In the following, referee comments are in black, while our responses are in green and added material is indicated in blue.

Referee 1

General Comments:

The authors are the developers of the Earth system model UKESM1, which is participating in CMIP6. In this study, they focused on the ocean physics and biogeochemical fields of the historical experiments using UKESM1, and analyzed the results by comparing them with various observations and other CMIP6 ESMs. They have done a comprehensive and good analysis of the UKESM1 performance, which I think provides useful information for those who are planning to do multi-model analysis using CMIP6 ESMs and for other ESM developers. I have the following questions and comments, which I hope will help improve the manuscript.

We would like to first thank the referee for their diligent reading of what's a rather long manuscript, made much longer by its supplementary material. Thank you!

Specific Comments:

Line 140 "Evaluation uses the period 2000-2009 of the CMIP6 Historical simulation and compares to corresponding periods of observational data." The analysis mainly uses the results of simulations and observations for the period 2000-2009. This period is known for its negative IPO index, so the observations include that signal. On the other hand, since the model takes 9 ensemble members, the effect of internal variability is expected to be negligible. This may contribute to the equatorial Pacific warm bias of UKESM1 (see Section 3.1), although it may be small. If possible, it would be better to extend the time period used in the analysis to 20 or 30 years. Even if this is difficult, it would be good to take a longer-term look at SST to confirm that the model bias of UKESM1 does not depend on how the average period was taken.

This is an excellent point. Our analysis focuses on an standardised recent period and does not consider the impact of variability, including important climatic modes. However, this point is worth mentioning, and we have added a further supplementary figure (S4) and the text below to address this.

"SST exhibits a number of major climate modes such as the Interdecadal Pacific Oscillation (IPO) and Atlantic Multidecadal Oscillation (AMO) that can introduce persistent and large-scale shifts in temperature that are of comparable magnitude to the model biases identified above. For instance, the IPO has a negative index (cooler than reference) during the time period shown in Figure 1, but a positive index (warmer than reference) during the preceding two decades (Salinger et al., 2001; Hu et al., 2018). Models also have climate modes, but these can be out of phase with those observed, and they may occlude or exaggerate biases. Supplementary Figure S4 partially addresses this by repeating the difference plot from Figure 1, but for the three preceding decades. The resulting patterns of model-observation difference are generally consistent between the decades and for both seasons, suggesting that they represent model biases rather than variability mismatch. In particular, persistent features include the strong cold bias in the western North Atlantic, warm biases in the equatorial Atlantic and Pacific basins (the latter seasonally), and a general warm bias in the Southern Ocean. As most other observational datasets used in the evaluation of UKESM1 properties are more restricted in the time periods they have available, similar analyses are more difficult. However, given the primary role of SST in many ocean processes, the apparent dominance of model bias in SST over its temporal

variability is suggestive that mismatches in major climatic modes is of secondary importance in our analysis."

Line 189 "In the Arctic, sea-ice is typically multi-year, and this positive bias in modelled area is accompanied by excessively thick sea-ice." I think that this sentence is confusing and needs to be rewritten.

We have amended this as follows:

"In the Arctic, sea-ice typically persists for multi-year periods, such that this bias towards excess ice area in UKESM1 is accompanied by sea-ice cover that is also excessively thick."

Line 230 "This is more pronounced in the Atlantic basin, in particular at tropical latitudes, where midwater (100-1000 m) biases up to 4 C are found in the model" I think that this is influenced by the high temperature bias in the formation region of the NADW. If this is the case, why don't we use Figure 1 to discuss the temperature bias in the deep layers in relation to the formation process of the NADW?

The largest temperature biases (4°C) are more associated with northward AAIW than southward NADW. This is perhaps clearer in the nitrogen nutrient figure where there is a distinct arm of high nutrient waters moving northward in the 100-1000 m region, and which appears associated with this positive temperature bias. Nonetheless, the warm bias in NADW appears associated with a corresponding bias in SST in the subpolar North Atlantic, so the following sentence has been added:

"The bias in southward-moving NADW (> 1000 m) is consistent with the warm bias in SST shown in its subpolar source regions in Figure 1."

Line 233 "These show the model ocean, particularly the Pacific basin, to be more stratified vertically compared to observations, with generally lower density surface waters (< 1000 m) overlying more dense deep waters." This may be due to inadequate and overall small parameterization of vertical mixing, the mechanism that transfers heat from the surface to the deeper layers. In fact, the bottom layer circulation is weakened accordingly (Figure 8). It would not be a bad idea to point out that the parameterization of vertical mixing is insufficient.

This clarification would be useful. We have added the following text to the end of this paragraph:

"This bias suggests that the model's parameterisation of vertical mixing may be insufficient, reducing the transfer of heat from the surface to deeper layers (and potentially weakening the deeper circulation; see below)."

Line 249 "Strong sinking around Antarctica, combined with a slightly weaker NADW, is consistent with the colder and fresher conditions shown for the deep ocean (particularly the Atlantic) in Figures 6 and 7," It is difficult to understand why "consistent" is used, so please rewrite it in more detail.

We would agree that the text is unclear on its use of "consistent". We have amended as below to more clearly note that the deep biases mentioned are consistent with a more dominant role for AABW in UKESM1:

"Stronger sinking in UKESM1 around Antarctica, combined with a slightly weaker NADW than observed, indicates a more dominant role for AABW in the model, and is consistent with the colder and fresher biases found in the deep ocean (particularly the Atlantic) in Figures 6 and 7, as well as biases in biogeochemical fields (see later)."

Line 409 "CFC-11" It is difficult to follow the discussion because of the sudden appearance of CFC-11 here; the description of CFC-11 in Figure 18 appears later in Line 421, so please reconsider the structure of this section.

We agree that this is an omission. We have added the following text to the model description section to clarify CFC-11's role:

"In addition to the biogeochemical tracers of MEDUSA-2.1, UKESM1 includes the chlorofluorocarbon tracer, CFC-11. This artificial tracer has an atmospheric time-history analogous to that of anthropogenic CO2, and can be used as a marker for recently ventilated watermasses (Key et al, 2004). It can be measured from seawater samples with high accuracy, and provides an additional measure here for evaluating simulated circulation."

Line 584 "the weaker deep overturning cell north of the ACC combined with a NADW cell which is slightly too weak (Figure 8) leads to a colder and fresher deep ocean" I would like to see a more detailed description of the mechanism of why the weaker deep overturning cell north of the ACC and the slightly too weak NADW cell lead to a colder and fresher deep ocean.

Yes, the reference to NADW is confusing here. We have rewritten this to reflect the clearer point that we intended to make, namely that the sluggish AABW cell in the Atlantic contributes to its positive nutrient and negative oxygen biases:

"For instance, the weak deep overturning AABW cell north of the ACC (Figure 8) reduces the ventilation rate of the abyssal Atlantic, and contributes to the build-up of nutrients and the corresponding depletion of oxygen."

Section 4.2 I would like to see a detailed description of the improvements in the model from the predecessor HadGEM2-ES, and what could be fixed to make it better. The distribution of silicic acid concentration seems to have improved significantly. (Figure S22)

We would agree that this is an omission in the manuscript. The relationship between HadGEM2-ES and UKESM1, and in particular their ocean components, is not made clear in the text. To address this, we have added the paragraph below to the model description section:

"UKESM1 is the successor model to its CMIP5 predecessor, HadGEM2-ES (Collins et al., 2011). Many of its components are evolved versions of those in the earlier model, including its land surface, physical atmospheric core, and atmospheric chemistry components (Sellar et al., 2019). However, in the specific case of the ocean in UKESM1, its dynamical core, grid domain, sea-ice, and marine biogeochemistry are wholly new and replace the corresponding components in HadGEM2-ES. Consequently, there is no direct traceability between the oceans of the two generations of CMIP model. Nonetheless, as part of the assessment of UKESM1, elements of its performance relative to that of HadGEM2-ES are examined in Section 4.2."

And, before the detailed description of the figure (Figure 23) in the main text, there is a detailed description of the supplement figures (Figures S21-S23), which seems odd to me, so please reconsider the structure of this section.

To avoid overcrowding the main body of the manuscript with figures, we decided to select a single example from the set of primary biogeochemical properties examined. These properties are introduced and discussed in the same order as earlier in the manuscript (hence the figure ordering), but we elected to choose primary production for the main body because of its high diversity in model representation.

Technical Corrections:

Line 6 “a new Earth system (ESM)” should be “a new Earth system model (ESM)”.

Amended as suggested.

Line 175 “% citepmarzocchi2015” typo.

Thanks for spotting this. Corrected.

Line 178 Typo. The umlaut is attached to the m.

Thanks for spotting this. Corrected.

Line 196 ”Much as with sea-ice extent itself, UKESM1 performs best in the Arctic,” I don’t think "best" is the right word.

We would agree. Changed "best" for "better".

Line 295 “Table 2” I think this should be “Table 3”. If we do that, then Table 2 will not be mentioned in the main text.

Thanks for spotting this. We have corrected this and have expanded the description of Tables 2 and 3.

Line 382 “the SO” The abbreviation "SO" is not used elsewhere except Line 499, so it should not be used here as well.

Thanks. Abbreviation removed.

Line 499 “SO” See the above comments on Line 382.

As above.

Line 727 “the other three cycles” C, Fe, and O2? It’s difficult to tell what cycles you are referring to, so please specify C, Fe, and O2.

Yes, this is correct. We have amended the text to:

“The model’s nitrogen, silicon and alkalinity cycles are closed and conservative (e.g. no riverine inputs), while the cycles of iron, carbon and oxygen are open.”

Line 759 “MOCSY-2.0 Orr and Epitalon (2015)” This should be ”MOCSY-2.0 (Orr and Epitalon, 2015)”

Thanks. Amended.

Figures

Please add labels to each panel in each figure.

We tried this, but couldn't get a consistent "finish" to the resulting Figure panels due to figures being a mix of single and multiple panels. However, we have implemented other suggestions around figures, and hope these make up for this omission. From the captions, it should be clear in all cases which panels are which.

Figure 1 It is easier to understand if observed, simulated, and differences are arranged in a single column or row.

We've rearranged figures so that observations, model and their delta are arrange in single columns corresponding to the two seasons shown. Other figures which don't have a delta have also been altered for consistency.

Figure 2 A color palette should be created that is white at 0.15 or less.

This is a good suggestion – we have amended the figure so that the colourscale shows white for values < 0.15.

Figure 4 It is easier to understand if observed, simulated, and differences are arranged in a single column or row. There is no explanation of the differences figures in the caption.

See previous response.

Figure 8 Contour interval should be noted in the caption.

Contours are every 2 Sv. This has been added to the caption as directed.

Figure 12 m-2 should be m^{-2} (superscript).

Fixed - thanks!

Figures 13 and 14 These include Hovmoeller diagrams, but aren't they unnecessary? There is no detailed description of the monthly variation in the main text.

The Hovmoller diagrams are not referred to specifically, but the description of both figures in the main text does discuss the seasonality of the properties, and the associated model biases. As such we have retained them. Also, the chlorophyll plot is not very flattering to our model, and deleting it feels like we'd be hiding poor results.

Figure 19 The first time it is mentioned in the main text is probably in Line 464. This is later than Figures 20, 21, and 22. The numbering of the figures should be in the order in which they are mentioned.

Thanks for spotting this. The mention of the figure got moved in manuscript drafting, but the figure itself did not. We have amended this.

Tables

Various values are listed in the tables, but in the main text, only AMOC in Table 1 (Line 268) and silicic acid in Table 2 (3?) (Line 295) seem to be mentioned. It is good to include various values in the tables, but the text should be revised so that it is not assumed that only a small part of the tables is covered in the main text.

We have added a few more references to Tables 1-2 in the main text so that readers are more clearly directed to the variability and trends that they report. Including:

"Table 1 lists the global means (or mean integrals) of these surface physical properties across both the full Historical period and the corresponding piControl period. For both of these simulation ensembles, the variability and ranges of each of these properties are given, together with the simple linear trend over the full 165 y period."

"The influx of CO2 into the surface ocean is also documented in Table 2's mean and trend statistics of surface DIC and air-sea flux, in particular how they compare with the corresponding piControl period."

Supplement Figures The numbering of the supplement figures should be in the order in which they are mentioned in the main text. Please check.

As well as Supplementary Figures being referred to out of order, we found two figures that were no longer mentioned in the text. These have been removed, and the order of Supplementary Figures has been amended to follow the order in which they are mentioned in the main text.

Figure S3 There is no explanation of the differences figure in the caption.

We have added reference to this in the caption.

Figure S4 It is easier to understand if observed, simulated, and differences are arranged in a single column or row. This figure is not mentioned in the main text? If not, please remove it.

Thanks for spotting this omission. Figure corrected and missing text on SSS added to the main text:

“Supplementary Figure S4 parallels Figure 1, showing the observed (WOA, 2013) and simulated sea surface salinity (SSS) for summer and winter, together with (model - observed) differences. UKESM1 shows a general negative bias in SSS (~1 PSU), but with significant regions of positive bias in the tropical Atlantic and Indian oceans (< 1 PSU). There are also “hotspots” of bias in the Bay of Bengal (positive), off the west (negative) and east (positive) coastline of equatorial South America, in the Yellow and East China seas (negative), and in the Arctic (both positive and negative). These regions are mostly located close to major riverine inputs, and likely reflect model inaccuracies in the precise location and magnitude of associated freshwater additions.”

Figure S6 ”Potential density anomaly in kg m 3 (minus 1000 kg m 3).” This description is confusing and needs to be rewritten.

We would agree. We have reworded this as:

"Potential density is shown as kg m-3 minus 1000 kg m-3 (i.e. the actual density range in the upper panels is 1025 to 1028 kg m-3)."

Figures S18 It would be good to add observed limiting nutrients (e.g., Moor et al., 2013, Table S2, DOI:10.1038/NGEO1765) for comparison.

We have amended to include the following text around nutrient limitation. However, the lack of biogeochemical detail in our model precludes a thorough comparison, and we have noted this.

“Corresponding observational patterns of nutrient stress are more sparsely available (Moore et al., 2013). However, UKESM1's nutrient limitation overlaps the major observed patterns, including widespread nitrogen stress in the Atlantic Ocean, and iron stress throughout the Pacific and Southern oceans, as well as at high latitudes in the North Atlantic (Moore et al., 2013). Nonetheless, the simplicity of MEDUSA prevents it from representing the limitation of phytoplankton found by Moore et al. (2013) for the macronutrient, phosphorus, and the micronutrients, cobalt, zinc and vitamin B12.”