LATE REQUEST FOR A SPECIAL PROJECT 2023–2025

MEMBER STATE: Italy

Principal Investigator[1]: Daniele Mastrangelo

Affiliation: CNR-ISAC

Address: via P. Gobetti 101

40129, Bologna, Italy

Other researchers: ……

……

Project Title: Globo Ensemble Reforecast 2023 – GLEREF23

If this is a continuation of an existing project, please state the computer project account assigned previously.	**SP** spitmast	
Starting year: (A project can have a duration of up to 3 years, agreed at the beginning of the project.)	2023	
Would you accept support for 1 year only, if necessary?	YES X	NO ☐

Computer resources required for the years: (To make changes to an existing project please submit an amended version of the original form.)		**2023**	**2024**	**2025**
High Performance Computing Facility	(SBU)	~~35000000~~	20000000	
Accumulated data storage (total archive volume)[2]	(GB)	~~13500~~	13500	

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: Daniele Mastrangelo

Project Title: Globo Ensemble Reforecast 2023 – GLEREF23

Extended abstract

Project description

Since 2015, the ISAC Institute of the Italian National Research Council (CNR) routinely operates a subseasonal forecasting system to issue 32-day ensemble forecasts on a weekly basis. The forecasting system is based on the Globo model (Malguzzi et al., 2011), an atmospheric general circulation model originally developed at ISAC as the global version of a pre-existing regional-scale hydrostatic model. The ensemble forecasts are contributed to the database of the WWRP/WCRP/THORPEX project on subseasonal to seasonal (S2S) prediction. This database is coordinated by ECMWF as a major goal of the S2S project (Vitart et al., 2017) and stores forecast and reforecast runs from 12 forecasting centres. Also, the ISAC real-time forecasts are provided to the National Civil Protection Department in the framework of a multi-annual cooperation between the two institutions (Mastrangelo et al., 2021), a cooperation that originally promoted the development of the ISAC S2S forecasting system.

Numerical atmospheric forecasts are affected by errors deriving from uncertainties in the initial conditions, from approximations intrinsic to the adopted numerical methods, and from the necessity to approximate the unresolved, subgrid physical processes. The error grows as the forecast lead time increases and is a function of the location and of the period of the year. A large number of simulations is necessary to evaluate the systematic model error and, subsequently, to calibrate the forecast fields by removing it, especially for long-term predictions as subseasonal or seasonal are (e.g., Kharin and Zwiers, 2003). A possibility to evaluate and minimize the model bias is given by the use of simulations of a past period for which the verifying observations (usually reanalysis gridded data) are already available. These simulations are the so-called reforecasts, retrospective forecasts generated with the same numerical model used to produce the real-time forecasts (e.g. Guan et al., 2022).

The current CNR-ISAC subseasonal forecasting system (Mastrangelo and Malguzzi, 2019) is based on a reforecast dataset covering the 1981-2010 period with 5-member ensembles initialized every 5 days (73 initialization days per year for a total of 10950 runs) with the ECMWF ERA-Interim reanalyses. This “fixed” dataset is routinely used to calibrate the real-time 41-member ensemble forecasts: a combination of the reforecasts initialized in the calendar days adjacent to the initialization date is adopted (Mastrangelo et al., 2012), with weights distributed following a Gaussian curve. This combination smooths the climatological fields to minimize the noise deriving from short-scale atmospheric dynamics and retaining the longer scale signal.

Recently, a new version of the Globo model was set up and preliminarily evaluated through a limited number of reforecasts initialized in a subset of winter and summer dates of the 2001-2020 period, selected as reference. The new version of Globo, as implemented in that test, resulted in a gain of predictive skill of approximately one day for the typical meteorological variables upon the old reforecasts initialized in the same dates. Following the positive outcomes of this test, the aim of this project, GLEREF23, is to produce the complete set of new Globo reforecasts as described in the next section.

The planned reforecasts will constitute a comprehensive dataset serving different aims in addition to real-time forecast calibration. Such a comprehensive dataset can be used (i) to evaluate the model performance and skill over a climatologically significant period (20 years), (ii) to test any possible algorithm (e.g., calibration technique, coupling with hydrological models, model output statistics techniques) that needs a training dataset spanning a climatological reference period, and (ii) as a reference for any new update of a model component or of the forecasting system. For instance, following one of the main goals of the S2S Prediction Project, more than 30 scientific papers—at our best knowledge—using the old Globo reforecast dataset have been published, mainly in the context of multi-model intercomparison studies on model forecasting skill (e.g. de Andrade et al., 2019) or process studies (e.g. Stan et al., 2022).

However, beyond the scientific goals, a prominent outcome of the project is operational: since the new reforecasts are used to calibrate the real-time forecasts, the new version of the CNR-ISAC subseasonal forecasting system, based on the updated Globo model, can be implemented operationally after having completed this new reforecast dataset. The new forecasting system will allow to continue all the active projects, the operational activity, and will be uploaded on the S2S database for the last phase of the project. For all these reasons, the simulations planned for GLEREF23 would be desirably completed as soon as possible.

Computational Approach

This project is based on the use of the atmospheric general circulation model GLOBO, coded in Fortran90 for a CPU-based computing architecture. The ideal computing environment needs Intel Fortran Compiler (ifort and mpiifort commands) and packages necessary for I/O such as ecCodes and CDO. The work flow is based on three steps performed by three different F90 files:

- pre-processing: external grib files are ingested to produce the initialization fields for the GLOBO grid;
- GLOBO simulation;
- post-processing: the binary GLOBO output is converted into grib2 files.

Pre- and post-processing jobs are serial.

The basic features of the new grid set up and reforecasting strategy are summarized in Table 1. For the planned new reforecast dataset, the regular lat-lon grid of the model will be implemented with 514 lon points, 362 lat points, 7 ground levels, and 70 vertical levels setting the top lid at about 1.4 hPa. The new vertical set up represents a relevant step forward as compared to the previous version of this model. This enhancement follows recent scientific outcomes in the S2S community: several publications have for instance highlighted how high-top models, with a greater number of model levels in the upper troposphere/lower troposphere, have minor model bias at that height, with a possible increase in the forecasting skill of the stratospheric dynamic and associated impact on the prediction of surface variables on the subseasonal scale (e.g., Lawrence et al., 2022).

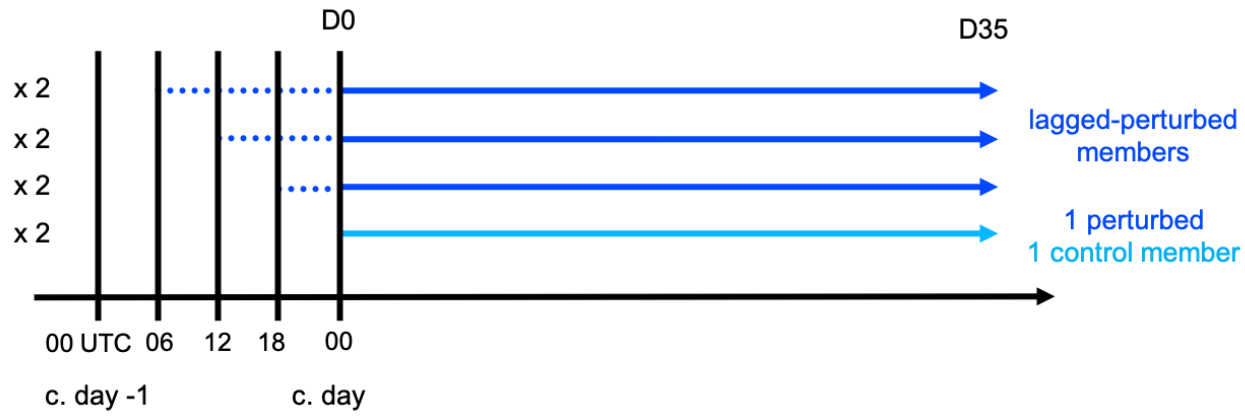
	Current reforecast dataset	**GLEREF23**
grid points: lon-lat	450 x 322	514 x 362
vertical levels	54	70
model top	~ 7 hPa	~1.4 hPa
initialization data	ERA Interim	ERA5
Members	5	8
reference period	30 y: 1981-2010	20 y: 2001-2020
initialization dates	every 5 d starting on 1 Jan	every 5 d starting on 1 Jan
runs	10950	11680

Table 1: main features of the model grid set up and reforecasting strategy between the reforecasts of the current and planned operational version.

ECMWF has recently introduced the new ERA5 reanalyses that, replacing ERA-Interim with various improvements, implements an ensemble data assimilation: for a selected field and output time, ERA5 consists of a higher resolution member and a reduced-resolution 10-member ensemble that provides an estimate of the relative, random uncertainty of the selected field (Hersbach et al., 2020). The possibility to have different

reanalysis fields can be used to initialize several members of ensemble numerical simulations over a potentially large period of the past (Vitart et al., 2019).

For the CNR-ISAC subseasonal forecasting system, the ERA5 dataset gives the opportunity to enlarge the reforecast ensemble: with 2 ERA5 ensemble fields for each synoptic hours of the 18-h period ending at 00 UTC, 8 Globo simulations will be run for each of the 73 nominal initialization dates (Fig. 1). To cover the 20-year reference climate period ending in 2020, a total of 11680 reforecast runs is therefore planned for GLEREF23.



- **Computational resources**

~~This new GLOBO set up has been already tested on the ECMWF'S high-performance facility (HPCF) installed in Bologna, Italy, thanks to user "ecff". We ported the same configuration recently and successfully optimized on the Italian cluster Galileo100 operated by CINECA (https://www.hpc.cineca.it/hardware/galileo100). The parallelization (MPI libraries) of the global domain is performed subdividing it on 16 (W-E direction) x 12 (S-N direction) subdomains. Since every HPCF node is made up of 128 physical cores, the resulting 192 GLOBO processes required 2 nodes on HPCF: a single run, consisting of 35 days of forecast range, took ~32 mins of run time with a total of ~2500 SBU. The basic time step is 300 sec, which has to be shortened to run possibly uncompleted simulations. No issues with the available memory were found.~~

~~The total number of 35-day reforecast comes out as follows: 8 members x 73 initialization dates x 20 years = 11680 runs. Therefore the basic amount of SBU would be 11680 * 2500 = 29.200.00. However, considering:~~

~~- resources to re-run possible uncompleted simulations that can usually sum up to 10% of all the runs,~~
~~- resources for pre- and post-processing output files~~

~~the total required computational resources = 35000000 SBU.~~

In the first phase of the project I could set up GLOBO as already planned: a single job (reforecast) features 192 processes requiring 2 nodes of the Atos HPCF. However, the computing resources needed for such a job are about 2/3 of what expected: a single run, consisting of 35 days of forecast range, takes ~20 mins with a total of ~1600 SBU. This performance is obtained with Intel Fortran compiler and mpiifort libraries as planned in the "Computational approach" Section. One of the main reason for this performance can be the use of the ifort instead of gfortran, the compiler adopted in the test mentioned in the original request.

The total number of 35-day reforecasts comes out as follows: 8 members x 73 initialization dates x 20 years = 11680 runs. However, as of 21 December 2023, the reforecasts initialized on 14 of the 73 initialization (calendar) dates have already been produced consuming ~ 4.000.000 of the 35.000.000 SBU initially requested for the whole project.

The remaining amount of simulations, therefore, comes out as follows: 8 members x 59 initialization dates x 20 years = 9440 runs. And the basic amount of SBU would be 9440 runs * 1600 SBU/run = 15.104.000 SBU.

Considering:

- resources to re-run possible uncompleted simulations that can usually sum up to 10% of all the runs,

- resources for pre- and post-processing output files,

the updated total computational resources required to complete this project is 20.000.000 SBU.

- **Storage resources**

Each GLOBO run produces two binary files: a "low res" of ~ 324 MB and a "high res" of 1.23 GB. These files are converted in grib2 files to become 240 MB and 0.9 GB respectively. The 11680 runs planned for the project will produce:

- low res files: ~ 2740 GB
- high res files: ~ 10512 GB

Therefore, I confirm that the total storage required for the whole project is 13500 GB.

A final note for this extension request of the pre-existing GLEREF23 late-2023 Special Project: considering that the whole simulating process is fully set up, it could also be good for me having the requested resources available over a period of only 4 full months, the amount of time I consider sufficient to complete the project.

References

de Andrade, F.M., Coelho, C.A.S. & Cavalcanti, I.F.A. Global precipitation hindcast quality assessment of the Subseasonal to Seasonal (S2S) prediction project models. Clim Dyn 52, 5451-5475 (2019). https://doi.org/10.1007/s00382-018-4457-z

Guan, H., Zhu, Y., Sinsky, E., Fu, B., Li, W., Zhou, X., Xue, X., Hou, D., Peng, J., Nageswararao, M. M., Tallapragada, V., Hamill, T. M., Whitaker, J. S., Bates, G., Pegion, P., Frederick, S., Rosencrans, M., & Kumar, A. (2022). GEFSv12 Reforecast Dataset for Supporting Subseasonal and Hydrometeorological Applications, Monthly Weather Review, 150(3), 647-665 https://doi.org/10.1175/MWR-D-21-0245.1)

Hersbach, H, Bell, B, Berrisford, P, et al. The ERA5 global reanalysis. Q J R Meteorol Soc. 2020; 146: 1999–2049. https://doi.org/10.1002/qj.3803

Kharin, V.V. and Zwiers, F.W. (2003) Improved seasonal probability forecasts. Journal of Climate, 16, 1684–1701.

Lawrence, Z. D., Abalos, M., AyarzagÃ1⁄4ena, B., Barriopedro, D., Butler, A. H., Calvo, N., de la CÃ¡mara, A., Charlton-Perez, A., Domeisen, D. I. V., Dunn-Sigouin, E., GarcÃ-a-Serrano, J., Garfinkel, C. I., Hindley, N. P., Jia, L., Jucker, M., Karpechko, A. Y., Kim, H., Lang, A. L., Lee, S. H., Lin, P., Osman, M., Palmeiro, F. M., Perlwitz, J., Polichtchouk, I., Richter, J. H., Schwartz, C., Son, S.-W., Statnaia, I., Taguchi, M., Tyrrell, N. L., Wright, C. J., and Wu, R. W.-Y.: Quantifying stratospheric biases and identifying their potential sources in subseasonal forecast systems, Weather Clim. Dynam., 3, 977-1001, https://doi.org/10.5194/wcd-3-977-2022, 2022.

Malguzzi, P., Buzzi, A., & Drofa, O. (2011). The Meteorological Global Model GLOBO at the ISAC-CNR of Italy Assessment of 1.5 Yr of Experimental Use for Medium-Range Weather Forecasts, Weather and Forecasting, 26(6), 1045-1055. https://doi.org/10.1175/WAF-D-11-00027.1

Mastrangelo D., Malguzzi P., Rendina C., Drofa O., and Buzzi A.: First outcomes from the CNR-ISAC monthly forecasting system, Adv. Sci. Res., 8, 77-82, 2012 doi:10.5194/asr-8-77-2012

Mastrangelo, D., & Malguzzi, P. (2019). Verification of Two Years of CNR-ISAC Subseasonal Forecasts, Weather and Forecasting, 34(2), 331-344. https://doi.org/10.1175/WAF-D-18-0091.1

Mastrangelo D., Delli Passeri L., Campione E., Malguzzi P., 2021: The contribution of S2S forecasts to the activities of the Italian Civil Protection Department. S2S Newsletter, No. 17 (Aug. 2021) http://s2sprediction.net/file/newsletter/Newsletter%2017_Aug%202021.pdf

Stan, C., Zheng, C., Chang, E. K., Domeisen, D. I., Garfinkel, C. I., Jenney, A. M., Kim, H., Lim, Y., Lin, H., Robertson, A., Schwartz, C., Vitart, F., Wang, J., & Yadav, P. (2022). Advances in the prediction of MJO-Teleconnections in the S2S forecast systems, Bulletin of the American Meteorological Society (2022). https://doi.org/10.1175/BAMS-D-21-0130.1

Vitart, F. and Coauthors,The subseasonal to seasonal (S2S) prediction project database, Bulletin of the American Meteorological Society, Volume 98, Number 1, p.163-173, (2017), doi: http://dx.doi.org/10.1175/BAMS-D-16-0017.1

Vitart, F., et al. Use of ERA5 to initialize ensemble re-forecasts. European Centre for Medium Range Weather Forecasts, 2019. http://doi:10.21957/w8i57wuz6