REQUEST FOR A SPECIAL PROJECT 2020–2022

MEMBER STATE: Sweden

Principal Investigator[1]: Jelena Bojarova

Affiliation: SMHI

Address: Folkborgsvägen 17

60176 Norrköping

Sweden

Other researchers: M.Lindskog, N. Gustafsson (SMHI), J. Barkmejier (KNMI), P.A.Escriba (AEMET), R. Azad (MET Norway)

Project Title: Development of Consistent HarmonVar-EPS system

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.)	2020	
Would you accept support for 1 year only, if necessary?	YES ☒	NO ☐

Computer resources required for 2020-2022: (To make changes to an existing project please submit an amended version of the original form.)		**2020**	**2021**	**2022**
High Performance Computing Facility	(SBU)	18M	18M	18M
Accumulated data storage (total archive volume)[2]	(GB)	60000	60000	60000

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: Jelena Bojarova

Project Title: Consistent HarmonVar-EPS system

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 Advisory Committee. 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 might receive a detailed review by members of the Scientific Advisory Committee.

The application is about the development of convective scale HarmonVar-EPS data assimilation system.

Introduction.

The HIRLAM (High Resolution Limited Area Model) international research programme is being carried out by the HIRLAM-C consortium (2016-2020) and is focused on further developments of the convective scale HARMONIE (HIRLAM ALADIN Research on Mesoscale Operational NWP In Europe) forecasting system (Bengtsson et al 2017). The HARMONIE system includes data assimilation, the forecasts model, the surface data assimilation and modelling and the high-resolution ensemble prediction system. At present the HARMONIE forecasting system is using a 3D-Variational data assimilation scheme for operational applications (Fischer et al 2005). The assumptions behind the 3D-Var, the limitations of the scheme and their consequences on the performance of HARMONIE forecasting system at meso-scales are discussed in Bojarova&Gustafsson, 2019. The development of the ensemble variational system, that would be able to consistently integrate together a variational data assimilation scheme and an ensemble prediction system, is one the strategic goals of the HIRLAM-C programme and one of the main research directions in the LAM meteorological community (Gustafsson et al, 2017). While most of individual components of such a system are already in place, including several schemes for generation of the initial conditions perturbations, the EnVar scheme will allow the use of flow-dependent ensemble based background error statistics and a computationally efficient multi-incremental 4D-VAR scheme with multiple outer loops will allow to handle weak non-linearities, the integration of these components together into a scheme powerful from both deterministic and probabilistic perspectives will require further developments and an extensive testing. This special project will concentrate on the inter-comparison of the available schemes for generation of initial condition perturbations both from the probabilistic point of view to predict severe weather events and on their ability to adequately sample the forecast error growth on meso-scales for a very short range ensemble. The further improvements of the variational schemes to allow them to utilize the full potential of the ensemble environment will be done in addition. The work will be carried out in a close coordination with the HIRLAM-C and the MetCoOp (Müller et al 2017) programmes. While the focus of HIRLAM-C lies on the algorithmic developments (advance 3DEnVar first to the level of 4D-VAR Hybrid and later on with 4DEnVar capabilities) and the focus of the MetCoOp lies on the computational efficiency of the algorithms (multi-incremental multiple resolution environment), the focus of this special project will be devoted to the validation and the final tuning of the scheme on its optimal performance, including such aspects as covariance localisation and space-scale dependent decomposition of the ensemble.

A) Methods for initial condition perturbations

The overview of the HarmonEPS system, its operational status and directions of further developments can be found in Frogner et al 2019. Within this special project we will concentrate on inter-comparison of 4 different schemes for generation of initial conditions perturbations (EDA, BRAND, LETKF and FORCING) . While the performance of the medium range forecast ensemble (18 h-36 h) is a focus of many on-going studies, the behaviour of the ensemble at a very short range (2h-12h) is less well understood, in particular at short convective scales. In addition to the generation of perturbations, there are indications that the behaviour of the forecast ensemble at the very short forecast ranges is strongly contaminated by the ensemble generation approach. The generation of perturbations is always associated with imbalances that can provoke spin-up processes, which are usually manifested as excess of the precipitation at early forecast ranges. The aim of our studies is to better understand the influence of the ensemble generation methods on the properties of the short range Ensemble Prediction Systems. The emphases will be done on the probabilistic forecasting of severe convective precipitation events. The initial condition perturbations that will be tested are listed below.

A1.) **EDA :**

The EDA system is a default scheme for initial condition perturbations in MEPS (MetCoOp EPS). The variability of EDA (Ensemble of Data Assimilation Runs) system originates from the variability imposed in observation space. The observations in the HarmonEPS EDA system are perturbed in a similar way as what is described in Isaksen at al, 2007. Conventional observations (SYNOP, SHIP, AIREP, Buoys, TEMP, Profiles) and radiances (AMSU-A, AMSU-B/MHS, IASI) for the Upper Air and T2m and RH2m for the surface are perturbed independently in different locations according to a Gaussian PDF with standard deviations of the assumed observation errors. Different realisations of such artificial perturbations are assimilated into different ensemble members. These members serve as initial conditions for EDA system. EDA system perturbations are computed against the ensemble mean or against the ensemble control run (assimilates original unperturbed observations) dependent on the configuration. There are indications that properties of the observing network, in particular its density and quality, may contaminate the EDA system at very early forecast ranges (Bojarova&Gustafsson, 2019).

A2.) **BRAND:**

BRAND perturbations are based on a randomization of the climatological background error covariance. A similar approach for perturbing initial conditions is applied in Raynaud &Bouttier, 2016. BRAND perturbations are generated as standard Gaussian random numbers in the entire control vector space and are transformed to the model space through the square-root of the background error covariance. In EPS mode the “i”-th BRAND perturbation is added to either the “i”-th ensemble member first-guess or the “i”-th ensemble member analysis dependent on configuration. Then a non-linear model is applied to propagate the BRAND system forward in time. The spread of the BRAND ensemble in EPS mode is controlled by assimilating the same observations to all ensemble members. There is a possibility to control how strongly the ensemble members are drawn to the observations. Although BRAND perturbations are drawn from the climatological background error covariance, the BRAND ensemble in the EPS mode reflects flow-dependency and has larger spread in the areas where the model evolution is sensitive to the initial conditions and smaller spread in the areas with the dense observation coverage (see Frogner et al, 2019, for more details).

A3.) **FORCING**

In stead of perturbing the initial conditions there is also the possibility to perturb the model tendency in a controlled manner. The computation of this type of tendency perturbations constrains the norm of the tendency in terms of total energy metric and looks for tendency perturbations that have the largest response in some prescribed metric after some lead time. During this lead time the same tendency perturbation is applied at every time step for some prescribed time after which the perturbation is switched off. These perturbations can be regarded as a generalization of the regular initial condition singular vectors. They have been used in various studies, such as in the context of climate runs in order to efficiently perturb non-linear model runs. More detail can be found in (Barkmeijer et al, 2003; Schrier et al , 2018)

A4.) **LETKF:**

The LETKF (Local Ensemble Transform Kalman Filter) scheme is an ensemble-based Upper Air data assimilation scheme implemented in the HARMONIE Forecasting system. In contrast to the variational data assimilation schemes, which are carried out in the spectral space, the LETKF is a grid-point scheme which performs an analysis locally at each point (Hunt et al, 2004). Because of its local nature, the LETKF scheme is highly scalable and numerically very efficient. The technical implementation of the LETKF in the HARMONIE forecasting system follows closely the one at the ECMWF. Details on the implementation and the performance of the algorithm can be found in Hamrud et al 2015 and Bonavita et al 2015. The LETKF scheme is powerful in sampling flow-dependent structures and identifying dynamically active directions. At the same time the local nature of the algorithm may lead to the excitation of the inertial-gravity waves in the short range forecasts. In the HARMONIE forecasting system the LETKF scheme is integrated into the Hybrid Gain environment where the LETKF perturbations (the ensemble member minus the ensemble mean) might be re-centred around the control analysis, which employs variational (spectral) scheme to perform the analysis.

All initial conditions perturbations will be tested in the Hybrid EnVar environment where they will be used to represent the flow-dependent error-of-the-day uncertainty. The Hybrid EnVar system allows to combine the full-rank climatological and ensemble-based flow-dependent background error statistics. The quality of the control forecast in the HybridEnVar system will serve as a measure of a quality of the initial conditions perturbation schemes in their ability to sample adequately the forecast error at the early forecast ranges. The “best-choice” initial perturbation scheme for convective-scale developments which will score best from probabilistic/deterministic point of view will be identified.

B.) Development of the HarmonVar-EPS framework

B1.) **HARMONIE Hybrid 3D-EnVar**

The HARMONIE Hybrid 3D-EnVar scheme is already implemented since CY40h1.1.1 and is phased into CY43 (develop branch). The HARMONIE Hybrid 3D-EnVar is based on the so-called alpha control variable approach (Lorenc 2003) and it closely follows the implementation described in Gustafsson et al 2014. In addition, the HARMONIE Hybrid 3D-EnVar system is extended with a possibility to treat multiple scales (as in Buehner&Shlyaeva 2015). The performance of the HybridEnVar scheme depends strongly on the quality of ensemble and its ability to sample enough forecast error growth directions. A small size ensemble can be a limitation. The covariance localisation is a known approach to remedy the severe rank-deficiency to a certain extent. The space-scale dependent localisation may help to make the scheme more flexible with a less severe localisation. The optimal setup of such a system has to be found through testing.

B2.) **HARMONIE 4D-VAR**

The multi-incremental multiple resolution HARMONIE 4D-VAR with several outer loops is implemented since CY40h. Recently it was extended with the large extension zone possibility during the minimization step. This option allows to run HARMONIE 4D-VAR on smaller domains without a detrimental effect of the horizontal wrap-around of the increment. Advantages of multi-incremental formulation with several outer loops for treatment of non-linearities is discussed in Gustafsson et al, 2012. The HARMONIE 4D-VAR scheme will be used as a reference system for the experiments.

B3.) **HARMONIE Hybrid 4D-VAR**

The HARMONIE 4D-VAR scheme is to be extended with a Hybrid environment, including possibility for multi-incremental minimization with multiple resolutions (HIRLAM-C Programme), as in Gustafsson et al 2014. The "best-choice" initial perturbation scheme will be tried in the HARMONIE Hybrid 4D-VAR framework. The optimal configuration of such a scheme including the size of ensemble, strategy for ensemble lagging, length of the 4D-VAR window, the length scales of the localisation and the ensemble forecast range has to be established from the experimentation. The extension of the HARMONIE 4D-VAR scheme with the Hybrid Environment seems to be a natural step when the convective scale ensemble is a part of the operational production. Even if the 4D-VAR scheme is known as computationally demanding, the multi-incremental minimization with multiple resolutions helps to reduce the computational costs significantly and make the scheme an affordable candidate for operational applications.

B4). **HARMONIE 4D-En-Var**

The HARMONIE 4D-En-Var scheme that utilizes the non-linear ensemble evolution instead of tangent-linear and adjoint model to describe the propagation of the increment will be developed on the basis of the HARMONIE Hybrid 4D-VAR. The main emphases will be on the optimal settings for the covariance localisation scheme including formulation of the space-time correlations for different localisation lengthscales. The 4D-En-Var scheme has been shown to be very successful for synoptic scale applications (Gustafsson&Bojarova, 2014) both due to reduced computational costs and due to the strength of consistent treatment of complex dynamical structures. At the same time the performance of the 4D-En-Var scheme is strongly dependent on the quality of ensemble.

Plan of the implementation and configuration of the experiments.

2020:

Setup optimal configuration of the Hybrid 3D-EnVar framework (spatial scale dependent decomposition of the ensemble, vertical localisation, form of the auto-correlation function for the covariance localisation). Test of the initial condition perturbations (EDA, BRAND, LETKF, FORCING) in the optimal HARMONIE Hybrid 3D-EnVar framework. The "best-choice" ensemble generation scheme to be selected.

In parallel. The development of the HARMONIE Hybrid 4D-VAR scheme

2021:

Test of the optimal HARMONIE Hybrid 4D-VAR configuration: ensemble size, lagging strategy, ensemble forecast range using the "best-choice" ensemble. HARMONIE 4D-VAR reference run for the performance of the control member. HARMONIE 3D-EnVar reference run with the optimal configuration of the ensemble.

In parallel, development of the HARMONIE 4D-En-Var scheme. The implementation of the HARMONIE 4D-En-Var requires a certain level of code refactoring. It is unclear at the moment if this level is available within the LAM release of CY43.

2022:

Test of the optimal configuration of the HARMONIE 4D-En-Var scheme (formulation of the space-time localisation for different spatial scales, size of ensemble, lagging strategy). Comparison of the meteorological performance versus costs between HARMONIE Hybrid schemes and the decision on HarmonVar-EPS system.

Justification of the computer resources needed.

DA-EPS experiments are usually expensive.

2020: We will run test experiments over a smaller domain that it used operationally (the so called DKCOEXP domain : 648x648 horizontal grid points, 65 vertical levels, 2.5km horizontal resolution). In order to save computer resources we will release 30 hours long forecasts twice a day. This makes approximately 160,000 SBUs for a 11 member (10 members + control run) ensemble for a day of experiment. In order to get statistically reliable information we have to run an experiment for at least a 2 week period, what makes 2,2MSBUs per experiment. During the first year, we consider to run at least 8 experiments (4 experiments to determine the optimal configuration of the HARMONIE Hybrid 3D-EnVar system + 4 experiments for different schemes for initial conditions perturbations). This makes 18 MSBUs in total.

We ask for the same amount of the resources for next 2 years even if the computational cost of the 4D-VAR system is higher.

2021: The work on HARMONIE 4D-VAR Hybrid and HARMONIE 4D-En-Var is in the less mature stage. Additional national resources and the resources of the HIRLAM-C programme will be used for the experimentations. Because 4D-Var experiments are more expensive we estimate 3MSBUs per experiment. 18MSBUs will allow us 6 experiment.

2022: Based on the outcomes of the test experiments, the optimal configuration of the HarmonVAR-EPS will be defined. The test of the HarmonVAR-EPS will be done on a larger domain (so called METCOOP B (or a corresponding one, used by other operational consortia) : 750x960 grid-points, 65 model levels, 2.5 km resolution). This results in 275TSBUs for a day of test, or ≈ 3,85 MSBUs for a two weeks experiment (using 10 ensemble members). 18MSBUs will allow us to run 4 experiments with increased number of ensemble members (2 alternative experiments over summer/winter period). The reference runs will be done utilizing additional national resources.

References

Barkmeijer, J., T. Iversen and T.N. Palmer, 2003: Forcing singular vectors and other sensitive model perturbations. Quarterly Journal of Royal Meteorological Society, 129, 2401-2423

Bengtsson, L., and Co-authors, 2017: The HARMONIE–AROME Model Configuration in the ALADIN–HIRLAM NWP System. Mon. Wea. Rev., 145, 1919–1935, https://doi.org/10.1175/MWR-D-16-0417.1

Bojarova J. and N. Gustafsson, 2019: Relevance of the climatological background error statistics for meso-scale data assimilation. Tellus A; Dynamical meteorology and Oceanography, 71:1, 1-22, DOI:10.1080/16000870.2019.1615168

Bonavita, M., M. Hamrud and L. Isaksen, 2015: EnKF and Hybrid Gain Ensemble Data Assimilation. Part II: EnKF and Hybrid Gain Results. Monthly Weather Review, 143, 4865-4882, DOI: 10.1175/MWR-D-15-0071.1

Buehner, M., and A. Shlyaeva, 2015: Scale-dependent background error covariance localisation. Tellus A., 67:1, DOI:10.3402/tellusa.v67.28027

Fischer, C., Montmerle, T., Berre, L., Auger, L. and Ştefănescu, S. E. 2005: An overview of the variational assimilationin the Aladin/France NWP system. *Q J. R Meteorol. Soc.* **131**, 3477–3492. doi:10.1256/qj.05.115

Frogner, I.-L., and Co-authors, 2019: HarmonEPS – the HARMONIE ensemble prediction system. Weather and Forecasting (under review)

Gustafsson, N., and J.Bojarova, 2014: Four-dimentional ensemble variational (4D-En-Var) data assimilation for HIgh Resolution Limited area Model. Non-linear Processes in Geophysics, 21, DOI:10.5194/npg-21-745-2014.

Gustafsson, N., J. Bojarova and O.Vignes, 2014: A hybrid variational ensemble data assimilation scheme for the HIgh Resolution Limited Area Model (HIRLAM). Non-linear Precesses in Geophysics, 21, DOI:10.5194/npg-21-303-2014.

Gustafsson, N., and Co-authors, 2017: Survey of data assimilation methods for convective scale numerical weather prediction at operational centres. Quarterly Journal of Royal Meteorological Society, 144, DOI:10.1002/qj.3179

Gustafsson, N., Huang, X.-Y., Yang, X., Mogensen, K., Lindskog, M., Vignes, O. and Thorsteinsson, S. 2012: Four-dimensional variational data assimilation scheme for a limited area model. Tellus A, 14985, DOI:10.3402/tellusa.v64i0.14985

Hamrud, M., M. Bonavita and L.Isaksen, 2015: EnKF and Hybrid Gain Ensemble Data Assimilation . Part I: EnKF implementation. Monthly Weather Review, 143, 4847-4864. DOI:10.1175/MWR-D-14-00333.1

Hunt , B. R., E.J. Kostelich and I. Szunyogh, 2007: Efficient data assimilation of spatiotemporal chaos: A local ensemble transform Kalman Filter. Physica D., 230,112-126. DOI:10.1016/j.physd.2006.11.008

Isaksen L, Fisher M, Berner J. 2007: Use of analysis ensembles in estimating flow-dependent background error variance. Pp. 65 – 86 in Proceedings of the ECMWF workshop on flow-dependent aspects of data assimilation, 11 – 13 June 2007. Available from ECMWF, Shinfield Park, Reading, RG2 9AX, UK (see also http://www.ecmwf.int/publications/library/do/references/list/14092007).

Lorenc, A., 2003: The potential of the ensemble Kalman Filter for NWP – a comparison with 4D-Var. Quarterly Journal of the Royal Meteorological Society, 129, 3183-3203, DOI: 10.1256/qj.02.132

Müller, M. and Co-authors, 2017: AROME-METCOOP-A Nordic Convective scale Operational Weather Prediction Model. Weather and Forecasting, 32, DOI: 10.1175/WAF-D-16-0099.1

Schrier, G. van den, L.M. Rasmijn, J. Barkmeijer, A. Sterl and W. Hazeleger. 2018: The 2010 Pakistan floods in a future climate Climatic Change, 2018, 148, 205-218, doi:10.1007/s10584-018-2173-7.