

Interactive comment on “Quantitative performance metrics for stratospheric-resolving chemistry-climate models” by D. W. Waugh and V. Eyring

D. W. Waugh and V. Eyring

Received and published: 21 July 2008

We thank Dr Morgenstern for his comments. He appears to not be "fond" of quantitative metrics of CCMs because it involves subjectivity and has the potential to be misused. We agree that subjectivity is still needed, but this does not mean we shouldn't be as objective (quantitative) as possible. Similarly, just because someone may misinterpret or misuse an analysis is not a reason not to do it. We believe that most scientists using this quantitative analysis will do so in a scientific manner, and that it will lead to a better understanding of models and their projections.

1. We agree that the results could depend on choices of parameters used, and that forming a single grade will depend on weighting used. However, we don't think that this

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper



should prevent these issues being explored. Yes there is a chance of "endless, futile debates over the details of the metric which are fruitless", however there is also the potential that this will lead to a better quantification of the differences between models (or versions of the same model) and better understanding of what the key parameters / diagnostics are. Even if no consensus is reached on parameters and weights it will not necessarily be a fruitless exercise, as some thing will be learned from the different tests and discussions.

2. If modelers are going to include non-physical fixes they will do this with or without studies like this one (and you claim this is already occurring), so we don't think this is a very good argument to not to do a quantitative analysis. The diagnostic tests applied in this study have been used because they are important for stratospheric processes. While we agree that developing a model so that it gets good grades in ONLY these diagnostic tests is not a good strategy, a model group, whether or not participating in a model intercomparison campaign, would be very well advised to evaluate the ability of the model to simulate the tape recorder, Cly, polar temperatures and the other diagnostics that are presented in this paper. More diagnostics will be included as a benchmark for CCM evaluation as part of the SPARC CCMVal report with the goal to develop a standard for CCM evaluation that models should fulfill before they are used with a large weight in stratospheric ozone projections. As pointed out by referee 1, the metrics based approach offers a fair approach to all model groups and has the advantage of being able to include all model projections in the ensemble, rather than having to exclude some of them from the start, which among other things is politically more palatable. While no approach to grading will be perfect or acceptable by all, having no accountability to meet certain standards for certain diagnostics (i.e. grading) is the least perfect of all approaches.

3: We disagree with your view that the current situation is not untenable. Please look at Figure 7b of Eyring et al. (2007) or Figure 6-12 of WMO (2007). Do you really think this is a good way to present stratospheric ozone projections to policy makers? Certainly

[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)[Discussion Paper](#)

not. We have used these figures in the above mentioned studies because at that time we didn't know how to do this better. However, it has strongly motivated us and others to develop a more quantitative approach to CCM evaluation. This approach has been discussed at the CCMVal workshop in Leeds in June 2007, see Eyring et al., Report on the Third SPARC CCMVal Workshop, SPARC Newsletter No. 30, p.17-19, 2008. While not everyone agreed that we have to build up this concept, the consensus was that after having developed a comprehensive list of diagnostics of model performance previously for radiation, dynamics, transport, and chemistry, the next step should be tackled and a more quantitative metric of performance (grade) for each model and each diagnostic should be developed. We provide here the first step towards reaching this goal.

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 10873, 2008.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper