Review of "Downscaling the climate change for oceans around Australia" by M.A. Chamberlain, C. Sun, R.J. Matear, M. Feng and S.J. Phipps

The paper presents a framework for dynamically downscaling climate change projections from a coarse resolution AOGCM onto an ocean mesoscale resolving model. The paper outlines a method of downscaling onto the high resolution model, evaluates the role of air-sea feedbacks, and presents evidence that downscaling projections onto a high resolution model can result in some differences in the near surface environment.

The topic of the paper is certainly of interest, but I have several difficulties with appreciating its value:

i) Is the focus on the method of downscaling or on the effects of downscaling on the Australian region? If the authors feel they are presenting a novel downscaling method, then the methods need to be more clearly justified and much better explained. The explanation of the methodology is incomplete, and spread throughout the paper in various sections. As it stands now, a reader would find it very hard to evaluate the merits of the methodology or employ it.

ii) If the focus is on the effects of the downscaling then the model simulations employed need to be better evaluated (see comments below).

iii) I think your experimental design is awkwardly described and hard to justify. CTRL, RELX, FREE and STRS should all be initiated from the same Spinup state and integrated for the same period of time. As it stands now different experiments are started from different control states and run for different lengths of time. I cannot see much value in the FREE and STRS experiments since they have been integrated for such a short period of time and are initiated from year 3 of RELX rather than a control state. This awkward methodology makes it very hard for the reader to see value in the results.

So from my point of view the paper doesn't do a very good job of justifying the downscaling methodology and the experimental design makes it difficult to evaluate the effects of the downscaling. I find it difficult to justify the publication of this study.

Specifics

Pg 426 Line 4: A small point, but all global climate models resolve boundary currents. They may not do it well, but they are present nonetheless.

Pg 426 Line 5: Perhaps note here for clarity that the global ocean model is coarse resolution except in the Australian region, where a mesoscale resolving model is nested.

Pg 426 Line 13: Differences have been shown, but significance has not. Are the differences significant relative to internal variability or drift in the ODM?

Pg 426 Line 19: For the framework to be attractive to others it needs to be more clearly outlined and compared to others. As presented it is difficult to determine if the methodology is sound.

Pg 426 Line 25: please clarify 'horizontal' rather than "spatial" resolution, or discuss both horizontal and vertical resolution.

Pg. 426, line 28: mesoscale resolving models are integrated for multi-decadal time-scales. Perhaps modify text to centurial and millennial time-scales.

Pg. 427, line 26: Specifics of the downscaling methods used by other 'limited-region' ODM studies should be discussed. How is your methodology for applying the climate change forcing to the ODM different? Is the only difference that your high resolution ODM is nested within a coarse resolution ocean model? Or is the surface boundary information applied differently?

Pg 427, line 27: I would move the paragraph starting at line 22 on pg 428 to here

Pg 428, line 24: again, coarse models have boundary currents, they are just not accurately resolved.

Pg 428, line 27: "By utilizing a global- scale ODM the passage of information from the AOGCM to the ODM changes from how we handled the open boundaries to how we initialise and force the ODM."

I don't really appreciate how much has 'changed' by using a global ODM with a high resolution nest. The high resolution nest still has a boundary with the coarse outer model. A discussion of the boundary between your high resolution nest and the outer model should be included. Are the surface forcings applied to both the outer model and the nest, or the nest only?

Pg 428, line 28: "This requires some consideration on how to incorporate the feedback of the ocean state on the atmosphere forcing fields in an ocean-only simulation.

Many ocean only models try to deal with the effects of atmosphere-ocean feedbacks on the surface forcing used to drive the ocean model. In section 2 of your paper you describe how you have chosen to handle these feedbacks, but this is a topic that has undergone much prior research. Some discussion as to how others have handled these feedbacks is required.

Figure 2: The caption states "Model grids", but the grids are not shown, rather the bathymetry is shown (missing units on the figure). I would also like to see an explicit figure of the high-resolution region. Also, if OFFAM is a global model than where is the Arctic Ocean?

Pg. 430, line 15: I can appreciate that OFAM has been used previously by many to study Australian ocean region. However, a reader of this paper requires some evidence of the models skill in this area. Some discussion of the models strengths and weaknesses in simulating the Australian ocean is required and perhaps a Figure or two demonstrating the skill (e.g. surface KE and EKE, T/S comparisons with observations).

Pg. 430, line 25: If the model was only spun-up for 16yrs then why does the 1990-2010 forcing need to be looped? Table 1 states that the CTRL was equilibrated for 26 yrs?

Table 1: The table information is difficult to understand. For example “End of year 3, RELX” – I assume implies the model was initiated from year 3 of the RELX. But why not start FREE and STRS experiments from the same initial state as RELX (e.g the CTRL simulation)? It seems odd that FREE has the heat and freshwater feedbacks of RELX, which are then just turned off – and run for another 3 years. This makes it very difficult to explicitly determine the effects of the feedbacks, since they were used in the initial conditions. The experimental design here seems odd.

Pg. 431, line 5: ‘quasi-stable solution’. At this point in the paper there is no evidence that the simulations are stable or near any sort of equilibrium. Given that some of you simulations are run for only 3 or 7 years, I doubt they are stable or near any sort of equilibrium. I actually have never seen such short simulations discussed in a scientific study.

Pg 431, line 10: Are you certain that the T/S anomalies are purely the result of anthropogenic forcing or is drift in the MK3.5 simulation T/S fields also a part of the anomalies? Are the T/S anomalies applied everywhere (3-D) in OFAM? How big are the applied anomalies compared to the OFAM T/S? I would like to see a plot of surface T and S (or surface density) in OFAM, MK3.5 cntrl, and the applied 2060 anomalies.

Pg. 431, line 15: The reader needs to see the forcings applied and how big they are relative to the control forcings. Please provide a figure of the surface forcings in control and the anomalies.

Section 2.3: If changing the surface fluxes are the essence of the the downscaling methodology, then the methodology you are employing needs some justification for others to employ it in their work. Section 2.3 tells us what you did, but not why. It also fails to show us the size of the relative terms in Equations 1-3. The use of the correction terms (HC and FC) should be discussed more thoroughly. I also don’t understand why the need for a separate diurnal variability term – is this not present in the control forcing?

Section 2.4: It is difficult for the reader to discern the experimental design when reading the paper – the methodology is discussed in various different sections (e.g Sections 2 and 3). Table 1 should be improved to clarify the differences between the spinup, cntrl and forcing experiments. It needs to clearly explain the integration timescales, initial conditions, forcings etc., and one section in the paper should clearly outline the complete experimental design.

Section 2.5, line 1: ‘All simulations’ except for the AOGCM coarse model.

Section 2.5.1: Heat exchanges are a function of more than just SST. For example winds and various coefficients are involved in the process.

You should show equations for how the Heat and FW Flux feedbacks are calculated and justify this equation. As it stands now, the reader could not replicate or understand the reasons for your methods. As I understand it, you restore the SSS and SST to 'guide fields' that include the AOGCM T/S anomalies. I think this connection between restoring terms and feedbacks maybe novel, but am uncertain if it is a sound method. You need to justify it to the reader better. Sometimes in the paper you interchange the use of 'feedback' and 'restoring' (eg. Pg 434, line 13) and this is unconventional to me. Show us how the restoring is calculated (equation) and show how large the restoring surface fluxes are in comparison to the AOGCM surface flux anomalies. A 30 day restoring timescale can produce quite strong fluxes.

Figure 4: Please show the GCM-ODM difference for the control states in the upper right panel. It appears that in the control state the SST signature of the EAC boundary current transport is more pronounced (i.e. a larger southward extent) than in the ODM. How well is the EAC resolved in the ODM? It would be interesting to see the modelled SST and SSS compared to observations.

Pg. 437, Line 2; Figure 5: It is not stated which year the annual averages are taken from. If one assumes it is averaged over the last year of the simulations, than given that the simulations are all integrated for different time periods the anomalies in Figure 5 seem meaningless to me. It is also not clear that the anomalies are significant relative to the internal or interannual variability of the models.

Pg. 437, line 5: "The amplitude of the difference in climate change for SST, as measured by the standard deviation over the region shown in Fig. 5 …" Please clarify the meaning of this statement. Perhaps you determined the spatial standard deviation of the SST patterns for the middle column of Figure 5 and found that FREE was more spatially variable. I don't understand how that relates to the "the amplitude of the difference in climate change for SST". Please clarify.

Pg. 437, ln 12: "The retention of the warmer and cooler regions in the three different ODM experiments suggests that the structure of the difference in climate change for SST is largely independent of feedback at the surface. " While the spatial patterns of change are similar between the experiments – the large difference in the amplitude of the SST anomalies between the FREE and RELX experiments in Fig.5 suggests to me that the feedbacks or "restoring terms" are very important. Please clarify.

Section 3.2: the above 2 comments also apply to this section.

Section 3.2: "However, even with this large freshwater flux feedback the consistency in pattern of the difference in climate change SSSs suggests the SSS difference pattern is **not very sensitive to air-sea fluxes.**" Perhaps you mean to say that it is not very sensitive to your feedback-restoring methods. SSS patterns are certainly sensitive to air-sea fluxes.

Figure 7: The far right shows the AOGCM and ODM stratifications for the 1990s. Please show the Levitus evaluation mentioned in the text. Also given that the

simulations are run for a short time period (e.g. 3 years) the near surface isopycnals are likely to still be under sever adjustment. Are the patterns shown in Figure 7 the result of the T/S anomalies applied, or the surface flux anomalies, or simply the result of model drift at depth?

Figures 5-8: It is also hard to evaluate the Figures since the models have all been run for different lengths of time and they are initiated from different control states.

Figure 8: Please evaluate the ODM versus AOGCM transports relative to observations. Is the ODM doing a good job or not? Furthermore, as noted in the caption Figure 8 shows an average velocity over the upper 250m. This is not a transport.

Pg 440, line 7: "Also demonstrated is the high spatial variability of transport in the ODM indicating the need for averages over long time or large spatial domains to quantify results." I agree completely. So why does the paper evaluate a 1 year average from a 3 year model run compared to a 1 year average from a 7 year model run?

Figure 9: It seems odd to me that model drift is not discussed until this late in the paper, and it is done so in a fairly abstract manner. Why not simply show a 2D plot of the linear trends salinity or temp? After reading the text describing Figure 9 several times I am still unclear if the methods to diagnose drift are robust. It certainly is not a common way of evaluating the role of model drift. This analysis also only pertains to the RELX simulation. It is difficult for me to believe that the SST and SSS in the shorter STRS and FREE simulations are insignificant.

Figure 10: The EAC transports show large interannual variability in the simulations. Given that most of the figures compare 1 year averages taken at different years, it is difficult to determine the relative roles of internal variability versus the forcing response.