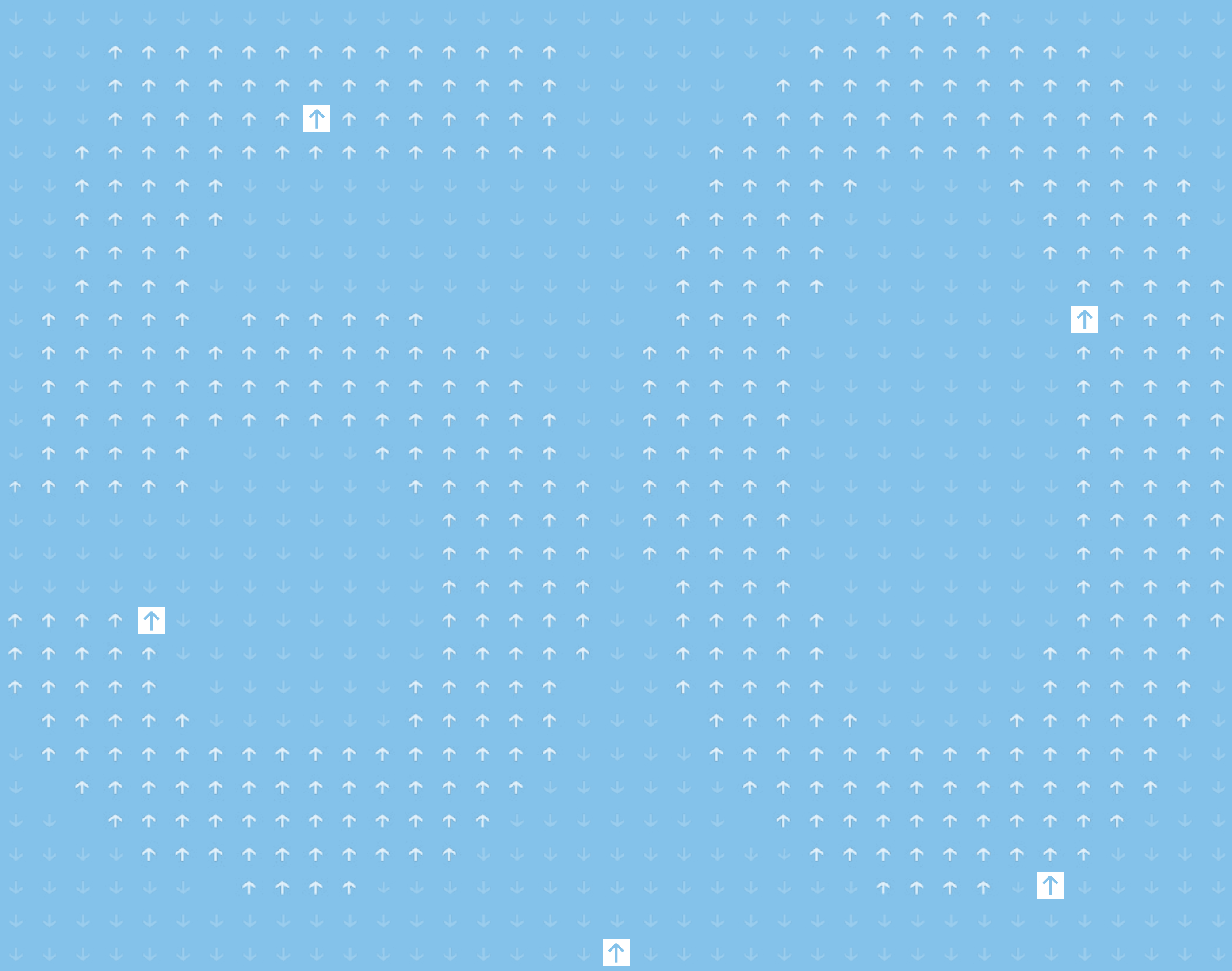
Fifty years of data assimilation at ECMWF



ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION

ECMWF | 50

PUBLICATION 02

Fifty years of data assimilation at ECMWF

ABSTRACT

Over the past 50 years, data assimilation (DA) has been a cornerstone of ECMWF's success in numerical weather prediction (NWP), enabling significant advancements in forecast accuracy and extending prediction lead times. Through pioneering research and strong collaborations with its Member States, European meteorological services, and space agencies such as the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency (ESA), ECMWF has led the operational adoption of variational DA techniques, transitioning from early methods to the groundbreaking implementation of four-dimensional variational data assimilation (4D-Var). This transformation has allowed the direct assimilation of satellite radiances, unlocking the full potential of spaceborne observations and revolutionising modern data assimilation and forecasting.

Further developments, such as the introduction of the Ensemble of Data Assimilations (EDA), have provided a more robust representation of flow-dependent errors, improving uncertainty quantification in initial conditions. ECMWF continues to drive innovation through the evolution of coupled DA, integrating atmospheric, ocean, and land observations to enhance Earth system modelling. The "all-sky, all-surface" approach has further optimised satellite data assimilation in complex conditions, ensuring the best possible use of European and international investments in space programmes. These advancements are made possible through sustained collaboration with national meteorological services, research institutions, and operational programmes such as the European Union's Copernicus programme.

Beyond weather forecasting, ECMWF's world-class DA infrastructure underpins the production of high-impact climate reanalysis datasets, such as ERA5, which have become essential for climate monitoring and research. Looking ahead, artificial intelligence (AI) and machine learning (ML) are set to reshape the DA landscape, offering unprecedented opportunities to enhance observation processing, error correction, and computational efficiency. As ECMWF prepares for future observing systems and AI-driven forecasting, its commitment to scientific excellence, strong partnerships, and collaboration with its Member and Co-operating States ensures that it remains at the forefront of meteorology and climate monitoring science.

INTRODUCTION →

The European Centre for Medium-Range Weather Forecasts (ECMWF) stands as a global leader in numerical weather prediction (NWP), renowned for its pioneering advancements and outstanding forecasting capabilities. Since its establishment 50 years ago, ECMWF has been at the forefront of meteorological research, providing critical weather forecasts that inform decisions for its Member and Co-operating States across a multitude of sectors and applications. Numerical weather prediction, which involves the use of physics-based models to simulate the Earth's atmosphere, has revolutionised our ability to predict weather, understand atmospheric dynamics and provide quantification of uncertainties about forecast products. At the heart of NWP lies data assimilation – a technique that seamlessly integrates observations from diverse sources into NWP models, ensuring that forecasts are optimally initialised to ensure best possible accuracy.

By incorporating observations from satellites, meteorological stations, buoys, aircraft and other platforms, data assimilation enables models to produce accurate initial conditions – a prerequisite for reliable weather forecasts. Throughout the last 50 years, ECMWF's innovations in this domain have set the global standard, pioneering significant improvements in forecast skill and extending the lead time of high-confidence predictions. This cornerstone of NWP is an example of the Centre's success in achieving scientific excellence and operational reliability (Rabier et al., 2000; Bauer et al., 2015).

The success of ECMWF, however, has not been achieved in isolation. Partnerships with national meteorological services, space agencies, and research institutions around the world have played a crucial role in establishing ECMWF as a world leader of weather forecasting. These collaborations have facilitated the sharing of expertise and ideas, resources, and observational data, creating a synergistic environment that has driven innovation. Through initiatives such as the European Union's Copernicus Earth observation programme and Destination Earth (DestinE), and the World Meteorological Organization's (WMO's) collaborative frameworks, ECMWF has greatly benefited from the power of international cooperation to push the boundaries of what is possible in NWP.

In this paper, we describe the fundamental role of data assimilation in ECMWF's success, its evolution alongside advancements in NWP, and the critical importance of partnerships in shaping ECMWF's trajectory as a global leader in meteorology.

HISTORICAL OVERVIEW OF DATA ASSIMILATION AT ECMWF →

ECMWF was founded in 1975 with the primary mission to produce ten-day weather forecasts using state-of-the-art NWP systems. From the outset, the Centre recognised the importance of data assimilation as the foundation for reliable model initialisation. In its early years, ECMWF implemented a basic three-dimensional optimal interpolation (OI) scheme (Lorenc, 1981), which provided a systematic approach to incorporating observational data into its models. This method, based on statistical interpolation, balanced observational data with prior forecast information (background data), weighting both sources of information according to their relative errors and spatial correlations and minimising the expected error variance of the resulting initial state to improve forecast accuracy. While OI was a significant step forward in data assimilation during its time, it had several shortcomings that limited its effectiveness compared to more advanced techniques like variational data assimilation. Significant limitations of the OI algorithm implemented at the time include the use of static, predefined error covariance matrices, the suboptimal use of observations not linearly related to the analysis variables (e.g. satellite radiances), the lack of model constraints in the analysis procedure, and the local nature of the solver, which can lead to numerical artefacts (e.g. discontinuities) in the resulting analysis fields.

> “ DURING THE 1990s, ECMWF PIONEERED THE USE OF VARIATIONAL DATA ASSIMILATION TECHNIQUES, SHIFTING FROM OPTIMAL INTERPOLATION WHICH HAD SIGNIFICANT LIMITATIONS.”

During the 1990s, ECMWF pioneered the use of variational data assimilation (Var) techniques, shifting from OI to more sophisticated approaches, in order to address the limitations described above. This effort was greatly facilitated by a proactive collaboration between ECMWF and Météo-France on what was called the Integrated Forecasting System (IFS)/ARPEGE project, which mobilised significant resources on both sides to address this new revolutionary (at the time) framework (Pailleux et al., 2014). Three-dimensional variational data assimilation (3D-Var) was operationally implemented at ECMWF in 1996 (following an earlier implementation at the US National Centers for Environmental Prediction (NCEP) in 1995). Worth noting is that prior to this implementation, and as described in Eyre et al. (2020), assimilating satellite observations as low-vertical-resolution retrieved profiles had at best a neutral impact in most NWP centres, exhibiting difficulties in specifying appropriate error statistics for the retrievals, contaminated by their climatological background.

An important intermediate step towards direct radiance assimilation was the assimilation of 1D-Var retrievals which used NWP short-range forecasts as background information. This removed large components of the climatological background from the retrieved profiles and was much closer to direct radiance assimilation than the assimilation of retrievals based on climatological information (Eyre et al., 2020). Indeed, the 1D-Var retrieval scheme (Eyre, 1989) used profiles from a short-range forecast as background, whereas other retrieval schemes used statistical background information (Reale et al., 1986). Even with very sophisticated techniques, it is unavoidable that errors in the selected background contribute to the retrieval error. The problem shows up as very systematic air-mass-dependent biases in the retrieved data (Andersson et al., 1991). The errors introduced by the retrieval process are characterised by horizontal correlations that vary with the meteorological conditions and are therefore difficult to accurately account for in the analysis. This problem is fully eliminated by incorporating the retrieval process within the analysis: a combined retrieval/analysis approach enables a more accurate combination of the information contained in the background, in the radiances and in the conventional data (Andersson et al., 1994). All data are analysed simultaneously in a single global inversion problem. The other major innovation of 3D-Var with respect to OI was the global nature of the solver of the analysis update equations, whose solution can be framed as an iterative minimum-finding algorithm of a global cost function. This allows certain issues (discontinuities, numerical artefacts) connected to the need to stitch together separate local analyses in OI to be avoided.

The transition from 3D-Var to 4D-Var at ECMWF was driven by the need to better incorporate time-evolving observations and improve the dynamical consistency of the atmospheric state produced by the analysis update (Andersson et al., 1994; Thépaut et al., 1996). While 3D-Var was a major advancement over OI, it still had at least two fundamental limitations. One is that it treated observations as if they all occurred at a single analysis time, ignoring the fact that weather systems evolve continuously. This meant that observations taken at different times within the assimilation window were not optimally used, leading to a less accurate initial state for the forecast model. The second, possibly more important, deficiency is that 3D-Var, like OI, is a purely statistical assimilation algorithm. This means that the forecast model plays no part in the solution of the analysis equations except for providing a background state. This means, among other things, that there is no guarantee that the resulting analyses are consistent with the model dynamics. This fact explains the importance at the time of "initialisation" techniques like Normal Mode Initialisation to suppress spurious high-frequency oscillations in the analysed fields (Temperton and Williamson, 1981). To address these problems, ECMWF implemented 4D-Var in 1997. Unlike 3D-Var, which only considers spatial relationships in the atmosphere, 4D-Var extends the assimilation process over a time window (initially 6 hours, later extended to 12 hours; see Figure 1). Instead of assuming the background state is static during this period, 4D-Var uses the numerical weather prediction model to evolve the atmospheric state forward in time.

One major advantage of 4D-Var over 3D-Var is its ability to extract more useful information from asynoptic (non-simultaneous) observations. Satellite and aircraft data, which are available at irregular times, could now be optimally incorporated by considering how they influenced the evolving atmospheric state. This resulted in more accurate initial conditions for forecasts, reducing errors and improving predictive skill, particularly for rapidly changing weather patterns.

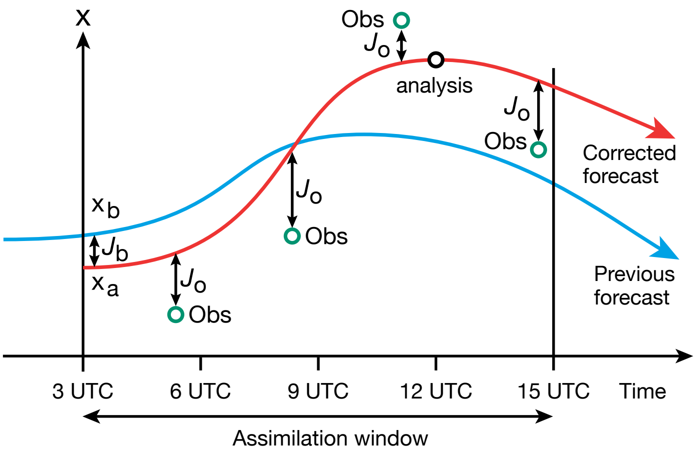


Figure 1: In the case illustrated here, for a single parameter x the observations are compared with a short-range forecast from a previous analysis over a 12-hour assimilation window. The model state x_b at the initial time is modified to achieve a statistically and dynamically based good compromise x_a by minimising a penalty function. The most important penalty terms are J_b, representing the fit to the previous forecast x_b, and J_o, representing the fit to all the observations within the assimilation window.

Another key improvement in 4D-Var is its ability to better control dynamic imbalances in the analysis. Since it uses the forecast model itself as a constraint, the final analysis is dynamically consistent, reducing unrealistic adjustments to temperature, wind, and pressure fields that could occur in 3D-Var. This leads to a smoother transition between the analysis and forecast phases, improving medium-range prediction accuracy.

Despite these advantages, the transition to 4D-Var also came with challenges. One major drawback was the computational cost. This necessitated advances in high-performance computing (HPC) to make 4D-Var operational. However, this was not enough, and incremental 4D-Var at ECMWF was introduced to address the computational challenges associated with full 4D-Var. Indeed, its original formulation was computationally expensive, requiring multiple integrations of the forecast model and its adjoint, involved in the process of minimising the distance between the model trajectory and the observations over the assimilation time window. To make 4D-Var operationally feasible, ECMWF adopted an incremental approach, first proposed by Courtier et al. (1994). This method allowed for a more efficient optimisation process by splitting the assimilation into multiple lower-resolution linear minimisation steps, known as outer and inner loops. Instead of solving the full nonlinear 4D-Var problem at once, incremental 4D-Var approximates it iteratively, first making a coarse-resolution estimate of how observations should be assimilated and then refining it through a series of linearised adjustments.

The implementation of incremental 4D-Var significantly reduced computational costs while maintaining the benefits of the full 4D-Var method. The outer loop operates at higher (finer) resolution, using the full nonlinear forecast model to update the control variables. The inner loop, where most of the optimisation occurs, uses a linearised (tangent-linear) version of the model at a reduced resolution to compute corrections more efficiently. This iterative refinement process ensures that the final analysis remains close to the optimal solution while avoiding the prohibitive expense of running a full-resolution nonlinear model at every iteration.

One of the primary advantages of incremental 4D-Var is its ability to make 4D-Var computationally affordable for operational use. Since the inner loop uses a reduced-resolution model, the overall cost is significantly lower compared to that of a full nonlinear 4D-Var system. Additionally, this approach improves numerical stability, as the assimilation increments remain small and are applied gradually, reducing the risk of introducing unrealistic changes to the atmospheric state. By approximating the analysis solution through successive iterations at increasing spatial resolution, incremental 4D-Var retains the ability to capture large-scale atmospheric corrections and, as the assimilation progresses, resolve smaller-scale features more effectively than a single direct minimisation. This makes it particularly useful for global numerical weather prediction at high resolution, which is an inherently multi-scale problem.

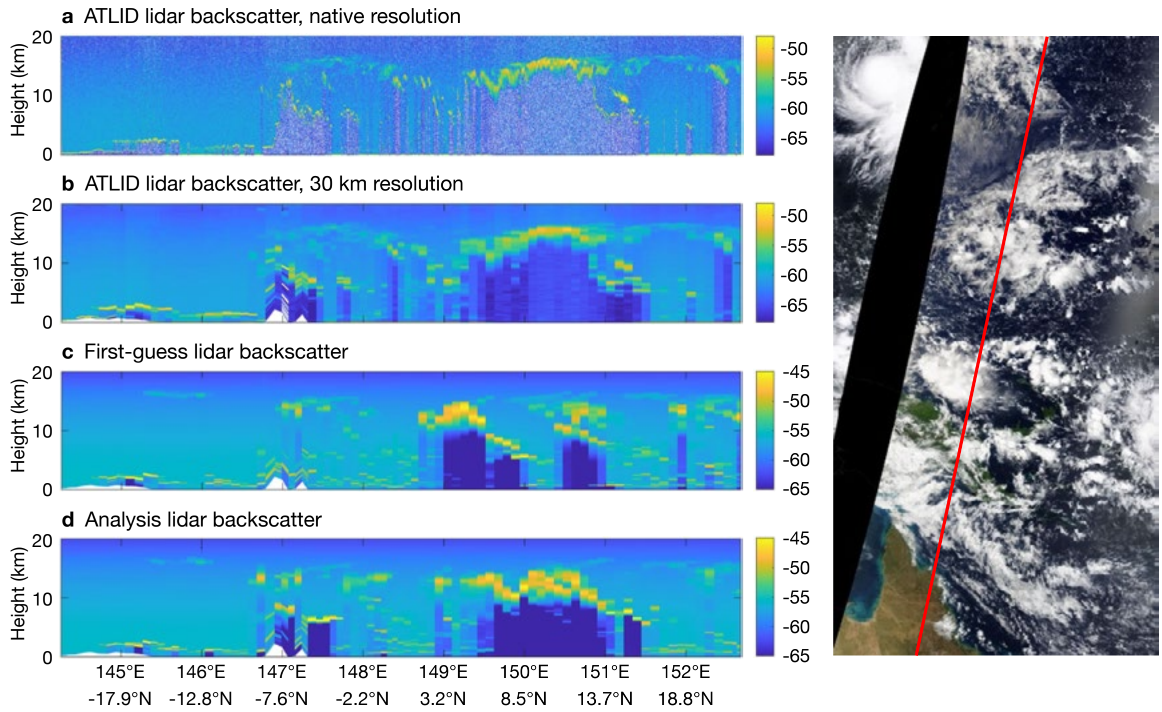
A final aspect of incremental 4D-Var that has allowed the algorithm to pass the test of time is its ability to deal efficiently with nonlinearities in the data assimilation system (Bonavita et al., 2018). As the model resolution increases and more observations are ingested that are nonlinearly related to the analysis variables, this capability of incremental 4D-Var has become increasingly important.

To deal with nonlinearities, an important development was related to incorporating increasingly sophisticated linearised physical parametrizations within the inner-loop minimisation process. In the standard formulation of 4D-Var, the inner loop uses a tangent-linear and adjoint model to propagate information about the state and its sensitivities. However, in early implementations, only the dynamical core of the forecast model was linearised, while physical processes such as radiation, convection, and boundary layer interactions were either ignored or represented in a very simplified manner (Mahfouf and Rabier, 2000). This limitation meant that some key atmospheric processes influencing cloud formation, precipitation, and turbulence were not properly accounted for in the assimilation, leading to suboptimal adjustments in the analysis.

To address this, ECMWF introduced linearised physics schemes within the tangent-linear and adjoint models, allowing physical processes to be considered during the minimisation of the 4D-Var cost function (e.g. Janisková et al., 2002). These schemes ensured that physical processes could be consistently represented within the assimilation cycle while maintaining computational efficiency. The introduction of these linearised physics schemes was particularly beneficial for the assimilation of cloud- and precipitation-affected satellite radiances, as well as for improving the representation of boundary layer and convection-related processes.

The development and refinement of these linearised physical parametrizations have continued as ECMWF has increased model resolution and improved satellite data assimilation. In later years, Janisková and Lopez (2013) expanded the use of linearised physics for variational cloud and precipitation assimilation. A recent achievement is the successful assimilation of lidar backscatter observations from the EarthCARE platform (see Figure 2, by Fielding et al., 2025), which would not have been possible without these continual developments.

Figure 2: Example for the assimilation of Atmospheric Lidar (ATLID) total backscatter on 3 August 2024. (a) ATLID total lidar backscatter at native resolution, averaged to the model grid, (b) ATLID total lidar backscatter at 30 km horizontal resolution, (c) first-guess total lidar backscatter, and (d) 4D-Var analysis total lidar backscatter. Backscatter is shown in units of 10 $\log_{10}$ (m^{-1} sr^{-1}). The red line in the satellite image shows the path of the satellite. From Fielding et al., 2025.



Incremental 4D-Var remains a cornerstone of ECMWF's data assimilation system, continuously evolving to take advantage of new computational capabilities and improved observational data. By balancing accuracy and efficiency, it has enabled ECMWF to maintain high forecast skill while integrating an ever-growing number of satellite and in-situ observations. The method has proved to be a crucial advancement in numerical weather prediction, allowing for more reliable forecasts and better representation of atmospheric processes.

> " THE INCREMENTAL FORMULATION OF FOUR-DIMENSIONAL VARIATIONAL DATA ASSIMILATION (4D-VAR) HAS PROVED TO BE A CRUCIAL ADVANCEMENT IN NUMERICAL WEATHER PREDICTION, MAKING THE METHOD ACCURATE AND AFFORDABLE, AND ALLOWING FOR MORE RELIABLE FORECASTS AND BETTER REPRESENTATION OF ATMOSPHERIC PROCESSES."

Benefiting from this established infrastructure, the Copernicus Atmosphere Monitoring Service (CAMS; described in the ECMWF 50th anniversary paper on Copernicus), which ECMWF operates on behalf of the European Commission, is able to integrate vast amounts of satellite and in-situ observations into its atmospheric composition models. Using 4D-Var, CAMS produces high-quality global analyses of aerosols, greenhouse gases and reactive gases, improving air quality forecasts and environmental monitoring.

It is also worth noting that the 4D-Var framework has enabled the development of the Forecast Sensitivity to Observation Impact (FSOI) methodology to assess the impact of observations on forecast quality. FSOI measures how individual observations influence forecast error reduction. Using the adjoint model, FSOI quantifies the gradient of forecast error with respect to each observation, showing whether a given observation has improved or degraded the forecast. This technique enables convenient and inexpensive real-time assessment of the usefulness of different observing systems, helping optimise data assimilation strategies by prioritising observations that contribute most to forecast improvement. (Cardinali, 2009; Dahoui et al., 2017).

This tool and others are widely used as what we call Observing System Experiments (OSEs) to inform observation providers (e.g. space agencies) about the usefulness of various observing systems, and ECMWF has played a crucial role in shaping the Global Observing System (GOS) through various contributions using its DA infrastructure. These include targeted observation experiments (Buizza et al., 2007) and the Concordiasi project (Rabier et al., 2013).

Since its implementation in 1997, many changes have been made in the 4D-Var system, and some of the advances and challenges are described in the following section.

ENSEMBLE OF DATA ASSIMILATIONS

The Ensemble of Data Assimilations (EDA) was introduced at ECMWF with two distinct but connected objectives. One was to provide improved initial conditions for the initialisation of the ECMWF Ensemble Prediction System (Buizza et al., 2008). The other was as a means to better estimate flow-dependent background error covariances within the variational data assimilation system. Before these EDA developments, the original implementation of incremental 4D-Var relied on static background error covariances, which were derived from climatological statistics. While these were carefully tuned, they did not evolve dynamically with the atmospheric flow. This limitation meant that background errors were often misrepresented, particularly in rapidly changing conditions such as during cyclogenesis, tropical cyclone development, or sudden stratospheric warmings (Bonavita et al., 2012).

Recognising the need for a more adaptive approach, ECMWF began developing the EDA in collaboration with Météo-France, which had been conducting pioneering work on ensemble-based estimation of background errors (e.g. Raynaud et al., 2008). Météo-France had explored the concept of using multiple realisations of the data assimilation cycle to diagnose errors dynamically, an approach that showed promise for improving the accuracy of background error covariance estimation. Inspired by these developments, ECMWF integrated the EDA into its operational 4D-Var system, creating an ensemble of perturbed data assimilation cycles to explicitly represent the uncertainties in the background state (Isaksen et al., 2010).

The introduction of the EDA marked a major advancement over the original incremental 4D-Var framework. In its traditional form, incremental 4D-Var minimised a cost function that included a background error covariance matrix (B-matrix), which had been computed from long-term statistics rather than evolving dynamically with the atmosphere. While this approach worked well in many cases, it struggled to correctly weigh observations in regions with high uncertainty, such as areas of active convection, frontal zones, or dynamically unstable regions. By using EDA-generated background errors, ECMWF was able to account for the flow dependency of forecast uncertainty, making the assimilation system much more responsive to the current state of the atmosphere (Bonavita et al., 2016).

The EDA works by running multiple independent 4D-Var analyses, each with stochastically perturbed observations and model states. These perturbations mimic the uncertainties in the observational data and model representation, creating an ensemble of analyses that reflects the possible range of atmospheric states. By computing the spread across the ensemble members, the EDA provides an adaptive estimate of background error covariances, which is then used in the main high-resolution 4D-Var assimilation. This allows the variational system to adjust its weighting of observations dynamically, giving more weight to observations in regions of high uncertainty and less weight where confidence in the background field is stronger (Isaksen et al., 2010).

Figure 3 shows the case of tropical cyclone Aere (north-eastern part of the Philippines on 8/9 May 2011) and is an illustration of how the errors diagnosed by the EDA, here for mean sea-level pressure, are, by design, constructed to estimate the real analysis errors, thus implicitly taking into account the observation network distribution and the model instabilities. In the present case, they act to extrapolate the observational information from the land-based stations into the more uncertain areas to the north-east of the cyclone, thus helping achieve a better positioning of the analysed storm.

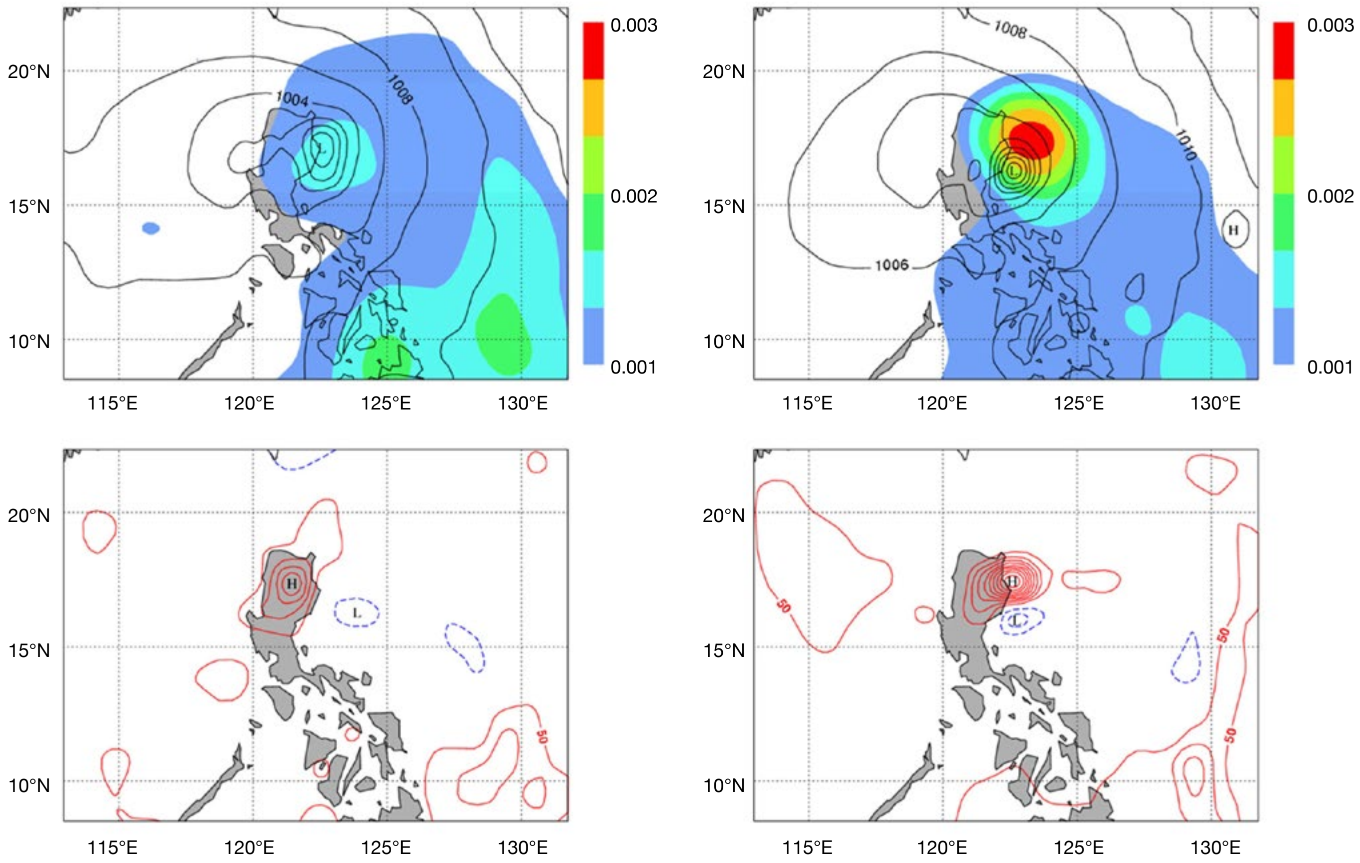


Figure 3: First line: Background mean sea-level pressure forecast valid on 9 May 2011 at 00 UTC (solid line, units: hPa) superimposed on background error estimates for the logarithm of surface pressure (shaded contours). Second line: Surface pressure analysis increments valid on 9 May 2011 at 00 UTC (solid lines indicate positive increments, dashed lines negative increments; isolines of 50 Pa). First column shows fields from the operational ECMWF analysis cycle at the time, with no EDA error estimate, second column from an experiment using EDA error estimates. From Bonavita et al., 2012.

A key benefit of incorporating the EDA into incremental 4D-Var was its impact on forecast sensitivity to observations. In a purely deterministic 4D-Var framework, the system assumes a fixed error distribution, which can lead to overconfidence in certain observations and underuse of others. With the EDA, the system continuously updates its understanding of error growth, leading to more accurate weighting of observational inputs (Bonavita et al., 2012). This proved particularly beneficial for satellite data assimilation, as it allowed ECMWF to assimilate more radiances dynamically, even in areas of high uncertainty, such as cloudy and precipitating regions (Geer et al., 2018).

Beyond its immediate impact on data assimilation, the EDA also played a crucial role in ensemble forecasting at ECMWF. By using EDA-based perturbations to initialise the Ensemble Prediction System (EPS), ECMWF was able to create more realistic ensemble spread, leading to better probabilistic forecasts. This dual application – improving both deterministic analysis and ensemble forecasting – solidified the EDA's place as a cornerstone of ECMWF's modern assimilation framework.

As computing power has increased, ECMWF has continued to refine the EDA, increasing the number of ensemble members and improving the perturbation methodologies. This has further strengthened the system's ability to represent uncertainty and make full use of the ever-expanding volume of satellite and in-situ observations (Lang et al., 2019). However, in recent years, increasing compute power at historical rates has become more challenging, and questions have been asked about the long-term future of investing so many compute resources in the EDA. The main justifications for running the EDA are i) not having to maintain a dedicated ensemble DA system separate from 4D-Var, ii) the 4D-Var is well tested, and iii) the skill of 4D-Var. The Object-Oriented Prediction System (OOPS, Bonavita et al., 2017) was developed to maintain multiple DA methods easily. It creates the possibility to run Ensemble Kalman Filter (EnKF) or Ensemble 4D-Var (EnVAR) for the EDA, but running a completely independent DA system for the deterministic analysis and the EDA is not desirable. At present, the 4D-Var algorithm has a higher skill level than any other algorithm tested in realistic NWP configurations. But it remains an open question if we can use other methods to replicate the contribution the EDA currently makes at much lower cost, without creating overheads in future support and testing. There is also a broader open question about the future of 4D-Var: whether it is computationally feasible at km-scale. At the time of writing, the answer to this is not clear, but OOPS definitely facilitates the implementation of alternative DA methodologies should these be needed.

> “ THE ENSEMBLE OF DATA ASSIMILATIONS HAS IMPROVED BOTH ANALYSIS ACCURACY AND ENSEMBLE FORECASTING – SOLIDIFYING ITS PLACE AS A CORNERSTONE OF ECMWF’S MODERN ASSIMILATION FRAMEWORK.”

WEAK-CONSTRAINT 4D-VAR

The implementation of weak-constraint 4D-Var at ECMWF was motivated by the need to address systematic model errors that limited the ability of 4D-Var to more effectively use various types of observations in the stratosphere and, more recently, at the surface. Traditional strong-constraint 4D-Var assumed that the numerical model used in data assimilation was perfect, neglecting the presence of conditional biases usually arising from deficiencies in model physics. However, systematic errors accumulated over time, particularly affecting stratospheric processes, boundary layer dynamics, and fast-evolving atmospheric phenomena. Weak-constraint 4D-Var allows the assimilation system to account for these errors dynamically, potentially leading to improved forecast accuracy and consistency (Trémolet, 2006).

One major improvement resulting from weak-constraint 4D-Var was the reduction in stratospheric temperature biases. Before its implementation, systematic biases in the stratosphere led to persistent temperature drifts, impacting the representation of the jet stream, planetary waves, and stratospheric circulation. Weak-constraint 4D-Var corrected these errors, producing a more realistic depiction of upper-atmospheric dynamics (Laloyaux et al., 2020). More recently, the development of a version of weak-constraint 4D-Var able to estimate time-varying error structures during the assimilation window has allowed its extension to the boundary layer and the surface, with tangible improvements in the use of surface observations (two-metre temperature, surface pressure, scatterometer winds).

ALL-SKY, ALL-SURFACE SATELLITE DATA ASSIMILATION

A particular enhancement of the data assimilation system at ECMWF is related to continual efforts to improve the observation operators (mapping the model into observation space) and the characterisation of observation errors, especially for satellite observations. These developments addressing better surface emissivity models, better representation of microphysics of snow and graupel particles in the microwave, inclusion of observation error correlation, etc. have led to a massive increase in satellite observation usage, in cloudy and rainy conditions, as well as over land, snow and sea-ice surfaces.

An example is shown in Figure 4 from Geer (pers. comm.), representing the progressively increasing usage of microwave radiances (here Advanced Microwave Scanning Radiometer 2, AMSR2) in the DA system, including after the implementation of a new major cycle of the IFS (Cycle 49r1, implemented in

November 2024). This cycle expanded the use of surface-sensitive microwave channels, for which a lot of data had previously been screened out due to surface types that are harder to simulate. This figure (bottom right) also shows the potential of a high-resolution all-sky/all-surface assimilation approach. The generalisation of the "all-sky, all-surface" approach is not restricted to microwave instruments but includes infrared ones, with high potential from advanced hyperspectral sounders such as the Infrared Atmospheric Sounding Interferometer (IASI) (Geer et al., 2019). These developments have largely benefited from the close partnership between ECMWF and EUMETSAT (see later section).

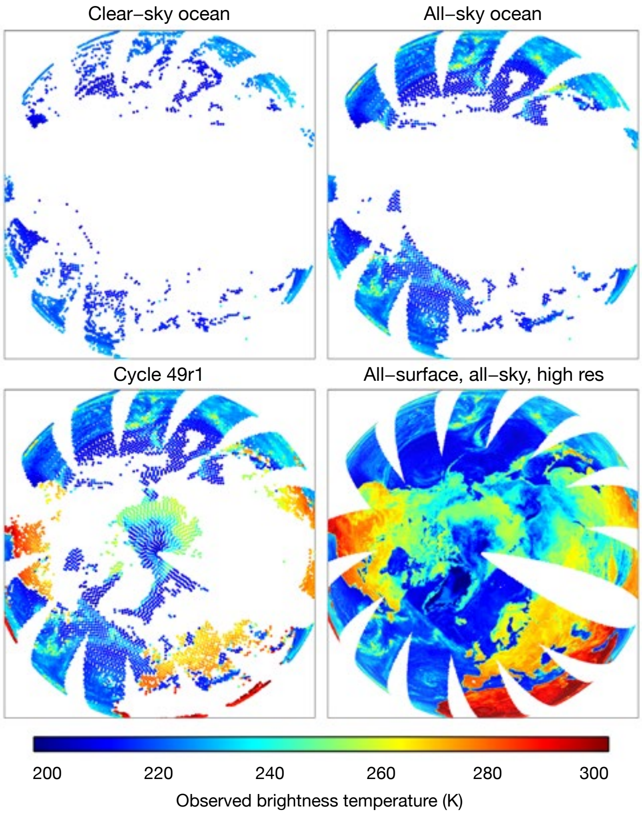


Figure 4: AMSR2 observed brightness temperatures in the 37 GHz v-polarised channel for the 12-hour DA window around 12 UTC, 17 November 2024, simulating the data coverage at earlier stages of DA development (clear-sky, all-sky, all-sky over sea ice/land after implementation of IFS Cycle 49r1). The bottom-right panel shows all data at the 40 km superobbing scale. In the other panels, the data has been thinned to 1 in every 8 superobs, giving effectively a 100 km spacing between observations. Data from multiple orbits has been allowed to overlap/superimpose.

TOWARDS A COUPLED DATA ASSIMILATION SCHEME

Historically, ECMWF's data assimilation systems for the atmosphere, ocean, and land operated independently, with the atmospheric 4D-Var system focusing on upper-air data, while ocean and land components were initialised separately.

For the ocean, ECMWF's ocean data assimilation began with the implementation of the NEMOVAR system, a variational data assimilation software developed collaboratively (CERFACS, ECMWF, Met Office, INRIA/Laboratoire Jean Kuntzmann) to integrate the NEMO ocean model. This system, operationalised in Ocean Analysis System 4 (Ocean-S4, implemented in 2011), used a multivariate three-dimensional variational (3D-Var) First Guess at Appropriate Time (FGAT) approach, assimilating temperature and salinity profiles alongside altimeter-derived sea level anomalies. Building upon this foundation, ECMWF introduced the Ocean ReAnalysis System 5 (ORAS5, introduced in 2017), which incorporated an ensemble generation technique to better represent uncertainties in ocean observations and model physics. ORAS5 provides improved initial conditions for coupled forecasts, thereby enhancing the skill of medium-range weather predictions and seasonal forecasts, the latter being used as an important component of the Copernicus Climate Change Service (C3S) offer. ORAS6 (to be implemented in 2025) further refines ocean reanalysis capabilities. ORAS6 is based on an ocean ensemble-based variational data assimilation system, offering flow-dependent background error variances and vertical correlation scales.

For the land, initially, the assimilation scheme was a two-dimensional Optimal Interpolation (2D OI) method for analysing screen-level parameters and snow depth, while soil moisture and temperature analyses used a one-dimensional OI (1D OI) approach. This framework, though foundational, had limitations in capturing the complex interactions between land surface variables and atmospheric processes. ECMWF introduced a simplified Extended Kalman Filter (EKF) for soil moisture analysis. This advancement allowed for a more dynamic and responsive assimilation of soil moisture data, improving the representation of land–atmosphere feedbacks (de Rosnay et al., 2013). The EKF approach facilitated the integration of various observational data sources, including satellite-derived soil moisture measurements such as those from the Soil Moisture and Ocean Salinity (SMOS) mission and Advanced Scatterometer (ASCAT) data.

The transition towards a fully coupled DA system at ECMWF involves several methodological advancements. One approach is the development of outer-loop coupling, where the coupled model is introduced at the outer-loop level of the assimilation process. This method allows for the simultaneous adjustment of atmospheric and oceanic states, ensuring consistency across the coupled system. Additionally, efforts are being made to enhance the assimilation of surface-sensitive observations, such as sea-surface temperatures and soil moisture, which are critical for accurately capturing the interactions between different Earth system components (de Rosnay et al., 2022).

OPERATIONAL CHALLENGES

Data assimilation for operational NWP is a computationally intensive task that needs to be run daily within strict timeframes on available hardware. This set of requirements poses challenges for DA system developers. Currently, most operational DA systems are run in a hybrid configuration with a high-resolution control analysis based on a global variational solver (either adjoint-based, 4D-Var, or ensemble based, EnVar) and an ensemble DA component for error estimation and cycling (again, either adjoint-based, Ensemble of Data Assimilations (see previous section), or ensemble based, EnKF and its variants). This schematic description already makes it apparent that while DA is conceptually a probabilistic estimation problem, the dimension of the control space for global NWP at current spatial resolutions (O(109)) limits the choice of viable algorithms to those that assume Gaussian errors and only weak nonlinearities in both the observations and the model evolution during the assimilation window (Bonavita et al., 2018).

From a computational perspective, ensemble-based methods (EnVar, EnKF) tend to have better scaling properties than adjoint-based methods, as the analysis sensitivities to observations are directly sampled from the ensemble background forecasts and the solver can be parallelised efficiently. On the other hand, localisation is a known performance limiter for these systems, and the need to sample from the ensemble forecasts requires their storage with fast memory access, which can become impractical for increasing spatial/temporal resolutions and ensemble size.

The adjoint-based methods (4D-Var and its ensemble DA system, EDA) use their ensemble component for background error covariance estimation, but the error evolution in the assimilation window is achieved through running their linearised and adjoint models. This means that for variational methods the main computational constraint comes from the requirement to run the forecast model and its linearised and adjoint versions efficiently and quickly at ever-increasing resolutions. This problem is compounded by the fact that solvers used in variational DA are intrinsically sequential and there is little scope for domain parallelisation. Ten years ago, this state of affairs led people to question the long-term viability of 4D-Var. However, new ideas have changed the picture in the last few years. One of these is continuous DA (Lean et al., 2019). Continuous DA is based on the incremental implementation of 4D-Var and the concept of letting fresh observations into the assimilation system while 4D-Var is running. In practice, this reduces the time-critical portion of 4D-Var to the duration of the last minimisation update instead of the duration of the whole algorithm (which currently runs with four minimisations). This concept will be further developed in the extending-window DA framework, where the length of the assimilation window itself will vary as a function of observation cutoff time, thus ensuring a more continuous update of the analysis and thus even better ability to describe and forecast fast-evolving weather events.

Another important aspect is that of computational efficiency. In the ECMWF DA system, the EDA is the most computationally demanding component, and efforts have been focused on reducing its cost while maintaining or even improving performance. A recent example of these developments is the soft-centred EDA concept (Hólm et al., 2022). This implementation of the EDA differs from the original one as the perturbed members are simplified, lower-resolution 4D-Var updates and the mean background forecast is re-centred on the unperturbed member background. In addition, the minimisations in the perturbed members start from an initial control vector and preconditioning that is inherited from the output of the first minimisation of the unperturbed member. The resulting EDA is approximately 30% cheaper to run and its performance is superior to that of the original version.

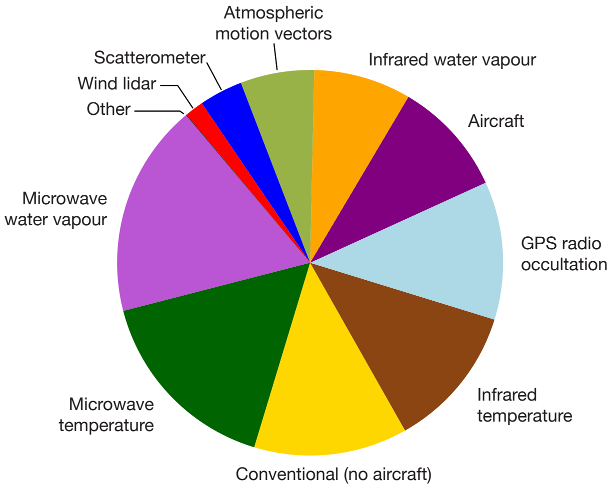
IMPACT OF ECMWF'S DATA ASSIMILATION ON WEATHER FORECASTING AND CLIMATE MONITORING →

As described in the previous sections, a driver for improved data assimilation at ECMWF has been the goal to make best use of the growing spaceborne observing system. This would not be possible without a very close partnership with space agencies, such as the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and the European Space Agency (ESA).

In particular, ECMWF and EUMETSAT have built a very efficient collaborative framework which is critical to enhance the use of satellite data in NWP and environmental monitoring. A key component of this partnership is ECMWF's active participation in EUMETSAT's Satellite Application Facilities (SAFs). These SAFs are specialised centres of excellence that focus on processing satellite data for specific applications, such as numerical weather prediction, climate monitoring, radio occultation and atmospheric composition, to name a few. By engaging with these facilities, ECMWF contributes its expertise in NWP to improve the processing and assimilation of satellite observations, thereby enhancing the accuracy of weather forecasts and climate analyses. Another significant aspect of this collaboration is the EUMETSAT Research Fellowship Programme, which places early-career scientists at institutions like ECMWF to develop innovative applications of satellite data. These seconded Fellows work on projects aimed at advancing the assimilation of satellite observations into ECMWF's forecasting models.

" A DRIVER FOR IMPROVED DATA ASSIMILATION AT ECMWF HAS BEEN THE GOAL TO MAKE BEST USE OF THE GROWING SPACEBORNE OBSERVING SYSTEM – MADE POSSIBLE THROUGH A VERY CLOSE PARTNERSHIP WITH SPACE AGENCIES, SUCH AS EUMETSAT AND ESA."

Figure 5: Relative impact of observing systems on the quality of the operational 24-hour forecast, estimated using their Forecast Sensitivity to Observation Impact (FSOI), and aggregated over the calendar years 2020 to 2024. The impact of microwave and infrared radiance sensors is separated by channel based on primary sensitivity to temperature or water vapour. Ground-based observations are separated into conventional (no aircraft) and aircraft. (Geer, pers. comm.)



For example, past Fellows have focused on improving the use of atmospheric motion vectors, radiances from geostationary satellites, and microwave radiance data from polar-orbiting satellites – all activities dedicated to maximising the impact of these observations in the ECMWF NWP suite. Furthermore, ECMWF and EUMETSAT jointly regularly conduct various flavours of OSEs (as mentioned in the previous section) to assess and optimise the impact of various satellite data on NWP (see Figure 5).

The insights gained from OSEs inform decisions on future satellite mission designs and data assimilation strategies, ensuring that ECMWF's models effectively exploit available satellite data and prepare for future missions (Healy et al., 2022). An advantage of this close cooperation with EUMETSAT is also the speed at which provision of feedback on data quality and evaluation of impact of data on the ECMWF system can be done. For example, EUMETSAT's polar-orbiting MetOp-C satellite was launched on 7 November 2018, and the EUMETSAT radio occultation (RO) team produced high-quality bending angle profiles by 13 November 2018, within only six days of launch, and made them available to the EUMETSAT Radio Occultation Meteorology Satellite Application Facility (ROM SAF) for evaluation. Within days, ECMWF was able to provide quality assessment of these new data in comparison with both the Metop-A and B measurements, by comparing them with NWP information mapped to observation space (see Figure 6, from Healy et al., 2019).

Figure 6: Observation-minus-background departure statistics. The bending angle observation-minus-background (O–B) departure statistics (standard deviation and mean) as a function of impact height for the three Metop satellites. The departures are normalised by dividing them by the bending angle noise values used when assimilating the data.
The statistics are computed for the period 27 November to 2 December 2018.

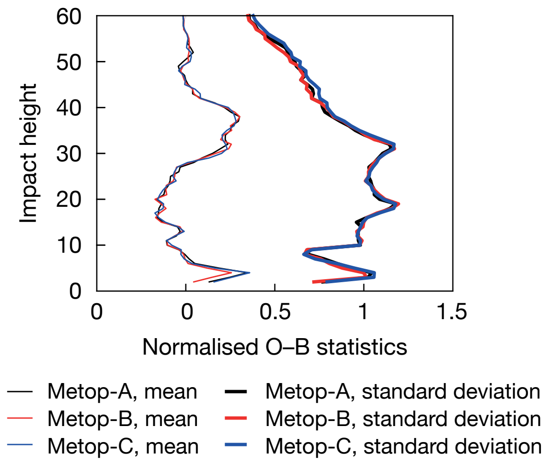
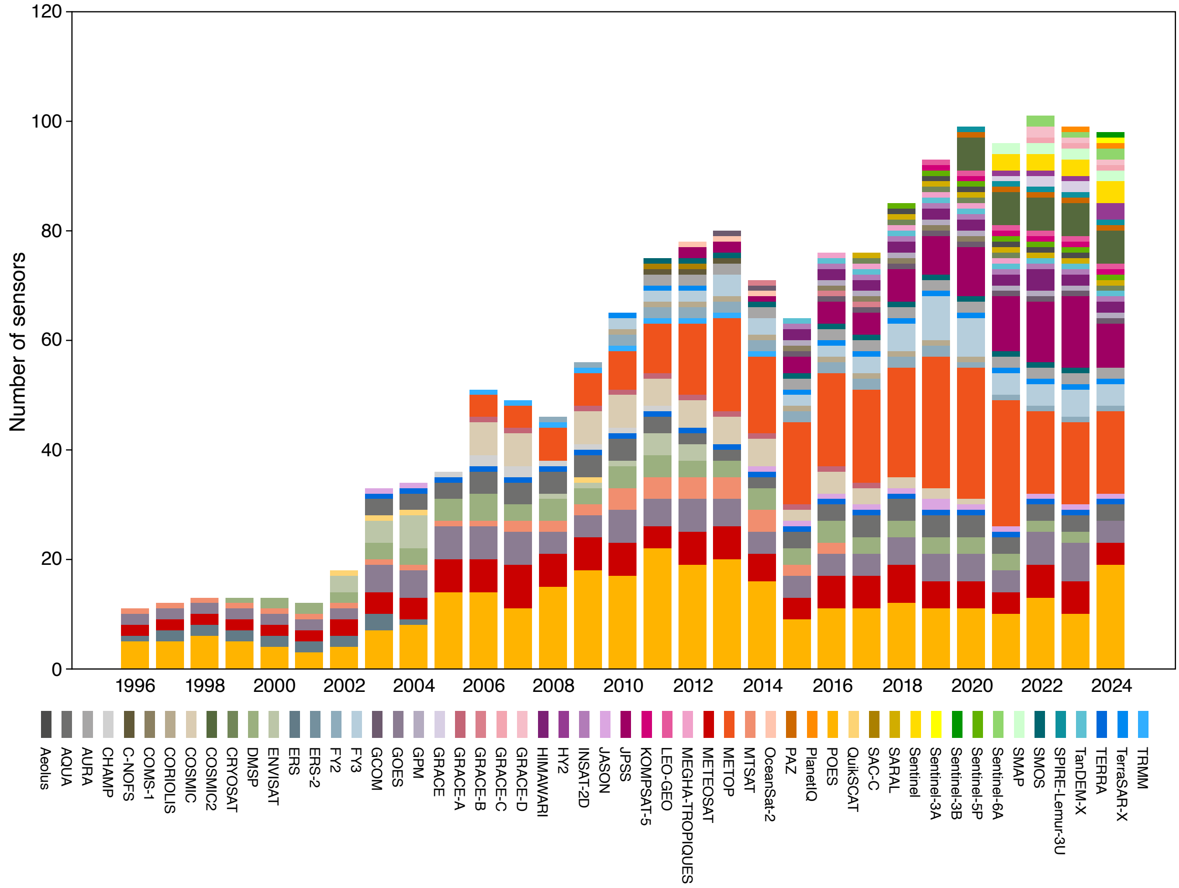


Figure 7: Increase in satellite sensors monitored at ECMWF from 1996 to 2024.

As a result, ECMWF has been a world leader at monitoring and assimilating satellite observations. Figure 7 shows how the data assimilation and model developments over nearly 30 years have enabled the number and diversity of satellite data instruments used to be massively increased.



IMPROVED FORECAST ACCURACY

Forecast skill improvements over the last 45 years have been achieved primarily through improvements to the forecast model, the quality and number of observations and the accuracy of the data assimilation method (Magnusson and Källen, 2013). It is challenging to attribute the contribution of each of these elements, but it is common to compare long-term trends in the performance of forecasts from the reanalysis system with trends in the forecasts from the operational system (see Figure 8). As a first approximation, we can say the reanalysis system shows improvements arising from changes to the observation system, and the trend in the operational system shows improvements from all components, so the difference in trends is an approximation of the combined contribution of model and data assimilation methodology changes.

The lead time at which the anomaly correlation of the 500 hPa geopotential height fell below 85% was 5 days in 2002 and 6.3 days in 2022, so a gain of 0.65 days per decade in this period, a drop from the 1 day per decade improvement reported by Magnusson and Källen (2013). The equivalent change for ERA5 was an increase from 5.5 days in 2002 to 5.9 days in 2022, so an increase of 0.2 days per decade. Therefore, in this 20-year period, we can say, approximately, that a gain of 0.2 days per decade arose from improvements in the Global Observing System, and a gain of 0.45 days per decade arose from improvements in the model and data assimilation. In this context, it is also worth noting ECMWF's current Artificial Intelligence

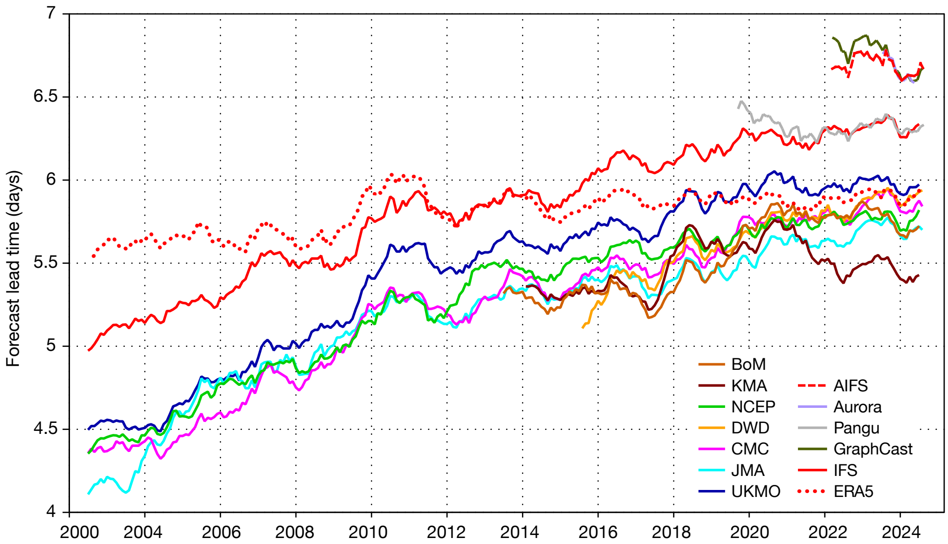


Figure 8: Forecast skill changes of various models, including ECMWF's IFS, AIFS and ERA5. The figure shows the lead time at which the anomaly correlation of 500 hPa geopotential height over the northern hemisphere extratropics falls below 85%.

> "FORECAST SKILL IMPROVEMENTS OVER THE LAST 45 YEARS HAVE BEEN ACHIEVED PRIMARILY THROUGH IMPROVEMENTS TO THE FORECAST MODEL, THE QUALITY AND NUMBER OF OBSERVATIONS AND ENHANCEMENTS OF THE DATA ASSIMILATION METHOD."

Forecasting System (AIFS; see ECMWF 50th anniversary paper on machine learning (in preparation)) configuration gains around 0.35 days over the best physics-based models in 2023–24, only marginally less than the model and data assimilation improvements for the last decade.

In considering these changes, the rapid changes in forecast skill of the IFS in 2005–2007, 2015–2017 and 2018–2020 with respect to ERA5 stand out. The main contributor to the forecast skill gain for the latter change was the introduction of continuous DA, which allowed for the ingestion of observations which arrived after the first minimisation in subsequent minimisations in the outer-loop 4D-Var configuration. Therefore, this gain can be attributed mainly to a change in DA methodology, though there were a number of other changes in this period. In 2015–2017, the changes were a mix of model, most notably increased horizontal resolution, and DA changes (and observation changes, but these would also impact ERA5, whose skill also rose during this period). For the older period, it is difficult now to attribute with high confidence, but a major change in background error formulation (Fisher, 2005) was introduced in 2005 and may have contributed to the large improvement seen in this period. Going further back, the transition to variational assimilation and direct radiance assimilation resulted in the largest changes to operational forecast scores at the end of the 1990s (see Figure 3 in the ECMWF 50th anniversary paper on Earth system modelling).

HIGH-IMPACT AND EXTREME WEATHER

In addition to monitoring the impact of data assimilation developments and improved observations on global scores, there have also been attempts to measure progress for high-impact and extreme weather. This is harder to study objectively, because by definition extreme events are rare and, therefore, it is challenging to test changes in a statistically robust way. The assimilation of satellite observations has repeatedly been shown to play a critical role in the accurate forecasting of individual severe weather cases, most notably that of Hurricane Sandy in October 2012 (McNally et al., 2014). Tropical cyclones (TCs) have been studied, most recently by Magnusson et al. (2025). They concluded that near TCs, observations are important for forecasts mainly up to one day ahead, with the dropsondes particularly helpful to reduce central pressure errors. However, at longer lead times, it is the microwave satellite radiances that are critical to the TC position, and also central pressure up to two days ahead. It was also shown that the development of all-sky microwave assimilation (see section on 'Major data assimilation enhancements' above and Geer et al., 2018) is increasing the impact of microwave radiances further, demonstrating that it is not just the observations, but the maturity of the data assimilation method which is important, especially in areas with persistent cloud cover such as TCs. Scatterometer observations were also shown to be of value, with increasing impact as data thinning is reduced. However, other observation types were not shown to have a strong impact on TC forecasts.

ECMWF also engages with partners to examine the impact of observations on forecasts of atmospheric river (AR) events (Lavers et al., 2024). These studies have examined the impact of targeted observations on forecasts of AR events, particularly through the Atmospheric River Reconnaissance (AR Recon) programme, which involves ECMWF and its Member States. In particular, they explore the value of field campaign dropsonde datasets, in the AR Recon seasons 2022/23 and 2023/24. These show where the dropsondes have value, which can be up to four days' lead time. ECMWF also played a pivotal role in supporting other field campaign experiments, particularly through its involvement in the THORPEX (The Observing System Research and Predictability Experiment) programme. ECMWF's contributions included providing targeted model runs and assimilating observations from these campaigns to enhance weather prediction accuracy in polar regions. ECMWF participated in the Concordiasi project, with data from Concordiasi being assimilated into ECMWF models, improving weather forecasts and reanalysis efforts in polar regions as well as evaluation of satellite data over difficult surfaces, particularly from the IASI on the MetOp-A satellite (Rabier et al., 2013).

CLIMATE AND ENVIRONMENTAL MONITORING

Both ERA-Interim and ERA5 reanalysis datasets, produced by ECMWF, rely on the 4D-Var system to integrate large volumes of observational data into a consistent, long-term dataset. ERA-Interim (Dee at al., 2011), covering the years 1979 to 2019, was based on an earlier version of 4D-Var with a 12-hour assimilation window and a coarser spatial resolution of approximately 79 km. In contrast, ERA5, the production of which was funded under the Copernicus programme, and covering from 1950 to the present, benefits from a more advanced weak-constraint 4D-Var, a higher resolution of approximately 31 km, and hourly output, providing a more detailed and accurate representation of atmospheric, land, and oceanic conditions (Hersbach et al., 2020). ERA5 also assimilates a broader range of satellite observations, including hyperspectral infrared and microwave radiances, with improved bias correction and error representation techniques. These enhancements result in a better depiction of stratospheric processes, and long-term climate trends for screen-level parameters (Simmons et al., 2021). Through the combination of state-of-the-art data assimilation and continuous improvements in observational data usage, 4D-Var in ERA5 continues to enhance the accuracy and reliability of climate reanalysis products, supporting a wide range of scientific, policy and business applications, generating a wide user base, as described in the ECMWF 50th anniversary paper on Copernicus. ERA5 is also crucial for initialising AI-based weather forecasting systems. It provides high-resolution, historical hourly atmospheric data used to train and initialise AI models, including ECMWF's AIFS. The AIFS leverages Graph Neural Networks (GNNs) trained on ERA5 and operational analyses to learn atmospheric patterns and improve predictions. By using ERA5 as initial conditions, AI models generate accurate forecasts of surface weather and extreme events which compete with forecasts from traditional models.

Last but not least, the CAMS reanalysis (EAC4) also benefits from ECMWF's advanced data assimilation infrastructure by integrating a vast array of satellite and in-situ observations into a consistent 20-year-long global dataset. This system ensures high-quality atmospheric composition reanalysis, improving accuracy in pollutants, greenhouse gases, and aerosols. Here also, the 4D-Var technique refines temporal consistency. This reanalysis is used for computing climatologies, studying trends, evaluating models, benchmarking other reanalyses, and most importantly, serving as boundary conditions for regional models covering past periods. These applications support policy-making and environmental monitoring efforts.

FUTURE DIRECTIONS AND PROSPECTS →

The development of ECMWF's data assimilation (DA) system will continue to be driven by the need for accurate initial conditions in Earth system modelling and optimal use of present and future observations to improve forecasts and climate data records. Over the next decade, the DA system will also support the training and initialisation of ECMWF's Artificial Intelligence Forecasting System (AIFS) and national forecasting efforts via the Anemoi initiative (Dramsch et al., 2024), consolidating ECMWF's collaborative efforts with its Member and Co-operating States on this critical issue. The focus will expand beyond initial conditions, using the DA system and observations to directly enhance forecast performance.

An important game changer in the next decade will be that AI and ML applications in DA will continue to rapidly evolve. ECMWF is already integrating ML to correct systematic model errors dynamically, extending beyond the capabilities of weak-constraint 4D-Var. Initial studies (Bonavita and Laloyaux, 2020) showed ML-based corrections can improve forecasts significantly. Recent results (Farchi et al., 2025) confirm forecasts based on this hybrid approach can match state-of-the-art data-driven models while retaining physical realism.

“ AN IMPORTANT GAME CHANGER IN THE NEXT DECADE WILL BE THE INCREASING ROLE OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING IN RESHAPING THE DATA ASSIMILATION LANDSCAPE.”

Machine learning also enhances hybrid observation modelling. As demonstrated by Geer (2024a and 2024b), ML supplements physics-based modelling of complex satellite observations, including radiative properties, hydrometeors, and surface interactions. Additional applications include monitoring observation systems (Dahoui, 2023) and developing latent spaces for variational DA (Melinc and Zaplotnik, 2024). These developments confirm ML's growing role in enhancing analysis accuracy and forecast skill within mathematically robust DA methodologies.

Within this likely revolution, it remains certainly true that the 4D-Var assimilation system (in a broad sense, and with all its peripheral components) will remain central to NWP and atmospheric composition. Efforts will therefore continue to improve observation and background error covariances while pushing computational resolution limits, leveraging experience from DestinE (Sandu, 2024). Extending assimilation windows will optimise performance and workflow efficiency, with potential benefits for time-critical boundary conditions in regional modelling.

The forthcoming transition to a hybrid HPC system with central processing units (CPUs) and graphical processing units (GPUs) will require code adaptation. Possible avenues could be the enhancement of tangent linear and adjoint calculations using machine-learned emulators to reduce computational costs in 4D-Var. The Ensemble of Data Assimilations, another cost-intensive system, will also benefit from these efficiency gains.

A key priority is improving consistency and efficiency across Earth system components, particularly through interface observations from satellites. The methodology successfully applied to ocean and sea ice observations will be expanded to land surfaces. Beyond initial conditions, DA will be used to learn meteorology-dependent errors for machine-learning-based corrections in medium-range forecasts, optimising model parametrizations, and training data-driven neural network models of the atmosphere.

Copernicus Services (CAMS and C3S) will continue to benefit from DA developments, and efforts to estimate emissions and surface fluxes of greenhouse gases and pollutants will intensify, leveraging satellite data such as Sentinel-5P for CH_4 and NO_2. In the context of the Paris Agreement and the monitoring of CO_2 emissions from space, the operational implementation is being prepared for the Copernicus CO_2 Monitoring (CO2M) mission, set for launch in 2027. The climate (ERA6) and atmospheric composition (EAC5) reanalyses will enter full production within the next few years, integrating scientific advancements with automated quality monitoring, while discussions on future reanalysis activities will begin, emphasising data-driven forecasting applications.

Maximising observational data usage remains a top priority and requires a proactive collaborative approach. Existing satellite observations will be assimilated in more challenging environments, such as complex land/sea ice surfaces and cloudy regions, in collaboration with EUMETSAT, ESA and Member and Co-operating States. Assimilation spatial and temporal resolution will increase, leveraging DestinE experience, and new methodologies will be developed to estimate spatial error correlations crucial for 4D-Var and EDA. The coupled DA framework will extract additional insights from observations at the interfaces between Earth system components.

Infrastructure enhancements will ensure the rapid adoption of new satellite observing systems. Early in the next decade, ECMWF aims for operational use of data from the Meteosat Third Generation Imaging (MTG-I) satellite Flexible Combined Imager (FCI) and Lightning Imager (LI) and from EarthCARE, working closely with EUMETSAT and ESA. The Centre will continue supporting EUMETSAT's mission advisory groups for MTG-S and EPS-SG, while expanding ESA collaborations through the DANTEX initiative (Bormann et al., 2025). Efficient integration of newly launched continuity satellites from the US, China, and Japan and evolving in-situ networks will maintain forecasting system performance.

ECMWF will continue engaging with private sector observation providers, particularly in radio occultation data, while working with EUMETSAT, ESA and the US National Oceanic and Atmospheric Administration (NOAA) to validate and acquire these datasets. Future Observing System Experiments and EDA impact assessments will guide network planning, including optimised conventional observation networks optimised as a result of the Systematic Observations Financing Facility (SOFF) initiative of WMO. This exemplifies the increasingly critical role of partnerships in the data assimilation strategy at ECMWF.

CONCLUSION

ECMWF has established itself as a global leader in data assimilation, consistently pioneering methodologies that have significantly improved numerical weather prediction. The transition from optimal interpolation to 4D-Var has been instrumental in enhancing forecast accuracy and extending lead times. The implementation of the Ensemble of Data Assimilations (EDA) has further strengthened uncertainty representation, refining initial conditions for both deterministic and ensemble forecasts. Advances in satellite data assimilation, particularly the integration of all-sky and all-surface observations, have maximised the use of spaceborne data, improving forecasts for extreme weather events. Continuous developments in weak-constraint 4D-Var have addressed systematic model errors, yielding more reliable analyses, especially in the stratosphere and at the surface. The coupled data assimilation framework is another milestone, promising enhanced Earth system modelling through the simultaneous assimilation of atmospheric, oceanic, and land observations.

Crucially, ECMWF's success is underpinned by the contributions of its Member and Co-operating States and strong partnerships with national meteorological services, space agencies, and research institutions. Collaborations with EUMETSAT, ESA and other agencies have ensured optimal use of satellite observations, while joint initiatives such as Copernicus as well as with the WMO have expanded the impact of ECMWF's advancements.

Looking ahead, ECMWF is at the forefront of integrating AI into data assimilation, exploring ML-based corrections to model biases and advanced observation handling. ECMWF is even pioneering radical research into producing forecasts directly from observations (Alexe et al., 2024 and McNally et al., 2024), essentially incorporating the DA step in a fully end-to-end AI-based forecasting system (called AI-DOP). The next decade will see increasing reliance on hybrid CPU-GPU architectures to optimise computational efficiency, ensuring that advanced DA techniques remain viable at higher resolutions. ECMWF's expertise will continue to shape future reanalysis products such as ERA6, reinforcing its role in climate monitoring and forecasting.

The Centre's commitment to international collaboration, particularly through its Member and Co-operating States and strategic partnerships, remains essential for optimising global observing networks. Additionally, ongoing research into continuous data assimilation and extended-window DA will further refine forecast initialisation, particularly for fast-evolving weather systems. With the impending launch of next-generation satellites and increased observational capabilities, ECMWF is well positioned to harness new data sources for even greater forecast improvements. As numerical weather prediction enters the AI era, ECMWF's data assimilation strategy ensures that both traditional physics-based models and emerging AI-driven approaches benefit from the most accurate initial conditions. By maintaining its focus on accuracy, efficiency, and scientific rigour, ECMWF is well positioned to define the next chapter in data assimilation and Earth system prediction, working hand in hand with its partners to push the boundaries of meteorological science.

CONTRIBUTORS

ECMWF would like to thank the contributors to this paper:

Jean-Noël Thépaut, Massimo Bonavita, Niels Bormann, Matthew Chantry, Stephen English, Alan Geer and Tony McNally.

REFERENCES →

Alexe, M., E. Boucher, P. Lean, E. Pinnington, P. Laloyaux, A. McNally, S. Lang, M. Chantry, C. Burrows, M. Chrust, F. Pinault, E. Villeneuve, N. Bormann & S. Healy, 2024: GraphDOP: Towards skilful data-driven medium-range weather forecasts learnt and initialised directly from observations. arXiv e-prints, arXiv:2412.15687. https://doi.org/10.48550/arXiv.2412.15687

Andersson, E., A. Hollingsworth, G. Kelly, P. Lönnberg, J. Pailleux & Z. Zhang, 1991: Global observing system experiments on operational statistical retrievals of satellite sounding data. Mon. Wea. Rev., 119, 1851–1865. https://doi.org/10.1175/1520-0493(1991)119<1851:GOSEOO>2.0.CO;2

Andersson, E., J. Pailleux., J.-N. Thépaut, J. R. Eyre, A. P. McNally, G. A. Kelly & P. Courtier, 1994: Use of cloud-cleared radiances in three/four-dimensional variational data assimilation. Q.J.R. Meteorol. Soc., 120, 627–653. https://doi.org/10.1002/qj.49712051707

Bauer, P., A. Thorpe & G. Brunet, 2015: The quiet revolution of numerical weather prediction. Nature, 525, 47–55. https://doi.org/10.1038/nature14956

Bonavita, M., L. Isaksen & E. Hólm, 2012: On the use of EDA background error variances in the ECMWF 4D-Var. ECMWF Technical Memorandum No. 664. https://doi.org/10.21957/3msfrh5zm

Bonavita, M., E.Hólm, L. Isaksen & M. Fisher, 2016: The evolution of the ECMWF hybrid data assimilation system. Q.J.R. Meteorol. Soc., 142, 287–303. https://doi.org/10.1002/qj.2652

Bonavita, M., Y. Trémolet, E. Hólm, S. Lang, M. Chrust, M. Janiskova, P. Lopez, P. Laloyaux, P. de Rosnay, M. Fisher, M. Hamrud & S. English, 2017: A strategy for data assimilation. ECMWF Technical Memorandum No. 800. https://doi.org/10.21957/tx1epjd2p

Bonavita, M., P. Lean & E. Hólm, 2018: Nonlinear effects in 4D-Var. Nonlin. Processes Geophys., 25, 713–729. https://doi.org/10.5194/npg-25-713-2018

Bonavita, M. & P. Laloyaux, 2020: Machine learning for model error inference and correction. Journal of Advances in Modeling Earth Systems, 12, e2020MS002232. https://doi.org/10.1029/2020MS002232

Bormann, N., P. de Rosnay, S. Healy, H. Zuo, S. E. English & F. Catapan, 2025: ECMWF and ESA start project to better exploit Earth system satellite data. ECMWF Newsletter No. 182, 10–12. https://doi.org/10.21957/m32jp51ad6

Buizza, R., M. Leutbecher & L. Isaksen, 2008: Potential use of an ensemble of analyses in the ECMWF Ensemble Prediction System. Q.J.R. Meteorol. Soc., 134, 2051–2066. https://doi.org/10.1002/qj.346

Cardinali, C., 2009: Monitoring the observation impact of the short-range forecast. Q.J.R. Meteorol. Soc., 135, 239–250. https://doi.org/10.1002/qj.366

Courtier, P., J.-N. Thépaut & A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. Q.J.R. Meteorol. Soc., 120, 1367–1387. https://doi.org/10.1002/qj.49712051912

Dahoui, M., L. Isaksen & G. Radnoti, 2017: Assessing the impact of observations using observation-minus-forecast residuals. ECMWF Newsletter No. 152, 27–31. https://doi.org/10.21957/51j3sa

Dahoui, M., 2023: Use of machine learning for the detection and classification of observation anomalies. ECMWF Newsletter No. 174, 23–27. https://doi.org/10.21957/n64md0xa5d

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut & F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q.J.R. Meteorol. Soc., 137, 553–597. https://doi.org/10.1002/qj.828

de Rosnay, P., M. Drusch, D. Vasiljevic, G. Balsamo, C. Albergel & L. Isaksen: 2013: A simplified Extended Kalman Filter for the global operational soil moisture analysis at ECMWF. Q.J.R. Meteorol. Soc., 139, 1199–1213. https://doi.org/10.1002/qj.2023

de Rosnay et al., 2022: Coupled data assimilation at ECMWF: current status, challenges and future developments. Q.J.R. Meteorol. Soc., 148, 2672–2702. https://doi.org/10.1002/qj.4330

Dramsch, J., B. Raoult, M. Chantry, T. García, L. Denby, F. Prill, N. Sokka, A. Vocino, J. Wijnands, T. Nipen, C. Osuna, S. Akodad, M. Van Ginderachter, D. Van den Bleeken, 2024: Introducing Anemoi: a new collaborative framework for ML weather forecasting. ECMWF Newletter No. 181, 6–8. https://www.ecmwf.int/en/newsletter/181/news/introducing-anemoi-new-collaborative-framework-ml-weather-forecasting

Eyre, J. R., 1989: Inversion of cloudy satellite sounding radiances by nonlinear optimal estimation. I: Theory and simulation for TOVS. Q.J.R. Meteorol. Soc., 115, 1001–1026. https://doi.org/10.1002/qj.49711548902

Eyre, J.R., S.J. English & M. Forsythe, 2020: Assimilation of satellite data in numerical weather prediction. Part I: The early years. Q.J.R. Meteorol. Soc., 146, 49–68. https://doi.org/10.1002/qj.3654

Farchi, A., M. Chrust, M. Bocquet & M. Bonavita, 2025: Development of an offline and online hybrid model for the Integrated Forecasting System. Q.J.R. Meteorol. Soc., e4934. https://doi.org/10.1002/qj.4934

Fielding, M., M. Janisková, S. Mason, R. Hogan, W. McLean & A. Benedetti, 2025: EarthCARE data begin to make an impact. ECMWF Newsletter No. 183, in preparation.

Fisher, M., 2005: “Wavelet” Jb – A new way to model the statistics of background errors. ECMWF Newsletter No. 106, 23–28. https://doi.org/10.21957/ef5601ak43

Geer, A. J., K. Lonitz, P. Weston, M. Kazumori, K. Okamoto, Y. Zhu, E. H. Liu, A. Collard, W. Bell, S. Migliorini, P. Chambon, N. Fourrié, M.-J. Kim, C. Köpken-Watts & C. Schraff, 2018: All-sky satellite data assimilation at operational weather forecasting centres. Q.J.R. Meteorol. Soc., 144, 1191–1217. https://doi.org/10.1002/qj.3202

Geer, A. J., S. Migliorini & M. Matricardi, 2019: All-sky assimilation of infrared radiances sensitive to mid- and upper-tropospheric moisture and cloud. Atmos. Meas. Tech., 12, 4903–4929. https://doi.org/10.5194/amt-12-4903-2019

Geer, A. J., 2024a: Combining machine learning and data assimilation to estimate sea ice concentration. ECMWF Newsletter No. 177, 14–21. https://doi.org/10.21957/agh93vs26

Geer, A. J., 2024b: Simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation. Journal of Advances in Modeling Earth Systems, 16, e2023MS004080. https://doi.org/10.1029/2023MS004080

Healy, S., N. Bormann & K. Lean, 2019: Operational assimilation of Metop-C data. ECMWF Newsletter No. 159, 3-5. https://www.ecmwf.int/en/newsletter/159/news/operational-assimilation-metop-c-data

Healy, S., N. Bormann, A. Geer, E. Hólm, B. Ingleby, K. Lean, K. Lonitz & C. Lupu, 2022: Methods for assessing the impact of current and future components of the global observing system. ECMWF Technical Memorandum No. 916. https://doi.org/10.21957/2f240fe55f

Hersbach, H., B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, A. Simmons, C. Soci, S. Abdalla, X. Abellan, G. Balsamo, P. Bechtold, G. Biavati, J. Bidlot, M. Bonavita, G. D. Chiara, P. Dahlgren, D. Dee, M. Diamantakis, R. Dragani, J. Flemming, R. Forbes, M. Fuentes, A. Geer, L. Haimberger, S. Healy, R. J. Hogan, E. V. Hólm, M. Janisková, S. Keeley, P. Laloyaux, P. Lopez, G. Radnoti, P. de Rosnay, I. Rozum, F. Vamborg, S. Villaume & J.-N. Thépaut, 2020: The ERA5 global reanalysis. Q.R.J. Meteorol. Soc., 146, 1999–2049. https://doi.org/10.1002/qj.3803

Hólm, E., M. Bonavita & S. Lang, 2022: Soft re-centring Ensemble of Data Assimilations, ECMWF Newsletter No. 171, 6–8. https://www.ecmwf.int/en/newsletter/171/news/soft-re-centring-ensemble-data-assimilations

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. https://doi.org/10.21957/obke4k60

Janisková, M., J.-F. Mahfouf, J.-J. Morcrette & F. Chevallier, 2002: Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. Q.J.R. Meteorol. Soc., 128, 1505-1527. https://doi.org/10.1002/qj.200212858306

Janisková, M. & P. Lopez, 2013: Linearized physics for data assimilation at ECMWF. In: Park, S., L. Xu (eds) Data Assimilation for Atmospheric, Ocean and Hydrological Applications (Vol. II). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35088-7_11

Laloyaux, P., M. Bonavita, M. Dahoui & E. Hólm, 2020: Towards an unbiased stratospheric analysis. I, 146, 2392–2409. https://doi.org/10.1002/qj.3798

Lang, S., E. Hólm, M. Bonavita & Y. Trémolet, 2019. A 50-member Ensemble of Data Assimilations. ECMWF Newsletter No. 158, 27–29. https://www.ecmwf.int/en/newsletter/158/meteorology/50-member-ensemble-data-assimilations

Lavers, D. A. et al., 2024: Advancing Atmospheric River Science and Inspiring Future Development of the Atmospheric River Reconnaissance Program. Bull. Amer. Meteor. Soc., 105, E75–E83. https://doi.org/10.1175/BAMS-D-23-0278.1

Lean, P., M. Bonavita, E. Hólm et al., 2019: Continuous data assimilation for the IFS. ECMWF Newsletter No. 158, 21–26. https://doi.org/10.21957/9pl5fc37it

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

Magnusson, L. & E. Källén, 2013: Factors Influencing Skill Improvements in the ECMWF Forecasting System. Mon. Wea. Rev., 141, 3142–3153, https://doi.org/10.1175/MWR-D-12-00318.1

Magnusson, L., S.J. Majumdar, M.L. Dahoui, N. Bormann, M. Bonavita, P.A. Browne et al., 2025: The role of observations in ECMWF tropical cyclone initialisation and forecasting. Q.J.R. Meteorol. Soc., 1–23. https://doi.org/10.1002/qj.4924

Mahfouf, J.-F. & F. Rabier, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. II: Experimental results with improved physics. Q.J.R. Meteorol. Soc., 126(564), 1171–1190. https://doi.org/10.1002/qj.49712656416

McNally, T., M. Bonavita & J. Thépaut, 2014: The Role of Satellite Data in the Forecasting of Hurricane Sandy. Mon. Wea. Rev., 142, 634–646. https://doi.org/10.1175/MWR-D-13-00170.1

McNally, T., C. Lessig, P. Lean, M. Chantry, M. Alexe & S. Lang, 2024: Red sky at night... producing weather forecasts directly from observations. ECMWF Newsletter No. 178, 30–34. https://doi.org/10.21957/tmc81jo4c7

Melinc, B. & Z. Zaplotnik, 2024: 3D-Var data assimilation using a variational autoencoder. Q.J.R. Meteorol. Soc., 150(761), 2273–2295. https://doi.org/10.1002/qj.4708

Pailleux, J., J.-F. Geleyn, M. Hamrud, P. Courtier, J.-N. Thépaut, F. Rabier, E. Andersson, D. M. Burridge, A. J. Simmons, D. Salmond, R. El Khatib & C. Fischer, 2014: Twenty-five years of IFS/ARPEGE. ECMWF Newsletter No. 141, 22–30. https://doi.org/10.21957/ftu6mfvy

Rabier, F., H. Järvinen, E. Klinker, J.-F. Mahfouf & A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. Q.J.R. Meteorol. Soc., 126, 1143–1170. https://doi.org/10.1002/qj.49712656415

Rabier, F., A. Bouchard, E. Brun, A. Doerenbecher, S. Guedj, V. Guidard, F. Karbou, V.-H. Peuch, L. E. Amraoui, D. Puech, C. Genthon, G. Picard, M. Town, A. Hertzog, F. Vial, P. Cocquerez, S. Cohn, T. Hock, H. Cole, J. Fox, D. Parsons, J. Powers, K. Romberg, J. VanAndel, T. Deshler, J. Mercer, J. Haase, L. Avallone, L. Kalnajs, C. R. Mechoso, A. Tangborn, A. Pellegrini, Y. Frenot, A. McNally, J.-N. Thépaut, G. Balsamo and P. Steinle, 2013: The Concordiasi infield Experiment over Antarctica: First Results from Innovative Atmospheric Measurements. Bull. Amer. Meteor. Soc., 94, ES17–ES20. https://doi.org/10.1175/BAMS-D-12-00005.1

Raynaud, L., L. Berre & G. Desroziers, 2008: Spatial averaging of ensemble-based background-error variances. Q.J.R. Meteorol. Soc., 134, 1003–1014. https://doi.org/10.1002/qj.245

Reale, A., D. Gray, M. Chalfant, A. Swaroop & A. Nappi, 1986: Higher resolution operational satellite retrievals. In Conference on Satellite Meteorology/Remote Sensing and Applications, 2nd, Williamsburg, VA (pp. 16-19).

Sandu, I., 2024: Destination Earth's digital twins and Digital Twin Engine – state of play. ECMWF Newsletter No. 180, 22–29. https://doi.org/10.21957/is1fc736jx

Simmons, A. J., H. Hersbach, J. Muñoz-Sabater, J. Nicolas, F. Vamborg, P. Berrisford, P. de Rosnay, K. Willett & J. Woollen, 2021: Low frequency variability and trends in surface air temperature and humidity from ERA5 and other datasets. ECMWF Technical Memorandum No. 881. https://doi.org/10.21957/ly5vbtbfd

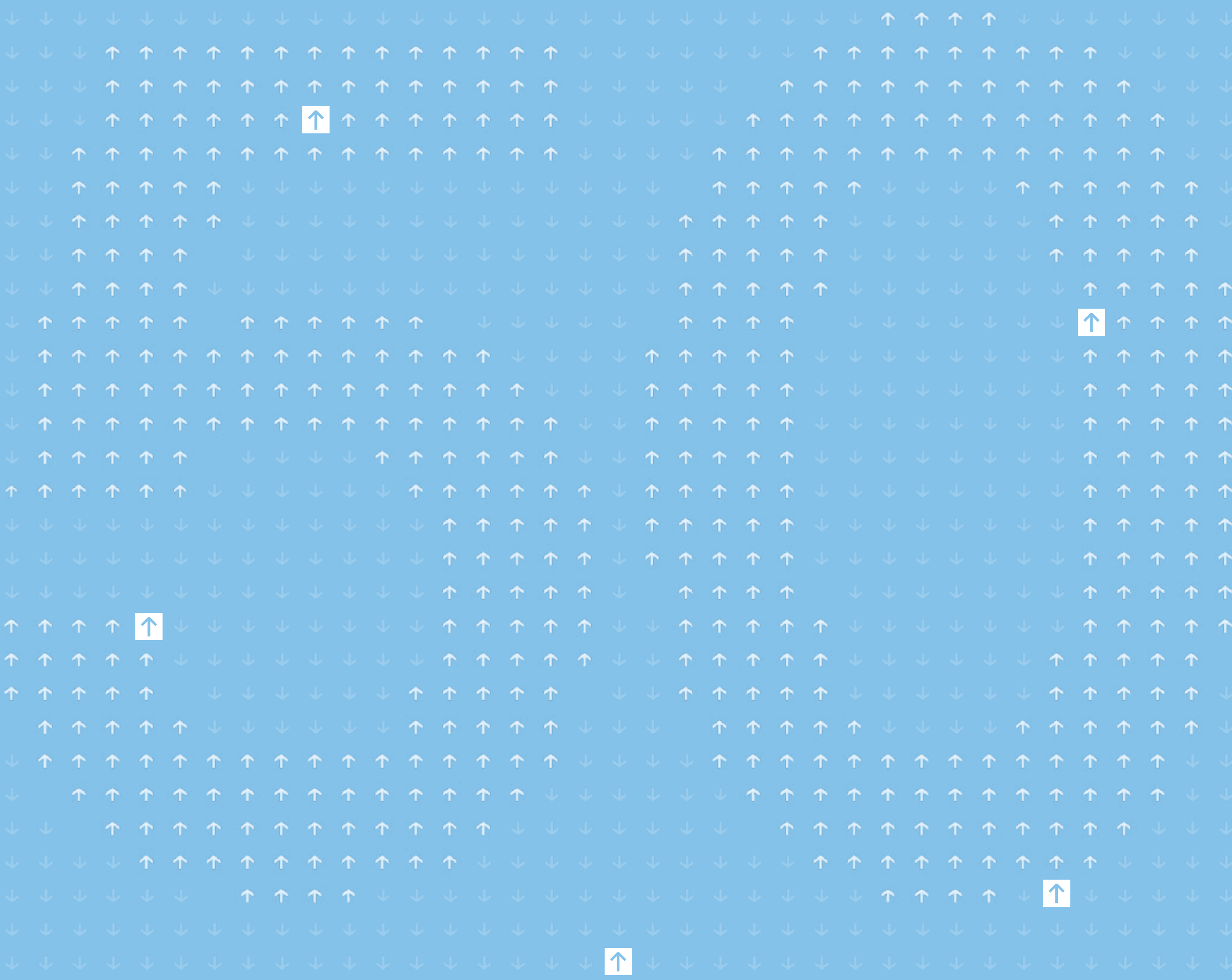
Temperton, C. & D. L. Williamson, 1981: Normal Mode Initialization for a Multilevel Grid-Point Model. Part I: Linear Aspects. Mon. Wea. Rev., 109, 729–743. https://doi.org/10.1175/1520-0493(1981)109<0729:NMIFAM>2.0.CO;2

Thépaut, J.-N., P. Courtier, G. Belaud & G. Lemaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. Q.J.R. Meteorol. Soc., 122, 535–561. https://doi.org/10.1002/qj.49712253012

Trémolet, Y., 2006: Accounting for an imperfect model in 4D-Var. Q.J.R. Meteorol. Soc., 132, 2483–2504. https://doi.org/10.1256/qj.05.224

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

Find out more about ECMWF at 50



ADVANCING WEATHER SCIENCE
THROUGH COLLABORATION