The use of inexact hardware in data assimilation for improved weather and climate prediction

Sam Hatfield*, Peter Düben, Aneesh Subramanian, Tim Palmer
University of Oxford, *samuel.hatfield@physics.ox.ac.uk, samhatfield.co.uk

Introduction

Conventional wisdom holds that, in Earth-System models, it is always best to use the highest numerical precision possible. However, several studies conducted thus far have found a significant tolerance in models to a reduction in this precision, and therefore a potential free source of computational resources. The aim of this project is to extend these investigations into data assimilation.

Motivation: reducing precision to improve data assimilation

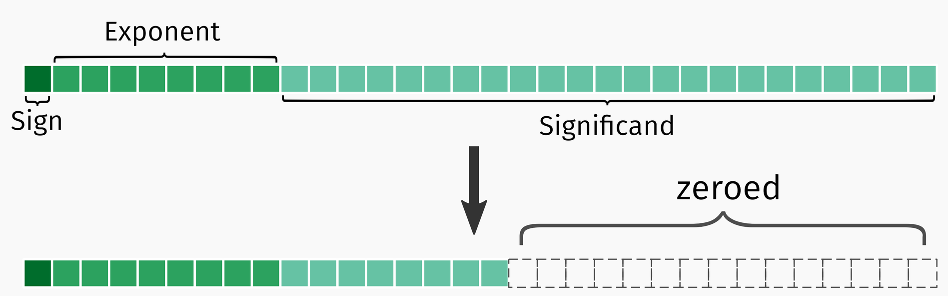

Figure 1 Procedure for emulating a reduced-precision floating-point number; in this case the trailing 15 bits of the significand are set to 0.

- The number of bits representing a variable in a model determines its precision
- Typically, 64 bits, or double precision, is used
- Given uncertainties in the model formulation, **double precision is often more precision than is necessary**
- If the number of bits can be reduced without having a significant impact on the model performance, then the computational resources thus saved can be reinvested in other more crucial areas, such as in the ensemble size
- This would lead to **an improvement in accuracy, for the same computational resources**
- For this study, we investigate how the quality of data assimilation is affected by a reduction in numerical precision
- We **emulate future variable-precision hardware** (Figure 1)
- By looking at the quality of analyses as precision is reduced, we can study trade-offs between ensemble size and numerical precision

Methods: model and assimilation setup

- We use the Lorenz '96 system: a simple, spatio-temporally chaotic toy model of weather (Figure 2)
- The variant used here has three "levels" analogous to different scales in the atmosphere
- The full, three-level version is integrated to give a truth run
- Observations of this truth are assimilated into a version with parametrized small scale dynamics
- To assimilate the observations, we use an **ensemble square root filter** (EnSRF)
- This is closely related to the ensemble Kalman filter, but the ensemble mean and ensemble spread are updated using different gain matrices
- Observations are processed one-at-a-time, to avoid inverting matrices

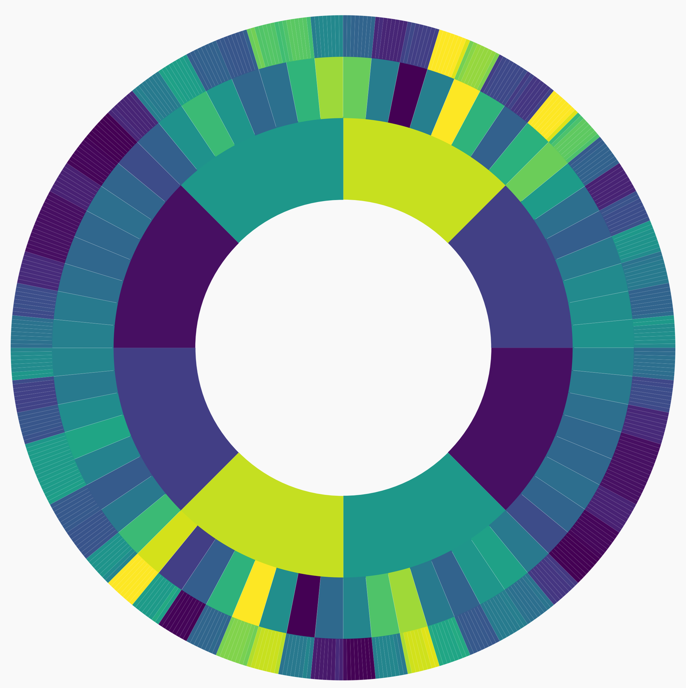

Figure 2 A schematic of the three-level variant of the Lorenz '96 model. The inner, middle and outer rings represent different scales of motion.

Results

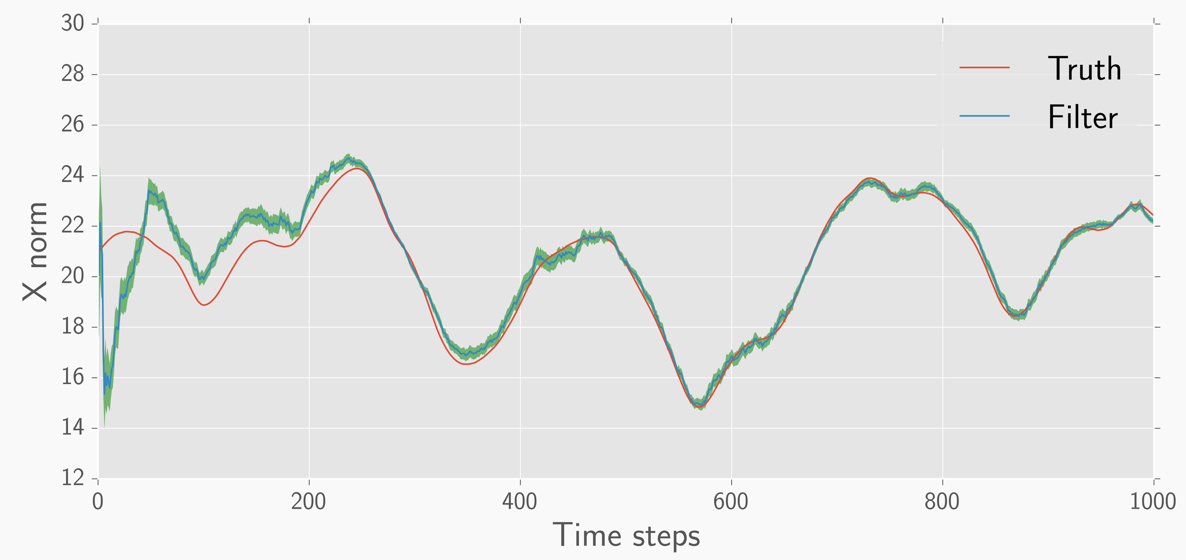

Figure 3 The trajectory of an EnSRF running at half precision.

- Figure 3 shows the trajectory of the EnSRF, assimilating observations from the truth run, running at half precision with 40 members
- This shows that **the filter can function at reduced precision without diverging**

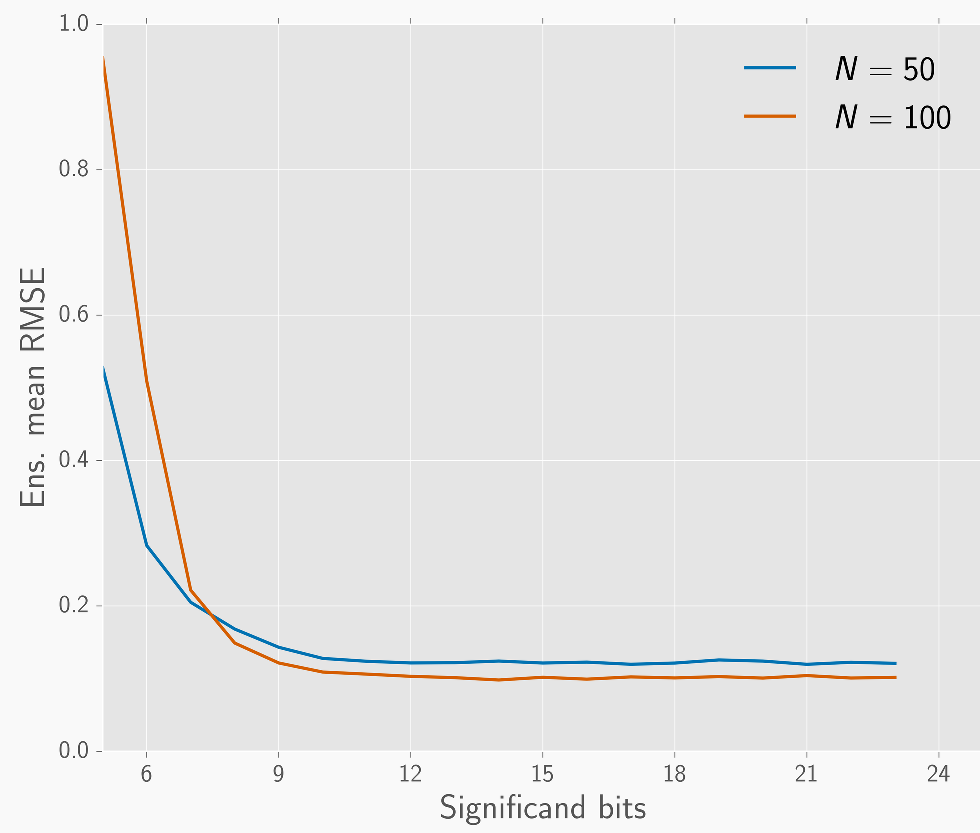

Figure 4 Analysis ensemble mean RMSE as a function of the number of significand bits.

- Figure 4 shows how the quality of the analysis is degraded as the number of significand bits used to code the filter is reduced (see Figure 1), for ensembles of 50 and 100
- The Lorenz '96/EnSRF is tolerant to a significant reduction in precision, with no discernible increase in error down to 10 significand bits
- This suggests that we can **successfully run the assimilation at reduced precision, with minimal increase in error**

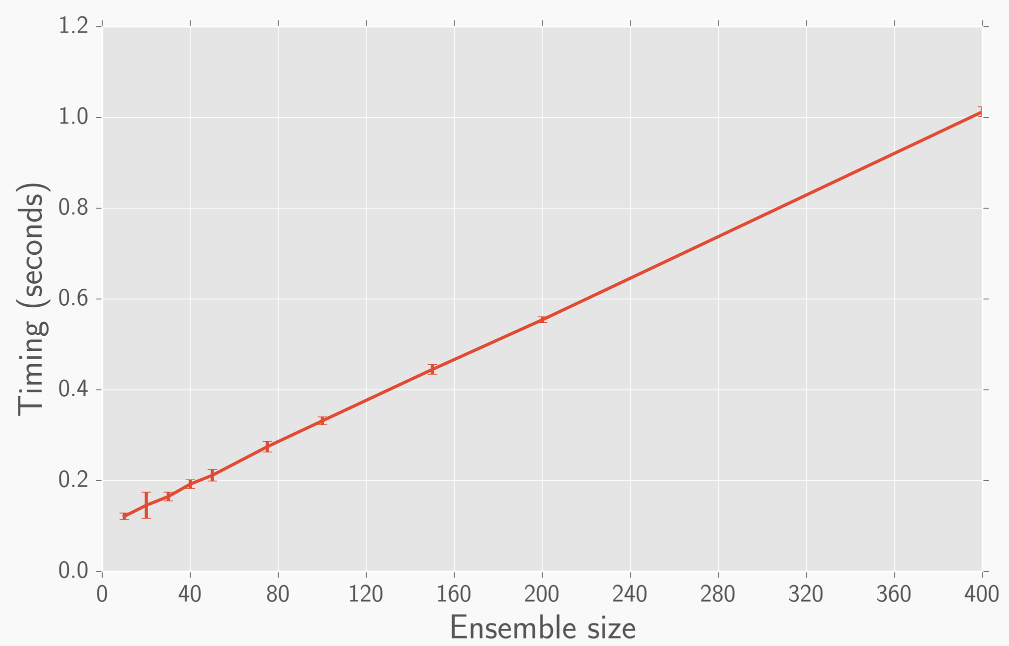

Figure 5 Wall-clock time as a function of ensemble size.

- Figure 5 illustrates the dependence of wall-clock time on ensemble size
- By assuming that wall-clock time is proportional to precision, we can estimate the relative computational cost of assimilation setups with different ensemble sizes and precision levels

	40	125	290
double	1.0	0.59	0.57
single	1.03	0.59	0.57
half	1.05	0.71	0.76

Table 1 Analysis ensemble mean RMSE as a function of ensemble size and precision, relative to "double 40".

- Table 1 shows the error of the assimilation as a function of the numerical precision and the ensemble size, relative to "double 40"
- Cells with the same colour represent scenarios that require the same computational resources
- Moving from "D40", double precision with 40 members, to S125 gives a **40% decrease in analysis error, while requiring no more computational resources**
- Also, by moving from D40 to H40, **compute-time can be quartered with only a 5% increase in error**
- Unfortunately, the error at half precision becomes quite significant as ensemble size is increased

Discussion and future work

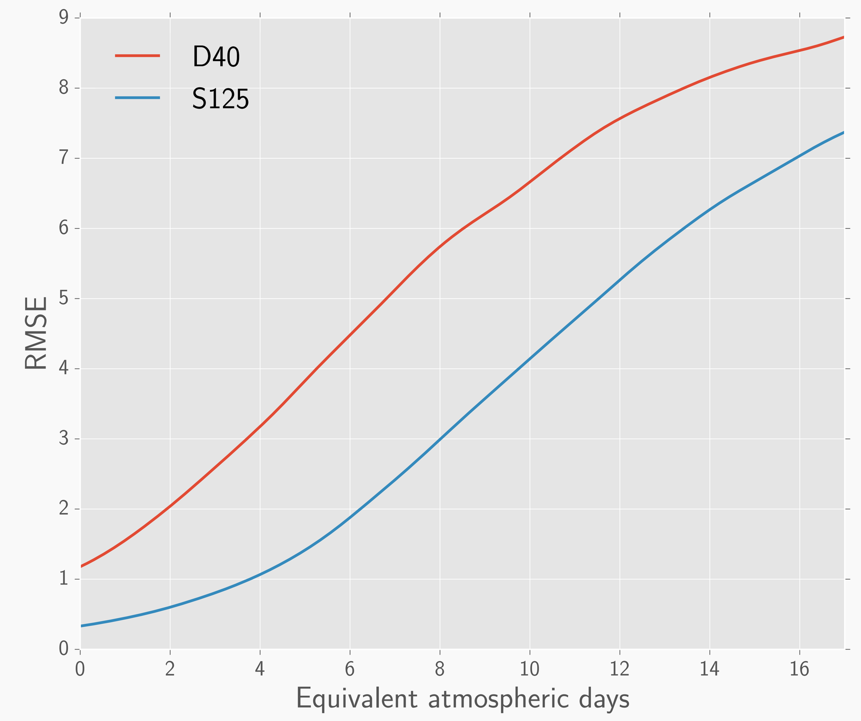

Figure 6 'Weather' forecasts initialised from conditions produced through two different assimilation systems.

- Comparing the chaoticity of the X-variables with that of the large-scale atmosphere gives a meaningful time scale for the Lorenz '96 system
- By this method, 1 model time unit is identified with 15 atmospheric days
- Figure 6 shows a comparison of forecasts initialised using the D40 and S125 assimilation setups using this time scale
- **An increase in the forecast horizon of ~4 days is observed**
- It remains to be seen whether the results presented here will also apply to a full atmospheric model
- This project will next consider an intermediate complexity model, such as SPEEDY, and eventually the Integrated Forecasting System of ECMWF
- The effect of increasing the number of observations will also be considered, as well as model resolution

Acknowledgments and references

Thornes et al., "On the Use of Scale-Dependent Precision in Earth System Modelling", *Q. J. Roy. Meteor. Soc.* 2016 (in print)
Whitaker and Hamill, "Ensemble Data Assimilation without Perturbed Observations", *Mon. Weather Rev.* **130** (2002)

Thanks to Andrew Dawson for help with the reduced precision emulator.

UNIVERSITY OF OXFORD

NERC SCIENCE OF THE ENVIRONMENT