Technical Memo

ECMWF
European Centre for Medium-Range Weather Forecasts

939

Forecast sensitivity to observations in OOPS

Marcin Chrust and Magnus Lindskog
(Research Department)

November 2026

Series: ECMWF Technical Memoranda

A full list of ECMWF Publications can be found on our web site under:
http://www.ecmwf.int/en/publications/

Contact: library@ecmwf.int

© Copyright 2026

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

Literary and scientific copyrights belong to ECMWF and are reserved in all countries. The content of this document is available for use under a Creative Commons Attribution 4.0 International Public License.
See the terms at *https://creativecommons.org/licenses/by/4.0/*.

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

Abstract

Forecast sensitivity-based observation impact (FSOI) estimates the contribution of each assimilated observation to the reduction of short-range forecast error. This memorandum describes its implementation in the Object-Oriented Prediction System (OOPS). It has two components: a new application, `ForecastSensitivity`, which computes the gradient of a forecast-error measure with respect to the initial state with the adjoint model, and an FSOI mode of the square-root minimisation of 4D-Var, which solves for the sensitivity of the forecast error to the observations and computes the impact of each observation. The gradient is a linear functional on increments (a dual vector), not a field: its grid-point values carry the quadrature weights of the grid, so it must be moved between grids with the adjoint spectral transforms rather than interpolated. Compared with the IFS, the energy norm and the gradients agree closely. OOPS includes the variational quality-control weights in the impact, so that the impacts sum to the linearised change in forecast error of the frozen-weight problem; the IFS leaves them out. With the weights, the OOPS total impact over 34 cycles is 0.78 of the IFS one; without them, 0.97 in the two cycles tested. The remaining differences for 2 m temperature and clear-sky radiances are traced, in a single-cycle test, to legacy settings of the IFS FSOI minimisation. The OOPS impacts, with the weights, and the IFS impacts, without them, add up to 0.82 and 1.06 of the Langland–Baker estimate they decompose, and to 0.80 and 1.06 with an error norm made consistent with the assimilation. The shortfall is associated with approximations of the FSOI problem shared by both systems. The IFS total is closer to the estimate because its impacts, without the weights, also credit observations that quality control weighted down with their full departure, which offsets the shortfall. Tests to quantify the contributions of these approximations are outlined as future work.

1 Introduction

The adjoint-based estimate of observation impact (Langland and Baker, 2004) measures how much each assimilated observation reduced the error of a short-range forecast, using the adjoint of the forecast model and of the data assimilation system. It complements observing system experiments (Gelaro and Zhu, 2009): it costs little more than one minimisation, gives the impact of every observation in a single experiment, and can be broken down by instrument, channel, variable, level, region or time. At ECMWF it has been used routinely to monitor the observing system (Cardinali, 2009).

In the IFS the calculation is made by a dedicated suite, FSOBS (forecast sensitivity to observations), with three steps: the gradient of a forecast-error measure with respect to the initial state, computed with the adjoint model in a dedicated IFS configuration; a minimisation that solves for the sensitivity of the forecast error to the observations; and the impact calculation. 4D-Var already runs in OOPS, whereas FSOBS still relies on the IFS implementation; moving FSOBS to OOPS removes the need to maintain the IFS code and configurations that only FSOBS still uses. It also lets FSOI use the same linear models, observation operators and quality control as the current 4D-Var (apart from the linearised surface scheme, switched off in the sensitivity runs), rather than configurations that have been kept for the sensitivity calculation only.

This memorandum describes the OOPS implementation. Section 2 summarises the method and the definition of the impact with variational quality control. Section 3 describes the implementation in OOPS and the IFS, including the transfer of the sensitivity gradient between grids. Section 4 compares OOPS with the IFS, including the impacts by observation type over the experiment period, and with the measured change in forecast error, and identifies the approximations shared by both systems. Section 5 concludes and outlines future work, beyond the port, that would explain the remaining deficiencies.

2 Method

2.1 Forecast-error measure and its gradient

Let $\mathbf{x}^f$ be a forecast valid at the verification time t_v, and $\mathbf{x}_v$ the verifying analysis (section 3). The size of the forecast error $\boldsymbol{\varepsilon} = \mathbf{x}^f - \mathbf{x}_v$ is defined as

$$e = \tfrac{1}{2}\,\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{C}\,\boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{C}$ is a norm matrix. Other norms are possible; as in the IFS FSOBS, $\mathbf{C}$ here defines the dry total-energy norm, averaged over the sphere S and integrated over all model levels,

$$\tfrac{1}{2}\,\boldsymbol{\varepsilon}^{\mathrm{T}}\mathbf{C}\,\boldsymbol{\varepsilon} = \frac{1}{2|S|}\int_S\int_0^{p_r}\left(u'^2 + v'^2 + \frac{c_p}{T_r}\,T'^2\right)\mathrm{d}p\,\mathrm{d}A + \frac{1}{2|S|}\int_S R_d\,T_r\,p_r\,(\ln p_s)'^2\,\mathrm{d}A, \tag{2}$$

with primes denoting components of $\boldsymbol{\varepsilon}$, c_p and R_d the specific heat at constant pressure and the gas constant of dry air, reference temperature $T_r = 350$ K and reference pressure $p_r = 1000$ hPa; the layer thicknesses are those of the reference atmosphere. Humidity does not enter the norm. With these definitions e has units of $\mathrm{J\,kg^{-1}\,Pa}$, and e/g, with g the gravitational acceleration, is the global-mean dry energy per unit area in $\mathrm{J\,m^{-2}}$.

Let $\mathbf{x}_a$ and $\mathbf{x}_b$ be the analysis and the background at the start of the assimilation window, t_0, and e_a, e_b the errors of the forecasts started from them and verifying at the same time. The gradient of e with respect to the initial state $\mathbf{x}_0$ at t_0 is

$$\mathbf{g} = \frac{\partial e}{\partial \mathbf{x}_0} = \mathbf{M}^{\mathrm{T}}\mathbf{C}\,(\mathbf{x}^f - \mathbf{x}_v), \tag{3}$$

where $\mathbf{M}^{\mathrm{T}}$ is the adjoint of the forecast model linearised around the forecast from $\mathbf{x}_0$ (Rabier *et al.*, 1996). We write $\mathbf{g}_a$ and $\mathbf{g}_b$ for the gradients along the forecasts from the analysis and from the background. Following Langland and Baker (2004), the change in forecast error due to the analysis increment $\delta\mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b$ is approximated by

$$\delta e = e_a - e_b \approx \tfrac{1}{2}\,(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}_a. \tag{4}$$

This is exact for a linear model, since e is quadratic; otherwise its error is of third order in $\delta\mathbf{x}_a$ (Errico, 2007), provided the adjoint is that of the nonlinear forecast model and each gradient is computed along its own forecast.

2.2 Sensitivity to the observations

In incremental 4D-Var (Courtier *et al.*, 1994) the analysis increment minimises

$$J(\delta\mathbf{x}) = \tfrac{1}{2}\,\delta\mathbf{x}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{x} + \tfrac{1}{2}\,(\mathbf{H}\,\delta\mathbf{x} - \mathbf{d})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{H}\,\delta\mathbf{x} - \mathbf{d}), \tag{5}$$

where $\mathscr{H}$ is the generalised observation operator, which includes the forecast model to the observation times, $\mathbf{H}$ its linearisation, and $\mathbf{d} = \mathbf{y} - \mathscr{H}(\mathbf{x}_b)$ the background departures. Without variational quality control the solution is $\delta\mathbf{x}_a = \mathbf{A}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{d}$, with the Hessian $\mathbf{A} = \mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}$. Substituting it in (4),

$$\delta e \approx \mathbf{d}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{z} = \sum_i d_i\left[\mathbf{R}^{-1}\mathbf{H}\mathbf{z}\right]_i, \qquad \mathbf{A}\mathbf{z} = \tfrac{1}{2}(\mathbf{g}_a + \mathbf{g}_b), \tag{6}$$

so the impact of observation i is $\delta e_i = d_i\,[\mathbf{R}^{-1}\mathbf{H}\mathbf{z}]_i$. Negative values mean the observation reduced the forecast error. The vector $\mathbf{z}$ is found by a minimisation with the Hessian of the 4D-Var cost function and the forecast gradient as right-hand side (Trémolet, 2008; Cardinali, 2009). Here and in the rest of this memorandum, an observation is a single datum, one value of one variable as stored in the IFS observation database; for satellite radiances it is one channel.

2.3 Variational quality control

The full 4D-Var cost function is non-quadratic for two reasons: the generalised observation operator $\mathscr{H}$ is nonlinear, and with variational quality control (Andersson and Järvinen, 1999) and/or the Huber norm (Tavolato and Isaksen, 2015) the observation cost is not quadratic in the residuals. Incremental 4D-Var treats both in the same way: in each outer loop, $\mathscr{H}$ is linearised around the current trajectory and the quality-control weights are evaluated from the departures from it, and both are then held fixed, so that every inner loop solves a quadratic problem, (5) with $\mathbf{R}^{-1}$ replaced by $\mathbf{W}\mathbf{R}^{-1}$, where $\mathbf{W}$ is the diagonal matrix of quality-control weights w_i, between 0 (observation effectively rejected) and 1 (full weight), with the Lanczos or conjugate-gradient algorithm. The FSOI minimisation is a problem of the same kind. Here and in the rest of this memorandum, the sensitivity to the observations and the impacts are those of this quadratic problem with the weights held fixed; they are not derivatives of the analysis with respect to the observations, which would also include the dependence of the weights and of the linearisation on the observations. We assume a diagonal $\mathbf{R}$, with σ_i the observation-error standard deviation of observation i. Write $\mathbf{r} = \mathbf{R}^{-1/2}(\mathbf{d} - \mathbf{H}\,\delta\mathbf{x})$ for the normalised residuals and $\psi(\mathbf{r}) = \mathbf{W}(\mathbf{r})\,\mathbf{r}$ for the derivative of the observation cost, with the weights $\mathbf{W}(\mathbf{r})$ written as functions of the residuals. If the outer loops have converged, the weights of the last inner loop are evaluated at the analysis residuals $\mathbf{r}_a$, and at convergence of that inner loop the gradient vanishes:

$$\delta\mathbf{x}_a = \mathbf{B}\,\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1/2}\,\mathbf{W}\,\mathbf{r}_a, \qquad \mathbf{r}_a = \mathbf{R}^{-1/2}(\mathbf{d} - \mathbf{H}\,\delta\mathbf{x}_a). \tag{7}$$

Substituting $\mathbf{r}_a$, and since $\mathbf{W}$ and $\mathbf{R}$ are diagonal and commute,

$$\delta\mathbf{x}_a = \mathbf{B}\,\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}(\mathbf{d} - \mathbf{H}\,\delta\mathbf{x}_a) \quad\Longrightarrow\quad \delta\mathbf{x}_a = \left(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\,\mathbf{d}. \tag{8}$$

The weights $\mathbf{W}$ in (8) are the variational quality-control secant weights, $w_i = \psi(r_i)/r_i$, used by the inner-loop minimisation. They should not be confused with the tangent weights $\psi'(r_i)$ that would arise from the second derivative of the full nonlinear cost. Thus $\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\mathbf{H}$ is not the exact Hessian of the nonlinear quality-control problem, but the Hessian of the quadratic problem obtained when the weights are frozen, the one minimised by the last inner loop. It is this frozen-weight problem that the FSOI calculation considered here differentiates.

OOPS applies the weights in a symmetric form, $\mathbf{W}^{1/2}\mathbf{R}^{-1}\mathbf{W}^{1/2}$, or $\mathbf{R}^{-1/2}\mathbf{W}\mathbf{R}^{-1/2}$ for observations with correlated errors. For diagonal $\mathbf{R}$ both equal $\mathbf{W}\mathbf{R}^{-1}$, so the derivation is unchanged. The second form is what the gradient condition above gives without the commutation, so for correlated errors the derivation holds with $\mathbf{W}\mathbf{R}^{-1}$ replaced by $\mathbf{R}^{-1/2}\mathbf{W}\mathbf{R}^{-1/2}$. The decomposition (4) then becomes

$$\delta e \approx \sum_i d_i \left[\mathbf{W}\mathbf{R}^{-1}\mathbf{H}\,\mathbf{z}\right]_i, \qquad \left(\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\mathbf{H}\right)\mathbf{z} = \tfrac{1}{2}(\mathbf{g}_a + \mathbf{g}_b). \tag{9}$$

With $\mathbf{W}$ held fixed, the weights appear both in the Hessian and in the impact, and the impacts add up to $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}_a$ for the increment of this linear problem. The IFS implementation keeps $\mathbf{W}$ in the Hessian but computes the impact as $d_i[\mathbf{R}^{-1}\mathbf{H}\mathbf{z}]_i$. An observation that variational quality control has weighted down, $w_i \approx 0$, then hardly changed the analysis but is still credited with an impact proportional to its background departure d_i, which is typically large. The OOPS implementation uses (9) by default and can reproduce the IFS form for comparison (section 3.4).

Strictly, (9) requires the Hessian and the weights at the analysis. Both the IFS and OOPS FSOBS configurations linearise around the background and evaluate the weights at the background departures, which are typically larger than the analysis residuals, and they solve at the resolution of the first outer loop of 4D-Var rather than the last. We return to this in sections 4.5 and 5.

3 Implementation

3.1 Overview

The OOPS FSOBS suite follows the structure of the IFS one. For each cycle it runs

1. a high-resolution trajectory and screening from the background, which give the background departures $\mathbf{d}$ and the quality-control decisions;

2. two sensitivity runs of the new OOPS application `ForecastSensitivity`, which compute the gradients $\mathbf{g}_a$ and $\mathbf{g}_b$ (3);

3. the FSOI minimisation, which solves (9) and writes the impact of each observation to the observation database (ODB);

4. the match-up (tasks `fc_sens_prepare` and `fc_sens_save`), which copies the impacts of the active observations into the feedback of the assimilation being analysed, matching observations by position, time, observed value, variable and entry number, and archives the result in the ECMWF archive (MARS) as `type=FSOIFB` (formerly `type=MFB`).

An observation of the assimilation that has no counterpart in the FSOI run, because its screening, thinning or quality control took a different decision, receives no impact, and an observation used only in the FSOI run is lost. The settings of the FSOI run must therefore reproduce the observation usage of the assimilation as closely as possible. In OOPS mode the two sensitivity runs replace the IFS adjoint tasks of the suite. The forecast errors ε are prepared as in the IFS suite, at T255, and the sensitivity window runs from the start of the assimilation window to the verification time, 24 hours after the analysis time and 27 hours after the start of the window. The analyses, forecasts and verifying analyses are here the operational ones (expver 0001). The operational system runs two 4D-Var assimilations (figure 1, table 1): the long-window assimilation (stream lwda) cycles, its background $\mathbf{x}_b$ being the 3-hour forecast from its previous analysis, whereas the early-delivery assimilation (stream oper), from whose analyses the operational forecasts start, starts from the same background without cycling. FSOBS runs no analysis of its own: it linearises the Hessian around $\mathbf{x}_b$, computes the departures $\mathbf{d}$ from it, and integrates the adjoint for both gradients along the forecast from the long-window analysis $\mathbf{x}_a$. The error norm, in contrast, verifies early-delivery forecasts (table 1). The error norm and the FSOI problem therefore refer to different assimilations, and the state of $\mathbf{x}_b^f$ at the start of the window, a 9-hour forecast, is not $\mathbf{x}_b$. The OOPS configuration reproduces these settings of the IFS one; section 4.5 quantifies their consequences.

Table 1: States used by each part of the FSOBS calculation, 12 UTC cycle.

Part of the calculation	State
Hessian, departures $\mathbf{d}$	$\mathbf{x}_b$: 3-hour forecast from the long-window analysis at 06 UTC
Adjoint for $\mathbf{g}_a$ and $\mathbf{g}_b$	forecast from the long-window analysis $\mathbf{x}_a$ at 09 UTC
$\mathbf{x}_a^f$	24-hour forecast from the early-delivery analysis at 12 UTC
$\mathbf{x}_b^f$	36-hour forecast from the early-delivery analysis at 00 UTC
$\mathbf{x}_v$	early-delivery analysis at 12 UTC on the next day

The trajectory runs at TCo1279. The IFS interpolates it to the resolution of the FSOI minimisation, whereas OOPS runs a low-resolution trajectory forced towards it.

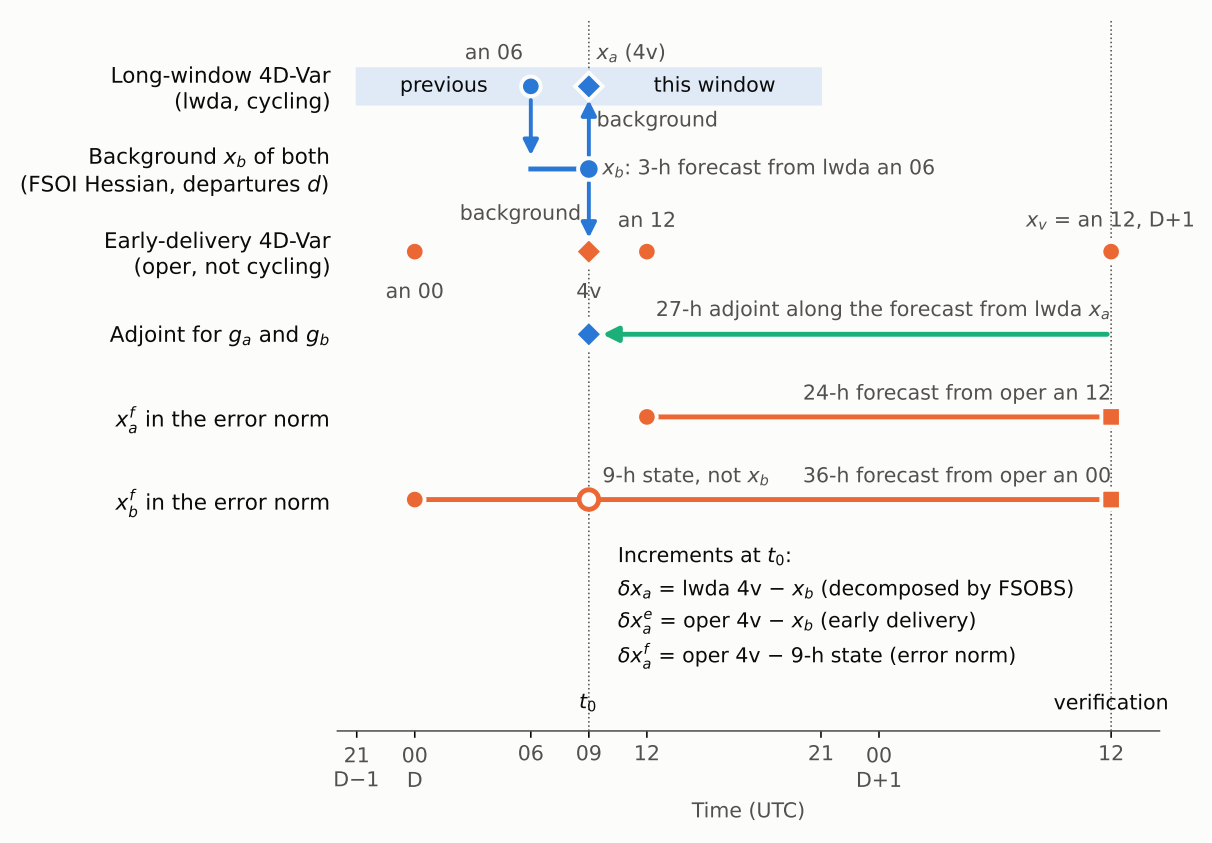

Figure 1: States used by FSOBS for the 12 UTC cycle of day D (long-window 4D-Var window from 09 to 21 UTC, verification at 12 UTC on D+1). Blue: long-window 4D-Var (lwda), which cycles; orange: early-delivery 4D-Var (oper), which starts from the long-window background $\mathbf{x}_b$*; green: adjoint integration.*

3.2 The `ForecastSensitivity` application

`ForecastSensitivity` computes one gradient (3). It

1. runs the nonlinear model from $\mathbf{x}_0$ (in FSOBS the long-window analysis $\mathbf{x}_a$ for both gradients, figure 1) over the sensitivity window and records the linearisation trajectory at the inner-loop resolution, as in the 4D-Var outer loop;
2. reads the forecast error at the verification time and applies the norm matrix $\mathbf{C}$;
3. integrates the adjoint model back to t_0 and writes $\mathbf{M}^{\mathrm{T}}\mathbf{C}\varepsilon$.

It needs no observations, background-error covariances or minimiser. Optional outputs are ε and $\mathbf{C}\varepsilon$ at the verification time, which were used to compare the norm with the IFS. The application is generic, and executables are also provided for the Lorenz 95 and quasi-geostrophic toy models.

In FSOBS, in the IFS as in OOPS, the runs for $\mathbf{g}_a$ and $\mathbf{g}_b$ differ only in the forecast error: both start from $\mathbf{x}_a$, so the same nonlinear forecast is run twice. The minimisation uses only $\frac{1}{2}(\mathbf{g}_a+\mathbf{g}_b)$ (10), and a single adjoint integration from $\frac{1}{2}\mathbf{C}(\varepsilon_a+\varepsilon_b)$, with ε_a and ε_b the errors of $\mathbf{x}_a^f$ and $\mathbf{x}_b^f$, would give the same right-hand side; one trajectory with two adjoint integrations would keep the two gradients, which section 4.5 uses separately. By its definition (3), however, $\mathbf{g}_b$ should be linearised along the forecast from $\mathbf{x}_b$, and the third-order accuracy of (4) (Errico, 2007) also requires each gradient along its own forecast. With the early-delivery forecasts this is not straightforward: $\mathbf{x}_b^f$ starts from the previous early-delivery analysis and does not pass through $\mathbf{x}_b$ (section 3.1). In the configuration of step E in section 5, with $\mathbf{x}_b^f$ the forecast from $\mathbf{x}_b$, $\mathbf{g}_b$ can be linearised along it.

The norm (2) is applied on the grid of the sensitivity runs (O256), with the Gaussian quadrature weights of the grid, whereas the IFS applies it in spectral space; appendix B compares the two forms. The linear model uses the same linearised physics as 4D-Var, except that the linearised surface scheme is switched off in the tangent-linear and adjoint models for stability over the 27-hour window; 4D-Var and the FSOI minimisation keep it.

3.3 The sensitivity gradient is a dual vector

The sensitivity runs and the minimisation do not use the same grid: the gradient is computed on the cubic octahedral grid of the sensitivity runs (O256) and used on the linear reduced Gaussian grid of the minimisation inner loop (N128). Both represent the same spectral truncation, T255. How the gradient is moved between the grids matters, because a gradient is not a field: what must not change between representations is the directional derivative $\mathbf{g}^{\mathrm{T}}\delta\mathbf{x}$ for every increment $\delta\mathbf{x}$, not the grid-point values of $\mathbf{g}$.

Let $\mathbf{S}$ be the inverse spectral transform (spectral to grid point) and $\mathbf{S}^{-1}$ the direct transform on a given grid. The tangent-linear model starts from a grid-point increment and applies $\mathbf{S}^{-1}$, so the adjoint model returns the grid-point gradient $\mathbf{g}=(\mathbf{S}^{-1})^{\mathrm{T}}\boldsymbol{\gamma}$, with $\boldsymbol{\gamma}$ the spectral gradient. With Gaussian quadrature, $\mathbf{S}^{-1}=\mathbf{S}^{\mathrm{T}}\mathbf{Q}$ with the diagonal matrix $\mathbf{Q}$ of quadrature weights (the transpose taken with the spectral scalar product (13) of appendix B), so $\mathbf{g}=\mathbf{Q}\mathbf{S}\boldsymbol{\gamma}$: each grid-point value carries its quadrature weight, of order $1/N$ on an N-point grid. These weights come from the change of variables at the start of the tangent-linear model and are independent of the weights in the norm.

Interpolating $\mathbf{g}$, or transforming it with the direct transform on one grid and the inverse transform on the other, keeps the size of the O256 values on the N128 grid. The correct N128 gradient is larger by about

the ratio of the numbers of points, $271\,360/88\,838 \approx 3.05$. A first version of the OOPS suite moved the gradient as a field, and its impacts were much smaller than those of the IFS.

The gradient is therefore stored as spectral coefficients and moved with the adjoint transforms. The sensitivity run writes $\mathbf{S}_1^{\mathrm{T}}\mathbf{g}_1 = (\mathbf{S}_1^{-1}\mathbf{S}_1)^{\mathrm{T}}\gamma = \gamma$ with the adjoint of the inverse transform on its grid, and the minimisation reads it with the adjoint of the direct transform on its own grid, $\mathbf{g}_2 = (\mathbf{S}_2^{-1})^{\mathrm{T}}\gamma$. The first step of the adjoint of the change of variable in the minimisation applies $\mathbf{S}_2^{\mathrm{T}}$, which returns $(\mathbf{S}_2^{-1}\mathbf{S}_2)^{\mathrm{T}}\gamma = \gamma$. The weights of each grid enter only through its own transform pair and cancel. The only requirement is that both grids transform the truncation exactly, $\mathbf{S}^{-1}\mathbf{S} = \mathbf{I}$, which holds for O256 and N128 at T255 to within the precision of the transforms on these reduced grids (appendix B). The stored file holds the same spectral gradient as the IFS, so the two can be compared directly. Humidity is stored as spectral coefficients, like the other variables, although it is a grid-point field in operational 4D-Var (section 4.4).

3.4 FSOI in the square-root minimisation

The FSOI mode of the square-root minimisation (`SQRTMinimizer`) is selected in its configuration. With the control variable $\mathbf{v}$, $\delta\mathbf{x} = \mathbf{U}\mathbf{v}$ and $\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$, it solves

$$\left(\mathbf{I} + \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\mathbf{H}\,\mathbf{U}\right)\mathbf{v} = \mathbf{U}^{\mathrm{T}}\,\tfrac{1}{2}(\mathbf{g}_a + \mathbf{g}_b), \qquad \mathbf{z} = \mathbf{U}\mathbf{v}, \tag{10}$$

with the Lanczos solver of 4D-Var, for a fixed number of iterations (62 in FSOBS, in the IFS as in OOPS). The right-hand side replaces the usual gradient of the cost function at the first guess; the J_b term is not added. The tangent-linear observation operators are then applied to $\mathbf{z}$, and for every J_o term the impact $d_i[\mathbf{W}\mathbf{R}^{-1}\mathbf{H}\mathbf{z}]_i$ is written to the observation space. Other cost terms produce no impact. The FSOI mode requires a single outer loop: the impact uses the background departures and the linearisation of the first outer loop. With a truncated solve after k iterations, the impacts add up exactly to $\mathbf{b}^{\mathrm{T}}\mathbf{v}_k$, with $\mathbf{b} = \mathbf{U}^{\mathrm{T}}\mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{R}^{-1}\mathbf{d}$, but only approximately to $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\tilde{\mathbf{x}}$, where $\delta\tilde{\mathbf{x}}$ is the increment of the FSOI linear problem, because the Krylov spaces built from the forecast gradient and from $\mathbf{b}$ differ. The increment $\delta\tilde{\mathbf{x}}$ is in general not the analysis increment $\delta\mathbf{x}_a$ (section 4.5).

A configuration switch selects the definition of the impact. By default $\mathbf{W}$ is included; otherwise $\mathbf{R}^{-1}$ is applied without the weights in the impact, while they stay in the Hessian, which reproduces the IFS definition (section 2.3).

3.5 Suite and scripts

The suite scripts provide a sensitivity mode for the two gradient runs and the FSOI configuration of the minimisation, and retrieve the two gradients at their own resolution in the order in which the minimisation reads them. The suite generator adds the OOPS sensitivity tasks instead of the IFS adjoint tasks when the experiment runs in OOPS mode. The new code paths are active only in OOPS FSOBS, and the changes are neutral for IFS FSOBS: an IFS-mode FSOBS experiment with the new code gives the same total impact as an earlier one in every cycle with the same observation input (15 to 20 May). Appendix A lists the configuration.

4 Comparison with the IFS

4.1 Experiments

The comparison uses an IFS reference experiment and an OOPS FSOI experiment, run with the same observations and the same cycle of the IFS (CY50R1) for the analyses of 15 to 31 May 2026 at 00 and 12 UTC, 34 cycles. The observation counts in the two agree to within 0.3% in every cycle. The single-cycle diagnostics of section 4.2 and the test of the IFS minimisation in section 4.4 were made for cycle 2026051512 in an earlier pair of development experiments, which reproduce the totals of the 34-cycle experiments to within 2% in the 12 cycles they share (15 to 20 May), except for cycle 2026051512 of the OOPS one, whose archive holds a test run with different settings. Table 2 gives the experiment identifiers. The systematic OOPS/IFS comparison covers 34 cycles; the attribution experiments of sections 4.2 and 4.4, and the configuration with a consistent error norm in section 4.5, are diagnostic tests over one to four cycles and should not be read as climatological estimates of the effects they isolate.

Table 2: Experiments. The 34 cycles are 2026051500 to 2026053112.

Configuration	Experiment	Cycles	Purpose
IFS reference	jdbs	34	IFS implementation
OOPS FSOI	jdbt	34	OOPS implementation, weights in the impact
OOPS without weights in the impact	jdit	2026051500, 2026051512	matched-definition comparison
IFS development	jbv7	2026051512	norm and gradient diagnostics; minimisation test
OOPS development	jbv8	2026051512	norm and gradient diagnostics; minimisation test
IFS, consistent error norm	jdkc	2026051500–2026051612	long-window error norm (section 4.5)
OOPS, consistent error norm	jdkd	2026051500–2026051612	long-window error norm (section 4.5)

4.2 Norm and sensitivity gradient

This section tests whether OOPS computes the same norm and sensitivity gradient as the IFS, in the development experiments for cycle 2026051512. Here and in section 4.4, IFS values a_j and OOPS values b_j (spectral coefficients or impacts per observation) are compared by their uncentred correlation $\sum_j a_j b_j / (\sum_j a_j^2 \sum_j b_j^2)^{1/2}$, which measures the agreement of the patterns, and by their amplitude ratio $(\sum_j b_j^2 / \sum_j a_j^2)^{1/2}$, the ratio of the root-mean-square values OOPS/IFS.

The norm agrees to four decimal places. OOPS wrote $\mathbf{C}\varepsilon$ at the verification time with the adjoint transform, and it agrees with $\mathbf{C}\varepsilon$ from the IFS norm routine with correlation and amplitude ratio 1.0000 for every variable and every band of model levels and wavenumbers (appendix B). The forecast errors of the two systems are identical.

The gradients agree closely at large and medium scales above the boundary layer: for total wavenumbers up to 120 and model levels 1–130, the correlation is at least 0.95 and the amplitude ratio 0.96–1.04 for every variable of table 3. Agreement decreases at the smallest scales and near the surface, and at the lowest model level the wind gradients are nearly uncorrelated (0.12–0.32). Part of the difference comes from

the linearised model: the IFS sensitivity configuration uses legacy linear-physics and dynamics settings, whereas OOPS uses the current 4D-Var settings. In a test for cycle 2026051500, setting the surface-layer switches of the IFS linearised physics as in 4D-Var (perturbation of the surface exchange coefficients and of the surface arrays, regularisation of the surface computation, surface trajectory, surface drag on the wind only: `LEKPERTS`, `LENOPERT`, `LREGSF`, `LESURFTRAJ`, `LSDRDS`) raised the correlation of the vorticity and divergence gradients at level 137 from 0.19 and 0.24 to 0.94 and 0.95, with amplitude ratios within 1%; `LSLPHY` alone had no effect. The dynamics settings of 4D-Var then raised the correlation of the temperature gradients at levels 131–136 from 0.60 to 0.71, but near the surface they remain the least well correlated (0.51 at level 137).

These differences hardly affect the impacts: in a test for cycle 2026051512, an OOPS minimisation started from the IFS gradient changed the total impact by less than 1%.

Table 3: Agreement of the IFS and OOPS sensitivity gradients for the forecast from the background, cycle 2026051512 (development experiments): correlation / amplitude ratio OOPS/IFS (section 4.2), per band of model levels and of total wavenumber n.

	Levels	n = 1–20	n = 21–120	n = 121–255
Vorticity	1–90	1.00 / 0.98	0.99 / 0.97	0.98 / 0.96
	91–130	0.99 / 0.98	0.98 / 0.99	0.92 / 1.01
	131–136	0.99 / 0.88	0.97 / 0.92	0.87 / 0.95
	137	0.12 / 0.76	0.14 / 0.96	0.16 / 1.50
Divergence	1–90	0.99 / 1.01	0.96 / 0.99	0.73 / 0.94
	91–130	0.98 / 1.00	0.95 / 1.04	0.75 / 0.98
	131–136	0.96 / 1.00	0.92 / 1.01	0.62 / 0.99
	137	0.32 / 1.68	0.23 / 1.72	0.25 / 1.48
Temperature	1–90	0.99 / 0.99	0.98 / 0.96	0.92 / 0.92
	91–130	0.99 / 0.98	0.96 / 1.04	0.90 / 1.03
	131–136	0.83 / 0.99	0.66 / 1.03	0.39 / 1.13
	137	0.77 / 0.79	0.69 / 0.78	0.42 / 0.77
Surface pressure		0.99 / 0.96	0.95 / 1.02	0.62 / 1.08

4.3 Impacts by observation type

Figures 2–5 and table 4 compare the impacts in the OOPS FSOI experiment with those in the IFS reference, by observation type, over the 34 cycles from 2026051500 to 2026053112. The impact of an observation type is the sum of δe_i over its observations, in the units of e (section 2); negative values mean a reduction of the 24-hour forecast error. Satellite radiances (clear-sky and all-sky), aircraft and GPS radio occultation have the largest impacts in both systems (figure 2). Because the OOPS impacts are smaller than the IFS ones by a larger factor for conventional than for satellite observations, the share of satellite observations in the total is larger in OOPS (figure 3). Per observation, drifting buoys have by far the largest impact in both systems. Apart from ground-based radar precipitation (observation type 14), which has few observations, they are followed in the IFS by surface stations, scatterometers and atmospheric motion vectors, and in OOPS by scatterometers and atmospheric motion vectors, since OOPS halves the impact of surface stations. Clear-sky radiances have the smallest impact per observation (figure 4). Since a radiance observation is a single channel (section 2.2), this is an impact per channel, and about three quarters of the clear-sky observations are channels of hyperspectral infrared sounders, which provide a hundred or more channels per spectrum. Counting the whole spectrum of a sounder as one observation would change the ordering by observation type considerably.

With the quality-control weights in the impact, the OOPS total impact is 0.778 of the IFS value over the 34 cycles, with little variation from cycle to cycle (0.72–0.82; figure 5). The ratio is 0.54–0.61 for conventional observations and 0.77–0.94 for satellite observations (table 4). In the two cycles examined in section 4.4, the main cause is the definition of the impact. Of the difference in the totals, conventional observations (aircraft, surface stations, radiosondes, drifting buoys, wind profilers and pilot balloons) account for 48%, all-sky radiances for 26% and clear-sky radiances for 18%.

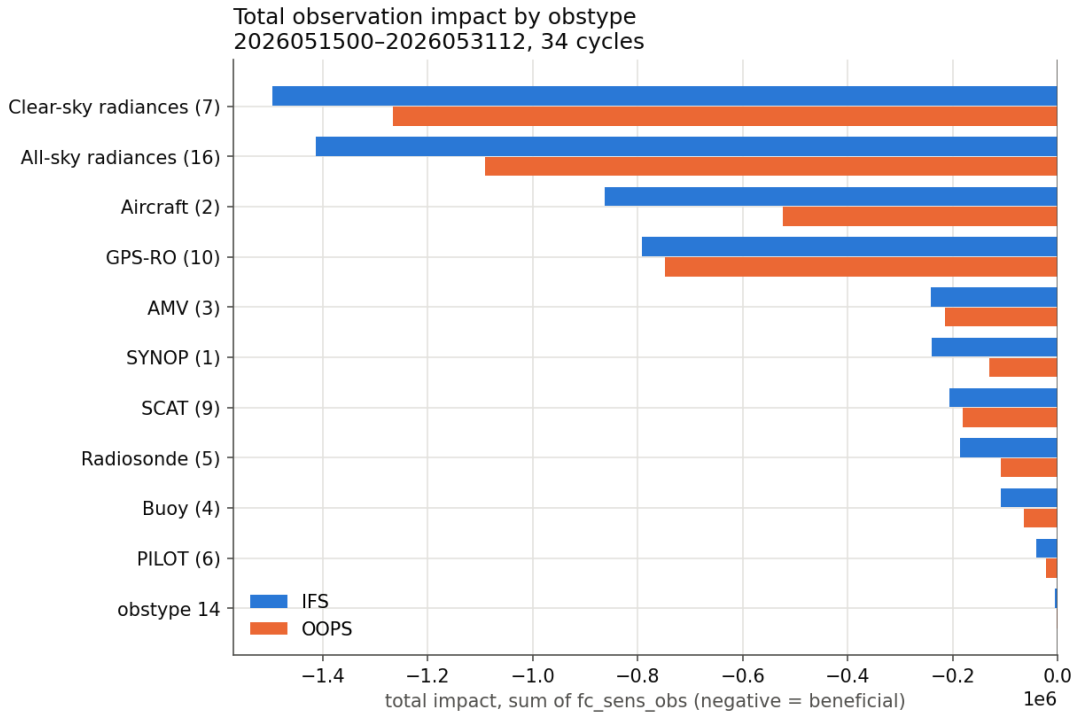

Figure 2: Total impact (sum of δe_i, units of e) by observation type, IFS (blue) and OOPS (orange), 34 cycles from 2026051500 to 2026053112. Negative values mean the observations reduced the forecast error. Numbers in parentheses are the IFS observation types and obstype 14 is ground-based radar precipitation.

4.4 Observation impact and its definition

This section establishes why the default OOPS and IFS impacts differ. The main result is that, with the same definition of the impact, the two systems agree closely in total over the two cycles tested: the ratio of the totals is 0.97, and every listed observation type that carries quality-control weights is within 0.91–1.08 of the IFS impact (table 5). Most of the difference in the default results therefore comes from the treatment of the variational quality-control weights, not from the implementation; the differences that remain, for clear-sky radiances and per observation, are discussed at the end of the section. The section compares the IFS reference with the OOPS calculation with the weights in the impact and without them (table 2), for the two cycles of 15 May; the per-observation diagnostics are for cycle 2026051512.

With the default definitions, the OOPS/IFS ratio decreases with the normalised departure $|d_i/\sigma_i|$, as expected from the quality-control weights, which decrease with the departure (section 2.3): for aircraft winds from 0.85 below 0.5 to 0.33 beyond 3 (table 6). Since the impact of an observation is proportional to its departure, observations with large departures carry much of the total. These weights are evaluated at the background departures (section 2.3). The analysis residuals are generally smaller, so the weights

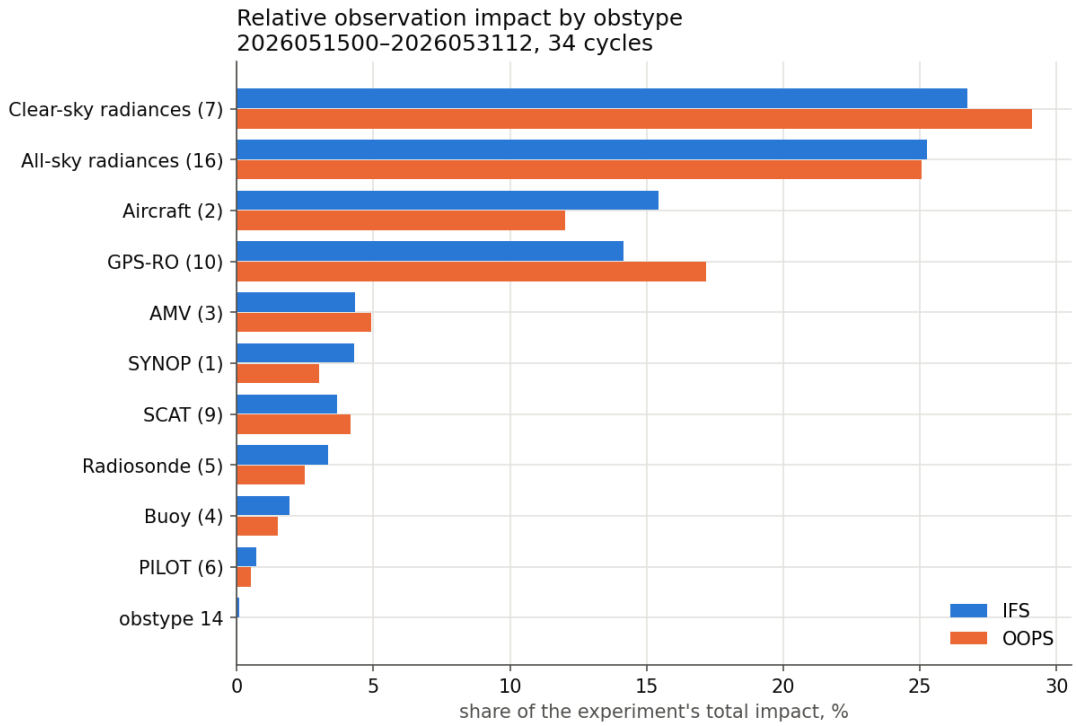

Figure 3: Share of each observation type in the total impact of each experiment, in per cent, IFS (blue) and OOPS (orange), 34 cycles from 2026051500 to 2026053112. Observation types as in figure 2.

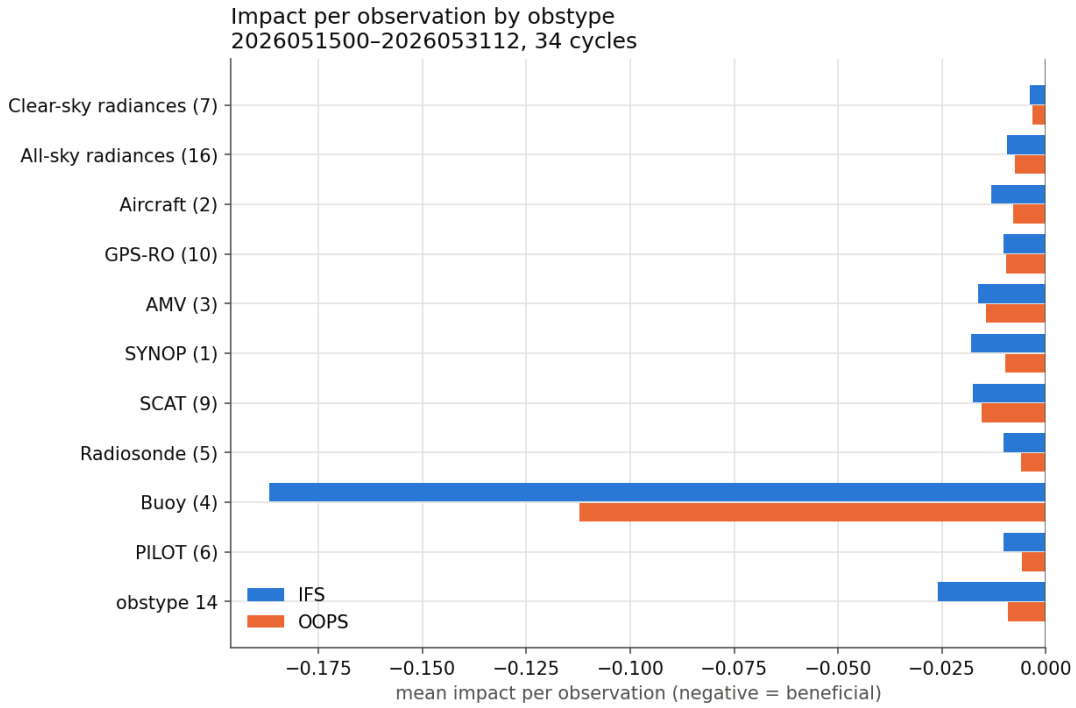

Figure 4: Mean impact per observation by observation type (units of e), IFS (blue) and OOPS (orange), 34 cycles from 2026051500 to 2026053112. Negative values mean a reduction of the forecast error. An observation is one datum (for radiances, one channel; section 2.2); observation types as in figure 2.

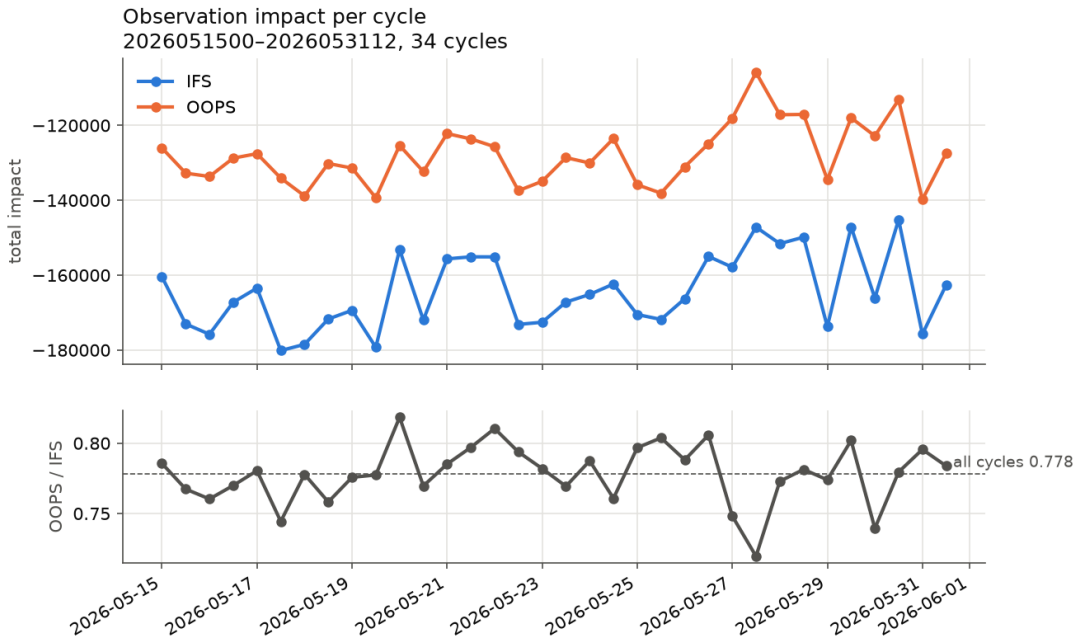

Figure 5: Total impact per cycle (units of e), IFS (blue) and OOPS (orange), top, and the ratio OOPS/IFS per cycle, bottom; the dashed line is the ratio of the totals over all 34 cycles.

Table 4: Total impact by observation type (units of e), IFS and OOPS, and the ratio OOPS/IFS, 34 cycles from 2026051500 to 2026053112. The total includes all observation types.

Observation type	IFS	OOPS	OOPS/IFS
Clear-sky radiances	−1 494 831	−1 266 036	0.85
All-sky radiances	−1 412 816	−1 090 113	0.77
Aircraft (AIREP)	−863 563	−523 474	0.61
GPS radio occultation	−792 089	−747 462	0.94
Atmospheric motion vectors (SATOB)	−241 493	−214 737	0.89
Surface stations (SYNOP)	−240 499	−130 841	0.54
Scatterometer (SCAT)	−205 808	−180 882	0.88
Radiosondes (TEMP)	−186 188	−107 747	0.58
Drifting buoys (DRIBU)	−108 167	−64 942	0.60
Wind profilers and pilot balloons (PILOT)	−40 310	−22 257	0.55
Ground-based radar precipitation (obstype 14)	−5 037	−1 751	0.35
Total	−5 590 800	−4 350 242	0.778

of the last inner loop of 4D-Var are larger, and the reduction of the impact of observations with large departures is probably overstated here; step B of section 5 would quantify this. Without the weights in the OOPS impact, the dependence on the departure disappears for aircraft winds (0.96–1.04 in every bin) and the OOPS total reproduces the IFS one: the ratio of the totals rises from 0.78 to 0.97 over the two cycles (from 0.79 to 0.96 at 00 UTC and from 0.77 to 0.99 at 12 UTC), and the observation types that carry the weights, conventional data and all-sky radiances, move to 0.91–1.08 of the IFS impact. Per observation, too, the conventional data agree much better without the weights, above all in amplitude, and the amplitude of the all-sky impacts comes close to that of the IFS, although their correlation stays low (table 7).

Table 5: OOPS/IFS ratio of the total impact by observation type, with the weights in the OOPS impact (default) and without them (matched definition), 15 May 2026, 00 and 12 UTC. Ground-based radar precipitation (obstype 14) is included in the total but not listed: in these two cycles it has only 2 913 observations and an IFS total impact of −1.1, too small for a meaningful ratio.

Observation type	Default	Matched
Surface stations (SYNOP)	0.47	0.91
Aircraft (AIREP)	0.61	0.99
Radiosondes (TEMP)	0.68	1.07
Drifting buoys (DRIBU)	0.68	1.01
Wind profilers and pilot balloons (PILOT)	0.51	0.92
Atmospheric motion vectors (SATOB)	0.89	0.98
Scatterometer (SCAT)	0.82	0.97
GPS radio occultation	0.95	1.00
All-sky radiances	0.74	1.08
Clear-sky radiances	0.86	0.86
Total	0.78	0.97

Table 6: OOPS/IFS ratio of the summed impact of aircraft winds by normalised departure $|d_i/\sigma_i|$, with the weights in the OOPS impact (default) and without them (matched), cycle 2026051512.

$\lvert d_i/\sigma_i \rvert$	Observations	Default	Matched
0–0.5	828 464	0.85	0.96
0.5–1	550 811	0.88	0.99
1–2	407 392	0.71	1.02
2–3	82 791	0.49	1.04
≥ 3	24 219	0.33	1.03

Table 7: Per-observation agreement of the OOPS and IFS impacts, with the weights in the OOPS impact (default) and without them (matched), cycle 2026051512: correlation / amplitude ratio OOPS/IFS (section 4.2). Conventional data: range over the variables and layers that carry at least 1% of the IFS total impact; their totals are 0.85–1.12 of the IFS ones without the weights.

Observations	Default	Matched
Conventional data	0.79–0.90 / 0.34–0.62	0.93–0.99 / 0.97–1.26
All-sky radiances	0.67 / 0.66	0.68 / 1.08
Clear-sky radiances	0.86 / 0.97	0.86 / 0.97
2 m temperature	0.64 / 0.86	0.64 / 0.86

Some differences remain once the impact is defined in the same way. The per-observation agreement for 2 m temperature, which has no quality-control weights, is poor (table 7). It comes from the linear

model of the IFS FSOI minimisation, which still runs the simplified tangent-linear and adjoint physics (vertical diffusion and surface drag, `LSPHLC`) on top of the full linearised physics, and `LSLPHY`; OOPS uses the linearised physics of 4D-Var without them. In a test for cycle 2026051512 in the development experiments, with both switched off in the IFS minimisation, the correlation for 2 m temperature rose from 0.73 to 0.97 and the amplitude ratio from 0.82 to 1.04 (these experiments have the weights in the OOPS impact, which makes no difference for 2 m temperature), while the other conventional observations and the OOPS/IFS ratio of the totals (0.76) did not change. Clear-sky radiances, which have no quality-control weights either, are at 0.86 of the IFS impact in both OOPS calculations. In the development experiments their OOPS/IFS ratio was 0.88 with the original IFS minimisation and 0.86 with the simplified physics switched off. With the IFS minimisation in the gridpoint-humidity configuration of 4D-Var instead (gridpoint humidity and `LSPRT`, which needed a change to the reading of the gradient in the IFS), started from the OOPS gradient and with the simplified physics kept, it rose to 1.03 on the same observations (IASI from 0.58 to 1.03), while the ratio of the totals moved only from 0.76 to 0.77; the gradient alone changes it by about 1% (section 4.2). The remaining difference for clear-sky radiances therefore points to the spectral humidity or `LSPRT` of the IFS FSOI minimisation; the two cannot be separated, since the configuration sets both. All-sky radiances agree in total with the matched definition (1.08) but less per observation (table 7). The two systems also linearise the minimisation around different trajectories (section 3.1): the IFS interpolates the high-resolution trajectory, whereas OOPS runs a low-resolution trajectory forced towards it, which may matter most in cloud; this has not been tested.

4.5 Comparison with the measured change in forecast error

This section asks whether the summed impacts reproduce the measured change in forecast error and, where they do not, whether the shortfall lies in the decomposition by observation or in the Langland–Baker estimate that is decomposed. The forecast errors e_a and e_b of (4) were computed from the forecast-error fields of each cycle with the IFS norm, which reproduces the values the IFS prints to a relative difference of 1.5×10^{-5} (in the two cycles whose logs are kept).

Because the error norm and the FSOI problem refer to different assimilations (section 3.1), the Langland–Baker estimate $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}$ can be evaluated with three increments at the start of the window (figure 1):

- $\delta\mathbf{x}_a = \mathbf{x}_a - \mathbf{x}_b$, the long-window increment, which the impacts decompose;
- $\delta\mathbf{x}_a^e$, the early-delivery 4D-Var analysis minus the same background, the increment of the analysis from which $\mathbf{x}_a^f$ starts;
- $\delta\mathbf{x}_a^f$, the early-delivery analysis minus the state of $\mathbf{x}_b^f$ at the start of the window, which corresponds to the measured change $e_a - e_b$.

We computed the three estimates with the OOPS gradients and the operational analyses and forecasts, at T255; the IFS impacts come from the IFS gradients, which change the total impact by less than 1% (section 4.2), so the same estimate serves for both. The spectral variables use the scalar product (13), and humidity, which is a grid-point field in the operational analysis, an area-weighted mean on a regular Gaussian grid; the ozone gradient is zero. Table 8 gives the results for the 34 cycles.

Do the impacts reproduce the measured change? No. Both definitions underestimate the measured reduction in forecast error, the OOPS one more than the IFS one, and so does the Langland–Baker estimate itself: with $\delta\mathbf{x}_a^f$, the increment that corresponds to the measured change, it accounts for 0.68 of

Table 8: Langland–Baker estimates for the three increments and the summed impacts, 34 cycles from 2026051500 to 2026053112 (units of e). Ratios of the sums, with the range of the per-cycle ratios in parentheses. Over the 34 cycles, $e_a - e_b = -6.65 \times 10^6$ and the summed impacts are -4.35×10^6 for OOPS with the weights (0.65 of $e_a - e_b$, per cycle 0.55–0.77) and -5.59×10^6 for the IFS without them (0.84, per cycle 0.71–1.00).

Increment	Estimate	/ $(e_a - e_b)$	OOPS / estimate	IFS / estimate
$\delta\mathbf{x}_a$ (long window)	-5.29×10^6	0.80 (0.69–0.95)	0.82 (0.78–0.86)	1.06 (1.00–1.13)
$\delta\mathbf{x}_a^e$ (early delivery)	-2.29×10^6	0.34 (0.30–0.41)	1.90 (1.65–2.10)	2.44 (2.12–2.76)
$\delta\mathbf{x}_a^f$ (error norm)	-4.51×10^6	0.68 (0.62–0.74)	0.96 (0.83–1.11)	1.24 (1.07–1.42)

it. Its gradients come from a 27-hour T255 adjoint integration along the forecast from the long-window analysis, with the linearised physics of 4D-Var except the linearised surface scheme (section 3.2). Shortfalls of this kind have been reported before, about 25% in the NRL system without moist processes in the adjoint (Langland and Baker, 2004) and about 15% in GEOS-5 (Gelaro *et al.*, 2007), though with other norms, forecast ranges and verifying analyses, so the fractions are not directly comparable. The estimates with the other two increments are not estimates of $e_a - e_b$, since neither increment separates the two forecasts that are verified.

Do the impacts add up to the estimate they decompose? Approximately. With the long-window increment $\delta\mathbf{x}_a$, the OOPS impacts add up to 0.82 of the estimate and the IFS ones to 1.06, and both follow it closely from cycle to cycle ($r = 0.93$ and 0.86). The definition with the weights falls 18% short. The FSOI problem is not the problem that produced $\delta\mathbf{x}_a$, and it differs from it in several approximations shared by the IFS and OOPS:

- it is linearised around the background, whereas the last outer loop of 4D-Var is linearised around $\mathbf{x}_{K-1}$, and it uses the background departures $\mathbf{d}$ rather than the effective departures of that loop (section 5);
- the quality-control weights are evaluated at the background departures, which are larger than the analysis residuals, so the weights are smaller than those of the last outer loop (section 2.3);
- it is a single minimisation at the resolution and time step of the first outer loop (TL255, 900 s), whereas the last outer loop of the operational 4D-Var runs at TL511 with a 450 s time step;
- the solve is truncated after a fixed number of iterations, and the impacts add up to $\mathbf{b}^{\mathrm{T}}\mathbf{v}_k$ rather than to $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\tilde{\mathbf{x}}$ (section 3.4).

Each of these can make the impacts differ from the estimate; which of them causes the shortfall has not been separated (section 5). The shortfall does not depend on the error norm: in a configuration consistent with the long-window assimilation (next paragraphs), the gradients change substantially, but the ratios are 0.80 and 1.06, against 0.83 and 1.07 for the same cycles in the original configuration.

The definition without the weights comes closer because it credits observations that variational quality control weighted down with their full departure, which offsets the shortfall. The two definitions use the same FSOI solution $\mathbf{z}$, and their sums differ by

$$\sum_i d_i(1 - w_i)\left[\mathbf{R}^{-1}\mathbf{H}\mathbf{z}\right]_i, \tag{11}$$

the impact credited to the part of each departure that quality control removed. Only the definition with the weights adds up, for a converged solve, to $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\tilde{\mathbf{x}}$ of the frozen-weight problem (9); without

them, the impacts differ from it by this term, whatever the approximations of the FSOI problem. Here the IFS impacts exceed the OOPS ones in every cycle, by 22–39% (1.29 over the 34 cycles), and since the two systems give nearly the same impacts with the same definition (section 4.4), this is the size of the term. Removing the approximations above would therefore bring the impacts with the weights towards the estimate, and leave those without them beyond it by this term. The closer total of the IFS definition is a compensation of two errors, not evidence for it, and the totals do not decide between the two definitions. The definition with the weights is the one consistent with the frozen-weight problem that the FSOI minimisation solves (section 2.3); neither is the sensitivity of the nonlinear, quality-controlled analysis. The close agreement of the OOPS impacts with the estimate for $\delta\mathbf{x}_a^f$ (0.96) is not a closure either: the impacts do not decompose that increment.

Contribution of $\mathbf{g}_a$. For the two early-delivery increments, $\mathbf{g}_a^{\mathrm{T}}\delta\mathbf{x}$ is at most 8% of $\mathbf{g}_b^{\mathrm{T}}\delta\mathbf{x}$ in every cycle (0% and 2% over the 34 cycles), but for the long-window increment it is 44%. Since $\mathbf{g}_a$ here combines the error of the early-delivery forecast with an adjoint along the forecast from the long-window analysis, this is not a clean test of stationarity; in the configuration consistent with the long-window assimilation of the next paragraph, $\mathbf{g}_a^{\mathrm{T}}\delta\mathbf{x}_a$ is only 2–8% of $\mathbf{g}_b^{\mathrm{T}}\delta\mathbf{x}_a$.

Consistency of the error norm and the FSOI problem. In the FSOBS configuration the impacts decompose the long-window increment while the error norm verifies early-delivery forecasts, and $\mathbf{x}_b^f$ does not start from the background. Configurations consistent with either assimilation are outlined in section 5 (step E). As a diagnostic, we ran a configuration consistent with the long-window assimilation for the four cycles of 15 and 16 May (experiments jdkc and jdkd), with $\mathbf{x}_a^f$ the 18-hour forecast from the long-window analysis valid 9 hours into the window, $\mathbf{x}_b^f$ the 30-hour forecast from the one valid 9 hours into the previous window, $\mathbf{x}_v$ the long-window analysis, and the OOPS gradients for the estimate. The estimate gives 0.67 of the measured change, close to the 0.70 of the increment that matches the original error norm for the same cycles, and the impacts add up to 0.80 of the estimate for OOPS and 1.06 for the IFS (per cycle 0.79–0.82 and 1.05–1.07), against 0.83 and 1.07 for jdbt and jdbs. A consistent error norm therefore changes neither how much of the measured change the estimate accounts for, about two thirds, nor how far the impacts are from the estimate they decompose.

Caveats. The comparison is for the total: $\frac{1}{2}(\mathbf{g}_a+\mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}$ cannot be split by observation, so it does not show whether the impacts of each observation type are right. And the measured reduction depends on the verifying analysis, here the operational analysis 24 hours later, from the same cycling as the analysis being verified: through the cycling its background descends from $\mathbf{x}_a$, so it is not independent of $\mathbf{x}_a$. Since $e_a - e_b = \frac{1}{2}(\mathbf{x}_a^f - \mathbf{x}_b^f)^{\mathrm{T}}\mathbf{C}(\mathbf{x}_a^f + \mathbf{x}_b^f - 2\mathbf{x}_v)$ is linear in the error of $\mathbf{x}_v$, a verifying analysis whose error is correlated with that of $\mathbf{x}_a$ (more than with that of $\mathbf{x}_b$) favours the forecast from the analysis, and the measured reduction would then be too large relative to an independent verification. The gradients $\mathbf{g}_a$ and $\mathbf{g}_b$ also depend on $\mathbf{x}_v$, but the impacts and the Langland–Baker estimates use the same gradients, so their ratios are much less sensitive to the verifying analysis than their ratios to $e_a - e_b$.

5 Conclusions and outlook

FSOI is now available in OOPS. The sensitivity gradient is computed by a generic OOPS application, and the sensitivity to the observations and the impacts by an FSOI mode of the square-root minimisation, with the same linear models, observation operators and quality control as 4D-Var, apart from the linearised

surface scheme in the sensitivity runs. The sensitivity gradient is a dual vector and must be moved between grids with the adjoint transforms. Compared with the IFS, the norm-weighted forecast error $\mathbf{C}\varepsilon$ agrees with correlation and amplitude ratio 1.0000, the gradients agree closely at large and medium scales, and, in the two cycles tested, the total impacts agree once they are defined in the same way (0.97).

With the quality-control weights in the impact, which OOPS uses by default, the impact of observations with large departures, mostly conventional, is smaller than in the IFS. Part of this reduction may come from evaluating the weights at the background departures rather than at the analysis residuals (step B). The remaining differences for 2 m temperature and clear-sky radiances are traced, in single-cycle tests, to legacy settings of the IFS FSOI minimisation that 4D-Var no longer uses: the simplified linearised physics for 2 m temperature, and spectral humidity with `LSPRT` off for clear-sky radiances, although the contributions of spectral humidity and `LSPRT` to the clear-sky difference cannot be separated. The per-observation agreement for all-sky radiances remains low (correlation 0.68 with the matched definition) and is not explained; the different linearisation trajectories of the two systems have not been tested. Neither definition reproduces the measured change in forecast error.

The comparison with the Langland–Baker estimate exposed an inconsistency in the FSOBS configuration, shared by the IFS and OOPS: the impacts decompose the long-window increment, whereas the error norm verifies forecasts from the early-delivery analyses. It explains neither of the two gaps that remain, which have about the same size in a configuration consistent with the long-window assimilation. Removing it (step E) is nonetheless the most important change needed for a consistent FSOBS, so that the impacts and the error norm refer to the same assimilation; whether that should be the early-delivery assimilation, whose analyses start the operational forecasts, or the long-window assimilation, which cycles, remains to be decided. The Langland–Baker estimate accounts for about two thirds of the measured change (0.68 with the increment that matches the error norm, 0.67 in the consistent configuration). Much of this gap may originate in the gradient calculation, which uses a T255 adjoint with the linearised physics, without the linearised surface scheme, along the forecast from $\mathbf{x}_a$ for both gradients (section 3.2), whereas the verified forecasts run at TCo1279. The impacts with the weights fall about 20% short of the estimate they decompose (0.82, and 0.80 in the consistent configuration, against 1.06 for the IFS); this shortfall is associated with approximations of the FSOI problem shared by both systems, whose contributions have not been separated: linearisation around the background, weights at the background departures, a single truncated minimisation, and the resolution of the first outer loop of 4D-Var (TL255, 900 s) rather than the last (TL511, 450 s). The impacts without the weights come closer only because they credit the part of the departures that quality control removed, which offsets this shortfall; removing the approximations would leave them beyond the estimate. The totals therefore do not decide between the definitions. The definition with the weights is the one consistent with the frozen-weight quadratic problem that the FSOI minimisation solves, the only one whose impacts add up to the linearised change of that problem, and this is why OOPS uses it by default. Neither definition is the sensitivity of the nonlinear, quality-controlled analysis (section 2.3).

The main limitation is shared by both systems. The analysis is produced by multi-incremental 4D-Var: several outer loops, each relinearising the observation operators and the model around the updated trajectory and re-evaluating the quality-control weights, with increasing resolution (TL255, TL319, TL399 and TL511 in the operational configuration). FSOBS runs a single minimisation at TL255 linearised around the background, with the weights at the background departures. If the outer loops have reached a fixed point, the inner loops are converged and all loops use the same resolution, the sensitivity of the analysis to the observations depends only on the last outer loop, linearised at $\mathbf{x}_{K-1}$, so the outer loops need not be repeated; what matters is the linearisation point and the resolution of that loop. (This is a Gauss–Newton argument: it gives the sensitivity of the final quadratic problem, neglecting the depen-

dence of $\mathbf{H}$ and of the weights on $\mathbf{x}_{K-1}$, and that of $\mathbf{x}_{K-1}$ on the observations through the earlier outer loops.) In that loop the increment $\mathbf{x}_a - \mathbf{x}_b$ solves the linear problem with the Hessian at $\mathbf{x}_{K-1}$, the weights at the analysis residuals, and the effective departures

$$\mathbf{d}_{\text{eff}} = \mathbf{y} - \mathscr{H}(\mathbf{x}_{K-1}) + \mathbf{H}_K(\mathbf{x}_{K-1} - \mathbf{x}_b) = \mathbf{d} + [\mathscr{H}(\mathbf{x}_b) - \mathscr{H}(\mathbf{x}_{K-1}) + \mathbf{H}_K(\mathbf{x}_{K-1} - \mathbf{x}_b)]\,. \tag{12}$$

The bracket is the nonlinearity residual of the observation operator, which may not be small. With $\mathbf{d}_{\text{eff}}$ and a converged FSOI solve, the impacts add up to $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}_a$ for the actual analysis increment. With the background departures $\mathbf{d}$ and the Hessian and the weights at the analysis, the impacts stay a decomposition by observation of the first-guess departures, but they no longer add up exactly. With truncated inner loops the exact sensitivity would require the adjoint of each minimisation through the outer loops (cf. Trémolet, 2008), which is only worth pursuing if linearisation at the analysis is not sufficient.

The following steps are possible future work, outside the scope of porting FSOI to OOPS. They would separate the approximations of section 4.5 and explain the remaining deficiencies. The current FSOI problem is that of the first inner loop of 4D-Var, linearised around the background at TL255 with the weights at the background departures, whereas $\delta\mathbf{x}_a$ results from the last one; steps B to D move it towards the last inner loop. The first three are single-factor tests for a few cycles, each changing one approximation of the FSOI problem and comparing the summed impacts with $\frac{1}{2}(\mathbf{g}_a + \mathbf{g}_b)^{\mathrm{T}}\delta\mathbf{x}_a$ and by observation type. The approximations may interact, so the effects found in single-factor tests need not add up to the whole shortfall. Step E is of a different kind: it removes the inconsistency between the error norm and the FSOI problem rather than testing an approximation. Step F concerns the IFS only.

A. Convergence of the FSOI solve: rerun the minimisation with, for example, half and twice the 62 iterations of the current configuration.

B. Quality-control weights at the analysis residuals instead of the background departures, with the Hessian otherwise unchanged. Because variational quality control has a large effect on the impact totals and the background departures are generally larger than the analysis residuals, evaluating the weights at the analysis may account for part of the shortfall.

C. Resolution and time step of the last outer loop of 4D-Var (TL511, 450 s) instead of the first (TL255, 900 s), for the sensitivity runs as well as for the minimisation. For the sensitivity runs this also tests how much of the gap between the Langland–Baker estimate and the measured change comes from the resolution of the gradients.

D. FSOI linearised around the analysis (as a proxy for $\mathbf{x}_{K-1}$): a high-resolution trajectory from the 4D-Var analysis to provide the linearisation for the FSOI minimisation, the residuals for the quality-control weights and the departures $\mathbf{d}_{\text{eff}}$ (12), while the background trajectory keeps providing the screening and $\mathbf{d}$. Together with C and a converged solve, this is the problem of the last inner loop, for which the impacts with the weights add up to the estimate. Assess the differences from the current impacts by observation type and as a function of the departure, for both $\mathbf{d}$ and $\mathbf{d}_{\text{eff}}$.

E. Make the error norm and the FSOI problem refer to the same assimilation. In either case $\mathbf{x}_b^f$ is the forecast from the background $\mathbf{x}_b$, and $\mathbf{g}_b$ is linearised along it (section 3.2), which the third-order accuracy of the Langland–Baker estimate requires. For the early-delivery assimilation, whose analyses start the operational forecasts and whose forecasts the error norm verifies now, decompose the early-delivery increment with the observations and window of that assimilation and the adjoint along the forecast from its analysis. For the long-window assimilation, whose increment FSOBS

decomposes now, verify forecasts from the long-window analyses against the long-window analysis, as in the diagnostic of section 4.5 (jdkc, jdkd). Which of the two FSOBS should follow remains to be decided.

F. In the IFS, update the FSOBS minimisation to the 4D-Var configuration (no simplified physics, gridpoint humidity), which needs the change to the reading of the gradient, and repeat the comparison of the impacts with OOPS over the full period, to confirm the single-cycle attribution of the 2 m temperature and clear-sky differences (section 4.4). This is worthwhile only while the IFS FSOBS remains in use alongside OOPS.

Acknowledgements

We thank Alexis Doerenbecher (Météo-France) for sharing the OOPS-based FSOI implementation at Météo-France and the associated technical documentation. We also thank Massimo Bonavita, Alan Geer, Bruce Ingleby and Cristina Lupu, whose review comments helped us improve this memorandum.

References

Andersson, E. and Järvinen, H. (1999). Variational quality control. *Q. J. R. Meteorol. Soc.*, **125**(554), 697–722, doi:10.1002/qj.49712555416.

Cardinali, C. (2009). Monitoring the observation impact on the short-range forecast. *Q. J. R. Meteorol. Soc.*, **135**(638), 239–250, doi:10.1002/qj.366.

Courtier, P., Thépaut, J.-N. and Hollingsworth, A. (1994). A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.*, **120**(519), 1367–1387, doi:10.1256/smsqj.51911.

Errico, R. M. (2007). Interpretations of an adjoint-derived observational impact measure. *Tellus A*, **59**(2), 273–276, doi:10.1111/j.1600-0870.2006.00217.x.

Gelaro, R. and Zhu, Y. (2009). Examination of observation impacts derived from observing system experiments (OSEs) and adjoint models. *Tellus A*, **61**(2), 179–193, doi:10.1111/j.1600-0870.2008.00388.x.

Gelaro, R., Zhu, Y. and Errico, R. M. (2007). Examination of various-order adjoint-based approximations of observation impact. *Meteorol. Z.*, **16**(6), 685–692, doi:10.1127/0941-2948/2007/0248.

Langland, R. H. and Baker, N. L. (2004). Estimation of observation impact using the NRL atmospheric variational data assimilation adjoint system. *Tellus A*, **56**(3), 189–201, doi:10.3402/tellusa.v56i3.14413.

Rabier, F., Klinker, E., Courtier, P. and Hollingsworth, A. (1996). Sensitivity of forecast errors to initial conditions. *Q. J. R. Meteorol. Soc.*, **122**(529), 121–150, doi:10.1002/qj.49712252906.

Tavolato, C. and Isaksen, L. (2015). On the use of a Huber norm for observation quality control in the ECMWF 4D-Var. *Q. J. R. Meteorol. Soc.*, **141**(690), 1514–1527, doi:10.1002/qj.2440.

Trémolet, Y. (2008). Computation of observation sensitivity and observation impact in incremental variational data assimilation. *Tellus A*, **60**(5), 964–978, doi:10.1111/j.1600-0870.2008.00349.x.

A Configuration

The implementation is in the branches of ifs-source, ifs-scripts, ifs-suites and OOPS for JIRA issue IFS-4937 (ecmwf-ifs/oops#22, ifs-source#275, ifs-scripts#409, ifs-suites#152). The experiments were built from the commits in table 9. The ifs-source commit of jdbt is five commits after that of jdbs; these commits do not change the results (bit-identical).

Table 9: Commits from which the experiments were built (ifs-bundle 87fba48dc1 for both).

Repository	jdbs (IFS)	jdbt (OOPS)
ifs-source	52e29a8a7e	9b90c2df8a
OOPS	96d96eb3da	de08821189
ifs-scripts	0d2178a6fc	0a3b5b567f
ifs-suites	e48ffb7375	e48ffb7375

In the IFS suite, the IFS adjoint tasks `J1` and `J1back` (IFS configuration 801) are replaced in OOPS mode by the `oopsvar` task in sensitivity mode (`mode=sensitivity`, `FSOI_MODE=an|fc`), running the executable `ifsForecastSensitivity`. The impacts are written to the ODB column `fc_sens_obs` and archived in MARS as `type=FSOIFB`. The main settings are listed in table 10.

Table 10: Main configuration settings of OOPS FSOBS.

Where	Key	Purpose
sensitivity run	`norm_matrix`	energy norm $\mathbf{C}$
sensitivity run	`output_initial_sensitivity`	gradient at t_0
sensitivity run, FieldsIFS	`write_spec_ad: true`	write with $\mathbf{S}^{\mathrm{T}}$
sensitivity run, NAMTRAJP	`LESURF2=.false.`	no linearised surface scheme
minimisation	`fsoi.forecast_sensitivity_an`, `fsoi.forecast_sensitivity_bg`	$\mathbf{g}_a$, $\mathbf{g}_b$
minimisation, FieldsIFS	`read_spec_ad: true`	read with $(\mathbf{S}^{-1})^{\mathrm{T}}$
minimisation	`fsoi.varqc_weights_in_impact`	$\mathbf{W}$ in the impact (default `true`)
minimisation	`fsoi.obs_sensitivity`	output of δe_i

B The energy norm in spectral and in grid-point space

The IFS sensitivity configuration applies the norm (2) to spectral coefficients, in the routine `cosjr`. OOPS applies it to grid-point fields, in `norm_matrix_mod`, so that it works with the grid-point fields of OOPS and on any grid. This appendix gives the two forms and the conditions under which they agree.

B.1 Spectral form

A field f at truncation T is represented by the coefficients f_n^m, $0 \le m \le n \le T$, each stored once as a real and an imaginary part. With the normalisation of the IFS spectral transforms, $\frac{1}{2}\int_{-1}^{1}(P_n^m)^2\,\mathrm{d}\mu = 1$, the

global mean of the product of two fields is

$$\overline{fh} = \sum_{m,n} \zeta_m \,\mathrm{Re}\left(f_n^m \overline{h_n^m}\right), \qquad \zeta_0 = 1,\ \zeta_{m>0} = 2, \tag{13}$$

which defines the spectral scalar product. For the wind, the mean kinetic energy follows from vorticity ξ and divergence D as $\overline{u^2+v^2} = \sum_{m,n\geq 1} \zeta_m a^2/[n(n+1)]\,(|\xi_n^m|^2 + |D_n^m|^2)$, with a the radius of the Earth. The spectral form of (2) is therefore

$$e = \frac{1}{2}\sum_k \sum_{m,n} \zeta_m \Delta p_k \left[\frac{a^2}{n(n+1)}\left(|\xi_{n,k}^m|^2 + |D_{n,k}^m|^2\right) + \frac{c_p}{T_r}|T_{n,k}^m|^2\right] + \frac{1}{2}\sum_{m,n} \zeta_m R_d T_r p_r\, |(\ln p_s)_n^m|^2, \tag{14}$$

where the wind terms are summed over $n \geq 1$ only, and Δp_k is the thickness of layer k in the reference atmosphere with surface pressure p_r. `cosjr` evaluates (14) and returns the coefficients multiplied by their weights, the gradient of e with respect to the spectral scalar product. The weights are diagonal in spectral space.

B.2 Grid-point form

On a grid with points j and Gaussian quadrature weights $w_j = \rho_\ell / N_\ell$, where ρ_ℓ is the Gaussian weight of latitude ℓ ($\sum_\ell \rho_\ell = 1$) and N_ℓ its number of points, the global mean is approximated by $\overline{f} \approx \sum_j w_j f_j$, and

$$e \approx \frac{1}{2}\sum_k \sum_j w_j \Delta p_k \left[u_{jk}^2 + v_{jk}^2 + \frac{c_p}{T_r} T_{jk}^2\right] + \frac{1}{2}\sum_j w_j R_d T_r p_r\, (\ln p_s)_j^2. \tag{15}$$

The norm matrix $\mathbf{C}$ is diagonal on the grid, with entries $w_j \Delta p_k$ for u and v, $w_j \Delta p_k c_p/T_r$ for T and $w_j R_d T_r p_r$ for $\ln p_s$; the humidity term has weight zero. The forecast error is read as spectral coefficients of T, $\ln p_s$ and the wind, and transformed to the O256 grid of the sensitivity run, where $\mathbf{C}$ is applied.

Because $\mathbf{C}$ contains the quadrature weights w_j, $\mathbf{C}\boldsymbol{\varepsilon}$ is the gradient of e with respect to the grid-point values, a dual vector (section 3.3), not a field.

B.3 Agreement of the two forms

The two forms agree if the grid-point quadrature integrates the products of the fields exactly, which is also the condition $\mathbf{S}^{-1}\mathbf{S} = \mathbf{I}$ of section 3.3, since $\mathbf{S}^{\mathrm{T}}\mathbf{Q}\mathbf{S}$ is the spectral representation of the quadrature. In latitude, Gaussian quadrature on the 512 latitudes of the O256 grid is exact for polynomials in $\mu = \sin\phi$ of degree up to 1023, more than the degree, at most 510, of a product of two fields at T255. In longitude, the product contains zonal wavenumbers up to twice the truncation, which the rows of a reduced grid near the poles, with $4\ell + 16$ points on the ℓ-th row of the octahedral grid, do not resolve; there, however, the associated Legendre functions of large m are very small. The two forms therefore agree to within the precision of the transforms rather than exactly by construction.

The spectral gradient follows from the grid-point one by the chain rule: if $\boldsymbol{\varepsilon}_g = \mathbf{S}\boldsymbol{\varepsilon}_s$, the gradient with respect to the spectral coefficients is $\mathbf{S}^{\mathrm{T}}\mathbf{C}\boldsymbol{\varepsilon}_g$, which is what the sensitivity run writes with the adjoint of the inverse transform (`write_spec_ad`). Written this way at the verification time, the OOPS $\mathbf{C}\boldsymbol{\varepsilon}$ was compared with the output of `cosjr` for the same forecast error (jbv7/jbv8, cycle 2026051512): correlation and amplitude ratio are 1.0000 for vorticity, divergence, temperature and surface pressure, over all coefficients and separately for $m = 0$ and $m > 0$ (and $n = 0$ for temperature and surface pressure; the $n = 0$ coefficient of vorticity and divergence is zero), and 1.00 in every band of model levels.

A grid-point norm also makes regional or masked norms straightforward, since a mask is a multiplication on the grid; the code has an option for this, which is not used here.