GCAT: An introduction

GCAT is a web-based tool for summarizing microbial growth curves using mathematical modeling. The user interface requires no programming and calls on an R package of the same name which processes input data files, models the curves, calculates important growth parameters from the fits, and returns both graphical and tabular output. This manual provides a detailed explanation of each step in the process of using GCAT.

Methodology

GCAT's basic function is to fit a growth curve to the cellular density of microbial cultures as a function of time, using non-linear regression techniques (thru the *nls* function in R). The resultant growth curve can be used to calculate parameters describing characteristics, including growth rate, theoretical plateau and lag time.

GCAT performs some transformations on cellular density data in order to better estimate relevant growth parameters. Cellular density is usually measured by optical means such as spectrophotometry, and is referred to interchangeably with optical density (). By fitting the logarithm of the optical density, , as a function of time, it is possible to estimate the maximal exponential growth rate (labeled "specific growth rate"), which is the growth rate of most interest in microbial growth experiments (Zwietering et al. 1990).

GCAT uses the Richards equation to model growth curves. In the event of algorithm failure, the logistic equation is substituted in and another fit is attempted. Curves drawn from both equations are monophasic and sigmoid, meaning that they are S-shaped and contain distinct lag, growth, and plateau phases. For a more detailed overview of the equations used, refer to Zwietering *et al.* (1990).

However, the above procedure relies on a stable measure of initial cell density. Multiple measurements of both the media background and initial culture are needed to make a good estimation of initial cell density, which is often a difficult task to accomplish under experimental constraints. If the initial cell density measurement is not reliable or accurate, the parameters may lose their meaning. Furthermore, with small numbers, a logarithmic transformation tends to exaggerate growth trends when numbers are low. Thus, GCAT offers a compensatory transformation, where n is an arbitrary constant with a default value of 1.

Caution should be exercised with this transformation as the term "specific growth" will no longer reflect a true exponential trend. However, use of this constant greatly stabilizes curve fitting with small values and the "specific growth" term can still be reliably used to compare and rank growth curves.

Experiment setup

GCAT is meant to analyze simultaneous growth experiments, where density is tracked over time (typically using measures of absorbance, or optical density) in multiple wells of microtiter plates. This means that input files need to provide well identifiers along with timestamp data and density measurements. In standard 96-well microtiter plates, wells are labeled alphanumerically (from A1 thru H12), and GCAT adopts this naming scheme for all experiments.

GCAT can also be used in large-scale experiments with multiple microtiter plates, in which case input files need to provide plate identifiers as well. Proper data formatting is described in the following section.

Using GCAT: Uploading data and specifying analysis parameters

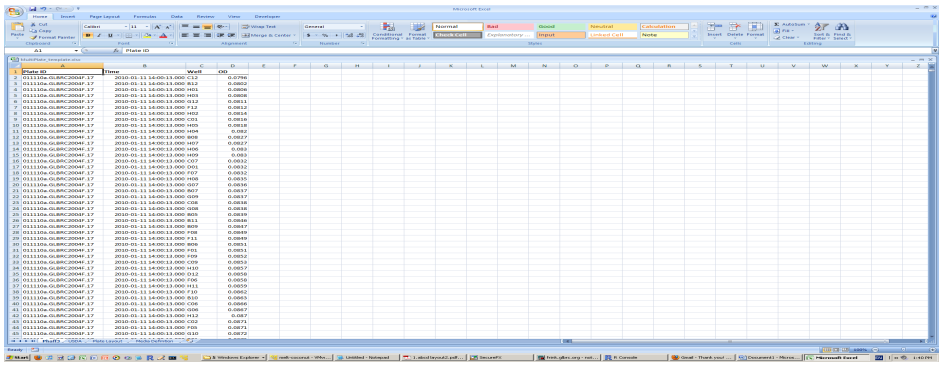
Input needs to be in a specific format in order to be read by GCAT. Files must be in **.csv** (comma-separated values) format, which can be easily edited and exported from any standard spreadsheet program. In this manual, files are displayed using Microsoft Excel® 2007.

Users have two choices for input format, depending on the nature of the experiment from which the data was generated. The multi-plate (long) format is used for data gathered by automated plate-handling systems, while the single-plate (wide) format is used for data gathered from a single plate reader. Both formats are described in detail below.

CGAT takes only one input file at a time and only standard 96-well microtiter plates will be supported in release 1.0/version 1.0

Multiple-plate (long) format:

Files in this format contain optical density reads in a long column, with additional columns identifying the plate ID, timestamps, and well names. These columns must be clearly labeled “Plate ID”, “Time”, “Well” and “OD”. Input does not need to be in a specific order, but please ensure that the Plate ID and Well ID columns are consistent and contains no misspellings, as GCAT relies on them to organize the data into separate growth curves.



Single-plate (wide) format:

Files in the single plate format and contain optical density reads in multiple columns (one for each well position), and one timestamp column, the header of which should

be left blank. The density reads must be labeled with the names of the wells, starting with “A1” and going through all columns (numbers) before advancing any rows (letters).

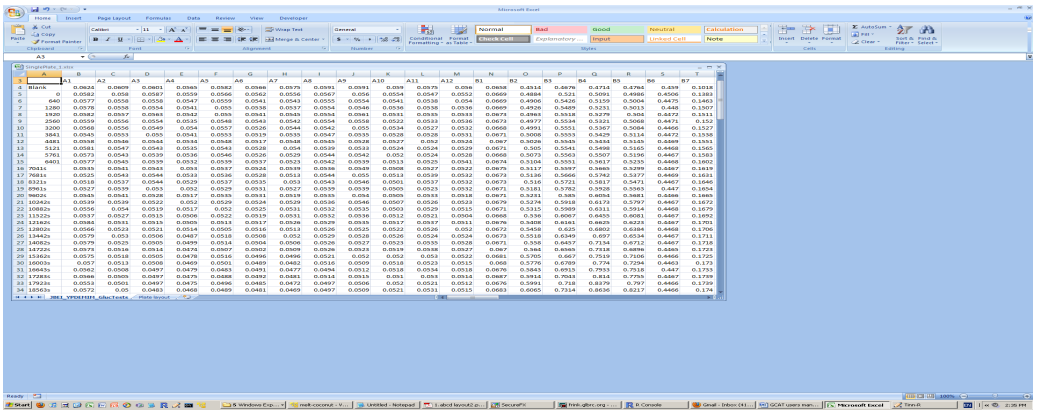
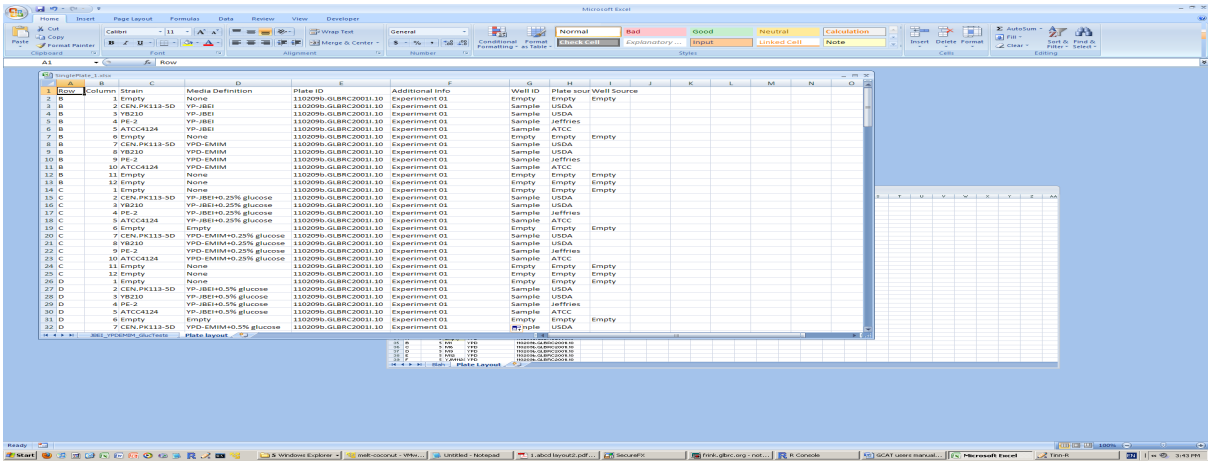


Plate layout information:

Users can also upload a companion **.csv** file containing identifying information for all wells in the experiment(s) to be analyzed, including strain identifiers and growth media (or growth environment) definitions, which are used for graphic output. The format for this file is shown below:



Note that to specify a strain as “Empty” lets GCAT know that the well should not contain any growing culture. If GCAT finds a trend to the contrary, the well will be flagged with an error (but still analyzed).

Using GCAT: Setting analysis parameters

On the analysis front page, several options are available for the user to manipulate. They are described in detail below.

Note: Point-specific parameters refer to the **index** (not the time!) of cellular density reads. GCAT automatically sorts the data in each well chronologically, so users may refer to the first timepoint of each well as point 1, the next (chronologically) as 2, and so on. The points are also numbered using their indices in the output graphics.

OD Transform

GCAT transforms optical density to before proceeding with model fits, as described in the *Methodology* section on page 1. Users can pick the default setting (), or specify a value for at their own discretion. Using zero for is not recommended unless growth curves display very accurate initial density measurements over a substantial amount of curve points, and including proper blanking.

Media background

If the user has taken their own detailed measurements of the absorbance of the media background, they may enter it here and it will apply across the entire experiment. Otherwise, if left blank, GCAT automatically takes the first timepoint of every curve to be a blank read and subtracts its value from the other measurements in the curve.

Inoculation timepoint

Users should enter the **index** of the point at which the wells are inoculated with culture. As a default, this is set to the second read (after the blank). This setting also applies across the entire experiment.

Growth Threshold

Users may enter a threshold to determine the presence of a growing culture. GCAT calculates the difference in optical density between all the points of each curve and the point specified as the inoculation point, above. If none of the points in a curve pass this threshold, the curve is discarded as an empty curve.

Points to ignore

If there are erroneous reads which pose a problem to analysis in an experiment, users can enter their **indices** (as a comma-separated list) to tell GCAT to ignore them across all wells.

Timestamp format

For Multiple-plate format, you will need to select the timestamp format from the select list that corresponds to the format used in the Time column of the input file. Single-plate format does not use this parameter.

Note: it is best to change the format of timestamps within the .csv files (Microsoft Excel allows mass formatting of timestamps) rather than enter a custom value for this parameter each time.

Below is a list of recognized single-letter placeholders for units of time:

S: Seconds as decimal number (00-61)

M: Minutes as an integer (00-59)

H: Hours as an integer (00-23) or (01–12) when used with **p**

p: AM/PM indicator in the locale

d: Day of month as an integer (00-31)

m: Month as an integer (01-12)

Y: Year with century

y: Year without century

The default time format for GCAT is **Y-m-d H:M:S**. If kept, this means that timestamps need to look like the following in the input data file: `2000-01-01 01:23:44.000` or GCAT will not proceed. Upon submission, GCAT will use this time format to convert all timestamps to the number of **hours** since the start of the experiment.

For Single-plate format, input time must be in seconds. The x axis of the well-by-well graphs is on a scale of hours.

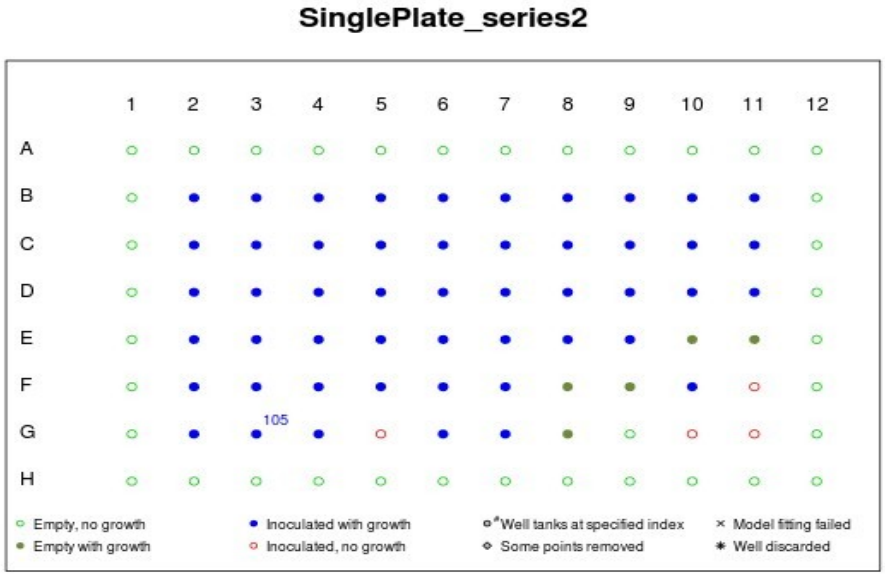
Using GCAT: Interpreting output

GCAT produces output in three formats: 1) a general graphic overview of the analyzed plates, 2) well-by-well graphs of cellular density vs. time with the model fits and calculated parameters drawn in (in .pdf format), and 3) a table containing all calculated values and identifying information.

Plate overview

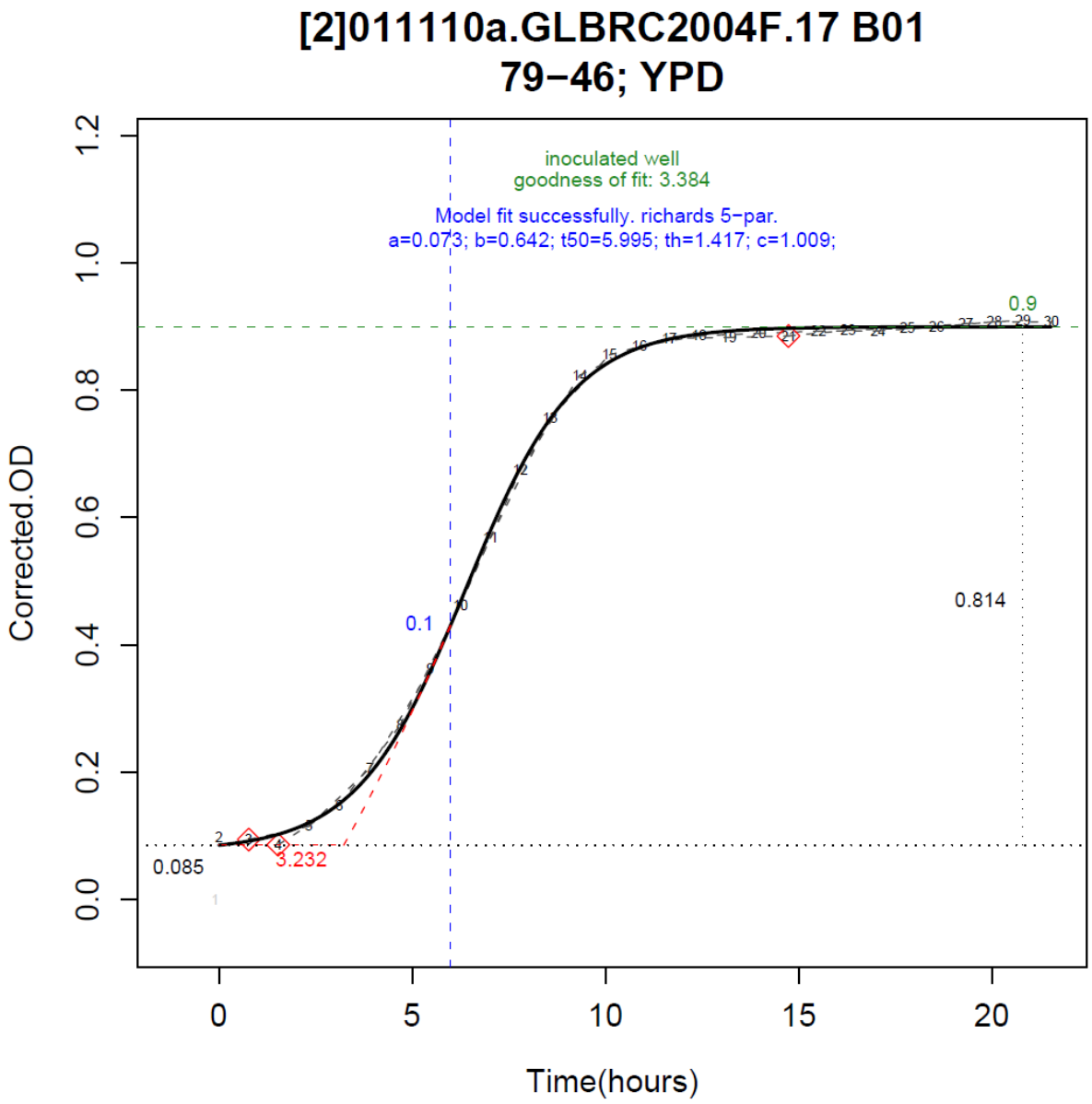
This graphic provides a visual overview of the plates, whether they were empty or inoculated (see *Plate format* on page 3) and whether the analysis confirmed it, the success or failure of the algorithm to produce a nonlinear fit, and various error flags. A legend is included for the points representing each well:

In general, an open point indicates a well that was found to be beneath the **growth threshold** (see page 4). If the well was marked empty in the analysis, it is colored green (no error). If the well was marked as containing a culture, it is colored red and marked as an error. Similarly, closed points represent inoculated wells displaying growth above the threshold, unless they are brown, in which case growth was found



in a well marked as empty (also an error).

Wells marked with an X indicate a failure of the algorithm to fit the curve (usually due to variant curve shapes). A number to the upper right corner of a well indicates the starting index of a tanking trend, if found. Wells marked by a diamond indicate the successful analysis after removal of an errant jumps (see page 4), whereas an asterisk indicates wells discarded for too many jumps.



Well-by-well graphs

The title of each graph is formatted as follows. On the first row, a bracketed number indicates the row index of the output table that corresponds to the graph, followed by the plate name and well identifier. The second row displays the strain identifier followed by media definition, if such information was made available in the input data.

The points are shown as small numbers, labeled by index as discussed on page 4. Grey points have been ignored for the analysis. Points surrounded by red diamonds indicate part of a tanking trend, while points surrounded by a triangle indicate a directional jump in density. All these symbols are used by GCAT to decide if any points need to be removed.

Note: tanking trends are only removed when the number of points in the trend reaches a certain threshold (this is not a user-defined parameter, but might be implemented in the future).

If the fit was successful, a graph of the model equation is shown, overlaid on the points. The following lines are also drawn and labeled with their numerical values:

Dotted black (horizontal at bottom): The fitted value for initial cellular density (after inoculation)

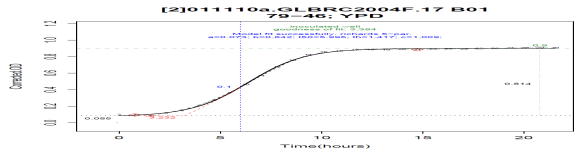
Dashed green (horizontal at top): The fitted upper limit for growth (plateau)

Dashed grey/light grey (vertical): The total change in OD reached by the end of the curve, or “final growth”. If this is significantly less than the plateau density (< 75%), the difference is also marked with a lighter grey line.

Dashed blue (vertical): The time at which the maximal (specific) growth rate is reached

Dashed red (tangent): The exponential growth trend at the point of maximal growth. The intersection of this line with the initial cellular density determines the lag time (labeled in red)

A legend of all symbols and lines is also included as the first page of each output PDF file for reference.



The text directly above each graph has three components.

1) Shows any errors detected by GCAT (growth in empty wells or no growth in inoculated ones, tanking trends, jumps, etc.) and will be red if there were any.

2) A goodness-of-fit metric, calculated as the inverse log of the normalized root mean square deviation for the fitted curve. A value above 2.75 (green) generally indicates a good fit, whereas a value below 2.0 (red) indicates an unreliable one that should be checked visually.

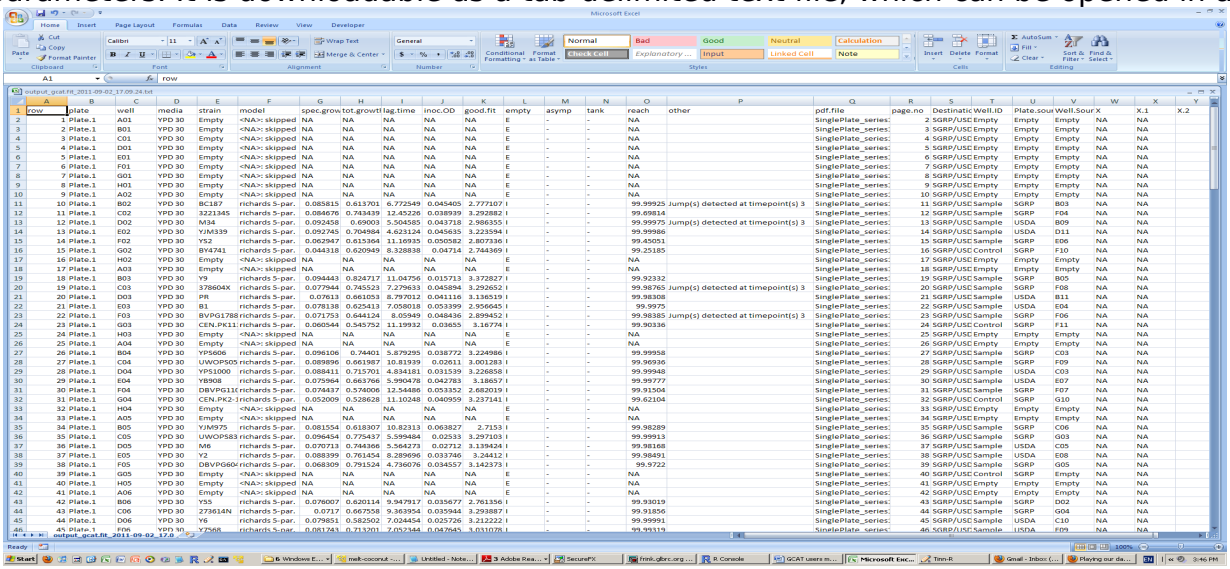
3) Lastly, the text indicates which model, if any, was successful in fitting the growth curve, and the parameters that were fit to it. The formula for the Richards and logistic equations are as follows:

Richards: Logistic:

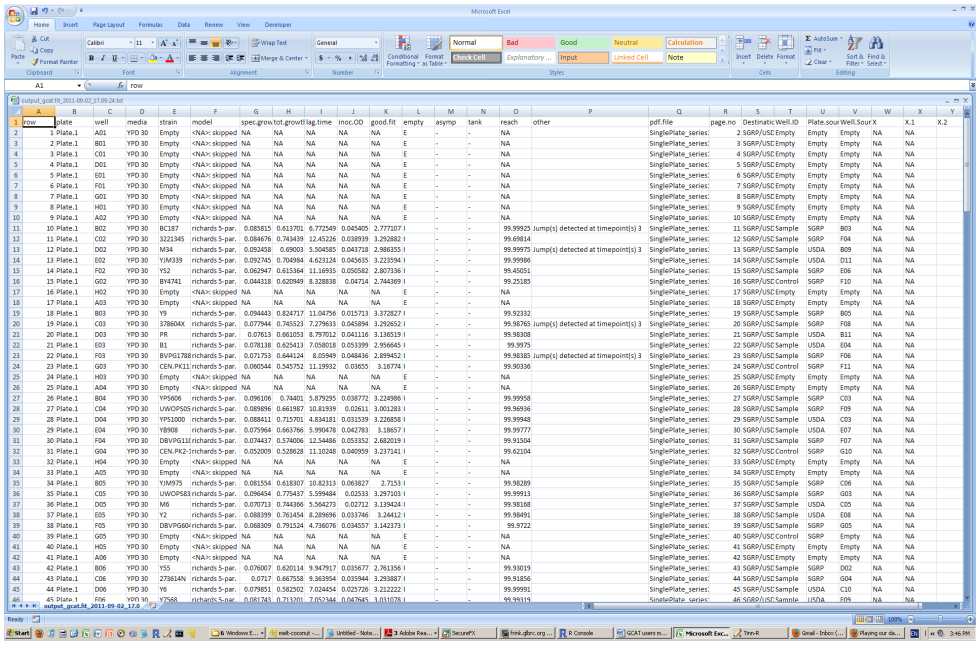
where log-cell density is a function of time , and and are the estimated parameters. By plugging in the displayed values, users can reconstruct the equation for the displayed fit.

Output table

The output table contains all the information about the curve fits and calculated parameters. It is downloadable as a tab-delimited text file, which can be opened in a



standard spreadsheet program such as Excel.



Each row contains the information and calculated parameters for one fitted well. Explanations of columns are as follows:

Column 1 (“row”): Lists the rows of the output table. This number should correspond to the bracketed number in the output graphs (see *well-by-well graphs*, above)

Columns 2-3 (“plate” and “well”): Plate ID and well position.

Columns 4-5 (“media” and “strain”): Media definition and strain ID, if the information was provided by the user in the input.

Column 6 (“model”): Which model (Richards or Logistic) was used successfully to fit the curve? If the curve was not fit successfully or skipped due to not having enough data or being under the user-defined growth cutoff (see *Specifying Analysis Parameters*), this column will read <NA>.

Columns 7-9 (“spec.growth”, “tot.growth”, “lag.time”): List the specific growth rate, total theoretical growth over baseline, and lag time calculated for each well from the fit parameters, and described previously (see *well-by-well graphs*). Note: “tot.growth” is the value of the theoretical plateau minus “inoc.OD” below.

Column 10 (“inoc.OD”): The calculated cell density at the time of inoculation. The specific timepoint at which this happens was specified by the user upon input (see *Specifying Analysis Parameters*).

Column 11 (“good.fit”): Goodness of fit statistic based on normalized RMSD (see *well-by-well graphs).*

Column 12 (“empty”): Flags for inoculation status. This has four possible values:

I – Well was marked as inoculated (normal)

E – Well was marked as empty (only containing media)

! – Well was marked as inoculated but contained no growth above the cutoff value.

E* - Well marked as empty by user, but had growth above the cutoff value.

Column 13 (“asymp”): Another flag, this one shows “L” if the bottom asymptote (density at inoculation) was not reached, and “U” if the upper asymptote (plateau density) was not reached. Whether a value was reached is determined by whether it is within a 10% margin of the cell density range of the entire curve. In either case, the flags indicate that there may not be enough data around the horizontal asymptotes to ensure an accurate fit.

Column 14 (“tank”): The tanking flag. If a number is shown, the curve “tanked” (showed a continued decrease in cell density measure) at the numbered index. Data after the index was automatically removed by GCAT and analysis was allowed to continue.

Column 15 (“reach”): This column displays the *percent* of total growth (column 8) reached by the fitted curve at its maximum (within the timepoints collected). Multiplying this percentage with “tot.growth” yields the actual increase in cell density seen within the time course of the experiment (which may be different from the theoretical increase in cell density fit by the algorithm).

Column 16 (“Other”): Additional flag column, displaying information about whether jumps in OD were detected and what was done about them (depends on user input as well).