On the Use of Regularization Techniques to Minimize the Oscillatory Behavior of Dynamic Response Surface Models

Jacob Santos-Marques

Table of Contents

Abstract

Dynamic response surface model (DRSM) methodology is a generalization on response surface models (RSM) that greatly facilitates the ability to determine reaction stoichiometries and rate information for a black box system. Activities reported in this thesis implement new strategies to tackle two of the weaknesses within the existing methodology: oscillatory end behavior and subjective determination of significant singular values. Regularization techniques were explored as remedies for the first weakness, while both empirical and statistical methods were explored to resolve the latter weakness. L_2 regularization, also known as ridge regression, was shown to reduce the oscillation-based error by over 50% for a targeted problem species in a pharmaceutical reaction system provided by Pfizer Inc. Reduction in oscillatory behavior stabilizes model extrapolation and downstream rate calculations. An f-test strategy that compared variances of singular values was determined best for objective determination of significant singular values, which is necessary to establish the number of expected independent reactions occurring in the black box system. A new iteration of the methodology is presented incorporating these improvements to overcome the two targeted weaknesses.

Introduction

Dynamic Response Surface Models have shown impressive promise in offering a new and efficient data-driven methodology by which an unknown time-evolving system can be modelled throughout time. The most impressive capability of this rising methodology is the subsequent ability to test and predict potential reaction stoichiometries to identify the independent reactions occurring in a black box. Additionally, once the reaction stoichiometries of the system are identified, it is possible to not only graph reaction rates vs time, but also estimate the rate constant of the reaction (Klebanov & Georgakis, 2015). These higher capabilities facilitated by DRSM methodology are not possible with other similar modelling techniques, such as response surface methodology.

First introduced by Klebanov and Georgakis, the methodology was applied to a simple batch process with two species, a semi-batch with five species, and a penicillin fermentation process (Klebanov & Georgakis, 2015). The viability of the methodology was then tested through a collaboration between Georgakis and Pfizer Inc. using simulated data from a reasonably complex pharmaceutical kinetic system consisting of ten species (Georgakis, 2016). In a continuation of that collaboration, Santos-Marques worked with Pfizer Inc. to examine the robustness and weaknesses of the methodology (Santos-Marques, 2016). Activities reported in this thesis detail possible improvements to the methodology that aim to tackle three of the major weaknesses identified in the methodology: oscillatory end behavior of the model, objective determination of significant singular values, and false positives on untrue reaction stoichiometries.

DRSM

The dynamic response surface model, DRSM, is a generalization of response surface models, RSMs, that vastly improves data analysis capabilities. DRSM methodology uses time-dependent concentration data to model a time-dependent output, unlike RSM which takes input from a specific time, typically the end of batch time, and models output at that same end of batch time. The quadratic response surface model general equation is as follows (Montgomery, 2013):

$$y = \beta_0 \sum_{i=1}^{n} \beta_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij} x_i x_j + \sum_{i=1}^{n} \beta_{ii} {x_i}^2$$

Equation 1

The general RSM equation defines a surface in (n+1) dimensional space with dimensionality corresponding to the factors, x_i, in the experimental design. The coefficients, β_{ij}, estimated by linear regression are the respective weights for each factor and factor pairing, indicating the importance of a factor on the output, y. If any given $\beta_{i,j}$ is zero, then the output does not depend on the corresponding factor pairing, creating a simpler than anticipated model.

Response surface methodology is an effective data driven modelling tool that allows one to predict species concentrations at end of batch for varying conditions after conducting well designed experiments. With these same well designed experiments, once can create a dynamic response surface model by measuring concentration data throughout the batch time. Creation of a DRSM requires the same time and experimentation as an RSM but boasts the capabilities to predict through time, while greatly facilitating the ability to predict stoichiometries and calculate rate constants. The general model for a quadratic dynamic response surface model is as follows (Klebanov & Georgakis, 2015):

$$y(\tau) = \beta_0(\tau) \sum_{i=1}^{n} \beta_i(\tau) x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \beta_{ij}(\tau) x_i x_j + \sum_{i=1}^{n} \beta_{ii}(\tau) {x_i}^2$$

Equation 2

The modelled concentration, y, and the coefficients, $\beta_{i,j}$, now depend on dimensionless time, τ.

$$\tau = \frac{t}{t_f}$$

Equation 3

The β coefficients are parametrized in time via the application of shifted Legendre polynomials, a polynomial expansion chosen for their orthogonality in the (0,1) interval of dimensionless time, τ. The first three shifted Legendre polynomials, SLPs, are as follows, accompanied by the recurrence relation by which higher ordered SLPs can be determined:

$$P_0(\tau) = 1$$
$$P_1(\tau) = -1 + 2\tau$$
$$P_2(\tau) = 1 - 6\tau + 6\tau^2$$

$$P_i(\tau) = \frac{(2i-1)(2\tau-1)P_{i-1} - (i-1) * P_{i-2}}{i}$$

Equation 4

The regression for a DRSM estimates the coefficients, γ_{ij}, of the SLPs used to parametrize the betas. When compared to an RSM, regression on a DRSM will need to estimate R times as many coefficients, where R is the number of SLP used to parametrize the betas.

$$\beta_0(\tau) = \sum_{r=0}^{R-1} \gamma_{0,r} P_r(\tau) \qquad \beta_i(\tau) = \sum_{r=0}^{R-1} \gamma_{i,r} P_r(\tau)$$

$$\beta_{ij}(\tau) = \sum_{r=0}^{R-1} \gamma_{i,j,r} P_r(\tau) \qquad \beta_{ii}(\tau) = \sum_{r=0}^{R-1} \gamma_{i,i,r} P_r(\tau)$$

Equation 5

Determination of R is vital to create an effective model. In the first methodology posited by Klebanov and Georgakis, a lack of fit (LoF) p-value was used to determine the significance of the model (Klebanov & Georgakis, 2015). If a model was determined to have significant lack of fit for a given R, then the number of polynomials would be increased. Significant lack of fit was defined as an f-test on lack of fit mean square over pure error mean square yielding a p-value below 0.05; this threshold represents a 95% confidence that the inaccuracies of the model are due to pure error, which cannot be modelled (Montgomery, 2013). More recently cross-validation has been successfully used by Santos-Marques to determine the best R for a given model across varying levels of error (Santos-Marques, 2016).

The true benefits of using DRSM methodology are downstream, after formation of the models themselves. Determining the reaction stoichiometries and calculating the respective reaction rate constants first requires an accurate evaluation of the number of singular values that are significant in describing the system. An (N) x (nS) matrix, RS, is formed containing the rate of appearance/disappearance of each species. RS is calculated via the obtained DRSMs created for each species, using derivatives of the SLPs with respect to τ. Each column of RS represents a different species (nS being the number of observed species), while the rows indicate the experimental conditions (with N total combinations).

Singular value decomposition if performed on the matrix RS.

$$RS = U\Sigma V^T$$

Equation 6

The diagonal matrix, Σ, contains entries representing the singular values of the matrix RS. A number of these values are determined to be significant (methods for doing so are discussed in *Significant Singular Values*). Significant singular values correspond to singular vectors in V that describe the data. Insignificant singular values indicate that the corresponding vectors in V represent error and do not contribute meaningful information to RS. If k singular values are deemed significant, then the first k rows of V^T, renamed V^T_k, are kept and used to form a projection matrix, P_k.

$$P_k = V_k V_k^T$$

Equation 7

This projection matrix is used to project potential stoichiometries onto the subspace spanned by the significant singular vectors, V^T_k.

With reaction stoichiometries identified, it is then possible to evaluate the respective reaction rates using the rates of appearance/disappearance of each species and the corresponding stoichiometric coefficients, assuming elementary kinetics.

Prior Work

Pfizer Inc. provided a set of simulated data by which the DRSM methodology would be tested. The data represented a true, reasonably complex pharmaceutical system of ten species and eight independent reactions.

RXN #	Species Stoichiometric Coefficients									
1	-1	-1	1	1	0	0	0	0	0	0
2	1	1	-1	-1	0	0	0	0	0	0
3	0	0	-1	1	1	0	0	0	0	0
4	0	0	0	0	-1	1	0	0	0	0
5	0	-1	0	-1	0	0	1	0	0	0
6	0	1	0	1	0	0	-1	0	0	0
7	0	0	0	1	0	0	-1	1	0	0
8	-1	0	0	0	0	-1	0	0	1	0
9	-2	0	0	0	0	0	0	0	0	1
10	0	-1	0	0	2	0	0	0	1	-1

Table 1: Species Stoichiometric Coefficients for Pfizer System

The above table shows the stoichiometric coefficients of the ten species in each of ten reactions. However, this reaction system is comprised of only eight linearly independent reactions because two of the reactions are reversible. This can also be shown via the rank of the stoichiometry matrix.

The data provided by Pfizer Inc. followed a face centered central composite design with three factors: temperature, initial concentration of species four, and initial concentration of species two. The face centered central composite design requires 17 experiments and has a cubic

design space in three dimensions where each dimension measures a factor in the design. The design requires experiments at the eight vertices of the cube, the center of the six faces, and three experiments at the center point.

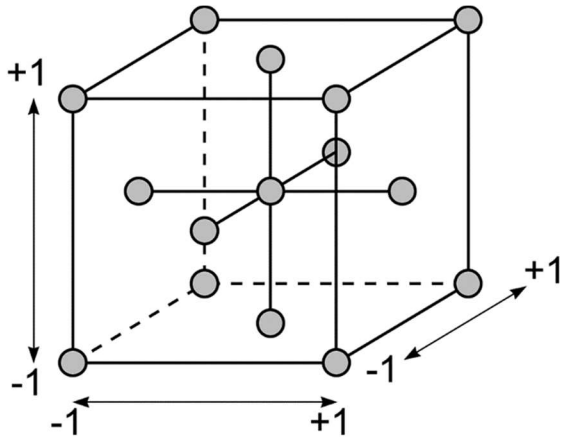
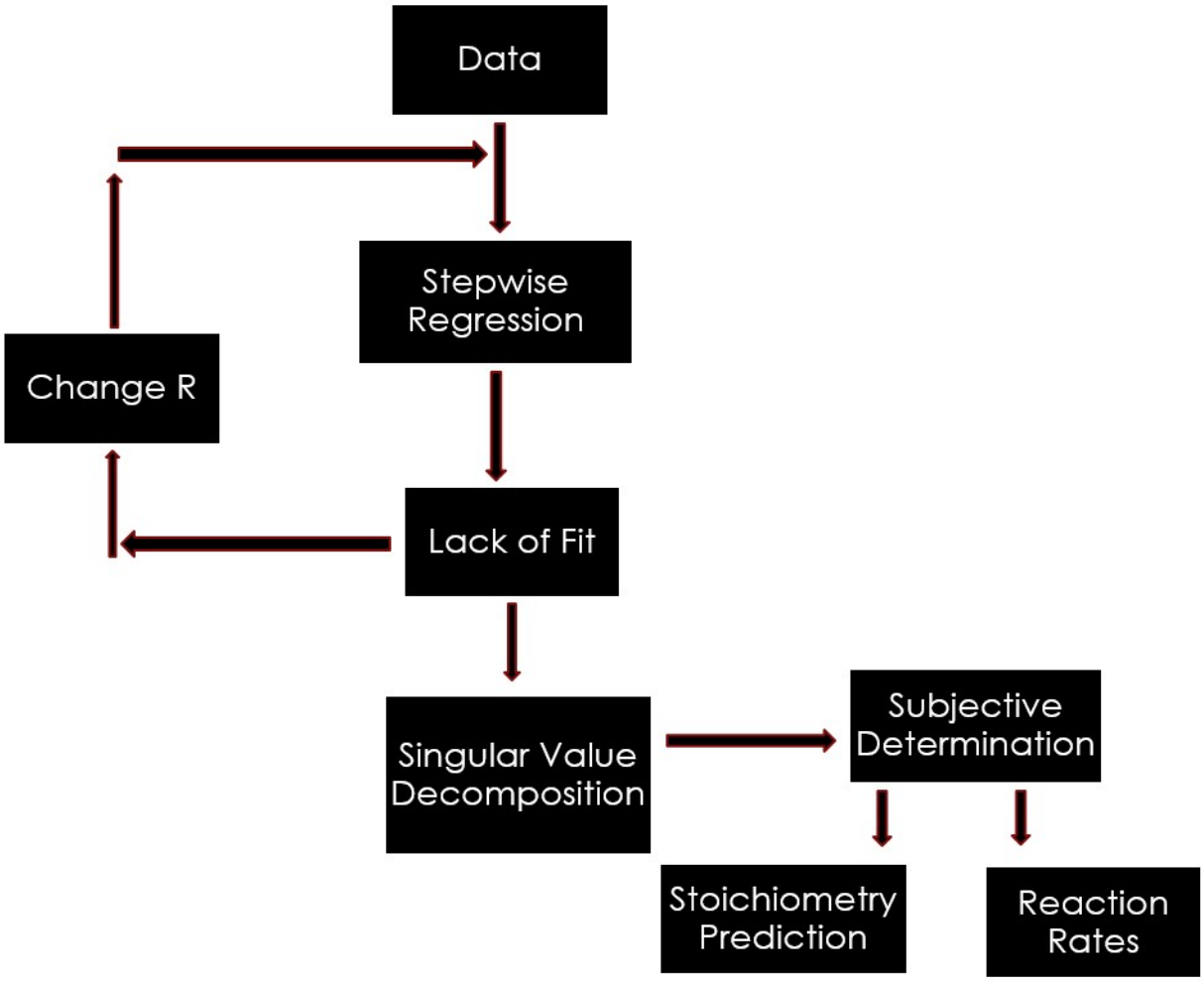


Figure 1: Face Centered Central Composite Design
http://manufacturingscience.asmedigitalcollection.asme.org/article.aspx?articleid=1746010

Blind DRSM Test

In a collaboration with Pfizer Inc, Georgakis was tasked with proving the capabilities of the DRSM methodology with respect to the reasonably complex simulated pharmaceutical system. A MatLab script was created that first established the recipe for DRSM creation and application. This script operated only for the provided Pfizer dataset, and utilized stepwise regression and a lack of fit p-value test for model validation (Georgakis, 2016).

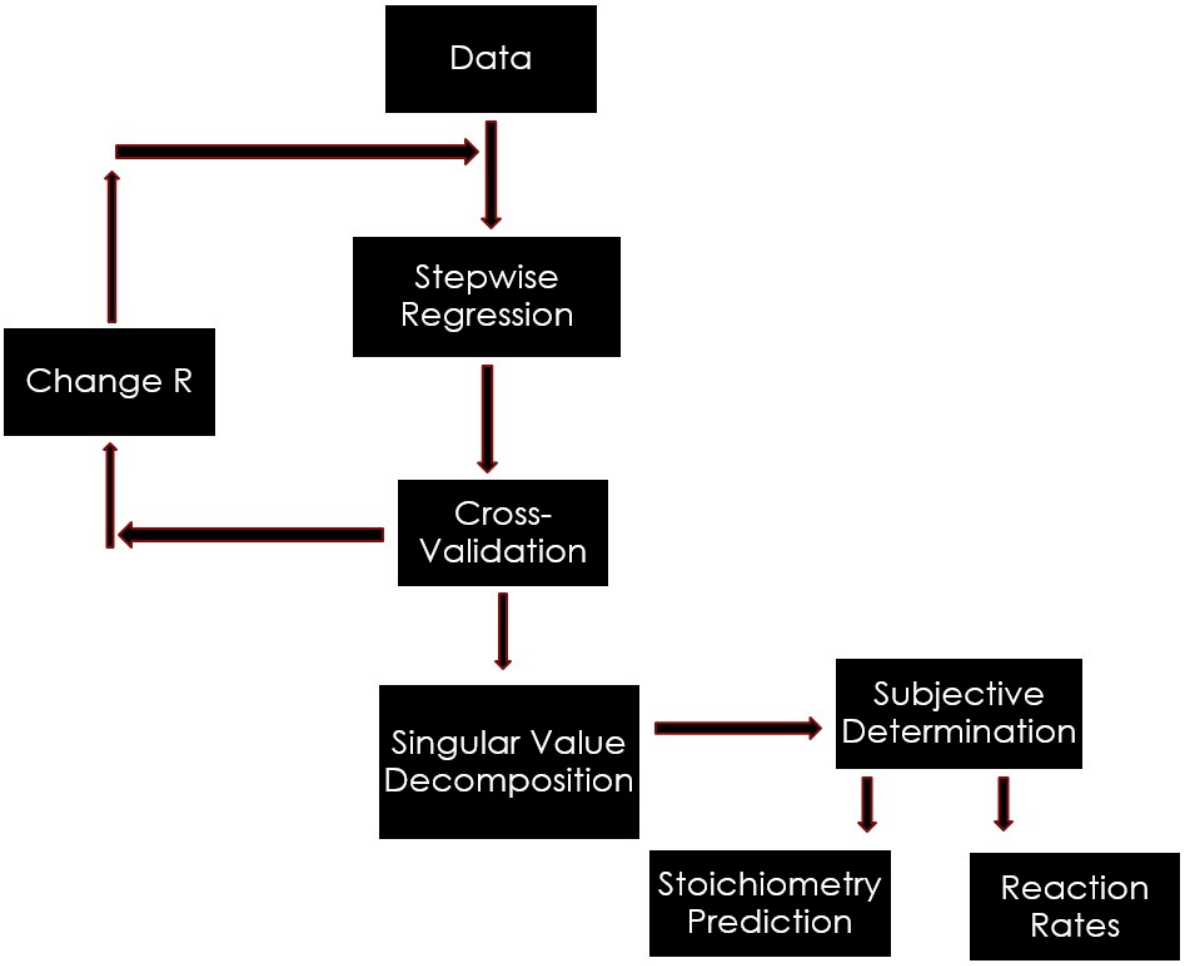
Figure 2: DRSM Methodology Flowchart, First Iteration, Georgakis, 2016



Cross-Validation Approach

After Georgakis' collaboration with Pfizer Inc., Santos-Marques, in a summer internship, was tasked with exploring the robustness of the methodology in greater detail. Tackling the same ten species, eight independent reaction system, the DRSM's versatility was tested with varying levels of error and different experimental designs. Because of this testing, a new MatLab script was created that replaced the LoF based polynomial determination with a leave one out (LOO) cross-validation method (Santos-Marques, 2016).

Figure 3: DRSM Methodology Flowchart, Second Iteration, Santos-Marques, 2016



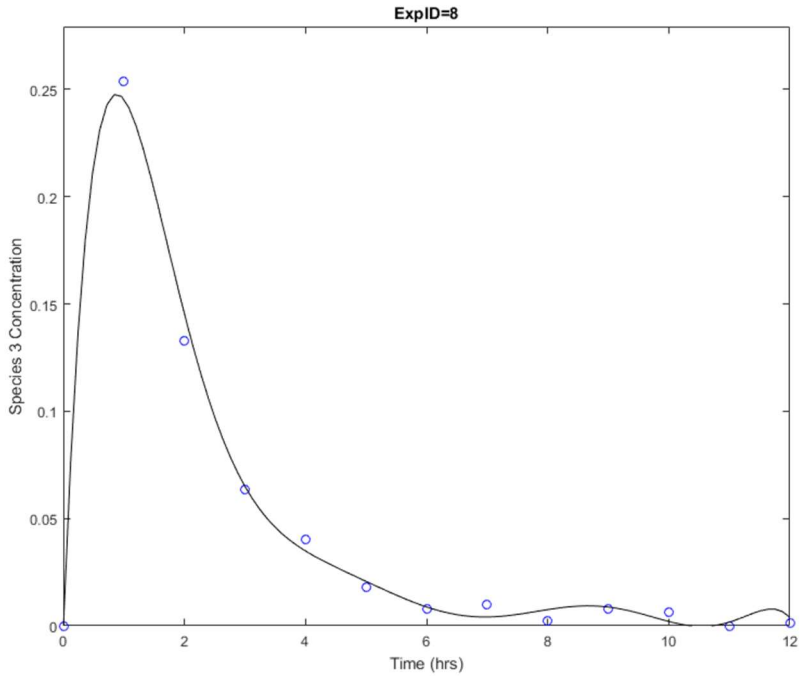
Weaknesses

Through extensive testing of the DRSM methodology, several problems were identified as foci for future research into DRSM improvements (Santos-Marques, 2016). Two of these identified weaknesses are the focus of this research, the oscillations and determining significant singular values. All major identified weaknesses are discussed briefly to provide contextual direction for the evolution of the DRSM methodology.

Oscillations

A recognizable trend in the DRSM predictions was oscillatory behavior in the prediction of some species concentrations, particularly species 3 and species 5. This trend occurred in areas of exponential decay, when the species concentration is approaching zero. Below is a prototypical example for species 3.

Figure 4: DRSM Prediction of Species 3 (in red) vs Trained Data (in blue)



The trained data, in blue, shows the simulated concentration with introduced error that was used to help create the species 3 DRSM. This wavy end behavior is not expected in a reaction system; one expects an exponential decay based on reaction engineering knowledge. This behavior also leads to severely inaccurate model extrapolation. This end behavior could be a result of overfitting, in which the DRSM is modelling error. The behavior was not resolved by decreasing the number of polynomials in the model. The above prediction was created via a nine-polynomial model, which was determined to be the best number of polynomials via cross-validation (Santos-Marques, 2016) and lack of fit (Georgakis, 2016).

Following are example modelled concentrations for species 2, 5, and 7. Species 5 shows oscillatory behavior and follows a similar trend to species 3, while the other species do not exhibit oscillations. Both species 3 and species 5 are intermediates as seen by Table 1.

.

Figure 5: DRSM Prediction of Species 2 (in red) vs Trained Data (in blue)

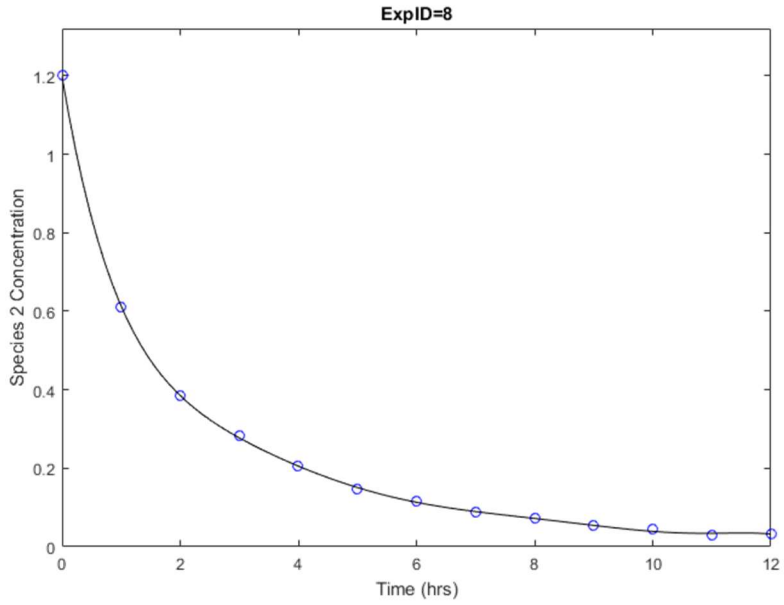


Figure 6: DRSM Prediction of Species 5 (in red) vs Trained Data (in blue)

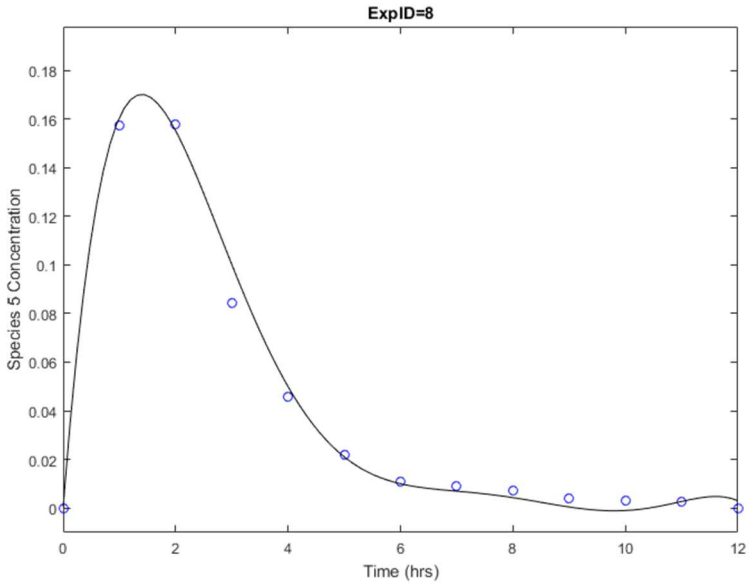
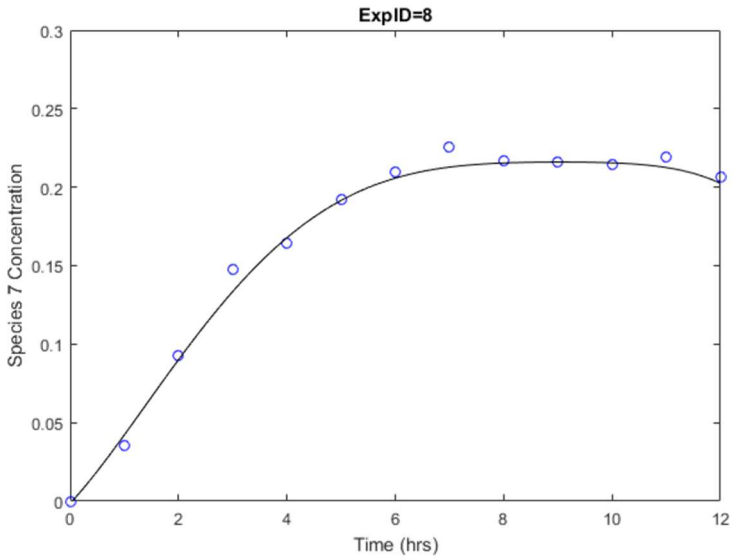


Figure 7: DRSM Prediction of Species 7 (in red) vs Trained Data (in blue)



Improving on this tendency to overfit when modelling such intermediates was determined to require methodological improvements on how the DRSM is created. Improving the modelling accuracy of the DRSM for such intermediate species is the focus of the research reported here.

Significant Singular Values

As previously alluded to, the determined number of significant singular values is vital in calculating the projection matrix, P_k. No objective method was incorporated within the methodology to determine the number of significant singular values. Previously, this determination was either done subjectively, or projections would be performed using every P_k with $1 \leq k \leq (nS-1)$. The singular values calculated using the first iteration of the DRSM methodology as applied to the Pfizer dataset were as follows (Georgakis, 2016):

19.4	10.7	5.5	3.8	0.99	0.53	0.00	0.00	0.00	0.00

Table 2: Singular Values Determined for Pfizer Dataset by Georgakis

The second iteration of the DRSM methodology created by Santos-Marques (2016) calculated singular values with a less obvious boundary between significant and insignificant:

135	47.6	17.3	13.5	7.7	4.8	4.4	2.5	0.9	0.5

Table 3: Singular Values Determined for Pfizer Dataset by Santos-Marques

This finding created the need to establish an objective function capable of determining the significant singular values.

Identifying Stoichiometries

A similar problem existed for the determination of "good" stoichiometric projections. At the time, projection score was calculated for each candidate reaction via the following formula:

$$n_r = P_k n_c \qquad \textit{Equation 8}$$

$$S(n_c) = 100\left(1 - \frac{\|n_r - n_c\|}{\|n_c\|}\right) \qquad \textit{Equation 9}$$

where n_c is the candidate stoichiometry vector. This method evaluates the deviation between n_c and n_r caused by the projection matrix (Georgakis, 2016). An accurate candidate reaction is

expected to be unaltered by the projection, resulting in a perfect score of 100. Because of introduced error and imperfect models, one cannot expect to obtain perfect scores for the accurate reaction stoichiometries. Therefore, there needs to exist a method by which one can objectively determine the validity of a stoichiometry based on its score. At the time of the report, only subjective methods existed.

False Positives

Further complicating the process by which stoichiometries were determined was the frequency of false positives encountered when scoring projected reactions that were not valid for the Pfizer system. This problem was identified by conducting an exhaustive test on every reaction stoichiometry with species coefficients spanning the integers: -2, -1, 0, 1, 2. It was found that an abundance of reaction stoichiometries, which were shown to be linearly independent from the true reaction system, scored very well. In some cases, the false reaction stoichiometries were even outperforming the true reaction stoichiometries (Santos-Marques, 2016). This result shows the need for a more refined and objective method by which candidate reactions are evaluated.

Data Restrictions

In the collaboration with Pfizer Inc., Santos-Marques improved upon the Matlab script first written by Georgakis that implemented the DRSM methodology. The initial script operated functionally only for the specific dataset and design provided by Pfizer. Santos-Marques created a script that generalizes the DRSM methodology so that it can be implemented across varying experimental designs and kinetic systems. Through this work to generalize the methodology for ease of implementation, three data problems were identified that would prevent the DRSM methodology from successfully operating. The following three weaknesses exploit the rigidity

of the performed regression; future attempts to resolve these problems will focus on the representation of the matrices used in the regression (Santos-Marques, 2016).

Experiments with Unequal Length

The first experimental design problem that could not be handled by the methodology was experiments with different lengths in time. In this situation, the data will be curtailed to the shortest duration experiment, throwing away any data collected in other experiments past that minimum duration. This creates a DRSM functional on the smallest time domain possible.

Experiments Collecting Data at Different Times

The second experimental design problem, which is directly related to the first, is the inability to handle data that is collected at different times for different experiments. In other words, the set of times at which data is collected must be identical for every experiment.

$$\boldsymbol{\tau}_{exp1} = \{\tau_1, \tau_2, \ldots \tau_n\} \qquad \textit{Equation 10}$$

$$\boldsymbol{\tau}_{exp1} = \boldsymbol{\tau}_{exp2} = \cdots = \boldsymbol{\tau}_{expm} \qquad \textit{Equation 11}$$

Missing Data

The third and final experimental design problem is inability to operate when specific observations are missing from the data set. Functionally, this problem is identical to the previous two. The $\boldsymbol{\tau}_{\text{exp}}$ will not be equivalent for every experiment if any of these three criteria exist in the dataset. In practice, these three problems pose separate restrictions on data collection. The first two restrict the freedom of the researcher to implement varying data collection strategies. The third imposes that missing a required observation will result in an error. These restrictions create the need for a rigid data collection strategy that does not allow room for mistakes. Overcoming these restrictions will require significant attention to be given to the regression, but would vastly improve the versatility of the methodology. The data restrictions are not the focus of the improvements made in this research.

Oscillatory Behavior

The simulated kinetic system features eight independent reactions; correctly identifying all reaction stoichiometries requires existence of eight significant singular values. Additionally, accuracy of the singular values is vital to stoichiometry prediction and estimation of rate constants, the primary advantages of the DRSM methodology. Improvements made on the singular values themselves would lead to improvements on the number of significant singular values, and therefore improve the downstream calculations estimating stoichiometries and rate constants. The identified weakness surrounding these downstream calculations, the false positives in reaction stoichiometries, were believed to be caused by error in the (nE * nT) x (nS) species rate of appearance/disappearance matrix, RS.

The matrix RS is calculated using an altered dynamic response model, based on the estimated gammas and the derivatives of the Shifted Legendre Polynomials (dSLP). This DRSM now models the change in concentration of a species over time.

$$P_1(\tau) = 2\tau - 1 \qquad P_2(\tau) = 6\tau^2 - 6\tau + 1 \qquad \textit{Equation 12}$$

$$\frac{dP_1(\tau)}{d\tau} = 2 \qquad \frac{dP_2(\tau)}{d\tau} = 12\tau - 6 \qquad \textit{Equation 13}$$

Thus, the RS matrix is extremely sensitive to the oscillatory end behavior of the DRSMs because this behavior translates to alternating positive and negative values in RS, reducing the reliability of singular value decomposition of RS. The species concentration being modelled is alternating between increasing and decreasing, particularly at experimental times when the species concentration is approaching zero. This is most evident when observing the DRSM for species 3.

This trend is caused by the DRSM overfitting simulated error when species concentration is near or approaching zero. Simulated data in these areas tends to, itself, oscillate due to no

longer having an even error distribution. The normal error distribution applied to the simulated data cannot create a negative concentration; negative concentration values are set to zero. Therefore, when concentrations are low enough that the error distribution would overlap into negative values, it is no longer functionally a normal distribution. The data will be skewed away from zero. A simple but easily understood example of this trend would be in filling a glass with water. A sufficiently large glass would have a normal error distribution around the target fill line. Imagine now that one is to fill the glass to only a depth of 1mm; one is much more likely to overfill than to underfill, so the error distribution is not centered at the fill line. The DRSM is modelling error when it comes to end behavior, most evident in species 3, and is therefore overfitting.

The overfitting in this case should not be confused with overfitting due to using too many SLP. The number of polynomials to use was determined via leave-one-out cross validation, which prevents overfitting by the polynomials. Instead, this overfitting was believed to be caused by the gammas in the model, the coefficients of the SLP. Stepwise regression needed to be improved upon to restrain the effect of the gammas on the end behavior.

Regularization

A mathematical approach especially used in machine learning algorithms that combats overfitting is regularization, which functions by adding a penalty or weight onto the least squares error minimization. This allows one to control the complexity of the model by choosing the regularization coefficient, $\lambda > 0$. The minimization of a regularized error function will yield smaller coefficients than linear regression; the exact effect of the regularization depends on the chosen λ and the chosen norm for regularization. The limit of regularized regression as λ goes to

zero is linear regression. The general form of the regularized error function is shown below (Bishop, 2006).

$$\min_{\gamma} \left(\frac{1}{N} \|y - X\gamma\|_2^2 + \lambda \|\gamma\|_p^p \right)$$

Equation 14

Two norms were evaluated in the application of regularization to the Pfizer data: the L_1 norm and the L_2 norm.

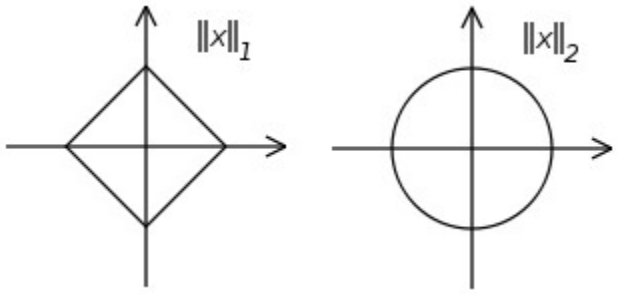


Figure 8: Unit Circles of L1 and L2 Norms
http://simonstechblog.blogspot.com/2013/04/transforming-points-in-unit-square-with.html

The L_2 norm is the Euclidean norm and has the unit circle most are accustomed to from trigonometry. Using the L_2 norm for regularization is called ridge regression or Tikhonov regularization. The L_1 norm's unit circle is a square with corners located on the axes. Utilizing the L_1 norm for regularization is called lasso (least absolute shrinkage and selection operator) (Bishop).

LASSO

Least absolute shrinkage and selection operator is the L_1 regularizer solving:

$$\min_{\gamma} \left(\frac{1}{N} \|y - X\gamma\|_2^2 + \lambda \|\gamma\|_1 \right)$$

Equation 15

The L_1 regularizer tends to shrink regularized coefficients, γ_{ij}, to zero. The tendency of lasso to shrink some γ_{ij} to zero can be understood by looking again at the L_1 unit circle. The regularization term has contours tracing a cross-polytope. This shape is best visualized in lower dimensions, but can be generalized as the convex object with vertices on each axis. Each dimension can be thought of as a γ_{ij}, with γ^* being a point in the space with coordinates describing the minimized gammas. Performing minimization with a constraint region of this shape will lead to γ^* often being located at a vertex of the constraint region. With γ^* on some axes, the model will be sparse and have set some γ_{ij} to zero.

The tendency of LASSO to set some gammas to zero when applying L_1 regularization is desirable. This creates a simpler model that more clearly identifies which factors play important roles in the model. Functionally, LASSO is the form of regularization most like stepwise regression because both will eliminate unneeded gammas from the model, but differ in the selection process. These two regression techniques use different target functions to determine which gammas are significant in the model.

Selection of the regularization coefficient is vital to creating an accurate model via LASSO, but is a delicate process. As stated, a λ of zero will lead to a linear regression based model. As λ is increased, fewer gammas will survive the regularization. If too many coefficients

are eliminated from the model, the prediction will be poor. Cross-validation error is a useful meter by which one can decide on an effective λ.

Results

New models were created to test the effectiveness of LASSO as a means of reducing oscillatory end behavior in the DRSM. As previously mentioned, oscillatory behavior was found primarily in the predictions for species 3 and species 5; species 3 will be looked at exclusively, being the largest source of oscillatory behavior and therefore the largest window for improvement in the methodology.

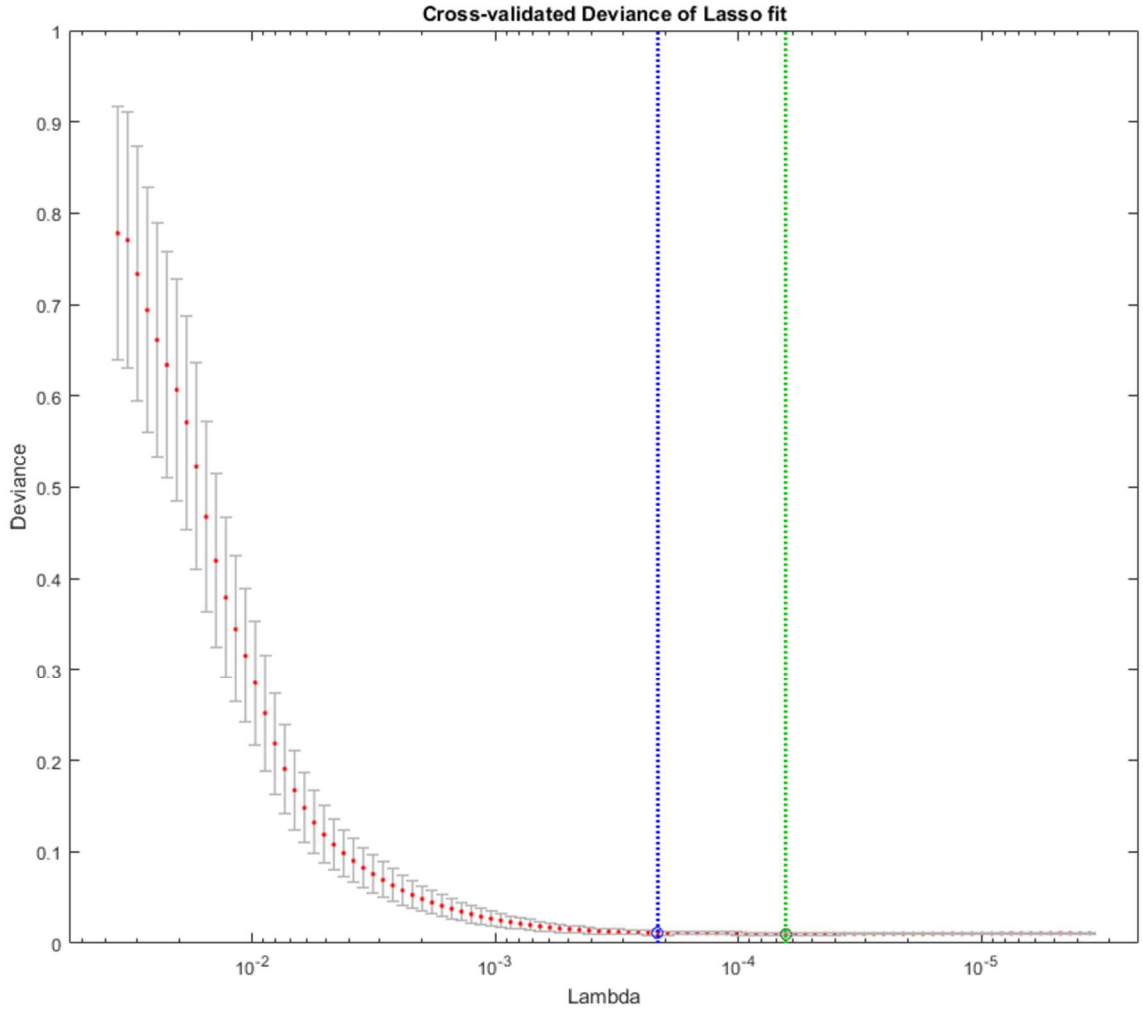
Introducing regularization into the methodology also introduced a second knob by which the model was controlled, the λ. To isolate the impact of λ on the model performance, the number of polynomials used will be kept constant and equal to the best number of polynomials as determined by both lack of fit (Georgakis, 2016) and cross-validation (Santos-Marques, 2016), which for species 3 is nine polynomials. Future work on this methodology should seek to vary both R and λ to explore a wider range of possible models, but the focus here is to demonstrate the efficacy and value of regularization as applied to the DRSM methodology.

To determine the best λ, a plot was created comparing the cross-validated deviance, theta, and the respective λ, shown below. Deviance shows the absolute error between the created model and the data used to create the model. This is done via cross-validation, a statistical technique by which some of the data is reserved for a testing set, while the rest, the training set, is used to create the model.

It is important to note that λ is increasing in value from right to left, and generally, as λ increases, the deviance increases. The cross-validated deviance is a normalized statistic that quantifies the error in the model's prediction. This deviance reaches a maximum when λ reaches

its highest value because as λ increases, fewer gammas are kept in the model. Elimination of too many gammas (which are responsible for curvature in the model) will lead to a linear prediction, which has poor performance. The opposite extreme on Figure 9, as λ shrinks toward zero, shows the limit of the deviance as linear regression is approached. The green line marks the λ for which a minimum deviance occurs, while the blue line indicates the largest λ value whose deviance falls within the bounds of error for the green demarcated value.

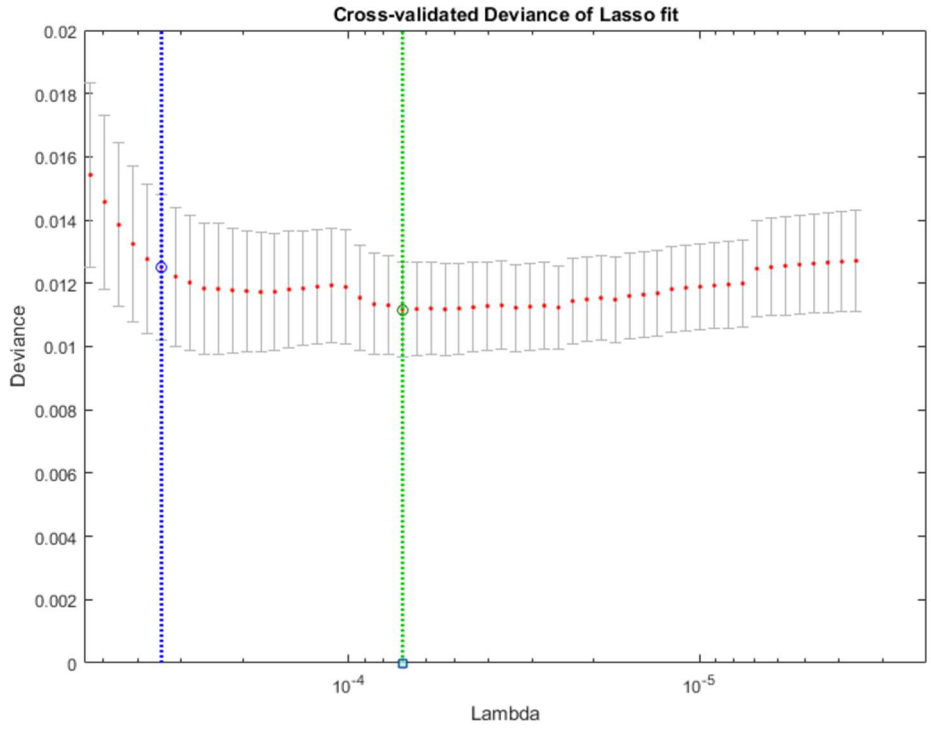
Figure 9:Cross-Validation Deviance vs Lambda for Species 3 DRSM



An interesting conclusion drawn from Figure 10 is that LASSO will not yield a significant reduction in model deviance. The magnitude of the deviance for the linear regression

model, which is the limit of this graph as λ goes to zero, is not significantly larger than the deviance of the λ marked in green.

Figure 10: Cross-Validation Deviance vs Lambda for Species 3 DRSM, closeup

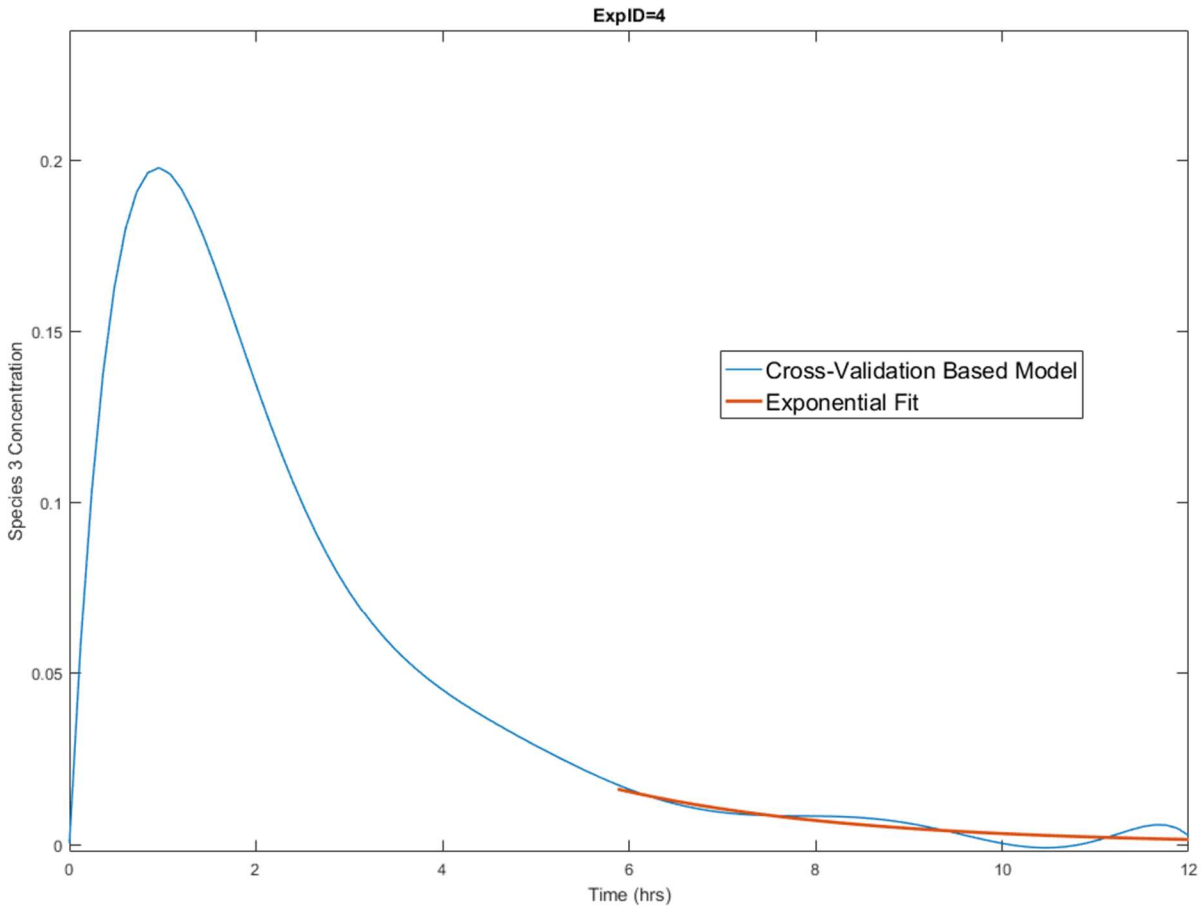


Another significant trend to notice is that Figure 9 appears to be monotonically decreasing when decreasing λ. The minimum in deviance marked by the green line is extremely subtle. Models that are best suited for improvement from LASSO have obvious, more pronounced minimums in deviance that indicate an ideal λ for regularization.

Figure 10 shows minimal improvement in deviance when using LASSO over linear regression, but the existing methodology utilizes stepwise regression, rather than linear regression, so it already improves upon linear regression. Going forward it is necessary to quantify the oscillatory behavior so that objective comparisons can be made and conclusions drawn on the effectiveness of LASSO in the DRSM methodology. An exponential curve was fit

to the DRSM prediction of the cross-validation based methodology (Santos-Marques, 2016) for the second half of the experimental duration, $\tau > 0.5$. This fit represents the possible exponential decay expected for species 3.

Figure 11: Example CV Model (Santos-Marques) and Exponential Fit for Species 3



The oscillatory behavior of each model was quantified by calculating the absolute difference between the model's prediction and this exponential fit for each of the seventeen experimental conditions in the face-centered central composite design. The absolute error was then summed to yield one error value per model. It was then possible to identify the best λ value for regularization based on minimizing the oscillation based error.

Model	Oscillation Error
$\lambda = 2.2 * 10^{-4}$	0.0046
Stepwise Regression	0.0048

Table 4: Oscillation Based Error Comparison for Stepwise Regression (Santos-Marques, 2016) and Best LASSO Regularization

The determined λ that minimized the oscillation based error can be seen in Table 4. This value for λ is close to the indicated blue λ in Figure 9. The improvement of LASSO regularization over stepwise regression was a ~5% reduction in oscillations.

Figure 12: Model Comparison; Stepwise Regression vs Best Determined LASSO Regularization

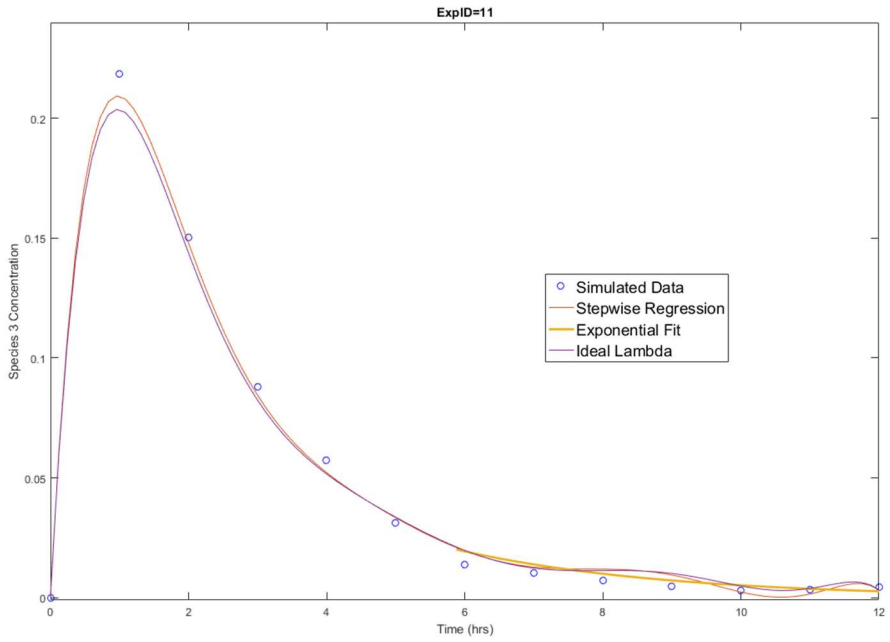
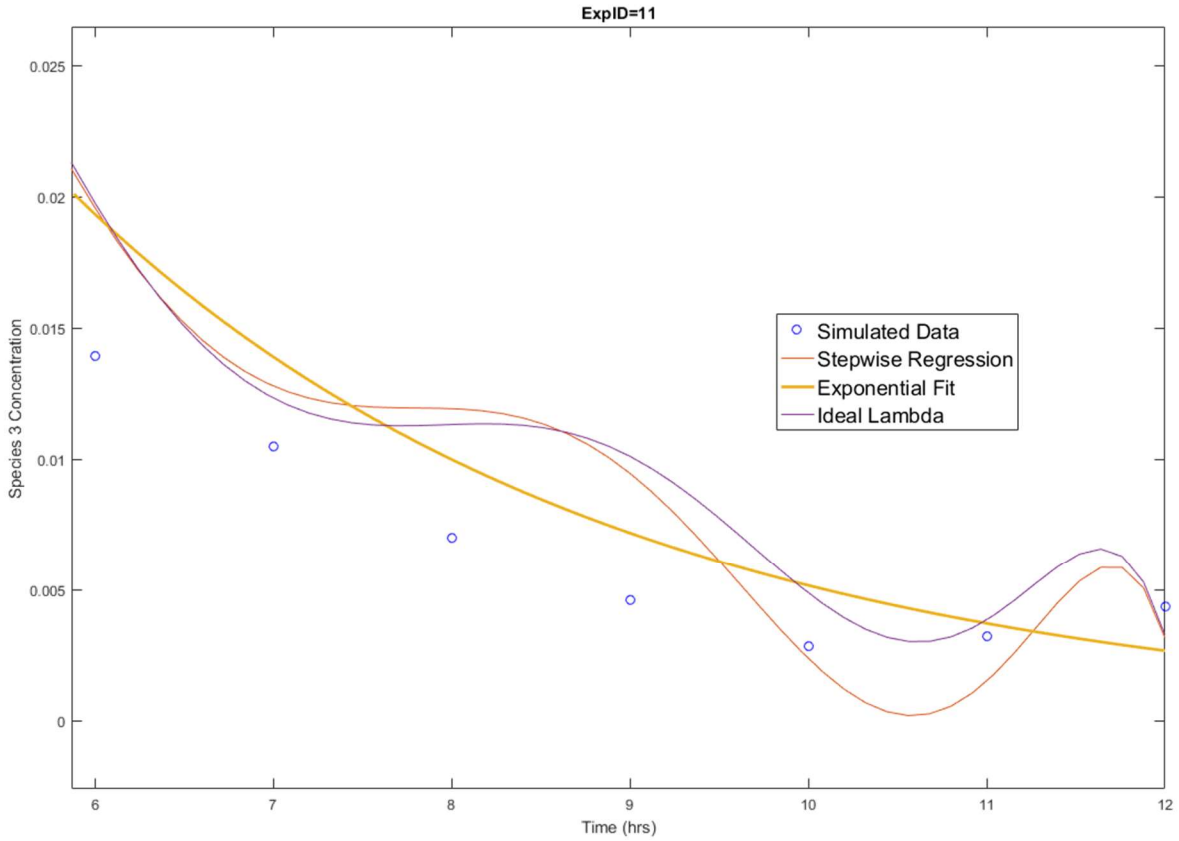


Figure 13: Closeup on Oscillatory Behavior Comparison; Stepwise Regression vs LASSO



As shown in Figure 13 and Table 4, the improvement made by LASSO in reducing oscillatory behavior was minimal. Most importantly, application of LASSO did not eliminate any oscillatory behavior, merely reduced the amplitude of the peaks. This means that the species rate of appearance/disappearance matrix will still suffer from a frequent change in sign for species 3, as modelled concentration oscillates between increasing and decreasing.

Figure 14: Oscillation Based Error Comparison Across Lambda Range

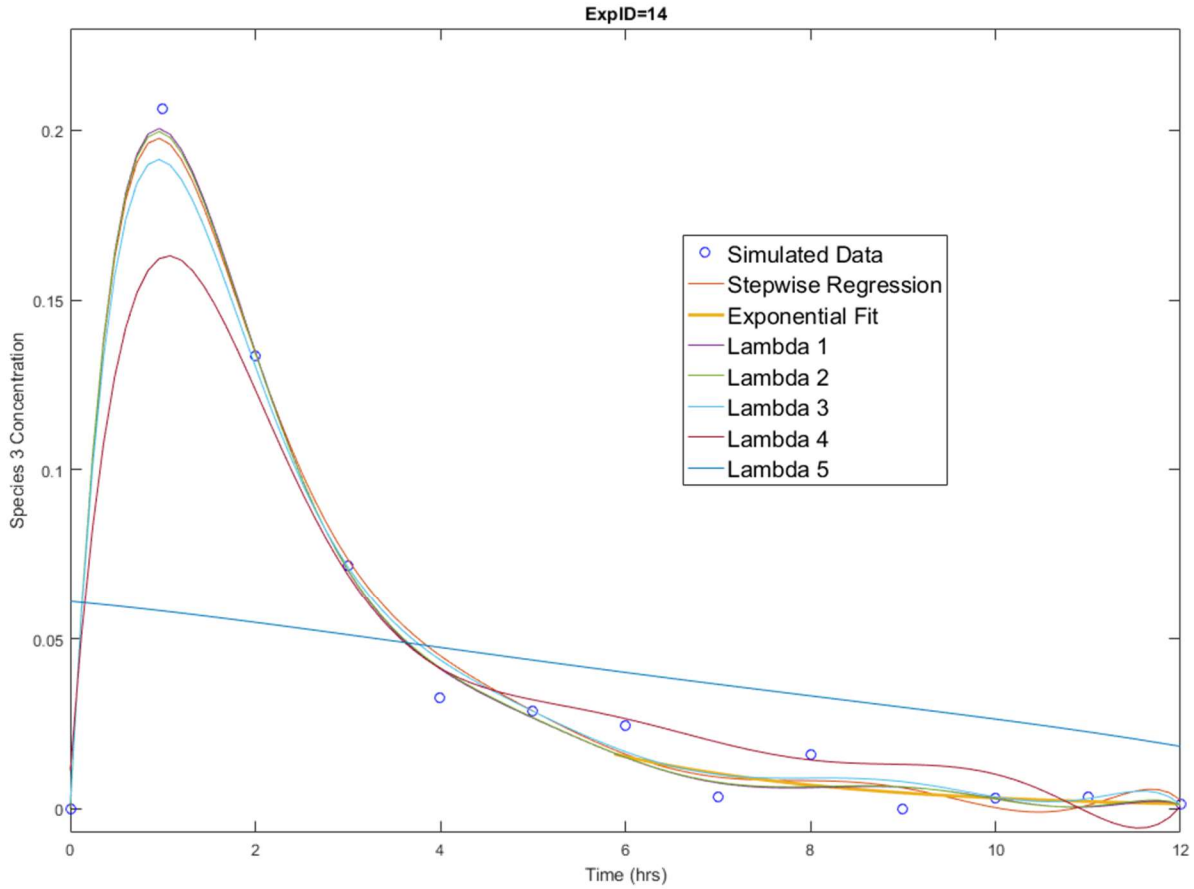
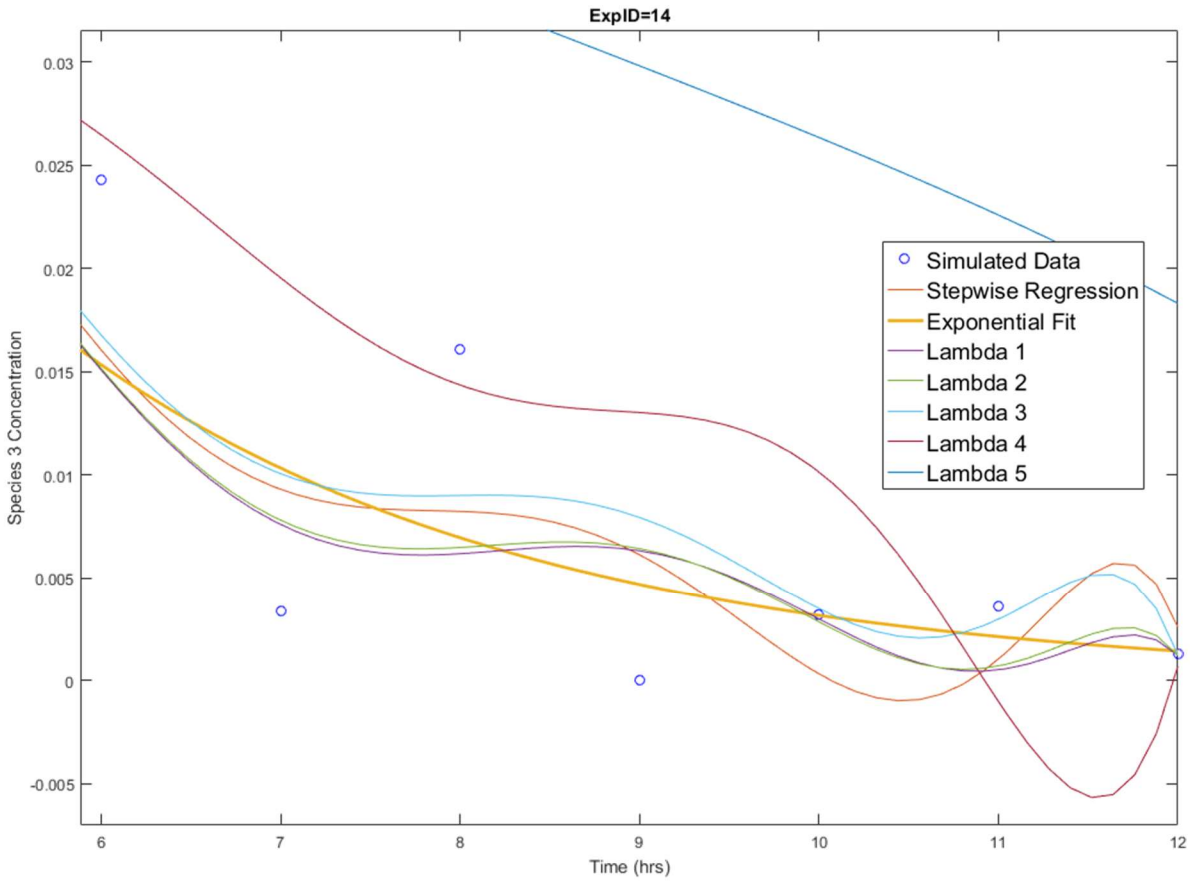


Figure 14 shows a broader range of λ values and their impact on the model for a given set of experimental conditions. The value of λ increases going from Lambda 1 to Lambda 5. The largest λ, Lambda 5, shows primarily a linear model due to too much reduction in model complexity. It can also be seen that as λ increases, the ability to predict the initial peak suffers. Lambda 4 appears to have the least significant oscillations, but its predictive ability is dramatically worse than stepwise regression. The end behavior, while steadier, is far off from the "ideal" exponential fit.

Figure 15: Closeup on Oscillation Comparison for Range of Lambdas



A closer look at the oscillatory behavior as λ is changed further demonstrates that LASSO is largely unable to eliminate oscillatory behavior across a large range of values. LASSO is capable of dampening existing oscillations to a degree. In Figure 15 one can see that Lambdas 1, 2, and 3 appear to exhibit oscillations to a lesser magnitude than the stepwise regression, but contain the same number of local extrema.

Ridge

Ridge regression, also known as Tikhonov regularization is the L_2 regularizer solving:

$$\min_{\gamma}\left(\frac{1}{N}\|y - X\gamma\|_2^2 + \lambda\|\gamma\|_2^2\right)$$

Equation 16

L_2 regularization does not explicitly perform covariate selection. The gammas in the regression will shrink in magnitude in proportion to the regularization coefficient selected, but are much more likely to stabilize at nonzero values, compared to L_1 regularization. Ridge regression can lead to zero gamma values, but this result is not as favored as in L_1. This can be seen when comparing the constraint regions of the regularizers. The cross-polytope shaping of the L_1 constraint region causes the optimization to more often lead to a vertex, on which some γ_{ij} will be zero. The unit circle for L_2 has no such vertices; minimization that causes a gamma to be zero is no more likely than minimization that leads to any reduced gamma value.

Ridge regression's inability to create a selective model (eliminate insignificant gammas) is a weakness when compared to LASSO or stepwise regression. It is desirable to determine causal relationships between the factors and the response. Understanding which factors play largest roles in the DRSM, and the polynomial dependence of those factors with time, provides insight into the kinetic system. Ridge, however, is a valuable tool for reducing error sensitivity in inverse problems in machine learning. Inverse problems, in this case regression, are extremely sensitive to error in the measurements. Thus, these problems are deemed ill-conditioned because the regression results are not stable. Improving the conditioning of the regression will decrease the model's dependence and smoothen oscillatory behavior.

Results

To test ridge regression on the Pfizer dataset, once again the number of polynomials was held at nine, as determined optimal by both lack of fit (Georgakis, 2016) and cross-validation (Santos-Marques, 2016). Species 3 will be the focus for analysis on oscillatory behavior.

Once again, determination of ideal regularization coefficient is key to creating the best possible model. Figure 16 shows the relative change in size of the gammas as the ridge parameter, α, is varied. The ridge parameter is related to the regularization coefficient, λ, originating from an alternative representation of L_2 regularization shown below.

$$\gamma = (X^T X + \alpha^2 I)^{-1} X^T y$$

Equation 17

If the ridge parameter is zero, Equation 17 will reduce back to ordinary least squares regression. The required value for α varies depending on the problem, and as seen in Figure 16 for species 3, it can become much larger than the utilized regularization coefficient, λ, in LASSO. This is a result of α's dependence on N, the number of data points utilized in the regression, which develops from the simplification of Equation 16 to Equation 17.

Figure 16 shows the change in standardized magnitude for the 90 gammas calculated via ridge regression for varying α. The general trend, which is expected and desired for ridge regression, is that the gammas are being reduced in magnitude as α increases. This can be seen much more clearly in Figure 17, in which only a handful of gammas are shown to better illustrate the change in magnitude as α increases.

Figure 16: Gamma Shrinkage vs Magnitude of Ridge Parameter for Species 3

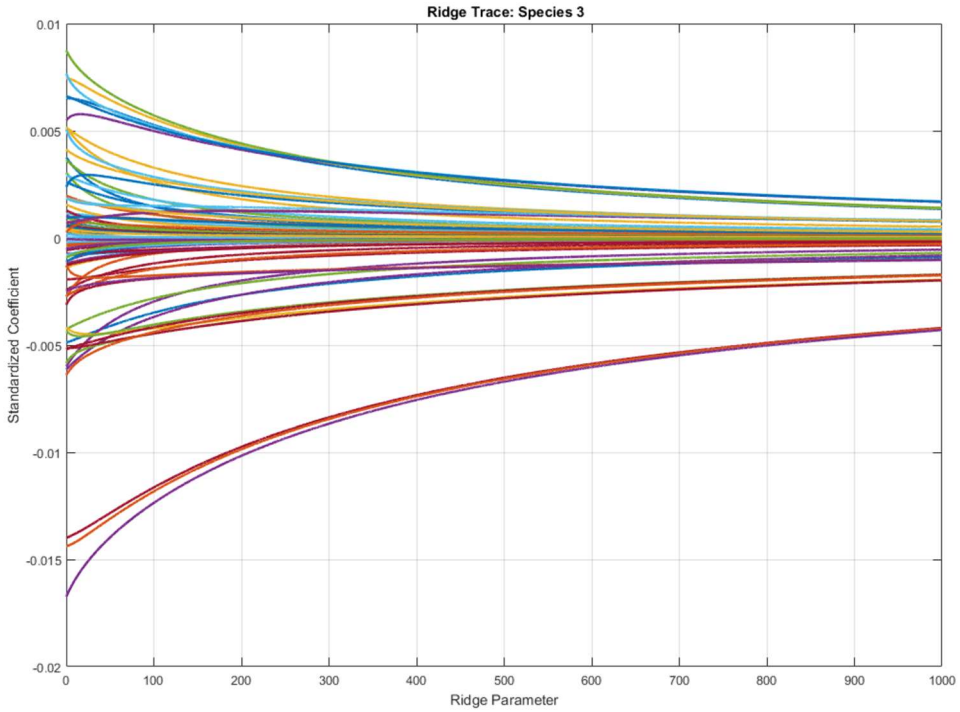
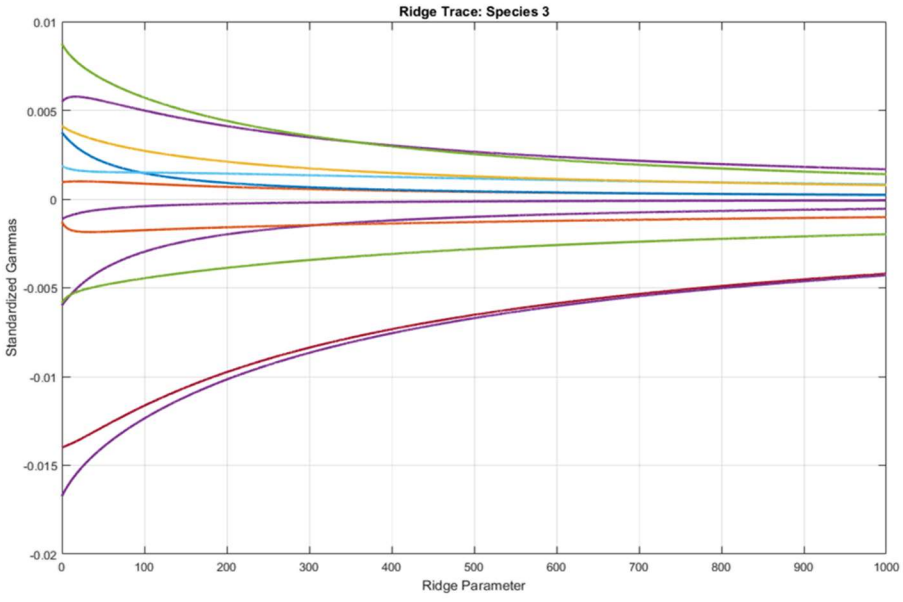


Figure 17: Selected Standardized Gammas vs Ridge Parameter for Species 3



Unlike λ of L_1 regularization, α can be increased indefinitely and still yield a model. In L_1 regularization, too large of a regularization coefficient eliminates all terms from the model. In L_2 regularization, the gammas will shrink in magnitude towards a steady-state, so a larger range of ridge parameters was analyzed then shown in Figure 16 and Figure 17.

Figure 18: Range of Ridge Parameters Considered for Regularization

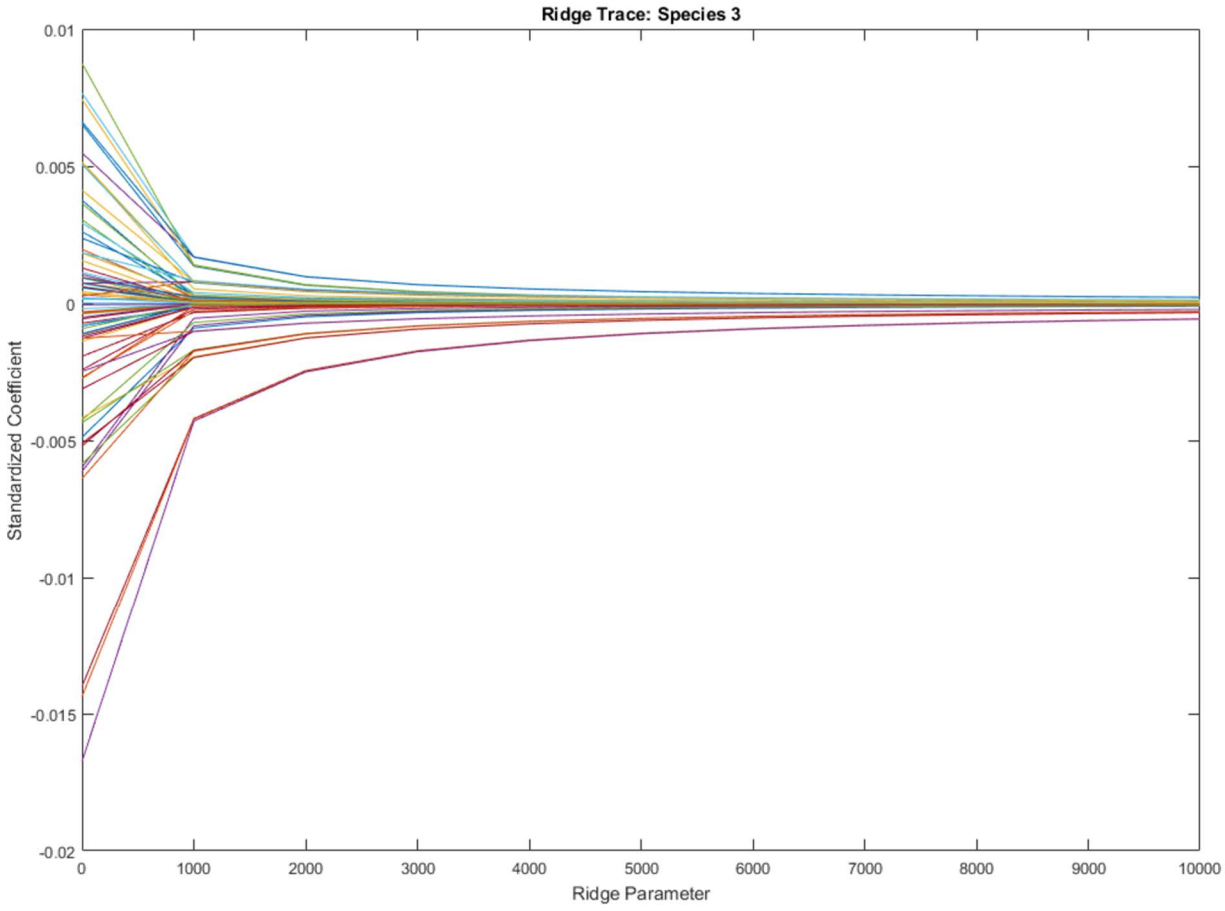


Figure 18 better shows the evolution of the gammas as they approach a steady-state value. Multiples of 1000, up to and including 10000, will be considered for the ridge parameter. It is also for this reason that Figure 18 lacks the curvature in the 0-1000 range as shown in Figure 16.

Figure 19: Ridge Regression Models, Ridge Parameters as Compared to Stepwise Regression

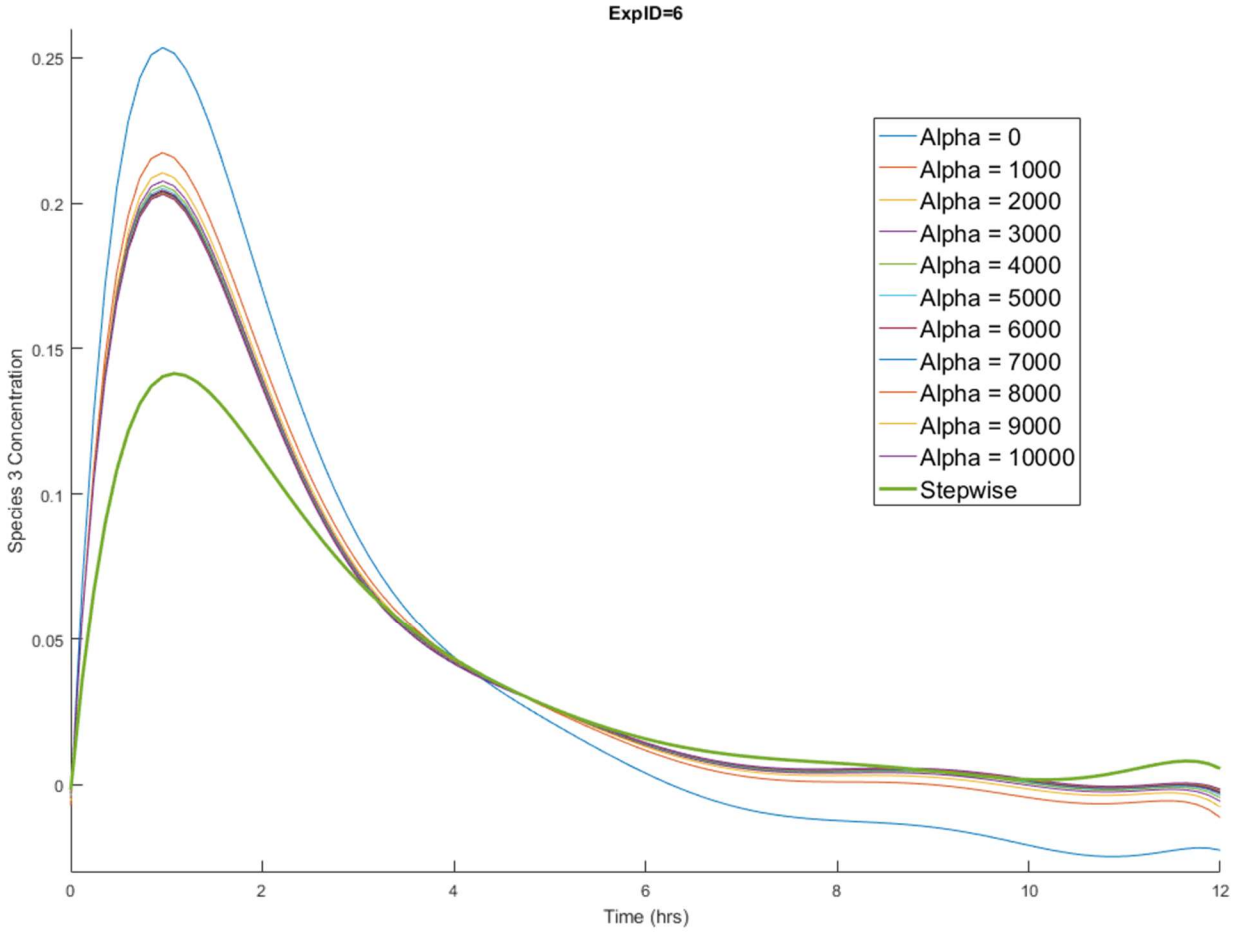


Figure 19 shows the impact of the ridge parameter for a given set of experimental conditions and exemplifies the general trends across other experimental conditions. First, looking at the first half of the experiments, one can see that ridge regression is overshooting the maximum concentration, especially at low ridge parameter values. Looking at the latter half of the experiment, one can see that the ridge based model will undershoot the stepwise model at insufficient α values. Generally, as α increases the ridge based model is pulled in closer to the stepwise regression model. This corrective behavior can be seen to have its limits. There is virtually no difference between the models with $\alpha = 4000$ and $\alpha = 10000$.

In a method, similar to that done for LASSO, the ability of ridge regression to reduce oscillatory behavior will be quantified by comparing the difference between the model and an exponential fit of the model. These results are shown in Table 5. A ridge parameter value of 4000 is deemed sufficient, past which there is no apparent significant improvement. Table 5: Oscillation Based Error vs Ridge Parameter

Regression Method	**Oscillation Based Error**
Stepwise Regression	0.0048
Alpha = 0	0.0371
Alpha = 1000	0.0097
Alpha = 2000	0.0051
Alpha = 3000	0.0026
Alpha = 4000	0.0022
Alpha = 5000	0.0021
Alpha = 6000	0.0020
Alpha = 7000	0.0020
Alpha = 8000	0.0020
Alpha = 9000	0.0020
Alpha = 10000	0.0020

Table 5: Oscillation Based Error vs Ridge Parameter

Impact on Singular Values

Stepwise	LASSO ($\lambda=2.2*10^{-4}$)	Ridge ($\alpha=4000$)
1285.1	1283.9	1285.9
448.8	443.9	453.9
133.3	133.5	134.9
85.1	83.9	92.4
41.1	40.1	43.6
18.5	17.8	18.8
11.8	11.8	11.3
9.8	9.7	9.4
5.5	5.5	5.5
2.8	2.8	2.8

Table 6: Comparison of Singular Values for Best Case Regularization of Species 3

The rate of species appearance/disappearance matrix, RS, was calculated using stepwise regression or regularization for species 3, the primary species exhibiting oscillatory behavior, to analysis what benefits regularization may have on the downstream calculations of the DRSM methodology. Table 6 shows the comparison for obtained singular values based on regression strategy for species 3. The DRSMs for the other nine species were calculated via stepwise regression for all three cases to isolate the impact of improving the behavior of species 3's DRSM.

RXN #	Stepwise	LASSO ($\lambda=2.2*10^{-4}$)	Ridge ($\alpha=4000$)
1	99.929	99.928	99.925
2	99.929	99.928	99.925
3	99.779	99.781	99.773
4	99.677	99.679	99.673
5	99.518	99.519	99.519
6	99.518	99.519	99.519
7	92.695	92.698	92.689
8	84.588	84.586	84.586
9	57.716	57.715	57.777
10	52.045	52.049	52.099

Table 7: Stoichiometry Projection Scores for Ten True Reactions as shown in Table 1

Table 7 shows the projection scores for the ten true reaction stoichiometries defining the chemical system as listed in Table 1. As was the case for singular values, the projection scores did not vary significantly when implementing regularization on species 3. Reduction of oscillatory behavior did not improve the scores of true reaction stoichiometries as expected. Additionally, Table 8 shows that regularization did not lower the projection scores for an example set of "false positive" stoichiometries.

RXN #	Stepwise	LASSO (λ=2.2*10^{-4})	Ridge (α=4000)
1	99.590	99.591	99.609
2	98.895	98.899	98.864
3	99.465	99.466	99.448
4	98.872	98.876	98.870
5	99.137	99.140	99.117
6	98.820	98.827	98.756
7	99.259	99.263	99.210
8	97.385	97.396	97.367
9	98.951	98.959	98.865
10	98.619	98.625	98.573
11	99.664	99.667	99.650
12	97.700	97.708	97.719
13	98.981	98.988	98.906

Table 8: Comparison of Example False Positives Based on Regression Technique

The thirteen reaction stoichiometries tested in Table 8 are shown in Table 9. These stoichiometries are each linearly independent from the true set of occurring reactions. These false positives are now shown to be unrelated to oscillatory behavior in the model.

RXN #	Species Stoichiometric Coefficients									
1	0	-1	0	-1	0	-1	1	0	0	0
2	-1	-1	1	-1	1	-1	0	0	0	0
3	-1	-1	1	0	1	-1	0	0	0	0
4	0	1	1	-1	0	0	-1	0	0	0
5	1	1	0	-1	0	-1	0	0	0	0
6	-1	-1	0	-1	1	-1	0	0	0	0
7	-1	-1	-1	1	0	1	0	0	0	0
8	0	-1	0	-1	0	1	0	0	0	0
9	-1	-1	-1	-1	1	-1	0	0	0	0
10	-1	0	1	-1	1	-1	-1	0	0	0
11	0	-1	1	-1	0	-1	1	0	0	0
12	0	1	-1	1	0	-1	0	0	0	0
13	0	0	-1	-1	0	0	0	0	0	0

Table 9: Example "False Positive" Stoichiometries Linearly Independent from True Set

Significant Singular Values

Determining which singular values, as shown in Table 6, are significant is an important step in the methodology that did not have an objective function behind it. Both previous strategies had a subjective determination of significance, as seen in Figure 2 and Figure 3. Presenting an objective strategy for determining significant singular objectives is the second goal of this research in improving the DRSM methodology.

As mentioned in the introduction, establishing the number of significant singular values plays an important role in the stoichiometry projection. The number of significant singular values indicates the number of independent chemical reactions occurring in the system that the collected data with its associated uncertainties can reveal. Without determining the number of significant singular values, and especially because of false positives, one cannot tell when to stop projecting more stoichiometries.

In this case, with ten observed species, there could be as many as nine independent reactions occurring, but no more than nine. The maximum number of independent reactions occurring will always be one less than the number of species observed. This is because nS reaction coefficients can only be arranged into, at most, nS linearly independent vectors of reaction stoichiometries. Additionally, every reaction must have at least two nonzero coefficients, therefore indicating at least one free variable. This reduces the maximum number of linearly independent reaction stoichiometries from nS to (nS-1).

The number of significant singular values indicates the dimensionality of the subspace whose basis is the linearly independent set of true reactions. Stoichiometries are projected into this subspace to gage how well they fit the system. True stoichiometries will

already exist in this subspace, and should hypothetically score 100%. Stoichiometries linearly *dependent* on the true reaction stoichiometries also exist in this subspace and should score 100%.

Propagating Error

Regression methods by which the DRSM's gammas are calculated have associated estimated error. Because there is a confidence interval associated with each model parameter estimated, there will be a prediction interval associated with the DRSM output. The gamma confidence interval was not considered when calculating the RS matrix. It was hypothesized that by propagating the gammas' uncertainty to the singular value calculation, one could obtain insight on which singular values are significant. Singular values cannot be negative, but it was thought that insignificant singular values may have an error range nearly incorporating zero.

To test this hypothesis, a simpler reaction system was considered for computational speed. This system featured only three reactions and five species. The gammas were normally distributed individually, with the mean being their respective calculated value and the standard deviation of their respective error. This was done 20,000 times, and the results are shown below.

Singular Values, No Error	**Singular Values, w/ Error**	**Standard Deviation**
53.59	53.56	1.42
9.57	9.53	0.27
3.72	3.73	0.24
1.62	1.59	0.10
1.00	0.99	0.04

Table 10: Singular Value Comparison when Propagating Error for Simpler Reaction System

As shown in Table 10, propagating the error from the gamma calculation had no significant effect on the singular values. The determined values showed narrow error bars for both significant and insignificant singular values.

IND Function

The factor indicator function is an empirical function suggested by Malinowski for its ability to "deduce dimensionality of the factor space" (Malinowski, 1977). This function was presented to determine significant eigenvalues, but here it is applied to singular values of RS, which are the square roots of the eigenvalues of $(RS)^T(RS)$. The dimensionality of the factor space is synonymous with the number of significant singular values of RS.

$$RE = \left(\frac{\sum_{j=n+1}^{j=c} \sigma_j^2}{r(c-n)} \right)^{\frac{1}{2}}$$
$$IND = \frac{RE}{(c-n)^2}$$

Equation 18: Factor Indicator Function (Malinowski, 1977)

The real error, RE, is a function of the data matrix's rows (r) and columns (c), and the singular values (σ). A number, n, of singular values to be "significant" is selected; the rest of the singular values, (n+1) to c, are incorporated into the real error, the difference between pure data and what was observed experimentally. Here, it is assumed that there are more rows than columns in the data matrix, hence the singular values being summed until the c^{th} value. This assumption is true for RS when using the Pfizer dataset.

The IND function will naturally minimize when n is equal to the number of significant singular values. Therefore, one can simply evaluate the function at all possible values of n and determine the number of significant singular values (Malinowski, 1977).

# Significant Singular Values	Stepwise	LASSO (λ=2.2*10^{-4})	Ridge (α=4000)
1	0.00537	0.00531	0.00544
2	0.00249	0.00248	0.00258
3	0.00205	0.00202	0.00221
4	0.00148	0.00145	0.00155
5	0.00121	0.00118	0.00120
6	0.00141	0.00140	0.00136
7	0.00203	0.00202	0.00196
8	0.00297	0.00297	0.00297
9	0.00770	0.00770	0.00770

Table 11: IND Function Value Comparison Between Regression Techniques

As shown in Table 11, the IND function is minimized when five singular values are significant, regardless of regression technique. This is less than expected based on knowledge of the chemical system. With eight total independent reactions, one would ideally expect eight singular values to be significant; error in the data will interfere with this, however. Interestingly, all three regression techniques fail to identify reactions 9 and 10, as seen in Table 7. These two reactions incorporate minor species that are largely obfuscated by introduced error. Perhaps the inability to accurately project two of the linearly independent true reactions is related this undershooting of the number of significant singular values.

Malinowski F-Test

The factor indicator test (IND function) relies on an empirical function of no statistical significance. Attention was given to search for a statistically significant method by which significant singular values can be ascertained. Malinowski first adapted the f-test for application in singular value decomposition in 1988.

If the error in the data follows a normal distribution, then so will the variance of the insignificant singular values. This is because insignificant singular values no longer represent meaningful information on the system, only error. It is possible then to compare the variance of a singular value to the pooled variance of the smaller singular values. If the two variances are statistically different, then the tested singular value is significant. A two-factor f-test will be applied to pass judgement on the respective variances (Malinowski, 1988).

$$F(1, c-n) = \frac{\sigma'^{2}_{n}}{\sum_{j=n+1}^{j=c} \sigma'^{2}_{j}}(c-n)$$

$$\sigma'^{2}_{n} = \frac{\sigma^{2}_{n}}{(r-n+1)(c-n+1)}$$

$$H_0: \sigma'^{2}_{n} = \sigma'^{2}_{pool}$$

$$H_a: \sigma'^{2}_{n} > \sigma'^{2}_{pool}$$

Equation 19

This f-test tests the null hypothesis that the variance of the nth reduced singular value is equal to the variance of the pooled reduced singular values. Reduction of the singular values is done by the large left fraction in Equation 19. This serves to magnify the significant singular values, while error singular values should have, statistically, the same reduced singular value (Malinowski, 1988).

N^{th} Singular Value	Stepwise	LASSO ($\lambda=2.2*10^{-4}$)	Ridge ($\alpha=4000$)
1	1.000	1.000	1.000
2	1.000	1.000	1.000
3	1.000	1.000	1.000
4	1.000	1.000	1.000
5	0.998	0.998	0.998
6	0.966	0.962	0.971
7	0.880	0.881	0.878
8	0.847	0.845	0.836
9	0.599	0.599	0.599

Table 12: Probability from F Cumulative Distribution Function for Varying Singular Values and Regression Techniques

Table 12 above shows the respective probabilities that each singular value is significant based on the f-test in Equation 19. Typically, a 95% confidence level is used; the smallest singular value above this confidence level is denoted with an underline.

The Malinowski F-Test determines that for each regression technique and with a 95% confidence level that six singular values are significant. Like the results of the IND function, this is lower than the expected number of eight significant singular values. Once again, this is likely related to the inability to correctly identify reactions 9 and 10. These reactions may be too minor to be statistically significant in the reaction system. However, it is also possible that inaccuracies arise from the sheer size of RS affecting the distribution (Malinowski, 2004).

Conclusion

The goal of this research was to suggest advancements in the DRSM methodology that tackle two problems: oscillatory end behavior and determination of significant singular values.

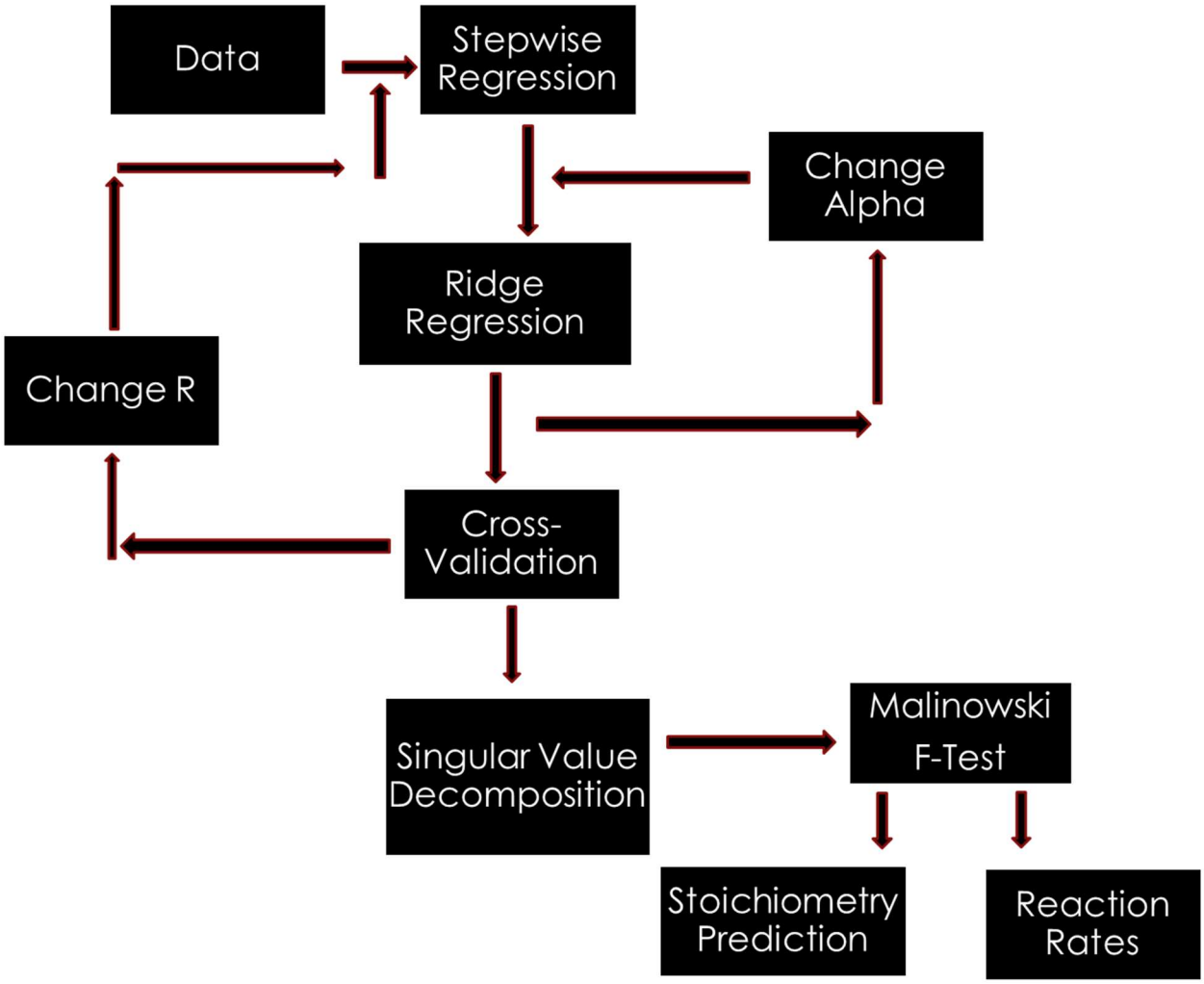
Regularization, particularly ridge regression, was shown to be able to greatly smoothen the model for species 3. Oscillation based error for species 3 was reduced by over 50% when implementing ridge regression with a large enough ridge coefficient. This reduction in oscillations did not lead to improvements in downstream weaknesses, such as stoichiometry projection, as was hoped.

Two potential methods were identified as functioning, objective strategies for determination of significant singular values; one of which is empirical in nature (IND function) while the other is statistically significant (f-test). Both methods determined fewer significant singular values than would be expected ideally for a reaction system with eight independent reactions. The Malinowski f-test is suggested as the better method to be implemented in the DRSM methodology for two reasons. First, the statistical basis for this technique is vital in such a mathematically driven modelling methodology. Second, the determination of six singular values via this f-test introduces peculiar implications on the inability to identify reactions 9 and 10 as shown in Table 1. The stoichiometry projection identified only six of the eight true independent reactions. It is possible that the two unidentified reactions are too minor to be considered by the methodology.

The new flowchart for DRSM methodology and subsequent kinetic calculations as established through this research is shown in Figure 20. This new strategy implements a second loop by which the optimal ridge coefficient is determined for a species, which will introduce a

second knob by which the model is controlled and can be tuned. Additionally, it is suggested that the use of stepwise regression in conjunction with ridge regression be explored. This could potentially overcome the performance issues encountered in using solely ridge regression. This increase in complexity is warranted by increasing model stability and reducing oscillatory end behavior.

Figure 20: Proposed Flowchart to Eliminate Two Previously Identified Weaknesses



Key

DRSM = dynamic response surface model

dSLP = shifted Legendre polynomials derived w.r.t. τ

k = number of significant singular values

LoF = lack of fit, the strategy by which the Georgakis iteration of DRSM methodology decided a suitable R for the model

LOO = leave one out, the cross-validation strategy by which the Santos-Marques iteration of DRSM methodology decided a suitable R for the model

N = total data points, nE*nT

nE = number of experiments in design

nS = number of species observed

nT = number of instants data was collected per experiment

p = norm chosen for regularization

P_k = projection matrix using k significant singular values

R = number of polynomials approximating β

RS = (N) x (nS) matrix showing rates of change for each species

SLP = shifted Legendre polynomials

X = matrix of covariates, (N) x (10R), products of design factors x_i and SLP

y = species concentration

α = ridge parameter, determines degree of regularization for ridge regression

β = coefficients of factors in DRSM, functions of τ, represented by weighted sums of SLP

γ = coefficients of SLP in DRSM estimated via regression

λ = regularization coefficient, determines degree of regularization

τ = dimensionless time, tau

References

Bishop, Christopher M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. Print.

Georgakis, C., From Composition Data Towards a Stoichiometric Model of a Pharmaceutical Reaction. *Pfizer Inc. Report.* **2016**

Klebanov, N.; Georgakis, C., Dynamic Response Surface Models: A Novel Data-Driven Approach for Time-Varying Process Outputs. *Ind. Eng. Chem. Res.* **2016 (in press)**

Malinowski, Edmund R. "Determination of the Number of Factors and the Experimental Error in a Data Matrix." *Analytical Chemistry* 49.4 (1977): 612-17. Web.

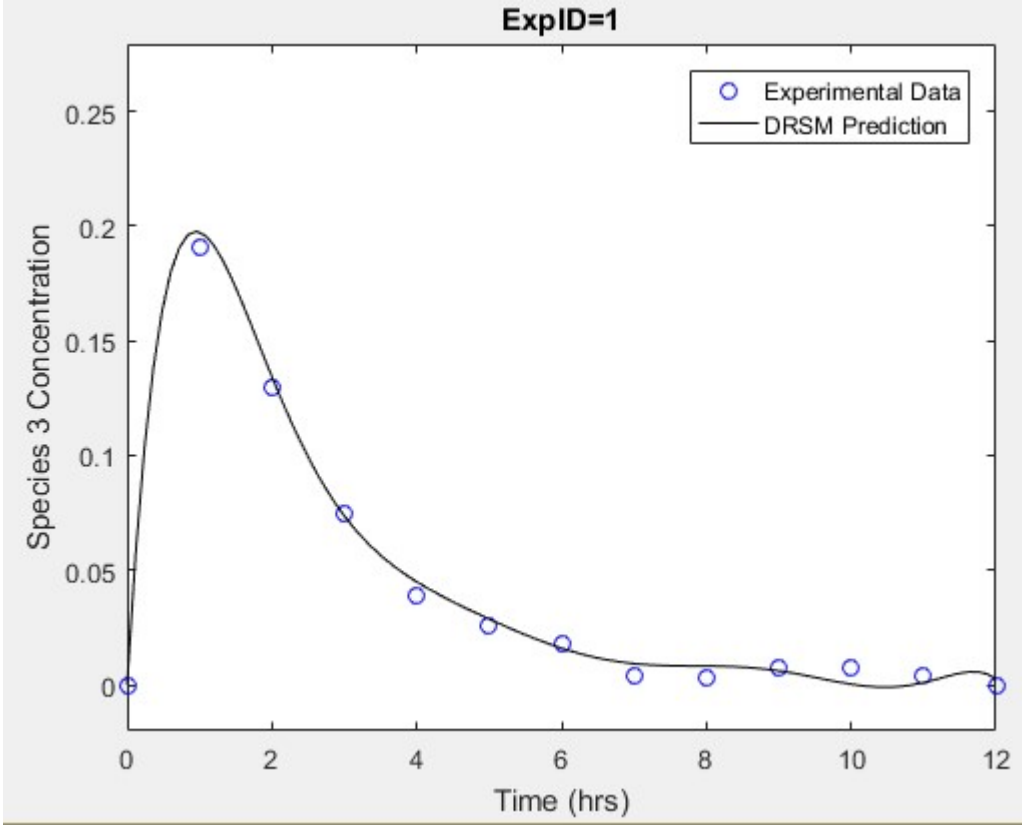
Malinowski, Edmund R. "Statistical F-Tests for Abstract Factor Analysis and Target Testing." *Journal of Chemometrics* 3.1 (1988): 49-60. Web.

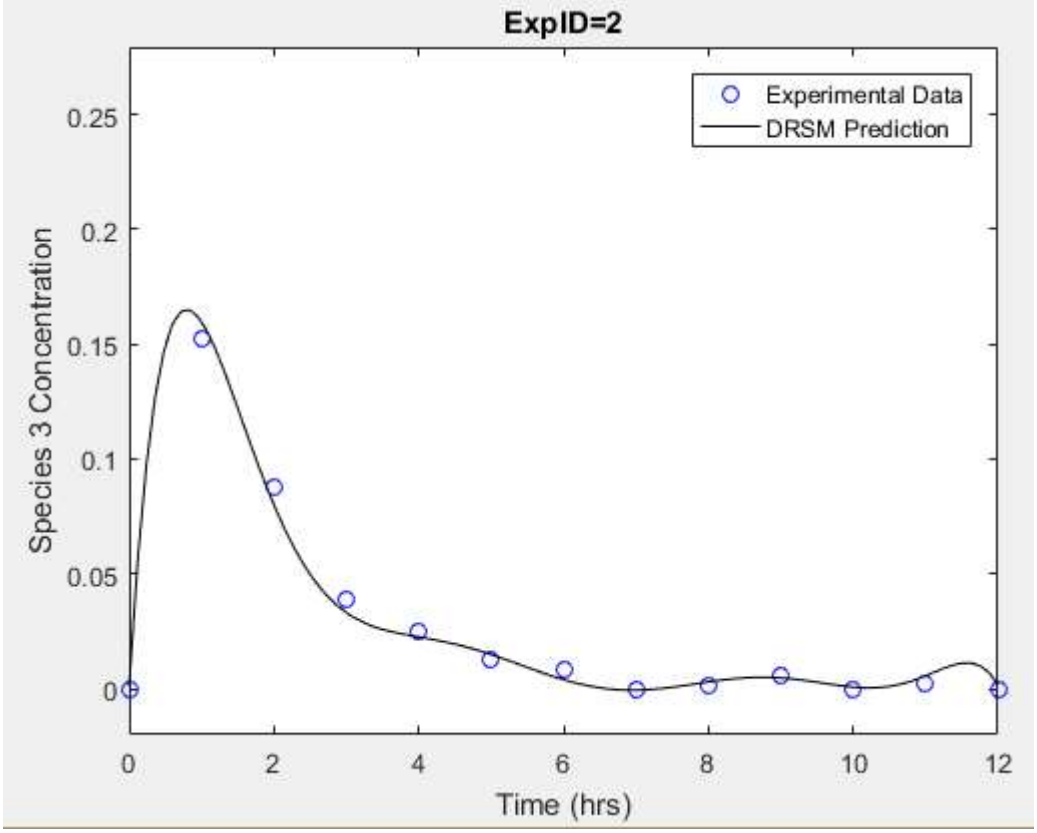
Malinowski, Edmund R. "Adaptation of the Vogt-Mizaikoff F-test to Determine the Number of Principal Factors Responsible for a Data Matrix and Comparison with Other Popular Methods." *Journal of Chemometrics* 18.9 (2004): 387-92. Web.

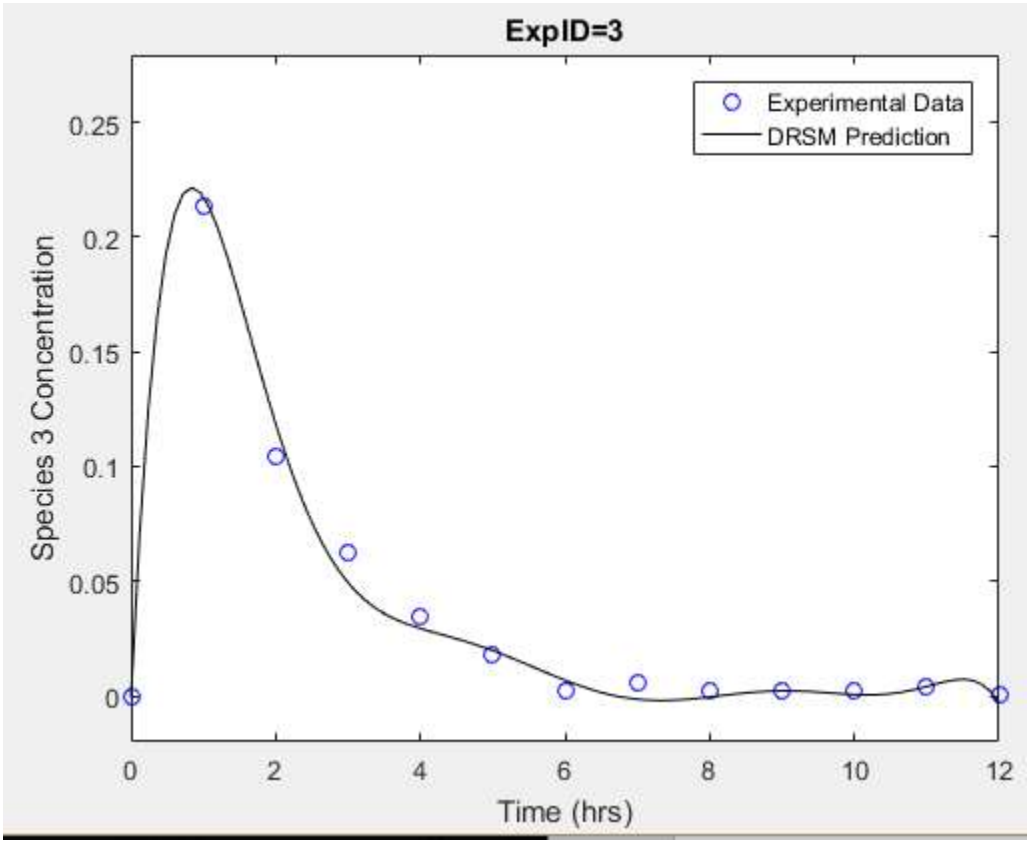
Montgomery, Douglas C. *Design and Analysis of Experiments*. 8th ed. Hoboken, NJ: John Wiley & Sons, 2013. Print.

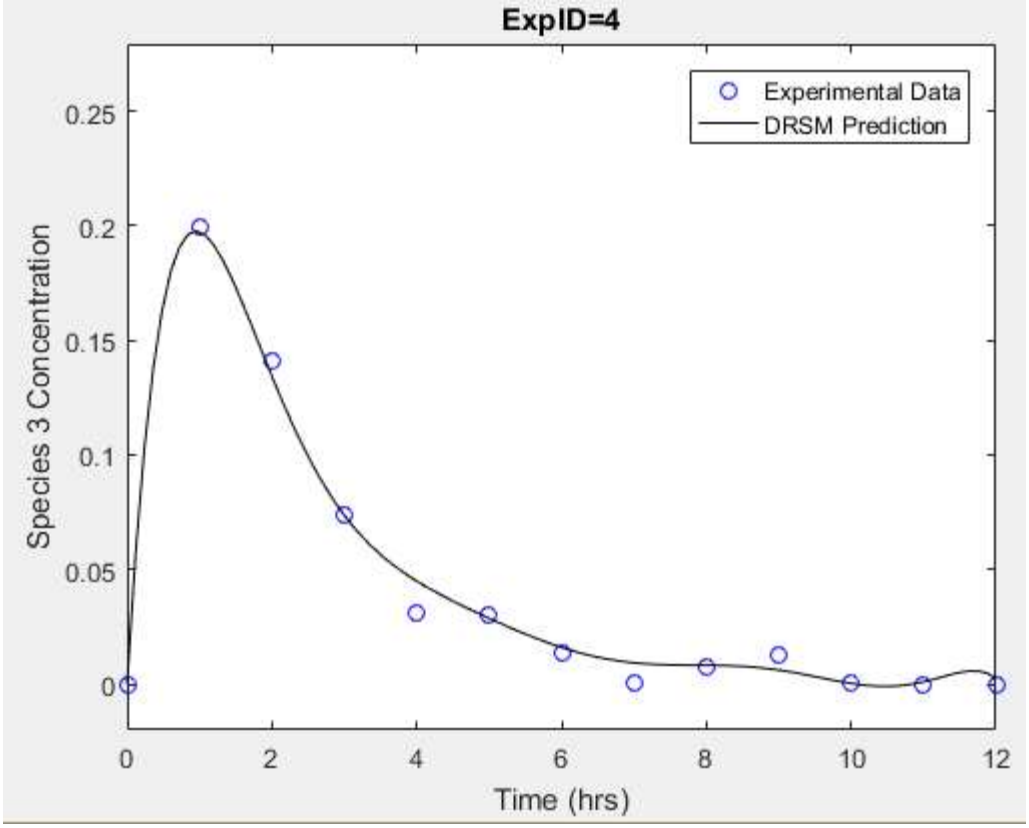
Santos-Marques, J., Exploring and Expanding the Application of Dynamic Response Surface Model Methodology. *Pfizer Inc. Report*. **2016**

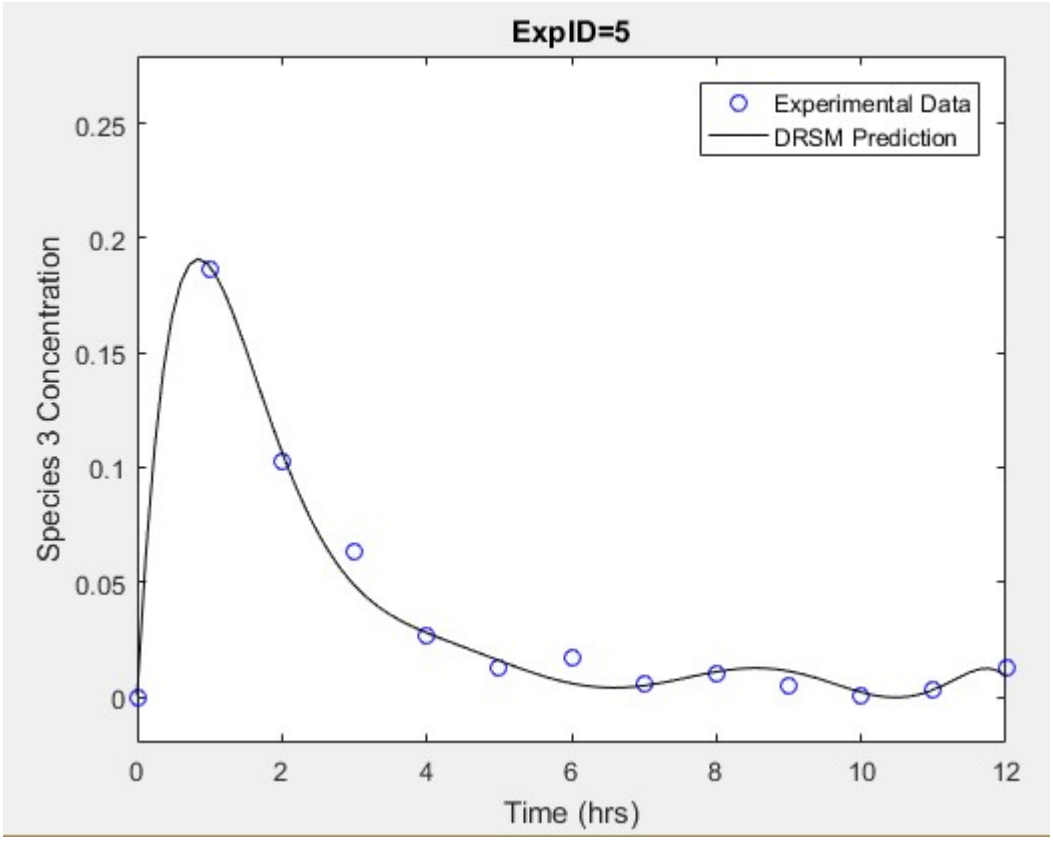
Appendix
Oscillations in Stepwise Regression DRSM

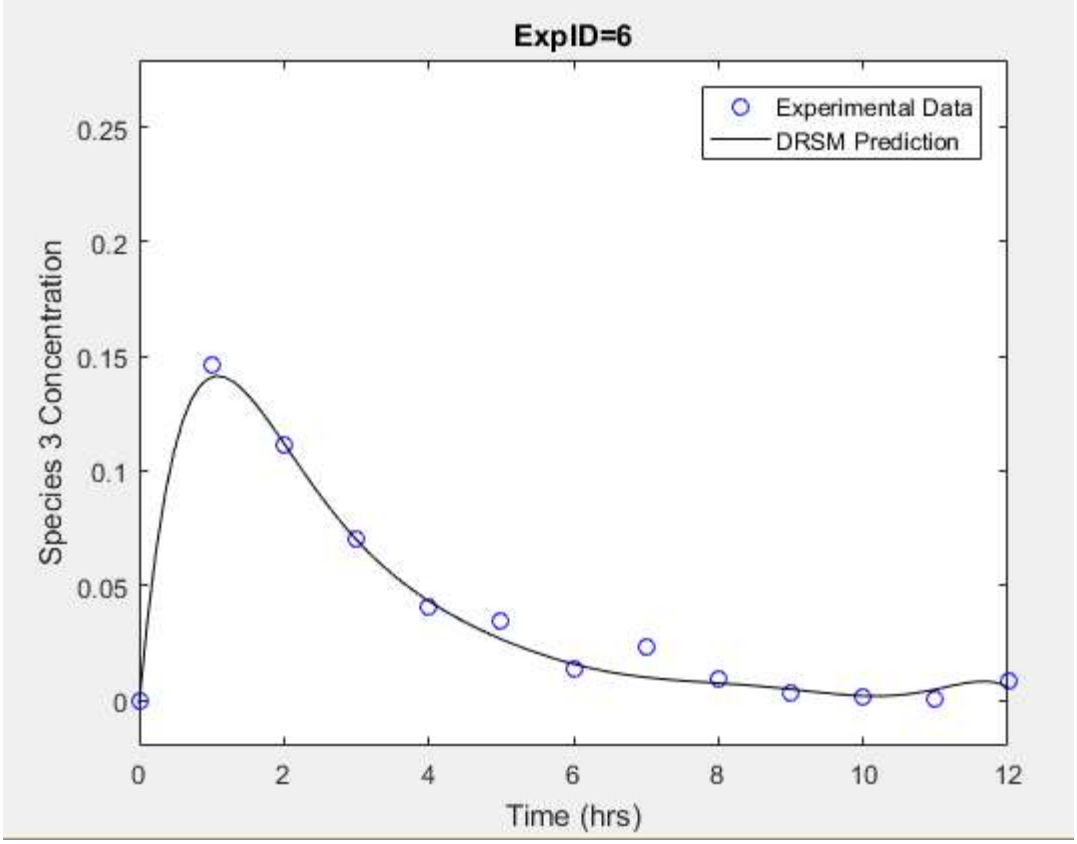


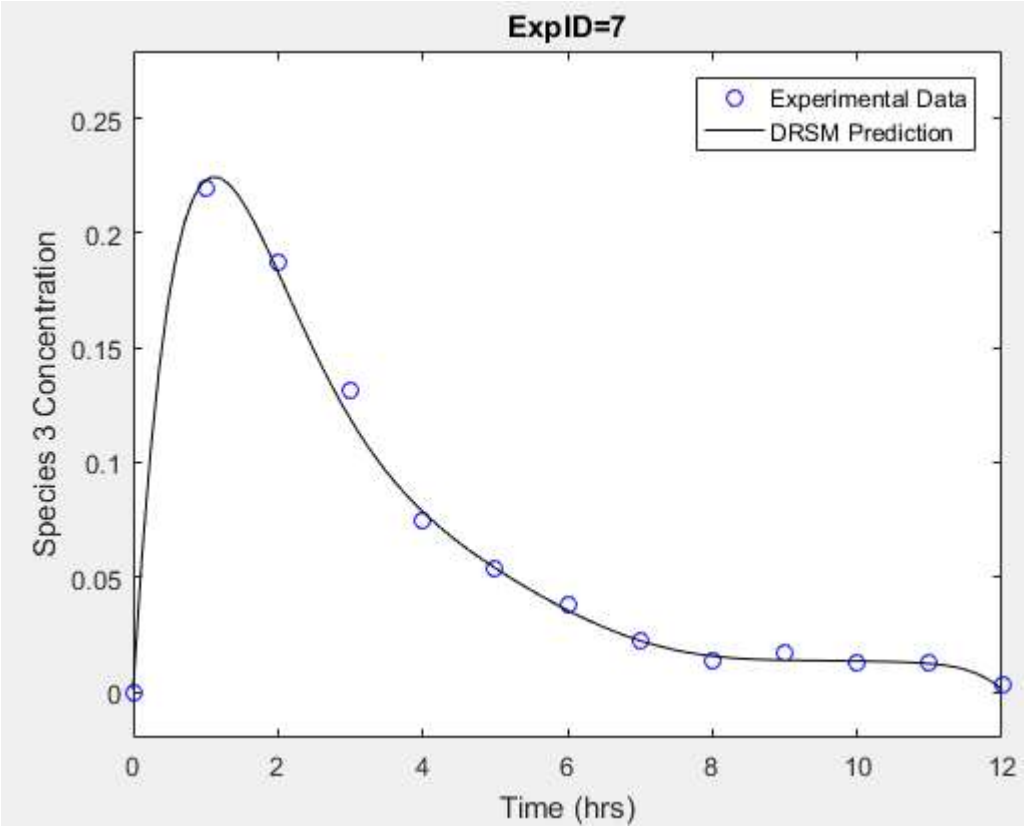

ExpID=2
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

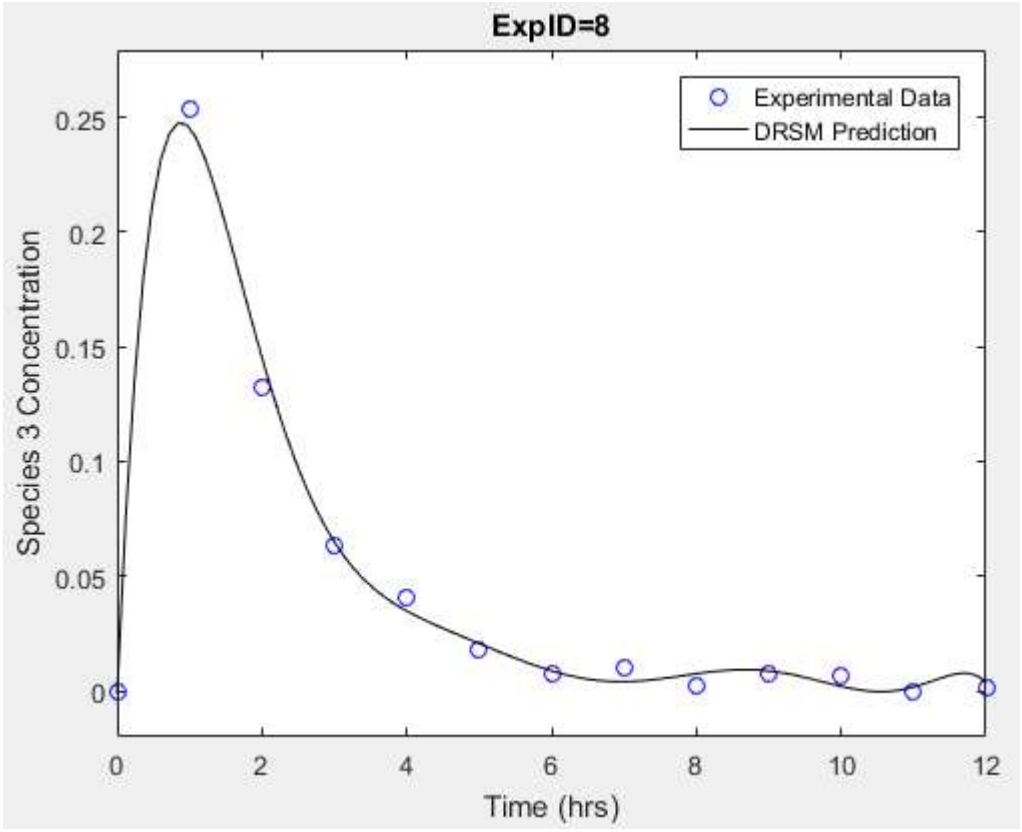

ExpID=3
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

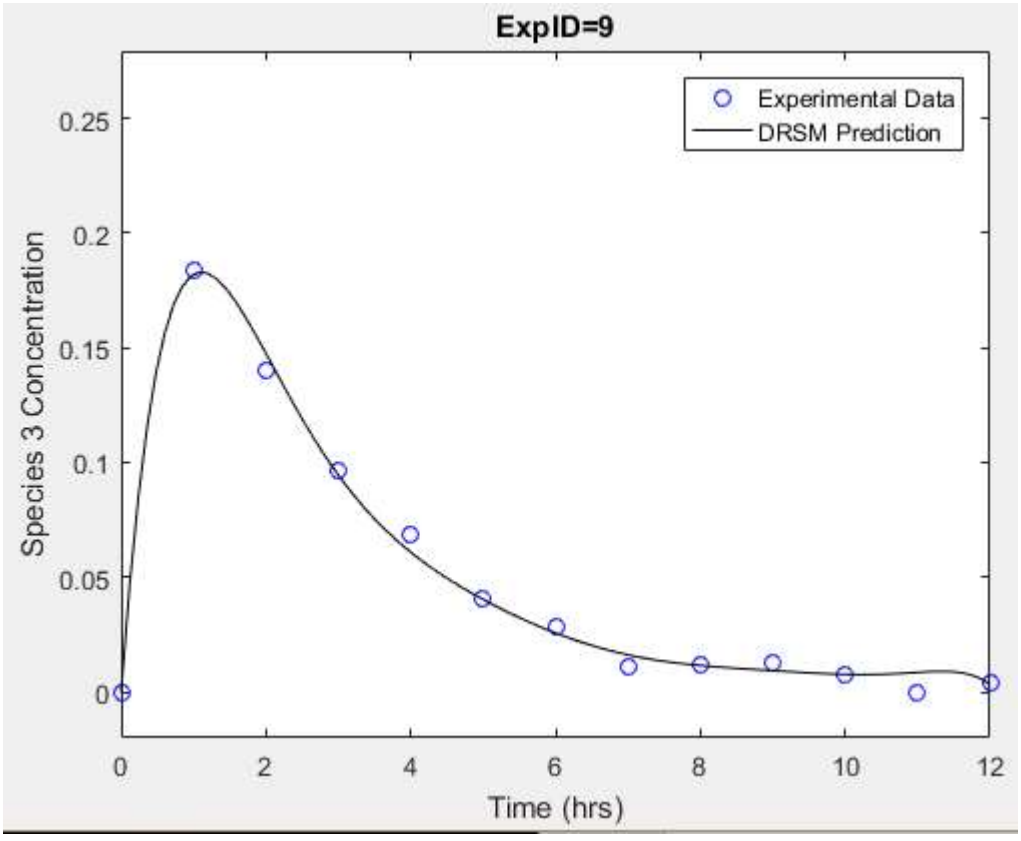

ExpID=4
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

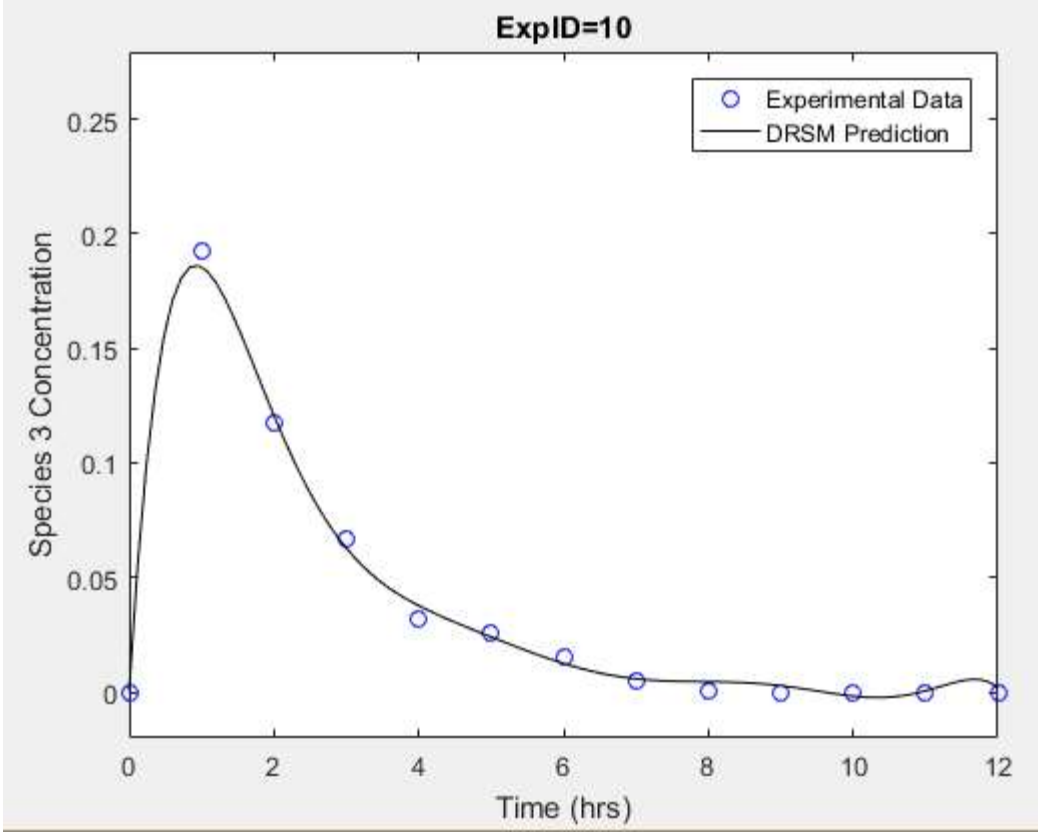

ExpID=5
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

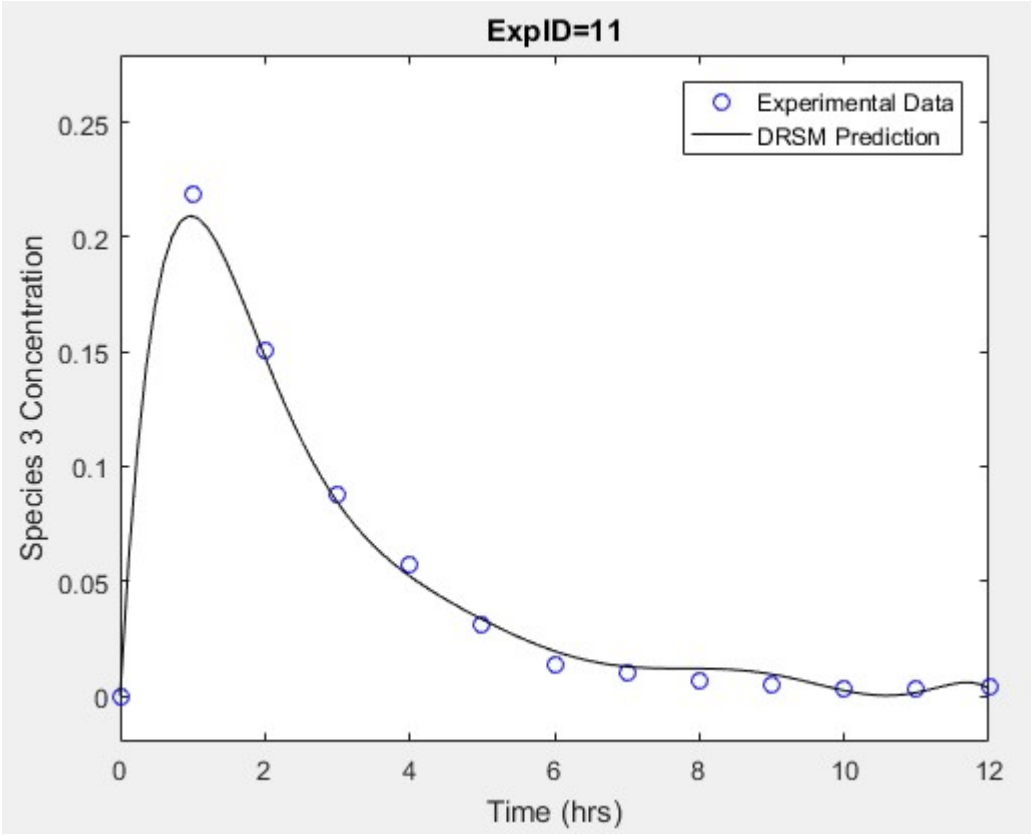

ExpID=6
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

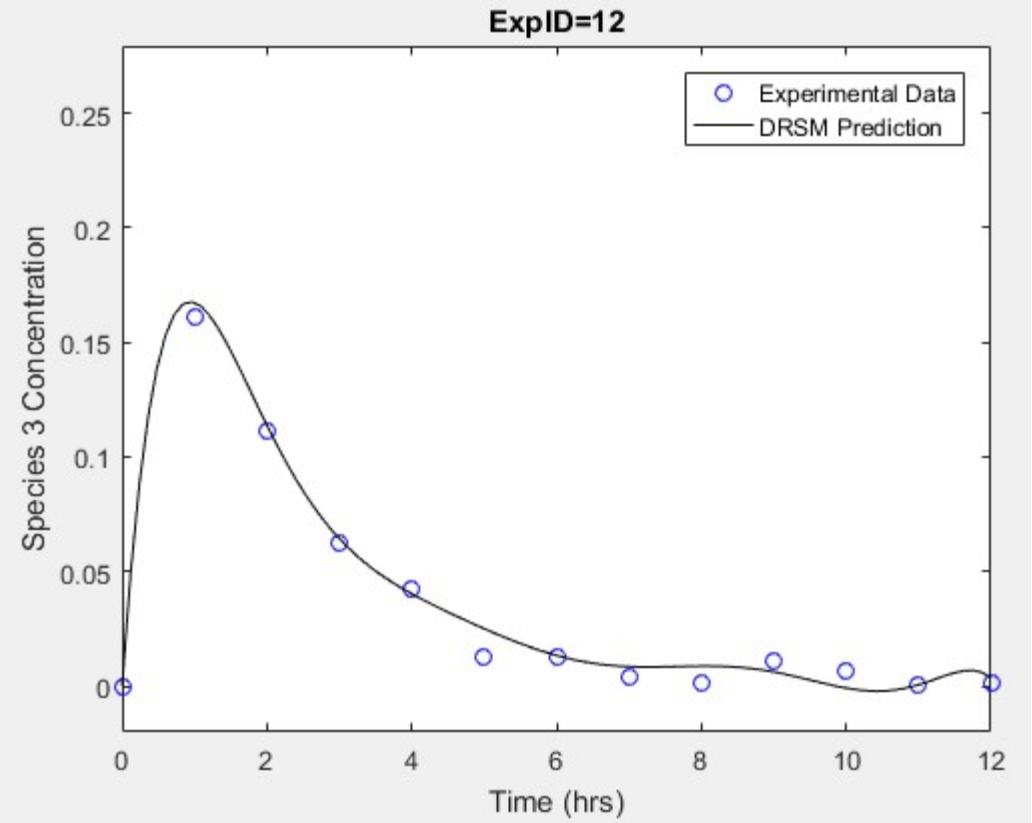

ExpID=7
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

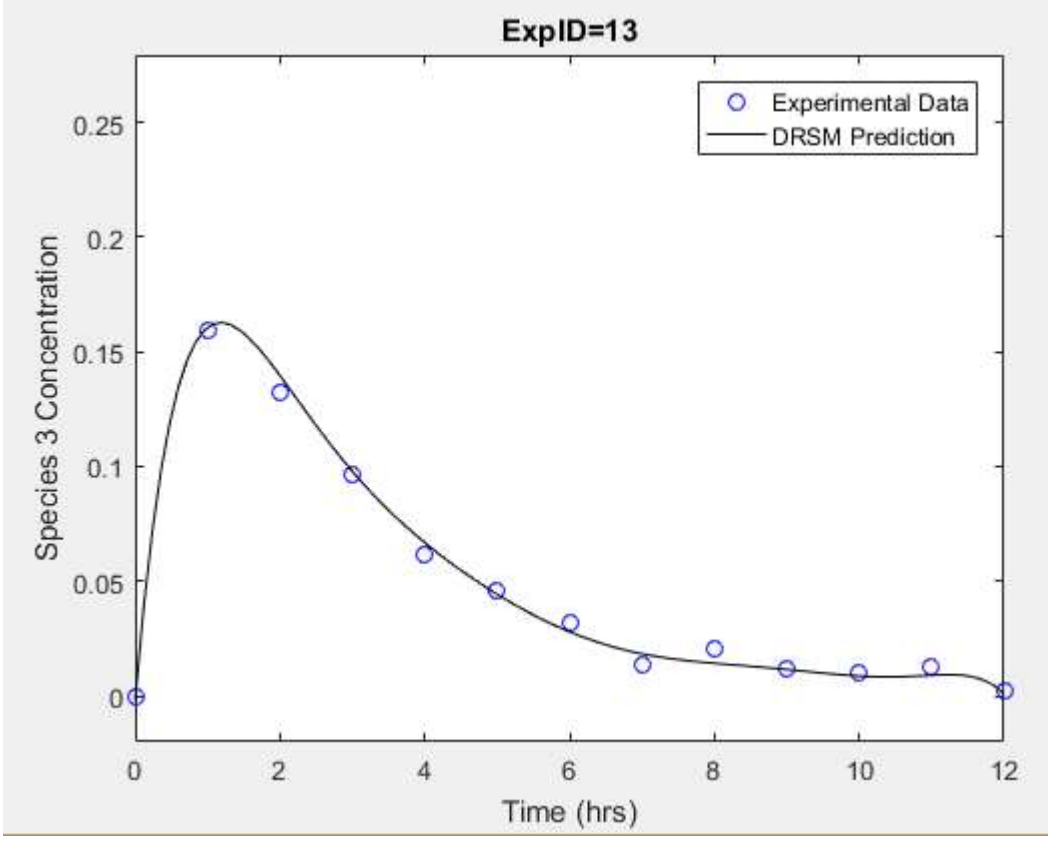

ExpID=8
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

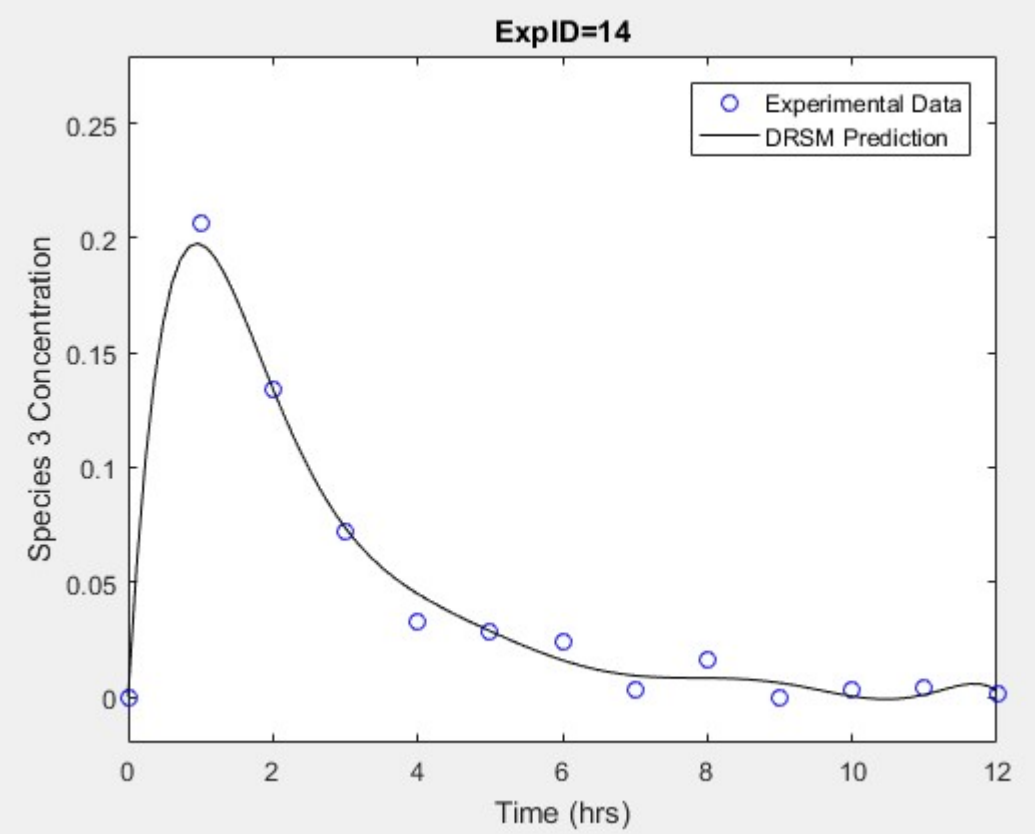

ExpID=9
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

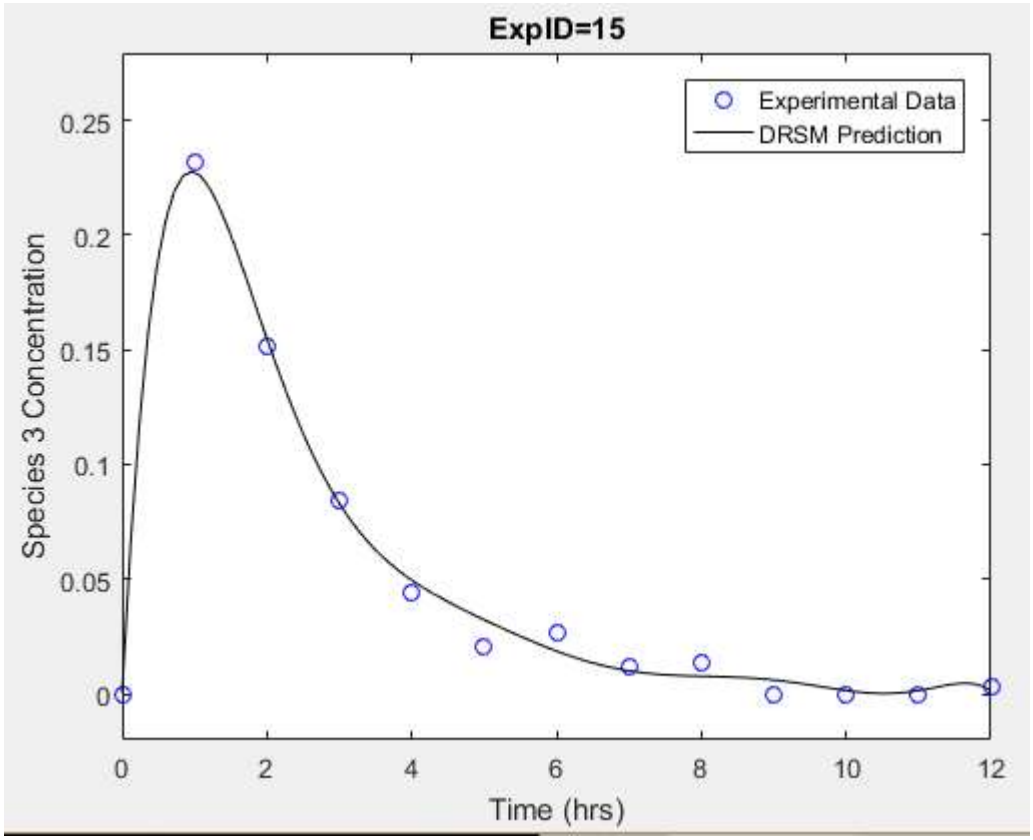

ExpID=10
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

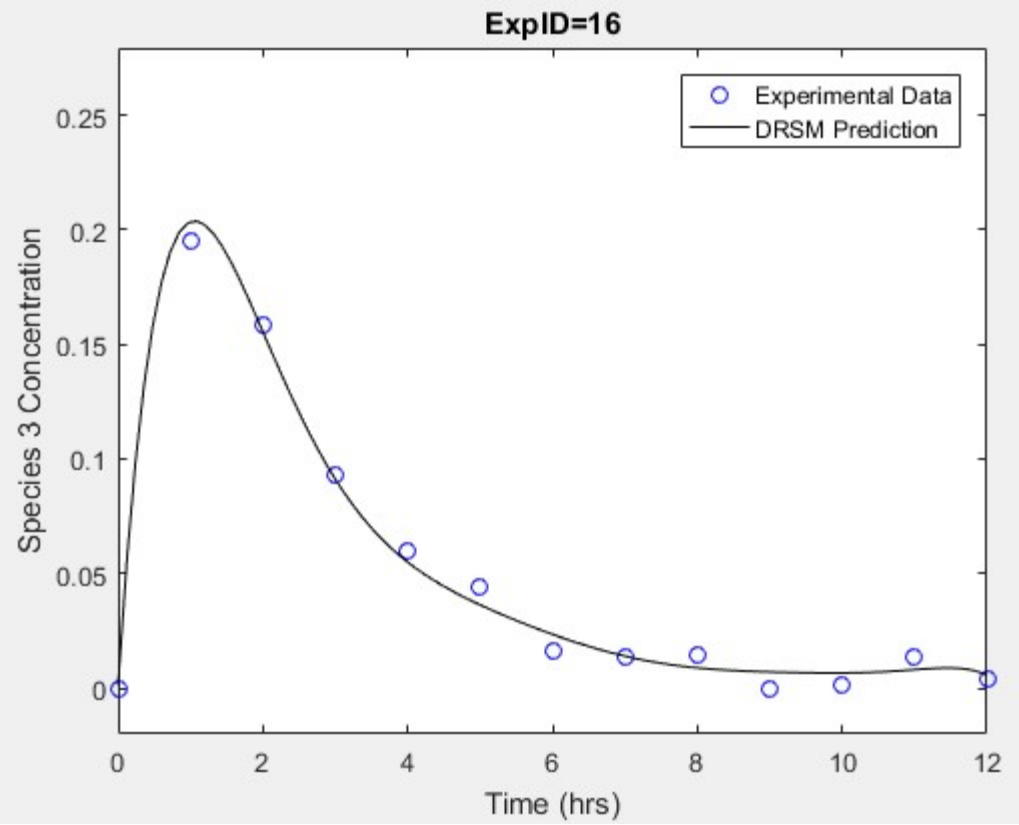

ExpID=11
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

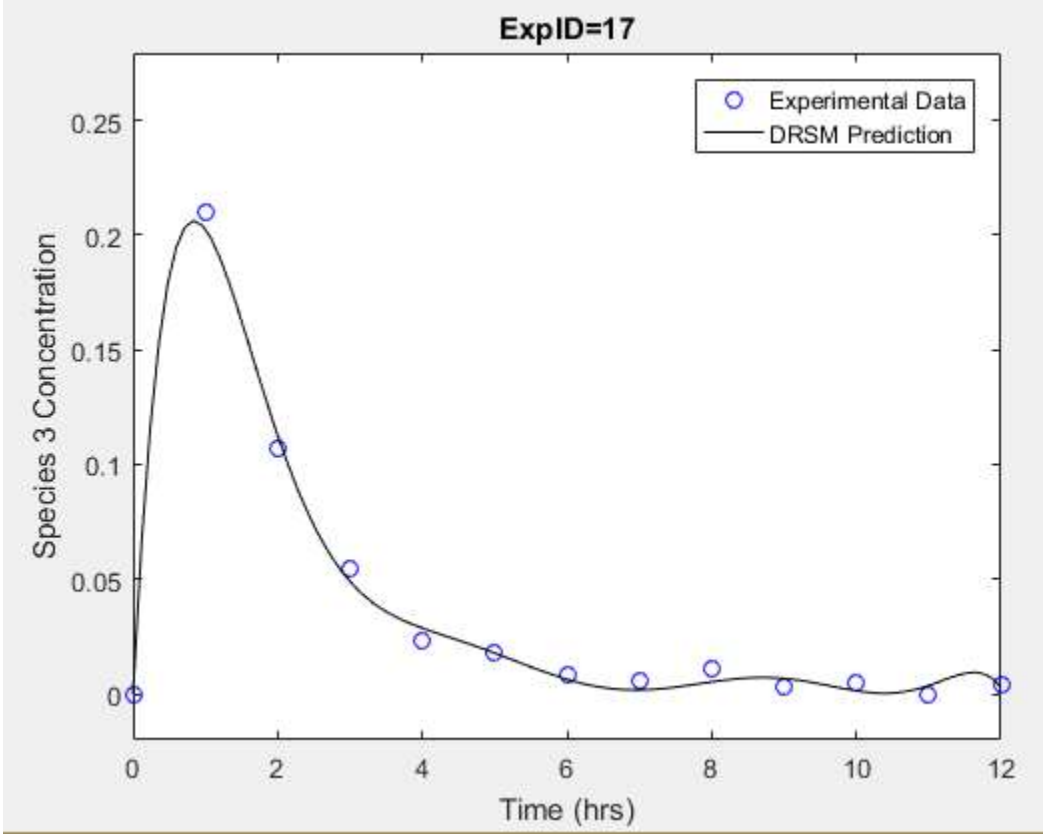

ExpID=12
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=13
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

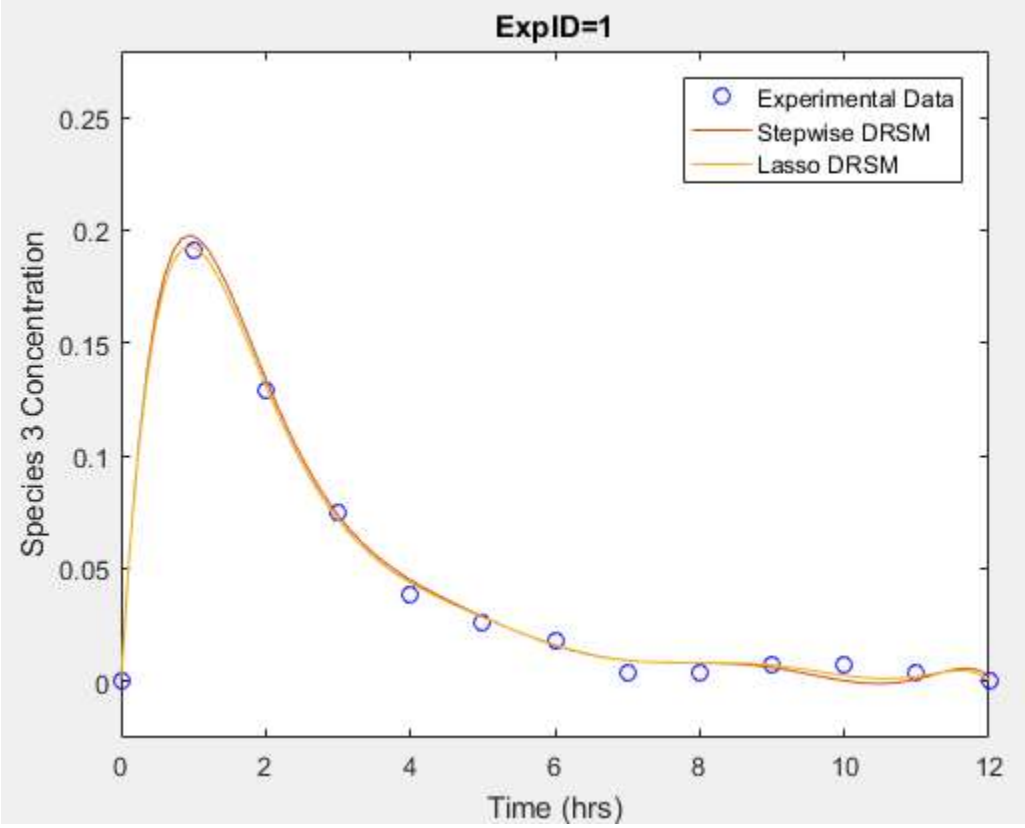

ExpID=14
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

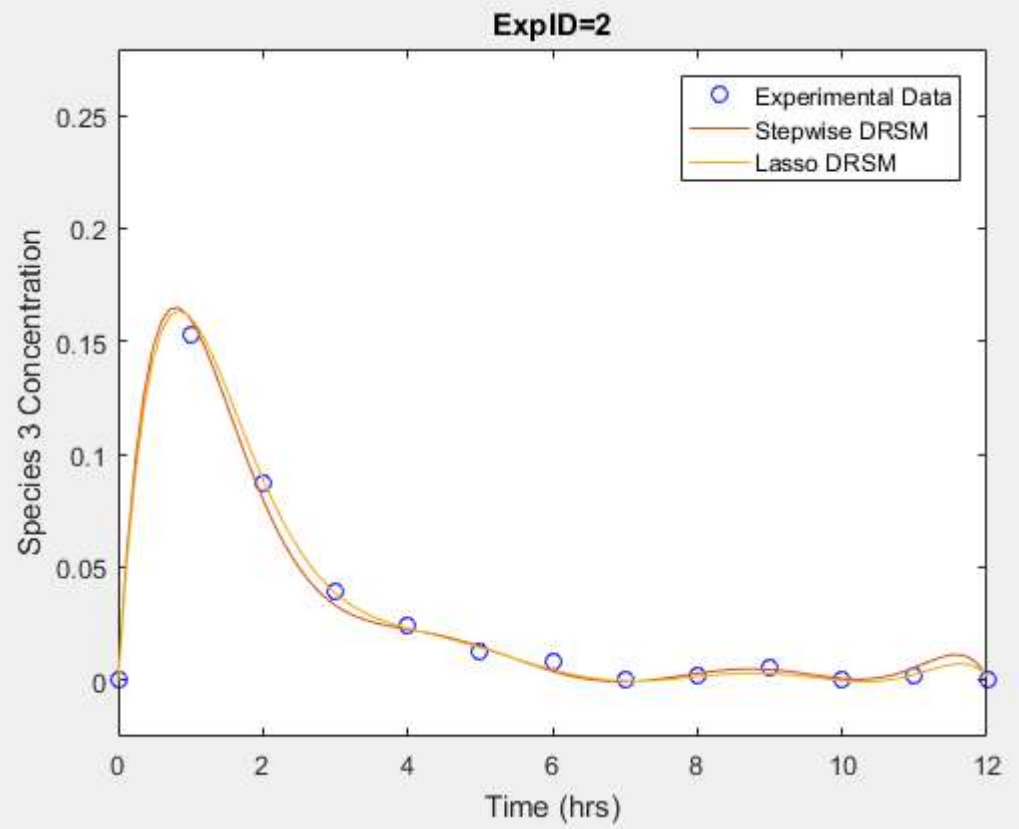

ExpID=15
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

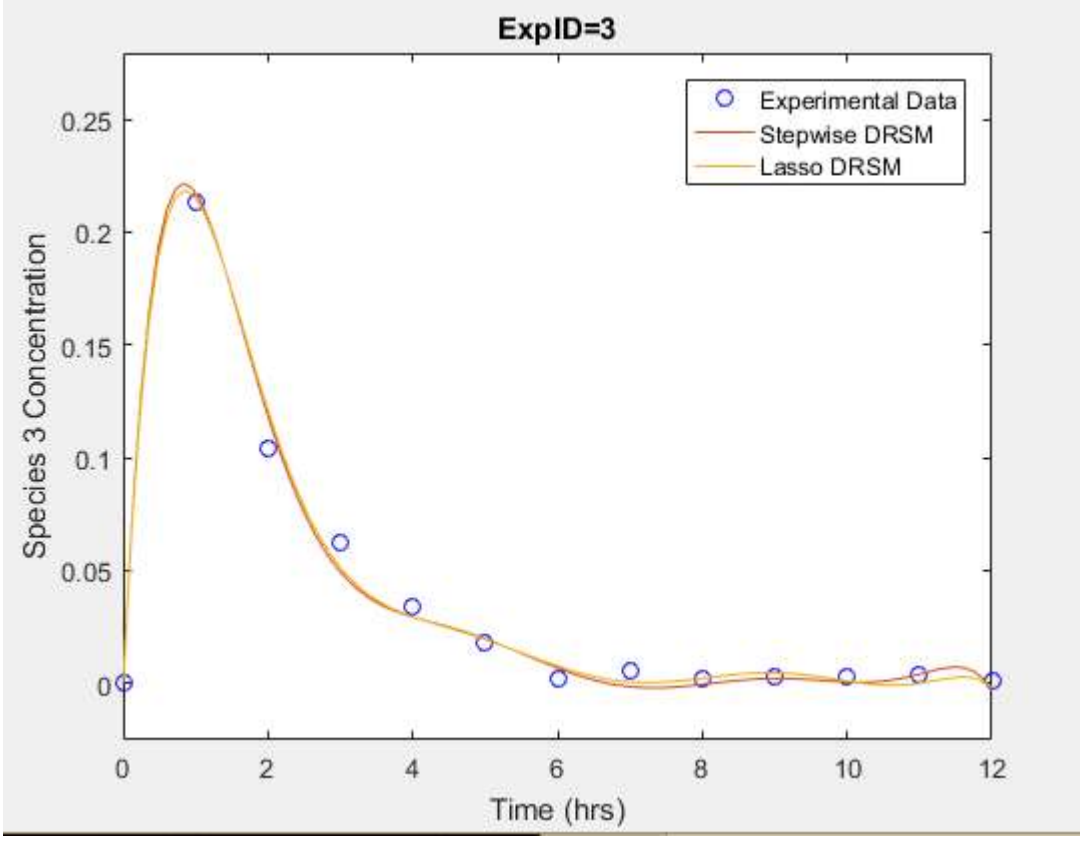

ExpID=16
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

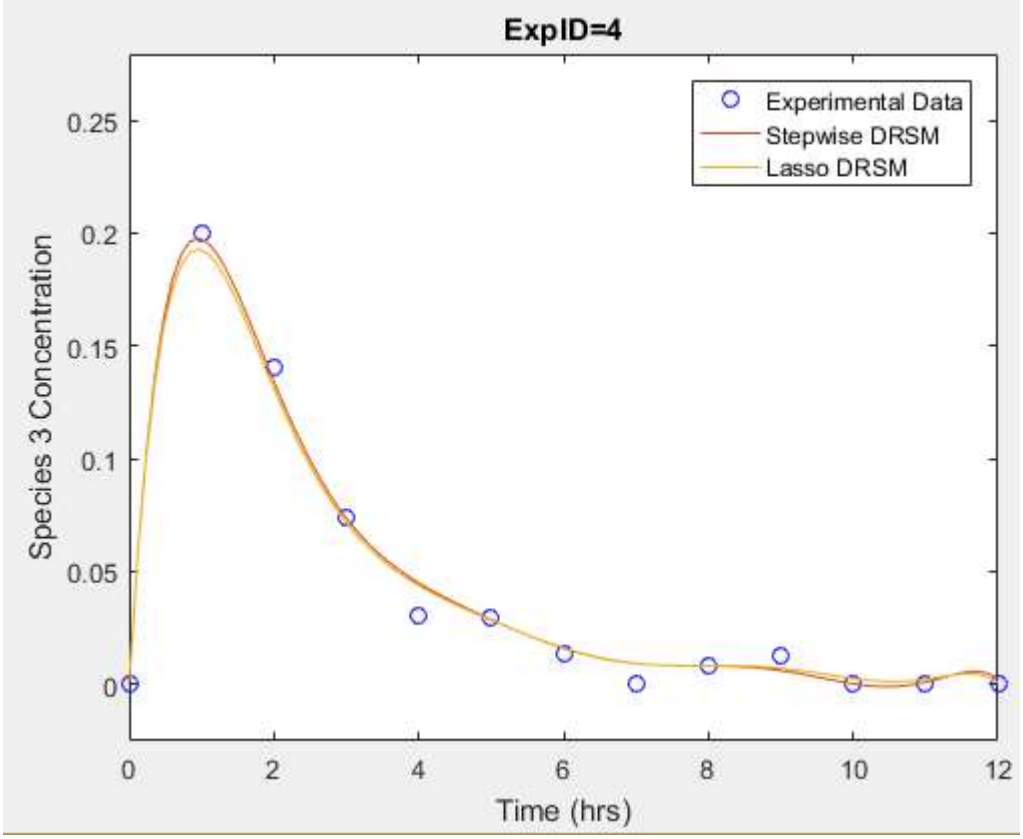

ExpID=17
Experimental Data
DRSM Prediction
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

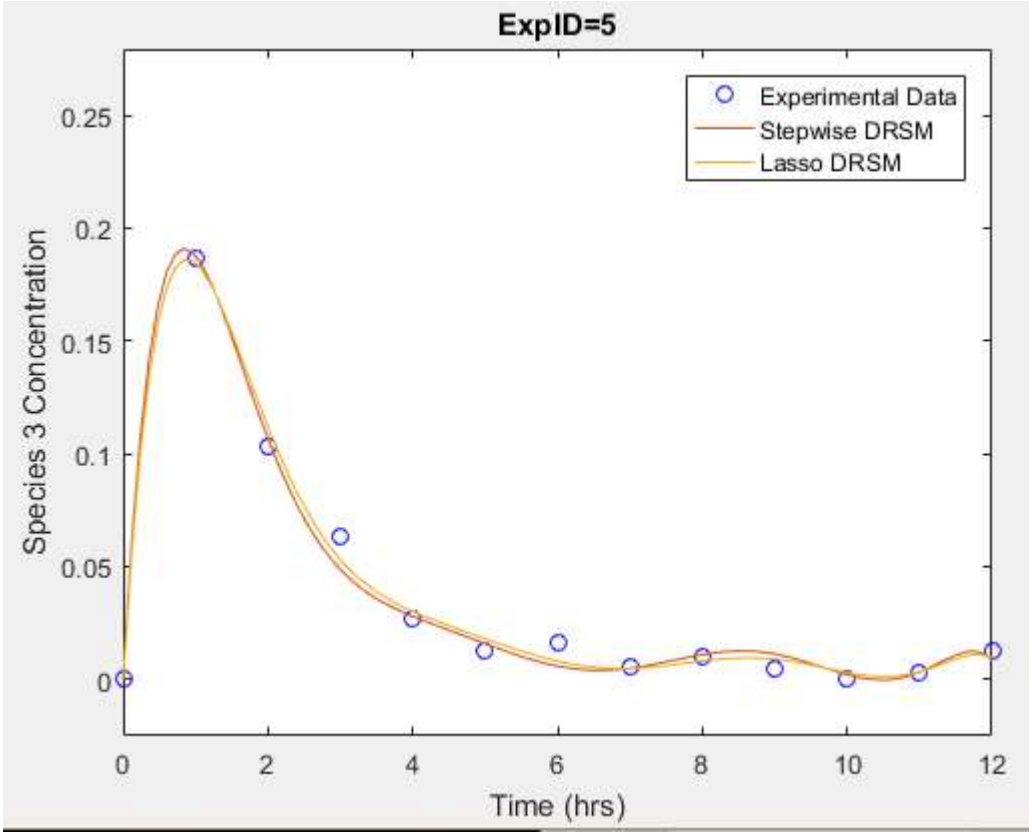
Stepwise vs LASSO, $\lambda=2.2*10^{-4}$

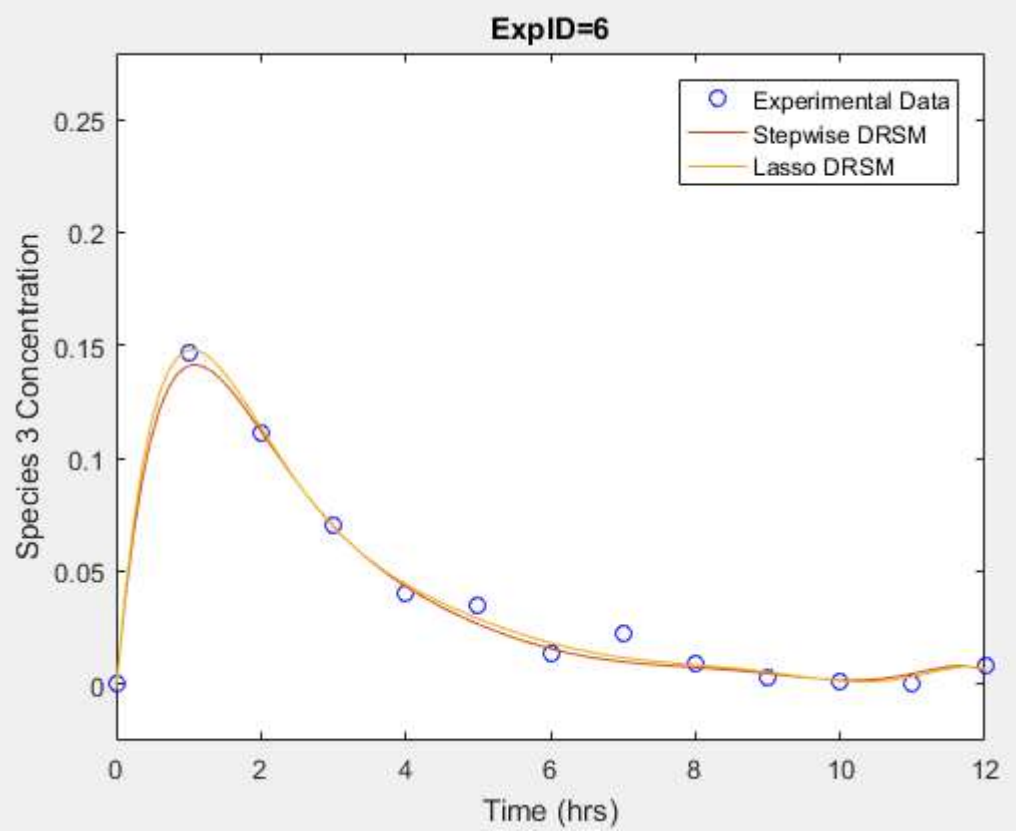


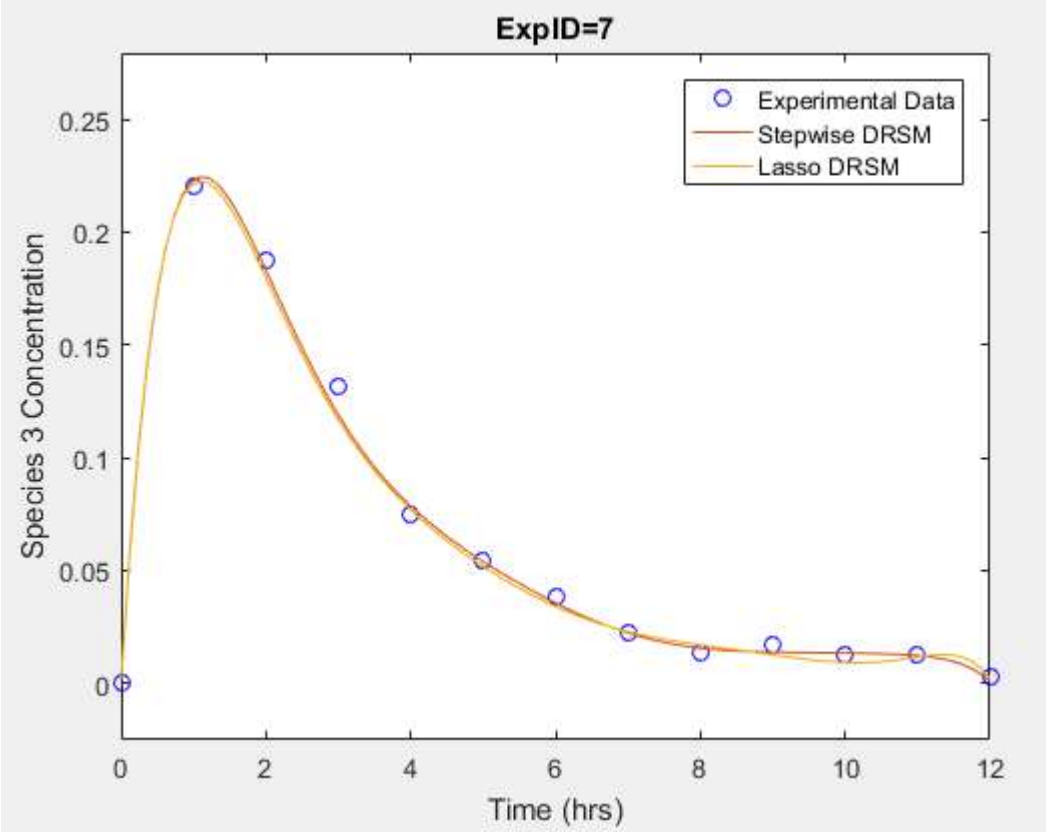

ExpID=2
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

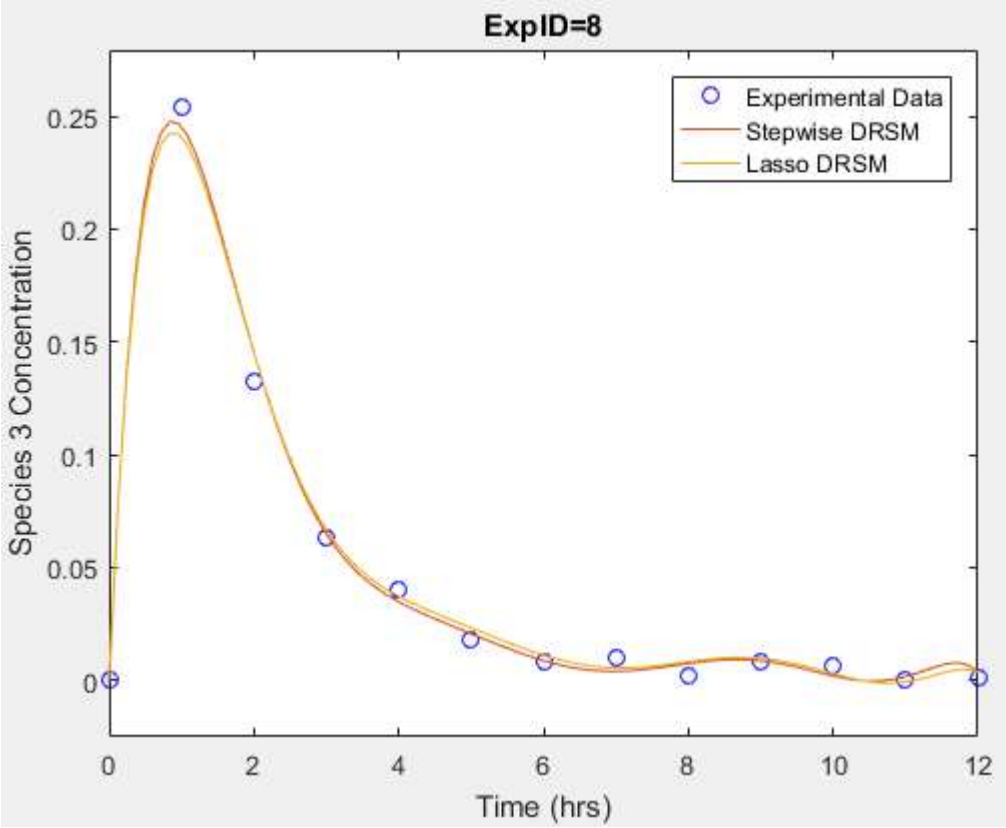

ExpID=3
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

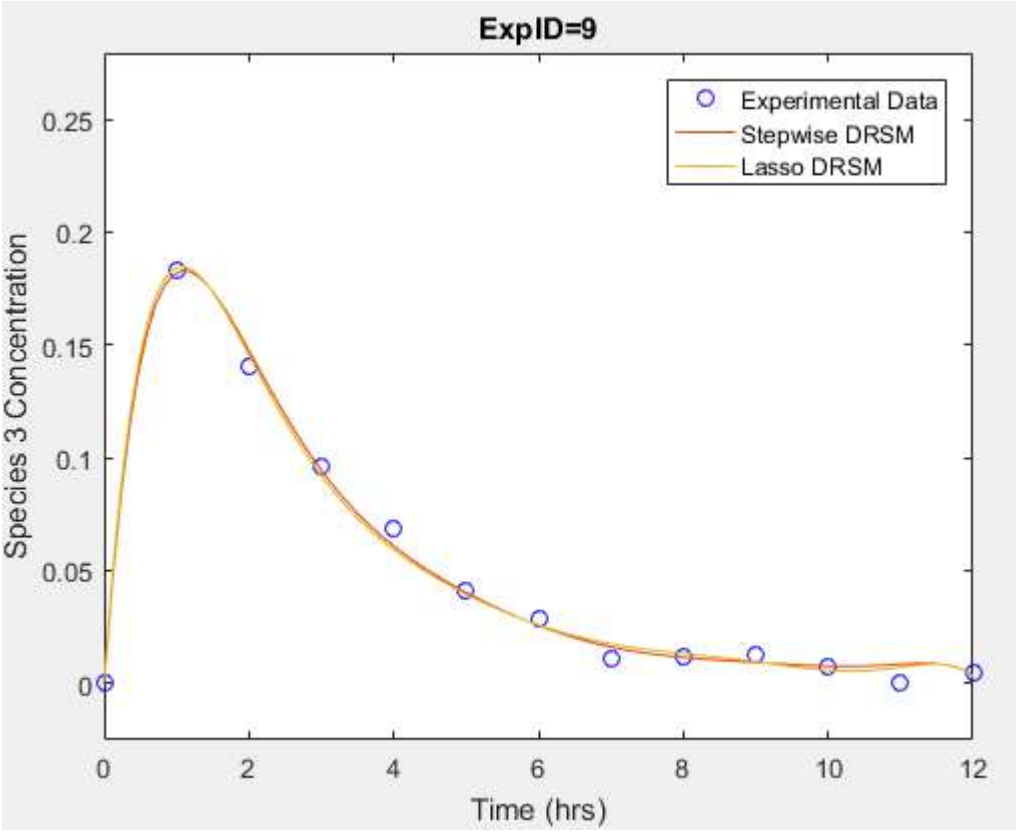

ExpID=4
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

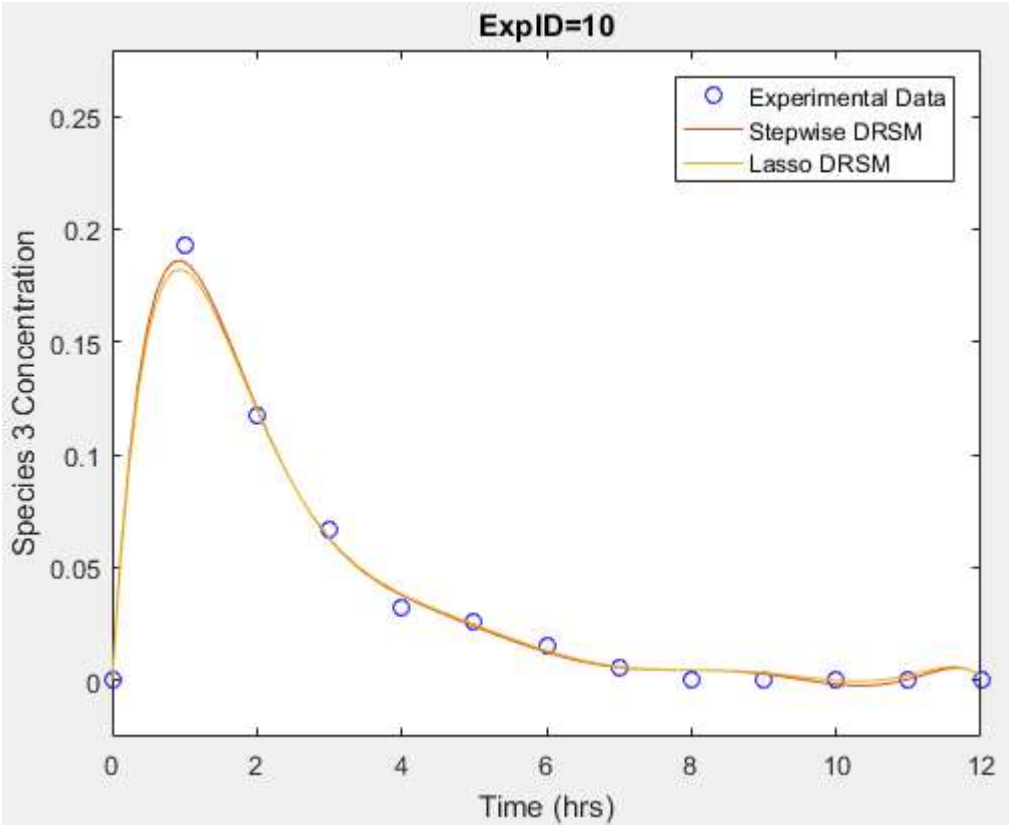

ExpID=5
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

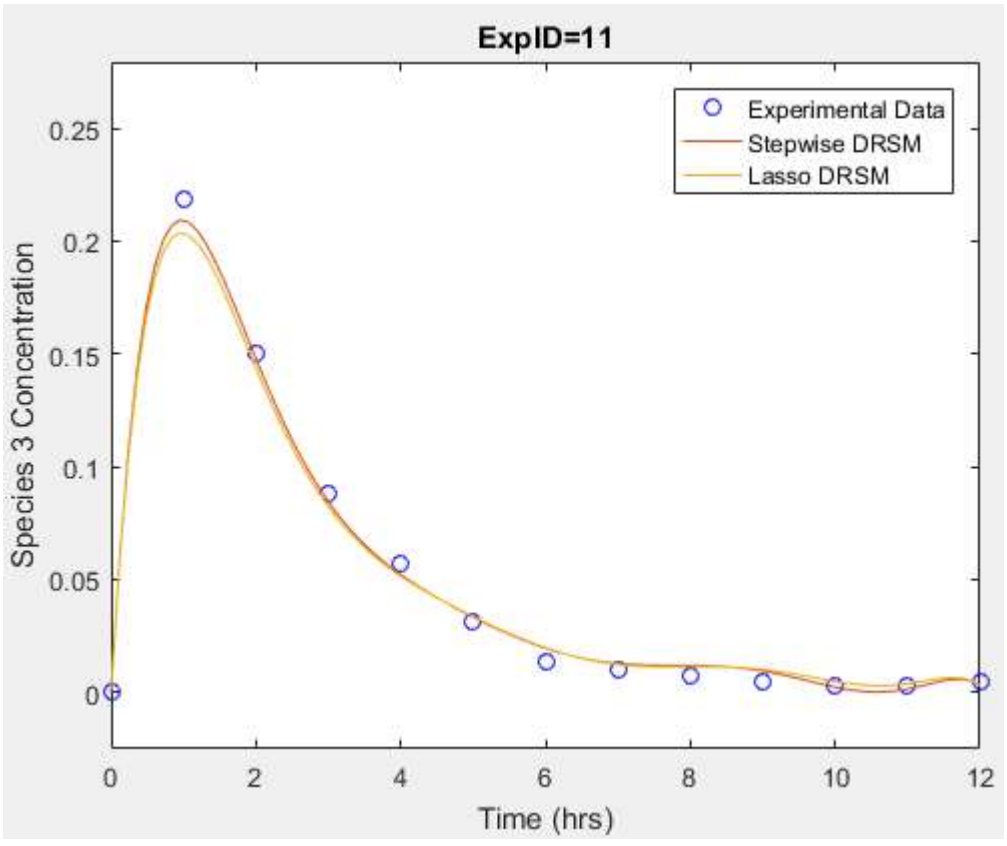

ExpID=6
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

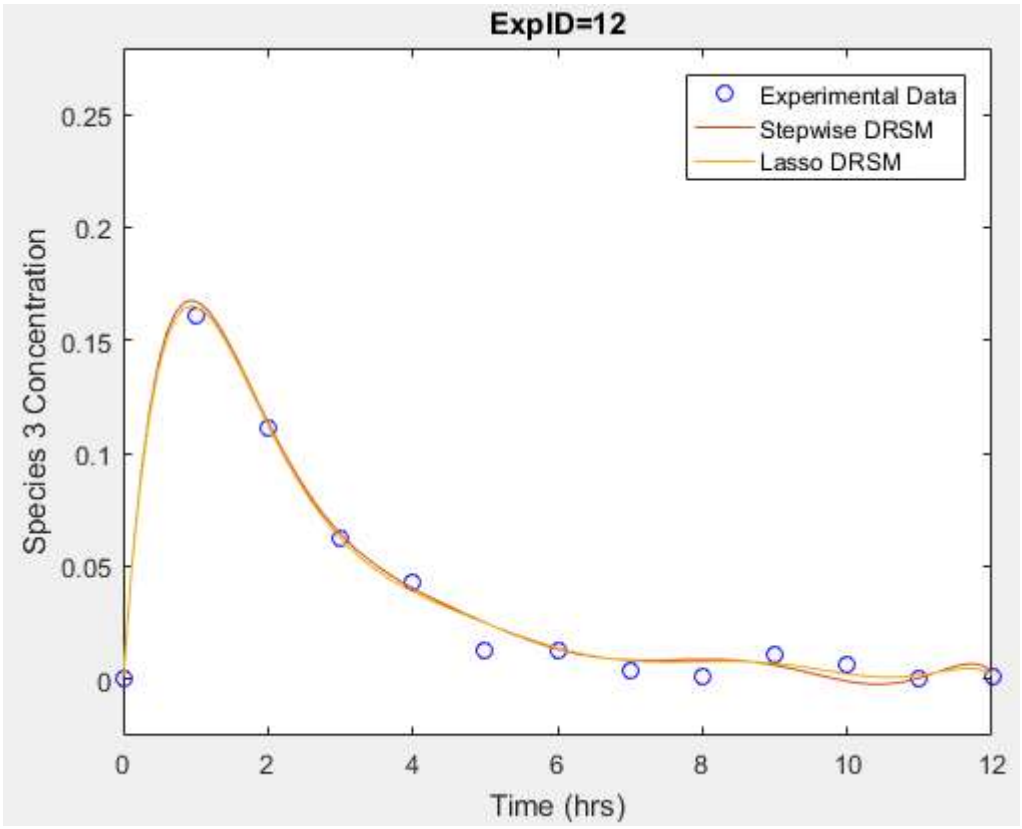

ExpID=7
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

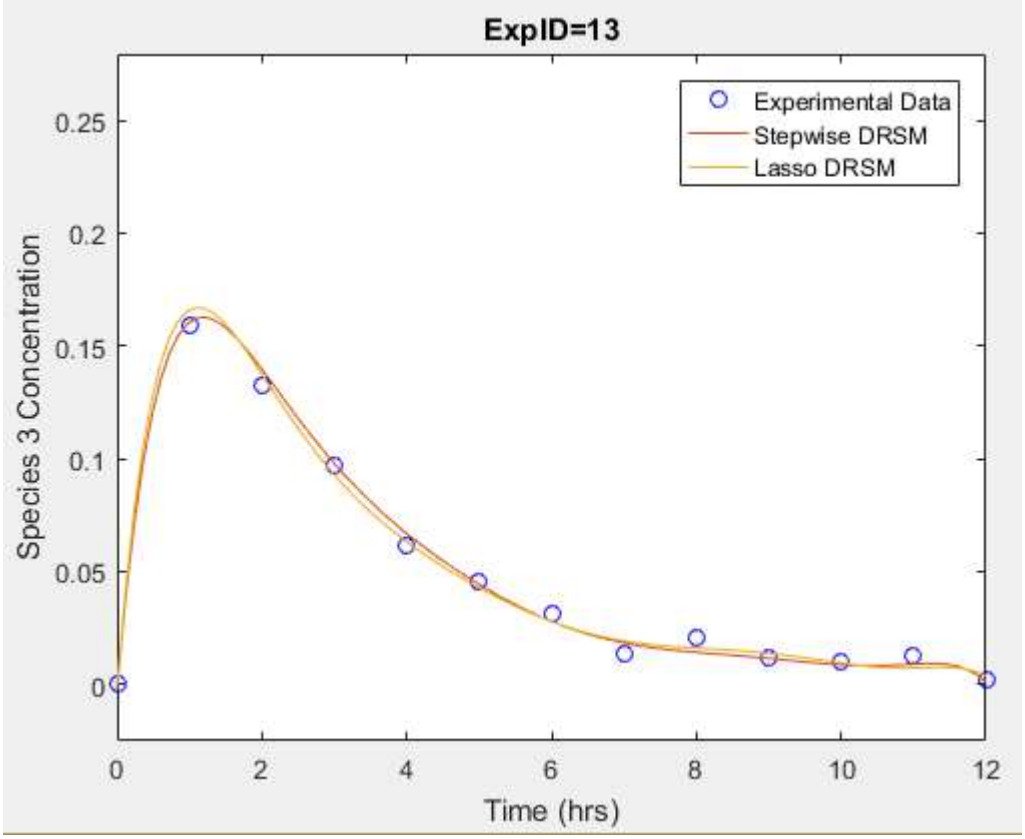

ExpID=8
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

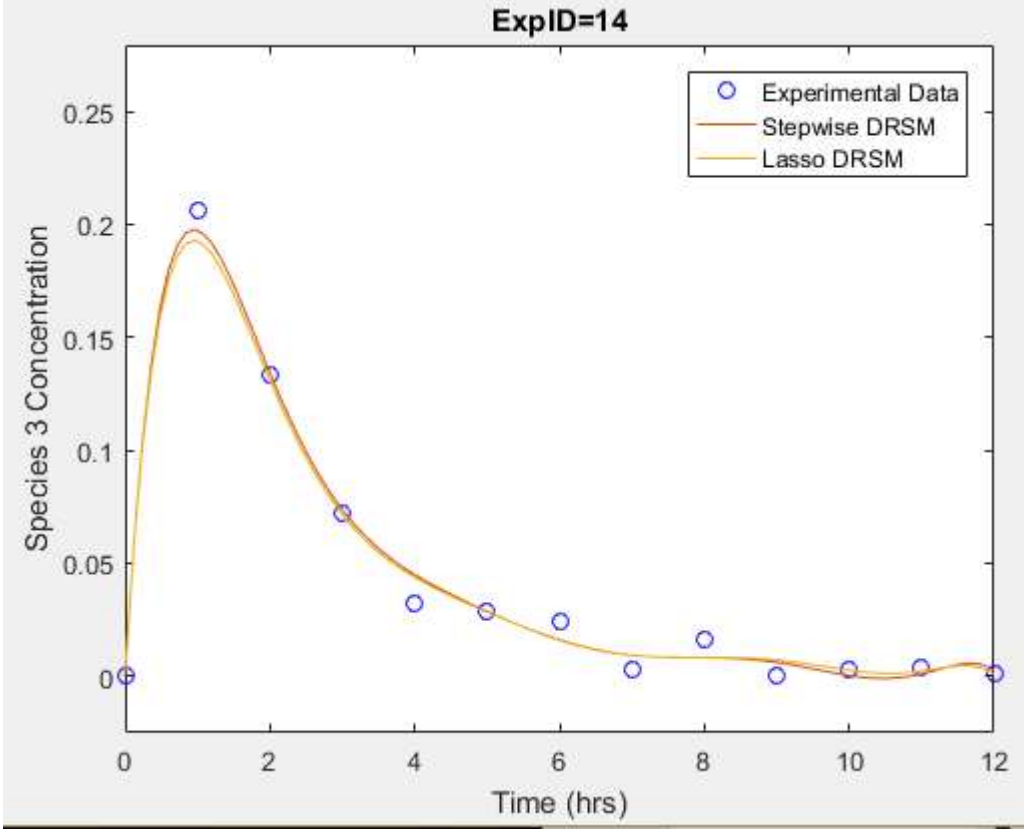

ExpID=9
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

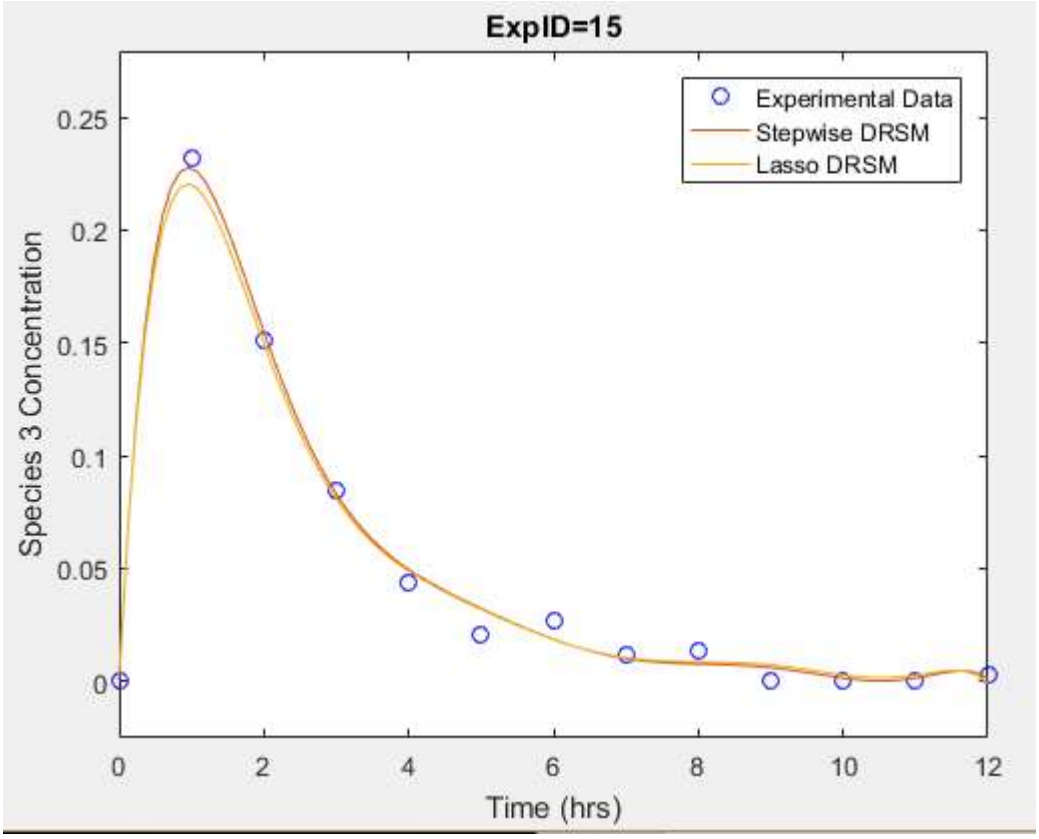

ExpID=10
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

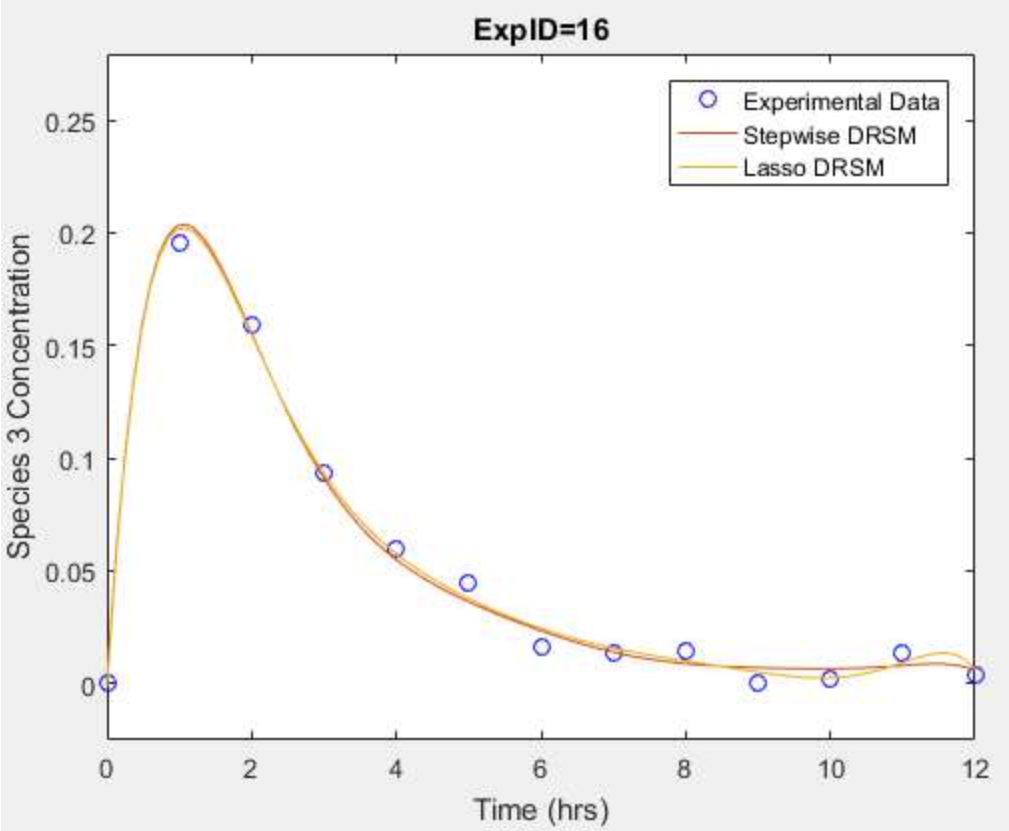

ExpID=11
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

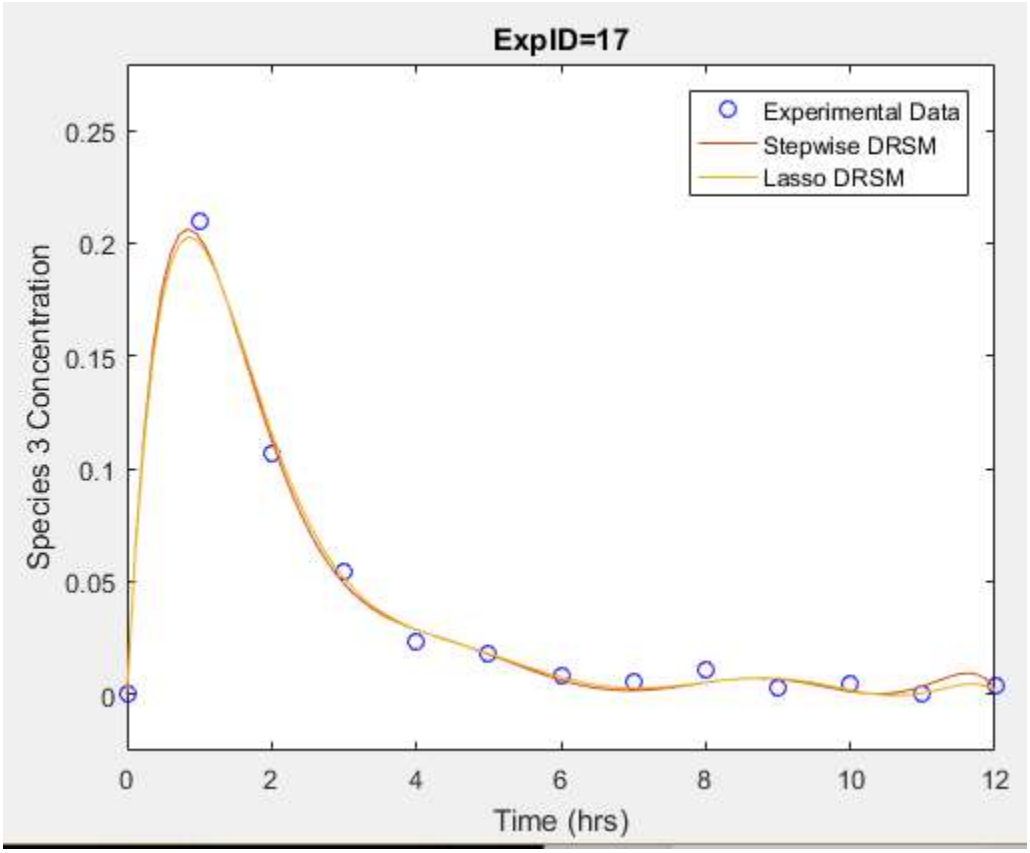

ExpID=12
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=13
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

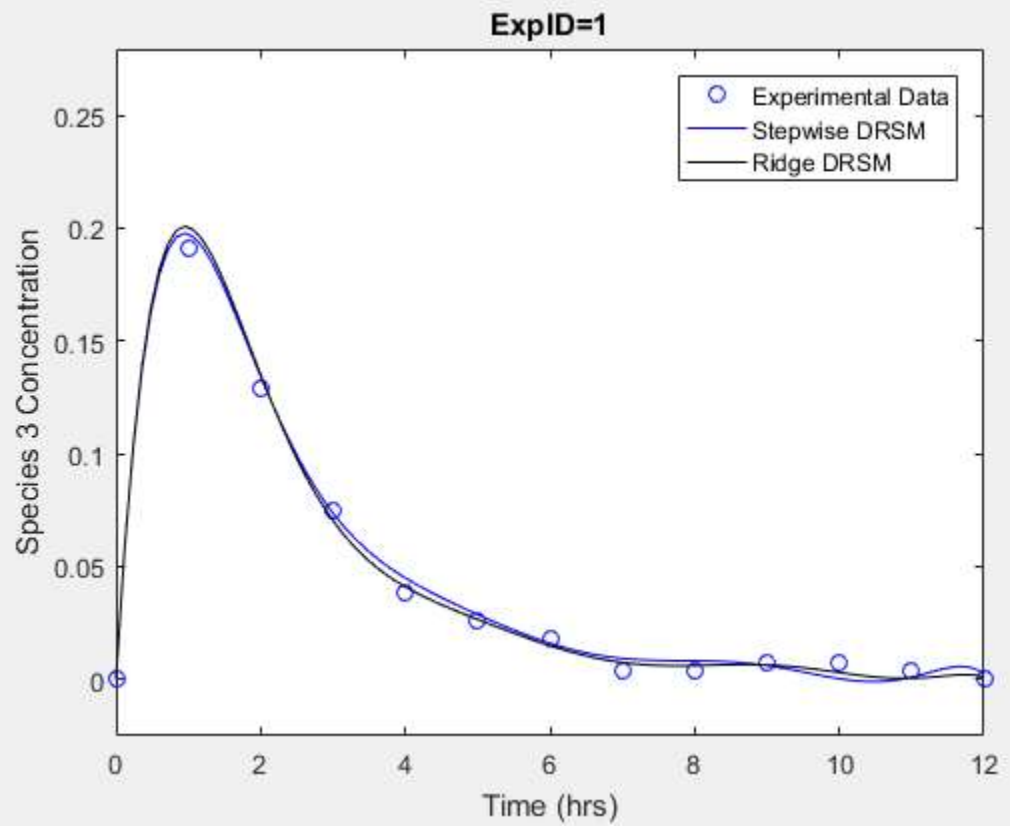

ExpID=14
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

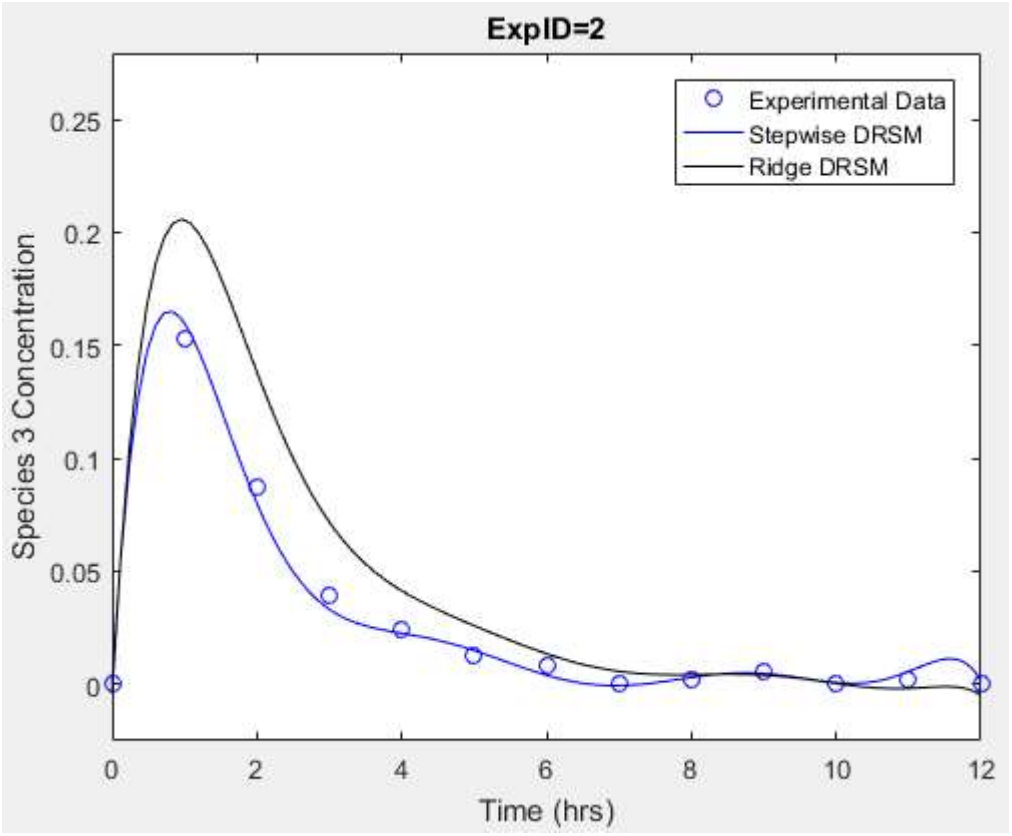

ExpID=15
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

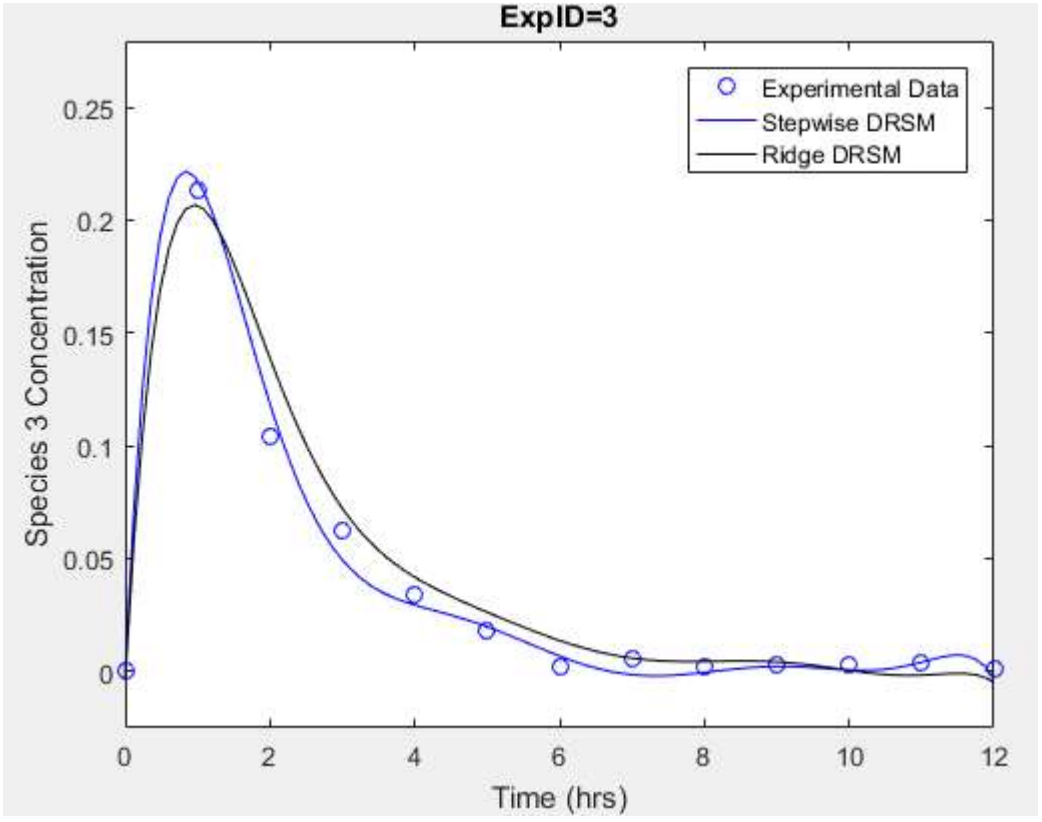

ExpID=16
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

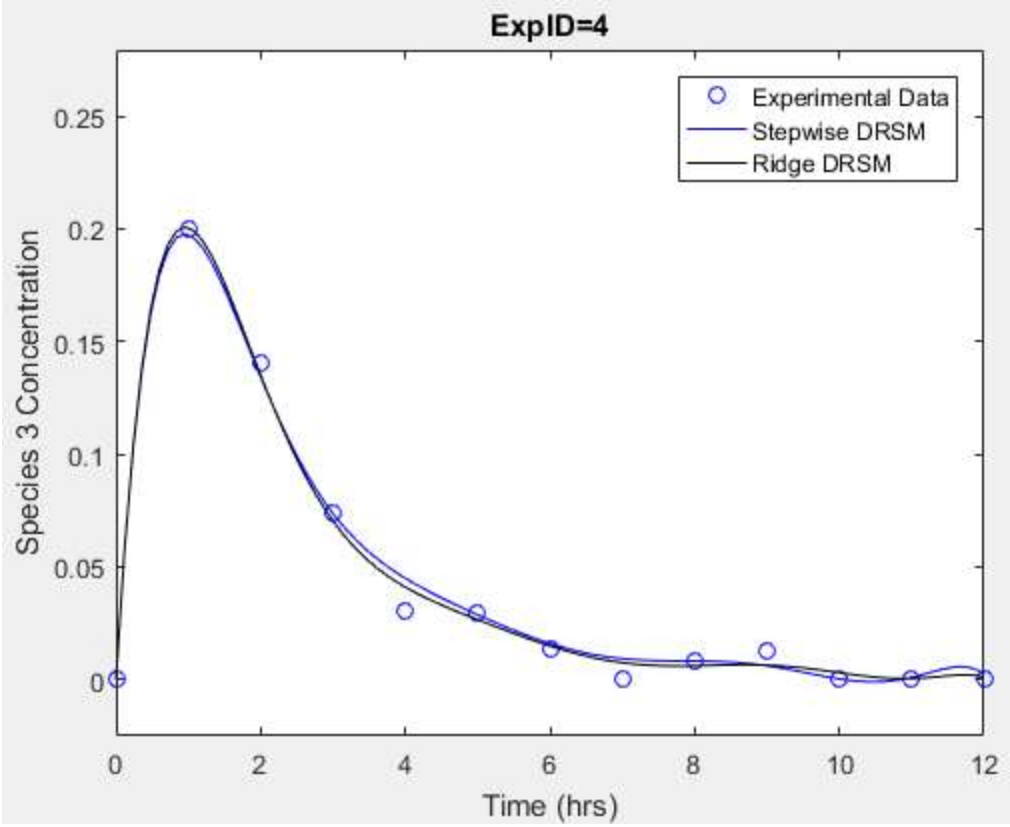

ExpID=17
Experimental Data
Stepwise DRSM
Lasso DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

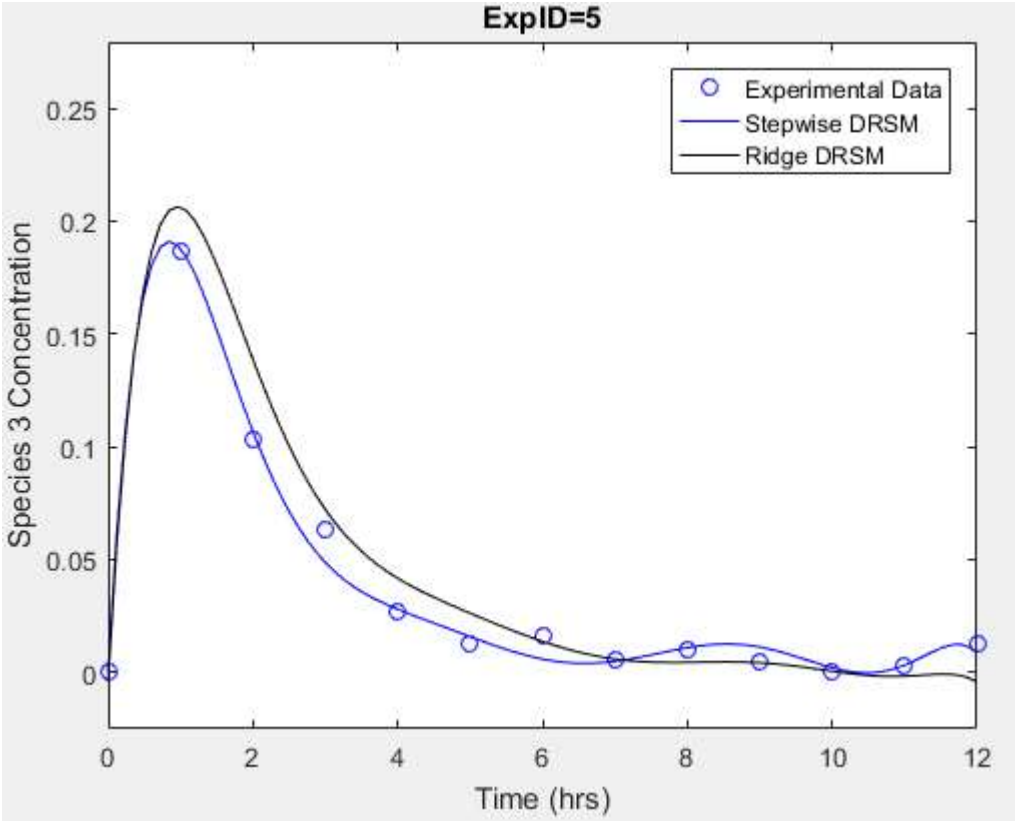
Stepwise vs Ridge, α=4000

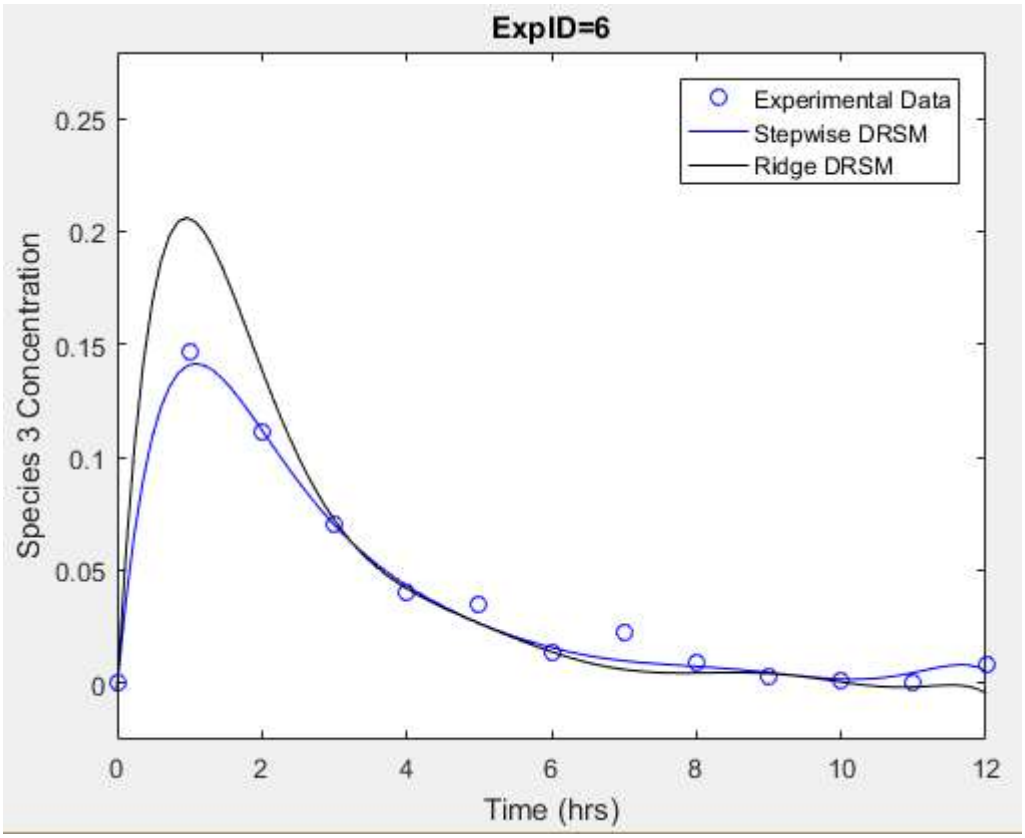


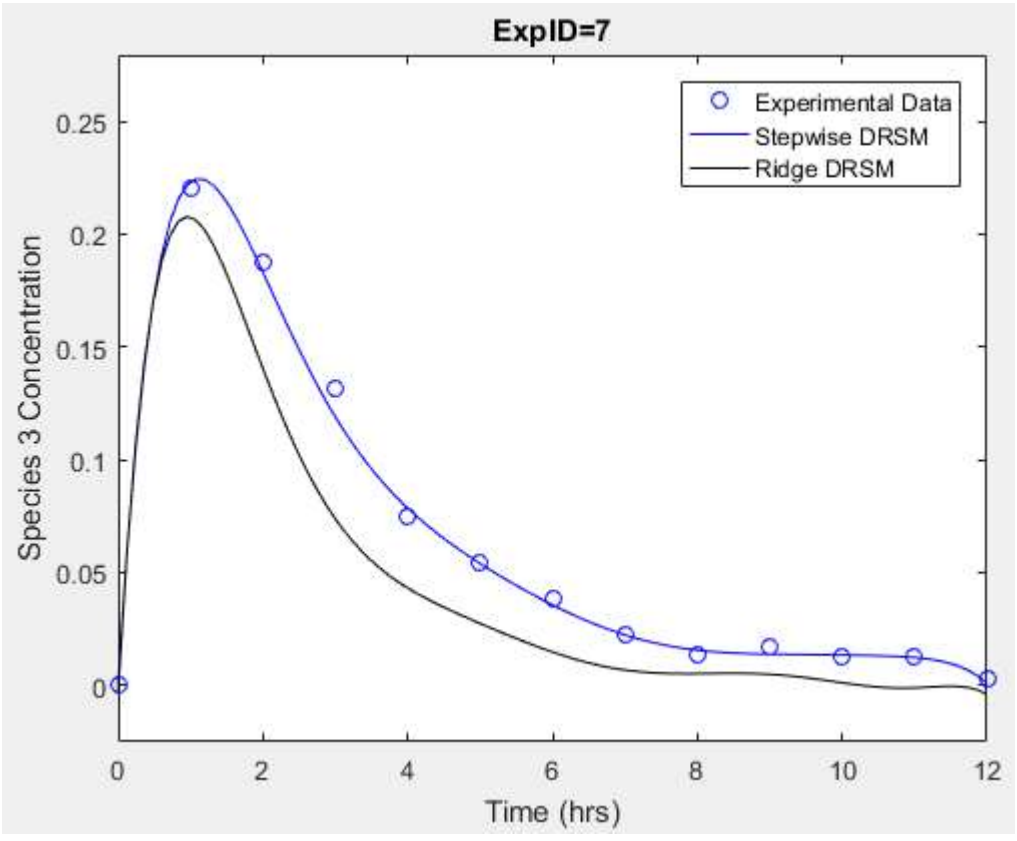

ExpID=2
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

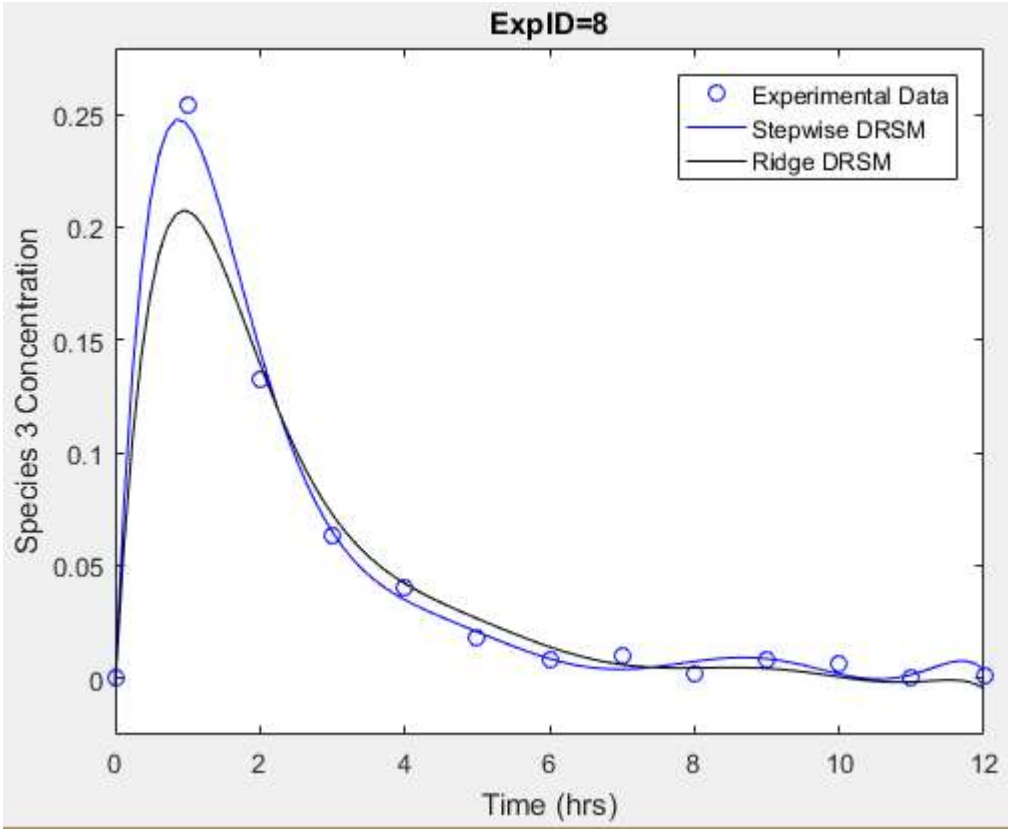

ExpID=3
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

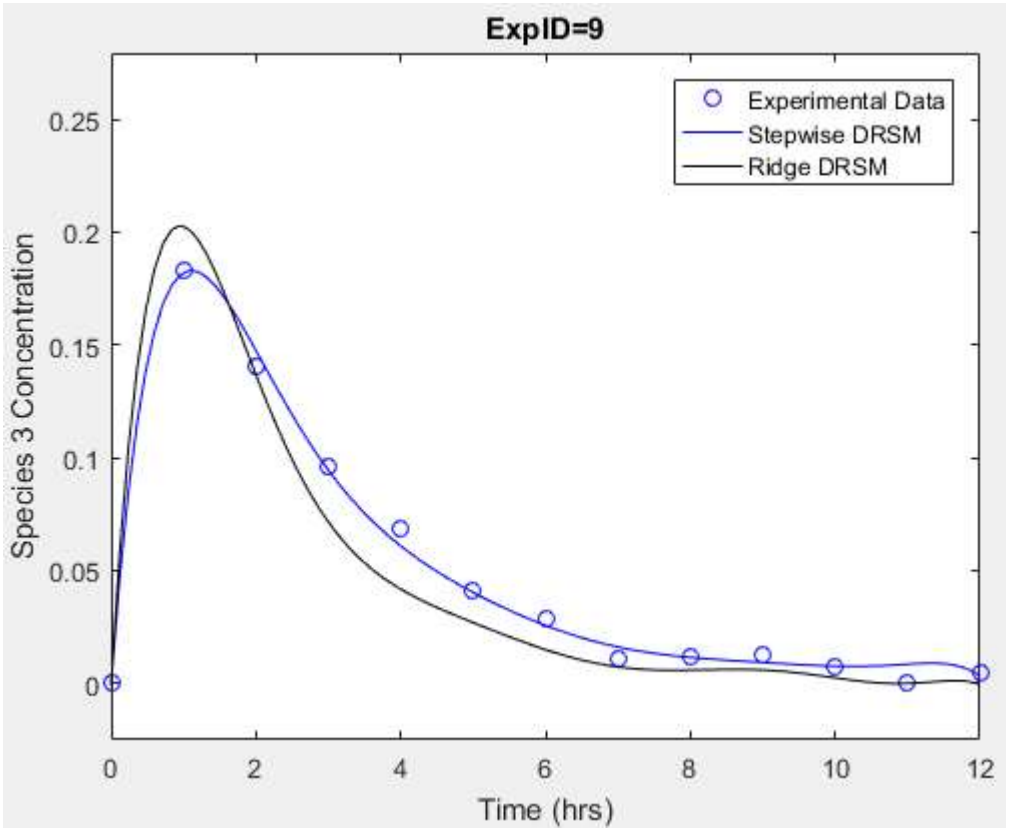

ExpID=4
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

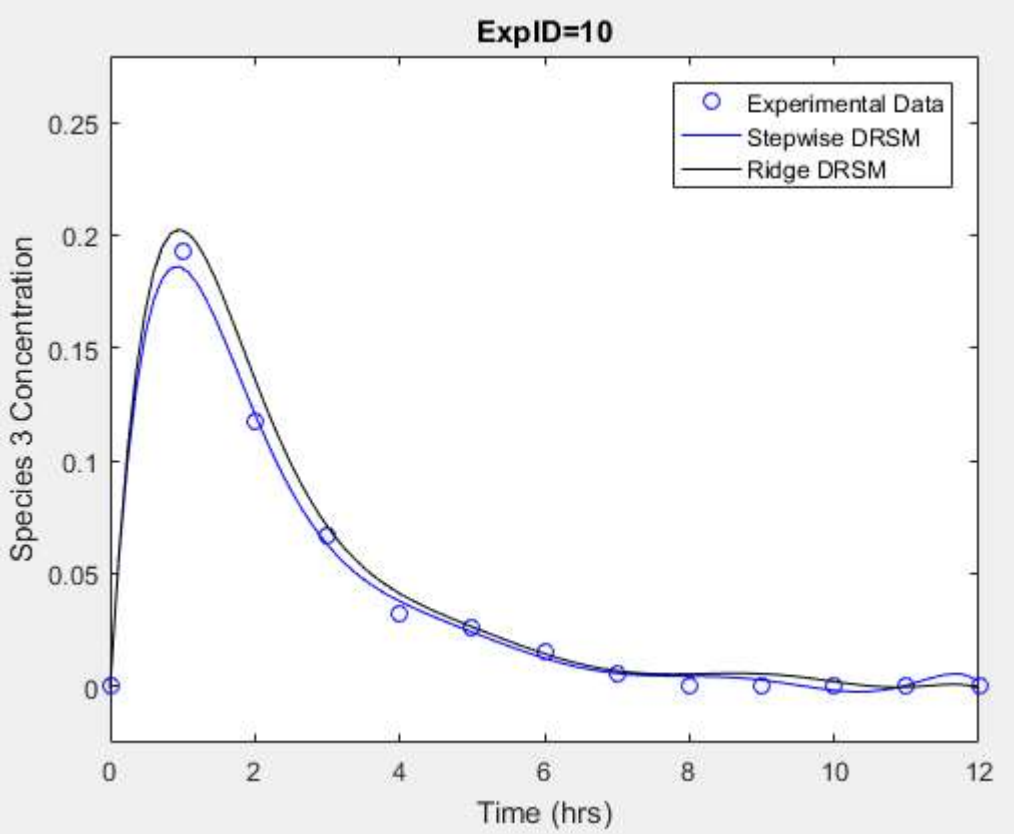

ExpID=5
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

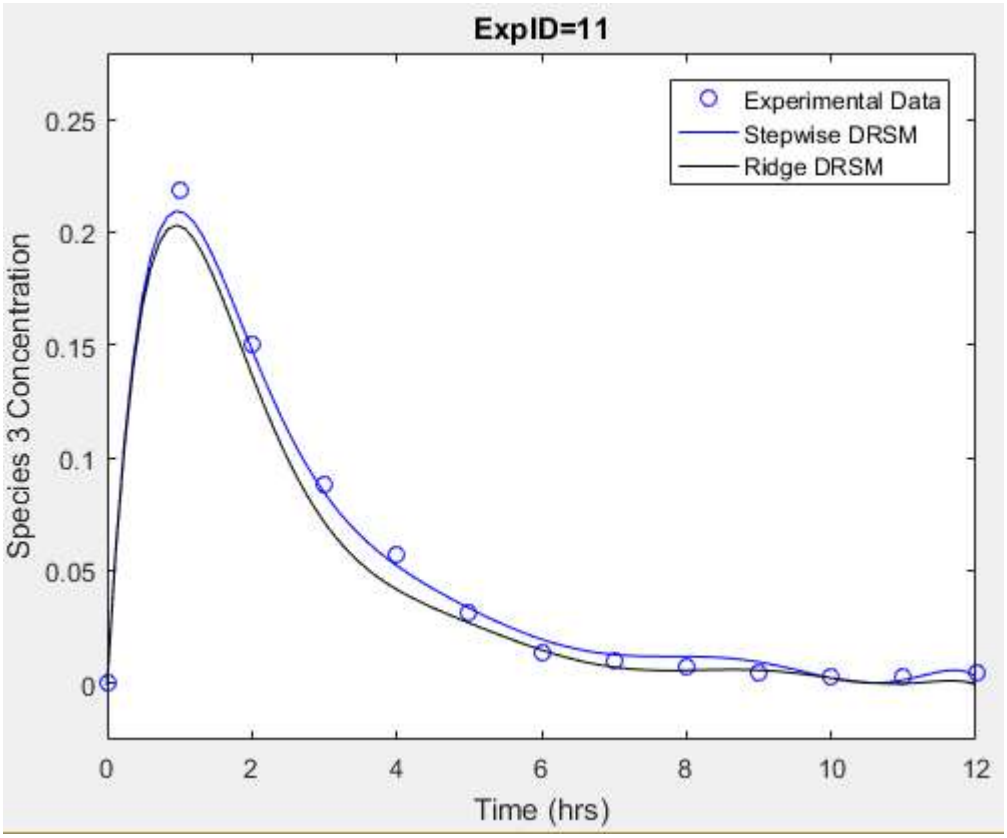

ExpID=6
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

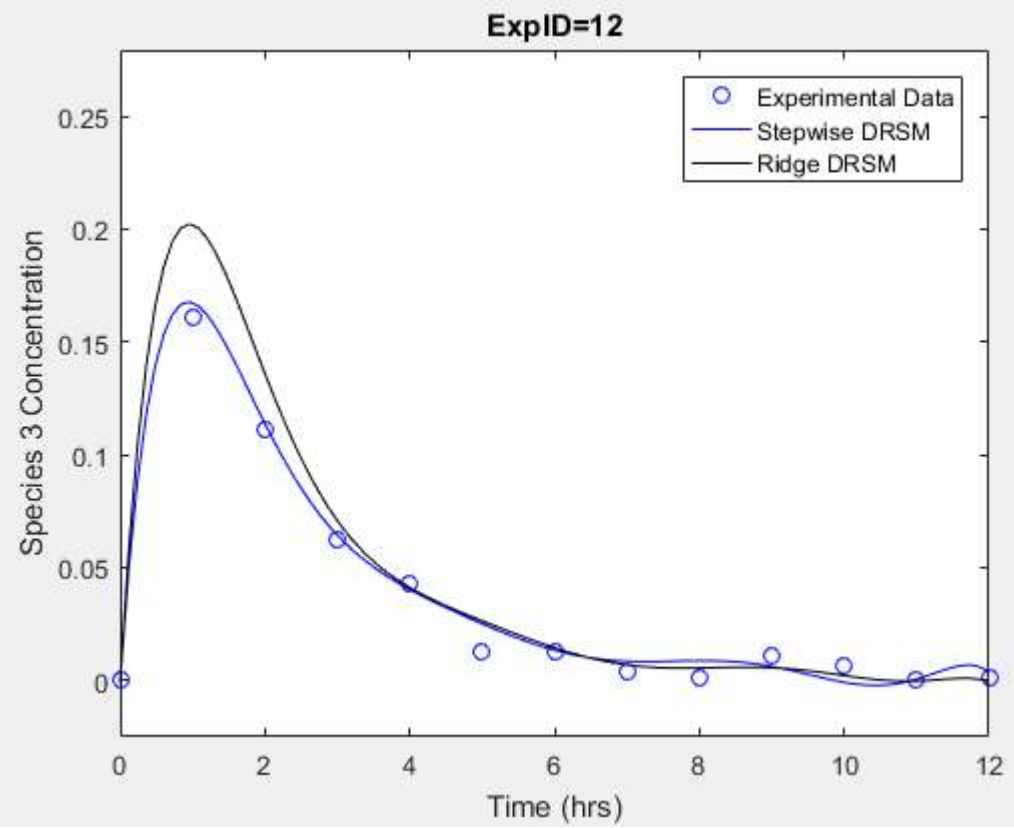

ExpID=7
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

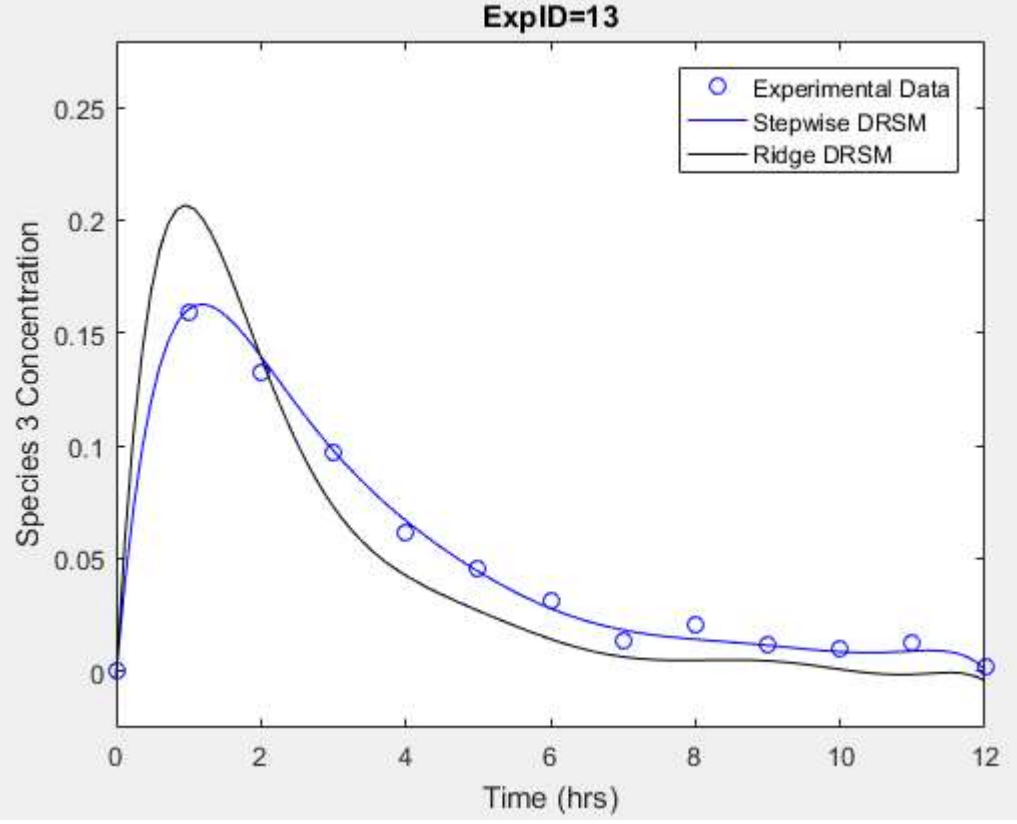

ExpID=8
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

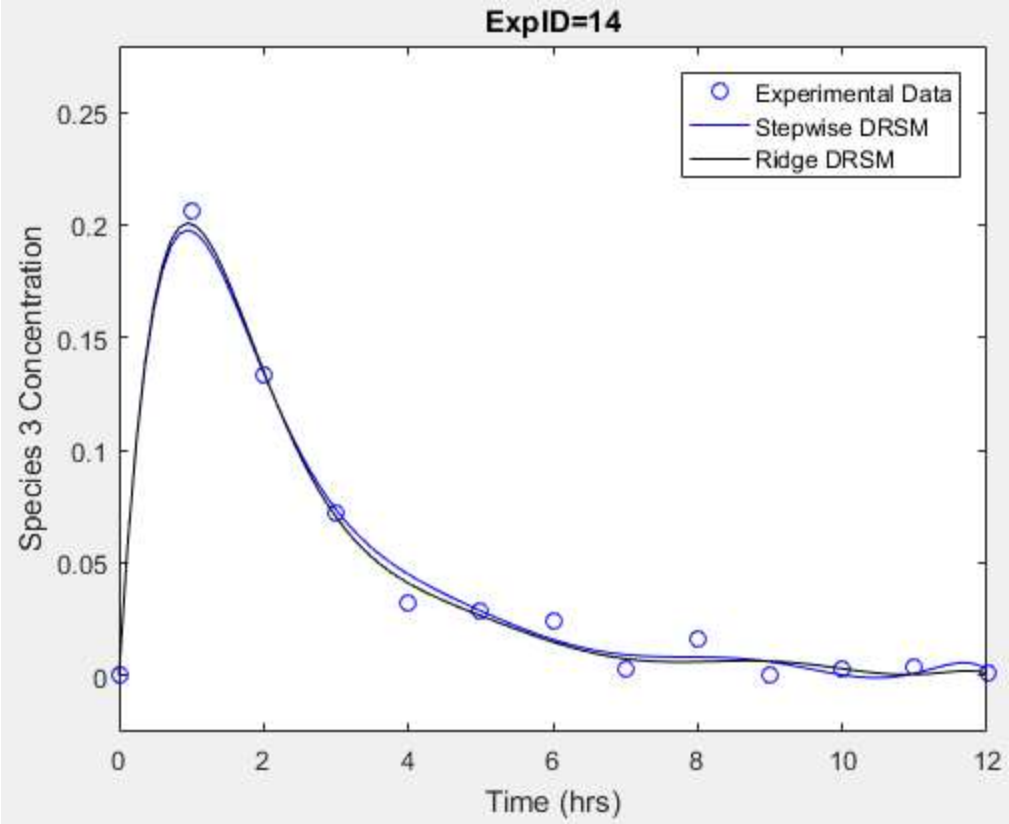

ExpID=9
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

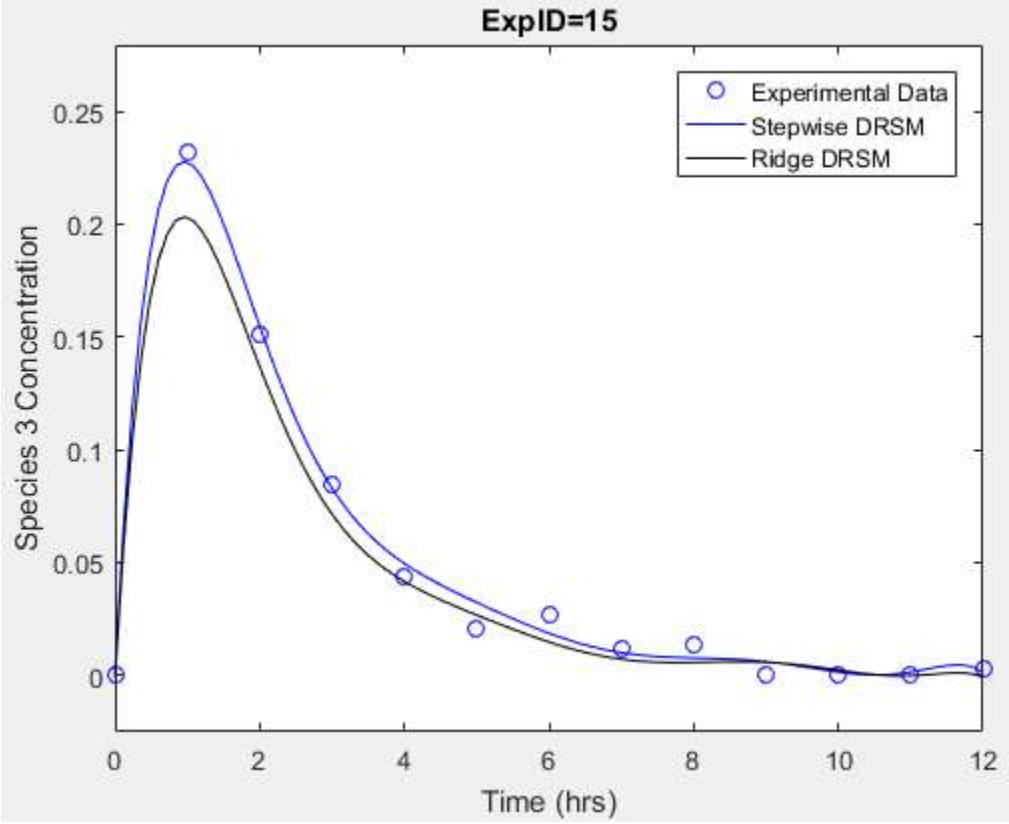

ExpID=10
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

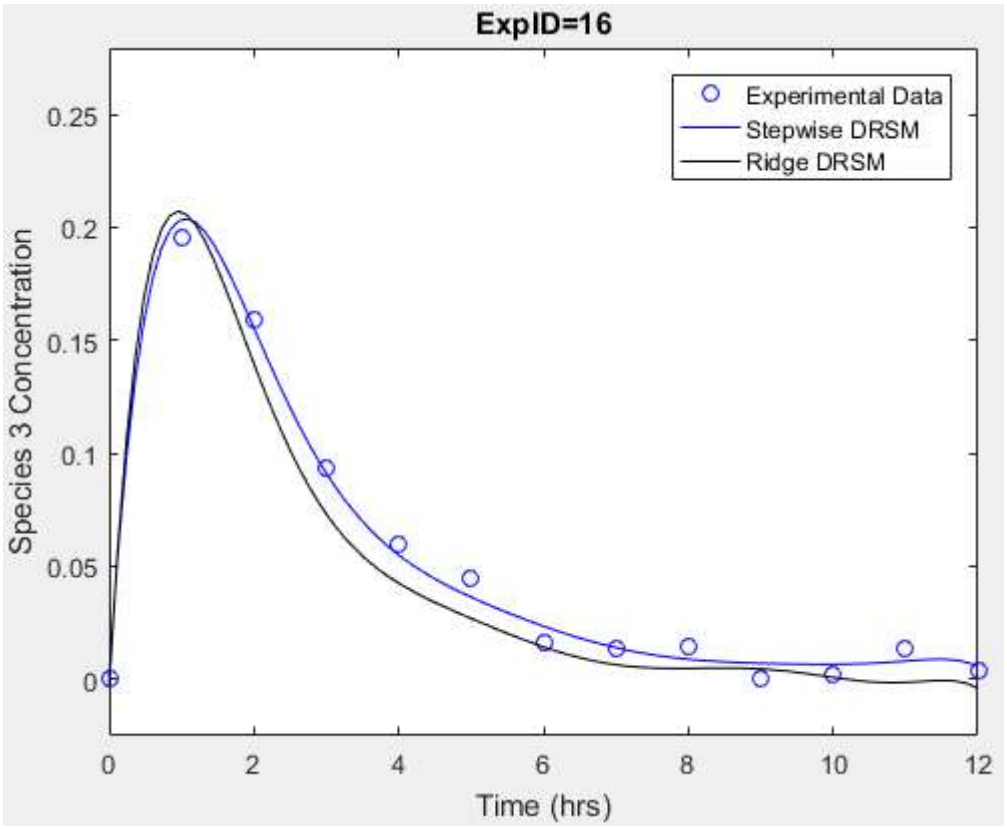

ExpID=11
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12

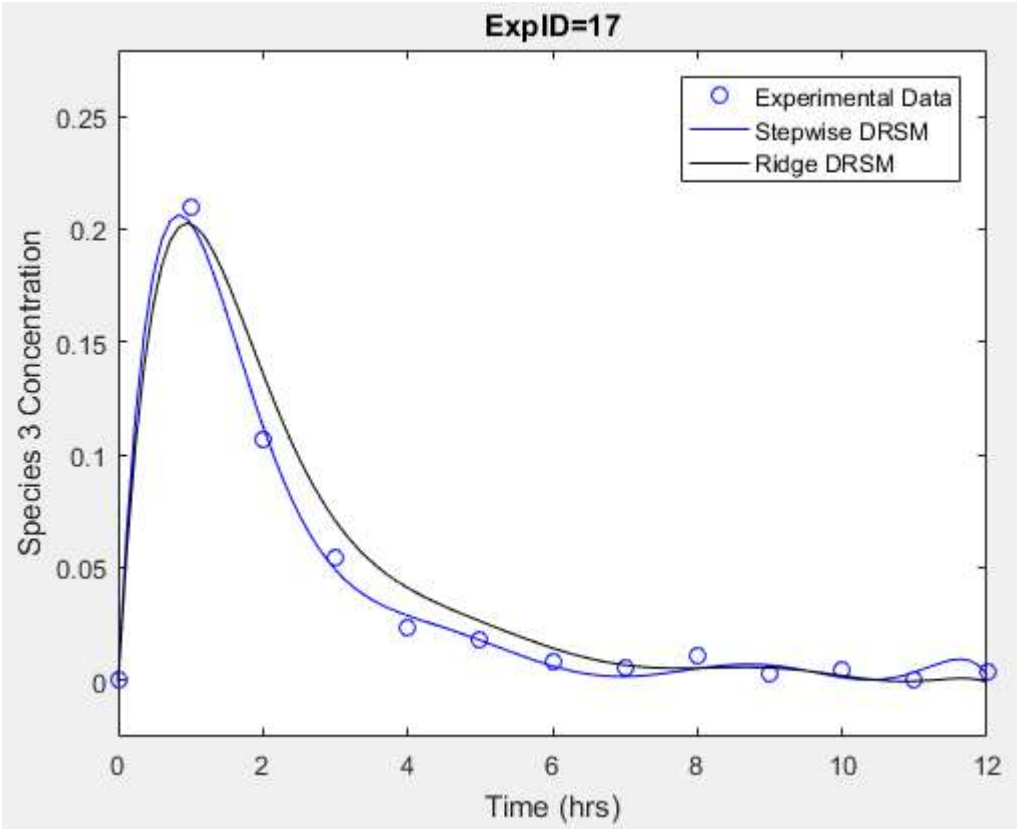

ExpID=12
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=13
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=14
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=15
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=16
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12


ExpID=17
Experimental Data
Stepwise DRSM
Ridge DRSM
Species 3 Concentration
Time (hrs)
0
0.05
0.1
0.15
0.2
0.25
0
2
4
6
8
10
12