Project name	High-Resolution Ensemble Sea Ice Reanalyses (Hi-Res SIR)
Research field	PE10_3 Climatology and climate change

Project leader

Title	Dr
Last name	Guemas
First name	Virginie
Organisation name*	Barcelona Supercomputing Center (BSC)
Department*	Earth Sciences Department
Group*	Climate Prediction Group
Country	Spain

Short CV of the research team
(Section limit: 2000 characters)

The **climate prediction group** of the Barcelona Supercomputing Center Earth Sciences department (ES-BSC) undertakes advanced research to forecast climate variations from one month to several years into the future (also known as **seasonal-to-decadal predictions**) and from regional to global scales. This objective relies on expanding our understanding of the climate processes through a deep analysis of state-of-the-art climate forecast systems in comparison with the most up-to-date observational datasets, and on exploiting these detailed analyses to refine the representation of climate processes in our climate forecast systems and their initialization. Emphasis is made on forecasting changes in **high-impact climate events** such as the persistent winds, floods, droughts and temperature extremes.

Many of the activities in modelization and prediction are based on research, development and predictions with the EC-Earth climate forecast system. Besides contributing to the fifth phase of the Coupled Model Intercomparison Project (CMIP5), which is one of the key datasets used to produce the UN Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5), global climate research activities of this group enabled production of historical global climate reconstructions and initial conditions for the EC-Earth community. Such data is critical for analysis of climate dynamics and initialization of seasonal-to-decadal climate predictions.

The group is a highly productive scientific entity which published **nearly 100 research articles in peer-reviewed journals over the last 5 years**, including **3 in prestigious high-impact journals:**

Massonnet F et al (2016) Science, doi : 10.1126/science.aaf6369

Guemas V et al (2013) Nature Climate Change, 3, 649-653, doi : 10.1038/nclimate1863

Doblas-Reyes F et al (2013) Nature Communications, 4, 1715, doi:10.1038/ncomms2704

For a complete list of publications:
https://earth.bsc.es/wiki/doku.php?id=publications:publications

List up to 5 recent relevant publications

F. Massonnet, P. Mathiot, T. Fichefet, H. Goosse, C. König Beatty, M. Vancoppenolle, T. Lavergne, 2013, A model reconstruction of the Antarctic sea ice thickness and volume changes over 1980-2008 using data assimilation, Ocean Modelling, 64 67-75, doi:10.1016/j.ocemod.2013.01.003

F. Massonnet, H. Goosse, T. Fichefet, 2014, Prospects for better seasonal Arctic sea ice predictions from multivariate initialization, Ocean Modelling 88 16-25, doi:10.1016/j.ocemod.2014

Guemas V, Blanchard-Wrigglesworth E, Chevallier M, Day J J, Déqué M, Doblas-Reyes F J, Fučkar N, Germe A, Hawkins E, Keeley S, Koenigk T, Salas y Mélia D, Tietsche S, 2015, A review on Arctic sea ice predictability and prediction on seasonal-to-decadal timescales, Quarterly Journal of the Royal Meteorology Society, doi:10.1002/qj.2401.

Guemas V, Chevallier M, Deque M, Bellprat O, Doblas-Reyes F J, 2016, Impact of sea ice initialisation on sea ice and atmosphere prediction skill on seasonal timescales, Geophysical Research Letters, doi:10.1002/2015GL066626, 43 (8), 3889-3896.

Guemas V, Doblas-Reyes F J, Mogensen K, Keeley S. , Tang Y., 2014, Ensemble of sea ice initial conditions for interannual climate predictions. Climate Dynamics, 43(9-10), 2813-2829, doi:10.1007/s00382-014-2095-7.

1. ***Research project***. Include discussion of the scientific questions that you are planning to address and the overall scientific goals of the project, as well as the background and significance of the topic. It is important that you describe the novelty, impact and timeliness of the proposal.

   Please include any previous results of relevance to the project, to demonstrate how the proposal contributes to the long-term goals of the proposer. Previous allocations (PRACE, nationals, or internationals, including reference codes) and the research publications that resulted from them should also be listed.

   If applicable, explain and justify the need for multi-year access.

   (**Section limit: 2 pages**)

The dramatic decline in Arctic sea ice has been an emblematic sign of ongoing global climate change. Profound reductions in sea ice areal coverage (Cavalieri and Parkinson, 2012) and thickness (Rothrock et al., 2008), among others, have already had devastating impacts on local ecosystems (Tynan, 2015; Kovacs et al., 2011), indigenous populations (Meier et al., 2014) and possibly lower-latitude climate (Cohen et al., 2015). These rapid changes are also unlocking economic and industrial opportunities. Thinner, younger ice facilitates operations of icebreakers in the High North. Increased marine accessibility promotes polar shipping (Stephenson et al., 2013; Smith and Stephenson, 2013) as an economically viable alternative to existing commercial routes (Liu and Kronbak, 2010). Ecotourism, resources extraction and industrial fishing are other examples of activities that can take place in an open Arctic Ocean. Further reductions in the sea ice cover (Massonnet et al., 2012; Stroeve et al., 2012) are expected in the near future and predicting the first Arctic-free summer is one of the current challenges of the scientific community.

A better understanding of the interactions between the long-term externally forced climate trend and the internal variability, which is essential to accurately estimate the amplitude of upcoming sea ice losses, requires long and continuous monitoring of polar climate variables. Unfortunately, before 1973, Arctic sea-ice data are limited to monthly estimates of sea ice extent (SIE), with complete cover assumed within the ice pack and a necessary treatment of missing data in the marginal seas (Walsh and Johnson, 1979). The situation is worse in the Antarctic where sea ice data is limited to estimates of extent climatologies over two distinct periods: 1929-1939 (Deutsches Hydrographisches Institute, 1950) and 1947-1962 (Tolstikov, 1966). From 1973, the US Navy, Canadian and Danish aerial reconnaissance provided quasi-weekly estimates of sea ice concentration (SIC; Knight, 1984). The advent of satellite microwave imagery in 1978 allowed for the retrieval of SIC at roughly 25 km resolution and a 2 day frequency, which was increased to a daily frequency in 1987 (Cavalieri et al., 1996).

The publicly distributed datasets National Snow and Ice Data Center (NSIDC; Cavalieri et al., 1996) and OSI-SAF (Eastwood et al., 2015) sea ice concentration fields (1978-present) and HadISST (1870-present; Rayner et al. 2003) stand as the best estimates of sea ice concentration obtained from combination, homogenization and extrapolation of these sparse observational data. High-quality and high-resolution satellite sea ice concentration products have also been distributed recently by the European Space Agency and cover the 1993-2008 period (Ivanova et al., 2015). The sea ice thickness (SIT) data are much scarcer (Kwok and Rothrock, 2009): the first unified dataset (Lindsay, 2010) was released in 2010. It combines Arctic observations by submarines from 1975, moored upward-looking sonar from 1990 and airborne or satellite electromagnetic measurements from the past decade.

Filling the critical gaps in sea ice observations is essential for monitoring the long-term evolution of the sea ice system but also to provide a complete description of the sea ice state to initialize seasonal-to-decadal climate predictions. Indeed, the knowledge of the initial climate system state has been shown to be a major source of information in seasonal forecasts (Balmaseda and Anderson 2009) as well as in decadal climate predictions (Smith et al. 2007; Keenlyside et al. 2008; Pohlmann et al. 2009). In particular, the spring Arctic sea ice thickness distribution has been shown to be a precursor of the September sea ice cover (Chevallier and Salas-Melia 2011) in a model study. A summer-to-summer re-emergence mechanism has been suggested whose memory lies in the sea ice thickness (Blanchard-Wrigglesworth et al. 2011a). Wang et al. (2013) also find a dependence of the sea ice prediction skill on the initial sea ice thickness. A complete and coherent description of the sea-ice state can only obtained through a physical extrapolation of the sparse observations, relying on the equations that describe the sea-ice dynamics and thermodynamics. This is, in short, the goal of data assimilation.

The first attempt to use data assimilation to provide such a complete description of the sea-ice state was performed by Lisaeter et al. (2003) by assimilating SIC data in a coupled ice–ocean model every week over more than 1 year using an ensemble Kalman filter (Evensen, 1994). Sea-ice velocity is well observed in the recent decades, and its assimilation can be used to improve the simulated SIT (Zhang et al. 2003), by applying an optimal interpolation procedure to assimilate buoy motion and satellite ice-motion data in an ice–ocean model. Lindsay and Zhang (2006) combined the velocity assimilation scheme of Zhang et al. (2003) with a nonlinear nudging scheme for SIC that substantially corrects the SIC when large differences between the model and the observational data are seen, i.e., along the sea-ice edges. The three-dimensional variational assimilation technique was tested by Caya et al. (2010) on a regional ice–ocean model. They showed significant improvements over a simple nudging technique when combining multiple sources of observational data. The propagation of the analysis update from the observed fields to the unobserved fields, such as the SIT, remains an issue. Using the Ensemble Kalman Filter, the propagation of the SIC information is performed through a model error covariance matrix which is estimated thanks to an ensemble of sea ice integrations (Massonnet et al. 2013). This approach has two advantages: 1) the technique to propagate the information is fairly simple and allows for a propagation to a large number of sea ice variables, 2) an ensemble of reanalyses is generated which allows to sample the uncertainty on the sea ice state.

Up until now, this approach has only been applied at a typical resolution of about 1°. However, resolving mesoscale ocean eddies would allow for more realistic representation of the ice drift and deformation and, consequently, of the Arctic open water percentage (Zhang et al. 1999; Gent et al, 2010). A better representation of the ocean circulation and associated heat transport toward the Arctic at high resolution as well as a better representation of the western boundary currents/frontal areas and their impact on the generation of storms which can break the Artic sea ice and lead to its melting can produce a more realistic representation of Arctic sea ice processes (Delworth et al. 2012).

Here, we propose to generate a 25-member ensemble of sea ice reanalyses at a resolution never achieved up until now in a data assimilation context. Furthermore, the large ensemble size will provide a robust evaluation of the related uncertainty. This reanalysis will use the deterministic Ensemble Kalman Filter (Sakov and Oke, 2008) to assimilate sea ice data from the OSISAF and ESA products and will cover the 1978-present period, thus becoming the longest sea ice reanalysis available. The version 3 of the Louvain-La-Neuve sea Ice Model (LIM3, Vancopenolle et al. 2009) is embedded in the version 3.6 of the Nucleus for European Modelling of the Ocean (NEMO), which is the latest version incorporated in several coupled climate models participating to the next Coupled Model Intercomparison Project Phase 6 (CMIP6). The grid resolution will be of 0.25° globally (approximately 25 km) with 75 vertical levels, with the thickness increasing from 1m below surface up to 500m in the deep ocean. The Drakkar Forcing Set Version 5.2 will be used as atmospheric surface forcing with perturbations to generate the different members of the sea ice reanalysis (Guemas et al. 2014).

The results of the project will be widely presented in scientific conferences and will feature at the meetings of the Scientific Steering Group (SSG) CLIVAR (Climate and Ocean Variability Predictability and Change) project. These results will contribute to the INTAROS (understanding of polar processes), APPLICATE (linkages between the polar and mid-latitude climate), and PRIMAVERA (added-value from increasing climate model resolution) projects funded by the European Commission in which the applying team is participating.

The applicant has obtained computing hours in the context of a previous PRACE Tier-0 project entitled “Land-Surface Initialization in High-resolution seasonal Prediction (LSIHP)” (12th call, reference # 2015133087). The goal of this project was to initialise a climate model with the best set of land-surface observations currently available and analyse how this additional initialization impacted the skill level of seasonal forecasts. These experiments showed that a realistic initialization of soil moisture improves the skill over Europe, a result that is robust for the two resolutions that were tested. Also, it was shown that the strong heat wave of 2010 could be predicted well in advance when the soil moisture was initialized in the model, whereas without that initialization, it could not. However, the specific prediction of the 2003 heatwave did not seem to depend strongly on the quality of that initialization. Two peer-review publications resulted from this work:

Prodhomme, C et al (2016) Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. Clim Dyn, 47, 919-935, doi:10.1007/s00382-015-2879-4.

Ardilouze, C et al (2016) Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability. Clim Dyn, under review.

Both proposals align with the applicant's goal (and that of her research team) to deepen our understanding of the physical processes leading to skilful climate forecasts.

2. ***Resource management***. Describe how you will manage the resources requested? Use a Gantt chart or equivalent to illustrate this, including tasks and milestones (**mandatory**). Please, detail the contribution of each institution to each task. In case of multi-year access, please provide also information for the second and third year in the Gantt chart or equivalent.

   (**Section limit: 1 page**)

A single-member spin-up experiment will first be started from the EN4 climatology (Good et al. 2013) and an ocean at rest with a sea ice thickness of 1m in the Antarctic and 3m in the Arctic. This spin-up will be carried out with NEMO3.6-LIM3 forced by DFS5.3 and will cover the 1958-1977 period. Two experiments will then be initialized from this spin-up, both also run with NEMO3.6-LIM3 forced by DFS5.2 and covering 1978-present: a 5-member control experiment without any sea ice data assimilation (but perturbation in the atmospheric forcing) and a 25-member sea ice reanalysis assimilating sea ice concentration from OSISAF over the 1978-1992 and 2009-present periods and ESA over the 1993-2008 period. The ESA product will be chosen over OSISAF whenever available due to its high resolution, which will benefit our high-resolution reanalysis. While the minimum number of members to use in ensemble data assimilation is a topic of scientific debate, running 25 should be enough to obtain a good sampling and estimation of the model error covariance matrix. In order to reduce the necessary amount of computational resources, only 5 members will be run for the control experiment. These experiments will follow the schedule indicated in Table 1. Table 2 summarises the resources requested and detailed in section 6.

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Spin-up	■	■										
Control			■	■	■							
Reanalysis				■	■	■	■	■	■	■	■	■

Table 1: Approximate schedule of the experiments to be performed. M5 stands for Month 5.

	Simulations (years)	Total Core-hours (millions)	Total archive (TB)
Spin-up	20	0. 276	0.72
Control	195	1.797	7.02
Reanalysis	975	9.045	35.1
Total	1190	11.119	42.84

Table 2: Resources requested. These estimates have been obtained by running NEMO3.6-LIM3 on MareNostrum III.

The Autosubmit software (Manubens-Gil et al., 2016), briefly described below, will be used to manage the workflow and ensure a uniform and optimal use of the resources. Whereas the 25 members of the reanalysis need to run simultaneously since the ensemble Kalman Filter is applied to all members at the end of each month of simulation, the control simulation offer flexibility as it comprises 5 independent 39-year long simulations which can run in parallel for an optimal use of the PRACE computing resources. The jobs will be managed, and packed in groups in a single big job if required, by Autosubmit to optimize the use of the machine and avoid collapsing the I/O system. The data storage and data transfer can be organized with a disk space of 10 TB in the “scratch” file system. This required scratch space is motivated by the large amount of output to be generated. These output data will be transferred immediately locally. Around 500 GB of “home” space will be required to host the code and its modified versions.

3. ***Methodology***. Describe the numerical methods and algorithms that you are planning to use, improve, or develop, the codes, packages or libraries that you need to undertake the project, and how these will enable the research to be achieved.

   (**Section limit: 1 page**)

A single executable needs to be built for the coupled NEMO3.6-LIM3 model as well as an executable for the extended Ensemble Kalman Filter. The resolution proposed (ORCA025: ~1,475,000 grid points) will help efficiently share calculations between 1000-1500 sub-domains, increasing the range of efficient compute-core usage. NEMO fully supports a parallel environment. NEMO generates NetCDF output and, at the end of a simulation, both NEMO and LIM generate restarts separately (also in NetCDF format).

For configuring and building the model executables, GNU make 3.81 or 3.81+, FORTRAN 77/90/95 compliant compiler with pre-processing capabilities and NetCDF4 deployed with HDF5 and SZIP are needed. A newly designed tool for automatic build configuration called “ec-conf” can be used. This useful tool requires Python 2.4.3 or 2.4.3+ (it does not work yet with Python 3.0+). For NEMO, the FCM bash and perl mechanism is essential. 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 HDF5, NETCDF CDFTOOLS v2, CDO, NCO and general configurations tools, such as PERL, PYTHON, AUTOCONF and AUTOMAKE.

4. ***Justification of Tier-0 needs***. Explain why this project needs to run on a Tier-0 system, why the machine you have requested is suitable for the project and how the use of the system will enable the science proposed. You should describe the architecture, machine/system name and the problem sizes that have been used to test for scaling and provide supporting evidence.

   Provide both a table and scaling plot (mandatory) such as the ones shown below with example data to illustrate the information requested.

   Please include as well performance analysis and representative benchmarks with explicit comparisons with state of the art in terms of time to solution, scaling and accessible and sustainable percentage of peak in the requested architecture.

   (**Section limit: 3 pages**)

For this project, we will be running the most recent version of NEMO using the highest resolution configuration currently available. Running NEMO with such configuration requires a system with high-level resources, which, when combined with the large number of ensemble members, requires computational resources available only on a small subset of machines worldwide. NEMO has already been deployed on MareNostrum III (a tier-0 system with 3056 nodes, 2x Intel SandyBridge-EP E5-2670/1600 8-core processors at 2.6 GHz, 8x4GB DDR3-1600 DIMMS (2GB/core) per node and Infiniband FDR10 network) and used as part of EC-Earth within the previous PRACE projects “HiResClim”, “HiResClim2” and “LSIHP”.

The configuration mentioned above requires hundreds of gigabytes of total memory for the computing nodes where the memory will run. A fat-node having more than 32 gigabytes of main memory is also recommendable to make the input / output server work properly and write all the needed outputs at this resolution. These requirements are difficult to find, even more so when we consider our need to produce a large ensemble of model members.

As explained in section 2, our experiment will be divided in different parts. First, we will perform a 20-year spin-up simulation, whose output will then be used to initialize the ensemble of control experiments and the ensemble of reanalysis experiments. Although these control and ensemble experiments are independent, they both rely on the output of the spin-up run. The latter represents only a small portion of the overall computing resources, but given that the spin-up belongs to the critical path of the experiment (all subsequent simulations depends on this spin-up), it is desirable to increase the amount of per-member resources to speed it up.

From the scalability-test performed on the model (Table 3 (right) and Figure 1 (red)), we know that it is possible to increase the model throughput by increasing the number of cores up to more than 2k, even if doing so also raises the cost of the simulations. As it is shown in Table 3, it is possible to use 1536 processes (or even more) with reasonable efficiency resulting in the reduction of the spin-up wall-clock time by a factor greater than two.

After the spin-up, both the control and reanalysis experiments can be performed independently, which means that 30 simulations should be executed in parallel to save time. The model can run at an acceptable efficiency using 1024 cores per member, so this part of the experiment could fill a 30k core machine during up to 20 days. It would also be possible to run the experiments using half the number of cores per member (i.e. 512). Although this set-up is slightly more efficient, it would require more wallclock time to produce the final output.

This strategy based in using fast-configurations for critical sections and throughput-configurations for other parts of the workflow when multiple independent instances can run concurrently is typical in climate simulations and has been described in the literature (Balaji et al. 2016).

All the information presented here comes from several performance analyses of the NEMO model. This type of analysis is used at BSC to drive model improvement. For example, a recent optimization, targeting inter-process communication, a critical issue in large supercomputers, had led us to improve the NEMO model throughput and to increase its efficiency by ~30%. The demand for high-resolution climate simulations has led us to invest much effort in this type of performance analyses, an effort that is necessary to optimize the usage of the computing resources made available to us. By actively researching how to improve the model's efficiency and scalability, we are not only able to take full advantage of modern supercomputers, but also to produce frontier research.

Number of Cores	Before-optimization (SYPD)	After-Optimization (SYPD)
256	1.36	1.34
512	2.17	2.33
768	2.80	3.09
1024	3.07	3.69
1280	3.27	4.20
1536	3.37	4.52
1792	3.53	4.80
2048	3.38	4.93

Table 3: NEMO 3.6 throughput in Simulated Years per Day (SYPD) while running ORCA025-LIM3 configuration with different number of processor-cores for the version without optimization and version including optimizations.

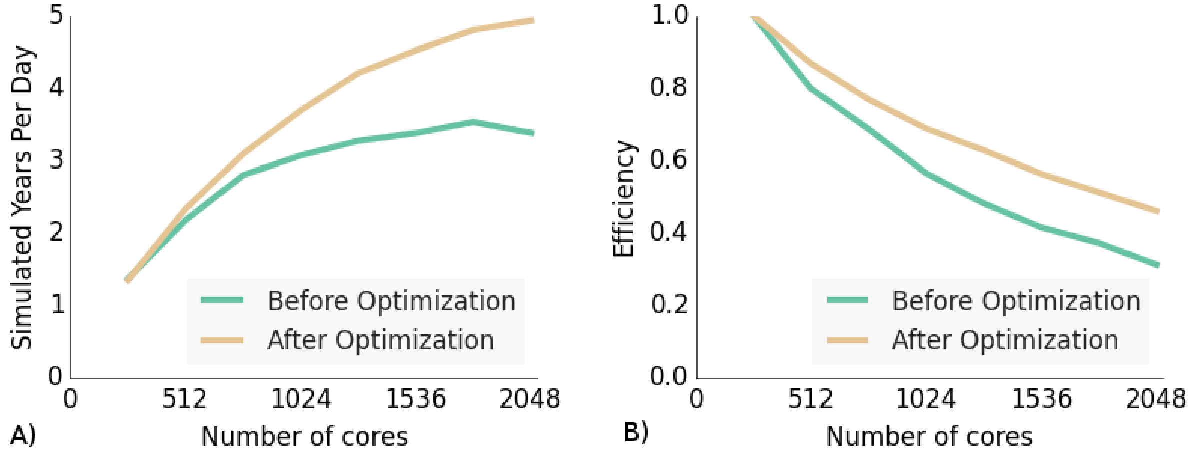

Figure 1: Simulated years per day (A) and relative efficiency (B) of NEMO 3.6 ORCA025-LIM3 simulations using different number of cores. Before/after comparison of the BSC optimization of intra-process communications.

5. ***Past Experience***. Describe your experience of using HPC resources in the past and how you will manage using a Tier-0 system. What other experience do you and your team bring to this project?
   (**Section limit: 1 page**)

The BSC Earth Sciences Department has been involved in the IS-ENES/PRACE-1IP working group focusing on the EC-Earth3 (which includes NEMO) adaptation to Tier-0 machines. It has tested a range of EC-Earth3 configurations, with atmospheric resolutions T255/511/799 on several HPC systems: SGI Altix 3500, NEC-SX6, Linux cluster with Intel Xeon, Dell PowerEdge 2900, IBM pSeries 575 Power6 and IBM Power PC. Finally, members of the BSC Earth Sciences Department coordinated the recent HiResClim, HiResClim2 and LSHIP projects supported by PRACE.

The BSC Earth Sciences Department has the privilege to rely on a team of experts on parallel model's performance, which specializes in analyzing parallel programming model codes using cutting-edge performance tools. This advantage in resources and expertise allows the department to always have the latest performance results and therefore be able to determine the optimum configurations for NEMO. Currently, the BSC Earth Sciences Department is a collaborator of the NEMO development team and a member of the NEMO HPC working group, providing the Consortium with performance reports and code optimizations. As a result of this collaboration, two optimization branches were created in the NEMO code repository by the BSC Earth Sciences Department (O. Tintó et al. 2015); both of them were merged in the NEMO 3.6 stable version and trunk, and constitute an indispensable basis for future developments. The impact of the implemented optimizations can be seen in table 3, and Figure 1, reaching up to a 40% improvement on the model throughput and significantly improving its efficiency.

Figure 2 illustrates an example of the performance tool's output. This is the picture provided by the Paraver tool from one single NEMO model execution. This image displays the communications pattern as a function of time (horizontal axis), while the vertical axis corresponds to the different processes executing the model. The different colours correspond to different MPI communication functions, except for the light blue which corresponds to no communication. This Paraver view is very useful to determine the communications within the model and to explain where the possible bottlenecks originate from. In this configuration, NEMO uses 32 processors. The image clearly shows the importance of the communications in NEMO, which for a large portion of the time is not performing effective computations. It also shows that a big portion of this time is devoted to global communications, which are painted in pink colour. Those collective communications belonged to the horizontal diffusion routine, inside the ice model (LIM) used in NEMO and was solved to a large extent by our targeted optimizations.

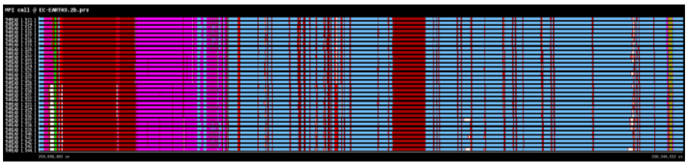

Figure 2: MPI call Paraver view of the NEMO component model in an EC-Earth's model execution. The horizontal lines give the behaviour of the different processes (513 to 544) as a function of time. Each colour corresponds to one different MPI communication function.

6. ***Justification of the amount of resources requested***. This should include information such as: run type, wall clock time per step, number of jobs per run type, the number of CPU cores and the total core hours per run type. This information should take the form of a table like the one shown below with example data (**mandatory**). Explain how the core hours requested will be used. In case of Multi-year access, please provide an estimation of the core hours requested for the second and third years in the table.

   (**Section limit: 1 page**)

Table 4 lists the experiments described in section 2, which will run by chunks of 1 month-length, i.e. 240 months/chunks of simulations for the spin-up and 468 months/chunks for the control and the reanalysis (multiplied by 5 and 25 members, respectively). The control and the reanalysis can all be run simultaneously, depending on the machine load, by Autosubmit. These experiments use the highest-resolution version of NEMO3.6-LIM3 configuration. The benchmarking exercise performed through internal queues suggests that, taking into account the average load of the MareNostrum III queues, the optimum performance for throughput is obtained using 512 procs, which requires a wall-clock time of 1.5 hours for a one-month simulation. For the spin-up run, which is the bottleneck of the entire experiment, the number of procs used will be 1536, which in this case requires a wall-clock time of 0.75 hours for one month of simulation. Finally, extrapolating the cost of the extended Kalman Filter at standard resolution, we estimate we will need 256 processors for 30min to perform the assimilation step which occurs at the end of each month of reanalysis (only once per 25 members). This last configuration generates 3GB of outputs per month of simulation.

Run type	# Runs	# Steps/Run	Walltime/Step	# CPU cores / Step	Total core hours
Spin-up	1	240	0.75	1536	276,480
Control	5	468	1.5	512	1,797,120
Reanalysis forecast	25	468	1.5	512	8,985,600
Reanalysis assimilation	1	468	0.5	256	59.904
Total					11,119,104

Table 4: Cost of the experiments proposed.

The final estimate is for a total request of 11.5 million core-hours, which includes the numbers described in Table 4 plus a small buffer of 5% to account for failing jobs that will need to be repeated.

The experiments will be run using Autosubmit, the launching and monitoring solution developed by the group of the applicant that allows the remote submission of NEMO experiments. Autosubmit includes in the workflow of the experiments a job that retrieves the data back to the Department data storage as soon as each chunk of simulation has completed. This means that the estimates for the archive are an absolute upper value in case the automatic download does not perform as expected. It is very likely that this number will approach a figure ten times smaller.

7. ***Data management plan***. Please indicate the data volume for input, output, and usage of parallel I/O tools. Include how long data needs to be held after grant period, needs to transfer to other sites, and other relevant information.
   (**Section limit: 1 page**)

   The type of simulations conducted during this project requires hosting a set of perturbed atmospheric forcings in order to maintain the ensemble spread. The Drakkar Forcing Set 5.2 consists in eight surface atmospheric variables: four at daily frequency (185 Mo each for one year of data) and four at 3-hr frequency (1.5 Go each for one year of data). The original forcing set has to be perturbed 25 times to create the ensemble. Hence the total disk space required to host the atmospheric forcing amounts to 2.5 To for a 1958-2015 period.

   Thanks to a well-organized workflow structure, the data produced by the integrations and the data assimilation will be downloaded to the local BSC-ES servers within 24 hours after completion. Hence, a standard two-week delay between the end of the project and the closing of the accounts is more than enough to let us clean the HPC repositories.

8. ***Additional needs***. Please indicate any additional need, like e.g.: visualisation, pre- and post-processing, etc.
   (**Section limit: 1 page**)

   The use of NCO, CDO, TOTALVIEW and NCVIEW is required to allow de-bugging in case a problem arises on the HPC.

9. ***Reviewers***. (optional) Please indicate **up to three** potential competitors that should be excluded as reviewers (name, affiliation and contact).

*10. **References***

Asif, M, A Cencerrado, O Mula-Valls, D Manubens, FJ Doblas-Reyes and A Cortés (2014). *Procedia Computer Science*, 29, 2370-2379, doi:10.1016/j.procs.2014.05.221.

Balaji, V, E Maisonnave, N Zadeh, BN Lawrence, J Biercamp, U Fladrich, G Wright (2016) *Geoscientific Model Development Discussions*, 1-25. doi:10.5194/gmd-2016-197

Balmaseda MA, D Anderson (2009) *Geophys Res Let* 36:L0170. doi:10.1029/2008GL035561

Blanchard-Wrigglesworth E, KC Armour, CM Bitz, E DeWeaver (2011) *J Clim* 24:231–250

Cavalieri DC, C Parkinson, P Gloersen, HJ Zwally (1996) National Snow and Ice Data Center. Digital media, Boulder

Cavalieri DJ, CL Parkinson (2012) *The Cryosphere*, 6, 881-889, doi:10.5194/tc-6-881-2012

Caya A, M Buehner, T Carrieres (2010) *J. Atmos. Oceanic Technol.* 27: 353–369, doi:10.1175/2009JTECHO701.1.

Chevallier M, D Salas-Melia (2011) *J Clim* 25:3025–3038. doi:10.1175/JCLI-D-11-00209.1

Cohen J, J Screen, JC Furtado, M Barlow, D Whittleston, D Coumou, J Francis, K Dethloff, D Entekhabi, J Overland, J Jones (2014) *Nature Geoscience*, 7 627-637, doi:10.1038/ngeo2234

Delworth, Thomas L, and Coauthors (2012) *J. Clim*, 25, 2755–2781.

Deutsches Hydrographisches Institute (1950) Atlas of ice conditions in the north atlantic ocean and general charts of the ice conditions of the north and south polar regions. Publ 2335, Hamburg

Eastwood, S, M Jenssen, T Lavergne, A Sorensen, and R Tonboe: EUMETSAT Ocean and Sea Ice Satelitte Application Facility. Global sea ice concentration reprocessing product (v1.2), Product user manual, Technical Report, Norwegian and Danish Meteorological Institutes, Oslo, Norway and Copenhagen, Denmark, 2015. 5526, 552.

Evensen G. (1994) *J Geophys Res* 99: 10143–10162, doi: 10.1029/94JC00572.

Gent, PR, SG Yeager, RB Neale, S Levis and DA Bailey (2010) *Clim Dyn*, 34, 819–833, doi:10.1007/s00382-009-0614-8.

Good, SA, MJ Martin and NA Rayner (2013) *Journal of Geophysical Research: Oceans*, 118, 6704-6716, doi:10.1002/2013JC009067

Guemas V, FJ Doblas-Reyes, K Mogensen, S Keeley, Y Tang (2014) *Clim Dyn*, 43(9-10), 2813-2829, doi:10.1007/s00382-014-2095-7.

Ivanova, N., et al. (2015) *The Cryosphere*, 9, 1797-1817. doi: 10.5194/tc-9-1797-2015.

Keenlyside NS, M Latif, J Jungclaus, L Kornblueth, E Roeckner (2008) *Nature* 453:84–88. doi:10.1038/nature06921

Knight, RW (1984) *Ann Glaciol* 5:81–84

Kovacs KM, C Lydersen C, JE Overland, SE Moore (2011) *Marine Biodiversity*, 41, 181-194. Doi: 10.1007/s12526-010-0061-0.

Kwok R, DA Rothrock (2009) *Geophys Res Lett* 36: L15501, doi: 10.1029/2009gl039035.

Lindsay RW (2010) *EOS Trans AGU* 91(44):405–416

Lisaeter KA, J Rosanova, G Evensen (2003) *Ocean Dyn* 53: 368–388. doi: 10.1007/s10236-003-0049-4.

Liu M, J Kronbak (2010) *Journal of Transport Geography*,18, p. 434-444.

Manubens-Gil, D., Vegas-Regidor, J., Prodhomme, C., Mula-Valls, O., & Doblas-Reyes, F. J. (2016). 2016 International Conference on High Performance Computing & Simulation (HPCS). doi:10.1109/hpcsim.2016.7568429

Massonnet F, T Fichefet, H Goosse, CM Bitz, G Philippon-Berthier, MM Holland, PY Barriat (2012) The Cryosphere, 6, 1383-1394.

Massonnet, F, P Mathiot, T Fichefet, H Goosse, C König Beatty, M Vancoppenolle, T Lavergne (2013) *Ocean Modelling*, 64 67-75, doi:10.1016/j.ocemod.2013.01.003

Meier WN, GK Hoverlsrud, BEH van Oort, JR Key, KM Kovacs, C Michel, C Haas, MA Granskog, S Gerland, D Perovich, A Akshtas, JD Reist (2014) *Review of Geophysics*, 52(3) 185-217, doi:10.1002/2013RG000431

Pohlmann H, JH Jungclaus, A Kohl, D Stammer, J Marotzke (2009) *J Clim* 22:3926–3938

Rayner NA, DE Parker, EB Horton, CK Folland, LV Alexander, DP Rowell, EC Kent, A Kaplan (2003) *J Geophys Res* 108:D14. doi:10.1029/2002JD002670

Rothrock DA, DB Percival, M Wensnahan (2008) *J Geophys Res*, C05003, doi:10.1029/2007JC004252.

Sakov P, P Oke (2008) *Tellus A*, doi:10.1111/j.1600-0870.2007.00299.x

Smith D, A Cusack, A Colman, C Folland, G Harris, JM Murphy (2007) *Science* 317:796–799. doi:10.1126/science.1139540

Smith LC and SR Stephenson (2013) *Proceedings of the National Academy of Sciences Plus*, 110, E1191-E1195, doi:10.1073/pnas.1214212110

Stephenson SR, LW Brigham, LC Smith (2013) *Polar Geography*, doi: 10.1080/1088937X.2013.845859

Stroeve J, V Kattsov, A Barret, M Serreze, T Pavlova, M Holland, W Meier (2012) *Geophys Res Lett*, 39, L16502, doi:10.1029/2012GL052676

Tintó Prims, O, M Castrillo, K Serradell , O Mula-Valls and FJ Doblas-Reyes (2015) BSC-CES Technical Memorandum 2015-002.

Tolstikov EI (1966) Atlas of the Antarctic. Glav Uprav Geod Kart MGSSSR, Moscow, vol 1, 126 pp

Tynan E (2015) *Nature Climate Change*, 5, 621, doi:10.1038/nclimate2708

Vancopenolle, M, T Fichefet, H Goosse, S Bouillon, G Madec and MAM Maqueda (2009) *Ocean Modelling*, 27(1):33–53.

Walsh JE, CM Johnson (1978*) J Phys Oceanogr* 9(3):580–591

Wang W, M Chen, A Kumar (2013) *Mon Weather Rev.* doi:10.1175/MWR-D-12-00057.1

Zhang, Y, W Maslowski and AJ Semtner (1999) *J Geophys Res*, 104, 18409-18429, doi:10.1029/1999JC900158.

Zhang J, D Thomas, DA Rothrock, RW Lindsay, Y Yu, R Kwok (2003) *J Geophys Res*, 108: 3170, doi: 10.1029/2001JC001041.