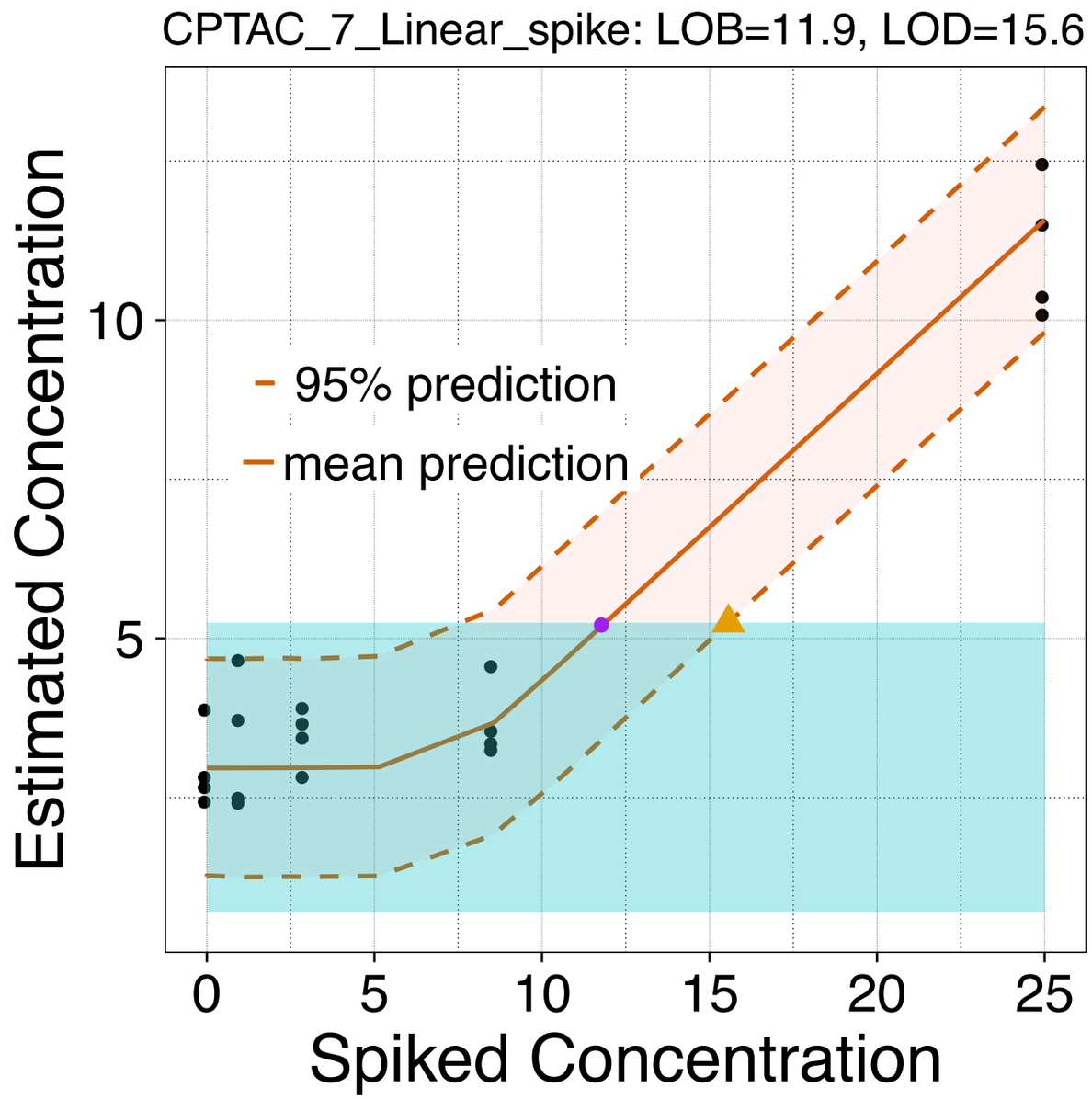



```
linear_quantlim(datain, alpha = 0.05, Npoints = 100, Nbootstrap = 500)
```

Arguments

`datain`	Data frame that contains the input data. The input data frame has to contain the following columns : CONCENTRATION, INTENSITY (both of which are measurements from the spiked in experiment) and NAME which designates the name of the assay (e.g. the name of the peptide or protein)
`alpha`	Probability level to estimate the LOB/LOD

MSstats R package: Nonlinear fit for LOB/LOD calculation: nonlinear_quantlim()



Usage

```
nonlinear_quantlim(datain, alpha = 0.05, Npoints = 100, Nbootstrap = 500)
```

Arguments

`datain`	Data frame that contains the input data. The input data frame has to contain the following columns : CONCENTRATION, INTENSITY (both of which are measurements from the spiked in experiment) and NAME which designates the name of the assay (e.g. the name of the peptide or protein)
`alpha`	Probability level to estimate the LOB/LOD

MSstats R package: Nonlinear vs Linear fit for LOB/LOD calculation