Detailed responses to reviewer 1 (reviewer comments are included in black, responses in blue font)

Overview

The study addresses the important question of whether the Earth system models can accurately simulate biogeochemical processes/fields on the Northwest North Atlantic shelf and if they are suitable for projecting the regional impact of climate change. The study ranks the CMIP5 and CMIP6 models in terms of capturing the observed seasonal variability in surface temperature, chlorophyll and nitrate in the Northwest North Atlantic shelf, highlighting that ESMs are not adequate to project accurately future changes in the NWA shelf. Based on the comparison with these ocean surface observations the study clearly demonstrates how a regional model outperforms the Earth system models in terms of capturing the seasonal variability in the region during the historical period. The study then provides plausible reasons for the better performance of some of the Earth system models and the improvement from the CMIP5 to CMIP6.The study is interesting, timely and I believe it is a substantial contribution towards improving our understanding and ability to provide accurate regional projections of climate change on the Northwest North Atlantic shelf, particularly in the context of biogeochemical fields. The methods and results are reasonable, justified and well presented. Hence, in my opinion the study fits well the Biogeosciences scope and should be consider for publication after addressing some very minor comments/revisions as I explain below.

Response: We appreciate the positive and constructive feedback. We provide detailed responses to the comments below.

General comments

Comment:

1. I agree that not all the ESMs are suitable to force regional model projections and should be selected carefully as not to force the regional model with unrealistic large-scale patterns. The choice of the 'parent model' should be based on its performance on capturing the large scale patterns/features that are introduced in the regional models such as: large scale circulation patterns, and biogeochemical and physical fluxes in-out of the regional model domain (the study highlights well this point in the introduction, lines 48-59 and in discussion, lines 343-351). Misrepresentation of the ESMs surface fields in the Gulf of Maine, Scotian Shelf and Grand Banks may be due to the models not being able to accurately represent on-shelf fine-scale biogeochemical and physical processes due to lack of resolution (which is inadequate to resolve the physical/biogeochemical processes on the shelves in all the ESMs, as mentioned in the manuscript), lack of accurate river discharge, biogeochemical model parameter optimisation for global use rather than the region of interest, or lack/misrepresentation of other local processes (e.g., enhance mixing associated with tides). The regional model is meant to address the above on-shelf problems, and as long as the large-scale patterns are introduced correctly at the open boundaries it should be able to perform well. Hence, I am a little reserved about the conclusion that 3 specific ESMs are most appropriate to force regional model projections in NWA (lines 21-23 and 77-79) based on the ESMs' performance on the shelf at the surface. In my understanding, there is no explicit analysis in the study to link the ESMs

ranking on the shelf with the large-scale physics/circulation patterns off-shelf, or across-shelf properties exchange beyond a moderate correlation of surface temperature and surface chlorophyll, and no explicit link between the ESMs performance on the shelf and their performance off the slope in the vicinity of the regional model boundaries. So in my opinion, I would suggest to further discuss and clarify this issue: does low ranking on the shelf translate directly to low ranking of the larger scale patterns in the region and specifically to low ranking in the quality of the ESMs exchange/input of physical and biogeochemical properties at the location of the open boundaries offshore the Scotian shelf, Gulf of Maine and Grand Banks? Why are the ESMs with high ranking on the shelf the ones that can provide the best open boundary conditions coming off the shelf?

Response: We agree with the reviewer that ESMs performance offshore, along a regional model boundary such as the ACM, is most important for regional downscaling and may differ from those on shelf, although we suspect that the performance on the shelf is related to the performance along the boundaries. To address this specific point we now provide an additional analysis of ESM performances along the ACM boundaries and compare those with the results from the shelf, see supporting Figures S10–13 and the following text on lines 272 to 281:

"*Model scores and ranking were also calculated along the boundaries of the regional model (see supporting Figure S10). The ranking shows that model performance on the shelf is not necessarily indicative of the performance along the boundaries of the regional model (supporting Figure S11, Table S2). Moreover, individual rankings are much more variable at the boundaries, even for the best performing models. The 8 best ESMs along the boundaries (22, 11, 30, 28, 16, 10, 26, 6) have an average rank of 9.2–10.5. There are no significant correlations between individual rankings, including temperature and salinity. Nonetheless, there is some agreement between the shelf and the outer boundary ranking for chlorophyll (ρ="0.80"), nitrate (ρ="0.81") and salinity (ρ="0.81" , supporting Figure S12 and Table S3). Interestingly, the agreement is better with CMIP6 models (Table S3). However, there is no agreement for temperature. A similar pattern is found for individual boundaries (Figure S13). In this case, and apart from temperature, the model ranks along the northeastern boundary agree the most with those from the shelf.*"

We also clarified the objectives and findings of our analysis and provide further context about regional applications using ESM output, which is not only used to provide boundary conditions for regional downscaling. For example, ESM projections are used to drive higher trophic level models and to assess societal impacts of climate change, such as fish catch, that affect mainly Exclusive Economic Zones (EEZ), i.e. coastal ecosystems. Our results indicate that the choice of ESM for these projections is very important. We added the following text on lines 60 to 64:

"*Despite these issues, CMIP historical simulations and future projections have been used to characterize biological responses to climate change in the NWA (e.g., Bryndum-Buchholz et al., 2020a; Greenan et al., 2019; Lavoie et al., 2019; Stortini et al., 2015; Wilson et al., 2019; Wilson and Lotze, 2019). ESM selection in these regional studies is either qualitative or based on either scenario outcomes (e.g. variability across models) or* *global assessments rather than on regional model performance. However, ESMs that poorly represent the dynamics of the NWA will affect the results of regional studies.*"

Specific comments

Comment:
2. Line 75 (typo): I think an 'on' is missing such that: 'based on the mismatch...'

Response: Corrected.

Comment:
3. Lines 147-148 (just a suggestion): The authors can refer and link R and Rbio symbols in Table 2 directly with the rank for temperature, chlorophyll and nitrate, and the rank of chlorophyll and nitrate only, respectively, to help the readers follow more easily the ranking metric.

Response: Thank you for the suggestion, we now refer to $\overline{\mathrm{R}}$ and $\overline{\mathrm{R}}^{\mathrm{bio}}$ as follows:

L160-162: "*The overall rank was determined by ranking models by the averages of their ranks for surface temperature, salinity, chlorophyll, and nitrate ($\overline{R}$). For models with equal averages the ranking was determined by the average of chlorophyll and nitrate ranks ($\overline{R}_{bio}$).*"

Comment:
4. Line 179: Please can you clarify which months you consider for the winter bias for nitrate? In my understanding winter is defined as December-February (based on Table3) but November-January are excluded from the observations for nitrate, so is it only February?

Response: Nov–Jan were excluded from the World Ocean Atlas (open/close circles in Fig 5g-i) and therefore in the Gulf of Maine there are only February data for the winter. However, for the Scotian Shelf and Grand Banks AZMP data were also available (black squares in Fig 5h-i) and are included in the winter (December-February) bias calculation. This was be clarified as follows in the captions of Figures 4 and 5:

"*Nov–Jan WOA nitrate data are excluded (open circles). Model comparison with observations in the Gulf of Maine is therefore only available from February to October. For the Scotian Shelf and the Grand Banks additional AZMP data are available. In case of multiple observations, the data are monthly averaged.*"

And in the main text as follows:

L199-200: "*Note that since Nov.–Jan. nitrate WOA observations were excluded from the analysis (see section 2.3), winter observations are only available in February in the Gulf of Maine and in December and January in Grand Banks.*"

Comment:
5. Line 233 (just a suggestion): Maybe if you could clarify that here you mean the best overall ESM in terms of nitrate and chlorophyll (the Rbio in Table 2) by adding

something along the lines: 'ACM and model 22 (the best overall ESM for combined nitrate and chlorophyll) indicates...'.

Response: The sentence was changed to:

L263-264: "*The gap between ACM and model 28 (ESM with best* $\overline{R}$ *and* $\overline{R}_{bio}$*, Table 2) indicates that none of the ESM performs best for all fields, especially for both chlorophyll and nitrate.*"

Note that the nitrate field was reprocessed for model 22 after we found an error in the original file and its nitrate score degraded slightly. The best ESM is now model 28.

Comment:
6. Lines 268-270: I agree that some of the errors in chlorophyll are linked with the temperature bias and the misrepresentation in general circulation. However, in my opinion i) the relative moderate correlation between chlorophyll and temperature r=0.51 and ii) the improvements in chlorophyll in CMIP6 relative to CMIP5 models that are not associated with any improvement in temperature (as discussed in lines 321-328) indicate that the errors in surface chlorophyll are also driven to a significant degree by a poor biogeochemical model component rather than by the ocean physics only. Hence, I suggest that the authors could modify this sentence to reflect this.

Response: This paragraph was modified as follows:

L311-322: "*The correlation between temperature and chlorophyll scores (and to a lesser extent salinity) and the concomitant poor scores in chlorophyll and temperature/salinity (i.e. Group C in Figure 7) indicate that errors in surface chlorophyll concentration are partly driven by a misrepresentation of the general circulation and, more generally, of ocean physics. The improvement in chlorophyll from CMIP5 to CMIP6 without an associated improvement in temperature suggest that the errors in surface chlorophyll were also driven to some extent by errors in the biogeochemical model component. Lavoie et al. (2019) indicated that the misrepresentation of primary production in the NWA may be associated with the misrepresentation of particulate organic matter sinking and remineralization in the subsurface layer. They found an annual subsurface nitrate peak in CanESM2, GFDL-ESM2M, NorESM1-ME, CESM1-BGC (models 2, 7, 18 and 3, respectively) similar to the high surface nitrate found in this study (supporting Figures S1 and S3). However, all these models had poor scores in our assessment and therefore do not provide an appropriate representation of the biogeochemistry on the NWA shelf (Figure 8) or along the ACM boundaries (Figure S11). However, it is not possible, and beyond the scope of this work, for us to draw conclusions about the source of the regional mismatch in surface chlorophyll and nitrate in the ESMs.*"

Comment:
7. Section 4.2: Although the authors mention it, in my opinion, they could highlight even more that the lack of correlation between resolution and rank is not surprising as none of the ESMs have the resolution to explicitly resolve the processes in shelf-scales rather than parameterise them. Maybe the authors could add a sentence after line 303 along the lines: 'This lack of correlation between model resolution and accuracy on the NAW shelf

and the primary control of performance by the model set-up is not surprising as all ESMs are coarse and do not explicitly resolve the shelf-scale processes but rather rely on their parameterisation.

Response: The following sentence was added:

L390-392: "*The lack of correlation between model resolution and performance on the NWA shelf is not surprising as all ESMs are still coarse and do not explicitly resolve shelf-scale processes but rather rely on their parameterisation. Much higher resolution will be necessary…*"

Comment:
8. Figure 8 caption (just a suggestion): The authors mention that the temperature rank is hidden for model 6, however, in my understanding ranks for other models are also hidden as they coincide (model 1, 30 and 18). Maybe just for clarity and to make the link between figure 8 and table 2 explicit, you could mention in the caption something along the lines 'Coinciding ranks as shown in table 2 are hidden'.

Response: The following sentence was added to Figure 8 caption:

L805: "*Hidden coinciding ranks (models 2, 3, 6, 10, 11, 18, 27, 28 and 30) are provided in Table 2.*"

Detailed responses to reviewer 2 (reviewer comments are included in black, responses in blue font)

General comments

Comment:

1. In this paper, the authors compare the output of CMIP5 and CMIP6 Earth System Models (ESMs) to observations in order to determine which models are suitable to build boundary conditions for projections. A ranking analysis was performed on a large array of ESMs. However, they are only looking at surface values of 3 variables and far away from the regional model boundaries, even though they mention on lines 44-46 that it is important to look at the information imposed at the boundaries. I think the objective stated on line 67 "Our objective is to assess the performance of a number of available ESMs in reproducing present conditions on the NWA shelf in contrast to a high-resolution regional model" is more in line with what is presented in the manuscript since there is no analysis at the boundaries.

Response: We agree with the reviewer that ESMs performance offshore, where the regional model boundary is located, is most important for regional downscaling and may differ from those on shelf, although we suspect that performances on the shelf is similar to performance along the boundaries. To address this specific point that was also raised by reviewer 1, we added an analysis of ESM performance along the ACM boundaries and compare those with the results from the shelf (see section 3.2 and supporting Table S2 and Figures S10–S13).

As mentioned in the response to reviewer 1's general comments, we also clarified the objectives and findings in the revised manuscript and now provide further context about the regional use of ESM data. ESM projections are used to drive higher trophic level models and to assess societal impacts of climate change, such as fish catch, that affect mainly Exclusive Economic Zones (EEZ), i.e. coastal ecosystems. Our results indicate that the choice of ESM for these projections is very important. We added the following text on lines 60 to 64:

"*Despite these issues, CMIP historical simulations and future projections have been used to characterize biological responses to climate change in the NWA (e.g., Bryndum-Buchholz et al., 2020a; Greenan et al., 2019; Lavoie et al., 2019; Stortini et al., 2015; Wilson et al., 2019; Wilson and Lotze, 2019). ESM selection in these regional studies is either qualitative or based on either scenario outcomes (e.g. variability across models) or global assessments rather than on regional model performance. However, ESMs that poorly represent the dynamics of the NWA will affect the results of regional studies.*"

Comment:

2. They are not discussing the processes that lead to the observed values in the region under study and they are not analysing if the models do represent these processes correctly. I believe salinity should be included, as surface temperature depends strongly on atmospheric forcing while salinity is more representative of the different water masses in that region.

Response: The objectives of our study are: 1) to provide users of ESM output with information about model performance, either for direct use in shelf regions (see previous comment) or for regional downscaling, and to compare with a high-resolution regional model. We now distinguish more clearly between the two types of usage of ESM output. We also discuss some potential sources of mismatch between models and observations but since only limited output is available from the ESMs (e.g., only monthly means of surface properties for CMIP5) we are not able to properly analyze the underlying reasons. We note that such an analysis is not the objective of this study and outside of the intended scope. We now discuss the findings of Lavoie et al (2019) about the parameterization of vertical fluxes and remineralization in the biogeochemical models. For instance, we added the following discussion:

L315-322: "*Lavoie et al. (2019) indicated that the misrepresentation of primary production in the NWA may be associated with the misrepresentation of particulate organic matter sinking and remineralization in the subsurface layer. They found an annual subsurface nitrate peak in CanESM2, GFDL-ESM2M, NorESM1-ME, CESM1-BGC (models 2, 7, 18 and 3, respectively) similar to the high surface nitrate found in this study (supporting Figures S1 and S3). However, all these models had poor scores in our assessment and therefore do not provide an appropriate representation of the biogeochemistry on the NWA shelf (Figure 8) or along the ACM boundaries (Figure S11). However, it is not possible, and beyond the scope of this work, for us to draw conclusions about the source of the regional mismatch in surface chlorophyll and nitrate in the ESMs.*"

L403-407: "*Lavoie et al. (2019) suggested that the PISCES biogeochemical model may underestimate subsurface remineralization in the CNRM and IPSL models, resulting in low surface nutrients where the Gulf Stream detaches from the coast. Our rankings (shelf and offshore) do not support this hypothesis; high surface nitrate concentrations were present in the CNRM models (throughout the region) and the IPSL-CM5A models (around the GoM) (Figures S1–4, S7, S9).*"

We included temperature in the comparison because it is an important variable for higher trophic level studies and climate change impacts. The fact that surface temperature is available at high spatial and temporal resolution on the shelf, similar to chlorophyll, is also important. Despite the tight control by atmospheric forcing, we did find significant differences in surface temperature across the ESMs. We believe these differences are of interest and relevant to many users.

Also, based on the Reviewer's suggestion, we have now included salinity as a fourth variable in our assessment (see Table 2, Figures 4–8, 12), that is discussed throughout the manuscript, such as in Sections 3.1 and 3.2:

L180–183: "*The range of simulated surface salinity is large (Figure 4d–f). Most models overestimate salinity in the GoM (bias = +1.46, Figure 4d). The mismatch is large on the SS and GB but not consistent among models, except for an annual, positive bias in CMIP6 models (bias = +1.42 and +0.76 respectively, Figure 4e–f). In the two latter* *regions, the biases in CMIP5 models compensate each other, resulting in an ensemble mean close to the observations*."

L215–217: "*For temperature and salinity, models 3, 20–21, and 24–25 have the largest discrepancy with observations and some clearly represent better the annual cycle than others. The best models for temperature (5–6, 14, 16 28) do not always match the best for salinity (5, 16, 27–28, 30).*"

L226–232: "*As observed previously in Figure 6, the scores of ESMs have a much larger range of variability for temperature (1.5–7.8), salinity (0.5–4.2) and nitrate (1.4–13.2) than for chlorophyll (0.81–1.42) due to the large mismatch observed with a few models (Figure 7, supporting Figures S1–S5). For temperature, 4 of the 6 poorest (largest) scores (> 4.5) are in the CMIP6 group. They all markedly overestimate temperature, especially in the GoM (see supporting Figures S1, S4–S5), except for model 4 that underestimates temperature in SS and GB. The other models have also the poorest scores with respect to salinity. They all largely overestimate salinity in the 3 regions and are clearly outliers with respect to their CMIP category.*"

Comment:
3. Moreover, it is very surprising that a similar study (Lavoie et al. 2019) in the exact same area, with the same purpose, and using some of the same ESMs is hardly mentioned at all. No comparison of the results of this study with the 2019 study is made.

Response: The study of Lavoie et al (2019) is different in that their main focus is on future projections, but it is carried out in the NWA and is therefore relevant, of course. We now discuss and cite their findings where appropriate and also included a recent report by Lavoie et al (2020) on regional downscaling in the NWA as well as a previous report by Lavoie et al (2013) that provided qualitative comparisons of CMIP5 models with observations.

Comment:
4. Also there is not enough details on the comparison with the data, they appear to be comparing different time periods (see detailed comments) or on how the ESMs were brought to a single grid.

Response: We provide the information about time range, averaging and spatial mapping in the Methods. Additional information is now added for completeness sake, as detailed in the responses to detailed comments 18–20, 27 and 30 below.

Comment:
5. There is only a vague mention of what the improvements are between the CMIP5 and CMIP6 models. What was improved should be stated (not only biogeochemistry of physics) so that the reader can judge on the potential impact on the ranking.

Response: See response to comment 2 above and comment 25 below. Model changes from CMIP5 to CMIP6 can be significant, including in the atmospheric and terrestrial realms. This study is not meant to pinpoint the sources of improvement in performances. Given the limited output available from these models, we can only speculate on the sources of improvement based on our results. The reader is referred to the specific papers listed in Table 1 for the list of changes in the models. To clarify this point, we added the following statement:

L424-425: "*For specific changes in the CMIP6 model versions, the reader is referred to the references listed in Table 1*."

Comment:
6. Increasing the model resolution in order to improve the representation of the circulation in the NWA has been mentioned by many authors (e.g. Loder, Brickman, Yool). Here it is stated that the resolution does not have an impact. This is a big statement, considering the general agreement, and it should be demonstrated. The authors could show the changes in circulation of a few models they are giving in example for this.

Response: Our findings are in line with previous work, including the ones cited above, see responses to detailed comments 9, 16, and 34–35.

Comment:
7. All these points should be addressed in order for the conclusions to be more convincing (ranking based on analysis of shelf surface conditions representative of boundary conditions).

Response: These points are addressed in the detailed comments below.

Comment:
8. Also, Lavoie et al. (2019) estimated that the boundary conditions obtained with the ESMs were not as reliable for the simulation of the conditions on the Scotian Shelf and in the Gulf of Maine. It would good to know if there was an improvement in this regard with the CMIP6 ESMs.

Response: An analysis of model performance along the ACM boundaries is now provided in the revised manuscript. We found similarities in the CMIP5 and CMIP6 rankings for all variable except temperature but with more variability in the ranking of individual variables. This further supports the findings of Lavoie et al (2019). The variability tends to decrease in the CMIP6 models, which suggests more reliability of boundary conditions obtained from the ESMs. However, given the disagreement between shelf and boundary ranking for temperature, we do not feel that we can go further in our discussion of the reliability of boundary conditions than we have.

<u>Specific comments</u>

Comment:
9. Line 11: Here you say that the coarse resolution is not appropriate to represent the circulation and elemental flux but later on you say that increasing the resolution does not matter. Is it important or not?

Response: The two statements are not in contradiction. It is well known that the coarse resolution of ESMs is an issue with regard to resolving shelf-scale processes and that a high resolution is necessary in these areas, as mentioned in comment 6 above. However, even the highest resolution ESM in the ensemble is too coarse to resolve shelf-scale processes and therefore it is not surprising that we do not yet see better performance with increasing ESM resolution. We have now clarified this point by adding the following sentence on Line 392:

"*The lack of correlation between model resolution and performance on the NWA shelf is not surprising as all ESMs are coarse and do not explicitly resolve shelf-scale processes but rather rely on their parameterisation. Much higher resolution will be necessary…*"

Comment:
10. Line 14: ability to reproduce surface observations...

Response: Done.

Comment:
11. Line 15: why is it particularly sensitive?

Response: We refer to the effects of climate change on the location and strength of the Gulf Stream and Labrador Sea currents. We modified the sentence as follows:

L13-15: "*The NWA region is biologically productive, influenced by the large-scale Gulf Stream and Labrador Current systems, and particularly sensitive to climatically induced changes in large-scale circulation.*"

Comment:
12. Line 16: The spatial mismatch in large-scale circulation was not demonstrated. There are references for CMIP5 but what about CMIP6. Changes, or not, in circulation should be shown/mentioned after an inspection of the ESMs results.

Response: We mentioned a warm bias in the Gulf of Maine that is in line with the results of Loder et al. (2015) and Saba et al. (2016) (Lines 365-266). Although smaller, a cold bias appears on Grand Banks in most models (Figures 4c and 5c). The biases suggest a mismatch in the large-scale currents. However, since we did not compare the position of the currents across the models, we rephrased the sentence L15-16 as follows:

"*Most ESMs compare relatively poorly to observed nitrate and chlorophyll and show differences with observed temperature and salinity that suggest spatial mismatches in their large-scale current systems*."

With the addition of salinity, we now have more support for this statement:

L305-307: "*A warm bias and a general overestimation of surface salinity in most models indicate a mismatch in the location of the Gulf Stream that influences conditions on the shelf, in line with the previous results of Loder et al. (2015) and Saba et al. (2016).*"

Comment:

13. Line22: How can we say just by looking at the surface temperature, nitrate and chl a that the top three models are appropriate for boundary forcing? The model boundaries are hundreds of meters deep (and more) and are not located in the regions analysed. It should be mentioned what are the tracers that will be downscaled at the boundaries? Salinity is certainly one of them, why was is not included in the analysis?

Response: The revised manuscript includes salinity and a comparison along the offshore boundaries of the ACM, see responses to comments 2 and 8 above, and response to comment 1 by reviewer 1. Based on the new results we modified our statement as follows:

L22–24: "*An additional evaluation of the ESMs along the regional model boundaries shows larger variability but is generally consistent with the ranking on the shelf. Overall, 11 ESMs were deemed satisfactory for use in the NWA, either directly or for regional downscaling*."

Comment:
14. Main text Line 71: why look only at three variables? What about salinity?

Response: Salinity is now included, see response to comment 2 above. Not all variables were available for all models (ESMs and ACM) and could be compared to observations, so we restricted the comparison to 4 variables (with the addition of salinity) in the revised manuscript. The selected variables are, arguably, the most important to potential users.

Comment:
15. Line 78: historical simulations are not used for projections. This should be rephrased.

Response: The sentence was rephrased as follows:

L85-88: *The comparison provides an overview of ESM performance in the NWA and shows sufficient confidence for only a third of the ESMs. The regional model clearly outperformed all the global models and regional downscaling using single ESM forcing (as opposed to an ensemble) is recommended.*

Comment:
16. Lines 115-116: The ESMs horizontal resolution in the region of interest should be given in Table 1. Some models have a variable resolution and it might not be that bad in the NWA.

Response: To give a sense of horizontal resolution that is easily comparable across models we have provided the number of grid cells in the three zones of interest in Table 1. This value also depends on the coverage, which can be very poor for coarse grids (e.g. IPSL–CM5), and therefore we believe this provides the most readily useful information to the reader. However, as suggested by the reviewer and since resolution is typically reported in degrees for ESMs, we added a column to Table 1 with average Δlon × Δlat on the NWA shelf. We also added the number of vertical levels.

Comment:

17. Line 117: MR and HR mean medium resolution and high resolution respectively. If they share the same grid where does the change in resolution come from?

Response: MR stands for Mixed Resolution and HR for Higher Resolution in the MPI model names. MPI-ESM-MR (CMIP5) and MPI-ESM1-2-HR (CMIP6) have the same ocean circulation model, but the horizontal resolution of the atmospheric component was improved from ~200 km (MPI-ESM-MR) to ~100 km (MPI-ESM1-2-HR). Thus, model names are not related to the ocean model, which can be confusing. A similar confusion can occur from the IPSL CMIP5 model names. In this case, the models share the same ocean model, but the horizontal resolution of the atmospheric model is higher in the medium resolution (MR) version compared to the low resolution (LR) version. To avoid the potential for confusion, the following text was added to the caption of Table 1:

"*Note that the IPSL-CM5 models share the same ocean component with a higher resolution atmospheric component in the MR version. Similarly, MPI-ESM-MR and MPI-ESM1-2-HR share the same ocean component with a higher resolution atmospheric component in the HR version*."

Comment:
18. Line 123: From where were the satellite data obtained. Who did the averaging?

Response: Links to the data were added as follows:

L132-137: "*1) satellite surface chlorophyll observations from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) as 8-day averaged maps at 1/12˚ resolution (1999–2010, https://doi.org/data/10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018), 2) surface nitrate from the World Ocean Atlas 2013 (WOA; Garcia et al., 2014) at 1˚ resolution, 3) daily surface temperature from the Operational SST and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012) at 1/20˚ resolution (2006–2016, https://doi.org/10.5067/GHOST-4FK01) and 4) surface salinity from the WOA at 1/4˚ resolution (Zweng et al., 2013). Monthly climatologies were calculated for each of these*."

The following references were added:

SeaWiFS. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Chlorophyll Data; NASA OB.DAAC, Greenbelt, MD, USA. doi:10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018. Accessed on 2014/03/12.

OSTIA. UK Met Office. 2005. GHRSST Level 4 OSTIA Global Foundation Sea Surface Temperature Analysis. Ver. 1.0. PO.DAAC, CA, USA. doi:10.5067/GHOST-4FK01. Accessed on 2019/12/06.

The original data were daily (OSTIA) and 8-day (SeaWiFS) maps which were converted to monthly climatologies, as mentioned on Lines 133–137.

Comment:

19. Line 128: Which data from the AZMP were used? Along the Halifax line only? Why were the data averaged seasonally and not monthly like the other data?

Response: See also response to comment 30. Yes, we used data along the Halifax Line where both high-resolution glider data and ship-based bi-monthly or seasonal data are available. The location of the data is presented in Figure 1. The glider missions and the AZMP data collection frequency along the Halifax Line were seasonal, which is why the spatially resolved dataset was averaged into seasons rather than months. This information was added in the paragraph (see above). At station 2, we were able to use a bi-weekly frequency for the AZMP climatology. We rephrased the text in the manuscript as follows for clarity:

L138-143: "*In addition, the regional model was validated using high-resolution in-situ observations along the Halifax Line (Figure 1) from the Atlantic Zone Monitoring Program (AZMP, 2000–2014, http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/azmp-pmza/index-eng.html) and glider transects between 2011 and 2016 (Ross et al., 2017). To enable a quantitative comparison between the glider and ACM data (Table 3), we spatially interpolated both datasets onto a transect following the Halifax Line (black line in Figure 1). Glider missions were seasonal and therefore both glider and AZMP transects data were seasonally averaged. For each mission, data were extracted at Station 2 to produce a monthly climatology.*"

Comment:
20. Line 132: So the model results are brought back onto 3 different grids, one for each variable. Are the time period also adjusted? For example SST goes from 2006 to 2016. The CMIP5 historical period ends in 2005. How can the two be compared then? Also, there are probably models that have a higher resolution than 1◦ (see my comment for lines 115-116), what is the impact of decreasing the resolution (converting from higher to lower resolution) and having on the ranking analysis. And how is the conversion of one grid to the other done?

Response: Since we used a heterogeneous data set, we brought the data and model to the same temporal (monthly) and spatial (observation grid) scale for comparison. We used a long-term climatology for robustness, so that our conclusions aren't affected by interannual variability. Ideally, we would use the same time range for observations and models, but this was not possible. All the ESMs used the same time range (30 years climatology, 1976–2005) so their intercomparison is robust. Note that Line 193 should read "(1976–2005)", not "(1975–2005)", which was corrected in the revised manuscript. Unfortunately, we could only run the ACM simulation for 15 years starting in 1999 so the ESMs and ACM simulations overlap for 6 years only. Since the CMIP6 historical simulations end in 2014 it was possible to use the range 2000–2014 with the CMIP6 models. However, the ESM intercomparison would have been less robust and we decided to use the same time range for all the ESMs. To assess the potential bias associated with the selected time range, we compared the scores of the CMIP6 models when averaged over the years 2000–2014, the same time range as for the regional model and similar to the observations of chlorophyll and temperature. Except for temperature in model 30, the scores and ranks are consistent with the previous, see Figures R1 and R2 below. Figure R1 is now included in the revised supplement (Figure S14) and we have now added a paragraph on scores uncertainties in the Discussion (L361-374).

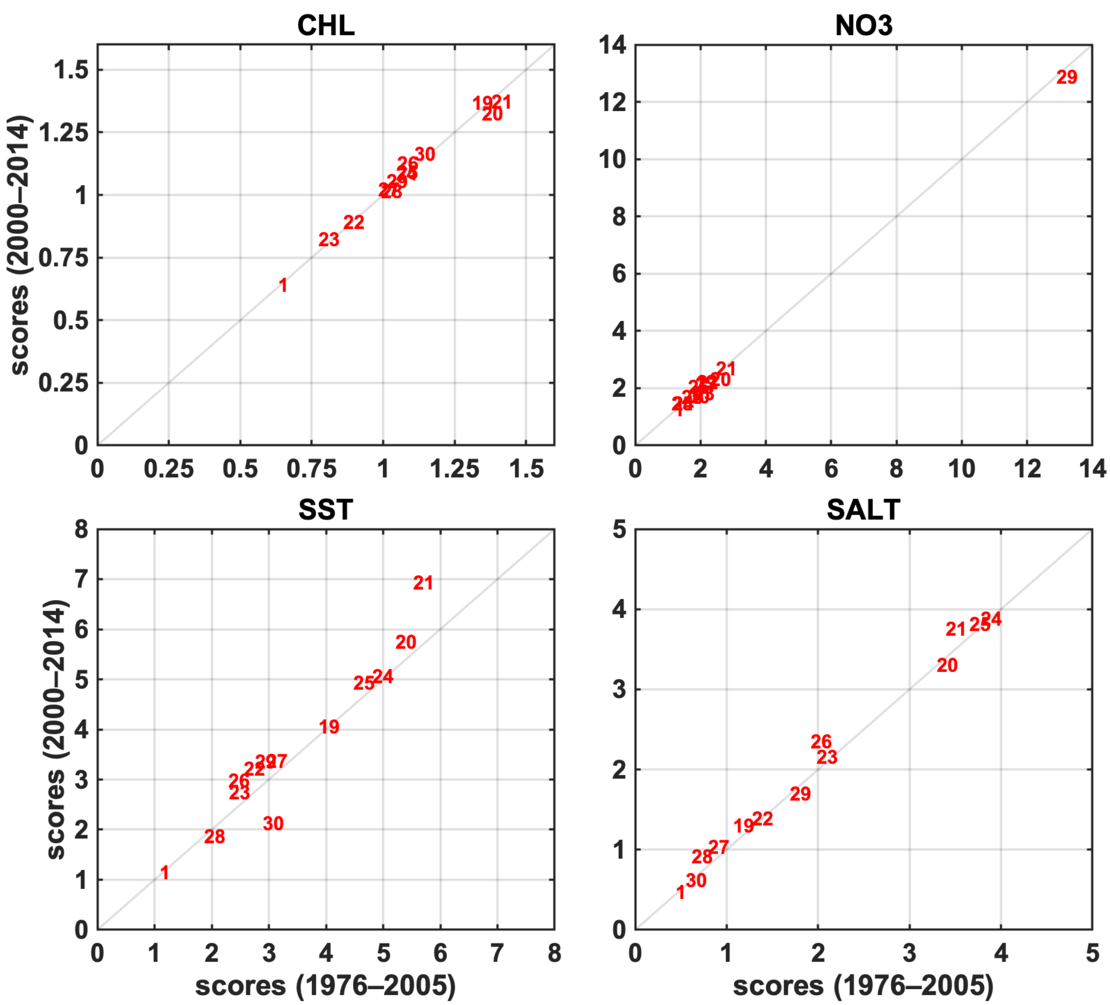


Figure R1. Relationships between scores from CMIP6 model climatologies averaged over 1976–2005 (x-axis) and 2000–2014 (y-axis). ACM (1) is indicated as reference.

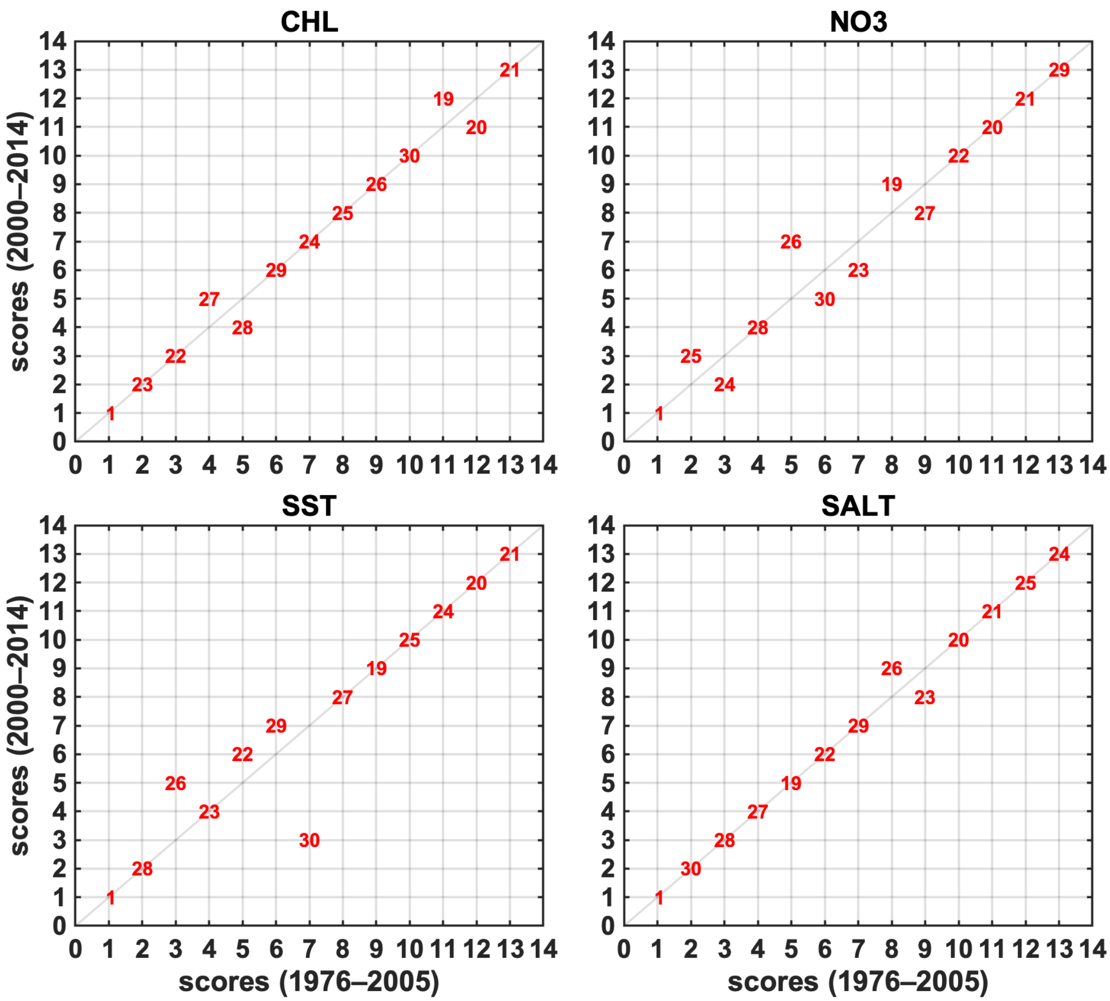


Figure R2. Relationships between ranks from CMIP6 model climatologies averaged over 1976–2005 (x-axis) and 2000–2014 (y-axis). ACM (1) is indicated as reference.

The conversion from grid to grid is simply a nearest neighbor interpolation onto the observations grid. This information is added as follows:

L145-146: "*For comparison with the observations, each model was mapped onto the SeaWiFS, WOA and OSTIA grids using a nearest neighbor interpolation.*"

Comment:
21. Also, how the thickness of the first grid cell compares between the different models?

Response: The ESMs have various vertical resolutions. For completeness, the number of vertical levels was added to Table 1.

Comment:
22. Line 175: What is the main difference between the CMIP5 and CMIP6 groups, why is it better? Improved BGC? Same question hold for nitrate.

Response: In the discussion, we speculate about the source of improvement in surface chlorophyll and nitrate fields (see section 4.6). The suggested sources of improvement refer to the literature as we cannot substantiate the reasons for these changes from the publicly available data.

Comment:
23. Line 200: Figure 6 does not show the annual cycle.

Response: Here we refer to the data that are used to calculate the RMSD. The sentence was modified to:

L221-222: "*some models are much better at representing the observed annual cycle (Figure 6), as indicated by the lower RMSD.*".

Comment:
24. Line 208: Could you explain why? From local atmospheric forcing or circulation change?

Response: We can only speculate about the reasons why some models have poor scores for temperature. In Lines 305–307 we mention the warm bias associated with a mismatch in the location of the Gulf Stream. These models also overestimate salinity, so this explanation is plausible. This information was added as follows:

L229-232: "*They all markedly overestimate temperature, especially in the GoM (see supporting Figures S1, S4–S5), except for model 4 that underestimates temperature in SS and GB. The other models have also the poorest scores with respect to salinity. They all largely overestimate salinity in the 3 regions and are clearly outliers with respect to their CMIP category.*"

However, we do not know why some CMIP6 models (CESM2, GISS-E2-1-G-CC) have a large temperature and salinity bias. These models already had poor scores for temperature and salinity in their CMIP5 version. The following sentence was added:

L417-429: "*Models with poor scores for temperature and salinity (CESM2, GISS-E2-1-G-CC) had already poor scores in their CMIP5 version and therefore the cause of their poor performance is likely the same.*"

Comment:
25. Line 209: What are the improvements in the CMIP6 models?

Response: Here "improvement" refers to the lower chlorophyll scores for the CMIP6 models 22 and 23 (CNRM-ESM2-1 and GFDL-ESM4). These models have the best chlorophyll scores after ACM. For clarity, "improvement" was removed and the sentence is now:

L232-235: "*The range of variability in chlorophyll scores did not reduce from CMIP5 to CMIP6 and given the relatively low scores of a few CMIP6 models (i.e. 22 and 23), the* *range is larger in the CMIP6 group (0.8–1.4, Figure 7, right panel) than in the CMIP5 group (1–1.4, Figure 7, left panel).*"

Comment:
26. Line 215: So this means that the model ranked 2 (and others) might not be ranked as high? What is the impact on the final choice?

Response: The numbers in the parenthesis correspond to the model ID. Since we excluded Nov–Jan from the WOA dataset, i.e. when nitrate is high at the surface, models with consistently low nitrate have lower scores than they would if all winter months were included, and it increases their rank with respect to nitrate. The overall rank is an average of the 4 variables so it should be less sensitive to this effect. Supporting Figures S1-S5j-l indicate that this might be the case for models 4, 8, 14, 19, and 26–27. Models 4, 8, 14 and 19 are part of the outer ensemble so the underestimation of the nitrate score has no effect on the final choice. However, for model 26–27, information on the underestimation of the nitrate scores should be provided. To reflect a possible bias in the ranking of these models we added an asterisk in Table 2 beside the overall rank of these models and the caption was updated accordingly.

Comment:
27. Line 218: Could the fact that you are using different time periods and different grid resolution for the three variables explain the lack of correlation?

Response: We used climatologies to remove the influence of time on the comparisons as much as possible. The new comparison of the scores calculated over the period 1976-2006 and 2000–2014 shows that time does not influence the scores (see response to Comment 20 above. Correlations are consistent between physical and biological variables, despite the different grid resolutions. Nitrate scores did not correlate with any of the other variables and therefore the source of errors for nitrate seems to differ from that of the other variables.

Comment:
28. Line 221: How do you explain that?

Response: We do not wish to speculate about the reasons for each models' individual ranking. As stated before, given the limited output that is available for each of the models we would have to guess, but this is outside the intended scope of this study. The objective is to report on the models' behaviour.

Comment:
29. Line 230: Why? Does it relate to temperature-dependant phytoplankton growth?

Response: Again, we can only speculate. Temperature-dependant phytoplankton growth is a possibility, large scale circulation is another. Except for models 14 and 19, all the models in Group C have poor temperature and salinity scores, and therefore their poor representation of chlorophyll is likely due to a mismatch in large scale circulation. Model 14 has poor salinity and chlorophyll scores but is the best ESM for temperature, whereas model 19 has poor temperature and chlorophyll scores but represents well salinity. Given the mismatch in error statistics for temperature and salinity in these 2 models, the influence of circulation is probably different. We added the following discussion about the source of mismatches in chlorophyll and nitrate:

L311-322: "*The correlation between temperature and chlorophyll scores (and to a lesser extent salinity) and the concomitant poor scores in chlorophyll and temperature/salinity (i.e. Group C in Figure 7) indicate that errors in surface chlorophyll concentration are partly driven by a misrepresentation of the general circulation and, more generally, of ocean physics. The improvement in chlorophyll from CMIP5 to CMIP6 without an associated improvement in temperature suggest that the errors in surface chlorophyll were also driven to some extent by errors in the biogeochemical model component. Lavoie et al. (2019) indicated that the misrepresentation of primary production in the NWA may be associated with the misrepresentation of particulate organic matter sinking and remineralization in the subsurface layer. They found an annual subsurface nitrate peak in CanESM2, GFDL-ESM2M, NorESM1-ME, CESM1-BGC (models 2, 7, 18 and 3, respectively) similar to the high surface nitrate found in this study (supporting Figures S1 and S3). However, all these models had poor scores in our assessment and therefore do not provide an appropriate representation of the biogeochemistry on the NWA shelf (Figure 8) or along the ACM boundaries (Figure S11). However, it is not possible, and beyond the scope of this work, for us to draw conclusions about the source of the regional mismatch in surface chlorophyll and nitrate in the ESMs.*"

L403-407: "*Lavoie et al. (2019) suggested that the PISCES biogeochemical model may underestimate subsurface remineralization in the CNRM and IPSL models, resulting in low surface nutrients where the Gulf Stream detaches from the coast. Our rankings (shelf and offshore) do not support this hypothesis; high surface nitrate concentrations were present in the CNRM models (throughout the region) and the IPSL-CM5A models (around the GoM) (Figures S1–4, S7, S9).*"

Comment:
30. Line 245: What are the years compared for the ACM and the glider data?

Response: Information on the ACM and glider data is provided in the Methods. The ACM data are the same as for the comparison with the ESMs, i.e. years 2000–2014 but presented as a seasonal (Figure 9) and daily (Figure 10) climatology to match the resolution of the glider data. The AZMP years are the same as ACM. The glider missions were carried out between 2011 and 2016 but were heterogeneous in time and space (see tracks on Figure 1). To enable a quantitative comparison between the glider and ACM data (Table 3), we spatially interpolated both dataset onto a transect following the Halifax Line (black line in Figure 1). The glider missions were seasonal, which is why the spatially resolved dataset was averaged into seasons (Figure 9). For each mission, data were extracted at Station 2 to produce a monthly climatology (Figure 10). ACM data were extracted at this location for comparison. We added this information as follows in the Methods section:

L140-143: "*To enable a quantitative comparison between the glider and ACM data (Table 3), we spatially interpolated both datasets onto a transect following the Halifax*

Line (black line in Figure 1). Glider missions were seasonal and therefore both glider and AZMP transects data were seasonally averaged. For each mission, data were extracted at Station 2 to produce a monthly climatology."

Comment:
31. Line 260: Correlation coefficients are high for nitrate despite having a large bias and RMSD. This should be explained.

Response: The correlation coefficient is a complementary measure to bias and RMSD. Correlation and bias are largely unrelated. The former is a measure the similarity in spatial or temporal variations but does not account for bias. In other words, the same correlation coefficient can occur for very different values of bias. Likewise, high correlation does not imply low RMSD. In a noisy data set the RMSD will be higher than in a data set that is smooth, while both might display the same correlation.

Comment:
32. Line 270: See my previous comments about time-period and grid differences. I think that a statement about a misrepresentation of ocean physics as the cause should be backed up since later on the cause for nitrate mismatch is stated as coming from the BGC behavior (line 279). There are refs for the CMIP5 models but was there an improvement in circulation with the CMIP6 group or not?

Response: Likely, the mismatch is partly associated with ocean physics and partly due to the BGC model. To reflect this, the sentence was modified as follows (see also the response to Comment 29 above):

L311-315: "*The correlation between temperature and chlorophyll scores (and to a lesser extent salinity) and the concomitant poor scores in chlorophyll and temperature/salinity (i.e. Group C in Figure 7) indicate that errors in surface chlorophyll concentration are partly driven by a misrepresentation of the general circulation and, more generally, of ocean physics. The improvement in chlorophyll from CMIP5 to CMIP6 without an associated improvement in temperature suggest that the errors in surface chlorophyll were also driven to some extent by errors in the biogeochemical model component.*"

Comment:
33. Line 288: So here again, the model-data comparison was made on a different grid than for the ESMs. Shouldn't it be done on the same grid for an appropriate comparison?

Response: No, the grid for comparison depends on the dataset. Here, since we have both high (glider) and low (AZMP) spatial resolution data, we mapped the data along the Halifax Line.

Comment:
34. Line 295: Lavoie et al. (2019) also point at the misrepresentation of the remineralisation depth in those models as a likely cause. This also explain why some models having a coarse resolution still have good results with biogeochemistry. But the statement made below that improving the model resolution does not improve the representation of circulation and main features in the models, such as the representation

of the Gulf Stream detachment point and flow around the Grand Banks should be demonstrated. There is a large consensus on that and it should not be stated lightly. The authors could actually show the mean currents between the two versions of a same model with improved resolution. Especially that you state that higher resolution is required to refine the projections on line 306. There is a contradiction here.

Response: See also response to comment 9 above. We agree with the Reviewer that there is a large consensus on the effect of grid resolution on large scale circulation and our discussion is in line with this consensus. The resolution of the CMIP models is much coarser than the resolution of the models used to study the effect of grid resolution on the large-scale current systems of the NWA. Therefore, as also pointed out by Reviewer 1 (see comment 7 by Reviewer 1), it is not surprising that current ESMs do not show the effect of grid resolution on model performances; much higher resolution will be necessary to see this effect. We clarified this point as follows after Line 390:

"*The lack of correlation between model resolution and performance on the NWA shelf is not surprising as all ESMs are still coarse and do not explicitly resolve shelf-scale processes but rather rely on their parameterisation. Much higher resolution will be…*"

We also added the following sentence Line 315:

"*Lavoie et al. (2019) indicated that the misrepresentation of primary production in the NWA may be associated with the misrepresentation of particulate organic matter sinking and remineralization in the subsurface layer. They found an annual subsurface nitrate peak in CanESM2, GFDL-ESM2M, NorESM1-ME, CESM1-BGC (models 2, 7, 18 and 3, respectively) similar to the high surface nitrate found in this study (supporting Figures S1 and S3). However, all these models had poor scores in our assessment and therefore do not provide an appropriate representation of the biogeochemistry on the NWA shelf (Figure 8) or along the ACM boundaries (Figure S11).*"

Comment:
35. Line 310: Here again it appears to be contradictory as you previously mentioned that BGC improvements we the cause for improvements in the CMIP6 ranking. There are likely different versions of the 4 BGC models mentioned, which should be specified in the table and considered in the analysis. Also, it could relate to the processes that control nitrate in the regions under study, they are different for your 3 regions. And how well are these processes represented by the ESMs?

Response: Here we refer to the general BGC component. There are not enough data available to compare specific model versions or parameterizations. This paragraph is meant to point out that, in our comparison, there was no relationship between the type of model and the overall performances. But we cannot go further, and this is not the objective of the study. For clarification the paragraph was modified as follows:

L395-402: "*Although model performance is likely influenced by the biogeochemical model structure, we did not find a clear relationship between the type of biogeochemical model and performance. Here we only refer to the model type because the same model may have different parameterizations when used by different groups. While the inner and* *outer ensembles share only 3 biogeochemical models (PISCES, HAMOCC, TOPAZ2) out of 13, there was no indication of consistently better performance for the biogeochemical models in the inner ensemble. For example, models using similar ocean biogeochemistry (e.g., PISCES: 5, 12–14 (CMIP5), 22 and 26 (CMIP6), and HAMOCC: 15–16, 18 (CMIP5), 28–29 (CMIP6)) had very different ranks, with no obvious relationship between overall model rank and the ocean biogeochemical model component. Moreover, 5 and 4 biogeochemical models were represented in the 5 best ranked ESMs on the NWA shelf and outer ACM boundaries, respectively, similar to previous findings by Rickard et al. (2016). Lavoie et al. (2019) suggested that the PISCES biogeochemical model may underestimate subsurface remineralization in the CNRM and IPSL models, resulting in low surface nutrients where the Gulf Stream detaches from the coast. Our rankings (shelf and offshore) do not support this hypothesis; high surface nitrate concentrations were present in the CNRM models (throughout the region) and the IPSL-CM5A models (around the GoM) (Figures S1–4, S7, S9).*"

Comment:
36. Line 335: What was updated in the ocean biogeochemistry?

Response: As mentioned in the response to previous comments above, our goal is not to discuss the details of the models. The HAMOCC biogeochemistry module includes a new parameterization of detritus sinking, which may influence surface chlorophyll and nitrate, as suggested by Lavoie et al (2019). However, this explanation is speculative and we do not think that it should be included here.

Comment:
37. Figure 4f: In the suppl. figures, there is more chl a in the model than in the obs. The opposite is shown here.

Response: The supplemental figures S1–S5 present the individual chlorophyll time series for the 29 ESMs, whereas Figure 4f shows ECM ensembles. The "all" ensemble time series are calculated with all the individual ESM time series in Figures S1–S5. Figure 4f shows that even though individual ESMs can be close to observations during the spring bloom (e.g. HadGEM2, Figure S2f) and even significantly larger (e.g. CESM2, Figure S4f), the ensemble of all ESMs underestimates the bloom.

Comment:
38. Figure 7: Maybe use ACM instead of ROMS.

Response: Done.

Comment:
39. Figure 8: Could specify that ACM has the same rank for the three variables (only see one point)

Response: The following sentence was added to the figure caption:

L805: "*Hidden coinciding ranks (models 2, 3, 6, 10, 11, 18, 27, 28 and 30) are provided in Table 2.*"

An observation-based evaluation and ranking of historical Earth System Model simulations in the northwest North Atlantic Ocean

Deleted: for regional downscaling

Arnaud Laurent[1], Katja Fennel[1], Angela Kuhn[2]

[1]Department of Oceanography, Dalhousie University, Halifax, Nova Scotia, Canada
[2]Scripps Institution of Oceanography, UC San Diego, USA

Correspondence to: Arnaud Laurent (arnaud.laurent@dal.ca)

Abstract. Continental shelf regions in the ocean play an important role in the global cycling of carbon and nutrients but their responses to global change are understudied. Global Earth System Models (ESM), as essential tools for building understanding of ocean biogeochemistry, are used extensively and routinely for projections of future climate states; however, their relatively coarse spatial resolution is likely not appropriate for accurately representing the complex patterns of circulation and elemental fluxes on the shelves along ocean margins. Here, we compared 29 ESMs used in the IPCC's Assessment Rounds (AR) 5 and 6 and a regional biogeochemical model for the northwest North Atlantic (NWA) shelf to assess their ability to reproduce surface observations of temperature, salinity, nitrate, and chlorophyll. The NWA region is biologically productive, influenced by the large-scale Gulf Stream and Labrador Current systems, and particularly sensitive to climatically induced changes in large-scale circulation. Most ESMs compare relatively poorly to observed nitrate and chlorophyll and show differences with observed temperature and salinity that suggest spatial mismatches in their large-scale current systems. Model-simulated nitrate and chlorophyll compare better with available observations in AR6 than in AR5, but none of the models performs equally well for all 4 parameters. The ensemble means of all ESMs, and of the five best performing ESMs, strongly underestimate observed chlorophyll and nitrate. The regional model has a much higher spatial resolution and reproduces the observations significantly better than any of the ESMs. It also simulates reasonably well vertically resolved observations from gliders and bi-monthly ship-based monitoring observations. A ranking of the ESMs indicates that only 1 ESM has good and consistent performances for all variables. An additional evaluation of the ESMs along the regional model boundaries shows larger variability but is generally consistent with the ranking on the shelf. Overall, 11 ESMs were deemed satisfactory for use in the NWA, either directly or for regional downscaling.

Deleted: climate change.

Deleted: due to

Deleted: circulation

Deleted: 3

Deleted: A ranking of the ESMs suggests that the top 3 models are appropriate as boundary forcing for regional projections of future changes in the NWA region

Formatted: English (CAN)

1 Introduction

Elemental fluxes along ocean margins, which are areas of complex physical and biogeochemical interactions, are important components of the global cycles of carbon (C) and nitrogen (N). For example, continental shelves host up to a third of oceanic primary production and over 40% of carbon burial in the ocean (Ducklow and McCallister, 2004; Muller-Karger, 2005; Walsh, 1991). They also are important sites of sediment denitrification leading to a net removal of fixed nitrogen (Fennel et al., 2006;

Seitzinger and Giblin, 1996). Many shelf regions are thought to be a significant sink for atmospheric CO_2 (Cai et al., 2006; Chen et al., 2013; Laruelle et al., 2018), including the eastern margin of North America (Fennel et al., 2019 and references therein), although there are significant discrepancies in available estimates. Despite their importance, the response of ocean margins to climate change is understudied relative to the open ocean.

Future projections of ocean biogeochemistry rely heavily on Earth System Models (ESMs). These are state-of-the-art, comprehensive representations of the major earth system components (including atmosphere, ocean, and land surface) and are routinely used to perform climate scenario projections. The spatial resolution of the CMIP-class ESMs typically ranges from 0.5 to 2° and is too coarse to resolve coastal ocean dynamics and interactions between shelf and the open ocean (Anav et al., 2013; Bonan and Doney, 2018; Holt et al., 2017). This leads to uncertainty in future projections, not only for margin regions, and a global underestimation of the high primary productivity in coastal regions (Bopp et al., 2013; Schneider et al., 2008).

Regional coupled circulation-biogeochemical models have been developed at much higher spatial resolution. These regional models have been used to investigate biogeochemical processes along ocean margins (Fennel et al., 2006, 2013; Lachkar and Gruber, 2011; Peña et al., 2019; Siedlecki et al., 2015; Zhang et al., 2019) and project future states resulting from climate change (Gruber et al., 2012; Hermann et al., 2016; Holt et al., 2016; Laurent et al., 2018). The regional models allow for the temporal and spatial resolution necessary to resolve mesoscale processes and can be regionally calibrated (e.g., Kuhn and Fennel, 2019; Mattern and Edwards, 2017). However, the dynamics of a regional model is strongly determined by information imposed along the model's open lateral boundaries, typically derived from a larger scale model, reanalysis product, or observation-based climatology. For future climate simulations, a regional model requires boundary information from future projections of large-scale models or ESMs.

Deleted:

Deleted: (Fennel et al., 2006, 2013; Lachkar and Gruber, 2011; Peña et al., 2019; Siedlecki et al., 2015; Zhang et al., 2019)

The northwest North Atlantic (NWA), located at the confluence of the subtropical and subpolar gyres, is particularly challenging to global ocean circulation models and highly sensitive to climate-induced modifications of the large-scale circulation, which are thought to be responsible for a multi-decadal deoxygenation trend in the region (Claret et al., 2018; Gilbert et al., 2010). While the CMIP models reasonably describe the large-scale climatological features of ocean physics in the NWA, the detailed current structure is poorly represented due to a mismatch in the location of the subtropical and subpolar gyres (Loder et al., 2015). The Gulf Stream usually extends too far north and the branch of the Labrador Current flowing southwest along the shelf edge tends to be missing (Lavoie et al., 2019; Loder et al., 2015). This leads to a warm bias in the NWA, a common feature among coarse resolution ESMs (Saba et al., 2016). The absence of the shelf-break current significantly impacts cross-shelf exchange with much larger shelf water residence times in a high-resolution regional model (Rutherford and Fennel 2018) compared to estimates from a global model (Bourgeois et al. 2016). These discrepancies have been attributed to the coarse resolution of the global models (Lavoie et al., 2019; Loder et al., 2015; Rutherford and Fennel, 2018; Saba et al., 2016). Despite these issues, CMIP historical simulations and future projections have been used to characterize biological responses to climate change in the NWA (e.g., Bryndum-Buchholz et al., 2020a; Greenan et al., 2019; Lavoie et al., 2019; Stortini et al., 2015; Wilson et al., 2019; Wilson and Lotze, 2019). ESM selection in these regional studies is either

Deleted: These discrepancies have been attributed to the coarse resolution of the global models (Loder et al., 2015, Saba et al. 2016, Rutherford and Fennel 2018, Lavoie et al., 2019).

qualitative or based on either scenario outcomes (e.g. variability across models) or global assessments rather than on regional model performance. However, ESMs that poorly represent the dynamics of the NWA will affect the results of regional studies. Increased coastal model resolution can be achieved by downscaling large-scale or global models, the so-called parent models, to high-resolution regional models, the child models (see, e.g. Hermann et al., 2019; Holt et al., 2016; Laurent et al., 2018; Lavoie et al., 2020). For future projections, the obvious approach is to downscale ESMs. Since simulation of the fine-scale processes in the child model is strongly influenced by the parent model, it is important to assess the skill of ESMs in reproducing historical observations prior to using them for downscaled future projections. Rickard et al. (2016) ranked ESMs based on their misfit with regional observations around New Zealand in order to discard models with significant errors and determine an ensemble of "best" models that can be used to study regional climate projections, either directly or indirectly through regional downscaling. Here, we take a similar approach.

Our main objective is to assess the performance of a number of available ESMs in reproducing present conditions on the NWA shelf in contrast to a high-resolution regional model. This is an important information for users of historical and future projections in the region. Additionally, we assess ESMs performance along the boundaries of the regional model. This information is necessary when downscaling with a regional model. More specifically, we compare 29 ESMs used in the two most recent IPCC Assessment Rounds (AR) as part of the Coupled Model Intercomparison Project 5 (CMIP5; Taylor et al., 2012) and its currently ongoing successor CMIP6 (Eyring et al., 2016). We carry out a systematic and quantitative assessment and ranking by comparing the CMIP5 and CMIP6 models against observed surface temperature, salinity, chlorophyll, and nitrate and perform the same comparisons for a regional biogeochemical model. The latter is the Atlantic Canada Model (ACM, Brennan et al., 2016; Rutherford and Fennel, 2018) with biogeochemistry (Bianucci et al., 2016; Kuhn and Fennel, 2019) and is intended for regional downscaling of ESM simulations in order to generate high-resolution future projections. For all models, we present statistical metrics based on the mismatch of each model with climatological surface observations of temperature, salinity, nitrate, and chlorophyll and a ranking based on these metrics. The regional model is further evaluated against in-situ measurements, including high-resolution cross-shelf glider transects. The comparison provides an overview of ESM performance in the NWA and shows sufficient confidence for only a third of the ESMs. The regional model clearly outperformed all the global models and regional downscaling using single ESM forcing (as opposed to an ensemble) is recommended.

2 Material and Methods

2.1 Models

2.1.1 Global models

The CMIP5 and CMIP6 framework provides state-of-the-art climate model datasets from the previous (AR5) and current (AR6) IPCC Assessment Rounds (Eyring et al., 2016; Taylor et al., 2012). Of all the ESMs, those that include ocean biogeochemistry with monthly outputs of surface temperature, salinity, chlorophyll and nitrate were included in our

Deleted: Increased coastal model resolution can be achieved by downscaling large-scale or global models, the so-called parent models, to high-resolution regional models, the child models (see, e.g. Hermann et al. 2019, Holt et al. 2016, Laurent et al. 2018). For future projections, the obvious approach is to downscale ESMs. Since simulation of the fine-scale processes in the child model is strongly influenced by the parent model, it is important to assess the skill of ESMs in reproducing historical observations prior to using them for downscaled future projections. Rickard et al. (2016) ranked ESMs based on their misfit with regional observations around New Zealand in order to discard models with significant errors and determine an ensemble of "best" models that can be used to study regional climate projections. Here, we take a similar approach.¶

Deleted: few

Deleted: that their historical simulations can be used for downscaled future projections in the NWA. The use of ESM ensembles is not recommended.

Deleted: ,

Deleted: , and temperature

comparison. A total of 29 such ESMs were available (Table 1), 17 from CMIP5 (models 2–18, downloaded from the Earth System Grid Federation (ESGF) data repository at https://esgf-node.llnl.gov/search/cmip5/) and 12 from CMIP6 (models 19–30, downloaded from the ESGF data repository at https://esgf-node.llnl.gov/search/cmip6/). These models vary in their horizontal and vertical resolution and include a total of 13 different ocean biogeochemical models of varying levels of complexity (Table 1 and references therein).

We accessed the historical simulations which were forced by observed atmospheric composition and land cover changes over the period ~1850–2005 (CMIP5) and ~1850–2014 (CMIP6). Monthly, spatially resolved climatologies of surface chlorophyll, nitrate, and temperature were calculated over 30 years (1975–2005) from each ESM historical simulation.

2.1.2 Regional model

The ACM is a high-resolution, regional configuration of the Regional Ocean Modeling System (ROMS, version 3.5; Haidvogel et al., 2008) for the NWA, nested within the larger ocean-ice model of Urrego-Blanco and Sheng (2012), that includes the Gulf of Maine, Scotian Shelf and Grand Banks (Figure 1). The coupled physical-biogeochemical model has 30 vertical layers and an average horizontal resolution of 9.5 km on the shelf (Table 1). Detailed descriptions and physical model validation are presented in Brennan et al. (2016) and Rutherford and Fennel (2018). The biogeochemical model is based on Fennel et al. (2006, 2008) but was expanded by splitting phytoplankton and zooplankton state variables into size-based functional groups, i.e. nano-micro-phytoplankton and micro-meso-zooplankton. The model was also modified by including temperature-dependent biological rates for nutrient uptake, phytoplankton and zooplankton mortality, grazing and zooplankton egestion and excretion (see supporting text). The model has 10 state variables: nitrate, ammonium, and two size classes each for phytoplankton, chlorophyll, zooplankton and detritus (Figure 2). This ecosystem structure is of intermediate complexity similar to the model of Aumont et al. (2015), which is used in 6 of the ESMs included in our study. Model parameters were optimized by Kuhn (2017) and are listed in supporting Table S1. The model description and equations are available in the Supporting Information.

Initial and open boundary conditions for nitrate (NO_3) were defined from a monthly climatology (Kuhn, 2017) based on in-situ observations and the World Ocean Atlas 2009 (Garcia et al., 2010). Other biological variables were set to 0.1 mmol N m^{-3} with a phytoplankton-to-chlorophyll ratio of 0.76 mmol N (mg Chl)$^{-1}$ (Bianucci et al., 2016). The model was initialized on January 1, 1999 and run through December 31, 2014. The first year was considered spin up. Monthly climatologies of surface chlorophyll, nitrate, and temperature were calculated for comparison with the ESMs.

2.1.3 Model resolution

The 30 models differ dramatically in their horizontal resolution and do not evenly cover the 3 regions of interest (Figure 3, Table 1). The regional ACM has a much higher resolution than any of the ESMs with about 16 times more horizontal grid cells than the highest resolution ESM and almost 300 times more than the lowest resolution ESM. Among the ESMs the highest resolution is achieved by models 16 and 28, which share the same grid. These two have more than twice the number of horizontal grid cells than the next highest resolution models (3, 18, 20–21). The lowest resolution ESMs are models 3 and 12–14 with only 26 horizontal grid cells within the NWA shelf resulting in a coarse representation, particularly in the SS region. The median number of grid cells in the NWA shelf region is 72 and 102 for the CMIP5 and CMIP6 models, respectively, compared to 6875 in the ACM.

2.2 Observations

Four types of observations were used in the model intercomparison: 1) satellite surface chlorophyll observations from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) as 8-day averaged maps at 1/12˚ resolution (1999–2010, https://doi.org/data/10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018), 2) surface nitrate from the World Ocean Atlas 2013 (WOA; Garcia et al., 2014) at 1˚ resolution, 3) daily surface temperature from the Operational SST and Sea Ice Analysis (OSTIA) system (Donlon et al., 2012) at 1/20˚ resolution (2006–2016, https://doi.org/10.5067/GHOST-4FK01) and 4) surface salinity from the WOA at 1/4˚ resolution (Zweng et al., 2013). Monthly climatologies were calculated for each of these.

In addition, the regional model was validated using high-resolution in-situ observations along the Halifax Line (Figure 1) from the Atlantic Zone Monitoring Program (AZMP, 2000–2014, http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/azmp-pmza/index-eng.html) and glider transects between 2011 and 2016 (Ross et al., 2017). To enable a quantitative comparison between the glider and ACM data (Table 3), we spatially interpolated both datasets onto a transect following the Halifax Line (black line in Figure 1). Glider missions were seasonal and therefore both glider and AZMP transects data were seasonally averaged. For each mission, data were extracted at Station 2 to produce a monthly climatology.

Deleted: Three

Deleted: 2009

Deleted: (WOA; Garcia et al., 2010) at 1˚ resolution, and 3)

Deleted:).

Deleted: In addition, the regional model was validated using high-resolution in-situ observations from the Atlantic Zone Monitoring Program (AZMP, http://www.meds-sdmm.dfo-mpo.gc.ca/isdm-gdsi/azmp-pmza/index-eng.html) and glider transects along the Halifax Line between 2011 and 2016 (Ross et al., 2017). Transect data from AZMP and gliders were seasonally averaged along the Halifax Line (Figure 1).¶

2.3 Comparison metrics

For comparison with the observations, each model was mapped onto the SeaWiFS, WOA and OSTIA grids using a nearest neighbor interpolation. Since some areas, such as the nearshore and the Bay of Fundy, are covered by only a few models, grid cells that are active in less than 85% of all models were excluded from the analysis to avoid biases. In the low-resolution WOA climatology, the months November to January were excluded because poor data availability in these months resulted in unrealistic patterns.

Deleted: .

Three zones were defined for a high-level comparison with the observations on the shelf: the Gulf of Maine (GoM), Scotian Shelf (SS), and Grand Banks (GB) (Figure 1). Subsequently, the term NWA shelf refers to the region covered by all 3 zones (GoM, SS and GB). An additional zone was also defined for a high-level comparison with the observations along the open boundaries of the ACM.

Following the method of Rickard et al. (2016), a score S was calculated for each model variable, υ (i.e., surface temperature, chlorophyll, and nitrate), for each month, t, in the climatology as the sum of the centered Root Mean Square Difference (RMSD) and bias between the observations (x) and the model (y), such that:

Deleted: is

$$S(t,v) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i(t,v) - \bar{x}(t,v)\right) - \left(y_i(t,v) - \bar{y}(t,v)\right)\right)^2} + \frac{1}{n}\left|\sum_{i=1}^{n}\left(x_i(t,v) - y_i(t,v)\right)\right|$$

where the index i refers to a grid cell and n is the total number of grid cells within the NWA shelf. The lower the score the better the match between model and observations. Annual mean scores $\bar{S}(v)$ were calculated for each model variable by averaging over t. For each variable, the models were ranked based on their annual mean score. The overall rank was determined by ranking models by the averages of their ranks for surface temperature, salinity, chlorophyll, and nitrate ($\bar{R}$). For models with equal averages the ranking was determined by the average of chlorophyll and nitrate ranks ($\bar{R}_{bio}$).

To facilitate the comparison with observations, the ESMs were grouped into CMIP5 and CMIP6 and the ensemble means of all models and of the 5 highest ranked models were calculated for each group.

3 Results

Models and model ensembles are first compared with observations to assess their ability to reproduce the annual cycles of surface temperature, salinity, chlorophyll and nitrate in the NWA region. Error statistics are then analyzed to understand how the models deviate from each observed variable and subsequently used to calculate the scores and then rank the models. Finally, additional, high-resolution comparisons between models and observations are presented to further assess the regional model's performance.

3.1 Model-data comparisons

First, we compare the spatially averaged climatological surface temperature (Figure 4&Figure 5a–c), salinity (Figure 4&Figure 5d-f), chlorophyll (Figure 4&Figure 5g–i) and nitrate (Figure 4&Figure 5j–l) in our 3 regions of interest. The ESMs reasonably reproduce the annual cycle of surface temperature, but the annual cycles of salinity, chlorophyll and nitrate are not simulated well in any of them (see supporting Figures S1- S5) and the range of simulated salinity and biological properties is large.

Temperature is relatively consistent between model ensembles (Figure 4a–c), but with large variability between models (Figure 5a–c). An annual, positive bias occurs in the GoM (bias = +2.30˚C, Figure 4a), whereas temperatures are overestimated in winter (Dec–Feb) on the SS and GB (bias = +1.95 and +0.94˚C respectively, Figure 4a–c) and underestimated in summer (Jun–Aug) on GB (–1.53˚C, Figure 4f).

The range of simulated surface salinity is large (Figure 4d–f). Most models overestimate salinity in the GoM (bias = +1.46, Figure 4d). The mismatch is large on the SS and GB but not consistent among models, except for an annual, positive bias in CMIP6 models (bias = +1.42 and +0.76 respectively, Figure 4e–f). In the two latter regions, the biases in CMIP5 models compensate each other, resulting in an ensemble mean close to the observations.

For surface chlorophyll, there is a large discrepancy between the model ensembles and observations (Figure 4g–i). Inter-model differences are largest for the time of maxima and magnitude of the spring and fall blooms (Figure 5g–i, supporting Figures

Deleted: .
Deleted: .
Deleted: -
Deleted: chlorophyll
Deleted:) and nitrate
Deleted: g-i
Deleted: GM
Deleted: 28˚C
Deleted: winter
Deleted: 2.32
Deleted: 1.26˚C
Deleted: (bias = +2.32 and –
Deleted: 93˚C
Deleted: d–f
Deleted: d–f

S1–S5g–i). Standard deviations for the magnitude of the spring bloom are large among ESMs in the 3 zones (SD=0.6, 0.81 and 0.83 mg m^{-3} in GoM, SS and GB, respectively). The maxima of the spring bloom also vary significantly in time among the models, with a standard deviation among ESMs for the time of maxima of the bloom of about 1.5 months (SD=1.15, 1.59 and 1.62 months in GoM, SS and GB, respectively). Most models in the CMIP5 group do not simulate a fall bloom, hence none is present in the ESM ensemble mean, but rather a fall/winter increase in chlorophyll concentrations. Among the CMIP6 group, only models 23–25 generate a fall bloom (see supporting Figures S4–S5 g–i). Overall, the ESMs underestimate annual surface chlorophyll concentrations (bias = –0.94, –0.50 and –0.29 mg m^{-3} for GoM, SS and GB, respectively, Figure 4g–i). The chlorophyll bias is about 20% smaller in the CMIP6 group compared to CMIP5. There are also large discrepancies between the model ensembles and observations for nitrate (Figure 4j–l), particularly in the CMIP5 group. The variability in nitrate concentrations among the ESMs is also large (SD = 2.80 mmol m^{-3}) but smaller by 29% in the CMIP6 group. Most of the models reproduce the seasonal variability of surface nitrate (Figure 5j–l, supporting Figures S1–S5j–l); however, the CMIP5 models tend to underestimate fall-winter concentrations (winter bias = –1.28 mmol m^{-3}), whereas the CMIP6 model group performs better but with some mismatches in the timing of the seasonal changes (spring, fall). Note that since Nov.–Jan. nitrate WOA observations were excluded from the analysis (see section 2.3), winter observations are only available in February in the Gulf of Maine and in December and January in Grand Banks. A few models markedly overestimate surface nitrate concentrations in the NWA shelf regions (see supporting Figures S1, S3–5), including within the CMIP6 group. Supporting Figures S6–S9 provide an illustration of the model variability for chlorophyll and nitrate in March (Figures S6 and S7) and October (Figures S8 and S9), i.e. around the time of the spring and fall blooms respectively. The regional ACM well reproduces the annual cycle of surface temperature (Figure 4a–c), salinity (Figure 4d–f), chlorophyll (Figure 4g–i) and nitrate (Figure 4j–l) in the three regions. The model correctly simulates the overall magnitude of temperature and chlorophyll biomass, the timing of the maxima of spring and fall blooms and the latitudinal variations in temperature, salinity, chlorophyll and nitrate, although the magnitude of the spring bloom in the GoM and GB regions is underestimated. Late summer surface salinity is slightly overestimated on the SS and GB.

3.2 Model statistics

Error statistics, i.e. RMSD and bias, are now analyzed and used to calculate the model scores. The distribution and relationships between scores are explored and then the ranks calculated.

Except for the relationship between temperature and salinity RMSD (r = 0.82, p <0.001), the RMSD between the spatially averaged climatological observations and models are not consistent between variables, as indicated by the increasing temperature RMSD in Figure 6. However, temperature and chlorophyll RMSD are weakly correlated (r = 0.50, p = 0.005). For temperature and salinity, models 3, 20–21, and 24–25 have the largest discrepancy with observations and some clearly represent better the annual cycle than others. The best models for temperature (5–6, 14, 16 28) do not always match the best for salinity (5, 16, 27–28, 30). For chlorophyll, the largest discrepancies with observations are in models 4, 8 14 and 19–21, but overall chlorophyll RMSD are relatively large and homogeneous, except for a few models that have lower RMSD (e.g.

Deleted: - S5
Deleted: varies
Deleted: 93
Deleted: 49
Deleted: GM
Deleted: a-c
Deleted: g–i
Deleted: g–i
Deleted: S5
Deleted: -c
Deleted: d-f
Deleted: g-i
Deleted: gradients
Deleted: GM
Deleted: The
Deleted: 51
Deleted: 0043

models 22–23). Interestingly, the magnitude of the spring bloom in model 18 (CMIP5 group) is somewhat close to the observations. However, the time shift of the bloom (May–June) results in a poor agreement with observations. The mismatch between observed and simulated nitrate is much higher for models 5, 7, 18 and 29 and some models are much better at representing the observed annual cycle (Figure 6), as indicated by the lower RMSD. The RMSDs of the ACM are about a third of the average RMSD of the ESMs for both chlorophyll (ESM RMSDs are ×2.0–4.1 that of the ACM) and nitrate (×1.4–11.4), a quarter for temperature (×1.1–10.4) and 13% for salinity (×1.3–15.5).

Model scores (see Sect. 2.3) represent the spatial and temporal mismatch within the NWA shelf region (Figure 7). In general, the scores provide similar results as the RMSDs in Figure 6, although groups tend to emerge from the score calculation. As observed previously in Figure 6, the scores of ESMs have a much larger range of variability for temperature (1.5–7.8), salinity (0.5–4.2) and nitrate (1.4–13.2) than for chlorophyll (0.81–1.42) due to the large mismatch observed with a few models (Figure 7, supporting Figures S1–S5). For temperature, 4 of the 6 poorest (largest) scores (> 4.5) are in the CMIP6 group. They all markedly overestimate temperature, especially in the GoM (see supporting Figures S1, S4–S5), except for model 4 that underestimates temperature in SS and GB. The other models have also the poorest scores with respect to salinity. They all largely overestimate salinity in the 3 regions and are clearly outliers with respect to their CMIP category. The range of variability in chlorophyll scores did not reduce from CMIP5 to CMIP6 and given the relatively low scores of a few CMIP6 models (i.e. 22 and 23), the range is larger in the CMIP6 group (0.8–1.4, Figure 7, right panel) than in the CMIP5 group (1–1.4, Figure 7, left panel). With the exception of model 29, which has a very poor (high) score for nitrate, the range of variability in nitrate is reduced in the CMIP6 group. In total, 5 models (3, 5, 7, 18, 29) have very poor scores for nitrate (> 4) strongly overestimating surface nitrate, except for model 3 in the Gulf of Maine (see supporting Figure S1j-l). The remaining models have more homogeneous nitrate scores (Figure 7) with the best (lowest) scores in models 25, 24, 9 and 6 (Table 2). Models that underestimate nitrate (2, 8, 14 and 19, see supporting Figures S1–S4) have a better score because they match the low nitrate observations in late spring–summer (Table 2). Overall, ACM has the best scores, $S_{(v)}$, for temperature (1.14), salinity (0.48), chlorophyll (0.64) and nitrate (1.27).

Among the 4 variables, and including the regional model, we found a correlation between the scores of temperature and salinity ($r = 0.74$, $p < 0.001$), as well as weak correlations between chlorophyll and temperature ($r = 0.53$, $p = 0.0025$) or salinity ($r = 0.42$, $p = 0.02$). There was no correlations between nitrate and chlorophyll ($r = 0.03$, $p = 0.86$), and nitrate and temperature ($r = 0.05$, $p = 0.78$) or salinity ($r = 0.003$, $p = 0.99$). As can be seen in Figure 6, the ESMs with a poor representation of nitrate are not necessarily performing poorly with respect to the other variables. Model 7 for instance has the poorest score for nitrate and a relatively poor score for temperature and salinity but the best score of the CMIP5 group for chlorophyll (Figure 7, left panel). Model 5 has a poor score for nitrate but among the best scores for temperature and salinity. In fact, only models 3 and 18 have poor scores for all variables. Similarly, models 24 and 25 have the best scores for chlorophyll but are among the worst for temperature and salinity. On average, models have worse scores in the GoM (3.99, 2.49, 1.73, 3.15) than on the SS (3.36, 2.35, 0.94, 2.22) and GB (2.53, 1.41, 0.72, 2.47) for temperature, salinity, chlorophyll and nitrate, respectively.

Deleted:). The models with lowest RMSD for all 3 parameters are models 22–23 (CMIP6 group).

Deleted:) and

Deleted: GM

Deleted:).

Deleted: improvement

Deleted: S1

Deleted: 3

Deleted:), but not

Deleted: 88) or

Deleted: 06

Deleted: 74

Deleted: chlorophyll or temperature.

Deleted: among the ESMs

Deleted: .

Deleted: GM

Deleted: 97

Deleted: 46

Overall, 4 groups emerge on the chlorophyll-nitrate space in Figure 7. This grouping is somewhat arbitrary but provides a "biological" focus on model performance that can be related to the biological ranking ($\mathcal{R}_{bio}$) in Table 2. It also follows the general ranking presented in Figure 8, with a few exceptions. Group A includes 11 of the 14 best models (5 CMIP5 and 6 CMIP6) except for model 9 and 24–25 whose rankings are degraded due to poor representation of temperature and salinity. Within the 14 best models, the 3 models that are not included in Group A are model 5 and 15–16, which have mid to poor nitrate scores but are among the best models for temperature and salinity. Group B includes 4 intermediate-score models with respect to biology (15, 16, 17, 2). Group C includes the 8 models with poor chlorophyll scores (5 CMIP5 and 3 CMIP6) and Group D the 5 models with poor nitrate scores (4 CMIP5 and 1 CMIP6). Most of the models with poor scores for temperature and/or salinity are included in Group C, i.e. with the poor chlorophyll scores.

The overall model ranking (average of temperature, salinity, chlorophyll and nitrate ranks) indicates the gap between ACM and ESMs, as well as within ESMs (Figure 8). As expected, ACM ranks first, following the best scores for both chlorophyll and nitrate. The gap between ACM and model 28 (ESM with best $\mathcal{R}$ and $\mathcal{R}_{bio}$, Table 2) indicates that none of the ESM performs best for all fields, especially for both chlorophyll and nitrate. This is also shown by the large range in individual ranks (dark grey lines in Figure 8) in most models. Group A includes the 8 best ranking models, 2 from CMIP5 (6, 10) and 6 from CMIP6 (28, 23, 22, 26, 27, 30, respectively). The most consistent in term of individual and overall ranking is model 28 (best ESM), the other ones having a relatively large spread. On the other side of the spectrum, models 18, 20, 3 and 21 (Groups C and D) have the poorest ranks because of their consistently poor scores. Model 2 has also consistent poor ranks for all variables. Despite its poor performance with respect to nitrate, model 29 is ranked within the mid-range of the ESMs because of the better performance with respect to the other variables (ranks 8–15); model 7 has consistently poor performances except for chlorophyll (rank 4).

Model scores and ranking were also calculated along the boundaries of the regional model (see supporting Figure S10). The ranking shows that model performance on the shelf is not necessarily indicative of the performance along the boundaries of the regional model (supporting Figure S11, Table S2). Moreover, individual rankings are much more variable at the boundaries, even for the best performing models. The 8 best ESMs along the boundaries (22, 11, 30, 28, 16, 10, 26, 6) have an average rank of 9.2–10.5. There are no significant correlations between individual rankings, including temperature and salinity. Nonetheless, there is some agreement between the shelf and the outer boundary ranking for chlorophyll (ρ = 0.80), nitrate (ρ = 0.81) and salinity (ρ = 0.81, supporting Figure S12 and Table S3). Interestingly, the agreement is better with CMIP6 models (Table S3). However, there is no agreement for temperature. A similar pattern is found for individual boundaries (Figure S13). In this case, and apart from temperature, the model ranks along the northeastern boundary agree the most with those from the shelf.

3.3 Additional model-data comparisons for regional ACM

While the resolution of the ESMs does not allow for a comparison at smaller spatial scales, we further compare the regional ACM to cross-shelf transects and station observations (Figure 9) along the Halifax Line (see Figure 1). The ACM reproduces

Deleted: 6
Deleted: 8
Deleted: 30
Deleted: ranking is
Deleted: .
Deleted: the
Deleted: ,
Deleted: and temperature
Deleted: 22 (the
Deleted: overall ESM)
Deleted: 5
Deleted: all
Deleted: 22,
Deleted: 25, 24,
Deleted: chlorophyll
Deleted: nitrate
Deleted: ,
Deleted: The best ranked CMIP5 models are 10 and 13.
Deleted: , 21
Deleted: 18
Deleted: chlorophyll and nitrate.
Deleted: their
Deleted: models 7 and
Deleted: are
Deleted: they are among
Deleted: best ESMs
Deleted: and 8, respectively
Formatted: Font: 9 pt, English (UK)

the seasonal variation and the vertical gradient in chlorophyll and nitrate along the transect (Figure 9), although the simulated distributions are smoother than the glider observations. The summer subsurface chlorophyll maximum is located at the appropriate depth (28 m simulated versus 32 m observed, on average). The ACM somewhat underestimates the depth of the nitracline in the offshore waters (34 m versus 43 m, $x > 150$ km) and overestimates surface nitrate in spring and fall, as seen in Figure 4.

Station 2, which is located nearshore on the Halifax Line (see Figure 1), provides additional, vertically resolved information with high temporal resolution that is useful for model validation (Figure 10). At this location, the ACM reproduces the annual cycle of chlorophyll and nitrate. Surface and subsurface nitrate and chlorophyll are qualitatively reproduced in all seasons except during the spring bloom, which is more pronounced and reaches deeper in the observations, although the magnitude and vertical distribution of chlorophyll concentration agree well with the glider observations at this time.

A quantitative, point-to-point comparison of the ACM with the time series and glider observations along the Halifax Line (Figure 9) and at Station 2 (Figure 10) is provided in Table 3. The comparison indicates relatively high correlations between the ACM and time series of chlorophyll (0.68–0.78) and nitrate (0.83–0.92) along the Halifax Line as well as glider measurements of chlorophyll (0.85–0.94) for all seasons. Correlations are high as well at Station 2 for nitrate time series and glider measurements of chlorophyll. The largest discrepancies with observations are found with the time series of chlorophyll in spring. These results indicate an overall good skill of the model to reproduce the seasonal, vertically resolved observations on the Scotian Shelf.

4 Discussion

4.1 Overall model performance on the shelf

There are significant discrepancies with observations and a large variability among ESMs in the representation of surface temperature, salinity, chlorophyll and nitrate in the NWA shelf (Table 2, Figure 6 and supporting Figures S1–S5). A warm bias and a general overestimation of surface salinity in most models indicate a mismatch in the location of the Gulf Stream that influences conditions on the shelf, in line with the previous results of Loder et al. (2015) and Saba et al. (2016). Chlorophyll concentration was also systematically underestimated, whereas surface nitrate concentration is relatively variable between models. These patterns agree with the qualitative assessment of Lavoie et al. (2013, 2019). The spring and fall blooms, which are characteristic annual features of the NWA region (Greenan et al., 2004, 2008) are absent in some and most models, respectively. The correlation between temperature and chlorophyll scores (and to a lesser extent salinity) and the concomitant poor scores in chlorophyll and temperature/salinity (i.e. Group C in Figure 7) indicate that errors in surface chlorophyll concentration are partly driven by a misrepresentation of the general circulation and, more generally, of ocean physics. The improvement in chlorophyll from CMIP5 to CMIP6 without an associated improvement in temperature suggest that the errors in surface chlorophyll were also driven to some extent by errors in the biogeochemical model component. Lavoie et al. (2019) indicated that the misrepresentation of primary production in the NWA may be associated with the misrepresentation of

Deleted: ,

Deleted: resulting from

Deleted: was present in most models

Field Code Changed

Deleted: Chlorophyll concentration was also systematically underestimated.

Formatted: English (UK)

Deleted: were

Formatted: English (US)

Deleted: indicated that errors in surface chlorophyll concentration were likely driven by the misrepresentation of the general circulation and, more generally, of ocean physics particulate organic matter sinking and remineralization in the subsurface layer. They found an annual subsurface nitrate peak in CanESM2, GFDL-ESM2M, NorESM1-ME, CESM1-BGC (models 2, 7, 18 and 3, respectively) similar to the high surface nitrate found in this study (supporting Figures S1 and S3). However, all these models had poor scores in our assessment and therefore do not provide an appropriate representation of the biogeochemistry on the NWA shelf (Figure 8) or along the ACM boundaries (Figure S11). However, it is not possible, and beyond the scope of this work, for us to draw conclusions about the source of the regional mismatch in surface chlorophyll and nitrate in the ESMs.

Following Rickard et al. (2016), who used a similar ranking procedure, the 29 ESMs can be divided into an inner and an outer model ensemble of the NWA shelf. The outer ensemble includes 18 models that clearly misrepresent surface conditions in the NWA shelf (models 2–5, 7–9, 11, 14–15, 17–21, 24–25 and 29) and were selected as follows. The 7 models with lowest ranks (2–4, 8, 18, 20–21) were included because they consistently misrepresent all surface fields on the NWA shelf. Models 7, 9, 17 had poor scores for three variables, Model 15 was also included in the outer ensemble because of the misrepresentation of surface nitrate, whereas models 24–25 misrepresented temperature and salinity. Since nitrate scores neither correlate with chlorophyll nor temperature, the mismatch with nitrate observations is more likely related to intrinsic biogeochemical model behaviour rather than to a mismatch in circulation, as suggested by Lavoie et al. (2019). Models with persistent positive or negative biases in surface nitrate (4–5, 7–8, 11, 14, 19 and 29, Figures S1–S5) were selected because they misrepresent the seasonal nitrate dynamics and therefore the other biogeochemical variables driven by nitrate are questionable. Seven of the outer models were different generations (CMIP5 and CMIP6) of the same model, i.e. CanESM (2, 19), CESM (3, 20–21) and NorESM (18, 29), which had also low ranks along the ACM boundaries. Their large scores imply that they have fundamental issues with representing biogeochemistry in the NWA.

The inner ensemble includes 11 models (6, 10, 12–13, 16, 22–23, 26–28, 30, Table 1). Can those be used as a multi-model (optimal) ensemble to characterize the future state of the NWA shelf region? Unfortunately, we found that an ensemble mean of the best CMIP5 or CMIP6 models poorly represent historical surface fields due to the large variability within the ensemble (Figure 5) and the biases in the ensemble surface temperature, salinity and chlorophyll concentration (Figure 4). Model 28 (MPI-ESM1-2-HR, CMIP6) was the only ESM with good performances for all variables and is therefore the most appropriate to represent surface conditions in the NWA shelf.

The regional model clearly outperformed the ESMs in our assessment, with a consistent representation of the surface and subsurface fields in all shelf areas. The high spatial resolution of the regional model also allowed for a fine scale model validation that was not possible for the ESMs. The complementary glider transects and time series stations provide a high-resolution dataset of in-situ chlorophyll and nitrate concentrations and shows that the regional model resolves seasonal and vertical variation in chlorophyll and nitrate on the Scotian Shelf, something that none of the ESMs were able to reproduce.

4.2 Model performance along the regional model boundaries

The assessment of an ESM's performance on the NWA shelf, as presented above, is necessary prior to using its results, for example, to estimate historical and future trends in physical and biogeochemical tracers (Lavoie et al., 2013, 2019) and their

Formatted: English (UK)

Formatted: Font: +Headings CS (Times New Roman), English (UK)

Deleted: .

Deleted: 17

Deleted: 8

Deleted: 16, 18

Deleted: 8

Deleted: Five of those were different generations (CMIP5 and CMIP6) of the same model, i.e. CanESM (2, 19) and CESM (3, 20–21). Their large

Deleted: imply that CanESM and CESM have fundamental issues with representing biogeochemistry in the NWA. Models 15–16 and 24–25 were…

Deleted: their

Deleted: and

Deleted: , respectively

Deleted: likely related to intrinsic biogeochemical model behaviour rather than to a mismatch in circulation.

Formatted: Font: +Headings CS (Times New Roman), English (UK)

Deleted: 12

Deleted: 9–

Deleted: 17

Deleted:).

Deleted: these models, and even of

Deleted: five

Deleted: ,

Deleted: represents

Formatted: Font: +Headings CS (Times New Roman), English (UK)

Deleted: 4.2

effects on upper trophic levels (e.g., Bryndum-Buchholz et al., 2020b; Stortini et al., 2015). For regional downscaling, an ESM's performance along the boundaries of the regional model are critical (e.g., Lavoie et al., 2020). We found significant differences between model performance on the shelf and along the ACM boundaries and more variability in model performance for the latter. At the boundaries, all models have at least 1 variable that is poorly represented (Figure S11). Surprisingly, there is no relationship between ESM ranking on the shelf and at the ACM boundaries for temperature. Given the importance of large-scale circulation in the region some agreement was expected. The mismatch could be explained by a lesser control of large-scale currents on shelf temperature, although ESM biases for temperature and salinity on the shelf indicate the influence of the Gulf Stream. The agreement is better for the other variables (Table S3). Among the 10 best ESMs along the ACM boundaries, 8 are included in the inner ensemble described above; the best overall ESM on the shelf (model 28) is ranked third at the boundaries. Similarly, models with poor performances on the shelf (3, 18, 20–21) had also poor scores at the boundaries. The inner ensemble can therefore be used as a guide for ESM selection in the NWA region.

4.3 Uncertainties in score calculations

We used a heterogeneous dataset to calculate error statistics. Also, the regional model simulated the period 2000–2014, whereas the time range 1976–2005 was used with the CMIP models, for consistency in their comparison. For surface salinity, chlorophyll and nitrate, Lavoie et al. (2013) found negligible historical trends (1970s–2000s) in a multi-model comparison. For surface temperature, they found an increase in temperature <0.5˚C over this period, which is very small in comparison to the inter-model differences (Figures S1–5a–c). Also, surface temperature is overestimated in the GoM, whereas the trend would result in an underestimate. Hence, the scores should not be affected by time differences between model and observation datasets.

Since the period 2000–2014 is available for the CMIP6 models, we calculated the scores over this period to be consistent with the regional model simulation and the chlorophyll and temperature observations. The 2000–2014 scores are in agreement with the 1976–2005 scores described in section 3.2 (see supporting Figure S14), showing the robustness of our calculations despite the heterogeneous dataset. The only significant differences are with models 30 and 21, which have improved and degraded 2000–2014 scores for temperature, respectively. Model 21 remains at the last rank (Table 2) but the overall rank of model 30 (UKESM1-0-LL) could be somewhat higher than indicated in Figure 8

4.4 Impact of spatial resolution

In general, the coarse horizontal resolution of the ESMs affects the representation of the NWA region in comparison to the regional model, particularly on the relatively narrow Scotian Shelf. The poor representation of coastal areas is a known limitation of global models (Holt et al., 2017) and results in a global underestimation of primary productivity in these regions (Bopp et al., 2013; Schneider et al., 2008).

There is no correlation between grid resolution and ESM rank (Figure 11) despite the fact that the best overall ESM (MPI-ESM1-2-HR) has also the highest resolution (Table 1). This result shows that higher grid resolution, as called for by Lavoie et

Deleted:

Deleted: ranked

Deleted: Table 2).

al. (2013) for the NWA and by McKiver et al. (2015) for the global ocean, is necessary but is not a guarantee for improved model performance at this time. In fact, some very coarse resolution models from the CMIP5 group were ranked as well or better than the other models and models with the second highest resolution (3, 18, 20–21) had all low ranks. The improved ranks at constant (e.g. models 22, 24, 25, 28) and even lower (model 29) ocean grid resolution in the CMIP6 group (Table 2, Figure 12) was also an indication that the discrepancies with observations, and the improvement in the CMIP6 models (see below), were not associated with the ocean grid resolution but rather resulted from the physical and biogeochemical setup of the models. Another hint at the lack of relationship between resolution and model rank is the limited improvement with the high-resolution MPI model in the CMIP5 group (MPI-ESM-MR), despite higher model grid resolution compared to its lower-resolution counterpart (MPI-ESM-LR, Table 2). The lack of correlation between model resolution and performance on the NWA shelf is not surprising as all ESMs are still coarse and do not explicitly resolve shelf-scale processes but rather rely on their parameterisation. Much higher resolution will be necessary to refine the projections in coastal areas (e.g., Holt et al. (2017), Saba et al. (2016)), which is not currently computationally feasible in ESMs (Holt et al., 2009, 2017).

4.5 Impact of biogeochemical model structure

Although model performance is likely influenced by the biogeochemical model structure, we did not find a clear relationship between the type of biogeochemical model and performance. Here we only refer to the model type because the same model may have different parameterizations when used by different groups. While the inner and outer ensembles share only 3 biogeochemical models (PISCES, HAMOCC, TOPAZ2) out of 13, there was no indication of consistently better performance for the biogeochemical models in the inner ensemble. For example, models using similar ocean biogeochemistry (e.g., PISCES: 5, 12–14 (CMIP5), 22 and 26 (CMIP6), and HAMOCC: 15–16, 18 (CMIP5), 28–29 (CMIP6)) had very different ranks, with no obvious relationship between overall model rank and the ocean biogeochemical model component. Moreover, 5 and 4 biogeochemical models were represented in the 5 best ranked ESMs on the NWA shelf and outer ACM boundaries, respectively, similar to previous findings by Rickard et al. (2016). Lavoie et al. (2019) suggested that the PISCES biogeochemical model may underestimate subsurface remineralization in the CNRM and IPSL models, resulting in low surface nutrients where the Gulf Stream detaches from the coast. Our rankings (shelf and offshore) do not support this hypothesis; high surface nitrate concentrations were present in the CNRM models (throughout the region) and the IPSL-CM5A models (around the GoM) (Figures S1–4, S7, S9).

4.6 Improvement from CMIP5 to CMIP6

Model performance improved in the new CMIP generation, but not uniformly across models and variables. We note that 2 of the 5 best models are from the CMIP5 for both the shelf and the ACM boundaries rankings. Therefore, with respect to historical conditions in the NWA region, CMIP6 models do not always have better performance. The average rank was not very different between the two CMIP groups, i.e. R̄ = 16.8 and 14.9 for CMIP5 and CMIP6, respectively (Figure 8, Table 2). The change in performance between the two generations of models can be assessed by evaluating the subset of models that are available for

Deleted: similar ranking of
Deleted: two
Deleted: models
Deleted: , MPI-ESM-LR and
Deleted: ,
Deleted: an important difference in
Deleted: (Figure 8).
Deleted: 3
Deleted: 4
Deleted: , NOBM
Formatted: English (US)
Deleted: ,
Formatted: English (US)
Deleted: ,
Deleted: ,
Formatted: English (US)
Deleted: ,
Formatted: English (US)
Deleted:)
Formatted: English (US)
Deleted: 4
Deleted: The 4
Deleted: ranked ESMs were
Deleted: group, although the
Deleted: 17.4
Deleted: 0

CMIP5 and CMIP6. There are nine such models (Figure 12). All CMIP6 models have improved overall ranks, indicating better performance (Figure 12). The overall improvement was large only for models that had average to low ranks in the CMIP5 group (ranks 15–22, x-axis in Figure 12). Temperature and salinity did not improve except for GFDL-ESM2M and NorESM2-LM (and CanESM5 for salinity) and degraded in some cases. Models with poor scores for temperature and salinity (CESM2, GISS-E2-1-G-CC) had already poor scores in their CMIP5 version and therefore the cause of their poor performance is likely the same. The change in ranking is therefore mainly associated with better surface fields for chlorophyll and nitrate. This is particularly the case for model pairs 3, 5, 6 and 8, which ranked much better for chlorophyll (+8.2) and nitrate (+11.0) in the CMIP6 group (Figure 12). The chlorophyll rank in model pair 4 improved significantly (+18) but this improvement was counteracted by degraded temperature and nitrate ranks. The lack of general improvement in surface temperature indicates that the temperature bias detected in the CMIP5 group was not solved in CMIP6, as seen in Figure 4.

Deleted: 12.7

We can only speculate about the source of improvement in the CMIP6 models. For specific changes in the CMIP6 model versions, the reader is referred to the references listed in Table 1. Kwiatkowski et al. (2020) recently showed that projected surface temperature, nitrate and net primary production differ significantly in CMIP5 and CMIP6 model ensembles. Higher climate sensitivity in CMIP6 models partly explains this difference but the source of change in primary production was not resolved. In the historical simulations, better surface chlorophyll and nitrate fields in CNRM-ESM2-1 may be associated with the transition from a climate model with ocean biogeochemistry to a fully coupled ESM, even though such transition may degrade historical simulations due to the replacement of observations by prognostic schemes that are poorly constrained (Séférian et al., 2019). Updated land and ocean biogeochemistry may have improved the representation of surface chlorophyll and nitrate in MPI-ESM1-2-HR (Müller et al., 2018), whereas the improvement in surface temperature and nitrate fields from GFDL-ESM2M to GFDL-ESM4 seem to be associated with the physical ocean component of the model, given that GFDL-ESM2G already performed well in the CMIP5 group. Danabasoglu et al. (2020) found a significant improvement for CESM2 at the global scale but a poor representation of the Gulf Stream–North Atlantic Current system, resulting in a large surface temperature bias. This is in line with our assessment for the NWA shelf where both physical and biological parameters had poor scores and the model was not found appropriate for shelf studies in the NWA.

Deleted: explain

4.7 Other coastal regions

Deleted: 5

Our results may also apply for other coastal regions, given the poor representation of coastal areas in ESMs, but the details are probably region specific. Discrepancies with observations in the NWA are partly driven by poor representation of large-scale circulation features such as the Gulf Stream and Labrador Current in most of the models. The representation of large-scale currents may improve (or worsen) in other regions, resulting in a different ranking there. For example, Rickard et al. (2016) found a different model selection in the inner model ensemble around New Zealand. Seven (out of 11) of their inner ensemble models (models 2–5, 7–8, 14) are not included in our inner ensemble. Model 3, perhaps the best model in their assessment, ranked 29 out of 30 in the NWA shelf region (Figure 8, supporting Figure S1). The representation of the dynamic NWA circulation is a known issue in ESMs and further regional comparisons will be necessary to assess if our results are representative for the global coastal ocean.

Formatted: Font: +Headings CS (Times New Roman)

5 Conclusions

We evaluated the CMIP5 and CMIP6 ESMs with biogeochemistry for the NWA shelf. Arguably, only 1 model (MPI-ESM1-2-HR) had a consistently good performance for all variables. 11 ESMs with satisfactory overall performance in their historical simulations of the NWA shelf were included in a ranked inner ensemble to guide the use of ESMs in the region. Apart for temperature, the ESMs evaluation along the boundaries of the regional model was relatively similar to the evaluation on the shelf but with more variability. Most of the highly ranked models can therefore be used either directly or for regional downscaling. We caution against using model ensembles that had poor agreements with observations on the NWA shelf. The regional model (ACM) clearly outperformed the global models and is a good candidate for downscaled projections in combination with one of the top ranked ESMs. Further refinement in the ACM should focus on the mechanisms that determine the magnitude of the spring bloom.

Deleted: 3 models (CNRM-ESM2-

Deleted: , GFDL-ESM4 and

Deleted:),

Deleted: from CMIP6, provided sufficient confidence

Deleted: .

Deleted: , either directly or in downscaling future projections for

Similar comparisons should be carried out in coastal areas before using CMIP model projections. While it is not clear how the presented model ranking will hold in other regions, it is highly likely that some models do not perform well in coastal areas generally and should not be used for regional investigations.

Given the lack of a direct relationship between model skill and horizontal resolution, it is unlikely that feasible grid refinement will significantly improve model performance in the NWA region. The improvement in scores from CMIP5 to CMIP6 shows that refining ocean biogeochemical components can improve the model performance.

Code and data availability. The ROMS code and the observations are available from the links referenced in the manuscript.

Supplement link. The supplement related to this article is available on-line at:

Author contribution. AL and KF conceived the study. AL and AK refined a previous version of the ACM model. AL conducted the analyses. AL wrote the manuscript with input from KF

Deleted: set up

Competing interests. The authors declare that they have no conflict of interest

Acknowledgements. The ACM was run on Compute Canada resources under the resource allocation project qqh-593-ac.

Financial statement. We acknowledge funding from the Canada First Research Excellence Fund, through the Ocean Frontier Institute, the MEOPAR Network of Centres of Excellence through the Prediction Core, and an NSERC Discovery Grant held by KF.

References

Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., Hosaka, M., Sakami, T., Yoshimura, H., Hirabara, M., Shindo, E., Tsujino, H., Mizuta, R., Yabu, S., Koshiro, T., Ose, T. and Kitoh, A.: Basic performance of a new earth system model of the Meteorological Research Institute (MRI-ESM1), Pap. Meteorol. Geophys., 64, 1–19, doi:10.2467/mripapers.64.1, 2013.

Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., Jones, C., Jung, M., Myneni, R. and Zhu, Z.: Evaluating the Land and Ocean Components of the Global Carbon Cycle in the CMIP5 Earth System Models, J. Clim., 26(18), 6801–6843, doi:10.1175/JCLI-D-12-00417.1, 2013.

Arora, V. K., Scinocca, J. F., Boer, G. J., Christian, J. R., Denman, K. L., Flato, G. M., Kharin, V. V., Lee, W. G. and Merryfield, W. J.: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases, Geophys. Res. Lett., 38(5), L05805, doi:10.1029/2010GL046270, 2011.

Aumont, O. and Bopp, L.: Globalizing results from ocean in situ iron fertilization studies, Global Biogeochem. Cycles, 20(2), GB2017, doi:10.1029/2005GB002591, 2006.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L. and Gehlen, M.: PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8(8), 2465–2513, doi:10.5194/gmd-8-2465-2015, 2015.

Bianucci, L., Fennel, K., Chabot, D., Shackell, N. and Lavoie, D.: Ocean biogeochemical models as management tools: a case study for Atlantic wolffish and declining oxygen, ICES J. Mar. Sci. J. du Cons., 73(2), 263–274, doi:10.1093/icesjms/fsv220, 2016.

Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, Science (80-.)., 359(6375), eaam8328, doi:10.1126/science.aam8328, 2018.

Bopp, L., Resplandy, L., Orr, J. C., Doney, S. C., Dunne, J. P., Gehlen, M., Halloran, P., Heinze, C., Ilyina, T., Séférian, R., Tjiputra, J. and Vichi, M.: Multiple stressors of ocean ecosystems in the 21st century: projections with CMIP5 models, Biogeosciences, 10(10), 6225–6245, doi:10.5194/bg-10-6225-2013, 2013.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., Davini, P., Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J., Guenet, B., Guez, L., Guilyardi, É., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N. and Vuichard, N.: Presentation and evaluation of the IPSL-CM6A-LR climate model, J. Adv. Model. Earth Syst., doi:10.1029/2019MS002010, 2020.

Bourgeois, T., Orr, J. C., Resplandy, L., Terhaar, J., Ethé, C., Gehlen, M. and Bopp, L.: Coastal-ocean uptake of anthropogenic carbon, Biogeosciences, 13(14), 4167–4185, doi:10.5194/bg-13-4167-2016, 2016.

Brennan, C. E., Bianucci, L. and Fennel, K.: Sensitivity of Northwest North Atlantic shelf circulation to surface and boundary forcing: a regional model assessment, Atmosphere-Ocean, 54(3), 230–247, doi:10.1080/07055900.2016.1147416, 2016.

Bryndum-Buchholz, A., Boyce, D., Tittensor, D., Christensen, V., Bianchi, D. and Lotze, H.: Climate-change impacts and fisheries management challenges in the North Atlantic Ocean, Mar. Ecol. Prog. Ser., 648, 1–17, doi:10.3354/meps13438, 2020a.

Bryndum-Buchholz, A., Prentice, F., Tittensor, D. P., Blanchard, J. L., Cheung, W. W. L., Christensen, V., Galbraith, E. D., Maury, O. and Lotze, H. K.: Differing marine animal biomass shifts under 21st century climate change between Canada's three oceans, edited by B. Favaro, FACETS, 5(1), 105–122, doi:10.1139/facets-2019-0035, 2020b.

Cai, W.-J., Dai, M. and Wang, Y.: Air-sea exchange of carbon dioxide in ocean margins: A province-based synthesis, Geophys. Res. Lett., 33(12), L12603, doi:10.1029/2006GL026219, 2006.

Chen, C.-T. A., Huang, T.-H., Chen, Y.-C., Bai, Y., He, X. and Kang, Y.: Air–sea exchanges of CO_2 in the world's coastal seas, Biogeosciences, 10(10), 6509–6544, doi:10.5194/bg-10-6509-2013, 2013.

Christian, J. R., Arora, V. K., Boer, G. J., Curry, C. L., Zahariev, K., Denman, K. L., Flato, G. M., Lee, W. G., Merryfield, W. J., Roulet, N. T. and Scinocca, J. F.: The global carbon cycle in the Canadian Earth system model (CanESM1): Preindustrial control simulation, J. Geophys. Res., 115(G3), G03014, doi:10.1029/2008JG000920, 2010.

Claret, M., Galbraith, E. D., Palter, J. B., Bianchi, D., Fennel, K., Gilbert, D. and Dunne, J. P.: Rapid coastal deoxygenation due to ocean circulation shift in the northwest Atlantic, Nat. Clim. Chang., 8(10), 868–872, doi:10.1038/s41558-018-0263-1, 2018.

Collins, W. J., Bellouin, N., Doutriaux-Boucher, M., Gedney, N., Halloran, P., Hinton, T., Hughes, J., Jones, C. D., Joshi, M., Liddicoat, S., Martin, G., O'Connor, F., Rae, J., Senior, C., Sitch, S., Totterdell, I., Wiltshire, A. and Woodward, S.: Development and evaluation of an Earth-System model – HadGEM2, Geosci. Model Dev., 4(4), 1051–1075, doi:10.5194/gmd-4-1051-2011, 2011.

Danabasoglu, G., Lamarque, J. -F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J. and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), J. Adv. Model. Earth Syst., 12(2), doi:10.1029/2019MS001916, 2020.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E. and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sens. Environ., 116, 140–158, doi:10.1016/j.rse.2010.10.017, 2012.

Ducklow, H. and McCallister, S.: The biogeochemistry of carbon dioxide in the coastal oceans, in The Sea, Vol 13: The global coastal ocean. Multiscale interdisciplinary processes, edited by A. R. Robinson and K. H. Brink, pp. 269–315, John Wiley & Sons, New York., 2004.

Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N. and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5., 2013.

Dunne, J. P.: Technical description of Tracers of Ocean Phytoplankton with Allometric Zooplankton version 2 (TOPAZ2) used in GFDL's ESM2M and ESM2G submitted as part of the coupled model intercomparison project phase 5., 2013.

Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., Stouffer, R. J., Cooke, W., Dunne, K. A., Harrison, M. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Phillipps, P. J., Sentman, L. T., Samuels, B. L., Spelman, M. J., Winton, M., Wittenberg, A. T. and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part I: Physical Formulation and Baseline Simulation Characteristics, J. Clim., 25(19), 6646–6665, doi:10.1175/JCLI-D-11-00560.1, 2012.

Dunne, J. P., John, J. G., Shevliakova, E., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., Milly, P. C. D., Sentman, L. T., Adcroft, A. J., Cooke, W., Dunne, K. A., Griffies, S. M., Hallberg, R. W., Harrison, M. J., Levy, H., Wittenberg, A. T., Phillips, P. J. and Zadeh, N.: GFDL's ESM2 Global Coupled Climate–Carbon Earth System Models. Part II: Carbon System Formulation and Baseline Simulation Characteristics*, J. Clim., 26(7), 2247–2267, doi:10.1175/JCLI-D-12-00150.1, 2013.

Deleted: Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., Krasting, J. P., Malyshev, S., Naik, V., Paulot, F., Shevliakova, E., A.Stock, C., Zadeh, N., Balaji, V., Blanton, C., Dunne, K. A., Dupuis, C., Durachta, J., Dussin, R., Gauthier, P. P. G., Griffies, S. M., Guo, H., Hallberg, R. W., Harrison, M., J. He, W. H., McHugh, C., Menzel, R., Milly, P. C. D., Nikonov, S., Paynter, D. J., Ploshay, J., Radhakrishnan, A., Rand, K., Reichl, B. G., Robinson, T., Schwarzkopf, D. M., Sentman, L. A., Underwood, S., Vahlenkamp, H., Winton, M., Wittenberg, A. T., Wyman, B., Zeng, Y. and Zhao, M.: The GFDL Earth System Model version 4.1 (GFDL-ESM4.1): Model description and simulation characteristics, J. Adv. Model. Earth Syst., 2019MS0020, n.d.¶

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J. and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9(5), 1937–1958, doi:10.5194/gmd-9-1937-2016, 2016.

Fennel, K., Wilkin, J., Levin, J., Moisan, J., O'Reilly, J. and Haidvogel, D.: Nitrogen cycling in the Middle Atlantic Bight: Results from a three-dimensional model and implications for the North Atlantic nitrogen budget, Global Biogeochem. Cycles, 20(3), GB3007, doi:10.1029/2005GB002456, 2006.

Fennel, K., Wilkin, J., Previdi, M. and Najjar, R.: Denitrification effects on air-sea CO2 flux in the coastal ocean: Simulations for the northwest North Atlantic, Geophys. Res. Lett., 35(24), L24608, doi:10.1029/2008GL036147, 2008.

Fennel, K., Hu, J., Laurent, A., Marta-Almeida, M. and Hetland, R.: Sensitivity of hypoxia predictions for the northern Gulf of Mexico to sediment oxygen consumption and model nesting, J. Geophys. Res. Ocean., 118(2), 990–1002, doi:10.1002/jgrc.20077, 2013.

Fennel, K., Alin, S., Barbero, L., Evans, W., Bourgeois, T., Cooley, S., Dunne, J., Feely, R. A., Hernandez-Ayon, J. M., Hu, X., Lohrenz, S., Muller-Karger, F., Najjar, R., Robbins, L., Shadwick, E., Siedlecki, S., Steiner, N., Sutton, A., Turk, D., Vlahos, P. and Wang, Z. A.: Carbon cycling in the North American coastal ocean: a synthesis, Biogeosciences, 16(6), 1281–1304, doi:10.5194/bg-16-1281-2019, 2019.

Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Zweng, M. M., Baranova, O. K. and Johnson, D. R.: World Ocean Atlas 2009, Volume 4: Nutrients (phosphate, nitrate, silicate), NOAA Atlas., edited by S. Levitus, U.S. Government Printing

Office, Washington, D.C., 2010.
Garcia, H. E., Locarnini, R. A., Boyer, T. P., Antonov, J. I., Baranova, O. K., Zweng, M. M., Reagan, J. R. and Johnson, D. R.: World Ocean Atlas 2013, Volume 4: Dissolved Inorganic Nutrients (phosphate, nitrate, silicate), S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 76., 2014.
Gilbert, D., Rabalais, N. N., Díaz, R. J. and Zhang, J.: Evidence for greater oxygen decline rates in the coastal ocean than in the open ocean, Biogeosciences, 7, 2283–2296, doi:10.5194/bg-7-2283-2010, 2010.
Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., Brovkin, V., Crueger, T., Esch, M., Fieg, K., Glushak, K., Gayler, V., Haak, H., Hollweg, H.-D., Ilyina, T., Kinne, S., Kornblueh, L., Matei, D., Mauritsen, T., Mikolajewicz, U., Mueller, W., Notz, D., Pithan, F., Raddatz, T., Rast, S., Redler, R., Roeckner, E., Schmidt, H., Schnur, R., Segschneider, J., Six, K. D., Stockhause, M., Timmreck, C., Wegner, J., Widmann, H., Wieners, K.-H., Claussen, M., Marotzke, J. and Stevens, B.: Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, J. Adv. Model. Earth Syst., 5(3), 572–597, doi:10.1002/jame.20038, 2013.
Greenan, B. J. ., Petrie, B. ., Harrison, W. . and Oakey, N. .: Are the spring and fall blooms on the Scotian Shelf related to short-term physical events?, Cont. Shelf Res., 24(4–5), 603–625, doi:10.1016/j.csr.2003.11.006, 2004.
Greenan, B. J. W., Petrie, B. D., Harrison, W. G. and Strain, P. M.: The onset and evolution of a spring bloom on the Scotian Shelf, , 53(5), 1759–1775, 2008.
Greenan, B. J. W., Shackell, N. L., Ferguson, K., Greyson, P., Cogswell, A., Brickman, D., Wang, Z., Cook, A., Brennan, C. E. and Saba, V. S.: Climate Change Vulnerability of American Lobster Fishing Communities in Atlantic Canada, Front. Mar. Sci., 6, doi:10.3389/fmars.2019.00579, 2019.
Gruber, N., Hauri, C., Lachkar, Z., Loher, D., Frolicher, T. L. and Plattner, G.-K.: Rapid Progression of Ocean Acidification in the California Current System, Science (80-.)., 337(6091), 220–223, doi:10.1126/science.1216773, 2012.
Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., Ohgaito, R., Ito, A., Yamazaki, D., Okajima, H., Ito, A., Takata, K., Ogochi, K., Watanabe, S. and Kawamiya, M.: Description of the MIROC-ES2L Earth system model and evaluation of its climate–biogeochemical processes and feedbacks, Rev., doi:10.5194/gmd-2019-275, 2019.
Hermann, A. J., Gibson, G. A., Bond, N. A., Curchitser, E. N., Hedstrom, K., Cheng, W., Wang, M., Cokelet, E. D., Stabeno, P. J. and Aydin, K.: Projected future biophysical states of the Bering Sea, Deep Sea Res. Part II Top. Stud. Oceanogr., 134, 30–47, doi:10.1016/j.dsr2.2015.11.001, 2016.
Hermann, A. J., Gibson, G. A., Cheng, W., Ortiz, I., Aydin, K., Wang, M., Hollowed, A. B. and Holsman, K. K.: Projected biophysical conditions of the Bering Sea to 2100 under multiple emission scenarios, ICES J. Mar. Sci., 76(6), 1937–1937, doi:10.1093/icesjms/fsz111, 2019.
Holt, J., Harle, J., Proctor, R., Michel, S., Ashworth, M., Batstone, C., Allen, I., Holmes, R., Smyth, T., Haines, K., Bretherton, D. and Smith, G.: Modelling the global coastal ocean, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 367(1890), 939–951, doi:10.1098/rsta.2008.0210, 2009.
Holt, J., Schrum, C., Cannaby, H., Daewel, U., Allen, I., Artioli, Y., Bopp, L., Butenschon, M., Fach, B. A., Harle, J.,

Pushpadas, D., Salihoglu, B. and Wakelin, S.: Potential impacts of climate change on the primary production of regional seas: A comparative analysis of five European seas, Prog. Oceanogr., 140, 91–115, doi:10.1016/j.pocean.2015.11.004, 2016.

Holt, J., Hyder, P., Ashworth, M., Harle, J., Hewitt, H. T., Liu, H., New, A. L., Pickles, S., Porter, A., Popova, E., Allen, J. I., Siddorn, J. and Wood, R.: Prospects for improving the representation of coastal and shelf seas in global ocean models, Geosci. Model Dev., 10(1), 499–523, doi:10.5194/gmd-10-499-2017, 2017.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H. and Núñez-Riboni, I.: Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations, J. Adv. Model. Earth Syst., 5(2), 287–315, doi:10.1029/2012MS000178, 2013.

Kuhn, A. M.: Integration of observations and models for an improved understanding of marine ecosystem dynamics, Dalhousie University. [online] Available from: http://hdl.handle.net/10222/73354, 2017.

Kuhn, A. M. and Fennel, K.: Evaluating ecosystem model complexity for the northwest North Atlantic through surrogate-based optimization, Ocean Model., 142, 101437, doi:10.1016/j.ocemod.2019.101437, 2019.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A. and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, Biogeosciences, 17(13), 3439–3470, doi:10.5194/bg-17-3439-2020, 2020.

Lachkar, Z. and Gruber, N.: What controls biological production in coastal upwelling systems? Insights from a comparative modeling study, Biogeosciences, 8(10), 2961–2976, doi:10.5194/bg-8-2961-2011, 2011.

Laruelle, G. G., Cai, W. J., Hu, X., Gruber, N., Mackenzie, F. T. and Regnier, P.: Continental shelves as a variable but increasing global sink for atmospheric carbon dioxide, Nat. Commun., 9(1), doi:10.1038/s41467-017-02738-z, 2018.

Laurent, A., Fennel, K., Ko, D. S. and Lehrter, J.: Climate Change Projected to Exacerbate Impacts of Coastal Eutrophication in the Northern Gulf of Mexico, J. Geophys. Res. Ocean., 123(5), 3408–3426, doi:10.1002/2017JC013583, 2018.

Lavoie, D., Lambert, N., den Mustapha, S. and van der Baaren, A.: Projections of Future Physical and Biogeochemical Conditions in the Northwest Atlantic from CMIP5 Global Climate Models., 2013.

Lavoie, D., Lambert, N. and Gilbert, D.: Projections of Future Trends in Biogeochemical Conditions in the Northwest Atlantic Using CMIP5 Earth System Models, Atmosphere-Ocean, 57(1), 18–40, doi:10.1080/07055900.2017.1401973, 2019.

Lavoie, D., Lambert, N., Rousseau, S., Dumas, J., Chassé, J., Long, Z., Perrie, W., Starr, M., Brickman, D. and Azetsu-Scott, K.: Projections of future physical and biogeochemical conditions in the Gulf of St . Lawrence , on the Scotian Shelf and in the Gulf of Maine, Can. Tech. Rep. Hydrogr. Ocean Sci., 334: xiii, 2020.

Lindsay, K., Bonan, G. B., Doney, S. C., Hoffman, F. M., Lawrence, D. M., Long, M. C., Mahowald, N. M., Keith Moore, J., Randerson, J. T. and Thornton, P. E.: Preindustrial-Control and Twentieth-Century Carbon Cycle Experiments with the Earth System Model CESM1(BGC), J. Clim., 27(24), 8981–9005, doi:10.1175/JCLI-D-12-00565.1, 2014.

Loder, J. W., van der Baaren, A. and Yashayaev, I.: Climate Comparisons and Change Projections for the Northwest Atlantic

from Six CMIP5 Models, Atmosphere-Ocean, 53(5), 529–555, doi:10.1080/07055900.2015.1087836, 2015.
Mattern, J. P. and Edwards, C. A.: Simple parameter estimation for complex models — Testing evolutionary techniques on 3-dimensional biogeochemical ocean models, J. Mar. Syst., 165, 139–152, doi:10.1016/j.jmarsys.2016.10.012, 2017.
Moore, J. K., Lindsay, K., Doney, S. C., Long, M. C. and Misumi, K.: Marine Ecosystem Dynamics and Biogeochemical Cycling in the Community Earth System Model [CESM1(BGC)]: Comparison of the 1990s with the 2090s under the RCP4.5 and RCP8.5 Scenarios, J. Clim., 26(23), 9291–9312, doi:10.1175/JCLI-D-12-00566.1, 2013.
Muller-Karger, F. E.: The importance of continental margins in the global carbon cycle, Geophys. Res. Lett., 32(1), L01602, doi:10.1029/2004GL021346, 2005.
Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., Ilyina, T., Kleine, T., Kornblueh, L., Li, H., Modali, K., Notz, D., Pohlmann, H., Roeckner, E., Stemmler, I., Tian, F. and Marotzke, J.: A Higher-resolution Version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR), J. Adv. Model. Earth Syst., 10(7), 1383–1413, doi:10.1029/2017MS001217, 2018.
Palmer, J. R. and Totterdell, I. J.: Production and export in a global ocean ecosystem model, Deep Sea Res. Part I Oceanogr. Res. Pap., 48(5), 1169–1198, doi:10.1016/S0967-0637(00)00080-7, 2001.
OSTIA. UK Met Office. 2005. GHRSST Level 4 OSTIA Global Foundation Sea Surface Temperature Analysis. Ver. 1.0. PO.DAAC, CA, USA. doi:10.5067/GHOST-4FK01. Accessed on 2019/12/06.
Peña, M. A., Fine, I. and Callendar, W.: Interannual variability in primary production and shelf-offshore transport of nutrients along the northeast Pacific Ocean margin, Deep Sea Res. Part II Top. Stud. Oceanogr., 169–170, 104637, doi:10.1016/j.dsr2.2019.104637, 2019.
Rickard, G. J., Behrens, E. and Chiswell, S. M.: CMIP5 earth system models with biogeochemistry: An assessment for the southwest Pacific Ocean, J. Geophys. Res. Ocean., 121(10), 7857–7879, doi:10.1002/2016JC011736, 2016.
Romanou, A., Gregg, W. W., Romanski, J., Kelley, M., Bleck, R., Healy, R., Nazarenko, L., Russell, G., Schmidt, G. A., Sun, S. and Tausnev, N.: Natural air–sea flux of CO2 in simulations of the NASA-GISS climate model: Sensitivity to the physical ocean model formulation, Ocean Model., 66, 26–44, doi:10.1016/j.ocemod.2013.01.008, 2013.
Ross, T., Craig, S. E., Comeau, A., Davis, R., Dever, M. and Beck, M.: Blooms and subsurface phytoplankton layers on the Scotian Shelf: Insights from profiling gliders, J. Mar. Syst., 172, 118–127, doi:10.1016/j.jmarsys.2017.03.007, 2017.
Rousseaux, C. S. and Gregg, W. W.: Recent decadal trends in global phytoplankton composition, Global Biogeochem. Cycles, 29(10), 1674–1688, doi:10.1002/2015GB005139, 2015.
Rutherford, K. and Fennel, K.: Diagnosing transit times on the northwestern North Atlantic continental shelf, Ocean Sci., 14(5), 1207–1221, doi:10.5194/os-14-1207-2018, 2018.
Saba, V. S., Griffies, S. M., Anderson, W. G., Winton, M., Alexander, M. A., Delworth, T. L., Hare, J. A., Harrison, M. J., Rosati, A., Vecchi, G. A. and Zhang, R.: Enhanced warming of the Northwest Atlantic Ocean under climate change, J. Geophys. Res. Ocean., 121(1), 118–132, doi:10.1002/2015JC011346, 2016.
Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., Bleck,

R., Canuto, V., Chen, Y.-H., Cheng, Y., Clune, T. L., Del Genio, A., de Fainchtein, R., Faluvegi, G., Hansen, J. E., Healy, R. J., Kiang, N. Y., Koch, D., Lacis, A. A., LeGrande, A. N., Lerner, J., Lo, K. K., Matthews, E. E., Menon, S., Miller, R. L., Oinas, V., Oloso, A. O., Perlwitz, J. P., Puma, M. J., Putman, W. M., Rind, D., Romanou, A., Sato, M., Shindell, D. T., Sun, S., Syed, R. A., Tausnev, N., Tsigaridis, K., Unger, N., Voulgarakis, A., Yao, M.-S. and Zhang, J.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, J. Adv. Model. Earth Syst., 6(1), 141–184, doi:10.1002/2013MS000265, 2014.

Schneider, B., Bopp, L., Gehlen, M., Segschneider, J., Frölicher, T. L., Cadule, P., Friedlingstein, P., Doney, S. C., Behrenfeld, M. J. and Joos, F.: Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models, Biogeosciences, 5(2), 597–614, doi:10.5194/bg-5-597-2008, 2008.

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., Decharme, B., Delire, C., Berthet, S., Chevallier, M., Sénési, S., Franchisteguy, L., Vial, J., Mallet, M., Joetzjer, E., Geoffroy, O., Guérémy, J., Moine, M., Msadek, R., Ribes, A., Rocher, M., Roehrig, R., Salas-y-Mélia, D., Sanchez, E., Terray, L., Valcke, S., Waldman, R., Aumont, O., Bopp, L., Deshayes, J., Éthé, C. and Madec, G.: Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate, J. Adv. Model. Earth Syst., 2019MS001791, doi:10.1029/2019MS001791, 2019.

Seitzinger, S. P. and Giblin, A. E.: Estimating denitrification in North Atlantic continental shelf sediments, Biogeochemistry, 35(1), 235–260, doi:10.1007/BF02179829, 1996.

SeaWiFS. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Chlorophyll Data; NASA OB.DAAC, Greenbelt, MD, USA. doi:10.5067/ORBVIEW-2/SEAWIFS/L3M/CHL/2018. Accessed on 2014/03/12.

Sellar, A. A., Jones, C. G., Mulcahy, J., Tang, Y., Yool, A., Wiltshire, A., O'Connor, F. M., Stringer, M., Hill, R., Palmieri, J., Woodward, S., Mora, L., Kuhlbrodt, T., Rumbold, S., Kelley, D. I., Ellis, R., Johnson, C. E., Walton, J., Abraham, N. L., Andrews, M. B., Andrews, T., Archibald, A. T., Berthou, S., Burke, E., Blockley, E., Carslaw, K., Dalvi, M., Edwards, J., Folberth, G. A., Gedney, N., Griffiths, P. T., Harper, A. B., Hendry, M. A., Hewitt, A. J., Johnson, B., Jones, A., Jones, C. D., Keeble, J., Liddicoat, S., Morgenstern, O., Parker, R. J., Predoi, V., Robertson, E., Siahaan, A., Smith, R. S., Swaminathan, R., Woodhouse, M. T., Zeng, G. and Zerroukat, M.: UKESM1: Description and evaluation of the UK Earth System Model, J. Adv. Model. Earth Syst., 2019MS001739, doi:10.1029/2019MS001739, 2019.

Siedlecki, S. A., Banas, N. S., Davis, K. A., Giddings, S., Hickey, B. M., MacCready, P., Connolly, T. and Geier, S.: Seasonal and interannual oxygen variability on the Washington and Oregon continental shelves, J. Geophys. Res. Ocean., 120(2), 608–633, doi:10.1002/2014JC010254, 2015.

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., Laufkötter, C., Paulot, F. and Zadeh, N.: Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO 2, J. Adv. Model. Earth Syst., 12(10), doi:10.1029/2019MS002043, 2020.

Stortini, C. H., Shackell, N. L., Tyedmers, P. and Beazley, K.: Assessing marine species vulnerability to projected warming on the Scotian Shelf, Canada, ICES J. Mar. Sci., 72(6), 1731–1743, doi:10.1093/icesjms/fsv022, 2015.

Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D. and Winter, B.: The Canadian Earth System Model version 5 (CanESM5.0.3), Geosci. Model Dev. Discuss., 1–68, doi:10.5194/gmd-2019-177, 2019.

Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An Overview of CMIP5 and the Experiment Design, Bull. Am. Meteorol. Soc., 93(4), 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., Seland, Ø. and Heinze, C.: Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM), Geosci. Model Dev., 6(2), 301–325, doi:10.5194/gmd-6-301-2013, 2013.

Urrego-Blanco, J. and Sheng, J.: Interannual Variability of the Circulation over the Eastern Canadian Shelf, Atmosphere-Ocean, 50(3), 277–300, doi:10.1080/07055900.2012.680430, 2012.

Vichi, M., Pinardi, N. and Masina, S.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory, J. Mar. Syst., 64(1–4), 89–109, doi:10.1016/j.jmarsys.2006.03.006, 2007a.

Vichi, M., Masina, S. and Navarra, a.: A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part II: Numerical simulations, J. Mar. Syst., 64(1–4), 110–134, doi:10.1016/j.jmarsys.2006.03.014, 2007b.

Vichi, M., Manzini, E., Fogli, P. G., Alessandri, A., Patara, L., Scoccimarro, E., Masina, S. and Navarra, A.: Global and regional ocean carbon uptake and climate change: sensitivity to a substantial mitigation scenario, Clim. Dyn., 37(9–10), 1929–1947, doi:10.1007/s00382-011-1079-0, 2011.

Voldoire, A., Sanchez-Gomez, E., Salas y Mélia, D., Decharme, B., Cassou, C., Sénési, S., Valcke, S., Beau, I., Alias, A., Chevallier, M., Déqué, M., Deshayes, J., Douville, H., Fernandez, E., Madec, G., Maisonnave, E., Moine, M.-P., Planton, S., Saint-Martin, D., Szopa, S., Tyteca, S., Alkama, R., Belamari, S., Braun, A., Coquart, L. and Chauvin, F.: The CNRM-CM5.1 global climate model: description and basic evaluation, Clim. Dyn., 40(9–10), 2091–2121, doi:10.1007/s00382-011-1259-y, 2013.

Walsh, J. J.: Importance of continental margins in the marine biogeochemical cycling of carbon and nitrogen, Nature, 350, 53–55, 1991.

Wilson, K. and Lotze, H.: Climate change projections reveal range shifts of eelgrass Zostera marina in the Northwest Atlantic, Mar. Ecol. Prog. Ser., 620, 47–62, doi:10.3354/meps12973, 2019.

Wilson, K. L., Skinner, M. A. and Lotze, H. K.: Projected 21st-century distribution of canopy-forming seaweeds in the Northwest Atlantic with climate change, edited by C. Sorte, Divers. Distrib., 25(4), 582–602, doi:10.1111/ddi.12897, 2019.

Yool, A., Popova, E. E. and Anderson, T. R.: MEDUSA-2.0: an intermediate complexity biogeochemical model of the marine carbon cycle for climate change and ocean acidification studies, Geosci. Model Dev., 6(5), 1767–1811, doi:10.5194/gmd-6-1767-2013, 2013.

Zhang, H., Fennel, K., Laurent, A. and Bian, C.: A numerical model study of the main factors contributing to hypoxia and its sub-seasonal to interannual variability off the Changjiang Estuary, Biogeosciences Discuss., Rev., doi:10.5194/bg-2019-341, 2019.

Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V., Boyer, T. P., Garcia, H. E., Baranova, O. K., Johnson, D. R., Seidov, D. and Biddle, M. M.: World Ocean Atlas 2013, Volume 2: Salinity, S. Levitus, Ed., A. Mishonov Technical Ed.; NOAA Atlas NESDIS 74., 2013.

Formatted: English (US)

Table 1. Information about the regional model and the 29 ESM models. For the CMIP5 models (2–18) the r1i1p1 ensemble was used. For the CMIP6 model (19–30) the r1i1p1f1 ensemble was used on the native grid when available, except for CNRM-ESM2-1, MIROC-ES2L and UKESM1-0-LL (r1i1p1f2), GFDL-ESM4 and NorESM2-LM (regridded), and GISS-E2-1-G (r101i1p1f1). The filled circles and open squares indicate the models that are part of the inner and outer ensembles, respectively. N indicates the number of vertical levels. Note that the IPSL-CM5 models share the same ocean component with higher resolution atmospheric component in the MR version. Similarly, MPI-ESM-MR and MPI-ESM1-2-HR share the same ocean component with higher resolution atmospheric component in the HR version.

Model			Shelf resolution (*n* cells)			Shelf resolution		Ocean BGC component	References
Name	ID	Σ	GoM	SS	GB	Δlon × Δlat (degree)	N		
ACM	1		1780	1366	3729	0.06×0.09	30	BIO_FENNEL	Brennan et al. (2016); Fennel et al. (2006)
CanESM2	2	□	11	14	29	1.4×0.9	40	CMOC	Arora et al. (2011); Christian et al. (2010)
CESM1-BGC	3	□	41	33	91	1.1×0.4	60	BEC	Lindsay et al. (2014); Moore et al. (2013)
CMCC-CESM	4	□	8	5	13	2×1.25	30	PELAGOS	Vichi et al. (2007a, 2007b, 2011)
CNRM-CM5	5	□	27	20	55	1×0.62	42	PISCES	Aumont and Bopp (2006); Voldoire et al. (2013)
GFDL-ESM2-G	6	●	20	15	39	1×1	50	TOPAZ2	Dunne (2013); Dunne et al. (2012, 2013)
GFDL-ESM2-M	7	□							
GISS-E2-H-CC	8	□	19	14	39	1×1	26	NOBM	Romanou et al. (2013) Schmidt et al. (2014)
GISS-E2-R-CC	9	□	15	12	29	1.25×1	32		
HadGEM2-CC	10	●	18	15	39	1×1	40	Diat-HadOCC	Collins et al., (2011); Palmer and Totterdell (2001)
HadGEM2-ES	11	□							
IPSL-CM5A-LR	12	●	8	5	13	2×1.25	31	PISCES	Aumont and Bopp (2006); Dufresne et al. (2013)
IPSL-CM5A-MR	13	●							
IPSL-CM5B-LR	14	□							
MPI-ESM-LR	15	□	23	23	73	0.8×0.5	47	HAMOCC 5.2	Giorgetta et al. (2013); Ilyina et al. 2013)
MPI-ESM-MR	16	●	136	87	193	0.4×0.3	95		
MRI-ESM1	17	□	40	29	80	1×0.5	50	MRI.COM3	Adachi et al. (2013)
NorESM1-ME	18	□	41	33	91	1×0.43	53	HAMOCC 5.1	Tjiputra et al. (2013)
CanESM5	19	□	27	20	55	1×0.62	45	CMOC	Swart et al. (2019)
CESM2	20	□	41	33	91	1×0.43	60	MARBL	Danabasoglu et al. (2020)
CESM2-WACCM	21	□					60		
CNRM-ESM2-1	22	●	27	20	55	1×0.62	75	PISCES	Aumont et al. (2015); Séférian et al. (2019)
GFDL-ESM4	23	●	20	15	39	1×1	75	COBALTv2	Stock et al. (2020)
GISS-E2-1-G	24	□	15	12	29	1.25×1	40	NOBM	Rousseaux and Gregg (2015)
GISS-E2-1-G-CC	25	□					40		
IPSL-CM6A-LR	26	●	27	20	55	1×0.62	75	PISCES	Aumont et al. (2015); Boucher et al. (2020)
MIROC-ES2L	27	●	20	18	43	1×0.77	62	OECO2	Hajima et al. (2019)
MPI-ESM1-2-HR	28	●	136	87	193	0.4×0.3	95	HAMOCC	Müller et al. (2018)
NorESM2-LM	29	□	25	20	57	1×0.6	70	HAMOCC	Müller et al. (2018)
UKESM1-0-LL	30	●	27	20	55	1×0.62	75	MEDUSA2	Sellar et al. (2019); Yool et al. (2013)

Deleted: CM4, GFDL-ESM4 and NorESM2-LM (regridded)... [1]
Formatted Table ... [2]
Merged Cells ... [3]
Merged Cells ... [4]
Split Cells ... [6]
Deleted: GM
Inserted Cells ... [5]
Inserted Cells ... [9]
Formatted Table ... [7]
Inserted Cells ... [8]
Inserted Cells ... [10]
Formatted ... [11]
Deleted: Brennan et al. (2016); Fennel et al. (2006)
Deleted: Arora et al. (2011); Christian et al. (2010)
Deleted: Lindsay et al. (2014); Moore et al. (2013)
Deleted: Vichi et al. (2007a, 2007b, 2011)
Deleted: Aumont and Bopp (2006); Voldoire et al. (2013)
Deleted: Dunne (2013); Dunne et al. (2012, 2013)
Merged Cells ... [12]
Inserted Cells ... [15]
Formatted Table ... [13]
Inserted Cells ... [14]
Deleted: Romanou et al. (2013) Schmidt et al. (2014)
Split Cells ... [16]
Merged Cells ... [17]
Formatted ... [18]
Deleted: Collins et al., (2011); Palmer and Totterdell (2001)
Inserted Cells ... [19]
Inserted Cells ... [20]
Merged Cells ... [21]
Inserted Cells ... [22]
Inserted Cells ... [23]
Deleted: Aumont and Bopp (2006); Dufresne et al. (2013)
Merged Cells ... [24]
Inserted Cells ... [27]
Formatted Table ... [25]
Inserted Cells ... [26]
Deleted: Giorgetta et al. (2013); Ilyina et al. 2013)
Split Cells ... [28]
Merged Cells ... [29]
Formatted ... [30]
Inserted Cells ... [32]
Formatted Table ... [31]
Inserted Cells ... [33]
Deleted: Adachi et al. (2013)
Deleted: Tjiputra et al. (2013)
Deleted: Swart et al. (2019)
Deleted: Danabasoglu et al. (2020)
Merged Cells ... [34]
Inserted Cells ... [37]
Formatted Table ... [35]
Deleted: Aumont et al. (2015); Séférian et al. (2019)
Inserted Cells ... [36]
Formatted ... [38]

Table 2. Annual model scores and ranking. **Ɍ represents the multi variable mean ranking and Ɍbio the chlorophyll and nitrate mean ranking. The final rank is provided in the right column. The asterisks beside the overall rank indicate a possible overestimation of the rank due to low nitrate concentrations (Figures S1–S5j-l).**

Ranked models			Scores				Ranks					
Name	ID	CMIP	Temp.	Salt.	Chl-a	NO3	Temp.	Salt.	Chl-a	NO3	Ɍ	Ɍbio
ACM	1	–	1.14	0.48	0.64	1.27	1	1	1	1	1.0	1.0
MPI-ESM1-2-HR	28	6	2.05	0.73	1.03	1.75	4	4	7	6	5.3	6.5
GFDL-ESM2G	6	5	2.12	1.33	1.17	1.67	5	8	20	5	9.5	12.5
GFDL-ESM4	23	6	2.49	2.10	0.81	2.10	9	16	2	12	9.8	7.0
CNRM-ESM2-1	22	6	2.74	1.39	0.90	2.21	12	10	3	17	10.5	10.0
HadGEM2-CC	10	5	2.58	2.02	1.02	2.11	11	13	6	13	10.8	9.5
IPSL-CM6A-LR	26	6	2.47	2.03	1.09	1.94	8	14	12	9	10.8	10.5
MIROC-ES2L	27	6	3.14	0.92	1.02	2.17	18	5	5	16	11.0	10.5
UKESM1-0-LL	30	6	3.08	0.67	1.15	1.96	17	3	17	10	11.8	13.5
CNRM-CM5	5	5	1.78	0.53	1.11	6.54	3	2	16	27	12.0	21.5
MPI-ESM-MR	16	5	2.14	1.22	1.09	2.57	6	7	14	21	12.0	17.5
IPSL-CM5A-LR	12	5	2.52	2.00	1.17	1.91	10	12	19	8	12.3	13.5
IPSL-CM5A-MR	13	5	3.07	2.37	1.09	1.80	16	18	13	7	13.5	10.0
MPI-ESM-LR	15	5	2.38	1.37	1.10	3.12	7	9	15	24	13.8	19.5
HadGEM2-ES	11	5	2.90	2.50	1.06	2.12	14	19	9	14	14.0	11.5
IPSL-CM5B-LR	14	5	1.51	2.64	1.36	2.03	2	23	26	11	15.5	18.5
NorESM2-LM	29	6	2.95	1.81	1.05	13.23	15	11	8	30	16.0	19.0
GISS-E2-1-G-CC	25	6	4.66	3.77	1.08	1.44	25	28	11	2	16.5	6.5
CanESM5	19	6	4.05	1.18	1.35	2.16	23	6	24	15	17.0	19.5

Formatted ... [43]
Formatted Table ... [44]
Inserted Cells ... [46]
Deleted:
Inserted Cells ... [45]
Formatted ... [49]
Formatted ... [54]
Formatted ... [56]
Formatted ... [47]
Inserted Cells ... [53]
Formatted ... [48]
Formatted ... [50]
Formatted ... [51]
Formatted ... [52]
Formatted ... [55]
Formatted ... [59]
Formatted ... [64]
Formatted ... [66]
Formatted ... [60]
Inserted Cells ... [61]
Formatted ... [57]
Formatted ... [58]
Formatted ... [62]
Formatted ... [63]
Formatted ... [65]
Deleted: 7
Moved (insertion) [1]
Deleted: 22
Deleted: 9
Deleted Cells ... [67]
Deleted Cells ... [72]
Deleted: 12
Inserted Cells ... [77]
Formatted ... [79]
Moved down [2]: 0.92
Inserted Cells ... [68]
Deleted: 74
Formatted ... [70]
Formatted ... [71]
Deleted: 91
Deleted: 3
Formatted ... [73]
Deleted: .0
Formatted ... [74]
Deleted: 6.0
Inserted Cells ... [76]
Inserted Cells ... [78]
Formatted ... [69]
Formatted ... [75]
Deleted: CNRM-ESM2-1
Formatted ... [83]
Inserted Cells ... [81]
Formatted ... [80]
Formatted ... [82]

GISS-E2-1-G	24	6	5.00	3.89	1.08	1.47	26	29	10	3	17.0		6.5
MRI-ESM1	17	5	2.78	2.63	1.15	2.53	13	20	18	20	17.8		19.5
GISS-E2-R-CC	9	5	3.84	3.00	1.19	1.62	21	25	22	4	18.0		13.0
GFDL-ESM2M	7	5	3.89	2.63	0.95	7.14	22	21	4	29	19.0	16.5	
GISS-E2-H-CC	8	5	3.64	2.07	1.35	2.29	19	15	25	18	19.3	21.5	
CanESM2	2	5	4.20	2.63	1.18	3.14	24	22	21	25	23.0		23.0
CMCC-CESM	4	5	5.18	2.15	1.40	2.39	27	17	29	19	23.0		24.0
NorESM1-ME	18	5	3.71	2.86	1.40	6.99	20	24	28	28	25.0		28.0
CESM2	20	6	5.40	3.42	1.38	2.61	28	26	27	22	25.8		24.5
CESM1-BGC	3	5	7.84	4.16	1.29	4.21	30	30	23	26	27.3		24.5
CESM2-WACCM	21	6	5.71	3.51	1.42	2.78	29	27	30	23	27.3		26.5

Deleted: 17
Deleted: 13
Formatted ... [224]
Formatted ... [225]
Deleted: 5
Deleted: 2.78
Deleted: 1.15
Deleted: 2.53
Deleted: 18
Deleted: 20
Formatted ... [222]
Inserted Cells ... [223]
Deleted: 19.0
Moved down [14]: MRI-ESM1
Moved (insertion) [14]
Deleted: 29
Inserted Cells ... [228]
Formatted ... [231]
Formatted ... [233]
Deleted: 6
Deleted: 98
Formatted ... [226]
Deleted: 1.04
Formatted ... [229]
Deleted: .26
Deleted: 15
Deleted: 8
Deleted: 30
Deleted: 0
Inserted Cells ... [227]
Formatted ... [230]
Formatted ... [232]
Deleted: NorESM2-LM
Deleted: 7
Deleted: GFDL-ESM2M
Deleted: 29
Deleted Cells ... [239]
Formatted ... [240]
Formatted ... [241]
Deleted: 7
Formatted ... [234]
Deleted: 89
Deleted: 0.95
Deleted: 7.14
Inserted Cells ... [235]
Inserted Cells ... [236]
Formatted ... [237]
Formatted ... [238]
Deleted: 16.5
Deleted: 3
Deleted: GISS-E2-H-CC
Formatted ... [246]
Deleted: 25
Deleted: 18

Table 3. Comparison statistics between ACM and AZMP and glider observations along the Halifax Line and at Station 2.

	RMSD				Bias				Correlation coefficient			
Season*	W	S	S	F	W	S	S	F	W	S	S	F
					Halifax Line							
Chlorophyll (time series)	0.25	0.37	0.39	0.36	0.08	0.22	0.28	0.13	0.68	0.78	0.71	0.75
Chlorophyll (Glider)	0.22	0.42	0.25	0.22	-0.14	0.13	0.17	0.04	0.88	0.78	0.94	0.85
Nitrate	2.99	2.73	2.13	1.77	0.76	2.03	0.74	1.27	0.90	0.83	0.85	0.92
					Station 2							
Chlorophyll (time series)	0.26	1.74	0.52	0.30	0.05	-0.56	0.26	0.01	0.64	0.22	0.48	0.82
Chlorophyll (Glider)	0.15	1.06	0.31	0.17	-0.03	-0.46	0.25	0.02	0.87	0.69	0.91	0.93
Nitrate	0.96	1.57	1.58	1.37	1.19	1.62	0.26	0.58	0.85	0.86	0.91	0.94

*Seasons are order sequentially and abbreviated as W (winter, Dec–Feb), S (spring, Mar–May), S (summer, Jun–Aug) and F (fall, Sep–Nov).

Formatted: Font: Not Italic

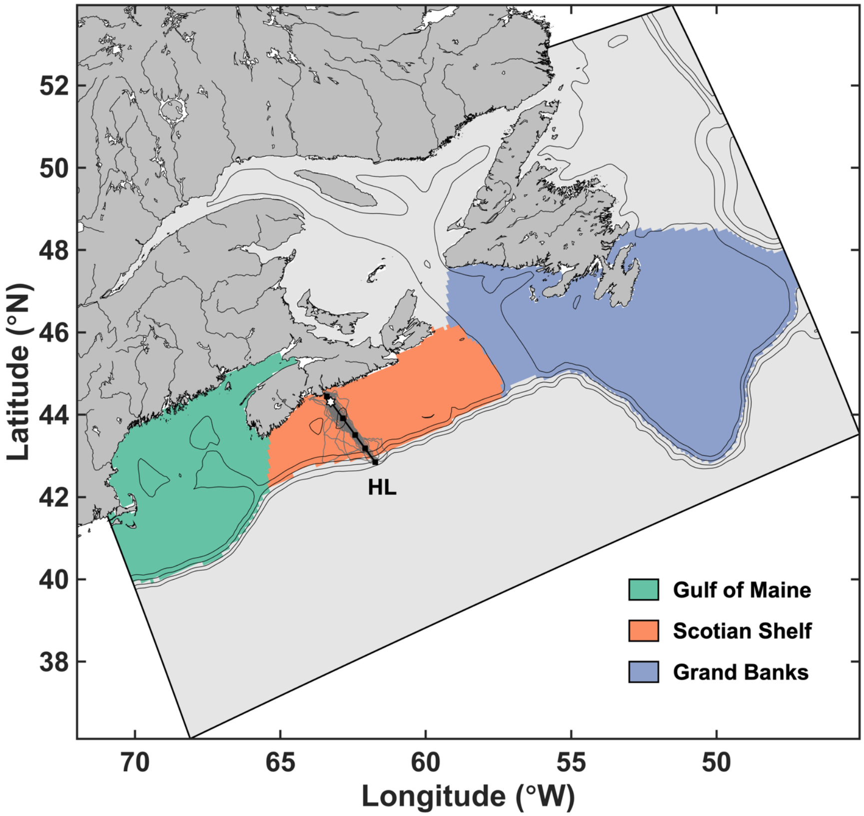
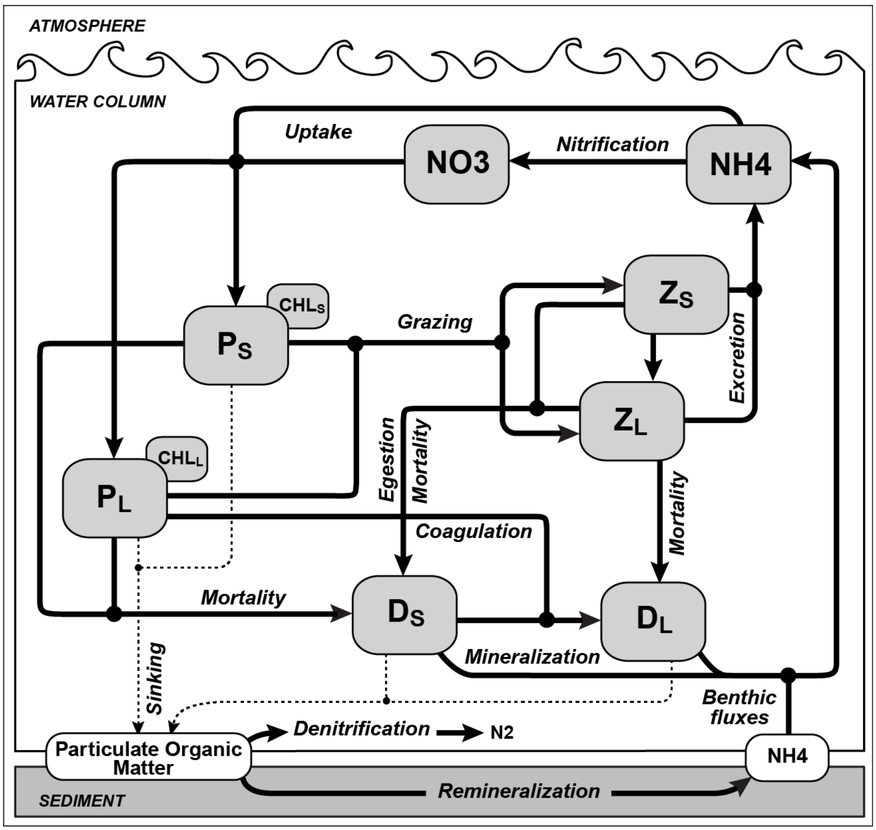


Figure 1. Study area indicating the 3 averaging zones, the limits of the ROMS grid and the location of the Halifax Line stations (squares) used in the analysis. The white star is Station 2 and the grey lines the gliders track.

1455



1460 **Figure 2. Schematic of the biogeochemical model used in ROMS. The state variables are small phytoplankton (P_S) and chlorophyll (CHL_S), large phytoplankton (P_L) and chlorophyll (CHL_L), small zooplankton (Z_S), large zooplankton (Z_L), slow-sinking small detritus (D_S), fast-sinking large detritus (D_L), nitrate (NO_3), and ammonium (NH_4). Dashed lines indicate sinking. Black dots represent the connections between paths.**

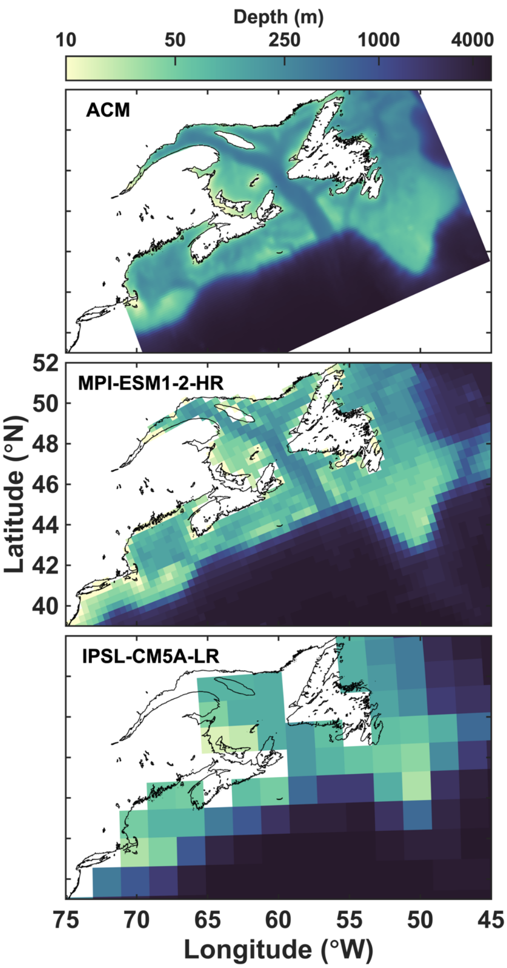


Figure 3. Bathymetry of the regional model (top), the highest resolution ESM (middle) and lowest resolution ESM (bottom).

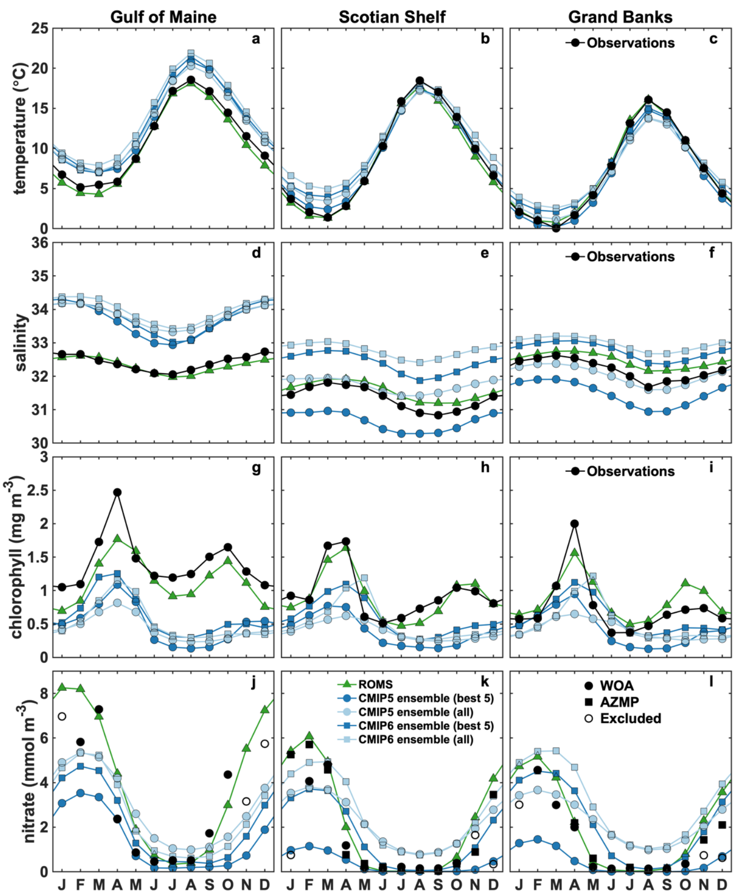
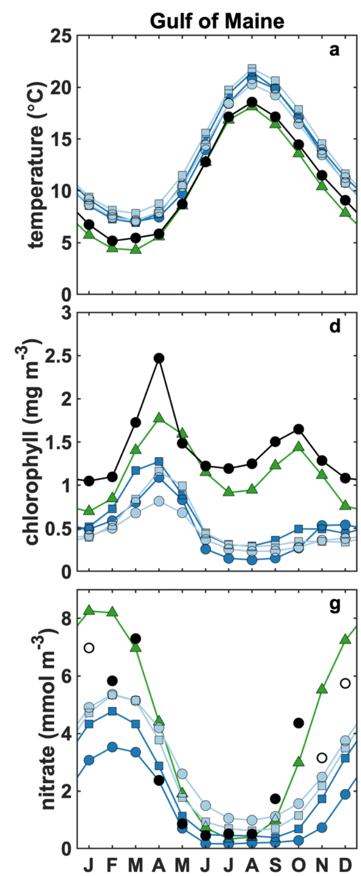


Figure 4. **Observed, ROMS and ensemble means area averaged surface temperature (a-c), salinity (d-f), chlorophyll (g-i) and nitrate (j-l) in the 3 NWA shelf regions.** Nov–Jan WOA nitrate data are excluded (open circles). Model comparison with observations in the Gulf of Maine is therefore only available from February to October. For the Scotian Shelf and Grand Banks additional AZMP data are available. In case of multiple observations, the data are monthly averaged.

Deleted:



Deleted: a-c),

Deleted: d-f) and temperature (g-i)

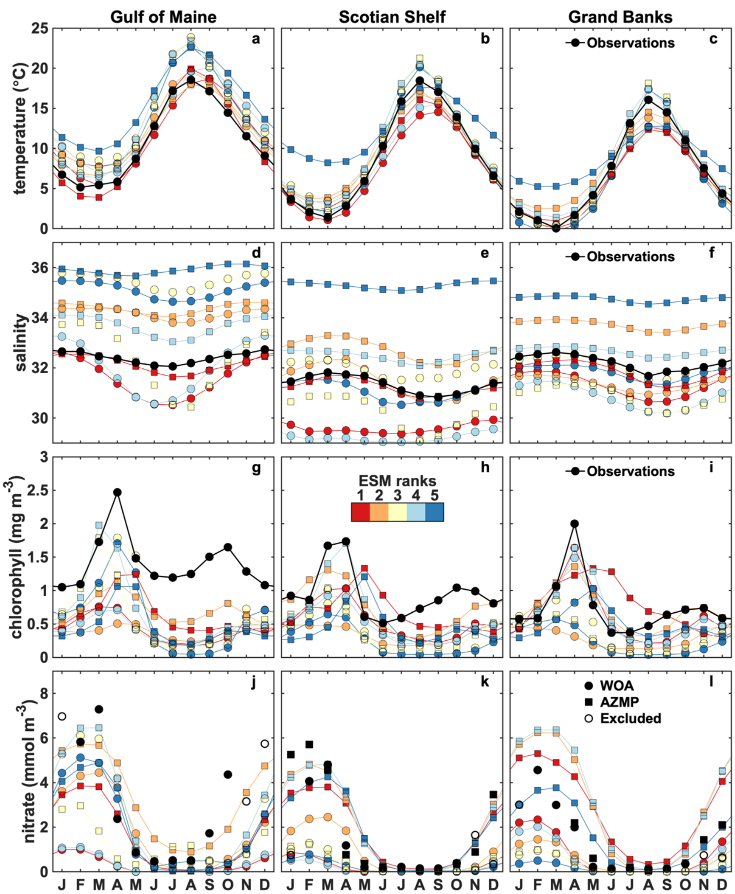
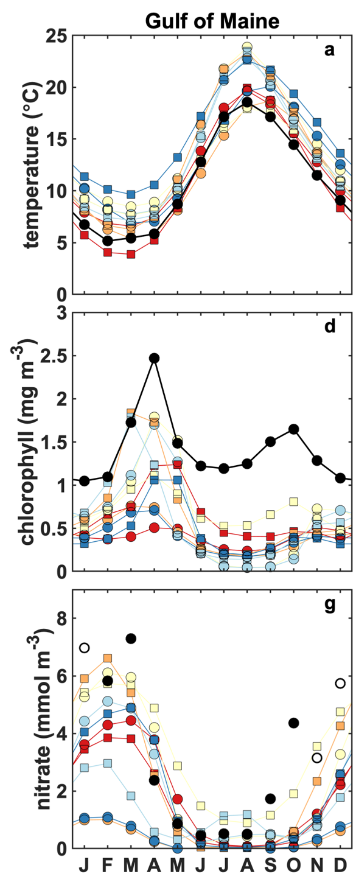


Figure 5. Observed (black dots) and best ESMs area averaged surface temperature (a-c), salinity (d-f), chlorophyll (g-i) and nitrate (j-l) in the 3 NWA shelf regions. The colored circles and squares indicate the CMIP5 and CMIP6 models, respectively. Nov–Jan WOA nitrate data are excluded (open circles). Model comparison with observations in the Gulf of Maine is therefore only available from February to October. For the Scotian Shelf and Grand Banks additional AZMP data are available. In case of multiple observations, the data are monthly averaged.



Deleted:

Deleted: a-c), nitrate (d-f) and temperature (

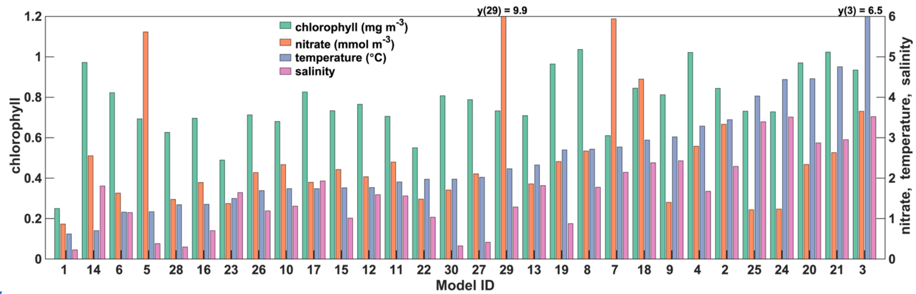
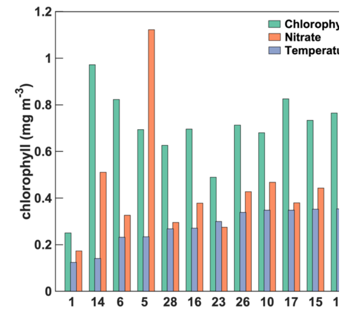


Figure 6. Root mean square difference between monthly, regionally averaged observations and models. Model numbers refer to the IDs in Table 1.

Deleted:



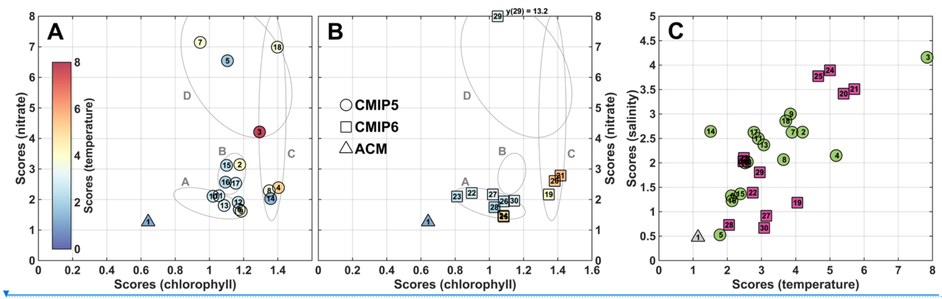
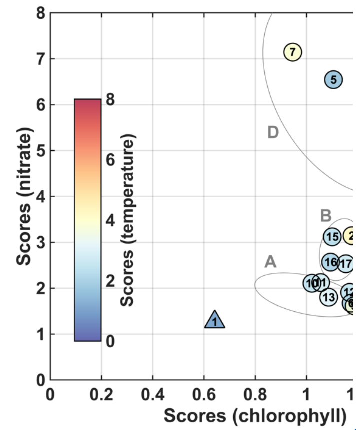


Figure 7. Model scores for surface chlorophyll (x-axis), nitrate (y-axis) and temperature (color scale) for the CMIP5 group (A, left panel), the CMIP6 group (B, middle panel) and the regional model (triangles). Salinity and temperature scores are compared in panel C (right). The grey ellipsoids indicate the groups A–D (see text) and are the same in panels A and B.

Deleted:



Deleted: right

Deleted: .

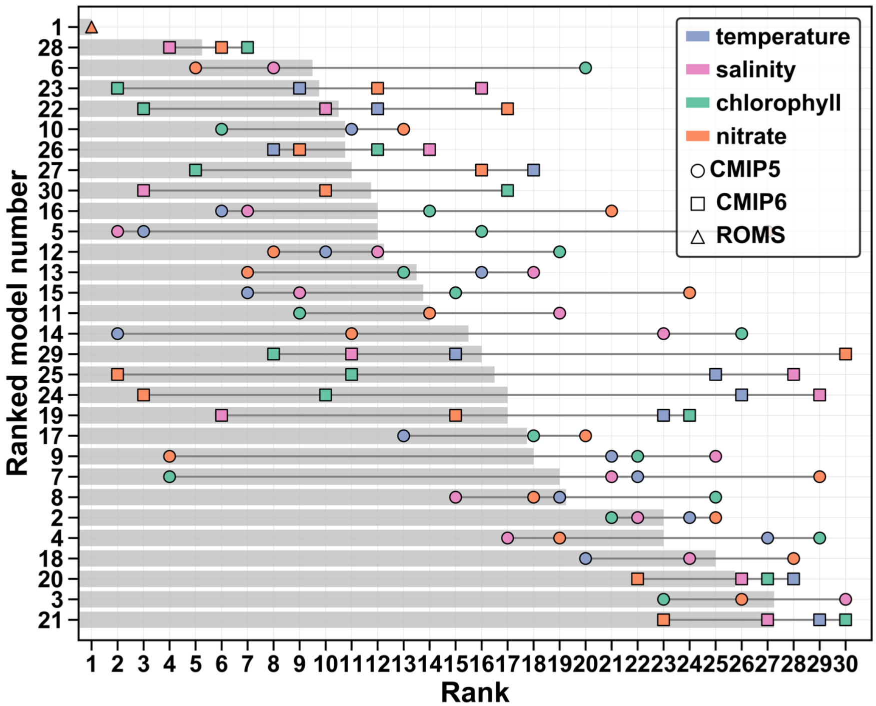
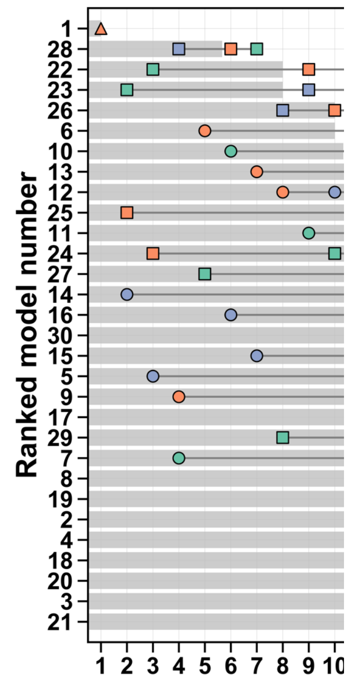


Figure 8. Model average (grey bars) and specific (dots) ranking. The final ranking is shown on the y-axis. Hidden coinciding ranks (models 2, 3, 6, 10, 11, 18, 27, 28 and 30) are provided in Table 2.



Deleted:

Deleted: The temperature rank for model 6 is hidden behind the nitrate rank (x = 5).

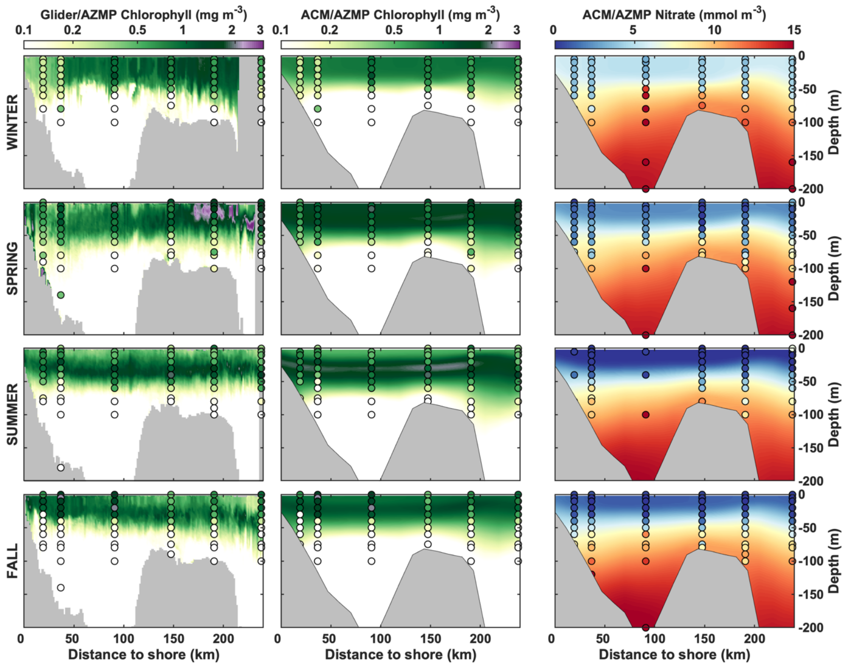


Figure 9. Comparison of gliders, AZMP and model seasonal climatologies of chlorophyll and nitrate along the Halifax line.

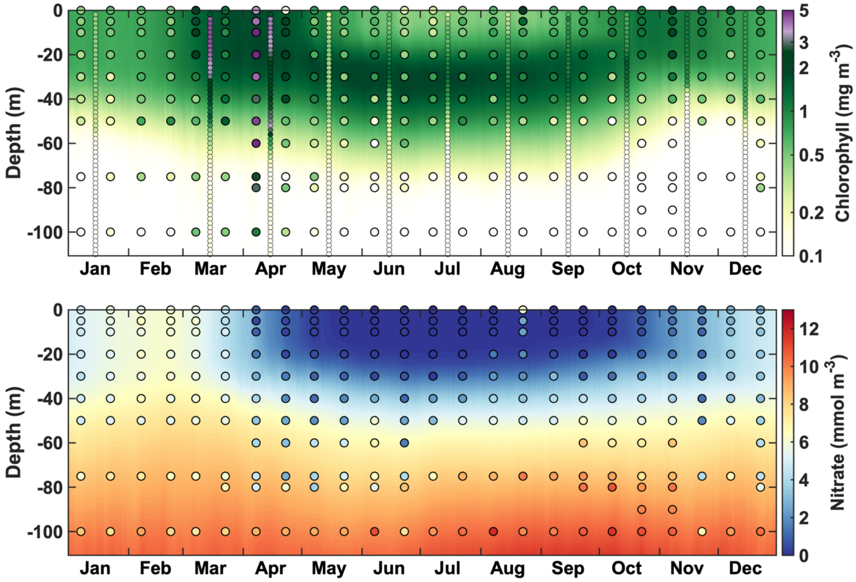


Figure 10. Comparison of vertically-resolved time series of chlorophyll (top) and nitrate (bottom) at Station 2 from the regional model (background), the glider transects (small dots) and the bimonthly sampling (large dots).

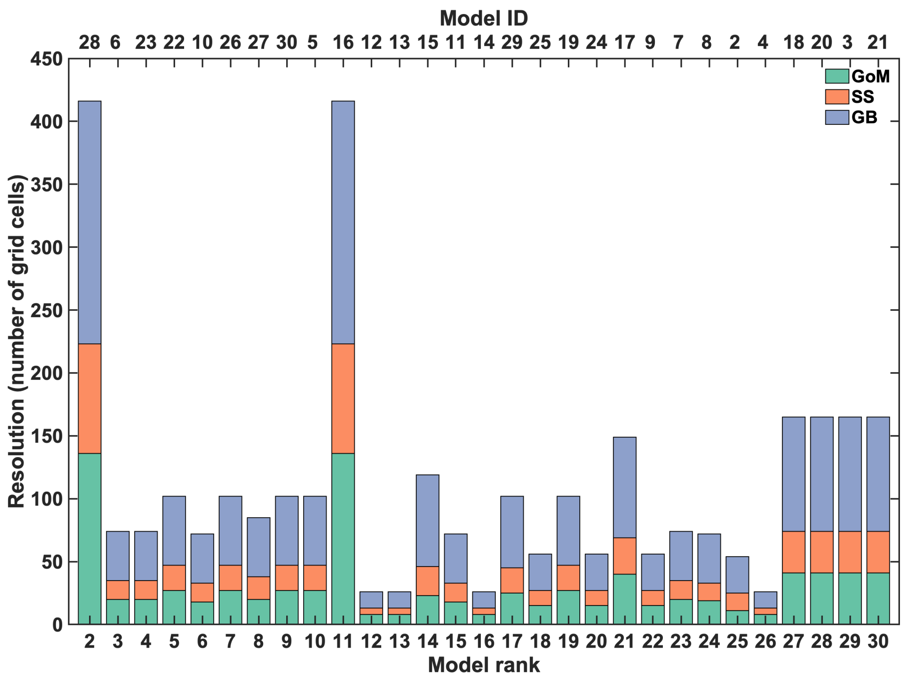
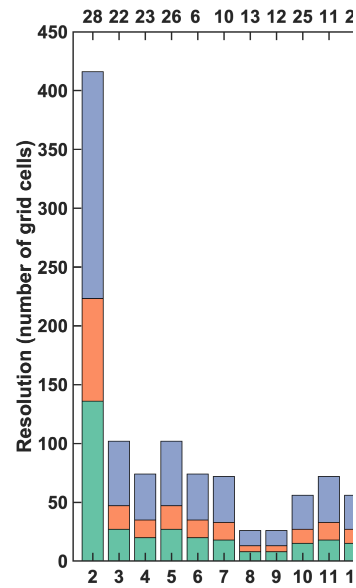


Figure 11. Resolution of the 29 ESMs ordered by their overall rank (see Figure 8).



Deleted:

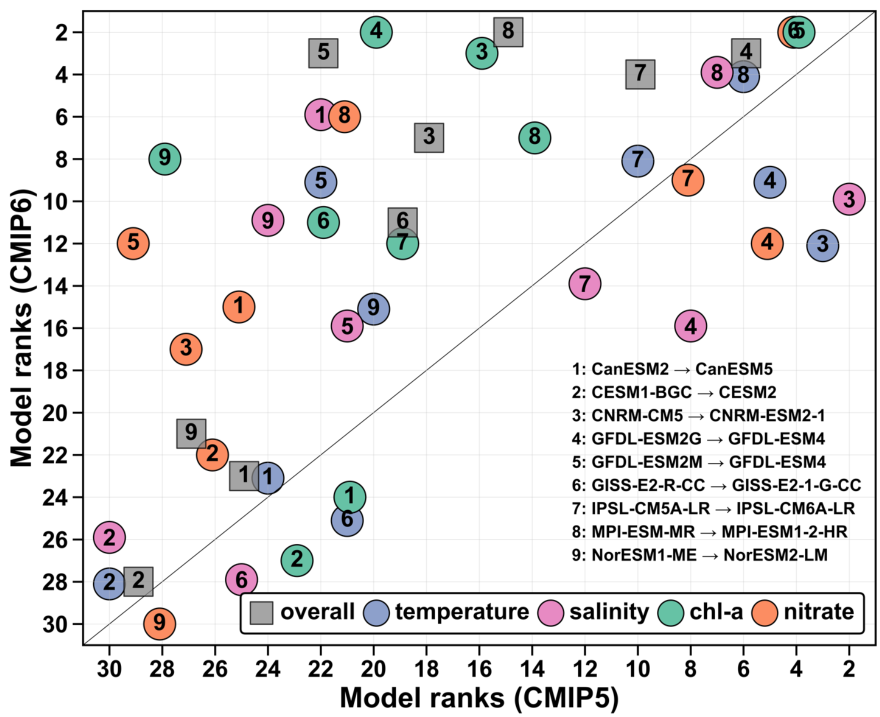
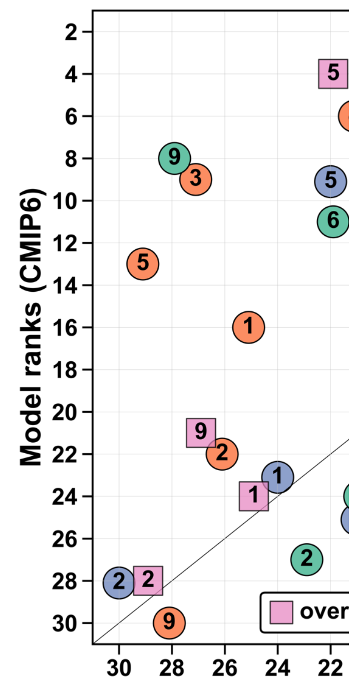


Figure 12. Comparison of the ranks of the former (x-axis) and current (y-axis) generations of ESMs. Grey squares represent the overall ranks, whereas the dots indicate temperature (blue), salinity (magenta), chlorophyll (green) and nitrate (orange) ranks. The numbers indicate the model (see legend). These numbers do not correspond to the original model IDs indicated in Table 1. The black line is the 1:1 line. Dots above this line indicate an improvement and dots below the line a worsening of the rank. Note that there were two CMIP5 GFDL models but only one in the CMIP6 group (model pairs 4 and 5).



Deleted:

Deleted: Blue dots indicate

Deleted: green dots the

Deleted: ranks

Deleted: the orange dots the

Page 25: [1] Deleted	Author	

Page 25: [2] Formatted Table	Author	

Formatted Table

Page 25: [3] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [4] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [5] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [6] Split Cells	Author	15/30/07 8:48:00 PM

Split Cells

Page 25: [7] Formatted Table	Author	

Formatted Table

Page 25: [8] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [9] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [10] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [11] Formatted	Author	

Centered

Page 25: [12] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [13] Formatted Table	Author	

Formatted Table

Page 25: [14] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [15] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [16] Split Cells	Author	15/30/07 8:48:00 PM

Split Cells

Page 25: [18] Formatted	Author	

Centered

Page 25: [19] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [20] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [21] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [22] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [23] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [24] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [25] Formatted Table	Author	

Formatted Table

Page 25: [26] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [27] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [28] Split Cells	Author	15/30/07 8:48:00 PM

Split Cells

Page 25: [29] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [30] Formatted	Author	

Centered

Page 25: [31] Formatted Table	Author	

Formatted Table

Page 25: [32] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [33] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Merged Cells

Page 25: [35] Formatted Table	Author	

Formatted Table

Page 25: [36] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [37] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [38] Formatted	Author	

Centered

Page 25: [39] Merged Cells	Author	15/30/07 8:48:00 PM

Merged Cells

Page 25: [40] Formatted Table	Author	

Formatted Table

Page 25: [41] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 25: [42] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [43] Formatted	Author	

Line spacing: single

Page 26: [44] Formatted Table	Author	

Formatted Table

Page 26: [45] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [46] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [47] Formatted	Author	

Font color: Auto

Page 26: [48] Formatted	Author	

Font color: Auto

Page 26: [49] Formatted	Author	

Font color: Auto

Page 26: [51] Formatted	Author	

Font color: Auto

Page 26: [51] Formatted	Author	

Font color: Auto

Page 26: [52] Formatted	Author	

Font color: Auto

Page 26: [53] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [54] Formatted	Author	

Font color: Auto

Page 26: [55] Formatted	Author	

Font color: Auto

Page 26: [56] Formatted	Author	

Left

Page 26: [57] Formatted	Author	

Font color: Auto

Page 26: [58] Formatted	Author	

Font color: Auto

Page 26: [59] Formatted	Author	

Font color: Auto

Page 26: [60] Formatted	Author	

Font color: Auto

Page 26: [61] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [62] Formatted	Author	

Font color: Auto

Page 26: [63] Formatted	Author	

Font color: Auto

Page 26: [64] Formatted	Author	

Font color: Auto

Page 26: [65] Formatted	Author	

Font color: Auto

Page 26: [67] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [68] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [69] Formatted	Author	

Font color: Auto

Page 26: [70] Formatted	Author	

Font color: Auto

Page 26: [71] Formatted	Author	

Font color: Auto

Page 26: [72] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [73] Formatted	Author	

Font color: Auto

Page 26: [74] Formatted	Author	

Font color: Black

Page 26: [75] Formatted	Author	

Font color: Auto

Page 26: [76] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [77] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [78] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [79] Formatted	Author	

Left

Page 26: [80] Formatted	Author	

Font color: Auto

Page 26: [81] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [82] Formatted	Author	

Font color: Auto

Font color: Auto

Page 26: [84] Formatted	Author	

Font color: Auto

Page 26: [85] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [86] Formatted	Author	

Font color: Auto

Page 26: [87] Formatted	Author	

Font color: Auto

Page 26: [88] Formatted	Author	

Font color: Auto

Page 26: [89] Formatted	Author	

Left

Page 26: [90] Formatted	Author	

Font color: Auto

Page 26: [91] Formatted	Author	

Font color: Auto

Page 26: [92] Formatted	Author	

Font color: Auto

Page 26: [93] Formatted	Author	

Font color: Auto

Page 26: [93] Formatted	Author	

Font color: Auto

Page 26: [94] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [95] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [96] Formatted	Author	

Font color: Auto

Page 26: [97] Formatted	Author	

Font color: Auto

Page 26: [99] Formatted	Author	

Font color: Auto

Page 26: [100] Formatted	Author	

Font color: Auto

Page 26: [101] Formatted	Author	

Font color: Black

Page 26: [102] Formatted	Author	

Font color: Auto

Page 26: [103] Formatted	Author	

Font color: Auto

Page 26: [104] Formatted	Author	

Left

Page 26: [105] Formatted	Author	

Font color: Auto

Page 26: [106] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [107] Formatted	Author	

Font color: Auto

Page 26: [108] Formatted	Author	

Font color: Auto

Page 26: [109] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [110] Formatted	Author	

Font color: Auto

Page 26: [110] Formatted	Author	

Font color: Auto

Page 26: [111] Formatted	Author	

Font color: Auto

Page 26: [112] Formatted	Author	

Font color: Auto

Page 26: [113] Formatted	Author	

Left

Page 26: [115] Formatted	Author	

Font color: Auto

Page 26: [116] Formatted	Author	

Font color: Auto

Page 26: [117] Formatted	Author	

Font color: Black

Page 26: [118] Formatted	Author	

Font color: Auto

Page 26: [119] Formatted	Author	

Font color: Auto

Page 26: [120] Formatted	Author	

Left

Page 26: [121] Formatted	Author	

Font color: Auto

Page 26: [122] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [123] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [124] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [125] Formatted	Author	

Font color: Auto

Page 26: [126] Formatted	Author	

Font color: Black

Page 26: [127] Formatted	Author	

Font color: Auto

Page 26: [128] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [129] Formatted	Author	

Font color: Auto

Page 26: [130] Formatted	Author	

Font color: Black

Deleted Cells

Page 26: [132] Formatted	**Author**	

Left

Page 26: [133] Formatted	**Author**	

Font color: Auto

Page 26: [134] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 26: [135] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 26: [136] Formatted	**Author**	

Font color: Auto

Page 26: [137] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [138] Formatted	**Author**	

Font color: Auto

Page 26: [138] Formatted	**Author**	

Font color: Auto

Page 26: [139] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [140] Formatted	**Author**	

Font color: Auto

Page 26: [141] Formatted	**Author**	

Font color: Auto

Page 26: [142] Formatted	**Author**	

Left

Page 26: [143] Formatted	**Author**	

Font color: Auto

Page 26: [144] Formatted	**Author**	

Font color: Auto

Page 26: [145] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [147] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [148] Formatted	**Author**

Font color: Auto

Page 26: [149] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 26: [150] Formatted	**Author**

Font color: Auto

Page 26: [151] Formatted	**Author**

Font color: Auto

Page 26: [152] Formatted	**Author**

Left

Page 26: [153] Formatted	**Author**

Font color: Auto

Page 26: [154] Formatted	**Author**

Font color: Auto

Page 26: [155] Formatted	**Author**

Font color: Auto

Page 26: [156] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [157] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [158] Formatted	**Author**

Font color: Auto

Page 26: [159] Formatted	**Author**

Font color: Auto

Page 26: [160] Formatted	**Author**

Font color: Black

Page 26: [161] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 26: [162] Formatted	**Author**

Left

Page 26: [164] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [165] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [166] Formatted	Author	

Font color: Auto

Page 26: [166] Formatted	Author	

Font color: Auto

Page 26: [167] Formatted	Author	

Font color: Auto

Page 26: [168] Formatted	Author	

Font color: Auto

Page 26: [169] Formatted	Author	

Left

Page 26: [170] Formatted	Author	

Font color: Auto

Page 26: [171] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [172] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [173] Formatted	Author	

Font color: Auto

Page 26: [174] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [175] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [176] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [177] Formatted	Author	

Font color: Auto

Page 26: [178] Formatted	Author	

Font color: Auto

Left

Page 26: [180] Formatted **Author**

Font color: Auto

Page 26: [181] Formatted **Author**

Font color: Auto

Page 26: [182] Formatted **Author**

Font color: Black

Page 26: [183] Formatted **Author**

Font color: Auto

Page 26: [184] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [185] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [186] Formatted **Author**

Left

Page 26: [187] Formatted **Author**

Font color: Auto

Page 26: [188] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [189] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [190] Formatted **Author**

Font color: Auto

Page 26: [191] Formatted **Author**

Font color: Auto

Page 26: [192] Formatted **Author**

Left

Page 26: [193] Formatted **Author**

Font color: Auto

Page 26: [194] Formatted **Author**

Font color: Auto

Page 26: [196] Formatted	Author	

Font color: Black

Page 26: [197] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [198] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [199] Formatted	Author	

Font color: Auto

Page 26: [200] Formatted	Author	

Left

Page 26: [201] Formatted	Author	

Font color: Auto

Page 26: [202] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [203] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [204] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [205] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [206] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 26: [207] Formatted	Author	

Font color: Auto

Page 26: [208] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [209] Formatted	Author	

Font color: Auto

Page 26: [210] Formatted	Author	

Font color: Black

Page 26: [211] Deleted Cells	Author	15/30/07 8:48:00 PM

Deleted Cells

Page 26: [213] Formatted	**Author**	

Left

Page 26: [214] Formatted	**Author**	

Font color: Auto

Page 26: [215] Formatted	**Author**	

Font color: Auto

Page 26: [216] Formatted	**Author**	

Font color: Black

Page 26: [217] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 26: [218] Formatted	**Author**	

Font color: Auto

Page 26: [219] Formatted	**Author**	

Font color: Auto

Page 26: [220] Formatted	**Author**	

Font color: Black

Page 26: [221] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [222] Formatted	**Author**	

Font color: Black

Page 27: [223] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [224] Formatted	**Author**	

Font color: Auto

Page 27: [225] Formatted	**Author**	

Left

Page 27: [226] Formatted	**Author**	

Font color: Auto

Page 27: [227] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [228] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Font color: Auto

Page 27: [230] Formatted **Author**

Font color: Black

Page 27: [231] Formatted **Author**

Font color: Auto

Page 27: [232] Formatted **Author**

Font color: Auto

Page 27: [233] Formatted **Author**

Left

Page 27: [234] Formatted **Author**

Font color: Auto

Page 27: [235] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [236] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [237] Formatted **Author**

Font color: Auto

Page 27: [238] Formatted **Author**

Font color: Auto

Page 27: [239] Deleted Cells **Author** **15/30/07 8:48:00 PM**

Deleted Cells

Page 27: [240] Formatted **Author**

Font color: Auto

Page 27: [241] Formatted **Author**

Left

Page 27: [242] Formatted **Author**

Font color: Auto

Page 27: [243] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [244] Inserted Cells **Author** **15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [246] Formatted	**Author**	

Font color: Auto

Page 27: [247] Formatted	**Author**	

Font color: Black

Page 27: [248] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 27: [249] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 27: [250] Deleted Cells	**Author**	**15/30/07 8:48:00 PM**

Deleted Cells

Page 27: [251] Formatted	**Author**	

Left

Page 27: [252] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [253] Formatted	**Author**	

Font color: Auto

Page 27: [254] Formatted	**Author**	

Font color: Auto

Page 27: [255] Formatted	**Author**	

Font color: Black

Page 27: [256] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [257] Inserted Cells	**Author**	**15/30/07 8:48:00 PM**

Inserted Cells

Page 27: [258] Formatted	**Author**	

Font color: Auto

Page 27: [259] Formatted	**Author**	

Font color: Black

Page 27: [260] Formatted	**Author**	

Left

Page 27: [261] Formatted	**Author**	

Font color: Auto

Page 27: [263] Formatted	Author	

Font color: Auto

Page 27: [264] Formatted	Author	

Font color: Auto

Page 27: [265] Inserted Cells	Author	15/30/07 8:48:00 PM

Inserted Cells

Page 27: [266] Formatted	Author	

Font color: Auto

Page 27: [267] Formatted	Author	

Font color: Auto

Page 27: [268] Formatted	Author	

Font color: Auto

Page 27: [269] Formatted	Author	

Font color: Auto

Page 27: [270] Formatted	Author	

Left

Page 27: [271] Formatted	Author	

Font color: Auto

Page 27: [272] Formatted	Author	

Font color: Auto

Page 27: [273] Formatted	Author	

Font color: Auto

Page 27: [274] Formatted	Author	

Font color: Auto

Page 27: [275] Formatted	Author	

Font color: Auto

Page 27: [276] Formatted	Author	

Font color: Auto

Page 27: [277] Formatted	Author	

Font color: Auto

Page 27: [278] Formatted	Author	

Font color: Auto

Left

Page 27: [280] Formatted **Author**

Font color: Auto

Page 27: [281] Formatted **Author**

Font color: Auto

Page 27: [282] Formatted **Author**

Font color: Auto

Page 27: [283] Formatted **Author**

Font color: Auto

Page 27: [284] Formatted **Author**

Font color: Auto

Page 27: [285] Formatted **Author**

Font color: Auto

Page 27: [286] Formatted **Author**

Font color: Auto

Page 27: [287] Formatted **Author**

Font color: Auto

Page 27: [288] Formatted **Author**

Left

Page 27: [289] Formatted **Author**

Font color: Auto

Page 27: [290] Formatted **Author**

Font color: Auto

Page 27: [291] Formatted **Author**

Font color: Auto

Page 27: [292] Formatted **Author**

Font color: Auto

Page 27: [293] Formatted **Author**

Font color: Auto

Page 27: [294] Formatted **Author**

Font color: Auto

Page 27: [296] Formatted	Author

Font color: Auto

Page 27: [297] Formatted	Author

Left

Page 27: [298] Formatted	Author

Font color: Auto

Page 27: [299] Formatted	Author

Font color: Auto

Page 27: [300] Formatted	Author

Font color: Auto

Page 27: [301] Formatted	Author

Font color: Auto

Page 27: [302] Formatted	Author

Font color: Auto

Page 27: [303] Formatted	Author

Font color: Auto

Page 27: [304] Formatted	Author

Font color: Auto

Page 27: [305] Formatted	Author

Font color: Auto

Page 27: [306] Formatted	Author

Left

Page 27: [307] Formatted	Author

Font color: Auto

Page 27: [308] Formatted	Author

Font color: Auto

Page 27: [309] Formatted	Author

Font color: Auto

Page 27: [310] Formatted	Author

Font color: Auto

Page 27: [311] Formatted	Author

Font color: Auto

Page 27: [313] Formatted **Author**

Font color: Auto

Page 27: [314] Formatted **Author**

Font color: Auto

Page 27: [315] Formatted **Author**

Left

Page 27: [316] Formatted **Author**

Font: Not Italic