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.

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 below.

Statistical 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 (through the *nls* function in R) or local regression (*loess* function in R). The resultant growth curve can be used to calculate parameters describing characteristics, including growth rate, theoretical plateau and lag time.

We have adapted GCAT to deal with certain peculiar qualities of growth curve data that are frequently encountered in microbiological research. These qualities include high optical density of the media at 600 nm (OD), low cell count inocula, varied and often non-standard growth curve shapes, and decreases in OD caused by cell settling, flocculation and death. The GCAT algorithm and these adaptations are described in the following paragraphs.

Microbial growth curves are usually expressed as $log(N/N_0)$ vs. time, where N is the cell number and N_0 is the number of cells at the beginning of the experiment, e.g. in the initial inoculate. The cell number is proportional to optical density (OD) at 600 nm, therefore N can be replaced by OD for growth curve analysis purposes, upon subtracting the background OD of the growth medium:

$$log(OD_{corr}) = b + f(t) \quad (1)$$

$$OD_{corr} = OD - OD_{blank}$$

$$b \approx log(OD_0 - OD_{blank})$$

where OD_{corr} is background-corrected optical density value at time t, b is the lower asymptote of the growth curve, $f(t)$ is a growth function (usually of sigmoid shape), OD is raw optical density, and OD_{blank} is optical density of blank medium.

However, background *OD* of media used in bioenergy research related experiments are often relatively high. This makes the initial *OD* of the cell culture difficult to measure, i.e. $OD_0 - OD_{blank}$ tends to 0 or is negative, and *b* tends to -∞ or is undefined. Therefore, GCAT allows the user to replace log-transform of OD with *log(x+0.1)* transform, which ensures that *b* is finite:

$$OD_{corr} = OD - OD_{blank} + 0.1 \tag{2}$$

$$b \approx log(OD_0 - OD_{blank} + 0.1)$$

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

To allow an even greater flexibility, GCAT also supports *log(x+δ)* transform for OD, where δ is any positive real value. Additionally, a GCAT user can use the first OD measurement in a well as the blank or supply a custom blank value. It is also possible to specify the time point at which the cultures were inoculated.

GCAT offers two alternative model options: sigmoid and LOESS. If the sigmoid option is chosen, GCAT picks the best-fitting of three widely-used sigmoid curve models: Richards, Gompertz and logistic. GCAT uses reparametrized formulas developed by Zwietering (Zwietering et al. 1990), which express growth in terms of lag time, specific growth rate and asymptotic growth value. This enables GCAT to directly estimate these essential growth curve characteristics, along with their standard errors. The Richards formula is expressed as follows:

$$y = b + A\left\{1 + v \cdot \exp\left(1+v\right) \cdot \exp\left[\frac{\mu_m}{A} \cdot \left(1+v\right)^{\left(1+\frac{1}{v}\right)} \cdot \left(\lambda - t\right)\right]\right\}^{\frac{-1}{v}} \tag{3}$$

Where *t* is time, *y* is response, i.e. the log-transformed corrected OD value, *b* is the lower asymptote, *A* is the upper asymptote, μ_m is the maximum specific growth rate, and λ is the lag time. The formula also includes the "shape" parameter *v*, which affects the position of the inflection point (Birch 1999). When *v* equals 1, the Richards formula becomes equivalent to Logistic:

$$y = b + \frac{A}{\left\{1 + \exp\left[\frac{4\mu_m}{A}\left(\lambda - t\right) + 2\right]\right\}} \tag{4}$$

Although the Richards formula is undefined when *v* equals 0, it tends towards the Gompertz formula as *v* tends to 0. The Gompertz formula is defined as follows:

$$y = b + A \cdot \exp\left\{-\exp\left[\frac{\mu_m \cdot e}{A}(\lambda - t) + 1\right]\right\} \quad (5)$$

Although, owing to the presence of a shape parameter, the Richards formula is able to fit a wider range of growth curves than Logistic or Gompertz, it has been criticized for being too flexible and poorly constrained by data (Zeide 1993). Indeed, in our own experience, the shape parameter often has wide confidence intervals. It can be argued that more parsimonious models should be preferred if they do not result in significantly worse fits, as they will be less affected by noise in the data and therefore have better predictive value. In case of linear regression, an F-test is often used to decide if excluding a certain model parameter results in a significantly worse fit. However, in our case, regression is non-linear, and it is not immediately obvious what model should be used if Richards were rejected, e.g. logistic or Gompertz. We therefore opted for the following model selection algorithm:

1. Fit the Richards model
2. Examine the shape parameter v and its standard error
3. If $v > 0.5$ and $|1 - v| < 2*SE_v$, use logistic model
4. Else if $v < 0.5$ and $|v| < 2*SE_v$, use Gompertz model
5. Else keep Richards model

SE_v refers to the standard error of the estimate of v, as returned by the *nls* function. The main idea behind this algorithm is as follows: if the shape parameter is close to 1, the data are well described by a logistic curve; if it is close to 0, they are well described by Gompertz curve, and if neither of the above is true, we are justified in using the Richards equation.

Since sigmoid curve equations are non-linear in their parameters, a non-linear least squares algorithm must be used to estimate parameter values. Such algorithms require an initial guess and may not converge if that prediction is too far away from the values that give the best fit. When analyzing a single growth curve, initial guess values can be supplied by hand. For example, IPMP 2013 software has an elegant implementation where a user can enter an initial guess and see how close the corresponding predicted curve is to the data. However, when dozens or hundreds of curves need to be analyzed in an automated fashion, such an implementation becomes unworkable. We have addressed this issue by fitting a LOESS local regression model first, estimating lag time, specific growth rate and asymptotic values from that, and supplying those estimates as the initial guess to the non-linear regression algorithm.

When using LOESS regression, GCAT estimates growth curve parameters as follows. First, LOESS regression is used to fit the log-transformed *OD* values as a function of time. Then, the entire time course is divided into 1,000 equal intervals, resulting in 1,001 time points. Estimated response values $\hat{y}$ are predicted by the LOESS model for each time point and derivatives $d\hat{y}/dt$ are computed numerically based on those values. The highest derivative value is maximum

specific growth rate μ_m, and the time point where it occurs is the inflection point. Lower and upper asymptote values b and A are estimated as the lowest and highest values of $\hat{y}$ respectively. The lag time is then estimated as the point at which the tangent to the growth curve at the inflection point intersects the lower baseline:

$$\lambda = t_{ip} - (\hat{y}_{ip} - b) / \mu_m \quad (6)$$

where λ is lag time, t_{ip} is the inflection point, and $\hat{y}_{ip}$ is estimated response (log-transformed *OD*) value at the inflection point (see Fig. 1 in Zwietering (Zwietering et al. 1990)).

In our experience, non-linear regression algorithms are prone to convergence failures. In the context of high-throughput data analysis, such failures must be handled in an automated way. We have been able to ensure convergence in most cases by using the port algorithm and setting appropriate constraints on allowable parameter values. If a fit still fails, GCAT reverts to an alternative growth curve model. Thus if fitting the Richards equation fails, a logistic model is used instead. If Richards model converges successfully but a simpler logistic or Gompertz model fails, GCAT reverts to Richards. Using these procedures we have been able to fit all experimental growth curves encountered so far.

Yeast cultures growing in complex media may undergo a diauxic shift. A common situation is for engineered, xylose-fermenting yeast growing on plant hydrolysate to switch to xylose consumption once they have consumed all available glucose. This results in complex growth curve shapes. For example, instead of reaching an asymptote, OD may continue to grow, albeit at a much slower rate than in the earlier growth stages (Werner-Washburne et al. 1993). If the culture is monitored long enough, a second plateau may be reached. Such curves can no longer be well modeled using traditional sigmoid equations. For complex curves that cannot be adequately modeled by a global equation, GCAT offers an option to use a local regression algorithm.

The local regression algorithm offered by GCAT is LOESS, as implemented in *loess* function in the R statistical computing environment (W. S. Cleveland, E. Grosse, and W. M. Shyu 1992). In order to obtain a good fit using LOESS, the user must set the *span* parameter, referred to as “smoothing parameter” in GCAT. This parameter determines the size of the local neighborhood used by the LOESS algorithm. The optimal value of the smoothing parameter depends on the data, but should be above $1/N$, where N is the number of points in the growth curve; GCAT's default value is 0.1. Values that are too high result in a poor fit, while values that are too low result in over-fitting. LOESS can fit a curve of an arbitrary shape. The estimation of growth curve parameters from a LOESS fit is identical to the procedure of providing initial guesses for sigmoid models, described above. In its current implementation, GCAT does not report standard errors for LOESS-estimated growth curve parameters. This would require the use of a resampling method, increasing computation times by several orders of magnitude.

Another issue often encountered in microtiter plate based cell culture growth experiments is declines in OD caused by cell settling, flocculation and death. Cell settling and flocculation

often occur in the first few minutes post-inoculation. They cause an initial decline in OD, which is followed by increase due to cell growth. To handle this phenomenon, GCAT allows the user to specify a set of time points to be dropped from the analysis, e.g. the user can drop several initial time points where OD is declining. Cell death results in a sharp decline of OD towards the end of the time course. GCAT can detect such a decline and disregard that portion of the growth curve.

Using GCAT: Uploading data and specifying analysis parameters

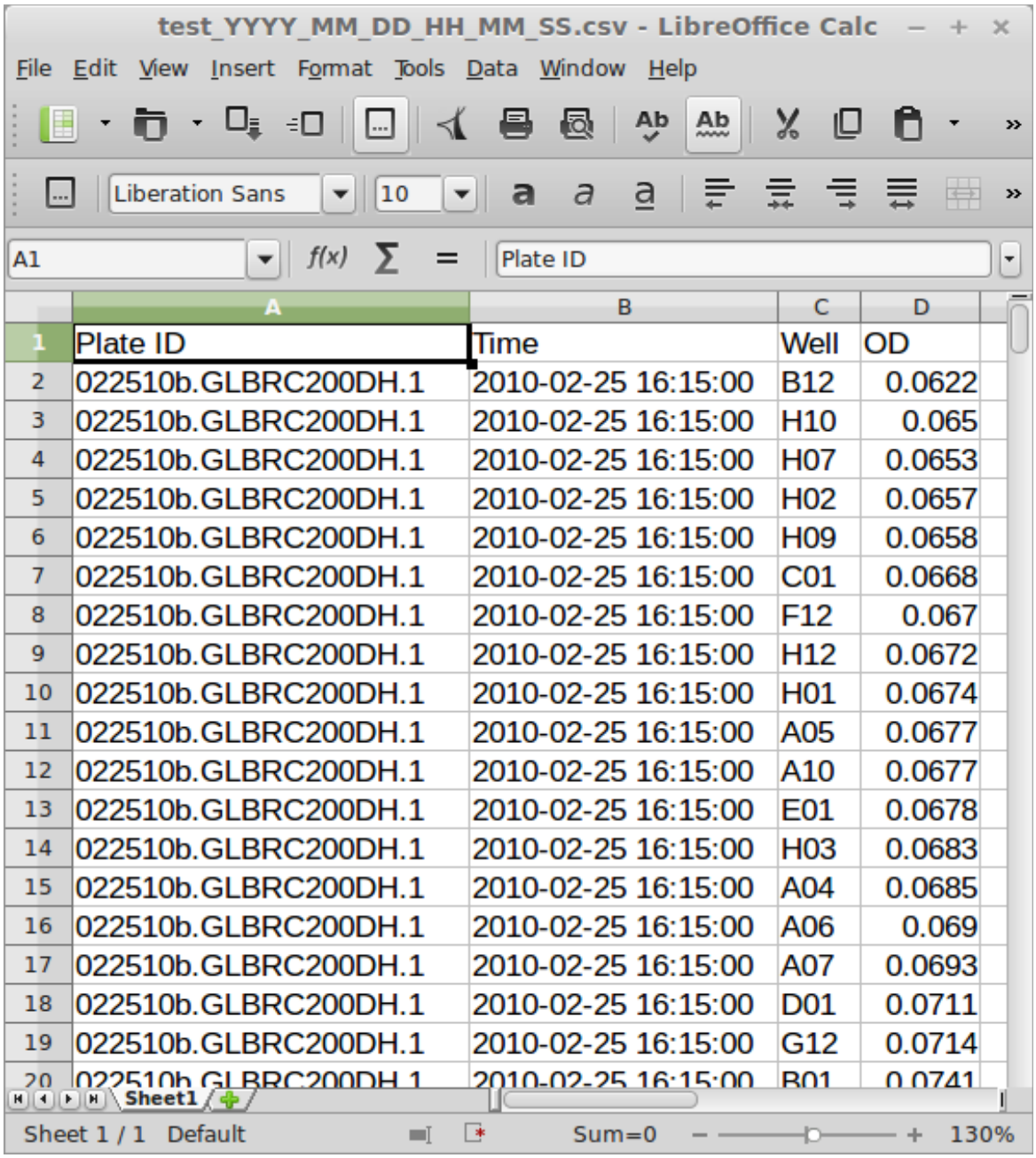
Input needs to be in a specific format in order to be read by GCAT. Files must be in **.csv** (comma-separated values) format, which can be easily edited and exported from any standard spreadsheet program. In this manual, files are displayed using LibreOffice.

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 are currently supported.

Multiple-plate (long) format

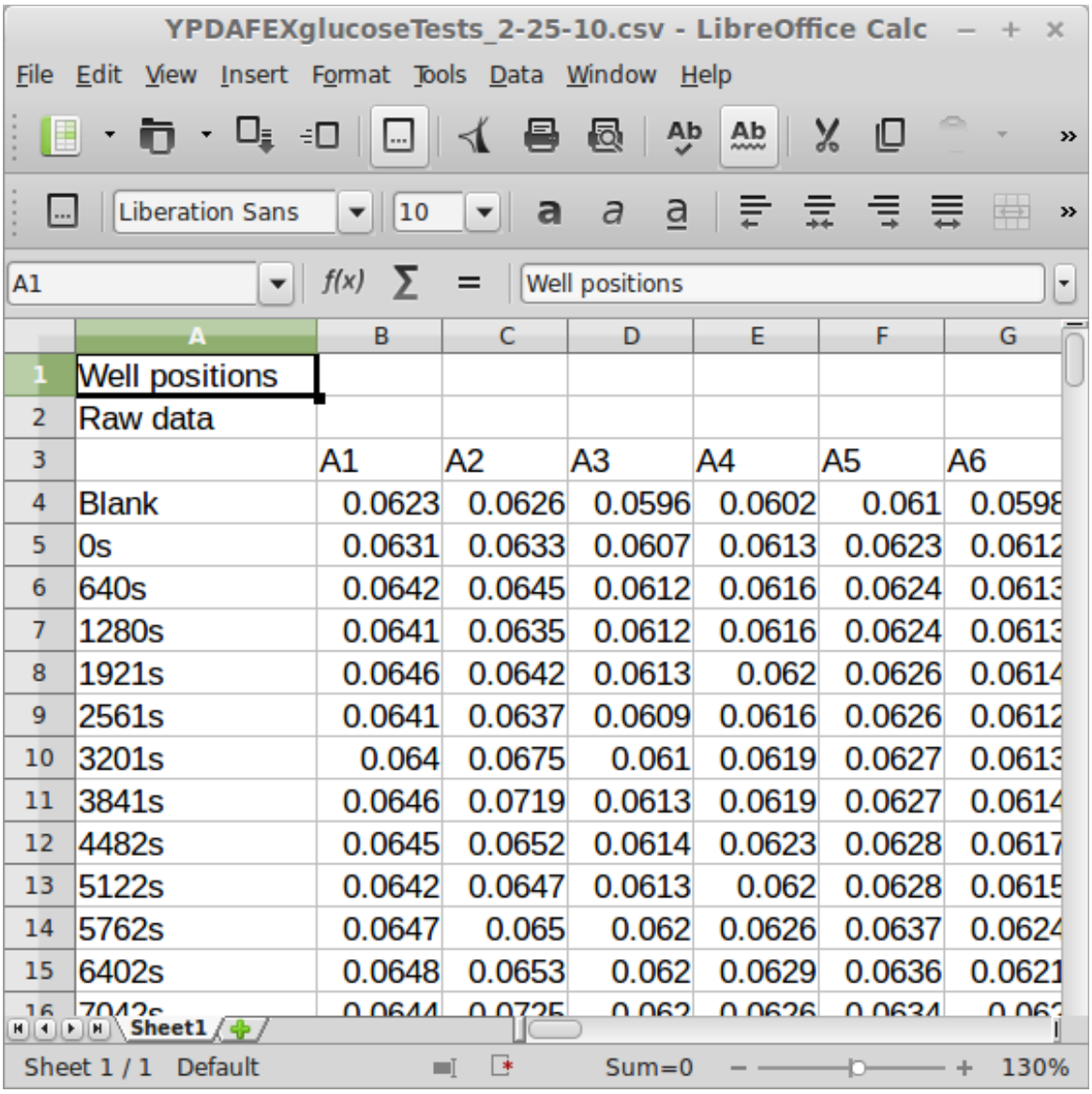
Files in this format contain optical density reads in a long column, with additional columns identifying the plate ID, timestamps, and well names. These columns must be clearly labeled “Plate ID”, “Time”, “Well” and “OD”. Input does not need to be in a specific order, but please ensure that the Plate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



	A	B	C	D
1	Plate ID	Time	Well	OD
2	022510b.GLBRC200DH.1	2010-02-25 16:15:00	B12	0.0622
3	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H10	0.065
4	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H07	0.0653
5	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H02	0.0657
6	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H09	0.0658
7	022510b.GLBRC200DH.1	2010-02-25 16:15:00	C01	0.0668
8	022510b.GLBRC200DH.1	2010-02-25 16:15:00	F12	0.067
9	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H12	0.0672
10	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H01	0.0674
11	022510b.GLBRC200DH.1	2010-02-25 16:15:00	A05	0.0677
12	022510b.GLBRC200DH.1	2010-02-25 16:15:00	A10	0.0677
13	022510b.GLBRC200DH.1	2010-02-25 16:15:00	E01	0.0678
14	022510b.GLBRC200DH.1	2010-02-25 16:15:00	H03	0.0683
15	022510b.GLBRC200DH.1	2010-02-25 16:15:00	A04	0.0685
16	022510b.GLBRC200DH.1	2010-02-25 16:15:00	A06	0.069
17	022510b.GLBRC200DH.1	2010-02-25 16:15:00	A07	0.0693
18	022510b.GLBRC200DH.1	2010-02-25 16:15:00	D01	0.0711
19	022510b.GLBRC200DH.1	2010-02-25 16:15:00	G12	0.0714
20	022510b.GLBRC200DH.1	2010-02-25 16:15:00	B01	0.0741

Single-plate (wide) format

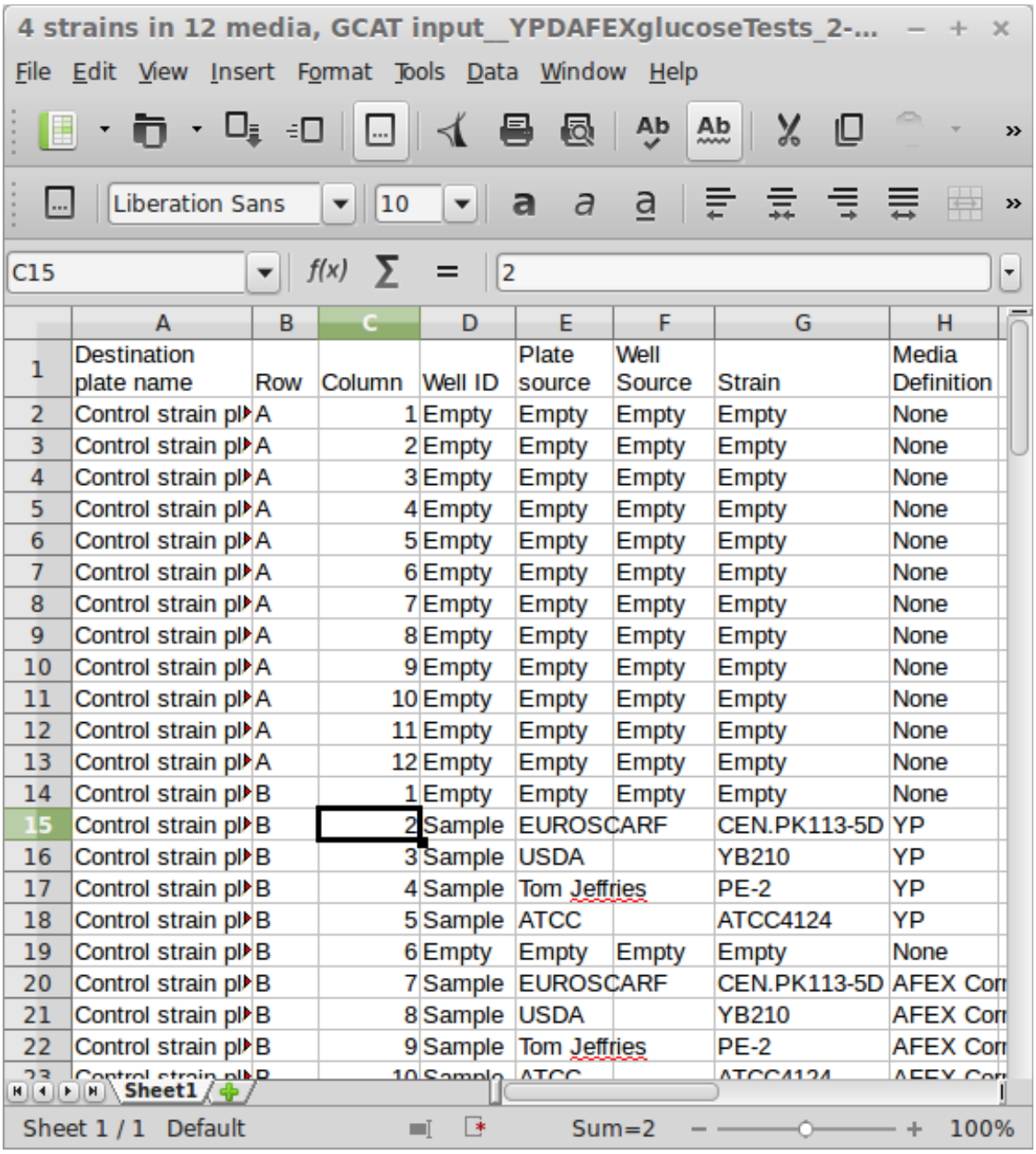
Files in the single plate format and contain optical density reads in multiple columns (one for each well position), and one timestamp column, the header of which should be left blank. The density reads must be labeled with the names of the wells, starting with “A1” and going through all columns (numbers) before advancing any rows (letters).



	A	B	C	D	E	F	G
1	Well positions						
2	Raw data						
3		A1	A2	A3	A4	A5	A6
4	Blank	0.0623	0.0626	0.0596	0.0602	0.061	0.0598
5	0s	0.0631	0.0633	0.0607	0.0613	0.0623	0.0612
6	640s	0.0642	0.0645	0.0612	0.0616	0.0624	0.0613
7	1280s	0.0641	0.0635	0.0612	0.0616	0.0624	0.0613
8	1921s	0.0646	0.0642	0.0613	0.062	0.0626	0.0614
9	2561s	0.0641	0.0637	0.0609	0.0616	0.0626	0.0612
10	3201s	0.064	0.0675	0.061	0.0619	0.0627	0.0613
11	3841s	0.0646	0.0719	0.0613	0.0619	0.0627	0.0614
12	4482s	0.0645	0.0652	0.0614	0.0623	0.0628	0.0617
13	5122s	0.0642	0.0647	0.0613	0.062	0.0628	0.0615
14	5762s	0.0647	0.065	0.062	0.0626	0.0637	0.0624
15	6402s	0.0648	0.0653	0.062	0.0629	0.0636	0.0621
16	7042s	0.0644	0.0725	0.062	0.0626	0.0634	0.062

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:



	A	B	C	D	E	F	G	H
1	Destination plate name	Row	Column	Well ID	Plate source	Well Source	Strain	Media Definition
2	Control strain pl	A	1	Empty	Empty	Empty	Empty	None
3	Control strain pl	A	2	Empty	Empty	Empty	Empty	None
4	Control strain pl	A	3	Empty	Empty	Empty	Empty	None
5	Control strain pl	A	4	Empty	Empty	Empty	Empty	None
6	Control strain pl	A	5	Empty	Empty	Empty	Empty	None
7	Control strain pl	A	6	Empty	Empty	Empty	Empty	None
8	Control strain pl	A	7	Empty	Empty	Empty	Empty	None
9	Control strain pl	A	8	Empty	Empty	Empty	Empty	None
10	Control strain pl	A	9	Empty	Empty	Empty	Empty	None
11	Control strain pl	A	10	Empty	Empty	Empty	Empty	None
12	Control strain pl	A	11	Empty	Empty	Empty	Empty	None
13	Control strain pl	A	12	Empty	Empty	Empty	Empty	None
14	Control strain pl	B	1	Empty	Empty	Empty	Empty	None
15	Control strain pl	B	2	Sample	EUROSCARF		CEN.PK113-5D	YP
16	Control strain pl	B	3	Sample	USDA		YB210	YP
17	Control strain pl	B	4	Sample	Tom Jeffries		PE-2	YP
18	Control strain pl	B	5	Sample	ATCC		ATCC4124	YP
19	Control strain pl	B	6	Empty	Empty	Empty	Empty	None
20	Control strain pl	B	7	Sample	EUROSCARF		CEN.PK113-5D	AFEX Cor
21	Control strain pl	B	8	Sample	USDA		YB210	AFEX Cor
22	Control strain pl	B	9	Sample	Tom Jeffries		PE-2	AFEX Cor

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.

Media background

This is the value that is subtracted from each OD measurement. By default, GCAT expects the user to manually enter a single background value that will apply across the entire experiment. This is appropriate when the media in all wells is the same.

Alternatively, GCAT can take the first time point of every curve to be a blank read and subtract its OD value from the other measurements in the curve. For example, if different wells have different growth media, a user may measure well ODs prior to inoculation and insert these values as the first row into the GCAT input spreadsheet.

GCAT also allows the user to set background OD to 0

OD Transform

GCAT transforms optical density to before proceeding with model fits, as described in the Statistical methodology section. *log(x)* is the default transform. *log(x)* is appropriate when OD of the inoculum is significantly higher than the background, empty medium value. Alternatively, *log(x+0.1)* or *log(x+δ)*, where δ is specified by the user, can be used if the OD of the inoculum is unknown or relatively small compared to the blank medium. In particular, if a user desires to reproduce GCAT's old default behavior, i.e. prior to version 5.0, they need to set δ to *1*.

Growth curve model

GCAT can fit a global sigmoid curve, choosing the best model as described in Statistical methodology, or a LOESS local regression curve. If LOESS is chosen, the user can set smoothing parameter value. This determines the proportion of data points used by the local regression algorithm. It should generally be $> 1/N$, where N is the number of points in the growth curve. Furthermore, the default value of *0.1* is not appropriate when N is relatively small, as it will not result in a smooth line. If this happens, try larger values to achieve smoothing.

Inoculation timepoint

Users should enter the **index** of the point at which the wells are inoculated with culture. By default, this is set to the first read (“1”). Set it to “2” if your first read is blank. This setting 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 aberrant 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.

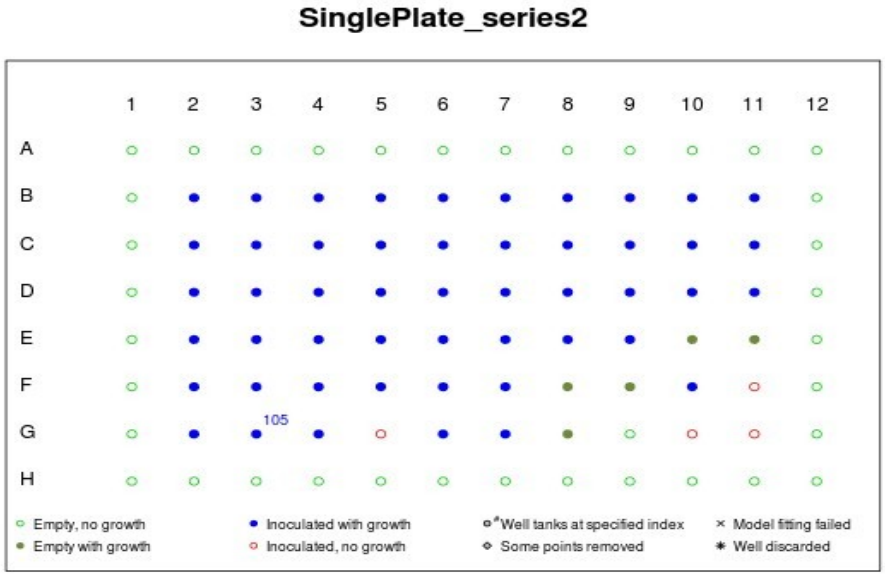
For Single-plate format, input time must be in seconds. The x axis of the well-by-well graphs is on a scale of hours.

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 model fit, and various error flags. A legend is included for the points

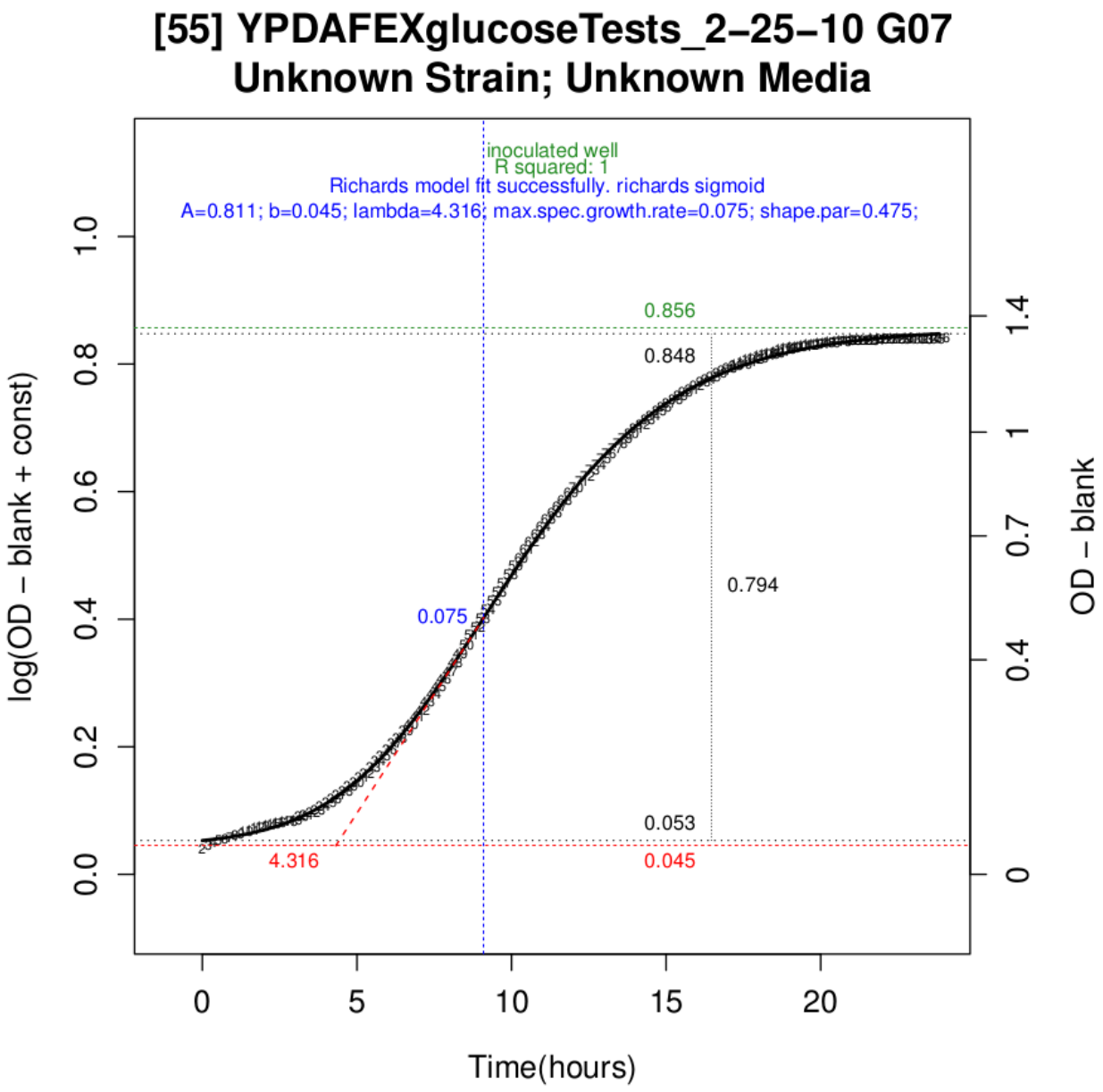


representing each well:

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 green, 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 aberrant jumps in OD between adjacent reads, 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. Grey points have been ignored for the analysis. If the fit was successful, a graph of the model equation is shown, overlaid on the points.

The plot is on drawn on the transformed logarithmic scale (with “const” signifying OD transform parameter δ). This shows the values that were actually used in growth curve analysis. An additional vertical axis on the right is back-transformed to the linear OD scale.

The following lines are also drawn and labeled with their numerical values:

Dashed red (horizontal at bottom): The baseline value (parameter *b*) of the fitted model

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

Dotted black (horizontal at top): The maximal fitted log OD value reached during the time course

Dashed green (horizontal at top): The fitted upper limit for growth (*plateau = A +b*)

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, i.e. inflection point

Dashed red (tangent): The exponential growth trend at the point of maximal growth. The intersection of this line with the baseline 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.

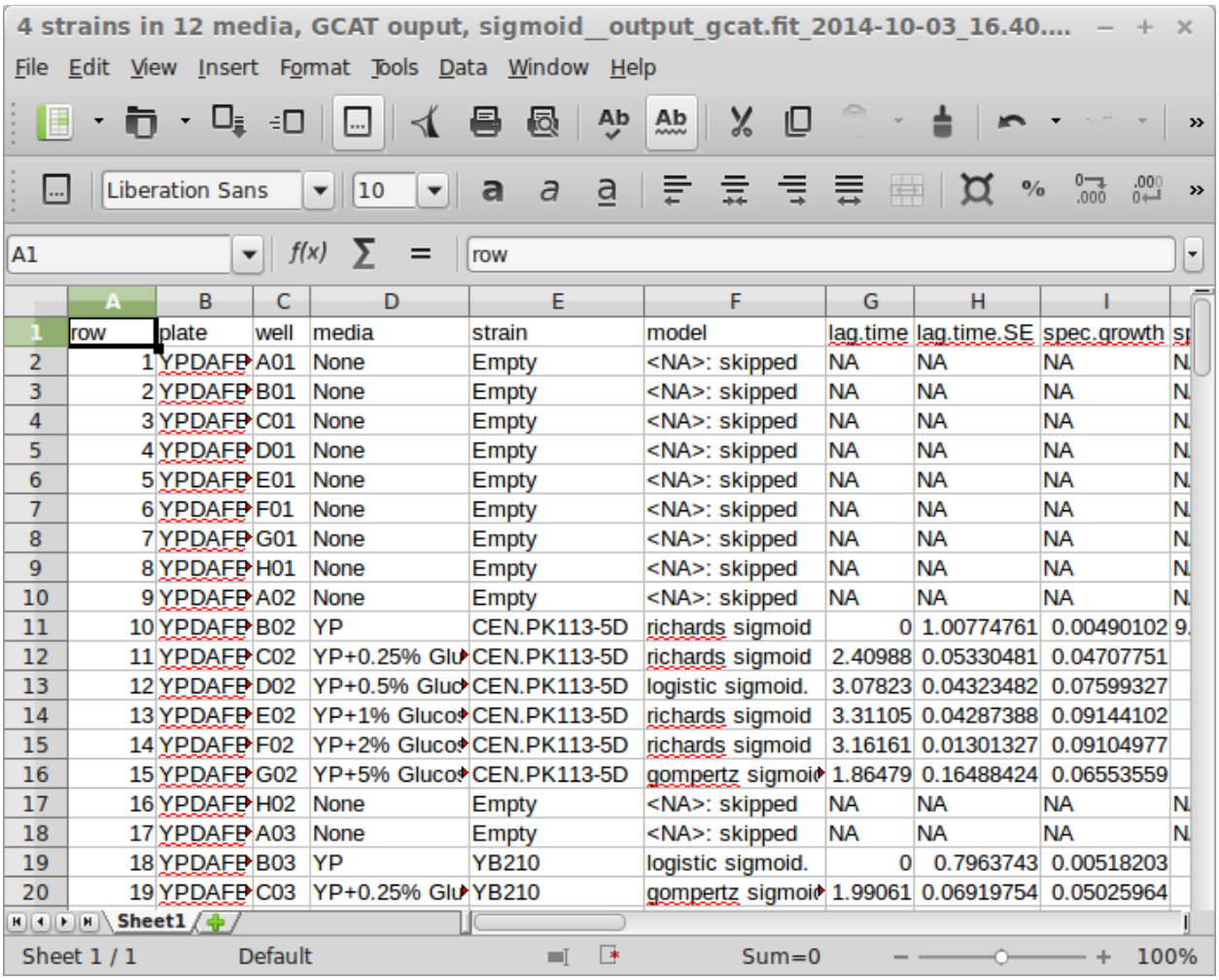
1) Shows any errors detected by GCAT (growth in empty wells or no growth in inoculated ones, tanking trends, jumps in OD, etc.) and will be red if there were any.

2) R^2 value. This is a goodness of fit metric. It should be close to 1.

3) Growth curve model and its parameters.

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.



	A	B	C	D	E	F	G	H	I	
1	row	plate	well	media	strain	model	lag.time	lag.time.SE	spec.growth	sp
2	1	YPDAFP	A01	None	Empty	<NA>: skipped	NA	NA	NA	N
3	2	YPDAFP	B01	None	Empty	<NA>: skipped	NA	NA	NA	N
4	3	YPDAFP	C01	None	Empty	<NA>: skipped	NA	NA	NA	N
5	4	YPDAFP	D01	None	Empty	<NA>: skipped	NA	NA	NA	N
6	5	YPDAFP	E01	None	Empty	<NA>: skipped	NA	NA	NA	N
7	6	YPDAFP	F01	None	Empty	<NA>: skipped	NA	NA	NA	N
8	7	YPDAFP	G01	None	Empty	<NA>: skipped	NA	NA	NA	N
9	8	YPDAFP	H01	None	Empty	<NA>: skipped	NA	NA	NA	N
10	9	YPDAFP	A02	None	Empty	<NA>: skipped	NA	NA	NA	N
11	10	YPDAFP	B02	YP	CEN.PK113-5D	richards sigmoid	0	1.00774761	0.00490102	9.
12	11	YPDAFP	C02	YP+0.25% Glu	CEN.PK113-5D	richards sigmoid	2.40988	0.05330481	0.04707751	
13	12	YPDAFP	D02	YP+0.5% Gluc	CEN.PK113-5D	logistic sigmoid.	3.07823	0.04323482	0.07599327	
14	13	YPDAFP	E02	YP+1% Glucos	CEN.PK113-5D	richards sigmoid	3.31105	0.04287388	0.09144102	
15	14	YPDAFP	F02	YP+2% Glucos	CEN.PK113-5D	richards sigmoid	3.16161	0.01301327	0.09104977	
16	15	YPDAFP	G02	YP+5% Glucos	CEN.PK113-5D	gompertz sigmoi	1.86479	0.16488424	0.06553559	
17	16	YPDAFP	H02	None	Empty	<NA>: skipped	NA	NA	NA	N
18	17	YPDAFP	A03	None	Empty	<NA>: skipped	NA	NA	NA	N
19	18	YPDAFP	B03	YP	YB210	logistic sigmoid.	0	0.7963743	0.00518203	
20	19	YPDAFP	C03	YP+0.25% Glu	YB210	gompertz sigmoi	1.99061	0.06919754	0.05025964	

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

- “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)
- “plate”: Plate ID
- “well”: well position.
- “media”: media definition, if the information was provided by the user in the input.
- “strain”: strain ID, if the information was provided by the user in the input.
- “model”: Which model 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>.
- “lag.time”: lag time estimate inferred from the fitted model. Lag time is estimated as the intersection of a tangent drawn to the growth curve at the inflection point with the lower

baseline of the sigmoid model. Ff LOESS is used, baseline is the same as the lowest predicted response value.

- “inflection.time”: inflection time point of the growth curve when drawn on the log scale.
- “max.spec.growth.rate”: maximum specific growth rate estimate inferred from the fitted model. Estimated as the first derivative of the growth curve at inflection time point when drawn on the log scale
- “baseline”: growth curve baseline on the log scale. If a global sigmoid model is used, this is parameter 'b' of the model. If LOESS is used, baseline is the same as the lowest predicted response value
- “amplitude”: difference between upper plateau and baseline values. If a global sigmoid model is used, this is parameter 'A' of the model. For LOESS, amplitude = max.log.OD - min.log.OD
- “plateau”: upper asymptote value of the fitted model for sigmoid models, i.e. $b+A$. Highest fitted value for LOESS.
- “inoc.log.OD”: log transformed adjusted OD value at inoculation. Estimated value from the fitted model is used, rather than the actual measurement
- “max.log.OD”: maximal log.OD value reached during the experiment. Estimated value from the fitted model is used, rather than the actual measurement
- “projected.growth”: maximal projected growth over inoculation value, i.e. plateau - inoc.log.OD, in global sigmoid models
- “achieved.growth”: maximal growth over inoculation value actually achieved during the experiment, i.e. max.log.OD - inoc.log.OD
- “shape.par”: the shape parameter ν of the Richards equation. Reported for Richards sigmoid models only
- “shape.par.SE”: standard error of the shape parameter ν of the Richards equation. Reported for Richards sigmoid models only
- “R.squared”: goodness of fit metric R^2, also known as “coefficient of determination”. R.squared is usually between 0 and 1. A value close to 1 indicates good fit.
- “RSS”: residual sum of squares, another goodness of fit metric. Smaller values indicate better fits.
- “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.
- “asymp.not.reached”: 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.
- “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.
- “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).
- “pdf.file”: name of the pdf file containing individual well plots
- “page.no”: the page in the pdf file that contains the well plot
- “Destination.plate.name”, “Well.ID”, “Plate.source”, “Well.Source”, “Media.Definition”: annotation columns from the layout file
- “.SE” columns: standard errors. Reported for those values that are estimated directly as parameters of global sigmoid models
- “.OD” columns: values back-transformed from logarithmic to linear, OD – blank, scale

References

Birch, Colin P. D. 1999. “A New Generalized Logistic Sigmoid Growth Equation Compared with the Richards Growth Equation.” *Annals of Botany* 83 (6): 713–23. doi:10.1006/anbo.1999.0877.

Werner-Washburne, M., E. Braun, G. C. Johnston, and R. A. Singer. 1993. “Stationary Phase in the Yeast Saccharomyces Cerevisiae.” *Microbiological Reviews* 57 (2): 383–401.

W. S. Cleveland, E. Grosse, and W. M. Shyu. 1992. “Local Regression Models.” In *Statistical Models in S*, edited by J.M. Chambers and T.J. Hastie. Wadsworth & Brooks/Cole.

Zeide, Boris. 1993. “Analysis of Growth Equations.” *Forest Science* 39 (3): 594–616.

Zwietering, M. H., I. Jongenburger, F. M. Rombouts, and K. van ’t Riet. 1990. “Modeling of the Bacterial Growth Curve.” *Applied and Environmental Microbiology* 56 (6): 1875–81.