Single-Precision in the Tangent-Linear and Adjoint Models of Incremental 4D-Var

SAM HATFIELD, ANDREW MCRAE, AND TIM PALMER

Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

PETER DÜBEN

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 1 September 2019, in final form 17 December 2019)

ABSTRACT

The use of single-precision arithmetic in ECMWF's forecasting model gave a 40% reduction in wall-clock time over double-precision, with no decrease in forecast quality. However, using reduced-precision in 4D-Var data assimilation is relatively unexplored and there are potential issues with using single-precision in the tangent-linear and adjoint models. Here, we present the results of reducing numerical precision in an incremental 4D-Var data assimilation scheme, with an underlying two-layer quasigeostrophic model. The minimizer used is the conjugate gradient method. We show how reducing precision increases the asymmetry between the tangent-linear and adjoint models. For ill-conditioned problems, this leads to a loss of orthogonality among the residuals of the conjugate gradient algorithm, which slows the convergence of the minimization procedure. However, we also show that a standard technique, reorthogonalization, eliminates these issues and therefore could allow the use of single-precision arithmetic. This work is carried out within ECMWF's data assimilation framework, the Object Oriented Prediction System.

1. Introduction

The development of efficient and scalable high-performance computing systems for Earth system simulation is becoming increasingly important for the advancement of Earth system science, as model users demand ever larger ensemble sizes, higher spatial and temporal resolution, and assimilation of vastly more data than before (Lawrence et al. 2018). This is especially the case for numerical weather prediction centers, such as the European Centre for Medium-Range Weather Forecasts (ECMWF), who must receive, process, and assimilate observations and then compute and distribute a forecast within a certain "time critical path" (around 2 h in the case of ECMWF). While model user demands increase, the nature of high-performance computing is itself shifting. Supercomputers are increasingly heterogeneous with, for example, graphics processing units becoming commonplace. Additionally, communication and memory transfer are rapidly becoming the main bottlenecks for Earth system simulations and methods must be devised to accelerate these processes if software developments are to keep pace with hardware developments.

The use of single-precision floating-point arithmetic instead of double-precision has been suggested in previous publications as a way to reduce the computational cost of Earth system simulation (Düben and Palmer 2014). Of course, single-precision computations are less precise than double-precision computations. However, any error introduced by lowering precision should always be compared with inherent, unavoidable sources of uncertainty, such as model error and initial condition error. A single-precision variant of the ECMWF model, the Integrated Forecasting System (IFS), compared favorably with the double-precision model, with a wall-clock time reduction of around 40% yet with minimal impact on forecast skill (Váňa et al. 2017); the model is currently undergoing testing to ensure its appropriateness for operational weather forecasting at ECMWF. Similar results were obtained with the COSMO model (Rüdisühli et al. 2014) and the NICAM model (Nakano et al. 2018).

Corresponding author: Sam Hatfield, samuel.hatfield@physics.ox.ac.uk

DOI: 10.1175/MWR-D-19-0291.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

The nonlinear forward model is only one part of an operational numerical weather prediction (NWP) system, however. The framework and algorithms for processing observational data and constructing initial conditions, known as data assimilation, form roughly half of the computational cost of NWP. During data assimilation, one makes use of an assimilation model as the model allows one to fill in gaps in observational data and propagate observational information forward through time. For some algorithms, such as the ensemble Kalman filter, the assimilation model is essentially the same as the forecasting model. Then, a substantial precision reduction, even below single-precision, can be accommodated (Hatfield et al. 2018b,a). For others, such as incremental 4D-Var, however, one makes use of the linearized forecast models: the tangent-linear model and its adjoint. The nature of these models is sufficiently different to that of the nonlinear forecasting model that it is not clear how they would be affected by the use of single-precision instead of double-precision. Given the potential for a similar 40% reduction in computational cost (and the possibility of reinvesting the saved resources to further improve the initial conditions), we believe this is an important question to answer.

In this paper, we aim to identify some of the issues that might be encountered when using single-precision floating-point arithmetic in 4D-Var. We do not tackle a full-complexity model such as the IFS but instead focus on a medium-complexity quasigeostrophic model in order to easily identify general issues. 4D-Var is essentially a gradient-descent minimization problem that is accelerated using a Krylov subspace method such as conjugate gradients or generalized minimal residual. The impact of rounding errors on Krylov methods is not a new subject [see, e.g., the review paper Meurant and Strakoš (2006)]. However, we are not aware of any study that directly assesses the impacts of rounding errors within the tangent-linear and adjoint models of 4D-Var on the convergence process. As such, our intention here is to highlight the relevant concepts from the field of Krylov methods that will likely be important when considering such a precision reduction. In section 2 we summarize the incremental 4D-Var algorithm and discuss the use of linearized models. In section 3 we introduce our experimental setup. In section 4 we investigate the standard tangent-linear/adjoint model consistency test as precision is reduced. In section 5 we show how reducing precision affects the rate of convergence and in section 6 we suggest how this might be remedied. Finally, we conclude in section 7.

2. Reducing precision in 4D-Var

Given a set of observations distributed over a window of time, 4D-Var informally finds the trajectory that minimizes a weighted sum of the distances between the trajectory and the observations and between the trajectory and an initial guess (the background), while remaining physically consistent with the known dynamical laws of the system. In this section, we focus on the incremental form of 4D-Var (Courtier et al. 1994), whereby a global, potentially nonconvex minimization problem is split into a series of simpler, quadratic minimization problems known as ''inner loops.'' We refer the reader to Lawless et al. (2005) for a summary of the incremental 4D-Var algorithm.

The incremental 4D-Var algorithm requires not just the nonlinear model operator, in order to compute the innovations across the assimilation window, but also its tangent-linear and adjoint, for propagating increments forward through time and gradients backward through time, respectively. If we assume that observations are distributed across the window equidistant in time with a spacing of m model timesteps, then the nonlinear model operator that advances the model state $\mathbf{x}_0$ from the beginning of the window up to the ith observation is denoted by M_{im}:

$$\mathbf{x}_{im} = M_{im}(\mathbf{x}_0). \tag{1}$$

The tangent-linear model is the Jacobian of this operator evaluated about $\mathbf{x}_0$, denoted by $\mathbf{M}_{im}(\mathbf{x}_0)$, and the adjoint is simply its transpose, denoted by $\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)$. Written component-wise, the tangent-linear model is given by

$$[\mathbf{M}_{im}(\mathbf{x}_0)]_{j,k} = \frac{\partial(\mathbf{x}_{im})_j}{\partial(\mathbf{x}_0)_k} = \frac{\partial[M_{im}(\mathbf{x}_0)]_j}{\partial(\mathbf{x}_0)_k}, \tag{2}$$

where $(\mathbf{x}_{im})_j$ and $(\mathbf{x}_0)_k$ are the jth and kth elements of the vectors $\mathbf{x}_{im}$ and $\mathbf{x}_0$, respectively. For large problems, $\mathbf{M}_{im}(\mathbf{x}_0)$ is not formed explicitly, but only the *action* of this on an initial perturbation $\boldsymbol{\delta}\mathbf{x}_0$ [i.e., $\mathbf{M}_{im}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{x}_0$]. The tangent-linear and adjoint models are usually implemented as separate models by differentiating the code underlying M_{im} line-by-line (Lawless 2013). They are formed through a chain-rule decomposition in time. For the tangent-linear model, this decomposition is given by

$$\begin{aligned}\mathbf{M}_{im}(\mathbf{x}_0) &= \mathbf{M}_1(\mathbf{x}_{im-1})\mathbf{M}_{im-1}(\mathbf{x}_0)\\ &= \mathbf{M}_1(\mathbf{x}_{im-1})\mathbf{M}_1(\mathbf{x}_{im-2})\cdots\mathbf{M}_1(\mathbf{x}_1)\mathbf{M}_1(\mathbf{x}_0).\end{aligned} \tag{3}$$

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

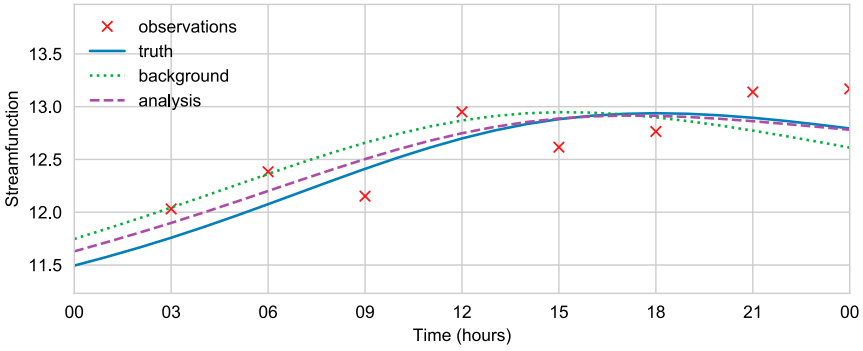

FIG. 1. The 4D-Var reference experiment. The streamfunction at a randomly chosen observation location is plotted.

In other words, between 0 and *im* we must evaluate the action of the Jacobian at each model time step. We use this chain-rule decomposition because, in reality, a numerical model only provides the model operator for a single time step: M_1. The Jacobian $\mathbf{M}_{im}(\mathbf{x}_0)$ is therefore not directly available, only $\mathbf{M}_1(\mathbf{x})$. The adjoint model is formed simply by taking the transpose of Eq. (3). Considering that the transpose of the product of two linear operators is the reversed product of the individually transposed operators, the adjoint model is given by

$$\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0) = \mathbf{M}_1^{\mathrm{T}}(\mathbf{x}_0)\mathbf{M}_1^{\mathrm{T}}(\mathbf{x}_1)\cdots\mathbf{M}_1^{\mathrm{T}}(\mathbf{x}_{im-2})\mathbf{M}_1^{\mathrm{T}}(\mathbf{x}_{im-1}). \quad (4)$$

Again, for large problems, the code evaluates the action of $\mathbf{M}_1^{\mathrm{T}}(\mathbf{x})$ on a vector, rather than forming $\mathbf{M}_1^{\mathrm{T}}(\mathbf{x})$ explicitly.

A Krylov subspace method such as conjugate gradients (CG) [see, e.g., Golub and Van Loan (1986)] is typically used to minimize the cost function of incremental 4D-Var. Notably, these methods only require matrix actions $\mathbf{Sw}$, where $\mathbf{S}$ is the Hessian of the cost function and $\mathbf{w}$ is an arbitrary vector, and do not require $\mathbf{S}$ to be formed explicitly. In incremental 4D-Var, this expression is given by

$$\mathbf{Sw} = (\mathbf{P}^b)^{-1}\mathbf{w} + \sum_{i=1}^{q} \mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\mathbf{M}_{im}(\mathbf{x}_0)\mathbf{w}, \quad (5)$$

where $\mathbf{P}^b$ is the background error covariance matrix, q is the number of observation vectors available across the window, $\mathbf{H}$ is the linearized observation operator, and $\mathbf{R}$ is the observation error covariance matrix (Haben 2011). For CG, the Hessian must be applied to the current search direction on every iteration of the minimization, which requires the tangent-linear and adjoint models to be run. This operation dominates the total computational cost. To save computational resources, the tangent-linear and adjoint models of the operational 4D-Var system of ECMWF are performed at a lower resolution than the full nonlinear model used to compute the innovations, with the resolution increasing as convergence is approached. We argue that using a lower precision in the tangent-linear and adjoint models could provide another way to lower the cost of the 4D-Var minimization. However, it is possible that the increased rounding errors from low-precision arithmetic could impact the convergence or stability of the minimization algorithm.

3. Experimental setup

To explore issues of reducing precision within 4D-Var, we use the Object Oriented Prediction System (OOPS) of ECMWF. This is a data assimilation framework written in C++, which is entirely model-agnostic—it can be applied both to idealized models and to full-complexity weather models such as the Integrated Forecasting System of ECMWF with relatively few code changes. We perform data assimilation experiments using a simple quasigeostrophic system as described below.

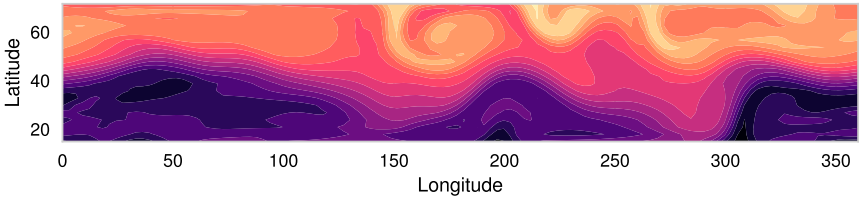

FIG. 2. A snapshot of top-layer quasigeostrophic potential vorticity q_1.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

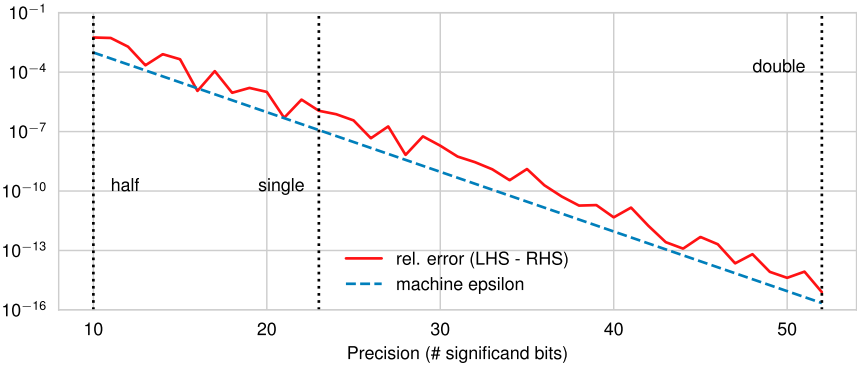

FIG. 3. The relative error between the left-hand side and right-hand side of the tangent-linear/adjoint model test, Eq. (10), as a function of numerical precision. The machine epsilon is also plotted as a function of precision for comparison.

a. Data assimilation setup

Our data assimilation setup consists of a strong-constraint incremental 4D-Var cost function minimized by the preconditioned conjugate gradient method (PCG) included in OOPS. For the background error covariance matrix $\mathbf{P}^b$ we choose the default matrix provided by OOPS—a standard correlation matrix with a horizontal length scale of 1000 km and a vertical correlation between the two layers of the model of 0.2. We also use $\mathbf{P}^b$ as the preconditioner matrix. We run the forecast model for 18 days to generate a nature run and perform assimilation over the final day. We assimilate observations of streamfunction, horizontal wind components, and wind speed that are generated at selected grid points by adding perturbations to the nature run drawn from a zero mean normal distribution with a diagonal observational error covariance matrix, $\mathbf{R}$. The observation error standard deviations for streamfunction, horizontal wind, and wind speed are 0.4, 0.6, and 1.2, respectively (these fields are nondimensionalized). The background is generated by drawing perturbations from a zero mean normal distribution with error covariance matrix $\mathbf{P}^b$. Figure 1 shows the assimilation window for the reference experiment, for which we use entirely double-precision. We perform three outer loop iterations and the inner loops terminate either when the norm of the cost function gradient has reduced by a factor of 100 or when 50 iterations have been performed. In none of the following experiments did the convergence process finish early due to the second termination criterion. Note also that, since we only consider the arithmetic of the tangent-linear and adjoint models when reducing precision, the computation of the reduction of the gradient norm is still performed with double-precision.

b. The two-layer quasigeostrophic system

We use the two-layer quasigeostrophic (QG) model provided by OOPS to generate the nature run and assimilate observations (Pedlosky 1979; Fandry and Leslie 1984). This model describes the advection of quasigeostrophic potential vorticity on two model layers q_1 and q_2 over a β-plane:

$$\frac{Dq_1}{Dt} = \frac{Dq_2}{Dt} = 0, \tag{6}$$

where

$$q_1 = \nabla^2\psi_1 - F_1(\psi_1 - \psi_2) + \beta y, \tag{7}$$

and

$$q_2 = \nabla^2\psi_2 - F_2(\psi_2 - \psi_1) + \beta y + R_s. \tag{8}$$

Here ψ_i is the streamfunction on layer i, β is the β-plane parameter, y is the y coordinate, R_s is a bottom forcing term representing orography, for example, and

$$F_i = \frac{f_0^2 L^2}{D_i g \Delta\theta/\overline{\theta}} \tag{9}$$

is the coupling term between the two layers defined from the Coriolis parameter at the center of the domain f_0, the length scale L, the depth of layer i D_i, the gravitational acceleration g, the difference in potential temperature across the layer interface $\Delta\theta$, and the mean potential temperature $\overline{\theta}$. There is a Gaussian mountain located in the bottom left of the model domain, with a height of 2000 m and a standard deviation of 1000 km. Note that all quantities in Eqs. (7) and (8) have been nondimensionalized by dividing by prescribed typical scales of, for example, time and length.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

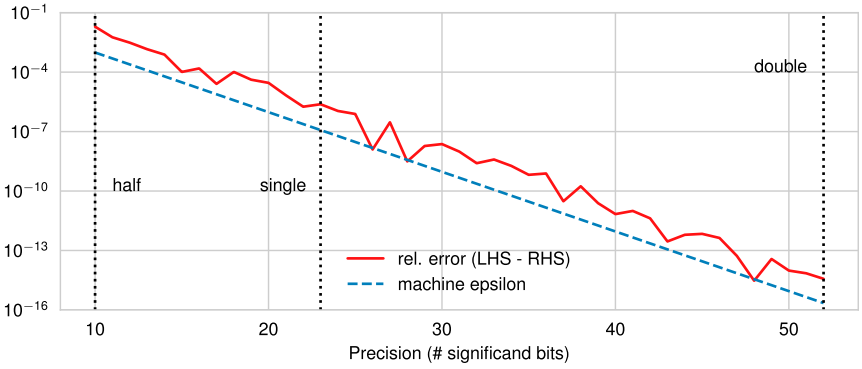

FIG. 4. As in Fig. 3, but the tangent-linear and adjoint models have been replaced with matrices of random numbers.

The spatial domain of one layer of the QG model is 120 by 20 grid points and the top and bottom layers are at depths of 5500 and 4500 m, respectively. A snapshot of the top-layer quasigeostrophic vorticity is illustrated in Fig. 2. The boundary conditions are cyclical in the zonal direction whereas the streamfunction values one grid point to the north and south of the grid are prescribed. The domain of the model is therefore equivalent to a zonal band on the sphere between approximately 15° and 72°N. The potential vorticity is advected by a semi-Lagrangian technique with bicubic interpolation to the departure point.

To investigate the impact of a precision reduction on the data assimilation procedure we use the Reduced Precision Emulator (rpe) (Dawson and Düben 2016, 2017) to emulate reduced-precision floating-point arithmetic. As a result, we were not able to measure any computational speed-up as we did not use hardware that natively supports such arithmetic. Here we only focus on the effects of low-precision arithmetic on the convergence of the 4D-Var minimisation.

Although the goal of this study is to investigate whether single-precision 4D-Var is feasible, we expect any effect from reducing precision to be evident and more easily identifiable if we consider levels of precision other than just double and single. Here we investigate the impact of reducing the significand width, which determines the precision of a calculation in terms of the number of significant digits of the output. Floating-point numbers also consist of an exponent bit field, which determines the exponent of the number when expanded in scientific notation in base 2. Reducing the exponent width reduces the range of representable numbers but does not affect the relative error of floating-point calculations taking place within this range. If a number goes outside of this range then an overflow or underflow floating-point exception will occur. For single-precision numbers, the maximum value of around 10^{38} usually provides enough range for most applications. However, for half-precision numbers, which only have 5 exponent bits, the maximum representable number is only 65 504 and therefore the probability of overflows is high. To avoid these effects and concentrate on the significand, we do not consider a reduction of the exponent below the double-precision width of 11 bits here. Nevertheless, when implementing such low-precision arithmetic in practice it will likely be necessary to rescale model variables to ''fit'' the calculations within the limited range available (Hatfield et al. 2019; Higham et al. 2019).

4. Tangent-linear/adjoint model asymmetry

Before beginning the data assimilation experiments, we first investigated how a precision reduction would affect the standard tangent-linear/adjoint model test. By definition, the relation

$$[\mathbf{M}_{im}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{x}]^{\mathrm{T}}\boldsymbol{\delta}\mathbf{y} = \boldsymbol{\delta}\mathbf{x}^{\mathrm{T}}[\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{y}], \tag{10}$$

where $\boldsymbol{\delta}\mathbf{x}$ and $\boldsymbol{\delta}\mathbf{y}$ are arbitrary vectors, relates the tangent-linear and adjoint models. The operators $\mathbf{M}_{im}(\mathbf{x}_0)$ and $\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)$ are in reality subroutines using floating-point arithmetic and are subject to different rounding errors when they operate on the input vector. The identity in Eq. (10) will therefore not hold exactly but will instead be violated slightly. At ECMWF, as a rule of thumb, the identity should be satisfied to at least the 10th decimal place to ensure convergence of 4D-Var (F. Váňa, June 2017, personal communication). Certainly, when testing the validity of the code of the adjoint model with respect to the tangent-linear model the identity should be satisfied to as many digits as possible, and the use of double-precision would ensure around 15 decimal places. In the following experiments

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

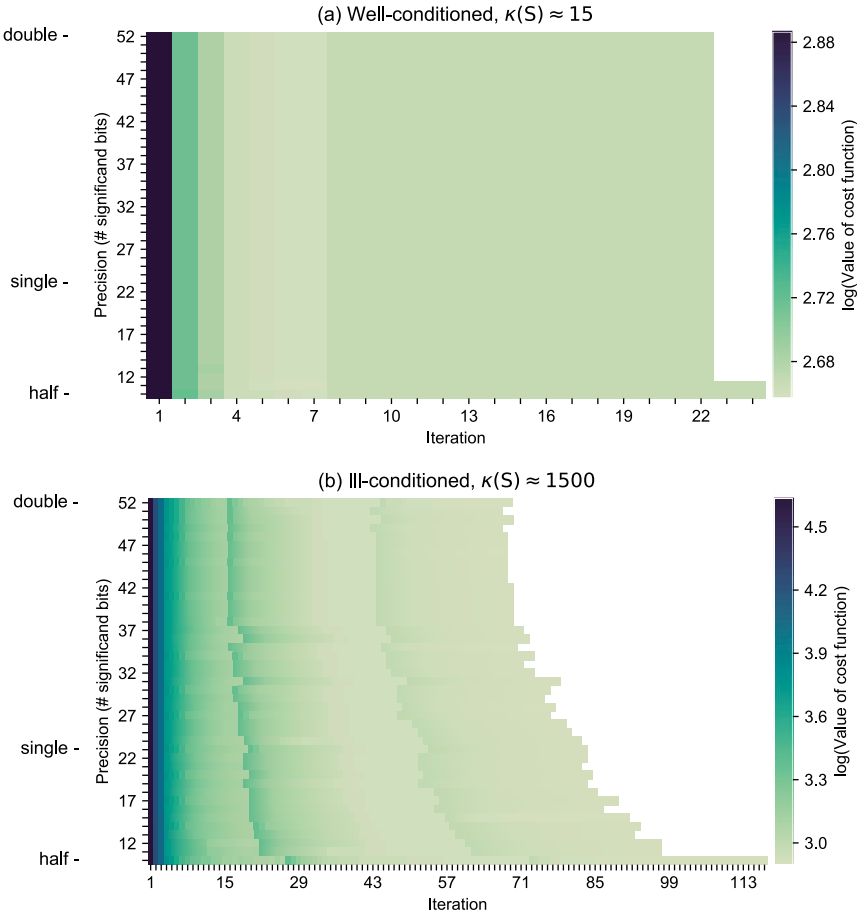

FIG. 5. An illustration of the behavior of convergence of the preconditioned conjugate gradient minimizer as precision is reduced in the tangent-linear and adjoint models, for (a) well-conditioned and (b) ill-conditioned problems. The number of iterations required to meet the convergence criteria is shown on the x axis. The color indicates the value of the cost function evaluated at the current iterate as the minimization proceeds from left to right. The condition number of the problem is given by $\kappa(\mathbf{S})$.

we will investigate how a violation of this identity beyond the seventh decimal place (i.e., single-precision and below) could prevent or slow down the convergence process of 4D-Var.

Figure 3 shows the relative error between the left-hand side and right-hand side of Eq. (10) for the QG model as a function of numerical precision, along with the machine epsilon, $\varepsilon = 2^{-p}$, where p is the number of significand bits. We define the relative error as

$$\frac{|\mathrm{LHS} - \mathrm{RHS}|}{|\mathrm{RHS}|} = \frac{|[\mathbf{M}_{im}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{x}]^{\mathrm{T}}\boldsymbol{\delta}\mathbf{y} - \boldsymbol{\delta}\mathbf{x}^{\mathrm{T}}[\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{y}]|}{|\boldsymbol{\delta}\mathbf{x}^{\mathrm{T}}[\mathbf{M}_{im}^{\mathrm{T}}(\mathbf{x}_0)\boldsymbol{\delta}\mathbf{y}]|}. \tag{11}$$

The relative error in a single floating-point operation is proportional to ε (Higham 2002) (i.e., when one significand bit is removed the floating-point truncation error doubles). However, Fig. 3 shows that this scaling is also valid for the tangent-linear and adjoint models overall. This is perhaps not surprising considering that they are, by definition, linear in the input. Essentially, even though the tangent-linear and adjoint models are formulated as computer subroutines, they behave like a matrix-vector multiplication. To see this, consider Fig. 4, which shows the same result as in Fig. 3 only the tangent-linear model has been replaced with a square matrix whose elements are random numbers and whose dimension is the same as the state vectors $\boldsymbol{\delta}\mathbf{x}$ and $\boldsymbol{\delta}\mathbf{y}$ (and the adjoint was replaced with the transpose of that matrix). The scaling of the relative error is identical to that in Fig. 3. It is reassuring that although the impacts of reducing precision may be deleterious, they are at least relatively predictable.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

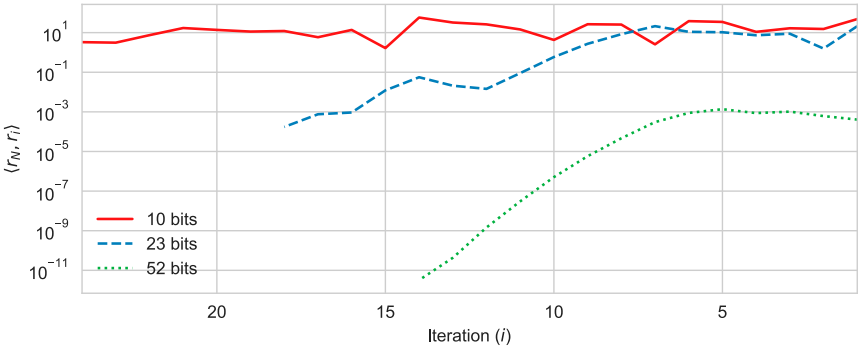

FIG. 6. The conjugate product between the final (Nth) residual and the ith residual for three levels of precision. The final residual is, for example, $N = 15$ for double-precision, and the increase in the conjugate product is visible as i goes from $i = 14$ to $i = 1$. This indicates a loss of conjugacy. The first inner loop from an assimilation experiment is shown.

5. Effects of reducing precision on convergence

We now investigate the effect of reducing precision in the tangent-linear and adjoint models on the convergence of an incremental 4D-Var minimization procedure. We do *not* reduce precision in any other components of the 4D-Var system, neither the conjugate gradient algorithm, the nonlinear model integration at the beginning of each inner loop nor any computations involving the background error covariance matrix. As mentioned earlier, the evaluations of the tangent-linear and adjoint models typically dominate the computational cost.

We consider both well-conditioned and poorly conditioned 4D-Var problems. The condition number of the Hessian, $\mathbf{S}$, is indicative of the number of iterations required to reach convergence—the higher the condition number, the longer a minimization algorithm typically takes to converge. We focus on the so-called spectral condition number, defined by

$$\kappa(\mathbf{S}) = \frac{\lambda_{\max}(\mathbf{S})}{\lambda_{\min}(\mathbf{S})}, \tag{12}$$

where $\lambda_{\max}(\mathbf{S})$ and $\lambda_{\min}(\mathbf{S})$ are the maximum and minimum eigenvalues of the Hessian, respectively (Tabeart et al. 2018). In fact, the relevant condition number is actually that of the Hessian of the preconditioned system, which by construction is simply equal to the largest eigenvalue of this Hessian, as preconditioning by $\mathbf{P}^b$ renders the smallest eigenvalue equal to unity (Haben 2011). The condition number of the Hessian contains information about the shape of the corresponding quadratic cost function – a condition number of unity indicates a perfectly symmetric quadratic "bowl" whereas larger condition numbers indicate increasing ellipticity.

Figure 5 illustrates the behavior of the convergence of incremental 4D-Var as precision is reduced in the tangent-linear and adjoint models, for well-conditioned and ill-conditioned problems, with condition numbers of around 15 and 1500, respectively. The color axis indicates the logarithm of the value of the cost function evaluated at the current iterate. We estimated the condition number of each problem through the Lanczos algorithm (Fisher 1998), which only requires matrix actions. The conditioning was changed by changing the observation error standard deviation: for the well-conditioned problem we used the baseline values mentioned in section 3a, whereas for the ill-conditioned problem we used values 10 times smaller than this. The ratio of the two condition numbers, 1500/15, is consistent with theory, which states that the condition number of the Hessian increases with the inverse of the square of the observation error standard deviation (Haben et al. 2011). Essentially, more accurate observations place a tighter restriction on the minimization of the cost function so that more iterations are required to reach the minimum within a given tolerance (Haben 2011). Note that the temporary increases in the value of the cost function as the iteration proceeds, most clearly visible for Fig. 5b as dark vertical lines, demarcate the beginning and end of subsequent inner loops.

Figure 5 illustrates how, for the well-conditioned problem, precision can be reduced significantly without impacting the convergence process. Going from double-precision (52 significant bits) to single-precision (23 significand bits) and beyond results in essentially the same solution. On the other hand, for the ill-conditioned problem, although the reduction of the cost function is more or less equivalent for the first 15 iterations, regardless of precision, more and more iterations are required to reach convergence as significand bits are removed. Specifically, using single-precision means that around 20% more iterations must be performed in order to reach the same solution as when using double-precision. The trend continues as precision is reduced further, with half-precision tangent-linear and

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

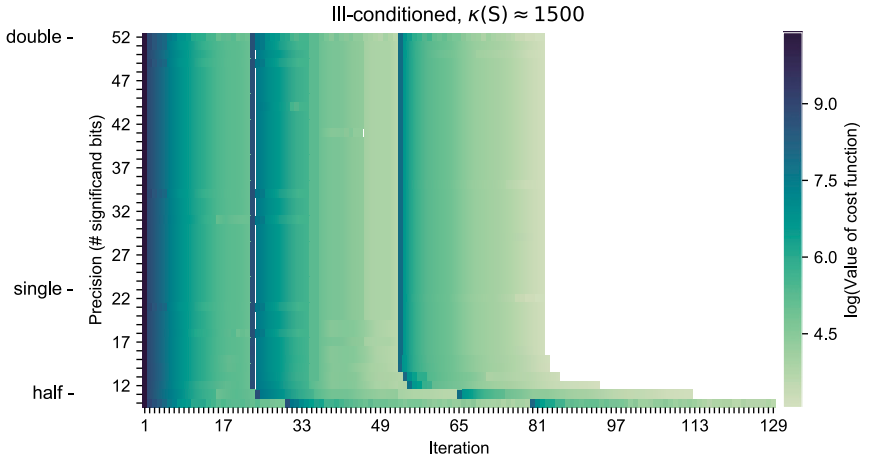

FIG. 7. As in Fig. 5b, but that the generalized minimal residual (GMRES) minimization technique is used.

adjoint models requiring around 70% more iterations than double-precision. Of course, if using single-precision results in a longer convergence process then this may negate any computational speed-up from using reduced-precision arithmetic. Nevertheless, we show in the next section how we might still be able to use such a low precision in 4D-Var successfully.

6. Reorthogonalizing the residuals

When the CG algorithm is used for minimization, it is common to manually reorthogonalize the residuals (which are used to build the search directions) with the Gram–Schmidt process. We omitted this in the previous discussion, but we believe this additional step makes the use of single-precision tangent-linear and adjoint models viable. It is well known that rounding errors can cause problems for the CG algorithm. Consider the incremental 4D-Var problem in its most simple form, to solve

$$\mathbf{S}\mathbf{w} = \mathbf{b} \tag{13}$$

for $\mathbf{w}$. Here, $\mathbf{b} \in \mathbb{R}^n$ is a constant vector that depends on the innovations [we omit the full form for brevity; see Haben (2011)]. The CG algorithm for solving Eq. (13) is outlined in Algorithm 1 (Trefethen and Bau 1997). At each step of the iteration we compute the residual $\mathbf{r}_n$, which is also the negative of the gradient of the cost function evaluated at $\mathbf{w}_n$. The CG algorithm therefore bears a resemblance to the simple gradient descent method. However, by using knowledge of the Hessian, $\mathbf{S}$, we are able to dramatically reduce the number of iterations required for convergence. The CG algorithm is appropriate when $\mathbf{S}$ is symmetric positive definite:

$\mathbf{w}_0 = 0$, $\mathbf{r}_0 = \mathbf{b}$, $\mathbf{p}_0 = \mathbf{r}_0$
for $n = 1, 2, 3, \ldots$ do
  $\alpha_n = (\mathbf{r}_{n-1}^{\mathrm{T}}\mathbf{r}_{n-1})/\mathbf{p}_{n-1}^{\mathrm{T}}\mathbf{S}\mathbf{p}_{n-1}$ step length
  $\mathbf{w}_n = \mathbf{w}_{n-1} + \alpha_n\mathbf{p}_{n-1}$ approximate solution
  $\mathbf{r}_n = \mathbf{r}_{n-1} - \alpha_n\mathbf{S}\mathbf{p}_{n-1}$ residual
  $\beta_n = (\mathbf{r}_n^{\mathrm{T}}\mathbf{r}_n)/(\mathbf{r}_{n-1}^{\mathrm{T}}\mathbf{r}_{n-1})$ improvement this step
  $\mathbf{p}_n = \mathbf{r}_n + \beta_n\mathbf{p}_{n-1}$ search direction
end

Algorithm 1: The conjugate gradient (CG) algorithm for $\mathbf{S}\mathbf{w} = \mathbf{b}$.

Mathematically, the residuals produced by the CG algorithm at successive iterations are all orthogonal to each other. However, the use of finite precision floating-point arithmetic causes a loss of orthogonality among these residual vectors, more so for low-precision floats than for high-precision floats. This can cause search directions to be explored multiple times (Trefethen and Bau 1997, Lecture 36), and we hypothesize that this loss of orthogonality is responsible for the retarded convergence process that we observe when reducing precision. In our case the situation is complicated by the presence of the preconditioner matrix, but the problem is very similar. Essentially, the residuals lose conjugacy with respect to the inverse of the preconditioner matrix. This is illustrated in Fig. 6, which shows the conjugate product (the inner product with respect to the inverse of the preconditioner matrix) between the final (Nth) residual and the preceding (ith) residuals for the first inner loop of ill-conditioned assimilation experiments using 10, 23 and 52 bit tangent-linear and adjoint models. Looking at double-precision (52 significand bits), for example, where $N = 15$, the conjugate product of the Nth residual with the ith residual increases as i tends toward 1 (the first residual). This is even more so the case for single-precision (23 significand bits) and half-precision (10 significand bits). These conjugate products should ideally be as close to zero as possible.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

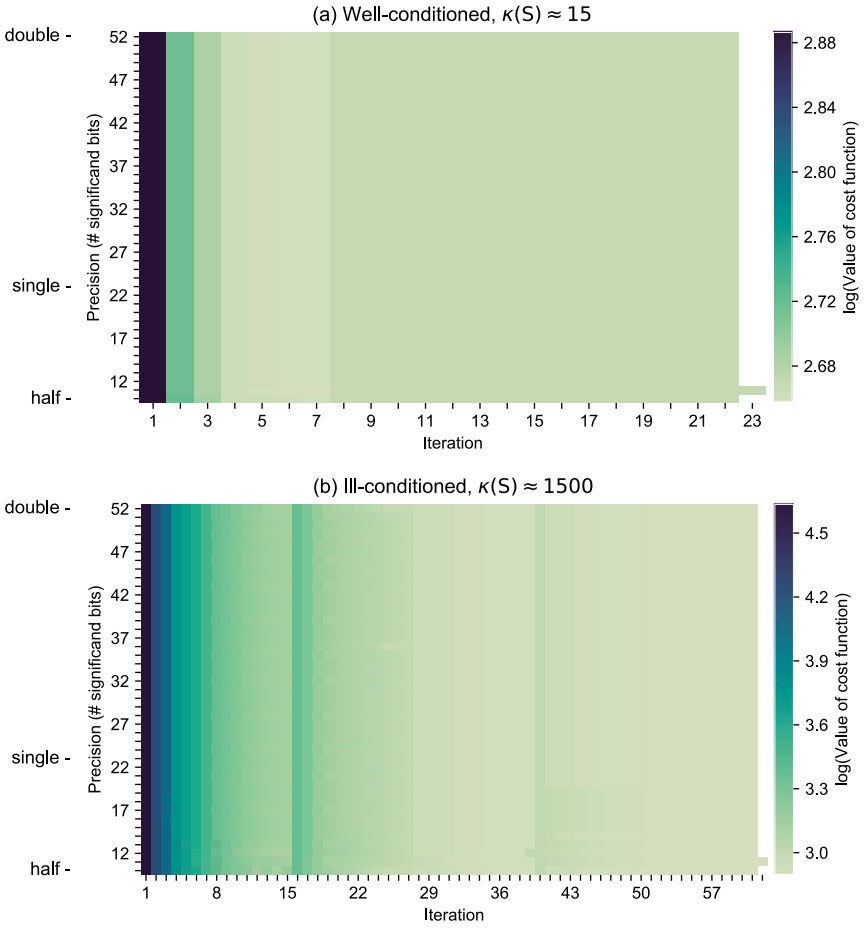

FIG. 8. As in Fig. 5, but that reorthogonalization is used in the preconditioned conjugate gradient minimizer.

To further support our hypothesis, we repeated the experiment of Fig. 5b but with the generalized minimal residual (GMRES) algorithm. Unlike CG, GMRES is designed for general (nonsymmetric) matrices, and the intermediate residuals are not orthogonal to each other. This algorithm therefore contains an explicit procedure that constructs an orthogonal basis, rather than using the residuals directly. GMRES does not suffer as much as the CG algorithm in the presence of rounding errors. Accordingly, as shown in Fig. 7, we were able to use as few as 15 significand bits without affecting the rate of convergence of the 4D-Var minimization when using the GMRES algorithm.[1]

In the case of CG, at each iteration we can therefore manually reorthogonalize the current residual against all of the previous residuals using the Gram–Schmidt process. For high-dimensional systems, using reorthogonalization is essential even with double-precision. It is used operationally at ECMWF (Fisher 1998) and also within the PCG minimizer of OOPS. By definition, the orthogonalization step adds some extra cost. In the context of solving linear systems, such orthogonalization (e.g., within GMRES) is considered a burden, and the increasing cost per iteration motivates the use of restarted methods. However, in our 4D-Var application, the matrix actions in each iteration require the execution of expensive tangent-linear and adjoint models over a number of timesteps, and reorthogonalization therefore has negligible cost. We were not able to detect a difference in the wall-clock time of our assimilation experiment when switching reorthogonalisation on and off, for example.

We repeated the experiment shown in Fig. 5 but with reorthogonalization and the results are shown in Fig. 8.

[1] Note that GMRES produces different results to CG at each iteration. They both explore the Krylov space $\{\mathbf{b}, \mathbf{Sb}, \mathbf{S}^2\mathbf{b}, \ldots, \mathbf{S}^{n-1}\mathbf{b}\}$, but GMRES minimizes $\|\mathbf{S}\mathbf{w}_n - \mathbf{b}\|_2 \equiv \|\mathbf{w}_n - \mathbf{w}^*\|_{\mathbf{S}^T\mathbf{S}}$, where $\mathbf{w}^*$ is the solution, while CG minimizes $\|\mathbf{w}_n - \mathbf{w}^*\|_{\mathbf{S}}$. The latter is closely related to minimizing the quadratic cost function.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

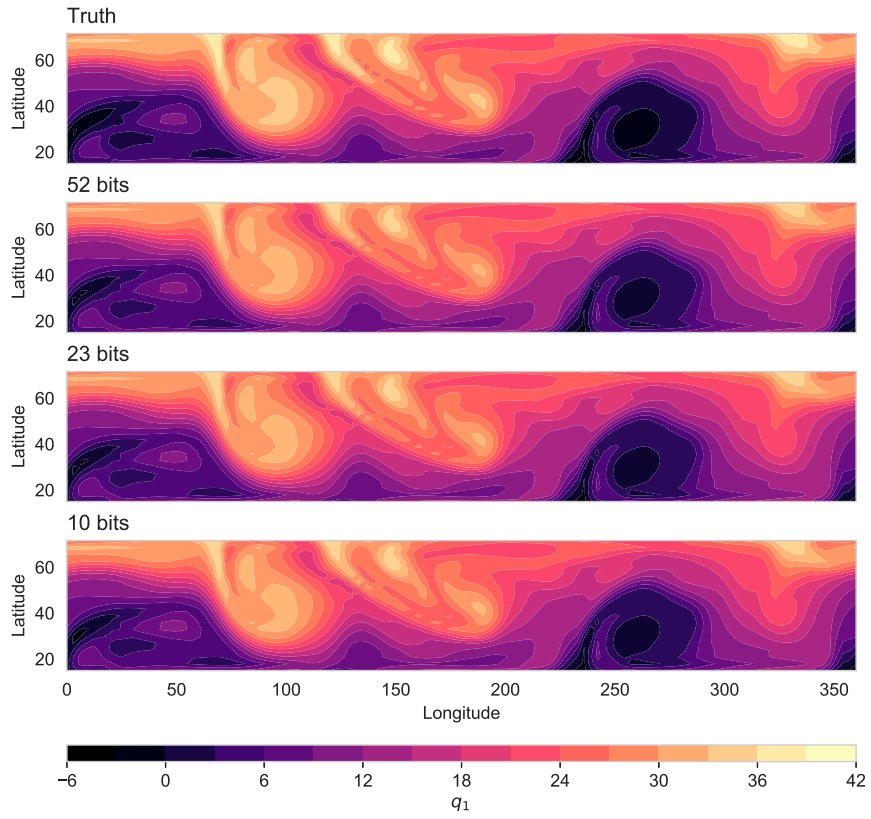

FIG. 9. Snapshots of the top-layer quasigeostrophic potential vorticity, q_1, for the truth and the analyses computed with tangent-linear and adjoint models at different levels of precision. The state at the end of the 24 h assimilation window is shown.

Clearly, reorthogonalization allows one to use a far lower precision in the tangent-linear and adjoint models than otherwise. Even with only 11 significand bits, which is almost half-precision, the convergence behavior of PCG remains the same. Figure 8 indicates that the asymmetry introduced between the tangent-linear and adjoint models (Fig. 3) when reducing precision does not preclude a successful minimization of the 4D-Var cost function. To support this statement, we also illustrate the analysis explicitly in Fig. 9. We show the top-layer quasigeostrophic potential vorticity q_1 at the end of the 12 h window for the truth and three levels of precision in the tangent-linear and adjoint models: double, single, and half-precision (with 10, 23, and 52 significand bits, respectively). Clearly all three levels of precision deliver extremely similar analyses and appear to be converging on the same minimum of the cost function.

7. Conclusions

The tangent-linear and adjoint models dominate the computational cost of 4D-Var. For this reason, they are typically run at a lower resolution than the full nonlinear model. Reducing precision in these models could provide another means to reduce their computational overhead. However, reducing precision in these models introduces a stronger asymmetry between them, compared with double-precision, which we can expect to disrupt the convergence process. Using a medium-complexity fluid simulation we have demonstrated that this is indeed the case for ill-conditioned problems, for which lowering precision results in a loss of orthogonality in the conjugate gradient residual vectors. This retards the convergence of the 4D-Var cost function minimization. However, by reorthogonalizing the conjugate gradient residuals, a well-established technique, which is already used operationally, we were able to minimize the 4D-Var cost function using linear models at a precision even lower than single-precision for the same number of iterations as with double-precision.

Our results are promising and indicate that the use of single-precision (or even lower precision) linear models in 4D-Var for operational numerical weather prediction may be viable. However, the 4D-Var problem as encountered in operational systems is somewhat more poorly conditioned than the one presented here. The

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

condition number of the Hessian in the 4D-Var system of ECMWF is typically around 15 000 (M. Chrust, July 2019, personal communication), compared with only around 1500 for our ''ill-conditioned'' problem. It remains to be seen to what extent reorthogonalization can cope with much higher condition numbers. Our simplified setup did not contain sufficiently tuneable parameters for investigating problems of such a high condition number. For this, we expect a similar study to ours but applied to a high-dimensional numerical weather prediction model to be valuable. We also did not include any model error in our data assimilation experiments. A previous study found that the presence of model error can allow precision to be reduced to a much lower level than otherwise (Hatfield et al. 2018a) and therefore we must include model error to make a truly fair test of reduced-precision models in 4D-Var.

A natural next test would be to consider a full-complexity model such as the IFS of ECMWF. Reducing precision throughout the tangent-linear and adjoint models would not be straightforward owing to the complexity of the model code. However, as a simple test one could choose a computationally expensive but algorithmically simple ''kernel'' in the tangent-linear and adjoint models and only reduce precision in that kernel. For example, one could use a single-precision matrix-matrix multiply in the Legendre transform, as in the half-precision experiments of Hatfield et al. (2019). This would increase the asymmetry between the tangent-linear and adjoint models and would therefore allow the reorthogonalization technique to be tested without making significant code changes to the IFS.

Acknowledgments. The authors thank Massimo Bonavita, Marcin Chrust, Olivier Marsden, and Elias Holm at ECMWF for their useful advice. Sam Hatfield is funded by the NERC under Grant NE/L002612/1. Peter Düben gratefully acknowledges funding from the Royal Society for his University Research Fellowship as well as funding from the ESiWACE and ESiWACE2 projects. ESiWACE and ESiWACE2 have received funding from the European Union's Horizon 2020 research and innovation programme under Grants 675191 and 823988, respectively. Andrew McRae and Tim Palmer received funding from the European Research Council (Grant 741112) under the European Union's Horizon 2020 research and innovation programme.

REFERENCES

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

Dawson, A., and P. D. Düben, 2016: aopp-pred/rpe: v5.0.0. zenodo, accessed 5 March 2020, https://doi.org/10.5281/zenodo.154483.

——, and ——, 2017: rpe v5: An emulator for reduced floating-point precision in large numerical simulations. *Geosci. Model Dev.*, **10**, 2221–2230, https://doi.org/10.5194/gmd-10-2221-2017.

Düben, P. D., and T. N. Palmer, 2014: Benchmark tests for numerical weather forecasts on inexact hardware. *Mon. Wea. Rev.*, **142**, 3809–3829, https://doi.org/10.1175/MWR-D-14-00110.1.

Fandry, C. B., and L. M. Leslie, 1984: A two-layer quasi-geostrophic model of summer trough formation in the Australian subtropical easterlies. *J. Atmos. Sci.*, **41**, 807–818, https://doi.org/10.1175/1520-0469(1984)041<0807:ATLQGM>2.0.CO;2.

Fisher, M., 1998: Minimization algorithms for variational data assimilation. *Proc. Seminar on Recent Developments in Numerical Methods for Atmospheric Modelling*, Reading, United Kingdom, ECMWF, 364–385.

Golub, G. H., and C. F. Van Loan, 1986: *Matrix Computations*. North Oxford Academic, 476 pp.

Haben, S. A., 2011: Conditioning and preconditioning of the minimisation problem in variational data assimilation. Ph.D. thesis, University of Reading, 196 pp., https://www.reading.ac.uk/web/files/maths/HabenThesis.pdf.

——, A. S. Lawless, and N. K. Nichols, 2011: Conditioning of incremental variational data assimilation, with application to the Met Office system. *Tellus*, **63A**, 782–792, https://doi.org/10.1111/j.1600-0870.2011.00527.x.

Hatfield, S., P. Düben, M. Chantry, K. Kondo, T. Miyoshi, and T. Palmer, 2018a: Choosing the optimal numerical precision for data assimilation in the presence of model error. *J. Adv. Model. Earth Syst.*, **10**, 2177–2191, https://doi.org/10.1029/2018MS001341.

——, A. Subramanian, T. Palmer, and P. Düben, 2018b: Improving weather forecast skill through reduced-precision data assimilation. *Mon. Wea. Rev.*, **146**, 49–62, https://doi.org/10.1175/MWR-D-17-0132.1.

——, M. Chantry, P. Düben, and T. Palmer, 2019: Accelerating high-resolution weather models with deep-learning hardware. *Proc. Platform for Advanced Scientific Computing Conf.—PASC '19*, Zurich, Switzerland, ACM Press, 11, https://doi.org/10.1145/3324989.3325711.

Higham, N. J., 2002: *Accuracy and Stability of Numerical Algorithms*. 2nd ed. Society for Industrial and Applied Mathematics, 710 pp.

——, S. Pranesh, and M. Zounon, 2019: Squeezing a matrix into half precision, with an application to solving linear systems. *SIAM J. Sci. Comput.*, **41**, A2536–A2551, https://doi.org/10.1137/18M1229511.

Lawless, A. S., 2013: Variational data assimilation for very large environmental problems. *Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*, M. Cullen et al., Eds., De Gruyter, 55–90.

——, S. Gratton, and N. K. Nichols, 2005: An investigation of incremental 4D-Var using non-tangent linear models. *Quart. J. Roy. Meteor. Soc.*, **131**, 459–476, https://doi.org/10.1256/qj.04.20.

Lawrence, B. N., and Coauthors, 2018: Crossing the chasm: How to develop weather and climate models for next generation computers? *Geosci. Model Dev.*, **11**, 1799–1821, https://doi.org/10.5194/gmd-11-1799-2018.

Meurant, G., and Z. Strakoš, 2006: The Lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numer.*, **15**, 471–542, https://doi.org/10.1017/S096249290626001X.

Nakano, M., H. Yashiro, C. Kodama, and H. Tomita, 2018: Single precision in the dynamical core of a nonhydrostatic global

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC

atmospheric model: Evaluation using a baroclinic wave test case. *Mon. Wea. Rev.*, **146**, 409–416, https://doi.org/10.1175/MWR-D-17-0257.1.

Pedlosky, J., 1979: *Geophysical Fluid Dynamics*. 1st ed. Springer-Verlag, 624 pp.

Rüdisühli, S., A. Walser, and O. Fuhrer, 2014: COSMO in single precision. *COSMO User Workshop*, Zurich, Switzerland, Federal Office of Meteorology and Climatology, MeteoSwiss, 28 pp., https://wiki.c2sm.ethz.ch/pub/COSMO/EventsCUW2013/Ruedisuehli.pdf.

Tabeart, J. M., S. L. Dance, S. A. Haben, A. S. Lawless, N. K. Nichols, and J. A. Waller, 2018: The conditioning of least-squares problems in variational data assimilation. *Numer. Linear Algebra Appl.*, **25**, e2165, https://doi.org/10.1002/nla.2165.

Trefethen, L. N., and D. Bau III, 1997: *Numerical Linear Algebra*. SIAM, 361 pp.

Váňa, F., P. Düben, S. Lang, T. Palmer, M. Leutbecher, D. Salmond, and G. Carver, 2017: Single precision in weather forecasting models: An evaluation with the IFS. *Mon. Wea. Rev.*, **145**, 495–502, https://doi.org/10.1175/MWR-D-16-0228.1.

Unauthenticated | Downloaded 01/21/22 06:31 PM UTC