REQUEST FOR ADDITIONAL RESOURCES IN THE CURRENT YEAR FOR AN EXISTING SPECIAL PROJECT

MEMBER STATE: ITALY

Principal Investigator[1]: Ines Cerenzia, Tiziana Paccagnella

Affiliation: ARPAE-Emilia Romagna, SIMC

Address: Viale Silvani 6, Bologna, Italy

E-mail: icerenzia@arpae.it

Other researchers: Andrea Montani (currently at ECMWF), Davide Cesari

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

Project account: **SPITCERE**

Additional computer resources requested for	24/06/19
High Performance Computing Facility (units)	10.000.000SBU
Data storage capacity (total) (Gbytes)	18.000 Gb

Continue overleaf

Technical reasons and scientific justifications why additional resources are needed

The production of the SPHERA reanalysis was revised and delayed by few months with respect to the project timetable, as already mentioned in the progress report released in June 2018. The reasons were:

1. Additional time spent to define the proper SPHERA setup. In particular, for the definition of the nesting modality to downscale ERA5 into SPHERA (use a one or two step nesting to go to the desired resolution). This issue was not considered originally because the initial plan was to use as driver COSMO-REA6, i.e. a 6-km resolution reanalysis based on COSMO. The specific test performed to see the impact of the double nesting took about 3-4 months.

2. The underestimation of the time required by a single-day COSMO model run as depicted in Table 1. This created a delay of about 2 months. A first reason for this erroneous estimate was an incorrect set up of the assimilation scheme in the preliminary tests performed to prepare the proposal in June 2017. The second and main reason was the necessity to run COSMO in double precision rather than in single precision as initially proposed; 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 time required to simulate each day (it was more than doubled) (Table 1).

	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	7500
Accumulated data storage (Gb) for 1 day	5.5	10.5

Table 1. First estimate and real use in terms of simulation time, memory storage and computing time requirements for 24hours of SPHERA reanalysis.

Due to these delays, the real production of the SPHERA dataset started in September 2018 rather than in April 2018, as originally estimated, and took longer than originally estimated. Consequently, it was not possible to produce the entire period foreseen for 2018, i.e. from 1995 up to 2016, but only a part of it. According to the proposal submitted in June 2017, almost the entire period was expected to be produced in 2018. Therefore, the majority of resources were allocated for 2018, and much less for 2019 and 2020 (respectively 50.240.000 SBU for 2018, 4.340.000 for 2019 and 2.420.000 for 2020).

In order to employ at best the resources allocated for 2018, the production period was sliced in four production trances of 4 years each (Figure 1), which could run at the same time. The trances are for (1) 2003-2006, (2) 2007-2010, (3) 2011-2014 and (4) 2015-2018. Each trance was preceded by 6 months of rerun needed for the soil spin-up. Years from 1995 to 1999 were not included because in September 2018 ERA5 had not been released yet for this interval. Due to the same reason, it was not possible to produce neither the years between 2000 and 2002. Indeed, the deep soil temperature provided to COSMO as bottom boundary condition is generated applying a temporal delay to the

three-years running mean of the deepest soil temperature of ERA5. Given that ERA5 was available at the time starting from 2000, it was not feasible to run the production earlier than in 2002.
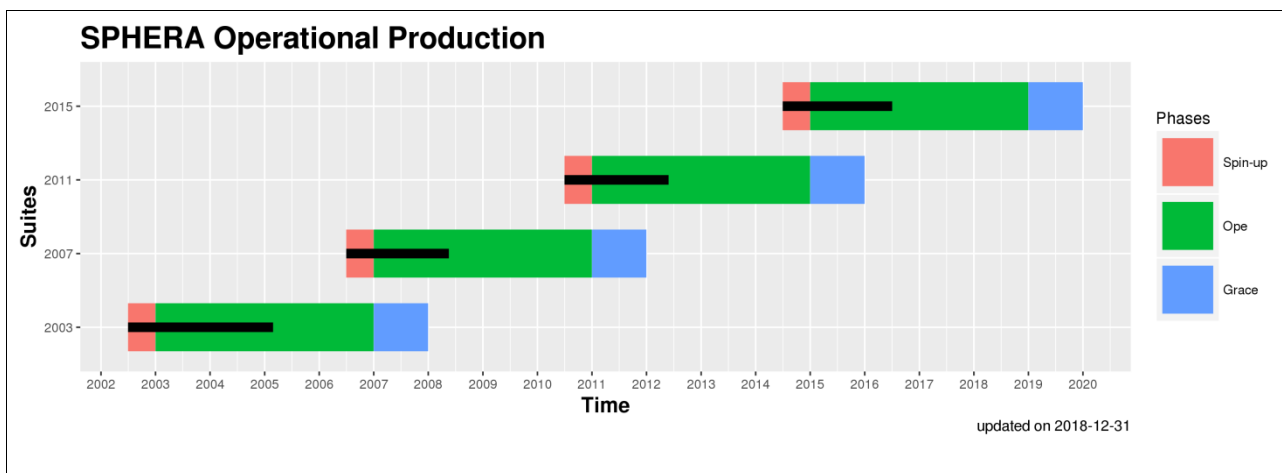

Figure1. Production advancement for the SPHERA dataset updated at 31/12/2018. Black line indicates the periods already produced, while the other colors represent the production phases ('Ope' stays for operational and indicates the production of final data, 'spin-up' indicates the initialization period, while 'Grace' indicates an optional rerun, only for comparison purposes).

In 2018, SPHERA was produced for about two years (or a bit less) plus the 6-months spin-up for each trance (Figure1). In 2018, the total computational resources used (including those used for the preliminary tests) were 42.671.566SBU, thus about 85% of the amount allocated. Thus, essentially this additional request is to employ in 2019 the resources that were allocated and not used in 2018, due to the mentioned delays and to the fact that ERA5 had not been released yet for years before 2000.

In 2019, the entire amount of allocated resources (i.e. 4.340.000SBU) has been used to continue the production of SPHERA along all the four trances. The production advancement is reported in Figure 2.
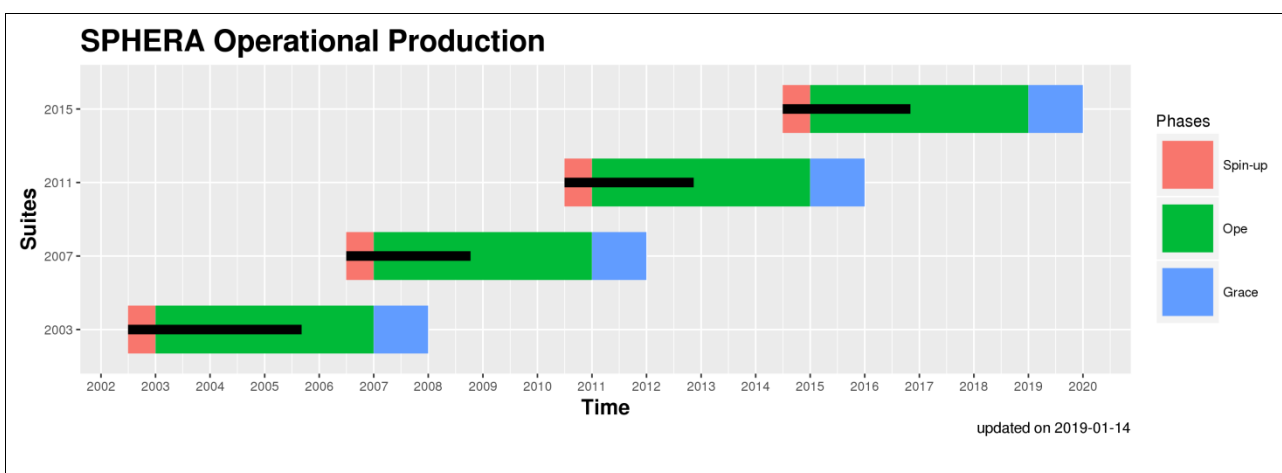

Figure 2. As in Figure 1, but updated at 14/01/2019.

At the current state, approximately 6.5 years are missing to close the gaps between the trances and fully cover the period 2003-2017. Considering that one year of rerun costs about 2.750.000 SBU, the requested additional 10millions SBU would cover at 56% the needs. Priority will be given to conclude the most recent years. The requested additional SBU would allow to fully cover the period 2011-2017. Notice that the computational cost was about 70% higher than the first estimate

(1.600.00 SBU per simulated year), again mainly due to the decision to use double instead of single precision COSMO version (Table 1).

In 2020, the already allocated resources of 2.420.000 SBU will be enough to extend of about one year the reanalysis dataset. However, 2,9 years will be still missing in the interval 2003-2010, as well as the years 2018-2019 and the period 1995-2002 in order to reach the originally planned extent from 1995 up to 2019. A modification to the special project for 2020 will be submitted attached to the present request in order to cover these remaining years.

Finally, the present form includes the request of some additional data storage for the 3.6 years that it will be possible to produce with the requested billing units. Currently the SPHERA dataset occupies 51.830 Gb. Other 12.000 Gb are occupied by the rerun performed in order to define the SPHERA setup (this specific space will be cleaned once the project will be concluded). According to the original proposal, the project in 2019 would occupy 60.000 Gb. The production of 3.6 years corresponds to a storage of 14.000Gb (Table 1). Therefore, additional 18.000Gb are requested (14.000Gb for the new production and other 4.000Gb to cover the small exceedence already done). As presented in Table 1, the really-used data storage was higher than the estimated one. Mostly this increment was 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.

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