Geoscientific

Model Development
Discussions

GMDD

Interactive

comment

Interactive comment on "Reverse engineering model structures for soil and ecosystem respiration: the potential of gene expression programming" *by* Iulia Ilie et al.

Iulia Ilie et al.

ilie@bgc-jena.mpg.de

Received and published: 1 March 2017

Response to Reviewer 2

In the following, we denote comments by the reviewer in **bold** and our own reponses in standard fonts.

Review of "Reverse engineering model structures for soil and ecosystem respiration: the potential of gene expression programming"

We would like to thank the reviewer for the evaluation and detailed comments on our manuscript. We further provide responses for the posed questions and details on how we intend to revise the manuscript.

In this manuscript Ilie et al. explore the use of "gene expression programming" (GEP) to select empirical models for soil and ecosystem respiration. The authors make a case that GEP is a technique for reverse engineering model structures by elucidating underlying mechanisms, rather than depending on hypothesis-driven experiments to identify these mechanisms.

- Indeed, this is our main motivation. But clearly also other methods for reverse engineering may be usable.

I have several concerns about the conceptual framework the authors used to present GEP. I am convinced that GEP is an interesting and worthwhile approach to automate model selection. However, I think it is over-reaching to suggest that GEP can 'reverse engineer' model development. It seems to me that the value of GEP is simply to automate the process of exploring a large number of regression models. I am not convinced that GEP reorganizes the model development process, because regression already is often the first step in model development.

- Thank you for challenging our fundamental ideas. The motivation of this work was indeed to automatize model development. And we believe that a GEP type of approach can help in such an endeavour. But we also agree that GEP is basically doing a selection after rejecting a large number of potential regression models. And this is still very different from classical model building. The choice of the regression model structure is not made directly by the analyst and rather by the algorithm. The analyst comes into play for deciding if and which solution proposed by GEP should be further used. The points discussed here were added to the revised manuscript. p3 l18-20.

Further, I find that the claim that GEP minimizes human influence and perception bias to be strong, as the authors seemingly arbitrarily select the driving

variables for the model, regardless of how the model's functional form is derived. From other work we know that selecting a single soil temperature at 5 cm soil depth can give a very different model from selecting a temperature from 15 cm soil depth (Graf et al., Biogeosciences doi:10.5194/bg-5-1175-2008). Similarly selecting to use VWC rather than a parameter like matric potential could be the difference between being able to predict rapid increases in flux with rainfall and not.

- We have provided an initial series of candidate predictors among and GEP automatically does a feature selection. Hence the model development remains a more objective approach. Moreover, GEP is meant to select not only the driver but also the model. Therefore, GEP should be able to deal with cases as the one suggested by the reviewer: different T_{soil} measurement depth can lead to different models. And this was clearly illustrated in the analysis with artificial data.

In the end, the functions selected by GEP suffer from the same problems as previously used formulae shown in Table 2. All of these functions tend to underestimate large fluxes ("hot spots" and "hot moments"). While the form of the functions may hold-up from training datasets to prediction datasets, the specific parameterizations often do not. I believe the authors have done a good job discussing limitations of GEP, and empirical approaches in general, in section 5.1.1. We know biogeochemical fluxes integrate multiple pools, reservoir dynamics and lags, and these are difficult to detect using semi-empirical models. The largest gains recently in representing soil respiration have come from simulating enzyme kinetics and solute diffusion (e.g. DAMM model) as well as simulating microbial growth dynamics. These advances have come from implementing expert knowledge, not from expediting regression model selection.

- We agree that we cannot show yet or beat expert knowledge as encoded e.g. in the DAMM model. Still, we believe that our paper is a first step in this direction.

And therefore it is important to showcase this opportunity to the relevant scientific community. The field of reverse engineering is young and cannot look back to half a century of experimental and conceptual work aiming at understanding soil respiration modelling.

Overall I would recommend that this manuscript be rejected in the current form, and the authors re-evaluate the presentation of the GEP method both in terms of creating certainty within the biogeosciences community that the approach is effective and accessible, as well as readily applicable to field data as was demonstrated with the data from Alice Holt.

- We do believe that our model approach is readily applicable and a novel tool offering the same accuracy as classical semi-empirical models but crucial with new opportunities of interpretation.

As was mentioned, I believe the GEP method has considerable potential, but as the manuscript is currently written my concern is that it will pass unnoticed by the community as a whole due to poor accessibility rather than scientific merit.

- We disagree with this comment, aligning with the other reviewer and also with the overall statement of the strong potential of this novel approach. However, the important step is to get this approach integrated into the modelling community (which is rather small) and allow it to be tested and modified. We do believe that a more general approach and presentation actually will promote its wider usage.

General Comments:

1. **I do not agree with Figure 1, that model development starts with expert knowledge. Expert knowledge does not come about on its own, but comes**

from observations, and regressions are critical to making sense of observations. By helping to identify which variables among a large number of potential explanatory variables correlate to a phenomenon, regression-type analyses lead to the second step in the scientific process: manipulative experiments to confirm hypothesized cause-and-effect relationships. Demonstrating cause-and-effect relationships limits the number of processes that need to be represented in models. I am not convinced that GEP provides a short-cut to this process.

- We thank the reviewer for his valuable point-of view. Maybe the question is rather what one would call "expert knowledge"? We do see observation as one key element of expert knowledge (Fig 1 now includes " including observations") , leading to a first empirically driven (i.e. regression style) approach to model formulation. Yet, once a model could not be immediately rejected it is propagated and used time and again and refined with including more processes etc. This is a tedious process. And here we see that GEP offers a considerable potential indeed. Maybe we have overstated the value of GEP in the manuscript and we will revise it accordingly, but once again - our motivation was thinking and exploring methods that elegantly bypass this approach. For instance, several of the co-authors have worked on the (Migliavacca et al 2011) paper to build a better model for ecosystem respiration in deciduous forests and come to the conclusion that this should be a job realized by a computer. Figure 1 is changed in the manuscript in order to capture and illustrate the points discussed here as well.

2. **Section 3.1 and 4.1, which outline artificial experiments with the GEP method could be strengthened considerably if the authors were to use a simple, mechanistic model of soil or ecosystem respiration rather than a seemingly random set of algebraic expressions. Using such a respiration**

model would allow the authors to attempt to recover the model basis functions and, if successful, enhance the reader's confidence with respect to the data from the site at Alice Holt.

- In this sections we mean to show the capacity of GEP to reconstruct functions from relatively simple example in order to shortly explore the effects of increasing non-linearity and number of variables. As ecological models tend to be more complex and the increase in non-linearity and complexity would no be so clear we chose to stick to some known genetic programming benchmark functions.

  Nevertheless we agree with the reviewer that adding a known ecological respiration model structure in the set of functions to be reconstructed would give more confidence in the application of GEP to ecological modelling. Thus the Q_{10} model is added to the GEP benchmark function set. (p4 l25 and p10 l27-28)

3. **I am concerned about the evaluations of GEP presented in Figs. 3 and 4. Fig 3 compares alternate machine learning techniques by comparing the MEF of the final model selected by each approach. It seems to me also important to compare the actual model structures, not just the fitness score. Did all the techniques recover the original models? If not, is variation in the MEF meaningful?**

  - In this study, GEP is the only approach which gives a readable model structure back. SVM, ANN, RF and KRR lack that property. Thus the comparison is done on the accuracy of predictions, by comparing the modelling scores and residuals.

4. **Figure 4c suggests that GEP was only able to recover about 30-55% of the correct number of parameters. If so, it seems GEP did NOT do a good job of recovering the original models.**

- We agree that at first glance, it would seem bad that the model retrieval with GEP based on the 3 different fitness functions gives a lower number of parameters than the initial number. However considering the high values of MEF when validating against original data, MEF> 0.96, we can draw the conclusion that the GEP performed a feature selection, eliminating "low impact" parameters and returned a more simple equivalent solution.

5. **Another major concern is the exercise shown in figure 7. The authors have examined whether summing predicted component fluxes gives predicted total fluxes that resemble observations. This is an interesting idea, but ultimately not that useful for two reasons:**

   (a) **The observed fluxes were not independently measured, e.g. Rauto was not measured independently, but was calculated by measuring the total flux (Rsoil) minus RH. I think you want to test whether all the variability simulated for the components can explain the variability observed for the total flux, but you don't have a measure of the component fluxes independent from the total flux.**

   (b) **We would like to see that the predictions for total flux are no worse than the predictions for the component fluxes. But in several cases the prediction for component fluxes are pretty poor. E.g. Predictions for RECO won't turn out any better than predictions for Rabove, which themselves were poor. That's not so interesting.**

   (a) We agree that because of learning from derived fluxes, it would be hard make a clear statement regarding the capacity of GEP to learn the variability of the studied sum and component fluxes.

   (b) We believe that nevertheless the exercise is useful as it shows that when we use GEP to learn models for each of the flux, sometimes the low-complexity pressure in the fitness functions make that the final solution has a lower

number of parameters and a slightly lower modelling capacity as well. However we see that when we sum up the models of the component fluxes and compare the predictions of these derived models with the original data, although the models have become more complex, the model performance is not significantly improved. This give us more confidence to state that the more simple models retrieved by GEP in the first place have a sufficient capacity to capture the meaningful information present in the data as well.

6. **The manuscript is figure heavy, consider condensing figures or removing. For example can Figures 5 and 9 be combined in an effective way? Are there other figures that may be unnecessary to the reader if they were described in the text or in a table?**

   - Although we agree that the manuscript contains many figures, we believe that they are necessary (or at least helpful) for reflecting the full picture presented in the text.

Specific comments:

- **Abstract is long, introduces a lot of terminology. Consider distilling to the most important take-homes, and make more approachable for a general audience.**

  The abstract will be shortened and simplified as suggested.

- **p.3 l. 8. The rationale for reordering should also be to try more options, things that people might miss**

  We would like to thank the reviewer for pointing this out. We agree that the increase in the option pool is a large aspect of our approach and somehow we believed that it would be self-explanatory, however it makes sense to state clearly as well. The aspect is added to the manuscript (p3 l8-9).

- **p. 3. L. 30. Why would we expect the functions to be portable across scales? Provide an ecological justification, otherwise this is not an interesting or useful exercise.**

  We believe that this would be more of a wider discussion of the way in which scaling of ecological models is at all interesting and relevant (Urban 2005).

  What we started exploring here is whether a larger grain model would be capable to capture some very strongly influential divers, even by losing specific information and if such processes indeed appear across scales.

- **p. 3. L. 22-35. When reading initially I found it difficult to understand what hypotheses the authors were testing. I think all of this information is there but needs to be re-organized to make it stand out to the reader.**

  Hypotheses and scope of the paper will be re-organized for clarity as it was suggested by other referee as well.

- **p.4 ll. 5. No need to introduce the conclusions. Consider shortening this to reduce repetition.**

  Thank you for you suggestion. Paragraph removed.

- **2.1 This section was not clearly written, I suggest more careful editing by co-authors. Please avoid including extra words in parantheses, they add complexity without clarity.**

  Section will be re-written for more flow clarity in the revised manuscript as suggested.

- **p.4 ll.15. Is the process of mapping operations to strings relevant to model fitting? I don't think so. Either this is excessive detail about the internal workings of GEP, or you need to explain how this is relevant.**

The process is relevant as it is one of the characteristics of the GEP approach. We apologize for not making this clear in the manuscript already, however this aspect and the effects of mapping are explained in more detail in the method section (2.1) of the revised manuscript.

- **p. 4. L. 20, what do you mean by "solution" The final selected model? Or the respiration predicted by that model? "Genes" and "chromosomes" should be presented in quotations initially.**

  Solution is the final selected model structure. Quotations are added as suggested.

- **p. 4 l. 30 I think you can shorten this paragraph to one sentence, simply state that in each generation, the best variants of a chromosome are determined by a fitness function described below.**

  The paragraph could be shortened, however the suggested line is not accurate as in a generation, there is only a variant for each chromosome , and the fitness function determines the ranking of all chromosomes in that generation.

- **p. 4 l. 32, what is an individual? Do chromosomes make up individuals?**
  An individual is a chromosome that encodes a mathematical formulation, made up by a set of strings called genes.

- **p. 5, l. 1 What is a hyper-parameter? Again, please try to avoid parenthetical phrases in this paragraph.**

  A hyper-parameter is a set of parameters which need to be set for the runs of a certain approach. Definition is added to glossary and further parentheses are avoided.

- **p. 5, ll. 12 "upon request" rather than "on demand".**

  Changed as suggested.

Interactive comment

- **p. 5, l. 11-14 most of this information doesn't appear useful, for example, does it actually matter that the cluster had 51 nodes? If someone ran it on a cluster with 12 nodes would it also work but be slower? Either explain the relevance of these details or remove them.**

  The description of the system on which all experiments should be relevant as the results might be influenced by the hardware set-up, due to the initialization of the random seed, speed of solution return and so on. Nevertheless, all non-necessary specification are removed.

- **p. 5, ll. 31 Consider omitting "derived from information-theoretic considerations".**

  Thank you for the suggestion. Omitted.

- **p. 6, ll. 20-25. I didn't understand the reason for this additional optimization. This sounds very much like ordinary regression model selection; does this undermine the unique value of GEP?** The original GEP gives a solution in the form of a general mathematical structure. For accurate scaling a further parameter optimization would be recommended. The value of GEP lays in the capacity of constructing the structure based on the on information found in the input data.

- **p. 6, ll. 27 Scaling noise with signal amplitude: This is good to include! This has been shown for soil respiration too (Lavoie et al. 2015, JGR-Biogeosciences, doi: 10.1002/2014JG002773)**

  Thank you for providing the reference. Added to paragraph.

- **Section 3.2.1 The first two paragraphs are repetitive in describing computation of GPP.Consider omitting or shortening the section on soil flux measurements, since these methods were reported previously.** Would not

remove as they are relevant to experiments and results presented, but will try to shorten.

- **Section 3.2.4 This paragraph can be removed to shorten. Figure 3c, consider omitting. It is repetitive, and the manuscript already has a large number of figures.**

Figure 3c removed. However we believe that the paragraph is needed for anticipating the comparison done on real observation between established models for terrestrial respiration in the community and the GEP based models.

- **p.12, l. 7 Sentence starting "We find that the global modelling performance. . ." Please reword, I don't understand this statement.**

Reworded as suggested.

"We found that when we compared the modelling performance of the models built as sum models from GEP built models for the component fluxes with the original GEP models built on the sum fluxes there not significant differences. However the total number of parameters is much larger for the sum models. This can be a result of the GEP approach eliminating the "low impact" drivers due to complexity pressure. We can see as well that the sensitivity of the sum fluxes to certain drivers can strongly manifest itself only in certain components which is why the drivers only get selected in the models built for those specific components." p12 l7-10.

- **Figure 12, is there a reason that this is presented in a polar plot? It seems on first glance that it could equally be presented as a 4-pane set of cartesian time series plots.**

By using polar plots, we reveal that the seasonal biases of the studied fluxes and the capacity of the models to capture/or not some of the variations in specific times of the year. But yes, it is a matter of taste as well.

References

M. Migliavacca, M. Reichstein, A. D. Richardson, R. Colombo, M. a. Sutton, G. Lass- lop, E. Tomelleri, G. Wohlfahrt, N. Carvalhais, A. Cescatti, M. D. Ma- hecha, L. Mon- tagnani, D. Papale, S. Zaehle, A. Arain, A. Arneth, T. A. Black, A. Carrara, S. Dore, D. Gianelle, C. Helfter, D. Hollinger, W. L. Kutsch, P. M. Lafleur, Y. Nouvellon, C. Reb- mann, R. Humberto, M. Rodeghiero, O. Roup- sard, M. T. Sebastia', G. Seufert, J. F. Soussana, and K. Michiel. Semiem- pirical modeling of abiotic and biotic factors con- trolling ecosystem respira- tion across eddy covariance sites. Global Change Biology, 17(1):390–409, jan 2011. ISSN 13541013. doi: 10.1111/j.1365-2486.2010.02243.x. URL http://doi.wiley.com/10.1111/j.1365-2486.2010.02243.x.

D. L. Urban. Modeling ecological processes across scales. Ecology, 86(8): 1996–2006, aug 2005. ISSN 0012-9658. doi: 10.1890/04-0918. URL http://doi.wiley.com/10.1890/04-0918.