



Reliable low precision simulations in land surface models

Andrew Dawson^{1,2} · Peter D. Düben^{1,2} · David A. MacLeod¹ · Tim N. Palmer¹

Received: 9 March 2017 / Accepted: 6 December 2017 / Published online: 14 December 2017
© The Author(s) 2017. This article is an open access publication

Abstract

Weather and climate models must continue to increase in both resolution and complexity in order that forecasts become more accurate and reliable. Moving to lower numerical precision may be an essential tool for coping with the demand for ever increasing model complexity in addition to increasing computing resources. However, there have been some concerns in the weather and climate modelling community over the suitability of lower precision for climate models, particularly for representing processes that change very slowly over long time-scales. These processes are difficult to represent using low precision due to time increments being systematically rounded to zero. Idealised simulations are used to demonstrate that a model of deep soil heat diffusion that fails when run in single precision can be modified to work correctly using low precision, by splitting up the model into a small higher precision part and a low precision part. This strategy retains the computational benefits of reduced precision whilst preserving accuracy. This same technique is also applied to a full complexity land surface model, resulting in rounding errors that are significantly smaller than initial condition and parameter uncertainties. Although lower precision will present some problems for the weather and climate modelling community, many of the problems can likely be overcome using a straightforward and physically motivated application of reduced precision.

Keywords Numerical precision · Floating-point arithmetic · Single precision arithmetic · Climate models · Land surface models · Deep soil processes

1 Introduction

Climate models require considerable computing power and energy to integrate. Historically the climate modelling community has been able to rely on ever increasing computing power, allowing a steady increase in model complexity with time. In more recent years this increase in computing power has come, not from more powerful processors, but from more processors working in parallel. This paradigm shift in high performance computing has caused problems with scalability of climate models; more parallelism means dividing work up into more “chunks”, with more communication between processors. As model resolution increases this communication becomes a limiting factor on the ability of a model to scale. Moreover, this situation seems unlikely

to change, with future proposals for HPC hardware tending towards more massive parallelism. Hence, climate model developers must do whatever they can to produce models that are efficient and scalable on current and future HPC hardware.

One option for improving model efficiency is reduced numerical precision. Often climate models will use 64 bits (double precision) to represent each real-number variable component in the simulation. If the number of bits used to represent each number is reduced to 32 bits (single precision), then a model could become more efficient due to both potentially faster arithmetic, and crucially a reduced volume of data being stored in memory and communicated between processors. Of course this efficiency improvement would not come for free; the reduction in the number of bits will correspond to a reduction in numerical precision in the model. However, it has been demonstrated that lower precision can still yield accurate simulations, particularly when computational savings are reinvested into model complexity; for example increasing resolution beyond what was affordable at double precision (e.g., Düben and Palmer 2014; Russell et al. 2015; Thornes et al. 2017; Vana et al. 2017).

✉ Andrew Dawson
andrew.dawson@ecmwf.int

¹ Atmospheric, Oceanic and Planetary Physics, Department of Physics, University of Oxford, Oxford, UK

² Present Address: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

Whilst there is certainly an emerging trend for lower precision computation, often stemming from the availability of hardware accelerators such as general purpose graphics processing units and similar technologies (e.g., Göttsche et al. 2007; Baboulin et al. 2009; Clark et al. 2010; Le Grand et al. 2013), the use of increased numerical precision as high as 128 and 256 bits has also been seen as useful in some parts of the scientific computing community (Bailey et al. 2012). This has led some in the weather and climate community to question how we should be handling numerical precision, should we go lower, or higher? Whilst increased precision may help with some areas of scientific computation, we argue that it isn't as relevant to weather and climate prediction due to the chaotic and uncertain nature of our models. Many Earth system model components (physical parametrizations, sub-models etc.) have a large amount of inherent uncertainty in their formulation. Often this uncertainty is much larger than implied by our choice of 64-bit precision, and in reality the information content of the numbers we are computing is likely to be considerably less than 64 bits. In fact we may have inadvertently over-engineered large parts of our models in terms of precision by defaulting to double precision without assessing whether it is necessary or appropriate for a particular computation.

In this paper we focus on land surface schemes. These schemes are used within weather and climate models to parametrize the interactions between the atmosphere and the land (soil, vegetation etc.). These schemes are representative of many parts of an Earth System model where the underlying physics is represented using simplified approximations, and there are relatively large uncertainties involved in the model formulation (e.g., Barrios and Francés 2012; Carsel and Parrish 1988). As is common in weather and climate modelling, these schemes are often computed in high precision, even though the actual information content in the variables being computed may be lower. Another reason to choose land surface schemes as a focus is the recent work of Harvey and Verseghy (2016), who studied the effect of single precision arithmetic on deep soil climate simulations. They found that single precision led to large errors in long-term simulations of deep soil temperature, and rightly raised concerns about the impact of reduced precision in climate models, but in doing so dismissed the idea that the weather and climate community should consider moving models to lower precision outright. This paper presents an alternative view of the problem discussed in *Harvey and Verseghy*, in which we can have both an accurate simulation, and save computational resources by using lower precision. Section 2 discusses the problems described in Harvey and Verseghy (2016), Sect. 3 demonstrates the application of low precision in an idealised land-surface representation, Sect. 4 presents low precision formulations of a real land surface scheme that preserve accuracy. Conclusions are presented in Sect. 5.

2 Low precision for land surface models

Whilst the application of reduced precision to numerical simulations looks promising for performance and computational cost, it is possible that it may also introduce problems due to larger rounding errors. An example of such a problem is given in Harvey and Verseghy (2016), where deep soil processes in a climate simulation are disrupted in the CLASS model (Verseghy 2000) when it is switched from 64-bit double precision to 32-bit single precision. Their key finding is that 32-bit precision is not sufficient to resolve small changes in temperature in a deep soil column over century time-scales.

This problem comes about due to the way that computers represent real numbers using a finite number of bits. The ubiquitous IEEE floating-point representation of real numbers divides the bits into three sections: 1 bit for the sign of the number; a sequence of bits that determine the magnitude of the number, called the exponent; and a sequence of bits that determine the digits of the number, called the significand (or sometimes the mantissa). When two floating point numbers are added together, the significand bits of the smaller number are shifted to the right until the two exponents match. This shifting can cause a loss of precision in the result. The worst-case scenario for precision loss is if one of the numbers is much smaller than the other, when the loss of precision can be total, meaning all the non-zero bits in the smaller number's significand are lost and the result of the addition will simply be the larger number. Harvey and Verseghy (2016) identify this as the reason their model fails to represent deep soil temperatures when using 32-bit arithmetic.

Harvey and Verseghy (2016) propose that the climate and weather modelling community avoid this potential problem by always using double, or even higher precision. They mention that the use of extended precision (128-bit, and even 256-bit arithmetic) is becoming more common for computations in other scientific disciplines, and that climate modelling should also be following this trend towards higher precision. In contrast, we argue that high precision may be appropriate for calculations where there is a precise deterministic solution that can be computed, but this is not the situation for weather and climate models. The atmosphere behaves as a chaotic dynamical system, which disallows us the ability to compute a precise solution for even a short forecast lead time. Given the error due to truncating our models at a particular spatial scale combined with using simplified approximations to represent subgrid-scale effects, there is considerable uncertainty in our model predictions. Indeed, this is the basis of the argument for moving away from deterministic prediction of weather and climate, and using probabilistic

methods instead. In the context of a probabilistic prediction we care less about the precise solution of any particular model realisation, and more about the statistics of an ensemble. From this perspective one might argue that investing more computing power to perform computations very precisely is a waste of resources.

Harvey and Verseghy’s suggestion is to simply increase precision everywhere in order to allow a very deep and therefore poorly represented layer in the land surface scheme to be integrated. Given our understanding of our models and their inherent errors and uncertainties, it seems paradoxical that in order to correctly resolve processes that make the smallest impact upon a state variable, we should require the highest and thus most expensive numerical precision. Our aim is to find a way to resolve this apparent paradox.

3 Low precision in an idealised land surface

The problem of not being able to resolve small increments to a state variable at low precision is likely to be common across many model components, particularly when considering long climate integrations where changes occur slowly. To address this problem we will start by considering a highly simplified numerical model, constructed such that it provides an analogue for the model studied by Harvey and Verseghy (2016). For this purpose the simplest possible representation of a land surface has been chosen, modelled by the one-dimensional heat equation

$$\frac{\partial T}{\partial t} = D \frac{\partial^2 T}{\partial z^2}, \tag{1}$$

where $T = T(t, z)$ is the temperature at time t and depth z and D is a thermal diffusivity that is assumed constant. The model is set up as a column of soil of depth H with an initially constant temperature profile $T(0, z) = 273.15$ K. At the initial time an instantaneous forcing is applied to the surface, holding it at a constant value $T(t > 0, 0) = 280$ K. The bottom boundary is treated as a perfect insulator implying that $\frac{\partial T}{\partial z} = 0$ on the bottom boundary $z = -H$. The thermal diffusivity constant is chosen to be $D = 7 \times 10^7 \text{ m}^2\text{s}^{-1}$ in an attempt to use a physically reasonable value for soil.

The equation is discretized over J evenly-spaced levels in the vertical separated by $\Delta z = 1$ m, and N discrete time steps of size $\Delta t = 1800$ s. The vertical discretization consists of a centred finite-difference approximation, and the time discretization is via a forward-Euler time-stepping scheme, giving:

$$T_j^{n+1} = T_j^n + \Delta t \left[D \frac{T_{j+1}^n - 2T_j^n + T_{j-1}^n}{(\Delta z)^2} \right]. \tag{2}$$

The choice of evenly spaced levels and a forward difference time-step are made to simplify the following discussion.

We note at this point that alternative formulations using unevenly spaced vertical levels as typically found in real land surface models such as CLASS, and a Crank–Nicolson implicit time-stepping scheme have been tested and do not change any of the findings in this section.

The numerical solution is computed first using double precision (64-bit) arithmetic; the evolution of temperature in this model is shown in Fig. 1a. A second solution is obtained by using single precision (32-bit) floating-point arithmetic, shown in Fig. 1b. The single precision solution is not able to resolve the small changes in temperature that occur at lower levels, and thus the solution is rather different from the double precision solution. This highly simplified model is exhibiting the same problem observed by Harvey and Verseghy (2016) in a real land-surface model, whereby it appears that 64-bit floating-point arithmetic is required in order to obtain the correct model solution.

The problem with the single precision solution is that the tendency term (right-hand-side of Eq. 1, and the term in square brackets in the discretized Eq. 2) can be much smaller than the soil temperature. This difference in size cannot be resolved by a single precision float with 24 significant bits, so adding the very small tendency to the current temperature actually results in adding zero to the current temperature. It is important to realise that the problem with the single precision solution is not that the tendency is too small to be represented by a single precision floating point number, but that its size relative to the current model state is so small that adding the two results in a complete loss of precision and hence no change to the state. Recognising this allows us to attempt to resolve this issue by computing the tendency term in single precision, whose 8 bit exponent should have plenty of range to represent the required small tendencies, but storing and updating the model state in double precision. This procedure is detailed in Algorithm 1.

Algorithm 1 Time-stepping with mixed precision. The

right-hand-side tendency term in Eq. 2 is evaluated using 32-bit precision, while the state variable update is done at higher 64-bit precision.

do

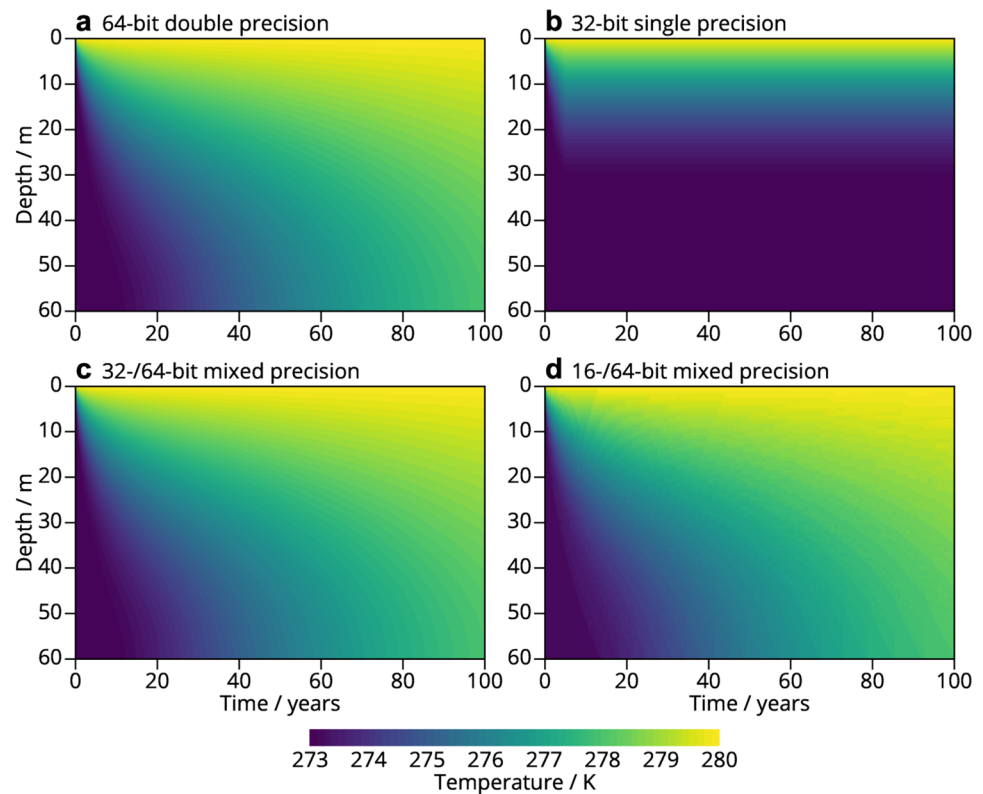
$T_{32bit} := T_{64bit}$ make a low precision copy of the model state

$\epsilon := \text{compute_tendency}(T_{32bit})$

$T_{64bit} := T_{64bit} + \Delta t * \epsilon$ update state in high precision

end

Fig. 1 Time evolution of temperature in the simple heat equation model (Eq. 2) in 4 configurations: **a** the whole model with 64-bit double precision, **b** the whole model with 32-bit single precision, **c** the model state updated in 64-bit double precision, but the tendency term computed in 32-bit single precision, and **d** the model state updated in 64-bit double precision, but the tendency term computed in emulated 16-bit half precision



The solution obtained using mixed precision is shown in Fig. 1c. This solution is the same as the double precision solution even though single precision was used to compute the tendency term. This demonstrates that a single precision float has the range required to store the tendency, which is then used to update a higher precision state in an operation which has the precision required to resolve the difference between the current state and the tendency. This point can be further demonstrated by running the model in mixed precision using the rpe software (Dawson and Düben 2016, 2017) to emulate IEEE 16-bit half-precision floating-point for the tendency term computations. The results of this model simulation are shown in Fig. 1d. In this simulation the temperature in the lower layers is evolving smoothly as in the double precision reference (Fig. 1a), indicating that even a 5-bit exponent is sufficient to represent the small tendencies. However, the solution is more noisy, particularly near the surface, suggesting that half-precision is probably too low a precision for this particular model implementation.

Harvey and Verseghy (2016) spend some time discussing the influence of the chosen time-step on the numerical stability of a low precision simulation. They reasoned that a smaller time-step is often used to solve numerical stability issues, and performed experiments with time-steps smaller than their default 1800s. They found that these experiments all produced worse results than with the default time-step. These experiments make sense in the context where we think

of numerical instability as being caused by incrementing a state variable by too much too quickly, which corresponds to our time-step being too long, but a loss of precision is actually the reverse scenario, and instead of taking shorter time-steps to resolve the issue we should actually be thinking about taking longer time-steps. Figure 2 shows the evolution of temperature at 30 m depth in the heat equation model. The solid black line is the solution obtained by integrating the model using double precision arithmetic. The dashed lines are the result of integrating the model using single precision arithmetic throughout, shown for a selection of time-steps from 30 min to 8 h. These results show that as the time-step is increased, we do a better job of representing the temperature evolution, which is consistent with our understanding that the problems at low precision are caused by trying to add temperature increments that are too small to be resolved relative to the current model temperature. The corollary is that we might expect the simulation to get even worse if we took time-steps shorter than 30 min, just as Harvey and Verseghy (2016) found. However, this result must be interpreted with caution. The heat equation model presented here is extremely forgiving in terms of time-step, we can take large time-steps and still maintain the desired evolution over a century time-scale at all depths. A more complex model of the land surface is unlikely to be so forgiving, and we would expect the larger time-step to introduce errors into the system, particularly near the surface where the forcing from the

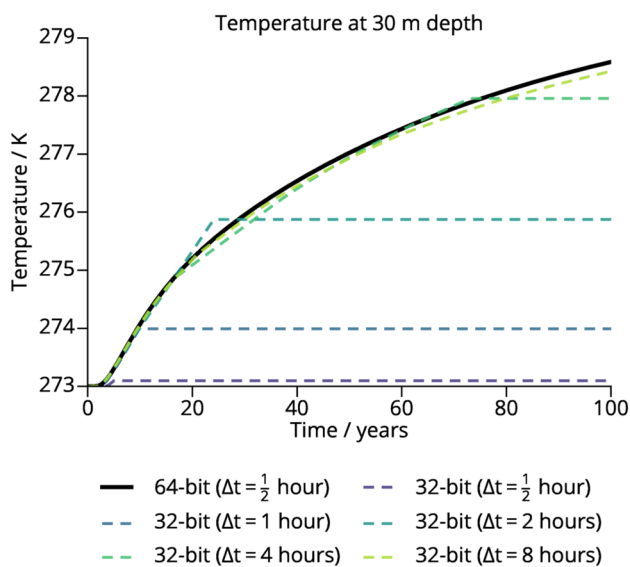


Fig. 2 Time evolution of temperature at 30 m depth in the simple heat equation model (Eq. 2). The solid black line is the double precision model configuration using a time-step $\Delta t = 1800$ s. The dashed curves are single precision configurations with time-steps starting at $\Delta t = 1800$ s and doubling up to $\Delta t = 8$ hours

atmosphere varies on short time-scales (e.g. we might lose our ability to represent the diurnal cycle). Given this, and other constraints on time-step due to coupled components, it is therefore unlikely that increasing the time-step will be a viable solution for enabling lower precision integrations in real land surface models.

We have demonstrated that a simple model that fails when using lower precision can, with only minimal changes, be modified to work correctly. This method relies on computing tendencies in lower precision but maintaining and updating the model state in a higher precision. In a more complex land surface model the computation of the tendency term is likely to be much more expensive than the update of the model state, which still leaves an opportunity for substantial computational savings to be made from using single (or even lower) precision, even if only for the tendency computations.

The addition of soil moisture in a land surface model may present an extra challenge to this methodology due to large variations in diffusivity/conductivity of wet and dry soils. However, it is likely that increments of temperature and soil moisture in such a model can be represented well by single precision, and that performing only the state update in higher precision can still prove effective. This is considered in more detail in Sect. 4.

4 Low precision in a real land surface model

In the previous section we outlined a potential solution to the problem of time-stepping a simple heat equation model with small increments using low precision, where we use lower precision for the majority of the model calculations, but use a higher precision to maintain and update the temperature state variable. The heat equation model used in Sect. 3 is extremely simplified compared to a real land surface model. In order to propose this as a solution for more complex models, we must test the theory on a larger model.

4.1 The HTESSEL model

In this section we examine the impact of reduced floating-point precision on the HTESSEL land-surface component of the ECMWF IFS model. The HTESSEL model is the Tiled ECMWF Scheme for Surface Exchanges over Land (TESSEL) with revised land-surface hydrology (Balsamo et al. 2009), including the van Genuchten formulation for hydraulic conductivity (van Genuchten 1980) and a spatially varying soil-type map. HTESSEL comprises a surface tiling scheme which allows each grid box to have a time-varying fractional cover of up to six tiles over land (bare ground, low and high vegetation, intercepted water and shaded and exposed snow) and two over water (open and frozen water). Each tile has a separate energy and water balance, which is solved and then combined to give a total tendency for the grid box, weighted by the fractional cover. The soil is discretized vertically into four layers below ground at 7, 21, 72 and 189 cm.

4.2 HTESSEL model configurations

The model experiments presented here are run offline without explicit coupling to an interactive atmosphere. Each experiment is initialized on the 1st May each year from 1981 to 2005 and run for a 4 month season, a time-scale for which land-surface processes are particularly relevant for predictability (e.g., Wang and Kumar 1998), using a time-step of 1800s. Initial conditions for soil temperature, moisture and ice temperature (where frozen soil is present) come from the ERA-Interim Land reanalysis (Balsamo et al. 2015), and other model parameters (e.g., albedo, vegetation cover, soil type) are initialized using the HTESSEL default values. The atmospheric forcing is from the WFDEI meteorological forcing data set, at 0.5° spatial resolution, created by using the Watch Forcing Data methodology applied to the ERA-Interim data (Weedon et al. 2014). The forcing comprises three hourly data for the following variables: longwave and shortwave radiation, rainfall, snowfall, surface pressure, wind speed, air temperature and humidity. HTESSEL was

run for land points on a regular grid at a 10° resolution. Each grid-point is picked from the original resolution of the initial conditions and forcing instead of being interpolated over the whole grid box. The grid points are chosen indiscriminately, in that we took every 10th point from the underlying data without considering a priori local characteristics. The total number of grid points simulated per year (~ 200) samples a wide range of possible surface and climatological conditions.

4.3 Experiments with HTESSSEL

Our first step is to try and reproduce in HTESSSEL the problems with lower precision noted by Harvey and Verseghy (2016) in CLASS and discussed in Sect. 2. The HTESSSEL model uses double precision arithmetic by default. However, HTESSSEL has only four fixed layers that are much shallower than the 60 m depth of the CLASS model configuration presented in Harvey and Verseghy (2016), and is run for seasonal rather than centennial time-scales. Because of this, the temperature increments made by HTESSSEL are large enough to be well resolved in all layers even at single precision. The implementation of HTESSSEL makes it difficult to change the vertical level distribution to allow for deeper soil structure. Instead we search for a numerical precision at which the HTESSSEL model can no longer produce the correct solution due to lack of precision, and produce qualitatively similar behaviour to CLASS at single precision.

The reduced precision emulator software rpe (Dawson and Düben 2016, 2017) was used to lower the numerical precision in the HTESSSEL model below single precision. Using a 16-bit significand to represent all floating-point numbers is sufficient to reproduce the problem in HTESSSEL. Figure 3 shows the evolution of temperature in model layer 3 (28–100 cm) over the first 30 days of a seasonal forecast initialized on the 1st May 1981, for a single model cell located at 50.25°N , 120.25°E . The grey line in Fig. 3 shows the evolution of temperature in the double precision reference HTESSSEL simulation. The purple dashed line shows the temperature evolution over the same period for the 16-bit significand HTESSSEL simulation. The low-precision model exhibits qualitatively similar behaviour to CLASS and the heat equation model at single precision, where the temperature does not change for long periods, as the low precision arithmetic is struggling to resolve the small state updates required.

We performed 3 additional experiments with the 16-bit significand model with increased time-steps of 1, 2 and 4 h, to compare with the results from the heat equation model in Sect. 3. The remaining dashed lines in Fig. 3 show the temperature evolution for these experiments. Increasing the time-step is effective at resolving the numerical precision problem, both the 7200 and 14400 s time-step experiments

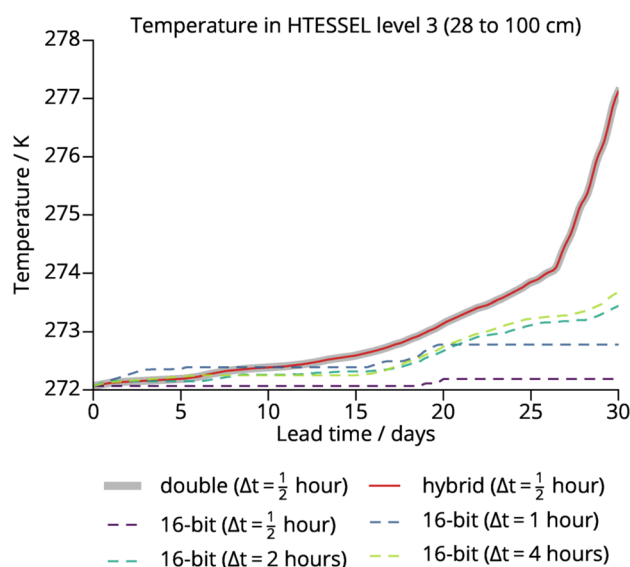


Fig. 3 Time evolution of temperature at model level 3 (28–100 cm depth) in a single cell in several HTESSSEL configurations. The solid grey line is the default double precision HTESSSEL configuration. The dashed lines are the 16-bit significand HTESSSEL simulations with varying time-steps from 30 min to 4 h. The solid red line is the hybrid precision HTESSSEL simulation with single precision for time-stepping and a 16-bit significand elsewhere

show a smooth evolution of temperature over time. However, as noted in Sect. 3, the increased time-step has also introduced significant error in the forecast temperature in layer 3. The longer time-step is also detrimental to the model's ability to represent the diurnal cycle in the upper layers (not shown). This suggests that increasing the time-step in models where low precision leads to unresolved state updates will introduce too much error for the model to remain useful, and is therefore not an acceptable way to adjust for lower numerical precision.

We now attempt to implement the algorithm described in Sect. 3 in HTESSSEL, whereby the tendency computation is performed in lower precision and the state update is done in higher precision. The HTESSSEL model was modified to use 16-bit significand precision for the low precision parts, and single precision for the high precision state update. The application of this technique in HTESSSEL is not quite straightforward, since the model does not use a simple forward Euler time-step, but an implicit time-stepping scheme, in which case we are not able to directly separate the tendency and the state in the update step. The majority of the compute time for HTESSSEL ($> 95\%$) is spent doing things other than the implicit solve for the time-stepping, which means that using low precision for everything but the implicit solve could still represent a significant saving of computational resources. This provides essentially the same functionality as the algorithm described previously, where

we attempt to do as much of the computation as possible in low precision, and doing only what is necessary at higher precision. The red line in Fig. 3 shows the temperature evolution in this hybrid precision simulation. This simulation is able to accurately reproduce the reference simulation, unlike the naïve 16-bit experiments.

In addition to examining temperature in a single grid cell, we can compare the hybrid precision simulation to the reference double precision simulation over all grid points and all May–August seasons from 1981 to 2005. We expect that the double precision and hybrid reduced precision simulations will produce different results due to accumulation of different rounding errors, so we must define what is an acceptable error. We will compute the root-mean-squared-error (RMSE) between the hybrid low precision HTESSEL model configuration and the double precision model, and compare to the ensemble spread from selected reference simulations. We consider an error to be acceptable if it is lower than the reference spread. The reference simulations are selected so that they provide a good estimate of the uncertainty in the HTESSEL model arising from initial conditions and model parameters.

Two reference simulations are chosen, the first is the ECMWF fully coupled land–atmosphere–ocean seasonal forecast System 4 (Molteni et al. 2011). This ensemble is generated by perturbing both initial conditions and model physics at run-time and is designed to represent a normal level of forecast uncertainty. This ensemble provides a good reference for errors in soil temperature, however soil moisture in the coupled model is under-dispersive, particularly for lower soil levels (Balsamo, private communication), making it a poor benchmark against which to compare errors in soil moisture. In addition to System 4 we also use the perturbed parameter ensemble described in MacLeod et al. (2016a, b). This perturbed parameter ensemble is created by perturbing two key hydrology parameters: the van Genuchten α , and the saturated hydraulic conductivity. The observational uncertainty in these parameters is large; the standard deviation of the parameters across soil samples can be as much as twice the mean (Carsel and Parrish 1988). These parameters are perturbed by picking a perturbation for each parameter in each simulation from the five-member set $0, \pm 40\%, \pm 80\%$ and applying this perturbation to the default value for each grid point. Perturbing both parameters in this way gives a 25 member ensemble. Though these perturbations are relatively large, the range of perturbations is within the range of observed variability in parameter values. We consider the spread of this perturbed parameter ensemble to be representative of the uncertainty associated with our representation of soil hydrology.

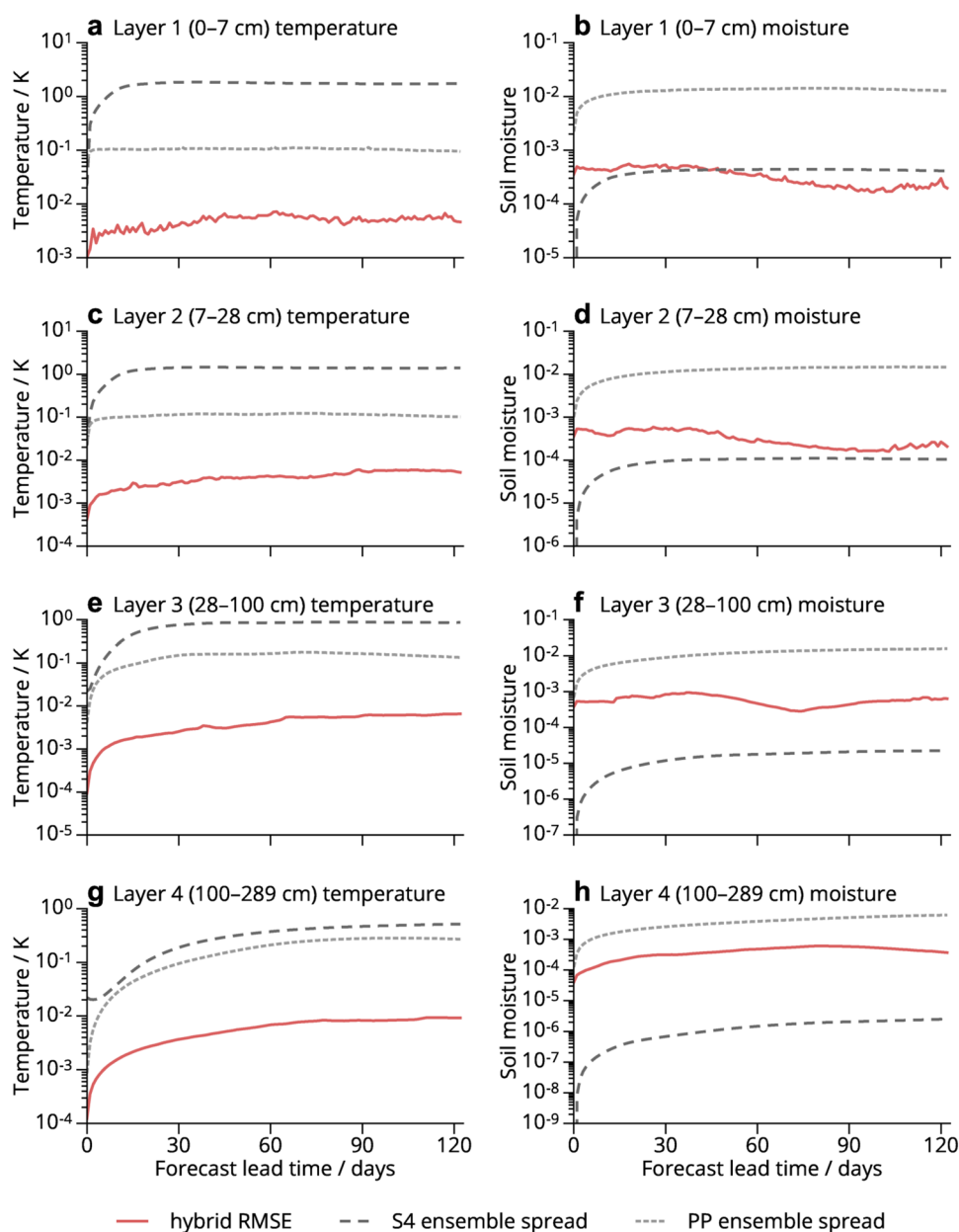
The red line in Fig. 4 shows the root-mean-squared-error in temperature and soil moisture with respect to the double precision reference for the hybrid precision HTESSEL

simulation as a function of forecast lead time. Also plotted are the ensemble spreads from the System 4 and the perturbed parameter reference experiments. For soil temperature the error in the hybrid precision HTESSEL model is well below the ensemble spread from both System 4 and the perturbed parameter ensembles in all soil layers. The error in soil moisture is consistently well below the perturbed parameter ensemble spread by ~ 1 order of magnitude. We note that the spread in soil moisture from System 4 is typically lower than the error in soil moisture produced by the hybrid precision HTESSEL model, which is expected given the suspected under-dispersive nature of soil moisture in System 4. These low errors in the hybrid precision HTESSEL model indicate that the reduced precision model is producing a physically reasonable state that is within initial condition and parameter uncertainties.

5 Discussion and conclusions

The impact of reduced numerical precision on land surface parametrization schemes has been studied. In particular we have focused on a numerical problem where a model has a quantity that is changing very slowly, which is unable to be computed correctly at low precision. The same problem is likely to occur in low precision implementations of many Earth system model components where we have quantities that vary only by small amounts over long time-scales. Harvey and Versegny (2016) claim that the weather and climate modelling community should restrict their use of precision to 64-bit double precision or higher to ensure these problems are avoided. This is a reasonable solution in the common paradigm of using one numerical precision for a whole model, but if we are willing to invest development time in exploring mixed precision approaches, the conclusion need not be so drastic. We have shown that in a simple heat diffusion model a mixed-precision approach allows the use of lower precision, provided that state variables are stored and updated in higher precision. We also demonstrated that the same basic approach can be applied to a real land surface scheme, using high precision for storing and updating state variables and lower precision for the remaining computations, to achieve a forecast that is within desired accuracy limits. This mixed-precision approach is likely generally applicable to many Earth system model components, and has been demonstrated in other areas (e.g., Düben et al. 2017; Dawson and Düben 2017). These results combined suggest that even when it appears lower precision may present problems, it could be possible to continue doing the majority of computations using lower precision and just use higher precision for selected computations that are particularly sensitive to precision. Limiting ourselves to double precision or

Fig. 4 Error in the time evolution of temperature and fractional moisture in the hybrid HTESSEL model simulation as a function of forecast lead time (red line). Also plotted are the ensemble spread from the ECMWF seasonal prediction System 4 (S4) and the MacLeod et al. (2016a, b) perturbed parameter ensemble



higher ignores the potential benefits of lower precision (e.g., it allows computer resources to be used for other more important purposes), and as we have shown, is likely an unnecessary course of action. In addition, using precision higher than double precision is a potentially expensive choice due to very limited hardware support. Harvey and Verseghy (2016) also suggested that whilst limiting ourselves to double precision as a minimum, model code should probably be tested at precisions higher than double precision to make sure double precision is sufficient. Whilst this is certainly a valid and potentially useful course of action, we suggest taking precautions in the opposite direction, that is model code should be tested with lower precision to identify code that may be problematic at lower

precision, and to check if double precision is actually more than is required for the particular problem.

The HTESSEL model configuration we have used in this study is quite different to the CLASS model discussed by Harvey and Verseghy (2016). Firstly it has fewer vertical levels, extending down to only a few metres below the surface, compared to the 60 m depth of the CLASS model, and secondly our integrations are over 4-month seasons, rather than century long integrations. However, we argue that our results are still applicable to the case of deep soil temperatures evolving over century time-scales. We have used a precision lower than the single precision examined in CLASS, which has the effect of tightening the constraint on minimum resolvable increment, and reproduces

the same problem seen in the deep soil in CLASS in the more shallow bottom layers of HTESSEL. We suggest that the CLASS model could likely make use of single precision for the majority of its computations and use double precision for time-stepping, provided that the size of the increments required at the lowest depth levels can be represented by a single precision float with an 8-bit exponent, which they likely can be. The situation may be more complex when soil moisture is considered in addition to heat diffusion due to the large variation in conductivity due to soil moisture variations. A successful implementation of a low precision land surface model should also aim not to introduce soil moisture biases for longer-range simulations. We cannot know from the relatively short simulations of HTESSEL that the reduced precision will not introduce soil moisture biases in longer range simulations, and understanding this requires further study.

We have discussed an approach to low precision in land surface models where lower precision is used for computing time-stepping tendencies, and higher precision for state updates. There are alternate approaches that may give desirable results, but that likely require more significant changes to model code. For example, one could choose to use low precision throughout the model, but increase the time-step with depth. The larger time-step at depth would ensure that increments in the lower layers where the state evolves more slowly are large enough to resolve even at lower precision, while retaining a relatively short time-step near the surface allows high frequency variability near the surface to be captured. This type of approach would likely require significant re-development of model numerics, and may not be worth it for smaller model components. Alternatively, one could vary numerical precision with depth, using higher precision to resolve the deeper layers that change more slowly. Whilst this may resolve an immediate numerical issue, it is at odds with the principle that we should use lower precision for variables that have the most uncertainty and the smallest information content.

One could argue that even if we can provide a significant improvement to the efficiency of a land surface model component, these components typically consume a very small fraction of the total computing resources required to integrate the whole forecast system and as such the performance benefit will go unnoticed. Whilst this may be true for some land surface schemes, the intent here is that the methodology is transferable to other model components, which either individually, or as a whole, likely represent a significant fraction of the total computing resource usage. In addition, as shown by Harvey and Verseghy (2016), sometimes the precision at which model components run can be dictated by external constraints (in the case of CLASS, the driving model system was changed to single precision in its entirety) and

understanding the behaviour of individual model components at lower precision can guide this type of model change.

Acknowledgements AD, PD and TP were funded by an ERC Grant (Towards the Prototype Probabilistic Earth-System Model for Climate Prediction, project number 291406). DM was funded by the EU-FP7 project SPECS (Grant agreement 308378). PD received funding from the ESIWACE project under Horizon 2020 (Grant agreement 675191). We thank two anonymous reviewers whose comments helped improve the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Baboulin M, Buttari A, Dongarra J, Krzak J, Langou J, Langou J, Luszczek P, Tomov S (2009) Accelerating scientific computations with mixed precision algorithms. *Comput Phys Commun* 180:2526–2533. <https://doi.org/10.1016/j.cpc.2008.11.005>
- Bailey DH, Barrio R, Borwein JM (2012) High-precision computation: mathematical physics and dynamics. *Appl Math Comput* 218:10106–20121. <https://doi.org/10.1016/j.amc.2012.03.087>
- Balsamo G, Beljaars A, Scipal K, Viterbo P, van den Hurk B, Hirschi M, Betts AK (2009) A revised hydrology for the ECMWF model: verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J Hydrometeorol* 10:623–643. <https://doi.org/10.1175/2008JHM1068.1>
- Balsamo G, Albergel C, Beljaars A, Boussetta S, Brun E, Cloke H, Dee D, Dutra E, Muñoz Sabater J, Pappenberger F, de Rosnay P, Stockdale T, Vitart F (2015) ERA-Interim/Land: a global land surface reanalysis data set. *Hydrol Earth Syst Sci* 19:389–407. <https://doi.org/10.5194/hess-19-389-2015>
- Barrios M, Francés F (2012) Spatial scale effect on the upper soil effective parameters of a distributed hydrological model. *Hydrol Process* 26:1022–1033. <https://doi.org/10.1002/hyp.8193>
- Carsel RF, Parrish RS (1988) Developing joint probability distributions of soil water retention characteristics. *Water Resour Res* 24:755–769. <https://doi.org/10.1029/WR024i005p00755>
- Clark MA, Babich R, Barros K, Brower RC, Rebbi C (2010) Solving lattice QCD systems of equations using mixed precision solvers on GPUs. *Comput Phys Commun* 181:1517–1528. <https://doi.org/10.1016/j.cpc.2010.05.002>
- Dawson A, Düben PD (2016) rpe v5.0.0. <https://doi.org/10.5281/zenodo.154483>. <http://github.com/aopp-pred/rpe/tree/v5.0.0>
- Dawson A, Düben PD (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 PD, Palmer TN (2014) Benchmark tests for numerical forecasts on inexact hardware. *Mon Weather Rev* 142:3809–3829. <https://doi.org/10.1175/MWR-D-14-00110.1>
- Düben PD, Subramanian A, Dawson A, Plamer TN (2017) A study of reduced numerical precision to make superparametrisation more competitive using a hardware emulator in the OpenIFS model. *J Adv Model Earth Syst*. <https://doi.org/10.1002/2016MS000862>
- Göddecke D, Strzodka R, Turek S (2007) Performance and accuracy of hardware-oriented native-, emulated- and mixed-precision solvers

- in FEM simulations. *Int J Parallel Emerg Distrib Syst* 22:221–256. <https://doi.org/10.1080/17445760601122076>
- Harvey R, Verseghy DL (2016) The reliability of single precision computations in the simulation of deep soil heat diffusion in a land surface model. *Clim Dyn* 46:3865–3882. <https://doi.org/10.1007/s00382-015-2809-5>
- Le Grand S, Götz A, Walker RC (2013) SPFP: speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. *Comput Phys Commun* 184:374–380. <https://doi.org/10.1016/j.cpc.2012.09.022>
- MacLeod DA, Cloke HL, Pappenberger F, Weisheimer A (2016) Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Q J R Meteorol Soc* 142:79–90. <https://doi.org/10.1002/qj.2631>
- MacLeod DA, Cloke HL, Pappenberger F, Weisheimer A (2016) Evaluating uncertainty in estimates of soil moisture memory with a reverse ensemble approach. *Hydrol Earth Syst Sci* 20:2737–2743. <https://doi.org/10.5194/hess-20-2737-2016>
- Molteni F, Stockdale T, Balmaseda M, Balsamo G, Buizza R, Ferranti L, Magnusson L, Mogensen K, Palmer T, Vitart F (2011) The new ECMWF seasonal forecast system (System 4), Tech. Rep. 656, European Centre for Medium-Range Weather Forecasts, Reading, UK
- Russell FP, Düben P, Niu X, Luk W, Palmer TN (2015) Architectures and precision analysis for modelling atmospheric variables with chaotic behaviour. In: 2015 IEEE 23rd annual international symposium on field-programmable custom computing machines, pp 171–178
- Thornes T, Düben PD, Palmer TN (2017) On the use of scale-dependent precision in Earth system modelling. *Q J R Meteorol Soc*. <https://doi.org/10.1002/qj.2974>
- van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44:892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Vana F, Düben PD, Lang S, Palmer TN, Leutbecher M, Salmond D, Carver G (2017) Single precision in weather forecasting models: an evaluation with the IFS. *Mon Weather Rev* 145:495–502. <https://doi.org/10.1175/MWR-D-16-0228.1>
- Verseghy DL (2000) The Canadian land surface scheme (CLASS): its history and future. *Atmos Ocean* 38:1–13. <https://doi.org/10.1080/07055900.2000.9649637>
- Wang W, Kumar A (1998) A GCM assessment of atmospheric seasonal predictability associated with soil moisture anomalies over North America. *J Geophys Res* 103:28637–28646. <https://doi.org/10.1029/1998JD200010>
- Weedon GP, Balsamo G, Bellouin N, Gomes S, Best MJ, Viterbo P (2014) The WFDEI meteorological forcing data set: WATCH forcing data methodology applied to ERA-Interim reanalysis data. *Water Resour Res* 50:7505–7514. <https://doi.org/10.1002/2014WR015638>