Combining DOE and First-Principles Science to Maximize Yield and Minimize Impurity with Fit Curve CDOE

JMP Discovery Summit - Americas

Brian Taylor, Chemical Development, Pharmaceutical Technology & Development, Operations, AstraZeneca, *Macclesfield, UK*

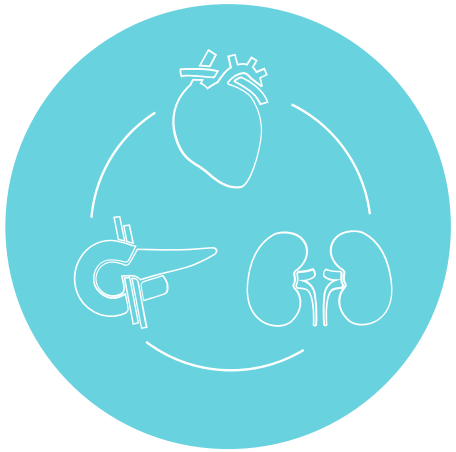
Chris Gotwalt, JMP

October 2021

1. AstraZeneca and where I work
2. Lab & DoE Backgound
3. Functional DoE background
4. Fit Curve Functional DoE
5. Analysis Steps
6. Chris Gotwalt Demonstration
7. Subject Matter Expert Guidance
8. Group Profiler: Linking to Process Insight
9. Analysis Approach Improvements
10. Summary

Focus on three main therapy areas and across key platforms

Oncology



Cardiovascular, Renal & Metabolism

Respiratory & Immunology

Combination of capabilities

Small molecules	Biologics	Protein Engineering	Other Emerging Drug Platforms	Diagnostics	Devices

PT&D in AstraZeneca

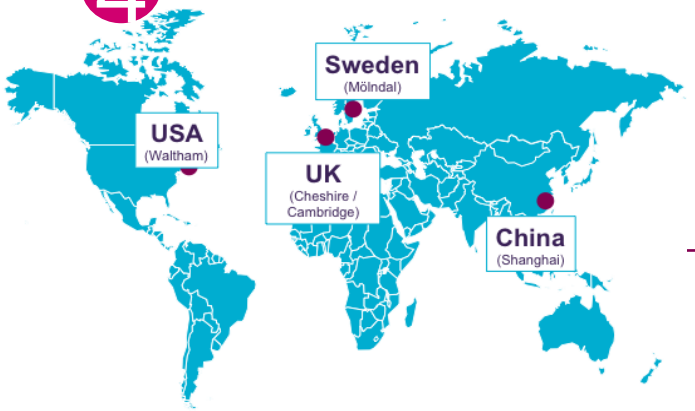
Chief Executive officer

BioPharmaceuticals R&D

Oncology R&D

Operations

4 Hubs



Pharmaceutical Technology & Development

~1,000 People

Drug Product Development

Chemical Development

~250 People

Process chemists, Analytical chemists, Computational chemists, Chemical engineers, Physical Organic chemists, Crystallisation scientists

Lifecycle of Medicine

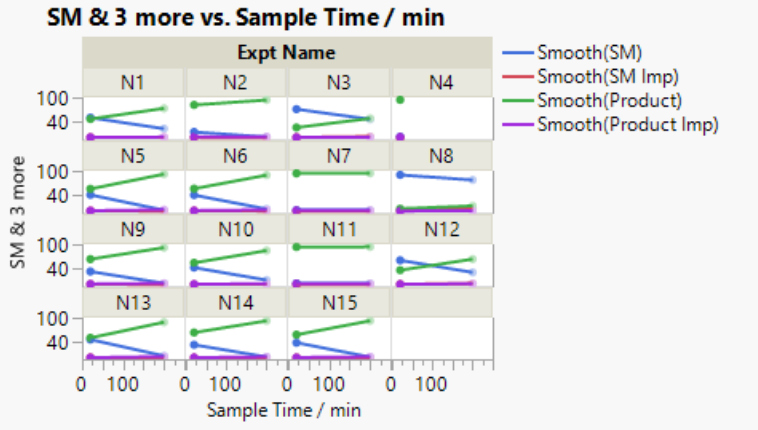
AstraZeneca spans the **entire life-cycle** of a medicine from research and development to manufacturing and supply, and the global commercialisation of primary care and speciality care medicines.

PT&D develops technologies to support products from **PhII clinical studies** through to **post product launch**

Chemical Development **Design, develop and optimise synthetic routes, processes and analytical methods for the Active Pharmaceutical.**

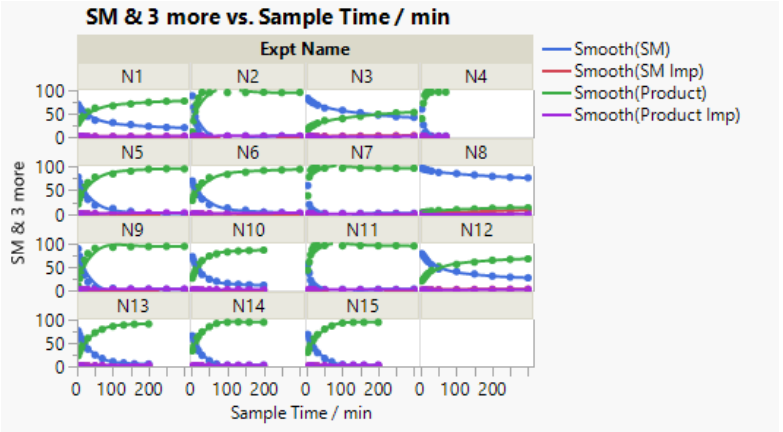
Lab Experimentation & DoE's Analysing Profiles

Early 2000s – 1or2 samples/Expt



More experiments are routinely sampled to produce a reaction profile or distribution

2020s – 12 Samples/Expt



- Analysing DoEs with reaction profiles from sampling experiments' over time
  - Strategies aim to reduce multiple response results into a single value
  - A common approach is to select a slice in time to analyse.
  - What process insight is lost from analysing a response in slices?
- What analysis approach can make use of all the time course reaction profile data?

What analysis approach can make use of all the time course reaction profile data?

Can we refine the analysis to consider subject matter knowledge?

Grand Presentation Title:
Combining DOE and First-Principles Science to Maximize Yield and Minimize Impurity with Fit Curve CDOE

What we do in JMP:
Analyse DoE Reaction Profiles with Fit Curve and CDOE

DoE Case Study Background

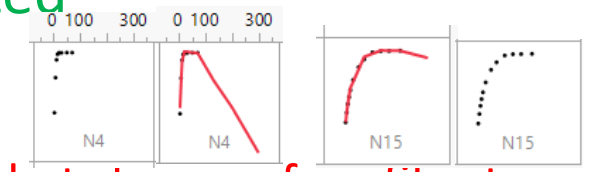
- 4 factor Definitive Screening Design (Conference Matrix Derived)
  - Solvent Volume 6.0 - 10.0 rel vols, Catalyst Charge 0.70 -1.30 mol%, Base Charge 4.50 -7.50 eq, Amine Starting Material 1.00 - 1.10 eq
- Sampling: 12 samples per experiment.
  - Time points selected to coincide every time 10% conversion expected to happen.
  - Each experiments samples time points can differ
- Process Response Target Criteria

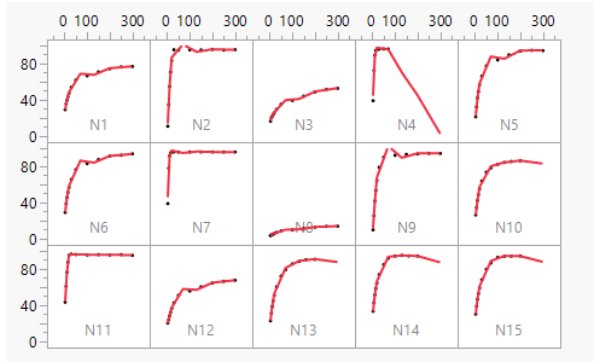
Responses	Target Criteria
SM – Starting Material	Minimise
SM Imp – Starting Material Imp	Minimise
Prod – Product	>95 %
Prod Imp – Product Impurity	<2 %
Reaction Time	120 – 240 Mins

  - Prod Imp priority to control

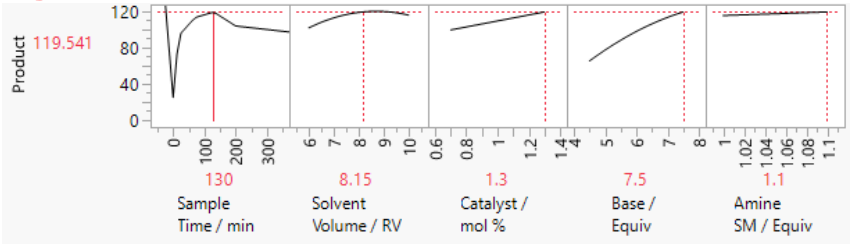
JMP Functional Data Explorer in combination with DoE to Chemical Reactions

- FDE flexible Spline fitting approach can be a strength and a weakness for fitting smooth continuous time course data
  - Countless shapes can be approximated
  - Unrealistic shapes are approximated
    - when domain specific knowledge infers what shapes of profiles to expect



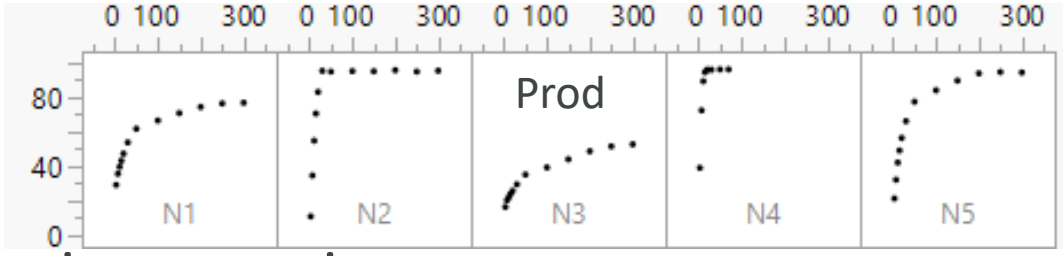


- Combining FDE and DoE has strengths and weakness
  - FDoE converts output to relate profiles to factors studied in the DoE
  - Unrealistic predictions estimated – negative values, values greater than theoretical maximum. Eg Prod Max 100%, prediction 119%.

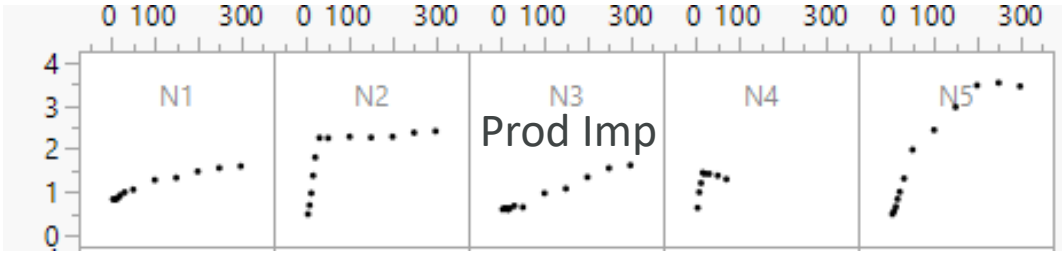


Chemical Reaction profile shape properties

- Natural Minimum and Maximum bound
  - Starting Materials and Products bounds: Minimum 0%, Maximum 100%.
- Profile Gradient
  - Montonic – increasing or decreasing
  - Plateau – travels to an optima and remains at optima over time



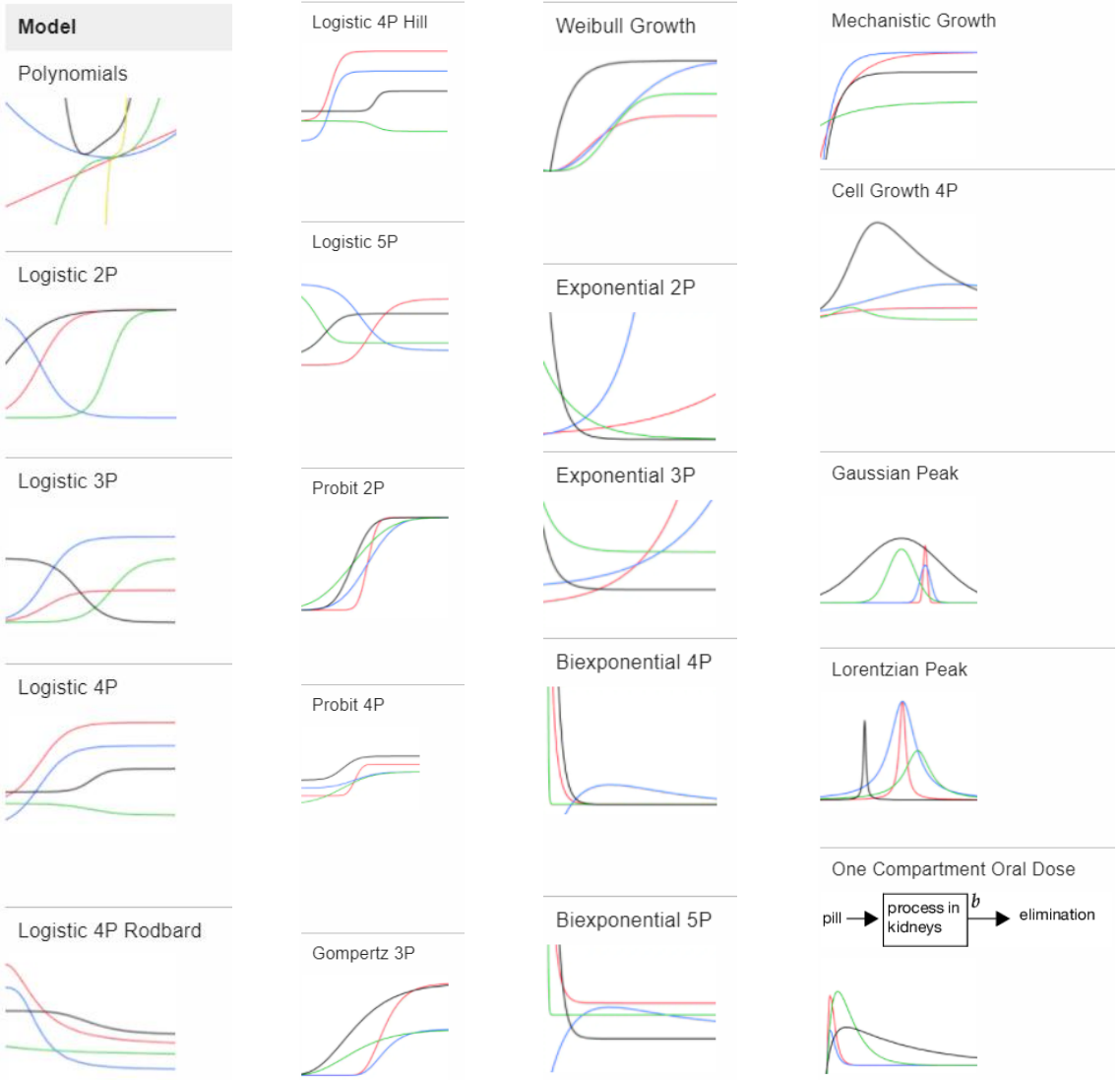
- Single Profile Peak/Trough
  - Product – Peak (maxima) reached, reaction held longer, decay occurs for some reason
- End of Reaction (EOR) can occur over different time scales (SM or Prod plateaus)
  - 50mins, 300 mins or 1200mins.
- Sum of Response Area % results add up to 100%.
  - SM + SM Imp + Prod + Prod Imp = 100%

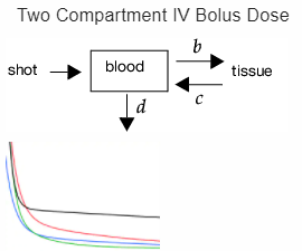


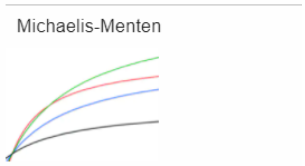
Functional DoE with Fit Curve

- Pre-loaded equations represent profiles more likely observed in Chemical Reactions
  - Some flexibility lost on fitting profile shapes
  - The profiles shapes fitted assumed to resemble more closely the underlying physical equations controlling chemical reaction profiles
  - Fitted curves compared to determine best fitted to use in DoE Modelling
    - Fit curve formula parameters used as responses in DoE Modelling

Fit Curve Pre-loaded Formulas







Select formulas to try that match the subject matter domain profile characteristics.

Analysis Steps
Analyse – Specialised Modelling – Fit Curve

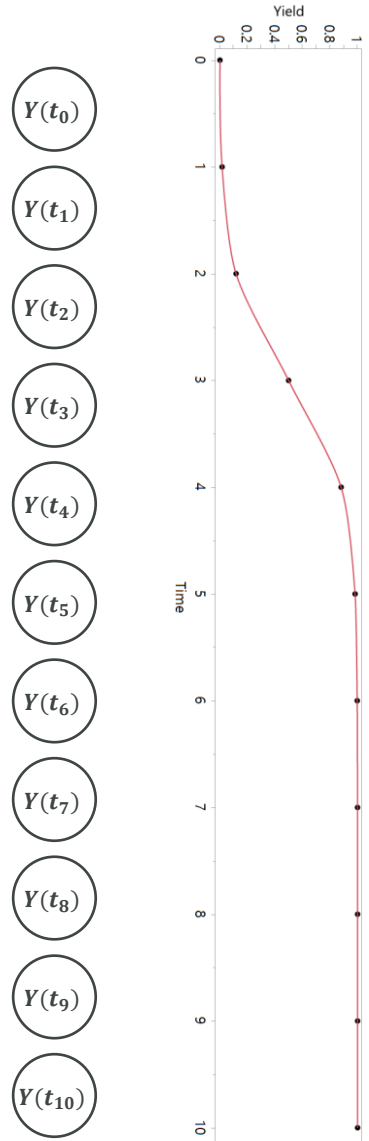
- Define the Data Structure for analysis
- Fit Multiple Curves to chemical reaction Response
  - Compare Curve fits and select preferred one
- Select Curve DoE on formula parameter Coefficients
  - Decide the DoE Analysis approach
    - Default is Forward Selection
  - Decide which Parameter Coefficients Distribution
    - Default is Gaussian
  - DoE Profiler for each chemical response
    - Compare multiple response profiles to understand reaction behaviour

Chris Gotwalt Fit Curve – CDoE Demonstration

Curve DOE Analysis Demo

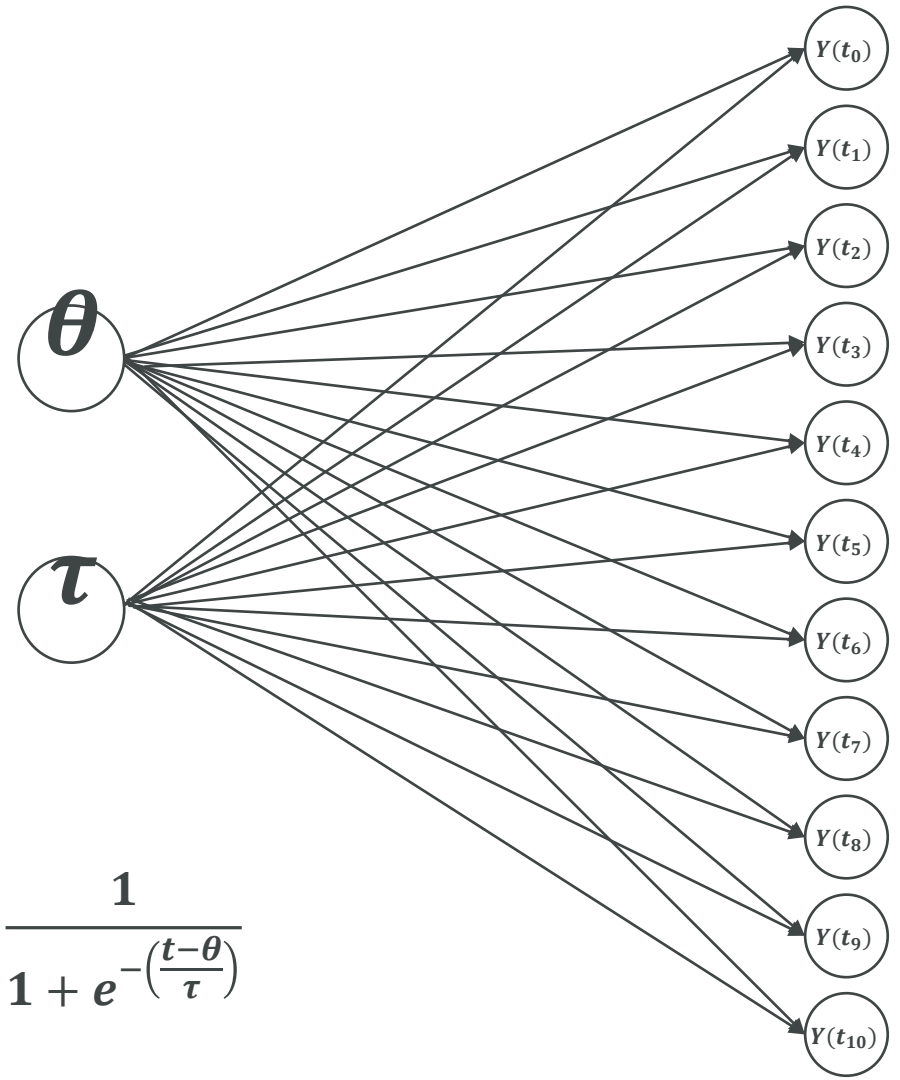
	Expt Name	Sample Time / min	Solvent Volume / RV	Catalyst / mol %	Base / Equiv	Amine SM / Equiv	Product	Product Imp	SM	SM Imp
1	N1	3	8	1.3	4.5	1	29.375	0.8351	69.623	0.1674
2	N1	7	8	1.3	4.5	1	35.85	0.8283	63.027	0.2945
3	N1	11	8	1.3	4.5	1	39.816	0.8428	58.857	0.4844
4	N1	15	8	1.3	4.5	1	43.438	0.8693	55.102	0.5912
5	N1	20	8	1.3	4.5	1	47.531	0.9292	50.879	0.6609
6	N1	30	8	1.3	4.5	1	53.969	0.9954	44.227	0.809
7	N1	50	8	1.3	4.5	1	61.919	1.0611	36.096	0.9243
8	N1	100	8	1.3	4.5	1	66.736	1.282	31.091	0.8916
9	N1	150	8	1.3	4.5	1	70.972	1.3323	26.764	0.9308
10	N1	200	8	1.3	4.5	1	74.555	1.4797	22.934	1.0305
11	N1	250	8	1.3	4.5	1	76.611	1.56	20.675	1.1537
12	N1	300	8	1.3	4.5	1	77.024	1.6041	20.252	1.1202
13	N2	3	8	0.7	7.5	1.1	11.069	0.4895	88.197	0.2446
14	N2	7	8	0.7	7.5	1.1	34.798	0.6968	63.886	0.619
15	N2	11	8	0.7	7.5	1.1	54.966	0.9749	43.416	0.6434
16	N2	15	8	0.7	7.5	1.1	70.737	1.3847	27.302	0.5759
17	N2	20	8	0.7	7.5	1.1	83.222	1.8063	14.562	0.4093

How Curve DOE Works

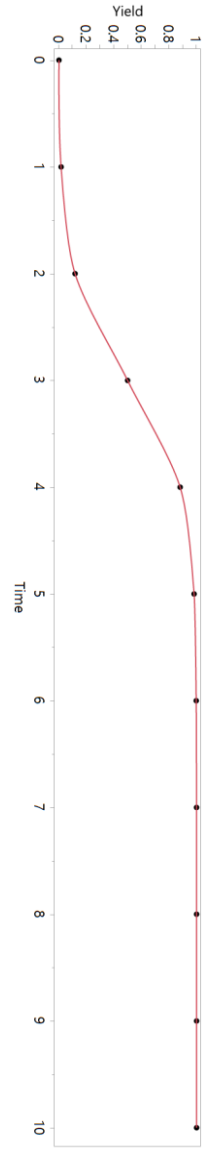


Responses

How Curve DOE Works

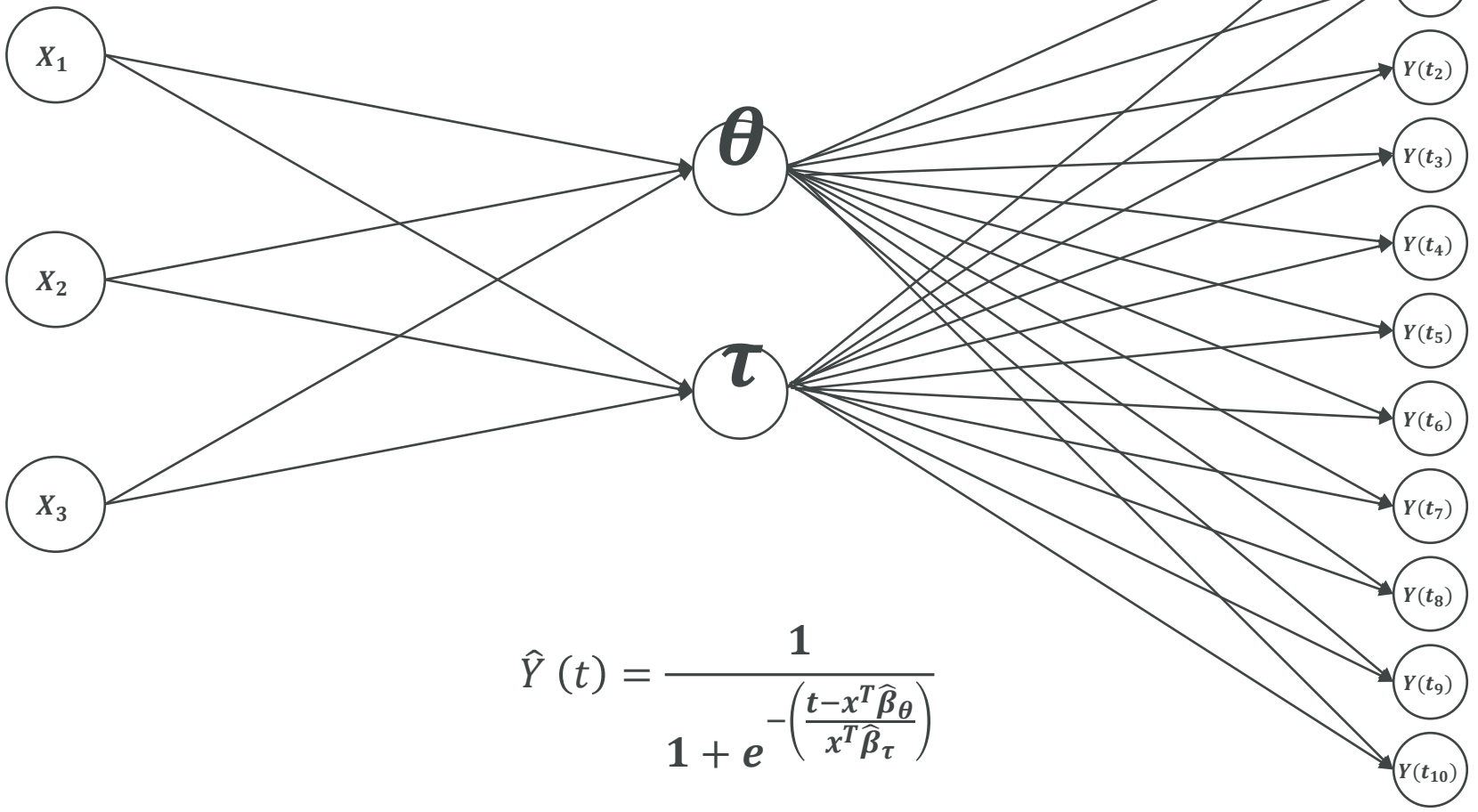


$$\hat{Y}(t) = \frac{1}{1 + e^{-\left(\frac{t-\theta}{\tau}\right)}}$$

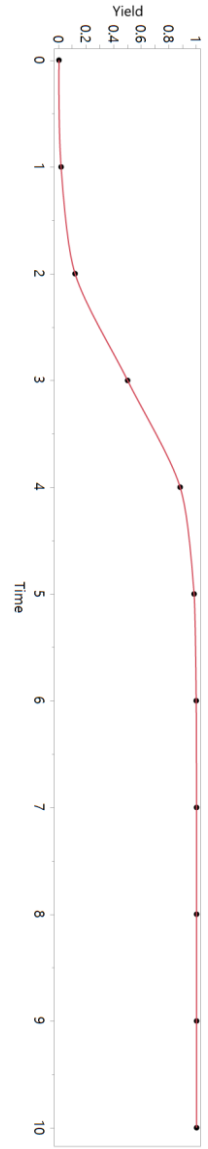


Responses

How Curve DOE Works



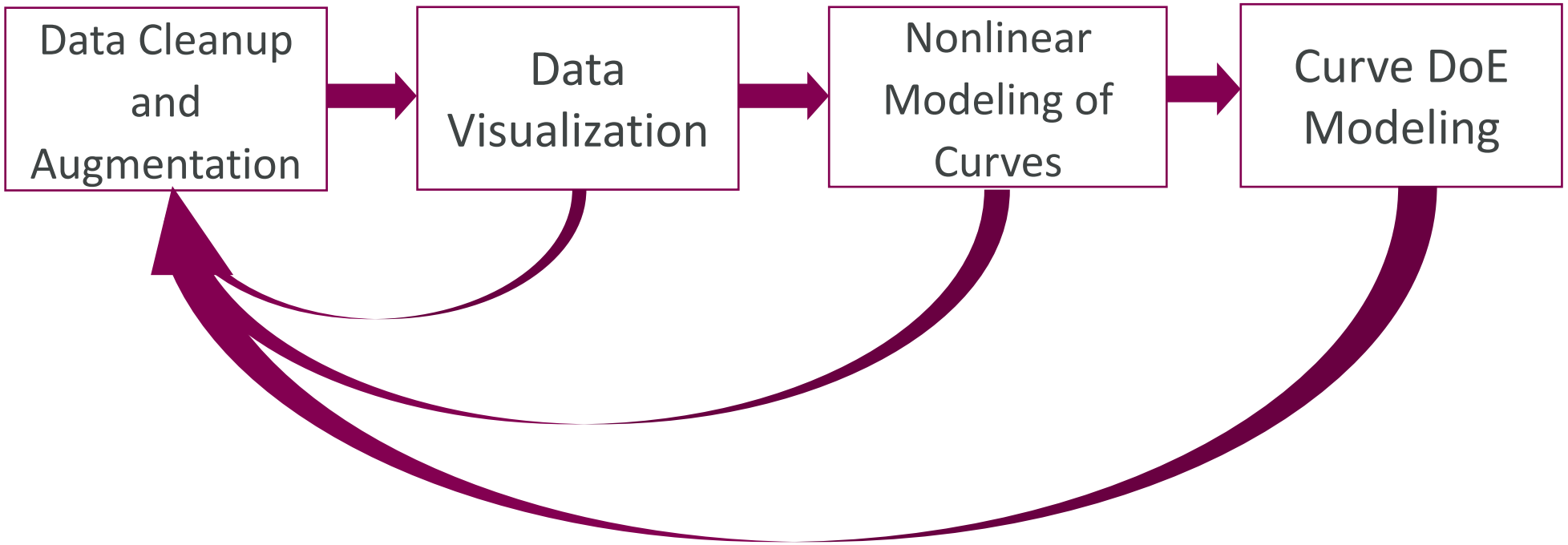
$$\hat{Y}(t) = \frac{1}{1 + e^{-\left(\frac{t - x^T\hat{\beta}_\theta}{x^T\hat{\beta}_\tau}\right)}}$$



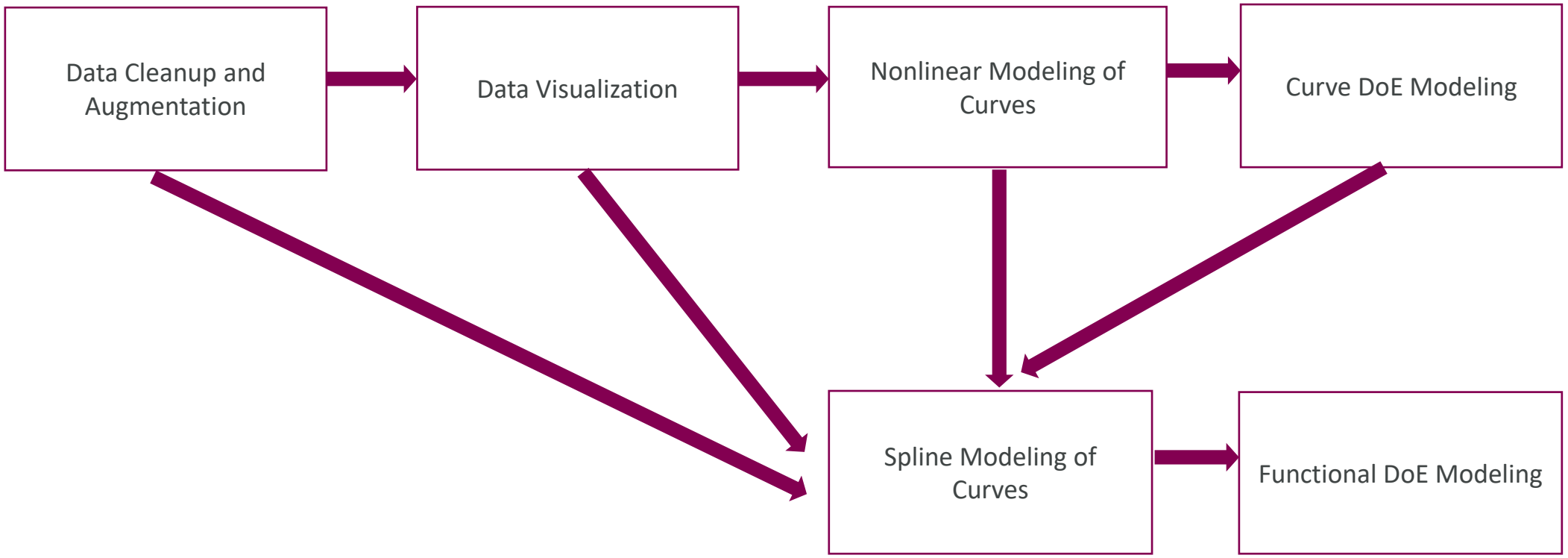
DOE Factors

Responses

Fit Curve Workflow



Fit Curve Workflow



Fit Curve or Functional Data Explorer?

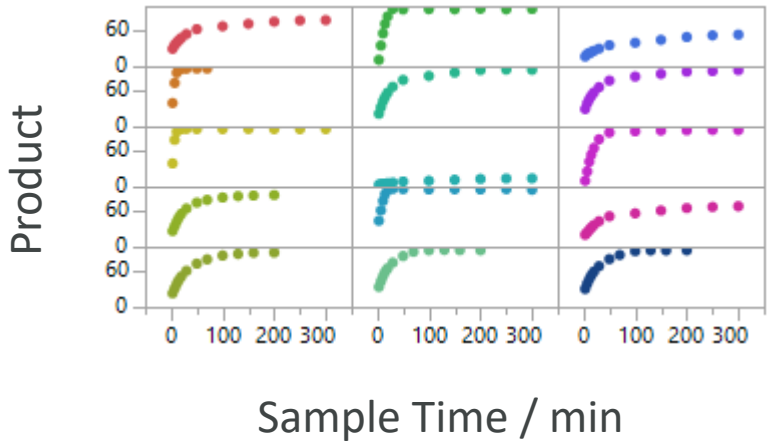
- Is there scientific basis for a particular nonlinear function?
- Do the trajectories fall into one (and only one) of the supported shapes?
- Does there seem to be a time-translation effect that would be nicely modelled with a location parameter?

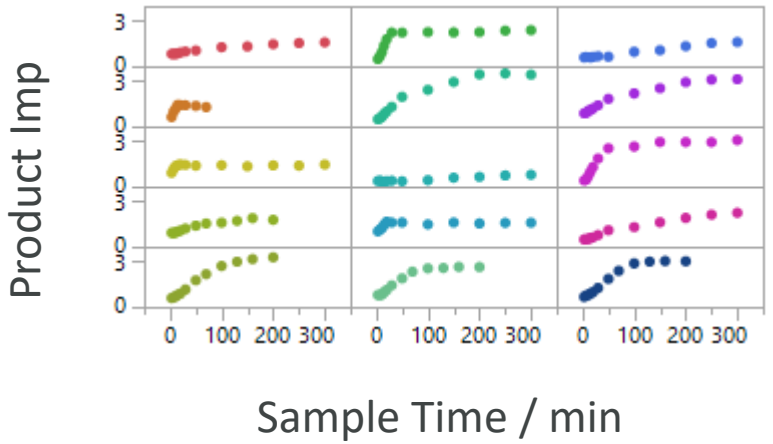
YES ⇒ Fit Curve

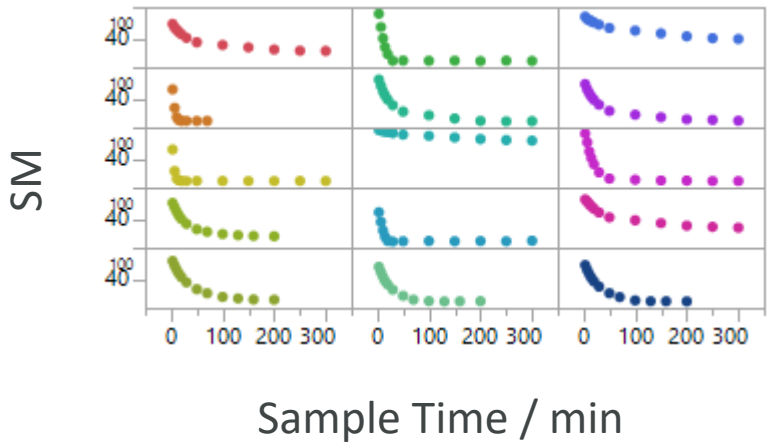
- Is there little first principles theory to work from?
- Are the trajectories highly non-homogenous?
- Do the curves have three or more “features”?

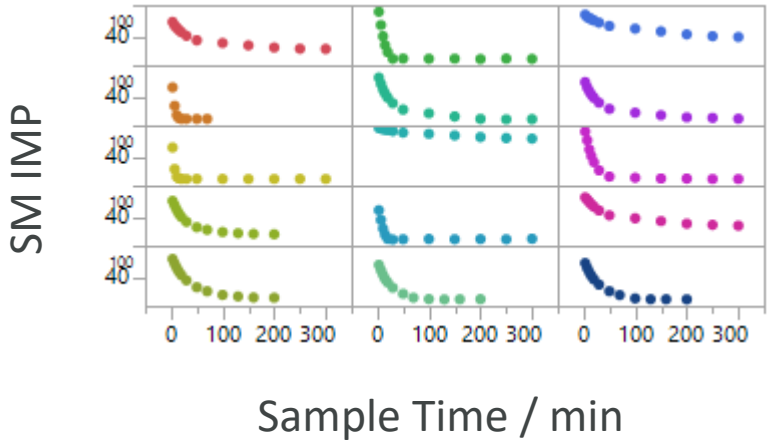
YES ⇒ Functional Data Explorer

The Four Response Curves

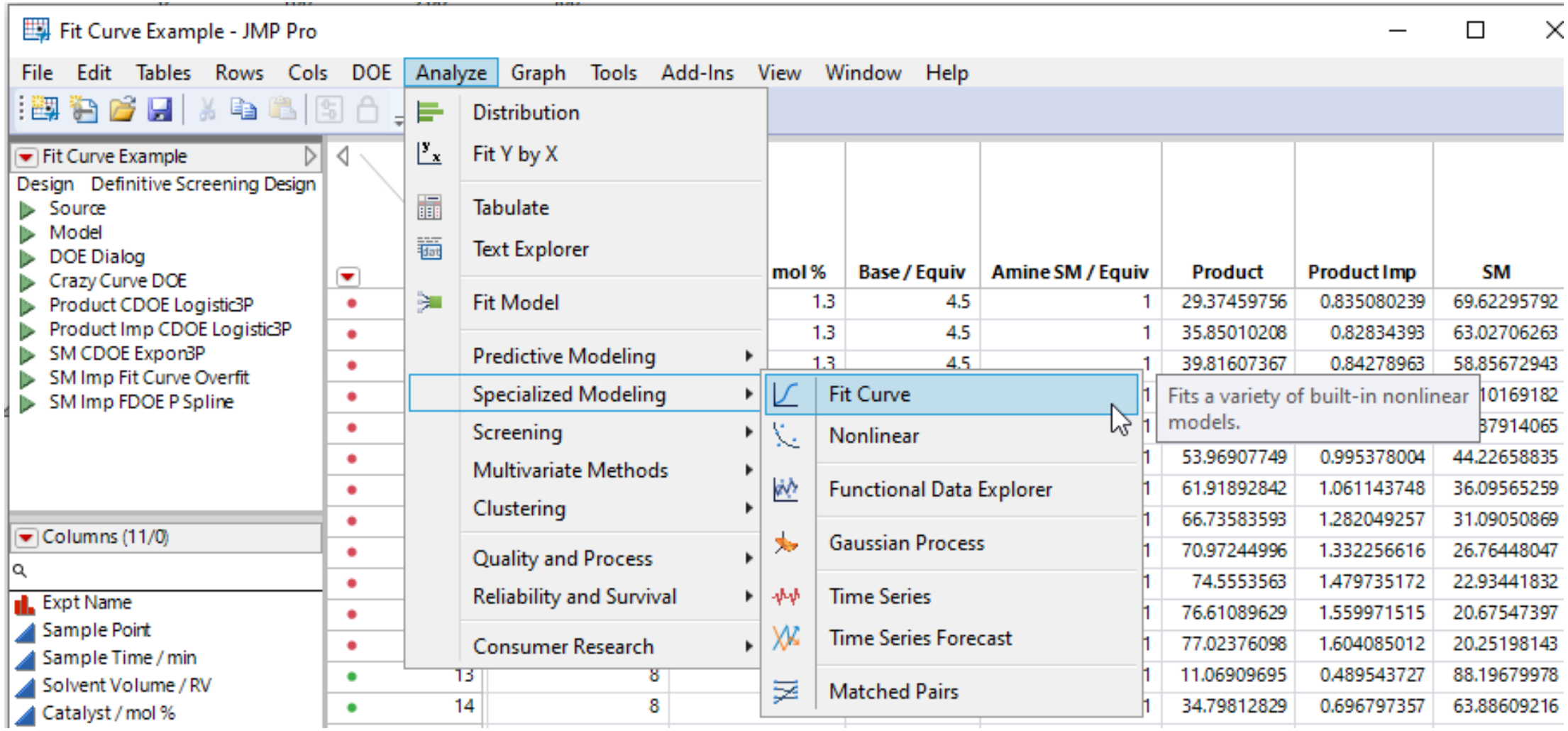




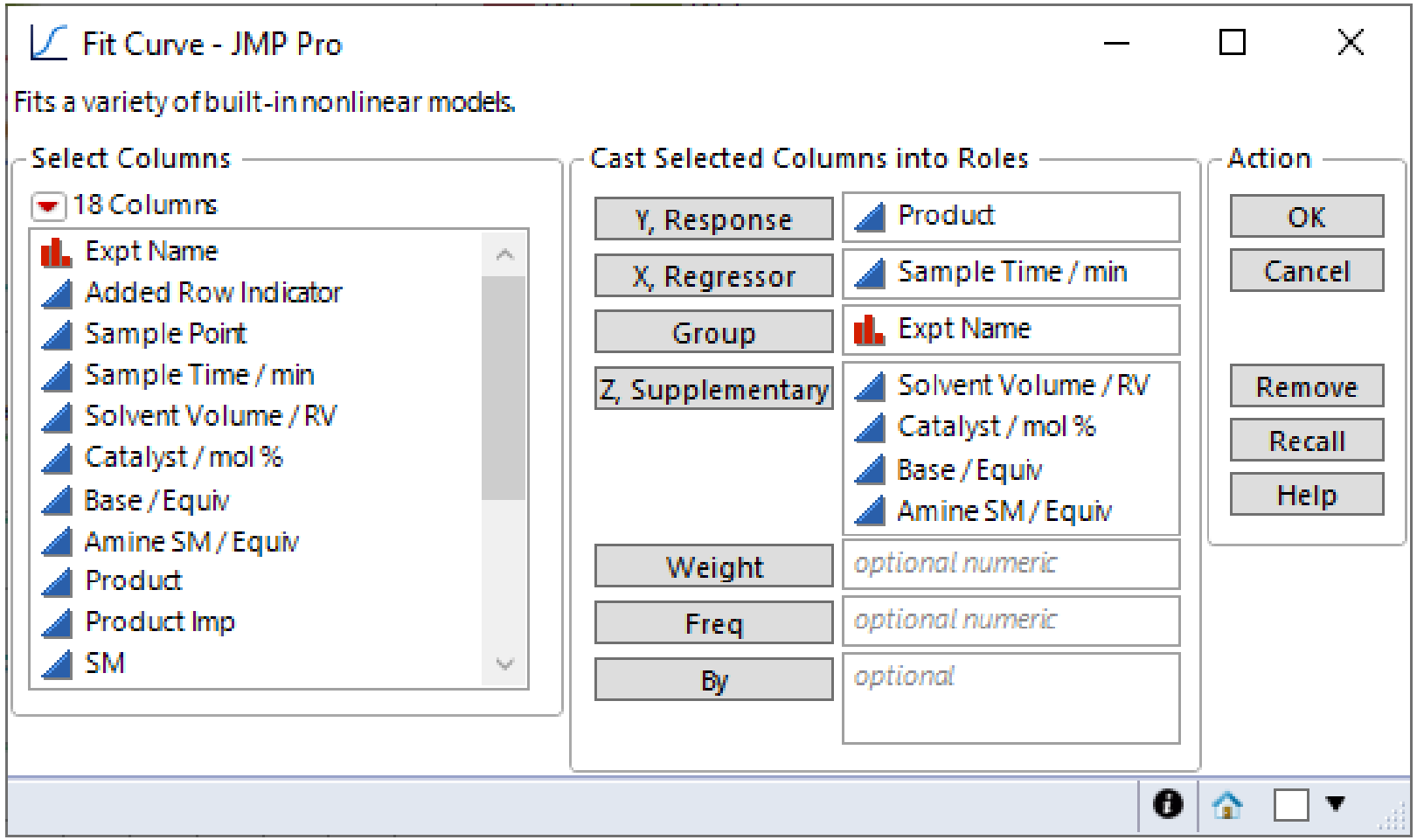




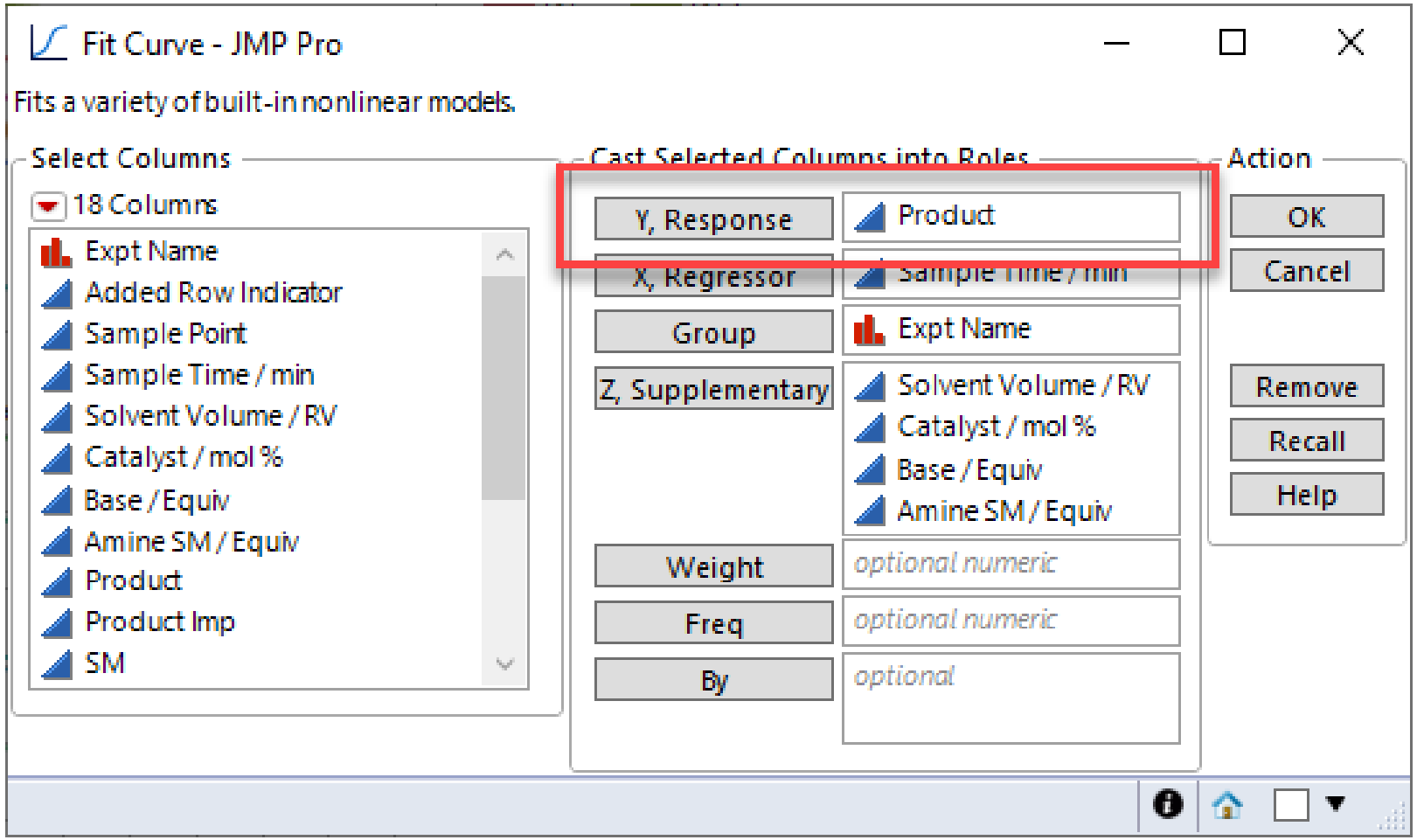
Launching Fit Curve



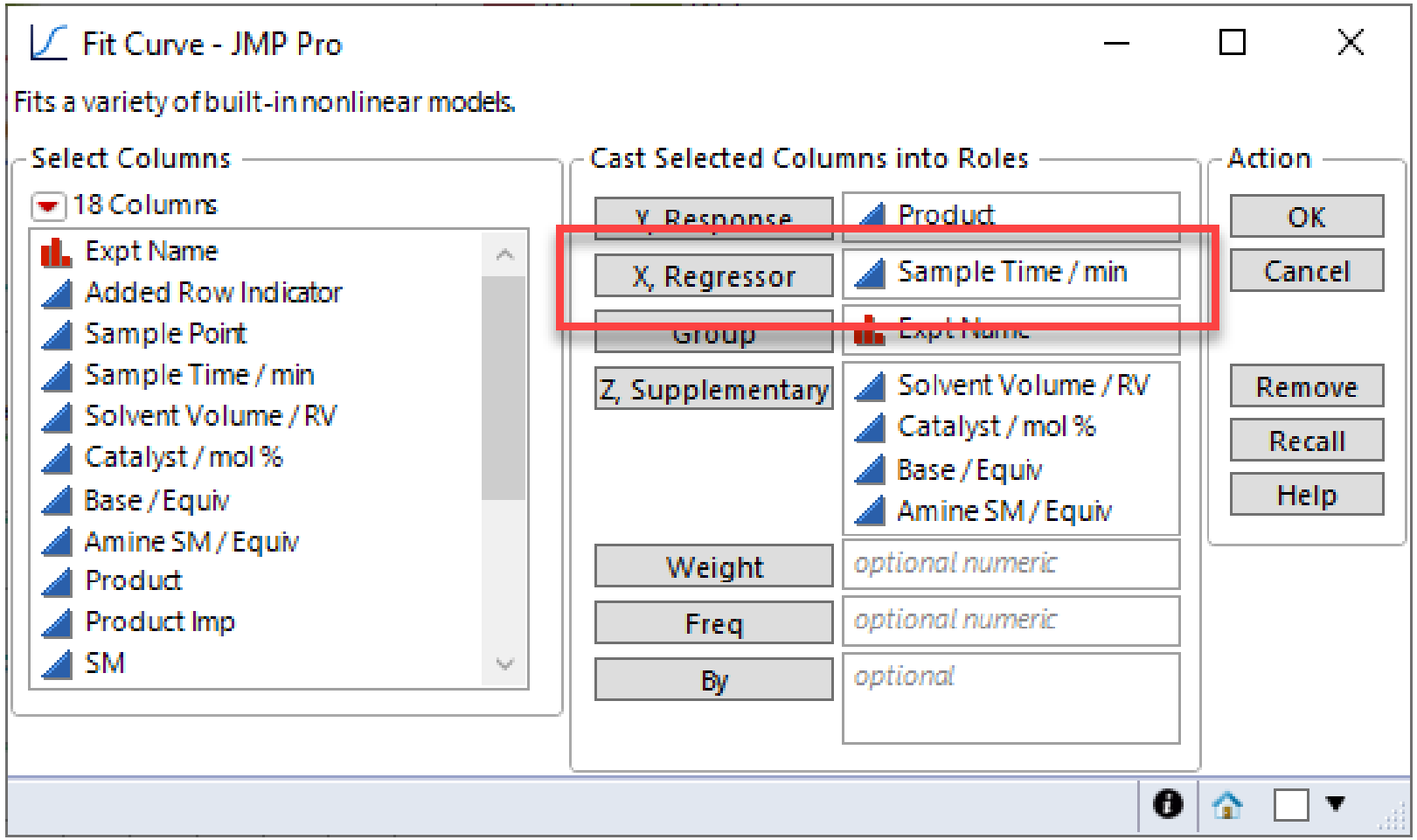
Launching Fit Curve



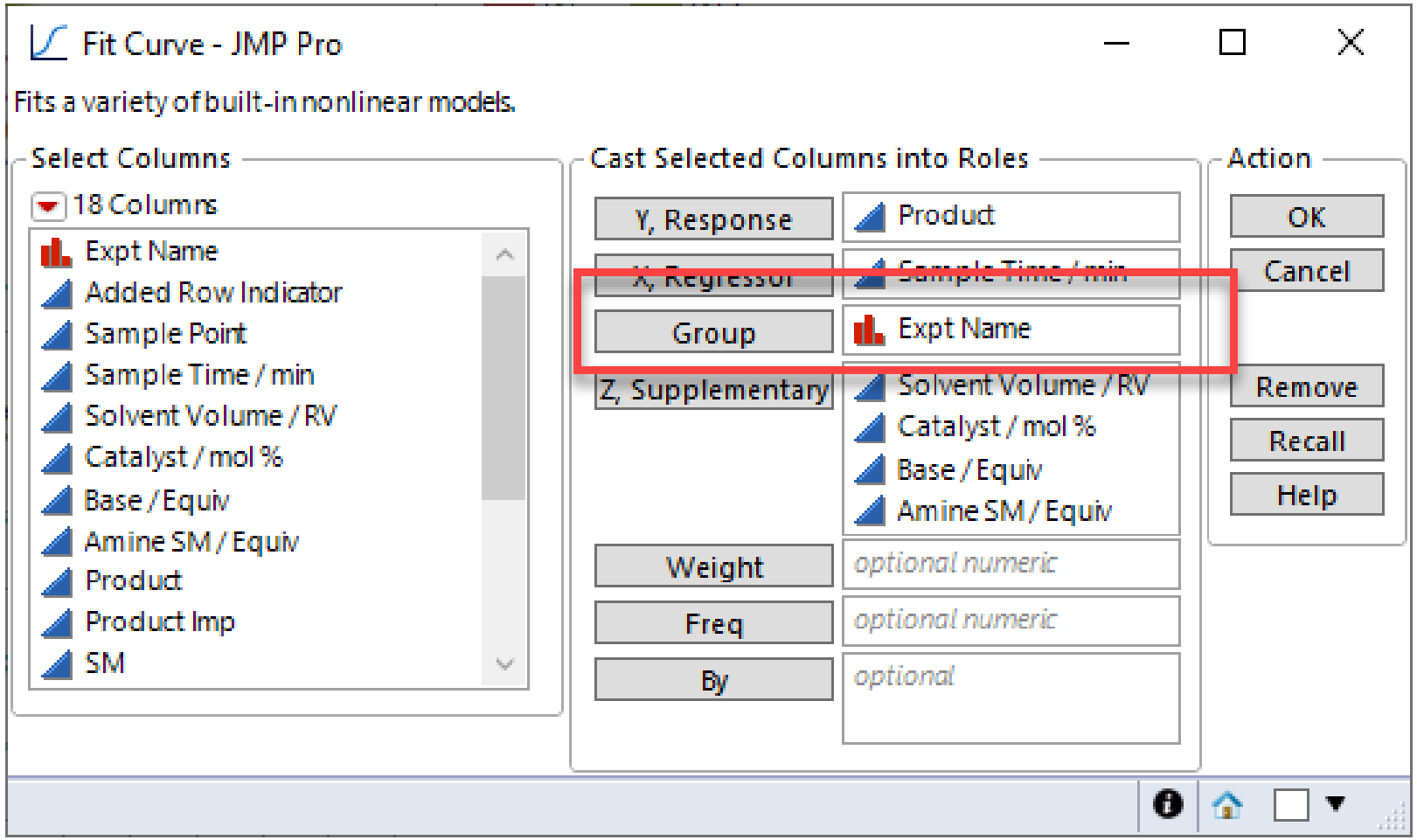
Launching Fit Curve



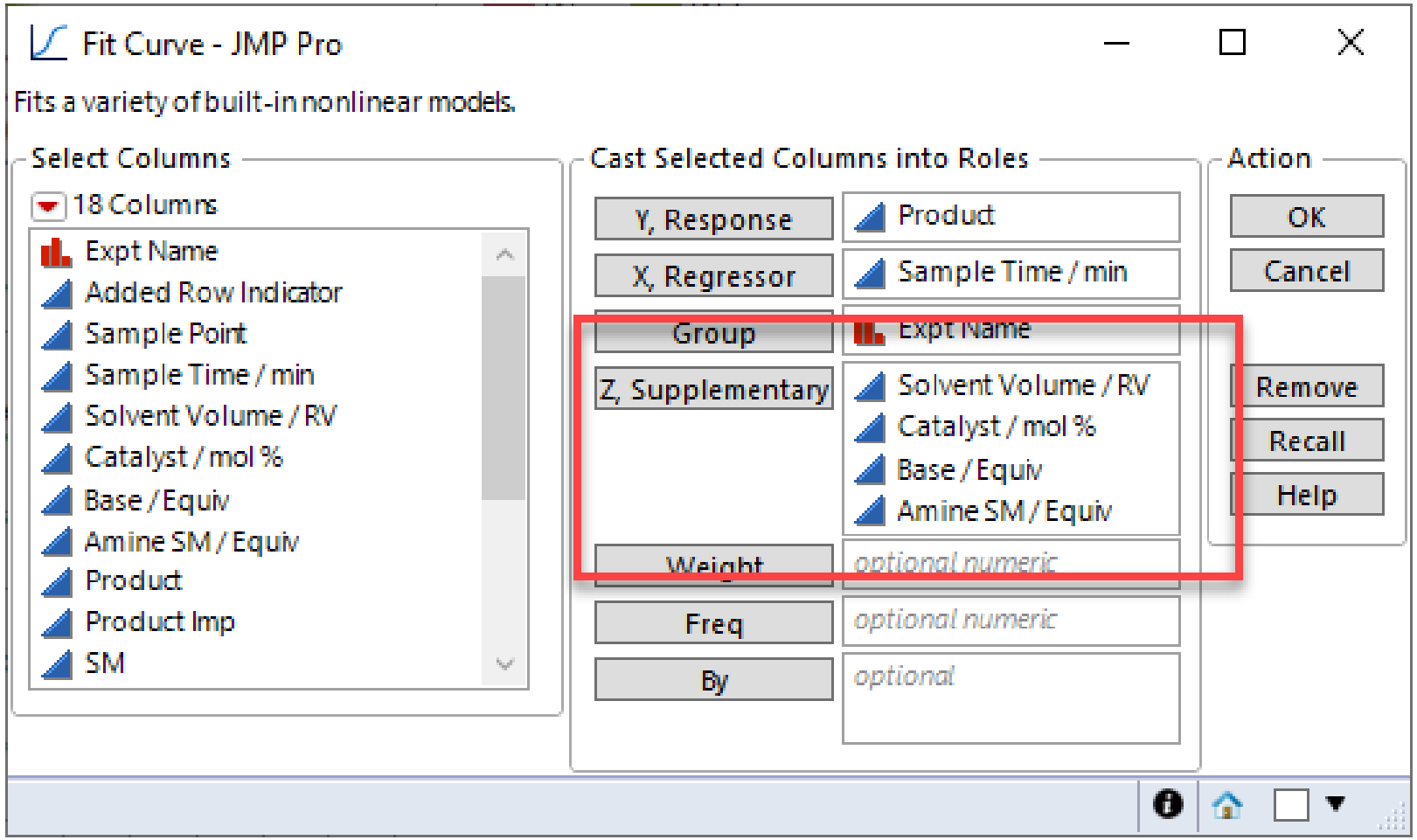
Launching Fit Curve



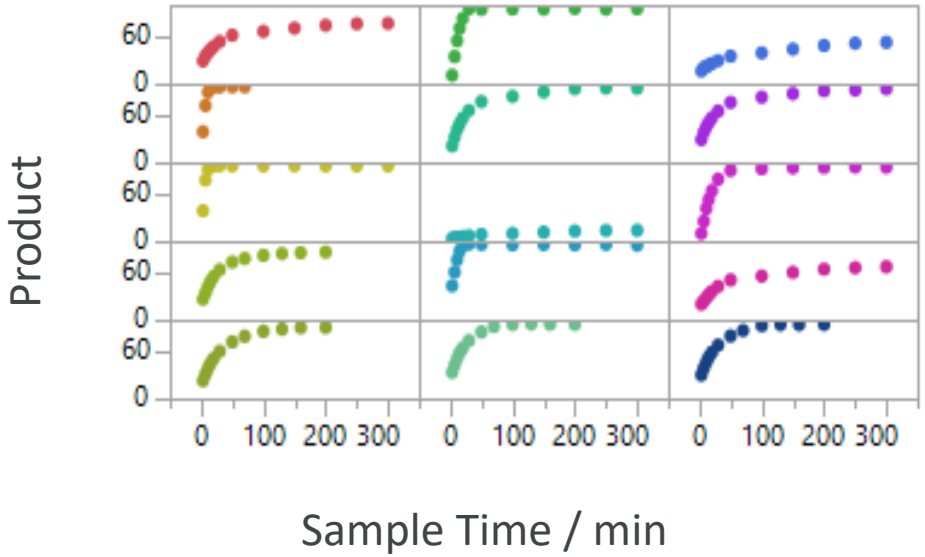
Launching Fit Curve

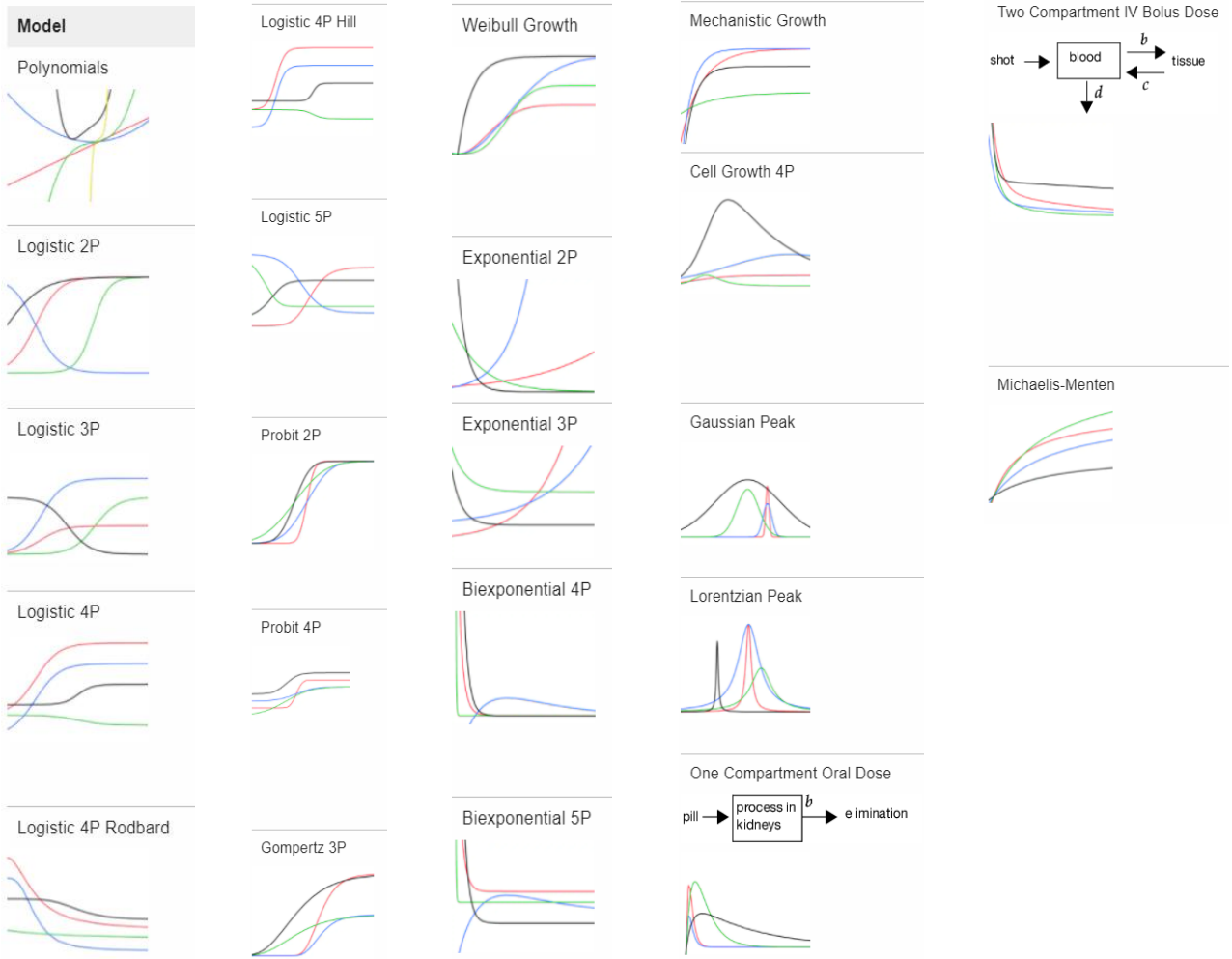


Launching Fit Curve

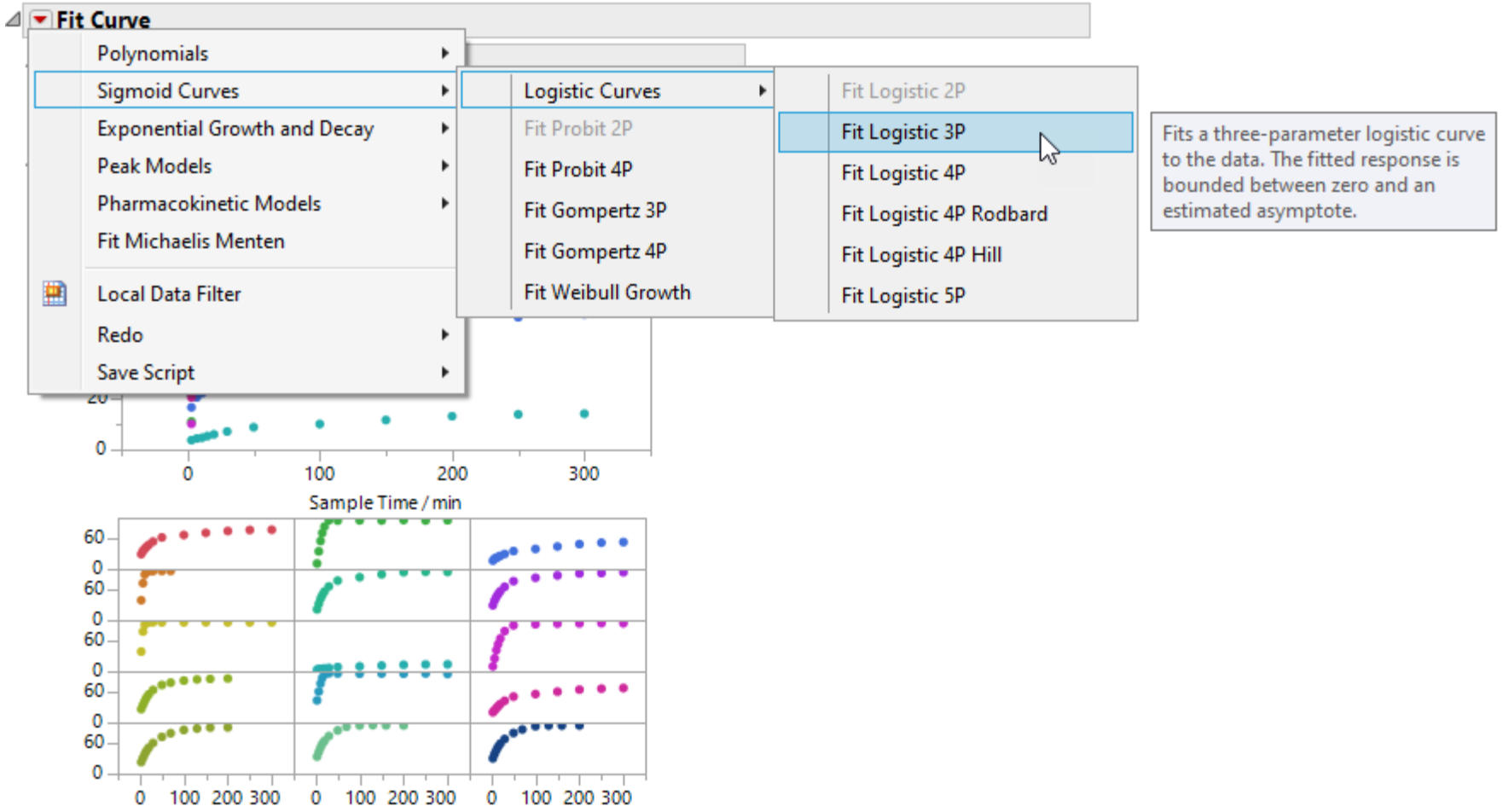


Data Visualization and the Nonlinear Model Key

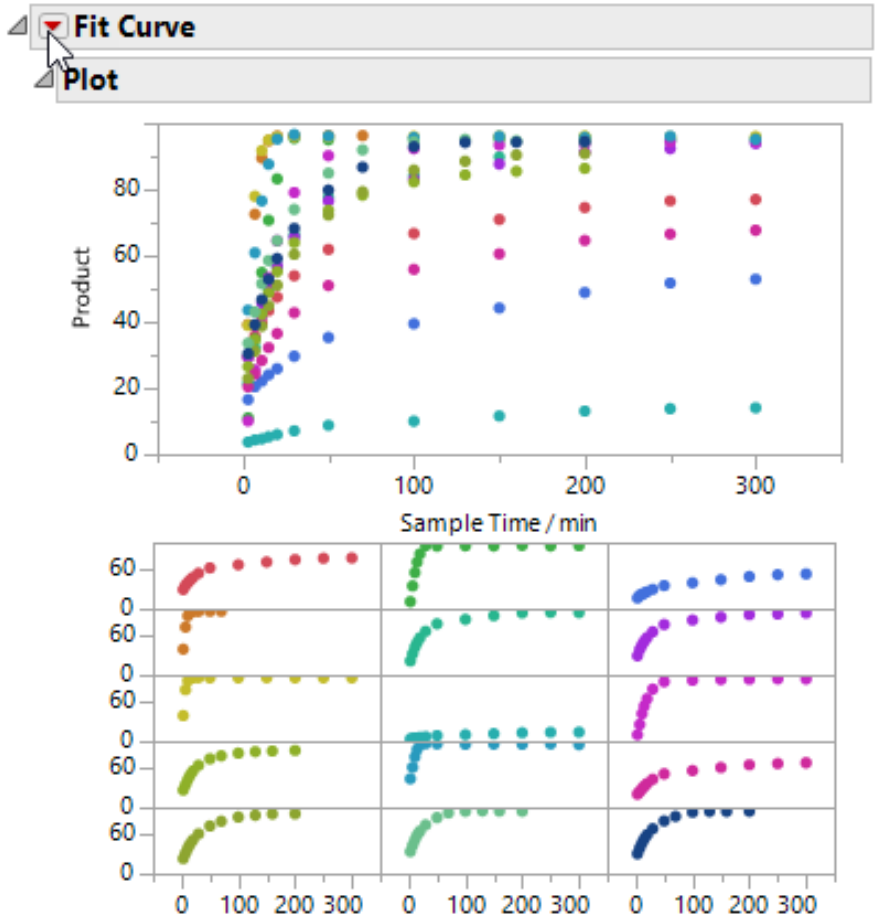


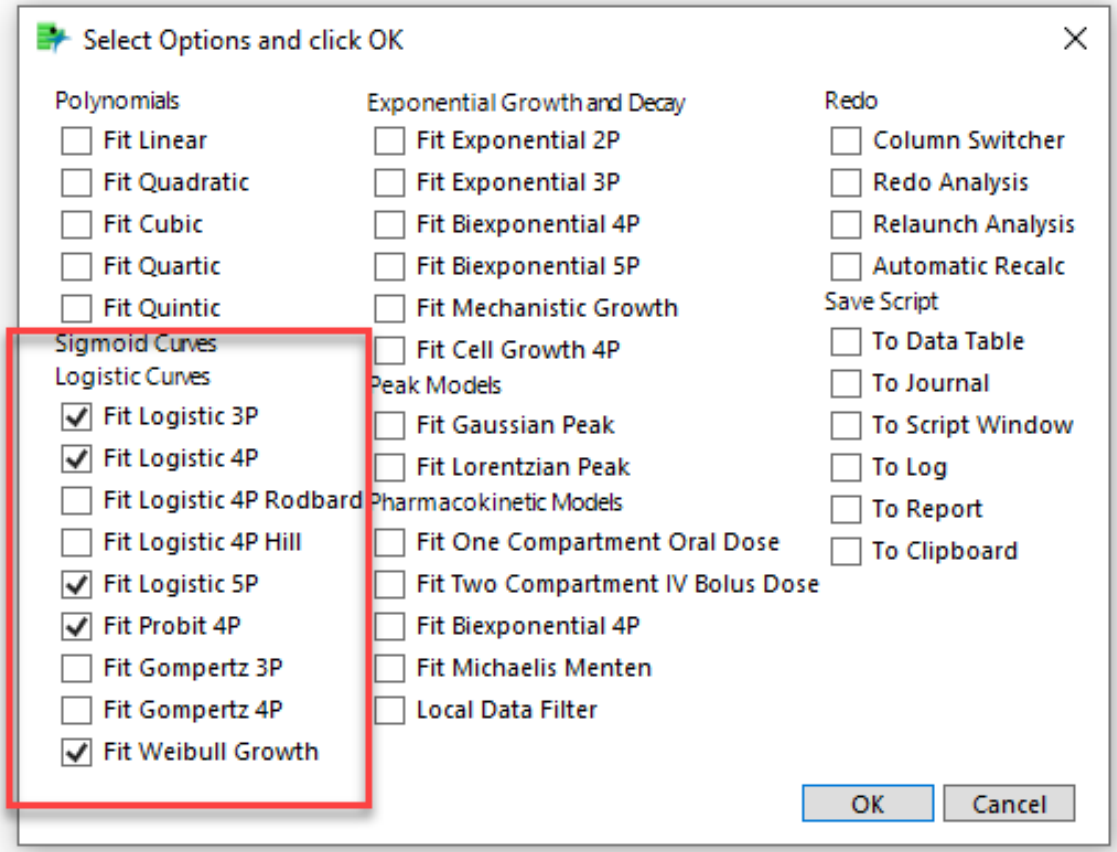


Fitting Nonlinear Models To the Data



Fitting Multiple Nonlinear Models To the Data





Alt+Right Click on red triangle to select multiple options

Fitting Multiple Nonlinear Models To the Data

Fit Curve

Model Comparison

Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Logistic 4P	774.21661	1		901.26526	287.54791	2.4788613	1.57444	0.9980418
Logistic 3P	925.40129	1.481e-33		1037.7242	970.09161	7.4052794	2.7212643	0.9933937
Weibull Growth	1132.045	1.989e-78		1244.3679	3138.5071	23.958069	4.8946981	0.9786268
Logistic 5P	1358.7337	1.19e-127		1481.4682	5000.6892	49.511775	7.0364604	0.9659454
Probit 4P	1463.7303	1.88e-150		1590.7789	14459.122	124.64761	11.164569	0.9015335

Fitting Multiple Nonlinear Models To the Data

Fit Curve

Model Comparison

Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Logistic 4P	774.21661	1		901.26526	287.54791	2.4788613	1.57444	0.9980418
Logistic 3P	925.40129	1.481e-33		1037.7242	970.09161	7.4052794	2.7212643	0.9933937
Weibull Growth	1132.045	1.989e-78		1244.3679	3138.5071	23.958069	4.8946981	0.9786268
Logistic 5P	1358.7337	1.19e-127		1481.4682	5000.6892	49.511775	7.0364604	0.9659454
Probit 4P	1463.7303	1.88e-150		1590.7789	14459.122	124.64761	11.164569	0.9015335

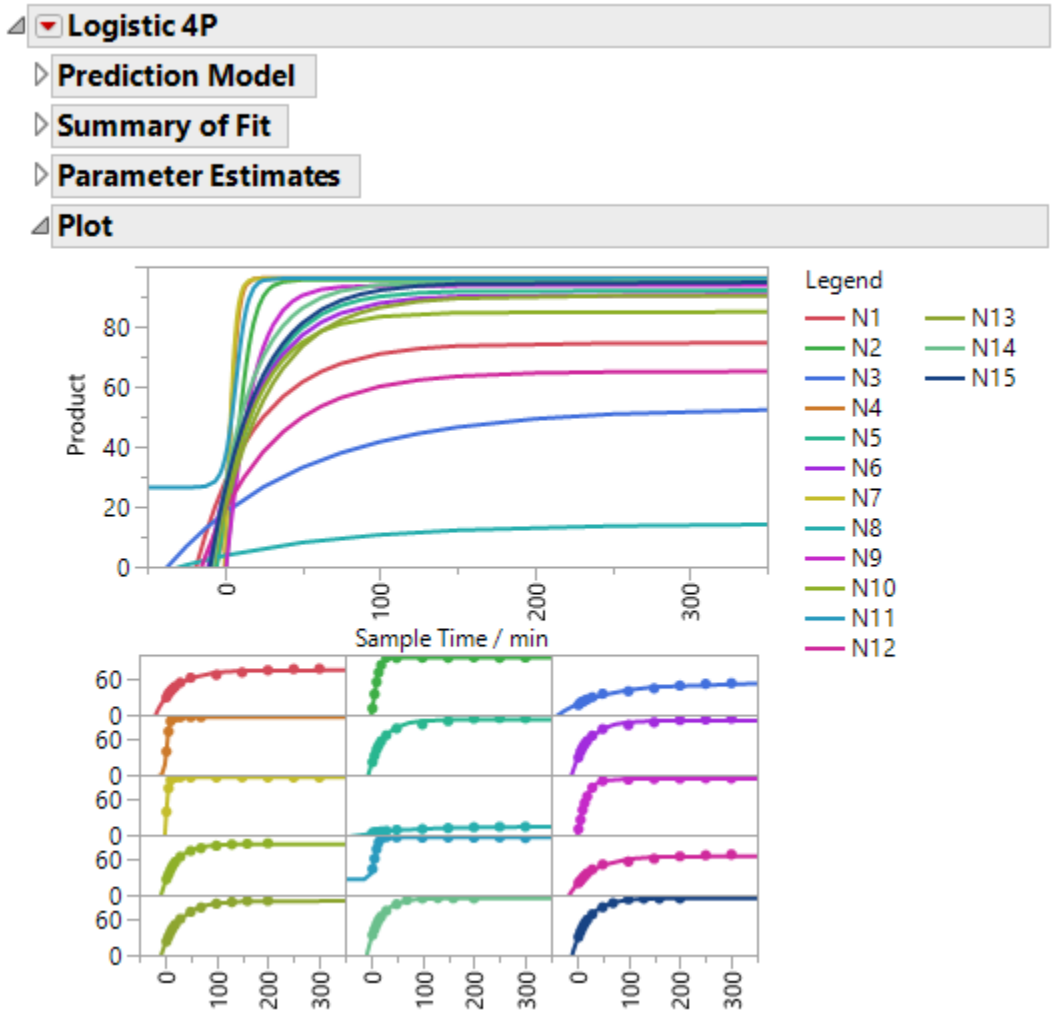
Fitting Multiple Nonlinear Models To the Data

Fit Curve

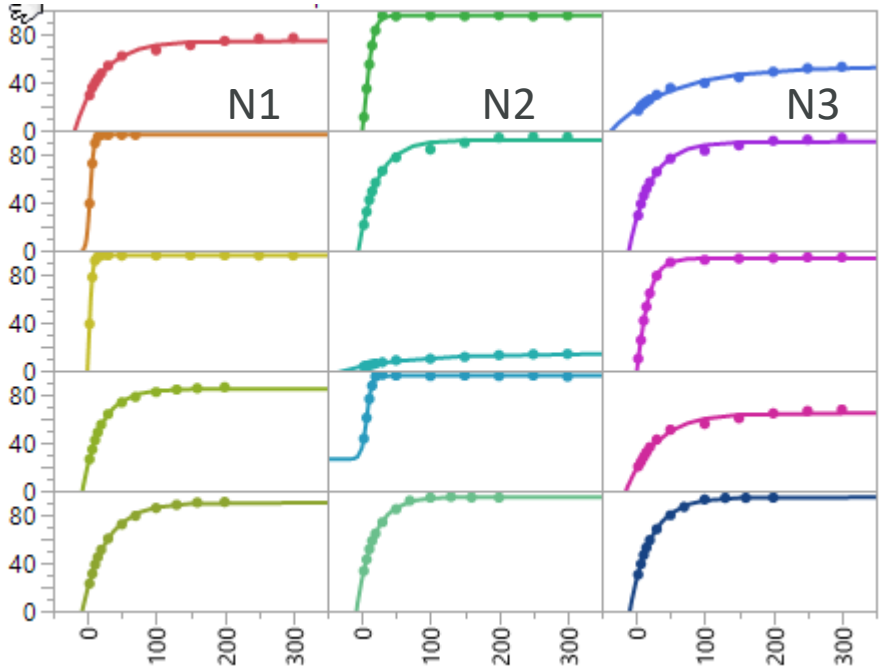
Model Comparison

Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Logistic 4P	774.21661	1		901.26526	287.54791	2.4788613	1.57444	0.9980418
Logistic 3P	925.40129	1.481e-33		1037.7242	970.09161	7.4052794	2.7212643	0.9933937
Weibull Growth	1132.045	1.989e-78		1244.3679	3138.5071	23.958069	4.8946981	0.9786268
Logistic 5P	1358.7337	1.19e-127		1481.4682	5000.6892	49.511775	7.0364604	0.9659454
Probit 4P	1463.7303	1.88e-150		1590.7789	14459.122	124.64761	11.164569	0.9015335

Logistic 4P Model



Logistic 4P Model



Logistic 4P

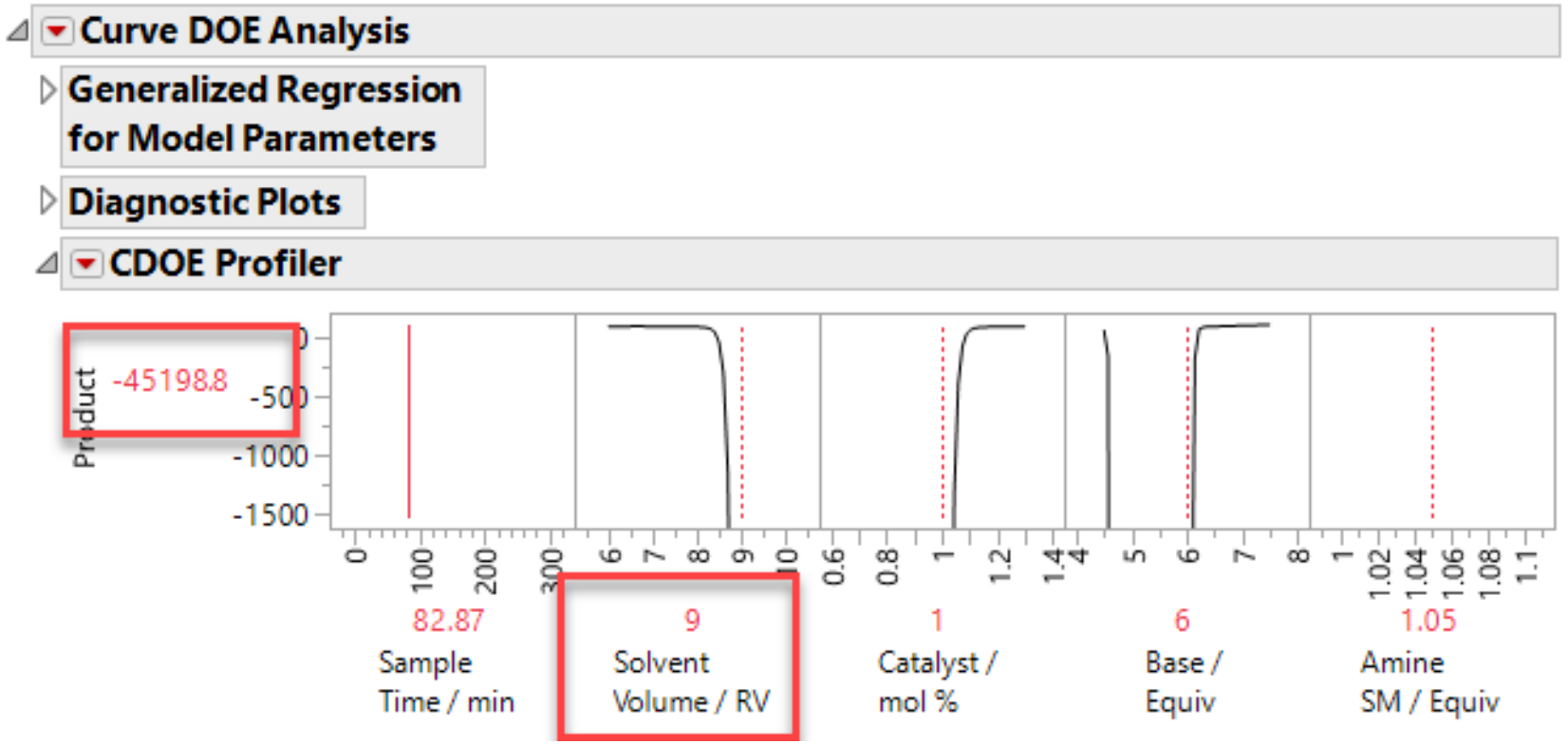
Prediction Model

Summary of Fit

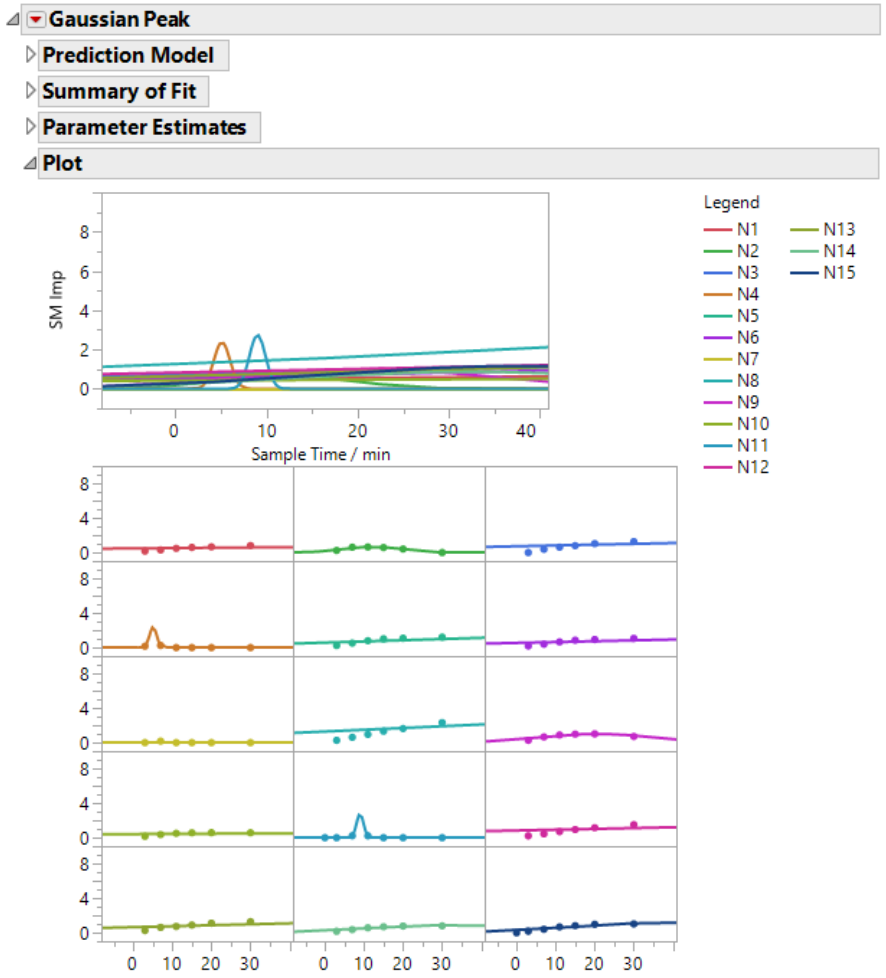
Parameter Estimates

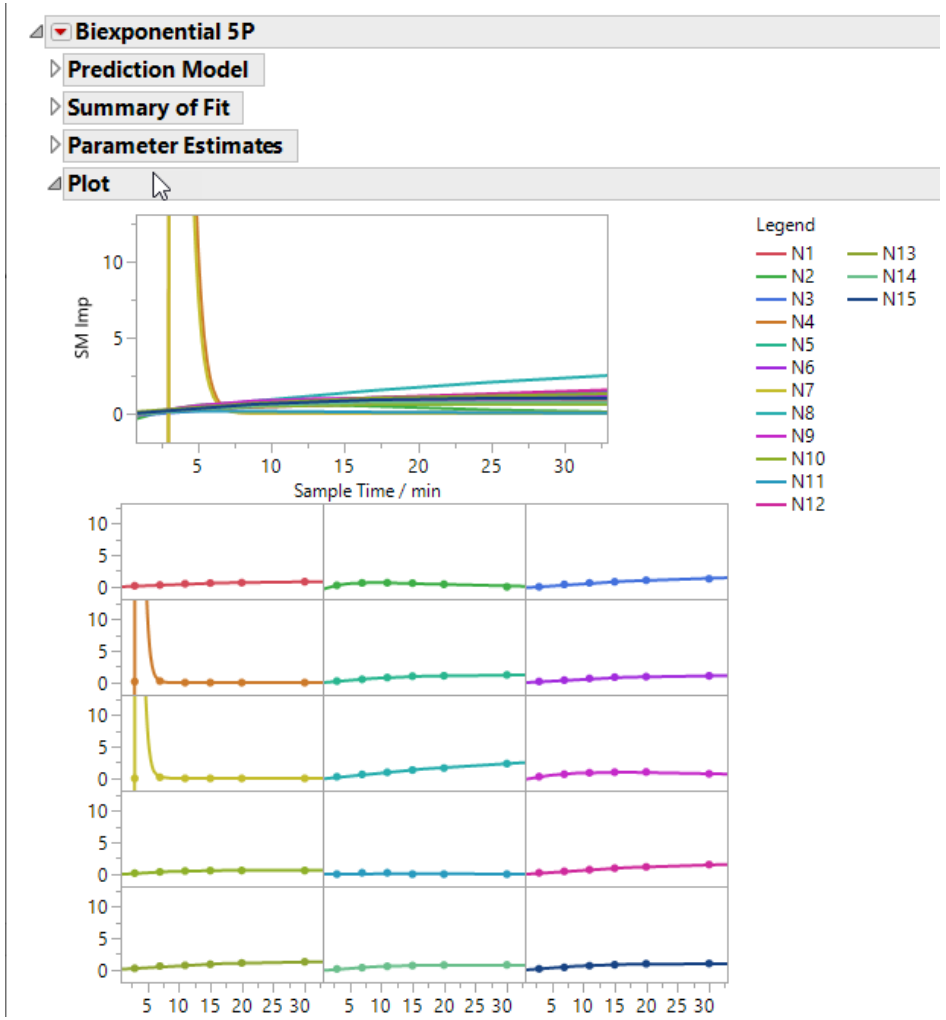
Parameter	Group	Estimate	Std Error
Growth Rate	N1	0.0254486	0.0080311
Inflection Point	N1	-431.7105	1141239
Lower Asymptote	N1	-2751999	7.992e+10
Upper Asymptote	N1	74.568627	0.9428258
Growth Rate	N2	0.1636571	0.0169077
Inflection Point	N2	5.4759384	1.9387716
Lower Asymptote	N2	-44.86278	21.420603
Upper Asymptote	N2	95.642648	0.6257901
Growth Rate	N3	0.0112441	0.0071275
Inflection Point	N3	-1159.956	33420084
Lower Asymptote	N3	-16149627	6.069e+12
Upper Asymptote	N3	92.970356	2.8297727
Growth Rate	N4	0.3668455	0.0869127
Inflection Point	N4	3.8772156	2.1243712
Lower Asymptote	N4	-2.131121	35.994322
Upper Asymptote	N4	96.311589	0.7976433

Logistic 4P Model

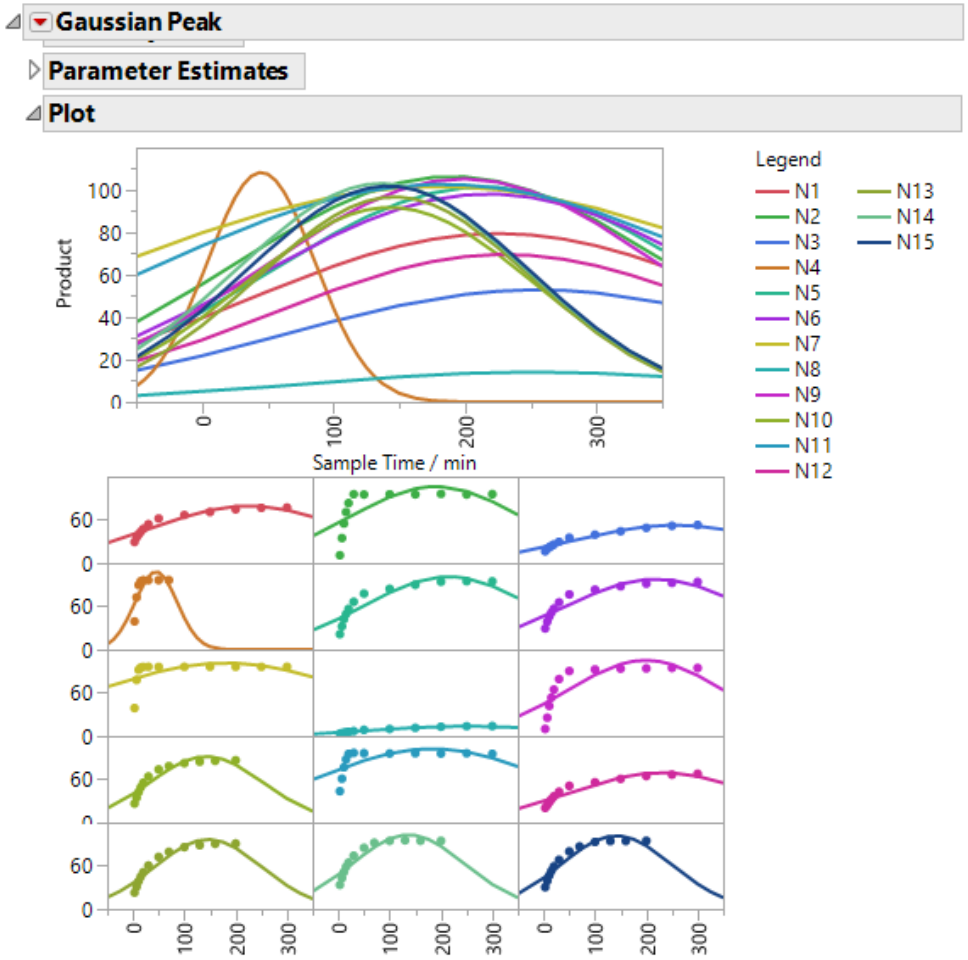


Checking the Nonlinear Model

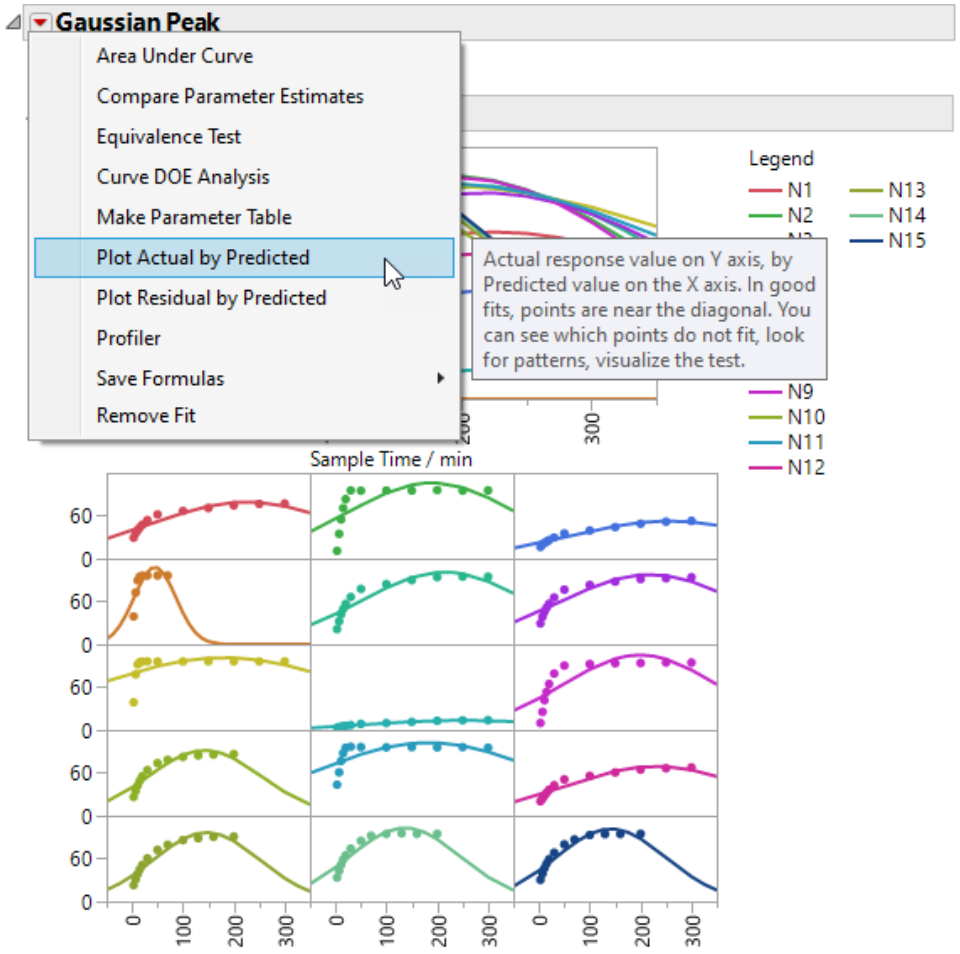


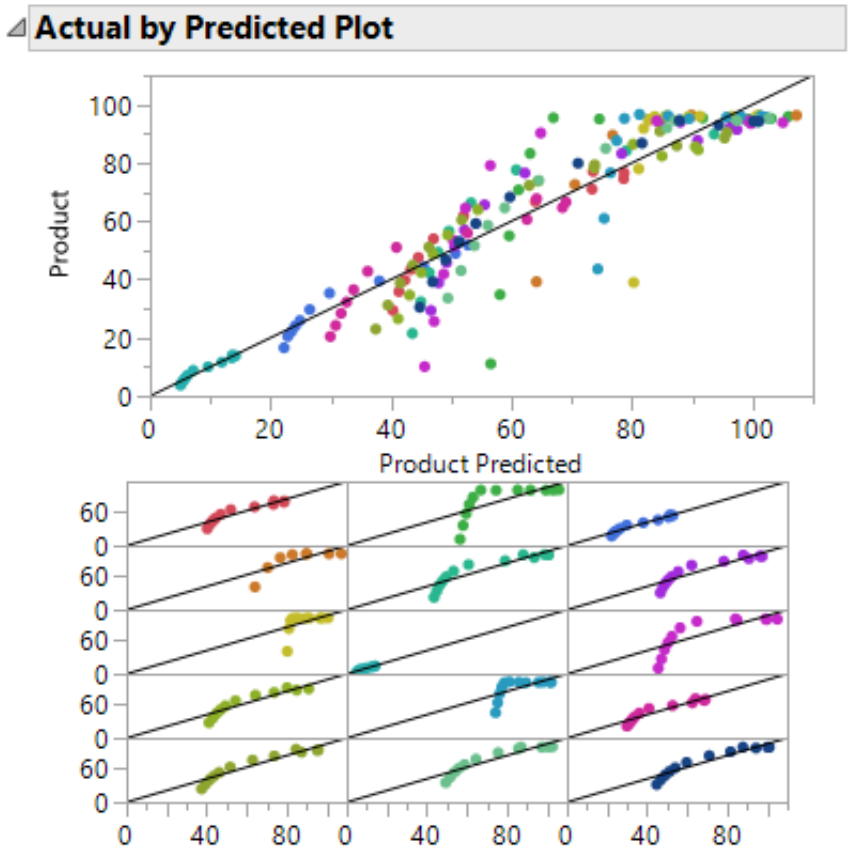


Checking the Nonlinear Model



Checking the Nonlinear Model





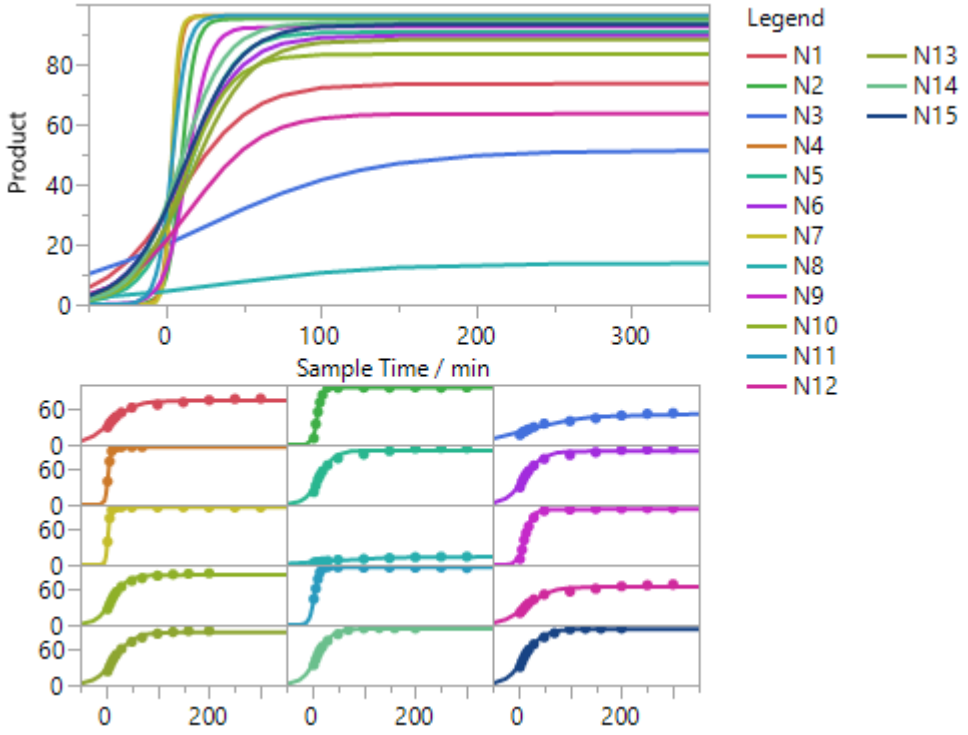
Three Parameter Logistic Fit

Logistic 3P

Prediction Model

$$\frac{c}{\left(1 + \text{Exp}\left(-a \cdot \left(\text{Sample Time / min} - b\right)\right)\right)}$$

a = Growth Rate
b = Inflection Point
c = Asymptote

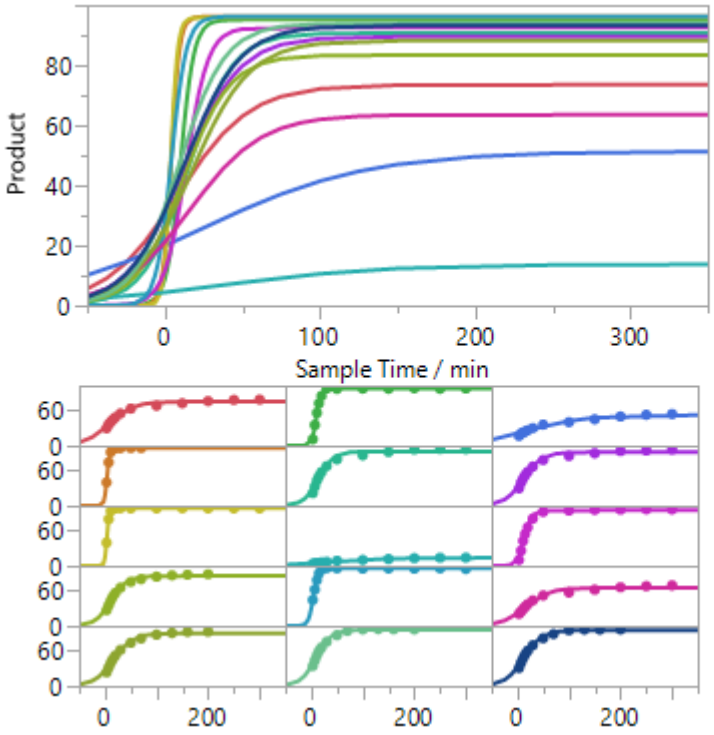


Three Parameter Logistic Fit

$$\frac{c}{\left(1+\mathrm{Exp}\left(-a\bullet\left(\mathrm{Sample\ Time}/\mathrm{min}-b\right)\right)\right)}$$

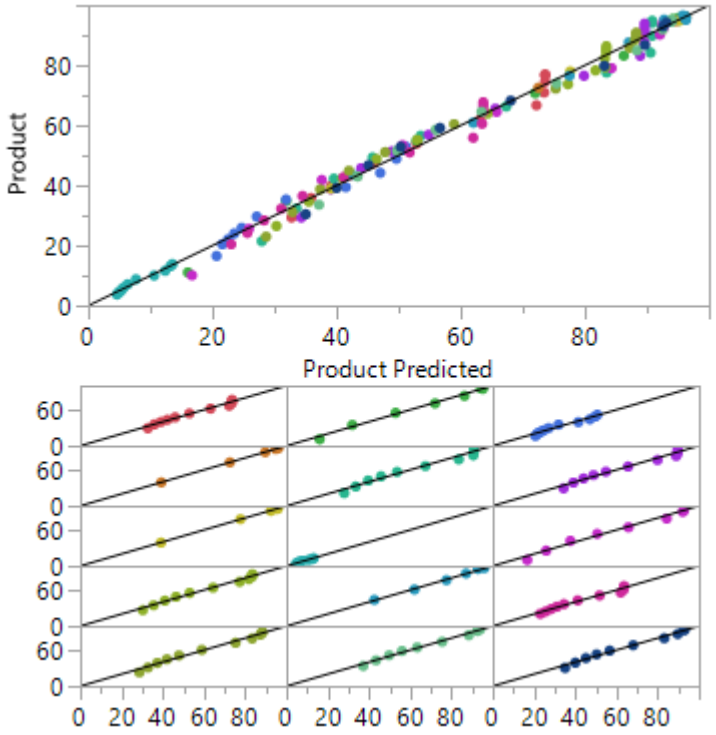
a = Growth Rate
b = Inflection Point
c = Asymptote

Regression Plot

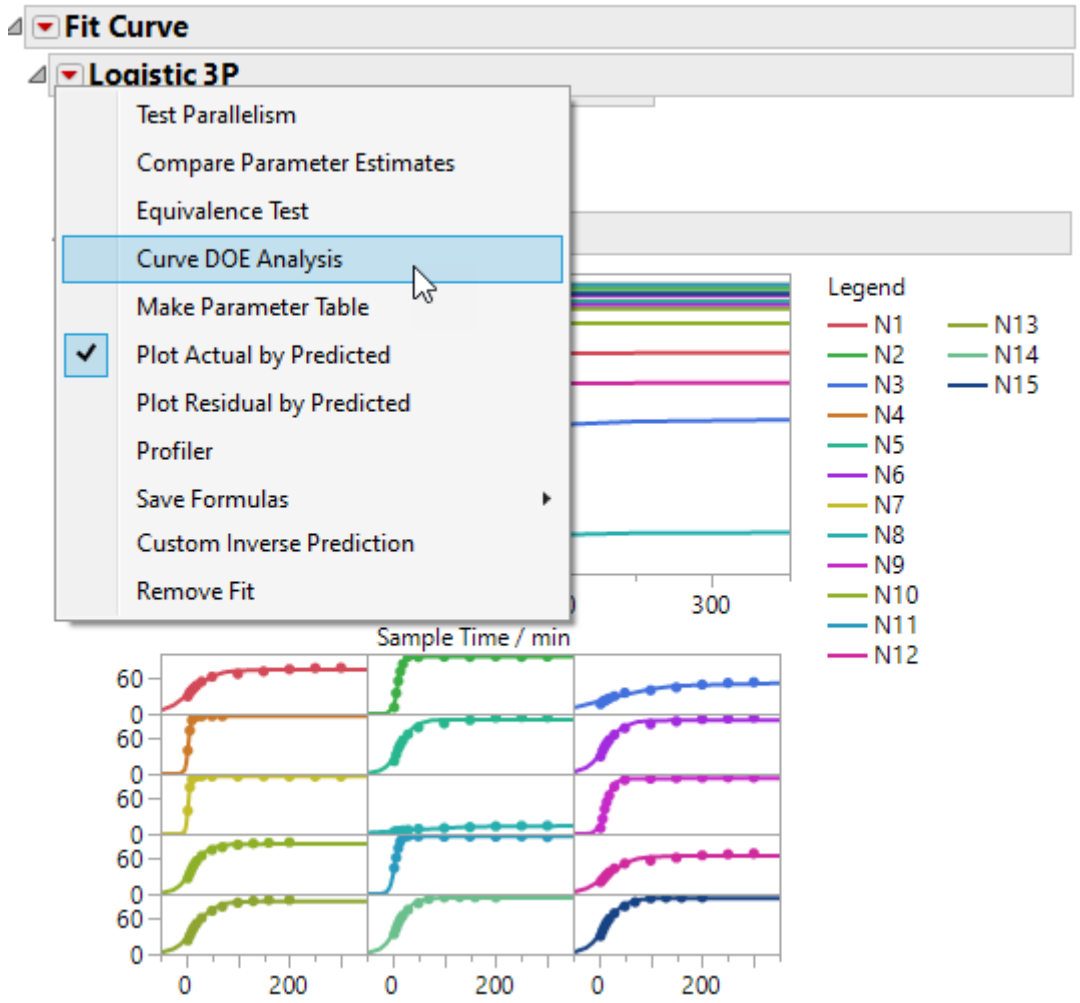


Actual By Predicted Plot

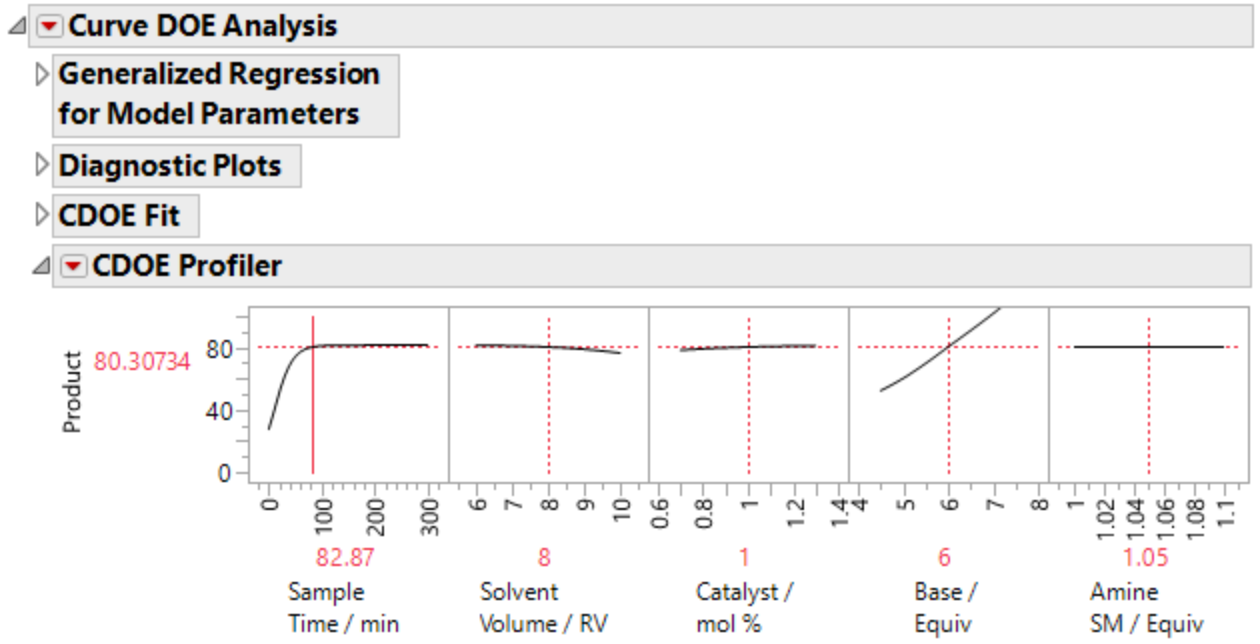
(Before CDOE)



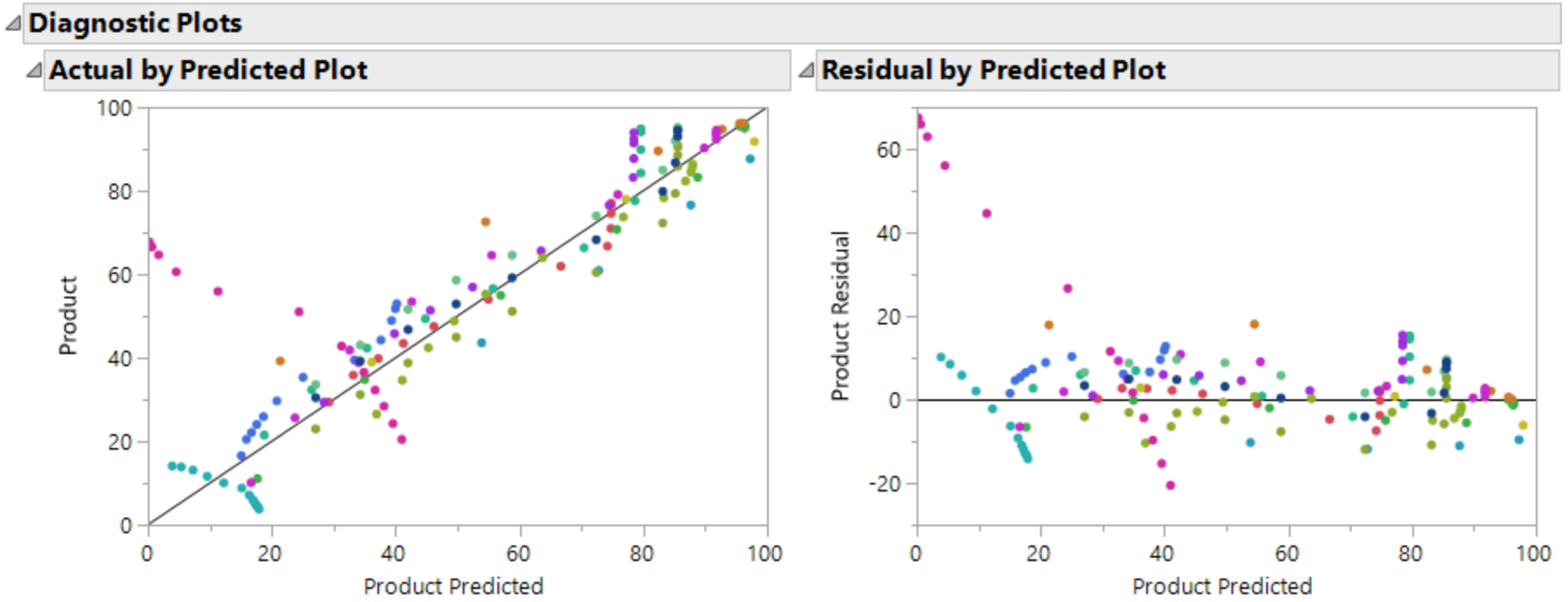
Three Parameter Logistic Fit



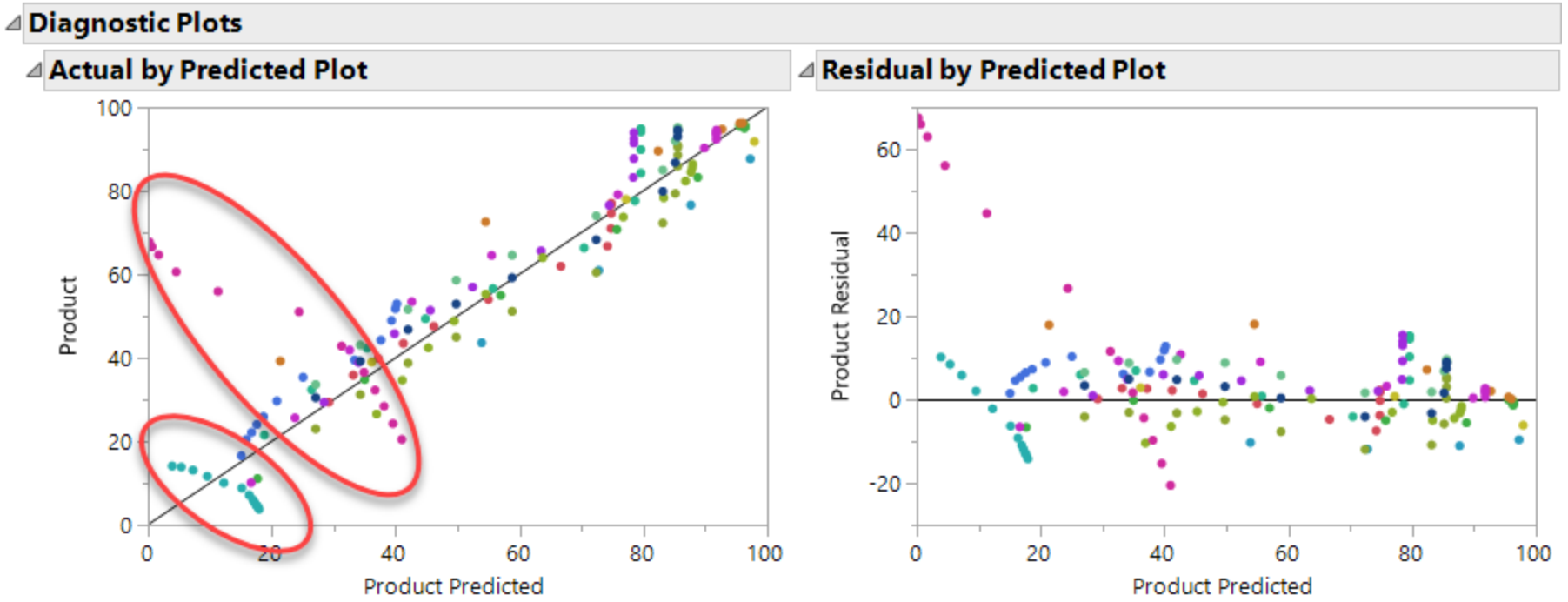
Three Parameter Logistic Fit



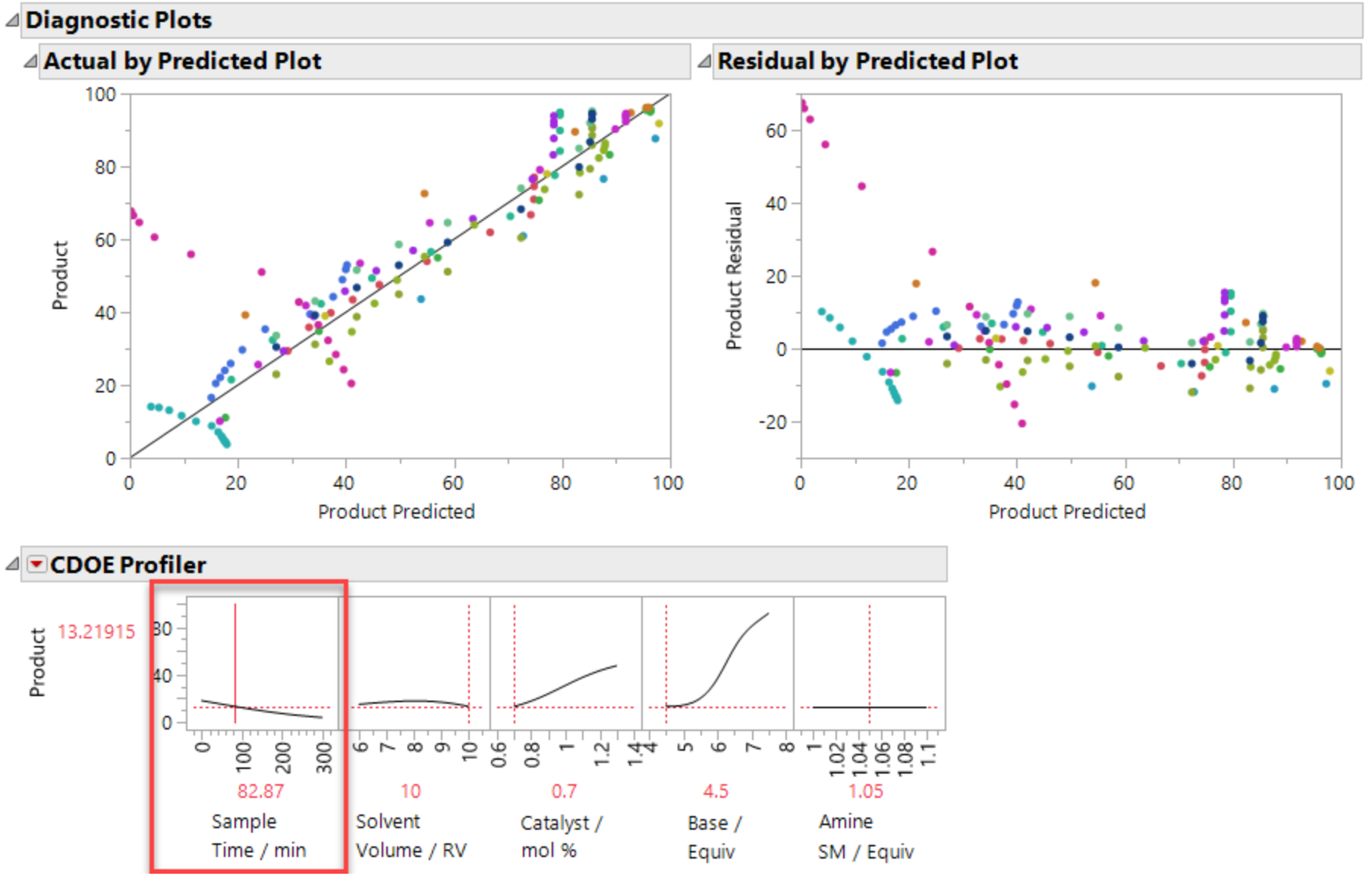
Three Parameter Logistic Fit



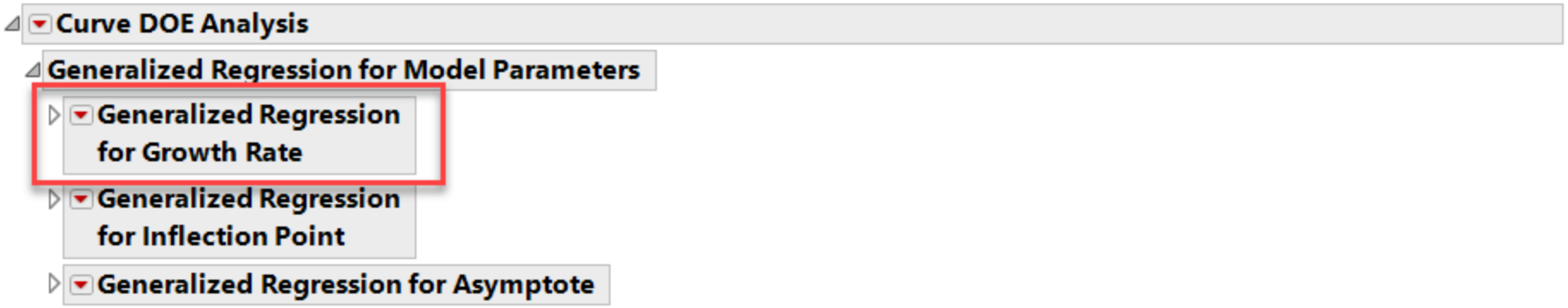
Three Parameter Logistic Fit



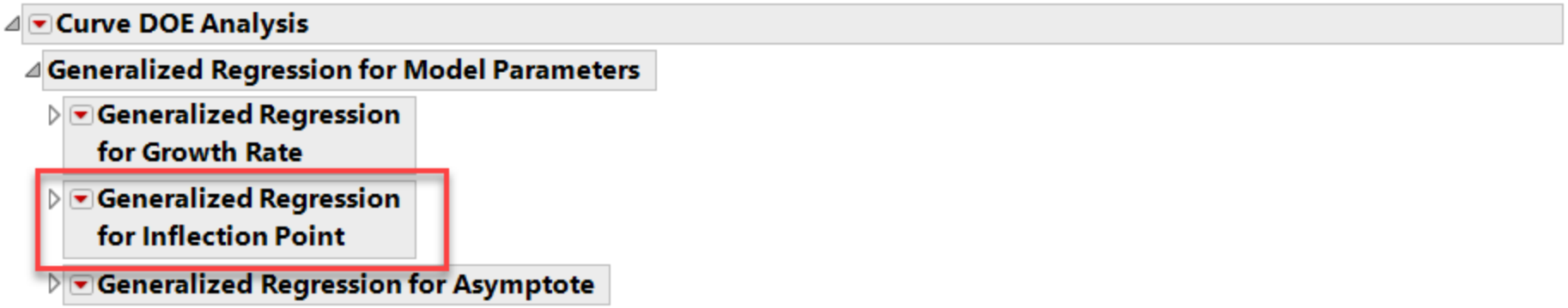
Three Parameter Logistic Fit



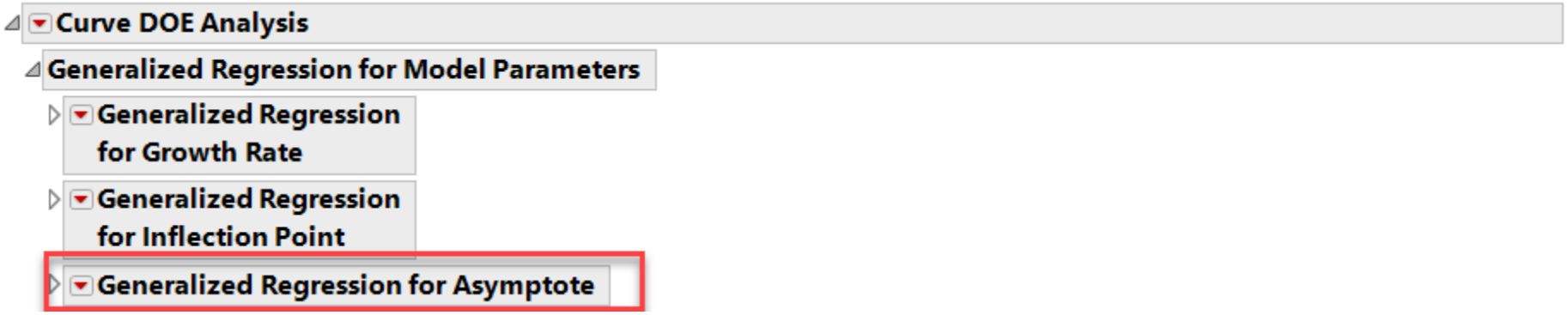
Three Parameter Logistic Fit



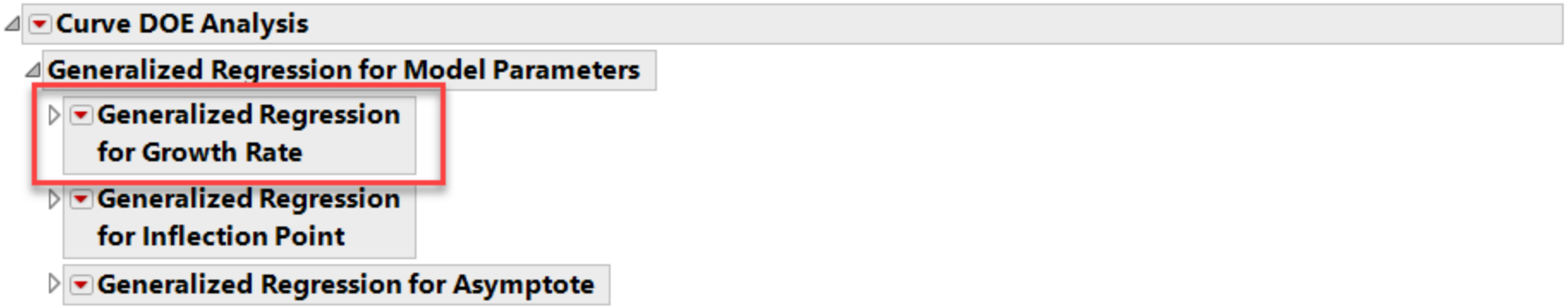
Three Parameter Logistic Fit



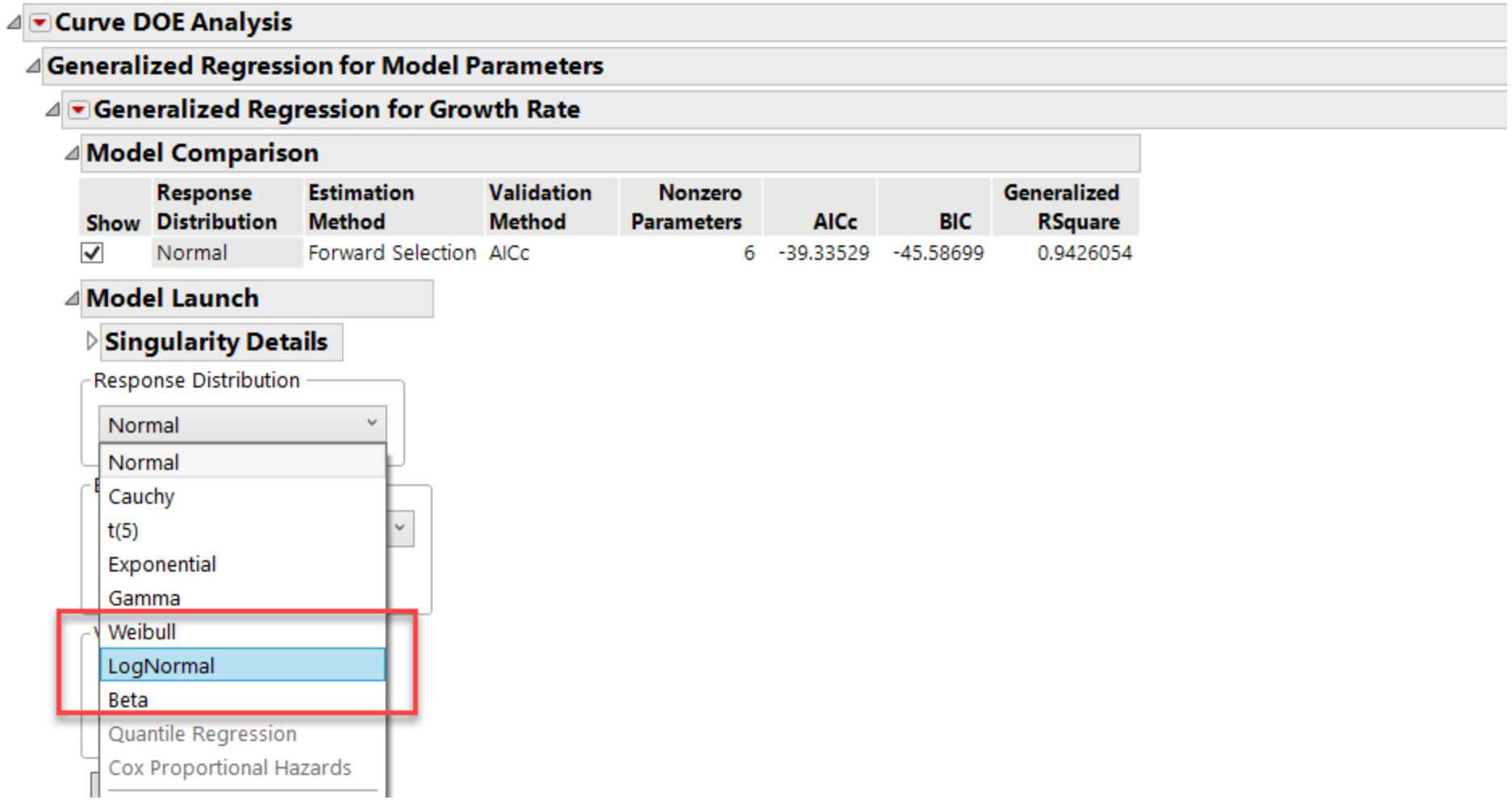
Three Parameter Logistic Fit



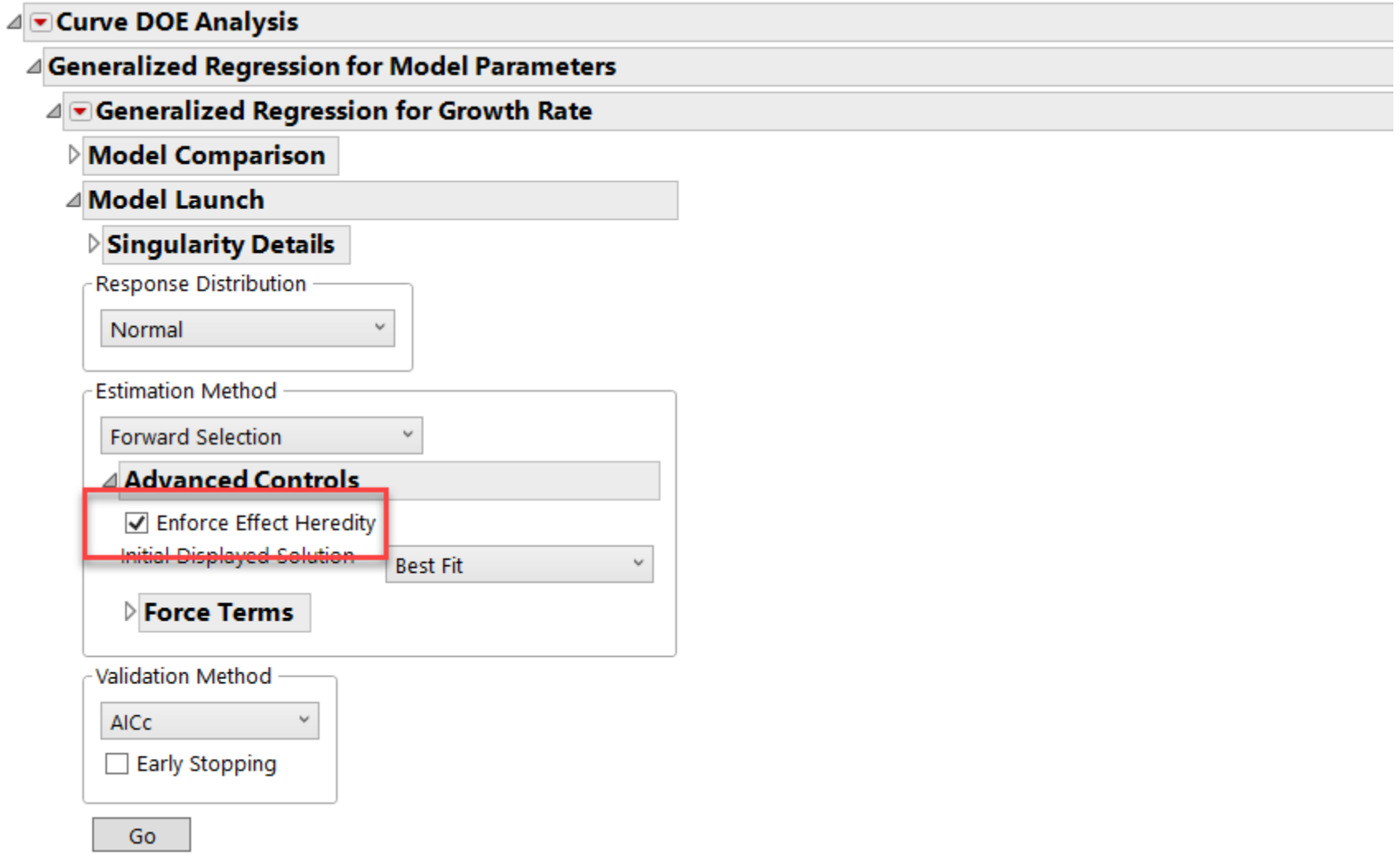
Three Parameter Logistic Fit



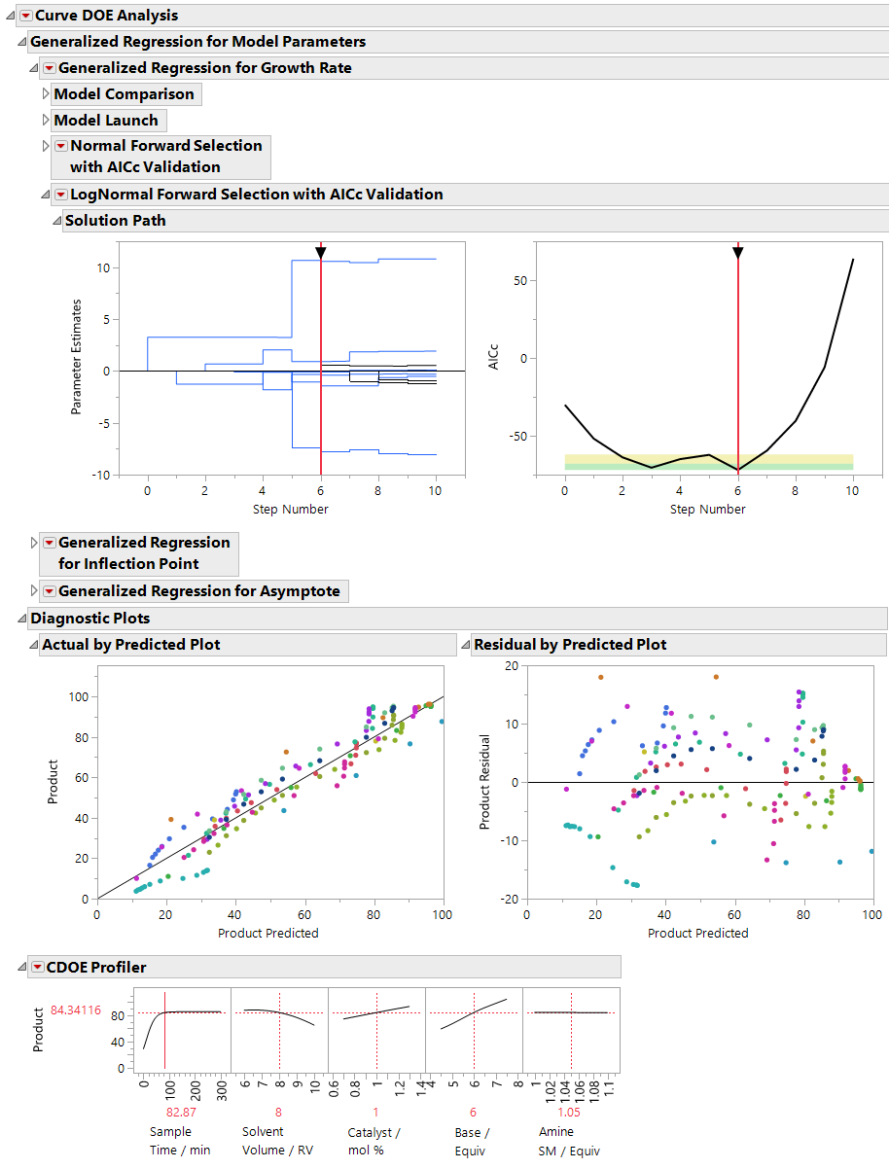
Three Parameter Logistic Fit



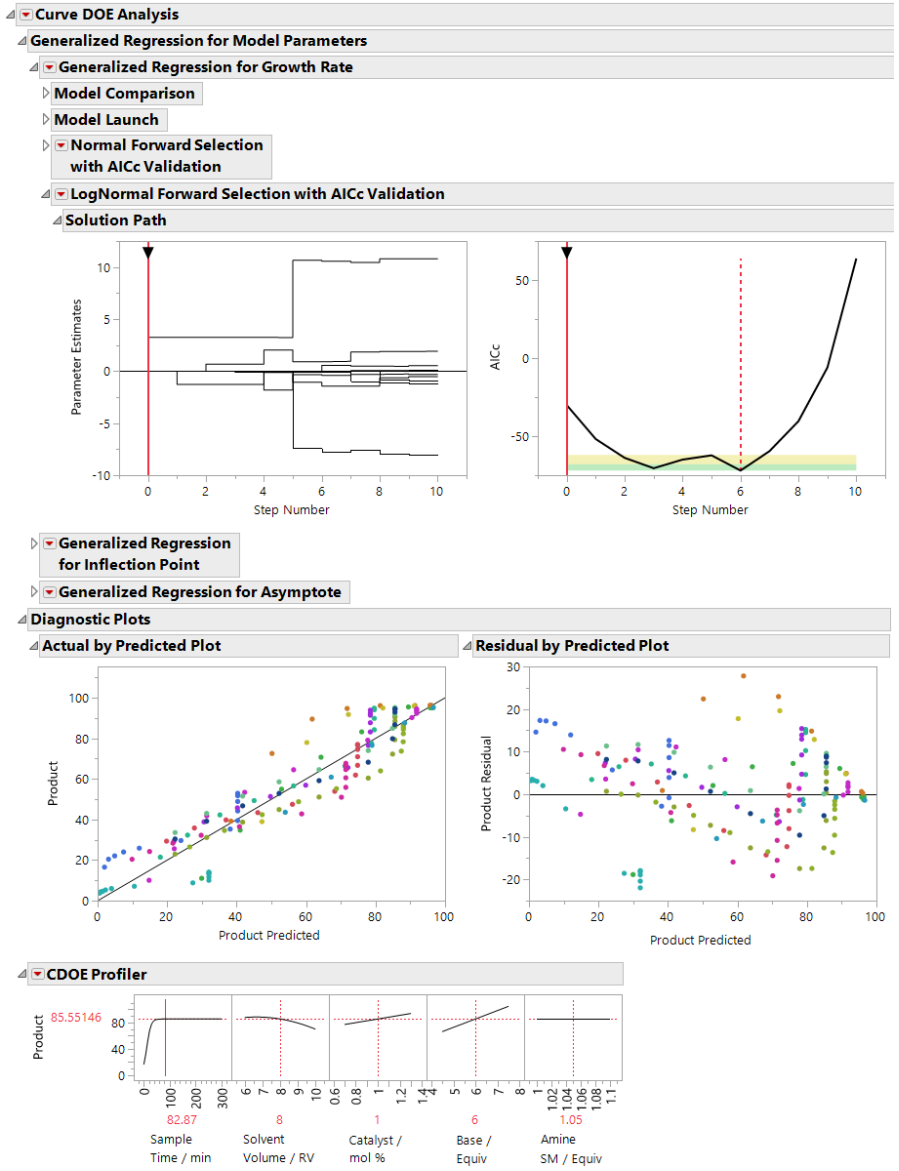
Three Parameter Logistic Fit



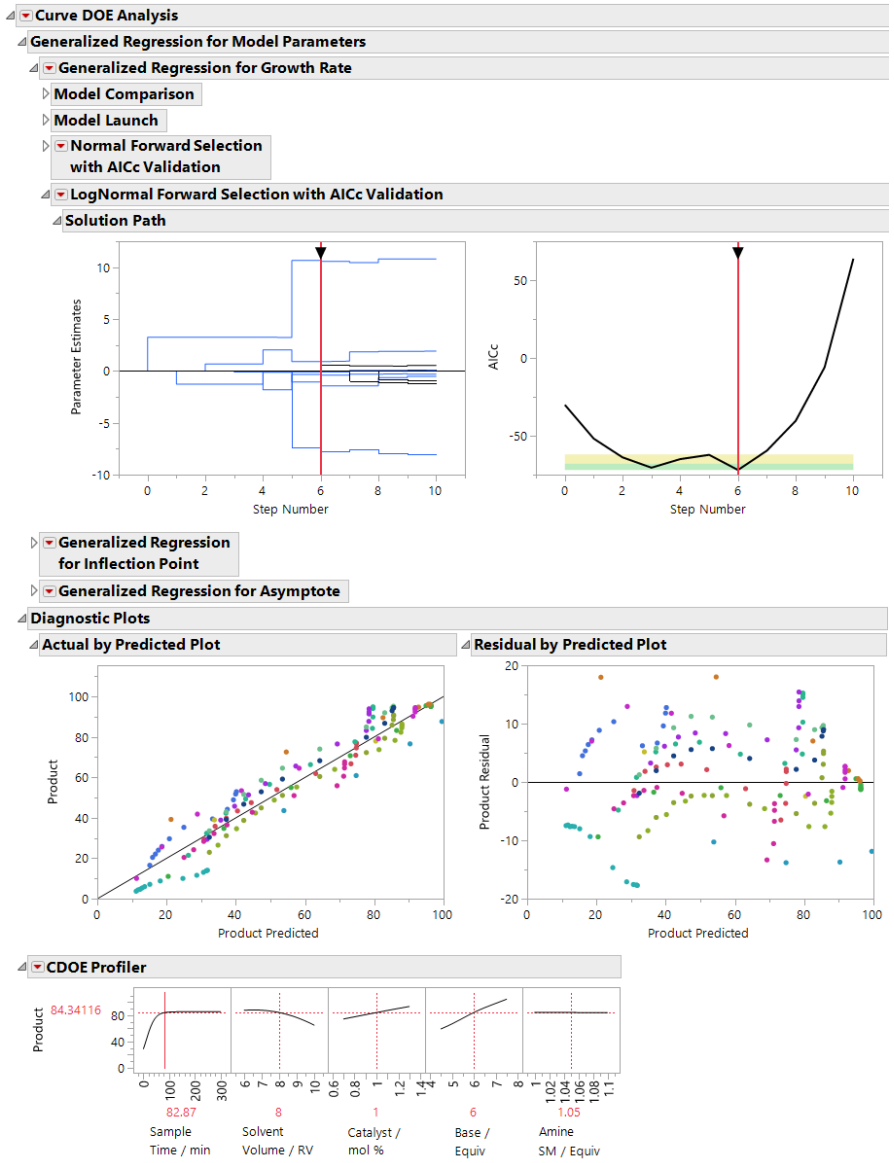
Three Parameter Logistic Fit



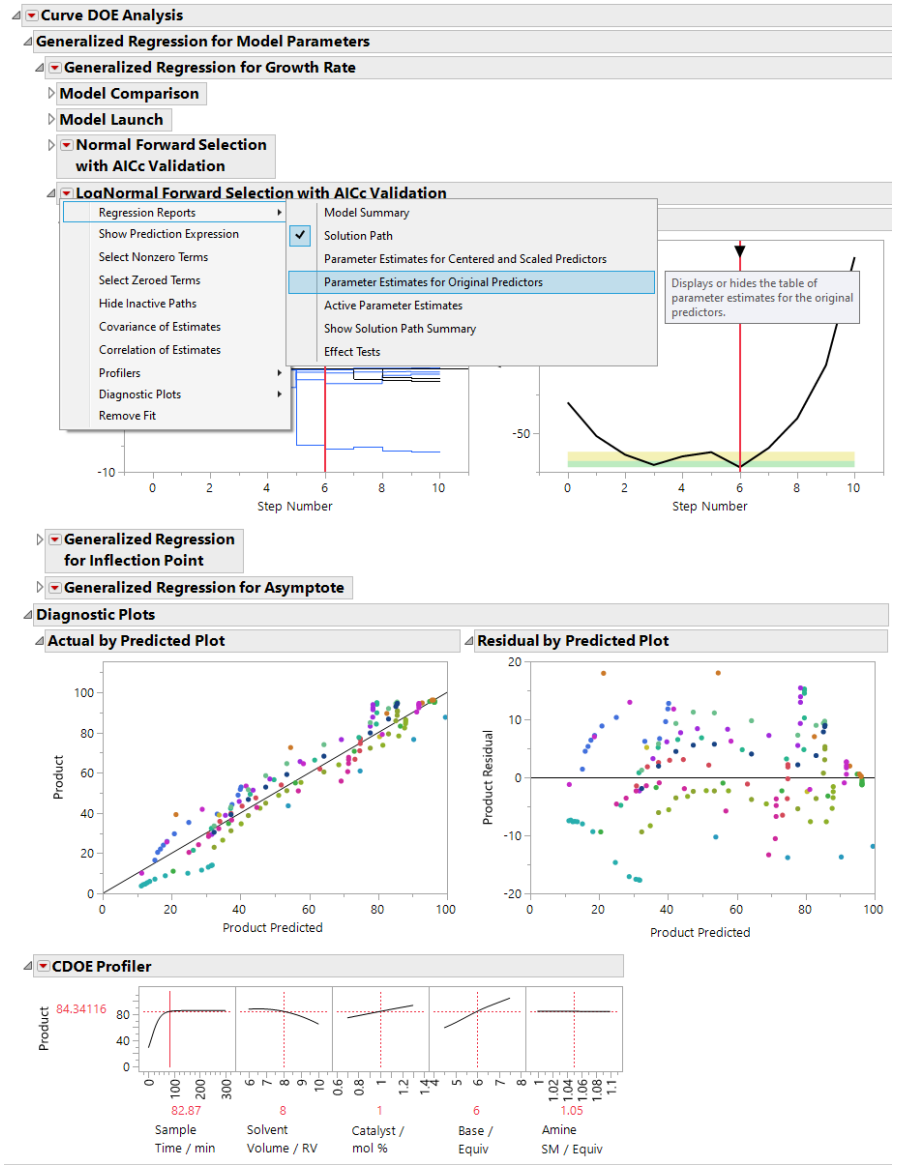
Three Parameter Logistic Fit



Three Parameter Logistic Fit



Three Parameter Logistic Fit



Three Parameter Logistic Fit

Curve DOE Analysis

Generalized Regression for Model Parameters

Generalized Regression for Growth Rate

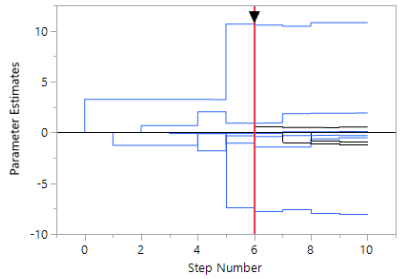
Model Comparison

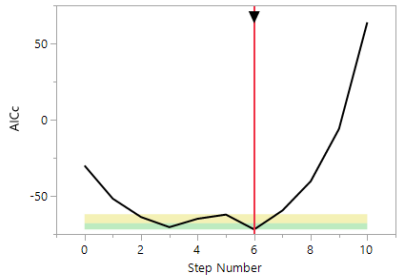
Model Launch

Normal Forward Selection with AICc Validation

LogNormal Forward Selection with AICc Validation

Solution Path



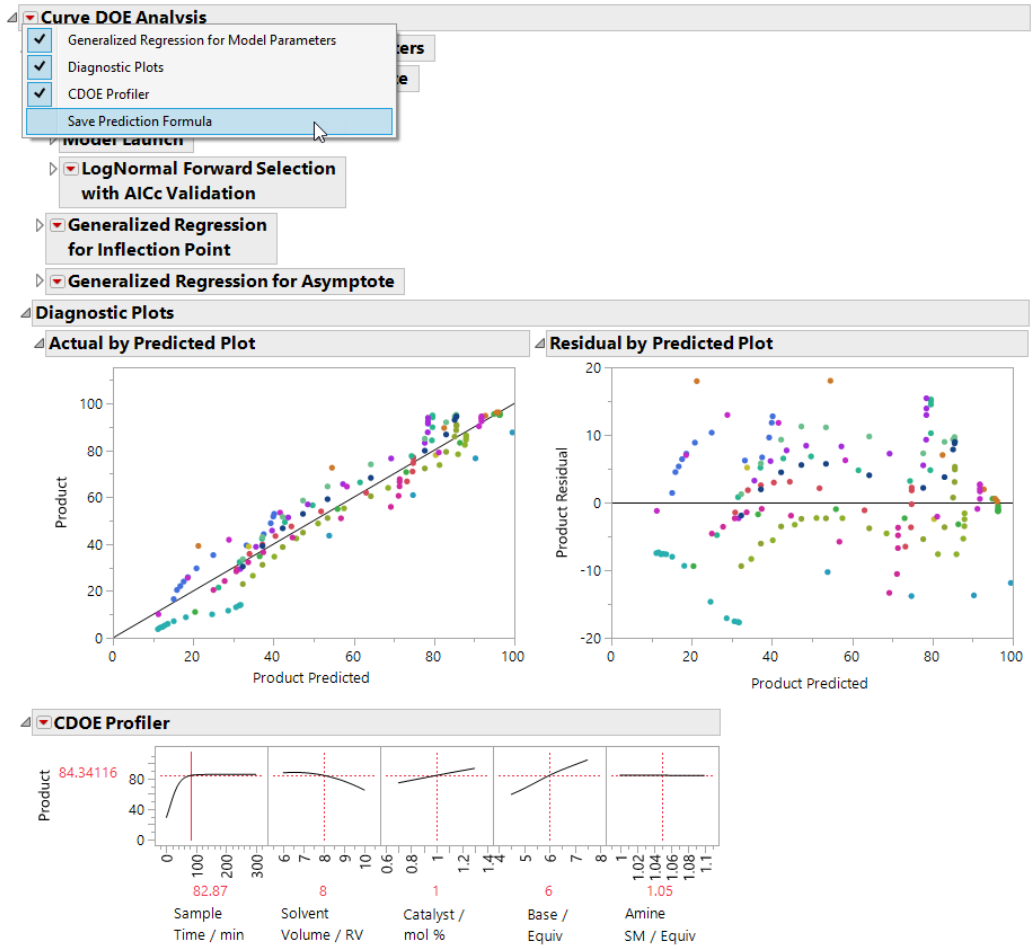


Parameter Estimates for Original Predictors

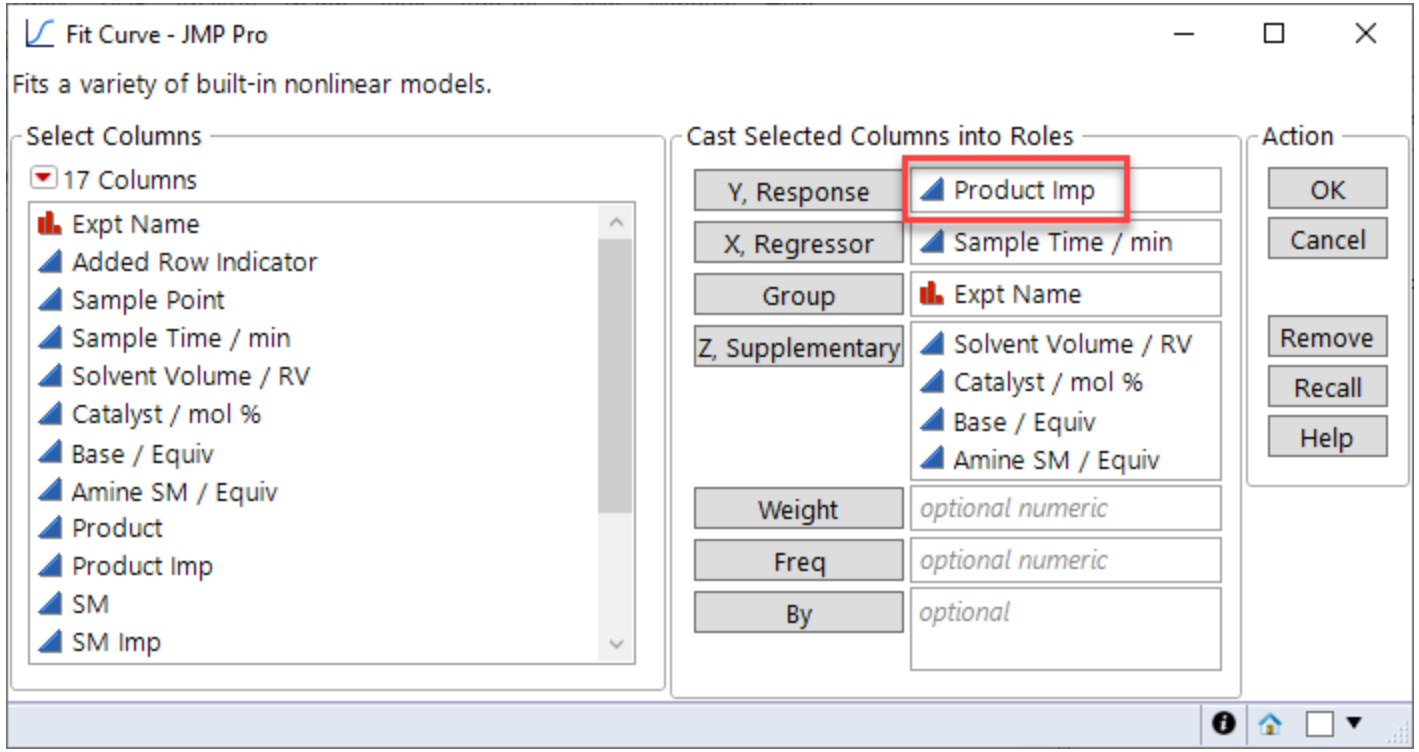
Term	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%	Singularity Details
Base / Equiv*Base / Equiv	0.1866392	0.0224568	69.073488	<.0001*	0.1426248	0.2306537	=17.111*Intercept-3
Solvent Volume / RV	-0.168498	0.0291016	33.524041	<.0001*	-0.225536	-0.11146	
Base / Equiv	-1.564481	0.2715392	33.195207	<.0001*	-2.096688	-1.032274	
Catalyst / mol %*Catalyst / mol %	0.4771401	0.1410395	11.444846	0.0007*	0.2007076	0.7535725	
Intercept	1.5977317	0.7678052	4.3301771	0.0374*	0.0928613	3.1026022	
Amine SM / Equiv	-0.59341	0.3939215	2.2692895	0.1320	-1.365481	0.1786622	=0.6796*Intercept+
Solvent Volume / RV*Catalyst / mol %	-0.033073	0.0329536	1.0072704	0.3156	-0.097661	0.0315147	
Solvent Volume / RV*Solvent Volume / RV	0	0	0	1.0000	0	0	
Solvent Volume / RV*Base / Equiv	0	0	0	1.0000	0	0	
Solvent Volume / RV*Amine SM / Equiv	0	0	0	1.0000	0	0	
Catalyst / mol %	0	0	0	1.0000	0	0	
Catalyst / mol %*Base / Equiv	0	0	0	1.0000	0	0	
Catalyst / mol %*Amine SM / Equiv	0	0	0	1.0000	0	0	
Base / Equiv*Amine SM / Equiv	0	0	0	1.0000	0	0	=-1.3037*Intercept-
Amine SM / Equiv*Amine SM / Equiv	0	0	0	1.0000	0	0	=0.3742*Intercept+

LogNormal Distribution Parameters	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Scale	0.072038	0.0100397	51.484979	<.0001*	0.0523605	0.0917154

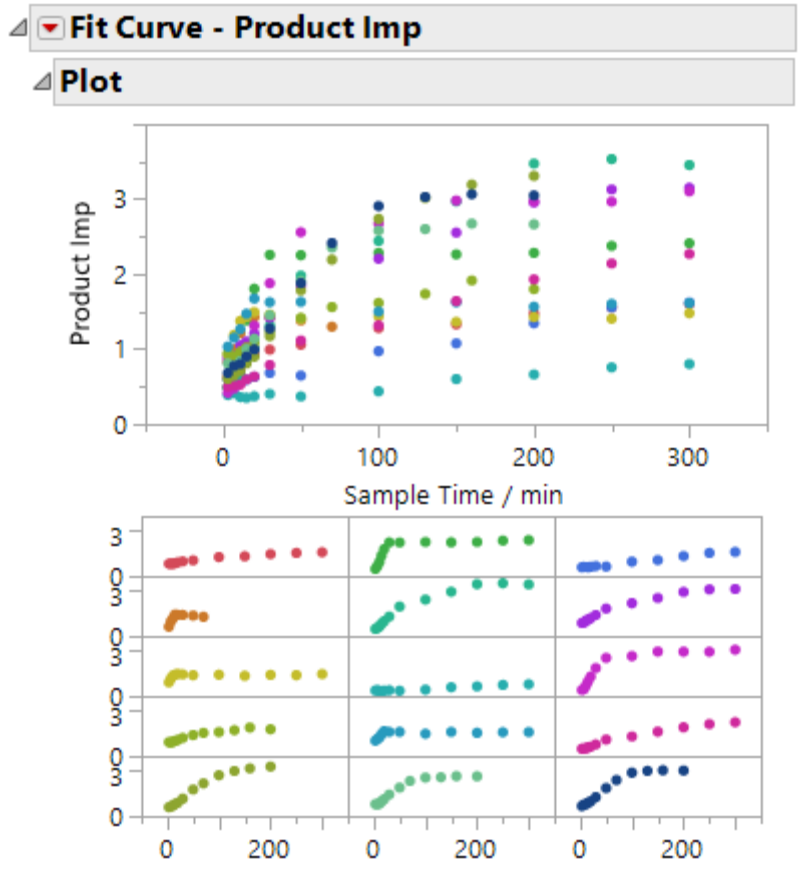
Three Parameter Logistic Fit



Product Impurity Model

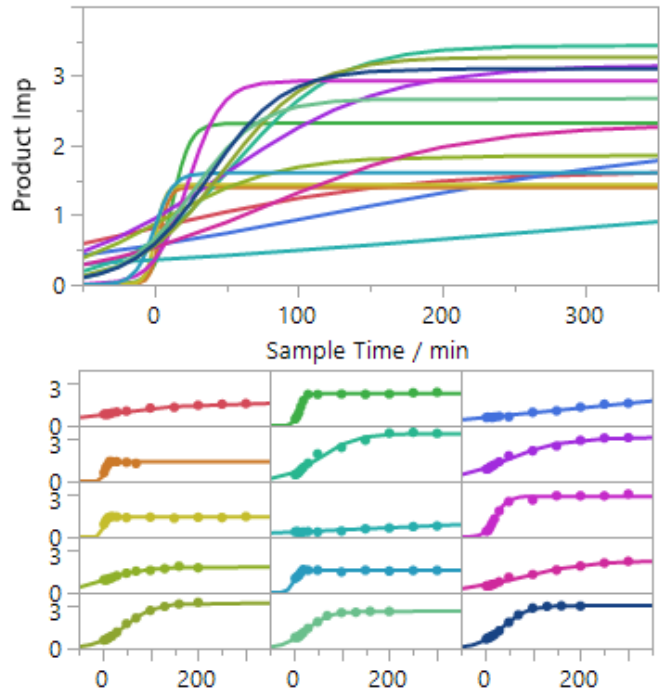


Product Impurity Model



Product Impurity Model

- Fit Curve - Product Imp
  - Model Comparison
  - Plot
  - Logistic 3P
    - Prediction Model
    - Summary of Fit
    - Parameter Estimates
    - Plot



Product Impurity Model

Curve DOE Analysis

Generalized Regression for Model Parameters

Generalized Regression for Growth Rate

Model Comparison

Model Launch

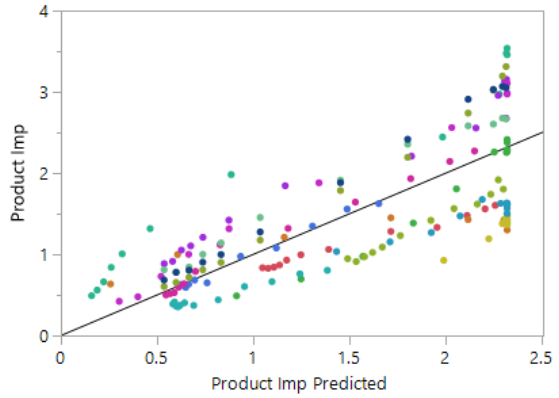
LogNormal Forward Selection with AICc Validation

Generalized Regression for Inflection Point

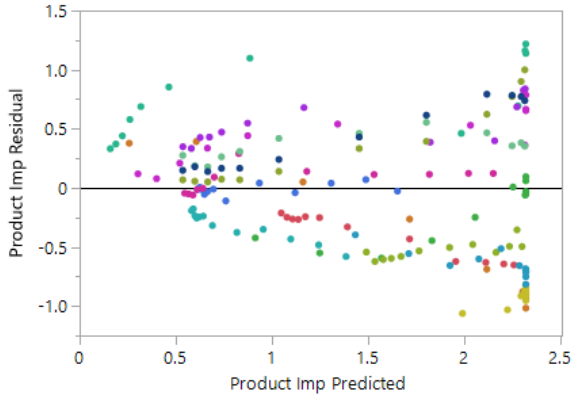
Generalized Regression for Asymptote

Diagnostic Plots

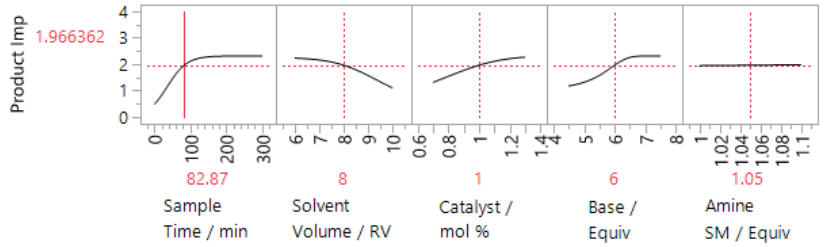
Actual by Predicted Plot



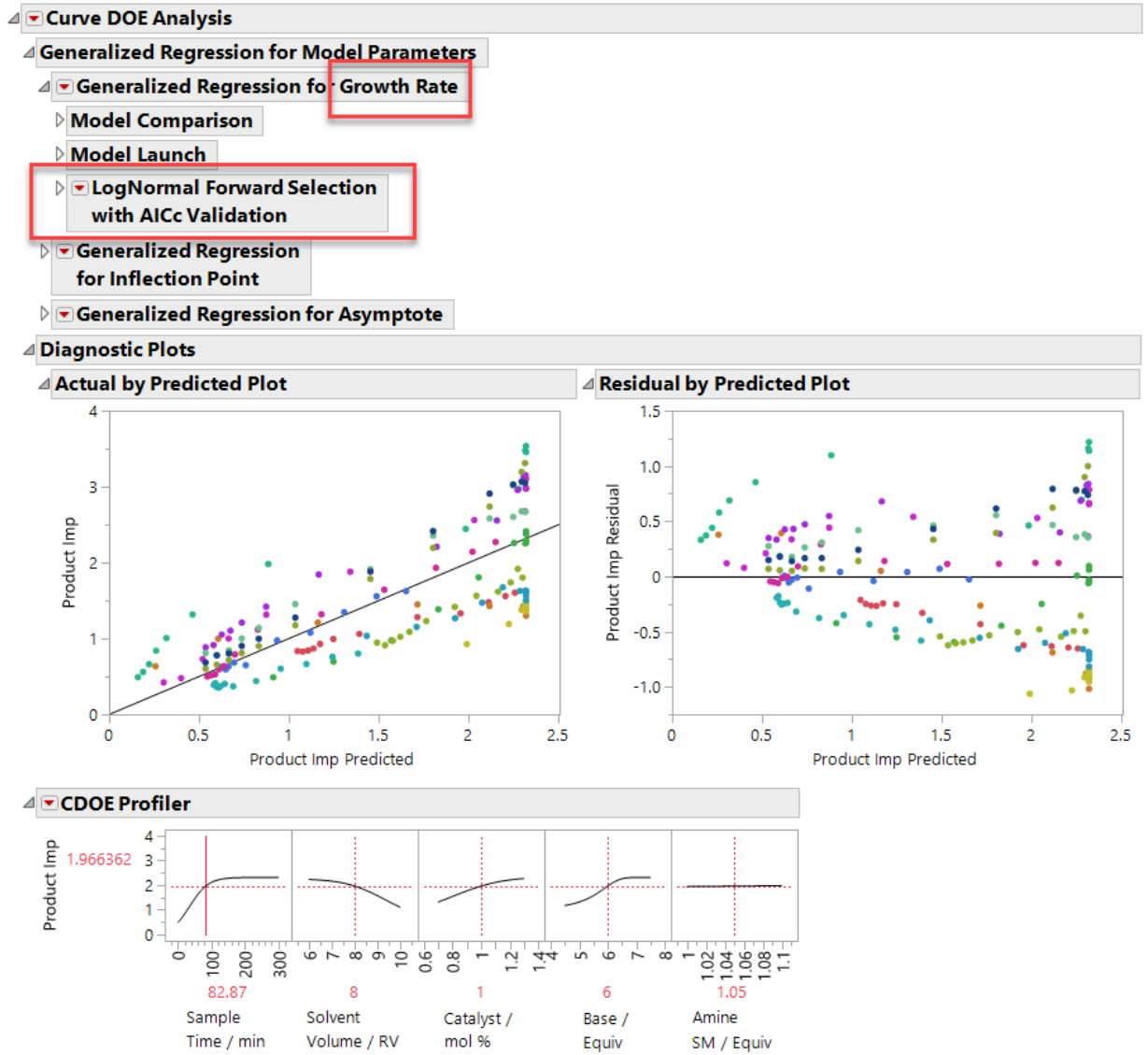
Residual by Predicted Plot



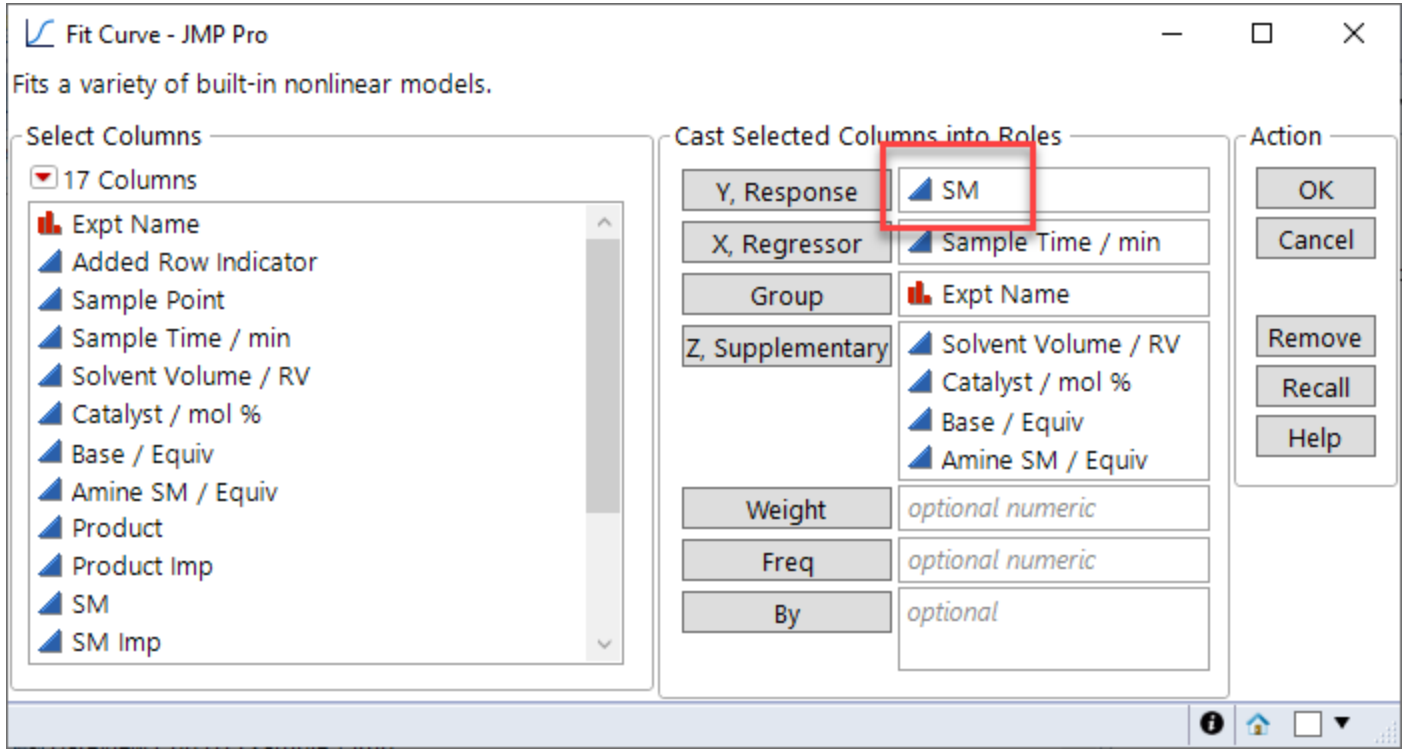
CDOE Profiler



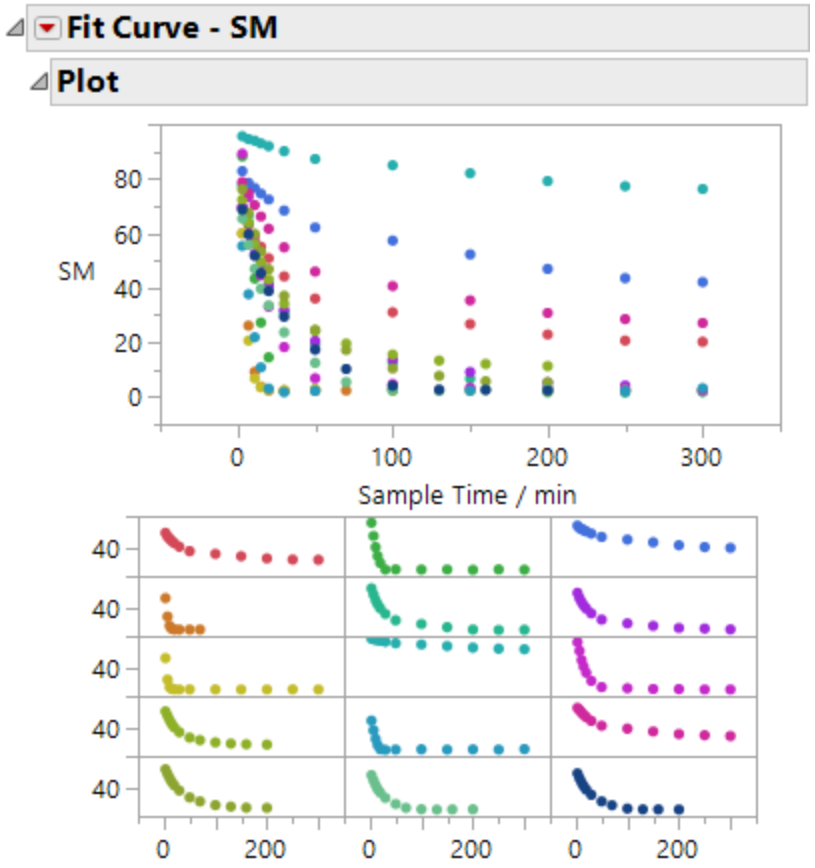
Product Impurity Model



Starting Material (SM) Model



Starting Material (SM) Model



Starting Material (SM) Model

Fit Curve - SM

Model Comparison

Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	781.18844	1		893.51132	427.51924	3.2635056	1.8065175	0.9971092
Logistic 4P	998.38685	6.854e-48		1125.4355	1027.7056	8.8595311	2.9764965	0.9930508
Exponential 2P	1012.3501	6.366e-51		1096.8573	2109.1502	14.446235	3.8008203	0.9857383
Logistic 3P	1532.4875	7.2e-164		1644.8104	30538.513	233.11842	15.268216	0.7935038

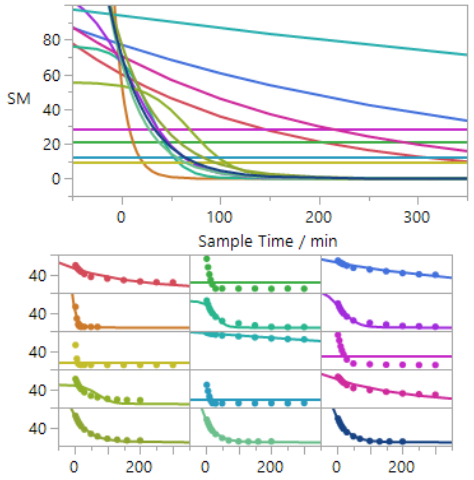
Logistic 3P

Prediction Model

Summary of Fit

Parameter Estimates

Plot



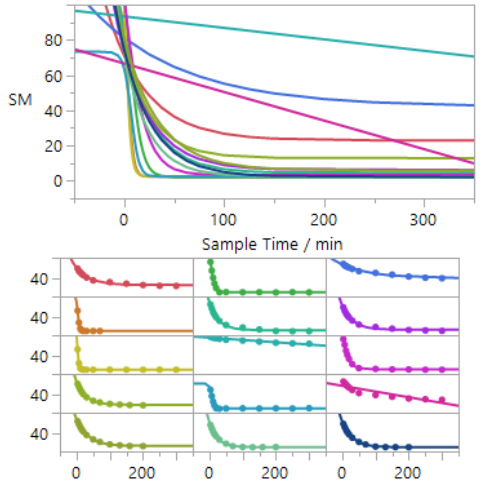
Logistic 4P

Prediction Model

Summary of Fit

Parameter Estimates

Plot



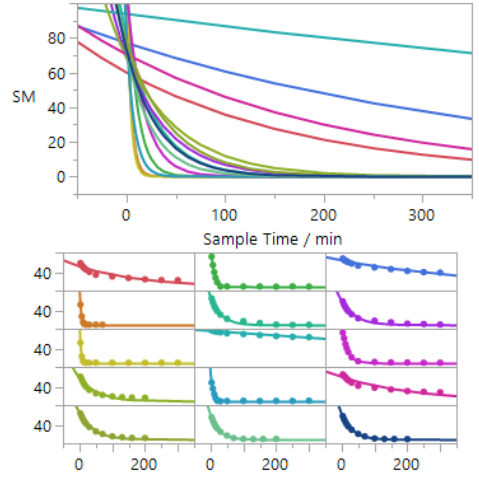
Exponential 2P

Prediction Model

Summary of Fit

Parameter Estimates

Plot



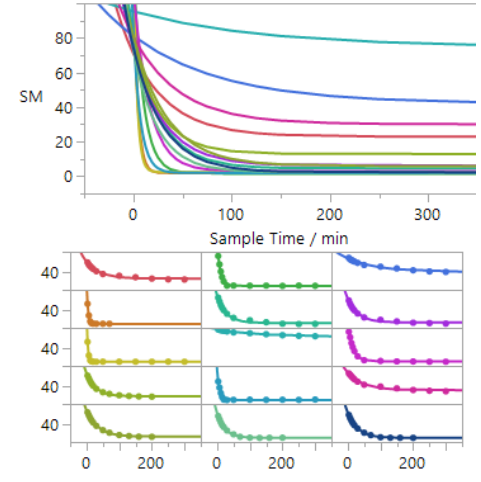
Exponential 3P

Prediction Model

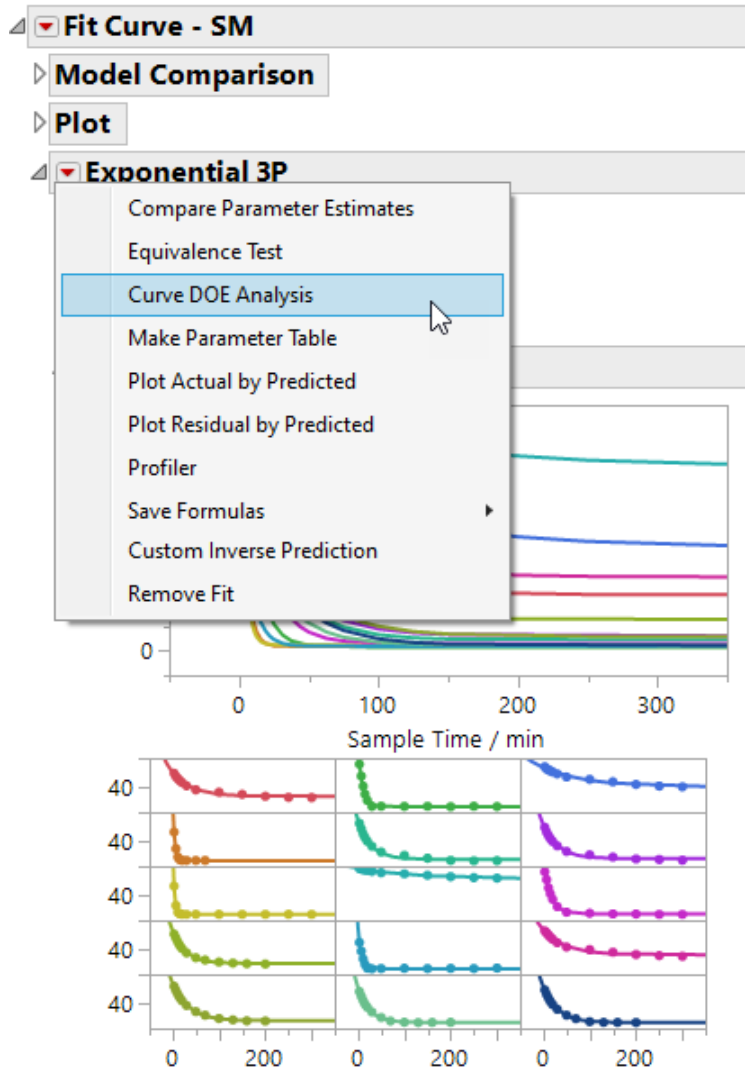
Summary of Fit

Parameter Estimates

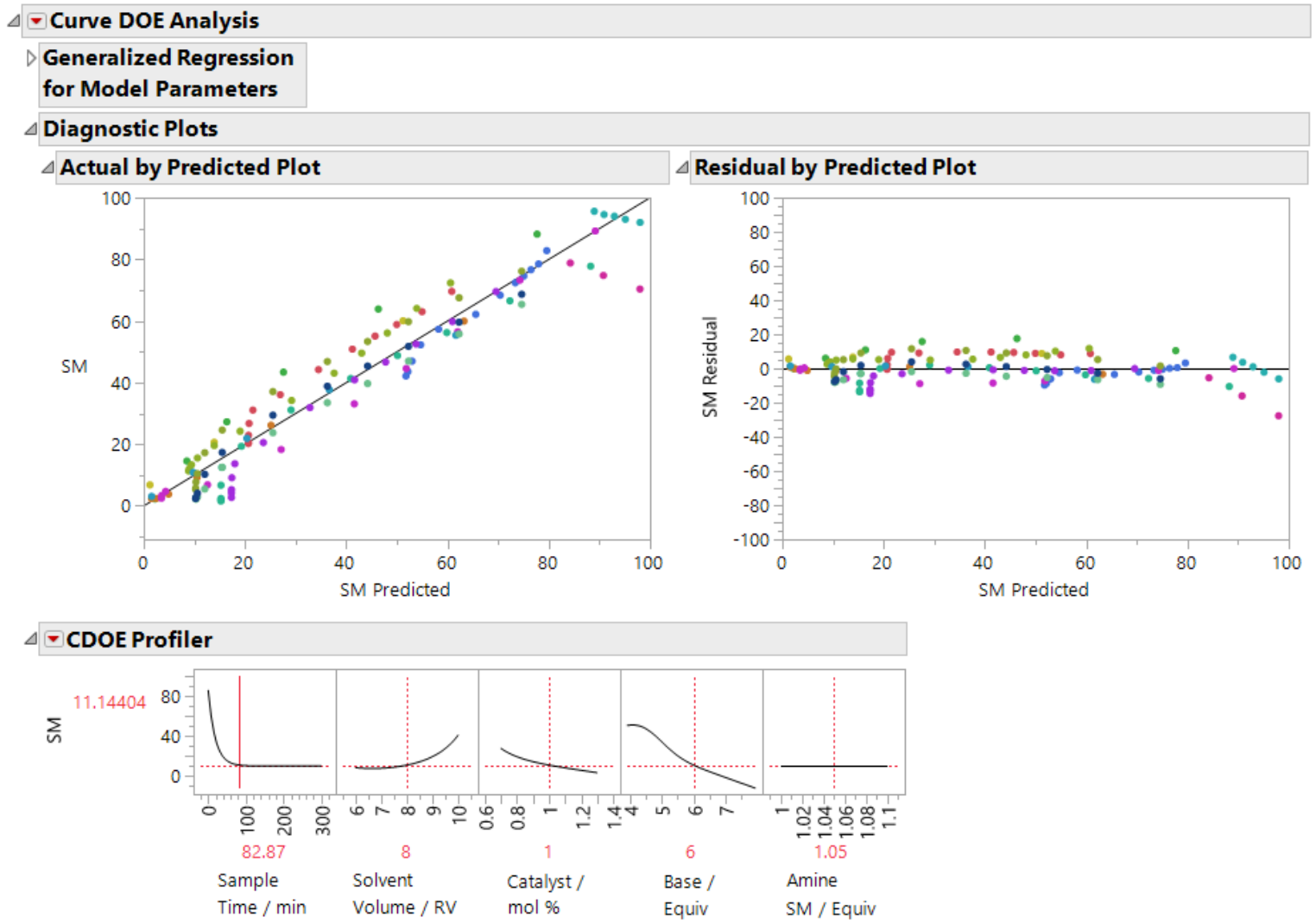
Plot



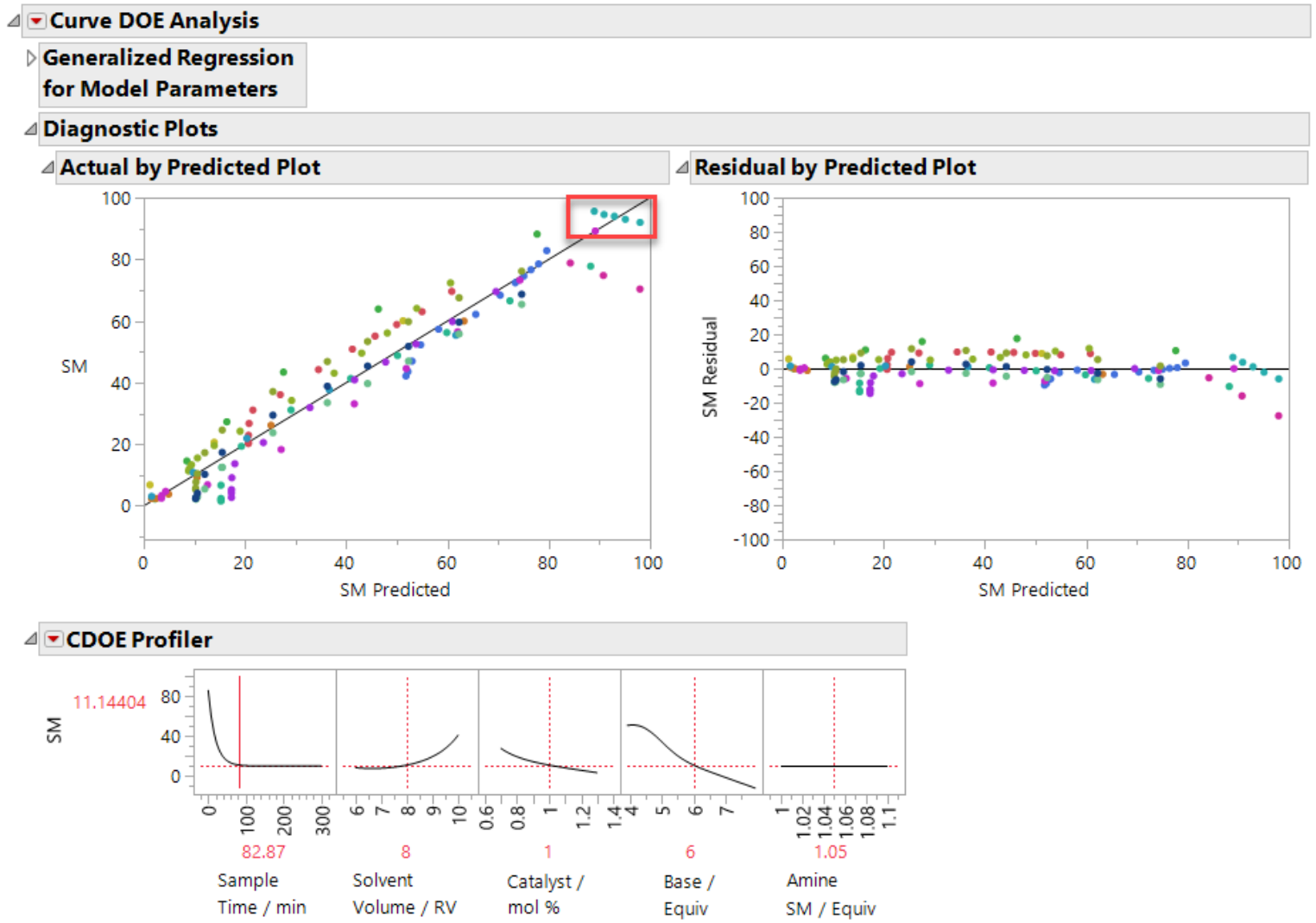
Starting Material (SM) Model



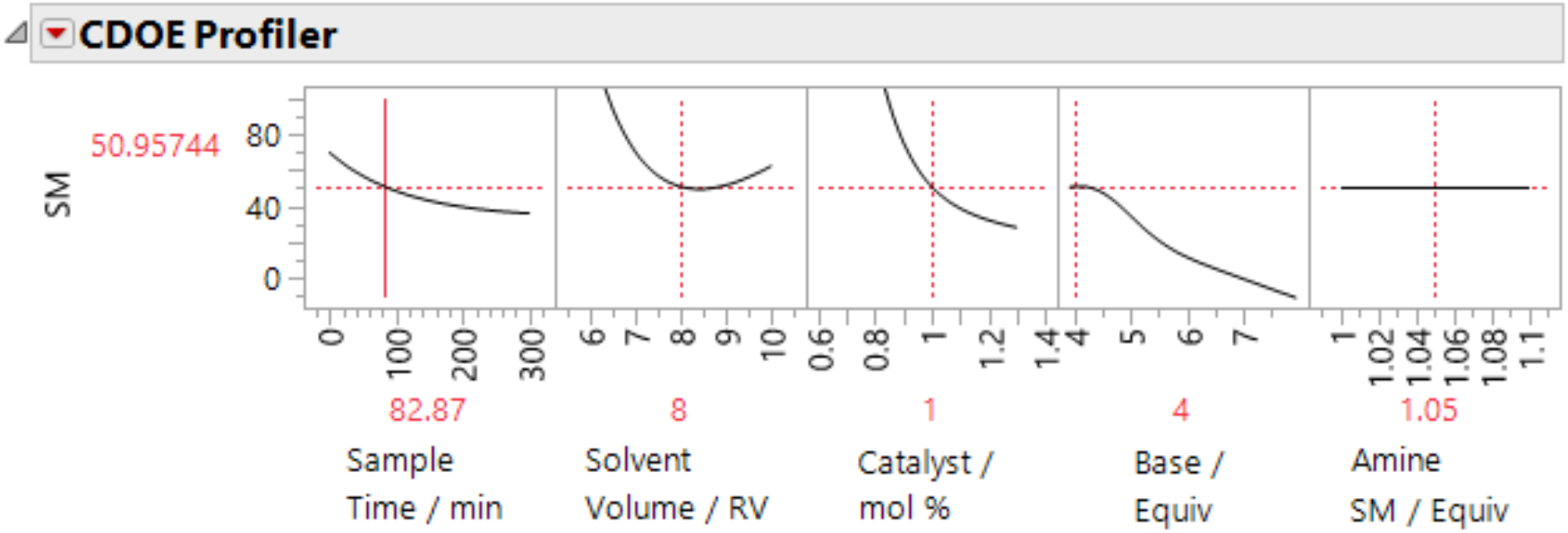
Starting Material (SM) Model



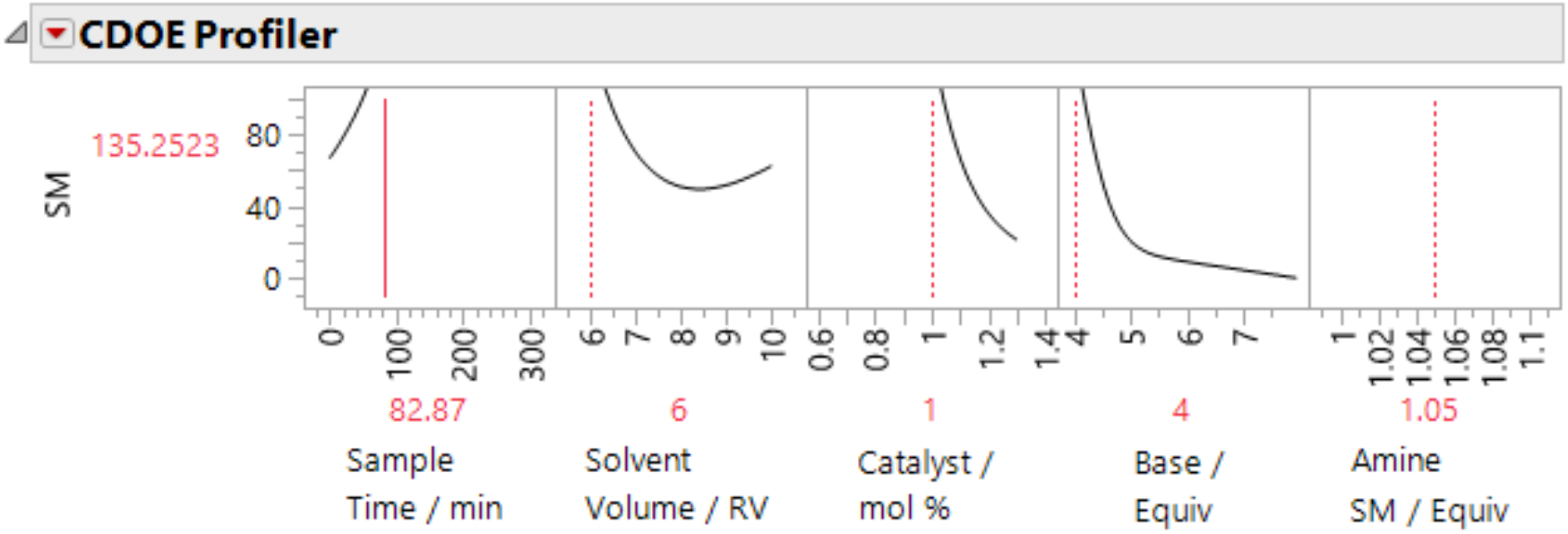
Starting Material (SM) Model



Starting Material (SM) Model



Starting Material (SM) Model



Starting Material (SM) Model

Exponential 3P

Prediction Model

$$a + b \cdot \mathrm{Exp}\left(c \cdot \text{Sample Time} / \text{min}\right)$$

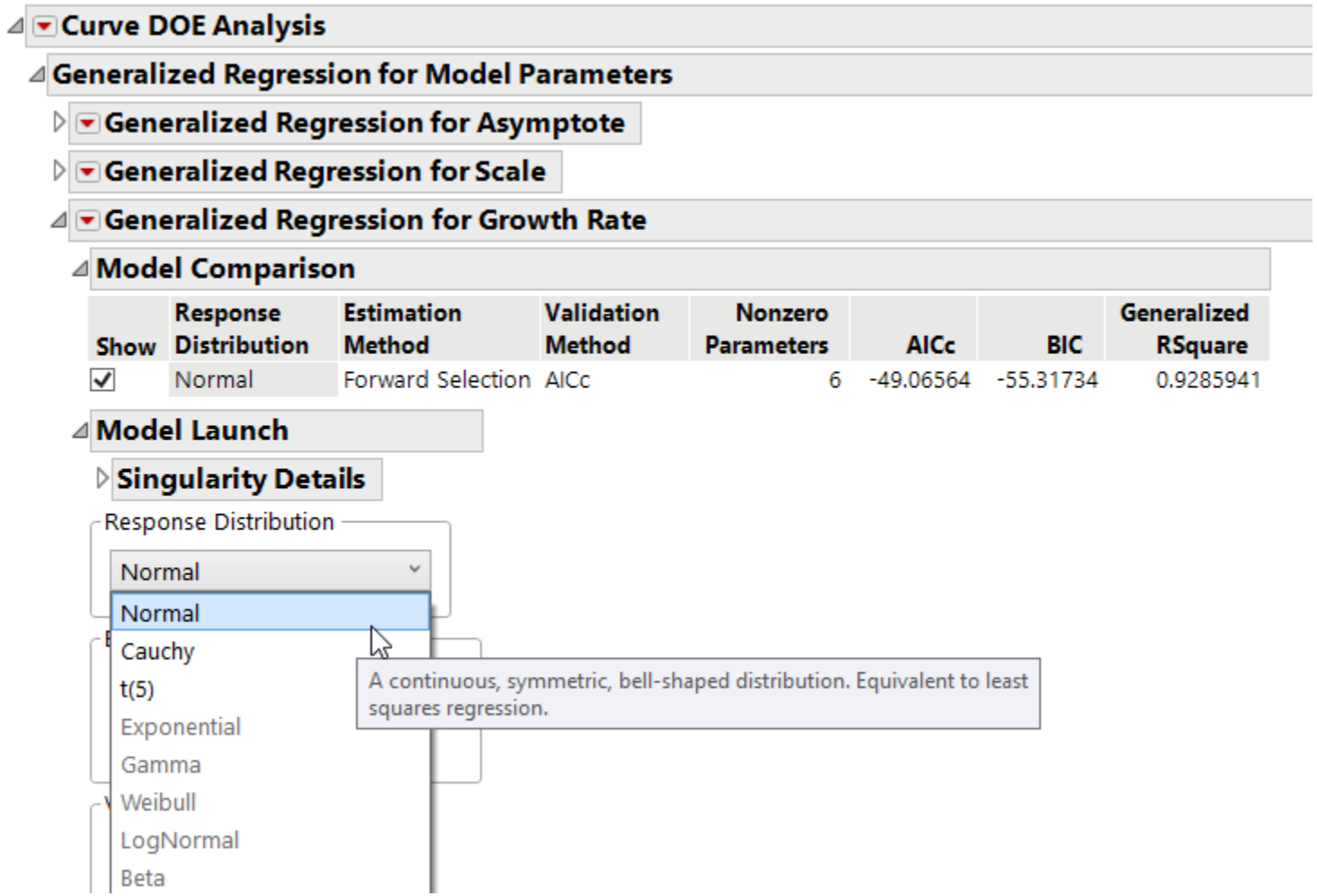
a = Asymptote

b = Scale

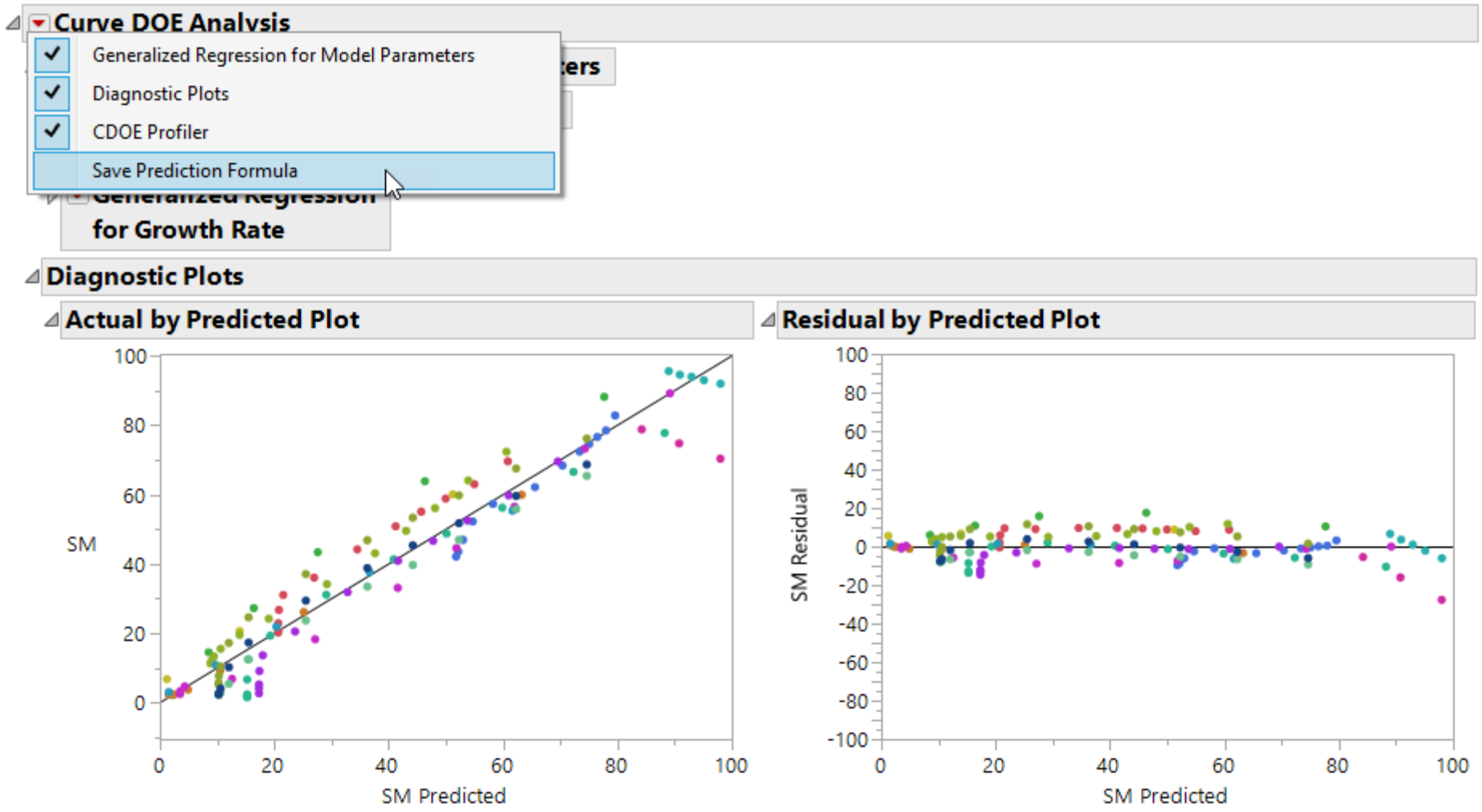
c = Growth Rate

	Group	Asymptote	Growth Rate		Scale
1	N1	22.867857654	-0.025230049	48	146326097
2	N2	1.4481425305	-0.101923433	12	.96878526
3	N3	41.996217734	-0.010872558	39	131356511
4	N4	1.7514216034	-0.236676633	11	.17315056
5	N5	4.5055740659	-0.036396107	79	418733351
6	N6	6.1012035335	-0.031638593	67	319873264
7	N7	2.4152494821	-0.29774875	1	1.1796014
8	N8	74.594130571	-0.007720537	20	725202973
9	N9	2.9957492718	-0.062309523	10	.65985095
10	N10	12.860706683	-0.036963998	65	615700289
11	N11	1.8686494638	-0.135956892	83	181872531
12	N12	30.116992898	-0.021342944	50	610337466
13	N13	5.8948190062	-0.028967266	75	115805735
14	N14	2.2783260949	-0.04095745	70	945997632
15	N15	2.2343745293	-0.03296952	72	208209911

Starting Material (SM) Model



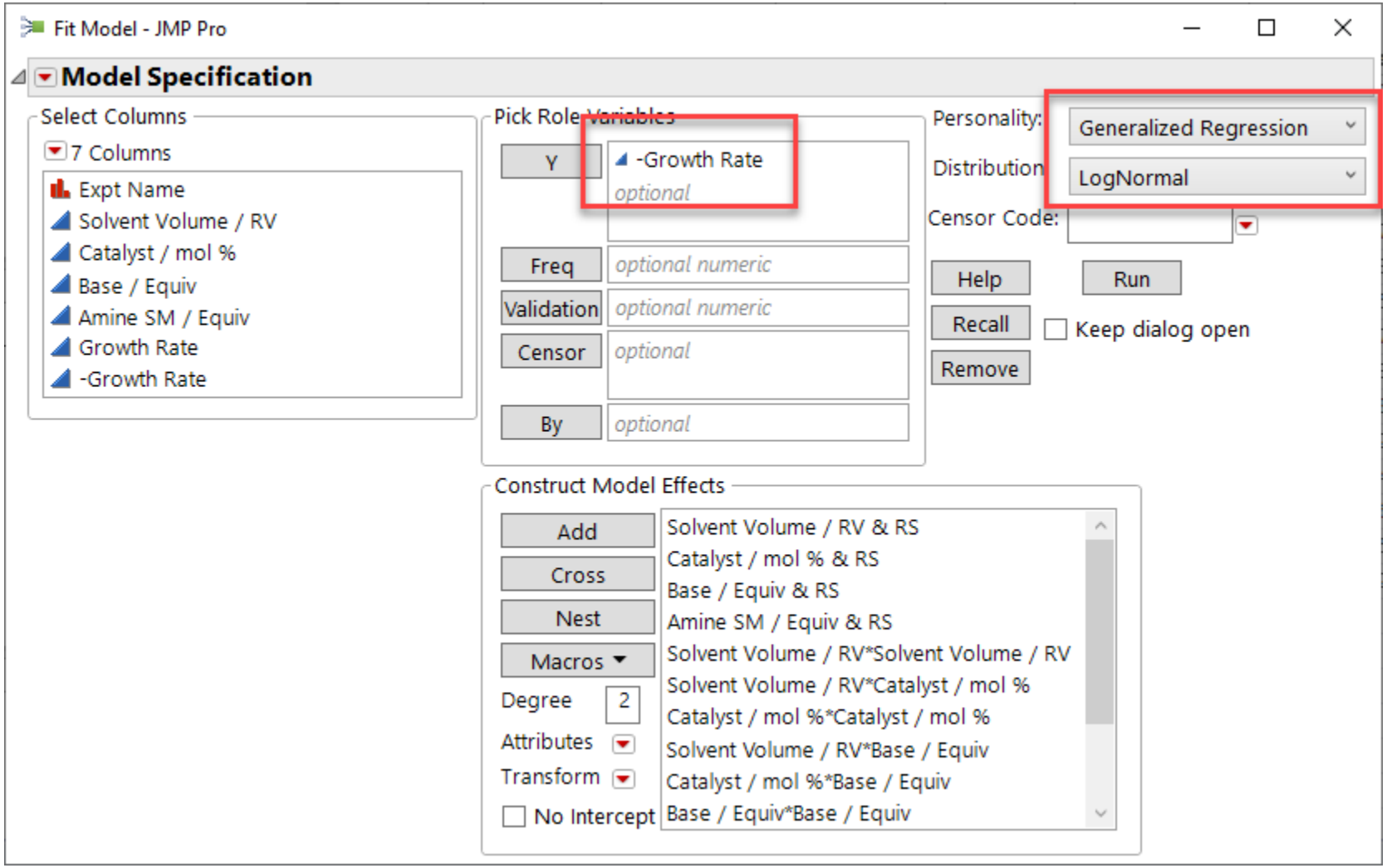
Starting Material (SM) Model



Starting Material (SM) Model

$$
\begin{aligned}
&4.8054227983 \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot 1.5113228449 \right) \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Base / Equiv} \cdot -3.380524086 \right) \\
&+ -22.51564084 \cdot \textit{Catalyst / mol \%} \\
&+ 15.580413139 \cdot \textit{Base / Equiv} \\
&+ \left(\begin{aligned} &34.165675219 \\ &+ \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot -0.437773887 \right) \\ &+ \textit{Base / Equiv} \cdot \left(\textit{Base / Equiv} \cdot 1.9283370001 \right) \end{aligned} \right) \\
&\quad \cdot \operatorname{Exp}\left(\left(\begin{aligned} &0.3868444112 \\ &+ -0.067025251 \cdot \textit{Solvent Volume / RV} \\ &+ \textit{Solvent Volume / RV} \cdot \left(\textit{Base / Equiv} \cdot 0.0143427946 \right) \\ &+ -0.099281912 \cdot \textit{Catalyst / mol \%} \\ &+ \textit{Base / Equiv} \cdot \left(\textit{Base / Equiv} \cdot -0.013697589 \right) \end{aligned} \right) \cdot \textit{Sample Time / min} \right)
\end{aligned}
$$

Starting Material (SM) Model



Starting Material (SM) Model

Generalized Regression for -Growth Rate

Model Comparison

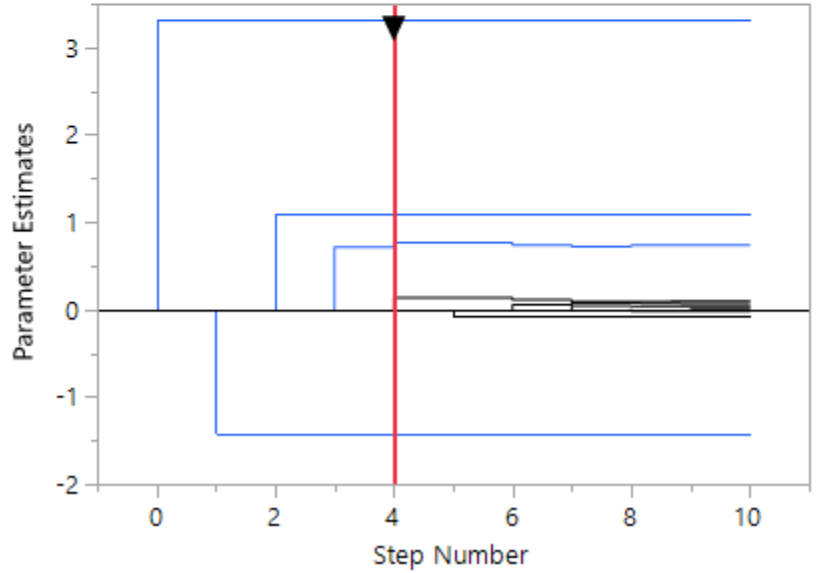
Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare
☑	LogNormal	Forward Selection	AICc	6	-105.8128	-112.0645	0.9941246
☑	Normal	Forward Selection	AICc	8	-63.39106	-81.72666	0.9914435

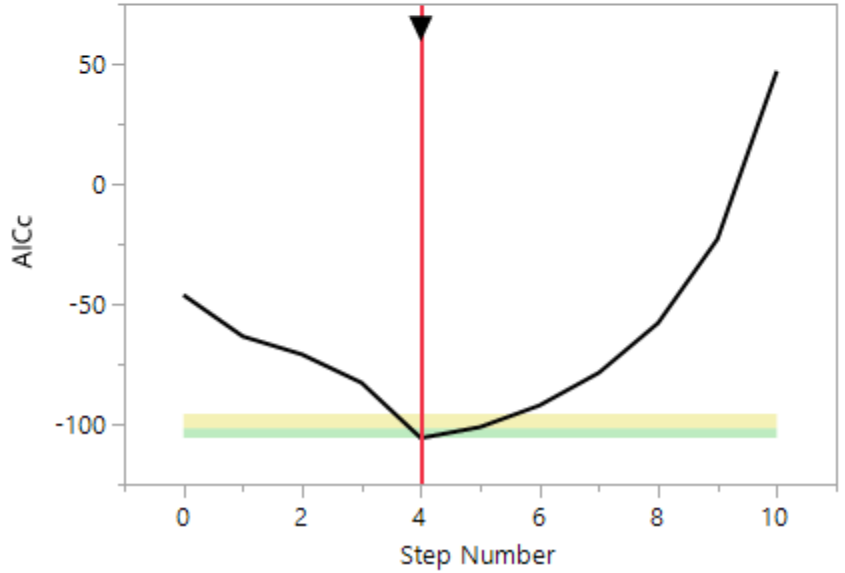
Model Launch

LogNormal Forward Selection with AICc Validation

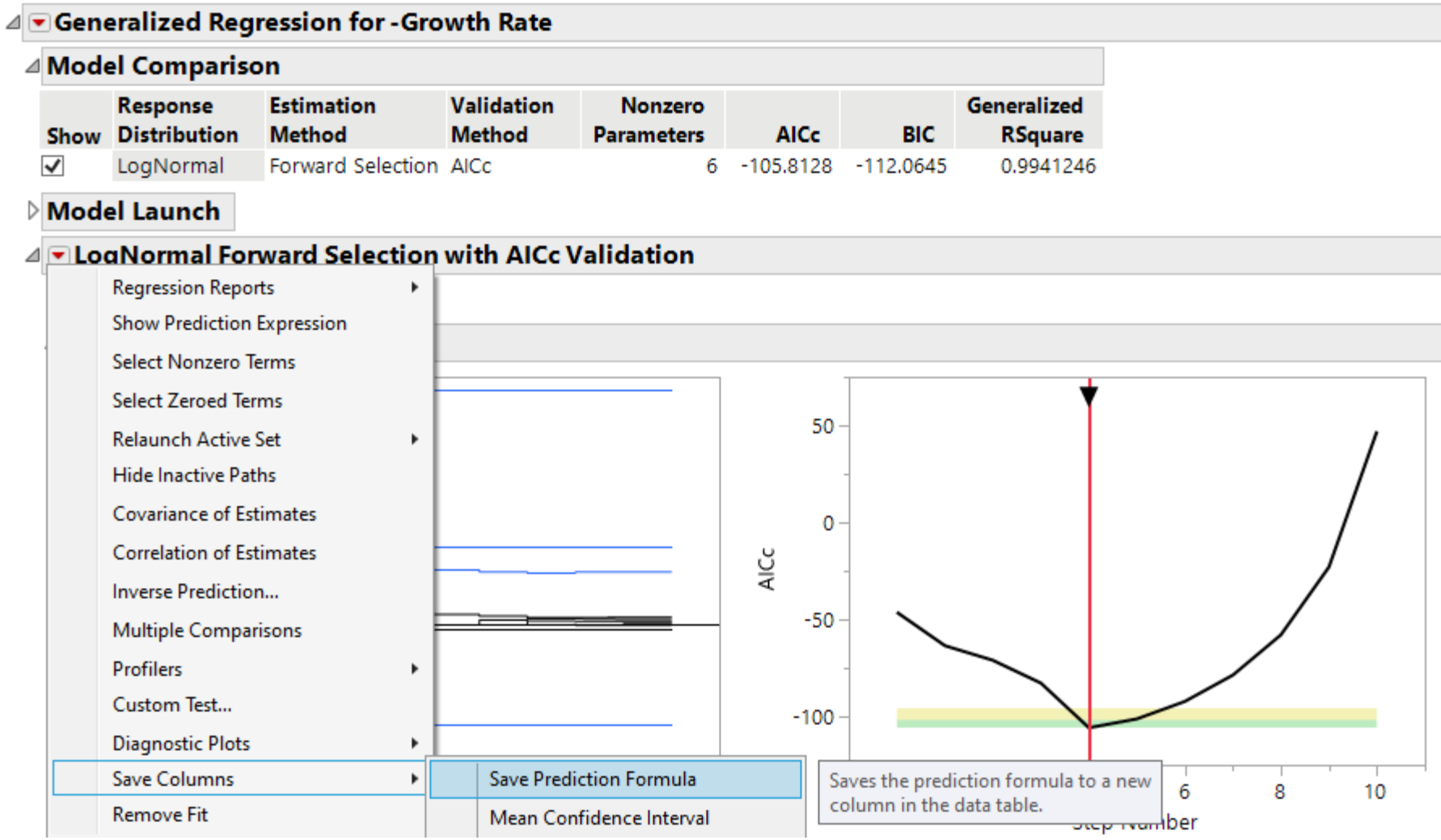
Model Summary

Solution Path





Starting Material (SM) Model



Starting Material (SM) Model

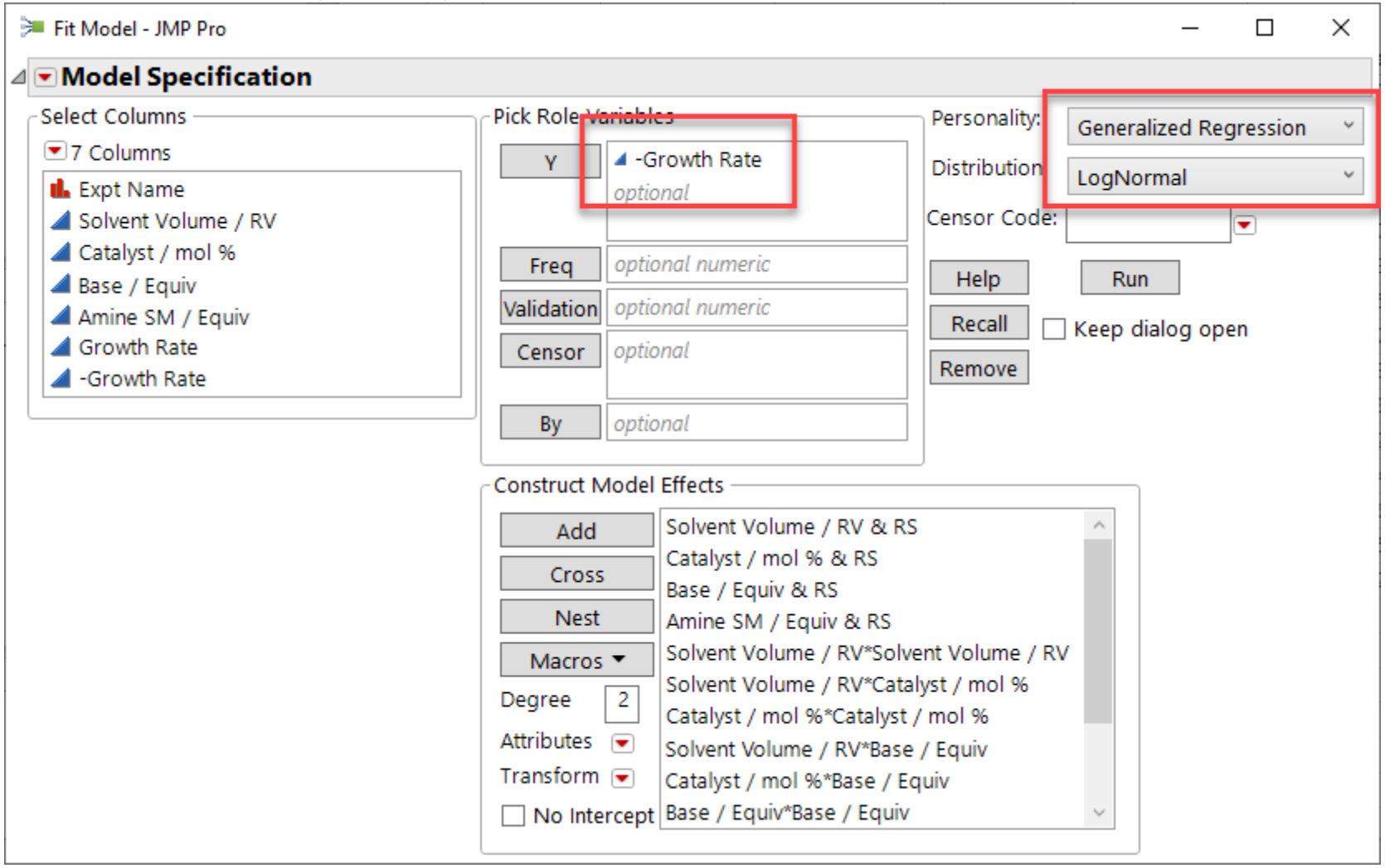
Parameter Estimates

Parameter	Group	Estimate	Std Error	Wald ChiSquare	Prob > ChiSquare	Lower 95%	Upper 95%
Asymptote	N1	22.867858	0.9429534	588.12622	<.0001*	21.019703	24.716012
Scale	N1	48.146326	1.6			4.966198	51.326454
Growth Rate	N1	-0.02523	0.0			0.029876	-0.020584
Asymptote	N2	1.4481425	0.7			.0294338	2.8668512
Scale	N2	120.96879	3.2			14.67599	127.26158
Growth Rate	N2	-0.101923	0.0				
Asymptote	N3	41.996218	1.8			8	
Scale	N3	39.131357	1.7			5.737057	42.525656
Growth Rate	N3	-0.010873	0.0			0.014141	-0.007604
Asymptote	N4	1.7514216	0.9			0.033233	3.5360763
Scale	N4	119.17315	8.0			03.45358	134.89272
Growth Rate	N4	-0.236677	0.0			0.273948	-0.199405
Asymptote	N5	4.5055741	0.8			.8453776	6.1657706
Scale	N5	79.418733	1.8			5.741327	83.09614
Growth Rate	N5	-0.036396	0.0			0.040312	-0.03248
Asymptote	N6	6.1012035	0.8			.3845257	7.8178814
Scale	N6	67.319873	1.			63.84646	70.793286
Growth Rate	N6	-0.031639	0.0020688	233.87727	<.0001*	-0.035693	-0.027584

Starting Material (SM) Model

	Expt Name	Solvent Volume / RV	Catalyst / mol %	Base / Equiv	Amine SM / Equiv	Growth Rate	-Growth Rate
1	N1	8	1.3	4.5	1	-0.025230049	0.0252300488
2	N2	8	0.7	7.5	1.1	-0.101923433	0.1019234335
3	N3	10	1	4.5	1.1	-0.010872558	0.010872558
4	N4	6	1	7.5	1	-0.236676633	0.2366766326
5	N5	6	0.7	6	1.1	-0.036396107	0.0363961067
6	N6	10	1.3	6	1	-0.031638593	0.0316385934
7	N7	6	1.3	7.5	1.05	-0.29774875	0.2977487499
8	N8	10	0.7	4.5	1.05	-0.007720537	0.0077205372
9	N9	10	0.7	7.5	1	-0.062309523	0.0623095232
10	N10	6	1.3	4.5	1.1	-0.036963998	0.0369639976
11	N11	10	1.3	7.5	1.1	-0.135956892	0.1359568915
12	N12	6	0.7	4.5	1	-0.021342944	0.0213429442
13	N13	8	1	6	1.05	-0.028967266	0.0289672656
14	N14	8	1	6	1.05	-0.04095745	0.0409574504
15	N15	8	1	6	1.05	-0.03296952	0.0329695197

Starting Material (SM) Model



Starting Material (SM) Model

Generalized Regression for -Growth Rate

Model Comparison

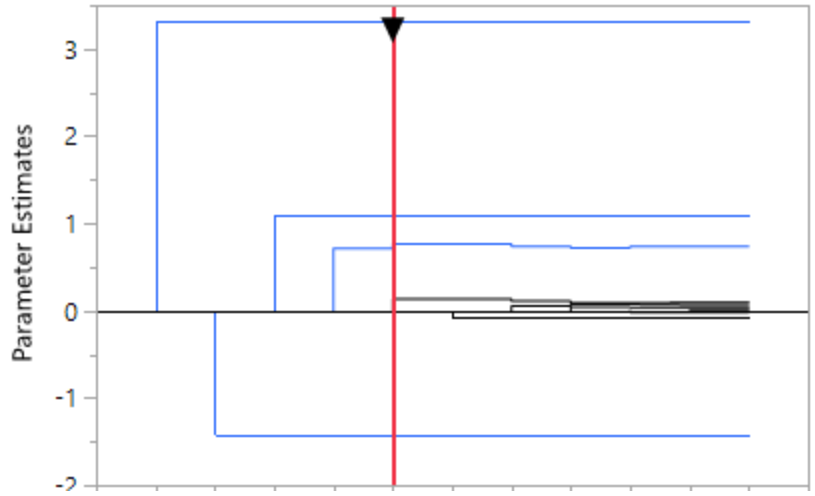
Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare
☑	LogNormal	Forward Selection	AICc	6	-105.8128	-112.0645	0.9941246
☑	Normal	Forward Selection	AICc	8	-63.39106	-81.72666	0.9914435

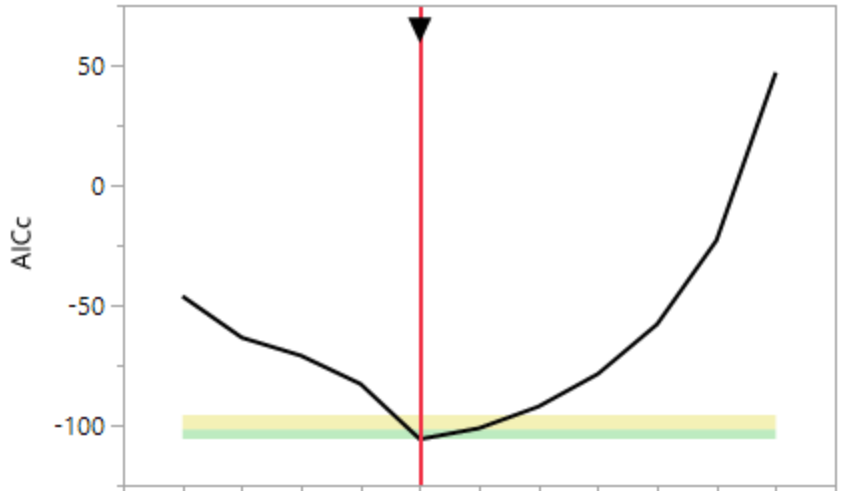
Model Launch

LogNormal Forward Selection with AICc Validation

Model Summary

Solution Path

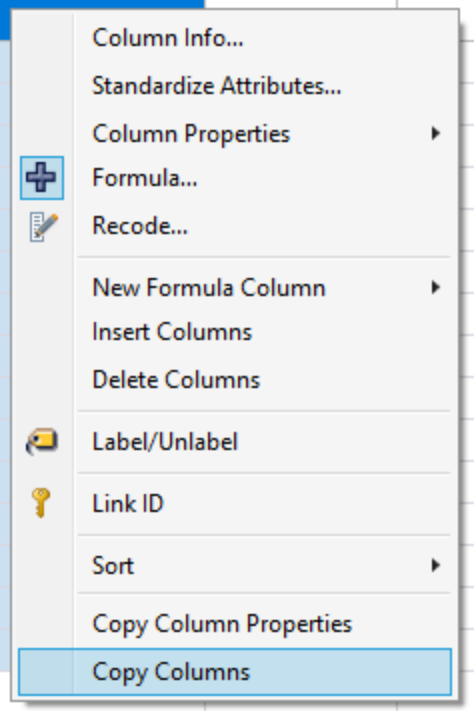




Starting Material (SM) Model

Starting Material (SM) Model

	Expt Name	Solvent Volume / RV	Catalyst / mol %	Base / Equiv	Amine SM / Equiv	Growth Rate	-Growth Rate	-Growth Rate Prediction
1	N1	8	1.3	4.5	1	-0.025...	0.0252...	
2	N2	8	0.7	7.5	1.1	-0.101...	0.1019...	
3	N3	10	1	4.5	1.1	-0.010...	0.0108...	
4	N4	6	1	7.5	1	-0.236...	0.2366...	
5	N5	6	0.7	6	1.1	-0.036...	0.0363...	
6	N6	10	1.3	6	1	-0.031...	0.0316...	
7	N7	6	1.3	7.5	1.05	-0.297...	0.2977...	
8	N8	10	0.7	4.5	1.05	-0.007...	0.0077...	
9	N9	10	0.7	7.5	1	-0.062...	0.0623...	
10	N10	6	1.3	4.5	1.1	-0.036...	0.0369...	
11	N11	10	1.3	7.5	1.1	-0.135...	0.1359...	
12	N12	6	0.7	4.5	1	-0.021...	0.0213...	
13	N13	8	1	6	1.05	-0.028...	0.0289...	
14	N14	8	1	6	1.05	-0.040...	0.0409...	
15	N15	8	1	6	1.05	-0.032...	0.0329...	



Starting Material (SM) Model

	Expt Name	Sample Time / min	Solvent Volume / RV	Catalyst / mol %	Base / Equiv	Amine SM / Equiv	Product	Product Imp	SM	Product Predictor	Product Imp Predictor	SM Predictor	-Growth Rate Prediction Formula
1	N1	3	8	1.3	4.5	1	29.37459...	0.835080...	69.6229...	29.5252...	0.78054...	60.823894854	0.0248591549
2	N1	7	8	1.3	4.5	1	35.85010...	0.828343...	63.0270...	32.5001...	0.80048...	54.961328808	0.0248591549
3	N1	11	8	1.3	4.5	1	39.81607...	0.842789...	58.8567...	35.5291...	0.82043...	49.954770117	0.0248591549
4	N1	15	8	1.3	4.5	1	43.43785...	0.869276...	55.1016...	38.5707...	0.84035...	45.679231092	0.0248591549
5	N1	20	8	1.3	4.5	1	47.53072...	0.929187...	50.8791...	42.3264...	0.86518...	41.201713038	0.0248591549
6	N1	30	8	1.3	4.5	1	53.96907...	0.995378...	44.2265...	49.3969...	0.91445...	34.50828182	0.0248591549
7	N1	50	8	1.3	4.5	1	61.91892...	1.061143...	36.0956...	60.4207...	1.01012...	26.956999806	0.0248591549
8	N1	100	8	1.3	4.5	1	66.73583...	1.282049...	31.0905...	71.0601...	1.21940...	21.546516392	0.0248591549
9	N1	150	8	1.3	4.5	1	70.97244...	1.332256...	26.7644...	72.6443...	1.37219...	20.794234497	0.0248591549
10	N1	200	8	1.3	4.5	1	74.55535...	1.479735...	22.9344...	72.8455...	1.47125...	20.689636071	0.0248591549
11	N1	250	8	1.3	4.5	1	76.61089...	1.559971...	20.6754...	72.8704...	1.53063...	20.675092546	0.0248591549
12	N1	300	8	1.3	4.5	1	77.02376...	1.604085...	20.2519...	72.8735...	1.56456...	20.673070392	0.0248591549

Starting Material (SM) Model

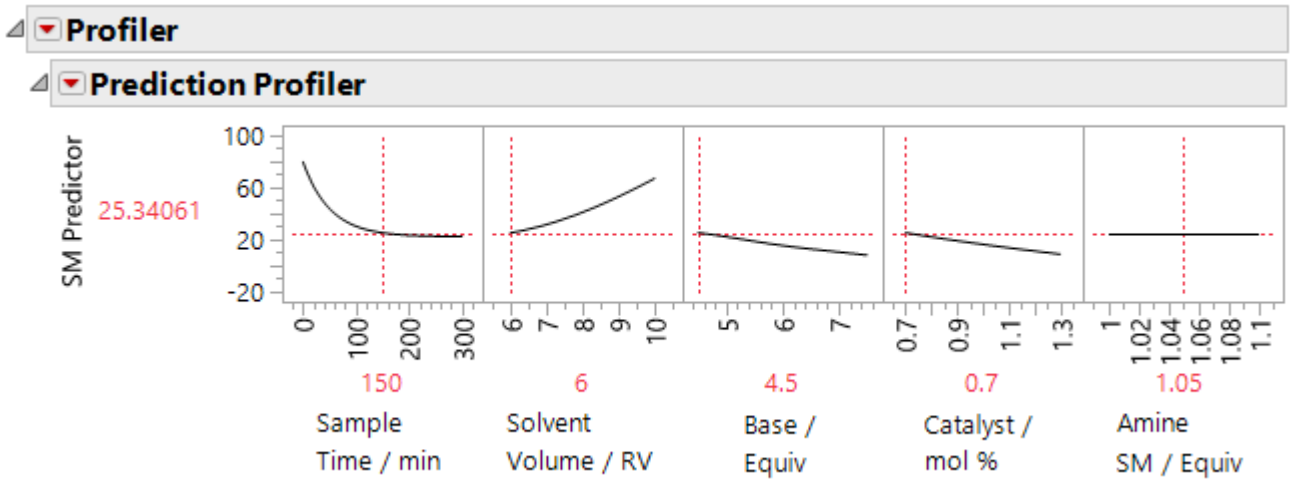
	Expt Name	Sample Time / min	Solvent Volume / RV	Catalyst / mol %	Base / Equiv	Amine SM / Equiv	Product	Product Imp	SM	Product Predictor	Product Imp Predictor	SM Predictor	-Growth Rate Prediction Formula
1	N1	3	8	1.3	4.5	1	29.37459...	0.835080...	69.6229...	29.5252...	0.78054...	60.823894854	0.0248591549
2	N1	7	8	1.3	4.5	1	35.85010...	0.828343...	63.0270...	32.5001...	0.80048...	54.961328808	0.0248591549
3	N1	11	8	1.3	4.5	1	39.81607...	0.842789...	58.8567...	35.5291...	0.82043...	49.954770117	0.0248591549
4	N1	15	8	1.3	4.5	1	43.43785...	0.869276...	55.1016...	38.5707...	0.84035...	45.679231092	0.0248591549
5	N1	20	8	1.3	4.5	1	47.53072...	0.929187...	50.8791...	42.3264...	0.86518...	41.201713038	0.0248591549
6	N1	30	8	1.3	4.5	1	53.96907...	0.995378...	44.2265...	49.3969...	0.91445...	34.50828182	0.0248591549
7	N1	50	8	1.3	4.5	1	61.91892...	1.061143...	36.0956...	60.4207...	1.01012...	26.956999806	0.0248591549
8	N1	100	8	1.3	4.5	1	66.73583...	1.282049...	31.0905...	71.0601...	1.21940...	21.546516392	0.0248591549
9	N1	150	8	1.3	4.5	1	70.97244...	1.332256...	26.7644...	72.6443...	1.37219...	20.794234497	0.0248591549
10	N1	200	8	1.3	4.5	1	74.55535...	1.479735...	22.9344...	72.8455...	1.47125...	20.689636071	0.0248591549
11	N1	250	8	1.3	4.5	1	76.61089...	1.559971...	20.6754...	72.8704...	1.53063...	20.675092546	0.0248591549
12	N1	300	8	1.3	4.5	1	77.02376...	1.604085...	20.2519...	72.8735...	1.56456...	20.673070392	0.0248591549

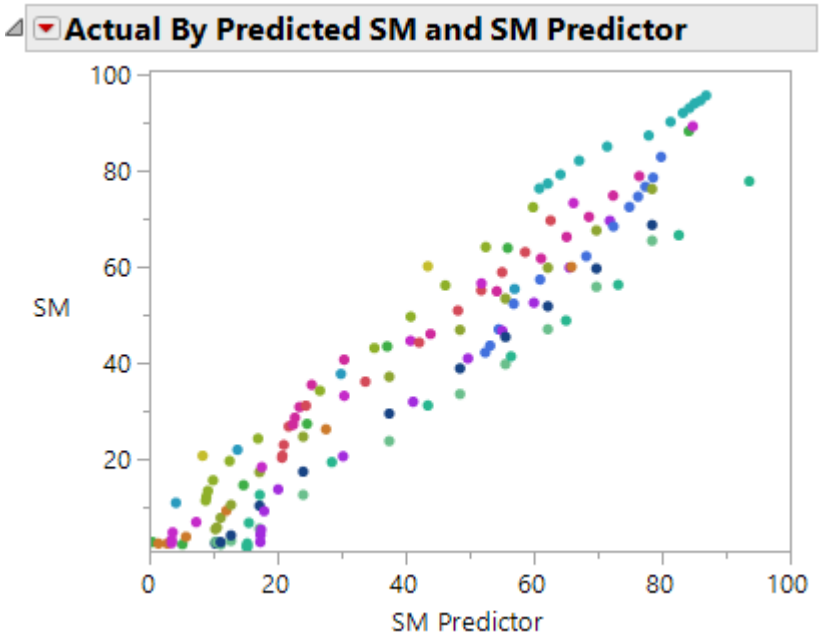
Starting Material (SM) Model

$$
\begin{aligned}
&4.8054227983 \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot 1.5113228449 \right) \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Base / Equiv} \cdot -3.380524086 \right) \\
&+ -22.51564084 \cdot \textit{Catalyst / mol \%} \\
&+ 15.580413139 \cdot \textit{Base / Equiv} \\
&+ \left(34.165675219 + \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot -0.437773887 \right) + \textit{Base / Equiv} \cdot \left(\textit{Base / Equiv} \cdot 1.9283370001 \right) \right) \\
&+ \left(\textit{Sample Time / min} \cdot \text{Exp} \left(0.3868444112 + -0.067025251 \cdot \textit{Solvent Volume / RV} + \textit{Solvent Volume / RV} \cdot \left(\textit{Base / Equiv} \cdot 0.0143427946 \right) + -0.099281912 \cdot \textit{Catalyst / mol \%} + \textit{Base / Equiv} \cdot \left(\textit{Base / Equiv} \cdot -0.013697589 \right) \right) \right)
\end{aligned}
$$

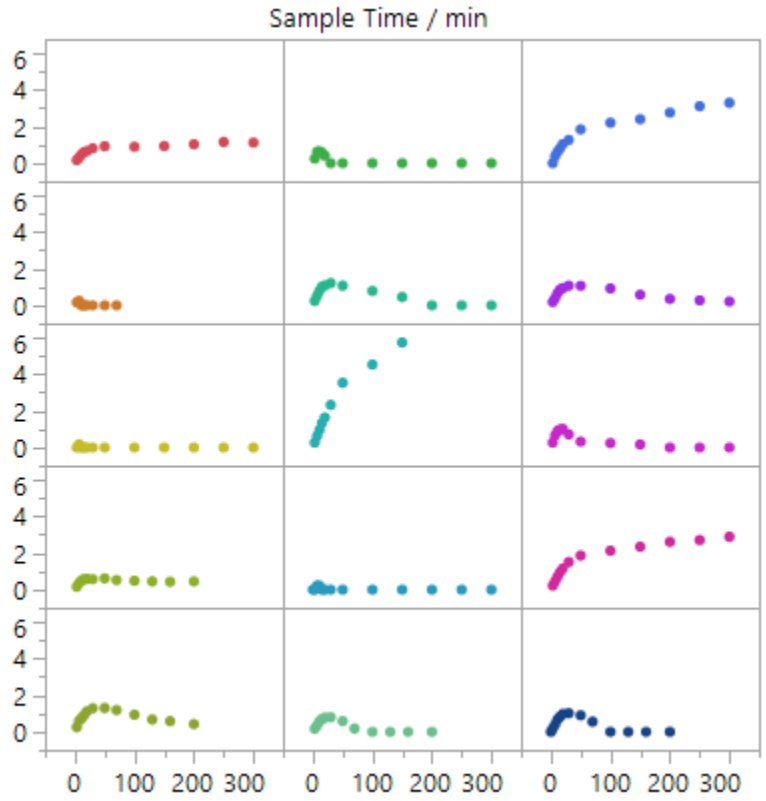
$$
\begin{aligned}
&4.8054227983 \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot 1.5113228449 \right) \\
&+ \textit{Solvent Volume / RV} \cdot \left(\textit{Base / Equiv} \cdot -3.380524086 \right) \\
&+ -22.51564084 \cdot \textit{Catalyst / mol \%} \\
&+ 15.580413139 \cdot \textit{Base / Equiv} \\
&+ \left(34.165675219 + \textit{Solvent Volume / RV} \cdot \left(\textit{Solvent Volume / RV} \cdot -0.437773887 \right) + \textit{Base / Equiv} \cdot \left(\textit{Base / Equiv} \cdot 1.9283370001 \right) \right) \\
&\quad \cdot \text{Exp} \left(- \textit{-Growth Rate Prediction Formula} \cdot \textit{Sample Time / min} \right)
\end{aligned}
$$

Starting Material (SM) Model

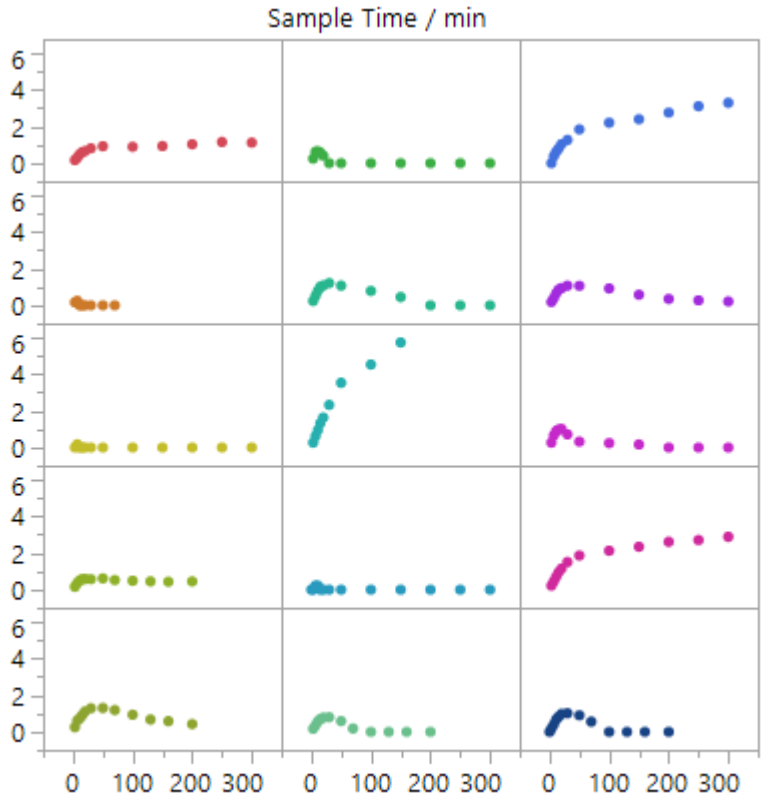


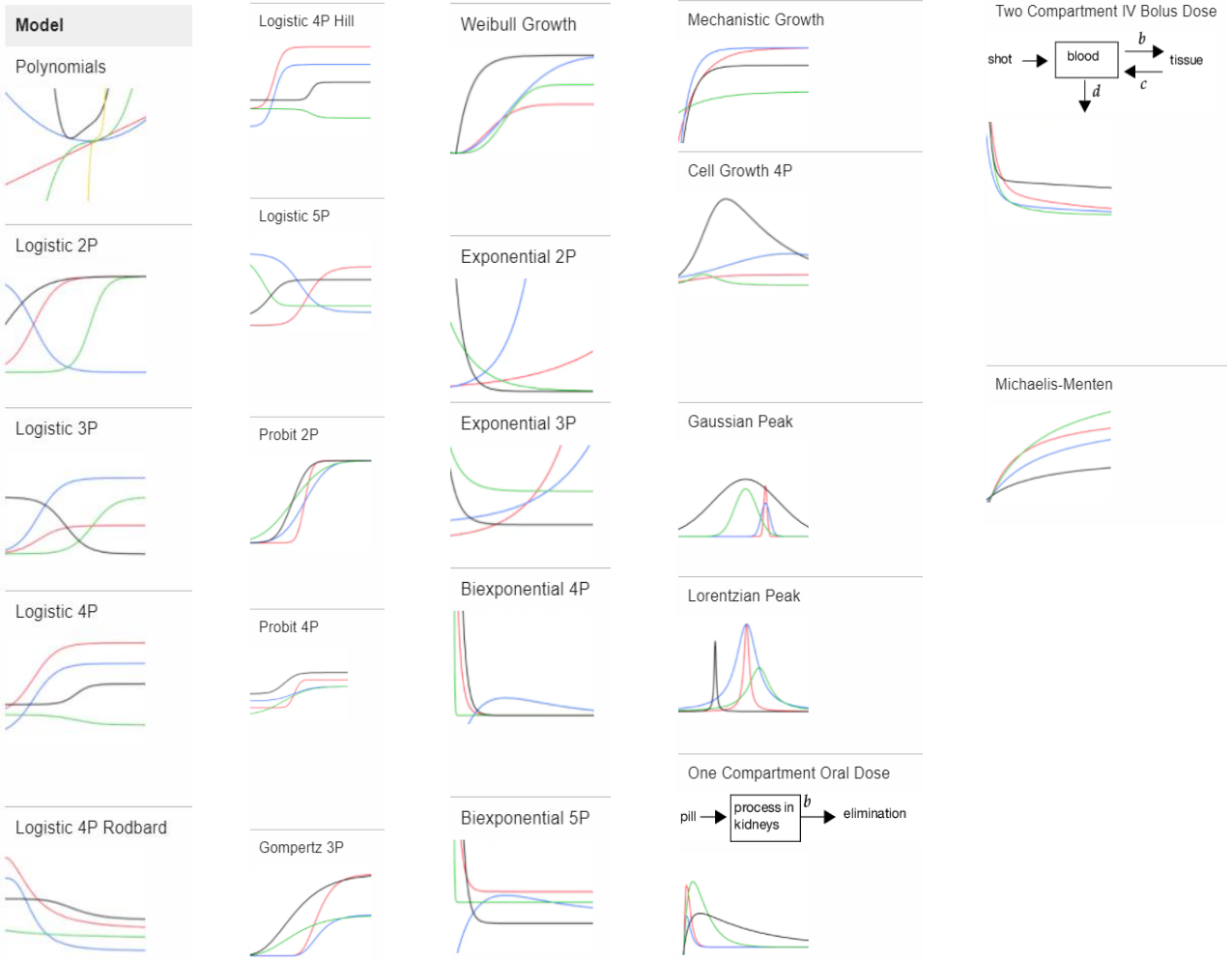


Starting Material Impurity (SM Imp) Model



Starting Material Impurity (SM Imp) Model





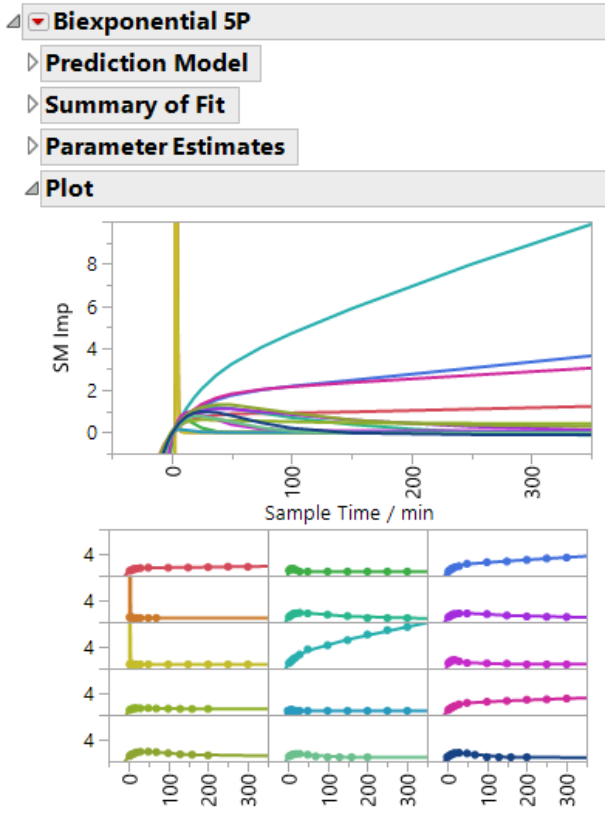
Starting Material Impurity (SM Imp) Model

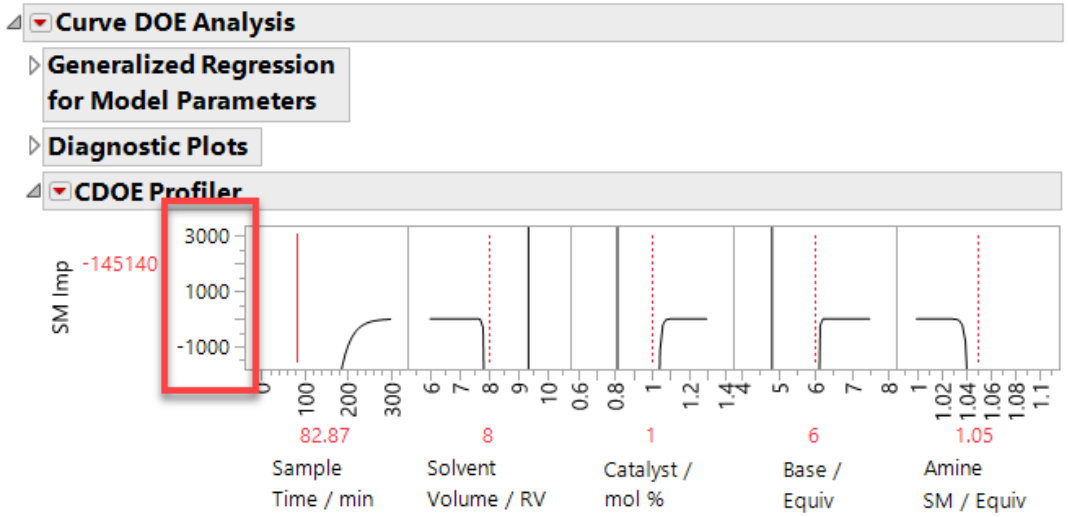
Fit Curve - SM Imp

Model Comparison

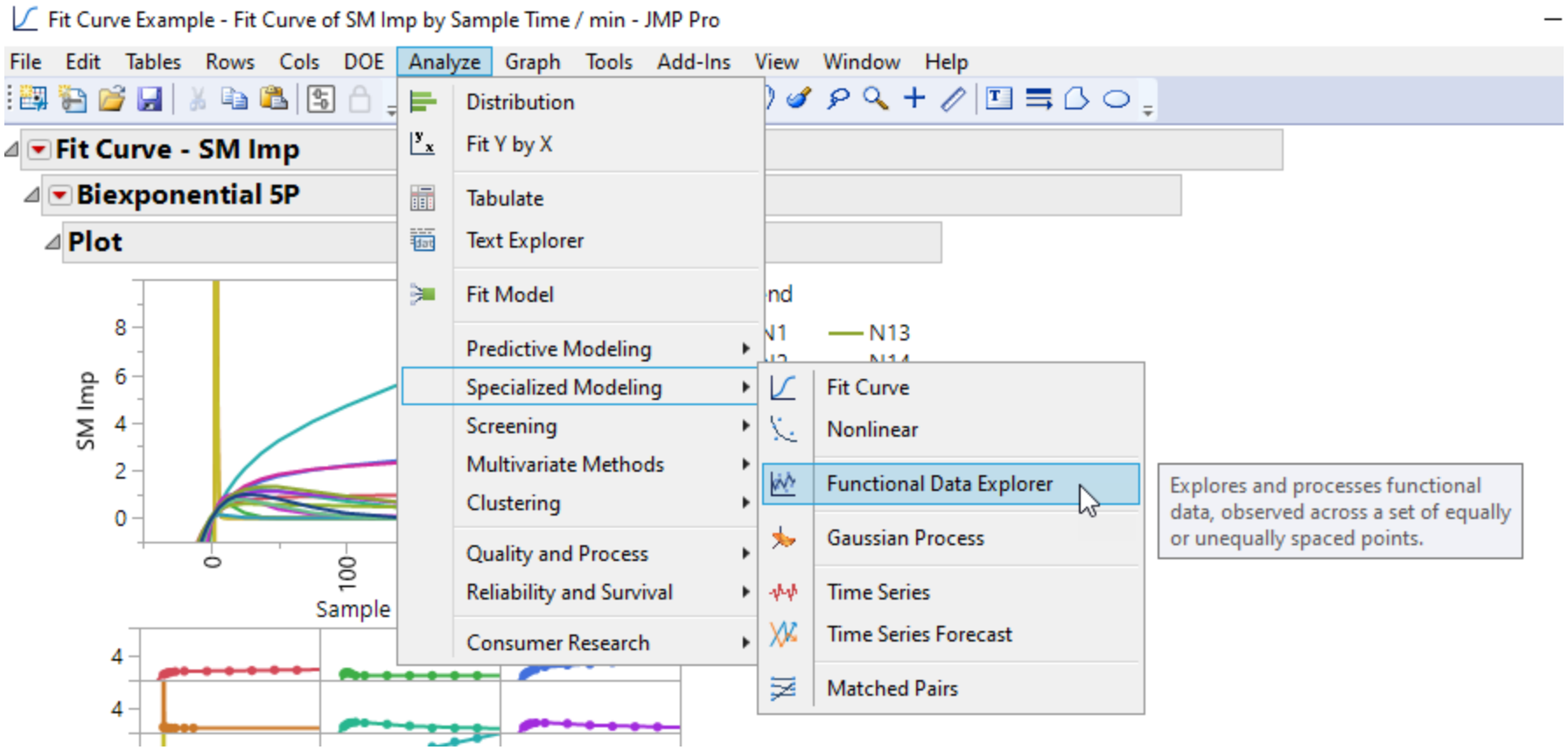
Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Biexponential 5P	-248.5568	1		-122.6224	0.5726787	0.00556	0.0745653	0.9980244
Biexponential 4P	-211.1367	7.4874e-9		-82.2548	1.1118453	0.0094224	0.0970691	0.9961644

Starting Material Impurity (SM Imp) Model

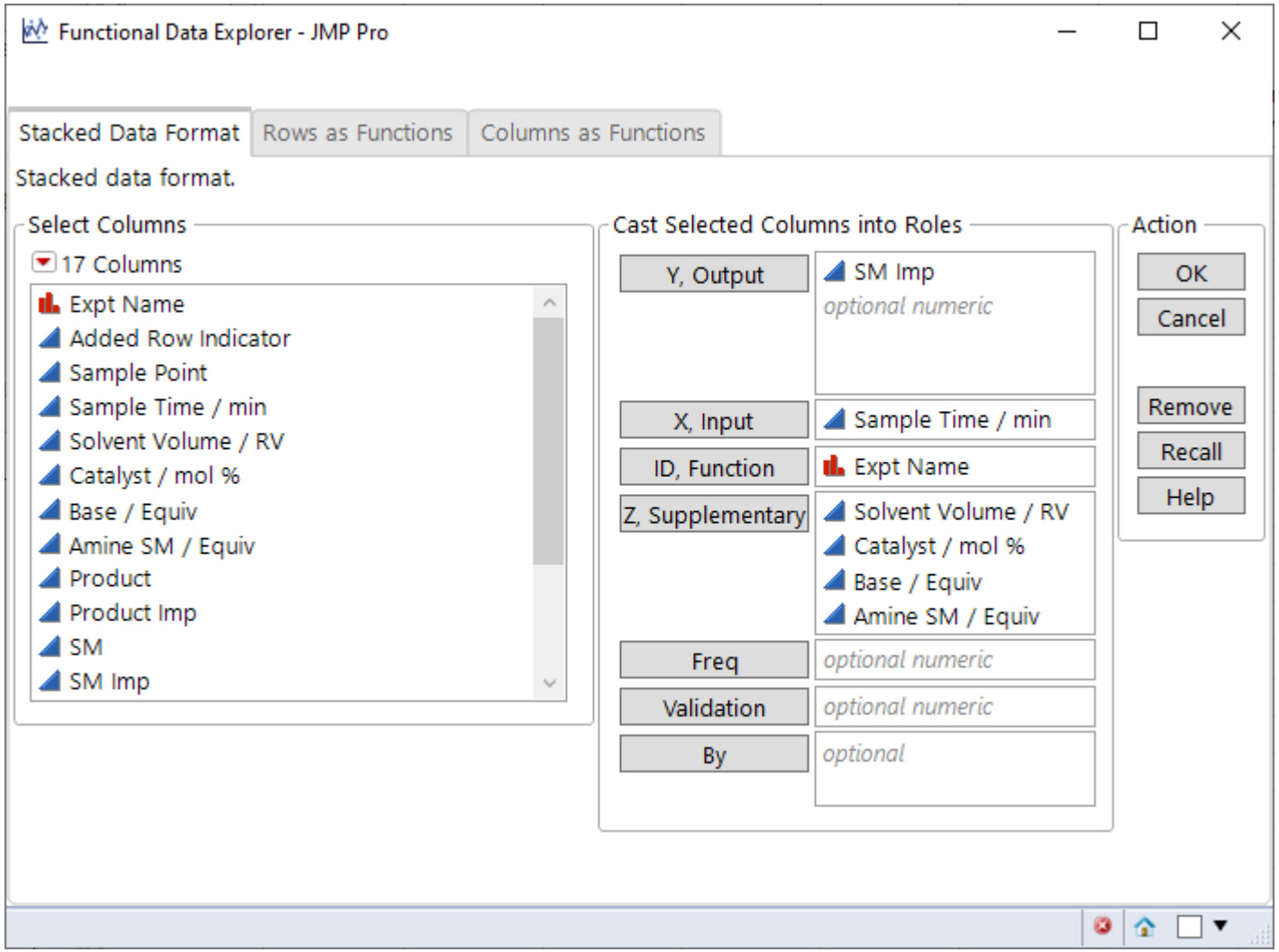




Starting Material Impurity (SM Imp) Model

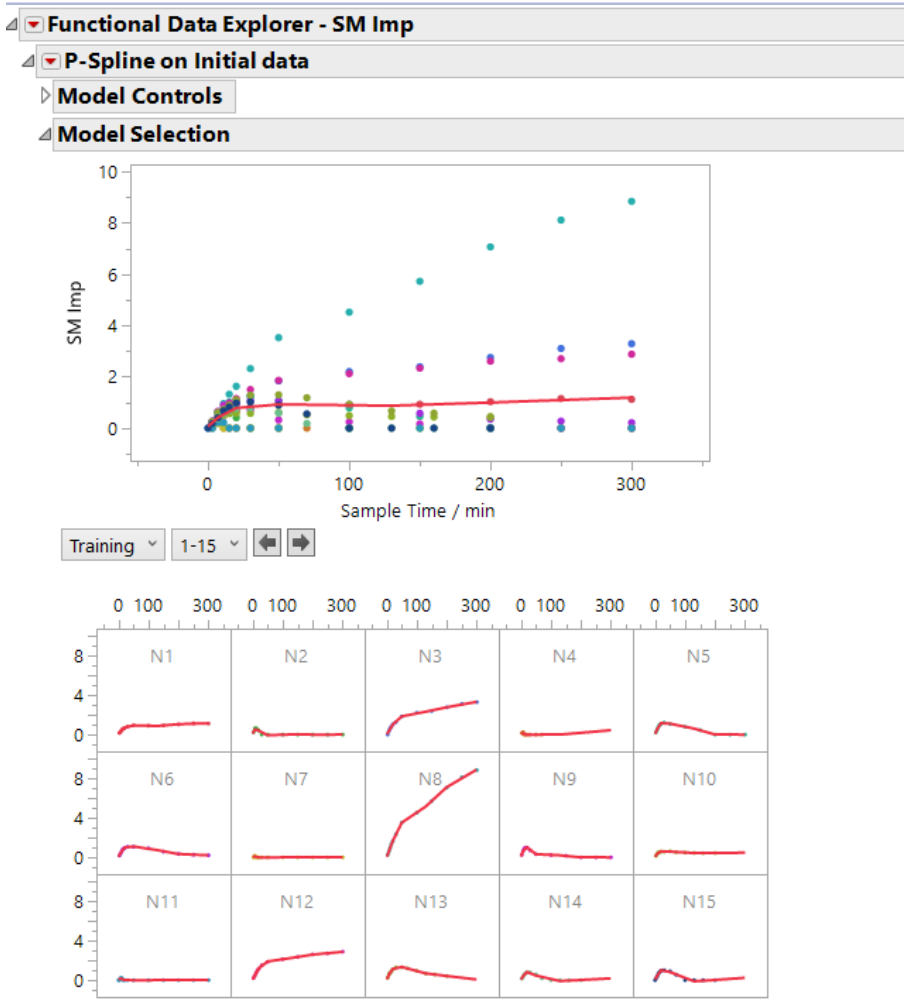


Starting Material Impurity (SM Imp) Model

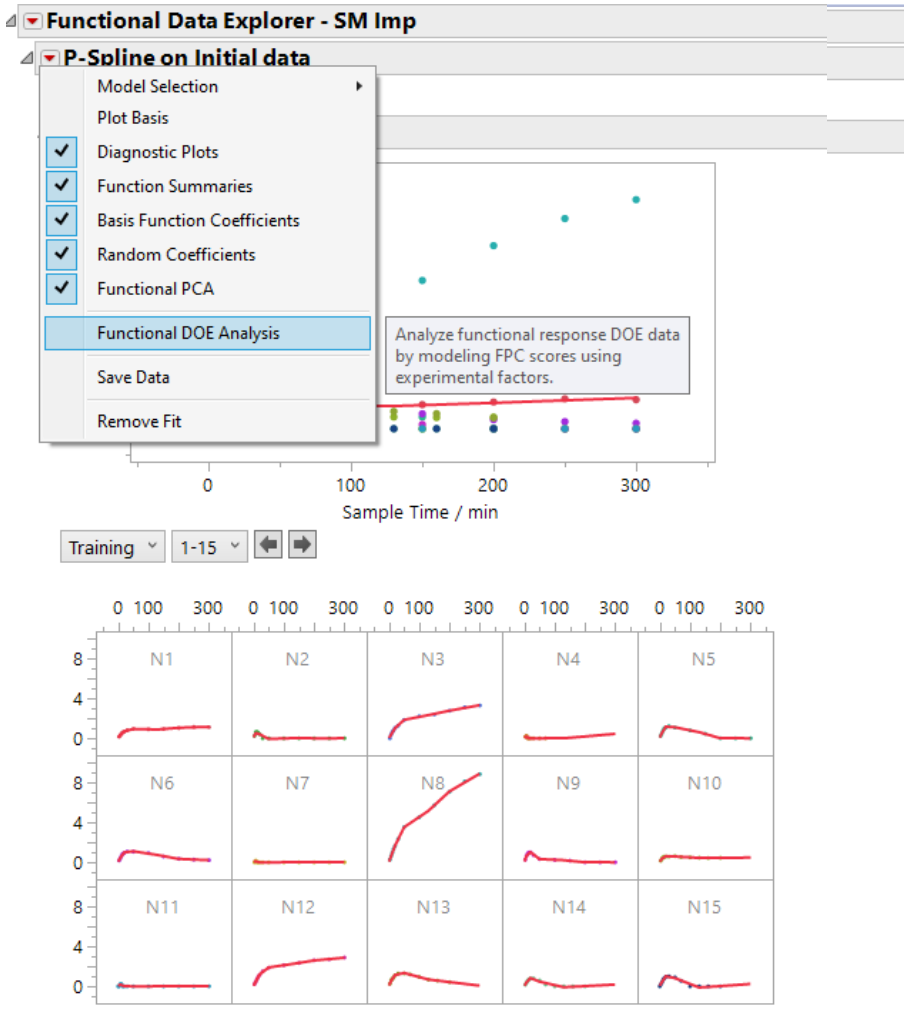


Starting Material Impurity (SM Imp) Model

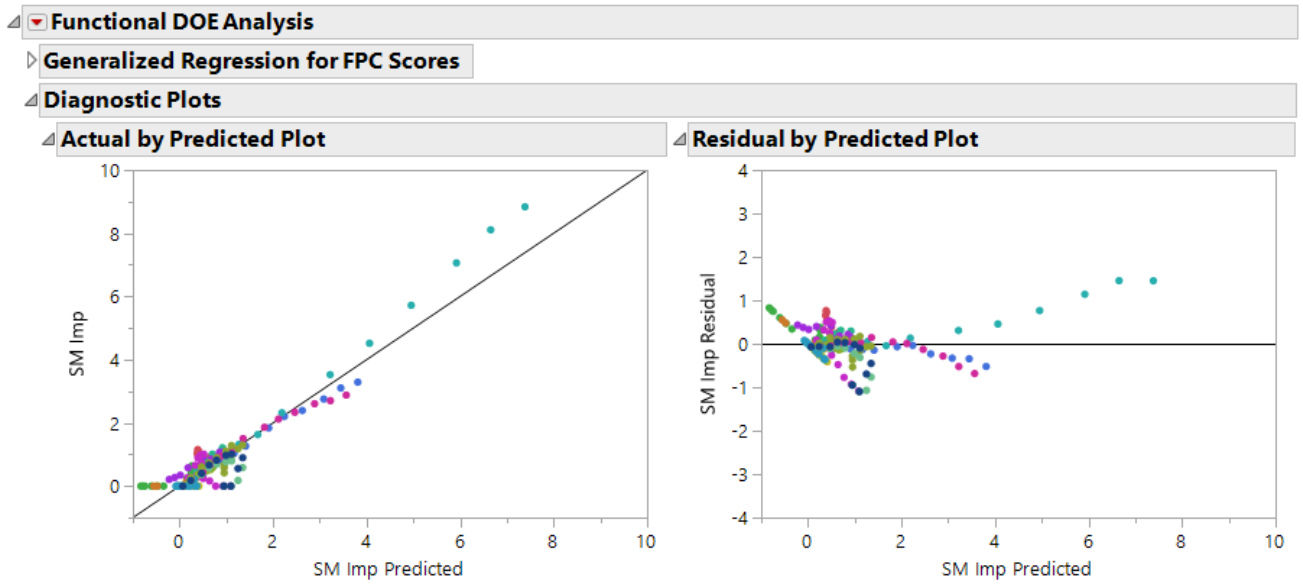
Starting Material Impurity (SM Imp) Model



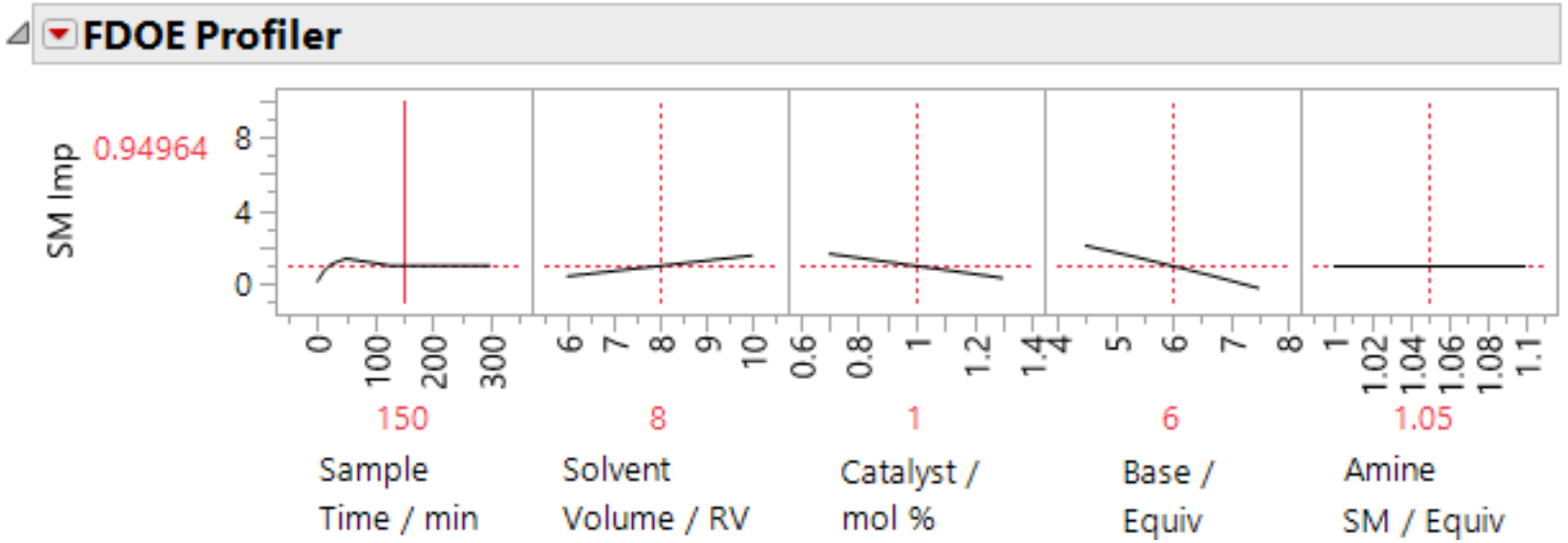
Starting Material Impurity (SM Imp) Model



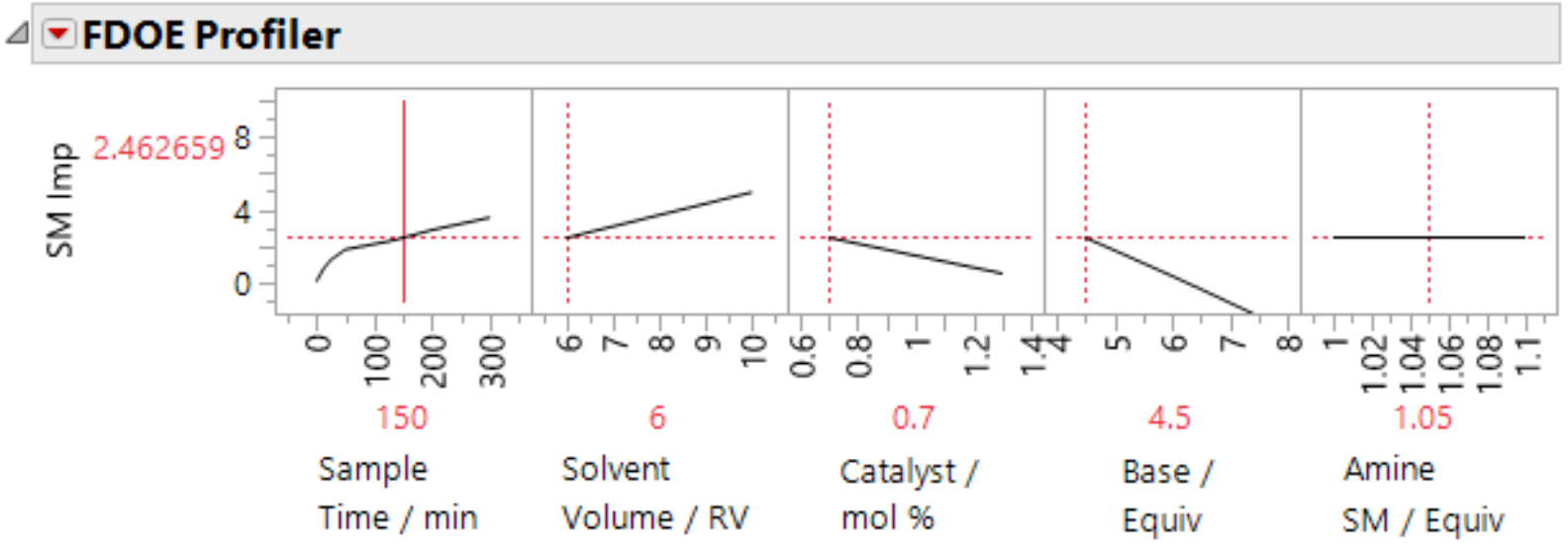
Starting Material Impurity (SM Imp) Model



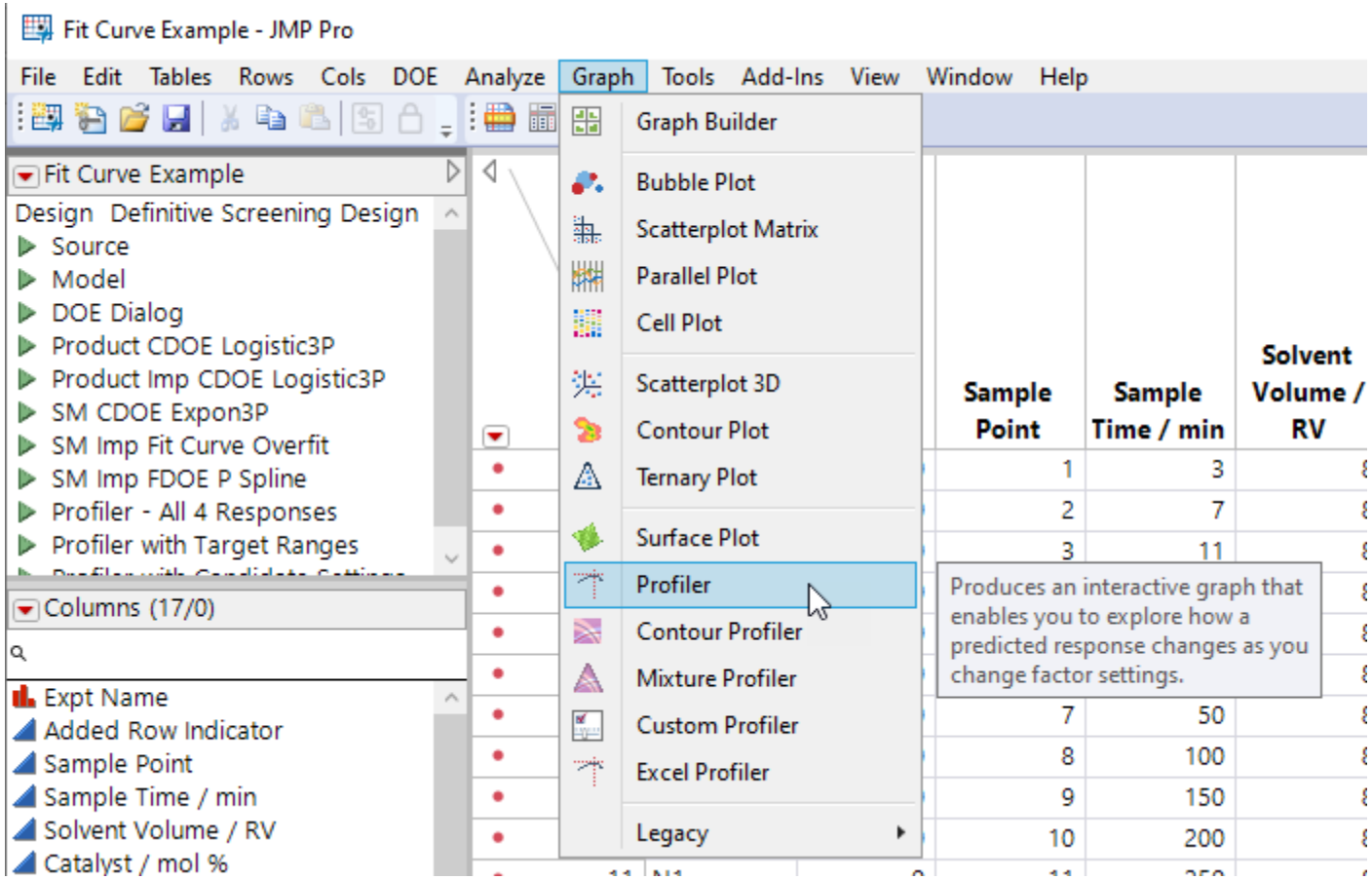
Starting Material Impurity (SM Imp) Model



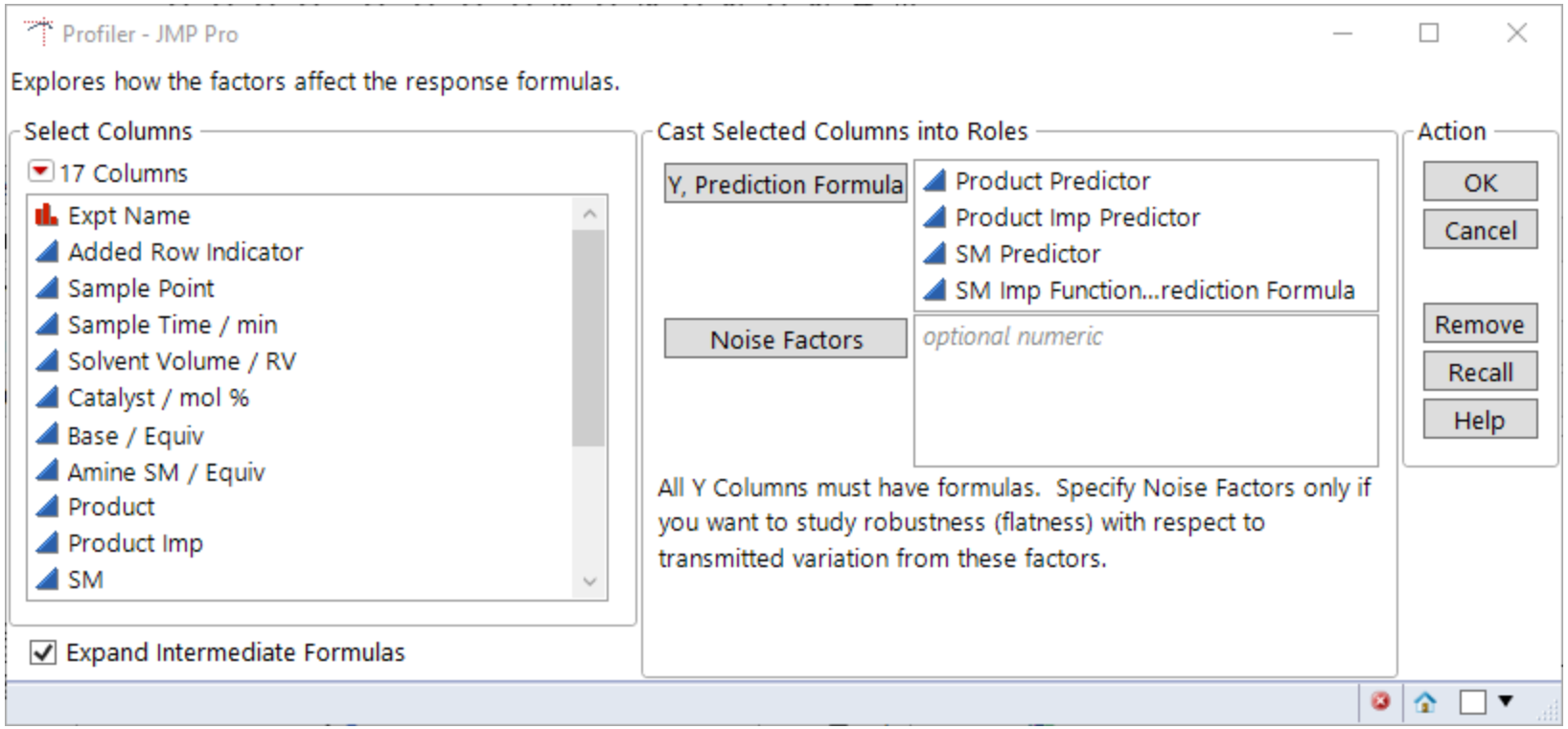
Starting Material Impurity (SM Imp) Model



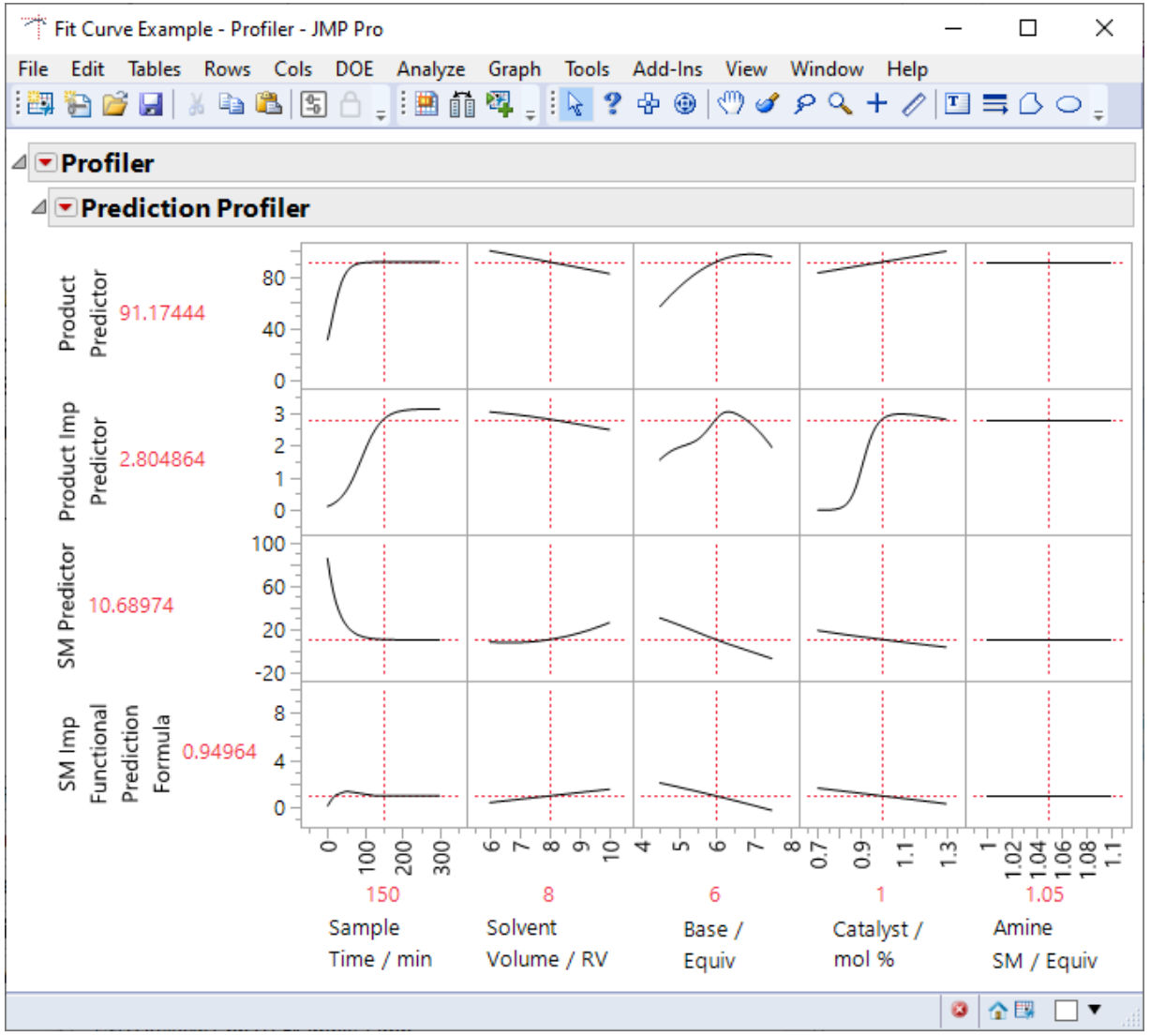
Profiler For All Four Responses



Profiler For All Four Responses



Profiler For All Four Responses



Fit Multiple Curves to Chemical Reaction Response Compare Model

- Fit Curve – Choose multiple curves to fit.
  - Review Model Comparison and Model Diagnostics to select a Model.

Fit Curve

Model Comparison

Model	AICc	AICc Weight	.2 .4 .6 .8	BIC	SSE	MSE	RMSE	R-Square
Logistic 4P	774.21661	1		901.26526	287.54791	2.4788613	1.57444	0.9980418
Logistic 3P	925.40129	1.481e-33		1037.7242	970.09161	7.4052794	2.7212643	0.9933937
Weibull Growth	1132.045	1.989e-78		1244.3679	3138.5071	23.958069	4.8946981	0.9786268
Logistic 5P	1358.7337	1.19e-127		1481.4682	5000.6892	49.511775	7.0364604	0.9659454
Probit 4P	1463.7303	1.88e-150		1590.7789	14459.122	124.64761	11.164569	0.9015335

- What diagnostic to assess to determine one model better than another?
- Response Fit Curve selections
  - Prod, Prod Imp – Logistic 3P, SM – Exponential 3P, SM Imp – FDE P Spline

Subject Matter Expert Guided Selection - Product

- May not always choose pre-loaded formula with best “Statistical” diagnostic results

Model Comparison

Model	AICc	AICc Weight	.2.4.6.8	BIC	SSE	MSE	RMSE	R-Square
Exponential 3P	766.00441	1		878.32729	392.18215	2.9937568	1.7302476	0.9973292
Logistic 3P	925.40129	2.44e-35		1037.7242	970.09161	7.4052794	2.7212643	0.9933937

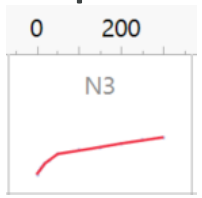
- For Product, Exponential 3P diagnostics better so why choose Logistic 3P
  - In this instance both very good. Subject matter proposal:
- The Exponential 3p is the same curve as kinetic reaction first order with fitted limits

Kinetic: $A = A_f + (A_0 - A_f)e^{-kt}$ Exponential $a + b \cdot \mathrm{Exp}(c \cdot \text{Sample Time / min})$

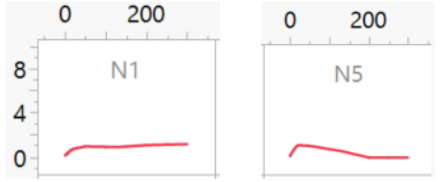
- Looking at the curves, the shape is more sigmoidal, suggesting more complex kinetics than first order so Exponential 3P lacks shape flexibility.
- Logistic 3P, lacks similarity to kinetic equations, more effective at describing sigmoidal curves.
- Taking more samples near time point 0 may have helped discriminate between formulas.

Subject Matter Expert Guided Selection – SM Imp

- SM imp can be formed and consumed (to form Product Imp). The pathways could occur at different rates to the extreme where one pathways is switched off.
  - N3, N8, N12 – SM Imp increases.
    - Pathways: Formation On, Consumption Off.



  - Other Expts – SM Imp rate plateaus or peaks and reduces.
  - Pathways: Formation On, Consumption On.



- None of the pre-loaded formulas could adequately fit the variety of profiles shapes created by the combined pathway effect.
- FDE P spline fitted making use of greater curve shape fitting flexibility.
- Ambitions is for First principle fitting approximation but in case ambition scaled back to empirical fitting approximation.

Select Curve DoE on formula parameter coefficients Compare Response Distributions

Curve DOE Analysis

Generalized Regression for Model Parameters

Generalized Regression for Growth Rate

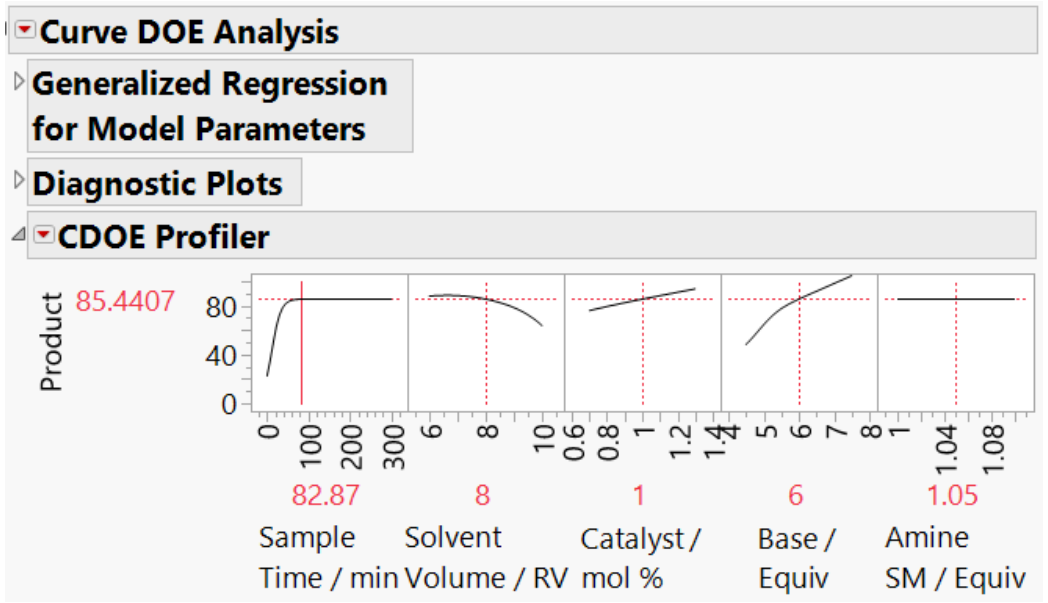
Model Comparison

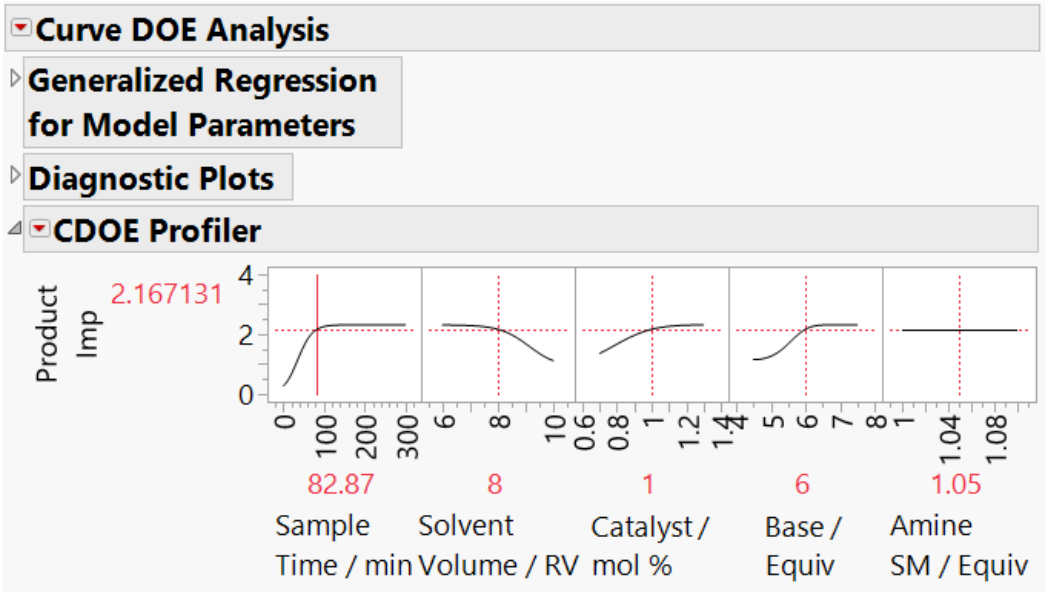
Show	Response Distribution	Estimation Method	Validation Method	Nonzero Parameters	AICc	BIC	Generalized RSquare
☑	Normal	Forward Selection	AICc	6	-39.33529	-45.58699	0.9426054
☑	Normal	Forward Selection	AICc	8	-38.70316	-57.03876	0.9813578
☑	Gamma	Forward Selection	AICc	6	-91.45945	-97.71115	0.9956105
☑	Weibull	Forward Selection	AICc	7	-80.04912	-91.09276	0.9944178
☑	Beta	Forward Selection	AICc	6	-88.25012	-94.50182	0.9948198

Model Launch

- What diagnostic to assess to determine one model better than another?

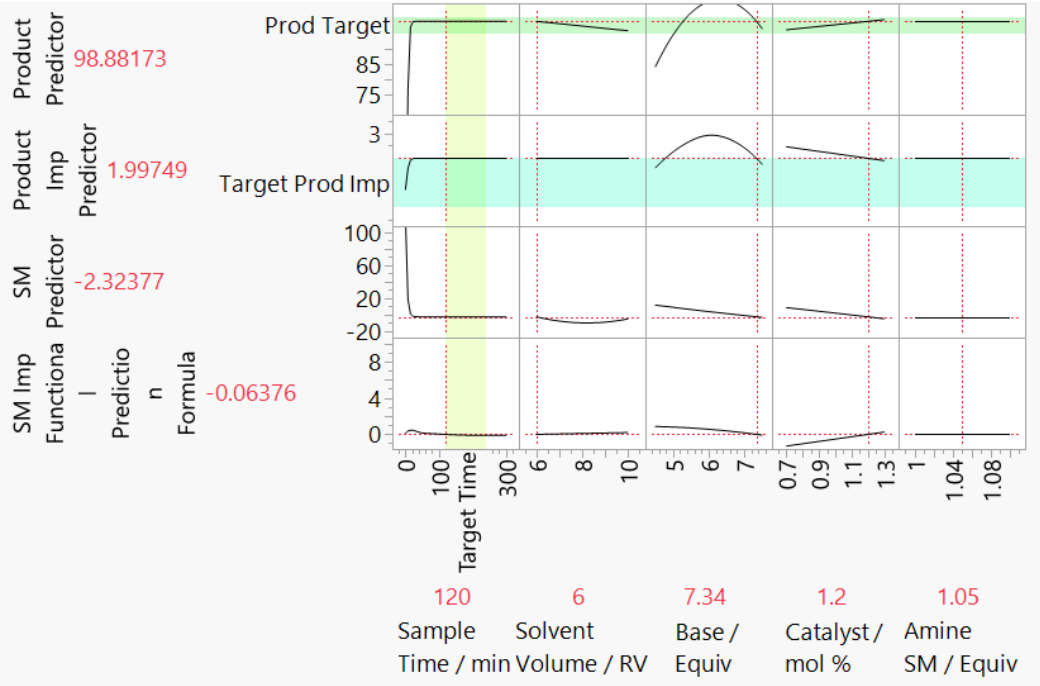
DoE Profiler for each reaction response

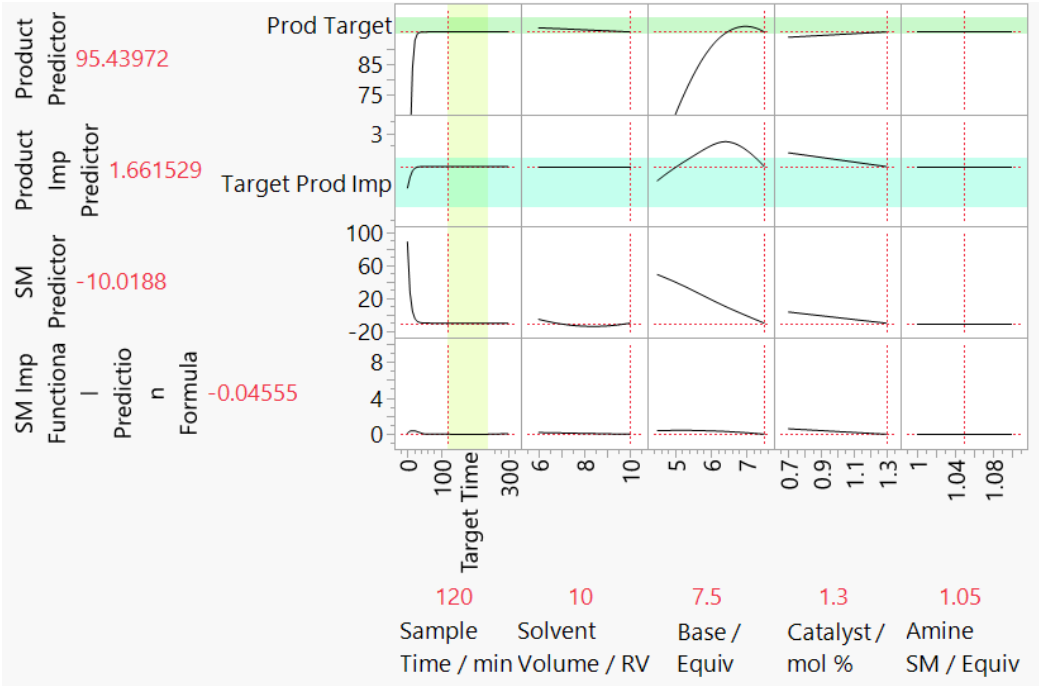




- Review profiles to understand factor influence on reaction response behaviour.
- Assess which factor settings have potential to comply with all response target criteria.

Group Profilers to facilitate Multi Target Criteria Evaluation





Response Target Criteria

Product, 95.0% or more (horizontal green)

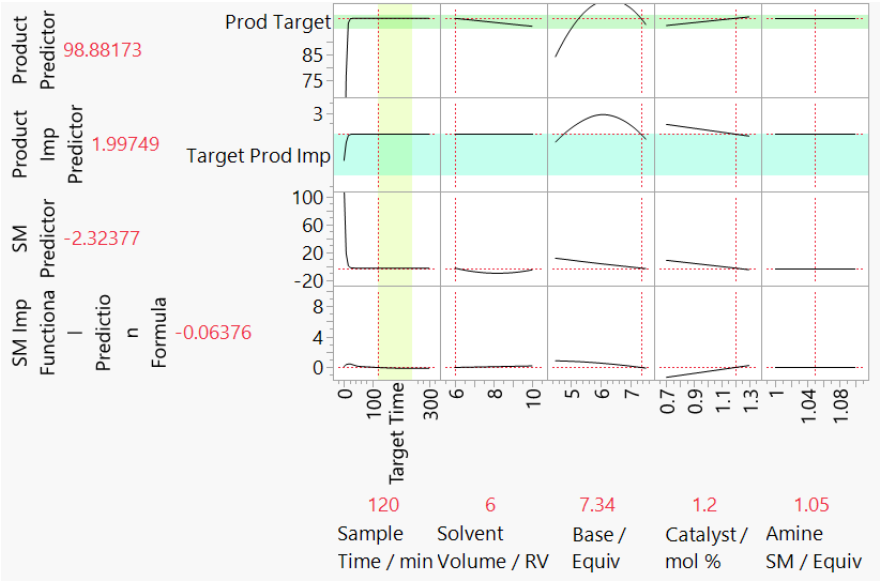
Product Imp, 2.0% or less (horizontal green/blue)

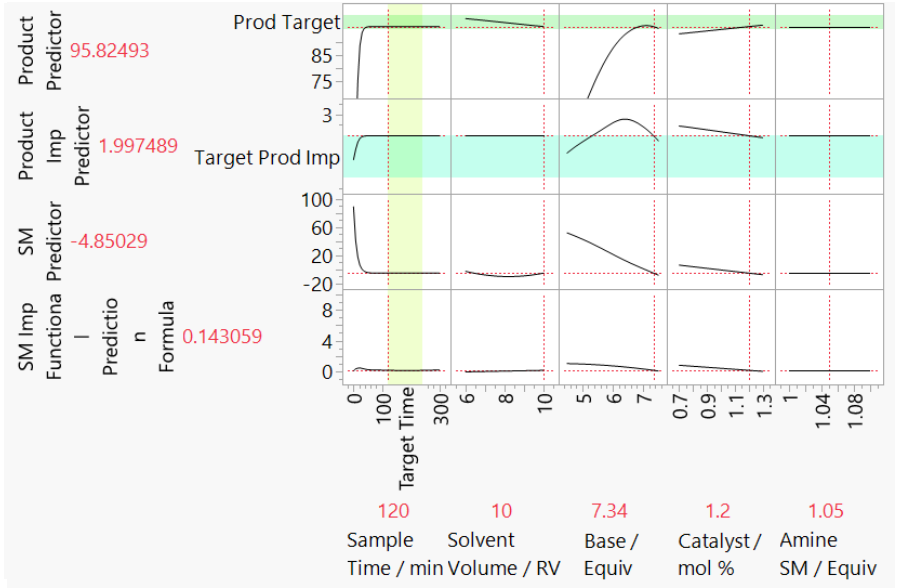
Reaction Time, 120 – 240 mins (vertical green)

Interested in the Group Profiler top 2 rows

Relating Group Profiler to Process Insight

- Key Responses: Prod and Prod Imp - Group Profiler top 2 rows
- Visualises impact over timeline of changing factor levels on key responses
- Factor combinations exist predicting Responses achieving target criteria
  - SM Amine not influential (Flat line), High Base minimise solvent influence (Flatter line)
  - Factor combinations limited due to sensitivity in Base and Catalyst acceptable levels
  - 12–240 mins, 6–10 vols, Base 7.34–7.5 eq, Catalyst 1.2–1.3 mol%, SM Amine 1.0–1.10 eq





Summary: Fit Curve Pre-loaded formula Improvements

Fit Curve Pre-loaded formulas

- User Specify lower and upper bounds to theoretical values
- User Specify lower and upper asymptotes to meet defined lower and Upper theoretical values
- Help with alerts to over-fitted pre-loaded equations
- Improved AICs performance to detect over fitting
  - Until then check Model Diagnostics - observed vs predicted results
  - How good should formula fit be on each experiment?
    - Majority of formulae parameter coefficients, for every experiment, statistically significant?

Summary: Curve DoE Improvements

Curve DoE

- For DSDs, default to 2-stage analysis approach.
- Bound original Y response to theoretical possible results eg 0 – 100% product.
- Select appropriate distribution to apply to fit curve parameter coefficients.
  - Gradient lies within 0 -1, therefore Beta distribution a better descriptor than a Gaussian(Normal) distribution?
  - Expand possible distribution options to cover distribution which reflect better features and number ranges observed in parameter coefficients estimates eg values always negative.
- What Design choices are most suited for reaction profile experimentation modelling – Factorials, Definitive Screening Designs, Response Surface Modelling, space filling?

Summary

- 2-step Analysis approach.

  i) Fit Curve on each experiments Chemical responses.

  ii) Curve DoE analysis on Fit Curve coefficients, converted back into Chemical responses and visualised in CDOE profiler.

- Tries to introduces subject matter knowledge into Analysis approach.
- Curve DoE gives insight into which factors influential on each response.
  - Solvent, Base, Catalyst factors at least influential on one response. Some factors interact, some are non-linear. The detail is available if needed.
  - SM Amine non-influential on responses.
  - The Group Profiler shows the factors combined net influence on responses.
- Unfinished approach, shows potential, we will continue to develop and refine approach.

Acknowledgements

AstraZeneca

- Rob Cox, Brendan Nixon