REQUEST FOR A SPECIAL PROJECT 2018–2020

MEMBER STATE: ITALY

Principal Investigator[1]: Ines Cerenzia, Tiziana Paccagnella

Affiliation: ARPAE SIMC

Address: Viale Silvani, 6 Bologna, Italy

Other researchers: Davide Cesari, Andrea Montani,Chiara Marsigli

Project Title: SPHERA (Special Project: High rEsolution ReAnalysis over Italy)

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** _____	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2018	
Would you accept support for 1 year only, if necessary?	YES ☒	NO ☐

Computer resources required for 2018-2020: (To make changes to an existing project please submit an amended version of the original form.)		**2018**	**2019**	**2020**
High Performance Computing Facility	(SBU)	50.240.000	4.340.000	38.550.000
Accumulated data storage (total archive volume)[2]	(GB)	55.000	60.000	132.200

An electronic copy of this form must be sent via e-mail to: *special_projects@ecmwf.int*

Electronic copy of the form sent on (please specify date): 30-06-2017

Continue overleaf

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

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

Principal Investigator: Ines Cerenzia, Tiziana Paccagnella

Project Title: SPHERA (Special Project: High rEsolution ReAnalysis over Italy)

Extended abstract

The developing interest of EU Community towards high resolution regional reanalysis is well expressed by the number of ongoing projects in this field: the recent call of Copernicus C3S_322 Regional Climate Reanalysis, the Framework Programme 7 projects UERRA (Uncertainties in Ensembles of Regional Reanalysis) and EURO4M (European Reanalysis and Observation for Monitoring).

In this framework, ARPAE-SIMC proposes the development of a high resolution atmospheric regional reanalysis over Italy, SPHERA, performed with the COSMO non-hydrostatic Limited Area Model. COSMO is developed in the framework of the COSMO (COnsortium for Small scale MOdelling, Schättler et al., 2011) consortium cooperation. It is used in the operational NWP suites in Italy, as well as in several other ECMWF Member States (Switzerland, Germany, Greece) and Co-operating States (Romania, Israel).

SPHERA will be performed by means of a dynamical downscaling of the COSMO-REA6 reanalysis (Bollmeyer et al., 2015) and by employing observational nudging during the model integration. SPHERA will cover 25 years and will produce three-dimensional hourly model output. COSMO-REA6 is a 6 km-resolution reanalysis covering the CORDEX European domain, driven by ERA-interim, whose development is currently on going thanks to a cooperation between the DWD, Deutscher Wetterdienst, the University of Bonn and Cologne within the Hans Ertel Center for Weather Research. Our effort will complement, even with some implementation differences, what is under development in Germany in the convection permitting reanalysis project COSMO-REA2 (Wahl et al., 2017).

The SPHERA reanalysis aims at:

- having a high resolution, space and time consistent, description of the past decades climate characteristics in term of some Essential Climate Variables (ECV, Bojinsky et al., 2004) over Italy and its surrounding Seas;
- evaluating the trends of these ECVs during the last decades also in connection to high impact severe weather event occurrences;
- providing accurate and long-term estimate of the atmosphere state on specific sites for specific purposes (e.g. for renewable energy or tracer dispersion applications);
- having a COSMO reference climate for COSMO applications in different scenarios associated to Climate Change over Italy;
- providing the atmospheric and surface boundary conditions to high-resolution model applications, e.g sea state and ocean modelling, marine sediment transport and coastal-erosion modelling, air pollution modelling, etc..
- having a COSMO model validation based on long term performance, to be used as a reference for the operational forecast in Italy and in the other Countries using COSMO operationally (previously mentioned). This point is particularly relevant since the COSMO model is currently undergoing deep revisions (mainly regarding the physical packages of turbulence and microphysics);
- calibrating the COSMO based forecasting systems (e.g. COSMO-LEPS, Montani et al., 2011).

In order to match the mentioned purposes, SPHERA archive will include the upper air variables: pressure, temperature, wind speed and direction, specific humidity and specific content of 6 micro-physical species (i.e. cloud water and ice, rain, snow, graupel and hail), cloud properties, turbulent kinetic energy, and the surface variables: air temperature, wind speed and direction, water vapour, pressure, precipitation, surface radiation budget, surface fluxes and soil temperature and soil moisture content of water and frozen water. The integration domain is reported in Figure 1.

As mentioned before, the assimilation of the conventional observations (PILOT, TEMP, AIREP AMDAR, ACARS, SYNOP, SHIP, DRIBU) will be performed by using the continuous nudging technique, which is the consolidated assimilation scheme of COSMO. The reanalysis time span is planned to range between 1995 and 2019; extensions to previous years will be evaluated during the project development, depending on the availability of the forcing model COSMO-REA6.

Figure 1. Integration domain of SPHERA reanalysis archive

SCIENTIFIC PLAN

1. Set up of the suite and preliminary tests

The first phase of the SPHERA project will aim at identifying a suitable model configuration for the reanalysis production. ARPAE-SIMC boasts a solid experience with the limited area model COSMO, given that it is a developing member of the COSMO Consortium and and given that the model has been run operationally at the Service since 2000. Based on this experience, the more critical issues related to a reanalysis archive over Italy will likely regard:

- the definition of the optimal soil surface boundary condition (the sea surface temperature, deep soil temperature and humidity)

- the implementation of a verification tool box
- the definition of an optimal set up of the model configuration
- the definition of the dataset to be assimilated

Specific experiments over a limited period of study (Apr-May-Jun 2007) will tackle these critical points. In particular the first point will need a special attention as small inaccuracies at the soil level can trigger systematic errors associated to the soil hydrological cycle. Due to the large soil inertia, tests to identify a potential deviation of the soil features will be also performed considering 24 months of reanalysis. A second phase will aim at developing an operational automatic system for the reanalysis production, which will run on the ECMWF supercomputer (HPC). The initial and boundary conditions will come from COSMO-REA6 archive, whose fields already site on HPC. Observational data will be extracted from MARS database.

The pre-processing and the model computation will be divided in 24h-long runs in order to facilitate the management. A 24h run will be performed in about 30min. After post-processing and compression, the final output will be stored on the ECFS storage space of ECMWF. The system of routines will be managed by ECFLOW package.

The completion of these preliminary activities is estimated by March 2018 if a late Special Project 2017 (i.e. SPHERA-PRE, Special Project: High rEsolution ReAnalysis over Italy – PREliminary study) has been approved.

2. Production

The time span of SPHERA will depend on the availability of the forcing dataset COSMO-REA6, whose production is still ongoing. At the moment of submitting this application, COSMO-REA6 covers the period 1995-2014. The forward years 2015 and 2016 will be already available in 2018 (Keller and Wahl, personal communication). Moreover, COSMO-REA6 will be progressively updated to near present, thus in 2020 the dataset will most certainly include also the interval 2017-2019 (Keller and Wahl, personal communication). Extension of COSMO-REA6 further into the past (back from 1995) are currently under discussion. The possibility to coherently extend SPHERA will be evaluated during this project development.

An estimate of the computational requirements, of the memory storage and of the computing time has been obtained by performing a 24h-long run of SPHERA on the HPC supercomputer. The results are reported in Table1, together with the estimated requirements for a 1year-long reanalysis.

	24hours	1year
High Performance Computing Facility (SBU)	4.400	1.600.000
Accumulated data storage (Gb)	5,5	2.007,5
Computing time required to simulate (hours)	0,5	182,5

Table 1. Computational, memory storage and computing time requirements for 24hours of SPHERA reanalysis and the estimates for 1 year.

2. a. High Performance Computing Facility requirement

The 24h-long reference example reported in Table 1 accounts for the extraction of observations from MARS archive, the extraction of the initial and boundary condition (COSMO-REA6) from HPC, the COSMO preprocessing, the COSMO simulation in continuous nudging and in single precision, the compression of the COSMO output and the storage on ECFS. To minimize the computational requirement in term of SBU, COSMO will be run in single precision rather than in double precision (but few routines), with a SBU saving of about 40%.

2. b. Data storage management

The final output of COSMO plus some post-processed output will consist of 13 full-3D variables, approximately 50 variables in 2D and 3 multi-level soil variables, covering a horizontal grid of 576x701 cells and 65 vertical levels (domain in Figure 1). Approximately 18Gb would be required every 24h to store this hourly data. In order to reduce this amount, a first JPEG2000 compression will be applied at each hourly grib output file, with a space saving of 70%. A second compression (less efficient) will be applied gathering the hourly output in a daily folder. In this way the space consumption per simulated day will be about 5.5Gb (Table 1).

3. Verification

SPHERA will be validated against several sources:

- long-term observational analysis based on high-resolution local networks (e.g. ERACLITO, Antolini et al. 2015; ARCIS, Pavan et al. 2013). These climatic datasets include the surface variables (total precipitation, temperature) from 1961 to near present, respectively over Emilia-Romagna region and over Northern-Italy;
- observational dataset based on high density local networks, including total precipitation, temperature, humidity and wind speed and direction. This dataset covers the Northern and Central Italy and it extends for more than 10 years in the past;
- European-based gridded observational dataset (e.g. E-OBS, Haylock et al. 2008).

4. Time table

SPHERA activities will be organized in the following manner:

	High Performance Computing Facility (SBU)	Accumulated data storage (Gb)
2018		
Jan-Mar: Set up of the suite, implementation of a verification tool box and preliminary tests (estimated tests correspond to a total of 5 years of reanalysis)	8.000.000	10.037,5
Apr-Dec: Reanalysis production for the period 1995-2016 (22 years of reanalysis)	42.240.000 (including 20% of margin for potential re-runs*)	44.165,0
2019		
Reanalysis production: extension to near future (2years)	3.840.000 (including 20% of margin for potential re-runs*)	4.015,0
Evaluation of the already generated reanalysis by means of the verification tool box	500.000	100,0
2020		
Evaluation of the full reanalysis archive by means of the verification tool box	500.000	100,0
Reanalysis production: extension to near future (1 years)	1.920.000 (including 20% of margin for possible re-runs*)	2.007,5

Table2. Time table of the SPHERA project. (*) a margin of 20% has been added to the computational requirement in order to account for potential re-run associated to errors in the reanalysis production.

5. Conclusion

The necessity to evaluate long term climatology is becoming more and more important for climatological evaluation itself but also for several other scientific, economical and social applications. Due to the sparseness and heterogeneity of observational data sources, the added value coming from numerical weather prediction models is crucial. The SPHERA reanalysis is meant to respond to this necessity in the most accurate and coherent way (e.g. having high spatial and temporal resolution, applying continuous nudging).

Moreover, improvements to this reference technique will be considered along the project development. A multiplicity of activities would be interested to SPHERA, from climate studies to operational forecast, to several downstream applications as ocean modelling, tracer dispersion, energy forecast and risk modelling.

6. Amend for 2020

As a first amend to this proposal, it was decided in 2018 to use ERA5 as driver for SPHERA, instead of COSMO-REA6 (in turn driven by Era-Interim) as initially proposed in this document. This decision was aimed at improving the final quality of SPHERA and followed from the fact that the first trance of ERA5 had been already published and successive trances were expected in a short time. This change arose a new question about the nesting modality of SPHERA in ERA5 (use a one or two step nesting to go to the desired resolution), and a specific experiment was performed in order to tackle the question. Details are given in the special project progress report of 2018.

In agreement with the planned timetable, the setup of SPHERA and part of its production were accomplished during 2018. However, due to some revisions and technical/scientific issues, the production accumulated few months of delay and costed more in terms of computing units. The real amounts of resources used in the production of SPHERA compared to those firstly estimated in this application are reported in Table 3.

	Estimated in the proposal of June 2017	Real use in the production
Computing time required to simulate 1 day (hours)	0.5	1.3 (+ 0.4 average time in queue)
High Performance Computing Facility (SBU) for 1 day	4400	7600
Accumulated data storage (Gb) for 1 day	5.5	10.5

Table 3. First estimate and real use in terms of simulation time, memory storage and computing time requirements for 24hours of SPHERA reanalysis.

Among the causes of delay and of increment of computing resources (details are in the application for additional resources for 2019), the most relevant point was the necessity to use the COSMO model in double precision instead than in single precision as initially proposed. Indeed, after the proposal submission severe bugs have been evidenced in the model code when running in single precision. They have been detected in the data assimilation scheme, in the turbulence scheme and in the soil scheme. In total, the decision to use double rather than single precision model version caused largest part of the increment of the simulation cost (70% higher than estimated) and of the time required to simulate each day (more than doubled), see Table 3.

Moreover, the really-used data storage was higher than the estimated one (Table 3). Mostly this increment is due to the storage of high frequency (sub-hourly) data for wind variables in the vicinity

of the surface (at different heights below 500m). They are intended for evidencing the added value, if any, of a sub-hourly wind reanalysis.

Due to these modifications, in 2018 it was possible to simulate 43% of the period 2003-2017, which is the interval for which it was possible to start the production at that time due to the availability of ERA5 (see the application for additional reasources for 2019 for more details). In 2019, the low amout of billing units allocated for SPHERA project allowed to bring forward the production only up to 57% of the period 2003-2017. The additional reasources required for 2019, if approved, would allow to reach 81% of the period 2003-2017. In this case, SPHERA would be completed for the interval 2011-2017, plus other non-continuous periods between 2003-2010 (see the application for additional reasources for 2019 for more details). The amended timetable for the SPHERA production is reported in Table 4.

The completion of the period 2003-2017 (2,9 years missing) would cost other 8,05 millions of SBU and would occupy other 12.000 Gb. Moreover, the full extent of SPHERA, as proposed in this document, ranges from 1995 up to 2019 (included). Therefore other 11 years still miss. They include 8 years from 1995 to 2002 plus 2 times 6 months of spin-up (for 2 production trances) and 2 years for the forward extension 2018-2019. They would need 30,5 millions of SBU to be generated and 42.160 GB to be stored. The SBU already allocated for 2020 are 2,4 millions. This amend proposes for 2020 the allocation of 38,55 millions SBU (instead of 2,4 millions) and of 54.160GB of the data storage, in order to complete the SPHERA archive in its planned extension (1995-2019). The simulation time should not be an issue, since each day to be simulated takes a bit less than 2 hours (considering also the time spent in queue, Table 3). The production has been segmented in several trances that can run independently (Figure 2). According to this plan, the longest trance would be 4 years long plus 6 months of spin-up. This trance would take about 4.5 months to be completed.

In conclusion, the SPHERA reanalysis dataset produced so far covers 57% of the period 2003-2017. An additional request of resources for 2019, also submitted with this request, would allow if approved to reach about 81% of the period 2003-2017. The present modification of the special project for 2020 would allow to fully cover the planned extent from 1995-2019.

	High Performance Computing Facility (SBU)	Accumulated data storage (Gb)
2018		
Jan-Aug: Set up of the suite, implementation of a verification tool box and preliminary tests (corresponding to 8,9 years of simulation)	24.700.000	12.000 (only 5 years are stored and with reduced amount of data)
Sep-Dec: Reanalysis production (43% of the period 2003-2017, corresponding to 6,5 years of simulation)	18.031.000	44.165
2019		
Jan: Reanalysis production (already reached 57% of the period 2003-2017, corresponding to other 2 years of simulation)	5.548.000 (so there was a bit of exceedence of the allocated amount)	7.667
Jul-Dec: Reanalysis production (if approved the application for additional resource in 2019, the period 2003-2017 will be covered for 81%, corresponding to other 3,6 years of simulation)	10.000.000	14.000

2020		
Reanalysis production (complete the period 2003-2017, corresponding to other 2,9 years of simulation)	8.044.600	12.000
Reanalysis production (complete the near future extension 2018-2019, 2 years of simulation)	5.548.000	7.667
Reanalysis production (complete the past extension 1995-2002 plus 2 times 6 months of spin-up, 9 years of simulation)	24.966.000	34.492

Table 4 Time table of the SPHERA project

Computer resources required for 2018-2020:	**2018**		**2019**			**2020**	
	Allocated	**Used**	**Allocated**	**Used**	**Applied for**	**Allocated**	**Applied for**
High Performance Computing Facility (SBU)	50.240.000	42.730.000	4.340.000	5.548.000	10.000.000	2.420.000	38.550.000
Accumulated data storage (GB,total archive volume)[2,]	55.000	56.165	60.000	63.835	78.000.000	62.500	132.200

Table 5. Computer resources required for 2018-2020, with the original, used and required amounts in the present amend and in the request for additional resources for 2019, submitted together with this amend.

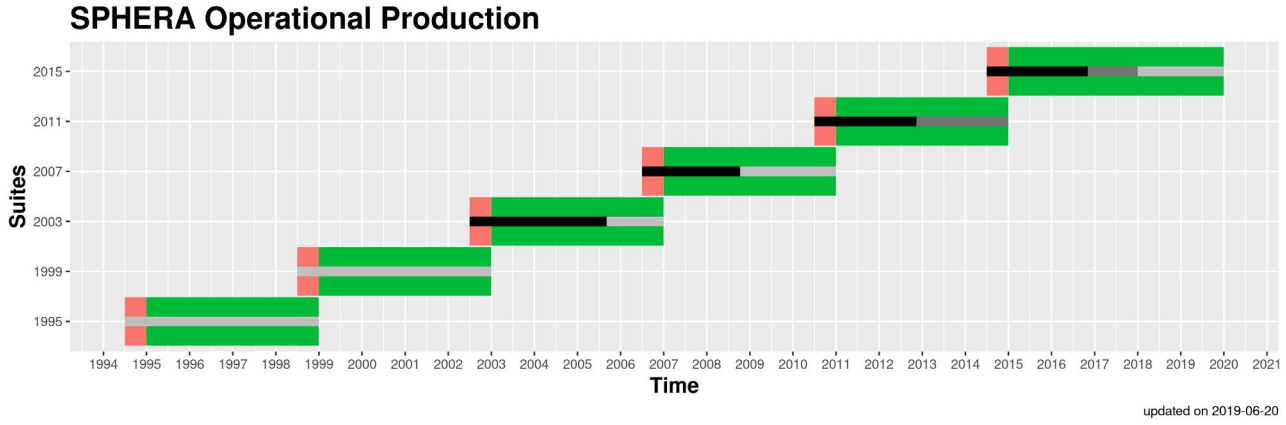

Figure2. Production advancement for the SPHERA dataset, including the already produced periods (black line), the periods that will be produced if the application to additional resources for 2019 will be accepted (dark grey line) and the periods that will be produced if the present amend to the original project will be accepted (light grey line). The colors represent the production phases ('red' indicates the initialization period, while 'green' the production of final data).

References

Antolini G., Auteri, L., Pavan, V., Tomei, F., Tomozeiu, R. and Marletto, V. , 2015. A daily high-resolution gridded climatic data set for Emilia-Romagna, Italy, during 1961–2010. Int. J. Climatol. 36,1970-1986

Arpe, K. 1991. The hydrological cycle in the ECMWF short range forecasts. Dyn. Atmos. Ocean. 16, 33–59

Bach, L., C Schraff, JD Keller, A Hense 2016. Towards a probabilistic regional reanalysis system for Europe: evaluation of precipitation from experiments

Betts, A. K. and Ball, J. H. 1999. Basin-scale surface water and energy budgets for the Mississippi from the ECMWF reanalysis. J. Geophys. Res. 104, 19293–19306.

Bollmeyer J. D. Keller, C. Ohlwein, S. Wahl, S. Crewell, P. Friederichs, A. Hense, J. Keune, S. Kneifel, I. Pscheidt, S. Redl and S. Steinkeet 2014. Towards a high-resolution regional reanalysis for the European CORDEX domain. Q. J. R. Meteorol. Soc. 141. 1-15

Haylock M. R., N. Hofstra, A. M. G. Klein Tank, E. J. Klok, P. D. Jones, and M. New. 2008: A European daily high-resolution gridded dataset of surface temperature and precipitation. J. Geophys. Res (Atmospheres), 113, D20119, doi:10.1029/2008JD10201

Montani A., Cesari D., Marsigli C., Paccagnella T., 2011. Seven years of activity in the field of mesoscale ensemble forecasting by the COSMO-LEPS system: main achievements and open challenges. Tellus, 63A, 605-624.

Pavan V., G. Antolini, G. Agrillo, L.Auteri, R. Barbiero, V. Bonati, F. Brunier, C. Cacciamani, O. Cazzuli, A. Cicogna,C. De Luigi, L. Maraldo,G. Marigo, R. Millini, E. Panettieri, S. Ratto, C. Ronchi, S. Saibanti, A. Sulis, F. Tomei, R. Tomozeiu, I. Torlai and G. Villani 2013. The ARCIS Project. Ital. J. Agromet.,2/2013, 51-55.

Schättler U., G. Doms, C. Schraff 2011. A description of the non-hydrostatic regional model LM – Part VII: Users' Giude., Technical report. Deutscher Wetterdienst: Offenbach, Germany.

Wahl S., C. Bollmeyer, S. Crewell, C. Figura, P. Friederichs ,A. Hense, J D. Keller, C. Ohlwein. 2017. A novel convective-scale regional reanalysis COSMO-REA2: Improving the representation of precipitation, Meteorologische Zeitschrift, doi: 10.1127/metz/2017/0824