SPECIAL PROJECT PROGRESS REPORT

All the following mandatory information needs to be provided. The length should *reflect the complexity and duration* of the project.

Reporting year	2019
Project Title:	The use of imprecise arithmetic to increase resolution in atmospheric models
Computer Project Account:	spgbtpia
Principal Investigator(s):	Palmer, T N
Affiliation:	University of Oxford
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Duben, P
Start date of the project:	2017
Expected end date:	2020

Computer resources allocated/used for the current year and the previous one

(if applicable)

Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	15,000,000	14,713,259	15,000,000	221,267
Data storage capacity	(Gbytes)	15,000	Low	15,000	Low

Summary of project objectives (10 lines max)

Investigate the possible benefits of reduced or variable numerical precision on weather and climate prediction. This builds on the work carried out in Oxford investigating reduced precision in simple models. Double precision is used as the default precision level in most weather and climate codes, yet these codes contain large sources of uncertainty and error. We are investigating if this high precision is necessary for an accurate forecast. We will examine a variety of kernels of the weather forecasting process to assess the viability of reduced-precision for operational weather forecasting.

Summary of problems encountered (10 lines max)

No notable problems encountered.

Summary of plans for the continuation of the project (10 lines max)

The remaining budget will be used to expand the work examining the importance of precision to physical parameterisation schemes. To date this component of the project has focused on a simplified GCM and very low resolution simulations with OpenIFS (T21 and T159). Once this preliminary investigation is complete, we will use the remaining budget to run higher resolution ensemble forecasts with varying precision across the parameterisation schemes to establish the impact on forecast quality.

List of publications/reports from the project with complete references

Chantry, M., Thornes, T., Palmer, T. and Düben, P., 2019. Scale-selective precision for weather and climate forecasting. Monthly Weather Review, 147(2), pp.645-655.

Hatfield, S., Chantry, M., Düben, P. and Palmer, T., 2019, June. Accelerating high-resolution weather models with deep-learning hardware. In Proceedings of the Platform for Advanced Scientific Computing Conference.

Summary of results

The main result from this year’s portion of the project is captured in the paper: Accelerating high-resolution weather models with deep-learning hardware, which was awarded the best paper prize at the Platform for Advanced Scientific Computing (PASC) conference. This paper investigates the use of reduced numerical precision for the Legendre transforms within IFS. The work carried out in year two of this special project, examining the lengthscale-dependence of necessary precision in the spectral space calculations of OpenIFS, was also published this year in Monthly Weather Review. Below we will describe the main results from the paper examining the Legendre transforms.

The Legendre transforms are already one of the most expensive kernels of the IFS and have expensive scaling which will impact future resolution increases, in particular the push towards 1km global forecasts. Previous work in this special project had demonstrated that a neighbouring kernel, the spectral space calculations, could be carried out with the vast majority of variables stored and calculated at half-precision or lower. This result suggests that reduced precision would be feasible for the Legendre transform calculations. This is of particular interest as this kernel is dominated by matrix-matrix multiplications. Driven by machine learning and other use-cases, specialist hardware has been developed to efficiently carry out reduced-precision matrix-matrix multiplications, e.g. the Volta hardware from NVidia and the tensor processing unit from Google.

Our paper tests the forecast quality of the IFS when reduced numerical precision, especially half-precision, is used for the Legendre transforms. We establish that almost all of the component matrix-matrix multiplications can be calculated at half-precision. Only the calculations which establish the global mean or very large lengthscale elements need higher precision to produce forecasts which perform as well as a double-precision equivalent when measured against the analysed state of the atmosphere. We demonstrate that these results hold even at T1279 resolution, the current deterministic forecast resolution, and for ensemble forecasts up to TCo399.
The results described could have a major impact on the future computing setup used for running the operational IFS forecast. Here we have established that the precision component of the available hardware has value for the Legendre transforms. We have achieved this through the use of a reduced-precision emulator which does not give the performance increase. Further work is required to write the code necessary to use this specialised hardware and understand the costs of moving the data between the CPU and specialised hardware.