REQUEST FOR A SPECIAL PROJECT 2019–2021

MEMBER STATE: Spain

Principal Investigator[1]**:** Víctor Homar Santaner

Affiliation: Universitat de les Illes Balears

Address: Cra. de 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

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.)	2019	
Would you accept support for 1 year only, if necessary?	YES ☒	NO ☐

Computer resources required for 2019-2021:
(To make changes to an existing project please submit an amended version of the original form.)

		2019	**2020**	**2021**
High Performance Computing Facility	(SBU)	12.000.000	12.000.000	12.000.000
Accumulated data storage (total archive volume)[2]	(GB)	20 000	20 000	20 000

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] 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: Víctor Homar Santaner

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

Extended abstract

The completed form should be submitted/uploaded at https://www.ecmwf.int/en/research/special-projects/special-project-application/special-project-request-submission.

All Special Project requests should provide an abstract/project description including a scientific plan, a justification of the computer resources requested and the technical characteristics of the code to be used.

Following submission by the relevant Member State the Special Project requests will be published on the ECMWF website and evaluated by ECMWF as well as the Scientific and Technical Advisory Committees. The evaluation of the requests is based on the following criteria: Relevance to ECMWF's objectives, scientific and technical quality, disciplinary relevance, and justification of the resources requested. Previous Special Project reports and the use of ECMWF software and data infrastructure will also be considered in the evaluation process.

Requests asking for 1,000,000 SBUs or more should be more detailed (3-5 pages). Large requests asking for 10,000,000 SBUs or more will receive a detailed review by members of the Scientific Advisory Committee.

Project Description

The Mediterranean region is characterized by an increasing demography, leading to urban sprawl along coastal lands. In a context of climate change, the population is increasingly exposed to challenging environmental impacts, such as short-time extreme events (flash floods, large hail, severe winds, storm surges or intense lightning) and long-term changes (frequency and intensity of extremes, energy and water resources availability). Resilient societies incorporate diverse mechanisms for living with, and learning from, changes and shocks, which must be fully understood before adaptive policies are adopted. In this context, the recently granted national 3-year project COASTEPS, led by the PI of this proposal, aims at providing basic and applied understanding on the main hydrometeorological and climatic threats that Western Mediterranean coastlands face in the current and projected climates. Increasing the lead time and accuracy of high resolution hazardous weather and flood warnings and forecasts, in order to reduce loss of life, injury, and damage to the economy, is a key challenge in the Mediterranean agenda of sustainable development. In recent years, the adoption of dynamical statistical methods to encompass the evolution of the atmospheric state has proven extremely valuable for structured risk management and civil protection protocols. However, the exact formulation of the uncertainties in the description of the atmospheric state is not well known. COASTEPS proposes to explore the potential of Kalman filtering ensemble data assimilation methods, stochastic physics and the use of sensitivity information as key directions towards better and more useful high-resolution forecasts. COASTEPS focusses the research efforts of an internationally renowned scientific team on improving the resilience of Western Mediterranean coastlands by exploring advanced forecasting tools and the challenges posed by Mediterranean weather extremes in the future climate.

The consequences of the natural hazards related to the water cycle that concentrate in the Mediterranean often reach catastrophic levels and, in terms of economic cost, a single event can exceed the impact of tens or hundreds of ordinary storms. The capability to predict such high-impact events at sub-kilometric scales over useful time-spans remains weak due to the contribution of very fine-scale processes, frequently undetected, and their non-linear interactions with larger scale processes. Advances in the identification of the predominant processes and particularly of their interactions at the different scales are needed in order to better forecast these events and reduce uncertainties on the prediction at short time scales. Improving the skill of numerical weather prediction systems combined with emerging model post-processing techniques and model diagnostic tools, will be the dominant source of better weather information for operational prediction offices during the next decade.
The general physical processes involved in the development of high impact phenomena affecting the coasts are well understood and described in a generally accepted conceptual model. Convective initiation often occurs over maritime bodies, posing a tremendous forecasting challenge addressed in this project. Convection-allowing ensemble data assimilation methods provide the best framework to approach the challenge of forecasting high impact weather phenomena with the current computational means. The enlightening study of Aksoy et al. (2010) offered a clear vision of

the endeavor this new paradigm supposes. How to represent initial condition uncertainty at the convective scales and to initialize ensemble prediction systems is still an open question. Ensemble Kalman Filters (EnKFs) use an ensemble to calculate a dynamical background error statistics. Current seamless integration of data assimilation with forecasting renders dynamically consistent initial ensembles. Model error is an important source of uncertainty in high-resolution forecasts. Despite the various approaches attempted in the past, stochastic perturbation to the model physics offers a rigorous scientific solution. The latest available version of WRF includes both the stochastically perturbed physics tendencies of Berner et al. (2015) and the stochastically perturbed parameter scheme of Jankov et al. (2016).

The proposed project aims to explore the potential of a promising assimilation and prediction strategy, including stochastic physics perturbations, to improve the short-range forecasts of weather phenomena affecting Mediterranean coastlands.

Scientific Plan

The ever-increasing societal demand for accurate phenomenological, spatial and temporal descriptions of future atmospheric states requires an increase in spatial and temporal resolution of numerical models. The conquest of the explicit model representation of convection involves challenges that are directly linked to predictability. Characterizing and quantifying the causes of limited predictability in mesoscale phenomena that produces high social impact in the Mediterranean poses an important challenge for the atmospheric community.

The atmosphere is modelled as a dynamic system whose state evolves over time. It must be assumed that the state of the atmosphere cannot be characterized with infinite precision. The mathematical entity that quantitatively describes inaccurate information is the probability function. Therefore, the mathematical representation of the present, past or future atmospheric state should be described in probabilistic terms. The evolution of the PDF of states of a deterministic dynamical system is governed by the Liouville equation (Leutbecher and Palmer, 2008). In atmospheric prediction science, the deterministic governing dynamics are the Navier-Stokes equations, together with the processes modelled in the physical parametrizations. However, the chaotic nature of such a set of equations, together with the uncertainties and approximations necessary in the numerical treatment and resolution of the equations, introduce errors in the solution, increasing the uncertainties associated with the predicted atmospheric state. Parameterization of the sources of error in the model can be treated with the Fokker-Plank equation (Penland 2003, Ehrendorfer 1994), which allows introducing stochastic terms in the evolution equations of the PDF. Even with current supercomputers, the solution of the Liouville or Fokker-Plank equations are still intractable for realistic problems due to the enormous dimensionality of the system.

Among the multiple sources of error that can contribute negatively to the predictability of the abovementioned phenomena, the most important are the following:

1) <u>Initial Condition problem:</u> EPS are still the best method to produce atmospheric probabilistic forecasts. The resolution required to adequately simulate high impact phenomena poses new challenges in terms of EPS generation strategies. Increased relative importance of nonlinear processes, increased number of degrees of freedom and reduced knowledge about relevant physical processes that are crucial for the unfold of the severe phenomena contribute to the challenge. In this context, bred vectors are hypothesized to efficiently and accurately sample the subspace of actual uncertainties in the initial conditions of a kilometric and subkilometric atmospheric systems. Traditional arithmetic bred vectors (Toth and Kalnay, 1993) have been recently challenged by logarithmic bred vectors (Primo et al. 2008) and rescaled bred vectors as more versatile and adjusted to the dynamic system nature of the evolution of perturbations. Furthermore, poor representation of the state of the atmosphere over maritime areas, where most of the severe weather affecting the coastal areas of the Mediterranean Sea initiates, poses a prediction challenge due mainly to the lack of *in-situ* observations. Data assimilation methods determine the most likely estimate of the atmospheric state by blending information form an initial guess (background) and a set of observations. In this proposed project we are going to use the Ensemble Kalman Filter algorithm, a linear and recursive estimator that produces unbiased minimum variance estimates, in a least square sense, of the atmospheric state, to assimilate conventional and remote sensing observations (radar, lidar, satellite and GPS among others). The advantage of the EnKF in comparison with other data assimilation methods, such as variational techniques (e.g., 3DVar or 4DVar) is that EnKF dynamically evolves the background error covariance matrix, improving upon static climatological estimates.

2) <u>Model Error:</u> The uncertainties inherent to physical parameterizations, either from incomplete process understanding or the dilemma of representing the impact of unresolved processes on the resolved scales, require a fundamentally different approach. Elements of parameterizations or entire schemes are likely to require components that appear statistical to the large scales because they are not fully determined by the resolved scales. How intense this approach needs to be is currently matter of active research across the field. Bauer et al. (2015) highlight the area of stochastic parameterizations as requiring much more attention in the future as it promises significant boosts of skill, but also involve substantial investments in scientific development and computing.

Relevant physical parametrizations for severe weather forecasting such as planetary boundary layer or convection have proved to increase forecast skill when physical parameters are adequately perturbed. The introduction of stochastic sources may create spatio-temporal imbalances in the solution, which must be prevented by using spatial and temporal correlations in random perturbations.

3) Observational error: Although the observational error is an important source of error to take into account, its study is beyond the scope of this project. Observational means used in this project are all quality controlled and issued by official weather services.

SPECIFIC ACTIVITIES

For the proposed project, the following specific activities are planned:

a) **PDF forecast by means of the Liouville equation**

Despite being intractable for complex fully realistic models, the currently most advanced HPCs begin to allow adventuring to solve for the first time the Liouville and Fokker-Plank equations for configuration of atmospheric interest. In particular, the simplification of an atmospheric evolution with a barotropic model and making use of dimension reduction by means of principal components makes the problem tractable. In this sense, this is a multi-parameter range-sweep study. Simple toy solutions of the barotropic model already reveal unanticipated characteristics such as a granulated support of the predicted PDF, which contradicts popular and widespread use and interpretation of current ensemble prediction systems (EPS). The specific calculations to be performed are a massive number of deterministic integrations of the nonlinear barotropic model for a low spectral representation of a historical extreme cyclogenetic event occurred in November 2001 over the western Mediterranean. In case ECMWF throughput does not allow for 10 dimensions and a parametric resolution of 1/100 over the range of plausible states over the entire simulation period, the number of PCs will be reduced. Indeed, the number of required simulations is ND^Np, with ND the number of dimensions and Np the number of discretization points in phase (parameter) space in each direction. For the proposed 10 dimensions and 100 points, the number of required parallel non simultaneous runs will be 1.E20. A first experiment will consist of solving the Liouville equation and analyzing the topological characteristics of the resulting 10-dimensional PDF. The second experiment will activate stochasticity and thus solve the Fokker-Plank equation with intermediate intensity noise. The comparison between both massively parallel experiments will shed light on the effects stochasticity plays on the topological characteristics of a probabilistic forecast of the atmospheric state.

These experiments are of uttermost interest for COASTEPS, as they explore the topological characteristics of the evolution of the most fundamental information expressing probabilistic forecasts in the atmosphere. The hypothesis that exact forecasted PDFs may show irregular topological features, even when initial condition PDFs have a compact support, would question most postprocessing and interpretation methods used nowadays for EPS for weather and climate.

In case the hypothesis is confirmed during this ECMWF project and irregular (e.g. granulated) pdf support is discovered, we will explore the role of the stochastic noise intensity on the forecast pdf and will also test the conclusions obtained for the case of November 2001 to other situations, such as Medicane formation or heavy precipitation cases.

b) **EPS generation: sampling IC error**

b1) Implementation of a high-resolution EnKF data assimilation system that incorporates *in-situ* conventional, Doppler radar (reflectivity and radial velocities) and satellite observations using an ensemble of 50 members. Conventional data information from 4 different platforms is available: radiosondes, METARS, maritime instruments, aircrafts. Such instruments measure both component of the wind, temperature, humidity and pressure through the entire atmosphere. Doppler radars data contain high temporal and spatial resolution reflectivity and radial velocity observations distributed in multiple tilted radar scans. Among the available set of miscellaneous products retrieved from meteorological instruments onboard satellites, assimilation of high resolution Rapid Scan Atmospheric Motion Vectors are expected to improve maritime severe weather events initiated and developed over the sea, providing a better representation of the environmental circulation through the entire atmosphere. One of the most important advantages of satellite observations is their high spatial coverage over areas where conventional and radar instruments cannot reach (e.g., maritime regions).

The proposed activity pursues the main objective of improving the above mentioned predictability challenges of severe weather events that mainly initiate and develop over the Mediterranean Sea and eventually affect populated coastal areas within the COASTEPS framework. This activity aims to continue exploiting the assimilation of satellite derived products over different COASTEPS case studies to study the predictability benefits for extreme events, with special attention on the Mediterranean Hurricanes (Medicanes). Current satellite instruments are very complex, measuring thousands of simultaneous correlated observations in the same area. The data assimilation community, use the "Thinning" and "Superobbing" techniques to reduce the effect of spatial correlated errors for such observations. Consequently, we will further exploit these techniques in the proposed activity. Finally, additional refined quality control algorithms will be developed with the main purpose of discarding no-physical observations assimilation.

Verification of the results obtained are divided in two phases. In the first phase, the performance of the data assimilation system using the EnKF will be evaluated through observation-space diagnostics. Root-mean-square innovations, total spread and consistency ratio. Once the resulting analyses show acceptable attributes, we will able to move on to the second phase, the verification of the forecast results. Probabilistic verification scores as the Brier Skill Score, Taylor diagrams, ROCs curves and reliability diagrams among others will be used to quantitatively assess the skill of the experimental forecasts.

Detailed tasks planned to be performed in this activity are listed below:

1) Hourly assimilation of available quality controlled conventional in-situ observations from NCEP-MADIS database (https://madis.noaa.gov/) to improve the representation of the dynamical and thermodynamical environment of the atmosphere. Investigate which kind of conventional observations (METARS, buoys, radiosondes, aircrafts, …) has a greater impact in the analyses and pay special attention on the horizontal and vertical covariance structures.

2) Study the impact of assimilating quality controlled reflectivity and radial velocities from Doppler radars in combination with conventional in-situ observations. Determine the potential effect of assimilating high spatial and temporal resolution radar observations.

3) Assimilation of raw Rapid-Scan Atmospheric Motion Vectors assimilation every 20 minutes through the entire atmosphere.

4) Sensitivity numerical experiments evaluating the impact of several 'superobbing' configurations. The first attempt will be to assess the impact of 'superobbing' using 90 km x 90 km x 25 hPa prisms. The results of this task will provide information of which is the best configuration of 'superobbing' for each type of satellite observation to transfer towards an operational system.

5) Quantitative assessment of the quality of the data assimilation system through root-mean square innovations, total spread, bias and consistency ratio. It will be evaluated as function of the total number of available observations and assimilated observations. These error statistics will provide information of the potential of the satellite data assimilation. Only when this step reveals promising results, short-range forecasts from the EnKF analyses may be performed.

6) Evaluate the potential of the observational error assignment. It is planned to test different algorithms that compute the observational error in function of the height. Additional quality control algorithms to discard no-physical values of the observations will be addressed during this task.

7) Sensitivity numerical experiments assessing the potential of the 'thinning' method using different criteria. Comparison against results obtained from 'superobbing' in TASK 1.

8) To determine the optimal horizontal and vertical covariance localization influence radius for this kind of satellite observations. Test different values of horizontal and vertical localizations.

9) Evaluate different values of the prior inflation factor for each assimilation cycle. This could improve the consistency ratio of the system.

<u>b2) Generation of bred-vector based hindcast of high-resolution forecasts:</u>

Specific tasks to be executed are the following:

1) Generation a chain of a number of arithmetic bred vectors for a test period covering 4 months (September to December 2014), for a domain over the western Mediterranean (see meteo.uib.cat/wrf/) and different rescaling using the ECMWF analysis dataset as a global reference.

2) Comparison of the reference experiment of task 1 against a bred generation experiment based on logarithmic scaling.

3) Generation of 72h forecasts every 12 hours for each of the two breeding techniques.

4) Application of the exponentially rescaling technique will be exploited to multiply ensemble size at no bred-generation cost.

5) Conduction of experiments focused on limits of the technique of task 4, regarding ensemble size and genuine bred vector diversity.

c) EPS generation: Sampling model error:

Generation of stochastic parameterization based climatologies of high resolution forecasts through the tasks listed below:

1) Generation of an EDA-EPS using stochastic physics using the same 2014 period considered in the previous activity b2) to generate runs of Stochastically Perturbed Physics Tendencies (SPPT) and Stochastically Perturbed Parameter (SPP) using boundary layer and surface model parameter perturbations.

2) Analysis of the optimal application of SPPT and SPP regarding spatial and temporal correlations.

3) Combination of the method proposed in activity b2) with stochastic parametrization techniques.

d) Exploring next-gen forecast sensitivity information:

Current methods for calculating sensitivity fields are based on the linearity hypothesis between predictors and predictors. For example, the linear attached model is the transposed operator of the corresponding linear tangent of the original nonlinear model. Similarly, the method of statistical sensitivities (Torn and Hakim, 2008) is based on linear covariances between predictors and predictors. The sensitivity field generated by the Ensemble Transform Kalman Filter method is also based on a linearity hypothesis. None of them takes into account the non-linear evolution of disturbances, nor the dependence of the effect of a local disturbance with the adjacent field (besides the linear covariances). However, sensitivity information is inherent and fundamentally non-linear.

We plan on performing the following computational tasks:

1) Experiments with a model of intermediate complexity: Adoption or programming of a simplified atmospheric model (baroclinic or quasi-geostrophic) for the realization of ideal experiments of perfect model. Programming its adjoint tangent linear model. Use of current methods for calculating sensitivities on this model, including the adjoint model, statistical sensitivities and ETKF methods. Performing idealized experiments to generate sensitivity information through atmospheric patterns using this intermediate complexity model. Development of supralineal sensitivity methods based on higher orders of approximation of the evolution of disturbances and by proximity phase-space cloud methods. Idealized experiments to generate sensitivity information by means of these methods.
2) Calculation and verification of sensitivities in high resolution prediction systems: Calculation of a catalog of sensitivity fields and information for each of the methods previously developed and a list of high impact cases considered. Design and implementation of Observation System Simulation Experiments (OSSE) for the comparative verification of these fields and sensitivity information. Verification of predictions based on society-sensitive predictions such as hourly precipitation, instantaneous wind intensity or intense convective activity.

References:

Aksoy, A., Dowell, D. C., & Snyder, C. (2009). A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Monthly Weather Review*, *137*(6), 1805-1824.

Aksoy, A., Dowell, D. C., & Snyder, C. (2010). A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part II: Short-range ensemble forecasts. *Monthly Weather Review*, *138*(4), 1273-1292.

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47.

Berner, J., Fossell, K. R., Ha, S. Y., Hacker, J. P., & Snyder, C. (2015). Increasing the skill of probabilistic forecasts: Understanding performance improvements from model-error representations. *Monthly Weather Review*, *143*(4), 1295-1320.

Ehrendorfer, M. (1994). The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I: Theory. *Monthly Weather Review*, *122*(4), 703-713.

Jankov, I., Berner, J., Beck, J., Jiang, H., Olson, J. B., Grell, G., ... & Brown, J. M. (2017). A Performance Comparison between Multiphysics and Stochastic Approaches within a North American RAP Ensemble. *Monthly Weather Review*, *145*(4), 1161-1179.

Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515-3539.

Penland, C. (2003). A Stochastic Approach to Nonlinear Dynamics: A Review (extended version of the article-" Noise Out of Chaos and Why it Won't Go Away"). *Bulletin of the American Meteorological Society*, *84*(7), 925-925.

Primo, C., Rodríguez, M. A., & Gutiérrez, J. M. (2008). Logarithmic bred vectors. A new ensemble method with adjustable spread and calibration time. *Journal of Geophysical Research: Atmospheres*, *113*(D5).

Toth, Z., & Kalnay, E. (1993). Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the american meteorological society*, *74*(12), 2317-2330.