**Determining Respiration, Reaeration and Primary Production Rates from Stream Metabolism Measurements**

**A User Manual for ModelMaker**

**Mike Grace**

This manual provides guidance on how to use the model for extracting stream metabolism parameters from diel dissolved oxygen data. The model was written by Dr Mike Harper and Dr Mike Grace of the Water Studies Centre at Monash University, Melbourne Australia. The model is written using the modelling package ModelMaker and it is assumed that the intending user has access to this modelling package. If not, it can be purchased from the distributor, ModelKinetix (AP Benson): http://www.modelkinetix.com/modelmaker/

Mike Grace gratefully acknowledges the feedback and advice provided in developing this manual, especially from Natalie Griffiths and Peter Levi from the research group of Assoc. Prof Jennifer Tank at Notre Dame University, South Bend, Indiana, USA.

**Description of Program**

The approach taken in this program to extract the 3 parameters controlling stream metabolism:

1. Gross Primary Production (GPP)
2. Community Respiration (CR)
3. Reaeration (E)

where DO = GPP - CR ± E (Equation 1)

is based on the ‘day-time’ regression method proposed by Kosinski (1984), where the photosynthesis term in Equation 1 is replaced by AIp, where A is a constant, I is the surface irradiance and p is an exponent reflecting the ability of the primary producers (predominantly benthic algae) to utilize the incident light (accounts for saturating photosynthesis).[[1]](#footnote-1) The reaeration term ‘E’ is replaced by KD where K is the reaeration constant and D is the oxygen deficit – the difference between the measured DO concentration and the 100% saturation value. (Butcher and Covington 1995) demonstrated that incorporating the temperature dependence of the reaeration and respiration terms resulted in much better agreement between modelled and measured DO concentrations, especially where there was a substantial diel fluctuation in temperature.

Equation 2, shown below, is the basis for the data fitting procedure and also incorporates the temperature dependence of reaeration (Kilpatrick *et al.* 1989).[[2]](#footnote-2) Several studies have used a temperature dependence factor, , of 1.072 for respiration based on the respiration rate doubling for every 10º temperature increase.

 (Equation 2)

**Data Considerations:**

There are three different data requirements to run the model:

1/ Dissolved oxygen expressed as % saturation

2/ Water temperature

3/ PAR

Ideally, each of these parameters has been recorded concurrently at exactly the same time interval, over at least a 24 hour period. The model will work for shorter periods but the uncertainty in parameter estimates becomes unusably high. It is recommended that data be collected at intervals of 5 minutes. Longer time intervals can provide good estimates of the metabolic parameters provided that the system under investigation is relatively productive. It is obviously critical that the data be as accurate as possible – biased data will always lead to incorrect parameter estimations. This subject is covered extensively in the Technical Manual for Stream Metabolism and won’t be discussed again here.

The model works using elapsed time in seconds, rather than time-of-day. There is no necessity that the data starts at a particular time of day. Common choices include the 24 hour period from midnight to midnight (which offers the attraction of being able to report metabolism for a calendar day), near dusk (or dawn) to near dusk (or dawn) which offers the advantage of analysis of a complete single day and single night and from noon to noon. Personal experience has found that the last option causes the most difficulties in obtaining good fits to the diel curve if the weather on the two days is different (especially if one day is clear and the next is very cloudy).

PAR is best measured on-site with appropriate light loggers, however, in the absence or malfunction of such loggers, light from a relatively near meteorological station can be perfectly adequate. One cautionary note is to be wary if the met station is too far away, then on cloudy days, there can be a lag between the apparent cloud cover according to the met data and the light experienced at the site and the resultant GPP effects on DO. This is most problematic in very productive sites. The units for PAR are not important (I usually use Es/m2/s, but other units are fine too –this will simply change the value of the pre-exponential term A in the GPP component of the model).

**Comments on data filtering.**

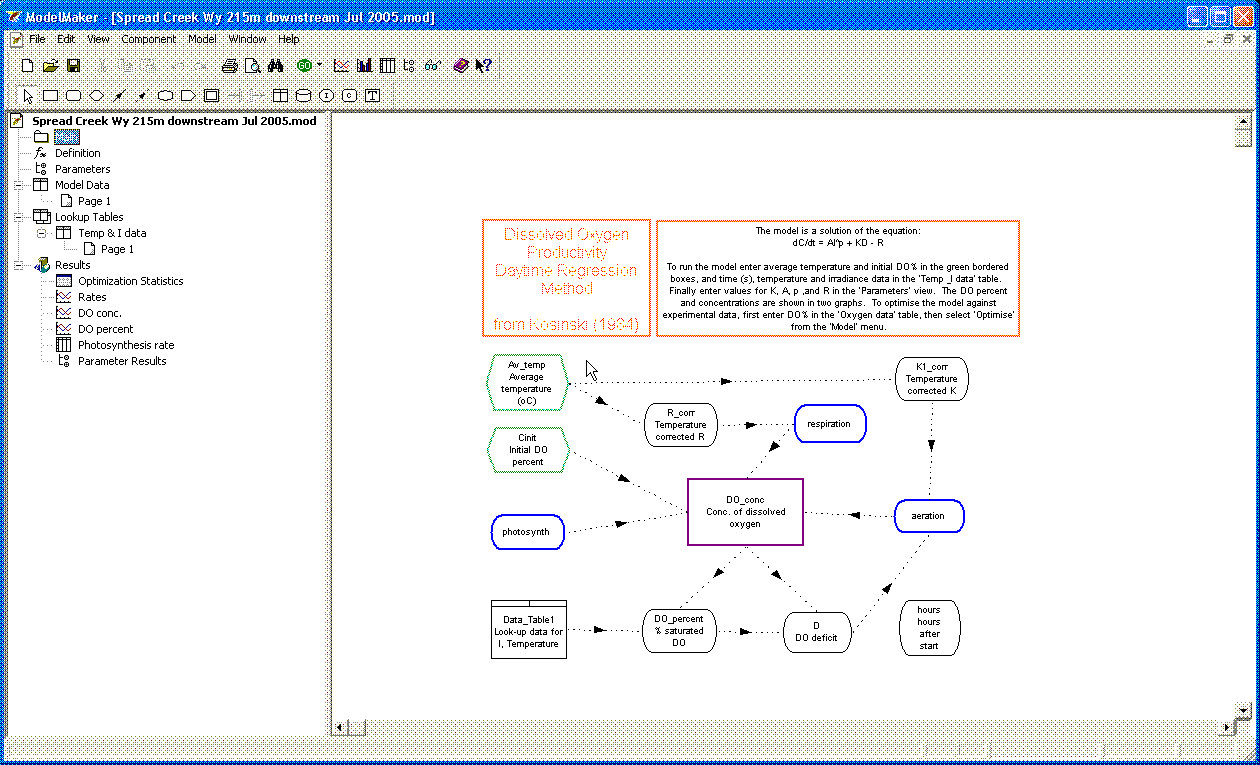
I’m currently investigating the impact of various data filtering and/or smoothing functions on the resultant estimates of GPP, CR and KO2. Data smoothing will often substantially decrease the time required to obtain a fit to the diel curve. ‘Outliers’ will also cause the fitting procedure to struggle in some cases (if the outlier is sufficiently different to the rest of the data) and can result in poor fits and erroneous data. The user is always recommended to closely examine the diel trace (simply plotting it in Excel is usually perfectly sufficient) to look for outliers and to remove them if possible. If it’s a single point, I usually just replace the outlier with the average of the point before and point after. If data sampling frequency is sufficiently high (e.g. 5 minutes), it is extremely unlikely that a single point would deviate dramatically from the points around it.

If several consecutive data points appear to be ‘off the smooth curve expected’, then much greater diligence is required. The deviation may be due to a problem with the probe (e.g. a leaf or other material has lodged in close proximity to the probe and affected the DO response) or it might be a real change in DO. Such ‘natural’ departures are most commonly observed in small, highly productive streams on days of alternating strong sunlight and intense cloud. Such circumstances are challenging to fit!

**Moving around the model and basic operations**

This section describes how to run the model, update the parameters and view the outputs. The operation of the ModelMaker software is a little confusing initially until the user recognises that there are two windows associated with the parameters –parameter input and parameter output. This will be hopefully made clear in the following description. This section uses screen captures from a working model to illustrate key points.

On first opening an existing model, the main screen will appear like:

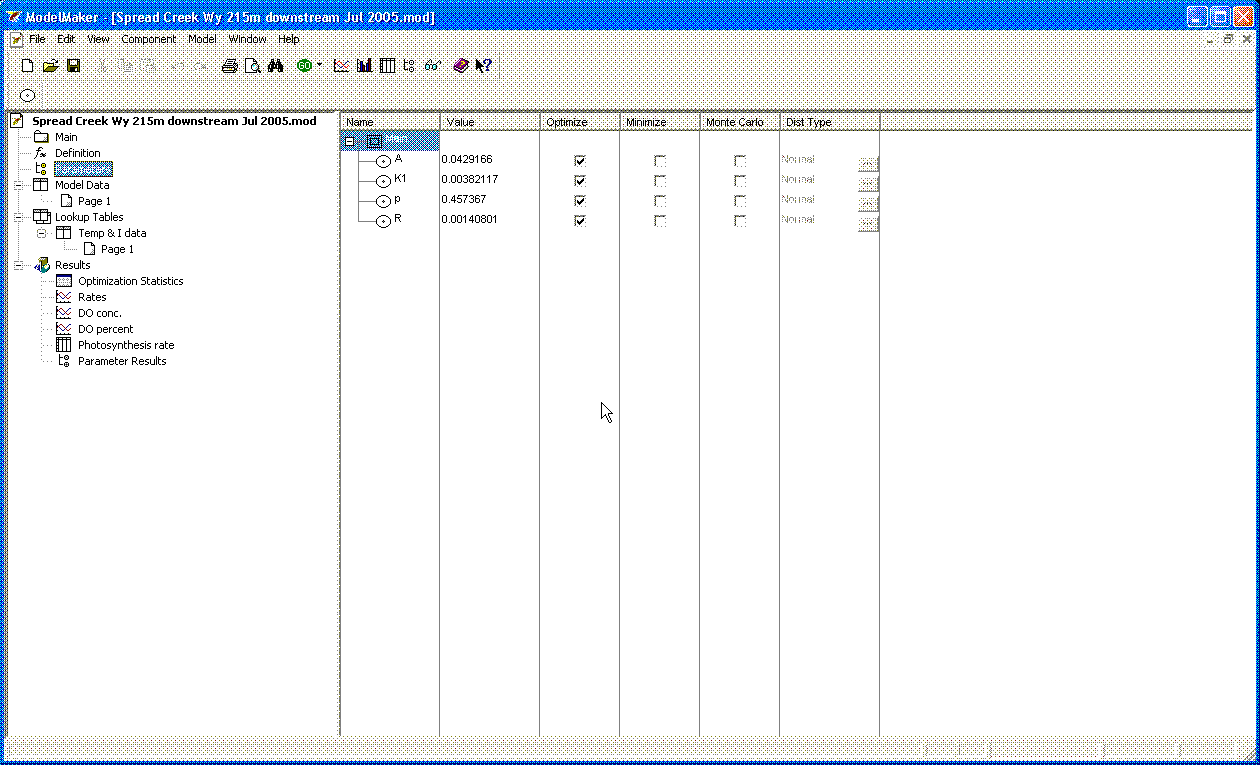


On the LHS of the figure, the various tables for data entry are listed, along with the links to the parameter windows and the graphical results displays. The RHS contains a view of the model structure (this is the first link, labelled ‘Main’ on the LHS below the name of the current model). Clicking on a link will open that link (table, window, pop-up box, graph) on the RHS.

Apart from seeing how the model is put together, the only time the user need access the ‘Main’ Window is to enter the data for the average temperature and the initial %DO (see below).

When you first open ModelMaker, there will be a “blank” model called ‘no name’. The first task is to open an existing diel curve analysis model. This is either available from me, or you can simply update one of yours. My strategy for a new data set is to open an existing model then immediately save it with the file name of the new model to be developed. This will prevent you accidentally overwriting the existing model. Then enter the appropriate data (see below) and save the model again.

Double click on the icon ‘Parameters’ – third item. Don’t confuse this with ‘Parameter Results’. You should now see on the RHS a small table, as shown below:



The first column, titled ‘Name’ (for the ‘Main’ model – which in this case is the only model) contains the names of the parameters to be fitted - A, p, K1 (i.e. KO2) and R. The second column contains the values for these parameters used in the most recent ‘run’ of the model (more on this later). The third column contains check boxes, where you can choose to optimize a particular parameter or not (should almost always be selected, unless you’re trouble-shooting, or needing to force a parameter to be constant to help sort out a particularly difficult curve). Unless you’re pretty experienced with running these models, don’t worry about the other columns – the last two allow you to do simulations.

Now double click on the LHS icon called “Page 1”. If it is not present, double click on Model Data.

There should be two marked columns – one is ‘time’ (in seconds) and the second is %DO (called DO\_percent). Copy the %DO data from the Excel spreadsheet and paste it into the second column. If data is already present, it can simply be cleared by clicking on the cell “DO\_percent (all cells in this column will then turn black) and pressing “delete” or cut. New data is then pasted in.

\* if the time interval in the first column is not correct, then create a column of time data in Excel and paste into the first column. A reminder, the time data should be 0, 300, 600, 900 etc assuming a 5 minute sampling interval.

Note that ModelMaker is very particular about the time data (it is used as the basis of the data referencing and analysis) so do not make any mistakes with this. A common trap to avoid is that the time data doesn’t extend as far as the DO (or temp or light) data. Easy to check! It does not matter if the time goes beyond the data, provided that later on when running and optimising the model, you select an upper time boundary that has all 3 other bits of data (i.e. DO, temp, light).

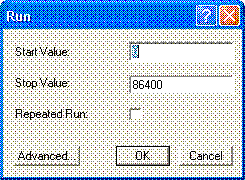
Now double click on the lower LHS icon called “Page 1”, which is under ‘Lookup Tables’ and ‘Temp & I data’. Enter the temperature and light data in the appropriate columns by cutting and pasting from the Excel spreadsheet. Note that the light data should contain 0s if there is no light reading. The model will throw an error if it finds a missing value in any of these tables.

Once the data entry into the tables is complete, double-click on the ‘Main’ box and then enter the values for the average water temperature during the period of deployment being analysed (easy to get this number form the Excel spreadsheet) and for the initial value of the %DO by double-clicking on the appropriately named green hexagons on the upper left side of the main sheet. The numbers are then entered into the box labelled ‘Equation’, followed by OK. It will be pretty obvious if the starting %DO is incorrect (the model fit will start from the wrong point), but having the wrong average temperature will not be so apparent.

Now you’re ready to run the model with the existing parameters (see later if you want to enter new values for these parameters). It is important to emphasize at this stage that there are two separate parts to fitting the data and extracting the model. They are interlinked but the distinction is crucial. *Running* the model simply uses the time, temperature and light data and the parameter values listed in the ‘Parameters’ window. It will calculate the diel curve from this information without any reference to the DO-percent values. It is this stage that will produce the ‘goodness of fit’ statistics and should always be the last step when optimization (see next) has been completed. *Optimization* uses one of two algorithms (Simplex or Marquadt) to improve the parameter estimates by iteratively changing these parameters to achieve the best possible convergence between the measured values of %DO and the calculated values based on the most recent parameter estimates. Chi-squared is used as the metric to determine the goodness of fit and convergence. ‘Running’ the model and ‘Optimizing the model’ require different steps and these are explained now.

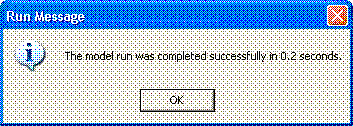
**Running the existing model (i.e. with current parameter values):**

Select the first item (‘Integrate’) under the ‘Model’ menu at the top of the screen. This will bring up a dialog box as shown below:



You then need to enter the starting time and the finishing time (in seconds) corresponding to the data sets in the respective ‘Table 1’. “86400” is listed in this diagram as it represents one day, so I’m asking the model to run for exactly one day from the first data point in the data set. It is easy to run the model at any other starting point, but please remember to then enter the appropriate values of average water temperature and the initial %DO (at the new starting time). A *very* common thing to forget!

You should hopefully then see a window such as the following:



Remember that running the model is very quick as there are no iterative calculations required.

Once the model has been run, have a look at the various windows under ‘Results’ on the LHS. Of particular interest is the Graph icon labelled “DO percent”. Click on this and you will get a plot showing the experimental values and the calculated values for %DO saturation. Remember (especially if the fit is poor at this stage), no optimization has yet been done.

Note that the ‘Optimization Statistics’ are only updated when you optimize the model. So don’t trust the values presented in this particular window (including r2) until you’ve run one optimization (more on this a bit later). At this stage as well, ignore the window labelled ‘Parameter Results’ as no parameter optimization has occurred.

**Optimizing model (i.e. improving the parameter estimates):**

Before starting on this section, I think it is worthwhile having a mental image of what’s happening in this stage. A useful image for me is to think of the solutions to the model (simply all values of all parameters) and the resultant measure of goodness of fit (sums of squares – basically difference between all the experimental data and the corresponding points)[[3]](#footnote-3) as being a landscape. Some combinations of parameters will result in really high sums of squares (i.e. poor fits), equating to ‘mountains’ in our landscape, while other combinations provide low points in the landscape. What the iteration process will do is to determine the lowest point in the landscape and that becomes the set of optimized values for the parameters.

Don’t forget to:

1. Enter the following information into the respective cells on the main model sheet:
   1. average temperature for the duration of the data that you are analysing (this may not be the average temp for the entire data set)
   2. The starting value for the %dissolved oxygen saturation (this is simply the first value in your data set to be analysed)

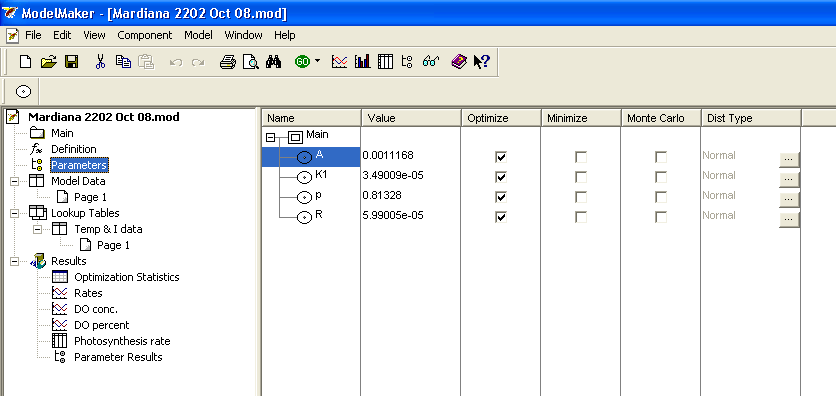
**Strategy for Data Fitting**

\* The key principle here is “Under no circumstances simply accept the results of the first data fitting”.

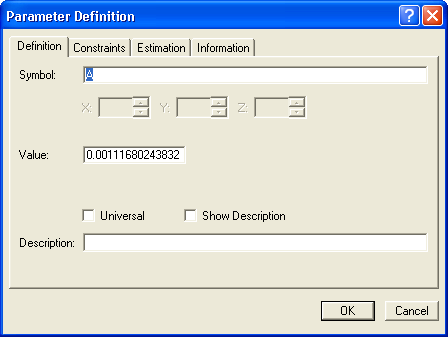
You should spend a bit of time ensuring that you have the real solution set and not a local minima in the solution space. As you become more experienced with the program, the changes you undertake should be fairly intuitive. However, that doesn’t help relatively new users. So here are some suggestions:

1. Most important! Think about the magnitudes of the parameters produced by the data fitting. Are they reasonable? I’ve included an appendix that lists ‘normal’ ranges for GPP, CR & E for different stream types (Tech. Manual, 2007). These values should be used as a guide only, as it is possible that you’re stream is behaving differently from the normal. e.g. an open agricultural stream with high nutrients, slow flow and warm-hot temperatures, might have extremely high rates of GPP and CR.
2. Start off with using the values in the appendix to provide initial estimates for the parameters R and K. If the stream is fast flowing and turbulent, choose a K value towards the high end of the range e.g. 50 d-1 for a stream. This equates to ca. 6 x 10-4 s-1. Similarly, select an R value [[4]](#footnote-4) of say 20 mg O2 L-1 d-1 corresponding to 2.4 x 10-4 mg O2 L-1 s-1. Allow p to begin at 0.1. This parameter is constrained to the range 0 – 1, so do not enter a value outside this range, otherwise you will get an error stating that parameter p has violated its range. A p value of one indicates no photoinhibition of photosynthesis (i.e. photosynthesis increases linearly with increasing light intensity) while a value of 0 reflects total inhibition (no photosynthetic activity whatsoever. Try an initial value of A = 0.001. I suggest you write down these initial estimates as one of the most important aspects of obtaining a reliable fit to the data is to run the model several times with quite different starting values for the parameters. If the model always converges to the same final values, then much greater confidence can be placed in the validity of the results. NEVER run the model only once and believe the results!! They may well be right, but this assumption must be tested.

To enter the initial parameter estimates, click again on the ‘Parameters’ icon (*not* Parameter Results). You will see a window like this:



Double click on each of the parameter labels in turn and enter the initial estimate. e.g. after double clicking on A (highlighted in blue above), the following window will appear:

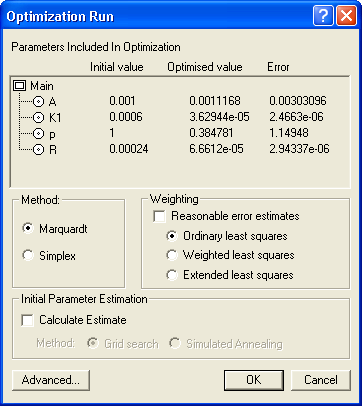


The desired initial estimate for ‘A’ is then entered into the ‘Value’ box. Click OK when done.

Repeat this procedure for each of the 3 remaining parameters. Note that scientific notation is replaced by the ‘e’ notation used in programs such as Excel. Hence 6 x 10-4 becomes 6e-4.

The initial estimates for the parameters are now loaded and the iteration can commence.

1. Under the ‘Model’ menu item, select ‘Optimization’. A window like the one below will appear (the values you have just entered for the initial estimates should appear in this window):



Select the ‘Marquadt’ method, if this is not already selected. The Simplex fitting routine is recommended only for advanced users (according to the ModelMaker User Manual).

Also ensure that the “ordinary least squares” weighting option is selected unless you have a strong reason to choose another option.

The click on the ‘OK’ button and the optimization will commence.

1. Assuming no error occurs, then the run will continue until convergence is reached. If the error “singular curvature matrix” is encountered, reset the initial ‘p’ value to be < 0.4 and try again. If that still causes the same error, then try changing the initial value of ‘A’ at least an order of magnitude (up or down). If p is < 0.4, and changing A an order of magnitude still fails, try increasing R??.
2. Natalie Griffiths and Peter Levi from Notre Dame University, Indiana recommend the following systematic approach to ensuring the correct ‘best fit’ and not some local minima is obtained:
   1. The approach involves doing at least 6 runs per data set - more if the results (r2, ER, GPP) are very different between runs, which seldom happens.
   2. start with a low p (~0.1), then a ‘high’ p (~0.4; if an initial p value much greater than this is used, the 'singular curvature matrix' error is often encountered)
   3. Repeat with low K followed by high K, and finally
   4. Using a low R then a high R (there is usually an order of magnitude between the low and high values of K and/or R used in the initial estimates).
   5. When a new data set is used, the whole approach is restarted – even if this data set is nominally similar (e.g. modelling the metabolism in the same stream on two successive days under similar meteorological conditions).
   6. Once some familiarity has been achieved with the program and the data analysis process, then the data fitting can be speeded up by simply cutting and pasting the new data into an existing model and using the parameter estimates from the previous converged run as the initial estimates for the new run. This approach is not recommended for less experienced users. Even experienced users are strongly recommended to try several different initial parameter sets.
3. Once the model has converged to a final set of parameters, these parameter values, with their associated uncertainties, should be written down.
4. As a final check, the model should be rerun, but this time allowing the program to select the initial parameter estimates using the ‘Calculate Estimate’ function and selecting ‘Simulated Annealing’.  The ‘temperature’ that comes up in the dialog box has nothing to do with the physical temperature of the water. It's a modelling term related to the mathematical annealing process used to do the non-linear fitting (it originates from the origin of "Annealing" - this is a temperature based process used in the treatment/hardening of metals, and was one of the first processes modeled using this mathematical technique).

To do this:

* + - From the model menu select optimize. In the dialog box, tick the check boxes ‘calculate estimate’ and ‘simulated annealing’.

1. Once the final convergence has been achieved, you need to rerun the model with these finalized parameters. To do this:
   1. Go to the parameter results window and right click on each parameter in turn. The option to update the parameter should be selected.
   2. Once all the parameters have been updated, the model can then be run by selecting integrate from the model menu.\* as above.
   3. Now that the model has been optimized and run, the model performance statistics should be noted (e.g. r2, F)
   4. A good fit to the data should produce an r2 of better than 0.9. Certainly in streams with large diel variation in %DO saturation, this should easily be achieved. Goodness of fit will sometimes be less satisfactory in small turbulent streams especially when they are also in forested or other low productivity situations.

**Calculating Daily CR and KO2**

Simply multiply the values from ModelMaker for R and K respectively, by 86,400 (number of seconds in the day).

*Note that the rates are calculated at the average temperature during the deployment.* Segments of Equation 2 can be used, if required, to convert these rates to rates at 25C (or any other desired temperature).

**Calculating Daily GPP**

Once values for A and p have been obtained from the fitting, GPP can be most simply obtained by returning to the original Excel spreadsheet and creating an extra column alongside the PAR column. The PAR is then converted to ‘instantaneous’ GPP by using the formula from Equation 2: i.e. GPP = AIp (where A and P are obtained from the final optimised run in ModelMaker, leaving the units as found in the ModelMaker output). All of these individual ‘instantaneous’ GPP values are summed over the daylight period (this really equates to the whole analysis period, given that non daylight hours are dark and thus have a PAR of zero?? Yes??)and then multiplied by the data sampling interval in seconds (e.g. 300 s). This will give the total GPP in the day. An example spreadsheet is available from me upon request.

**Possible Problems:**

Sometimes the optimization procedure stops with a message saying that the “Optimization was stopped on run (#) of (#) because the limit of machine accuracy was reached”. This is not a concern as it simply means that the program is making incredibly small adjustments to the parameters in an attempt to get an even better fit and these changes are less than the level of precision of numbers in the computer – see the ModelMaker Reference Manual page 53 for a little more detail. This condition is effectively the same as convergence being reached. Parameters can be updated and the model then run.

**References:**

**Appendix: Typical values for GPP, CR and KO2**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Stream Type | Typical Discharge (ML d-1) | Typical Range for KO2 (d-1) | Typical Range for GPP (mg O2 L-1 d-1) | Higher GPP values (mg O2 L-1 d-1) | Typical Range for CR (mg O2 L-1 d-1) | Higher CR values (mg O2 L-1 d-1) |
| Rivers | 20 – 2000 | 0.2 – 20 | 0.1 – 20 | Up to 80 | -1 to -20 | Up to - 40 |
| Creeks | 0.1 – 40 | 5 – 250 | 0.2 – 50 | Up to 280 | -8 to -100 | Up to - 200 |

If data analysis produces parameters outside the typical ranges (and especially if the “higher limits” are exceeded), then the following points need to be considered to determine whether these parameters are correct or an artifact arising from problems with the data analysis:

1. KO2 - if the stream is large and slow flowing (especially if also deep), then low values of KO2 are expected. Values higher than 20 d-1 would only occur if there water is very turbulent (lots of air entrainment)
2. KO2 - values < 5 d-1 in small streams would be highly unusual unless stream flow was slow and nearly laminar with no visible turbulence, riffle zones, cascades etc.
3. KO2 - values > 100 d-1 have frequently been reported in small, turbulent streams. The issue in these circumstances is the impact that any uncertainty in KO2 will have on the estimates of GPP and CR.
4. GPP - values are usually < 15-20 mg O2 L-1 d-1 in larger streams and < 50 mg O2 L-1 d-1 in smaller streams. Values larger than this should be treated with suspicion unless there is significant visible evidence of primary producers in the study reach (e.g. phytoplankton, benthic algae, significant stands of macrophytes – submerged or emergent). It is strongly recommended that when very high GPP values are expected or recorded in the absence of macrophytes, then water samples (and if appropriate, sediment samples) be collected for chlorophyll-a analysis. Conversely, if large amounts of primary producers are observed on days with clear skies and GPP is < 2 mg O2 L-1 d-1, then doubt is cast on this GPP value. Further investigation would be required as other factors might be responsible for the low GPP.
5. CR - values are generally confined within a relatively narrow range of -1 to -40 mg O2 L-1 d‑1. Although GPP can often be negligible due to severe nutrient and/or light limitation, CR values smaller than -1 mg O2 L-1 d‑1 are considered unlikely and would be an immediate cause for investigating the data analysis. Values in excess of -50 mg O2 L-1 d‑1 have been recorded but tend to be limited to streams with highly visible large stores of organic matter (e.g. decaying vegetation, large standing stocks of benthic algae). Again a visual check of the study reach (which should be done anyway!) will determine whether a ‘high’ CR value is sensible. If a high CR value is obtained without any large and obvious source of organic matter, then two aspects should be investigated before discarding the data (but after checking the data analysis has been performed correctly): a) is there a large source of dissolved organic carbon in the stream (perhaps from a pollution event or continuous discharge of effluent)?; or b) is there a significant hyporheic flow in the study reach with a large rate of hyporheic respiration? This possibility should be checked by salt addition. As noted in that section, if groundwater inflow accounts for more than 5%, then the study reach may be considered unsuitable as a monitoring site.

1. I have also used the alternate formulation of Jassby and Platt and this yields essentially the same result. [↑](#footnote-ref-1)
2. I also have a version of this model that determines the temperature dependence of the reaeration rate, . I’m happy to provide this version of the model, but users need to be very aware that adding another parameter to be fitted will inevitably improve the final fit (and hence can’t a priori be used as a reason for including ). Justification for inclusion of the term can best be done using approaches such as the Aikake Information Criterion. [↑](#footnote-ref-2)
3. What I’m stressing here is the importance of the mental image rather than the strict correctness of the mathematical description. It is not really necessary to understand the mathematical detail of the optimization process! Users interested in this maths are directed to the Reference Manual provided by ModelMaker (Chapter 6). [↑](#footnote-ref-3)
4. The initial R value to be entered into the program is positive. The equation used (Eqn 2) recognizes that respiration reduces the DO. [↑](#footnote-ref-4)