SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2021 (the second year of the project)
Project Title:	The development of consistent HybridVar-EPS system
Computer Project Account:	spseboja
Principal Investigator(s):	Jelena Bojarova
Affiliation:	SMHI
Name of ECMWF scientist(s) collaborating to the project (if applicable)	
Start date of the project:	November 2019
Expected end date:	October 2022

Computer resources allocated/used for the current year and the previous one

(if applicable)

Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	18MCTUs	18MCPUs	18MCPUs	0
Data storage capacity	(Gbytes)	60000GB		60000GB	

Summary of project objectives (10 lines max)

The objective of the project is to develop, validate ad perform final tuning of the HarmonieVAR-EPS system on its optimal performance from both probabilistic and deterministic points of view. This include choice of an optimal approach for generation of initial conditions perturbations (BRAND/EDA/LETKF/FORCING); advancing the scheme to the Hybrid 4DVar/Hybrid 4DEnVar levels; customizing covariance localisation by further development of vertical, space-scale and time dimensions; validation of the performance of the scheme through numerical efficiency of the scheme, probabilistic and deterministic verification scores.

Summary of problems encountered (10 lines max)

The main problem encountered during the project was a development of instabilities in the HARMONIE EnVar system that were leasing to the numerical explosion of the Forecast runs. These numerical instabilities were track down to a particular geographic location in the vicinity of Sogn Fjörd in Norwegian Mountains associated with extremely steep orography. Small size of ensemble (10 members) chosen to save resources turned out to be unrealistically small and resulted in a low quality performance of HybridEnVar scheme. The larger size ensemble has to be used even if this leads to higher computational costs. Certain technical issues has to be solved in order to be able to use the LETKF and FORCING perturbations schemes in CY43 of common codes.

Summary of plans for the continuation of the project (10 lines max)

The restricted HybridEnVar system with FORCING and LETKF initial perturbations will be conducted and the performance of the system will be evaluated using deterministic and probabilistic scores. The results will be compared with restricted HybridEnVar BReND and HybridEnVar EDA systems taking into account computational efficiency of the systems. In parallel the Hybrid 4DVAR and the Hybrid 4DEnVar schemes will be further developed including advanced covariance localisation options. The “best-choice” initial perturbation scheme with at least 20 ensemble members will be used for the final tuning of the HarmonieVar-EPS system on its optimal performance.

List of publications/reports from the project with complete references

Summary of results

If submitted **during the first project year**, please summarise the results achieved during the period from the project start to June of the current year. A few paragraphs might be sufficient. If submitted **during the second project year**, this summary should be more detailed and cover the period from the project start. The length, at most 8 pages, should reflect the complexity of the project. Alternatively, it could be replaced by a short summary plus an existing scientific report on the project attached to this document. If submitted **during the third project year**, please summarise the results achieved during the period from July of the previous year to June of the current year. A few paragraphs might be sufficient.

The main objectives of this special project are
1) to perform an 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
2) further improve variational environment to obtain a consistent ensemble variational framework

In the HARMONIE system several methods are available for generation of the initial condition perturbations .The EDA system is a default scheme for sampling of initial conditions in MEPS (MetCoOp EPS). EDA scheme is based on the variability imposed in observation space. Over LAM domain the observing networks vary during a day. This may impose undesirable variations in the ensemble generation scheme. Another problem is that the variability of the EDA system depends on the density of the observing network. In the current experiments we have been using a conventional observing network. It was difficult to sample enough variability for moisture (specific humidity) using EDA perturbations in our experiments. The lack of variability is a serious drawback for the sapproach, because in the hybrid ensemble variational framework variability of the ensemble determines amplitude of the analysis increment. BRAND, LETKF and FORCING perturbations are other techniques available in the HARMONIE system for the generation of initial conditions. For the comprehensive summary of different ensemble perturbations schemes in HARMONIE see Frogner et al, 2019.
In our experiments so far we have concentrated on the inter-comparison of EDA and BRAND schemes. There are still some technical issues remaining with the implementation of the LETKF and FORCING perturbations schemes, in CY43 in HIRLAM repository, the version of the common codes that was selected for the experiments of the special project.
BRAND perturbation scheme is based on the randomization of the B-matrix covariance. The variability is imposed in the control vector space. It does not depend on the properties of the observing network and samplse all scales consistently following structures determined in the background error covariance. The same unperturbed observations are assimilated by all ensemble members in order to prevent a divergence of ensemble. The main drawback of the BRAND scheme is that adding the climatological perturbation to an analysed field introduces a chock that pushes the obtained ensemble member away from observations. The skills of such ensemble members measured through the distance to observations become heavily degraded in comparison to those that the unperturbed member (the control) manage to achieve. The spread of the very short range ensemble overestimate uncertainty. Ensemble Variational scheme draws forecasts too close to observations degrading the ensemble quality. This can be noticed by a rapid increase of the distance to observation for the perturbed ensembles members already after three hours.

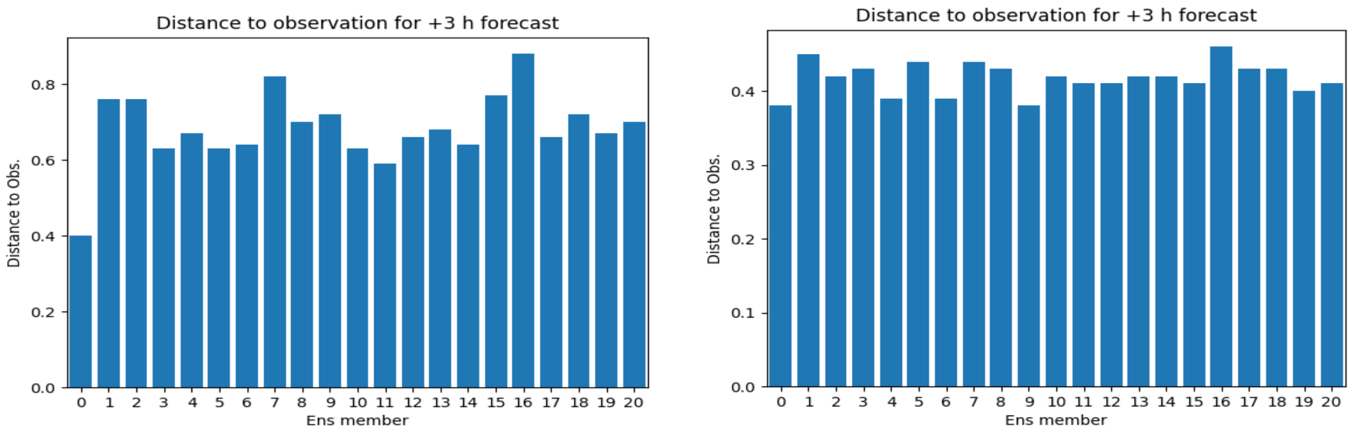

Figure 1. Relative distance to observations (per observed quantity) of the +3h forecasts for BRAND (left plot) and BREND(right plot) ensembles. Notice difference in scales for right and left plots.

A new flavour of the BRAND scheme, a so-called BREND perturbations, BRAND in the EPS mode, were developed in this special project. BREND perturbed ensemble members retain a similar forecast quality measured as a squared distance to observations to one of the unperturbed control. Figure 1 present a relative squared distance to observations (per observed quantity) for the original BRAND scheme (left plot) and newly developed BREND scheme (right plot) Some members may even marginally outperform the control run. The reason is the even assimilating the same observations the ensemble does not collapse because this is a specification of the observation error variance that determines how close an ensemble member allows to come to observations. During the two weeks of experiment conducted in this special project a very small amplitude perturbation was enough after each analysis to manage a vital spread of the ensemble. It is interesting to notice that the variability among the ensemble members is the largest in the areas with the coarse observation coverage. The amplitude of the added perturbation is defined as 1/sqrt(ensemble_size). For ensemble size of 20 members the amplitude of the perturbation is around a quarter of a climatological standard deviation. An incremental perturbation scheme is considered to even further improve possibilities of the ensemble to represent the error growth in an adequate way for very short forecast ranges. The implementation of the scheme need to be done, the algorithm is based on the incremental analysis update which already exist in the codes.

Figure 2 shows a temperature field at approximately 700 hPa (model level 47) produced by HybridEnVar scheme using BREND ensemble for generation of initial conditions perturbations.
Top left plot shows the +3h control forecast (unperturbed ensemble member) , top right plot shows analysis increment, and three bottom plots show the +3h forecasts of first three ensemble members . Ensemble is driven by IFS model EPS_EC ensemble on the boundaries. The size of ensemble is 20 members plus a control run. Valid time 2019 06 26 12 UTC

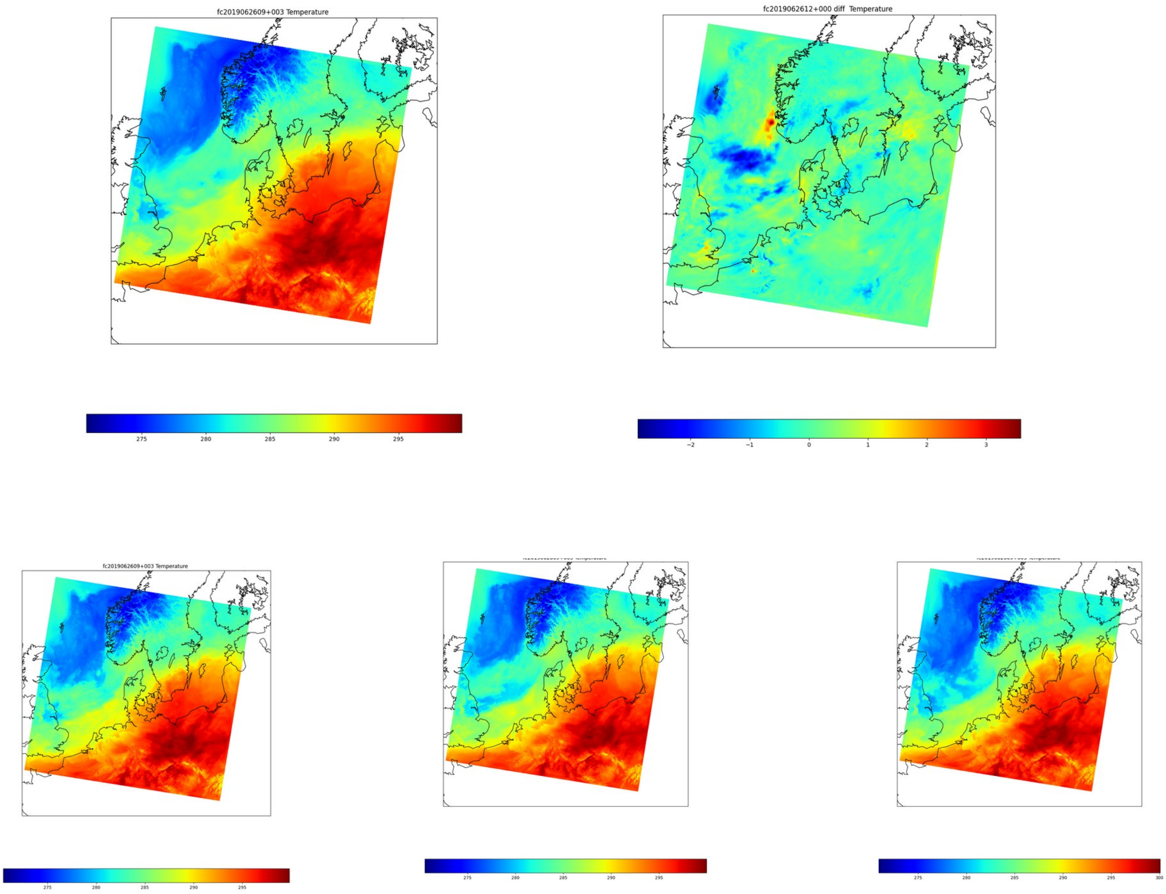

Figure 2 . Temperature fields at 700 hPa (model level 47) from HybridEnVAR + BREND ensemble The +3h control forecast (unperturbed ensemble member, top left plot), analysis increment (top right plot), the +3h forecasts for first three ensemble members (three bottom plots). Ensemble is driven by IFS model EPS_EC ensemble on the boundaries. The size of ensemble 20 members + control. Valid time 2019 06 26 12 UTC

The quality of the ensemble variational data assimilation is heavily dependent on the ensemble size and the localisation length scales. It turned out that the initially considered minimal size of ensemble (10 members) is too small to demonstrate the potential of the HybridEnVar scheme. We have had to double ensemble size to obtain the forecast skills comparable to those of the default (3DVAR) scheme. Two different localisation length scales were tried using HybridEnVAR BREND configuration with 20 ensemble members. Use of shorter localisation scales (length scale 100 km; Gaussian auto-correlation function) destroys flow-dependent structures for large-scale variables, such like wind and mass field. Use of longer localisation scales (length scale 500 km, Gaussian auto-correlation function) retains flow-dependent structures in wind and mass field but provides too noisy humidity field. In the current version of the codes the same spatial local weights are used on all vertical levels for all variables. The space-scale dependent localisation is implemented in the code, but some technical work for ensemble decomposition still need do be done. The work is still in progress. It is expected that space-scale localisation will allow to efficiently apply different localisation lengths for different variables. BREND scheme generates variable dependent variability. Energy dominates small scales for humidity field and large scales for surface pressure field.

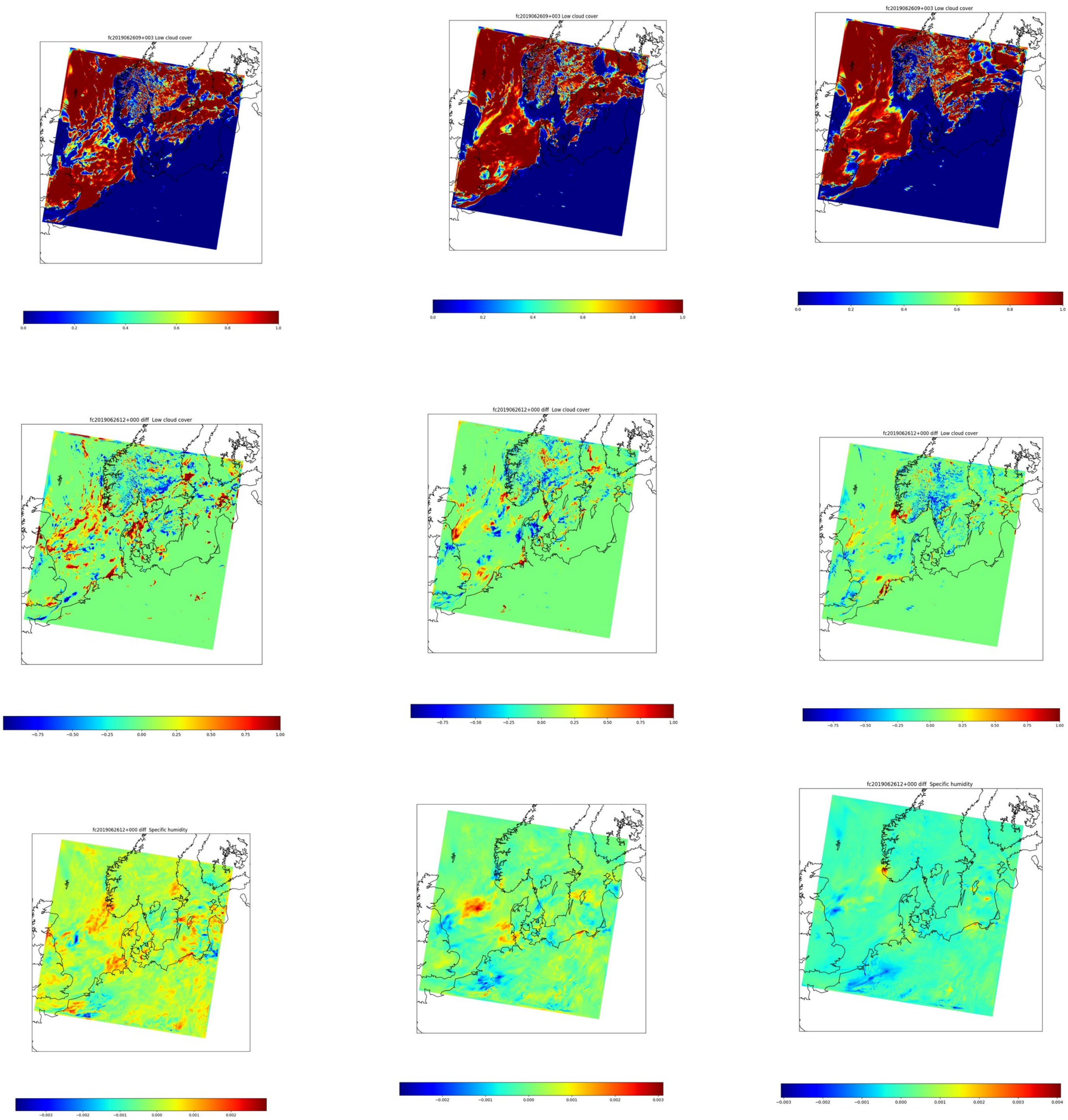

Figure 3. The +3h low cloud cover field forecast (top plots), the low cloud cover field increment (middle plots) and the corresponding specific humidity increment at model level 47 (approximately 700hPa) (bottom plots) for control member in HybridEnVar experiments using BREND ensemble with 500 km localisation scale (plots to the left), using BREND ensemble with 100 km localisation scale (plots in the middle) and using EDA ensemble with 100 km localisation scale (plots in the right). All experiments are conducted using 20 + 1 ensemble members. Valid time 2019 06 26 12 UTC

Figure 3 illustrates the dependency of the HybridEnVar scheme on the length of the localisation scale. Top plots show the +3h low cloud cover forecast , middle plots show the low cloud cover field increment. The bottom plots show the corresponding specific humidity increments at model level 47 (approximately 700hPa) . All results are shown for the control member in HybridEnVar experiments using BREND ensemble with 500 km localisation scale (plots to the left), using BREND ensemble with 100 km localisation scale (plots in the middle) and using EDA ensemble with 100 km localisation scale (plots in the right). The experiments are conducted using 20 ensemble members plus a control run. Validity time is 2019 06 26 12 UTC. Longer localisation scales results in a more noisy humidity increment . More energy kelp on shorter scales in a humidity field results in a higher spatial variability of low level cloud cover.

The standard verifcation scores (the bias and the standard deviation against SYNOP surface observations and TEMP profiles) show that HybridEnVAR system is able to improve skills of the wind forecast in the lower part of the troposphere and mass and humidity fields in the middle of the troposphere while degrading the quality of the winds field in the free atmosphere and mass/humidity fields near surface. We expect that the space-scale dependent localisation will help to improve performance of the HybridEnVAR Scheme.

Table 1. The summary of experiments conducted within “the development of consistent HarmonVAR-EPS system ”.

Experiment name	minimization	Ensemble	Size	Loc. Scale	Account
Hybrrid3DEnVAR_EDA_rest	restricted	EDA	10	300km	spseboja
jelena_EnVar_dev	full	BRAND	10	100km	sesmhi
Hybrid3DEnVAR_BRAND_EPS	restricted	BRAND	20	300km	sesmhi
Hybrid3DEnVAR_BRAND_rest	restricted	BRAND	10	100km	spseboja
Hybrid3DEnVAR_BRAND_rest_20	restricted	BRAND	20	100km	spseboja
Hybrid3DEnVAR_BREND_rest_20	restricted	BREND	20	100km	spseboja
Hybrid3DEnBAR_EDA_rest_20	restricted	EDA	20	100km	hlameps
Hybrid3DEnVAR_BREND_rest_20_500	restricted	BREND	20	500km	sesmhi

It is worth to mention that all HybridEnVar experiments with different flavours of initial condition perturbations were conducted using restricted minimization. The analysis increment is restricted to the longest 100 1D-waves (from 320 available). The information on shorter ways is kept unchanged extracting it from the first guess field. This is done as a remedy to a pathological behaviour of the HARMONIE forecasts in case of a very steep orography using a linear grid. A false convection was developed by a model in a particular geographical location in a vicinity of Sogn Fjord in Norway. Extremely heavy catabalic winds were generated that caused the explosion of the forecast run. The investigations have shown that the development of numerical instabilities happens in this area even in the 3DVAR system. However, the variational data assimilation with homogeneous and isotropic structure functions has produced analysis fields insensitive to orography properties and in such way have helped to smooth out instabilities. For the longer forecast lengths horizontal diffusion removes energy on the shortest scales and prevents instabilities to grow. The same happens if the shortest scales are taken from a “mature” forecast field.