Project: **HiResClim: High Resolution Ensemble Climate Modeling**
Research field: **Global Climate Modelling**
Project leader: **Dr Colin Jones**, **Swedish Meteorological and Hydrological Institute**, **Research Department, Rossby Centre**, **Sweden**

Research project: Science Aims and description

Over the past few years the international climate modeling community has collaborated around the 5th Coupled Model Intercomparison Project (CMIP5, http://cmip-pcmdi.llnl.gov/cmip5 and Taylor et al. 2012), producing a coordinated set of simulations to support the IPCC 5th Assessment Report. For the first time, CMIP5 promotes 2 experiment protocols; climate **projections** and **initialized** climate **predictions**. The climate projection protocol follows standard methods for estimating future climate conditions. In short; a coupled GCM is integrated for many hundreds of years, using estimates of pre-industrial greenhouse gas (GHG) and aerosol concentrations, solar intensity, volcanic aerosols and land-use. Once the model attains a stable equilibrium with the forcing, a number of model states are saved that sample the simulated climate variability. Each of these states is used as an initial condition to extend the pre-industrial simulation through the historical past (in CMIP5 1850-2005), using observation-based estimates of GHGs, aerosols, land-use, volcanoes and solar variability. These simulations are extensively evaluated against observations to determine the overall quality of present-day GCMs. The historical runs are then extended into the future, using different GHG, aerosol and land-use scenarios that sample a range of plausible socio-economic futures up to ~2100. These are referred to as Representative Concentration Pathways (RCPs, Moss et al. 2010). Application of the same RCPs to different GCMs for the period 2006-2100 supports the generation of a multi-GCM ensemble of climate projections. Climate prediction differs from projection by using observation-based initial conditions to start a GCM simulation from a known, recent date, with the initialized ocean providing predictability beyond typical atmospheric timescales. Climate prediction has recently emerged as a possibility due to the convergence of a number of important improvements in; (i) ocean observational coverage (ii) GCM simulation quality (iii) our understanding of processes providing inter-annual predictability.

Both HiResClim GCMs contributed to the projection and prediction protocols of CMIP5, using atmospheric spectral resolutions of; ARPEGE (T127, ~150km), EC-Earth (T159, ~125km). Both models used the NEMO ocean at a resolution of ~1° (~0.3° near the equator). While such resolutions compare favorably with other CMIP5 models, they remain poor in terms of model resolution required for an accurate simulation of important modes of climate and weather variability, such as; the Gulf Stream and Atlantic Meridional Overturning Circulation (AMOC), the El Nino–Southern Oscillation (ENSO), the Madden-Julian Oscillation (MJO), monsoon circulations, mid-latitutde blocking anticyclones and both tropical and extra-tropical cyclones. Potential changes in these modes of climate and weather variability, in response to future GHG and aerosol forcing, will likely impact society and ecosystems far more than a slow change in the time-mean climate. It is therefore important GCMs realistically simulate such variability. In HiResClim we will address this through a set of coordinated experiments, based on the CMIP5 protocols, employing significantly increased resolution in the two models. Specifically, the ARPEGE atmosphere will use T359 and 31 vertical levels and EC-Earth T511 and 91 levels. Both models will use the NEMO ocean, at 0.25° horizontal resolution and 75 vertical levels.

In a core set of experiments both models will repeat a CMIP5 pre-industrial control run, of ~700 simulated years, at high resolution. Ten different model states from each of the GCM pre-industrial runs will then initialize a 10-member ensemble of CMIP5 historical simulations (1850-2005) with each GCM. Each historical run will spawn 2 future projections, following, respectively the RCP 4.5 and 8.5 CMIP5 scenarios, resulting in 20 projections with each GCM covering 2006-2100. Such a large ensemble is necessary to establish robust estimates of future changes in the important, but highly variable, weather and climate phenomena listed above. HiResClim, therefore, directly address two of the leading challenges for future climate modeling; higher model resolution and increased ensemble size. Compared to the original CMIP5 simulations, horizontal resolution is increased by a factor of ~3 (ARPEGE) and ~3.5 (EC-Earth), with an increased number of vertical levels in the ocean (both models) and atmosphere (EC-Earth) also being important for improved simulation quality.

Core results will all be saved in the CMIP5 archives, thus becoming openly available to the wider research community. An important question to address with the HiResClim ensemble is the benefit such an increase in model resolution implies for simulating the aforementioned climate and weather variability and, in particular, whether future estimates of this variability differ when higher model resolution is employed. The core set of experiments are also an important prerequisite to planned, high-resolution initialized climate predictions/hindcasts, to be made with both models in the recently-funded EU FP7 project SPECS. In particular, a high-resolution EC-Earth present-day coupled climatology is required for the observation-based anomaly initialization procedure developed for EC-Earth at SMHI. In SPECS, such high-resolution, anomaly-initialized EC-Earth predictions will be compared to predictions using full-field initialization with the same model. The ensemble of historical CMIP5 simulations also provide an important set of uninitialized model runs, against which predictability in the initialized hindcasts, made with the same model for the same time period, can be evaluated. The high-resolution, initialized hindcasts, planned in SPECS, will also be compared to lower resolution hindcasts already made with the two models in CMIP5.

As well as the core experiments, we also define a set of optional experiments, important for improving our future ability to make reliable climate simulations. As a result of the number of institutes in HiResClim using EC-Earth, the optional runs will primarily be made with this model, using a T511L91/0.25°L75 configuration. One set of experiments will use sea surface temperatures (SST) and sea ice concentrations (SIC) derived from the original EC-Earth CMIP5 projections, made at T159L62/1.0°L45. From 5 of these runs, covering 1950-2100 and following both RCP 4.5 and 8.5, we will extract monthly SST and SIC values and using observations, monthly mean, climatological bias-corrections for 1979-2010 will be calculated for each EC-Earth SST and SIC field. The bias correction fields will then be applied to all monthly EC-Earth SST and SIC fields for the period 1950-2100 of each ensemble member, resulting in 5 sets of monthly, bias-corrected SST and SIC fields, consistent with the original T159/1.0° runs. These fields will be applied as a lower boundary condition for the T511L91 version of EC-Earth and a set of atmosphere-only, historical and projection runs made. Comparison of these with the original T159/1.0° runs will indicate the benefit of increased atmospheric resolution in simulating important modes of climate and weather variability. Comparing the atmosphere-only runs to the T511L91/0.25°L75 coupled runs will indicate the importance of coupling a high-resolution ocean model.

In the second year of the project, the T511L91/0.25°L75 version of EC-Earth will perform a control simulation starting from a restart file representative of present-day conditions, derived from an earlier projection run. This simulation will be for 300-500 years and use GHG and aerosol forcing fixed at present-day values, producing a high-resolution, long-term equilibrium realization of the present climate. Finally, depending on computational throughput in year 1 of the project, in the second year we will extend the pre-industrial control simulation for a further ~300 years. Analysis of both the extended pre-industrial and present-day forcing runs will aid in understanding slower modes of variability in the climate system and the ability of EC-Earth to accurately simulate these modes.

Describe how you will manage the resources requested?

As both models have already been run on the requested system, the planned simulations can begin early in the project. The first year is dominated by the core 700-year control experiments of both models. These are necessary for the subsequent historical and future simulations. Based on earlier runs, we estimate the pre-industrial runs will take ~175-210 wall-clock days of computing. We therefore expect these jobs will take the entire 1st year and may stretch into the beginning of year 2. Model states for use as initial conditions for the historical runs will be drawn from the last ~150-200 years of the control runs, hence the first historical runs can likely start towards the end of year 1,. CERFACS (ARPEGE) and SMHI (EC-Earth) will be responsible for the control simulations, monitoring their progress and taking care of data produced. The mode of job-integration is detailed in response to Question 6.

In year 1, atmosphere-only runs, using bias-corrected SST/SIC fields will begin. These are independent of the 700-year control and of each other. They can, therefore, be run in parallel. IC3 will be responsible for these. Late in year 1 and the first half of year 2, the 10 historical simulations will be made. These runs are independent of each other and can also run in parallel, subject to availability of compute nodes. One historical run we requires ~70-100 wall-clock days. Depending on the extent we cane parallelize these jobs we estimate they take 8 months, if access to compute nodes becomes a problem, the planned ensemble size will be reduced. CERFACS will make all ARPEGE runs, SMHI and KNMI will share the EC-Earth runs. The RCP scenarios are planned for months ~16-24. If necessary the ensemble size of the RCP runs will also be reduced. CERFACS makes the ARPEGE runs, KNMI and SMHI share the EC-Earth runs. Depending on experiences in the first year of the project, the remaining EC-Earth optional experiments will be made in year 2. SMHI makes the pre-industrial extension and IC3 the present-day control run. Figure 1 illustrates approximate timings of both the core and optional experiments in HiResClim.

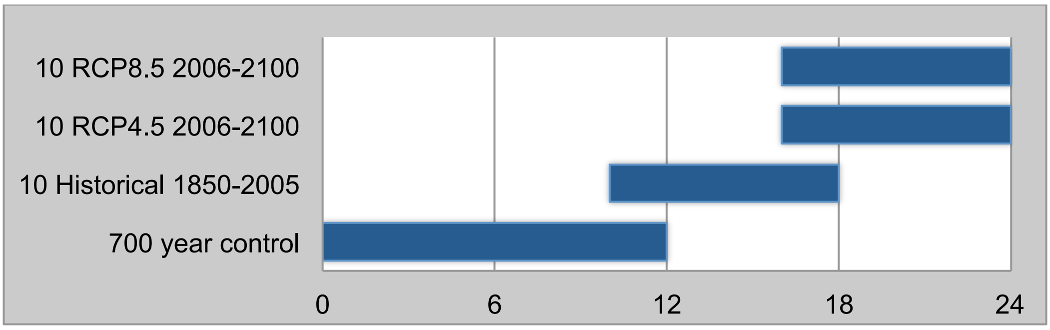

Figure 1. Approximate timing of core experiments in HiResClim

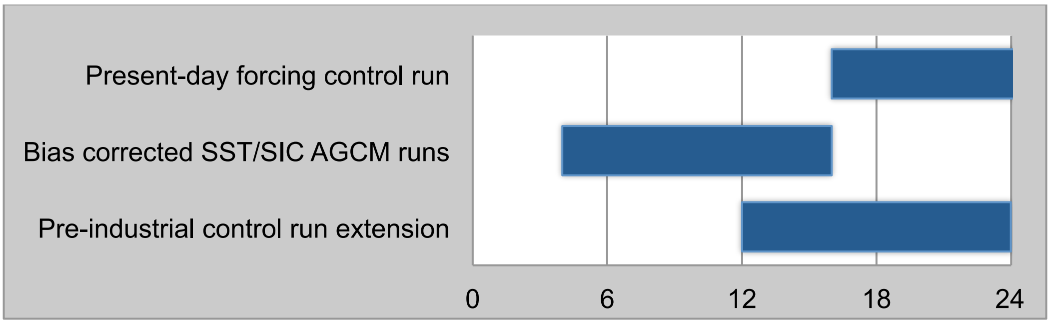

Figure 2. Approximate timing of optional (EC-Earth) experiments in HiResClim

Describe the numerical methods and algorithms that you are planning to use, improve, or develop, the codes, packages or libraries that you need to undertake the project, and how these will enable the research to be achieved. (1 page)

The atmosphere-land component of ARPEGE is the operational weather forecast model of Météo-France, while EC-Earth is based on cycle 36R4 of the ECMWF Integrated Forecast System (IFS), also used in System 4 for seasonal forecasting at ECMWF. The basic code infrastructure of IFS and ARPEGE has been shared for the past ~15 years allowing close collaboration on development targeting optimum performance on a range of parallel computing systems. The main concern for optimum use on Tier-0 machines is the problem size, combined with the ability to run numerous independent ensemble members in parallel. The resolution proposed for HiResClim (ARPEGE: T359, ~300.000 horizontal grid points and EC-Earth: T511, ~425,000 points) will help efficiently share calculations between 1000-1500 sub-domains, increasing the range of efficient compute-core usage per model executable. The need to make a large (~10 members per model) ensemble of simulations, each independent of the other, also greatly increases the efficient use of parallel architectures.

NEMO, developed by a European consortium, is the ocean component of both the ARPEGE and EC-Earth models, while both models use the OASIS coupler, developed and maintained by CERFACS, for coupling atmosphere-land and ocean models. NEMO has already been adapted for Tier-0 machines by several is-ENES laboratories, including all 4 in this consortium, in collaboration with the PRACE-1IP support team. This group was created to optimize a very high-resolution (T799/ORCA025) version of EC-Earth. On different architectures, we intended to optimize the core mapping of MPI tasks (John Donners, SARA), implement dynamical memory allocation and load-balanced domain decomposition (Andrew Porter, STFC) or design optimal coupling interface and strategy (CERFACS). The chosen problem size for NEMO (ORCA025, 1442 x 1021 horizontal grid points) enables distribution to 1000 sub-domains. An important development for HiResClim is that we plan to use the latest version of OASIS coupler (OASIS-MCT), which enables a direct and parallel exchange of coupling fields between every model process, avoiding the previous coupler-centric bottleneck. HiResClim provides an opportunity to measure OASIS-MCT performance in 2 modeling systems running at high-resolution. This will provide important information to a number of other European modeling groups also using the OASIS software.

In HiResClim the minimum amount of necessary diagnostics will be stored on disk and immediately post-processed for transfer to final local storage, through the RENATER network. During the project, thanks to CERFACS/IPSL collaborations (ANR PULSATION Project, Curie thin node Grand Challenge) the newly developed XIOS IO server may be tested in ARPEGE-NEMO, including an option to save diagnostic data on a reduced grid. Experience from this effort will inform plans by other European modeling groups to investigate the XIOS IO server within the EU project is-ENES2.

Explain why this project needs to run on a Tier-0 system, why the machine requested is suitable and how use of the system will enable the science proposed.

This proposal is an important step in furthering the ENES (European Network for Earth System Modeling) preparations for efficient use of Tier-0 and beyond HPC systems. The EU FP7 large-scale coordinating project, is-ENES2, due to start in late 2012, has as one central goal the technical development of European Climate and Earth System Models for efficient application on high-end HPC systems. In HiResClim two of the ENES models will begin to address this aim, concentrating on the 2 leading science requirements for accurate and reliable climate forecasting, namely significantly increased model resolution and the use of a large ensemble of simulations. The former is critical for models to realistically represent the observed power spectrum of climate and weather variability, while the latter is necessary for developing robust estimates of future change in this variability and for determining near-term climate predictability. In HiResClim the significant increase in model resolution, combined with the required reduction in model timestep, increases the computational cost of 1 simulated year by ~25 (ARPEGE) and ~35 (EC-Earth), compared to that recently experienced for CMIP5. Combined with the aim to develop a suitable ensemble size and the need for long coupled control runs for generating ensemble initial conditions, the proposed set of core experiments are simply not feasible on Tier 2 (local) or Tier 1 (national) facilities. They are most efficiently addressed on a major Tier 0 system. In year 1 of the project we envisage requiring ~ 38 million core hours on the Curie Thin Node system, increasing to ~ 112 million core hours in year 2, with a minimum of 2 jobs (~4-5 thousand cores) and a maximum of up to 10 jobs (~20-30 thosuand cores) running in parallel for extended periods (e.g. of a few months in duration). Both models have been extensively tested in terms of technical performance (short integration on climate timescales) on the Curie Fat Node system (EC-Earth) and both Curie architectures (ARPEGE). Figure 3 and Table 1, provide scaling details of the ARPEGE-NEMO T359/ORCA025 model on the Curie Thin nodes over the core range 128 to 8192. Figure 4 and Table 1 provide similar statistics, derived from the is-ENES-PRACE1IP evaluation of EC-Earth at T799/ORCA025 on the Curie Fat Node system. Here scaling is presented between 650 and 3500 cores for a single model job.

# cores	Absolute timing (s) per 1 month long run	speedup
128	10158	1
256	5190	1.95
512	2926	3.47
1024	1865	5.44
2048	1776	5.71
4096	2453	4.14
8192	2324	4.37

Table 1: Absolute timings for a 1 month simulation of the ARPEGE-NEMO T359/0.25° configuration on the Curie Thin node system

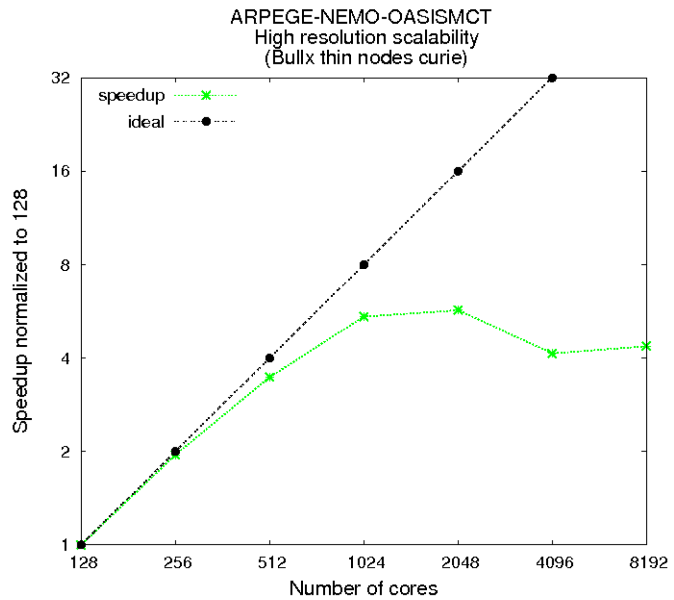

Figure 3: Speed up of ARPEGE-NEMO T359/0.25° configuration relative to 128 core performance on the Curie thin node system. Also shown is a perfect scaling line

No. of cores	Secs per model timepstep			Time for One day simulation	1-day speed up scaled to 650 cores	Ideal Speed-up
	All step average	I/O steps average	Compute steps ave			
650	3.20	5.92	2.55	383.88	1.00	1.00
778	2.55	4.76	2.03	306.09	1.25	1.20
906	2.15	4.05	1.72	258.55	1.48	1.39
1034	1.90	3.54	1.53	228.24	1.68	1.59
1162	1.74	3.24	1.40	209.01	1.84	1.78
1290	1.59	2.95	1.28	190.52	2.01	1.98
1418	1.44	2.68	1.17	173.19	2.22	2.18
1546	1.35	2.53	1.09	161.82	2.37	2.38
1610	1.38	2.62	1.11	165.07	2.33	2.48
1866	1.31	2.62	1.02	157.59	2.44	2.87
2634	1.20	2.86	0.83	144.43	2.66	4.05
3402	1.21	2.87	0.81	145.08	2.65	5.23
3482	1.17	2.66	0.82	139.82	2.75	5.36

Table 2. Timings of EC-Earth T799/0.25° on the Curie Fat Node system, relative to 650 cores. Also shown are the average time (in secs) for a single model timestep (All step) and for only I/O steps and only compute steps (with no I/O occurring).

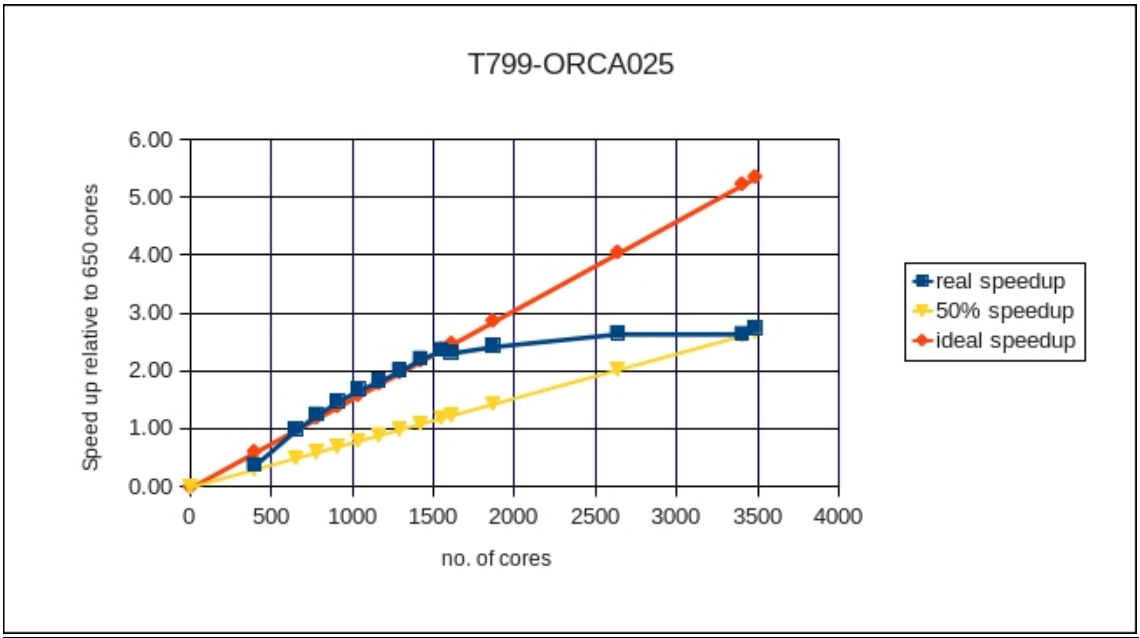

Figure 4: Speed up of EC-Earth T799/0.25° configuration on the Curie Fat Node system expressed relative to 650 cores. Also shown is ideal speed-up and 50% speed up curves

It is important to emphasize that in HiResClim many of the planned simulations are independent of each other, allowing an ensemble of simulations to be bundled into a single executable for highly efficient use of parallel architectures. IC3 recently tested the autosubmit system they have developed, on a Cray XE6, achieving efficient use of ~10,000 cores by a single executable containing an ensemble of high-resolution EC-Earth integrations. The autosubmit option will be used primarily in year 2 of the project, where 20 historical runs (10 per GCM) can be run in parallel, spawning 40 RCP future projection runs (2RCPs x 10 members per GCM) all of which can also integrate in parallel. Experience gained in making such a coordinated, 2-GCM ensemble of long, high-resolution coupled climate simulations will be of enormous importance to the wider ENES community. In particular, this experience will feed directly into a major JRA in is-ENES2, that brings together the majority (5-6) of European GCMs around a common multi-GCM, multi-member integration protocol to address bottlenecks to efficient GCM use of high-end HPC systems. A second EU FP7 large-scale project, SPECS, due to begin in late 2012, has a primary focus on improving European seasonal to decadal climate prediction capabilities. SPECS also has a work package addressing the development and application of a high-resolution multi-GCM, multi-member ensemble of seasonal to decadal predictions. In contrast to is-ENES2, the SPECS activity addresses scientific and predictability benefits of increased model resolution combined with an ensemble approach. All 5 institutes in HiResClim are partners in is-ENES2, with two of them (SMHI and CERFACS) co-leading the JRA on high-resolution, multi-GCM, multi-member modeling. Four of the HiResClim institutes are also partners in SPECS, with IC3 being the coordinating institute. SMHI leads the SPECS workpackage on improving high-resolution, multi-model approaches to climate prediction.

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

All members of the project are active participants in the is-ENES and is-ENES2 EU FP7 projects and have more than 2 decades experience developing and running GCMs on a range of HPC systems. Is-ENES is the main European collaboration on the technical development of European climate models. Both of the project models made extensive contributions to the climate projection and prediction protocols in CMIP5. ARPEGE-NEMO simulations were a collaboration between CERFACS and CNRM-Météo-France, while KNMI, SMHI and IC3 were 3 of the most active EC-Earth institutes. The partners have extensive collaborative experience in coupled climate modeling, contributing to a large number of past and on-going EU projects, including; DEMETER, PREDICATE, DYNAMITE, PRUDENCE, ENSEMBLES, COMBINE, is-ENES, EMBRACE (coordinator: Jones, SMHI), as well as 3 large-scale projects due to begin in late 2012: SPECS (coordinator Doblas-Reyes IC3), EUPORIAS and is-ENES2.

CERFACS has developed the OASIS coupler for ~20 years (Valcke 2006; Redler et al. 2010) and are currently leading major improvements to the OASIS code (OASIS-MCT) to make optimal use of the PRACE tier-0 machines (Valcke). The OASIS-MCT version enables direct and parallel exchange of coupling fields between different component model processes, avoiding previous coupler-centric bottlenecks. Maisonnave and Valcke, within the FP7-IS-ENES project, have extensively worked on GCM coupling on HPC resources and provided support on OASIS-based coupling to several European laboratories. Last year CERFACS won PRACE preparatory access on Curie the coupled ARPEGE-NEMO configuration for this proposal and was developed targeted to Tier-0 machine (Maisonnave et al. 2011). In collaboration with PRACE-2IP, a study on NEMO model fault-tolerance compliance has been initiated (Maisonnave, 2012).

All 5 of the HiResClim institutes were involved in the IS-ENES/PRACE-1IP working group focusing on EC-Earth climate model adaptation to Tier-0 machines (Donners et al. 2012). The KNMI team developed the first climate version of EC-Earth (version 2.3) that was subsequently used for all CMIP5 simulations by the EC-Earth consortium. SMHI since led development of the latest version of EC-Earth (version 3) which will be used in this proposal. SMHI have recently tested a range of EC-Earth configurations, in the resolution range T255/511/799 on a number of HPC systems. IC3 lead development of the auto-submit system; a python-based wrapper software, which bundles numerous EC-Earth jobs into a single executable, performing job-control on this parallel set of simulations throughout their execution. Autosubmit will be evaluated within is-ENES2 by a number of ENES GCM groups. NSC has significant experience operating large-scale HPC resources as well as analyzing, tuning, parallelization and data storage of parallel applications. NSC, with SMHI, will host EC-Earth data through an ESG node.

Justify the number of core hours requested.

Table 3 lists the core and optional runs planned for year 1 of the project. These assume a restart frequency (single job run length) of 3 simulated months for both models. Table 4 lists the core experiments planned for year 2 of the project. Depending on progress in year 1, some year 2 simulations may be able to start towards the end of year 1. In year 1, all simulations can, in principle, run in parallel. In practice it is likely the two 700 year spin up runs operate in parallel and the 5 SST/SIC optional EC-Earth runs are made as a single executable using the autosubmit software. Similarly in year 2, the 20 historical simulations can operate in parallel, as can the subsequent 40 RCP projection runs.

Run type	# Runs	# Steps/Run	Walltime/Step	# CPU cores	Total core hours/Type Run
700yr control **core** ARPEGE-NEMO	1	2800	5328	2048	8,500,000
700yr control **core** EC-Earth	1	2800	8892	2048	14,250,000
5 SST/SIC EC-Earth 1950-2100 **optional**	5	3000	8892	2048	15,250,00
Total					**38,000,000**

Table 3: Core and optional runs planned for year 1.

Run type	# Runs	# Steps/Run	Walltime/Step	# CPU cores	Total core hours/Type Run
10 Historical (1850-2005) **core runs** ARPEGE-NEMO	10	620	5328	2048	19,000,000
10 Historical (1850-2005) **core runs** EC-Earth	10	620	8892	2048	31,400,000
20 RCP projection core runs (2006-2100) ARPEGE-NEMO	20	380	5328	2048	23,100,00
20 RCP projection core runs (2006-2100)	20	380	8892	2048	38,500,000
Total					**112,000,000**
Extend 700 year pre-industrial run: EC-Earth, optional	1	1200	8892	2048	6,000,000
400 year present-day control: EC-Earth optional	1	1600	8892	2048	8,000,000

Table 4: Core runs planned for year 2.

References

Hazeleger, Wilco, and Coauthors, 2010: EC-Earth: A Seamless Earth-System Prediction Approach in Action. Bull. Amer. Meteor. Soc., 91, 1357–1363. doi:http://dx.doi.org/10.1175/2010BAMS2877.1

Donners, J., Basu, C., McKinstry, A., Asif, M., Porter, A., Maisonnave, E., Valcke, S., Fladrich, U., 2012: Performance analysis of EC-EARTH 3.1, PRACE White Paper, to be published

Maisonnave, E., Coquart, L. and Valcke, S., 2011: PRACE preparatory access for high resolution ARPEGE-NEMO porting on Bullx TGCC platform ,Working Note, WN/CMGC/11/77, SUC au CERFACS, URA CERFACS/CNRS No1875, France

Meurdesoif, Y., Denvil, S., Masson, S., 2012: Production et dissémination de données à l'IPSL, 29th ORAP Forum, 2012, Paris

Moss Richard H, Jae A. Edmonds, Kathy A. Hibbard, Martin R. Manning, Steven K. Rose, Detlef P. van Vuuren, Timothy R. Carter, Seita Emori, Mikiko Kainuma, Tom Kram, Gerald A. Meehl, John F. B. Mitchell, Nebojsa Nakicenovic, Keywan Riahi, Steven J. Smith, Ronald J. Stouffer, Allison M. Thomson, John P. Weyant & Thomas J. Wilbanks 2010: The next generation of scenarios for climate change research and assessment *Nature* 463, 747-756(11 February 2010) doi:10.1038/nature08823

Redler, R., Valcke, S. and Ritzdorf, H., 2010: OASIS4 - A Coupling Software for Next Generation Earth System Modelling, Geoscience Model Development, 3, 87 - 104, DOI:10.5194/gmd-3-87-2010.

Taylor, Karl E., Ronald J. Stouffer, Gerald A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. Bull. Amer. Meteor. Soc., 93, 485–498.
doi: http://dx.doi.org/10.1175/BAMS-D-11-00094.1