

Interactive comment on “A community diagnostic tool for Chemistry Climate Model Validation” by A. Gettelman et al.

Anonymous Referee #2

Received and published: 6 July 2012

The paper describes a modular analysis and plotting tool mainly aimed at climate and chemistry-climate models. It explains well the workflow and the requirements for its application. Examples are provided to illustrate the capabilities of the tool.

General:

I am sure the tool will prove useful in tracing and characterising model changes (as are some other tools), but I would have preferred a slightly more neutral way of describing this fact. Not every model change will result in a model improvement (a word slightly overused in this paper); I would encourage the use of the more neutral word change instead of improvement.

Given the undoubted abilities of the tool to summarise and compare physical quanti-

C399

ties I found the focus on performance metrics slightly odd. Quantifying model changes and illustrating model improvements are non-trivial tasks, and even though those are helped by tools such as the one described here, the overall model assessment will still rely on the choice of diagnostics used. This choice will hopefully be influenced/guided by the “fit for purpose” principle. Therefore, instead of focusing on performance metrics I would have preferred more details about the practical implementation of derived physical diagnostics. For example another example like the tropopause trend calculation would be nice, maybe deriving streamlines from wind components, or similar. Even though I do not doubt the statement that “The code can produce performance metrics and is designed to enable comparison of models to observations.” I find emphasising “quantitative grades” (relative to observations) slightly irritating in the context of describing this tool. Obviously, the results of such calculation do not depend on the tool discussed here, but on the observational data used. Some observational data seems to be part of the package, but I wasn’t quite clear about which and how error characterisations are considered, please clarify. Consequently I would like to ask the authors to de-emphasise the quantitative grades and to highlight the ability to characterise changes between model versions. Maybe a comparison of two very similar runs could be shown, where the significance of the zonal-mean zonal wind changes (or another quantity) is assessed.

Minor comments:

When talking about the input data for the tool it would be helpful to spell out in a little more detail the link between the CF-netCDF and CCMVal-2 format conventions.

Small technical question regarding Figure 2: I assume pressure labels could be defined as non-overlapping? Presumably the data plotted was on a pressure grid and the height scale was converted on the fly using a constant scale height?