REQUEST FOR A SPECIAL PROJECT 2022–2024

MEMBER STATE: SPAIN

Principal Investigator[1]: Etienne Tourigny

Affiliation: Barcelona Supercomputing Center – Earth Sciences Department

Address: Plaça Eusebi Güell, 1-3, 08034 Barcelona (Spain)

Other researchers: Aude Carreric, Pablo Ortega, Miguel Castrillo, Pierre-Antoine Bretonnière, Eric Ferrer, Roberto Bilbao, Juan Acosta Navaro, Vladimir Lapin

Project Title: High-resolution ocean reconstructions for initializing decadal climate predictions

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SPESICCF**_____
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2022
Would you accept support for 1 year only, if necessary?	YES ☒ NO ☐

Computer resources required for 2022-2024: (To make changes to an existing project please submit an amended version of the original form.)	**2022**	**2023**	**2024**
High Performance Computing Facility (SBU)	63,6M	0	0
Accumulated data storage (total archive volume)[2] (GB)	20k		

---

[1] The Principal Investigator will act as contact person for this Special Project and, in particular, will be asked to register the project, provide annual progress reports of the project's activities, etc.

[2] These figures refer to data archived in ECFS and MARS. If e.g. you archive x GB in year one and y GB in year two and don't delete anything you need to request x + y GB for the second project year etc.

Principal Investigator: Etienne Tourigny

Project Title: High-resolution ocean reconstructions for initializing decadal climate predictions

Extended abstract

I.   Scientific plan

Decadal climate prediction (DCP), which includes timescales from a year up to a decade, is a research field of growing interest due to its key role for decision-making purposes. This interest spans many sectors, including agriculture and food security, health, energy, water management and disaster risk reduction, highlighted in the context of climate change.

A fundamental aspect in DCP is its prediction skill. A prediction is only useful if it provides information beyond what we expect from a mean season/year. Climate predictability arises from both externally forced influences (natural and anthropogenic sources) and from internally generated variability (i.e., slow processes in the climate system, usually involving the ocean). The premise of DCP is that such internal variability processes, when adequately modelled and initialized, can improve our predictive ability (Meehl et al., 2014, 2021). A prominent example is the North Atlantic region where surface temperature has been found to be skilfully predicted several years in advance (Smith et al., 2019) but where the neighbouring continents show poor skill. This poor skill might be due to the coarse model resolutions used to make the predictions, (typically involving grid spacings of ~100 km), which fail to represent the key dynamical and physical processes.

Using an eddy-permitting configuration of the climate model will enable the representation of previously unresolved interactions from ocean eddies, which could improve the representation of key teleconnection mechanisms (Mahajan et al., 2018). We can therefore expect a higher prediction skill, in particular for extreme events. We plan to produce a DCP system with the latest version of the coupled model EC-Earth3.3 (Döscher et al., 2021) in its high-resolution (HR) configuration (Haarsma et al., 2020 ; of approximately 25km). This version has been appropriately tuned to reduce the model biases and improve process-representation in the Equatorial Pacific and the North Atlantic, two key regions for DCP skill. One particularly interesting feature of this HR model version is the improvement in the simulation of the deep convection in the Labrador Sea and the Atlantic Meridional Overturning Circulation (AMOC) compared to the standard resolution version (SR, of approximately 100 kms in both the atmosphere and the ocean). Indeed, a decadal prediction system based on EC-Earth3.3-SR has been shown to experience a consistent collapse of the Labrador Sea convection (Bilbao et al., 2021). This collapse is caused by an initialization shock and induces a quick degradation of the predictive skill in the Subpolar North Atlantic, a source region of decadal variability and predictability (Smith et al., 2020). Some preliminary tests show that this problem is not present in EC-Earth3.3-HR, for which the Labrador Sea convection remains active and stable, positively impacting the strength of the AMOC.

The tuning of EC-Earth3.3 at HR and the production of a single-member ocean reconstruction used to initialize HR seasonal predictions (from 1981-2020 only, whereas decadal prediction hindcasts will be done from 1960-2020) have been realized within the framework of the "Near-term Climate Prediction at High Resolution" (spesiccf-2020) ECMWF special project, and will allow the production of a HR seasonal forecast system during the second year of the project. For more details refer to the spesiccf-2020 2021 progress report submitted to ECMWF.

Additional improvements of this future DCP system can be achieved through the initialization procedure and in particular with the in-house reconstructions used to initialize it. We developed a

new refined strategy, adjusted to make the most of the new state-of-the-art ocean and atmosphere reanalyses from ECMWF (i.e., ORAS5 and ERA5), both provided at high resolution (approximately 25km). We improved our protocol to produce these oceanic conditions in such a way that avoids the non-stationary bias reported in the North Atlantic subsurface for the ECMWF seasonal forecast system SEAS5, as a consequence of initialising the ocean with ORAS5 (Tietsche et al., 2020). This new setup has been developed and successfully tested with the SR version of the model via a series of reconstructions with the ocean-sea ice standalone SR version of EC-Earth forced with ERA5 surface fluxes, each one following a different assimilation approach. The tests included using nudging coefficients of different strength, disabling the nudging at specific regions and/or depths, and combining ORAS5 with other products. The best strategy was selected by performing a series of reduced seasonal predictions, each one initialised from one of the tested reconstructions, and computing the associated skill. The best performing approach resulted from combining the assimilation, through nudging of ORAS5 temperature and salinity at the surface and of EN4 temperature and salinity below the mixed layer. As an illustration, Figure 1 shows the skill for predicting ocean temperature and salinity at 500m, for a subset of seasonal experiments, in the region where the non-stationary bias in SEAS5 was reported. The selected strategy (dark green line) shows persistently high correlation values in the region.

Thus, the oceanic and sea ice initial conditions of the DCP system will come from HR ocean-sea ice-only simulations with NEMO3.6 (the corresponding ocean-sea ice components of EC-Earth3.3 which will be used for the HR DCP system), driven by ERA5 surface atmospheric forcings (e.g, radiation, precipitation, temperature) and assimilating sea temperature and salinity at the surface from the ORAS5 ocean reanalysis and 3D ocean temperature and salinity below the mixed layer from the EN4 ocean reanalysis.
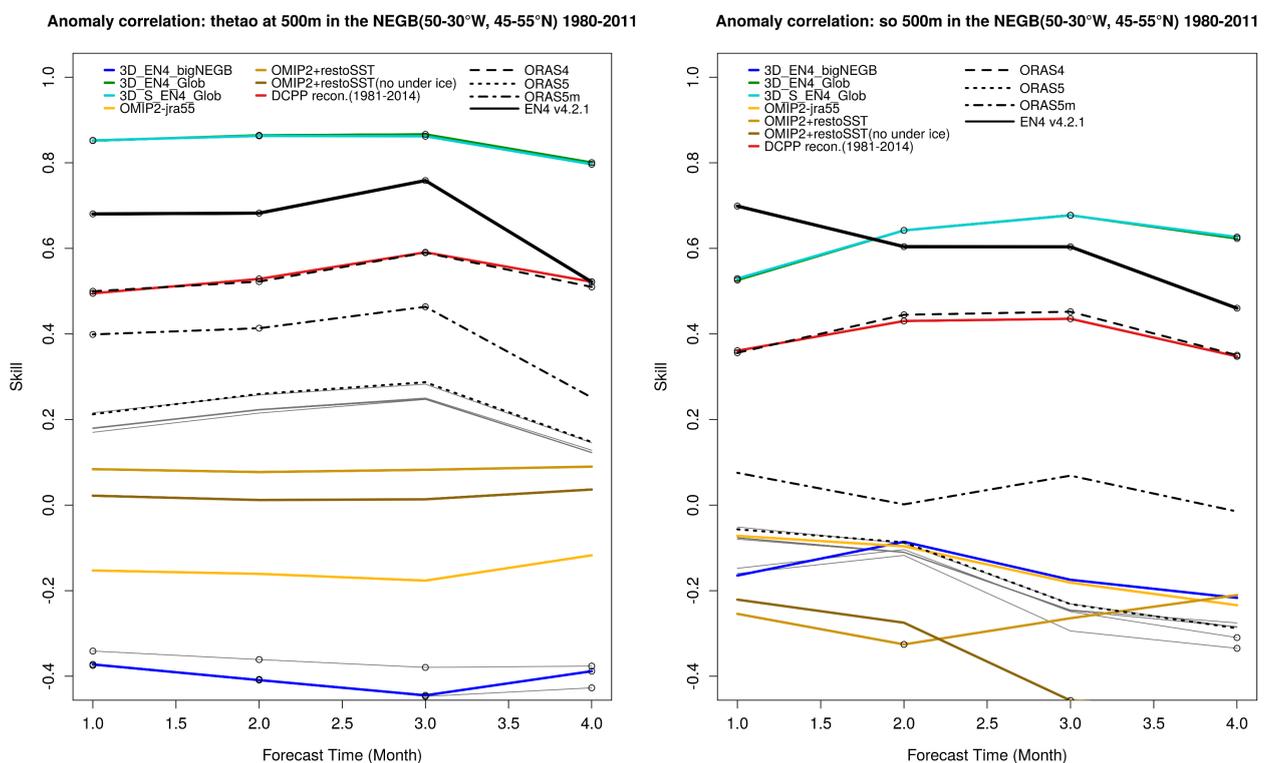
Figure 1: Anomaly correlation coefficient at each forecast month (November through February) of the ocean temperature (left) and salinity (right) in a region North East of the Great Banks in which ORAS5 presents a non-stationary biases, for a selection of seasonal prediction systems initialised following different strategies in the ocean. The skill is evaluated against observations from the Ishii et al. (2005) subsurface temperature and salinity analysis. All predictions are initialised on the 1st of November, have an ensemble size of 5 members and cover the reforecast period 1980-2011.

We plan to produce five different members of this HR ocean-sea ice only simulation (historical reconstructions), covering the period 1959-2020. The ensemble members will be generated by forcing NEMO3.6 with perturbed atmospheric forcings (following the method of Massonet et al., 2015) and by initializing each member using the December 31st conditions of each of the last five years of a HR NEMO3.6-LIM3 spin-up. This spin-up will be run with constant 1959 ERA5 forcing and assimilation of ocean observations, and for a period of 30 years to allow for the model to equilibrate. The HR reconstructions will be subsequently used to initialize the HR decadal hindcasts, which are computationally expensive and will therefore require further study and special attention, beyond the scope of this project.

This project is part of the recent efforts made in the community to produce decadal predictions at higher resolution. Coordinated experiments are being produced to evaluate the predictive skill of DCP systems, and enable multi-model comparisons. The BSC-ES is particularly involved in these joint activities, including the Decadal Climate Prediction Project (DCPP, Boer et al., 2016), an initiative contributing to CMIP6 and that will inform the next Intergovernmental Panel on Climate Change (IPCC) assessment report, the European Climate Prediction (EUCP) H2020, project and acting as an official Global Producing Center of Near-Term Climate Prediction designated by the World Meteorological Organization (WMO) (https://decadal.bsc.es). The HR simulations will contribute to the EUCP project, whose purpose is to develop an ensemble climate prediction system based on a new generation of improved and higher-resolution climate models, designed to support practical and strategic decision-making on adaptation to and mitigation of global warming. These HR experiments are computationally challenging, but their production is expected to bring decadal prediction into a new level.

This special project fosters collaboration with European research centers (mainly the EC-Earth consortium partners collaborating on climate prediction, such as SMHI, DMI, CNR/ISAC) which could benefit from the improved ocean reconstructions used to initialize their own decadal/seasonal predictions. The topic is also of high relevance for ECMWF activities which include ocean reanalyses and seasonal prediction, and will foster further collaboration between BSC, ECMWF and the EC-Earth consortium. This work will also be beneficial for the development of EC-Earth4, which is based on more recent IFS developments, also leading to enhanced collaboration.

II.   Justification of the computing resources requested

The oceanic reconstructions that will be used to initialize the decadal prediction system will be run with a tuned version of NEMO3.6 in its high-resolution configuration ORCA025L75 (described in section III).

Several tests have already been performed within the framework of the spesiccf project in order to evaluate the performance of the model, evaluating different metrics such as speedup and efficiency. The results of these preliminary tests show that optimum performance is obtained when using 1210/36 cores for the NEMO/XIOS components respectively for the HR atmosphere-forced ocean-only experiments. This configuration provides a compromise between the required amount of resources and the computational cost, taking into account the scalability of the parallel application and the average load of the platform. Using this configuration, one year of simulation requires ~170,000 SBUs and the throughput is approximately 3.2 SYPD (simulated years per day).

The different members of the ocean reconstruction will cover the period 1959-2020, for a total of 62 years. A fixed forcing ocean reconstruction will be run for 30 years and used to initialise the 5 members from the December initial conditions of its last 5 years. Our simulations offer flexibility as

they comprise 5 independent 62-year long simulations which can run in parallel for an optimal use of the computing resources which we estimate to:

62 years * 5 members * 170,000 SBUs/year = 52,7M SBUs

30 years * 1 member * 170,000 SBUs/year = 5,1M SBUs

This final estimate of 63,6 million SBU includes a buffer of 10% to account for output post-processing and repeated jobs due to hardware and software (numerical instabilities) failures and data management issues on the HPC platform.

The experiments will be run using the Autosubmit (Manubens-Gil et al., 2016) workflow manager developed by the BSC Computational Earth Science (CES) group, allowing the remote submission of EC-Earth and NEMO experiments. The Autosubmit workflow includes tasks to send processed model output to the BSC data storage and clean the HPC storage space within reasonable time limits. In practice, the actual storage space in the "scratch" file system that is required corresponds only to 3 years of simulation for each simulation, including the necessary restarts, i.e. about 2.1 TiB for each experiment. Knowing that we can run the different members concurrently, the total ECFS storage required is about 10.5 TiB, in case we are not able to transfer in due time the output to our local HPC. We therefore request a total tape storage of 20 TiB to accommodate for other files we need to store on our ECFS space, but would find 10 TiB acceptable. Around 500 GiB of disk space will be required to host the code and its modified versions.

The EC-Earth3.3 model and Autosubmit workflow have been developed and optimized for the Marenostrum4 HPC at BSC and Cray XC40 (cca, ccb) HPCF at ECMWF. The high usage of Marenostrum4 incurs long queuing times to run these costly reconstructions, which also involve long production times due to the need of running them sequentially in time. We request ECMWF computing resources to conduct the reconstructions, speeding up their production process. In order to avoid risks associated with delays in porting the model on the new ATOS (Atos Sequana XH2000) HPCF at ECMWF, as well as delays in the availability of the ATOS HPCF for actual production, we plan on running these experiments on the current Cray XC40 platform during the first months of 2022. We will also begin the porting of the model to the AMD architecture on the TEMS/ATOS HPCF until May 2022 when the ATOS HPCF is scheduled to replace the Cray XC40 for production. Therefore, should the production of part of the reconstructions be delayed beyond the availability of XC40, we will complete them on the ATOS HPCF. The HR decadal predictions initialized from these reconstructions, too costly for a ECMWF special project, will be performed on the Marenostrum4 supercomputer, combining both internal resources and competitive projects of supercomputing access to the Spanish supercomputing network and PRACE.

III.   Technical characteristics of the code to be used

The oceanic reconstructions that will be used to initialize the decadal prediction system will be run with a tuned version of NEMO3.6 in its high-resolution configuration ORCA025L75. This tuned version has been set from a tuning exercise of the coupled model EC-Earth3.3 in its high-resolution configuration, using NEMO3.6 and IFS (ORCA025L75-T511L91). The future decadal prediction system that will be initialised with these different ocean-only reconstructions will also be run with this latest version of the EC-Earth coupled model. The ORCA025L75 grid of NEMO3.6 has a resolution of 0.25° globally (approximately 25 km) with 75 vertical levels whose thickness increases from 1m below surface up to 500m in the deep ocean.

NEMO is an ocean model, developed by the European consortium and used for oceanographic research, operational oceanography, seasonal forecasting and climate research studies, that represents a large community with more than 1,000 registered users. It is based on the Navier-Stokes equations. The core of the NEMO model is OPA, a primitive equation model adapted to regional and global

ocean circulation problems down to kilometric scale. In the horizontal direction, the model uses a curvilinear orthogonal grid and in the vertical direction, a full or partial step z-coordinate, or s-coordinate, or a mixture of the two.

NEMO fully supports a parallel environment. For configuring and building the model executable, GNU make 3.81 or 3.81+, FORTRAN 77/90/95 compliant compiler and C++ 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" will 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, FCM, bash and perl are essential. GNU date (64-bit) is also required for executing the model with the run scripts.

As the experiment will have a complicated workflow in certain phases, the Autosubmit software will be used to manage the workflow and ensure a uniform and optimal use of the resources. 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. The hardware that best fits the needs of the model is made of nodes with a general-purpose core. Due to lack of source code adaptations, any usage with an accelerator or any other computing device would not take full advantage of these resources.

IV.   References

Bilbao, R., S. Wild, P. Ortega, J. Acosta-Navarro, T. Arsouze, P.-A. Bretonnière, L.-P. Caron, M. Castrillo, R. Cruz-García, I. Cvijanovic, F.J. Doblas-Reyes, M. Donat, E. Dutra, P. Echevarría, A.-C. Ho, S. Loosveldt-Tomas, E. Moreno-Chamarro, N. Pérez-Zanon, A. Ramos, Y. Ruprich-Robert, V. Sicardi, E. Tourigny and J. Vegas-Regidor (2021). Assessment of a full-field initialised decadal climate prediction system with the CMIP6 version of EC-Earth. Earth System Dynamics, 12, 173-196, doi:10.5194/esd-2020-66.

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., Mueller, W. A., Taylor, K. E., Zwiers, F., Rixen, M., Ruprich-Robert, Y., and Eade, R. (2016) The Decadal Climate Prediction Project (DCPP) contribution to CMIP6, Geosci. Model Dev., 9, 3751–3777, https://doi.org/10.5194/gmd-9-3751-2016.

Döscher, R., M. Acosta, A. Alessandri, P. Anthoni, A. Arneth,T. Arsouze, T. Bergmann, R. Bernadello, S. Bousetta, L.-P. Caron, G. Carver, M. Castrillo, F. Catalano, I. Cvijanovic, P. Davini, E. Dekker, F.J. Doblas-Reyes, D. Docquier, P. Echevarria, U. Fladrich, R. Fuentes-Franco, M. Gröger, J. v. Hardenberg, J. Hieronymus, M.P. Karami, J.-P. Keskinen, T. Koenigk, R. Makkonen, F. Massonnet, M. Ménégoz, P.A. Miller, E. Moreno-Chamarro, L. Nieradzik, T. van Noije, P. Nolan, D. O'Donnell, P. Ollinaho, G. van den Oord, P. Ortega, O. Tintó Prims, A. Ramos, T. Reerink, C. Rousset, Y. Ruprich-Robert, P. Le Sager, T. Schmith, R. Schrödner, F. Serva, V. Sicardi, M. Sloth Madsen, B. Smith, T. Tian, E. Tourigny, P. Uotila, M. Vancoppenolle, S. Wang, D. Wårlind, U. Willén, K. Wyser, S. Yang, X. Yepes-Arbós and Q. Zhang (2021). The EC-Earth3 Earth system model for the Climate Model Intercomparison Project 6. Geoscientific Model Development Discussion, doi:10.5194/gmd-2020-446.

Haarsma, R., M. Acosta, R. Bakhshi, P.-A. Bretonnière, L.-P. Caron, M. Castrillo, S. Corti, P. Davini, E. Exarchou, F. Fabiano, U. Fladrich, R. Fuentes Franco, J. García-Serrano, J. von Hardenberg, T. Koenigk, X. Levine, V. Meccia, T. van Noije, G. van den Oord, F.M. Palmeiro, M. Rodrigo, Y. Ruprich-Robert, P. Le Sager, E. Tourigny, S. Wang, M. van Weele and K. Wyser (2020). HighResMIP versions of EC-Earth: EC-Earth3P and EC-Earth3P-HR. Description, model

performance, data handling and validation. Geoscientific Model Development, 13, 3507-3527, doi:10.5194/gmd-2019-350.

Ishii, M., M. Kimoto, K. Sakamoto, and S. Iwasaki (2005) Subsurface Temperature And Salinity Analyses. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. https://doi.org/10.5065/Y6CR-KW66.

Mahajan, S., Evans, K. J., Branstetter, M. L., & Tang, Q. (2018). Model resolution sensitivity of the simulation of North Atlantic Oscillation teleconnections to precipitation extremes. *Journal of Geophysical Research: Atmospheres*, 123, 11,392– 11,409. https://doi.org/10.1029/2018JD028594

Manubens-Gil, D., J. Vegas-Regidor, C. Prodhomme, O. Mula-Valls and F.J. Doblas-Reyes (2016). Seamless management of ensemble climate prediction experiments on HPC platforms. Proceedings of the 2016 International Conference on High Performance Computing & Simulation (HPCS), Innsbruck, 895-900, doi:10.1109/HPCSim.2016.7568429.

Massonnet, F., Fichefet, T., & Goosse, H. (2015). Prospects for improved seasonal Arctic sea ice predictions from multivariate data assimilation. Ocean Modelling, 88, 16-25.

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., Corti, S., Danabasoglu, G., Doblas-Reyes, F., Hawkins, E., Karspeck, A., Kimoto, M., Kumar, A., Matei, D., Mignot, J., Msadek, R., Navarra, A., Pohlmann, H., Rienecker, M., Rosati, T., Schneider, E., Smith, D., Sutton, R., Teng, H., van Oldenborgh, G. J., Vecchi, G., & Yeager, S. (2014). Decadal Climate Prediction: An Update from the Trenches, *Bulletin of the American Meteorological Society*, *95*(2), 243-267, doi:10.1175/BAMS-D-12-00241.1

Meehl, G.A., Richter, J.H., Teng, H. *et al.* Initialized Earth System prediction from subseasonal to decadal timescales. *Nat Rev Earth Environ* **2,** 340–357 (2021). https://doi.org/10.1038/s43017-021-00155-x

Smith, D.M., Eade, R., Scaife, A.A. *et al.* Robust skill of decadal climate predictions. *npj Clim Atmos Sci* **2,** 13 (2019). https://doi.org/10.1038/s41612-019-0071-y

Smith, D.M., Scaife, A.A., Eade, R. *et al.* North Atlantic climate far more predictable than models imply. *Nature* **583,** 796–800 (2020). https://doi.org/10.1038/s41586-020-2525-0

Tietsche, S., M. Balmaseda, H. Zuo, C. Roberts, M. Mayer and L. Ferranti (2020) The importance of North Atlantic Ocean transports for seasonal forecasts. Clim Dyn 55, 1995–2011. https://doi.org/10.1007/s00382-020-05364-6