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

Please email the completed form to special_projects@ecmwf.int.

MEMBER STATE: Spain

Principal Investigator[1]: Víctor Homar Santaner

Affiliation: Universitat de les Illes Balears

Address: Cra. Valldemossa, km 7.5
07122 – Palma de Mallorca (Illes Balears)

Other researchers: Diego Saúl Carrió Carrió,
Alejandro Hermoso Verger,
Aina Maimó Far

Project title: Exploring the Predictability Limits of Severe Weather in the Western Mediterranean: Use of Ensemble Data Assimilation Systems, Stochastic Techniques and Sensitivity Calculation Methods

Project account: **SPESHOMA**

Additional computer resources requested for		2020
High Performance Computing Facility	(units)	10000000
Data storage capacity (total)	(Gbytes)	20000

Continue overleaf

[1] The Principal Investigator is the contact person for this Special Project

Technical reasons and scientific justifications why additional resources are needed

This special project is focused on the predictability of severe weather in the Mediterranean region by means of cutting-edge ensemble generation strategies techniques.

The forecast of socially sensitive aspects of extreme events, such as the distribution, span, timing, mode, and intensity of heavy precipitation requires high spatial resolutions in order to accurately represent these features at catchment scale and produce valuable hydrometeorological forecasts. Furthermore, the greater number of degrees of freedom involved in small scale processes demands the use of ensembles of large size to adequately sample the uncertainty present in initial conditions and model formulation and produce profitable probabilistic forecasts.

Computational resources granted for this year have been used to study the predictability of a heavy precipitation and flash flooding event occurred on 12 and 13 September 2019 in Eastern Spain. The performance of data assimilation based on Ensemble Kalman Filter (EnKF), bred vectors and stochastic physics different ensemble generation strategies has been evaluated. The model used to perform the experiments is the WRF-ARW with a domain that covers the Western Mediterranean region (750x500 grid points), 2.5 km horizontal resolution domain and 50 vertical levels, and the ensembles are composed by 50 members. Given the long duration of the episode, forecasts are divided into several assimilation and forecast cycles, multiplying the computational cost and limiting the number of HPC infrastructure candidates to run the experiments promptly.

Results obtained from the completed experiments motivate the proposal of the following specific activities:

1) New ensemble experiments based on tailored bred vectors: Preliminary results of an ensemble based on tailored bred vectors do not show the expected improvement in skill with respect to a downscaling ensemble generated from the 50 members of the global ECMWF ensemble. It is expected that perturbations generated at convective scale (2.5 km) using the same model used to produce the forecasts represent errors present in initial conditions better than downscaled perturbations generated at a lower horizontal resolution (~18 km).

Some hypotheses to explain this unexpected behaviour are that small scale perturbations grow fast during the first hours of simulation but saturate early so that perturbation growth is low during the last part of the simulations. In addition, the use of orthogonalization during the breeding cycle combined with the large number of perturbations generated can give rise to perturbations in areas where the errors are not relevant for this episode.

In this sense, a new ensemble experiment with bred perturbations with a larger horizontal scale is proposed to analyse the effect of modifying the perturbation scale in ensemble performance.

2) Stochastic parameterizations experiments: Experiments performed with the SPPT scheme show a similar performance than multiphysics and downscale ensembles for this episode, although some significant differences are obtained during some phases of the episode. Additional experiments will allow the introduction of some improvements in this technique and evaluate their impact on this episode. For instance, the tendency obtained for each independent parameterization can be stochastically perturbed rather than perturbing the full physics tendency, which includes effects of all tendencies which can be substantial but opposite, yielding a practically zero resulting physics tendency, which is not affected by the SPPT methodology. In addition, the introduction of stochastic perturbations into the microphysics parameterization, which are not currently available in WRF will be tested.

3) EnKF experiments: The proposed experiments consist in implementing a high-resolution EnKF system assimilating conventional observations and Atmospheric Motion Vectors to reduce large scale errors and assimilating vertical profiles of humidity and temperature from Polar satellites (POES) to reduce convective scale errors. In order to run such experiments, we start our simulation 12 hours before the episode took place, in which the first 6 hours are used to deal with the spin-up effect of the numerical weather model and the following 6 hours are intended to assimilate the abovementioned observations using 20-min data assimilation cycles. From the last data assimilation cycle, it is obtained the new initial conditions that we will use to initiate our 24 h ensemble forecast.

4) Shorter lead time experiments: Accurate hydrometeorological forecasting relies on the ability of meteorological forecasts to produce adequate precipitation estimates at the catchment scale. For this episode, the small scale of the affected catchments (100-1000 km^2) requires high spatial resolution, which leads to a greater influence of nonlinear processes and shorter predictability horizons, so that lead times of 24-30 h may be inadequate for this episode. Indeed, some phases of the episode are not accurately reproduced with simulations which begin 12-18 hours before the initiation of the convective systems involved. This circumstance motivates the design of new experiments with shorter lead times with the different techniques used (downscaling, multiphysics, stochastic, bred vectors, EnKF) in order to investigate the predictability of the different phases of this episode.

The above-mentioned research and specific activities and jobs heavily rely on the use of the ARW-WRF. We plan on running multiple instances of the ARW-WRF model, well known in the HPC community for its scalability and up-to-date standards in scientific programming and performance. The simulation domain covers the entire Western Mediterranean basin with 50 ensemble members at 2.5 km and 50 vertical levels, so this piles up to ~750x500x50x50 grid points. In addition, the stochastic algorithms for the physical parameterizations make use of Fast Fourier Transform to analyse and determine the spectral properties of the perturbations. In order to achieve significant statistical results on the ensemble generation proposed, ensemble size must be moderately large. This factor, combined with the use of bred vector calculation and stochastic physics parameterization results in an extremely demanding experimental setup. Contrary to standard simulation configurations, which might ultimately be run on smaller-scale computers with impracticably longer running times, the proposed experiments at useful and socially competitive scales can only be run on supercomputing facilities such as ECMWF HPCF.

Regarding the EnKF data assimilation experiments, apart from the use of the ARW WRF model we also will use the suite of programs used to implement the Ensemble Kalman Filter, which is the Data Assimilation Research Testbed (DART), from the NCAR/UCAR group (Colorado, US). The DART filter is an extremely demanding algorithm that needs the entire statistical description of the atmosphere (e.g. 50 ensemble members) in memory to perform the ensemble assimilation step. Contrary to the forecasting step, which might ultimately be run on smaller-scale computers, the assimilation steps require heavy use of memory, disk access and CPU.

All algorithms and codes on which this project relies are already tested and fully functional in ECMWF HPCF. The computational cost of each ensemble experiment for this episode is approximately 1 million SBU. This ambitious proposal requires considerable computational resources to complete all the proposed activities and have sufficient resources to determine the optimal configurations of the different techniques proposed. The estimated resources required for each of the activities proposed is the following:

1) 1 million SBU to perform an additional experiment based on bred vectors with a larger perturbation scale

2) 2 million SBU to test new stochastic configurations based on independent perturbations for different parameterizations and microphysics perturbations.

3) 2 million SBU to perform additional data assimilation experiments modifying the tested configurations and the observations assimilated.

4) 5 million SBU to run shorter lead time experiments, dividing the episode in different phases and running an ensemble for the different techniques analysed: bred vectors, stochastic parameterizations, EnKF and the standard downscaling and multiphysics approaches.