Dear Editor of Geoscientific Model Development,

We have modified the manuscript according to the remaining comments and queries from the two anonymous reviewers. The main points that were modified concern the inclusion into the main text of a part of the information contained in the supplementary materials and clarification of some misleading information. To fulfill these recommendations, we have:

1) Moved Figure 4 of the Supplement into Section 4.2.2 or 4.2.3 according to the referee suggestion. This latter presents sensitivity of the sea-ice model to coupling frequency. Although not central in the study, this information is valuable to track change in skill of the present model (CNRM-ESM1) compared to the previous version (CNRM-CM5 and CNRM-CM5.2).

2) Detailed how the water conservation has been improved in the current with the respect to the previous model versions.

3) Given some details on carbon cycle components (e.g., the implicit nitrogen limitation used in the ISBA land carbon cycle module, redfield ratio in the carbon pools simulated in the PISCES marine biogeochemical model)

4) Included further evaluation diagnostics in the supplementary materials (Figure comparing skill of CNRM-ESM1 versus CNRM-CM5.2) to support between-model comparison.

5) Included intercomparison diagnostics following Anav et al. (2013) to strengthen the evaluation of CNRM-ESM1 and to present how it compares with other IPCC-class Earth system models.

Please find a detailed response to each questions/comments point by point below in blue (text fragments are in blue italics).
Reviewer #1:
The authors present the development and a first evaluation of the CNRM earth system model, which evolved from the climate model CNRM-CM5 by coupling modules for the land and ocean carbon cycle to the model. They evaluate the main physical drivers of the land and ocean carbon cycle and compare these results to the previous model version. A number of ecosystem parameters, biogeochemical tracers, and carbon fluxes are evaluated against observation based estimates. The paper is of high scientific relevance for the earth system modelling community and definitely within the scope of the journal. The model-model and model-data comparisons are sound and the paper is overall well written. There are, however, a number of issues listed below that should be addressed by the authors. I recommend to accept the manuscript for publication in GMD after these minor revisions have been addressed by the authors.

General comments:

Comment 1: The authors find a quite large sensitivity of sea ice area to the coupling time step of the model (24h vs. 6h), but they somewhat hide this result in the supplement. Since this issue is mentioned several times in the manuscript, I suggest moving Figure 4 of the Supplement into Section 4.2.2 or 4.2.3 and discussing it there. Also, a discussion as to why the coupling time step causes these differences would certainly be of interest for the audience (although I see that this could be beyond the scope of this model description).

We agree with the referee. Please see responses to comment 1 and 2 below.

Comment 2: In 4.2.2 there is a discussion about deep convection in the model and the fact that CNRM-ESM1 simulates open ocean polynyas, but not in the region where the so-called wedell polynya was observed in 1974-1976. CNRM-CM5 in contrast simulated polynyas in the wedell sea, and the authors state that "The use of GELATO6 in CNRM-ESM1 compared to GELATO5 in CNRM-CM5 in addition to the change in coupling frequency might be at the origin of this model-data disagreement."

While it is interesting that open ocean polynyas have been observed and are not only seen in model simulations, the number of occurrences is a bit low to call this a "model-data disagreement" in my opinion. Is there any evidence that the difference in location is due to the different version of the ice model and/or the timestep? Isn’t there a couple of other possible reasons in the model set-up? Altogether I find this paragraph on polynyas a bit too vague for a model evaluation paper, and I would suggest either deleting this discussion from the manuscript or go much more into the details as to why the location of polynyas is where it is in the different model versions (but this might as well be beyond the scope of this model description).

We agree with the referee on both these comments.
This point is not central in the paper since it could be developed and further analyzed in another study, focused on the ocean and the sea ice. However, we have chosen to present the impact of the coupling frequency on sea-ice distribution to track changes in skill between the various CNRM models. In the present model, the representation of the sea
ice has been degraded with respect to the previous model version. That said, we agree with the referee on the fact we have brought information that is not in the scope of the current paper, e.g., impact of the open ocean polynia or the impact on the deep ocean ventilation. Therefore, we have removed the above-mentioned paragraph and presenting the Figure S4 in the main text as an evaluation material to explain the difference in skill of CNRM-ESM1 compared to CNRM-CM5.

Comment 3: p 5696, l 14-15: "...the model-data mismatch is likely related to the decision of Takahashi et al. (2010) to exclude observations from El Niño years from their analysis [...] This hypothesis is validated when comparing model results against recent data products derived from statistical Monte-Carlo Markov Chain or Neural Network gapfilling methods (Landschützer et al., 2014; Majkut et al., 2014)." If the comparison with the latter data products gives better or different results, these should be presented here. Why compare to a data product with a known bias and then discuss that if we would compare to another available data product results would look better?

The reviewer is right but we have preferred to keep ‘direct’ observations from Takahashi et al. (2010) rather than observation-derived estimate like that of Landschützer et al., 2014. Nonetheless, we have provided Figure R1 as comparison in the supplementary materials.
Figure R1: Zonally-cumulated ocean carbon fluxes (fgCO$_2$) averaged over 1986-2005 from observation-derived estimates and as simulated by CNRM-ESM1. Observation-derived estimates from Takahashi et al. (2010) and Landschützer et al. (2014) are represented with blue and green solid lines, respectively. Result from CNRM-ESM1 is given with a solid red line.

Comment 4: Summary, p 5702, l 5-8: "This change is attributed to [...] a higher coupling frequency that induces a stronger northward flow of deep water masses from the Southern Ocean." In the manuscript, the authors do not provide any evidence or model sensitivity tests to support this hypothesis. Unless they do so in a revised version of the manuscript, I suggest deleting the statement on the coupling frequency here.

We apologize for this lack of clarity. Some results and discussions have been provided in the section 4-2-2 of the submitted version of the manuscript:

“In the Southern Ocean, the flow of Antarctic bottom water (AABW) is about 11.6±1 Sv"
in CNRM-ESM1 averaged over the 1850-2005 period. This flow of AABW is in agreement with the deep flow of waters compared to the observed estimate of 10 ± 2 Sv (Orsi et al., 1999). Consequently, the flow of deep water masses in CNRM-ESM1 is stronger than that of CNRM-CM5 (Séférian et al., 2013; Voldoire et al., 2013). “However, we have omitted to provide evidence of this improvement in order to compare our results to previous model version.

The text has been replaced by:

“This flow of AABW is in agreement with the deep flow of waters compared to the observed estimate of 10 ± 2 Sv (Orsi et al., 1999). Consequently, the flow of deep water masses in CNRM-ESM1 has been improved in regards that of CNRM-CM5 which ranges between 3.4 and 6.2 Sv over the same period (Séférian et al., 2013; Voldoire et al., 2013).”

In the summary, we have followed referee suggestion and reword the statement on the Southern Ocean circulation as follows.

Submitted:
“This change is attributed to improved water conservation in the ocean-sea ice model as well as a higher coupling frequency that induces a stronger northward flow of deep water masses from the Southern ocean.”

Revised:
“This change is attributed to a stronger northward flow of deep water masses from the Southern ocean.”

Specific comments:

In the abstract, please state to which time period the uptake values (2.2PgCyr-1 land sink, 1.7PgCyr-1 ocean sink) refer to.

Done and acknowledged

p 5673,l 6-10: "The models of this class ... primarily through their contribution to the concentration- and emission-driven experiments that compose CMIP5." I don’t really understand what the authors want to say here. The really new point with ESMs is that we can run emission driven (something that traditional climate models cannot)

We agree with the reviewer but ESM including carbon cycle components enables biophysical interactions like the impact of rising CO2 on evapotranspiration or the shading of ocean biota on light transfer and hence heat trapping. As such, ESM components, even under concentration-driven scenario, can generate feedback that traditional climate models do not simulate. Besides, ESM components have been investigated in terms of impacts like ocean acidification, variability etc… in various IPCC reports based on working groups 1 and 2.
Therefore, we chose to keep the sentence as it is.

p 5673,l 18: Please consider replacing "ensemble" by something else. "Ensemble" should
be reserved for different realisations of model runs of the same model, which the authors do not mean here, as far as I understand.

Done and acknowledged

p 5680, l 25-29: The nitrogen limitation of the Land BGC model is mentioned and a reference is given (Calvet et al.2008). Nitrogen limitation is a critical point in ESM simulations, and CMIP5 has shown that including it can potentially alter the reaction of land carbon uptake to enhanced CO2 and climate change dramatically. Therefore, I think it would be good to provide a brief summary of how the nitrogen limitation is implemented in the CNRM-ESM.

We agree with the reviewer. We have amended the text as follows.

Submitted:
“IT results that nitrogen dilution occurs as soon as the increase in total biomass of a plant under rising CO2 relative to growth under ambient CO2 is greater than the corresponding increase in total nitrogen.”

Revised:
“ISBA uses an implicit nitrogen limitation parameterization which is based on the meta-analysis of leaf nitrogen measurement under CO2 enrichment condition (Yin et al., 2002). This simple implicit nitrogen limitation is based on the nitrogen dilution hypothesis, which assumes that internal nitrogen content of a plant decrease under rising CO2 due to the accumulation of non-structural carbohydrates. It results that nitrogen dilution occurs as soon as the increase in total biomass of a plant under rising CO2 relative to growth under ambient CO2 is greater than the corresponding increase in total nitrogen. In current version of ISBA, a linear decrease between specific leaf area index and nitrogen to carbon ratio in leaves is used to mimic this mechanism (Calvet et al., 2008), and hence to limit the net assimilation of atmospheric CO2.”

This simple implicit nitrogen limitation is used to relate internal concentration of nitrogen to biomass and to limit the net assimilation of atmospheric CO2.

p 5682, l 21-22: "Only the internal concentrations of iron, silicon and calcite inside the sinking particles are prognostically simulated." I think this can be misleading: The amount of carbon transported down by sinking is also prognostically simulated, right? Only the ration C:P and C:N is fixed. Please clarify.

The information is given above in the main text:
“The ratios between carbon, nitrate and phosphate are kept constant to the values proposed by (Takahashi et al., 1985) in all living and nonliving pools of organic matter.” We chose to keep the sentence as it is.

p 5683, l 2-5: Nitrogen fixation is mentioned, (“... nitrogen fixation should balance...”) but it remains unclear how it is implemented. Could the authors please clarify.

We have clarified this sentence as follows:
Importantly, to ensure conservation of nitrogen in the ocean, annual total nitrogen fixation is adjusted to balance losses from denitrification following (Lipschultz et al., 1990; Middelburg et al., 1996; Soetaert et al., 2000). For the other macronutrients, alkalinity and organic carbon, the conservation is ensured by tuning the sedimental loss to the total external input from rivers and dust. Therefore, carbon and nitrogen cycles are decoupled to a certain degree.”

Section 3.1: In my opinion "Equilibrium strategy" is not a very good title for this section. The authors describe the spin-up of the model her, so "Spin-up strategy" or just "Model Spin-up" would be more suitable perhaps. Also, in line 7 a 320 year online spin-up is mentioned. Then towards the end of the section there is "an online adjustment ... for 400 years". Is this something different? Or is this the same as 320 years, but one of both is a typo? Please clarify, and (if the same) consider mentioning the online spin-up only once in this section.

We have changed the title as suggested by the referee
Besides, the review is right, the word ‘online’ (by opposition with offline) is not useful here because all the ESM components were spun-up within a coupled ESM.

p 5685 l 16: "...performed with the NCAR model (Ammann et al.,2007)." The "NCAR model" should be CCSM.

Done and acknowledged

p 5685 l 24: "...20th century reconstruction of the NCAR model (Ammann et al.,2003)." The "NCAR model" would be the CCSM, but in Ammann et al.(2003) they don’t use CCSM or another model developed by NCAR. Please make sure that the model name and reference are correct.

We apologize for this misleading information. Both preindustrial and historical simulations rely on the reconstruction from Ammann et al., (2007) performed with the NCAR Community Climate System Model.

p 5687, l 4-6: I do not understand how the land C flux could be "explained by missing processes in ISBA such as the [...] riverine-induced carbon transport from land to oceans". A riverine transport could close the carbon cycle in the sense that it would transport the excess C taken up by land into the ocean eventually. This would still result in a positive land C-flux (but the land pool size would be constant in equilibrium which cannot be the case in the current model setup).

The Figure represents the NBP, which is decomposed as follows:
With each flux >0
In the version of ISBA used in CNRM-ESM1, NBP = GPP – TER.
Therefore, missing processes like Fire, weathering or land-use land-cover change lead to overestimate NBP.

p 5688 l 3-4: "...appears to amplify the global average cold bias of 0.8°C (with biases of −0.7 and -1°C in boreal winter and summer, respectively)" I do not understand these numbers: Are these the differences between CNRM-CM5 and CNRM-ESM? The ESM biases have already been state above? Please clarify.

We have clarified our statement as follows: 
“Small deviations between CNRM-CM5 and CNRM-ESM1 mean state can be essentially attributed to the land carbon cycle, which appears to amplify the global average annual cold bias of 0.8°C (with seasonal differences between CNRM-ESM1 and CNRM-CM5 of -0.7°C and -1°C in boreal winter and summer, respectively).”

p 5688, l 20: "Summer PR is similar between the two models with non-significant changes in simulated values (< 10−5 mm day−1 )..." I find it difficult to believe that PR changes are that small between the two different models, particularly since during boreal winter bias is reduced by 0.014 mm day-1. Please check and/or clarify.

We thank the referee for his/her thoughtful comments. We have amended the text for sake of clarity.
We have changed following statement accordingly.
Submitted:
“The mismatch between simulated and GPCP observed PR over continents is slightly improved during the boreal winter with a bias reduced by 0.014 mm day−1 in CNRM-ESM1. Summer PR is similar between the two models with non-significant changes in simulated values (<10−5 mm day−1) but their geographical patterns have been slightly degraded in CNRM-ESM1 compared to CNRM-CM5. Although weak, changes induced by the ISBA biophysical coupling slightly improve the representation of the seasonal cycle in PR over Northern mid-latitude continents by amplifying the seasonal maximum and shifting it from June in CNRM-CM5 to July in CNRM-ESM1 (supplementary materials).”
Revised:
“Although weak, changes induced by the ISBA biophysical coupling slightly affect the representation of the seasonal cycle in PR over the vegetated regions (Figure S1). These lead to improve the simulated PR in CNRM-ESM1 compared to CNRM-CM5 over some vegetated regions during the growing season (spring-summer). Between 30°N and 60°N, the averaged error in simulated PR compared to GPCP is reduced by 0.12 mm day−1 in CNRM-ESM1 compared to that of CNRM-CM5. Over the tropics (30°S-30°N), simulated PR is also improved in CNRM-ESM1 compared to CNRM-CM5 but to a lesser extent with a reduction of the averaged error by 0.06 mm day−1 with respect to GPCP. Although PR have been improved over some region, their geographical patterns have been degraded in CNRM-ESM1 compared to CNRM-CM5, especially during the winter.”

Section 4.2.2, evaluation of MLD: The terms MLD_max and MLD_min are a bit
ambiguous. I guess it is not the real min and max over all model timesteps but rather the min and max of the mean over some period (probably monthly)? Is this time window comparable to the time resolution of the Sallee data? Also please consider to call the Sallee et al data "observation based estimate" or similar, not "observations" in the figure caption of Figure 7.

The reviewer is right both model and data are used with a monthly temporal frequency. The figure caption has been amended accordingly.

p 5693, l 8: "...but strongly overestimates the spatial variations of this field." From Figure 9 i read a standard deviation < 1 for MLD in both cases, which means that variability is underestimated compared to observations. Regarding Figure 9: Since the small symbols for seasonal values are there, they should be discussed (why are there two seasons where MLD is very different in the two models)?

We apologize for this misleading information. Erroneous computations were performed for the mixed-layer depth at seasonal timescale. They are now performed accurately and Figure 9 has been updated accordingly. Besides, the referee is right. CNRM-ESM1 underestimates the depth of the seasonal mixed-layer depth. We have corrected this mistake accordingly.

p 5693, l 14-15: "...difference in simulated SSS between the two models can be attributed to the revised water conservation interface and erroneous distribution of sea-ice cover. Besides, changes in coupling frequency (i.e. 24 to 6 h) might be at the origin of differences in skills between the two models." I do not understand what the "revised water conservation interface" is. If this leads to important differences it must be at least briefly described Section 2. Is there any evidence that the coupling frequency changes the model skill in reproducing SSS?

Following referee remarks, we have modified this statement as follows.
Submitted:
“Since CNRM-CM5.1, the coupling between NEMO and GELATO has been revised in order to improve the conservation of water and salt.”
Revised:
“Since CNRM-CM5.1, the coupling between NEMO and GELATO has been revised in order to improve the conservation of water and salt. In the previous model version, CNRM-CM5.1, there was a large drift in salinity (~0.011psu/century) and in sea level (~21 cm/century). These were caused by (1) the melting of land glaciers (other than Antarctic and Greenland) that was not routed to the ocean and (2) an erroneous coupling between sea-ice and ocean models. The coupling did not take into account the fact that sea ice is levitating over the ocean in this version of NEMO. Although not severe, it resulted in a loss of water in the model. These errors have been fixed in CNRM-CM5-2 and CNRM-ESM1 and hence reducing the residual drifts in salinity to +0.001psu/century and in sea level to +1.2cm/century.”
p 5699, l 5: replace "climate change" with "climate forcing". The climate change is seen in the indices looked at, which react to a forcing.

Done and acknowledged

Figure 2: The text says, land carbon flux is 0.75 PgCyr-1. From the Figure I would guess that the number must be more around 1.25. So I assume that either the Figure axis or the number in the text is wrong. Please check.
The reviewer is right but we are comparing two different period:
Land sink is about 0.75 Pg C y^-1 over the preindustrial period in our model. This value has been computed from an average flux over 400-year long piControl simulation.
The land sink in the early 20th century is about 1.2 Pg C y^-1 (in 1901-1910). This value accounts for both low frequency fluctuations (in response to climate mode of variability) and to a lesser extent rising of atmospheric CO2. This latter has risen by about 11 ppm from 1850 to 1901.

Technical corrections:
When referring to the supplement, could the authors please give the figure number, rather than just "(Supplement)".

Done and acknowledged

p 5675, l 3: acronym AOGCM not explained at first use

Done and acknowledged

p 5675, l 6: "...(ARPEGE-Climat, SURFEX, NEMO, GELATO, respectively)" this addition makes no sense without further explanation. Since all components are discussed below, I would just delete this.

Done and acknowledged

p 5675, l 8: "...was based on version 5.2. This version of the atmospheric code derives from cycle 37 of..." I think the authors mean their version 6.1, but as the sentence is constructed "This" refers to version 5.2. Please consider rephrasing to avoid confusion.

We agree with the referee. To avoid confusion, we have modified this sentence as follows.
Submitted:
"The atmospheric component is based on version 6.1 of the global spectral model ARPEGE-Climat whereas CNRM-CM5.1 was based on version 5.2. This version of the atmospheric code derives from cycle 37 of the ARPEGE-IFS (Integrated Forecast System) numerical weather prediction model developed jointly by Météo-France and the European Center for Medium-range Weather Forecast."

Revised:
"The atmospheric component is based on version 6.1 of the global spectral model ARPEGE-Climat which corresponds to an updated version of the atmospheric code used in CNRM-CM5.1. This updated version of the atmospheric code derives from cycle 37 of the ARPEGE-IFS (Integrated Forecast System) numerical weather prediction model developed jointly by Météo-France and the European Center for Medium-range Weather Forecast."

p 5675, l 23-24: "The main difference from the CNRM-CM5.1 atmospheric model is the improved treatment of volcanic aerosols." But this is not a difference in ARPEGE, right? If so, please consider rephrasing something like "The interactive chemistry module already used in CNRM-CM5.1 has been updated with an improved..."

We apologize for this misleading statement. Improvements concern only coefficients for the stratospheric chemistry module, not the atmospheric code. We have deleted this statement from the main text.

p 5678, l 1: "The coupling between the atmosphere and the surface models is implicit..." I do not understand what "implicit" should mean here. I guess, the land surface is a submodel of the atmosphere (organised as a subroutine call)? Please reword this sentence.

We have amended this sentence as follows.
Submitted:
"The coupling between the atmosphere and the surface models is implicit and occurs every atmospheric timestep (i.e., 30 minutes) while the coupling between the atmosphere and the ocean models is handled by the OASIS coupler (Valcke, 2013) and occurs every 6 hours."

Revised:
"In CNRM-ESM1, exchanges of momentum, water and energy between the atmosphere and the surface models occurs every atmospheric timestep (i.e., 30 minutes) because SURFEX is a submodel of the atmospheric code. The coupling between the atmosphere and the ocean models is handled by the OASIS coupler (Valcke, 2013) and occurs every 6 hours."

p 5678, l 21: "... the recommendations of the JPL-2003-25 report (Sander et al., 2006)." The "JPL-2003-25" is not helpful for the reader, consider rephrasing: "...the recommendations of Sander et al. (2006)." Please also correct the reference Sander et al. (2006) in the References section.

Done and acknowledged

p 5678, l 21-23: "Photochemical production and loss rates of ozone rely on the main gas-phase reactions driving the NOx, HOx, ClOx, BrOx catalytic cycles." I think it would be easier to understand, if this sentence would come before the previous one.

Done and acknowledged
In the present concentration-driven experiments...

In the concentration-driven experiments presented here...

...provided by CMIP5 according to the CMIP5 protocol

3 non-vegetated surfaces -> 3 non-vegetated surface types

This is a key advantage of this approach as most the...

A key advantage of this approach is that most of the ... without any additional parameters needed.

The ocean biogeochemical model... or similar

Dependence of growth on...

Dependence of growth on

into -> in

either "following the formulation of Geider et al..." or remove "formulati-

tion".

What is "Princeton atmospheric forcing"? Please provide a reference here.

detailed in (Voldoire et al., 2013)

detailed in Voldoire et al., 2013

acronym GPCP should be explained
Section 4.2.3: CNRM-CM5.2 is used while CNRM-CM5.2 is used in figure caption. Please use consistent names.

p 5694, l 6: "Figure 10 shows that the amplitude of annual mean GPP as..." Fig 10 shows the annual mean, I suggest deleting "the amplitude of".

p 5694, l 9: "...patterns of high GPP (values)." Please add values or remove the "(values)"

p 5694, l 18: "...Princeton university forcings (REF)..." Yes, please add a reference.

p 5700, l 5: "... uptake of CO2 of about 2.1 and 1.7 Pg C y$^-1$ for land and ocean..." Which year do these numbers correspond to? Is it 2005, or a mean over the last decade?

Figure 6d: The scale could be narrower (perhaps -2 to 2 psu?), since larger biases seem to occur only at the surface and the surface bias is already depicted in 6b. As it is now it is difficult to see the structures of the salinity biases at depth.

Figure 9a: In the figure legend it says PAR is given in the figure caption it says RSDS. Please check and correct.

Figure 13: This figure is really difficult to read, since the different sizes of the symbols are difficult to distinguish. I think it would be much easier for the reader to have one panel per depth with all 4 tracers.
Reviewer #2:
The article by Séférian et al. presents the first Earth-System Model developed at CNRM. It is based on the CNRM-CM5 coupled climate model (Voldoire et al. 2013) and includes the ISBA module for a representation of the land surfaces in the carbon cycle. The ocean biogeochemistry (PISCES) was already implemented in CNRM-CM5. The subject of the paper particularly fits within the scope of GMD. The overall structure of the paper and the quality of the science are very good, and the paper is well written. The main features of the physical behavior of the model are presented and evaluated. The protocol of the preparation of the model and the realization of the simulations (spin-up, pre-industrial control and historical simulation) are rigorously done. The amount of work and knowledge is impressive. Without a doubt, the authors are true experts in their fields.

We appreciate the reviewer’s careful reading and suggestions for corrections. Most of his/her suggestions and comments are addressed in the revised manuscript.

However, the interpretation of the evaluation results deserves revisions (specific comments below). The authors use words like “realistically”, “moderate” or “reasonably well-simulated” to describe the behavior of the model compared with observations. This is totally subjective and does not have its place in such a paper. I understand what the authors mean by this, and it would be acceptable in an oral presentation, but not in a scientific paper. These statements should be supported by more objective quantifications or comparisons with other models. For instance, how can we say that a model is “realistic”, or an agreement is “moderate”? Two different paths can be considered: either using statistics that objectively quantify the agreement (for instance, “my model explains XX% of the variance of the observations, and we assume that above [a meaningful threshold], the model meaningfully reproduces the observations on this diagnostic”), or by comparing the results with other models that will be used as benchmark. Both approaches are used in the paper (notably the comparison with CNRM-CM5) but not systematically. For instance, Figures 6, 7 and 8 of the supplementary materials are among the most interesting figures for they put the model in the multi-model intercomparison context. I strongly recommend taking advantage of the availability of the CMIP5 ESMs outputs to strengthen the evaluation of CNRM-ESM1 and see how it compares with other models. One way to do it is a portrait diagram showing synthetic metrics of model-data agreement (see Anav et al. 2013, Fig18) with the results of other ESMs for the carbon cycle variables. Nevertheless, this is only a possibility and the most important point to me is to remove, as much as possible, the subjectivity from the presentation of your results. In the specific comments I point out the issues that need more robust support.

One truly annoying point is the use of the supplementary materials. The authors do not specify which figure is pointed out by the reference to the supplementary materials, and the reader has to guess which figure he should look for. I still don’t
know if you use Figure 1 or not. Following this, I recommend the publication of the paper in GMD after major revisions.

We agree with the referee’s comments and we have strengthened the model evaluation using quantitative evaluation in the revised manuscript.

Specific comments

Abstract: the end of the abstract is about the too strong flow of North Atlantic Deep Water and the accumulation of anthropogenic carbon in the deep ocean. Meanwhile, those points are not reported in the summary and conclusion section (section 5). I suggest either presenting these points more comprehensively in section 5 or removing them from the abstract (maybe focus on something else) so that the abstract and last section are aligned in terms of contents and highlights. As well, it is a shame to end the abstract the way it is right now, with a quite negative statement. A focus on the good performances of the model would be more appropriate.

We understand the referee point of view. However, to our point of view, it is fundamental to highlight model weaknesses in an evaluation paper and hence to point out future leads of improvements. We have thus chosen to include a small paragraph on the biases in anthropogenic carbon storage in the conclusion section.

Submitted:
“This change is attributed to improved water conservation in the ocean-sea ice model as well as a higher coupling frequency that induces a stronger northward flow of deep water masses from the Southern ocean. While the simulated anthropogenic carbon storage agrees with 1994 observation-based estimates, the ocean carbon sink falls within the lower range of the combination of observation and model estimates over the recent years (Le Quéré et al., 2014).”

Revised:
“This change is attributed to a stronger northward flow of deep water masses from the Southern ocean which improves the vertical distribution of biogeochemical tracers. However, the strengthening of the meridional flow of deep water masses has also distorted the vertical structure of some carbon-related fields. Indeed, the unrealistic flow of North Atlantic deep water of about 26.1 Sv tends to deplete the stock of anthropogenic carbon storage between surface and 1200 m (Figure 16c) and consequently to increase it at depth. Since biases in anthropogenic carbon storage compensate across the water column, the simulated anthropogenic carbon storage agrees with 1994 observation-based estimates. Regarding the ocean carbon sink, CNRM-ESM1 simulates a global ocean carbon sink that falls within the lower range of the combination of observation and model estimates over the recent years (Le Quéré et al., 2014).”

Page 5688, line 20: the expression ‘non-significant’ suggests that there is a statistical support behind this statement. Change it for something like ‘very
small’ and add an estimate of this change in percentage (of the average field for example).

Please see the response below.

Page 5688, line 25: please specify which figure in the supplementary materials is targeted by this mention. I assume you mean Figure 2, but add it for the convenience of the reader. Additionally, I’m not convinced by this figure. Where am I supposed to see an improvement between CM5 and ESM1? Either I don’t understand the meaning of the figure therefore I assume it is not correctly explained or there is no comparison with observations on this figure and I don’t see how it should be able to show any improvement. Please clarify.

According to the comments of the both referees, we have changed the wording of this section for sake of clarity.

Submitted:
“*The mismatch between simulated and GPCP observed PR over continents is slightly improved during the boreal winter with a bias reduced by 0.014 mm day$^{-1}$ in CNRM-ESM1. Summer PR is similar between the two models with non-significant changes in simulated values ($\leq 10^{-5}$ mm day$^{-1}$) but their geographical patterns have been slightly degraded in CNRM-ESM1 compared to CNRM-CM5. Although weak, changes induced by the ISBA biophysical coupling slightly improve the representation of the seasonal cycle in PR over Northern mid-latitude continents by amplifying the seasonal maximum and shifting it from June in CNRM-CM5 to July in CNRM-ESM1 (supplementary materials).*”

Revised:
“*Although weak, changes induced by the ISBA biophysical coupling slightly affect the representation of the seasonal cycle in PR over the vegetated regions (Figure S1). These lead to improve the simulated PR in CNRM-ESM1 compared to CNRM-CM5 over some vegetated regions during the growing season (spring-summer). Between 30°N and 60°N, the averaged error in simulated PR compared to GPCP is reduced by 0.12 mm day$^{-1}$ in CNRM-ESM1 compared to that of CNRM-CM5. Over the tropics (30°S-30°N), simulated PR is also improved in CNRM-ESM1 compared to CNRM-CM5 but to a lesser extent with a reduction of the averaged error by 0.06 mm day$^{-1}$ with respect to GPCP. Although PR have been improved over some region, their geographical patterns have been degraded in CNRM-ESM1 compared to CNRM-CM5, especially during the winter.*”

Page 5693, line 16; page 5694, line 12, Line 662, Lines 687-688: specify which figure in the supplementary materials is targeted by this mention for the reader’s convenience.

We apologize for mistakes. Supplementary figures are now clearly mentioned in the revised manuscript.
Page 5689, lines 16-18: what I see on Figure 6ab is a bias map, i.e. a quantification of model errors. I have no particular mean to tell if those biases are sufficiently small to tell that the model “realistically simulates” the observed climate (same for “moderate” on line 18). I suggest either adding the results of CNRM-CM5 (either on this figure or another one) to put this statement in perspective of a model that can be used as a benchmark, or add an objective quantification of the model-data agreement that could make more sense than qualifications like ‘realistic’.

We agree with the referee. Yet, this section is essentially dedicated to the skill-assessment of CNRM-ESM1 ocean drivers. We include a basic metrics from CNRM-CM5 to strengthen and support our model-data assessment. The objective intercomparison of CNRM-ESM1 to previous model version is detailed 4.2.3. In agreement with referee comments, we have included new supplementary figures comparing model-data agreement for CNRM-ESM1 and CNRM-CM5 to support statistical metrics summarized in the Taylor diagram (Figure 9 in the submitted manuscript).

Page 5690, line 1-2: I don’t understand how you can say that the vertical structure of S matches better with WOA2013 than T. They don’t have the same units, and we have no idea about the uncertainty linked with the diagnostic to judge if the difference is smaller for S than T. Either provide objective elements to support your statement, or remove this sentence.

According to the referee’s comments, we have amended the section as follows.

Submitted:
“At depth, the vertical structures in T and S present moderate deviations from WOA2013 observations. T is underestimated by ~2°C within the first 1000 m of the Atlantic and Pacific oceans, except in the deep water formation zone (North Atlantic, North Pacific and Southern Ocean), where the model displays positive biases. Contrasting to T, the vertical structure of S matches well with WOA2013 observations. Deviations from the observed S profile are found in deep water formation zones where haline biases tend to compensate for the warm bias in T, enabling deep convection of water masses.”

Revised:
“At depth, the vertical structures in simulated T and S display biases from those estimated from WOA2013 observations. T is underestimated by ~2°C within the first 1000 m of the Atlantic and Pacific oceans, except in the deep water formation zone (North Atlantic, North Pacific and Southern Ocean), where the model displays positive biases. The largest deviation in vertical structure of simulated S from that estimated from WOA2013 are found in deep water formation zones where haline biases of about ~1 psu tend to compensate for the warm bias in T, enabling deep convection of water masses.”

Page 5691, lines 3-4: What are your criteria to say that the model ‘reproduces well’ the pattern of the observations? Replace ‘CNRM-ESM1 reproduces well the regional pattern of MLDmax’ with ‘CNRM-ESM1 simulates the main regional features of MLDmax’.
Done and acknowledged.

Page 5691, lines 19-20: This sentence is too vague. Either remove it, or put this result in perspective with another model (if so, we need to see the figure).

Right, we have amended the section as follows.

Submitted:
“Compared to the observation-derived estimates, the MLD_{min} is reasonably well simulated in CNRM-ESM1. Compared to the previous model versions (Séférian et al., 2013; Voldoire et al., 2013), CNRM-ESM1 fails at reproducing the deepest values of mixing in the Southern Ocean and the Tropics. These are tightly linked to the current parameterization of the ocean mixing employed in CNRM-ESM1 implying a contribution of the surface wind energy to the mixing below the MLD.”

Revised:
“Compared to the observation-derived estimates, CNRM-ESM1 captures the main regional pattern of MLD_{min} but the model fails at reproducing the deepest values of mixing in the Southern Ocean and the Tropics. This bias might be linked to the current parameterization of the ocean mixing employed in CNRM-ESM1 because previous model version using this parameterization also exhibited similar patterns of errors as detailed in (Séférian et al., 2013; Voldoire et al., 2013).”

Section 4-2-3: I have an issue with this section. You talk about skills that are ‘different’, or ‘similar’. But objectively, on your plot, you don’t have the means to say that two correlations are truly different (statistically speaking). It is even more the case when you compare the behavior of the correlations for two different variables, or two different seasons, since they are not computed on the same number of degrees of freedom. Your statement “the difference in simulated SSS between the two models can be attributed to [...]” is too strong compared with the true ability of those metrics to demonstrate what you say. In my opinion, metrics are mainly good tools to highlight outliers, but not so much to demonstrate any physical link (even if, sometimes, we can ‘understand’ their behavior). I would be much more convinced by a set of well-chosen maps of the biases of SSS, PR and sea ice, with both model versions (CM5 and ESM) side to side. They would show a much more reliable proof of the difference between the models and where does it come from. By the way, Figure 3e of the supplementary materials looks weird to me. On panel (a) mainly, there is sea ice in the Labrador Sea (between 0-10%). There is no sea ice in this region in panel (b), thus the difference should not be zero in panel (c) in the Labrador Sea. You might have an issue with the display of your results.

We thank the referee for his/her careful reading.
Indeed, correlation coefficients computed with at annual time scales have to be compared with each other, not with those computed at seasonal time scales. With that said, the referee is right a Taylor diagram has to be supported with maps of model-data errors. We have therefore included several supplementary Figures of model-data errors.
for temperature, salinity and mixed-layer depth from both CNRM-ESM1 and CNRM-CM5.2 to support the Taylor diagram presented in the manuscript.

Then, in agreement with first reviewer comments, we have moved Figure S3 into the main text as Figure 10. This latter will help the reader to better understand the influence of the ocean-atmosphere coupling frequency on sea-ice modeling. This Figure has been updated and corrected in agreement with above-mentioned comments concerning Figure 8.

Page 5696, lines 10-21: What about removing the El-Nino years before the computation of your climatology of CNRM-ESM1 to try to match the methodology of Takahashi et al. (2010)? Either provide a justification for not doing it, or provide an additional figure with the El-Niño years removed.

In agreement with the comments and suggestions of the first referee, we have chosen to provide a comparison with another data product which includes El-Niño influence on sea-air carbon fluxes (Figure R1). Figure R1 is also provided in supplementary materials.
Figure R1: Zonally-cumulated ocean carbon fluxes (fgCO$_2$) averaged over 1986-2005 from observation-derived estimates and as simulated by CNRM-ESM1. Observation-derived estimates from Takahashi et al. (2010) and Landschützer et al. (2014) are represented with blue and green solid lines, respectively. Result from CNRM-ESM1 is given with a solid red line.

Minor points

Page 5673, line 6-7: suggest adding a reference to Flato (2011)

Done and acknowledged

Page 5693, line 8: looking at the Taylor Diagram, the standard-deviation ratio for MLD (on average) is around 0.5. Therefore, it shows that the model underestimates (rather than overestimates) the spatial variations of this field compared with the observation-derived MLD product of Sallée et al. 2010.

Done and acknowledged

Page 5690, line 5: suggest changing “Thanks to” with “Because of”

Done and acknowledged

Page 5690, line 27: define in this sentence what MLDmax and MLDmin are (for convenience), or explicitly refer to the caption of the figure.

Done and acknowledged

Figure 8: the color bar seems to be incoherent with the contour line for the model. If I’m correct, the dashed contour line highlights the isoline 15 but it remains within the 0-10 range of the color field. This is really a minor point but you might have a look at it to correct it.

We appreciate reviewer careful reading. There was indeed an error in the sea concentration labeling. This error is now corrected in the revised manuscript.

Page 5693, lines 5-6: add ‘apart from a tendency to show higher variability (standard-deviation ratio).

This point is especially true for the mixed-layer depth and was already mentioned in this subsection of the submitted manuscript.

Page 5694, line 11: replace “lesser extend” with “lower extent”

Done and acknowledged
Figures 10, 11, 12: the color palette with white in the middle should be used for differences (as on Figure 5), not for a full field. Change it for a meaningful color palette (i.e. in the same way as the color palette you use for sea ice cover). I would also suggest adding the difference maps for all those fields. These to avoid leaving the reader playing a game of ‘guess where it’s greener’.

We agree with the referee regarding the choice of the color palette. We have updated the Figures accordingly.

Regarding the general comments of the referee, we do not want to include too much materials in the manuscript since an objective skill-assessment of CNRM-ESM1 versus other IPCC-class Earth system model is now provided in the main text to support the discussion section.

Page 5694, line 18: it looks like the authors forgot to change (REF) with the right reference of the Princeton University Forcing

Done and acknowledged

Page 5695, line 10: replace “compared to” with “than”

Done and acknowledged

Page 5696, line 8: I’m not too sure about the expression “carbon cycling”. . . I suggest double-checking that it is truly correct.

This is a correct wording.

Page 5696, line 8: replace ‘In term of’ with ‘In terms of’

Done and acknowledged

Page 5698, lines 24-26: need for a reference to support what you say about the origin of the biases in the equatorial upwelling systems.

This point is discussed in Landshutzer et al. (2014) which is already mentioned in this section. To support our statement, we have included comparison to Lanschutzer et al. data product for sea-air CO2 fluxes as supplementary figure.

Page 5699, line 19: replace “following” with “after” to avoid repetition

Done and acknowledged

Page 5701, line 24: replace “discrepancies similar to” with “similar performances as” (more positive)
Done and acknowledged
References:


Development and evaluation of CNRM Earth-System model
– CNRM-ESM1

R. Séférian¹, C. Delire¹, B. Decharme¹, A. Voldoire¹, D. Salas y Melia³, M. Chevallier¹, D. Saint-Martin¹, O. Aumont², J.-C. Calvet¹, D. Carrer¹, H. Douville¹, L. Franchistéguy¹, E. Joetzjer¹, S. Sénési¹

¹ CNRM-GAME, Centre National de Recherches Météorologiques-Groupe d’Etude de l’Atmosphère Météorologique, Météo-France/CNRS, 42 Avenue Gaspard Coriolis, 31057 Toulouse, France

² Sorbonne Universités (UPMC, Univ Paris 06)-CNRS-IRD-MNHN, LOCEAN-IPSL Laboratory, 4 Place Jussieu, F-75005 Paris, France

³ Montana State University, Institute on Ecosystems, Department of Ecology, 111 AJM Johnson Hall, Bozeman, Montana 59717, USA

Abstract:

We document the first version of the Centre National de Recherches Météorologiques Earth system model (CNRM-ESM1). This model is based on the physical core of the CNRM-CM5 model and employs the Interactions between Soil, Biosphere and Atmosphere (ISBA) and the Pelagic Interaction Scheme for Carbon and Ecosystem Studies (PISCES) as terrestrial and oceanic components of the global carbon cycle. We describe a preindustrial and 20th century climate simulation following the CMIP5 protocol. We detail how the various carbon reservoirs were initialized and analyze the behavior of the carbon cycle and its prominent physical drivers. Over the 1986-2005 period, CNRM-ESM1 reproduces satisfactorily several aspects of the modern carbon cycle. On land, the model captures the carbon cycling through vegetation and soil, resulting in a net terrestrial carbon sink of 2.2 Pg C y⁻¹. In the ocean, the large-scale distribution of hydrodynamical and biogeochemical tracers agrees with a modern climatology from the World Ocean Atlas. The combination of biological and physical processes induces a net CO₂ uptake of 1.7 Pg C y⁻¹ that falls within the range of recent...
estimates. Our analysis shows that the atmospheric climate of CNRM-ESM1 compares well with that of CNRM-CM5. Biases in precipitation and shortwave radiation over the Tropics generate errors in gross primary productivity and ecosystem respiration. Compared to CNRM-CM5, the revised ocean-sea ice coupling has modified the sea-ice cover and ocean ventilation, unrealistically strengthening the flow of North Atlantic deep water (26.1±2 Sv). It results in an accumulation of anthropogenic carbon in the deep ocean.

1. Introduction

Earth system models (ESMs) are now recognized as the current state-of-the-art models (IPCC, 2013), expanding the numerical representation of the climate system of the 4th Assessment Report (IPCC, 2007). They enable the representation of subtle non-linear interactions and feedbacks of different magnitude and signs of various biogeochemical and biophysical processes with the climate system. The latter contribute, in addition to the atmospheric radiative properties and global climate dynamics, to determine the Earth’s climate variability (Arora et al., 2013; Cox et al., 2000; Friedlingstein and Prentice, 2010; Schwinger et al., 2014; Wetzel et al., 2006).

Although there is no uniformly accepted definition, ESMs generally bring together a global physical climate model and land and ocean biogeochemical modules (Bretherton, 1985; Flato, 2011). As such, they enable the representation of the global carbon cycle. The models of this class have played a larger role in the 5th IPCC report than in previous reports, primarily through their contribution to the concentration- and emission-driven experiments that compose CMIP5.

Even if the concept of Earth system modeling is being extended to include further processes and reservoirs (e.g., nitrogen cycle, aerosols) (Hajima et al., 2014), there are still large uncertainties in the representation of the carbon cycle and its interactions with climate (Anav et al., 2013a; Friedlingstein et al., 2013; Piao et al., 2013). To reduce them, there is a need for improvements of both physical and ecophysiological
parameterizations (Dalonech et al., 2014), and for the development of observation-based methods to constrain model projections (Wenzel et al., 2014). But the reduction of carbon cycle-climate uncertainties also requires a greater number and diversity of ESMs. This path is promoted and followed by various international initiatives like the Global Carbon Budget (http://www.globalcarbonproject.org/) that sequentially incorporate more and more models into their analyses (Le Quéré et al., 2013; 2015).

This manuscript documents the first IPCC-class ESM developed at Centre National de Recherches Météorologiques and provides a basic evaluation of the model’s skill. This model is based on the CNRM-CM5.1 climate model jointly developed by CNRM and Cerfacs (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique), which has contributed to the fifth phase of the Coupled Model Inter-comparison Project (CMIP5) (Voldoire et al., 2013). CNRM-CM5.1 did not include a representation of the global carbon cycle but accounted for chemical-climate interactions with an interactive stratospheric chemistry module (Cariolle and Teyssedre, 2007). While this configuration of CNRM-CM5 contributed to the CMIP5 results publicly released, a first intermediate version of the CNRM ESM was developed with the inclusion of the marine biogeochemistry model PISCES (Aumont and Bopp, 2006). This model version was evaluated against modern oceanic observations (Séférian et al., 2013) and employed in various studies (Frölicher et al., 2014; Laufkötter et al., 2015; Schwinger et al., 2014; Séférian et al., 2014).

A terrestrial carbon cycle module is being developed at CNRM since the 2000s (Calvet and Soussana, 2001; Calvet et al., 2008; 2004; Gibelin et al., 2008; 2006), but it has never been coupled to an atmosphere-ocean model. This carbon cycle module evolved from the physically-based ISBA model (Noilhan and Mahfouf, 1996; Noilhan and Planton, 1989) and is able to simulate the surface carbon fluxes and the terrestrial carbon pools. The carbon fluxes module was extensively tested over France and Europe (Sarrat et al., 2007; Szczypa et al., 2012), and the carbon cycle module was tested for temperate and high latitude regions (Gibelin et al., 2006; 2008) and was used more recently in studies of carbon cycling over the Amazon basin (Joetzjer et al., 2015; 2014), permafrost regions (Rawlins et al., 2015) and at global scale (Carrer et al., 2013b). In this work, this terrestrial carbon cycle module is coupled to a global climate model for the first time.
Here, we present a first evaluation of the CNRM-ESM1. In section 2, we describe the
model, focusing on the Earth system’s components and aspects of the climate model
that are particularly relevant to the global carbon cycle. We describe in section 3 the
pre-industrial control and 20\textsuperscript{th} century experiments that we conducted, together with
the forcings used and how the experiments were initialized. In section 4, we present
and discuss the results of these experiments. We summarize the results in section 5
and present conclusions.

2- CNRM-ESM components

2-1 The physical core

CNRM-ESM1 is based on the physical core of the CNRM-CM5.1 Atmosphere-Ocean
General Circulation Model, extensively described in Voldoire et al. (2013), which
accounts for the physical and dynamical interactions occurring between atmosphere,
land, ocean and sea-ice.

The atmospheric component is based on version 6.1 of the global spectral model
ARPEGE-Climat which corresponds to an updated version of the atmospheric code
used in CNRM-CM5.1. This updated version of the atmospheric code derives from
cycle 37 of the ARPEGE-IFS (Integrated Forecast System) numerical weather
prediction model developed jointly by Météo-France and the European Center for
Medium-range Weather Forecast. In CNRM-ESM1, the geometry, parameterizations
and dynamics have been chosen to match the choices made for CNRM-CM5.1. Thus
differences are mainly due to debugging and recoding. The atmospheric physics and
dynamics are solved on a T127 triangular truncation that offers a spatial resolution of
about 1.4\degree in both longitude and latitude. Consistently to CNRM-CM5.1, CNRM-
ESM1 employs a “low-top” configuration with 31 vertical levels that extend from
the surface to 10 hPa in the stratosphere. The layers are unevenly distributed with 6
layers below 850 hPa except in regions of high orography, nine layers above 200 hPa
and four layers above 100 hPa. The dynamical core of the model, the radiative scheme
for longwave and shortwave as well as the physical parameterization for deep and
shallow convection are identical to those employed in CNRM-CM5.1. The reader is
referred to Voldoire et al. (2013) for the original description of the atmospheric model parameterizations.

The land-surface component is an updated version of the SURFface EXternalisé modeling platform (SURFEx7.3) (Masson et al., 2013b) associated with the Total Runoff Integrating Pathways (TRIP) river routing model (Oki and Sud, 1997). SURFEX was designed so that the same code could be run offline or coupled to a GCM, to allow easy transfer from offline improvements to the coupled model and to be able to compare online and offline runs.

This model prognostically computes the exchange of energy, water and carbon between the atmosphere and three types of natural surfaces: Land, free water bodies, oceans or seas. The energy, water and carbon balances are calculated separately for each surface type and area-averaged over each atmospheric grid cell. The natural land surfaces are represented by the module originally developed by Noilhan and Planton (1989). This module solves the surface energy and soil water budgets using the force-restore method and a composite soil-vegetation-snow approach. The version used here is the same as for CNRM-CM5.1: e.g. the soil hydrology uses 3 vertical layers (Boone et al., 1999) while soil temperature is solved using 4 vertical layers. In CNRM-ESM1, land surface albedo benefits from an improved spatial representation derived from MODIS satellite measurements (Carrer et al., 2013a) except for the area covered by snow for which the albedo is prognostically computed following Douville et al. (1995). Over water bodies and oceans, we use the CNRM-CM5.1 parameterization for momentum and energy fluxes except for the sea-to-air turbulent fluxes that are computed from the COARE scheme (Fairall et al., 2003). Interactions between the land surface energy and water budgets and the terrestrial carbon cycle module are detailed in section 2.3.1.

The ocean component uses version 3.2 of the NEMO model (Madec, 2008) in the ORCA1L42 configuration. This configuration offers a horizontal resolution from 1° to 1/3° near the equator and 42 levels in depth. The vertical discretization uses a partial-step formulation (Barnier et al., 2006), which ensures a better representation of bottom bathymetry and thus stream flow and friction at the bottom of the ocean. Ocean dynamics and physics is solved using a timestep of 1 hour. Vertical physics
relies on the parameterization chosen for the CNRM-CM5.1 climate model. The mixed layer dynamics is parameterized using a double diffusion process (Merryfield et al., 1999), Langmuir cell (Axell, 2002) and account for the contribution of surface wave breaking (Mellor and Blumberg, 2004). A parameterization of bottom intensified tidal-driven mixing similar to Simmons et al. (2004) is used in combination with a specific tidal mixing parameterization in the Indonesian area (Koch-Larrouy et al., 2010; 2007). Finally, CNRM-ESM1 benefits from an improved Turbulent Kinetic Energy (TKE) closure scheme (Madec, 2008), based on the Blanke and Delecluse (1993) TKE. This parameterization allows a fraction of surface wind energy to penetrate below the base of the mixed layer ensuring a better coupling between surface wind and subsurface mixing. The main difference from the CNRM-CM5.1 ocean model is the explicit modulation of the radiative shortwave penetration into the ocean by marine biota (Lengaigne et al., 2009; Mignot et al., 2013), which is further detailed in section 2.3.2.

The sea-ice model used in CNRM-ESM1 is GELATO6. This model employs the same horizontal grid as NEMO and solves sea-ice dynamics and thermodynamics every 6 hours. This model represents an updated version of the former sea-ice model used in CNRM-CM5.1 (Voldoire et al., 2013). In GELATO6, sea-ice dynamics is computed using the Elastic-Viscous-Plastic scheme proposed by Hunke and Dukowicz (1997) formulated on an Arakawa C-grid (Bouillon et al., 2009). To simulate the response of sea ice to convergence-divergence movements, GELATO6 employs a redistribution scheme derived from Thorndike et al. (1975). This scheme ensures the representation of the rafting phenomenon for the slab of sea ice thinner than 0.25 m and of ridging for the slab thicker than 0.25 m. GELATO6 includes a thermodynamic scheme that resolves the evolution of four ice thickness categories (0–0.3, 0.3–0.8, 0.8–3 and over 3 m). These four slabs of sea ice are modeled with 10 vertical layers unevenly distributed across the slab thickness. An enhanced resolution at the top of the slab is used to better represent the evolution of sea ice in response to the high frequency variability of the atmospheric thermal forcing. Besides, all sea-ice slabs may be covered with one snow layer. In GELATO6, the snow layer is considered to occult the transfer of light across the snow-sea ice-ocean continuum. This snow layer can age or form ice using the formulation described in Salas y Mélia (2002). Since CNRM-CM5.1, the coupling between NEMO and GELATO has been
revised in order to improve the conservation of water and salt. In the previous model version, CNRM-CM5.1, there was a large drift in salinity (-0.011 psu/century) and in sea level (-21 cm/century). These were caused by (1) the melting of land glaciers (other than Antarctic and Greenland) that was not routed to the ocean and (2) an erroneous coupling between sea-ice and ocean models. The coupling did not take into account the fact that sea ice is levitating over the ocean in this version of NEMO. Although not severe, it resulted in a loss of water in the model. These errors have been fixed in CNRM-CM5-2 and CNRM-ESM1 and hence reducing the residual drifts in salinity to +0.001 psu/century and in sea level to +1.2 cm/century.

In CNRM-ESM1, exchanges of momentum, water and energy between the atmosphere and the surface models occurs every atmospheric timestep (i.e., 30 minutes) because SURFEX is a submodel of the atmospheric code. The coupling between the atmosphere and the ocean models is handled by the OASIS coupler (Valcke, 2013) and occurs every 6 hours. In CNRM-ESM1, the frequency of coupling between the ocean and atmosphere models has been increased compared to CNRM-CM5 in order to better resolve the dynamics of the sea-ice, which is resolved at this timestep (i.e., 6 hours).

2-2 Atmospheric chemistry

The atmospheric chemistry scheme in CNRM-ESM1 consists of an interactive linear ozone chemistry model MOBIDIC (Cariolle and Teyssedre, 2007) including a representation of the three-dimensional atmospheric CO$_2$ mixing ratio.

As in CNRM-CM5, the ozone mixing ratio is treated as a prognostic variable with photochemical production and loss rates climatology computed by a full chemistry scheme. That is, the net photochemical production in the ozone continuity equation is solved using a first-order Taylor series around the local value of the ozone mixing ratio, air temperature, and the overhead ozone column. Ozone destruction terms are used to parameterize the heterogeneous chemistry as a function of the equivalent chlorine content prescribed for the actual year. All Taylor coefficients of this linearized scheme were determined using a two-dimensional chemistry scheme with
56 constituents, 175 chemical reactions, and 51 photoreactions (Cariolle and Brard, 1985). Photochemical production and loss rates of ozone rely on the main gas-phase reactions driving the NO\textsubscript{x}, HO\textsubscript{x}, ClO\textsubscript{x}, BrO\textsubscript{x} catalytic cycles. In this version, the gas-phase chemical rates were upgraded according to the recommendations of Sander et al. (2006). While the ozone mixing ratio is fully described across the atmospheric column, the linear ozone scheme was especially designed to resolve its evolution in the stratosphere for the sake of radiative transfer calculation. Therefore, some tropospheric chemical reactions are not taken into account in this scheme. The reader is referred to a manuscript by Eyring et al. (2013) for an extensive evaluation of the linear scheme versus TOMS satellite measurements and intercomparison with other CMIP5 models.

In CNRM-ESM1, the atmospheric CO\textsubscript{2} mixing ratio can be treated as a prognostic tracer. It responds interactively to natural CO\textsubscript{2} exchange from land and ocean every 30 min and 6 h, respectively, while anthropogenic carbon emissions are prescribed in this model version. The CO\textsubscript{2} mixing ratio can affect the physical climate by impacting the atmospheric radiative transfer computations and both terrestrial and marine carbon uptake. In the concentration-driven experiments presented here, the CO\textsubscript{2} mixing ratio is however prescribed to the global yearly average atmospheric concentrations according to the CMIP5 protocol.

2-3 The biogeochemical components

2-3-1 Land biogeochemical model

In CNRM-ESM1, the interactions between climate and vegetation are handled by the ISBA scheme embedded in the SURFEX surface model. The land biogeochemical module in ISBA represents land surface physics, plant physiology, carbon allocation and turnover, and carbon cycling through litter and soil (Calvet and Soussana, 2001; Calvet et al., 1998; Gibelin et al., 2006; 2008). The land cover is represented by 9 plant functional types (PFT, given in Figure 1) and 3 non-vegetated surface types that are determined spatially by the ECOCLIMAP physiographic database (Masson et al., 2013a).
ISBA uses a semi-mechanistic treatment of canopy photosynthesis and mesophyll conductance following the Jacobs et al. (1996) and Goudriaan et al. (1985) photosynthesis model. Mesophyll conductance in this framework corresponds to the rate of photosynthesis under light-saturated conditions (Jacobs et al., 1996). As such, this scheme does not explicitly account for Michaelis-Menten kinetics of the Rubisco enzyme found in Farquhar et al. (1980) and Collatz et al. (1992) models. ISBA includes a representation of the soil water stress. Key parameters of the photosynthesis model respond to the soil water stress, permitting the representation of drought-avoiding and drought-tolerant responses to drought. For low vegetation and trees, the response to drought is based on the meta-analyses of Calvet (2000) and Calvet et al. (2004) respectively.

The model simulates a ratio of intercellular CO\textsubscript{2} to atmospheric CO\textsubscript{2} that depends on leaf-to-air saturation deficit, leaf temperature, and soil moisture. Assimilation is calculated from this ratio, air CO\textsubscript{2} concentration, leaf temperature, and solar radiation considering plant photosynthetic pathways: C\textsubscript{3} or C\textsubscript{4} (Calvet et al., 1998; Gibelin et al., 2006). Stomatal conductance, which represents the vegetation control on gas transfer (here, CO\textsubscript{2} and water vapor) between the leaves and the atmosphere, is finally deduced from the assimilation rate. Leaf dark respiration is taken as a fraction of maximum CO\textsubscript{2} limited rate of assimilation. Standard Q\textsubscript{10} response functions determine the temperature dependencies of mesophyll conductance, CO\textsubscript{2} compensation point, maximum photosynthetic rate and hence photosynthesis and respiration.

ISBA simulates the evolution of 6 reservoirs of biomass including leaf, wood, and roots, and assumes the existence of metabolic/structural reservoirs of biomass (Gibelin et al., 2008). Vegetation biomass is simulated interactively based on the carbon assimilated by photosynthesis, and decreased by turnover and respiration. The autotrophic respiration combines the respiration from all these reservoirs except the woody reservoir that is supposed not to respire (Gibelin et al., 2008). In this model, the vegetation phenology results directly from the carbon balance of the leaves. Therefore, phenology is completely driven by photosynthesis and no growing degree-day model is used. A key advantage of this approach is that most of the soil and atmospheric drivers (the abiotic drivers) of phenology are accounted for without any additional parameters (Szczypta et al., 2014). Leaf area index (LAI) is determined
from the leaf biomass and the specific leaf area index, which varies as a function of leaf nitrogen concentration and plant functional type (Gibelín et al., 2006). ISBA uses an implicit nitrogen limitation parameterization which is based on the meta-analysis of leaf nitrogen measurement under CO₂ enrichment condition (Yin et al., 2002). This simple implicit nitrogen limitation is based on the nitrogen dilution hypothesis, which assumes that internal nitrogen content of a plant decrease under rising CO₂ due to the accumulation of non-structural carbohydrates. It results that nitrogen dilution occurs as soon as the increase in total biomass of a plant under rising CO₂ relative to growth under ambient CO₂ is greater than the corresponding increase in total nitrogen. In current version of ISBA, a linear decrease between specific leaf area index and nitrogen to carbon ratio in leaves is used to mimic this mechanism (Calvet et al., 2008), and hence to limit the net assimilation of atmospheric CO₂.

The soil organic matter and litter module in ISBA follows the soil carbon part of the CENTURY model (Parton et al., 1988). Four pools of litter are represented. They are differentiated by their location above- or below-ground and their content of lignin. The litter pools are supplied by the fluxes of dead biomass from each biomass reservoir (turnover) as described in Gibelin et al. (2008). The 3 soil organic matter reservoirs (active, slow and passive) are characterized by their resistance to decomposition with turnover times spanning from a few months for the active pool to 240 years for the passive pool. Heterotrophic respiration and hence the flux of CO₂ released to the atmosphere is the sum of respiration from the litter and soil organic matter reservoirs. The rate of decomposition of organic matter is determined essentially by soil moisture and temperature using a Q₁₀ dependence following the formulation of Krinner et al. (2005). The rate of decomposition (by respiration) depends also on the lignin fraction and the soil texture following Parton et al. (1988).

Changes in the carbon balance of the vegetation affect the energy and water balance, and hence the climate, through changes in stomatal conductance and LAI. Through its control on leaf transpiration, stomatal conductance affects latent heat flux and the surface energy balance. LAI on the other hand affects evapotranspiration because it is used to scale leaf-level to canopy level transpiration and evaporation from the interception reservoir (water intercepted by leaves).
In CNRM-ESM1, except for crops, changes in LAI don’t affect the albedo of the land-surface, as it is the case in some other models. As mentioned earlier, albedo is derived from satellite observations corrected in the presence of snow, but does not depend on the changes in LAI calculated by the model. This limits the biophysical feedback from vegetation change to the atmosphere.

### 2-3-2 Ocean biogeochemical model

The ocean biogeochemical model of CNRM-ESM1 is PISCES (Aumont and Bopp, 2006). This model simulates the biogeochemical cycles of oxygen, carbon and the main nutrients with 24 state variables. Macronutrients (i.e., nitrate and ammonium, phosphate, silicate) and micronutrient (i.e., iron) ensure a better representation of the phytoplankton dynamics, because these 5 nutrients contribute to the nutrient limitation process (Aumont et al., 2003). PISCES represents two size-classes of phytoplankton (i.e., nanophytoplankton and diatoms) Dependence of growth on temperature is parameterized according to Eppley et al. (1969). Growth rate is also limited by the external availability in nutrients using Michaelis-Menten relationships. Diatoms differ from nanophytoplankton by their need in silicon, by higher requirements in iron (Sunda and Huntsman, 1997) and by higher half-saturation constants because of their larger mean surface-to-volume aspect ratio. Zooplankton is represented by two size-classes: microzooplankton and mesozooplankton.

PISCES can be considered as a Monod model (Monod, 1942) since it does not represent the internal concentration of nutrients in the cells. The ratios between carbon, nitrate and phosphate are kept constant to the values proposed by Takahashi et al. (1985) in all living and nonliving pools of organic matter. However, internal concentrations of iron in both phytoplankton and of silicon in diatoms are prognostically simulated. They depend on the external concentration of these nutrients, on the potential limitation by the other nutrients and on light availability.

Phytoplankton chlorophyll concentration is prognostically simulated following Geider et al. (1998). PISCES simulates semilabile dissolved organic matter, small and big sinking particles, which differ by their sinking speeds (i.e., 3 m d$^{-1}$ and 50 to 200 m d$^{-1}$).
only the internal concentrations of iron, silicon and calcite inside the sinking particles are prognostically simulated. In addition to exchange with organic carbon, dissolved inorganic carbon is also altered by the production and dissolution of calcite. Carbon chemistry in seawater is computed from the distribution of dissolved inorganic carbon and alkalinity. Calcite is prognostically simulated following Maier-Reimer, (1993) and Moore et al. (2002). Alkalinity includes the contribution of carbonate, bicarbonate, borate and water ions. Oxygen is prognostically simulated using two different oxygen-to-carbon ratios, one accounting when ammonium is converted to or mineralized from organic matter, the other when oxygen is consumed during nitrification. For carbon and oxygen pools, air-sea exchange follows the Wanninkhof (1992) formulation. Importantly, to ensure conservation of nitrogen in the ocean, annual total nitrogen fixation should balance denitrification following Lipschultz et al. (1990), Middelburg et al. (1996) and Soetaert et al. (2000). For the other macronutrients, alkalinity and organic carbon, the conservation is ensured by tuning the sedimental loss to the total external input from rivers and dust. Therefore, carbon and nitrogen cycles are decoupled to a certain degree.

The boundary conditions account for nutrient supply from three different sources: atmospheric dust deposition for iron and silicon (Jickells and Spokes, 2001; Moore et al., 2004; Tegen and Fung, 1995), rivers for carbon (Ludwig et al., 1996) and sediment mobilization for sedimentary iron (de Baar and de Jong, 2001; Johnson et al., 1999). In CNRM-ESM1, riverine input of carbon has been revised from Ludwig et al. (1996) in accounting for the interannual variability of runoff estimated with an offline SURFEX simulation over the 1948-2010 period using the global atmospheric forcing from Princeton University (PGF, Sheffield et al., 2006).

In CNRM-ESM1, the marine biophysical feedback is induced by changes in the penetration of downward irradiance in response to marine biota chlorophyll concentration. This feedback mimics the fact that light absorption in the ocean indeed depends on particle concentration and is spectrally selective (Morel, 1988). The implementation of this mechanism is fully described in Lengaigne et al. (2006) and Lengaigne et al. (2009) for an ocean forced configuration and Mignot et al. (2013) for a current ocean coupled configuration. It is derived from an accurate 61 spectral bands
formulation proposed by Morel (1988) using three large wavebands: blue (400–500 nm), green (500–600 nm) and red (600–700 nm). These three bands correspond to the spectral domain of maximum absorption for chlorophyll. The chlorophyll-dependent attenuation coefficients depend on the three-dimensional chlorophyll field predicted by PISCES. They are computed at each time step from a power-law relationship fitting to the coefficients computed from the full spectral model of Morel et al. (1988). This biophysical feedback represents a major evolution from the ocean component used in Voldoire et al. (2013) and Séférian et al. (2013).

3- Experimental set-up

3-1 Spin-up strategy

The CMIP5 specification requires each model to reach its equilibrium state before kicking off formal simulations, especially for long-term control experiments. To obtain the initial conditions for CNRM-ESM1 preindustrial steady state at year 1850, we first initialize the various physical and biogeochemical components of the model as described below and perform a 400-year-long spin-up simulation using CNRM-ESM1 with all 1850 external forcings (Taylor et al., 2009).

Initialization of the physical components of CNRM-ESM1 relies on previous model outputs from CNRM-CM5.1. This latter model was first initialized from World Ocean Atlas 2005 observations for salinity and temperature (Antonov et al., 2006; Locarnini et al., 2006) and spun-up for 200 years. The 801th year of the centennial long CMIP5 preindustrial run from CNRM-CM5.1 was employed as initial condition for CNRM-ESM1 preindustrial state.

Marine biogeochemical reservoirs were initialized from fields of a previous preindustrial simulation of CNRM-CM5.1 coupled to PISCES. In this previous simulation, PISCES state variables were initialized from World Ocean Atlas 1993 observations for nitrate, phosphate, silicate, and oxygen (Levitus et al., 1993) and the Global Ocean Data Analysis Project (Key et al., 2004) for alkalinity and preindustrial
dissolved inorganic carbon (DIC). From this initialization, this intermediate version of
the ESM was integrated online for 1100 years.

Land biogeochemical reservoirs were initialized from zero and spun-up using an
acceleration approach for soil carbon and wood during the first century of the spin-up
simulation. This approach consists in updating the wood growth, the litter and soil
biogeochemistry modules several times per time step with constant incoming carbon
fluxes and physical conditions allowing to fill up much faster the various reservoirs of
carbon. As a result of this approach, soil carbon and wood reservoirs were
respectively spun-up for 21800 and 1200 years.

Finally, both physical and carbon cycle components of CNRM-ESM1 benefit from an
physical adjustment under 1850 preindustrial control conditions for 400 years. Section
4.1 describes the residual drifts of the model at quasi-equilibrium state.

3-2 CMIP5 preindustrial control and historical simulations

Following CMIP5 specifications (Taylor et al., 2009), CNRM-ESM1 has performed
several CMIP5 long-term core experiments and part of the tier-1 experiments.

The preindustrial control simulation, piControl, is integrated for 250 years using
constant external forcing prescribed at 1850 conditions and starting from the last year
of the on-line adjustment simulation. That is, atmospheric concentrations of
greenhouse gases are set to 284.7 ppmv, 790.9 ppbv, and 275.4 ppbv for CO₂, CH₄
and N₂O, respectively. Those of CFC-11, CFC-12 are set to zero. Influence of natural
aerosols is prescribed using the optical depths of five types of tropospheric aerosols
(black carbon, sea salt, sulfate, dust and particle organic matter) from an LMDZ-INCA simulation forced with CMIP5 prescribed emissions (Szopa et al., 2013).

Stratospheric volcanic aerosols are prescribed similarly but using a long-term average
climatology from a last millennium simulation performed with the NCAR Community
Climate System Model (Ammann et al., 2007).

The 20th century experiment, historical, is performed from year 1850 to 2005. This
simulation starts from the CNRM-ESM1 states of the last year of the on-line
adjustment simulation. The modern evolution of the external forcings of both atmospheric greenhouse gases and incoming solar irradiance follows the recommended yearly average observations (Taylor et al., 2009). The monthly temporal and spatial variability of the five tropospheric aerosols also rely on a LMDZ-INCA simulation (Szopa et al., 2013) while those of stratospheric sulphate aerosol concentrations from explosive volcanoes are derived from a 20th century reconstruction of the NCAR Community Climate System Model (Ammann et al., 2007).

Note there is no land-cover change related to anthropogenic land use in the above-mentioned simulations. The fraction of vegetal cover is set to the present-day state using the in-house ECOCLIMAP database (Masson et al., 2013a). Therefore, changes in physical and biogeochemical properties of the vegetation due to actual land-cover changes are excluded by design.

4- Results

4-1 Model equilibrium in the preindustrial control simulation

To illustrate the stability of CNRM-ESM1 at the end of the spin-up simulation, we show the global average values of a few variables during the 250 years of the piControl simulation (Figure 2) and their drifts (Table 1).

In terms of energy balance, the global mean top-of-atmosphere (TOA) net radiative balance is about 3.57±0.23 W m\(^{-2}\), while the global mean net surface radiation flux (NSF) is 0.87±0.24 W m\(^{-2}\) (Figure 2a). The imbalance in the energy budget between the surface and TOA (about 2.7 Wm-2) is predominantly due to the non-conservation of energy of the spectral atmospheric model and, to a lesser extent, its coupling with the ocean model. Taking apart this non-conservation offset in TOA net radiation flux, there is no discernible deviation between year-to-year fluctuation between the TOA and NSF net radiation fluxes.

In terms of global-scale climate indices, the global mean surface temperature (T\(_{2m}\)) and sea surface temperature (SST) over the piControl period are 12.52±0.15 and
17.76±0.1°C respectively (Figure 2b). They both display almost no drift over the duration of the piControl simulation (Table 1). We use soil wetness index (SWI) and sea surface salinity (SSS) to evaluate the stability of the simulated water cycle (Figure 2c). These both have almost no drift (Table 1), confirming that the water cycle is closed. Also, there is no drift in both Northern Hemisphere and Southern Hemisphere sea-ice volume (NIV, SIV, respectively) for which long-term means are respectively 20.88 and 6.25 $10^3$ km$^3$ (Figure 2d).

Regarding the simulated global carbon cycle, Figure 2e shows that the natural carbon cycle is stable over the piControl simulation with terrestrial and oceanic carbon fluxes of $0.75±0.57$ and $-0.94±0.13$ Pg C y$^{-1}$, respectively. Both terrestrial and oceanic components of the simulated carbon cycle exhibit drifts smaller than $10^3$ Pg C y$^{-1}$ demonstrating that soil and deep ocean carbon stocks have reached a steady state. Deviation from zero in the terrestrial carbon flux is essentially explained by missing perturbations or processes in ISBA such as fire-induced CO$_2$ emissions or riverine-induced carbon transport from land to oceans (Battin et al., 2009; Regnier et al., 2013). Natural ocean carbon outgasing falls within the upper range of ocean inverse estimates (Jacobson et al., 2007; Mikaloff Fletcher et al., 2007).

### 4-2 Late 20th century climatology

#### 4-2-1 Land physical drivers

In the following, we focus on the physical drivers of the global carbon cycle. From a land perspective, surface temperature ($T_{2m}$), precipitation (PR) and photosynthetically active radiation (PAR) are the prominent factors controlling the rate of photosynthetic activity as well as the rate of autotrophic and heterotrophic respiration, and hence the net land-air exchange of carbon.

Compared to the CRUTV4 dataset (Harris et al., 2013) over the period 1986-2005, CNRM-ESM1 displays a global annual averaged bias of -3°C in $T_{2m}$ over continents. In Northern Hemisphere winter (DJFM, Figure 3a) simulated $T_{2m}$ is generally lower than the observations except for some regions (e.g., North East Siberia, South of
Australia and part of Argentina). The mean bias over continents in boreal winter is about -4°C and can reach up to -6°C over mountain regions. Figure 3b shows that simulated summer (i.e., JJAS) $T_{2m}$ is also generally colder than the observations (-0.8°C in global average) over a large fraction of the continents. Only the most Northern domains of the Northern Hemisphere display a warm bias that can reach up to 3°C in the North of Canada. The geographical structure of the $T_{2m}$ bias compares well with those detailed in Voldoire et al. (2013). Such an agreement in the bias structure for $T_{2m}$ was expected since both models rely on the same physical parameterizations for both the atmosphere and land surface physics. Small deviations between CNRM-CM5 and CNRM-ESM1 mean state can be essentially attributed to the land carbon cycle, which appears to amplify the global average annual cold bias of 0.8°C (with seasonal differences between CNRM-ESM1 and CNRM-CM5 of -0.7°C and -1°C in boreal winter and summer, respectively). This cooling is due to the enhanced evapotranspiration by the interactive terrestrial biosphere compared to the fixed one in CNRM-CM5.

Figure 4 shows the regional structure of the precipitation (PR) bias of CNRM-ESM1 with respect to the Global Precipitation Climatology Project (GPCP) observations (Adler et al., 2003). Over continents, CNRM-ESM1 slightly underestimates the amount of the seasonal PR except over Asia, the Western coast of America and Australia. The major regional bias in seasonal PR is found over Amazonia, where PR is underestimated by 2 and 5 mm day$^{-1}$ in boreal summer and winter, respectively. Similar to state-of-the-art Earth system models, CNRM-ESM1 displays an excess of precipitation over the oceans. This excess is especially strong in the Southern part of the tropical oceans and is associated with the overestimated seasonal latitudinal migration of the ITCZ. The land biosphere biophysical coupling induces small but noticeable changes in the global hydrological cycle between CNRM-CM5 and CNRM-ESM1. Although weak, changes induced by the ISBA biophysical coupling slightly affect the representation of the seasonal cycle in PR over the vegetated regions (Figure S1). These lead to improve the simulated PR in CNRM-ESM1 compared to CNRM-CM5 over some vegetated regions during the growing season (spring-summer). Between 30°N and 60°N, the average error in simulated PR compared to GPCP is reduced by 0.12 mm day$^{-1}$ with CNRM-ESM1 compared to that of CNRM-CM5. Over the tropics (30°S-30°N), simulated PR is also improved in...
CNRM-ESM1 but to a lesser extent with a reduction of the average error by 0.06 mm day$^{-1}$ with respect to GPCP. Although PR have been improved over some regions, their geographical pattern has been degraded in CNRM-ESM1 compared to CNRM-CM5, especially during the winter.

Compared to SRB satellite-derived observations (Pinker and Laszlo, 1992), CNRM-ESM1 overestimates the photosynthetically active radiation (PAR) globally (Figure 5). Major biases are found over continents except for some regions in the Tropics. The magnitude of the seasonal biases is weaker in Northern Hemisphere winter than in summer when regional biases reach up to 20-30 W m$^{-2}$ over the Western border of the continents. Regions where PAR is underestimated match reasonably well with those showing too intense precipitations compared to the GPCP dataset (Figure 4). The general overestimation in PAR is due to the substantial underestimation in low cloud cover in CNRM-ESM1 consistent with CNRM-CM5. Biases in PAR are also found over ocean upwelling system and are linked with an underestimated fraction of stratocumulus.

4-2-2 Ocean physical drivers

From an oceanic perspective, temperature is as important as over land surface because it sets the marine biota’s growth rate, playing a large role in the biological-mediated processes (e.g., export, soft tissue pump). In addition, both temperature (T) and salinity (S) control the solubility of CO$_2$ into seawater and the chemical-mediated air-sea exchanges of carbon. The mixed-layer depth (MLD) and the sea-ice cover (SIC) are also critical drivers of the ocean carbon cycle as they both contribute to the nutrient-to-light limitation in the high latitude oceans (Sarmiento and Gruber, 2006).

In the following, we assess the representation of these drivers.

Compared to WOA2013 data products (Levitus et al., 2013), CNRM-ESM1 realistically simulates both the mean annual sea surface temperature and sea surface salinity, both in terms of amplitude and spatial distribution, as shown in Figure 6ab. Moderate positive biases in sea surface temperature and sea surface salinity are found in the Southern Ocean and in the Eastern boundary upwelling systems. Strong biases
in sea surface salinity are found in the Labrador and Arctic Seas. While most of these biases are related to an overestimated atmospheric surface heating, biases in the Labrador Sea and in the Arctic are essentially due to erroneous representation of the mixed-layer depth and the Arctic sea-ice cover. These points will be further detailed below.

At depth, the vertical structures in simulated $T$ and $S$ display biases from those estimated from WOA2013 observations. $T$ is underestimated by ~2°C within the first 1000 m of the Atlantic and Pacific oceans, except in the deep water formation zone (North Atlantic, North Pacific and Southern Ocean), where the model displays positive biases. The largest deviation in vertical structure of simulated $S$ from that estimated from WOA2013 are found in deep water formation zones where haline biases of about ~1 psu tend to compensate for the warm bias in $T$, enabling deep convection of water masses. Because of this compensating mechanism, the flow of North Atlantic deep waters (NADW) fueling the Atlantic meridional overturning circulation is about 26.1±2 Sv at 26.5°N in CNRM-ESM1 averaged over the 1850-2005 period. This value is stronger than the observations-derived estimate of 18 ± 5 Sv (Talley et al., 2003) or the observations from RAPID-MOCHA monitoring array over 2004-2007 estimating the flow at about 18.5±4.9 Sv (Johns et al., 2011). In the Southern Ocean, the flow of Antarctic bottom water (AABW) is about 11.6±1 Sv in CNRM-ESM1 averaged over the 1850-2005 period. This flow of AABW is in agreement with the deep flow of waters compared to the observed estimate of 10 ± 2 Sv (Orsi et al., 1999). Consequently, the flow of deep water masses in CNRM-ESM1 has been improved in regards that of CNRM-CM5 which ranges between 3.4 and 6.2 Sv over the same period (Séférian et al., 2013; Voldoire et al., 2013). As detailed in several intercomparison studies (de Lavergne et al., 2014; Heuzé et al., 2013; Sallée et al., 2013; Séférian et al., 2013), CNRM-CM5 substantially underestimated the flow of AABW leading to an erroneous distribution of hydrodynamical and biogeochemical fields at depth. Here, although stronger than the observation-based estimates, the flow of NADW and AABW improves the deep ocean ventilation as well as the distribution of tracers at depth (section 4-2-5).

As mentioned above, an accurate representation of spatial and temporal MLD is essential for numerous ocean biogeochemical processes. For example, winter mixing
entrains carbon- and nutrient-rich deep waters to the surface, which play an important role in the transfer of CO$_2$ across the sea-to-air interface. In summer, MLD contributes to the nutrient-to-light limitation of the phytoplankton growth in high-latitude oceans. The maximum and minimum mixed-layer depth (hereafter $MLD_{\text{max}}$ and $MLD_{\text{min}}$) are respectively used as a proxy of the winter and summer MLD since mixing occurs randomly during seasons in response to numerous environmental factors (wind, stratification, local instability etc…) that present a large spatiotemporal variability.

Figure 7 presents composites of yearly $MLD_{\text{max}}$ and $MLD_{\text{min}}$ as simulated by CNRM-ESM1 in averaged over the 1986-2005 period and as derived from observations (Sallée et al., 2010). Figure 7ab shows that CNRM-ESM1 reproduces the main regional pattern of $MLD_{\text{max}}$ compared to the observation-derived estimates. However, the model tends to simulate too large and too deep mixing sites in the North Atlantic, the North Pacific and the Southern Ocean. In the North Atlantic, the larger than observed mixed volume of surface dense waters (combination of surface area and depth of the mixing zone) is at the origin of the strong flow of NADW simulated in CNRM-ESM1. In the Southern Ocean, although open ocean polynyas were observed from space in the past decades (Cavaliere et al., 1996; Comiso, 1999), their locations are erroneous in CNRM-ESM1 similarly to several other CMIP5 Earth system models (de Lavergne et al., 2014). CNRM-ESM1 simulates open ocean polynyas in the Indian basin and close to the Ross Sea but not in the Atlantic basin as observed from space between 1974 and 1976.

Compared to the observation-derived estimates, CNRM-ESM1 captures the main regional pattern of $MLD_{\text{min}}$ but the model fails at reproducing the deepest values of mixing in the Southern Ocean and the Tropics. This bias might be linked to the current parameterization of the ocean mixing employed in CNRM-ESM1 because previous model version using this parameterization also exhibited similar patterns of errors as detailed in Séférian et al. (2013) and Voldoire et al. (2013).

Similarly to the MLD, SIC is an important driver of the ocean carbon cycle. It constitutes a physical barrier for exchange of CO$_2$ between the ocean and the atmosphere leading to an accumulation of carbon-rich waters below the sea ice (Takahashi, 2009). It also plays a large role in the seasonal timing of algal blooms (Wassmann et al., 2010). Compared to the MLD, seasonal variations of sea ice are...
strongly and directly responsive to the seasonal fluctuations of atmospheric forcing. Therefore, it matters that the model is able to accurately capture the spatial distribution and timing of annual minimal and maximal sea ice covers in both Hemispheres. For this purpose, we evaluate differences between composites of simulated and observed SIC (Cavalieri et al., 1996) for September and March over the 1986-2005 period (Figure 8). In the Arctic Ocean, CNRM-ESM1 underestimates SIC in the Beaufort, Chukchi and East Siberian seas in September, while too much sea ice tends to be present in the Barents Sea (Figure 8a). In March, SIC is largely overestimated in the Barents and Nordic seas, as well as in the Bering and Okhotsk seas on the Pacific Ocean side, showing that the simulated winter sea ice edge spreads too far South and East in these regions (represented with iso-15% in Figure 8c). On the contrary, SIC is slightly underestimated in the Labrador Sea and Baffin Bay in March (Figure 8c). This too far North ice edge comes along with positive SST biases in this region (Figure 6a), and explains why the simulated deep convection zone is too large and shifted northward in CNRM-ESM1 as shown in Figure 7.

In the Antarctic Ocean, Figure 8b shows that the spatial structures of SIC biases mirror somehow the model-data mismatch in MLD as shown in Figure 7b. That is, in austral winter, CNRM-ESM1 underestimates SIC where erroneous open ocean deep convection zones are located, namely offshore Wilkes Land in the Indian Ocean sector (Figure 8b). Conversely, too much sea ice is simulated in the Atlantic Ocean sector. As in CNRM-CM5.1, simulated summer Antarctic SIC is strongly underestimated, with very little sea ice surviving summer melt in the Weddell and Ross Seas (Figure 8d).

### 4.2.3 Comparison with previous model version

In the following, we compare the skill of CNRM-ESM1 to the closest version of CNRM-CM5 climate model, called CNRM-CM5.2. Figure 9 summarizes skill-assessment metrics for CNRM-ESM1 and CNRM-CM5.2 in terms of major physical drivers of the global carbon cycle (field maps and patterns of errors are presented in Figures S2 to S7).
The Taylor diagram for land surface physical drivers clearly demonstrates that CNRM-ESM1 and CNRM-CM5 display comparable skills except for PR (Figure 9a). Most of the differences in skills are indeed not significant at a 95% confidence level; models differ solely in terms of PR for which CNRM-ESM1 produces slightly weaker correlation coefficients.

Over the ocean, Figure 9b shows further differences between both models. The weakest difference in skill concerns SST for which both models display good agreement withWOA2013. Regarding the MLD, CNRM-ESM1 displays a slightly better agreement than CNRM-CM5 with observation-derived MLD (Sallée et al., 2010) in terms of correlation but strongly underestimates the spatial variations of this field. Major differences are noticeable for SSS. CNRM-ESM1’s skill is clearly lower than that of CNRM-CM5. To investigate this difference, we have computed skill of PR over the ocean since this latter contributes to the spatiotemporal distribution of the SSS concomitantly to the runoff and the sea-ice seasonal cycle. Skill in PR over the ocean is similar for both models (blue diamonds on Figure 9b). A similar finding is noticed for simulated runoff (not shown). Therefore the difference in simulated SSS between the two models can be attributed to the revised water conservation interface and erroneous distribution of sea-ice cover. In addition, changes in coupling frequency (i.e. 24h to 6h) might be at the origin of differences in skills between the two models since it impacts sea-ice cover (Figure 10).

From the small differences in skill between the two models, we can assume that the inclusion of the global carbon cycle and the biophysical coupling have not noticeably altered the simulated mean-state climate in CNRM-ESM1 compared to that of CNRM-CM5.

### 4-2-4 Terrestrial carbon cycle

Now that the physical drivers of the global carbon cycle have been evaluated, we assess the ability of CNRM-ESM1 to replicate available modern observations of the terrestrial carbon cycle. We focus on gross primary productivity (GPP), vegetation autotrophic respiration (Ra) and soil organic carbon content (cSoil) that control the
net natural fluxes of CO$_2$ on land. Simulated budget of vegetation biomass and total ecosystem respiration (TER, sum of autotrophic and heterotrophic respirations) are evaluated against available published estimates. While we can assess the capability of CNRM-ESM1 to fix and emit carbon on land, it is important to note that the CO$_2$ fluxes due to land-use changes are not taken into account in this analysis.

To evaluate CNRM-ESM1 GPP, we rely on two streams of data, namely the FluxNet-MTE (Jung et al., 2011) and the MOD17 satellite-derived observations (Running et al., 2004). Figure 1 shows that the annual mean GPP as simulated by CNRM-ESM1 is slightly too strong compared to the observed estimates. The largest model-data mismatch is found in the Tropics between 10°N and 20°S, where CNRM-ESM1 simulates erroneous patterns of high GPP. Over Amazonia, CNRM-ESM1 fails to reproduce the zonal gradient of GPP. Regions of high GPP are in association with overestimated PAR and, to a lesser extent, underestimated PR in summer (Figure 4 and Figure 5, respectively; see also Figure S8). The geographical structure of simulated GPP fits the observed over the African and Asian rain forest but its amplitude is overestimated by about 3 gC m$^{-2}$ day$^{-1}$. This regional overestimation impacts both the zonal and global GPP budget, which are larger than the published estimates except >60°N (Table 2). This stronger-than-observed GPP constitutes a systematic bias of the current version of ISBA. In an offline simulation, (Carrer et al., 2013b) show that ISBA forced with PGF overestimates global GPP by 60 Pg C y$^{-1}$. Regional biases in GPP are partly compensated by overestimated Ra (Figure 12). Simulated Ra agrees reasonably well with satellite-derived estimates except in the Tropics. This bias compensation between GPP and Ra is analyzed in detail by Joetzjer et al. (2015). In this study, the authors demonstrate that the current parameterizations of Ra and water stress in ISBA are not adequate for tropical broadleaf trees (Figure 1). Considering that these results were deduced from offline simulations forced with in situ observations, we can assume here that biases in GPP and Ra result from a combination of erroneous ecophysiological parameterizations and biases in physical drivers in CNRM-ESM1.

Despite these biases, the global partitioning between vegetation biomass and soil carbon is realistic with 596.7 and 2105 Pg C compared to the observed estimates of 560±94 (DeFries et al., 1999) and 1750±250 Pg C (Houghton, 2007), respectively.
Furthermore, the geographical structure of cSoil agrees well with Harmonized World Soil Database (JRC, 2012) except in the Northern Hemisphere (Figure 1). Although several processes are missing in ISBA to accurately simulate high-latitude carbon stock (e.g., permafrost dynamics, bacterial degradation of the litter, fire-induced turnover etc…), a part of cSoil underestimation can be attributed to the summer warm bias in near-surface temperature (Figure 3b). This latter tends to enhance heterotrophic respiration of the soil, reducing the soil organic matter (R>0.6, Figure S2).

Table 2 shows that CNRM-ESM1 overestimates globally terrestrial ecosystem respiration (TER) when compared to the up-scaled measurements of FluxNet-MTE. In the Tropics, simulated TER fluxes are 32% higher than the FluxNet-MTE estimates. As mentioned above, this bias is essentially due to an unrealistic Ra, which amounts to 72% of TER over the sector in the model. Table 2 shows that the simulated TER is 126.9 Pg C y\(^{-1}\), larger than estimates published by Jung et al. (2011) of 96.4±6.0 Pg C y\(^{-1}\). Nevertheless, the simulated net land carbon sink (LCS), which can be estimated by subtracting TER from GPP, is 2.19 Pg C y\(^{-1}\) in average over the 1986-2005 period and remains within the range of values estimated from various observation-based methods (IPCC, 2007; 2013; Le Quéré et al., 2014).

4-2-5 Ocean carbon cycle

Compared to the terrestrial carbon cycle, the ocean carbon cycle has already been implemented in previous versions of CNRM-CM5 (Séférian et al., 2013). The modeled marine biogeochemistry components have already benefited from detailed evaluation against modern observations (Frölicher et al., 2014; Séférian et al., 2013), analyses of future projections (Laufkötter et al., 2015) and sensitivity benchmarking (Schwinger et al., 2014). The major difference between CNRM-ESM1 and previous versions of CNRM-CM5 including a marine biogeochemistry module lies in the representation of ocean tracers in the deep ocean. Figure 14 shows that the representation of oxygen, phosphate, nitrate and silicate fields was improved in CNRM-ESM1 at depth, except around 1000 m where the strong flow of NADW tends to alter the distribution of tracers. Below 1500 m, the tracer distribution is in
reasonable agreement with the observations with correlation coefficients ~0.8. This represents a noticeable improvement with respect to the CNRM-CM5 oxygen distribution (R~0.4). In addition to nutrients, the vertical distribution of carbon-related fields like dissolved inorganic carbon has been substantially improved in CNRM-ESM1 compared to CNRM-CM5 (Figure S9), showing a much better agreement with GLODAP observations (Key et al., 2004; Sabine et al., 2004).

In terms of carbon cycling into the ocean, Figure 1 shows the simulated mean annual sea–air CO$_2$ fluxes over 1986-2005 together with observation-based estimates by Takahashi et al. (2010) using 2000 as a single reference year. While the model broadly agrees with the observations in terms of spatial variation for regions of carbon sink (i.e., North Atlantic, North Pacific and between 50°S-40°S), it displays a too strong source of carbon to the atmosphere in the equatorial Pacific and in the Southern Ocean. In the equatorial Pacific, the model-data mismatch is likely related to the decision of Takahashi et al. (2010) to exclude observations from El Niño years from their analysis. Since surface ocean pCO$_2$ of the Eastern tropical Pacific during El Niño events tends to be lower than the long-term mean, the Lamont-Doherty Earth Observatory (LDEO) climatology tends to underestimate outgasing of CO$_2$ in the equatorial Pacific over the 1986-2005 period. This hypothesis is validated when comparing model results against recent data products derived from statistical Monte-Carlo Markov Chain or Neural Network gapfilling methods (Landschützer et al., 2014; Majkut et al., 2014. Figure S10). In the Southern Ocean, the model-data mismatch is especially pronounced south of 60°S. This bias in fgCO$_2$ is associated with overestimated mixing (Figure 7), which tends to bring too much deep carbon-rich water masses to the surface, enhancing the outgasing of CO$_2$. CNRM-ESM1 results display similar discrepancy when compared to other observation-derived data products which agree in a moderate CO$_2$ outgasing south of 60°S (Figure S10). That said, simulated patterns of sea-to-air carbon fluxes in this domain qualitatively agree with the data in showing a combination of source and sink regions.

The storage of anthropogenic CO$_2$ by the oceans (CO$_2^{ANTH}$, Figure 16) provides a complementary view of the ocean carbon fluxes by revealing the chronology of the ocean CO$_2$ uptake from preindustrial to modern state. Here, we have chosen to stick to the available observation-derived estimates, GLODAP, which use year 1994 as a
single reference year (Key et al., 2004; Sabine et al., 2004). In order to account for the interannual variability of the simulated fields, we chose to analyze yearly average results from CNRM-ESM1 over 1990-2005 (Figure 1). Besides, computation of $\text{CO}_2^{\text{ANTH}}$ is not straightforward since natural and anthropogenic pools of carbon are not treated separately in PISCES. We approximate consequently $\text{CO}_2^{\text{ANTH}}$ from the difference between modern and preindustrial stocks of dissolved inorganic carbon. Negative values were set to zero in the computation since they are essentially generated from differences in simulated interannual variability. Ideally, this computation would have required a historical simulation with constant preindustrial atmospheric $\text{CO}_2$ for the sea-to-air $\text{CO}_2$ fluxes. Figure 1 shows that the maximum $\text{CO}_2^{\text{ANTH}}$ is concentrated in the North Atlantic region. This feature is linked to the large-scale circulation in the surface layer of the ocean, which converges in the North Atlantic, before being exported to depth with the flow of NADW (Pérez et al., 2013). The Southern Ocean also stores a large fraction of $\text{CO}_2^{\text{ANTH}}$ in association with the subduction of modal and intermediate water masses (Sallée et al., 2012). Compared to this global view, CNRM-ESM1 displays features that are broadly consistent with the $\text{CO}_2^{\text{ANTH}}$ estimates. However, the stronger flow of NADW and AABW leads to a depletion of the stock of $\text{CO}_2^{\text{ANTH}}$ between 0 and 1200 m (Figure 1c). This mechanism leads to an increase in the stock of $\text{CO}_2^{\text{ANTH}}$ at depth. Over the 1850–1994 period, the model takes up a total of 100.8 Pg C, which is in agreement with the observations that suggest a net uptake of 106 ± 17 Pg C over the same period (Khatiwala et al., 2013; Sabine et al., 2004).

4-2-6 Ecosystem dynamics

In this section, we assess the performance of CNRM-ESM1 in terms of two ecosystem dynamics parameters, namely the peak leaf area index ($\text{LAI}_{\text{max}}$) and the ocean surface chlorophyll (Chl). Both parameters are monitored continuously from space since the 1980s and the 1990s, respectively, providing a suitable set of indirect observations to assess the simplified ecosystem representation embedded in Earth system models.
Regarding LAI$_{\text{max}}$, Figure 17 shows that the model agrees well with satellite-derived observations (Zhu et al., 2013) except over Africa and Asia with overestimated values. As such, this ecosystem parameter behaves similarly to GPP and Ra, responding to biases in PR and PAR. In the Northern mid-latitudes, LAI$_{\text{max}}$ is slightly overestimated compared to the satellite-derived observations but remains in the low range of values simulated by other CMIP5 Earth system models evaluated in Anav et al. (2013b). Using an offline simulation forced with atmospheric reanalyzes (Szczypta et al., 2014) shows similar biases in LAI over Northern Europe as those noticed in CNRM-ESM1. It is thus likely that missing processes like forest and crop management or fire-induced disturbance might induce an overestimated LAI$_{\text{max}}$.

Regarding ocean Chl, Figure 18 shows that CNRM-ESM1 displays a reasonable agreement with satellite-derived observations (O'Reilly et al., 1998). Although regional patterns of Chl concentrations were improved compared to that of CNRM-CM5 (Séférian et al., 2013), major model discrepancies are found in oligotrophic gyres and equatorial upwellings. Biases are more pronounced in the Southern Hemisphere where the model fails to produce very low Chl in the Southern Pacific gyres. CNRM-ESM1 also fails at capturing Western border high Chl concentrations in relation with the equatorial upwelling. Underestimated Chl concentrations in upwelling systems are essentially due to biases in surface wind forcing but also to the coarse horizontal and vertical resolution of the ocean model. This model limitation partly explains why Chl concentrations are underestimated in high-latitude oceans. In these domains, high coastal concentrations are captured from satellite sensors but cannot be resolved by the model due to its coarse resolution.

### 4.3 Recent evolution of the Climate system

In the present section, we analyze the transient response of various climate indices to the recent climate forcing from 1901 to 2005. We focus on the near-surface temperature ($T_{2\text{m}}$), the September Arctic sea-ice extent (SIE), the 0-2000 m ocean heat content (OHC) as well as the land and ocean carbon sinks (LCS, OCS, respectively). Over this period, these climate indices are analyzed with their nominal values except for $T_{2\text{m}}$ and OHC that are represented with respect to the 1961-1990 and the 1955-
2005 periods, respectively. Figure 19 illustrates how these various climate indices evolve from 1901 to 2005 and Table 3 summarizes their mean-state, interannual variability (IAV) and decadal trends over the 1986-2005 period.

Figure 19 shows that the transient response of $T_{2m}$ agrees reasonably well with modern observations (Morice et al., 2012). At the end of the last decades of the historical simulation (i.e. 1986-2005), CNRM-ESM1 overestimates the $T_{2m}$ increase, a discrepancy widely shared by other CMIP5 Earth system models (Huber and Knutti, 2014; Kosaka and Xie, 2013; Meehl et al., 2011; Watanabe et al., 2013). The amplitude of the simulated recent IAV is in line with the observations (Table 3). In particular, the model simulates strong cooling followed by stronger warming after the 1991 mount Pinatubo eruption. Contrasting with temperature, the simulated SIE poorly agrees with observation-based estimates (Cavalieri et al., 1996; Comiso, 1999; Rayner et al., 2003). Indeed, CNRM-ESM1 underestimates the mean-state SIE by about 2 $10^6$ km$^2$ and overestimates not only the IAV but also the decadal decrease in extent (Table 3). Therefore, in terms of Arctic sea ice, the skill of CNRM-ESM1 is similar to CNRM-CM5 as detailed in Massonnet et al. (2012). A better agreement is found for OHC for which CNRM-ESM1 results agree with observation-based estimates in term of mean-state and decadal trends (Figure 19, Table 3). Only the recent IAV in OHC is underestimated by the model, but this latter is poorly constrained by the observations in regards of the little amount of data available below 1000 m (Levitus et al., 2012; 2009; Willis et al., 2004).

The recent evolution of LCS and OCS agrees with the range of observation-based and model-derived estimates (Le Quéré et al., 2014; Takahashi et al., 2010) with an uptake of CO$_2$ of about 2.1 and 1.7 Pg C y$^{-1}$ for land and ocean, respectively (Table 3). Underestimation in mean-state OCS is essentially due to the stronger river-induced offshore outgasing of CO$_2$ which is about 0.9 Pg C y$^{-1}$ in the model and assumed to be of 0.45 Pg C y$^{-1}$ in the observation-derived estimates. Both OCS and LCS IAV are underestimated in CNRM-ESM1 compared to the estimates. For OCS IAV, this behavior is found in most ocean biogeochemical models as shown in Wanninkhof et al. (2013). Indeed, simulated IAV from biogeochemical models substantially contrasts with the large IAV estimated from atmospheric inversion which also contributes to the mix of observations and model reconstructions that compose the data (Le Quéré et
For the land carbon cycle, underestimated LCS IAV may be related to the under-sensitivity of ISBA to climate variability in contrast with the over-sensitivity to the rising CO$_2$, a behavior shared with other land surface process-based models (Piao et al., 2013). Note that differences in phase between simulated and estimated LCS were expected since the land sink of carbon is approximated from the difference between atmospheric growth rate, land-use emissions and ocean carbon sink (Friedlingstein et al., 2010).

5. Summary & conclusions

In this manuscript, we evaluate the ability of the Centre National de Recherches Météorologiques Earth system model version 1 (CNRM-ESM1) to reproduce the modern carbon cycle and its prominent physical drivers. CNRM-ESM1 derives from the atmosphere-ocean general circulation model CNRM-CM5 (Voldoire et al., 2013) that has contributed to CMIP5 and to the fifth IPCC assessment report. This model employs the same resolution and components as CNRM-CM5 although it uses updated versions of the atmospheric model (ARPEGE-CLIMAT v6.1), surface scheme (SURFEXv7.3) and sea-ice model (GELATO6) in addition to a 6-hour coupling frequency. Several biophysical coupling processes are enabled in CNRM-ESM1 thanks to the terrestrial carbon cycle module ISBA (Gibelin et al., 2008) and the marine biogeochemistry module PISCES (Aumont and Bopp, 2006). They consist of the land biosphere-mediated evapotranspiration feedback and the ocean biota heat-trapping feedbacks.

Since an earlier version of CNRM-CM5 including the marine biogeochemistry module PISCES was distributed and used in several studies (Frölicher et al., 2014; Laufkötter et al., 2015; Schwinger et al., 2014; Séférian et al., 2013), the inclusion of the terrestrial carbon cycle module ISBA constitutes the major advancement in the CNRM-ESM1 development. Although the ISBA terrestrial carbon cycle module was developed at CNRM in the 2000s, it had never been coupled to an atmosphere-ocean model and run for long climate simulations. Here, we show that ISBA embedded in CNRM-ESM1 reproduces the general pattern of the vegetation and soil carbon stock over the last decades. Although the photosynthesis scheme in ISBA differs from the
other state-of-the-art process-based models (e.g., Dalmon et al., 2014), the model displays similar behavior. That is, it overestimates both the land-vegetation gross primary productivity and the terrestrial ecosystem respiration. The compensation between these two fluxes leads to a correct land carbon sink over the modern period that agrees with the most up-to-date estimates (Friedlingstein et al., 2010; Jung et al., 2011; Le Quéré et al., 2014). The largest model-data mismatch is found in the Tropics where the gross uptake of CO$_2$ from the vegetation is strongly compensated by an overestimated autotrophic respiration. Maybe apart from this compensating mechanism, our analysis demonstrates that the terrestrial carbon cycle module of CNRM-ESM1 displays similar performances as other IPCC-Class vegetation models (Figures 20 and Figures S11 and S12, see also details in Anav et al. (2013a) and Piao et al. (2013)). The future effort in development will be oriented towards a better parameterization of the carbon absorption and respiration by the vegetation in association with a better representation of ecophysiological processes as detailed in Joetzjer et al. (2015). Further processes like fire-induced disturbance, mortality or linked with permafrost will also be included in order to improve the representation of the live biomass and soil carbon pool.

Regarding the marine biogeochemistry component, CNRM-ESM1 produces results in terms of biogeochemical variables that are comparable to other IPCC-class ocean biogeochemical models (Figure 20). The global distribution of biogeochemical tracers such as oxygen, nutrients and carbon-related fields has been improved with respect to an earlier model version presented in Séférian et al. (2013) (Figure 14 and Figure S9). This change is attributed to a stronger northward flow of deep water masses from the Southern ocean, which improves the vertical distribution of biogeochemical tracers. However, the strengthening of the meridional flow of deep water masses has also distorted the vertical structure of some carbon-related fields. Indeed, the unrealistic flow of North Atlantic deep water of about 26.1 Sv tends to deplete the stock of anthropogenic carbon storage between surface and 1200 m (Figure 16c) and consequently to increase it at depth. Since biases in anthropogenic carbon storage compensate across the water column, the simulated anthropogenic carbon storage agrees with 1994 observation-based estimates. Regarding the ocean carbon sink, CNRM-ESM1 simulates a global ocean carbon sink that falls within the lower range of the combination of observation and model estimates over the recent years (Le
Quéré et al., 2014). This slightly underestimated carbon sink is attributed to larger outgasing of natural CO$_2$ induced by the riverine input, which fits the upper range of values documented in the fifth IPCC assessment report (IPCC, 2013). Future development will target a better representation of this flux of carbon in close relationship with the recent development on the land surface hydrology (Decharme et al., 2013).

We show that CNRM-ESM1 displays results comparable to those of CNRM-CM5 in spite of the inclusion of the global carbon cycle and various biophysical feedbacks. Simulated near-surface temperature, precipitation, incoming shortwave radiation over continents as well as temperature, salinity and mixed-layer depth over oceans broadly agree with observations or satellite-derived product. Except for the salinity and the mixed-layer depth, CNRM-ESM1 display quite similar skill at simulating physical drivers of the global carbon cycle compared to CNRM-CM5. Such a comparison demonstrates the reliability of this model to produce suitable simulations for future climate change projection and impacts studies.

In addition to preindustrial control and historical simulations discussed in this manuscript, several other simulations were performed with CNRM-ESM1 following both the CMIP5 and GeoMIP experimental design. The CNRM-ESM1 model outputs (referred as “CNRM-ESM1”) are available for download on ESGF under CMIP5 and GeoMIP projects.

**Code availability:**

A number of model codes developed at CNRM, or in collaboration with CNRM scientists, is available as Open Source code (see https://opensource.cnrm-meteo.fr/ and http://www.nemo-ocean.eu/). However, this is not the case for the Earth system model presented in this paper. Part of its code is nevertheless available upon request from the authors of the paper.

**Acknowledgement:**
We thank the two anonymous reviewers for their constructive comments and suggestions of the discussion paper. This work was supported by Météo-France, CNRS and CERFACS. We particularly acknowledge the support of the team in charge of the CNRM-CM climate model. Supercomputing time was provided by the Météo-France/DSI supercomputing center. Data are published thanks to the ESGF and IS-ENES2 projects. Finally, we are grateful to C. Frauen for her kind advices on the nuances of the English language.

References:


Calvet, J.-C., Rivalland, V., Picon-Cochard, C. and Guehl, J.-M.: Modelling forest transpiration and CO2 fluxes—response to soil moisture stress, Agricultural and


Decharme, B., Martin, E. and Faroux, S.: Reconciling soil thermal and hydrological


Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Kasting, J. P. and Winton, M.: Dominance of the Southern Ocean in anthropogenic carbon and heat...


1231 Levitus, S., Antonov, J. I., Baranova, O. K., Boyer, T. P., Coleman, C. L., Garcia, H.
E., Grodsky, A. I., Johnson, D. R., Locarnini, R. A., Mishonov, A. V., Reagan, J. R.,
Sazama, C. L., Seidov, D., Smolyar, I., Yarosh, E. S. and Zweng, M. M.: The World

1235 Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R.
A., Mishonov, A. V., Reagan, J. R., Seidov, D., Yarosh, E. S. and Zweng, M. M.:
World ocean heat content and thermosteric sea level change (0-2000 m), 1955-2010,

1239 Levitus, S., Antonov, J. I., Boyer, T. P., Locarnini, R. A., Garcia, H. E. and
Mishonov, A. V.: Global ocean heat content 1955–2008 in light of recently revealed
2009.

1243 Levitus, S., Conkright, M. E., Reid, J. L., Najjar, R. G. and Mantyla, A.: Distribution
of nitrate, phosphate and silicate in the world oceans, Progress in Oceanography,

J. W.: Bacterial transformations of inorganic nitrogen in the oxygen-deficient waters
of the Eastern Tropical South Pacific Ocean, Deep-Sea Res, 37(10), 1513–1541,

1250 Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E.,

1253 Ludwig, W., Probst, J. and Kempe, S.: Predicting the oceanic input of organic carbon


1256 Institut Pierre-Simon Laplace (IPSL). [online] Available from: http://www.nemo-

1258 Maier-Reimer, E.: Geochemical cycles in an ocean general circulation model.

1259 Preindustrial tracer distributions, Global Biogeochem. Cycles, 7(3), 645,

1261 Majkut, J. D., Sarmiento, J. L. and Rodgers, K. B.: A Growing Oceanic Carbon
1262 Uptake: Results from an inversion study of surface pCO2 data, Global Biogeochem.

1265 Database of Land Surface Parameters at 1-km Resolution in Meteorological and
Climate Models, J. Climate, 16(9), 1261–1282, doi:doi: 10.1175/1520-

1268 Masson, V., Le Moigne, P., Martin, E., Faroux, S., Alias, A., Alkama, R., Belamari,


Wetzel, P., Maier-Reimer, E., Botzet, M., Jungclaus, J., Keenlyside, N. and Latif, M.:
Effects of ocean biology on the penetrative radiation in a coupled climate model, ... of climate, 19(16), 3973–3988, 2006.


<table>
<thead>
<tr>
<th>TOA</th>
<th>NSF</th>
<th>T_{2m}</th>
<th>SST</th>
<th>SWI</th>
<th>SSS</th>
<th>NIV</th>
<th>SIV</th>
<th>LCF</th>
<th>OCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>[W m^{-2}]</td>
<td>[W m^{-2}]</td>
<td>[°C]</td>
<td>[°C]</td>
<td>[-]</td>
<td>[psu]</td>
<td>[10^{3}]</td>
<td>[10^{3}]</td>
<td>[Pg C y^{-1}]</td>
<td>[Pg C y^{-1}]</td>
</tr>
</tbody>
</table>
Table 1: Drift in climate indices used to evaluate the equilibrium of CNRM-ESM1’s physical and biogeochemical components. The drifts are computed over the 250-year long preindustrial simulation of CNRM-ESM1 for the top of the atmosphere net radiative balance (TOA), the net surface heat flux (NSF), the near-surface temperature ($T_{2m}$), the sea surface temperature (SST), the sea surface salinity (SSS), the soil wetness index (SWI), the northern and southern sea-ice volume (NIV and SIV, respectively) as well as the land and ocean global carbon fluxes (LCF and OCF).

<table>
<thead>
<tr>
<th>Regions</th>
<th>Drift [units 10^{-4} century^{-1}]</th>
<th>CNRM-ESM1</th>
<th>MTE-FluxNet</th>
<th>CNRM-ESM1</th>
<th>MTE-FluxNet</th>
</tr>
</thead>
<tbody>
<tr>
<td>High latitude north (&gt;60°N)</td>
<td>4.4</td>
<td>4.5</td>
<td>-1.2</td>
<td>9.6</td>
<td>-1.6</td>
</tr>
<tr>
<td>Mid-latitude North (20°N-60°N)</td>
<td>37.9</td>
<td>34.8±2.7</td>
<td>36.2</td>
<td>3 (52%)</td>
<td>29.9±2.7</td>
</tr>
<tr>
<td>Tropics (20°S-20°N)</td>
<td>73.2</td>
<td>62.3±1.9</td>
<td>72.5</td>
<td>8 (72%)</td>
<td>54.8±1.9</td>
</tr>
<tr>
<td>Mid-latitude South (20°S-60°S)</td>
<td>16.1</td>
<td>9.3±0.6</td>
<td>15.6</td>
<td>9 (56%)</td>
<td>8.5±0.6</td>
</tr>
<tr>
<td>Global</td>
<td>130.0</td>
<td>111.3±6.0</td>
<td>126.9</td>
<td>9 (64%)</td>
<td>96.4±6.0</td>
</tr>
</tbody>
</table>

Table 2: Regional and global budget of gross primary production (GPP) and terrestrial ecosystem respiration (TER) as simulated by the CNRM-ESM1 and
estimated from the FluxNet-MTE data product. Values in brackets indicate the ratio between the autotrophic respiration (Ra) and TER. The uncertainties for the FluxNet-MTE data product derives from the regional partitioning of global mean uncertainties published in (Jung et al., 2011). GPP and TER fluxes are determined from a yearly average over 1986-2005.

<table>
<thead>
<tr>
<th></th>
<th>CNRM-ESM1</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>T₂m [°C]</td>
<td>mean 0.43</td>
<td>0.30±0.08 (Morice et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>IAV 0.13</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>trend 4.0 10⁻²</td>
<td>2.6 10⁻²</td>
</tr>
<tr>
<td>SIE [10⁶ km²]</td>
<td>mean 4.68</td>
<td>6.70±0.26 (Comiso, 1999; Fetterer et al., 2002; Rayner et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>IAV 1.14</td>
<td>0.46</td>
</tr>
<tr>
<td></td>
<td>decadal -16 10⁻²</td>
<td>-6.8 10⁻²</td>
</tr>
<tr>
<td>OHC [10²² J]</td>
<td>mean 3.34</td>
<td>3.50±1.42 (Levitus et al., 2012)</td>
</tr>
<tr>
<td></td>
<td>IAV 0.69</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>trend 0.44</td>
<td>0.50</td>
</tr>
<tr>
<td>LCS [Pg C y⁻¹]</td>
<td>mean 2.19</td>
<td>2.06±1.0 [models] (Le Quéré et al., 2014)</td>
</tr>
<tr>
<td></td>
<td>IAV 0.59</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>trend 0.3 10⁻²</td>
<td>1.8 10⁻²</td>
</tr>
<tr>
<td></td>
<td>mean 1.65</td>
<td>1.87±0.4 [models] (Le Quéré et al., 2014)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.15±0.5 [obs.-models combination]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.0±0.7 (Takahashi et al., 2010)</td>
</tr>
</tbody>
</table>
Table 3: Modern mean-state, interannual variability (IAV) and decadal trends of various global climate indices: the near-surface temperature ($T_{2m}$), Arctic September sea-ice extent (SIE), 0-2000m ocean heat content (OHC) as well as the land and ocean carbon sinks (LCS and OCS respectively). For LCS and OCS, positive values indicate an uptake of CO$_2$ by land and ocean. All metrics are computed over the 1986-2005 period for both model and observations. Decadal trends are estimated from linear regression over the 1986-2005 period. IAV is estimated from the standard deviation of the detrended time series.

<table>
<thead>
<tr>
<th></th>
<th>IAV</th>
<th>Trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>OCS [Pg C y$^{-1}$]</td>
<td>0.09</td>
<td>$4.5 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>$1.8 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

Figure 1: Fraction of dominant vegetation type as prescribed in SURFEX. This fraction results from aggregation of the various ECOCLIMAP's vegetation types at 1 km resolution over the T127 CNRM-ESM1 horizontal grid (~$1.4^\circ$ nominal horizontal resolution).

Figure 2: Time series of various climate indices along the 250-year long control simulation. (a) Net radiative fluxes at the top of the atmosphere (in red, left y-axis) and surface (in blue, right y-axis) are used to assess the stability of the climate energy flow in the model; (b) Near-surface global average temperature (in red, left y-axis) and global averaged sea surface temperature (in blue, right y-axis); (c) Soil wetness index (in red, left y-axis) and sea surface salinity (in blue, right y-axis) are used as proxy of the hydrological cycle; (d) Sea ice volume in the Northern Hemisphere (in red, left y-axis) and in the Southern Hemisphere (in blue, right y-axis) are used to
evaluate the stability of the cryosphere component in CNRM-ESM1; (e) Global carbon fluxes over land (in red, left y-axis) and over ocean (in blue, right y-axis) are used to assess the equilibration of the global carbon stock. For carbon fluxes, positive (negative) fluxes indicate an uptake (outgasing) of CO$_2$ by land or ocean.

Figure 3: Biases in simulated near-surface temperature ($T_{2m}$) compared to the CRUTV4 observations (Harris et al., 2013) averaged 1986-2005. Winter (a) and summer (b) periods are computed from DJFM and JJAS months.

Figure 4: Biases in simulated precipitation (PR) compared to the GPCP observations (Adler et al., 2003) averaged over 1986-2005. Winter (a) and summer (b) periods are computed from DJFM and JJAS months.

Figure 5: Biases in simulated photosynthetically available radiation (PAR) compared to the SRB satellite-derived observations (Pinker et al., 1992) averaged over 1986-2005. Winter (a) and summer (b) periods are computed from DJFM and JJAS months.

Figure 6: Annual bias patterns of simulated temperature T and salinity S averaged over 1986-2005 compared to the WOA2013 observations (Levitus et al., 2013). Surface biases for sea surface temperature (a) and salinity (b) are represented using the same colorbar. Vertical structure of biases for temperature (c) and salinity (d) are estimated using zonal-average biases from WOA2013 across the Atlantic and Pacific oceans.

Figure 7: Composite of yearly extremum of mixed-layer depth over 1986-2005. Left panels represent the maximum mixed-layer depth ($MLD_{max}$) for (a) observations
Sallée et al., 2010 and (b) CNRM-ESM1. Right panels represent the minimum mixed-layer depth (MLD\text{min}) for observations (c) and CNRM-ESM1 (d).

**Figure 8:** Sea ice cover (SIC) as simulated by CNRM-ESM1 averaged over 1986-2005. Top panels represent composite of September sea ice cover, while bottom panels are for March. Iso-15\% of SIC serves as comparison between model results and NSIDC observations (Cavalieri et al., 1996) averaged over 1986-2005; model results and observations are indicated with dashed and solid black lines, respectively.

**Figure 9:** Taylor diagrams showing the correspondence between model results and observations for CNRM-ESM1 and CNRM-CM5.2. Near-surface temperature (T\text{2m}), precipitation (PR) and incoming short-wave radiation (RSDS) are used to assess model performance over land surface. Sea surface temperature (SST), sea surface salinity (SSS), mixed-layer depth (MLD) and precipitation (PR) are used to assess model performance over ocean. Filled and empty symbols indicate skills for CNRM-ESM1 and CNRM-CM5.2, respectively. The size of the symbols indicates whether statistics were computed from annual mean climatology or seasonal average (JFM, AMJ, JAS, OND) over 1986-2005.

**Figure 10:** Impact of coupling frequency on sea ice cover (SIC) as simulated by CNRM-ESM1 averaged over 1986-2005. Top panels represent composite of September sea ice cover, while bottom panels are for March. Iso-15\% of SIC serves as comparison between model results using a 6-h coupling frequency (dashed lines) and those using a 24-h coupling frequency (solid lines).

**Figure 1:** Annual-mean terrestrial gross primary production (GPP). Values are given for (a) observation-derived MTE-FluxNet (Jung et al., 2009) averaged over 1986-
2005, (b) satellite-derived observation from MODIS over 2000-2013 and (c) CNRM-ESM1 over 1986-2005.

**Figure 1:** Annual-mean autotrophic respiration (Ra) as estimated from MODIS over 2000-2013 (a) and as simulated by CNRM-ESM1 (b) over 1986-2005. Panel (c) represents the zonal-cumulated Ra in function of latitude for both satellite-derived estimates (in blue) and CNRM-ESM1 (in red).

**Figure 2:** Stocks of modern soil organic carbon (cSoil) as estimated from FAO/IIASA/ISRIC/ISSCAS/JRC (2012) Harmonized World Soil Database (a) and as simulated by CNRM-ESM1 (b) averaged over 1986-2005. Panel (c) represents the zonally-cumulated soil organic stock in function of latitude for both observation-based estimates (in blue) and CNRM-ESM1 (in red).

**Figure 3:** Taylor diagrams showing the correspondence between model results and observations for CNRM-ESM1 and CNRM-CM5.2 (Séférian et al., 2013).

Climatological distribution over 1986-2005 of simulated Oxygen (O$_2$), phosphate (PO$_4$), nitrate (NO$_3$) and silicate (SiO$_2$) concentrations are assessed against WOA2013 data product. Filled and empty symbols indicate skills for CNRM-ESM1 and CNRM-CM5, respectively. The size of the symbols indicates the depth at which statistics have been computed.

**Figure 4:** Annual-mean ocean carbon fluxes (fCO$_2$) as estimated by the Takahashi et al., (2010) database (a) and as simulated by CNRM-ESM1 averaged over 1986-2005 (b). Panel (c) represents the zonal-cumulated carbon fluxes in function of latitude for both observation-based estimates (in blue) and CNRM-ESM1 (in red). Negative (positive) fluxes indicate an uptake (outgasing) of CO$_2$. 

**Figure 5:**
Figure 16: Annual-mean zonal-average anthropogenic carbon (CO$_2$$_{ANTH}$) across the Atlantic and Pacific oceans as simulated by CNRM-ESM1 averaged over 1990-2005 (a) and as estimated from the GLODAP database compiling data up to 1994 (b). Panel (c) represents the mean-annual bias in zonal structures between model and observation-based estimates in CO$_2$$_{ANTH}$.

Figure 17: Composite of yearly maximum of leaf area index (LAI$_{max}$) as estimated from AVHRR satellite observations of Zhu et al. (2013) (a) and as simulated by CNRM-ESM1 (b) over 1986-2005. Panel (c) represents the zonal-average LAI$_{max}$ in function of latitude for both observation-based estimates (in blue) and CNRM-ESM1 (in red).

Figure 18: Annual-mean surface chlorophyll concentrations (Chl) as estimated from SeaWiFS over 1997-2010 (a) and as simulated by CNRM-ESM1 (b) over 1986-2005. Panel (c) represents the zonal-averaged Chl in function of latitude for both satellite-derived estimates (in blue) and CNRM-ESM1 (in red).

Figure 19: Time series of various climate indices as monitored from available observations (blue solid line) and as simulated by CNRM-ESM1 (red solid line) since 1901 with global near-surface temperature (a), September arctic sea-ice extent (b), 0-2000m ocean heat content (c), land carbon flux (d) and ocean carbon flux (e). Hatching represents the ±2σ estimated from the ensemble deviation between the 100 members of the HadCRUT4 database (Morice et al., 2012) for near-surface temperature, the standard deviation between NSIDC Fetterer et al. (2002), Comiso (1999) and Hadisst (Rayner et al., 2003) databases, the pentadal variability of the observed ocean heat content (Levitus et al., 2012) and spread between Global Carbon Project reconstructions for both land and ocean (Le Quéré et al., 2014). For both OCS and LCS, positive (negative) fluxes indicate an uptake (outgasing) of CO$_2$. 
Figure 20: Skill-score matrix based on (a) spatial correlation and (b) globally averaged root-mean squared error for relevant fields of the simulated carbon cycle from current generation Earth System models. Leaf area index (LAI), gross primary productivity (GPP), autotrophic respiration (Ra), heterotrophic respiration (Rh) and soil carbon (cSoil) are used to assess model skill in terms of modern mean-state terrestrial carbon cycle. Sea-air carbon flux ($f_{gCO_2}$), surface chlorophyll (Chl) and surface concentrations of oxygen ($O_2$), nitrate ($NO_3$), phosphate ($PO_4$) and silicon (Si) are used to evaluate the skill of the current models at replicate modern mean-state ocean carbon cycle. Both models and observed fields are averaged over time from 1986 to 2005 to determine skill score metrics, except for cSoil, $O_2$, $NO_3$, $PO_4$, Si observations (only a modern mean-state climatology is available). Black squares indicate that models fields are not available (implying that these fields are either not simulated by the model or not published on the ESGF).