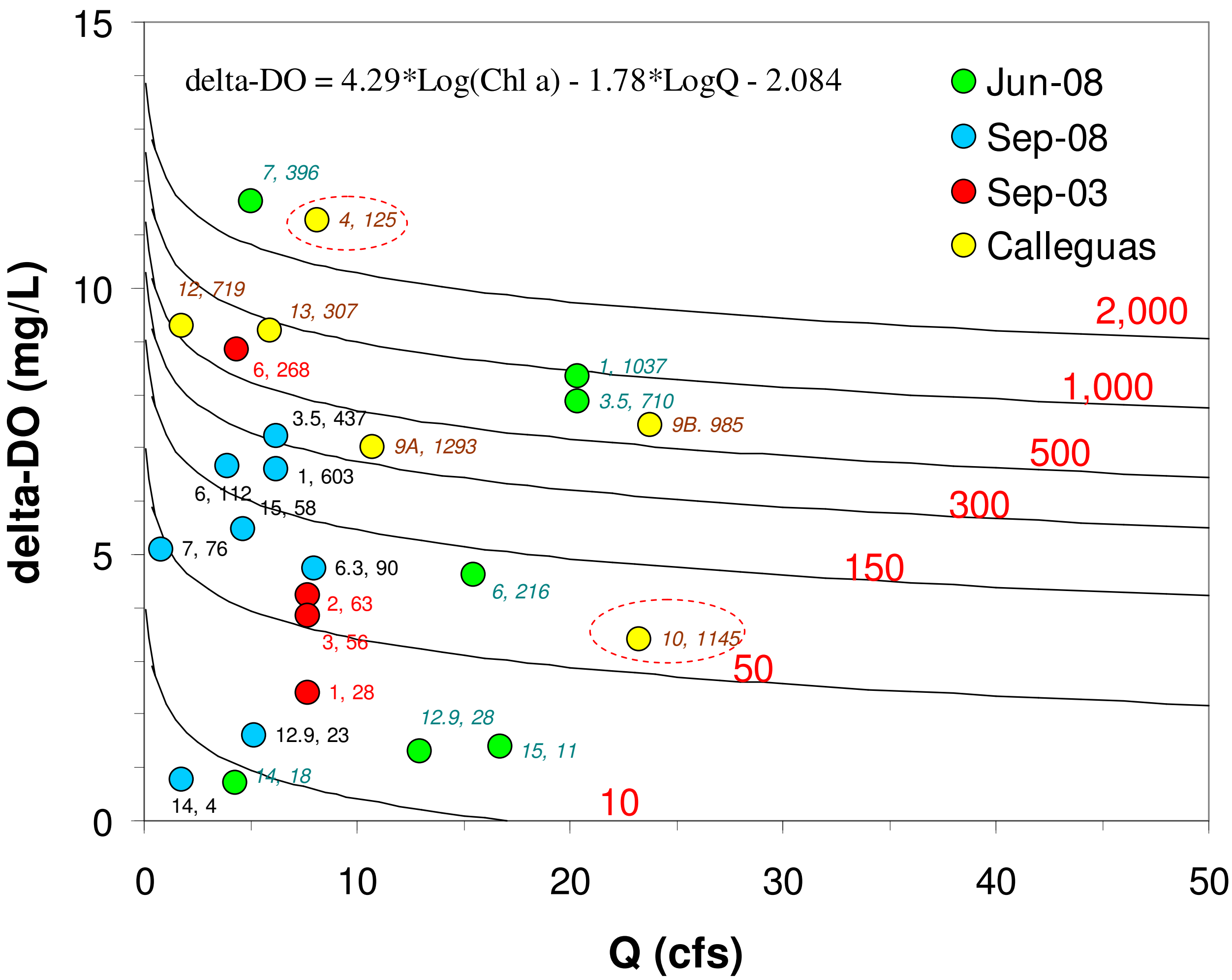

The above chart shows the graphical model developed from the UCSB-LTER study (Chl-a data from Kristie, delta-DO from SBCK and flow from SBCK, USGS and Ventura County gauging records). The regression relationship used in the model is shown on the graph ($p < 0.001$, r-squared = 0.81). On the chart, plotted in green and blue, are the data used to derive the equation; the identifying numbers show the SBCK site codes, followed by measured mean Chl-a densities. By comparing actual measured Chl-a density with the model's lines of equal density plotted on the graph (the red numbers indicate Chl-a in mg/m^2) the accuracy of the model can be judged. In my opinion it does a reasonable job, but one with a large associated error – a large error, in spite of a very high r-square value, being typical for a variable extending over a range of almost three orders-of-magnitude estimated from a log or semi-log regression equation.

Also plotted on the graph are some additional data: 4 values from work done by Julie and myself on the Ventura in September 2003 (in red) (unfortunately, the only occasion we measured both Chl-a and delta-DO within the same time frame), and 6 values from Diana's sonde study on Calleguas Creek in May 2008 (in yellow). The September 2003 data produced one of those "Eureka!" moments; as you can see it fits very well.. (I should mention that this exercise results from a question Ben raised of "how can we verify the model?" To which I'll return at the end of this exercise.

Unfortunately, Diana's data did not fit as well: two, possibly three points, not fitting at all. And that's an understatement for the two points I'll now discuss (#4 – Revlon Slough & #10 – Hill Canyon Reach of Conejo Creek). I've circled both these points on the graph and you can see how anomalous they are: at Revlon Slough negligible algal densities produce a very high delta-DO while at Hill Canyon high algal concentrations produce little diel oxygen variation. Another point (#9A – Lower Conejo Creek) also appears appreciably displaced: having lots of Chl-a but only a modest delta-DO.

As to what might be happening, my first thought about Revlon Slough was phytoplankton, but Diana also measured water-column Chl-a and at no site would including phytoplankton have meaningfully altered her overall mean values for benthic and floating algae. So it remains a puzzlement: no algae, but a large delta-DO. Peak DO concentrations are very high, 15-16 mg/L, and the peak occurs around 11 AM. Like what's happening there? But it's a slough and lots of questions remain unanswered: What kind of admixture of different waters flow at the monitoring location, and how might that mix have varied over the week to ten days the sensor was in place (unfortunately the conductivity sensor was erratic and often gave impossible readings – which might be a further indication of strange happenings)? How accurately was flow determined – and by how much did it vary over the days when the sonde was emplaced? We just don't know.

Site #10 was equally strange, but in the opposite direction: lots of algae but low delta-DO. Again, I don't know what might be happening here, but this location and the third half-weird location, #9A (which also had reduced delta-DO for the amount of measured Chl-a), share an interesting commonality – they are both located relatively short distances downstream from major WWTPs (Hill Canyon WWTP at #10, Camarillo WWTP at #9A). I'm going to take the easy way out and simply claim that these locations, dominated by waste water discharges that we know all too little about (e.g., how flow might vary over a 24 hour period, and how the timing of the flow peak varied in relation to minimum and maximum DO) are simply anomalous and can be disregarded.

To illustrate just how strange some of Diana's data are I did a multiple regression on log-transformed values and ended up with an unusual relationship: one that was not statistically valid (p-value = 0.22) but had a reasonable amount of explanatory power (r-square = 0.4); explaining 40 % of the delta-DO variation between her locations. Eliminating the two most problematic points gave me an even better relationship (p-value = 0.25, but r-value = 0.81), at least as far as explaining variations went. However, both equations were of the form

$$\text{delta-DO} = \text{constant}/[(\text{Chl-a})^x Q^y]$$

In other words, increasing flow (Q) would decrease delta-DO, which sounds good, while increasing the amount of algae (Chl-a) would also decrease delta-DO. Which not only doesn't sound so good, but is not actually possible. I don't fault Diana's competence and field-craft, on the contrary, but this does emphasize a point I've been trying to make all along – determining the mean algal density in a reach is not rocket science. Not even close.

In the table that follows I list Julie's and Diana's data showing mean values of total algal density, the standard error of the mean (SE, in mg/m^2) and 2-times the standard error of the mean (2*SE) measured as a percent of the mean value. Twice the standard error is

	mean Chl-a mg/m^2	SE mg/m^2	2*SE %
Julie's 2003 Ventura data			
Foster Park	268.3	25.0	19%
Shell Road	55.5	7.2	26%
Stainley Drain	63.2	13.1	41%
Main Street	28.0	6.7	48%
Diana's 2008 Callegues data			
9A-Howard	1293.3	477.3	74%
9B-Adolfo	984.9	347.3	71%
10-Gate	1145.2	219.6	38%
12-Park	718.8	148.9	41%
13-BeltPress	306.6	116.4	76%
Revolon Slough	125.3	33.0	53%

approximately the 95 % confidence interval (c.i.) above and below the mean. Thus, using Diana's "Howard" data (#9A) as an example, the 95 % c.i. would be 1293 mg/m^2 ± 74 % or 1293 ± 957 mg/m^2 or between Chl-a values of 336 to 2250 mg/m^2. In other words, a pretty wide range. Julie's data has a lower standard error, and Kristie's (which I haven't shown, but which was used to develop the delta-DO/Chl-a/Q relationship) is somewhere between the two: twice the standard error varying from 20 to 66 %. The relative magnitude of these standard errors, of course, says nothing about the individuals doing the work; Calleguas Creek was a much tougher proposition than the Ventura River. And a low standard error doesn't necessarily signify a more accurate determination of the mean either: luck and coincidence can play their nasty little roles, and sloppy work all too often produces much nicer statistics.

My next step was to revise the model by incorporating this new data. There are actually two versions of the original model. In the version shown on the opening page values of delta-DO, which were more or less normally distributed, were not log-transformed prior to developing the regression model. In the other version delta-DO values were log transformed. As a result of this review I've decided that models in which delta-DO values were not log-transformed gave the best fit (slightly increased R-value and increased p-values for each of the equation parameters), and that I originally erred in including the most problematical of Kristie's points: a Sept. Chl-a value of 396 mg/m^2 on lower San Antonio Creek, associated with a delta-DO and Q of 11.6 mg/L and 5.1 cfs, respectively. I'm not sure which of these three measurements is in error, but since the model which includes this point in the regression estimates probable Chl-a density at 1500 mg/m^2 while the estimate from the model excluding it is 3600 mg/m^2, one or more of these values obviously is (i.e., this point is similar to the Revlon data discussed above, relatively low Chl-a producing a huge delta-DO).

Figure 1 shows results from four models. All use Kristie's original data (both with and without the problematic point) with either log-transformed or non-transformed delta-DO values. There's not a lot of difference between them at lower Chl-a values (bottom panel), but at higher values there is a tendency for all except the O*, nTF model (bright green circles; Kristie's data minus the S.

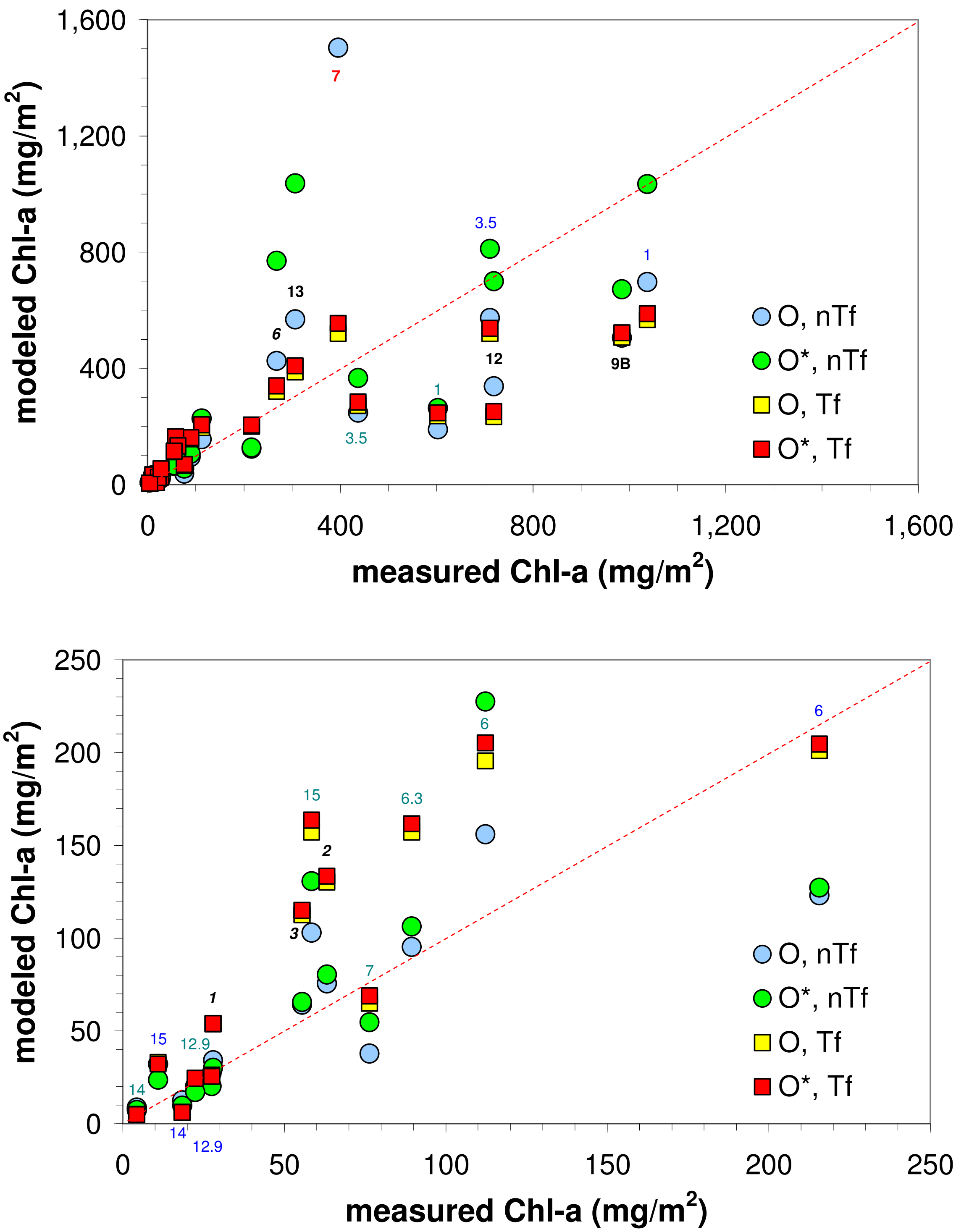

Figure 1. In the upper panel estimated values of Chl-a (y-axis) are shown plotted against measured values for 4 versions of the Chl-a model. They are coded as follows: O = Kristie's original data, O* = Kristie's original data but removing a problematic S. Antonio point, nTf means delta-DO values were not log transformed, and Tf means they were. I've identified each vertical stack of modeled results with site codes: Kristie's June data in blue, Sept. data in green, and bold black numbers for Julie's and Diana's data; the red 7 marks the San Antonio point eliminated in the O* models. The lower panel is an expanded view of Chl-a values in the range of 0-250 mg/m^2.

Antonio pt. with non-transformed delta-DO) to seriously underestimate Chl-a. This held true as I added first Julie's data and then Diana's 3 valid points to the regression model. The r-squared values for models using the non-transformed delta-DO values remained higher (~0.88 v. ~0.80) and the p-value of the weakest variable (Log Q) improved (from ~0.06 to ~0.02) making the models using non-transformed values preferable. Figure 2 shows how the O* - nTf results are modified as, first Julie's, and then Diana's points are added to the regression. The final model, which includes all the available data (except, of course, the removed S. Antonio pt. and Diana's 3 weird points) does just about as good a job as O* - nTf and is statistically better. It is the one I intend using from here on out. Or at least until I can acquire additional data. This revised model is shown in Figure 3. (And here I'll like to appeal to any masochists who might have actually read this far, to please send me any additional data – or rumors of data – that they might have access to or knowledge of.

Finally, to return to Ben's question, more data would both improve and continue to test the model. Especially data collected specifically for this purpose. By that I mean data collection that placed an equal emphasis on measuring flow and delta-DO as well as Chl-a. Typically, effort is expended on measuring only one, or at most two, of these parameters at any one time or by any one investigator. Ideally, field collection of Chl-a samples (which takes at least half a day at any one location) would coincide with emplacing a sonde to measure variations in DO at a gauging station location where the gauge is checked by at least one reasonably accurate field measurement. Lacking a gauging station location (I did say *ideally*) multiple flow measurements (with, perhaps, a sonde that included a pressure transducer to track variations in stage) would be almost as good.

At the other extreme, we would need, at least, the data used here: field collection of Chl-a samples with preservation and later analysis; manual measurement of DO at the estimated times of maximum and minimum concentration; and some determination of flow (from on-site measurement or gauging station records). Hopefully, all three activities occurring within some reasonable time frame – at least within the same week if not within 2 or 3 days of each other. Simply stating these opposites shows where the pitfalls of this low-intensity option might lie. And where improvements can be made.

The Chl-a measurement effort is the same in both cases, but, as I've shown earlier, these results, howsoever carefully done, will have a large associated error. This, paradoxically, makes this intensive collection effort compatible with relatively crude and error prone measurements of delta-DO and flow; however crude or haphazard they are not likely to be accompanied by standard errors greater than 50 %. With a little more time and effort both measurements can be significantly improved. Greater accuracy in flow measurement can be achieved by simply measuring flow more than once at the same location: moving slightly upstream or down and repeating the measurement. Averaging of multiple measurements would both improve accuracy and provide an estimate of probable error. Multiple measurements of delta-DO would also help accuracy, but more importantly, might avoid the one big pitfall of doing this by hand: What if the sun does not come out on the day of measurement?

Pre-dawn measurements are rather foolproof; rain or shine, results will be similar. However, mid-afternoon peaks are likely to be very different on days of overcast, or when coastal fog never lifts. Not to mention bias in the data if fog or overcast continues over one part of the catchment while another basks in sunshine during the time of measurement. So there needs to be some kind of fall-

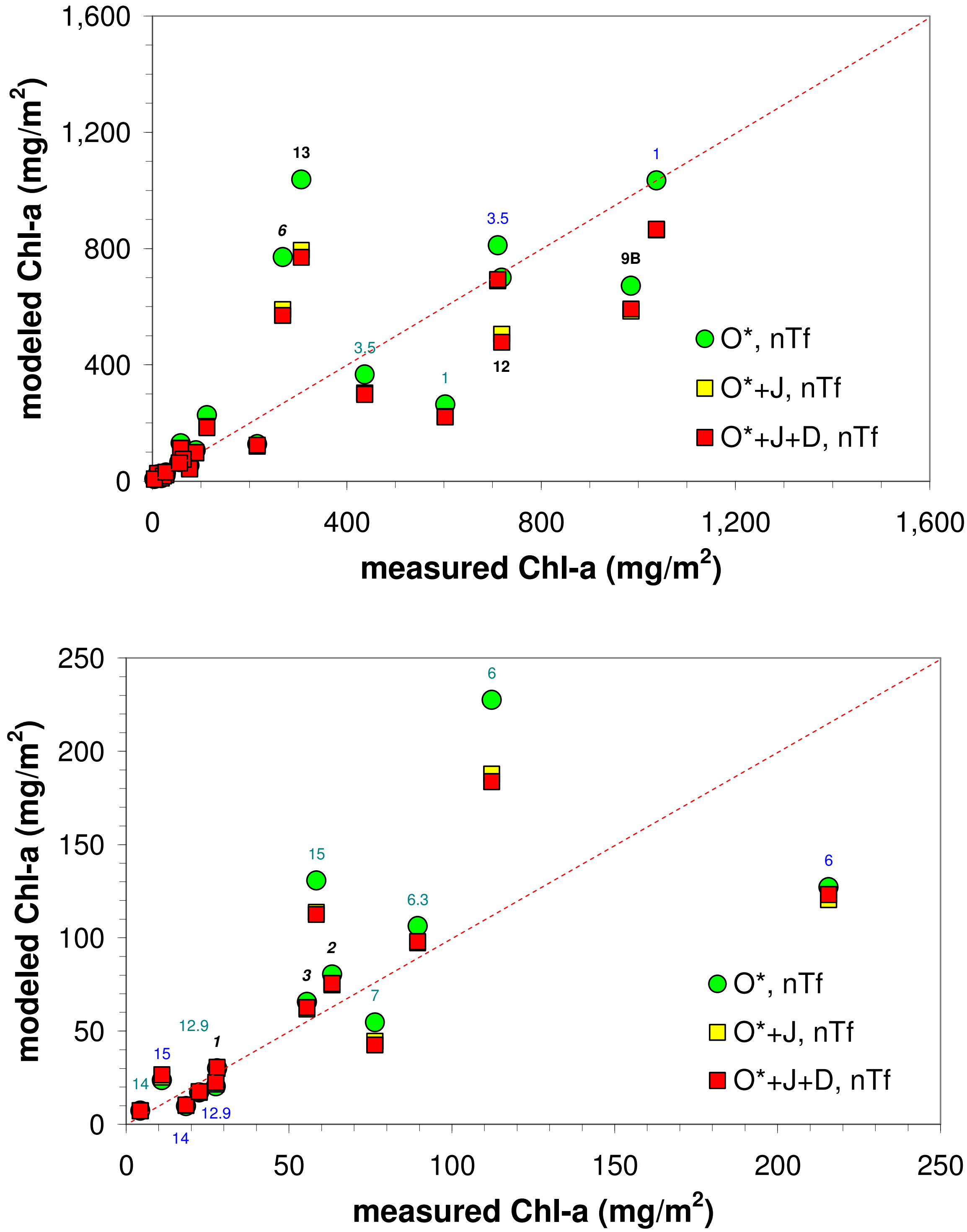

Figure 2. In the upper panel estimated values of Chl-a (y-axis) are shown plotted against measured values for 3 versions of the Chl-a model. They are coded as follows: O* = Kristie's original data without the problematic S. Antonio point, +J = Julie's data added, +D = Diana's data added, and nTf means delta-DO values were not log transformed. Each vertical stack of modeled results is numbered with the site code: Kristie's June data in blue, Sept. data in green, and bold black numbers for Julie's and Diana's data; the red 7 marks the San Antonio point eliminated in the O* models. The lower panel is an expanded view of Chl-a values in the range of 0-250 mg/m^2.

-back plan to repeat these measurements on an alternate day. This is especially important since this is the easiest and least time-consuming measurement that needs to be made. And not getting it roughly right will make the determination of flow (which takes more effort) and Chl-a (which takes the most effort and expense) a total waste of time for purposes of extending or testing this relationship.

Since Channelkeeper is already making delta-DO and flow measurements, and sondes can be purchased or rented to improve and automate this effort, a final question remains: Can Channelkeeper, or any volunteer group, make Chl-a measurements? The answer is a qualified yes. Chl-a determinations are divided into two parts: (1) collecting samples and (2) analyzing samples. Field collection is a laborious and pain-staking pain-in-the-ass, but it’s not rocket science. You don’t need to know how to identify algae, simply how to collect samples. A half-day lesson by a qualified instructor should do the trick. Hire Diana, or Kristie. They’d probably appreciate a few extra bucks. The tools required are simple to make, but practice is needed to use them effectively. It’s time-consuming, but not hard. And algae are easy to catch.

Analyzing samples is a problem of a different order. The equipment is more involved, and more expensive. You need a fluorometer, which is pretty big bucks (we’re talking a couple of thousand here). But it can be done. You need some practice to get it right, but it also isn’t rocket science – it’s more like making soup in your kitchen. There are alternatives. Having someone else do the analysis, or part of the analysis (the fluorometer business for example). Labs will do it for money, and perhaps a university will also. Or maybe a professor UCSB would make their lab available on off-hours. (A conversation with Scott Cooper might be worth someone’s time.) Or maybe Diana wants to buy a house.

Table 1. Channelkeeper (SBCK) site codes for monitored Ventura River locations.

location name	no.	location name	no.
Lagoon, east side	0e	Lion Canyon	8
Lagoon, west side	0w	Pirie Creek	9
Main Street	1	upper S.Antonio	10
Stanley Drain	2	Santa Ana Blvd.	11
Shell Road.	3	Hwy. 150	12
above C.Larga confluence	3.5	below the diversion	12.4
Canada Larga	4	above the diversion	12.4
Upper C. Larga	5	Camino Cielo	12.9
Foster Park	6	Matilija below dam	13
above OVSD	6.1	N.Fork Matilija	14
above S.Antonio confluence	6.3	Matilija above dam	15
S.Antonio at confluence	7c	middle S.Antonio	17
lower S.Antonio	7		

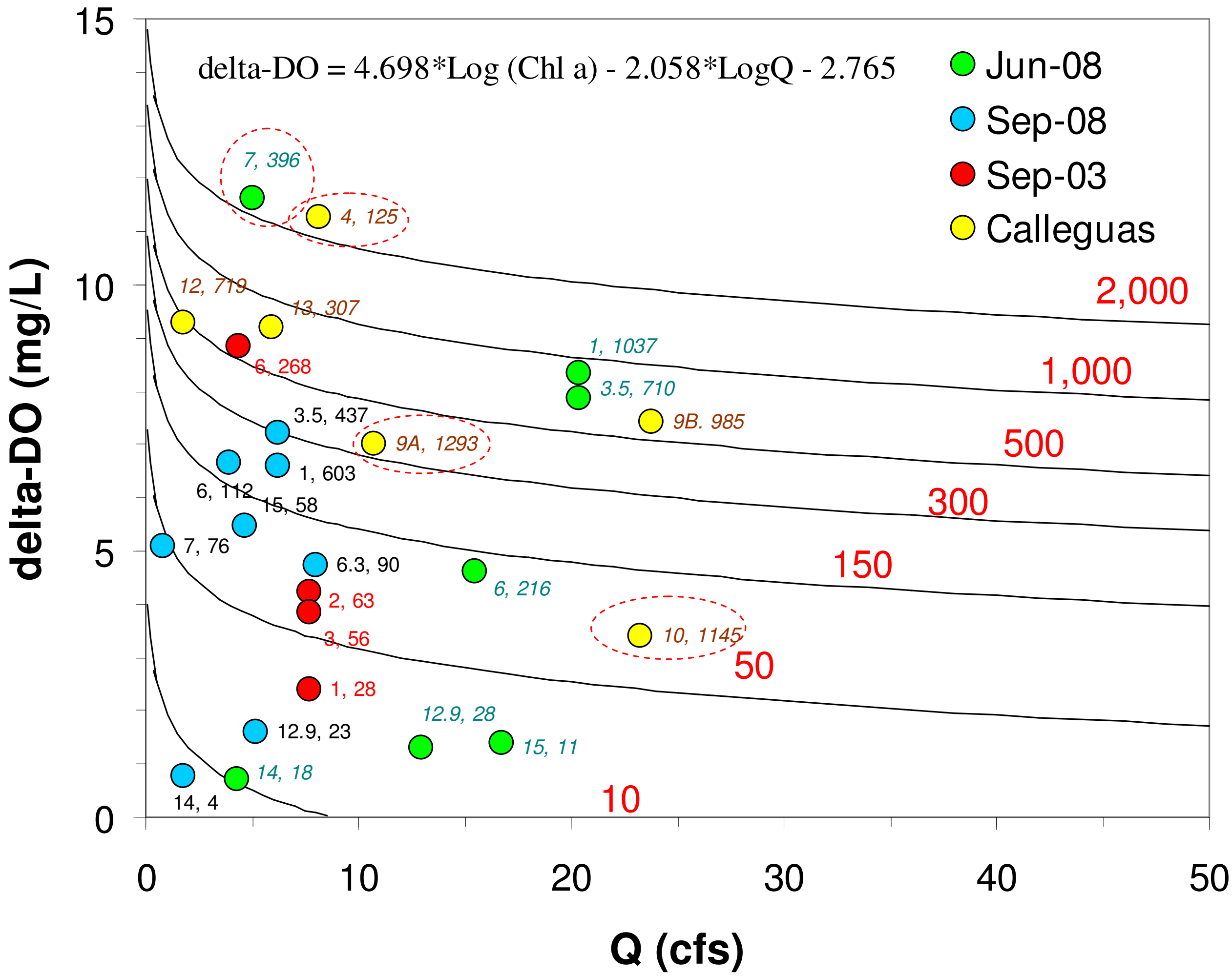

Figure 3. The above chart shows the revised graphical model developed using data from the UCSB-LTER study, Ventura data collected by Julie and myself in 2003, and 2008 Calleguas Creek data collected by Diana. The regression relationship used in the model is shown on the graph ($p < 0.001$, r-squared = 0.87). Although the regression equation shows delta-DO as the dependent variable (required to plot the graph in this fashion, with Q and delta-DO on the axis', and modeled Chl-a as lines labeled in red with density in mg/m^2) it was initially derived using Chl-a as the dependent variable. On the chart all currently available data are shown as points: the identifying numbers show either SBCK or Diana's site code, followed by measured mean Chl-a densities. By comparing actual measured Chl-a density with the lines of equal density plotted on the graph (the red numbers indicate Chl-a in mg/m^2) the accuracy of the model can be judged. Points circled by dashed lines indicate problematic data that were not included in deriving the regression model for reasons discussed in the text.