ECMWF Feature article

from Newsletter Number 135 – Spring 2013

METEOROLOGY

Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations

www.ecmwf.int/en/about/news-centre/media-resources

doi:10.21957/q99zv6da

This article appeared in the Meteorology section of ECMWF Newsletter No. 135 – Spring 2013, pp. 20–24.

Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations

Sean Healy, Florian Harnisch, Peter Bauer, Steve English

In *'The ECMWF Strategy 2011–2020'* one of the complementary goals is to contribute towards the optimisation of the Global Observing System (GOS) so that it meets the future needs of weather forecasting and climate monitoring. The evolution of the GOS should reflect updated user requirements, and the emergence of new – or recently demonstrated – technologies, which may complement the more established measurement techniques. The satellite component of the GOS is composed of a diverse set of complex observing systems, each with particular strengths, so optimising the future GOS requires finding a reasonable balance of these distinct measurement types.

Global Navigation Satellite System Radio Occultation (GNSS-RO) measurements complement the information provided by satellite radiances, and these measurements are of particular value in the upper troposphere and stratosphere because of their high vertical resolution. More information about the GNSS-RO measurement technique is given in Box A.

A question that has arisen in the context of designing the future GOS is how the impact of GNSS-RO measurements is likely to be enhanced if the number can be increased significantly above the 2,500 to 3,000 observations that are currently available. This is particularly relevant because new GNSS systems, such as Galileo, could increase the GNSS-RO observation number. ECMWF has recently completed a Galileo Science Study, funded by the European Space Agency (ESA), to address this question. This involved investigating the impact of up to 128,000 GNSS-RO observations per day (*Harnisch et al.*, 2013).

The Galileo Science Study supports the case for a considerable increase in the number of GNSS-RO measurements above the present day levels. More specifically, the results indicate that we should be aiming for at least 16,000 to 20,000 operational GNSS-RO measurements per day, in the coming years. These results have helped inform a recent revision of the WMO *'Vision for the Global Observing System in 2025'*.

Method

Estimating the impact of increasing the GNSS-RO observations requires the ability to simulate the measurements accurately, and then assess them within the context of a state-of-the-art numerical weather prediction (NWP) system. Their impact must be given relative to a reasonable GOS baseline, in order to obtain reliable and robust results that will remain relevant in the coming years.

One well-established approach for investigating the impact of the simulated data is to perform Observing System Simulation Experiments (OSSEs) (*Andersson & Masutani,* 2010). However, OSSEs require the simulation of all the observation types that are assimilated, and they are extremely computationally expensive. In this study we have used an alternative approach based on an Ensemble of Data Assimilations (EDA).

An EDA can be used to derive flow-dependent theoretical analysis and short-range forecast error covariance information (See *Isaksen et al.*, 2010). An ensemble of ten 4D-Var assimilations is run in parallel, but with perturbed observations, model physics and sea surface temperatures. If the perturbations are consistent with the actual error statistics of these variables, then the EDA will provide an accurate estimate of the analysis and short-range forecast error statistics – or uncertainty – of the NWP system. The uncertainty is estimated from the 'spread' of the ensemble, which is the standard deviation of the ensemble members about the mean of the ensemble. It is sometimes stated that ensembles provide 'errors of the day', but this is shorthand terminology and it can be misleading. More precisely, the ensembles provide estimates of the 'error statistics of the day'.

An EDA has been operational at ECMWF since July 2010, and it has had a significant impact on forecast scores (*Bonavita et al.*, 2012). Furthermore, *Tan et al*. (2007) have shown that the EDA can be used to assess the impact of new measurements. They demonstrated that simulated wind profiles from ESA's Earth Explorer Atmospheric Dynamics Mission (ADM-Aeolus) can be assimilated alongside real data in an EDA, and that the EDA spread values can be used to estimate the impact of the simulated data. The central idea is that if the new observations are going to have a positive impact, the estimate of the uncertainty should be reduced and thus the EDA spread values should be reduced.

 doi:10.21957/q99zv6da

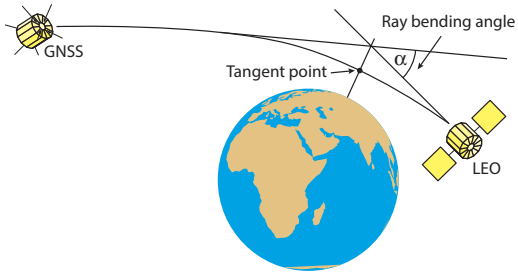
The GNSS-RO measurement technique

A

The figure shows the geometry of the GNSS-RO measurement technique. A radio signal is transmitted by a GNSS satellite, passes through the atmosphere and is measured with a receiver placed on a LEO satellite, such as the GRAS instrument on Metop-A. The path is bent as a result of gradients in the refractive index of the atmosphere, which in turn can be related to gradients in the temperature and humidity.

The ray bending angle, α, can be derived from the time required for the radio signal to propagate between the GNSS and LEO satellites, and the motion of the LEO satellite enables the variation of α as a function of tangent height to be determined. Temperature, humidity and surface pressure information can be retrieved directly from the bending angle profiles within the 4D-Var framework. The limb sounding geometry of the measurements means that they have good vertical resolution, focussed on a few hundred kilometres around the horizontal location of the tangent point.

The measurements provide particularly good temperature information in the vertical interval between ~200 hPa to ~5 hPa. They are less sensitive to biases than satellite radiance measurements, because they are derived from a precise measurement of a time delay with an atomic clock. Consequently, the GNSS-RO measurements are assimilated without bias correction, and so they can 'anchor' the bias correction of satellite radiances.



The geometry of a GNSS-RO measurement. A radio signal is emitted by the GNSS satellite and measured with a receiver placed on the low earth orbit (LEO). The path of the radio signal is bent as a result of refractive-index gradients in the atmosphere. The motion of the LEO satellite enables the variation of ray-bending with tangent height to be derived.

EDA experiments

The EDA experiments have been performed for a six week period from 1 July to 15 August 2008. This period was chosen because it enabled comparison with a detailed set of GNSS-RO observing system experiments (*Bauer et al.*, 2013). The EDA spread values have been used to estimate the impact of 2,000, 4,000, 8,000, 16,000, 32,000, 64,000 and 128,000 simulated GNSS-RO observations, relative to a baseline observing system that assimilates all the non GNSS-RO observations that were used operationally during the period of the experiment. It is worth noting that although 128,000 observations per day is considered a very dense network of GNSS-RO measurements, the mean spacing between observations in an hour is still about 300 km.

Two additional EDA experiments have also been performed for comparison purposes: one assimilating real GNSS-RO measurements, and the other assimilating neither simulated nor real GNSS-RO measurements.

The simulated GNSS-RO observations are generated using ECMWF operational analyses (see Box B). They are assumed to be randomly distributed in both space and time. This simplification means that we do not perform any satellite orbit modelling in order to calculate the observation locations. Random noise is added to the simulated data to simulate observation error.

The simulated GNSS-RO measurements are subsequently assimilated in the EDA experiments assuming the same observation error statistics that are used operationally at ECMWF. The EDA spread values depend on the assumed observation error statistics used in the EDA system, rather than the actual observation error statistics. This may seem a rather subtle point, but it has been key to understanding many aspects of the study. For example, it explains why – unlike OSSEs – it is possible to use a mix of real and simulated data in the EDA approach, and also why the ECMWF analyses can be used for simulating the measurements.

The EDA provides a theoretical estimate of the analysis and short-range forecast error statistics, which are valid for the assumed observation error statistics. Clearly, the hope must be that the assumed and actual errors statistics are reasonably close for real measurements, and the fact that the real data is assimilated successfully would support this assumption. However, the error statistics used in the EDA for the simulated GNSS-RO measurements represent the characteristics we hope to obtain when the corresponding real data becomes available, rather than the actual, complex error statistics of the simulated data that is assimilated.

doi:10.21957/q99zv6da

Some remarks, technical aspects and interpretation

B

The technical details of the EDA experimentation are described in *Harnisch et al*. (2013). The simulated observations are produced with a two-dimensional (2D) observation operator, which accounts for the real limb geometry of the measurement by using the NWP information in a plane to compute the bending angles. The simulated GNSS-RO measurements are subsequently assimilated in the EDA experiments assuming the same observation error statistics that are used operationally at ECMWF, and using a one-dimensional (1D) operator which uses only the NWP profile information at the location of the observation.

The obvious inconsistency of simulating the observations with a 2D operator, but assimilating with a 1D operator was deliberate. Firstly, it reflects how the GNSS-RO measurements are currently used at most NWP centres. Furthermore, it was originally expected that this inconsistency would introduce correlated forward model errors, which would influence the spread values. However, this is not correct because the spread values are determined by the observation error statistics assumed in the EDA system that are used to generate the perturbations, rather than the actual errors on the simulated measurements. This fact can be related to well-known theory. For example, the theoretical error statistics provided by a linear Kalman filter depend only on the assumed observation and model error covariance matrices.

In general, we note that there are some clear similarities between this type of EDA calculation and the one-dimensional information content studies routinely used satellite meteorology. One-dimensional variational (1D-Var) retrievals usually provide a solution error covariance matrix. For example, these are often used in radiance channel selection studies. In the 1D-Var case, the covariance matrices can be computed explicitly, because they have relatively small dimensions, of order 100. In the 4D-Var case, the covariance matrices are much larger, because they relate to a full three-dimensional atmospheric state estimate, so they have to be approximated using the ensemble methods. We interpret the EDA spread results as a 4D-Var information content study.

Main results with simulated GNSS-RO data

Figure 1 shows the time series for the EDA spread values for 12-hour temperature forecasts at 100 hPa, averaged over the (a) northern hemisphere extra-tropics, (b) tropics and (c) southern hemisphere extra-tropics. This is a fairly typical EDA time series, irrespective of the chosen variable or level. There is generally a spin-up period of roughly one week before the spread values reach a reasonably stable state, which is then maintained for the remainder of the experiment. The increase in the EDA spread during the early stages is because the EDA members are 'cold started' from the same initial forecast state. The subsequent stability of the EDA spread values is because they approximate the forecast error statistics, rather than the forecast errors themselves. This stability of the EDA spread values has important practical implications, suggesting that it is not necessary to run extremely long experiments to test new measurements with this approach.

The vertical profiles of the EDA spread values of the temperature analyses are shown in Figure 2. These are averaged for the period 8 July to 15 August 2008 to remove the spin-up period. The inclusion of the either real or simulated measurements reduces the spread values compared to a control experiment, where GNSS-RO measurements are not assimilated. In general, there is good agreement between the EDA spreads produced by the ~2,500 real GNSS-RO observations (EDA_real) and 2,000 simulated measurements (EDA_02). The small differences in these spread values can be attributed to the random spatial sampling assumed when generating the simulated data, which produces a slightly different latitudinal distribution for real and simulated data, as shown in Figure 3. The largest reductions in the spread values are in the upper-troposphere and lower/middle-stratosphere, where the GNSS-RO information content is known to be greatest. The spread values continue to reduce as more GNSS-RO measurements are added, even for observation numbers greater than 32,000 per day (EDA_32).

We have been able to compare the EDA results with a comprehensive set of GNSS-RO observing system experiments (OSEs) conducted over the same period. These OSEs were designed to investigate whether there is any indication of saturation of forecast impact with the currently available GNSS-RO observation numbers, and they looked at the impact of assimilating 0, 5, 33, 66 and 100% of the available observations. The 100% coverage in the OSEs represents an average of 2,450 real observations per day. Figure 4 shows a comparison of the OSE results with the EDA spread values for the 24-hour temperature error at 100 hPa. In both cases the results are 'normalized' with their result when the GNSS-RO measurements are not assimilated. There is reasonably good agreement between the two approaches for low observation numbers and this result provides us with additional confidence in the EDA results.

doi:10.21957/q99zv6da

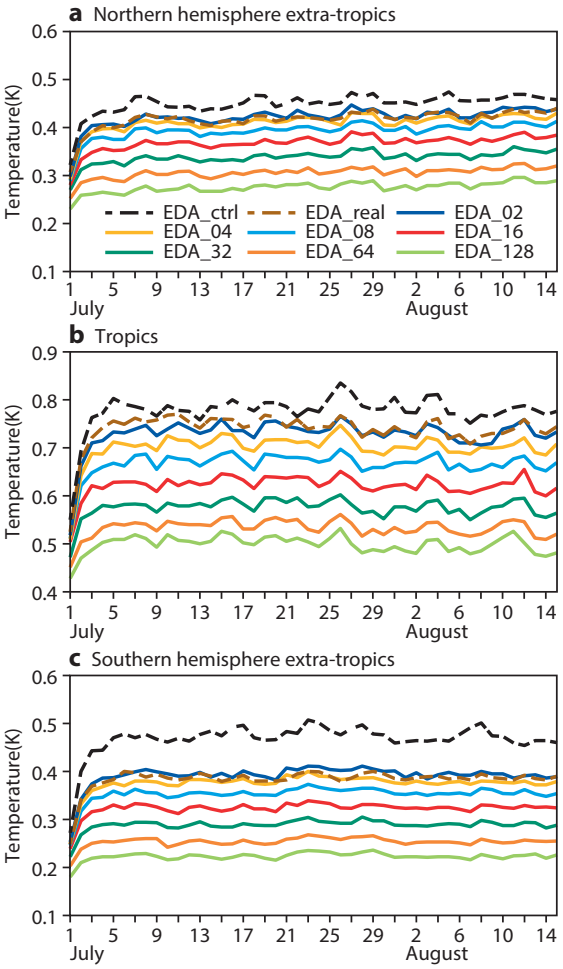


Figure 1 Time series for July to August 2008 of daily 12-hour forecast EDA spread values for temperature (K) at 100 hPa for (a) northen hemisphere extra-tropics, (b) tropics and (c) southern hemisphere extra-tropics. 'EDA_n', refers to the EDA experiment with 'n' thousand simulated GNSS-RO observations. EDA_ctl is the control experiment with no GNSS-RO measurements assimilated and EDA_real is with the real GNSS-RO data.

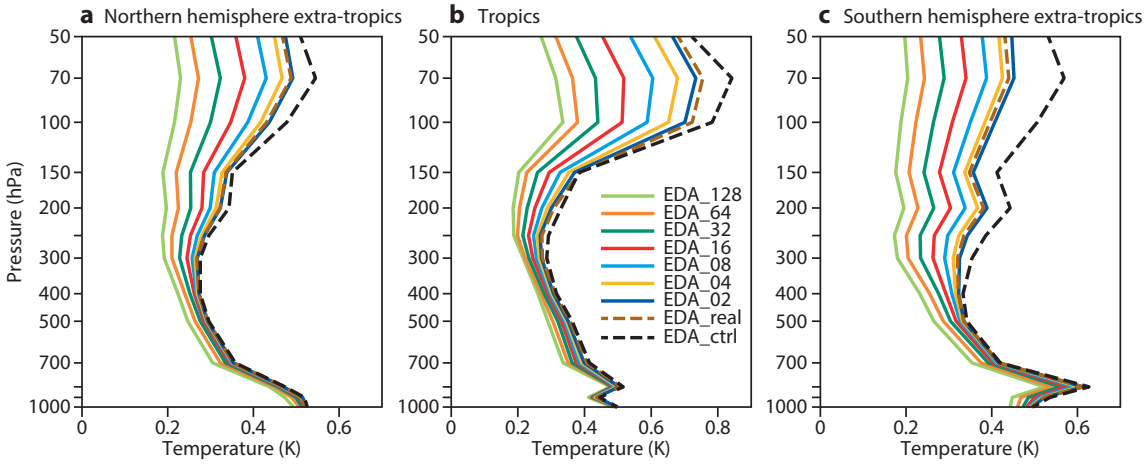


Figure 2 EDA spread for temperature (K) at analysis time averaged over (a) northern hemisphere, (b) tropics and (c) southern hemisphere. Results are for the period 8 July 2008 to 15 August 2008. As with Figure 1, 'EDA_n', refers to the EDA experiment with 'n' thousand simulated GNSS-RO observations, EDA_ctl is the control and EDA_real is with the real GNSS-RO data.

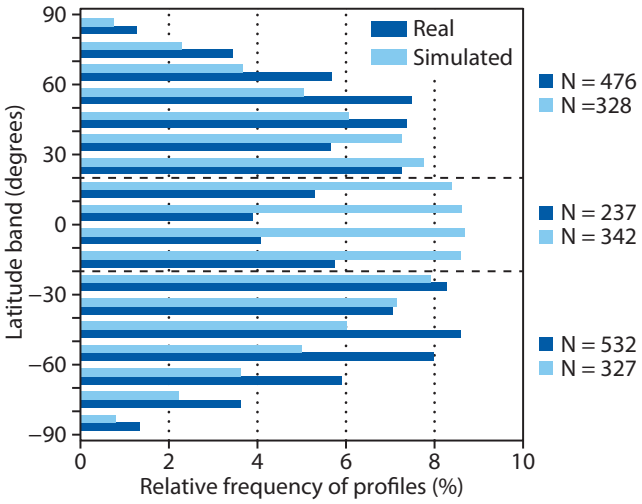


Figure 3 Zonal distribution of real (black) and simulated (grey) GNSS-RO profiles in 10 degree latitude bins for EDA_real (black) and EDA_2 (grey). The relative frequency with respect to the total number of 1245 for EDA_real and 997 for EDA_2 is shown. The horizontal dashed lines separate the southern hemisphere, from the tropics and the northern hemisphere. The bending angle numbers are shown on the right side for real data (black) and simulated data (grey). Results are the average values over the 00 UTC and 12 UTC assimilation cycles in the period 1 July 2008 to 15 August 2008.

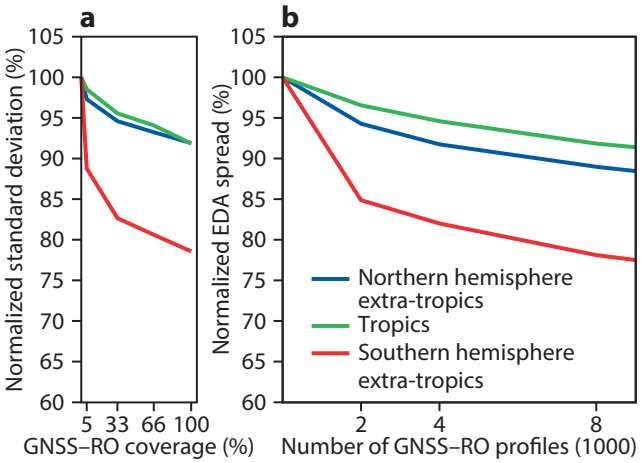


Figure 4 (a) Normalized standard deviation of 24-hour forecast error of observing system experiments as function of the assimilated percentage of real GNSS-RO data and (b) normalized EDA spread (%) at the 24-hour forecast range as function of the assimilated number of simulated GNSS-RO profiles. Results are for temperature at 100 hPa for the period 8 July 2008 to 15 August 2008.

Figure 5 shows the normalized EDA spread as a function of GNSS-RO observation number for temperature, geopotential height and relative humidity on various fixed pressure levels. Again, the normalization is performed by dividing each spread value by the no GNSS-RO result. The continued reduction in the spread values is clear when the results are plotted in this way. However, one useful feature that has been noticed is that in general 16,000 observations provide about 50% of the spread reduction that is obtained with 128,000 observations, and 16,000 observations is on a steep part of the curve.

It was originally anticipated that there would be some evidence of saturation of observation impact, whereby additional observations produce little additional impact for the high observation numbers. Perhaps surprisingly, this is not the case. However, once the link between EDA spread values and the theoretical error statistics is acknowledged, this result is an obvious consequence of standard data assimilation/retrieval theory.

It is useful to distinguish between having correlated observations and correlated observation errors. Highly correlated observations with uncorrelated errors will always reduce the theoretical analysis/retrieval errors, even if the additional measurements do not effectively add new information to the NWP system. This is simply a result of reducing random measurement noise by repetition, and it is standard experimental practice in a laboratory. Exactly the same arguments apply to the EDA approach. Even though additional measurements might not contain significant new information, they can reduce the spread values of the system by effectively reducing measurement noise through repetition, as long as uncorrelated observation errors are assumed. Therefore, we should not expect the curves shown in Figure 5 to plateau and have a zero gradient, until the spread hits a lower limit determined by the assumed model error statistics.

We have investigated if the reduction of the EDA spread at large observation numbers is solely because the additional observations are repetitions of the previous measurements, but this appears not to be the case. An additional EDA experiment was performed using 64,000 observations, but doubling the assumed observation error variances. We found that the 64,000 ‘degraded’ observations produced lower spread values than 32,000 observations with the standard errors.

doi:10.21957/q99zv6da

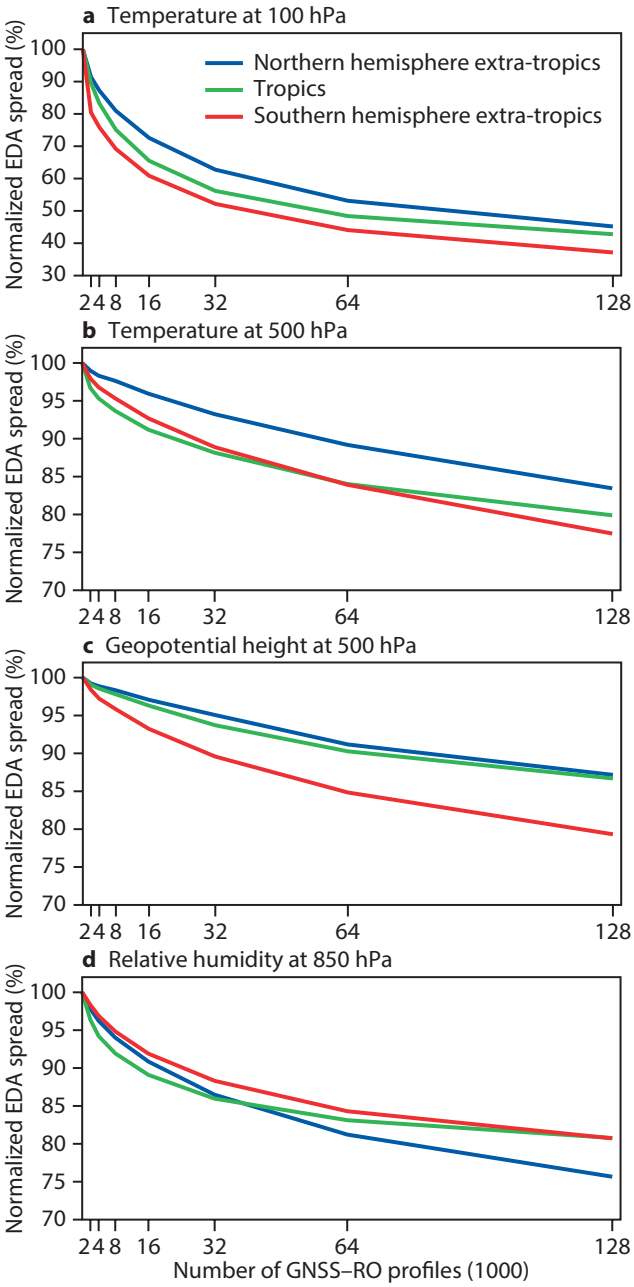


Figure 5 Normalized EDA spread (%) at analysis time as function of the assimilated number of simulated GNSS-RO profiles for (a) temperature at 100 hPa, (b) temperature at 500 hPa, (c) geopotential height at 500 hPa and (d) relative humidity at 850 hPa. Results are for the period 8 July 2008 to 15 August 2008.

EDA spread-skill relationship when using simulated observations

As noted earlier, in standard EDA experiments using real observations, there is a clear relationship between the EDA spread and analysis and short-range forecast error statistics. The EDA spread values will provide useful information about the actual error statistics of the NWP system, provided the assumed observation and model error statistics are reasonable.

This spread-skill relationship breaks down with simulated data because the assumed GNSS-RO observation errors are clearly incorrect. In reality, the simulated observations provide no new information about the atmospheric state because they have been derived from ECMWF analyses. Therefore, it should be noted that the assimilation of the simulated GNSS-RO measurements cannot improve the actual skill of the NWP forecasts, even if the spread values have been reduced (see *Harnisch et al.*, 2013).

Summary and future work

We have used the EDA system to estimate the impact of increasing the number of GNSS-RO measurements from their current levels of ~2,500 per day up to 128,000 profiles per day. We do not reach a saturation of forecast impact, but broadly speaking 16,000 observations per day have around half the impact of 128,000 observations on the EDA spread values. The results suggest that there will be a benefit from increasing the number of the GNSS-RO measurements, and we believe the aim should be to provide at least 16,000 to 20,000 operational measurements per day.

The EDA has proved to be a very useful and efficient tool for this study. In future work, it might be possible to refine the results and look at the impact of different observation sampling patterns. It would also be interesting to compare EDA results with OSSEs for a common period.

Further reading

Andersson, E. & **M. Masutani**, 2010: Collaboration on Observing System Simulation Experiments (Joint OSSE). *ECMWF Newsletter* **No. 123**, 14–16.

Bauer, P., G. Radnoti, S. Healy & **C. Cardinali**, 2013: GNSS radio occultation constellation observing system experiments. *ECMWF Tech. Memo. No. 692*.

Bonavita, M., L. Isaksen & **E. Hólm**, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. Q. J. *R. Meteorol. Soc.*, **138**, 1540–1559.

Isaksen, L., J. Haseler, R. Buizza & **M. Leutbecher**, 2010: The new Ensemble of Data Assimilations. *ECMWF Newsletter* **No. 123**, 17–21.

Harnisch, F., S.B. Healy, P. Bauer & **S.J. English**, 2013: Scaling of GNSS radio occultation impact with observation number using an ensemble of data assimilations. *ECMWF Tech. Memo. No. 693*.

Tan, D.G.H., E. Andersson, M. Fisher & **l. Isaksen,** 2007: Observing-system impact assessment using a data assimilation ensemble technique: application to the ADM-Aeolus wind profiling mission. *Q. J. R. Meteorol. Soc.*, **133**, 381–390.

© Copyright 2016

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

The content of this Newsletter article is available for use under a Creative Commons Attribution-Non-Commercial-No-Derivatives-4.0-Unported 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.

 doi:10.21957/q99zv6da