AGENCIA ESTATAL DE INVESTIGACIÓN - Convocatorias 2018
Proyectos de I+D de GENERACIÓN DE CONOCIMIENTO y Proyectosde I+D+i RETOS INVESTIGACIÓN

AVISO IMPORTANTE - La memoria no podrá exceder de 20 páginas.Para rellenar correctamente esta memoria, lea detenidamente las instrucciones disponibles en la web de la convocatoria.Es obligatorio rellenarla en inglés si se solicita más de 100.000,00 €.

IMPORTANT – The research proposal cannot exceed 20 pages. Instructions to fill this document are available in the website.If the project cost exceeds 100.000,00€, this document must be filled in English.

IP 1 (Nombre y apeliidos):Mario César Acosta Cobos

IP 2 (Nombre y apellidos):

TÍTULO DEL PROYECTO (ACRÓNIMO): Optimización de Modelos del sistema Terrestre en el camino a la nueva generación de Sistemas "Exascale" de alto rendimiento. (OEMES)

TITLE OF THE PROJECT (ACRONYM): Optimization of Earth system Models in the path to the new generation of Exascale high performance computing Systems (OEMES)

1. PROPUESTA CIENTÍFICA - *SCIENTIFIC PROPOSAL*

1. Introduction and state-of-the-art

In the last decade, our understanding of climate change has increased (Merchant et al, 2017), as the societal has need for pulling through to advice and policy. However, whilst there is a great confidence in the fact that climate change is happening, there remain uncertainties (Doblas-Reyes et al, 2013). For example, the levels of greenhouse gas emissions and aerosols likely to be emitted, or perhaps even more significantly uncertainties on the degree of warming and the likely impacts (Mudryk et al, 2017). Other example in Figure 1 (Flato et al, 2013) shows three main sources of uncertainty (global temperature) in projections of the fifth phase of the Coupled Model Intercomparison Project (CMIP5): due to future emissions (scenario uncertainty, green), due to internal climate variability (orange), and due to inter-model differences (blue). The figure shows that the spread between RCP scenarios is the dominant source of uncertainty at the end of the century, but internal variability and inter-model uncertainty are more important for the near-term.Increasing the capability and comprehensiveness of 'whole Earth system' models, in order to represent an ever-increasing realism and detail of new scenarios for our future climate, is the only way to reduce these latter uncertainties (Palmer, 2014).

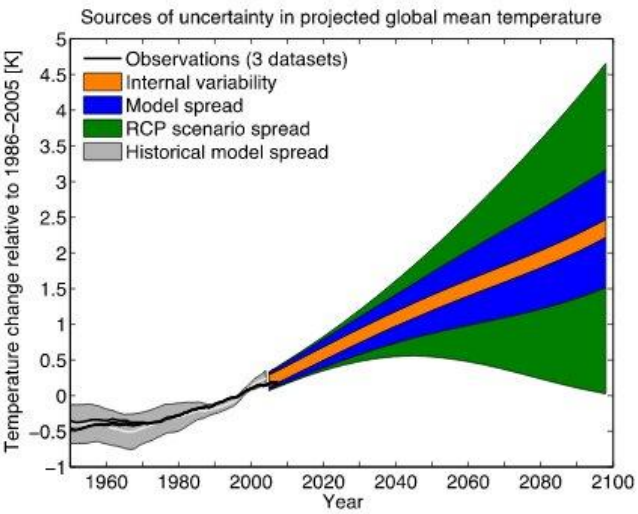

Figure 1. The sources of uncertainty in global decadal temperature projections, expressed as a 'plume' with the relative contribution to the total uncertainty colored appropriately. The shaded regions represent 90% confidence intervals.

This means that, even though the scientific community has little doubt that climate is sensitive to mankind's activity, many questions remain unsolved at the quantitative level (Stevens et al, 2013; Liu et al, 2015). There is a need to better qualify and quantify uncertainty of the predictions, estimate the probability of extreme events, quantify the feedbacks between climate and biogeochemical cycles such as carbon dioxide and methane, and therefore identify the impacts of climate change on marine and terrestrial ecosystems and on societies (Lenton et al, 2003; Doblas-Reyes et al, 2013; Flato et al 2013;

Hartman et al, 2013; Mauristsen et al, 2013). These are some of the reasons why **the final goal of OEMES is to increase the capability and comprehensiveness of 'the whole Earth system' models, in order to represent an ever-increasing realism and detail of new scenarios for our future climate, being the only way to reduce these latter uncertainties in the future.**

However, the increase of the capability of ESMs is strongly linked to the amount of computing power and data storage capacity available. The scientific community ask for increased model resolution, large numbers of experiments, increased complexity of ESMs, and longer simulation periods compared to the current state of climate models (Palmer, 2014). For example, Prodhomme et al (2016) showed the benefits when the horizontal resolution of a global ESM is increased for seasonal climate prediction. This proves that there is a vital need for HPC in order to predict the evolution of the climate and answer key societal questions about the impact of global warming on human activities (Palmer 1999; Palmer 2014). Actually, sustained computing power of 1 Pflop/s or more is already required today for Europe to maintain its scientific weight in climate change research worldwide.

HPC has evolved in the last years from a technology crucial to the academic research community to a point where it is acknowledged as a key piece of the numerical modelling (Casanova et al, 2011), being its application a real need for Earth System Modelling (Bastoul et al, 2004). HPC is currently undergoing a major change as the next generation of computing systems ('exascale systems') is being developed to be ready in a near future (Broekema et al, 2012). These new systems pose numerous challenges, from a 100-fold reduction of energy consumption to the development of programming models for computers that host millions of computing elements, while addressing the data challenge presented by the integration of both observational and simulation/modelling data (Alexandrov et al, 2014; Palmer, 2015). These challenges cannot be met by mere extrapolation but require radical innovation in several computing technologies and numerical algorithms.

Most applications targeting exascale machines require some degree of rewriting to expose more parallelism, and many face severe strong-scaling challenges if they are effectively to progress to exascale, as it is demanded by their science goals. There is an on-going need of support for software maintenance, tools to manage and optimise workflows across the infrastructure, and visualisation. Support for the development and maintenance of community code bases is recognised as enhancing research productivity and take-up of HPC (Palmer, 2014). There is an urgent need for algorithm and software development to be able to continue to exploit high-end architectures efficiently to meet the needs of science, industry and society. **This project aims (as starting hypothesis) at figuring out how to increase the computational performance of ESMs and how to adapt them for the new generation of supercomputers, when the increase of computer power will be mandatory to reduce the uncertainties of climate simulations.** This proposal is split into the following sections:

1. Earth System Modelling and the path to exascale challenge in Europe
2. The path to exascale challenge for EC-Earth, the European ESM
3. Efforts of Europe, BSC and the Performance Team to the exascale challenge for EC-Earth and ESMs in general
4. Societal challenges
5. Objectives
6. Background of the work team and other groups working on it
7. Methodology and planning
8. Technical resources
9. Human resources

1. Earth System Modelling and the path to exascale challenge in Europe

Weather, Climatology and Earth Sciences (WCES) encompass a wide range of disciplines ranging from the study of the atmosphere to the oceans. They are all part of Earth system sciences or geosciences. Earth system sciences address many important societal issues, from weather prediction to air quality, ocean prediction and climate change to natural hazards such as seismic, volcanic and tsunami hazards. The development and the use of high performance computing plays a crucial role to accomplish these societal issues, allowing the execution of numerical models with the computer power needed to obtain useful results (Stocker et al, 2013; Taylor et al, 2013; Li et al 2014).

Research in the fields of weather, climatology and Earth sciences is of key importance for Europe. Obviously, one of the main concerns for Europe is the climate change. Increasing the capability and comprehensiveness of 'whole Earth system' models that represent in ever increasing realism and detail scenarios for our future climate, is the only way to reduce the current uncertainties (Palmer, 2014). A further challenge is to provide more robust predictions of regional climate change at the decadal, multi decadal and centennial timescales to underpin local adaptation policies. In many regions of the world, there is still considerable uncertainty in the model predictions of the local consequences of climate at different timescales and model resolution plays therefore a key role. A dual track approach should be considered involving multi member multi model comparisons at the current leading edge model resolution (about 20 km, limited to a few decades) alongside the longer term, aiming to develop a global convective resolving model (down to 1 km resolution). Reducing these uncertainties in climate projections requires a coordinated set of experiments, taking advantage of existing and new HPC platforms (Alexandrov et al, 2014).

As Guest (2014) explains, modelling the climate system is a challenge because it requires the simulation of a myriad of interacting and complex processes as well as their analysis at different time and spatial scales. Climate system modelling requires furthermore sophisticated numerical models, due to the inherently non-linear governing equations (Randall et al, 2007). In addition, huge computational resources are needed to solve billions of individual equations describing the physical processes at different scales. Indeed, model simulations are required to represent both, the modification of the larger scale, global state (inside which extreme events are developing) and the fine temporal scale and spatial structure of such events (storms, cyclones, intense precipitation, etc).

Currently, global climate models running in an efficient way (from a computational point of view) have typical grid spacing of 100–200 km and are limited in their capacity to represent processes such as eddies, clouds, orography effects, small scale hydrology, etc. The latest generation of models, under development or just starting to be used, have grid spacing in the 20-50 km range, and there are evidences that a number of important climate processes are better represented at this resolution (Prodhomme et al, 2016) (e.g. aqua planet, ENSO, satellite validation, blocking, tropical storm numbers, etc.). A priority for **OEMES is then to continue the development of coupled models at such high resolution and to use them in ultra-high resolution experiments (1-10 km range) focused on key climate processes.**

Very high resolution global models are expected to improve our predictions and understanding of the effect of global warming on high impact weather events on seasonal, decadal and century timescales (Zhang et al, 2015). Another issue is the simulation of regional scale climate features, of crucial importance for assessing impacts on society and economic activities (farming, fisheries, health, etc.) and for which improved regional models, embedded in global climate models, are necessary.

Increasing model resolution down to 1 km requires increases by factors of at least 100 to 1,000 in computing power compared to the current state. It should be noted that each increase of the spatial resolution by a factor 2 in each direction mandates at least an eightfold increase in computing power, depending on the time step length that needs to be decreased with increasing resolution. More computing power means more resources of a supercomputing used in parallel. However, the more resources in parallel are used, the more overhead is introduced to achieve the parallelization. **One of the main goals of OEMES is for instance to achieve the good scalability of the models, when the parallelization will be extreme.**

Climate models fundamentally scale with difficulty on supercomputers because the problems they represent are connected, physically and algorithmically (Broekema 2012). This requires significant communication results in an increasing overhead with increasing domain decomposition (Irony et al, 2004). There are applications for which distributed systems would provide good performance (Dongarra et al, 2014), but these cases generally depend on models with very good portability and with relatively low input/output volumes, criteria that are not fulfilled in general by ESMs. Additionally, systems especially suited to ESM high performance computing applications need to provide both capability and capacity with a good balance between computer power and energy efficiency (Pavel et al, 2014). **One of these ESMs, which need urgently the improvement of his capability and capacity in the path to exascale, is EC-Earth.**

The path to exascale challenge for EC-Earth, the European ESM

EC-Earth is a project, a consortium and a model system. The EC-Earth consortium consists of 24 academic institutions and meteorological services from 11 different countries in Europe, including the

"Agencia Estatal de Meterología" (AEMET). EC-Earth is used in many EU FP7 and H2020 projects, including IS-ENES2, ESiWACE, PRIMAVERA and others.

ESMs, such as EC-Earth, are currently the only way of providing information on the future climate to society (Hazeleger et al, 2007). EC-Earth generates reliable in-house predictions and projections of global climate change, which are a prerequisite to support the development of national adaptation and mitigation strategies. EC-Earth is developed as part of a Europe-wide consortium thus promoting international cooperation and access to wide knowledge and data base. It further enables fruitful interactions between academic institutions and the European climate impact community. EC-Earth makes successful contributions to international climate change projections such as CMIP5 or the future CMIP6 (Taylor et al, 2015; Stocker et al 2013; Zhang et al, 2015). The development by the consortium will ensure more reliable projections, which can be offered to decision and policy makers at regional, national and international levels.

EC-Earth can be used for predictions across time scales, ranging from weather to climate. EC-Earth can be used to study climate feedbacks, which govern the global and regional response. The scientific questions at stake include the dynamics of both, the atmosphere and the ocean, at all spatial and temporal scales, with the many interactions between the various scales, which are significant from an energetic and/or transport points of view (Palmer et al, 2008). This includes progressing in the numerical simulation at higher spatial resolution, so that the effects of unresolved scales can be made as small as possible. However, this increase in the resolution of the grids involves the use of thousands of parallel resources of a supercomputer. The execution of these models are complex, and the lack of scalability of the multi component (including the simulation at the same time of the atmosphere, ocean, land, coupler, etc.) climate models is just one reason why they have not been run often on complex machines with good energy efficiency. EC-Earth is one of the ESMs, which has this kind of problems, preliminary studies proved that the improvement of the energy efficiency of EC-Earth is mandatory before thinking in the execution of ultra-high resolution experiments, where the computer power needed will increase dramatically (Yepes et al, 2016). Figure 10 in Yepes et al (2016) shows the scalability (SpeedUp) and parallel efficiency of EC-Earth using standard resolution (the most common resolution used for the institutions these days), **this figure shows that the parallel efficiency (how the supercomputers resources are exploited in comparison to a sequential running) is less than 30% for a number of 300-400 cores, similar to the efficiency achieved using other higher resolutions and more cores. This efficiency is well below the minimum required in some supercomputing centres to use their resources.** For example K-Computer (Fifth in the Top 500, Japan) requires a minimum parallel efficiency of 50% to use its resources.

The EC-Earth model is a global, coupled climate Earth Model that consists of three main components: IFS for the atmospheric model, NEMO for the ocean model and the coupling between them using OASIS3-MCT. It has other sub-components which are not explained here. A brief description of the main components and computational challenges follow as:

1.1. IFS atmospheric component - Weather forecasting

The Integrated Forecasting System (IFS) as atmosphere model: this is an operational global meteorological forecasting model developed by the European Centre of Medium-Range Weather Forecasts (ECMWF). The dynamical core of IFS is hydrostatic, two-time-level, semi-implicit, semi-Lagrangian and applies spectral transformations between grid-point and spectral space.

Weather forecasting applications such as IFS cannot be run globally for climate because of the prohibitive cost of associated computing resources and limits in model scalability. Improve scale resolution is necessary for reliable predictions of some important aspects of regional climate change (Shepherd et al, 2014). Developing such very high resolutions will require developing scalable and more efficient dynamical cores and resolve explicitly physical processes, whose modelling still today relies on phenomenological parametrisations.

1.2. NEMO ocean component – Ocean modelling

The Nucleus for European Modelling of the Ocean (NEMO) is a state-of-the-art modelling framework for oceanographic research, operational oceanography seasonal forecast and climate studies. It discretizes the 3D Navier-Stokes equations, being a finite difference, hydrostatic model, with a free sea surface and a non-linear equation of state.

In ocean applications such as the NEMO model, the progress is intricately linked to the computing power and energy efficiency available due to the need for increasingly higher model resolutions, many more simulations, and greater complexity in ocean system models. Operational oceanography is a new and rapidly growing sector, providing key assessments for coastal water quality, fisheries and marine ecosystems, offshore, military, transport, etc. A key concern in the ocean, such as in the atmosphere, is eddies, which can be addressed only by building upon recent advances in ocean modelling to construct more accurate, high resolution models. Yet it remains a challenge to run realistic global or regional ocean/sea ice models at resolutions high enough to ensure dynamical consistency over a wide range of resolved scales.

1.3. OASIS3-MCT coupler - Coupling among Earth System components

The OASIS3 coupler is a library to be linked to the component models, whose main task is to interpolate and exchange the coupling fields between them (using the same or different grids).

A climate model is the assembly of different numerical models, which can be independently developed, interacting in a coupled way. This is a complex process among different components that can all run on different grid configurations, supporting a variety of spatial resolutions and time scales and exchanging boundary data with each other through the coupler (Palmer, 2014). The huge computational cost comes because some of the variables must be coupled in a conservative way so that the total flux or energy in the source and target grids must be the same, increasing the complexity of the coupling algorithm. These issues represent serious practical implications for the numerical implementation and the computational expense of the coupling. The main problem will come when higher resolutions are used, where thousands of cores have to communicate the information of the coupling, converting this in a bottleneck.

1.4. Robustness and validation of Earth System Models

Finally, an important aspect to be addressed is robustness for EC-Earth. Robustness has several aspects including: fault resilient algorithms which enable them to run on large and error-prone systems, reproducibility of numerical results and numerical stability of algorithms, and accuracy and validation of the results compare to observational data (Palmet et al, 2008; Reichler and Kim, 2008). Additionally, robustness enables many opportunities to trade performance for accuracy and reproducibility, which makes it possible to consider approximate computing scenarios where the exactitude of some computations can be turned to maximise compute or memory access performance, whilst keeping the results' quality within an acceptable margin.

Taking into account the computational challenges of EC-Earth (IFS, NEMO, OASIS and the robustness of the Earth System) in the path to use more computing power with massive parallel platforms. And considering that the challenges cannot be met by mere extrapolation but that require radical innovation, **OEMES focuses on three specific challenges for EC-Earth:**

(1) Analysing the main bottlenecks of the atmospheric, ocean and coupling components when an extreme parallelization is used.

(2) Trying to exploit high-end architectures efficiently, reducing the energy consumption of these models through novel mathematical methods and computational algorithms.

(3) Evaluating if massive parallel execution and the new methods implemented could impact the quality of the simulations or lose reproducibility.

2. Efforts of Europe, BSC and the Performance Team to the exascale challenge for EC-Earth and ESMs in general

These objectives are common to the European path through the initiative EuroHPC. However, as it will be explained, OEMES will fulfill these objectives providing a novel approach that has not been taking into account by the community, which main efforts are done through EuroHPC.

EuroHPC is a joint collaboration between European countries and the European Union about developing and supporting exascale supercomputing by 2022/2023. The EuroHPC is a legal and funding entity which will enable pooling of the Union's and National resources on High-Performance Computer (HPC) with the aim of: (1) acquiring and providing a world-class pre-exascale supercomputing infrastructure to Europe's scientific and industrial users, matching their demanding application requirements by 2020 and (2) developing exascale supercomputers based on competitive EU technology that the Joint Undertaking could acquire around 2022/2023, and that would be ranking among the top three places in the world. To complete these tasks, several European projects born under EuroHPC and the H2020 framework, related to achieve the required exascale computing for ESMs with different societal challenges. Some of the relevant projects are ESiWACE, ENES and ESCAPE from the HPC side and CMIP and PRIMAVERA from the scientific side:

ESiWACE, this H2020 project aims at "building a critical mass and expertise to increase the weather and climate community impact on hardware development towards the extreme scale and create international Exascale initiatives". Other goal of ESiWACE is to establish demonstrator simulations, which will be run at the highest affordable resolutions to estimate the computability of configurations that will be sufficient to address key scientific challenges in weather and climate prediction. However, the results also show that the energy efficiency of these configurations is poor, the production is close to 0.5 Simulations Year Per Day (SYPD), a real bad number considering that the community considers an acceptable configuration when this number is higher than 5 SYPD. Some strategies are also studied through ESiWACE to improve the energy efficiency of these models, being the Domain Specific Language (DSL) the most promising. Such language is bundled with a reliable Source-to-Source translator that converts DSL code into fully compatible Fortran code, where the computation details are expressed. The goal is not only the adaptation to the architectures; but also aiming for capturing the semantics of climate models. However, it is not clear yet how well the models will be optimized once the configurations change critical parameters such as the resolution. The results is that new profiling analyses and the introduction of additional optimizations will be needed again, if not to the fortran coded directly, at least during the transformation done by the translator.

ESCAPE, the goal of this H2020 project is the "Future improvements in predictive skills for both, weather and climate that will eventually originate from enhanced spatial resolution, the better representation of more complex physical and chemical processes". To achieve this, ESCAPE defines and encapsulates the fundamental algorithmic building blocks (known as 'Weather & Climate Dwarfs') as small components which are doing only a small part of the calculations of a real model (such as the physical calculations of an atmospheric model). This is the prerequisite for any subsequent co-design, optimization, and adaptation efforts done to the Dwarfs. Additionally, some of the Dwarfs are implemented not only for general purpose architectures, but also accelerators (GPUs, FPGAs) or the new DSLs. However, it is not clear again which is the best methodology to do the profiling analysis of the Dwarfs executions and which optimizations will be really valuable for real models.

ENES is the European Network for Earth System modelling. It provides information and services on the European Earth System Models and some associated tools, on data produced by internationally coordinated climate model experiments and on the use of high-performance computing facilities to run complex climate simulations. ENES provides the future HPC infrastructure needed for the simulation of complex ESMs compound by different components (such as ocean, atmosphere, ice, vegetation...) and the coupling algorithmic to achieve this. However, the creation of the new infrastructure able to achieve this needs an analysis of the complex ESMs which will be used. This models are compounded by several independent components running in parallel, where the correct load balance of all of them at the same time represent a great challenge to maintain a good efficiency and scalability.

PRIMAVERA, this H2020 project is in charge of "simulating and predicting regional climate with unprecedented fidelity, for the benefit of governments, business and society in general". *PRIMAVERA* stands for "Process-based climate simulation: Advances in high-resolution modeling and European climate risk assessment". It aims at developing a new generation of advanced and well-evaluated high-resolution global climate models.

The Coupled Model Intercomparison Project (CMIP) is a comparison of a handful of global coupled climate models performing experiments using atmosphere models coupled to a dynamic ocean, a simple land surface, and thermodynamic sea ice. The objective of CMIP is to better understand past, present, and future climate change arising from natural, unforced variability or in response to changes in radiative forcings in a multi-model context. Its increasing importance and scope is a tremendous successful, but this very success poses challenges for all involved. The new iteration, CMIP6, continues the pattern of evolution and adaptation characteristic of previous phases of CMIP. CMIP6 comprises 21 individual CMIP6-Endorsed MIPs and the DECK and CMIP6 historical simulations.

Apart from the European community, BSC is also leading activities related to the application of HPC to the improvement and optimization of applications. In particular, BSC at Computer Science (CS) department is leading POP, a Centre of Excellence in Computing Applications in the area of Performance Optimisation and Productivity. POP offers the service of precisely assessing the performance of computing application of any sort, from a few hundred to many thousand processors. POP targets code owners and users from all domains, including infrastructure operators, academic and industrial users. To achieve this, BSC has developed the open access BSC Tools, a performance analysis tools (similar to others as Vampyr but more powerful) which allow application developers to identify and characterise the inefficiencies that caused a poor performance (**https://tools.bsc.es/**). However, since POP is targeted for all kind of applications, it can not provide analysis oriented to one complex field such as Earth Science, being the analysis more related to general technical issues.

To supply this, the Performance Team inside the Earth Science (ES) Department is compounded by experts with knowledge in both, Computer Science and Earth Science fields, to provide the interdisciplinary knowledge. The Performance Team is collaborating with the CS department at BSC to extend the national research started at CS department to the Earth Science community, providing: (1) knowledge about the mathematical and computational side of Earth System Applications and about the specific needs in HPC and (2) researching about HPC methods specifically used for Earth System Applications. Some preliminary analyses have been done for EC-Earth [Acosta et al. 2016, 2017; Yepes et al. 2016, 2017] using BSC Tools and presented to the community in international conferences. However, these tools are not integrated or used by the community as a standard yet.

During the collaboration between the Performance Team and the CS department, others applications from the CS department have been proved as valuable for ESM applications. Dynamical Load Balance (DLB, **https://pm.bsc.es/dlb**) is a library devoted to speedup hybrid parallel applications, which improves the efficient use of the computational resources inside a computing node. DLB is useful for ESMs because the classical domain decomposition done for MPI and OpenMP paradigms produce subdomains with more computational work than others (i.e. latitudes close to the poles in a global model) which are impossible to solve in a static way. Other one is Cassandra/Pandora (**https://www.bsc.es/research-and-development/software-and-apps/software-list/pandora-hpc-agent-based-modelling-framework**), a program developed to analyse the results generated by a simulation created with the library. Cassandra allows the user to visualize the complete execution of simulations using a combination of 2D and 3D graphics, as well as statistical figures, which will be very useful for the Earth System postprocessing, reducing the operations needed to create diagnostics and removing some storage operations. Finally, the last one is OmpSs (**https://pm.bsc.es/ompss**), to extend OpenMP with new directives to support asynchronous parallelism and heterogeneity (devices like GPUs). OmpSs will be useful for the complex workflows used by ESMs, creating data-dependencies of the tasks and improving the performance automatically.

3. Societal challenges

The interdisciplinary character of the actors involved in Earth System Modelling in Europe (Weather and climate prediction, numerical modelling, computer science and HPC) requires <u>comprehensive solutions that include</u> technological development, structural changes (e.g. optimisation of infrastructures to achieve exascale computing) and the adaptation of scientific models which are not evolved in the last decades. OEMES can provide this <u>interdisciplinary framework</u> to evolve these ESMs in order to do the ultra-high resolution experiments, which could solve the future evolution of the weather and climate and answer key societal questions about the impact of global warming.

All in all, OEMES's results are expected to provide better scientific understanding of climate change in Spain and Europe, whereby matching the premises of one challenge in the "Plan Estatal de Investigación Científica, Técnica y de Innovación 2013-2016": '5º Reto en acción sobre el cambio climático y eficiencia en la utilización de recursos y materias primas', in particular with the topic "I. Cambio Climático".

OEMES will help to enhance the quality of the results of EC-Earth, contributing to the challenge specified, through the ability to exploit high-end architectures efficiently, reducing the energy consumption of EC-Earth in the simulation of ultra-high resolution experiments. These experiments will contribute to the reduction of uncertainties around the climate change study.

OEMES is also tightly following the H2020 societal challenge "**Climate action, environment, resource efficiency and raw material**", via its specific objectives "Climate action", "Resource efficiency" and "Earth Observations". In the interdisciplinary space, it is also related to the H2020 call in "**Future and Emergent Technologies for High Performance Computing**". In particular, "**FETHPC-02-2017: Transition to Exascale Computing**", whose specific challenge is **to take advantage of the full capabilities of exascale computing**, through high-productivity programming environments and new mathematics and algorithms for extreme scale HPC systems for existing or visionary applications.

OEMES framework contributes also to the objectives defined by the "Estrategia Española de Ciencia y Tecnología y de Innovación 2013-2020" because of its international profile and the possibilities that will be offered to the postdoctoral scientists and the doctoral student to collaborate with the extermal members of the team who are part of recognized institutions that lead large international projects. These types of collaborations are the ideal ground from where a successful career network is built.

Moreover, the European community is working on the development of new infrastructures to achieve exascale computing and the adaptation of the scientific models to improve the computational efficiency. However, there is no a clear methodology about how to do this and validate results once some optimizations are implemented. Additionally, there are some BSC research lines explained in 1.3 which has not been proved in real ESMs but will be useful in the path to improve the efficiency of these models going to the pre-exascale goal. Although the creation of the Performance Team and the leadership of the PI was only three years ago, the team has published several documents (Acosta et al. 2016, 2017; Yepes et al. 2016, 2017) showing the great utility of the BSC performance tools for the profiling analysis of ESMs and how these tools and the knowledge between Earth and Computer Science provide an evaluation of the ESMs never done by the community in the past. This means that OEMES has the potential to push the BSC (the host institution) and Spain at the forefront of the international scientific community for profiling analysis applied to ESMs. This means to integrate a novel methodology to analyze the computational performance and validation of ESMs and create a new standard followed for all the EC-Earth community. If it is funded, OEMES can consolidate the profile of the PI and the rest of the research team within the community, opening a direct channel to future European funding and introducing the Spanish research and tools inside these possible projects. The successful establishment of a research sub-group in profiling analysis and optimization within ES department will complete the spectrum of activities in which BSC is a reference player at the international level, making OEMES fit for objective 2.6 ("Fortalecimiento institucional").

4. Objectives

The overal scientific objective of OEMES is <u>to develop a new methodology for profiling analysis and validation which will be a reference for all the HPC European earth system community, extending the national tools and research lines to the community. Secondly and an example of success, optimize the energy efficiency of EC-Earth, one of the main European Earth System Model, to take advantage of future exascale computers in a efficient way. The final goal is to achieve a minimum acceptable efficiency for EC-Earth to simulate ultra-high resolutions experiments, increasing the capability and comprehensiveness of 'the whole Earth system' model, in order to produce ever-increasing realism and detail scenarios for the future climate and reduce the uncertainties in climate change simulation.</u>

This goal will be achieved through four complementary <u>objectives</u>, which define the project's methodological approach and cover all the possible optimizations for EC-Earth:

-**Developing a novel profiling method to highlight numerical and computational problems** of numerical models, executed in extreme parallel platforms and **creating a new standard for all the community**. (WP1)
-**Developing a novel method to validate and reproduce the results of Earth System Models**, taking into account a trade-off among accuracy, reproducibility and performance and **creating a new standard for all the community**. (WP2)
-Being ready to take advantage of future pre-exascale computers. To accomplish this, the next sub-goals are needed (WP3-4-5):

-Preparing ultra-high resolution experiments for extreme parallelization, **integrate the new profiling methodology developed and evaluate the profiling results**.

-Improving energy efficiency of the main components of EC-Earth. A minimum parallel efficiency of 50% using more than 20,000 cores will be achieved for:

-The atmospheric component of EC-Earth, IFS. (WP3)

-The ocean model of EC-Earth, NEMO. (WP4)

-The coupling among Earth System Components, OASIS. (WP5)

-Integrating the new validation method and achieving a trade-off among the new optimizations (performance), the better results for the higher resolution (accuracy) and the chaotic nature of ESMs (reproducibility). (WP6)

The completion of OEMES's scientific objectives is expected to make a leap forward in our understanding of weather and climate predictions. The developments within OEMES are expected to be key on the energy efficiency improvement of the ESMs. There are different EU-wide strategies designed to accomplish the main goal of several European projects, in order to move ESMs to exascale platforms in the future, trying to reduce the thousands and thousands of hours that these models spend in our supercomputers. OEMES project will contribute greatly to this goal with a novel and unique approach.

5. Background of the research team and other groups working in the topic

Experience of the PI: During his PhD, Dr Acosta contributed greatly to HPC applied to ESMs and therefore developed expertise in this field. The general knowledge acquired was to improve the computational efficiency of existing computational dynamic fluids models to address the study of circulation, transport and mixing of water components, obtaining the results in an efficient way. This expertise includes wide knowledge in numerical models (governing equations, numerical algorithms and computational implementation) and how to adapt them efficiently to actual and new HPC resources. He has moved to BSC three years ago, assuming the leadership of the Performance Team. Dr Acosta's research lines are well embodied in those related to HPC applied to ESMs in the BSC-Es. As leader of the performance team inside the department, he is the supervisor of one PhD student and one master student, who are developing theses in the HPC topic. Only during the last year, Dr Acosta has collaborated in the publication of three peer-reviewed papers, the submission of five papers (two accepted), five oral communications, four posters in international conferences and four technical memoranda, all within the framework of five H2020 and FP7 projects. He has also several formal collaborations with international institutions, the Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS, OASIS developers), and the ECMWF (IFS developers), where Acosta did a post-doctoral visit of one week and of two months (Severo Ochoa mobility grant) of in 2017 respectively at CERFACS and ECMWF. During his career, the applicant has participated in 9 national and international projects. He is author of 6 peer-reviewed articles in international journals. He presented his work at 15 congresses and workshops, and reviewed some manuscripts for peer-review journals. He is part of the group of experts of European eXtreme Data and Computing Initiative: Weather, Climatology and Solid Earth Sciences (EXDCI/WCES). Thanks to the good trajectory of the Performance Team in the last three years, Dr. Acosta is working in the HPC WP of ESiWACE2 and ESCAPE2 and the leader of the HPC WP of IS-ENES3 as BSC PI. Moreover, he is also the PI of a research project with KNMI related to HPC of the model Armonie-Arome.

Experience of Alicia in the Research Group: Dr. Alicia Sanchez is currently co-leading the CES group within the Earth Sciences Department. She has a long experience in developing computational techniques based on numerical methods for physics modelling as well as data analysis algorithms to deal with huge amount of data. She has participated on several international Scientific Collaborations and has helped with her interdisciplinary knowledge to consolidate important European research projects within the framework of High Precision Energy technologies. Currently, her interest focuses on developing robust and efficient numerical recipes for the improvement of ESM.

Experience of Roberto in the Research Group: Dr. Roberto Bilbao is currently a postdoctoral scientist in the CP group within the ES Department. He has a strong background in physical oceanography and in ocean and climate modeling. He has experience in seasonal/decadal forecasting and climate projections, with a particular focus on ocean dynamics. He is involved in the EC-Earth Ocean working group, which, among other tasks, coordinates the tuning of the ocean model component of EC-Earth.

He is involved in European projects such as PRIMAVERA, where his role is to investigate if the new higher simulations predicte regional climate with unprecedent fidelity and its validation.

Experience of Kim in the Research Group: He is the leader of the CES group at the Earth Sciences department in the BSC. In the last years, he has been in charge for the system administration of all the computational resources of the department and he was also responsible of supervising the operational runs of the NMMB/BSC-Dust model and CALIOPE Air Quality System in the HPC infrastructures of the BSC. In that sense, he was also involved in the analysis of the models to improve their performance and developed strong skills of HPC applied to ESMs as NEMO. In March 2014, he wrote his Master Thesis in the “Analysis, Developments and optimizations on the NMMB/BSC model”, focusing his research on applying many different techniques to improve the model performance in an HPC environment. Furthermore, he's focused on deploying ESMs (dust transport, climate or weather forecast) required by the department in a wide range of HPC architectures. He applied with success these skills in projects like IS-ENES (1 & 2), ESiWACE, SDS-WAS or BDFC or CONSOLIDER. He also teaches in different HPC schools oriented to the Earth Sciences and Climate Model as PRACE Advanced Training Course (2012 to 2016) and the IS-ENES Summer Schools (2014, 2016). He is the BSC PI of H2020 projects (ESiWACE1&2, ESCAPE2, IS-ENES2). He is the leader of WPs about HPC and evaluation of ESMs in ESiWACE, ESCAPE2 and IS-ENES3.

At national level, several research groups work using EC-Earth, among them:

-Earth Science department at Barcelona Supercomputing Centre (BSC-ES). The department was granted several H2020, FP7, Copernicus, ESCAPE European projects, apart from other projects funded by the Ministerio de Economía y Competitividad (MINECO), all related to Earth System Modelling researching. During that same period, BSC-ES also participated in 21 RES and 4 PRACE projects. The BSC-ES has participated in climate services initiatives like the Climate Services Partnership (CSP). Members of the BSC-ES participate in committees of the World Climate Research Programme (WCRP), such as the CLIVAR Scientific Steering Group or the Working Group on Seasonal to Interannual Prediction (WGSIP). Two research groups work using EC-Earth, the Computational Earth Science (CES) and the Climate Prediction (CP) groups, both of them in the Earth Science department.

CES group provides expertise and guidance to the other scientists in the technical issues and develops a framework for the most efficient use of HPC applied to ESMs. CES is divided in three teams and research lines: Models and workflow development, Data and Diagnostics and Performance. The performance team aims at providing feedback on model efficiency to modellers around Europe and improve them by applying optimizations by using novel and research methods. Many of these activities are related to EC-Earth. Some members of CES (including Mario Acosta, Alicia Sánchez and Kim Serradell) are part of the technical working group of EC-earth and NEMO consortium. There is also an official collaboration with the main developers of IFS (ECMWF) and OASIS (CERFACS), being Mario Acosta the contact point. From CES and the Performance Team, the people with the correct knowledge and the availability to accomplish the OEMES goals will be involve in the project (Alicia Sanchez, Kim Serradell and Mario Acosta).

On the other hand, CP group (including Roberto Bilbao) 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. Many of the activities in modelling and prediction are based on research, development and predictions with the EC-Earth climate forecast system. This group is already active in the planning of the next phase of Coupled Climate Model Intercomparison project, CMIP6, and is preparing to make core contributions including the ground-breaking high-resolution global climate simulations with EC-Earth. Moreover, the group is also contributing to H2020 European projects, planning a set of experiments using exascale platforms for the future, with horizontal spacing close to 1km for the atmosphere and ocean for PRIMAVERA project.

The environment of the department provides a unique opportunity where experts from both groups can collaborate together, pursuing the same goal. The expertise in climate prediction will be needed to set up ultra-high resolution experiments, validate the results and evaluate the reproducibility of the experiments, in order to take advantage of the future exascale platforms to simulate patterns which cannot be possible using the actual computer power. The expertise in HPC will be needed to evaluate which areas of the code need a real optimization and improve the energy efficiency of these experiments, obtaining results in a logical time. The team formed by the PI, Dr. Roberto, Dra. Alicia and Mr. Kim provide an interdisciplinary and complimentary expertise beyond the computer science, HPC and climate modelling backgrounds.

-La Agencia Estatal de Meteorología (AEMET). They work actively using EC-Earth, their main goal is to build a fully coupled Atmosphere-Ocean- Land-Biosphere system model, usable from seasonal to decadal climate prediction and climate projections.

At international level, many different research centres and universities departments work on the topic (EC-Earth consortium is compound by 24 universities and institutions around Europe, all of them use EC-Earth actively). Listed below are the most relevant ones for the project:

-The European Centre for Medium-Range Weather Forecasts (ECMWF), they develop the atmospheric model (IFS) used by EC-Earth.

-The University of Exeter, they are one of the main developers of the ocean model (NEMO) used by EC-Earth.

-The Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (CERFACS), they develop the coupling library (OASIS) used by EC-Earth.

-The Swedish Meteorological and Hydrological Institute (SMHI). They are one of the main developers of EC-Earth.

The Working Group includes four PhD researchers from the international groups commented. Dr. Nils Wedi is the head of Earth System Modelling at the ECMWF (https://www.ecmwf.int/en/about/who-we-are/staff-profiles/nils-wedi).Their expertise includes all aspects of scientific and computational performance related to weather prediction. He is the ECMWF scientific coordinator of ESCAPE project and a member of the WMO working group on numerical experimentation (WGNE). Dr. Sophie Valcke is the leader of the OASIS code coupler development team at the CERFACS (https://cerfacs.fr/annuaire/valcke/). His expertise includes the knowledge on high-resolution atmosphere-ocean-ice coupled modelling. She is the CERFACS PI for ESiWACE1&2, IS-ENES1&2&3 and METAFOR projects and plays a key role in the International Working Committee on Coupling Technologies. Dr. Martin Schreiber is a proleptic lecturer at the University of Exeter (UK), in the Mathematics Department, and member of the technical working group of NEMO (https://www.martin-schreiber.info/). His expertise includes the knowledge of performance analysis and optimization of ocean models on HPC architectures. Dr. Uwe Fladrich is a scientific software developer at the SMHI, in the Rossby Centre, and leader of the technical working group of EC-Earth (https://www.smhi.se/en/research/research-departments/climate-research-rossby-centre2-552/uwe-fladrich-rossby-centre-1.74857). His expertise includes the knowledge on efficient software development processes and numerical aspects of climate models. He possesses comprehensive experience in the coordination of distributed software development and project activities within the EU funding framework, being SMHI PI of IS-ENES3.

Methodology and planning

As envisaged from its conception, OEMES has been designed with the aim of tailoring the objectives, the Research Group's background, the Working Group's expertise and the facilities of the host institution, in order to ensure its successful completion. The work plan reflects this idea, where the scientific program can be implemented from head-to-tail and in autonomy. The project is divided in seven work packages (WP): two for research (WP1-2), five for research and development (WPs3-6) and one for management and dissemination (WP7).

WP1: Develop a profiling method to highlight numerical and computational problems of numerical models executed in extreme parallel platforms.

Mario Acosta, PhD (20%)	Kim Serradell (20%)
Roberto Bilbao, PhD (10%)	Alicia Sánchez (20%)

The goal of this WP is to create a profiling methodology personalized for ESMs. This methodology will be used for WPs from 3 to 5. Moreover, the new methodology could be used for scientists in other ESMs to facilitate the effort in order to improve the energy efficiency of the models, using a methodology oriented to improve this kind of models for platforms with thousands of cores. It comprises task 1.1 and will be monitored by M1, M3 and D1.

Task 1.1. Creation of a profiling method specific for ESMs

In the ESMs case, an in-depth performance analysis can lead to feasible and productive solutions that do not require a full rewrite of the code while effectively improving the performance. Preliminary studies proved that an exhaustive profiling analysis of EC-Earth will be useful to improve its performance (Yepes et al, 2017). Understanding of the behaviour of applications, running in HPC systems, is not straight forward, this task will develop a methodology to undertake this analysis, with the aim of helping scientists to cope with this issue.

The methodology intends to be useful to scientists that are running EC-Earth on HPC systems and are willing to understand its computational performance, uncovering issues that are potentially hindering its efficiency. This methodology will be specific for the needs of weather and climate models. For this purpose, the next methodology will be extended to study each of the components of EC-Earth.

1) Mathematical study
2) Computational study
3) Profiling study applied to ESMs
4) Identification of main bottlenecks

The interdisciplinary environment of the research group will provide the basis to complete the three first steps from different point of views, including the requirements of each field. The third step will be completed taking into account ESMs and computational metrics, following a novel path that foresees to revolute the profiling analysis for ESMs.

The main outcome of this task will be a novel methodology to highlight problems not previously identified, explaining the reasons for their occurrence and propose optimizations to substantially improve the computational performance of the model.

WP2: Achieving a method to validate and reproduce the results of Earth System Models, taking into account a trade-off among accuracy, reproducibility and performance.

Mario Acosta, PhD (10%)
Kim Serradell (10%)
Roberto Bilbao, PhD (45%)
Alicia Sánchez(10%)

The goal of this WP is to prove that the optimizations developed for the different components of EC-Earth do not reduce the accuracy and reproducibility of the results. We will propose a specific methodology for ESMs, taking into account the chaotic nature of weather and climate models and the round off errors introduced by parallel executions. This new methodology could be used by scientists in other ESMs and environments in order to prove that new optimizations for exascale computing do not affect the quality of the results. It comprises tasks 2.1 and will be monitored by M2-M3 and D2.

Task 2.1. Creation of a validation and reproducibility method to evaluate computational optimizations

This task will develop a method to evaluate all the possible optimizations for EC-Earth, taking into account the chaotic nature of ESMs and the introduction of round off errors produced by parallel computing. The methodology will use statistical methods to evaluate ensemble experiments (same experiment run several times with small perturbations in the initial conditions) and compare the results when the optimizations are included or not.

The main goal of this task is to create a novel methodology which could be used by both climate and computer scientists to evaluate the accuracy and reproducibility of the results. To achieve this, the process will compare automatically the results among ensemble experiments and between different parallel executions. The comparison will include data from CMIP5 to evaluate the accuracy of the simulations. The computational performance of each experiment, including some optimizations on a different scale of aggressiveness will be also evaluated. The final result will be to achieve a trade-off among accuracy (comparing to CMIP5 data), reproducibility (comparing different parallel executions) and performance (using performance metrics among executions).

WP3: Integration and analysis of the new profiling methodology for IFS

Mario Acosta, PhD (60%)
Uwe Fladrich, PhD

Computer Engineer (30%) Nils Wedi, PhD

The goal is the integration of the new profiling methodology developed in the task 1.1, in order to analyse and improve the energy efficiency of the atmospheric model, in order to adapt the model for pre-exascale platforms with thousands and thousands of cores, where the parallel overhead in the algorithms could be a handicap. The development will be integrated into the official version of IFS and EC-Earth. This will ensure an automatic process where IFS users could use the new methodology by activating a parameter. The responsible of this WP (Mario Acosta) will work directly with one of the main coordinators of IFS and EC-Earth codes (Nils Wedi and Uwe Fladich). Both of them will provide the support and tools needed to integrate the new development to the official repositories. Additionally, they will help in the evaluation of the profiling analysis and the interest of the new optimizations for the IFS and EC-Earth users. Preliminary work proved that a profiling analysis for IFS can localize some bottlenecks of the model (Acosta et al, 2017). This work, which has started in collaboration with ECMWF, it will follow through two tasks: 3.1 Integration of the methodology and highlight the main bottlenecks of the model using the methodology proposed in WP1, and 3.2, where some novel optimizations will be implemented in order to improve the energy efficiency. It comprises tasks 3.1-3.2 and will be monitored by M4-M6 and D3.

Task 3.1. Integration of the new profiling methodology and analysis of IFS
This task will set up all parameters needed for the ultra-high configuration and will provide the integration of the new methodology for profiling analysis of the IFS component of EC-Earth. The goals for this task will be to understand perfectly the mathematical and computational algorithm used by the model in order to evaluate if the implementation is optimal and which numerical methods could be changed. The profiling tools will be used to highlight the main bottlenecks of the model for the parallel execution using a large number of resources and ultra-high resolution grids, pointed out with parts of the code should be optimized before running operational experiments with ultra-high resolution grids on future pre-exascale platforms.

Task 3.2. Improving energy efficiency of IFS to achieve a minimal acceptable efficiency for future pre-exascale computers
This task will develop different optimizations in order to improve the scalability of the model when ultra-high resolution grids and supercomputers with thousands of cores are used. Preliminary tests proved that IFS has different areas of the code which could be improved. Some of the strategies that can be developed and tested for IFS are:

- Overlapping communications and calculations for the transposition and transformation stages. This could reduce the communication process in parallel executions after the calculations phases. A novel technique which could be used here is the use of Ompss (see Section 1.2) to do this process, which could improve the performance of the model dramatically.
- Optimizing the shared memory implementation (OpenMP), using DLB methodology (see Section 1.2) to improve the load balance of the threads implementation.
- Evaluate new output processes, comparing two input/output parallel servers, Xios (widely used by the earth science community) and Cassandra (see Section 1.2).

WP4: Integration and analysis of the new profiling methodology for NEMO

Kim Serradell (60%) Uwe Fladrich, PhD
Computer Engineer (30%) Martin Schreiber, PhD

The goal of this WP is similar to the WP3 but for NEMO. In this case, the responsible of the WP (Kim Serradell) will collaborate with Martin Scheiber and Uwe Fladrich, who will provide the support and knowledge needed for the integration of the development into the official repositories of NEMO and EC-Earth and the evaluation of the results. Preliminary work proved that optimizations for NEMO can increase the computational performance (Tintó et al, 2018). This work, which has started in collaboration with the NEMO consortium, it will follow the same strategy that WP3, with similar tasks. It comprises tasks 4.1-4.2 and will be monitored by M4-M6 and D4.

Task 4.1. Integration of the new profiling methodology and analysis of NEMO
This task will integrate and create the profiling of NEMO, using the methodology developed in the task 1.1, explicitly oriented to ocean models. The strategy is similar to Task 3.1 but for NEMO.

Task 4.2. Improving energy efficiency of NEMO to achieve a minimal acceptable efficiency for future exascale computers

Apart from the creation of the ultra-high resolution grids for NEMO, the profiling methodology will reveal the main bottlenecks of the code, similar to the strategy explained in task 3.2. Preliminary tests proved that NEMO has different areas of the code which could be improved. Some of the strategies that can be developed and tested for NEMO are:

- The reduction of the accuracy of the variables from double to single precision. This task will determine which computational phases of NEMO can use single precision.
- Optimizing the domain decomposition of the model during the calculation phases to reduce the unload balance in the parallel execution.

WP5: Integration and analysis of the new profiling methodology for OASIS

Alicia Sánchez, PhD (60%)	Uwe Fladrich, PhD
Computer Engineer (30%)	Sophie Valcke, PhD

The goal of this WP is similar to WP3 and WP4, but for OASIS. In this case, the responsible of the WP (Alicia Sánchez) will collaborate with Sophie Valcke and Uwe Fladrich, who will provide the support and knowledge needed for the integration of the development into the official repositories of OASIS and EC-Earth and the evaluation of the results. Preliminary analysis proved that a correct configuration for OASIS can increase the computational performance (Acosta et al, 2016). This work, which has started in collaboration with CERFACS, it will follow the same strategy that WP3 and WP4, with similar tasks. It comprises tasks 5.1-5.2 and will be monitored by M4-M6 and D5.

Task 5.1. Integration of the new profiling methodology and analysis of coupling using OASIS

This task will integrate and create the profiling evaluation of coupling, using the methodology developed in the task 1.1, explicitly oriented to the coupling using OASIS. The strategy is similar to Task 3.1 but for the coupler.

Task 5.2. Improving energy efficiency of coupling using OASIS to achieve a minimal acceptable efficiency for future exascale computers

In analogy to previous tasks, the application of the profiling methodology will help to address some problematic points within the code, similar to the strategy explained in Task 3.2. Preliminary tests proved that coupling between EC-Earth components has different areas of the code which could be improved. Some of the strategies that can be developed and tested for the coupling used by EC-Earth are:

- Optimizing the load balance between components. A process to optimize the synchronization of the execution time of each component executed in parallel will improve the performance of the coupled model.
- Exploring the different methods for the conservative methods available, in order to determine the best one for EC-Earth, taking into account a trade-off among accuracy, reproducibility and energy efficiency.

WP6: Integration of the new reproducibility methodology and validation of the optimized code

Roberto Bilbao, PhD (45%)	Mario Acosta, PhD (10%)
Kim Serradell (10%)	Alicia Sánchez, PhD (10%)
Computer Engineer (10%)	Uwe Fladrich, PhD

The goal of this WP is to validate and prove the utility of the execution of large and complex experiments using supercomputers with thousands of cores. This WP will integrate the new methodology developed in task 2.1. The responsible of this WP (Roberto Bilbao) will work directly with one of the main coordinators of EC-Earth code Uwe Fladich. He will provide the support and tools needed to integrate the new development to the official repositories. This will ensure an automatic process where EC-Earth users could use the new methodology by activating a parameter and following the instructions. It comprises tasks 6.1 and will be monitored by M5-M6 and D6.

Task 6.1. Proving the improvement in the simulation results of higher resolution and the validation of the optimizations introduced

On the one hand, it will use a formula to find a trade-off between computational performance, reproducibility and accuracy, which will make possible to consider different computing scenarios where the exactitude of some computations can be turned to maximise, whilst keeping the results' quality within an acceptable margin. To achieve this, ensemble simulations will be done to compare the spread of the results, comparing the optimizations developed in the others WPs, among them and to observational data. On the other hand, simulations results will be evaluated in other to show to the community the great improvement that the ultra-high resolution, comparing the accuracy of the new simulations and other operational lower resolution simulations done in the past.

WP7. Project management and dissemination of results.

This WP will ensure the appropriate management of the project and broadly disseminate the outputs throughout its duration. It will be feasible thanks to the Project Management Department at the host institution (BSC) and the strong group of "Earth System Services" established at the department level (BSC-ES). WP7 will monitor the progress of the project, ensure timely preparation of scientific reports (milestones and deliverables) and outreach activities, facilitate communication among the Research, Working Group members' institutions (BSC, ECMWF, SMHI, CERFACS, Exeter University). There will be visits and project meetings to ensure the collaboration and communication with the external members, to discuss the results, plan additional analysis, ensure the integration of the methodology and ensure the exploitation of the results by all the community [M3-M4-M5;V1-V2-V3-V4-V5]. Papers to ensure the dissemination of the novel profiling and validation methodology and the success case in EC-Earth and the optimizations implemented [Pa1-Pa2]. The project will also undertake a final report that, in addition to the summary of the scientific achievements, will identify priority research lines to enhance the energy efficiency of EC-Earth to take advantage of pre-exascale platforms [D7].

Planning

Details are provided bellow about the methodology plan, including milestones and project meetings (Ms), deliverables (Ds) and visits (Vs). Note that the expected date is indicated in terms of the corresponding month, thereafter *pm*. The schedule of the tasks described above is presented in the following chronogram.

Project month (PM)		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
WP1	T1.1																		
WP2	T2.1																		
WP3	T3.1																		
	T3.2																		
WP4	T4.1																		
	T4.2																		
WP5	T5.1																		
	T5.2																		
WP6	T6.1																		
WP7	Miles												M1						
	Delive						D1						D2						
	Meetings													M3					
	Papers												Pa1						
	Visits													V1					

Project month (PM)		19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
WP1	T1.1																		
WP2	T2.1																		
WP3	T3.1																		
	T3.2																		
WP4	T4.1																		
	T4.2																		
WP5	T5.1																		
	T5.2																		
WP6	T6.1																		
WP7	Miles																		M6
	Delive			D3	D4	D5									D3	D4	D5	D6	D7
	Meetings		M4												M5				
	Papers																		Pa2
	Visits			V2		V3									V4				V5

List of milestones (tracking progress)

- M1- New profiling and validation methodology for ESMs
- M2- First project meeting
- M3- Second Project Meeting
- M4- New optimized version of EC-Earth

List of visits (tracking achievements)

V1, V3- 2 two-day visit of ECMWF at BSC
V2, V5 - 2 two-day visit of SMHI at BSC
V4- 1 five-day visit at CERFACS

- M5- Third project meeting
- M6- Validation of the results

List of deliverables (tracking achievements)

- D1- Assessment of profiling methodology for ESMs
- D2- Assessment of reproducibility methodology for ESMs
- D3- Assessment of profiling and optimization work for IFS
- D4- Assessment of profiling and optimization work for NEMO
- D5- Assessment of profiling and optimization work for coupling of OASIS3-MCT
- D6- Assessment of work for validation of the results
- D7- Final scientific report and recommendations

6. Technical resources

OEMES will have access to the BSC HPC facilities to run the experiments in all WPs. The simulations will require a large number of resources for each test. WP3-4-5 and 6 will require a big numbers of cores in order to do the tests to evaluate the energy efficiency and optimizations developed for each of the EC-Earth components. This means that more than 10,000 cores are needed per tests. However, mostly of these tests will not require long period of simulations. Only reproducibility tests will require experiments of at least 100 years of simulation in order to evaluate the accuracy and reproducibility of the results. Thanks to the computational resources of BSC, with the new MareNostrum IV (165,888 cores, 11.15 PetaFlops) the tests will be possible though it will be demanding. However, the completion of the OEMES would be ensured thanks to the use of Marenostrum IV. Moreover, MareNostrum IV is formed of clusters of three different technologies. These are technologies currently being developed in the US and Japan to accelerate the arrival of the new generation of pre-exascale supercomputers. Marenostrum IV includes a cluster which consists of IBM POWER9 processors and NVIDIA Tesla GPUs (1.5 PetaFlops), a cluster made up of Intel Knights Hill (KNH) processors (0.5 PetaFlops) and a cluster formed of 64 bit ARMv8 processors in a prototype machine (0.5 Petaflops). This structure presents a unique opportunity to fulfill the OEMES goals.

Additional infrastructure is needed for a successful implementation of OEMES. Due to the unusually large data needed to do the profiling analysis and evaluate the parallel execution of the model, additional local storage will be needed. The profiling results are expected to generate files with a size of 200 Gb per execution. The output of ultra-high resolution grids (for the reproducibility experiments) should be also taken into account, from 150GB to 300GB per experiment. For the amount of data that the project is expected to generate, we need around 80 TB of raw space. That will require the acquisition of 20 4TB disks. Additionally, the project would need a workstation (PC) for the new Computer Engineer demanded to work in autonomy. Finally, and in order to facilitate an efficient execution/development of the activities in missions outside the host institution, e.g. attendance to meetings, OEMES would require a laptop with an UNIX-based OS.

7. Human resources

The need for a Computer Engineer for the optimal achievement of the tasks in WP3, WP4, WP5 and WP6 is justified by the evaluation of the model performance and the implementation of different optimizations that contribute to the improvement of each of the components of EC-Earth (IFS, NEMO and OASIS). There are three complex components (an atmospheric model, an ocean model and a coupler) which require being analyzed and optimized at the same time (from pm 13 to pm 30). This requires enough human resources, not only to analyze test results or develop the optimizations, but also to do the integration, all the scalability and reproducibility tests and take the results, namely the preparation and execution of hundreds of tests which can be automatized only partially. The additional computer engineer will help the scientists of the Research and Working Group towards a better completion of these objectives, taking over the preparation, execution and analysis of the tests, developed for three components of EC-Earth in parallel. The candidate should be a recent master graduate in computer science, physical or mathematical, preferably with a background in HPC.

Bibliography

Acosta M. et al, Performance analysis of EC-Earth 3.2.0: Coupling - Barcelona Supercomputing Center - Public Wiki, 2016.

Acosta M., Profiling and computational performance of IFS using BSC Tools - Barcelona Supercomputing Center - Public Wiki, 2017.

Alexandrov V. Scalable Stochastic and Hybrid Methods and Algorithms for Extreme Scale Computing, Procedia Computer Science, 2014.

Bastoul C. et al, Putting polyhedral loop transformations to work, Languages and Compilers for Parallel Computing., 2004.

Broekema C. et al, Exascale High performance Computing in the Square Kilometer Array, AstroHPC, 2012

Casanova H. et al, Parallel algorithms, CRC Press, 2011

Doblas-Reyes F. et al, Seasonal climate predictability and forecasting: status and prospects. Wiley Interdisciplinary Reviews: Climate Change,2013.

Dongarra J. et al, Applied Mathematics research for Exascale Computing, Report, US Department of Energy, 2014

Flato G. et al, Evaluation of climate models. In Climate change 2013: the physical science basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, 2013

Guest, M. The scientific case for high performance computing in Europe 2012-2020. Technical report, FP7 Project PRACE RI-261557, 2014.

Hartmann D. et al, Observations: Atmosphere and Surface, Cambridge University Press, 2013.

Hazeleger W., EC-Earth, science and implementation plan, Internal document, 2007

Irony D. et al, Communication lower bounds for distributed-memory matrix multiplication, Journal of Parallel and Distributed Computing, 2004.

Lenton T., Land and ocean carbon cycle feedback effects on global warming in a simple Earth system model. Tellus, 2003

Li G. et al, Tropical biases in CMIP5 multimodel ensemble: the excessive equatorial cold tongue and double ITCZ problems, J. Clim., 2014

Liu W. et al, Extended reconstructed sea surface temperature version 4 (ERSST.v4): part II. Parametric and structural uncertainty estimations. Journal of Climate, 2015.

Mauritsen T. et al, Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models, Nat. Geosci., 2015

Merchant C. et al, Uncertainty information in climate data records from Earth observation. Earth System Science Data Discussions, 2017.

Mudryk L. et al, Snow cover response to temperature in observational and climate model ensembles. Geophysical Research Letters, 2017.

Palmer T. et al, Toward Seamless Prediction: Calibration of Climate Change Projections Using Seasonal Forecasts, Bull. Amer. Meteor. Soc., 2008

Palmer T., A nonlinear dynamical perspective on climate prediction, J. Clim., 1999

Palmer T., Build high-resolution global climate models, Nature, 2014

Palmer T., Build imprecise supercomputers, Nature, 2015

Pavel K. et al, Changing computing paradigms towards power efficiency, Philosophical Transactions A, 2014.

Prodhomme C. et al, Benefits of resolution increase for seasonal forecast quality in EC-Earth, Journal of Climate, 2016.

Randall D. et al, Climate Models and Their Evaluation in the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, 2007

Rauser F. et al, Rethinking the default construction of multimodel climate ensembles, Bull. Am. Meteorol. Soc., 2015

Reichler T. et al, Uncertainties in the climate mean state of global observations, reanalyses, and the GFDL climate model. Journal of Geophysical Research, 2008.

Shepherd T. Atmospheric circulation as a source of uncertainty in climate change projections, Nat. Geosci., 2014

Stevens B. et al, What are climate models missing? Science, 2013

Stocker T. et al, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge Univ. Press, 2013

Taylor K. et al, An overview of CMIP5 and the experiment design, Bull. Am. Meteorol. Soc., 2013

Tinto O. et al, Detecting, analyzing and optimizing MPI communication bottlenecks in Earth System models, Journal of Computational Science, 2018

Yepes X.. et al, Performance analysis of EC-Earth 3.2.0: Using a new metric approach, Barcelona Supercomputing Center - Public Wiki, 2016.

Yepes X. et al, Simulation-based Performance analysis of EC-Earth 3.2.0 using Dimemas, Barcelona Supercomputing Center - Public Wiki, 2017.

Zhang X. et al, Double ITCZ in coupled ocean-atmosphere models: from CMIP3 to CMIP5, Geophys. Res. Lett., 2015

2. IMPACTO ESPERADO DE LOS RESULTADOS - *EXPECTED RESULTS IMPACT*

Social/economic impact

OEMES has the potential to increase the capability and comprehensiveness of global ESMs, in order to represent an ever-increasing realism and detail of new scenarios for our future climate and reduce the uncertainties in the study of climate. Such capability is a fundamental asset for policy makers as governments are debating at present how to generate reliable in-house predictions and projections of climate change. Something which is a prerequisite to support the development of national adaptation and mitigation strategies. Because 5 years is within the lifespan of a political career, potentially unpopular measures aimed at reducing emissions need to be backed by reassurance from the climate community that they will be effective. Results will have to be usable towards the public opinion to demonstrate the effectiveness of such measures in terms of creation of effective mitigation startegies.

Moreover, being able to provide reliable climate preditions using higher resolutions combined to an optimized earth system model from the computational point of view translates into a reduction of the energy cost of generating simulation results. Such cost is estimated here to be millions € for Spain alone according to the cost energy of supercomputers (1.2 millions € for Marenstrum IV per year for example) . In this sense, OEMES offers a potentially significant return on investment for the Spanish society where less energy will be needed to do the same work. Furthermore, OEMES will generate a methodology which allows to highlight the real problems of the ESMs to scale in an easy way that can be used to continuously advise policy makers in planning the transition to a future sustainable.

At present, though there are several institutions working in HPC for ESMs, there is none in the community besides BSC who has the appropiate tools to provide the profiling analysis needed to optimize and validate the ESMs in an efficient way for the future exascale platforms. Moreover, BSC is one of the few institutions around Europe with the capability to take advantage of the new ultra high simulation results for climate predictions. This places BSC and Spain in a strategic position at the forefront of the international climate science community. OEMES will contribute to enhance Spain's role within the EU as plans of mitigations strategy that will be decided organically within the union.

Scientific impact

OEMES fosters the integration of advanced HPC techniques specifically oriented for the assessment and modelling of weather and climate predictions with the idea of testing innovative technological options and strategies to improve our climate prediction. OEMES will cause a step forward in the development of numerical models for massive parallel platforms and its application to produce ultra-high resolution simulations, transferring knowledge on a short, medium and long-term to scientific community and environmental institutions. This will offer to the climate science community valuable indications on which parts of the code should be given priority for future exascale platforms. From an institutional strategic point of view, at the national level, the scientific community is lacking a research group dedicated to do profiling analysis and optimizations of ESMs. OEMES is meant to sustain the PI's efforts in establishing a permanent and successful group focused on this topic, providing also an optimized ESM which can be use for the CP group at BSC and AEMET for improved climate predictions. At the international level, within the EC-Earth community, HPC activities are also being carried out without a clear coordination and lacking a leadership. If it is funded, OEMES will help the PI and BSC to become the natural reference for these activities within the EC-Earth community.

Dissemination and Internationalization of results:

At least two manuscripts will be submitted to international journals, of high-impact whenever possible. Being the outcome of public research activity, results will not be subject to commercial use. Instead, results will be widely presented in scientific conferences, from the conferences related to EC-Earth users (e.g. EC-Earth and NEMO meetings) to general conferences (e.g AGU, EGU, JLESC, H2020 assamblies). Frequent video-conferences will be organized with our international collaborators to discuss results and to plan future collaborations. The results of OEMES will feature at meetings of the TWG of NEMO and EC-Earth where we are members, and the leaders of IFS and OASIS code development, who we are collaborating. They will also contribute to the PRIMAVERA and CMIP meetings through the CP group, where several members are part.

Communication and popularization of science:

To respond to the demand of the general public on information about topics such as global warming or gasses emissions, what changes to expect in the near-term and how to mitigate their effects, we will organize at least two presentations intended to non-specialists, as already done by our group in the past (e.g. the leader of the CP, V. Guemas at a local public library). Results from OEMES will also be accessible through our group website. Moreover, members of the Climate Prediction and Computational Group often receive extensive press coveragein national and international journals (e.g. El Periodico, Scientific American, La Vanguardia) also thank to a dedicated Communication Unit at BSC that is responsible for media relations, digital communication including the BSC website and social media. OEMES will benefit from the experience of this unit and the ever expanding network being created with members of the press, TV and radio dedicated to the general public.

Transfer of results:

The project deliverables and the new optimized code will be shared at national level with AEMET and at international level with EC-Earth, NEMO, IFS and OASIS users, which are the natural user of the weather and climate forecast information, to keep them informed about the progress of this sort of predictions and HPC development, remembering that we are talking about more than 24 institutions around Europe (see section 1.2). Furthermore, although a specific transfer of knowledge is not planned at this stage, BSC can count on the Earth System Services group which is dedicated to bridge the gap between climate information and end users in key sectors of society (energy, urban development, infrastructure, transport, health andagriculture) via tailored services to societal actors. They develop these services in collaboration with public administrations, private companies, funding agencies and spin-off companies that could exploit operational opportunities. OEMES will count on the expertise of this group to explore any possible way to exploit the results generated.

3. CAPACIDAD FORMATIVA - *TRAINING CAPACITY*

Training program

The Earth Sciences Department of BSC offers the opportunity for prospective PhD students to work in an international multidisciplinary scientific environment. In the context of this project we would like to recruit and enroll a PhD student in the Computer Science doctoral program (https://www.uab.cat/web/postgrau/doctorats/tots-els-doctorats/informacio-general/informatica) of the University Autonoma of Barcelona (UAB), which was recognized with MEC Excellence Mention RD1393/2007. **The goal for the PhD student could be, taking into account the main bottlenecks localized by our novel methodology, which of these areas could be ported to GPUs for a hybrid computation (CPUs+GPUs), and develop this approach instead of optimizing the CPU code. This should be a very attractive goal for the student and the community, since accelerators are also very important for ESMs and an alternative solution. Additionally during the PhD, a visit of three months could be done to the ECMWF and the supervisor would be Nils Wedi, thanks to the open collaboration and similar to the postdoctoral visit done by the PI in 2017.**

The background of the PI (supervisor of one PhD thesis and one master thesis), the experience of others PhDs in the Research Group and the head of the computational group provides a great opportunity for training new experts (see CVs attached for more research information). The training plan will be oriented towards the acquisition of competences in HPC applied to Earth Science. The student will be involved in the research group activities, including periodical internal and external seminars, journal clubs, and invited speaker conferences held at BSC. In addition, he/she will have the opportunity to share the results of his/her research in different national and international scientific forums and he/she will be encouraged to co-author scientific research articles of high impact. Complementary skills required for efficient research execution and communication will be fostered through the student participation on multiple training activities offered by BSC.

Selected Previous PhD students in the group (last 5 years):

1. **Vincenzo Obiso**, "Assessment of Dynamic Aerosol-Radiation Interaction in Atmospheric Models". 07/03/ 2018.
2. **Lluís Vendrell,** November 2017, "Modeling the dust life cycle and its associated meteorological processes from global to regional scales".
3. **Víctor Manuel Valverde Morales**. "Characterization of atmospheric pollution dynamics in Spain by means of air quality modelling". 08/04/2016. Publications: Valverde et al. (2015, Int. J. Clim.); Valverde et al. (2016a, 2016b, Sci.Total Env.).
4. **Raül Marcos**, "Improvement of seasonal forecasting techniques applied to water resources and forest fires" February 2016,
5. **Robert** Banks. "Assessment of planetary boundary-layer schemes with advanced remote sensing instruments and air quality modelling". 04/04/2016. Publications: Banks et al. (2015, Bound.-lay. Meteorol.); Banks and Baldasano (2016, Sci.Total Env.); Banks et al. (2016, Atm. Res.)
6. **Luis Lage Rodrigues**. "Calibration and combination of seasonal climate predictions in tropical and extratropical regions". 22/01/2016. Publications: Lage Rodrigues et al. (2014, Clim. Dyn.); Lage Rodrigues et al. (2014, J. Geophys. Res. Atm.);
7. **Danila Volpi.** "Benefits and drawbacks of different initialization techniques in global dynamical climate predictions". 01/03/2015. Publications: Volpi et al. (2017a, Clim. Dyn.); Volpi et al. (2017b, Clim. Dyn.); Carrassi et al. (2016, Clim. Dyn.); Carrassi et al. (2014, Nonl. Proc. In Geophys.);
8. **Alba Badia Moragas.** "Implementation and development of a gas-phase chemical mechanism within the global/regional atmospheric model chemical non-non-hydrostatic multiscale model (NMMB/BSC-CHEM)". 12/12/2014. Publications: Badia and Jorba (2015, Atm. Env.); Badia et al. (2017, Geosc. Model Dev.).
9. **Simone Marras**. "Variational Multiscale Stabilization of finite and spectral elements for dry and moist atmospheric problems ". 10/12/2012. Publications: Marras et al. (2012a, 2012b, J. Comp. Phys.); Marras et al. (2013, J. Comp. Phys.);

10. **Karsten Haustein**."Development of an atmospheric modeling system for regional and global mineral dust prediction". 31/01/2012. Publications: Haustein et al. (2012, Atm. Chem. Phys);
11. **Sara Basart Alpuente**."Mineral dust model validation through ground based and satellite observation in North Africa and Europe" 30/01/2012. Publications: Basart et al. (2009 Atm. Chem. Phys.), Basart et al. (2012 Atm. Chem. Phys.), Basart et al. (2012 Tellus);
12. **María Teresa Pay Pérez**. "Regional and urban evaluation of an air quality modelling system in the European and Spanish domains". 22/11/2011. Publications: Pay et al. (2010ab, Atm. Env.), Pay et al. (2011, Atm. Env.), Pay et al. (2012, Atm. Env.).

Current students:

1. Verónica Torralba. "Seasonal climate prediction for the wind energy sector: methods and tools for the development of a climate service". Expected date: November 2018.
2. Jaime Pérez Benavides. "Development and evaluation of an air quality modelling system over Barcelona: from regional to street scale". Expected date: July 2019.
3. Daniel Rodríguez Rey. "Climate Change: climate prediction. Different time scales and from regional to global" Expected date: July 2020.
4. Bianca Mezzina. "ENSO influence on the North Atlantic-European winter: mechanisms and implications for predictability". Expected date: November 2020
5. Oriol Tintó. "Computational Improvement of ocean models: The path to a more efficient NEMO model". Expected date: June 2019 (**Supervised by the PI Mario Acosta**)

Scientific and professional development of previous students

Former postdocs and Ph.D. students hosted at the ES-BSC hold/have held positions in several well known scientific institutions and energy companies around the globe, such as:

María Teresa Pay - Assistant professor at Universitat de Barcelona (Departamento de Genética, Microbiología y Estadística); Karsten Hausetin - Associate Researcher at Stanford University (Department of Geophysics; Alba Badia - Research Associate at University of East Anglia (School of Environmental Sciences); Robert Banks - Researcher at Delft University of Technology (Department of Geoscience and Remote Sensing); Victor Valverde - Researcher at Joint Research Center; Luis Ricardo Lage Rodrigues - permanent position at the Center for Earth System Sciences (CCST) from the Brazilian National Institute for Space Research (INPE), Danila Volpi, research scientist in Météofrance (Toulouse, France).

Scientific and formative context:

The Earth Sciences department at BSC (BSC-ES) conducts multi-facet research in Earth system modelling. The designation of Prof Francisco J. Doblas-Reyes as Director of the BSC-ES in 2014 initiated the merging of the BSC-ES with the Climate Forecasting Unit of the Institut Català de Ciències del Clima (IC3-CFU), which he was leading at the time and who had become in a short time a main European actor in the development of climate predictions and climate services.

The BSC is a highly productive scientific entity that has published more than 150 research articles in peer reviewed journals over the last 5 years, including 5 in prestigious high-impact journals and with a very dense international collaborative network counting at least 50 institutes worldwide. During last 6 years (2012-2018), BSC-ES was granted 19 EU H2020 projects, 5 EU Copernicus projects, 7 projects funded by the Ministerio de Economía y Competitividad (MINECO), 2 projects funded by the European Space Agency, 1 project funded by the French Ministry of Sciences and 3 projects from ERA-NET ERA4CS. In the same period, BSC-ES participated in 21 RES and 4 PRACE projects.

BSC-ES department international activity includes the coordination of the two World Meteorological Organisation (WMO) regional centres specialised in sand and dust warning and forecasting, as well as the participation in climate services initiatives like the Climate. Additionally, the BSC has been awarded with the Severo Ochoa's Centre of Excellence project of the Spanish government since its first call (2011). Also the BSC has been awarded by the beneficiary of Marie Skłodowska-Curie Action COFUND program for postdoctoral fellows (STARS; H2020-MSCA-COFUND-754433).

4.IMPLICACIONES ÉTICAS Y/O DE BIOSEGURIDAD - *ETHICAL AND/OR BIOSAFETY IMPLICATIONS*

Not applicable