Barcelona Supercomputing Center
Centro Nacional de Supercomputación

RED ESPAÑOLA DE SUPERCOMPUTACIÓN

AECT-2017-1-0006 Applying an ensemble data assimilation technique to generate a high-resolution regional dust analysis

1. General Information

Activity Id
AECT-2017-1-0006

a) Activity Title
Applying an ensemble data assimilation technique to generate a high-resolution regional dust analysis

b) Area
Astronomy, Space and Earth Sciences

2. Research Project Description

a) Is this a Test Activity?
No

b) Is this a Long Term Activity that will extend over several application periods?
No

c) Brief description of the Project
Over the past decade, there has been a growing recognition of the crucial role of sand and dust storms (SDS) on weather, climate and ecosystems, along with their important adverse impacts upon life, health, property, economy. 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 (NAMEE) with activity specialisation on Atmospheric Sand and Dust Forecast: the Barcelona Dust Forecast Center (BDFC). The BDFC mission is to enhance the delivery of timely and quality SDS forecasts to end users.

Dust numerical prediction has become an important activity at many research and operational weather centres in the past decade due to growing interest from a diverse set of stakeholders, such as air quality regulatory bodies, aviation and military authorities, solar energy plant managers, and health professionals. Data assimilation offers a mathematical framework to incorporate observational information into models for the production of the best estimate of dust concentrations (a dust analysis) with the aim to improve dust monitoring and predictions. Nowadays, most operational aerosol centres with aerosol forecasting capabilities run systems which include aerosol assimilation.

While the general assimilation tools can be ported with some dedicated effort to any atmospheric variable, there are some specific challenges in dust assimilation which are mainly related to the paucity of the suitable observations available for assimilation and the complexity of extracting specific dust signals from satellite radiances which are affected by all aerosol species and other atmospheric quantities.

By objectively combining model simulations with satellite observations, the present project aims to prepare the operational implementation of an ensemble data assimilation technique to generate an improved high-resolution regional dust forecast for the NAMEE domain. The proposed improved daily dust forecast 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 operational aerosol forecasts (with data assimilation) have been thought mainly 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).

The recent development of satellite retrieval algorithms using blue wavelengths (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. Di Tomaso et al. (2016) have recently shown that theassimilation of Deep Blue retrievals from the MODIS satellite sensor has a positive impact on the model analysis and forecasts.

The novelty of the present project will be the generation of a dust forecast at an unprecedented high-resolution using the state-of-art NMMB/BSC-Dust model and its advanced data assimilation capabilities (Pérez et al., 2011; Di Tomaso et al., 2016) with the assimilation of satellite products over source regions with specific observational constraints for dust. This system will be the basis for a future production of an innovative dust reanalysis (a consistent 3-dimensional dust concentration dataset over a long period) for the NAMEE domain.

The present activity forms part of different international and national initiatives on the generation of improved dust products based on data assimilation techniques:
- ACTRIS-2 (grant agreement number 654109): a European project aimed at integrating ground-based stations (for analysis but also for model operational purposes) equipped with advanced atmospheric probing instrumentation for aerosols, clouds and short-lived gas species.
- ESA Aerosol CCI: a project aimed at producing a set of global aerosol products from a set of European satellite instruments.
- WMO Barcelona Dust Forecast Center (http://dust.aemet.es): an initiative under the umbrella of WMO managed by AEMET and BSC and withthe main objective to enhance the delivery of timely and quality SDS forecasts to users.

d) Grant References

ACTRIS-2 (http://www.actris.eu/)
ESA CCI (http://cci.esa.int)
WMO BDFC (http://dust.aemet.es)

e) Brief description of the Project (if this Activity takes place in the context of a Technology or Industrial Project)

Not applicable.

f) Specific Activity proposed

The present project aims to prepare the operational implementation of an ensemble data assimilation

technique to generate an improved high-resolution regional dust forecast for the WMO Barcelona Dust Forecast Center using the NMMB/BSC-Dust model. This aim will be achieved finding a solution to challenges related to the representation of model uncertainty for the specific numerical model used, and to the treatment of the observations, and observation error correlation statistics, for the specific spatial and temporal resolution used.

The Earth Sciences Department of BSC (ES-BSC; http://www.bsc.es/ESS) operates daily regional dust forecasts based on the NMMB/BSC-Dust model and conducts intensive modelling research for short-term operational prediction. The NMMB/BSC-Dust model is participating in the WMO SDS-WAS model intercomparison and is the model operational at the WMO Barcelona Dust Forecast Center. The NMMB/BSC-Dust has been developed at the ES-BSC in collaboration with NOAA/National Centers for Environmental Prediction (NCEP), NASA Goddard Institute for Space Studies and the International Research Institute for Climate and Society (IRI). The NMMB/BSC-Dust model provides operational dust forecast (without data assimilation) over North Africa-Middle East-Europe and global regions. The most relevant characteristic of the NMMB/BSC-Dust model is its on-line coupling with the Non-hydrostatic Multiscale Model (NMMB). This provides a unique framework to simulate/predict weather and air quality in a wide range of scales from global to mesoscale applications (from 100 to 1 km), and allows interactions among meteorology-dust-chemistry processes. Recently, the model has been coupled to a data assimilation scheme based on a Local Ensemble Transform Kalman Filter (LETKF; Di Tomaso et al., 2016) which has been implemented and tested so far only for the global domain, an old model version, and at a coarse spatial resolution.

To achieve the project's objective, a series of tuning experiments will be carried out with a recently improved model physical parameterization to iteratively adjust, on a regional basis source and sink parameters for a single execution of the model. Moreover, different assimilation parameters need to be calculated to the model's resolution and updates. Data assimilation 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, longer simulations (one month period) will be run and compared with respect to the operational products published in the BDFC website (http://dust.aemet.es/). We will use a 24-hour assimilation window, where observations are aggregated (limited to satellite overpasses) into 6-hour slices centred on the nominal valid time of the analysis.

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. The meteorological fields will be downscaled from NCEP/GFS. 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. The data assimilation scheme used will be the LETKF implementation with a four-dimensional extension as described in detail in Di Tomaso et al. (2016).

The configuration mentioned above requires extensive computational resources to produce a large enough ensemble of model members. Otherwise, from the previous scalability test performed on the

NMMB/BSC-Dust model, 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. It is possible to use 256 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. Therefore, the main assimilation cycle for the production of the dust analysis and forecast has to be executed on HPC facilities.

The benchmarking exercise performed taking into account the average load of the MareNostrum III queues showed that the optimum performance for throughput is obtained using 256 procs, which requires a wall-clock time of 0.23 hours for a one-day simulation following the BDFC model configuration (at 0.1ºx0.1º and 40 vertical layers). In the data assimilation mode, multiple NMMB/BSC-Dust 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 unprecedented high-resolution of the ensemble simulations. 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 the model's experiments. The final estimate is for a total request of 144,128 core-hours, which includes the model configuration (100 experiments of one-single model daily run) and the data assimilation configuration experiments indispensable to build the operational system (100 experiments of 24-member daily run).

The type of simulations conducted during the present project requires hosting a set of gridded satellite dust observations in order to produce the analysis (i.e. 2 GB per month 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. 30 GB per month of simulation considering NCEP/GFS forecast). 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 input data) and the code amounts to 2.5TB in the home file system. Results are written periodically every three hours of simulation. As a result, each experiment (daily run) will include nine binary files corresponding to a day of simulation (24 hours) on a 3-hourly basis. 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 resulting raw binary files are post-processed to NetCDF files. The default output variables have an estimated size of each binary file up to 4.5GB.

g) Computational algorithms and codes outline

NMMB/BSC-Dust is the dust module of the NMMB/BSC-Chemistry Transport Model (Pérez et al., 2011; Jorba et al., 2012; Spada et al., 2013; Badia and Jorba, 2016), which 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 a 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 model as well as an executable for the Local Ensemble Transform Kalman Filter (LETKF). The model fully supports a parallel environment. For the present project, we will be running the most recent version of the NMMB/BSC-Dust model using the settings of the operational run of the WMO Barcelona Dust Forecast Centre (BDFC), with a horizontal resolution of 0.10ºx0.10º, and a vertical resolution of 40 hybrid sigma-pressure layers (~ 29,000,000 grid points).

hybrid sigma-pressure layers (~29,000,000 grid points).

The research group has a long-term experience in using HPC platforms, with most of their members working in HPC environments, both developing and using numerical models. Regarding the numerical model, the group has performed benchmarks in different HPC platforms, namely Marenostrum, CURIE, and Minotauro. At the proposed resolutions and configuration the models scale efficiently with 256 nodes.

3. Software and Numerical Libraries

Software components that the project team requires for the activity.

a) Applications + Libraries

HDF5, NETCDF, R, OPENMPI, UDUNITS, NCO, INTELMPI

b) Compilers and Development Tools

INTEL

c) Utilities + Parallel Debuggers and Performance Analysis Tools

PYTHON, IMAGEMAGICK, NCVIEW, AUTOCONF

d) Other requested software

CDO

e) Proprietary software

Not applicable.

4. Research Team Description

a) Personal Data

Name of Team Leader	Sara Basart
Institution	Barcelona Supercomputing Center - Centro Nacional de Supercomputación
e-mail	sara.basart@bsc.es
Phone	+34 934134038
Nationality	Spain

b) The employment contract of the activity leader with the research organisation is valid at least 3 months after the end of the allocation period.

Yes

c) Curriculum Vitae of the Team Leader

Dr Sara Basart born in Barcelona (Spain, 13th May 1978). She is Bachelor in Physics (2005) and Master in Science in Meteorology (2008) from Barcelona University (Barcelona, Spain). Dr Basart obtained her PhD degree in Engineering Environmental (Degree of European Doctor) at Technical University of Catalonia (UPC) in January 2012 while doing her research at different research centers in Spain (like the Izana Atmospheric Research Center of the Meteorological State Agency of Spain, CN-AEMET) and

the Izaña Atmospheric Research Center of the Meteorological State Agency of Spain, CAIAEMET) and France (Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL and Laboratoire de Météorologie Dynamique, LMD).

At present, Dr Basart is postdoctoral researcher in the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS, http://www.bsc.es/earth-sciences/mineral-dust-forecast-system). Her main research background covers mineral dust modelling, air quality and aerosols. She is scientist in charge of the WMO Sand and Dust Storm Warning Advisory and Assessment System (SDS-WAS; http://sds-was.aemet.es/) Regional Center for Northern Africa, Middle East and Europe, and first Regional Specialized Meteorological Center with activity specialization on Atmospheric Sand and Dust Forecast, the Barcelona Dust Forecast Center (BDFC; http://dust.aemet.es/), hosted in BSC-CNS. She also participates in international projects as the International Cooperative on Aerosol Prediction (ICAP) initiative and ACTRIS and ACTRIS-2. She is also leading the BSC-CNS participation in Copernicus (CAMS-84). She has authored or co-authored more than 30 peer-reviewed publications in international journals and book chapters and over 60 communications to international and national conferences.

ResearchGate: https://www.researchgate.net/profile/Sara_Basart
ORCID record at http://orcid.org/0000-0002-9821-8504

d) Names of other researchers involved in this activity

Dr Enza Di Tomaso (BSC, enza.ditomaso@bsc.es)
Dr Carlos Pérez García-Pando (BSC, carlos.perez@bsc.es)
Mr Kim Serradell (BSC, kim.serradell@bsc.es)

e) Relevant publications

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

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.

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.

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.

5. Resources

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

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

Requested machine	Finis Terrae II ((BULL with 7712 cores Intel Haswell 2680v3, Infiniband Fat-Tree FDR, 44TB memory, 1 PB disk)
Interprocess communication	Null

Typical Job Run				
Number of processors needed for each job		256.00		
Estimated number of jobs to submit		100.00		
Average job durations (hours) per job		0.23		
Total memory used by the job (GBytes)		78.00		
Largest Job Run				
Number of processors needed for each job		256.00		
Estimated number of jobs to submit		100.00		
Average job durations (hours) per job		5.40		
Total memory used by the job (GBytes)		78.00		
Total disk space (Gigabytes)	**Minimum**	600.00	**Desirable**	2000.00
Total scratch space (Gigabytes)	**Minimum**	3000.00	**Desirable**	7000.00
Total tape space (Gigabytes)	**Minimum**	1500.00	**Desirable**	3000.00
Total Requested time (Thousands of hours)		144.13		

If this activity is asking for more than 2Million CPU hours, you need to justify the amount of resources requested for the activity (max 1000 characters)

INFORMATION: The estimated cost of the requested hours, considering only the electricity cost, is 1647.4059 euros.

The architectures selected for the requested resources are only a suggestion. If no hours in this machine/these machines are available, please grant resources in any other similar architecture where the codes used for the application may run efficiently.
** this option implies that if no hours in this machine/these machines are available, the acces committee will reject the full application

6. Abstract for publication

There is an increasing need for accurate predictions of sand and dust storms because of its impact on life, health, property, environment and economy in many countries. In alignment with the mission of the first Regional Specialized Meteorological Center on Atmospheric Sand and Dust Forecast, the present project aims to prepare the operational implementation of an ensemble data assimilation technique to generate an improved high-resolution regional dust forecast for Northern Africa, Middle East and Europe. The novelty of the proposed project will be the generation of a dust forecast at an unprecedented high resolution using a state-of-art dust model and its advanced data assimilation

unprecedented high-resolution using a state-of-art dust model and its advanced data assimilation capabilities with the assimilation of satellite products over source regions with specific observational constraints for dust.

7. Contact with CURES during last year

Information about the RES Users Committee (CURES).

a) User has contacted the CURES during last year

No

b) If not, indicate why you have not contacted the CURES

Because this is my first application to RES.

Barcelona Supercomputing Center, 2016