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 2020……………………………………………….……

Project Title: Regional Reanalysis Spin Up ……………………….……

……………………………………………………….……

Computer Project Account: spseschi …….……………………………….……

Principal Investigator(s): Semjon Schimanke, Martin Ridal, Adam El Said …….……

……………………………………………………….……

Affiliation: SMHI and Météo-France ...…………………………….……

Name of ECMWF scientist(s) collaborating to the project
(if applicable)
……………………………………………………….……

……………………………………………………….……

Start date of the project: 01/01/2020 .…………………………………….……

Expected end date: 31/12/2021 .…………………………………….……

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	(SBU)	-	-	55,000,000	56,441,834
Data storage capacity	(Gbytes)	-	-	350,000	235,617

Summary of project objectives (10 lines max)

The original idea of the project was to investigate appropriate lengths of spin-up periods for reanalysis production. In the FP7 project UERRA, the project team applied spin-up periods of 4 months. For the production of CERRA (Copernicus European regional reanalysis), we decided to run the spin-up for one complete year. This comes of course with high computing costs.

Based on that, the first research question was to find the minimum spin-up length, which ensures that the quality of each stream is more or less indistinguishable from a continuous run. In addition, the impact of the timing of the onset on the spin-up, e.g. at the beginning of the year or in autumn, should be investigated.

Summary of problems encountered (10 lines max)

For the production of CERRA, a strategy for the spin-up periods was needed to be in place before this special project became available. A way to reduce the costs of the spin-ups was found as outlined in the result section but there is more to be done about it.

On the other hand, we faced problems in the way we produced the flow-depending B-matrixes, which needed to be solved immediately. Optimizing these procedures needed a lot of SBU which are not available in the C3S contract and we decided to use the resources of this special project. Encountered open questions are e.g. (a) forecast length (3h vs 6h), (b) daily, monthly, yearly and seasonal temporal variabilities, (c) resolution (and subsequent interpolation considerations), and (d) observations.

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

We cannot afford additional experiments in 2020 since the special project's SBU resources are consumed. However, a lot of experiments were performed and investigations of the data have only started. Analyses will deal with questions as length of moving average of daily variabilities and the update frequency of daily component (every day or every two days).

For 2021, the available SBU resources might be used to conduct a series of shorter spin-up experiments as outlined in the original research plan. Here, the idea is to run spin-up experiments initialized during different seasons of the year to identify the best time of the year to start the spin-up. Otherwise, investigations on the optimal B-matrix might continue. We would like to understand what proportions of "static" and "dynamical" parts (e.g. 80/20, or 50/50) give best results.

List of publications/reports from the project with complete references

El-Said, A. (2020): Estimation of uncertainties in the model system, deliverable D322_Lot1.1.1.7 of the Copernicus service C3S_322_Lot1, contact: info@copernicus-climate.eu

Summary of results

Spin-up strategy

SBU needs for the production of the spin-up periods are significant and the SBU resources available in the C3S contract to produce CERRA are limited. Therefore, we had to decide on a spin-up strategy before this special project became available. The production of one year CERRA data (HiRes) consumes approximately 12 MSBU. In addition, we produce an ensemble with 10 members with somewhat lower resolution (CERRA-EDA), which provides information required for uncertainty quantification as well as the continuously updated B matrix used in the HiRes run. The production of one year CERRA-EDA needs about 15 MSBU. Hence, in total, about 27 MSBU are needed to run one reanalysis year with the applied systems. The production of the regional reanalysis is split up into six streams and each stream needs to spin-up. Consequently, a total of about 162 MSBU (6*27 MSBU) would be needed for the spin-ups if the model systems would be used straight away. These resources were not available so we decided to use lighter model versions for the spin-up.
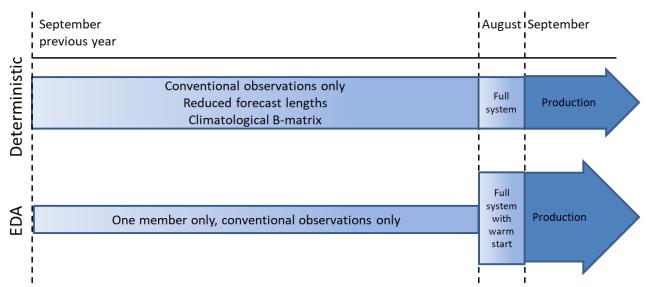

Figure 1: Schematic setup for the spin-up periods to reduce SBU costs.
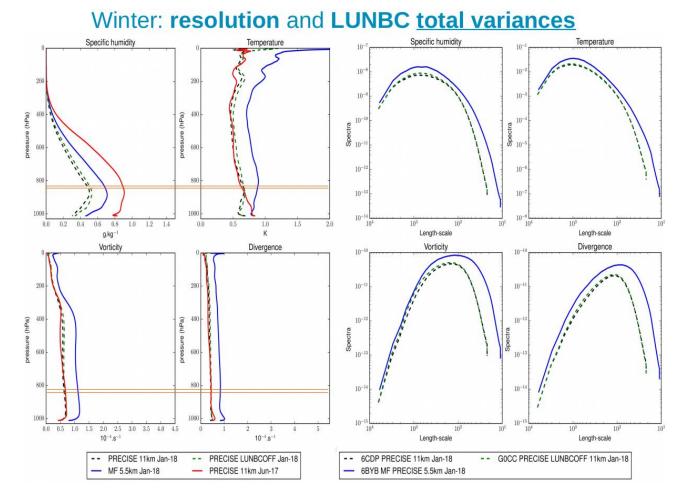
For first 11 months of the Hires spin-ups, we decided to use only conventional observations as the processing of satellite data consumes resources. In addition, we run shorter forecasts lengths as well as the archiving is turned off. Then, only the last month of the spin-up is run with the full system. The taken measures reduce the SBU needs for the HiRes runs to 7.3 MSBU compared to 12 MSBU with the full system.

For the EDA system, we run the first 11 months with one member only as well as we use only conventional observations. For the final month, we use the full system. That reduced the SBU costs for the spin-up year to 3 MSBU compared to 15 MSBU when running the full system for a year.

Hence, applying these rather simple measures the SBU costs for the spin-up period could be reduced from 27 MSBU to roughly 10 MSBU for each stream.
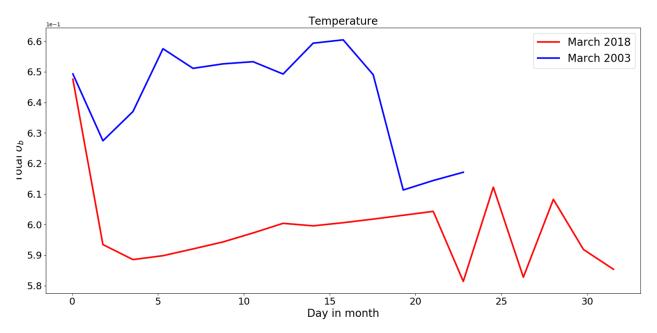
Investigations of B-matrix computation

The EDA system for this project was based on a preliminary EDA system developed by Pierre Brousseau and Adam El-Said at Météo-France using the OLIVE-ALADIN scripting system. Results developed over time during the assimilation process between the OLIVE-ALADIN (5.5km, 5 member EDA) and HARMONIE-ALADIN (11km, 10 member EDA) configurations, such as switching off large-scale or spectral boundary layer mixing protocols (LSMIX/LUNBC) in HARMONIE for example. We also showed results that the potential variations in the statistics of the B-matrix due to; daily and seasonal variabilities, observation densities (1990's obs config vs present-day), and forecast resolutions, were significant enough to be taken into account in the final method used to construct the B-matrix. We also showed results for investigating the effects of; forecast lengths for generation of differences for B-matrix statistics (3h vs 6h), inclusion of satellite observations, LUNBC=on/off, LSMIX, removing TEB (town wind-shear calculations), monthly and yearly variabilities, showing no cause for consideration, and developed the B-matrix method accordingly. Below is an example of the effects of switching off LUNBC in HARMONIE (labelled PRECISE below) and resolution (5.5km vs 11km):

We also made significant changes to the HPC resource allocation scripts which for 'FESTAT', which is the script that generates the B matrices after the interpolation ('EE927') and differencing ('FEMARS') procedures. This was absolutely necessary (and expensive) because 'FESTAT' was configured sub-optimally. As an illustrative example: previously the HARMONIE scripts responsible for queueing FESTAT resulted in the FESTAT sub-tasks occupying far too many nodes, which both costs a lot of SBU's and is extremely inefficient as it leaves a lot of HPC node memory unused. Adam El-Said liaised with Dominique Lucas and members of the HPC team extensively, to come up with organised scripting variables that require the least amount of input from users, while simultaneously utilising HPC resources as effectively as possible. This has also been successfully implemented outside CERRA, and has been of great benefit to HARMONIE users in general.

Subsequently, the B-matrix method we have developed now comprises two components, utilising an approximate 80/20 split, to account for seasonal and daily variabilities respectively. The 80% proportion of the split represents seasonal variability, and is generated by a linear pooling function, takes a varying number of differences based on what time of year it is. So for example, it would pool more summer forecast differences as the time of year approaches summer, and vice versa. The 20% weighting accounts for the daily variabilities from the last 2.5 day moving average.
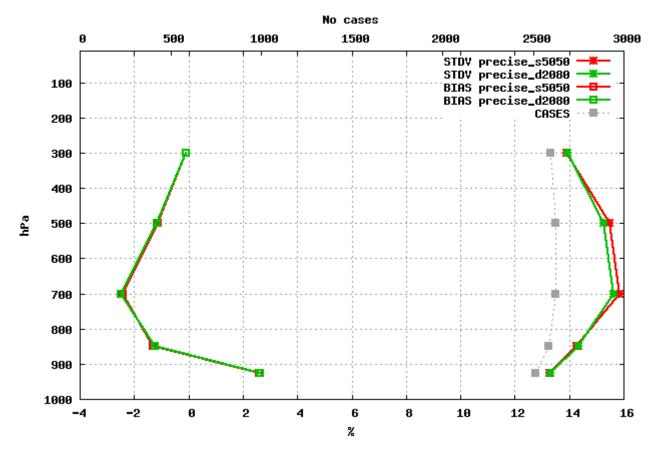
We have begun investigations into the effect of changing the proportion of each component of the B matrix (from 80/20 to 50/50 for example) and investigating the effect of differing weather regimes in selected years. Here we have an example of the differences in B-matrix statistics for the total sigma_b value for temperature (spatially and vertically summed) between the entirety of March-2003 and March-2018:

where March-2003 was dominated by a blocking regime and Mar-2018 was dominated by negative NAO.

During both periods we have also investigated how the different weightings of the B matrices affect the results. Not only the B-matrix itself but also the resulting forecasts, using the generated B-matrix when running experiments with the high resolution deterministic model version of the CERRA system. So far the latter has only been performed for two experiments for March 2018 where the first uses a "static" mix of summer and winter with 50% of each period for all the month. This is the "traditional" method and used in many operational models. In the second experiments we used 80% summer/winter mix that scales with the time of the year and 20% from a 2.5 day running average from EDA runs. This is updated with a 48 hour interval which means that we get a new B matrix every second day. Preliminary results show a very small but positive impact from the "dynamic" B matrix. One example of a vertical profile of relative humidity is shown in the next figure. The red curves represent an experiment using the "static" B while the green curves represent the "dynamic" B. What is shown is forecasts of 12 hour relative humidity, valid at midday, from a 25 day experiment. A small positive impact can be seen in the standard deviation (right most lines) while the bias is very similar between the two.

119 stations Selection: ALL
Relative humidity   Period: 20180301-20180325
Statistics at 12 UTC   Used {00} + 12

More experiments are planned to be performed testing different combinations when constructing the B such as more inclusion of the day to day variability from the EDA runs e.g. 50% instead of the current 20% or even 100%. The two different periods will also be compared to investigate how a B matrix computed for a different weather regime will influence the results of a deterministic run for a completely different weather situation.