RED ESPAÑOLA DE SUPERCOMPUTACIÓN

Application Id: AECT-2016-1-0014

1.- Activity Title: Ensemble sea ice data assimilation for polar prediction
Area: Astronomy, Space and Earth Sciences

2.- Research Project Description:

a) Is this a Test Activity? No

b) Is this a Long Term Activity that will extend over several application periods? Yes

c) Brief description of the Project:
This Activity takes place in the framework of several national and international projects (all ongoing or funded): PICA-ICE and Juan de la Cierva Formación, funded by the Ministry of Economy and Competitiveness (MINECO); SPECS, funded by the European Commission under the framework of the FP7 program; the Polar Prediction Project, the Year of Polar Prediction and the Southern Ocean Region Panel, which are international initiatives in which the applicant is deeply involved.

PICA-ICE (Previsión Interanual de la Cubierta de hielo marino del Ártico y su Impacto en el Clima de Europa ; Jan. 2013 - Dec. 2015; lead by V. Guemas) was a first step in generating better sea ice conditions for Arctic prediction and beyond. Recognizing the need to refine this area of research, the applicant François Massonnet decided to pursue the efforts developed in PICA-ICE and obtained the competitive Juan de la Cierva Formación funding (Nov. 2015-Oct. 2017). His research project is now entirely devoted to implement and maintain advanced sea ice data assimilation methods for sea ice initialization, presented below in detail, and for which extensive computational resources are required.

The SPECS project (Seasonal-to-decadal climate Prediction for the improvement of European Climate Services, Nov. 2012-Jan. 2017; lead by F. Doblas-Reyes) aims at advancing regional prediction capabilities and prepare the new generation of forecast systems in Europe. SPECS considers at the same time practical aspects of the climate prediction by e.g. bringing together all the actors involved in the operational prediction (including scientists and stakeholders). It is also raising important scientific questions, such as the impact of sea ice initialization on seasonal polar and extra-polar predictions.

Finally, the Polar Prediction Project, the Year of Polar Prediction and the Southern Ocean Region Panel are three initiatives of the World Weather and Climate Research Programmes. The applicant François Massonnet is member of these organizations, in which data assimilation stands as a top priority for state estimation and initial condition generation.

All aforementioned projects have in common the desire to better understand the role of, and improve the representation of, sea ice initial conditions in state-of-the art climate models. Recent results obtained from the applying team (Guemas et al., 2014; Massonnet et al., 2015b) and other international groups (Chevallier and Salas-Mélia, 2012) have stressed the potential importance of sea ice initial conditions for predictability. It is now time to devise initialization systems able to incorporate a growing number of observations in climate models with increasing resolution, in order to generate top-quality sea ice initial conditions for polar climate predictions. This is the objective of the present project, that largely rests on the use of high-performance computing.

Grant References:
GA 308378CGL2012-31987

d) Brief description of the Project (If this Activity takes place in the context of a Technology or Industrial Project):
Not applicable

e) Specific Activity proposed:
At high latitudes, a significant part of the oceans is covered by sea ice that forms when seawater freezes. Sea ice is a major player in the global climate system. It is also a primary indicator of climate change, as it responds rapidly to variations in surface air temperatures and wind regimes. The past decades have been the scene of two contrasting stories at the two ends our planet. In the Arctic, dramatic decline in sea ice area (Cavalieri and Parkison, 2012) and thickness (Rothrock et al., 2008) have sparked scientific and public awareness about the possible consequences of a “blue Arctic Ocean” on ecosystems (Kovacs et al., 2011), local populations (Meier et al., 2014), as well as the potential benefits of these changes for the shipping industry and economy (Liu and Kronbak, 2010; Smith and Stephenson, 2013) and possible impacts on the climate in lower-latitude countries (e.g. Cohen et al., 2014). By contrast, the Antarctic sea ice cover has slightly expanded since the late 1970s (Parkinson and Cavalieri, 2012) and even reached record high levels last year (Massonnet et al., 2015a). The root causes for this increase in a warming world remain elusive.

The sparseness of reliable sea ice thickness measurements explains in great part the current difficulty to predict the evolution of Arctic sea ice even a few months in advance (Stroeve et al., 2015). A great deal of the memory of summer Arctic sea ice indeed resides in its spring thickness (Chevallier et al., 2012). The lack of multi-decadal thickness records is also a problem for identifying the origins of past Antarctic sea ice changes. Sea ice thickness is indeed the missing piece that would allow to close the sea ice mass budget (sea ice areal coverage is easily observed from space).

In either case, models and remote-sensed observations share the secrets of this puzzle. Climate models have the well-known advantage to simulate all possible variables at high temporal frequency and fine spatial resolution, as well as the dynamical links between them. Notwithstanding these advantages, models are prone to systematic biases and their use in isolation is generally not sufficient. Observations are in this sense complementary to climate models: they are incomplete snapshots of the real world. How to combine these two sources of information (and their respective uncertainties) as to produce an optimal synthesis? This is the objective of “data assimilation” (DA hereafter), a field of research that has received increased interest in the domain of geosciences and climate prediction, owing to its robust mathematical formulation and demonstrated ability to resolve practical problems.

Past studies have demonstrated the relevance and appropriateness of DA in the field of polar sciences and, in particular, sea ice research (Lisaeter et al., 2003; Mathiot et al., 2012; Massonnet et al., 2013; 2014; 2015b; Barth et al., 2015). One critical aspect systematically highlighted by these studies is the need to accurately estimate the “model error covariance matrix”. The correct specification of this matrix allows to transfer information from the observable variables (e.g., sea ice concentration) to unobservable variables (e.g., sea ice thickness). A common way to correctly estimate this matrix while retaining its spatial- and time-dependence is to run “ensemble” simulations: a number (usually 30 at least) of replicae of the same model are run in parallel, but each one is subject to a different perturbation. This sample helps, at any given time, to characterize the statistical structure of model error. Owing to the increased computational cost implied by this approach, only regional, stand-alone or coarse versions (~2°) of these models are currently run in DA mode.

Polar climate research thus faces a dilemma: how to better estimate poorly-observed variables such as sea ice thickness using effective DA methods while taking full advantage offered by the latest generation of climate models at high resolution? This equation can only be resolved by considering the use of high-performance computing. Meeting such a timely challenge can be achieved through this project covering three successive 4-month periods: (1) DA in standard resolution forced runs (~1°), (2) DA in standard resolution coupled runs (~1°), and (3) DA in high-resolution coupled runs (~0.25°). We largely document below our planned procedure for the first four months (period 1) but also present our longer-term strategy for consistency.

Period 1: DA in standard resolution forced runs.
Task 1.1: Reference simulation (up to month 0.5)
The Nucleus for European Modelling of the Ocean (NEMO) consortium has recently released its reference configuration that will form the ocean-sea ice basis of several climate models participating to the next Coupled Model Intercomparison

Project (CMIP6), including the state-of-the-art climate model EC-Earth3. The Climate Prediction Group (CPG) of the Barcelona Supercomputing Center Earth Science Department (BSC-ES) is currently calibrating EC-Earth3 in standalone configuration using prescribed atmospheric forcing (Drakkar Forcing Set 5.2; Dussin et al., 2014). A control run (CTRL hereafter) will be conducted over 1979-2015 at standard (ORCA1, ~1°) resolution during the first two weeks of Period 1 and will be the reference for any later simulation in this Activity.

Task 1.2 Calibration of the sea ice data assimilation method (up to month 1.5)

The ensemble Kalman filter (EnKF; Evensen, 2003) will be downloaded from the Nansen Environmental and Remote Senting Center (NERSC) repository (http://enkf.nersc.no) with which the CPG has established firm relationships. The main investigator of the present project (F. Massonnet) has a long experience working with this filter and collaborators at NERSC, so that its compilation and installation on local HPC facilities should be a matter of days. Next, the daily sea ice concentration products from the European Space Agency (ESA) will be assimilated into the model over the year 2007. This year was the first one of a series of extremely low sea ice extents in the Arctic and is therefore a good case to test the DA method (e.g., Massonnet et al., 2015b). Three sensitivity experiments will be conducted varying the time interval between data assimilation steps (from one day to one month), the strength of perturbations in the atmospheric forcing used to generate ensemble spread, and the minimum size of the ensemble required to preserve essential relationships between sea ice concentration and thickness and maintain sufficient ensemble spread. The optimal configuration will be chosen in order to ensure computational efficiency while keeping a statistically-consistent ensemble throughout the year. It is indeed important that the EnKF remains a tractable tool, given the use that will be made of it in subsequent tasks.

Task 1.3 Sea ice reanalysis (up to month 4)

A full 1979-2015, global sea ice reanalysis will be conducted following the procedure described in Task 1.2. Only sea ice concentration will be assimilated but, if time permits, the additional assimilation of passive microwave sea ice thickness products will be considered over the last decade, depending on the availability of such products at that time. The CPG has established contacts with the Institut de Ciències del Mar (Barcelona) that is currently processing the Soil Moisture and Ocean Salinity (SMOS) products. Our reanalysis will be assessed against independent in-situ Arctic and Antarctic sea ice thickness data (Rothrock et al., 2008; Worby et al., 2008) and compared to the widely-used PIOMAS sea ice reanalysis (Schweiger et al., 2011). Our data will be made publicly available in order to profit the whole community of polar researchers in quest of better constrained estimates of sea ice thickness and volume in both hemispheres. This will place Barcelona, and Spain, in the short list of institutions proposing global sea ice reanalyses.

Period 2 (a detailed implementation plan will be submitted if the Period 1 is granted).

During the second period of 4 months, the EnKF will be implemented in the newest version of the EC-Earth GCM at standard resolution (just released) that will have been tested and installed on MareNostrum3 by then. The coupled reanalysis will be run over 2005-2015 and its suitability for initializing seasonal sea ice prediction will be assessed in comparison to existing methods. This will be the first time that such a comprehensive sea ice data assimilation will be implemented in a coupled operational system, making EC-Earth one of the reference tool for polar prediction.

Period 3 (a detailed implementation plan will be submitted if Period 2 is granted).

The final objective of the Long Term Activity is to have the CPG equipped with a resilient data assimilation system for sea ice initialization in the high-resolution (~1/4° in the ocean, ~30 km in the atmosphere) version of EC-Earth. While it is difficult to anticipate practical issues such as storage or RAM allocation that could arise when running ensembles of high-resolution simulations, we plan to perform initial test cases to examine the feasibility of DA in high-resolution. The full accomplishment of a high-resolution reanalysis will probably require many more months of simulations.

Sea ice data assimilation is an extremely challenging objective that only a few centers in the world can afford to tackle. The Barcelona Supercomputing Center is one of them, thanks to the unique alchemy of technical and scientific expertises, as well as the large computational resources available. Considering sea ice data assimilation for climate prediction is not an option: it is a necessary step to keep the CPG and Spain at the forefront of polar research for the next five years.

References:
https://docs.google.com/document/d/1RTye4lURrqJyQr0dvi5rx_3Rvbgu0DXG22fhz7MvuWA/edit?usp=sharing

f) Computational algorithms and codes outline:

EC-Earth3 comprises three major components: IFS, NEMO and OASIS3. For this first phase only NEMO will be considered.

For configuring and building the model executable, GNU make 3.81 or 3.81+, FORTRAN 77/90/95 complaint compiler with preprocessing 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+ (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 are needed for IFS. To test the model with the run scripts, GNU date (64-bit) is also required.

The standard configuration will be ORCA1L75 (~1° resolution with 75 vertical levels) which has already been tested on various super-computing platforms, Marenostrum3 among them. A benchmarking performed previously suggests that, taking into account the average load of the Marenostrum queues, optimum performance is obtained when using 48 procs. These configurations generate 0.2 GB output per month of simulation. The high resolution will be ORCA025L75 (~0.25° resolution with 75 vertical levels) which has been tested and validated by the applicant on Marenostrum3. Benchmarking reveals that an optimal use of the machine is made with 512 procs. About 1 GB output per month of simulation is created.

Experiments in this project involve the use of large ensembles (~30 members). Each member of the ensemble is running on 48 processors, and one could think, a priori, to use local clusters to achieve this goal. However, this would be untractable because local clusters at our disposal have a limited number of CPUs so that only two or three members could run at the same time, which would lengthen the experiment to a year at least (!). At the same time, the memory resources demanded by this project are not possible in a small cluster. 8000-10000 Gigas of Scratch space is difficult to achieve in a small cluster and the RAM memory needed during the execution is larger than the usual RAM memory found in a local cluster (usually 6-12 Gigas per node).

3.- Software and Numerical Libraries

Software components that the project team requires for the activity

Applications + Libraries:

BLAS
FFTW
HDF5
LAPACK
NETCDF
R
SCALAPACK
OPENMPI
UDUNITS
NCO
SZIP
MKL

Compilers and Development Tools:

GCC
TOTALVIEW
INTEL
MVAPICH2

POE
GDB

Utilities + Parallell Debuggers and Performance Analysis Tools:
CMAKE
PERL
PYTHON
VALGRIND
GNUPLOT
NCVIEW
NCL
AUTOCONF
GREASY
PARAVER

Other requested software:
GRIB_API, GRIBEX, CDO, CDFTOOLS

Propietary software:
Not applicable

4.- Research Team Description

a)Team Leader:
Name: François Massonnet
Institution: Earth Sciences Department, Barcelona Supercomputing Center
E-mail: francois.massonnet@bsc.es
Phone: +34 638 136 518
The employment contract of the activity leader with the research organisation is valid at least 3 months after the end of the allocation period: Yes

Curriculum Vitae of the Team Leader:
Dr Massonnet obtained the Juan de la Cierva Formación of the Spanish Ministry of Economy and Competitiveness (MINECO). His research lines are seasonal-to-decadal climate predictability and prediction, the role of sea ice in the climate system, data assimilation, climate model evaluation and high-to-mid- latitude teleconnections.

He is a member of the WCRP/CLIVAR/CliC/SCAR Southern Ocean Region Panel. He has also contributed to the IPCC (Fifth Assessment Report), has participated in the EU FP7 COMBINE Project (2009-2013) and is involved in the H2020 project PRIMAVERA. He is (co-)author of 20 articles/chapters published in international peer-reviewed journals and authored more than 30 presentations or posters in national or international conferences. Three of his achievements were prized, and he has been invited four times in international conferences or workshop to present his work.

During his PhD, he developed expertise in three topics. First, he developed an extensive set of metrics to measure the performance of Arctic and Antarctic sea ice model simulations. Second, he investigated and discussed the importance of initial conditions, model physics and parameters in climate simulations of sea ice, giving clear directions to the community as to where efforts should be placed in the future. Finally, he implemented a sophisticated data assimilation scheme (the EnKF) in a state-of-the-art ocean-sea ice model at low resolution. He proposed an original method to calibrate sea ice model parameters, as well as reconstructing changes of Antarctic sea ice volume.

Dr Massonnet is the contact point for sea ice-related aspects of EC-Earth3. He is also at the forefront of sea ice prediction activities as member of the Sea Ice Prediction Network (SIPN). He has recently published an original study in the special issue of the Bulletin of the American Meteorological Society (BAMS) to explain the record sea ice extent in Antarctic in September 2014. This study was emphasized at BSC (https://www.bsc.es/about-bsc/press/bsc-in-the-media/%C2%BFpor-qu%C3%A9-crece-el-ant%C3%A1rtico-mientras-el-%C3%A1rtico-est%C3%A1-perdiendo-hielo) and in the media (BBC, El Periódico, El Mundo) (http://www.bbc.com/mundo/noticias/2015/11/151109_hielo_antartica_artico_lp; http://www.elperiodico.com/es/noticias/sociedad/por-que-crece-hielo-marino-antartida-vientos-bsc-4649073; http://elmundo.sv/por-que-crece-el-antartico-mientras-el-artico-esta-perdiendo-hielo/)

URL: www.climate.be/u/fmasson

d) Names of other researchers involved in this activity:

François Massonnet (francois.massonnet@bsc.es)
Neven Fuckar (neven.fuckar@bsc.es)
Virginie Guemas (virginie.guemas@bsc.es)
Domingo Manubens (domingo.manubens@bsc.es)
Muhammad Asif (muhammad.asif@bsc.es)
Danila Volpi (danila.volpi@bsc.es)
Chloé Prodhomme (chloe.prodhomme@bsc.es)
Omar Bellprat (Omar.Bellprat@bsc.es)
Martin Ménégoz (martin.menegoz@bsc.es)
Mario Acosta (mario.acosta@bsc.es)
Eleftheria Exarchou (eleftheria.exarchou@bsc.es)
Francisco Doblas-Reyes (francisco.doblas-reyes@bsc.es)

e) The five most relevant publications, in the last five years,from the members of the research team:

Guemas V, Doblas-Reyes FJ, Andreu-Burillo I, Asif M (2013), Retrospective prediction of the global warming slowdown in the past decade, Nature Climate Change, 3, 649-653, doi : 10.1038/nclimate1863.

Massonnet F, Goosse H, Fichefet T (2014), Prospects for better seasonal Arctic sea ice predictions from multivariate initialization, Ocean Modelling 88 16-25, doi:10.1016/j.ocemod.2014.12.013

Massonnet F Guemas V, Fuckar NS, Doblas-Reyes FJ (2015), The 2014 all-time high record of Antarctic sea ice extent, *invited* contribution to the Bulletin of American Meteorological Society Special Issue "Climate Extremes of 2014", doi:10.1175/BAMS-D-15-00093.1

Guemas V, Blanchard-Wrigglesworth E, Chevallier M, Day JJ, Déqué M, Doblas-Reyes FJ, Fuckar N, Germe A, Hawkins E, Keeley S, Koenigk T, Salas y Mélia D, Tietsche S (2014). A review on Arctic sea ice predictability and prediction on seasonal-to-decadal timescales. Quarterly Journal of the Royal Meteorological Society, doi:10.1002/qj.2401

Doblas-Reyes FJ, Andreu-Burillo I, Chikamoto Y, García-Serrano J, Guemas V, Kimoto M, Mochizuki T, Rodriguez LRL, van Oldenborgh GJ (2013), Initialized near-term regional climate change prediction, Nature communications 4, 1715, doi:10.1038/ncomms2704

f) Does any of the researchers involved in the activity apply to the mobility program? No

5.- Resources

a) Estimated resources required for the Activity fot the current Application Period:

Requested machine: MareNostrum 3
Interprocess communication: Null

Typical job run:

Number of processors needed for each job: 48.00
Estimated number of jobs to submit: 49200.00
Average job durations (hours) per job: 0.20
Total memory used by the job (GBytes): 15.00

Largest job run:
Number of processors needed for each job: 48.00
Estimated number of jobs to submit: 285.00
Average job durations (hours) per job: 2.00
Total memory used by the job (GBytes): 150.00

Total disk space (Gigabytes):
Minimun: 1000.00
Desirable: 2000.00

Total scratch space (Gigabytes):
Minimun: 8000.00
Desirable: 10000.00

Total tape space (Gigabytes):
Minimun: 0.00
Desirable: 0.00

Total Requested time: (Thousands of hours): 500.00

INFORMATION: The estimated cost of the requested hours, considering only the electricity cost, is 5715 euros

The required resources have to be executed in the selected machines, the other architectures do not fit the requirements to execute the proposal (this option implies that if no hours in this machine/these machines are available, the acces committee will reject the full application).

b) Estimate of the total resources that the Activity will require until it is completed (including the present and all the following Application Periods):

Number of application periods expected to complete this Activity: 3
Total Requested Time (thousands of hours) expected to complete this Activity: 2500.00

Abstract for publication:
The specification of initial conditions is crucial in the field of seasonal climate prediction. However, in remote regions such as the Poles, observations are sparse. How to optimally combine sparse observations with model output is a research topic, called "data assimilation".

This long-term activity aims at implementing and testing the ensemble Kalman filter (a data assimilation method) in the European climate model EC-Earth3 for better estimation of sea ice state. Using ensembles of simulations guarantees a realistic approximation of the model error covariance matrix, so that observational information is consistently transferred in all parts of the model.

Improvement in initial conditions is expected to greatly improve prediction skill in the Arctic, the Antarctic and lower latitudes, with benefits for the climate services sector.