Response to Dr. Graham D. Farquhar's comments

Overall response: Comparing the re-revised manuscript with the original manuscript, we realized how much help we have received from the review process and how fortunate we have been. Thank you very much.

1. Comment: The manuscript has been improved. But it is still too long. I ask that the manuscript be no more than 40 pages.

Response: the manuscript has been shortened.

2. Comment: I ask the authors to *closely* address the second round of review comments as well as those that came earlier. Some key ones are that test data should not be error free, and should not have perfect spacing in the X space. Weaknesses and limitations as well as strengths need to be included. The interdependence of gamma* and the kinetic parameters is important, although the parameter alpha is also needed if not all glycolate is photorespired (in other words the 0.5 in the reviewer 2's comments would not necessarily be correct). I ask the second and subsequent authors to help the first author pay particular attention to the comments on writing style mentioned by Reviewer 2.

Response: We have revised the manuscript so that the 100 synthetic A/Ci curves now use random Ci spacing. Because of this revision, Figures 6 and 7 (previously Figures 5 and 6) are now completely different.

We understand that it is important to test fitting methods with synthetic data that contain introduced errors. But this current paper is not the best place to do it. It would be difficult to attribute without ambiguity the cause of failure in parameter estimation if error-containing synthetic A/Ci curves are used. The failure could be attributable to improperly introduced errors or to an inadequate fitting method. Further, how measurement noise affects parameter estimation critically depends on the number of datum points the dataset contains and the distribution of these points along the Ci axis. This itself is a complicated topic and a rigorous treatment would be required if error-containing synthetic data are used. This current paper focuses on methodological issues and does not have enough room to do justice to this topic.

In this current paper, the new approach is tested against both synthetic data and chlorophyll fluorescence measurements. The test against chlorophyll fluorescence measurements is more powerful than the test against the synthetic data, with or without the introduced error.

To conduct a test with error-containing synthetic data, we will need to consider what distributions we draw error from for A and Ci, how the errors of A and Ci are correlated, how the dataset is structured along the Ci axis, and how many points to sample. These issues if included in the present paper will not only greatly increase the length of the paper but also distract us from the main tasks of the paper. Even if we use synthetic data with error, we would still need to apply error-free synthetic data in order to achieve some key objectives of the paper. This paper covers multiple tasks – demonstrate why existing methods don't work, understand why it is so difficult to estimate FvCB parameters, and propose a new method with its rationale being explained. One cannot use synthetic data with error to demonstrate why a fitting method does not work because the failure could be caused by the error improperly introduced to the synthetic data. It is much more straightforward to understand why a fitting method does not work with error-free synthetic data. However, we are planning a separate paper on measurement issues and sampling strategies in which error-containing synthetic data will be used.

The interdependence among different parameters is important biologically. But this issue is very different from the question whether different parameters can be independently estimated from a given A/Ci dataset, which is one of the concerns of the paper. Whether a set of parameters can be independently estimated is determined by the nature of the data, the characteristic of the model, and the fitting method. If indeed two parameters are interdependent biologically, then their independently estimated values should faithfully reflect this biological interdependency. We provided more detailed explanations in our responses to relevant comments made by Reviewer No. 2.

<u>3. Comment</u>: I have some additional queries:
Page 2 line 1 "by decreasing the light level" - where is this discussed in the text, apart from on page 21, as I don't understand how one avoids the need for an estimation of Jmax and theta (see Farquhar & Wong, AJPP, eq A3)"

Response: For typical A/Ci curve measurements, the light level is kept constant at all sampling levels of Ci. J is a function of Jmax and light level. Light absortance and theta are the parameters of the function. Because this function does not contain Ci as a variable and is the only relationship in which light absorptance and theta are used, it is impossible to estimate these two parameters when the light level is kept constant. In order to estimate light absorptance and theta, multiple levels of light have to be used. In this study, we only consider typical A/Ci curve measurements.

Because for typical A/Ci curve measurements, the light level is kept constant, J could be treated as if it is a constant parameter. Once J is estimated from the A/Ci data, Jmax could be calculated outside the optimization process from the estimated J, the light level, and the prescribed values of light absorptance and theta. Of course, Jmax could be estimated directly as a parameter in the optimization process, but then prescribed values of light absorptance and theta as well as the light level would have to be used within the optimization process to calculate J.

If multiple levels of light are used, J is not constant any more and thus cannot be treated as a parameter to be optimized. In this case, Jmax has to be estimated directly as a parameter and if more than three levels of light are used, then the parameters of absortance and theta can also be estimated.

<u>4. Comment</u>: There are places where the co-authors with English as the native tongue do not appear to have read carefully. So on page 15, '*Kco* is OF the order of 60 Pa'.

‘indicates that TO a good approximation’; ‘all have *Cc* values much too small compared to *Kco*.’ instead of "much too small" I think you mean "negligible" or "very small". ‘This would occur if the sample values of *Ci* of Rubisco-limited points are too small (*Cc* is less than *Ci*).’ I don't think you mean "too small", as it has the connotation of the experimenter having made a mistake or something. Also it reads as if Cc being less than Ci (the normal case in the light) is the definition or explanation of Ci being too small.

Response: Change made. Thanks for pointing out these errors.

5. Comment: Page 21 ‘the RuBP regeneration-limited state is the only state that does not suffer from the problem of overparameterization’ - I am unclear about this. Do the authors take into account that normally one has to fit Jmax and theta, as well as effective absorptance?

Page 23 ‘One way to increasing the number of RuBP regeneration-limited points is to decreasing the light level.’ – again, then need Jmax, theta, & effective absorptance by PSII.

Page 45 where is Jmax used?

Response: It is impossible to estimate theta and effective absorptance because only a single level of light is used for typical A/Ci curves. When we say ‘the RuBP regeneration-limited state is the only state that does not suffer from the problem of overparameterization’, we refer to the expression A=Wc(1-gamma*/Cc)-Rd where J is a parameter to be estimated directly. Obviously if Jmax, theta and effective absorptance are all treated as parameters to be estimated from A/Ci curves, which is a wrong thing to do unless multiple levels of light are used in the measurement, then the RuBP regeneration-limited is overparameterized. We have revised the manuscript so that it is clearly understood that we are dealing with typical A/Ci curves for which measurements are made at a single light level and therefore we are treating J as the only limitation state-specific parameter for RuBP regeneration. The good news is that when multiple light levels are used, one can prove, in theory, that the RuBP regeneration-limited state is still not overparameterized and therefore Jmax, light absorptance and theta can all be independently estimated from A/Ci/light curves.

6. Comment: In the references, the Farquhar et al. paper, should have the 2 and the 3 subscripted.

Page 52 line 1 ‘and thus are not used in the fitting’ – should that read ‘preliminary fitting?’

Response: Changes made.

Response to comments by Reviewer #1

Overall response: This review is extremely helpful to us. As a result of responding to it, we have reorganized the whole manuscript and substantially rewritten some sections. We are so grateful for all the insightfulness and generosity this reviewer has shown in reviewing the current as well as the previous version of our manuscript.

We believe our new approach is the definitive solution to A/Ci curve fitting. However, how A/Ci measurements are made are also important. There is no magic solution if the dataset does not contain sufficient information to begin with. A fitting method can only be as good as the data.

We believe we followed the principle of objectivity and impartiality in testing and comparing different A/Ci fitting methods. We have been working on A/Ci fitting methods for years. The values of the parameters and the spacing of the Ci values used to generate the synthetic A/Ci curve in Figure 1 were from a previous unpublished effort in comparing existing methods **BEFORE** we conceived the idea for our new approach. The values of the parameters for the curves in Figure 5 and 6 were randomly generated although the same Ci spacing as in Figure 1 was used. These synthetic curves were not designed to favor one method over the others. In the revision, we also let the Ci spacing to vary randomly. This should leave no doubt that the methods are tested against the synthetic data on an equal footing.

1. **Comment (the following two paragraphs)**:

The method of estimating the parameters of the FvCB model proposed in this manuscript does indeed include significant improvements over other methods, and is based on a simple and brilliant intuition, supplemented by hard work on the logic of the model. However it is not a definitive solution, and its presentation seriously undermines the manuscript's impact.

This reviewer found that the current revision presented enough detail to make the proposed method understandable, which the first version did not. However, understanding the method is still a struggle, and once the reader understands it, it becomes clear that the explanations offered in both the first and second versions are much more complex than necessary, and hinder comprehension more than they help. The revision falls short of what was necessary. The authors did not address previous questions satisfactorily, and did not understand how extensive the rewriting needs to be. The reviewer apologizes for having failed to make this clear enough, and the rest of this review will thus address problems more directly, and without going into page-by-page detail.

Response: The previous manuscript was disorganized. With the new reorganization and revision, we hope clarity has improved.

Among other changes, we now distinguish different fitting methods based on how their cost functions are formulated. In this way, the similarity and differences among different methods can be seen clearly. It also lets us to explain, with relative ease, why certain methods lead to poorly-shaped cost functions and thus cannot guarantee successful minimization. Finally, it allows us to introduce the problem of inadmissible fitting more naturally. We discarded the term "simultaneous estimation method" as it is confusing in the context of this manuscript.

We firmly believe the new approach is a definitive solution to fitting A/Ci curves. But we are mindful that the estimation of parameters in a model is affected by three factors:

1. The quality and quantity of the data
2. The structure of the model
3. The fitting methods.

For any model parameter estimation, these three factors are intermingled to some degree. But for the FvCB parameter estimation, they are intermingled to an exceptionally high level. It is not our intention to claim that our new method can always successfully estimate the eight parameters in the FvCB model regardless the quality and quantity of the data. For example, it would not make sense to estimate parameters specific to a limitation state when there are no points in the dataset that represent this state. It is a problem of the dataset, not the fitting method. No fitting method can solve an inherent problem in the dataset.

This reviewer thinks our new approach is not a definitive solution probably because three was that one synthetic curve (out of 100) for which the new approach was not able to retrieve the parameters very accurately. It was a case where the sampled points were poorly structured and were dominated by points in the Rubisco- and TPU- limited states. These two states are structurally overparameterized. The synthetic curve had two points in the RuBP regeneration-limited state. Thus in theory our new approach should be able to retrieve the true parameters. But any parameter estimation has to face the practical issue of setting a convergence criterion in the code for iteration to stop. The criterion can only be the same for fitting the same model to different datasets. If the convergence criterion is too large, the iteration stops too early; if it is too small, the computer is stuck. So it should be 'reasonably small'. For that particular curve, the convergence criterion was too large because it had barely enough information to constrain all eight parameters. When we used a smaller convergence criterion, the estimated values of parameters became more accurate. However one cannot guarantee accurate parameter estimation by using an infinitesimally small convergence criterion. A much better way is to improve the information content of the data.

The problem with previous methods is that no matter how good the A/Ci dataset is, they cannot guarantee the successful estimation of FvCB parameters because they have failed to recognize the structural characteristics of the FvCB model and cannot use the information in the dataset correctly and to its fullest extent. Our new method solves this problem so that its only limitation is one imposed by the data, not anything else. It is in that sense that we believe our new method is a definitive solution.

2. Comment (the following three paragraphs):

The manuscript could be made considerably shorter, easier to understand, and more effective, by changing its general outline and perspective in two ways. First, after details of the model have been explained, the manuscript should show the continuity between the approach it proposes and previously published ones, and then highlight how it improves on them. Second, the manuscript should lay out the advantages and limitations of the method clearly, without overstating the former or disguising the latter.

The method proposed in this manuscript belongs to the same family of solutions as those used or studied by, in chronological order, Dreyer et al., Dubois et al., Miao et al., and Yin et al., and it appears to improve significantly on them, partly by fixing the transition points iteratively (the enumeration phase), and partly by substituting a segment-wise cost function to the conventional model-wise one. In its current form, the manuscript creates confusion by both bundling previous 'SEM' approaches together, when they are in fact quite different, and setting itself in contrast to them. It is much easier to understand that the method presented here falls under the label 'SEM', along three other ones, and that each of them has its particularities. The main difficulty in statistical estimation of the parameters of the FvCB model is that typical gas exchange data (ACi curves) do not carry enough information to permit estimation of all parameters. Yin et al., following a suggestion of Dubois et al., attempted to bring in fluorescence information; light response data could also be brought in. 'Disjunct' estimation methods, on the other hand, have relied on fixing the transition point(s), which amounts to addition of information as well. However several articles, including this manuscript, argue that 'disjunct' estimation really should be improved upon. The improvements proposed in this manuscript are based on the following central intuitions: fixing the transition point(s) does indeed add information into the estimation procedure, and could thus increase the number of estimable parameters, but the value at which a transition point is fixed does not have to be permanent. It is entirely feasible to successively try out different values for the transition point(s)-the enumeration phase-, and at each hypothetical value of the transition(s), estimate the other parameters. This may seem similar to the approach of Manter and Kerrigan, but the authors combine this with a segment-wise cost function, which apparently holds the potential to remove even further some of the overparameterization that limits the number of estimable parameters when using model-wise cost functions. In some data sets, the combination of fixed transition points and segment-wise cost function seems likely to lead to reliable estimation of more than three parameters. In addition, the authors provide some interesting arguments regarding which parameters should be fixed, when not all are estimable, and they have clearly given a lot of thought and work to the software implementation of their ideas.

Unfortunately, the manuscript greatly exaggerates the capabilities of the method, and actively obfuscates its limitations. The authors claim that up to 8 parameters can be estimated, and provide 'guidelines for informative measurements'. These guidelines are really limitations: the method breaks down if data do not follow them, and quickly falls back to estimating only 4 parameters. See table 5. The presentation would be a lot more effective if the authors did not overstate their case. The manuscript should begin by stating that this method can estimate 4 parameters, and that there are reasons to believe that up to 8 possibly might be, but only if the Ci range is sampled carefully. In its current form, many readers will be left with the impression that the manuscript is not entirely forthcoming regarding the capabilities of the method, and this will distract from the real advances that are being presented. Openly discussing strengths and weaknesses, as well as unresolved questions (how many parameters might be estimated using realistic data) would be far more effective.

Response: We have reorganized and rewritten the manuscript largely along the line suggested by reviewer. We now use the formulation of the cost function to distinguish between methods and to relate methods to each other. Confusing terms are dropped.

The ultimate criterion to evaluate any parameter estimation method is whether the method is able to use the information in the data correctly and to its fullest extent. A good fitting method should allow the number of estimable parameters and the accuracy of the estimated parameters to be limited only by the data, not anything else. We have provided both theoretical analyses and example demonstrations to show that all previous A/Ci fitting methods have various methodological problems that prevent them from using the information in A/Ci measurements correctly and to the fullest extent. These methods themselves are a limiting factor in the number of estimable parameters and the accuracy of the estimated parameters. What we are doing in this study is to remove this unnecessary methodological limiting factor so that A/Ci analysis is truly limited only by the measurements.

The reviewer is correct in pointing out that the enumeration of all possible limitation state distributions and the use of a distribution-wise cost function are two crucial components in our new method. Other components such as treatments of structural overparameterization and inadmissible fits are also important. But we need to clarify that enumerating limitation state distributions is not equivalent to fixing the transition point iteratively. For each distribution, the limitation type of each point is always fixed but the transition point is not fixed during the iterative process of minimizing the cost function and is not determined until the parameters are optimized for the distribution. In fact, it is not even absolutely necessary to calculate the transition point in our approach.

Both our new approach and those of Dreyer et al., Dubois et al., Miao et al. and Yin et al. do not assume the transition point is fixed. But the new approach contains two nested optimizations: the optimization for the limitation state distribution and the parameter optimization. There is no optimization for the limitation state distribution in the approach of Dreyer et al., Dubois et al., Miao et al. and Yin et al. The new approach takes into consideration the fact that the FvCB model is a change-point model; those previous methods regard the FvCB model as a regular model because they use what this reviewer calls “the model-wise cost function”. The “model-wise cost function” of a change-point model has intractable shapes for parameter estimation. Statisticians have known this since the 1970s. The way the cost function is formulated in our new approach does not suffer from this problem because the if-then conditions in the FvCB model is not used within the parameter optimization process. In the revised manuscript, the similarity and differences are stated explicitly.

As compared with existing methods, the only weakness we can think of our new approach is that it is more computationally demanding because all possible distributions of limitation states have to be tried before the optimal limitation state distribution is found. Thus there are many regressions to be done for a single curve fitting. But with new computer technologies, this should not be a major problem.

We firmly believe our new approach is a definitive solution to estimating FvCB parameters from A/Ci curves. Only time can tell whether we are correct or wrong with this statement. We may have been very enthusiastic about our new approach, but we made no attempt to overstate the problems of existing approaches or disguise the problems of our new approach. Thus we reject that “Unfortunately, the manuscript greatly exaggerates the capabilities of the method, and actively obfuscates its limitations” and “These guidelines are really limitations: the method breaks down if data do not follow them, and quickly falls back to estimating only 4 parameters”. What a fitting method can do cannot exceed what is in the data. For example, if the dataset contains no points in the Rubisco-limited state, then Vcmax and Kco cannot be estimated by any method. It is not a fitting method problem. It is a measurement problem. Also, if all points are in a straight portion of the curve, no hope the parameters can be estimated accurately even though the points are produced by the model because these points can be fit by any number of sets of parameters to a degree of accuracy within the convergence criterion of the code and the rounding error of the computer.

Nevertheless, we should have done a better job in explaining which parameters can be estimated in theory and which can be in practice and there is a big difference between theory and practice. Thus we have adopted the suggestions made in the third paragraph of the Comment No.2.

3. <u>Comment:</u>
Proving that a method of parameter estimation works, and that it works well, is very difficult. The proof of the pudding is in the eating, and testing how well the method can retrieve known parameters from synthetic data is more or less the only truly convincing way to obtain it. The main purpose of using synthetic data to test the performance of the procedure is to test how it handles error, or in other words, how well it can retrieve a known signal through the noise. Foregoing error misses the purpose. In real data, we cannot know with certainty what the signal was; the ‘true’ values remain forever unknown, although we can quantify how probable the estimates are. Given a believable model such as FvCB, we use regression to estimate its parameters, whose ‘true’ value is not known. This assumes that each individual data point (every A,Ci pair) in a measured ACi curve is a realization of the model, to which random error has been irretrievably added, with the error for each individual point being drawn from the same distribution, and being uncorrelated with the error on the other points. When we use synthetic data to test a regression method without including error, we can obtain a sort of theoretical absolute upper limit of the number of parameters that can be estimated, as well as a check that the mechanics of the procedure are working, but we learn nothing about how many parameters can in fact be estimated using real data, or realistic synthetic data. Using 100 error-free data sets does not provide any more information on estimable parameters than

one error-free set did. In fact, it only serves to reinforce the erroneous suggestion that the method can yield estimates for 8 parameters. This usage of error-free data is not acceptable. The method may very well be able to estimate 8 parameters from real data under some circumstances, but this ‘simulation’ tells us almost nothing about it.

Response:
We understand that it is important to test fitting methods with synthetic data that contain introduced errors. But this current paper is not the best place to do it. It would be difficult to attribute without ambiguity the cause of failure in parameter estimation if error-containing synthetic A/Ci curves are used. The failure could be attributable to improperly introduced errors or to an inadequate fitting method. Further, how measurement noise affects parameter estimation critically depends on the number of datum points the dataset contains and the distribution of these points along the Ci axis. A treatment of this dependency, which would be required if error-containing synthetic data are used, is beyond the scope of this present study and is better done in another paper.

In this current paper, the new approach is tested against both synthetic data and chlorophyll fluorescence measurements. The test against chlorophyll fluorescence measurements is more powerful than the test against the synthetic data, with or without the introduced error.

The test against error-containing synthetic data is best done in conjunction with sampling strategies. To do a good job with such a test, we will need to consider what distributions we draw error from for A and Ci, how the errors of A and Ci are correlated, how the dataset is structured along the Ci axis, and how many points to sample. These issues are not simple and if included in the present paper will not only greatly increase the length of the paper but also distract us from the main tasks of the paper. Even if we use synthetic data with error, we would still need to apply error-free synthetic data in order to achieve some key objectives of the paper. This paper covers multiple tasks – demonstrate why existing methods don’t work, understand why it is so difficult to estimate FvCB parameters, and propose a new method with its rationale being explained. One cannot use synthetic data with error to demonstrate why a fitting method does not work because the failure could be caused by the error improperly introduced to the synthetic data. It is much more straightforward to understand why a fitting method does not work with error-free synthetic data.

<u>4. Comment (the following two paragraphs)</u>: Furthermore, the manuscript demonstrates that the range and density of sampling in the X space (the range of Ci) have a great deal of influence on the number of parameters estimable. So, to get to the upper limit of that number, it is not only necessary to use error-free data, but the data must also have a carefully chosen spacing between Ci values. This is very useful information: any method or instrument has its limits, and it is important to know them before using it. Unfortunately, the ‘guidelines for informative measurements’ are an attempt to pass off these limitations as a defect in measurements, when they really are limitations of the method. Likewise, when one synthetic data set out of the hundred used in the manuscript results in poor estimates, blaming poorly structured data is exactly backwards. If the model whose parameters we try to retrieve has indeed been used to generate data sets, and some of them lead to poor estimates, the failure gives us useful information on the limits of the method, not the data. Based on the awareness of those limits, we can then tailor our measurement protocol to maximize the chance of getting estimates for more than the 4 parameters listed in table 5, but this must come after having explained the method, its strengths and limitations.

The problems with the use of synthetic data extend to comparisons with other methods. What the manuscript proposes improves on previous methods, but is also in continuity with the ones it gathers under the label ‘SEM’. However, the methods labeled ‘SEM’ are also very different from each other. Treating them as a whole invalidates the comparison. For example, Yin et al. make use of fluorescence data, while the others do not. By lumping Yin’s method with Dubois’ and Miao’s, the manuscript ends up rejecting it, when it has not even tested it. Likewise, much of Dubois et al.’ argument is that ACi curves are insufficient to support estimation of more than three parameters. Setting up a comparison where that method is made to estimate more four or five parameters, and then claim that it does not work, is disingenuous. Setting up a comparison using data that are tailored to the limitations of this manuscript’s method is no less disingenuous: synthetic data can be created that would allow Yin’s, Dubois’, Miao’s, and Sharkey’s methods to yield good estimates of four, five, possibly more parameters. The graphs of the cost function are more convincing, in this sense that they can give some indication that those methods simply cannot estimate more than a certain number of parameters, while this one might. They are used disingenuously, however, when it is suggested that other methods simply do not work. The cost functions for those methods would look quite acceptable if fewer parameters were being estimated. A much better comparison would show how many parameters it takes before each method breaks down. This method would still be clearly better than the others, but the comparison would be easier to understand, and more candid.

Response: Because the Rubisco and TPU-limited states are overparameterized, there are situations in which the set of parameters that produces the best fit is not unique. This is a theoretical fact and no fitting method can avoid it. Table 1 in the manuscript documents these situations. If a given A/Ci curve contains all three biochemical limitation states and if the RuBP regeneration limited state contains at least two points, then in theory all eight parameters could be estimated by our method, assuming there are sufficient numbers of points in the Rubisco- and TPU-limited points. In the one case that our approach was not able to estimate the parameters very accurately, there were two RuBP regeneration-limited points in the synthetic curve, the theoretical minimum required for accurately estimating all 8 parameters. This theoretical minimum is still very close to overparameterization. This situation is analogues to solving a system of linear equations whose coefficient matrix is very close to singular. A computer cannot hold an infinitive number of digits and any iterative optimization program has to be given some convergence criterion so that it knows when enough is enough and to stop. With the new approach, we could produce more accurate estimation for this particular case by simply using a more strict convergence criterion (but with previous methods, a more strict convergence criterion would still not help because of the problems in the logic of their optimization procedures). But this still does not change the fact that accurate parameter estimation ultimately is constrained by the information in the dataset.

As parameter estimation is affected by the quality and quantity of the data, the structure of the model, and the fitting method, we need to distinguish between two situations in which parameter estimation fails. First, by analyzing the structure of the FvCB model, we know for certain patterns of the A/Ci curves, the FvCB model is overparameterized regardless of the fitting method used. Second, we know for certain fitting methods, the FvCB model is overparameterized regardless of the patterns of the dataset. For the first situation, there is nothing we can do except to improve the measurements. For the second situation, we have to improve the fitting method and that is why we have introduced the new method. The guidelines for informative A/Ci measurements address the first situation and thus are not an attempt to pass off the limitations as a defect in measurements.

Ineffective A/Ci datasets can be produced both in the real world and in the computer world (synthetic data). For an actual A/Ci curve, we don't know a priori whether the first situation would arise because we don't know the actual limitation states of individual data points. But we can have some idea after the fitting. If two sets of parameters fit the dataset equally well, then we know overparameterization due to measurement limitation (the first situation) occurred for the set with more parameters and we should trust only the values of the parameters in the set with less parameters. This is the strategy we applied in the code.

We have made no attempt to design synthetic A/Ci curves to favor one method over the other so we don't understand why words such as 'disingenuous' were used in the comment. In the original 100 curves, the Ci spacing is the same for every curve but the actual values of parameters were randomly selected. So, that one case whose parameters were not accurately estimated was an isolated random case. Following the suggestion from the editor, we introduced a degree of randomness into the spacing for Ci values. A complete random Ci spacing is not needed as it does not make sense to have all points in a Ci range from, for example, 1 to 5 Pa, which could happen if a complete random scheme is used. For all 100 new synthetic curves, our approach worked well. But that does not mean deficient synthetic curves cannot be produced with randomly spaced Ci values or our new approach can guarantee success even for deficient A/Ci curves.

We grouped Yin's, Dubois' and Miao's methods together because they all use the formulation of model-wise cost function and thus will suffer the same types of problems associated with the model-wise cost function. These problems will occur even when fluorescence data or variable light data are used together with A/Ci measurements.

Many factors can affect the shape of a cost function in many different ways. The number of parameters affects the relative "flatness" of the cost function but do not change a cost function from bowl-shaped to monotonic or from continuous to discontinuous. It is the monotonic and discontinuous shapes of the cost function, not the number of estimable parameters, that are the core problem for the model-wise cost function. For methods that use the formulation of model-wise cost function, the key question is not how many parameters can be estimated but whether any parameters can be estimated with confidence at all.

Existing A/Ci curve fitting methods may work sometimes. But they **cannot guarantee** successful parameter estimation even for A/Ci curves that have no inherent deficiencies. Our new approach will fail when the measurements are not adequately made and thus contain insufficient information to constrain all parameters, but it **can guarantee** success for A/Ci curves that have no inherent deficiencies. No amount of testing can prove or disprove a method and ultimately what decides is logic. Existing methods do not take into consideration the structural characteristics of the FvCB model. They have problems in logic. That is the fundamental reason why we must reject them. Our new method is logically consistent, that is why we have confidence in it.

5. Comment: In the absence of a definitive demonstration of this method superiority, which, again, is a difficult challenge for any method, the illustrations of cost functions in Fig. 7 are the most convincing argument the manuscript offers. They are quite startling; were they really generated while none of the parameters were fixed? They would indicate that the method has at least the potential to yield very reliable estimates of more parameters than others. Of course, where one chooses to 'slice' the domain of the function with respect to other parameters can produce very different patterns, so three dimensional plots would be much more informative. They should be substituted to the current two-dimensional ones. Again, it should be made clear that the graphs do not indicate how many parameters can actually be estimated using real data.

In addition, the claim that the adoption of a piecewise cost function is responsible for the better behavior of the minimization is at least partially incorrect. The simple fixing of the transition point(s) in itself eliminates some flat areas, and increase the curvature of the cost function with respect to one or more dimensions. Can the authors show how the model-wise error mean square compare with the piecewise error mean square, within their two-phase estimation approach?

Response: It is really not a surprise why the model-wise cost function and segment-wise cost function differ so much in their shapes. This has nothing to do with how many parameters to be estimated. To calculate the model-wise cost function, you have to apply the **IF-THEN** condition in the FvCB model. Anytime you use the **IF-THEN** condition to do any mathematical calculation, you introduce discontinuity. That is why the model-wise cost function is not smooth to any order of its derivatives. In contrast, no if-then condition is used in the calculation of the segment-wise cost function so the value of the cost function changes smoothly with its parameters, to any order of derivatives. The absolute flatness in some portions of the model-wise cost function is caused by the fact that the FvCB model selects the minimum carboxylation rate to calculate the assimilation rate so that limitation state-specific parameters always have no influences in some portions of the cost function. 3-D shapes are nice but even 3-D shapes can be shown for only two parameters so they probably don't help much.

The new method does not really fix the transition points. What is fixed is the limitation state of every data point during the parameter estimation phase. We have experimented with a different approach that uses transition points as parameters to replace one or two parameters in the FvCB model but the results are not as reliable as the fixed limitation state approach. We have also experimented with the idea of enumeration for the model-wise cost function. It helps but cannot solve the fundamental problems inherent with the model-wise cost function.

6. Comment: The authors are also partly mistaken as to why it is beneficial to sample in the portion of the response with the most curvature. The reason is likely that by sampling more closely in the transition regions, one simply increases the likelihood that the transition points will be fixed near the best-fit value. See also number 6. In addition, the idea that regions of less curvature are not as important in general in obtaining good estimates is incorrect. Different regions of the curve give information on different parameters. A simple heuristic way to illustrate this is to see the effect of varying the sampling density in various parts of the curve on the confidence intervals of various parameters, using hyperbolic functions similar to the sub-functions of FvCB, when. See Gallant, or Ratkowsky's Handbook of Nonlinear Regression Models. Also, the usage of the terms 'nonlinear' and 'nonlinearity' on p22 is very confusing. No part of the curve is more or less nonlinear, in the sense of 'parameter nonlinearity', which is the relevant one here.

Response: In the revised manuscript, the terms 'straight' and 'curved' are used in the context discussed here to avoid confusion. We agree that by sampling more closely in the transition regions, the transition points will more likely be fixed near the best-fit value. We actually made similar arguments in the initial version of the manuscript. But the Reviewer No. 2 strongly disagreed with the idea so we removed it. In this version of the manuscript, we put this idea back in with the concern of the Reviewer No. 2 noted.

We are not suggesting that the more straight portions of the A/Ci curve should not be sampled. In fact, we specifically emphasized that the overall range should be covered. What we are suggesting is this: if we can only take a very limited number of samples, the curved portion should be covered relatively more densely. We eliminated the general discussion on the relative importance of the points in straight vs. curved portions of a nonlinear curve.

7. Comment: The manuscript ignores an obvious reason for the difficulties other 'SEM' methods encounter: the TPU-limited phase may start in a lower range of Ci than assumed, such that only a few data points, located in the range of greatest curvature, correspond to the RuBP-limited phase. In other words, they fit a two-phase model to data that were generated by a three-phase model. This also contributes to the importance of sampling the region of greatest curvature. Apologies if this is what the authors meant to say.

Response: Yes this is another reason why other 'SEM' methods don't work as well. But this problem is not as insolvable as the problems associated with the model-wise cost function. For example, they could borrow the idea from our approach by fitting the seven scenarios in Table 1 separately and then choose the one with the best fit. But this cannot solve the shape problems of the cost function, which can make the fitting fail even when they fit the right model to the right data.

In testing against the synthetic A/Ci curves, we fit the right (three-phase) model to the right (three-phase) data so this is not an issue for the comparison.

8. Comment: In conclusion, the method in this manuscript is very likely superior to previous ones. It is regrettable that in their enthusiasm, the authors end up presenting misleading arguments. A simple reordering of parts, and reframing of the arguments, would result in a much more effective presentation. It would also become clearer what details of this revision are truly needed for comprehension, and which can be left out. Forty clearly written pages would be entirely sufficient. The use of synthetic data, if they are going to be error-free, should be clearly defined as a way to gage the putative maximum number of estimates that might be possible, not the number that actually is possible, and one data set is more than enough for that purpose.

Response: Although we don't think we presented any misleading arguments in the previous versions of this manuscript, we regret they were not as clear as they should be. With the latest rewriting and reorganization, we hope we have done a better job. Using more than one error-free synthetic curve helps the discussion because if it is just one curve, one can always say the good estimation is due to a good initial guess.

9. Comment: Please note that specific comments from the first review remain valid. Here are a few additional ones.

Response: We have taken the new comments as well as the initial comments into consideration in the revision.

10. Comment: Discussion of the J/Vcmax ratio: the ratio is not fixed. Its range is constrained, and this is a characteristic of the model, not of the estimation method.

Response: It is entirely possible that the Jmax/Vcmax ratio is constrained for natural leaves. But a reasonable estimation method should let the data tell the story. As explained in the manuscript, Vcmax, Jmax and TPU have to be correlated with each other if all three limitation states occur in the same curve. But the FvCB model does not say all three limitation states have to occur in the same curve. In fact, it is very easy to use the model to generate curves with only one limitation state (that is, over the whole range of Ci, the assimilation rate is limited by one biochemical state). Therefore that the Jmax/Vcmax is constrained is not a characteristic of the model.

In another manuscript under preparation, we show that by fixing the transition (inflection) point between limitation states, a linear relationship is created between Jmax and Vcmax when there is none in the artificial A/Ci curves (ResponseFigure.1). The slope increases when the inflection point is fixed at higher Ci values. For a given A/Ci curve, the ratio of estimated Jmax/Vcmax also depends on the distribution of Ci values sampled. That is why there is some scattering in the estimated Jmax/Vcmax in each of the plots representing different choices of the inflection point (a, b and c) in ResponseFigure.1.

11. Comment: Discussion of overparameterization: move demonstrations to appendix.

Response: Suggestion adopted.

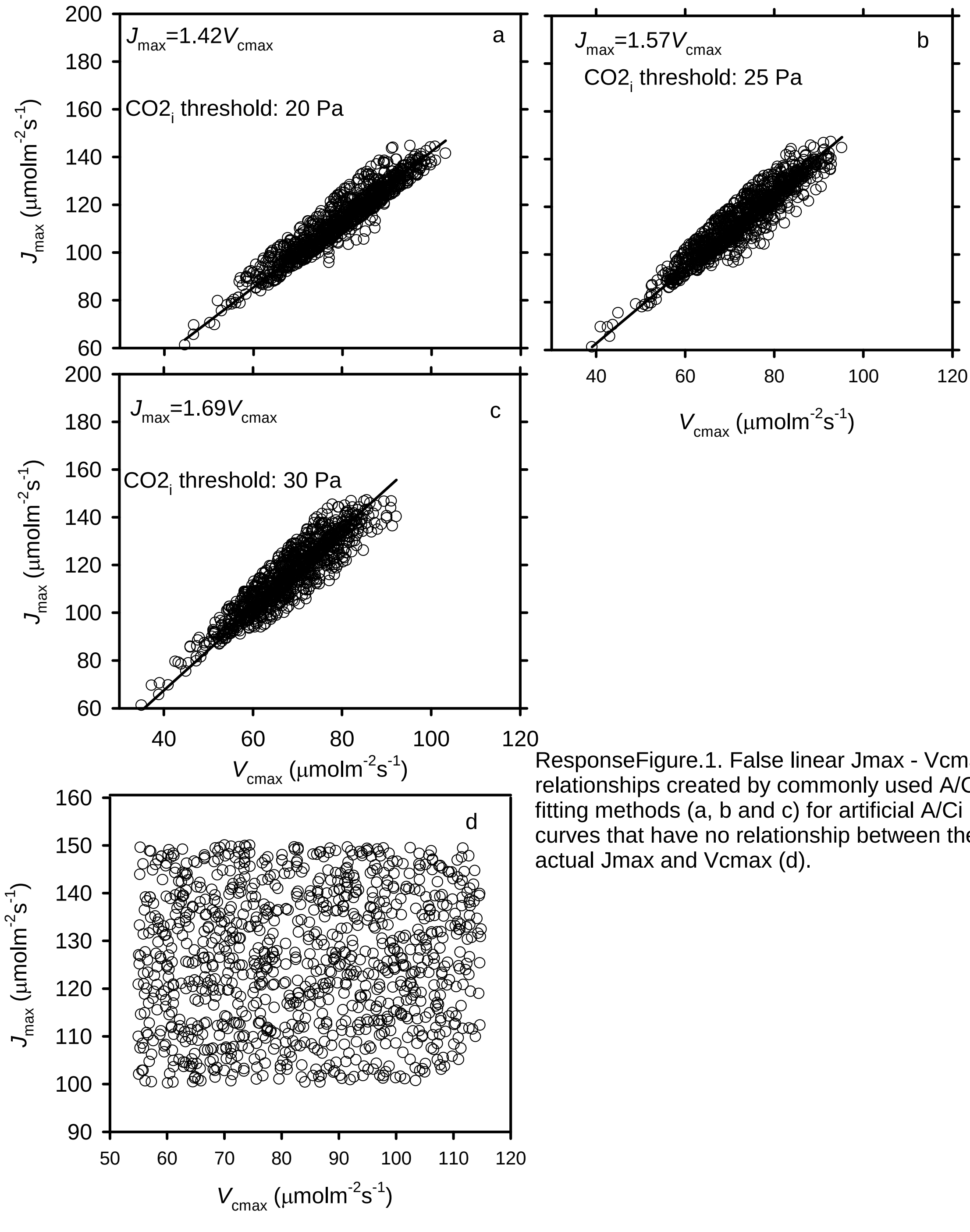
12. Comment: This demonstration overlaps with earlier discussion, see p. 8. The range of the J/Vcmax and J/Tpu, and Vcmax/Tpu ratios are indeed constrained. These are characteristics of the model.

Response: See the Response to Comment No. 10.

13. Comment: Much of the originality of this work, and some of the potential strength of this method, resides in the use of a piecewise cost function. Can you provide any reference on piecewise cost functions? This reviewer is very unsure of the consequences of this choice with respect to inference. The authors declined to address the matter of inference in their response, on practical grounds. This reviewer is aware that confidence intervals and hypothesis testing has received essentially no attention in the context of estimating FvCB, yet the prospect of outputting estimates without any idea of probability is unsettling.

Response: The use of enumeration and segment-wise (or piecewise) cost function makes the minimization of the cost function more reliable. But it does not fundamentally change the fact that parameter estimation is done by minimizing the mismatch between the data and model. In the statistics literature, we have seen any change-point model with a structure similar to that of the FvCB model. It is not just that different portions of the curve are fit with different submodels. It is also that in any portion that is fit with a submodel, the other two submodels must produce larger results in that portion. Also see the Response to Comment No. 5.

The inference issue could be addressed with the Monte Carlo simulation if we have reliable estimations on the real errors in A and Ci and their correlations. We agree that this is an important issue and the community needs to pay attention to it.



ResponseFigure.1. False linear Jmax - Vcmax relationships created by commonly used A/Ci fitting methods (a, b and c) for artificial A/Ci curves that have no relationship between the actual Jmax and Vcmax (d).

This figure is a response to relevant comments made by both reviewers.

Response to comments by Reviewer No. 2

Overall response: We have benefitted tremendously from these new comments, just like what we had from the initial comments. Thank you very much.

1. Comment: Writing style: While much improved over the last draft, I do find that the excessive use of parentheses to be distracting. If it is important to include in the manuscript, then please integrate the comments into the text. These 'side notes' become quite distracting. The manuscript reads more like a lecture, and somewhat disorganized at that. There are too many cases of the authors stating that a topic will be further discussed later. The authors also adopted a writing style that is overly familiar with numerous statements that start with "One might question…" or "Note that…" or "now let us…". I don't think this is necessary. Finally, I don't see a major reason to why this manuscript can't adopt a format that is more appropriate for PC&E, such that there is an Introduction, Methods, Results and Discussion. Obviously, an additional section on model development would be appropriate.

Response: The handling editor and the other reviewer also raised issues with our writing style. So clearly we had a problem. We realized that some paragraphs read more like free expressions of our thoughts than texts for others to read. In addition, the manuscript was indeed disorganized. We have substantially revised the manuscript and hopefully the new manuscript is better.

2. Comment: Manuscript Content: The authors seem quite distanced from previous A/Ci analyses and tend to overly criticize previous curve fitting regimes. The examples of "bad" curve fitting approaches (e.g., Fig 1 and Fig 3) would never make it out of my lab. As outlined below, that some of the problems they identify are over-stated.

Response: With all due respect, we disagree with the assessment that we overstate some problems with previous A/Ci fitting methods. ***In fact, we believe problems with previous A/Ci curve fitting methods may have been underestimated by the leaf gas exchange study community. As part of another manuscript under preparation, we show that previous methods are capable of creating tight but wrong relationships when nothing exists (ResponseFigure.1)***. Previous studies based on A/Ci curve analyses will have to be reexamined.

A/Ci curve fitting is more than finding a set of parameters that fits a curve as tightly as possible. Consistency is at least as important as the minimum of the sum of squared errors. "Consistency" has three meanings here:

- Being consistent with the information contained in the available dataset,
- Being consistent with the FvCB model structure,
- Being consistent with optimization principles.

A typical A/Ci curve contains less than 15 points while the FvCB model contains at least eight parameters. As a result, one can always fit a curve well regardless whether or not the obtained parameters are biologically meaningful. Complicating the problem is that the

FvCB model is not a typical nonlinear model for parameter estimation, it is a change-point model whose estimation is much more challenging than a regular nonlinear model. Thus unless one pays close attention to the issue of consistency, one has no confidence on the parameters obtained. No previous methods meet all three aspects of the consistency requirement identified above.

3. Comment: This manuscript still doesn't address the issue of heterogeneity across the leaf. I think the authors made a good argument in the comments to my initial review, but somehow this didn't make it into the paper. While under most circumstances this might not be a big issue, there are circumstances when it will be. In fact, this leads me to the biggest problem with this type of analysis.

Response: Thanks for reminding us this issue. The relevant discussion in our response to this reviewer's initial comments is now incorporated into the revised manuscript.

4. Comment: With a "black box" approach to curve fitting there is a large risk of the research community failing to identify the limitations of their data. While the authors might argue that their curve fitting model isn't a "black box", I use the analogy of current gas exchange systems. While all manufacturers provide the details and equations associated with their gas exchange systems, researchers may neither fully understand the limitations of the data being provided nor understand that they are making incorrect measurements. This doesn't mean that a fully objective statistical approach isn't needed – I think the approach provided by the authors is excellent. But I do think that an improved user interface is needed and that the authors shouldn't identify their method as perfect but should be used as a tool with proper verification based on the known biology of the system. Proper visual verification is still needed.

Response: We agree with this reviewer on this comment with no reservation. Even a perfect tool can be misused. Our fitting method can only be as good as the data provided. We have carefully investigated every possible scenario we can think of. That is part of the reason why the manuscript has so many parentheses and side notes. We are currently testing a visualization tool on leafweb and we expect it to be publicly available soon.

5a. Comment: Finally, the authors completely neglect the fact that some of the parameters, namely those that are COMPLETELY determined by the kinetic properies of Rubisco are treated as independent parameters. Kc, Ko and Γ* are all interrelated, and I discuss this in more detail below.

5b. Comment: Pages 12-13: It seems that in their description of the curve fitting model, the authors neglect the fact that Γ* is dependant on Kc, Ko and Vcmax. Thus, fixing one a priori and allowing the model to calculate the other neglects the fact that:

$$\Gamma^* = 0.5 \cdot O \cdot Kc \cdot Vomax / Vcmax \cdot Ko$$

Perhaps the authors should explain how their curve fitting resolves this issue. It is commonly accepted and has been shown that under a wide range of conditions the ratio of Vcmax to Vomax will be constant. But if the term Kco is allowed to change while Γ* is kept constant, then this will create a condition that completely disagrees with what we know of Rubisco kinetics.

Response: Treating two parameters as mathematically independent in the optimization process does not necessarily mean they have to be biologically independent. Conversely, two biologically independent parameters might not be able to be treated as mathematically independent in the optimization process. Whether two parameters can be treated as mathematically independent parameters is determined by how they appear in the model and by the characteristics of the dataset available. Table 1 in the manuscript reflects this point. For example, if the dataset contains Rubisco-limited points only, then Vcmax and Rd cannot be treated as independent parameters for estimation, regardless how these two parameters are related to each other biologically. Only the expression Vcmax-Rd can be treated as an independent parameter for estimation. But if the dataset contains RuBP regeneration limited points, then Vcmax and Rd can be treated as independent parameters for the estimation purpose, regardless of their biological relationship.

If two parameters are mathematically independent and thus estimable from a given A/Ci dataset, it would be better not to make any assumption about their biological relationship unless such relationship can be specified accurately and explicitly. In fact, if there is a biological relationship, then we should be able to retrieve this relationship from an adequately measured A/Ci dataset. Let the data tell the story.

Of course, if the biochemical relationship can be specified accurately and explicitly, then this is no point to estimate both parameters, just estimate one and calculate the other from the known relationship.

For the dependence of Γ* on Kc, Ko and Vcmax, ideally, the dataset should contain enough information so that both Γ* and Kco can be estimated independently. The estimated values of Γ*, Kco, and Vcmax should obey the relationship $\Gamma^* = 0.5 \cdot O \cdot Kc \cdot Vomax / Vcmax \cdot Ko$, assuming somehow Kc and Ko could be solved from the estimated Kco with additional information, the value of Vomax is available, and the kind of leap of faith can be made as discussed in Woodrow and Berry, Annu. Rev. Plant Phys. Mol. Biol. 39: 533-394 (1988).

If the dataset does not allow both Γ* and Kco to be estimated, then at least the value of one of the two parameters has to be provided from somewhere else and the error in the provided value will unavoidably introduce errors to other parameters estimated from the data. We show that it is better to estimate Kco while the value of Γ* is provided. If the provided value of Γ* is fairly accurate and the dataset contains enough information for Kco to be estimated accurately, then the relationship discussed above should still hold for the estimated parameters.

The dataset may be such that the values of both Kco and Γ* have to be provided *a priori*. For the application in leafweb, we actually conduct the fitting with two scenarios: one in which both Kco and Γ* are provided *a priori* and the other in which both Kco and Γ* are estimated. One will have to examine whether the estimated values of the parameters make biological sense in order to determine which scenario is realistic. There is no magic solution. As we stated earlier, our approach can only be as good as the data upon which it acts.

6. Comment: Page 5: I agree with the text in the bottom two paragraphs. However, given the length of the paper, I question whether such an explanation is necessary given the difficulty in obtaining gas exchange measurements at these concentrations of Cc.

Response: We have shortened the two paragraphs somewhat. It may be true that it is difficult to obtain measurements at very low Cc concentrations. But as the optimization process searches the parameter space for the optimal parameters, these low Cc concentrations may be reached at certain stages. When this happens, troubles can be created if an incorrect form of the model is not used. When optimizing parameters in a model like FvCB, if something could happen in theory, it will happen in reality. This is a hard lesson we have learned.

7. Comment: Page 7: I feel that the section header is adequate to describe the content of the section. The header paragraph is redundant and unnecessary.

Response: Change made.

8. Comment: Page 8: The potential for fixing the inflection Ci does exist and it might be a problem among many researchers. But under most circumstances it does not force Jmax to Vcmax. By arbitrarily identifying a point where the inflection between Vc and Vj occurs, the curve fitting program is constraining data points to one limiting process vs. another. This is because the final analysis is based on the data points that occur at lower or higher Ci relative to that inflection point and that there may or may not be a data point exactly where that arbitrary inflection point exists. The "true" (used loosely here) inflection point can be calculated from the results of the curve fitting from all of the points. This is still a problem, but at worst it would fix the ratio of Jmax to Vcmax as a constant. More common would be a situation where the ratio of Jmax to Vcmax would be constrained within a certain range. To support my point, numerous papers showing A/Ci analyses have been published using a variety of curve fitting techniques. But reported values of Vcmax/Jmax show that even within a treatment there is quite a bit of variability in this ratio. If, strictly speaking, previous A/Ci analyses were "arbitrarily forcing Jmax to be proportional to Vcmax and their ratio to be a constant", then the Jmax/Vcmax should be constant in these previous papers.
So, while I completely agree with the authors that the arbitrary and/or subjective establishment of the inflection point is a problem, I do not think it is as simple as they state it. This is important to resolve since it is one of their fundamental arguments that previous methods are inadequate.

Response: In another manuscript under preparation, we show that by fixing the transition (inflection) point between limitation states, a linear relationship is created between Jmax and Vcmax when there is none in the artificial A/Ci curves (ResponseFigure.1). The slope increases when the inflection point is fixed at higer Ci values. For a given A/Ci curve, the ratio of estimated Jmax/Vcmax also depends on the distribution of Ci values sampled. That is why there is some scattering in the estimated Jmax/Vcmax in each of the plots representing different choices of the inflection point (a, b and c) in ResponseFigure.1.

In previous studies, the values of Jmax/Vcmax estimated from A/Ci curve analyses, when corrected to a common reference temperature of 25^{o}C, vary beteween 1 to 3. Three factors could be responsible to this variation:

1. There are true variations in the ratio Jmax/Vcmax among species and leaves.
2. Different researchers tend to fix the transition (inflection) points at different Ci values.
3. Different A/Ci curves tend to be sampled at different values of Ci.

An objective of our new approach is to eliminate the effects of the last two factors.

It is true “there may or may not be a data point exactly where that arbitrary inflection point exists”. More relevant is how many points in the A/Ci curve are assigned a wrong limitation type if an inaccurate inflection point is used. There may be situations in which an inaccurate inflection point is used but the limitation types of all datum points are assigned correctly. For these situations, the parameters may be estimated correctly if the problem of overparameterization in the FvCB model is taken care of. More likely to happen when an inaccurate inflection point is used is one or more sample points will be assigned a wrong limitation type. Because the number of sample points in a typical A/Ci curve is not overwhelmingly larger than the number of unknown parameters in the FvCB model, the values of parameters estimated can be far away from the true values even if only one point is assigned a wrong limitation type. Thus the key in A/Ci curve fitting is to determine the limitation types of all points in the dataset as accurately as possible. The most important contribution of this study is that we found a way to do just that.

9. Comment: Page 19, I do notice that in a few locations the authors finish a paragraph or a line of reasoning by stating that an idea will be further developed later. This works much to the detriment of the flow of the manuscript. If it is important to be discussed, then discuss it completely.

Response: Change made.

10. Comment: Page 20, The authors give an excellent example of how the Sharley method of curve fitting can induce error. But as a scientist, I would never have allowed this fit to make it to a finalized dataset. It is clear from the data that something odd was occurring and I would ensure that this issue was resolved before moving on to the next curve. Perhaps all researchers may not have the extensive curve fitting experience that I have, but I would give more credit to the curve fitting community. Just because something has the potential to cause a problem doesn't necessarily mean that it will always be problematic.

Response: This comment is about the issue of inadmissible fits discussed in the paper. We came to know this issue by chance and at first we thought it was rare. Later by reading statistical literatures, we realized this issue is common for change-point models and has been recognized by the statistics community for a long time. So we conducted more tests and carefully analyzed the fittings. We found that inadmissible fits do occur fairly frequently in A/Ci curve fitting, often in a way that is not as obvious as the one shown in the manuscript. Yes, it might not always happen. But when it happens, it causes a consistency problem and the estimated parameters will not make sense. So the community needs to be aware of this problem. The new approach provides a solution to it.

11. Comment: Page 33, seldom in a manuscript are the differing roles of the authors clearly stated. Is it necessary to state who is the senior author here?

Response: Change made. These words actually were not in the original manuscript. We added them because the other reviewer made a required-to-response comment about the details of the optimization code.

12. Comment: Page 38, I have two main problems with the verification against actual data. First is that if the authors aren't going to follow their own recommendations with the data analysis, then why include it in the discussion. Shouldn't the authors compare their 'idealized' data collection compared with a non-idealized? Second, in the two paragraphs immediately prior to the conclusions, the authors delve into the methods.

Response: We were using measurements we had in hand for the analysis and they are not in an idealized pattern for estimating all eight parameters. We reorganized the manuscript so that the discussion on informative A/Ci measurements occurs near the end of the manuscript. We also revised the texts in relevant sections. These changes should resolve both issues discussed in this comment.