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	2021
Project Title:	Imprecise approaches to accelerate weather forecasts
Computer Project Account:	spgbtpia
Principal Investigator(s):	Tim Palmer
Affiliation:	University of Oxford
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Peter Duben, Sam Hatfield, Inna Polichtchouk
Start date of the project:	2020
Expected end date:	2022

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	5,946,010	15,000,000	12,000,000
Data storage capacity	(Gbytes)	15,000	10,000	15,000	17,000

Summary of project objectives (10 lines max)

Investigate the applications of reduced numerical precision and reduced numerical precision hardware to improving the performance of weather and climate modelling. Reduced numerical precision is being increasingly supported by new hardware, in particular GPUs and similar machine learning devices. This hardware offers significantly increased performance if the numerical precision errors can be kept below forecast errors and uncertainty. Machine learning itself offers another approach to accelerate weather forecasting, whereby algorithms can be emulated at reduced cost using a machine learning algorithms (e.g. neural networks). We explore the tolerance of kernels of weather and climate forecasting to reduce precision calculations or to emulation by neural networks.

Summary of problems encountered (10 lines max)

No notable problems were encountered in this year of the project.

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

On reduced numerical precision, we have started an assessment on the performance of radiation schemes, thus far in offline testing. We will continue this research by coupling the reduced precision radiation schemes to IFS and testing the online forecast skill. We also plan to explore the use of reduced numerical precision in data communication, specifically the MPI communications used to parallelise weather and climate forecasting systems across many computer nodes. On machine learning emulation, we will build upon the successes in gravity wave drag emulation and explore the emulation of other physical processes, particularly those that are key to climate forecasting.

List of publications/reports from the project with complete references

Chantry, Hatfield, Duben, Polichtchouk & Palmer, Machine learning emulation of gravity wave drag in numerical weather forecasting, submitted to JAMES, https://arxiv.org/abs/2101.08195

Hatfield, Chantry, Duben, Lopez, Geer & Palmer, Neural networks as the building blocks for tangent-linear and adjoint models in variational data assimilation, submitted to JAMES, https://www.essoar.org/doi/10.1002/essoar.10506310.1

Summary of results

Our use of special project this year has been focused on the emulation of physical parameterisation schemes to accelerate weather forecasting. The goal is to accelerate said schemes without degrading forecast quality. We will now briefly summarise the results of two submitted papers that cover the research in the project.

In the first paper, Chantry et al. 2021 (currently under review), we use neural networks to emulate the non-orographic gravity wave drag scheme within ECMWF's IFS model. To achieve this, we generate a dataset of the tendencies produced by the existing gravity wave drag scheme and train neural networks to emulate the tendencies given the current state of the atmosphere. Here we find that the ability to emulate the existing scheme offline (i.e. decoupled from the IFS) increases as the degrees of freedom within the neural network increases (see figure 1 in Chantry et al. 2021). Offline testing is key for training neural networks, but provides an unreliable guide for online performance, particularly for long integrations (Brenowitz et al. 2020). Therefore, a second phase of testing is required, where the emulator is coupled to the forecasting system.
To thoroughly test our neural network emulators while coupled to the IFS we undertook a range of forecasting problems, ranging from medium range to multi-year timescales, at a range of different spatial resolutions. These were run using the resources on this special project. Through this testing

phase we were able to find a stable configuration that produces accurate forecasts on all timescales. In figure 4 of Chantry et al. we show that our networks are neutral compared with the current gravity wave drag scheme when assessed against the analysis. In figure 5 we show that in a series of long range forecasts our neural network model is able to recreate the behaviour of the QBO. These results show the ability of our neural networks to accurate predict tendencies with no persistent bias or flaw.

Our purpose for creating neural network emulators was to accelerate uncertain areas of parameterised physics. Using special project resources, we assessed the computational cost of our networks when coupled to the IFS. Currently we find that our emulators take comparable solution time to the existing scheme but find that large accelerations are possible if the hardware infrastructure included GPU devices. Both the existing scheme and our neural network emulators perform much slower when coupled to IFS, suggesting that data movement is a dominant cost in the calculation. Moving parameterisation schemes to dedicated GPU devices might alleviate this issue if sufficiently many parameterisation schemes can be emulated. Further work is required to understand and optimise performance.

In our second paper, Hatfield et al 2021 (under review), we build on the above work to test our neural network emulators within data assimilation. At ECMWF (and many other forecasting centres) 4D-var is the data assimilation framework of choice. In this framework an optimal initial condition is found through a minimisation procedure. This procedure propagates increments to the atmospheric state forward in time through a tangent-linear equivalent of the IFS code, and then propagates gradients backwards through an adjoint of the IFS code. This requires the maintenance of two additional versions of the forecasting model, with additional care required to remove large gradients (Janisková & Lopez 2013). With a neural network emulator, building tangent linear and adjoint equivalents of the nonlinear scheme is very easy as the complexity lies not within the equations but the weights of the neural network. We test a neural network trained on the nonlinear versions of the non-orographic gravity wave drag scheme to produce accurate and stable tangent-linear and adjoint variants. We find that our tangent-linear and adjoint codes can be stably used within the full 4D-var data assimilation framework over many continuous cycles of assimilation. There is no notable degradation in the accuracy of forecasts that use our neural networks for the tangent and adjoint calculations.

Key to both of these papers will be building upon the results across a suite of parameterisation schemes. In particular the non-orographic gravity wave scheme is not the most expensive or impactful scheme. Therefore, testing these approaches in schemes such as radiation or cloud microphysics will provide great insight into the possibility of a predominately neural network parameterisation suite.

Chantry, Matthew, et al. "Machine learning emulation of gravity wave drag in numerical weather forecasting." *arXiv preprint arXiv:2101.08195* (2021).

Hatfield, Sam et al. "Building tangent-linear and adjoint models for data assimilation with neural networks" *ESSOAR preprint* https://www.essoar.org/doi/10.1002/essoar.10506310.1 (2021).

Brenowitz, Noah D., et al. "Machine Learning Climate Model Dynamics: Offline versus Online Performance." *arXiv preprint arXiv:2011.03081* (2020).

Janisková, Marta, and Philippe Lopez. "Linearized physics for data assimilation at ECMWF." *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*. Springer, Berlin, Heidelberg, 2013. 251-286.