Geosci. Model Dev. Discuss., 4, C429–C436, 2011
www.geosci-model-dev-discuss.net/4/C429/2011/

**Geoscientific
Model Development
Discussions**

Interactive comment on "iGen: a program for the automated generation of models and parameterisations" *by* D. F. Tang and S. Dobbie

D. F. Tang and S. Dobbie

daniel.tang@manchester.ac.uk

Received and published: 12 July 2011

The authors thank the referees for their constructive comments. Both referees recognised the importance of the technique described in this paper and offered suggestions for minor revisions to improve the paper's presentation. Based on these suggestions, a revised version of the paper has been written in which the following changes have been made:

Anonymous referee 1 ————————-

"1. I do not like the introduction. First, I do not agree that when it comes to developing parameterizations, the main problem is in the "model assumptions"."

- We were using the term in a broad sense but because of the potential for confusion

we no longer use this term.

"I think the problem is that parameterizations represent our understanding of the system and ability to reduce the system behavior to simple rules. For that, one needs to understand the system in the first place. This may come from the observations (i.e., isolating the forcings and finding the response), or from model simulations, where the forcings can be better controlled and the responses perhaps easier to isolate. The fact the models have their limitations (through the model assumptions or, more generally, model imperfections, due to limited spatial resolution, for instance) is a secondary issue in my mind. The key point is the thought processes that leads to the understanding of model behavior and to the designing simple rules that form the key component of a parameterization. If I think about the parameterization of deep convection, for instance, an area where I have some expertise, I can clearly see how this thought process works and why it is so difficult to develop a robust parameterization because of the complexity of the system. Note that the parameterization can be just a simple rule or even another model (simple or complicated), but the key element is that such a model does encapsulate our understanding. And this is where the iGen comes in: it can analyze a complex model and create a set of simple rule that can be quickly use as a parameterization. But it also has a drawback because by using iGen one can come up with rules without physical understanding, and the understanding is in my mind the key element of progress."

- While it is true that iGen can be used to avoid the need for physical understanding (if all one wants is a functional relationship) iGen can equally be used as a tool to aid the development of physical understanding (in fact, referee 2 mentions that "the ability to simplify equations of a high-resolution model by making approximations is part of the process of understanding a complex system"). This matter is now discussed in the last paragraph of the introduction.

"...I personally would prefer to see the introduction built around such discussions."

- the introduction has now been rewritten centred around a discussion of these points.

"2. The reference to superparameterization (SP) is actually a good example of what I refer to above. Since the convective atmosphere is a complex system involving multiscale dynamics, why not to use another model (that we know does a decent job in representing deep convection) to replace convective parameterization? This is the idea for SP. (BTW: SP was suggested by Grabowski and Smolarkiewicz in 1999; the reference to Khairoutdinov et al. does not seem right here; perhaps a better reference would be 2003 BAMS paper by Randall et al.)"

- This reference has been changed to the original Grabowski and Smolarkiewicz.

"...IGen can analyze such a model and create a simpler set of rules that can replace the subgrid-scale model. I am sure the authors would like to embark on such a project, although analyzing 3D moist fluid flow model seems quite challenging. Perhaps a comment on that would be desirable."

- A discussion of superparameterisation has been added to the introduction.

"3. P. 846, l. 6: I think this is the average KE of an atom that is proportional to temperature, not the other way around."

- Surely if A is proportional to B then B is also proportional to A.

"4. It is not clear to me what the parameter r is on p. 851, line 7. Can this be explained a little better?"

- An explanation has been added

"5. P. 862, the discussion of the entrainment in stratocumulus simulations. I do not understand what "variance per second" and "variance of entrainment per second" are. Please clarify. Overall, this section can be removed from the manuscript and replaced by a reference to the second paper as the details of the results are very sketchy. Perhaps more importantly, I am not sure if comparison between 2D simulations and ob-

servations is that easy. I am sure I will have some comments on this for the second paper."

- This section has been reduced in size by removing the incidental details of the experiment. However, I think it is important to communicate the fact that iGen can be used to analyse models of realistic complexity so the essential details of the experiment have been retained to make this point, interested readers can then read the companion paper.

The "variance of entrainment per second" was calculated as the variance of the entrainment over a 4 hour period divided by 14400 seconds (4 hours). When the variance per second is multiplied by the duration of a model's timestep we get the variance of entrainment over the duration of that timestep. This point is explained in the companion paper but to prevent confusion here we have replaced "variance per second" with "variance over a 4 hour period", the distinction is not important for the purposes of this paper.

Vaughan Phillips ————–- "This is a seminal paper, establishing a new approach in the problem of parameterizing complex algorithms. It may lead to better statistical representation of sub-gridscale processes that are too small to represent explicitly in large-scale atmospheric models and other models. It may prove useful for financial modeling. For this reason, I think the authors should invest effort in improving the overall style of the paper. Phrases like "notoriously" or "fear" or "successfully" or "the authors firmly believe" are not really scientific."

- The language used in the paper was imprecise in places and I have spent quite some time improving this. The final sentence has been left unchanged: It is up to the Scientific community to decide whether or not iGen is an important tool for model development, but readers may be intereseted in our opinion on the matter, given our experience with and knowledge of iGen.

"There needs to be a "Description of iGen" section, with an overview of the general

structure of it and perhaps a flow-chart. At present crucial details are in the introduction instead of being in such a description section."

- this section has been added and a section moved from the introduction

"Also, practical details of implementation of the algorithm would be good to include, perhaps in a new section."

- An 'implementation' section has been added

"It would be good to compare the proposed method with existing alternatives for creating cheap algorithms, such as neural networks."

- a paragraph has been added explaining why iGen differs from curve fitting and neural networks.

"Perhaps when there are too many inputs for iGen to cope with, neural networks or some other approach, would be better."

- the discussion section now contains comments on the case when there are too many variables.

"The validation section is fine. It would be good to show how iGen converges as the specified error to be accepted is diminished and the sparse grid resolution is enhanced."

- The convergence properties of the sparse grid used by iGen has been discussed at length in the literature. References have been added.

"It would be useful to see how iGen performs when given a 1D, or 2D lookup table (e.g. for droplet-droplet collision efficiencies). Can it cope with hundreds of if statements in a lookup table?"

- Yes, iGen can cope with lookup tables. The speed of the analysis would depend very much on the data in the table and the context of the rest of the program. It would

be very difficult to make provable general statements about iGen's performance in this case.

"A topic to tackle concerns accessibility to the iGen algorithm and reproducibility of the results. Where and when is iGen going to be available to the community, if at all ?"

- iGen source code is not currently available to the public but should be released very soon. Interested parties should contact the lead author at the email address shown on the paper.

"Finally, the limit on the number of input variables for iGen to function properly seems quite problematic. It would be good to comment on how implementation of iGen could circumvent such difficulties, (for example by selective use of iGen for certain parts of a complicated program that are identified as being more tractable with more inputs)."

- this is now covered in the discussion section.

"Page 844, line 9-10: A key issue about the utility of this proposed technique is whether the bounds on the error converge as iGen is altered towarsd a higher resolution. Does the approximate solution tend towards the original exact solution as the specified error is reduced ? It would be good to show this on the validation plots."

- The convergence properties of the sparse grid used by iGen has been discussed at length in the literature. References have been added.

"A potential problem is the number of variables of the parameterising model and whether the approximate solution has a realistic sensitivity with respect to perturbations of the input variables. Could the approximate solution exhibit unrealistically wide transitions from one state to another in sensitivity tests ?"

- I think the referee is referring here to the Gibbs phenomenon or ringing effect at discontinuities. Polynomials never converge to discontinuous functions. However, discontinuities can be formally removed from a wrapped model by adding uncertainties on the inputs (i.e. by adding a random number to the input to represent the finite precision

of the input). This is now mentioned in the implementation section.

"Page 845, Line 6-9: This is an important point and could be developed more in the paper. The ability to simplify equations of a high-resolution model by making approximations is part of the process of understanding a complex system (See Yano$-1òùs papers that do this for deep convection; see Isaac Held$-1òùs philosophy in BAMS about the necessity to create analytical simplified equations in order to understand a system). By creating approximate formulae, iGen can advance scientific understanding of the essence of a system."

- A paragraph on this has been added to the introduction

"Page 845, line 11: The description of the overall structure of iGen is not as lucid as it could be. It needs to be stressed that the input to iGen is the source code of the highresolution model and the output from iGen is more simplified source code, to replace the original source code. A flow-chart or similar diagram would summarize clearly the overall structure of iGen."

- this section has been re-written and is now much clearer.

"The paper needs to be structured better. At this point, the introduction needs to be wrapped up and the description section should start. I think towards the end there need to be some notes about practical details on implementation of iGen. What programming language must the original source code be in ? Do any intricate parts of the original source code need to be simplified (E.g. by iGen) to prepare it in advance ? What to do if the error bounds are large or if there are too many input variables ? Is there ever advantage to combining use of iGen with neural networks or other techniques for creating simplified algorithms, and how does iGen complement these alternatives ? Try this: 1 Introduction 2 Description of iGen program for automatic param. 2.1 Overview [with flow chart indicating the very general sequence of steps that iGen takes] 2.2 Symbolic analysis applied by iGen: operations defined 2.3 random numbers 2.4 fixed loops etc C364 3. Validation of iGen 4. Practical Implementation of iGen 5. Discussion

6. Conclusions."

- the structure has now been improved, with 'description' and 'implementation' sections added, as suggested. All the points raised have been dealt with in the appropriate sections.

"Page 857, line 13: "without fear" is not a scientific phrase in this context; try to use the passive tense of the verb instead of "we", so as to focus on what is being described."

- I (the lead author) do not subscribe to the dogmatic view that the passive tense should be preferred in scientific writing. The sentence in question would have to be changed from

'We therefore restrict ourselves to the consideration of the basic recursive functions without fear that this will be a problem for our proposed application.'

to

'Therefore, the analysis technique is restricted to the consideration of the basic recursive functions but this is not considered to be a problem for the proposed application.'

The meaning and precision of the two, in context, is the same in all important respects yet the original is less staid and is easier to read and understand.

"Page 861, line 10: I think the full equation for the scattering cross-section of this droplet size distribution needs to be written."

- There is no single analytic function that describes the scattering cross section, the most concise description is in terms of recursively defined functions and would require 12 equations. Mie theory is discussed in the literature and a reference is given for the interested reader. An exact, numerically calculated value is given in the plot.

Interactive comment on Geosci. Model Dev. Discuss., 4, 843, 2011.