Received: 4 May 2020 | Revised: 18 August 2020 | Accepted: 21 September 2020

DOI: 10.1002/qj.3917

Quarterly Journal of the Royal Meteorological Society RMetS

RESEARCH ARTICLE

# Continuous data assimilation for global numerical weather prediction

P. Lean[1] | E. V. Hólm[1] | M. Bonavita[1] | N. Bormann[1] | A. P. McNally[1] | H. Järvinen[2]

[1]European Centre for Medium-Range Weather Forecasts, Reading, UK

[2]Institute for Atmospheric and Earth System Research (INAR), University of Helsinki, Finland

**Correspondence**
P. Lean, European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK.
Email: peter.lean@ecmwf.int

**Abstract**

A new configuration of the European Centre for Medium-Range Weather Forecasts (ECMWF) incremental 4D-Var data assimilation (DA) system is introduced which builds upon the quasi-continuous DA concept proposed in the mid-1990s. Rather than working with a fixed set of observations, the new 4D-Var configuration exploits the near-continuous stream of incoming observations by introducing recently arrived observations at each outer loop iteration of the assimilation. This allows the analysis to benefit from more recent observations. Additionally, by decoupling the start time of the DA calculations from the observational data cut-off time, real-time forecasting applications can benefit from more expensive analysis configurations that previously could not have been considered. In this work we present results of a systematic comparison of the performance of a Continuous DA system against that of two more traditional baseline 4D-Var configurations. We show that the quality of the analysis produced by the new, more continuous configuration is comparable to that of a conventional baseline that has access to all of the observations in each of the outer loops, which is a configuration not feasible in real-time operational numerical weather prediction. For real-time forecasting applications, the Continuous DA framework allows configurations which clearly outperform the best available affordable non-continuous configuration. Continuous DA became operational at ECMWF in June 2019 and led to significant 2 to 3% reductions in medium-range forecast root mean square errors, which is roughly equivalent to 2–3 hr of additional predictive skill.

**KEYWORDS**

4D-Var, continuous data assimilation, data assimilation, operational numerical weather prediction

## 1 | INTRODUCTION

From the time of the first operational numerical weather prediction (NWP) in the 1950s until the late 1990s, models were initialised using analyses created primarily from conventional (non-satellite) observations, such as synoptic surface reports and radiosonde ascents. Since these observations were mainly taken at synoptic times (0000, 0600, 1200, 1800 UTC), the daily schedule of analysis and forecasting was synchronized accordingly. However, over the past two decades, the nature of the Global Observing System (GOS) has changed dramatically with satellites

  

and aircraft now providing the bulk of observations in a near-time-continuous stream. The analysis quality is thus less critically coupled to the synoptic observing times. While many short-range, limited-area NWP systems now run more frequently (Benjamin *et al.*, 1994; 2016, Milan *et al.*, 2019), most operational global NWP centres still produce their forecasts at the synoptic times only.

Similarly, the algorithms used to perform data assimilation (DA) have undergone considerable change. Until the mid-1990s, atmospheric analyses at ECMWF and in many other weather centres were produced using Optimum Interpolation (OI) Lorenc (1981). A feature of OI, and Kalman filtering in general, is that it solves the observation weight matrix in a single computation. The only option to ensure maximum observational input in operational OI systems is to wait as long as possible for late arriving data (until the so-called 'cut-off' time) and then complete the assimilation and forecast computations as fast as possible to meet the fixed forecast dissemination schedule. Clearly, such a system maximises time-critical computations and is inherently inflexible.

Within this framework, where the DA and forecast need to complete in the time available between a fixed observation cut-off time and a required forecast dissemination time, the design of the operational NWP configuration becomes an optimisation problem. Modifications to the system in terms of resolution/complexity and the computer resources dedicated to analysis and forecast are made to maximise the skill of the forecasts that are available at the required dissemination time. In particular, the scheduling of the cut-off time involves a trade-off between the available number of observations and the time allowed for DA and forecast computations. For example, a later cut-off time allows more observations to be assimilated, but also means that less time is available to perform the DA and model forecast calculations if the fixed dissemination schedule is to be met.

However, OI has now been widely replaced with four-dimensional variational data assimilation (4D-Var) Rabier *et al.*, (2000). In 4D-Var the observations constrain the model state better the more abundant they are, thus favouring the use of longer assimilation windows. However, as the length of the assimilation time window is increased, the cost function contours begin to foliage due to the chaotic nature of the geophysical flow, resulting eventually in multiple minima (Gauthier, 1992). Therefore, it is important to ensure the model state remains close to the true minimum and avoids one of the fallacy minima with associated loss of predictability.

Variational analysis algorithms are iterative by nature and allow greater flexibility to configure computations in an operational environment. The iterative nature of 4D-Var allows gradual extension of the assimilation window length at consecutive minimisation steps where new observations can be introduced in each successive outer loop. Pires *et al.*, (1996) studied the Lorenz (1963) system and suggested quasi-static variational assimilation. They concluded that gradual time-extension of the assimilation window improves predictability because the solution lies closer to the true minimum and is more strongly constrained by observations. Earlier, (Järvinen *et al.*, 1995) and (Järvinen *et al.*, 1996) had suggested an identical concept of quasi-continuous variational DA to reduce the amount of time-critical computations to improve operational resilience, as well as allowing the very latest arriving observations to affect the solution and to perform a more accurate iterative search of the solution. Their experiments with the ECMWF pre-operational 4D-Var at T21 resolution confirmed that it is possible to compute preliminary solutions with incomplete observational input and have a good approximation of the atmospheric state at the time when the very latest observations arrive.

The concept of quasi-continuous (Järvinen *et al.*, 1995, 1996) or quasi-static (Pires *et al.*, 1996) data assimilation thus exploits the fact that most of the observations are available long before the cut-off time. In these configurations, the cut-off time becomes far less categoric. The start time of the DA computations becomes decoupled from the time at which the last observations arrive. Alternative DA configurations can then be explored to optimise performance of real-time applications. It is worth noting that the quasi-continuous DA concept is unrelated to the similar sounding Variational Continuous Assimilation technique (Derber, 1989; Böker, 2010) which is a form of weak-constraint DA.

Apart from experimentation (Veersé and Thépaut, 1998; Choi *et al.*, 2013), the Meteorological Service of Canada was the first to apply this idea in operations, albeit in a limited sense (Gauthier *et al.*, 2007). They implemented an incremental 4D-Var (Courtier *et al.*, 1994) such that the first inner loop was started before the cut-off time, and observations collected during these computations were added to the assimilation for the second inner loop. This choice was justified by the long time required to run 4D-Var on the available computing resources and the necessity to include a large number of late-arriving observations. However, the impact of this choice was not discussed, and this configuration is no longer used.

While the early work on quasi-continuous DA showed clear promise, it has not been investigated in detail in a full-complexity, state-of-the-art global DA and forecasting system. One reason for this is that the concept requires the observational data processing software to accommodate continuous streams of data. At ECMWF this has been a major undertaking over recent years (e.g., Bauer *et al.*, 2020). Nevertheless, the concept has the potential to allow

modern DA systems to operate more continuously and to better utilize the present form of the GOS.

This paper introduces the new quasi-continuous DA configuration (herein referred to as 'Continuous DA') which has been operationally implemented in the Integrated Forecasting System of ECMWF. The benefits from using more expensive and accurate DA configurations in this framework are investigated.

## 2 | CONTINUOUS DATA ASSIMILATION CONCEPT

4D variational DA aims to determine the model trajectory that best fits in a least-square sense all the observations available during a given time window, according to their perceived accuracy and the perceived accuracy of the initial background state. This concept naturally leads to the formulation of the standard strong-constraint 4D-Var cost function:

$$\begin{aligned} J(\mathbf{x}_0) &= J_{\mathrm{B}}(\mathbf{x}_0) + J_{\mathrm{O}}(\mathbf{x}_0) \\ &= \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_{\mathrm{b}}) \\ &\quad + \frac{1}{2}\sum_{k=0}^{K}\{\mathbf{y}_k - G_k(\mathbf{x}_0)\}^{\mathrm{T}}\mathbf{R}_k^{-1}\{\mathbf{y}_k - G_k(\mathbf{x}_0)\}. \end{aligned} \quad (1)$$

In Equation 1, $\mathbf{x}_0$ is the control vector at the start of the assimilation window; $\mathbf{x}_{\mathrm{b}}$ and $\mathbf{B}$ are the background and its expected error covariance matrix (typically constructed using a combination of errors of the day and climatological errors, e.g., Bonavita *et al.*, 2016); $\mathbf{y}_k$ and $\mathbf{R}_k$ are the set of observations presented to the analysis in the $k$th sub-window and their expected error covariances; and $G_k$ is a generalised observation operator (or forward model) that produces the model equivalents of the observations $\mathbf{y}_k$ by first integrating the forecast model $M$ from $t_0$ to $t_k$ and then applying the standard observation operator $H_k$ to the propagated fields, that is:

$$G_k = H_k \circ M_{t_0 \to t_k}. \quad (2)$$

The solution of Equation (1) represents the general nonlinear weighted least-square solution of the assimilation problem using the forecast model as a strong constraint during the assimilation window. As in similar optimisation problems, the cost function in Equation (1) cannot be solved efficiently by standard methods for realistic NWP DA systems, given the size of the control vector $\mathbf{x}_0$ ($\mathcal{O}(10^9)$). A possible solution, first proposed in the meteorological community by Courtier *et al.*, (1994), under the name of 'incremental 4D-Var', is to simplify the solution of Equation (1) through the application of an approximated form of the Gauss–Newton method (Lawless *et al.*, 2005). This consists of approximating the minimisation of the nonlinear cost function in Equation (1) as a sequence of minimisations of linearised, quadratic cost functions defined in terms of perturbations around a sequence of progressively more accurate trajectories (i.e., nonlinear model integrations). Typically, the iterations of the linear minimisation algorithm are referred to as the 'inner loop' while the repeated re-linearisation around the new nonlinear model trajectories are known as the 'outer loop'. The cost function linearised around a guess trajectory $\mathbf{x}^{\mathrm{g}}$ can be expressed as an exact quadratic problem in terms of the increment $\delta\mathbf{x}_0$ at the initial time:

$$\begin{aligned} J(\delta\mathbf{x}_0) =& \frac{1}{2}(\delta\mathbf{x}_0 + \mathbf{x}_0^{\mathrm{g}} - \mathbf{x}_{\mathrm{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\delta\mathbf{x}_0 + \mathbf{x}_0^{\mathrm{g}} - \mathbf{x}_{\mathrm{b}}) \\ &+ \frac{1}{2}\sum_{k=0}^{K}\{\mathbf{d}_k - \mathbf{G}_k(\delta\mathbf{x}_0)\}^{\mathrm{T}}\mathbf{R}_k^{-1}\{\mathbf{d}_k - \mathbf{G}_k(\delta\mathbf{x}_0)\}. \end{aligned} \quad (3)$$

In Equation (3), $\mathbf{d}_k = \mathbf{y}_k - G_k(\mathbf{x}_0^{\mathrm{g}})$ are the observation departures around the latest model trajectory and $\mathbf{G}_k = \mathbf{H}_k\mathbf{M}_{t_0 \to t_k}$ is the linearisation of the generalised observation operator around the defined trajectory (where $\mathbf{H}_k$ is the linearised observation operator and $\mathbf{M}_{t_0 \to t_k}$ is the tangent linear of the forecast model). While convergence of incremental 4D-Var cannot be guaranteed in general, in practice it has been found to work well in the ECMWF operational DA system, with visible improvements in analysis accuracy for up to six outer-loop re-linearisations (Bonavita *et al.*, 2018).

If we call $\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_K)^{\mathrm{T}}$, the concatenation of the vector of observation departures in the assimilation window, in standard incremental 4D-Var the size of this vector remains unchanged during the minimisation procedure, while the values of its elements change due to the evolving trajectories and the nonlinear response of the generalised observation operator. In this sense, incremental 4D-Var repeatedly solves the same statistical problem taking advantage of progressively more accurate starting points (i.e., the successive trajectories). In contrast, in the continuous implementation of incremental 4D-Var, the size of $\mathbf{d}$ varies from one minimisation to the next as new observations are made available to the analysis algorithm.

From a conceptual point of view, letting additional observations into the assimilation system between one minimisation and the next implies that we are solving slightly different optimisation problems in different outer-loop iterations. In this sense, the only 'real' analysis takes place in the last outer loop–inner loop minimisation, and the role of the previous iterations is to present this analysis with the best starting point possible given the available observations up to that point and the background.

In practice, the number of additional observations available between successive outer-loop re-linearisations is large in absolute terms (around 500,000 observations) but small in relative terms (only 3 to 4% of the total number of observations). This implies that the topology of the cost function does not change substantially between successive minimisations and the information on the Hessian of the cost function gained during one minimisation remains relevant for accelerating the convergence of the following minimisation using standard pre-conditioning techniques (Fisher, 1998). This is confirmed by the results of the experiments described in the following, where no significant change to the converge characteristics of 4D-Var was noted during the continuous DA experimentation.

A more important practical consequence of the fact that in continuous DA the observations count increases at each outer loop is the fact that quality control ('screening') algorithms on the available observations need to be exercised at each outer-loop re-linearisation. The screening algorithms effectively implement nonlinear quality control (QC) decisions on the available observations based on some form of measure of the distance of the observation to the relevant model equivalent (e.g., mean squared distance normalised by assumed observation error variance, estimated presence/absence of cloud in either observed or simulated radiance, etc.). On the one hand, re-taking these screening decisions at each outer loop, taking advantage of a more accurate trajectory, increases the chances that the correct QC decision is made, especially for observations nonlinearly related to the model field and thus more sensitive to the results of this QC procedure. On the other hand, the re-screening may allow feedback processes, where incorrect screening decisions bias the solution of one minimisation, making departure-based screening decisions in subsequent minimisations potentially less reliable. Experiments conducted to examine the use of 're-screening' on its own show that overall the impact is largely neutral. The results in this paper show that any potential negative feedbacks are clearly outweighed by the other improvements that Continuous DA brings.

In the following two sections, we compare the performance of a continuous DA system against two non-continuous baselines. The final operational implementation of Continuous DA in the ECMWF system will be described in detail in Section 5.

## 3 | EXPERIMENT SET-UP

In Section 1 we described how NWP scientists optimise the configuration of their systems in such a way as to maximise the skill of the forecasts issued at a fixed dissemination time. In the following set of experiments, we vary just one parameter, the number of outer loops, and systematically compare the performance of a Continuous DA configuration against two more traditional baselines. The number of outer loops was chosen as the parameter to be varied as it has previously been shown to have a strong impact on forecast skill (Bonavita *et al.*, 2018) and because it fits rather well with the Continuous DA concept since additional outer loops provide more opportunities for late arriving observations to be introduced.

The primary aims of these experiments are:

1. To find evidence of any detrimental impacts caused by changing the number of observations in each successive outer loop;
2. To determine if a Continuous DA configuration can out-perform a non-continuous DA baseline for real-time NWP.

All experiments were carried out using ECMW's Integrated Forecasting System (IFS). The three different configurations are illustrated schematically in Figure 1, and defined as:

(a) *Continuous DA* – Newly arrived observations are introduced in each outer loop.
(b) *Realtime Baseline* – Non-continuous DA with the same data cut-off as the first outer loop of the Continuous DA configuration, that is, this configuration would be feasible in real-time forecasting applications.
(c) *Offline Baseline* – Non-continuous DA with the same data cut-off as the final outer loop of the Continuous DA configuration, that is, all observations are available in all outer loops.

Note from Figure 1 that the time at which the analysis is complete is much later in the Offline Baseline configuration than in the Realtime Baseline and Continuous DA configurations. The Offline Baseline is intended to be representative of applications that do not need to run in time-critical situations, such as reanalysis. In contrast, the Realtime Baseline is representative of a realistic non-continuous DA configuration that could be used in real time to deliver a forecast that meets the required dissemination schedule.

For each configuration, separate experiments were run with different numbers of outer loops (1, 2, 3, 4, 5, 6, 8 and 10 outer loops).

For the Offline Baseline configuration, the data cut-off time was set to the end of the assimilation window. In the Realtime Baseline configuration, the observation cut-off time was adjusted based on the number of outer loops, mimicking the fact that additional outer loops would require an earlier cut-off time in any real-time application to allow for the additional computation time. For each

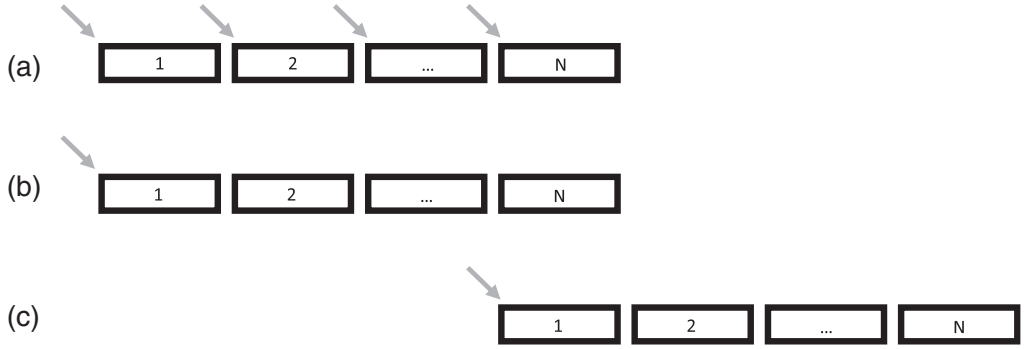


**FIGURE 1** Schematic of experiment configurations: (a) Continuous DA, (b) Realtime Baseline and (c) Offline Baseline. Boxes represent outer loops of 4D-Var. Time progresses from left to right. Arrows indicate when new observations are inserted into the DA system. All observations that have arrived up until that time are inserted so their number also increases from left to right

additional outer loop, the cut-off time is moved back by 20 min (approximately equal to the runtime of each outer-inner loop minimisation in operations). In the Continuous DA configuration, the observations were fed in at each outer loop using a cut-off time which was moved 20 min later for each outer loop. The cut-off time for the final outer loop of the Continuous DA configuration was the same as that for the Offline Baseline.

The experiments were run for a single season between 1 December 2017 and 28 February 2018 using a 12-hr assimilation window and 12-hr cycling of 4D-Var. The outer-loop model integrations used a triangular cubic octahedral grid (TCo) truncated at 639 spherical harmonic modes (*approx*18 km) while the inner loops used a reduced Gaussian linear grid (TL) truncated at 399 modes (*approx*50 km). In the Continuous DA configuration, quality control was re-run at each outer loop on all the observations using the most up-to-date trajectory fields.

The chosen configuration has been designed to systematically investigate the concepts used for Continuous DA, at affordable computational resources for the required experimentation. Note that this configuration differs from the operational configuration of 4D-Var at ECMWF which uses an "early-delivery" and "delayed cut-off" cycling scheme as well as higher spatial resolution and different resolutions for each inner loop. The details of the final operational configuration of Continuous DA implemented at ECMWF will be described in Section 5 along with the impacts in an operational setting.

## 4 | EVALUATION OF THE CONTINUOUS DA CONCEPT

### 4.1 | Observation usage and analysis convergence

The total number of observations ingested into the DA system and the number of assimilated observations in the Continuous DA and Offline Baseline configurations with four outer loops are shown in Table 1. As is typical in current DA systems, only a small fraction (around 7%) of the available observations are actively assimilated as the data are thinned to remove issues related to correlated observation errors and many observations are rejected by the quality control (for example to remove cloud-contaminated radiances in the clear-sky processing stream). As the Offline Baseline configuration only had observations introduced at the beginning of the assimilation, the number of active observations remained constant in each outer loop of 4D-Var. In contrast, in the Continuous DA configuration, the number of observations increased by around 3–4% in each outer loop.

The total number of available observations in the final outer loop was approximately equal to the number in the Offline Baseline (by design). The cut-off time for these two configurations was identical; the small 0.1% difference in total numbers was caused by a technical detail in the way that some of the pre-processing applications (which perform preliminary QC of the data prior to 4D-Var) handle data at the boundaries of the incoming data files. When the data are processed in multiple batches, the number of observations passed through to 4D-Var increases slightly compared to when the data are processed in a single batch. It is interesting to note that the number of assimilated observations was around 2% higher in the final outer loop of the Continuous DA configuration compared to the Offline Baseline, despite the number of incoming observations being approximately equal. Additional sensitivity tests have shown that this is primarily caused by the re-screening of observations in each outer loop against an increasingly accurate first guess (results not shown here).

One of the aims of the experiment was to find evidence of any issues caused by the introduction of new observations in successive outer loops. Table 2 shows the number of inner-loop iterations needed to reach the

**TABLE 1** Mean number of observations (in millions) in each outer loop in the Offline Baseline and Continuous DA configurations

| | Available observations | | Active observations | |
|---|---|---|---|---|
| Outer loop | Offline baseline | Continuous DA | Offline baseline | Continuous DA |
| 1 | 273.7 | 248.2 | 20.6 | 18.8 |
| 2 | — | 257.4 | 20.6 | 19.7 |
| 3 | — | 265.7 | 20.6 | 20.4 |
| 4 | — | 273.9 | 20.6 | 21.1 |

**TABLE 2** Mean number of iterations required to reach the specified convergence criteria for each outer loop in the Offline Baseline and Continuous DA configurations

| | Iterations | Iterations |
|---|---|---|
| Outer loop | Offline baseline | Continuous DA |
| 1 | 33.7 | 33.4 |
| 2 | 31.4 | 30.1 |
| 3 | 30.5 | 28.2 |
| 4 | 28.3 | 27.3 |

specified convergence criteria of the solver from the four outer-loop experiments. Compared to the Offline Baseline experiment, the Continuous DA configurations was able to converge in slightly fewer iterations, so the additional observations clearly do not cause a poorer convergence. The improved convergence appears to be related to the presence of fewer outliers after QC decisions are rerun using the latest trajectory.

## 4.2 | Analysis and forecast performance

Observation–background departure statistics are a useful indication of the quality of the short-range forecast. Figure 2 shows the standard deviation of the background departures for the three configurations as a function of the number of outer loops. Statistics were only calculated for the first 8 hr of the assimilation window to ensure consistent sampling was used for each configuration. For example, the cut-off time for the ten outer-loop Realtime Baseline configuration was 3 hr and 20 min before the end of the 12 hr assimilation window. The different panels show departures for different observation categories which are sensitive to different meteorological variables. For example, ATMS channel 7 is sensitive to mid-tropospheric temperature, while channel 22 is primarily sensitive to upper-tropospheric humidity.

In the Offline Baseline experiments, as the number of outer loops increases, the background departures were reduced, indicating an improved analysis and hence an improved background forecast provided to the next cycle. The improvement was dramatic between one and two outer loops. These improvements come from the re-linearisation around a new trajectory provided by running the full nonlinear model. However, the improvement beyond six outer loops was limited.

The departures from the Realtime Baseline experiments clearly show the degradation caused by moving the cut-off time progressively earlier to accommodate the extra outer loops within the schedule. The trade-off alluded to in Section 1 between the number of available observations and time available for performing DA calculations is apparent. The optimal fit was found with three outer loops. For four outer loops and above, the degradation caused by the loss of incoming observations outweighed the benefits provided from running with more outer loops. Indeed, the operational ECMWF system used three outer loops up to June 2019, in a non-continuous configuration.

Finally, the departure statistics from the Continuous DA experiments are very close to that of the Offline Baseline. This suggests that Continuous DA can achieve an analysis quality that is comparable to that from the Offline Baseline which is a configuration unfeasible for real-time NWP.

Further evidence that the Continuous DA configuration provides an analysis of comparable quality to that of the Offline Baseline is provided in Figures 3 and 4 which show the root mean square error at T+72 and T+144 hr, respectively, as a function of the number of outer loops for the three configurations. Once again, the Realtime Baseline forecast quality degrades when running with more than three or four outer loops as the loss of incoming observations outweigh the benefits of running with more outer-loop iterations. The errors from the Continuous DA experiments and the Offline Baseline experiments are statistically indistinguishable in the plots shown here.

These results are significant as they clearly demonstrate that, for real-time forecasting applications, a Continuous DA system can provide better quality analyses and forecasts than the best possible non-continuous DA configuration. This point is further demonstrated in Figures 5

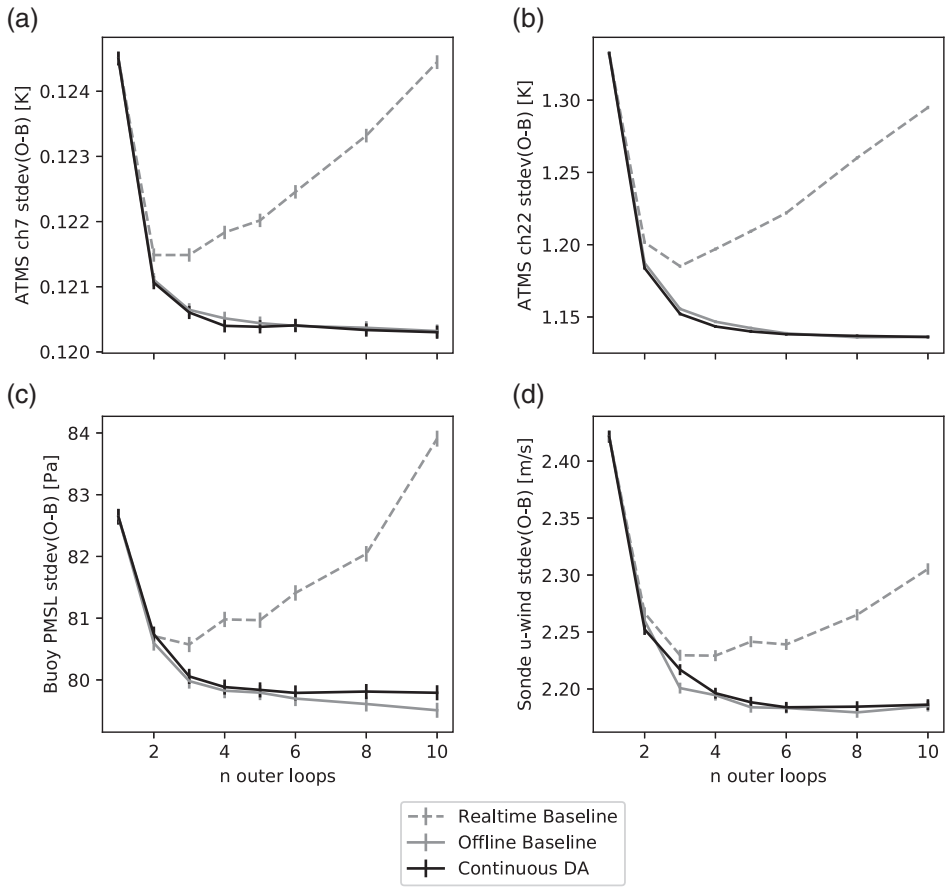


**FIGURE 2** Standard deviation of observation minus background departures for data selected from the screening trajectory as a function of number of outer loops for Realtime Baseline (grey dashed), Offline Baseline (grey solid) and Continuous DA (black) configurations, for (a) ATMS channel 7 sensitive to mid-tropospheric temperature, (b) ATMS channel 22 primarily sensitive to upper-tropospheric humidity, (c) mean sea level pressure measured from buoys, and (d) *u*-component of the wind at 500 hPa measured by radiosondes. Error bars indicate 95% confidence intervals. Only the first 8 hr in each assimilation window were considered in order to obtain a consistent sample for the three configurations

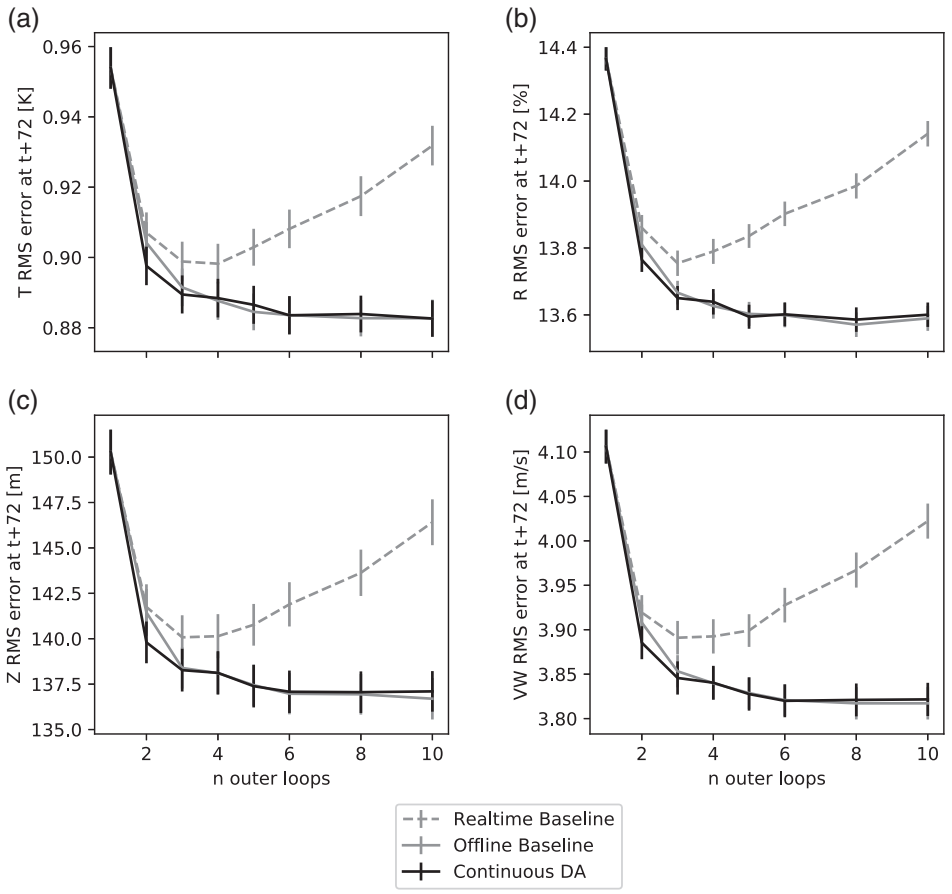


**FIGURE 3** Root mean square forecast error at T+72 hr for (a) temperature, (b) relative humidity, (c) geopotential height and (d) vector wind at 500 hPa for the three different configurations Realtime Baseline (grey dashed), Offline Baseline (grey solid) and Continuous DA (black). Verification against the operational analysis

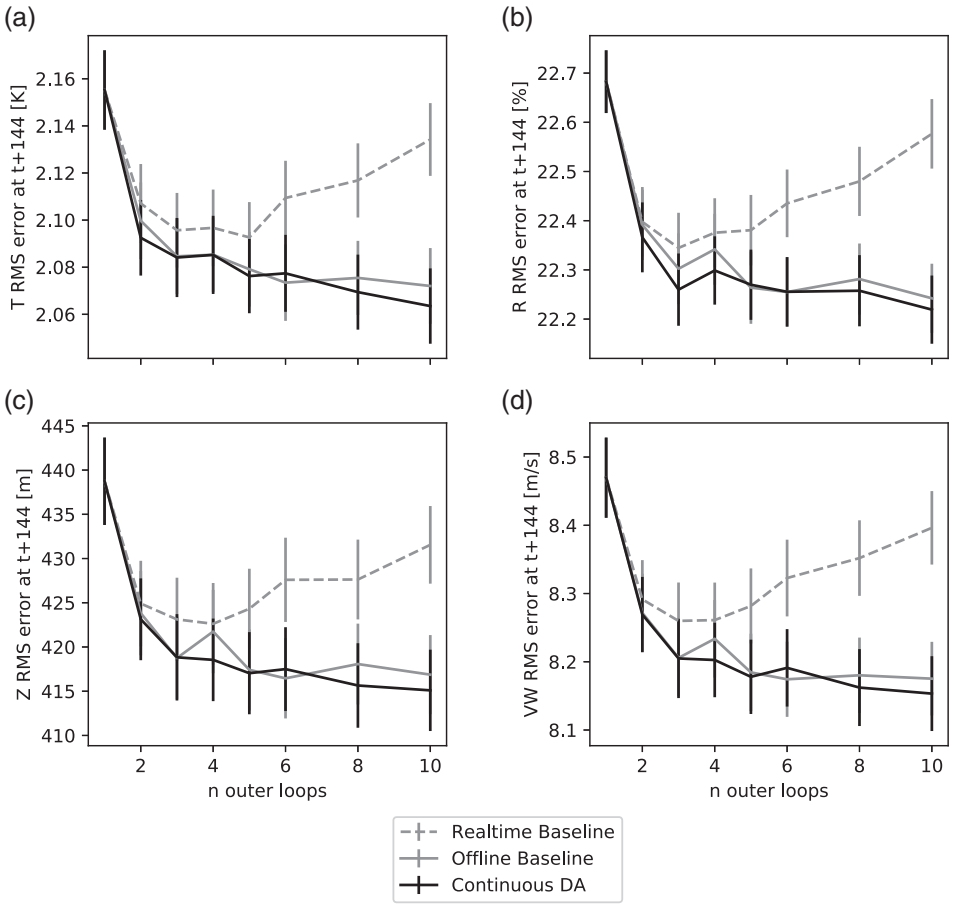


**FIGURE 4** As Figure 3, but results are shown at T+144 hr

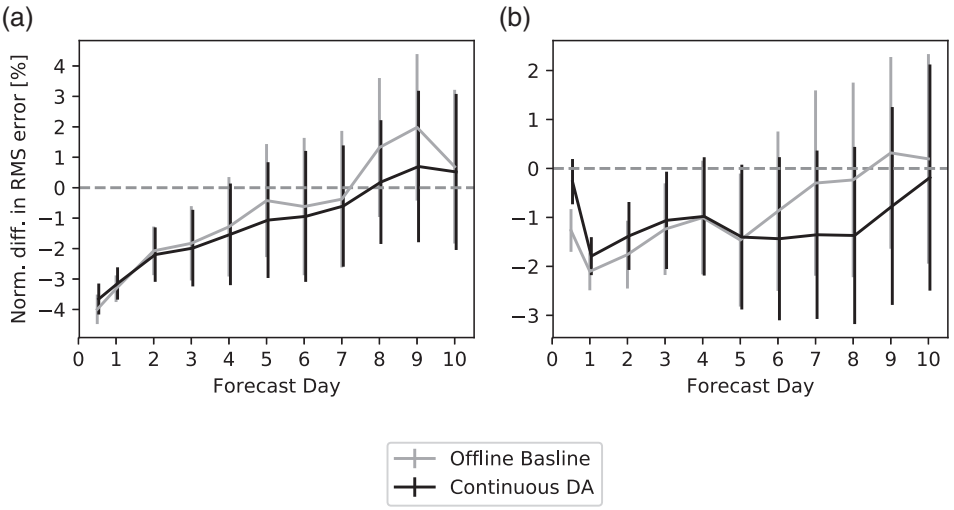


**FIGURE 5** Normalised change in root mean square error for geopotential height at 500 hPa as a function of lead time for the four outer loop Offline Baseline (grey solid) and four outer loop Continuous DA (black) using the three outer loop Realtime Baseline (grey dashed) as a control, for (a) Southern Hemisphere (20°–90°S) and (b) Northern Hemisphere (20°–90°N). Verification i9s against the operational analysis

and 6 which show a comparison of the normalised difference in root mean square error as a function of forecast lead time for the Continuous DA and the Offline Baseline experiments using four outer loops. The control for these two experiments is the Realtime Baseline experiment with three outer loops, that is, the best available non-continuous DA configuration that could be afforded for real time forecasting. Statistically significant forecast improvements are seen out to around day 3 or day 4.

A small apparent increase in the bias of temperature forecasts (of order 0.01 K) is found in the tropical mid-troposphere (Figure 7). This leads to a corresponding change bias in the geopotential height forecasts in

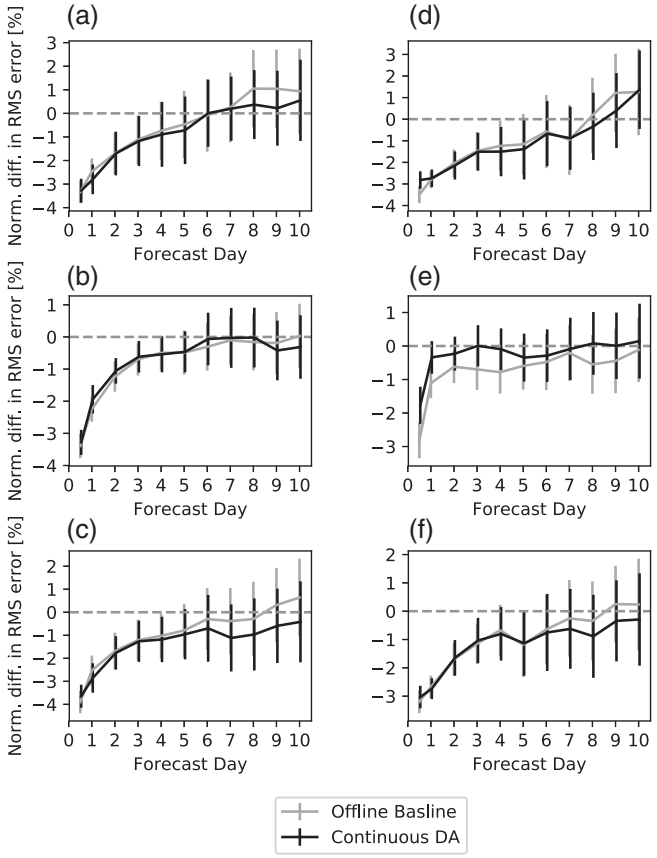


**FIGURE 6** As Figure 5, but scores for (a, b, c) vector wind at 200 hPa and (d, e, f) temperature at 500 hPa for (a, d) the Southern Hemisphere (SH, 20°–90°S), (b, e) Tropics (TR, 20°S–20°N) and (c, f) Northern Hemisphere (NH, 20°–90°N)

the tropical mid and upper troposphere. Sensitivity experiments (not shown) suggest that this may be related to the cloud screening of infrared satellite data, arising from feedback processes in which occasional incorrect cloud screening in one minimisation affects the screening in subsequent minimisations. However, this is considered a small effect that is clearly outweighed by the otherwise positive impacts of the Continuous DA configuration. Also, given the uncertainty in the reference analysis and the very small magnitude of the change in forecast bias, although we can be confident that the bias has changed, we cannot be sure that it is a degradation.

## 5 | OPERATIONAL IMPLEMENTATION OF CONTINUOUS DA

Continuous DA was introduced into operations at ECMWF as part of cycle 46r1 (Haiden *et al.*, 2019) on 11 June 2019. This section provides details of the configuration used which differs slightly from the simplified system used in the experiments in the previous two sections.

Since 2004, ECMWF has used a two-stage early-delivery and delayed cut-off suite (Haseler, 2004), as shown in Figure 8a. The main cycling analysis suite consists of consecutive 12-hr assimilation windows (two per day), with a relatively long observation cut-off time 4 hr after the end of each assimilation window. The relatively long cut-off time means that all observations with timeliness of up to 4 hr can be included in the assimilation. The analyses resulting from this cycle are also referred to as "delayed cut-off analyses". By providing the background fields for subsequent cycles, they are the backbone of the DA cycle, in the sense that they carry forward in time the observational information from past observations. To allow an earlier delivery of the operational forecasts, separate analyses are run which have a much tighter cut-off time and a more time-critical production schedule. The background information for these comes from short-range forecasts from the delayed cut-off analysis, and the window length is only 6 hr to help with the required scheduling. The resulting analyses are referred to as "early-delivery analyses". These early-delivery analyses are used to produce the main operational high-resolution forecast, and they are hence critical for achieving good forecast skill.

To benefit from late-arriving observations, Continuous DA was activated only in the more time-critical early-delivery cycles. There was little benefit in activating it in the delayed cut-off cycles as the number of available observations remains approximately constant in all outer loops. An additional benefit of this choice is that the small temperature bias discussed in Section 4 does not develop.

The operational schedule previously had a data cut-off time at 0400 UTC. However, the 6-hr assimilation window ended at 0300 UTC. This meant that some observations that had already arrived were not being assimilated as they were beyond the end of the assimilation window. Therefore, the length of the assimilation window in the early-delivery cycles was increased from 6 to 8 hr to ensure that any observation that had arrived could be assimilated.

Finally, to take advantage of the fact that Continuous DA allows us to benefit from more expensive DA configurations, the number of outer loops was increased from three to four while still delivering forecasts at the same time as before. The new configuration is shown in Figure 8b.

The resolution of the current operational assimilation is TCo1279 outer loops (approximately 9 km) with three inner loops at resolutions TL255, TL319 and TL399 (approximately 78, 63 and 50 km respectively). The additional fourth outer loop in Continuous DA is also at TL399/50 km.

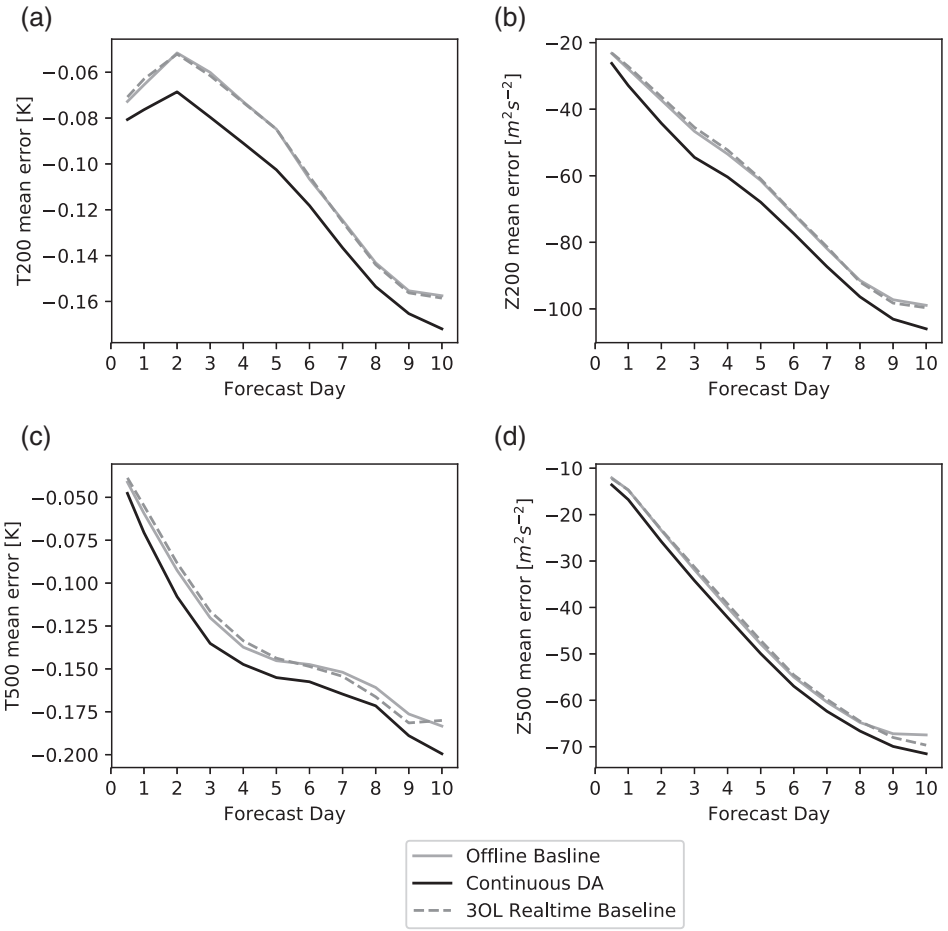


**FIGURE 7** Mean error as a function of forecast lead time for (a, c) temperature (K) and (b, d) geopotential height (gpm) at (a, b) 200 hPa and (c, d) 500 hPa for the four outer loop Continuous DA configuration (black), the four outer loop Offline Baseline (grey solid) and the three outer loop Realtime Baseline (grey dashed)

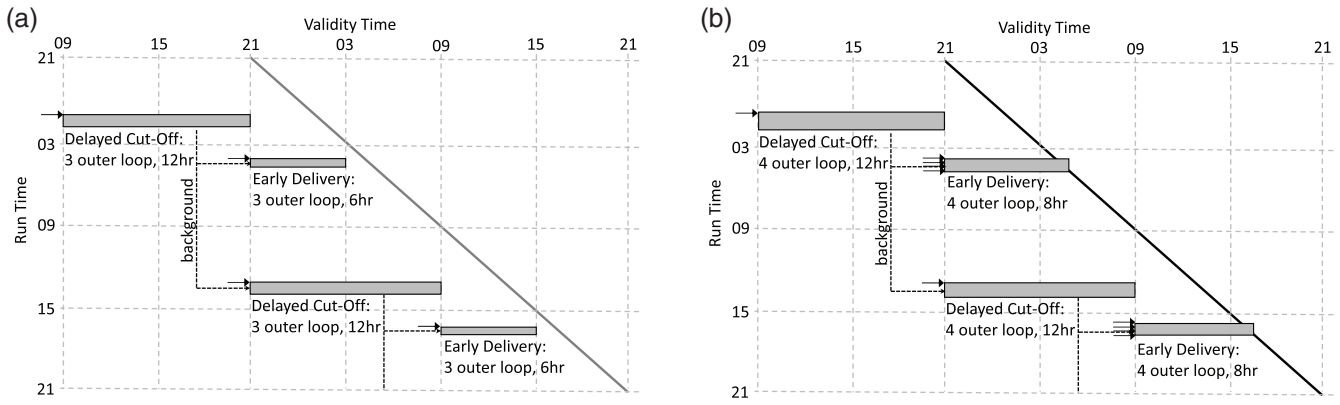


**FIGURE 8** Schematic representation of the two-stage early-delivery and delayed cut-off suite used operationally at ECMWF for (a) the previous operational configuration and (b) the new Continuous DA configuration. The *x*-axis represents the observation time and the *y*-axis the time at which the assimilation calculations are run. Solid arrows indicate when observations are inserted into the system. Dashed arrows indicate the source of the background used in each cycle. All times are UTC

The addition of new observations in later outer loops (effectively a later cut-off time), combined with the extended assimilation window, allowed an extra 85 min of observations to be assimilated in the early-delivery cycles of 4D-Var.

Table 3 compares the number of assimilated observations in each outer loop of 4D-Var between the (previously operational) three outer loop 6-hr window control and the new Continuous DA four outer loop, 8-hr window configuration. Overall the number of assimilated

| Outer loop | Active observations Control – 3 outer loops | Active observations Continuous DA – 4 outer loops |
|---|---|---|
| 1 | 10.6 | 10.5 |
| 2 | 10.6 | 11.1 |
| 3 | 10.6 | 11.4 |
| 4 | — | 11.7 |

**TABLE 3** Mean number of assimilated observations (millions) in each outer loop of 4D-Var for the three outer loop 6-hr window control experiment and the four outer loop 8-hr window Continuous DA experiment

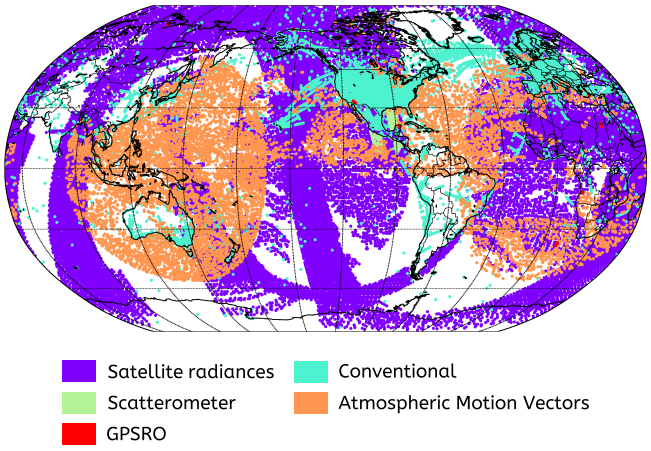


**FIGURE 9** Example of extra observations assimilated in a single Continuous DA cycle compared to the control. Different colours are used to indicate different observation types

observations was increased by around 10%. The majority of these additional observations were very recent observations near the end of the assimilation window. At the time that the analysis is complete, the most recent assimilated observation is only 35 min old, compared to 120 min in the control. Figure 9 shows the coverage of the additional observations, illustrating that these consist of a wide range of observation types, including aircraft data, observations from geostationary satellites, and parts of orbits from polar-orbiting satellites. The availability of observations with excellent timeliness obviously determines how many additional observations can be included, and the new Continuous DA configuration will benefit even more from any improvements in observation timeliness.

The improvement in forecast scores for geopotential height at 500 hPa are shown in Figure 10. The root mean squared errors were typically reduced by around 2% at day 3 for most variables.

The relative contributions of the different aspects of the Continuous DA configuration are shown in Figure 11. The impact of adding new observations in each outer loop without the other changes was found to be relatively modest (black dashed line). This is because with a 6-hr assimilation window, many of these new observations would be beyond the end of the window and not assimilated. Similarly, using the 8-hr assimilation window on its own (grey line) had a small impact as very few observations were available in this extension. It was only by combining the Continuous DA with the 8 hr window (grey dashed line) that the full benefit of the extra observations could be realised. Finally, the addition of an extra outer loop (black line) added a further improvement on top of the previous results. In a non-continuous DA configuration, the addition of a fourth outer loop would have been possible only by moving the cut-off time earlier which would have led to a loss of observations and a corresponding degradation of the forecast scores (as was demonstrated in Section 4).

## 6 | DISCUSSION

### 6.1 | Observation timeliness

The success of Continuous DA is critically linked to the timeliness of the available observations. This is particularly crucial to achieve good observational coverage towards the end of the early-delivery analysis cycle. Figure 12 shows an example of the observation coverage in the last populated half-hour of an early-delivery analysis cycle. The coverage here is relatively sparse, as a stringent timeliness of 25 min or better is required. Observations that achieve such timeliness are typically aircraft data, data from the GMI (Global Precipitation Measurement Microwave Imager) satellite, or satellite data from the DBNet (Direct Broadcast Network) initiative coordinated by WMO (WMO, 2017). Timeliness of satellite data is limited by the opportunities for data downlink, which for low-earth-orbit satellites is often confined to one or two polar reception stations and requires an overpass of the satellite over the reception station. For GMI, satellite-to-satellite relay systems are used instead to allow faster downlinks to a reception station without having to wait for an overpass. In contrast, the DBNet initiative uses local reception stations to complement the global processing, aiming at a timeliness of 20 min. The benefit of such an initiative in the Continuous DA context is highlighted in Figure 13. DBNet observations, which in general arrive

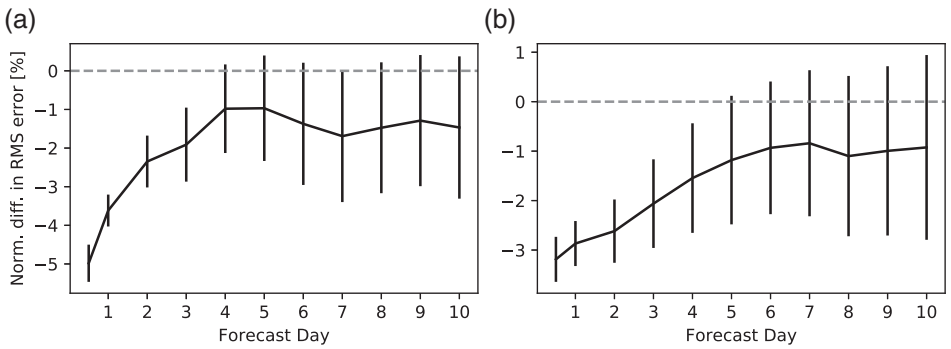


**FIGURE 10** Normalised change in root mean square error for geopotential height at 500 hPa as a function of lead time for (a) the Southern Hemisphere and (b) the Northern Hemisphere for the four outer loop 8-hr window Continuous DA configuration compared to the three outer loop 6-hr window control. Results are shown from two 3-month periods of testing; 1 December 2016–28 February 2017 and 1 June 2017–31 August 2017

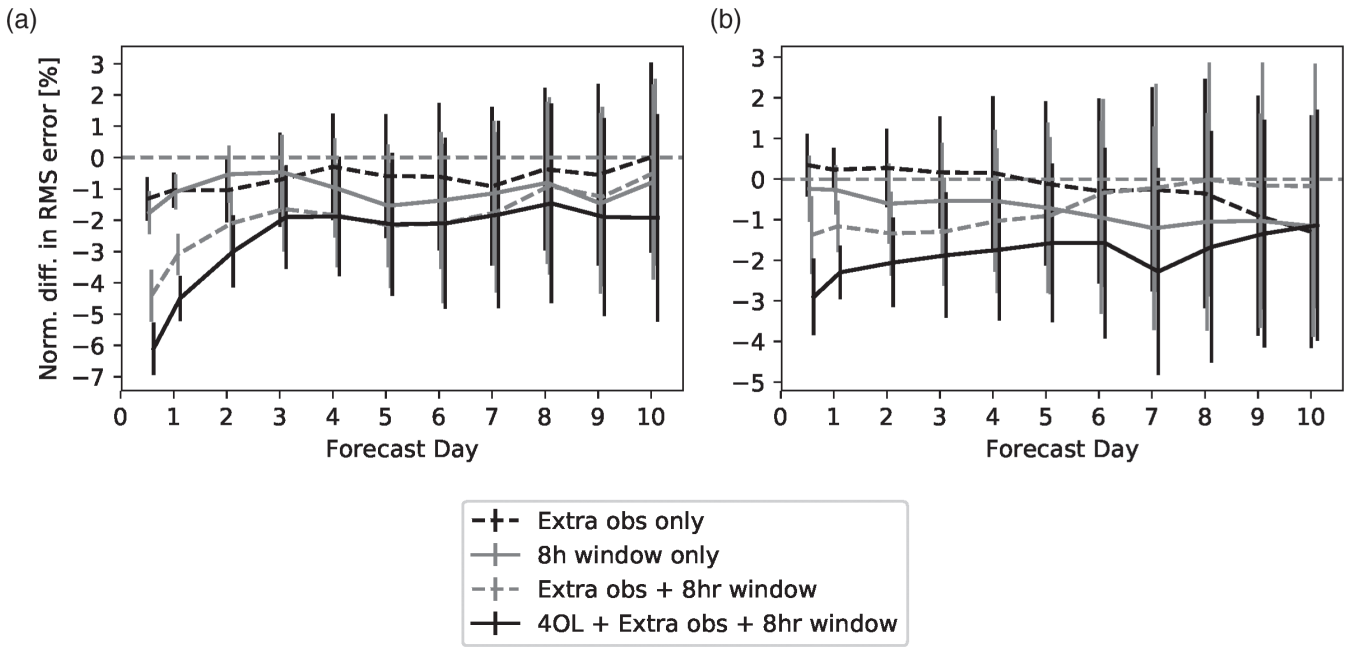


**FIGURE 11** Change in root mean square error in geopotential height at 500 hPa for the (a) Southern Hemisphere and (b) Northern Hemisphere with respect to a three outer loop non-continuous DA control experiment. Results are shown from experiments with extra observations in each outer loop (black dashed), 8 hr assimilation window (grey solid), extra observations and 8 hr assimilation window (grey dashed) and three outer loops with extra observations and 8 hr assimilation window (black solid). The experiment was run for a single season from 1 June 2017 to 31 August 2017

within 30 min of observation time, are crucial to populate the final hours of the assimilation window as many of the globally processed observations do not arrive before the cut-off time, and this is even more the case in the Continuous DA configuration.

Observations towards the end of the assimilation window play a particularly crucial role in 4D-Var assimilation (McNally, 2019). Not only are they the most up-to-date information for the subsequent forecast, but they are also the observations most exposed to growth of errors in the background during the assimilation window, and they are crucial in providing dynamical information through the 4D-Var tracing mechanism. Continuous DA ensures the best possible coverage for observations towards the end of the assimilation window, and this is a critical factor in the success of Continuous DA. At the same time, with Continuous DA there is an increased incentive to further improve the timeliness for many observations, as any improvement will directly lead to increased data usage.

## 6.2 | Relationship with Rapid Update Cycling DA

The work presented in this paper has clear parallels with the field of Rapid Update Cycling (RUC) DA. For regional

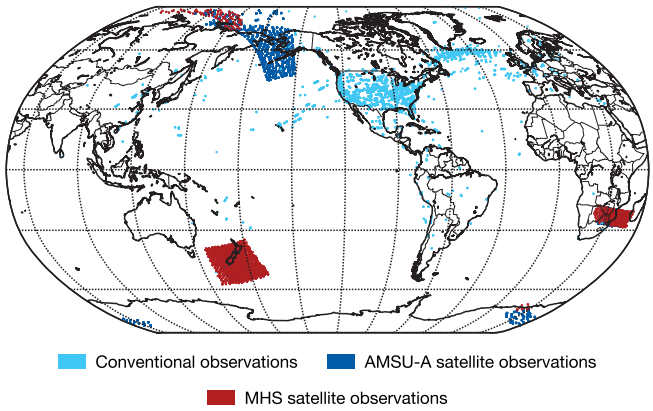


**FIGURE 12** Example of assimilated observations made between 0400 and 0425 UTC with a data cut-off time of 0425 UTC

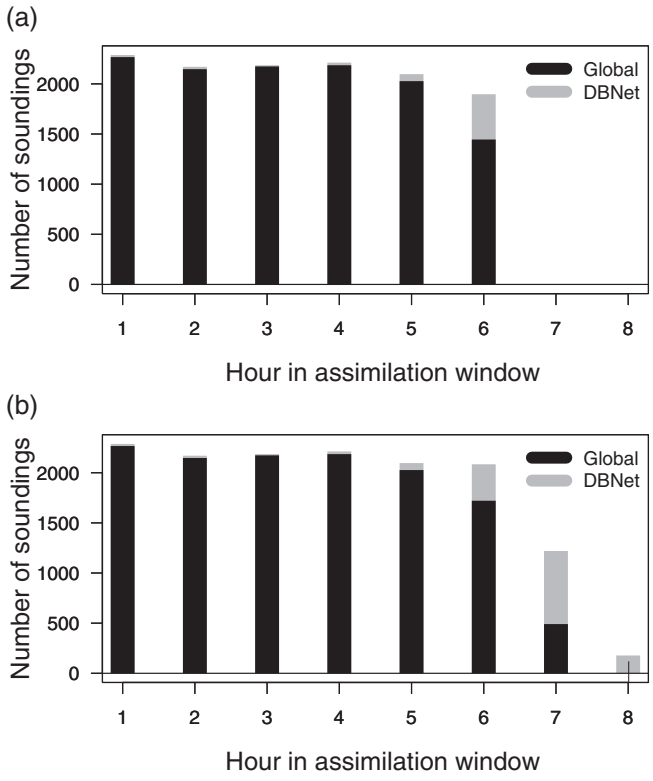


**FIGURE 13** Number of assimilated NOAA-19 AMSU-A soundings in each hour of the early-delivery assimilation window for (a) the control and (b) Continuous DA

NWP systems which are intended to forecast sub-synoptic and mesoscale features that have correspondingly short predictability time-scales, there is a clear benefit to generate analyses and forecasts frequently based upon the latest observations. Typically, these use hourly cycling data assimilation systems with a 1-hr assimilation window (Benjamin *et al.*, 1994; Benjamin *et al.*, 2016; Milan *et al.*, 2019).

Payne (2017) discussed how a rapid update cycle could be adapted for use in global NWP systems using both overlapping and non-overlapping assimilation window configurations. The paper highlighted the benefits from having more frequently updated analyses which can provide more up-to-date and accurate forecasts in the hours between the standard synoptic analyses.

In this paper, we have not considered more frequent analyses and forecasts, but instead shown how the analysis and forecast quality at the standard synoptic times can be improved by feeding in recently arrived observations during the DA calculations. Having said that, the Continuous DA approach does lend itself well to use within a RUC system and this possibility could be explored further.

Similarly, the question of whether Continuous DA would be beneficial for limited-area NWP would be worthy of further study. The time constraints in regional NWP are even more pressing than those in global medium-range NWP. On the one hand, DA at the mesoscale and convective scale can be highly nonlinear and the analysis quality may benefit from using multiple outer loops which Continuous DA might make feasible. On the other hand, the cut-off times in these systems are already extremely tight with little scope for assimilating additional late-arriving observations. It might be instructive to start by studying the number of late-arriving observations assuming that only the last minimization was time critical.

## 6.3 | Scalability

The results in this paper have demonstrated that a continuous DA framework can provide an improved quality analysis by starting the assimilation calculations earlier and using more late-arriving data. In addition, continuous DA may also bring computational benefits.

As the rate of increase of computer processor speed slows, operational centres are forced to use an ever increasing number of compute nodes to run more expensive and accurate NWP systems. Concerns about the power required to run these systems become increasingly prominent. Consequently, the scalability of the 4D-Var algorithm has received a great deal of attention in recent years (e.g., Fisher and Gürol, 2017; Bousserez *et al.*, 2020). If an application scales perfectly then the runtime is expected to be reduced in inverse proportion to the number of compute nodes upon which it is run. However, the 4D-Var algorithm does not scale perfectly. As an example, the strong scaling characteristics of the 4D-Var system used in this paper are shown in Figure 14. Although the time to solution is faster when running on more nodes, the computational efficiency gets worse.

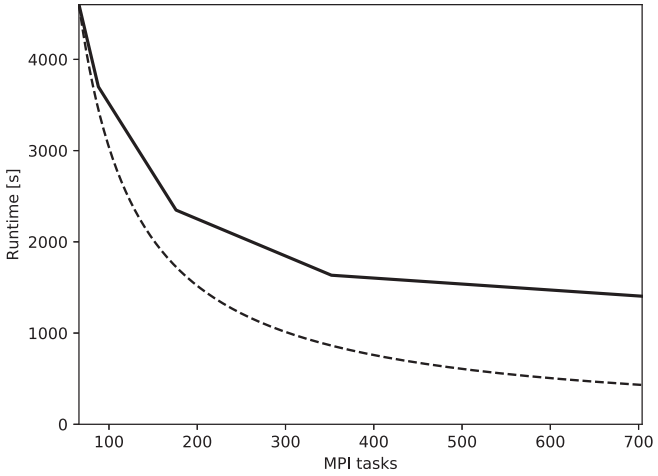


**FIGURE 14** Runtime of 4D-Var (black solid) as a function of the number of MPI tasks used compared with the theoretical runtime that could be achieved by an application with perfect strong scaling characteristics (black dashed)

The Continuous DA framework has the potential to alleviate some of these concerns. For example, a more efficient usage of finite compute resources would be to start the assimilation calculations earlier and run on fewer nodes. Previously, starting earlier would have led to a degradation in the quality of the analysis as the earlier cut-off time would have led to fewer observations being received. However, within a continuous assimilation framework, this becomes feasible. One possible configuration would be to run all but the final outer loop on a small number of nodes with a final time-critical minimisation running on a larger number of nodes in order to maintain the existing cut-off time. In the example in Figure 14, the efficiency is improved by a factor of 3 by running on 100 MPI tasks instead of running on 700 MPI tasks. It also has the additional benefit of reducing peaks in the workload on the high-performance computing facilities as the computations are distributed more evenly in time.

Taken to its logical extreme, one can envisage a continuously cycling DA system running on a relatively small number of nodes. As new observations arrive, they are introduced into the next outer loop, helping to further refine the analysis. While the simplicity of this concept is enticing, the results in this paper indicate that the benefits of running with more than six outer loops may be limited. Moreover, making the operational system more expensive has drawbacks in terms of the cost required to test changes prior to implementation, and on the ability to reproduce operational results in a research environment. On the other hand, having a frequently updated analysis has benefits in terms of the resilience of the system to failures as a reasonably up-to-date analysis will always be available from which a forecast can be initialised when required.

### 6.4 | Limitations of this study

Continuous DA allows real-time forecasting applications to benefit from more expensive DA configurations. This study has explored just one way in which this could be exploited: increasing the number of outer loops. Further work could be conducted to vary other parameters (e.g., higher DA resolution, shorter timesteps or stricter convergence criteria) to see if further improvements in the forecasts are possible within this framework.

In addition, due to limitations in the existing configuration of the system, the results in this paper unfairly penalised the Continuous DA configuration as the surface analysis is tied to the first outer loop of 4D-Var. This means that the number of observations available to the surface analysis is reduced as the cut-off time is moved earlier for experiments with more outer loops. There is no fundamental reason why the surface analysis cannot be made to run later in the assimilation and have access to more observations.

## 7 | CONCLUDING REMARKS

Despite being proposed in the mid-1990s, the continuous data assimilation concept has had limited uptake from the NWP community. Motivated by the aim of making use of more late-window observations, ECMWF has explored this concept further.

The Continuous DA framework exploits the fact that most of the observations arrive long before the cut-off time. It allows real-time forecasting applications to benefit from more expensive and accurate DA configurations while simultaneously using more late-arriving observations, without delaying the time at which the analysis is ready for use by the forecast model.

This paper has presented results which indicate that the analysis produced in a continuous assimilation framework closely approximates the solution provided by the (unaffordable) Offline Baseline configuration which had access to all the observations in each outer loop. For real-time forecasting applications, continuous DA allows configurations which clearly outperform the best available non-continuous DA configuration. No issues were found related to the ability of the minimisation algorithm to converge when increasing the number of observations in each outer loop. A small bias in the analysed temperature

in the Tropics merits further attention, but is minor when compared to the improvements that Continuous DA provides.

In the final operational configuration, around 10% more observations are assimilated. The majority of these are valuable end-of-window observations. By the time that the analysis is complete, the most recent assimilated observation is now only 35 min old, compared to 120 min previously. The time at which forecasts are issued remains unchanged, but the root mean square error of the medium-range forecasts is reduced by around 2–3%. The benefits were shown to have come from both the use of more recent observations, and from the addition of a fourth outer loop. The results also highlight the critical importance of observation timeliness for real-time NWP applications.

Further work is being undertaken to explore new configurations made possible by Continuous DA which may help to further improve forecast scores as well as the resilience and computational efficiency of the operational DA system.

Continuous DA went into operations at ECMWF as part of cycle 46r1 on 11 June 2019.

## ACKNOWLEDGEMENTS

The results presented in this paper have relied on the work of many people at ECMWF. In particular, the successful operational implementation in cycle 46r1 would not have been possible without the considerable efforts of Axel Bonet, Anna Mueller-Quintino, John Hodkinson, Cristiano Zanna and Enrico Fucile. Useful feedback on various aspects of the performance of the new configuration were provided by Cristina Lupu, Reima Eresmaa and Simon Lang. We also wish to thank Stephen English who provided encouragement to pursue this work from its early stages all the way through to operations. Finally, we would like to thank two anonymous reviewers whose comments helped improve this manuscript.

## ORCID

*P. Lean* https://orcid.org/0000-0002-3662-5382

## REFERENCES

Bauer, P., Quintino, T., Wedi, N., Bonanni, A., Chrust, M., Deconinck, W., Diamantakis, M., Düben, P., English, S., Flemming, J., Gillies, P., Hadade, I., Hawkes, J., Hawkins, M., Iffrig, O., Kuehnlein, C., Lange, M., Lean, P., Marsden, O., Mueller, A., Saarinen, S., Sarmany, D., Sleigh, M., Smart, S., Smolarkiewicz, P., Thiemert, D., Tumolo, G., Weihrauch, C. and Zanna, C. (2020). The ECMWF Scalability Programme: Progress and Plans. Technical Memorandum 857, ECMWF, Reading, UK.

Benjamin, S.G., Brundage, K.J., Miller, P.A., Smith, T.L., Grell, G.A., Kim, D., Brown, J.M. and Schlatter, T.W. (1994). The Rapid Update Cycle at NMC. pp. 566–568 in preprints for 10th Conference on Numerical Weather Prediction, Portland, OR. American Meteorological Society, Boston, MA.

Benjamin, S.G., Weygandt, S.S., Brown, J.M., Hu, M., Alexander, C.R., Smirnova, T.G., Olson, J.B., James, E.P., Dowell, D.C., Grell, G.A., Lin, H., Peckham, S.E., Smith, T.L., Moninger, W.R., Kenyon, J.S. and Manikin, G.S. (2016) A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Monthly Weather Review*, 144, 1669–1694. https://doi.org/10.1175/MWR-D-15-0242.1

Bonavita, M., Hólm, E.V., Isaksen, L. and Fisher, M. (2016) The evolution of the ECMWF hybrid data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 142, 287–303. https://doi.org/10.1002/qj.2652

Bonavita, M., Lean, P. and Hólm, E.V. (2018) Nonlinear effects in 4D-Var. *Nonlinear Processes in Geophysics*, 25, 713–729. https://doi.org/10.5194/npg-25-713-2018

Bousserez, N., Guerrette, J.J. and Henze, D.K. (2020) Enhanced parallelization of the incremental 4D-Var data assimilation algorithm using the Randomized Incremental Optimal Technique. *Quarterly Journal of the Royal Meteorological Society*, 146, 1351–1371. https://doi.org/10.1002/qj.3740

Böker, J. (2010) On variational data assimilation in continuous time. *Quarterly Journal of the Royal Meteorological Society*, 136, 1906–1919. https://doi.org/10.1002/qj.695

Choi, Y., Lim, G.-H. and Lee, D.-K. (2013) Radar radial wind data assimilation using the time-incremental 4D-Var method implemented to the WRFDA system. *Tellus*, 65A, 1–17. https://doi.org/10.3402/tellusa.v65i0.19677

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. (1994) A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120, 1367–1387. https://doi.org/10.1002/qj.49712051912

Derber, J.C. (1989) A variational continuous assimilation technique. *Monthly Weather Review*, 117, 2437–2446. https://doi.org/10.1175/1520-0493(1989)117<2437:AVCAT>2.0.CO;2

Fisher, M. (1998). Minimization algorithms for variational data assimilation, pp. 364–385 in Proceedings of Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling. ECMWF, Reading, UK.

Fisher, M. and Gürol, S. (2017) Parallelization in the time dimension of four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 143, 1136–1147. https://doi.org/10.1002/qj.2997

Gauthier, P. (1992) Chaos and quadri-dimensional data assimilation: a study based on the Lorenz model. *Tellus*, 44A, 2–17. https://doi.org/10.3402/tellusa.v44i1.14938

Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S. and Morneau, J. (2007) Extension of 3DVAR to 4DVAR: implementation of 4DVAR at the Meteorological Service of Canada. *Monthly Weather Review*, 135, 2339–2354. https://doi.org/10.1175/MWR3394.1

Haiden, T., Janousek, M., Vitart, F., Ferranti, L. and Prates, F. (2019). Evaluation of ECMWF forecasts, including the 2019 upgrade. Technical Memorandum 853, ECMWF, Reading, UK.

Haseler, J. (2004). Early-delivery suite. Technical Memorandum 454, ECMWF, Reading, UK.

Järvinen, H., Thépaut, J.-N. and Courtier, P. (1995). Quasi-continuous variational data assimilation. Technical Memorandum 210, ECMWF, Reading, UK.

Järvinen, H., Thépaut, J.-N. and Courtier, P. (1996) Quasi-continuous variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 122, 515–534. https://doi.org/10.1002/qj.49712253011

Lawless, A.S., Gratton, S. and Nichols, N.K. (2005) Approximate iterative methods for variational data assimilation. *Numerical Metods in Fluids*, 47, 1129–1135. https://doi.org/10.1002/fld.851

Lorenc, A.C. (1981) A global three-dimensional multivariate statistical interpolation scheme. *Monthly Weather Review*, 109, 701–721. https://doi.org/10.1175/1520-0493(1981)109<0701:AGTDMS>2.0.CO;2

Lorenz, E.N. (1963) Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20, 130–141. https://doi.org/10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2

McNally, A.P. (2019) On the sensitivity of a 4D-Var analysis system to satellite observations located at different times within the assimilation window. *Quarterly Journal of the Royal Meteorological Society*, 145, 2806–2816. https://doi.org/10.1002/qj.3596

Milan, M., Macpherson, B., Tubbs, R., Dow, G., Inverarity, G., Mittermaier, M., Halloran, G., Kelly, G., Li, D., Maycock, A., Payne, T., Piccolo, C., Stewart, L. and Wlasak, M. (2019) Hourly 4D-Var in the Met Office UKV operational forecast model. *Quarterly Journal of the Royal Meteorological Society*, 1–18. https://doi.org/10.1002/qj.3737

Payne, T.J. (2017) Rapid update cycling with delayed observations. *Tellus*, 69A, 1–17. https://doi.org/10.1080/16000870.2017.1409061

Pires, C., Vautard, R. and Talagrand, O. (1996) On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus*, 48A, 96–121. https://doi.org/10.3402/tellusa.v48i1.11634

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. and Simmons, A. (2000) The ECMWF operational implementation of four-dimensional variational assimilation. I: experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126, 1143–1170. https://doi.org/10.1002/qj.49712656415

Veersé, F. and Thépaut, J.-N. (1998) Multiple-truncation incremental approach for four-dimensional variational data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 124, 1889–1908. https://doi.org/10.1002/qj.49712455006

WMO (2017). Guide to the Direct Broadcast Network for near-real-time relay of Low Earth Orbit Satellite Data. Attachment to the Guide to the WMO Information System (WMO-No. 1061) 1185, WMO, Geneva, Switzerland.

**How to cite this article:** Lean P, Hólm EV, Bonavita M, Bormann N, McNally AP, Järvinen H. Continuous data assimilation for global numerical weather prediction. *Q.J.R. Meteorol. Soc.*. 2020;1–16. https://doi.org/10.1002/qj.3917