Reply to anonymous Referee #1

Referee comment: *What are the sources for the forcings, the boundary and the initial conditions (both physics and biogeochemistry? (page 8404)*
Author's response: We agree that this information should be added.
Changes to the manuscript: This information has been added to section 2.2 Experimental setup.

Referee comment: *Model has been spun up for three years. Is there any reference to show that this time is enough, particularly for the deeper part of the domain? Or did the authors make some test to check this? (page 8405)*
Author's response: This was tested in an earlier version of the model and this is described in Hansen (2008), where it was shown that the model did not have any significant drift beyond the first three years. Reference: Hansen, C. (2008): Simulated primary production in the Norwegian Sea – Interannual variability and impact of mesoscale activity, PhD Thesis.
Changes to the manuscript: The reference to Hansen (2008) has been included in the text.

Referee comment: *After equation 3 (page 8404) authors write the value for "g" for Meso-zooplankton. Although the equivalent reference value for microzooplankton is reported in table 2, I would recommend the authors to report the standard value here as well for clarity*
Author's response: We agree with the referee
Changes to the manuscript: The value for microzooplankton is provided in the text as well.

Referee comment: *Authors did not explain the meaning of "mu_z" in equation 5. I assume this is the maximum mortality rate. If this is the case, I suggest authors to use the latin letter "m" instead of the greek letter "mu" because authors (in agreement with most literature) already use the greek letter "mu" for growth of the phytoplankton (equation 1 and 2) therefore it is misleading having the same symbol representing growth in one equation and mortality in another one.*
Author's response: Here we followed the symbols used in the ECOHAM user guide, but we agree that it does make more sense to use m for mortality,
Changes to the manuscript: The symbol has been changed to m in the equation and table 2 and defined in the text below eq. 5 as well.

Referee comment: *It is unfortunate that authors do not take full advantage of the spatial resolution of the model and the data and presents results lumped for the whole domain or at most for just 2 areas. I acknowledge that spatial coverage of the data can be limited particularly in winter, but figure 3 highlights that spatial pattern of uncertainty could be investigated in a more detailed way than the one authors already discussed contrasting two subdomains (NWS and BAS). Furthermore authors describe different performance of the model in the deeper domain (below 500m) compared to the upper domain (page 8407): nevertheless figures 4 and 5 show only the synthesis of the model-data comparison across the whole domain. I suggest authors to show also the outputs from upper and bottom domain separately, particularly because authors state "[Upper] Silicate has no skill in the years 1999 and 2000." while figure 4 shows good to very good skill for silicate in those years. This leads to think that the good results for nutrient simulation highlighted in figures 4 and 5 could be biased by good initialisation of the model in the deep basins where the dynamics are limited in a 10 years period. I would also suggest authors to mention the residence time of the basin, in order to give the opportunity to readers that are not expert of the area (like myself) an idea of the relative importance of endogenous dynamics versus boundary forcings. Similarly, authors show vertical profiles for Chl and nitrate, but they discuss also phosphate and silicate. I suggest authors to add similar plot to figures 9 and 10 for P and Si (perhaps as supplementary information or removing the June panel from figure 9 and 10 to limit the number of figures).*

Author's response: Thank you for this suggestion, the spatial component is indeed very interesting and should be taken greater advantage of. And the paragraph on page 8407 was confusing and in the last sentence about silicate it should have explained that the skill for silicate is lower in 1999 and 2000 compared the year before and after. We have removed this sentence in the new version.
Changes to the manuscript: As a result of your suggestion we changed figure 6 and 7 (figure 7 and 8 in the new manuscript) to be spatial maps of bias and model efficiency respectively. We chose to include only the upper 100 meters in these figures. For nutrients this is where we see the improvement and although the bias is better at depth, we do not see much difference between the model performances for different parameters (see figures R1 and R2 below). Figures R1 and R2 are for the depths 100 to 500 meters, for the depth 500 to 1500 meters there was almost no differences between the runs. Part of the good results at depth is probably due to good initial conditions and relaxation at the boundary. Paragraph 3.1 has been modified for more clarity and the information about residence time has been included. In addition Figures 9 & 10 has been updated to include phosphate and nitrate as the reviewer suggested. Here we also found an error in the code for the plots (the previous plots showed accumulated nutrients from the surface to depth, not nutrient concentration in a certain depth interval), we have corrected that as well, the correction resulted in less smooth curves than in the previous figures, it also becomes clear that there are no clear improvements at depth, so those claims have been removed from the manuscript.

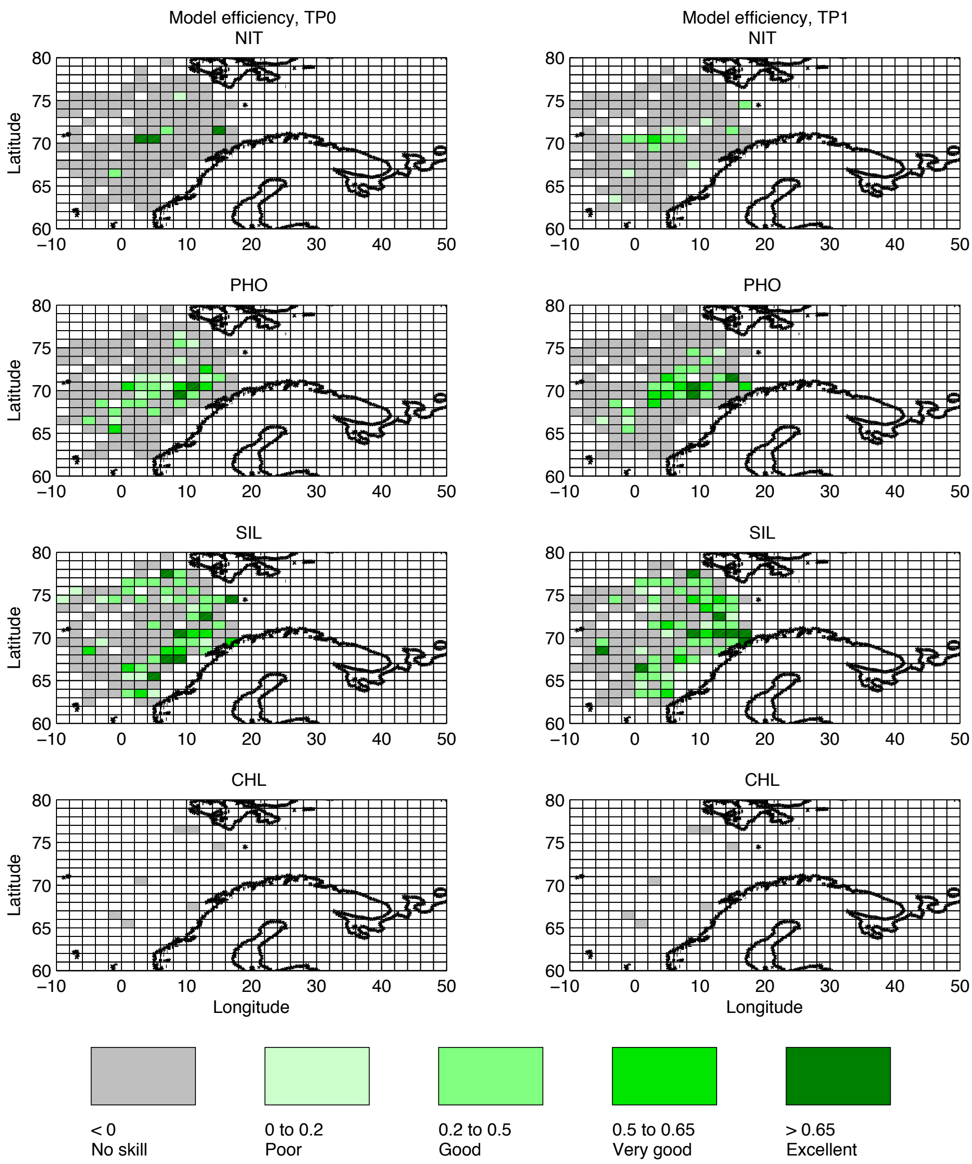

Figure R1. Percentage bias in the 100-500 meters intervall for the model simulations compared to all available observations from the period 1998-2001 in 2x1 degree boxes from the simulations with the fine-scale model with the original (TP0) and final set of parameters (TP1).

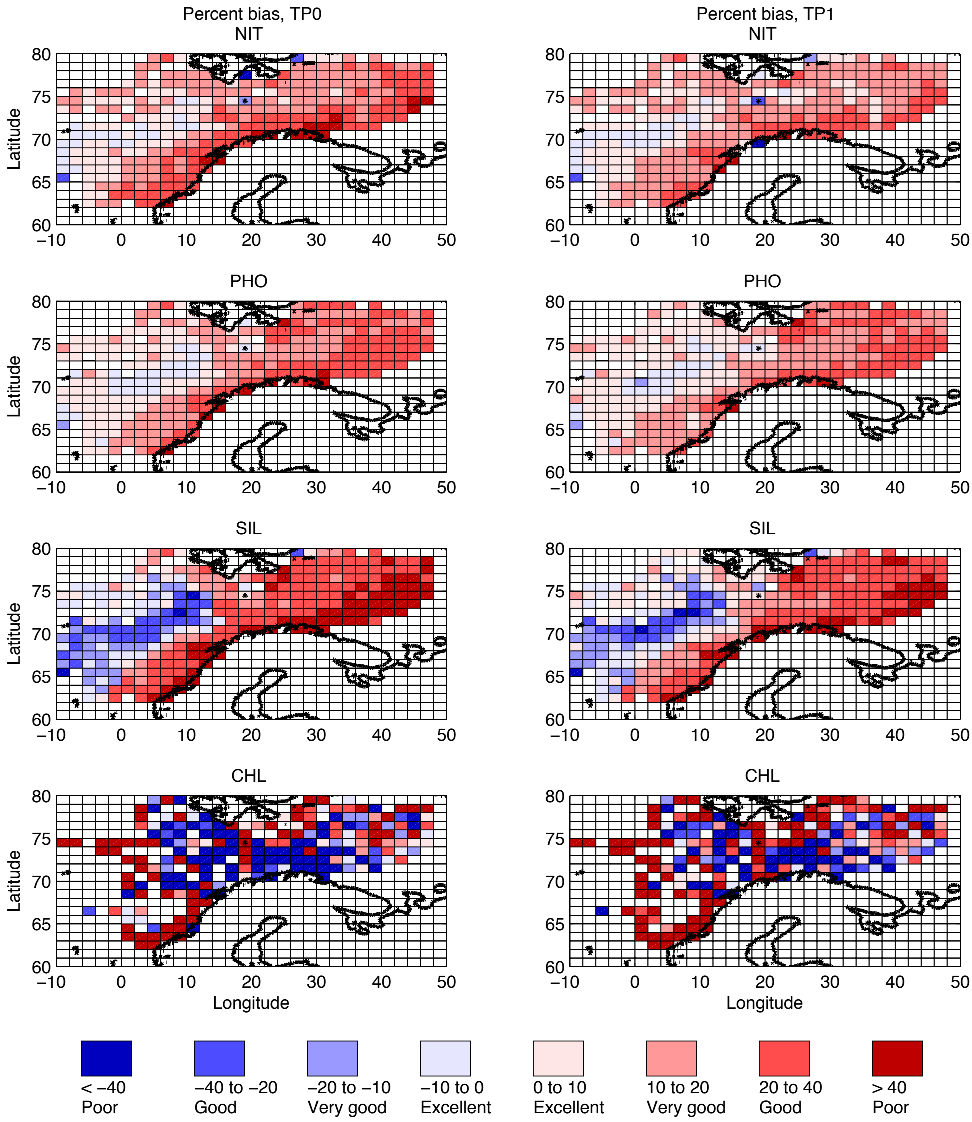

Figure R2. Model efficiency in the 100-500 meters intervall for the model simulations compared to all available observations from the period 1998-2001 in 2x1 degree boxes from the simulations with the fine-scale model with the original (TP0) and final set of parameters (TP1)

Referee comment: In the discussion section, the main items are discussed, but I would suggest authors some changes in order to make the section more clear: The introductory paragraph is redundant, as it simply summarizes the entire workflow and this is already well clear from the previous sections. This section looks more like a conclusion than an opening of discussion.
Author's response: We agree that this paragraph is redundant.
Changes to the manuscript: This paragraph has been removed.

Referee comment: The title of section 4.1 is misleading. In the paper the model is not validated, as this would require to compare the model output with a completely independent dataset from the one used for calibration/tuning. From my understanding of section 2.3 all data have been used for tuning therefore the model has not been validated. I would suggest to title this section "uncertainty connected to observation" that is an appropriate title for the discussion in this section. I would also suggest authors to refer to Stow et al., 2009 (http://dx.doi.org/10.1016/j.jmarsys.2008.03.011) for comprehensive analysis of this topic.
Author's response: What the referee points out is correct; the same dataset has been used both for tuning and comparison of the model results.
Changes to the manuscript: The title has been changed to "Uncertainties connected to observations", we think this more accurately describes this section. We have

also added a sentence about the same dataset being used for both tuning and validation and referred to Stow et al, 2009.

Referee comment: *The first sentence in section 4.1 is arguable. The quality of measures does not depend (only) on their abundance: a broken thermometer will always give the wrong temperature. I suggest authors to reformulate this sentence and in particular to be more clear with the meaning of "quality of measure" for them.*
Author's response: The referee has a good point.
Changes to the manuscript: 'Quality of the measurements' was changed to 'representativity of the measurements'

Referee comment: *Authors state (page 8412) that fluorometer Chl-a may vary with a factor of 3-4 compared to HPLC Chl-a. I may agree with this, but I recommend authors to clarify if they refer to in situ fluorometer Chl-a or Chl-a measured with a fluorometer in the lab from extracted pigments. In the latter case the error is expected to be much lower than the one suggested by author. A reference to back up their estimate would be needed as well.*
Author's response: We did mean fluorometer measurements from the field.
Changes to the manuscript: It has been specified that we mean fluorometer measurements from the field in the text.

Referee comment: *In section 4.2 authors state that changes in mortality of zooplankton produced little effect, contrarily to the expectation but they did not provide any potential reason for this.*
Author's response: Without further analysis, an answer to this will be quite speculative, but given that we may be allowed to speculate: For example an increase in zooplankton mortality would decrease the population of zooplankton and decrease the grazing mortality of phytoplankton. This would again increase the phytoplankton population and in turn increase nutrient consumption and decrease nutrients. Here there are two potential feedback loops: increased phytoplankton will be available as food for the remaining zooplankton, increasing their growth or because a large part of dead zooplankton will sink out of the system as detritus, the decreased rate of remineralization from zooplankton would decrease nutrient concentrations and limit the growth of phytoplankton. The opposite would be the case for a decreased zooplankton mortality. In order to pin down the exact reason we would have to rerun the model with different zooplankton mortality parameters and output all rates of transfer between different variables for these runs.
Changes to the manuscript: We feel that this is speculative and prefer not to write a possible reason into the paper, but rather say that we are not sure about the reason.

Referee comment: *In the same section, authors rightly interpret the lack of effect of the change in N:Chl ratio on the model performance on simulating Chl with a compensatory mechanism. This mechanism should lead to a different distribution of phytoplankton along the water column. I suggest authors to bring this evidence to corroborate their hypotheses and to discuss the potential consequences*
Author's response: Looking closer into these results we realized that this is the result of a mistake in the processing of the model results, and in retrospect we realize that we should have been more suspicious towards these results, while the nutrients are not very sensitive to the change, the change in chlorophyll is actually quite big. Our theory about different vertical distribution of phytoplankton as a result of altering this parameter was still correct. However, the effect was small, changing concentration of phytoplankton (expressed in units of mg N/m^3) only by about 3-5% difference between the runs with ratios 13.7 and 6.3, compared to the effect on chlorophyll concentration (or phytoplankton expressed in units of mg Chl/m^3) in two runs that was about double in the run with ration 6.3 compared to 13.7 (N10 and N08).

Changes to the manuscript: New values have been added in figure 4 and 5 and the text have been modified to take into account the new results. We have not written that there was a mistake in the discussion paper thinking that it would only be confusing to the reader that has not read the discussion paper.

Referee comment: I totally agree on the limitation due to computational constraint highlighted in the last paragraph of section 4.2. Authors could state while they chose to run the sensitivity test on using the entire 3D model instead of running those in faster 1D set-up (maybe in contrasting environment in the domain), particularly since authors do not show spatial pattern of sensitivity (by the way, this information could be really informative and would increase the impact of the paper).

Author's response: We do have a 1-D version with GOTM coupled to NORWECOM, but the GOTM-NORWECOM gives quite different results from the full 3-D model, it does not overestimate the magnitude of the spring bloom and the duration of the bloom is much shorter, it underestimates the chlorophyll concentration during July when the model presented here overestimates it. The GOTM-NORWECOM model could probably have been used for a parameter sensitivity analysis, but for the sake of the tuning we found that the results were too different for it to be useful.

Referee comment: In few occasions (e.g. beginning of page 8415), errors in the simulation of physics have been used to explain errors in the biogeochemistry. The explanation given are perfectly reasonable, but the general performance of the physical model has not been shown, nor adequate reference has been given in support of authors' hypotheses.

Author's response: Our colleagues have compared the model to the hydrography from the Svinøy section (and other regular sections) we see that the model rarely places the fronts in the correct position. In the case of the Svinøy section, which is upstream of station M, the model places Atlantic water – often defined as having salinity greater than 35 - too far to the west (Figure R3), but unfortunately there are no publications where these results are shown.

Changes to the manuscript: Since we do not have any publications showing the (mis)placement of fronts, but we know his is true, we added 'not shown' to the manuscript. With regards to a late development of MLD leading to a late spring bloom, we have used Samuelsen et al. (2009) as a reference.

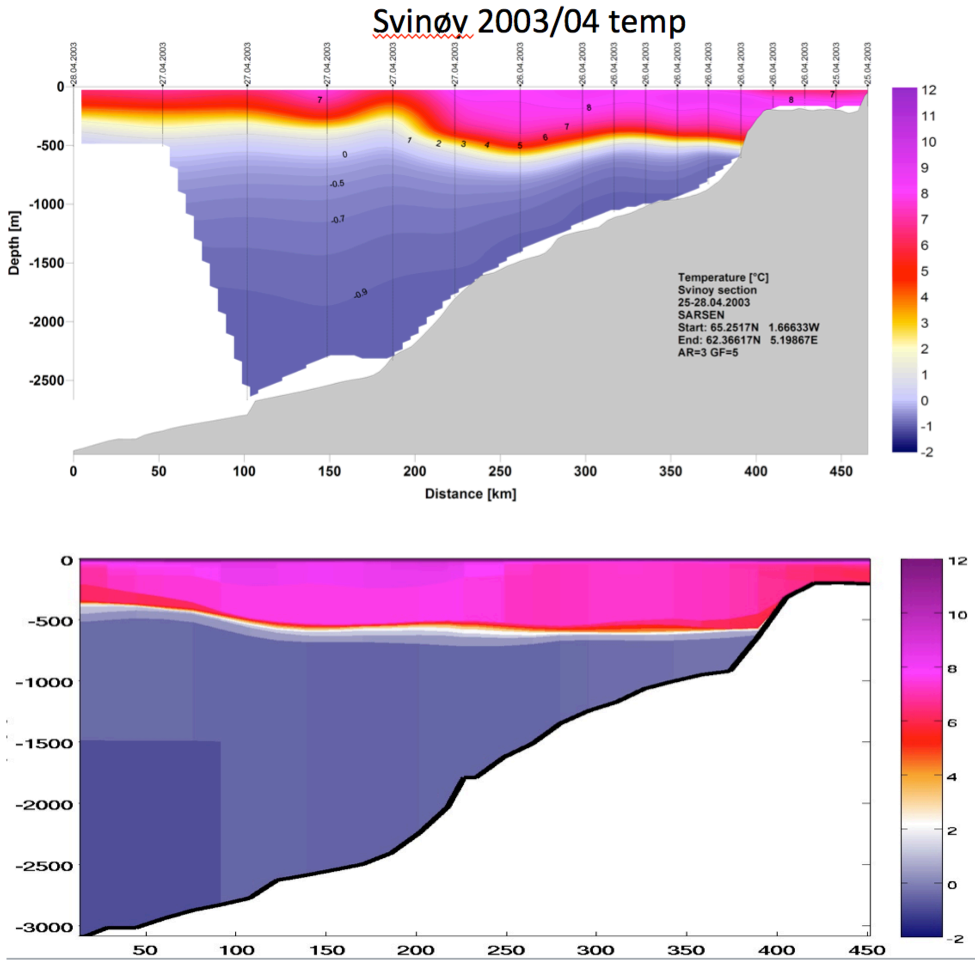

Figure R3. Salinity section from the Svinøy-section from observations (above) and the model referred to as high-resolution in this paper (below).

<u>Referee comment</u>: *Similarly, bad simulation of bloom initiation has been suggested as potential error in Chl-a simulation, however comparisons between bloom initiation timing (model vs. data or model vs. model during the sensitivity test) have not been provided. Such a way, these statements remains quite speculative.*
<u>Author's response</u>: We have looked at the timing of the bloom in this and previous studies, and none of our efforts to adjust the timing have not been successful in mowing it more than 3-5 days back or forth.
<u>Changes to the manuscript</u>: 'not shown' has been added behind the sentence "The model is consistently late in its initiation time and none of the parameter alterations significantly affected the timing of the spring bloom" to indicate that we have actually looked at this. When the same thing is mentioned earlier in the manuscript, the reference to the paper Samuelsen et al. 2009 has been included. This paper shows the timing issue quite clearly in figure 3.

<u>Referee comment</u> *From line 15 of page 8415 authors do not discuss regional differences in performance but they discuss the general performance of the model, therefore this part should go under a different header (either a 4.4 header or a generic Conclusion).*
<u>Author's response</u>: This referee is right.
<u>Changes to the manuscript</u>: This part of the manuscript is now under the heading '5. Conclusions'

<u>Referee comment</u> *Finally, but I appreciate that this is a personal opinion, I would remove any dubitative form when authors states that model could be improved in closer collaboration with empiricists. I believe that this is the way forward without any doubt if modellers want to build reliable model that describe the main ecological principle and pathway and are up-to-date to the more recent*

understanding of marine ecosystems.
Author's response: We agree with the referee.
Changes to the manuscript: The word 'perhaps' has been removed
-
Further minor corrections suggested:
Referee comments
1.Page 8401, l 24-26: For clarity I suggest to write: "The HYCOM-NORWECOM model was tested against local in-situ data and derived gridded climatology of nutrients, as well as satellite data, however. . ."
2. Page 8402, l10: add a comma between salinity and temperature
3.Page 8402 l24: I'm not native English, however I believe that "provide" is a better word than "proved" in this context
4.Page 8406, L20: the standard Taylor diagram show standard deviation, correlation coefficient and centered RMS not variance (see Taylor, 2001 figure 2)
Changes to the manuscript: The text has been changed according to the reviewer suggested.

Referee comment: Page 8410, l21: "profiles in the upper 1000m of the water column IN THE NORWEGIAN BOX. . ."
Changes to the manuscript: We added 'in the Norwegian Sea box' in this sentence

Referee comment: Page 8411, l22: I would rewrite the sentence starting with "It is howewver.." like this: "different requirements for geographical coverage, number of stations and frequency are needed depending on the different issues addressed, parameters measured and the area complexity (e.g. Ottersen et al., 1998)"
Changes to the manuscript: The sentence now reads: "Depending on the issues addressed, there will be different requirements for geographical coverage, number of stations, frequency and parameters measured (e.g. Ottersen et al., 1998)."

Referee comment: Table 2: I believe that there is a typo for case N12: it should be maximum microzooplankton grazing rate, and not grazing preferences for microzooplankton
Changes to the manuscript: Yes, this was a typo, it has been corrected.

Referee comment: Table 3: this could be moved in the supplementary information
Changes to the manuscript: The other reviewer suggested to move it to an appendix, after reading the journals definition of supplementary material, we found that appendix was more fitting, so it has been moved to an appendix.

Referee comment: Figures 4,5,6,7: I really like the colour coded system, as it is really communicative and easy to understand. However, at the same time it can be also misleading: e.g., in figure 4 a big improvement from 0.2 to 0.49 will not be highlighted at all, whilst a small improvement from 0.49 to 0.51 will stand out. I would suggest authors to write also the value of the different metrics inside the coloured box.
Author's response: We like this suggestion.
Changes to the manuscript: We have added the numbers in figure 4 and 5. After the reviewers suggestion above to look at the spatial pattern, we have changed figure 6 and 7 (now 7 and 8) to show this, the boxes on those figures are too small to contain numbers, so we did not include them there. A typo on the label of the model efficiency figures has also been corrected.

Referee comment: Figure 8: is the Taylor diagram showing relative SD (i.e. SDmodel/SDdata) or the absolute value? In any case I would plot the dotted circumference passing through the DATA point, to highlight when model and data have the same standard deviation.
Author's response: The Taylor diagram shows relative SD.
Changes to the manuscript: A dotted line for SD=1 has been added to the Taylor diagram and the information that it is showing the relative SD is added to the figure label.