Overall Comments:

I have used one of the IGCM3 versions extensively over a period of years and find it an excellent workhorse of a model, ideally suited to long integrations and to coupling to other climate components. Both the model and the Forster et al. (2000) paper which serves as reference for the model description are over a decade old and an update in the literature is past due. I was excited to receive this manuscript.

Unfortunately, in its present state the manuscript gives the impression of a hastily-assembled overview for a small group of insiders. I do not recommend publication of the manuscript because, in its present state, it fails to fulfil the requirements of the journal (see www.geoscientific-model-development.net/submission/manuscript_types.html), specifically with regard to the following points:

> "The papers should be detailed, complete, rigorous, and accessible to a wide community of geoscientists. "
>
> "(...) ideally, the description should be sufficiently detailed to allow in principle the reimplementation of the model by others, and so all technical details which could substantially affect the numerical output should be described. "
>
> "The model description should be contextualised appropriately. For example, the inclusion of discussion of the scope of applicability and limitations of the approach adopted is expected."

Let me make clear that I am still excited by the promise contained in this first draft, and look forward to reading a properly complete version. General, specific, and technical comments and suggestions follow.

Who is the intended audience?

1) Current users only?
In its present state the manuscript can only serve those who already use the IGCM in one of its forms and are familiar with its strengths and with extant online resources. If current users are your target, don't put yourself through time-consuming peer review; write an internal technical report.

2) New users?
You have here a model that is fast and flexible; that can be readily harnessed to other components of the climate system (full ocean, stratospheric chemistry); that can be efficiently run for very long periods of time (multiple centuries) even on a single cpu and thus allows e.g. robust ensemble statistics or studies on orbital-forcing timescales which state-of-the-art CGCMs cannot; that therefore lends itself ideally to process-oriented studies of the climate system without sacrificing too much accuracy: say so! Line 87 announces that T42L20 is "the standard configuration for studies of the troposphere and climate". This sentence on its own is simply not defensible today (see Table 1 of Bracegirdle et al. (2013) for resolution and lid heights of CMIP5 models). Users have a large choice of highly accurate models. The niche filled by the IGCM must be fully and clearly spelled out, starting in the abstract.

What is the goal of this paper?

1) Highlighting improvements?
The end of the abstract and lines 49-50 in the Introduction suggest showing the improvements made with respect to version IGCM3 are a goal. Improvements are thereafter not mentioned again and are never shown graphically. If this is truly a goal, "before and after" plots and a detailed discussion of improvements seen in the modelled basic-state climate would be needed. I recommend you simply drop "improvements" in favour of a solid description and validation of the IGCM4 as it is. The description could be accompanied by a table or bullet-point list that details the technical changes from IGCM3 to IGCM4 for those familiar with IGCM3.

2) Description of the IGCM4?
It seems from the manuscript that this is the true intent of the paper. But at present both the description of components and the validation section are too perfunctory to be generally useful.

Specific comments:

Introduction:

Provide a thorough motivation for a model like the IGCM within a more detailed explanation of the hierarchy-of-models concept. What are the strengths and limitations of a simpler model? What does it contribute that state-of-the-art models cannot and who would benefit from such a model? Ending the Introduction with the customary reader-orientation statement outlining which topics are covered in each section would be helpful.

Section 2:

This should be a complete, thorough description of the model configuration and ALL its physics. Simply referring readers to the Appendix from Forster at al. (2000) leaves too many gaps. Include all relevant information and highlight updates as appropriate.

2.1 If three "standard configurations" are named, results from all three should be shown. It seems to me from all your graphs that T42 is the standard horizontal resolution. It can be mentioned that T170L20 is an additional possibility which has been tested.

2.2 This does not warrant a separate section; parallelisation is standard in GCMs. The pet name of the UEA's cluster is of rather less interest than the machine architecture would be! Which compiler do you use, what is your timestep, and, consequently, what is performance speed not only on multiple processors (which you give), but also on a single processor for comparison? Are production benchmarks available upon request or on a website? Of the three websites given in the text, two are the same and none relates explicitly to the IGCM4. Listing reference websites should come at the end, not replace proper descriptions.

2.3 Rewrite as a complete description of the IGCM4 model as is. It is worth making a sub-section for surface processes (including the new snow and sea ice) and a separate one for convection.

2.4 Name all radiating species. In keeping with my "sell the model" message, this is an opportunity to point out how easy it is to modify the concentration of any species via namelist in the rundeck, for anyone envisaging climate sensitivity experiments. Discussion of aerosol treatment (or of the compensation for its absence) belongs in this section, not in the Conclusions (lines 339–347).

2.5 This sub-section describes what is in my view one of the biggest improvements to the model. It would be desirable to find the same level of care and detail in the other model description paragraphs.

Section 3:

This section is the most disappointing. Rather than illustrating model strengths by presenting the sort of robust statistical analyses enabled by very long runs, or by a large ensemble of runs, you present a few graphs from 50 years of the most basic model setting. It is unclear to me why you show a T42L20 model at all, when only a 50-year integration is presented. Furthermore, once the surface fields are validated with the T42L20 version, you switch to T42L35 for the zonal-mean fields, before finally adding on the q-flux ocean for a brief mention in the Conclusions. I agree that results from a coupled full ocean warrant their own paper, but given how important even a slab ocean is for achieving correct stratospheric vortex driving (see Winter and Bourqui, 2011), and how comparatively inexpensive it is to run, such an integration would constitute a valuable part of the Results section.

I recommend that each of the fields you wish to validate be shown for each of three model versions: T42L20, T42L35, and T42L35 coupled to the slab ocean. Do at least 100 years (as you have done already in the T42L35 setting), with the running times for each one given in the Configuration subsection. The depth of the slab ocean and length of spin-up time required should also be stated. The variability of each configuration could be illustrated with graphs of the variances of selected fields.

Use contours of the "truth" fields overlaid on shaded model fields; it allows much easier comparison than seeing the plots side-by-side does. I expect that the surface temperature will not change between the three IGCM4 versions, but perhaps precipitation will: this will be interesting to analyse. In the zonal mean, it will be interesting to see the difference the lid height makes to the wind field. What role does the GWD play here? What difference does the slab ocean make between versions having the same lid height? I am aware that the manuscript is categorized as a "model description paper" to be published in a model development journal, but readers and future users will find value in seeing at a glance which version best lends itself to whatever climate study they intend to perform. The level of detail given to the annual count of stratospheric sudden warmings is inconsistent with the very sparse treatment of everything else, and is not particularly well backed up by graphs.

Section 4:
New concepts such as climate sensitivity should not be introduced in the Conclusions. This can be addressed in Section 3, with a short definition of the term and references to Joshi et al. (2003). Similarly, the treatment of aerosols should be addressed in the model description section.

Bibliography:
Check that order is alphabetical throughout. You cite three papers from the 1990s using early versions of the IGCM to study Mars and zero papers done in the last five years using late versions of the IGCM coupled to a slab ocean to better study the Earth system!

Technical comments:

Line 86: When the truncation wavenumber is 42, then 2.8 degrees represents the grid spacing of the transform grid. It is not the resolution of the model; the resolution is half of one of those 42 smallest waves, or about 4.3 degrees.

Lines 183-188: What is the reference for this new GWD scheme? Is it still Lindzen (1981)?

Line 190: 3×10^{-6} **or** 3 ppmv but not both!

Lines 213-215: Reference needed.

Line 219: Reference needed.

Line 222: "Results" is perhaps not the best term, since no experiment is being performed. "Model validation" would be more accurate.

Lines 225-226: Explicitly state which climatologies are used.

Lines 240-241: This is inconsistent with the figure caption and the text printed on the figure itself. See comments under Figures.

Figures:

All figures: more complete captions. Data sources should be explicitly named and there's no harm in repeating the data sources from one figure to the next. Which years from the reanalysis datasets are used, and why?

1. Not sure a figure is warranted here, especially when it gives the impression that L20 levels are indexed 16:35. Perhaps a table with exact pressure values as well as approximate heights in km?

2. Colourbars! Or much larger and bolder labels on the curves.

3. The "observations" are in fact reanalysis data; give complete sources in the figure captions. See my comments above for combining reanalysis contours with IGCM shading, and for showing all three of the IGCM4 settings discussed in the manuscript. The bias over northern Africa in both seasons, Australia in DJF, and high-latitude North America in JJA is widely greater than 10 K, in contradiction with the text. Instead of shrugging this off as "reasonable compared to CMIP5", it would be more helpful to have a more careful analysis of possible reasons. How much is this related to the simplicity of the soil scheme and the ensuing surface humidity errors? What does this bias do to convection?

4 and 5. These are confusing, in part because of inconsistencies between the caption and the text, especially the top right panel, described as "difference between IGCM4 and observations" in the caption but not in the text or in the small title under the panel itself. The centre and right bottom panels are overkill – why not just show the variance of the model precipitation? The top left panel contains the colour green and the colourbar at the bottom does not – are all panels really on the same scale?

6. Provide a complete caption.

7 and 8. This is minor, but Figure 2 has North on the right at the bottom of every panel, and these Figures have North on the left on the bottom panels only. As a kindness to the reader, you could draw attention to this reversal. (My first impression was that the figures had accidentally been mislabelled.) Otherwise these in particular should be shown for all settings of the IGCM4 that you discuss, with ERA-40 contours overlaid. I am interested to see whether the boreal winter tropospheric jet axis shifts slightly in the meridional between the high-top and the low-top versions.

9. The caption says this is for DJF and the graph title says DJFM. Which is it? It should be the same months as the zonal-mean wind and temperature plots. The text says the "zonally asymmetric component of the circulation is apparent" (line 295); it would be more apparent if you indeed plotted only the zonal anomaly of the geopotential height. If you were to show this Z' field at the 200 hPa level as well, it could nicely tie in with your later discussion of the stratospheric vortex variability, since the wave-1 and wave-2 stationary patterns that mainly force the stratospheric circulation would stand out well.

10. Here it will be very interesting to see the differences between versions of the IGCM4. I'm not sure what is contributed by differentiating between splitting and displacement events, since the manuscript does not go into the mechanisms required for either or address why the model "misses" when it does.

11. A more complete caption here; presumably these are annual, global averages?

References:

Bracegirdle et al. (2013), JGR-Atmospheres, vol. 118, pp. 547–562
Forster at al. (2000), Clim. Dyn., vol. 16, pp. 833–849
Joshi et al. (2003), Clim. Dyn., vol. 20, pp. 843–854
Winter and Bourqui (2011), Geophys. Res. Lett., vol. 38, 10.1029/2011GL047011