Prior to uncovering Model S buried in the calibration data, a signal correction method for bacterial predictions was chanced upon when trying to understand the ratio of absorbances (B:C) for the predicted fractions of Model 3. While visualizing the data of the bacterial predictions in a table that conforms to the 96-well position (**FiguredRedTable**) a slight ‘bump’ in the data was noticed in row H columns 6-8 of the bacterial predictions. At row H there is no bacteria in any of the wells but a Chlamy gradient exists across that row. With perfect models the value at each row would be the same and a flat line at the known concentration would be observed **figure with flat lines**. (Model S is used to illustrate the correction method found with Model 3). This plot, however, shows how the errors in bacterial fraction actually behave: the presence of Chlamy imparts noise to the prediction with a severity that depends on the concentration of bacteria. As the concentration of bacteria increases the under-predictions become more severe with respect to Chlamy content. To observe this net effect of Chlamy only the predictions at column 12 (where no Chlamy is present) is removed from each prediction at its respective row effectively shifting the parabolas downward and forcing them pass through the origin **FigureTruncated**. These behavior gave rise to an idea that allowed the manipulation of the predictions, albeit with **one/two** key concession being made, but which result in an increased performance over the stock prediction of the modeled calibration data and is expected to work with other Tecan measurements as well; however this is yet to be determined.

To elaborate it is best to begin with the parts of a parabolic function **[cite math book]**:

|  |  |
| --- | --- |
|  | Parabola standard form |
|  | Parabola vertex form: vertex (h, k) |
|  | Relationship between the Equations above |
|  |  |
|  | Truncated parabola |
|  |  |
|  |  |

**Explain the abvoe**

In order for the manipulations to work, the concession has to be made that the observed parabolic trends seen in **FigureUntruncated** all come from a parabola having a unique breadth or **a2** parameter (i.e. the rate at which it flares open), which is not an unreasonable first approximation as seen in **TruncatedPlot. Once** this concession is made the next step is to model the observed relationship between each fitted parabolas vertex (h, k) at **h** and the known bacterial concentration: **h** moves leftward as a function of B. Objectively, this relationship is true, there is an apparent movement of fitted parabolas vertex’s h value leftwards in FigureUnctruncated. Its not unreasonable to imagine that at any concentration of bacteria B, a theoretical parabola exists whose vertex lies along the trace and outlines the path of all possible predictions given Chlamy—lower lying parabolas have a lower rate of of change than higher lying parabolas with respect to Chlamy concentration. When these theoretical parabolas are subsequently forced to pass through the origin, the Chlamy concentrations (predicted with a good degree of accuracy) determines the extent of this influence and this value is either added or removed from the original prediction depending whether it lies below or above the x-axisas seen in truncated.

To grant the concession the data is fitted to find the average parabolic breadth which yields -0.4187, but alternatively the median can be used. Subsequently, it was found that the average over all the corrections Models 1-S (global **a2** average, -0.4995) gave the best performance. Next, the path that the parabolas’ vertices trace must be found and is the slope **h/B.** This parameter had to found numerically from each of the fitted parabolas as the set of returned parameters **[a2,a1, a0]** do not yield an h using the analytical function to obtain (h, k). Through Nympy’s **poly1d[cite]** function, which takes in a curves parameters to create a predictive linear model, a list of 100K numbers, representing a range of **h** values (-1,1), is used obtain the **h** value at each apparent vertex (using the argmax property of Numpy lists). The large list of number is necessary approximate the true **h** as best as possible. These **h** are then plotted against the known B concentration of each row and results in the following regression **Fig\_hBRegres.** The parameter representing this line are used to compose a line **f(pB) = -m(pB) + b** which yield a **h** of (h, k) that corresponds to that particular parabola in the space of all possible parabolas which lie along that path.

Then for any predicted concentration pB, this function is used to yield an **h** of a theoretical parabola whose only other known attribute known is an averaged breadth (a2) parameter. To fit a parabola to this h, **a1 is necessary** and can be derived from the mathematical relationships above. Once **a1** is obtained, the parameters **a2** and **a1** are fitted to conform to a truncated parabola in the form **[a2, a1, 0]. Using the poly1D** function the resulting equation is a parabola which takes in **pC** and yields the net effect of that Chlamy concentration for that particular parabola. The yielded value is then added subtracted from the given pB. The resulting corrections are shown in **FigureCorrected.**

|  |
| --- |
|  |
|  |

While at first it may seem that this finding is an effect of circular logic, take for example the fact that in the above uncorrected graph above any horizontal line would cross more than one point in the graph (e.g Bpred = .213) each would yielding the exact **h** from **f(pB)** and thus the same parabola. However, when these parabola are truncated and subsequently fed the pC which different in all cases, the signal which is removed is different. Thus it seems that indeed there is some truth behind the mathematical manipulation of the data. Moreover, it is possible that without the accurate predictions of Chlamy it would not be possible to achieve this kind of correction