Seamless Evaluation of Stochastic Physics Parametrizations

Submitted by Claudio Sánchez to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics, September 2014.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: .....

Abstract

A substantial segment of the error in numerical weather prediction and climate projections comes from the intrinsic uncertainties of General Circulation Models of the atmosphere. Stochastic physics schemes are one of the preferred methods to represent the model uncertainty in Ensemble Prediction Systems, where different realizations of the same forecast are created to quantify the probabilities of different outcomes in the atmospheric flow.

Stochastic physics schemes have been successfully employed in medium-range and seasonal forecasting systems, as they increase the skill of probabilistic forecasts. Similarly it has been demonstrated than these schemes can improve certain aspects of the model's climate. However, it is still not clear whether they are a truthful representation of the model uncertainties they aim to represent.

In this thesis, a collection of stochastic physics schemes are evaluated using a seamless approach. It is found that they can improve the representation of the tropical climate and extra-tropical cyclones, but they degrade the individual representation of these processes deteriorating the deterministic skill of the model. Some important features of the model can be degraded by the stochastic physics schemes, like energy and moisture conservation on climate scales. Some closures to the schemes are proposed and successfully tested to remove or reduce some of the problems found. Alternative approaches in the development of stochastic parametrizations are also investigated.

Stochastic physics schemes have some benefits but still require further development to produce a realistic representation of model error. It is also recommended that evaluation methodologies must be expanded to include process-based diagnostics to display the realism of its perturbations.

Acknowledgements

The research documented in this thesis has been financially supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). The academic partnership between the Met Office and the University of Exeter has provided the institutional framework to develop this PhD project.

The person who deserves more credit on the development of this thesis and my personal development as a scientist is Keith Williams, my manager at the Met Office and also one of the two supervisors of this PhD. He gave me the opportunity to work for the Met Office, shared with me his enthusiasm and has had an endless patience with my bizarre questions. Thanks to him I can even make a living from atmospheric modelling!

The supervisor from the Exeter University is Matt Collins. He gave me the possibility to carry out this PhD, and he has also provided very valuable comments on the different stages of this work. I should also appreciate his very lax attitude towards the university bureaucracy, which has helped me to focus on the science and forget about the tortuous student portal of the Exeter University.

The intellectual foundations of this work lie in the endless, richful and very stochastic conversations with Glenn Shutts. His encyclopedic knowledge, ruthless criticism of stochastic physics schemes and good mood have been fundamental for my personal induction and progress in this very "noisy" field.

In order to produce the scientific material shown here, I have adapted and developed several algorithms from other people. They have been the target of many of my silly and probably annoying questions. Therefore I acknowledge the patience and help from Warren Tennant for SKEB2 and MOGREPS; Kevin Hodges, Lizzie Froude and Ruth McDonald for TRACK and its matching technique; Kalli Furtado for the Mixed-Phase scheme and Paul Earnshaw, Tim Hinton, Prince Xavier and Dan Copsey for different diagnostics and model configurations.

Many other people have helped me to forge my scientific skills and passion for the atmospheric sciences. I would not have been able to carry out this PhD without the previous dedication and scientific nourishment from Donal Murtagh and Joakim Urban. I hope they forgive me for the projects we have never finished. I am very grateful to Ricardo Garcia, Natalia Calvo and several professors and colleagues I had during my Physics degree at the Universidad Complutense de Madrid.

Now it comes the people who won't understand this thesis, and even dare to ask where they can find the weather forecast for next weekend in here. Without their support, I would not have had the courage to write a single word of this thesis.

In the first place are my parents, Angel Sanchez and Marisol Ortiz de Zarate. They have always been there supporting and encouraging me with every single step I have taken in my life. Thanks to them I have learnt to fight to the end for what I believe. The fondness and support of my sister Rita has been equally important in my life, as well as the growling of Lolo, my parent's dog. Thanks as well to my extended family: Nuria, Nacho, Marisa, Miriam, Manolo, Raquel and the adopted ones like Andres Garea, Silvia, Pilar and Juan Andres.

Despite the constant availability of my family, when you live overseas your friends become your family. My life in Exeter would have been unbearable, and hence this project doomed to fail, if it wasn't for the help of the following people amongst many others: Katka Rigdova for her dedicated friendship. James and Dorleta, so welcoming and inclusive that they even let me destroy their house from time to time! Ivan, Fede and Ian for their companionship and high frequency of interesting activities. Ben for his weirdness and careful listening to my raving nonsense. Alex Rose and Caroline Nye for making me feel at home when we shared a house. Mat Probert, Daska, Ben Rothwell and Jim for their sympathy and fun outings. Slavka for her attentiveness. Tara the killer dog, and her no less dangerous owners Kalina and Howard for their guaranteed entertainment. Natalia, Georgios and Joana for their caring and inclusive attitude. Dani Garcia and Senen Germande for their interesting discussions and lectures on pharmacology. Chloe, Liana, Trijsje and other capoeristas for their fun.

Contents

List of Figures

List of Tables

Declaration

The work carried out in this thesis includes the production of all the scientific material shown. In addition, the cyclone matching technique was developed for MetUM and the TRACK algorithm for NWP output. The majority of process-based diagnostics were adapted for my scientific needs (e.g. blocking, CCEW power-spectra or MJO time-lag composites). Some of the stochastic schemes investigated in this thesis were also developed. These are the biharmonic dissipation rate for SKEB2, SPT, SIPS and the different stochastic approaches for MPC.

Chapter 1

Introduction

The diversity of different phenomena occurring in the atmosphere is astonishingly complex and beautiful. The engine of the atmosphere, the latitudinal gradient of solar radiation, gives birth to a great diversity of clouds, precipitation events or scattering processes which produce the rich and colourful images we observe in the sky.

In addition to this beauty, the prediction of the evolution of the atmosphere is one of the most important and challenging scientific problems of our times. It is important because the accurate prediction of devastating phenomena like hurricanes, severe droughts or runaway effects in the climate system is paramount for our well-being. And it is challenging because of the number of actors able to guide and modify weather and climate patterns whose interactions encompass a wide range of temporal and spatial scales:

- Microscales of a few micrometers (snowflakes and raindrops)
- Convective scales of about one kilometer and few hours (updraft, downdraft and clouds)
- Mesoscales of a one week and a hundred kilometers (squall lines and rainfall bands)
- synoptic scales of weeks and $\sim$ 1000 km. (equatorial waves, cyclones)
- planetary scales of months and 10^4 kilometers [the Hadley circulation, Walker circulations or Madden Julian Oscillation (MJO)]

- intradecadal variability [El Niño Southern Oscillation (ENSO), solar cycle, changes in Green House Gas (GHG) concentrations]

In order to provide useful and as accurate as possible prediction of the future state of the atmosphere, the physical laws of fluid dynamics are combined with diabatic processes such as latent heat, radiative absortion/reflection, interactions with boundary conditions such as mountains, land, sea-ice and oceans, and last but not least the anthropogenic forces that have emerged in the last two centuries. All these different processes are mingled to form a General Circulation Model of the Atmosphere (GCMA or commonly abbreviated as GCM), an extremely complex tool that has been developed since the invention of the digital computer in the middle of the 20^{th} century.

In the early days, weather forecasting and climate prediction were separate sciences, the former was defined as Numerical Weather Prediction (NWP) and was crucially dependent on defining an accurate initial state and running at the highest possible resolution, whereas climate modeling and prediction sought to incorporate more complexity to capture the feedbacks amongst different Earth subsystems, like the ocean, sea-ice, atmospheric gaseous composition or the biosphere. The representation of these subsystems allowed climate models to capture the evolution of past, present and future climates. These climate models are often described now as Earth System Models (ESMs).

The division between NWP and climate modeling remained until the resolution of the latter was high enough to represent important process of day-to-day variability and the former started to incorporate more complexity and modeling elements of the Earth System such as atmospheric composition (Milton *et al.*, 2008). Since weather and climate are built over the same physical process in the atmosphere, the idea of merging climate and weather models into a "seamless" model was developed (Palmer *et al.*, 2008; Senior *et al.*, 2011).

Higher computational resources and a better understanding and representation of atmospheric processes have produced a dramatic progress in NWP forecasting and climate simulation over the last decades (Simmons and Hollingsworth, 2002; Reichler and Kim, 2008). However, many of predictions are still deficient in many aspects, models have problems to simulate important atmospheric phenomena such as blocking (D'Andrea *et al.*, 1998), the Madden Julian Oscillation (Zhang, 2005)

or the diurnal cycle of convection (Yang and Slingo, 2001).

The set of equations that defines a GCM can not be solved analytically with the exception of a few very idealized cases. The way to obtain the most accurate solution of a GCM involves the transformation of the GCM's continuous equations into a discrete set of algebraic equations, which inevitably introduces errors in the resolved spatial and temporal scales, so-called large-scale flow or resolved flow. The discretization carries a substantial problem, all the atmospheric phenomena with an impact on momentum, heat or moisture that occur in spatial scales below the truncation limit and in timescales shorter than the timestep can no longer be solved, they need to be represented with their average effect for a given large-scale forcing, the representation of these averaged processes in the GCM is defined as parametrizations.

Atmospheric parametrizations adopt the notions of statistical mechanics, where the bulk effects of the molecular motions are represented by the thermodynamic properties of the gas. However, the separation between macro and micro scales in the atmosphere is an arbitrary choice based on the computational capacity to solve the GCM. There is not a true separation of scales in the atmosphere, since its energy spectrum is quasi-continuous on all observable length scales from the planetary scales down to a few kilometers (Nastrom and Gage, 1985), and in many cases the number of subgrid scale events per gridbox is not large enough to permit the existence of a meaningful statistical equilibrium (Williams, 2005). These events are phenomena such as gravity waves, perturbations with wavelengths of 1-10km caused by orography or a rapid release of latent heat, convective clouds of a few kilometers that can organize to form large scale clusters larger than 500km in the tropics, and small-scale turbulence in the Boundary Layer caused by shear between the atmosphere and the land. Many atmospheric centers are developing GCMs with a high horizontal resolution of the order of one kilometer or less, which are able to fully or partially resolve some of these processes, these models are defined as "convection permitting models", the comparison between these high resolution GCMs with low resolution ones highlights the deficiencies of current parametrization schemes (Holloway *et al.*, 2013).

The uncertainty of a weather forecast not only comes from an incomplete description of the atmosphere in a GCM, even if we had a perfect model, the predictions

would be uncertain after week two. The atmosphere is a chaotic system, so it is highly sensitive to initial conditions (Lorenz, 1963). Earth Observing systems, like instruments on-board of Earth orbiting satellites, atmospheric radars, radiosondes or weather stations, provide initial conditions for the GCM, but these instruments have spatial and temporal sampling limitations as well as intrinsic uncertainties. The small errors in the initial state would cascade up to larger scales increasing the forecast error and decreasing the flow-dependent predictability (Lorenz, 1969).

The limitations of the predictability of a deterministic NWP forecast led the development of the Ensemble Prediction System (EPS), where a number of forecasts are produced with slightly different initial conditions determined by the observational uncertainty. Some of these systems have recently started to adopt methodologies to account for the model error as well. The combination of the different outcomes of the ensemble of forecasts can be used to produce probabilistic predictions to obtain the most likely weather evolution and the uncertainty of the forecast. The added value of an EPS in comparison to a single forecast covers different fields, the EPS can produce statistics of harmful extreme events (Hamill *et al.*, 2012; Neal *et al.*, 2013), provide the potential economic value of an EPS user action (Buizza, 2008) or produce probabilities of flood events (Cloke and Pappenberger, 2008).

The skill of the EPS depends on the reliability of the ensemble. Figure 1.1 helps to illustrate the concept of ensemble dispersion and reliability. The individual ensemble members are represented by blue lines. They must diverge rapidly enough to cover all possible outcomes that the real atmosphere could lead to, or in other words, the blue lines must diverge and produce a blue dark envelope, the forecast uncertainty. This envelope should be large enough to sample all the current climatology, the light blue envelope. The forecast uncertainty is also defined as ensemble spread or ensemble dispersion. A system that produces little dispersion for a high degree of uncertainty (small dark blue envelope for a large light blue envelope) is defined as underdispersive, whereas a system that systematically produces too much spread in comparison to its error (dark blue envelope bigger than light blue) is defined as overdispersive.

The techniques to obtain different initial conditions for the ensemble members aim to perturb short-term numerical forecasts. The perturbations are based on the the uncertainty of observations and push the model towards directions that

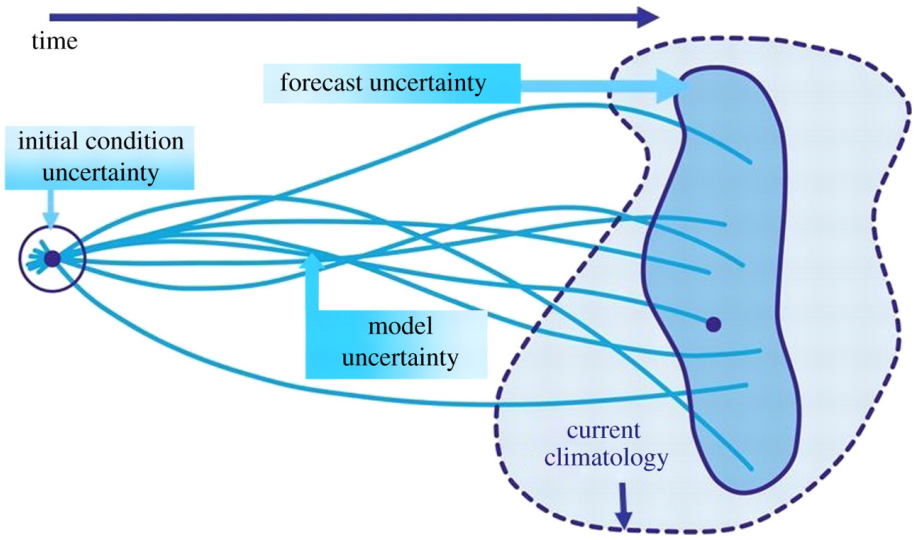

Figure 1.1: Schematic description of an Ensemble Prediction System (EPS) using initial condition and model uncertainties. Blue lines show different trajectories of the individual forecasts of the ensemble, The dark blue envelope represents the range of solutions sampled by the EPS, whereas the lighter blue envelope represent the range of possible states that the real atmosphere could encompass. *From Slingo and Palmer (2011)*

produce stronger divergence amongst members. This technique of blending the statistics from the observational uncertainty and short-range forecasts is known as Data Assimilation (Kalnay, 2002). Even with these techniques to account for observational uncertainty, the operational EPS are underdispersive and thus they can not predict the full range of the forecast uncertainty (Buizza *et al.*, 2005).

The inability of the EPS to produce enough dispersion in their predictions accentuated the need to represent model errors as well as the uncertainty of the initial state. An important source of model error, probably the most important, is the misrepresentation of the multiscale coupling between resolved and subgrid scale process, or in other words, the absence of fluctuations around the most likely effect of subgrid process on the resolved scales. There is enough evidence of the impact of unresolved convective variability in the large-scale flow (Cohen and Craig, 2006; Shutts and Palmer, 2007). On an EPS context, the subgrid effects on the large-scales should be represented by a probable value rather than the mean value (Teixeira and Reynolds, 2008a). There have been three different approaches to represent model error in an EPS:

- *Multimodel or multiparametrization ensemble:* Different members are run with

different parametrization approaches, or even with different models. Although this technique provides clear benefits (Charron *et al.*, 2010; Berner *et al.*, 2011), it does not explicitly represent model error because the ensemble of different parametrizations have not explicitly been developed to simulate the subgrid variability of the processes they simulate. The technical maintenance of such ensemble is a costly task that many atmospheric centres cannot afford.

- *Perturbed Parameter ensemble:* Key parameters in physical parametrizations, whose value is uncertain, are different for each member of the ensemble, therefore the ensemble can represent the uncertainty of the parametrization (Murphy *et al.*, 2004). However, as the previous example of the multiparametrization, this approach although useful, can not fully represent the subgrid variability as it does not address the structural uncertainty of the parametrizations nor the representation of the subgrid variability.

- *Stochastic-dynamic parametrization schemes:* The subgrid fluctuations around the statistical mean of the parametrized process are represented by a stochastic forcing (Palmer, 2001; Palmer *et al.*, 2005). This approach has been very competitive against the other two approaches in EPS from medium-range, prediction up to two weeks (Charron *et al.*, 2010; Berner *et al.*, 2011) to seasonal (Weisheimer *et al.*, 2011) and annual (Doblas-Reyes *et al.*, 2009). However their formulation is often very simple and ad-hoc (Charron *et al.*, 2010).

Stochastic-dynamic parametrization schemes aim to represent the uncertainty in physics, that is why they are defined as "stochastic physics" schemes. They also represent some of the underlying uncertainties of the dynamical core such as the diffusion schemes, which are typically utilized to filter spatial and temporal small scales to control the growth of numerical instabilities. There are two approaches in the development and implementation of stochastic physics schemes:

- The stochastic forcing terms are included in the model itself rather than in the parametrizations. These schemes have been develop mainly to increase the ensemble spread. They represent the subgrid variability by adding a pseudo-Gaussian stochastic forcing to the total physical tendencies (Buizza *et al.*, 1999), or the upscale kinetic energy from subgrid processes which is not properly represented by current parametrizations (Shutts, 2005). The impact of

these schemes on EPS has been quite positive, increasing various properties of the reliability of an EPS in addition to the ensemble spread (Berner *et al.*, 2008, 2009, 2011; Charron *et al.*, 2010; Tennant *et al.*, 2011). They have been implemented in many operational EPS around the globe.

- Stochastic features are introduced in various elements of parametrizations: Given a Probability Density Function (PDF) of a parametrized phenomenon, the stochastic scheme draws a random value from the PDF, e.g. mass flux at cloud base (Plant and Craig, 2008) or gravity wave packets (Eckermann, 2011). The stochasticity could be also added in a parameter of the closure such as an extra term in the equation for the updraught mesh fraction (Bengtsson *et al.*, 2013). Although these schemes are more physically-based, they are still under research and none of them have been made operational.

Another advantage of stochastic physics schemes is their ability to improve the mean climate through the process of noise-induced drift, where physically consistent stochastic perturbations drive the model away from its preferred chaotic attractors and make it explore different regions of the phase space, hence increasing the ensemble spread by creating a more realistic variability (Slingo and Palmer, 2011). The stochastic noise-induced drift concept is well known in simple and idealized models (Wilks, 2008; Arnold *et al.*, 2013). Theoretically it could also benefit a GCM, but given its complexity and non-linearity is not yet clear how. There have been several studies which have highlighted the impact of stochastic schemes on climate processes:

- Stochastic fluctuations to air-sea fluxes can improve the century mean sea surface temperature, atmospheric Hadley circulation and net upward water fluxes (Williams, 2012).

- A Stochastic representation of the Kinetic Energy Backscatter (SKEB) was able to simulate a better occurrence of weather regimes in the North Pacific (Jung *et al.*, 2005a), and improve tropical seasonal mean rainfall (Berner *et al.*, 2008).

- Several versions of the SKEB schemes can improve the frequency of blocking events in the Northern hemisphere (Palmer *et al.*, 2005; Tennant *et al.*, 2011; Berner *et al.*, 2008, 2012).

- A stochastic parametrization of gravity waves can produce a better stratospheric climate and representation of the Quasi Biennial Oscillation (QBO), one of the most important manifestations of wave-mean flow interactions in the atmosphere (Piani *et al.*, 2004; Lott *et al.*, 2012)

- For a given configuration of the Integrated Forecasting System (IFS) of the European Centre for Medium-range Weather Forecasting (ECMWF), a SKEB scheme is able to produce better improvements than increasing the horizontal resolution, but slightly worse improvements than an upgraded physics package (Berner *et al.*, 2012).

The potential benefits of different stochastic schemes on the representation of the atmosphere are encouraging. However, many of these schemes have been developed following a very simple formulation, they pose large uncertainties in their formalism, or given the operational demands of an EPS to produce a sizeable ensemble spread, they follow ad-hoc approaches to maximize the spread but not necessarily for well-established scientific reasons. Additionally, the scientific underpinning of some of these schemes such as the SKEBs have been challenged, arguing that a deterministic scheme might be more suitable to represent the missing process (Shutts, 2013).

The research field of stochastic parametrizations is very open and it has not been sealed off against any particular development framework or evaluation methodology yet. Stochastic physics schemes should increase the ensemble spread and improve the representation of climate processes. If we consider the example of a tropical cyclone, the different ensemble members would give different trajectories and the average of the ensemble of tracks should be the best forecast of the final track. Over longer timescales, a better representation of tropical cyclone tracks would indicate dangerous areas where the seasonal or decadal likelihood of land-hitting hurricanes is high, and also an improved tropical climate due to the improvements in the particular processes in tropical cyclones. Notwithstanding this example, there are very few studies where stochastic physics are evaluated on one particular atmospheric process. The stochastic schemes are generally evaluated under EPS skill scores, complex mathematical formulae whose values are very difficult to translate to real world weather or climate events.

There are many gaps on the understanding of how the stochastic perturbations affects the representation of important processes of the atmosphere in the tropics

and mid-latitudes. There are also many open questions whose answers have barely started to be explored. A workshop on stochastic physics held at the headquarters of the Spanish Meteorological Agency (AEMET) in Madrid, June 2013, selected the following questions as the most relevant for the development and evaluation of stochastic physics schemes:

1. Are current stochastic physics schemes perturbing the right spatial and temporal scales?

2. Do stochastic physics schemes represent physical process correctly or there are fundamental flaws?

3. Should there be stochasticity in the deterministic model outside the ensemble forecasting context?

4. Should uncertainty representations be developed alongside the physical parametrizations or added a posteriori by model error schemes?

5. How can we develop stochastic physical parametrizations in the presence of compensating model errors/heavily tuned models?

6. Which priorities are the most relevant for the development of stochastic physics?

1.1 Objectives of the thesis:

The work of this thesis aims to dispense some answers to the several dialectics where stochastic physics is often included, either as a great tool for the development of future probabilistic GCMs (Palmer, 2012), or as a mere ad-hoc artifact to bolt on EPS to account for the lack of dispersion. Some of most popular stochastic physics schemes are studied across timescales, resolutions and different atmospheric processes like blocking, convectively coupled equatorial waves or extra-tropical cyclones. In addition, some improvements for these schemes are developed and tested, as well as the development of new a scheme that perturbs the parametrization's initial state, or the transformation of a present deterministic parametrization to a stochastic physics scheme using different approaches. This thesis contains two introductory chapters, a description of the methodology employed, three chapters where the results of the different investigations carried out are reported, and final chapter

for the conclusions of the work. A brief description of each of these chapters is given below.

Chapter 2 explains the motivation for stochastic physics and describes the schemes employed in this thesis, the SKEBs schemes in sect. 2.2, the Stochastic Perturbation of Physical Tendencies (SPPT) in 2.3 and the Random Parameter approach in 2.4. A review of the main impacts of these schemes on atmospheric processes is given in sect. 2.5. Other stochastic schemes that are not evaluated in the thesis are described in 2.6, these are relevant because of their interesting formulation or results.

The main techniques to evaluate a GCM are explained in chapter 3. It contains NWP traditional verification scores (sect. 3.1), techniques to evaluate the probabilistic forecasts from EPS (sect. 3.2) and methodologies to evaluate relevant atmospheric processes, such as Mid-Latitude cyclones (3.3), Blocking (3.4), Convectively Coupled Equatorial Waves (CCEW, sect. 3.5) and the Madden-Julian Oscillation (3.6).

The methodology employed in this thesis is described in chapter 4. It contains a brief description of the GCM employed for the study, the United Kingdom Met-Office Unified Model (MetUM), with its different systems and configurations (sect. 4.1). The chapter also includes a description of the products employed to evaluate the experiments carried out in the thesis: Reanalysis datasets (sect. 4.2), the Global Precipitation Climatology Project (GPCP, sect. 4.3) and the Cloud and the Earth's Radiant Energy System (CERES, sect. 4.4).

In the first chapter of results, chapter 5, the Stochastic Kinetic Energy Backscatter version 2 (SKEB2) is evaluated across different timescales, 5 day NWP forecasts (sect. 5.1) and 20 year climate simulations (sect. 5.2). In addition, the impacts of SKEB2 are compared to deterministic improvements in the processes it aims to represent (e.g. a less diffusive dynamical core) and across two different model configurations (sect. 5.3). It is found that SKEB2 degrades deterministic forecasts but on average it creates stronger cyclones and more organized convection events, which lead to improvements in the representation of the mean tropical climate. These improvements are consistent across different model cycles and resolutions. However, there are some side effects as well, the scheme's forcing do not scale well across resolutions and it produces a spurious westward tropical wave. Some of the results reported in this chapter have been published in Sanchez *et al.* (2014).

On the chapter 6, some changes for SKEB2 are proposed and tested to minimize the errors described in chapter 5. The first one of them is to reduce the scheme's forcing on low wavenumbers (sect. 6.1), which helps the propagation of atmospheric tropical waves. The second improvement proposed is the use of a different methodology to represent the energy dissipated by the dynamical core (sect. 6.2). This new method is found to scale better across different horizontal resolutions and provides a better climatology. The last one is an ad-hoc modulation of the convective dissipation rate, which in combination with the other two changes improves the representation of mid-latitude cyclones and tropical climate. However, as the deterministic model improves and the horizontal resolution increases, the physical justification for SKEB2 becomes less clear.

Chapter 7 explores different approaches to represent the stochasticity for physical parametrizations. A scheme that perturbs the initial state for parametrizations is developed and evaluated in section 7.1, it produces a marginal impact. A SPPT scheme is developed and compared to present schemes such as SKEB2 or Random Parameters v2 (RP2, see 2.4) in section 7.2, its performance is very positive but on long timescales it has a considerable impact on the moisture and energy budgets. Different versions of the SPPT scheme are developed to remove clear sky radiation and conserve water-vapour and moist static energy, they both reduce the negative aspects of the scheme. The final section makes a deterministic mixed-phase scheme stochastic by using different approaches (sect. 7.3), it is found that the tendency perturbation approach is the most effective and the RP2 is equivalent to adding white-noise.

The main results reported in this thesis are summarized in the conclusions chapter, no. 8. Each of the questions described above as the most relevant for the development and evaluation of stochastic physics is addressed in section 8.1, where some answers or suggestions for future work are provided based on the evidence gathered in the thesis.

The Appendix B is added to briefly describe the foundations of the atmospheric prediction problem. It contains a brief history of atmospheric models (in section B.1) and the main components of a GCM: the dynamical core (sect. B.2) and the different parametrizations (sect. B.3), with an emphasis on their main uncertainties.

Chapter 2

Stochastic Physics Schemes

A physical system is classified as stochastic when its state is determined by a probability function given by the statistics of the collective behaviour of random processes, as described by Penland (2003) "macroscopic stochasticity is a manifestation of microscopic chaos". The etymology of the word comes from the Greek "stokhastikos" ($\sigma\tau\chi\alpha\sigma\tau\iota\kappa o\varsigma$), which means aim at a target.

The level of randomness in the system is subjected to the effects of the microscopic fluctuations in the macroscopic representation. Thermodynamics for instance is a fully deterministic field. However, quantities such as "heat", "temperature" or "internal energy" describe the macroscopic state of a very large ensemble of unresolved random motions of particles. Given the large population of particles within a small volume ($N \sim 10^{23}$), the Probability Density Function (PDF) of their macroscopic effects can be approximated by a deterministic delta functions with negligible error.

The definition of stochastic processes resembles the representation of many processes in an atmospheric model, like individual subgrid features such as eddies, gravity waves or cumulus clouds. These individual processes are unpredictable, but on large populations their statistical effects can be represented by some flow-dependent variables, therefore they are good candidates to be represented as stochastic processes. The introduction of stochastic elements in atmospheric models is described in the next section.

2.1 History of Stochastic physics

The idea of using stochastic physics to represent the uncertainty in atmospheric models goes back to the study of Epstein (1969), where a stochastic-dynamic forecasting technique involving Monte-Carlo approximations did improve the short-range prediction. It also provided specific information on the nature and extent of the uncertainty of the forecast. The technique, although expensive, was used in the early Ensemble Prediction Systems (EPS, Leith 1974). The concept of a stochastic climate model was suggested by Hasselman (1976), where the time-scales of the model are well separated in slow changing (climate scales) and fast changing (weather scales). The latter is represented by a first-order Markov process.

Stochasticity was applied to represent turbulent processes. Mason and Thomson (1992) developed the idea of "stochastic backscatter" to stochastically re-introduce the kinetic energy flow from subgrid-scales in a Large-Eddy Simulator (LES), and thus improve the representation of the energetic multiscale transfer (see sect. B.2.2). The scheme was later employed by Frederiksen and Davies (1997) in a two-dimensional barotropic vorticity equation of the sphere. It did outperform eddy viscosity parametrization in the representation of kinetic energy transfers across spatial scales.

Despite the continuous development of atmospheric models in terms of improvements in resolution, complexity or the quality of observations, many model uncertainties and errors were persisting. Palmer (2001) pointed to the formulation of parametrizations, as they are unable to couple scales above and below the truncation limit, e.g. the effects of convection show substantial fluctuations around the parametrized large-scale driven mean (Xu, 1992). Palmer (2001) and other studies (Williams, 2005; Palmer *et al.*, 2005) suggested the development and implementation of new "stochastic-dynamic" schemes to stochastically simulate parametrized processes. These stochastic physics schemes would improve the large-scale representation of subgrid processes and increase the capacity of the model to simulate different internal modes of variability, thus increasing the spread of possible solutions.

The concept of a "stochastic-dynamic" scheme took form in a simple parametrization to stochastically perturb the parametrized tendencies (Buizza *et al.* 1999, described in detail in section 2.3). Despite its simplicity, the scheme was quite suc-

cessful to alleviate the lack of ensemble spread of a medium range EPS. The adaptation of the stochastic backscatter scheme to a state-of-the-art EPS (Shutts 2005, described in sec 2.2) showed similar improvements. Soon these schemes were developed in other centres and were made operational (Teixeira and Reynolds, 2008a; Bowler *et al.*, 2009; Charron *et al.*, 2010; Reynolds *et al.*, 2011; Berner *et al.*, 2011). These stochastic schemes are also competitive with other representations of model error such as multimodel or perturbed-parameters on seasonal to annual ensemble forecasts (Doblas-Reyes *et al.*, 2009) and superior to those for monthly to seasonal forecast ensembles (Weisheimer *et al.*, 2011). They can also produce notable improvements in the representation of climate processes like the frequency of weather regimes (Jung *et al.*, 2005a), or tropical rainfall (Berner *et al.*, 2008). These improvements are described in detail in section 2.5.

Despite the stochastic physics schemes developed for EPS having been quite successful representing model error in probabilistic forecasts and in some cases the representation of atmospheric processes, their physical basis is poor or rather uncertain (Shutts, 2013; Hermanson *et al.*, 2009). An alternative approach is to represent a specific processes within the parametrization by a stochastic formulation. Some examples include:

- Perturbations to the Convective Available Potential Energy (CAPE) were able to generate part of the total convective variance in a quasi-equilibrium tropical circulation model (Lin and Neelin, 2000). Longer correlation timescales lead to a better matching of observed and simulated power spectra of equatorial daily mean precipitation.

- Lin and Neelin (2003) explored two physical pathways to represent small-scale variability in the convection scheme as stochastic processes: The effects of random variations of cloud-base mass flux on CAPE, and the impacts of random variations in the vertical structure of the resolved heating to represent different levels of entrainment or differences in squall line organization due to vertical shear. Both schemes help to increase the daily variability emerging from small-scale processes and enhance the power of Kelvin waves (see sect. 3.5 for a description of Kelvin and other tropical waves).

- A different approach was taken by Lin and Neelin (2002). Instead of parametriz-

ing the physics of unresolved processes using a stochastic framework, they parametrize the statistics of one of the parametrization's large-scale variables using a stochastic framework (e.g. convective heating Q). In other words: The value of Q is drawn from a constrained PDF driven by the large-scale values, rather than using an unknown PDF determined by the stochastic representation of internal processes. The scheme was able to increase the intraseasonal variability, but the impact on the climatology was modest. The distribution of convective heating was more sensitive to the large-scale effects of model dynamics than the parameters affecting the PDF of Q. Therefore they concluded that it is not very prudent to develop stochastic physics schemes outside an atmospheric model framework.

A theoretical basis for a stochastic parametrization of cumulus convection was provided by Craig and Cohen (2006), where the equilibrium fluctuations of a field of cumulus clouds under homogeneous large-scale forcing were derived statistically. The theory agrees with results from a Cloud Resolving Model (CRM) reported in Cohen and Craig (2006). A new convection scheme was developed to adapt this theory (Plant and Craig 2008, described in section 2.6.1), the scheme is based on the Kain-Fritsch scheme (Kain and Fritsch, 1990), and computes the cloud base mass fluxes from random sampling of cumulus clouds from a spectrum based on the equilibrium exponential distribution of Craig and Cohen (2006). The scheme is effective in producing equivalent distributions of convective variability to the theory using a Single Column Model (SCM) as described in Plant and Craig (2008), and in a three dimensional model of Radiative-Convective Equilibrium (Keane and Plant, 2012)

Using a similar approach, Eckermann (2011) developed a scheme where the large-scale effects of non-orographic gravity waves are represented by a packet of random waves from a prescribed spectrum. The scheme improves the representation of the stratosphere climate and variability (Lott *et al.*, 2012). It is briefly described in section 2.6.4.

Another approach consist on the division of the grid into a lattice where the phenomena could be stochastically simulated. Each site evolves following probabilistic transition rules that determine the type of cloud, like the scheme described in Khouider *et al.* (2010), section 2.6.5 of the thesis, or the region of convective

activity (Bengtsson *et al.* 2011, 2013, and sect. 2.6.3).

The stochasticity has been extended to ocean processes and the coupling to the atmosphere as well, Brankart (2013) uses random walks to represent unresolved temperature and salinity fluctuations. The ocean stochastic parametrization improves the large-scale circulation of the ocean, especially in the regions of intense mesoscale activity. Williams (2012) shows that stochastic fluctuations to air-sea fluxes of heat and moisture can improve the climate mean of surface sea temperature, atmospheric Hadley circulation and net upward water flux.

The representation of unresolved processes does not need to be explicitly included in the model, it can also be represented by "faulty" chips that generate imprecise solutions to the equations. These chips allow a different tolerance to errors. Using a spectral model, the processors with the highest concentration of faulty chip would compute the small-scales on high wavenumbers, whereas the precise processors would deal with representation of large-scales on the low wavenumbers. The system would run efficiently a very high resolution GCM with realistic but imprecise low-scale variability (Düben *et al.*, 2013).

2.2 Stochastic Kinetic Energy Backscatter (SKEB)

Current GCMs have a deficient representation of the kinetic energy spectra. The $k^{-5/3}$ slope at the higher scales is absent, as discussed in section B.2.2. The main hypothesis for this absence is the incapacity of physical parametrizations to provide a realistic representation of the impact of kinetic energy fluctuations below the truncation scale, and the excessive diffusion of dynamical cores. The impact of kinetic energy subgrid fluctuations may have important consequences for the predictability of the synoptic and large-scale flow and it is often described as one of the main causes for the lack of spread in EPS (Palmer, 2012).

The Stochastic Kinetic Energy Backscatter schemes (SKEBs) are designed to represent the upscale transfer of kinetic energy from unresolved or highly diffused scales, a physical process defined as "energy backscatter". The effects of the energy backscattered were mostly unknown across different scales at the time these schemes were developed, therefore their representation is stochastic.

In practical terms, the kinetic energy injection is introduced in the large scale

flow through perturbations in the stream-function across chosen scales. In some schemes these perturbations are modulated by a dissipation mask, an estimation of the energy dissipated by different process like implicit or explicit numerical diffusion, organized convection or sub-grid gravity wave breaking. Some SKEB schemes also include a temperature forcing modulated by an estimated available potential energy loss rate. The representation of SKEB normally follows eq. 2.1, where F_Ψ is the streamfunction forcing, D_{TOT} is the total dissipation mask, F is the forcing pattern and b_R is the backscatter ratio, which is the percentage of energy lost backscattered upscale.

$$F_\Psi = b_R \sqrt{D_{TOT}} \cdot F \tag{2.1}$$

In this section, a theoretical description of energy backscatter is given in 2.2.1. A brief description of the history of backscatter and the different schemes developed can be found in 2.2.2. The different ways to compute a dissipation mask and their impact on the schemes are discussed in 2.2.3, and the forcing pattern details in 2.2.4.

2.2.1 Energy Backscatter

One of the effects of increasing horizontal resolution is to enhance variability at large scales. Seiffert and Von Storch (2008) compared the effects of increasing horizontal resolution on the spectral ECHAM model (Roeckner *et al.*, 2003) from T31 to T63. They found that the standard deviation of the spectral coefficients at low wavenumbers were up to 1.3 times higher for the T63 version. This effect is higher for the temperature field than vorticity or divergence, the other dynamical variables of the model (see Figure 1.j,k,l of Seiffert and Von Storch 2008).

Using a simple model based on the barotropic vorticity equation, Thuburn *et al.* (2013) computed spectral energy transfers for different horizontal resolutions. The difference in the spectral energy tendency between the reference calculation and truncated versions (at 170, 85 and 42 wavelengths) shows a transfer of energy to low wavenumbers from 5 to 20, in addition to the removal of energy close to, but smaller than the truncation wavenumber. None of the combinations between the typical turbulent and diffusion schemes represented a realistic energy backscatter. The study concludes that the current SKEB schemes possibly achieve their results by repairing damage to the energy spectrum caused by truncation errors and excessive

dissipation rather than modelling a realistic backscatter.

Other way to identify the kinetic energy backscatter is through the vorticity equation. Shutts (2013) coarse-grained the vorticity equation of the IFS from T1279 to T159. The energy backscatter into low wavenumbers is evident in the vorticity flux divergence associated with the rotational wind, this term is also responsible for the energy sink close to the truncation limit. The contributions to the spectra from SKEB were also included. Its input is very noisy at low wavenumbers and then gradually decreases towards a constant input for $n > 40$.

Given that the backscatter of kinetic energy appear to be constrained to small wavenumbers, as shown by Thuburn *et al.* (2013) and Shutts (2013), there is a suggestion that a kinetic energy backscatter scheme should be deterministic at the planetary scales, leaving the current SKEB schemes to counteract the spurious model dissipation.

2.2.2 History and description of SKEB schemes

The SKEB scheme was first developed for Large Eddy Simulators (LES) by Mason and Thomson (1992), in order to represent a flow of energy from subgrid-scales to explicitly resolved scales poorly simulated by the Boundary Layer (BL) turbulence scheme. Their SKEB formulation was successfully able to excite energy at scales close to the model truncation limit. Frederiksen and Davies (1997) developed a SKEB scheme for a two-dimensional barotropic vorticity equation on the sphere, and found that it did provide a more realistic energy spectra than conventional eddy viscosity parametrizations.

After the positive results shown for a LES and an idealized model such as the barotropic vorticity equation, Shutts (2005) adopted the SKEB algorithm for a full GCM to represent a source of model error in an EPS. It successfully improved the representation of the kinetic energy power spectra and the ensemble dispersion of the EPS. Numerous centres developed their own version of SKEB. Here follows a brief description of each of them:

- *CASBS:* The Cellular Automaton Stochastic Backscatter Scheme (CASBS) was the first SKEB to be implemented in the Integrated Forecasting System (IFS, Jung *et al.* 2010a and references therein). The CASBS employed a

Cellular Automata as a forcing pattern (see 2.2.4), and a global dissipation rate built from estimates of dissipation from numerical diffusion, convection and gravity wave drag. The scheme substantially increased the kinetic energy at smaller scales, making the IFS able to simulate the mesoscale $k^{-5/3}$ tail in the Energy spectra. The scheme was also quite effective generating spread associated with model error and a small but consistent positive impact on skill.

The CASBS scheme was also found to be quite positive for the seasonal forecasting system (Berner *et al.*, 2008). It produced remarkable reductions in the error of tropical seasonal mean rainfall, extratropical blocking frequency and Sea Surface Temperature (SST) drift in the tropical Pacific. The skill of the model was notably improved with the addition of the CASBS for ensemble-mean accuracy of tropical pacific SST variability, and in terms of probabilistic seasonal predictions of temperature, precipitation and mean-sea-level pressure.

- *SSBS (a.k.a. SPBS):* The internal characteristics of the CA, like the spectral power law, were difficult to modify (Berner *et al.*, 2009). The forcing pattern was thus replaced by a spectral decomposition where each spherical harmonic forcing evolves following a first order auto-regressive process, this gives full control of the spectral characteristics of the perturbations. The scheme was renamed as Spectral Stochastic Kinetic Energy Backscatter (SSBS) or Spectral Stochastic Backscatter (SPBS) as defined in Palmer *et al.* (2009) and it is operational in present IFS cycles. The distribution of spectral power in the streamfunction forcing mimics the power-law found by Shutts and Palmer (2007) and Shutts (2008a) using a coarse-graining methodology. The temperature forcing was removed in this version. SSBS produces a more skilful probabilistic forecast to a control run without any model error scheme, a better match between the kinetic energy spectra and a better representation of flow-dependent predictability. The positive effects are stronger in the tropics, where the operational ensemble is more underdispersive.

- *MSC SKEB*: The Meteorological Service of Canada (MSC) has developed a stochastic kinetic energy backscatter algorithm for its updated EPS (Charron *et al.*, 2010). Their implementation follows Shutts (2005) but the algorithm

injects energy at smaller scales, from wavenumber 40 to 128. It also uses a more elementary temperature forcing without assuming a specific balance and a capped spherical harmonic decomposition for the forcing pattern (Li *et al.*, 2008). The scheme improves the dispersion and reliability of the ensemble, but it also introduces a low-level temperature bias compared to the former EPS.

- *NOGAPS SKEB*: The Navy Operational Global Atmospheric Prediction System (NOGAPS; Peng *et al.* 2004) has also developed their own version of SKEB following Berner *et al.* (2009). Their formulation does not employ any dissipation mask, and the decorrelation time of the forcing pattern is slightly less than the SSBS, 30 min (for further details see Reynolds *et al.* 2011). Their version of SKEB increases the spread substantially and decreases the number of extraneous outliers at all lead times. However, it does not produce any significant impact on the tropical cyclones track errors. Reynolds *et al.* (2011) also includes some preliminary results where a convective dissipation mask is employed, having small improvements in the ensemble mean of the RMSE in the mid-latitudes.

- *MOGREPS SKEB1*: The scheme was implemented in the Met Office Global and Regional EPS (MOGREPS, Bowler *et al.* 2008) and it is described in Bowler *et al.* (2009). It employed a three dimensional random pattern generator with spatial and temporal correlation lengths determined by the coarse-graining work of Shutts (2008a). It is described in the Appendix E of Evensen (2003). The numerical dissipation mask was proportional to the local kinetic energy of the flow and scaled to a global average value of $0.75 W m^{-2}$. The scheme created the $k^{-5/3}$ spectral slope as in Shutts (2005) and increased the spread of the ensemble in terms of wind variables.

- *MOGREPS SKEB2*: The MOGREPS SKEB scheme was upgraded to SKEB2 (Tennant *et al.*, 2011), and it is employed in the operational version of MOGREPS. The SKEB2's forcing pattern is based on an spherical harmonic decomposition following Berner *et al.* (2009), but forcing wavenumbers between 5 and 60, the optimal choice for the greatest growth of ensemble spread and minimal impact of forecast skill. The dissipation rate have also been updated, SKEB2 includes a convective dissipation rate and the numerical dissipation

rate is now based on energy loss estimated by the 2-D Smagorinsky-Lilly turbulence closure (Smagorinsky 1963, explained in section 6.2.1). There is a consistent improvement of nearly all aspects of the MOGREPS performance with the introduction of SKEB2. An additional feature of SKEB2 is the forcing of divergent modes through the velocity potential field, this feature increases the growth rate of spread and a produces a marginal reduction of the ensemble-mean error.

- *WRF SKEB*: There is a SKEB scheme developed for a Limited Area Model (LAM), the Air Force Weather Agency (AFWA) Joint Mesoscale Ensemble (JME, Hacker *et al.* 2011), a limited area ensemble system based on the Weather Research and Forecasting (WRF) model. Its SKEB version has no dissipation mask and a different forcing pattern based on 2-D Fourier modes instead of spherical harmonics (Berner *et al.*, 2011). Overall the scheme outperforms the ensemble system that utilizes multiple combinations of different physics schemes, especially for winds in the free atmosphere. However, for temperature at surface, the multiphysics ensemble produces better probabilistic forecasts.

As described in the previous list, the number and diversity of SKEB schemes is large and this could lead to potential confusion amongst them. Table 2.1 shows the main characteristics of each of the SKEB schemes developed until now, indicating properties of the schemes such as the type of numerical dissipation, forcing pattern, or whether it includes a temperature forcing. In terms of the results, all of the schemes have a very positive impact on the ensemble spread and reliability of the probabilistic forecasts, even those with no dissipation mask.

	Forcing Pattern	Temperature forcing	Numerical Dissipation Rate	Convective dissipation rate	Gravity wave dissipation rate
CASBS	Cellular Automata	APE backscatter included in the numerical dissipation rate.	Proportional to Biharmonic diffusion.	Rate of variation of convectively-generated kinetic energy to resolved flow.	Product between the parametrized and resolved horizontal wind.
SSPB	Spherical harmonic decomposition	No	Same as above	Same as above	Same as above
SKEB1	Pseudo-random field generator	No	Proportional to the total kinetic en. field	No	No
SKEB2	Spherical harmonic decomposition	No	Smagorinsky $2D$ diffusion scheme	Proportional to CAPE and the vertical gradient of mass flux	No
MSC's SKEB	Capped Spherical harmonic decomp.	Proportional to the product between the Forcing pattern and dissipation rate.	Modulus of the inner product of the vectorial tendencies and horizontal wind		
NOGAPS's SKEB	Spherical harmonic decomposition	No	No	Under development	No
JME's SKEB	2-D Fourier nodes	No	No	No	No

Table 2.1: Schematic description of the main characteristics of the SKEB schemes developed, see section 2.2.2

2.2.3 Estimation of the dissipation rates

There are different sources of kinetic energy dissipation in GCMs. The SKEB schemes include dissipative processes associated to systematic energy losses, such as the diffusion schemes and upscale error growth though physical parametrizations. The connection of the stochastic perturbation to an instantaneous dissipation rate makes the SKEB scheme physically motivated, as it takes flow dependence into account. However, there are considerable uncertainties in observational estimates of atmospheric energy dissipation rates (Shutts, 2005), so their formulation across different schemes varies. There are SKEBs schemes that do not include any representation of the dissipation (SKEB schemes at NOGAPS or JMS), or the dissipative processes are represented with a universal formula (SKEB at MSC). The different sources of numerical and physical dissipation are:

- *Numerical dissipation rate:* A substantial part of the kinetic energy is lost through explicit diffusion or implicitly by the Semi-Lagrangian interpolation error, like the smoothing introduced by the interpolation to the departure point (see section B.2.1). Part of this energy is not dissipated and a small faction of it is backscattered towards the resolved scales. Shutts (2005) provides an extensive summary of the different studies carried out to make estimations of total energy dissipated by the explicit diffusion schemes and the Semi-Lagrangian (SL) advection scheme. SSBS assumes that the implicit numerical dissipation of the SL scheme is equivalent to biharmonic diffusion (McCalpin, 1988), whereas SKEB2 uses the Smagorinsky non-linear diffusion scheme (Smagorinsky, 1963), a two dimensional simplification of the more general Smagorinsky-Lilly turbulent closure equation (Mason and Thomson, 1992). The differences between these two approaches are investigated in chapter 6.

- *Convective Energy dissipation:* A small faction of the convective turbulence, created around convective events, might be cascaded upwards towards the larger scales, and thus become an important part of the observed $k^{-5/3}$ spectrum (Lilly, 1983; Peng *et al.*, 2014). Current parametrizations can not reproduce this release of turbulent kinetic energy (see sect. B.3.5 for a description of the uncertainties of convective parametrizations).

- *Mountain drag dissipation:* Most of the 3D turbulence schemes dissipate as

heat the turbulent effects from eddies on the vicinity of mountains, whereas a small fraction could be available to backscatter into the near-grid scale of GCMs. Subgrid scale Gravity Wave breaking also generates 3D turbulence that might contribute to an upscale transfer of energy.

- *Dissipation of Available Potential Energy (APE):* Horizontal explicit and implicit diffusion also affects the temperature field. Shutts (2005) adds the rate of loss of APE, crudely estimated using a quasi-geostrophic definition, to the numerical dissipation rate. Charron *et al.* (2010) explicitly adds a temperature forcing proportional to the dissipation rate and the forcing pattern.

Convective dissipation rate is larger over areas of deep convection around the Inter-Tropical Convergence Zone (ITCZ) or storm tracks in the mid latitudes (see Fig. 2.1.b), the numerical dissipation rate is more focused on the mid-latitudes where advection is larger (Fig. 2.1.c), and the Mountain dissipation rate sits over high mountain ranges such as Himalayas, Rockies or Andes (Fig. 2.1.d). The combination of these three sources includes many aspects of weather processes, its geographical reach is ample so as it is vertically (Fig. 2.1.a), as numerical dissipation tends to occur more in the low troposphere whereas deep convection dissipation is more focus on mid to high troposphere.

The total dissipation field produced is generally quite noisy, it contains large gradients which may be a threat for the stability of the model, therefore smoothing is required. The total dissipation rate of the CASB and SPBS for the IFS is smoothed retaining only the first 30 wavenumbers (Berner *et al.*, 2009), SKEB2 in MOGREPS uses a $1-2-1$ spatial filter 5 times (Tennant *et al.*, 2011). This operation ensures that scheme also spreads the uncertainty to adjacent grid-points (e.g. a cyclone being slightly misplaced in reality).

The simplified backscatter schemes with no dissipation mask are an improvement over the models without SKEB, but their performance is not as good as the scheme with the full dissipation rate. Berner *et al.* (2009) compared an operational ensemble of IFS without SSBS, a similar system with SSBS with a constant dissipation rate, SSBS with no convective dissipation and SSBS with the full dissipation rates. The experiments with simplified dissipation mask outperforms the control, in the extratropics the experiment with the most complex dissipation produces small improvements but significant at 95%, with the scheme without convective dissipation

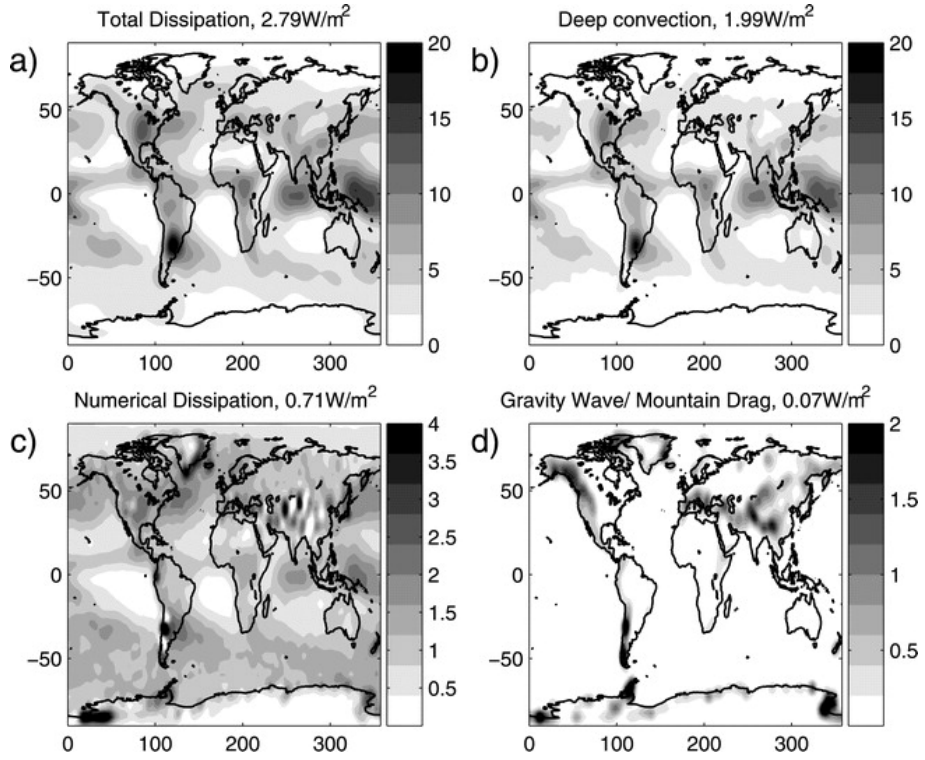

Figure 2.1: Vertically integrated annual-mean total dissipation rate per unit area (W/m^2). (a) Total dissipation which is the sum of (b) Deep convection, (c) numerical dissipation, and (d) gravity/mountain wave drag. *From Berner* et al. *(2008)*

performing better than the one with a constant dissipation rate. In the tropics there is a similar pattern for low level winds skill scores, but with little differences between the constant dissipation SPBS and the one without convective dissipation (see Figure 14 of Berner *et al.* 2009). A similar conclusion can be drawn from Reynolds *et al.* (2011), the addition of a provisional convective dissipation mask reduces the low level winds mean Root Mean Error Square (RMSE) of the ensemble mean in the mid-latitudes.

2.2.4 Forcing patterns

There have been two different methods to compute the stochastic field, so-called forcing pattern, each has its own characteristics in terms of spatial, temporal and scale interaction. The following methods have been used:

- Cellular Automata (CA): The use of CA to describe subgrid variability was suggested by Palmer (2001), as it has spatial scales equivalent to mesoscale cellular convection. The CA is based on the Conway's Game of life (Gardner,

1983). It divides the globe in cells smaller than the gridboxes, each cell could be alive or dead depending on probabilistic rules which could depend on the state of neighbour cells. The memory of the cells makes the scheme attractive to represent convective clusters driven by model flow phenomena such as fronts or Mesoscale Convective Systems (MCS). A comprehensive description of the CA details is given in Shutts *et al.* (2008b), the CA scheme has also been used inside a convection parametrization closure (see section 2.6.3).

- Spherical harmonic pattern with time evolution by autoregressive process: The pattern is based on a truncated expansion in terms of spherical harmonics. The spectral coefficients evolve following an autoregressive process (eq. 2.2), where m and n are the zonal wavenumber and spherical harmonic degree respectively, α is a damped autoregressive parameter with $\alpha\,\epsilon[0,1]$, and it is determined by the model timestep and autocorrelation timescale τ (eq. 2.3), g_n is the power law function to provide the desired power spectrum and $r^t_{m,n}$ is a random number for a given time and spectral point. The power law have been deduced by Shutts and Palmer (2007) and is proportional to $n^{-1.27}$. The SPBS draws the random number from a Gaussian white-noise process with mean zero and variance σ_z (Berner *et al.*, 2009) and the SKEB2 draws it from an uniform distribution with range $\epsilon[-0.5, 0.5]$ (Tennant *et al.*, 2011). The first term of eq. 2.2 correspond to the memory of previous timesteps and the second is the noise term regulated by power law, a full description and derivation of the perturbations' energy given by the spectral forcing pattern can be found in Berner *et al.* (2009) and its appendix.

$$f^{t+1}_{m,n} = [1 - \alpha(n)]f^t_{m,n} + \sqrt{\alpha(n)}g(n)r^t_{m,n} \tag{2.2}$$

$$\alpha(n) = 1 - exp\,(-\Delta t/\tau(n)) \tag{2.3}$$

2.3 Stochastic Perturbation of Parametrized Tendencies

The errors associated to physical parametrizations (reported in sect. B.3) can be represented by a simple scheme, the Stochastic Perturbation of Parametrized ten-

dencies (SPPT, Buizza *et al.* 1999). The scheme perturbs the prognostic variables horizontal wind u, meridional wind v, temperature T and humidity q; adding a multiplicative noise to the total tendency due to parametrized physical processes. The multiplicative noise factor, or forcing pattern, was on its original implementation a random number r sampled from a uniform distribution between 0.5 and 1.5 and uniformly distributed over horizontal boxes. For a given time a new random number was drawn for each box. Buizza *et al.* (1999) evaluated the performance of the scheme on the IFS, finding that the scheme improves the spread of the ensemble and its performance, in particularly for the probabilistic prediction of precipitation. It also explored the sensitivity of the results for different box sizes and temporal autocorrelations of the random number.

The original forcing pattern for SPPT was somewhat unphysical, as there were large discontinuities when r changed in space, from the edge of one box to the neighbouring box, and in time. A new version of the scheme has been developed (described in Palmer *et al.* 2009), its main upgrades are:

- A different forcing pattern, based on a spherical harmonic decomposition that varies smoothly in space and time (see section 2.2.4) but the SPPT has a quasi-Gaussian power law.

- The same r is applied to the different perturbed variables, making the scheme univariate. Perturbations generate a more physically-based balance amongst model variables.

- For reasons of numerical stability and physical realism, perturbations are tapered to zero in the lowermost atmosphere, as tendencies in the Boundary Layer are quite large and variable, from 300 to 1300 metres. A similar tapering is used in the stratosphere.

- Supersaturation correction, which is used when changes to T and q result in condensation.

The new SPPT scheme substantially outperforms the one evaluated by Buizza *et al.* (1999). It generates larger ensemble spread and increases the skill scores, more notably in the tropics. Palmer *et al.* (2009) also show that the best performing SPPT configuration is based on a multiscale forcing pattern built from two independent

patterns, the first one represent fast evolving synoptic scale errors and the second slower evolving planetary scale errors. On present cycles of the IFS a 3^{rd} pattern has been added (Shutts G, 2010, *personal communication*). Table 2.2 shows the properties of the different patterns.

Pattern	Standard deviation of r	Temporal autcorrelation	Spatial Autocorrelation
Fast	0.52	6 hours	500 km
Medium	0.18	3 days	1000 km
Slow	0.06	30 days	2000 km

Table 2.2: Properties of patterns that form the multiscale pattern of SPPT

A different version of the SPPT has been built for the Canadian EPS (Charron *et al.*, 2010). It forces planetary and synoptic scales to maximize the impact. The decorrelation time τ is 3 hours, but it is reported in Charron *et al.* (2010) that the results are insensitive to a choice of τ between 3 and 12 hours. The standard deviation of the forcing pattern is close to 0.23 and the value or r lies in the range $[0.5, 1.5]$. When the SPPT scheme is removed from the EPS, the dispersion of high level dynamical variables is degraded for most of the forecast lead times (see Figure 9 Charron *et al.* 2010). It also has a negative impact on the Continuous Ranked Probability Score (CRPS, Jolliffe and Stephenson 2003), which is significantly degraded during the second week of the forecast. The scheme helps to improve many of the EPS skill scores after the first week of the forecast, although the performance of the system on these timescales is relatively low.

The concept of adding stochastic perturbations proportional to the tendencies has been extended to a high resolution LAM, where many processes still need to be parametrized (like shallow convection or turbulent eddies). Bouttier *et al.* (2012) describe the impact of a SPPT algorithm in the Application of Research to Operations at Mesoscale (AROME, Seity *et al.* 2011) of $2.5km$ horizontal resolution. The implementation follows Palmer *et al.* (2009), including the vertical tapering and the univariate approach. The scheme enhances ensemble spread and the probabilistic skills scores, although it slightly degrades the forecast skill of the ensemble members. The perturbations are physically consistent, as observed in one case study, where SPPT enlarges the area where fog spread is non-zero. The spread, although positive

within the forecast time-range, is believed to become overdispersive if it was run for ranges much longer than one day. The study of Bouttier *et al.* (2012) also includes some SPPT tuning experiments where the following results are reported:

- If space or time correlations are reduced, the impact of SPPT is weaker, suggesting that the AROME model error is not limited to small scales only.
- Doubling the time correlation produces a bigger impact on the ensemble skill scores than doubling the space correlation.
- There is almost a linear relationship between the impact of SSPT in ensemble spread and the SPPT standard deviation parameter.

2.4 Random Parameters

The Random Parameter scheme (RP, Bowler *et al.* 2008) aims to account for the uncertainty associated to the empirical parameters in bulk-formula parametrizations, and simulate the non-deterministic processes not explicitly accounted for by the different parametrizations. Each of the random parameters evolves in time following an autoregressive process given by eq. 2.4, bounded to a given range, P_{max} and P_{min}, estimated by the experts.

$$P_t = \mu + r(P_{t-1} - \mu) + \epsilon \tag{2.4}$$

Where P_t is the parameter value at time t, μ is the default value of the parameter in the deterministic parametrization, r is the autocorrelation of P and ϵ is the random term sampled for a uniform distribution from the range $\pm(P_{max} - P_{min})/3$ to ensure that parameters values are evenly distributed.

The parameters included in the scheme are:

- *Entertainment rate*, in convection parametrization. Scales the rate of mixing between environmental air and the convective plume (further explained in sect. B.3.5).
- *Time-scale for destruction of CAPE*, convection parametrization. The convective instability, represented by CAPE is removed by mass fluxes, the time-scale of CAPE destruction determines how intense convective mass flux needs to be

in order to eliminate the CAPE in a given time (also futher explained in sect. B.3.5).

- *Flux-profile parameter*, boundary layer parametrization. The stability-dependence of turbulent mixing coefficients are dependent on this parameter, so it can enhance or reduce mixing in the stable boundary layer.

- *Asymptotic neutral mixing length*, boundary layer. It sets the magnitude of the mixing lengths and hence the turbulent mixing coefficients.

- *Gravity wave constant*, gravity wave drag parametrization. Defines the magnitude of the wind's parametrized tendencies from the scheme.

- *Critical Froude Number*, gravity wave drag. It controls the proportion of the drag attributed to flow blocking and gravity wave drag respectively. The larger the critical Froude number, the larger the proportion attributed to flow blocking.

- *Critical relative humidity*, condensation parametrization. The threshold for relative humidity for cloud formation, it has a vertical dependence, the perturbation scales to the model level without changing the vertical structure of the parameter.

- *Ice-fall speed*, condensation parametrization. It modifies the fall speed of ice nuclei.

The Random Parameter 2 scheme (RP2) also includes:

- *Charnock Parameter*, boundary layer. It influences the surface wind stress over the oceans.

The scheme does not have a negative impact on the skill of deterministic forecasts, indicating that it provides a veridical source of spread based on model uncertainties rather than stochastic noise (Bowler *et al.*, 2008). It produces a significant increase of the ensemble spread in the tropics, and a small increase elsewhere (see Fig. 4. of Tennant *et al.* 2011).

As explained in the introduction (chapter 1), the perturbation of parameters does not sample all the structural uncertainty nor the subgrid variability of the

physical parametrizations. However, it has been used extensively to sample the model uncertainty in climate change predictions in projects such as Quantifying Uncertainty in Model Predictions (QUMP, Murphy *et al.* 2004). A comparison between the random parameter and perturbed tendencies approaches is carried out in chapter 7 of the thesis.

2.5 Evaluation of Stochastic Physics outside EPS scores

The evaluation of SKEB, SPPT and RP2 schemes has been focused on the impacts on the spread and error skills of EPS from short-range to seasonal scales. There have been very few studies focused on the effects of the scheme on key processes in the mid-latitude and tropical weather. Some of these studies have looked at the effects of SKEB schemes on weather phenomena and mean systematic error of boreal winters, nearly all of them use the schemes CASB or SPBS developed for the IFS for the ECMWF model. The reported results for different processes are

- The CASBS scheme reduces the westerly wind bias over the boreal winter in the North Pacific against ERA40, improving the geopotential height at 500mb, described hereafter as Z_{500} (Fig. 1 of Jung *et al.* 2005a). They use the IFS cycle CY26R3 (see Jung *et al.* 2005a, 2010b for a review of the model systematic biases and differences between different cycles). On the newer model cycle CY32R1 the SPBS produces smaller improvements for the whole Northern hemisphere (Fig. 3.b of Berner *et al.* 2012), although on the newer cycle the mean bias of Z_{500} has been notably reduced. The impact of different versions of the SPPT on IFS cycle CY35R1 is small but significant, with the 2-scales pattern being the most effective (Fig. 14 of Palmer *et al.* 2009).

- Improvements in the frequency distribution of weather regimes over the North Pacific. Jung (2005b) evaluated the capacity of CASBS and SPPT schemes to force the model towards a more realistic distribution of weather regimes in the North Pacific wintertime. The IFS on cycle CY26R3 and $T_L \sim 95$ considerably misrepresented the frequency of geopotential at 500mb clusters to those found in ERA-40, CASBS was able to produce a more realistic frequency of regimes,

in particular the underestimation of blocking events, driven by a reduction of the westerly wind bias. The scheme was also found to be more effective than SSPT. On a general sense, the CASBS projection of sub-synoptic vorticity onto the large scale lead to a reduction on the overpopulation (under-population) of the more (less) stable regimes.

- There is a consistent improvement in the representation of blocking events on the NH wintertime over the North Pacific for CASBS at cycle CY26R3 (Fig. 17 Palmer *et al.* 2005) and cycle CY29R2 (see Figure 5 of Berner *et al.* 2008). SPBS shows equivalent results for cycle CY32R1 (Fig. 4 Berner *et al.* 2012). On the other hand, improvements on the underestimated blocking frequency over Central Europe are small and not significant at the 95% level for all the cases. Outside the IFS, Tennant *et al.* (2011) reported an improvement on blocking frequency over the East Siberia and North Pacific.

- Seasonal mean tropical precipitation improvements in DJF. CASBS on the IFS cycle CY29R2 produces a reduction of the wet bias over the ITCZ over the oceanic regions, Maritime continent and Northern Australia (see Figure 4 of Berner *et al.* 2008), similar improvements are found for SPBS on cycle CY32R1 (Figure 5 of Berner *et al.* 2012). SPBS in combination to SPPT, produces a weaker impact in a newer model cycle CY35R1 (see Fig. 15 of Palmer *et al.* 2009).

- SPBS on cycle CY32R1 is able to reduce the power of westward propagating tropical waves, more in agreement with observations (Figure 6.f of Berner *et al.* 2012)

- SPBS is able to produce a peak in the tropical divergence wind spectra around 60 days for the period 1990-2005, but it vanishes and gives way to a red spectrum when averaged over more years (Figure 7e,f of Berner *et al.* 2012).

- SPBS on CY32R1 increases the Z_{500} systematic error over the Southern Hemisphere, likely caused by a overactive tropical divergence and synoptic activity over the Southern Ocean (Berner *et al.*, 2012)

The SPBS improvements found by Berner *et al.* (2012) are generally higher than those observed when the horizontal resolution increases from T96 to T511 (210km

to 40km), but lower than the upgrades produced by an upgrade in the physics from cycle CY32R1 to cycle CY36R1.

The combination of SPBS and SPPT is able to excite the right growing modes of Tropical cyclone error. Lang *et al.* (2012) investigated the impact of different EPS perturbation methods on the ensemble spread of track and intensity of Tropical cyclones (TC). The Ensemble of Data Assimilations (EDA) exhibits the larger spread for TC tracks, with SPBS, SPPT and Singular Vectors having a similar impact. In terms of central pressure, SPBS leads after day 3 but the differences to the other experiments are barely significant. SPBS perturbations have a larger amplitude in the outer region of the TC whereas SPPT perturbations are more focused on the TC core and upper levels.

2.6 Other stochastic schemes

There are other important stochastic schemes, their formulation and results are noteworthy and therefore they deserve a brief description, which is given below. Unfortunately, these schemes are not employed in the investigations carried out for the present thesis, their development is costly and may well take longer than the time allocated for the research presented in the thesis.

2.6.1 Plant-Craig scheme

The deterministic deep convective parametrizations are driven by the large-scale effects on the mean mass flux. In the Plant and Craig (2008) scheme, the ensemble effects of a collection of independent clouds is derived from the statistical mechanics theory of Craig and Cohen (2006) and CRM results from Cohen and Craig (2006). In this theory each individual cloud has a mass flux given by the Probability Density Function (PDF) in eq. 2.5, angle brackets denote an ensemble average. The stochastic sampling of the mass flux PDF produces a statistical distribution that represents the convective fluctuations around the mean.

$$p(m)dm = \frac{1}{< m >} e^{-m/<m>} dm \tag{2.5}$$

Plant and Craig (2008) carried out Single-column tests to investigate the functioning of the Plant-Craig (PC) scheme. They found that the mean profiles of

temperature and humidity produced by the scheme at various grid-lengths were comparable to those obtained from CRM simulations.

The scheme was tested further in a three dimensions Radiative-Convective Equilibrium model by Keane and Plant (2012). The scheme was shown to produce the correct scaling for the variability of rainfall for several gridlengths, outperforming conventional deterministic deep convection schemes.

Groenemeijer and Craig (2011) implemented the scheme in the Consortium for Small-scale Modelling model (COSMO, Schattler 2011). They found that the PC scheme produces a substantial increase of the ensemble spread. In weather patterns that are weakly (strongly) forced, the relative impact of the stochastic scheme is high (low). However, they do not show any verification of the probabilistic forecasts yielded by the scheme.

2.6.2 Stochastic Convection

A relatively straightforward stochastic convection parametrization was designed for the NOGAPS-EPS. It is based on a stochastic perturbation proportional to the convection tendencies (Teixeira and Reynolds, 2008a; Reynolds *et al.*, 2008). The stochastic convection scheme perturbs the tendencies of horizontal winds and temperature, these perturbations are proportional to the tendencies produced by the moist convection parametrization times a random number. Although it might look similar to the SPPT (sect. 2.3), the stochastic convection has no spatial or temporal auto-correlation and humidity is not directly perturbed. The scheme's perturbations alone produce a significant increase of the ensemble spread at 500-hPa geopotential height, 250- and 850-hPa winds, and 850hPa temperature. At the early stage of the forecast, the ensemble perturbations are larger in the tropics. The spread on the extra-tropics is less prominent than in the tropics and is mainly driven by perturbations in the synoptic scales coming from the tropics.

Snyder *et al.* (2010) studied the effect of the stochastic scheme on the tropical cyclone genesis and evolution of four different cases. The scheme increases the ensemble spread of tropical cyclone tracks and the fraction of ensemble members predicting genesis. On the other hand it also increases the rate of false alarms. However, the increase in correct genesis predictions is greater than the increase in false alarms

2.6.3 Cellular Automata for convection

Deterministic convective parametrizations represent the relationship of the ensemble average of subgrid convection and the instantaneous state of the atmosphere in a vertical grid-box column. This approach leaves behind any parametrization for the horizontal transport of heat, moisture or momentum. Huang (1990) showed that mass transport due to gravity waves that propagate in the horizontal can trigger new convective cells.

A Cellular Automata (CA, see 2.2.4 for a brief description) was embedded in the Aire Limitée Adaptation/Application de la Recherche à l'Opérationnel (ALARO, Benard *et al.* 2010 and references therein) model convective parametrization to represent horizontal communication and memory (Bengtsson *et al.*, 2011, 2013). The CA acts on a subgrid mesh to represent individual convective events. With the right choice of CA rules, it can be expected to mimic the effect of gravity wave propagation on convective organization, forming clusters of spatial scales larger than the truncation scale of the model.

A probabilistic set of rules is explored by Bengtsson *et al.* (2013), where birth and survival of cells are based on probabilities given by the number of neighbours with an active state, it generates CA patterns less artificial that look more like clusters of convection (see Figure 2 of Bengtsson *et al.* 2013). These probabilistic rules can be adjusted to match observations.

The scheme was first tested in a simplified version for the Shallow-Water model in Bengtsson *et al.* (2011). They made the mass source term Q proportional to the CA fraction σ, which describes the fraction of active cells within the model gridbox. The proposed CA scheme produces a kinetic energy backscatter from the smallest to the largest atmospheric scales. It also slows the phase speed of Kelvin waves in regions where convergence is large, in agreement with observations. Subgrid cells can organize and propagate against the mean flow, a feature not seen in conventional deep convection parametrizations.

On more recent research, the CA has been coupled to the ALARO convection scheme (Bengtsson *et al.*, 2013). It is coupled as an extra term for the equation for the updraught mesh fraction σ_u (equation 4 of Bengtsson *et al.* 2013), the additional term acts as an independent source of information on potential convective activity,

in addition to the large-scale moisture convergence. They show an exceptionally well defined squall line episode in France in August 2010 enhanced by the CA implementation, precipitation intensity and convective organization are in a better agreement with radar observations. The impact on large-scale skill scores is small, but on situations with strong convective activity the CA provides a significant source of spread where random errors are thought to occur. It also improves the ratio of hit rate and false alarms, in particular with the CA version that uses probabilistic rules.

2.6.4 Explicit Stochastic Parametrization of Non-orographic Gravity Wave Drag

Temperature and velocity perturbations of gravity waves come from a wide spectrum of waves with very similar spectral shapes (Kim *et al.*, 2003). Usually non-orographic gravity wave parametrizations are based on a quasi-invariant global background spectrum of many waves from indistinct tropospheric sources, although new schemes are based on physical models of gravity wave generation from specific sources such as deep convection (e.g. Charron and Manzini 2002).

A different approach has been taken by Eckermann (2011), where the parametrization randomly picks up wave-packets from the spectrum. The typical GCM timesteps of 1-60 min and horizontal resolutions of 10-1000km do not appear to be either large or long enough respectively for the full wave-ensemble to emerge within a gridbox. At any given GCM timestep, subgrid-scale wave fluxes would vary as individual grid boxes contain different subsets of sporadic sources and wave-field members of the broader wave-ensemble. The scheme has no spatio-temporal correlation of wave properties between adjacent grid boxes or model timesteps. The scheme reduces the computational time by an order of magnitude to the deterministic parametrization, it exhibits explicit gravity wave intermitency, which is parametrized by a bulk formula in the deterministic scheme. The stochastic scheme variability can realistically increase the ensemble spread and produce a better representation of the stratospheric climate.

An additional improvement of the scheme is described in Lott *et al.* (2012). The new scheme permits to launch a few monochromatic waves at each model timestep and distributes their tendencies over several timesteps. Therefore at a given time, there are different waves acting together. Such scheme improves the representation

of large-scale equatorial waves and therefore of the Quasi-Biennial-Oscillation (QBO, Baldwin 2001).

2.6.5 Stochastic Multicloud Model (SMCM)

Purely deterministic parametrizations of convection are found to be inadequate for the representation of the highly intermittent and organized tropical convection (Palmer, 2001). Some of these deficiencies are associated to the progressive deepening of tropical convection on multiple scales (see Xavier 2012 and 3.6). The Stochastic Multi-Cloud Model (SMCM, Khouider *et al.* 2010) aims to improve the representation of cloud transitions in the tropics.

The SMCM is based on a lattice model where each lattice site is occupied by a cloud of a certain type (congestus, deep or stratiform) or clear sky. The evolution of the cloud is driven by random transitions between the four states. These transitions are governed by intuitive probability transition rates depending on the large-scale variables: CAPE and a proxy for middle troposphere dryness. One example of these transitions is given by Frenkel *et al.* (2012), if there is a large instability (CAPE >> 400 J/kg) and the column's middle troposphere is moist, the transition rate from clear-sky site to a congestus is high, similarly a deep convective site will turn into a stratiform site with high probability.

Frenkel *et al.* (2012) coupled the SMCM scheme to a simple two-layer atmospheric model capable of capturing the main characteristics of tropical convection and associated wave features, and compared the scheme to a usual deterministic convective parametrizations. SMCM yields highly intermittent solutions that capture the progressive deepening of tropical convection on multiple scales. A medium and coarse resolution model can produce more variability with the stochastic parametrization than with the deterministic scheme over a medium resolution grid. Peters *et al.* (2013) compares results from SMCM to tropical observations and concludes that the scheme is able to reproduce the dependencies of convective variability, but fractions of deep convective and stratiform clouds compare better to observations when using the convection proxy related to convergence (i.e. vertical velocity at 500 hPa) rather than those related to stability (CAPE).

Chapter 3

Evaluation of atmospheric models

Unlike other physical systems like ocean tides or the dynamics of the solar system, it is extremely complicated to make accurate predictions of the evolution of the atmosphere. It is a chaotic system (Lorenz, 1963) and the tool we employ to develop these predictions, a General Circulation Model of the atmosphere (GCM), has many and large uncertainties, as it is described in Appendix B. Despite these difficulties, it is possible to predict certain aspects of the weather and climate, a condition that is defined as "predictability".

Over the last decades, numerical weather prediction (NWP) has focus on the predictability of synoptic weather events such as the position of cyclones and fronts in the mid-latitudes, and Mesoscale Convective Systems (MCS) in the tropics. The quality of NWP forecasts is evaluated using numerical scores that compare modelled fields with observations. These fields could be geopotential height at 500mb (Z_{500} hereafter), which gives a picture of how well simulated are the main geostrophic structures in the mid-latitudes, or low (high) level winds that indicate converge (divergence) driven by tropical convection. Problems with the sampling of observations and/or their spatial interpolation have lead to the development of "analysis", a model realization driven by assimilated observations. Analyses provide a better spatial coverage and allow simple operational evaluation of the modelled field.

The mean state of climate simulations is evaluated using similar scores against observational products and reanalyses (see examples of these in the methodology chapter, no. 4). These techniques reveal the existence of model biases, e.g biases on radiative fields denote problems with the representation of clouds, or biases in Z_{500} indicate problems on the representation of mid-latitude synoptic variability

(e.g. cyclones are too weak, lack of blocking events). These classical scores are still widely used and are defined as "traditional verification techniques". They are described in section 3.1 of the thesis.

Equivalent techniques were developed to evaluate Ensemble Prediction Systems (EPS). These verify the distribution of the ensemble as the sample from a probability distribution and compare its shape to the probabilities of past events, so-called "reliability of the ensemble". Some of these techniques are described in section 3.2.

The emergence of the seamless model also brought additional evaluation techniques. Despite the upscale transfer of error from small scales to the large-scales (Lorenz, 1969; Tribbia and Baumhefner, 2004), the predictability of certain timescales is associated to processes occurring on these timescales (Hoskins, 2013). Therefore in combination with the traditional verification techniques, there is a new branch of model evaluation focused on "process-based techniques", where the main characteristic of atmospheric processes are evaluated to indicate its degree of predictability. Figure 3.1 illustrates the seamless prediction idea where different timescales are associated to particular processes such as cyclones, blocks, Madden-Julian Oscillation (MJO), North Atlantic Oscillation (NAO), El Niño Southern Oscillation (ENSO), Quasi-Biennial Oscillation (QBO), Pacific Decadal Oscillation (PDO) or Atlantic Multi-decadal Oscillation (AMO).

The idea of the seamless prediction is also defined as "predictability chain". Rodwell *et al.* (2013) provides an illustrative example of the predictability chain. They link poor forecasts for the onset of blocking events for spring over Europe to a poor representation of shot-lived but intense MCS over the US plains. Hoskins (2013) provides another example where large anomalies in the winter atmospheric vortex gives some predictive power over the troposphere for the following month, but such anomalies are forced by anomalous tropospheric wave flow, which is driven by short scales.

For the evaluation of the stochastic physics schemes carried out in this thesis, we make use of the Transpose Atmospheric Model Inter-comparison Project (Transpose-AMIP, Williams *et al.* 2013) approach, where the impacts of these schemes on model biases are investigated through an extensive set of diagnostics for 5 day forecasts and centennial simulations. The set of diagnostics includes traditional verification and

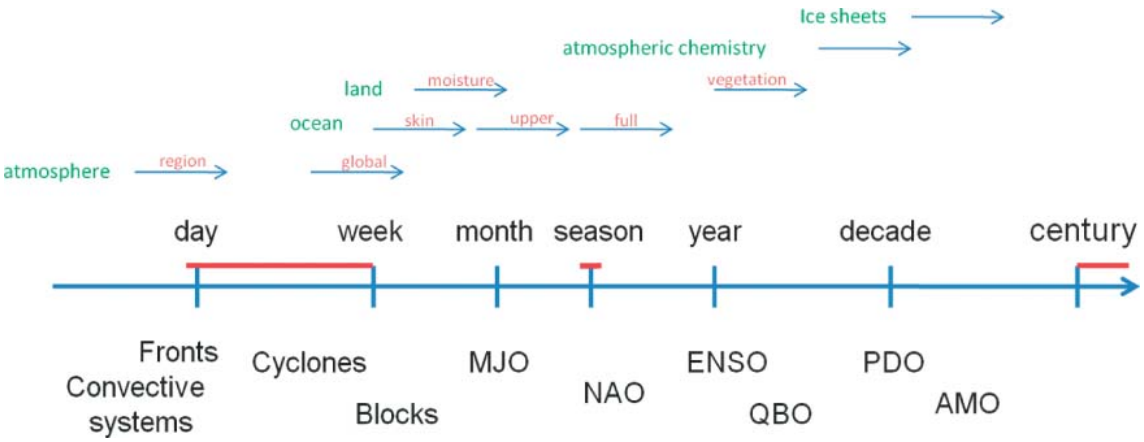

Figure 3.1: Schematic representation of the seamless weather-climate prediction problem. Timescales are shown along the horizontal axis in the middle. The focus for prediction in the recent decades is indicated by red lines on the axis. Some phenomena and their time-scales are shown at the bottom (acronyms are given in the text). Components of the Earth system that need to be represented in the GCM or ESM are included on the top of the figure. *From Hoskins (2013).*

process based techniques. We focus on important processes for the short-to-mid range predictability which are also major drivers of the climate variability. Such processes are mid-latitude cyclones (sect. 3.3), blocking (3.4), Convective Coupled Equatorial Waves (CCEW, sect. 3.5) and the MJO (3.6).

3.1 Traditional verification techniques

Verification indexes and scores are a valuable tool to provide a general picture of the system. However, they might hide compensating errors or show good scores for the wrong reasons. An improvement on the verification skill could be related to a poorly simulated process, e.g. a good simulation of the Outgoing Longwave Radiation (OLR) at the Top Of the Atmosphere (TOA) may be caused by an unrealistic brightness of high level clouds. There are a wide variety of these indexes and few sources to understand their advantages and disadvantages, Jolliffe and Stephenson (2003) and Wilks (2006) are reference textbooks with an ample description of these weather verification techniques. In order to evaluate the impact of the stochastic physics schemes within the scope of this thesis, the indexes and techniques we employed are:

- *Mean Bias:* It is the raw difference between a modelled and observed field, as described by eq. 3.1, where F is the forecast and O the observations for N forecasts. It is a useful metric to show systematic errors of the model. Although it is an illustrative quantity, when horizontally or vertically averaged it could hide compensating errors.

$$Bias = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i) \tag{3.1}$$

- *Root Mean Square Error (RMSE):* It is one of the most widely used forecast scores. It is the square root of the mean squared difference between observed and forecasted field (eq. 3.2).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \tag{3.2}$$

- *Anomaly Correlation Coefficient (ACC):* It is the correlation between forecasted and observed anomalies to the climatology (eq. 3.3). It measures how well the forecast captures the magnitude of anomalies from climatology. As a correlation, it ignores biases in the forecast anomalies. It has been found empirically that $ACC = 60\%$ corresponds to the range up to which there is synoptic skill for the largest scale weather patterns. $ACC = 50\%$ corresponds to forecast for which the error is the same as for a forecast based on a climatological average (Jolliffe and Stephenson, 2003).

$$ACC = \frac{\sum_{i=1}^{N}(F_i - C_i)(O_i - C_i)}{\sqrt{\sum_{i=1}^{N}(F_i - C_i)^2 \sum_{i=1}^{N}(O_i - C_i)^2}} \times 100 \tag{3.3}$$

3.1.1 Weaknesses of traditional verification scores

Traditional verification scores are very useful to measure the skill of the model for large-scale and synoptic scales over the Z_{500} field. However, they are not ideal to verify other fields in high resolution forecasts as many of the features resolved are very detailed and highly unpredictable. These techniques can not distinguish between a "near miss" and much poorer forecast, as point-by-point comparison do not account for the intrinsic spatial correlation between forecasts and observations.

A common problem of these scores is the "double penalty". When there is a slightly offset on the position of a weather event between a forecasts and the observations, the forecast is thus penalized because it fails to predict the event where it occurred and also to predict where it did not occur. RMSE shows better skill for a field with a wide but weaker feature that enclose the observed one than for a forecast that simulates the right shape and intensity of the feature but slightly misplaced (for a visual example of the double error penalty see Figure 6.3 of Jolliffe and Stephenson 2003).

In order to provide other methods for spatial verification, there have been a development of a few novel techniques to reward the realism of the spatial structures of high resolution forecast fields such as rainfall. Many of these techniques are reviewed by Casati *et al.* (2008). One of these is the Method for Object-based Diagnostic and Evaluation (MODE, Davis *et al.* 2006a,b), which identifies objects in the precipitation field. Mittermaier and Bullock (2013) have extended this technique to cloud fields.

3.2 Ensemble Prediction System verification techniques

The additional members of the ensemble add an extra degree of complexity to the verification of the EPS. Besides having a good skill, the ensemble members must have the right spread (how well it represents the uncertainty of the forecast), see Fig. 1.1.

The ensemble members can be characterized as finite random samples of a probabilistic distribution. The collective behaviour and statistical properties of the ensemble are compared against a long term record to estimate the ensemble's "reliability", which quantifies the degree to which the forecast probabilities are consistent with the relative frequency of the observed outcomes. In an unbiased ensemble, the observations are statistically indistinguishable from ensemble members.

There are a few methodologies to measure the skill and the realism of the ensemble's Probability Density Function (PDF), with different weaknesses and strengths. Generally speaking there is no clear superiority of one validation methodology over the others, as some of them are very complex or can hide compensating errors. It

is hard to link the EPS scores to real weather events. The evaluation of EPS is still an open field or research where new techniques emerge such as the verification of the ensemble of mid-latitude cyclone tracks (Froude *et al.* 2007a, see section 3.3.3). In this thesis the traditional EPS metrics employed to evaluate the capacity of stochastic physics to represent the model uncertainty are:

- *Deterministic verification of the ensemble mean:* The Ensemble Mean (EM) is the average of all ensemble members over a given point (eq. 3.4 where the subindex j denotes the ensemble member of M realizations), EM is the "deterministic" realization of the EPS. Although it ignores the statistical properties of the ensemble, it is a good estimate to compare to deterministic models using traditional deterministic scores such as RMSE or ACC. One disadvantage of the EM is that it filters out small scales when averaging, so the EM holds no information on important sub-synoptic processes.

$$EM = \sum_{j=1}^{M} F_j \tag{3.4}$$

- *Ensemble spread-error:* This is the most familiar metric in the evaluation of EPS, it associates the predicted uncertainty and the accuracy of the forecast. The spread or dispersion of the ensemble is the standard deviation of the members of the ensemble (eq. 3.5, where F_j are individual ensemble members and $\overline{F}$ is the EM). A necessary condition for the ensemble variability is that the RMSE of the EM should be identical the ensemble spread. If the RMSE is higher, then the system is "under-dispersive", differences amongst ensemble members are not high enough and the truth is outside of the ensemble's PDF. If the RMSE is lower than the ensemble spread then the system is "over-dispersive", differences amongst different members are bigger than the differences in the observations. The ensemble spread and error are normally presented on regional averages (Northern hemisphere from 20N-90N for example).

$$Spread = \sqrt{\frac{1}{M} \sum_{j=1}^{M} \left(F_j - \overline{F}\right)^2} \tag{3.5}$$

- *Reliability diagram:* It is a plot of the observed frequency against forecast probability of a particular event (e.g. a location warmer than X degrees), it

measures the consistency between observed frequencies and predicted probabilities. The range of forecast probabilities is divided into K bins, then probabilities of the observed event are computed for each bin of the forecasts. The perfect reliability is indicated by the 1:1 line. If the observed-forecast curve lies below the 1:1 line, it indicates the system is over-forecasting the event, it occurs more frequently in forecasts than in reality. If it is above the 1:1 line the system is under-forecasting, the EPS underestimates the frequency of the event.

3.3 Mid-Latitude cyclones

One of the main processes in the mid latitudes is the formation and development of extra-tropical synoptic cyclones. These systems control winds, cloudiness and precipitation. At longer time-scales they transport heat, momentum and water vapour from the equator to the poles. It is hard to quantify the skill of the models in the representation of mid-latitude cyclones using conventional methods such as RMSE or ACC of Z_{500} (Froude *et al.*, 2007a). A more useful tool to diagnose cyclones is the TRACK algorithm (sect. 3.3.1), as it gives direct information about the location and characteristics of individual cyclones, which in large numbers could provide a useful body of statistics to diagnose model deficiencies.

3.3.1 TRACK algorithm

TRACK identifies features in the spatial scales of interest from temporally sliced fields. Then it tracks these, linking the features together to form trajectories. Although in this thesis TRACK is employed to diagnose extra-tropical cyclones, its versatility enables it to track other atmospheric cyclones such as Tropical cyclones/ African easterly waves (Hodges *et al.*, 2003; Bengtsson *et al.*, 2004), clouds (Hodges and Thorncroft, 1997; Hodges, 1998) and ocean eddies (Hodges, 1999b). A general description of the the different sections of the algorithm is given below; for further details the reader is referred to Hodges (1994, 1995, 1999a).

For extratropical cyclones a minimum frequency of 6 hourly data is required to accurately compute the storm tracks. Cyclones can be identified from different variables like relative vorticity at the 850-hPa level (ξ_{850}) or Mean Sea Level Pressure

(MSLP). Fast moving MSLP features can be masked by a strong background flow, so they are difficult to identify on the earlier stages of their development. On the contrary ξ_{850} focuses on smaller scale features and it is less dependent on the background flow, therefore features can be identified at an earlier stage of development. Figure 3.2 shows an illustration of the differences between MSLP and ξ_{850}. There are features (denoted by dots) from weak vorticity features that are not apparent in the MSLP field.

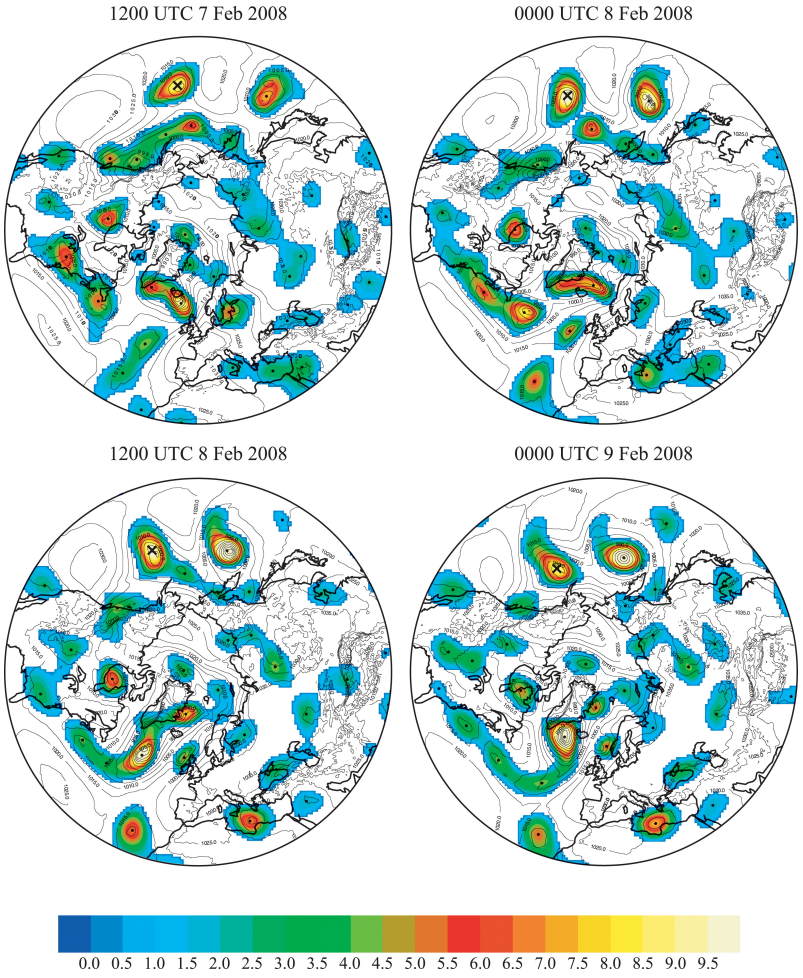

Figure 3.2: European Centre for Medium-range Weather Forecast (ECMWF) analysis MSLP field (contours) and T42 filtered vorticity field at 850hPa (coloured shading). For 12Z 7/2/2008, 0Z 8/2/2008, 12Z 8/2/2008 and 0Z 9/2/2008. Dots denote the vorticity features identified using Hodges (1995, 1999a) methodology. Vorticity units are $10^5 s^{-1}$ (relative to background field removal). *From Froude (2010).*

Another disadvantage of tracking over MSLP fields is that it is an extrapolated field and may therefore be sensitive to how the extrapolation is performed (e.g. the representation of the orography in the model). See Hoskins and Hodges (2002) for

a discussion of the advantages and disadvantages of different fields for mid-latitude cyclone tracking. In this thesis we employ ξ_{850} to track extra-tropical storms

The tracking algorithm identifies and tracks cyclones following four main stages: data filtering, object identification, feature point identification and tracking:

- *The filtering stage:* In order to isolate the synoptic scales of the field, a spectral filter is applied to the raw fields. The filter is first applied to remove wave numbers less than or equal to 5 (as in Hoskins and Hodges 2002). For ξ_{850} fields the lower-wave filtering has very little impact and is not very sensitive to the choice for the lowermost wavenumber (Anderson *et al.*, 2003). The spatial filter is also applied to truncated wavenumbers above 42 (T42) to remove mesoscale noise. This is necessary for fields such as ξ_{850} to avoid identifying very small scale structures. The Hoskins filter (Sardeshmukh and Hoskins, 1984) is also applied to reduce the Gibbs phenomena. It acts like a quadratic hyperdiffusive ∇^4 smoothing and its amplitude is reduced to 10% of its value on the smallest retained scale.

- *Object-identification:* The algorithm identifies objects as regions encircling extrema in the filtered field. Cyclonic objects are identified as positive (negative) anomalies in the Northern Hemisphere (Southern Hemisphere). The object points are first classified as objects if they are above a user defined threshold T, usually set to 1.0 $10^{-5}s^{-1}$ for ξ_{850}. The object points are then agglomerated into distinct sets (see Hodges 1994 for details on the labelling algorithm). All objects below some user-defined size are filtered. The segmentation process reduces the object information to a small subset of the original data.

- *Feature point identification:* This section identifies the extrema within each of the objects. The maxima and minima of each object would be adequate for tracking in a high resolution grid, but for a reduced grid at T42, this will significantly limit the potential smoothness of the tracks computed in the final stage of the method. Interpolation algorithms described in Dierckx (1981, 1984) are used to interpolate or smooth the data so that extrema can be located within the grid boxes. The maximization algorithm of Goldfarb (1969) is then used to identify the feature points. For further details of the feature point identification see Hodges (1995).

- *Tracking stage:* This stage determines the correspondence for the feature points amongst different time-slides, linking them and creating the tracks. The method has been adapted from the algorithm of Sethi and Jain (1987) based on Salari and Sethi (1990), but it has been extended to work on a spherical domain (Hodges, 1995). In the initial procedure the feature points are linked to each other using a nearest neighbour distance between points in subsequent time-frames. The displacement between feature points on tracks is within a upper-bound displacement d_{max}, which regionally varies between 6 and 3 geodesic degrees. The cost function is minimized by a modified version of the Greedy exchange algorithm of Sethi and Jain (1987) and Salari and Sethi (1990). This is an iterative optimization method, which proceeds forwards and backwards in time, swapping those pairs of points on tracks that give the greatest increase in track smoothness (see Hodges 1999a, for details of the algorithm).

Once the tracks have been computed, they are filtered so that only those tracks that last at least 2 days, travel further than 1000 km and have a majority of their lifecycle in 20N - 90N or 20S - 90S are retained for further analysis. This stage is necessary so that only mobile, meteorologically significant, extratropical cyclones are considered in the statistical analysis.

3.3.2 TRACK statistics

A wide variety of spatial statistics can be obtained from the ensemble of feature tracks. They can be computed differently, from the basic counting and averaging of each grid-point, which are subjected to biases (Taylor, 1986), to more some sophisticated weighting functions based on analytic functions (Murray and Simmonds, 1991). In this thesis the approach described in Hodges (1996) is used, where statistics are computed directly on the sphere using spherical kernel estimators with local kernel functions. These statistics are:

- *Track density:* Number of feature points, using one point for each track, closest to the estimation point.

- *Genesis density:* Density of systems in their source areas. It is computed from the starting points of the tracks and excluding any tracks starting at the first

timestep

- *Lysis density:* Density of systems on their dying areas. It is computed from the end point of the tracks excluding all the tracks ending in the last timeframe.

- *Intensity:* The maximum vorticity of the object on the filtered field. It does not include regions where the feature track density is below a suppression threshold equal to 1.0 $10^6 km^2$ per month. It is computed from the average of attributes

- *Velocity:* Average distance between two adjacent points in the track divided by the time-lapse of frames, 6 hours.

- *Growth rate*: Rate of intensity change of a cyclone per time-lapse of frames.

In order to provide intelligible numbers over a global domain, the raw density is scaled to one month and 5°spherical cap ($\sim 10^6\ km^2$).

3.3.3 Storm matching technique

The TRACK algorithm has been developed into an object-oriented verification technique, in order to evaluate how individual forecasted cyclones deviate from their analyzed counterparts with increasing lead time. It uses a matching technique, which is a systematic method to determine which track of the forecast corresponds to which track of the analysis. There is a detailed description of the development of the matching algorithm for TRACK in Bengtsson *et al.* (2004) and Froude *et al.* (2007a). In this thesis only storms are considered, although the algorithm could be equally employed to anti-cyclones.

A simulated storm is said to match an analyzed storm if the two tracks satisfy certain predefined spatial and temporal criteria. There is no restriction on the difference in intensity, as this could make the conditions too severe and thus reduce the population of the matched storms. The conditions to match an analyzed and forecasted storms are:

1. At least a percentage, equal to T, of their points are overlapping in time, as shown in eq. 3.6, where n_A and n_B denote the total number of points in the analysis and forecast tracks respectively, and n_M denotes the number of points in the analysis track that overlap in time with the forecast track.

$$100 \cdot |\frac{2n_M}{n_A + n_B}| \geq T \tag{3.6}$$

2. The geodesic separation distance d between the first k points of the forecast track and the corresponding points in the analysis track is less than S.

A sensitivity study for the values of k, T and S is present in Froude *et al.* (2007a), where six different sets are explored. When $k = 1$ the algorithm produces more forecast storms match, but many of these might be incorrect. As the number of k points increases, then less storms match, making it difficult to produce a reasonable body of statistics. When $k = 4$ it produces a good balance between right matching and number of statistics. Although the number of forecast tracks that match the analysis varies depending on the matching criteria, differences in the diagnostics produced from these matched storms are marginal (see Figure 4 of Froude *et al.* 2007a). The default set of parameters for the matching technique is: $k = 4$, $T = 60\%$ and $S = 4^{\circ}$.

A schematic representation of the spatial matching when $k = 4$ is given in Figure 3.3. The tracks from the model that match the analyzed storm are A, B and C because their first four points are less than S from the analysis and share more than 60% of their points with the analysis. Storm D does not match because at its origin the separation is greater than S.

There is an additional constraint in the method. Only storms whose genesis occurs in the first 3 days are considered. Results from Bengtsson *et al.* (2005) show that there is little skill in the prediction of tracks beyond the third day.

The matching technique has been employed to study the short-range predictability of storms for the EPS included in the The Observing System Research and Predictability Experiment (THORPEX) Interactive Grand Global Ensemble (TIGGE) for NH (Froude, 2010), and SH (Froude, 2011). EPS with stochastic model error schemes (such as described in chapter 2) have more impact in increasing the ensemble spread of cyclone intensity. Nevertheless all models have common problems, the cyclones are too slow and the ensemble spread of the storm intensity is underdispersive.

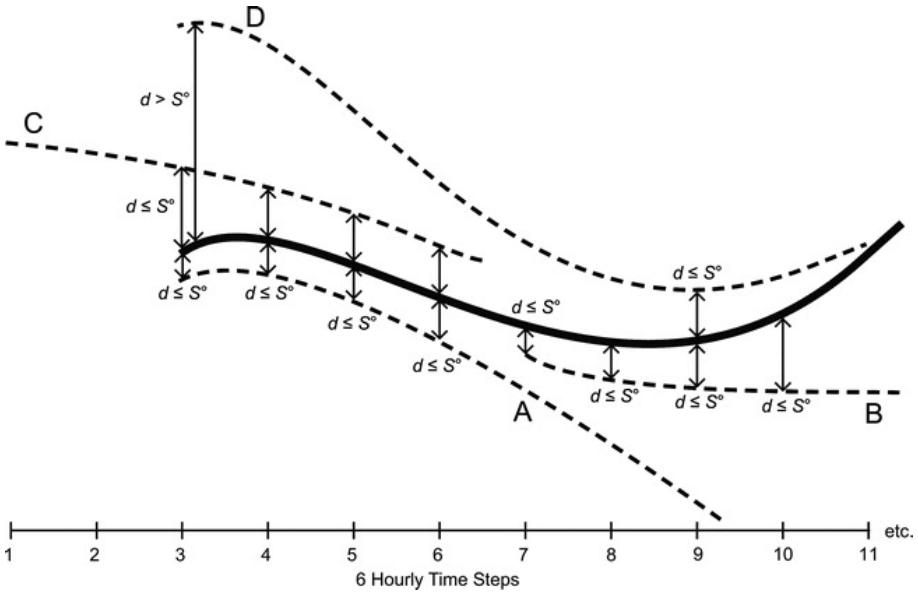

Figure 3.3: Schematic representation of the spatial matching with $k = 4$. Solid curve represents the storm track from analysis and dashed from model. A,B,C tracks do match but D does not. *From Froude* et al. *(2007a).*

3.4 Blocking

Blocking is one of the most important and complex weather process in the mid-latitudes. It is associated to extreme weather, in particular to heat waves or extreme cold spells in Europe (Hoskins and Sardeshmukh, 1987). A Blocking event is a quasi-stationary area of high pressure that "blocks" the usual mobile weather systems of the middle latitudes, thus the usual westerly winds are replaced by easterlies over the region "blocked". Their lifetime spans from few days to ~ 10 days. Blocking is an example of an emergent phenomenon implicitly driven by dynamical and physical processes in the model. Previous studies have identified upscale feedbacks that help to maintain the large scale blocking structures (Shutts, 1986; Lau, 1988b).

GCMs generally underestimate the frequency of blocking, specially those with low horizontal resolutions and thus low Eddy Kinetic Energy (EKE), which translates in weaker storm tracks. Studies with very high horizontal resolution models show a significant increase in blocking frequency (Matsueda 2009). It is therefore a useful phenomena to test the ability of the model representing the atmosphere in the mid-latitudes.

Blocking has been one of the first targets in the evaluation of stochastic physics schemes, as these schemes attempt to simulate low-scale fluctuations and push the

model away from their preferred attractors into other less explored regimes like blocking. Numerous studies with different stochastic schemes and models show that stochastic physics schemes increase the frequency of occurrence of otherwise too weakly populated blocking regimes in the Pacific basin and have a marginal impact on the Atlantic basin of the Northern hemisphere (Palmer *et al.*, 2005; Tennant *et al.*, 2011; Berner *et al.*, 2008, 2012).

3.4.1 Blocking indexes

A number of different indices have been proposed to diagnose atmospheric blocking. Most of them are based on the detection of a meridional gradient of a blocking sensitive variable such as geopotential at 500hPa (Z_{500}) or potential temperature (θ) on a potential vorticity (PV) surface.

One of the most employed indexes is the one described in Tibaldi and Molteni (1990). It diagnoses blocking by the presence of a meridionally oriented dipole of high Z_{500} on the north side and low Z_{500} of the south side of a variable latitude ϕ_0 (a mathematical description is given in the appendix of Scaife *et al.*, 2010).

Another index based in a similar idea is the one proposed by Pelly and Hoskins (2003) but it uses θ over an isosurface of 2 PV (the definition of dynamical tropopause). A large-scale blocking episode occurs at a particular longitude ϕ if the dipole based quantity B (eq. 3.7) is positive for at least a longitude section of 15°, and 4 consecutive days. Figure 3.4 provides a schematic representation for the B blocking index.

$$B = \frac{2}{\Delta\phi}\int_{\phi_0}^{\phi_0+\Delta\phi/2} \theta d\phi - \frac{2}{\Delta\phi}\int_{\phi_0-\Delta\phi/2}^{\phi_0} \theta d\phi \tag{3.7}$$

Tibaldi and Molteni (1990) index produces fewer blocking events than Pelly and Hoskins (2003), but the variable latitude allows the longitudinal profile of the blocking frequency to resemble the PV computed profile (see Fig. 11 of Pelly and Hoskins 2003). Therefore Tibaldi and Molteni (1990) index is adequate to represent a qualitative comparison of the blocking frequency of a GCM.

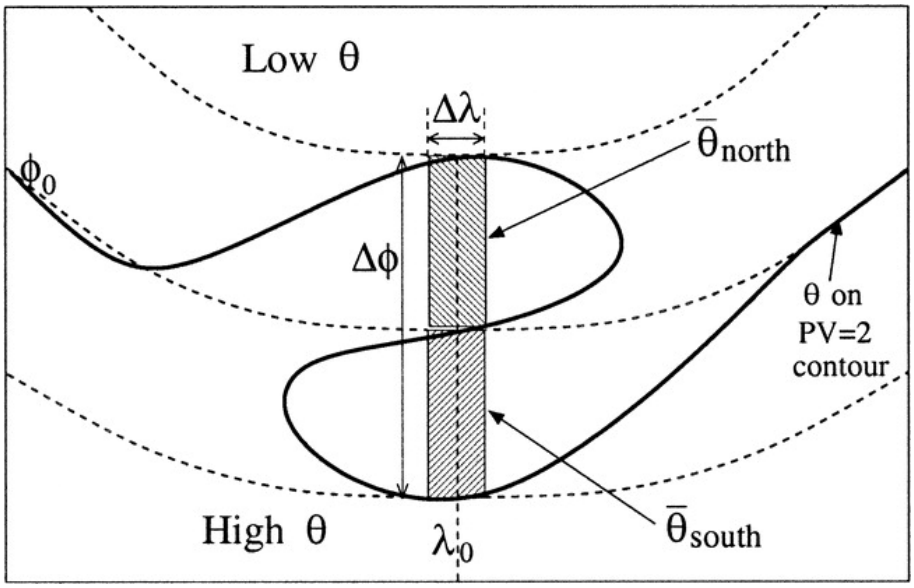

Figure 3.4: Schematic representation of the calculation of the PV-theta blocking index B at a given longitude ϕ_0 Thick line denotes theta on a PV = 2 contour *From Pelly and Hoskins (2003).*

3.5 Convectively Coupled Equatorial Waves

A substantial fraction of the tropical large-scale variability in convection on timescales less than 30 days is organized by waves that move eastward or westward along the equator. These waves are know as Convectively Coupled Equatorial Waves (CCEW) and they are fundamental components to understand the interaction of convection and other physical processes with the dynamics. Kiladis *et al.* (2009) review the different types of CCEW, their main characteristics, theoretical derivation and observational evidence.

The basic structure and dispersive characteristics of CCEWs are described by the wave solutions of the shallow water equations. These govern the vertically independent motions of a single thin layer of incompressible homogeneous density fluid on a rotating sphere. The derivation of the phase relationship and wave solutions of these set of equations was first made by Matsuno (1966). It is assumed that the coriolis parameter f is linearly proportional to distance from the equator ($f = \beta y$). Solutions to the wave equation that decay away from the equator must satisfy the dispersion relationship given in equation 3.8, where ω is the frequency, k is the wavenumber, g is acceleration due to gravitiy, h_e is the depth of the undisturbed layer of fluid and n is a positive integer equal to 0, 1 or 2.

$$\frac{\sqrt{gh_e}}{\beta}\left(\frac{\omega^2}{gh_e} - k^2 - \frac{k}{\omega}\beta\right) = 2n + 1 \tag{3.8}$$

There are three wave solutions for ω in the cubic equation 3.8, Eastward Intertio Gravity waves (EIG), Westward Inertio Gravity wave (WIG), and Equatorial Rossby waves (ER). The solution for $n = 0$ is the Mixed Rossby-Gravity wave (MRG). An additional solution is the latitudinally symmetric Kelvin waves, represented by $n = -1$ in the dispersion equation (eq. 3.8). All these solutions are presented in Figure 3.5 and their horizontal structures in Figure 3.6. The Inertio-Gravity and Kelvin waves tend to be more divergent in character whereas the MRG and ER waves are more rotational.

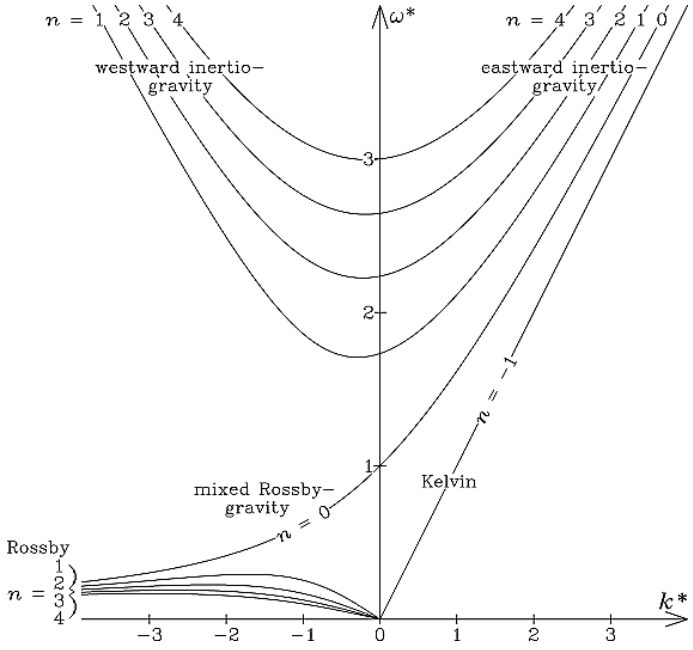

Figure 3.5: Dispersion curves for the propagating wave solutions to the shallow water equation, represented by functions of nondimensional frequency $\omega* = \omega/\left(\beta\sqrt{gh_e}\right)^{1/2}$ and wavenumber $k* = k\left(\sqrt{gh_e}/\beta\right)^{1/2}$. Westward propagating waves are denoted by negative wavenumber and eastward with positive wavenumber. *From Kiladis* et al. *(2009)*

Despite the crude approximations made by Matsuno (1966), considering a dry atmosphere with no vertical structure, these waves are present in the atmosphere coupled with convection. Wheeler and Kiladis (1999) calculated the global space-time spectrum of tropical cloudiness and found prominent spectral peaks along the dispersion curves shown in Figure 3.5 (see section 3.5.2 for details on the analysis of CCEW). Similar results where obtained using a more direct measure of satellite-derived rainfall (Cho *et al.*, 2004).

A description of the main characteristics of the Kelvin waves is given in the next section, as it is one of the main CCEW and is poorly represented by models. It is followed by some remarks about the simulation of CCEW in current models, and the techniques most applied in such evaluations.

3.5.1 Kelvin waves

Many of the short-period synoptic scale active convective cells moving eastwards in the MJO are actually Kelvin waves (see 3.6 for a description of the MJO). The dynamical structure of the Kelvin waves is quite similar to the idealized solution shown in Figure 3.6.f (see Figure 7 and 8 of Kiladis *et al.* 2009). Winds at $850hPa$ are easterly to the east of the negative anomaly of T_b (brightness temperature, a proxy for convection) and westerly to the west, causing low-level convergence and convection.

Composites of Kelvin waves show that humidity increases in the lower troposphere since 2 days prior to the lowest T_b and then it propagates vertically through the full tropospheric column. After the wave has passed by the observing point, the lower troposphere rapidly dries and cools while the upper troposphere remains moist and warm. This is also associated with the cycle of cloudiness: shallow convection progresses to deep convection on day 0, and it ends with an upper tropospheric stratiform cloud after the convective signal has passed (see Straub and Kiladis 2002 for a more detailed example).

There is strong evidence suggesting that Kelvin waves are initiated indirectly by wave forcing from the extra-tropics (Yang *et al.*, 2007; Straub and Kiladis, 2003). The signal from the convective ascent east of the westerly seems to stretch from the SH extra-tropics toward the equatorial region (see Figure 4 of Yang *et al.* 2007). The extratropical forcing of the equatorial Kelvin waves has also a theoretical underpinning. Hoskins *et al.* (2000) using both Gill-type and primitive equation models showed that a moving higher-latitude vorticity forcing in the winter hemisphere is surprisingly effective in triggering the equatorial Kelvin wave.

3.5.2 Simulation of CCEW

The MJO attracts most of the attention in the study of the interactions between convection and dynamics in the tropics. Nevertheless, there have been some studies

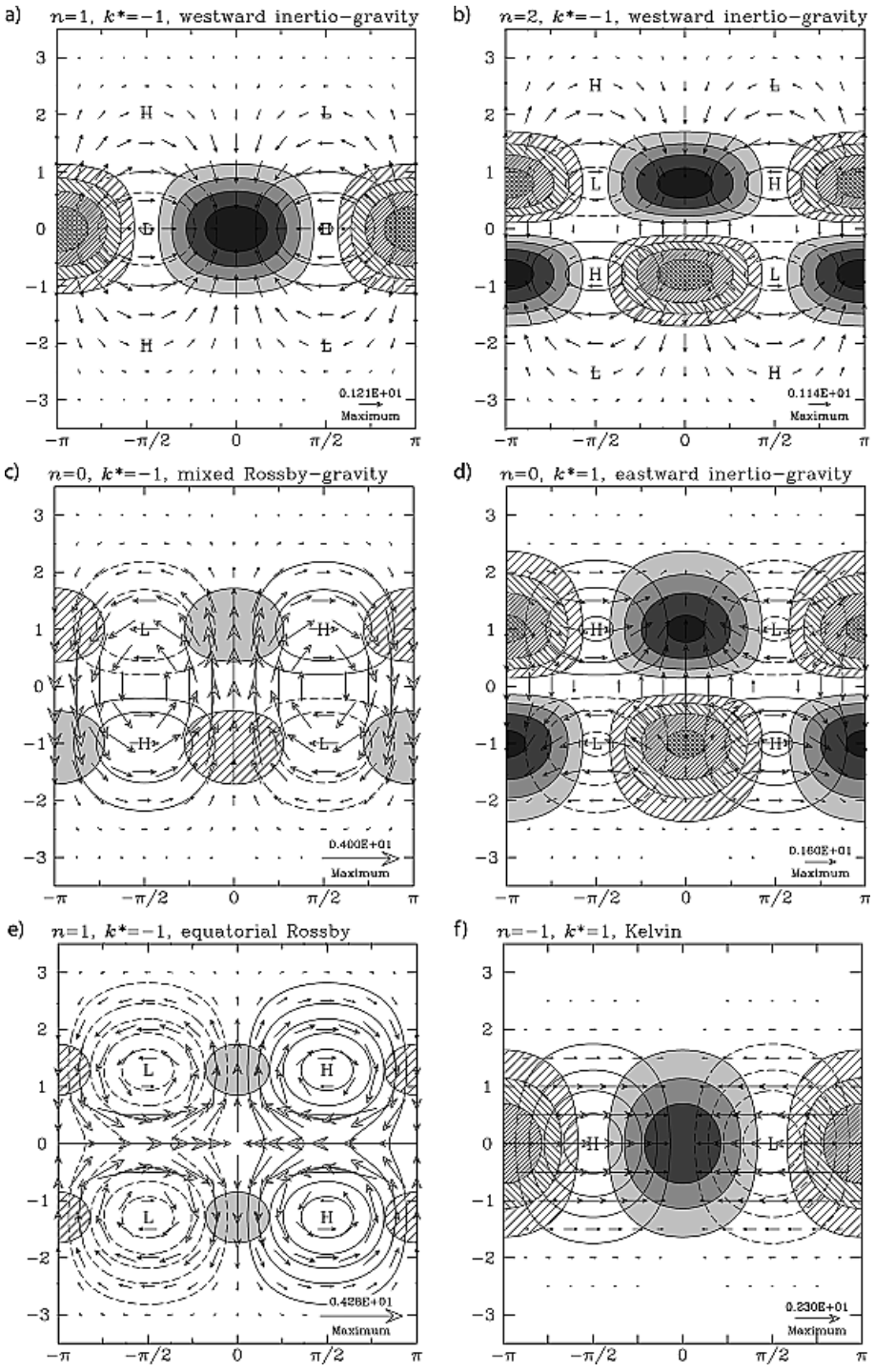

Figure 3.6: Horizontal structures of a subset of wave solutions, each is shown for a non-dimensional wavenumber $k = \pm 1$. All scales and fields are nondimensionalized. Hatching is for divergence and shading for convergence, with a 0.6 unit interval between successive levels. Unshaded contours are geopotential, with a contour interval of 0.5 units. Negative contours are dashed. *From Kiladis* et al. *(2009)*

investigating the characteristics of CCEW in GCMs. In a similar way to the MJO, the convection-dynamics coupling is not good enough and many models have a weak variance for all waves, except for the eastward intertio-gravity waves (Lin *et al.*, 2006).

The Equatorial wave activity of the Hadley centre Global Atmospheric Model (HadGAM1, Martin *et al.* 2006), the predecessor of MetUM at climate scales (see section 4.1), is reported by Yang *et al.* (2009). Their main conclusion is that the model performs well for equatorial waves coupled with off-equatorial convection, but it performs poorly for waves coupled with equatorial convection. HadGAM1 fails to simulate the near-surface anomalous equatorial zonal wind, together with intensified equatorial convection in phase with westerly winds of Kelvin and Rossby $n = 1$ waves.

3.5.3 Techniques to identify and analyse CCEW

One of the most popular diagnostics to evaluate the representation of CCEW in a GCM is the background-removed power spectra of a given field. Such field must have a strong dynamics-convection coupling such as winds at $850hPa$ or $250hPa$, Outgoing Longwave Radiation (OLR) or precipitation. The process to obtain this diagnostic is described in Wheeler and Kiladis (1999). A brief explanation is given in the following paragraph.

A double Fourier Transform (in space and time) is applied to the latitudinal average of the tropical belt of the field chosen. As seen in Figure 3.6, equatorial waves are either symmetric or antisymmetric about the equator. In order to have this in mind, we decompose the spectra into a symmetric part where the latitude averaging is given by $F(\phi) = [F(\phi) + F(-\phi)]/2$ and an asymmetric part (where $F(\phi) = [F(\phi) - F(-\phi)]/2$). The wavenumber-frequency field is obscured by the red noise present, so a background power spectra is built by averaging the power of the symmetric and anti-symmetric component and smoothing it 10 times with a 1-2-1 filter in frequency and wavenumber. Dividing the individual raw power-spectra by the background power-spectra yields a plot where the wave signals are clear (Figure 3.7) and situated above the dispersion curves given by eq. 3.8.

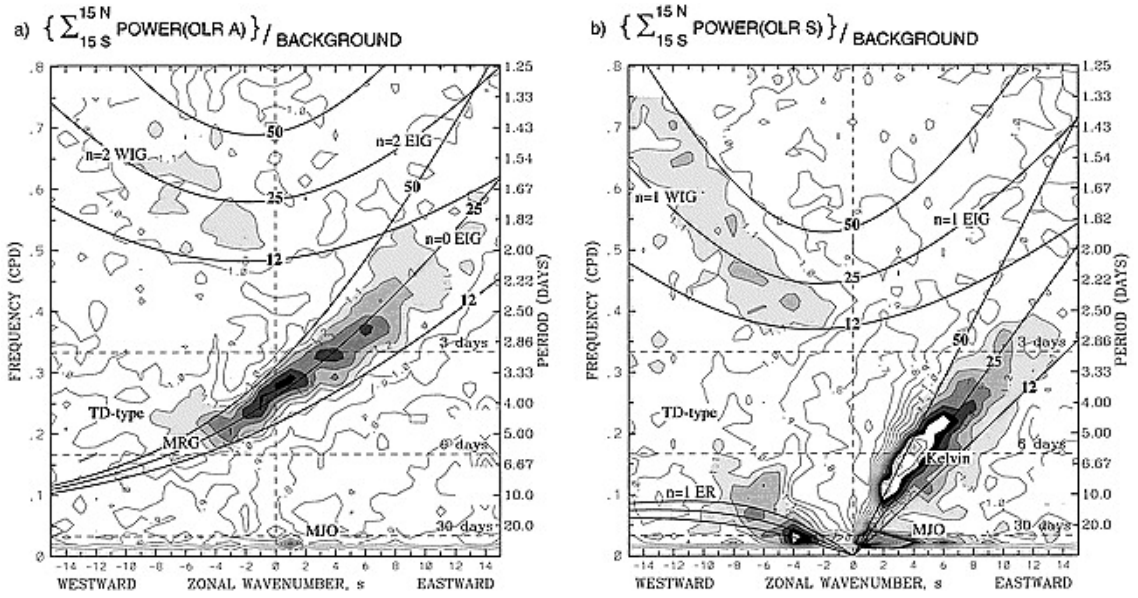

Figure 3.7: Power spectrum OLR divided by background power-spectrum (a) Anti-symmetric (b) Symmetric. Dataset comes from 18 years twice-daily record of satellite observed OLR over the 15N-15S domain. *From Wheeler and Kiladis (1999).*

3.6 Madden Julian Oscillation

In the equatorial Indian and Western Pacific ocean there is a propagation of a wave-like phenomena defined as the Madden-Julian Oscillation (MJO). It features a large-scale eastward moving center of strong deep convection and precipitation (active phase), flanked by regions of weak deep convection and precipitation to the east and west (inactive or suppressed phase). It travels with an average speed of $5\,m/s$ and has a local intra-seasonal period of $30 - 90\,days$.

The MJO is the main component of the intra-seasonal variability in the tropical atmosphere and encompass a wide range of interactions from mesoscale to large scale in the spatial domain. Since it was first documented by Madden and Julian (1971, 1972), the interest of the atmospheric community on the MJO has intensified because of its extensive interactions with other components of the climate systems like modulation of tropical cyclones (Liebmann *et al.*, 1994), the onsets and breaks of the Asian-Australian Monsoon (Yasunari, 1979) or modulation of the timing and evolution of El Niño (Lau and Chan, 1998a). The MJO's influence extends via teleconnection mechanisms to the extra tropics and its weather patterns (Weickmann and Berry, 2011). The following subsections details the MJO characteristics (sect. 3.6.1), the GCM's capacity to simulate them (sect. 3.6.2) and the set of diagnostics

employed for its evaluation (sect. 3.6.3). For further information the reader is referred to the Zhang (2005) literature review.

3.6.1 Main features of the MJO

The two phases of the MJO, active and suppressed, are connected by overturning zonal circulations that extend vertically through the entire troposphere. In the lower troposphere about the 850hPa level and near the surface, there is an anomalously strong convergence of zonal winds (easterlies to the east and westerlies to the west). Zonal winds reverse directions in the upper troposphere, typically at 200hPa level.

The behaviour of the MJO across its different stages is illustrated in Figure 3.8. An MJO event starts with a negative pressure anomaly over East Africa and the Indian Ocean at stage F, by stage G the pressure anomaly has spread eastwards along with the eastern circulation cell, at stage H, the zonal circulation cells indicate that the centre of large-scale convection has moved eastwards across Indonesia. By the stage A, the two circulation cells are nearly symmetric. At B the western cell shrinks and pressures rise over the Indian Ocean, signalling the weakening of the convection and the emergence of an inactive phase in the Indian Ocean. Weak convection is signalled at stage C, on stage D there is no convergence in the lower troposphere for the active cell, located now in the Atlantic. The MJO finishes at stage E when the inactive phase reaches Indonesia and there are two nearly symmetrical circulation cells.

The eastward moving active phase is actually a myriad of high frequency small-scale convective systems moving in all directions (Nakazawa, 1988). Generally new convective systems develop eastwards of the previous ones, moving the large convective centre of the MJO to the east. The most noticeable high-frequency variability within the large-scale ensemble of cloud clusters are eastwards propagating synoptic-scale disturbances at the speed of coupled Kelvin waves.

3.6.2 Representation of the MJO in GCMs

A realistic simulation of the MJO in a GCM would benefit the medium and long term prediction beyond the synoptic scale systems and help to gain confidence about the impact of climate change in diverse processes. However, GCMs have serious

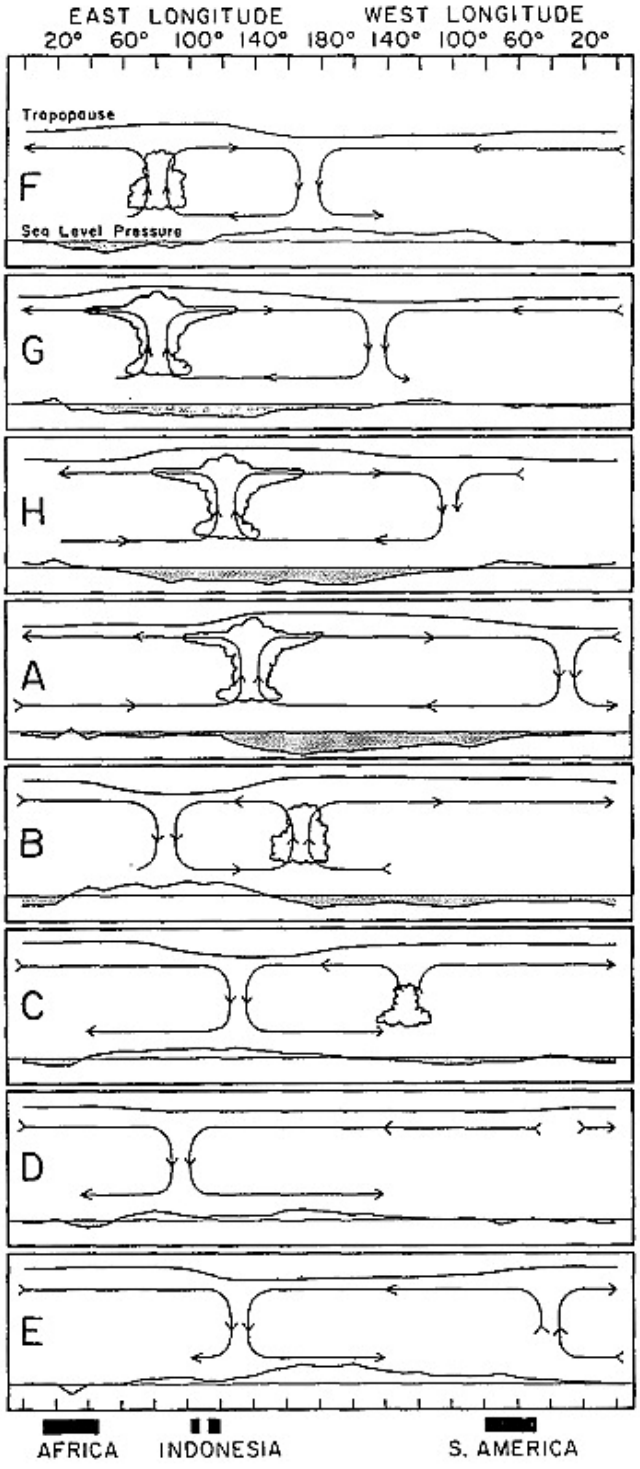

Figure 3.8: Schematic diagram illustrating the different life stages of an MJO event along the equator. Cloud symbols represent the convective "active" phase, arrows indicate the zonal wind circulation and curves above and below the circulation represent perturbations in the upper troposphere and sea level pressure. *From Madden and Julian (1972).*

drawbacks representing the MJO characteristics. The eastward propagating signals are too weak, their propagation speed too fast and their spatial distribution and seasonal cycles are unrealistic. Some models can reproduce the dynamics of the MJO adequately but fail to couple dynamics and convection well, generating precipitation in the wrong places (Zhang, 2005).

The deficiencies of cumulus convection parametrizations are normally responsible for the inability of models to simulate the MJO properly. The MJO is more realistic when the GCM has a realistic convection-moisture relationship. The majority of the models tend to moisten the troposphere uniformly at much lower precipitation thresholds than observed (Xavier, 2012).

A good representation of the climatological mean state is important. Simulated MJO signals tend to be stronger in models whose mean seasonal cycles are stronger and whose mean precipitation is more realistically distributed with respect to Sea Surface Temperature (Slingo, 1996).

3.6.3 MJO diagnostics: CLIVAR diagnostics

The CLIVAR MJO working group has developed a series of diagnostics to evaluate the boreal winter and summer MJO (Waliser *et al.*, 2011), it is composed in two sets of diagnostics:

- *Level 1:* They are meant to provide an initial assessment of the model's spatial and temporal intraseasonal variability and the most basic features of the MJO that can be easily calculated and understood by a non-MJO expert, like maps of intraseasonal variance to reveal whether the model produces the right intraseasonal variability and its correct seasonality.

- *Level 2:* These diagnostics provide a more comprehensive diagnostics of the MJO through multivariate analysis, wavenumber frequency spectral decomposition and composite analysis of MJO lifecycles.

The variables used for level 1 and 2 diagnosis are OLR, precipitation and zonal wind at 850 and 200 hPa. Because in boreal summers systems travel NortEast from West-Africa to the Indian Subcontinent, the MJO diagnostics are calculated separately for boreal summer and winter.

Level 1 and 2 set of diagnostics were applied to three coupled and five uncoupled models in Kim *et al.* (2009), they found that generally, the MJO signal in the large-scale circulation variables (like horizontal wind at 850hPa) is better represented than in convection (precipitation), the intraseasonal variability of precipitation and low-level horizontal wind (U_{850}) is stronger than observed in the majority of GCMs, and often the dominance time-scale of the MJO modes is outside the 30-80 day band, showing that strong MJO events last shorter in models than observations.

The CLIVAR MJO diagnostics are quite useful to show shortcomings in the ability of models to simulate the MJO. However, they do not directly indicate which physical processes are most important or responsible for the quality of the MJO. There is a need to explore and developed more process-oriented diagnostics, like vertical profiles of diabatic heating from different parametrizations to gain new insight into the convective interactions necessary for MJO simulations, like the project described by Petch *et al.* (2011).

Chapter 4

Methodology

The main concepts to understand stochastic physics schemes and the evaluation diagnostics employed in this thesis have been described in the previous chapters. It is also necessary to specify and describe the model we employ and its characteristics as well as the observational products we utilize to evaluate the realism of the impacts of stochastic parametrizations.

The statistical significance of the differences between the results obtained from our experiments and the observational dataset is obtained from a student t-test. Confidence intervals are given by $1.96\sigma/\sqrt{N}$; where σ is the standard deviation for the chosen diagnostic of the phenomenon sampled (e.g. RMSE or the distance error of mid-latitude cyclones), N is the number of observations and the 1.96 gives the 95% statistical significance.

In the present chapter, a description of the GCM we employ in this thesis is given in section 4.1. It is the Met Office Unified Model (MetUM) and we make use of the Global Atmosphere configuration 3 (GA3, described in sect. 4.1.2) and GA6 (sect. 4.1.3). The chapter also includes a description of the products employed to verify the model; the reanalyses (sect 4.2), as well as specific products for precipitation, such as the Global Precipitation Climatology Project (GPCP, sect. 4.3) and the radiative effects of clouds, such as the Clouds and the Earth Radiant Energy System (CERES, sect. 4.4).

4.1 Met Office Unified Model

The United Kingdom Meteorological Office (abbreviated as "Met Office" or UKMO) developed the first version of the Met Office Unified Model (MetUM) in the early nineties, under the need to create a global GCM which could be used for climate and NWP activities. The model was a merge of the dynamics of the previous NWP model, as it incorporated an efficient integration scheme into a conservative finite-volume dynamical formulation (Cullen and Davies, 1991), and most of the physical parametrization package from the climate model. It took an additional two years until the climate model was consider acceptable. The performance of this first version of MetUM was documented in Cullen (1993).

Under this unified model approach, changes to the model were progressively developed and adapted to both configurations. A new Semi-Implicit Semi-Lagrangian (SL, see sect. B.2.1) formulation was developed to enhance the non-hydrostatic capability of the model (Davies *et al.*, 2004). Major physics development in one system were usually introduced to the others within a year or two, like the development of a new orographic drag scheme or the Prognostic Cloud fraction and Prognostic Condensation scheme (PC2, Wilson *et al.* 2008).

Different prediction systems were developed for different purposes, like seasonal prediction, short-range probabilistic prediction or centennial Earth System studies. The multiplicity of prediction systems revealed one of the main setbacks of the seamless prediction paradigm, which is that one change could be difficult to integrate in one particular system because of a degradation of performance in the other. One example is the low resolution climate model used for paleo-climate and Earth System studies. The Semi-Lagrangian formulation of the present MetUM version is too diffusive for low resolutions (see sect. B.2.1), therefore the previous version of this system, the Hadley Centre Climate Model v3 (HadCM3, Gordon *et al.* 2000), is still employed in many of these studies and it is still an important component on the "MetUM family" of prediction systems. Until recently the operational Decadal Prediction System DePreSys (Smith *et al.*, 2007) was built from HadCM rather than the last MetUM version for similar reasons.

Since 2010 onwards, the Met Office has taken the seamless concept one step further. It provides a yearly configuration with the same dynamics and physics

for all the model's systems, the "Gobal Atmosphere" configuration (Walters *et al.*, 2011). Any potential change in the model configuration is tested across NWP and climate scales. If the impacts are positive, this change is combined with others and tested again until a final configuration is reached.

MetUM with a GA configuration provides a framework where different prediction systems can be developed with different timescales, resolutions and prediction purposes. Figure 4.1 shows a schematic description of some of these systems for different timescales, atmospheric grid resolution and complexity.

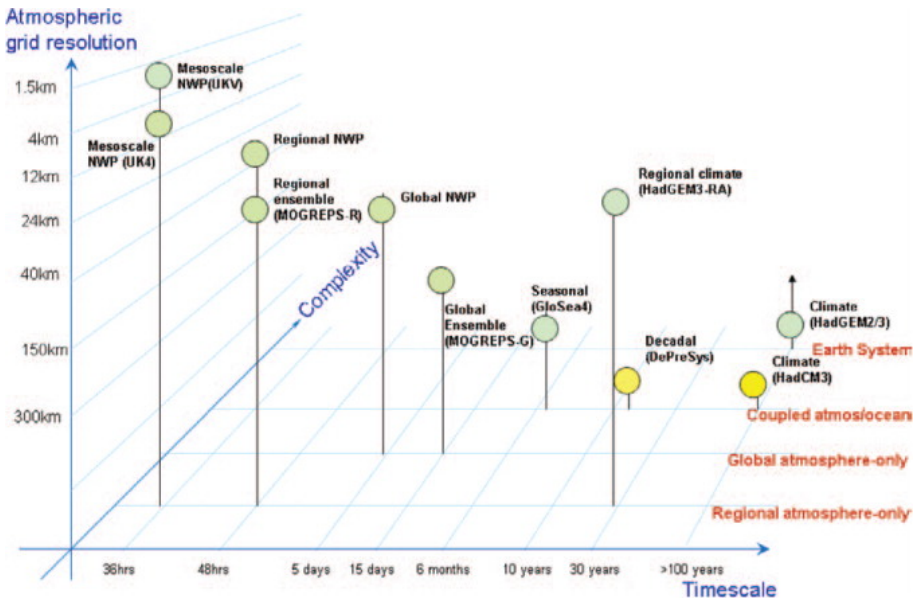

Figure 4.1: Schematic description of MetUM systems across different timescales (x axis), atmospheric grid resolution (y axis) and complexity (z axis). All the systems shown in green use GA configurations. The two yellow circles indicate the two exceptions to the seamless approach, HadCM3 and DePreSys (see text for details). *From Brown* et al. *(2012)*

The GA definition also includes a configuration for the Land surface scheme, which simulates the exchange of moisture, heat and momentum with the Boundary Layer (see B.3.4 for a brief description on Land surface schemes). MetUM is coupled to the surface Joint UK Land Environment Simulator (JULES, Best *et al.* 2011). The main characteristics of JULES at GA3 are described in section 4 of Walters *et al.* (2011) and it is active in all the MetUM systems.

Although many of the MetUM configurations of Limited Area Models (LAMs) are also within the GA framework, in the course of the thesis we only employ global

configurations. By convention their horizontal grid-mesh is defined on $2n$ longitudinal points and $1.5n + 1$ latitudinal points, this choice makes the grid-spacing approximately isotropic in the mid-latitudes. The integer n defines the resolution and represents the maximum number of zonal 2 grid-point waves that can be represented by the model. A model with $n = 96$, the typical resolution for decadal to centennial experiments, is defined as N96 and has a grid-length of 145 km in the mid-latitudes.

The seamless nature of the MetUM model enables a comprehensive study of the impacts of stochastic physics schemes across different horizontal resolutions and timescales, using probabilistic and deterministic forecasts to understand the nature of their perturbations on the atmospheric flow. The systems we employ in this thesis are described in the next subsection:

4.1.1 MetUM systems and their setup

- *Short-range deterministic forecasts (defined as "NWP forecasts"):* They are employed to make predictions of the evolution of the atmosphere for less than one week ahead. The system is useful to indicate the predictability of day-to-day weather events like cyclones, squall lines or fronts. It is also useful to diagnose the sources of model error, which normally spin up early on in the forecasts. The system uses prescribed evolving SST, and dynamic soil moisture and temperature given by JULES. It has 70 levels in the vertical (50 in the troposphere and 20 in the stratosphere), and they can be easily adapted to several different horizontal resolutions like N96 ($\sim 145\,km$ in the mid-latitudes), N216 ($\sim 65\,km$) or N320 ($\sim 45\,km$).

  In this thesis we run a set of 200 forecasts with this system. These forecasts are started in the dates shown in Table 4.1. They consist on the original 20 forecast dates employed for routine evaluation plus dates starting 3 days earlier, and 3,6 and 9 days later to obtain enough statistics for cyclone tracking. The years included range from 2008 to 2012.

- *The Met Office Global and Regional Ensemble Prediction System (MOGREPS):* A Short-range EPS, normally run for 3 days on research mode. It consist of 24 members, where 23 of them are perturbed by The Ensemble Transform

5 JUN	8 JUN	11 JUN	14 JUN	17 JUN
19 JUN	22 JUN	25 JUN	28 JUN	1 JUL
6 JUL	9 JUL	12 JUL	15 JUL	18 JUL
20 JUL	23 JUL	26 JUL	29 JUL	1 AUG
3 AUG	6 AUG	9 AUG	12 AUG	15 AUG
5 DEC	8 DEC	11 DEC	14 DEC	17 DEC
19 DEC	22 DEC	25 DEC	28 DEC	31 DEC
2 JAN	5 JAN	8 JAN	11 JAN	14 JAN
16 JAN	19 JAN	22 JAN	25 JAN	28 JAN
31 JAN	2 FEB	5 FEB	8 FEB	11 FEB

Table 4.1: Start-dates for the deterministic NWP set of forecasts. Start time is 12Z. Forecast are done from 2008 to 2012.

Kalman Filter (ETKF, Bowler *et al.* 2008) and stochastic physics schemes. See Bowler *et al.* (2008, 2009) for a comprehensive description of the system. It uses the same model framework as the deterministic NWP forecasts. This system is the natural habitat for the stochastic schemes and where they have been extensively evaluated. Therefore MOGREPS is used to diagnose the probabilistic scores of new developed schemes or changes to the present ones for the different investigations carried out in the thesis.

The MOGREPS probabilistic forecasts employed in this thesis have the following setup:

- 11 ensemble members without perturbations to the initial conditions, the ETKF is switched off. All the ensemble spread is generated by the stochastic physics schemes.
- The forecasts are run every 6hr between 18Z of 6 November 2012 to 12Z of 13 December 2012.
- The horizontal resolution is N216, model configuration uses similar characteristics as the deterministic NWP model. Forecast are initialized from the Met-Office Analysis (MO-AN).

As described in the list above, the experimental design of the EPS does not include the ETKF nor any other method to perturb the initial conditions. The

nonlinear feedback between stochastic schemes and the ETKF could produce a different ensemble spread and hence a difference in the perturbations to the initial conditions (Tennant W, 2013, *personal comms.*). These interactions would create differences on the ensemble characteristics that are hard to associate to the stochastic physics scheme alone. Additionally, there are technical constraints that made us exclude these initial condition perturbation schemes. They normally need a higher number of ensemble members than 11, that would increase the computational cost of the EPS experiments carried out for the investigations of this thesis, and therefore make some of the experiments nonviable.

The evaluation of these MOGREPS suites is done against the Met Office Analysis rather than ECMWF reanalysis or observations. The low number of ensemble members and forecasts might produce noisy results with poor statistical significant if compared to a reduce sample of observations, we are also interested in the global structure of the ensemble spread, therefore we employ an analysis to evaluate our experiments. The choice of MO-AN comes from the fact that it provides the model's initial state and if another analysis was used, it would add an extra source of uncertainty which would slightly contaminate our spread-error relationships.

- *MetUM Climate configuration:* Also known as Hadley Centre Global Earth Model v3 (HadGEM3) for continuity with previous models developed by the the Hadley Centre, the climate "branch" of the Met Office. HadGEM3's setup follows the Atmospheric Model Inter-comparison Project experimental design (AMIP, Gates *et al.* 1999). A general framework for climate model evaluation which uses prescribed SST, sea-ice and atmospheric composition from November 1981 to Decemeber 2001. Experiments are normally run for 20 years, which provide a useful comparison to see any noise-induced drift in the climate mean provoked by the stochastic schemes. Climate experiments also provide a large population of the atmospheric processes described in chapter 3 (e.g. mid-latitude cyclones or Convectively Coupled Equatorial Waves). Thus the experiments are also useful to understand the effects on stochastic schemes on the representation of these processes, major drivers of the atmospheric variability at synoptic scales. Unlike NWP systems, the climate system

includes 15 more vertical levels in the stratosphere, in order to represent the non-negligible effects of the middle atmosphere dynamics in the climate.

There are other systems that are unfortunately not included due to its computational requirements. The seasonal system is an obvious example, The Global Seasonal system (GloSea Arribas *et al.* 2011) at GA3 configuration uses a lagged ensemble of 42 members at N216 (Machlanan2014 *et al.*, 2014). Similarly a coupled climate model could provide useful information on how stochastic noise-induced drifts on the lower troposphere feed back on the ocean circulation.

4.1.2 Global Atmosphere 3.0/3.1 (GA3) configuration

The first GA cycle was GA3 which uses 3.0 for climate systems and 3.1 for NWP for operational purposes. All our experiments use 3.0 configuration for a clean test between the different timescales. The GA3 setup for MetUM is extensively described in Walters *et al.* (2011). The main components are:

- *Dynamical core:* GA3 uses a Semi-Implicit Semi-Lagrangian (SI-SL) formulation as described in Davies *et al.* (2004). The climate experiments use the less diffusive Quasi-Cubic interpolation scheme for the interpolation of the departure point whereas NWP uses Cubic (see sect. B.2.1 for an explanation of different types of departure points).

- *Radiation:* The radiation scheme of Edwards and Slingo (1996) is used in GA3 with a configuration based on Cusack *et al.* (1999). It employs 21 k-tems for the the major gases in the Short-Wave (SW) bands (H_2O, O_3, CO_2 and O_2) and 47 k-terms for the major gases in the Long-Wave (LW) bands. The sub-grid cloud structure is represented using the Monte-Carlo Independent Column Approximation (McICA) as described in Hill *et al.* (2011)

- *Cloud microphysics:* Uses the scheme based on Wilson and Ballard (1999) that incorporates modifications to the particle-size distribution and minimum cloud water content for autoconversion (Abel *et al.*, 2010), fall velocities (Abel and Shipway, 2007), and the inclusion of substeps for a more accurate representation of drizzle processes.

- *Large scale cloud:* MetUM at GA3 uses the Prognostic Cloud fraction and Prognostic condensate (PC2, Wilson *et al.* 2008) to determine the fraction

of cloud cover and the amount and phase of condensed water. PC2 is embedded in each parametrization that has an influence in clouds, e.g. it deals with the production of condensate from radiative heat within the radiation parametrization.

- *Orographic gravity wave drag:* The flow blocking and gravity wave effects due to subgrid orography are represented by a scheme based on Webster *et al.* (2003).

- *Non-orographic gravity wave drag:* MetUM at GA3 uses the scheme described in Scaife *et al.* (2002), it provides a spectrum of gravity waves in 4 azimuthal directions and represents the process of wave-generation, propagation and dissipation.

- *Atmospheric boundary layer:* The MetUM parametrization of the turbulent processes extends to the full depth of the troposphere. It is a first-order turbulent closure mixing scheme as described in Brown *et al.* (2008).

- *Convection:* MetUM uses a mass flux convection scheme based on Gregory and Rowntree (1990) with various modifications. Deep convection uses a different Convective Available Potential Energy (CAPE) closure than Gregory and Rowntree (1990) and shallow convection uses a different closure. The scheme also includes a convective momentum transport based on Gregory and Allen (1991).

4.1.3 Global Atmosphere 6 (GA6) configuration

The GA6 configuration for MetUM has been operational in the Met Office since May 2014. The main changes to GA3 include the new dynamical core Even Newer Dynamics for General Atmospheric Modelling of the Environment (ENDGame, described below), and several changes to the model physical parametrizations, such as:

- *Radiation:*

  - Improved CO_2 and O_3 LW absorption which improves heating/cooling the in the stratosphere (Zhong and Haigh, 2000).

  - Radiation timestep reduced from three hours to one hour, improving the accuracy of the radiation scheme.

- *Cloud scheme (PC2):*

  - Improved cloud erosion method and numerical definition for mixed-phase cloud.

- *Microphysics:*

  - Implementation of improved drizzle size distribution following Abel and Boutle (2012), which improves the representation of light rain.

- *Convection:*

  - Increased Entrainment rate in deep convection, following results from Klingaman and Woolnough (2013) which shows improvements in tropical variability such as the Indian Monsoon, Tropical cyclones or MJO.

- *Gravity Wave Drag (GWD):*

  - Introduction of a new version of the GWD scheme which includes a cut-off mountain approach to diagnose mountain wave-drag.

ENDGame:

The new dynamical core ENDGame shares many aspects with its predecessor, the New Dynamics scheme (Davies *et al.*, 2004). They both employ a SI-SL finite difference discretization and use same staggering for the grid variables. The details of the ENDGame discretization are reported in Wood *et al.* (2013). The most significant differences between ENDGame and New Dynamics are:

- ENDGame uses a nested iterative timestep structure, where the advection and fast physics (convection and boundary layer) are iterated to get a better estimate of the departure point. This also improves the numerical stability of the model.

- The increased stability allows the the semi-implicit time-weights to be closer to the time-centred value of 0.5. This reduces the damping of the explicit solution and thus improves the accuracy of the model.

- The horizontal grid is shifted half a grid length, so scalars are no longer held at the poles, improving the scalability.

- The continuity equation uses a SI-SL discretization instead of the previous Eulerian approach on the New Dynamics, this further improves the accuracy and stability of the dynamical core, but at the cost of losing the mass conservation properties. A mass-fixer scheme as described in Zerroukat (2012) is employed.

ENDGame improves the scalability, stability and accuracy of the dynamical core in comparison to New Dynamics (Walters *et al.* 2015, *in preparation*). It enables future upgrades in resolution that would have been computationally unaffordable. It significantly reduces the failure rate in high resolution climate simulations that sample a wide range of years and synoptic conditions. It improves the departure point calculations, producing a more accurate simulation of the intensity of systems like mid-latitude cyclones.

4.2 Reanalysis

Retrospective analysis or reanalysis is one of the most important means to validate climate models and study climate processes. They are produced by constraining a NWP system by long-term observations into an unvarying data assimilation system. This produces a 4-dimension homogeneous output data. Thus it makes easier to validate models than inhomogeneous raw observations from different sources. Reanalysis are a powerful tool for atmospheric research but their products have also uncertainties. In addition to the uncertainty of the observing systems, data assimilation can fill the gaps by adding physically meaningful information from forecast models, but the model dominates in regions with spare observations such as the Southern Ocean.

The need for reanalysis was advocated by Trenberth *et al.* (1988), they found significant discontinuities in operational analysis, that were related to changes in the forecast model and analysis systems. Shortly after, the first generation of reanalysis was developed, where the NCEP-NCAR reanalysis (Kalnay *et al.*, 1996) and European Re-Analysis project (ERA-15, Gibson *et al.* 1997) were the most popular. The production of reanalyses is an ongoing program in several NWP centres with

new production of reanalyses as model and data assimilation methods are improved. The reanalyses used in this thesis to compare the effects of stochastic physics are:

- *ERA-Interim* (ERA-I, Dee *et al.* 2011) The primary goal for ERA-Interim has been to address the problems found by its predecessor ERA-40 (Uppala *et al.*, 2005). These problems are mainly related to the representation of the hydrological cycle, the quality of the stratospheric circulation, and the consistency in time of reanalyzed geophysical fields. Compared to its predecessor ERA40, ERAI has got a higher horizontal resolution, from T159 to T255 and a wide variety of changes in the physics package described in section 3 of Dee *et al.* (2011). A four-dimensional variational (4DVAR) data assimilation system with 12-h (Thepaut *et al.*, 1996) cycling is used with output every 6 hours. The number of observations assimilated in ERA-I has increased from approximately 10^6 per day in the early 1989 to 10^7 per day in 2010. Figure 4.2 shows the number of different observations incorporated into ERAI split into the observed fields. ERAI covers the period from 1st January 1989 onwards with an extension from 1979 to 1988 in preparation. The main limitation in the ability of ERAI to describe the evolution of the atmosphere lies in the quality and availability of observations of humidity, winds and the distribution of aerosols in large parts of the atmosphere.

- *The NASA Modern-Era Restrospective analysis for Research and Applications* (MERRA, Rienecker *et al.* 2011). MERRA, as ERA-I, was also developed with the aim to improve the representation of the hydrological cycle, and also to incorporate the observations from NASA's Earth Observing System (EOS) satellites. MERRA is generated from the Goddard Earth Observing System version 5 (GEOS) model and its data assimilation system (Rienecker *et al.*, 2008). GEOS-5 uses finite-volume dynamics with a resolution of 1/2 degrees of latitude and 2/3 degrees of longitude ($\sim$55 km in the mid latitudes) with 70 levels. The data assimilation system is the Three Dimensional Variational data assimilation (3DVAR) based on the Gridpoint Statistical Interpolation scheme (GSI, Wu *et al.* 2002), with a 6 hourly update cycle. Differences to ERA-I and other new generation reanalysis have been reduced from previous generations. However, there are still substantial differences in poorly constrained quantities such as precipitation and surface fluxes due to differences in assimilating mod-

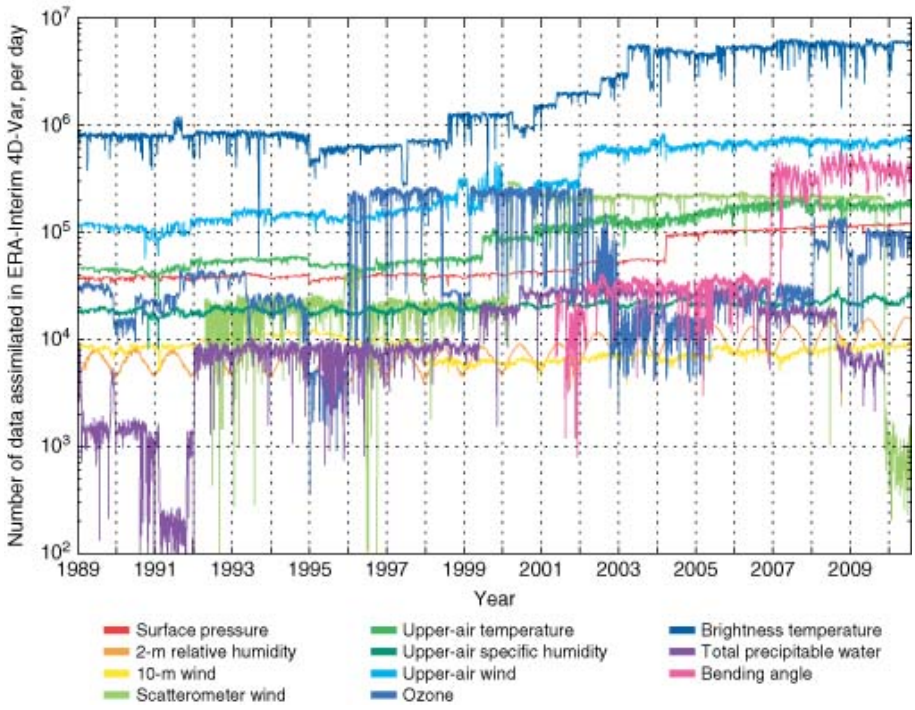

Figure 4.2: Number of daily observations assimilated into the atmospheric analysis of ERAI, on a logarithmic scale. *From Dee* et al. *(2011).*

els and differences in the number and quality of observations assimilated, for instance ERAI incorporates the Global Positioning System Radio Occulation (GPS-RO) whereas MERRA does not assimilate these observations (Dee *et al.*, 2011). MERRA covers the time period from 1979 onwards.

4.3 Global Precipitation Climatology project (GPCP)

Precipitation is a key variable to examine in the evaluation in a GCM. It is very sensitive to large-scale disturbances like tropical waves (section 3.5) or extra-tropical cyclones (section 3.3), and feedbacks on these by the release of latent heat. In addition, the impacts of precipitation on our society are quite evident (e.g. water management, agricultural and electricity generation).

A large portion of precipitation occurs over the oceans, where there is a lack of ground-based observations, and satellite observations have limited temporal sampling (especially low-Earth orbiting ones). In order to overcome these setbacks and produce a high quality precipitation database, the GPCP project was launched by the Global Energy and Water cycle Experiment (GEWEX). It is a merged analysis

that incorporates precipitation estimates from low-orbit satellite microwave data, geosynchronous-orbit infrared data and surface rain-gauge observations. Details of the merging algorithm are described by Huffman *et al.* (1997). GPCP data used in this thesis is from version 2 (Adler *et al.*, 2003). It has a spatial resolution of 2.5°x 2.5°latitude-longitude box and its temporal span covers from January 1979 to the present.

As an example of is products, the total average of precipitation from 1979 to 2001 is shown in Figure 4.3. The main aspects of the global distribution of precipitation are clear: The maxima in the tropics along the Intertropical convergence zone (ITCZ), the South Pacific convergence zone (SPCZ), dry areas in the eastern parts of the subtropical oceans. In the mid latitudes the storm track signature of precipitation is clear in the Atlantic and Pacific Oceans on the NH and the Southern Ocean (SO) storm track in the SH.

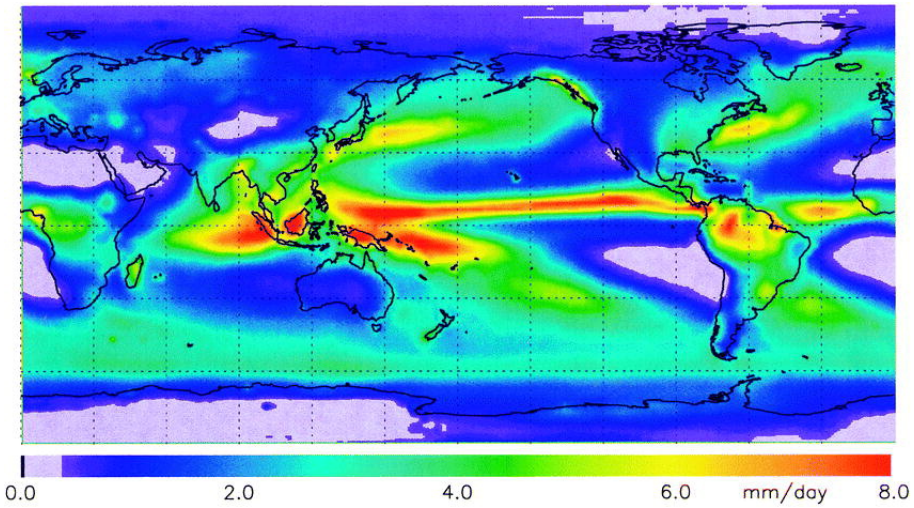

Figure 4.3: The 23 year (1979-2001) annual mean precipitation (mm / day). *From Adler* et al. *(2003)*

Because of the desire to obtain the longest record, the GPCP is inhomogeneous in terms of its inputs and datasets. This inhomogeneity is minimized by the calibration of different datasets of different temporal periods. In this thesis there is no study of global or regional climate trends of precipitation, therefore this apparent setback does not invalidate the main conclusions. There are some other know problems with GPCP. Polar precipitation estimates over land are derived solely of cloud information, GPCP underestimates precipitation over regions with orographic features (Nijssen *et al.*, 2001), which could be related to the relative lack of rain gauges

in mountainous regions, or because satellite observations both passive microwave and IR also have difficulty detecting shallow orographic precipitation. Pfeifroth *et al.* (2013) compared GPCP v2 along MERRA, ERAI and other precipitation products against PACRAIN, a ground-based rainfall dataset in the West Pacific (Greene *et al.*, 2008). GPCP shows the highest correlation and lowest monthly deviations with reference PACRAIN station data in comparison with other products.

4.4 Cloud and the Earth's Radiant Energy System (CERES)

One of the major sources of uncertainty of GCMs lies in the impact of clouds upon the radiative energy transfer (see sect. B.3.2). CERES is an instrument onboard of the Earth Observing System satellites, whose measurements of radiative fluxes are suitable for examining the role of clouds in the radiative heat balance of the climate.

The CERES instrument consist of a three-channel scanning broadband radiometer, a detailed description of the instrument is given in Wielicki *et al.* (1996). The first channel measures the thermal radiation emitted from the Earth's surface in the $8 - 12\ \mu m$ "window", to provide accurate measurements of clear-sky longwave measurements. The remaining two CERES spectral bands measure shortwave ($0.2 - 5\ \mu m$) and total ($0.2 - 100\ \mu m$) broadband radiation. The three telescopes are co-aligned such as they share a 98% common field of view with a spatial resolution of 20 km at nadir.

Errors in measuring radiative fluxes come from three different sources: Instrument calibration and stability, insufficient sampling of the angular variation of radiation and the inability to adequately sample the large diurnal variation of solar-reflected and earth-emitted radiation. The last two errors are reduced by the use of three spacecraft: Tropical Rainfall Measuring Mission (TRMM) launched in 1997, Terra (launched in 1999) and Aqua (launched in 2002). In October 2011, another CERES instrument was launched aboard the Suomi National Polar-Orbiting Partnership (NPP) while the instrument aboard of TRMM ceased to operate in March 2000.

The filtered broadband radiances are split into shortwave and outgoing longwave,

using the approach described in Loeb *et al.* (2001). The radiances are then converted to a radiative flux using empirical angular directional models (ADMs, Loeb *et al.* 2007 and references therein), which are defined according to various surface, cloud, and atmospheric properties. CERES' final product is the observed Top of the Atmosphere (TOA) broadband reflected SW, LW, and downward net fluxes at 3-hourly, daily, and monthly temporal scales.

Chapter 5

Evaluation of SKEB2

The evaluation of the operational stochastic physics schemes have been heavily focused on their capacity to produce ensemble spread, a very useful feature given that nearly all of the state-of-the-art Ensemble Prediction System (EPS) do not produce enough spread (Buizza *et al.*, 2005). However, many questions regarding the impacts on the representation of processes in well resolved scales, such as planetary scales or synoptic scales on the short range, have not been investigated.

One of the most successful stochastic physics schemes, across different operational EPS, is the Stochastic Kinetic Energy Backscatter (SKEB). Previous studies have reported that the family of SKEB schemes is able to increase the ensemble spread, improve EPS skill scores and reproduce the $k^{-5/3}$ slope on the kinetic energy power spectra (see section 2.2 for details in their formulation and results). Despite its physically based implementation driven by the dissipation masks, the scheme has many internal parameters whose values seems to be chosen to maximize the EPS skill scores. One of these parameters is the backscatter ratio b_R, which is the fraction of energy unrepresented by the model that is backscattered to the chosen scales (see eq. 2.1). b_R is an amplitude factor to modulate the forcing of the scheme, the higher b_R the higher is the forcing to the streamfunction and the impact of the scheme. The current value is set to 0.0275 which is the optimal value to produce enough realistic spread and a realistic kinetic energy spectra, as reported by Berner *et al.* (2009). However this value has been increased to 0.2 to enlarge the ensemble spread of the seasonal system of the Integrated Forecast System (IFS). This increase seems to cause a too strong activity over the Southern Ocean and too much power on the kinetic energy spectra from wavenumber 30 (Fig. 2 of Berner *et*

al. 2012).

Another choice when implementing SKEB schemes is the range of scales where the forcing pattern is active. As described in sect. 2.2, the Spectral Stochastic Backscater Scheme (SPBS) forces all wavenumbers. SKEB2 forces wavenumbers $5 < n < 60$ and the Canadian SKEB forces smaller scales with $40 < n < 128$. Tennant *et al.* (2011) reported that the range of wavenumbers taken for SKEB2 was the optimal for the greatest growth of ensemble spread and minimal impact of forecast skill. But there are strong doubts about the effects of perturbations on the streamfunction at energy containing scales ($n < 10$), whose evolution is well resolved by current deterministic models on the short-range until they are infected from error upscaling from smaller scales (Tribbia and Baumhefner, 2004). Strong perturbations at these large scales might be detrimental to the skill of deterministic forecasts, constraining the use of these schemes to EPS only.

In the present chapter, we carry out a comprehensive evaluation of the Stochastic Kinetic Energy Backscatter v2 (SKEB2, described in 2.2.2) using the MetUM seamless framework across different scales, using the evaluation tools described in chapter 3. The research is done using MetUM GA3 configuration for NWP 5-day deterministic forecasts and climate simulations (all systems are described in sect. 4.1). The effects of the scheme over different model configurations (GA3 and GA6) are also investigated, as well as a similar comparison to the one carried out by Berner *et al.* (2012) to diagnose whether the improvements of stochastic schemes could compete with model improvements and increased resolution. Different arbitrary values of the backscatter ratio are used to show the effects of a too overactive scheme, these include the default value of 0.0275, 0.1, 0.2 and 0.3 which translates as the 30% of the energy lost is backscattered into the energy spectra. The main results of this chapter have been published in Sanchez *et al.* (2014).

The first section of the chapter explores the impact of SKEB2 in short-range 5-day deterministic NWP forecasts (5.1). The deterministic NWP system is described in section 4.1.1. Forecasts are carried out using different horizontal resolutions, N96 ($\sim 145\,km$ in the mid-latitudes), N216 ($65\,km$) and N320 ($45\,km$) with and without SKEB2. There are additional experiments done with the lower resolution with the different values of the b_R.

The evaluation methodology for these NWP forecasts covers the use of traditional

verification scores for deterministic simulations (sect. 3.1), and the representation of mid-latitude cyclones using the TRACK algorithm (sect. 3.3.1), which adds useful information on the impacts of the SKEB2 on one of the main drivers of weather in the mid-latitudes.

In addition to the NWP forecasts, climate simulations with SKEB2 using the range of different values for b_R are used to investigate the effects of the scheme in the mean climate and variability in sect. 5.2. Looking at processes such as blocking, mid-latitudes, CCEW or the organization of convection. Results of the model are compared against ERAI, GPCP and CERES (see chapter 4 for a description of these products). The results from this research help to indicate whether there is any particular physical process where the effects of SKEB2 are detrimental, as well as the processes which are improved by the stochastic forcing of the streamfunction. Another important question that is also investigated relates to the capacity of SKEB2 to generate spread at climate timescales, and therefore be a useful tool for future probabilistic climate models (Palmer, 2012).

Given the positive results found by Berner *et al.* (2012) where the impacts of the SPBS were superior to those made by a higher resolution and slightly inferior than those made by improved physics. A similar comparison is done amongst simulations with the GA3.0 system at N96, N96 with SKEB2, N216 and an additional N96 experiment with the GA6.0 system (see sect. 4.1.2 and 4.1.3 for a description of both model configurations); such comparison also provides a useful set of results to see if the effects of the stochastic backscatter are comparable to a less diffusive dynamical core and improvements in the representation of convection included in GA6.0 system. A comparison of the effects of SKEB2 in GA3.0 and GA6.0 is also carried out, showing the sensitivity of the schemes to the model used and whether the scheme modulates itself if the model error decreases, e.g. using ENDGame, a less diffusive dynamical core (detailed in sect. 4.1.3).

A set of final conclusions is drawn from these three sections (sect. 5.4). Highlighting the main benefits of the SKEB2 scheme but also their detrimental impacts, e.g. worsening the deterministic scores, impacting the wrong scales or producing a poor simulation of physical processes. On the chapter 6 of the thesis, some different approaches will be investigated to minimize the negative impacts found.

5.1 Results from NWP forecasts

In order to determine the realism of the SKEB's representation of the missing energy, a set of 200 deterministic forecasts are run at different resolutions (see Table 4.1 for the start-dates of forecasts). They provide evidence on how the scheme reduces itself when moving to higher resolutions where more small scales are resolved. Also, different values of the backscatter ratio b_R are used to investigate how the effects of the scheme scale up with a stronger amplitude. The effects of the scheme are evaluated using traditional skill scores and the statistics of tracked mid-latitude cyclones.

5.1.1 NWP skill scores

A general perspective of the SKEB2 performance can be gain using traditional verification scores such as the Root Mean Error Square (RMSE) or Anomaly Correlation Coefficient (ACC, see section 3.1 for a description of both scores). The skill simulating the large-scales in the mid-latitudes could be reveal by variables such as geopotential at 500hPa (defined as Z_{500}) and Pressure at Mean Sea Level (PMSL). Dynamical variables like winds or temperature could indicate structural model uncertainties in the representation of atmospheric processes, like feedbacks from turbulence or radiative processes in the large-scale flow.

The set of 200 MetUM forecasts is compared to the ECMWF analysis, Since SKEB2 forces the streamfunction, we start looking at the impacts on the horizontal winds (modulus of the zonal wind u and meridional wind v). Figure 5.1 shows the temporal evolution of the mean RMSE for different regions and two levels: low (250hPa) and high (850hPa). These are the right levels to monitor the jet stream and cyclones in the mid latitudes, and convergence (divergence) at low (high) levels in the tropics. The mean RMSE has been averaged across all the forecasts. RMSE increases almost linearly, and the values are higher at 250hPa, as winds are stronger at this level. The RMSE of the forecasts with SKEB2 is always above the control ones, with a more detrimental impact on the lower level, where forecasts with SKEB2 loose about 5 hours of predictive skill. Even the skill of the forecasts of the high resolution model with SKEB2 is worse than the low resolution without the scheme in the tropics at 850hPa (Fig. 5.1.e). Forecasts of the high resolution model for day 2 are nearly as bad with SKEB2 as those at day 3 without it.

Another interesting result from Figure 5.1.e is the fact that the mean RMSE of the forecasts with SKEB2 increases with resolution, in disagreement with the rest of regions and levels. In the Tropics, low level winds are controlled by convergence around convective cores. SKEB2's kinetic energy release is proportional to the vertical gradient of mass fluxes, therefore the stochastic scheme could modify these convection episodes and the winds around them. This effect could stronger at high resolution as the vertical gradient of mass fluxes is likely to increase.

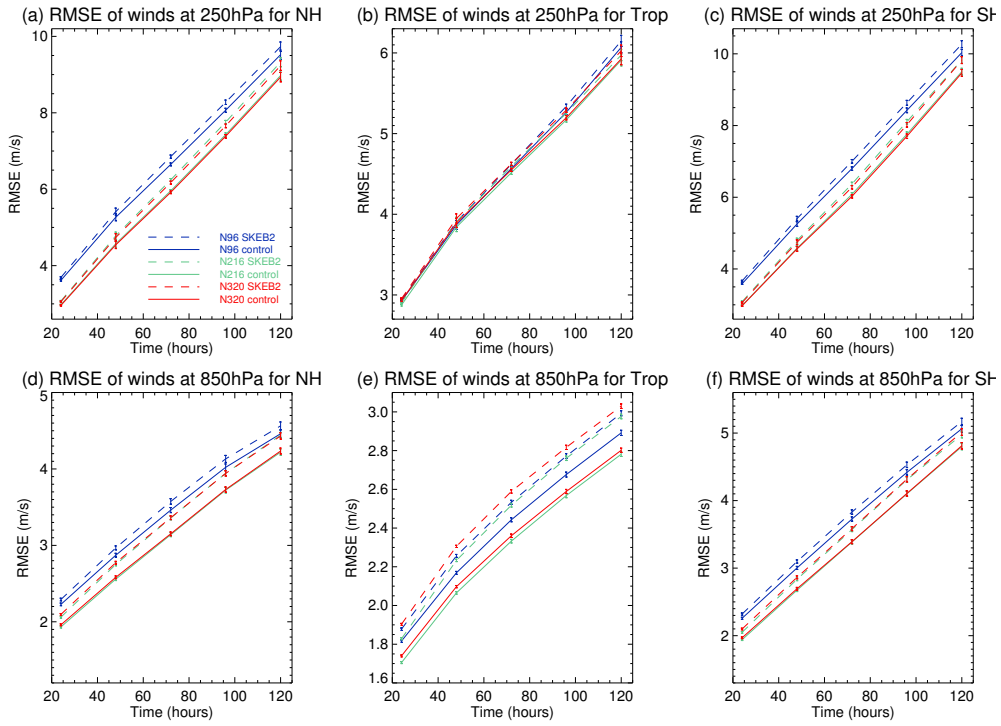

Figure 5.1: Mean value of the Root Mean Square Error (RMSE) for wind (modulus of u and v) forecasted over two different levels: 250hPa for (a), (b) and (c), 850hPa for (d), (e) and (f) and three different regions: NH (20N-90N) for (a) and (d). Trop. (20S-20N) for (b) and (e). SH (20S-90S) for (c) and (f). Blue shows N96 horizontal resolution, green N216 and red N320. Continuous lines denote the control runs across different resolutions and dashed line the runs with SKEB2. Confidence intervals are obtained using the methodology explained in the introduction of chapter 4.

The degradation of RMSE of wind fields by stochastic schemes have been reported in other studies. The original version of the Stochastic Parametrization of Perturbed Tendencies (SPPT, described in sect. 2.3) in the IFS did degrade winds in the NH and tropics at 850hPa as shown in Figure 11 of Buizza *et al.* (1999). A version of SKEB2 in a previous model configuration had similar impacts on the

global 250hPa winds with a degradation of a couple of hours in terms of predictive skill (see Fig. 11 of Tennant *et al.* 2011), although it also shows that such degradation is smaller than the one done by the perturbations to the initial condition.

The deterioration of the forecasts with lead time is clear in Fig. 5.1, but it is hard to see whether SKEB2 negative impacts increase or decrease with forecast lead time, levels, regions or resolutions. The ratio between the averaged RMSE of the forecasts with SKEB2 and the averaged RMSE from control forecast is shown in Figure 5.2. The ratio is always greater than 1.0 and errors at 850hPa are higher than at 250hPa. When resolution increases errors become more than 5% higher with SKEB2, an important degradation in terms of skill. The issue of the double penalty of RMSE becomes apparent here, as higher resolution can have a sharper representation of features such as fronts or tropical squall lines, therefore impacting negatively on the RMSE score if these are better resolved but slightly misplaced. The ratio between the errors increases until day 2 of the forecast and then it drops, more notably in the high resolution simulations.

If we increase the amplitude of SKEB2 on the low resolution model, the average RMSE of winds increases dramatically as shown in Figure 5.3; up to 35% for $b_R = 0.3$ in the tropical winds at 850hPa. The same pattern as Fig. 5.2 is observed, the increase of the RMSE ratio at level 850hPa is higher than at 250hPa and errors peak at day 2. The higher the SKEB2 forcing the more severe is the damage to the forecast skill in terms of point by point differences in the horizontal wind field.

Results for Z_{500}, a variable that depicts the large scale features driven by the geostrophic theory, are also quite negative. Table 5.1 shows the Z_{500} value of RMSE averaged amongst all forecasts for day 5 for all the different experiments in the Northern and Southern Hemisphere. As it does for winds, the increase in b_R deepens the degradation of Z_{500} RMSE.

The ACC of winds at 850hPa decreases when SKEB2 amplitude increases (Table 5.2), implying than it has also a negative effect in the temporal correlation of the weather low-wind structures in addition to the spatial mismatch.

There is a clear deterioration of the deterministic model when the SKEB2 is included. The Met Office Global NWP index, a combination of the RMSE of geopotential height and winds at different levels and locations is degraded by an equivalent

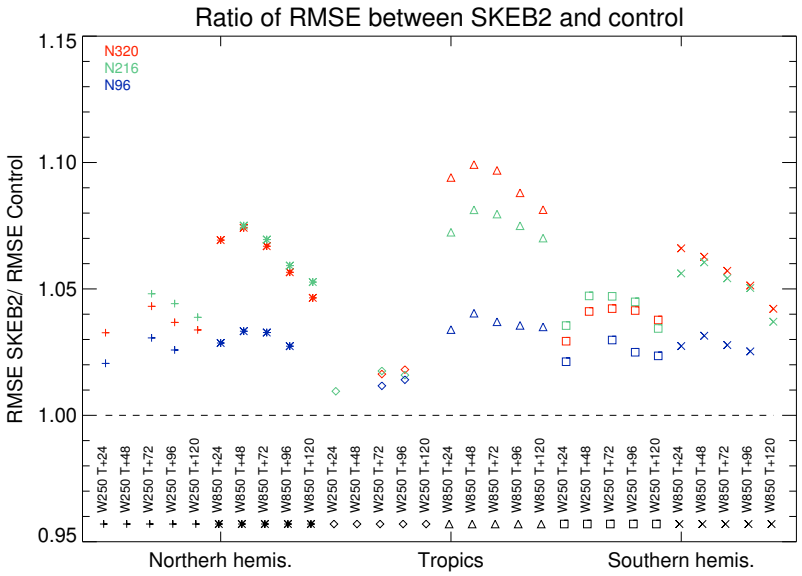

Figure 5.2: Ratio between the RMSE of SKEB2 and RMSE of control averaged across all forecasts for several forecast times. Pluses denote winds at 250hPa and asterisks winds at 850hPa for NH, diamonds winds at 250hPa and triangles at 850hPa for Tropics, squares winds at 250hPa and crosses winds at 850hPa for SH. Red is for N320 SKEB2/control ratio, green for N216 and blue for N96. Only differences above the 95% of statistical significance are shown.

of three years of model development work (Walters 2011, *personal communication*). The stochastic forcing is more detrimental on the first 2 days of the forecast, the short-range timescales where the model shows high predictability driven by the large-scales. On those timescales, the effects of small-scale errors have not contaminated the skill of the forecast fully (as shown in Figure 2 and 4 of Tribbia and Baumhefner 2004). SKEB2 aims to represent the small-scale energy dissipation, but on its current formulation it seems to force small wavenumbers which could introduce errors on the large scales.

5.1.2 NWP bias

Despite the detrimental effects of SKEB2 in individual forecasts, the averaged bias across all the forecasts of winds at 850hPa for day five shows an improvements (Figure 5.4). The winds are too weak in the mid latitudes (Fig. 5.4.a), probably because of the internal diffusion of the model, where two processes are probably the most responsible: the interpolation to the departure point and the excessive off-

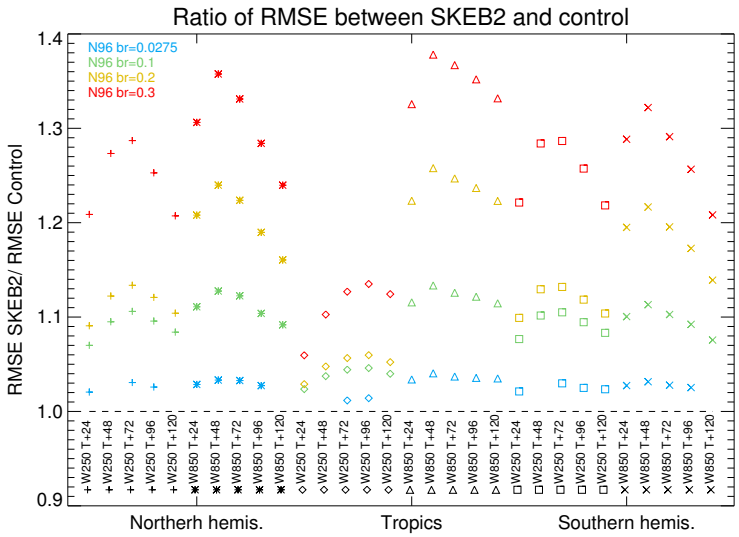

Figure 5.3: Same as Figure 5.2 but for different SKEB2 backscatter ratios b_R at N96. The ratio of SKEB2 with the default backscatter ratio (equal to 0.0275) is shown in blue , $b_R = 0.1$ in green; $b_R = 0.2$ in yellow and $b_R = 0.3$ in red.

centering towards the implicit solution in the Semi-implicit scheme (Woods N, 2014, *personal comm.*). When we increase the amplitude of SKEB2 the winds become stronger, which removes the biases in the mid-latitudes. However, at high b_R it also creates too strong winds in the Tropical West Pacific and Maritime continent. For the highest b_R, the convective part seems to be producing too strong winds, and this is probably an indication that the scheme is backscattering too much kinetic energy around convective cores, whereas in the mid-latitudes SKEB2 has not compensated all the energy dissipated by the implicit diffusion, there are still weak winds over the Southern Ocean (as shown in Figure 5.4.e)

The mean field of winds at 850hPa shows lower RMSE when SKEB2 backscatter increases until $b_R = 0.2$. For the highest backscatter ratio, the negative effects of over-active winds in the tropics compensates the benefits of stronger winds over the mid-latitttudes, producing similar values of RMSE.

Biases at higher resolutions also improve with SKEB2 although at a lesser magnitude because numerical diffusivity decreases when resolution increases (not shown). Tennant *et al.* (2011) also showed in their Figure 11 that SKEB2 decreases the bias of weak high level winds in a former configuration of MetUM.

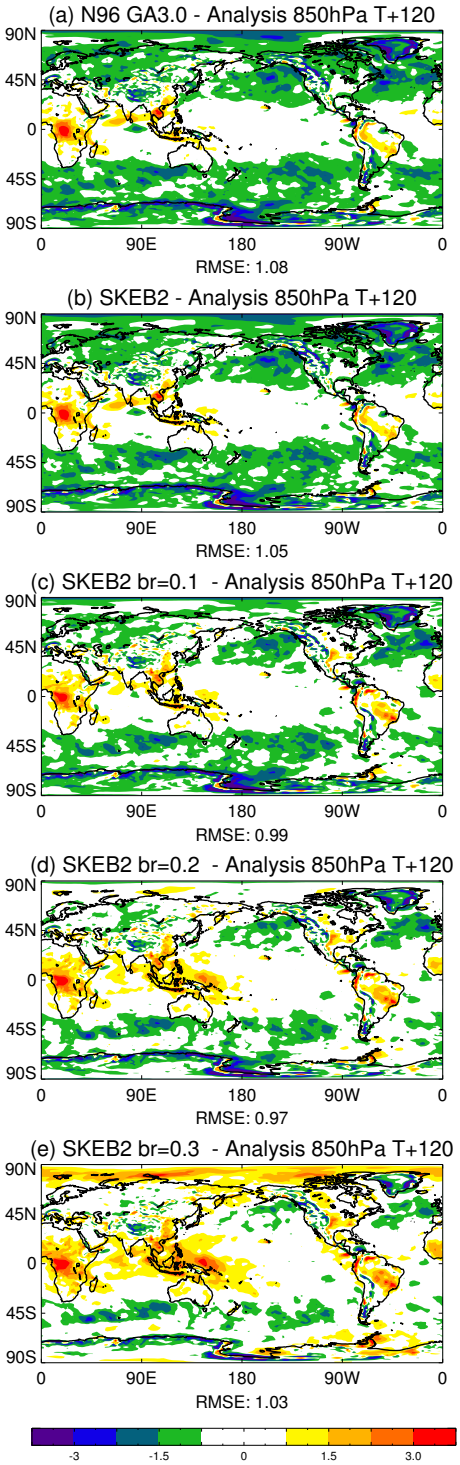

Figure 5.4: MetUM bias (model – ECMWF analysis) of the modulus of winds at 850hPa (m/s) for (a) N96 control (b) N96 SKEB2 default b_R (c) N96 SKEB2 $b_R = 0.1$ (d) N96 SKEB2 $b_R = 0.2$ (e) N96 SKEB2 $b_R = 0.3$; subtitle shows RMSE of the model averaged field versus the averaged analysis.

	NH	SH
N96 control	49.7	58.4
N96 SKEB2	51.6	60.5
N96 SKEB2 $b_R = 0.1$	57.8	65.1
N96 SKEB2 $b_R = 0.2$	65.4	72.0
N96 SKEB2 $b_R = 0.3$	72.9	79.1
N216 control	44.0	52.0
N216 SKEB2	47.2	54.5
N320 control	43.4	51.0
N320 SKEB2	46.0	53.7

Table 5.1: Geopotential at 500hPa RMSE (m) for day 5 (T+120 hours), NH and SH as defined in Figure 5.1.

	DJF	JJA
Control	0.50	0.47
SKEB2 $b_R = 0.0275$	0.47	0.45
SKEB2 $b_R = 0.1$	0.42	0.38
SKEB2 $b_R = 0.2$	0.36	0.33
SKEB2 $b_R = 0.3$	0.31	0.27

Table 5.2: Global average Anomaly Correlation Coefficient (ACC) of winds at 850hPa and T+120 for low resolution simulations with SKEB2 and increasing b_R.

5.1.3 Location / intensity errors of extra-tropical cyclones

The SKEB2's representation of mid-latitude low level winds is worse in terms of traditional skill scores as shown in the previous subsection, but on the other hand it improves the wind bias over the storm track. Features in the mid-latitudes could be in general better represented by SKEB2 but individually diverted away from their correct path. One of the main features in the mid-latitude are cyclones, commonly described as "storms", whose position and intensity can be tracked using the TRACK algorithm (see section 3.3 for a description of the algorithm). An additional method can pair tracked storms from the model to those in the analysis making possible to obtain intensity and positional errors (sect. 3.3.3). We apply this tracking algorithm and its matching technique to all our experiments. The number of storms

matched for each forecast day is shown in Table 5.3. Only storms developed before day 3 are taken into account, so after day 3 the number of storms drops as they decay and are not replaced. There are quite small differences amongst the number of storms matched for each experiment.

	NH	SH
T+24	2160	2210
T+48	2574	2603
T+72	2379	2366
T+96	1774	1698
T+120	1135	1087

Table 5.3: Number of storms matched for each hemisphere and forecast day. Numbers shown are the average amongst the different 5-day NWP experiments.

The mean intensity and distance errors of the matched storms for the experiments across different resolutions with and without SKEB are shown in Figure 5.5. The error increases with lead-time, although at N96 the intensity error saturates at day 2 after a severe drop in the forecast intensity in comparison to the other resolutions. These errors improve with resolution, and they are slightly larger in the SH in terms of position and intensity. Mid-latitude cyclones are transported by the barotropic flow which is not difficult to resolve even at coarse resolutions. In contrast, the intensity, which is strongly influenced by subgrid processes (e.g. convection or microphysics), and the implicit diffusion of the advection scheme that smooths sharp gradients of vorticity such as the storms produce. Therefore errors in the storm intensity are more sensitive to horizontal resolution than errors in storm location.

SKEB2 is beneficial increasing the mean intensity of storms, although it also introduces a positional error in all resolutions, but this error is proportionally small to the intensity improvement. The intensity increase done by SKEB2 is equivalent for both hemispheres and resolutions. There is a clear problem with the way SKEB2 modulates its perturbation across resolutions. The SKEB2 dissipation rate should be much higher and thus create a higher impact on the intensity at N96 than at N320, since dissipation is higher at lower resolutions, but results shows that the intensity increase made by SKEB2 is similar at both resolutions.

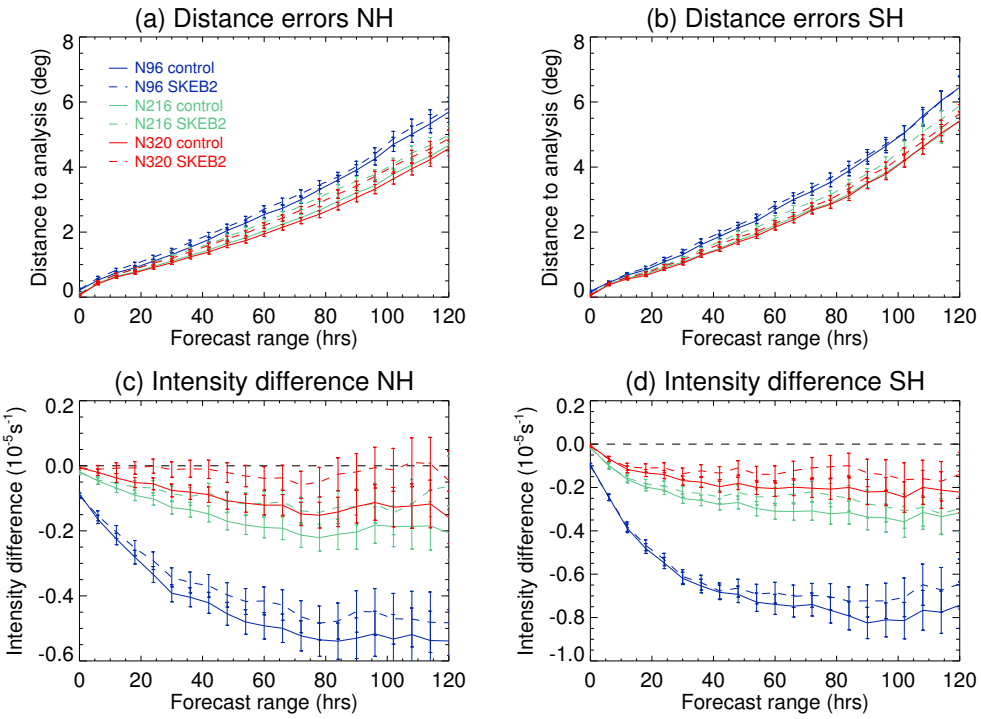

Figure 5.5: Distance (a,b) and intensity (c,d) differences of simulated storms matched to analyzed ones. Continuous lines denotes storms from control and dashed storms from the forecasts with SKEB2. Blue is for N96, green N216 and red for N320. First column (a,c) for the Northern Hemisphere and second (b,d) for the Southern hemisphere. Confidence intervals are obtained using the methodology explained in the introduction of chapter 4.

A similar plot for the low resolution experiments with the different backscatter ratios at N96 is shown in Figure 5.6. Distance errors are amplified and intensity errors reduced when b_R increases. For $b_R = 0.3$ at the NH, the storms become too active with an intensity higher than in the analysis (Fig. 5.6.c). The default amplitude factor $b_R = 0.0275$ is optimal to simulate storms at the adequate intensity in the Northern hemisphere (Figure 5.4.c) at N320, the operational horizontal resolution of current NWP models and EPS. Nevertheless, at lower resolutions this factor is too low, Figure 5.6.d indicates that the factor could be increased and the intensity in the Southern hemisphere would still be low (also shown in Figure 5.3.e over the Southern Ocean).

The absolute intensity error shows that errors growth with lead time when b_R increases, and they are always above the control (not shown). Although the ratio between the absolute intensity of the control and SKEB2 with $b_R = 0.3$ is smaller than the ratio of absolute errors in distance (Fig. 5.6).

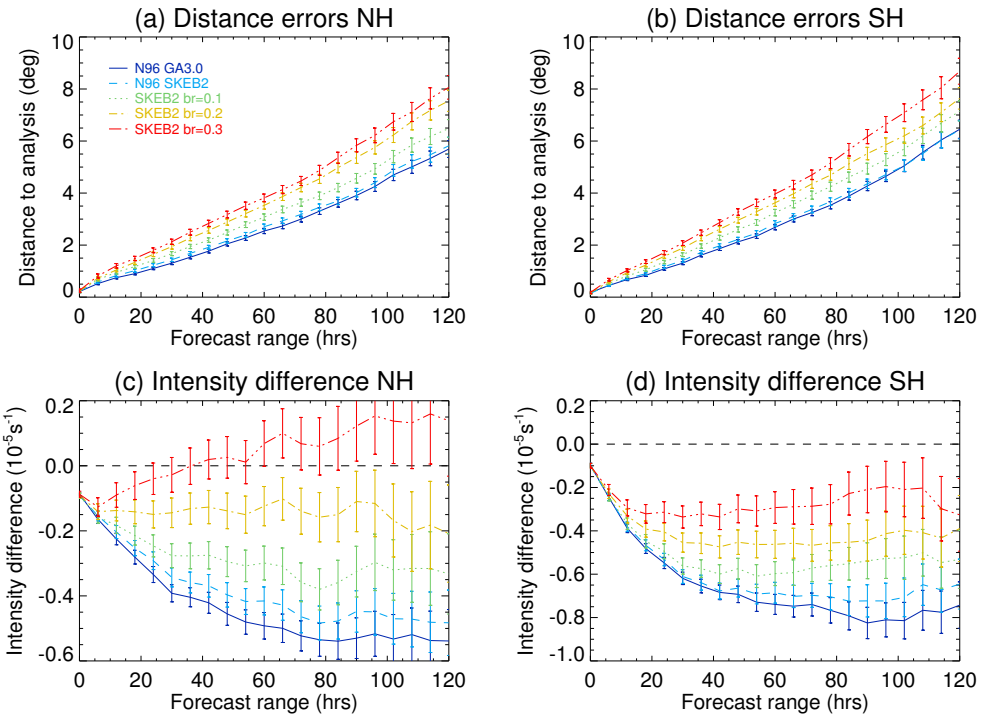

Figure 5.6: Same as Figure 5.5 but for different amplitudes of SKEB2 at N96. Dark blue is for control, pale dashed line for SKEB2 default ($b_R = 0.0275$), green dotted line for $b_R = 0.1$, yellow dash-dotted for $b_R = 0.2$ and red long-dashed dotted for $b_R = 0.3$.

The extra kinetic energy backscattered by SKEB2 could strength the divergence (convergence) at high (low) levels. This strengthening is beneficial to produce stronger vertical motions which triggers convection and latent heat release. These lead to an enhancement of the intensity of storms, as it is shown in Figure 5.5.c,d and 5.6c,d and partially fixes the problem of weak winds in the mid-latitudes (Figure 5.4). However, an excessive divergence could slow down the displacement of baroclinic systems such as storms, increasing the distance error of the storms as the simulated ones lag behind the analyzed. Slow storms would also increase the RMSE and ACC of the forecasts.

The average storm speed bias can be obtained from the difference between two successive points on the storm track divided by the time interval (6 hours). Simulated mid-latitude cyclones are generally too slow in most of the current models (Froude, 2010, 2011), SKEB2 speeds up the cyclones in the Northern hemisphere substantially (Fig. 5.7.a), such increase in the speed of the storms for the SH is so not clear (Fig. 5.7.b). The absolute error of speed shows again that errors increase when b_R increases. Mean speed rates for N216 and N320 are still low but slightly

higher than those at N96 (not shown).

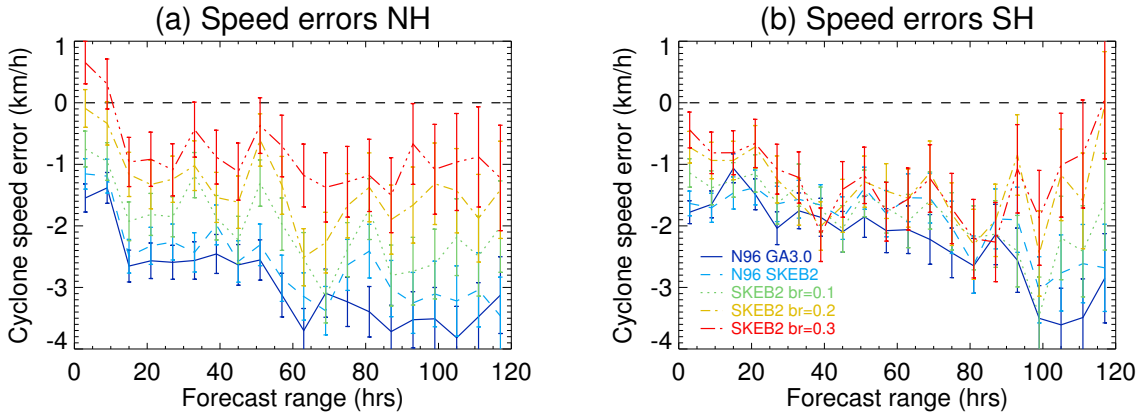

Figure 5.7: Mean propagation speed bias of storms for the low resolution experiments (km/h) (a) Northern hemisphere (b) Southern hemisphere. Same colour scale as Fig. 5.6.

Another interesting result from Froude *et al.* (2007a) and Froude (2010, 2011) is the inability of EPS to predict the fast growth rates associated with intense storms. The growth rate is the rate of intensity change of cyclones at each timestep. It does not show much variation across the lead time as it is essentially related to changes to model deficiencies in parametrizations (e.g. convection). The mean intensity growth rate is calculated from the storms population with a positive growth in both analyzed and simulated storms. Table 5.4 shows the temporal average of the mean growth rate from the initial date of the forecast until day 3, as new born storms are not tracked thereafter. The simulated storms do not grow at the same rate as the storms in the analysis do. In the Southern hemisphere the problem is deeper. The bias of mean growth rate is reduced when horizontal resolution increases, but SKEB2 is nearly as effective as resolution increases in generating growth for storms. At high b_R the growth rate bias has become positive, storms develop quicker and deeper than in reality. Froude (2010) indicated that the lack of growth is an important setback in producing quality advice of storm surges to weather services.

The results obtained from the mid-latitude cyclone tracking suggest that SKEB2 has a negative impact on individual cyclones, introducing errors on their intensity, distance and speed. However on average, the extra vorticity added by the backscatter kinetic energy alleviates the high diffusivity of the storm intensity. It also helps to speed up storms and generate higher rates of intensity growth.

Our analysis has been focus on the mean characteristics of storms, but we could

	NH	SH
N96 control	-0.46	-0.67
N96 SKEB2	-0.36	-0.60
N96 SKEB2 $b_R = 0.1$	-0.16	-0.40
N96 SKEB2 $b_R = 0.2$	0.08	-0.20
N96 SKEB2 $b_R = 0.3$	0.32	0.01
N216 control	-0.28	-0.37
N216 SKEB2	-0.16	-0.28
N320 control	-0.22	-0.29
N320 SKEB2	-0.11	-0.21

Table 5.4: Mean bias of growth rates of storm intensity ($10^{-5}\ s^{-1}\ day^{-1}$) for the different experiments and hemispheres. The rates are averaged between $T + 0$ and $T + 72$.

obtain more information looking at particular ranges of the storm intensities. For weak storms (intensity lower than 3 $10^{-5}\ s^{-1}$) there is an obvious bias in our technique: The weak modelled storms are stronger than the analyzed ones, because if the model would weaken these storms below the analyzed intensity, they would not be strong enough to be classified as a storm by the tracking algorithm.

The relation between modelled and analyzed intensities is shown in Figure 5.8, the average of intensity of simulated storms whose analyzed intensity lies between a certain range of intensities. The average of the modelled intensities for weak analyzed storms is slightly higher than the average of analyzed storms as described previously. Intense storms are weaker in the model than analysis, but these improve when resolution increases. On the other hand SKEB2 with the highest backscatter increases the storm intensity across the whole spectrum of intensities rather than at high intensities where storms are more diffused. Again this may be a side-effect of the numerical dissipation rate being unable to adequately scale up the dissipation of sharp vorticity gradients. Similar results are seen for the growth rate and speed (not shown). SKEB2 does not spin up the growth of storms at high growth rates, where the simulated growth is weaker than in the analysis. Similarly, fast storms are too slow in the model, but these are faster at high resolutions. SKEB2 is more active accelerating storms on the lower side of the speed range.

The fact that averaged storms show better behaviour than individual ones indi-

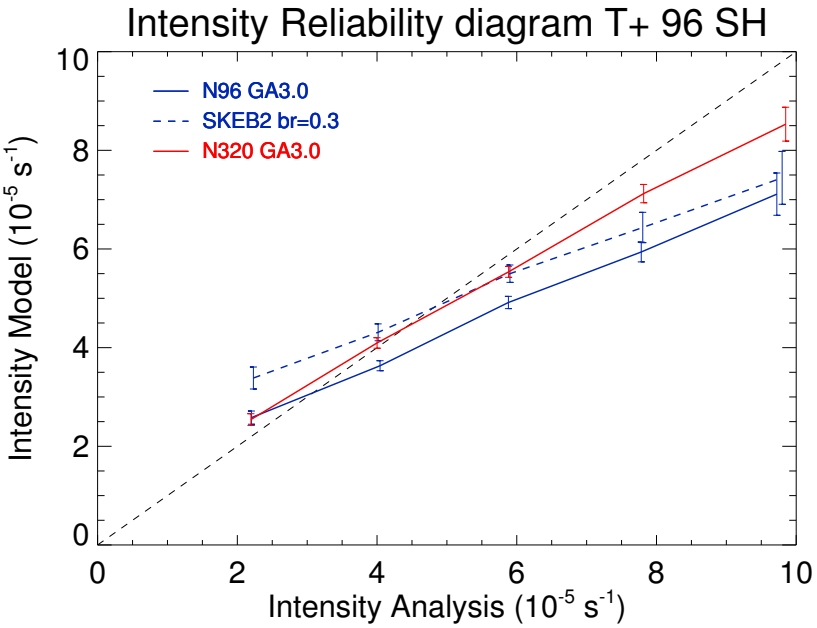

Figure 5.8: Reliability diagram for the intensity of storms. The mean intensity of simulated storms versus the mean intensity of analyzed storms within a given intensity segment of $2 \cdot 10^{-5} s^{-1}$. Confidence intervals are the standard deviation of the population divided by the square root of the length of the sample. Blue line is N96 control, dashed blue line N96 with SKEB2 $b_R = 0.3$ and red line is N320 control.

cates than the dissipation masks are doing their job, helping to increase the intensity and growth rate of storms, which are too low at all resolutions tested. However, their scaling across resolution and intensity ranges is quite bad. It does not seem to address well all true sources of dissipation of sharp vorticity gradients. The numerical dissipation may have a bigger contribution from the damping done by the excessive off-centering than the interpolation to the departure point. The nature of the off-centering damping could be constrained to divergent modes rather than rotational modes. However the SKEB2 scheme has a divergent component too, although it may have different spatial and temporal timescales than the effects of eddies damped by the Semi-Implicit scheme.

Despite that the representation of the numerical dissipation is not perfect, the model exhibits a large drop in the storm intensity (Fig. 5.5c,d) on the first two days of the forecast and then it seems to saturate, so there is a true need to backscatter some kinetic energy and, in principle, that should not degrade the simulation of an individual storm. Hence the spotlight goes onto the scales chosen to inject the forcing. In a perfect SKEB scheme the energy would be injected on the right scales,

and later upscaled towards larger scales. Therefore the predictability of the cyclone would not be affected so much as it does with SKEB2. An excessive backscatter ratio, as those employed here, might be adding too much energy on scales that have little dissipation, degrading the forecast and its representation of storms.

Another problem may come from the excessive smoothing of SKEB's dissipation masks (see sect. 2.2). Its physical justification is aimed to account for the spatial uncertainty of phenomena such as cyclones or fronts. Adding kinetic energy stochastically on points adjacent to these system tells the model that maybe there is a contribution of this system on this gridpoint. This representation is perhaps a bit coarse and could help these systems to change their "deterministic" trajectories towards a less realistic path.

5.2 Results of low resolution climate simulations

When we include stochastic variability into a physical model, there might be a noise-induced drift towards a different mean state. The stochastic perturbations push the system away from its own attractors, increasing the frequency of rare events poorly modelled by state-of-the-art models, e.g. the effects of CABS on weather regime frequency shown in Jung *et al.* (2005a) and described in sect. 2.5. Therefore the stochastic schemes may help to produce a better representation of variability and mean climate.

In addition to the noise-induced transitions, if the stochastic noise is strong and targeted to the drivers of variability at the timescales of our simulation, the stochastic physics schemes should also produce different outcomes for their climate mean, or in other words, increase the ensemble spread on climate scales.

In order to estimate the capacity of SKEB2 to improve the mean climate and variability of MetUM and generate spread at climate timescales. We have set up the experiments shown in table 5.5 with an ensemble of SKEB2 and other simulations with increasing b_R to study how the scheme modifies the climate when the SKEB2's forcing increases.

The two ensembles of SKEB2 with different b_R are compared with an ensemble of climate simulations of the climate model's predecessor, the Hadley Centre Global Environment Model v2 (HadGEM2, Collins *et al.* 2011). HadGEM2 ensemble consist

Name of the experiment.	No. of ensemble members
Control	1
SKEB2 default	5
SKEB2 $b_R = 0.1$	3
SKEB2 $b_R = 0.2$	1
SKEB2 $b_R = 0.3$	1

Table 5.5: Description of different 20 year climate simulations and ensembles with different b_R for SKEB2

of 7 members with initial perturbations to soil variables, in addition two members swap their soil variables at the start of each month.

There is an important difference between the model we have used for this climate studies and the one we used for NWP deterministic prediction. The climate system employs a quasi-cubic interpolation scheme to the departure point for the advection scheme (see sect. B.2.1), whereas the short-range prediction runs had cubic scheme. Quasi-cubic is less diffusive although less accurate with gradients (Woods N, 2011, *personal comm.*), therefore the counter-diffusion properties of SKEB2 may be masked by the decrease of the internal diffusivity done when changing the interpolation scheme from Cubic to Quasi-cubic.

5.2.1 SKEB2 capacity to generate climate spread

In order to quantify the capacity of the SKEB2 scheme to generate spread at climate timescales, the ensemble of climate simulations with SKEB2 is compared to the HadGEM2 ensemble. There are differences on the mean climate amongst the five simulations of the SKEB2 ensemble. This ensemble is built of simulations whose only differences are produced by SKEB2 with the default backscatter ratio value of 0.0275 for all members. An additional ensemble of 3 members with $b_R = 0.1$ is also used and exhibits different mean states.

The differences found amongst different climate simulations with SKEB2 are significant for several variables, e.g: Z_{500}, low level (850hPa) winds or temperature. However, these differences are far from the model bias and could be produced by the internal variability of the model rather than driven by SKEB2. In order to estimate

the significance of these differences, a statistical f-test is used with the global and hemispheric mean of the standard deviation of the SKEB2 ensemble against the HadGEM2 ensemble. Results show low percentages of significance, quite below the acceptable levels of 90% or 95%. Thus the spread of SKEB2 ensembles with default b_R or $b_R = 0.1$ is not statistically different than the spread in the HadGEM2 ensemble, which has no stochastic physics scheme. Therefore the capacity of SKEB2 to create spread at climate scales is quite small, although the number of members of the ensemble used is quite low. It would be desirable to repeat the experiment with more members if computing resources were available.

The spread of the mean climate amongst the different members of SKEB2 with default b_R slightly increases in comparison to the spread generated by the HadGEM2 ensemble (Figure 5.9), it is larger in DJF than JJA. The differences are located in the Storm track regions of the Southern Ocean, North Pacific and North Atlantic. Both seasons and ensembles show the same areas of maximum spread. In addition to the storm tracks, the West Indian ocean during boreal summer is also a region of large low level wind variability due the passage of active cells of the Madden Julian Oscillation (MJO, see sect. 3.6 for a description) and other tropical disturbances, but the models used (MetUM and HadGEM2) are not able to generate differences in the climate mean over this region.

Differences in spread at climate scales for Z_{500} are also quite reduced, Table 5.6 shows its standard deviation of wintertime on the extra-tropical hemispheres for SKEB2 def. b_R and HadGEM2 ensemble. The standard deviations of SKEB2 and HadGEM2 are not are significantly different even at 90% level.

	NH DJF	SH JJA
SKEB2 b_R def.	63.6	62.0
HadGEM2	63.9	60.

Table 5.6: Z_{500} standard deviation of the ensembles (m). See text for details

A deeper analysis of the low level wind differences amongst SKEB2 ensemble members, control run and ERAI reanalysis is presented in Figure 5.10. It shows for the different experiments the standard deviation, confidence interval and climate mean averaged over the boxes shown in Figure 5.9. These boxes are situated over

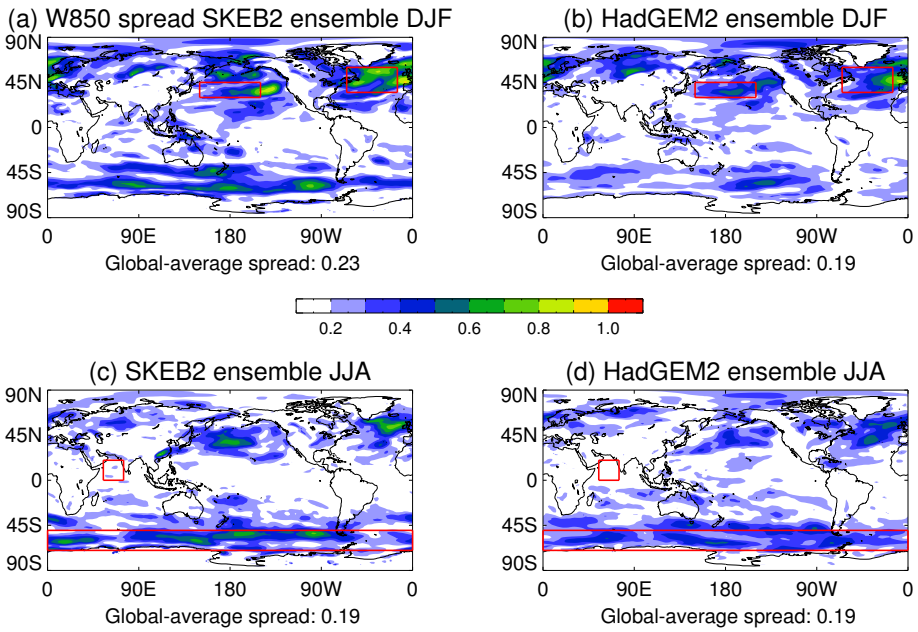

Figure 5.9: Winds (m/s) at 850hPa standard deviation of the mean climate amongst SKEB2 ensemble of default b_R (a) and (c), HadGEM2 ensemble (b) and (d). Boxes denote the regions shown in Figure 5.10. Subtitles show the Global mean spread.

regions with large climate spread in the ensemble plus the West Indian Ocean region. The main results drawn from Figure 5.10 are:

- The low level winds in regions of high storm activity in the Northern Hemisphere, like those in the boxes of Figure 5.10.a and 5.10.b are well simulated by MetUM. SKEB2 adds little value in these regions. The SKEB2's spread on climate scales is over-dispersive producing members that simulate average winds outside the confidence interval of ERAI, like the first member of the SKEB2 default b_R ensemble for the North-Atlantic region in the boreal winter (Fig. 5.10.a). When the scheme is amplified, the differences in the spread across different SKEB2 simulations are not noticeable for these regions in the NH.

- Low level winds over the Southern Ocean (SO) in the austral winter (JJA) are too low in MetUM (Figure 5.10.c). This is an important bias of many GCMs. The addition of SKEB2 slightly improves the simulated winds, as the value of some of members is outside the confidence interval of the control run and between reanalysis and control, like the 2^{nd} and 4^{th} members of the b_R default ensemble and 3^{rd} member of the $b_R = 0.1$ ensemble. But there is no additional

benefit when there is an increase of the energy backscattered driven by b_R.

- For the case of the West Indian Ocean, where the Asian monsoon occurs in JJAS, there is a large standard deviation of low level winds. The mean value simulated by MetUM is again too low. This is an important systematic bias of many models, likely caused by a poor representation of convective processes. SKEB2 ensemble spread on climate scales is very low, as seen on Figure 5.9.c and thus the members of the SKEB2 b_R default ensemble are almost indistinguishable. Nevertheless, when the backscatter increases, there is a climate shift and the averaged winds produced by the model within the box increases and for $b_R > 0.1$ the model goes beyond the reanalysis value.

5.2.2 SKEB2 improvements in the mean climate

Some positive results of the SKEB2 performance on climate scales have already been shown in Figure 5.10. The scheme increases low level winds over the Southern Ocean region and over the West Indian Ocean. Winds in the latter region seems to be substantially affected by SKEB2. Its perturbation can flip the sign of the bias when the backscatter ratio becomes too strong.

A different mean state produced by SKEB2 with increasing backscatter ratio would be evident if their global skill scores such as the RMSE were different with and without the scheme. The RMSE is a well suited metric for a preliminary assessment of the global effects of the scheme in different model fields, despite its disadvantages (as described in section 3.1). Table 5.7 shows the global RMSE of the control and SKEB2 experiments for large-scale variables such as PMSL or Z_{500}. Overall the impact of the scheme is neutral to positive. At the largest amplitudes results are positive for nearly all seasons and variables. There is not a preferred season where the model's simulation is better with SKEB2.

The RMSE of the climate mean of dynamic variables such as winds is shown in Table 5.8, SKEB2 improves the RMSE for winds at 850 and 200 hPa levels for all seasons. These improvements seem stronger in JJA than in other seasons for both levels, there are no proportional improvements in RMSE when the backscatter ratio is increased, maybe an excessive b_R could flip the side of the bias whereas a smaller

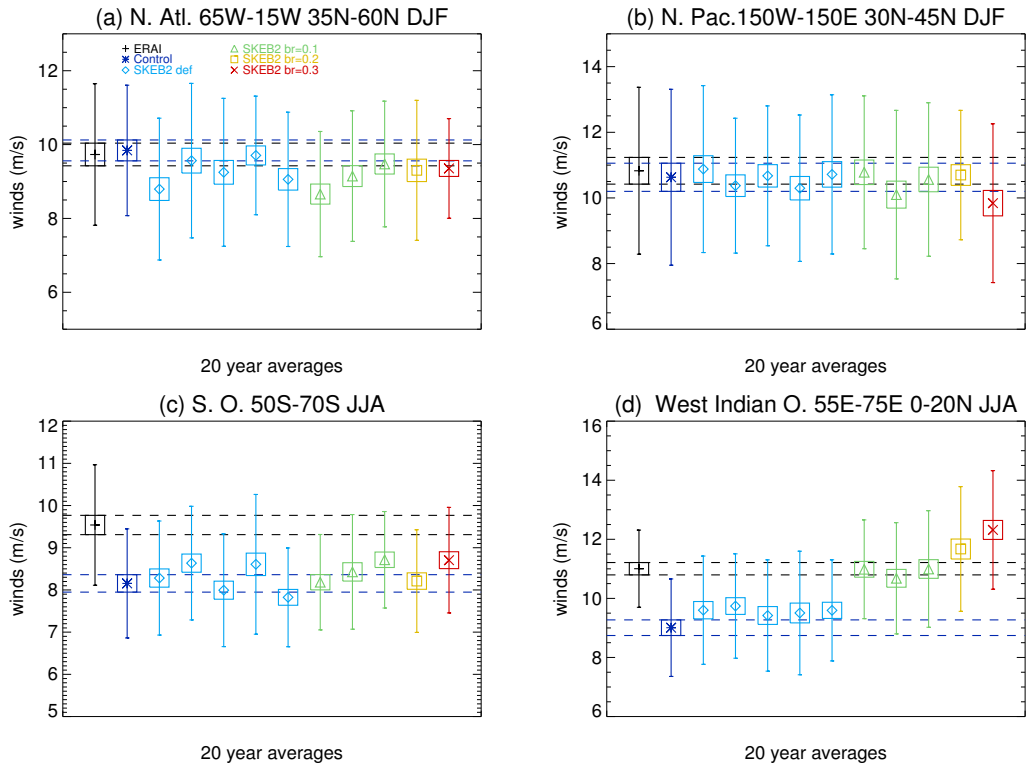

Figure 5.10: Average values (symbols), standard deviation (bars) and confidence intervals (boxes) of monthly mean winds at 850hPa for all the different climate runs plus ERAI reanalysis, period from Jan 1989 to Dec 2001 (a) North Atlantic section (65W-15W, 35N-60N) for DJF (December, January and February) (b) North Pacific (150W-150E, 30N-45) for DJF (c) Southern Ocean (whole longitudinal domain, 50S-70S) for JJA (June, July and August) (d) West Indian Ocean (55E-75E, 0-20N) for JJA. Black with pluses is for ERAI, dark-blue with asterisk for control, pale blue with diamonds for the ensemble of SKEB2 default amplitude, green with triangles for the ensemble of SKEB2 with $b_R = 0.1$; yellow with squares $b_R = 0.2$ and red with crosses for $b_R = 0.3$ scheme backscatter ratio. Dashed lines denote the confidence interval of control and ERAI (see introduction of chapter 4).

	Control	SKEB2 $b_R = 0.0275$	SKEB2 $b_R = 0.1$	SKEB2 $b_R = 0.2$	SKEB2 $b_R = 0.3$
PMSL DJF	2.08	2.12 (-2%)	2.21 (-6%)	1.84 (+11%)	1.86 (+10%)
PMSL MAM	2.10	2.04 (+2%)	2.18 (-3%)	2.12 (0%)	2.04 (+2%)
PMSL JJA	2.66	2.63 (+1%)	2.52 (+5%)	2.58 (+2%)	2.65 (0%)
PMSL SON	1.98	1.92 (+3%)	1.82 (+7%)	1.81 (+8%)	1.80 (+9%)
Z_{500} DJF	2.96	2.91 (+2%)	3.13 (-4%)	2.69 (+9%)	2.76 (+7%)
Z_{500} JJA	2.71	2.77 (-2%)	2.62 (+3%)	2.58 (+4%)	2.54 (+6%)

Table 5.7: Global RMSE scores for different synoptic metrics like PMSL (hPa) and Z_{500} (dam) against ERAI, in brackets there is the percentage of improvement to control: (RMSE(exp) - RMSE (control)) / RMSE (control), positive (negative) percentage implies improvement (degradation).

value would reduce it, as shown in Fig. 5.10 for low level winds in the West Indian Ocean.

The energy backscatter by SKEB2 produces fluctuations in the large-scale flow, affecting the parametrization output. Table 5.9 shows the RMSE of fields related to the convection, radiation and condensation parametrizations such as total precipitation and Outgoing Longwave Radiation (OLR) at the Top of the Atmosphere (TOA). Again we see a clear improvement for JJA that scales up with increasing b_R even at large values. Other seasons show neutral to slightly positive results and do not seem to be affected by the different amplitudes of SKEB2 perturbations.

As described in section 3.1, RMSE could hide double penalty errors and it is hard to interpret what process are behind the improvements or degradations of the score. A further analysis is needed to understand the course of the SKEB2 induced changes seen in RMSE. On the following subsections we investigate low level wind improvements in JJA and OLR improvements for the longer tropical season JJAS.

Mean low level wind improvements in JJA

The different members of SKEB2 ensembles have different climate averages for low level winds, as shown in Figure 5.9. Therefore the RMSE shown in tables 5.7, 5.8 and 5.9 could be an extreme value of the ensemble rather than the most likely. There could also be compensating errors amongst the different regions. A good

	Control	SKEB2 $b_R = 0.0275$	SKEB2 $b_R = 0.1$	SKEB2 $b_R = 0.2$	SKEB2 $b_R = 0.3$
W850 DJF	1.41	1.37 (+3%)	1.40 (0%)	1.32 (+6%)	1.32 (+6%)
W850 MAM	1.33	1.30 (+2%)	1.32 (0%)	1.30 (+3%)	1.29 (+3%)
W850 JJA	1.66	1.60 (+4%)	1.56 (+6%)	1.61 (+3%)	1.58 (+6%)
W850 SON	1.43	1.38 (+4%)	1.41 (+2%)	1.38 (+4%)	1.36 (+5%)
W200 DJF	3.05	2.84 (+6%)	3.07 (0%)	3.02 (+1%)	3.16 (-3%)
W200 MAM	2.65	4.45 (+7%)	2.22 (+16%)	2.24 (+10%)	2.50 (+5%)
W200 JJA	4.40	4.39 (0%)	3.81 (+13%)	3.97 (+9%)	3.91 (+11%)
W200 SON	2.76	2.76 (0%)	2.69 (+2%)	2.80 (-1%)	2.79 (0%)

Table 5.8: Global RMSE scores for winds at 250 and 850hPa against ERAI (m/s). Same format as Table 5.7.

	Control	SKEB2 $b_R = 0.0275$	SKEB2 $b_R = 0.1$	SKEB2 $b_R = 0.2$	SKEB2 $b_R = 0.3$
Precip DJF	1.39	1.35 (+2%)	1.39 (0%)	1.35 (+2%)	1.38 (0%)
Precip MAM	1.26	1.23 (+2%)	1.21 (+3%)	1.24 (+1%)	1.28 (-1%)
Precip JJA	1.87	1.80 (+3%)	1.63 (+12%)	1.60 (+14%)	1.60 (+14%)
Precip SON	1.34	1.34 (0%)	1.31 (+2%)	1.29 (+3%)	1.32 (+1%)
OLR DJF	10.2	9.99 (+2%)	10.5 (+1%)	9.96 (+3%)	9.95 (+3%)
OLR JJA	11.5	11.2 (+3%)	9.85 (+15%)	9.45 (+18%)	9.32 (+19%)

Table 5.9: Global RMSE scores for variables heavily influenced by physical parametrizations like precipitation to GPCP (mm/day) or OLR at TOA to CERES (W/m^2). Same format as Table 5.7.

performance in the tropics could be masked by a poor performance in the mid-latitudes.

The ratio of RMSE between all the members of the SKEB2 ensembles and the control run is shown in Figure 5.11. Different subplots show different seasons for low and high level winds and each subplot shows different levels and regions. The improvements at the low level winds are more localized in the Tropics and SH, whereas NH shows a more neutral performance or even negative for the case of MAM (Fig. 5.11.b). For high level winds, the errors and spread of the mean climate of the different simulations are larger than for the low level winds. Winds at 250hPa depict the position of the jet stream in each hemisphere. A slight displacement in its representation could produce large increases of RMSE. There is not a clear relation of an increase or decrease of the RMSE ratio when the amplitude of the SKEB2 increases (with the exception of the tropical high level winds in JJA).

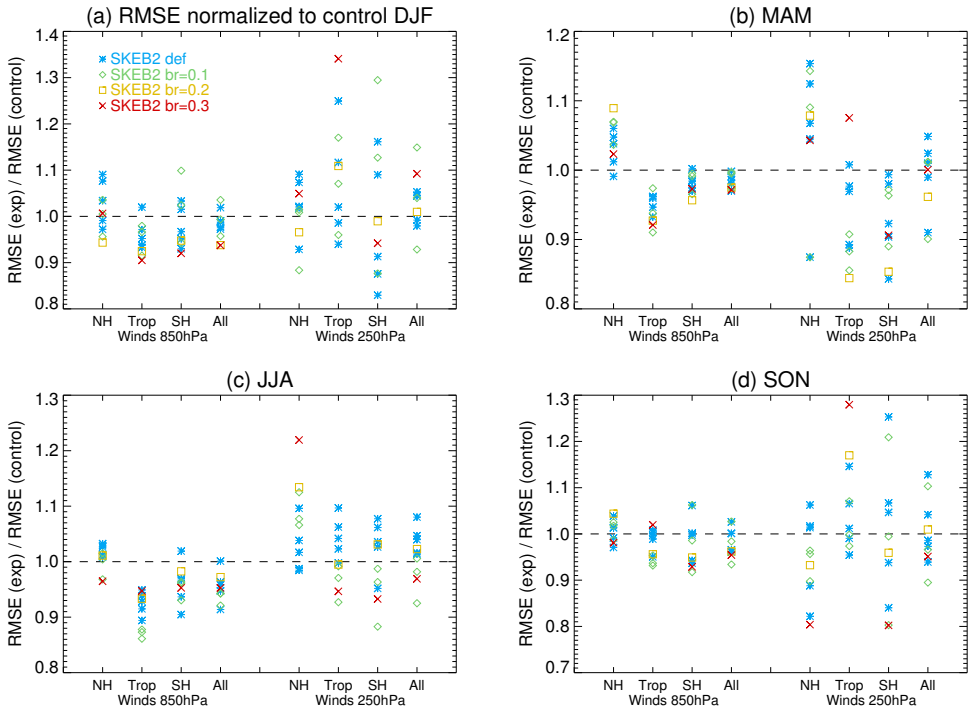

Figure 5.11: RMSE ratio between SKEB2 experiments and control for winds at two different levels, colour scale follows same classification as Figure 5.10. SKEB2 default amplitude is represented pale blue asterisks; SKEB2 with b_R=0.1 with pale green diamonds; SKEB2 b_R=0.2 and a yellow square and SKEB2 b_R=0.3 with red crosses. Globe regions corresponds to NH: 90-30N, Trop: 20N-20S, SH 30-90S and All is the full globe.

The spatial structure of the low level wind bias for JJA, the season with the

SKEB2's largest impact, is shown in Figure 5.12. The MetUM biases are larger for the Southern Ocean where the storm track is displaced equatorwards, or more precisely, the equatorward side of the storm tracks is too strong whereas the poleward side is too weak. Northeast Asian monsoon winds are too weak over the West Indian Ocean. South Atlantic Trade winds are slightly misplaced equatorwards over the Guinea gulf. Biases over Antactica and Greenland could be related to a different methodology to derive the pressure levels between model and observations and are thus ignored.

When SKEB2 is included in the model, it increases (decreases) low level winds on the equatorward (poleward) side of the Southern Ocean track, improving its representation for all seasons (Fig. 5.12). It also increases winds over the Indian Monsoon region in agreement with Figure 5.10. When the backscatter ratio increases, biases in the West Indian Ocean are reduced and flips the sign of the bias as seen in Figure 5.10.d (Fig. 5.12.c,d,e,f). Winds also increase in the Maritime continent and central Pacific, degrading the mean climate.

On the high levels, winds over the Indian Monsoon region during JJA move westwards, but too slow in the model. SKEB2 accelerates winds over this region, more strongly over the Maritime continent, reducing the bias but making winds in the West Pacific too strong. It also decelerates winds of the SH jet stream over South America for JJA, improving the model in the East Pacific (not shown).

The described improvements in low level winds are likely caused by changes in the intraseasonal variability. Sections 5.2.3 and 5.2.4 describe the investigation of the intraseasonal variability and details its realism for the areas under study: Southern Ocean, West Indian Monsoon and Maritime continent.

OLR at TOA biases

MetUM produces too much upper level divergence in key tropical areas such as the Pacific and Atlantic basins of Central America and equatorial Indian Ocean (not shown). Divergence is associated with deep convection. OLR at TOA is a good proxy for convection, as the convection triggers and maintains cumulonimbus. These clouds trap the Long-Wave (LW) radiation from the Earth's surface, thus LW rad. is not transmitted to the space, therefore regions with low OLR measured from satellites are convectively active in the low latitudes. The excessive divergence of

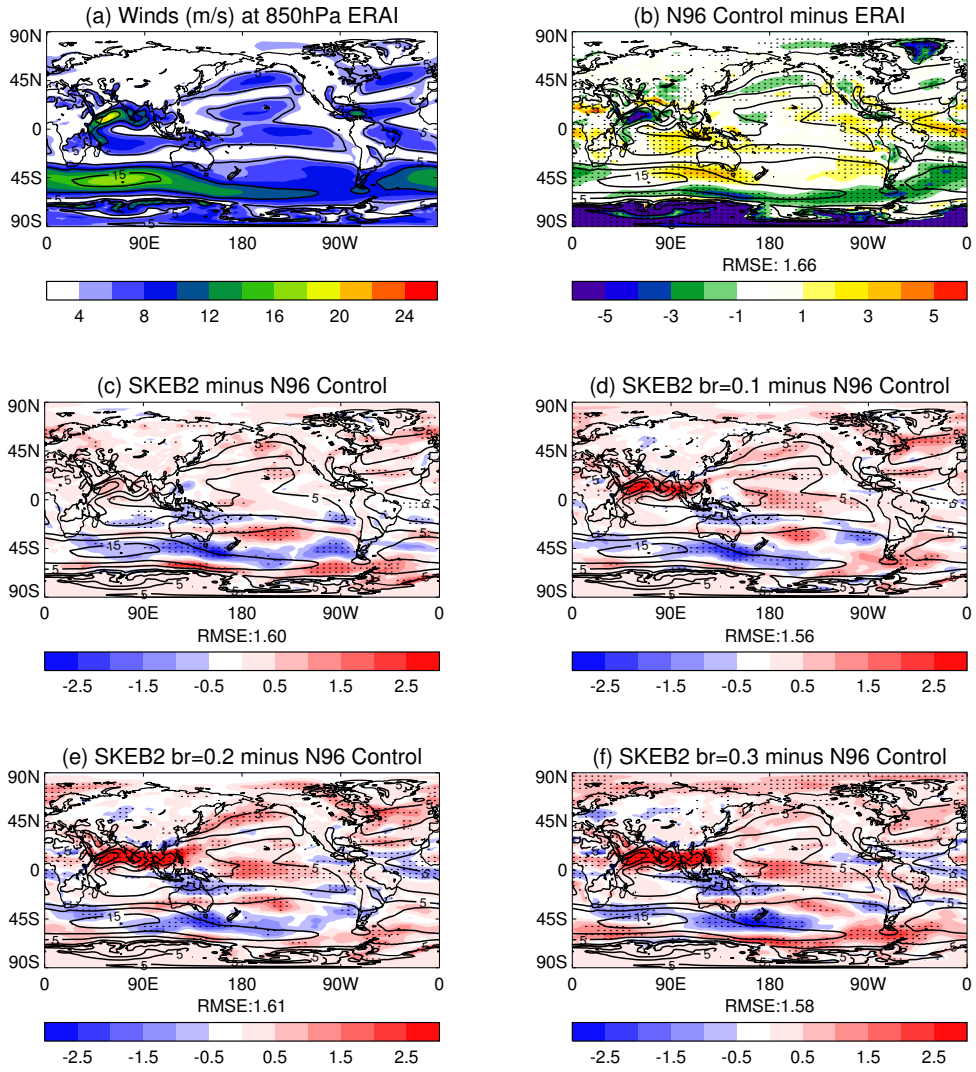

Figure 5.12: Modulus of winds at 850hPa (m/s) in boreal summer (JJA). (a) ERAI reanalysis averaged winds during the interval Jan 1989-Dec 2001. (b) Control biases to ERAI. (c) One member of the SKEB2 with default $b_R = 0.0275$ ensemble minus Control. (d) One member of SKEB2 with $b_R = 0.1$ ensemble minus Control. (e) SKEB2 with $b_R = 0.2$ minus Control. (f) SKEB2 with $b_R = 0.3$ minus control. Dotted areas denote significance above a 95% level using a t-student test. Contours show ERAI reanalysis winds each 5 m/s. Subtitles show RMSE of the climate mean field versus ERAI.

MetUM is associated with too much convection and thus thicker clouds with lower OLR. Figure 5.13.a shows the OLR biases of the control run to CERES (described in sect. 4.4). There is too much convection over the Inter-tropical Convergence Zone (ITCZ) in both sides of Central America, over the West Pacific Ocean and too little convection over the Indian subcontinent and Maritime Continent.

All simulations with SKEB2 reduce the OLR biases over these high-convective areas. This reduction seems to be proportional to the b_R backscatter ratio (Fig. 5.13), it also seems to erase the dipole on the Indian monsoon, where there is too much convection on the West Indian Ocean and too little over the Indian subcontinent and Maritime continent. This could be related to the improvements in low levels winds shown in Figure 5.12. Although SKEB2 induced improvements look very positive, the default value for the backscatter ratio produces barely significant differences to control (Figure 5.13.b), thus it should be increased to see a more beneficial impact in the tropics.

Similar improvements over the same regions are also observed for the standard deviation of OLR at TOA (not shown). The model produces excessive variability of OLR over the Tropical Indian Ocean and both sides of central America, and too little over the North-East side of the Indian subcontinent (similar pattern as Fig 5.13.a). SKEB2 stochastic perturbations reduces the excessive variance over those regions, having a variability pattern more in agreement with CERES.

The Precipitation field also shows a large improvement in the JJA season (Table 5.9). Precipitation mostly occurs in the tropics (see Figure 4.3), where it is driven by convective processes, as water condensates on upgrdaught plumes and precipitates heavily. Figure 5.14 shows the precipitation field and changes to GPGP of the experiments done. The control experiment shows excessive precipitation over the same regions where OLR biases are located (Figure 5.13). SKEB2 with increasing b_R reduces the precipitation over these regions. Additionally, precipitation over the West Indian Ocean extends towards the west coast of India, reducing the dry bias there. There is also a degradation in the precipitation field on the West Pacific where it becomes too strong.

The notable improvements found the Indo-Pacific region for the boreal summer could be related to many different causes, it is necessary to look at the intraseasonal variability to discern some possible mechanisms which could lead to such improve-

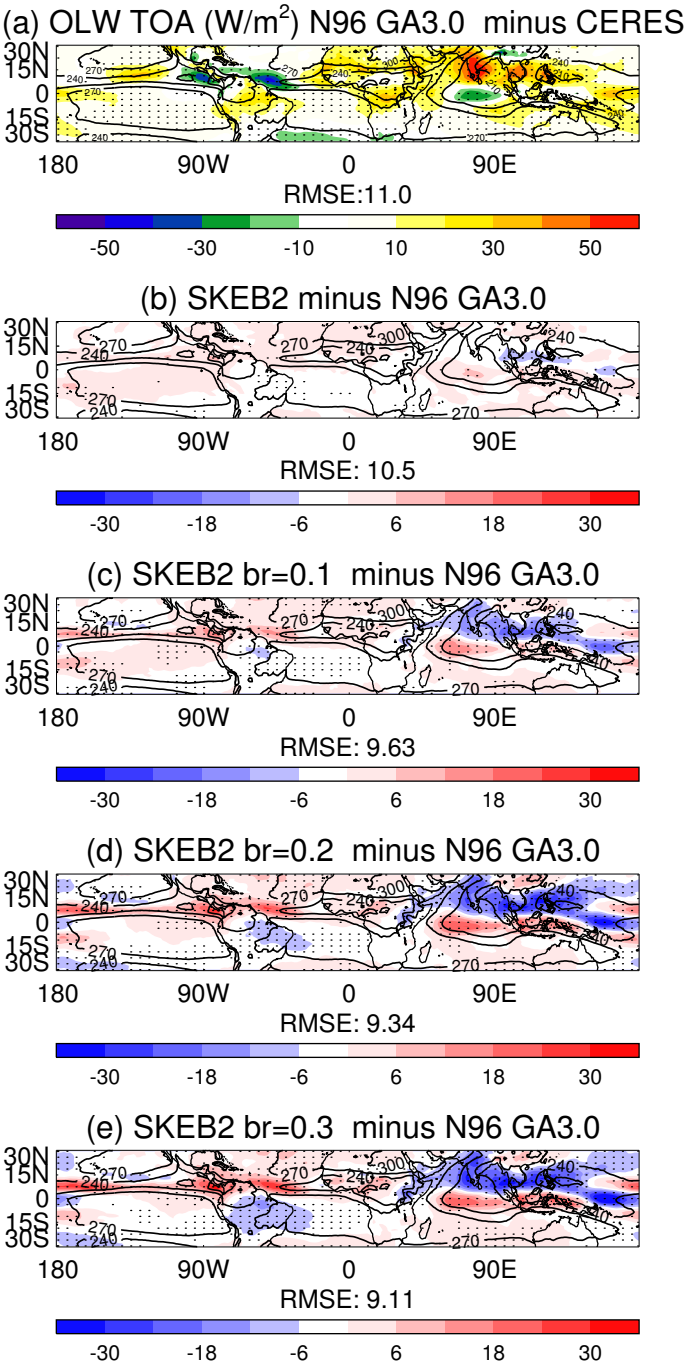

Figure 5.13: Outgoing Longwave Radiation (OLR) at the Top Of the Atmosphere (TOA) for JJAS. (a) Control minus CERES (b) SKEB2 Ensemble mean with default b_R=0.0275 minus control, (c) SKEB2 Ensemble mean with default b_R=0.1 minus control, (d) SKEB2 with default b_R=0.2 minus control, (e) SKEB2 with default b_R=0.3 minus control. Dotted areas denote statistical significance above a 95% level using a t-student test. Contours show CERES values each 30 W/m^2. Subtitles show RMSE of the climate mean to CERES.

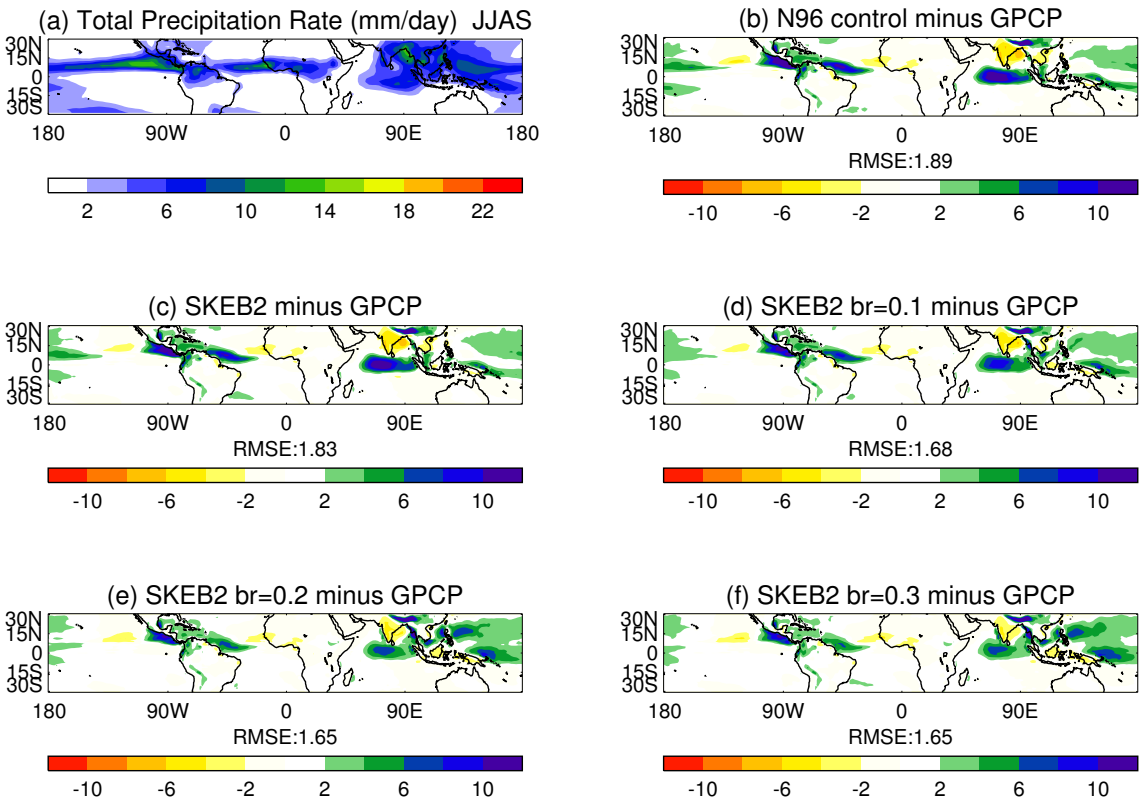

Figure 5.14: Precipitation fields (mm/day) for (a) control. (b) Differences control minus GPCP. (c) SKEB2 def. b_R minus GPCP. (d) SKEB2 def. b_R =0.1 minus GPCP. (e) SKEB2 b_R =0.2 minus GPCP. (f) SKEB2 b_R =0.3 minus GPCP. Subtitles show RMSE to GPCP

ments. In section 5.2.4 there is an investigation of the tropical variability in terms of CCEW's power spectra, MJO diagnostics, and the spectral decomposition of precipitation at particular locations.

5.2.3 Mid-latitude variability

The main drivers of the mid-latitude climate are synoptic and large-scale processes such as cyclones or blocking events. The improvements seen on the averaged Z_{500} in both hemispheres (Table 5.7) or low level winds over the Southern Ocean (Figure 5.12) are probably linked to a better representation of some aspects of these processes. In order to estimate the effects of SKEB2 perturbations on them, the TRACK algorithm (section 3.3.1) is employed to track mid-latitude cyclones and obtain statistics from the climatological sample, in combination with 2D blocking frequency indexes (section 3.4).

The clustering of weather regimes was also explored. Following Jung *et al.*

(2005a), a number of k clusters were computed for the North-Atlantic and North-Pacific using the clustering algorithm described in Fereday *et al.* (2008). There is little significance amongst the clusters for daily PMSL for different k numbers. Members of one cluster's population were not significantly different from other members in other clusters. Therefore the frequency of occurrence is not reliable. Dawson *et al.* (2012) have shown that a low resolution model produces clusters of low significance, in contrast to the higher resolution model where the significance is clear and cluster resembles those found in a combination of ERA40 and ERAI reanalysis.

Storm tracks

The average density of storms per season is computed as described in section 3.3.2. The boreal winter (DJF) storm track density is shown in Figure 5.15. There are two preferred regions for Storm tracks as depicted in Fig. 5.15.a. These are the North-Atlantic, where storms generated in the eastward lee of the Rockies cluster together across Labrador and Newfoundland on their way to Eurasia, dying right before reaching the North Urals in Siberia. Storms emerge also from the northwards lee of the Himalayas and East China and go across the North Pacific Ocean through Japan, dying in the West Coast of North-America. The Mediterranean Sea is another region with a noticeable number of storms.

The control simulation of MetUM simulates few storms in the majority of the areas with high density, such as Newfoundland, North of Urals and East China. It also generates too many storms in the Mediterranean (see Hoskins and Hodges 2002 and Froude 2010 for a general description of model biases in the representation of storms). SKEB2 has a negative effect on the deviation of individual storms from the analyzed path for NWP results (sect. 5.1.3). However, it does not affect negatively the regional distribution of storms as the main biases of the model remains unchanged with the inclusion of SKEB2. On the other hand the track density does not benefit too much from the high backscatter improvements seen in other regions such as the Tropics (Figure 5.10.a, 5.10.b).

SKEB2 could produce some positive changes in the Southern Ocean low level winds (Figure 5.10.c and Figure 5.12). The scheme's perturbations displace the SH storm track polewards in some of the members of the ensemble. Figure 5.16 shows the track density for JJA in the Southern Ocean for ERAI, control and simulations

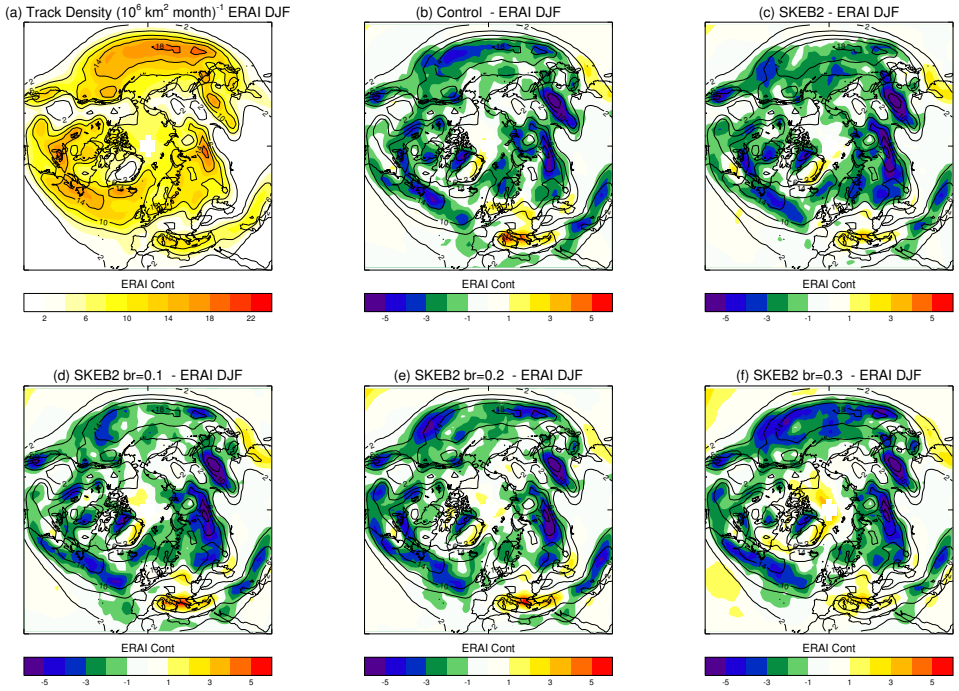

Figure 5.15: Northerm Hemisphere DJF track density of storms within an area equal to 10^6 km^2 per month. (a) ERAI (b) Control minus ERAI (c) SKEB2 default b_R - ERAI (d) SKEB2 $b_R = 0.1$ - ERAI (e) SKEB2 $b_R = 0.2$ - ERAI (f) SKEB2 $b_R = 0.3$ - ERAI. Contours show ERAI values for every 2 storms within 10^6 km^2 per month.

with an increasing SKEB2 amplitude. There is an area of high density of storms by the edge of the Antarctic coast over the Pacific side of the Southern Ocean. MetUM simulates poorly the concentration of storms over this area, underpredicting storm density by a factor of 1/4. Some simulations with SKEB2 could reduce this bias, but on the other hand it also decreases the storms south of the African continent, degrading the model.

In terms of intensity, SKEB2 with increasing b_R shows that it can minimize the bias of low storm intensities at NWP scales, improving the averaged low level winds over the mid-latitudes (Figures 5.4 and 5.6). In a climate context, this improvement should be lower as it uses a less diffusive interpolation scheme to the departure point in the Semi-Lagrangian scheme (described in sect B.2.1). The intensity of the storms in climate simulations against the ERAI storms is shown in Figure 5.17. The control simulation shows that model weakens storms in areas of strong storms in NH for DJF (Fig. 5.17.a) such as North of Iceland, central North Pacific, and Southern Ocean for JJA (Fig. 5.17.f). SKEB2 gradually decreases the lack of storm intensity

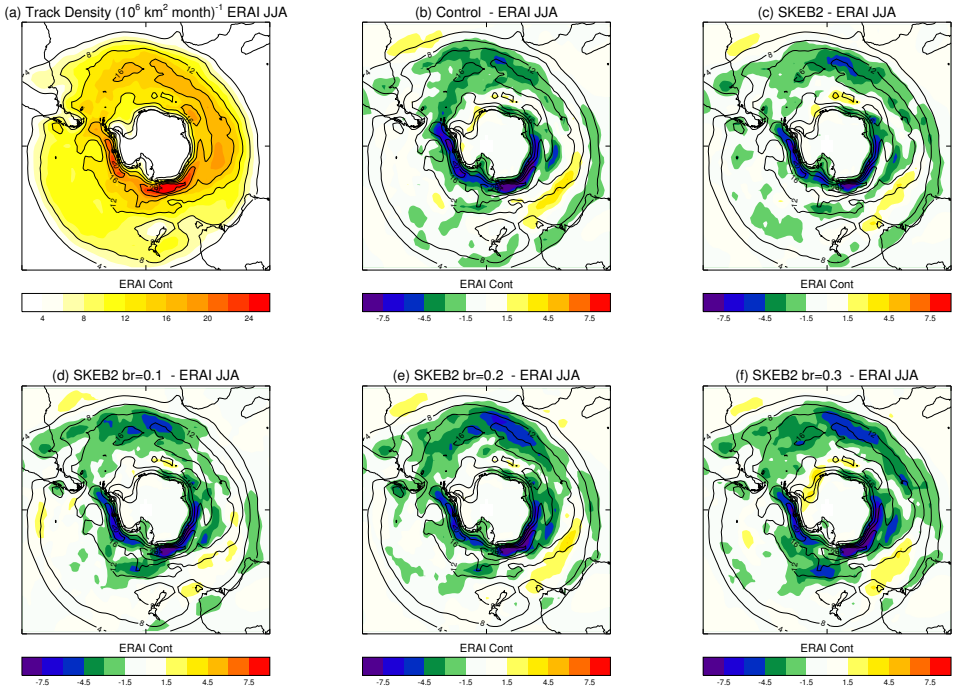

Figure 5.16: Southern Hemisphere JJA track density of storms. Same as Figure 5.15.

over these regions and flips its sign when the backscatter increases, generating too strong storms in the North Pacific on the East side of Japan for DJF in the NH, or the Drake Passage in the SO for JJA.

Blocking

Many GCMs underpredict the frequency of blocking events, and some studies attribute this problem to the lack of small-scale eddies that help to sustain this large-scale phenomena (see section 3.4). The stochastic forcing of schemes such as SPBS or SKEB2 has been demonstrated to lead to a higher frequency of blocking events over the North Pacific (see sect. 2.5). However, these improvements were shown in 1D plots of latitudinal mean of blocking frequency. These plots can mask latitudinal dipoles in the error and therefore provide erroneous results, see Scherrer *et al.* (2005) for a discussion between 1D and 2D blocking maps. In this thesis 2D maps are employed instead of 1D latitudinal mean plots.

Blocking frequency of control and SKEB2 simulations versus ERAI for DJF in the NH are shown in Figure 5.18. Contours show that there are two preferred regions for blocking: over the Bering Sea and the North Atlantic Ocean, which also encircles Central Europe. MetUM control simulation underpredicts the frequency of

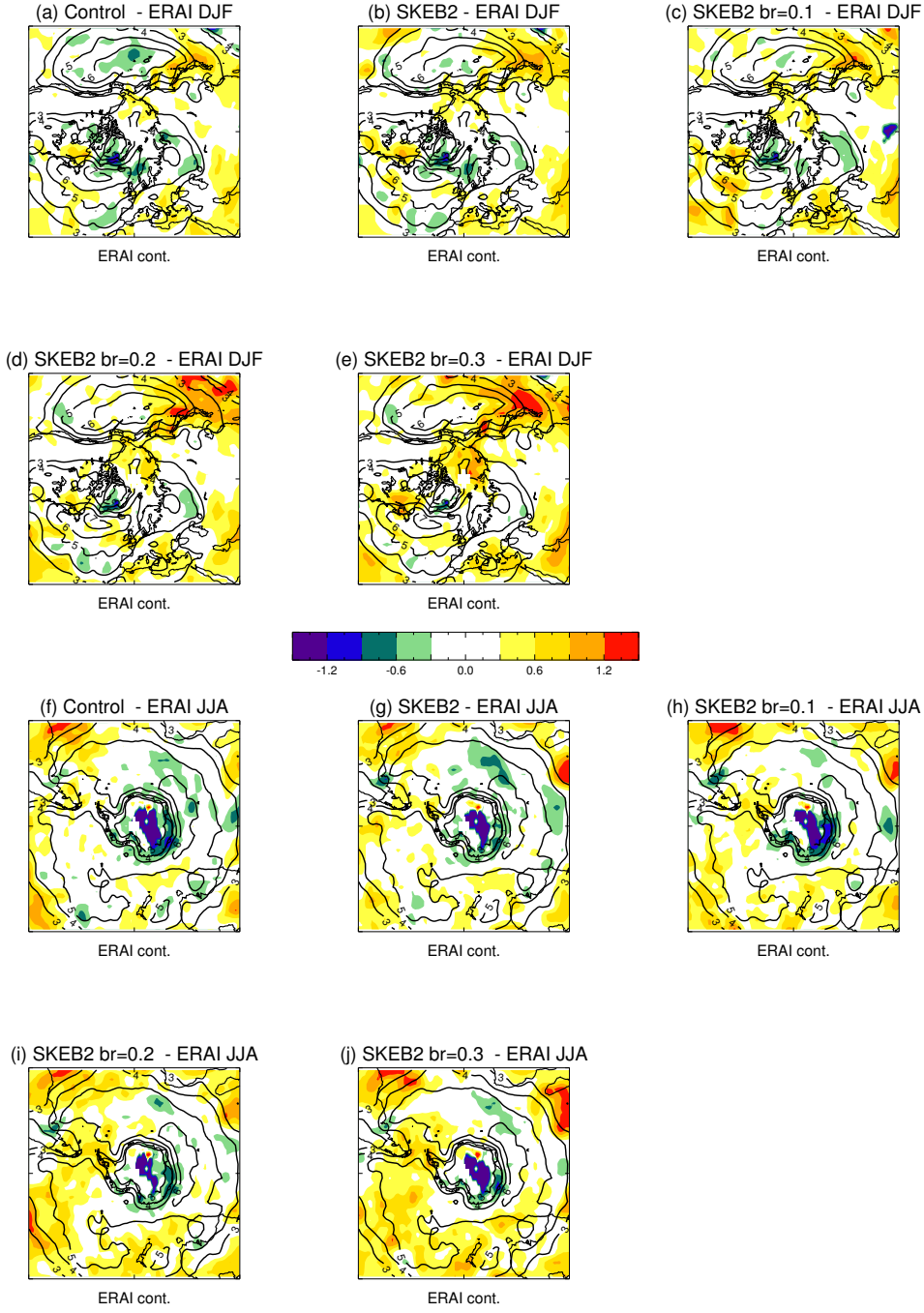

Figure 5.17: Mean Intensity of climate simulations (10^{-5} s^{-1}), the maximum of relative vorticity within the feature storm and then filtered T42. For NH in DJF: (a) Control minus ERAI (b) SKEB2 default b_R minus ERAI (c) SKEB2 $b_R = 0.1$ - ERAI (d) SKEB2 $b_R = 0.2$ - ERAI (e) SKEB2 $b_R = 0.3$ - ERAI. For SH in JJA: (f), (g), (h), (i) and (j), same configuration as the row above. Contours are ERAI values each 10^{-5} s^{-1} from 2 10^{-5} s^{-1}.

blocking over these regions, and over-predicts it over East Siberia and the Labrador Sea. The perturbations caused by SKEB2 helps to simulate more blocking over the Kamchatka Peninsula and Sea of Okhotsk (Fig. 5.18,c,e,g,i), but the frequency in this region is too high compared to ERAI (see effects of simulation with b_R =0.3 over this region on Fig. 5.18.h). Over other regions, the differences caused by SKEB2 are quite small and flips the sign depending on the amplitude or ensemble member. The amplitude of the scheme does not affect the degradation of blocking for DJF. In MAM all simulations with SKEB2 produce about 0.003 day^{-1} more blocking frequency over the Scandinavian peninsula (not shown). If we would make use of a 1D plot, it would probably tell us that SKEB2 improves the blocking index over the North Pacific on DJF in agreement with the other studies, but the 2D plots show that this improvement is slightly southwards of the main area of blocking frequency underestimation.

5.2.4 Tropical variability

The simulation of the tropical variability is one of the critical problems of state-of-the-art GCMs. The majority of these models have serious setbacks to represent adequately the intraseasonal tropical variability (Lin *et al.*, 2008). This problem is often associated to a poor representation of the spatial and temporal organization of convection, a key process for the development and propagation of Convectively Coupled Equatorial waves (CCEW, see 3.5). CCEWs are an important component for the adequate representation of the main mode of variability in the tropics, the Madden Julian Oscillation (section 3.6).

The kinetic energy backscatter by SKEB2 has a very positive effect on the mean tropical climate of OLR at TOA and the precipitation fields during the summer season in the Indo-Pacific region (JJAS) as shown in Figures 5.13 and 5.14. These benefits become more positive with an increasing b_R, so it is worth to investigate whether there could be some processes of the intraseasonal variability better represented by the strong stochastic forcing made by SKEB2 with high backscatter ratios. We are particularly interested in the impacts on CCEW, MJO and the organization of convection.

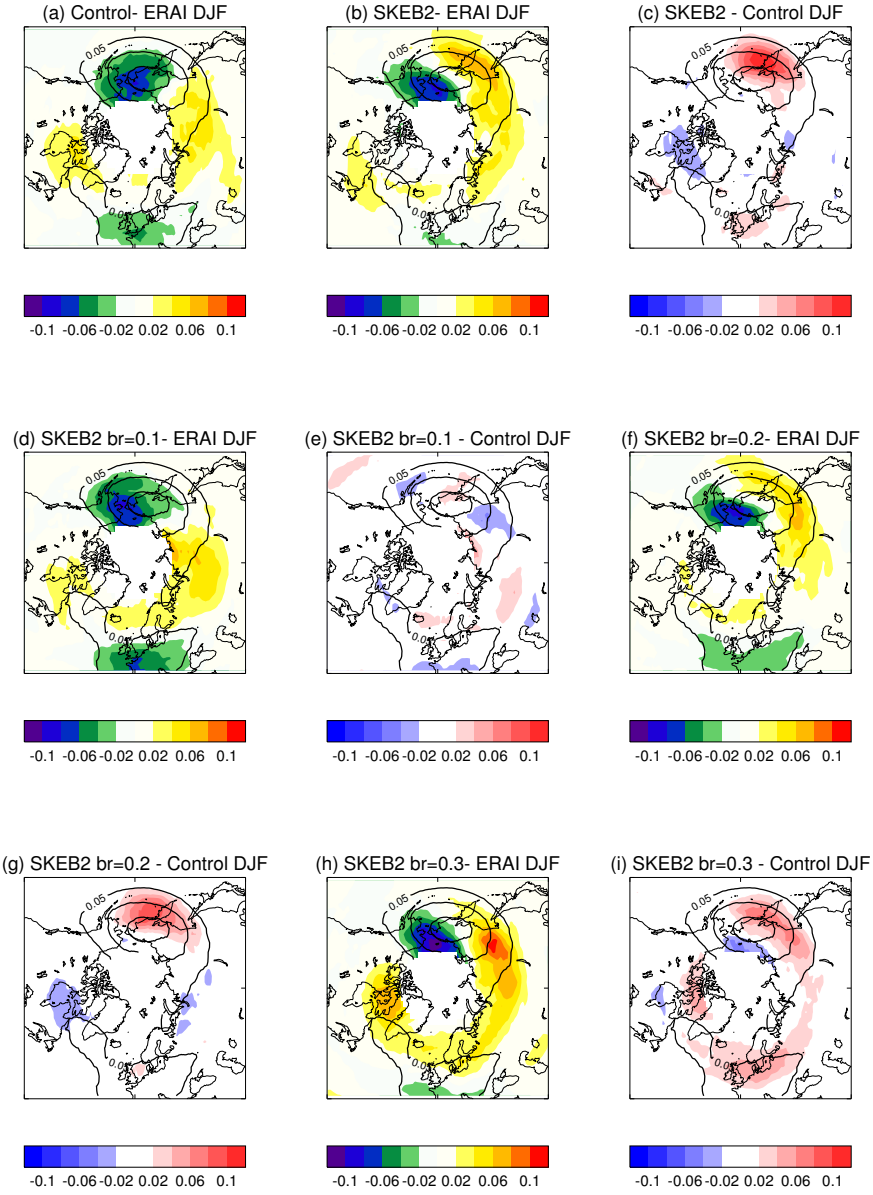

Figure 5.18: Blocking frequency for NH in DJF. Contours show ERAI from 0.1 day^{-1} for each 0.05 day^{-1}. Coloured, differences to ERAI (a) Control minus ERAI (b) SKEB2 default b_R- ERAI (c) SKEB2 default b_R - Control (d) SKEB2 $b_R = 0.1-$ERAI (e) SKEB2 $b_R = 0.1-$Control (f) SKEB2 $b_R = 0.2-$ ERAI (g) SKEB2 $b_R = 0.2-$Control (h) SKEB2 $b_R = 0.3-$ ERAI (i) SKEB2 $b_R = 0.3-$ Control.

Convective Coupled Equatorial Waves

The background removed power spectra of tropical low level winds indicates where the spectral and temporal variability is strong, and in combination with the dispersion curves of the idealized waves, it also indicates possible deficiencies on the simulation of specific CCEWs. The procedure to obtain the background removed power spectra is described in section 3.5.3.

The background removed power spectra of the tropical low level winds for the control run shows that some of the CCEWs are poorly represented or absent in comparison to ERAI (Figures 5.19.a,f and 5.20.a,f). The power spectra of westward propagating Kelvin waves, key components for the MJO propagation, is too weak against observations for long periods and too strong for periods shorter than 3 days. When we increase SKEB2's b_R, the power spectra of Kelvin waves slightly decreases as shown in the symmetric spectra for zonal winds at 850hPa (Figure 5.19). The Anti-symmetric part of the spectra shows that SKEB2 leads to the emergence of a westward propagating tropical wave with frequency lower than 3 days and wavenumber 5 (Figure 5.20). This node is not observed in the observations and control (Fig. 5.20.a,f), other variables such as OLR shows this unrealistic power for this spurious westward wave (not shown).

The emergence of the spurious westward wave (Fig. 5.20.c,d,e) could be related to the divergent component of the SKEB2 scheme, which forces the velocity potential besides the streamfunction, more related to rotational modes. As shown previously, the improvements in the OLR comes from the reduction of excessive divergence at upper levels (Fig. 5.13). The vertical tilt of the forcing pattern could be artificially stopping divergence at high levels and therefore improving the OLR. It may also degrade the power spectra of Kelvin waves, associated to divergence. We conduct another climate simulation at N96 with the highest backscatter ratio for SKEB2, but this time with no forcing to the velocity potential. Differences on the mean OLR at TOA between the simulation with the velocity potential forcing and the one without it are very small (Figure 5.21) and non-significant at the 95% level. The anti-symmetrical background-removed power spectrum of horizontal winds at 850hPa of the simulation without the velocity potential also shows the spurious westward wave (not shown).

Another possible cause of the spurious wave (Fig. 5.20,c,d,e) could be the conse-

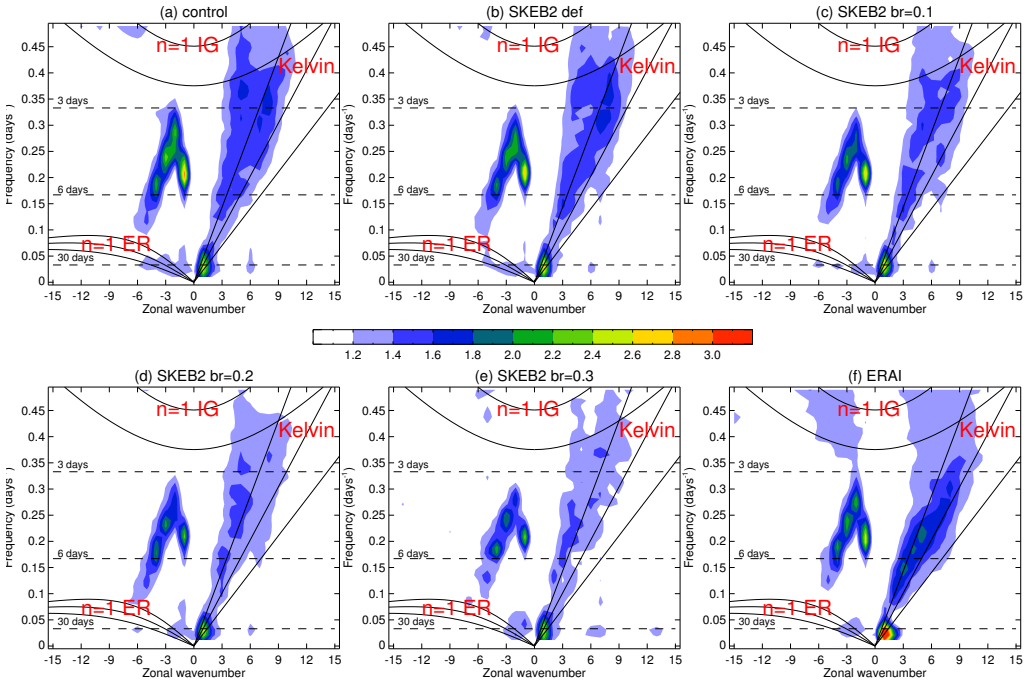

Figure 5.19: Background-removed symmetric power spectra for horizontal wind at 850hPa. Positive (negative) zonal wavenumbers correspond to waves propagating Eastwards (Westwards). Horizontal dashed lines indicate wave periods for 30,6 and 3 days. Idealized solutions of the tropical waves are shown for equivalents depths of 12m, 25m and 50m, for $n = 1$ Equatorial Rossby waves (ER), $n = 1$ Intetio-Gravity waves and Kelvin wave. (a) Control (b) SKEB2 default $b_R = 0.0275$ (c) SKEB2 $b_R = 0.1$ (d) SKEB2 $b_R = 0.2$ (e) SKEB2 $b_R = 0.3$ (f) ERA Interim.

quence of the unbalanced perturbations added by SKEB2, as it only perturbs winds through the streamfunction and velocity potential, but it does not perturb the temperature field. SKEB2 perturbations may radiate away as gravity waves mainly westwards, as the prevailing tropical high level winds are easterlies. The period of 3 days and their anti-symmetric nature remain inexplicable to us.

Madden-Julian Oscillation

Many aspects of the representation of MJO in current GCMs are poorly simulated (see section 3.6.2). These do not depend uniquely on the convective parametrization, they also depend upon the complex interactions of convection with other physical processes in the model.

The SKEB2's impacts on the tropical atmosphere are ambiguous so far. The

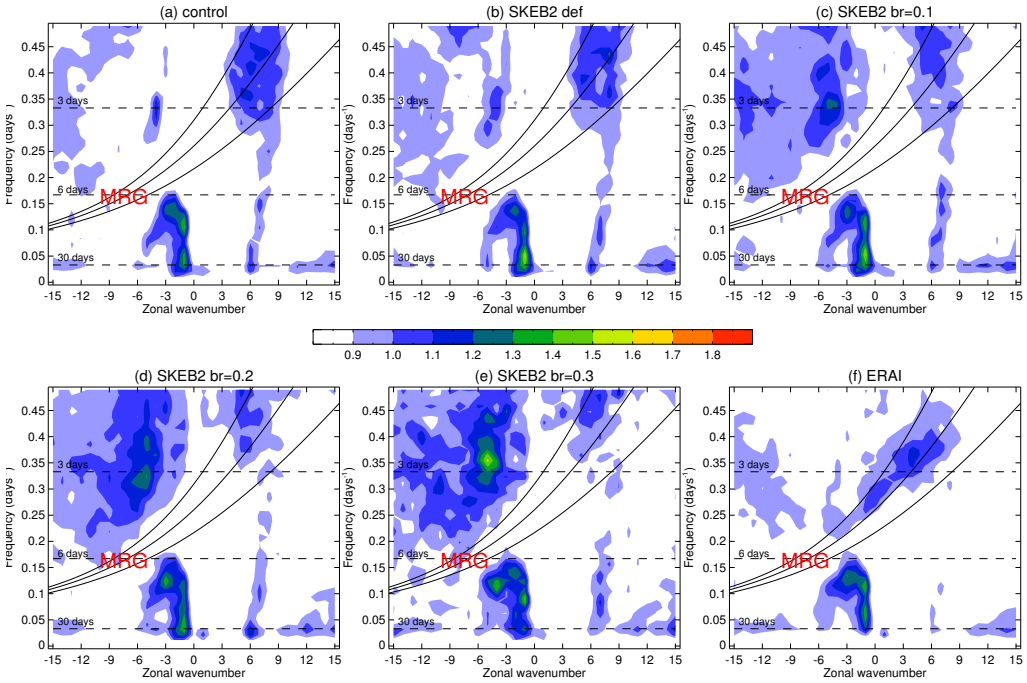

Figure 5.20: Background-removed anti-symmetric power spectra for horizontal wind at 850hPa, idealized solution for a Mixed Rossby-Gravity wave for n = 0 (MRG) is displayed with equivalent depths of 12m,25 and 50m. Same order as Figure 5.19.

improvements in the mean climate are clear, but CCEW seem travel with the wrong dispersion relationship. It is thus important to look at events that could provide information about the propagation of systems in the Indo-Pacific ocean. The use of lag-correlation plots is a good tool for these studies, as it shows the propagation of any particular anomaly across time and space.

The lag-correlation technique as described in Lin *et al.* (2008) is employed for the 30-60 day filtered precipitation to highlight the active phase of the MJO. We first look at the North-South propagation in the West Tropical Indian Ocean. Hence we average the filtered precipitation anomaly between longitudes 70-100E and correlated it to the value at the point 12.5N 85E, the main area of Northwest propagation of low level winds. The result of this operation is shown in Figure 5.22, where the GPCP observations show a clear propagation of rain northwards, followed by a lack of precipitation on the same latitude (Fig. 5.22.a). MetUM only shows significant correlation with the precipitation event between day -5 and day 10 (Fig. 5.22.b). The perturbations of SKEB2 seems to increase the correlation and in some cases like SKEB2 with $b_R = 0.3$ extends it back to day -10. A similar study is done for

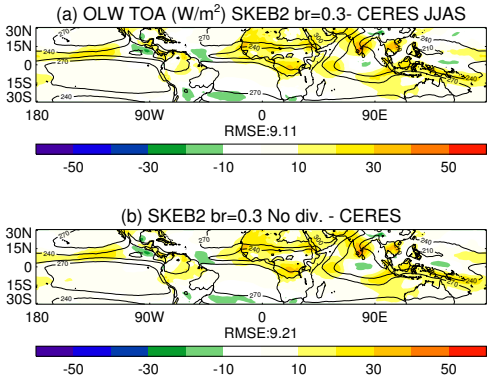

Figure 5.21: OLR at TOA (W/m^2) climate average differences between (a) SKEB2 $b_R = 0.3$ with velocity potential forcing (b) SKEB2 $b_R = 0.3$ without velocity potential. Same contours as Figure 5.13.

longitudinal propagation using the correlation between the 5-25N to itself at 15N 95E, SKEB2 impacts are negligible (not shown).

Temporal organization of convection

Despite the emergence of the erroneous kind of variability seen on the horizontal wind spectrum, SKEB2 can produce improvements in the temporal distribution of the tropical rain. The power spectra of daily rain for the tropical summer (JJAS) averaged over 20 years and latitude bands between 5S and 5N shows that MetUM produces too much power at low frequencies and too little variability at high frequencies in comparison to GPCP (Figure 5.23). The peak of the MJO between 30 and 60 days is missing, SKEB2 produces a better representation of these frequency nodes over the Indian Ocean (45-90E), although the power is too large over the Maritime continent.

Another deficient aspect of tropical convection is the intermittency the convection scheme. The scheme is triggered sporadically and it shows an unrealistic on-off behaviour, which kills the organization and propagation of convective systems (Martin G, 2012, *personal communication*). In order to understand whether SKEB2 produces more long-lasting episodes of precipitation, we output the convective rain during one season (JJAS for 1981, the start year of the climate AMIP run) for each timestep from control and the SKEB2 simulation with the highest backscatter ratio. The power spectra for periods of one day over the box 60-70E and 5S-2.5N, a region

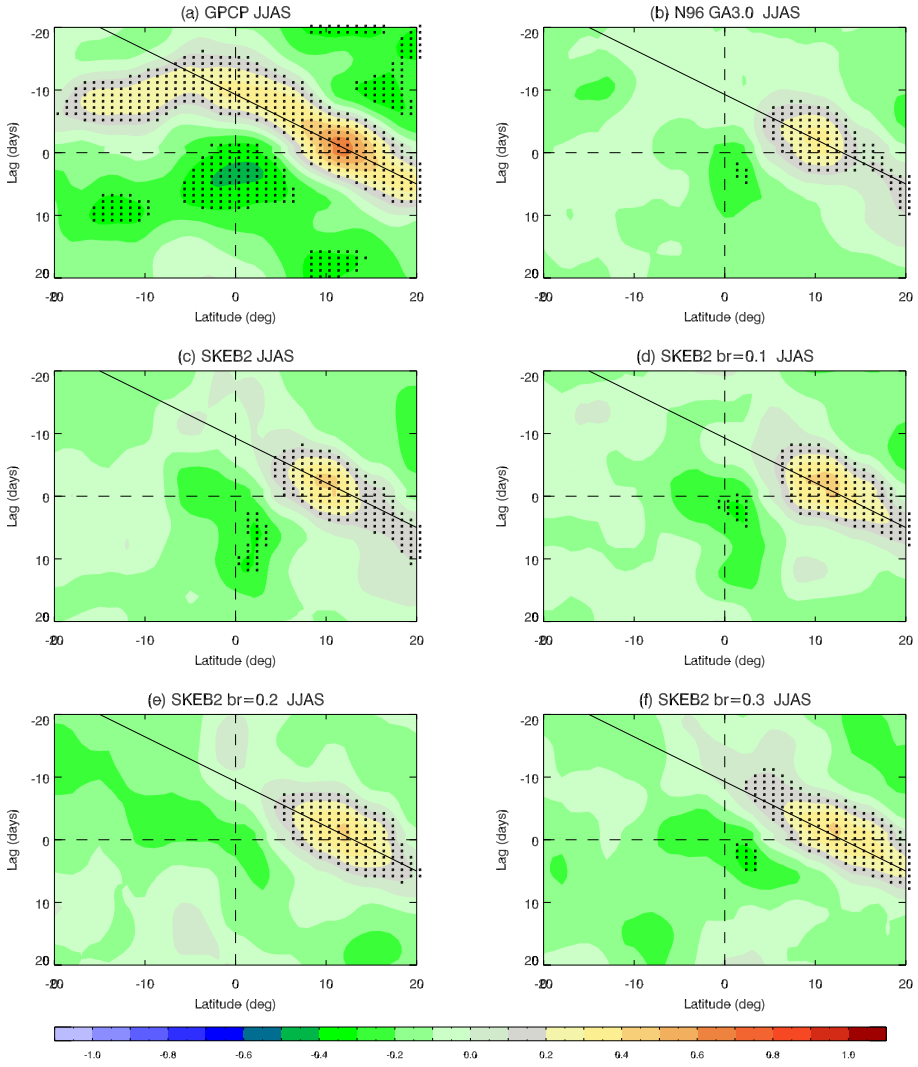

Figure 5.22: Lag correlation of the 30-60 day precipitation anomaly averaged between 70E and 100E with respect to itself at 12.5N and 85 E for JJAS (a) GPCP (b) Control (c) SKEB2 def. b_R (d) SKEB2 $b_R = 0.1$, SKEB2 $b_R = 0.2$ and SKEB2 $b_R = 0.3$. Dotted regions denote lag correlation above 95% confidence level. Diagonal thick line correspond to phase speed of 1.8 m/s.

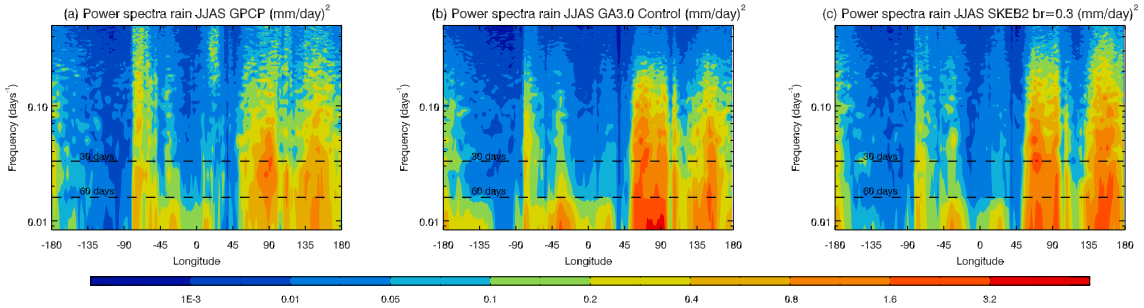

Figure 5.23: Latitudinally averaged (5S-5N) power spectra for precipitation at different longitudes for (a) GPCP (b) Control (c) SKEB2 with $b_R = 0.3$; Tropical summer (JJAS) 1 day averaged dataset. Dashed lines indicates frequencies equivalent to periods of 30 and 60 days.

where there is a large bias in OLR (Figure 5.13), is shown in Figure 5.24. For high frequencies (low number of timesteps) SKEB2 produces less power than control, indicating the convective precipitation is less intermittent and more organized.

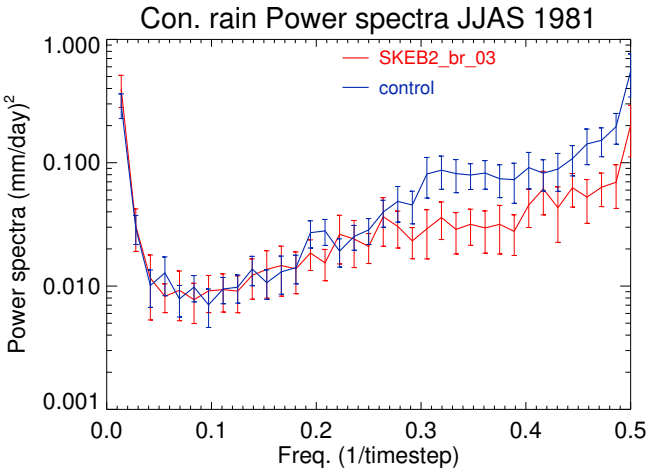

Figure 5.24: Convective rain power spectrum of one day intervals on a box over the West Indian Ocean (60-70E and 5S-2.5N), dataset is all timesteps of JJAS for 1981. Blue line is control and red one SKEB2 with the highest backscatter ratio. Confidence intervals are obtained using the methodology explained in the introduction of chapter 4.

Intermediate b_R values for SKEB2 indicate a gradual transition in the results shown in Figures 5.23 and 5.24 from control to SKEB2 with the highest backscatter ratio. The scheme increases high frequency and decreases red noise at low frequency in the seasonal variability, as well as decreases the high frequency inter-diurnal convective rain.

The scheme diminishes the effects of the on-off behaviour of the deterministic parametrization, making convection episodes to last longer and be more organized. These effects leads to an improvement in the seasonal variability of convection, propagation of precipitable systems and thus the seasonal mean of convection related fields such as OLR or precipitation.

5.3 Comparison to GA6 configuration

The GA6 configuration for MetUM includes the less diffusive dynamical core ENDGame and some changes to convection like an increase of the entrainment rate, which enhances the tropical variability (see 4.1.3 for further details on GA6 configuration). The differences between GA6 and GA3 are thus a good test-bed to show how SKEB2 perturbations compare to model improvements in the processes the stochastic scheme aims to represent. It is also interesting to investigate the behaviour of SKEB2 on the new configuration, where it is supposed to scale down its forcing given the GA6 improvements in the internal variability of the model.

Some of the results obtained from the NWP and climate experiments will be compared to those obtained from the GA6 and GA6 plus SKEB2 NWP forecasts and climate simulations, in particular to the mid-latitude processes such as extra-tropical cyclones and the impacts on tropical variability.

5.3.1 Impacts on NWP and mid-latitude cyclones

A similar set of 200 forecasts (see Table 4.1 for details) is run with the GA6 configuration at the lowest resolution, which we define as "GA6 control", another set of forecasts with GA6 configuration also includes SKEB2 at the default b_R and $b_R = 0.2$ to highlight the effects of SKEB2 in the new configuration. Climate simulations are run at GA6 with a control configuration and SKEB2 with $b_R = 0.2$ (see sect. 4.1.1 for a description of these systems).

Increasing the horizontal resolution is more effective in the reduction of RMSE of low level winds than improving the model, as shown in Table 5.10 for different regions. Changes made by the inclusion of the SKEB2 in the system degrade individual forecasts slightly more in GA6 than in GA3 for all regions. Similar results are found for Z_{500} (not shown). An increased horizontal resolution can resolve small

scale features than enables the predictability of mid-latitude systems, whereas improvements in the model dynamical core and physics can produce modest increases that are not significant at the 95% level.

	NH	Trop	SH
N96 GA3	4.46 ± 0.1	2.89 ± 0.03	5.06 ± 0.1
N96 GA3 SKEB2	4.56 ± 0.1	2.99 ± 0.03	5.17 ± 0.1
N216 GA3 control	4.22 ± 0.08	2.78 ± 0.02	4.78 ± 0.09
N96 GA6 control	4.40 ± 0.09	2.89 ± 0.02	5.02 ± 0.09
N96 GA6 SKEB2	4.52 ± 0.1	3.02 ± 0.03	5.16 ± 0.1

Table 5.10: Average of RMSE for the set of 200 forecasts, winds 850hPa. Confidence intervals at 95% are also included.

The results from cyclone tracking show a similar pattern: The intensity of the storms increases with SKEB2, a bit more with a less diffusive model implementation and the higher resolution produces the highest increase (Figure 5.25). When all these options are combined, the intensity bias becomes quite negligible, and even the intensity of cyclones becomes a bit over-active in the NH (not shown). In terms of cyclone position, improvements in high resolution produce better tracks than model changes (Figure 5.26). SKEB2 slightly degrades the position for both model configurations.

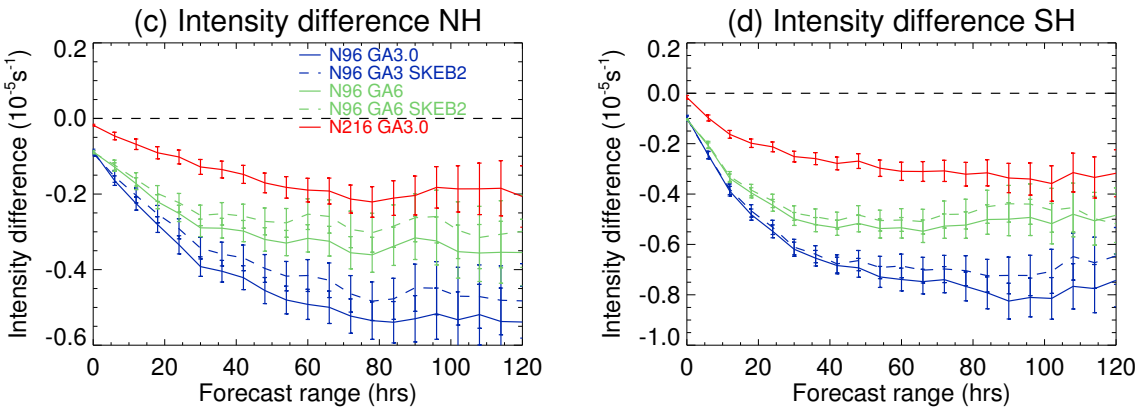

Figure 5.25: Average Intensity bias of tracked storms for (a) NH (b) SH. Blue are N96 at GA3, green N96 at GA6 and red N216 at GA3. Continuous indicate control runs and dashed forecasts with SKEB2.

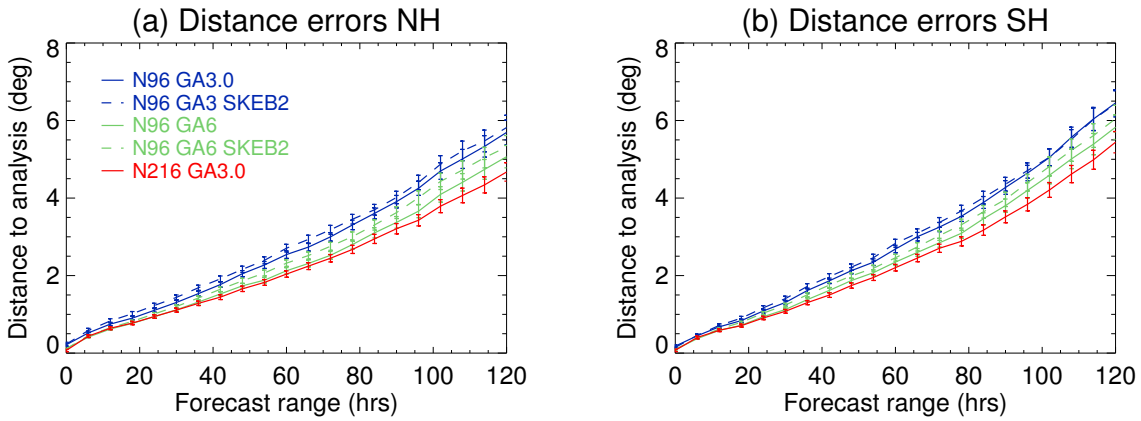

Figure 5.26: Mean distance error of mid-latitude cyclones to analysis. Same colour and line type classification as Figure 5.25.

The growth rate of the intensity of storms is underrepresented in GA3 (see Table 5.4), but a higher horizontal resolution and the inclusion of SKEB2 help to produce stronger growth rates. The upgrades on the model dynamical core and physics also improve the representation of the growth rate (Table 5.11). For the new configuration the increase is nearly similar to the increase in resolution from N96 to N216 and higher than the increment produced by the default version of SKEB2. Again SKEB2 seems to be quite insensitive to model changes as it produces similar increments of growth across different model versions and hemispheres. In fact, when the b_R increases to 0.2, the increments seen in the eddy growth rate are stronger at GA6 than GA3, which it is detrimental for the NH as the growth rate of storms becomes too high.

	NH	SH
N96-GA3	-0.46	-0.67
N96-GA3 SKEB2	-0.36	-0.60
N96-GA3 SKEB2 $b_R = 0.2$	-0.08	-0.20
N216-GA3	-0.28	-0.37
N96-GA6	-0.33	-0.50
N96-GA6 SKEB2	-0.22	-0.42
N96-GA6 SKEB2 $b_R = 0.2$	0.29	0.05

Table 5.11: Mean Growth Rate of the storm intensity in the GA3.0 and GA6.0 forecasts. Same format as table 5.4

The reliability diagram for the intensity of storms shows that improvements in the model increase the intensity of the "right storms", those with high intensity values which are more diffused by the model (Figure 5.27). The intensity of storms from the higher range of intensities are simulated with equivalent skill at the low resolution model with new configuration and high resolution model with the former configuration. The extra KE added by SKEB at GA6 configuration does not target the right intensity ranges as seen for the GA3 model (Fig. 5.8).

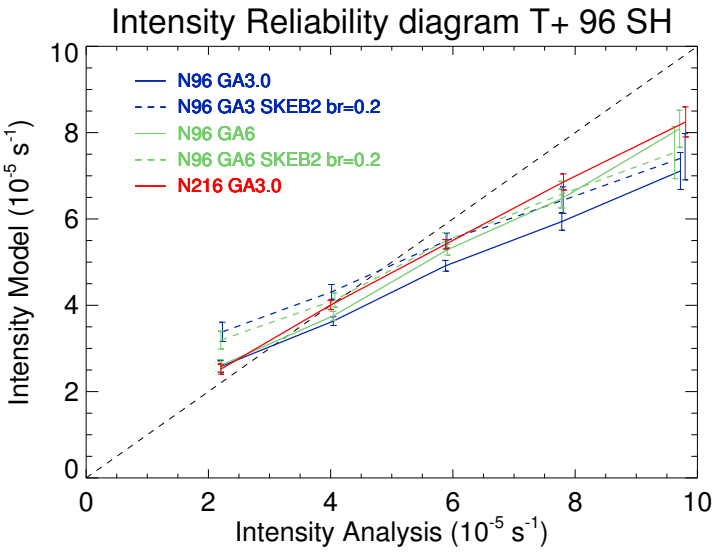

Figure 5.27: As Fig.5.8, but for N96 GA3 (blue), N96 GA6 (green) and N216 GA3 (red), continuous lines represent the control forecasts and dashed lines the forecasts with SKEB2.

The comparison across model configurations reveals that SKEB2 does not scale down its forcing at GA6, despite it works alongside a less diffusive dynamical core. SKEB2 produces a similar impact in terms of individual RMSE, cyclone intensity and positional errors across the two different MetUM configurations GA6 and GA3. In some cases, SKEB2's impact is more negative for the newest version of the model when the backscatter ratio is high.

5.3.2 Impacts on tropical climate

The tropical climate is very sensitive to changes in the convection parametrization. Some of the GA6 implementation includes changes to some of these parameters, such as the entrainment rate which leads to clear improvements in the representation of the MJO (Klingaman and Woolnough, 2013).

A high backscatter ratio b_R is still able to produce a better climatological representation of the Indian monsoon in the newest model configuration. The biases in precipitation over the West Tropical Pacific, and both sides of central America are reduced by SKEB2 for both model's configurations in JJAS (Figure 5.28,a,b). However, in GA6 there is already too much precipitation over the West Pacific and the inclusion of SKEB2 makes the bias over this region worse (not shown). Therefore the difference of the tropical precipitation RMSE is smaller between SKEB2 and control simulations for GA6 than GA3 (Table 5.12).

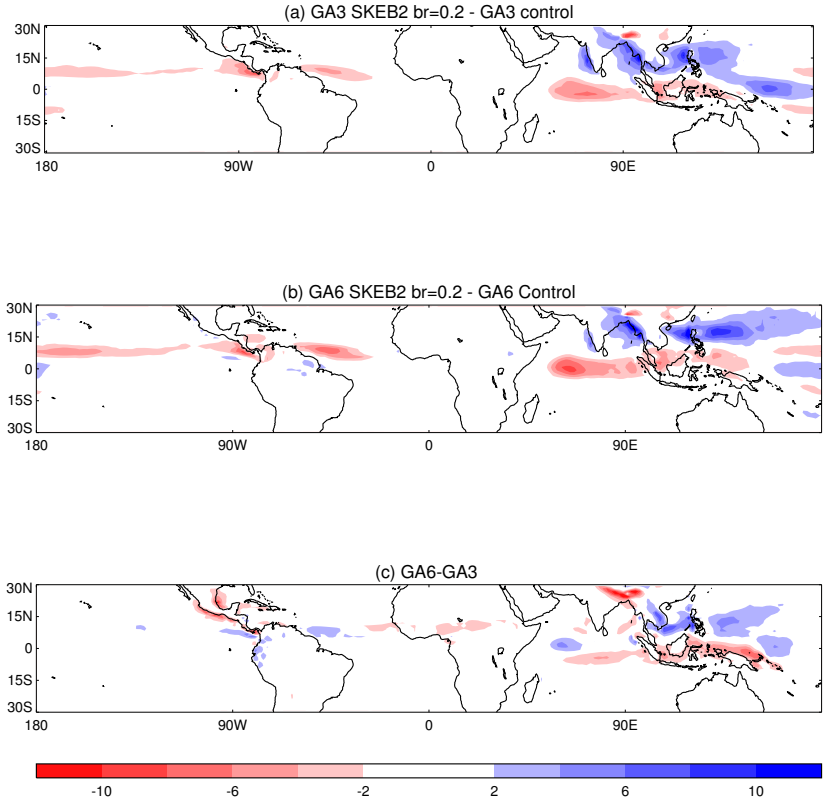

Figure 5.28: Differences in precipitation biases for JJA, (a) GA6 SKEB2 with $b_R = 0.2$ at GA3 (b) same but for GA6 (c) Differences between GA6 and GA3.

Over the Maritime continent the changes in precipitation caused by the model improvements from GA3 to GA6 are quite similar to those done by SKEB2, although the latter has a deeper impact (see Fig. 5.28). The excessive precipitation East of Phillipines occurs over a region where the standard deviation of low level winds, precipitation and OLR is too high. The increase of variability in this area could indicate that there are more sporadic events of local or large-scale precipitation, e.g. earlier phases of the MJO may travel beyond the Maritime continent enhanced by the modulated variability of SKEB2 and changes in the entrainment rate, but there

is no such clear signal on the MJO composites of the different phases (not shown).

	GA3	GA6
Control	2.54	2.62
SKEB2 b_R=0.2	2.17	2.50

Table 5.12: RMSE of the tropical precipitation (30S-30N) to GPCP (mm/day) for different model configurations.

The effects on high level winds, controlled by the divergence around convective plumes is shown in Figure 5.29. The GA3 control shows two major biases (Fig 5.29.a), the easterlies over the Tropical Indian Ocean and Maritime continent are too weak and the SH jet stream is too weak but its equatorward branch is too strong. Upgrades included for GA6 degrade these biases (Fig. 5.29,e).

As described in the SKEB2 climate results (sect. 5.2), the stochastic scheme produces more organized convection and reduces climate biases for OLR and precipitation (Figs 5.13 and 5.14). These effects invigorate the westwards winds over the Indian Ocean and Maritime continent improving the model (Fig. 5.29.b,d). However, SKEB2 also reduces the speed of the SH jet stream, which is positive to resolve model biases in the Tropical East Pacific, but negative over the Atlantic ocean and African continent. At GA6 the deceleration of the jet stream made by SKEB2 is stronger.

Despite the degradation of high level winds and West Pacific rainfall climate biases, the improvements seen in the GA6 control are clear for tropical variability (see also Walters D, 2014, *in preparation*). The power spectra of Kelvin waves increases for GA6 in comparison to GA3 (Fig. 5.30.a,b). The impact of SKEB2 with a high backscatter is quite similar across the different model versions (Fig. 5.30.c,d).

The representation of the backscattered Kinetic Energy by SKEB2 is quite homogeneous over the tropical belt. It improves the climatology of the Indo-Pacific region but with adverse effects on the representation of CCEW and the SH jet stream. The power introduced by SKEB2 might be too much on top of GA6 and its enhanced tropical variability. This excess of energy may radiate extra-tropical Rossby waves which impact the course of the SH Jet stream, some studies such as

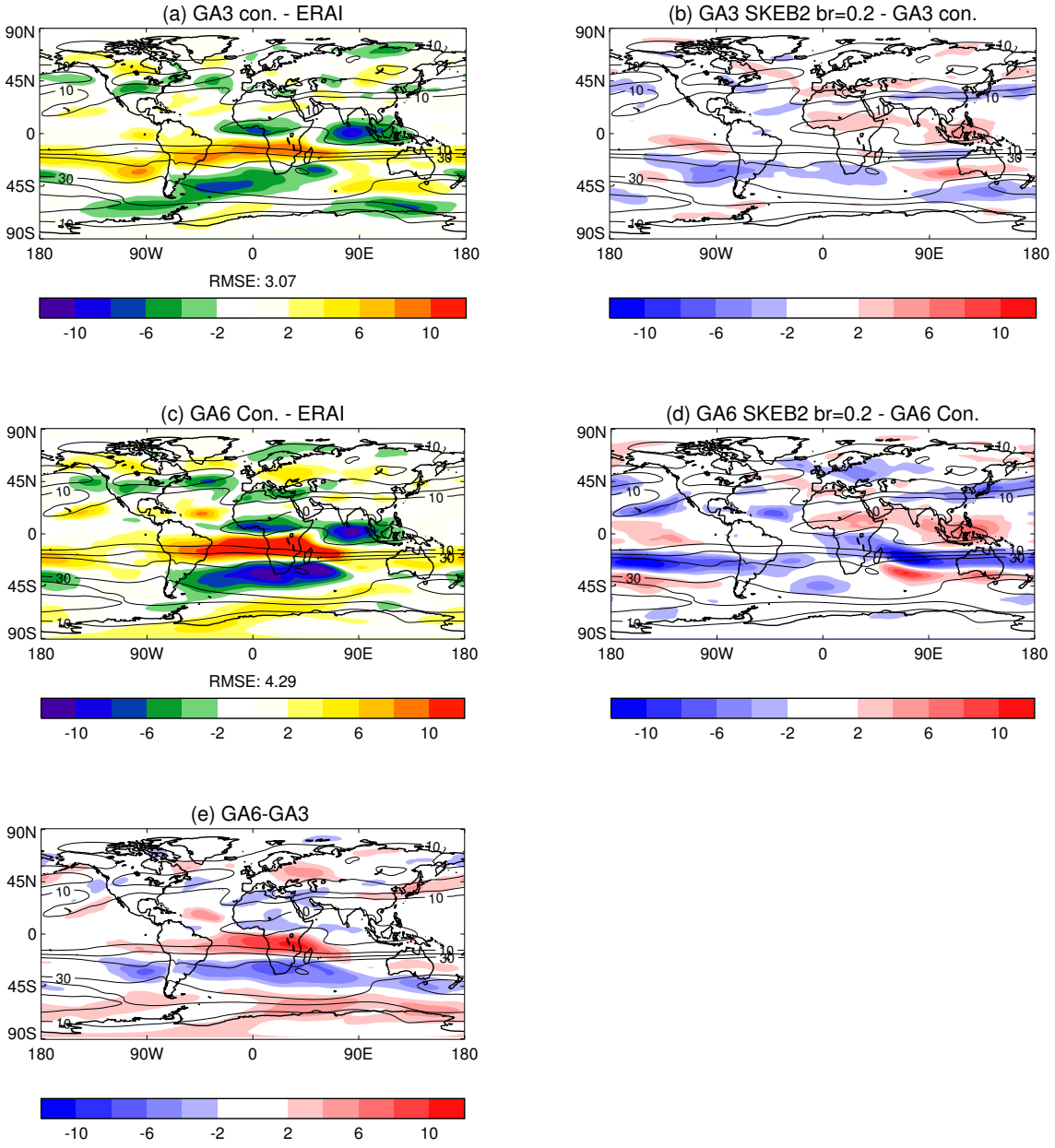

Figure 5.29: Differences (m/s) in 250hPa winds for JJA (a) GA3 against ERAI (b) GA3 SKEB2 $b_R = 0.2$ - GA3 (c) GA6 - ERAI (d) GA6 SKEB2 $b_R = 0.2$ - GA6 (e) GA6 - GA3. Contours show ERAI 10 m/s isolines from 10 m/s.

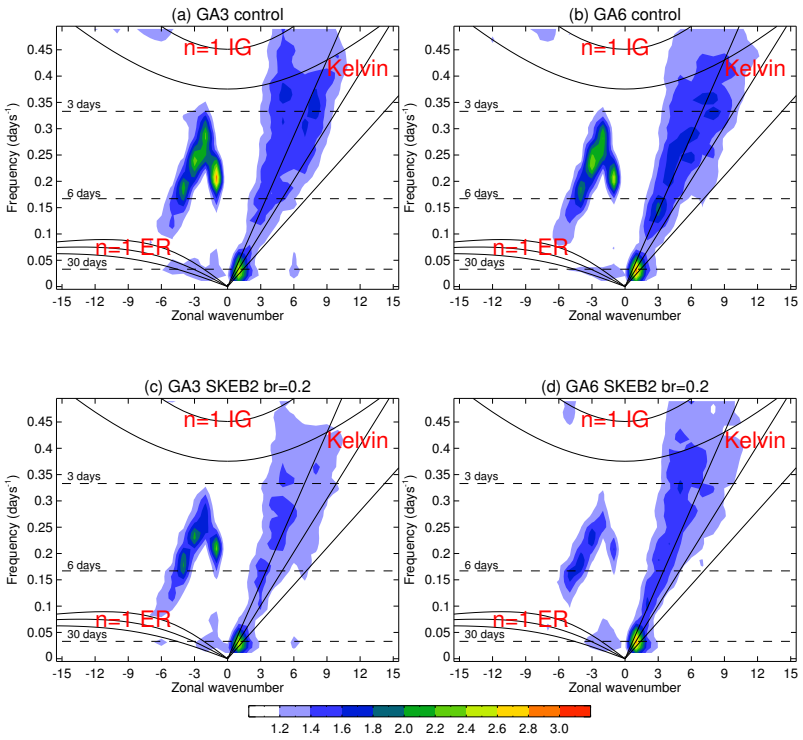

Figure 5.30: Symmetric Power spectra for (a) GA3 control, (b) GA6 control (c) GA3 SKEB2 b_R=0.2 (d) GA6 b_R=0.2 SKEB2. Same plot setup as in Figure 5.19

Lee *et al.* (2013) establish links between enhanced tropical convection and southern subtropical anticyclones.

5.4 Conclusions

A comprehensive evaluation of the SKEB2 scheme has been carried out using a very diverse set of methodologies. These have been applied to different timescales and regional processes. The most interesting results from this evaluation are:

- SKEB2 degrades deterministic forecasts of Z_{500} and horizontal winds at low and high levels. Skill scores such as RMSE or ACC are worse when the scheme is included and the stronger the forcing, the higher the degradation in forecast skill. The stochastic scheme changes the trajectory, intensity, speed and growth rate of mid-latitude cyclones pushing them away from analyzed cyclones. This change is proportional to the amplitude of the scheme controlled by the backscatter ratio b_R.

- Despite that individual systems are diverted by SKEB2, the mean effects of the scheme are positive. It helps to increase the intensity of mid-latitude

cyclones and strengthens weak low level winds. The average speed and growth of cyclones are also benefited from the Kinetic Energy released by the scheme.

- SKEB2 decreases the intermittency of tropical convection and increases the power spectra of daily precipitation, more in agreement with observations.

- The effects of SKEB2 on the tropical climate are positive for the boreal summer. A more organized convection regulated by the KE released by the scheme reduces climate biases in divergence which leads to improvements in OLR, precipitation and cloud fields. The higher the forcing, the stronger the bias reduction.

- The improvements in the tropical climate are not however driven by the right kind of synoptic variability. For high b_R, the scheme produces a spurious westward anti-symmetric wave and decreases the power of Kelvin waves.

- There is an increase in the blocking frequency over the North Pacific, but southwards of the model bias. Therefore the beneficial effects of SKEB2 in the representation of atmospheric blocking reported in other studies (e.g. Berner *et al.* 2012) may be a compensating error in the latitudinal distribution of blocking events over the area.

- The capacity of the scheme to generate spread at climate scales is negligible, therefore it may not be a useful tool to generate spread in predictions of uncertainty in climate change experiments or decadal prediction.

On average, the impact of SKEB2 is beneficial counteracting the internal diffusivity of the model and helping to organize convective events through the release of KE around convective cells. But the scheme has three major setbacks of the scheme:

- SKEB2 does degrade the predictability of short-range forecasts, suggesting that the scheme may be putting energy in scales that are well resolved and thus displacing weather structures such as mid-latitude cyclones out of their "natural path"

- SKEB2 perturbations do not seem to scale well across horizontal resolutions. The increase in storm intensity is similar if not higher at the high resolution N320 than at the low resolution N96. Moreover, the scheme increases the

intensity of storms uniformly and it does not produce stronger intensities on average to extreme storms that are highly diffused. The scheme's numerical and convective dissipation do not seem to compute adequately the high diffusion of vorticity gradients such as the ones produced by strong storms or the same storm across different resolutions.

- If a high backscatter ratio is employed, the perturbed streamfunction might be unbalanced with other dynamical fields such as temperature, leading to the emergence of spurious variability such as the westward node seen in the tropical anti-symmetric spectra (Fig. 5.19).

The results gathered in the present chapter describe a different narrative to the ones described in various studies that have employed the CASBS or SPBS in the IFS (Jung *et al.*, 2005a; Berner *et al.*, 2008, 2012). The SKEB2 improvements are not consistently better than those made by an increase in horizontal resolution, nor they are always lower than improvements made by physics. For the GA3 configuration of MetUM, the representation of the intensity of mid-latitude cyclones is greatly improved by increases in resolution and then modestly by changes in the model configuration and the addition of the SKEB2 scheme. On the other hand, improvements in the climatological tropical precipitation are better for SKEB2 than for the model upgrades.

The changes produced by SKEB2 are similar when the model is upgraded, even if the internal diffusivity of the model decreases because of upgrades like ENDGame, a less diffusive dynamical core. However, the stochastic scheme should modulate and scale down if its "raison d'être" becomes weaker. This is another reason to suspect the poor scaling of the scheme given by its dissipation masks and maybe by the forcing of large-scales. Different approaches to remedy these deficiencies will be explored in the next chapter.

Chapter 6

SKEB2 Improvements

In chapter 5 several deficiencies of the Stochastic Kinetic Energy Backscatter v2 scheme (SKEB2, sect. 2.2) were reported. These are the emergence of a spurious westward tropical wave when the amplitude of the scheme increases or poor the scaling of the scheme when resolution increases. It was suggested that these deficiencies could be caused by the implicit forcing of the large-scales through the forcing pattern and a poor construction of the dissipation rates. The present chapter investigates three different changes to SKEB2 to improve these deficiencies:

- We remove the large scales of the forcing pattern by increasing the lowermost wavenumber of the spherical harmonic decomposition, defined as N_1, from 5 to 20. Deterministic forecasts, probabilistic forecasts from MOGREPS and climate simulations are employed to determine the scheme's large-scale forcing on the degradation of skill, ensemble spread and tropical variability (see sect. 4.1.1 for a description of the different MetUM systems employed).

- A Biharmonic numerical dissipation is developed and compared to the current approach based on Smagorinsky (1963). The Biharmonic scheme is employed by other schemes such as the Spectral Backscatter Scheme (SPBS, sect. 2.2.2), and theoretically it produces a better matching to the internal dissipation of the interpolation to the departure point (McCalpin 1988 and sect. B.2.1). The contribution of different terms to the Biharmonic dissipation is investigated along the effects of both schemes on different vertical ranges. MOGREPS probabilistic forecasts are employed alongside climate simulations.

- A damping factor to control the convective dissipation for high horizontal

resolutions is added. It controls the perturbations from the convective part of the SKEB2 scheme, mainly acting in the tropics. The factor should amplify the SKEB2's effects at low resolution and reduce them at high resolutions.

As we want to minimize the emergence of other sources of model error from the SL internal diffusivity that are not related to the interpolation to the departure point, we employ GA6 MetUM configuration (described in sect. 4.1.3) to reduce the error caused by the excessive off-centering of the Semi-Lagrangian scheme.

In the first section of the chapter, the lowermost wavenumber of the SKEB2's forcing pattern is increased from 5 to 20 (sect. 6.1). We first analyse the structure of the forcing pattern and SKEB2 perturbations to assess it realism (subsection 6.1.1). The impact on the ensemble spread is quantified and some extra diagnostics are employed to determine the realism of the ensemble spread, some of these techniques are the cyclone tracking or the spectral decomposition of error and spread (subsect. 6.1.3). Effects on deterministic skill scores and tropical variability are also evaluated (subsect. 6.1.2 and 6.1.4 respectively).

The sensitivity of the model to the choice of numerical dissipation mask is detailed in section 6.2. The concept of numerical dissipation and the different methodologies to estimate it are described in subsection 6.2.1, which also includes a comparison between these. Results from probabilistic forecasts made by MOGREPS with the different dissipation masks are reported in subsect. 6.2.2, and climate simulations in subsect. 6.2.3.

The last section of the chapter describes the damping factor to the convective dissipation rate (sect. 6.3). Its impacts, in combination to an improved numerical dissipation mask, are reported in subsection 6.3.1 for NWP forecasts and 6.3.2 for the tropical climate.

6.1 Increasing N_1 from 5 to 20

The stochastic forcing of large-scales is not a desirable effect of the SKEB2 scheme. Stochasticity should be introduced at the truncation scales and then implicitly upscaled towards the large-scales by the model. However, the excessive diffusivity at small scales and other problems in the spectral energy transfer makes it very difficult for such a scheme to have any noticeable impact. Therefore the SKEBs schemes act

on a wide range of scales to maximize the impact on the ensemble spread without degrading the forecast skill too much.

Many of the SKEB schemes include low wavenumbers forcing, e.g: Spectral Backscatter Scheme (SPBS) that forces all wavenumbers, including those $n < 10$ which are well resolved in short-range forecasts. SKEB2 forces wavenumbers in the 5-60 range to maximize the spread (Tennant *et al.*, 2011). Higher wavenumbers than 60 are not forced as they are not within the energy containing scales and thus have infinitesimal effects. One exception is the SKEB scheme at the Meteorological Service of Canada (Charron *et al.*, 2010), which only forces small scales from wavenumber 30 to 128, but they use a longer decorrelation timescale of 36 hours instead of 6. Thus their perturbations on smaller scales last longer, and may be able to effectively force the flow to a similar magnitude as it was forced by a large scale pattern. For further details on the different SKEB schemes see section 2.2.

The stochastic contribution could produce too much noise on the backscattered signal at low wavenumbers. Coarse-grained studies of the vorticity equation made by Shutts (2013) have shown that the backscatter on low wavenumbers is constant, and the SPBS backscatter is too noisy at those scales. Therefore the effects of increasing the lowermost wavenumber in SKEB2 are investigated in the next subsection.

6.1.1 Structure of the forcing pattern

Increasing the lowermost wavenumber produces a less homogeneous and weaker pattern. Figure 6.1 shows the forcing pattern F_Ψ for both configurations. When the large-scales (wavenumber 5 to 20) are included, the pattern looks more homogeneous and with a deeper amplitude, whereas without the large-scales it looks more spotty and the amplitude halves. The size of the nodes in the $N_1 = 5$ forcing pattern (Fig. 6.1.a) could encompass structures such as synoptic cyclones, therefore the SKEB2 could potentially shape its features (e.g position, intensity, growth) as shown in section 5.1.

The power-spectra of the global streamfunction forcing $\hat{F_\Psi}$, averaged over all levels, is shown in Figure 6.2. SKEB2 with $N_1 = 5$ shows a quasi-constant forcing between wavenumber 5 and 60, the range chosen to backscatter the KE, then it drops to zero on the smaller scales. On the very low wavenumbers, the forcing made by the scheme with $N_1 = 20$ is about 25 times smaller than with $N_1 = 5$.

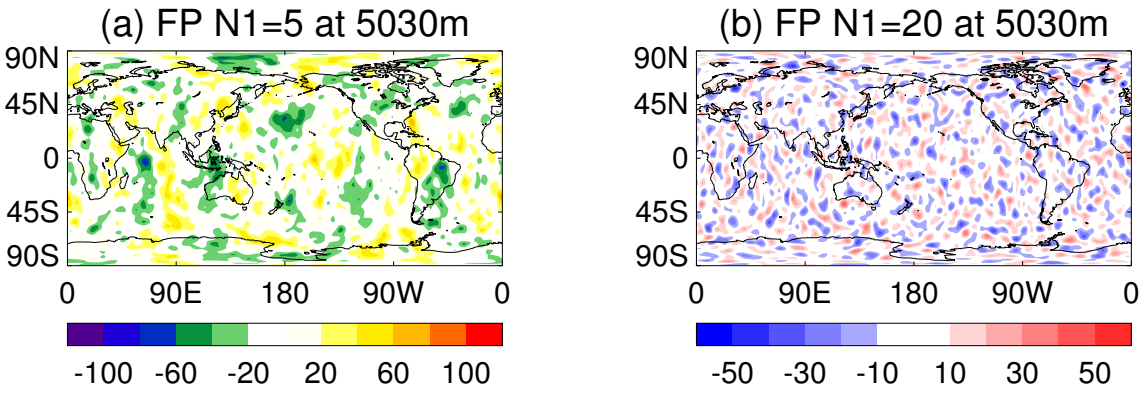

Figure 6.1: Forcing Pattern (F_Ψ, adimensional variable) for 12Z 02/09/1981 at the MetUM level corresponding to 5030m elevation from terrain for. (a) N_1=5 (b) $N_1 = 20$.

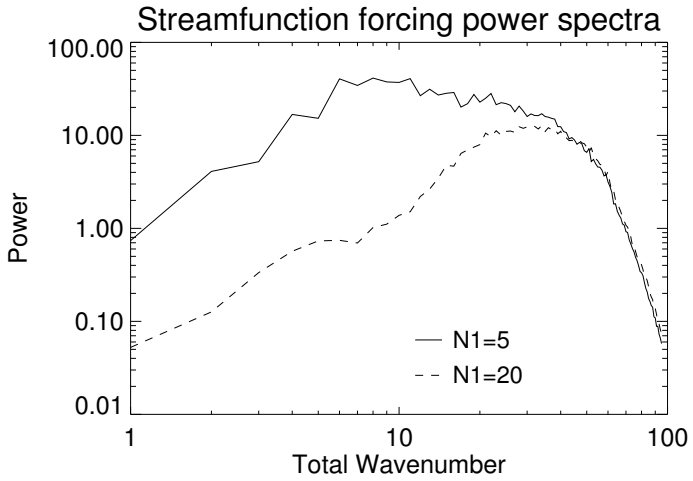

Figure 6.2: Power spectra of the vertically integrated Streamfunction forcing ($\hat{F_\Psi}$). Continuous line shows $N_1 = 5$ and dashed line $N_1 = 20$.

6.1.2 NWP impacts

The impact of a SKEB2 which does not force the large-scales is more beneficial in terms of individual RMSE across different regions and forecast-lead times (see Tables 6.1 and 6.2). The skill of the forecast with $N_1 = 20$ nearly halves the mean RMSE of the $N_1 = 5$ to control. The error introduced by the SKEB2 scheme on deterministic forecasts is still significantly high.

The reduction of the negative effects of the scheme on the skill of the forecast may be produced by a less active scheme, rather than a more realistic one that samples the uncertainty of the model in a better way. As shown in Figure 6.1, the amplitude of the forcing pattern is smaller, so the SKEB2 perturbations are also smaller. A

	T+24	T+72	T+120
GA6 Control	2.20	3.37	4.40
SKEB2 $N_1 = 20$	2.25	3.45	4.47*
SKEB2 $N_1 = 5$	2.28	3.53	4.52*

Table 6.1: RMSE of winds at 850hPa for different forecast lead times and experiments over the Northern Hemisphere (20N-90N). * denote values that are not statistically significant at the 95% to the experiments to control.

	T+24	T+72	T+120
GA6 Control	1.80	2.43	2.89
SKEB2 N_1=20	1.84	2.49	2.95
SKEB2 N_1=5	1.86	2.54	3.02

Table 6.2: Same as Table 6.1 but for Tropics (30N-30S)

less active scheme might have serious problems to generate ensemble spread in a Ensemble Prediction System (EPS), the major task of SKEB2. Therefore the impact on EPS performance of these two configurations of the scheme is investigated in the next sub-section.

6.1.3 Impact on ensemble spread

Two different MOGREPS set of forecasts are run with $N_1 = 5$ and $N_1 = 20$, in order to assess the impact on the ensemble spread and skill. The description of the probabilistic forecast setup is in sect. 4.1.1.

The classic Ensemble Mean (EM) error vs ensemble spread plot for winds is shown in Figure 6.3. The RMSE of the control run (with no SKEB2 perturbations) is higher than the EM, as the latter have been smoothed by averaging the 11 members of the ensemble. The spread is clearly lower than the error indicating that the ensemble is underdispersive, something which we expect as the EPS does not include any perturbation to the initial conditions. The spread produced by the $N_1 = 5$ version is higher in all regions for low and high level horizontal winds, and its Ensemble Mean (EM) error is slightly lower.

The error and spread curves show a distinct behaviour over the mid-latitudes and

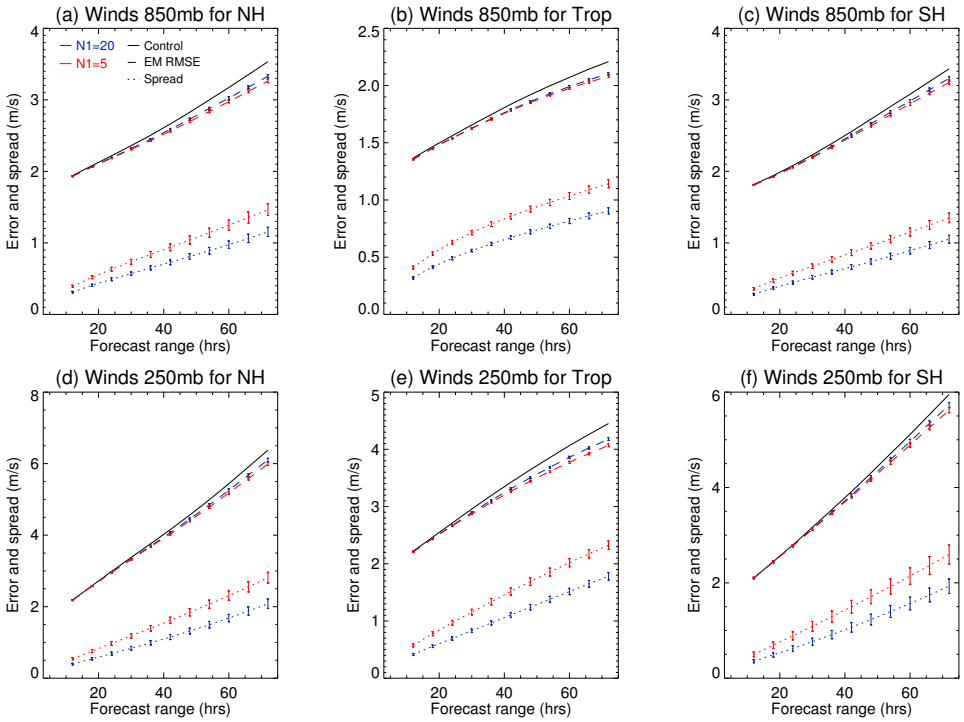

Figure 6.3: RMSE vs spread plots for winds at 850hPa (top row) and 250 (bottom row). Red line represents $N_1 = 5$ and blue $N_1 = 20$. Continuous black line shows the mean RMSE of the control member, which is similar for the two experiments, dashed line shows Ensemble Mean (EM) RMSE and dotted ensemble spread. (a) and (d) for NH (30-90N), (b) and (e) Tropics (20S-20N), (c) and (f) SH (90-30S).

tropics. Over the mid-latitudes error and spread are quite linear (Fig. 6.3 .a,c,d,f), with the former growing faster than the later, therefore the model is diverging from the observation's manifold and the perturbations from the scheme are not strong enough to push the model towards the real state. Over the tropics (Fig. 6.3.b,e), the error and spread show a parabolic behaviour, they grow faster on the first two days of the forecast and their seem to saturate afterwards, from day 2 to 3 of the forecast, the error does not seem to grow more than the spread as it does in the extra-tropics. The shorter timescales of the processes governing tropical weather such as inter-diurnal convection saturates the error in the first two days of the forecast. The geopotential at 500hPa (Z_{500}) in the mid-latitudes exhibits the same behaviour although the improvements of the EM to the control are less evident (not shown).

Error growth across spatial scales:

As suggested previously, the SKEB2 with $N_1 = 5$ could be forcing scales with low error (well resolved), creating the spread seen in Fig. 6.3. Thus it is useful to investigate the power spectra of the ensemble error and spread as done in Berner *et al.* (2009). It was shown that the characteristics of error growth across spatial scales were better captured by an ensemble with SPBS and reduced initial perturbations in comparison to an ensemble with large initial perturbations and no SPBS.

The Z_{500} power spectra of the ensemble mean and spread across different forecast lead times is shown in Figure 6.4. The ensemble error grows faster across forecast time in the synoptic scales (wavenumbers 5-40) and there are marginal differences between both SKEB2 experiments, the ensemble spread grows across all scales independently of the lowermost scale of the forcing pattern defined by the N_1 parameter. At the third day the spread at large scales is closer to the error than synoptic or meso-scales (Fig. 6.4.d), which is a worrying aspect of the scheme. The differences in spread between both N_1 experiments are nearly similar in all scales although at the end of the forecast both converge on the mesoscale ($n > 50$).

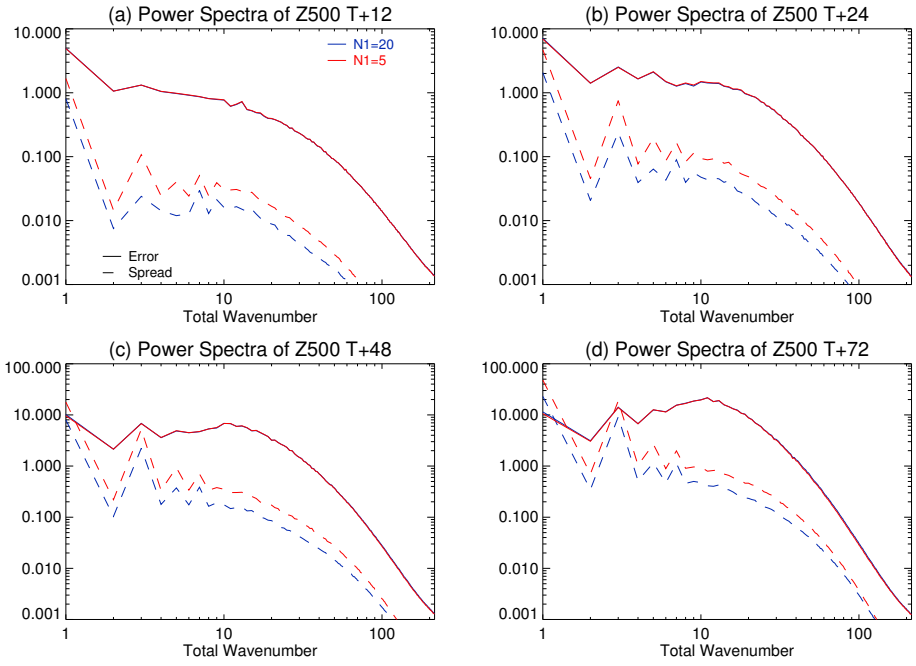

Figure 6.4: Power spectra of ensemble mean error (solid) and ensemble spread (dashed) for different forecast times (a) T+12 (b) T+24 (c) T+48 and (d) T+72. $N_1 = 20$ is show in blue and $N = 5$ in red.

A forcing pattern that includes large-scales produces more spread and slightly

less error on the ensemble for winds and Z_{500}, but not necessarily on the right scales where the error grows. In the mid-latitudes, mid-latitude cyclones are the main drivers of variability on the synoptic scales, whose error growth is poorly simulated. It is thus interesting to observe the growth of the spread and error produced in the representation of storms.

Impact on mid-latitude cyclones

The ensemble spread is an indication of the capacity of the model to represent its uncertainty, but it does not provide much information about the sources of error and how well the spread matches those. There is one methodology available to understand the dispersion of the ensemble in terms of the representation mid-latitude cyclones, one of the main drivers of variability in the extra-tropics. Such technique is the cyclone tracking and matching for EPS developed by Froude *et al.* (2007b) and described in section 3.3.3 of the thesis.

The matching technique tracks all the ensemble members. The statistics from the ensemble of tracks show how well an EPS can simulate the uncertainty of the position and intensity of storms. If SKEB2 failed to produce a significant impact, one particular storm would be very similar across different ensemble members and its spread would be low. If the spread was too high, it would be over the error of the control member. If the scheme was forced by the wrong mechanisms, the error of the tracks from the perturbed members would be higher than the control.

The population of storms for the different N_1 experiments is around 1400 for NH, and 1100 for SH. The spread in the trajectory of the storm is small and does not change much with a different N_1 (Fig. 6.5.a,b). For storm intensity, the experiment with the lowest N_1 produces more spread (Fig. 6.5.c,d). The ensemble shows higher diversity in terms of storm evolution, but still lower than the absolute error of the EM and control. In absence of Data-Assimilation and initial condition perturbations, storms are quite similar across both hemispheres, with NH having slightly more error and spread, probably driven by the hemispheric differences in orography and land-sea contrast.

Unlike it is shown in Figure 6.3, where the control from both experiments is similar, Figure 6.5 shows different statistics for the control member for the two experiments. The matching technique for EPS demands that at least one perturbed

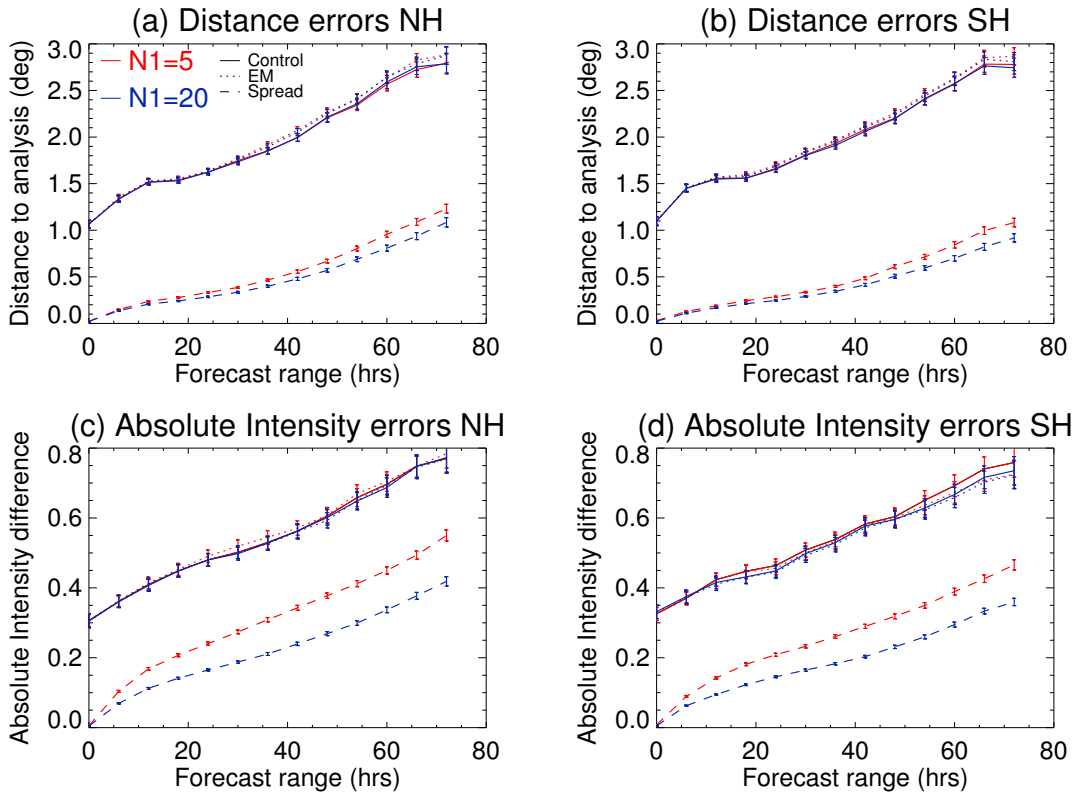

Figure 6.5: Properties of the mean storm representation in the EPS across forecast times, (a) Distance error in the NH (b) Distance error in SH (c) Absolute intense in the NH (d) Absolute intensity in the SH. Solid line represents control, dotted ensemble Mean and dashed Ensemble spread. $N_1 = 20$ is show in blue and $N = 5$ in red.

member of the ensemble must also match the analysis, therefore the number of storms matched to the analysis can be slightly different across experiments. Despite the "control" storm intensity and distance is different in both experiments, these differences are statistically indistinguishable at the 95% level.

The EM of the storm intensity does not provide any predictability increase over the control member, in disagreement with many EPS within the TIGGE ensemble (see Figures 3 and 4 of Froude 2011). As shown in the previous chapter (see sect. 5.3), the inclusion of a less diffusive dynamical core in GA6 makes the diffusion of storm intensity smaller, therefore at medium to high resolutions the representation of the mean intensity is quite realistic and is not seriously affected by the numerical dissipation. The MOGREPS members have a horizontal resolution of N216, which could be able to simulate storms with the right intensity. The mean absolute intensity of the storm population does not really show if the model produces storms that are too weak or too strong. Therefore the mean intensity error of the EM

and control is shown in Figure 6.6 for both hemispheres. The intensity is very well simulated by the control in the NH but on the SH it drops as shown extensively during the evaluation of NWP forecasts in the last chapter (sect. 5.1). The addition of a strong forcing from SKEB2 might create storms that are too strong in the NH, leading to a degradation of the EM skill.

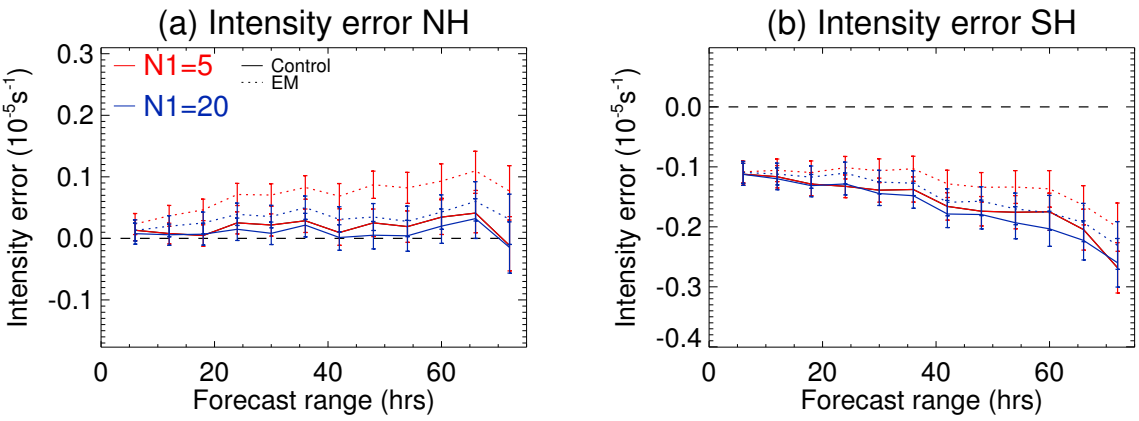

Figure 6.6: Mean Intensity of tracked storm for control (solid) and EM (dashed), red line shows $N_1 = 5$ and blue $N_1 = 20$. (a) NH (b) SH.

Another interesting aspect of the ensemble evaluation is the reliability, the capacity of the ensemble to simulate large spread when the error is large and little when there is a high degree of predictability. In the context of storm evaluation, it is very valuable to provide the right probabilities for storms as these are routinely used to set warnings for weather hazards (Neal *et al.*, 2013).

The reliability of spread vs EM error for storm position is shown in Figure 6.7. The population of EM storm positional error is distributed into bins of 1.5°, the spread of the storms for each of these bins is averaged, then it is possible to quantify if there model produces too much (low) spread for small (large) error. The reliability of the ensemble is far from the $1:1$ line, which indicates perfect reliability, the lack of initial condition perturbations does not push the storms away from their preferred trajectories early in the forecast, having little spread for large errors. The reliability improves a bit with forecast time, as the reliability curve has a higher slope and it's closer to the $1:1$ line. Differences between the two experiments are not significant in the majority of the bins, and the behaviour in both hemispheres is quite similar.

For the reliability of storm intensity, the spread induced by SKEB2 should be

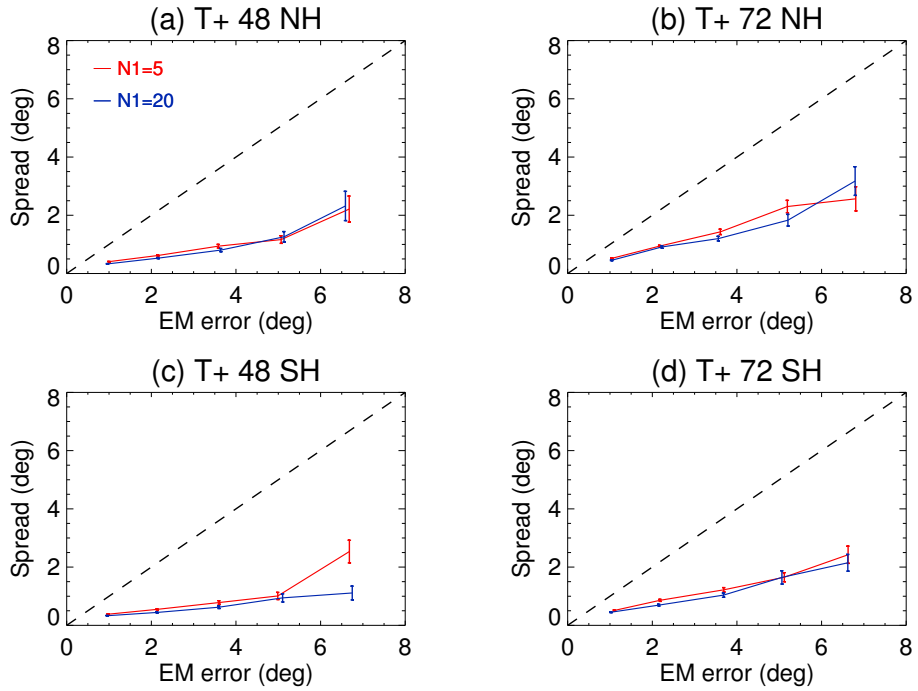

Figure 6.7: Spread-RMSE EM error reliability diagram for the position of storms (a) T+48 NH (b) T+72 NH (c) T+48 SH (d) T+72 SH. Red line shows $N_1 = 5$ and blue $N_1 = 20$ dashed 1:1 line shows perfect reliability. Bins from EM error are taken every 1.5 deg.

higher to match the error of the control one. Figure 6.8 indicates that spread is quite constant independently of how far the model is from the real state. The spread increases with forecast time and slightly becomes a bit more sensitive to the EM error. The flat line seen in the early stage of the forecast (Figure 6.8a,c) is something expected as the error in such timescales mainly comes from a poor representation of the initial state in the analysis, and the SKEB2 is not suited to represent this source of error. However, at day 3 the model error is evident as detailed in the evaluation of SKEB2 with NWP forecasts with SKEB2 (sect. 5.1), and the SKEB2 seem to have little sensitivity to these large errors. With the forcing pattern acting on scales $5 < N_1 < 20$ the spread of intensity is slightly overdispersive for small errors (< 1 10^{-5} s^{-1}). If forecasts were longer than 3 days, the spread of storm intensity for low errors would grow well above the 1:1 line, producing unreliable probabilities of the intensity of storms that are well predicted.

An EPS built with the solely contribution of SKEB2 fails to produce enough dispersion and sample the error in the representation of extra-tropical cyclones. The scheme's perturbations do not grow enough when error becomes larger in terms of position and intensity. The SKEB2 with $N_1 = 5$ produces overactive storms in

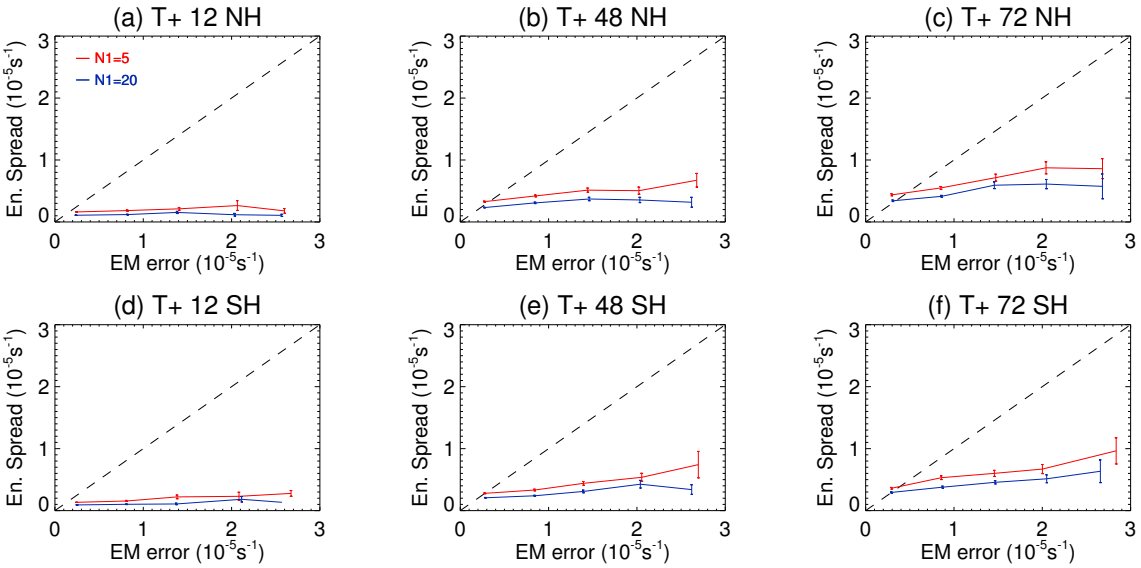

Figure 6.8: Spread-RMSE EM error reliability diagram for the intensity of storms, EM error is sorted in bins of 0.6 10^{-5} s^{-1}. (a) T+12 NH (b) T+48 NH (c) T+72 NH (d) T+12 SH (e) T+48 SH (f) T+72 SH. Same colour and line properties as Figure 6.7.

the NH that may not produce a better mean-track than the control.

Maybe three day forecasts are not long enough to allow SKEB2 to produce a sizeable impact closer to the forecast error. The lack of perturbations in the initial conditions is partially responsible for the poor spread and ensemble skill described. However, the model error emerging from deterministic forecasts is clear for the intensity of storms, which are too weak from the first day of the forecast (Fig. 5.5). The sources of the diffusivity of storm's intensity could be attributed to different systemic characteristics of the dynamical core, such as the excessive off-centering or the error in the interpolation to the departure point (described in section B.2.1). Unlike climate scales, where the effects of these sources of model error are well known (Stratton, 2004; Sanchez *et al.*, 2013), there is no clear understanding of what fraction of error correspond to each source in the short-range weather forecasting, and whether their attributes are different. Therefore, with the current knowledge about the uncertainties arousing from numerical dissipation, it is hard to tell whether SKEB2 is doing a good job on the representation of some of the sources of model error it aims to simulate.

6.1.4 Tropical variability

Climate simulations with different N_1 are carried out, in order to understand how the forcing on large scales changes the impacts seen on the mean climate (sect 5.2). These experiments use the AMIP setup described in section 4.1.1, with the only difference of having a different value for N_1. The simulation of the Convectively Coupled Equatorial Waves (CCEW) improves with the removal of the large-scales of the forcing pattern. It removes the spurious westward wave from the anti-symmetric background removed power-spectra (Fig. 6.9). The power of Kelvin waves is quite similar for both simulations (not shown).

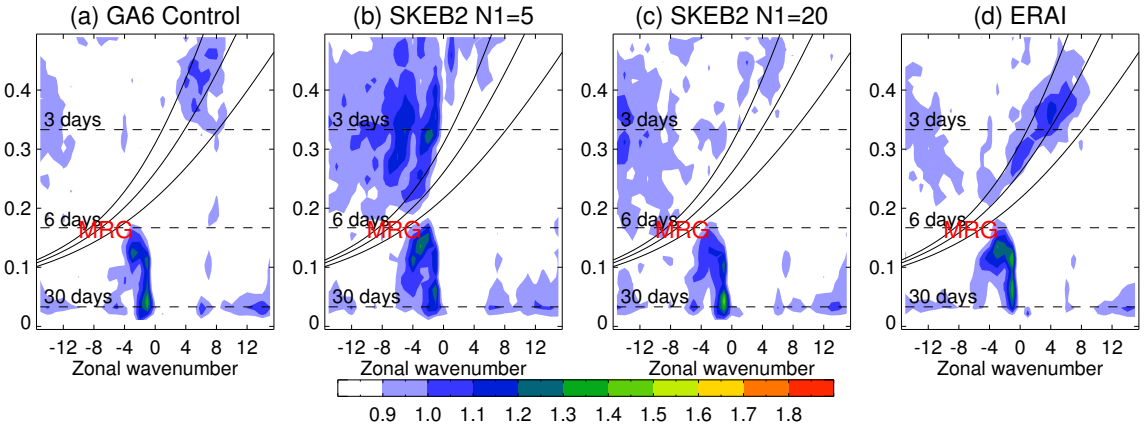

Figure 6.9: Background removed anti-symmetric spectra for zonal wind at 850hPa for (a) GA6 control (b) SKEB2 with $N_1 = 5$ at GA6.0 (c) SKEB2 with $N_1 = 20$ at GA6.0 and (d) ERAI. Same structure as Figure 5.20. SKEB2's experiments uses a $b_R = 0.2$ to highlight their impacts.

The nodes of SKEB2's forcing pattern with $N_1 = 5$ have a large extent in the tropics, with an oscillating behaviour with a wavenumber 5 (Fig. 6.1). The consistent forcing around these areas with a decorrelation time of 6 hours could produce the emergence of spurious patterns that propagate westwards following the flow in the high level winds, it is not clear why westward high level winds drives these anomalous patterns more than eastward low level winds, where the SKEB2 forcing is stronger. When the large-scales are excluded from the forcing pattern composition, the small scale forcing might produce similar perturbations but these are geographically smaller and maybe removed by the flow. However, these small scale perturbations could enhance small scale waves, Fig. 6.9.c shows the emergence of a wave of wavenumber 12 and period of 3 days, whose power is slightly higher than in

the control and observations (Fig. 6.9). This small wave seems to be a much weaker version of the spurious westward wave seen in Fig. 6.9.b, and it also has a shorter wavelength.

The dissolution or severe weakening of the spurious westward node does not diminish the SKEB2 improvements seen in the simulation of the tropical mean climate. Figure 6.10 shows the Outgoing Longwave Radiation (OLR) biases to the control simulation (Fig. 6.10.a) and the differences of the experiments done with the two N_1 values to control (Fig. 6.10.b,c). The impact of the SKEB2 with $N = 20$ is smaller, but the improvements are over the same geographical locations and the RMSE of the tropical mean is still quite beneficial to the control.

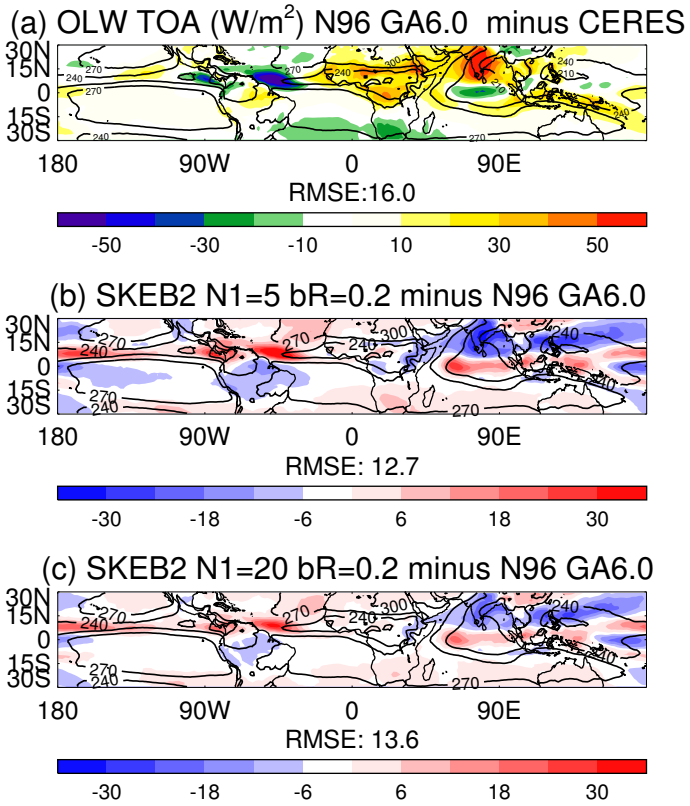

Figure 6.10: Outgoing Longwave Radiation (OLR) at The Top of the Atmosphere (TOA) for (a) N96 GA6 control minus CERES, (b) SKEB2 $N_1 = 5$ minus N96 GA6 control and (c) SKEB2 $N_1 = 20$ minus N96 GA6 control. Subtitle denotes tropical RMSE (30S-30N) and contours show OLR values for each $30\,W/m^2$ from $210\,W/m^2$. Simulations with SKEB2 employ a $b_R = 0.2$.

6.2 Biharmonic numerical dissipation

In the last chapter it was found that SKEB2 enhances the intensity of storms, which in principle is positive as it does offset the model's dissipation of such important weather systems. However, its increments are quite constant independently of the horizontal resolution, the storm intensity range, or their error (see Fig. 5.5, 5.8 and 6.8). In the Semi-Lagrangian dynamics (SL, see sect. B.2.1 for details), the diffusion introduced by the interpolation to the departure point diminishes when the horizontal resolution increases (Stratton, 2004), similarly strong storms are enhanced by a higher interpolation scheme (Sanchez *et al.*, 2013). SKEB2's numerical dissipation rate should decrease when horizontal resolution increases, as there is less energy diffused by the interpolation to the departure point.

SKEB2 numerical dissipation rate is based on the Smagorinsky subgrid non-linear diffusion scheme (Smagorinsky, 1963). There are other methods to compute the numerical dissipation such as the biharmonic dissipation, based on results from McCalpin (1988), which is included in the Spectral Stochastic Backscatter (SPBS, see sect. 2.2). A brief explanation of the methodologies to compute the numerical dissipation is given in the next subsection.

6.2.1 Numerical dissipation schemes

Local energy dissipation from unresolved turbulent process is driven by the turbulent transport and local dissipation. These two terms can be understood following an example with the one-dimensional advection-diffusion equation of $u(x,t)$ shown in 6.1, where the diffusion coefficient is defined as $K(x)$.

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} = \frac{\partial}{\partial x}\left[K\frac{\partial u}{\partial x}\right] \tag{6.1}$$

The energy equation is obtained by the product of eq. 6.1 and u, and is shown in eq. 6.2.

$$\frac{\partial \frac{1}{2}u^2}{\partial t} = -u\frac{\partial \frac{1}{2}u^2}{\partial x} + \frac{\partial}{\partial x}\left[K\frac{\partial \frac{1}{2}u^2}{\partial x}+\right] - K\left(\frac{\partial u}{\partial x}\right)^2 \tag{6.2}$$

The local change of kinetic energy, represented by eq. 6.2, is thus driven by the advection, the flux divergence of turbulent kinetic energy, and a negative-definite

dissipation term. Numerical dissipation schemes aim to estimate the effects of the subgrid dissipation term using different formulae.

Smagorinsky dissipation rate

The Smagorinsky dissipation rate is built from the viscous force **F** due to lateral stresses, equations 6.3 and 6.4 show the horizontal components of the force.

$$F_x = (k_H \Delta)^2 \left[\frac{\partial}{\partial x}(|D| D_T) + \frac{\partial}{\partial y}(|D| D_S) \right] \tag{6.3}$$

$$F_y = (k_H \Delta)^2 \left[\frac{\partial}{\partial x}(|D| D_S) - \frac{\partial}{\partial y}(|D| D_T) \right] \tag{6.4}$$

Where Δ is the grid-length, k_H is a numerical factor, D_S and D_T are the shearing and tension strains defined in equations 6.5 and 6.6 respectively, and D is the modulus of both strains (eq. 6.7).

$$D_S = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y} \tag{6.5}$$

$$D_T = \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \tag{6.6}$$

$$D = \sqrt{D_S^2 + D_T^2} \tag{6.7}$$

The rate of work of the viscous force is given by $uF_x + vF_y$, it is the equivalent of eq. 6.2 and thus is equal to the sum of the advection term, a flux divergence term and a pure dissipation term D_{num}, given by eq. 6.8. the "Smagorinsky dissipation rate" we employ for SKEB2 is given by D_{num}, where the k_H factor has been tuned to produce a global-mean energy dissipation of 0.7 $W\,m^{-2}$, the estimated energy dissipated at N216 and N144 (90km in the mid latitudes) horizontal resolutions (Tennant *et al.*, 2011).

$$D_{num} = (k_H \Delta)^2 D^3 \tag{6.8}$$

The vertical stress is ignored on the computation of the Smagorinsky numerical dissipation for SKEB2, but it is included in the more general Smagorinsky-Lilly turbulence closure (see section 4 of Smagorinsky 1963 for a discussion and evaluation of the different terms). The Smagorinsky turbulence scheme is widely used on

convection permitting models, models with grid-lengths of few kilometers able to partially resolve convective events and gravity waves. The seamless nature of the MetUM model enabled a straightforward adaptation of the Smagorinsky scheme from the high-resolution prediction system to compute the numerical dissipation for SKEB2.

The fact that the Smagorinsky scheme is employed for such fine resolution models poses a question about its adequacy for SKEB2, as the scheme is employed for systems such as MOGREPS or Glosea4 in MetUM, whose subgrid diffusion occurs at larger spatial scales. Furthermore, these MetUM systems do not include the Smagorinsky-Lilly turbulent scheme, so the estimation of the kinetic energy dissipated is not based on the model's formulation of subgrid turbulence, and its equivalence to the dissipation produced by the interpolation scheme is unclear.

There is the need to develop another way to estimate the numerical dissipation, which is more in agreement to the dissipation produced by the interpolation to the departure point, and operates in scales typical of an intermediate resolution GCM. The Biharmonic dissipation rate is a good candidate, its formulation is briefly explained here.

Biharmonic dissipation rate

The study carried out by McCalpin (1988) compared the dissipation inherent to the SL advection to more traditional forms of dissipation such as Laplacian or Biharmonic eddy viscosity. The magnitudes of the amplification factor λ for various interpolation schemes were used to compute a more traditional measures of viscosity, finding that the linear interpolation to the departure point results in dissipation which is effectively Laplacian, while quadratic and cubic result in Biharmonic dissipation.

The SPBS uses the Biharmonic diffusion of rotational modes, given by eq. 6.9, where ξ is the relative vorticity, K is the Biharmonic diffusion coefficient and α_{num} is a factor to scale the dissipation rate, currently set to 3 (Palmer *et al.*, 2009).

$$D_{num} = \alpha_{num} K |\nabla \xi|^2 \tag{6.9}$$

Prior to the implementation of the dissipation rate, we explore the dissipation from the interpolation to the departure point of different sources:

- Rotational flow: The large-scales are driven by ageostrophic motions that are mostly rotational in the mid-latitudes, where the SKEB2 perturbations may be quite beneficial to counteract the numerical dissipation of storms (as shown in 5.1.3). The dissipation term is proportional to $|\nabla\xi|^2$.

- Divergent flow: It is more associated to mesoscales. Some studies associate the lack of the $k^{-5/3}$ spectra to divergence motions (see sect. B.2.2). Their dissipation by the SL scheme is proportional to $|\nabla\chi|^2$, where χ is the divergence horizontal wind. It is obtained following two different formulae, the pure divergence computation as in eq. 6.10, which may introduce second order errors, or the more accurate computation of divergence through the continuity equation as shown in eq. 6.11, where ρ represents the density.

$$\chi = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \tag{6.10}$$

$$\chi = -\frac{1}{\rho}\frac{\partial}{\partial z}(\rho w) \tag{6.11}$$

- Vertical motions: Defined by the vertical velocity w, it is controlled by parametrized processes such as convection so in principle it may have important effects. Their formulation follows McCalpin (1988) and is given as the Finite Difference Equation 6.12, where k is the model level, i denotes the horizontal dimension and α is the fractional Courant number, $\alpha = w\,\partial t/\partial z$.

$$D_Z = \alpha\frac{(u_{i,k+1} + u_{i,k-1} - 2u_{i,k})^2}{\Delta x_i^4} \tag{6.12}$$

The three terms are multiplied by the diffusion tensor K_i (as in eq. 3.4 of McCalpin 1988). $K = \frac{3}{128}\frac{\Delta x_i^4}{\Delta t}$

The contribution from these terms in comparison to the Smagorinsky rate are given in Figure 6.11, the zonal mean of the numerical dissipation temporally averaged for 3 days on an arbitrary day like the 1^{st} of September 1988. A N96 MetUM simulation have been employed to compute the dissipation rates. The Smagorinsky numerical dissipation scheme (Fig. 6.11.a) produces large values on the high latitudes and over the jet levels ($\sim 10km$). The rotational biharmonic dissipation

estimates less dissipation throughout the troposphere with the exception of the subtropics on the NH (Fig. 6.11,b). The addition of the dissipation from divergent motions increases the dissipation rate everywhere with a strong impact in the tropics (Fig. 6.11.c), where the Smagorinsky dissipation is too weak. Both methods to estimate the divergent component produce similar results (Fig. 6.11.c,d). The effects of the vertical interpolation on the total Biharmonic dissipation rate are very weak and mainly concentrated right below the Tropical Tropopause (Fig. 6.11.e,f).

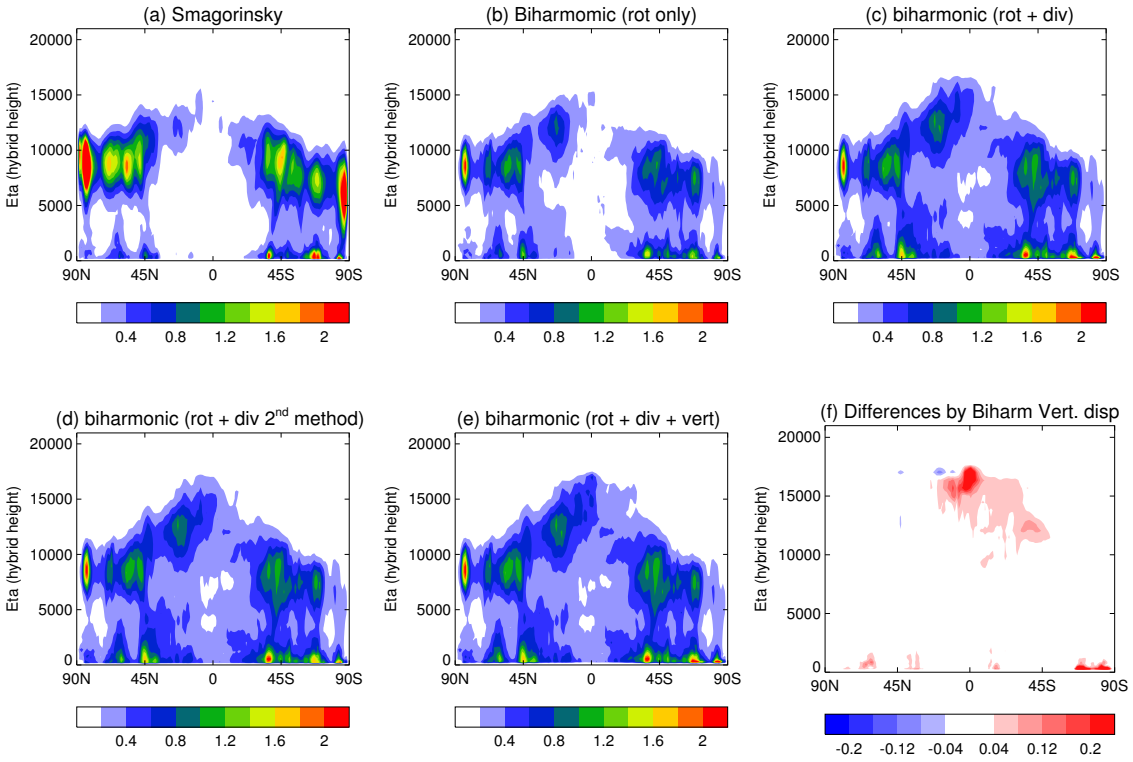

Figure 6.11: Zonally averaged Numerical dissipation rate ($10^{-3}\, m^2\, s^{-3}$) for 3 day averaged between 1/09/1988 and 4/09/1988, (a) Smagorinsky (b) Biharmonic rotational term, (c) Biharmonic rotational plus divergent terms (d) Biharmonic rotational plus divergent (based on eq. 6.11) terms (e) Biharmonic all terms (div base on 6.11) (d) Biharmonic vertical dissipation contribution as differences between (e) and (d).

The results obtained from the comparison of different dissipation terms recommend us to ignore the vertical dissipation for the computation of the total Biharmonic dissipation rate. Their effects are weak, moreover the interpolation scheme employed is subjected to changes to enhance the performance of the model over the Tropical Tropopause Layer (TTL), a very sensitive area that controls the Brewer-

Dobson circulation, input of water vapour in the stratosphere and other important drivers of the middle-atmosphere climate. The vertical interpolation of the departure point has recently been changed from tri-cubic to a more accurate Hermite-cubic to remove a temperature bias over TTL in the GA6 configuration (Walters D, 2013, *personal communication*). The dissipation of the divergent flow produced by both methods is very similar for the example given, but we prefer to use the most accurate description (eq. 6.11). The combination of the rotational and divergent dissipation rates forms the final biharmonic dissipation. It is compared to the Smagorinsky rate in the next section.

Comparison of both dissipation rates

One of the most positive features of a numerical dissipation rate (D_{num} hereafter) would be its scalability across different horizontal resolutions. This could enhance the storm intensity at low resolutions, and diminish it at high resolutions, improving their representation. The resolution sensitivity could be measured by the global-average of the vertically integrated D_{num}, so we set 4 forecasts of winter and summer with 6 hourly output (~ 80 points) of vertically integrated D_{num}. Figure 6.12 shows the D_{num}- resolution relationship. Biharmonic produces a stronger dissipation rate for the low resolutions, especially for the NH (Fig. 6.12.c).

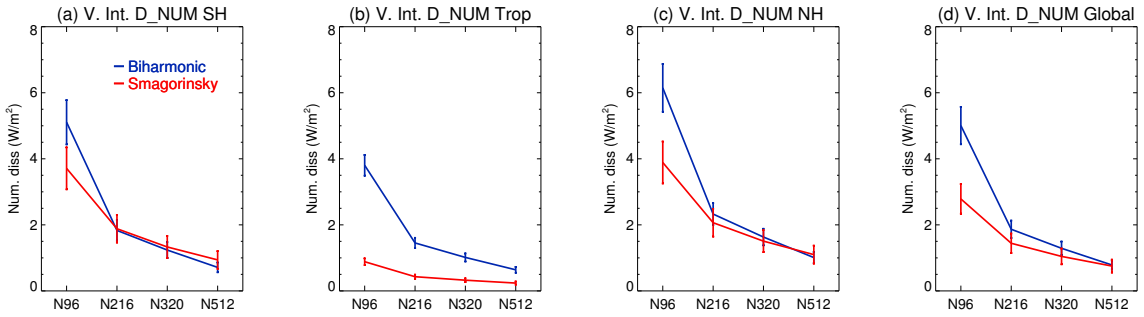

Figure 6.12: Vertically Integrated global-average Numerical dissipation rate ($W\,m^{-2}$) for Biharmonic (blue) and Smagorinsky (red), (a) SH, (b) Trop, (c) NH, (d) Global. See text for details.

The vertical distribution of D_{num} is heterogeneous, as shown in Fig. 6.11, therefore high dissipation in the BL could be masking low dissipation over the jet levels. Figure 6.13 shows the D_{num} over two different level ranges, the Boundary layer (BL, $z < 1.5km$) and the jet stream levels ($6 - 9\,km$). For the Smagorinsky rate the

contribution from the jet levels is higher at low resolution and drops faster with increasing resolution, whereas for the Biharmonic dissipation the BL contributions dominates over the jet levels. D_{num} at both vertical ranges drop with similar slopes when resolution increases.

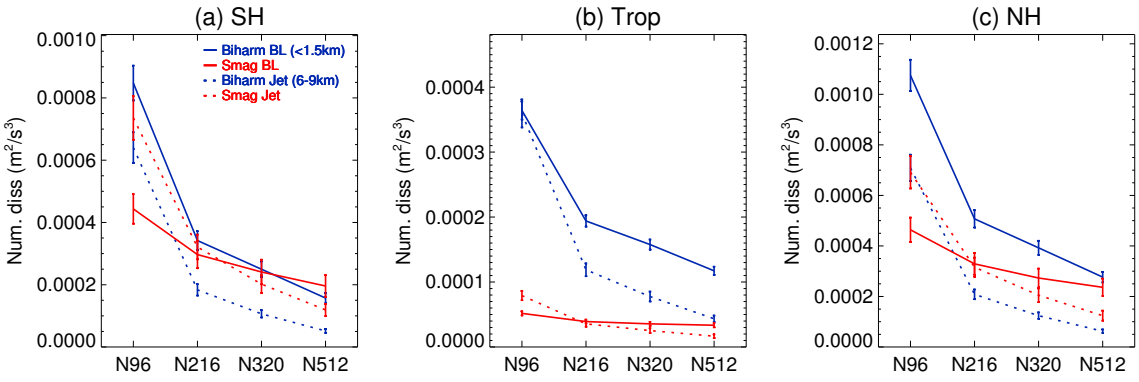

Figure 6.13: Global mean numerical dissipation rate averaged between two level ranges, Jet (solid) and BL (dotted). (a) SH (b) Trop (c) NH.

The biharmonic dissipation scheme provides some benefits to the Smagorinsky; better scalability across resolutions, and higher rates over the Boundary layer, where it is ought to occur more Kinetic Energy dissipation due to a higher density small scale eddies driven by the mixing in the Boundary Layer. Although it is not clear how the interpolation to the departure point could dissipate these.

For model stability reasons, the SKEB2 scheme have a logarithm tapering for low levels from 2km, so it could essentially reduce the impact of the scheme if the Biharmonic dissipation is taken in. A series of MOGREPS experiments with the different dissipation rates are carried out to find out if there is a significant impact on the ensemble spread (see sect. 4.1.1 for a description of the system and its setup).

6.2.2 Impact on MOGREPS scores

The Smagorinsky and Biharmonic numerical dissipation rates could produce differences in the ensemble spread, as their maximum estimation of energy losses occurs over different areas, Biharmonic "sees" more energy in the lower levels and Smagorinsky does it over the jet stream levels. In order to investigate their differences and impact on the EPS performance, several MOGREPS experiments are run with a different vertical extent. SKEB2's full vertical range ($z < 12\,km$), forcing restricted to

the Boundary Layer (BL, $z < 1.5\,km$), or Jet levels ($6 < z < 9\,km$). All the experiments have $N_1 = 20$ as it has been shown to be more positive for the representation of the mean intensity of mid-latitude cyclones and produce a better tropical climate with high b_R (see sect. 6.1).

The above mentioned differences of the dissipation rates around BL or Jet levels clearly produce and impact on the ensemble spread of low and high level winds, as shown in Figure 6.14. Biharmonic produces more spread than Smagorinsky for both levels in the tropics, as well as a higher spread in the boundary layer and lower spread at jet levels.

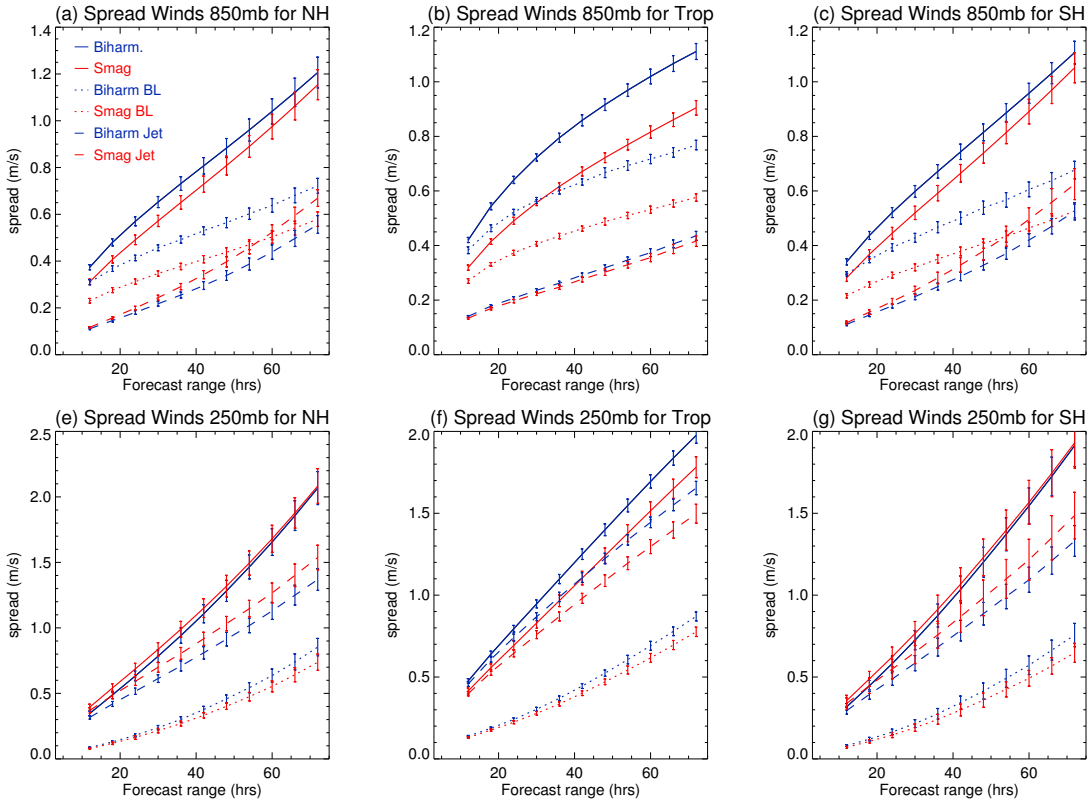

Figure 6.14: Ensemble spread for different SKEB2 experiments, blue includes Biharmonic dissipation and red Smagorinsky, solid line indicates SKEB2 operating across its full vertical range, dotted line shows experiments where SKEB2 is active only on the boundary layer and dashed on Jet levels (see text for details). (a) NH low level (850hPa) winds (b) Tropical 850hPa winds (c) SH low level winds (d) NH upper level (250hPa) winds (e) Trop. upper level winds (f) SH upper level winds.

In the low level winds, the contribution from the Boundary Layer has a higher impact over the spread on the first two days of the forecast (slightly longer for the tropics), afterwards the spread from the Jet levels is higher (Fig. 6.14). In the high

level winds, the difference between the spread produced in the BL and Jet levels is substantial, although the additive effect of the extra-spread done by the Biharmonic in the boundary layer compensates its lack of spread in the jet levels, with an overall spread slightly smaller than Smagorinsky for the mid-latitudes.

Although the spread in high level winds could be similar between Biharmonic and Smagorinsky dissipation rates, the spread at Z_{500} is lower from the former (Fig 6.15). The large-scale structures such as cyclones or blocks are more perturbed by the Smagorinsky rate than Biharmonic, as these are mainly driven by perturbations over levels where the jet stream flows.

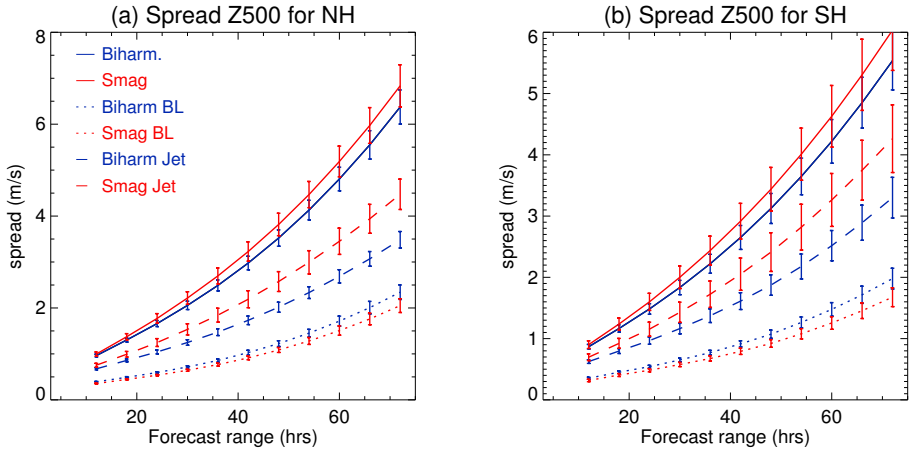

Figure 6.15: Ensemble spread for Z_{500} (a) NH and (b) SH. Same line distribution as Figure 6.14.

The biharmonic dissipation impact is smaller over mid-latitude large-scale structures as the spread decreases for Z_{500}. However, the error of the ensemble does not seem to degrade when perturbations are only applied to the BL (not shown). The structure of the power spectra of the ensemble spread of Z_{500} is quite similar for both rates, with higher amplitude for the Smagorinsky rate (not shown).

Another useful technique which we could include in future work is the "relaxation" (a.k.a. nudging). A particular prognostic field like winds is relaxed towards the observations in a given area or level. By relaxing the model towards reality in certain regions, such as the Indian Ocean, or on the Boundary Layer, it would be easier to localize the sources of error and spread in the system, providing useful information about which processes need to be perturbed to produce a reliable source of model error. A nice example where the relaxation technique is used to investigate

the origin of extended-range predictability of the Integrated Forecast System (IFS) is given in Jung *et al.* (2010a), their study also provides a detailed description of the technique.

6.2.3 Impact on climate scales

Despite the lack of spread in the Z_{500} field for SKEB2 with the biharmonic rate, the RMSE of the mean climate shows slightly better scores except for SH JJA (Table 6.3), the only case where SKEB2 degrades the control. This improvement may be caused by the stronger forcing of the Biharmonic dissipation at low resolution, in comparison to Smagorinsky (Fig. 6.13.a,c). Results from cyclone tracking (see 3.3.1 for a description on the technique) do not show any major difference amongst the three experiments (not shown).

	NH DJF	NH JJA	SH DJF	SH JJA
Control	32.6	26.3	23.8	38.1
SKEB2 Smagorinsky	31.6 (+3%)	25.3 (+4%)	22.3 (+6%)	39.3 (−3%)
SKEB2 Biharmonic	30.3 (+7%)	24.6 (+6%)	22.2 (+7%)	39.8 (−4%)

Table 6.3: Z_{500} RMSE of the climate mean for different experiments, seasons and hemispheres. In brackets difference to control, pluses denote an improvement and minus a degradation.

In terms of tropical climate, the capacity of the biharmonic dissipation to detect energy dissipation in the tropics produces a larger forcing which benefits the organization of convection, and the global-average fields of high level divergence, OLR, clouds and precipitation following the same mechanisms as described in sect. 5.2. These effects of the biharmonic dissipation are detailed when the new numerical dissipation rate is combined with the convective dissipation factor in the next section.

6.3 Convective dissipation factor

The modulus of the numerical dissipation rate and the convective dissipation rate forms the D_{TOT} function, which SKEB2 uses as an estimate for the total energy

dissipated or missing by the model (eq. 2.1). The inclusion of the Biharmonic numerical dissipation makes the scheme more scalable across horizontal resolutions (see section 6.2), but the impact of the scheme in the tropical high resolution is too high (see Fig. 5.1), so additional sources to modulate the scheme over these regions are desirable.

The simplest solution to modulate further the impact of SKEB2 in the tropics is to add an amplitude factor F_{CONV} for the convective dissipation rate. This factor is proportional to the root square ratio of the horizontal resolution, denoted by N, to the standard resolution of 216, as shown in eq. 6.13. this formulation was preferred over the linear one as the dissipation is a two dimensional field, the 216 value is chosen because it is the operational resolution for MOGREPS and thus the factor has no impact on the operational system. The definition of the convective factor F_{CONV} is ad-hoc. Future research should focus on building a proper representation of unrepresented KE released by the convective parametrization. One possible candidate is the coarse-graining of the divergence field (Shutts G, 2014, *personal comm.*)

$$F_{CONV} = \sqrt{\frac{216}{N}} \tag{6.13}$$

The inclusion of F_{CONV} helps to modulate the convective dissipation rate, as it decreases when horizontal resolution increases (see Fig. 6.16,a,b,c,d). The rate without the factor produces a similar estimation independently of resolution. A new "improved" version of the scheme includes the convective dissipation factor and the Biharmonic dissipation, it produces Kinetic Energy (KE) perturbations whose global values scale better across resolutions (Fig. 6.16,e,f,g,h), specially in the tropics, where it is constant across resolutions for the "default" scheme. This is probably related to the fact that the Smagorinsky rate is very small in the tropics, whereas Biharmonic could add more resolution dependency to the energy estimation.

A configuration of SKEB2 with the Biharmonic dissipation and F_{CONV} could potentially help to offset the excessive dissipation of wind seen for the low resolution, or reduce the impact of the scheme at high resolution where there are some symptoms that SKEB2 could be backscattering too much energy, such as the too strong winds over the Maritime continent (see sect. 5.1 for a description and discussion of these

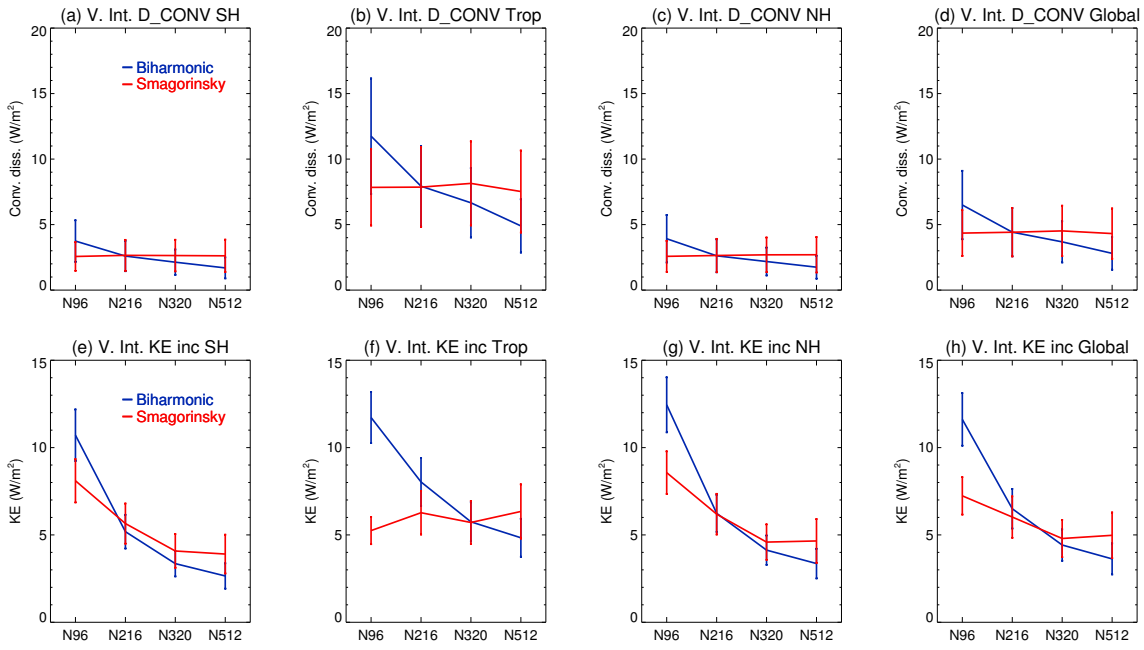

Figure 6.16: Global-average of the vertically integrated convective dissipation (a), (b), (c), (d) and KE increment by SKEB2 (e),(f),(g),(h)

results).

In principle, a stochastic convective parametrization should produce stronger forcing for higher resolutions as the equilibrium assumption is weaker. Smaller gridboxes contain fewer convective clouds and thus their averaged effects are less consistent and more prone to fluctuations (e.g. PDF of mass fluxes becomes wider when resolution increases as shown in Fig. B.5). However, the KE backscattered from the misrepresentation of missing eddies around convective elements should be lower when resolution increases, as there is less energy unrepresented per gridbox.

The combination of biharmonic and convective dissipation, in addition to the explored increase of the lowermost wavenumber of the forcing pattern from 5 to 20, is defined as "SKEB2 improved" and is explored in this section. We make use deterministic forecasts and climate simulations at N96 resolution (these systems are described in sect. 4.1.1). For comparison, the default version also includes the increase of the lowermost wavenumber N_1, although it conserves the Smagorinsky dissipation rate and the convective dissipation rate is not modulated by the resolution dependent factor F_{CONV}.

6.3.1 Effects on NWP forecasts

The skill of the NWP forecasts is represented by the RMSE of the low level winds (Fig. 6.17). For the tropics, the high resolution forecasts with SKEB2 are no longer the worst performing case, the increase of N_1 helps to decrease the SKEB2 forcing on the tropics (Fig. 6.17.b). In general terms the improved version of SKEB2 generates more error for low resolution, as the scheme's forcing is higher, but for high resolutions the level of error is similar for both versions of the SKEB2 scheme, since the KE increments are equivalent or slightly lower for the improved version (see Fig. 6.17.e,g,h).

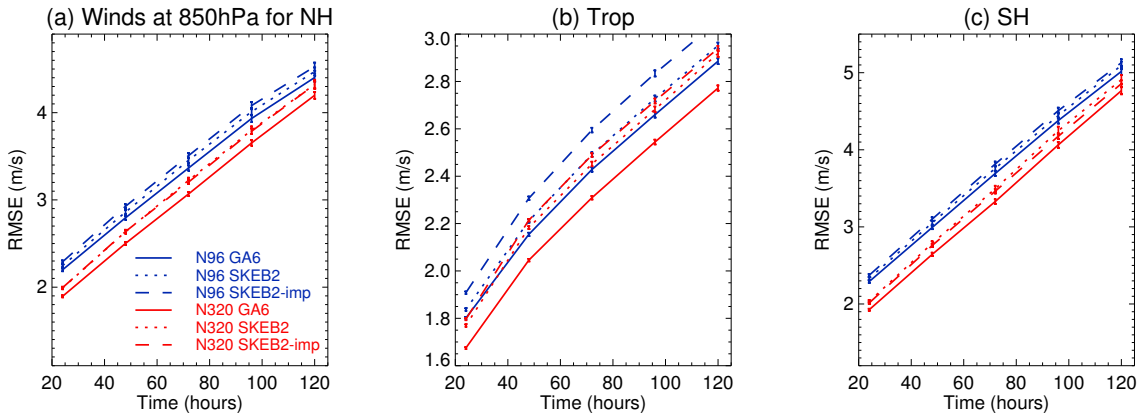

Figure 6.17: RMSE of winds at 850hPa for (a) NH (b) Tropics (c) SH. Red line shows N320 and blue N96. Solid line shows control, dotted SKEB2 with the default version and dashed improved version.

The intensity of extra-tropical cyclones shows better sensitivity with the new version of the SKEB2. The N_1 increase reduces the intensities of all storms, therefore differences between different experiments are no longer significant due to the high numbers of experiments required to obtain significant statistics from the storm matching technique. The storms produced by the GA6 simulations are still too weak for the low resolution, and the improved-SKEB2 increases their intensity more than the default version, this effect is clearer in the NH (Fig. 6.18.a). The forecasts done at high resolution no longer have problems in simulating the mean intensity of cyclones for the NH, their storm intensities match the analysis very well. However, SKEB2 increases the intensity making the storms slightly over-active, the improved version of SKEB2 reduces their intensity (Fig. 6.18.a), but for SH where there is still some dissipation and the difference between the two configurations is marginal.

The improved version of SKEB2 shows better sensitivity to the dissipation of storms intensities, in terms of resolution and in terms of their intensity range, where there is a slightly increase of intensity for the stronger storms (not shown).

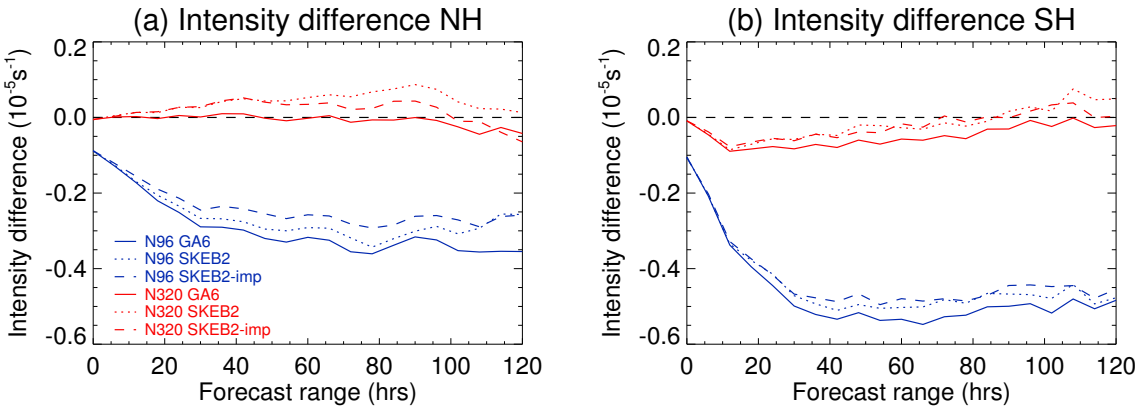

Figure 6.18: Mean cyclone intensity difference to EC analysis (a) NH, (b) SH. Same line distribution as Figure 6.17.

6.3.2 Effects on tropical climate

The scheme's impact on the tropics has substantially increased for low resolutions, as the total dissipation rate D_{TOT} and thus the KE increments are larger (Fig. 6.16.b,f). In the last chapter, it was found that a higher impact leads to a better representation of the divergent tropical flow, leading to an improved OLR and other convective-coupled fields such as precipation or clouds. However, it also provoked the emergence of a spurious westward antisymmetrical wave (see section 5.2 for more details), the spurious wave is removed when a large-scales are removed from the forcing pattern (see sect. 6.1). It is therefore useful to quantify the impact of the scheme in terms of the representation of the climate mean OLR and antisymmetric CCEW,

The representation of tropical anti-symmetric waves is slightly degraded (Fig 6.19), the signal of the spurious westward wave introduced by the unbalanced perturbations to the scheme is present with the improved version of the scheme, which is more powerful than the default version (Fig. 6.19.b,c). In addition to an excessive power of high wavenumber waves of 3 day period, the representation of eastward Mixed Rossby-Gravity waves of high temporal frequency is weaker, degrading the representation of a wave node which is already too weak for the control. The im-

proved version of SKEB2 diminishes the power of Kelvin waves (not shown).

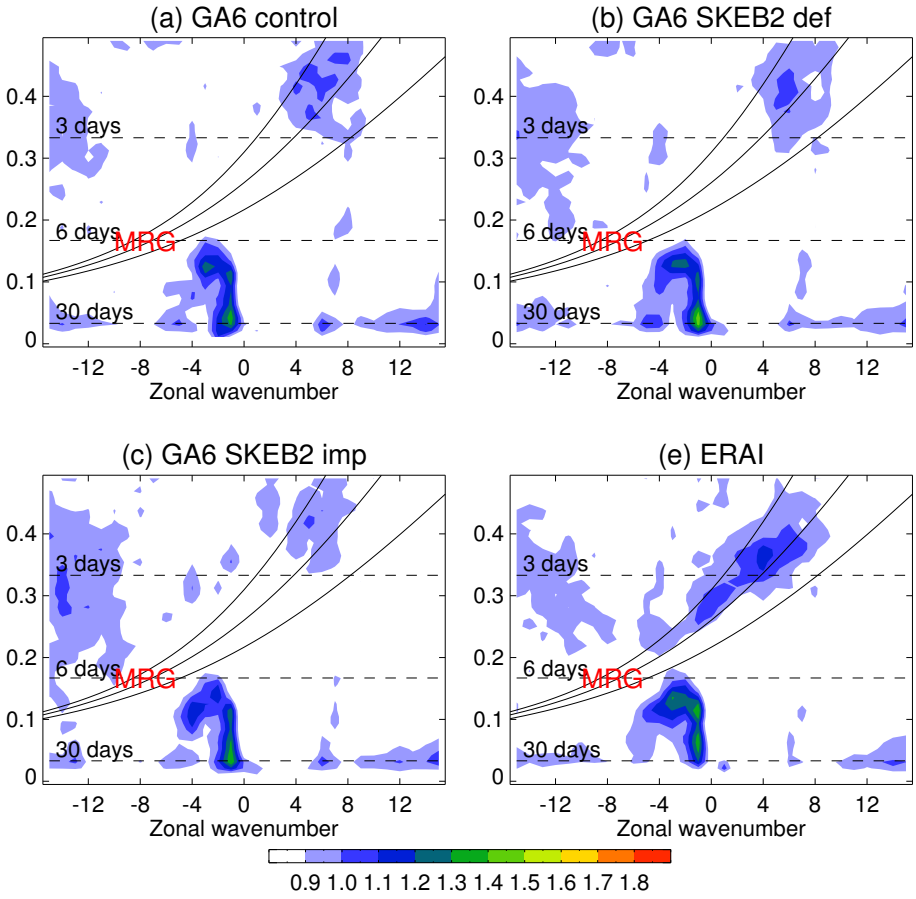

Figure 6.19: Background removed anti-symmetric power spectra of low level winds, as Figure 6.9, (a) GA6 control (b) SKEB2 default version for GA6 (c) improved version of SKEB2 for GA6 (e) ERAI reanalysis,

The improved version of SKEB2 is able to reproduce the improvements in the tropical climate seen over the last chapter (see sect. 5.2). Figure 6.20 shows the OLR for the control, SKEB2 default and SKEB2 improved. The impacts of the improved version are not very high in respect to the default version, but its effects over the excessively dry Indian subcontinent and South-East Asia are clearer and statistically significant, as well as the drying over the tropical West Atlantic, and area with excessive convection. Improvements over similar regions occurs for precipitation and high-level clouds (not shown). These are beneficial except east of Philippines where the model is already too moist.

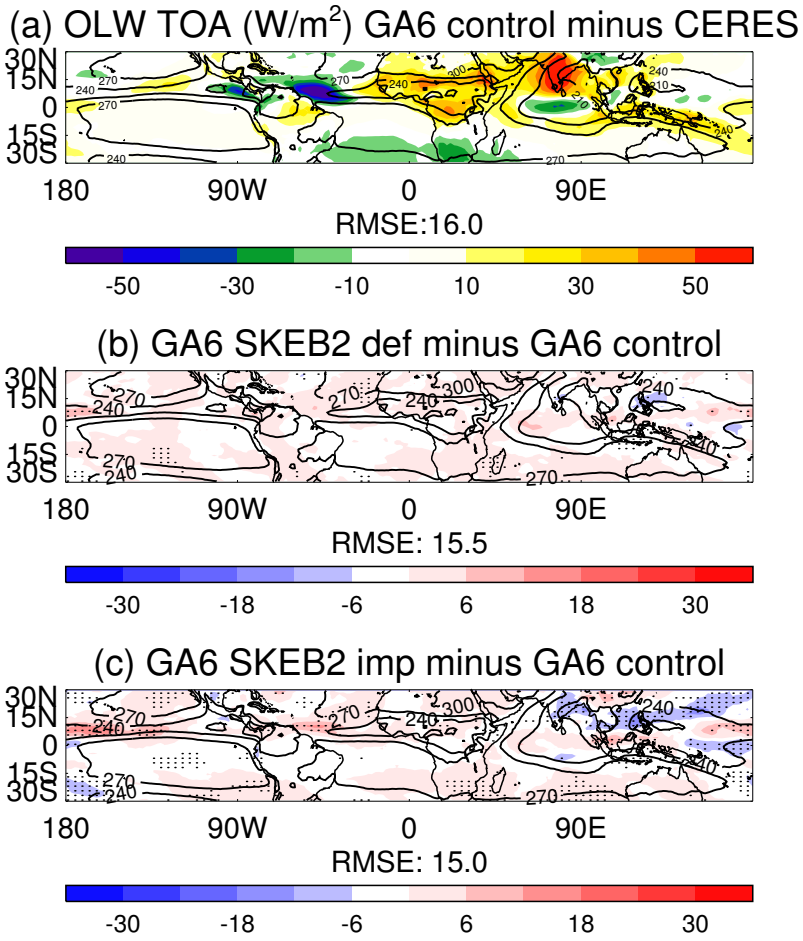

Figure 6.20: mean OLR at TOA for (a) GA6 Control minus CERES (b) GA6 SKEB2-default minus GA6 control (c) SKEB2-improved minus GA6 control. Dotted denotes differences significant at 95% using a t-student test.

6.4 Conclusions

The evaluation of SKEB2 carried out in chapter 5 highlighted some of its deficiencies, like the emergence of a spurious westward tropical wave when the forcing of the scheme is too high, or the deterioration of the deterministic skill. In the present chapter, we have investigated different solutions to mitigate and if possible remove these problems, aiming to develop a newer version of the scheme that improves the simulation of kinetic energy backscatter and thus produces a more realistic representation of model error.

SKEB2's perturbations on well resolved planetary scales on the short term is one of our main concerns, therefore we explore the effects of the scheme's large-scale forcing by removing lower wavenumbers in the spherical harmonic decomposition of the forcing pattern. We explore the differences of two sets of experiments with

different N_1, the lowermost wavenumber for the spherical harmonic forcing pattern. The control experiment uses N_1 equal to 5, its default value when large-scales are included. For the second experiment, N_1 is set to 20, following recommendations from Shutts (2013). The main results are:

- The forcing pattern with $N_1 = 20$ is spatially less homogeneous than with $N_1 = 5$, and its amplitude halves.

- The reduction of the scheme's amplitude leads to a reduction of the deterministic RMSE for the $N_1 = 20$ experiment.

- The ensemble spread is reduced when N_1 is set to 20, slightly increasing the ensemble mean error of winds. The spread grows in different scales than the error, which grows faster on synoptic scales whereas the spread grows in the large-scales.

- The representation of storms for the ensemble with $N_1 = 5$ is poor. The spread is too little in comparison to the positional and intensity errors, and the intensity of well predicted storms is nearly overdispersive at the end of the forecast. The experiment with $N_1 = 20$ does not clearly improve these aspects, although it reduces the ensemble mean intensity of storms in the NH, which is too high.

- The anti-symmetric westward spurious tropical wave is removed in the $N_1 = 20$ experiment, although it seems that it produces another spurious wave with lower wavelength and weaker power. The background removed anti-symmetric power spectra is more realistic with $N_1 = 20$.

- The $N_1 = 20$ decreases the positive impacts on the tropical mean climate, although these are still very beneficial in comparison to the control.

The perturbations from SKEB2 scheme are not very sensitive to resolution, error or the magnitude of the diffusivity. The scheme uses masks to estimate the location of numerical diffusion and the unrepresented creation of convective kinetic energy, but these masks do not seem to be giving the right magnitude for the local energy backscattered. Therefore a new method to estimate the numerical energy dissipation is developed and compared to the current Smagorinsky formulation. The new

method is made from contributions of Biharmonic dissipation of the rotational and divergent flow, which according to McCalpin (1988) are equivalent to the dissipation produced by the cubic interpolation to the departure point in the Semi-Lagrangian scheme. A vertical dissipation component is also explored but their contribution is negligible. The main results from the comparison between both dissipation rates are:

- The vertically integrated global mean of the Biharmonic rate is more sensitive to horizontal resolution, it produces higher values for low resolution and also a stronger forcing in the tropics.
- Biharmonic produces a larger impact over the low levels, whereas Smagorinsky has a deeper impact for levels where the Jet stream flows.
- On climate scales, the Z_{500} climate RMSE is slightly more positive for the Biharmonic dissipation for nearly all seasons and hemispheres.

SKEB2 produces impacts that are too large on the tropical belt at high resolution, degrading the skill of deterministic forecasts. The KE upscaled from eddies around convective cores should decrease when horizontal resolution increases, as less energy is contained by the unresolved scales. A simple ad-hoc numerical dissipation factor is included in the scheme. It is based on the ratio between the horizontal resolution and a standard resolution. A version of the stochastic scheme that incorporates this factor, plus the Biharmonic dissipation, can produce the following differences to a default version with the Smagorinsky rate.

- The new version increases the deterministic forecast error for low resolution, as the impact of the scheme is higher. It slightly improves the intensity of mid-latitude cyclones. For high resolution the impacts are neutral, but it slightly decreases the intensity of storms, which they are a bit over-active.
- The representation of the tropical waves is slightly deteriorated with the new scheme, as it weakens Kelvin and Mixed Rossby-Gravity waves, it also increases the power of short waves of 3 day period excessively.
- The representation of the tropical climate is slightly improved, differences between the new version and control are statistically significant over regions where there are severe biases for the divergent flow, OLR or precipitation.

The improvements proposed removes the major setbacks of the SKEB2. The spurious westward tropical wave is removed when the large waves are not included in the forcing pattern, and the resolution sensitivity of the scheme improves with the Biharmonic numerical dissipation and the resolution dependent factor for the convective rate. However, these solutions are also limited and do not improve many other aspects of the flow:

- SKEB2 produces ensemble spread over the wrong scales, more in the large-scale than on synoptic scales, where the ensemble mean error grows. The new improved version is unable to fix this, even if the scheme with the biharmonic rate is only active in BL levels.

- The scheme still has a detrimental impact on deterministic forecasts. Their RMSE are still high, although it has been reduced for high resolutions from the scheme's version employed in chapter 5.

- There are small differences between the new and default version of SKEB2 for the representation of mid-latitude cyclones, many deficiencies remains such as positional errors.

- Although the spurious westward wave has been removed, the scheme still dissipates the power of Kelvin waves, in contrast to the SPBS as shown by Berner *et al.* (2012) and described in sect. 2.5 of the thesis.

It is hard to tell how good the SKEB2 representation of Kinetic Energy Backscatter is. Despite of recent studies by Shutts (2013) and Thuburn *et al.* (2013), described in section 2.2.1, there is little understanding of its internal mechanisms. The scheme could be improved by using a more realistic convective dissipation rate, built from coarse-graining studies. Another option to improve the scheme could be the addition of a temperature forcing like CASB or the Canadian SKEB (described in sect. 2.2.2).

The importance of the SKEB schemes seems to wane as the atmospheric community are employing models with higher resolutions and better dynamics and physics. The numerical dissipation does not seem to be an issue anymore for resolutions of N216 onwards for GA6, as it simulates well the mean intensity of storms, making

unnecesary the numerical contribution of SKEB2. It is therefore important to include other aspects of model error in our research, such as deficiencies in the physical parametrizations of key processes.

In the next chapter, the uncertainty of physical parametrizations is explored following different ideas, such as stochastically perturbing physical tendencies or stochastically perturbing parametrization's internal parameters.

Chapter 7

Stochasticity in physical processes

The major source of model error comes from crude assumptions made in the construction of physical parametrizations, some of them described in Appendix B of the thesis. Several stochastic schemes have been developed to represent the uncertainty emerging from the internal deficiencies of parametrizations, the so-called "internal uncertainty" schemes. These are the Random Parameters scheme v2 (RP2, see sect. 2.4) or Cellular Automata (CA, see 2.6.3). Other stochastic schemes aim to represent the structural uncertainty of the parametrization, aiming to break general assumptions such as the quasi-equilibrium approximation, such schemes are the Stochastic Perturbation of Physical Tendencies (SPPT, see sect. 2.3) or the Stochastic Convection (SC, see 2.6.2). There is also the possibility to perturb what the parametrization "sees", the initial state.

In the present chapter we explore different options to represent model error and compare their positive and negative impacts. We aim to estimate the realism of the physical perturbation and their effects on the ensemble dispersion and climate processes, following the methodology applied throughout the thesis (see chapter 3 for a description). The different options considered in the chapter are:

- Development and evaluation of a scheme that perturbs the initial state (section 7.1). It represents uncertainty coming from previous timesteps. This methodology leaves the structure of parametrization untouched and therefore it conserves its internal framework. The scheme is defined as Stochastic Initial State for Parametrizations (SISP).

- In section 7.2, the Stochastic Perturbed Tendencies (SPT) is developed following the SPPT template (sect. 2.3) and compared to other stochastic schemes.

We explore different options for the scheme, such as applying conservation constrains (described in Appendix A), or leaving clear sky radiation unperturbed.

- The uncertainty of one scheme, the Mixed-Phase Cloud scheme (Field *et al.*, 2014), is explored using different approaches to randomly perturbed important parameters that shape the Probability Density Function (PDF) of the supersaturation distribution, or simply perturbing the cloud fraction and cloud water content following the SPPT approach (sect. 7.3).

Some of these experiments and comparisons are new to the research field of stochastic-physics. There is no comparison between the Random Parameter approach and Stochastic Tendency perturbation at a general level or to a single scheme. This is an important comparison because these schemes are starting to be developed for convective-resolution models (Bouttier *et al.*, 2012), and therefore employed to predict very detailed weather.

7.1 Perturbing the Initial State

Physical parametrizations have been built using some crude approximations and poorly constrained assumptions. However, it has been a substantial amount of work put on their development and thus they have the capacity to produce a realistic first-order representation of physical processes. Schemes such as SPPT neglect most of the complexity and the internal mechanisms of the parametrizations, in addition they do not represent the uncertainty in the triggering of local processes (e.g. convection) as it just amplifies or diminishes the local tendency.

There is one possible stochastic scheme that would respect the structure and assumptions of parametrizations and represent subgrid fluctuations. It is a scheme where the initial state is perturbed. Parametrizations may see different initial states where convection, non-orographic gravity wave or rainfall could be triggered in different ways and locations, which might lead to increments in the local tendencies that could represent the uncertainty in the location of these events. A simple version of such scheme has been developed during the work of this thesis and is defined as Stochastic Initial State for Parametrizations (SISP).

The SISP scheme inherits a spherical harmonic decomposition forcing pattern from SKEB2 (see sect. 2.2.4 for its description). The power law is modified to be Gaussian (given in eq. 7.1), where $< F_{\Psi}^2 >$ is the standard deviation of the forcing pattern, σ is the random number variance, $\Sigma(N)$ a normalization factor and β a relation between the radius of the Earth and the decorrelation scale equal to 500 km.

$$g(n) = \frac{< F_{\Psi}^2 >}{2\sigma\Gamma(N)} exp[-\beta n(n+1)] \tag{7.1}$$

Each prognostic variable of the initial state, horizontal winds u and v, specific humidity q and potential temperature θ, is perturbed using a simple formulation as shown in eq. 7.2, where X' is the perturbed initial state, X is the original state and F_P is the forcing pattern. Experiments are carried out where two different forcing patterns are used for slow (gravity wave drag, microphysics and radiation) and fast physics (convection and boundary layer). The second forcing pattern is a mathematical 180°longitudinal translation of the original forcing pattern. The perturbations are applied between level 7 (400 m) and level 42 (12040 m) to prevent numerical instabilities close to the surface and stratosphere, as well as to reduce the scheme's computational cost.

$$X' = (1 + F_P)\, X \tag{7.2}$$

If the perturbed relative humidity q is above saturation, the super-saturation is converted to specific liquid water ratio following the Classius-Claperyon equation, and the cloud ratios are modified within the PC2 scheme (Wilson *et al.*, 2008). In a second version of the SISP scheme, different $< F_{\Psi}^2 >$ could be used for different variables, targeting more those that are more uncertain or have less physical constrains, like winds in opposition to relative humidity.

In order to explore the impacts of this new scheme in a EPS context, we employ a previous version of MOGREPS at GA3 (more precisely GA3.1, see section 4 of Walters *et al.* 2011 for a description of GA3.0/GA3.1 differences). Horizontal resolution is N216. Initial conditions are perturbed using the Ensemble Transform Kalmar Filter (ETKF, Bowler *et al.* 2009). Probabilistic forecasts are run every 12 hours from 20 December 2010 to 9 January 2011. Several flavours of the SISP are

tested. The power law $g(n)$ (eq. 2.2) has been modified to include a fixed standard deviation for the forcing pattern $< F_\Psi^2 >$ (eq 7.1) for all variables equal to 0.01, a second experiment uses a "scaled" version where T and horizontal wind components have got $< F_\Psi^2 >$ equal to 0.05 and q equal to 0.01 for stability reasons. A final experiment with SISP uses the same scaled configuration but with a second forcing pattern for fast physics as described above. For comparison a extra experiment with no stochastic physics is included, as well as one experiment with the default version of SKEB2 (as defined in chapter 5). Experiments are verified against a combination of observations made by ground stations, radio-sondes and instruments onboard commercial airplanes.

7.1.1 SISP results

The SISP scheme has a very mild effect on the large-scale structures. The ensemble spread for the Northern Hemisphere (NH) of the Mean Sea Level Pressure (MSLP) shows little impact for the different SISP experiments in comparison to the run with no stochastic physics (Figure 7.1). The increments are very weak in comparison to those made by SKEB2. Amongst the different SISP setups, the scaling version is the superior one, although differences amongst them are not clearly significant. In terms of RMSE of the Ensemble Mean (EM), the scaled version is barely indistinguishable from the experiment with no stochastic physics (not shown). SISP results from mid- and high-level fields such as geopotential at 500hPa or winds at 250hPa also show a very small impact on the ensemble spread (not shown).

In terms of surface variables, the SISP scheme is able to perturb the structure of the Boundary Layer (BL) and therefore it has a positive impact on the ensemble spread (Figure 7.2), with the version that forces the fast physics with a different forcing pattern having the strongest impact. However, the SKEB2 spread does grow quicker and at the third day of the forecast it nearly overtakes the best performing version of the SISP experiments. The relative increase of spread from SISP is also reduced for larger forecast time. At T+12 the best performing SISP configuration is about 1/3 higher in spread than the experiment with no stochastic physics. At the third day, the increment is about 20% higher. Similar impacts are observed for surface winds (not shown). The oscillatory behaviour of Figure 7.2 is driven by the different number and quality of observations being assimilated by the ETKF at 00Z

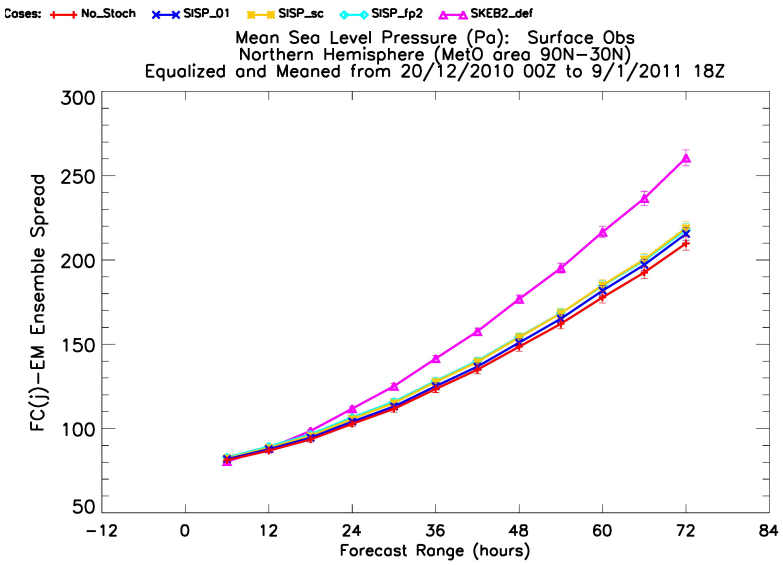

Figure 7.1: Ensemble spread of NH (90N-30N) Mean Sea Level Pressure (PMSL) for MOGREPS experiments. Red shows experiment with no stochastic physics, dark blue SISP with $< F_{\Psi}^2 >$ =0.01, yellow with scaled SISP (see text for details), light blue scaled SISP with a different forcing pattern for fast physics and purple shows experiment with SKEB2.

and 12Z (same applies for Figure 7.3).

The SISP increases the dispersion of surface variables, but its effects in terms of error are not beneficial as shown in Figure 7.3, where the ratio of RMSE EM between the SISP and SKEB2 experiments to the experiment with no stochastic physics is shown for the Southern Hemisphere (SH) 2 metre Temperature. The error in the SISP experiments increases in comparison to the version with no stochastic physics, unlike the SKEB2 in which it decreases and is negative (meaning an improvement in the forecast). A similar result was found by Tompkins and Berner (2008), where humidity perturbations to the initial state of the convective parametrization lead to a degradation of the EM error and probabilistic scores. They suggest that zero-mean perturbations to the input does not lead to zero-mean parametrization tendencies, and hence their scheme introduces biases in the convection scheme.

Overall, perturbing the initial condition with a simple scheme does not seem to produce notable improvements on the ensemble spread and error of an EPS, like other schemes such as SPPT or SKEB have done. Differences in the initial state, what the parametrization sees, does not seem to produce very different results from the deterministic physical parametrization, and even less to perturb large-scale

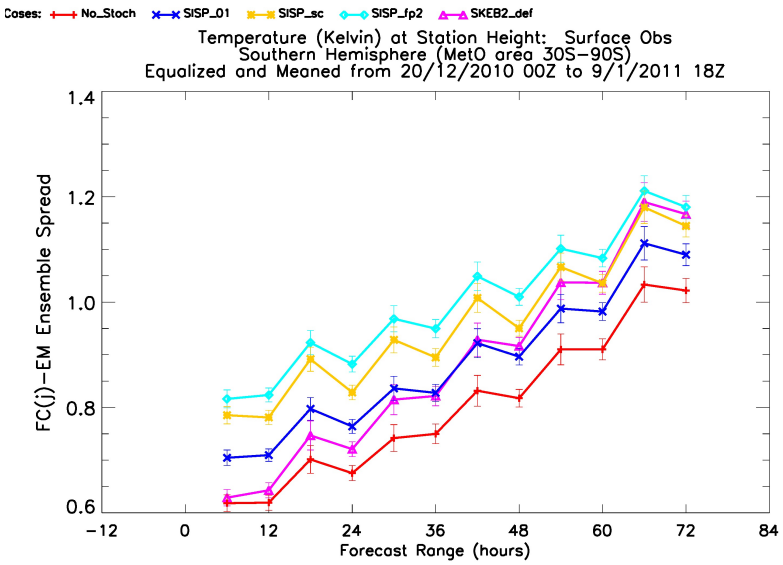

Figure 7.2: Ensemble spread of Southern Hemisphere (30S-90S) Temperature at 2 meters. Same line distribution as Figure 7.1

structures as seen in Figure 7.1 where the PMSL field is quite indifferent to the SISP perturbations. There is an exception in the Boundary layer, as temperature and wind fields at the surface seem to be substantially different across the ensemble with perturbed initial states. Nevertheless, such divergence does not grown with forecast time, indicating that it saturates quickly. In addition, these perturbations also increase the error of the EM, maybe because they lack any BL consistency in terms of balance amongst different fluxes. The new scheme, although original, has proven to have a limited capacity to improve the probabilistic forecasts.

7.2 Stochastic Perturbation of Tendencies (SPT)

One of the most successful approaches to represent uncertainty in physical parametrizations is to perturb the physical tendencies, as done in schemes such as SPPT in the IFS or the Canadian Model (see sect. 2.3 for more details on these schemes and their results). In the present thesis, we implement a similar scheme in MetUM and quantify its impacts across timescales. The scheme is compared to other stochastic schemes such as Random Parameters 2 (RP2, sect. 2.4) and Stochastic Kinetic Energy Backscatter v2 (SKEB2, described in 2.2 and evaluated in chapter 5).

The scheme created is named as Stochastic Perturbation of tendencies (SPT). The main characteristics that define it are:

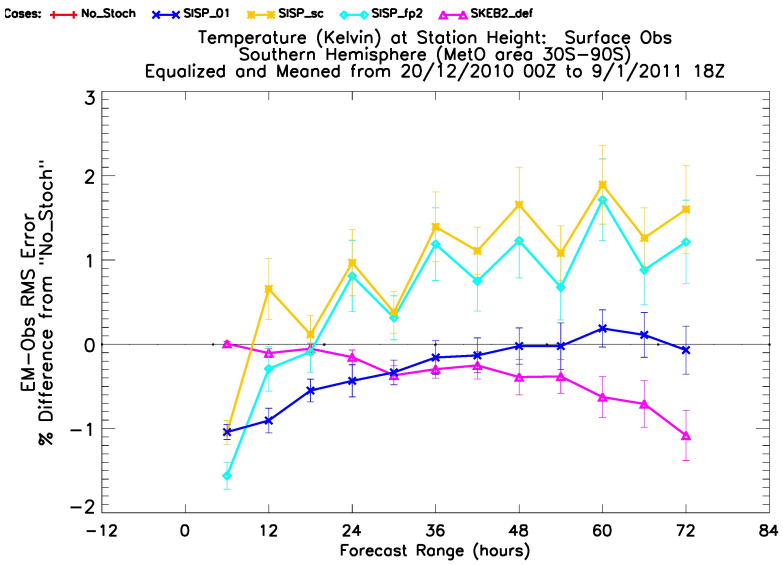

Figure 7.3: Ratio of the RMSE of the Ensemble Mean between the different experiments and the experiment with No Stochastic physics (red line in Figures 7.1 and 7.2). Variable is 2m. Temperature for the SH. Same line distribution as Figure 7.1.

- A spherical harmonic forcing pattern similar to the SISP scheme (described in sect. 7.1), whose vertical structure follows SKEB2 (as described in Tennant *et al.* 2011).

- It follows an univariate approach where all tendencies are forced with a standard deviation $\sigma = 0.5$ with the exception of the Gravity wave drag, whose perturbed tendencies use a $\sigma = 0.42$ for stability reasons. This is a similar setup as the fast pattern of SPPT (see Table 2.2).

- Tendencies from Boundary Layer are not included for stability reasons. The scheme employs a similar tapering as SPPT, with the SPT increments ramping up linearly from level 9 ($\sim 600\,m$) to level 15 ($\sim 1.6\,km$), and ramping down from level 41 ($\sim 11.5\,km$) to level 45 ($\sim 14.8\,km$).

- Slow physics (radiation, microphysics and gravity wave drag) and fast physics (convection) use different forcing patterns, the latter is a 180°longitudinal displacement of the former one.

- An option has been included to remove clear sky radiation from the perturbed tendencies. The MetUM large-scale cloud scheme, the Prognostic Cloud fraction and Prognostic Condensate (PC2, sect. 4.1.2) can output the tendencies

of T and q generated as a result of condensation due to radiative processes. This option only perturbs the radiation tendencies where clouds are present, leaving the well-resolved clear sky tendencies unperturbed.

- An option to conserve water and Moist Static Energy (MSE) has been developed for the scheme. It conserves the vertically integrated water vapour in the column. In addition, the temperature tendencies depends on q perturbations to conserve MSE. Appendix A describes these conservation constrains and their implementation.

The numerical problems associated to BL perturbations mainly occur on regions with high standard deviation of the orography, such as the Kashmir region or the Peruvian section of the Andes. Over these regions, the tendencies from the Gravity Wave drag (GWD) are quite large ($\sim 5\ m/s$). When the SPT acts against those tendencies (with a negative forcing pattern), the GWD tendencies become larger to force the model towards the state it would be if its tendencies weren't forced. This creates a feed-back process that produces larger GWD tendencies which eventually imbalance the boundary Layer, provoking a spurious warning of more than 10K, triggering an instability that provokes a grid-point storm. In a recent version of the scheme (not included in this thesis), the stability of the model with SPT has been substantially improved if the forcing to GWD tendencies is switched off in areas where the standard deviation of subgrid orography is higher than 500m and when the forcing pattern greater than 0.5 (ad-hoc values to ensure the stability of the model)

A set of different experiments is employed in different systems to investigate the impacts on the different flavours of the SPT. These experiments are described in table 7.1. The capacity of the SPT experiments to improve a probabilistic forecasts in comparison to other stochastic schemes is reported in section 7.2.1, this section also includes results from short-range NWP forecasts. Further assessment of the impacts of the scheme in the long term are evaluated employing 20 year climate simulations, reported in section 7.2.2. A description of these systems is described in section 4.1.1.

RP2	Includes RP2 scheme only
SKEB2	Includes SKEB2 with the improved configuration described in chapter 6.
SPT	Includes the version of the SPT scheme with no additional constrains.
SPT_no_csky	Includes a SPT version where Radiative tendencies from clear sky are not perturbed.
SPT_cons	Includes the MSE and water vapour column conservation.

Table 7.1: Description of the MOGREPS and climate experiments carried out for this section.

7.2.1 Impacts on Short-Range forecasts

The capacity of SPT to improve the ensemble is first evaluated on the Z_{500} field, a proxy for the representation of mid-latitude Large-Scale (LS) structures such as cyclones and blocking (described in chapter 3). Figure 7.4 shows the RMSE of the Ensemble Mean (EM) and spread for Z_{500} for NH and SH. The EM RMSE is very similar amongst the different experiments, with *SPT* and *SPT_no_csky* slightly worse than control. In terms of spread, in both hemispheres the *SPT_cons* produces similar values to *RP2*, *SPT_no_csky* produces larger spread and *SPT* has the strongest impact, even superior to *SKEB2*.

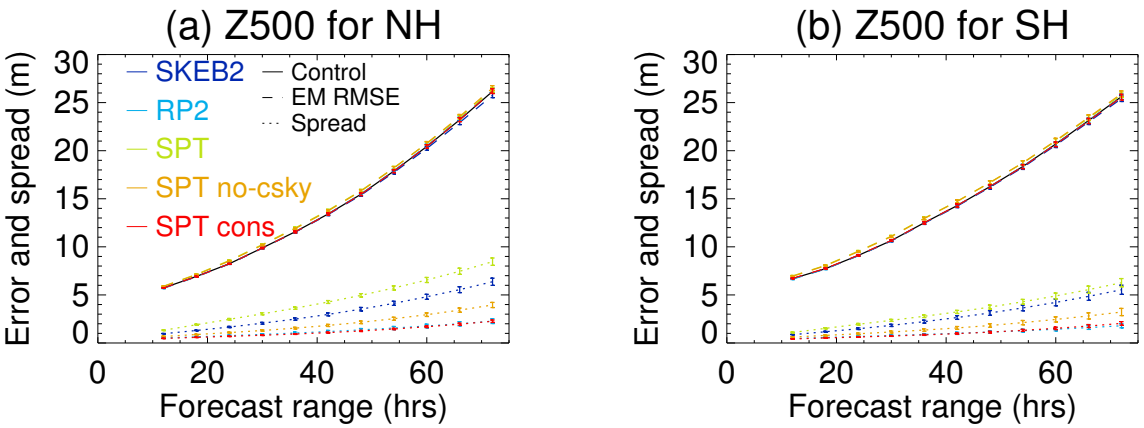

Figure 7.4: RMSE EM-spread of Z_{500} for: SKEB2 (dark blue line), RP2 (light blue), SPT (green), SPT with no clear sky forcing (orange) and SPT with conservation for water and MSE (red). Dark line shows control RMSE, dashed lines EM RMSE and dotted lines spread. (a) NH (90N-30N) (b) SH (30S-90S).

Unlike SKEB2, SPT is not aimed at representing specific processes like energy

dissipation by the dynamical core and convection. It coarsely represents the subgrid variability of important small scale processes such as the evolution and effects of clouds or the latent heat release of rain. Despite the simplicity of the SPT scheme, it can produce a better calibrated ensemble than SKEB2 for LS fields like Z_{500}, as SPT produces more spread and an equivalent level of EM RMSE than SKEB2.

The large-scale structures in mid-latitudes and tropics are driven by temperature and winds, in addition to humidity. The dispersion of temperature at different levels can provide some idea about the level of spread in the representation of highly parametrized processes such as clouds or moist convection. Figure 7.5 shows the EM error and spread for low (850hPa) and high (250hPa) level temperature. The greatest differences are in the tropical region, where *SPT_cons* clearly outperforms *RP2* and *SPT* is clearly superior to *SKEB2* generating spread (Fig. 7.5.b,e) and EM RMSE in the low level. For the mid-latitudes, *SPT_cons* produces more spread in the temperature than *RP2* with the exception of low level in NH (Fig. 7.5,a). Similar results for the SKEB-SPT comparison are found for the Integrated Forecasting System (IFS) and reported in section 2.3 of Palmer *et al.* (2009).

The spread and error of the ensemble for winds show very similar characteristics than it does for temperature (Figure 7.6). The only notable difference is that *SKEB2* does generate a similar level of spread to *SPT* at low level. The SKEB2 scheme forces low level winds directly, whereas spread in temperature and high level winds is generated by indirect effects. The SPT can have a higher direct impact on the convection and the properties of clouds, affecting the vertical distribution of energy. Thus these SPT changes have a greater effect in the ensemble spread of high level fields and low level temperature than the indirect effects of SKEB2.

The improvements on the EM RMSE of the tropical low level winds are better for *SPT* than for *SKEB2* with equivalent levels of spread (Fig. 7.6,b). This is also observed in low resolution deterministic forecasts. Figure 7.7 shows the ratio of winds RMSE between the different experiments and the control (as done in sect. 5.1). The error produced by SKEB2 is higher than all the other experiments at both levels and all regions, with the exception of high level winds in the tropics and in the SH for day 3 onwards. The impact on the deterministic forecasts of *RP2* and *SPT_cons* is not significant for all cases, nor it is for *SPT_no_csky* for many regions and levels. Similar results are found for N216 forecasts (not shown).

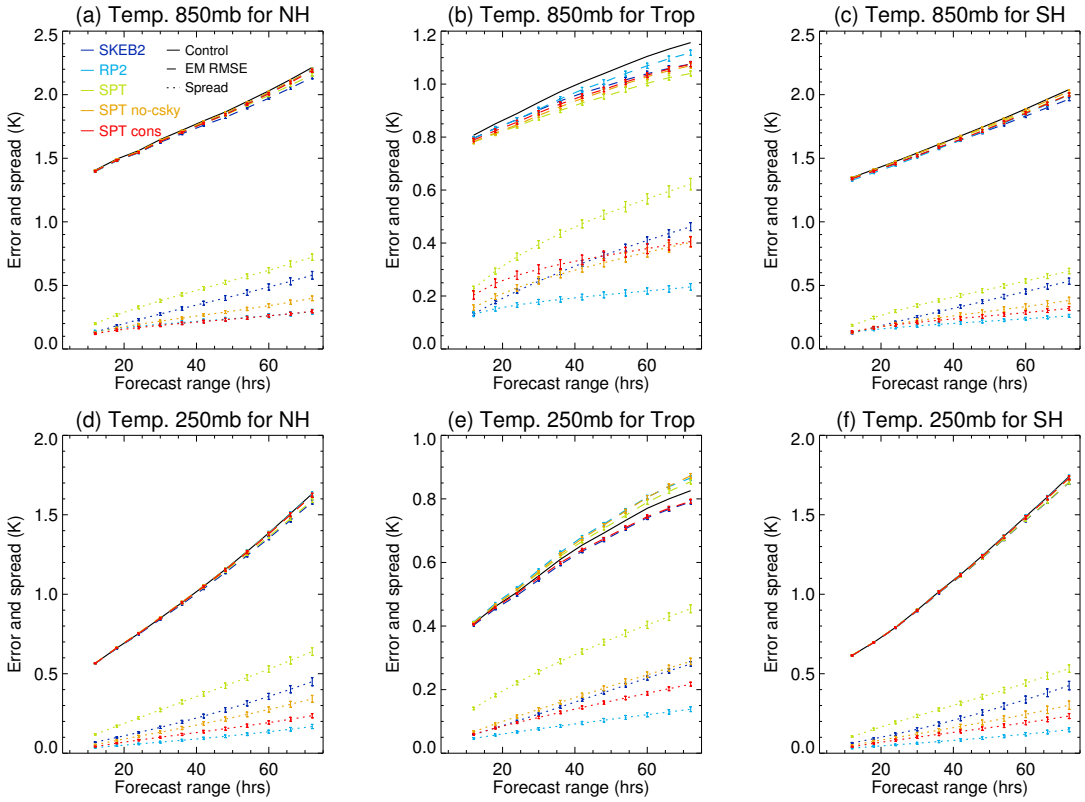

Figure 7.5: Error-spread plots for temperature, 850mb (a,b,c) and 250mb (d,e,f) winds. NH (a,d), Tropics (20N-20S) (b,e) and SH (c,f). Same line distribution as Figure 7.4.

A coarse representation of the subgrid variability is more beneficial in generating ensemble spread than stochastically sampling the internal uncertainty of key parameters within parametrizations (as done by RP2) for nearly all the cases. If additional constrains in SPT are added to conserve water vapour in the column and MSE, the impact of the scheme decreases, but it is still competitive with RP2 in terms of EM RMSE and spread in the mid-latitudes and clearly superior in the tropics, with a non-significant impact on the RMSE of deterministic forecasts. When the radiation from clear sky is not perturbed, there is less spread amongst ensemble members and it is inferior to the forecasts produced using SKEB2. This option also reduces the error of deterministic forecasts.

7.2.2 Impact on climate scales

Impact on global budgets

One of the main concerns about the impacts of SPT on climate scales is the impact it has on the delicate balance of energy and moisture budgets. The SPT randomly

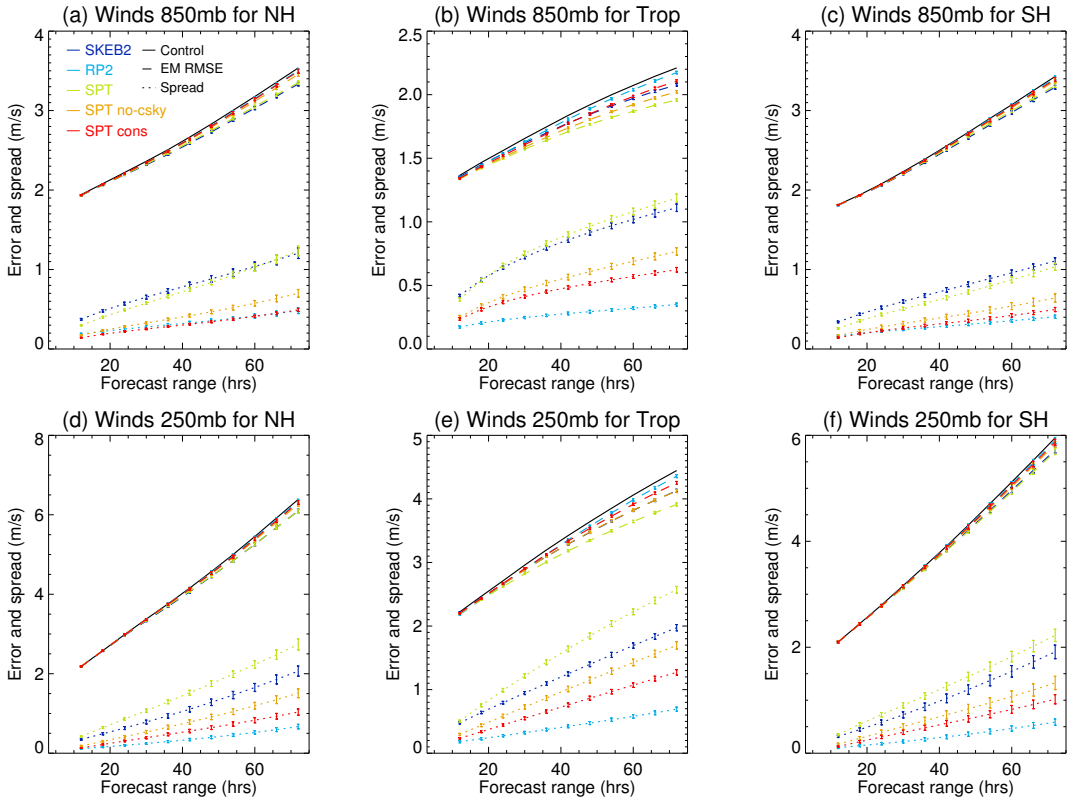

Figure 7.6: RMSE-spread plot of winds, same levels, distribution and line colours as Figure 7.5.

modifies moisture, kinetic energy and internal energy. Although many of the atmospheric parametrizations have little constrains to avoid unbalances in these budgets, SPT can potentially amplify large tendencies which could worse the problem of energy and water conservation. In order to check the unbalance in energy and water provoked by SPT perturbations, we make use of the diagnostics provided by the Energy Conservation scheme (EC, see Appendix A) and other global quantities such as total Precipitation minus Evaporation (defined as $E - P$), the total mean water vapour and the net energy flux at the Top of the Atmosphere (TOA).

The different versions of SPT have different effects on the conservation of energy, water and balance of $E - P$. Table 7.2 shows the global figures averaged over time, including the control run (with no stochastic physics). *SPT* increases substantially the energy correction term added to the model, as a positive (negative) EC term indicates energy supplied (extracted) by the EC scheme to the model. *SPT* also creates more water vapour and produces more precipitation than evaporation. The effect on $P - E$ is quite negative and would indicate a major problem if the scheme was introduced in an Earth System Model (ESM) with a full representation of the

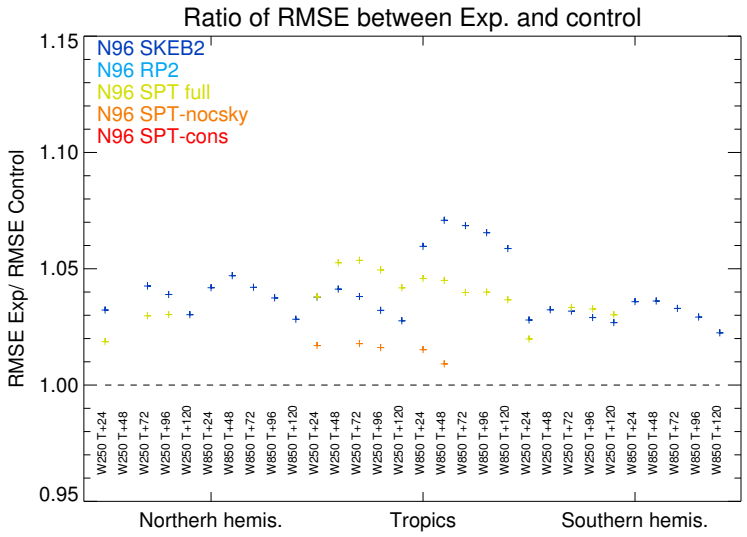

Figure 7.7: Averaged Ratio of RMSE from experiment and control at N96. Same colour scale as Figure 7.4 and distribution as Figure 5.2.

water cycle. The *SPT_no_csky* produces slightly closer numbers to equilibrium, although they are still too high. *SPT_cons* values are similar to control and slightly better than *RP2*.

	Control	*RP2*	*SPT*	*SPT_no_csky*	*SPT_cons*
EC (W/m^2)	-0.79	-0.75	5.74	5.44	-0.75
Total q_v (g/kg)	23.25	23.17	24.03	23.98	23.26
Total P-E (mm/day)	-0.0002	-0.0003	0.1728	0.1711	-0.0001
Net energy flux at TOA (W/m^2)	-0.0043	-0.0018	-0.0084	-0.0077	0.0001

Table 7.2: Global values (averaged over time) of conservation properties such as Energy Conservation increments (EC), total mean water vapour (q_v), or total Precipitation minus evaporation and net energy flux at TOA.

The last row of table 7.2 shows the net energy flux at TOA. Ideally this should be zero in order to maintain the Earth in a thermodynamic equilibrium. Negative values imply that the Earth releases more energy than it receives. *SPT* doubles the imbalance of energy at TOA, radiating away more energy and thus cooling the atmosphere, which triggers the addition of energy through the *EC* term. When clear sky temperature tendencies are not perturbed in *SPT_no_csky*, the net energy flux is still too high.

The spurious creation of water vapour may produce different feedbacks that lead to a cooling of the atmosphere. Table 7.3 shows global quantities averaged over time to provide a narrative for the mechanism for the cooling. One plausible hypothesis is that the excessive water changes the reflective properties of clouds, enhancing reflectivity of clouds in the upper levels of the troposphere. However, the downward short-wave flux at the surface SW^{surf}_{down} is only $1\,W/m^2$ smaller than the control for the *SPT*. The imbalance is much higher for *RP2* and this experiment does not produce such a large cooling. Another possible mechanism is a global increment of precipitation given the extra water vapour available for condensation and precipitation, this would thicken the low level clouds, blocking downward long-wave flux at the surface LW^{surf}_{down}. Also the excessive saturation would reduce the Latent Heat from the surface. Values in table 7.3 agree with this hypothesis.

	Control	*RP2*	*SPT*	*SPT_no_csky*	*SPT_cons*
Precipitation (mm/day)	3.03	3.02	3.10	3.10	3.01
Latent Heat (W/m^2)	87.65	87.50	84.88	84.82	87.25
LW^{surf}_{down} (W/m^2	336.78	338.71	338.81	338.73	336.84

Table 7.3: Global values averaged over time for precipitation, Latent Heat (LH, see sect. B.3.4), short-wave (SW) downwards flux over surface (SW^{surf}_{down}) and long-wave downwards flux over surface (LW^{surf}_{down}).

The EC scheme adds an uniform temperature increment to correct the energy budget (see eq. A.7 in Appendix A). MetUM has a cold bias in the tropospheric zonal temperature for JJA and DJF (Fig. 7.8,a,d). This bias is partially removed by the EC action when SPT is included. This impact is clearest for JJA (Fig. 7.8,e). However, the SPT causes an important warming of the Tropical Tropopause Layer (TTL) when combined with the associated warming caused by the EC (Fig. 7.8,b,e). The warming caused by both factors is greater in the stratosphere and specifically in the TTL region (Fig. 7.8,c,f). The SPT is not active above the upper troposphere so the warming could be a feedback from the vertical moist transport from below, and the impacts of the radiation from the uniform EC warming. Similar biases are found for the specific and relative humidity over the TTL (not shown). The excessive amount of warming needed by the *SPT* experiment to correct the energy balance is another setback for the inclusion of SPT in an ESM. The coupling

between the chemistry and atmosphere is very sensitive to the TTL. Biases over the TTL could affect the distribution of water vapour in the stratosphere and trigger important radiative feedbacks (Solomon *et al.*, 2010; Dessler *et al.*, 2013).

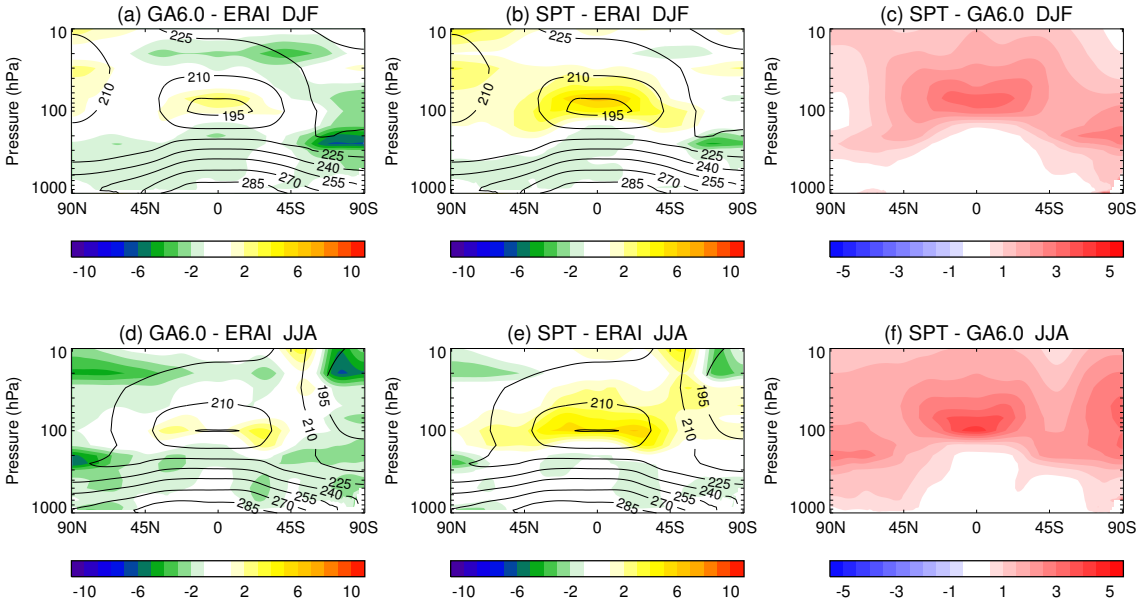

Figure 7.8: Zonal Temperature biases (K) in colours and ERAI zonal temperature in contours for DJF (a,b,c) and JJA (d,e,f). (a,d) Control - ERAI (b,e) *SPT* - ERAI (c,f) *SPT* minus control.

Although the spurious increment of temperature could blur the general impacts of the SPT scheme, it is worth looking at how SPT changes climate variability processes in comparison to the other stochastic schemes.

Impact on tropical climate processes

The global increment in tropical precipitation for summer (JJAS) is shown in Figure 7.9. As discussed in chapters 5 and 6, in GA6 control there are tropical areas like the equatorial Atlantic Ocean or Indian Ocean with excessive divergence, which leads to negative biases in Outgoing Longwave Radiation (OLR) and too much precipitation. The SKEB2 reduces these biases, although the impact of the default version is minimal (Fig. 7.9,e). The impact of *SPT_cons* is even smaller (Fig. 7.9,d). *RP2* reduces the biases, but also increases precipitation over the West Pacific, in a bigger proportion than SKEB2 (Fig. 7.9,b). The impact of *SPT* is the most beneficial. It is also noteworthy that the precipitation is reduced over West Africa on the *SPT* experiment (Fig. 7.9,c), another prominent bias where other schemes

have little effect.

The improvements in tropical precipitation over Africa may come locally from a better coupling to the surface fluxes which triggers local convection. However there is also a large-scale effect from the SH Jet structure over Africa and the Indian Ocean. Figure 7.10,a shows the ERAI wind field at 250hPa on the Tropics and SH for summer. The westerly jet structure is clear over the subtropical SH plus the easterly winds in the tropical Indian Ocean. MetUM at GA6 has weak easterlies over the tropics, its SH jet extends equatorwards over Africa and slows down in the centre of the jet over the same longitudes (Fig. 7.10,b). These biases may be linked to a poor representation of some aspects of convection, like the diurnal cycle (Bechtold *et al.*, 2004). *RP2* worsens the high level wind biases over Africa and this significantly worsens the dry bias there (Fig. 7.10), whereas *SPT* reduces these biases notably. There are local processes such as African Easterly waves (AEW, Kiladis *et al.* 2006) whose representation improves by *SPT* (not shown). These waves are important drivers of Alantic Tropical Cyclones (Thorncroft and Hodges, 2001).

Overall there are similarities between the SPT and SKEB2 in the way they couple dynamics and physics. They both improve the representation of the tropical dynamics, decreasing divergence and improving the SH Jet stream over summer (Fig. 5.29 and Fig. 7.10). This leads to improvements in the representation of clouds and precipitation (Fig. 5.28 and 7.9), and intermittency of convection (Fig. 5.24). There are also some common setbacks. Both schemes reduce the power of high frequency Kelvin waves (not shown).

Impact on mid-latitude variability

A better distribution of the jet also improves the SH mid-latitude variability. Figure 7.11,a shows the ERAI storm track density for austral winter (JJA), where the storms spiral from the South Atlantic towards the East Pacific side of the Antarctica (Hoskins and Hodges, 2005). There are fewer storms on the Atlantic side of the Southern Ocean (SO) and at the end of the storm tracks in the control (Fig. 7.11,b). *SPT* decreases the bias as it has got more storms in the Atlantic Ocean and at the end of the storm track in the west coast of South America. However, it also decreases the density of storms around the Antarctic continent (Fig. 7.11,c,d). The

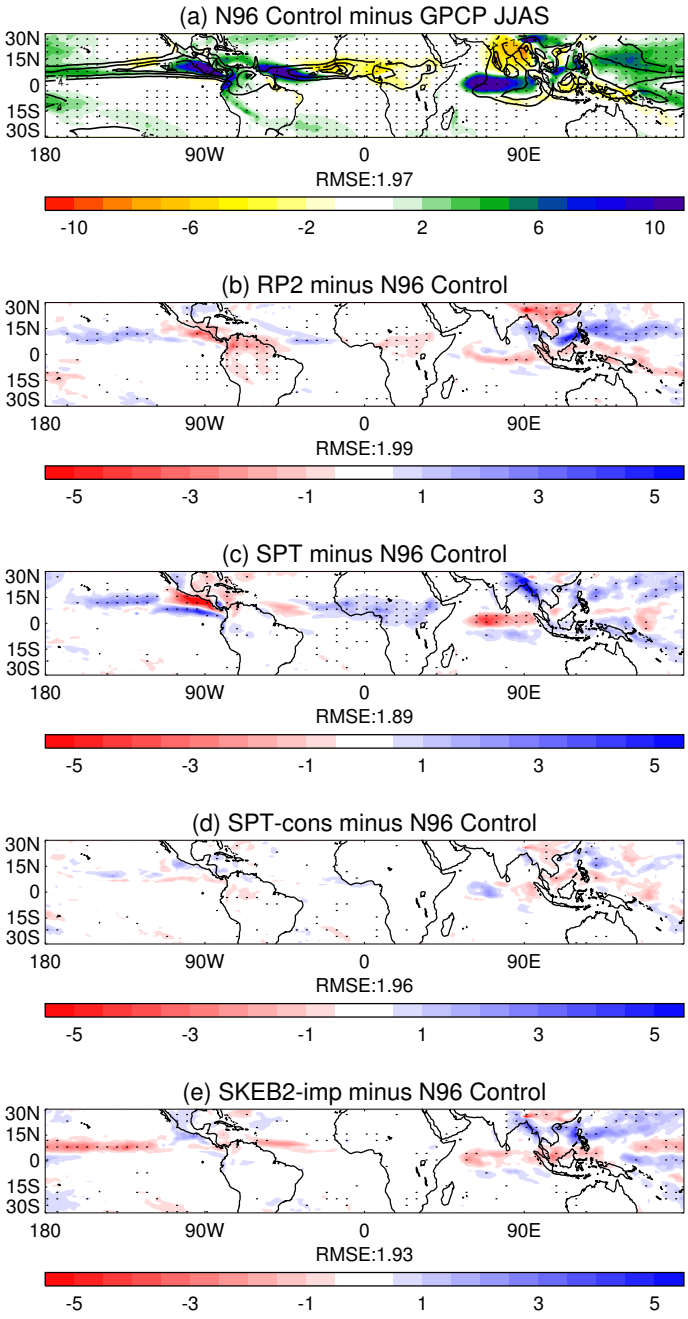

Figure 7.9: Total precipitation biases (mm/day) for JJAS (a) Control minus GPCP, contours show GPCP (b) RP2 minus Control (c) SPT minus Control (d) SPT-cons minus Control (e) SKEB2-imp minus control. Dots denote statistically significant differences above 95% level of confidence using a student t-test.

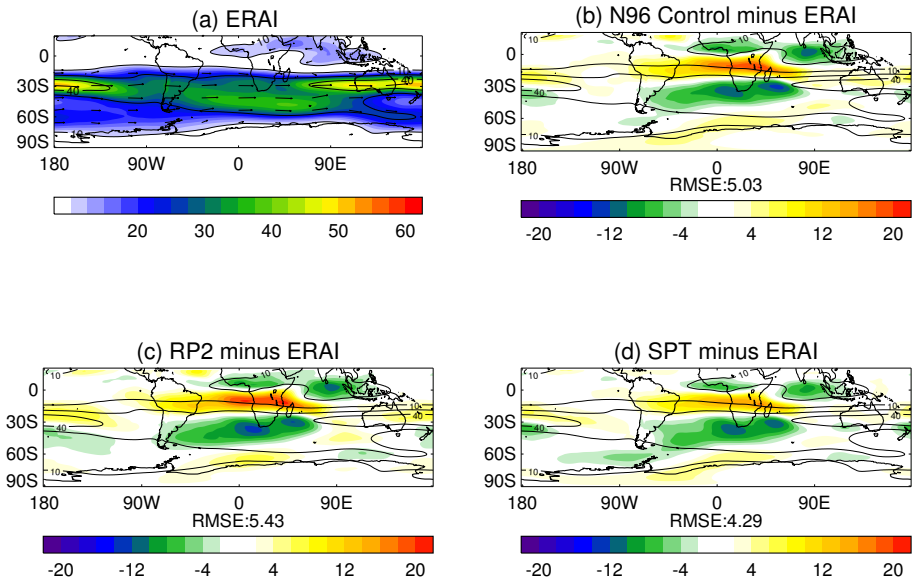

Figure 7.10: Winds at 250hPa for JJA, (a) ERAI field, arrows show direction of the wind, (b) Control minus ERAI, (c) RP2 minus ERAI and (c) SPT minus ERAI.

simulation of the summer SH storm track is less realistic for *SKEB2*, as there are more subtropical storms coming from South America, fewer over the Pacific side of the SO and more South of Australia. The impact of the other experiments is negligible.

The magnitude of the impact of the SPT looks far superior to the one produced by RP2. The former produces more ensemble spread and stronger impacts on the climate system, some of them undesired as the high imbalances in conservation quantities. However, this study has limitations as a comparison between the stochastic perturbed tendency approach and stochastic perturbed parameters. The forcing pattern is different, RP2 one has no spatial or vertical structure. In addition, they don't exactly perturb the same processes as SPT does not perturb BL and RP2 does not include all parameters that may play a role in each parametrized tendency. A comparison of the different approaches for stochastic physics within the same framework could provide a clearer comparison. The next section provides such comparison.

7.3 Stochastic Mixed Phase Parametrization

In the last section of this chapter, we explore the impacts of different methodologies to transform a deterministic bulk-formula parametrization into a stochastic

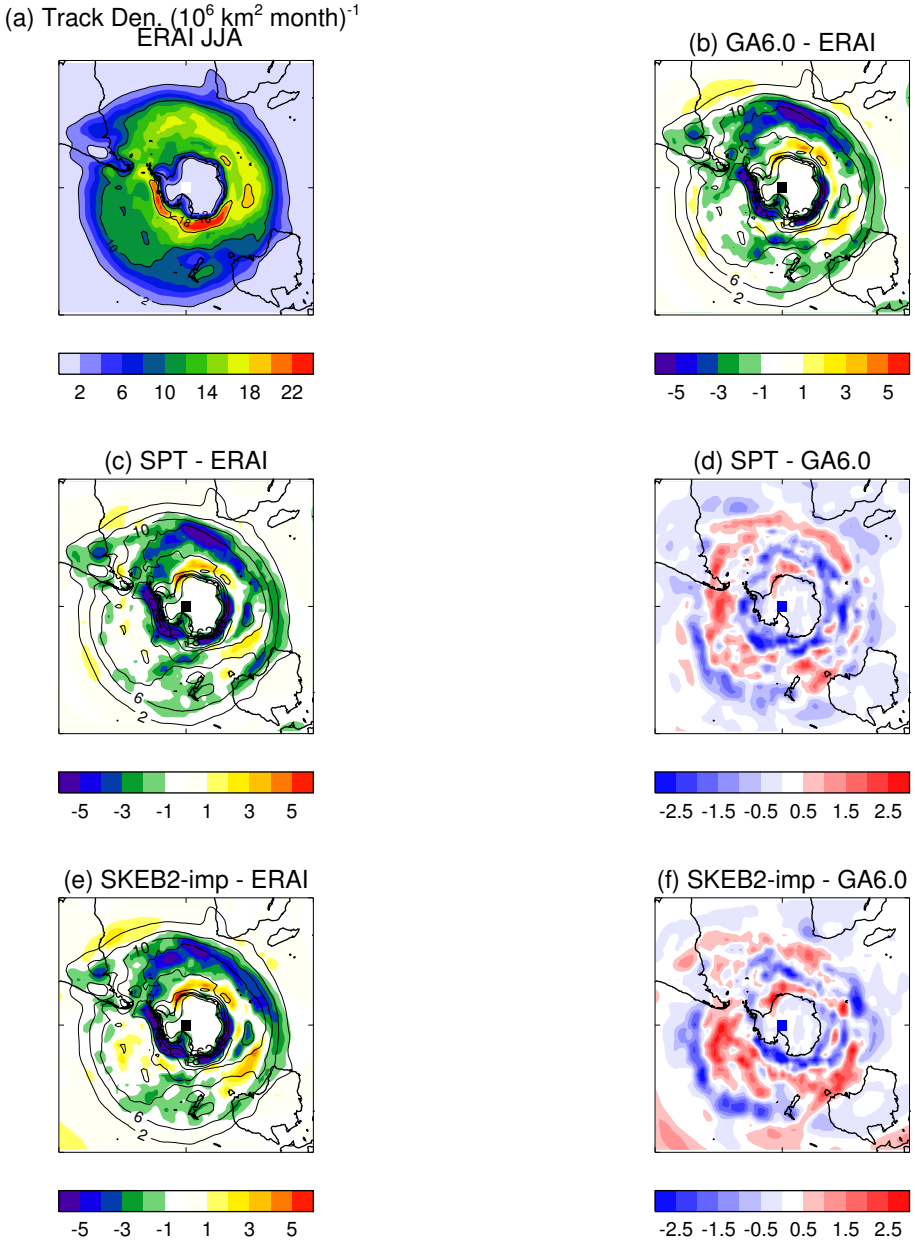

Figure 7.11: Storm track density for JJA over SH. (a) ERAI tracks, (b) GA6, (c,d) SPT (e,f) SKEB2. (b,c,e) Colours shows biases to ERAI and contours are ERAI tracks. (d,f) Biases to GA6.

parametrization. The parametrization we choose is the Mixed Phase Cloud (MPC) parametrization of Field *et al.* (2014), which is based on highly stochastic subgrid turbulent motions within icy clouds.

A MPC is a type of cloud that contains ice and supercooled liquid water in an icy environment within the cloud. When an air parcel rises through the freezing level droplets do not instantly freeze, in fact supercooled water droplets may continue to exist even at $-40C$. MPC can form in the updraft of convection, large-scale ascent or turbulence in icy clouds which locally give enough ascent to form liquid water. These clouds emerge more frequently in the polar regions, where they cover large areas throughout the year (Morrison *et al.*, 2012). MPC can also exist in mid-latitudes and tropics (Riihimaki *et al.*, 2012). Their impact on the radiative fluxes and energy balance is significant in many regions (Shupe and Intrieri, 2004).

The MPC parametrization we employ is based on Large-Eddy Simulations (LES) detailed in Hill *et al.* (2013). The parametrization computes a distribution of supersaturation S from the turbulent environment and the ice-cloud properties. The parametrization's tendencies, condensed water amount q and cloud fraction C, are the integral of the S distribution over a given threshold (e.g. supersaturation with respect to ice for water saturation). See section 2 of Field *et al.* (2014) for details on the scheme.

The S distribution is the solution of a stochastic differential equation, where the turbulent motions are represented as random up- and down-draughts (eq. 5 in Field *et al.* 2014). The stochastic equation has some parameters whose values are uncertain. The parametrization experts have suggested to perturb three of these parameters within a given range to sample the uncertainty of the scheme. These parameters represent the variance of the turbulent vertical velocity fluctuations, the vertical extent of the turbulent zone and the first moment of ice distribution.

Four different approaches to represent the uncertainty of the MPC scheme are investigated in order to estimate which one produces the best calibrated ensemble and the better improvements in climate fields (by noise-induced drifts). The first three methods explore the perturbed parameter approach, perturbing the 3 parameters above mentioned in the following way:

- Parameters at each grid-point in each level are perturbed using white noise. This experiment is defined as WN

- Parameters are perturbed using the RP2 approach, where they temporally vary following a first order auto-regressive process. Thus the forcing is homogeneous in the horizontal and vertical. This experiment is defined as $1AR$

- Parameters are perturbed by a Spherical Harmonic Forcing Pattern (SH-FP) similar to SPT (see 7.2 and 2.2.4 for a more general description of SH-FP). The experiment is defined as $SH - FP$

The spatio-temporal PDF for the forcing pattern of these 3 methodologies is shown in Figure 7.12, where one day runs with one hourly output have been employed. The $SH - FP$ shows the expected Gaussian behaviour and it is perhaps too narrow to fully sample the range of likely values of $[0, 1]$. The WN has a constant probability for all possible ranges, and the $1AR$ produces limited discrete values, given its uniformity in space.

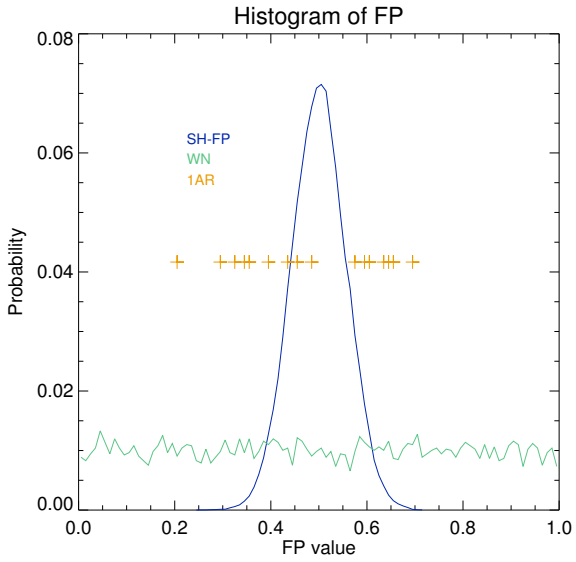

Figure 7.12: Distribution of the 3 different Forcing Patterns (FP) to perturb the chosen parameters within the MPC scheme. Green line shows the white noise method, yellow crosses 1AR values and blue line SH-FP method. See text for details.

A fourth extra scheme is developed following the SPPT approach. C and q tendencies from the MPC parametrization are perturbed using the SPT forcing pattern, with a standard deviation for the forcing pattern equal to 0.5 and without vertical tapering and moisture or energy conservation constrains. This experiment is defined as $lSPT$.

MOGREPS experiments are carried out with the four different approaches to represent the MPC stochastically. Their setup follows the description given in the Methodology (chapter 4). The uncertainties of the MPC are tiny in comparison to the total model error, therefore the impact on spread will be negligible for large-scale fields. However, the impact on temperature should be more clear as the supercooled water modifies the radiative fluxes in the atmosphere.

The EM RMSE and spread for temperature of the stochastic MPC schemes is shown in Figure 7.13 for different regions and low (850mb) and high (250mb) levels. The $lSPT$ is clearly the version that produces the largest spread in all regions and levels. In terms of EM RMSE it is also the best performing with a clear improvement for the tropics. In the low level SH, $lSPT$ experiment shows better EM RMSE than control whereas the EM RMSE of the other experiments is worse than control.

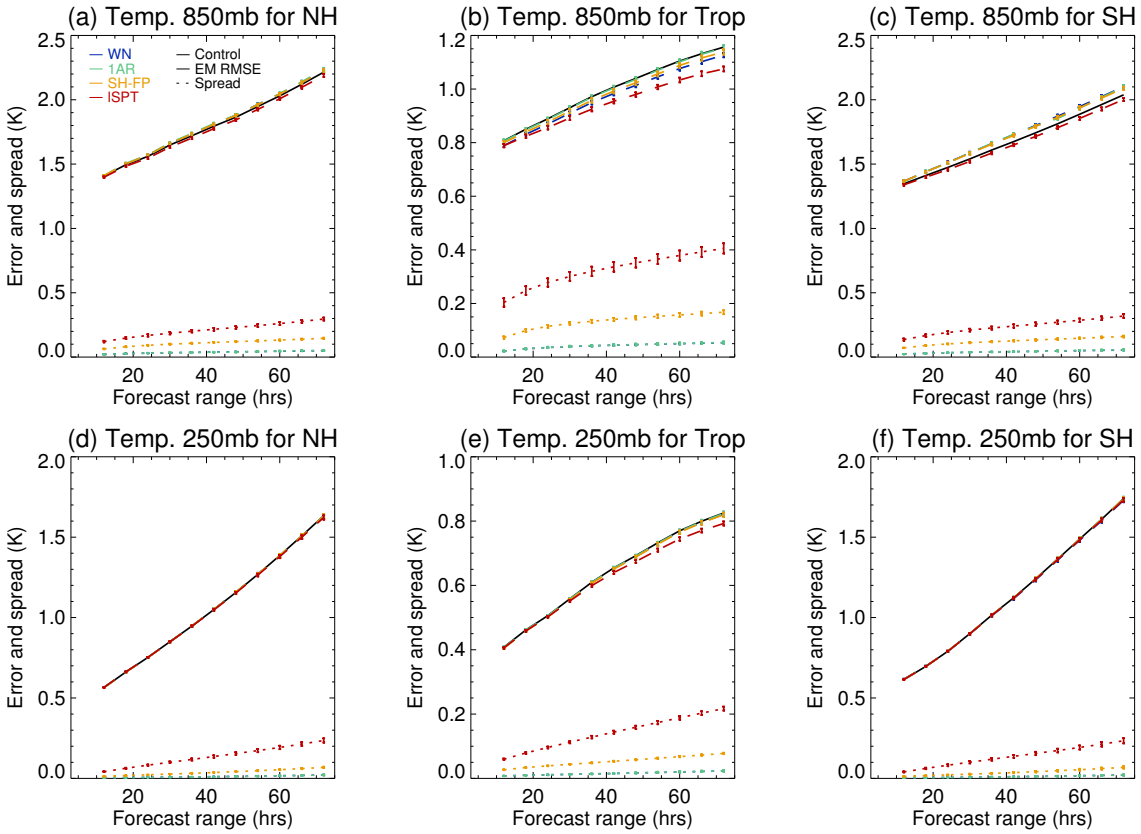

Figure 7.13: Ensemble Mean RMSE-spread for Temperature, same layout and format as Figure 7.4 but with the colours described in Fig. 7.12 and red as the SPT-like approach.

The three different methodologies to perturb the uncertain parameters produce a smaller impact in the dispersion of temperature within the ensemble (Fig. 7.13). Nevertheless, there are interesting differences amongst them. The $SH-FP$ exper-

iment produces higher spread than the other two for all regions and levels. Giving spatial and vertical structure to a forcing pattern is clearly beneficial for increasing the dispersion in an EPS. The spread generated by $1AR$, a similar approach than RP2, is quite poor in comparison to the one generated by a fully 3-D SH-FP in exp. $SH-FP$. $1AR$ is almost indistinguishable from WN for all levels and regions and the RMSE EM is worse in low level tropical winds.

In order to look for climate noise induced-drifts that may improve the representation of climate processes, five AMIP climate simulations are performed. The four stochastic flavours of the MPC plus a control with the original version of the scheme. Differences in the mean climate amongst these experiments are marginal for radiative fields such as Outgoing Shortwave Radiation (OSW) or OLR (not shown). The capacity of the different stochastic schemes to produce different tendencies for the cloud evolution does not change the radiative representation of clouds in climate scales.

The effects of the stochastic perturbations for the MPC scheme could be noticeable on the level-by-level radiative distribution of clouds. There is an ideal process-based diagnostic to assess this aspect of the climate, the satellite-simulator. It is an algorithm that transforms the modelled radiative field into a radiative beam at different wavelengths that a satellite would observe. Such simulators are described in Bodas-Salcedo *et al.* (2011) and Pincus *et al.* (2012) and will be employed in future work to estimate the effects of stochasticity on MPC clouds.

Different approaches to stochastically sample the uncertainty of a MPC schemes can generate different levels of spread and EM RMSE in a short-range EPS, with the tendency perturbation approach having the largest impact. However, these approaches can not produce significant differences on radiative fields of climate simulations. The impact of a single scheme is very limited and detailed process-based diagnostics must be employed.

7.4 Conclusions

Different methods to introduce stochasticity in physical parametrizations have been explored in this chapter. These cover from novel approaches like perturbing the initial state, to more consolidated ones like perturbing physical tendencies. A final

section explores different ways to stochastically represent the internal uncertainty of a Mixed-Phase Cloud (MPC) parametrization.

A scheme that perturbs the initial state of physical parametrizations, The Stochastic Initial State for Parametrizations (SISP), is developed in the course of this thesis. This scheme respects the internal structure and constrains of parametrizations while allowing fluctuations in subgrid processes (e.g. change the conditions for the emergence of deep convection, change water content or temperature within a cloud). The SISP produces a marginal dispersion in the ensemble within the free atmosphere. In the boundary layer, the spread is higher but it grows at the same rate as an experiment with no stochastic physics and thus it is quickly overtaken by SKEB2. In addition SISP degrades the RMSE of the EM, probably caused by implicit changes on the mean BL and convective tendencies. An additional problem for SISP lies in the difficulty in estimating individually the uncertainty in the initial state. In the present form of SISP, a Spherical Harmonic Forcing Pattern (SH-FP) perturbs the prognostic variables θ, q, u and v using an arbitrary amplitude. Coarse-graining studies of the prognostic variables at the beginning of the timestep are required to obtain adequate ranges for the amplitude and scales of the SH-FP.

Stochastically perturbing physical tendencies is a popular approach, with successful schemes such as the Stochastic Perturbation of Parametrized Tendencies (SPPT, see 2.3 section for details). However, their current formulation poses a few issues, such as the perturbation of well resolved clear-sky radiative tendencies or imbalances on the moisture and energy budgets. In this thesis, the Stochastic Perturbation of Tendencies (SPT) is developed following the SPPT approach. It includes an option to remove radiation from clear sky, and another to conserve water vapour in the column and Moist Static Energy (MSE). A comparison of the SPT with these options and other stochastic schemes such as SKEB2 or RP2 shows:

- The SPT is quite at effective generating ensemble spread. It is superior to SKEB2 in many fields such as high level temperature. However, it also produces a slight deterioration of the EM RSME for Z_{500} in both hemispheres.

- Even though than SPT produces higher or equivalent levels of spread for winds to SKEB2, it produces smaller RMSE than SKEB2 for individual forecasts.

- When the SPT is included in a climate model, the energy correction needed

to balance the energy budget is too high, of the order of $5\,W/m^2$. The scheme also produces spurious water vapour and more precipitation than evaporation. These imbalances translate into a cooling of the atmosphere, compensated by an homogeneous warming by the Energy Conservation scheme. The temperature increment makes the temperature in the Tropical Tropopause Layer (TTL) unacceptable for ESM studies.

- Despite conservation issues, there is a huge improvement in the climate produced by an AMIP run with SPT. It notably reduces the biases on the representation of the SH Jet stream in JJA, leading to a better precipitation field in the tropics and storm track over the Southern Ocean. These improvements outperform those made by SKEB2.

- A version in which the clear-sky tendencies are not perturbed produces less spread, but better RMSE for deterministic forecasts, even non-significant from control in the mid-latitudes. The beneficial impacts on the tropical climate are slightly reduced. This version also has an undesirable impact on the conservation of moisture and energy.

- A version that conserves MSE and water vapour in the column produces acceptable drifts in MSE and moisture budgets, with minor improvements in the climate. For EPS, this version is equivalent to RP2 for Z_{500} and superior for winds and temperature, with the exception of NH in low levels.

- All SPT versions and SKEB2 have positive or neutral impact on the tropical climate, whereas RP2 shows a slight degradation of tropical precipitation and representation of SH jet stream.

Using a MPC scheme as a template, we explore different approaches to represent stochastically the internal uncertainty of the scheme. Perturbing the tendencies produces more dispersion in temperature than perturbing key internal parameters, and produces a clearly improvement for the EM RMSE in the tropics and low level SH. If key parameters are perturbed, the best method to produce ensemble spread is to employ a SH-FP. Using a similar approach as RP2 is equivalent to employ white noise to perturb the parameters. The climate means of radiative fields are indistinguishable amongst the different approaches and the control run. Future research using cloud-simulators and other cloud-based diagnostics is needed

to observe improvements in the representation of atmospheric processes made by the stochastic MPC schemes.

Overall we observe that perturbing parametrizations internally (e.g. perturbing the initial state or key parameters in the MPC) is not as effective as perturbing tendencies to produce a better calibrated ensemble. However, for long timescales, the tendency perturbation approach needs a closure to avoid an spurious generation or drain of moisture and energy.

Perturbing parameters with a spatially and vertically homogeneous forcing pattern, as done in RP2, is not a successful approach in comparison to others. It generates little spread and the impacts on climate scales are not as positive as the SPT with conservation constrains or SKEB2. When it is included in an MPC, it is clearly inferior to a perturbed tendency approach. As suggested by the experiments with the MPC scheme, it could improve its performance if it would use a SH-FP, although the spatial characteristics of the pattern will need to be investigated using coarse-graining techinques.

The number of results presented in the present and previous chapters is enough to try to provide some conclusive answers about the role and future of stochastic physics schemes. In chapter 8, the last one of the thesis, the results are summarised and recommendations are made for the development and evaluation of stochastic physics.

Chapter 8

Conclusions

Weather forecasts and climate projections are often imprecise as they come from an imperfect tool, a General Circulation Model of the atmosphere (GCM). The imperfections arise from highly uncertain assumptions included in the model, such as the bulk representation of important processes occurring on scales smaller than the truncation scale. Stochastic physics is one method to represent some of these underlying uncertainties in a probabilistic forecast, where the importance lies on the reliability of the different forecasts of the ensemble rather than on the accuracy of a single member of the ensemble. Therefore stochastic schemes have been developed mainly to produce divergent outcomes in the ensemble, the so-called ensemble spread. These schemes can divert the flow away from the model's preferred state, improving the climate and atmospheric processes in certain cases. However, it is not perfectly clear if they are a simple tool to increase the ensemble spread or if they are a physical representation of the model uncertainties, and thus are able to improve the representation of intrinsic atmospheric processes.

One of the most popular stochastic schemes is the Stochastic Kinetic Energy Backscatter (SKEB). It puts back into the model the Kinetic Energy drained by diffusion and missing from in convective processes. The research undertaken with SKEB2, the MetUM version of a SKEB scheme, over different prediction systems, atmospheric processes and model resolutions is reported in chapter 5 of the thesis. SKEB2 degrades deterministic forecasts, diverting away the trajectory, intensity and speed of mid-latitude cyclones. Nevertheless, the averaged effects of the scheme are positive. It increases the intensity, speed and growth rate of cyclones, and strengthens the weak winds over the mid-latitudes at low resolutions. The scheme

helps to maintain convective episodes over time, which causes a positive effect on the divergent flow. This improves the mean representation of radiation, clouds and precipitation over the Tropical belt (with the exception of the West Pacific where the scheme makes the model moister). On the other hand, the scheme produces the emergence of a spurious westward tropical wave. These results are enhanced if we increase the amplitude of the backscatter, controlled by the backscatter ratio b_R.

When the model is upgraded by improving processes such as the internal diffusivity or subgrid variability of convection (processes that SKEB2 represents), there are several improvements in atmospheric phenomena that are on the same direction as those made by SKEB2. For example, an increase of mid-latitude cyclone intensity. However, the impacts of the scheme remains constant where they should decrease as its "raison d'être" becomes weaker. This is observed for mid-latitude cyclones, SKEB2 is unable to spin-up more those storms that are more diffused, generating a quasi-constant forcing for the intensity of storms across resolutions and intensity ranges.

Despite the usefulness of SKEB2 in producing spread at NWP and seasonal timescales, the scheme is unable to produce spread at climate scales. An ensemble of climate experiments with SKEB2 produces minor differences in the mean climate amongst the members, and these differences are not statistically significant from other ensembles with no stochastic physics.

The SKEB2 scheme needs to be improved to eliminate the problems found in chapter 5 (mainly the westward spurious tropical wave and lack of sensitivity to model error). In chapter 6, some ideas are developed to offset these deficiencies. These are: (i) Removing the large-scales from the forcing pattern, (ii) using an alternative method to estimate the numerical dissipation by the dynamical core and (iii) the combination of this alternative method with a factor to modulate the convective dissipation rate.

(i) The removal of the large-scales in the forcing pattern of the scheme is done by increasing the lowermost truncation wavenumber of the spherical harmonic expansion from 5 to 20. This generates weaker perturbations that reduce the RMSE of deterministic forecasts, the ensemble spread and the improvements of tropical climate biases, although there is still a beneficial impact in comparison to the control with no stochastic physics. This change reduces the power of the spurious westward

equatorial wave and its zonal extent. The ensemble spread generated by SKEB2 does have some undesired properties that this change does not help to fix. The power spectra of the ensemble spread grows faster at large-scales than the error and the spread of storms shows little sensitivity to their error.

(ii) A different method is employed to compute the estimation of energy loss in the interpolation to the departure point (described in sect. B.2.1), the contribution of different terms is explored and compared to the original version. The new method is more sensitive to horizontal resolution and produces larger impacts over the low levels rather than mid to high levels. The higher activity at low resolution improves the Z_{500} climate and tropical convection-coupled fields like precipitation or Outgoing Longwave Radiation (OLR).

(iii) The convective dissipation mask also has a weak resolution dependency. In order to alleviate this problem, a resolution dependent amplitude factor F_C is added to the convective mask. The combination of the new convective and numerical dissipation masks is compared to the default scheme. The new version makes the scheme more active at low resolutions, degrading deterministic forecasts but improving the intensity of mid-latitude cyclones, as well as important tropical fields in a climate simulation. However, it also degrades the representation of CCEW by weakening the power of Kelvin waves.

At high resolutions, the impact of SKEB2 is minimal as the numerical diffusivity is negligible in comparison to the error coming from physical processes. In chapter 7, different approaches to represent uncertainty from physical parametrizations are explored.

A scheme that perturbs the initial state of parametrizations is developed. It respects their internal structure and theoretically would help to produce spread in the diagnostics of physical processes (e.g. triggering deep convection). Such scheme produces marginal spread in the free atmosphere and increases the ensemble mean RMSE in the Boundary Layer. This result indicates that current parametrizations are insensitive to changes in the initial state, as they tend to provide similar tendencies for physical processes.

A perturbed tendency approach, named as Stochastic Perturbation of Tendencies (SPT), is developed and compared to other schemes. It is found that SPT is

the best performing in terms of spread generation, and its error in the deterministic forecasts is lower than SKEB2. On climate scales, the improvements are very positive for the SH and Tropical fields. However, the scheme has a serious impact on the conservation of energy and moisture. This makes the scheme unsuitable for Earth System Models, where these budgets are fundamental.

An option to leave clear sky radiation tendencies unperturbed is developed for the SPT. It reduces the spread and produces better deterministic forecasts. Nevertheless, the impact in the climate budget is still unacceptable. Another option enables the scheme to conserve water vapour in the column and Moist Static Energy (MSE). This option minimizes the problems with conservation to acceptable levels and reduces the impact of the scheme, although it is still competitive to a random parameter approach in its present formulation.

Different approaches for the stochastic representation of physical uncertainties are explored under the same framework in the final section of chapter 7. Using a perturbed-tendency approach generates more spread and less ensemble error than perturbing parameters. For the latter, having a spatial and vertically correlated forcing pattern is clearly beneficial in comparison to adding white noise, or perturbing parameters just on the temporal scale. No improvements on the long term mean of radiative fields are found for any of these experiments.

The improved configuration of SKEB2 and the SPT with MSE and water conservation have been combined and put forward as the new stochastic physics setup for future GA releases (Sanchez C, 2015, *in preparation*). Presently this configuration has been included in the GA7 prototype for EPS and climate systems, the latter will be employed in the Coupled Model Intercomparison Project Phase 6 (CMIP6), a comparison of climate models for climate-change assessments (Meehl *et al.*, 2014).

8.1 Remaining Questions and Future Research

There are several questions issued on the introduction (chapter 1) as the most relevant for the development and evaluation of stochastic physics. The results of this thesis provide some some evidence to answer some of these questions, or suggest guidelines for future work to obtain a more precise response.

Are current stochastic physics schemes perturbing the right spatial and temporal scales?

No. Current schemes like SPT or SKEB2 degrade the skill of deterministic forecasts. Although they provide a good first-order approximation to represent model uncertainty (e.g. SKEB2 counter diffuses mid-latitude cyclones), they are too simple (SPT) or their formulation uncertain (SKEB2's dissipation masks). SPT produces a less harmful impact than SKEB2, despite having a similar or superior impact on the ensemble, indicating that maybe forcing physics coarsely is less harmful than forcing dynamics.

There are large uncertainties within stochastic physics schemes. Spatial and temporal correlation scales are normally "tuned up" for performance in an EPS. It requires laborious work to study their realism in terms of each of the processes they aim to represent. One example is the range of spatial scales where the SKEB2 forcing is active. Shutts (2013) found that the SKEB2's backscatter on the large-scales was not realistic. Investigations in sect. 6.1 have shown that that spurious artefacts may appear if these scales are forced (e.g. spurious westward equatorial wave for a highly tuned up version of the scheme).

Do stochastic physics schemes represent physical process correctly or there are fundamental flaws?

We have found examples where stochastic physics can introduce flaws in the representation of atmospheric processes.

- SKEB2 produces a spurious westward equatorial when the amplitude of the scheme is high and the forcing pattern includes large scales (sect. 5.2).
- SKEB2 and SPT degrades the power of already weak Kelvin waves (sect. 5.3 and 7.2).
- SPT has a serious impact on the energy and moisture budgets on climate simulations (sect. 7.2).

Internal closures need to be developed within Stochastic physics schemes to prevent the emergence of systemic flaws in the representation of atmospheric processes (e.g. a closure in SPT to conserve vapour and MSE). When the emergence of these

problems is unclear, there is a need to develop process-based diagnostics to understand the mechanisms that lead to the poor representation of these processes (e.g. Kelvin waves composites).

Should there be stochasticity in the deterministic model outside the ensemble forecasting context?

Stochasticity in a deterministic model is clearly an oxymoron! By deterministic model we mean a deterministic forecast where single events must have the highest predictability. With current model error schemes like SPT or SKEB2 there should not be. It has been shown that SKEB2 deviates individual storms in terms of position, intensity, speed and growth (sect. 5.1). Therefore it provides a poor deterministic forecast. SPT has also a detrimental impact on the deterministic skill, although much smaller.

For climate scales, stochasticity could be useful, as climate is an ensemble of weather states with no need for accurate predictability on the synoptic scales. The improvements found on the climate simulations (sect. 5.2), like the reduction of tropical biases, could improve the model on these long timescales.

The different ideas explored to improve the realism of the stochastic parametrizations have lead to a decrease of the ensemble spread, e.g the use of a more physically justified numerical dissipation mask for SKEB2 (sect. 6.2), or the inclusion of the conservation constraints in SPT (sect. 7.2). The development of stochastic physics schemes for a seamless model should employ different diagnostics, as done in the present thesis, to improve the representation of atmospheric processes.

Should uncertainty representations be developed alongside the physical parametrizations or added a posterior by model error schemes?

In an ideal world, stochastic parametrizations should be designed to represent atmospheric process implicitly, e.g. horizontal transport of convection represented by the Cellular Automata (sect. 2.6.3), or the stochastic fluctuation of mass-fluxes by the Plant-Craig scheme (sect. 2.6.1). These schemes provide a more physically-based representation of the error they aim to represent. However, in practical terms, EPS need urgent solutions to increase the ensemble spread. The research done with the mixed-phase scheme shows that the representation of internal processes, such as the

uncertainty of internal parameters, produces a lower impact in the ensemble than perturbing the tendencies (sect. 7.3). The advantage of stochastically perturbing physical tendencies over the perturbing internal parameters is also evident in sect. 7.2, where SPT clearly outperforms the present version of the random parameter scheme, although the comparison is not clear as they have very different forcing patterns.

In order to achieve an equivalent level of spread to model error stochastic schemes, several implicit stochastic physics schemes for different processes should be developed and combined. Hopefully new schemes such as the Stochastic Multi- Cloud Model (SMCM, sect. 2.6.5), or the Stochastic non-orographic gravity wave drag scheme (sect. 2.6.4) are developed further to replace current bulk-formula deterministic parametrizations.

How can we develop stochastic physical parametrizations in the presence of compensating model errors/ heavily tuned models?

A careful approach must be taken. One clear example is the study carried out in sect 5.3, where the impact of the SKEB2 across different configurations was evaluated. The scheme helps to offset the diffusivity of the interpolation to the departure point (see sect. B.2.1). This improves the intensity of storms and even fully removes the bias if the scheme is tuned up. Improvements in the Semi-Implicit scheme produce similar or larger impact on the same diagnostics.

The community of atmospheric scientist needs to estimate and understand the contributions from the different sources of error across timescales, in order to avoid stochastic schemes introducing compensating errors. Understanding the sources of error is not an easy task in a GCM, but with the help of process-based diagnostics, model biases could be traced down to particular processes.

Which priorities are the most relevant for the development of stochastic physics?

The results gathered in this thesis show that stochastic model error schemes (SKEB2 and SPT) can provide substantial benefits for probabilistic NWP and climate simulations. However, they also degrade short-range deterministic forecasts, indicating that their representation of model error is limited. These schemes have also large

uncertainties or are to simple. We believe that the development of implicit stochastic parametrizations (e.g. Plant-Craig, Stochastic Multi-Cloud Model) should be the way forward to represent the subgrid effects stochastically, as these schemes can provide a more physically based representation of the process the aim to represent. Such schemes could potentially improve deterministic forecasts as well.

The development of implicit stochastic schemes is an idealistic idea that could be challenged by the operational need to produce a sizeable ensemble spread. If this is the case, closures for current model error schemes should be developed. In this thesis, we have explored ideas like a more adequate numerical dissipation mask for SKEB2 or the conservation of water vapour and energy in the SPT. Future ideas could include a more physically-based convective dissipation mask and/or coarse-graining studies of physical tendencies to calibrate SPT (e.g. Shutts and Palmer 2007). The schemes with these closures can produce a more realistic representation of model error and reduce the negative impacts on deterministic forecasts, even though their benefits in the ensemble spread may be reduced.

The results found in this thesis are hard to extrapolate to other models, some of them contradict results found in previous studies made with the IFS (see sect. 5.4). The formulation of GCM and parametrizations may substantially differ (including stochastic physics schemes like SKEB). Although there are similar systematic biases in different models (e.g. tropical precipitation biases), their magnitude and structure may be quite different. It could be useful to carry out a comparative study of common stochastic schemes like SKEB2 or SPPT, such comparison can reveal how these schemes interact with parametrizations and their coupling with the atmospheric flow.

There is also a need to develop and incorporate new process-based diagnostic to understand the sources of model error, and its representation by the stochastic physics schemes. The use of cyclone tracking has been very useful to diagnose the effects made by the SKEB2's forcing. It would be interesting to employ similar tracking algorithms for mesoscale convective systems, cloud simulators to understand the impact on clouds, or the use of composites of MJO events, Kelvin waves, or Rossby waves to understand the nature of the tropical improvements and their connections to the extra-tropics.

The development of stochastic physics in a seamless framework should be under-

pinned by its physical realism, making them a key component of future prediction systems of the atmospheric challenges yet to come in a changing climate. The Seamless evaluation employed in this thesis aspires to ensure this crucial task.

Appendix A

Conservation issues of SPT

One of the main concerns about the Stochastic Parametrization of Tendencies (SPT, described in sect. 2.3 and evaluated in sect. 7.2) is the fact that its perturbations may substantially affect the energy and moisture balance, as these schemes randomly add temperature, moisture and wind increments with no physical constrains. In order to reduce the impact of the scheme on the global energy and water balance, an option for the SPT is developed to enforce water conservation and Moist Static Energy (MSE). In order to diagnose the impact of these new features of the SPT, diagnostics from the global energy correction scheme are included and are explained in the last section of the present appendix.

Water conservation

The vertically integrated water vapour is computed for all the model columns following eq. A.1, where z_{top} and z_{bot} are the upper and lower limit of the SPT tapering. A similar operation is performed after the SPT forcing is added (eq. A.2), where Δq is the humidity increment introduced by SPT.

$$Q = \int_{z_{bot}}^{z_{top}} \rho q dz \tag{A.1}$$

$$Q_{SPT} = \int_{z_{bot}}^{z_{top}} \rho (q + \Delta q) dz \tag{A.2}$$

Humidity perturbations that conserve the total Q are defined as $\Delta\widehat{q}$, and the ratio of vertically integrated water vapour after and before adding the SPT increments is defined as α, so $Q(i,j) = \alpha(i,j) \cdot Q_{SPT}(i,j)$. Equalizing both definitions of Q gives equation A.3.

$$\int_{z_{bot}}^{z_{top}} \rho(q + \Delta\widehat{q})dz = \alpha \int_{z_{bot}}^{z_{top}} \rho(q + \Delta q)dZ \tag{A.3}$$

To satisfy the relationship of eq. A.3, the water content must be equal in all levels such that $q + \Delta\widehat{q} = \alpha(q + \Delta q)$. Therefore the perturbations to the humidity that conserve total water are given by eq. A.4.

$$\Delta\widehat{q}(i, j, k) = \alpha(i, j)\Delta q(i, j, k) + q(i, j, k)(\alpha(i, j) - 1) \tag{A.4}$$

Conservation of MSE

In order to conserve the MSE, the temperature perturbations must related to the definition $MSE = c_pT + gz + Lq$, where c_p is the heating content at a constant pressure and L the latent heat of vaporization. The potential term $g\,z$ vanishes when T perturbations are linked to local q perturbations on the same gridpoint and level. Therefore, MSE conserving temperature perturbations are defined by equation A.5.

$$\Delta T(i, j, k) = -\frac{L}{C_p}\Delta q(i, j, k) \tag{A.5}$$

Energy conservation diagnostics

Climate configurations in MetUM make use of an Energy Correction (EC) scheme to avoid spurious drifts in the atmosphere's total energy. The EC scheme computes the energy error and places it back in form of an uniform temperature increase or decrease. The computation of the energy error ϵ for a given time equal to Δt follows eq. A.6, where ΔE is the change in total atmospheric energy in Δt and F is the energy flux into the atmosphere during Δt.

$$\epsilon = \Delta E - F \tag{A.6}$$

The total energy E is given in eq. A.7. It is the integral of energy over a spherical volume between the Earth surface and the Top of the Atmosphere (TOA), ϕ is the latitude, λ the longitude and r the radius of the Earth. The Energy is a sum where the first term represents the internal energy of an atmospheric parcel in virtue of its temperature, the second term is the potential energy represented by its influence to the Earth's gravitational field, the third is the kinetic energy given

by the momentum within the parcel and the fourth is moist energy and represents energy released or absorbed by water phase changes.

$$E = \int_{surf}^{TOA} \int_{-\pi}^{\pi} \int_{0}^{2\pi} \rho \left[c_p T + gz + \frac{1}{2}(u^2 + v^2 + w^2) - (L_c m_{cl} + (L_c + L_f) m_{cf}) \right] r^2 cos\phi d\lambda d\phi dr \tag{A.7}$$

The energy flux into the atmosphere is given by the sources and sinks of energy fluxes as shown in eq. A.8. The main source is the incoming radiation from the sun SW_{in}^{TOA}. Other components that increase the energy content in the atmosphere are the sensible heat sh_{in}^{surf} and latent heat represented by the rainfall r_{surf} and snowfall s_{surf} at the surface. Sinks of energy include outgoing radiation terms such as SW_{out}^{TOA} and LW_{out}^{TOA} plus radiation absorbed by the surface $SW_{netdown}^{surf}$ and $LW_{netdown}^{surf}$. a is the radius of the Earth.

$$F = \int_{\Delta t} \int_{-\pi}^{\pi} \int_{0}^{2\pi} [SW_{in}^{TOA} - SW_{out}^{TOA} - LW_{out}^{TOA} - SW_{netdown}^{surf} - LW_{netdown}^{surf} + sh_{in}^{surf} + L_c r_{surf} + (L_c + L_f) s_{surf}] a^2 cos\phi d\lambda d\phi dt. \tag{A.8}$$

If there is too much dissipation of energy, as in the case of models with Semi-Lagrangian Semi-Implicit dynamical cores (sect. B.2.1), the kinetic energy term will decrease and thus ΔE would be lower and the energy correction slightly negative, therefore the model will add a positive increment and thus the ϵ diagnosed will be positive.

Appendix B

Description of a Global Circulation Model

Atmospheric prediction is an initial-value problem, run (integrated) over time for days in the case of short-range Numerical Weather Prediction (NWP), months for seasonal prediction, or centuries for the case of long climate experiments. The ability to make a skilful NWP forecast requires a realistic representation of the atmosphere in the equations (good model) and an accurate initial conditions (good observations). The latter is less relevant in climate models as they tend to simulate the average effects of the flow rather than provide a prediction for a given time and location as done for NWP. The present Appendix provides a description of the principles and components of a NWP and climate model.

The weather prediction model was envisioned by Bjerknes (1904). He listed seven unknowns that govern the evolution of the atmosphere: Pressure P, temperature T, density ρ, humidity q and the three components of velocity vector $\mathbf{v} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$. The model's equations is a set of 7 independent equations: The hydrodynamic equations of motion or equation of momentum for the three coordinates (eq. B.1), continuity equation or conservation of mass (B.2), the equation of state for ideal gases (B.3), the first law of thermodynamics or conservation of energy (B.4) and a conservation equation for water mass (B.5)

$$\frac{d\boldsymbol{v}}{dt} = -\frac{1}{\rho}\boldsymbol{\nabla}(P) - \boldsymbol{\nabla}(\phi_e) + \boldsymbol{F} - 2\boldsymbol{\Omega} \times \boldsymbol{v} \tag{B.1}$$

$$\frac{\partial \rho}{\partial t} = -\boldsymbol{\nabla}(\rho \boldsymbol{v}) \tag{B.2}$$

$$P = \rho RT \tag{B.3}$$

$$\frac{Q}{T} = C_p \frac{dlnT}{dt} - R\frac{dlnP}{dt} = C_p \frac{dln\theta}{dt} = \frac{dS}{dt} \tag{B.4}$$

$$\frac{dq}{dt} = E - C \tag{B.5}$$

In addition to the prognostic variables described above, the model incorporates other diagnostic variables and parameters. The equation of momentum (B.1) contains $\boldsymbol{\nabla}(\phi_e) = g\boldsymbol{k}$, the Newtonian gravitational potential force, the frictional force $\boldsymbol{F}$ and the angular velocity of the rotation of the Earth $\boldsymbol{\Omega}$. The equation of state of an ideal gas (B.3) includes the parameter R, the gas constant for air or a corrected version to account for the effects of humidity. The thermodynamic equation is defined by the diabatic heat Q or the rate of change of Entropy S where C_p represents the coefficient of specific heat at constant pressure, there is a more elegant way to show the equation of energy conservation as the change of potential temperature θ (given by eq. B.6, where P_0 is a reference pressure), and shows that potential temperature is conserved in absence of diabatic heating. The conservation of water vapour mixing is governed by the difference between the sources of Evaporation (E) and the sinks of Condensation (C).

$$\theta = T(P_0/P)^{R/C_p} \tag{B.6}$$

These equations lack a general analytical solution. In some idealized cases wave-solutions can be found, such as Rossby waves, gravity waves or tropical waves. Other approximations filter the low-order magnitude terms (quasi-geostrophic model), takes the vertical motion of the flow as small perturbations about a mean height (shallow water approximation), or neglects sources of vertical acceleration other than pressure differences (hydrostatic approximation). For a general derivation of the model's equations and their wave solutions under these cases, the reader is refereed to general atmospheric dynamics text-books such as Holton (1972) or Kalnay (2002).

In order to find a numerical solution, equations need to be discretized over a given grid-size (of the order of $1 - 200\,km$) and timesteps (of a few minutes to hours),

additional terms need to be included to represent the sources and sinks of processes below truncation. The nature of these terms is explained with the following example: Consider the conservation equation for water vapour written in flux form, expanding the total derivate and doing the operation $q \times$ eq. (B.2) $+\rho \times$ eq. (B.5).

$$\frac{\partial(\rho q)}{\partial t} = -\boldsymbol{\nabla}(\rho \boldsymbol{v} q) + \rho(E - C) \tag{B.7}$$

In the real atmosphere, the variables contain scales that are resolved by the model's grid and smaller subgridscales. We represent the spatial average over a grid by an over-bar and the primes as the sub-grid perturbation. The prognostic variables are then (the effect of subgrid fluctuations in ρ is neglected):

$$u = \bar{u} + u' \quad q = \bar{q} + q' \tag{B.8}$$

The "Reynolds" averaging procedure is applied in the equation (B.7), by definition the grid-box average of subgrid quantities is zero, e.g. $\overline{u'\bar{q}} = \bar{u'}\bar{q} = 0$. Also grid-average remains unchanged with the Reynolds averaging, e.g: $\overline{\bar{u}\bar{q}} = \bar{u}\bar{q}$. The grid-average equation of (B.7) is thus:

$$\frac{\partial(\rho\bar{q})}{\partial t} = -\nabla(\rho\overline{\boldsymbol{v}}\overline{q}) - \nabla(\rho\overline{\boldsymbol{v}'q'}) + \rho(\bar{E} - \bar{C}) \tag{B.9}$$

The first term on the right hand side of eq. (B.9) is the grid-scale or resolved advection term of moisture. The second term is the divergence of the eddy fluxes of moisture or turbulent moisture transport. The last term is the molecular-scale phase transitions of water.

The resolved terms are defined as "dynamics" or "large-scale flow" and can be computed explicitly. The second and third terms on the r.h.s. of eq. B.9 are the sinks and source terms of water vapour in the gridbox, they are defined as "subgrid" or "parametrized" terms and they are quite important, without them the model integrations are not realistic after the first or second day (Kalnay, 2002).

The large-scale terms are resolved using discretization methods, this process is defined as "dynamical core". Their implementation and sources or errors is described in section B.2. The subgrid-scale terms are provided by physical parametrizations or simply "physics", like the diabatic radiative effects (Q), water phase-transitions (E and C), convective adjustment of vertical motions, turbulence of heat, momentum

and water vapour caused by the interaction between the atmosphere and the lower boundary condition and the effects of orography on the flow. A brief explanation of the different parametrization schemes can be found in section B.3. In the next section a brief history of the weather and climate modeling is given, which details in the development of NWP, climate models and lately the "seamless" model

B.1 History of atmospheric prediction and climate modelling

A weather forecast using Bjerknes model (eq. 1-5) was undertaken by Richardson (1922). He applied a finite difference method to solve the system, but the results after 6 hours were completely unrealistic due to an imbalance in the initial data used (Lynch, 2006). Additionally a large timestep was employed, which was later discover to be too large for the spatial domain, as it breached the Courant Friedrichs and Lewy (CFL) stability criteria (Courant *et al.* 1928), see section B.2 for details). The existence of these destabilizing high waves made impractical to integrate the full set of Berjknes equations, a.k.a. "primitive model" equations.

The primitive equations were simplified using the technique of scale analysis to filter high frequency components (Charney, 1951). The resulting equations are know as the "quasi-geostrophic system"and can produce reasonable 24 hour forecasts of large-scale structures. Nevertheless, these forecasts weren't accurate enough and research using the primitive equations continued. The first application of the primitive equations was a success, producing a good simulation of the development and occlusion of a frontal structure (Hinkelmann, 1951).

As primitive equation models became global, they were run for the whole sphere (e.g. Kasahara *et al.* 1967). Although essentially they were the same models, they were named differently as General Circulation Models (GCMs).

The effects of unresolved vertical motions were included by Manabe *et al.* (1967), with the development of a dry and a moist convective parametrizations, the latter was quite important as it also represented the radiative effects of clouds. Turbulent motions were represented by a parametrization based on the similarity theory of Monin and Obukhov (1954), and the effects of small gravity waves through the vertical flow by Palmer *et al.* (1986).

The discovery of the weather as a chaotic system by Lorenz (1963) lead to the development of Ensemble Prediction Systems (EPS, Leith 1974), a combination of different realizations of the same model. It went alongside the development of Data-Assimilation techniques to represent error in initial conditions (see Kalnay 2002 for an overview). More recently, EPS have also incorporated stochastic physics schemes to represent model uncertainties (Palmer, 2001).

On a separate area of the atmospheric sciences, climate models were built from low resolution GCMs adding other components of the Earth system such as Ocean or Sea-Ice. Described as "Coupled Models", many of them needed a flux adjustment of heat and moisture on the exchange layer to produce good simulations (Lynch, 2007). The climate models gained more and more complexity, adding some extra processes such as the effects of aerosols or land-use changes (e.g. Collins *et al.* 2011). These models are now defined as Earth System Models (ESM).

The complexity and resolution of NWP and climate models have been continuously growing. Nowadays climate models are able to simulate the day-to-day weather phenomena, a very valuable capability to produce studies of extremes under climate change (May, 2008). On the other hand NWP benefits from the representation of Earth-System processes such as the radiative effects of biomass-burning, an important process driving short-range cloudiness and precipitation in tropical areas (Milton *et al.*, 2008). All these benefits lead to the conclusion that a unique model could be employed for all timescales, such model is defined as seamless model, with the additional schemes are bolted on particular systems if required, i.e DA for EPS or sea-ice representation for long-timescales simulations.

B.2 Dynamical core

Before the dynamical core gets into action to solve the primitive equations, a few choices must be made about the "resolution" of the model: the length of the horizontal grid-size, number of vertical levels and the timestep. The sensitivity of the results to these choices is usually very high, with the horizontal resolution being the most important, as higher resolution enables more processes to be resolved. Pope and Stratton (2002) found that some features of the climate of a GCM converge when horizontal resolution increases (e.g. mid-latitude variability), while a number

of significant features do not, such as rain or moisture processes. However, Demory *et al.* (2013) has showed that the global hydrological cycle can converge at higher resolutions (below 60km).

The choice of resolution is generally given by the computational availability rather than a well-defined scale separation, which is absent in the atmosphere. This lead to important errors below the truncation limit, which can cascade upward to large scales contaminating the skill of the forecast (Tribbia and Baumhefner, 2004). The non-linear feedbacks across different scales is one of the main motivations for stochastic physics schemes. Therefore the present section contains a short description of the dynamical core and the structural uncertainties, which may affect feedbacks between resolved and parametrized scales.

The numerical solutions for the primitive equations can be obtained by finite difference methods, where values from the previous or next timesteps and adjacent points are taken to compute the derivates, or spectral methods, where the variables are expanded in terms of a finite series of orthogonal functions (generally spherical harmonics). The former provides a better representation of local discontinuous behaviour (Cullen and Davies, 1991), whereas spectral models are faster to compute at medium resolutions and do not have the problem of the singularity at the poles.

When a finite difference method is applied to a Partial Differential Equation (PDE) it becomes a Finite Difference Equation (FDE). The solutions of the FDE may not converge to the PDE if the wrong choices are taken. There are different methods to substitute the derivates, like using the difference between the previous and future timestep (centered approach), the difference between the present and previous timestep (explicit) or between the future and present timestep (implicit). The centered approach is generally the most accurate as its truncation error is of second order, but a combination of the centered and implicit, known as "Semi-Implicit", is the most reliable (Simmons *et al.*, 1978). In addition to the truncation errors, some choices for the model setup could introduce computational wave-like phenomena and unstable solutions. One example is the CFL criteria. We consider the example of a one-dimensional advection equation (B.10), and produce the FDE applying the "upstream scheme" method (B.11).

$$\frac{\partial u}{\partial t} = -c\frac{\partial u}{\partial x} \tag{B.10}$$

$$\left(\frac{\partial u}{\partial t}\right)_n \sim \frac{u_{n+1} - u_{n-1}}{2\Delta t} \tag{B.11}$$

The solution of (B.11) is given in equation (B.12), where $\mu = c\Delta t/\Delta X$ is know as the Courant number. If the condition $0 <= \mu <= 1$ is not satisfied, the solution is not bounded and after a few timesteps solutions will blow up, this is the CFL condition (see sec. 3.2.3 of Kalnay 2002 for details).

$$u_j^{n+1} = (1-\mu)\, u_j^n + \mu\, u_{j+1}^n \tag{B.12}$$

The constrain of the CFL became an important burden as higher resolutions demanded shorter timesteps. The need for a GCM which could employ longer timescales lead the development of the Semi-Lagrangian technique (SL, see review of Staniforth and Cote 1991), which does no longer need to comply with the CFL stability criterion and also posses a more accurate treatment of advection (Bermejo, 1990). Most of the state-of-the-art GCMs employ the Semi-Implicit Semi-Lagrangian (SI-SL) scheme as their dynamical core (Ritchie *et al.*, 1995; Davies *et al.*, 2004). However, the SL method has two main disadvantages: (i) Absence of formal conservation properties, (ii) a higher internal diffusivity than previous schemes. The dissipation created by the SL scheme is part of the physical motivation for one of the most employed stochastic physics schemes, the Stochastic Kinetic Energy Backscatter (SKEB, see sec. 2.2), which is widely tested in this thesis. Therefore a brief explanation of the SI-SL scheme and its diffusive properties is given in section B.2.1. On section B.2.2 one of the main consequences of the implicit dissipation of energy is explained, the dissipation of the spectra of Kinetic energy on the mesoscale.

B.2.1 Semi-Lagrangian scheme

Semi-Lagrangian advection is based on the interpolation of fields from a "departure point" most often using a backward Lagrangian trajectory. Considering the first-order prognostic equation of a scalar field F with a source term Ψ:

$$\frac{dF}{dt} = \Psi \tag{B.13}$$

This equation may be integrated between times $t^n = n\Delta t$ and $t^{n+1} = t^n + \Delta t$ following the parcel of air that arrives at gridpoint x_a at time t^{n+1}, the gridpoint x_a is called the "arrival point". The location of the parcel at time t^n is represented by

x_d and defined as "departure point"of the parcel, which is generally not a gridpoint. The change in F between time t^n and t^{n+1} is simply the integral of Ψ along its trajectory over the relevant time interval, where $F^{n+1} = F(x_a, t^{n+1})$ and $F_d^n = F(x_d, t^n)$.

$$F^{n+1} - F_d^n = \int \Psi dt = \bar{\Psi}\Delta t \tag{B.14}$$

Errors are inevitably introduced via the estimation of the departure point x_d, estimation of departure point value F_d^n and estimation of the trajectory time-average $\bar{\Psi}$. These estimations require interpolation. To obtain accurate results from a SL integration scheme it is necessary to choose the order of interpolation carefully. Interpolation using high degree polynomials is more accurate and gives much less damping, but on the other hand it has an additional computational cost.

Linear interpolation is adequate for the terms used in the evaluation of the trajectory, but more accurate interpolation is essential for the terms evaluated at the departure point (Staniforth and Cote, 1991). There are several interpolation schemes to the departure point, a list from the more diffusive to the more accurate follows:

- Linear interpolation, it is not suitable as the damping is too large for all the scales.

- Quadratic Lagrange, it is more viable and thus it was used in the early studies of the scheme.

- Cubic interpolation gives very little damping and it is mostly at small scales.

- Quasi-cubic interpolation is a blend between linear and cubic interpolations (see Figure 2 of Ritchie *et al.* 1995 for a schematic illustration the interpolation method). It is faster but it can sharp gradients erroneously producing spurious extrema (Wood N, *unpublished results*).

- Quintic interpolation: It is one of the highest order (5^{th}) schemes. It has a positive impact on the accuracy but additional cost. Sanchez *et al.* (2013) found it was about 8% more expensive than the Quasi-cubic for a low resolution climate run.

The choice of interpolation scheme can have a large influence on the energetics of low resolution models below grid-sizes of 200km (Chen *et al.*, 1997; Stratton, 2004). They lead to a weak mid-latitude variability (Greeves *et al.*, 2006; Sanchez *et al.*, 2013). McCalpin (1988) computed the amplification factors $|\lambda|^2$ of the SL scheme, a $|\lambda|^2 > 1$ would lead to computational instabilities, whereas a $|\lambda|^2 < 1$ damps the solution. He found that linear interpolation has the same spectral characteristics as Laplacian viscosity (∇) and Quadratic and Quintic are equivalent to biharmonic viscosity (∇^2).

B.2.2 Diffusion of the Kinetic Energy Spectra

The Kinetic Energy (KE) of the atmospheric motions is dissipated by state-of-the-art GCMs, either explicitly to prevent numerical instabilities or implicitly by the dynamical core. This excessive energy dissipation may result in insufficient variability and underdispersive ensembles (Thuburn *et al.*, 2013).

An important impact on the KE dispersion is believed to affect the energy transfer at smaller scales. An observational study of Nastrom and Gage (1985) showed that KE spectra follows a k^{-3} dependence on a large scale, where k is the wavenumber. The k^{-3} regime is dominated by rotational modes. Around wavelengths of 400 km there is a transition to a shallower $k^{-5/3}$ dependence dominated by divergence modes. Their study shows a similar curve for horizontal wind, vertical wind and potential temperature (Fig. B.1). The KE spectra was computed from observations made by instruments on-board commercial airlines.

The k^{-3} dependence on the large-scale is well explained by 2-D turbulence, but the cause of the $k^{-5/3}$ is still under discussion. Tung and Orlando (2003) argue that the -5/3 slope is produced primarily from a forward energy cascade (downscale, from larger to smaller scales), whereas Lilly (1989) argues that a small amount of energy injected at small scales (from convection or other sources) cascades upscale, producing the shallow spectra.

GCMs with low resolutions have problems in simulating the $k^{-5/3}$ slope in the kinetic energy spectra. Terasaki *et al.* (2009) showed the spectra of different horizontal resolutions for the Nonhydrostatic Icosahedral Atmospheric Model (NICAN, Satoh *et al.* 2008). They conclude that a grid size less of 10 km is needed to reproduce the $k^{-5/3}$ power spectrum for NICAN. Similar results have been reported running the

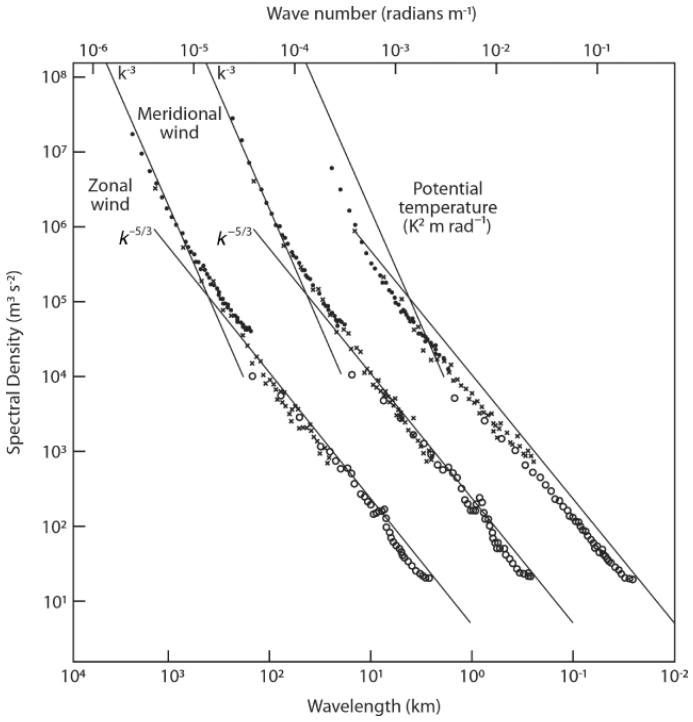

Figure B.1: Variance power spectra of wind and potential temperature based on aircraft observations. Lines with slopes -3 and $-5/3$ are included for each variable, for comparison. *From Nastrom and Gage (1985)*

Met Office Unified Model (MetUM, sect. 4.1) across different horizontal resolutions (Roberts M, 2013, *personal communication*). The inability of atmospheric models to simulate the $k^{-5/3}$ slope has important repercussions in their predictability, as error in mesoscales could propagate upscale towards the large scales within one or two days (Tribbia and Baumhefner, 2004).

B.3 Parametrizations

The problem of representing the bulk effects of features smaller than the gridbox on the resolved scales is not a trivial one. These features are quite complex and too diverse to be easily generalized, like clouds, or the effect of orographic mountains over the flow. In addition, these features are fundamental drivers of atmospheric flow, e.g. the radiative heating or cooling caused by molecular absorption and scattering, heating effects of the cumulus clouds over the tropics. Despite its importance, the formulation of parametrizations is restricted to represent the essential aspects of the physical processes, as they should ensure mathematical, computational and conceptual tractability (McFarlane, 2011).

Parametrized processes include gravity waves caused by orography or a rapid release of latent heat in deep convective updraughts (described in sect. B.3.6), radiative effects of solar and terrestrial radiation on the Earth's surface and atmosphere (sect. B.3.1), the formation and evolution of clouds and precipitation (B.3.2 and B.3.3), different kinds of turbulence (B.3.4) and the most uncertain of these processes, the moist convection events where water ascends to compensate an energy imbalance (B.3.5).

There are important links amongst parametrizations. They strongly influence each other through their changes in the large-scale variables. When clouds are developed they alter the heating profile of the column, leading to changes in convection and the internal composition of the cloud, which might lead to various forms of precipitation. Many of the physical processes naturally re-arrange energy in the vertical column. Therefore Parametrizations generally focus on the effects of subgrid physical processes within the vertical column. However, there are important subgrid phenomena that emerge from the horizontal transport of heat, moisture or momentum (Huang, 1990). These interactions are neglected by the current formulation of physical parametrizations.

As cited above, the development of realistic parametrizations is paramount for an adequate representation of the atmosphere, therefore they are subjected to a very intense research that covers many different areas of the parametrized processes. The present section gives a simple explanation of the different parametrizations with an emphasis on its uncertainties. For a more detailed description the reader is referred to textbooks about parametrizations like Stensrud (2007) or Trenberth (1992).

B.3.1 Radiation

The electromagnetic radiation interacts with several atmospheric components through absorption, emission and scattering at molecular and micro scales. These interactions are usually the major influence on local and global temperature. Given that the source of radiation is quite different from solar and terrestrial sources, their radiative effects in the atmosphere are consider separately.

- Short-Wave (SW) radiation ($\lambda \leq 3.7\,\mu m$). It comes from the Sun, and is absorbed by Ozone in the stratosphere and scattered by clouds and aerosols in the troposphere.

- Long-Wave (LW) radiation ($\lambda \geq 3.7\,\mu m$). It is emitted by the Earth surface and atmosphere, dominated by absorption of water vapour and clouds in the troposphere and Carbon dioxide in the stratosphere.

The parametrization of radiation provides a fast and accurate method to determine the heat released or absorbed by radiative processes, represented by the Q_{rad} term, the major component of the Q term in the energy equation (eq. B.4). It is the sum of the net SW and LW in a given point of the atmosphere (eq B.15), where dF^{net} is the divergence between the upward and downward radiative fluxes, and it is computed by the radiative transfer equation.

$$Q_{rad} = \frac{g}{c_p}\left(\frac{dF^{net}_{SW}}{dp} + \frac{dF^{net}_{LW}}{dp}\right) \tag{B.15}$$

Net fluxes are computed by the radiative equation, an integral of the absorption over the frequency spectra determined by the amount of the absorber and the absorption coefficients $k_\nu(p,T)$. The most used method to integrate the radiative equation is the correlated-k method (Fu and Liou, 1992), where wavelengths with similar $k_\nu(p,T)$ are binned together into "quadrature points" defined as k. This approximation introduces a marginal error in comparison to more complete "line by line models" where the absorption is computed for each spectral line ν. The integration of the radiative equation is one of the most expensive bits of a GCM given that the radiative equation must be integrated for all the quadrature points in each gridpoint on the 3-D field. The radiation code is thus called every 2-3 hours, a timescale longer than GCM timesteps.

The approximations made to obtain the radiative fluxes are quite accurate for clear skies, where radiation only interacts with gases and aerosol particles. However, when clouds appear they change the radiative fluxes of SW and LW substantially. Normally the radiation parametrization follows the simple approach given in eq. B.16, where F_{clr} is the clear sky radiation, F_o is the overcast sky and C is the fractional cloud amount.

$$F = F_{clr}\,(1 - C) + F_o\,C \tag{B.16}$$

Radiative fluxes from clouds (F_o) are computed following the assumption that clouds are randomly overlapped in a column, which is very simplistic to real cloud-radiation interactions. These assumption lead to important biases on the heating

rates (Barker *et al.*, 1999). There are also large uncertainties about the radiative properties of the cloud composition, e.g. ice water has complicated scattering relationships due to its high density of shapes (Fu, 2007).

A different approach to represent the radiative effects of cloud inhomogeneity was taken by Pincus *et al.* (2003). They developed the Monte Carlo Independent Column Approximation (McICA), a scheme that employs stochastic approaches to calculate the radiative fluxes in vertical subcolumns. Some of the details of the scheme are explained in the next subsection.

Monte Carlo Independent Column Approximation (McICA)

The McICA scheme aims to minimize the cost of the Independent Column Approximation scheme (ICA), described in Cahalan *et al.* (2004); Barker *et al.* (1999). The ICA scheme splits the GCM column into a number N of subcolumns with the column's water content distributed to different "overcast" subcolumns. The radiative transfer calculation is performed for each subcolumn independently.

The ICA method provides accurate domain-average fluxes. However, their computational expense is far too high for operational GCMs, given that it needs to perform N times more radiative-transfer calculations. The McICA scheme randomly chooses a set of subcolumns for each quadrature point, minimizing its cost but introducing conditional random errors. The subgrid structure of the cloud is simulated by the algorithm of Räisänen *et al.* (2004), which provides vertical overlap cloud fraction and cloud condensate.

The effects of the McICA noise have been extensively studied for GCMs across timescales. Barker *et al.* (2008) reported that the noise has statistically insignificant effects on 6 GCMs. Hill *et al.* (2011) tested McICA in MetUM. The original version of McICA gave worse short-range forecasts of near-surface temperature than the previous parametrization of radiation which had cloud random overlap assumptions. Including two methods to reduce the McICA noise (restricting the random sampling of subcolumns to the cloudy ones and assigning the most important k-terms to each subcolumn), the surface temperature forecasts were improved.

Despite of its stochastic nature, the McICA scheme is different from current stochastic physics schemes such as those described in chapter 2. Its distinction reflects the fact that radiation is a well understood process and there is very little

room to represent the uncertainty in this aspect. McICA is designed to reduce model error by providing approximate solutions to the full problem rather than exact solutions to some approximate problem.

B.3.2 Large-Scale Cloud scheme

Clouds are the manifestation of phase changes in the atmosphere within a complex turbulent flow and chemical interactions with aerosols (Bodenschatz *et al.*, 2010). Clouds are a fundamental component of the atmosphere, as they strongly interact with solar and terrestrial radiation changing the local temperature and thus influencing the general circulation. They are also one of the main drivers of the hydrological cycle, through condensation and precipitation. Clouds have therefore a very important role on weather forecasts and climate predictions.

The representation of the effects of clouds is artificially divided in a GCM. The moist convective parametrization simulates cumulus clouds caused by strong vertical ascent (see sect. B.3.5), whereas the more usual stratiform clouds are simulated by the Large-scale cloud schemes. The internal composition of the clouds and precipitation processes are also treated separately in microphysics schemes (a.k.a. large-scale rain schemes, see sect. B.3.3).

The main purpose of a Large-scale cloud scheme is to calculate the amount of condensate q_{cl} and fractional coverage of cloud within a gridbox C_l. Quantities that are handed to the microphysics scheme to estimate the composition and properties of water and ice, and later to the radiation parametrization to compute the radiative effects of these particles.

In cloud schemes, the local variations of moisture and temperature from the gridbox mean, defined as s, can be described by a Probability Density Function (PDF) defined as $G(s)$ (Sommeria *et al.*, 1977). The gridbox values are weighted by $G(s)$ as shown in eq. B.17 for C_l, and B.18 for $\overline{q_{cl}}$, where Q_c represents the mean condition of the gridbox.

$$C_l = \int_{s=-Q_c}^{\infty} G(s)ds \tag{B.17}$$

$$\overline{q_{cl}} = \int_{s=-Q_c}^{\infty} (Q_c + s)G(s)ds \tag{B.18}$$

The cloud parametrization problem is essentially one of how to parametrize the form of $G(s)$. Some schemes use a parametrized PDF such as Smith (1990), others diagnose the PDF from the cloud's prognostic variables (Tiedke, 1993), e.g. cloud erosion by entrainment reduces the width of the PDF. Wilson *et al.* (2008) provides a discussion of the different strategies to determine $G(s)$.

B.3.3 Microphysics

The composition of clouds determines their radiative properties (e.g. brightness), the internal turbulence that drives circulation around clouds, and precipitation events of different kinds (drizzle, graupel, snow). Therefore the parametrization of microphysical processes is fundamental for a GCM. Nevertheless, large uncertainties remain. Clouds are one of the primary sources of uncertainties for the climate predictions (Heintzenberg and Charlson, 2009). Different NWP models can produce a large range of simulated ice water paths (Waliser *et al.*, 2009).

The difficulties in building a realistic microphysics scheme lie in the large number of interactions amongst cloud particles. Figure B.2 shows a flow diagram of these interactions in the state-of-the-art microphysics scheme of Seifert and Beheng (2006), a two dimension scheme that computes the particle mixing ratio and concentration. In addition to the interactions, there is also a wide variety of the properties of one particle, e.g. there are 80 different types of ice habits or shapes (Pruppacher and Klett, 2010). A wrong choice of habit can lead to errors in the radiative properties, formation of precipitation and the evolution of the cloud as a whole (Khvorostyanov and Curry, 2005).

Modeled precipitation is also unrealistic in many aspects (Stephens *et al.*, 2010). One of the most common errors is the excessive drizzle produced by strato-cumulus clouds (Bodas-Salcedo *et al.*, 2011). The precipitation is controlled by the raindrop size distribution (larger drops fall faster) and follows Marshall and Palmer (1948) simple gamma function, given in eq. B.19, where $N(D)$ is the density of droplets as a function of diameter D, N_0 is the intercept parameter and λ the slope parameter. In many microphysics schemes N_0 is assumed to be a constant, although observations show that it may vary over 3 orders of magnitude for ice (Field *et al.*, 2005). A more complex relationship for N_0 could improve the representation of drizzle (Abel and Boutle, 2012).

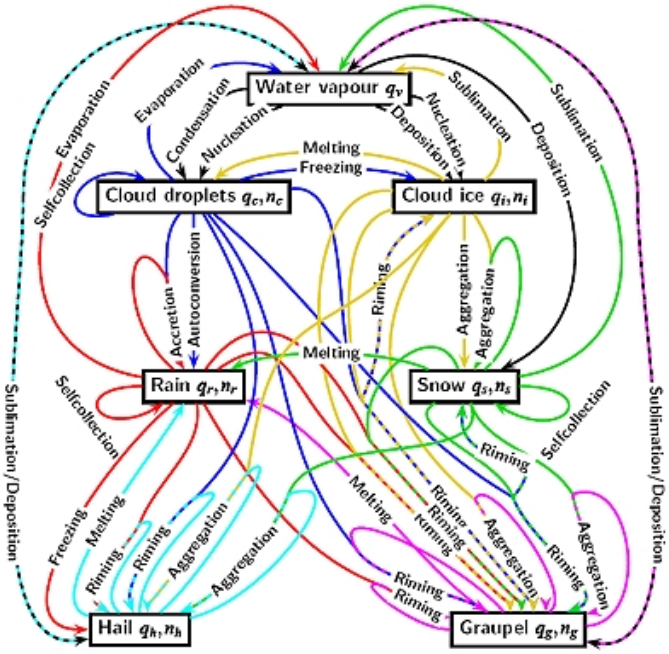

Figure B.2: Flow diagram of microphysical processes as implemented in the Seifert and Beheng (2006) two-moment microphysics scheme. *From Karlsruhe Institute of Technology website*

$$N(D) = N_0 e^{-\lambda D} \tag{B.19}$$

The effects of aerosols in clouds and precipitation are also important and not well simulated by present models. They influence radiative aspects such as brightness of clouds and water aggregation. Many of the Earth Model Systems include interactive aerosols schemes where the evolution of certain species is dependent of photo-chemical processes and atmospheric thermodynamics. The impacts of aerosols in NWP timescales are also important (Milton *et al.*, 2008). For a detailed description of an aerosol scheme, its impacts and their degree of complexity, the reader is referred to Mulcahy *et al.* (2014).

B.3.4 Planetary Boundary Layer (PBL)

The atmosphere is in a radiative imbalance. Outgoing radiation from the Earth's surface and Ocean does not compensate for the absorption of SW, the energy balance is compensated by an additional transfer of energy from the surface to the atmosphere know as "sensible heat", and evaporation of moisture from the surface cover

to the atmosphere known as "latent heat". In addition, there is also a surface flux of momentum caused by wind shear. The lower section of the atmosphere, where the interactions amongst these fluxes and the atmospheric flow take place, is know as the Planetary Boundary Layer (PBL or BL). It can be as shallow as 100m during night to few thousand meters when the atmosphere is heated from the surface.

The mechanism of the atmosphere to redistribute the surface fluxes throughout the PBL is turbulence. It has timescales from seconds to few hours and spatial scales of centimeters to few hundred meters (Stensrud, 2007). Therefore they are small scale and should be parametrized. The PBL parametrization determines together with the surface parametrization the surface fluxes and their mixing over the boundary layer.

The PBL and land-surface parametrizations are crucial for climate, as they determine the radiation reflected (albedo) by the surface and ocean, besides the hydrological cycle through evaporation. They also interact with radiation through stratocumulus clouds and fog, and control the triggering of convection through instabilities from the PBL. For weather forecasts, the realism of processes occurring in the PBL is also important as many forecast products are needed near the surface (e.g. fog forecast for airports, surface temperature for icy-roads).

Turbulent terms are characterized by the correlations of subgrid variables, such as the turbulent moisture term $\overline{v'q'}$ of equation B.9. The traditional approach has been to represent these terms as turbulent diffusion (eq. B.20), which is know as "first-order approximation"(Louis, 1979).

$$\overline{\boldsymbol{v}'q'} = K\frac{\partial \overline{q}}{\partial z} \tag{B.20}$$

In the MetUM boundary layer scheme, the nature of the K coefficients is given by the different types of boundary layer (Lock *et al.*, 2000). There are 6 different types which are classified according to its stability and presence of clouds. Figure B.3 provides a schematic representation of the six types of boundary layer with the profiles of virtual potential temperature θ_v and $K(z)$. For stable types, the first-order approximation works well. However, more complex derivations for K, or even higher order closures are required for unstable conditions. For a detailed description of the PBL and its parametrization, reader is referred to Stull (1988) textbook.

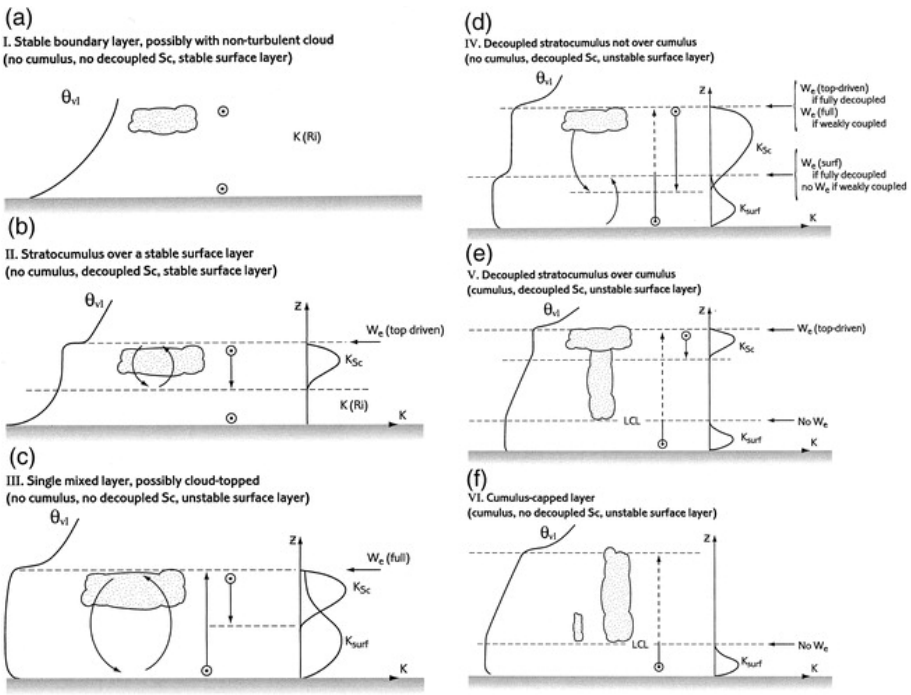

Figure B.3: Schematic representation of the six boundary layer types, with profiles for surface turbulence K_{surf} and Stratocumulus turbulence K_{Sc}. *From Lock* et al. *(2000).*

In spite of some advances, the boundary layer is not represented realistically in models (Teixeira *et al.*, 2008b). There are some important issues that need to be addressed such as the representation of subgrid vertical fluxes, cloud fraction and water for the development of more general parametrization that represents all types of boundary layers.

In terms of land-surface parametrizations, the representation of surface roughness, vegetation or soil types is controlled by many uncertain or unphysical parameters. The reader is refereed to Overgaard *et al.* (2006) for a review on land-surface schemes and their uncertainties.

B.3.5 Convection

Surface fluxes of energy and moisture have the potential to perturb the atmospheric vertical profile of temperature and thus give rise to buoyancy forces that provoke a vertical ascent of air masses. When these air masses reach their level of condensation, they release latent heat and feed back on the column instability. This phenomena is know as moist convection and is one of the most challenging process to parametrize given its uncertainties and importance. Moist convection is associated

to the emergence of cumulus clouds, therefore the convection parametrization is also known as cumulus parametrization.

The strength of the convective instability determines its vertical extend. Weak episodes are defined as shallow convection. They are constrained to the boundary layer and give birth to low-level stratocumulus, the most prominent type of cloud in the subtropical oceans. Strong convection is defined as deep convection, where vertical displacements may reach the tropopause. Deep convection is a crucial aspect of the tropical climate and an important driver of the large-scale circulation in the mid-latitudes. Convection is also associated to a downdraft (or subsidence) around the cloud to conserve mass, mixing of environmental air in the cloud, known as entrainment, and cloudy air with the environment, known as detrainment.

Moist convection is an amalgam of updrafts and downdrafts across different scales. Figure B.4 shows a schematic description of the intrinsic cumulus processes whose formulation poses large challenges. In addition to the intrinsic uncertainties in the representation of one particular convective cloud episode, the representation of the large scale effects of an ensemble of these clouds is even more challenging.

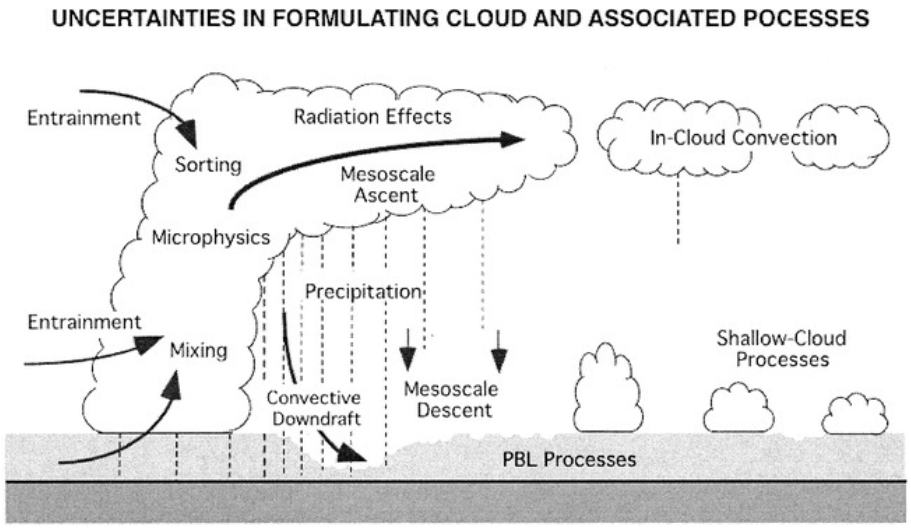

Figure B.4: Cumulus cloud associated processes where major uncertainties in formulation exits. *From Arakawa (2004).*

The strength of convection is normally represented by the Convective Available Potential Energy (CAPE), the maximum energy available to an ascending parcel (Emanuel, 1994). CAPE is related to the terms that represent convection in the large-scale equations, the apparent heat source Q_1, the apparent moisture sink Q_2

and the apparent momentum sink Q_3. Following Yanai *et al.* (1973), the budget of the apparent source for moist static energy $h = C_pT + gz + L_vq$ is given in eq. B.21.

$$Q_1 - Q_R - Q_2 = -\frac{\partial\overline{\omega' h'}}{\partial p} \tag{B.21}$$

There are different approaches to the cumulus problem. A few review studies explain their main advantages and disadvantages (Arakawa, 2004; Plant, 2010). Probably the most used one is the mass-flux approach of Arakawa and Schubert (1974) and Tiedtke (1989), where convection is characterized by an ensemble of plumes (cumulus) over an area of tolerably uniform forcing. These plumes are characterized by the convective mass flux (eq. B.22), where σ is the fractional area covered by cumulus, ω^c is the bulk vertical velocity of the convective plumes on pressure coordinates and g is the gravity constant.

$$M = -\frac{\sigma\omega^c}{g} \tag{B.22}$$

The vertical eddy fluxes that define Q_1, Q_2 and Q_3 are proportional to the difference between the convective plume and environment times the mass flux (eq. B.23), where χ is a given field which could be moisture q, static energy $s = C_pT + gz$ or velocity v.

$$\overline{\omega\chi} = -gM^c(\chi^c - \overline{\chi}) \tag{B.23}$$

The mass flux approach assumes that there is an exchange of mass between cloud and environment by entrainment ϵ and detrainment δ. This is one of the main setbacks of the mass flux formulation, as there is not a universal formulation of entrainment rates applicable to all convective situations. They normally follow a vertical profile, although there is extensive research to produce adaptive formulations for entrainment and detrainment (Derbyshire *et al.*, 2011; Stirling and Stratton, 2012). Some of these approaches have produced substantial improvements in the representation of tropical variability (Bechtold *et al.*, 2008).

The apparent heat source of convection is given in equation B.24, where s is the static energy. Its physical interpretation is that convection affects the large-scale environment by heating through compensating subsidence, detrainment of cloud air into the environment and evaporation of cloud and precipitation.

$$Q_1 = Q_R - gM^c\frac{\partial \overline{s}}{\partial p} + \delta(s^c - \overline{s}) - LE \tag{B.24}$$

There are several pitfalls in the representation of convective events in present GCMs. As horizontal resolution of models increases, the number of convective plumes within a gridbox decreases and fluctuations on their bulk effects become larger (Craig and Cohen, 2006; Shutts and Palmer, 2007). Figure B.5 shows the mean mass flux from very a high resolution atmospheric simulation able to resolve convection (more details in Plant and Craig 2008 and Cohen and Craig 2006). For typical horizontal resolutions of climate models ($\sim 100\,km$), the Probability Density Function (PDF) of the mean mass fluxes can be approximated to a delta function, but for higher resolution there are many other likely outcomes for the "resolved" mass flux. Another problem of the convection schemes is their lack of communication between model columns. Convection can organize in large-scale phenomena such as Mesoscale Convective System (MCS), that spans spatial scales up to 500km. Present schemes have severe difficulties to represent MCS adequately. The emergence of these uncertainties have made convective parametrization the ideal Trojan horse to introduce stochasticity in atmosphere models, this issue is broadly discussed in chapter 2.

B.3.6 Gravity Wave Drag

GCMs in the early 1980 had excessively strong mid-latitude westerly (easterly) stratospheric winds in the winter (summer), making obvious that a sink of momentum to balance the meridional transport was missing (Palmer *et al.*, 1986). There are atmospheric waves, known as Gravity Waves (GW), that can deposit momentum and exert a drag on the flow. Although drag implies a deceleration, forces produced by GW dissipation can either accelerate or decelerate atmospheric winds.

GW are generated by lower atmospheric sources like flow over irregular orography such as mountains and valleys, and uneven distribution of diabating heating associated with convective sources, fronts or jet imbalances. The spatial scales of these waves are on the range of 5-500 km horizontally and therefore their effects on the main flow need to be parametrized in GCMs. Current GW parametrizations are divided between orographic GW and non-orographic GW and are extensively described and evaluated in review papers like Kim *et al.* (2003) and Alexander *et*

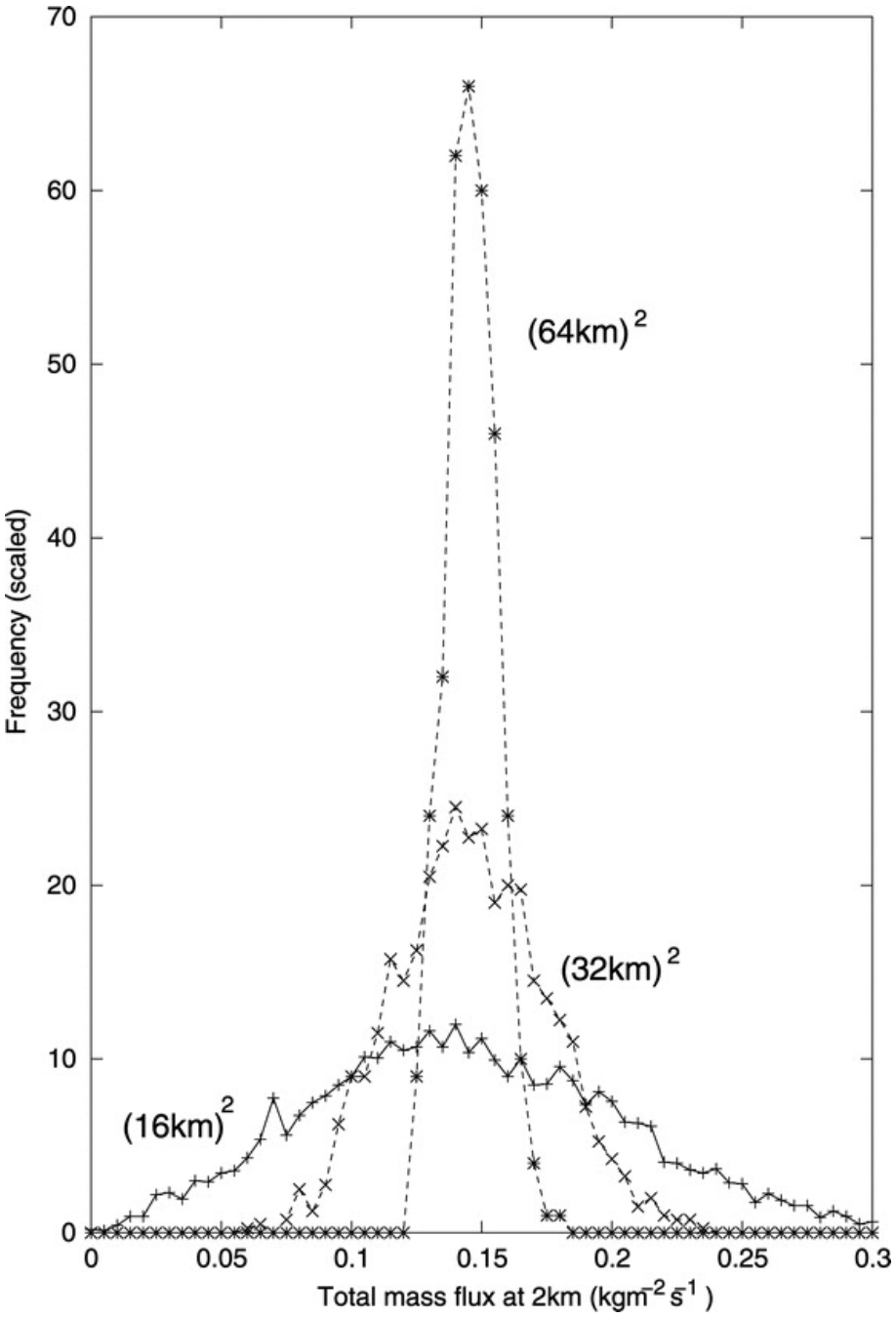

Figure B.5: Frequency plot of total convective mass flux per unit area (at a height of 2 km) obtained from CRM simulations of Cohen and Craig (2006). *From Plant and Craig (2008)*

al. (2010).

GW parametrizations are fundamental for the representation of the middle-atmosphere and thus they are widely used in state-of-the-art GCMs. They provide a function for the vertical flux of pseudo momentum flux and three dimension wave-propagation properties. However, they include a number of important simplified assumptions about the basic gravity-wave dynamics such as:

- Simulated gravity waves propagate only in the vertical up to a height where they break and deposit momentum.
- All GW parametrizations employ tunable parameters which act to scale the wave drag and/or change the breaking heights, these parameters remain difficult to quantify with observations (Alexander *et al.*, 2010).
- The representation of sources of non-orographic waves is largely ad-hoc, as they don't fully interact with the sources of gravity waves (McFarlane, 2011).

Bibliography

Abel SJ and Shipway BJ (2007): A comparison of cloud-resolving model simulations of trade wind cumulus with aircraft observations taken during RICO. *Q.J.R. Meteorol. Soc.*, **133**, 781–794.

Abel, SJ, Walters DN and Allen G (2010): Evaluation of stratocumulus cloud prediction in the Met Office forecast model during VOCALS-REx, *Atmos. Chem. Phys.*, **10**, 10541-10559.

Abel SJ and Boutle IA (2012): An improved representation of the raindrop size distribution for single-moment microphysics schemes. *Q.J.R. Meteorol. Soc.*, **138**, 2151–2162.

Adler RF and Coauthors (2003): The Version-2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979–Present). *J. Hydrometeor*, **4**, 1147–1167.

Alexander MJ, Geller M, McLandress C, Polavarapu S, Preusse P, Sassi F, Sato K, Eckermann S, Ern M, Hertzog A, Kawatani Y, Pulido M, Shaw TA, Sigmond M, Vincent R and Watanabe S (2010): Recent developments in gravity-wave effects in climate models and the global distribution of gravity-wave momentum flux from observations and models. *Q.J.R. Meteorol. Soc.*, **136**,1103–1124.

Anderson D, Hodges KI and Hoskins BJ (2003): Sensitivity of feature-based analysis methods of storm tracks to the form of background field removal. *Mon. Wea. Rev.*, **131**, 565–573.

Arakawa A, Schubert WH (1974): Interaction of a cumulus cloud ensemble with the large-scale environment. Part I. *J Atmos Sci.*, **31**,674–701.

Arakawa A (2004): The Cumulus Parameterization Problem: Past, Present, and Future. *J. Climate*, **17**, 2493–2525.

Arnold HM, Moroz IM and Palmer TN (2013): Stochastic parametrizations and model uncertainty in the Lorenz 96 system. *Phil. Trans. R. Soc. A*, **371**, 20110479.

Arribas A and Coauthors (2011): The GloSea4 Ensemble Prediction System for Seasonal Forecasting. *Mon. Wea. Rev.*, **139**, 1891–1910.

Baldwin MP and Coauthors (2001): The quasi-biennial oscillation, *Rev. Geophys.*, **39(2)**, 179–229.

Barker HW, Stephens GL and Fu Q (1999): The sensitivity of domain-averaged solar fluxes to assumptions about cloud geometry. *Q.J.R. Meteorol. Soc.*, **125**, 2127–2152.

Barker HW, Cole JNS, Morcrette JJ, Pincus R, Räisänen P, von Salzen K and Vaillancourt PA (2008): The Monte Carlo Independent Column Approximation: an assessment using several global atmospheric models. *Q.J.R. Meteorol. Soc.,* **134**,1463–1478.

Bechtold P, Chaboureau JP, Beljaars A, Betts AK, Köhler M, Miller M and Redelsperger JL (2004): The simulation of the diurnal cycle of convective precipitation over land in a global model. *Q.J.R. Meteorol. Soc.*, 130. 3119–3137.

Bechtold P, Koehler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F and Balsamo G (2008): Advances in simulating atmospheric variability with the ECMWF model: From synoptic to decadal time-scales. *Q. J. R. Meteorol. Soc.*, **134**, 1337–1351.

Bermejo R (1990): On the Equivalence of Semi-Lagrangian Schemes and Particle-in-Cell Finite Element Methods. *Mon. Wea. Rev.*, **118**, 979–987.

Bénard P, Vivoda J, Mašek J, Smolíková P, Yessad K, Smith C, Brožková R and Geleyn JF (2010): Dynamical kernel of the Aladin–NH spectral limited-area model: revised formulation and sensitivity experiments. *Q. J. R. Meteorol. Soc.*, **136**, 155–169.

Bengtsson L, Hagemann S and Hodges KI (2004): Can climate trends be calculated from reanalysis data? *J. Geophys. Res.*, **109**, D11111.

Bengtsson L, Hodges KI and Froude LSR (2005): Global observations and forecast skill. *Tellus*, **57A**, 515–527.

Bengtsson L, Körnich H, Källén E and Svensson G. (2011): Large-scale dynamical response to subgrid scale organization provided by cellular automata. J. Atmos. Sci., **68**, 3132–3144.

Bengtsson L, Steinheimer M, Bechtold P and Geleyn JF (2013): A stochastic parametrization for deep convection using cellular automata. *Q.J.R. Meteorol. Soc.*, **139**, 1533–1543.

Berner J, Doblas-Reyes FJ,Palmer TN, Shutts G and Weisheimer A (2008): Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Phil. Trans. R. Soc.*, **366**, 2559–2577.

Berner J, Shutts GJ, Leutbecher M and Palmer TN (2009): A Spectral Stochastic Kinetic Energy Backscatter Scheme and its impact on flow-dependent predictability in the ECMWF Ensemble Prediction System. *J. Atmos. Sci.*, **66**, 603-626.

Berner J, Ha S-Y, Hacker JP, Fournier A and Snyder C (2011): Model Uncertainty in a Mesoscale Ensemble Prediction System: Stochastic versus Multiphysics Representations. *Mon. Wea. Rev.*, **139**, 1972–1995.

Berner J, Jung T and Palmer TN (2012): Systematic Model Error: The Impact of Increased Horizontal Resolution versus Improved Stochastic and Deterministic Parameterizations. *J. Climate*, **25**, 4946–4962.

Best MJ and coauthors (2011): The Joint UK Land Environment Simulator (JULES), model description: Part 1: Energy and water fluxes, *Geosci. Model Dev.*, **4**, 677–699

Bjerknes V (1904): Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik, *Meteor*, **21**,1-7.

Bodas-Salcedo A, Webb MJ, Bony S, Chepfer H, Dufresne J-L, Klein SA, Zhang Y, Marchand R, Haynes JM, Pincus R and John VO (2011): COSP: satellite simulation software for model assessment. *Bull. Amer. Meteorol. Soc.*, **92**, 1023–1043.

Bodenschatz E, Malinowski SP, Shaw RA and Stratmann F (2010): Can we understand clouds without turbulence?, *Science*, **327**, 970–971.

Bouttier F, Vié B, Nuissier O, Raynaud L (2012): Impact of Stochastic Physics in a Convection-Permitting Ensemble. *Mon. Wea. Rev.*, **140**, 3706–3721.

Bowler NE, Arribas A, Mylne KR, Robertson KB, Beare SE (2008): The MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **134**, 703–722.

Bowler NE, Arribas A, Beare SE, Mylne KR and Shutts GJ (2009): The local ETKF and SKEB: Upgrades to the MOGREPS short-range ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **135**, 767–776.

Brown AR, Beare RJ, Edwards JM, Lock AP, Keogh SJ, Milton SF and Walters DN (2008): Upgrades to the boundary-layer scheme in the Met Office numerical weather prediction model. *Boundary-Layer Meteorol.*, **128**, 117–132.

Brown A, Milton S, Cullen M, Golding B, Mitchell J, Shelly A (2012): Unified Modeling and Prediction of Weather and Climate: A 25-Year Journey. *Bull. Amer. Meteor. Soc.*, **93**, 1865–1877.

Brankart JM (2013): Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modeling, *Ocean Model.*, **66**, 64–76.

Buizza R, Miller M and Palmer TN (1999): Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **125**, 2887–2908.

Buizza R, Houtekamer PL, Toth Z, Pellerin G, Wei M, and Yuejian Z (2005): A comparison of the ECMWF, MSC and NCEP global ensemble prediction systems, *Mon. Wea. Rev.*, **133**, 1076-1097

Buizza R (2008): The value of probabilistic prediction. *Atmosph. Sci. Lett.*, **9**,36–42.

Cahalan RF, Ridgway W, Wiscombe WJ, Gollmer S and Harshvardhan (1994): Independent Pixel and Monte Carlo Estimates of Stratocumulus Albedo. *J. Atmos. Sci.*, **51**, 3776–3790.

Casati B, Wilson LJ, Stephenson DB, Nurmi P, Ghelli A, Pocernich M, Damrath U, Ebert EE, Brown BG and Mason S (2008): Forecast verification: current status and future directions. *Met. Apps*, **15**, 3–18.

Catto JL, Shaffrey LC and Hodges KI (2010): Can Climate Models Capture the Structure of Extratropical Cyclones?. *J. Climate*, **23**, 1621–1635.

Charney JG (1951): Dynamic forecasting by numerical process, Compendium of Meteorology, *American Meteorological Society*, Boston, pp. 470–482

Charron M and Manzini E (2002): Gravity waves from fronts: Parameterization and middle atmosphere response in a general circulation model. *J. Atmos. Sci.*, **59**, 923–941.

Charron M, Pellerin G, Spacek L, Houtekamer P, Gagnon N, Mitchell H and Michelin L (2010): Toward random sampling of model error in the Canadian ensemble prediction system. *Mon. Wea. Rev.*, **138**, 1877–1901.

Chen M, Rood RB and Takacs LL (1997): Impact of a semi-Lagrangian and an Eulerian dynamical core on climate simulations. *J. Climate*, **10**, 2374–2389.

Cho HK, Bowman KP and North GR (2004): Equatorial waves including the Madden-Julian Oscillation in TRMM rainfall and OLR data, *J. Clim.*, **17**, 4387–4406.

Collins WJ and Coauthors (2011): Development and evaluation of an Earth-system model–HadGEM2, *Geosci. Model Dev. Discuss.*, **4**, 997–1062,

Cloke HL and Pappenberger F. (2008): Evaluating forecasts of extreme events for hydrological applications: an approach for screening unfamiliar performance measures. *Meteorological Applications*, **15**, 181–197.

Cohen BG, Craig GC (2006): Fluctuations in an Equilibrium Convective Ensemble. Part II: Numerical Experiments. *J. Atmos. Sci.*, **63**, 2005–2015.

Courant R, Friedrichs KO and Lewy H (1928): Uber die partiellien Differenzengleichungen der mathematischen Physik, *Math. Annalen*, **100**, 32–74.

Craig GC and Cohen BG (2006): Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.*, **63**, 1996–2004.

Cullen MJP and Davies T (1991): A conservative split-explicit integration scheme with fourth-order horizontal advection. *Q.J.R. Meteorol. Soc.*, **117**,993–1002.

Cullen MJP (1993): The unified forecast/climate model. *Meteor. Mag.*, **122**, 81–94.

Cusack S, Edwards JM and Crowther JM (1999): Investigating k distribution methods for parameterizing gaseous absorption in the Hadley Centre Climate Model, *J. Geophys. Res.*, **104(D2)**, 2051–2057.

D'Andrea F, Tibaldi S, Blackburn M, Boer G, Deque M, Dix MR, Dugas B, Ferranti L, Iwasaki T, Iwasaki T, Kitoh A, Pope V, Randall D, Roeckner E, Straus D, Stern W, van den Dool H and Williamson D. (1998): Northern Hemisphere atmospheric blocking as simulated by 15 atmospheric general circulation models in the period 1979–1988. *Climate Dynamics*, **14**, 385–407.

Davies T, Cullen MJP, Malcolm AJ, Mawson MH, Staniforth A, White AA and Wood N (2005): A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q.J.R. Meteorol. Soc.*, **131**,1759–1782.

Davis C, Brown B and Bullock R (2006): Object-based verification of precipitation forecasts. Part I: methodology and application to Mesoscale Rain Areas. *Mon. Wea. Rev.*, **134**, 1772–1784.

Davis C, Brown B and Bullock R. (2006): Object-based verification of precipitation forecasts. Part II: application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795.

Dawson A, Palmer TN and Corti S (2012): Simulating regime structures in weather and climate prediction models, *Geophys. Res. Lett.*, **39**, L21805.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN and Vitart F (2011): The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, **137**, 553–597.

Demory ME, Vidale PL, Roberts M, Berrisford P, Strachan J, Schiemann R an Mizielinski MS (2013): The role of horizontal resolution in simulating drivers of the global hydrological cycle, *Clim. Dynam.*, *in press.*

Derbyshire SH, Maidens AV, Milton SF, Stratton RA and Willett MR (2011): Adaptive detrainment in a convective parametrization. *Q.J.R. Meteorol. Soc.*, **137**,1856–1871.

Dessler AE, Schoeberl MR, Wang T, Davis SM and Rosenlof KH (2013): Stratospheric water vapor feedback, *Proc. Natl. Acad. Sci.*, **110**, 18,087–18,091

Dierckx P (1981): An algorithm for surface fitting with spline functions. *SIAM J. Num. Anal.*, **19**, 1286–1304.

Dierckx P (1984): Algorithms for smoothing data on the sphere with tensor product splines. *Computing*, **32**, 319–342.

Doblas-Reyes F and Coauthors (2009): Addressing model uncertainty in seasonal and annual dynamical seasonal forecasts. *Q. J. R. Meteorol. Soc.*, **135**, 1538–1559.

Düben PD, McNamara H and Palmer TN (2013): The use of imprecise processing to improve accuracy in weather and climate prediction. *Journal of Computational Physics*. **271**, 2–18,

Eckermann SD (2011): Explicitly Stochastic Parameterization of Nonorographic Gravity Wave Drag. *J. Atmos. Sci.*, **68**, 1749–1765.

Edwards JM and Slingo A (1996): Studies with a flexible new radiation code. I: Choosing a configuration for a large-scale model. *Q.J.R. Meteorol. Soc.*, **122**, 689–719.

Emanuel KA (1994): Atmospheric Convection. *Oxford University Press* , New York; 580 pp.

Epstein ES (1969): Stochastic dynamic prediction. *Tellus*, **21**, 739–759.

Evensen G (2003): The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dynamics*, **53**: 343–367.

Fereday DR, Knight JR, Scaife AA, Folland CK, Philipp A (2008): Cluster Analysis of North Atlantic-European Circulation Types and Links with Tropical Pacific Sea Surface Temperatures. *J. Climate*, **21**, 3687-3703.

Field PR, Hogan RJ, Brown PRA, Illingworth AJ, Choularton TW and Cotton RJ (2005): Parametrization of ice-particle size distributions for mid-latitude stratiform cloud. *Q.J.R. Meteorol. Soc.*, **131**, 1997–2017.

Field PR, Hill AA, Furtado K and Korolev A (2014): Mixed-phase clouds in a turbulent environment. Part 2: Analytic treatment. *Q.J.R. Meteorol. Soc.*, **140**, 870-880.

Frederiksen JS and Davies AG (1997): Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.*, **54**, 2475–2492

Frenkel Y, Majda AJ and Khouider B (2012): Using the Stochastic Multicloud Model to Improve Tropical Convective Parameterization: A Paradigm Example. *J. Atmos. Sci.*, **69**, 1080–1105.

Froude LSR, Bengtsson L and Hodges KI (2007a): The predictability of extratropical storm tracks and the sensitivity of their prediction to the observing system. *Mon. Wea. Rev.*, **135**, 315–333.

Froude LSR, Bengtsson L and Hodges KI (2007b): The prediction of extratropical storm tracks by the ECMWF and NCEP Ensemble Prediction Systems. *Mon. Wea. Rev.*, **135**, 2545–2567.

Froude LSR (2009): Regional differences in the prediction of extratropical cyclones by the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **137**, 893–911.

Froude LSR (2010): TIGGE: Comparison of the prediction of Northern Hemisphere extratropical cyclones by different ensemble prediction systems. *Wea. Forecasting*, **25**, 819–836.

Froude LSR (2011): TIGGE: Comparison of the Prediction of Southern Hemisphere Extratropical Cyclones by Different Ensemble Prediction Systems. *Wea. Forecasting*, **26**, 388–398

Fu Q and Liou KN (1992): On the Correlated k-Distribution Method for Radiative Transfer in Nonhomogeneous Atmospheres. *J. Atmos. Sci.*, **49**, 2139–2156.

Fu Q (2007): A New Parameterization of an Asymmetry Factor of Cirrus Clouds for Climate Models. *J. Atmos. Sci.*, **64**, 4140–4150.

Gardner M (1983): The game of life, parts I–III. In Wheels, life, and other mathematical amusements. New York, NY: W.H. Freeman.

Gates WL and Coauthors (1999): An overview of the results of the Atmospheric Model Intercomparison Project (AMIP I). *Bull Am Meteor Soc*, **80**,29–55.

Gibson JK, Kållberg P, Uppala SM, Nomura A, Hernandez A, Serrano E. (1997): ERA description. *ERA-15 Report Series*, No. **1**, ECMWF: Reading, UK.

Goldfarb D (1969): Extension of Davidon's variable metric method to maximization under linear inequality and equality constraints. *SIAM J. Appl. Math*, **17**, 739–764.

Gordon C, Cooper C, Senior CA, Banks H, Gregory JM, Johns TC, Mitchell JFB and Wood RA (2000): The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16**, 147–168.

Greene JS, Klatt M, Morrisey M and Postawko S (2008): The Comprehensive Pacific Rainfall Database. *J. Atmos. Oceanic Technol.*, **25**, 71–82.

Greeves CZ, Pope VD, Stratton RA and Martin GM (2006): Representation of Northern Hemisphere winter storm tracks in climate models. *Clim. Dynam.***28**, 683–702.

Gregory D and Rowntree PR (1990): A mass flux convection scheme with representation of cloud ensemble characteristics and stability-dependent closure. *Mon. Weather Rev*, **118**, 1483–1506.

Gregory D and Allen S (1991): The effect of convective scale downdrafts upon nwp and climate simulations, in Ninth Conf. Numerical Weather Prediction, *Am. Meteorol. Soc.*, Denver, Colorado, pp. 122–123.

Groenemeijer PH and Craig GC (2011): Ensemble forecasting with a stochastic convective parametrization based on equilibrium statistics. *Atmos. Chem. Phys.*, **12**, 4555–4565.

Hacker JP, Ha SY, Snyder C, Berner J, Eckel FA, Kuchera E, Pocernich M, Rugg S, Schramm J and Wang X (2011): The U.S. Air Force Weather Agency's mesoscale ensemble: scientific description and performance results. *Tellus A*, **63**, 625–641.

Hamill TM, Brennan MJ, Brown B, DeMaria M, Rappaport ED and Toth Z (2012): NOAA's Future Ensemble-Based Hurricane Forecast Products. *Bull. Amer. Meteor. Soc.*, **93**, 209–220.

Hasselmann K (1976): Stochastic climate models Part I. Theory. *Tellus*, **28**, 473–485.

Heintzenberg and Charlson RJ (2009): Clouds in the Perturbed Climate System, *MIT Press*, Cambridge.608 pp.

Hermanson L, Hoskins B and Palmer TN (2009): A comparative method to evaluate and validate stochastic parametrizations. *Q.J.R. Meteorol. Soc.*, **135**, 1095–1103.

Hersbach H (2000): Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems, *Wea. Forecasting*, **15**, 559–570.

Hill PG, Manners J and Petch JC (2011): Reducing noise associated with the Monte Carlo Independent Column Approximation for weather forecasting models. *Q.J.R. Meteorol. Soc.*, **137**, 219–228.

Hill AA, Field PR, Furtado K, Korolev A and Shipway BJ (2014): Mixed-phase clouds in a turbulent environment. Part 1: Large-eddy simulation experiments. *Q.J.R. Meteorol. Soc.*, **140**,855-869.

Hinkelmann K (1951): Der Mechanismus des meteorologischen Larmes, *Tellus*, **3**, 285–296.

Hodges KI (1994): A general method for tracking analysis and its application to meteorological data. *Mon. Wea. Rev.*, **122**, 2573–2586.

Hodges KI (1995): Feature tracking on the unit sphere. *Mon. Wea. Rev.*, **123**, 3458–3465.

Hodges KI (1996): Spherical nonparametric estimators applied to the UGAMP model integration for AMIP. *Mon. Wea. Rev.*, **124**, 2914–2932.

Hodges KI and Thorncroft CD (1997): Distribution and statistics of African mesoscale convective weather systems based on the ISCCP Meteosat imagery. *Mon. Wea. Rev.*, **125**, 2821–2837.

Hodges KI (1998): Feature-point detection using distance transforms: Application to tracking tropical convective complexes. *Mon. Wea. Rev.*, **126**, 785–795.

Hodges KI (1999a): Adaptive constraints for feature tracking. *Mon. Wea. Rev.*, **127**, 1362–1373

Hodges KI (1999b): Extension of spherical nonparametric estimators to nonisotropic kernels: An oceanographic application. *Mon. Wea. Rev.*, **127**, 214–227

Hodges KI, Hoskins BJ, Boyle J and Thorncroft C (2003): A Comparison of Recent Reanalysis Datasets Using Objective Feature Tracking: Storm Tracks and Tropical Easterly Waves. *Mon. Wea. Rev.*, **131**, 2012–2037.

Holloway CE, Woolnough SJ, Lister GMS (2013): The Effects of Explicit versus Parameterized Convection on the MJO in a Large-Domain High-Resolution Tropical Case Study. Part I: Characterization of Large-Scale Organization and Propagation. *J. Atmos. Sci.*, **70**, 1342–1369.

Holton JR (1972): An Introduction to Dynamic Meteorology. *Academic Press*, New York, 319 pp.

Hoskins BJ and Sardeshmukh PD (1987): A Diagnostic Study of the Dynamics of the Northern Hemisphere Winter of 1985-86. *Q.J.R. Meteorol. Soc.*, **113**, 759–778.

Hoskins BJ and Yang GY (2000): The equatorial response to higher-latitude forcing. *J. Atmos. Sci.*, **57**, 1197–1213.

Hoskins BJ and Hodges KI (2002): New perspectives on the Northern Hemisphere winter storm tracks. *J. Atmos. Sci.*, **59**, 1041–1061.

Hoskins BJ, Hodges KI (2005): A New Perspective on Southern Hemisphere Storm Tracks. *J. Climate*, **18**, 4108–4129.

Hoskins B (2013): The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q.J.R. Meteorol. Soc.*, 139, 573–584.

Huang XY (1990): The organization of moist convection by internal gravity waves. *Tellus A*, **42**: 270–285.

Huffman GJ and Coauthors (1997): The Global Precipitation Climatology Project (GPCP) combined precipitation datasets. *Bull. Amer. Meteor. Soc.*, **78**, 5–20.

Jolliffe IT and Stephenson DB (2003): Forecast Verification: A Practitioner's Guide in Atmospheric Science. *John Wiley and Sons*, Chichester, UK; 292pp.

Jung T, Palmer TN and Shutts GJ (2005): Influence of a stochastic parameterization on the frequency of occurrence of North Pacific weather regimes in the ECMWF model, *Geophys. Res. Lett.*, **32**, L23811.

Jung T (2005b): Systematic errors of the atmospheric circulation in the ECMWF forecasting system. *Q.J.R. Meteorol. Soc.*, **131**, 1045–1073.

Jung T, Miller MJ and Palmer TN (2010a): Diagnosing the Origin of Extended-Range Forecast Errors. *Mon. Wea. Rev.*, **138**, 2434-2446.

Jung T, Balsamo G, Bechtold P, Beljaars ACM, Köhler M, Miller MJ, Morcrette JJ, Orr A, Rodwell MJ and Tompkins AM (2010b): The ECMWF model climate: recent progress through improved physical parametrizations. *Q.J.R. Meteorol. Soc.*, **136**, 1145–1160.

Kain JS and Fritsch JM (1990): A one-dimensional entraining/detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802.

Kalnay E, Kanamitsu M, Kirtler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetma A, Reynolds R, Jenne R, Joseph D. (1996): The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteorol. Soc.*, **77**, 437–471.

Kalnay E (2002): Atmospheric Modeling, Data Assimilation, and Predictability. *Cambridge University Press,* 342 pp.

Kasahara A, Washington WM (1967): NCAR global general circulation model of the atmosphere, *Mon. Weather Rev.*, **95**, 389–402.

Keane RJ and Plant RS (2012): Large-scale length and time-scales for use with stochastic convective parametrization. *Q.J.R. Meteorol. Soc.*, **138**,150–1164.

Khouider B, Biello J and Majda A (2010): A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216

Khvorostyanov, VI and Curry JA (2005): Fall Velocities of Hydrometeors in the Atmosphere: Refinements to a Continuous Analytical Power Law. *J. Atmos. Sci.*, **62**, 4343–4357.

Kiladis GN, Thorncroft CD, Hall MJ (2006): Three-Dimensional Structure and Dynamics of African Easterly Waves. Part I: Observations. *J. Atmos. Sci.*, **63**, 2212–2230.

Kiladis GN, Wheeler MC, Haertel PT, Straub KH and Roundy PE (2009): Convectively coupled equatorial waves,*Rev. Geophys.*, **47**, RG2003.

Kim YJ, Eckermann SD and Chun HY (2003): An overview of the past, present, and future of gravity-wave drag parameterization for numerical climate and weather prediction models. *Atmos.–Ocean*, **41**, 65–98

Kim D and Coauthors (2009): Application of MJO Simulation Diagnostics to Climate Models. *J. Climate*, **22**, 6413–6436.

Klingaman NP and Woolnough SJ (2013): Using a case-study approach to improve the Madden-Julian oscillation in the Hadley Centre model. *Q. J. R. Meteorol. Soc.*, in press

Lang STK, Leutbecher M and Jones SC (2012): Impact of perturbation methods in the ECMWF ensemble prediction system on tropical cyclone forecasts. *Q.J.R. Meteorol. Soc.*, **138**, 2030–2046.

Lau KM and Chan PH (1988a): Intraseasonal and interannual variations of tropical convection: A possible link between the 40–50 day oscillation and ENSO? *J. Atmos. Sci.*, **45**, 506–521.

Lau NC (1988b): Variability of the observed midlatitude storm tracks in relation to low-frequency changes in the circulation pattern. *J. Atmos. Sci.*, **45**, 2718–274.

Lee SK, Mechoso CR, Wang C, Neelin JD (2013): Interhemispheric Influence of the Northern Summer Monsoons on Southern Subtropical Anticyclones. *J. Climate*, **26**, 10193-10204.

Leith CE (1974): Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Li X, Charron M, Spacek L and Candille G (2008): A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443–462.

Liebmann B, Hendon H and Glick J (1994): The relationship between tropical cyclones of the western Pacific and Indian oceans and the Madden-Julian Oscillation, *J. Meteorol. Soc. Jpn.*, **72**, 401–411.

Liebmann B and Smith CA (1996): Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.

Lilly DK (1983): Stratified turbulence and the mesoscale variability of the atmosphere. *J. Atmos. Sci.*, **40**, 749–761.

Lilly DK (1989): Two-Dimensional Turbulence Generated by Energy Sources at Two Scales. *J. Atmos. Sci.*, **46**, 2026–2030.

Lin JL, Weickman KM, Kiladis GN, Mapes BE, Schubert SD, Suarez MJ, Bacmeister JT and Lee MI (2008): Subseasonal Variability Associated with Asian Summer Monsoon Simulated by 14 IPCC AR4 Coupled GCMs. *J. Climate*, **21**, 4541-4567.

Lin JWB and JD Neelin (2000): Influence of a stochastic moist convective parameterization on tropical climate variability, *Geophys. Res. Lett.*, **27**, 3691-3694.

Lin JWB, Neelin JD (2002): Considerations for Stochastic Convective Parameterization. *J. Atmos. Sci.*, **59**, 959–975.

Lin JWB and JD Neelin (2003): Toward stochastic deep convective parameterization in general circulation models, *Geophys. Res. Lett.*, **30**, 1162.

Lin, JL and Coauthors (2006): Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals, *J. Clim.*, **19**, 2665–2690,

Lock AP, Brown AR, Bush MR, Martin GM, Smith RNB (2000): A New Boundary Layer Mixing Scheme. Part I: Scheme Description and Single-Column Model Tests. *Mon. Wea. Rev.*, **128**, 3187–3199.

Loeb NG, Priestley KJ, Kratz DP, Geier EB, Green RN, Wielicki BA, Hinton POR and Nolan SK (2001): Determination of unfiltered radiances from the Clouds and

the Earth's Radiant Energy System (CERES) instrument. *J. Appl. Meteor.*, **40**, 822–835.

Loeb NG, Kato S, Loukachine K, Smith NM and Doelling DR (2007): Angular distribution models for top-of-atmosphere radiative flux estimation from the Clouds and the Earth's Radiant Energy System instrument on the Terra satellite. Part II: Validation. *J. Atmos. Oceanic Technol.*, **24**, 564–584.

Lorenz EN (1963): Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

Lorenz EN (1969): The predictability of a flow which possesses many scales of motion, *Tellus*, **21**, 289–307.

Lott F, Guez L Maury P (2012): A stochastic parameterization of non-orographic gravity waves: Formalism and impact on the equatorial stratosphere, *Geophys. Res. Lett.*, **39**, L06807.

Louis JF (1979): A parametric model of vertical eddy fluxes in the atmosphere. *Bound.-Layer Meteor.*, **17**, 187–202.

Lynch P (2006): The Emergence of Numerical Weather Prediction: Richardson's Dream, *Cambridge University Press*, Cambridge, 2006, 279pp.

Lynch P (2007): The origins of computer weather prediction and climate modeling, *J. Comput. Phys.*, **227**,3431–3444.

MacLachlan C, Arribas A, Peterson D, Maidens A, Fereday D, Scaife AA, Gordon M, Vellinga M, Williams A, Comer RE, Camp J and Xavier P (2013): GloSea5: the Met Office high resolution seasonal forecast system. *Q. J. R. Met. Soc.*, doi: 10.1002/qj.2396

Madden RA and Julian PR (1971): Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific, *J. Atmos. Sci.*, **28**, 702–708.

Madden RA and Julian PR (1972): Description of global-scale circulation cells in the tropics with a 40–50 day period, *J. Atmos. Sci.*, **29**, 1109–1123.

Manabe S, Smagorinsky J and Strickler RF (1965): Simulated Climatology of a general circulation model with a hydrologic cycle. *Mon. Wea. Rev.*, **93**, 769–798.

Marshall JS and Palmer W (1948): The distribution of raindrops with size. *J. Meteorol.*, **5**, 165–166.

Martin GM, Ringer MA, Pope VD, Jones A, Dearden C and Hinton TJ (2006): The Physical Properties of the Atmosphere in the New Hadley Centre Global Environmental Model (HadGEM1). Part I: Model Description and Global Climatology. *J. Climate*, **19**, 1274–130

Mason PJ and Thomson DJ (1992): Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.*, **242**, 51–78.

Matsueda M, Mizuta R and Kusunoki S (2009): Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model. *J. Geophys. Res.*, **114**, D12114.

Matsuno T (1966): Quasi-geostrophic motions in the equatorial area, *J. Meteorol. Soc. Jpn.*, **44**, 25–43.

Matthews AJ (2000): Propagation mechanisms for the Madden-Julian Oscillation, *Q. J. R. Meteorol. Soc.*, **126**, 2637–2651.

May W (2008): Potential future changes in the characteristics of daily precipitation in Europe simulated by the HIRHAM regional climate model, *Climate Dyn.*, **30**, 581–603.

McFarlane N. (2011), Parameterizations: representing key processes in climate models without resolving them. *WIREs Clim Change*, **2**, 482–497

McCalpin JD (1988): A Quantitative Analysis of the Dissipation Inherent in Semi-Lagrangian Advection, *Mon. Wea. Rev.*, **116**, 2330-2336

Meehl GA, Moss R, Taylor KE, Eyring V, Stouffer RJ, Bony S and Stevens B (2014): Climate Model Intercomparisons: Preparing for the Next Phase, *Eos Trans. AGU*, **95(9)**, 77-84.

Milton SF, Greed G, Brooks ME, Haywood J, Johnson B, Allan RP, Slingo A and Grey WMF (2008): Modeled and observed atmospheric radiation balance during the West African dry season: Role of mineral dust, biomass burning aerosol, and surface albedo. *J. Geophys. Res.*, **113**, D00C02.

Mittermaier MP and Bullock R (2013): Using MODE to explore the spatial and temporal characteristics of cloud cover forecasts from high-resolution NWP models. *Met. Apps*, **20**, 187–196.

Monin AS and Obukhov AM (1954): Basic laws of turbulent mixing in the surface layer of the atmosphere (in Russian), *Contrib. Geophys. Inst. Acad. Sci. USSR*, **151**, 163–187.

Morrison H, Boer G, Feingold G, Harrington J, Shupe MD and Sulia K (2012): Resilience of persistent Arctic mixed-phase clouds. *Nat. Geosci.*, **5**, 11-17.

Mulcahy J, Walters DN, Bellouin N and Milton SF (2014): Impacts of increasing the aerosol complexity in the Met Office global NWP model, *Atmos. Chem. Phys*, **14**, 4749-4778.

Murphy JM, Sexton DH, Barnett DN, Jones GS, Webb MJ, Collins M and Stainforth DA (2004): Quantification of modelling uncertainties in a large ensemble of climate change simulations *Nature*, **430**: 768-772

Murray RJ and Simmonds I (1991): A numerical scheme for tracking cyclone centers from digital data. Part II: Application to January and July GCM simulations. *Aust. Meteor. Mag.*, **39**, 167–180.

Nakazawa T (1988): Tropical super clusters within intraseasonal variations over the western Pacific, *J. Meteorol. Soc. Jpn.*, **66**, 823–836.

Nastrom GD and Gage KS (1985): A climatology of atmosphere wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, **43**, 857–870.

Neal RA, Boyle P, Grahame N, Mylne K and Sharpe M (2013): Ensemble based first guess support towards a risk-based severe weather warning service. *Met. Apps*, **21**, 563–577.

Nijssen B, O'Donnell GM, Lettenmaier DP, Lohmann D and Wood EF (2001): Predicting the discharge of global rivers. *J. Climate*, **14**, 3307–3323.

Overgaard J, Rosbjerg D and Butts MB (2006): Land-surface modelling in hydrological perspective: a review. *Biogeosciences*, **3(2)**, 229-241.

Palmer TN, Shutts GJ and Swinbank R (1986): Alleviation of a systematic westerly bias in general circulation and numerical weather prediction models through an orographic gravity wave drag parametrization. *Q.J.R. Meteorol. Soc.*, **112**, 1001–1039.

Palmer TN (2001): A non-linear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q.J.R. Meteorol. Soc.*, **127** 279–304.

Palmer TN, Shutts G, Hagedorn R, Doblas-Reyes FJ, Jung T and Leutbecher M (2005): Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163–193.

Palmer TN, Doblas-Reyes FJ, Weisheimer A and Rodwell MJ (2008): Toward Seamless Prediction: Calibration of Climate Change Projections Using Seasonal Forecasts. *Bull. Amer. Meteor. Soc.*, **89**, 459–470.

Palmer TN, Buizza R, Doblas-Reyes FJ, Jung T, Leutbecher M, Shutts G, Steinheimer M and Weisheimer A (2009): Stochastic parametrization and model uncertainty. *Tech. Rep. ECMWF RD Tech. Memo.*, **598**, 42 pp.

Palmer TN (2012): Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Q.J.R. Meteorol. Soc.*, **138**. 841–861.

Penland C (2003): Noise out of chaos and why it won't go away. *Bull. Amer. Meteor. Soc.*, **84**, 921–925

Pelly JL, Hoskins BJ (2003): A New Perspective on Blocking. *J. Atmos. Sci.*, **60**, 743–755.

Peng MS, Ridout JA and Hogan TF (2004): Recent modifications of the Emanuel convective scheme in the Navy Operational Global Atmospheric Prediction System. *Mon. Wea. Rev.*, **132**, 1254–1268.

Peng J, Zhang L, Luo Y and Zhang Y (2014): Mesoscale Energy Spectra of the Mei-Yu Front System. Part I: Kinetic Energy Spectra. *J. Atmos. Sci.*, **71**, 37–55.

Petch J, Waliser D, Jiang X, Xavier PK and Woolnough S (2011): A global model intercomparison of the physical processes associated with the Madden–Julian os-

cillation. GEWEX News, International GEWEX Project Office, *Silver Spring* MD, Vol. 21 (3), 3–5

Peters K, Jakob C, Davies L, Khouider B and Majda AJ (2013): Stochastic Behavior of Tropical Convection in Observations and a Multicloud Model *J. Atmos. Sci.*, **70**, 3556–3575.

Pfeifroth U, Mueller R, Ahrens B (2013): Evaluation of Satellite-Based and Reanalysis Precipitation Data in the Tropical Pacific. *J. Appl. Meteor. Climatol.*, **52**, 634–644.

Piani C, Norton WA and Stainforth DA (2004): Equatorial stratospheric response to variations in deterministic and stochastic gravity wave parameterizations, *J. Geophys. Res.*, **109**, D14101

Pincus R, Barker HW and Morcrette JJ (2003): A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields, *J. Geophys. Res.*, **108**, 4376.

Pincus R, Platnick S, Ackerman SA, Hemler RS and Hofmann RJP (2012): Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators. *J. Climate*, **25**, 4699-4720.

Plant RS and Craig GC (2008): A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, **65**, 87–105.

Plant RS (2010): A review of the theoretical basis for bulk mass flux convective parametrization. *Atmos. Chem. Phys.*, **10**, 3529–3544.

Pope VD and Stratton RA (2002): The processes governing horizontal resolution sensitivity in a climate model. *Climate Dyn.*, **19**, 211–236.

Pruppacher HR and Klett JD (2010): Microphysics of Clouds and Precipitation, Atmospheric and oceanographic sciences library, *Kluwer Academic Publishers*, Cambridge, MA, 954 pp.

Räisänen P, Barker HW, Khairoutdinov MF, Li J and Randall DA (2004): Stochastic generation of subgrid-scale cloudy columns for large-scale models. *Q.J.R. Meteorol. Soc.*, **130**, 2047–2067.

Reichler T and Kim J (2008): How Well Do Coupled Models Simulate Today's Climate?. *Bull. Amer. Meteor. Soc.*, **89**, 303–311.

Reynolds CA, Teixeira J and McLay JG (2008): Impact of Stochastic Convection on the Ensemble Transform. *Mon. Wea. Rev.*, **136**, 4517-4526.

Reynolds CA, McLay JG, Goerss JS, Serra EA, Hodyss D and Sampson CR (2011): Impact of Resolution and Design on the U.S. Navy Global Ensemble Performance in the Tropics. *Mon. Wea. Rev.*, **139**, 2145–2155.

Richardson LF (1922): Weather Prediction by Numerical Process, *Cambridge University Press*, Cambridge, xii+236pp.

Rienecker MM and Coauthors (2008): The GEOS-5 Data Assimilation System—Documentation of versions 5.0.1 and 5.1.0, and 5.2.0. NASA Tech. Rep. Series on Global Modeling and Data Assimilation, NASA/TM-2008-104606, Vol. 27, 92 pp.

Rienecker MM and Coauthors (2011): MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *J. Climate*, **24**, 3624–3648.

Riihimaki LD, McFarlane SA, Comstock JM (2012): Climatology and Formation of Tropical Midlevel Clouds at the Darwin ARM Site. *J. Climate*, **25**, 6835-6850.

Ritchie H, Temperton C, Simmons A, Hortal M, Davies T, Dent D and Hamrud M (1995): Implementation of the semi-Lagrangian method in a high resolution version of the ECMWF forecast model. *Mon. Weather Rev.*, **123**, 489–514.

Rodwell MJ, Magnusson L, Bauer P, Bechtold P, Bonavita M, Cardinali C, Diamantakis M, Earnshaw P, Garcia-Mendez A, Isaksen L, Källén E, Klocke D, Lopez P, McNally AP, Persson A, Prates F and Wedi N (2013): Characteristics of Occasional Poor Medium-Range Weather Forecasts for Europe. *Bull. Amer. Meteor. Soc.*, **94**, 1393–1405.

Roeckner E and Coauthors (2003): The atmospheric general circulation model ECHAM5: Part I. Model description, Tech. Rep. 349, *Max Planck Inst. for Meteorol.*, Hamburg, Germany.

Scaife AA, Butchart N, Warner CD, Swinbank R (2002) Impact of a Spectral Gravity Wave Parameterization on the Stratosphere in the Met Office Unified Model. *J. Atmos. Sci.*, **59**, 1473–1489.

Scaife AA, Woollings T, Knight JR, Martin G and Hinton T (2010): Atmospheric blocking and mean biases in climate models, *J. Clim.*, **23**, 6143–6152,

Salari V and Sethi KI (1990): Feature point correspondence in the presence of occlusion. *IEEE Trans. PAMI*, **12**, 87–91.

Sanchez C, Williams KD, Shutts GJ, McDonald RE, Hinton TJ, Senior CA and Wood N (2013): Towards the development of a robust model hierarchy: investigation of dynamical limitations at low resolution and possible solutions. *Q.J.R. Meteorol. Soc.*, **139**, 75–84.

Sanchez C, Williams KD, Shutts G and Collins M (2014): Impact of a Stochastic Kinetic Energy Backscatter scheme across time-scales and resolutions. *Q.J.R. Meteorol. Soc.*, doi:10.1002/qj.2328.

Sardeshmukh PD and Hoskins BJ (1984): Spatial smoothing on the sphere. *Mon. Wea. Rev.*, **112**, 2524–2529

Satoh M, Matsuno T, Tomita H, Miura H, Nasuno T and Iga S (2008): Nonhydrostatic Icosahedral Atmospheric Model (NICAM) for global cloud resolving simulations. *Journal of Computational Physics*, the special issue on Predicting Weather, Climate and Extreme events,**227**, 3486-3514,

Schattler U, Doms G and Schraff C (2011): A description of the non- hydrostatic regional COSMO-model, part VII: user's guide, http: //www.cosmo-model.org

Scherrer SC, Croci-Maspoli M, Schwierz C and Appenzeller C (2006): Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region. *Int. J. Climatol.*, 26, 233-249.

Seifert A and Beheng KD (2006): A two-moment cloud microphysics parameterization for mixed-phase clouds. Part I: Model description. *Meteorol. Atmos. Phys.*, **92**, 45–66.

Seiffert R and Von Storch JS (2008): Impact of atmospheric small-scale fluctuations on climate sensitivity, *Geophys. Res. Lett.*, **35**, L10704,

Seity Y, Brousseau P, Malardel S, Hello G, Bénard P, Bouttier F, Lac C and Masson V (2011): The AROME-France convective-scale operational model. *Mon. Wea. Rev.*, **139**, 976–991.

Senior CA, Arribas A, Brown AR, Cullen MJP, Johns TC, Martin GM, Milton SF, Webster S and Williams KD (2011):'Synergies between numerical weather prediction and general circulation climate models.' In The Development of Atmospheric General Circulation Models: Complexity, Synthesis and Computation, Donner L, Schubert W and Sommerville R (eds). *Cambridge University Press*, Cambridge, UK.

Sethi IK and Jain R (1987): Finding trajectories of feature points in a monocular image sequence. *IEEE Trans. PAMI*, **9**, 56–73.

Shaffrey LC and Coauthors (2009): U.K. HiGEM: The New U.K. High-Resolution Global Environment Model—Model Description and Basic Evaluation. *J. Climate*, **22**, 1861–1896.

Shupe MD, Intrieri JM (2004): Cloud Radiative Forcing of the Arctic Surface: The Influence of Cloud Properties, Surface Albedo, and Solar Zenith Angle. *J. Climate*, **17**, 616-628.

Shutts GJ (1986): A case study of eddy forcing during an Atlantic blocking episode. *Advances in Geophysics*, **29**, Academic Press, 135–162.

Shutts GJ (2005): A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. R. Meteorol. Soc.*, **131**, 3079-3102.

Shutts GJ and Palmer TN (2007): Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, **20**, 187–202.

Shutts GJ (2008a): The forcing of large-scale waves in an explicit simulation of deep tropical convection. *Dyn. Atmos. Oceans*, **45**, 1–25.

Shutts G, Allen T and Berner J (2008b): Stochastic parametrization of multiscale processes using a dual-grid approach. *Phil Trans R Soc A*, **366**,2625–2641.

Shutts GJ (2013): Coarse Graining the Vorticity Equation in the ECMWF Integrated Forecasting System: The Search for Kinetic Energy Backscatter. *J. Atmos. Sci.*, **70**, 1233–1241.

Simmons AJ, Hoskins BJ and Burridge DM (1978): Stability of the Semi-Implicit Method of Time Integration. *Mon. Wea. Rev.*, **106**, 405–412.

Simmons AJ and Hollingsworth A (2002): Some aspects of the improvement in skill of numerical weather prediction. *Q.J.R. Meteorol. Soc.*, **128**,647–677.

Slingo JM (1996): Intraseasonal oscillations in 15 atmospheric general circulation models: Results from an AMIP diagnostic subproject, *Clim. Dyn.*, **12**, 325–357.

Slingo J, Palmer TN (2011): Uncertainty in weather and climate prediction. *Phil. Trans. R. Soc. London A.*, **369**, 4751–4767.

Smagorinsky J (1963): General circulation experiments with the primitive equations. I: The basic experiment. *Mon. Wea. Rev.*, **91**, 99-164

Smith RNB (1990): A scheme for predicting layer clouds and their water contents in a general circulation model. *Q. J. R. Meteorol. Soc*, **116**, 435–460.

Smith DM, Cusack S, Colman AW, Folland CK, Harris GT and Murphy JM (2007): Improved surface temperature prediction for the coming decade from a global climate model, *Science*, **317**, 796-799.

Snyder A, Pu Z and Reynolds CA (2011): Impact of Stochastic Convection on Ensemble Forecasts of Tropical Cyclone Development. *Mon. Wea. Rev.*, **139**, 620–626

Solomon S, Rosenlof KH, Portmann R, Daniel J, Davis S, Sanford T and Plattner GK (2010): Contributions of stratospheric water vapor to decadal changes in the rate of global warming, *Science*, **327**, 1219–1223

Sommeria G, Deardorff JW (1977): Subgrid-Scale Condensation in Models of Nonprecipitating Clouds. *J. Atmos. Sci.*, **34**, 344–355.

Staniforth A and Cote J (1991): Semi-Lagrangian Integration Schemes for Atmospheric Models—A Review. *Mon. Wea. Rev.*, **119**, 2206–2223.

Stensrud DJ (2007): Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models. *Cambridge University Press*, 488 pp.

Stephens GL, L'Ecuyer T, Forbes R, Gettlemen A, Golaz J-C, Bodas-Salcedo A, Suzuki K, Gabriel P and Haynes J (2010): Dreary state of precipitation in global models. *J. Geophys. Res.*, **115**, D24211.

Stirling AJ and Stratton RA (2012): Entrainment processes in the diurnal cycle of deep convection over land. *Q.J.R. Meteorol. Soc.*, **138**, 1135–1149.

Stratton RA (2004): The energetics of a semi-Lagrangian dynamical core. *Hadley Centre Technical Note*, **52**.
http://www.metoffice.gov.uk/archive/hadley-centre-technical-note-52

Straub KH and Kiladis GN (2002): Observations of a convectively coupled Kelvin wave in the eastern Pacific ITCZ, *J. Atmos. Sci.*, **59**, 30–53.

Straub KH and Kiladis GN (2003): Extratropical forcing of convectively coupled Kelvin waves during austral winter, *J. Atmos. Sci.*, **60**, 526–543.

Stull RB (1988): An Introduction to Boundary Layer Meteorology, *Kluwer Academic Publishers*, xiii + 666 pp. Dordrecht, Boston, London.

Taylor KE (1986): An Analysis of the Biases in Traditional Cyclone Frequency Maps. *Mon. Wea. Rev.*, **114**, 1481–1490.

Teixeira J and Reynolds C (2008a): Stochastic nature of physical parameterizations in ensemble prediction: a stochastic convection approach. *Mon. Weather Rev.*, **136**, 483–496.

Teixeira J, Stevens B, Bretherton CS, Cederwall R, Doyle JD, Golaz JC, Holtslag AAM, Klein SA, Lundquist JK, Randall DA, Siebesma AP and Soares PMM (2008b): Parameterization of the Atmospheric Boundary Layer: A View from Just Above the Inversion. *Bull. Amer. Meteor. Soc.*, **89**, 453–458.

Tennant WJ, Shutts GJ, Arribas A and Thompson SA (2010): Using a Stochastic Kinetic Energy Backscatter Scheme to Improve MOGREPS Probabilistic Forecast Skill. *Mon. Wea. Rev.*, **139**, 1190-1206.

Terasaki K, Tanaka HL and M Satoh (2009): Characteristics of the Kinetic Energy Spectrum of NICAM. *SOLA*, **5**, 180-183.

Thépaut JN, Courtier P, Belaud G and Lemaître G (1996): Dynamical structure functions in a four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.*, **122**, 535–561.

Thorncroft C, Hodges K (2001): African Easterly Wave Variability and Its Relationship to Atlantic Tropical Cyclone Activity. *J. Climate*, **14**, 1166–1179

Thuburn J, Kent J and Wood N (2013): Cascades, backscatter and conservation in numerical models of two-dimensional turbulence. *Q.J.R. Meteorol. Soc.*, **140**, 626–638.

Tibaldi S and Molteni F (1990): On the operational predictability of blocking. *Tellus*, **42A**, 343–365.

Tiedtke M (1993) Representation of clouds in large-scale models. *Mon. Weather Rev.*, **121**, 3040–3061.

Tiedtke M (1989): A comprehensive mass-flux scheme for cumulus parameterization in large-scale models. *Mon. Weather Rev.*, **117**, 1779–1800.

Tompkins AM and Berner J (2008): A stochastic convective approach to account for model uncertainty due to unresolved humidity variability, *J. Geophys. Res.*, **113**, D18101.

Trenberth KE, Olson JG (1988): An Evaluation and Intercomparison of Global Analyses from the National Meteorological Center and the European Centre for Medium Range Weather Forecasts. *Bull. Amer. Meteor. Soc.*, **69**, 1047–1057.

Trenberth KE (1992): Climate System Modeling, *Cambridge Univ. Press*, 788 pp.

Tribbia JJ and Baumhefner DP (2004): Scale Interactions and Atmospheric Predictability: An Updated Perspective. *Mon. Wea. Rev.*, **132**, 703–713.

Tung KK and Orlando WW (2003) The k^{-3} and $k^{-5/3}$ energy spectrum of atmospheric turbulence: Quasi-geostrophic two-level model simulation. *J. Atmos. Sci.*, **60**, 824–835

Uppala SM, Kållberg PW, Simmons AJ, Andrae U, Da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, Van De Berg

L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Hólm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf JF, Morcrette J-J, Rayner NA, Saunders RW, Simon P, Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P and Woollen J (2005): The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012.

Waliser D and Co-authors (2009): Cloud ice: A climate model challenge with signs and expectations of progress, *J. Geophys. Res.*, **114**, D00A21

Waliser D and Coauthors (2009): MJO simulation diagnostics. *J. Climate*, **22**, 3006–3030

Walters DN and Coauthors (2011): The Met Office Unified Model Global Atmosphere 3.0/3.1 and JULES Global Land 3.0/3.1 configurations. *Geosci. Model Dev.*, **4**, 919–941.

Webster S, Brown AR, Cameron DR and Jones CP (2003): Improvements to the representation of orography in the Met Office Unified Model. *Q. J. R. Meteorol. Soc*, **129**, 1989–2010.

Weickmann KM and Berry EB (2007): A synoptic–dynamic model of subseasonal atmospheric variability. *Mon. Wea. Rev.*, **135**, 449–474.

Weisheimer A, Palmer TN and Doblas-Reyes FJ (2011): Assessment of representations of model uncertainty in monthly and seasonal forecast ensembles, *Geophys. Res. Lett.*, **38**, L16703.

Wheeler M and Kiladis GN (1999): Convectively coupled equatorial waves: Analysis of clouds and temperature in the wavenumber-frequency domain, *J. Atmos. Sci.*, **56**, 374–399.

Wheeler M and Hendon HH (2004): An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction, *Mon. Weather Rev.*, **132**, 1917–1932.

Wielicki BA, Barkstrom BR, Harrison EF, Lee RB, Smith GL, Cooper JE (1996): Clouds and the Earth's Radiant Energy System (CERES): An Earth Observing System Experiment. *Bull. Amer. Meteor. Soc.*, **77**, 853–868.

Wilks DS (2006): Statistical methods in the atmospheric sciences (second edition), *Academic Press*, 627pp

Wilks DS (2008): Effects of stochastic parametrization on conceptual climate models. *Philos. Trans. Roy. Soc.*, **366A**, 2475–2488.

Williams PD (2005): Modelling climate change: The role of unresolved processes, *Philos. Trans. R. Soc. A.*, **363**, 2931–2946.

Williams PD (2012): Climatic impacts of stochastic fluctuations in air–sea fluxes. *Geophys. Res.Lett.*, **39**, L10705.

Williams KD, Bodas-Salcedo A, Deque M, Fermepin S, Medeiros B, Watanabe M, Jakob C, Klein SA, Senior CA and Williamson DL (2013): The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models, *J. Clim.*, **26**, 3258–3274.

Wilson DR and Ballard SP (1999): A microphysically based precipitation scheme for the Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc*, **125**, 1607–1636.

Wilson DR, Bushell AC, Kerr-Munslow AM, Price JD and Morcrette CJ (2008): PC2: A prognostic cloud fraction and condensation scheme. I: Scheme description. *Q. J. R. Meteorol. Soc.*, **134**, 2093–2107.

Wood N, Staniforth A, White A, Allen T, Diamantakis M, Gross M, Melvin T, Smith C, Vosper S, Zerroukat M and Thuburn J (2013): An inherently mass-conserving semi-implicit semi-Lagrangian discretization of the deep-atmosphere global non-hydrostatic equations. *Q.J.R. Meteorol. Soc.*. **140**, 1505–1520.

Wu WS, Purser EJ and Parrish DF (2002): Three-dimensional variational analysis with spatially inhomogeneous covariances. *Mon. Wea. Rev.*, **130**, 2905–2916.

Xavier PK (2012): Intraseasonal Convective Moistening in CMIP3 Models. *J. Climate*, **25**, 2569–2577.

Xu Q (1992): Formation and Evolution of Frontal Rainbands and Geostrophic Potential Vorticity Anomalies. *J. Atmos. Sci.*, **49**, 629–648.

Yanai M, Esbensen S and Chu J-H (1973): Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *J. Atmos. Sci.*, **30**, 611–627.

Yang GY and Slingo J (2001): The Diurnal Cycle in the Tropics. *Mon. Wea. Rev.*, **129**, 784–801.

Yang GY, Hoskins B and Slingo J (2007): Convectively coupled equatorial waves. Part III: Synthesis structures and their forcing and evolution, *J. Atmos. Sci.*, **64**, 3438–3451

Yang GY, Slingo J and Hoskins B (2009): Convectively coupled equatorial waves in high resolution Hadley centre climate models, *J. Clim.*, **22**, 1897–1919.

Yasunari T (1979): Cloudiness fluctuations associated with the Northern Hemisphere summer monsoon, *J. Meteorol. Soc. Jpn.*, **57**, 227–242.

Zerroukat M and Allen T (2012): A three-dimensional monotone and conservative Semi-Lagrangian scheme (SLICE-3D) for transport problems. *Q.J.R. Meteorol. Soc.*, **138**, 1640–1651.

Zhang C (2005): Madden-Julian Oscillation, *Rev. Geophys.*, **43**, RG2003.

Zhong W and Haigh JD (2000): An efficient and accurate correlated-k parameterization of infrared radiative transfer for troposphere–stratosphere–mesosphere GCMs. *Atmosph. Sci. Lett.*, **1**, 125–135.