General comments:

This is a very interesting paper addressing the skill assessment of the newly coupled REcoM2-FESOM1.3 model. The coupling of a global biogeochemical model to a finite element ocean model is to my knowledge a scientific novelty and therefore of high scientific significance. The methods utilized for the skill assessment are valid and adequate and the assessment is carefully conducted. Given the new usage of a finite element ocean mode it would be beneficial to explain in more detail how the finite element ocean model improves the skills of the model. Otherwise the results are presented in a comprehensive and clear manner. I support publication of this paper in GMD, subject to minor revisions.

Specific comments:

-The article places its focus on the Southern Ocean since the coupled model FESOM-REcoM2 is meant to simulate biogeochemical processes in the Southern Ocean (p. 4168, line 22-24). The authors should explain why a global model, especially this model, is appropriate for usage in the Southern Ocean. Is the model MITgcm-REcoM1/2 performing especially well in the Southern Ocean? Is FESOM expected to perform especially well in the Southern Ocean? It would be advantageous if the Authors would describe the reason for the model´s focus on the Southern Ocean in more detail.

-The oceanic regions utilized for the model-observation comparison (as defined in Figure 3) don´t consider any region north of 70N even though there is data available (see Figures 5, 7, 8, and 11). Is there any specific reason for leaving this area out? Please add an explanation (I guess the data coverage is too sparse for the evaluation of seasonal cycles. Does that mean the data north of 70N is also left out of the Taylor diagram for spatial-seasonal distributions (Figure 4b)?)

- The authors compare the model output against data products for NPP, EP, chlorophyll, MLD, and nutrients (Fe, DIN, DSi). Still it looks like different comparison-methods where used for different variables and it is not obvious why that is the case. I think the authors should explain why Fe is the only variable missing in the Taylor-diagram (Figure 4) and why there are comparisons of spatial distributions of MLD (Figure 5), DIN (Figure 7), DSI (Figure 8), chlorophyll (Figure 11), NPP (Figure 12), and EP (Figure 14), but not for Fe. Also seasonal cycles are only presented for MLD (Figure 6), and NPP (Figure 13). Probably these choices are dependent on the scarcity of the data, but it would be beneficial to add explanations for these choices.

- The text shows very clearly that the availability of Fe-data is very low (globally and specifically in the Southern Ocean). The authors put a lot of effort in comparing the available Fe-data with the model. Even though the model reproduces the spatial distribution of iron reasonably well, its iron concentrations are clearly too low. The

latter is discussed in detail, comprising different explanations and possible solutions of the problem. In order to complete the discussion it would be beneficial to see how good the model performs in comparison with other models in terms of iron concentration.

- The discussion focuses mainly on the Fe-distribution as the main deficiency of the model. I think that should be clearly stated at the beginning of the discussion. Also the first part of the discussion doesn't follow the rest of the text and should be placed later, where the results of FESOM-Recom are discussed.

-Several passages claim that normalized standard deviations higher/lower than 1 (model output) indicate that the concentration gradients are too steep/low when compared to the observations. While this is in general true, it is not possible to confirm this finding with model-observation differences since the standard deviation is a measure with respect to the mean value. Therefore a confirmation with model-observation differences would only be possible when considering anomalies (with the mean value as reference value). Please correct these statements.

-When referring to the NPP data product, the authors refer to Behrenfeld and Falkowski, 1997. To my knowledge, Behrenfeld and Falkowski did not provide a NPP data product, but a model (vertically Generalized Production Model – VGPM) for the estimation of NPP. The latter relies on different variables which can be provided for example by a satellite-product such as SeaWiFs. In order to be accurate it would be good to refer to the data as "*Net Primary Production estimated with SeaWIFS chlorophyll and the VGPM-model*" or something similar. The authors should also clarify if they calculated the data-product themselves or retrieved it from a research group or a website. The same holds probably for EP.

- p. 4162, line 7: the MLD-criterion of de Boyer Montegut is cited incorrectly. The mixed layer depth is defined as the *first* depth at which the difference between the potential density at 10m depth and *the potential density at deeper lying reference levels is greater than* 0.03 kg m-3.

 -p. 4165, line 9: please mention the values provided by Schneider et al.

 -p. 4166, line 46: the text claims that the seasonal cycle is closest to the satellite-based estimate between 10-45N and S. Figure 13 indicates that this statement is not correct for the North Indian region. Please correct the statement accordingly.

-Table 3: The table should be expanded to be more consistent with the text. Please include the values of Behrenfeld and Falkowski 1997 as well as Schneider 2008 for global NPP, Siegel 2014 for global EP, and Carr 2006 for NPP in the Southern Ocean.

-Conclusions: When declaring that the modeled spatial fields are on average better in the Southern Ocean than on the global scale, the authors should state which variables are performing better (according to Figure 15 this seems to be the case for

DIN and Si) and which variables are performing worse (according to Figure 15 this seems to be the case for MLD and Chl).

Technical comments:

-Figures 4, 6, and 13: please enlarge the labeling. Especially the labeling of Figures 6 and 13 is almost unreadable

-Figures 10, 17, and 18: please mention in the figure-captions that the illustrated variables are modeled values

- p. 4156, line 16: please add "for explicit scientific studies" or something equivalent after "Before using a newly coupled biogeochemical-ocean model" (since the model was already used for the model-data comparison)

- p. 4157, line 18: replace "grid points" with "levels" or "layers"

- p. 4157, lines 21-22: please add references to the terms "Redi diffusion" and "Gent and McWilliams parameterization"

- p. 4159, line 26: please add "an" after "gives"

- p. 4167, line 14: please confine the statement to "The EP of the model".

-p. 4168, line 12: consider replacing "the differences between the fields are especially clear" with "the differences between the fields are especially visible"

 -p. 4171, line 24: please correct the sentence to "We will now examine the roles of MLD and iron concentration in explaining the seasonal variability of NPP".