



## ***Interactive comment on “The ACCENT-protocol: a framework for benchmarking and model evaluation” by V. Grewe et al.***

**Anonymous Referee #1**

Received and published: 6 February 2012

General Comments: This paper summarises a comprehensive framework for undertaking model evaluation. The use of example questions with which to test the models is very useful as is the discussion of uncertainty. It would be nice to see the results of this framework put into practice as a proof of context. Perhaps that is to follow in a future paper.

Specific Comments: Table 1. I don't think it is necessary to include this as a table as these questions and their abbreviations are clearly stated in section 2.1.

Section 2.1: It might be worth stating here that asking whether a model is 'good' in the general sense is not as useful or perhaps even sensible as there are too many variables to assess, with likely too much uncertainty to constrain any results. Having a specific specific aspect (e.g. ozone layer recovery) provides much greater power for

C1500

the validation. You quite clearly use very specific questions as examples of 'purpose' but its probably worth saying that a more general question is not a useful 'purpose' if that is the case.

Section 2.5: Can you include some brief discussion on a useful format in which to provide quality information? Is it a text description or a machine readable set of fields on model grid resolution or something entirely different? It is vital that this information is easily accessible and understandable to the scientific community and perhaps even those end users of products for decision making. Many different institutions could produce a variety of different formats of assessment. This could lead to confusion and in the worst case scenario it may lead to users ignoring such assessments when using the data. This is perhaps not something you can go into detail with here but a little discussion of the issues would be useful, especially if there is a preferred format that you are aware of.

Measurement uncertainty in in situ observations of climate variables such as temperature, precipitation, pressure etc. is not well documented. There are many different instruments with differing levels of precision. Calibration of these has not always been regularly conducted and any adjustments not always documented. Any metadata that do exist are rarely in a digital format. I hope the situation for atmospheric composition data is much better and as this protocol is aimed at that data perhaps the original statement is ok. However, its worth being aware that the picture is not so rosy for all climate data.

Page 3607 line 22 - Do you really mean 'smaller' here? For the models to be significantly different from reality would they not have to lie outside the uncertainty range given by the observations? Perhaps rewording this sentence may make its meaning clearer - I may have misunderstood.

Section 2.6: I'm not quite clear on the equations here but that might be my lack of familiarity with the specifics. Is 't' time?

C1501

Section 2.8: Given the subjectivity of benchmarking can you add a comment here stressing the importance of clarity in how the assessment was conducted. Anyone using the quality information or grading should understand exactly what was assessed and how. Also, if different model products are to be compared using this assessment then it is important that the same assessment was conducted on all products otherwise a comparison of the validation/assessment is not really useful. If assessments were to differ then this should be clear.

Technical Comments: (suggested changes are capitalised) Section 2.2 I don't think you need 'As discussed in the last section' as this follows very nicely anyway. Perhaps just saying 'The IDENTIFIED purpose...' would suffice.

Page 3603: line 11 - 'extend' should be 'extent' line 20 - 'input parameters AND boundary conditions such as emissions...' sounds more logical?

Page 3604: line 16 - 'There ARE a huge variety...' instead of IS.

Page 3607: line 3 - 'has impacts' I think instead of 'have an impact' lines 5-8 - I don't understand this sentence. Please can you reword it.

Colons needed before lists: p3602 l26, p3605 l12, p3608 l5

Page 3609: line 12-13 - 'leads to the conclusion OF which parameters' sounds better. line 21 - 'feedbacks AND altitude of aircraft emissions...'

Page 3611: line 1 - 'ozone grade and a temperature grade ARE not comparable' lines 14-18 - 'an example which' no comma needed lines 14-18 - 'for the quality indicators AND THEN using its lower...'

Page 3612: line 15 - no 'Especially' needed just 'The uncertainties..' line 27 - Perhaps use 'BENCHMARKING data' here as opposed to 'observational data' for consistency.

---

Interactive comment on Geosci. Model Dev. Discuss., 4, 3599, 2011.