Project scope and plan

Project name

TOwards a **P**redictive **Sy**stem for atmospheric CO_2 and **Cl**imate fe**ed**backs (**TOPSyCled**).

Research field

PE10_3 Climatology and Climate change

Principal Investigator (PI)

Dr.

Bernardello

Raffaele

Barcelona Supercomputing Center

Earth Sciences

Climate Prediction Group

Spain

1 Key scientific/societal/technological contribution of the proposal (200 words max.)

The so-called climate warming hiatus over the first decade of the 21st century took the research community by surprise, prompting a wave of skepticism among the public. A similar situation could occur again, if for example, atmospheric CO_2 concentration was changing at a rate that would seem inconsistent with reported trend in global emissions. Even on a year-to-year time scale, changes in atmospheric CO_2 growth rate, primarily caused by natural fluctuations are often misinterpreted as reporting of emissions growth rates. Such misunderstandings could be addressed with the development and deployment of a near-term carbon cycle prediction system.

TOPSyCled will assess the predictability of the carbon cycle system, via ESM control experiments and decadal hindcasts over the last 60 years. We will then explore predictions of the climate-carbon system in the near-future, assuming anthropogenic emissions follow the United Nations Framework Convention on Climate Change (UNFCCC) Nationally Determined Contributions (NDCs) and quantify the direct impact of emission reductions on CO_2 concentrations, accounting for the natural variability of the climate system and the carbon sinks. This represents a major step towards the verification of near-term emission trends, and the timing of any emissions peak, providing policy-relevant analysis for the UNFCCC global stocktakes.

2 Detailed proposal information (Maximum 14 pages, graphs and tables included)

2.1 Justification for the importance of the scientific problem and the requested resources (~2 pages)

The **Intergovernmental Panel on Climate Change 5th Assessment Report (IPCC-AR5)** concluded that "Cumulative emissions of CO_2 largely determine global mean surface warming by the late 21st century and beyond", unambiguously identifying the causal link between anthropogenic emissions of carbon dioxide (CO_2) and global warming[1], based on recent advances in Earth system modelling. The IPCC AR5 also assessed the positive feedback between climate change and the carbon cycle stating: "there is high confidence that the feedback between climate and the carbon cycle is positive in the 21st century. As a result more of the emitted anthropogenic CO_2 will remain in the atmosphere". These findings highlight the central role of the carbon cycle in the global climate system.

In 2015, under the **United Nations Framework Convention on Climate Change (UNFCCC)** the 21st Conference of the Parties (**COP21**) signed a historical agreement known as the Paris Agreement on climate **(PA hereafter)**. The PA reaffirms the goal of limiting global temperature increase to well below 2°C by the end of the 21st century. Under Article 4 of the PA, parties "aim to reach global peaking of greenhouse gas emissions as soon as possible" and to "undertake rapid reductions thereafter in accordance with the best available science". Such ambitious goal is pursued by requiring all parties to put forward their best efforts through **"Nationally Determined Contributions" (NDCs)** and to strengthen these efforts in the years ahead. Under Art. 14 of the PA, a **"global stocktake"**, to take place in 2023 and every 5 years thereafter, "will assess collective progress toward achieving the purpose of the Agreement". The global stocktake outcome "will inform parties in updating and enhancing their actions and support and enhancing international cooperation on climate action." Hence the first major global milestone in the implementation of the PA is for emissions to reach a peak and start decreasing. According to the AR5 emission scenarios, this must occur within the next decade. Even identifying whether emissions have peaked is both a detection and a prediction challenge: emissions must be observed to fall, and predicted to continue to do so. Given current uncertainties in our understanding of the carbon cycle, the “best available science” at present would be unable to detect with confidence that emissions have peaked until one decade or more after they had actually done so.

To this end the newly funded **H2020 project CCiCC** (Climate-Carbon interactions in the Coming Century – start date June 2019) aims at reducing this uncertainty as an essential contribution to the UNFCCC global stocktake process integral to the implementation of the Paris Agreement. So far, there has been little attempt by the scientific community to predict the near-term evolution of the carbon cycle, and in particular what would be the near-term growth rate of atmospheric CO_2 in the next decade if all countries follow their PA ambitions on emissions reduction. **There is an urgent need to develop the capability to simulate and assess the near-term evolution of the global carbon cycle and the climate system in response to different near-term emission trajectories.** CCiCC addresses directly this need in work package 2 of which the PI of this proposal is co-leader. CCiCC is closely linked to two of the World Climate Research Programme’s (WCRP) Grand Chanllenges**: 1) Near-term Climate Prediction and 2) Carbon Feedbacks in the Climate System.** This proposal describes part of BSC’s contribution to CCiCC. Such contribution would not be possible without the support of **PRACE** and the access to **tier-0** computational resources.

The **overall objective** of **TOPSyCled** (this proposal) is **to develop new tools and methods to predict, for the first time, the evolution of global carbon cycle variability over the coming decade, including atmospheric CO_2, land and ocean carbon sinks, and climate response to track the overall progress towards the goal of the Paris Agreement**. This will be achieved along with an improved understanding of decadal variability of the climate-carbon cycle system, accounting for forced response and natural variability. We will use the **EC-Earth3 Earth System Model** to develop and continually improve initialization techniques via validation against available observational products. The best-performing initialization technique will be used to perform future near-term predictions assuming anthropogenic emissions follow the UNFCCC NDCs ambitions and, as a baseline scenario, the RCP4.5. This will allow anticipating and explaining the near-term evolution (up to 2030) of atmospheric CO_2 increase and climate response, in time for the first global stocktake in 2023.

Decadal predictions require to be performed in ensembles as large as possible to make sure the spread of the members is representative of the uncertainty of the observations used as initial conditions[2]. Furthermore, retrospective predictions covering a period as long as possible ensure the robustness of the bias correction (that accounts for inherent model biases) that will be used also to correct future predictions. **TOPSyCled** will be divided in 4 tasks. **Task-1** will explore the mechanisms driving decadal predictability of the global carbon cycle by performing idealized process-oriented predictions **(7.99 Mcore-hours)**. **Task-2** will reconstruct the recent past (1958-present) to provide initial conditions for the retrospective predictions **(0.46 Mcore-hours)**. **Task-3** will assess the predictability of our modelling system with retrospective predictions over the period 1981-present **(22.37 Mcore-hours)**. Finally, **Task-4** will provide unique emission-driven predictions of the near-future (2019-2028) evolution of atmospheric CO_2, carbon cycle and climate when emissions follow either the UNFCCC NDCs or a baseline scenario **(0.61 Mcore-hours)**.

Such an ambitious exercise requires appropriate access to tier-0 computing resources and the associated support offered by PRACE. **A total of 34.57 Mcore-hours** are requested in this proposal, including a 10% overhead to account for failing jobs that will need to be repeated. The detailed justification of this amount is presented in Sections 2.2 and 2.6. An ambitious set of analyses will be performed. The project will be carried out by members of the Earth Sciences Department of the Barcelona Supercomputing Center. The data produced will be made publicly accessible via the EUDAT collaborative data infrastructure of which BSC is a partner and node-hosting member.

2.2 Overview of the project (~4 pages)

The slowdown in global warming that emerged over the first decade of the 21st century took the research community by surprise. Climate scientists were not able to account for what was happening[3], which prompted a wave of climate scepticism among the public. A similar "climate surprise" could occur again in the near future, if for example, atmospheric CO_2 concentration was changing at a rate that would seem inconsistent with reported trend in global emissions. The research community needs to be able to understand the processes at play and to clearly communicate these, in order to enable appropriate policy responses. Even on a year-to-year time scale, the reporting of changes in atmospheric CO_2 growth rate (Figure 1), primarily caused by natural fluctuations of land and ocean carbon sinks[4,5], is often misinterpreted as reporting of emissions growth rates[6]. Such misunderstandings could be addressed with the development and deployment of a near-term carbon cycle prediction system. This would allow us to analyze the relative importance of the natural variability and the degree of uncertainty that needs to be taken into account to detect the impact of NDCs. In fact, these are defined thinking in the long term for climate stabilization but the climate trend is not constant and every year is different. As a consequence, the resulting atmospheric CO_2 following the implementation of NDCs might not be the expected on a decadal timescale, due to natural variability. Decadal predictions provide the ability to anticipate and explain such possible outcomes.

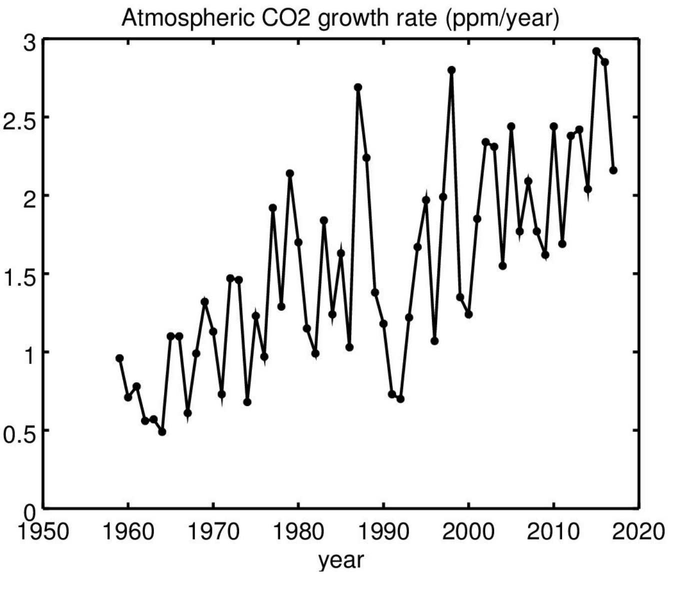


Figure 1: Global averaged atmospheric CO_2 growth rate (source: NOAA)

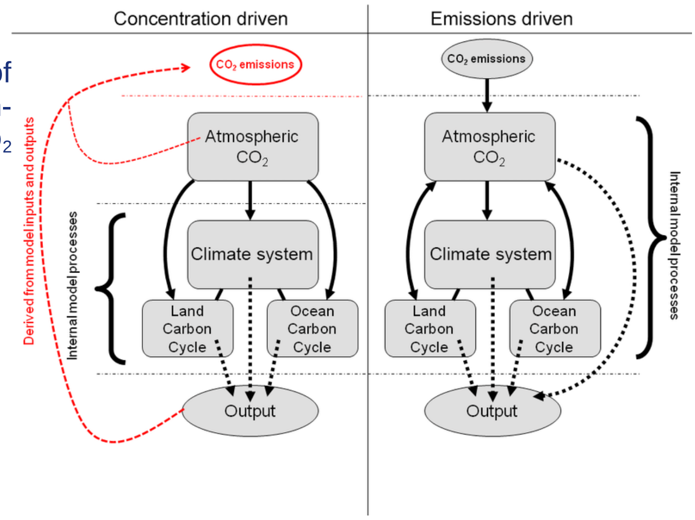
Our best tools to predict the near-term future evolution of atmospheric CO_2 and climate response are Earth System Models (ESMs). These are complex numerical representations of the primitive equations of fluid motion for both the atmosphere and ocean. Coupled to these are representations of other climate-relevant processes (i.e. land hydrological cycles, sea-ice and cloud formation etc.) and of biogeochemical cycles of elements influencing the climate like carbon.

Over the past decade, near-term climate predictions have emerged as rapidly improving tools at the service of society and decision-makers. The CMIP5 model experiment suite included a set of such predictions that proved skilful at regional scales[7]. Moreover, near-term climate predictions have proven their ability to predict global-scale variability mechanisms like, for example, the fluctuations in the strength of the North Atlantic sub-polar gyre[8] and the Atlantic meridional overturning circulation[9] .

Future near-term climate is the result of two components a) change in atmospheric radiative forcing and b) the natural variability of the climate system. ESMs can be used to simulate historical and future climate by prescribing the radiative forcing based on observed data and future emission scenarios. These simulations do not attempt to phase the model with the observed natural variability of the climate and are thus useful only in a statistical sense on centennial timescales. ESMs can also produce climate reconstructions where, besides the radiative forcing, the natural variability is also taken into account by continuously constraining the model's solution towards the observed state of the climate through numerical techniques commonly referred to as data assimilation (or nudging). Finally, ESMs can be used to perform near-term climate predictions where the radiative forcing is still prescribed throughout the simulation but only the simulation's initial state is constrained towards the observed climate through data assimilation, a procedure referred to as initialization. The evaluation of the ability of a model and a particular initialization technique to produce skilful near-term climate predictions is normally assessed by comparing retrospective climate predictions with available observations. Retrospective climate predictions are near-term predictions of the past climate initialized using only contemporaneous information available at the time of starting the simulation[7]. Here we use near-term and decadal as synonyms.

While these predictions can be performed using state-of-the-art ESMs with a complete description of the carbon cycle, the predictability of the carbon cycle received little attention so far. Thus, the extension of this exercise beyond the physical climate is an emerging and promising topic which has been explored only in a few models so far. As of now however, no modelling group has attempted ESM-based initialized near-term predictions of the global carbon cycle driven by CO_2 emissions. Furthermore, the only decadal predictions of the climate system planned for CMIP6 are based on a middle of the road greenhouse gas scenario (RCP4.5), not highly relevant in the context of the PA. **TOPSyCled** will produce unique decadal predictions of the global carbon cycle with one state-of-the-art CMIP6 ESM driven by near-term NDCs compliant emissions, allowing us to develop the capability to assess the success of NDCs implementations in terms of expected atmospheric CO_2 and climate response, accounting for the natural variability of the carbon and climate system, ultimately providing unique policy relevant information for the global stocktake. Such predictions will be important tools to indicate potential early warning of systematic errors in emission reporting, or carbon cycle response.

Figure 2: Schematic of CO_2 concentration-driven and CO_2 emission-driven simulations[10] .



At this point, it is neccessary to clarify what is meant by **"concentration-driven"** and **"emission-driven"** predictions. In the first case, atmospheric CO_2 concentration is prescribed and it does not respond to the variability in the land and ocean carbon sinks. These are quantified and, combined with the prescribed atmospheric CO_2 growth rate, allow to compute the so-called compatible emissions. In the second case, atmospheric CO_2 concentration is interactively calculated by the model as a response to the ocean and land carbon sinks while emissions of CO_2 are prescribed instead of concentrations. In this case the carbon feedback is taken into account and, importantly, atmospheric CO_2 concentration is a prognostic variable that can be predicted (Figure 2). As emission-driven near-term predictions are still an unexplored ground, CCiCC partners decided to adopt a dual strategy where most of the simulations planned will be carried out both in concentration-driven and emission-driven modes.

To achieve the overall objective of TOPSyCled (see Section 2.1) we have identified four **specific objectives** as follows:

1) - To understand and quantify the potential predictability of atmosphere-land and atmosphere-ocean carbon fluxes.

2) - To test initialization techniques for retrospective and future predictions.

3) - To quantify the predictability of the carbon cycle and climate systems.

4) - To perform future predictions of atmospheric CO_2 using NDCs-based CO_2 emissions.

Activities in TOPSyCled have been organized in tasks appositely designed around these 4 specific objectives. Below we provide a detailed explanation of these tasks and of the resources needed to carry them out successfully.

Task-1. A classic way to quantify potential predictability is through a "perfect model approach"[11] in which we assume that the model reproduces all the processes driving the predictability of a given variable and that such representation is not affected by model biases. This potential predictability is a measure of how long the memory of an initial state drives the evolution of a given variable in a given region. Or in other words, it quantifies the ability of the model to predict itself. Such an approach is well suited to establish a theoretical framework of predictability assessment across different models, as well as to investigate and better understand the processes driving predictability of the global carbon cycle.

In our case, the perfect model framework will be used to assess the potential predictability of key variables for the carbon cycle for which we have insufficient observations. This allows for a better understanding of the mechanisms driving low frequency (interannual to multi-decadal) variability of land and ocean CO_2 exchange with the atmosphere. To improve our understanding of the processes driving the low-frequency variability we will also estimate the contribution of the specific regions to the potential predictability of the climate-carbon system.

We will select 5 starting dates from an existing multi-century preindustrial control simulation and will run for each starting date a 15-member ensemble of 10 years, with slightly perturbed initial conditions. Starting dates will be selected from the preindustrial control to cover a variety of climate modes.

We will re-run the same predictions but in a pacemaker setup focusing on the key regions that determine the variability of surface carbon fluxes. In such pacemaker experiments, only the regions of interest will be subject to perturbed initial conditions, enabling us to quantify the impact of such hot spot regions on the potential predictability of atmospheric CO_2. This variability hotspot includes for instance, the North Atlantic, the Southern Ocean, or the equatorial Pacific. Such simulations will enable us to assess the regional vs. global effects on the carbon cycle and the pathways in the climate-carbon system through which the variability drivers are expressed. All experiments will be carried out in both concentration-driven and emission-driven mode.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
3	5	15	10.00	768	1.99	**3,438,720.00**

Table 1. Concentration-driven experiments in perfect model framework.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
3	5	15	10.00	720	2.81	**4,552,200.00**

Table 2. Emission-driven experiments in perfect model framework.

Task-2. Already existing CMIP6 "historical" simulations (1850-2015) although not in phase with the observed climate, provide a good starting point to produce reconstructions where ocean and atmospheric physical fields are nudged to observations to phase the model's climate with the observed variability. We will not directly assimilate any observation for the carbon cycle allowing instead both land and ocean biogeochemical models to evolve following the physical constraints of the ocean and atmospheric forcings. We will let the model's climate adjust to the observations for two decades (1958-1978) and then consider initial conditions for retrospective near-term predictions.

The reconstructed fields are used as a compromise between the inherently biased model solution and the observed state of climate and carbon cycle. We will test several options (up to 5) for producing the reconstructed fields by varying, for example, the weight of the nudging towards observations. This will translate in several possible solutions for the initialization of the retrospective predictions that we will assess against existing observations to compute metrics that will allow highlighting strengths and weaknesses of our reconstructed carbon cycle.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
5	1	1	60.00	768	1.99	**458,496.00**

Table 3. Reconstructions of historical climate and global carbon cycle. Up to 5 different options will be tested.

Task-3. We will perform decadal retrospective predictions using the initial conditions that best approximate observations among those tested in Task-2. We will use both CO_2 concentration-driven and newly developed CO_2 emission-driven simulations. The latter will enable us to simulate prognostic atmospheric CO_2 concentrations along with the evolution of land and ocean carbon sinks, therefore accounting for emerging climate-carbon feedbacks and the climate response all together. Retrospective predictions will be bias-corrected following procedures that have been tested and investigated over the past decade in the context of seasonal-to-decadal climate predictability[12]. Our retrospective predictions will be performed every year for the period overlapping the availability of observations (see Section 2.3) to allow for a more meaningful statistical reliability of results, drift treatments, as well as predictive skill assessment.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
1	42	15	10.00	768	1.99	**9,628,416.00**

Table 4. Retrospective predictions in CO_2 concentration-driven mode.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
1	42	15	10.00	720	2.81	**12,746,160.00**

Table 5. Retrospective predictions in CO_2 emission-driven mode.

Task-4. With the new knowledge acquired on the mechanisms responsible for the climate-carbon system predictability and with the improved predictive systems developed during CCiCC we will attempt to forecast the future near-term evolution of atmospheric CO 2 over the stocktake period using for the first time ever emission-driven ESMs for decadal predictions. These will be performed assuming CO_2 emissions follow NDC ambitions while non- CO_2 emissions will be prescribed according to the sustainability narrative SSP1. These predictions will be compared to “baseline” decadal predictions that use SSP2-4.5 CO_2 emissions and non- CO_2 prescribed concentrations, allowing us to assess the skills to detect and attribute changes in atmospheric CO_2 and climate response due the implementation of the UNFCCC NDCs.

N. Experiment	N. start dates	N. members	Length (years)	N. cores	Wall clock (hr/year)	Tot. core-hours
2	1	15	10.00	720	2.81	**606,960.00**

Table 6. Future predictions of atmospheric CO_2 following NDCs emissions and RCP4.5 emission scenario as a baseline.

By the end of the project, we have the ambition to provide robust annual to decadal predictions of atmospheric CO_2, land and ocean carbon sinks and climate response, in order to inform on the possible outcome of the implementation of the UNFCCC compliant anthropogenic emissions (NDCs) in time for the 2023 global stocktake. Up to now, ESMs driven by CO_2 emissions have been only used for long-term projections[13], but not in the context of near-term decadal prediction. We will move beyond state-of-the-art by attempting for the first time to predict the evolution of the coupled climate-carbon cycle on decadal timescale, with the additional value of using a fully interactive emission-driven ESM. This will allow us to inform on the possible changes in atmospheric CO_2 and climate resulting from both emissions policies and internal climate and carbon cycle variability.

TOPSyCled’s expected outcomes and impacts are:

-supporting major international scientific assessments such as the IPCC and the UNFCCC global stocktake in 2023;

-increase confidence in climate predictions by advancing our understanding of decadal variability and by improving initialization techniques;

-increase interest in climate predictions by broadening their scope to the prediction of biogeochemical variables like (but not only) atmospheric CO_2;

-providing added-value to decision and policymakers;

-sustaining Europe's leadership in climate science.

2.3 Validation, verification, state of the art (~2 pages)

2.3.1 Validation

Retrospective predictions of ocean and land carbon sinks (Task-3) will be validated using an observation-based reconstruction of air-sea CO_2 flux (1981-present)[14] and an upscaling product of evapotranspiration and CO_2 fluxes over land (1950-2014)[15]. Reconstructions of climate and biogeochemical conditions (Task-2) will be evaluated using a variety of observational products: GO-SHIP ocean interior carbon (1994-present)[16], ocean CFC-11, CFC-12 and SF6 (1982-present)[17], terrestrial water storage (GRACE 2003-present)[18]. Furthermore, emission-driven retrospective predictions will be validated against observed atmospheric CO_2 concentrations provide by NOAA. Finally, using the same source, for the first time future predictions of atmospheric CO_2 will be also validated, at the end of the project.

2.3.2 Verification

On-line monitoring of simulations directed at key diagnostics will allow to detect errors in the experiment setup or numerical instabilities. This will ensure the best use of computational resources as problematic experiments can be stopped as soon as a problem is detected. Key variables are for example: mass conservation of biogeochemical elements, surface temperature bias with respect to observations etc.

EC-Earth3-ESM has been tested for reproducibility on MN4. We use restart files for all components that allow bit reproducibility when the number of cores and MPIconfigurations are maintained. This will be the case for this project as all experiments will be performed with the same configuration. However, if for any reason we have to change our configuration we have developed a tool that allows us to test reproducibility based on statistical significance. This tool was developed to ensure consistency of simulations performed on several platforms across Europe because EC-Earth3 is a community model.

2.3.3 Sensitivity analysis and uncertainty quantification

In 2017, the Global Carbon Project (GCP), led by several CCiCC project partners, assessed for the first time each term of the global carbon budget independently, with an estimate of the land sink based on an ensemble of land carbon cycle models[19]. This new approach allows quantifying the carbon budget imbalance (B_{IM}) when combined with the rest of the terms estimated (fossil fuel emissions, land-use change, atmospheric CO_2 increase and ocean CO_2 uptake). B_{IM} is a measure of imperfect closure of the global carbon budget and hence it offers a quantitative measure of the community's level of understanding of the contemporary carbon budget. The B_{IM} shows annual absolute errors of 0.7 $GtCyr^{-1}$ on average, with large year-to-year variability of 1-2 $GtCyr^{-1}$ (corresponding to 10-40% of fossil emissions), and also longer, semi-decadal anomalies of 0.5-1 $GtCyr^{-1}$. The B_{IM} does not show any clear bias, with both long-term mean and the trend close to zero. Given the very low year-to year variability in anthropogenic emissions[20], we expect the B_{IM} variability to be primarily due to errors in the understanding of processes driving land and ocean carbon sinks, and their responses to climate variability, as represented in models.

The magnitude of the B_{IM} severely limits our capability to detect any near-term changes in atmospheric CO_2, and therefore to correctly attribute such changes to emission mitigation efforts or to internal natural variability of the climate-carbon system[6]. Given the magnitude of B_{IM} in the near future, it would take 10 to 20 years to detect a 1% change in the increase of CO_2 emissions at the 68% confidence level (e.g. from the 1% per year increase of the past few years to a 0% per year, i.e. emissions stabilization). To reduce this detection time and thus provide meaningful near-term predictions of atmospheric CO_2 and the carbon cycle, we will use observational constraints in two ways: first to assist in the choice of the initial conditions that best approximate observations of the land and ocean carbon reservoirs; and second to provide stronger constraints, and potentially apply bias corrections, on the temporal evolution of the simulated land and ocean carbon sinks.

2.3.4 Comparison with state of the art

The predictability of the carbon cycle has received little attention so far. Thus, the extension of this exercise beyond the physical climate is an emerging and promising topic which has been explored only in a few models so far to investigate the predictability of oceanic primary production over the tropical Pacific[21] or the global ocean[22] and of the carbon uptake over the North Atlantic[23] or the global scale in a perfect model set up[24].

These studies have shown potential predictability of the global land and ocean carbon uptake of up to six years, with a median predictability horizon of four years. Predictability of global carbon uptake is driven by the ocean's predictability because of its stronger capability to generate low-frequency fluctuations in carbon flux[24]. More recently, other authors[25] have found that variations of the global ocean CO_2 uptake are predictable up to 2 years in advance with however, evidence for a higher predictive skill up to 5 years regionally.

As of now however, no modelling group has attempted ESM-based initialized near-term predictions of the global carbon cycle driven by CO_2 emissions.

2.4 Software and Attributes (~2 pages)

2.4.1 Software

The EC-Earth3 GCM (Global Climate Model) comprises three major components: the atmospheric model IFS (Integrated Forecasting System) Cy36r4, the ocean model NEMO 3.6[26], which also includes the LIM3 sea-ice model[27], and OASIS3 that couples the main components. IFS is an operational global meteorological forecasting model developed and maintained by the European Centre of Medium-Range Weather Forecasts (ECMWF). NEMO is a state-of-the-art modelling framework for the ocean used for oceanographic research, operational oceanography, seasonal forecasting and climate research studies. The ESM (Earth System Model) version of EC-Earth3 includes additional components, also coupled via OASIS3: LPJ-GUESS dynamic vegetation model[28], PISCES ocean biogeochemistry model[29] (as a NEMO module) and TM5 global atmospheric transport model[30]. LPJ-GUESS is used to simulate the evolution of the land vegetation and carbon fluxes, PISCES is used to simulate ocean biogeochemistry and CO_2 fluxes with the atmosphere and TM5 is used for atmospheric chemistry and transport of trace gases such as CO_2. In this activity we will use the T255-ORCA1 configuration, which corresponds to a spatial resolution of 80 km in the atmosphere/land and 100 km in the ocean, and 3x2 degrees with 10 vertical levels, CO_2-only configuration for TM5.

In TOPSyCled we will use EC-Earth3-ESM in two different configurations. For CO_2 emission-driven simulations the full configuration is needed, including atmospheric transport module TM5. On the other hand, for CO_2 concentration-driven simulations, a reduced configuration is sufficient (no TM5 needed) resulting in a considerably less computationally demanding implementation. A detailed optimization and scalability analysis of the two configurations will be given in Section 2.6.

2.4.2 Particular libraries

For configuring and building the model executables, GNU make 3.81 or 3.81+, FORTRAN 77/90/95 compliant compiler and C++ compiler with pre-processing capabilities, NetCDF4 deployed with HDF5 and SZIP, as well as HDF4 libraries are needed. A tool for automatic build configuration called "ec-conf" can be used. This useful tool requires Python 2.4.3 or 2.4.3+ (although it does not work yet with Python 3.0+). For NEMO, the FCM bash and perl mechanism is essential, as it is the I/O GRIB_API 1.9.9 or 1.9.9+ and GRIBEX 370 mechanism that are needed for IFS. To test the model with the run scripts, GNU date (64-bit) is also required.

The simulations will require MPI libraries and runtime facilities (MPICH2, MPICH-MX, HP-MPI, OpenMPI), optimization and data handling tools, such as BLAS, LAPACK, HDF4, HDF5, NETCDF, PARMETIS, SCALAPACK, P-NETCDF, UDUNITS, GRIB_API, CDFTOOLS v2, CDO, NCO and general configurations tools, such as PERL, PYTHON, AUTOCONF and AUTOMAKE.

2.4.3 Parallel programming

The EC-Earth3 model is composed of NEMO for the ocean, IFS for the atmosphere, OASIS for the coupling and XIOS for I/O management of NEMO files. All the components are parallelised in the space dimension by using MPI. IFS supports also OpenMP, but within EC-Earth this model runs only in MPI mode.

2.4.4 I/O requirements

NEMO's I/Os are handled by a parallel I/O library called XIOS that reads and writes NetCDF3 and NetCDF4 files, using one or more “server” processes dedicated exclusively to the IO management. With XIOS, the I/Os are managed in an asynchronous way so that the read and write doesn't penalize the computational part. IFS writes files directly in GRIB format while OASIS3 does not generate any output. At the end of a simulation the three components always generate restarts separately (IFS in binary, and NEMO and OASIS3 in NetCDF format).

Each year of simulation is estimated to produce about 30 GB of output in about 70 files (one per each ocean, sea ice, atmospheric or biogeochemical variable) to be stored long term. The size of each file varies because some files contain 2D and others 3D variables. This output will be post-processed on the HPC platform to compute a small number of 1D climate indices for the experiment scientific monitoring and to format the files according to international standards.

To make sure that any chunk can be repeated should an output file discovered to be corrupted, the restart files (containing a snapshot of the atmosphere, ocean, sea-ice, coupler, land and ocean biogeocemistry as well as atmospheric CO_2 distribution) of a total size approaching 20 GB will be also kept in the long-term repository. In particular, the restarts of the reconstruction runs will act as initial conditions for the retrospective predictions. Chunks can be repeated at any time by the workflow manager for as long as the PRACE accounts remain open to solve any contingency that corrupted the data.

Assuming that at the production peak there will be ten ensemble members of the retrospective predictions running simultaneously, each year of ten-member ensemble simulation will produce 300 GB of output and 200 GB of the corresponding restarts. Taking into account that the next group of ensemble members will start running as the post-processing of the previous one is underway, the typical amount of data that will reside simultaneously in the working file system should be at least three times that produced by the ten members mentioned: one for the set of simulations running, another one for the set being post-processed and, finally, a third one for the set being transferred to the BSC local archive. This approach requires around 1.5 TB of working space. Transfer rates between Marenostrum4 and the BSC local archive of up to 5 TB per day have been reached, which should fit the plan mentioned. However, as the transfer rate is not yet known because it depends on the platform where the request is allocated, the production rate will have to be adjusted to this factor. To cater for maintenance activities in the BSC local storage or an increase in the line traffic that might temporarily lower the transfer rate, the space requested for handling the output data is increased to 4 TB, which offers a buffer of one day of theoretical maximum production that cannot be transferred immediately.

2.5 Data: Management Plan, Storage, Analysis and Visualization (~1 page)

2.5.1 Data Management Plan covering

The data stored in the BSC local archive will be managed and curated by the Data and Diagnostics Team of BSC-ES. They have developed a framework to store all the simulations and the observational data required by the BSC-ES researchers that offers access to all the data with a strict documentation, organisation and, of course, formatting. The data produced in TOPSyCled will be made accessible via the EUDAT collaborative data infrastructure, of which BSC is a partner and node-hosting member. The terms of access will be fully public and accept commercial use of the data to better link to the future activities in decadal prediction of the Copernicus programme, which is arguably one of the most efficient ways to link to a wide range of climate data users.

2.5.2 Project workflow

Thanks to the well-defined workflow structure and the use of an adapted workflow manager like Autosubmit (see Sec. 2.5.3), the data produced by each simulation will be transferred to the permanent storage using a dedicated transfer node. The analysis of results is run locally as soon as the post-processing of a chunk is completed, which should happen on average around 3 hours after the chunk started running, although this time depends on the queue capacity for the different types of jobs. In this context, a standard two-week delay between the end of the project and the closing of the PRACE accounts is more than enough for the team members to clean the HPC repositories.

2.5.3 Software workflow solution

The Autosubmit software will be adopted to manage the workflow and ensure a uniform and optimal use of the resources. Autosubmit was developed by and is maintained at BSC. It is a python-based tool to create, manage and monitor experiments by using Computing Clusters, HPC's and Supercomputers remotely via ssh. It has support for experiments running in more than one HPC and for different workflow configurations. The jobs will be managed, and packed in groups in a single big job whenever required, by Autosubmit to better manage the I/O system while maximising the use of the machine.

2.5.4 I/O requirements

Most of the analysis will be run locally after the transfer of output is completed. Only limited post-processing is performed on the supercomputer to diagnose the correct functioning of the experiment. For this reason, we do not have any special I/O requirements related to analysis and visualization of results.

2.6 Performance of Software (Maximum 3 pages)

2.6.1 Testing of your code on the requested machine

EC-Earth3 has been tested and extensively used in production in MareNostrumIV for previous tier-0 PRACE calls under projects "HiResClim", "HiResClim2", "HiResSIR", "LSIHP", "Glob15km" or "HiResNTCP", as well as Spanish Supercomputing Network (Red Española de Supercomputación, RES) calls under projects AECT-2019-1-0011, AECT-2018-3-0023, AECT-2018-3-0006, AECT-2018-2-0011, AECT-2018-1-0006, AECT-2018-1-0020 and AECT-2017-2-0008.

2.6.2 Quantify the HPC performance of your project

The full configuration of the EC-Earth3-ESM required a new study, with respect to the EC-Earth3-GCM configuration, to evaluate the performance of the complete Earth System Model. New parallel components have been included for this Multiple Program, Multiple Data (MPMD) application, increasing the complexity for different components which run independently in parallel, each one with different execution times and parallel algorithms. After a preliminar analysis, we remarked that the performance of the complete coupled model was reduced by 50% once one of the new components (TM5) was coupled to the complete system. This is due to two factors: the coupling of IFS to TM5 limits the number of cores allowed for the IFS model to 256, and the coupling requires an important number of field exchanges from IFS to TM5 with a 6-hour frecuency. We decided to start our analysis for the coupled version using only IFS and TM5. For this study, a profiling tool called Extrae was used to instrumentalize the execution of the model and Paraver was used for visualization. The next figure shows an example of the work done.

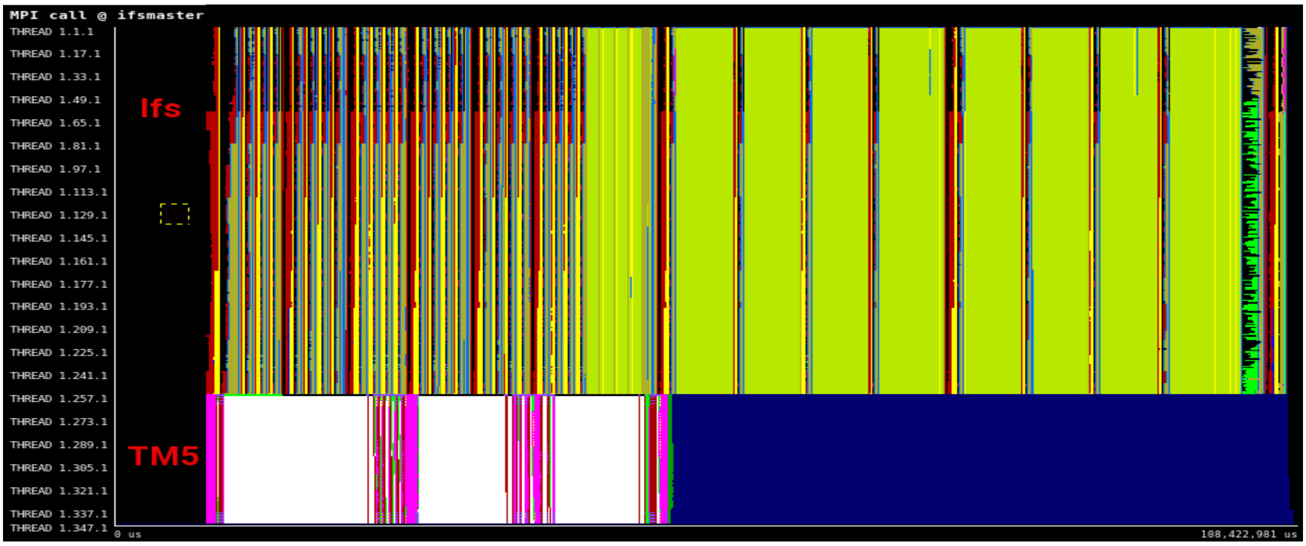


Figure 3: Paraver visualization of profiling obtained with Extrae applied to the coupled system IFS-TM5.

The figure shows the coupled execution of IFS and TM5 on Marenostrum4. The y-axis represents the MPI processes used for the parallel execution, using 256 MPI processes for IFS and 90 for TM5. The x-axis represents the execution of the model along the time, running 1 day of simulation (32 time steps). The trace shows different MPI events and their duration along the time. For example, pink color represents MPI reduces and yellow color broadcast operations. Additionally, black color represents calculation time (doing pure computation) and blue color (at the end of TM5) represents waiting time, since TM5 finishes before IFS, even though the computation done by IFS at that time is related to coupling information.

The trace analysis and the profiling study done proved that the main overhead of this execution is coming from MPI operations (more than 30% for IFS and 80% from TM5). The solution to increase the computational performance without modifying the algorithms was to optimize the Intel MPI library parameters related to the network of Marenostrum4 and the use of binding (affinity) for MPI processes, alternating TM5 and IFS processes to reduce network saturation and memory bounding. Thanks to these improvements, we managed to increase the maximum SYPD by 40%. At the same time, an analysis of the execution time and the trace analysis revealed the maximum parallel efficiency achieved for each component. This allowed to us to reduce the parallel resources by 20%, improving the efficiency without losing computational performance.

A second step in the optimization of the code was to reduce the vertical resolution of TM5 from the full 34 levels to 10 levels which allowed to double the throughput. However, this modification required a scientific assessment to make sure that the solution with reduced vertical resolution was still acceptable for our purposes. In Figure 4 we show the surface CO_2 concentrations after one year of simulation for the chosen set of vertical levels. It was decided that the 10 level configuration offered the best performance while preserving realistic spatial patterns of CO_2 concentrations.

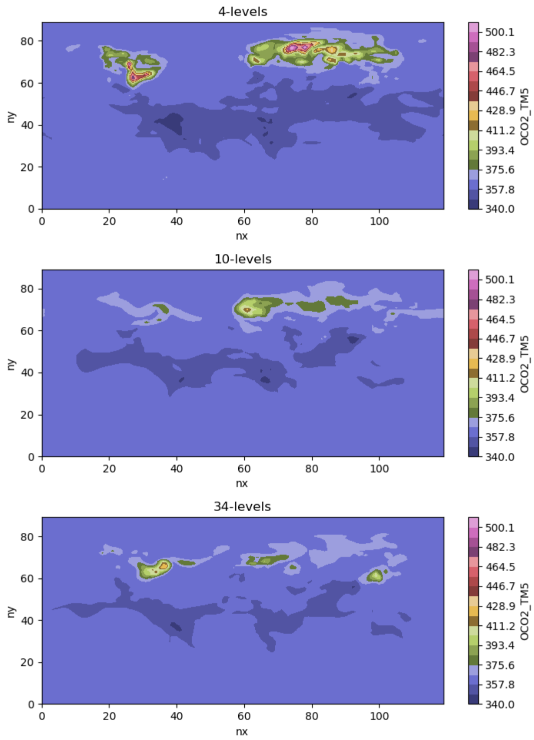


Figure 4: Surface concentration of atmospheric CO_2 as simulated by TM5 with different vertical resolution.

2.6.2.1 *Strong and weak scalability*

Scalability is reported for the two configurations that will be used in TOPSyCled. For Earth System applications, weak scalability metrics are not applicable since the problem to solve is fixed in space (i.e. surface of the earth in the horizontal dimension and height of the atmosphere and depth of the oceans in the vertical dimension) and therefore only strong scaling metrics are used. Note that for the Emission-driven configuration, the use of TM5 strongly limits the lower and upper bounds of processor configurations for IFS, which explains the unusual parallel efficiency numbers. The optimization of the code setup and the maximization of the throughput was carried out through other means as explained in the previous section. In the following tables the setup highlighted in red was the chosen one.

Concentration-driven configuration (IFS+NEMO+LIM3+PISCES+LPJ-GUESS)

N Proc	Time to solution (m)	Ideal time to solution (m)	Parallel efficiency (fraction)
192	361.8	361.8	1.00
768	119.4	90.4	0.76
1344	86.4	51.7	0.55

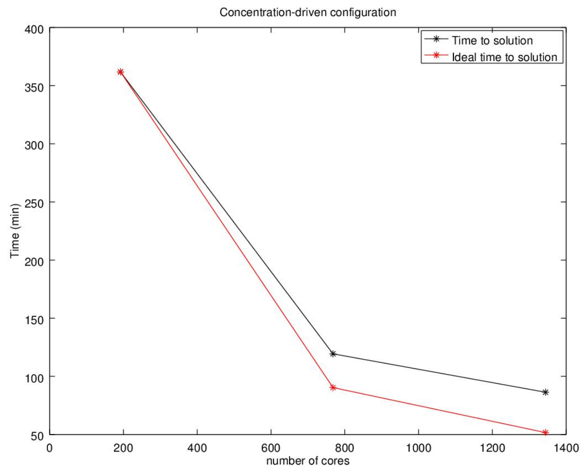


Figure 5: Scalability plot for the concentration-driven configuration.

Emission-driven configuration (IFS+NEMO+LIM3+PISCES+LPJ-GUESS+TM5)

N Proc	Time to solution (m)	Ideal time to solution (m)	Parallel efficiency (fraction)
576	247.5	247.5	1.00
624	214.8	228.5	1.06
720	168.8	198	1.17
768	168.8	185.6	1.10
816	170.6	174.7	1.02

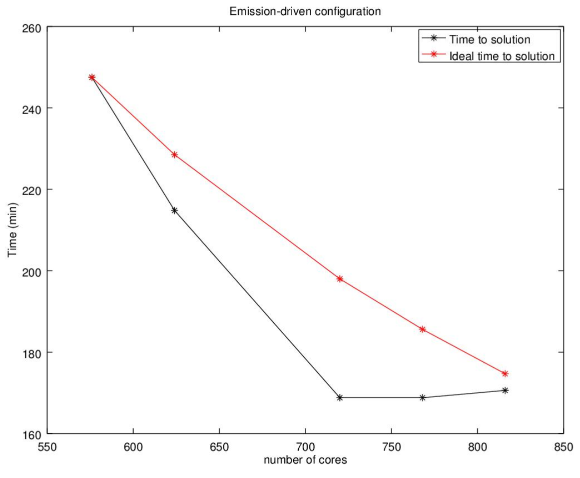


Figure 6: Scalability plot for the emission-driven configuration. The use of TM5 strongly limits the lower and upper bounds of processor configurations for IFS, which explains the unusual parallel efficiency numbers. The optimization of the code setup and the maximization of the throughput was carried out through other means (see Section 2.6.2).

2.6.2.2 Precision reported

All operations will be carried out with double-precision.

2.6.2.3 Time-to-solution

We will run experiments in two different configurations. Times to solution are given for both and summed.

Ti*_C1=1.99x768=1528.3

Ti*_C2=2.81x720=2023.2

Tf*_C1=1.99x768x8850=13525632

Tf*_C2=2.81x720x8850=17905320

2.6.2.4 System scale

Results were measured on full-scale system.

2.6.2.5 Measurement mechanism

Performance measurements were carried out through the usage of timers.

2.6.2.6 Memory usage

We will need full access to each occupied node's memory (96 Gb) for both configurations used.

2.6.2.7 OPTIONAL: Percentage of available peak performance

Not reported.

3 Milestones (quarterly basis) (Maximum 1 page)

Milestones:

M1- Completion of concentration-driven predictions in perfect model framework

M2- Completion of 5 reconstruction cases

M3- Completion of emission-driven predictions in perfect model framework

M4- Competion of retrospective decadal predictions

M5- Completion of future forecast of atmospheric CO_2.

3.1 Gantt Chart

Months since start	1	2	3	4	5	6	7	8	9	10	11	12
Task-1		M1			M3			EGU	P1			
Task-2			M2			OS						
Task-3									M4			P2
Task-4											M5	P2

3.2 Communication plan

At least two publications (P1 and P2 in Gantt chart) in international peer-reviewed journals will be prepared summarising the results relative to potential predictability and actual predictability. It is important that these publications are submitted soon after the results are obtained for the results to contribute to the activities of dissemination and knowledge transfer planned in WP4 of CCiCC. Our approach is to translate existing and new knowledge into a language that facilitates an effective dialogue with decision and policymakers. Several tools will be used to achieve an effective dialogue. We will initially prepare decision and policymakers by providing fact sheets on the core concepts to be investigated in CCiCC. Building on this, CCiCC findings will be translated into usable formats (e.g., executive summaries, policy brief), which also put the findings in context with other emerging literature. These targeted contents will be communicated in several formats (text, info-graphics, etc.). In addition, carbon outlooks will build on the annual release of the Global Carbon Budget to present annual and decadal forecasts developed in TOPSyCled.

The results will also be presented at two scientific conferences: Ocean Sciences (OS in the chart) in February 2020 and the European Geosciences Union general assembly (EGU in the chart) in April 2020. Results will also be submitted for presentation in the PRACEDays event of the years in which the project is active.

4 Personnel and Management Plan (~1/2 page)

Activities within TOPSyCled will be carried out by the PI together with a team of researchers already part of the BSC-ES. These are:

Dr. Etienne Tourigny who has a PhD in Meteorology from the Instituto Nacional de Pesquisas Espaciais (INPE-CPTEC, Brasil) and a M.Sc. in Atmospheric Science from the Université du Québec à Montréal (UQAM). Dr. Tourigny has a strong multi-disciplinary background, having studied physics, computer science, atmospheric science and biosphere-atmosphere interactions. After obtaining a Marie-Curie fellowship, Dr. Tourigny joined the climate prediction group at BSC where he is developing a new research line on seasonal predictions of wildfires while, at the same time, actively contributing to the development of the CMIP6 version of the EC-Earth3 ESM. Dr. Tourigny will contribute to the setup, execution and analyses of all experiments.

Dr. Valentina Sicardi who holds a Masters Degree in Environmental Sciences with majors in Oceanography and Meteorology from the Univerisita' Parthenope, Napoli, (Italy) and obtained a PhD degree from the University of Hamburg for her research at the Max Planck Institute for Biogeochemistry, Jena, (Germany). Since 2010 she has worked at the Earth Sciences Department of the Barcelona Supercomputing Center and at present she is conducting research on the optimal strategy to initialize the climate model EC-Earth for seasonal-to-decadal predictions while also contributing to the generation of initial conditions for biogeochemical predictions. Dr. Sicardi will contribute to the setup, execution and analyses of all experiments.

Pierre-Antoine Bretonnière who holds a Masters Degree in "Mathematical and Mechanical Modelling" from the Matmeca engineer school in Bordeaux (France). Graduated in 2010, he has worked in several climate research institutes (CERFACS -Toulouse - France, Catalan Institute of Climate Sciences - Barcelona - Spain and the Earth Sciences Department of the Barcelona Supercomputing Center). His work focuses on climate models outputs and diagnostics, data management and model coupling. Mr Bretonnière will coordinate the data management and will assist in the post-processsing of results.

Dr. Mario Acosta holds a PhD in Computer Science from the University of Granada and currently leads the performance team at the BSC-ES. He has an extensive experience in optimization of all components of EC-Earth and has been instrumental in the testing and development of the Earth

System Model configuration of EC-Earth. Dr. Acosta will assist in the coordination of the experiments.

5 Previous Allocations and Results (~1/2 page)

BSC-ES members are involved in a range of projects in the broad spectrum of climate sciences. These efforts have been supported by PRACE computing resources in previous calls. For instance, members of the department have led PRACE projects: "HiResClim", "HiResClim2", "HiResSIR", "LSIHP", "Glob15km" or "HiResNTCP" over the past 5 years. The PI of this proposal was awarded a project with the “Red Española de Supercomputación” for a total amount of 4.8 Mcore-hours (AECT-2019-1-0011). The project is still undergoing and results are being analyzed. Moreover, BSC-ES participates in one of the most interesting tasks of the Grand Challenge on Near-Term Climate Prediction: the provision of decadal predictions in real time for the elaboration of WMO’s Annual-to-Decadal Climate Outlook. Should the results of the experiments proposed in TOPSyCled suggest strong predictability for future atmospheric CO_2, the PI, together with CCiCC partners will propose to include this variable among those provided to WMO.

6 References (Maximum 30)

1-IPCC. in Climate Change 2013: The Physical Science Basis. Summary for Policymakers. (ed T.F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley) 1-29 (Cambridge University Press, 2013).

2-Boer, G. J., et al. Geosci. Model Dev., 9, 3751-3777, (2016).

3-Marotzke, J. & Forster, P. M. Nature 517, 565, (2015).

4-Conway, T. J. et al. Journal of Geophysical Research - Atmospheres 99, 22831-22855, (1994).

5-Keeling, C. D. et al. Nature 375, 666-670, (1995).

6-Peters, G. P. et al. Nature Climate Change, (2017).

7-Doblas-Reyes, F. J. et al. Nature Communications 4, 1715, (2013).

8-Wouters, B. et al. Geophys. Res. Lett. 40, 3080-3084, (2013).

9-Matei, D. et al. Science 335, 76-79, (2012).

10-Jones, C. D. et al. Geosci. Model Dev. 9, 2853-2880, (2016).

11-Boer, G. J. Climate Dynamics 36, 1119–1133, (2011).

12-García-Serrano, J. & Doblas-Reyes, F. J. Climate Dynamics 39, 2025-2040, (2012).

13-Friedlingstein, P. et al. J. Clim. 27, 511-526, (2014).

14-Landschützer, P. et al. Nature Climate Change 8, 146-150, (2018).

15-Tramontana, G. et al. Biogeosciences 13, 4291-4313, (2016).

16-Talley, L. D. et al. Annual Review of Marine Science 8, 185-215, (2016).

17-Olsen, A. et al. Earth Syst. Sci. Data 8, 297-323, (2016).

18-Luthcke, S. B. et al. Journal of Glaciology 59, 613-631, (2017).

19-Le Quéré, C. et al. Earth Syst. Sci. Data 10, 405-448, (2018).

20-Raupach, M. R. et al. Proceedings of the National Academy of Sciences of the United States of America 104, 10288-10293, (2007).

21-Séférian, R. et al. Proceedings of the National Academy of Sciences 111, 11646-11651, (2014).

22-Yeager, S. G. et al. Bulletin of the American Meteorological Society, (2018).

23-Li, H. et al. Nature Communications 7, 11076, (2016).

24-Séférian, R. et al. Geophys. Res. Lett. 45, 2455-2466, (2018).

25-Li et al., Sci. Adv. 2019; 5 : eaav6471

26-OPA 8 Ocean General Circulation Model - Reference Manual (G. Madec, P. Delecluse, M. Imbard, C. Levy), Technical report, LODYC/IPSL Note 11, 1998.

27-Vancoppenolle, M., T. et al. Ocean Modelling, 27, 33-53, (2009).

28-Smith, B., Prentice, I.C. & Sykes, M.T. Global Ecology & Biogeography 10, 621-637, (2001).

29-Huijnen, V. et al. Geosci. Model Dev., 3(2), 445-473, (2010).

30-Aumont, O. et al. Geosci. Model Dev. 8, (2015).

Appendix 1: Track Record of the PI

Curriculum vitae and list of publications of the Principal Investigator

Dr. Raffaele Bernardello (M) has a PhD in Oceanography from the Universitat Politecnica de Catalunya-BarcelonaTech. He is a senior researcher in the climate prediction group at BSC where he coordinates all the activities related to the global carbon cycle. He will be the principal investigator for BSC in H2020 project CCiCC (795k euros) and co-leader of WP2. His expertise and research interests are in the broad context of the interactions between climate dynamics and global carbon cycle. As part of his Marie-Curie fellowship, Dr. Bernardello worked on the assessment of the decadal predictability of biogeochemical properties in the upwelling systems of the Atlantic Ocean. He has participated to 3 national projects (Spain:OAMMS-CTM2008-03983 UK: BATMAN-NE/K015613/1 USA: NOAA-NA10OAR4320092), one FP6 project (SESAME-36949) and one ESA project (ENVISAT-A0290). At present, Dr. Bernardello supervises one postdoctoral researcher, one PhD student and he is the PI of a Spanish project (DeCUSO-CGL2017-84493-R; 114k euros) dedicated at investigating the decadal predictability of carbon uptake in the Southern Ocean, serving at the same time, as an external collaborator in the UK project CUSTARD with focus on Southern Ocean biogeochemical processes.

Selected publications:

Bernardello, R. *et al.* Factors controlling interannual variability of vertical organic matter export and phytoplankton bloom dynamics – a numerical case-study for the NW Mediterranean Sea. *Biogeosc.* **9,** 4233–4245 (2012).

Bernardello, R. *et al.* Impact of Weddell Sea deep convection on natural and anthropogenic carbon in a climate model. *Geophys. Res. Lett.* **41**, (2014b).

Bernardello, R. *et al.* Response of the Ocean Natural Carbon Storage to Projected Twenty-First-Century Climate Change. *J. Clim.* **27,** 2033–2053 (2014a).

Bernardello, R. *et al.* Using Preformed and Remineralized Nutrients to Map Spatial Variation in remineralization of organic particles. *Submitted to Glob. Biogeochem. Cycles.*

de Lavergne, C. et al. Cessation of deep convection in the open Southern Ocean under anthropogenic climate change. Nature Clim. Change, **4**, (2014).

Cabre, A., et al. Oxygen Minimum Zones in the tropical Pacific across CMIP5 models: mean state differences and climate change trends. Biogeosciences, accepted September 2015.

Granted patents and other measures for the relevance of the work

Prior allocation history in PRACE, national calls, as well as international programs such as INCITE of the US DoE

Dr. Bernardello was recently awarded a tier-1 project by the “Red Española de Supercomputación” to carry out simulations with the EC-Earth3-ESM as a contribution to the international activity C4MIP (part of CMIP6). Ref.: AECT-2019-1-0011 The Coupled Climate-Carbon Cycle Model Intercomparison Project (C4MIP). BSC contribution with the Earth System Model EC-Earth. Tot: 4.8 Mcore-hours.

Participation by team members in other European Commission (EC) actions, such as ERC or Marie Skłodowska-Curie EC grants, etc.

Dr. Bernardello and Dr. Tourigny have been Marie-Curie fellows within the BSC-ES.

Previous presentations at PRACEdays

Appendix 2: Resources from the European ICEI project (FENIX research infrastructure) – Pilot Phase

This 18th Call includes a pilot phase to incorporate a fraction of the resources from the Fenix Research Infrastructure, funded by the European ICEI project. Applicants interested in using the additional ICEI resources are requested to provide here the following information:

Amount of required ICEI resources

<Enter your text here>

Description of the software and services that are planned to be executed within the ICEI infrastructure

<Enter your text here>

Description of special needs, e.g. in terms of third-party software

<Enter your text here>