Project name	High-resolution regional dust reanalysis based on ensemble data assimilation techniques (eDUST)
Research field	PE10_1 Atmospheric chemistry, atmospheric composition, air pollution PE10_3 Climatology and climate change

Project leader

Title	Dr.
Last name	Basart
First name	Sara
Organisation name*	Barcelona Supercomputing Center – Centro Nacional de Supecomputación (BSC-CNS)
Department*	Earth Sciences Department
Group*	Atmospheric Composition
Country	Spain

*Example
- *University of Amsterdam*
- *Department of Physics*
- *Group of Molecular Physics*

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

The Earth Sciences Department at BSC (ES-BSC) has developed into a reference institution in Europe in the field of climate predictions, air quality and atmospheric composition modelling. The **Atmospheric Composition group** at ES-BSC aims at **better understanding and predicting the spatiotemporal variations of atmospheric pollutants along with their effects** on air quality, weather and climate and contributes to a variety of forecasting activities. The group has established several collaborations with national research teams: AEMET, CIEMAT, CSIC-IDAEA, CSIC-ICM, CEAM, Technical University of Catalonia, University of Granada and University of Murcia. On an international level regular collaborations are established with: University of Aveiro, National Observatory of Athens, National Meteorological Agency of Turkey, National Centers for Environmental Predictions (NCEP; USA), NASA Goddard Institute for Space Studies (GISS, USA), NASA Geophysical Fluid Dynamics Laboratory (GFDL, USA), Finnish Meteorological Institute (FMI, Finland), International Research Institute for Climate and Society (IRI, USA), University of California Irvine (UCI; USA).

A core activity of the group is **sand and dust storm modelling and forecasting** from regional to global scales. As a result of its excellence, the BSC hosts both the WMO SDS Warning Advisory and Assessment System (SDS-WAS), and more recently of the first Regional Specialized Meteorological Center for Northern Africa, Middle East and Europe with activity specialization on Atmospheric Sand and Dust Forecast, the Barcelona Dust Forecast (BDFC), in a close and successful collaboration with the Spanish State Meteorological Agency (AEMET). Moreover, from October 2016, ES-BSC is hosting an AXA Chair on Sand and Dust Storms. This 15-year research programme is not only intended to support the two WMO SDS Regional Centers based at BSC, but also to widen the scope and relevance of the mineral dust research at BSC.

List up to 5 recent relevant publications (in the last year)

Basart, S., L. Vendrell, J.M. Baldasano (2016). High-resolution dust modelling over complex terrains in West Asia. Aeolian Research, 23, 37-50, doi:10.1016/j.aeolia.2016.09.005.

Di Tomaso, E., Schutgens, N. A. J., Jorba, O. & Pérez García-Pando, C. (2016). Assimilation of MODIS Dark Target and Deep Blue observations in the dust aerosol component of NMMB/BSC-CTM version 1.0, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-206.

Gkikas, A., S. Basart, N. Hatzianastassiou, E. Marinou, V. Amiridis, S. Kazadzis, J. Pey, X. Querol, O. Jorba, S. Gassó and J.M. Baldasano (2016). Mediterranean intense desert dust outbreaks and their vertical structure based on remote sensing data. Atmospheric Chemistry and Physics, 16, 8609-8642, doi:10.5194/acp-16-8609-2016.

Huneeus, N., S. Basart, S. Fiedler, J.-J. Morcrette, A. Benedetti, J. Mulcahy, E. Terradellas, C. Pérez García-Pando, G. Pejanovic, S. Nickovic, P. Arsenovic, M. Schulz, E. Cuevas, J. M. Baldasano, J. Pey, S. Remy and B. Cvetkovic (2016). Forecasting the northern African dust outbreak towards Europe in April 2011: a model intercomparison, Atmospheric Chemistry and Physics, 16, 4967-4986, doi:10.5194/acp-16-4967-2016.

Pérez García-Pando, C., Miller, R. L., Perlwitz, J. P., Rodríguez, S., & Prospero, J. M. (2016). Predicting the mineral composition of dust aerosols: Insights from elemental composition measured at the Izaña Observatory. Geophysical Research Letters, 43(19).

For a complete list of publications of the Earth Sciences Department at BSC:
https://earth.bsc.es/wiki/doku.php?id=publications:publications

IMPORTANT NOTICE

Please **upload a single document**, using this template, in PDF format. This pdf file has to address **ALL of the points below**, and must not exceed 8 MB. The minimum font size that you can use is 12 point; the minimum margin is 2 cm. The **maximum number of pages you may submit is 15 pages, including cover page and references**. Reviewers will be instructed not to consider pages out of the limit.

Note. The information above can be deleted in the proposal

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

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

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

(**Section limit: 2 pages**)

Over the past decade there has been a growing recognition of the crucial role of **sand and dust storms (SDS)** on weather (e.g. Pérez et al., 2006), climate (e.g. Miller et al., 2014) and atmospheric chemistry (e.g. Bauer and Koch, 2005) and ecosystems (e.g. Jickells et al., 2005), along with their important **adverse impacts upon life, health, property, economy** and other strategic sectors (e.g. Pérez García-Pando et al., 2014; Rodríguez et al., 2001; Schroedter-Homscheidt et al., 2013; Stefanski and Sivakumar, 2009). Reacting to the concerns on SDS by its most affected member states, the World Meteorological Organization (WMO) endorsed the launch of the SDS Warning Advisory and Assessment System (SDS-WAS), and more recently of the first Regional Specialized Meteorological Center for Northern Africa, Middle East and Europe with activity specialization on Atmospheric Sand and Dust Forecast, the Barcelona Dust Forecast Center (BDFC). The SDS-WAS mission is to enhance the delivery of timely and quality SDS forecasts, observations, information and knowledge to users through an international partnership of research and operational communities.

Understanding, managing and mitigating SDS risks requires fundamental and cross-disciplinary knowledge underpinned by state-of-the-art scientific research, the availability of **reliable information on SDS past trends and current conditions**, the provision of skilful forecasts and projections tailored to a diversity of users, and the capacity to use the information effectively. At present, all these requirements are confronted by **major challenges**. These challenges include the **lack of reliable dust information in many countries affected by SDS** and the very **limited integration of dust information and forecasts into practice and policy**.

A major obstacle to reconstruct comprehensive dust information of the past is the scarcity of historical and routine in-situ dust observations, particularly in the countries most affected by SDS. Model simulations can be used to "fill in the blanks" and overcome the sparse coverage, low temporal resolution and partial information provided by measurements. By objectively combining model simulations with satellite observations, **eDUST will produce an advanced decadal high-resolution dust reanalysis for Northern Africa, Middle East and Europe** that will contribute to the formulation of management and/or mitigation plans for different socio-economic sectors. The proposed dust regional reanalysis will be built on three pillars: a state-of-art dust model and data assimilation system, quality observations and understanding of their respective uncertainties, and flow-dependent uncertainties reflected by the ensemble simulations. So far current reanalysis have been thought for the global domain (missing dust processes associated to finer spatiotemporal scales) and are based on the assimilation of total aerosol optical properties (lacking observational constraints on the model individual aerosol components) (e.g. NAAPS reanalysis; Lynch et al., 2016; MACC-II reanalysis, Cuevas et al., 2015; Inness et al., 2013; MERRAaero, Buchard et al., 2015; Buchard et al., 2016). The novelty of eDUST will be the generation of a dataset at an unprecedented high resolution spanning the satellite era of quantitative aerosol observations, the assimilation of satellite products over source regions with specific observational constraints for dust.

The recent development of algorithms retrieving dust properties from satellite reflectances in the blue range of the visible spectrum (Deep Blue algorithm) has led to a further extension of quantitative dust information to the source regions that are often characterised by bright surfaces and is potentially the most useful for dust applications. One of our researchers (Di Tomaso et al., 2016) has recently shown that the assimilation of MODIS Deep Blue retrievals from Collection 6 has a positive impact on the model analysis and forecasts downwind from the strongest dust sources of Sahara and the Arabian peninsula compared to the assimilation of Dark Target retrievals only. A combined MODIS Deep Blue and Dark Target-based dust analysis was produced at BSC by coupling the NMMB/BSC-Dust model with a Local Ensemble Transform Kalman Filter (LETKF) scheme (Figure 1.1). **The eDUST project will produce an unprecedented high-resolution regional dust reanalysis using the state-of-art NMMB/BSC-Dust model and its advanced data assimilation capabilities**. The NMMB/BSC-Dust model is operational at the WMO BDFC. It is the dust module of the NMMB/BSC-Chemical Transport Model (NMMB/BSC-CTM; Pérez et al., 2011; Haustein et al., 2012; Jorba et al., 2012; Spada et al., 2013; Badia and Jorba, 2016), which is an online multi-scale atmospheric model designed and developed at BSC in collaboration with NOAA/NCEP and the NASA Goddard Institute for Space Studies. The model includes an assimilation scheme based on LETKF (Hunt et al., 2007; Miyoshi and Yamane, 2007). LETKF is a development of the ensemble-based transform Kalman filter (ETKF, Bishop et al., 2001) and of the local ensemble Kalman filter (LEKF, Ott et al., 2004), and is particularly suited for the assimilation of aerosol information since it has been observed by Anderson et al. (2003), Shinozuka and Redemann (2011), and Schutgens et al. (2013) that aerosol fields have limited spatial correlations (~100km). Within MACC-II, the NMMB/BSC-Dust model was already used to produce a 30-year regional long-term dust simulation (without data assimilation), which was crucial to study the seasonal incidence of meningitis in Niger (Pérez García-Pando et al., 2014).

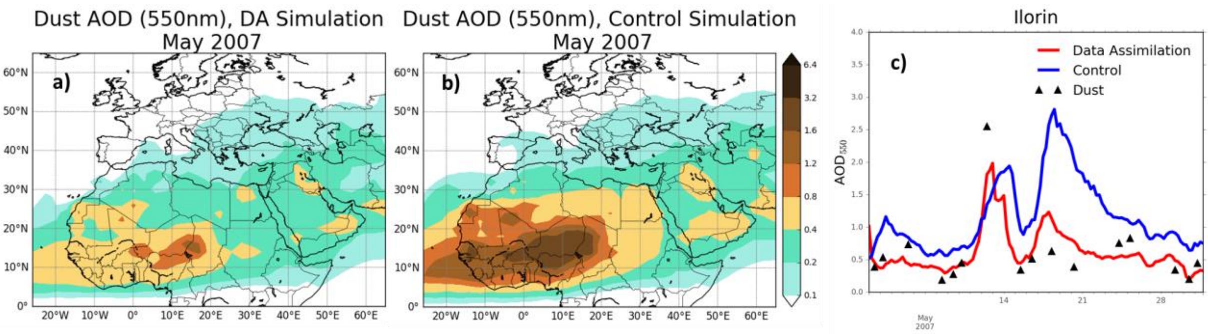

Figure 1.1 a) Monthly values of dust optical depth at 550 nm for May 2007 with the assimilation of MODIS Deep Blue and Dark Target. b) Same as in a), but without data assimilation. c) Comparison between the model simulations in a) and b) and AERONET observations in Ilorin (Niger). Simulations are based on the NMMB/BSC-Dust model. Extracted from Di Tomaso et al. (2016).

The high-resolution dust reanalysis will describe with accuracy the dust variability and trends, and provide extensive information for the socio-economic evaluation of major events, and their short (direct) and long-term (induced) impacts on society. These results will contribute to the activities of the WMO SDS-WAS project.

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

   (**Section limit: 1 page**)

The decadal **dust reanalysis using the NMMB/BSC-Dust model coupled with a LETKF assimilation scheme** (Pérez et al., 2011; Di Tomaso et al., 2016) will be run during the Earth Observation System satellite era from 2002 to 2016 when two polar-orbiting MODIS sensors can provide a good spatial and temporal coverage (globally two daily samples) and retrieval quality.

The baseline model configuration will follow the settings of the operational run of the Barcelona Dust Forecast Centre (BDFC), with a horizontal resolution of 0.10ºx0.10º, and a vertical resolution of 40 hybrid sigma-pressure layers. The model domain will cover Northern Africa, Middle East, Europe, and the North Atlantic up to about 50ºW (this will ensure that Atlantic arch episodes affecting Europe in spring are captured in our reanalysis). The meteorological fields will be downscaled from ERA-Interim reanalysis. The dust model has been recently updated with high-resolution source identification from MODIS Deep Blue, and improvements in the emission scheme and size distribution. A series of tuning experiments will be carried out with the model to iteratively adjust, on a regional basis, source and sink parameters (Milestone M1 in Table 2.1). Different assimilation parameters need to be adjusted to the model's resolution and updates. Some short experiments will be run to choose a close-to-optimal configuration for representativeness error (to be added to the instrument error component), observation density, covariance localisation and background error statistics. The method for generating the ensemble is also crucial since the ensemble represents the uncertainty in the model background. Calibration of model and observation error parameters will be guided by internal assimilation diagnostics: statistics on first-guess and analysis departures, comparison between the analysis root mean square error and ensemble spread, ensemble rank histograms. Once the system configuration is selected (Milestone M2 in Table 2.1), longer simulations will be run and validated. We will use a 24-hour assimilation window, where observations are a subset (limited to satellite overpasses) to 6-hour slices centred on the nominal valid time of the analysis (at 0, 6, 12, and 18 UTC). We will use the LETKF implementation with a four-dimensional extension as described in detail in Di Tomaso et al. (2016). These experiments will follow the schedule indicated in Table 2.1.

Table 2.1. Approximate schedule of the experiments to be performed.

Simulations	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Model configuration	M1											
Data assimilation configuration			M2									
Dust reanalysis												

The **Autosubmit software** (Manubens-Gil et al., 2016), briefly described below, will be used to manage the workflow and ensure a uniform and optimal use of the resources. Our ensemble data assimilation system offer an extensive computing flexibility as they are a set of daily simulations (24 members, 15 years) which can be run independently and therefore in parallel for an optimal use of the PRACE computing resources. They will be managed and packed in groups in a single big job if required, by Autosubmit to optimise the use of the machine and avoid collapsing the I/O system.

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

   (**Section limit: 1 page**)

NMMB/BSC-CTM is a parallel MPI application designed to run on both global and local scale. It subdivides the domain of operation into horizontal tiles and assigns them to computational units. Thus, the parallelization is addressed on a subdomain basis approach. Since the global latitude-longitude grid deforms toward the pole region, the nodes closer to the poles perform an additional filtering step using fast Fourier transform (FFT) to be applied to keep the integration stable using a time step of decent size. However, the polar filtering based on conventional fast Fourier transform requires transpositions involving extra communications and thus limits scaling.

NMMB/BSC-CTM is a coupled model constructed over the Earth System Framework model (ESMF) coupling framework; this implies that in between the execution of each module (dynamics, physics, chemistry, aerosol) the model performs a coupling step to exchange information. The numerical methods employed within the model are: the Adams-Bashforth Scheme for horizontal advection, the Crank-Nicholson scheme to compute vertical advection tendencies, the forward-backward scheme for horizontally propagating fast waves, and an implicit scheme for vertically propagating sound waves. Additionally, the chemistry module applies an Euler-Backward Iterative scheme to solve the ordinary differential equations of the stiff system of gas-phase chemistry. The I/O strategy of the system is designed for the setup of dedicated writing nodes. This results in a partition between computational and I/O nodes.

A single executable needs to be built for the NMMB/BSC-CTM model as well as an executable for the Local Ensemble Transform Kalman Filter (LETKF). The NMMB/BSC-CTM model fully supports a parallel environment. For the eDUST project, we will be running the most recent version of the NMMB/BSC-CTM model using the settings of the operational run of the Barcelona Dust Forecast Centre (BDFC), with a horizontal resolution of 0.10^{o}x0.10^{o}, and a vertical resolution of 40 hybrid sigma-pressure layers (~ 29,000,000 grid points).

For configuring and building the model executables, FORTRAN 77/90/95 complaint compiler with pre-processing capabilities and NetCDF-3 are needed. The simulations will require MPI libraries and runtime facilities (ESMF, wgrib, makedep90, bacio, sigio, sp), data handling tools, such as HDF4, HDF5, NETCDF3, NETCDF4, GRIB_API, R, CDO, NCO and general configurations tools, such as PERL, PYTHON, AUTOCONF and AUTOMAKE.

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

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

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

   (**Section limit: 3 pages**)

For the eDUST project, we will be running the most recent version of the NMMB/BSC-CTM model using the settings of the operational run of the Barcelona Dust Forecast Centre (BDFC), with a horizontal resolution of 0.10ºx0.10º, and a vertical resolution of 40 hybrid sigma-pressure layers. Running the NMMB/BSC-CTM model with such configuration requires a system with high-level resources, which, when combined with a large number of ensemble members (around 24 members) needed for the data assimilation system, requires computational resources available only on a small subset of machines worldwide. NMMB/BSC-CTM has already been deployed on MareNostrum III (a tier-0 system with 3056 nodes, 2x Intel SandyBridge-EP E5-2670/1600 8-core processors at 2.6 GHz, 8x4GB DDR3-1600 DIMMS (2GB/core) per node and Infiniband FDR10 network).

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

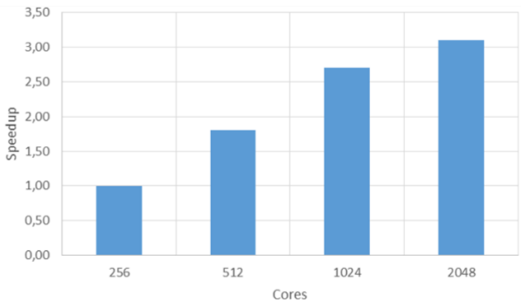

Figure 4.1 Speedup of the NMMB/BSC-CTM model for the global domain, 24km resolution. These estimates have been obtained by running NMMB/BSC-CTM on MareNostrum III.

As it is described in Section 2, the present HPC experiment is divided into different parts. A first part that includes the model configuration and the data assimilation configuration; and later once the system configuration is defined a longer dust reanalysis simulation. The configuration experiments represent only a small portion of the overall computing resources but they are needed to choose a close-to-optimal configuration for generating the ensemble used during the reanalysis.

From the scalability test performed on the NMMB/BSC-CTM model (Figure 4.1), we know that it is possible to increase the model throughput by increasing the number of cores up to more than 2k, even if doing so also raises the cost of the simulations. As it is shown in Figure 4.1, it is possible to use 512 processes (or even more) with reasonable efficiency. The dust forecasts used to generate the ensemble for the data assimilation can be performed independently, which means that 24 simulations (i.e. the number of member in the ensemble) should be executed in parallel to save time. The model can run acceptable efficiency using 1024 cores per member of the ensemble.

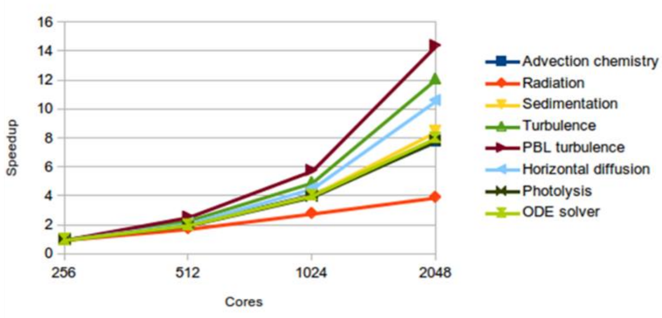

Figure 4.2 The most significant functions of the NMMB/BSC-CTM model for global domain, gl24km resolution. These estimates have been obtained by running NMMB/BSC-CTM on MareNostrum III.

In Figure 4.2 is presented the scalability of some significant functions for the chemistry configuration. Although most of the functions provide good scalability, some of them such as radiation, photolysis, and ODE solver have limited scalability in comparison to the rest ones. Investigation on the NMMB/BSC-CTM model using the best parallel model performance tools are still ongoing in two directions to reach an optimum scalability, in collaboration with the BSC Computer Sciences department. These actions include not only an adjustment of the model configuration and a balance of the number of cores dedicated to each of its components, but also modifications to the code itself and work on the parallel programming models.

Beyond the ongoing performance analysis, currently, there is a stable version of the NMMB/BSC-CTM model that it is used in the operational daily forecast of the Barcelona Dust Forecast Center and it is the model version that will be used in the present project. The results shown in Figure 4.1 and Figure 4.2 correspond to the model version that will be used in the present project.

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

Members of the BSC Earth Sciences Department (ES-BSC) have also contributed extensively to a large number of past and on-going EU projects (CLIM-RUN, QWeCi, DENFREE, PREFACE, EUCLEIA, IS-ENES2, SPECS, which is coordinated by BSC, and EUPORIAS). ES-BSC has been involved in the IS-ENES/PRACE-1IP working group focusing on the EC-Earth3 adaptation to Tier-0 machines. It has tested a range of model configurations, in the atmospheric resolutions T255/511/799 on several HPC systems: SGI Altix 3500, NEC-SX6, Linux cluster with Intel Xeon, Dell PowerEdge 2900, IBM pSeries 575 Power6 and IBM Power PC. ES-BSC leads the development of Autosubmit, a python-based wrapper software that can manage any type of climate simulation workflow. It can also bundle numerous jobs into a single big job for performing job-control on parallel sets of simulations throughout the execution. Finally, members of ES-BSC coordinated the recent HiResClim and HiResClim2 projects supported by PRACE and two RES projects (AECT-2015-1-0007 and AECT-2016-2-0018), which demonstrates its ability to lead this kind of projects.

ES-BSC has also expertise in analysing parallel programming model codes using cutting-edge tools. This allows continuously carrying out performance analyses to reach the optimum configurations for NMMB/BSC-CTM. Collaboration has been established with the NMMB developers for sharing the performance reports and code optimisation tools and techniques.

Initial work on the CURIE system (PRACE Access projects 2010PA0419 and 2010PA0627) with an early version of the model pointed out some issues. Although most of the functions provide good scalability, some of them such as radiation, photolysis, and ODE solver have limited scalability in comparison to the rest and will need a deeper analysis for their performance to be improved. As described in the previously, FFT transformations are applied to maintain the model stable. These routines are communication intense and when running high-resolution configurations with a large number of cells in the mesh, can have an impact on the efficiency of the model.

Figure 5.1 illustrates an example of such performance analyses. These are two different views as shown by the Paraver tool of the execution of one routine (exch4) of the NMMB/BSC-CTM model with one and two threads respectively. Paraver uses different types of plots to represent the values stored in a trace file. A trace file is a file holding information of an execution to perform a subsequent analysis. When we use more and more cores for the execution of an application, the communication can be more complicated. Moreover, many scientists agree that when we double the number of the cores, a new bottleneck is observed. That’s why we tried to decrease the communication cost for a stencil communication routine called exch4. In the example, the speedup is 1.76 which means that when we doubled the number of the cores the communications routine was improved by 1.76 times.

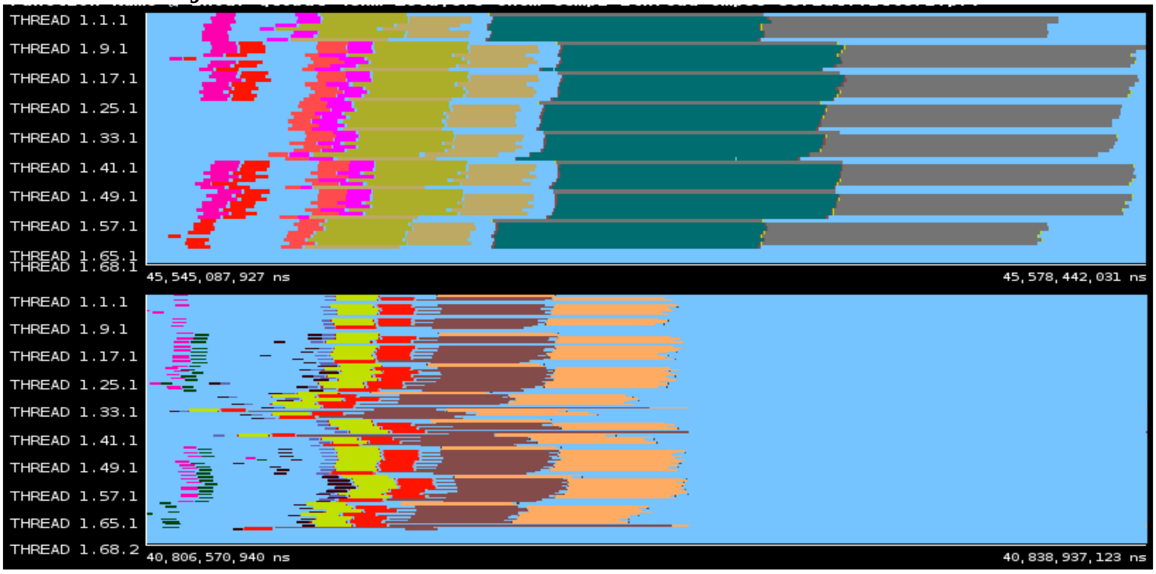

Figure 5.1. Executing exch4 routine of the NMMB/BSC-CTM model with one thread (top) and two threads (bottom).

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

   (**Section limit: 1 page**)

Table 6.1 lists the experiments described in Section 2, which will run, by chunks of 1 day-length on 3-hourly basis, i.e. 9 hours/chunk of simulations for all the experiments. The benchmarking exercise performed through internal queues suggests that, taking into account the average load of the MareNostrum III queues, the optimum performance for throughput is obtained using 1024 procs, which requires a wall-clock time of 0,15 hours for a one-day simulation. The model configuration and data assimilation configuration experiments include a set of sensitivity runs to define the final configuration used in the dust reanalysis as it is described in Section 2. The computational time estimated for one-year execution of the high-resolution dust reanalysis which is perform using a 24 member ensemble data assimilation system following the BDFC model configuration (at 0.1ºx0.1º and 40 vertical layers) is approximately of 1.2 million CPU hours.

Run type	# Runs	# Steps/Run	Walltime/Step	# CPU cores	Total core hours/Type Run
Model configuration	500	9	0,15	1024	691.200,00
Data assimilation configuration	1,000	9	0,15	1024	1.382.400,00
Dust reanalysis	131,765	9	0,15	1024	18.215.193,60

forecast					
TOTAL					**20.288.793,60**

Table 6.1: Cost of the experiments proposed.

The final estimate is for a total request of 21 million core-hours, which includes the numbers described in Table 6.1 plus a small buffer of 3% to account for failing jobs that will need to be repeated. The experiments will be run using Autosubmit, the launching and monitoring solution developed by the group of the applicant that allows the remote submission of NMMB/BSC-CTM experiments. Autosubmit includes in the workflow of the experiments a job that retrieves the data back to the Earth Sciences Department at BSC data storage as soon as each chunk of simulation has completed.

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

The main assimilation cycle for the production of the reanalysis will be executed on the HPC facilities. In the data assimilation mode multiple (independent) NMMB/BSCCTM simulations are run to account for model uncertainty in the calculation of the data assimilation corrections. The needed computational power is directly dependent on the unpreceded high-resolution of the ensemble simulations.

eDUST proposes a simulation over more than one decade (a 15-years reanalysis, 2002-2016) with the necessary set-up experiments. The proposed dust reanalysis consists of daily runs for 15-years reanalysis period.

The type of simulations conducted during the eDUST project requires hosting a set of gridded satellite dust observations in order to produce the analysis (i.e. 23 GB per year of simulation) as well as the global meteorological input data files used as initial meteorological conditions and boundary conditions at intervals of 6 h (i.e. 90 GB per year of simulation considering ERA-Interim reanalysis). Around 500 GB of *home* space will be required to host the code. Hence the total disk space required hosting the input data (dust observations and meteorological initial and boundary conditions for the complete simulation) and the code amounts to 2.5TB in the *home* file system.

The required *scratch* space is motivated by the requirement to perform many independent simulations at the same time (i.e. 24 independent-members in the ensemble data assimilation system) each producing up to 50GB of raw data per day of simulation. The total disk space required to hold these data in the workspace can be organised with a disk space of 120 TB in the *scratch* file system. These data will be held in the workspace before being post-processed to reduce the data volume to less than a fifth of their original size (from 1TB per day of simulation in the execution to 12GB per day of simulation), which will then be transferred to the *archive* file system and finally locally. Hence the total disk space required hosting the resulting dust reanalysis for 15-years period and the corresponding set-up experiments will amount 90Tb.

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

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

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

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

Olivier Boucher (olivier.boucher@lmd.jussieu.fr), CNRS research director at the Laboratoire de Météorologie Dynamique (LMD), Paris, France.
Slodoban Nickovic (nickovic@gmail.com), Senior Scientist, Republic Hydrometeorological Service of Serbia, Belgrade, Serbia.
Angela Benedetti (angela.benedetti@ecmwf.int), Senior Scientist, European Centre for Medium-Range Weather Forecasts, ECMWF, Reading, United Kingdom.

10. ***References***

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., & Holmén, K. (2003). Mesoscale Variations of Tropospheric Aerosols. Journal of the Atmospheric Sciences, 60(1), 119-136.

Badia, A., & Jorba, O. (2016). Gas-phase evaluation of the online NMMB/BSC-CTM model over Europe for 2010 in the framework of the AQMEII-Phase2 project. Atmospheric Environment, 115, 657-669.

Bauer, S. E., & Koch, D. (2005). Impact of heterogeneous sulfate formation at mineral dust surfaces on aerosol loads and radiative forcing in the Goddard Institute for Space Studies general circulation model. Journal of Geophysical Research: Atmospheres, 110(D17).

Bishop, C. H., Etherton, B. J., & Majumdar, S. J. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. Monthly weather review, 129(3), 420-436.

Buchard, V., da Silva, A. M., Colarco, P. R., Darmenov, A., Randles, C. A., Govindaraju, R., ... & Spurr, R. (2015). Using the OMI aerosol index and absorption aerosol optical depth to evaluate the NASA MERRA Aerosol Reanalysis. Atmos. Chem. Phys, 15(10), 5743-5760.

Buchard, V., da Silva, A. M., Randles, C. A., Colarco, P., Ferrare, R., Hair, J., Hostetler, C., Tackett, J., & D. Winker (2016). Evaluation of the surface PM2.5 in Version 1 of the NASA MERRA Aerosol Reanalysis over the United States, Atmospheric Environment, 125, 100-111, doi:10.1016/j.atmosenv.2015.11.004.

Cuevas, E., Basart, S., Baldasano Recio, J. M., & Berjon, A. (2015). The MACC-II 2007-2008 reanalysis: atmospheric dust evaluation and characterization over northern Africa and the Middle East. Atmospheric chemistry and physics, 15(8), 3991-4024.

Di Tomaso, E., Schutgens, N. A. J., Jorba, O. & Pérez García-Pando, C. (2016). Assimilation of MODIS Dark Target and Deep Blue observations in the dust aerosol component of NMMB/BSC-CTM version 1.0, Geosci. Model Dev. Discuss., doi:10.5194/gmd-2016-206.

Haustein, K., Pérez, C., Baldasano, J. M., Jorba, O., Basart, S., Miller, R. L., ... & Washington, R. (2012). Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model Part 2: Experimental campaigns in Northern Africa. Atmospheric Chemistry and Physics, 12(L03812), 2933-2958.

Hunt, B. R., Kostelich, E. J., & Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. Physica D: Nonlinear Phenomena, 230(1), 112-126.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., ... & Flemming, J. (2013). The MACC reanalysis: an 8 yr data set of atmospheric composition. Atmospheric chemistry and physics, 13(8), 4073-4109.

Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., ... & Kawahata, H. (2005). Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), 67-71.

Jorba, O., Dabdub, D., Blaszczak-Boxe, C., Pérez, C., Janjic, Z., Baldasano, J. M., ... & Gonçalves, M. (2012). Potential significance of photoexcited NO2 on global air quality with the NMMB/BSC chemical transport model. Journal of Geophysical Research: Atmospheres, 117(D13).

Lynch, P., Reid, J. S., Westphal, D. L., Zhang, J., Hogan, T. F., Hyer, E. J., Curtis, C. A., Hegg, D. A., Shi, Y., Campbell, J. R., Rubin, J. I., Sessions, W. R., Turk, F. J., and Walker, A. L. (2016) An 11-year global gridded aerosol optical thickness reanalysis (v1.0) for atmospheric and climate sciences, Geosci. Model Dev., 9, 1489-1522, doi:10.5194/gmd-9-1489-2016.

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

Miller, R. L., Knippertz, P., García-Pando, C. P., Perlwitz, J. P., & Tegen, I. (2014). Impact of dust radiative forcing upon climate. In Mineral Dust (pp. 327-357). Springer Netherlands.

Miyoshi, T., & Yamane, S. (2007). Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. Monthly Weather Review, 135(11), 3841-3861.

Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., Kostelich, E. J., Corazza, M., ... & Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. Tellus A, 56(5), 415-428.

Pérez García-Pando, C., Stanton, M. C., Diggle, P. J., Trzaska, S., Miller, R. L., Perlwitz, J. P., ... & Thomson, M. C. (2014). Soil Dust Aerosols and Wind as Predictors of Seasonal Meningitis Incidence in Niger. Environmental Health Perspectives, 122(7), 679.

Pérez, C., Haustein, K., Janjic, Z., Jorba, O., Huneeus, N., Baldasano, J. M., ... & Perlwitz, J. P. (2011). Atmospheric dust modeling from meso to global scales with the online NMMB/BSC-Dust model–Part 1: Model description, annual simulations and evaluation. Atmospheric Chemistry and Physics, 11(24), 13001-13027.

Pérez, C., Nickovic, S., Pejanovic, G., Baldasano, J. M., & Özsoy, E. (2006). Interactive dust-radiation modeling: A step to improve weather forecasts. Journal of Geophysical Research: Atmospheres, 111(D16).

Rodríguez, S., Querol, X., Alastuey, A., Kallos, G., & Kakaliagou, O. (2001). Saharan dust contributions to PM10 and TSP levels in Southern and Eastern Spain. Atmospheric Environment, 35(14), 2433-2447.

Schroedter-Homscheidt, M., Oumbe, A., Benedetti, A., & Morcrette, J. J. (2012). Aerosols for concentrating solar electricity production forecasts: requirement quantification and ECMWF/MACC aerosol forecast assessment. Bulletin of the American Meteorological Society, 94(6), 903-914.

Schutgens, N. A. J., Nakata, M., & Nakajima, T. (2013). Validation and empirical correction of MODIS AOT and AE over ocean. Atmospheric Measurement Techniques, 6(9), 2455-2475.

Shinozuka, Y., & Redemann, J. (2011). Horizontal variability of aerosol optical depth observed during the ARCTAS airborne experiment. Atmospheric Chemistry and Physics, 11(16), 8489-8495.

Spada, M., Jorba, O., Pérez García-Pando, C., Janjic, Z., & Baldasano, J. M. (2013). Modeling and evaluation of the global sea-salt aerosol distribution: sensitivity to size-resolved and sea-surface temperature dependent emission schemes. Atmospheric Chemistry and Physics, 13(23), 11735-11755.

Stefanski, R., & Sivakumar, M. V. K. (2009). Impacts of sand and dust storms on agriculture and potential agricultural applications of a SDSWS. In IOP Conference Series: Earth and Environmental Science (Vol. 7, No. 1, p. 012016). IOP Publishing.