Investigating the Potential for Improving the Accuracy of Weather and Climate Forecasts by Varying Numerical Precision in Computer Models

Thesis submitted for the degree of

Doctor of Philosophy

by

Tobias Thornes

Oriel College

University of Oxford

Hilary 2018

Abstract

Accurate forecasts of weather and climate will become increasingly important as the world adapts to anthropogenic climatic change. Forecasts' accuracy is limited by the computer power available to forecast centres, which determines the maximum resolution, ensemble size and complexity of atmospheric models. Furthermore, faster supercomputers are increasingly energy-hungry and unaffordable to run.

In this thesis, a new means of making computer simulations more efficient is presented that could lead to more accurate forecasts without increasing computational costs. This 'scale-selective reduced precision' technique builds on previous work that shows that weather models can be run with almost all real numbers represented in 32 bit precision or lower without any impact on forecast accuracy, challenging the paradigm that 64 bits of numerical precision are necessary for sufficiently accurate computations. The observational and model errors inherent in weather and climate simulations, combined with the sensitive dependence on initial conditions of the atmosphere and atmospheric models, renders such high precision unnecessary, especially at small scales. The 'scale-selective' technique introduced here therefore represents smaller, less influential scales of motion with less precision. Experiments are described in which reduced precision is emulated on conventional hardware and applied to three models of increasing complexity. In a three-scale extension of the Lorenz '96 toy model, it is demonstrated that high resolution scale-dependent precision forecasts are more accurate than low resolution high-precision forecasts of a similar computational cost. A spectral model based on the Surface Quasi-Geostrophic Equations is used to determine a power law describing how low precision can be safely reduced as a function of spatial scale; and experiments using four historical test-cases in an open-source version of the real-world Integrated Forecasting System demonstrate that a similar power law holds for the spectral part of this model. It is concluded that the scale-selective approach could be beneficially employed to optimally balance forecast cost and accuracy if utilised on real reduced precision hardware.

Contents

Acknowledgements

The investigation presented in this thesis was based on the ideas of the author in discussion with Tim Palmer and Peter Düben, who supervised this work and to whom I owe my principal thanks. I would also like to thank Peter Düben and Andrew Dawson for developing and providing permission to use the reduced precision emulator used in the majority of the experiments; Hannah Christensen (nee Arnold) for her advice regarding the Lorenz ’96 system, parameterisation schemes and weather regimes; Shafer Smith for providing the original SQG code on which my SQG program was based; the ECMWF for making available their IFS code in the form of the Open IFS; and Matthew Chantry for his help and advice regarding the installation and running of Open IFS and his implementation of the reduced precision emulator therein. This work was carried out at the Atmospheric, Oceanic and Planetary Physics sub-department at the University of Oxford, which I thank for providing the facilities and environment that made these experiments possible.

All the work presented in this thesis is my own, except where explicitly stated.

Publications

Some of the work presented in this thesis has been published in peer-reviewed scientific journals, as described below. It is hoped that the other parts of this investigation will also be published in the future. The wording and format of the published work in all cases differs from the text presented in this thesis.

Chapter 2 (Modified Lorenz '96 System experiments)

The construction of this system and the reduced precision results are published in:

Thornes, T., 2016. Can Reducing Precision Improve Accuracy in Weather and Climate Models? *Weather*, 71, pp.147-150.

Thornes, T., Düben, P. D. & Palmer, T., 2017. On the Use of Scale-Dependent Precision in Earth System Modelling. *Q. J. R. Meteorol. Soc.*, 143, pp.897-908.

Chapter 3 (Error propagation in the SQG System)

These results have been written up into a paper, an it is hoped that they will be published in future.

Chapter 4 (Reduced Precision in the SQG System)

The results of these experiments have been published online, awaiting print edition:

Thornes, T. Düben, P. D. & Palmer, T., 2018. A Power Law for Reduced Precision at Small Spatial Scales: Experiments with an SQG Model. *Q. J. R. Meteorol. Soc.*, in press.

Chapter 5 (Reduced Precision in Open IFS)

The results presented here have been combined with those of other researchers in the Predictability group at the University of Oxford to form a paper currently in preparation.

List of Abbreviations

2D	Two-Dimensional
2DV	Two-Dimensional barotropic Vorticity
3D	Three-Dimensional
4D-Var	Four-Dimensional Data Assimilation
BS(S)	Brier (Skill) Score
CET	Central England Temperature
CPU	Central Processing Unit
D	Double-precision model with one resolved tier (modified Lorenz ‘96)
DD	Double-precision model with two resolved tiers (modified Lorenz ’96)
ECMWF	European Centre for Medium-range Weather Forecasts
EDT	Error Doubling Time
ENSO	El Nino Southern Oscillation
EOF	Empirical Orthogonal Function
EWP	England and Wales Precipitation
FFT	Fast Fourier Transform
FFTW	‘Fastest Fourier Transform in the West’
FPGA	Field Programmable Gate Array
GCM	Global Climate Model
GMT	Greenwich Mean Time
GPU	Graphical Processing Unit
HPC	High Performance Computing
IEEE	Institute of Electrical and Electronics Engineers
IFS	Integrated Forecasting System
IGCM	Intermediate Global Climate Model
IGN(SS)	IGNorance (Skill) Score
IPCC	Intergovernmental Panel on Climate Change

KS	Kolmogorov-Smirnov
MA	Multiplicative-Additive parameterisation
MARS	Meteorological Archive and Retrieval System
MTU	Model Time Unit
NaN	Not a Number
NWP	Numerical Weather Prediction
PC	Principal Component
pdf	probability density function
QBO	Quasi-Biennial Oscillation
QG	Quasi-Geostrophic
RMSE	Root Mean Square Error
RPS(S)	Ranked Probability (Skill) Score
Setup L	Long-term forecast setup (SQG system)
Setup S	Short-term forecast setup (SQG system)
SD	State-Dependent parameterisation
SH	Single-precision X and Half-precision Y model (modified Lorenz '96)
SP	Stochastic Processor
SST	Sea Surface Temperature
SQG	Surface Quasi-Geostrophic
WRF	Weather Research and Forecasting

1. Introduction

For many years, it has been assumed that for forecasts of weather and climate to be accurate, the computer models used to produce these forecasts must be run at a relatively high numerical precision. The default 'double-precision' standard used for most computer programs is conventionally applied to models of Earth's atmosphere with no discrimination between parts of these models where representing values very exactly is more important or less important to the overall output. By 'numerical precision' is meant the number of bits of information – 0s and 1s – used to represent numbers that are stored on or used by computers. This essentially determines the number of decimal places to which a number can be represented, and the range of values that it can take. But physical intuition, backed up by recent studies, suggests that many of the variables used in numerical models of the atmosphere are unlikely to take values that span such a large range, or to be observable or predictable to such a high degree of accuracy, that it is necessary to represent them in double-precision for a forecast to be accurate. Doing so may hence incur unnecessary computational costs. One especially intriguing possibility is that smaller-scale quantities are less accurately known and less important to the model output, and therefore may require less numerical precision than larger-scale quantities. The aim of this thesis is to test this possibility, using a range of different models as analogues to determine whether a scale-selective approach to numerical precision could be used to improve the computational efficiency of real-world forecasts.

This section introduces the fundamental premise behind this thesis and the questions that it is intended to answer in more detail. It begins (Section 1.1) with a theoretical justification for taking a scale-selective approach to atmospheric modelling, by exploring the scale-dependent nature of the Universe as a whole. It then explains why reduced precision weather and climate forecasts could be of practical importance, and why it is that the world needs more efficient forecasts in the first place (Sections 1.2 and 1.3). An outline of different methods that can be used to reduce precision (Section 1.4) and how two of these ('Mixed Precision' and 'Stochastic Processors') will be implemented in this study (Section 1.5) is given, alongside a brief summary of what can be gleaned from other experiments of a similar kind that were conducted prior to the

investigation that will be presented in the remaining chapters of this thesis (Section 1.6). Finally, methods used in subsequent chapters to analyse the quality of forecasts are described (Section 1.7).

1.1. The Importance of Scale-Selectivity

Our universe is inherently scale-dependent. Although the fundamental forces that govern physical interactions are not believed to vary in space or time, because the relationship between the force strength and the separation of the interacting entities is not the same for different forces, different processes predominate and different physical behaviour emerges on larger scales than at smaller scales. On the largest scales – those of stars, planets and galaxies – the universe is governed by gravity, through which matter interacts in a deterministic (albeit not always easily predictable) way described by well-tested laws of Newtonian and relativistic mechanics.

At the smallest scales, meanwhile – the scales of sub-atomic particles – it is the theory of quantum mechanics that makes the most accurate predictions, and the exact position and momentum of any given particle cannot be defined simultaneously: on these scales the universe is not deterministic. Our everyday reality on Earth, including the atmosphere that surrounds us, comprises such a large aggregation of interacting fundamental particles that quantum mechanical effects are not ordinarily observed, yet nor is our own mesoscale existence entirely deterministic. Our reality is at the fulcrum between the large and the small, the predictable and the unpredictable, and is dominated by chaos.

The theory of chaos tells us that nothing in the everyday world is indefinitely predictable. Applied to Earth's atmosphere, the physical material with which this thesis is concerned, it means that the weather cannot be predicted with any certainty in a very specific place or very far into the future. The atmosphere, in other words, has a finite predictability that can vary according to the state of the weather. Longer-term, 'climate', forecasts can be produced, but these only predict how the average state of the atmosphere or the frequency of particular events will change over time; not the exact state of the atmosphere (the weather) at a specific distant time. The extent of

predictability is reflected in the 'skill' of a weather forecast over the course of the following days, which is to say the extent to which it is able to successfully predict the observed state of the atmosphere, and the number of days out to which the forecast contains any skill at all, and depends upon scale-dependent phenomena. Large-scale systems such as fronts (on the order of tens of kilometres in size), arising because of the baroclinic instability associated with large-scale temperature gradients in a rotating atmosphere, increase the predictability of the atmosphere in a given region because they move and evolve in a largely deterministic manner: one can watch the approach of a storm on RADAR or satellite images and calculate by eye the rough path that it is going to take over the course of minutes to hours. Unpredictability, on the other hand, wells up from the smallest scales – the scales of individual particles, eddies and gusts – so that, like the often cited butterfly flapping its wings, in the absence of large-scale systems damping down smaller-scale effects tiny unpredictable fluctuations can push the atmosphere into one of any number of states. These fluctuations may have their origins in quantum uncertainty or may arise from deterministic events that only appear random through lack of information regarding their origins. An analogy for such 'quasi-random' deterministic events is the colour of the cars passing an observer stationed at the side of a busy road: the sequence of colours may appear random to the observer, but in principle it could be predicted if that observer knew the actions and intentions of all the drivers using that road, and the colours of their vehicles. In any case, such fluctuations are too small to be observed accurately on a global scale – at least, with today's technology – and their influence is hence unforeseeable.

Earth's atmosphere therefore exhibits very different behaviour on large- and small-scales, which to some extent mimics the fundamental dichotomy between large-scale predictability and small-scale unpredictability in the universe overall. This fundamental physical property provides the ultimate motivation for the investigation presented in this thesis, which questions the use of uniform numerical precision to represent phenomena occurring on all scales in the atmosphere, and probes whether more efficient use could be made of finite computing resources by representing smaller spatial scales with less precision in computer models.

This idea becomes all the more plausible when one considers the limitations inherent within all numerical models of the atmosphere. Edward Lorenz, the pioneer of chaos theory, showed explicitly that weather forecasts could not accurately predict the weather at more and more distant lead-times indefinitely, simply because the observations used to initialise these models will never be perfect (Lorenz 1993). A completely accurate picture of the present state of the atmosphere would, in principle, allow a completely accurate forecast to be made out to infinite time, ignoring quantum effects. But as soon as there is any observational error at all – even on the scale of a butterfly's wing-flap that the observations fail to include – that error will grow exponentially and scupper the entire forecast within two or three weeks at the most. Such is the nature of chaos. The 'Error Doubling Time' (EDT), the time taken for an initial error to double in size, is for this reason much larger in our atmosphere on larger scales than on smaller ones, where it takes longer for the error to grow. Thus, unless our observations are perfect at the smallest scales, representing them at very high precision may be rendered purposeless by the fact that any observational errors will immediately begin to grow and dominate those scales.

Furthermore, the paucity, mixed quality and non-uniformity of the observational network, even in the satellite era, must constrain the precision of a model's initial conditions, and the numerical precision should be adjusted to reflect this. If we can only be sure of the temperature at a given grid-point to one decimal place but represent it in our models to ten decimal places, most of that data is meaningless. All of this applies even to a 'perfect' model that can exactly represent all physical processes. In reality, of course, no numerical model of the atmosphere is perfect. Firstly, the equations we use to describe reality are just that – a description based on what we observe, not a prescription that nature will obey exactly, especially as we approach the more uncertain, smaller scales. Secondly, computers are not infinitely powerful, and do not have the resources to solve these equations down to very small scales: all atmospheric models have a fixed 'resolution', the smallest scales that they can capture, and if the effects of phenomena that affect the atmosphere on smaller scales than this – convective clouds, trees and other surface features, and so on – are to be included at all, they must be 'parameterised', as explained in Section 1.3. Parameterisation is not at all accurate compared to resolving features explicitly, and again renders the smallest scales in the

model – those that depend most closely on the impact of still smaller scales that cannot be resolved – especially uncertain. All of these four reasons – scale-dependent Error Doubling Times, observational uncertainties, model errors and limited model resolution – mean that in numerical models, as in the universe as a whole, smaller-scale phenomena cannot be known with as great a degree of precision as larger-scale ones.

1.2. Why do we Need Good Models of the Atmosphere?

The dream of being able to predict the weather hours, days or even weeks in advance has been with humanity for millennia. Our day-to-day activities have always been profoundly influenced by the weather, and the benefits of knowing the conditions that we are likely to be living or working in in the near-future are obvious. Moreover, the more advanced the warning we have of extreme weather events such as heat waves and floods, the better placed we are to minimise the damage they may cause to our lives, livelihoods and property. Longer-term forecasts of the average conditions in coming seasons and years, referred to as 'climate' as opposed to 'weather' forecasts, have the potential to be of substantial benefit to society, and especially to economic activities that are particularly dependent upon temperature, wind and rainfall.

Despite numerous attempts to explain the weather by both natural and supernatural means, for most of human history this dream has remained far from a reality. At least as far back as Biblical times, it was known that a red sky at night would often mean fair weather to follow, whereas a red sky in the morning suggests rain. But because the atmosphere is such a large and complicated system, we were unable to progress beyond such tentative observations before the invention of communications technology such as the telegraph and reliable measuring instruments such as the thermometer and barometer between the seventeenth and nineteenth centuries. The new possibilities opened up by this technology and the desire to know the weather in advance – especially at sea – led to the establishment of the British Meteorological Office under Robert Fitzroy in 1854 and the first attempts at daily weather forecasts were made shortly thereafter.

But until the middle part of the twentieth century, such forecasts remained very unreliable. They rested on the inaccurate assumption that Earth's atmosphere is a periodic and macroscopically deterministic system. Weather maps were produced using observations of variables such as temperature and pressure at a finite number of surface weather stations scattered across the globe, and of Sea Surface Temperatures (SSTs) recorded by ships. Forecasts were made by looking up similar conditions in archives of previous weather maps to those presently observed and assuming that the atmosphere would evolve similarly to the way in which it did before. However, such forecasts proved to be unreliable. It was realised as early as the 1920s that, just as a given configuration of stars, planets and satellites never occurs twice in the astronomical record, so the atmosphere never 'repeats itself' exactly, even when the grid of observations is coarse (Richardson 1922). These early 'statistical' weather forecasts could therefore never be accurate enough to be of much use to society beyond vague storm warnings for ships.

The lack of repeatability in the atmosphere stems from its inherent chaos, described in Section 1.1. A chaotic system is one that appears, from the point of view of the observer, to be random, so that even very slight differences between two initially very similar states cause them to diverge over time until all similarity disappears (Lorenz 1963). Even given the coarse resolution of observing stations, the atmosphere has never been observed to repeat exactly the same state across the entire globe, and even if it did there would still be unobservable but significant differences in conditions in between the stations (Lorenz 1993). The atmosphere is a dynamical system, meaning that it evolves approximately deterministically (ignoring any true randomness that arises from quantum uncertainty) over time as described by fixed laws, which implies that it ought to be predictable. But because it is chaotic, not periodic or constant, those small unobservable differences grow rapidly over time and can lead to a profoundly different state to what would be predicted based on similar – but not identical – conditions that have occurred before. This obviously restricts the accuracy and applicability of forecasts based on statistics drawn from historical observations. But it also renders a completely analytical forecast, based on deterministic equations that describe the atmosphere's behaviour in terms of physical 'laws', impossible to realise in practice. Such a

forecast would require the position and behaviour of every constituent particle to be known exactly and simultaneously, something that is far beyond human capabilities to achieve.

It was Bjerknes who, in 1904, pioneered the first analytical description of the atmosphere's behaviour based on the principles of Newtonian mechanics and classical thermodynamics. He recognised that the atmosphere can be treated as a fluid, the motion of any small 'parcel' of which can be described using four key equations (Tribbia 1997). The first is a Newtonian equation of motion relating the change in velocity vector $\vec{V}$ over time t (the acceleration) of a parcel of air of density ρ to the forces acting on that parcel, which are the parcel's own weight under gravitational acceleration $\vec{g}$, the Coriolis force resulting from Earth's rotation at angular velocity $\vec{\Omega}$, the pressure gradient in the atmosphere $\vec{\nabla}p$ and the frictional force $\vec{F}$. Newton's Third Law of motion tells us that the net acceleration of a particle is given by the sum of the acceleration due to each of the forces acting on that particle, so that,

$$\frac{d\vec{V}}{dt} + 2\vec{\Omega} \times \vec{V} = -\frac{1}{\rho}\vec{\nabla}p + \vec{g} + \vec{F}. \tag{1.1}$$

The second key equation is the continuity equation, which encapsulates the fundamental physical principle that the fluid parcel's mass must be conserved. Because the parcel is moving, with velocity components that may change in magnitude at different rates in different directions, it can be stretched or shrunk over time, changing its density. In a mathematical sense, *divergence* (symbol $\vec{\nabla}$) describes the extent to which vector components change at different rates in different directions. To conserve the parcel's mass, the rate of change of its density must therefore depend upon the divergence of the velocity, yielding,

$$\frac{\partial \rho}{\partial t} = -\vec{\nabla} \cdot (\rho \vec{V}). \tag{1.2}$$

Thirdly, there is the equation of state of an ideal gas with dry air gas constant R, which describes how the pressure p exerted by any (ideal) gas depends upon its density and temperature T,

$$p = \rho R T. \tag{1.3}$$

Finally, one must account for the conservation of energy, which is described by the First Law of Thermodynamics. In this context, this law tells us that if some energy in the form of heat flows at

rate Q into the fluid parcel, either the internal energy of the parcel (proportional to the temperature) must increase or the parcel must do 'work' on another system or entity, thereby passing on this energy elsewhere. The heat capacity C_v describes the amount of energy a substance absorbs at constant volume per unit temperature increase, when no work is done on or by the parcel, whilst at constant pressure the rate at which work is done depends upon the rate of change of the parcel's volume, which is proportional to the rate of change of the reciprocal of the density. Hence,

$$C_v \frac{dT}{dt} + p \frac{d}{dt}\left(\frac{1}{\rho}\right) = Q. \tag{1.4}$$

A fifth equation was added by L F Richardson to describe the evolution of the specific humidity q in the atmosphere according to the sum of humidity sources and sinks, S,

$$\frac{\partial q}{\partial t} + \vec{\nabla} \cdot \left(q\vec{V}\right) = S, \tag{1.5}$$

to complete the set of five equations that remain the basis of all numerical atmospheric models. In principle, these five equations can be used to forecast the atmospheric wind velocity, pressure, temperature and density at any point in space and time given some initial state, and further equations can be derived to predict additional quantities from these. However, because they represent a coupled system of non-linear partial differential equations, in the absence of perfect observations, either graphical or numerical methods must be employed to approximate their solutions (Tribbia 1997).

Modern-day weather and climate forecasts are based on the numerical methods pioneered by L F Richardson during the First World War. Richardson envisaged that some sixty-four thousand human computers would be required to work continuously to produce real-time global weather forecasts by solving these equations numerically with pencil and paper, a 'staggering figure' (Richardson 1922) which was of course unfeasible. Hence, it was only with the advent of the digital computer that his methods could be implemented effectively and Numerical Weather Prediction (NWP) began to make real positive gains over statistical methods of forecasting. The forecasts produced and publically disseminated today depend on numerical Global Climate Models (GCMs) run using some of the most powerful computers in the world, and the huge energy demand they entail makes it essential that such models are made as accurate and efficient as possible.

Nonetheless, the chaotic nature of the atmosphere means that even the most powerful models will never be able to predict future conditions perfectly.

Today, the need for accurate weather and climate forecasts is more acute than ever because of the threat posed by climatic change. Whilst the threat itself is undoubtable, there remain considerable uncertainties regarding some of the feedback mechanisms that could worsen or ease its extent on regional scales (Palmer 2014). The coming decades will see melting ice, rising sea-levels, more instances of extreme floods and heatwaves and the extinction of species and cultures because of the climatic change caused by global warming, which will cause intense physical and emotional stress to humans and other animals (Hansen et al. 2013). But this suffering can be minimised if people are sufficiently prepared to take preventative action in advance, whether that involves moving populations, farm animals and grain to higher ground ahead of a flood or changing long-term agricultural practices to reflect altered temperatures and precipitation in the future. Accurate forecasts of specific extreme weather events and of longer-term climatic trends in the frequency and intensity of these on local scales are vital if societies are to be able to adapt in these and numerous other ways to protect their most vulnerable members. To produce such forecasts, we need our numerical models to be as good as they can be.

1.3. Current Limitations of Atmospheric Computer Models

There are two types of error associated with numerical models of the atmosphere. ‘Observational uncertainty’ in the initial conditions with which a model is primed arises inevitably because of our limited ability to observe the atmosphere exactly. But this is compounded by ‘model uncertainty’ associated with an incomplete understanding of atmospheric physics or a failure to model it accurately, which is in some cases the larger source of error – especially for climate forecasts where some physical processes or feedbacks that are not obvious may not be accounted for and model predictions are difficult to test. Furthermore, even were the physical equations known perfectly, significant model uncertainty would still arise from the need to discretise the differential

equations being solved. Numerical models must represent physical processes that are in reality continuous in space and time on a discrete grid of points, the spacing between which is referred to as the spatial or temporal grid-point 'resolution' of the model. Each of these grid-points corresponds to an area of space on the globe, and the higher the resolution, the smaller the grid-points and the more spatially specific the forecast can be. An infinitely accurate model would need to have infinitely small grid-points, but in practice the grid-points must always have a finite size because of the finite computing power available: the most advanced global weather models today have grid-boxes on the order of 10 km long on each side, and for climate forecasts it is closer to 100 km. This means that the effects of 'unresolved' phenomena that occur on smaller scales than the grid-boxes have to be imperfectly approximated, and are usually treated as statistical functions of the atmospheric variables that are explicitly resolved, using so-called 'parameterisation schemes' (Wilks 2005).

Improvements to weather and climate models can be made from two directions simultaneously, therefore: improving the parameterisation schemes used to approximate sub-grid-scale variables (where the 'art' comes into modelling) and increasing the resolution of the models so that the forecast is less dependent upon these approximations. Both deep and shallow convection, which lift air parcels to heights exceeding or limited to the 500 hPa level respectively, are especially difficult to parameterise. Ultimately, no parameterisation scheme will be able to represent clouds or deep convection accurately enough to reliably forecast precipitation. Reducing the grid-cell sizes of global models to less than 1 km, the size of individual convective clouds, will be a necessary step towards representing the dynamics of the atmosphere sufficiently to make useful forecasts of climatic change (Palmer 2014), and avoiding the need to parameterise deep convection will require still higher resolutions. Given the current operational resolution of global weather and climate models, and that every halving of resolution typically requires a six- to eight-fold increase in computational power (Mitchell et al. 2012), bringing this about will require an enormous increase in computational capacity.

As typical computer speeds and memory sizes have increased over recent decades, the resolution of the models used by global forecasters such as the European Centre for Medium-range

Weather Forecasts (ECMWF) has improved, which has been a key factor (alongside increased ensemble size and better physics schemes) in the production of increasingly accurate forecasts (Bauer et al. 2015). However, the resolution of all GCMs remains too coarse for very important features such as convective clouds and eddies in the atmosphere and oceans to be resolved, making the prediction of localised rainfall especially difficult (Tribbia 1997). Meanwhile, in spite of great improvements in forecast quality over the past few decades associated with better observations as well as increased resolution (Simmons and Hollingsworth 2002), the quality of the weather and climate forecasts that are needed for societies to be able to prepare for and adapt to climatic change remains limited. The ECMWF's latest 'System 4' model for predicting seasonal climate, for example, produced 'dangerously' unreliable forecasts when run retrospectively on 1981-2010 weather data for some parts of the world (Weisheimer and Palmer 2014).

There are two key constraints imposed on models run on any computer system: the computer speed (or 'flop rate') and memory. More memory means that more values can be computed and stored by a computer, whereas a greater speed enables those values to be computed more rapidly. If we increase the spatial or temporal resolution of forecasting models whilst still producing the forecast ahead of time, there will be more values to compute in the same amount of time and more data to store, and both the speed and the memory must therefore be increased. Until now, improvements in the resolution of weather and climate models have largely relied upon transistors becoming ever smaller in accordance with Moore's Law, allowing the density of transistors and the speed of computer processors to continually increase. But in recent years, because fundamental limits on the density of computer transistors that can be employed without overheating are being approached, a trend towards linking up multiple processors in parallel has begun, to increase the overall speed and memory of a computer whilst the speed and the energy requirements of each individual processor remain the same (Düben et al. 2014a). This means that there is now an almost linear relationship between computer speed and overall power consumption.

Although the fastest computer in the world may reach the 'exascale' level technically needed to resolve convective clouds in the 2020s, such a machine would likely not be affordable to forecast centres for many years after that, and it would take still more time for existing computer

models to be optimised to make full use of the parallel architecture. For example, substantial savings in computational cost have also been demonstrated using Graphics Processing Units (GPUs) instead of conventional Central Processing Units (CPUs) but, aside from other difficulties, just re-writing the model code to take advantage of the additional performance that GPUs' parallel structures could provide will be a lengthy and effortful process, and the fruits of any such labour many not be realised for many years (Mitchell et al. 2012). Furthermore, exascale computers will use vast quantities of electricity, potentially contributing significantly to the very problem – climatic change – which their use in forecasting is intended to compensate for. Hence, hardware, software and economic barriers (Kogge et al. 2008) conspire to render cloud-resolving global models impossible to run operationally for the foreseeable future. To run ensemble forecasts, which allow forecasters to assign probabilities to weather events, at such a resolution would be even more difficult. These difficulties are compounded by the wide variety of computer systems and models currently in operation, each of which may require code to be written differently to obtain peak performance.

1.4. Reduced Precision Methodology

Recently, new methods of improving the efficiency of models by switching to 'inexact hardware' have been proposed. Although much of this hardware is yet to be manufactured, it is within the capability of our current technology to produce it, and its implementation could therefore swiftly deliver results. Of course, all computer hardware is 'inexact' in that it inevitably has to round off numbers at some (usually high) level of precision and the calculations it carries out can never be entirely immune to error. But explicitly 'inexact' hardware allows this error to increase relative to conventional hardware in order to decrease the computational cost of each calculation. If this is done in parts of a model where conventional precision is unjustified anyway because quantities are not very precisely known, the accuracy of the program's output may be unchanged.

In a weather and climate context, this would allow models to be run at higher resolution for the same overall computational speed and memory requirements by sacrificing the precision of at least a subset of the variables (Düben et al. 2014b) to reduce computational costs. This is especially useful for variables at small spatial scales, which are known with relatively low precision because of observational constraints and the limited accuracy of forecasts close to the 'truncation scale' where parameterisation becomes necessary (see Section 1.1). On these scales, there is good reason to believe that any rounding errors associated with reducing the precision will be smaller than the model error, and more than compensated for by the increase in resolution that is obtainable in reduced precision for the same computational cost, to give a more accurate forecast overall.

Inherent observational and model error is already accounted for in many models of the atmosphere through the use of stochastic or 'random' components in the parameterisation schemes that they include, which can be used to provide some compensation for those schemes' inability to accurately represent unresolved processes by running a whole ensemble of forecasts, each of which contains a different random draw from the pool of possible influences of the unresolved variables, whose true value could be anywhere within this range. None of these members can be assumed to be more 'accurate' than the others. Computational savings from reduced precision could therefore be procured with negligible degradation of the forecast accuracy if the variables remain within the range of the uncertainty described by the ensemble spread, and indeed it is a waste of resources and physically unjustifiable to represent any one of the ensemble members' outputs to many more decimal places than the uncertainty interval that this spread implies. The spread increases the further into the future the model is being used to forecast, which should mean that the precision can be further decreased over time without affecting the accuracy. In light of this, the most useful forms of inexact hardware will be able to reduce the precision more severely for smaller-scale variables or after a certain forecast lead-time, in a 'scale selective' fashion (Düben et al. 2014b).

There are several ways of reducing precision in a computer model that have the potential to improve the efficiency. Four that have shown sufficient promise to have been subject to considerable research in recent years are 'Probabilistic Pruning', 'Stochastic Processors', 'Mixed Precision' floating point arithmetic and 'fixed-point' arithmetic. This section will summarise each

of these methods. For weather and climate forecasts, the methods most likely to be suitable and for which real supercomputer hardware exists or may soon be developed are Stochastic Processors and Mixed Precision arithmetic, and it is hence these methods that are emulated in the experiments presented in this thesis.

1.4.1. Probabilistic Pruning

'Probabilistic pruning' was put forward by Lingamneni and colleagues in 2011 as a way of improving the efficiency of hardware by physically deleting or 'pruning' parts of the circuits that have a low probability of being used in any given operation (Lingamneni et al. 2011). The aim is to obtain a circuit that uses the minimum number of components possible without the error introduced by deleting components crossing beyond the acceptable limit. Each logic gate or collection of logic gates in the circuit is assigned a significance value according to how important the bits to which it is connected are to the accuracy of the circuit's output. These components are ranked, and the least important ones are pruned away (Düben et al. 2014a).

Experiments on parallel processors demonstrated a halving of the computational cost with very small errors introduced, but the potential for savings in serial processors (where there are fewer alternative paths to prune) is more limited (Lingamneni et al. 2011). In general, the fractional cost savings are proportional to the fraction of the circuit that is pruned, independent of the type of circuit being used, although the corresponding error in a computer model being run on the pruned hardware will depend upon how the model is coded. A recent study that emulated the effect of pruned hardware on the idealised, Lorenz '96 model (which is described in Chapter 2) suggested that it could yield considerable energy savings of up to two thirds if applied appropriately (Düben et al. 2014a). Pruning can be applied in combination with other inexact hardware techniques, but whilst the initial results using emulators are promising, pruned hardware suitable for HPC applications would be difficult to manufacture and its implementation remains a long way off.

1.4.2. Stochastic Processors

The most basic method of reducing precision is to decrease the voltage supplied to computer circuits, an approach known as 'voltage over-scaling'. Conventionally, a high enough voltage is supplied to ensure that all calculations proceed successfully and accurately. Using over-scaling, the energy costs of the computations are decreased but at the risk that some of the calculations will yield inaccurate results or fail entirely. Essentially, the decreased power means that some of the bits of information used to represent numbers in the calculations will spontaneously 'flip' with some non-zero probability p, and the accuracy of the values will therefore decrease in proportion to p. Exactly where and when such bit-flips will occur is unpredictable, so processors that use over-scaling are referred to as 'Stochastic Processors'. For these to produce useful results, the voltage must remain high enough for the calculations not to fail entirely (Düben et al. 2014b).

In a recent study by Düben and colleagues, a Stochastic Processor was emulated by allowing one 'bit' of information to flip with a probability p at every calculation carried out outside the initialisation, output and testing parts of the code for idealised models of the atmosphere. The position of the bit flip was decided randomly, so that at some times the changed bit would be more significant to the output of the calculation than at other times. The degradation in the accuracy of the models was small for small fault rates ($p = 0.01$) but became very significant when the fault rate was increased (to $p = 0.05$ or 0.1), suggesting that the capacity for savings in computational cost by this method is limited. The authors noted that much more research into the effects of high fault rates on different parts of an atmospheric model is needed before this method can be widely used in real forecasting models (Düben et al. 2014a). Indeed, because Stochastic Processors can potentially access any of the bits used to represent a number, regardless of their significance, they can easily cause imbalances that strongly degrade the model performance (Düben and Palmer 2014). It may be necessary to add more complicated design features to computer architecture that can compensate for some of the errors introduced before this inexact method can be of any benefit, such as those described in a recent paper by Sartori and Kumar (2011).

1.4.3. Mixed Precision

A third method for inexact computing is that of Mixed Precision floating point arithmetic. Instead of allowing bit-flips to occur randomly at any point within the bit-wise representation of a floating point number, this method involves truncating that representation to a specified number of bits. For example, a 64 bit 'double-precision' number might be represented instead with 32 bits ('single-precision') or 16 bits ('half-precision'), which would decrease the number of decimal places and the range of values that the number could take. To optimise the accuracy and efficiency of a model, different parts of the model and spatial scales of the variables can be truncated to different degrees, an approach sometimes termed 'flexible' precision.

The advantages of this method are that the reduction in precision is much more controllable than that obtained with Stochastic Processors, since the position of the 'faulty' bits is decided by the manufacturers and users rather than left to chance, and that computational savings are possible using existing hardware and without much modification of existing model code. It is true that either new code designed for use with existing Field Programmable Gate Array (FPGA) hardware (see Section 1.5.5), or new computer architectures capable of delivering Mixed Precision that existing models can take advantage of would have to be adopted to gain the maximum possible benefit in terms of computational speed. These possibilities will be considered in more detail in Section 1.5.5. But even using forecast centres' existing computer CPUs, Mixed Precision could allow considerable savings in terms of memory (since each number will require fewer bits to store), which is often the key bottleneck on model performance, to be made (Düben and Palmer 2014).

Although hardware optimised for Mixed Precision is not yet available, benchmark tests using emulated hardware (Düben and Palmer 2014) or FPGAs (Russell et al. 2015) to run idealised models suggest that the computational cost savings using Mixed Precision could be considerable. Indeed, Mixed Precision produced significantly more accurate results than stochastic processing in the former study, which found that most of the dynamical core of a GCM could operate using very heavily reduced precision with negligible loss of accuracy. This might allow the memory required

to store floating point numbers to be reduced so drastically that processing units could be replaced by look-up tables. For all these reasons, it is considered to be the most promising method of quickly and effectively implementing inexact computing (Düben and Palmer 2014) to improve computational efficiency.

1.4.4. Fixed-point Arithmetic

Computer models almost always use 'floating-point' arithmetic, which means that the decimal point is allowed to 'float' relative to the significant digits of the number. For example, the digits 1, 2 and 3 can be used to represent the numbers 123, 12.3, 1.23, 0.123 and so on in the same model without restriction. Most computer processors, however, are not capable of carrying out floating-point arithmetic. Instead, they carry out 'fixed-point' calculations, which means that all numbers are represented as integers that are scaled by a fixed amount, for instance the integer 123 in the example above. Special algorithms are then used to emulate the floating-point calculations. These algorithms carry a computational cost, so it is less costly to instead use the explicitly supported fixed-point arithmetic in the model itself, avoiding the need to emulate floating-points. However, this reduces both the range and the numerical precision of the variables in the model, potentially reducing its accuracy (Oberstar 2007). This is because a fixed-point number can only be an integer multiple of a pre-determined factor, and the number of possible such integers depends on the number of bits available. For example, if the pre-factor is 1000, the representable numbers would be 1000, 2000, and so on up to a maximum of $n \times 1000$ where n is fixed by the total number of bits allocated to the number. In this case, numbers less than 1000 would not be representable at all. To increase the resolution of available numbers one could reduce the pre-factor to, for example, 100, but this would reduce the range to numbers between 100 and $n \times 100$.

Fixed-point arithmetic is therefore likely to be very efficient in cases where there is a small dynamic range. For instance, if some variable is only observable to one decimal place and always lies between 10.0 and 20.0, fixed-point arithmetic could be used to represent it with a pre-factor of

0.1 and just enough bits to represent integers between 1 and 100. For many variables in weather and climate models, including the idealised systems investigated here, this is not the case and the dynamic range is too large for all the possible values to be represented in fixed-point without critical significant figures being lost. For this reason, although fixed-point arithmetic may be useful in specific parts of forecast models in specific circumstances, it would be difficult to implement successfully and was not investigated further in this study.

1.5. Implementing Reduced Precision

The purpose of this thesis is to investigate whether the promise of better forecasts using scale-dependent inexact techniques is well-founded. To do this, it is necessary to carry out tests of weather and climate forecasting ability for models of comparable computational cost run with and without inexact hardware. However, the hardware required for Stochastic Processors or for flexible-precision (beyond double- single- and half-precision) arithmetic does not yet exist, and is only likely to be developed if the benefits of inexact computing can be demonstrated. Therefore, in this study inexact hardware had to be emulated using conventional hardware. This section explores how this can be done for Mixed Precision and Stochastic Processors in turn, then discusses the state of development of real hardware capable of implementing these techniques and the potential savings in computational cost such hardware could provide at a given resolution.

1.5.1. What are Double-, Single- and Half-Precision Floating Point Arithmetic?

Computers represent numbers using binary ‘bits’ of information, each of which can take either of the values 0 and 1, and use these to carry out a form of integer arithmetic. In binary code, each bit is associated with a power of 2 (the first bit is 2^0, the second 2^1 and so on), which is either active or inactive depending on whether the bit is in state 0 or 1. When a bit is set to 1, the corresponding, active power of 2 is included in the number; if the bit is 0, it is not. So, the number 1 can be

represented by setting the first bit to 1 and all the rest to 0, since this is just $2^0 + 0 = 1$. The number 3 is represented by setting the first two bits in the sequence to 1, yielding $2^0 + 2^1 + 0 = 3$. The largest number representable in this way depends upon the number of bits in the sequence used to represent it.

The binary code used in computers to represent 'floating point' numbers in calculations is more complicated than this. Following the standards set by the Institute of Electrical and Electronics Engineers (IEEE 2008), each number is split into three parts – the sign, the exponent and the significand – each of which takes a share of the total number of bits. The precision with which a number can be represented and stored depends upon the number of bits the computer hardware gives to each.

Most computers are built to carry out floating point arithmetic in 'double-precision'. Here, each number is represented by 64 bits, of which 1 bit represents the sign of the number (+1 or -1), 11 bits represent the exponent E (the highest power of two that the number is greater than) and 52 bits represent the significand S (the number's exact multiple, somewhere between 1 and 2, of this power). The overall number N is given by,

$$N = \pm S \cdot 2^E. \tag{1.6}$$

Where the value of each bit in the significand, b_i, or the exponent, b_j, is either 0 or 1, the significand and exponent are given by,

$$S = 1 + \sum_{i=0}^{51} b_i 2^{-i}, \tag{1.7}$$

$$E = \sum_{j=0}^{10} b_j 2^j - 1023. \tag{1.8}$$

This means that the significand is a fraction between 1 and 2. Note that the 1 in Equation 1.7 is sometimes referred to as an 'implicit' bit, in which sense the number has a 53-bit significand, but the extra bit is always 1, never 0, so is not stored explicitly. The number of terms in the significand and exponent sums is equal to the number of bits in the significand and exponent respectively, with each bit corresponding to a power of 2 beginning at 2^0. The largest resolvable number is obtained by setting all the bits in the significand part to 1 and all except the first bit in the exponent part to 1, which gives $E_{\max} = (0 + 2^1 + \cdots + 2^{10}) - 1023 = 1023$ and $N_{\max} = \pm 1.99\ldots \times 2^{1023}$. The

smallest resolvable number is obtained by setting all the bits in the significand to 0 and all except the first bit in the exponent to 0, giving $E_{\min} = 2^0 - 1023 = -1022$ and $N_{\min} = \pm 1 \times 2^{-1022}$. All bits in the exponent being set to 0 gives zero; all being set to 1 is defined as infinity if the significand bits are all zero, or 'Not a Number' (NaN) if some are non-zero.

All 'normal' numbers representable in double-precision therefore lie between 2^{-1022} and 2^{1023}; any number outside this range will be rounded to zero or infinity. The spacing between resolvable numbers increases as the numbers get larger because of the finite number of bits in the significand: the spacing for numbers between 2^E and 2^{E+1} is 2^{E-52} for a 52-bit significand (Regan 2015). Hence, the largest possible fractional rounding error for numbers that aren't explicitly resolvable (sometimes called the 'machine epsilon') must be half the fractional spacing, so for a number of any size 2^E, the maximum fractional rounding error is $2^{E-52}/(2 \times 2^E) = 2^{-53}$. This error applies to numbers that lie exactly half way between two resolvable values. If the hardware is configured correctly it is also possible to represent 'subnormal' numbers as small as $2^{-1022-52} = 2^{-1074}$ by taking some of the 52 bits used for the significand and using them for the exponent instead. Values at an equal spacing of 2^{-1074} between 2^{-1022} and 0 can then be represented.

Precision Level	Total bits	Sign bits	Exponent bits	Significand bits	'Normal' spacing	'Normal' minimum	'Subnormal' minimum	Maximum
Double-Precision	64	1	11	52	2^{E-52}	2^{-1022}	2^{-1074}	2^{1023}
Single-Precision	32	1	8	23	2^{E-23}	2^{-126}	2^{-149}	2^{127}
Half-Precision	16	1	5	10	2^{E-10}	2^{-14}	2^{-24}	2^{15}

Table 1.1: The number of bits in total and in each of the sign, exponent and significand used to represent a number in each of double-, single- and half-precision floating point arithmetic according to IEEE standards (IEEE 2008) are listed. Also given are the smallest representable 'normal' and 'subnormal' numbers, the largest representable number and the spacing between numbers of order 2^E and 2^{E+1} for exponent E.

Double-precision is generally favoured in High Performance Computing (HPC) applications because it minimises the risk of round-off error in calculations and it allows one to work with very large or very small numbers that require an 11-bit exponent to represent. For some

physics applications where there is a precise deterministic solution that is very sensitive to numerical errors, such as calculating planetary orbits and simulating supernovae, carrying out some computations in double-precision or even higher precision may be necessary (Bailey and Borwein 2009). Until recently, it was assumed that weather and climate models should be run entirely in double-precision because of those few components within the code that genuinely require it, such as land surface schemes that involve very small temperature changes (Harvey and Verseghy 2016) and the setting up of Legendre Transformations, which requires precision-sensitive recurrence relations for the computation of Legendre polynomials (Vana et al. 2017).

However, because of the chaotic nature of the atmosphere and the uncertainties inherent in weather and climate models, it is unlikely that close to double-precision is strictly necessary except in a few specific parts of the code. Calculations can potentially be performed faster, and numbers will take up less memory, if each number is represented in less than double-precision – what we might call 'reduced precision'. In principle, one may choose any number of bits for the significand and the exponent, so long as one bit is retained for the sign part. Reducing the number of bits in the exponent reduces the range of resolvable numbers, whereas reducing the number of bits in the significand increases the spacing between them (hence reducing the 'precision'). There are existing IEEE Mixed Precision formats, 'single-precision' and 'half-precision', which represent each number with 32 or 16 bits respectively, with these bits split between component parts as shown in Table 1.1. Equations 1.6-1.8 still hold for these, so long as the upper limits on the summations are changed to represent the appropriate numbers of bits that are afforded to the significand and exponent in each case. The spacing between normal (and subnormal) numbers and the lowest resolvable subnormal number also change, as Table 1.1 shows. So long as the appropriate parts are protected, precision in many atmospheric models can be reduced significantly without penalty. For example, Dawson et al. (2017) showed that so long as the storage of state variables takes place in high precision, a land surface scheme can be run with a 16 bit significand without penalty.

1.5.2. The Fortran Language

Most operational forecast models are written in the Fortran computer language, which is favoured because of its relative simplicity and resultant speed; it is designed explicitly for the purpose of translating scientific formulae into code for numerical solution. The commonly used 1990 release, Fortran 90, is the language in which all the experiments presented in this thesis were carried out. A Fortran program must be 'compiled' into an executable file; the compilers used here are the GNU Compiler 'gFortran' (Free Software Foundation Inc. 2016) version 4.9.2 and, when emulating Stochastic Processors, the Intel compiler 'ifort' (Intel 2016), version 14.0.

Fortran compilers require all variables to be named and categorised into 'types' at the start of the program. Types that the language natively supports include integers and floating point real numbers in either quad- (128 bit), double- or single-precision. It also includes operations (addition, multiplication and so on) and inbuilt functions that act on each variable according to its type. It is possible to define new 'types' with their own sets of rules for how these operations and functions should treat them. A 'reduced precision' type, for example, can be created which is treated in all operations as though it were resolvable only at some specified level of precision (see below).

1.5.3. Emulating Mixed Precision on Conventional Hardware

Carrying out computations with numbers represented in single- or half-precision instead of double-precision reduces their cost, allowing the resolution of a forecast to be increased, for example, but also may decrease their accuracy because of the truncation of the significand. The computations may not work at all – the computer model crashes – if the exponent is truncated so heavily that some of the numbers lie outside the representable range. Because of these conflicting factors, it is by no means clear whether reducing the precision will increase or decrease the accuracy of a computer model overall. Therefore, it is important to be able to emulate different levels of reduced precision for different segments of code within a model so that any effect on the accuracy can be

explicitly analysed using conventional hardware and a variety of configurations can be tested without having to construct new hardware each time. The results of these tests can then be used to build reduced precision hardware that is optimised from the outset to maximise the accuracy of the computer model that it is designed to run.

One of the key advantages of Mixed Precision hardware over other methods of improving the efficiency of computer models is that it does not require the model code to be altered substantially: all that needs to be added are instructions telling each part of the code which precision hardware to use during calculations. The same advantage applies when this hardware is being emulated. To emulate Mixed Precision hardware, the main program treats variables as special Fortran 'types' that are defined by a separate, emulator program configured alongside the main program. Whenever these variables are used in computations, the main program calls the emulator program and passes it the variables' double-precision values and a 'flag' that indicates how many bits should be allocated to the significand and exponent. The emulator program then rounds the values accordingly, carries out the required calculation and returns what the 'reduced precision' output should be in the form of a rounded double-precision floating point number. In this way, the main program always runs using double-precision floating point numbers and double-precision hardware throughout, but its output is the same as the output that would be obtained using Mixed Precision hardware. If the emulator encounters values that lie outside the specified range of the exponent, the calculation will yield 'Not a Number' (NaN) and the main program will fail.

The process of passing numbers out to a subsidiary program, rounding them before the calculation and passing them back to the main program is computationally expensive, which is why emulated Mixed Precision computing is actually slower than standard double-precision arithmetic. Since the output of the main program is in double-precision, there is also no saving in memory when using an emulator. Real Mixed Precision hardware would not go through this process of rounding numbers and would therefore entail no such costs. Hence, the emulator cannot be used to analyse the potential computational cost savings that such hardware would yield, only the effect that it might have on the accuracy of the model being run.

The Mixed Precision emulator used here rounds numbers according to a 'round to nearest' principle that differs slightly from the 'round to nearest, tie to even' standard used by IEEE. The difference only affects numbers for which the most significant bit to be rounded away is a 1 and all the bits less significant than this are 0, in which case the IEEE scheme would require that the result of the rounding retained 0 as its least-significant bit, but the emulator used here does not. This small difference is not expected to have any discernible effect on the precision required for accurate simulations.

1.5.4. Emulating Stochastic Processors on Conventional Hardware

Although the two techniques have very different requirements to be implemented in practice, emulating a Stochastic Processor is done in a similar way to emulating Mixed Precision hardware. Again, a 'flag' is passed from the main computer program to the emulator program telling it whether to perform a given calculation inexactly or not. This time, though, the user can choose not the level of bit-truncation but instead the probability of a bit 'flip' within the emulator. A different flag can be used for different scales or types of variables so that a different probability p of bit-flipping can be applied to each. Alongside the 'flag' is passed the number to be altered. If $p = 0$, the emulator has no effect and the output equals the input. If $p = 1$, there will definitely be a bit-flip in one of the sixty-four bits that make up the number in double-precision. Ordinarily, $0 < p < 1$ and a random number r_1 is generated between 0 and 1 each time the emulator is called. If $r_1 < p$, the emulator randomly flips one of the bits, changing the number by a small or large amount depending on which bit is flipped.

Changing the sign or the exponent of the number would of course change its value very significantly, and even a few changes of this type could be disastrous for the model being run, potentially causing a crash if the calculated value moves far outside the acceptable range. Therefore, in the emulator to be used here it is assumed that the exponent and sign bits can be protected from bit flips. For a double-precision floating point number, there are 52 bits in the

significand. If a bit flip occurs (that is, $r_1 < p$), a second random number r_2 between 0 and 1 is generated, and the multiple of $1/52$ closest to this number is identified as the index of the bit to be flipped. Very low numbers will have a negligible effect on the output because they correspond to very small powers of 2. Flipping bit 0, for instance, will change the output number by a factor of 2^{-52}, whereas flipping bit 52 changes the output by a factor of $2^{-2} = 0.5$, which is substantial, although the chance of this occurring for any given number passed to the emulator is only $p/52$. For very small p, therefore, the model accuracy should not be affected by the Stochastic Processor; for very large p the effect could be very significant. The emulator must be tried with multiple values of p to determine the optimal choice of p that maximises the efficiency.

1.5.5. Inexact Hardware

The form of inexact hardware perhaps closest to realisation is what will here be referred to 'Mixed Precision' hardware, capable of allowing selected parts of a given model to be implemented in bit-truncated reduced precision. Whilst Stochastic Processors and the hardware required to carry out probabilistic pruning (see Section 1.4) are still in the early stages of development, Mixed Precision hardware is available now, in the form of Field Programmable Gate Arrays (FPGAs). These highly customisable circuits can be used to reduce the number of bits used to represent floating-point numbers in certain parts of an atmospheric model and for variables pertaining to specific spatial scales according to user specifications. However, just as for GPUs and other forms of parallel processing, the model code has to be completely rewritten in order to be used on an FPGA (Düben and Palmer 2014), which would require considerable effort in the case of a full-scale global model.

Unlike standard CPU and GPU hardware, which support only double- and single- (and sometimes half-) precision, FPGAs allow the precision to be reduced to almost any specified level. When the idealised Lorenz '96 model, which supports two scales of variables, was coded for a real FPGA setup, Peter Düben and colleagues found that reducing the precision in the significands of the large- and small-scale variables to only 12 bits or 9 bits, depending on the set parameters in the

model, yielded a more accurate output than parameterising the small-scale variables in long-term forecasts, for example. It was also found that beginning the model in single-precision and reducing the precision further later on had no impact on the accuracy because the uncertainty arising from the stochasticity of the system increases with the forecast lead-time. Such tests using FPGAs demonstrate the potential value of Mixed Precision hardware, but FPGAs are only one possible form of such hardware and it is unlikely that full-scale GCMs could be re-written to run on FPGAs.

Therefore it is important that, in order to obtain the advantages in speed as well as memory that Mixed Precision has the potential to provide, new hardware is developed. This hardware must be capable of running each part of the model for which it is designed at the optimal level of precision to achieve the maximum overall accuracy. Since this optimal level is obtained by a trade-off between resolution (achieved through the reduced computational cost that comes with reduced precision) and precision, it will probably be different for different variables and will vary across different spatial scales and different parts of the model. Some parts – perhaps including the core time-stepping scheme – will need to be kept entirely in double-precision, whilst others may benefit from a reduction in precision only after a certain optimal amount of time has passed.

Hence, Mixed Precision hardware will need to contain a number of components, each operating at a particular level of precision, which can be called upon to carry out a sub-set of calculations. This may at first seem a difficult task, but hardware that operates at single- or half-precision already exists: standard CPUs can usually run in single-precision, and the operational Integrated Forecasting System (IFS) model of the European Centre for Medium-range Weather Forecasts (ECMWF) has already been successfully run almost entirely in single-precision in a study that found that this could save up to 40% of computational costs (Vana et al. 2017). In April 2016, NVidia launched its ‘Pascal’ GPU architecture, which is capable of carrying out operations at double-, single- and half-precision, with a proportional increase in maximum processor speed when precision is reduced (Walton 2016). Machine learning applications increasingly require low-precision architectures, which may create a demand for CPUs capable of running at half-precision or lower; NVidia’s ‘Volta Tensor Core’, designed for ‘deep learning’ applications, can run in Mixed Precision, with calculations carried out in half-precision up to ten times faster than they can

be in double-precision (NVidia 2018). Creating such hardware should hence be technically straightforward; greater difficulty may lie in coordinating the hardware to apply the appropriate precision in the appropriate place, but the FPGA example shows that this is not beyond current capabilities. Certainly, Mixed Precision hardware should be more straightforward to engineer than probabilistic pruning hardware or Stochastic Processors, a clear design architecture for which is yet to be agreed upon (Düben et al. 2014b).

But there may also be the potential to integrate both Stochastic Processors and Mixed Precision architectures together, yielding perhaps greater efficiency models than either technique could provide in isolation. One could imagine, for instance, implementing stochastic parts of a model such as stochastic parameterisation schemes using Stochastic Processors, obviating the need to call a random number generator by using the errors in the hardware as a natural source of noise. Because this stochasticity is itself illustrative of uncertainty about the physical process being parameterised, one might not expect the accuracy to be degraded by simultaneously reducing the number of bits used to represent these numbers.

The computational cost savings that might be achievable with real reduced-precision hardware would depend upon how much of an operational forecast model could be reduced in precision and to what extent, and upon the computational 'overhead' associated with this. For instance, there might be some cost involved in instructing the program where and how to reduce precision. Given the scarcity of real reduced precision hardware to test, it is difficult to estimate this, and the estimates made in this thesis will therefore assume the theoretical maximum benefit obtainable with reduced precision, ignoring such overheads. This approach may seem unjustifiably optimistic, but an automated implementation of probabilistic pruning and reduced precision explored by Lingamneni et al. (2012) suggests that near-zero overhead is not an unrealistic possibility for some types of hypothetical hardware.

1.6. Previous Experiments with Reduced Precision

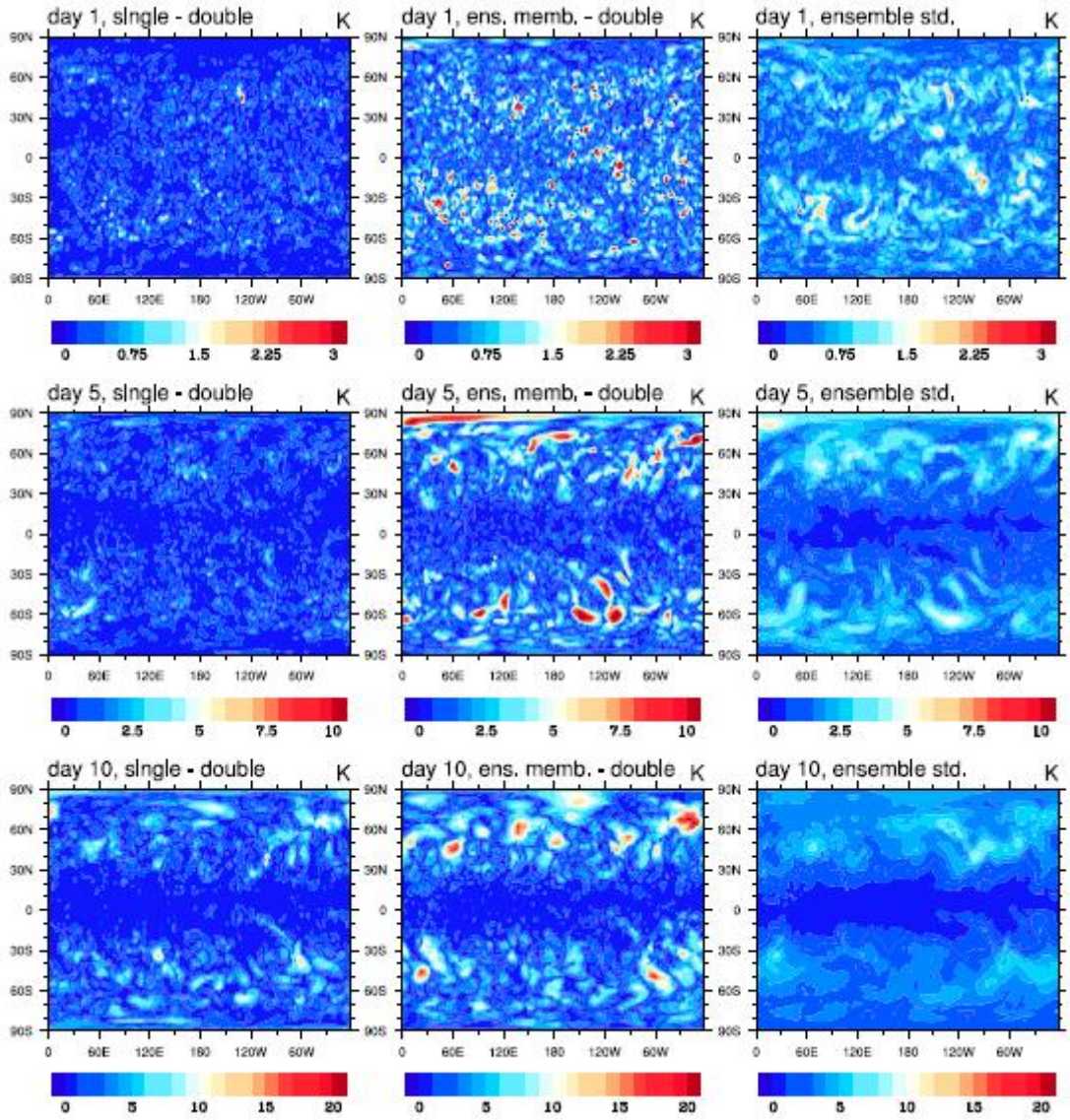

Figure 1.1: Plots reproduced from Düben and Palmer (2014; their Figure 14) to demonstrate the potential for single-precision to be used in the ECMWF Open IFS model without loss of forecast accuracy. They plot the absolute difference in temperature at 850 hPa between a double-precision control run and both a single-precision simulation (left) and one member of the ECMWF's operational ensemble forecast (middle) alongside the ensemble standard deviation (right), after 1, 5 or 10 days from the start of the simulations.

There is a growing body of literature describing experiments that aim to assess the potential for inexact techniques to be applied in the context of weather and climate forecasting. These studies have begun to reduce the uncertainty regarding exactly which parts of weather and climate models

could beneficially be performed in reduced precision, and what the optimal degree of precision reduction is for each. Because this task remains far from accomplished, and because, in the absence of real inexact hardware on a scale applicable to real-world GCMs, it has been necessary to emulate such hardware on conventional processors, it remains very difficult to estimate from these studies the potential savings in computational cost – or, equivalently, the potential gain in accuracy – that could be brought about using Mixed Precision hardware to run such models. But studies have already provided tantalising clues that the benefits of inexact techniques could be substantial.

Using Graphical Processing Units (GPUs) to take advantage of parallel processing techniques, Michalakes and Vachharajani (2008) demonstrated a 1.23-times speed-up relative to standard CPUs on running the Weather Research and Forecasting (WRF) model, and up to seven-times speed-up was obtained by running physical parameterisation schemes on GPUs by Lapillonne and Fuhrer in the Consortium for Small Scale Modelling forecast model (2014). Although these experiments were done in double-precision, GPUs already exist in forms capable of executing single-, double- and half-precision, which would further increase the efficiency.

Indeed, there are promising results that suggest the possibility of single-precision operational forecasts. Figure 1.1 is a reproduction of Figure 14 in Düben & Palmer's 'benchmark' study in which the ECMWF's Open Integrated Forecasting System (Open IFS) global atmospheric model was run in single-precision. The figure demonstrates that the difference between the single- and double-precision control runs was smaller than that between the double-precision control and one of the ensemble members used to produce an ensemble forecast for temperature at 850 hPa for 1, 5 and 10 days' lead-time at 32 km resolution (a comparison was also carried out between forecasts of the geopotential height at 500 hPa and yielded similar results). That the single-precision predictions fell within the envelope of the ensemble forecast implies that running the ECMWF model in single-precision has a negligible effect on the forecast accuracy, although the authors note that this result may not hold true in other flow regimes (Düben and Palmer 2014).

Tests are yet to be carried out to see whether precision might be reduced even further in certain parts of the Open IFS model, something which this study aims to address. However, Düben and Palmer (2014) were able to apply reduced precision to specific parts of the simpler

Intermediate Global Climate Model (IGCM): parts of the model that accounted for 45 and 53 per cent of the overall computational cost were run using an emulator with numbers truncated to have 6 and 8 bits in the significand respectively (considerably less than half-precision), and the forecast error in the geopotential height was smaller than that for a control run at coarser resolution. The authors conclude that 98% of this model could be run with an 8 bit significand and yield better results than the lower resolution double-precision alternative with equivalent computational costs, and that the entire dynamical core of the model could be run with 15 bits in the significands of all floating point numbers with negligible effect on the accuracy. The authors project that further savings could be achieved, even given computational overheads, using heavily reduced precision at high wavenumbers in high-resolution spectral models, suggesting that parts of a spectral model dealing with small-scale phenomena could be represented in 8 bits with no degradation in accuracy.

However, most models – and most of the computations in the ECMWF model that Düben and Palmer (2014) investigated – carry out calculations in grid-point space, where the benefits are less profound because precision must be reduced uniformly across all spatial scales. One advantage of having the same precision for all variables is that the computational overhead would likely be small, although the model might still need to be instructed to protect more sensitive parts of the code such as time-stepping schemes, but the potential advantages of reduced precision are limited because it is not possible in grid-point space to apply different precision levels to small-scale and large-scale dynamics without considerably altering the models to utilise a complicated mixture of coarser and finer grids simultaneously, a potentially interesting avenue of future research but one which is difficult to put into practice operationally at present. Also, Düben and Palmer did not investigate the effect that simultaneously truncating the exponent parts of the numbers would have.

Francis Russell and colleagues addressed both of these limitations in their study (Russell et al. 2015), in which the Lorenz ’96 idealised model atmosphere (see section 2) was run on FPGA hardware. The exponents and significands of the large- and small-scale variables were truncated separately and the Hellinger distance was used as a measure of the long-term forecast accuracy of the reduced precision models. Models with the small-scale variables represented using close to single- or half-precision performed similarly well to a double-precision model in which the initial

conditions were changed. Because the study used real reduced precision hardware, rather than an emulator, the reduction in computational cost could also be assessed, and was found to be substantial. For example, a model with slightly more than half-precision in the large-scale variables and slightly less than half-precision in the small-scale variables (labelled ‘DFE4’ in that paper) had a speed-up factor of 72.7 over a single-core CPU in double-precision. The Russell et al. (2015) investigation was only carried out in an idealised atmosphere, however, so no light could be shed upon the extent to which such speed-ups could be realised in a full-complexity system.

Previous studies have largely focussed on how far different parts of idealised or real-world models can be processed inexactly without a considerable impact on performance. Some have explored the extent to which restricting lower precision to only smaller spatial scales could be a way of retaining high forecast skill with a lower cost, but no systematic study of scale-selective precision in a real-world system has been made. This thesis will build on the insights obtained in previous investigations but will move beyond them in two key respects. First, it will explicitly address whether reducing precision can lead to more accurate forecasts through the potential to re-invest computational cost savings in higher-resolution models. This will be the focus of the investigation using the extended Lorenz ’96 system described in Chapter 2. Second, it will apply scale-selective reduced precision systematically to the spectral components of systems that are closer to the real world, namely the SQG system and Open IFS introduced in detail in Chapters 3 and 5 respectively.

1.7. Evaluating the Performance of Forecasts

In order to test the effect of reducing the precision on a model’s capacity to forecast the weather and climate accurately, it is necessary to develop robust methods of assessing different models’ forecasting ability. Because Earth’s atmosphere is chaotic, the ability of a model to forecast the ‘true’ state of the weather at any given time will degrade to zero after a few days. However, according to the notion of ‘seamless’ weather and climate forecasts, a model that produces accurate

weather forecasts will also be able to successfully predict the climate – that is, the average behaviour of the atmosphere – for many weeks, months or years ahead. In practice, the relative quality of weather and climate forecasts may be different for different models, and different inexact techniques may affect one or the other more strongly. Thus, it is important that the accuracy of both kinds of forecasts is assessed for the models in this study, all of which are chaotic systems replicating key aspects of the real atmosphere's physical properties. Several methods of assessing model skill will be presented in this section; all of these can be affected by random variability or changes to the mean conditions, meaning that it is beneficial to average these scores over multiple runs of a model when assessing its accuracy.

1.7.1. Measures of Short-term Forecast Skill

The most straightforward measure of a weather forecast's ability to match the 'true', observed state of the atmosphere – which is generally taken to mean the forecast's reliability and resolution (Wilks 2006) – is the Brier Score (BS), which is applicable when there are two possible outcomes, such as 'rain' or 'no rain'. If J is the number of possible outcomes, such situations can be said to have $J = 2$. If an ensemble forecast is made with n observation points, and for the kth point such an outcome (for example, that it will rain) is forecast with probability y_k, for $k = 1, \dots, n$, the Brier Score measures the mean square error between these predicted probabilities y_k and the corresponding proportions of occasions on which the event is observed to occur, which is given the symbol o_k, according to (Wilks 2006),

$$\mathrm{BS} = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2. \tag{1.9}$$

This can be converted into a Brier Skill Score (BSS) that provides a more meaningful comparison between different forecasts by invoking a 'reference' forecast with Brier Score $\mathrm{BS}_{\mathrm{ref}}$, where,

$$\mathrm{BSS} = 1 - \mathrm{BS}/\mathrm{BS}_{\mathrm{ref}}. \tag{1.10}$$

For a real weather forecast, the 'reference' would be a forecast based upon climatological probabilities for each forecast category (for example, the climatological probability that it will rain or not rain). In an idealised system it takes the form of an analytical probability density function (pdf) describing what the probabilities ought to be in theory, according to the mathematical properties of the system: for example, the idealised system may be such that all the categories are of equal likelihood, in which case the reference probability function will be a flat line.

For forecasts with $J > 2$ possible ordinal outcomes with a clear order of outcome categories, such as the total accumulated rain in 1mm bins, each outcome can be given an index j, and the Ranked Probability Score (RPS) is more appropriate than the BS. Here, for each forecast category j, at each observation point the cumulative probabilities $Y_j = \sum_{l=1}^{j} y_l$ and $O_j = \sum_{l=1}^{j} o_l$ are defined before the mean square error is computed, yielding,

$$\mathrm{RPS} = \sum_{j=1}^{J} \left(Y_j - O_j\right)^2. \tag{1.11}$$

This is then averaged over the n observation points. The corresponding Ranked Probability Skill Score (RPSS) is,

$$\mathrm{RPSS} = 1 - \langle \mathrm{RPS} \rangle / \langle \mathrm{RPS} \rangle_{\mathrm{ref}}. \tag{1.12}$$

The reference $\langle \mathrm{RPS} \rangle_{\mathrm{ref}}$ comes from the system climatology. To obtain a smooth evolution of this score through time, it is usual to average the score over multiple runs of the ensemble forecast with independent initial conditions; the <> indicates this averaging. Because the RPSS uses the cumulative probabilities for each forecast category, it takes into account the predicted probability across all categories, not just the 'true' category. Therefore, it is referred to as a 'non-local' score (Arnold 2013). This score is especially useful because it is approximately proportional to the societal value of a real forecast. However, the probabilities y_k must be computed across an ensemble of forecasts with at least ten members sharing the same initial conditions to avoid a negative bias that otherwise implies that the skill of the model is worse than that of a forecast based on climatological probabilities (Kumar et al. 2001).

An alternative approach is to measure the amount of information or 'entropy' that a forecast contains, which is the motivation behind the Ignorance Score (IGN). Here, for n forecasts, the forecast probability that the kth point will have outcome i is $f(k)_i$, and if the actual observed outcome is $j(k)$ then (Roulston and Smith 2002),

$$\mathrm{IGN} = -\frac{1}{n}\sum_{k=1}^{n} \log_2 f(k)_{j(k)}. \tag{1.13}$$

Ignorance is a measure of the information deficit of someone possessing only the forecast model. The smaller the ignorance, the lower the entropy and the more useful (truthful) information the forecast contains. Using a reference, climatological Ignorance Score $\langle \mathrm{IGN} \rangle_{\mathrm{ref}}$, the Ignorance Skill Score can be defined as,

$$\mathrm{IGNSS} = 1 - \langle \mathrm{IGN} \rangle / \langle \mathrm{IGN} \rangle_{\mathrm{ref}}. \tag{1.14}$$

As for the other skill scores, it is usual to average the IGNSS over multiple forecast runs. Because the ignorance score depends only on the forecast probability $f(k)$ of the observed outcome, like the BSS it is a 'local' score that does not take into account the predicted shape of the pdf away from the 'truth' (Arnold 2013). This makes it more sensitive than non-local scores such as the RPSS, which uses the cumulative probability over all possible outcomes, to outliers that may happen to fall into the 'truth' category despite most of the ensemble's forecasts lying far away: the IGNSS penalises under-dispersive forecasts (those where all the ensemble members are in close agreement) much more heavily than over-dispersive ones (where there are many outliers).

All the above scores are proportional to the difference between the model prediction and the 'truth' observation, so the smaller the score, the better the forecast. They can each be evaluated at every timestep of the model independently. It is often useful to compare these skill scores against the absolute error in the forecast $E(t)$ at time t, which is given by,

$$E(t) = \langle |X_{\mathrm{model}}(t) - X_{\mathrm{truth}}(t)| \rangle, \tag{1.15}$$

where $X_{\mathrm{model}}(t)$ is value of the variable $X(t)$ to be forecast, $X_{\mathrm{truth}}(t)$ is the 'true' value of $X(t)$ at this time, and < > signifies an average over several ensemble members, model runs or both. A larger value of $E(t)$ indicates a less accurate result. This statistic is simply a measure of the

difference between the model and the truth, and it is difficult to gauge from this whether the model has performed well or poorly as compared to a climatological forecast. However, it does enable comparisons to be drawn between different models of a given atmospheric system, and is expected to be proportional to the skill scores for each, providing a useful check on their proficiency as indicators of the forecast accuracy. Another, similar metric that amplifies the effects of larger values and therefore enables a better detection of outlying datapoints is the Root Mean Square Error (RMSE). This is defined as,

$$\mathrm{RMSE}(t) = \sqrt{\langle \left(X_{\mathrm{model}}(t) - X_{\mathrm{truth}}(t) \right)^2 \rangle}. \tag{1.16}$$

1.7.2. Measures of Long-term Forecast Skill

The longer-term, 'climate' forecasting skill of a model describes its ability to match the 'true' mean characteristics of the atmosphere over an extended period of time. To assess this skill, we can crudely compare the mean and standard deviation of the predicted quantity $X(t)$ over an extended timeseries with the 'truth' defined from either a high-resolution run of an idealised atmosphere or from observations of the real atmosphere as appropriate. A more accurate measure of the climate forecast skill, though, is obtained by comparing the cumulative probability density function (pdf) of $X(t)$ predicted by the model, Ψ_{model}, with the 'true' cumulative pdf Ψ_{truth}. The pdf is a function describing the fraction of data-points falling into each of a set number of forecast categories (or 'bins'), and the cumulative pdf is defined for each category by summing the pdfs of all the previous bins. The Kolmogorov-Smirnov (KS) statistic, D_n, comprises the maximum difference between the truth and model cumulative pdfs (Wilks 2005),

$$D_n = \max(|\Psi_{\mathrm{truth}} - \Psi_{\mathrm{model}}|). \tag{1.17}$$

Hence, this not a truly localised statistic (it does not take into account the place and time at which the value of an X-variable falls into any particular category) but provides a measure of the maximum distance between the 'truth' and 'model' distributions. The statistic does not account for

the extent to which the distributions differ between the two datasets away from this maximum, so may be swayed by anomalies if there is not a sufficiently large number of data-points.

A measure of the climate forecast skill that avoids this problem is the Hellinger distance H, obtained by integrating the difference between the 'truth' and the model over all categories, rather than only considering the one for which the difference is maximum, so that (Arnold et al. 2013),

$$H^2 = \frac{1}{2}\int\left(\sqrt{\Psi_{\text{truth}}} - \sqrt{\Psi_{\text{model}}}\right)^2 dx. \tag{1.18}$$

This gives a better measure than the KS statistic of how consistently the two distributions disagree with one another across all bins, but tends to mask sharp disagreements that only occur in a very small part of the distribution but could be very relevant in a real forecast. For this reason, it is useful to use both statistics when analysing climate forecasts. For both, a better forecast will have a score closer to zero because this will indicate a smaller difference from the 'truth'.

1.7.3. Reproducing Regime Behaviour

Another test of the proficiency of a given model of an atmosphere is its ability to replicate the 'regimes' of behaviour that the atmosphere occupies. These regimes constitute different regions of the system's attractor – the set of states that the system can be in. Some chaotic systems possess 'strange attractors' that are shaped in such a way that allowable states are separated by impossible states in phase space, the most famous of these being the 'butterfly'-shaped attractor discovered by Edward Lorenz (1993). The state of the system can lie on either of the two 'wings' of the butterfly but not in the space in between; these two wings can be thought of as two regimes between which the system can suddenly switch. Earth's atmosphere indeed possesses such regimes, for example El Nino and La Nina phases of the El Nino Southern Oscillation (ENSO), although its size and complexity mean that such regimes can be difficult to identify.

A dynamical system can be said to exhibit regime behaviour if three criteria are met, established by Lorenz (2006a). First, the system must inhabit at least two distinct regions in phase space, which is to say two alternating values of some fundamental quantity such as the total energy.

Second, it must be capable of transitioning rapidly between these regions. Finally, the average length of time between transitions must be large compared to the oscillations that the system is undergoing. Only systems that are close to the transition point from periodic to chaotic behaviour exhibit regimes: too much chaos leads to rapid and brief switching of behaviour, whereby what Lorenz refers to as ‘events’ occur and one or more of the regimes are entered only very briefly and sporadically, whereas too little prevents switching between regimes altogether as the system is unable to explore new regions of phase space. A proficient model of a system, if run for a sufficiently long time to exceed the regime transition time, would be expected to reproduce the regimes of behaviour exhibited by the true system, and this hence provides a further test of the quality of models used for long-term forecasts.

2. Reducing Precision in an Idealised System 1: Modified Lorenz ’96

The aim of this chapter is to provide a preliminary demonstration of the potential for scale-dependent numerical precision to improve the accuracy of forecasts by saving computational costs. For this initial study, the complications associated with full-complexity models of the real-Earth atmosphere are avoided by introducing for the first time a modified version of an ‘idealised’ atmospheric system that was developed in its original form by Edward Lorenz in 1996 (described in Section 2.2). This new system comprises three rather than the original two spatial scales, and is used to compare the proficiency of reduced precision, high-resolution forecasts with that of low-resolution double-precision forecasts of a similar computational cost, yielding results that inform the reduced precision studies in more complicated systems described in subsequent chapters. In this chapter, the parameterisation schemes employed and the system variables chosen to carry out these experiments are outlined (Sections 2.3 and 2.4) and the results are explained in full (Section 2.5). It is shown that representing large scales in single-precision and medium scales in half-precision yields results that are much closer to the double-precision ‘truth’ than those obtainable with the medium scales parameterised.

2.1. Building an Idealised Atmosphere

This thesis aims to demonstrate the extent to which Mixed Precision computing could be used to improve the efficiency and accuracy of the global models that inform real-world weather and climate forecasts. But such complex models are computationally costly to run and to verify against real observations of Earth’s atmosphere, which limits the number and nature of Mixed Precision experiments that can be carried out. Therefore, it is useful to perform a wider variety of

experiments using lower-cost systems that model 'idealised atmospheres', the results of which will help to determine where best to focus in experiments on the more complex models.

An idealised atmosphere is not a representation of the real Earth atmosphere. Rather, it involves the dynamical evolution of a set of non-dimensional variables that evolve in space and time similarly to real atmospheric equivalents such as temperature or pressure. There are two key advantages to using a highly idealised system. First, because the boundaries of the system can be clearly defined, the dynamical equations that define the system are known with complete certainty and none of the variables are too small to be observed, it is possible to calculate the 'true' values of all the variables at any given time. The accuracy of models of the system that involve parameterising or reducing the precision in some of the variables can hence be determined by comparing their predictions directly against this 'truth'. Second, so long as the system is small enough, it is computationally cheap to determine both the 'true' state and the predicted state. This means that a large number of experiments can be carried out in a relatively short amount of time.

It is important that an idealised system resembles the real atmosphere in some critical respects if the results of these experiments are to be relevant to real-world forecasts. The system must be dynamical, so that there exist differential equations describing its behaviour. Unlike the real atmosphere, these equations will be exactly known and will explicitly control – not just describe – the evolution of the system, meaning that the idealised atmosphere will not contain any true randomness. However, the system must also be chaotic so that its behaviour appears random, resembling the real atmosphere in that introducing very slight changes in the initial conditions will cause it eventually to evolve into an entirely different state.

A system's chaoticity can be measured by evaluating the 'Error Doubling Time' (EDT) – that is, the time taken for the error introduced by a perturbation to the initial conditions to double, a key way of characterising an idealised atmosphere. This quantity is inversely proportional to the Lyapunov exponent, a positive value of which indicates that the system is chaotic (Lorenz 2006b), and is measured in the 'Model Time Units' (MTUs) with which computer models of the idealised system are run. In his 1996 paper (republished in 2006), Edward Lorenz put forward a way of measuring this doubling time for a system of K large-scale variables. The true initial state of the

system for the k^{th} variable is defined as X_{k0}, and a perturbed initial state X'_{k0} is created by adding to it a random number e_{k0} so that,

$$X'_{k0} = X_{k0} + e_{k0}. \tag{2.1}$$

When the system is run forward over N timesteps, the error between the observed and 'true' states after $n \leq N$ timesteps of size Δt have elapsed is $e_{kn} = X'_{kn} - X_{kn}$. The average error across all K variables at any time $\tau = n\Delta t$ is given by $e(\tau)$. This procedure is repeated for multiple 'runs' of the model, each with its own set of initial conditions X_{k0} and X'_{k0} set to be equal to the final conditions X_{kN} and X'_{kN} of the previous run. An average $E(\tau)$ across all M runs is obtained by averaging the logarithm of the square of this quantity, so that (Lorenz 2006b),

$$\log\big(E^2(\tau)\big) = \langle \log\big(e^2(\tau)\big) \rangle_M. \tag{2.2}$$

To compute the Error Doubling Time, $\log\big(E(\tau)\big)$ is plotted against τ. At small τ, an exponential growth curve is observed in the form of a straight diagonal line, the gradient of which is $m \approx \Delta \log(E) / \Delta\tau$. Because it takes time $\Delta\tau$ for the error to increase by a factor E, this gradient must be related to the doubling time t_d (that is, the time for the error to increase by a factor of 2) through,

$$t_d = \frac{\log(2)}{\log(E)} \Delta\tau = \frac{\log(2)}{m}. \tag{2.3}$$

The ability to calculate the EDT is very important when using an idealised system because it provides a link between this system and models of the real atmosphere by relating the value of an MTU to a specific number of real atmospheric days. If the chaoticity of the idealised system is to be the same as that of the real atmosphere, errors in the initial conditions must grow at the same rate in both systems. Therefore, if the most advanced GCMs have an Error Doubling Time of 1.5 days (Lorenz 2006b), a doubling time of, for example, 1.5 MTUs in the idealised system would imply that 1 MTU $=$ 1 day, whereas if the idealised system's doubling time was measured to be 0.3 MTUs, this would imply that 1 MTU $=$ 5 days.

Although it may at first seem counterintuitive, the Error Doubling Time of real atmospheric models has decreased, not increased, since the 1960s, meaning that the models have

become increasingly sensitive to the initial conditions. This is because chaotic effects are damped out at the smallest resolved scales in models by diffusion, and increasing the model resolution shifts this diffusive regime away from synoptic scales. Also, thunderstorms, for example, evolve much more rapidly than larger-scale phenomena such as the Quasi-Biennial Oscillation (QBO). Although both are chaotic features of the atmosphere, slight uncertainties in the initial conditions will make the thunderstorm's evolution difficult to predict even a few hours ahead, whereas the state of the QBO ordinarily follows a regular pattern and can hence be known relatively certainly many months in advance (Lorenz 1993), although, being part of a chaotic system, even the QBO confounded expectations by deviating from its regular behaviour in 2015 (Newman et al. 2016). For this reason, improving the model resolution may beyond a certain level start to yield less and less significant improvements to forecasts. A useful idealised atmosphere should replicate this increase in chaoticity when the resolution is increased.

2.2. The Lorenz '96 System

In 1996, Edward Lorenz introduced an idealised atmosphere in which there are two tiers (scales) of variables evolving under dynamical differential equations. This system has since been used repeatedly by model developers to test, for example, parameterisation schemes (Wilks 2005; Arnold et al. 2013) and data analysis methods (Ott et al. 2004) pertinent to real GCMs because of its ability to replicate the chaotic, multi-scale dynamics of Earth's atmosphere in a low complexity, one-dimensional space. Variables are discretised onto grid points in space and time just as for a real atmospheric model, and the number of variables can be altered according to complexity and computational affordability requirements. Hence, this system has already been used to study reduced precision techniques by other authors (Düben et al. 2014a; Russell et al. 2015, 2017). However, those studies suffered from the drawback that there is no way of increasing the resolution of the model when precision is reduced in the small-scale variables to assess whether the accuracy of a forecast improves. To do this, it is helpful to extend the system to include a third tier of

variables on an even smaller scale to create a new system; this three-tier 'Modified Lorenz '96' system was created and used for the first time in this investigation.

2.2.1. The Two-Tier System

Lorenz' original 1996 system comprises either one or two tiers of variables: K large-scale X-variables, labelled X_k for $k = 1, \dots, K$, each of which is (optionally) coupled to J small-scale Y-variables, labelled Y_{jk} for $j = 1, \dots, J$. The X-variables represent the values of some hypothetical atmospheric quantity at each point around a one-dimensional latitude circle, and each is cyclically coupled to its neighbours so that $X_{k+K} = X_k$. The Y-variables represent the values of this quantity at a smaller scale, and although each is coupled to a single X-variable they also form a higher-resolution latitude circle so that the small-scale variable at the edge of the region covered by the kth large-scale variable X_k couples to the first small-scale variable belonging to the next region, covered by the $(k+1)$th variable X_{k+1}, which means that $Y_{j+J,k} = Y_{j,k+1}$.

Two differential equations govern the time evolution of the two sets of variables; these equations are not derived by approximating the behaviour of the real, continuous atmosphere, but are designed to give a simple representation of coupled variables that evolve chaotically over time. They differ categorically from the Navier Stokes Equations in that they are ordinary, rather than partial, differential equations and are therefore much more tractable for numerical solution. The first describes how each of the large-scale variables X_k evolves under the influence of three factors: the neighbouring X variables, any large-scale forcing F, and the kth sub-group of small-scale variables Y_{jk}. The second equation describes how each of the small-scale variables Y_{jk} evolves in a similar way. Three parameters describe the relationship between the different scales of variable: c, b and h as described below. The two equations are (Lorenz 2006b),

$$\frac{dX_k}{dt} = X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}, \tag{2.4}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k. \tag{2.5}$$

It is necessary to choose the values of F, b, c and h to suit a particular experiment. F represents the 'forcing' on the large-scale variables – it is essentially an energy input to the system, which is necessary for the system to behave chaotically. If F is too small, the system is not chaotic but periodic and poorly represents the real atmosphere. Making F very large, on the other hand, produces a system that is so chaotic that 1 MTU corresponds to very many atmospheric days and very small timesteps are necessary to replicate the time-scales of realistic weather. b is the ratio of the amplitudes of the large- and small-scale variables: it is this that differentiates them spatially, and b is conventionally set to 10 so that the two differ by an order of magnitude. But the large-scale variables also evolve more slowly than the small-scale ones: the parameter c represents the ratio of their rates of evolution. If c is very large there is too sharp a divide between large- and small-scale variability; the real atmosphere contains no such clear division, so smaller values of c are more realistic, but these also make the system more chaotic. Lorenz found that setting $F = 20$ and $c = 10$ produced an atmosphere suitable for carrying out experiments pertinent to the real atmosphere. It had a doubling time of approximately 0.3 MTUs, making 1 MTU equal to five atmospheric days (Lorenz 2006b). The final parameter, h, determines the amount of influence the large- and small-scale variables have on one another, and is conventionally set to 1.

When this system is solved on a computer, the values of each of the X_k variables are updated after each timestep of length Δt, which is a chosen fraction of the Model Time Unit, using a fourth-order Runge-Kutta (RK4) scheme. Both the timestep and the numerical method used can affect the time-evolution of the equations for a given set of initial conditions, potentially causing the numerical solution to differ from the analytical one even in double-precision. Here, it was assumed that the double-precision numerical solution was approximately accurate, with the justification that the system appeared to be stable and behaved in accordance with expectations in this and the previous studies in which it was also employed. The RK4 scheme is an iterative, numerical method for integrating a first-order differential equation of the form $dy/dx = f(x, y)$ that one wishes to solve to find the function y. It is possible to calculate the value of y at timestep $i + 1$ from its value at timestep i by using a Taylor expansion to evaluate $y_{i+1} = y(x_i + \delta)$ where

$x_i + \delta$ is the value of x at timestep $i + 1$. The higher the order of the expansion, the more accurate it becomes; to fourth order, the expansion would be (Riley et al. 2006),

$$y(x_i + \delta) = y(x_i) + \delta f(x_i, y_i) + \frac{\delta^2}{2}\left(\frac{df}{dx}\right)_{x_i} + \frac{\delta^3}{6}\left(\frac{d^2 f}{dx^2}\right)_{x_i} + \frac{\delta^4}{24}\left(\frac{d^3 f}{dx^3}\right)_{x_i}. \quad (2.6)$$

But this involves high-order derivatives that are not straightforward to compute. Under the Runge-Kutta method this is approximated as,

$$y_{i+1} = y(x_i) + \frac{1}{6}(c_1 + 2c_2 + 2c_3 + c_4), \quad (2.7)$$

Where c_1, c_2, c_3 and c_4 are prescribed functions of x_i, y_i and δ.

This method must be applied in turn to all the variables that are sought explicitly at each time-step, which in the fully resolved two-level system means K X-variables and JK Y-variables. The cost associated with this can be reduced, however, by 'parameterising' the smaller-scale Y-variables. This parameterisation means that only the X-variables need to be explicitly solved for, and the effect of the unresolved Y-variables on these is approximated from what the X-variables were at the previous timestep. The parameterisation function that describes this influence is denoted $U_p(X_k)$, which requires a much less costly calculation to compute. The only equation to solve by the comparatively costly iterative method is,

$$\frac{dX_k}{dt} = -X_{k-1}(X_{k-2} - X_{k+1}) - X_k + F - U_p(X_k), \quad (2.8)$$

2.2.2. Extension to a Three-Tier System

To enable two different 'resolutions' of parameterised models to be compared, for this study a new, three-level system based on the Lorenz '96 one was created by adding a third tier of smaller-scale variables labelled Z_{ijk} that relate to the Y_{jk} medium-scale variables in a similar manner to that in which the Y_{jk} relate to the large-scale X_k variables. There are I small-scale variables labelled $i = 1, \ldots, I$ for every medium-scale variable, giving rise to a total number of $IJK + JK + K$ variables. The system is governed by three differential equations, the first of which is identical to that of the two-level system that describes the evolution of X_k, the second of which introduces a

term describing the influence of Z_{ijk} on the evolution of Y_{jk} and the third of which describes the evolution of the Z_{ijk}, each of which is directly coupled to one Y_{jk} variable but not to any of the X_k:

$$\frac{dX_k}{dt} = X_{k-1}(X_{k+1} - X_{k-2}) - X_k + F - \frac{hc}{b}\sum_{j=1}^{J} Y_{j,k}, \tag{2.9}$$

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k - \frac{fd}{e}\sum_{l=1}^{L} Z_{i,j,k}, \tag{2.10}$$

$$\frac{dZ_{i,j,k}}{dt} = deZ_{i-1,j,k}\left(Z_{i+1,j,k} - Z_{i-2,j,k}\right) - g_Z dZ_{i,j,k} + \frac{fd}{e}Y_{j,k}. \tag{2.11}$$

The three tiers of variables are cyclically linked in space as shown in Figure 2.1, where $I = J = K = 8$. Although each Z-variable is one of a set of eight belonging to a given Y-variable, those at the edges of the set (with i =1 or 8) interact with those at the edges of adjacent sets, which is to say that $Z_{8,j,k}$ is linked to $Z_{1,j+1,k}$ and so on, and the same is true for the Y-variables (with j =1 or 8) at the edges of each X-variable set. This means that waves can propagate around the whole circular atmosphere on all three spatial scales independently.

To make the relationship between the small- and medium-scale variables as symmetric as possible with that between the medium- and large-scale variables, for this study $b = d$ and $c = e$ always. Lorenz' original paper and numerous studies since – see for example Arnold et al. (2013) and Wilks (2005) – have used $h = 1$, $b = 10$ and $c = 10$, which is followed here for simplicity. Those studies took $K = 8$ and $J = 32$, giving 264 variables overall. In this study, $K = J = I = 8$, yielding 584 variables, so that the system is not much more computationally complex than those two-dimensional ones and is symmetric between all three spatial scales. This also means that at successively smaller scales the number of variables increases by a factor of ~9, close to the 8-fold increase in complexity that doubling the horizontal resolution in a real-world GCM would induce.

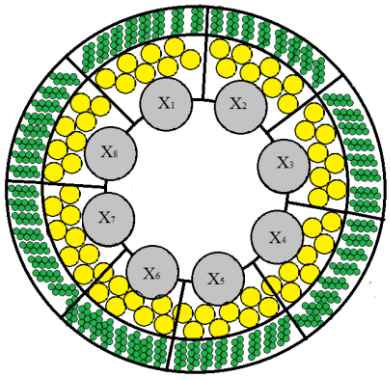

Figure 2.1: Diagrammatic representation of the three-level Lorenz '96 system created in this study. The 8 large-scale variables X_k (labelled and shaded in grey), are each coupled to eight medium-scale variables Y_{jk} (not labelled, yellow) which are coupled to eight small-scale variables z_{ijk} (not labelled, green).

For the three-level scheme, it is possible to formulate 'Low Resolution' parameterised models in which, just as for the two-level scheme, Y_{jk} variables are parameterised, the X_k variables evolve according to Equation 2.8 and the Z_{ijk}-variables are ignored altogether. But it is also possible to create more computationally costly 'High Resolution' parameterised models in which the X_k-variables evolve according to Equation 2.9, but the Z_{ijk} variables are parameterised and so the Y_{jk}-variables, though explicitly resolved, evolve according to the equation,

$$\frac{dY_{j,k}}{dt} = -cbY_{j+1,k}\left(Y_{j+2,k} - Y_{j-1,k}\right) - cY_{j,k} + \frac{hc}{b}X_k - U_p\left(Y_{jk}\right). \quad (2.12)$$

Now, the influence of the smallest-scale variables is approximated as a function of the medium-scale variables, which is denoted $U_p\left(Y_{jk}\right)$.

2.3. Parameterisation Schemes

In operational weather and climate models, processes occurring at unresolved scales are parameterised, so that their effects on the large-scale flow are approximated as a function of the large-scale parameters. This function may be entirely deterministic, or include stochastic noise to account for the uncertainty associated with the small-scale dynamics which in reality do not, in a chaotic system, take unique values for a given large-scale configuration and are hence not explicitly predictable from the larger-scale flow.

The two-level Lorenz '96 system has been used in previous studies to evaluate the performance of parameterisation methods or 'schemes' that mimic the sorts of functions that could be used in real-world atmospheric models. The aim when developing parameterisation schemes is to devise the most accurate approximation function, in this case $U_p(X_k)$ or $U_p\left(Y_{jk}\right)$ for use in Equation 2.8 or 2.12, that can be evaluated for the smallest computational cost. U_p will only ever

be an approximation of the true influence at any given time of variables on a smaller scale than that to which the model has been truncated, which is known as the true 'sub-grid tendency' and denoted $U(t)$. The simplest approximation for $U(t)$ is a fixed, deterministic function in which $U_p(X_k)$ will always have the same value for a given value of X_k. In the idealised Lorenz '96 system, this function can be obtained by evaluating the 'truth' timeseries for many different sets of initial conditions and plotting the true sub-grid tendency $U(t)$ against X_k, then finding the best-fit polynomial to the resulting curve. This deterministic function for U_p is denoted U_{det} to distinguish it from those produced by more sophisticated methods. For example, when parameterising the Y_{jk}-variables, fitting the points in the X_k plot to a quartic polynomial yields (Wilks 2005),

$$U_p(X_k) = U_{\text{det}}(X_k) = b_0 + b_1 X_k + b_2 X_k^2 + b_3 X_k^3 + b_4 X_k^4, \tag{2.13}$$

where $b_0, \ldots, b_4$ are constants.

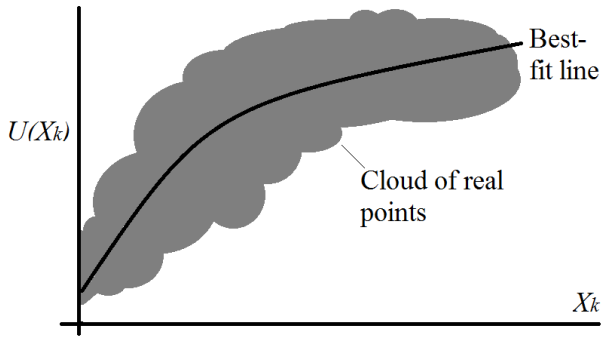

Figure 2.2: Illustration of the 'cloud' of real points observed when sub-grid tendency $U(X_k)$ is plotted as a function of the variable X_k, and how the deterministic best-fit tendency relates to this.

For such a deterministic scheme to be accurate, all the points on the graph of $U_p(X_k)$ versus X_k or $U_p(Y_{jk})$ versus Y_{jk} would have to lie on the curve described by Equation 2.13, so that the effect of small-scale variables on each large-scale variable X_k of the system were determined entirely by the value of X_k. In reality, there will be many combinations of the small-scale variables all giving the same large-scale configuration, and a cloud of points will surround the deterministic curve when $U(X_k)$ is plotted against X_k, as illustrated in Figure 2.2. This means that the deterministic scheme is not likely to be very accurate, motivating the inclusion of a random error

term $e_k(t)$ to account for the deviation of the curve from its deterministic form to produce more sophisticated (although also more computationally costly) 'stochastic' parameterisation schemes. For these, (Arnold et al. 2013),

$$U_p(X_k) = U_{\text{det}}(X_k) + e_k(t). \tag{2.14}$$

Previous studies of the two-level Lorenz '96 system have used five different types of stochastic parameterisation scheme, each of which essentially differs in the form of $e_k(t)$. Wilks (2005) proposed a first-order autoregressive stochastic model, for which,

$$e_k(t) = \phi e_k(t - \Delta t) + \sigma_e(1 - \phi^2)^{1/2} z(t). \tag{2.15}$$

Here the autoregressive parameter ϕ is the autocorrelation between $e_k(t)$ and $e_k(t + \Delta t)$, a best fit for which can be obtained from 'truth' runs of the system. A random number $z(t)$, picked from a zero-mean and unit variance distribution of values between -1 and 1, represents the uncertainty in the exact value of U_p for a given value of X_k, and σ_e is the standard deviation in the $e_k(t)$. Rather than remaining static, the parameterisation term evolves over time, which offsets the state-dependent structural error (Hansen and Penland 2007). Two stochastic parameterisation schemes were proposed on this basis: a 'white noise' parameterisation for which the sub-grid tendencies are assumed not to be temporally correlated so that $\phi = 0$ and a 'red noise' one where ϕ is measured from the autocorrelation of $e_k(t)$. In both cases, $\sigma_e = s_e$, where s_e is the standard deviation measured in the 'truth' run values of $U_p(X_k)$. Both the red- and white-noise schemes constitute 'additive' noise parameterisation because $e_k(t)$ is simply added to $U_{\text{det}}(X)$.

Arnold et al. (2013) built on Wilks' work to produce three new parameterisation schemes for the Lorenz '96 model. The first is 'State-Dependent' (SD) noise, which remains additive but disposes of the unwarranted assumption that the standard deviation in the noise σ_k must equal that measured in the sub-grid tendency $U(t)$ and is independent of the large-scale variable X_k. Instead,

$$\sigma_e = \sigma_1|X(t)| + \sigma_0, \tag{2.16}$$

where the constants σ_0 and σ_1 are deduced by binning the standard deviations measured in the 'truth' run values of $U(t)$ according to the magnitude of X_k and finding the curve best fitting the resulting discretised plot of $\sigma_e(X_k)$.

There is no reason to suppose that the noise in $U_p(X_k)$ should be independent of the magnitude of U_p itself. Therefore, a second alternative parameterisation was proposed by Arnold et al. that assumed 'Multiplicative' (M) rather than additive noise. Here, Equation 2.15 is still used to evaluate $e_k(t)$, but Equation 2.14 is modified so that,

$$U_p(X_k) = U_{\text{det}}(X_k)[1 + e_k(t)]. \tag{2.17}$$

This scheme, however, is unlikely to represent the true uncertainty well because it implies that the deviation of U_p from U_{det} should approach zero for very small U_{det}, something that is not at all apparent in the sub-grid tendency plot illustrated in Figure 2.2. Therefore, a fifth, 'Multiplicative-Additive' (MA) parameterisation was also considered that contains both types of noise, so that,

$$e_k(t) = |U_{\text{det}}|\epsilon_m(t) + \epsilon_a(t), \tag{2.18}$$

where ϵ_m and ϵ_a are separate multiplicative and additive sources of noise. The two noise terms are each related to a single random number $\epsilon(t)$, with the corresponding scaling constants σ_m and σ_a evaluated through a best-fit to the observed noise profile. Then,

$$e_k(t) = \epsilon(t)[\sigma_m|U_{\text{det}}| + \sigma_a]. \tag{2.19}$$

The modulus $|U_{\text{det}}|$ must be used here to ensure that the two sources do not cancel one another out.

2.4. System Setup

In this investigation, the modified Lorenz '96 system equations were solved numerically in the Fortran 90 language. 'Truth' runs were performed by integrating Equations 2.9-2.11 simultaneously, resolving all three scales of variables. The parameters b, e, f and h were set as outlined in Section 2.2.2. The remaining three parameters – the forcing F and the relative rapidities c and d – were varied during tests of Mixed Precision to simulate slightly different 'atmospheres'.

Initial tests were used to verify that these values yielded a suitably chaotic atmosphere, and showed that three atmospheres should provide ample variety: one with relatively high chaoticity (small EDT) for which $F = 20$ and $c = d = 10$; a slightly less chaotic amtmosphere with $F = 10$ and $c = d = 4$; and a still less chaotic (relatively high EDT) atmosphere with $F = 10$ and $c = d = 10$. However, in reduced precision experiments it was found that there was no significant difference between the $c = d = 10$ and $c = d = 4$ cases. Hence, to keep the discussion concise, detailed results are not included in what follows for the $c = d = 4$ setup.

Model Type	Bits in X	Bits in Y	p_X	p_Y	Relative Cost
D	64	-	-	-	1
DD	64	64	-	-	9
DS	64	32	-	-	5
DH	64	16	-	-	3
SH	32	16	-	-	2.5
SP1	64	64	0	0.05	5.5
SP2	64	64	0	0.1	5
SP3	64	64	0.001	0.05	5.5
SP4	64	64	0.001	0.1	5
SP5	64	64	0.005	0.01	6

Table 2.1: Characteristics of the ten model types compared in this study. Double-precision and Mixed Precision models are labelled 'D' for 'double', 'S' for 'single' and 'H' for 'half'-precision using two letters to represent the precision in the X- and Y- variables respectively. Stochastic processor models are labelled 'SP' and numerated in order of increasing bit-flip probability. The number of bits for each variable, the probability of a bit-flip occurring for the X-variables (p_X) and Y-variables (p_Y) and the estimated relative computational costs are listed. Changes in p_X alone affect the cost very little because there are eight times as many Y-variables as there are X-variables.

Since the equations were to be solved on standard, double-precision hardware, it was necessary to emulate reduced precision hardware as detailed in Section 1.4. The 'truth' was compared to low-resolution double-precision models in which the medium-scale Y –variables were parameterised and high-resolution, reduced precision models in which the small-scale Z-variables were parameterised. The parameterisation schemes were initialised with parameters tuned to each setup separately, as outlined in Section 2.4.2. For all three setups, experiments were performed to

compare the 'truth' to a range of lower-resolution model types, the characteristics of which are detailed in Table 2.1. Models in double-precision (D) at high and low resolution were compared to those with single- (S) or half- (H) precision emulated at specified spatial scales or those for which Stochastic Processor (SP) bit-flips were emulated.

2.4.1. Estimated Potential Savings in Computational Cost

Artificially emulating reduced precision is computationally costly, so explicit measurements of the models' relative computational costs cannot be made. However, Table 2.1 includes rough estimates of the costs (relative to type D) that might be expected with hypothetical hardware. Although the reduced precision methods all have higher cost estimates than type D, most are considerably cheaper than type DD. The estimates are based on the 'effective number of variables' resolved by each model type, a function of the actual number of variables and the effective computational cost of each variable relative to double-precision. The small costs associated with initialisation and output, which are left in double-precision; with switching between different processors to exploit different levels of precision; and with the increased number of links between interconnected variables in High Resolution models, are ignored. Hence, the computational cost savings given here are slight overestimates of what could realistically be achieved.

In Mixed Precision, it is assumed that the cost required to compute any single X- Y- or Z-variable is independent of spatial scale but is proportional to the number of bits with which that variable is represented. Model type D explicitly resolves 8 variables in double-precision so the effective number is also 8. Type SH uses half the number of bits to resolve each of the 8 X-variables and a quarter the number to resolve each of the 64 Y-variables, so its effective number is $8/2 + 64/4 = 20$, yielding a cost relative to type D of $20/8 = 2.5$.

For the Stochastic Processor models, the fractional voltage saving V is estimated for a given probability of bit-flips using Figure 4 of Sloan et al. (2010), which is based on measurements of FPGA hardware with a stochastic 'fault injector'. This is converted to an energy saving E

through the relationship $E \propto P \propto V^2$ where P is the dynamic power consumption (Karakonstantis and Roy 2011), a very crude approximation (Snowdon et al. 2005) that gives only a rough estimate. Then, the effective number of variables is computed by multiplying the number of X- and Y-variables by the value of $(1-E)$ appropriate to each spatial scale. For instance, Model type SP1 has a probability $p_X = 0$ of flips in the 8 X-variables and $p_Y = 0.05$ of flips in the 64 Y-variables which reduces the cost of each Y-variable by E=44% according to the figure in Sloan et al. (2010), so the effective number of variables is $8 \times 1 + 64 \times 0.56 \approx 44$. The relative cost is $44/8 = 5.5$.

2.4.2. Formulating Parameterisation Schemes

The best-fit parameters necessary for parameterising Y_{jk} and Z_{ijk} using each of the five schemes were computed from the 'truth' following the method of Wilks (2005). The sub-grid tendencies $U(X_k)$ and $U(Y_{jk})$ can be explicitly calculated for each of the eight X_k and thirty-two Y_{jk} respectively at each timestep by evaluating Equations 2.8 and 2.12 and computing the difference between these and the same equations with $U_p(X_k)$ and $U_p(Y_{jk})$ omitted. Best-fit quartic polynomials are then computed for plots of $U(X_k)$versus X_k and $U(Y_{jk})$ versus Y_{jk} to obtain the parameters $b_0, \ldots, b_4$ in each case. The values obtained here are listed in Table 2.2. These are all the parameters required to implement a deterministic parameterisation scheme.

	$F = 20, c = d = 10$		$F = 10, c = d = 10$		$F = 10, c = d = 4$	
	$U(X_k)$	$U(Y_{jk})$	$U(X_k)$	$U(Y_{jk})$	$U(X_k)$	$U(Y_{jk})$
b_0	0.10730	0.03767	0.07433	0.03097	0.04440	0.01587
b_1	0.24372	0.23570	0.35570	0.3382	0.10781	0.09132
b_2	0.00017	0.11910	-0.00071	0.2411	0.00157	0.10477
b_3	-0.00051	-0.05953	-0.00293	-0.43198	-0.00036	-0.07299
b_4	0.00001	0.00184	0.00014	0.20960	0.00000	-0.00311

Table 2.2: The best-fit parameters to the quartic polynomial (Equation 2.13) used to implement parameterisation schemes for the Lorenz '96 system by approximating $U(X_k)$ (to parameterise Y_{jk}) or $U(Y_{jk})$ (to parameterise Z_{ijk}).

	$F = 20, c = d = 10$		$F = 10, c = d = 10$		$F = 10, c = d = 4$	
	$U(X_k)$	$U(Y_{jk})$	$U(X_k)$	$U(Y_{jk})$	$U(X_k)$	$U(Y_{jk})$
ϕ	0.96254	0.987499	0.98787	0.993647	0.99621	0.99838
σ_e	1.49890	0.150887	1.07395	0.117636	0.46383	0.044804
σ_0	0.62227	0.03217	0.41570	0.07225	0.25083	0.09737
σ_1	0.01028	0.0213	-0.01260	0.08541	-0.00241	0.08867
σ_a	0.63612	0.08260	0.41550	0.05380	0.25910	1.12650
σ_m	0.01400	0.18610	-0.03382	0.17720	-0.04150	-0.01838

Table 2.3: The best-fit parameters obtained to implement stochastic parameterisation schemes (see Section 2.3) to approximate $U(X_k)$ (to parameterise Y_{jk}) or $U(Y_{jk})$ (to parameterise Z_{ijk}).

To apply stochastic schemes, it is necessary to deduce a number of additional parameters. The additive schemes require the standard deviation σ_e and, if 'red noise' is to be simulated, also the lag-1 autocorrelation ϕ of the timeseries $X_k(t)$ or $Y_{jk}(t)$, which can be straightforwardly be computed from a long 'truth' run. Here, a run of 2 million timesteps (each of length $\Delta t = 0.005$ MTUs) was used. For the state-dependent scheme, the state-dependent standard deviation parameters σ_0 and σ_1 were computed by binning X_k or Y_{jk} into bins of width 0.1 (see Section 2.3). For the Multiplicative Additive scheme, σ_m and σ_a were computed by best-fitting Equation 2.19 to the sub-grid tendency plots. The parameters obtained are listed in Table 2.3.

2.4.3. Turbulence in Lorenz '96

The chaoticity of an atmospheric model is proportional to the inverse of the Error Doubling Time (EDT), which was computed for the X-variables, as outlined below, to be 0.8 and 0.4 MTUs for the $F = 10$ and $F = 20$ setups respectively. Comparing these values to the large-scale EDT of around 1.5 days that Lorenz (2006b) identified for the most advanced models of the real atmosphere (in 1996), it was deduced that 1 MTU ≈ 2-4 real atmospheric days for these setups.

The EDTs were computed here following the method of Lorenz (2006b) as outlined in Section 2.1. Fifty pairs of 5 MTU runs of the truth system with random, independent initial conditions were carried out. Each pair consisted of one unperturbed run and one run in which the initial conditions were randomly changed (independently for each of the 8 X-variables, the 64 Y-

variables or the 512 Z-variables as appropriate) by up to $\pm 0.1\%$. If the values of the X-variables, for example, at each timestep n for runs with and without a perturbation are denoted X'_{kn} and X_{kn} respectively, the 'error' in each variable is,

$$e_{kn} = X'_{kn} - X_{kn}. \tag{2.20}$$

The square of this error was then averaged across all eight X-variables and the logarithm of this was averaged over all fifty runs to define $\log\left(E^2(t)\right)$, from which $E(t)$ was calculated. The Error Doubling Time t_d was computed from the gradient of the initial exponential growth using Equation 2.3. A similar procedure can be followed for the EDTs in the Y- and Z-variables.

In an atmospheric model, chaos takes different forms depending on the scale-dependency of the turbulence it produces. Two forms that are important when describing the real atmosphere are 'two-dimensional' (2D) and 'three-dimensional' (3D) turbulence (Leith 1971). 2D turbulence acts to concentrate the energy E of the system at large spatial scales. This diminishes the importance of the smallest scales so that small errors do not rapidly propagate upwards (the turbulence is not 'scale-selective') and the error in a forecast of the system can be made arbitrarily small by minimising the uncertainty in the initial conditions. With 3D turbulence, the chaoticity is greater on smaller spatial scales and small-scale errors grow rapidly, imposing a lower limit on the forecast error regardless of the (inevitably non-zero) error in the initial conditions (Lorenz 1969).

Ideally, it is desirable that the model setups being investigated mimic one of these turbulence regimes. Lorenz demonstrated 3D turbulence in his setup of the two-level system by showing that the EDT was greater for the X- than for the Y-variables (Lorenz 2006b). Lorenz' method to compare EDTs at different scales independently could not be applied to the three spatial scales here because perturbing the Z-variables alone to measure the Z EDT knocks the system off its 'attractor' (the subset of states that the system is capable of exploring), which causes the error in Z to initially shrink while the system rebalances. Hence, the leading Lyapunov exponent, a direct measure of the chaoticity of the system as a whole, was measured instead by perturbing all the system variables simultaneously. The values obtained with all three tiers (λ_{XYZ}) were compared to those obtained using systems containing only X- and Y-variables (λ_{XY}) or only X-variables (λ_X).

To compute the leading Lyapunov exponents, a single, long run of the system was compared to an equivalent system that was perturbed using the 'breeding vector' method (Toth and Kalnay 1993) to ensure that the perturbed system always remained on the system's attractor. Following the initial spin-up, all variables in the system were perturbed by random amounts, with mean 0 and standard deviation 0.01; for the Y- and Z-variables the size of perturbations was reduced in proportion to their average magnitudes relative to the X-variables. The overall initial error magnitude $E(0)$ is,

$$E(0) = \frac{\sqrt{\sum_i \left(X_i - X_i'\right)^2 + \sum_{ij}\left(Y_{ij} - Y_{ij}'\right)^2 + \sum_{ijk}\left(Z_{ijk} - Z_{ijk}'\right)^2}}{K+J+I}, \tag{2.21}$$

where the non-primed and primed quantities are 'truth' and 'perturbed' values respectively.

The truth and perturbed systems were then evolved for 10000 timesteps. Every ten timesteps (an interval chosen by trial and error), the perturbed system was re-scaled so that the overall error magnitude, $E(10)$, matched that of the initial perturbation, $E(0)$. Except in the first 1000 timesteps (which were excluded to make the results independent of the exact perturbation used), the error magnitudes were measured at the end of each ten-timestep interval prior to rescaling and their average over all intervals, $\langle E(10) \rangle$, was computed. Under the assumption that the error grows according to,

$$\langle E(t) \rangle = E(0) \exp(\lambda t), \tag{2.22}$$

the Lyapunov exponent λ can be estimated by setting $t = 10$. This process was repeated five times and the results were averaged for each λ calculated. The procedure was the same for the one-tier and two-tier systems, except that only the X- or X- and Y-variables were included respectively.

For the $F = 10$ setup it was found that $\lambda_X = 2.2$, $\lambda_{XY} = 12.8$ and $\lambda_{XYZ} = 9.8$, whilst for the $F = 20$ setup, $\lambda_X = 4.5$, $\lambda_{XY} = 23.5$ and $\lambda_{XYZ} = 20.3$. These results imply that, regardless of setup, 3D (scale-selective) turbulence is simulated on large spatial scales, but introducing the Z-variables barely changes the chaoticity, which is more consistent with 2D (non-scale-selective) turbulence on small scales. 3D turbulence on all spatial scales could be obtained by reducing the damping parameter g_Z in Equation 2.11: this might be expected to make the Z-variables less sensitive to a reduction in numerical accuracy, since their behaviour would be more chaotic.

However, because the reduced precision models always parameterise the Z-variables, this was found to have a negligible effect on the reduced precision analysis, and $g_Z = 1$ was retained in the experiments presented below for maximum consistency with the two-tier system.

2.5. Scale-Selective Reduced Precision Experiments

The accuracy of the two double-precision and eight reduced precision model types (see Table 2.1) was assessed using four metrics: the models' ability to predict the short-term evolution of the system ('weather' simulations); the system's longer-term mean state ('climate' simulations); its chaoticity; and its tendency to oscillate between different behavioural 'regimes'. The models' predictions for the large-scale X-variables were compared to the 'true' X-variables evaluated using all three tiers of variables; because Low Resolution models cannot resolve Y-variables, none of the models were assessed on their ability to predict these. Models of type DS, DH, SP2 and SP4 are not discussed here because they behave almost identically to DD, SH, SP1 and SP3 respectively.

2.5.1. Short-term Forecasting Ability

The short-term forecasting ability of the models was assessed using the Absolute Error score and the Ranked Probability Skill Score (RPSS), each averaged over the 8 X–variables and over 250 independent sets of initial conditions to smooth the plots. The Ignorance Skill Score (IGNSS) was also used for some of the tests, but was found to give almost identical findings to the RPSS; IGNSS results will therefore not be included here for conciseness. For each initial condition, an ensemble of 50 forecasts was created that differed only in the random numbers invoked in the stochastic parameterisation schemes, and the Absolute Error was measured as the difference between the value forecast by the model and the 'truth', averaged over the 50 ensemble members.

To compute the RPSS (See Section 1.7.1), the cumulative pdf P_l for each model under investigation was created by binning the outputs predicted by the ensemble members into ten bins

of possible X-variable values, labelled using the index l, which were defined so that they had equal populations when the whole timeseries of the 'truth' run was binned. P_l was normalised by dividing by the number of ensemble members. The Ranked Probability Skill (RPS) of the ensemble forecast was defined by comparing this to the 'true' system's cumulative pdf T_l at this timestep, which was 0 for all bins below to the 'true' value and 1 thereafter. The RPSS was obtained from this by comparing the mean RPS across all 250 sets of initial conditions, $\langle \mathrm{RPS} \rangle$, to the RPS of a reference climatology in which all the bins are equally populated, $\langle \mathrm{RPS} \rangle_{\mathrm{ref}}$.

Figures 2.3 and 2.4 show the Absolute Error and RPSS scores of the models for the $F = 20$ and $F = 10$ setups respectively. The forecast skill improves when the resolution is increased (from D to DD) in all cases. But, under additive parameterisation, SH performs just as well as DD (any small differences would most likely disappear if the averaging was done over more runs). Under multiplicative parameterisation, SH begins to fall in accuracy relative to DD but only after around 1 MTU in both setups, which suggests that a lower-cost model that begins in single-precision but switches to half-precision once the forecast reaches a lead-time τ might not exhibit this decline.

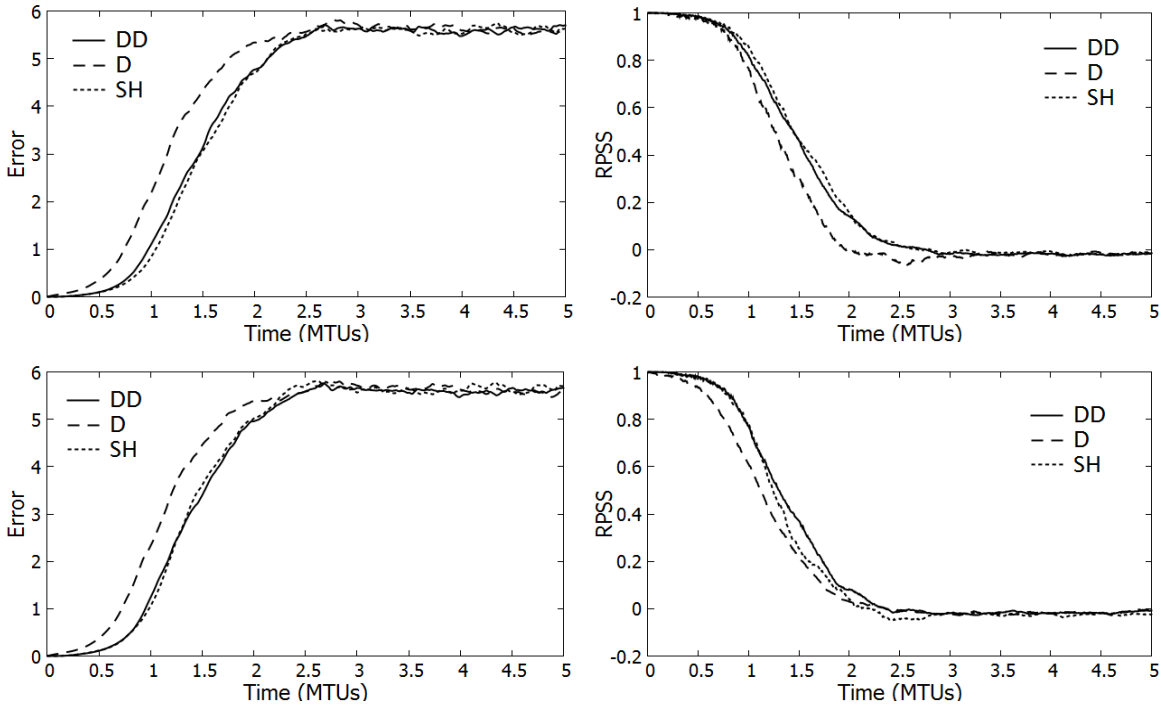

Figure 2.3: The Absolute Error (left) and RPSS (right) are plotted for models DD, D and SH of the $F = 20$ system using additive (top plots) and multiplicative (bottom plots) stochastic parameterisation schemes.

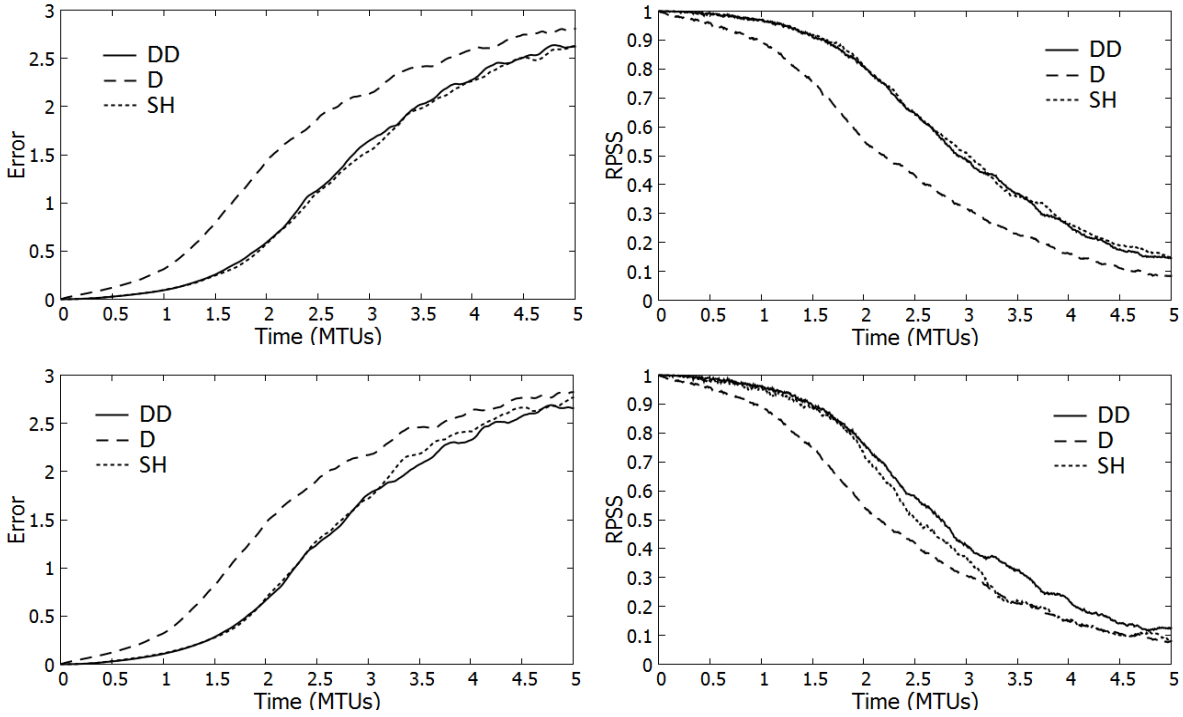

Figure 2.4: The Absolute Error (left) and RPSS (right) are plotted for models DD, D and SH of the $F = 10$ system using additive (top plots) and multiplicative (bottom plots) stochastic parameterisation schemes.

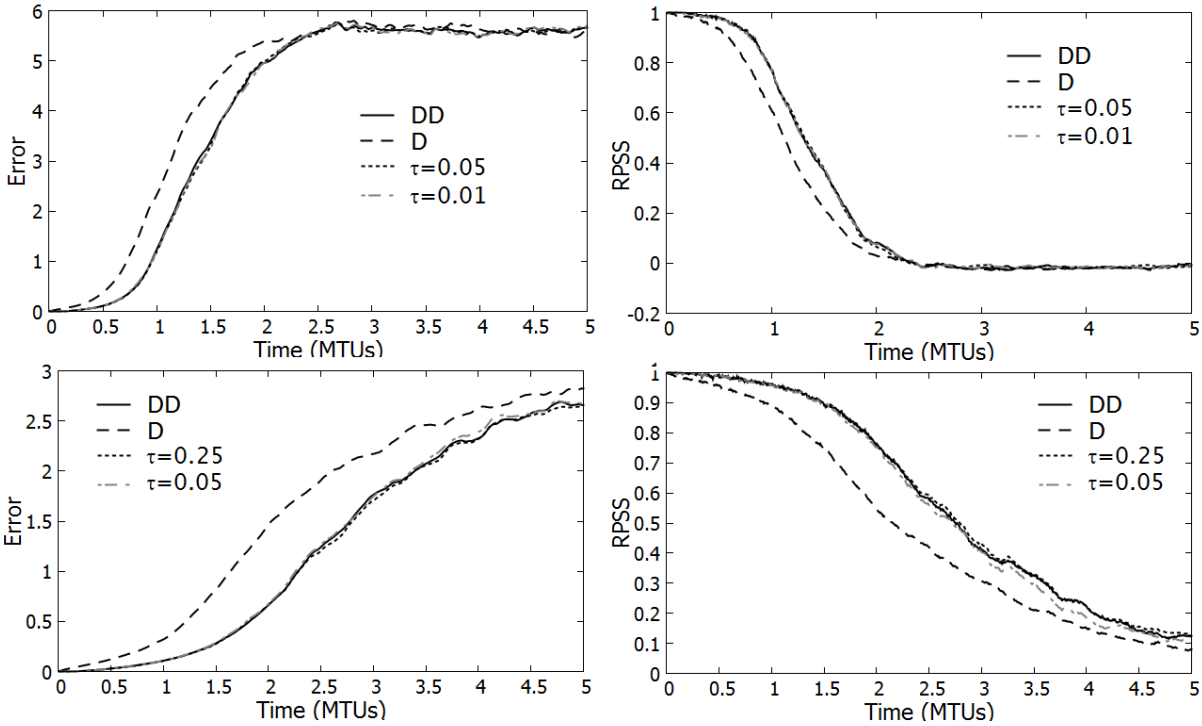

Figure 2.5: The Absolute Error (left) and RPSS (right), averaged over 250 runs, are plotted for models DD and D and selected models for which X and Y-variables respectively begin in single-precision but the Y-variables are switched to half-precision (SH) after a delay τ in MTUs, using multiplicative parameterisation for the $F = 20$ (top plots) and $F = 10$ (bottom plots) setups.

To test this idea, and to ascertain the smallest value of the delay τ that can be used to produce a forecast with negligible reduced precision error, High Resolution models were run again using the multiplicative stochastic parameterisation scheme but now with the precision in the Y-variables switched from single- to half- precision at values of τ ranging from 0.01 to 1 MTU. The Error and RPSS timeseries, again averaged over 250 runs, are shown in Figure 2.5 for selected values of τ. For both setups, $\tau = 0.25$ MTUs (~1 real atmospheric day) was sufficient for the reduced precision model to match Model DD exactly. τ could be reduced to as little as 0.01 MTUs (~1 hour) for the $F = 20$ setup with no noticeable decline in accuracy, whilst errors associated with half-precision became evident for $\tau \lesssim 0.05$ MTUs for the $F = 10$ setup. That implementing such a delay might be beneficial was indeed suggested by a previous study that used FPGAs to explicitly implement delayed-onset reduced precision in the two-tier Lorenz '96 system (Düben et al. 2015). This can be attributed to the onset of errors associated with half-precision being delayed until after these errors become very small compared to the overall model error, which increases with forecast lead-time. Implementing a similar delay in operational weather models could be a viable way of reducing the rounding errors in low-cost, reduced precision runs.

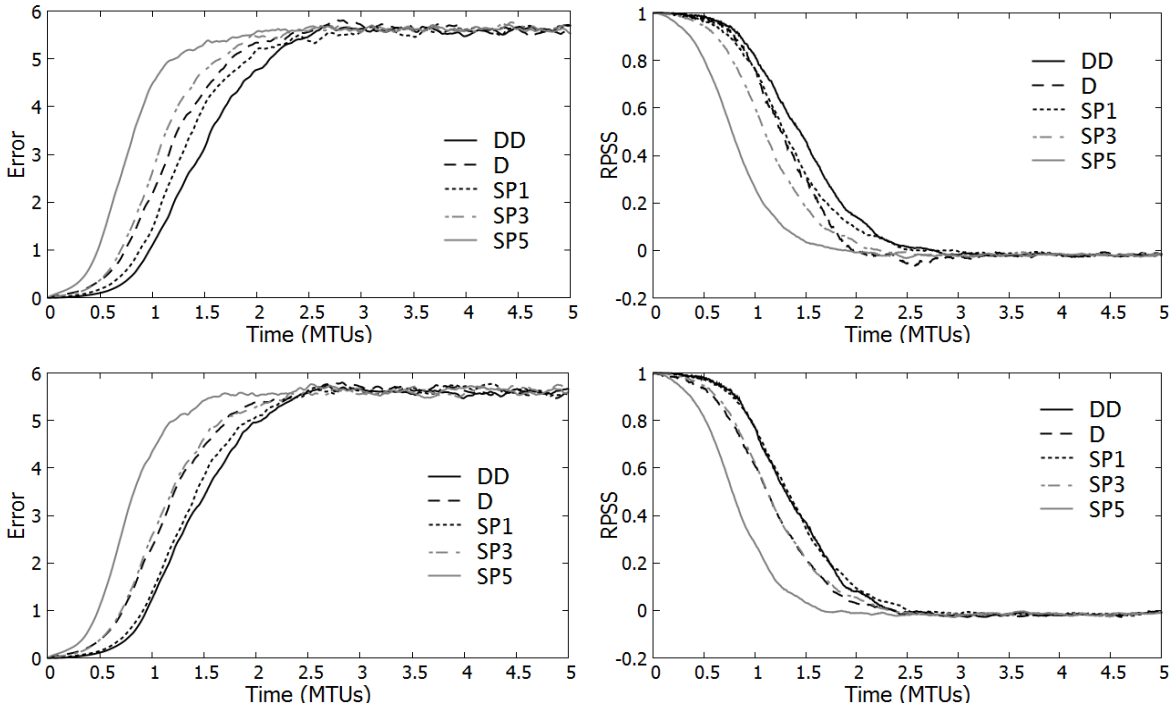

Figure 2.6: The Absolute Error (left) and RPSS (right) are plotted for Stochastic Processor models of the $F = 20$ system using additive (top plots) and multiplicative (bottom plots) parameterisation schemes.

Figures 2.6 and 2.7 show Error and RPSS plots for the 'Stochastic Processor' models. They illustrate that using a Stochastic Processor is only worthwhile if the large-scale variables are protected from bit-flips. Model type SP1, for which bit-flips were constrained to the Y-variables only (see Table 2.1), performs better than type D but not as well as the double-precision control type DD, showing that even a modest probability of faults in the calculations can have a deleterious effect on the results. Model type SP3, with a higher probability of bit-flips, performs no better than the low-resolution type D, whilst SP5 performs much worse, because it incorporates bit-flips in the X-variables: a fault rate of just 0.5% in the X-variables yields a very poor forecast. Further tests revealed that the forecasting ability of SP3 significantly improved when the two most important significand bits were protected from bit-flips, which could be a way of making Stochastic Processors more viable as a means of safely reducing computational costs, but it is unclear how such difficult-to-isolate bits could be protected in real hardware.

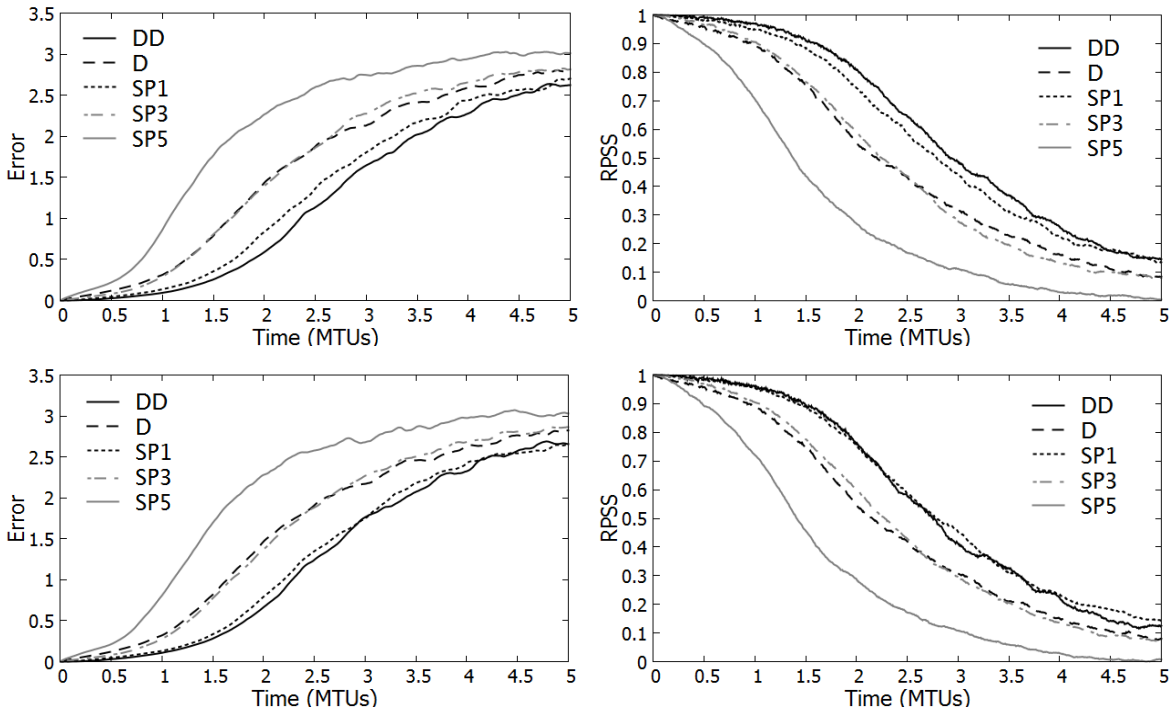

Figure 2.7: The Absolute Error (left) and RPSS (right) are plotted for Stochastic Processor models of the $F = 10$ system using additive (top plots) and multiplicative (bottom plots) parameterisation schemes.

2.5.2. Long-term Forecasting Ability

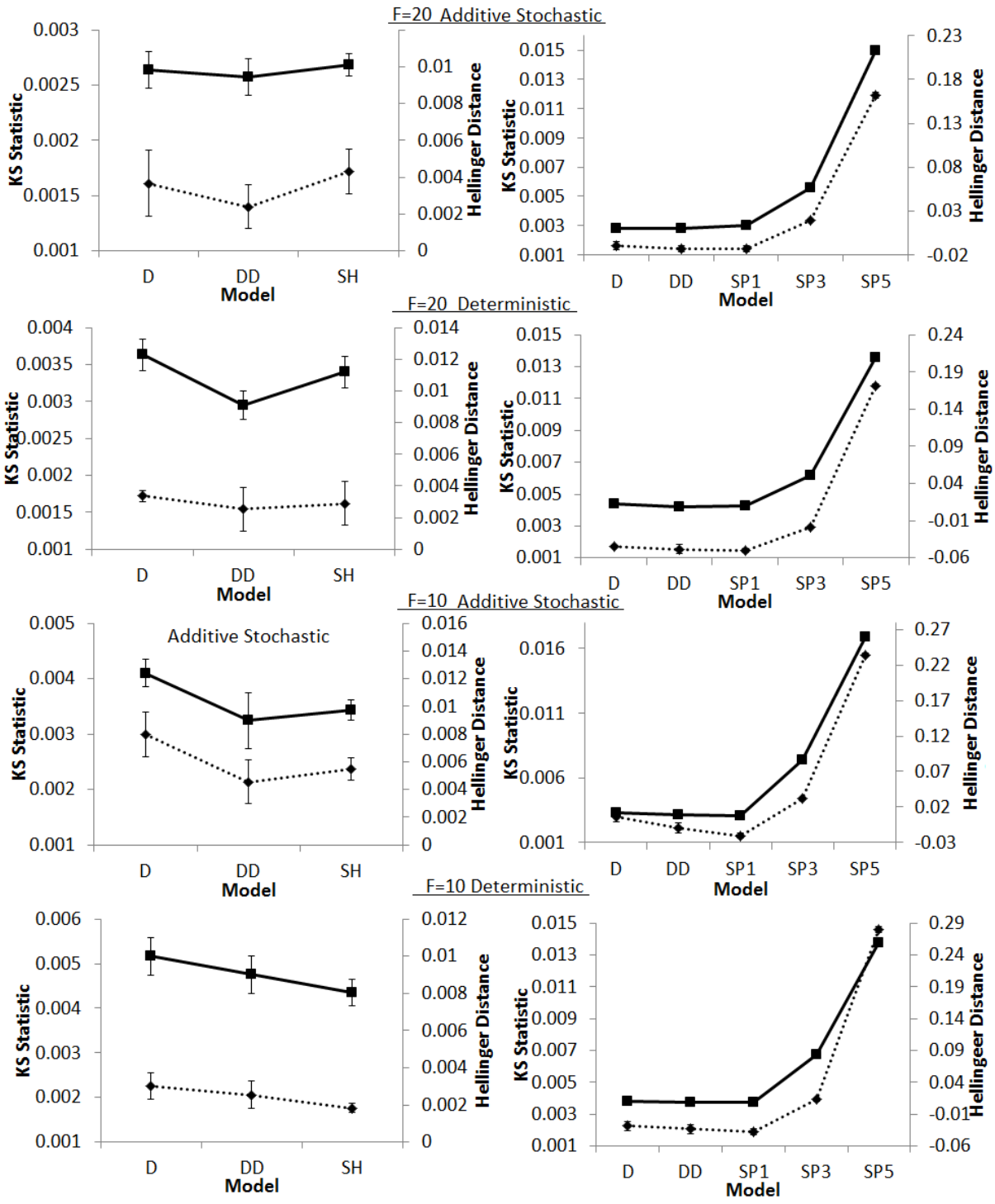

Figure 2.8: KS statistics (dotted lines) and Hellinger Distances (solid lines) for models DD, D and SH (left plots) or SP1-5 (right plots) of both setups, using additive stochastic (top), and deterministic (bottom) parameterisation schemes. Lines are drawn between the points to make the differences between models clearer. Error bars (not always visible) represent the standard error in the mean.

To evaluate the long-term forecasting ability of the models, the KS statistic D_n and Hellinger Distance H (see Section 1.7.2) were averaged across five long (50 000 MTU) integrations with independent initial conditions. The standard error in the mean (Hughes and Hase 2010) was used to quantify the uncertainty. Figure 2.8 shows the results for both setups, using additive stochastic and deterministic parameterisation schemes to elucidate the influence of the scheme stochasticity. In both setups the double-precision and Mixed Precision models generally lie within the error bars of one another, although reducing resolution (type D) increases the Hellinger distance significantly when using the deterministic scheme in the $F = 20$ setup and the stochastic schemes in the less chaotic $F = 10$ setup.

Amidst all the reduced precision models, only SP1, under additive parameterisation with $F = 10$, exceeds type D in accuracy by a significant amount. Most importantly, though, there is no significant loss of accuracy relative to DD in any of the reduced precision cases given the error, except when bit-flips are allowed in the X-variables (SP3-5). This means that, in general, reducing precision has no negative impact on the long-term forecasts' quality, and that, on the other hand, resolution is less critical for long-term than for short-term forecasts and therefore does not always affect D_n and H by a significant amount. There is hence little difference between high resolution and high precision options for long-term forecasts of the Lorenz '96 system; it is however unlikely that this conclusion would also be true of more complicated systems.

2.5.3. Atmospheric Chaoticity

An accurate model of the real-Earth atmosphere should conserve the tendency for small perturbations to grow over time – the chaoticity, which can be measured using the Error Doubling Time as described in Sections 2.1 and 2.4.3. Figure 2.9 plots the EDTs computed from the timeseries for the 'truth' and model integrations of both setups by comparing 'unperturbed' runs of the model with 50 'perturbed' runs for which the initial conditions were slightly altered by a different amount each time. For the stochastic parameterisation schemes, some of the deviation

contributing to the difference between the perturbed and unperturbed runs is due not to this initial perturbation but to the stochasticity of the schemes themselves. Hence, the EDTs are larger – because the perturbed and unperturbed runs deviate from one another less rapidly – using the deterministic parameterisation scheme where all differences are attributable to the initial perturbation. Except in the deterministic case, model type D exhibits smaller EDTs further from the truth than those of DD, implying that increasing the resolution genuinely improves the accuracy of stochastic simulations.

Model types SH and DD both perform well, matching the true doubling time for the $F = 20$ setup (a value of 0.24) exactly under additive parameterisation, where model D underestimates it by 20% (with a value of 0.19). Among the Stochastic Processor models, SP1 – which contains no bit-flips in the X-variables and moderate bits-flips in the Y-variables (See Table 2.1) – sometimes simulates the true chaoticity well, even matching DD for accuracy in, for example, the case of stochastic parameterisation in the $F = 10$ setup, but in other instances performs poorly even relative to type D, indicating that even with a low probability of bit-flips Stochastic Processors do not reliably capture the fundamental characteristics of the simulated system. SP5, where the bit-flips are most likely, performs much worse than the other models.

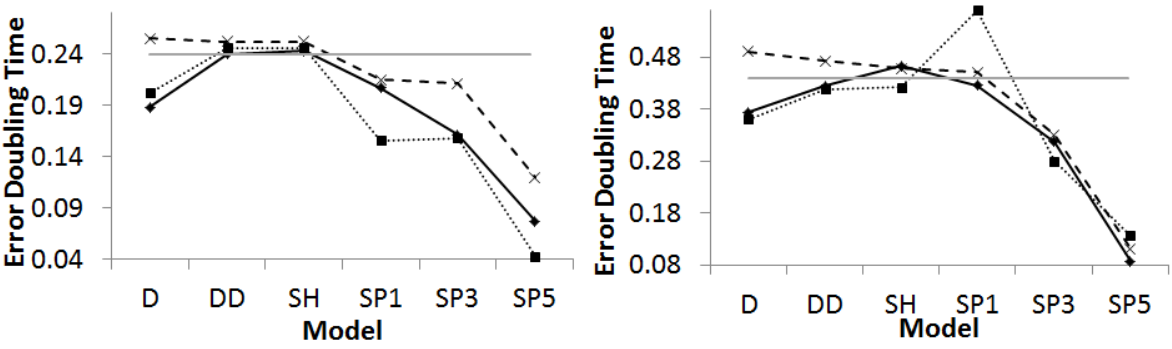

Figure 2.9: Error Doubling Times in MTUs for the 'truth' (grey lines), and models of type DD, D, SH and SP1, 3 and 5 using additive stochastic (solid lines), multiplicative stochastic (dotted lines) and deterministic (dashed lines) parameterisation schemes for the $F = 20$ (left) and $F = 10$ (right) setups. Lines are drawn between points to make the differences between models clearer.

2.5.4. Regime Behaviour

Christensen et al. (2015) have previously identified one example of dual-regime behaviour (see Section 1.7.3) in the two-tier Lorenz '96 system whereby the spatial covariance Σ of the eight large-scale X-variables, which describes how the X-variables are linked in space, switches between positive and negative states. To compute this, the individual covariance C for the four pairs of opposing X-variables was added up, according to,

$$\Sigma = C(1,5) + C(2,6) + C(3,7) + C(4,8). \tag{2.23}$$

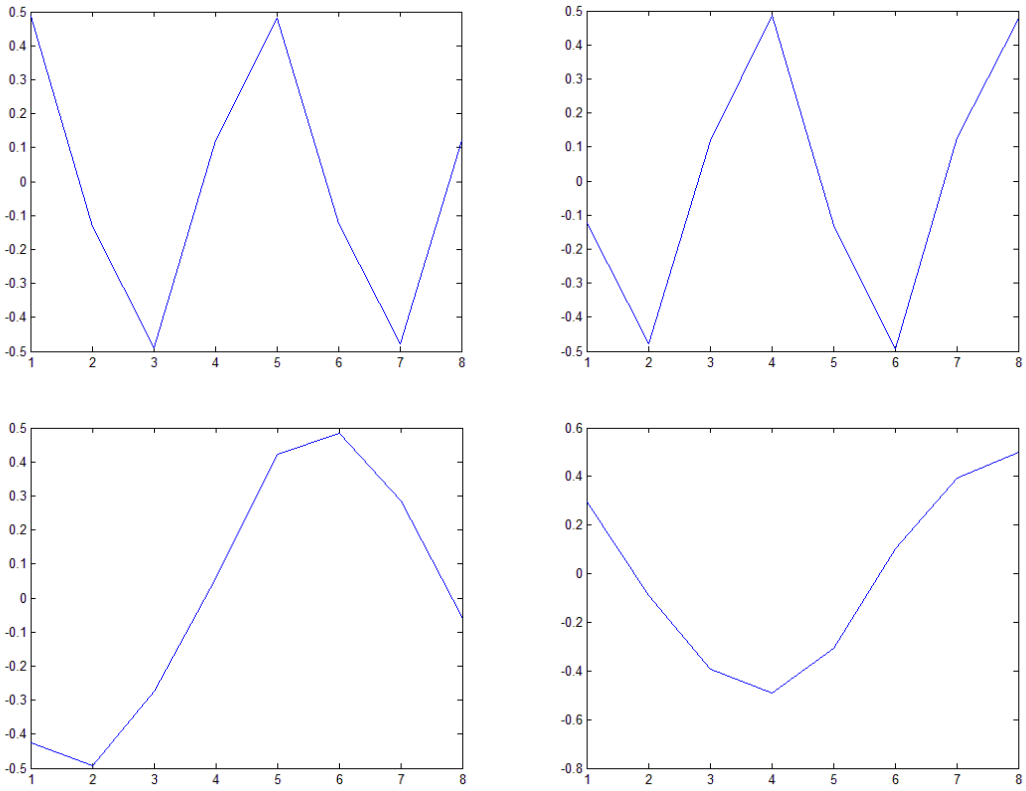

Figure 2.10: Principal Components 1 (top left), 2 (top right), 3 (bottom left) and 4 (bottom right) are plotted for the eight X_k variables in the three-level Lorenz '96 system introduced in Section 2.2. PCs 1 and 2 and PCs 3 and 4 constitute $\pi/4$ and $\pi/2$ out-of-phase pairs respectively.

The regimes were defined on the basis of whether this total is less than or greater than zero. Christensen et al. (2015) found a clear distinction between times at which opposite X-variables were in phase and $\Sigma > 0$ (which they dubbed Regime A) and those for which they were out of

phase and $\Sigma < 0$ (dubbed Regime B), distinguishing two regimes that satisfied Lorenz' criteria for regimes. They carried out a 50000 MTU (700 atmospheric 'years') run to define the 'truth', and the Principal Components (PCs) were computed for the entire time-series using the covariance matrix C, which is symmetric (it is a $K \times K$ matrix when there are K large-scale X variables) and can therefore be diagonalised, in which form it consists of $K = 8$ eigenvalues that correspond to $K = 8$ eigenvectors, each of which is a different 8-component Empirical Orthogonal Function (EOF). This matrix of diagonalised EOFS, V, is used to define the PCs at each timestep through,

$$PC_{ij} = \sum_k V_{jk} X_{ik}, \tag{2.24}$$

where, $i = 1, \ldots n$ for n timesteps in the series, $j = 1, \ldots, K$ and $k = 1, \ldots, K$ for $K = 8$ X-variables.

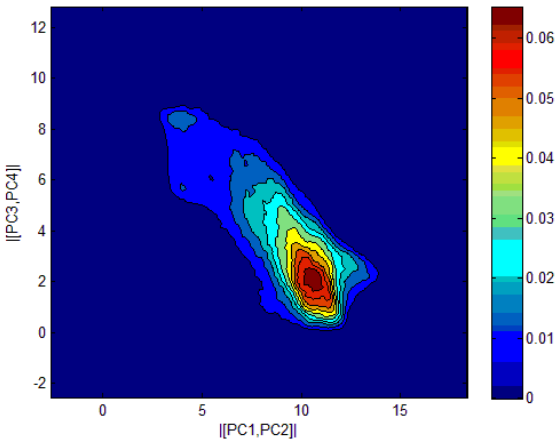

Figure 2.11: The probability density function (pdf) is plotted for a 2-level Lorenz '96 system similar to that used by Christensen et al. (2015) for a 5000 MTUs run with timesteps of 0.005 MTUs.

So long as the mean value has been subtracted from the original time-series, the first two Principal Components (PC1 and PC2) are $\pi/4$ out-of-phase dual-peak ('wave-2') sine functions, while the next two (PC3 and PC4) are $\pi/2$ out-of-phase and single-peak ('wave-1'). These four PCs are represented graphically in Figure 2.10, which was produced to replicate the findings of Christensen et al. (2015). They identified the dominance of PCs 1 and 2 at any given time with Regime A, when opposing X-variables are in phase, and the dominance of PCs 3 and 4 with Regime B behaviour, when opposing X-variables are out of phase. These two regimes could be clearly distinguished when the pdf was plotted as a contour plot in phase space as a function of

both pairs of similar Principal Components' magnitudes: that is to say, the pdf on the z-axis was plotted as a function of |PC1,PC2| on the x-axis and |PC3,PC4| on the y-axis, where the vertical scale represents the magnitudes of PCs 1 and 2 and those of PCs 3 and 4 being added in quadrature. This pdf represents the cumulative likelihood, across all timesteps, of the Principal Component pairs taking on each of their possible combinations of values. Combinations of values that are very rarely taken have low values on the contour plot, whereas those that are very popular exhibit large values. To illustrate this, Figure 2.11 shows a reproduction of the same regime behaviour in PC space that Christensen et al. (2015) observed for the two-level Lorenz '96 system.

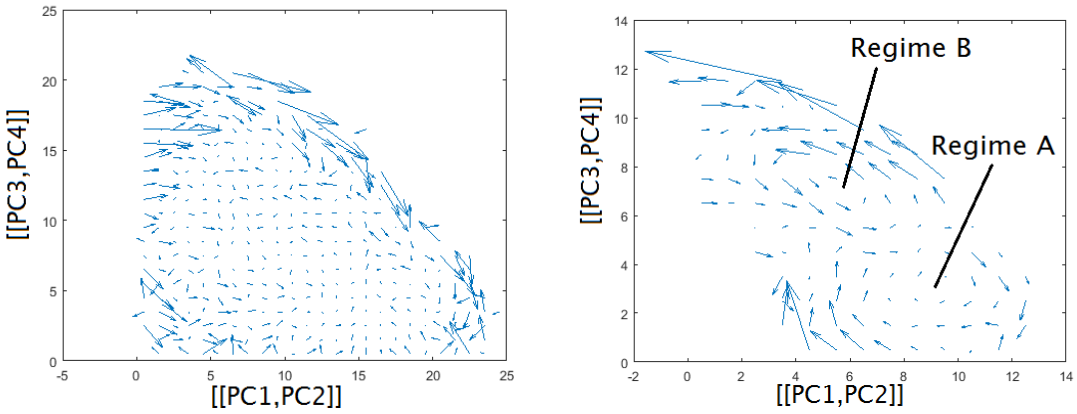

Figure 2.12: Vector plots for the $F = 20$ (left) and $F = 10$ (right) setups' truth timeseries. At each timestep, the system is at a different point in Principal Component (PC) space. Across all timesteps, these plots show the average direction in which all points in PC Space within a radius of 0.5 Principal Component Units from each arrow's tail move in 1 timestep. The size of the arrow reflects the average magnitude of the motion. The two centres of rotation in the $F = 10$ plot correspond to regimes A (clockwise) and B (anticlockwise). At regime centres, the system's behaviour is so persistent that the arrows disappear entirely. The $F = 20$ plot does not demonstrate the presence of such clear regimes.

When a similar analysis was carried out for the three-level modified Lorenz '96 system used here, it was found that the system exhibits such regimes only in the less chaotic, $F = 10$ case. For the 'truth' and each model, PCs 1-4 were computed at each timestep for long runs of 5000 MTUs (approximately 40 atmospheric years) each. The average value was subtracted and the data were smoothed by applying a rolling average over 0.2 MTUs to make the regimes as visible as

possible, then two-dimensional pdfs were plotted as a function of |PC1,PC2| and |PC3,PC4|. The presence or absence of regimes was confirmed using 'vector plots' of the average trajectory at each point in PC space, akin to those plotted by Christensen et al. (2015). Any regimes are visible as the centres of the orbits of the Principal Components. These plots are shown for the 'truth' runs of the setups in Figure 2.12, and for selected double-precision and reduced precision models of the $F = 10$ setup in Figure 2.13. The Principal Component pdfs themselves are provided for the 'truth' integrations and for a selection of six models in Figures 2.14 and 2.15 respectively.

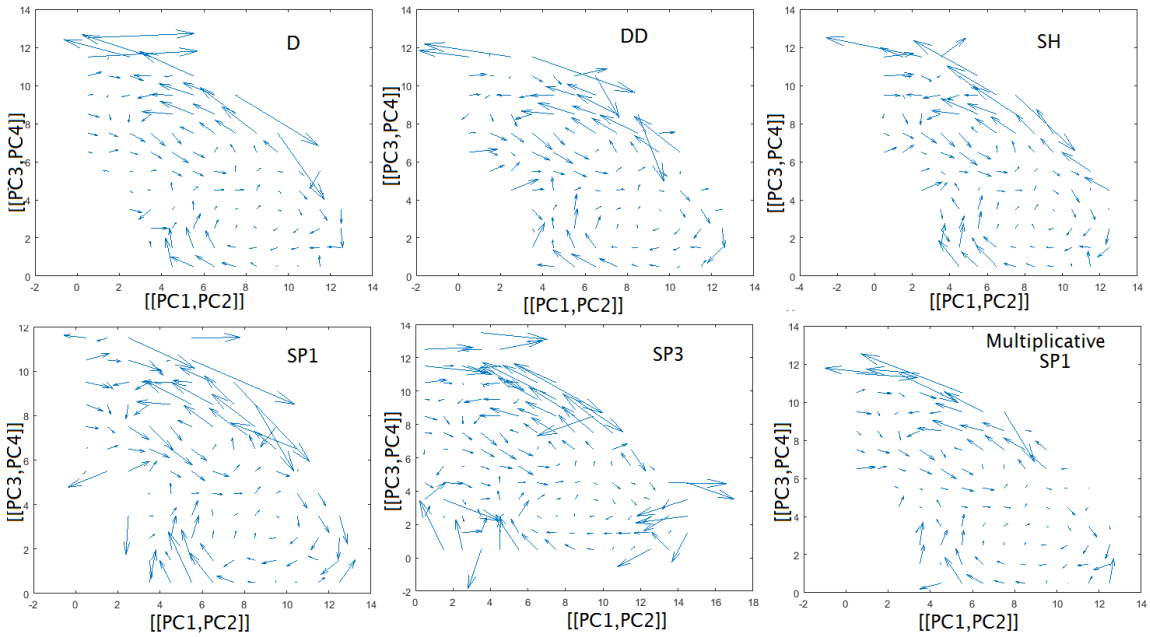

Figure 2.13: Vector plots for selected models (labelled) of the $F = 10$ setup under additive stochastic parameterisation and model SP1 under the multiplicative stochastic scheme. Plots for all other multiplicative and deterministic models were very similar to the additive scheme equivalents and are not shown. The regimes are visible but not as dominant as the 'true' regimes in all cases shown except SP3. In the additive SP1 model, Regime B is beginning to break down: there are large arrows contrary to the anticlockwise motion around the regime centre. The contrary arrows are much smaller for the multiplicative SP1 model, but larger for the additive model SP3 where regime behaviour breaks down more completely.

Nearly all the models failed to capture the small Regime B peak visible in the 'truth' pdf for $F = 10$, or merged it with the much more dominant peak for Regime A, with the exceptions of model types D and SH under additive stochastic parameterisation and model type SP1 under

multiplicative stochastic parameterisation; all of the plots pertaining to these are shown. It is possible that these cases possess a near-optimal level of noise, contributed to by rounding errors, bit flips and stochastic parameterisation schemes, which allows the models to explore a sufficient breadth of phase space to replicate both regimes of the system. Increasing this noise further (as in SP2-5 and SP1 in the additive stochastic case) may mask the regimes because random perturbations carry the system away from the true attractor too often for its structure to be accurately reproduced. For all parameterisation schemes model SP3 reproduced regimes much less well than did SP1, and SP5 performed even worse. But using Mixed Precision (model type SH) did not significantly alter the pdfs relative to the double-precision control (type DD), whilst lowering the resolution (type D) degraded the accuracy.

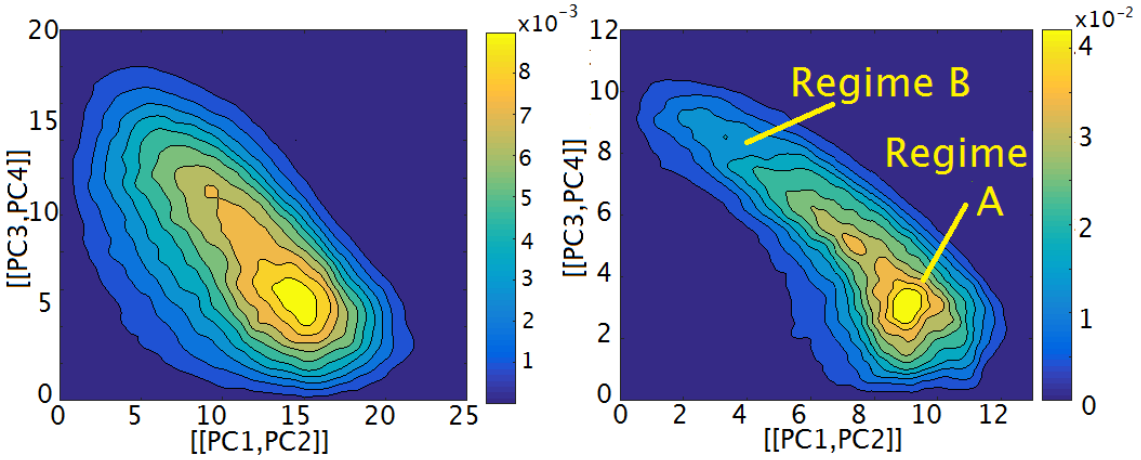

Figure 2.14: Principal Component pdfs for the $F = 20$ (left) and $F = 10$ (right) setups' truth timeseries. Two regimes are clearly distinguishable only in the less chaotic $F = 10$ case.

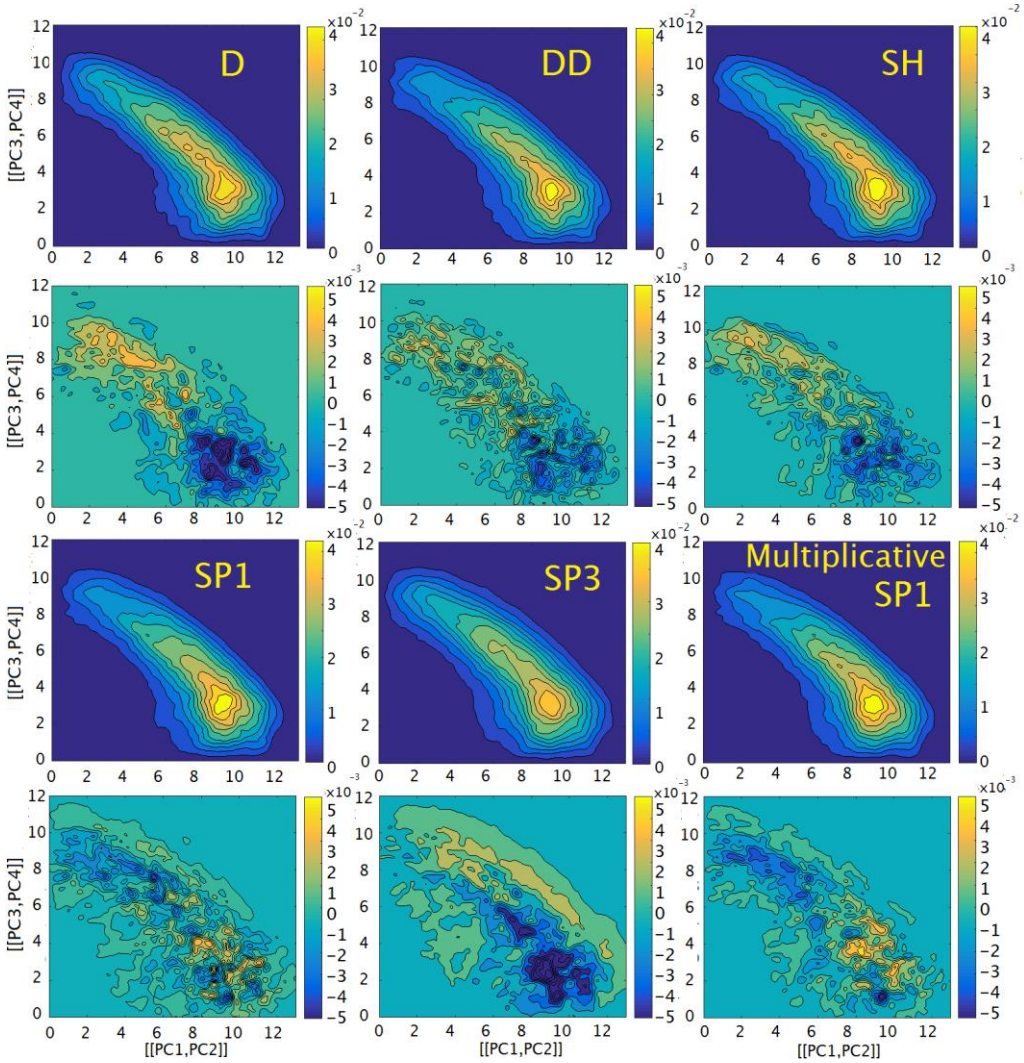

Figure 2.15: Principal Component pdfs are plotted for selected models (labelled) of the $F = 10$ setup under additive stochastic and model SP1 under multiplicative stochastic parameterisation, the same cases as in Figure 2.13. Beneath each plot is shown the difference from the ‘truth’ pdf (shown in Figure 2.14, right).

2.6. Conclusions from Modified Lorenz ’96

By extending the Lorenz ’96 system to three tiers of variables, it was possible to analyse the effects of reducing precision in a scale-selective way. The three-tier system exhibits the chaoticity and non-linear interactions between scales that are characteristic of the real Earth atmosphere, but its low complexity and the possibility of defining the ‘truth’ exactly meant that the accuracy of cheap Low Resolution models and expensive High Resolution models in double-precision, and cheap High Resolution models in emulated reduced precision could be straightforwardly compared. In

this way, it was found that by trading numerical precision for resolution, better forecasts of the true state of the system could be made for similar computational costs.

Scale-selective Mixed Precision demonstrated considerable promise. For short-term forecasts, reducing the X-variables to single-precision and the Y-variables to half-precision (model type SH) had no significant effect on the accuracy when additive stochastic parameterisation was used, even though the estimated computational cost was reduced from 9 times (for the control) to 2.5 times (for SH) that of the Low Resolution model. A loss of accuracy for this model type under multiplicative stochastic parameterisation could be remedied by keeping the model in higher precision to begin with and switching to half-precision only after a small delay, retaining most of the computational cost saving. Computational constraints mean that the resolution of operational weather forecast models is often decreased at long lead-times at the expense of increased error (Buizza 2010); perhaps reducing precision instead would improve medium-term forecasts.

For forecasts of long-term mean conditions, atmospheric chaoticity and regime behaviour, Mixed Precision models (SH) were as proficient as the double-precision controls (DD). KS Statistics and Hellinger Distances were consistent for DD and SH and their Error Doubling Times sometimes exactly matched the 'truth', which the low resolution alternative (type D) underestimated by up to 20%. The SH additive forecast exceeded the ability of DD to reproduce atmospheric regimes. At no point did the Mixed Precision models crash, so the maximum half-precision exponent cannot have been exceeded anywhere even in these long runs. Overall, the results suggest that resolving atmospheric variables at scales that are currently parameterised or close to the truncation scale in half-precision and at larger scales in single-precision would improve forecast accuracy for a given computational cost.

Stochastic Processor model types SP1 and 2, where the large-scale variables are protected from bit-flips, performed well across all metrics, but they were estimated to be around twice as expensive as SH (see Table 2.1). Introducing low-probability bit-flips in the X-variables (types SP3-5) greatly reduced the accuracy for no cost saving, demonstrating the importance of a multi-scale approach. SP2 and SP4, which differ from SP1 and SP3 only in having a larger probability of

bit-flips in the Y-variables, occasionally crashed in the longer runs, showing that the bit-flip probability must be kept small even for variables close to the truncation scale.

These findings demonstrate the potential benefits that could be attained were single-precision and even half-precision to be more widely used in weather and climate forecasting. However, the applicability of results obtained in this idealised system to full-complexity models is limited. These models contain many more variables and may therefore be more likely to crash in reduced precision, which should hence be applied with care. Some of these variables may exhibit larger dynamic ranges than others, and would be thus more likely to exhibit faults with a reduced precision exponent; a successful forecast may be more dependent on the exact value of some variables than of others. Furthermore, the real atmosphere does not, of course, possess three clearly defined spatial scales of phenomena, and the observed continual increase in chaoticity towards smaller spatial scales that fluids in the real universe exhibit may make it more appropriate to reduce precision in a continuous fashion towards smaller scales in a real-world model. To investigate such a scheme, it is necessary to emulate reduced precision in a more sophisticated model that more closely resembles the real atmosphere, with a much wider range of spatial scales.

3. Exploring the Nature of the Surface Quasi-Geostrophic System

Before investigating the implementation, and potential effects, of scale-selective reduced precision in a full-complexity atmospheric model, it is first necessary to explore an intermediary system that is more realistic than the Lorenz '96 environment but retains a relative simplicity and tractability so that detailed tests can be carried out without incurring large computational costs. This chapter will describe the properties of the Surface Quasi-Geostrophic (SQG) system (Section 3.1), an idealised atmosphere based on the same equations that describe the physics of the real atmosphere but with a number of important simplifications, and outline the particular formulation of this system used in this thesis (Section 3.2). Before applying reduced precision in this context, further experiments were carried out to probe in further detail the nature of the SQG system, most importantly the energy and enstrophy scaling that it exhibits, the results of which are also presented here (Section 3.3). These results allow speculations to be made regarding the relevance of this system to more advanced forecasting models (Section 3.4), and to some extent can be used to justify the results of the reduced precision experiments that will be described in the next chapter.

3.1. Surface Quasi-Geostrophic Theory

Although strictly two-dimensional, and hence relatively computationally cheap to solve, the Surface Quasi-Geostrophic (SQG) equations (Blumen 1978; Pierrehumbert et al. 1994; Held et al. 1995) describe the flow of an idealised rapidly-rotating fluid on a surface that can be extrapolated to predict the dynamics throughout the whole of a three-dimensional fluid. Hence, a numerical system based on these equations may replicate properties of the real atmosphere, such as the $E \propto k^{-5/3}$ relation between energy and wavenumber. Other 2D systems such as the 2D Euler model display quite different behaviour, such as an $E \propto k^{-3}$ spectrum (Palmer et al. 2014). Unlike

2D Euler, the SQG Equations can emulate frontogenesis in the real atmosphere (Constantin et al. 1994); indeed, the dynamics of the ocean surface (Le Traon et al. 2008) and the tropopause (Held et al. 1995) can be more accurately described through SQG than by using the more general and less tractable Quasi Geostrophic (QG) equations.

SQG theory is rooted in the Navier-Stokes equations that are used to describe flow in real-world weather and climate models, which represent a balance between pressure gradient $\vec{\nabla}p$, Coriolis, gravitational and viscous drag forces acting on any given parcel of fluid, as described in Chapter 1. Where $\vec{g} = g\hat{z}$ is the acceleration due to gravity, $\vec{u} = u\hat{x} + v\hat{y} + w\hat{z}$ is the velocity of the fluid particle, $\vec{F_v}$ is the viscous drag force per unit mass and ρ is the fluid density, the Navier-Stokes Equations in reference frame rotating with angular velocity $\vec{\Omega} = \Omega\hat{z}$ are,

$$\frac{d\vec{u}}{dt} + (\vec{u} \cdot \vec{\nabla})\vec{u} = -\frac{\vec{\nabla}p}{\rho} + \vec{g} + \vec{F_v} - 2\vec{\Omega} \times \vec{u}. \tag{3.1}$$

In this general form, these non-linear equations are not a very efficient way of representing atmospheric dynamics in a computer simulation. However, in the real atmosphere it is not unrealistic to assume that the viscous drag term is close to zero, and that two convenient force-balance assumptions apply. Hydrostatic balance implies that gravity is balanced by the vertical component of the pressure gradient, so that the parcel is falling at terminal speed with no net vertical force, whilst geostrophic balance implies that the Coriolis and horizontal pressure gradient forces balance, leaving no net horizontal force. These two assumptions give rise to the equations,

$$\frac{1}{\rho}\frac{dp}{dz} = g, \qquad \frac{1}{\rho}\left(\frac{\partial p}{\partial x}\hat{x} + \frac{\partial p}{\partial y}\hat{y}\right) = -2\vec{\Omega} \times \vec{u}. \tag{3.2}$$

The two forms of balance lead to a system that will tend towards stasis, described by,

$$\frac{d\vec{u}}{dt} + (\vec{u} \cdot \vec{\nabla})\vec{u} = 0. \tag{3.3}$$

Quasi-Geostrophic (QG) theory describes departures from perfect geostrophic balance in a rapidly rotating, stably stratified fluid – hence the term ‘quasi’-geostrophic (Charney and Eliassen 1949; Held et al. 1995). This theory allows for a non-zero vorticity ζ, a measure of the local rotation of the flow; and buoyancy Θ, which describes the upward force opposing gravity on the fluid. In essence, this means that there are local net horizontal and vertical accelerations, departing

from hydrostatic and geostrophic balance. These quantities are most easily expressed in terms of the horizontal streamfunction ψ, which is related to the velocity of the flow through,

$$(u, v) = \left(-\frac{\partial \psi}{\partial y}, \frac{\partial \psi}{\partial x}\right). \tag{3.4}$$

The vorticity and buoyancy are then defined by the equations,

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2}, \qquad \Theta = f\frac{\partial \psi}{\partial z}. \tag{3.5}$$

Under the Boussinesq approximation (neglecting density fluctuations), QG theory predicts that these quantities evolve according to (Held et al. 1995),

$$\frac{d\zeta}{dt} = \frac{\partial \zeta}{\partial x}\frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x}\frac{\partial \zeta}{\partial y} + f\frac{\partial w}{\partial z}, \tag{3.6}$$

$$\frac{d\Theta}{dt} = \frac{\partial \Theta}{\partial x}\frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x}\frac{\partial \Theta}{\partial y} - N^2 w. \tag{3.7}$$

Here, N is the buoyancy frequency of a reference state and $f = 2\Omega \sin(\phi)$ is the Coriolis parameter at the fluid parcel's latitude ϕ, which can be assumed to be fixed if the area over which calculations are being carried out is approximately flat; this 'f-plane' assumption is accurate at mid-latitudes in the real atmosphere.

Further simplifications can be applied as outlined in Held et al (1995). Equation 3.6 can be conveniently rewritten in terms of the potential vorticity, which, unlike vorticity, is conserved except where heat exchange with the surroundings or frictional drag takes place. This is defined as,

$$q = \frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial y^2} + \frac{\partial}{\partial z}\left(\epsilon \frac{\partial \psi}{\partial z}\right) \tag{3.8}$$

where if $\epsilon = (f/N)^2$. If f and N are constant the vertical coordinate z can be rescaled to Nz/f, in which case $q = \nabla^2 \psi$. Eliminating the vertical flow, in terms of q Equation 3.6 becomes,

$$\frac{dq}{dt} = \frac{\partial q}{\partial x}\frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x}\frac{\partial q}{\partial y}, \tag{3.9}$$

Meanwhile, Equation 3.7 can also be simplified, in the case that there is zero vertical velocity. Then, $N^2 w = 0$ and the evolution of the buoyancy is described entirely by the remaining terms, which represent advection by the geostrophic wind – transport that conserves fundamental properties of the fluid parcel such as temperature and density. Hence,

$$\frac{d\Theta}{dt} = \frac{\partial \Theta}{\partial x}\frac{\partial \psi}{\partial y} - \frac{\partial \psi}{\partial x}\frac{\partial \Theta}{\partial y}. \tag{3.10}$$

In other words, the vorticity and buoyancy evolve equivalently and either of these could be used to drive a QG model. The Surface Quasi-Geostrophic (SQG) system is a special case of a QG system for which the potential vorticity is assumed to be constant, $q = q_0$. If q is constant at the surface, the potential vorticity equation (3.8) can be decomposed into horizontal and vertical components, and applying Fourier decomposition (Tulloch and Smith 2006) and the assumption that for $z > 0$ the solution decays with z (Held et al. 1995), one obtains,

$$\psi(k, z) = \psi(k, 0)\frac{\epsilon}{k}\exp(kz/\epsilon). \tag{3.11}$$

This means that the surface ($z = 0$) potential temperature dictates the streamfunction at all heights and the SQG system uses a single vertical layer ($z = 0$) of the stably stratified fluid, with $w = 0$. The buoyancy and streamfunction at the surface can be used to describe the whole three-dimensional fluid flow, and are related through (Majda and Bertozzi 2002),

$$\psi = -\int \frac{1}{|y|}\Theta(x + y, t)dy. \tag{3.12}$$

3.2. Establishing an SQG System Setup

To convert the SQG theory into a working model, the SQG equations must be written as numerical code to be solved on a computer. In this investigation, a Fortran SQG program was created based on code originally written by Shafer Smith (2009), which was extensively adapted to create appropriate setups with which to carry out experiments. To be able to run many reduced precision experiments and, at the same time, to obtain results that are as pertinent as possible to real-world models, it is desirable to maximise the model's resemblance to the real atmosphere at the same time as minimising the complexity and computational cost. A willingness to balance these two requirements is what motivated the introduction of the realistic topography, hyperviscous damping and wind-like forcing that will be utilised here, as well as the optimisation of key numerical parameters to create two different setups, 'Setup L' and 'Setup S', suited to long- and short-term simulations respectively as described below.

Once these setups were established, the program could be adapted to carry out two sets of experiments, the aim of which was firstly to empirically establish the effect on model accuracy of reducing precision to different degrees on different spatial scales, and secondly to establish a theoretical justification for scale-selective precision through an analysis of the degree of turbulence exhibited by the model at different spatial scales. The methods used and results obtained in this latter investigation will be outlined later in this chapter, and will form a background upon which to view the results of the reduced precision experiments described in Chapter 4.

The numerical version of the SQG system applied here is mostly solved in spectral space, meaning that the values of the prognostic variables are not computed separately at different grid-points. Instead, the variables are decomposed into distinct wavelengths, characterised by the wavenumber k, which describe the quantities' value at each of a series of spatial scales. Large features on the model's domain have small wavenumbers, whilst smaller features have large wavenumbers. Most of the model's calculations are carried out in this spectral space, where the scale-dependency of different properties can be explicitly measured. However, whilst the left-hand-side of the driving equation (Equation 3.10) is stepped forward in spectral space, the right-hand-side must be evaluated in grid-point space because the spatial differentials are difficult to evaluate in spectral space. This means that, at every timestep, the model uses a Fast Fourier Transform (FFT) algorithm to change between spectral and grid-point representations of the variables. All the variables in the SQG code are dimensionless, and can be related to real-world scales only by considering the size of the weather phenomena they simulate.

3.2.1. Replicating Previous Studies

Before adding additional features to the model code to create setups optimised for this investigation, it was necessary to verify that the SQG behaviour observed in previous experiments could be replicated. A number of SQG test cases were presented by Held et al. (1995), including one in which an otherwise flat topography was interrupted by an elliptical vortex in the centre of

the field, which the authors described as an 'isolated mountain' (topography will be discussed in detail in Section 3.2.2). In their experiment, a filament of the flow was observed to fragment into smaller vortices due to flow around this feature driven by a uniform horizontal wind. A similar flow was observed in tests of the code to be used in this investigation, under a similar setup to that of Held et al., which implies that the model used here simulates the same SQG dynamics. The vorticity observed after 200000 (dimensionless) timesteps is shown in Figure 3.1 alongside the equivalent plot from Held et al. (1995), illustrating the same vortex-shedding phenomenon.

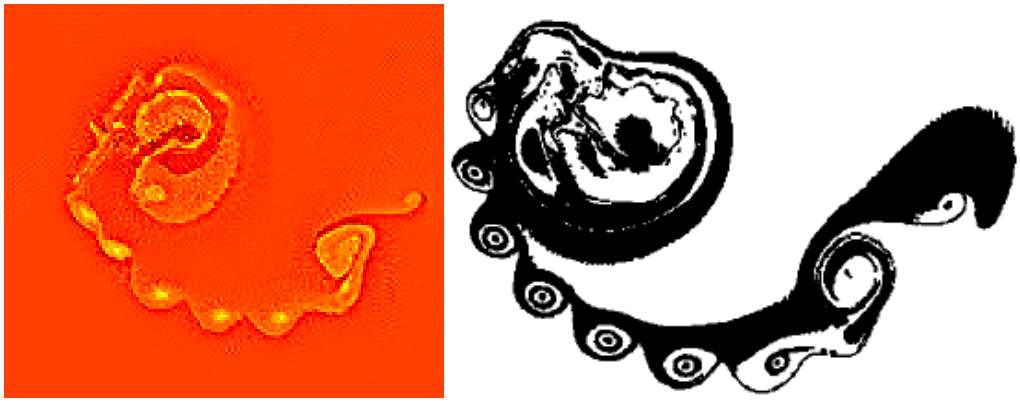

Figure 3.1: The vorticity field after 200000 timesteps obtained using Held et al.'s setup in the SQG code used in this investigation with $k_{\max} = 255$ resolution (left); and the surface potential temperature (which behaves equivalently to vorticity) plotted by Held et al. (1995) for the same setup (right).

3.2.2. Bottom Topography

The quasi-three-dimensional character of the SQG equations means that they can be used to describe flow over a non-flat bottom topography to simulate the effect on the atmosphere of, for instance, a mountain range. Introducing this topography can be achieved by using a height term $h(x, y)$ in Equation 3.9 (Held et al. 1995), so that,

$$\frac{d\Theta}{dt} = \frac{d\psi}{dy}\frac{d}{dx}(\Theta + N^2 h) - \frac{d\psi}{dx}\frac{d}{dy}(\Theta + N^2 h). \tag{3.13}$$

The function $h(x, y)$ can, in general, take any form. Since the aim here is to simulate a situation resembling as closely as possible the weather over Earth's surface, it is preferable to devise an Earth-like topography. Quasi-random spatial oscillations resembling natural undulations can be produced using a fractal Brownian motion technique previously used for computer simulations of

imaginary mountainous regions (Fournier et al. 1982). Here, 'red' noise (where the spectral energy density is inversely proportional to the frequency) is simulated by a 'midpoint displacement' algorithm. The algorithm begins with sixteen grid-points and works sequentially down to smaller and smaller scales. Unknown values mid-way between known values are recursively calculated by averaging the two or (depending on the position on the grid) four surrounding points then randomly offsetting by an amount proportional to the depth of the recursion. If the range of random variability is halved at each step, the resulting stochastic terrain resembles red noise (Olsen 2004).

The mid-point displacement algorithm produces a topography that undulates with roughly equal amplitude across the whole domain of the model. To simulate a mountain on Earth, which would penetrate deep into the atmosphere and thereby potentially produce interesting blocking or channelling alterations to the flow, it is necessary to amplify regions of the field artificially. This produces a 'mountainous' topographical field such as that shown in Figure 3.2.

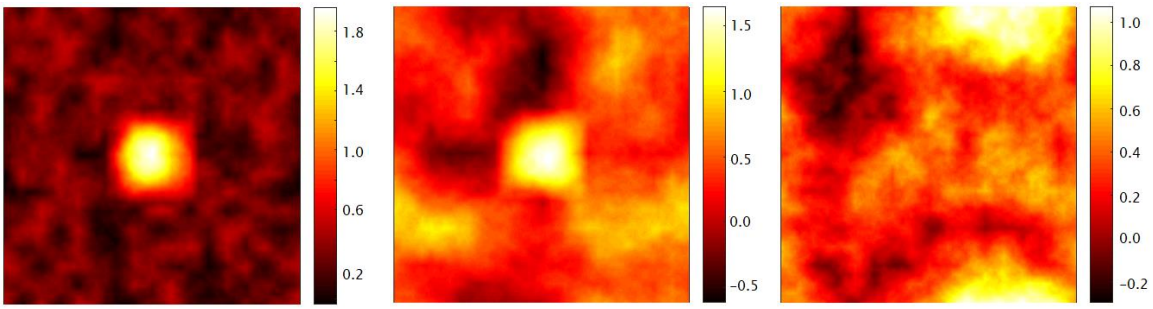

Figure 3.2: Scale-selective bottom topography of amplitude 1.0 plotted with a mountain of height 0.75 above the background topography (left), one of height 0.5 (centre) and with no mountain (right) in the centre of the field, for runs of resolution $k_{\max} = 127$. The latter two cases correspond to Setups S and L respectively.

3.2.3. Forcing and Damping

In order to replicate the behaviour of the real atmosphere, it is necessary for the streamfunction ψ – which is the prognostic variable – to evolve chaotically over time. Some randomness is provided by the fractal nature of the topography, but this alone is not sufficient to ensure chaotic, non-linear flow. One method of obtaining such flow is to introduce a random forcing term. Because in the real atmosphere energy is added at large scales and removed at small scales, it is justifiable to run

experiments applying forcing at large-scale wavenumbers between $k = 6$ and $k = 8$, following Pierrehumbert et al (1994). For these wavenumbers, the forcing of magnitude f is added to the right hand side of Equation (3.10). Thus,

$$\frac{d\Theta}{dt} = -v \cdot \nabla\Theta + f\sqrt{1-\lambda^2}\sum_{k=6}^{8} e^{2\pi i r_k} - d, \tag{3.14}$$

where d is a drag term which recycles energy from the small spatial scales back to the larger scales, r_k is a random number called separately for each wavenumber and λ is the forcing correlation. Similarly to Held et al (1995) a constant mean wind can be introduced alongside this.

An alternative means of simulating chaotic flow is to introduce shear in the form of a background wind field applied directly to the right hand side in Equation (3.10), replacing the constant f in Equation (3.14) with some function $u(x, y)$. This was the approach taken in the present study, where a sinusoidal zonal wind was introduced so that $u(x, y) = \sin(y)$, where y is the normalised latitudinal distance from the 'equator', pushing the flow 'eastwards' (i.e. from left to right) in the system's northern hemisphere and 'westwards' in its southern hemisphere, and sloping to zero at the centre ('equator') and at the periodic top and bottom boundaries.

Whichever forcing is applied, hyper-viscous drag is necessary to remove excess energy; this is akin to viscous dissipation at small spatial scales in the real atmosphere. In the SQG system, the hyperviscous damping magnitude, d, can be adjusted to ensure that energy does not become too strongly concentrated at high wavenumbers over time. To implement this, a filter is applied to the prognostic variable that reduces the large wavenumbers' magnitude but leaves the small wavenumbers almost unchanged. If $\Theta(k)$ is the buoyancy at wavenumber k, it is reduced once at every timestep according to,

$$\Theta(k) = \Theta(k)\left[1 + d\left(\frac{4\pi k}{k_{\max}}\right)^8\right]^{-1}. \tag{3.15}$$

This is referred to as 'hyperviscosity' because of its stronger dependence on wavenumber than ordinary viscosity. Ordinary viscosity would be proportional to k^2; any damping term proportional to k^p where $p > 2$ is 'hyperviscous'. The higher the value of p, the more strongly biased towards small spatial scales the filter becomes. Here, $p = 8$ was chosen on the basis of empirical findings that in most real-world models $p = 8$ or $p = 16$ tends to remove sufficient energy at small scales

to keep the model stable without much affecting the large scales (Smith and Hammett 1997). When the system is re-run at different resolutions, the hyperviscosity strength must be altered in proportion to the resolution to prevent higher-resolution cases from becoming too heavily damped, hence the inclusion of the $k_{\max}$ term. One caveat is that hyperviscous damping is prone to generating 'ripples' on small scales, potentially introducing small-scale inaccuracies. Because of the rapid downscale propagation of errors in the SQG system, however, this is not expected to affect the results of the SQG experiments in this thesis.

3.2.4. Finalising the Setups

Experiments were conducted in which a 'mountain' was seeded by increasing the initial topographical height with a sinusoidal function, to at most three times the maximum height of the surrounding topography, in the centre of the field, so as to observe the effect of this mountain on the streamfunction flow. The topographic height and wind forcing magnitude are the most influential parameters in the SQG system, and since it is non-dimensional it is their relative magnitude that is important. With a sinusoidal wind forcing of 0.5, when the topographical height was set to 1.0 with a mountain of additional height 0.5 the topography was sufficiently prominent for the streamfunction to flow around the mountain, rather than passing over it with minimal disturbance. Increasing the topographical height to more than this would cause the topography to become so dominant that it prevented flow entirely.

However, including the mountain in the topographical forcing meant that the energy of the system was observed to grow gradually over time for millions of timesteps, which is not realistic, presumably because of the additional topographic forcing that this entails. This should not be problematic for short-term forecasts, as the change is slight and would not be expected to change the behaviour of 'weather' in the flow field, but it may cause problems with longer-term forecasts, where the average conditions might drift over time, complicating the calculation of long-term climate mean properties. Hence, it was decided to use two different setups: 'Setup S', with scale-

selective topography including a mountain of height 0.5, with which to conduct experiments on short timescales; and 'Setup L', with no mountain, wind forcing of size 0.02 and the energy settling to a fixed level more quickly, for longer-term simulations. Figure 3.3 shows the total energy as a function of time for Setup S and Setup L on the timescales at which they are used in this paper, spinning-up from zero at zero timesteps.

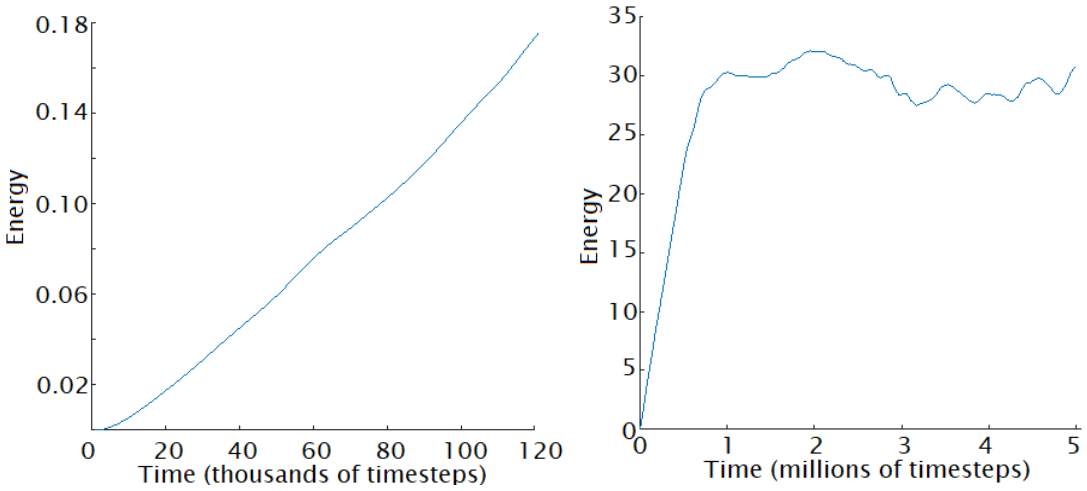

Figure 3.3: Total energy is plotted as a function of time for a single run of Setup S for 120000 timesteps (left) and one of Setup L for 5 million timesteps (right); these timescales reflect those on which the two setups are employed in the experiments presented in this paper. The total energy does not stop growing in Setup S, even when that setup is run for a comparable length of time to Setup L.

Figure 3.4 shows the streamfunction and vorticity spectra exhibited by Setup S between 50000 and 120000 timesteps, which is the range of timesteps within which most of the experiments presented here will fall. The streamfunction spectrum shows that its magnitude varies between approximately 10^{-3} and 10^{-1}, which is well within the range representable by half-precision, something that will become important in the reduced precision experiments presented in the next chapter. The streamfunction follows approximately (within experimental error) a $k^{-5/3}$ power law with wavenumber similar to that observed for the kinetic energy as will be shown in Section 3.3.1. The vorticity, on the other hand, does not display a single spectral slope; at moderate wavenumbers

it obeys approximately a k^{-1} power law, whilst at very high wavenumbers it exhibits a $k^{-10/3}$ power law; for most of the spectrum, the vorticity values lie between 0 and 10.

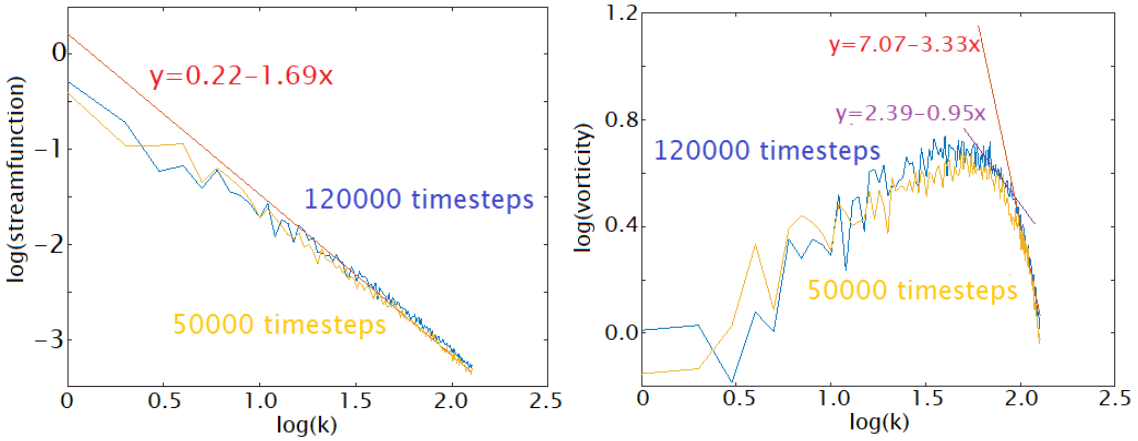

Figure 3.4: Base-10 logarithmic spectra are plotted for the streamfunction (left) and vorticity (right) for setup S spun up from a flat (zero) initial field for 50000 (yellow) and 120000 (blue) timesteps. Best-fit straight lines are shown for both cases; in the vorticity case, separate fits are made for separate parts of the curve, and the equations of these lines are colour-coded to match the lines to which they correspond.

3.3. Turbulence and Energy Scaling in the SQG System

It is important to be able to replicate the turbulent energy cascade observed in the real atmosphere in order to produce accurate forecasts. Turbulence governs the propagation of energy between different spatial scales, which is what allows small initial errors in atmospheric models to grow rapidly. This is of paramount importance to scale-selective reduced precision experiments because the propagation of errors upwards from small scales could lead those scales to have a larger impact on the overall predictability than they otherwise would. This section explores the turbulence exhibited by SQG and its implications, both in theory and as observed in experiments.

3.3.1. Turbulence in the SQG System

Turbulence, whether exhibited in the real atmosphere or in an atmospheric model, can be characterised by plotting the energy and enstrophy power spectra as a function of wavenumber k,

which describe how these quantities are distributed between the large scales (small values of k) and small scales (large values of k). In the real atmosphere, observations suggest that there is an energy cascade at scales below about 400km (depending on height in the atmosphere), and that the energy spectrum there has a spectral slope of -5/3, so that $E \propto k^{-5/3}$ where E is the energy per unit wavenumber (Nastrom and Gage 1985). General circulation models with a -5/3 energy spectrum display a similar downscale energy cascade (Augier and Lindborg 2012). This is referred to as 'three-dimensional (3D) turbulence' as it is thought to be associated with three-dimensional interactions present in the flow at small scales. At larger scales than this, it is observed that $E \propto k^{-3}$, a spectral slope that implies an 'inverse energy cascade' with energy moving to larger scales and which some have attributed to the essentially two-dimensional nature of large-scale atmospheric flow; hence, such a slope is associated with 'two-dimensional (2D) turbulence' (Rotunno and Snyder 2008). At these large scales, the energy inverse-cascade is dominated by a downscale (forward) cascade of enstrophy G.

The SQG system is one of a family of two-dimensional models based on the QG equations that exhibit 2D or 3D turbulence, as measured by the steepness of the spectra. The spectral slope can be altered by changing the coefficient α in the equation that relates ψ and Θ in spectral space (equivalent to grid-point space Equation 3.12 when $\alpha = 1$), which is,

$$\psi(\vec{k}) = -|\vec{k}|^{-\alpha}\Theta(\vec{k}). \tag{3.16}$$

The 2D Euler and SQG systems have $\alpha = 2$ and $\alpha = 1$ respectively. In the 2D Euler case, this leads to non-local interactions between structures at different spatial scales and Θ is equivalent to the potential vorticity. This system exhibits turbulence that is truly two-dimensional in that energy 'inverse-cascades' from small scales to large scales at small wavenumbers (large spatial scales), the 'energy-cascading range', although enstrophy cascades from large to small scales at large wavenumbers (small spatial scales), the 'enstrophy-cascading range'. In the $\alpha = 1$ SQG case, meanwhile, local interactions between similar spatial scales dominate (Pierrehumbert et al. 1994). The potential vorticity is then zero and Θ describes the temperature field on the lower boundary. Although the turbulence is still strictly two-dimensional (because the model is two-dimensional),

the expected energy and enstrophy spectral slopes in the SQG system are akin to those associated with two- or three-dimensional turbulence in the real atmosphere in different parts of the spectra. These spectra can only be seen clearly in high-resolution runs (Capet et al. 2008), so a maximum wavenumber of at least $k_{\max} = 127$ was used throughout this investigation.

In a well-balanced SQG system there are two conserved quantities: the total energy $E = \langle \psi\Theta \rangle/2$, where $\langle\ \rangle$ denotes an average across the surface ($z = 0$), and the temperature variance or enstrophy $G = \langle \Theta^2 \rangle/2$. Here, E and G are the energy and enstrophy per unit wavenumber, as it is these quantities that can be directly measured in practice in the real atmosphere. The total energy $\overline{E}$ and enstrophy $\overline{G}$ can be computed from these by integrating over all wavenumbers, so that, for domain area A,

$$\overline{E} = \int E(k)dk = -\frac{1}{A}\iint \psi\Theta\, dx\, dy, \tag{3.17}$$

$$\overline{G} = \int G(k)dk = \frac{1}{A}\iint \Theta^2\, dx\, dy. \tag{3.18}$$

At low wavenumbers (large scales), SQG exhibits an energy cascade for which the energy and enstrophy spectra approximately obey $E \propto k^{-2}$ and $G \propto k^{-1}$, which is slightly less steep than the real atmosphere. At larger wavenumbers (smaller spatial scales), just as in the real atmosphere the spectral slope changes, but in SQG the spectra follow $E \propto k^{-8/3}$ and $G \propto k^{-5/3}$ (Pierrehumbert et al. 1994), so that the enstrophy cascade is similar to the energy cascade in the real atmosphere but there is also an energy cascade that is slightly steeper than in the real atmosphere but is nonetheless downscale. Because of this similarity, the SQG model is often considered a better analogue for turbulence in the real atmosphere than are other QG models. The large-scale $E \propto k^{-2}$ and small-scale $G \propto k^{-5/3}$ parts of the spectrum correspond to the 'energy cascade' and 'enstrophy cascade' regions respectively in the SQG model.

To verify that the setups used here are likely to exhibit the type of turbulence expected in SQG, an attempt was made to replicate the results of Pierrehumbert et al. (1994) by plotting the enstrophy and energy density spectra after 'spinning up' the model for 50000 timesteps with the hyperviscous filter switched on and scale-selective topography applied. Those authors used a flat topography with forcing applied to wavenumbers between $k = 6$ and $k = 8$ and observed an

enstrophy density slope in the energy cascading region close to the theoretically predicted value of -5/3. Using similar forcing to theirs at the same resolution ($k_{\max} = 255$), similar results were obtained, shown in Figure 3.5, in accordance with the theoretical predictions. The transition between enstrophy and energy cascade regions occurs at $k \approx 10$, which implies that in this setup $\epsilon/H \approx 10$ (Tulloch and Smith 2006), where H is the characteristic height scale of the system and $\epsilon = (f/N)^2$ (see Section 3.1).

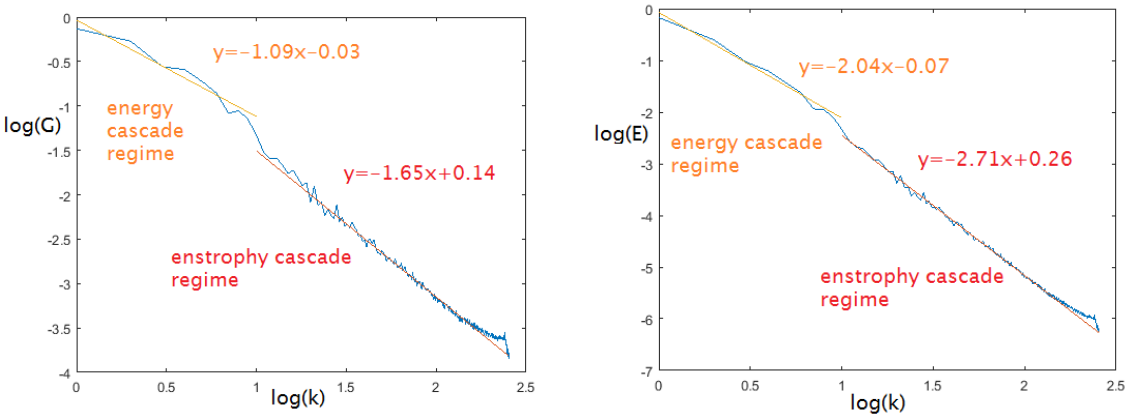

Figure 3.5: Enstrophy (left) and energy (right) density spectra obtained for an SQG ($\alpha = 1$) setup with scale-selective topography, hyperviscous filtering and large-scale forcing (between $k = 6$ and 8) reproducing approximately the results of Pierrehumbert et al. (1994). Both plots use base-10 logarithms.

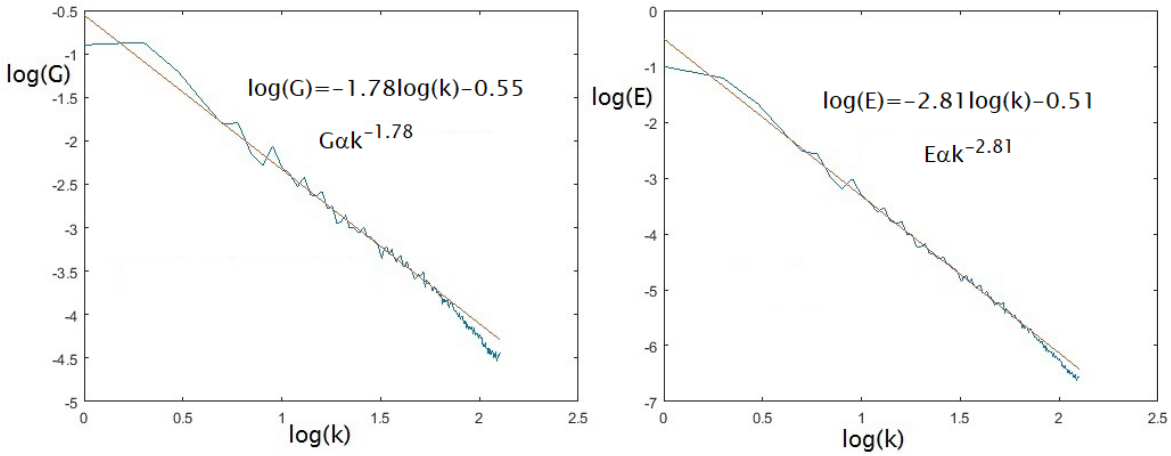

Figure 3.6: Enstrophy (left) and energy (right) density spectra for SQG Setup S, which includes sinusoidal wind forcing (rather than artificial forcing of wavenumbers 6 to 8) with scale-selective topography including a 'mountain' and hyperviscous filtering. Best-fit straight lines are plotted and the equation for each is given, alongside the relationship that this implies between the spectral enstrophy or energy and the wavenumber.

For each of the setups S and L used here, however, a sinusoidal wind forcing will be employed as described in Section 3.2.3, rather than the arguably less physically justified forcing of wavenumbers 6 to 8. Changing the forcing means that the large-scale energy cascading regime, if present, is no longer dominant over a sufficiently large range of spatial scales to be visible, possibly because the energy is no longer perfectly conserved. However, at smaller spatial scales the expected spectral slopes of $-5/3$ and $-8/3$ for the spectral enstrophy and energy respectively are again approximately observed, as shown in Figure 3.6 for Setup S, for which a mountain has also been introduced in the centre of the field (Setup L, identical except for the mountain, does not differ considerably in spectral slope). Hence, these setups of the SQG system indeed appear to emulate the three-dimensional turbulence of the real atmosphere. That the low-wavenumber regime is not observed with these setups implies that the characteristic height scale H is very large (it can be treated as infinite), and that both the transition wavenumber between the two regimes and the Rossby number UL/H for characteristic velocity and length scales U and L tend towards zero.

3.3.2. Error Propagation in Multi-Scale Systems

In 1969 Lorenz analysed the predictability of flow in a system possessing multiple scales of motion and argued that, for such a deterministic system, an arbitrarily small initial observational error will rapidly grow in size so much that the states of the true and observed (equivalent to unperturbed and perturbed) systems will eventually differ so greatly as to appear to be entirely unrelated to each other, as though the system was non-deterministic (random) on long enough time-scales. Any such system – including, Lorenz proposed, Earth's atmosphere – is only predictable out to a finite forecast time that cannot be reduced by reducing the initial error. Where the spectrum of the cascading quantity (energy in Earth's atmosphere) follows a slope proportional to $k^{-\lambda}$, this limited predictability applies to systems for which $\lambda < 3$; for a spectrum with $\lambda \geq 3$, the range of predictability can be made arbitrarily large by making the initial error sufficiently small.

If the initial error in a system of limited predictability is confined to the smallest scales, it will gradually propagate upwards to affect, eventually, even the largest scales. Error growth is faster on smaller scales, so if all scales begin with the same relative initial error this will grow to reach its saturation level (that at which the 'observed' system is no more closely related to the 'truth' at this scale than is any random configuration of variables) at smaller scales first, then at successively larger scales. In the two-dimensional barotropic vorticity model (2DV) resolving wavenumbers from 1 to 21, Lorenz showed that if the largest scales remained predictable for around 2 weeks (mimicking real-world forecasts), the smallest scales would reach error saturation after only two or three minutes, regardless of whether the small initial error is constricted to the largest or smallest scale (Lorenz 1969).

This model is far from perfect as a representation of the real atmosphere but if these results can be extrapolated to real-world models, they imply that observational errors in features smaller than 1 km could destroy the predictability of scales exceeding 10 km in the atmosphere within a few hours: the so-called 'Butterfly Effect' (Palmer et al. 2014). A butterfly flapping its wings could by this principle set off a series of changes to the atmosphere that eventually result in a tornado, in a way that could not be predicted without observing every wing-flap and incorporating it into a near-perfect atmospheric model. On the other hand, more recent studies have suggested that the upscale spread of small-scale errors (such as the failure to observe a butterfly's wing-flap) to the larger-scales may be of less practical importance to weather forecasts than the uniform growth of observational errors on all spatial scales. This is because observational uncertainties are by no means restricted to the smallest scales, and initially large-scale errors can rapidly propagate downscale, overwhelming any initial small-scale errors, before propagating back upscale. Also, in the real atmosphere some mesoscale features are driven by larger scale phenomena (weather fronts, for example), and may therefore inherit predictability from the larger scales, halting the cascade of errors up from the smaller scales (Durran et al. 2013).

Analysing the error propagation in SQG is therefore important if we are to determine which conception of atmospheric error growth it simulates and hence assess its potential relevance – or not – to real-world atmospheric models. Such an analysis could also provide a theoretical

explanation for the effects of scale-dependent precision. If small-scale errors are found to propagate rapidly upscale, lowering the precision on smaller scales to induce an error greater than the observational and model error in those scales might be deleterious, as this error would rapidly come to affect the larger scales. If, however, large-scale error growth is dominated by initial large-scale errors with negligible cross-wavenumber propagation, or indeed errors propagate downscale and quickly obliterate the predictability of the small scales, the precision on the smallest scales might be made arbitrarily small without affecting the ability to predict the evolution of larger-scale phenomena but it might still be important to maintain high precision in the large-scale variables.

Because of its $k^{-5/3}$ power spectrum for the cascading quantity (the enstrophy in the SQG system), which is characteristic of three-dimensional turbulence and is thought to match the energy spectrum of the real atmosphere on scales below around 400 km (Nastrom and Gage 1985), the propagation of errors between different spatial scales in the SQG system is of theoretical interest and has been the subject of a number of studies (Rotunno and Snyder 2008; Durran and Gingrich 2014). That the SQG system has $\lambda = 5/3 < 3$ for the enstrophy in the enstrophy-cascade region should mean that it exhibits the behaviour predicted by Lorenz. It has been suggested that the SQG system is a better analogue for the real atmosphere than other two-dimensional models such as 2DV, which follow a k^{-3} spectrum, although the lack of baroclinic instability and convective clouds in the SQG system nevertheless limit its ability to simulate large- and meso-scale error growth. Durran and Gingrich (2014) repeated Lorenz' experiments for the SQG case, and found that large-scale errors tend to propagate downscale rapidly, causing the error in the small scales to saturate and then initiating an upscale error cascade identical to that which would occur had the errors initiated at the smallest scales. This implies that initial errors in the large scales are more important than those in smaller scales, regardless of the model dynamics (i.e. SQG behaves in qualitatively the same way as 2DV). The authors therefore warn against falsely assuming that the most significant errors initiate at the small scales (Durran and Gingrich 2014): so long as the observational error is not restricted to the smallest scales, the effect of a butterfly flapping its wings will be lost beneath that of errors propagating downscale from the lower wavenumbers.

The notion of smaller-scale errors growing more rapidly can be reconciled with that of the dominance of large-scale initial errors by recognising the difference between relative and absolute errors. One might assume that the same absolute observational error has the same effect on predictability regardless of the scale at which that error is introduced. This might be a very small relative error on large scales (a small proportion of a large number) but could be a very large relative error on small scales (a large proportion of a small number, enough to saturate those scales). If so, this would mean that the requirement for an accurate and precise representation of values is more stringent on the larger scales, as the relative error there must be much lower than it needs to be on the smallest scales to achieve the same predictability. It is likely that real-world observations cannot often meet this requirement, so that initial large-scale errors drive the evolution of errors later in the forecast (Durran and Gingrich 2014). Hence, one would expect that sufficiently high precision to record the observed state of the system exactly would be of crucial importance on the largest scales, and of diminishing importance towards smaller and smaller scales where the same relative initial error has less and less bearing on the large-scale predictability. This implies that increasing the precision of large-scale observations may be more important than increasing the resolution of the observations when it comes to improving the overall quality of forecasts.

The aim of the experiments presented in this Chapter is to check whether these findings hold in the context of the above formulation of the model, and thereby to speculate on the degree to which turbulence will render high precision superfluous on smaller spatial scales. This will be done via a somewhat different approach to that of the previous studies, which focussed on the evolution of a 'spectral density' quantity to represent the theoretically-derived 'error' at each wavenumber. In the code used here, the streamfunction ψ is the fundamental prognostic variable that is solved for, and here the error is therefore measured empirically as the difference between the 'truth' and the perturbed model ψ field in spectral space. In this way, one can track the evolution of initial errors in space and time and compute the Error Doubling Time at each spatial scale.

3.3.3. Experiment 1: Error Propagation Between Scales

For this first experiment, a deterministic 'truth' model starting with a flat, zero streamfunction field was 'spun up' for 50000 timesteps under Setup S to generate initial conditions. From this initial point, alongside a continuing 'truth' run, a model was run that was identical except for a random perturbation to the streamfunction introduced at a single spatial scale, and the growth in the streamfunction Absolute Error, defined as the grid-point averaged absolute difference between the unperturbed 'truth' and perturbed runs, was tracked across all spatial scales for a further 50000 timesteps. In this way, the spread of the error from its initial wavenumber – the 'wavenumber of perturbation' k_r – to adjacent wavenumbers could be monitored. 25 ensemble members were used, with the magnitude of the initial perturbation varying randomly each time, and the ensemble average prediction for ψ was compared to the 'truth'. This process was then repeated for many different wavenumbers of perturbation k_r between 0 and 127.

The SQG system is two-dimensional, with wavenumbers in the x- and y-directions being described by k_x- and k_y-components of the overall wavenumber 'radius' $k_r = \sqrt{k_x^2 + k_y^2}$. In this experiment, perturbations were introduced in k_x- and k_y-space, in the centre of the plane so that $k_x = k_y$ and $k_r = \sqrt{2k_x^2}$. The resulting error at each wavenumber 'radius' k_r was computed at each timestep by measuring the streamfunction difference between the perturbed and unperturbed simulations. An existing function in the SQG code was used to convert errors computed in this way in k_x- and k_y-space into a function of k_r by carrying out a 'ring integral' to add together at each k_r the error contributions of all the k_x and k_y components for which $k_x^2 + k_y^2 \approx k_r^2$.

The perturbations introduced at different spatial scales (values of k_r) all had the same relative perturbation size: the maximum magnitude of the random perturbation was 5 per cent of the magnitude of the initial streamfunction at each spatial scale, meaning that the initial value was multiplied by a random number between 0.95 and 1.05. Experiments showed that the exact size of the perturbation had no impact on the overall results, so long as the perturbation was not so large as

to cause the error in the smallest spatial scales to saturate immediately. Tests also showed that the error growth was independent of the exact location of the perturbation: introducing the perturbation at $k_x = k_r$, $k_y = 0$ instead of at $k_x = k_y$ made little difference to the results.

Figure 3.7 shows the size of the error in the streamfunction as a function of wavenumber and how this evolves over time when the perturbation is applied at relatively large scales ($k_x = 9$), medium scales ($k_x = 36$) and small scales ($k_x = 72$). In each case, three plots are shown: one at high temporal resolution for just the first 2000 timesteps; one at coarser temporal resolution for the first 20000 timesteps; and one at a still coarser temporal resolution spanning the full 50000 timesteps after spin-up. The data are plotted logarithmically to make the growth over time more easily discernible. As an alternative way of visualising the same data, they are also plotted as Hovmoller diagrams in Figure 3.8, where the colour scale indicates the magnitude of the error plotted as a function of both wavenumber and time for the highest and lowest temporal resolutions.

In all cases, the error almost instantaneously spreads to all scales, but the error propagates downscale more rapidly than it does upscale. When the initial error is in the larger scales, it rapidly spreads to small scales, after which the error at all scales continues to increase until each scale reaches saturation, which happens sooner at the smaller spatial scales than at the larger ones. In the plot in the top right of Figure 3.7, where $k_x = 9$, the error in wavenumbers larger than 25 has ceased to grow 50000 timesteps after the perturbation. Errors have ceased to grow at this point only for wavenumbers larger than 80 for $k_x = 30$ and the error is still growing at all scales for $k_x = 72$, the bottom right plot. By this time, the error is becoming increasingly concentrated at the larger spatial scales as a consequence of the smaller scales approaching saturation. Although it was not practical to run the model for many more timesteps at all wavenumbers, further tests revealed that the error at the very largest scales alone was still growing after 150000 timesteps when $k_x = 9$, as shown in Figure 3.9. The maximum error was increased by a factor of 10 relative to the 50000 timesteps case (Figure 3.7), but only for the smallest wavenumbers.

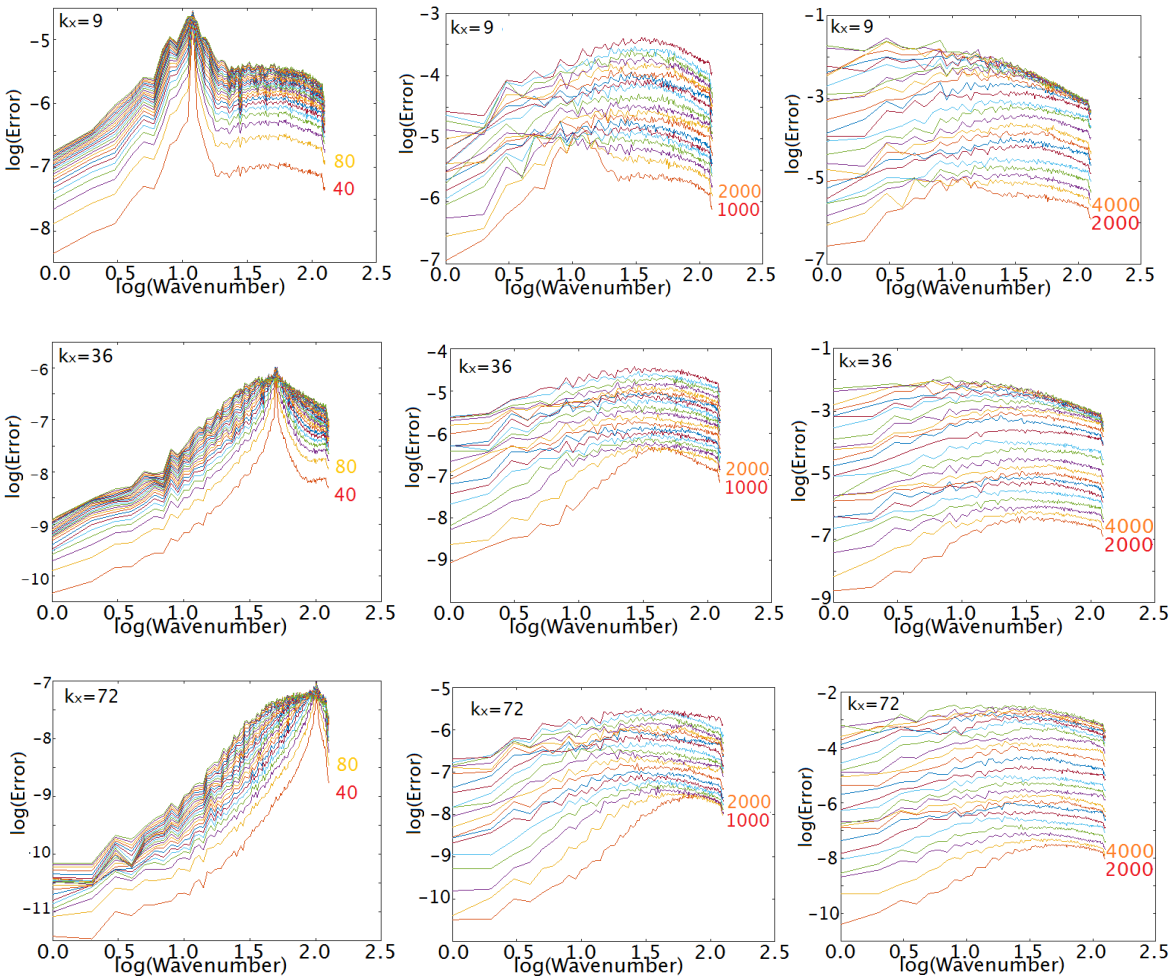

Figure 3.7: The Error in the streamfunction is plotted logarithmically as a function of wavenumber when a perturbation is introduced at large scales ($k_x = 9$, $k_r = 13$, top row), medium scales ($k_x = 30$, $k_r = 42$, middle row) and small scales ($k_x = 72$, $k_r = 102$, bottom row). The left-hand figures show the error every 40 timesteps for the first 2000 timesteps after the perturbation; the middle figures show the error every 500 timesteps for the first 20000 timesteps and the right-hand figures show the error every 1000 timesteps for the first 50000 timesteps. In each case, time increases for subsequent curves moving up the plot, with a different colour for each timestep shown, and the first two timesteps are labelled in the colour corresponding to each. All results are averaged over 25 ensemble members.

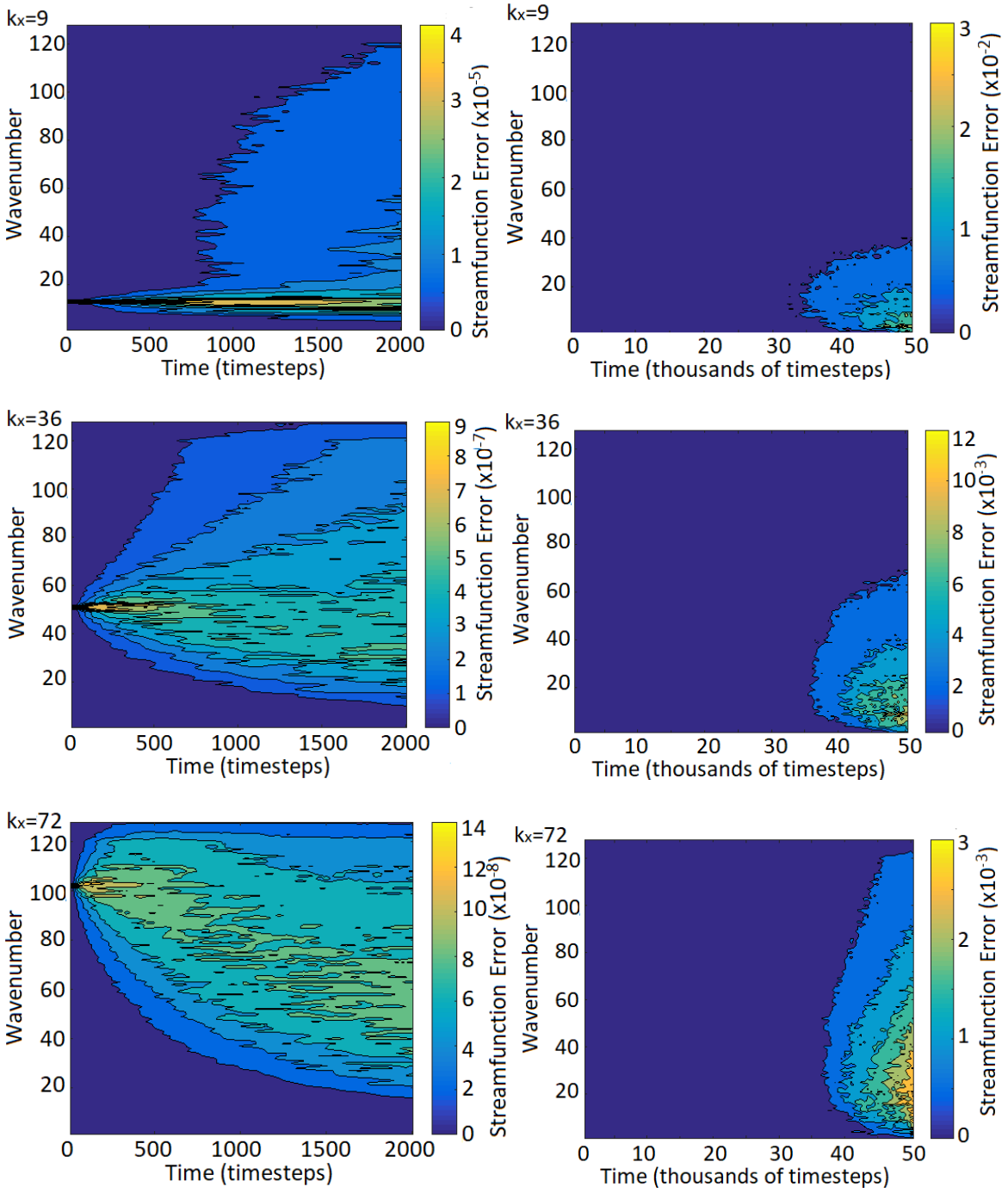

Figure 3.8: Hovmoller diagrams showing the error for the first 2000 timesteps with high temporal resolution (left column) and the error for the first 50000 timesteps for a lower temporal resolution run (right column) after a perturbation is applied, for the same three perturbations as in Figure 3.8. The colour scale on each plot indicates the extent of the error, measured as a function of time on the x-axis and wavenumber on the y-axis.

When the initial perturbation is introduced at very small scales (bottom row, Figure 3.7) it is more gradually shifted upscale, so that the error in the perturbation scale shrinks and before 2000

timesteps have elapsed the error is as great in the medium scales as it is in the smaller scales, as shown by the flattening of the curves in the bottom row of Figure 3.7 and the more even spread of error across all wavenumbers in the bottom row of Figure 3.8 relative to the other plots. The same process occurs for moderate-scale perturbations (middle row left in Figure 3.7): here, the error moves slightly more readily towards smaller wavenumbers than towards larger ones. After around 20000 timesteps (middle plots) the graphs look very similar regardless of the perturbation scale k_x, with large-scale error starting to take over from small-scale error in all three cases. The system is moving towards a -5/3 power scaling law for error, as is revealed by fitting a straight line to the saturated scales in the longer-term $k_x = 9$ log-log plot, Figure 3.9. The best-fit line has a gradient of $-1.6 \approx -5/3$ and an intercept of $0.23 \approx \log(1.7)$. This implies that the streamfunction error $\delta\psi$ varies with wavenumber according to $\delta\psi \approx 1.7k^{-5/3}$.

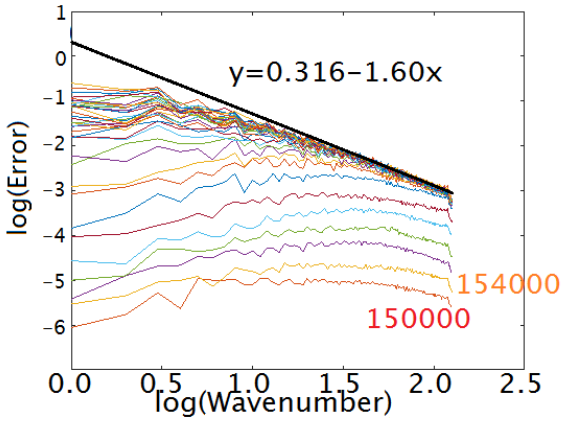

Figure 3.9: The base-10 logarithm of the streamfunction error is plotted against that of the wavenumber at 4000 timestep intervals for 150000 timesteps after the initial perturbation at $k_x = 9$. For $\log(k) \gtrsim 1.5$ ($k \gtrsim 30$) the error is saturated and is no longer growing; a best-fit curve is plotted for this part of the graph.

3.3.4. Experiment 2: Error Doubling Times

The high temporal resolution output for a single ensemble member from Experiment 1 was used to estimate the EDT at each spatial scale, and the relative size of the initial perturbation was hence the same for all spatial scales tested. For each wavenumber, the initial growth in streamfunction error

in the perturbed wavenumber after the introduction of the perturbation was measured by fitting a straight line to the error growth curve and computing its gradient m. The initial error (that at 50000 timesteps, when the perturbation was introduced) ϵ_0 was also measured, and the time taken for this error to double in size – that is, for the error to increase by this much again – was taken to be the EDT. Therefore, the equation used to calculate the EDT was,

$$t_d = \frac{\epsilon_0}{m}. \tag{3.19}$$

Figure 3.10 shows examples of an error growth curve used to compute the Error Doubling Time at large scales ($k_x = 20$) and small scales ($k_x = 60$). In all cases, there was an initial decrease in the error as the model settled into its new state, meaning that ϵ_0 is not equivalent to the y-axis intercept of the best-fit straight line.

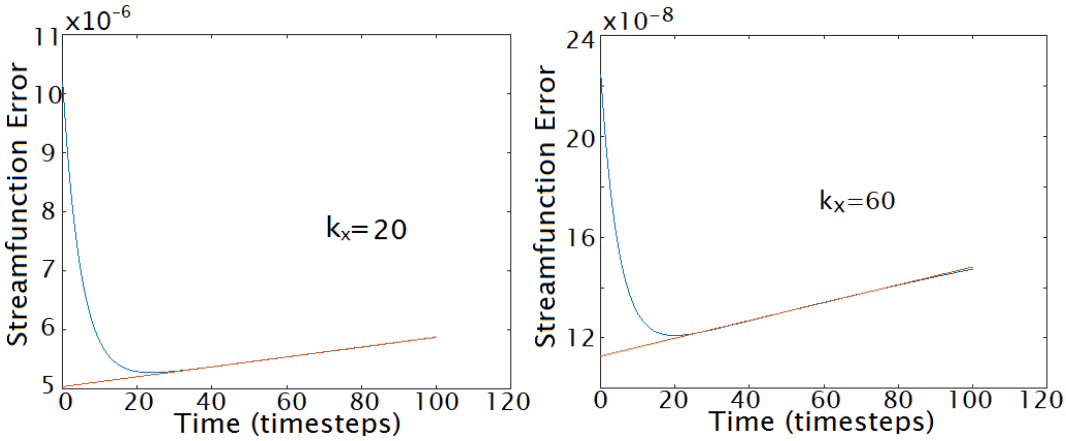

Figure 3.10: The streamfunction Error as a function of time after an initial random perturbation for $k_x = 20$ and 60, corresponding to $k_r = 28$ and 85. In each case, the perturbation size was measured from where the curve crosses the y-axis, and the gradient of the best-fit straight line to the first sustained period of growth (orange line) was used to deduce how long it would take this error to double (the EDT) if this initial growth were to be sustained.

Table 3.1 lists the results up to $k_r = 100$. The EDT decreases rapidly as k_r increases until $k_r \gtrsim 40$, whereupon the slope shallows as shown when the results are plotted in Figure 3.11. It was found that the best-fitting model for the Error Doubling Times is that of a polynomial relationship with wavenumber, whereby $\tau_d = ck^m$, so that $\ln(\tau_d) = m\ln(k) + \ln(c)$. To find the gradient and intercept of the best fit line, $\ln(\tau_d)$ was plotted against $\ln(k)$as shown in Figure 3.12. The points lie

almost exactly in a straight line, with a gradient of $m = -1.23 \pm 0.03$ and an intercept $\ln(c) = 10.9 \pm 0.1$; these errors represent the error in fitting to the supplied data as output by the GNUPLOT fitting program, and do not account for uncertainties in the EDTs themselves. This corresponds to the relationship,

$$\tau_d = (5.42 \times 10^4) k^{-4/3}\,, \tag{3.20}$$

similar to the $-5/3$ power law that is observed for the enstrophy spectrum, implying that – at moderate wavenumbers at least – errors propagate between spatial scales according to a form of three-dimensional turbulence, which accords well with the results of Experiment 1.

k_x	k_r	EDT	k_x	k_r	EDT	k_x	k_r	EDT	k_x	k_r	EDT	k_x	k_r	EDT
1	1	55941	16	23	1143	31	44	615	46	65	394	61	86	229
2	3	12603	17	24	1267	32	45	439	47	66	234	62	88	247
3	4	7954	18	25	1059	33	47	468	48	68	375	63	89	292
4	6	5607	19	27	998	34	48	490	49	69	272	64	91	191
5	7	5897	20	28	607	35	49	483	50	71	291	65	92	68
6	8	5515	21	30	804	36	51	482	51	72	284	66	93	196
7	10	3333	22	31	760	37	52	354	52	74	265	67	95	158
8	11	3165	23	33	666	38	54	408	53	75	352	68	96	252
9	13	1950	24	34	796	39	55	394	54	76	255	69	98	232
10	14	1955	25	35	626	40	57	327	55	78	290	70	99	448
11	16	1458	26	37	649	41	58	454	56	79	220	71	100	251
12	17	1591	27	38	558	42	59	321	57	81	211			
13	18	1638	28	40	526	43	61	439	58	82	217			
14	20	1615	29	41	668	44	62	254	59	83	241			
15	21	1107	30	42	480	45	64	279	60	85	317			

Table 3.1: Error Doubling Times in timesteps measured for initial perturbations to the streamfunction at $k_x = k_y$, for values of k_r ranging between 1 and 100. The perturbations were random, but the maximum relative size of the perturbation was the same for all spatial scales.

The EDTs obtained here can also be used to relate the model's dimensionless time units to the real world. Operational atmospheric models have an EDT of around 1.5 real Earth days on the spatial scale at which they forecast (Lorenz 2006b). It is difficult to relate this to a spatial scale in this non-dimensional model, but assuming that $k_r = 20$, which has an EDT of $\tau_d \approx 1000$ timesteps, corresponds roughly to the scale at which a model with resolution $k_{\max} = 127$ would be used to make predictions akin to real-world weather forecasts implies that 1 model timestep is

roughly equivalent to 1.5×10^{-3} days, or about 2 minutes. This rough approximation will be used to contextualise the times-scales used in the SQG system throughout the rest of this thesis.

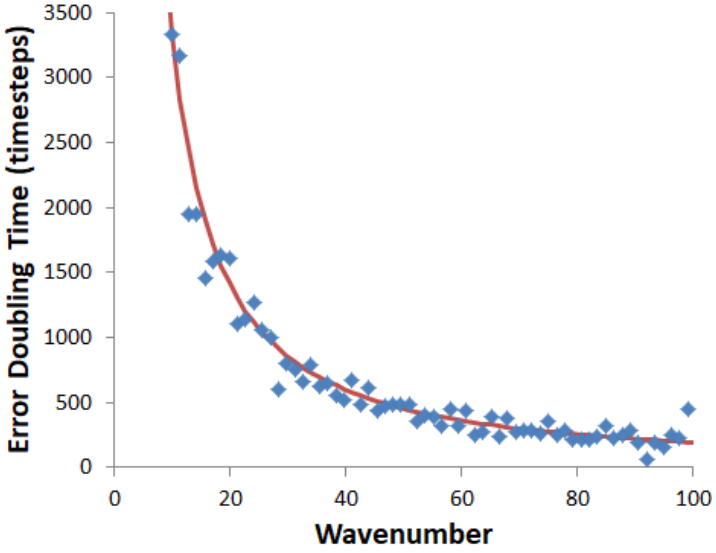

Figure 3.11: The measured Error Doubling Times are plotted against overall wavenumber k_r. A best fit curve is also shown (red), determined from the log-log plot for EDT and wavenumber (Figure 3.12), where the relationship between these is modelled as a polynomial.

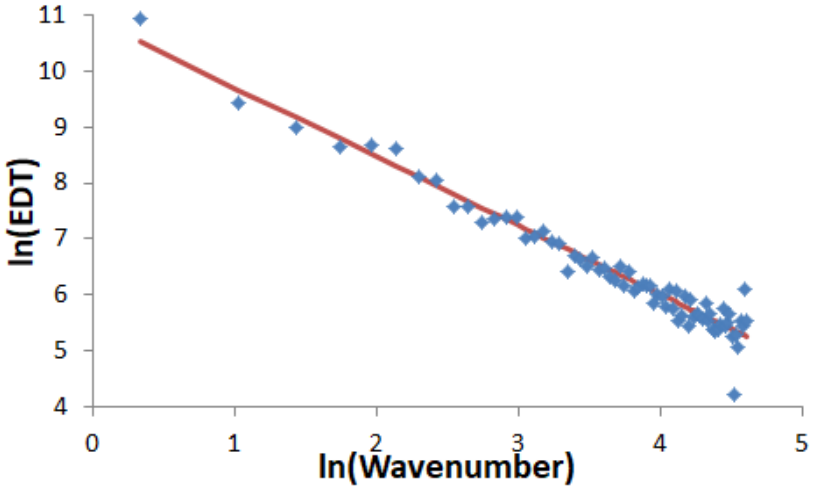

Figure 3.12: The EDT is plotted against wavenumber k_r with both variables on logarithmic scales, with a best-fit straight line (red).

3.3.5. Experiment 3: The Butterfly Effect

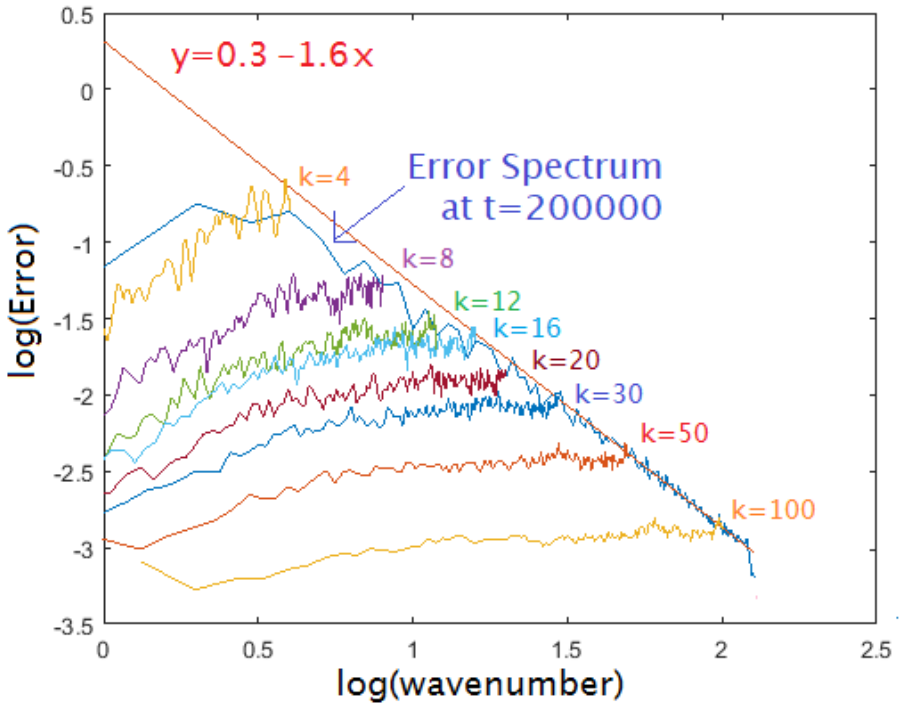

Figure 3.13: The base-10 logarithm of the error is plotted as a function of wavenumber (blue curve, labelled 'error spectrum') for a 200000 timestep run of Setup S, after a perturbation was introduced at 50000 timesteps at wavenumbers above $k_x = 20$. The red line is the best fit to this curve, showing that it follows approximately a -5/3 gradient. The other eight curves superimposed on this refer to eight different wavenumbers (labelled) at which the error was measured as a function of time; the leftmost point of each of these curves corresponds to time $t = 0$, whereas the rightmost point of each corresponds to time $t = 200000$. These curves are each stretched by different amounts so that they all meet the overall spectrum (the blue curve) at 200000 timesteps; flatter curves saturate at their final value more rapidly.

If a system exhibits the 'Butterfly Effect', first discovered by Lorenz in the 1960s (Lorenz 1963, 1969), when even a small perturbation to the initial conditions is introduced in the smallest scales, this should rapidly propagate up-scale to induce errors relative to the equivalent unperturbed system that eventually saturate at an error corresponding to zero skill. This implies that even if large-scale observations of the atmosphere are perfect, any errors in small-scale observations will inevitably limit the forecast lead-time for which the system remains predictable, regardless of how

small those errors are. This has two key implications for a system where the error at different wavenumbers saturates according to a -5/3 spectrum such as the SQG system.

First, Lorenz showed that whilst the small-scale errors may reach their saturation point within a matter of minutes, larger scales (the important scales for real-world forecasts) reach saturation within a few days, which is hence the limit of useful predictability. In that case, for a system with three-dimensional turbulence the time taken $\tau(k)$ for errors at wavenumber k to have a significant effect on (larger-scale) wavenumber $k/2$ should be proportional to $k^{-2/3}$ (Palmer et al. 2014). This prevents the limit of predictability from being pushed back appreciably by increasing the resolution of accurate observations, and implies that representing the very small scales beyond a certain level of precision will not be very important for the accuracy of forecasts of larger-scale quantities. For a system displaying true two-dimensional turbulence, on the other hand, τ should be indepdendent of k for large enough k. Second, and as a consequence of this, after a given amount of time has elapsed before saturation, the error in the large scales (e.g. $k_L = 2$) should be the same, regardless of the exact wavenumber k at which the perturbation was introduced, for large k. The aim of this experiment was to test both of these predictions in the SQG system.

To do this, since only the propagation of errors upscale – not downscale – was relevant, the methodology of Experiment 1 was adapted slightly so that rather than perturbing a single wavenumber, the whole spectrum above a specified wavenumber of perturbation was perturbed. The perturbation was created by sampling the high wavenumber spectrum of a long run of the same system and replacing the high wavenumbers of the perturbed run with these values from the sampled run, thereby preserving a realistic system state. The wavenumber of perturbation k_r was varied and the error growth was tracked across all wavenumbers k. Figure 3.13 plots the error spectrum logarithmically as a function of wavenumber after 200000 timesteps have elapsed for $k_x = 20$ ($k_r = 28$), with a best-fit line showing that this follows approximately a $-5/3$ power law. On top of this are plotted further curves intended to show how the error grows over time for eight different wavenumbers on a logarithmic scale. Each of these curves corresponds to the same number of timesteps (from $t = 0$ on the left to $t = 200000$ on the right), but they are stretched by

different amounts to meet the overall spectrum at 200000 timesteps. The length for which each curve is flat therefore indicates the period of time for which the error at that scale is saturated. This saturation occurs much earlier for the higher wavenumbers, as implied by Experiment 1. The saturation timescale is hence indeed wavenumber-dependent and is larger for smaller wavenumbers.

To investigate further, for a variety of values of k_r the error after 10000 and 20000 timesteps after the introduction of the perturbation at the very largest scale $k = 1$ was computed. Figure 3.14 plots the logarithm of this error in both cases, and shows that the exact wavenumber k_r at which the perturbation is applied is less and less important in determining the error at large scales as k_r increases. Both curves begin with a large error in $k = 1$ when the perturbation is introduced at small k_r, but the error decreases as k_r is increased. Eventually, the curves' gradient diminishes at high k_r, which is not quite consistent with the Butterfly Effect theory as neither curve appears to be entirely levelling off. The curves shallow over a wider range of wavenumbers at longer lead-times after the perturbation; this is to be expected, as after a longer time the model error will have grown, masking the effect of the initial perturbation.

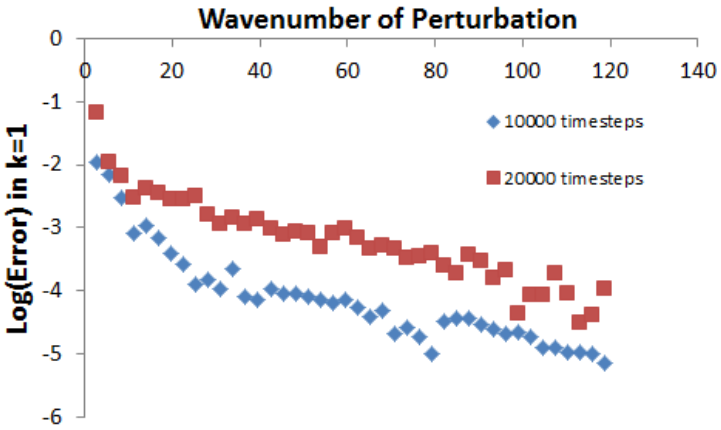

Figure 3.14: The logarithm of the error in the largest scale ($k = 1$) is plotted 10000 (blue dots) or 20000 (red squares) timesteps after a perturbation is introduced for all wavenumbers greater than or equal to the 'wavenumber of perturbation' k_r, plotted against the wavenumber of perturbation.

Finally, the dependency of the time τ for errors to propagate from scale k to scale $k/2$ was measured explicitly. Several perturbed models were run for 150000 timesteps after spin-up, at

$k_{\max} = 255$ resolution to avoid dissipative effects which occur at the smallest resolved scales from affecting the results near $k = 127$, and compared to the 'truth'. All wavenumbers were perturbed above k_t, but the different models used different values of k_t ranging between 0 and 248 in intervals of 4. For each k_t, the Absolute Error in scale $k_t/2$ was calculated every 200 timesteps; the error in large-scale wavenumbers k_L ranging between $k_L = 1$ and $k_L = 10$ was also computed. The whole process was repeated four times using independent sets of initial conditions, and the errors at each timestep were averaged over the five runs. The time τ taken for the error in $k_t/2$, and the time T for the error in each of the k_L, to reach a threshold value was then measured using a temporal resolution of 200 timesteps. This measurement was made for threshold error values of between 10^{-6} and 10^{-3}, which correspond to around 0.1% and 100% of the large-scale average streamfunction magnitude ($\sim 10^{-3}$) respectively.

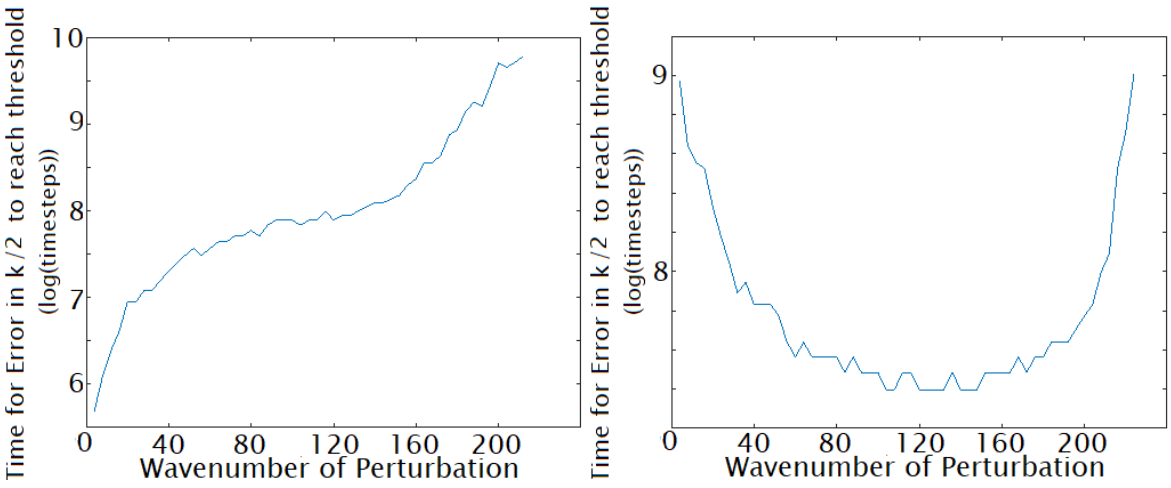

Figure 3.15: The natural logarithm of the time taken τ for the streamfunction error to reach a specified threshold of 1×10^{-5} (left) or to reach a scale-dependent threshold of $(k_t/2)^{-5/3} \times 10^{-3}$ (right) in wavenumber $k_t/2$ is plotted, where the initial error is restricted to wavenumbers above each wavenumber of perturbation k_t, at $k_{\max} = 255$ resolution.

If the SQG system behaves like a $k^{-5/3}$ 3D system, as its enstrophy spectrum suggests it should, we would expect $\tau \propto k_t^{-3/2} G^{-1/2} \propto k_t^{-2/3}$, whereas T should be independent of k_t at large k_t (Palmer et al. 2014). Figure 3.15 shows an attempt to measure τ, using a fixed error threshold of

1×10^{-5} for all values of $k/2$, and shows that τ by this measure increases with wavenumber for low k before levelling out. However, this result is unlikely to be accurate, because the error threshold that can be considered 'significant' should be larger at smaller wavenumbers: an error of 1×10^{-5} would be very substantial relative to the average streamfunction magnitude at $k/2 = 50$ (see Figure 3.4), for instance, but is only around 1% of the average magnitude at $k/2 = 1$. Any attempt to correct for this by using a scale-dependent threshold – for example, $k^{-5/3} \times 10^{-3}$ would yield an error threshold curve with approximately equal magnitude to the streamfunction at $k/2 = 1$ – risks muddying the results by artificially imposing a power law that is difficult to disentangle from the observed trend for τ. It is more useful, therefore, to use T to determine whether the Butterfly Effect is in operation, because here the time taken for a threshold error to be induced in the same single wavenumber or group of wavenumbers is measured for all the values of k_t under investigation.

Figure 3.16 plots T against k_t for $k_L = 1$, this time using ten runs with different initial conditions, with the resulting error spectra having been averaged before measuring T. The figure demonstrates that T is independent of k_t when the time taken to reach any of four different thresholds is measured in wavenumber $k_L = 1$; Figure 3.17 demonstrates the same effect when the error growth is averaged over all wavenumbers k_L between 1 and 10 for the same ten runs. The latter case yields smoother plots, especially for an error threshold of 1×10^{-3} which is not reached at all in $k_L = 1$ for initial perturbations restricted to wavenumbers above 160. This latter threshold is the most significant, as it represents roughly the average magnitude of the streamfunction at large spatial scales ($k_L \approx 1$). Hence, the results show that the levelling-off at high wavenumbers is most prominent for streamfunction errors that are significant in size relative to the background magnitude, and is still visible for smaller error thresholds.

Although in no instance do the curves completely flatten-off as might be expected in theory (which may be because of non-linear interactions between non-adjacent wavenumbers), this result may nonetheless be compatible with the Butterfly Effect, implying that the time taken for an initial error to significantly affect the largest scales becomes almost independent of that error's spatial

scale when that scale is small (above $k \approx 50$). If this is indeed the case, it would be impossible to indefinitely extend the accuracy of a forecast of the SQG system by restricting initial observational errors to smaller and smaller spatial scales.

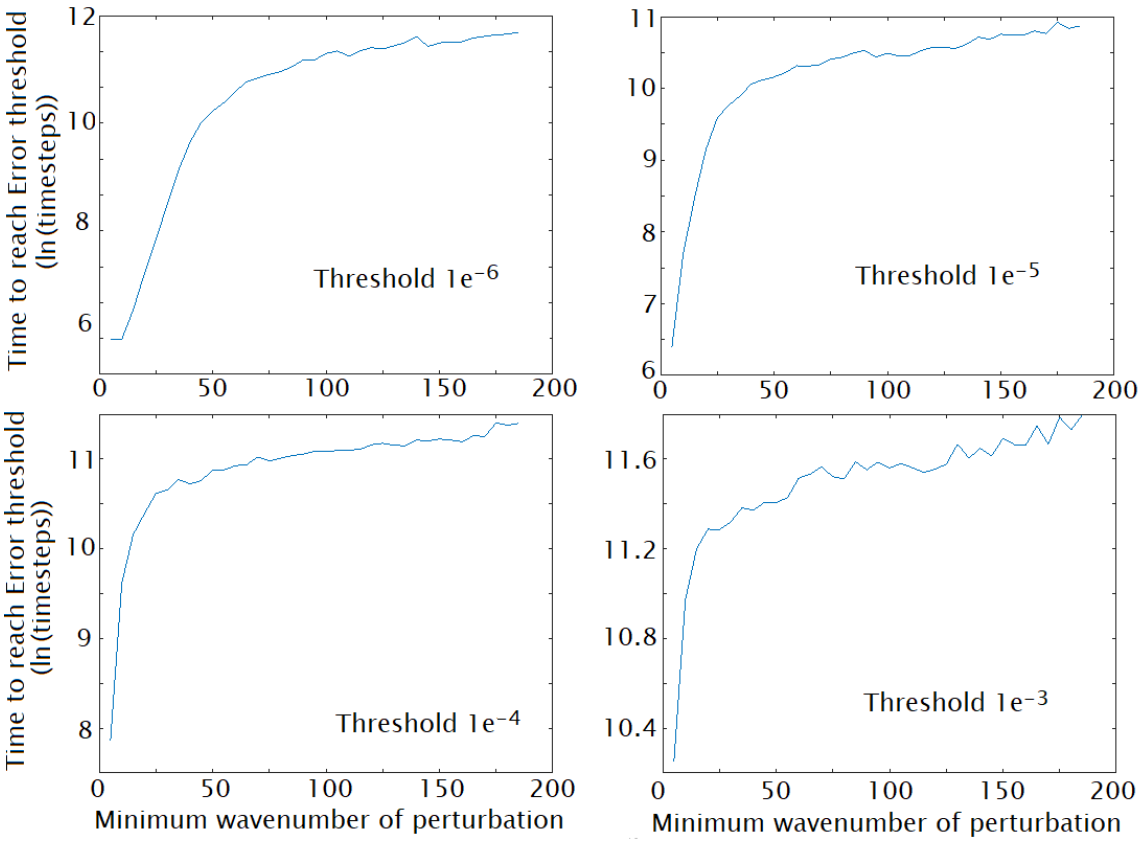

Figure 3.16: The natural logarithm of the time taken T for the streamfunction error to reach a specified threshold, ranging from 1×10^{-6} (top left) to 1×10^{-3} (bottom right) in wavenumber $k_L = 1$ is plotted as a function of the minimum wavenumber above which initial condition perturbations are applied. Simulations were carried out at $k_{\max} = 255$ resolution, and the error spectrum from which T was calculated was averaged over ten runs with different initial conditions of 150000 timesteps each after spin-up.

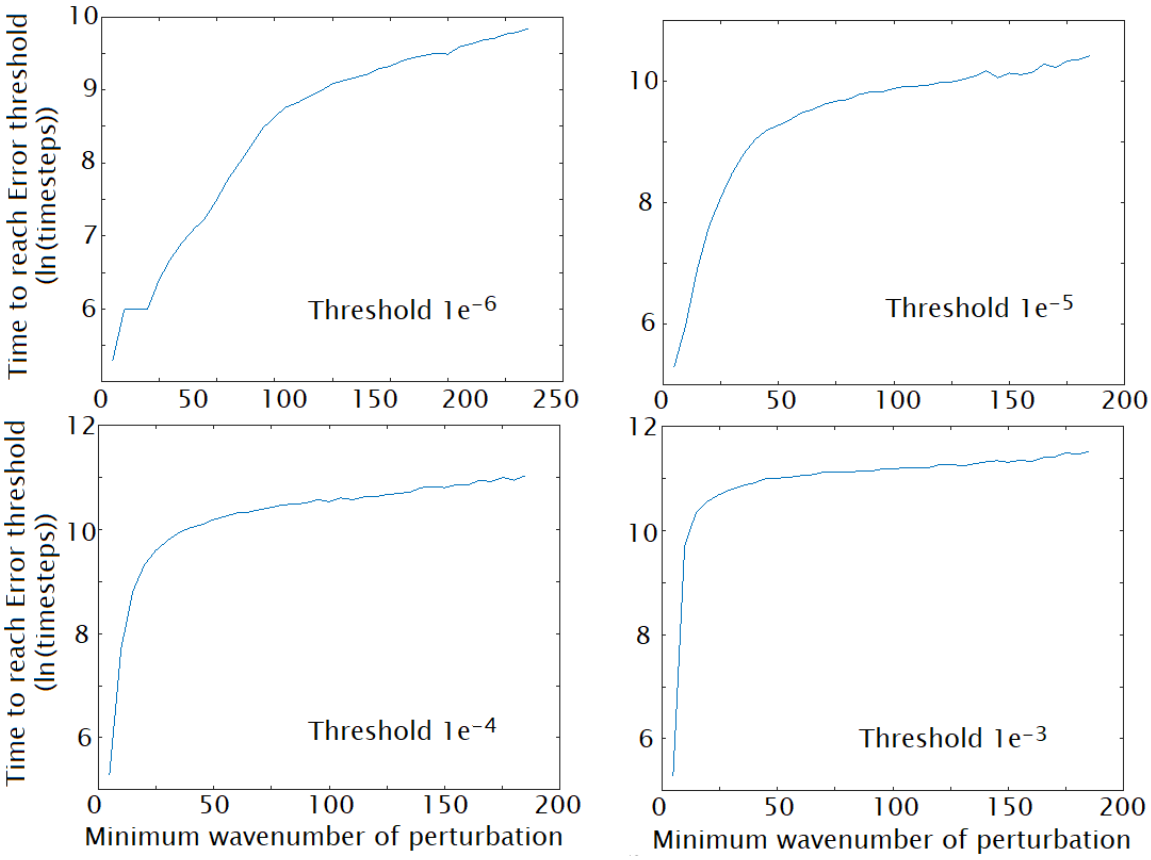

Figure 3.17: The natural logarithm of the time taken T for the streamfunction error to reach a specified threshold, ranging from 1×10^{-6} (top left) to 1×10^{-3} (bottom right) for error growth curves averaged across all wavenumbers between $k_L = 1$ and $k_L = 10$ is plotted as a function of the minimum wavenumber at which initial condition perturbations are applied. Simulations were carried out at $k_{\max} = 255$ resolution, and T was computed by averaging over 10 runs of 150000 timesteps each after the initial 50000 timestep spin-up, with different bottom topography giving rise to different initial conditions for each.

3.4. Implications for Reduced Precision Experiments

The results of the above experiments accord with Lorenz' (1969) observation that errors at smaller spatial scales should grow more rapidly (have smaller Error Doubling Times) than those at larger scales, and that an initial error at very small scales should rapidly propagate upscale, for a system with $\lambda < 3$ such as the SQG system. However, the results suggest that downscale error propagation from an initial error in the largest scales is more rapid still, so that if errors with the same *relative* size are introduced on both large and small scales the former will soon spread to all scales, making

an initial condition error at small scales irrelevant in the long-term unless the relative initial error at all larger scales is much smaller. This finding at first seems somewhat counterintuitive, given that the EDT is larger at larger scales, but can be explained when one takes into account three important considerations. Firstly, the EDT applies to the *absolute* value of the error, whereas the error's absolute size will be larger at larger spatial scales when the *relative* error is the same at all scales; secondly, the EDT describes only the growth in error at the scale where the perturbation is introduced, not its propagation to other scales. Finally, the results of Experiment 2 show that the error saturates more rapidly at smaller spatial scales, with the maximum error at each scale also proportional to the -4/3 power of the wavenumber, implying that the capacity for errors to develop on small scales is more limited regardless of their higher EDTs.

Hence, these results are in agreement with Durran & Gingrich's (2014) hypothesis that minimising the absolute observational error at larger spatial scales, to ensure that the relative error is small at these scales, is more important for an accurate forecast than is extending accurate observations down to smaller scales. Retaining high precision to avoid rounding error should therefore be more important at the larger scales. This may have particular relevance for reduced precision studies in the SQG context, since reducing precision by truncating the significand would not incur the same *absolute* error on all spatial scales, but would induce a similar *relative* error across all spatial scales. Therefore, it would be expected that the number of significand bits required for an accurate forecast should be smaller at higher wavenumbers, following a similar trend to the -4/3 power law observed here in the EDTs. This power law can be attributed to the three-dimensional turbulent nature of the energy cascade that is observed in the real atmosphere and the enstrophy cascade that is exhibited by the SQG system.

4. Reduced Precision in an Idealised System 2: The Surface Quasi-Geostrophic Model

Having analysed the propagation of errors between scales in the SQG system, it is now possible to attempt to answer the question that motivated its investigation in this thesis in the first place: what is the impact on the accuracy of forecasts of the SQG system of reducing numerical precision at different spatial scales? This accuracy is assessed in this chapter through a series of experiments in which reduced precision models are used to make forecasts using the setups introduced in Section 3.2.4. The implementation of the Mixed Precision emulator is described (Section 4.1), and the results of experiments carried out using this emulator are given in detail in weather (Section 4.2) and climate (Section 4.3) contexts. It is found that for the spectral part of the system it is possible to formulate a power-law function describing the required numerical precision at each wavenumber, and that the required precision is less at smaller spatial scales in accordance with the hypothesis presented in Section 1.1. An attempt is made to estimate the potential computational cost savings that might be associated with such a scale-selective approach (Section 4.4) and the implications are discussed (Section 4.5).

4.1. Emulating Reduced Precision in the SQG System

In a system such as the SQG one that is solved numerically in spectral space, spectral variables are represented on the computer as arrays whose different components correspond to different wavenumbers (spatial scales) of those variables. Hence, to carry out scale-dependent precision experiments in the SQG system, a version of the Mixed Precision emulator introduced in Section 1.5.3 was used for which the number of bits in the significand could be specified for each wavenumber component of a given spectral variable. This meant that the level of precision could be specified for each spatial scale without having to define independent variables for each scale.

Although by default in the emulator used here the exponent remains at the double-precision standard (11 bits), no matter how many significand bits are used, it is also possible to run in 'half-precision mode', whereby the number of bits in the exponent is reduced to the 5 bits required by the half-precision IEEE standard when 10 significand bits are used. However, in none of the experiments presented below was true half-precision (10 bits in the significand, 5 bits in the exponent) observed to yield results any different from those obtained with 10 bits in the significand and 11 bits in the exponent.

This is in accordance with the results obtained in the modified Lorenz '96 experiments, where the exponent precision could be reduced to as low as 4 bits before there was a degradation in accuracy. Subnormal numbers are used by the emulator, so the truncated variables must be larger than $2^{-24} \approx 10^{-11}$ for the lower bound of the half-precision range not to be exceeded; indeed, the spectra of the SQG system variables reveal that this is feasible across all scales for most spectral quantities. In the results that follow, a distinction is made between '10 bit significand' and 'half-precision' cases, with the exponent truncated only in the latter case, but in all instances the two yield equivalent results. There is no IEEE standard for reducing bits below half-precision, and it can be assumed that 5 exponent bits could be used uniformly in real Mixed Precision hardware.

In light of the results from Lorenz '96 experiments, which showed that the efficacy of Stochastic Processors was less than the potential efficacy of Mixed Precision processors of a smaller estimated computational cost, a program to emulate a 'Stochastic Processor' in a similar scale-selective way for the SQG case was not developed, and only the Mixed Precision technique was investigated. There is no reason why Stochastic Processors could not in principle be applied in future studies given a suitable emulator.

4.2. Experimental Results: Short-term Forecasts

Using Setup S, the efficacy of scale-selective reduced precision for short-term forecasting in SQG was measured in five different experiments, and this section presents the results. The first

experiment is restricted to the double-, single- and half-precision standards that were employed in the simpler Lorenz ’96 model. The impact of reducing the prognostic variables to half-precision is assessed, and an attempt is made to determine over what range of spatial scales half-precision can be implied with impunity. But the multiscale nature of the SQG system makes a more dynamic approach to precision attractive, going beyond the standard precision paradigms to investigate exactly how much precision is necessary for an accurate forecast. Experiment 2 therefore explores the lowest number of bits in the significand that is permissible at each spatial scale in turn without compromising the model accuracy, and the results obtained from this are used in Experiment 3 to run the model with ‘optimal’ precision at each spatial scale.

The above experiments are all concerned solely with the precision used to carry out calculations in the grid-point and spectral parts of the model, not with the transitions between them. In Experiment 4, the degree to which precision can be reduced when carrying out Fast Fourier Transforms (FFTs), which are necessary to change between spectral and grid-point space but are especially computationally costly, is analysed. FFTs are performed in double-precision throughout most of the other experiments because emulating reduced precision in this part of the model very significantly reduces the run speed. In Experiment 5, it is verified that reduced precision in the spectral, grid-point and FFT components simultaneously can yield an accurate forecast, with the precision optimised for efficiency across the whole model. Finally, in Experiment 6, high-resolution, low-precision runs and low-resolution, high-precision runs are compared to see whether increasing resolution by reducing precision is beneficial, using parameterisation to make the analysis more relevant to real-world models.

In the initial comparisons of the quality of short-term, ‘weather’ forecasts between double-precision and reduced precision models (Experiments 1, 2, 4 and 6), only the Absolute Error metric was used. This metric is simple to compute, comprising the absolute difference between the model and the ‘truth’ evaluated in grid-point space and averaged over all grid-points, but provides a straightforward means of quickly identifying the impact of rounding errors induced by reducing precision. In Experiments 3 and 5, which draw together the findings of previous experiments to test models with optimised precision, further metrics are used to verify the models’ accuracy. These are

the Root Mean Square Error (RMSE), the Ranked Probability Skill Score (RPSS) introduced in Section 1.7.1 and the difference in ensemble spread between different models.

4.2.1. Experiment 1: Single- and Half-precision

For this first experiment, the effect of reducing the spectral variables' components that correspond to large spatial scales to single-precision and those pertaining to smaller scales to half-precision was analysed. All runs were performed at the same resolution, with maximum wavenumber $k_{\max} = 127$, using Setup S and the grid-point and FFT components were kept in double-precision.

The 'truth' system was spun up from an initially flat streamfunction field to 150000 timesteps to define the 'initial conditions', then run for a further 2000 timesteps, corresponding to around 3 days in the real atmosphere according to the estimate made in Section 3.3.4 (about 2 minutes per timestep). For these last 2000 timesteps, this 'truth' was compared to a set of models, each with a different truncation wavenumber k_t determining the point at which 'large' and 'small' scales were divided, with single-precision for $k \leq k_t$ and half-precision for $k > k_t$. Each model was run fifty times to form an ensemble forecast, with each member of the ensemble beginning with initial conditions prescribed by the spin-up of the truth system but for which every component of the streamfunction was perturbed by a different random amount up to 5% of the 'true' values in grid-point space to represent observational uncertainty. Because the perturbations were applied in grid-point space they were not scale-dependent, and should not therefore have affected the relative precision required for an accurate forecast for different spatial scales.

The Absolute Error in the streamfunction, averaged over all ensemble members and all points in grid-point space, is plotted for all the models with k_t between 1 and 9 in Figure 4.1. There is no loss of accuracy at all for $k_t = 8$ or 9 in half-precision; only for $k_t \leq 3$ does the error begin to grow considerably. Hence, an application of half-precision to all wavenumbers greater than 8 in the SQG system would not be expected to affect the accuracy of the model's output, given the simulated observational error that is encapsulated in the ensemble spread.

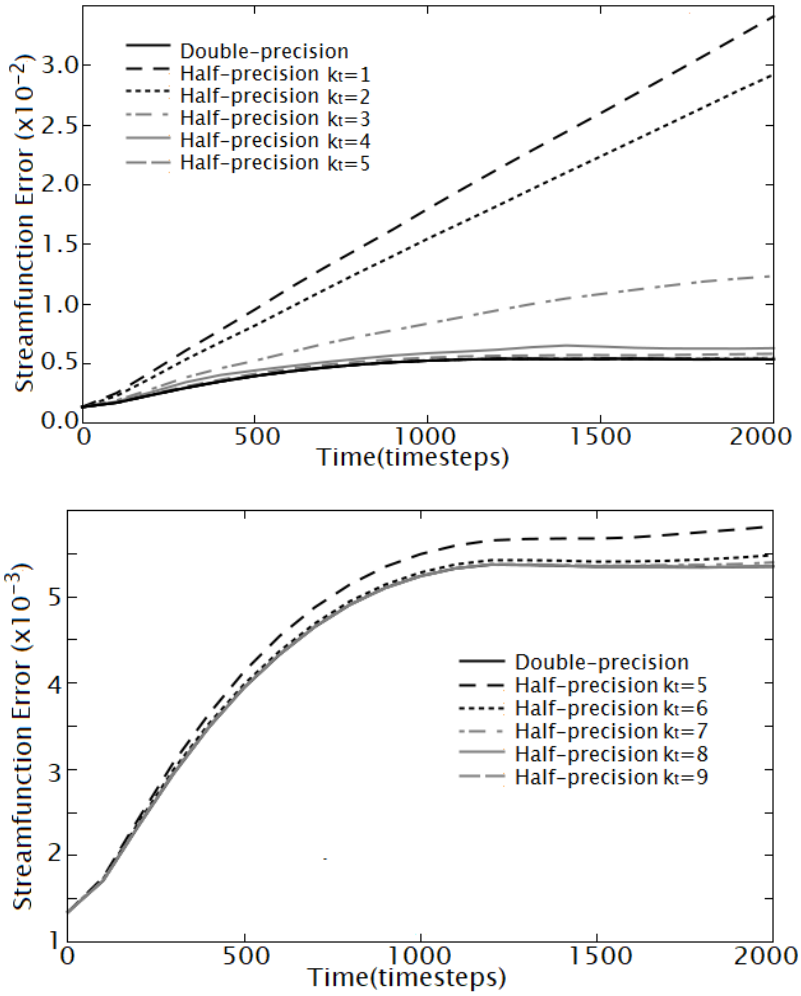

Figure 4.1: The streamfunction Absolute Error is plotted for nine half-precision models with the truncation wavenumber varying between $k_t = 1$ and $k_t = 5$ (top), and between $k_t = 5$ and $k_t = 9$ (bottom), as compared to the double-precision control run (solid black line, hidden behind the $k_t = 8$ and $k_t = 9$ lines in the bottom plot). The plots were averaged over fifty ensemble members for each model, with members differing only in the perturbation (by a random amount up to 5%) to the initial conditions that they used.

4.2.2. Experiment 2: The Limits of Reduced Precision

Experiment 1 showed that truncating to half-precision all wavenumbers above $k_t = 8$ had a negligible effect on the accuracy. The aim of this second experiment was to see exactly how far precision could be reduced during calculations for all wavenumbers above k_t before errors became evident, for each spatial scale (that is, for each possible value of k_t) in turn. To do this, many different reduced precision models were created, each with a single-precision 23 bit significand

used for the large scales ($k \leq k_t$) but with a lower precision applied to the small-scales ($k > k_t$). The small-scale precision was varied between 4 and 23 bits in the significand, and the wavenumber of truncation k_t was varied between 0 and 127. Runs with $k_t = 0$ had all spatial scales in small-scale, reduced precision; those with $k_t = 128$ had all scales in large-scale, single-precision. Precision was reduced in all the spectral quantities used in calculations at each timestep. The unperturbed 'truth' system was first spun-up for 50000 timesteps; then the truth, a double-precision control ensemble and multiple reduced precision ensembles were run forward for a further 5000 timesteps (around 7 atmospheric days). Each ensemble had fifty members with perturbed initial conditions as in Experiment 1.

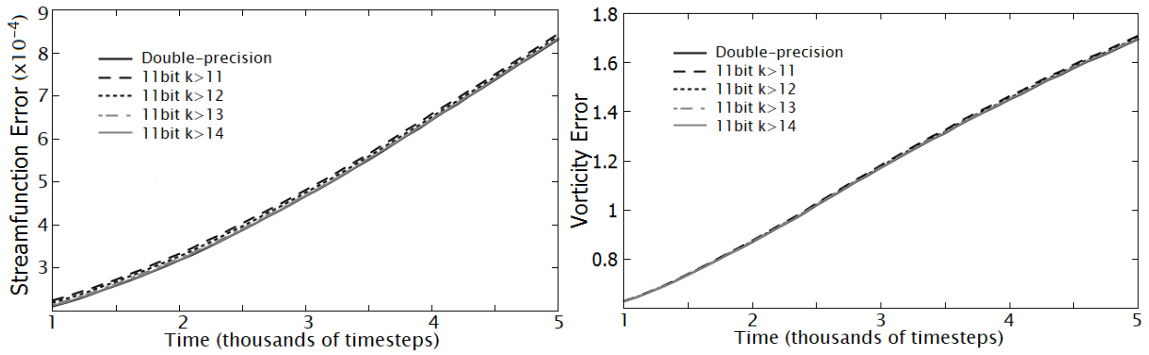

Figure 4.2: The Error measured in the streamfunction (left) and vorticity (right), averaged over all spatial scales and over 50 ensemble members, between the 'truth' and firstly a double-precision control run (black lines), and secondly runs with 23 bits in the significand for wavenumbers $k \leq k_t$ and 11 bit precision for $k > k_t$ for k_t varied between 10 and 14 (see the keys on each chart). The curves with $k_t \geq 12$ for the streamfunction and $k_t = 14$ for the vorticity are indistinguishable from double-precision.

At 3000 timesteps (4 days), the Absolute Error – averaged across all grid-points and all fifty ensemble members – in both prognostic variables (the streamfunction and the vorticity) was measured relative to the 'truth', for both the double-precision control and the reduced precision models. For each number of bits in the significand s to be tested, several reduced precision models were run, with the truncation scale k_t being gradually increased for each successive model, and the first value of k_t for which there was no significant difference in Error between the double-precision model and the reduced precision model at 3000 timesteps was determined. This was repeated until

all relevant combinations of k_t and s had been investigated. To save computer running time, it was possible to reduce the number of necessary runs by roughly identifying the range of values of k_t across which the model would change from inaccurate to accurate for a given level of significand precision s. The difference in Absolute Error was deemed to be 'significant' when the control and reduced precision errors differed by more than the experimental uncertainty, which was taken to be the standard error in the mean of the 50 control run members, and was typically less than 0.5% of the control ensemble's mean Absolute Error relative to the 'truth'. It was necessary to determine the optimal combinations of s and k_t separately for the streamfunction and vorticity fields.

Figure 4.2 shows the results obtained in an example where there are $s = 11$ bits in the significand and k_t is varied between 10 and 14 in successive reduced precision models, with separate plots for the vorticity and the streamfunction. The error curve is indistinguishable from the control at 3000 timesteps after spin-up for $k_t = 12$ and $k_t = 14$ for the vorticity and streamfunction respectively. Similar curves could be plotted for any of the levels of small-scale precision investigated here. Table 4.1 shows the results in full. These suggest that the streamfunction is more sensitive to reduced precision than is the vorticity at large spatial scales: the precision in all variables could be reduced to 16 bits at all spatial scales without incurring any additional error for this 4-day forecast relative to the control in the vorticity, whereas the streamfunction required this reduction in precision to be restricted to $k > 3$. At the smallest spatial scales, however, it is the vorticity that is more sensitive, requiring 5-bit precision to be restricted to $k > 123$ for unaffected forecasts as opposed to $k > 111$ for the streamfunction.

This may be because the streamfunction field is relatively smooth, equivalent to some slowly-varying quantity in the real atmosphere with limited small-scale detail, so that the small-wavenumber components of ψ are in general stronger than the high-wavenumber components, as illustrated in the streamfunction spectrum shown in Section 3.2.4. Hence, the smallest scales have a smaller absolute error associated with them than the large scales and contribute much less to the overall error score, so very low precision can be used for a relatively large range of the very high wavenumbers (small spatial scales) so long as the precision is not so low as to induce large errors,

but relatively high precision is necessary in a large range of the low wavenumbers (large spatial scales). This large-scale bias is present but less strong in the vorticity, which is more evenly distributed across spatial scales than the streamfunction, with the medium scales being the strongest (see again the spectra in Section 3.2.4). The vorticity displays a more intricate structure and corresponds to a more difficult-to-predict and spatially variable quantity in the real atmosphere. This lowers the contribution of the largest scales and increases the contribution of the smallest scales to the overall Error, relative to the streamfunction, and thus makes the vorticity less sensitive to reduced precision at large scales and more so at small scales. No attempt was made to counter the implicit weighting towards the larger scales here because it is the larger scales which have more impact on the evolution of real-world weather forecasts, where the values of variables on scales close to the model's truncation scale are often not included in the output at all.

Significand bits	Minimum k_t for accurate ψ	Minimum k_t for accurate ζ
4	NONE	NONE
5	111	123
6	84	94
7	54	60
8	36	38
9	25	24
10	17	16
11	14	12
12	9	8
13	7	5
14	6	3
15	4	3
16	3	0
17	2	0
18	2	0
19	0	0

Table 4.1: The minimum wavenumber of truncation k_t above which all wavenumbers in the streamfunction ψ or vorticity ζ can be represented with the specified number of bits in the significand for the Error of simulations in reduced precision to differ from that of a double-precision control run by less than the standard deviation from the mean of the 50 control ensemble members, under Setup S, as measured at 53000 timesteps. When 4 significand bits were used there was significant error regardless of the truncation scale.

This experiment involved reduced precision not only in ψ and ζ but also in other quantities such as the array representing the right-hand-side of the governing equation. Further tests showed that there is no difference in the range of wavenumbers to which reduced precision can be safely applied when precision is instead only reduced for ψ and ζ with all other variables kept in double-precision throughout all calculations, which implies that the overall rounding error is dictated by the rounding error in these quantities.

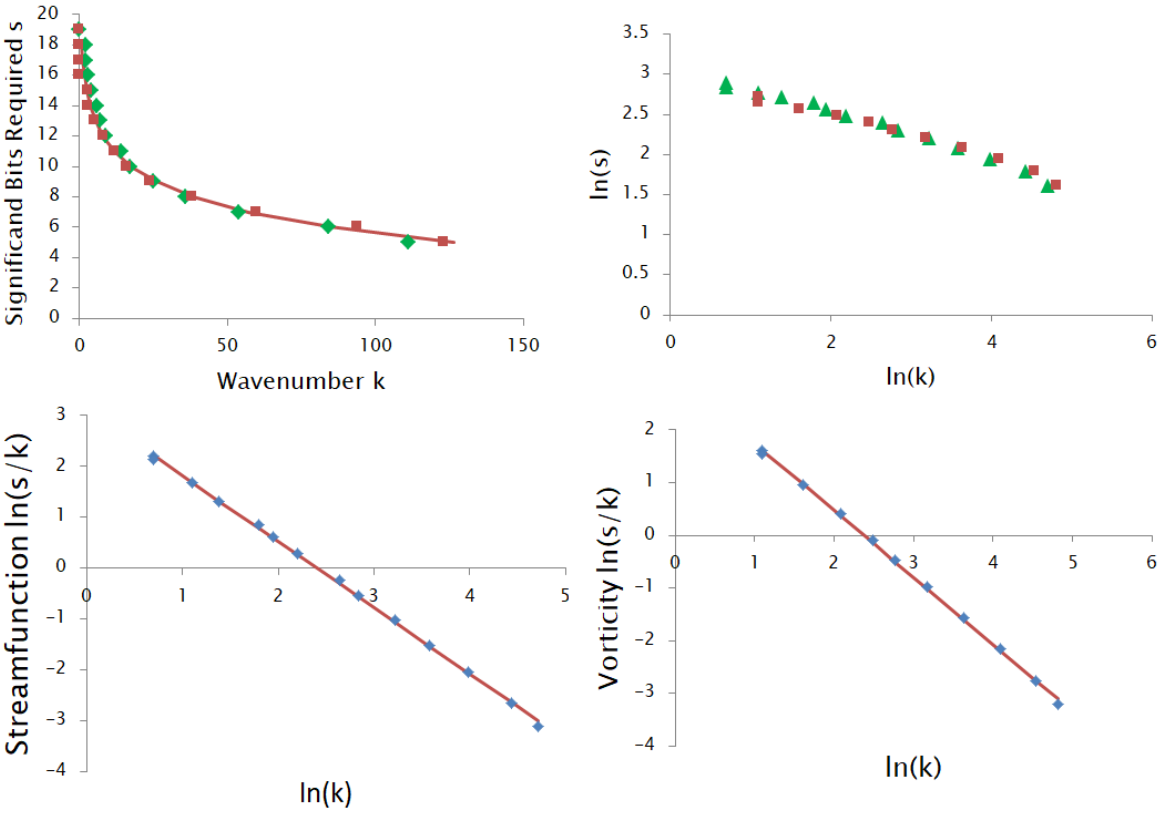

Figure 4.3: The number of significand bits s required for zero error in the streamfunction (green) and vorticity (red, with best-fit trend line) at 53000 timesteps is plotted as a function of wavenumber k (top left) alongside the same data plotted in natural logarithms (top right), demonstrating a near straight line. The natural logarithm of s/k is plotted as a function of k for the streamfunction (left) and vorticity (right) with best-fit straight lines in red, in the bottom two plots.

The observed levels of required precision in both quantities from Table 4.1 are plotted as a function of wavenumber in Figure 4.3; in order to obtain a function describing the extent to which precision can be reduced at each spatial scale, the natural logarithms of the variables are also plotted. These plots reveal a tendency for the precision to curve away from a straight line at high

wavenumbers, implying that a more linear relationship might be obtained by dividing the precision by the wavenumber. Figure 4.3 therefore also includes the natural logarithm of the number of required significand bits per unit wavenumber s/k against $\ln(k)$, which indeed yields an approximately straight line with a best-fit gradient $m = -1.30 + 0.01$ and intercept $c = 3.12 \pm 0.03$ for the streamfunction (these errors are those output by the GNUPLOT fitting program and include the uncertainty in the fit, not in the datapoints' positions) so that,

$$\ln(s/k) = -1.31 \ln(k) + 3.17. \tag{4.1}$$

Very similar results were obtained on plotting the precision required to minimise error in the vorticity, where $m = -1.27 \pm 0.01$ and $c = 3.01 \pm 0.04$.

This result implies that the required precision follows a power-law relationship with wavenumber, whereby, taking exponentials of Equation 4.1,

$$s/k \approx 24k^{-\frac{4}{3}}. \tag{4.2}$$

Plotting the precision per unit wavenumber s_k allows for a more satisfying straight-line fit, but there is no scientific justification for this division by wavenumber, and it may be more useful to those who might potentially implement reduced precision in a real-world setting to notice that this implies $s \propto k^{-1/3}$, a power law that could be directly applied to determine which levels of precision to use. These power law relationships are somewhat akin to the power law behaviour exhibited by the cascading enstrophy per unit wavenumber (see Section 3.3.1), which also describes the relative magnitude of the streamfunction as a function of wavenumber. The dropping off at higher wavenumbers means that smaller scales should make a smaller relative contribution to the overall streamfunction value when the Absolute Error is computed. On the other hand, the EDT also follows a $-4/3$ power law as a function of wavenumber; if errors grow more rapidly at smaller scales, initial errors in these scales might be expected to be more important than those at larger scales. It could be conjectured that these two factors partially counter one another, but that the former factor wins out to give $s \propto k^{-1/3}$ (and hence $s_k \propto k^{-4/3}$). It may also be relevant that the magnitude of the rounding error e_r relative to double-precision induced when the number of bits in the significand is reduced to level s goes as $e_r \propto 2^{64-s}$, a relationship that is not linear.

4.2.3. Experiment 3: Optimised Spectral Precision

The results of Experiment 2 suggest that it may be possible to choose an optimum level of precision to use at each spatial scale during calculations. The precision could either be reduced to the same degree for all variables, or reduced to a different degree for different variables at a given spatial scale. Since leaving the variables other than ψ and ζ in double-precision made no difference to the results in Experiment 2, it is unlikely that the latter approach, which would be more complicated to realise both in emulations and with real-world hardware, would yield much benefit. For this experiment, hence, precision in all the variables was reduced at each spatial scale to the level required for there to be no rounding-error either in the streamfunction or in the vorticity at that scale according to the results of Experiment 2. It was expected that this would induce no additional overall error relative to the double-precision control.

As was the case for Experiment 2, a 'truth' integration was spun up for 50000 timesteps, but for this experiment this was compared to control and reduced precision models for a further 70000 timesteps (around 100 atmospheric days), all of them using the 'true' state after 50000 timesteps as initial conditions. Three types of model were run and tested against the 'truth', all at $k_{\max} = 127$ resolution: a double-precision control model; the scale-selective model; and a model with all wavenumbers reduced to 15 bits in the significand, the computational cost of which would be higher than that for the scale-selective model (see Section 4.4). In each case the results were averaged over fifty ensemble members, each with different random perturbations to the initial streamfunction as in Experiment 1. The Absolute Error relative to the 'truth' was calculated at each timestep in the streamfunction and vorticity separately, in the same way as for Experiment 2, and the results are plotted in Figure 4.4.

The scale-selective model produces far less error than the model with all spatial scales at 15 bit precision, despite the fact that the smallest scales are represented with only 5 bits. When the scale-selective error is optimised for accuracy in the vorticity, there is no error relative to the double-precision control for the first 30000 timesteps (40 days) in the streamfunction, or 20000

timesteps (25 days) in the vorticity; this shows that precision in the smallest scales is indeed much less critical, at least in the case of a short-term forecast, than that in the larger scales and that it is therefore more efficient to use less precision at smaller spatial scales for short- and medium-term forecasts (days or weeks). Thereafter both curves approach the 15 bit significand error, however. The error in the vorticity saturates after around 40000 timesteps, whilst that in the streamfunction continues to grow.

When the precision is optimised to instead avoid error in the streamfunction, the streamfunction predicted by the scale-selective model remains as accurate as that of the double-precision control for the entire simulation apart from a brief period in the middle of the run. The accuracy in the vorticity, meanwhile, rapidly degrades. This is consistent with the notion that optimising precision for the vorticity leaves the streamfunction vulnerable to error at the low wavenumbers, whereas the vorticity requires a higher precision than the streamfunction at high wavenumbers. In light of this, a model with scale-selective precision reduced to the lowest level that incurs no error in either streamfunction of vorticity according to Table 4.2 was also run, producing the results shown on the bottom row of Figure 4.4. However, this yielded no real improvement: the error evolves similarly to the streamfunction-optimised case. This implies that the low wavenumbers (those for which the streamfunction requires a higher precision) dominate in determining the overall error.

Although the scale-selective model performed well in the Absolute Error test for a single run, it is important to verify more robustly that this model accurately predicts the state of the SQG system. For this reason, the Absolute Error, Root Mean Square Error (RMSE) and Ranked Probability Skill Score (RPSS, see below) were computed for five different runs with distinct initial conditions and 10 ensemble members each; Figure 4.5 shows the results averaged over all five runs for the case where precision was optimised in the vorticity.

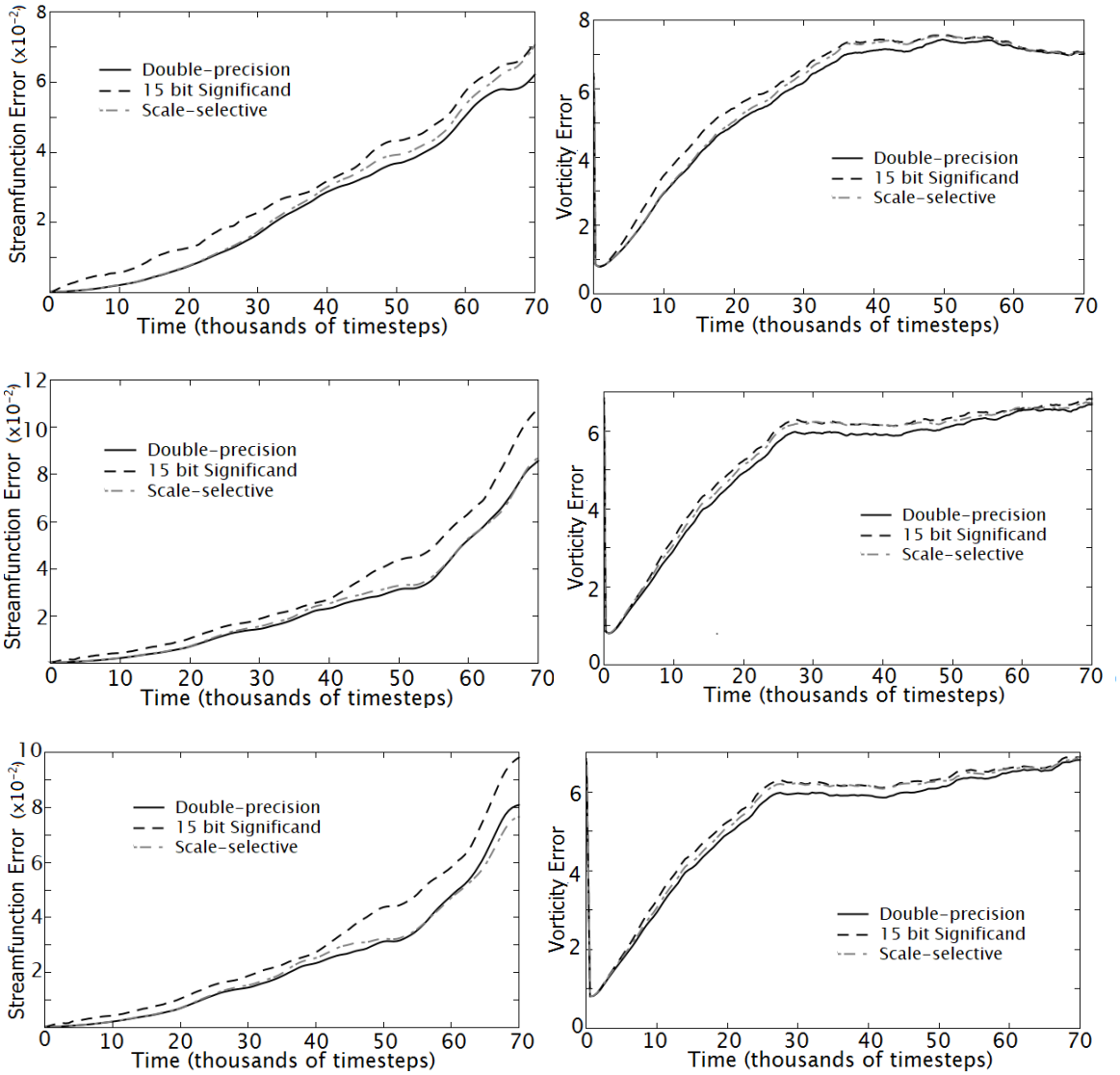

Figure 4.4: The Error in the 25-member ensemble mean streamfunction (left) and vorticity (right) is plotted for 70000 timesteps following a spin-up of 50000 timesteps, for double-precision, 15-bit significand and scale-selective precision optimised to avoid error in the vorticity (top row), the streamfunction (middle row), or set to the lowest level at each wavenumber that incurs no error in either the streamfunction or the vorticity (bottom row) according to Table 4.2, where the Error is measured relative to a single unperturbed 'truth' integration. Each ensemble member differs from the others through a random perturbation to the initial (50000 timestep) streamfunction of up to 5% the initial value, averaging to zero. The sharp initial drop in vorticity error comes as a result of the vorticity field adjusting to this streamfunction perturbation. Results are averaged over five runs with different bottom topography and hence different initial conditions after spin-up.

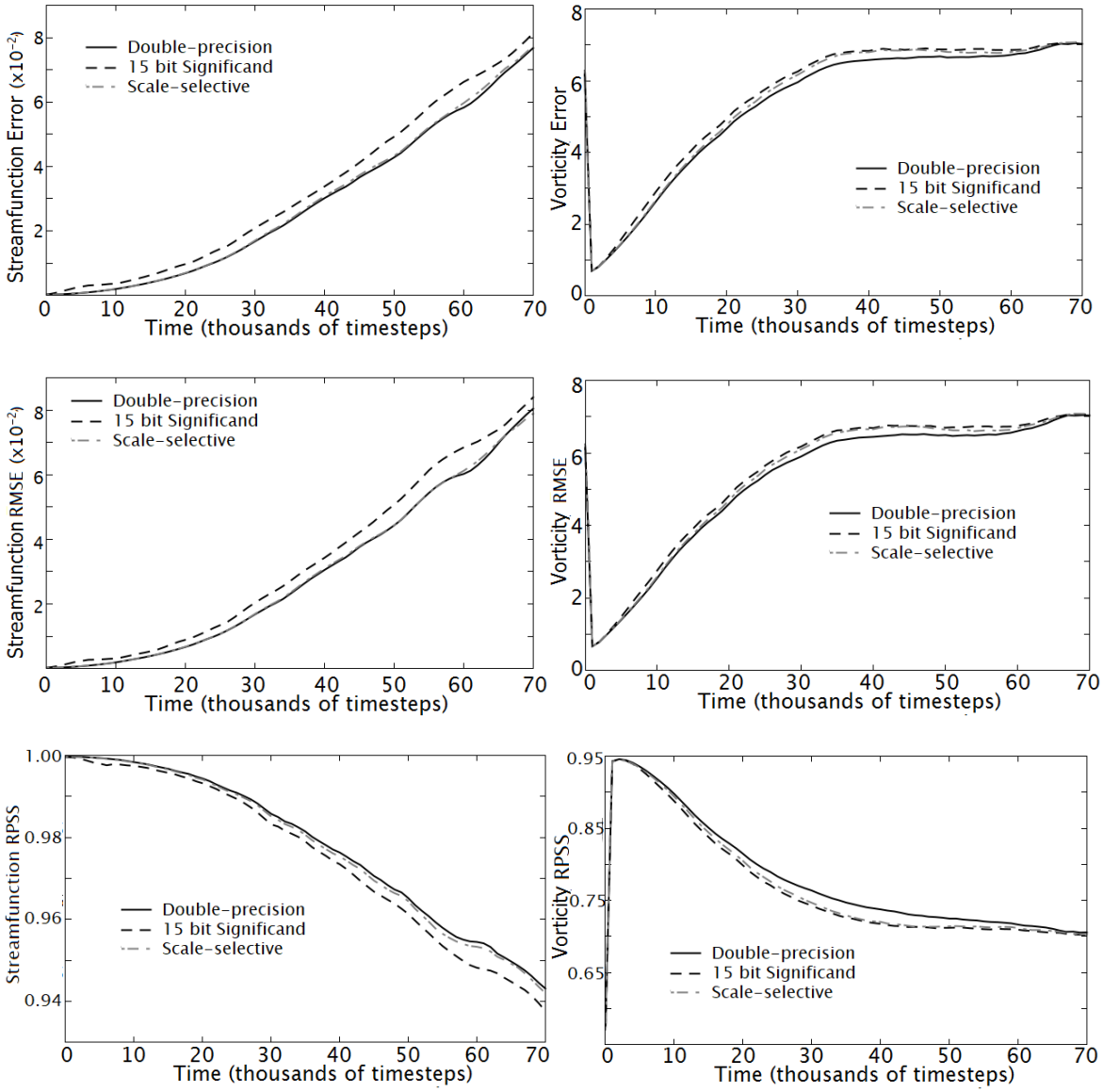

Figure 4.5: The mean Absolute Error (top row), Root Mean Square Error (RMSE, middle row) and Ranked Probability Skill Score (RPSS, bottom row) for the 25-member ensemble are plotted for the streamfunction (left) and vorticity (right) for the double-precision, scale-selective and 15 bit uniform reduced precision models, averaged over 5 runs with different initial conditions.

The Ranked Probability Skill Score (RPSS), outlined in Section 1.7.1, in general provides a more robust assessment of the actual skill, over and above that of a climatological (statistical) forecast, of a dynamical model. Here, the RPSS was applied using ten equally-populated forecast categories to bin the streamfunction or the vorticity values predicted by the double-precision control and each model ensemble, and compare these to the 'true' value. However, the RPSS also

requires a reference forecast, which corresponds to a typical, climatological, distribution of the forecast quantity into the ten categories. Because producing this using a higher-resolution forecast would be computationally costly, and in the SQG system changing the resolution introduces considerable shocks to the system state (see Section 4.2.6), the reference was instead produced using persistence – that is, the 'truth' initial conditions after spin-up but before the streamfunction was perturbed.

The results of the three error or skill metrics averaged over five runs are largely in accordance with those obtained using the single run in Figure 4.4 (the top row is the comparable plot), albeit slightly smoothed by the averaging. The RMSE yields almost identical results to the Absolute Error, and both of these suggest that the scale-selective model streamfunction remains accurate throughout the 70000 timesteps after spin-up, whereas the vorticity begins to lose accuracy relative to the double-precision control after around 20000 timesteps. The RPSS in the streamfunction is consistent with the scale-selective and double-precision models having equal skill, surpassing that of the 15 bit model, throughout, but the scale-selective model loses skill in the vorticity quite rapidly, at around 10000 timesteps. This may mean that small-scale features begin to suffer from reduced precision rounding errors a few 'days' into the forecast.

To check whether a forecast of the 'weather' of the SQG system would be degraded by scale-selective precision, it is also interesting to compare the visual fields output by the models. Figure 4.6 shows the vorticity field at 30000 timesteps after spin-up for the unperturbed 'truth', and the double-precision, scale-selective (optimised to avoid error in the vorticity) and 15 bit models, for the example of the first perturbed ensemble member in the simulations utilised to produce Figure 4.5. The difference between the double-precision perturbed member (top right in Figure 4.6) and the 'truth' (top left) gives an indication of the difference between ensemble members at this timestep. There is little appreciable difference between the scale-selective and double-precision plots and most small-scale features are retained, showing that reducing precision has not degraded the quality of the forecast, so the decline in vorticity RPSS may not be indicative of anything greatly detrimental. The plot for 15 bits in the significand also closely matches the double-precision plot, but not quite as well as does the scale-selective plot.

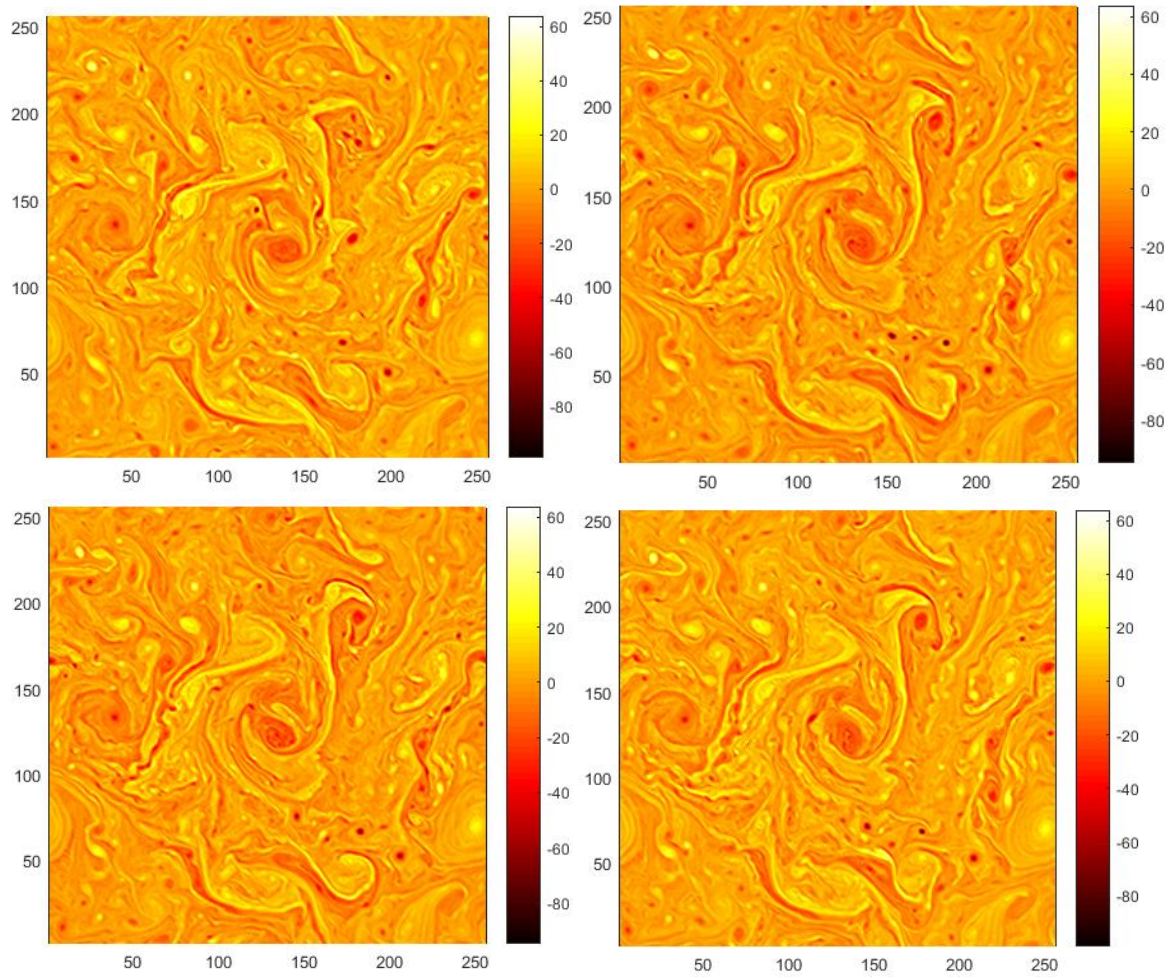

Figure 4.6: Vorticity plots at 30000 timesteps after spin-up for the unperturbed 'truth' (top left), and for a single perturbed ensemble member of the double-precision (top right), scale-selective (bottom left) and 15 bit significand (bottom right) models.

To verify that the optimal level of precision was being applied on each spatial scale, the scale-selective model optimised to reduce the error in the vorticity (Model 'scale-selective') was also compared against a model where the number of bits in the significand was reduced by 1 relative to the scale-selective model at every wavenumber (Model 'scale-selective -1'). For comparison, a model was also run with half-precision in all but the first 4 wavenumbers ($k = 0$ to $k = 3$), which were in single-precision. Figure 4.7 plots the results for the first 5000 timesteps. The scale-selective precision model demonstrably outperforms all the others, implying that it is indeed optimised for the minimum precision required for an accurate output.

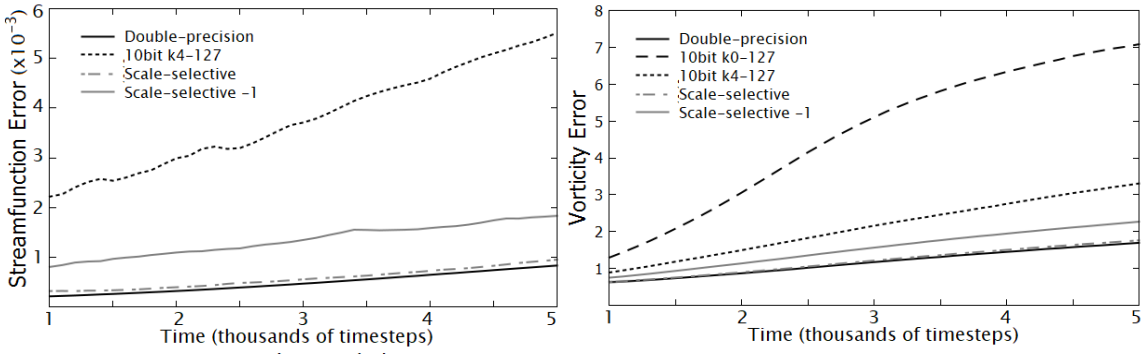

Figure 4.7: The scale-selective model (optimised to minimise error in the vorticity) is compared to a model with one fewer bit in each spatial scale ('scale-selective -1'), a model with 10 bits uniformly (omitted because the error is so great in the streamfunction case), one with 10 bits for smaller spatial scales ($k = 4$ and above) and the double-precision reference, for error in the streamfunction (left) and vorticity (right).

As was noted above, preliminary results indicated that the runs with a half-precision significand (10 bits) were equally accurate regardless of whether a half-precision exponent (5 bits) or single-precision exponent (8 bits) was used. For simplicity, a single-precision exponent was therefore used throughout, and it was assumed that a half-bit exponent could be used for all parts of the model that use 10 or fewer significand bits. However, a greater computational cost saving could be achieved by reducing the exponent as far as possible across the entire model. Figure 4.8 shows the results of running the scale-selective model (conservatively optimised for zero error in both vorticity and streamfunction), or a model with 19 bits in the significand, with in both cases all spatial scales reduced to 5 bits in the exponent. There is no noticeable increase in error in either the streamfunction or the vorticity (not shown) relative to runs with a single-precision exponent (see Figure 4.6), suggesting that the half-precision exponent can indeed be applied to all wavenumbers without penalty, and that in the SQG system the dynamic range of spectral quantities is sufficiently small for all numbers to be representable using just 5 bits, so that their magnitudes must usually lie between 6×10^{-5} and 7×10^{4}. This is consistent with the streamfunction and vorticity spectra plotted in Section 3.2.4. Figure 4.8 also shows the results of running the scale-selective model with just 4 bits in the exponent for spatial scales for which the significand was less than or equal to 10, 6 or 5 bits. A 4 bit exponent makes no difference to the Error, provided that it is restricted to

variables with 5 bits or fewer in the significand, which corresponds to the highest 5 wavenumbers in this case; larger spatial scales evidently require 5 bits in the exponent.

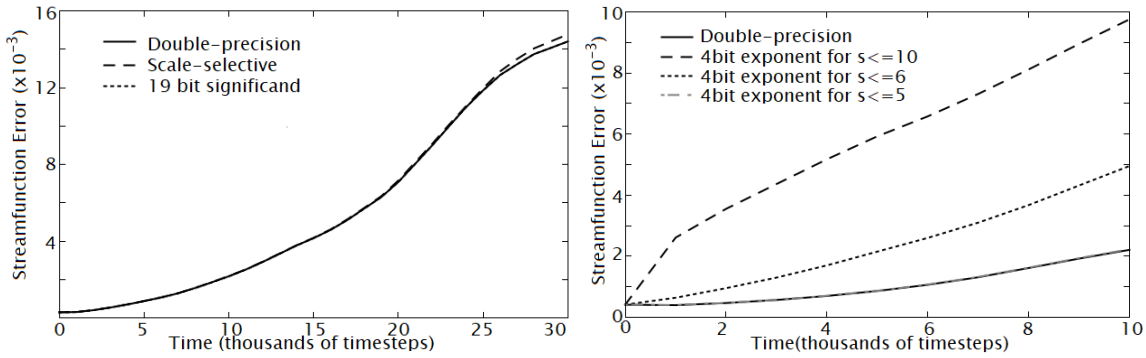

Figure 4.8: Error in the streamfunction for scale-selective and 19 bit significand models with 5 bits used for the exponent (left), and for scale-selective models with 4 bits in the exponent applied to all wavenumbers for which the number of significand bits s is less than 10, 6 or 5 (right), compared to the double-precision control. Vorticity plots are similar and are not shown for conciseness. The 30000 timestep timescale in the left-hand plot is the same as for Figure 4.6; the right-hand plot focusses on the first 10000 timesteps to show the initial divergence of the different models.

Although these experiments were designed to test scale-selective precision in a test-case that resembles, as far as possible, a real-world atmospheric model, it is also interesting to investigate the effects on better-known test-cases, such as the Held et al. (1995) Figure 4 test-case that was reproduced in Section 3.2.1. It must be remembered that the precision optimisation used to construct the scale-selective model above is not strictly applicable to any other setup than Setup S, and that the Held et al. test-case must be run for many more timesteps than Setup S to produce interesting vortex-shedding behaviour, as it has a flat bottom topography except for the central elliptical mountain, reducing the forcing relative to Setup S. Nonetheless, the visual output for this test-case in double-precision at $k_{\max} = 127$ was compared to that of the scale-selective model, and to models with 1 or 2 bits added to the significand across all wavenumbers relative to this, labelled ‘scale-selective +1’ and ‘scale-selective +2’. For comparison, models with a uniform reduction in significand precision were also run, as well as a low-resolution model with $k_{\max} = 63$.

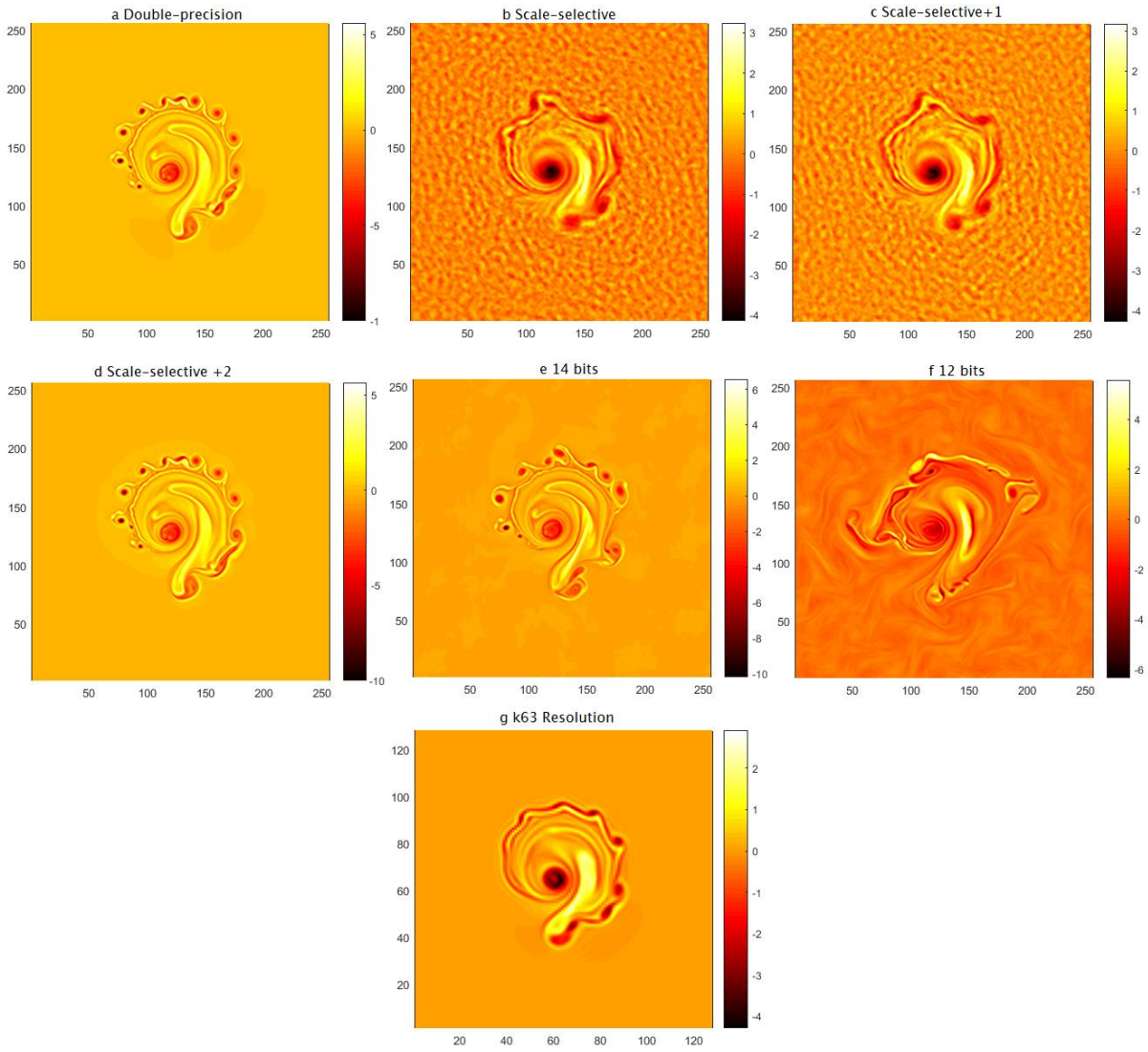

Figure 4.9: The high-resolution 'truth' (a) is compared to six different models forecasting the behaviour of the vorticity field in the Held et al. (1995) test-case (their Figure 4, see section 3.2.1) after 150000 timesteps. The scale-selective model and its variants (b, c and d) are optimised conservatively, with the lowest precision used at each wavenumber without incurring significant Error either in the streamfunction or in the vorticity according to Table 4.1.

Figure 4.9 shows the vorticity fields produced after 150000 model timesteps have elapsed (as no Error Doubling Time experiment has been carried out for this setup, this cannot be converted into model 'days'), when the vortex shedding behaviour is most visible, for the double-precision, scale-selective, scale-selective +1 and scale-selective +2 models, and models with 12 or 14 bits in the significand for all wavenumbers. The scale-selective model reproduces the 'truth' very poorly,

and there is negligible improvement when 1 extra bit is added at each wavenumber, but adding 2 bits (Figure 4.9d) yields an almost perfect agreement with the 'truth', one which continues for at least another 150000 timesteps. This result is better than that achievable with all wavenumbers reduced to 14 bits significand precision (Figure 4.9e), and than the low-resolution alternative (Figure 4.9g). The Figure demonstrates that good results cannot be obtained by ignoring the small scales entirely, but scale-selectively reducing precision can yield better results for the same or better computational cost savings relative to high-resolution double-precision, so long as the scale-selective optimisation is tuned to the test-case in question.

4.2.4. Experiment 4: Fast Fourier Transforms

Not all the calculations within an atmospheric model take place in spectral space. For the efficiency of computing resources to be maximised, it is desirable that reduced precision be also applicable to other components, which often constitute the majority of the computational cost, albeit without the possibility of exploiting distinctions between different spatial scales as is possible in spectral space. This experiment therefore explores reducing precision in the Fast Fourier Transform (FFT) component of the SQG system, used to transfer information between the spectral and grid-point calculations carried out at each timestep, which is responsible for more than thirty per cent of the computational cost of the system. Calculations within spectral space were fixed at single-precision for $k \leq 10$ and half-precision for $k > 10$, which Experiment 1 showed should have no effect on the short-term forecast accuracy, but in contrast to previous experiments, the number of bits in the significand was also changed for the FFT parts of the model to 23 (single-precision), 16, 14, 12, 11 or 10 bits. A half-precision run with the exponent reduced to 5 bits for a 10 bit significand was also performed. In each case, the 'truth' was compared to ten ensemble members, again each with perturbed initial conditions, and Absolute Error scores were computed for runs of 2000 timesteps.

To reduce the precision in the FFT, it was necessary to use a different algorithm to the very fast 'FFTW' script usually employed, because FFTW is not compatible with the reduced precision

data types used by the emulator. Instead, a Fortran program based on an existing algorithm (Press et al. 1992) was written to perform the FFT in a way that enabled precision to be reduced but which runs considerably more slowly than FFTW, although it produces an identical output in double-precision. Hence, running-time constraints restricted the ensemble size to 10 members, with perturbations applied to a single initial condition, and meant that not all of Experiments 1-3 could be repeated in conjunction with a reduced precision FFT. This would not be a problem with real reduced precision hardware, which would not require an emulator data type.

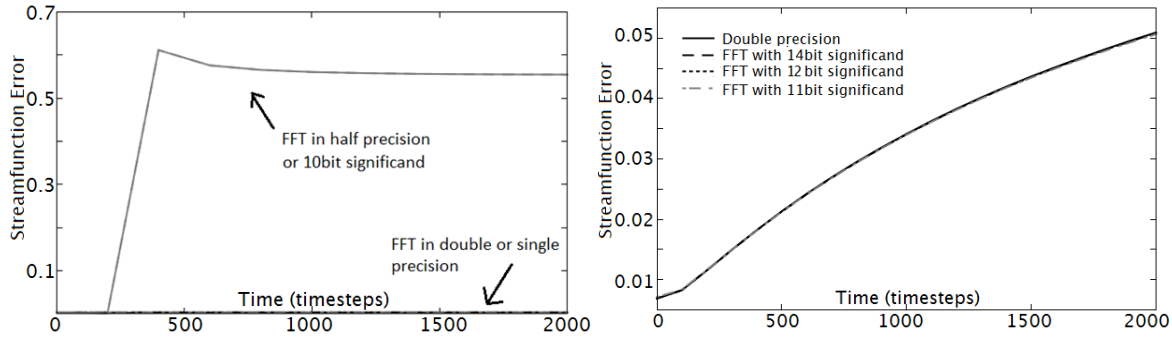

Figure 4.10: The Absolute Error in the streamfunction induced by carrying out only the Fast Fourier Transforms (and inverse transforms) in the model at various levels of precision are plotted. In all cases, the spectral variables in the model are resolved in half-precision above wavenumber $k_t = 10$ and in single-precision below this and grid-point calculations are in single-precision. Data are plotted every 100 timesteps. In the left plot, the FFT in double-and single-precision and with the significand reduced to 10 bits are compared. The double-precision control and single-precision errors (lower curve) are identical, and are much smaller than those induced by truncating to half-precision or 10 bits in the significand (upper curve), which are very great after 300 timesteps have elapsed. In the plot on the right, the significand is reduced to 14, 12 or 11 bits for the FFT; clearly this makes no difference relative to the control, indicating that the FFT works well so long as it is performed with more than 10 bits in the significand.

Figure 4.10 shows the results. Reduction to single-precision made no difference to the error score, and indeed the streamfunction field when plotted visually appeared identical to the double-precision case. There was virtually no error induced with the significand reduced to as little as 11 bits. Carrying out the transforms using 10 bits, on the other hand, gave rise to considerable errors after 300 timesteps had elapsed. The delay implies that not all the FFTs in the model require more

than 10 bits in the significand, but a critical point in time is reached after which at least one FFT requires a higher precision than is available, giving rise to considerable error that is an order of magnitude larger than the initial condition error exemplified by the double-precision control and hence immediately reaches saturation point, with no further error growth possible. This does not occur any earlier for the exponent-truncated half-precision case, suggesting that the FFTs do not require more than 5 bits in the exponent to avoid crashes.

4.2.5. Experiment 5: Optimised Spectral, Grid-point and FFT Precision

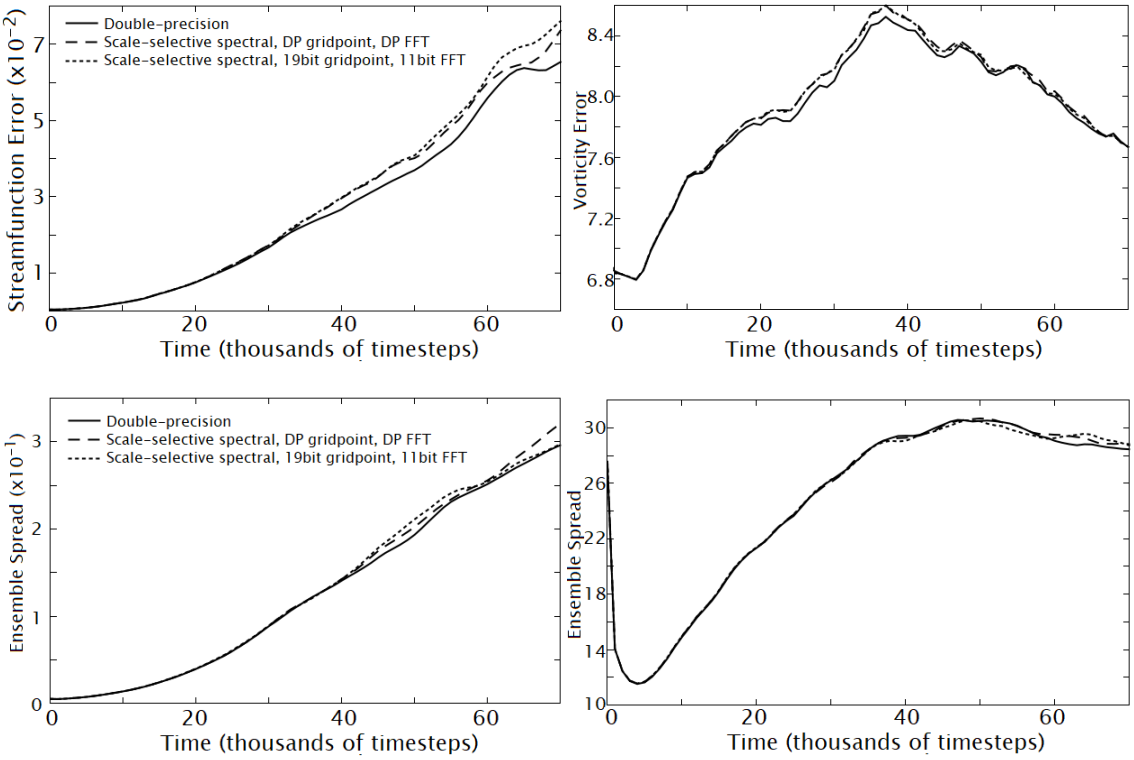

Figure 4.11: The Absolute Error (top row) in the five-member ensemble mean and the ensemble spread (bottom row) in the streamfunction (left) and vorticity (right) is plotted for runs in double-precision (black lines), with scale-selective precision in the spectral component and the FFT and grid-point components in double-precision (dashed lines) and with scale-selective spectral precision and reduced precision in the FFT and grid-point components (dotted lines).

The above experiments suggest that precision can be reduced in the spectral part of the SQG model in a scale-selective way, in the grid-point part to 19 bits (the precision required for the lowest

wavenumber) because scale-selectivity cannot be applied there, and in the Fast Fourier Transform part to 11 bits in the significand. For Experiment 5, a run was performed to verify that the model can successfully function with all three components in reduced precision simultaneously. Five ensemble members were used, each run for 70000 timesteps after spin-up, which was sufficient to assess the performance of the model. Figure 4.11 shows the streamfunction and vorticity Absolute Error and ensemble spread for runs in double-precision and scale-selective precision with reduced precision in the FFT and grid-point components switched on or off (similar to Figure 4.6). The graphs demonstrate that switching this on introduces little additional error relative to the run with only scale-selective precision in spectral space, and that reducing precision in all three components of the model does not significantly affect the ensemble spread, yielding a forecast that matches double-precision out to at least 30000 timesteps, which is of the order of 40 atmospheric days – beyond the range of accurate real-world weather forecasts. RMSE plots were almost identical to the Absolute Error plots and are omitted for conciseness; the equivalent streamfunction RPSS plot is provided in Figure 4.12 and shows similar results.

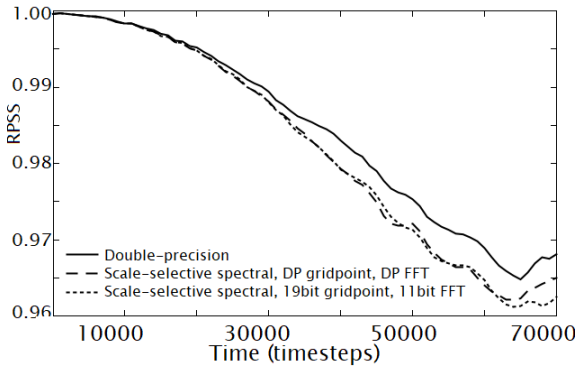

Figure 4.12:The Ranked Probability Skill Score (RPSS) is plotted for the same models as in Figure 4.11, for the streamfunction, with persistence used as the reference forecast. The equivalent plot for the vorticity is omitted because a vorticity RPSS reference forecast could not be generated successfully.

4.2.6. Experiment 6: Attempting to Trade Resolution against Precision

In the Lorenz ’96 experiments, it was demonstrated that a high-resolution Mixed Precision model with medium-scale variables in half-precision could outperform a lower-resolution model in

double-precision for similar estimated computational costs (see Section 2.5). To make this test relevant to real-world operational analogues, unresolved variables were parameterised in these models; without this, lowering the resolution would have an unrealistically deleterious effect. It is interesting to explore whether a similar result holds for the SQG system, but because this system is of a very different nature to Lorenz '96, involving variables that evolve across multiple spatial scales, the parameterisation approach in the SQG case must be adjusted relative to Lorenz '96.

In the SQG system, 'high' and 'low' resolution models can be obtained by using different ranges of wavenumbers – for example, by resolving wavenumbers between $k = 0$ and $k_{\max} = 127$ in a high-resolution model and only resolving wavenumbers up to $k_{\max} = 63$ in a low-resolution alternative. If the first 64 wavenumbers are initially identical in both models, they will both represent the same atmospheric situation, but will begin to deviate from one another over time because of the absence of the smallest scales in the low-resolution model. This deviation can be reduced by using parameterisation to 'correct' the lower wavenumbers in the low-resolution model at every timestep towards the values that the high-resolution model would be expected to take. Empirical parameterisation schemes of this kind are tuned by running the model at low resolution for a short period of time over which it is computationally affordable to also run a high-resolution model from the same initial conditions and comparing the two. The parameters required to correct the low-resolution model in this run are then assumed to be the same as those required to correct low-resolution runs in all similar circumstances.

This methodology was employed to develop a deterministic parameterisation scheme for SQG akin to those developed for the Lorenz '96 case by Wilks (2005) and Arnold (2013). The method is described in the Appendix, alongside some of the caveats that apply for the SQG system. The 'truth' was defined by running the model at a high resolution, with $k_{\max} = 127$. After a 150000 timestep spin-up, this 'truth' run continued for 2000 timesteps, over which it was compared to three different types of model. First, there was a medium-resolution ($k_{\max} = 63$) control model in double-precision with parameterisation employed to represent the expected unresolved effect of wavenumbers 64 to 127. Second, there were equivalent, parameterised models but in scale-selective reduced precision, with wavenumbers below the truncation wavenumber k_t represented in

single-precision and those above k_t represented in half-precision. k_t could be varied between 1 and 63 for different models of this kind. Third, there was an equivalent model in double-precision but at low-resolution ($k_{\max} = 31$) and wavenumbers above 31 parameterised. The same three model types were also run with parameterisation switched off for comparison. The models were compared to the 'truth' by computing the Absolute Error of the ensemble mean streamfunction in grid-point space, averaged across all resolved grid-points. A fifty-member ensemble was used, with initial conditions perturbed by up to 5% as in Experiment 1.

The left-hand plot in Figure 4.13 shows the results for $k_t = 1$ with parameterisation carried out in spectral space. The figure shows that even with all wavenumbers $k \geq 1$ reduced to half-precision there is no loss of accuracy for scale-selective precision compared to the double-precision control, whereas reducing the resolution has a very profound effect. This suggests that errors introduced by lowering from the high resolution of the 'truth' to medium resolution are more important to the overall error in reduced precision parameterised models of the system than is any error induced by reducing precision. The fact that parameterisation is necessary for a fair test is shown by the right-hand plot in Figure 4.13, which demonstrates that with parameterisation switched off the low-resolution model performs very poorly indeed.

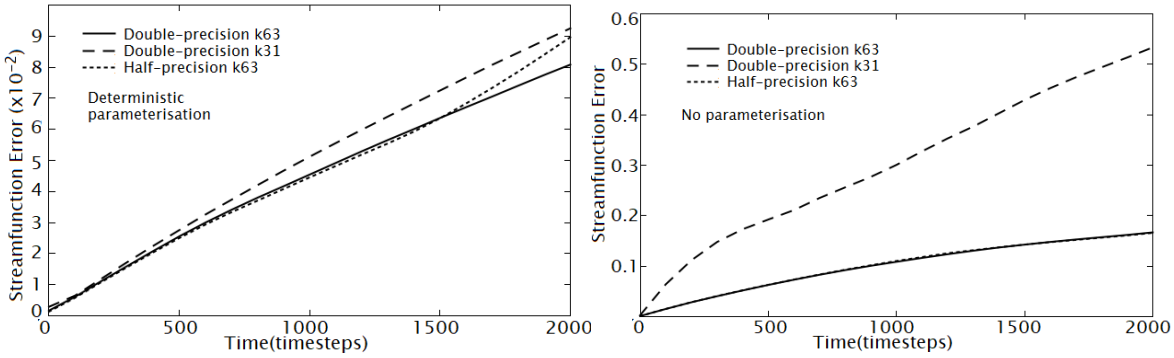

Figure 4.13: The streamfunction ensemble mean Absolute Error score is plotted for models of Setup S in $k_{\max} = 63$ resolution double-precision, $k_{\max} = 31$ resolution double-precision and $k_{\max} = 63$ resolution single- (for $k < k_t$) and half- (for $k \geq k_t$) precision with $k_t = 1$, under deterministic parameterisation (left) or no parameterisation (right), applied in spectral space.

In order to analyse the relative computational cost of the models compared here, it can be assumed that the cost is proportional to the number of wavenumbers resolved explicitly, and to the number of bits used for each wavenumber. Because there are two dimensions in wavenumber space, halving the overall resolution $k_{\max}$ means resolving roughly one quarter of the number of one-dimensional wavenumbers. Where the double-precision control cost, with $k_{\max} = 63$, is 1, the low-resolution double-precision models, where $k_{\max} = 31$, have a cost of roughly 0.25. The medium-resolution low-precision models with $k_t = 1$ have only 1 combination of one-dimensional wavenumbers ($k_x = k_y = 0$ so that $k_r = 0$) in single-precision (32 bits total) and the remainder of the 64^2 one-dimensional wavenumbers in half-precision (16 bits per wavenumber), so the cost is $(1/64^2) \times (32/64) + (1 - 1/64^2) \times (16/64) = 0.25$. This means that this scale-selective precision model would have the same approximate cost as the low-resolution model on maximally efficient Mixed Precision hardware, assuming no overheads associated with reducing the precision.

An important caveat to the above results is that parameterisation was carried out in spectral space and that all wavenumbers were treated equivalently, despite referring to different spatial scales that have different magnitudes and ought theoretically to require independent 'corrections' according to other studies (Malardel and Wedi 2016), as described in the Appendix. Attempts to derive a wavenumber-dependent spectral parameterisation scheme, or to parameterise the variables in grid-point space, both of which would have a greater physical justification, were unsuccessful. It is puzzling that the spectral parameterisation of variables worked so well here despite not being dependent on the wavenumber, but it is beyond the scope of this study to explore this question in more detail. Since no explanation can here be offered for this, nor any guarantee that the parameterisation schemes used here are optimal, confidence in the results is limited and the conclusion that higher-resolution models outperform higher-precision ones in the SQG system remains tentative.

4.3. Experimental Results: Long-term Forecasts

An investigation into the ability of scale-selective precision to improve the efficiency of longer-term forecasts was performed using Setup L, in which there was no central mountain in the bottom topography and the long-term properties of the system (such as the total energy) were constant. Owing to the time required to run the model at a high enough resolution for long integrations, the optimum precision levels at each wavenumber for good long-term forecasts were not explicitly measured. Instead, the levels obtained in Experiment 2 were used to create a 'scale-selective' model, which could then be compared against models at other levels of precision.

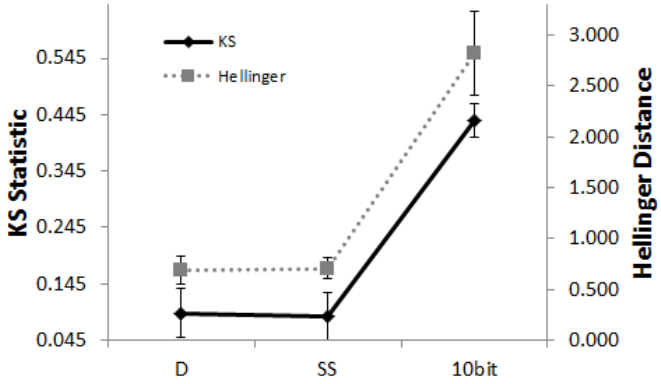

Figure 4.14: The Kolmogorov-Smirnov (KS) and Hellinger Distance statistics (see Section 1.7.2) measures of the difference between the model and 'truth' climatology, computed by comparing model and truth pdfs across 250000 timesteps (around 1 year) for three different ensemble members, are plotted for double-precision (D), scale-selective (SS) and 10bit reduced precision models of the system. In each case, error bars were computed by averaging over four runs of the system with different randomly-forced bottom topography that supply the initial conditions after spin-up.

Similarly to the short-term forecast investigation, the model was first spun up for 50000 timesteps in double-precision. This time, however, it was then run for a further 500000 timesteps to create a 'truth' timeseries. To create a control double-precision ensemble 'forecast' of this truth, to the initial conditions after the spin-up a random perturbation amounting to up to five per cent of the unperturbed values was introduced in each wavenumber component of the streamfunction separately for each of three different ensemble members. Reduced precision forecasts were created

in the same way, except that the emulator was now switched on. Using all timesteps in the second half of the run after spin-up, the streamfunction values at each wavenumber and for each ensemble member were binned by magnitude using 1000 bins of width 0.001, which spans the typical orders of magnitude of this variable. All wavenumbers were distributed amongst the same set of bins. This allowed a pdf, taking into account all ensemble members, to be created for each model, providing a measure of the long-term mean properties.

By comparing these pdfs to the pdf computed for the 'truth', the Hellinger Distance and Kolmogorov-Smirnov (KS) statistics could be calculated to measure the difference from the truth as outlined in Section 1.7.2. These were computed for four different runs of the whole system, differing in the bottom topography used, and the results were averaged across these four runs. The standard error was calculated as a measure of uncertainty in the quoted values. The first half of the run (the first 250000 timesteps) was not included in this calculation so as to allow the perturbed models to acclimatise before any statistics were measured; the remaining 250000 timesteps that were included equate to approximately one atmospheric year.

Figure 4.14 shows the results for the double-precision control; the scale-selective model tuned to optimise precision in the streamfunction according to the results of Experiment 2, where the highest wavenumbers are represented with just 5 bits; and a model with the significand reduced uniformly in precision to 10 bits. The KS and Hellinger statistics give the same outcome: the scale-selective model performs as well as the double-precision control within experimental error, whereas the 10 bit model performs much worse, with pdfs much further from the truth. This demonstrates that scale-selectively reducing precision need not have very negative impacts on the ability of a model to forecast long-term mean conditions, which ability might be more strongly degraded by reducing uniformly to half-precision (10 bits in the significand). In this experiment, the exponent part of number was kept at 8 bits throughout; only the significand was truncated. Because of the expense of performing such long runs, a half-precision model (with the exponent reduced to 5 bits as well as the significand to 10 bits) was not tested, but as explained in Section 4.2 it is not expected that this would perform very much worse than the 8 bit exponent model.

4.4. Estimated Potential Savings in Computational Cost

To quantify the potential savings in computational cost (in terms of computer run-time) that could be achieved by applying reduced precision without reducing the accuracy of the model output, the 'gprof' software profiling program (Fenlasen and Stallman 1997) was used to estimate the time spent on different operations when the SQG main program was run without the reduced precision emulator. It was found that around 50% of the computational time is spent carrying out calculations in spectral space; 15% in grid-point space and 35% in FFTs between these spaces. Each FFT or inverse FFT contributes around 5% of the overall time to run the program.

The potential computational cost savings achievable with scale-selective reduced precision in the SQG setup S were computed for each of these three components separately. It was assumed always that there would be no overhead cost associated with specifying which level of precision to use in real Mixed Precision hardware, an assumption justified by the fact that, since such hardware does not yet exist, no realistic estimates of this overhead can be made, and it could in principle be close to zero. The savings estimates presented here would be valid for the most efficient imaginable hypothetical hardware, which would perhaps be realisable in the far future.

To calculate the cost saving of Mixed Precision in spectral space, each of the 128 wavenumbers was assigned to a 'band' according to the minimum number of significand bits s with which it could be represented without faults according to the results of Experiment 3. It was assumed that in double-precision all wavenumbers k_r make a contribution to the overall cost in proportion to the number of combinations of 1D wavenumbers k_x and k_y for which $k_x^2 + k_y^2 \approx k_r^2$, which is proportional to the circumference $2\pi k_r$ of the circle of radius k_r in 2D wavenumber space. Thus, the fractional contribution of each band can be computed by summing the wavenumbers in that band and dividing by 8256, which is the sum of all numbers from 1 to 128 (1 is added to all wavenumbers from 0 to 127 to account for wavenumber 0). Multiplying this by the fractional cost C of calculations for the appropriate value of s for that band then yields the cost of the band. This calculation is illustrated in Table 4.2, where also the overall cost is computed by

summing over all wavenumber bands. It is assumed that the exponent can be reduced to 5 bits with impunity for all wavenumbers, that the sole sign bit is preserved, and that reducing the significand to s bits would thus correspond to a cost saving relative to the 64-bit double-precision control of,

$$C = (s + 5 + 1)/64. \tag{4.3}$$

The number of significand bits permissible at each scale is different for the streamfunction and vorticity, but both yield roughly the same overall computational cost, 20% of the double-precision control. Were all the spectral components to be reduced to the same level of precision, a similar 80% cost saving would require them all to be represented with just 10 bits (that is, roughly 20% of the 52 bits used in double-precision) in the significand. Section 4.2.3 showed that even a model with a uniform reduction to 15 bits in the significand would incur significant errors, let alone 10 bits, so the scale-selective approach offers a much higher computational efficiency.

s	C	ψ Σk	Cost %	ζ Σk	Cost %
19	0.391	3	0.01	0	0
18	0.375	0	0	0	0
17	0.359	3	0.01	0	0
16	0.344	4	0.02	6	0.02
15	0.328	11	0.04	0	0
14	0.313	7	0.03	9	0.03
13	0.297	17	0.06	21	0.08
12	0.281	60	0.20	42	0.14
11	0.266	48	0.15	58	0.19
10	0.250	172	0.52	164	0.50
9	0.234	341	0.97	441	1.25
8	0.219	819	2.17	1089	2.89
7	0.203	2085	5.13	2635	6.48
6	0.188	2646	6.01	3161	7.18
5	0.172	2040	4.25	630	1.31
Sum		**8256**	**19.58**	**8256**	**20.07**

Table 4.2: For each number of bits in the significand s, the fractional cost C relative to a 64-bit control is computed, assuming 5 bits in the exponent and 1 bit for the sign of the reduced precision number, according to Equation 4.3. The sum of the wavenumbers, Σk, represented at each level of precision in the scale-selective models developed in Experiment 2 (see Table 4.1) is listed for cases where either the error in the streamfunction ψ or that in the vorticity ζ is used to determine at what range of wavenumbers each level of precision can be applied. This number as a fraction of 128 is multiplied by the fractional cost column to obtain the percentage cost associated with each band relative to the total cost across all wavenumbers in double-precision; the total cost is the sum of this over all bands, shown at the bottom of the table.

For the FFT part of the model, where scale-selective precision cannot be applied, Experiment 4 revealed that precision could be reduced to 11 bits in the significand and 5 bits in the exponent with no noticeable effect on the model output. According to the row for 11 bits in the significand in Table 4.1, this would reduce the cost to 26.6% of the double-precision control cost. This is almost as great a saving as the 20% achieved with scale-selective precision in the spectral part of the model, but these could not be represented with 11 bits in the significand without incurring considerable errors, requiring scale-selective precision to obtain a similar saving.

In order to apply scale-selective reduced precision to calculations carried out in grid-point space, it would be necessary to transform the variables computed in grid-point space into spectral space before and after each calculation, and apply scale-selective reduced precision to the transformed quantities. However, carrying out additional FFTs incurs considerable computational cost, and it is more computationally efficient to instead uniformly reduce the precision of variables in grid-point calculations to 19 bits in the significand, a level that rendered zero error in either ψ or ζ for even the largest-scale variables according to the scale-selective analysis of Experiment 2, in which case the grid-point calculations would require just 39% of the double-precision cost.

Under the assumption that the measured computer time spent in spectral space, grid-point space and FFTs is proportional to the computational costs of these components, these figures can be combined to yield an overall computational cost relative to double-precision according to the calculation represented diagrammatically in Figure 4.15. Overall, up to 75% of the running cost of the model could be saved by performing spectral calculations in scale-selective precision, grid-point calculations with 19 bits in the significand and FFTs with 11 bit significands.

Reduced precision would also have an impact on the fields stored by the model. If these fields are stored in spectral space, there would be an 80% saving in the storage space required, assuming that no additional storage space is needed to decipher which wavenumbers have which level of precision. This information would not change between runs, and could therefore be stored as a single, small file, ready to be utilised for conversion of the output data if it was needed in a non-scale-selective manner. If data are stored in grid-point format, reducing to 19 bits in the

significand across all spatial scales would reduce the amount of space required by 61% compared to double-precision.

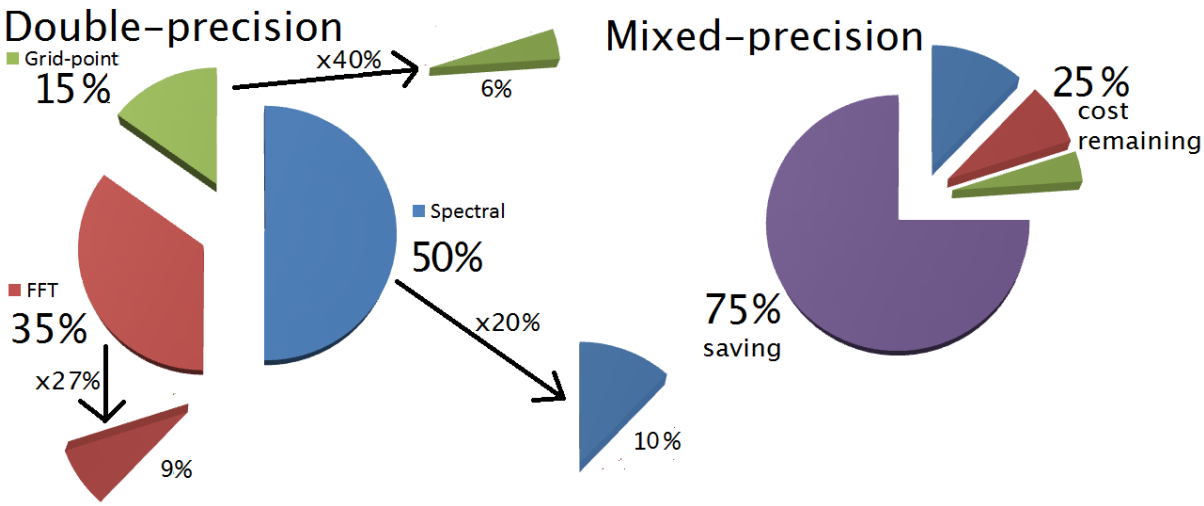

Figure 4.15: Diagram representing how the computational cost associated with the three components (spectral, grid-point and FFT) of the program used to solve the SQG equations can be reduced using 5 bits in the exponent, scale-selective precision in the significand for the spectral part, and reduction to 19 and 11 bits in the significand for the grid-point and FFT parts respectively, assuming no overhead costs.

4.5. Conclusions from the SQG System

Before exploring the impact of reduced precision in the SQG system, an analysis was carried out into the nature of scale-dependent turbulence and error propagation in that system using a codification based on that of Smith (2009) and a system setup designed to simulate realistic weather phenomena with real-world-like bottom topography (see Chapter 3). By analysing the propagation of errors between spatial scales and measuring the Error Doubling Time (EDT) $\tau_d(k)$ as a function of wavenumber k (Section 3), it was confirmed that the SQG model exhibits a -5/3 power law in the saturation error with wavenumber, and that this is reflected in the EDTs through the relationship $\tau_d(k) = (5.42 \times 10^4)k^{-4/3}$. Initial errors therefore grow faster (the time for error to double is smaller) at smaller scales than at larger scales, but also have a smaller maximum value. This power law is akin to that exhibited by the enstrophy per unit wavenumber as a function of

wavenumber in the SQG system (Pierrehumbert et al. 1994), and by the energy per unit wavenumber above 400km in the real atmosphere, and is characteristic of three-dimensional turbulence. A similar power law, $\delta\psi(k) \approx 1.7k^{-5/3}$, was observed for the maximum (saturation) error $\delta\psi(k)$ induced at each spatial scale. Using these results, it was estimated that 1 timestep in the SQG setups used here is equivalent to approximately 2 minutes of real atmospheric time.

In light of such results, it was expected that errors of a similar relative size would be more detrimental to the output of the model on weather forecasting timescales if introduced at larger spatial scales, and that high precision would be less critical for the higher wavenumber components of the prognostic variables. To test this hypothesis, experiments were carried out into the effects of reduced precision on short- and medium-range forecasts of the model's prognostic variables.

In spectral-space calculations, it was found that, whilst half-precision could be applied to all spatial scales above $k = 8$ without incurring error, a more efficient forecast could be achieved by scale-selectively varying precision in the prognostic variables, with 19 significand bits at the largest scales but as few as 5 bits in the significand for the smallest scales. The minimum precision required in the streamfunction s varied with k according to $s \approx 24k^{-1/3}$. If such a scale-selective model for the spectral calculations were to be combined with a 19-bit significand for grid-point calculations and an 11-bit significand for FFTs, the minimum precision required for no error, the computational cost of the model could be reduced by an estimated maximum of 76% with no loss of short and medium-term (days or weeks) forecasting accuracy. The cost saving relative to double-precision was only 61% in grid-point space, as compared to 80% in spectral space, because of the impossibility of scale-selectively reducing precision in the former case. Whilst it has been generally assumed that grid-point calculations become increasingly computationally advantageous at higher resolutions, because in double-precision the cost scales more steeply with resolution for spectral models, the efficiency offered by scale-selective precision in spectral space that is demonstrated here could mean that spectral models would instead be favoured even at very high resolutions were Mixed Precision hardware to be used.

It is interesting that the required precision follows a power law very similar to the EDT, saturation error and enstrophy spectrum power laws when these quantities are plotted against wavenumber. This result implies that to retain accuracy the precision should be reduced less steeply than the increase in turbulence towards smaller spatial scales, suggesting that turbulence is not the sole determinant of how much precision is required. Nevertheless, the result conforms with the hypothesis that precision is less critical on the scales where errors grow more rapidly, a finding that could be of interest to users of real-world spectral models and that incentivises the construction of real hardware capable of implementing scale-selective precision. The next stage in this investigation will be to scale-selectively reduce precision in a real-world model and to analyse whether similar results are obtained.

5. Experiments in a Full-Complexity System: The Open IFS

Chapters 2 and 4 of this thesis have demonstrated that reduced precision techniques could be used (with appropriate hardware) to reduce the cost of simulations without degrading the accuracy, or allow forecasts to be made at a higher resolution than would otherwise be affordable, in idealised models. For any efficiency saving to be translated into an operational context, however, so that real benefits can be brought about for users of weather and climate forecasts, it is necessary to demonstrate the efficacy of reduced numerical precision in a full-complexity forecasting system. The scale-selective approach with which this thesis is primarily concerned is only applicable to models that have a spectral component, such as the Integrated Forecast System (IFS) developed and used operationally by the European Centre for Medium-range Weather Forecasts (ECMWF). Implementing reduced precision in such a full-complexity operational model of the real atmosphere is a natural progression from the experiments presented so far in this thesis.

In this Chapter, the IFS is described (Section 5.1) alongside previous attempts to implement reduced precision in this context (Section 5.2) and the mechanism by which the reduced precision emulator is applied (Section 5.3). Forecasts of four different test-cases are compared in single-precision and reduced precision to identify how low the precision can be reduced at each wavenumber without incurring significant errors, using the standard deviation between ensemble members in the ECMWF forecasts as a baseline uncertainty. This enables scale-selective reduced precision models to be constructed that follow a similar power law relating precision to spatial scale to that observed in the SQG system, and it is demonstrated that these models can be used to accurately forecast weather conditions pertinent to real-world scenarios (Section 5.4). It is found that delaying the onset of reduced precision may in some cases be beneficial. These results could have an important bearing on how to implement real scale-selective hardware in operational forecasting models in the future (as outlined in Sections 5.5 and 5.6).

5.1. The ECMWF Open IFS

The ECMWF has produced ten-day global forecasts operationally (for real-time use) since August 1979. The numerical computer model on which these forecasts are based is the Integrated Forecast System (IFS), which is under a continual process of revision and improvement by researchers at ECMWF and Meteo-France. At the time of writing, the latest operational model is version 43r3, released in July 2017, which produces a single ten-day global forecast at 9 km horizontal resolution with 137 vertical levels and a fifteen-day fifty-member ensemble global forecast at 18 km horizontal resolution with 91 vertical levels (ECMWF 2017a). All versions of the IFS since 1991 use some form of reduced Gaussian grid, which reduces the number of grid-points near Earth's poles to maintain a roughly constant distance between grid-points at all latitude circles, preventing numerical instabilities (Hortal and Simmons 1991).

The model is run using one of two Cray supercomputers installed at ECMWF in 2016, which each have over 3000 compute nodes and give a peak performance of 8.5 petaflops (ECMWF 2017b), requiring a total power consumption of 1.9 MW each (Top500.org 2016). Increasing the efficiency of the computations would enable the resolution, the number of ensemble members or the complexity of the IFS model to be increased, potentially improving the forecast accuracy, without needing to increase computer energy consumption and running costs. In particular, it is a key source of forecast error that the model cannot resolve individual convective clouds, which are of the order of 1 km in size, and phenomena such as these have to be parameterised instead.

Although the IFS itself can only be run at the ECMWF, the centre makes available a portable version of the model, called 'Open IFS', which is available for licensed researchers to download and use remotely. The version of Open IFS that is employed here is based on the '38r1' release of IFS, which was used operationally until June 2013, in a fully-functional format except for the data assimilation and ensemble-forecasting components. The user can select which resolution and output parameters to use, and may specify the start-date of the forecast by generating initial condition files on the ECMWF computer system and copying these to use with their local

installation. When run on a local machine, it is inevitable that the IFS may have to use a lower resolution than is possible operationally; in this study, for example, the maximum resolution at which Open IFS is run is 'TL255', which corresponds to 256 wavenumber components (from 0 to 255) in spectral space and a triangular-linear reduced-Gaussian grid of approximately 78 km horizontal resolution. By contrast, from January 2010 to March 2016 (a timespan covering most of the test-cases explored here) the ECMWF's operational releases of IFS used a 'TL1279' resolution, which corresponds to 1280 wavenumber components and roughly 16 km horizontal resolution. In March 2016 the model was upgraded to 'TCo1280' resolution, using a triangular-cubic-octahedral reduced Gaussian grid to bring about 9 km resolution forecasts with essentially the same number of spectral modes. The lower resolution means that a forecast produced for a given start date using the Open IFS may be less accurate than the equivalent forecast issued operationally.

The Open IFS uses a semi-Lagrangian, semi-implicit numerical scheme to solve the Navier-Stokes Equations for the momentum, surface pressure and temperature of atmospheric fluid parcels at each timestep, from which can be derived diagnostic quantities such as the geopotential and vertical velocity. The model is vertically discretised into a series of levels that follow the topography in the lower troposphere and gradually transition to follow pressure levels in the upper troposphere using Finite Element discretisation, and the levels are more finely spaced near Earth's surface than towards the top of the atmosphere. In the dynamical core, time-stepping takes place for some prognostic variables in spectral space because this makes differentiation more straightforward, and Legendre followed by Fourier Transforms are used to map the provisional tendencies for these quantities into grid-point space, where parameterisation schemes are used to simulate the effect of unresolved physical processes that might alter these tendencies (ECMWF 2012a). These processes include radiative transfer to and from space, turbulent mixing, drag from unresolved orography and gravity waves, interactions with the soil and land surface features, moist convection, and clouds, all of which affect the atmosphere at scales smaller than can be resolved by the model. Parameterisation computations are carried out across the vertical levels of each grid box individually (ECMWF 2012b).

The ECMWF produces high-resolution 10-day forecasts every twelve hours. Initial conditions are prescribed using surface observations, aircraft measurements, buoys and satellites, which have variable resolution and quality across the globe and are assimilated onto the model's grid using Four-Dimensional Data Assimilation (4D-Var). This process uses the previous data assimilation cycle's output to create, over a 12 hour period, a sequence of model states that fit as closely as possible with the observations. These then provide the best possible starting conditions for the next assimilation cycle (ECMWF 2015).

In addition to the high-resolution control forecast, the ECMWF also produces lower-resolution ensemble forecasts, which enable the uncertainty in the control to be assessed. Each ensemble member forecasts the same time period with slightly different initial conditions, to represent observational uncertainty, and different stochastic perturbations to the parameterisation schemes, which represent model uncertainty. A wide range of outcomes across the ensemble suggests that the atmospheric state may not be very predictable, and the percentage of ensemble members predicting a given outcome can be used to estimate the probability of that outcome occurring (ECMWF 2015). The high-resolution forecast data output since 1985 and the ensemble forecast output since 1995 can be downloaded from the MARS data archive available online to registered users (ECMWF 2017c).

5.2. Previous Reduced Precision Studies in Open IFS

Reducing precision in the Open IFS was first investigated by Düben and Palmer (2014), who carried out 'benchmark' tests running single-precision in the dynamical core of the system at resolutions ranging between T159 and T639, the latter being the operational resolution of the ECMWF's ensemble forecasts. The authors compared the 500 hPa geopotential and 850 hPa temperature predicted by double- and single-precision simulations and found that the difference between these was less than that between the double-precision simulation and the output of one ensemble member of the ECMWF's operational ensemble forecast. Hence, this study was instrumental in demonstrating the potential for single-precision to provide considerable cost

savings in the operational IFS model without degrading the forecast accuracy. This potential was confirmed by Vana et al. (2017), whose demonstration of a forty per cent speed-up relative to double-precision when most of the IFS was run in (scale-independent) single-precision may lead to single-precision being used operationally for ensemble forecasts in the near-future (Vana et al. 2016). This would equate to saving many millions of pounds in electricity per year, or alternatively the cost saving could be reinvested to substantially increase the model resolution or complexity.

A study of the extent to which precision could be reduced further than single-precision in parts of the Open IFS was carried out by Düben et al. (2017). Four test-cases were investigated, spanning a wide range of weather regimes. Using the reduced precision emulator to simulate the effects of reduced precision on conventional hardware, an automated method was employed to determine the lowest precision that could be utilised for each field, parameter and subroutine without inducing forecasting errors, and it was found that considerable cost savings would likely be realised were precision to be optimised in this way on real-world Mixed Precision hardware. The authors also found that the use of low precision in so-called 'super-parameterisation' schemes that explicitly model globally unresolved processes in specific areas of interest might make those schemes cheaper and therefore more attractive as a means of representing unresolvable processes in the IFS. That study therefore introduced the concept of applying different levels of precision to different parts of a model to the Open IFS, but it did not investigate varying numerical precision with spatial scale by applying the same idea to different wavenumbers within the spectral part of the dynamical core. That will be the focus of the experiments presented in this chapter, which apply explicitly scale-selective precision to the Open IFS for the first time.

It is difficult to estimate the computational cost saving that could be attained with a scale-selective approach to precision. Much of the power consumption for a typical machine is associated with cooling and other idle processes, but the 70 per cent of the energy that is typically used in performing calculations and moving data to memory (Song et al. 2009) is vulnerable to savings by reducing precision. However, a scale-selective approach is only possible in the spectral part of the model, which Vana et al. (2016) estimated would account for 14 per cent of the overall calculations' cost in double-precision at the resolution they used (TL399). Furthermore, scale-

selective precision may not be applicable to some of the parameterisation schemes and transformation (between spectral and grid-point space) routines, which together constitute some 80% of the IFS model's cost according to Vana et al. (2016).

5.3. Emulating Reduced Precision in Open IFS

This investigation will involve reducing precision in the spectral computations part of the Open IFS, where scale-selectivity can be straightforwardly applied. Reducing precision for grid-point calculations and transforms would require further changes but is not scale-selective and will not be investigated here. The Open IFS is much more complex than the Lorenz '96 and SQG systems. It has all the functionality of the operational IFS including all the parameterisation schemes, and contains some half a million lines of code distributed across more than two thousand files (Düben and Palmer 2014). However, the changes that need to be made to introduce the reduced precision emulator capable of implementing wavenumber-dependent precision (see Section 4.1) in spectral calculations are minimal.

The reduced precision module, 'rp_emulator.mod' is added, and the initial configuration files are changed to ensure that this is compiled during the model's installation. An additional subset of the namelist (the file containing specifiable parameters to direct how the model is run) entitled 'nam_rpe' allows namelist flags to be used to switch on the emulator and specify the number of significant bits to use for global (scale-independent) variables in spectral space, as well as what precision to use in up to three wavenumber bands for the scale-dependent quantities, and the model communicates these to the emulator via modifications to the linking program 'sudyn.F90'. There is also the capability to represent numbers that have 10 bits of precision in the significand with a 5 bit exponent, which is necessary for IEEE half-precision. The model is instructed to call the emulator every time a calculation is made in spectral space, and the emulator program rounds the input and output numbers for that calculation at each wavenumber range

individually as though they were represented with the specified number of bits. All these modifications were carried out by Matthew Chantry and Peter Düben in Oxford.

For this thesis, the author introduced additional slight modifications: most importantly, the number of bands of wavenumbers for which precision levels can be individually set was increased from three to nine, allowing a more flexible scale-selective approach to reduced precision to be implemented, through additional namelist flags and modifications to the reduced precision module. In all test cases, it was found that no more than nine different wavenumber bands with different levels of precision were required.

5.4. Test Cases

The proficiency of reduced precision forecasts was tested using four case studies centred upon the European region. Focussing on a region of this size meant that the ability to forecast the progress of specific historical weather phenomena in the reduced precision Open IFS could be tested, and the more extreme case studies were accompanied by studies of more 'ordinary' weather conditions over the same region for comparison. Initial data for the start date applicable to each test case were obtained from the ECMWF archive and converted into a format suitable for Open IFS. This model was then run with the grid-point and transform components in single-precision, and the spectral component in both single-precision to provide a 'truth' control and in reduced precision; it was assumed on the basis of previous studies (see Section 5.2) that single- and double-precision forecasts are of comparable quality and single-precision was used to save computational costs.

The aim throughout all these case studies was to determine how far precision could be reduced at each spatial scale (characterised by the wavenumber k) before the forecast ceased to be 'accurate'. Therefore, many different scale-selective models were run. In each run, ten different variables were analysed, as listed in Table 5.1. These variables were chosen because they represent a range of different quantities that are important for real-world forecasts with varying degrees of spatial variability. For simplicity, only quantities measured at the surface and at a single

representative level of the atmosphere – 500 hPa – were investigated. The methodology was as follows: first, the number of significand bits was uniformly reduced across all spatial scales ($k \geq 0$) until significant error was induced to indicate that the forecast was no longer accurate, as defined using the metrics outlined below. Then, the largest scale ($k = 0$) alone was kept at the lowest level of uniform precision required for an accurate forecast whilst precision was reduced further for wavenumbers $k \geq 1$, to check the hypothesis that smaller-scale quantities can be represented less precisely without penalty. Once the minimum level of precision for $k \geq 1$ had been ascertained, the process was repeated for $k \geq 2$ and so on up to the maximum wavenumber (either 159 or 255) resolved in each test-case, always keeping the lower wavenumbers at $k = 0$ precision. To save time, not all wavenumbers were explicitly tested: it was assumed that, for example, if wavenumbers 50 and 60 both require a minimum of 5 bits in the significand, all wavenumbers between them also require 5 bits. Within the spectral part of Open IFS, there are some non-scale-selective quantities, the precision in which can be set separately; a similar procedure was followed to determine the minimum precision for such 'global' quantities.

The two metrics used to analyse the accuracy of the reduced precision runs were as follows. For Metric A (the 'Averaged Area' metric), the ten output quantities were averaged over the whole European area at each timestep and the 500 hPa geopotential was compared across the different runs. 'Accurate' scale-selective forecasts were deemed to be those that remained close to the single-precision control throughout the forecast. To estimate the window within which forecasts could be regarded as accurate by this metric, the original forecasts produced by each of the fifty members of the ECMWF forecast ensemble at the time of each test case were downloaded from the ECMWF MARS archive. These were used to compute the ensemble standard deviation at each timestep of the original ECMWF forecast, averaged across all European grid-points; this quantity was then used to form error bars either side of the single-precision forecast. A reduced precision forecast was considered to be 'accurate' if it matched the single-precision forecast to within these error bars across the entirety of a five-day forecast. The geopotential at 500 hPa was used to determine the minimum precision required at each spatial scale because it was found to be the most

sensitive of the variables tested to reduced precision. The other variables were then checked to ensure that these were indeed forecast accurately alongside the geopotential.

Because this metric considers only spatially-averaged quantities, Metric G (the Grid-point metric) was devised to check the grid-point-by-grid-point accuracy of the scale-selective models. Here, the standard deviation was computed, again using MARS data, for each grid-point, and the fraction of grid-points for which scale-selective precision models differed from the single-precision control by more than the standard deviation was measured. The forecast was considered 'accurate' at any given time if this fraction was below 0.32, the fraction (to 2 significant figures) of points that would ordinarily lie outside one standard deviation of the mean in a Gaussian distribution (Riley et al. 2006), which is here assumed to hold. Because the standard deviation between ensemble members is very small for many grid-points near the start of each forecast, this test is overly strict at early times; therefore, for the present purposes the errors introduced by reduced precision were considered to be acceptable if the fraction of erroneous grid-points was consistently below 0.32 for lead-times of one day onwards. To complement this test, the output variables were also compared visually in grid-point space over Europe, to see whether the weather conditions and the evolution of extreme weather events were comparable between high- and low-precision runs.

Quantity	Code	Level	Units	Quantity	Code	Level	Units
Geopotential	z	500hPa	m^2s^{-2}	Sea-level pressure	msl	surface	Pa
Temperature	t	500hPa	K	10m Wind u-component	10u	surface	ms^{-1}
2m Temperature	2t	surface	K	10m Wind v-component	10v	surface	ms^{-1}
Total precipitation	tp	surface	m	Specific humidity	q	500hPa	$kg\ kg^{-1}$
Relative vorticity	vo	500hPa	s^{-1}	Relative humidity	r	500hPa	$kg\ kg^{-1}$

Table 5.1: Output quantities that were used to assess the accuracy of reduced precision forecasts in this investigation, together with the ECMWF code (or label), the level in the atmosphere and the SI measurement units pertinent to each.

Metrics A and G were used to determine the optimal levels of precision in spectral space, and on this basis a 'scale-selective' model was constructed that represented each wavenumber using the lowest level of acceptable precision. This scale-selective model was tested using the same metrics, alongside modified scale-selective models in which the precision levels had been altered,

to verify that the scale-selective model genuinely utilises the lowest-precision possible at each spatial scale that is necessary for an accurate forecast.

Two case studies of potential interest were identified in the St Jude's Day storm that brought widespread disruption to northern Europe in October 2013 and the stormy, wet weather that caused floods in the UK in February 2014. Both of these were recent enough to have been predicted by the current generation of ECMWF ensemble forecasts and are examples of events that it is especially important to forecast accurately. For a comparison to more 'normal' conditions, the weather conditions precisely four years before each of these events were also forecast in two further case studies; choosing this interval meant that February 2010 – a time of record-high predictability for extratropical weather – was included. Throughout all case studies a much lower spatial resolution was used for the model than would be used operationally at the ECMWF because of computational constraints, but running separate experiments at T159 and T255 resolution, corresponding to 160 and 256 resolved wavenumbers respectively, meant that the effects of changing the resolution on the required precision could be assessed. Throughout these test-cases, times are quoted in 24 hour format in Greenwich Mean Time (GMT).

5.4.1. October 2013: The St Jude's Day Storm

On St Jude's Day (28th October) 2013 a storm tracked across southern England, brining abnormally high winds of up to 80mph that led to the collapse of several trees, four deaths, and disruption to power supplies. These were the strongest winds since 2002. Heavy rain also led to localised flooding, and a number of transport links were disrupted. Although it was not as devastating as notable twentieth-century storms, the storm's recent occurrence meant that it was observed and forecast using modern-day techniques, including ensemble forecasts, which make it a particularly interesting case with which to test the proficiency of reduced precision forecasts. If a reduced precision forecast is able to predict the passage of the storm as accurately as the double-precision alternative, the two forecasts are equally useful for the purposes of preparing for extreme weather.

According to the Met Office, the storm was predicted as early as 24th October, when severe weather warnings were issued, and the timing, location and magnitude of its impacts were forecast correctly from at least 26th October (Met Office 2013). Therefore it was decided here to test the accuracy of reduced precision forecasts over Europe initialised on 26th October 2013 at midnight (00:00 GMT).

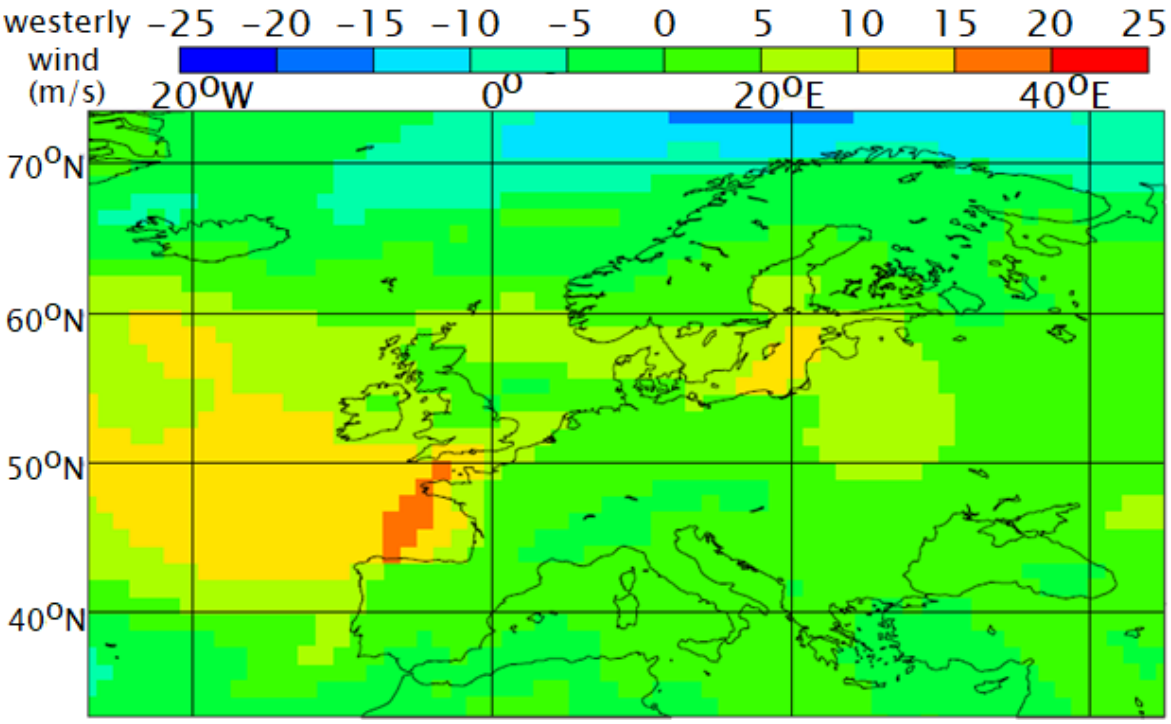

Figure 5.1: The 10m wind predicted over the European region at 09:00 GMT on 28th October by a single-precision global run of the Open IFS model initiated with the initial conditions for midnight on 26th October 2013. The passage of the St Jude's Day storm, which in reality occurred slightly earlier and further northward, is visible.

According to observations, the St Jude's Day storm took the form of a deepening depression that moved across southern England from the Bristol Channel to the Wash, with the strongest surface winds across the south coast reaching up to 71kn (37 ms^{-1}), and gusts of up to 91kn (47ms^{-1}) in the Dover Straits (Eden 2013). The pressure recorded at Heathrow Airport reached a minimum of 978hPa at 05:20 GMT as the storm passed over. The storm then tracked across Scandinavia, reaching southern Sweden by 18:00 GMT where 46kn (24ms^{-1}) gusts were measured (Villiers and White 2014). It was followed by unsettled weather, with a calmer day on

31st October followed by further storms from 2nd November (Eden 2014a). Figure 5.1 shows the 10m wind predicted by a single-precision run of the Open IFS initiated at midnight on 26th October for what this model predicts to be the height of the storm, 09:00 on 28th October. The storm is clearly visible, although its predicted arrival is slightly later than what it would be in reality, and the observed maximum wind speed was in the event higher than that predicted here.

Comparing between different output quantities in Table 5.1, it was found that two of these quantities, the 500 hPa geopotential and to a lesser extent the mean sea level pressure, were much more sensitive to reductions in precision than all the other quantities. The geopotential was therefore chosen to be the quantity in which the accuracy should be tested in the process of optimising precision, because an accurate geopotential implies that the other quantities are also accurate when the model is run in reduced precision. The geopotential is equivalent to the negative of the potential energy per unit mass, and is a key prognostic for determining the accuracy of a forecast. It is perhaps surprising that the precipitation, which is especially difficult to predict amongst the variables tested, was less sensitive to reduced precision. The explanation for this may lie in the fact that precipitation is shorter-lived and more spatially localised than geopotential, meaning that the low wavenumbers, where rounding errors are less likely to be masked by model errors, are less important for precipitation, and thus rendering this quantity less vulnerable to rounding error.

Table 5.2 lists the results of a precision optimisation analysis carried out using runs initialised at midnight on 26th October. In the experiments used to determine these levels, runs with precision reduced in some wavenumbers were compared to the single-precision run used to generate the plot in Figure 5.1. The table shows the minimum wavenumber at which precision is reduced to each number of significand bits in turn in three different optimised scale-selective models for each resolution.

The precision levels for Model A were determined using Metric A to find the lowest level of precision achievable at each wavenumber without incurring significant error in the geopotential averaged across the whole European region. The precision levels for Model G, meanwhile, were set using Metric G, which compares each grid-point individually. Both models were then assessed for

accuracy using the same metrics used to produce them. Because when the precision levels were combined to form Model G this model was not accurate according to Metric G, models of type G* were created by adjusting the precision levels by hand by trial and error to produce an approximately accurate forecast according to Metric G.

$k_{\max} = 159$ resolution				$k_{\max} = 255$ resolution			
Significand bits s	Minimum wavenumber			Significand bits s	Minimum wavenumber		
	Model A	Model G	Model G*		Model A	Model G	Model G*
13	0	0	0	14	0	0	0
8	1	1	1	9	2	1	1
7	3	6	8	8	3	7	9
6	7	9	11	7	3	9	11
5	15	70	80	6	9	44	64
4	103	121	131	5	123	131	161
3	117	139	139	4	162	184	224
2	-	-	-	3	242	243	253

Table 5.2: The minimum wavenumber k at which the number of bits in the significand can be reduced to s whilst satisfying the requirements for an 'accurate' five-day forecast for Models A, G and G* is listed for the St Jude's Day 2013 test-case with a resolution of $k_{\max} = 159$ (left columns) and 255 (right columns).

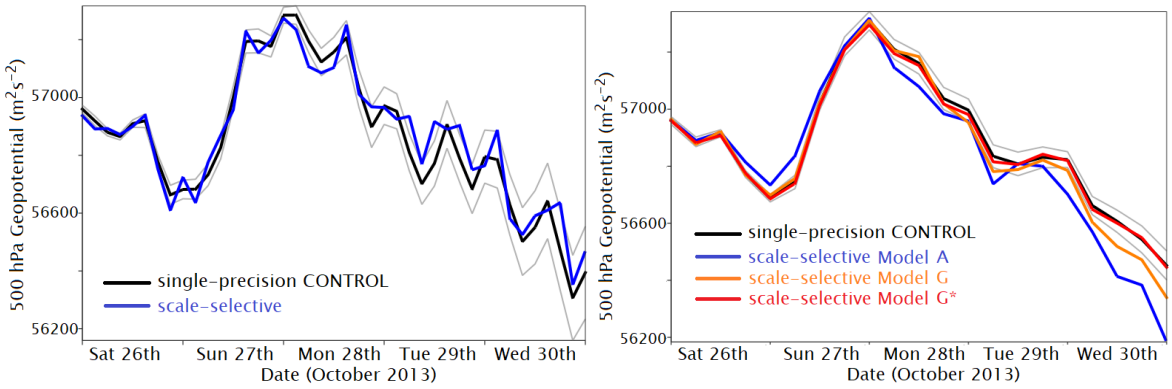

Figure 5.2: The scale-selective Model A (blue line) is compared to single-precision (black line) at $k_{max} = 159$ resolution (left) and at $k_{max} = 255$ resolution (right) using Metric A over a 5 day forecast period for the European region. The grey lines enclose a region that is one standard deviation of the ECMWF operational ensemble (downloaded from MARS) either side of the single-precision model. For $k_{max} = 255$ (the right hand plot) Models G and G* are also tested using Metric A, because Model A does not yield an accurate forecast in this case.

In the 'accurate' Model G*, for both resolutions some of the low k thresholds had to be increased by 2 wavenumbers and the high k thresholds by 10 wavenumbers relative to Model G, but in the high-resolution case it was also necessary to increase the medium k thresholds by 20 or 30 wavenumbers to produce an accurate model. Metric G is evidently harsher as a measure than Metric A because it takes into account local differences between the single-precision and reduced precision runs, but both metrics demonstrate a similar decrease in the required precision at higher wavenumbers. No adjustment was required at wavenumber 0; this requires much higher precision than the other wavenumbers because it represents globally constant properties, and reducing precision too far in these can easily begin to violate conservation laws for quantities whose constant global totals are defined in high precision.

Figure 5.2 shows the Metric A forecast accuracy tests at both resolutions. In the low-resolution $k_{\max} = 159$ case, whilst Model A was accurate according to Metric A, the Metric G model was inaccurate according to Metric G; hence the need for adjustment to form Model G*. For $k_{\max} = 255$ Model A is inaccurate, and the too-rapid fall in geopotential it exhibits after Tuesday 29th could not be substantially reduced by increasing the wavenumber thresholds slightly or by keeping the largest scales in single-precision. Only by adjusting the precision levels as in Model G* could an accurate forecast for the five days be produced.

It is evident in these plots that the standard deviation between ensemble members grows gradually during the forecast, and appears to do so more rapidly than the reduced precision models diverge from the single-precision forecast. However, this divergence – the presence of which is expected in a chaotic system – is by no means entirely absent, and at longer lead-times the different models do chaotically diverge, as is illustrated in Figure 5.3. This figure plots the geopotential Root Mean Square Error (RMSE) between each of Models G and G* and single-precision, averaged across all grid-points for both the European region and the globe as a whole, alongside the ensemble forecast geopotential standard deviation. The RMSE demonstrably grows over time but only after the standard deviation reaches a comparable magnitude to the initial model error. The 'bump' of low forecast accuracy around 31st October in the European region corresponds well with

an increase in ensemble spread on this date. This was associated with a calmer but apparently less predictable interval between stormier conditions on the surrounding days. The standard deviation between ECMWF ensemble members is much smaller when averaged over the whole globe rather than Europe only, and whilst on a global scale the models' agreement with single-precision degrades more smoothly, both models are very far from the standard deviation in this case.

Figure 5.4 shows tests of the models using Metric G, to demonstrate that Model G* is accurate by this measure. In the $k_{\max} = 159$ case, although Model G* remains above the 0.32 grid-point fraction threshold for the first two days of the forecast, by the time of the storm (Monday 28th) it dips consistently below this threshold. The apparently low skill for days 0-2 results from the very small ensemble spread in the operational model for small lead-times, and therefore does not indicate that the scale-selective models were further from the single-precision control at low lead-times (see Figure 5.3). However, it does mean that these models, which are optimised here for five-day forecasts, would introduce significant errors relative to the operational ensemble spread if they were used for next-day or two-day forecasts. A separate optimisation would have to be performed to create a model with no degradation in skill for such forecasts, but in this study the focus is on forecasting storms 3-5 days in advance, and a more accurate forecast for days 1-2 could not be produced without substantially changing the wavenumber thresholds.

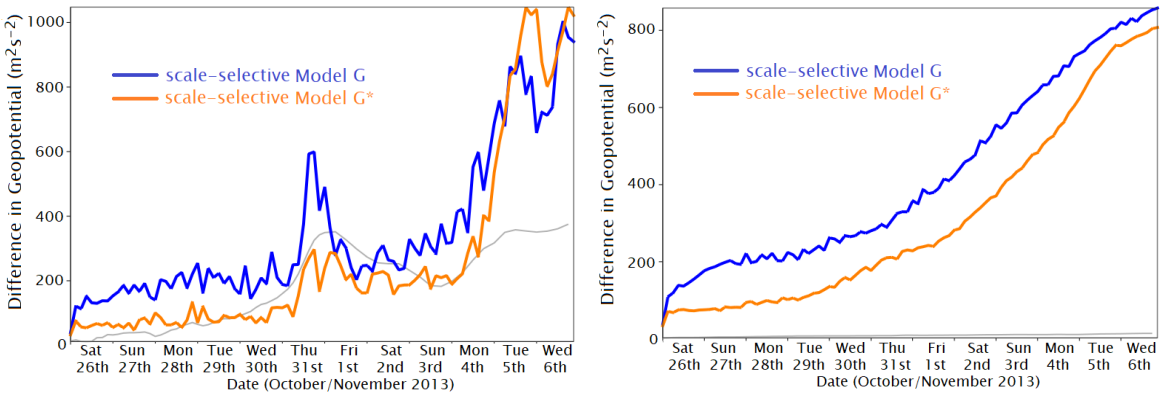

Figure 5.3: The Root Mean Square Error, averaged across all European grid-points (left) or the entre globe (right), between Model G (blue line) or Model G* (yellow line) and single-precision, is plotted for a 12 day forecast at $k_{\max} = 159$ resolution. The grey line is the ensemble standard deviation downloaded from MARS. This is barely visible in the global-averaged case, as the ensemble spread is near-zero in many grid-points.

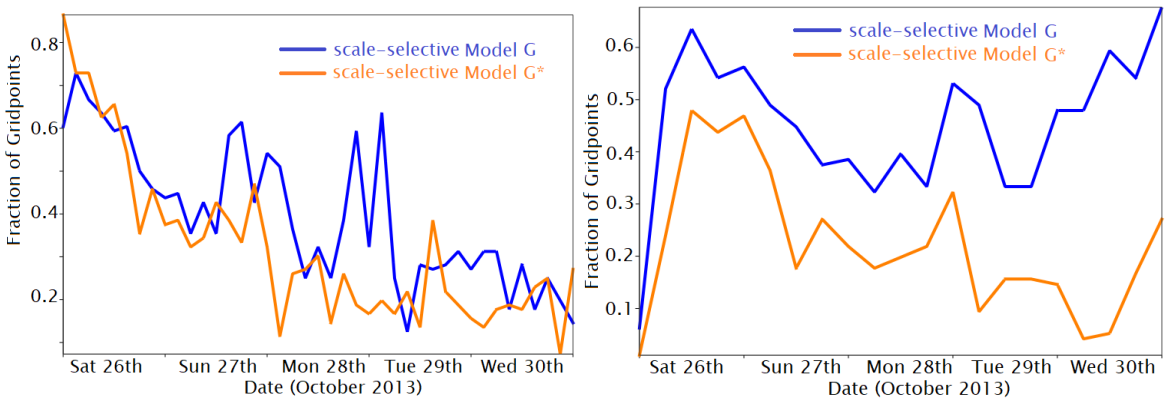

Figure 5.4: Scale--selective Models G (blue lines) and G* (yellow lines) are compared using Metric G, which measures the fraction of European grid-points for which each model agrees with single-precision to within the MARS ensemble standard deviation, for $k_{max} = 159$ (left) and $k_{max} = 255$ (right) resolutions.

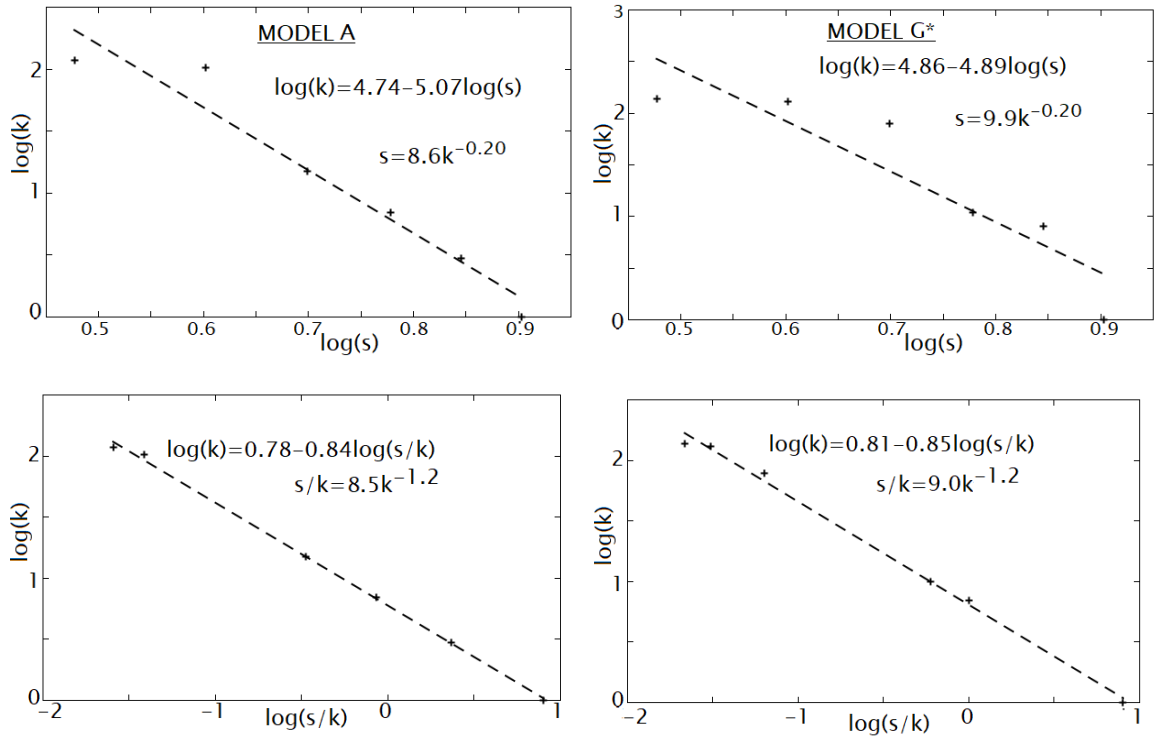

Figure 5.5: The minimum wavenumber k is plotted logarithmically against the lowest number of significand bits s that can be used to represent that wavenumber whilst satisfying Metric G (top plots) or against s/k (bottom plots) for Model A (left) and Model G* (right) for the $k_{max} = 159$ case. Best-fit lines to both sets of points are shown (dashed lines), and the equations describing these lines and the corresponding polynomial relationship between s or s/k and k are given on each graph.

For the $k_{\max} = 255$ case the fraction for Model G* is below 0.32 from day 2 onwards. Notice that the initial error begins close to zero in this case before jumping at the first timestep; this indicates that the reduced precision error originates in the calculations, not the initial conditions. It was found that significant initial condition error is induced when the number of bits used for the first wavenumber, $k = 0$, is less than 14; because this occurs in the $k_{\max} = 159$ case, the error begins relatively high and there is no jump at the first timestep for $k_{\max} = 159$. Raising $k = 0$ to 14 bits or greater does not affect the forecast from day 2 onwards, however, which is why Model G* could have fewer than 14 bits for $k = 0$ and still satisfy the accuracy criteria used here. The same phenomenon is exhibited in many of the other plots in this chapter when models are run with fewer than 14 bits in wavenumber zero.

In Figure 5.5, the minimum wavenumber k is plotted logarithmically against the minimum number of bits s in the significand that can be used for that wavenumber without error at $k_{\max} = 159$ resolution, yielding a roughly straight line. Although Model G* has a higher wavenumber threshold for any given level of precision, both graphs have approximately the same slope, which corresponds to the relationship $s \propto k^{-0.2}$. As in the SQG case, plotting k logarithmically against s/k instead, also shown in Figure 5.5, gives a line that is much straighter. There is no physical justification for doing this, except that it allows the straight line of the s versus k plot to be established more accurately, as the observed relationship $s/k \propto k^{-1.2}$ is much more robust than, but is also entirely consistent with, the s versus k relationship. For this reason s/k is plotted against k in the remainder of the test cases presented in this chapter. Furthermore, because the slope is the same for both metrics, but Metric G is the more pertinent to forecasts produced for local regions and would hence be more useful in stormy weather than region-averaged forecasts, only Model G* results will be analysed for those test cases.

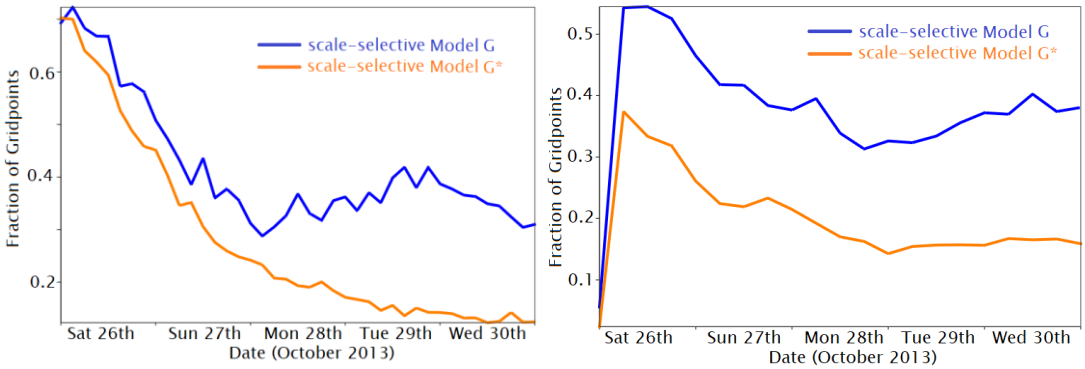

Figure 5.6: Models G and G* are tested for their forecasting accuracy using Metric G* but for the entire globe, rather than the European region. The fraction of grid-points for which each model lies within the standard deviation of the single-precision simulation is plotted, for resolutions of $k_{\max} = 159$ (left) and $k_{\max} = 255$ (right).

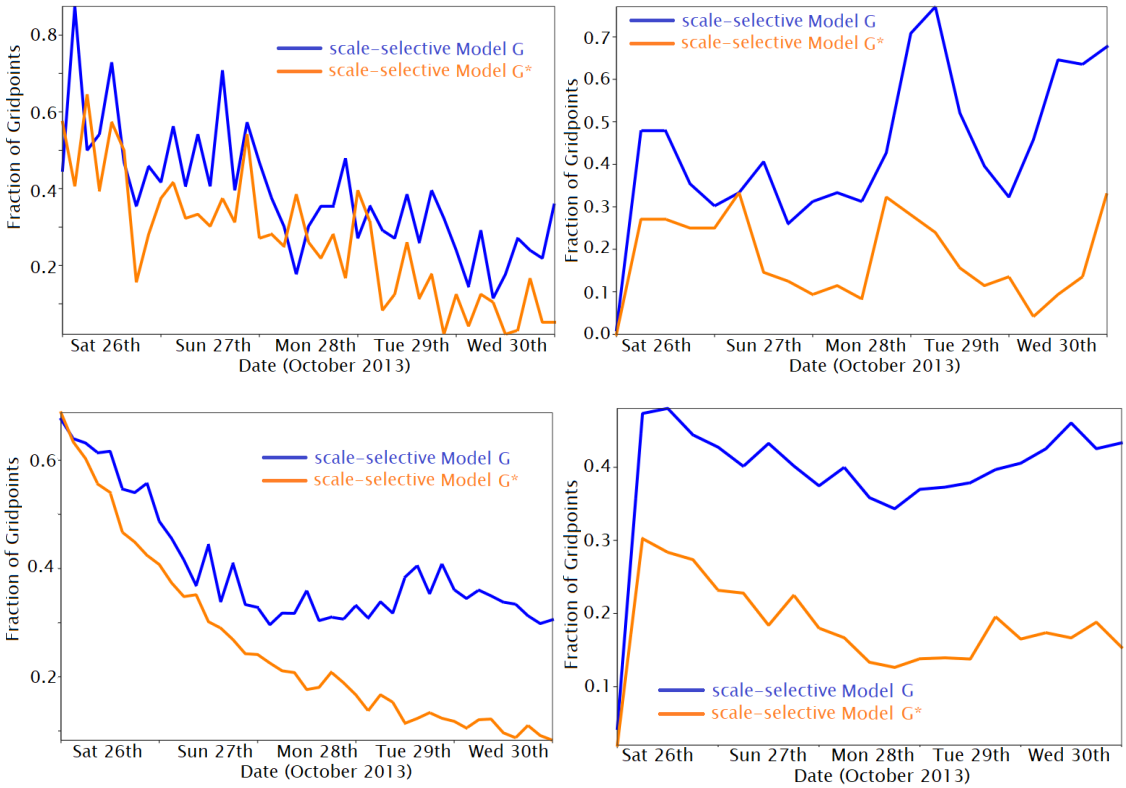

Figure 5.7: The accuracy of Models G and G* is measured using Metric G for the mean sea-level pressure variable, in the European (top row) and global (bottom row) cases, for resolutions $k_{\max} = 159$ (left column) and $k_{\max} = 255$ (right).

The results presented so far have been restricted to the European region, looking solely at the geopotential at 500 hPa. To demonstrate the accuracy of the scale-selective simulations on a

global scale, Figure 5.6 plots the geopotential forecast accuracy measured using Metric G but for all grid-points across the entire globe, for both resolutions. Though the curves are generally smoother, since there are more grid-points to count, the results are consistent with those where the analysis was restricted to the European region: again, Model G* is accurate according to Metric G. Figure 5.7 shows European and global forecasts for the mean sea-level pressure (mslp), which was the only parameter other than geopotential amongst those tested (see Table 5.1) to show significant sensitivity to a reduction in precision to the levels investigated here. Again, Model G* performs well for predictions of the mslp, just as it does for the geopotential.

The results of the Lorenz '96 experiments in Chapter 2 suggested that accurate forecasts could be produced more cheaply by reducing precision more at longer lead-times, when the accuracy naturally begins to degrade so that a precise representation of the model fields is less necessary. For this reason, simulations were also performed at both resolutions where the precision was kept at single-precision for the first one or two days of the forecast, then reduced to the Model G (that is, unadjusted) scale-selective precision levels. As explained above, Model G did not yield accurate forecasts when it was used from the outset; the aim here was to investigate whether it could become accurate if its implementation was delayed slightly. The two 'dynamic-precision' runs were tested using Metric G, and the results are shown in Figure 5.8.

At low resolution, over the European region there is no advantage to delaying the implementation of scale-selective reduced precision by one or two days: almost as soon as the reduction in precision is implemented, the accuracy relapses to that of the run with reduced precision switched on from the outset. This implies that the reduced precision rounding errors are not cumulative over time. However, at higher resolution, delaying the reduction in precision by one day yields a slight advantage out to at least day 5 of the forecast. On a global scale, furthermore, at both resolutions there is a considerable advantage to delaying the onset of reduced precision, in that Model G, which is not accurate if implemented from day 0, becomes accurate when implemented after day 1. There is no significant advantage to delaying the reduction in precision to day 2, suggesting that the positive effect comes from preserving high-precision when the ensemble spread is small and the single-precision forecast is more certain, during the first day of simulations. The

ensemble spread is larger for the European region than for the globe as a whole, as was shown in Figure 5.3, which may explain why the positive effect of delaying the reduction in precision was not seen in that region. Further tests revealed that the results above do not change significantly when the first 1 or 2 days are implemented with 16 bits in the significand, as opposed to single-precision (23 bits significand).

To verify that the scale-selective reduced precision run was capable of successfully predicting the storm, forecasts for the different variables were also compared by eye over the European region. Figure 5.9 shows, for both low- and high-resolution cases, the 10m wind predicted at 15:00 on Sunday 27th October, which according to the single-precision control forecast was expected to be the windiest time, for four different models: the single-precision model; a model with precision uniformly reduced to one bit below that usually required in the highest wavenumber (that is, 12 bits for $k_{\max} = 127$ and 13 bits for $k_{\max} = 255$); and Models G and G*. It is clear from a comparison by eye of these plots that the difference between the reduced precision models is slight, and that none of them produces a significantly different forecast to the single-precision control. Indeed, the models at each resolution differ by much less than the difference between the high- and low-resolution single-precision forecasts. This applies even for the 12- and 13-bit significand forecasts, which were considered inaccurate according to Metric G based on the geopotential, illustrating the fact that the 10m wind is less sensitive to precision reductions and implying that the scale-selective models are suitable for accurately predicting windy conditions. A similar agreement between the four models at each resolution is also observed in the 10m wind predicted for the end of the five-day forecast (not shown here).

Overall, this case-study has shown that the St Jude's Day Storm could be predicted by a model where the spectral part is run in scale-selective reduced precision, with as few as 3 bits used at the highest wavenumbers, with no significant degradation in accuracy relative to single-precision. According to the results presented here, it is likely that models run at higher resolutions would require slightly higher precision at the lowest wavenumbers in order to be accurate, but that the precision at high wavenumbers – which account for most of the computational cost – might nevertheless be reduced to the order of 3 bits without penalty, with the caveat that the minimum

wavenumber at which a given number of bits can be safely applied tends to be higher when the maximum resolved wavenumber is higher.

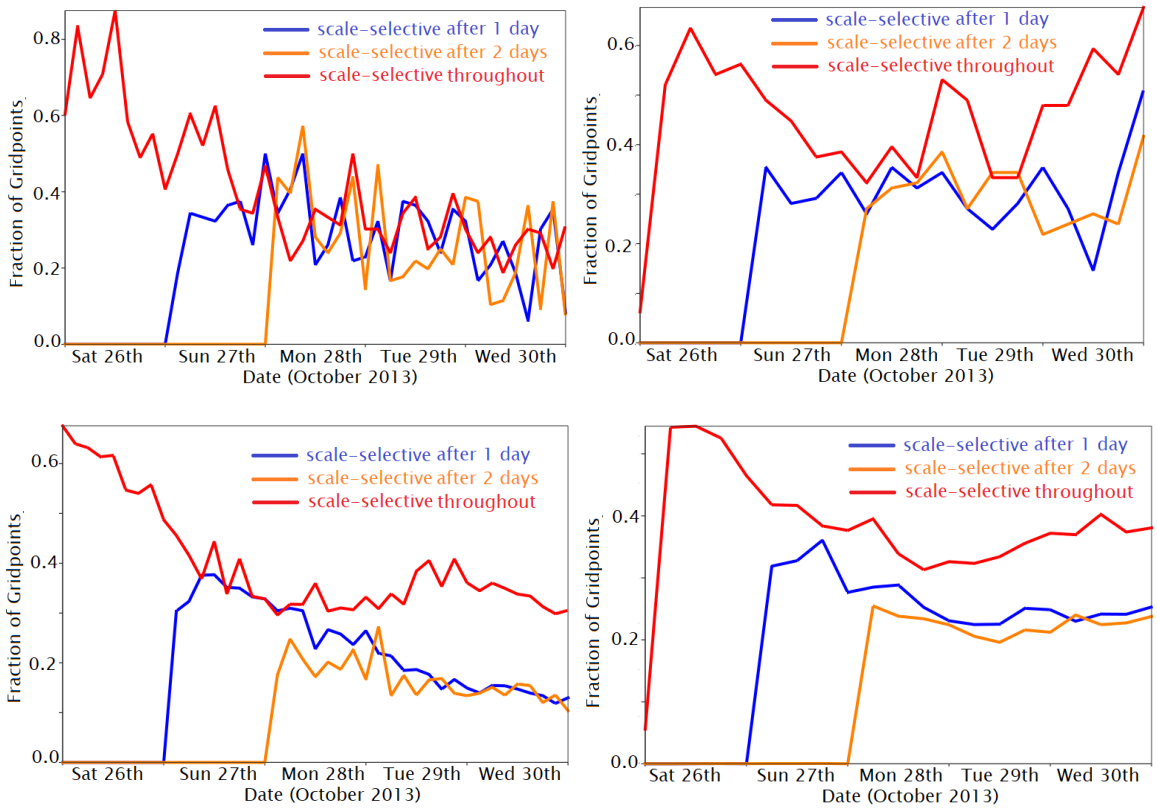

Figure 5.8: The fraction of grid-points for which the Geopotential at 500 hPa lies outside the standard deviation from the single-precision value is plotted for runs with Model G scale-selective precision throughout (red line) or introduced after a lead time of 1 (blue line) or 2 (orange line) days, with single-precision used before this. The left-hand plots are for $k_{\max} = 159$ resolution; the right-hand plots for $k_{\max} = 255$; the top row shows results for the European region, and the bottom row for the entire globe.

This may be explained if a given spatial scale (wavenumber) has a stronger effect on the overall model output when it is further from the model's truncation scale: wavenumber 150 would not be very influential to a forecast at resolution $k_{\max} = 155$, for instance, relative to its importance at resolution $k_{\max} = 255$. This is partly because of the model numerics: aliasing, whereby spectral energy builds up at small scales, and the horizontal diffusion used to counteract this strongly affect numerical values at the scales that are closest to the resolution limit in the Open IFS (Wedi et al.

2013), so that the reduced precision rounding error can be made very great before it exceeds the model error at the smallest resolved scales, lowering the 'effective resolution' in each case.

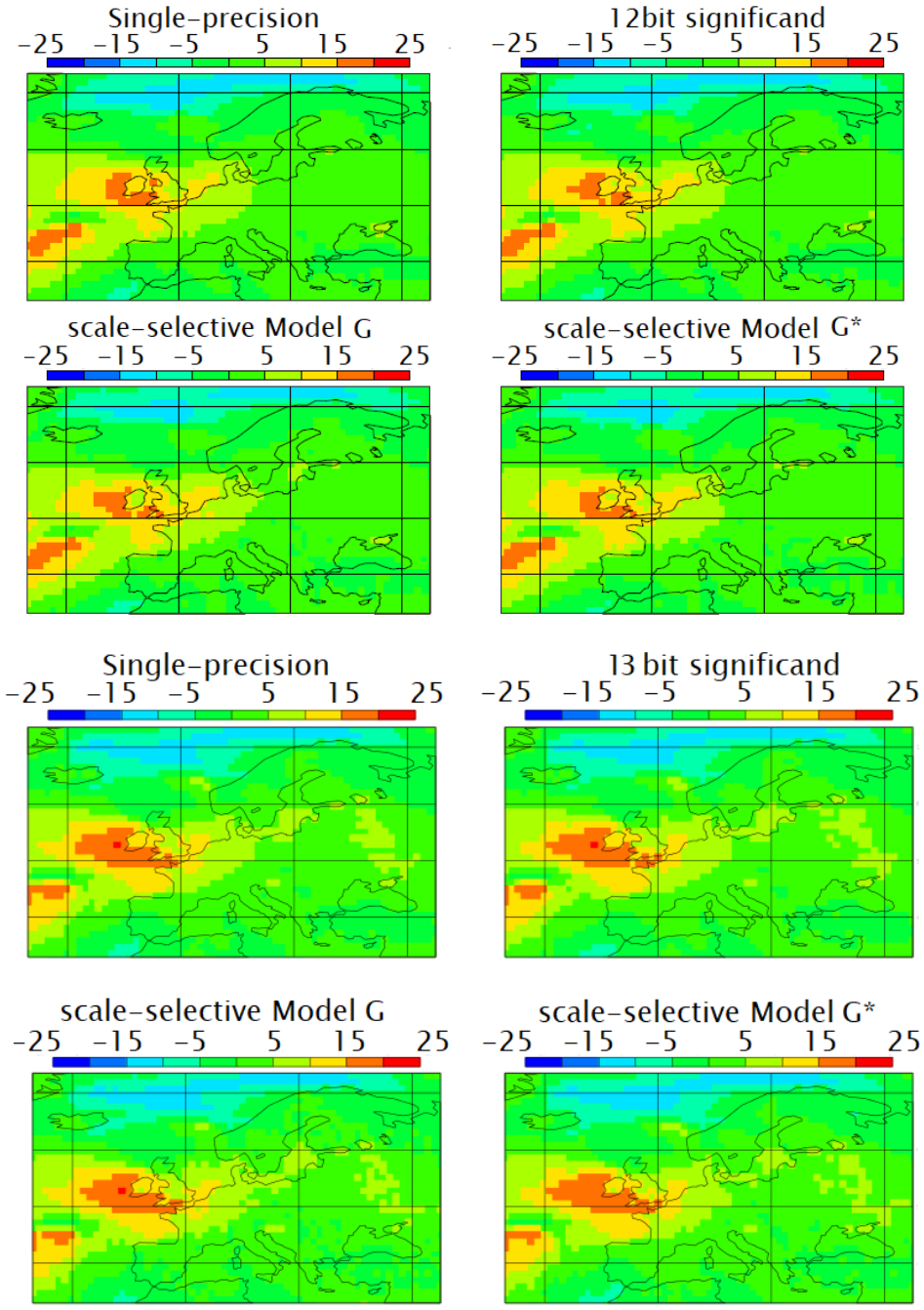

Figure 5.9: The 10m wind predicted for 15:00 on Sunday 27th October 2013 by four different models, all initialised at 00:00 on Saturday 26th October 2013, for the $k_{\max} = 127$ (top four plots) and $k_{\max} = 255$ (bottom four plots) cases.

5.4.2. October 2009: 'Normal' conditions

To provide a contrast with the St Jude's Day test-case and to determine the degree to which the results of that study were dependent upon the stormy conditions in the forecast, the same analysis was repeated for a start date four years earlier, 00:00 on 26th October 2009. In the UK, October 2009 was less gusty than the average October, and the final ten days – including the period examined in this study – were especially warm, reaching a noteworthy 20.2°C on 27th October in Caernarfon, with extensive cloud cover and occasional rain but no storms (Eden 2009), which contrasts well with the 2013 case.

$k_{\max} = 159$ resolution			$k_{\max} = 255$ resolution		
Significand bits s	Minimum wavenumber		Significand bits s	Minimum wavenumber	
	Model G	Model G*		Model G	Model G*
15	0	0	16	0	0
10	1	1	10	1	1
9	3	3	8	7	7
7	10	10	7	10	10
6	21	21	6	72	92
5	78	78	5	146	166
4	111	126	4	199	219
3	137	152	3	244	249

Table 5.3: The minimum wavenumber k at which the number of bits in the significand can be reduced to s whilst satisfying the requirements for an 'accurate' five-day forecast for Models G and G*, for the October 2009 test-case with resolutions of $k_{\max} = 159$ (left columns) and 255 (right columns).

In light of the results from the St Jude's Day study, only Metric G was used to analyse the accuracy for this test case. The precision levels required at each wavenumber for an accurate forecast are shown in Table 5.3 (Model G), alongside the adjusted precision levels required to form a scale-selective model that was accurate (Model G*) , for both resolutions at which the model was run. For this test-case, only the higher wavenumber bands had to be adjusted to produce an accurate Model G*, by up to 15 wavenumbers in the low-resolution and 20 wavenumbers in the high-resolution case. Again, just 3 bits of significand precision were required at the very highest wavenumbers for an accurate forecast, but in this test-case 15 or 16 bits were required for

wavenumber 0, representing globally invariant quantities. This implies that the necessary numerical precision for global quantities may be context-specific in the Open IFS.

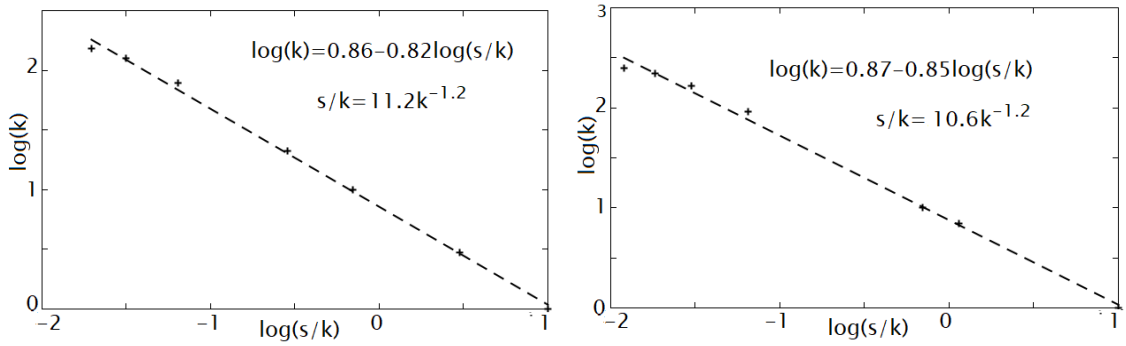

Figure 5.10: The minimum wavenumber k to which precision can be reduced to s bits in the significand is plotted logarithmically as a function of s/k (using a base-10 logarithm) for the $k_{max} = 159$ (left) and $k_{max} = 255$ (right) resolution cases. The equation of the best fit line is shown on each plot, and is translated into a power-law equation relating s/k to k in each case.

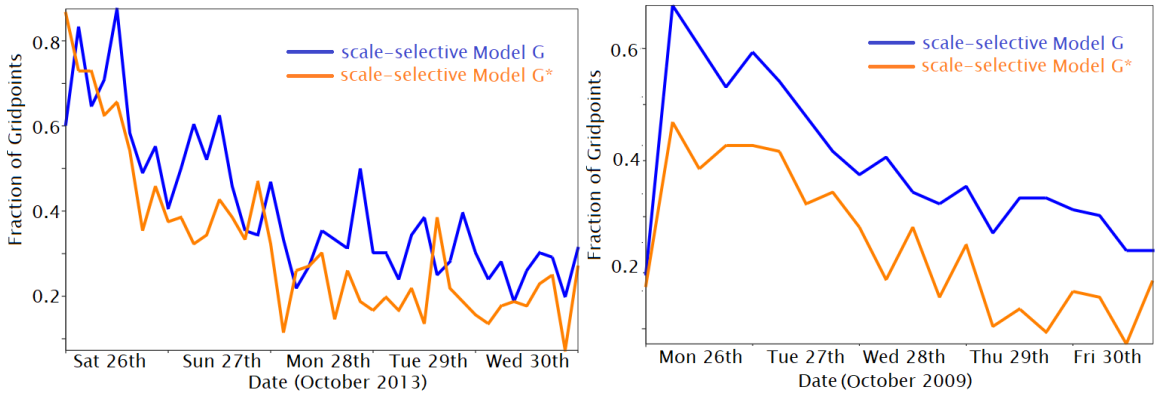

Figure 5.11: The 500 hPa Geopotential accuracy of the scale-selective Models G and G* in the low-resolution (left) and high-resolution (right) cases is assessed using Metric G in the European region.

The minimum wavenumber in each 'band' of precision for Model G* is plotted against the number of significand bits required per unit wavenumber in Figure 5.10 at both resolutions, yielding a very similar power law to the St Jude's Day case, of $s/k \propto k^{-1.2}$ which corresponds to $s \propto k^{-0.2}$. The European and Global geopotential predicted by Model G* is assessed using Metric G in Figure 5.11. A delayed start-time for reduced precision was again investigated, yielding the plots in Figure 5.12; again, at low resolution there is no advantage to delaying the onset of reduced

precision, but in this test-case a significant improvement is visible at higher resolution, especially when the forecast quality is analysed for the whole globe rather than for Europe only.

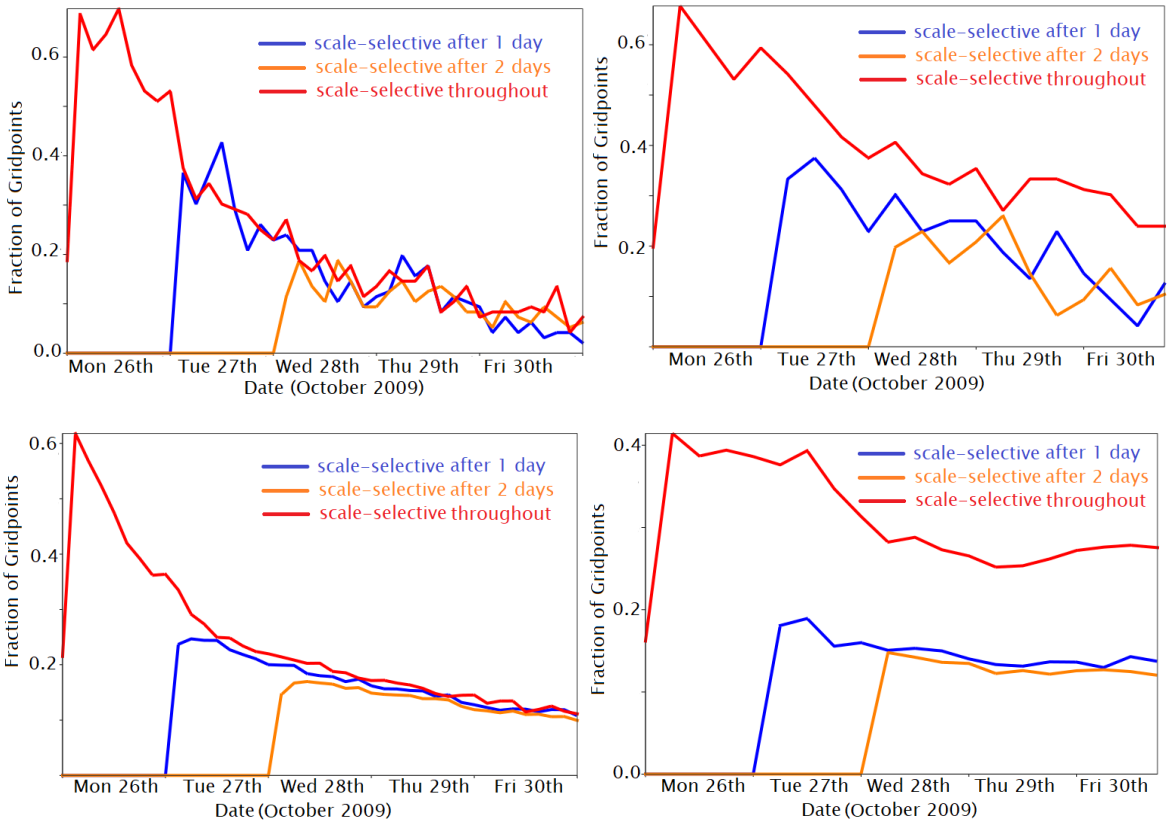

Figure 5.12: The 500 hPa Geopotential accuracy in the European region (top row) and for the whole globe (bottom row) is assessed using Metric G for runs with Model G scale-selective precision throughout (red line) or introduced after a lead time of 1 (blue line) or 2 (orange line) days, with single-precision used before this. The left- and right-hand plots show low- and high-resolution cases respectively.

5.4.3. February 2014: February Storms

February 2014 exhibited the lowest mean sea-level pressure over the British Isles of any month on record, and the first half of the month was characterised by successive storms and enhanced rainfall, leading to localised flooding in the UK. Overall, February 2014 was exceptionally mild and wet, with a Central England Temperature (CET) anomaly of 6.3°C, which is 1.8°C above the 1981-2010 average for February, and England and Wales Precipitation (EWP) totalling 213% of

the February average. One of the strongest storms was that which crossed from Southwest to Northwest England on 12th, which was followed by a similar strength storm further east on 15th before calmer and very mild conditions set in for the second half of the month (Eden 2014b). The five-day forecast period chosen for this test case begins at midnight at the beginning of 11th and encompasses both these storms and the transition to less stormy conditions. The aim here is hence to evaluate the ability of low-precision forecasts to accurately predict the storms in advance of their arrival.

$k_{\max} = 159$ resolution			$k_{\max} = 255$ resolution		
Significand bits s	Minimum wavenumber		Significand bits s	Minimum wavenumber	
	Model G	Model G*		Model G	Model G*
13	0	0	13	0	0
10	1	1	10	1	1
9	7	7	9	3	3
8	8	8	8	8	10
7	12	12	7	9	11
6	31	36	6	38	78
5	71	76	5	132	162
4	102	112	4	187	217
3	134	144	3	245	245

Table 5.4: The minimum wavenumber k at which the number of bits in the significand can be reduced to s whilst satisfying the requirements for an 'accurate' five-day forecast for G and G*, for the February 2014 test-case with a resolution of $k_{\max} = 159$ (left columns) and 255 (right columns).

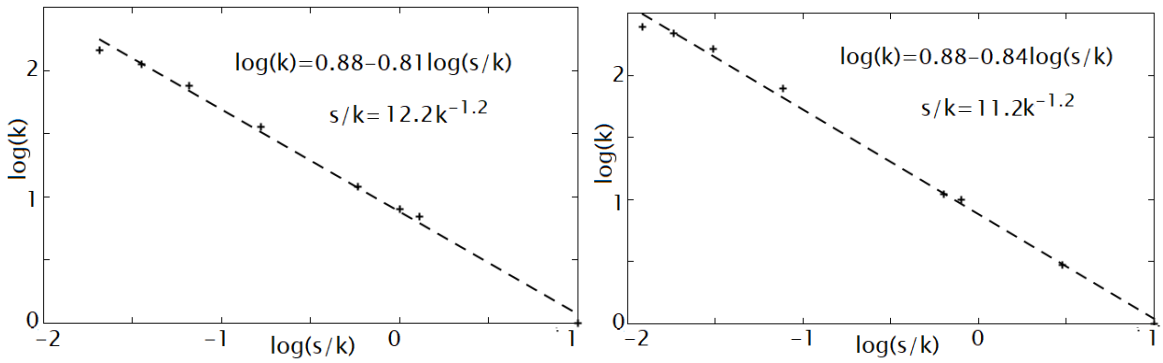

Figure 5.13: The minimum wavenumber k to which precision can be reduced to s bits in the significand is plotted logarithmically as a function of s/k (using a base-10 logarithm) for the $k_{max} = 159$ (left) and $k_{max} = 255$ (right) resolution cases. The equation of the best fit line is shown on each plot, and is translated into a power-law equation relating s/k to k in each case.

Table 5.4 shows the precision levels obtained for Models G and G*, determined using the same method as was employed for the St Jude's Day Storm test-case. The precision required per unit wavenumber as a function of minimum wavenumber is plotted for both resolutions in Figure 5.13, again so as to produce the most accurate straight-line fit, which suggests that $s \propto k^{-0.2}$ in accordance with the other test-cases. The performance of these models, assessed using Metric G, is shown in Figure 5.14, and an analysis of the performance of Model G when the reduction in precision is delayed by one or two days in shown in Figure 5.15.

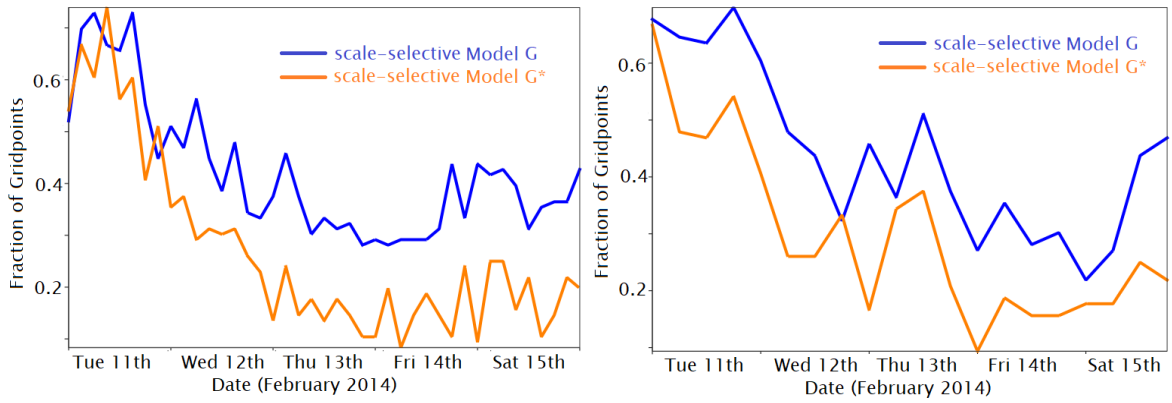

Figure 5.14: The 500 hPa Geopotential accuracy of the scale-selective Models G and G* is assessed using Metric G for the low-resolution (left) and high-resolution (right) cases in the European region for the February 2014 test-case.

At low-resolution, the February 2014 case yields very similar results to those of the St Jude's Day 2013 example, requiring just 13 bits in the significand at the largest scales and 3 bits at the smallest twelve scales for an accurate forecast. The only significant difference between the two cases comes at higher resolution, where in general slightly less precision is required for an accurate Model G* in the February 2014 example. When the onset of reduced precision is delayed, in the February 2014 case there is a slight advantage relative to scale-selective precision throughout for the Europe-only analysis, rendering Model G 'accurate' by the criteria of Metric G, but this time only in the low-resolution case. The advantage of delaying the reduction in precision is considerably greater for the global analysis.

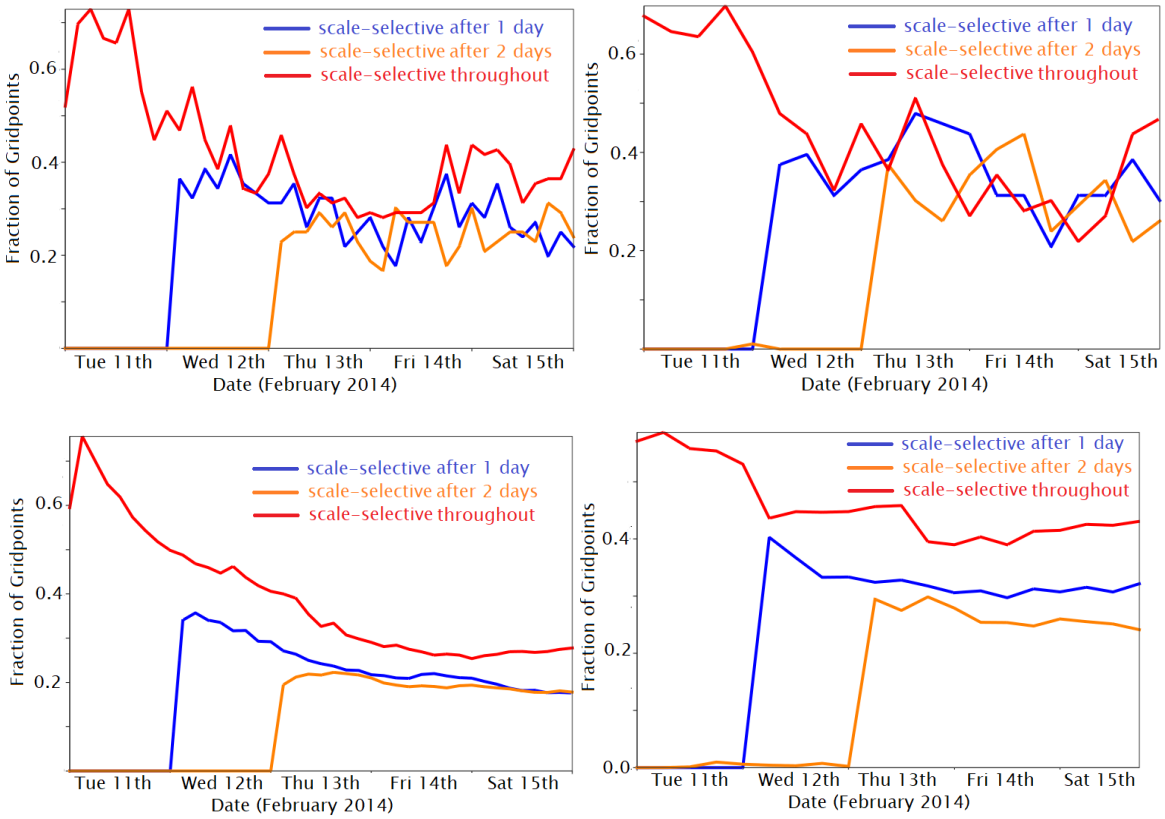

Figure 5.15: The 500 hPa Geopotential accuracy in the European region (top row) and for the whole globe (bottom row) is assessed using Metric G for runs with Model G scale-selective precision throughout (red line) or introduced after a lead time of 1 (blue line) or 2 (orange line) days, with single-precision used before this. The left- and right-hand plots show low- and high-resolution cases respectively.

5.4.4. February 2010: High extratropical predictability

Exactly four years prior to the mild, stormy weather of 2014, February 2010 was cold and calm, with no gusts recorded across the United Kingdom. The monthly CET was 1.8°C below average, and in the latter half of the month there were heavy accumulations of snow in England, Wales and Scotland (Eden 2010). What makes this month especially interesting is that it corresponded to a peak in predictability for the ECMWF forecast system: IFS forecasts in February 2010 were more accurate than those in any month before or since, for both European and Northern Hemisphere extratropical forecasts (Richardson et al. 2010). This might be related to the predominance of easterly flows over Europe, which brought the anomalously cold and snowy conditions; in

historical records, only the Februaries of 1986, 1947 and 1895 have exhibited more persistently easterly winds. This type of weather extreme is clearly forecast relatively well by the ECMWF model, and it could be said hence that the February 2010 test-case represents the IFS at its best.

The same reduced precision analysis was carried out as for the previous three test-cases, and the results are listed in Table 5.5. The precision required per unit wavenumber as a function of minimum wavenumber is plotted for both resolutions in Figure 5.16; an assessment of the performance of Models G and G* is shown in Figure 5.17 using Metric G; and Figure 5.18 shows the performance of Model G when the reduction in precision is delayed by one or two days.

$k_{\max} = 159$ resolution			$k_{\max} = 255$ resolution		
Significand bits s	Minimum wavenumber		Significand bits s	Minimum wavenumber	
	Model G	Model G*		Model G	Model G*
13	0	0	14	0	0
8	1	1	9	1	1
7	5	5	8	6	6
6	12	12	7	8	8
5	62	62	6	56	56
4	105	125	5	136	136
3	122	142	4	191	211
2	-	-	3	244	254

Table 5.5: The minimum wavenumber k at which the number of bits in the significand can be reduced to s whilst satisfying the requirements for an 'accurate' five-day forecast for G and G*, for the February 2010 test-case with a resolution of $k_{\max} = 159$ (left columns) and 255 (right columns).

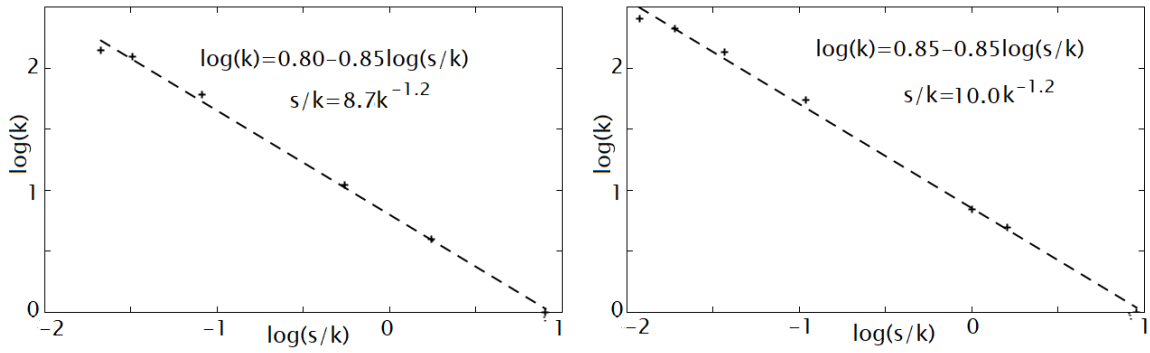

Figure 5.16: The minimum wavenumber k to which precision can be reduced to s bits in the significand is plotted logarithmically as a function of s/k (using a base-10 logarithm) for the $k_{max} = 159$ (left) and $k_{max} = 255$ (right) resolution cases.

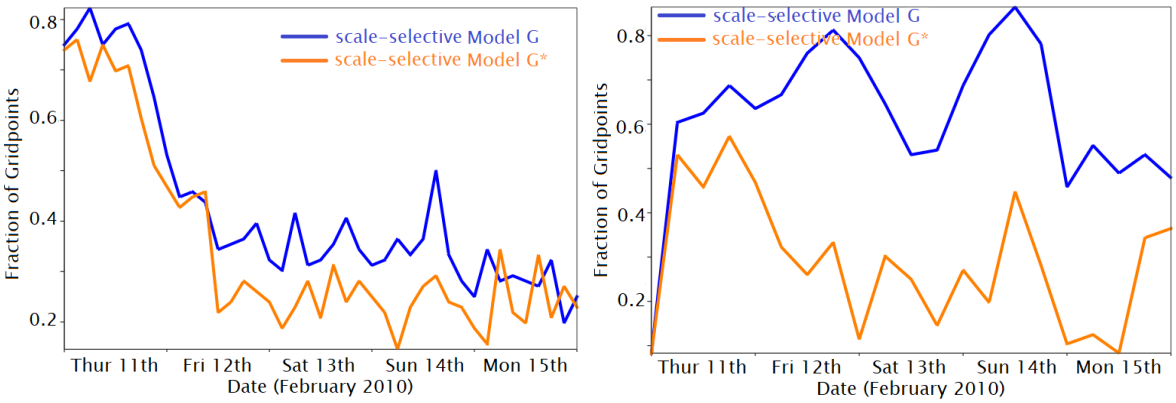

Figure 5.17: The 500 hPa Geopotential accuracy of the scale-selective Models G and G* is assessed using Metric G for low-resolution (left) and high-resolution (right) in the European region for February 2010.

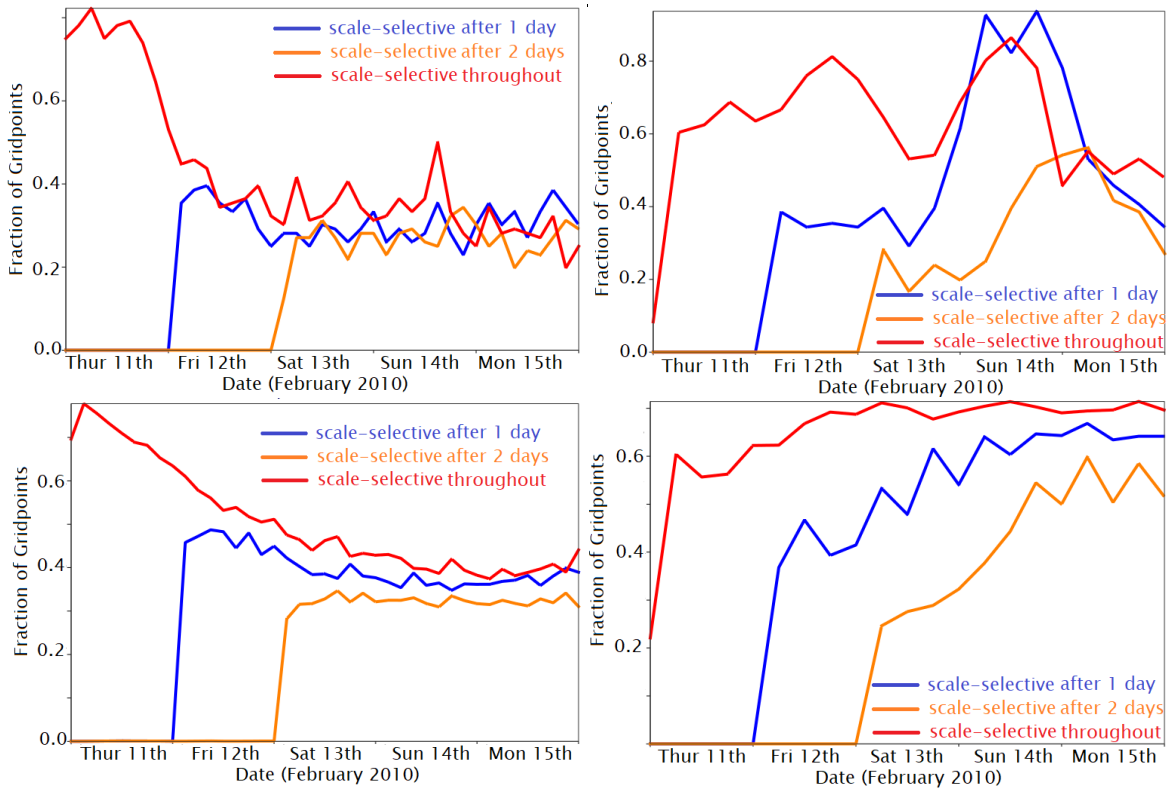

Figure 5.18: The 500 hPa Geopotential accuracy in the European region (top row) and for the whole globe (bottom row) is assessed for the February 2010 test-case using Metric G for runs with Model G scale-selective precision throughout (red line) or introduced after a lead time of 1 (blue line) or 2 (orange line) days, with single-precision used before this. The left- and right-hand plots show low- and high-resolution cases respectively.

The significand precision required for an accurate forecast in the February 2010 case is remarkably similar to that required in the St Jude's Day 2013 case at very high and very low

wavenumbers for both high and low resolutions, as may be seen by comparing the Model G* requirements in Tables 5.2 and 5.5. However, a higher precision is notably required in the Jude's Day low-resolution case for some quite low wavenumbers: 7 bits can be safely applied in the St Jude's case for wavenumbers 8 and higher, but will work for wavenumbers 5 and higher in the February 2010 case. As for February 2014, in February 2010 there is some advantage to delaying the onset of reduced precision in the Europe-only analysis (Figure 5.18), though this advantage disappears by day 5 in the low-resolution case and only occurs for day 5 onwards in the high-resolution case, suggesting that it may not be significant. For the global analysis, delayed onset is again somewhat beneficial in the low-resolution case, but fails to improve the forecast accuracy in the high-resolution case by a sufficient amount for the Metric G fraction of grid-points to fall below the 'accurate' threshold of 0.32.

5.5. Overall Results

The test cases explored in Section 5.4 demonstrated that the level of precision required for an accurate forecast at each wavenumber in the Open IFS is dependent on the model resolution, but that, although there are some slight variations, levels of precision are similar for a given model resolution for all four test cases. Figure 5.19 plots the number of significand bits as a function of the minimum wavenumber at which each can be applied in Model G* for both resolutions to illustrate this. The most consistent, albeit slight, differences between cases suggest that a relatively high precision is required for the 'normal' conditions (October 2009), at both low- and high-resolution, and a relatively low precision is required for the St Jude (October 2013) and high atmospheric predictability (February 2010) examples. This small sample therefore suggests that the 'storminess' of the atmosphere does not have a strong impact on the precision required for an accurate forecast. It is somewhat surprising that the required precision is lower for the high-predictability February 2010 case, where the ensemble spread would be especially small and the Metric G criterion for an accurate forecast (that no more than one third of the grid-points lie further from single-precision than the ensemble standard deviation) ought to be more difficult to meet.

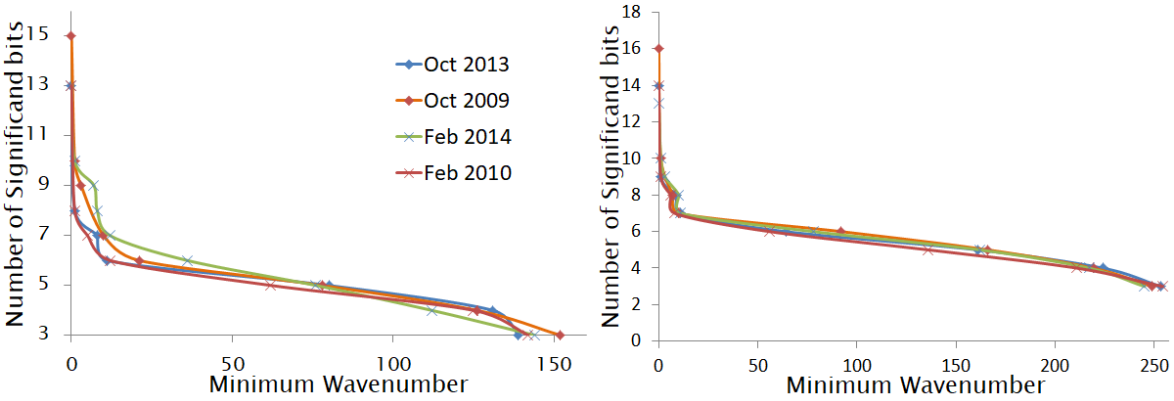

Figure 5.19: The number of significand bits is plotted as a function of the minimum wavenumber at which that number of bits can be applied for an accurate forecast according to the Model G* formulation for all four test-cases, for the $k_{\max} = 159$ (left) and $k_{\max} = 255$ (right) resolutions. Coloured lines are added to improve the legibility of the plots (see key). The blue and purple lines, corresponding to October 2013 and February 2010, are consistently low, tentatively suggesting that these test cases require a slightly lower precision, but the test cases in general yield similar results.

Nevertheless, for a given resolution there was rarely more than a one bit difference in the optimal precision at any given wavenumber between the four test-cases. This similarity suggests that it is not unreasonable, hence, to combine the four cases' results and compute the 'average' precision required for each wavenumber at each resolution that was investigated. Doing so is motivated by the fact that the results for individual test-cases are not of great practical use to anyone attempting to make operational forecasts in reduced precision, since it may be difficult to determine how stormy the atmosphere is likely to be over the course of a given forecast, and therefore what level of precision would be necessary, ahead of that forecast being made. Furthermore, it was not possible to estimate the uncertainty of the precision levels obtained for any given test-case, and it may be that the trend towards higher precision in some test-cases was mere coincidence. By combining the results of the four test-cases, it is possible to compute not only a continuous curve (not restricted to integer values of s) to show the average precision s required at each wavenumber k, but also to estimate upper and lower bounds on the resulting average precision using the maximum and minimum precision seen across the four examples. A forecaster conservatively seeking to avoid rounding errors could derive from the resulting plot the upper

bound precision regardless of the atmospheric conditions, whilst one seeking to minimise the computational cost even at the risk of introducing a few small errors might use the lower bound estimates.

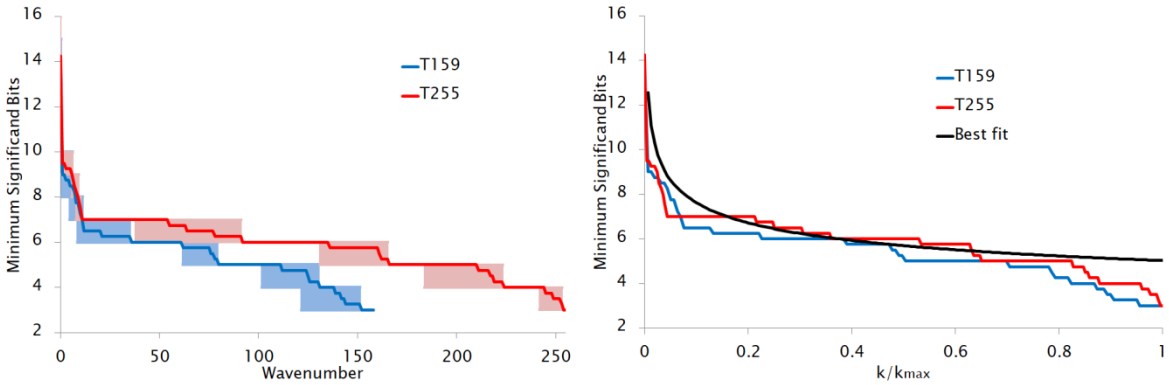

Figure 5.20: The minimum number of significand bits required for an accurate forecast in each wavenumber, as used in Model G*, is plotted, averaged over all four test-cases, for the $k_{\max} = 159$ (blue) and $k_{\max} = 255$ (red) resolutions. The left hand graph plots this directly as a function of wavenumber, and shading indicates the spread between maximum and minimum values at each wavenumber; the right-hand graph plots as a function of wavenumber k divided by the maximum resolved wavenumber $k_{\max}$, with a first-order polynomial best-fit line using the average best-fit parameters for the two resolutions.

In light of this, Figure 5.20 plots the average, maximum and minimum number of significand bits required across all four test-cases, for both $k_{\max} = 159$ and $k_{\max} = 255$ resolutions, using the optimised precision levels obtained for Model G*. The precision that a forecaster might, on average, be expected to require at a given wavenumber according to Metric G can be deduced by reading off the corresponding value of s from the graph and conservatively rounding up to the next largest integer. As has already been shown in Section 5.4, the slopes of the curves are similar at both resolutions, which motivates a still more widely applicable approach: if the wavenumber is plotted as a fraction of the resolution $k_{\max}$, also shown in Figure 5.20, the curves for both resolutions tested here almost overlap. Best-fit polynomial functions of the form $s = ak^{-b}$ for constants a and b were fitted to the $k_{\max} = 159$ and $k_{\max} = 255$ datapoints separately and yielded $b = -0.20$ and $b = -0.17$ respectively. Averaging the coefficients for the two resolutions yields the equation $s = 4.6(k/k_{\max})^{-0.18}$, which is plotted as a black line in Figure

5.20 and is consistent with the $s \propto k^{-0.2}$ relationship observed in the individual test-cases given the experimental uncertainty.

The fact that such a similar power is observed at both resolutions implies that the same power law applies at all forecast resolutions once scaling to account for effects such as horizontal diffusion and aliasing that affect smaller scales more strongly is applied, but this conjecture is beyond the scope of this study to test. Indeed, this result was obtained for a relatively low resolution compared to operational forecasts: ECMWF's operational IFS uses 1280 wavenumbers for high-resolution 10 day forecast and is triangular-cubic-octahedral (with a higher grid-point resolution than triangular linear grids would have for the same spectral resolution), and 680 wavenumbers for the 15 day ensemble forecast (ECMWF 2017a). Explicit tests would need to be carried out at operational resolutions to see whether the same power law holds for these.

5.6. Conclusions from Open IFS

The experiments described in this chapter were designed to test whether accurate results could be obtained with the ECMWF's Open IFS when scale-selective precision was applied to the model's spectral component. Weather forecasts for the real world were produced for four different start-dates that encompassed extremely stormy (St Jude's Day 2013, February 2014), moderate (October 2009) and relatively calm (February 2010) conditions at two different resolutions, both of which were considerably lower than those currently used operationally at ECMWF. The double-precision forecasts were compared to those with one or more wavenumbers in reduced significand precision, to see how low the precision in each wavenumber could be made without degrading the quality of the overall forecast.

An 'optimal' scale-selective model, labelled Model G*, was produced for each test-case at each resolution independently, and in each case represented an 'accurate' model with as few bits as possible assigned to the significand at each wavenumber. A model was deemed 'accurate' if the proportion of grid-points for which the model's prediction differed from the double-precision forecast by more than the standard deviation of the operational ECMWF ensemble members at that

grid-point was below the proportion expected by random chance, assuming a Gaussian distribution, from day 2 of the forecast onwards. This test was made using only grid-points that fell in the European region, since the models' efficacy at predicting European weather was the focus of this study, but it was also demonstrated that the models performing well in Europe successfully forecast global conditions; the precision required for accurate forecasts appeared to be similar whether European or global metrics of forecast quality were used. Of eight different meteorological variables investigated, the geopotential at 500hPa was found to be the most sensitive to reductions in model precision, and was therefore the variable used throughout the precision optimisation process. The behaviour of the model at different height levels would, of course, also need to be investigated were reduced precision to be implemented operationally, but this study was restricted to exploring those variables that might be important for an accurate five-day forecast of European near-surface weather, so that the high atmosphere was beyond its scope.

The results were broadly consistent across all four test-cases, yielding a similar constitution for the optimised Model G* for each. Between 13 and 16 bits of precision were required for the largest scales (wavenumber 0) for an accurate forecast, with higher precision necessary when the resolution was higher. As few as 3 bits were required for the very smallest scales (highest wavenumbers) close to the maximum resolvable wavenumber at each model resolution. The fact that the required precision should be resolution-dependent is not unexpected, since features that had negligible influence because they were close to the model's truncation scale in a low-resolution model can be fully resolved and more influential when a higher resolution is used. At the low wavenumbers far from the truncation scale, however, both high- and low-resolution models demonstrated a similar drop-off in the number of significand bits required s as the wavenumber k was increased: all four test-cases displayed an $s \propto k^{-0.2}$ relationship. It was beyond the time-constraints of this study to repeat the precision-optimisation analysis for another part of the globe and investigate further the relationship between the required precision and the predictability of the atmosphere in a particular geographical location or behavioural state. Precision thresholds may change in different parts of the globe where the atmosphere is more or less chaotic, for instance.

By combining the results of different test cases, it was seen that the required precision as a function of wavenumber divided by the maximum resolvable wavenumber was approximately independent of the resolution used, which may mean that the observed $s \propto k^{-0.2}$ relationship is independent of model resolution. However, verifying this would require tests at higher resolutions than could be achieved here.

For the European region, in no instance was a substantial improvement in the forecast quality at days 3, 4 and 5 achieved by delaying the onset of scale-selective reduced precision and leaving the first one or two days of the forecast in single-precision, though this did occasionally yield slight improvements. When the forecast quality metric was extended to the whole globe, some advantage to delaying the reduction in precision by one day was observed, especially in the October 2009 case, but there was no additional advantage gained by delaying for a second day.

No attempt was made to reduce precision in the grid-point or transform parts of the model, and because the spectral space calculations comprise less than 10% of the computational cost when the model is run operationally (at higher resolutions), the overall cost reduction that the scale-selective models used here would bring about in operational forecasts is small and was not explicitly calculated. Nonetheless, this study has demonstrated that using double- or even single-precision is not necessary for the Open IFS model to produce accurate forecasts. Were real reduced precision hardware to be adopted, a scale-selective approach in the spectral part of the model should be combined with a fixed reduction in precision in the grid-point and transform components to bring about the most efficient possible use of numerical precision.

6. Conclusions

This thesis sought to examine the potential for reduced numerical precision in computer models to yield more efficient, and ultimately more accurate, forecasts of the weather and climate. Specifically, this work has focussed for the first time on the possibility of reducing precision to different degrees on different spatial scales, motivated by the idea that smaller-scale variables tend to be known less precisely than larger-scale ones. To test this, three successively more complex and realistic numerical systems were considered, and in each case experiments suited to that system were carried out either to determine how low precision could be reduced without incurring significant rounding errors, or to demonstrate whether higher-resolution, low-precision forecasts can be more accurate than lower-resolution double-precision ones. Without access to real reduced precision hardware which – with the exception of single- and (less common) half-precision processors – remains still to be developed on a scale applicable to HPC, reduced precision was emulated on conventional hardware using a special computer program. This prevented the explicit measurement of the actual computational cost savings or performance improvements that could be realised with a scale-selective precision computer architecture, but the estimates made here suggest that these benefits could be substantial in some circumstances.

In this chapter, the key findings of this study will be outlined in summary (Section 6.1), its limitations will be discussed (Section 6.2) and the implications for future studies and the future of weather and climate forecasts will be explored (Section 6.3).

6.1. Summary of Key Findings

The key findings of this thesis were as follows:

Resolution is more important than precision. In the modified Lorenz ’96 system, for similar computational costs, high-resolution reduced-precision models outperformed low-resolution

double-precision models, implying that it is often beneficial for numerical precision to be traded-off in favour of increased resolution.

A Scale-selective approach to numerical precision leads to the most efficient forecasts. In modified Lorenz ’96, reducing to half-precision only in the medium-scale variables, and keeping the large-scale variables in single-precision, led to more accurate forecasts than putting everything in half-precision, but to less costly forecasts than are possible with uniform double- or single-precision. No benefit was obtained from Stochastic Processors when bit-flips were allowed in the large-scale variables, but a modest benefit was obtainable when bit-flips were restricted to medium-scale variables. In the SQG system and the Open IFS a scale-selective approach, whereby the precision of each wavenumber was tuned individually, meant that as few as 3 significand bits could be used in some parts of the model whilst larger-scale variables required as much as 19 bits. In the SQG system, this could save an estimated eighty per cent of the cost of spectral calculations; combining this with uniform precision reductions in the grid-point and transform components of the system led to an estimated overall cost saving of seventy-five per cent. The scale-selective approach only works in spectral models, however, which provides a motivation for forecasters to continue to favour such models over entirely grid-point-based alternatives.

Initial condition errors propagate more rapidly downscale than upscale under 3D Turbulence, so smaller scales require less precision. Introducing errors with the same relative size (for example, one per cent) on large and small scales in the SQG system is akin to introducing much larger absolute errors on large scales. Therefore, although the Error Doubling Time is greater on the small scales, large-scale initial errors may be of more practical importance. This is part of the explanation for why precision can be reduced more on smaller spatial scales: reduced precision rounding errors there are rapidly overwhelmed by errors propagating down from the larger scales. Observational and modelling errors are in practice likely to have a larger relative size at smaller spatial scales, but because for real-world forecasts these other errors would usually be more

significant than rounding errors on the smallest scales, this does not affect the conclusion that less precision is necessary at smaller spatial scales.

The number of bits in the exponent can be reduced to half-precision standards without penalty for variables with 10 bits or fewer in the significand in the Lorenz '96 and SQG systems. In the Lorenz '96 system, there was negligible difference between half-precision models (10 bit significand, 5 bit exponent) and models with a half-precision significand and single-precision exponent (10 bit significand, 8 bit exponent). In the SQG system, likewise, wavenumbers accurately represented with 10 significand bits or fewer required only 5 bits for the exponent. Reducing to fewer than 5 bits caused crashes in both these models, suggesting that retaining at least 5 bits in the exponent may be necessary even when there are fewer bits than this in the significand. This finding is specific to these systems: in other systems, a 5 bit exponent may not be sufficient to capture the required dynamic range even at the smallest scales, and could cause crashes.

A similar power law describes the relationship between the number of significand bits required and the wavenumber in both spectral models investigated. In the SQG system, it was found that the number of significand bits required s scales with wavenumber k according to $s \propto k^{-1/3}$. In the Open IFS, the relationship was $s \propto k^{-0.2}$ across all four test cases, which can be averaged to prescribe of the value of s required at each wavenumber.

This power law may be independent of the resolution. The same power law was observed at both resolutions tested in the Open IFS. When s was plotted against the wavenumber scaled to the resolution, $k/k_{\max}$, where higher resolutions have a larger $k_{\max}$, the curves were almost identical for $k_{\max} = 159$ and $k_{\max} = 255$. This information could one day be used to carry out optimised-precision runs of the operational IFS independent of the resolution or atmospheric conditions. The similarity of the power laws for the Open IFS and the SQG system may mean that the relationship between wavenumber and minimum applicable significand precision is universal.

Delaying the onset of reduced precision can be beneficial in some circumstances. In the Lorenz ’96 system, one of the parameterised reduced-precision models did not perform as well as the same model in double-precision when implemented from the outset, but the skill improved substantially when the onset of scale-selective precision was delayed by as little as the equivalent of one hour. In the real-world Open IFS test cases, a model that was not accurate when implemented from the outset became accurate if implemented after a delay of one day for some test cases, especially when the system was run at a higher resolution or global forecasts were considered. The two systems, despite being very different in their complexity, hence demonstrated very similar results in this respect.

Extreme weather can be forecast accurately using scale-selective precision. For forecasts of four Open IFS test cases, two of which pertained to storms of recent years, the scale-selective model of Open IFS was almost indistinguishable from a single-precision alternative at predicting 10m wind by eye, and the twelve quantities that were tested were forecast accurately in scale-selective precision, falling within the ensemble standard deviation from a single-precision control.

6.2. Limitations and Scope for Future Research

Although the degree to which high numerical precision is less necessary on smaller spatial scales is of theoretical interest, the primary aim of this thesis is a practical one: to demonstrate the efficacy of reduced precision forecasts so that operational predictions of the weather and climate can be made more efficiently, with real-world benefits either in terms of energy savings or improvements in forecast accuracy. A study such as this must face inevitable limitations in the degree to which it is able to satisfy such an aim. Scale-selective reduced precision was tested here using both idealised and real-world models, each of which has its own drawbacks. An idealised model is cheaper to run and easier to analyse but is limited in its applicability to real-world forecasters. None of the results

obtained using the modified Lorenz ’96 or SQG systems can be directly applied to operational forecasts; they can only act as pointers to suggest possible ways in which it might be useful to test reduced precision in real-world models. Further research would be necessary to carry out such tests before reduced precision could be applied operationally.

An attempt was made in this thesis to carry out some such real-world tests, using the ECMWF’s Open IFS, but whilst it is directly applicable to real-world forecasters, the computational expense of such a complex model limits the number of tests that can be carried out within a limited time. One particular metric was used here to analyse the efficacy of scale-selective reduced precision forecasts for four specific test cases, and although an effort was made to ensure that these represented as wide a range of possible weather scenarios as possible, the investigation was not sufficiently rigorous to guarantee that the scale-selective precision levels obtained would work for all the important output variables in all weather conditions. More test cases would be needed to confirm whether the required levels of precision at different spatial scales are genuinely independent of the atmospheric conditions, as the results of Chapter 5 tentatively suggest they are. Other quantities than the geopotential at 500 hPa (which was chosen here because it was the most sensitive to reduced precision of the twelve quantities explored) and other metrics than the percentage of grid-points falling outside the standard deviation from the ‘truth’ should be used to verify that precision is not degraded so much as to lower the accuracy of forecasts. Furthermore, this study was not able to explore the precision reductions that could be safely applied to the grid-point and transform components of the Open IFS.

Another theoretical contribution of this thesis, aside from the work on reduced precision, was the exploration of error propagation between scales in the SQG system described in Chapter 3. This work is not intended to be a definitive study of the SQG system, but insofar as that system, which remains a useful analogue of the real atmosphere because of its cascading behaviour for the spectral energy and enstrophy, has been investigated relatively little since its 1970s inception except for a few sporadic studies, the work presented in this thesis can constitute another stage in the ongoing exploration of its properties and their applicability (or not) to Earth’s atmosphere. Here, the propagation of errors between scales was only explored to the extent of explaining the

success of a scale-selective approach to reduced precision in the SQG system; further studies would be necessary to investigate the details of that system's properties and their implications.

Reduced precision in the exponent of numerical variables was explored in the Lorenz '96 and SQG cases, but the dynamical range of variables is likely to be very different for real-world atmospheric models, and will differ between different quantities. It is therefore by no means certain that numbers that can be represented with 10 significand bits or fewer in a real-world model could also tolerate 5 bits in the exponent, as was the case with the idealised systems investigated here. Further work would be needed to successfully implement a reduced precision exponent in the Open IFS, and to apply reduced precision to the grid-point part of that model. Future studies might also investigate the potential to re-scale variables in idealised or operational models to bring them within the dynamic range representable with fewer than 5 bits in the exponent. Furthermore, there is more work to be done exploring in detail the effects of delaying the reduction in precision in either the significand or the exponent in Open IFS, a topic that was only touched upon in this thesis because of constraints in time and computational resources.

The scale-selective approach to reduced precision with which this thesis was largely concerned can only be applied in spectral space, which renders it completely inapplicable to entirely grid-point models, such as finite volume models, and limits its usefulness in models for which a large part of the computational cost is associated with grid-point calculations. Further work might explore the possibility, however, of scale-selectively reducing numerical precision in a multigrid context, whereby lower precision could be utilised for a finer mesh embedded within selected grid-squares of a larger-scale, higher-precision grid.

Perhaps the most serious limitation of most theoretical studies of reduced precision, including this one, is the inability to test the technique using real-world hardware. Although the emulators used in this thesis for Mixed Precision and Stochastic Processors were designed to replicate as closely as possible the effects that real-world hardware would have, it was impossible to measure explicitly the computational speed-up that could practically be achieved. The emulator program used additional computer power to run on standard double-precision hardware, rather than saving resources, and all the cost savings estimated here were speculations that assumed the

maximum possible efficiency for reduced precision hardware that in large part remains hypothetical. FPGAs can provide some hints of the speed-ups that might be achieved, but can only be used with fairly basic programs. Until hardware that is developed on which large-scale models can be run with fewer than 23 bits in the significand, it will be difficult to demonstrate decisively what the benefits of scale-selective reduced numerical precision could be in terms of cost savings and the increased model resolution or complexity into which these savings might be translated.

6.3 Towards a Scale-Selective Reduced Precision Future for Weather and Climate Forecasts

Today the world faces an unprecedented century of uncertainty. How quickly, if at all, can greenhouse gas emissions be curtailed? How rapidly and by how much will the planet warm as a result of these emissions, and will natural feedbacks be initiated that serve to stymie or accentuate this? How will local climates change, and different species suffer, because of global warming, and by how much will global sea-levels rise? Can a method be found to extract carbon dioxide from the atmosphere, or will more drastic ways of artificially engineering the planet's climate system need to be sought? Answering these and myriad other questions regarding the future of the planet in a changing climate will be difficult, but the task will be made easier by reliable and accurate climate forecasts. It is expected that, whatever else may happen, extreme weather will become increasingly prevalent across the globe, and accurate weather forecasts will therefore be increasingly essential to protect human lives and infrastructure from the worst possible effects.

In light of this, it is imperative that we develop weather and climate forecast models that are as accurate as possible, which means that they will need to be as efficient as possible. The world's energy supplies are limited, and the majority of the global electricity demand is still met by the very same fuels whose combustion causes the climatic change that necessitates accurate climate predictions and makes weather predictions all the more difficult. We will not have indefinite resources with which to power the supercomputers of tomorrow, so will need to use those resources

that are available very carefully in order to utilise their maximum potential. ‘Business as usual’ will be no more excusable in the weather and climate forecasting arena than it is in the wider economy; old assumptions will have to be challenged, and old inefficiencies will have to be remedied.

Challenging the double-precision paradigm is just one component of the forecasting revolution that will need to take place over the coming years. Given the results presented in this thesis, it seems to be an unjustifiable waste of resources that we cannot spare, to represent with 64 bits of precision every number in every computer model, even when initial conditions are observed to a much lower precision than this or reducing precision in some places does not influence the accuracy of an ensemble forecast. Reduced precision hardware will enable the same physics to be computed much faster and with less of an energy cost, or much better forecasts to be produced for the same energy costs, than today’s hardware designed for double-precision by default. Such hardware has not yet been developed, but even if the impetus provided by the weather and climate community is not sufficient to merit its manufacture, machine-learning algorithms that require a similar bitwise efficiency are already creating a demand for a new paradigm in High Performance Computing whereby high-precision is not taken for granted. Specifiable precision for each component of computer code could become a possibility within the next decade, and ought to become normality within a quarter of a century if computing is to play its part in bringing about a more sustainable world. This could give rise to very considerable cost savings in the spectral part of forecast models, brought about by applying the wavenumber-dependent precision approach explored in this thesis. In addition, a ‘multigrid’ approach that uses finer grids embedded within coarser ones to bring about scale-selectivity in grid-point space could enable even greater savings in a wider range of atmospheric models, and is something that should be explored.

But reduced precision hardware will bring no benefits without software that can exploit it. The emulator used in this thesis and elsewhere has been used to demonstrate that highly specifiable precision is not only possible to program on the software side, it can also be implemented without introducing significant errors relative to double-precision computer models. This thesis has demonstrated that a scale-selective approach to reduced precision could enable the maximum possible efficiency in forecasts: that is, the optimal balance between high forecast accuracy and low

computational cost. It was shown that variables need much less precision at the smallest scales than at large scales, and that even global-scale variables can be represented with much less than double-precision without an impact on the model accuracy. It was even possible to formulate a power law relationship that could conceivably be used to predict the amount of precision required for each spatial scale in a full-complexity atmospheric model. If forecasting models continue to solve the equations that describe Earth's atmosphere in spectral space – or indeed switch to using spectral space for reasons of computational efficiency – the scale-selective approach could become an integral part of the forecasts of the future. It is my hope that this thesis has played a small but not insignificant part in moving the weather and climate community towards this new horizon.

Bibliography

Arnold H. 2013. Stochastic Parameterisation and Model Uncertainty. DPhil Thesis,

Arnold H, Moroz I, Palmer TN. 2013. Stochastic Parameterisations and Model Uncertainty in the Lorenz ’96 System. *Philos. Trans. R. Soc. A* **371** : 20110479. doi:10.1098/rsta.2011.0479

Augier P, Lindborg E. 2012. A New Formulation of the Spectral Energy Budget of the Atmosphere, with Application to Two High-Resolution General Circulation Models. *J. Atmos. Sci.* **70** : 2293–2308.

Bailey DH, Borwein JM. 2009. *High-Precision Computation and Mathematical Physics*. Berkeley, CA; 1-14 pp. http://crd-legacy.lbl.gov/~dhbailey/dhbpapers/dhb-jmb-acat08.pdf.

Bauer P, Thorpe A, Brunet G. 2015. The quiet revolution of numerical weather prediction. *Nature* **525** : 47–55.

Blumen W. 1978. Uniform Potential Vorticity Flow: Part I. Theory of Wave Interactions and Two-Dimensional Turbulence. *J. Atmos. Sci.* **35** : 774–783.

Buizza R. 2010. The Value of a Variable Resolution Approach to Numerical Weather Prediction. *Mon. Weather Rev.* **138** : 1026–1042. doi:10.1175/2009MWR3077.1

Capet X, Klein P, Hua BL, Lapeyre G, McWilliams JC. 2008. Surface Kinetic Energy Transfer in Surface Quasi-Geostrophic Flows. *J. Fluid Mech.* **604** : 165–174. doi:10.1017/S0022112008001110

Charney JG, Eliassen A. 1949. A Numerical Method for Predicting the Perturbations of the Middle Latitude Westerlies. *Tellus* **1** : 38–54. doi:10.3402/tellusa.v1i2.8500

Christensen HM, Moroz IM, Palmer TN. 2015. Simulating Weather Regimes: Impact of Stochastic and Perturbed Parameter Schemes in a Simple Atmospheric Model. *Clim. Dyn.* **44** : 2195–2214. doi:10.1007/s00382-014-2239-9

Constantin P, Majda AJ, Tabak E. 1994. Formation of Strong Fronts in the 2D Quasigeostrophic Thermal Active Scalar. *Nonlinearity* **7** : 1495–1533.

Dawson A, Düben PD, MacLeod DA, Palmer TN. 2017. Reliable Low Precision Simulations in Land Surface Models. *Clim. Dyn.* 1–10. doi:10.1007/s00382-017-4034-x

Düben PD, Palmer TN. 2014. Benchmark Tests for Numerical Weather Forecasts on Inexact Hardware. *Am. Meteorol. Soc. Mon. Weather Rev.* **142** : 3809–3829. doi:10.1175/MWR-D-14-00110.1

Düben PD, Joven J, Lingamneni A, McNamara H, De Micheli G, Palem K V, Palmer TN. 2014a. On the Use of Inexact, Pruned Hardware in Atmospheric Modelling. *Philos. Trans. R. Soc. A* **372** : 20130276. doi:10.1098/rsta.2013.0276

Düben PD, McNamara H, Palmer TN. 2014b. The Use of Imprecise Processing to Improve Accuracy in Weather and Climate Prediction. *J. Comput. Phys.* **271** : 2–18. doi:10.1016/j.jcp.2013.10.042

Düben PD, Russell FP, Niu X, Luk W, Palmer TN. 2015. On the Use of Programmable Hardware and Reduced Numerical Precision in Earth System Modelling. *J. Adv. Model. Earth Syst.* **7** : 1393–1408. doi:10.1002/2015MS000494

Düben PD, Subramanian A, Dawson A, Palmer TN. 2017. A study of reduced numerical precision to make superparameterization more competitive using a hardware emulator in the OpenIFS model. *J. Adv. Model. Earth Syst.* **9** : 566–584. doi:10.1002/2016MS000862

Durran DR, Reinecke PA, Doyle JD. 2013. Large-scale Errors and Mesoscale Predictability in Pacific Northwest Snowstorms. *J. Atmos. Sci.* **70** : 1470–1487. doi:10.1175/JAS-D-12-0202.1

Durran DR, Gingrich M. 2014. Atmospheric Predictability: Why Butterflies are Not of Practical Importance. *J. Atmos. Sci.* **71** : 3476–2488. doi:10.1175/JAS-D-14-0007.1

ECMWF. 2012a. *IFS Documentation - Cy38r1 Part III: Dynamics and Numerical Procedures*. Reading; 1-29 pp.

ECMWF. 2012b. *IFS Documentation - Cy38r1 Part IV: Physical Processes*. Reading; 1-189 pp.

ECMWF. 2015. *User Guide to ECMWF Forecast Products*. Reading; 1-129 pp.

ECMWF. 2017a. Operational Configurations of the ECMWF Integrated Forecasting System. *Doc. Support* https://www.ecmwf.int/en/forecasts/documentation-and-support (Accessed November 10, 2017).

ECMWF. 2017b. Supercomputer. *ECMWF* https://www.ecmwf.int/en/computing/our-facilities/supercomputer (Accessed November 29, 2017).

ECMWF. 2017c. MARS Catalogue. http://apps.ecmwf.int/ (Accessed December 6, 2017).

Eden P. 2009. Weather Log: October 2009. *Weather* **64** : I – IV.

Eden P. 2010. Weather Log: February 2010. *Weather* **65** : I – IV.

Eden P. 2013. Weather Log: October 2013. *Weather* **68** : I – IV.

Eden P. 2014a. Weather Log: November 2013. *Weather* **69** : I – IV.

Eden P. 2014b. Weather Log: February 2014. *Weather* **69** : I – IV.

Fenlasen J, Stallman R. 1997. *GNU-gprof: the GNU profiler*. Boston; https://ftp.gnu.org/old-gnu/Manuals/gprof-2.9.1/.

Fournier A, Fussell D, Carpenter L. 1982. Computer Rendering of Stochastic Models. *Commun. ACM* **25** : 371–384.

Free Software Foundation Inc. 2016. Welcome to the Home of GNU Fortran. *06/02/2016* https://gcc.gnu.org/fortran/ (Accessed July 15, 2016).

Hansen J, Penland C. 2007. On Stochastic Parameter Estimation using Data Estimation. *Phys. D Nonlinear Phenom.* **230** : 88–98. doi:10.1016/j.physd.2006.11.006

Hansen J … Zachos JC. 2013. Assessing "dangerous climate change": required reduction of carbon emissions to protect young people, future generations and nature. *PLoS One* **8** : e81648. doi:10.1371/journal.pone.0081648

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.

Held IM, Pierrehumbert RT, Garner ST, Swanson KL. 1995. Surface Quasi-Geostrophic Dynamics. *J. Fluid Mech.* **282** : 1–20. doi:10.1017/S0022112095000012

Hortal M, Simmons AJ. 1991. Use of Reduced Gaussian Grids in Spectral Models. *Mon. Weather Rev.* **119** : 1057–1074. doi:10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2

Hughes I, Hase T. 2010. *Measurements and their Uncertainties*. 1st ed. Cambridge University Press: Cambridge;

IEEE. 2008. IEEE Standard for Floating-Point Arithmetic. *IEEE Stand.* 1–70.

Intel. 2016. Intel Fortran Compiler. https://software.intel.com/en-us/fortran-compilers (Accessed July 15, 2016).

Karakonstantis G, Roy K. 2011. Voltage Over-Scaling: A Cross-Layer Design Perspective for Energy Efficient Systems. *20th European Conference on Circuit Theory and Design*, Linköping, Sweden, 548–551

Kogge P … Yelick K. 2008. *Exascale Computing Study: Technology Challenges in Acheiving Exascale Systems*. Notre Dame, Indiana; 1-278 pp.

Kumar A, Barnston A, Hoerling M. 2001. Seasonal Predictions, Probabilistic Verifications, and Ensemble Size. *Am. Meteorol. Soc. J. Clim.* **14** : 1671–1676.

Lapillonne X, Fuhrer O. 2014. Using Compiler Directives to Port Large Scientific Applications to GPUs: An Example from Atmospheric Science. *Parallel Process. Lett.* **24** : doi:10.1142/S0129626414500030

Leith CE. 1971. Atmospheric Predictability and Two-Dimensional Turbulence. *J. Atmos. Sci.* **28** : 145–161. doi:10.1175/1520-0469(1971)028<0145:APATDT>2.0.CO;2

Lingamneni A, Enz C, Nagel J-L, Palem K, Piguet C. 2011. Energy Parsimonious Circuit Design through Probabilistic Pruning. *Des. Autom. Test Eur. Conf. Exhib.* 1–6. doi:10.1109/DATE.2011.5763130

Lingamneni A, Kirthi K, Enz C, Karp RM, Palem K V, Piguet C. 2012. Algorithmic Methodologies for Ultra-efficient Inexact Architectures for Sustaining Technology Scaling. *Proceedings of the 9th Conference on Computing Frontiers*, 3–12

Lorenz E. 1963. Deterministic Non-Periodic Flow. *J. Atmos. Sci.* **20** : 130–141.

Lorenz E. 1969. The Predictability of a Flow which possesses Many Scales of Motion. *Tellus* **21** : 289–307. doi:10.3402/tellusa.v21i3.10086

Lorenz E. 1993. *The Essence of Chaos*. University of Washington Press: Seattle;

Lorenz E. 2006a. Regimes in Simple Systems. *Am. Meteorol. Soc. J. Atmos. Sci.* **63** : 2056–2073. doi:10.1175/JAS3727.1

Lorenz E. 2006b. Predictability - a Problem Partly Solved. *Predictability of Weather and Climate*, R. Palmer T & Hagedorn, Ed., Cambridge University Press: eds. T. N. Palmer & R. Hagedorn, 40–58

Majda AJ, Bertozzi AL. 2002. *Vorticity and Incompressible Flow*. Cambridge University Press:

Cambridge;

Malardel S, Wedi NP. 2016. How does Subgrid-scale Parametrization influence Nonlinear Spectral Energy Fluxes in Global NWP Models? *J. Geophys. Res. Atmos.* **121** : 5395–5410. doi:10.1002/2015JD023970

Met Office. 2013. "St Jude"s Day' storm - October 2013. http://www.metoffice.gov.uk/about-us/who/how/case-studies/st-judes-day-storm-oct-2013 (Accessed July 25, 2017).

Michalakes J, Vachharajani M. 2008. GPU Accelleration of Numerical Weather Prediction. *Parallel Process. Lett.* **18** : 531–548. doi:10.1142/S0129626408003557

Mitchell J, Budich R, Joussaume S, Lawrence B, Marotzke J. 2012. Infrastructure Strategy for the European Earth System Modelling Community 2012-2022. *ENES Rep. Ser. 133*

Nastrom GD, Gage KS. 1985. A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft. *J. Atmos. Sci.* **42** : 950–960.

Newman PA, Coy L, Pawson S, Lait LR. 2016. The Anomalous Change in the QBO in 2015-2016. *Geophys. Res. Lett.* **43** : 8791–8797.

NVidia. 2018. *Training with Mixed Precision*. 1-19 pp. http://docs.nvidia.com/deeplearning/sdk/pdf/Training-Mixed-Precision-User-Guide.pdf.

Oberstar EL. 2007. Fixed-point Arithmetic & Fractional Math. *Superkits Whitepapers* http://www.superkits.net/whitepapers/Fixed Point Representation & Fractional Math.pdf (Accessed November 1, 2016).

Olsen J. 2004. *Realtime Procedural Terrain Generation*. University of Southern Denmark;

Ott E … Yorke JA. 2004. A Local Ensemble Kalman Filter for Atmospheric Data Assimilation. *Tellus* **56** : 415–428.

Palmer TN. 2014. More Reliable Forecasts with Less Precise Computations: A fast-Track Route to Cloud-Resolved Weather and Climate Simulators? *Phys. Trans. R. Soc. A* **372** : 20130391. doi:10.1098/rsta.2013.0391

Palmer TN, A D, Seregin G. 2014. The Real Butterfly Effect. *Nonlinearity* **27** : R123–R141. doi:10.1088/0951-7715/27/9/R123

Pierrehumbert RT, Held IM, Swanson KL. 1994. Spectra of Local and Non-Local Two

Dimensional Turbulence. *Chaos, Solitons & Fractals* **4** : 1111–1116. doi:10.1016/0960-0779(94)90140-6

Press WH, Flannery BP, Teukolsky SA, Vetterling WT. 1992. Fast Fourier Transform. *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing*, Cambridge University Press: Cambridge, 490–529

Regan R. 2015. The Spacing of Binary Floating-Point Numbers. *Explor. Bin.* http://www.exploringbinary.com/the-spacing-of-binary-floating-point-numbers/.

Richardson D, Bidlot J, Ferranti L, Ghelli A, Hewson T, Janousek M, Prates F, Vitart F. 2010. Verification statistics and evaluations of ECMWF forecasts in 2009-2010. *ECMWF Tech. Memo.* **635** : 2–3.

Richardson LF. 1922. *Weather Prediction by Numerical Process*. Cambridge University Press;

Riley KF, Hobson MP, Bence SJ. 2006. *Mathematical Methods for Physics and Engineering*. 3rd ed. Cambridge University Press: Cambridge;

Rotunno R, Snyder C. 2008. A Generalisation of Lorenz' Model for the Predictability of Flows with Many Scales of Motion. *J. Atmos. Sci.* **65** : 1063–1076. doi:10.1175/2007JAS2449.1

Roulston MS, Smith LA. 2002. Evaluating Probabilistic Forecasts using Information Theory. *AMS Mon. Weather Rev.* **130** : 1653–1660. doi:10.1175/1520-0493(2002)130<1653:EPFUIT>2.0.CO;2

Russell FP, Düben PD, Niu X, Luk W, Palmer TN. 2015. Architectures and precision analysis for modelling atmospheric variables with chaotic behaviour. *Proc. - 2015 IEEE 23rd Annu. Int. Symp. Field-Programmable Cust. Comput. Mach. FCCM 2015* 171–178. doi:10.1109/FCCM.2015.52

Russell FP, Duben PD, Niu X, Luk W, Palmer TN. 2017. Exploiting the Chaotic Behaviour of Atmospheric Models with Reconfigurable Architectures. *Comput. Phys. Commun.* **221** : 160–173. doi:10.1016/j.cpc.2017.08.011

Sartori J, Kumar R. 2011. Architecting Processors to allow Voltage/Reliability Tradeoffs. *CASES Proc. 14th Int. Conf. Compil. Archit. Synth. Embed. Syst.* 115–124. doi:10.1145/2038698.2038718

Simmons AJ, Hollingsworth A. 2002. Some Aspects of the Improvement in Skill of Numerical Weather Prediction. *Quaterly J. R. Meteorol. Soc.* **128** : 647–677.

Sloan J, Kesler D, Kumar Rakesh RA. 2010. A Numerical Optimisation-based Methodology for Application Robustification: Transforming Applications for Error Tolerance. *IEEE/IFIP Int. Conf. Dependable Syst. Networks* 161–170.

Smith S. 2009. QGModel. http://www.cims.nyu.edu/~shafer/tools/index.html (Accessed February 28, 2017).

Smith SA, Hammett GW. 1997. Eddy Viscosity and Hyperviscosity in Spectral Simulations of 2D Driftwave Turbulence. *Phys. Plasmas* **4** : 978–990. doi:10.1063/1.872210

Snowdon D, Ruocco S, Heiser G. 2005. Power Management and Dynamic Voltage Scaling: Myths and Facts. *Proceedings of the 2005 Workshop on Power Aware Real-time Computing*, Vol. 12 of, Jersey City, USA, 1–7

Song S, Ge R, Feng X, Cameron KW. 2009. Energy Profiling and Analysis of the HPC Challenge Benchmarks. *Int. J. High Perform. Comput. Appl.* **23** : 265–276. doi:10.1177/1094342009106193

Top500.org. 2016. Cray XC40, Xeon E5-2695v4 18C 2.1GHz, Aries interconnect. *Top 500* https://www.top500.org/system/178749 (Accessed March 23, 2017).

Toth Z, Kalnay E. 1993. Ensemble Forecasting at NMC: The Generation of Perturbations. *Bull. Am. Meteorol. Soc.* **74** : 2317–2330. doi:10.1175/1520-0477(1993)074<2317:EFANTG>2.0.CO;2

Le Traon PY, Klein P, Hua BL. 2008. Do Altimeter Wavenumber Spectra Agree with the Interior or Surface Quasigeostrophic Theory? *J. Phys. Oceanogr.* **38** : 1137–1142. doi:10.1175/2007JPO3806.1

Tribbia JJ. 1997. Weather Prediction. *Economic Value of Weather and Climate Forecasts*, R.W. Katz and A.H. Murphy, Eds., Cambridge University Press, 1–18

Tulloch R, Smith KS. 2006. A theory for the atmospheric energy spectrum: Depth-limited temperature anomalies at the tropopause. *PNAS* **103** : 14690–14694. doi:10.1073/pnas.0605494103

Vana F, Düben PD, Lang S, Palmer T, Leutbecher M, Salmond D, Carver G. 2017. Single Precision in Weather Forecasting Models: An Evaluation with the IFS. *Mon. Weather Rev.* **195** : 495–502. doi:10.1175/MWR-D-16-0228.1

Vana F, Carver G, Lang S, Leutbecher M, Salmond D, Düben P, Palmer T. 2016. Single Precision IFS. *ECMWF Newsl. 148* 20–23.

Villiers M, White D. 2014. Near-surface winds and wind shear at four airports during the St Jude's day storm. *Weather* **69** :

Walton M. 2016. NVidia Unveils First Pascal Graphics Card, the Monstrous Tesla P100. *ArsTechnica* http://arstechnica.co.uk/gadgets/2016/04/nvidia-tesla-p100-pascal-details/ (Accessed April 11, 2016).

Wedi N, Hamrud M, Mozdzynski G. 2013. The ECMWF Model: Progress and Challenges. *Seminar on Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling*, 1–14

Weisheimer A, Palmer TN. 2014. On the Reliability of Seasonal Climate Forecasts. *J. R. Soc. Interface* **11** : 20131162. doi:10.1098/rsif.2013.1162

Wilks D. 2005. Effects of Stochastic Parameterisations in the Lorenz '96 System. *Q. J. R. Meteorol. Soc.* **131** : 389–407. doi:10.1256/qj.04.03

Wilks D. 2006. *Statistical Methods in the Atmospheric Sciences*. Second. Elsevier Academic Press: Cambridge, Massachusetts;

Appendix: Parameterising the SQG System

Parameterisation of the SQG system follows the same general method of Wilks (2005) used in the Lorenz '96 system. It takes place in the following three stages. First, low- and high-resolution models of the SQG system are compared to determine typical differences between them. The model is spun up for a certain number of timesteps, T, which is referred to hereafter as the 'spin-up time'. T is chosen to match the spin-up time of the models to which the parameterisation is to be applied; here, the short-term forecasting setup (Setup S) is used, with $T = 50000$. With this point as the initial condition, high- and low- resolution simulations are run, and at each timestep for 100 timesteps the difference in the right hand side of Equation (3.10) is measured between them. Either this can be done for every wavenumber in spectral space, or the right-hand-side can be converted to grid-point space to evaluate the difference at each grid-point. The difference between the high- and low-resolution run at a given timestep and grid-point or wavenumber is referred to as the 'observed sub-grid tendency' – that is, the direction in which the low-resolution model tends to deviate from a higher-resolution model that includes processes not visible at the lower resolution – labelled U. Where $\mathrm{rhs}_\mathrm{L}(t)$ and $\mathrm{rhs}_\mathrm{H}(t)$ are arrays containing the right-hand-side values for the low-and high-resolution runs respectively at each low-resolution grid-point or wavenumber, at timestep t,

$$U(t) = \mathrm{rhs}_\mathrm{H}(t) - \mathrm{rhs}_\mathrm{L}(t). \tag{A.1}$$

Second, a formula is constructed to predict the sub-grid tendency as a function of rhs_L. The sub-grid tendencies are plotted at 100 timesteps after the spin-up, $U(T + 100)$ (the y-axis) against $\mathrm{rhs}_\mathrm{L}(T + 100)$ (the x-axis), producing a scattering of points as shown in Figure A.1 for both spectral and grid-point examples. 100 timesteps was chosen because this is how long it takes the model to move away from the fixed initial conditions sufficiently for any bias caused by the change in resolution relative to the 'truth' spin-up of the low-resolution model to be fully realised. The fixed coefficients $a_0, .., a_3$ are then evaluated that best fit the data points through a cubic polynomial of the form,

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3. \tag{A.2}$$

Since the right hand side, like all the prognostic variables in the SQG system, is a complex quantity, this must be done for its real and imaginary parts separately.

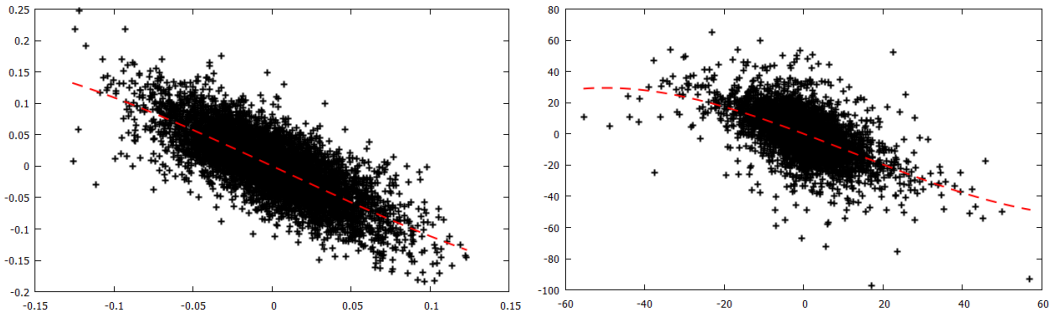

Figure A.1: Examples scatter-plots showing the sub-grid tendency after 100 timesteps, $U(T+100)$, computed from the difference between the right-hand-side of Equation (3.10) for $k_{\max}=127$ 'high resolution' and $k_{\max}=63$ 'low resolution' models, plotted against the $k_{\max}=63$ right-hand-side, $\mathrm{rhs}_{\mathrm{L}}$. In the left plot, each datapoint represents a different wavenumber component in spectral space; in the right plot, each pertains to a different grid-point when the quantities are computed in grid-point space. The plots were produced using Setup S with $T=50000$ timesteps.

The third and final step to parameterise the model is to use the deterministic equation (A.2) to correct low-resolution operational models, pushing them back towards the high-resolution 'truth'. During operational running, the unknown true sub-grid tendency is approximated using the 'parameterisation sub-grid tendency' $U_p(\mathrm{rhs}_{\mathrm{L}}) \equiv y(x) \approx U$, so that the corrected low-resolution right-hand side is,

$$\mathrm{rhs}_L^*(t) = \mathrm{rhs}_{\mathrm{L}}(t) + U_p(t) \approx \mathrm{rhs}_H(t) = \mathrm{rhs}_{\mathrm{L}}(t) + U(t). \qquad \text{(A.3)}$$

The low-resolution right-hand-side must be corrected in this way every timestep that the model continues to run in low resolution because the same 'shock' arising from the difference in resolution consistently applies.

When constructing parameterisation schemes it is assumed that the sub-grid tendency will obey the same formula for all timesteps sufficiently close to T. That this assumption was valid was verified by making best-fits to analogous scatter plots for $U(T+50)$, $U(T+80)$ $U(T+90)$ and $U(T+95)$. The best-fit coefficients for these get increasingly close to the $U(T+100)$ case, but for $U(T+90)$ onwards they match it exactly and cease to change to within the experimental error.

It is also assumed that the sub-grid tendency will depend upon the value of $\mathrm{rhs_L}$, but will not vary between different grid-points (when parameterising in grid-point space) or different wavenumbers (when parameterising in spectral space). The grid-point, spatial homogeneity is justified by the observation that the difference between the different-resolution runs is not correlated with the topography, so that on the periodic domain all grid-points are equal. This can be demonstrated by plotting $U(T+100)$ against the topographic height, whereupon no trend is observed. The assumption of scale homogeneity in spectral space is more difficult to justify theoretically, however, as one might expect the different wavenumbers to follow different relationships between U and $\mathrm{rhs_L}$, as is implied by results such as those obtained by Malardel and Wedi (2016). For this reason, initially parameterisation was attempted only in grid-point space. To apply the parameterisation scheme in the operational model, at each timestep $\mathrm{rhs_L}$ was transformed into grid-point space, $U_p(\mathrm{rhs_L})$ was added to each grid-point, then the result $\mathrm{rhs_L^*}$ was transformed back to spectral space. However, as Figure A.1 demonstrates, the polynomial fit is much less convincing in grid-point than in spectral space, and results suggested that parameterising in grid-point space was not effective in decreasing the error associated with reducing the resolution. The spectral-space parameterisation schemes, meanwhile, very effectively reduced the error induced by reducing the resolution, and could be used in combination with reduced precision to yield the results presented in Section 4.2.6, even though a theoretical justification for treating all wavenumbers the same was absent.

Since it cannot be assumed that the tendencies are scale-invariant, these findings motivated the construction of a parameterisation scheme in spectral space that can vary with wavenumber (that is, with different spatial scales). Computing the sub-grid tendency as a function of wavenumber is, however, complicated by the fact that the model is two-dimensional: a spectral field such as the sub-grid tendency that varies with k_x and k_y can be collapsed into a function of a single wavenumber $k=\sqrt{k_x^2+k_y^2}$ only by averaging over regions in phase space for which

$\sqrt{k_x^2 + k_y^2}$ is closer to k than it is to $k + 1$ or $k - 1$ according to some assumed criterion. k acts as an index for arrays in a numerical model and must therefore take integer values.

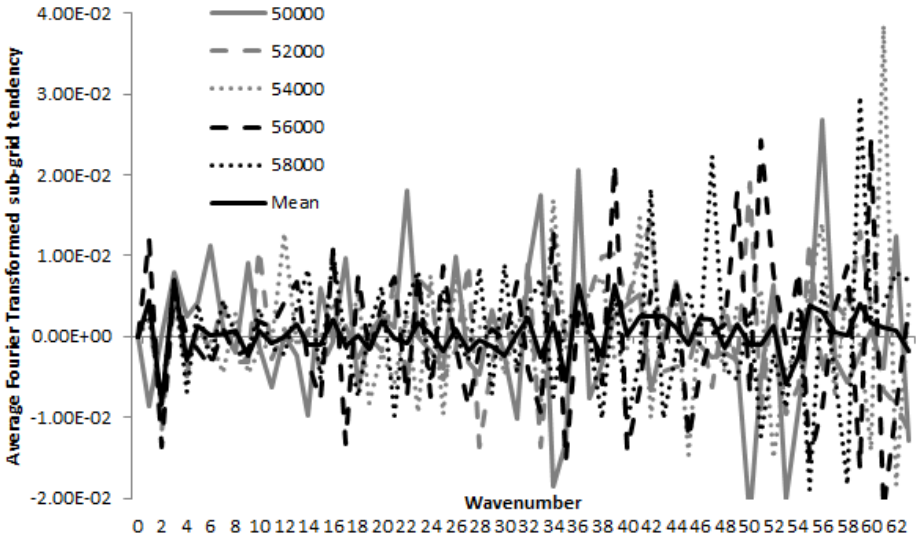

Figure A.2: The sub-grid tendency, Fourier Transformed into spectral-space after averaging over all rows and columns of grid-point space, is plotted as a function of wavenumber for each of five start times, each separated by 2000 timesteps, after spin-up during a single run of the model. Observations were made at eleven start times each separated by 1000 timesteps, but only five start-times are shown for clarity. The tendencies are averaged over the 100 timesteps following each start time. The solid black line shows the average across all eleven start times for which this was done.

To avoid making any such assumptions, a procedure was carried out to convert the tendencies observed in grid-point space (where isotropy in the x- and y-directions can be justifiably assumed) into one-dimensional spectral functions for each row and column in turn. To do this, the system was run with only the high-resolution model actively evolving forwards in time; the low-resolution right-hand-side, rhs_L, was computed at each of the 100 timesteps independently from the streamfunction output by the time-evolving high-resolution model, by truncating this variable to only the first 64 wavenumbers. A truncated version of the high-resolution right-hand-side, denoted rhs_H, was also produced at each timestep by simply taking the first 64 wavenumbers of the high-resolution right-hand side. Then, both rhs_L and rhs_H were converted into grid-point space, with each corresponding to a 128×128 grid at each timestep. Sub-grid tendencies U were

computed in grid-point space through the equation $U = \mathrm{rhs_H} - \mathrm{rhs_L}$ to produce another 128×128 grid. Power spectra showing the dependency of the tendency on wavenumber (spatial scale) were produced by computing the one-dimensional Fourier Transform of each row and column of this grid independently, yielding $128 + 128 = 256$ vectors, each with 64 components. These were averaged over all columns and rows (exploiting the assumption that the model is isotropic in the x- and y-directions in grid-point space) and over all 100 timesteps to yield a single power spectrum.

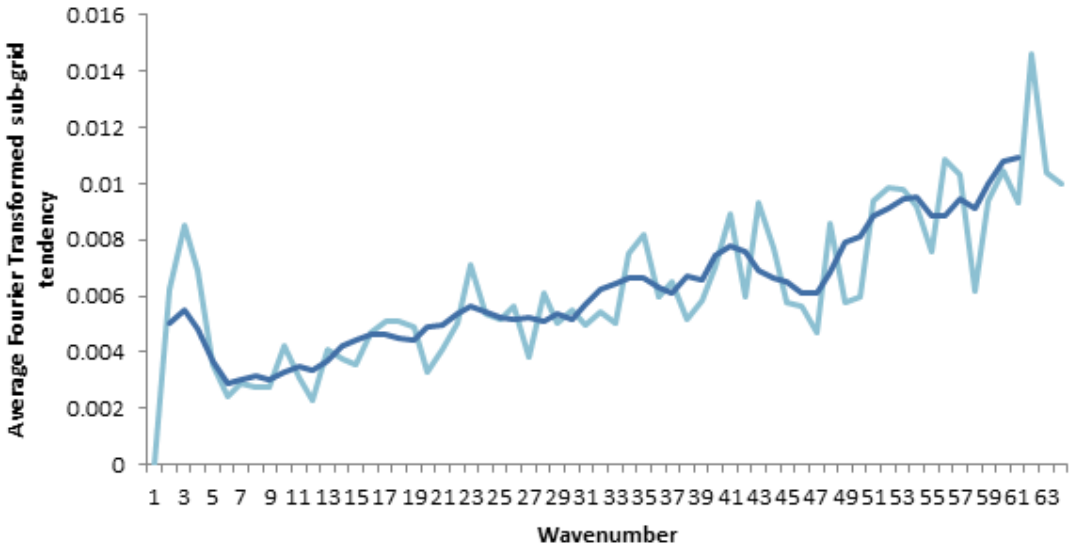

Figure A.3: The absolute values of the Fourier transformed sub-grid tendencies (Figure A.2) averaged over all eleven analysis points (pale blue line) and smoothed through a rolling average over each three wavenumbers (dark blue line).

This power spectrum was averaged over all 100 timesteps at each of eleven start times separated by 1000 timesteps between T and $T + 10000$, which are far enough apart to be uncorrelated to one another. The results, shown in Figure A.2, suggest that the tendency does indeed tend to be slightly larger in magnitude for higher wavenumbers. This is perhaps better illustrated by Figure A.3, which plots the absolute value of the sub-grid tendency at each wavenumber, averaged over all eleven start times and smoothed by averaging over every three wavenumbers. However, it is not straightforward to take this trend into account and to build a wavenumber-dependent parameterisation scheme because the tendencies oscillate randomly about zero (see Figure A.2), so the sign of the required correction cannot be predicted. Although attempts

were made to produce wavenumber-dependent parameterisation schemes by constructing functions to emulate the curve shown in Figure A.3, no improvement in accuracy of low-resolution models was thereby obtained, and an investigation to improve on this would lie beyond the scope of this study. It was hence necessary for the purposes of this investigation to treat, without justification, the sub-grid tendency as independent of wavenumber and use spectral parameterisation schemes built with this assumption, on the pragmatic basis that such schemes work well.