Fitting curves using non-parametric approaches

Keith R. Shockley, NIEHS

October 24, 2017

OUTLINE:

Parametric modeling

Non-parametric modeling

Estimating point of departure from fitted curves

Parametric modeling

Parametric models

- Pre-specified model form
  - Linear model: $f(x) = mx + b$
  - Hill model: $f(x) = f_0 + f_{max} * x^h/(AC_{50}^h + x^h))$
  - Cubic polynomial: $f(x) = a_3x^3 + a_2x^2 + a_1x^1 + a_0$
- Contain parameters, some of which might be useful
  - Slope and y-intercept
  - AC_{50}, f_{max}, f_0, h
  - a_3, a_2, a_1, a_0

Parametric models (pros and cons)

<u>Pros</u>

- ❑ Reduce unknown (and possibly complicated) function *f(x)* to a simple form with few parameters
- ❑ Can produce consistent results when the curve fits the data well
- ❑ May have familiar and useful parameters

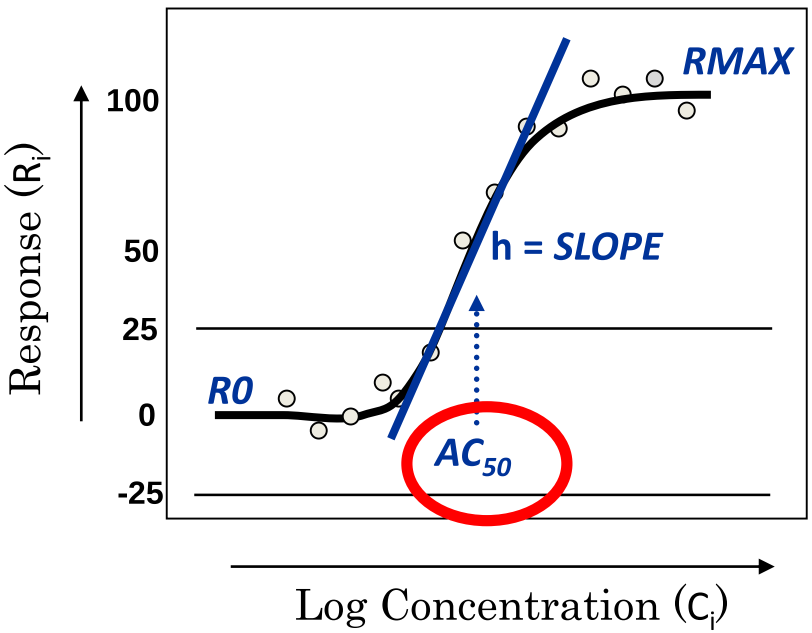
<u>Cons</u>

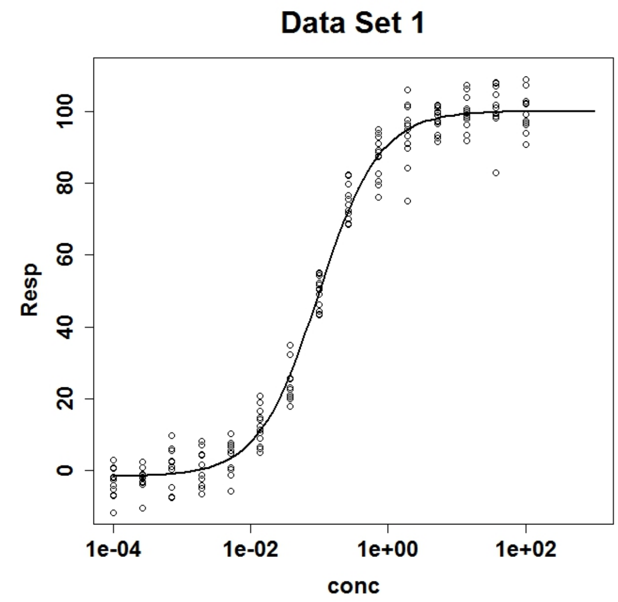
- ❑ A pre-specified parametric model may not fit the data well
- ❑ Carry distributional assumptions (e.g., Normality)
- ❑ Different parametric models may produce different *BMD* estimates, reflecting model uncertainty
- ❑ Model averaging can be helpful when true function is not on edge of model averaging space

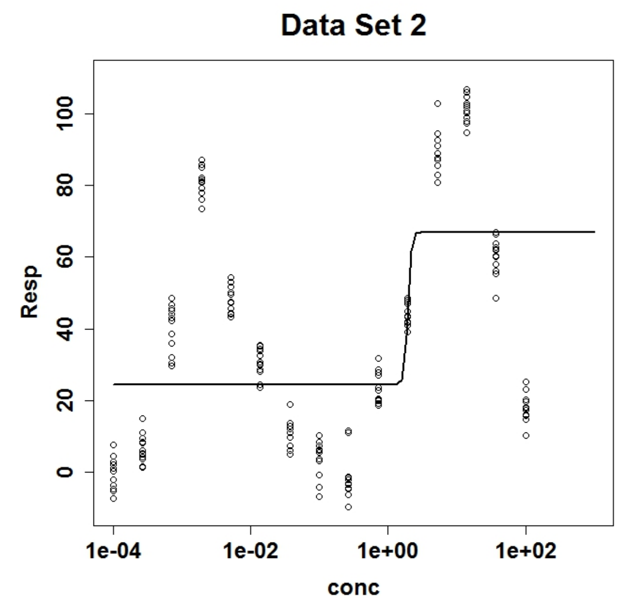
Hill model

Hill equation

$$R = R_0 + \frac{R_{max} C^h}{AC_{50}^{\ h} + C^h}$$







Non-parametric modeling

Non-parametric models

- Flexible model form
  - Interpolation
  - LOESS (nonparametric local regression)
  - Splines (continuous piece-wise polynomials between knots)

- Parameters may not be readily interpretable
  - Interpolation – estimates values that lie between data points
  - LOESS – fits segments of the data at each point in the range of the data set by calculating many polynomial coefficients using weighted least squares, but doesn’t estimate a single “coefficient” for a global model
  - Splines – calculates many “polynomial coefficients” between intervals of data

Non-parametric models (pros and cons)

Pros

- ❑ Makes fewer assumptions about *f(x)*
- ❑ Uses the data to learn about the potential shape of *f(x)*
- ❑ Should fit the data very well

Cons

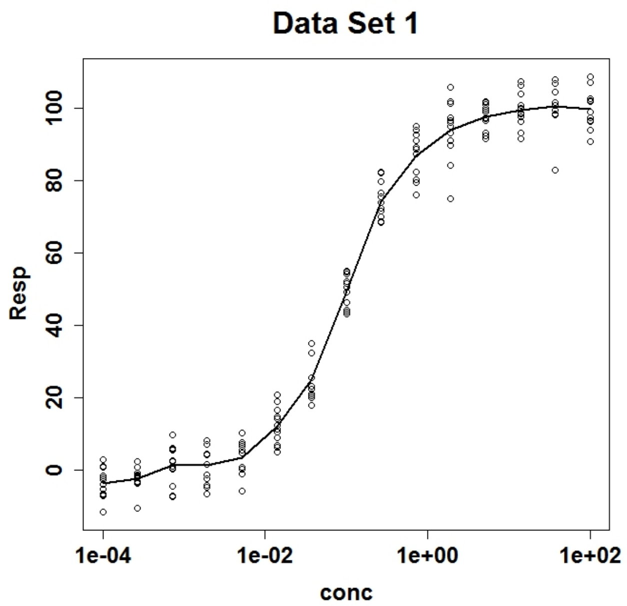
- ❑ Parameters may not be readily interpretable
- ❑ Carry distributional assumptions too
- ❑ May be computationally intensive
- ❑ May not be as familiar as parametric approaches

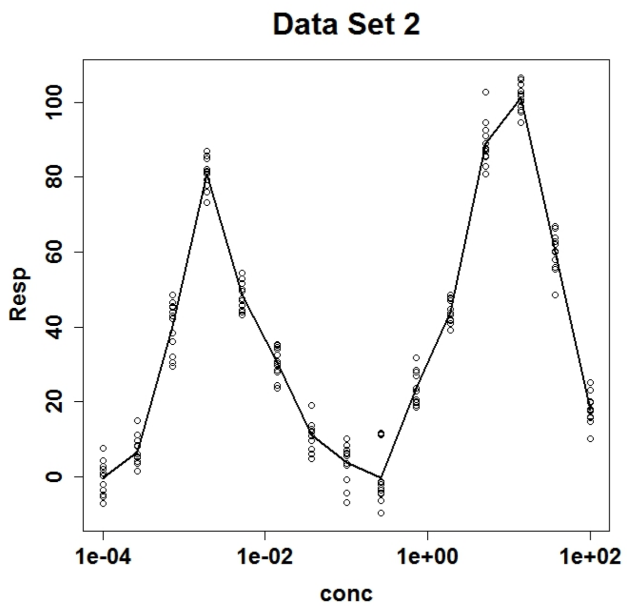
B-spline

A basis spline (B-spline) is a piecewise polynomial function, where the pieces meet at the knots. Any spline can be expressed as a linear combination of B-splines.

For each interval $[x_i, x_{i+1}]$:

$B(x) = a_3x^3 + a_2x^2 + a_1x^1 + a_0$

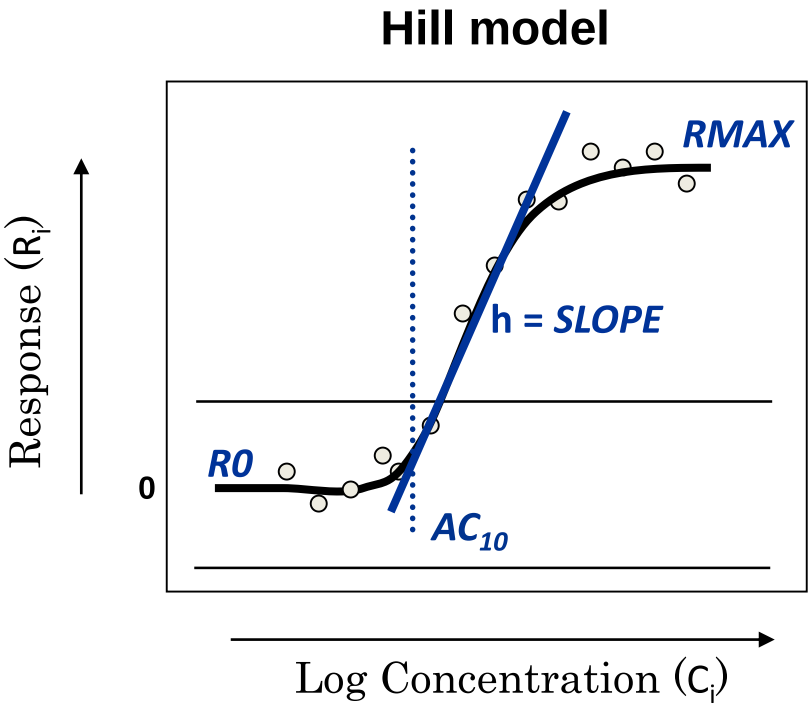




Estimating point of departure from fitted curves

Curve Fitting and Potency Estimation (Case 1)

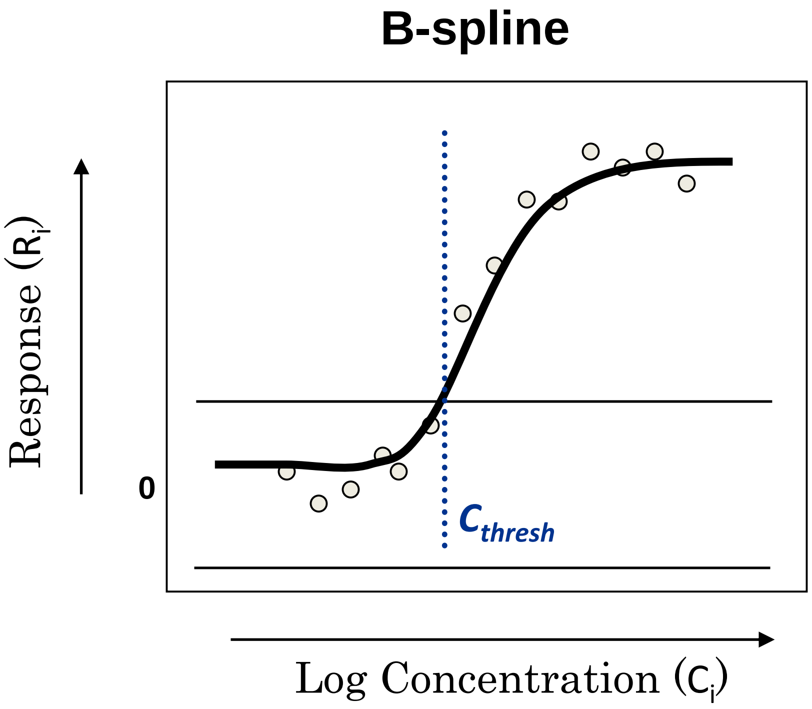
- Case 1: Hill equation model and AC_{10} parameter



1. Fit the Hill model to the data.
2. AC_{10} is the point of departure.

Curve Fitting and Potency Estimation (Case 2)

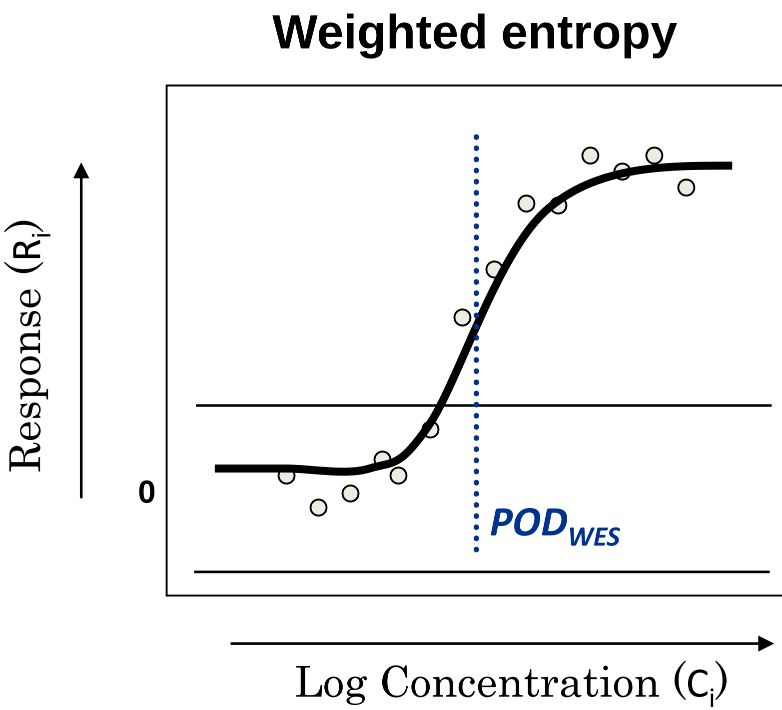
- Case 1: Hill equation model and AC_{10} parameter
- Case 2: B-spline and concentration that curve crosses a response threshold



1. Fit a B-spline to the data.
2. The point of departure is the concentration at which the curve crosses the detection band.

Curve Fitting and Potency Estimation (Case 3)

- Case 1: Hill equation model and AC_{10} parameter
- Case 2: B-spline and concentration that curve crosses a response threshold
- Case 3: Polynomial interpolation and entropy-based point of departure



1. Fit an interpolation curve to the data.
2. Calculate a "weighted entropy" along the curve (*Shockley, 2014*).
3. The *POD* is the concentration at which the change in entropy is maximal (*Shockley, 2016*).

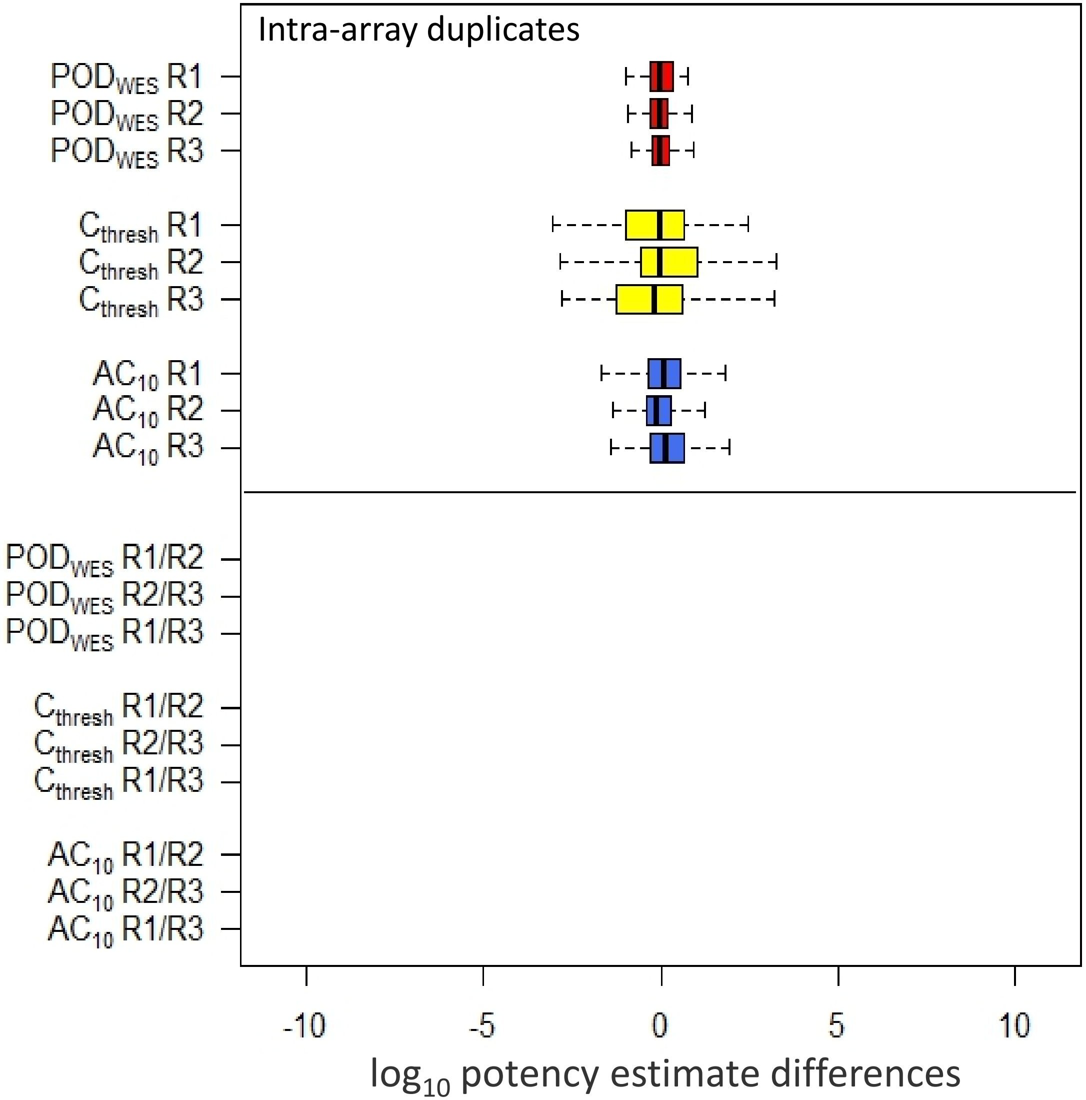
Simulation Study

Data Set Parameters				Estimator		
True $\|RMAX\|$	True AC_{10}	True C_{thresh}	True POD_{WES}	AC_{10}	C_{thresh}	POD_{WES}
5% error (15% Detection Limit)						
25	0.0001	0.001	0.003	**17.51 (-2.84)**	**4.08** (+1.45)	1.26 (+0.11)
25	0.01	0.2	0.4	**4.13** (-1.36)	**2.78** (+0.31)	1.30 (-0.11)
25	1.1	15	43.9	**10.64** (+0.07)	**4.37** (-1.63)	1.28 (-0.27)
50	0.0001	0.0004	0.001	**5.80** (-1.32)	**4.46** (+1.27)	1.03 (-0.03)
50	0.01	0.04	0.2	1.34 (-0.96)	1.63 (-0.32)	1.14 (-0.27)
50	1.1	4.3	10.2	1.59 (-0.79)	**3.67** (-0.98)	1.23 (-0.17)
100	0.0001	0.0002	0.0004	1.37 (-1.00)	**4.68** (+0.85)	1.06 (+0.06)
100	0.01	0.02	0.07	0.61 (-0.96)	0.99 (-0.07)	1.20 (-0.17)
100	1.1	1.8	3.7	0.56 (-0.95)	**3.15** (-0.65)	1.53 (-0.07)
10% error (30% Detection Limit)						
50	0.0001	0.0004	0.003	**8.39** (-1.48)	**4.50** (+1.77)	1.34 (+0.11)
50	0.01	0.04	0.3	1.43 (-0.97)	**2.65** (-0.46)	1.30 (-0.04)
50	1.1	4.3	39.3	**8.49** (-0.58)	**3.91** (-1.31)	1.29 (-0.24)
100	0.0001	0.0002	0.001	1.18 (-1.01)	**4.77** (+1.48)	1.00 (-0.03)
100	0.01	0.02	0.2	0.64 (-0.96)	**1.38** (-0.25)	0.92 (-0.32)
100	1.1	1.8	9.1	1.17 (-0.93)	**3.44** (-0.97)	0.97 (-0.16)

Log_{10}Precision (log_{10}Bias) in 15-point concentration response data simulated from 10,000 Hill model curves, with $R_0 = 0$ and $h = 1$. *Adapted from Shockley (2016).*

Repeatability of Potency Estimates

(Tox21 Phase II BG1 estrogen receptor agonist)

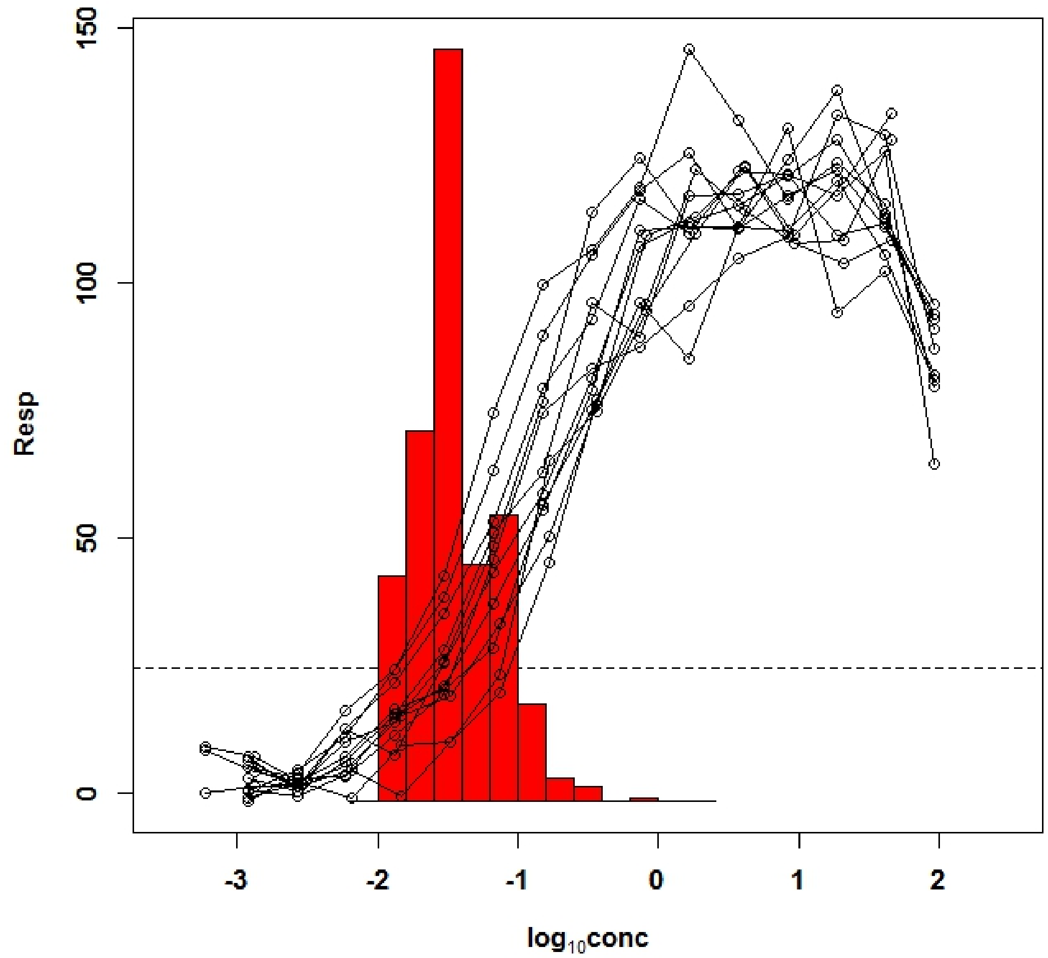


adapted from Shockley, 2016

Examples: Uncertainty in POD_{WES} (bootstrapping)

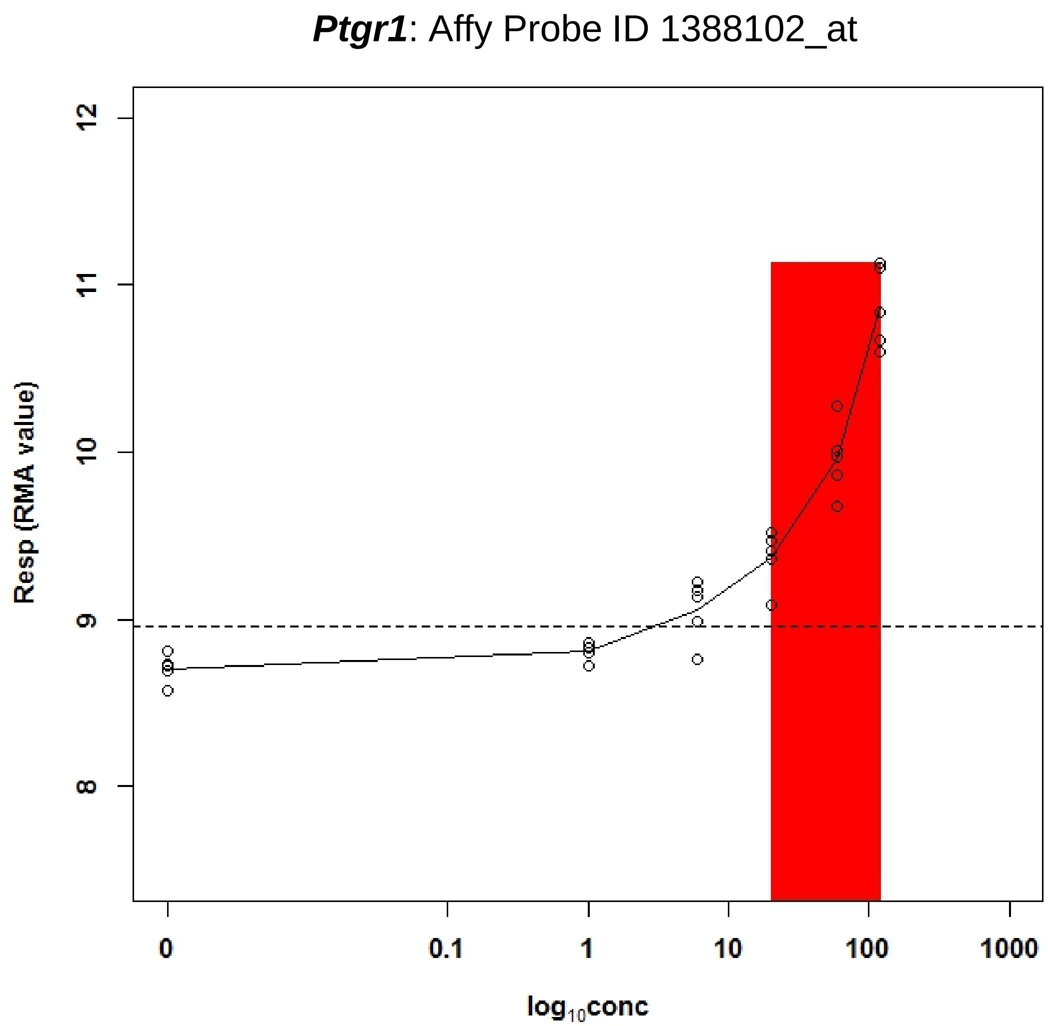
(Tox21 Phase II BG1 estrogen receptor agonist)

Oxymetholone: CASRN 434-07-1



Examples: Uncertainty in POD_{WES} (bootstrapping)

(Dunnick et al., *Arch. Toxicol.*, 2017)



Summary

- Parametric modeling requires pre-specifying the model, but is more familiar and may have interpretable parameters.

- Nonparametric modeling is more flexible, but may be less familiar and may not have readily interpretable parameters.

- Simulation studies and repeatability of experimental results can be used to evaluate the performance of proposed modeling approaches.