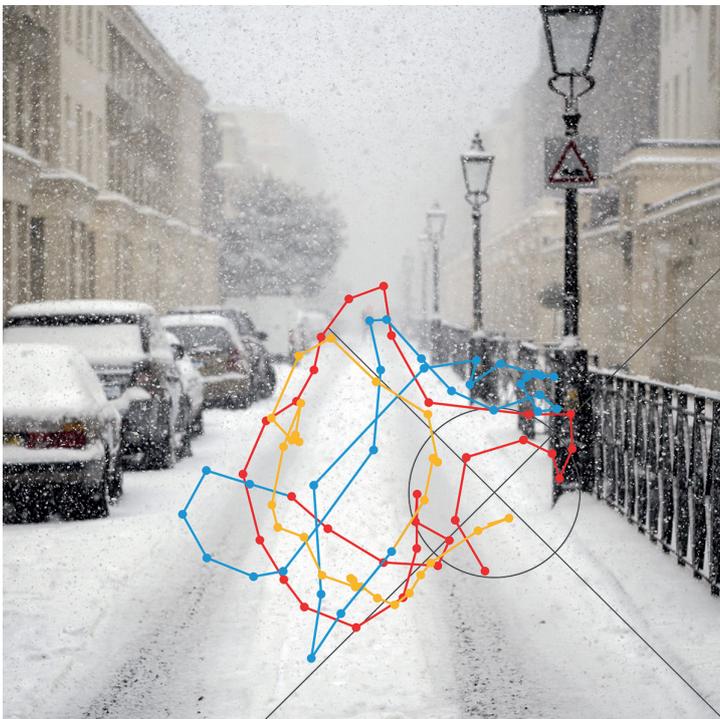


METEOROLOGY

A 50-member Ensemble of Data Assimilations



Cover image: aley.gunebakanli/fstock/Thinkstock

This article appeared in the *Meteorology* section of *ECMWF Newsletter No. 158 – Winter 2018/19*, pp. 27-29.

A 50-member Ensemble of Data Assimilations

Simon Lang, Elías Hólm, Massimo Bonavita, Yannick Trémolet (JCSDA, US)

Since 2010, ECMWF has run an Ensemble of Data Assimilations (EDA) to help determine the initial conditions for its ensemble forecasts and its higher-resolution deterministic forecast. The EDA is an ensemble of 4D-Var data assimilations that reflects uncertainties in observations; atmospheric boundary conditions, such as sea-surface temperature; and the model physics. The role of the EDA is twofold: it contributes to the high-resolution initial conditions (the high-resolution analysis) by providing flow-dependent estimates of the errors in the short-range forecasts (the background) used in the data assimilation system; and it helps to determine the initial conditions for ensemble forecasts by providing EDA-based perturbations to the high-resolution analysis. In 2013, the number of EDA members was increased from 10 to 25. We have now developed a new, optimised 50-member EDA configuration that has a comparable computational footprint to the current operational 25-member configuration. The increase in ensemble size improves both the high-resolution analysis and the EDA-based perturbations to the initial conditions for the 50-member ensemble forecast. The change is due to be implemented in the next upgrade of ECMWF's Integrated Forecasting System (IFS Cycle 46r1), scheduled for 2019.

Benefits of more EDA members

The EDA uses a Monte Carlo approach to simulate the impact of observation, boundary condition and model uncertainties on the 4D-Var data assimilation system (Isaksen et al., 2010). This makes it possible to quantify uncertainty in the background and in the analysis. Since each EDA member is an independent 4D-Var data assimilation with multiple outer loops, the EDA is capable of sampling non-Gaussian posterior probability density functions. The maintenance cost of the EDA is low, as there is no need to support a separate data assimilation system for the ensemble assimilation component, and it is straightforward to propagate new developments in the high-resolution 4D-Var to the EDA. The EDA scales perfectly with ensemble size and contributes significantly to the skill of ECMWF's forecasts, but its computational cost is high. As a result, the EDA members are run at lower resolution than the high-resolution 4D-Var and the number of EDA members is currently limited to 25. However, it is advantageous to increase the number of EDA members, for two reasons. First, a bigger ensemble size improves the sampling of flow-dependent errors of the background used in ECMWF's 4D-Var system. Improved estimates of background errors in the data assimilation system in turn enable an improved high-resolution analysis. Second, increasing the EDA ensemble size to 50 makes it possible to assign independent EDA-based initial perturbations to each of the 50 ensemble forecast members. The resulting exchangeability of ensemble members is particularly important for research experiments involving ensemble forecasts.

Figure 1 illustrates the impact of the new 50-member EDA on the uncertainty estimates used in ECMWF's 4D-Var system. Due to better sampling, the ensemble standard deviation of the 50-member EDA background forecasts is spatially smoother than the ensemble standard deviation of the 25-member EDA background forecasts. Going to 50 EDA members reduces the sampling noise. As a result, the number of points with very small and very large standard deviations is reduced. This in turn improves the use of observations, as observations in regions with unrealistically small (unrealistically large) standard deviations are given too little (too much) weight in the 4D-Var system.

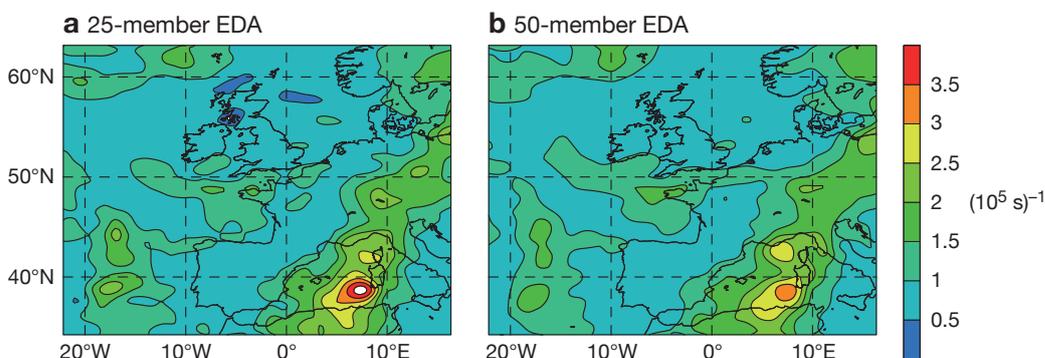


Figure 1 EDA 3-hour background forecast ensemble standard deviation of relative vorticity at model level 100 starting at 18 UTC on 31 August 2017 for (a) the 25-member EDA and (b) the 50-member EDA.

Computational efficiency

The computational cost per individual member in the 50-member configuration has been reduced so that the overall cost is the same as for the current 25-member configuration. This has been achieved by extensive optimisation, for example in the preconditioning of the first 4D-Var minimisation of the perturbed members with information from the minimisation of the unperturbed control member; the blacklisting of passive observations in the perturbed members; and improved observation thinning and optimisation of the observation operator setup for lower-resolution inner loops. In addition, non-essential tasks have been deactivated. For example, by default we now run a reduced number of final trajectories, which are merely used for diagnostic purposes, and the surface analysis is run with an optimal interpolation scheme instead of the operational simplified extended Kalman filter of the high-resolution analysis. Extensive tests have shown that the optimisation does not negatively impact the skill of individual EDA members or the EDA's uncertainty estimates.

A welcome side effect of the optimisation work is that the observation-related changes will also reduce the computational cost of all standard deterministic, low-resolution research experimentation and hence make it possible to achieve a higher throughput for research experiments.

Exchangeability

The EDA members are used together with singular vectors to generate perturbations which are added to the high-resolution analysis to construct the perturbed initial conditions for ensemble forecasts. Currently the initial perturbations of the ensemble forecasts have a plus–minus symmetry to distribute the 25 EDA members between the 50 ensemble forecast members: the EDA perturbation generated from the first EDA member is assigned to the first ensemble forecast member, the first EDA perturbation with the sign reversed is assigned to the second ensemble forecast member, and so on. A disadvantage of this scheme is that the ensemble forecast members are not exchangeable. Exchangeability is desirable for research ensemble forecast testing with small ensemble sizes (Leutbecher, 2018) to estimate the skill of the full 50-member ensemble. By increasing the number of EDA members to match the number of ensemble forecast members, it is possible to assign one individual EDA perturbation to each ensemble forecast member. As a result, the plus–minus symmetry of the initial perturbations is no longer required and the ensemble forecast members become exchangeable.

Figure 2 shows the mean absolute difference between pairs of members in an ensemble forecast experiment. In the ensemble forecast experiment that has the plus–minus symmetry of the initial perturbations from 25 EDA members, there is a distinct difference between pairs of members that share the same initial perturbation with only the sign reversed on the one hand, and pairs of members with independent initial perturbations on the other. In contrast, the members of the ensemble forecast experiment with perturbations from 50 EDA members and without the plus–minus symmetry are indistinguishable.

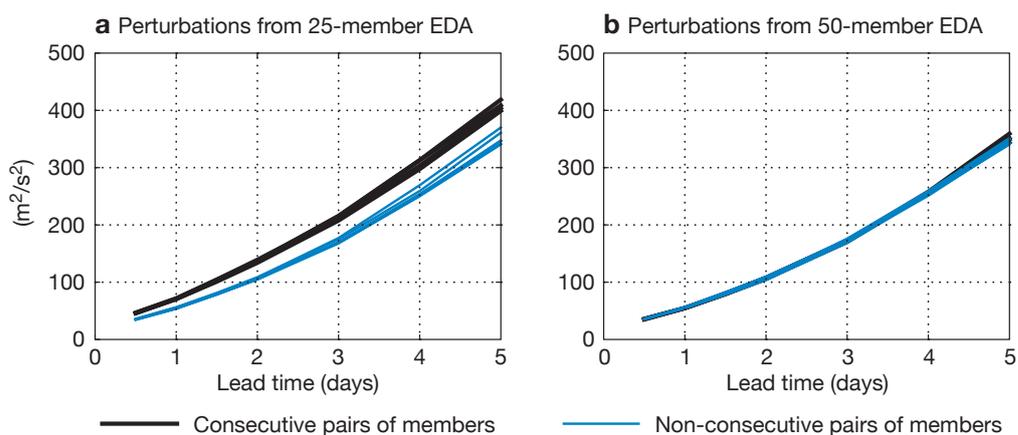


Figure 2 Mean absolute difference of 500 hPa geopotential in the northern extratropics between consecutive pairs of ensemble forecast members and non-consecutive pairs for (a) an ensemble forecast experiment with plus–minus symmetry of the initial perturbations (based on a 25-member EDA) and (b) an ensemble forecast experiment without such symmetry (based on a 50-member EDA). The differences are averaged over a period of 41 days.

Research configuration

In addition to the new 50-member configuration, we have introduced computationally inexpensive research and testing EDA configurations that run with only 10 members. They will reduce the cost of EDA experimentation and streamline the testing and validation process for new model cycles. Specifically, from the next IFS upgrade, the operational EDA will have 50 members with outer loops at TCo639 resolution and two inner loops at TL191 resolution. The research configuration, on the other hand, has only 10 members, TCo399 outer loops and TL95/TL159 inner loops. Only the control member is produced using the same 4D-Var configuration as the standard low-resolution (TCo399) deterministic research experiments used for model development. This makes it considerably cheaper to assess the impact of model changes on the EDA and the feedback this has on the analysis. The cost of the EDA research configuration is only approximately three times that of a standard deterministic research experiment. The cost of equivalent experimentation with the currently operational 25-member EDA is over 20 times higher than a deterministic research experiment.

For research and development, it is desirable to run experiments at lower resolution to reduce computational cost. However, it is important for conclusions derived from research experiments to be relevant to the full-resolution operational system. For example, a new development that improves forecast quality in a research experiment compared to a reference experiment should also improve the high-resolution operational system. In order to test if this is indeed the case for the research EDA configuration, we look at a change to the Stochastically Perturbed Parametrization Tendencies (SPPT) scheme, which is used in the EDA to simulate model uncertainties. The focus here is not a decrease or increase in forecast skill but whether the full operational EDA configuration and the research EDA configuration react to the change in the same way. Figure 3 shows the impact on the skill of forecasts based on the unperturbed EDA control member resulting from a change to the SPPT. The change influences the EDA ensemble standard deviation and thus the uncertainty estimates used in the 4D-Var system. Consequently, it impacts the forecast skill of each EDA member, including the unperturbed EDA control member. The plot shows the results obtained in two sets of forecast skill experiments, one run with the currently operational 25-member EDA configuration – which is equivalent in cost to the new optimised 50-member configuration – and one run with the new 10-member research configuration. For each configuration, the skill of the unperturbed control forecasts is compared to a reference experiment which is run without the change to the stochastic physics scheme in the EDA. The difference between the experiment with the stochastic physics change and the reference experiment is very similar regardless of whether the 25-member configuration or the 10-member research configuration is used. This shows that it is possible to derive robust results from the research EDA configuration for a fraction of the cost of running the full operational configuration. However, it is essential to compare like with like and not to mix the different configurations.

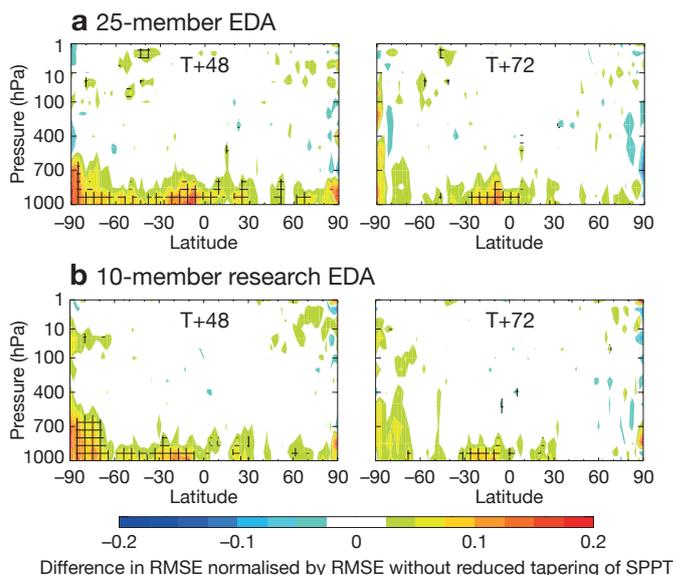


Figure 3 Impact of reduced tapering of SPPT (stochastic physics) in the boundary layer on the root-mean-square error (RMSE) in temperature for forecasts at different lead times based on the EDA control members in (a) the current EDA configuration with 25 members and (b) the proposed research test configuration with 10 members and lower-resolution outer loops and lower-resolution inner loops for the perturbed EDA members. Green to red shading indicates larger RMSE with reduced tapering of SPPT.

Outlook

The current method to initialise ensemble forecasts is to apply EDA-based perturbations and singular vectors to the high-resolution deterministic analysis. That analysis is more accurate than the EDA analyses thanks to its higher resolution and the fact that it contains six extra hours of observations compared with the EDA. The long-term goal, however, is to develop a high-resolution EDA which runs in parallel with the high-resolution 4D-Var and includes the same observations, and which can provide the initial conditions for ensemble forecasts directly. In addition to the benefits described in this article, the development of a computationally efficient 50-member EDA is a step towards that goal.

Further reading

Isaksen, L., M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher & L. Raynaud, 2010: Ensemble of Data Assimilations at ECMWF, *ECMWF Technical Memorandum No. 636*.

Leutbecher, M., 2018: How many ensemble members are desirable? *ECMWF Newsletter No. 157*, 5.

© Copyright 2019

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, England

The content of this Newsletter is available for use under a Creative Commons Attribution-Non-Commercial-No-Derivatives-4.0-ported Licence. See the terms at <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

The information within this publication is given in good faith and considered to be true, but ECMWF accepts no liability for error or omission or for loss or damage arising from its use.