4 Results & Analysis

The data produced by the first calibration (SS) could not be employed for the modeling as technical issues with the Tecan had caused noise to signal issues beyond acceptable ranges. After the instrument's issues were resolved, new calibration data was obtained (SynCom (SC) calibration data), which serves as the primary model creation and testing dataset. After the creation of the models, the predictions of the bacterial composition required a method to be developed to reduce the noise in those predictions. Further analysis of these models and their behavior led to considerable improvement and provides a future strategy for modeling the phycosphere system.

4.1 Modeling III

Although the visualization in Figure 4.1 of the SS calibration data does initially not suggest that the Tecan's measurement data are flawed, upon comparison with the SC data (Figure 4.6) differences in their predictions during analysis suggest otherwise. Although the SS calibration models are not used for the analyses, they nonetheless served to establish much of the code and methods for handling and investigating the SC calibration data. The early calibration data also helped guide the selection of one regressor from the one-hundred measured wavelengths and cut down the time spent on model selection. The initial coding base and methods, along with the selected regressor, are applied to the SC calibration and the appropriate models are built for analysis. The resulting models show that their application to the cell-based absorbance data is not as trivial as in chemical or macromolecuar data. Therefore, a solution was engineered to account for the observed decrease in bacterial signal predictions.

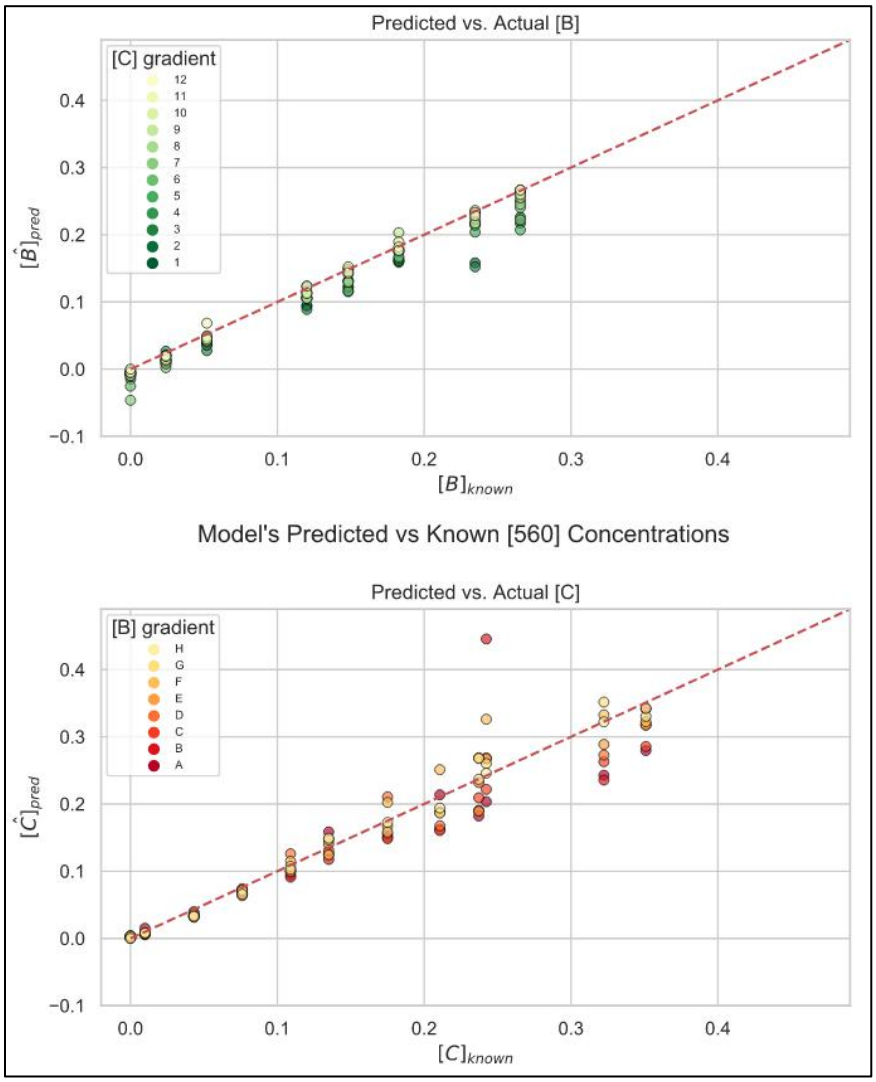

Figure 4.1: Single strain calibration prediction results. Models with these strains exhibited noise which did not allow their accurate application.

4.1.1 Single-Strain Calibration Contributions

The first models constructed were used for the analysis of screening experiment 4 (S4) data. During the preliminary analysis of S4 it was unknown which wavelength would serve best as a proxy concentration and all wavelengths of the measured spectrum were modeled $(A_\lambda \sim A_{680} + A_{750})$. Browsing through the resulting model's predictions against the known concentration, it becomes evident that one regressor wavelength would suffice since only subtle differences in the range of the predictions themselves are present in the majority of these models (Supplementary Materials). To avoid choosing a regressor wavelength arbitrarily, the scientific literature was consulted for information regarding the optimal wavelength to serve as a proxy concentration measurement. For Chlamydomonas, this meant

using A_{680} [37, 42], which the spectrum data in Figure 4.3 confirms, as there is a spike at 680nm, as is expected given Chlamydomonas's chlorophyll content. Although 680nm is not the maximally absorbing peak, it is the sharpest peak as the two other peaks are overlapping and broad. Continued probing for a representative absorbance wavelength to use for subsequent analyses reveals that neighboring wavelengths produce similar values as noted in the continuous nature of the absorbance spectrum(Figure 4.3)**.** The wavelength with the least correlation to both A_{680} and A_{750} with respect to each of S4's wavelengths was obtained using the Pearson's correlation function from the Python's Pandas package. A special function was written to find the sums of minimums correlation values which reveal the minima at A_{560} as seen in Figure 4.2. In the absorption spectrum of Chlamydomonas (Figure 4.3): 560nm is the lowest measured absorbing wavelength, the lowest point between the two maximum absorbing peaks, and the furthest from the two main measured wavelengths (680 and 750nm) for the S4 models. Given these observations, A_{560} is chosen to serve as the absorbance wavelength at which to represent concentration for all other work conducted.

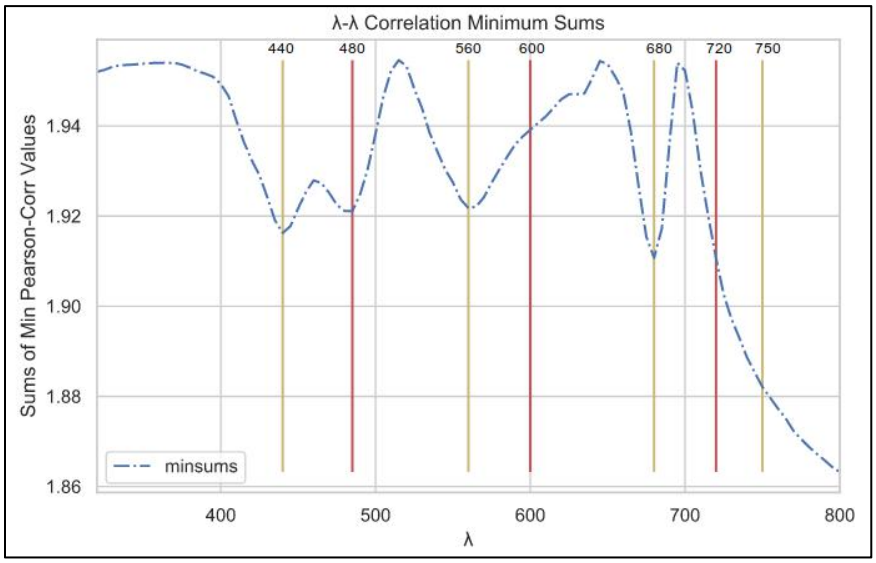

Figure 4.2: Sums of Minimum Correlation Values to A_{680} A_{750}. Helps establish default wavelength to use as the regressor for the models using SS data.

4.1.2 SynCom Calibration

After the new calibration data (SC) set was obtained, models were created to be applied to the data. These models are derived from System 2 below and solved for the concentration of Chlamydomonas ($[C]$) and Strains ($[B]$) as illustrated with System 1 in Methodology. These variables are denoted with the 560 subscript to symbolize that the predicted concentrations are absorbance measures at that particular wavelength. Each data set (Screens or PBR) is

measured at explicit wavelengths and also dictates which wavelengths to extract extinction coefficients from and subsequently model Table 2.1 The R^2 value does not fall below 0.995 for any of the coefficient determination regressions, however, this metric is only good when the individual regressions are used to regress back to each measured component. Nevertheless, they do inform on the extinction coefficients themselves being calculated precisely.

$$A_\lambda = k_{C\lambda}[C]_{560} + k_{B\lambda}[B]_{560} \qquad \text{System 2}$$

$$A_\lambda = k_{C\lambda}[C]_{560} + k_{B\lambda}[B]_{560}$$

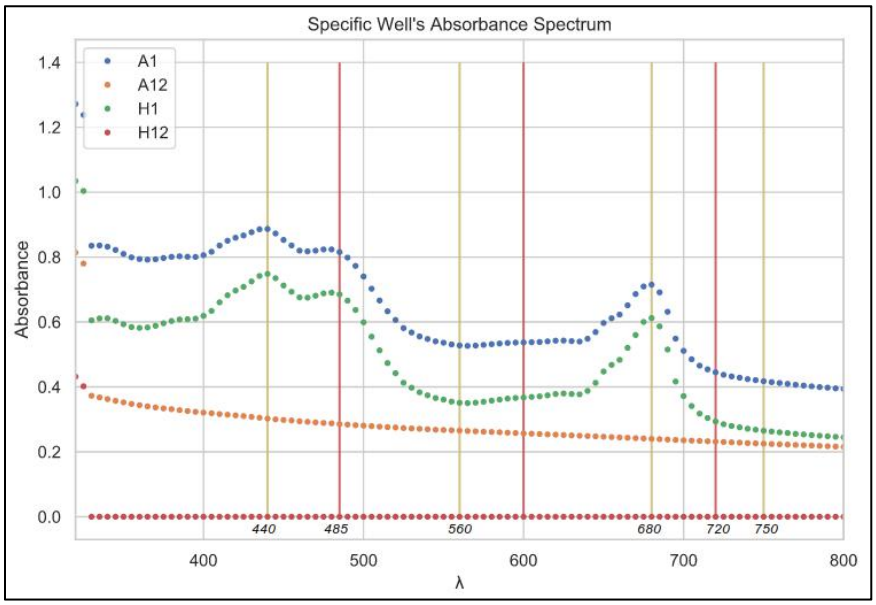

Figure 4.3: Absorbance spectrum of select Chlamydomonas:bacteria solutions. A12 and H1 denote the highest concentrations of either bacteria or Chlamydomonas respectively. A1 is the mixture of the both of these highest concentrations of these two components.

Models	Applicable Data Set	Model #
$A_{560} \sim A_{680} + A_{750}$	S4 Data	Model 1
$A_{560} \sim A_{680} + A_{730}$	PBR-to-Tecan Conversion	Model 2
$A_{560} \sim A_{680} + A_{720}$	PBR 1-4 Data	Model 3
$A_{560} \sim A_{600} + A_{750}$	S1,S3	Model 4
$A_{560} \sim N/A + A_{750}$	S2	Model 5

Table 2.1: Models Needed to be generated for application

The models created are first applied to the calibration data which is not used for the regressions (A-G:1-11). As is explained in Methodology 3.4.4, those mixtures have a known

concentration of Chlamydomonas and Strain as measured by A_{560} and correspond to the values used for the SLR regressions. The Beer-Lambert law states that the models should return the known and respective values at row H and column 12 as predictions and that their sums ($[B]_{560} + [C]_{560}$) should be equal to the empirical measurements (A_{560}) of their respective wells. When the sums of the predicted values are plotted against the empirical measurements, they should line up along the $x = y$ line as seen in Figure 4.4. For Model 1, the sums of the component predictions have an R^2 value of 0.9972; on average, each predicted sum is only slightly overestimated by 0.023 absorbance units. However, the actual component predictions are quite variable for the bacterial component and suggests Chlamydomonas being responsible for the discrepancy. As seen in Figure 4.6, the gradient of Chlamydomonas in each row, represented by green color gradient, has the general effect of forcing the models to both over-predict and under-predict the Strain component's content.

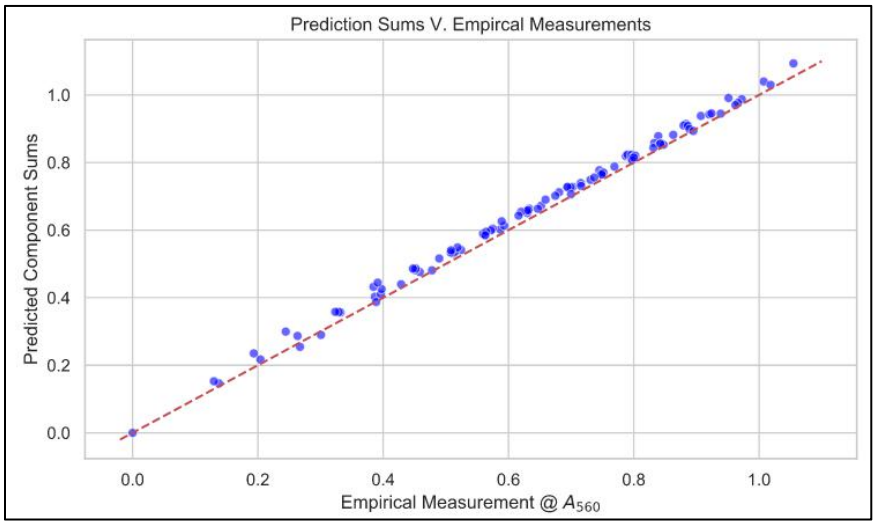

Figure 4.4: Predicted component sums vs. Empirical measurements. The predicted sums of Model 1 show adherence to the Beer-Lambert law.

The predicted values in Figure 4.6 are expected to line up along the red $x = y$ line for each respective component. Instead of doing so, the predictions for $[B]_{560}$ are distributed along the y-axis with an average standard deviation of 0.033 absorbance units. The predictions for $[C]_{560}$ fare much better and have an average standard deviation of 0.006 absorbance units. Given the relatively small magnitude of absorbance measurements in general, the standard deviations for the bacterial components do not warrant their use for the analysis of the given data, as any prediction would have an associated error range of about 10^8 cells per cell count regressions. Although the application of the models is not meant to provide a quantitative measure of the contents of the phycosphere, it would be preferable if they provided a sound

estimate to form and test hypotheses. These would allow to quickly iterate over experimental designs without constantly measuring the cell content. A general solution is devised and expanded upon in section 4.3.2 that corrects the observed dependence.

4.1.3 Analysis of General Model Behavior

Figure 4.6 shows that the interaction of light with two components of the SynCom is not entirely linear for the Strains, as the Beer-Lambert law-based models should have produced. This observation is true for Models 1-3 and it is not unexpected since the usual application of these types of models is towards determining concentrations of chemical compounds or macromolecules,[37] which are less complex and orders of magnitude smaller than cells. As is the case in any system studied through spectroscopy, light will not only interact with the analyte of interest; instead, every compound and macromolecule comprising the system up to and including the cell's internal contents interact with the incident wavelength of light. The level of interaction or the affinity with which an analyte absorbs light is determined by its respective extinction coefficient [38]. The high correlation between the real and predicted Chlamydomonas concentration predictions are largely due to the chlorophyll content, as it is responsible for absorbing a large fraction of light at 680nm, represented by the larger slope in Figure 4.5. It is important to note that even at the chosen concentration (A_{560}), three of the extinction coefficients are almost equivalent. As the models apply the coefficients to the given input, the similar light absorption of those wavelengths and Chlamydomonas's larger size have the net effect of attributing absorption to Chlamydomonas, causing the range of bacterial predictions seen in Figure 4.6.

4.1.4 Models 4 & 5

Visualization of Model 4 on the test data revealed the lack of accuracy of the predicted values (Figure 4.7). Possible explanations could include the use of flawed data to build the model or the inherent stochasticity of complex biological mixtures. Previous inspection of the data had not revealed any issues and therefore the relationship of the coefficients in the denominator was suspected to be affecting the predictions.

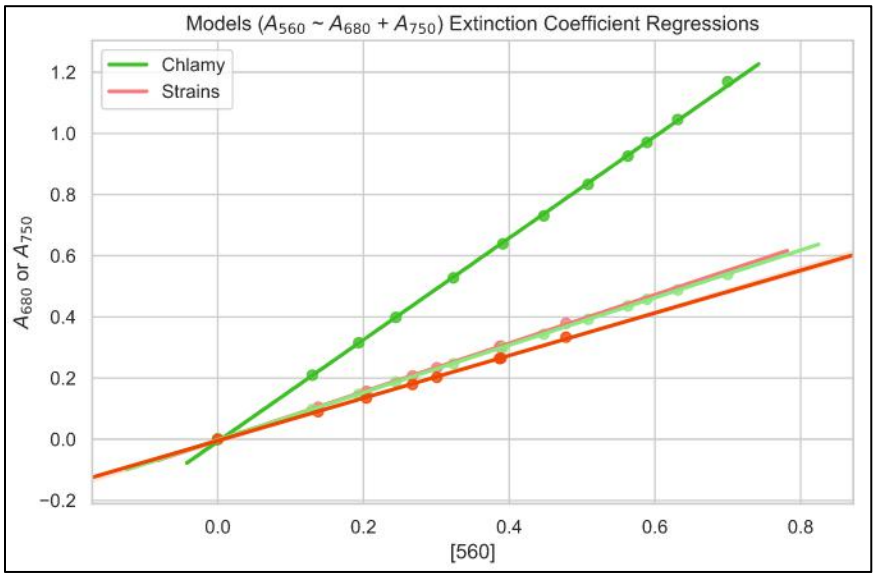

Figure 4.5: Extinction coefficients. The extinction coefficients of Model 1 are plotted against one another. Green and red coefficients represent Chlamydomonas and bacteria coefficients.

In Model 4[B] the coefficients in the denominator are composed of opposing member and wavelength values. The difference of these products and the extinction coefficient in the numerator are multiplied by the empirical measurements of the modeled wavelengths (600nm & 750nm for Model 4). The calculation of the denominator K yields 0.0079 for this model: The resulting proximity to zero indicates why the values are much larger than the predictions of Model 1 (Figure 4.6) This observation played a critical role in determining an optimized model and is further expanded up in Section 4.3.1. Given the inaccurate results of this model, it cannot be applied to S1 and S3 data; consequently, S1 and S3 data are no longer considered for further study. Moreover, S2 was measured at F_{680} and A_{750} –seeing that at least two absorbance measurements are required to generate these types of models, S2 is also excluded from further study.

$$\widehat{[B]}_\lambda = \frac{A_{600}k_{C750} - A_{750}k_{C600}}{k_{B750}k_{C600} - k_{B600}k_{C750}} = \frac{A_{600}k_{C750}}{K} - \frac{A_{750}k_{C600}}{K} \qquad \text{Model 4 [B]}$$

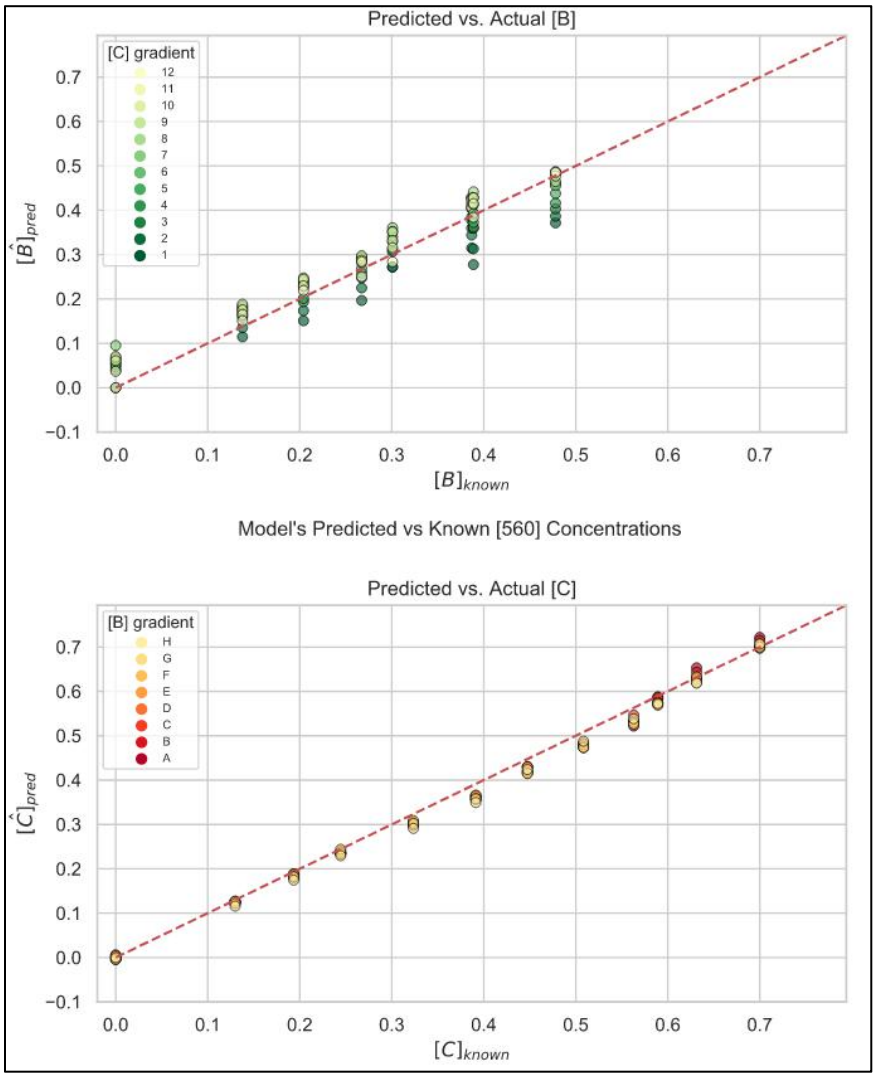

Figure 4.6: Component Predictions of Model 1: Predictions of Model 1show large variation the prediction of bacterial fractions [B], while the predictions for Chlamydomonas are much more accurate.

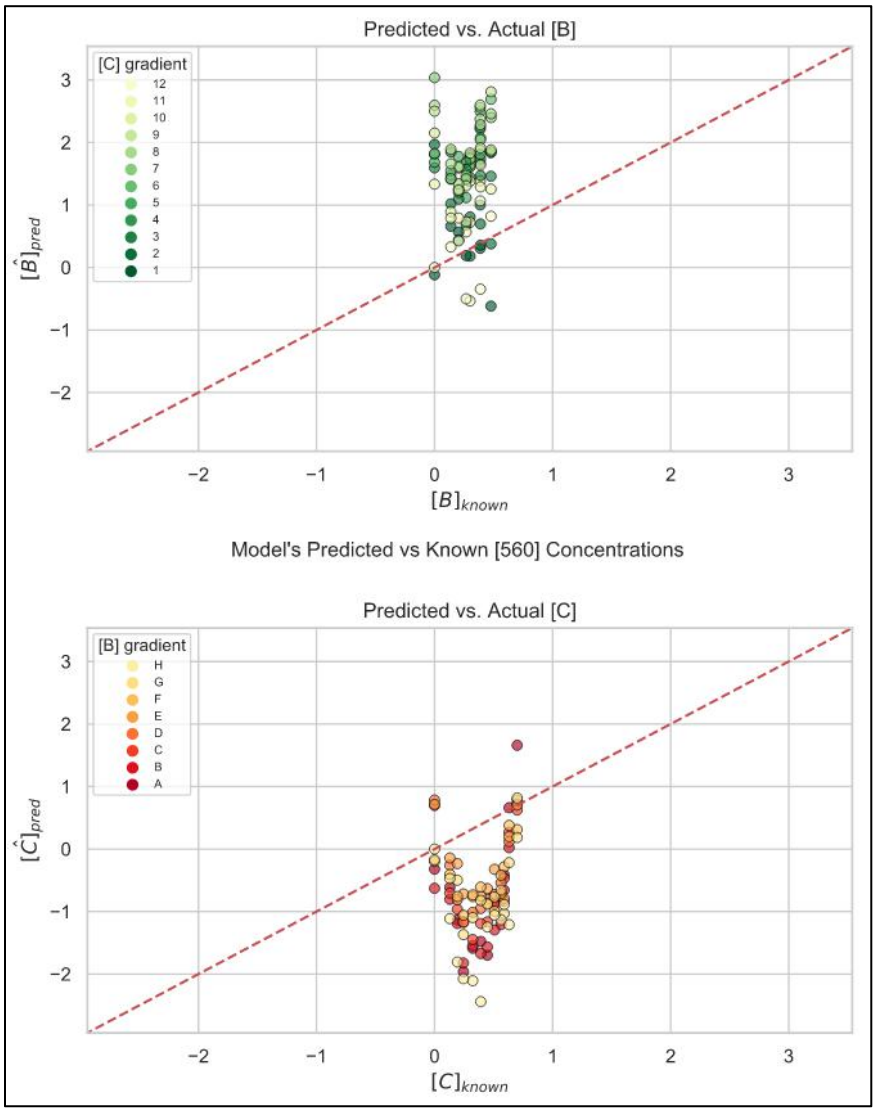

Figure 4.7: Component predictions of Model 4: The numerator's approach to zero in Model 4 force the predictions of this model to over-predict greatly.

4.2 Pivoting the Thesis Scope

Several factors forced an exhaustive search for an optimization of the Beer-Lambert models. First, the predictions for the Strain fractions had a relatively large standard deviation. Second, preliminary application of these models to PBR and S4 resulted in predictions falling under the Tecan's absorbance measurement limits ($A_\lambda < 0.1$) [41]. Last, the provided relative abundance measurements in PBR4's Metadata could not be directly compared to the preliminary ratios composed of absorbance predictions from Model 3: This comparison is necessary to validate the usability of the method for monitoring the phycosphere. The two types of organisms under study do not absorb light equivalently per unit cell making their absorbance ratios largely uninterpretable. This problem can be addressed by regressing

absorbances (A_{560}) to Coulter cell count estimates, and creating ratios from the regressed counts instead. To facilitate comparison of cellular fractions and relative abundances a regression table is provided (Appendix 1). However, the relative small values predicted from preliminary model application and the associated large standard deviation problems remained. These issues forced the project, originally centered around gleaning information for the respective taxa in the screening and PBR measurements, to change its scope from an analytical study to an exercise in mathematics and software engineering to improve the interpretation of future measurements.

4.3 Modeling IV: Engineering Solutions

After failing to derive meaningful insights from the application of Models 1-3, it became necessary to perform a more exhaustive study of the results of the SC calibration. The continued probing of the data yielded two key insights. The first was uncovering the nature of the interdependency of the Strain fraction predictions to the known concentrations of Chlamydomonas and Strains, which allowed the optimization of the Models 1-3. The second insight was derived from Model 4 and its predictions that stem from the difference of the products of the extinction coefficients, K, in the denominator of the models. The combination of these two results are combined into a general solution and modeling strategy.

4.3.1 Special K

Incorrect predicted values of Model 4 (Figure 4.7**)** could be due to small values in the denominator and the properties of zero. Comparison to Model 1's K (0.554230) pointed towards the existence of a better overall model. Specifically, it was hypothesized that if predictions diverted from the real values as K approaches zero, there must be a K_s in the space of all K that these data can generate, which would result in more accurate predictions. And since two coefficients K are produced for every model $(A_z \sim A_x + A_y)$, equal in magnitude but opposite in sign, then K_s should have the largest magnitude in this space. Furthermore, it must be noted that resulting K from models which conform to the following $(A_z \sim A_x + A_y)$ and $(A_y \sim A_x + A_z)$ are not equal.

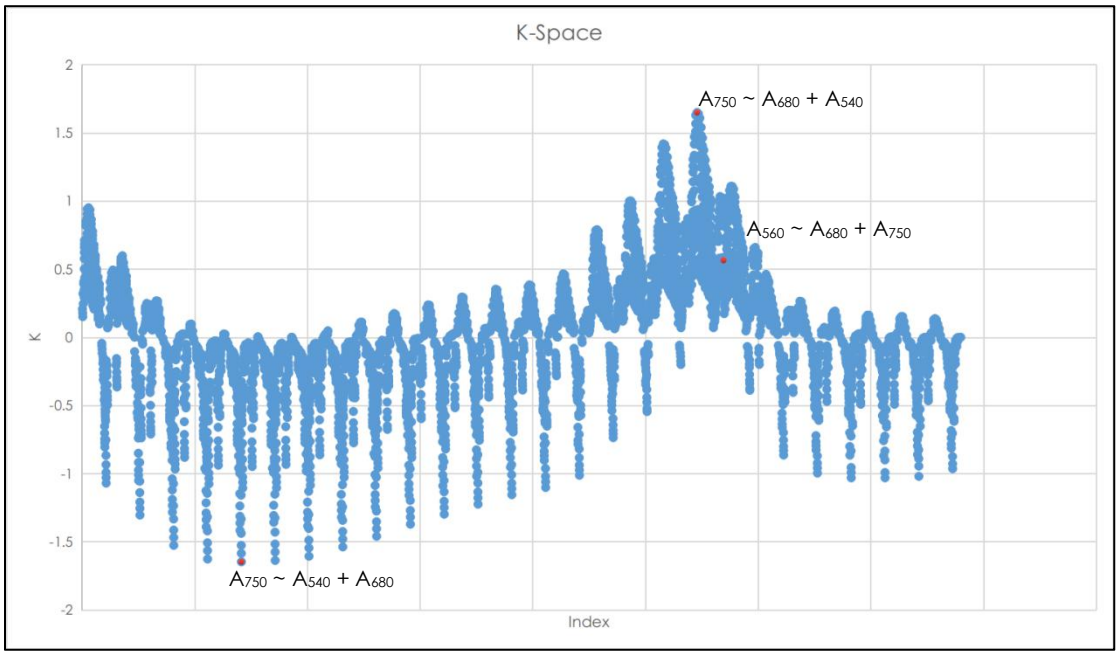

Figure 4.8. K-Space. The absorbance measurements at all wavelengths from were modeled against all others to derive the denominator of each model. The lowest and highest dots indicate the highest K. Model 1's relative position is also shown.

A brute force approach was used to extract the four coefficients required to calculate the K. The resulting K_s has a value of 1.64876 and is yielded by model $A_{750} \sim A_{680} + A_{540}$ (Model S) (Figure 4.10). In contrast, the models required for the analysis of the given data have the form $A_{560} \sim A_{680} + A_{7*0}$. Even though the wavelengths are similar, the impact that the regressor wavelength has on the predictions is better. The required models have similar wavelength to Model S but have a better separation of coefficients (Figure 4.9) in comparison to coefficients of Model 1 (Figure 4.5).

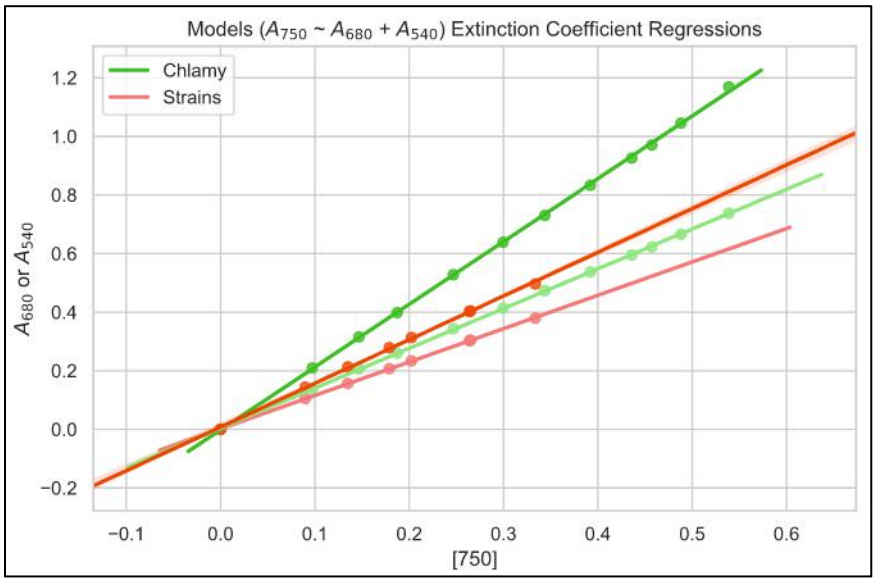

Figure 4.9: Model S coefficient Plots: The coefficients of Model S have a better separation than those of Models 1-5 and lead to the observation that the strength of the Chlamydomonas effect on the predictions of Strain content is proportional to the true concentration of the Strains in the System

Comparison of Models S and 1 show that the incorrect predictions of Model 1 are an effect of the wavelengths used to model the data as three of those coefficients are very similar to one another (Figure 4.5). While predictions of Model S do not fall directly on the x=y line, this model primarily underestimates the concentrations. These values indicate that the lack of accuracy is a function of Chlamydomonas' concentration on respective bacterial loads. The greater bacterial concentration is, the greater the fraction attributed to Chlamydomonas, lowering the Strain predictions. Second, there is considerably less variability for low concentrations of bacteria even at high Chlamydomonas concentrations, as noted by the predictions at [B]=0. These observations point to a clear relationship in how the models attribute signal to the opposing member, first noticed with Model 1 and further explained in Section 4.3.2.

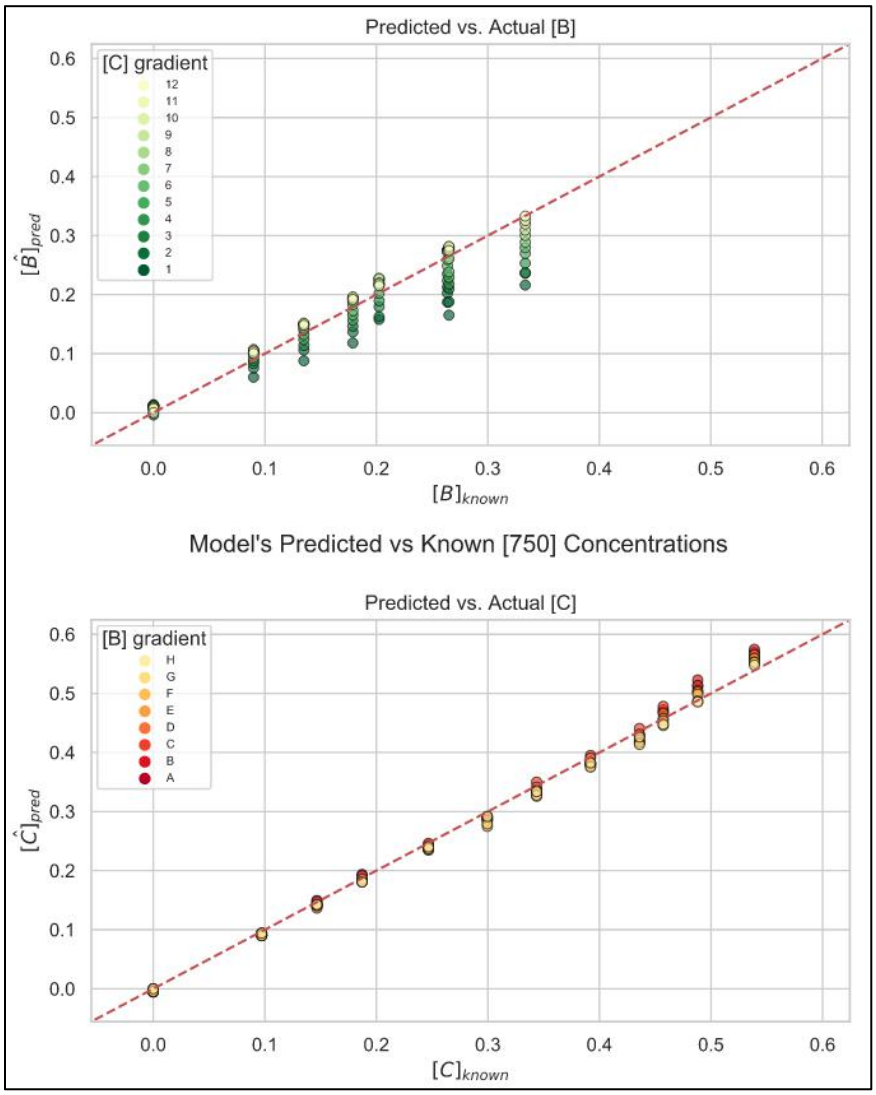

Figure 4.10: Model S Predictions: The resulting predictions on the testing data, show that Model S is far superior at distinguishing the true concentration of the Strains at low Chlamydomonas concentrations

4.3.2 Optimization of Model S

While visualizing the data of the bacterial predictions in a table that conforms to the 96-well position (Table 2.2) a slight 'bump' in the data was noticed in row H columns 6-8 of the Strain predictions—Row H has no bacteria, and only a Chlamydomonas gradient exists across the row and predictions at those wells for bacteria should be zero. With perfect models, the predicted value at any given row would be equal and a flat line at the known concentration would be observed (Figure 4.11). The parabolic trends were first observed with Model 3, but the following makes use Model S to illustrate the correction method. In Figure 4.11, the presence of Chlamydomonas imparts an error to the prediction with a severity that is proportional to the concentration of bacteria. As the true concentration of bacteria increases

(purple lines), the under-predictions become more severe with respect to Chlamydomonas content. To observe the net effect of Chlamydomonas on each of the predictions, the predictions at column 12 (where no Chlamydomonas is present) are subtracted from each prediction of the respective row, effectively shifting the parabolas downward and forcing them to pass through the origin (Figure 4.12). This allows the improvement of the predictions and if validated could be applied to Tecan measurements.

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.2161	0.2361	0.2371	0.2532	0.2695	0.2794	0.2886	0.2999	0.3091	0.3194	0.3256	0.3329
B	0.187	0.2025	0.2127	0.2235	0.2347	0.2485	0.2589	0.2747	0.2759	0.2741	0.2723	0.2738
C	0.1652	0.1878	0.2098	0.2178	0.229	0.2383	0.2606	0.2757	0.2783	0.2804	0.2816	0.2745
D	0.1579	0.1625	0.1794	0.1895	0.202	0.2159	0.2248	0.2267	0.2275	0.2191	0.2189	0.215
E	0.118	0.1366	0.1464	0.1562	0.1652	0.1726	0.1831	0.1882	0.1943	0.1928	0.1961	0.1917
F	0.0878	0.106	0.1136	0.1236	0.1319	0.1403	0.1435	0.1489	0.1487	0.1507	0.1516	0.1479
G	0.06	0.0755	0.0833	0.0881	0.0928	0.1001	0.1038	0.1036	0.107	0.1042	0.1046	0.1005
H	-0.004	0.0045	0.0096	0.0112	0.0123	0.0135	0.0122	0.0114	0.0099	0.0085	0.0068	-0

Table 2.2: A_{560} **Strain fraction prediction.** In row H there is a visible increase towards the middle values. This shows that Chlamydomonas imparts a parabolic shaped noise to the predictions.

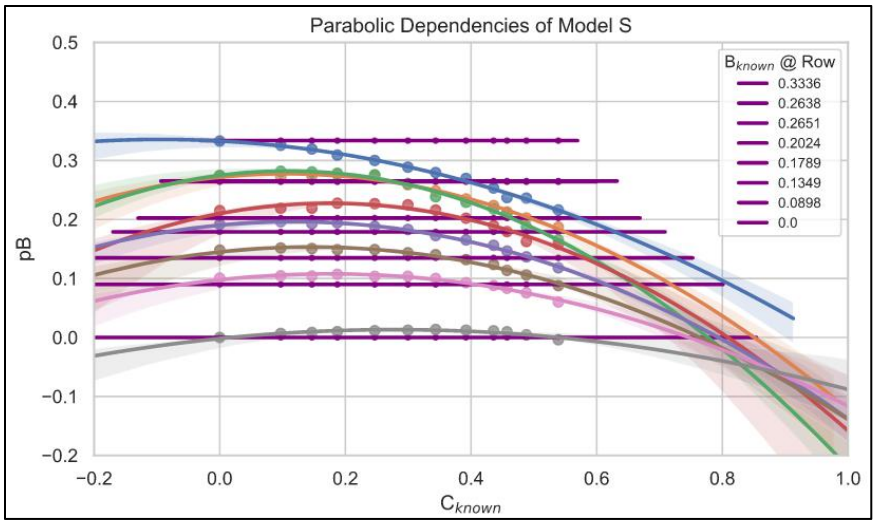

Figure 4.11: Model S Predictions. Known Strain [B] concentrations (purple) superimposed on the predictions [pB] yielded by the with respect to know Chlamydomonas concentration.

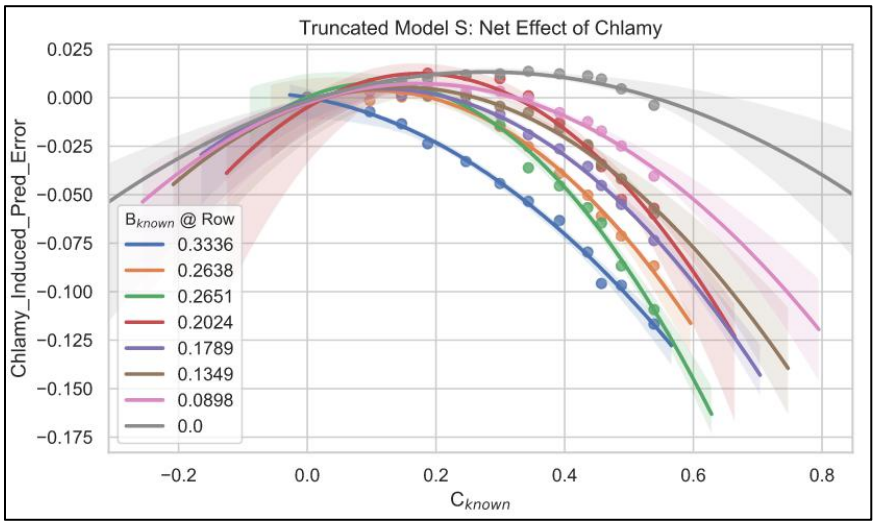

Figure 4.12: Model S Predictions Truncated at origin. The Strain fraction predictions of column 12 is removed from teach of the predicted values, forcing the fitted parabolas to intercept at the origin

$f(x) = a_2x^2 + a_1x + a_0$	Standard Equation of a Parabola
$f(x) = a_2(x-h)^2 + k$	Vertex Form of Parabola Equation: V = (h, k)
$f(x) = a_2x^2 - 2a_2hx + a_2h^2 + k$	Relationship between the Equations above
$f(x) = a_2x^2 + a_1x + 0$	Truncated parabola
$a_1 = -2a_2h$ $h = a_1/2a_2$	Relationship of h and a_1

Table 2.3. Relevant parts of parabolic functions.

The modifications to the model require making the assumption that the observed parabolic trends (Figure 4.12) come from a parabola having a unique breadth parameter (a_2). The next step is to model the observed relationship between each fitted parabola's vertex (h, k) where h is a theoretical Chlamydomonas concentration and k is the predicted bacterial fraction (Figure 4.11) with an apparent linear relationship (Section 3.7). The apparent movement of the fitted parabolas' vertices h (Figure 4.11) which traces a path up and leftwards, validates the linear relationship. By extension, at any true concentration of bacteria $[B]$, a theoretical vertex exists of a parabola which lies along the path which represents all the possible predictions that can be made (Equation 12 & 13, Methodology). Given that true bacterial concentration and a gradient of Chlamydomonas—lower resting Parabolas thus have a lower rate of change than higher resting ones with respect to Chlamydomonas concentration. When these theoretical parabolas are subsequently forced to pass through the origin, the Chlamydomonas concentrations help determine the extent of this influence and the value is

either added or removed from the original prediction depending on whether it lies below or above the x-axis as seen in Figure 4.12.

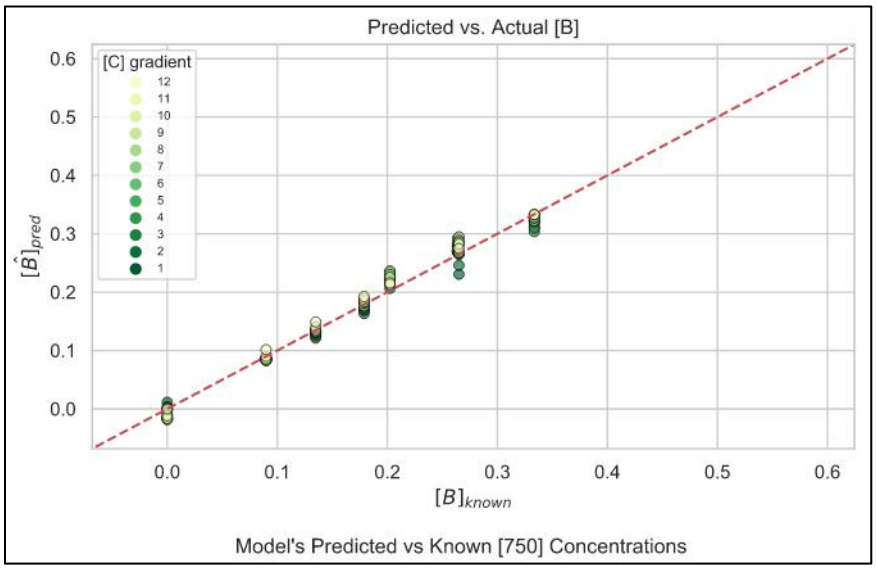

Figure 4.13: Corrected Results of Model S. There is a marked gain of predictive power in the models especially beyond the Tecan's 0.1 detection limit

4.4 Pivoting Back to Analysis

Strain predictions tend to be either over-predicted or under-predicted by the Chlamydomonas gradient with a strength proportional to the true Strains concentration (Section 4.3.2). The optimization of Models 1-S is meant to increase the accuracy of those model's predictions. Correction factors resulting from the model optimization are yet to be validated experimentally, nevertheless in the following analyses they are applied to the models' predictions. As the calibration data pointed out, the under-predictions were likely caused by Chlamydomonas effect and made the analysis at that time prone to errors

Using the correction factors, the analysis of SC4, PBR2-4 and PBR4's Metadata was conducted**.** Analysis of the SC4 data consists of first resolving the signal between the components and searching for the Strains that exhibit interesting interactions with Chlamydomonas, given a community or binary context and a control consisting of Chlamydomonas's supernatant. The PBR data serves to create a tool which can resolve the signals of the component's concentrations in real time as PBR data is collected continuously. Before creating a tool, the models' applicability to the PBR data must first be established and shown to provide similar predictions on the same sample's measurement by either Tecan or PBR. PBR2-3 data are used to establish the transferability of the models. Establishing this is

key in analyzing PBR4's Metadata which consists of daily cell counts and 16S sequencing relative abundance measurements. The goal is to show that through the successful application of the models to the PBR data, the predictions can be mapped to the observed relative abundances from the 16S sequencing. This would reduce the need to perform sequencing and lead to optimized experimental design.

4.4.1 Analysis of Screen 4

This project was originally centered on finding Strains that exhibited noteworthy trends, such as helping Chlamydomonas grow or finding those that grew best with Chlamydomonas. The expectation was that resolving their signals would yield such trends and a more nuanced experimentation with the resulting Strains could be undertaken. Towards this extent the Beer-Lambert law is employed; however, as it has been established, the resulting models' usability was not as straightforward. During the preliminary analysis of SC4, the predicted values from these models were considered too low ($A_{560} < .1$) to allow for a significant analysis and required finding a solution in order to assess the data for Strains of interest. Once the correction factors were calculated, the models' predictions were corrected. Unfortunately, only SC4 can be analyzed in this manner (Section 4.14).

The S4 experiment was conducted to answer two major questions in determining potential candidates for further experimentation: How does the Strain-to-Chlamydomonas cell count inoculation ratios affect either component's growth; and if the individually tested Strains' depend on the presence of Chlamydomonas or other bacterial strains. It was expected that at least some of the Strains should exhibit differential growth and would show up as higher in the predictions using these models.

Upon comparing the corrected and uncorrected versions of the predictions for the data, no significant changes are observed between them (Figure 4.14). Considering the low concentration of Chlamydomonas in the system, the observed Chlamydomonas effect only manifests slightly making the corrections consistent to what is expected. Furthermore, the models have a tendency to over-predict Strain content at low known Strain and Chlamydomonas concentrations. For example, Model 1 ($A_{560} \sim A_{680} + A_{750}$, Figure 4.6), used for this analysis, consistently over-predicts Strain concentrations; at low concentrations of Chlamydomonas ($A_{560} < .1$), however, the predictions are minimally over-predicted, making these small corrections also consistent with the calibration data.

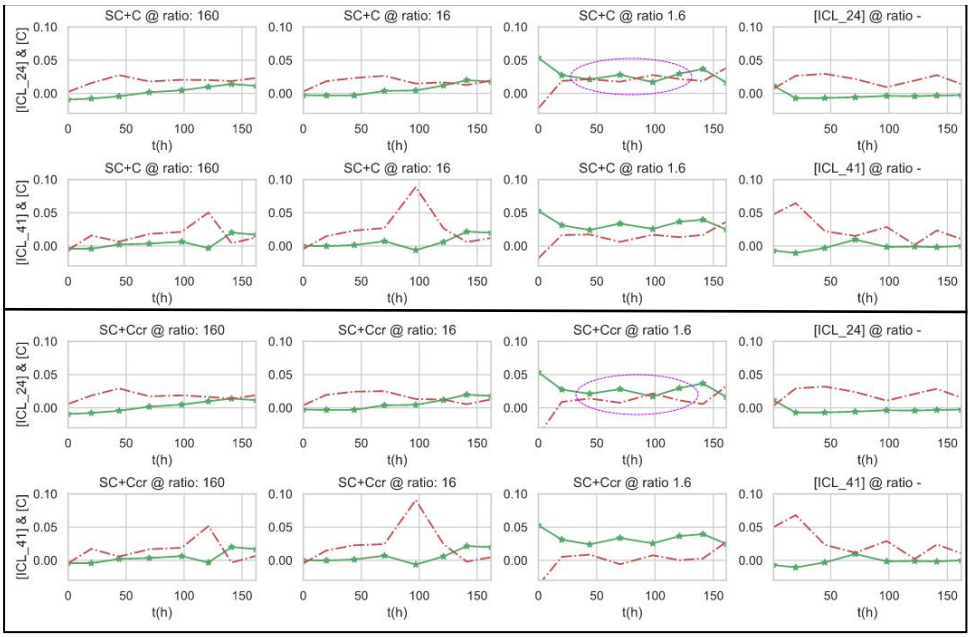

Figure 4.14: Select screening results. Top represents uncorrected values, bottom are the corrected values. At the levels measured the correction factors push the bacterial concentration lower as they tend to be over-predicted (purple ellipse). Green curves represent Chlamydomonas

Thirty bacterial strains were tested and their predictions are for the most part under 0.050 absorbance units. Given the almost unchanged nature of the corrected results, the data itself is further reviewed and points to the issue of low predictions being a product of the empirical measurements themselves. The empirical absorbance data (A_{680}, A_{750}) falls into the range of 0 - 0.192 and 0 - 0.181 absorbance units respectively, however, the vast majority of measurements are considerably lower than this and have an average of 0.030 and 0.015 absorbance units. The spikes in certain results (e.g. ICL_41 at 100Hrs, ratio=16; Figure 4.15) likely represent errors in the measurements as they only appear during at that time for many other Strains and immediately shoot back down in the subsequent measurement. Had growth truly been observed, the signal increase would have continued over for more than a single measurement.

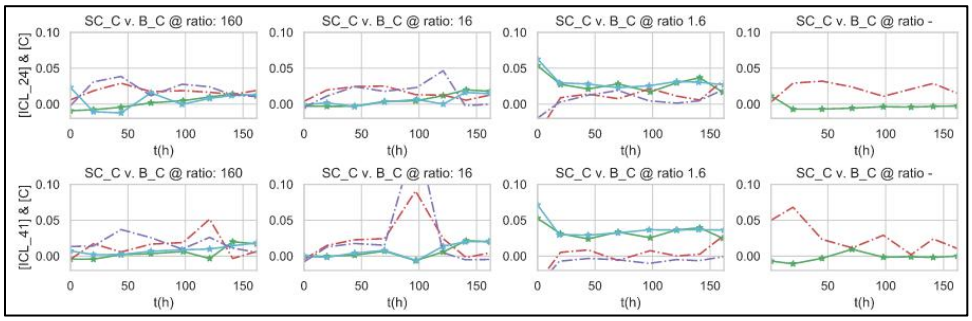

Figure 4.15: SynCom vs Binary Composition. This experiment does not show growth. Green and teal lines represent Chlamydomonas and red and purple represent Strain.

4.4.2 Model Transferability to PBR

Verification of the transferability of Model 3 to PBR data is conducted to ensure that the envisioned monitoring tool will accurately assess the system's contents. Gaining the ability to do so would assess in real time the dynamics of the phycospheres. Here, models for the PBR are created using the coefficients from the Tecan measurements. Per the Beer-Lambert law, any calibration, whether through the PBR or Tecan, should yield the same extinction coefficients once corrected for the path length. Correction of the PBR's much longer path length requires the Tecan-derived coefficients to be multiplied by the path-length (2.7 cm) of the vessels as demonstrated in Model 2PBR. However, when this was applied to the PBR measurements, the data was severely under-predicted. It is likely that the system also makes use of a path-length correction algorithm much like the Tecan to correct the path length internally. Hence, the extinction coefficients are applied as obtained from Tecan and Model 2 is used without modification, instead.

$A_{720} = 2.7 \bullet k_{B720}[B]_{560}$ $A_{680} = 2.7 \bullet k_{B680}[B]_{560}$ $A_{720} = 2.7 \bullet k_{C720}[C]_{560}$ $A_{680} = 2.7 \bullet k_{C680}[C]_{560}$	Model 2PBR System

The system is solved for each of the two components (omitting the path length) and the data is entered into the models to retrieve the predictions. In theory, the predicted values from PBR measurements at sampling time should equal the sample's predicted values given its Tecan measurements since the aliquot taken at sampling time retains the same concentrations of the components. Once the models have been applied, the predicted values for each component should line up along the $x = y$ line, have a positive correlation if growth is continuous, or alternatively, have strong precision about the $x = y$ line if the system's growth is constant.

After applying the model to selected vessels from PBR2, the results show growth in those vessels and indicate that the predictions on both systems are similar. However, the limited amount of data used to validate the transferability of the models requires that more data is used to establish this more accurately.

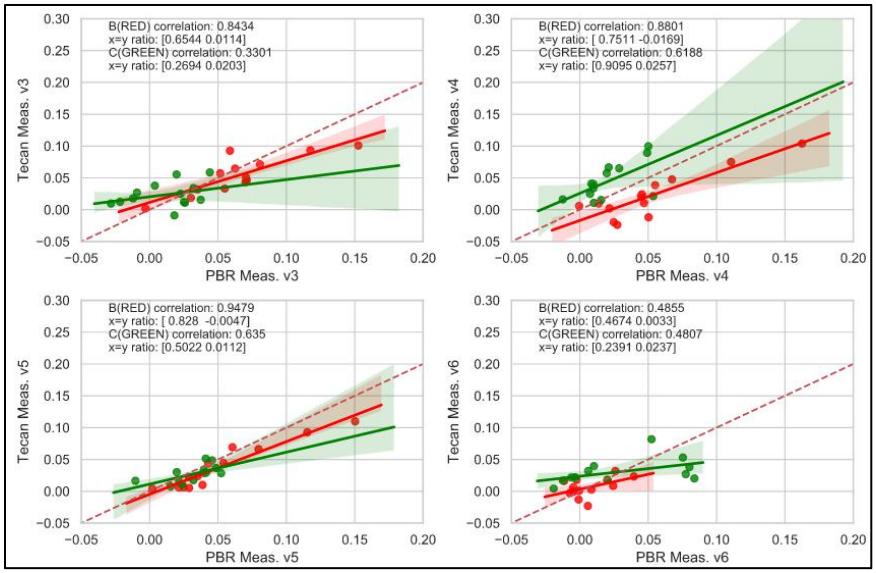

Figure 4.16: Select results from PBR2 transferability experiments. The results indicate that indeed the models are able to be taken from the Tecan to the PBR.

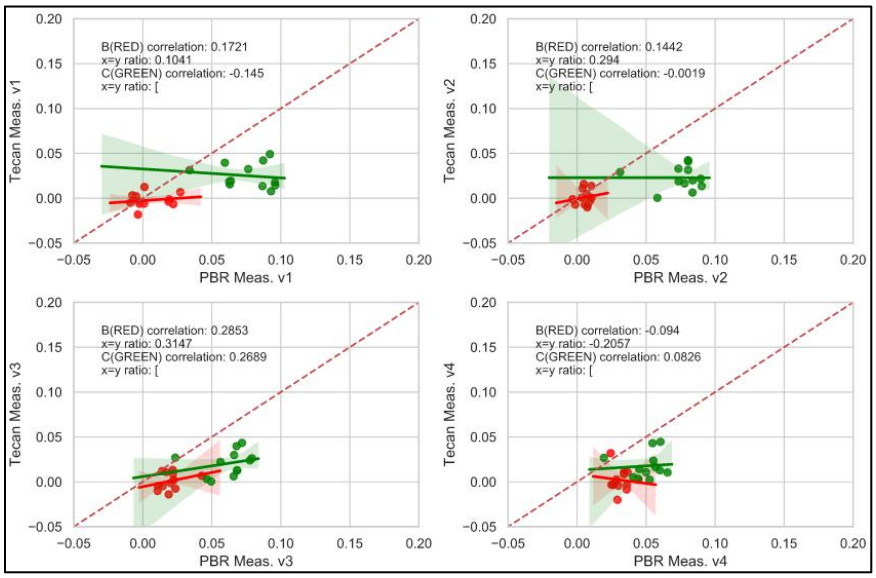

Figure 4.17: Select results from PBR3 transferability experiments. This experiment was stopped two-weeks post inoculation. The results seem to indicate constant growth.

Not only does the limited amount of data limit the validation of this method for monitoring the phycospheres, but the data was also measured at different wavelengths than the ones used to create the model. PBR2-3 data was collected before the start of this project and were measured at the Tecan plate reader at 730nm instead of the 720nm wavelength used by the PBR, however the effects of this discrepancy should be minimal. Predictions according to each data's respective constraints are as follows: Model 2 ($A_{560} \sim A_{680} + A_{730}$) is used for the Tecan measurements and Model 3 ($A_{560} \sim A_{680} + A_{720}$) is used for the PBR measurements. Since the predictions do not stem from the same model, the SC calibration data was used to conduct a similar analysis for comparison. Those results indicate that bacterial fraction predictions should be, on average, 0.05 absorbance units higher for PBR-measured data; this is in agreement with the observed results (Figure 4.15). Chlamydomonas' predictions are only slightly affected and both systems should yield similar values. Incorporating more accurate data into the model creation, could result in more reliable predictions after applying the appropriate corrections.

4.4.2 PBR Data Modeling

The suitability of transferring models generated with Tecan measurements is as of now inconclusive given the low yielding results collected from PBR2 & 3 as well as the different wavelengths measured. Furthermore, the latest iteration of PBR experiments does not have Tecan measured values as the instrument was out of commission during the data collection stage of this work. Given its observed stability, planned Tecan-measured samples will be key in demonstrating the viability of the models in monitoring the PBR as well as verifying whether the observed Chlamydomonas effect is manifested in the phycosphere models given their larger volumes. Nevertheless, PBR4 data (Figure 4.18) is modeled to illustrate how the proposed monitoring could facilitate its analysis.

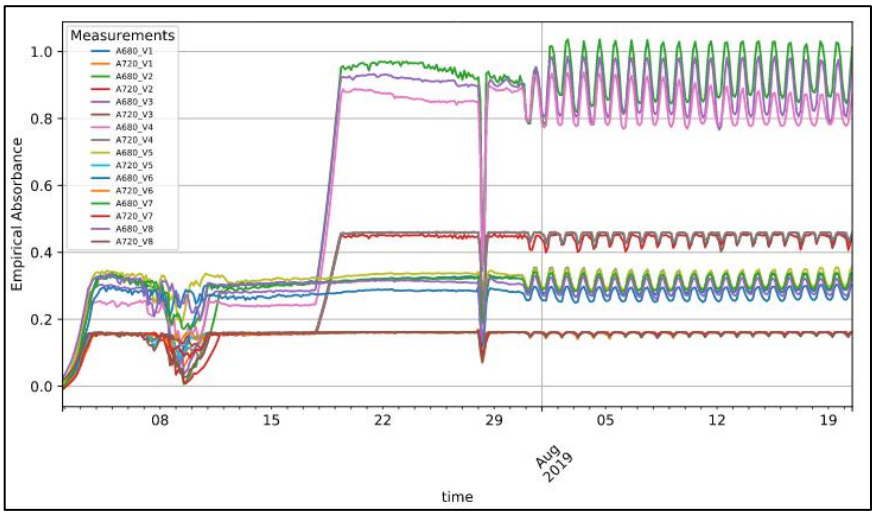

Figure 4.18: PBR4 raw data. For simplicity all the raw data from the experiment is visualized in this figure. Each vessel is measured at A_{680} & A_{720}. Oscillations correspond to the latest experiment where the system is submitted to light and dark cycles.

Each PBR experiment consists of 8 vessels each inoculated with a simplified mixture of strains from the ICL (SynCom) and Chlamydomonas. Once the instrument is connected to the network, it begins taking measurements; through its control software, the measurements can be viewed in real-time and monitored remotely. However, monitoring only shows the absorbance measurements at 680 and 720nm for each individual vessel and thus solely informs on the overall growth of the system. Alternatively, the system also measures the input of fresh medium over time, which continuously dilutes the system to maintain a constant absorbance and serves as a proxy to estimate the total community growth[45]. The main interest of the research group is to identify the state of both components in the system in real time and this part of the project tries to validate the Beer-Lambert law as a viable approach. Given the specific instrument model currently in use, the predictive models that can be constructed are limited to one: Model 3 ($A_{560} \sim A_{680} + A_{720}$). At these wavelengths, the calibration data indicates that it is particularly prone to the Chlamydomonas effect (Figure 4.19). Fortunately, the corrections devised for this model should bring the predictions back into focus as seen in the corrected version (Figure 4.19, bottom graph). These preliminary corrections, still in need of validation for both the PBR and Tecan, are nevertheless applied to PBR4 in order to illustrate their use.

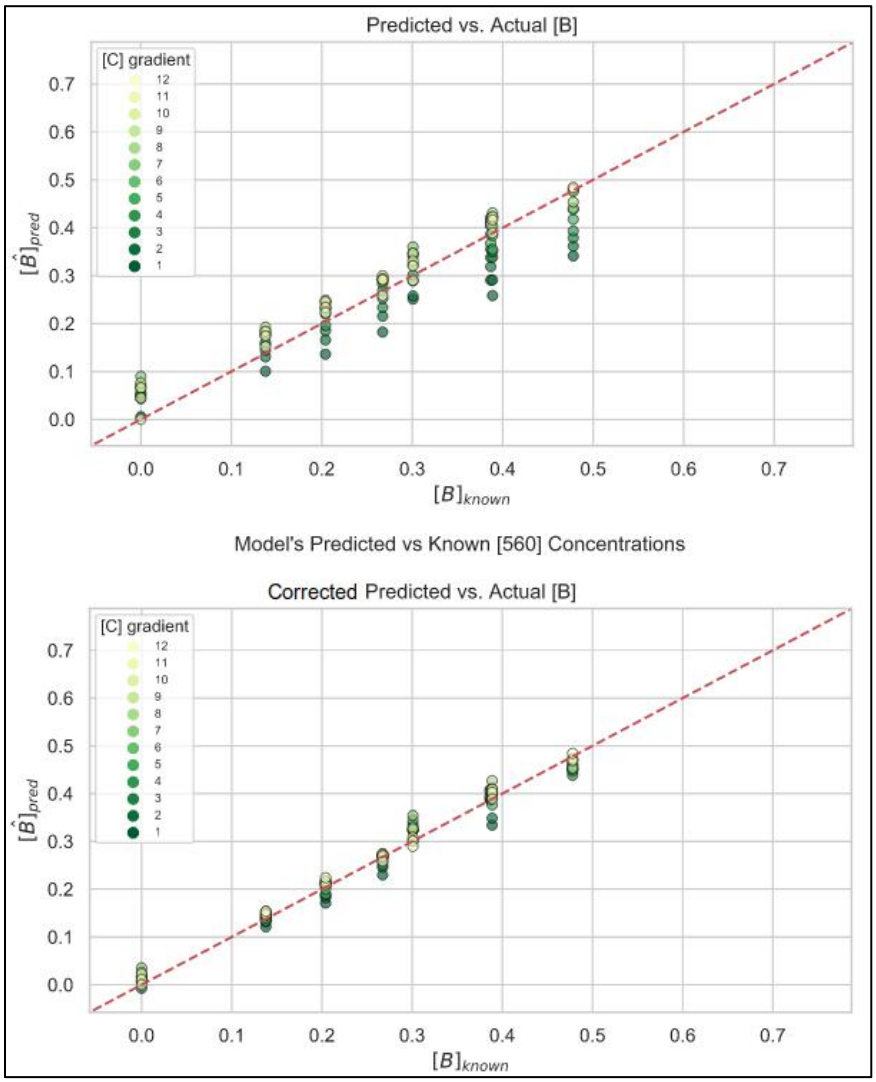

Figure 4.19: Component predictions of Model 3. Top graph shows the predictions on the testing wells in their uncorrected form. The bottom graph shows a marked gain in accuracy.

Figures 4.20 and 4.21 show the data re-sampled at a three hour window after removal of negative measurements , which occur mainly post-inoculation and are a calibration issue of the instrument. The PBR vessels (Figure 4.20) exhibit growth within the first few days post-inoculation, although it is unclear as to whether the observed growth is due to Chlamydomonas or the SynCom. The calibration data established that the system can in fact be modeled, albeit inaccurately due to the Chlamydomonas effect, and corrected. In order to illustrate the model and correction factor, the phycosphere model at vessel 4 is chosen as it has the highest range of empirical measurements as well as corrected and uncorrected prediction forms. The rest of the modeled vessels are to be found in Supplementary Material.

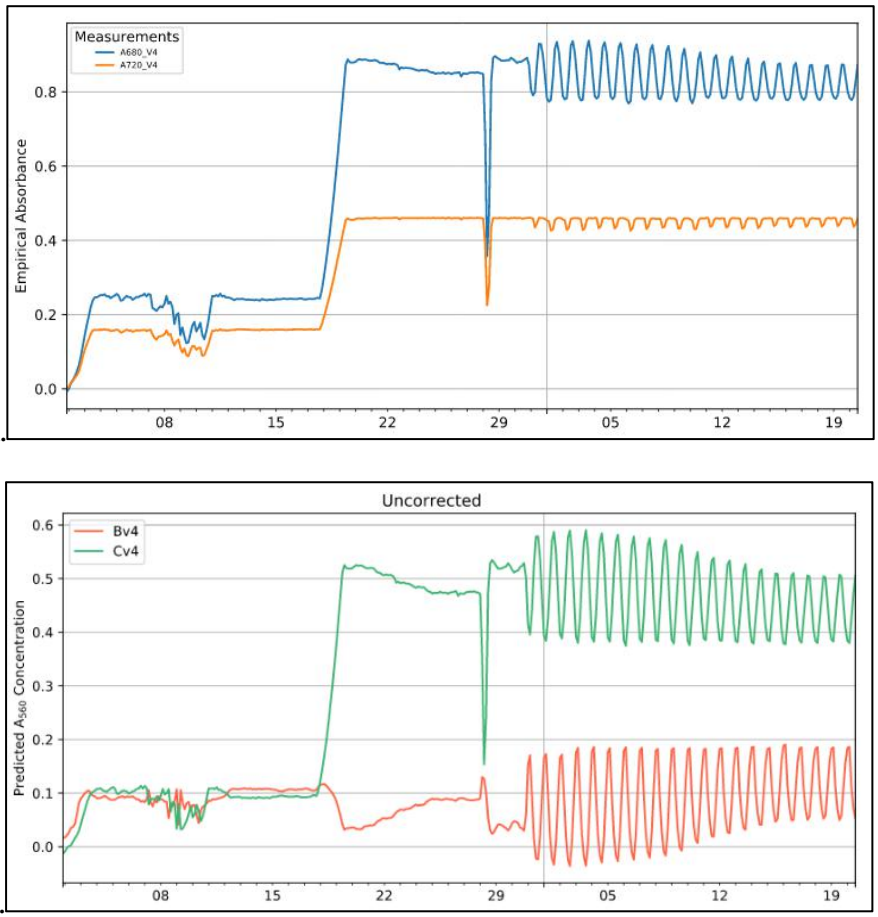

Figure 4.20: Predicted values vs raw data for PBR4-Vessel4 over time. A. Absorbance measurements **B.** Predicted values, green represents Chlamydomonas and red represents Strains

Raw values shows the growth and stabilization of the community (Figure 4.20A) . The constant concentration values are enforced on the system by programming it to maintain a constant absorbance level as measured by A_{720}. Hence, the observed growth beginning on July 17th is due to the instrument receiving a new and higher directive, i.e. to grow up to $A_{720} = .45$ absorbance units. Another crash is then observed on all vessels, per Figure 4.18, on July 28th and indicates either instrument malfunction or a forced dilution of all vessels. The system then quickly obtains its directed growth and a new experiment is begun on August 1st. That experiment, symbolized here by the oscillating curves, is simulating the light and dark cycles over a 24 hour period and shows how the systems adapts to those conditions. The valleys of the Chlamydomonas predictions (in green) line up with the peaks of bacterial predictions. Whether these peak-to-valley predictions are a mathematical artefact or a true phenomenon should be observable in the empirical cell count data or relative abundances, but multiple samples would need to be obtained over the 24 hour period this is observed.

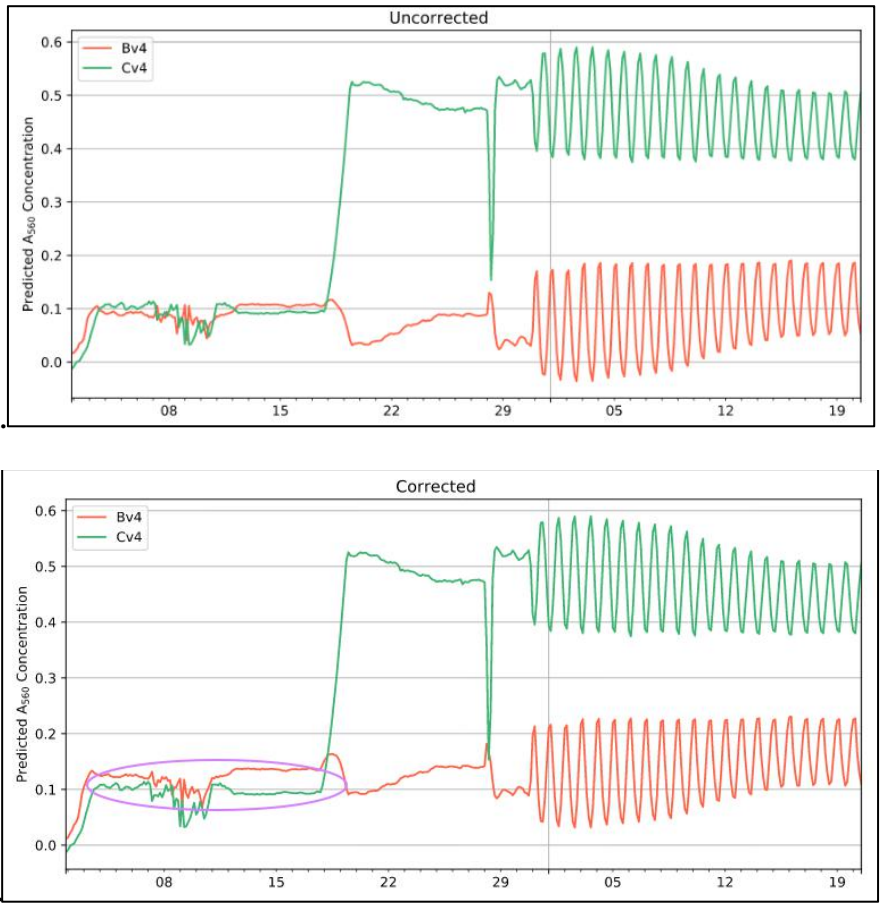

Figure 4.21: Predicted uncorrected values and corrected values for PBR4-Vessel4 A. Uncorrected values show a stretch of predictions which dip into negative values. **B.** Corrected version largely fixes these negative values. The increase in the signal is as much as 0.1 absorbance units, given that at Chlamydomonas concentration its effect is stronger

The uncorrected predictions (Figure 4.21A) indicate that growth is about the same for the two components during the initial two weeks post inoculation. In the context of absorbance measurement and this particular SynCom, the "same or similar" predictions translate to roughly 250x more bacterial cells relative to Chlamydomonas per the A_{560} -to-cell-count regression. After the system receives the directive to increase total system concentration Chlamydomonas cells accelerate their growth. The divergence of the growths immediately after the change in directive appears similar to what is observed in the diverging-converging oscillations in the latter part of the experiment. However, in this case, both empirical measurements are going upwards and the divergence in the predictions is likely due differences in light components absorption per their extinction coefficients. Note that three out of four coefficients have similar rates of absorption for both wavelengths, and that Chlamydomonas's absorption of 680nm light is much more efficient than the others hence its perceived faster rate of growth (Figure 4.5). The correction factor seems to corroborate this as

the purple ellipse indicates an increased correction of Strains. When the system reaches the directed absorbance near July 17th, it forces the system to maintain a constant absorbance and fresh medium is introduced to the system as necessary. The constant A_{720} measurement allows the observation of slight bacterial growth and as the Strains continue to grow more medium is introduced. Given that the bacteria divide faster than Chlamydomonas, the dilution might affect Chlamydomonas concentrations more.

The Beer-Lambert law states that a measured absorbance at any wavelength is equal to the sum of the individual absorbance of each of the components at that wavelength. Thus, for any wavelength measured or predicted, this must be true. However, as seen with the calibration data, this does not apply to individual components in this system as Chlamydomonas' concentration impacts the overall prediction of the bacteria. While Chlamydomonas's measurements tend to be correctly predicted by these models (when true bacteria concentrations are A_{560} < 0.45) the same is not true for the Strains' predictions. However, with the correction factors generated from the calibration data these bacterial fractions can be re-estimated (Figure 4.21). The resulting corrections are mostly unchanged and while the elevated levels of Chlamydomonas do indicate a stronger effect, a low prediction of bacteria is likely indicating that the system truly does have a low concentration of them. The negative predictions in the uncorrected results are mostly brought back to positive ranges as no negative values should exist. Negative values occur when the Strain fractions are below the limit of detection for that particular instrument. The slight correction of these data indicates that at these concentrations, bacteria and Chlamydomonas are already relatively well predicted.

To get a more objective estimation of the systems contents it might be useful to use the theoretical cell-counts to monitor the system (Figure 4.22). Since the instrument used to measure cell counts and regressions is not meant for quantitative measurements on this cell size range, these are by no means the true concentrations of cells in the system. Nevertheless, it provides a first approximation in more intuitive units of measurement for the monitoring of the system. The discrepancy of the absorbance by Chlamydomonas relative to the Strains (Figure 4.22) show that their quantities remain fairly constant throughout the experiment while the bacterial fraction fluctuates as the system tries to maintain a constant A_{720} during the day night cycle. It also appears that the bacteria are beginning to converge on to a particular value as noted by the amplitude of the oscillations diminishing.

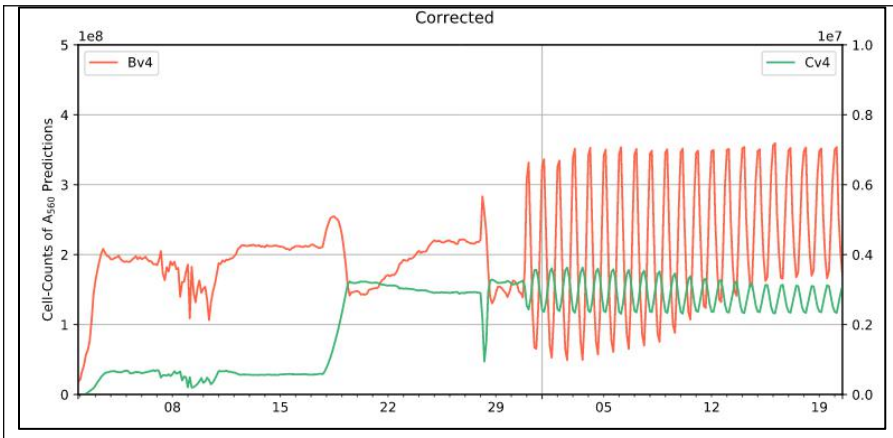

Figure 4.22: Predictions as cell counts. Predicted absorbace measures to cell counts.

Lastly, it was hoped that cell count regressions would provide a reliable first approximation of comparison to the 16S abundances (Figure 4.23) however, it does not seem that using cell count data is a good proxy for this. Given the relative abundances in this Figure 4.23 it would appear that more genetic material can be extracted from the lower quantities of Chlamydomonas than bacteria[48].

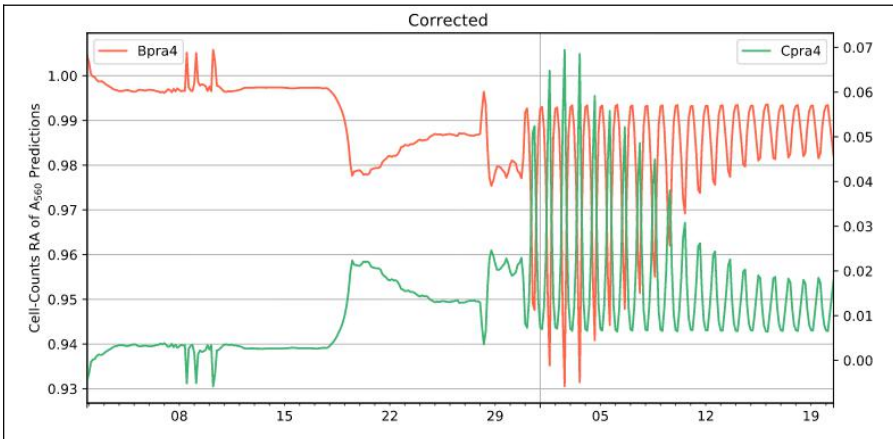

Figure 4.23: Relative abundance of PBR4 across time. The relative abundance of cell counts in PBR4

Given the results for the relative abundances, it does not seem that the Metadata consisting of 16S data and Coulter cell counts allow itself to be probed in this manner. It may be that another regression can be made on counts and 16S sequencing data, but that remains to be tested.

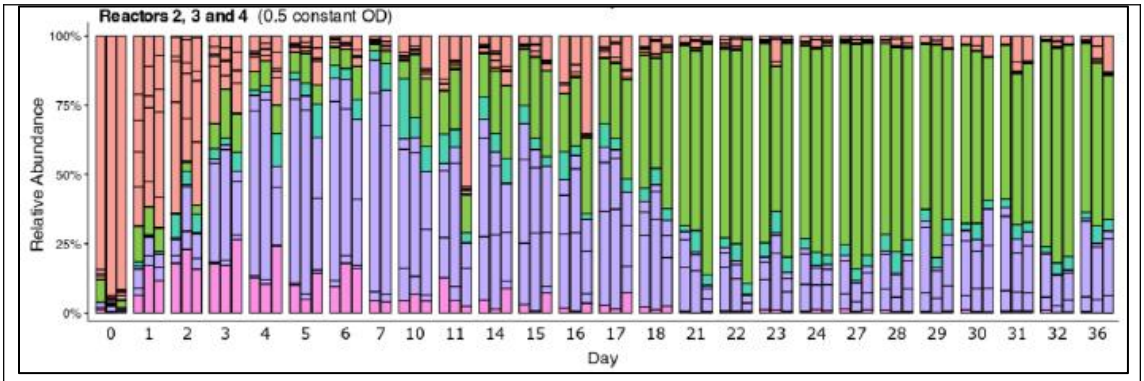

Figure 4.24: Relative abundances bar graph of 16S sequencing from PBR4. Green indicates Chlamydomonas. Other colors represent different bacterial families. Provided by Dr. Jose Flores and Dr. Ruben Garrido-Oter.