SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Representing uncertainty in weather and climate prediction
Computer Project Account:	spgbtpuc
Start Year - End Year :	2012 - 2014
Principal Investigator(s)	Tim Palmer Hannah Christensen (*nee* Arnold)
Affiliation/Address:	Atmospheric, Oceanic and Planetary Physics, Clarendon Laboratory, University of Oxford, Oxford, U.K.
Other Researchers (Name/Affiliation):	Antje Weisheimer, ECMWF

The following should cover the entire project duration.

Summary of project objectives

The aim of the project is to develop and test representations of model uncertainty in a state-of-the-art NWP model. Following the "seamless prediction" paradigm, it is hoped that these developments in NWP will then feed into a probabilistic earth system model. Specifically, this project will study the representation of model uncertainty in the ECMWF convection parametrisation scheme. At present, the Stochastically Perturbed Parametrisation Tendencies (SPPT) scheme does not distinguish between the different physics schemes, which likely have vastly different error characteristics. Deep convection is generally acknowledged to be the parametrisation to which weather and climate models are most sensitive, so uncertainty originating in its representation must be well represented. Through studying the effect of the SPPT scheme applied only to the convection parametrisation tendency, and comparing the results with a perturbed parameter approach, we hope to ascertain how to accurately represent model uncertainty in the ECMWF convection scheme.

Summary of problems encountered

When we come to run each new cycle, there are usually a few minor problems with setting experiments running as an external Member State user. However the user support department has always sorted out these issues rapidly.

Experience with the Special Project framework

Our experience has been extremely positive. The application procedure is transparent, and yearly progress reports are a reasonable requirement. In particular, we have found the user support department to be an invaluable resource - Paul Dando in particular has always gone out of his way to help us with technical and scientific problems, making enquiries on our behalf in the research department for anything he did not already know.

Summary of results

It is now acknowledged that representing model uncertainty in atmospheric simulators is essential for the production of reliable probabilistic forecasts, and a number of different techniques have been proposed for this purpose. During this special project, we considered both perturbed parameter and stochastic approaches to representing model uncertainty.

Firstly, we developed new perturbed-parameter schemes for use in the ECMWF convection scheme. Two types of scheme are developed and implemented. Both schemes represent the joint uncertainty in four of the parameters in the convection parameterization scheme, which was estimated using the Ensemble Prediction and Parameter Estimation System (EPPES). The first scheme developed is a fixed perturbed parameter scheme, where the values of uncertain parameters are varied between ensemble members, but held constant over the duration of the forecast. The second is a stochastically varying perturbed parameter scheme. The performance of these schemes was compared to the ECMWF operational stochastic scheme, stochastically perturbed parameterization tendencies (SPPT), and to a model that does not represent uncertainty in convection. The skill of probabilistic forecasts made using the different models was evaluated. While the perturbed parameter schemes improve on the stochastic parameterization in some regards, the SPPT scheme outperforms the perturbed parameter approaches when considering forecast variables that are particularly sensitive to convection. Overall, SPPT schemes are the most skillful representations of model uncertainty owing to convection parameterization.

Secondly, we considered a generalisation to the existing ECMWF SPPT scheme. We investigated the impact of perturbing the five IFS physics schemes using independent random fields instead of using the same random field. In the operational SPPT, the uncertainty is assumed to be proportional to the total net tendency, whereas this generalisation to SPPT assumes that the errors from the different parametrisation schemes are uncorrelated, and that the uncertainty in the forecast is proportional to the individual tendencies. If the physics tendencies tend to act in opposite directions, this "independent SPPT" (SPPTi) will acknowledge the large uncertainty in the individual tendencies and will increase the forecast ensemble spread. The SPPTi scheme had a significant positive impact on the spread of the ensemble forecasts at T159 resolution. The improvement in ensemble spread was also observed at T639, though this was accompanied by an increase in RMSE for some variables. It was found that the impact could largely be attributed to perturbing the convection scheme separately from the other physics schemes, though independently perturbing the cloud or radiation tendencies also had a large impact.

Please see the attached extract (chapters 5 and 6) from Arnold (2013) that describes the results in full:

> Arnold, H. M., 2013. Stochastic Parametrisation and Model Uncertainty, PhD thesis, University of Oxford.

List of publications/reports from the project with complete references

Arnold, H. M., 2013. *Stochastic Parametrisation and Model Uncertainty*, PhD thesis, University of Oxford.

Christensen, H. M., Moroz, I. M. and Palmer, T. N., 2015, Stochastic and perturbed parameter representations of model uncertainty in convection parametrization. J. Atmos. Sci., 72, 2525-2544

Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new Special Project.)

The improvement in forecast reliability when SPPTi is used in the EPS is very promising, though the scheme also resulted in an increase in forecast error for some variables. Nevertheless, we feel the scheme has merit. Our key conclusion is that the only way to improve SPPTi is to systematically measure the characteristics of the error that SPPTi seeks to represent. How correlated are the errors between the different parametrisation schemes? What are the spatial and temporal scales on which they vary? To answer these questions, we propose a series of coarse graining experiments be carried out on high-resolution model data and the characteristics of the error be measured. This is also the central aim of a new special project for 2015 - "Constraining stochastic parametrisation schemes through coarse graining" (spgbtpcs).

The skill of stochastic parametrisation schemes at representing uncertainty in medium range and seasonal forecasts is well established. An exciting new area of research considers the performance of stochastic parametrisation schemes in uncoupled and coupled climate-length integrations. This is the basis of a second special project which also started in 2015: "The Impact of Stochastic Parametrisations in Climate Models: EC-EARTH System Development and Application" (spgbtpsp), where we are in the process of implementing stochastic parametrisation schemes in the EC-EARTH climate model.

Stochastic Parametrisation and Model Uncertainty

Hannah Mary Arnold

Jesus College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity Term 2013

Stochastic Parametrisation and Model Uncertainty

Hannah Mary Arnold, Jesus College

Submitted for the degree of Doctor of Philosophy, Trinity Term 2013

Abstract

Representing model uncertainty in atmospheric simulators is essential for the production of reliable probabilistic forecasts, and stochastic parametrisation schemes have been proposed for this purpose. Such schemes have been shown to improve the skill of ensemble forecasts, resulting in a growing use of stochastic parametrisation schemes in numerical weather prediction. However, little research has explicitly tested the ability of stochastic parametrisations to represent model uncertainty, since the presence of other sources of forecast uncertainty has complicated the results.

This study seeks to provide firm foundations for the use of stochastic parametrisation schemes as a representation of model uncertainty in numerical weather prediction models. Idealised experiments are carried out in the Lorenz '96 (L96) simplified model of the atmosphere, in which all sources of uncertainty apart from model uncertainty can be removed. Stochastic parametrisations are found to be a skilful way of representing model uncertainty in weather forecasts in this system. Stochastic schemes which have a realistic representation of model error produce reliable forecasts, improving on the deterministic and the more "traditional" perturbed parameter schemes tested.

The potential of using stochastic parametrisations for simulating the climate is considered, an area in which there has been little research. A significant improvement is observed when stochastic parametrisation schemes are used to represent model uncertainty in climate simulations in the L96 system. This improvement is particularly pronounced when considering the regime behaviour of the L96 system — the stochastic forecast models are significantly more skilful than using a deterministic perturbed parameter ensemble to represent model uncertainty. The reliability of a model at forecasting the weather is found to be linked to that model's ability to simulate the climate, providing some support for the seamless prediction paradigm.

The lessons learned in the L96 system are then used to test and develop stochastic and perturbed parameter representations of model uncertainty for use in an operational numerical weather prediction model, the Integrated Forecasting System (IFS). A particular focus is on improving the representation of model uncertainty in the convection parametrisation scheme. Perturbed parameter schemes are tested, which improve on the operational stochastic scheme in some regards, but are not as skilful as a new generalised version of the stochastic scheme. The proposed stochastic scheme has a potentially more realistic representation of model error than the operational scheme, and improves the reliability of the forecasts.

While studying the L96 system, it was found that there is a need for a proper score which is particularly sensitive to forecast reliability. A suitable score is proposed and tested, before being used for verification of the forecasts made in the IFS.

This study demonstrates the power of using stochastic over perturbed parameter representations of model uncertainty in weather and climate simulations. It is hoped that these results motivate further research into physically-based stochastic parametrisation schemes, as well as triggering the development of stochastic Earth-system models for probabilistic climate prediction.

Abbreviations

1DD 1 Degree Daily YOTC dataset
A Additive noise stochastic parametrisation used in Chapters 2 and 3
ALARO Aire Limitée Adaptation/Application de la Recherche à l'Opérationnel (
AMIP Atmospheric Model Intercomparison Project
AR(1) First Order Autoregressive
BS Brier Score
BSS Brier Skill Score, usually calculated with respect to climatology.
CA Cellular Automaton
CAPE Convectively Available Potential Energy
CASBS Cellular Automaton Stochastic Backscatter Scheme
CCN Cloud Condensation Nuclei
CIN Convective Inhibition
CMIPn Climate Model Intercomparison Project, Phase n
CONV IFS Convection parametrisation scheme
CONVi CONV perturbed independently using SPPT
CRM Cloud Resolving Model
DEMETER Development of a European Multimodel Ensemble system for seasonal to inTERannual prediction
ECMWF European Centre for Medium-Range Weather Forecasts
EDA Ensembles of Data Assimilation
ENSO El Niño Southern Oscillation
EOF Empirical Orthogonal Function
EPS Ensemble Prediction System
EUROSIP European Seasonal to Interannual Prediction project
GCM General Circulation Model
GLOMAP Global Model of Aerosol Processes
GPCP Global Precipitation Climatology Project
IFS Integrated Forecasting System — the ECMWF global weather forecasting model
IGN Ignorance Score
IGN$_{\mathbf{L}}$ Ignorance Score calculated following Leutbecher (2010)
IGNSS Ignorance Skill Score, usually calculated with respect to climatology
IPCC AR4 Intergovernmental Panel on Climate Change's fourth assessment report
ITCZ Intertropical Convergence Zone
KL Kullback-Leibler Divergence
KS Kolmogorov-Smirnov Statistic
LES Large Eddy Simulation
LSWP IFS Large Scale Water Processes (clouds) parametrisation scheme
LSWPi LSWP perturbed independently using SPPT
L96 The Lorenz '96 System — the second model described in Lorenz (1996)
M Multiplicative noise stochastic parametrisation used in Chapters 2 and 3
MA Multiplicative and Additive noise stochastic parametrisation used in Chapters 2 and 3
MME Multi-Model Ensemble

MOGREPS Met Office Global and Regional Ensemble Prediction System
MTU Model Time Units in the Lorenz '96 system. One MTU corresponds to approximately five atmospheric days.
NAO North Atlantic Oscillation
NCEP National Centers for Environmental Prediction
NOGW IFS Non-Orographic Gravity Wave Drag parametrisation scheme
NOGWi NOGW perturbed independently using SPPT
NWP Numerical Weather Prediction
PC Principal Component
pdf Probability Density Function
PPT Precipitation
RDTT IFS Radiation parametrisation scheme
RDTTi RDTT perturbed independently using SPPT
REL Reliability component of the Brier Score
RMS Root Mean Square
RMSE RMS Error
RPS Ranked Probability Score
RPSS Ranked Probability Skill Score, usually calculated with respect to climatology
SCM Single Column Model
SD State Dependent additive noise stochastic parametrisation used in Chapters 2 and 3
SKEB Stochastic Kinetic Energy Backscatter
SME Single-Model Ensemble
SPPT Stochastically Perturbed Parametrisation Tendencies
SPPTi Independent Stochastically Perturbed Parametrisation Tendencies
TCWV Total Column Water Vapour
TGWD IFS Turbulence and Gravity Wave Drag parametrisation scheme
TGWDi TGWD perturbed independently using SPPT
THORPEX The Observing-System Research and Predictability Experiment
T159 IFS spectral resolution - triangular truncation of 159
T850 Temperature at 850 hPa
U200 Zonal wind at 200 hPa
U850 Zonal wind at 850 hPa
UM Unified Model — the U.K. Met Office weather forecasting model
WCRP World Climate Research Programme
WWRP World Weather Research Programme
YOTC Year of Tropical Convection
Z500 Geopotential height at 500 hPa

5

Experiments in the IFS: Perturbed Parameter Ensembles

But as the cool and dense Air, by reason of its greater Gravity, presses upon the hot and rarified, 'tis demonstrated that this latter must ascend in a continued stream as fast as it Rarifies

– Edmund Halley, 1686

5.1 Introduction

In Chapters 2 and 3, the impact of perturbed parameter and stochastic representations of model uncertainty on forecasts in the L96 system was considered. The results from that simple system indicated that the best stochastic parametrisations produced more skilful forecasts than the perturbed parameter schemes. However, the perturbed parameter ensembles were skilful in forecasting the weather of the system, and performed better than many of the sub-optimal stochastic schemes, such as those which used white noise. This chapter will extend the earlier work in the Lorenz '96 system by comparing the performance of a stochastic and perturbed parameter representation of model uncertainty in the ECMWF convection scheme. Convection is generally acknowledged to be the parametrisation to which weather and climate models are most sensitive (Knight et al., 2007), and it is therefore imperative that the uncertainty originating in the parametrisation of convection is well represented.

In Section 5.2, the ECMWF model, the Integrated Forecasting System (IFS), is described,

and its parametrisation schemes are outlined. In Section 5.3, a generalisation to SPPT is formulated which allows the stochastic perturbations to the convection tendency to be turned off and replaced with alternative schemes. Section 5.4 describes the perturbed parameter representations of model uncertainty which have been developed for this study. In Section 5.5, the experimental procedure and verification techniques are described, and results are presented in Section 5.6. In Section 5.7, the results are discussed and some conclusions are drawn.

5.2 The Integrated Forecasting System

The IFS is the operational global weather forecasting model developed and operated by ECMWF. The following description refers to model version CY37R2, and the configuration which was operational in 2011. The IFS comprises several components (Anderson and Persson, 2013; ECMWF, 2012). The **atmospheric general circulation model** consists of diagnostic equations describing the physical relationship between pressure, density, temperature and height, together with prognostic equations describing the time evolution of horizontal wind speed (zonal, U, and meridional, V), temperature (T), humidity (q), and surface pressure. The model dynamical equations describe the evolution of the resolved-scale variables, while the effect of sub-grid scale processes is included using physically motivated, but statistically derived, parametrisation schemes. The atmospheric model contains a number of these parametrisation schemes, which will be discussed further in Section 5.2.1. Each scheme operates independently on each atmospheric vertical column. Two stochastic parametrisation schemes can be used to represent model uncertainty: SPPT (Section 1.4.3.1) and SKEB (Section 1.4.3.2).

The atmospheric model is numerically integrated using a semi-Lagrangian advection scheme combined with a semi-implicit time integration scheme. Together, these provide stability and accuracy, enabling the use of larger time steps to reduce integration time. Horizontally, the IFS is a dual spectral/grid-point model. The dynamical variables are represented in spectral space to aid the calculation of horizontal derivatives and the time-stepping scheme. The physical parametrisations are spatially localised, so are implemented in grid-point space on a reduced Gaussian grid. The model then converts back and forth between grid-point and spectral space. Vertically, the atmospheric model is discretised using sigma co-ordinates. This is a hybrid co-ordinate system; near the surface, the sigma levels follow the orographic contours whereas higher in the atmosphere the sigma levels follow surfaces of constant pressure.

Physics Parametrisation Scheme	Abbreviation
Radiation	RDTT
Turbulence and Gravity Wave Drag	TGWD
Non-orographic Gravity Wave Drag	NOGW
Convection	CONV
Large Scale Water Processes	LSWP

Table 5.1: Physical parametrisation schemes in the IFS atmospheric model

The atmospheric model is coupled to a **land surface model** — the H-TESSEL scheme (Hydrology-Tiled ECMWF Scheme for Surface Exchange over Land). The land within each grid box is represented by up to six different types of surface, with which the atmosphere exchanges water and energy. The atmospheric model is also coupled to an **ocean wave model** called "WAM". The coupling allows the exchange of energy between wind and waves in both directions. Persisted SST anomalies are used out to day ten; the atmosphere and the ocean are coupled through exchanges of heat, momentum and mass.

Data assimilation is used to calculate the starting conditions for the IFS forecasts. The four-dimensional variational data assimilation (4DVar) system combines information from observations with the physical description of the atmosphere contained in the model. This generates a physically reasonable estimate of the state of the atmosphere. However, this method produces no flow-dependent estimate of the uncertainty in the analysis. To estimate this, an ensemble of data assimilations (EDA) is generated: ten equally likely analyses are calculated at a resolution of T399 (Isaksen et al., 2010). They differ from each other due to the introduction of small perturbations in the observations and SST, as well as perturbations from SPPT.

Operationally, the model is used to produce both a high resolution deterministic forecast and a lower resolution ensemble forecast out to a lead time of fifteen days. The deterministic model is run at a spectral resolution of T1279 (16 km) with 91 levels in the vertical and a time step of 10 minutes. A single forecast is made using unperturbed initial conditions from the 4DVar system, without the SPPT and SKEB parametrisation schemes in the forecast model. The EPS is operationally run with a spectral resolution of T639 (30 km) and 62 vertical levels, and with a time step of 20 minutes. The ensemble has fifty perturbed members and one control member. Initial condition uncertainty is sampled using the EDA system, combined with perturbations from the leading singular vectors. The stochastic parametrisations are activated to sample model uncertainty.

5.2.1 Parametrisation Schemes in the IFS

There are five main parametrisation schemes in the IFS, shown in Table 5.1. The physics tendencies from these five schemes are combined with the dynamical tendencies using a technique called "fractional stepping" (Wedi, 1999). The schemes are called sequentially, and schemes called later use updated variables. This has the disadvantage of introducing intermediate time steps (hence "fractional stepping") at which the tendencies are updated by each parametrisation scheme in turn. However, it has the advantage of ensuring a balance between the different physical parametrisation schemes.

The first scheme to be called is the radiation scheme (RDTT), including both a long- and short-wave radiation calculation. The full radiation scheme is very expensive, so it is calculated on a coarser grid than the other parametrisation schemes, and the resultant tendencies interpolated to the required resolution. Furthermore, the scheme is not run at every time step: for the high resolution forecast, it is run once an hour, and for the EPS it is run once every three hours. The radiation scheme interacts with clouds: incoming short wave radiation is reflected by clouds, and the clouds emit long wave radiation. Since 2007, the Monte Carlo Independent Column Approximation (McICA) approach has been used to account for clouds. The grid box is divided into a number of sub-columns, each of which has a cloud fraction of '0' or '1' at each vertical level. Instead of calculating the sum over all columns in a grid box and over all radiation intervals (which would be prohibitively expensive), a Monte Carlo approach is used whereby the radiative transfer calculation is performed for a single randomly selected sub column only. This will introduce unbiased random errors into the solution.

The second scheme called accounts for vertical exchange of energy, momentum and moisture due to turbulence and gravity wave drag (TGWD). The scheme accounts for turbulent exchange between the surface and the lowest atmospheric levels. Atmospheric momentum is also affected by sub-grid scale orography. Orography exerts a drag on the atmospheric flow both from blocking the flow in the lowest levels, and due to reflection and absorption of gravity waves.

The third scheme is the non-orographic gravity wave drag scheme (NOGW). Non-orographic gravity waves are generated by convection, the jet stream and frontogenesis. They are particularly important in the stratosphere and mesosphere, where they contribute to driving the Brewer-Dobson circulation, and the quasi-biennial and semi-annual oscillations.

The convection parametrisation (CONV) is based on the mass-flux scheme of Tiedtke

(1989). The scheme describes three types of convective cloud: deep, shallow and mid-level. The convective clouds in a column are represented by a pair of entraining and detraining plumes of a given convective type, which describe updraft and downdraft processes respectively[1]. The choice of convective type determines certain properties of the cloud (such as the entrainment formulation). The mass flux at cloud base for deep convection is estimated by assuming that deep convection acts to reduce convectively available potential energy (CAPE) over some specified (resolution dependent) time scale. Mid-level convection occurs at warm fronts. The mass flux at cloud base is set to be the large scale vertical mass flux at that level. For shallow convection, the mass flux at cloud base is derived by assuming that the moist static energy in the sub-cloud layer is in equilibrium.

Finally, the large scale water processes (LSWP, or "cloud") scheme contains the prognostic equations for cloud liquid water, cloud ice, rain, snow and cloud fraction. It builds on the scheme of Tiedtke (1993), but is a more complete description, including more prognostic variables and an improved representation of mixed phase clouds. Whereas the convection scheme calculates the effect of unresolved convective clouds, the cloud scheme calculates the impact of clouds which are resolved by the model. This means that the same cloud could be represented by a different parametrisation scheme if the resolution of the model changed.

The IFS also contains parametrisations of methane oxidation and ozone chemistry. The tendencies from these schemes do not affect the variable tendencies perturbed by SPPT, so the schemes will not be considered further here.

5.3 Uncertainty in Convection: Generalised SPPT

The operational SPPT scheme addresses model uncertainty in the IFS due to the physics parametrisation schemes by perturbing the physics tendencies using multiplicative noise; the word 'tendency' refers to the change in a variable over a time step. SPPT perturbs the sum of the parametrisation tendencies:

$$T = \frac{\partial X}{\partial t} = D + K + (1 + e) \sum_{i=1}^{5} P_i. \tag{5.1}$$

[1]Entrainment is the mixing of dry environmental air into the moist convective plume, while detrainment is the reverse.

where T is the total tendency in X. D is the tendency from the dynamics, K is horizontal diffusion, P_i is the tendency from the ith physics scheme in Table 5.1, and e is the zero mean random perturbation. The scheme perturbs the tendency for four variables: T, U, V and q. Each variable tendency is perturbed using the same random number field. The perturbation field is generated using a spectral pattern generator. The pattern at each time step is the sum of three independent random fields with horizontal correlation scales of 500, 1000 and 2000 km. These fields are evolved in time using an AR(1) process on time scales of 6 hours, 3 days and 30 days respectively, and the fields have standard deviations of 0.52, 0.18 and 0.06 respectively. It is expected that the smallest scale (500 km and 6 hours) will dominate at a 10 day lead time — the larger scale perturbations are important for monthly and seasonal forecasts.

SPPT does not distinguish between the different parametrisation schemes. However, the parametrisation schemes likely have very different error characteristics, so this assumption may not be valid. In particular, this chapter considers alternative, perturbed parameter representations of model uncertainty in convection. In order to test a perturbed parameter scheme in convection, it is necessary to be able to 'switch off' the SPPT perturbations for the convection parametrisation tendency. A generalised version of SPPT was developed for this chapter, building on earlier work by Alfons Callado Pallares (AEMET)[2]. In this scheme, the multiplicative noise is applied separately to the tendencies from each physics parametrisation scheme,

$$T = D + K + \sum_{i=1}^{5} (1 + e_i) P_i, \tag{5.2}$$

where the stochastic field, e_i, for the convection tendency can be set to zero. In order to detect an improvement in the representation of uncertainty in the convection scheme, the uncertainty in the other four schemes must be well represented. In this experiment, SPPT is used to represent uncertainty in the other four schemes, applying the same stochastic perturbation to each scheme. The stochastic perturbations are three-scale fields with the same characteristics as used in operational SPPT. The SKEB scheme (Section 1.4.3.2) represents a process that is otherwise missing from the model, so will be used in these experiments.

The results of Chapters 2 and 3 indicate that a multiplicative stochastic parametrisation scheme is a skilful representation of model uncertainty in the Lorenz '96 system. Using SPPT

[2]See Section 1.9 for an outline of the code changes to the IFS which have been incorporated from Callado Pallares, and the changes which have been developed as part of this thesis.

to represent convective uncertainty is therefore a good benchmark when testing the perturbed parameter schemes outlined below. The Lorenz '96 system also indicated that multiplicative and additive noise stochastic schemes produced skilful forecasts. It would be interesting to test an additive noise scheme in addition to SPPT for the convective tendencies. Additive noise represents uncertainty in the convection tendency when the deterministic tendency is zero. This uncertainty could be due to the model discretisation, or from errors in the formulation of the convection parametrisation scheme which cannot be captured by a multiplicative noise scheme. However, additive noise schemes will not be investigated further here. Implementing an additive noise scheme in the IFS is problematic in the context of convection (Martin Leutbecher, pers. comm., 2013). The deterministic convection parametrisation acts to vertically redistribute heat and moisture in the atmosphere, drying some levels, and moistening by an equivalent amount at others. A multiplicative noise term does not disrupt this balance. However, an additive term would disrupt this balance, and developing and implementing an additive scheme which preserves the balance is outside the scope of this thesis.

5.4 Perturbed Parameter Approach to Uncertainty in Convection

5.4.1 Perturbed Parameters and the EPPES

When developing a parametrisation scheme, parameters are introduced to represent physical processes within the scheme. For example, in the entraining plume model of convection, the degree to which dry environmental air is turbulently mixed into the plume is assumed to be proportional to the inverse of the radius of the plume, with the constant of proportionality defined to be the entrainment coefficient. This is a simplification of the true processes involved in the convective cloud, and because of this and the sparsity of the required environmental data, physical parameters such as the entrainment coefficient are poorly constrained. However, the evolution of convective clouds and the resultant effects on weather and ultimately global climate are very sensitive to these parameters, and to the entrainment coefficient in particular (Sanderson et al., 2008). Because of this, perturbed parameter models have been proposed to represent the uncertainty in predictions due to the uncertainty in these parameters.

In a perturbed parameter ensemble, the values of a selected set of parameters are sampled

from a distribution representing the uncertainty in their values, and each ensemble member is assigned a different set of parameters. These parameters are fixed globally and for the duration of the integration. The parameter distribution is usually determined through "expert elicitation" whereby scientists with the required knowledge and experience of using the parametrisation suggest upper and lower bounds for the parameter (Stainforth et al., 2005). No information about the relationships between parameters is included in the ensemble, though unrealistic simulations can be removed from the ensemble later (Stainforth et al., 2005).

The poorly constrained nature of these physical parameters can have adverse effects on high-resolution deterministic integrations. Tuning the many hundreds of parameters in atmospheric models is a difficult, lengthy, costly process, usually performed by hand. An attractive alternative is the use of a Bayesian parameter estimation approach. This seeks to provide the probability distribution of parameters given the data, and provides a framework for using new data from forecasts and observations to update prior knowledge or beliefs about the parameter distribution (Beck and Arnold, 1977). One specific proposed technique is the Ensemble Prediction and Parameter Estimation System (EPPES) (Järvinen et al., 2012; Laine et al., 2012), which runs on–line in conjunction with an operational ensemble forecasting system. At the start of each forecast, a set of parameter values for each ensemble member is sampled from the parameters' joint distribution. The joint distribution is updated by evaluating the likelihood function for the forecast and observations after the verifying observations are available. Note that this may be many days after the forecast was initialised, so other perturbed parameter ensemble forecasts will have been initialised in the meantime. In this way, the EPPES approach differs from a Markov Chain Monte Carlo method, which updates the parameter distribution before each new draw.

My collaborators, Peter Bechtold (ECMWF), Pirkka Ollinaho (Finnish Meteorological Institute) and Heikki Järvinen (University of Helsinki), have used this approach with the IFS to better constrain four of the parameters in the convection scheme: ENTRORG, ENTSHALP, DETRPEN, RPRCON:

- ENTRORG represents organised entrainment for positively buoyant deep convection, with a default value of $1.75 \times 10^{-3}\,\mathrm{m}^{-1}$.
- ENTSHALP $\times$ ENTRORG represents shallow entrainment, and the default value for ENTSHALP is 2.

- DETRPEN is the average detrainment rate for penetrative convection, and has a default value of $0.75 \times 10^{-4}\,\mathrm{m}^{-1}$.

- RPRCON is the coefficient for determining the conversion rate from cloud water to rain, and has a default value of 1.4×10^{-3}.

The likelihood function used was the geopotential height at 500 hPa for a ten day forecast. The resultant optimised value of each parameter was used in the high resolution deterministic forecast model, and many of the verification metrics were found to improve when compared to using the default values (Pirkka Ollinaho, pers. comm., 2013). This is very impressive, since the operational version of the IFS is already a highly tuned system.

The EPPES approach also produces a full joint pdf for the chosen parameters. Since Gaussianity is assumed, this takes the form of a covariance matrix for the four parameters. This information is useful for model tuning as it can reveal parameter correlations, and can therefore be used to identify redundant parameters. However, the joint pdf also gives an indication of the uncertainty in the parameters. I have been provided with this information, which I have used to develop a perturbed parameter representation of uncertainty for the ECMWF convection scheme.

5.4.2 Method

The EPPES approach was used to determine the posterior distribution of four parameters in the ECMWF convection scheme at T159. This was calculated in terms of a mean vector, $M(i)$ and covariance matrix with elements $\Sigma_{i,j}$, where $i = 1$ represents ENTRORG, $i = 2$ represents ENTSHALP, $i = 3$ represents DETRPEN, and $i = 4$ represents RPRCON.

$$M = \begin{pmatrix} 0.182804e-02 \\ 0.214633e+01 \\ 0.778274e-04 \\ 0.151285e-02 \end{pmatrix}$$

$$\Sigma = \begin{pmatrix} 0.9648e-07 & -0.2127e-04 & -0.4199e-09 & -0.1839e-07 \\ -0.2127e-04 & 0.9255e-01 & 0.1318e-05 & -0.3562e-04 \\ -0.4199e-09 & 0.1318e-05 & 0.5194e-10 & -0.1134e-08 \\ -0.1839e-07 & -0.3562e-04 & -0.1134e-08 & 0.4915e-07 \end{pmatrix}$$

By comparison with the default values, the M vector indicates the degree to which the parameters should be changed to optimise the forecast. The off-diagonal terms in the Σ matrix indicate there is significant covariance between parameters. This highlights one of the problems with using "expert elicitation" to define parameter distributions — such distributions contain no information about parameter inter-dependencies.

5.4.2.1 Fixed Perturbed Parameter Distribution

The usual method used in perturbed parameter experiments is a fixed perturbed parameter ensemble (Murphy et al., 2004; Sanderson, 2011; Stainforth et al., 2005; Yokohata et al., 2010). Each ensemble member is assigned a set of parameter values which are held constant spatially and over the duration of the integration. Such ensembles are traditionally used for climate–length integrations. It will be interesting to see how well such an ensemble performs at representing uncertainty in weather forecasts.

The multivariate normal distribution supplied by Bechtold, Ollinaho and Järvinen in October 2012 was sampled to give N sets of the four parameters, where the number of ensemble members, $N = 50$. The procedure for this is as follows. N sample vectors, z_n $(1 \leq n \leq N)$, are drawn from the four-dimensional standard multivariate normal distribution ($M = \mathbf{1}$, $\Sigma = \mathbf{I}$). The Cholesky decomposition is used to find matrix A, such that $AA^T = \Sigma$:

$$A = \begin{pmatrix} 3.1062E-4 & 0 & 0 & 0 \\ -6.8490E-2 & 2.9641E-1 & 0 & 0 \\ -1.3518E-6 & 4.1332E-6 & 5.7470E-6 & 0 \\ -5.9200E-5 & -1.3387E-4 & -1.1492E-4 & 1.2048E-4 \end{pmatrix}$$

The samples from the standard multivariate distribution are transformed to samples from the correct parameter distribution, x_n, using the transformation:

$$x_n = M + Az_n \tag{5.3}$$

Two types of fixed perturbed parameter ensemble are considered here. The first uses the same fifty sets of four parameters for all starting dates ("TSCP"). Sampling of the parameters is performed offline: Latin hypercube sampling is used to define fifty percentiles at which to sample the standard multivariate normal distribution, before (5.3) is used to transform to parameter space. This technique ensures the joint distribution is fully explored. The covariance of the resultant sample is checked against the EPPES covariance matrix; 10,000 iterations found a sample whose covariance matrix differed by less than 5% from the true matrix. The sampled parameter values are shown in Table 5.2. The second type of fixed perturbed parameter ensemble uses N new sets of parameters for each initial condition ("TSCPr"). This sampling is performed online, and the samples are not optimised. However, when forecasts from many starting conditions are taken together, the ensemble is sufficient to fully sample the joint pdf.

5.4.2.2 Stochastically Varying Perturbed Parameter Distribution

Khouider and Majda (2006) recognise that a problem with many deterministic parametrisation schemes is the presence of parameters that are "nonphysically kept fixed/constant and spatially homogeneous". An alternative to the fixed perturbed parameter ensemble described above is a stochastically varying perturbed parameter ensemble ("TSCPv") where the parameter values are varied spatially and temporally following the EPPES distribution. However, the EPPES technique contains no information about the correct spatial and temporal scales on which to vary the parameters. Since the likelihood function is evaluated at day ten of the forecast, the set of parameters must perform well over this time window to produce a skilful forecast; this indicates that ten days could be a suitable temporal scale. The likelihood function evaluates the skill of the forecast using the geopotential height at 500 hPa. The likelihood function will therefore focus on the midlatitudes, where the geopotential height has high variability. A suitable spatial scale could therefore be $\sim 1000\,\mathrm{km}$. The SPPT spectral pattern generator is a suitable technique for stochastically varying the parameters in the convection scheme. It generates a spatially and temporally correlated field of random numbers.

Selected Permutations				
Number	ENTRORG	ENTSHALP	DETRPEN	RPRCON
1	1.8203e-3	2.3121e+0	8.7197e-5	1.3731e-3
2	2.4122e-3	1.7785e+0	6.8076e-5	1.5406e-3
3	1.6561e-3	1.9652e+0	6.8473e-5	1.6606e-3
4	1.1054e-3	2.4199e+0	8.0037e-5	1.5757e-3
5	2.0785e-3	1.4016e+0	7.5603e-5	1.4454e-3
6	1.7574e-3	1.8280e+0	6.4998e-5	1.6689e-3
7	1.9311e-3	2.2220e+0	8.9559e-5	1.1177e-3
8	1.6377e-3	2.4955e+0	7.8693e-5	1.4802e-3
9	1.8987e-3	1.9838e+0	7.6484e-5	1.3189e-3
10	1.5554e-3	2.4064e+0	7.5846e-5	1.4879e-3
11	1.4116e-3	2.1398e+0	6.7459e-5	1.7171e-3
12	1.5061e-3	2.2247e+0	8.3208e-5	1.4336e-3
13	1.9821e-3	2.3726e+0	8.0642e-5	1.3901e-3
14	1.9148e-3	2.3462e+0	8.5138e-5	1.3017e-3
15	2.0575e-3	2.3785e+0	7.5289e-5	1.4298e-3
16	1.2438e-3	2.2379e+0	7.3377e-5	1.9182e-3
17	1.4471e-3	2.1781e+0	8.0664e-5	1.5861e-3
18	1.6914e-3	2.6139e+0	7.6817e-5	1.5926e-3
19	1.3696e-3	1.9646e+0	7.4865e-5	1.8839e-3
20	1.5985e-3	2.0827e+0	7.8839e-5	1.8027e-3
21	2.1007e-3	2.7757e+0	9.9625e-5	9.2852e-4
22	1.9999e-3	1.8012e+0	7.8752e-5	1.5595e-3
23	1.7084e-3	2.2402e+0	8.1817e-5	1.4210e-3
24	2.5506e-3	2.0698e+0	7.2315e-5	1.6063e-3
25	1.8514e-3	2.1935e+0	7.3820e-5	1.6587e-3
26	1.8047e-3	1.9516e+0	7.9388e-5	1.6842e-3
27	1.8358e-3	2.1223e+0	7.5268e-5	1.6807e-3
28	1.9477e-3	2.1125e+0	7.7636e-5	1.3370e-3
29	1.7890e-3	2.3366e+0	7.9808e-5	1.5194e-3
30	1.6740e-3	2.1128e+0	7.6250e-5	1.6079e-3
31	1.7733e-3	2.0756e+0	7.1866e-5	1.8394e-3
32	2.3390e-3	2.5212e+0	8.4617e-5	1.1419e-3
33	1.8828e-3	1.9526e+0	8.0540e-5	1.4474e-3
34	2.2864e-3	1.8812e+0	7.4853e-5	1.5983e-3
35	2.1500e-3	1.5878e+0	7.2808e-5	1.6111e-3
36	1.9647e-3	2.6737e+0	8.4574e-5	1.2773e-3
37	1.5776e-3	2.5651e+0	8.6837e-5	1.2864e-3
38	1.6185e-3	2.5899e+0	8.1101e-5	1.3479e-3
39	2.2445e-3	1.6171e+0	6.7066e-5	1.7770e-3
40	1.5317e-3	1.9515e+0	8.3194e-5	1.3346e-3
41	2.0375e-3	2.2471e+0	7.9687e-5	1.3254e-3
42	2.2090e-3	2.3014e+0	7.0050e-5	1.4242e-3
43	1.7413e-3	2.2027e+0	8.2246e-5	1.4821e-3
44	2.1779e-3	1.6718e+0	5.7394e-5	1.6127e-3
45	1.4782e-3	2.5573e+0	9.0479e-5	1.2180e-3
46	2.1244e-3	1.9506e+0	7.9765e-5	1.3744e-3
47	1.3171e-3	1.8954e+0	8.4434e-5	1.6384e-3
48	1.8671e-3	2.1600e+0	7.8113e-5	1.5982e-3
49	2.0184e-3	2.2348e+0	7.7212e-5	1.4772e-3
50	1.7250e-3	1.6116e+0	6.8595e-5	2.1027e-3

Table 5.2: Chosen perturbed convection parameters for the fixed perturbed parameter experiment.

In this experiment the standard SPPT settings were used in the pattern generator. A three scale composite pattern is used which has the same spatial and temporal correlations as used in SPPT. The standard deviations of these independent patterns are 0.939 (smallest scale), 0.325 and 0.108 (largest scale) to give a total standard deviation of 1. These settings vary the parameters faster and on smaller spatial scales than the scales to which EPPES is sensitive, as estimated above. However it will still be useful as a first test, and when combined with the fixed perturbed parameter ensemble (varying on an ∞ spatial and temporal scale), it can provide bounds on the skill of such a representation of model uncertainty.

The SPPT pattern generator is used to generate four independent composite fields with mean 0 and standard deviation 1. The correct covariance structure is introduced using the transformation matrix, A: $x_{i,j,t} = M + Az_{i,j,t}$, where the indices i, j refer to latitude and longitude and t refers to time. The parameters do not vary as a function of height since the convection parametrisation is applied columnwise in the model. The resultant four covarying fields are used to define the values of the four convection parameters as a function of position and time.

5.5 Experimental Procedure

Parameter estimation was carried out with the EPPES system using IFS model version CY37R3 for 45 dates between 12 May 2011 and 8 August 2011, with forecasts initialised every 48 hours. The same model version is used here for consistency. Different initial dates to those used to estimate the joint pdfs must be selected (an out of sample test), but taken from the same time of the year since the EPPES estimated pdfs may be seasonally dependent. In order to detect improvements in the model uncertainty representation, it is important that initial condition uncertainty is well represented in the ensemble forecast. The best technique possible will be used, which for the IFS involves using hybrid EDA/singular vector estimates for the perturbations. The initial dates used must be after June 2010, when the EDA system became operational. The selected dates for the hindcasts are therefore from Summer 2012. The parametrisation schemes will be tested at T159 (1.125°) using a fifty member ensemble forecast. The schemes are tested using ten-day hindcasts initialised every five days between 14 April and 6 September 2012 (30 dates in total). Persistent SSTs are used instead of a dynamical ocean. The high-resolution ECMWF 4DVar analysis is used for verification.

	Other Four Tendencies: SPPT
Convection: Zero	TSCZ
Convection: SPPT	TSCS
Convection: Perturbed Parameters (constant)	TSCP
Convection: Perturbed Parameters (resampled)	TSCPr
Convection: Perturbed Parameters (varying)	TSCPv

Table 5.3: Proposed experiments for investigating the representation of uncertainty in the ECMWF convection parametrisation scheme.

Five experiments are proposed to investigate the representation of model uncertainty in the convection scheme in the IFS (Table 5.3). In each experiment, the uncertainty in the other four parametrisation tendencies (radiation, turbulence and gravity wave drag, non-orographic gravity wave drag, large scale water processes) is represented by SPPT ("TS"). In the first experiment, there is no representation of uncertainty in the convection tendency ("CZ"). In the second, SPPT is used to represent uncertainty in the convection tendency ("CS" — equivalent to the operational SPPT parametrisation scheme). In the final three, uncertainty in the convection tendency is represented by a static perturbed parameter ensemble, with and without resampling of parameters for different start dates ("CPr" and "CP" respectively), and by a stochastically varying perturbed parameter ensemble ("CPv").

In order to compare the different representations of convection model uncertainty, the SPPT scheme must correctly account for uncertainty in the other four tendencies. Therefore verification will be performed in a two stage process. Firstly, the calibration of the ensemble will be checked in a region with little uncertainty due to convection, i.e., where there is little convective activity. The five experiments in Table 5.3 should perform very similarly in this region as they have the same representation of uncertainty in the other four tendencies. Secondly, a region where convection is the dominant process will be selected to test the different uncertainty schemes. Given that model uncertainty has been accounted for in the other four parametrisations using SPPT, and that a region has been selected where the model uncertainty is dominated by deep convection, a scheme which accurately represents uncertainty in deep convection will give a reliable forecast in this region, and any detected improvements in forecast skill can be attributed to an improvement in representation of uncertainty in the convection scheme.

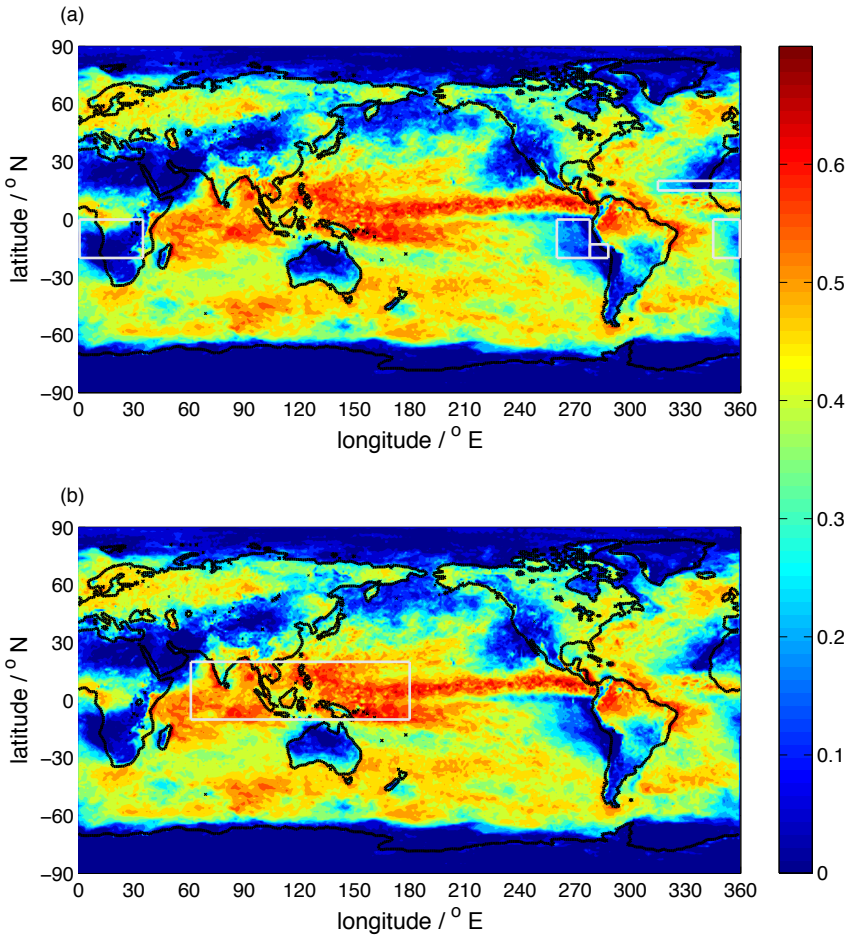

Figure 5.1: Convection diagnostic (colour) derived from the IFS tendencies calculated as part of the YOTC project (see text for details). (a) Regions where the diagnostic is close to zero (bounded by grey boxes), indicating there is little convection. (b) Regions where the diagnostic is large (bounded by grey box), indicating convection is the dominant process.

5.5.1 Definition of Verification Regions

The regions of interest are defined using the Year of Tropical Convection (YOTC) dataset from ECMWF. YOTC was a joint WCRP and World Weather Research Programme/The Observing-System Research and Predictability Experiment (WWRP/THORPEX) project which aimed to focus research efforts on the problem of organised tropical convection. The ECMWF YOTC dataset consists of high resolution analysis and forecast data for May 2008 — April 2010. In particular, the IFS parametrisation tendencies were archived at every time step out to a lead time of ten days.

The 24-hour cumulative temperature tendencies at 850 hPa for each parametrisation scheme are used. Forecasts initialised from 30 dates between 14 April and 6 September 2009 are selected, with subsequent start dates separated by five days. To identify regions where convection is the dominant process, the ratio between the magnitude of the convective tendency and the sum of the magnitudes of all tendencies is calculated, and is shown in Figure 5.1. This diagnostic

can be used to define regions where there is little convection (the ratio is close to zero) or where convection dominates (the ratio greater than 0.5). Since the forecasting skill of the IFS is strongly latitudinally dependent, both the regions with little convection and with significant convection are defined in the tropics (25°S–25°N). Both regions are approximately the same size, and cover areas of both land and sea. Any differences in the forecast verification between these two regions will be predominantly due to convection.

5.5.2 Chosen Diagnostics

Four variables of interest have been selected which will be used to verify the forecasts. Temperature and zonal wind at 850 hPa (T850 and U850 respectively), correspond to fields at approximately 1.5 km altitude, and falls above the boundary layer in many places. The geopotential height at 500 hPa (Z500) is a standard ECMWF diagnostic variable. It is particularly useful in the extra-tropics where it shows characteristic features corresponding to low and high pressure weather systems. The zonal wind at 200 hPa (U200) is particularly interesting when considering convection. This is because 200 hPa falls close to the tropopause, where deep convection is capped. Convective outflow often occurs at this level, which can be detected in U200.

For each variable, the impact of the schemes will be evaluated using a number of the diagnostics described in Chapter 1:

- Bias (Section 1.7.2.1)
- RMSE compared to RMS spread (Section 1.7.3.1)
- RMS error-spread graphical diagnostic (Section 1.7.3.1)
- Forecast skill scores: RPSS, IGNSS, ESS (Sections 1.7.1.2 and 1.7.1.3, and Chapter 4 respectively)

In convecting regions, precipitation (PPT) and total column water vapour (TCWV) will also be considered. PPT is a parametrised product of the convection and large scale water processes parametrisation schemes, so an improvement to the convection scheme should be detectable by studying this variable. The convection scheme effectively redistributes and removes moisture from the atmosphere, so an improvement in TCWV could be indicative of an improvement in the convection scheme.

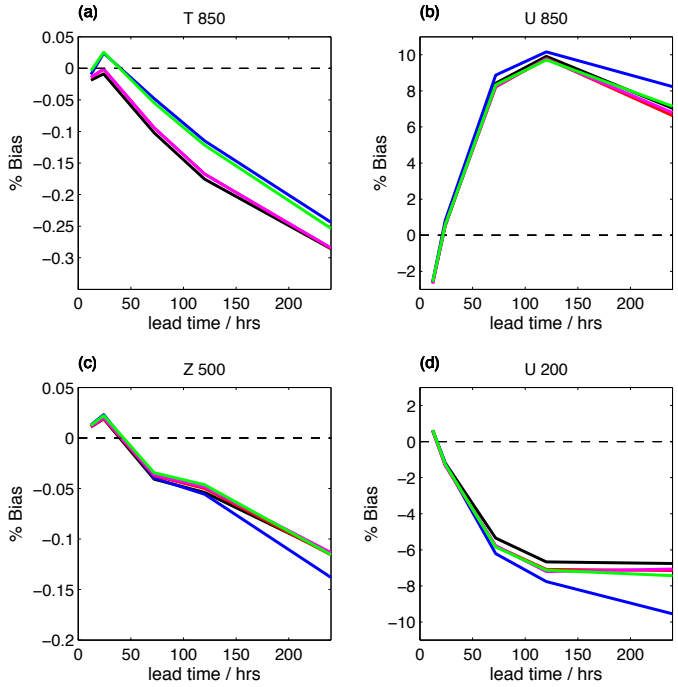

Figure 5.2: Percentage forecast bias in tropical regions with **little convection** as a function of time for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The bias is calculated as described in the text, and given as a percentage of the root mean square of the analysis in the region of interest. The red line is obscured by the magenta line in each figure.

5.6 Verification of Forecasts

5.6.1 Verification in Non-Convecting Regions

Firstly, the impact of the different representations of model uncertainty will be considered in the non-convecting regions defined in Figure 5.1(a). Figure 5.2 shows the percentage bias of forecasts, calculated following (1.24), for regions of little convection. This is a useful diagnostic, as it can indicate the presence of systematic errors in a forecast. A small change is observed in the bias when different uncertainty schemes are used, and in particular the TSCPv scheme (green) performs well for all variables considered. As expected, the TSCP and TSCPr schemes (red and magenta respectively) perform similarly — the bias in the ensemble mean is unaffected by whether the parameter perturbations are resampled for each initial condition. For U850, Z500 and U200, the TSCZ scheme (black) outperforms the TSCS scheme (blue), a result which will be discussed in Section 5.7.

The impact of the uncertainty schemes on the calibration of the ensemble can be summarised by evaluating the RMS error in the ensemble mean and the RMS ensemble spread as a function of time within the region of interest, which should be equal for a well calibrated ensemble.

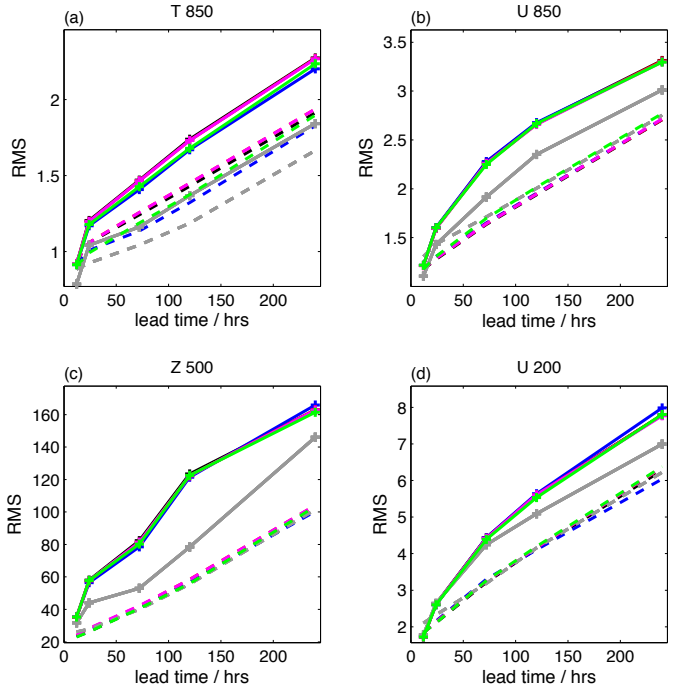

Figure 5.3: Temporal evolution of root mean square ensemble spread (dashed lines) and root mean square error (solid lines) for regions with **little convection** for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The grey curves indicate the results for the operational (T639) EPS forecasts for comparison. The red line is obscured by the magenta line in each figure.

Figure 5.3 shows this diagnostic for the regions with little convection. Forecasts are observed to be slightly under-dispersive for all variables. The underdispersion is large for Z500, but is small for the other variables (note that the y-axes do not start at zero). For comparison, the RMS spread and error curves are also shown for operational ensemble forecasts at T639 (grey). At this higher resolution the ensemble spread is similar, but the RMSE is smaller. The low resolution (T159) used in the five test experiments is responsible for the higher RMSE and therefore contributes to the under-dispersive nature of the ensemble.

A more comprehensive understanding of the calibration of the ensemble can be gained by considering the RMS error-spread diagnostic. The forecast-verification pairs are collected for each spatial point in the region of interest for each starting condition. These pairs are ordered according to their forecast variance, and divided into 30 equally populated bins. The RMS spread and RMSE are evaluated for each bin and displayed on scatter plots. Figure 5.4 shows this diagnostic for regions with little convection at lead times of one, three and ten days. The scattered points should lie on the one-to-one diagonal, shown in black, for a statistically consistent ensemble following (1.26). *The diagnostic indicates a large degree of flow-dependent spread in the ensemble forecasts, with scattered points lying close to the one-to-one line.* The

T850 forecasts are particularly well calibrated, and the spread of the U850 and U200 forecasts are also a skilful indicator of the expected error for all five experiments. The Z500 forecasts show little flow dependency at short lead times, but improve when the longer ten-day forecasts are considered. *As expected, the results from the five experiments are very similar*, and show moderately under-dispersive but otherwise well calibrated forecasts.

5.6.2 Verification in Convecting Regions

The previous section indicates that the model uncertainty in the other four tendencies is sufficiently well represented by SPPT. This section considers forecasts for the strongly convecting regions defined in Figure 5.1(b) to evaluate the impact of the new uncertainty schemes. Figure 5.5 shows the percentage bias for forecasts of T850, U850, Z500 and U200 in this region for the five different schemes considered. The bias is similar for all schemes; no one scheme is systematically better or worse than the others.

Figure 5.6 shows the RMS error and spread as a function of time averaged over all cases for all points within the region of interest. The RMS error in the forecast is similar for all experiments — the perturbed parameter ensembles have not resulted in an increase in error over the operational scheme, except for a slight increase for T850. *However, the fixed perturbed parameter ensemble (red/magenta) has resulted in an increase in spread over the operational TSCS forecast (blue).* This is especially large for T850, where the observed increase is 25% at long lead times. Interestingly, *the TSCZ 'deterministic convection' forecasts of T850 also result in an increase in ensemble spread over TSCS.* This is a counter-intuitive result, as it is expected that using a stochastic parametrisation would increase the spread of the ensemble. This result will be discussed in Section 5.7, and motivates the experiments carried out in Chapter 6. For comparison, the results for the operational EPS are also shown in grey. As is the case in regions with little convection, some of the ensemble under-dispersion at T159 is due to an increased forecast RMSE compared to the operational T639 forecasts, though the forecasts are under-dispersive at both resolutions.

Figure 5.7 shows the RMS error-spread graphical diagnostic for the five forecast models in regions with significant convection. The impact of the different schemes is slight. However, there is a larger difference than in regions with little convection (see Figure 5.4). All schemes remain well calibrated, and do not show large increases in error compared to the operational

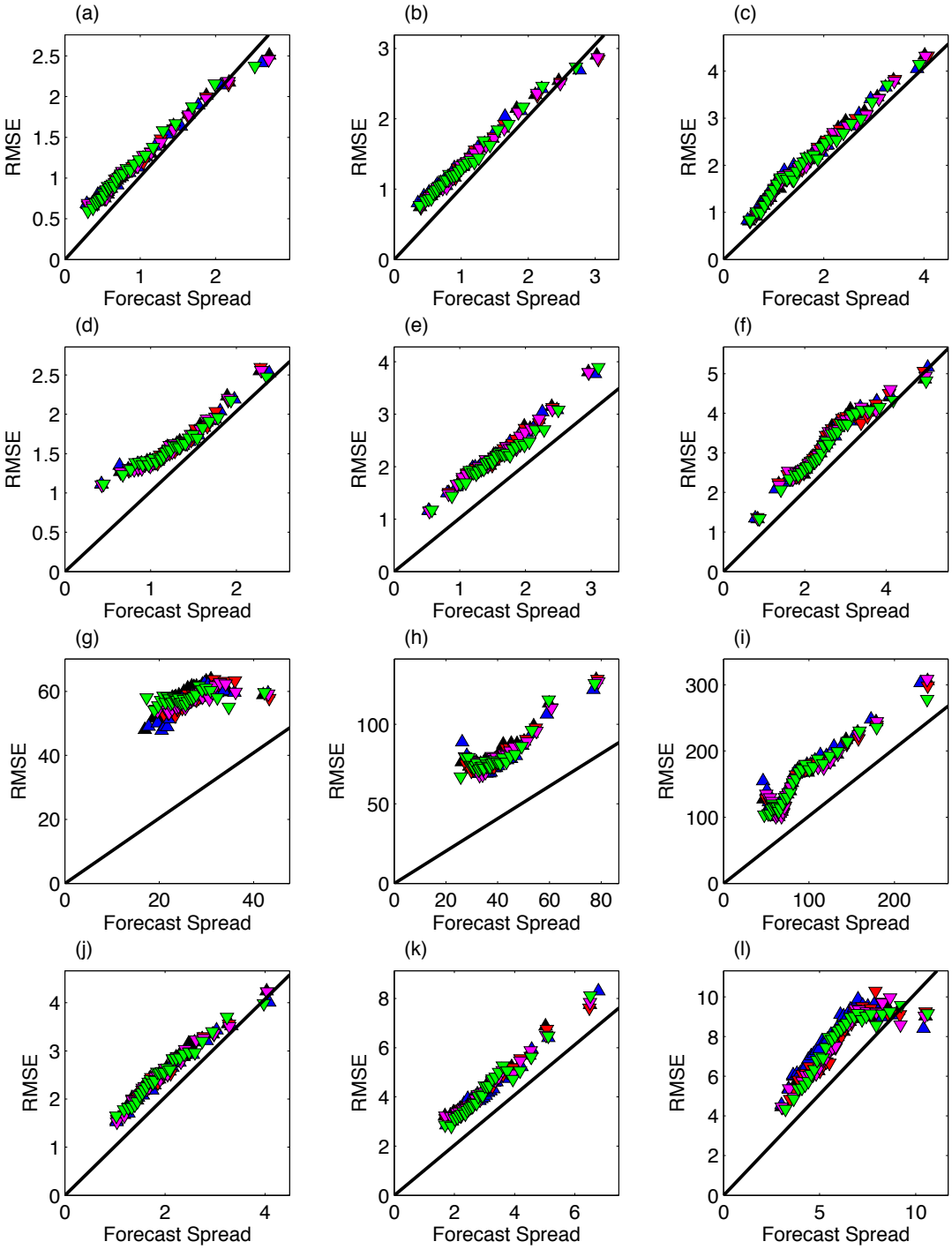

Figure 5.4: Root mean square error-spread diagnostic for tropical regions with **little convection** for (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200, at lead times of 1 day (first column), 3 days (second column) and 10 days (third column) for each variable. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. For a well calibrated ensemble, the scattered points should lie on the one-to-one diagonal shown in black.

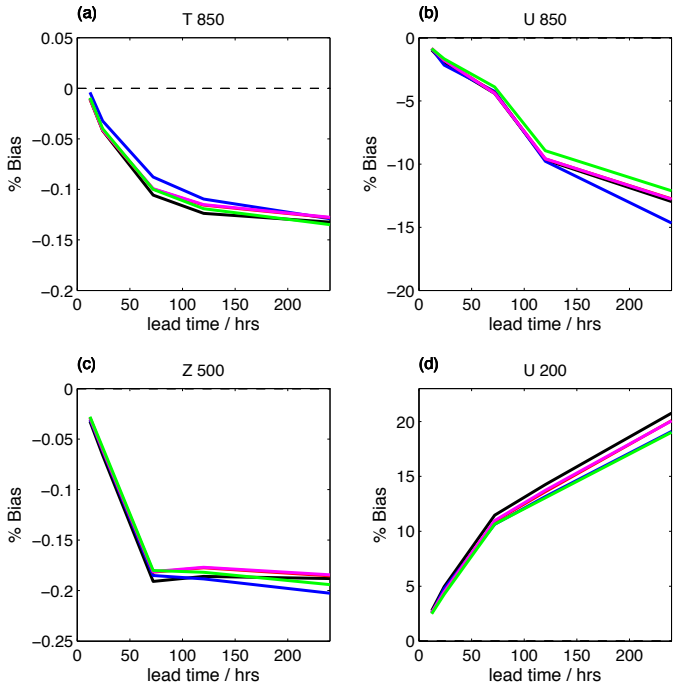

Figure 5.5: Percentage forecast bias in tropical regions with **significant convection** as a function of time for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The bias is calculated as described in the text, and given as a percentage of the root mean square of the analysis in the region of interest. The red line is obscured by the magenta line in each figure.

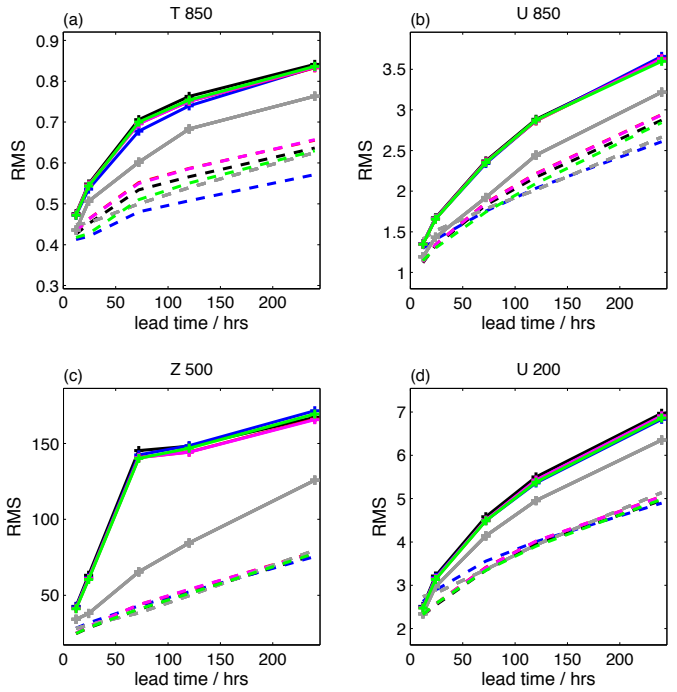

Figure 5.6: Temporal evolution of root mean square ensemble spread (dashed lines) and root mean square error (solid lines) for regions with **significant convection** for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The grey curves indicate the results for the operational (T639) EPS forecasts for comparison. The red line is obscured by the magenta line in each figure.

TSCS forecasts. The fixed perturbed parameter schemes (red and magenta) have larger spread than the other schemes, which is most apparent for T850 in Figures 5.7(a–c). TSCS (blue) has the most under-dispersive ensemble at long lead times, though is better calibrated than the other experiments at short lead times. The TSCPv experiment has intermediate spread, improving on TSCS but under-dispersive compared to the TSCP experiments.

The skill of the forecasts is evaluated using the RPS, IGN and ES as a function of lead time, and the skill scores evaluated with respect to the climatological forecast for the convecting region. The results are shown in Figure 5.8 for each variable of interest. There is no significant difference between the TSCP and TSCPr forecasts according to the skill scores. *The TSCP/TSCPr schemes score highly for a range of variables according to each score*: they perform significantly better than the other forecasts for T850 according to RPSS and IGNSS, for U850 according to IGNSS and ESS, and for Z500 according to all scores (see Appendix A for details of significance testing). For U200, the TSCS forecasts are significantly better than the other forecasts, and the TSCZ forecasts are significantly poorer. However for the other variables, TSCS performs comparatively poorly, and often produces significantly the worst forecasts. This is probably due to the poorer forecast ensemble spread.

5.6.2.1 Precipitation Forecasts

The impact of the different model uncertainty schemes on forecasts of convective precipitation is a good indicator of improvement in the convection scheme. However, it is difficult to verify precipitation forecasts as measurements of precipitation are not assimilated into the IFS using the 4DVar or EDA systems, unlike T, U and Z. One option is to use short-range high resolution (T1279) deterministic forecasts for verification. However, there are known problems with spin-up for accumulated fields like precipitation — the model takes a few time steps to adjust to the initial conditions (Kaallberg, 2011). Instead, the Global Precipitation Climatology Project (GPCP) dataset is used for verification of precipitation forecasts. The GPCP, established by the WCRP, combines information from a large number of satellite and ground based sources to estimate the global distribution of precipitation. The data set used here is the One-Degree Daily (1DD) product (Huffman et al., 2001), which has been conservatively re-gridded onto a T159 reduced Gaussian grid to allow comparison with the IFS forecasts.

Figure 5.9 shows the RMS error-spread diagnostic for convective precipitation. All forecasts

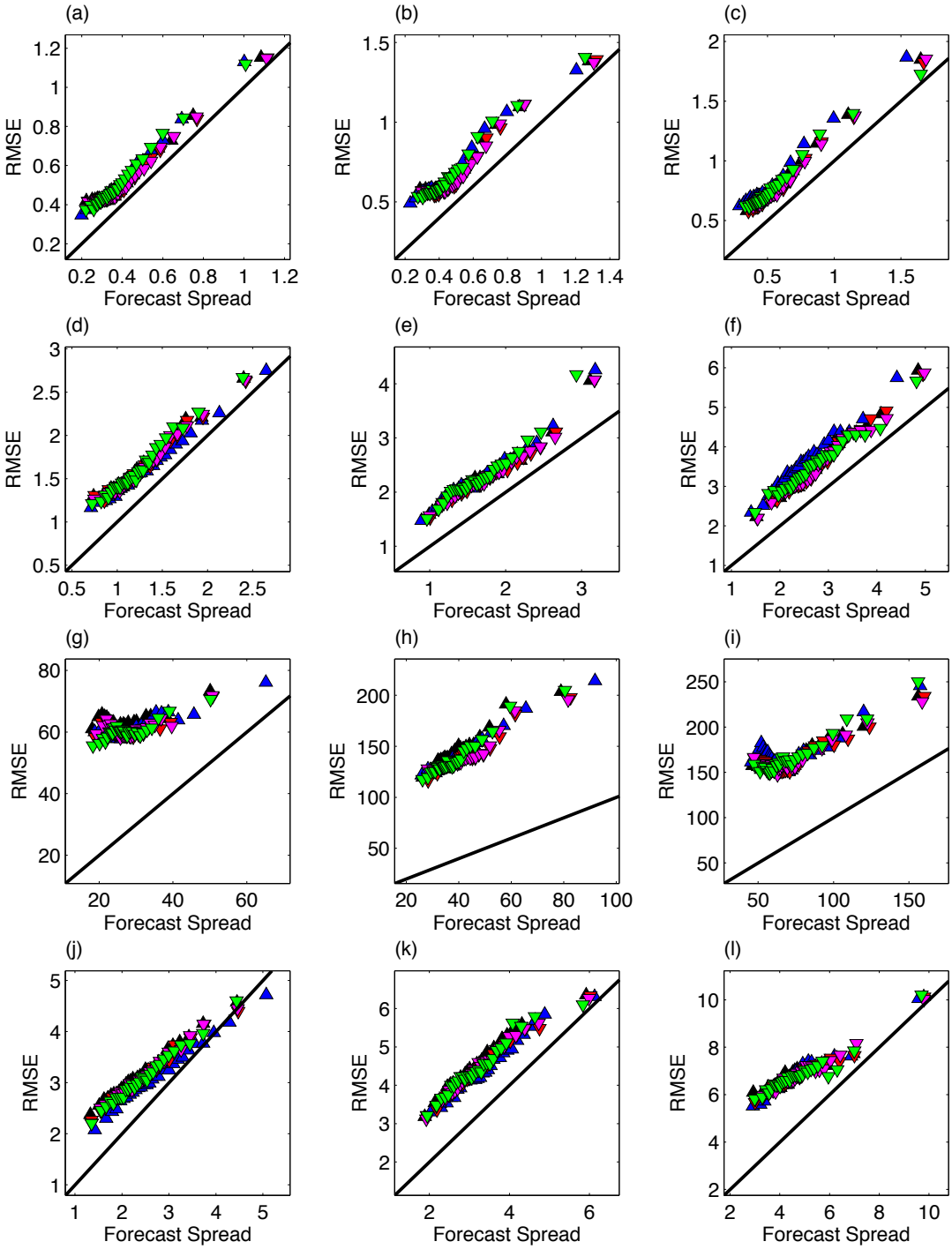

Figure 5.7: Root mean square error-spread diagnostic for tropical regions with **significant convection** for (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200, at lead times of 1 day (first column), 3 days (second column) and 10 days (third column) for each variable. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. For a well calibrated ensemble, the scattered points should lie on the one-to-one diagonal shown in black.

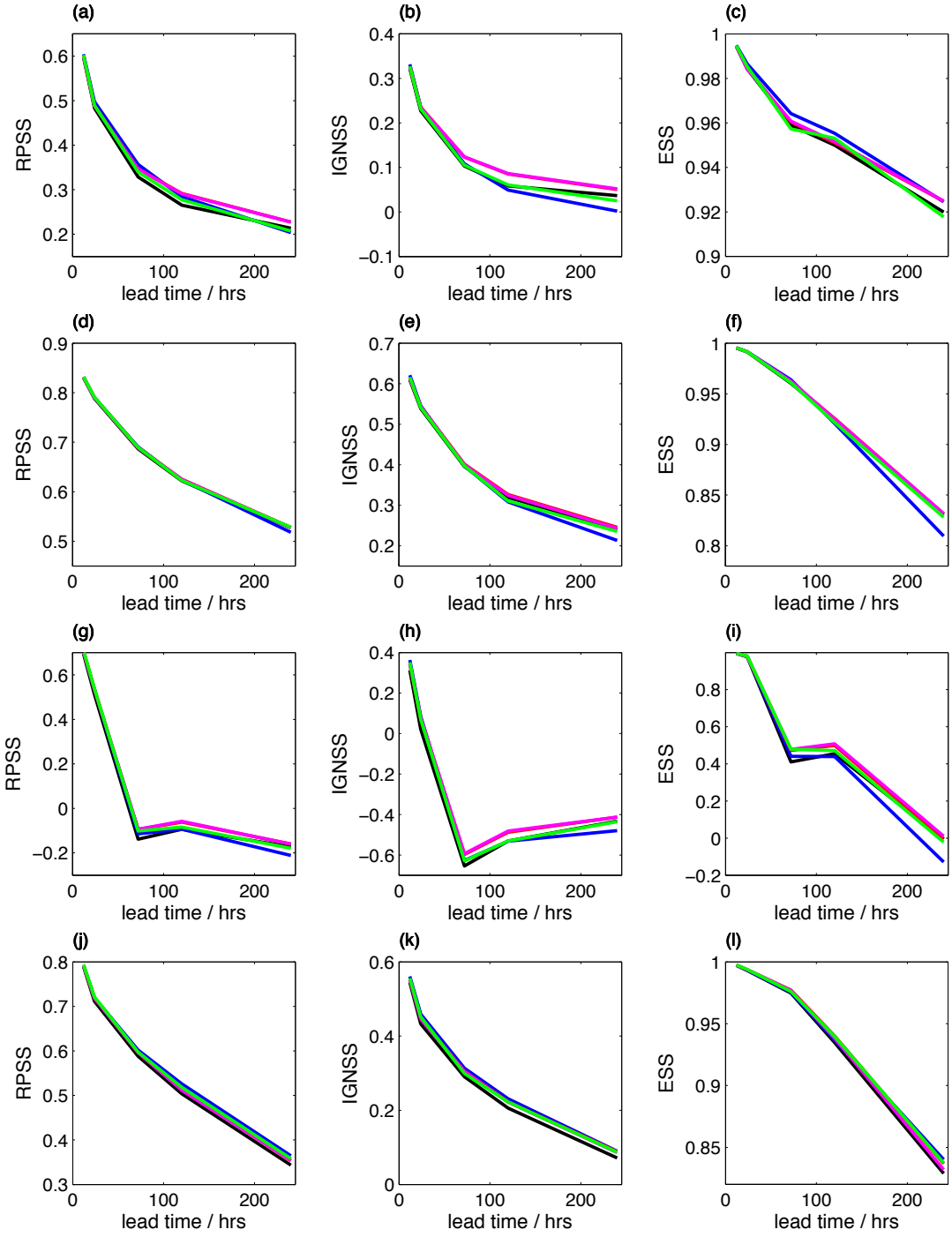

Figure 5.8: Ensemble forecast skill scores calculated for tropical regions with **significant convection**. First column: Ranked Probability Skill Score. Second column: Ignorance Skill Score. Third column: Error-spread Skill Score. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The red line is obscured by the magenta line in each figure.

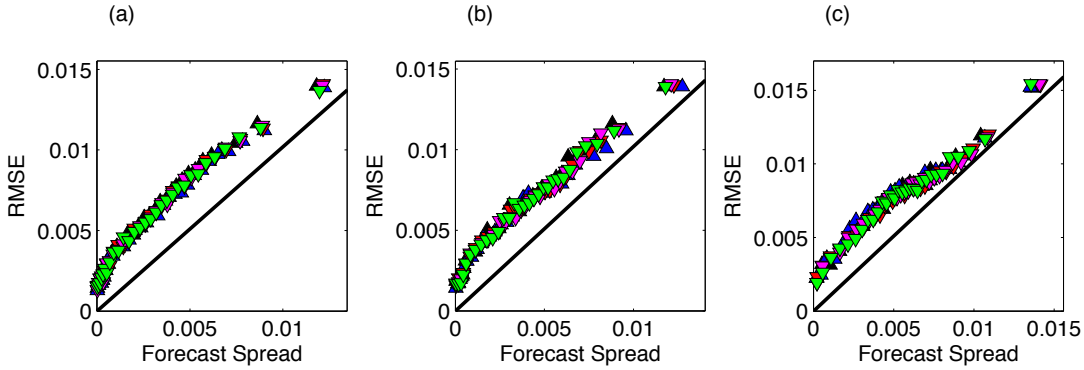

Figure 5.9: RMS error-spread diagnostic for cumulative convective precipitation for the 24 hour window before a lead time of (a) 1 day, (b) 3 days and (c) 10 days. The diagnostic is calculated for tropical regions with **significant convection**. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. For a well calibrated ensemble, the scattered points should lie on the one-to-one diagonal shown in black.

are under-dispersive, and the different uncertainty schemes have only a slight impact on the calibration of the ensemble. Figure 5.10(b) indicates more clearly the impact of the different schemes on the ensemble spread and error. On average, the TSCZ scheme is significantly the most under-dispersive and has a significantly larger RMSE. The two stochastic schemes, TSCS and TSCPv, have significantly the smallest error. TSCS has significantly the largest spread at short lead times, and TSCP and TSCPr have significantly the largest spread at later lead times. Figure 5.10(a) shows the bias in forecasts of convective precipitation. The stochastic schemes, TSCS and TSCPv, have the smallest bias over the entire forecasting window. Figures 5.10(c)–(e) show the forecast skill scores for convective precipitation. TSCS is significantly the best between days three and five according to RPS and ES. TSCZ is significantly the poorest according to RPS, but the other schemes score very similarly. ES and IGN also score TSCZ as significantly the worst at early lead times, but at later lead times, no one scheme is significantly different to the others.

It is important for a model to capture the spatial and temporal characteristics of precipitation. The global frequency distribution of rain rate (in mm/day) was considered for the different forecast models and compared to the GPCP 1DD dataset. The results are shown in Figure 5.11. All five forecast models perform similarly well, and no one model performs particularly well or poorly compared to the others. All forecasts under-predict the proportion of low rain rates and over predict the proportion of high rain rates when compared to the GPCP data set (grey), but overall predict the distribution of rain rates well.

The spatial distribution of cumulative precipitation (convective plus large scale) was also

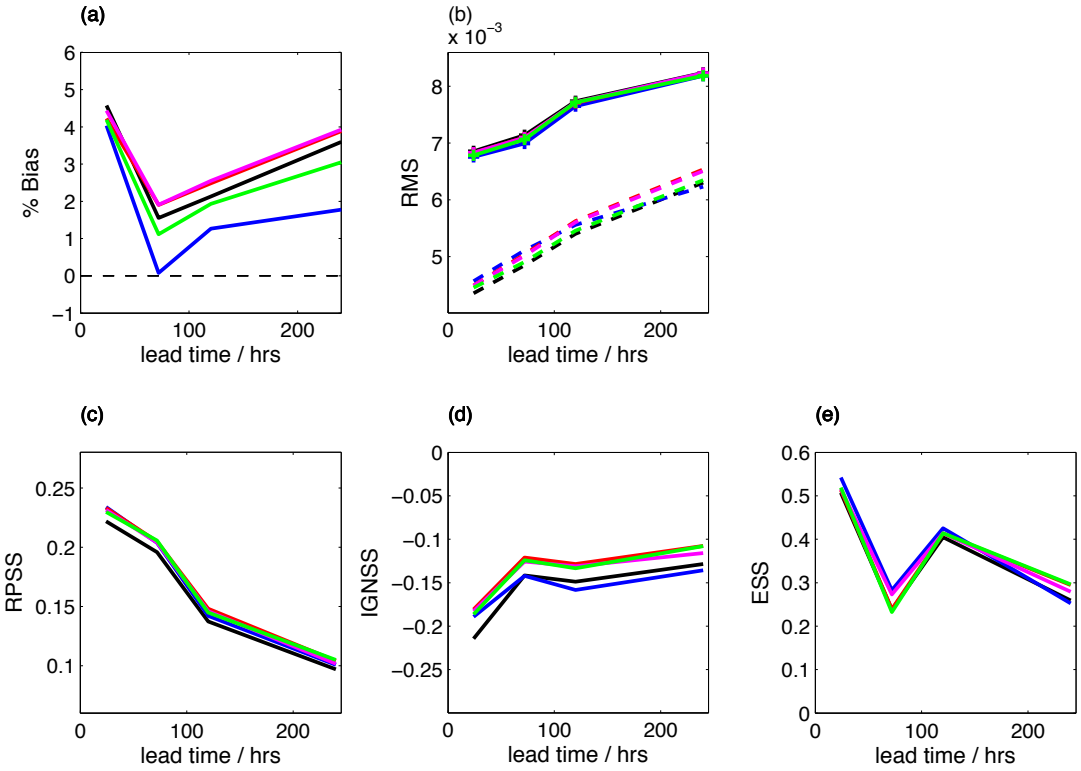

Figure 5.10: Summary forecast diagnostics for 24 hour cumulative convective precipitation (prior to the indicated lead time) in tropical regions with **significant convection**. (a) Percentage bias. (b) Temporal evolution of RMS ensemble spread (dashed lines) and error (solid lines) averaged over the region. (c) Ranked Probability Skill Score. (d) Ignorance Skill Score. (e) Error-spread Skill Score. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The red line is obscured by the magenta line in each figure.

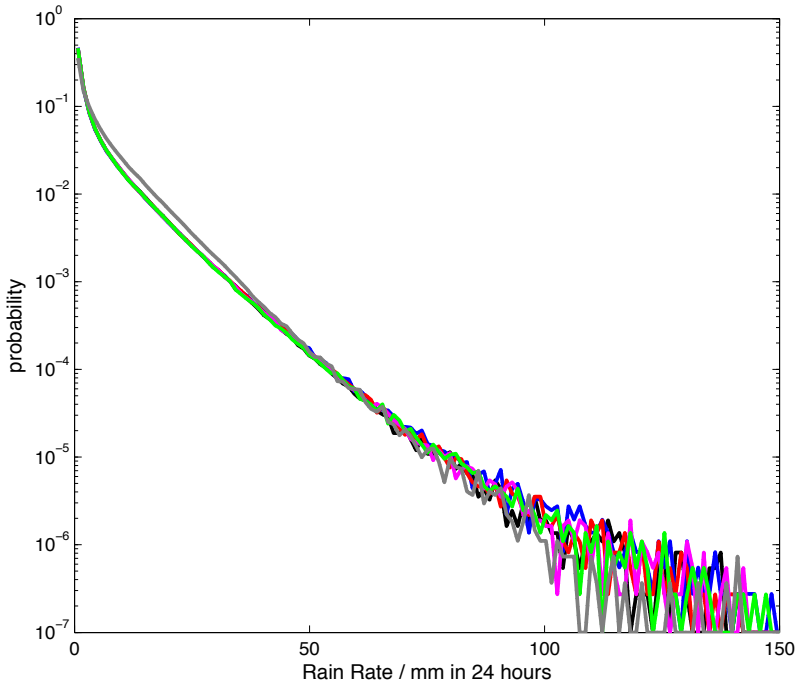

Figure 5.11: Probability distribution of rain rate (mm/24hrs) evaluated globally. The distribution has been normalised to 1, given that rain is observed in each 24 hour window. The observed result from the GPCP dataset (grey) is compared to the five experiments: TSCZ (solid black), TSCS (blue), TSCP (red), TSCPr (magenta) and TSCPv (green).

considered for the different forecast models. All schemes performed equally well (not shown). When compared to the GPCP data, all showed too much precipitation over the ocean, and in particular forecast intensities of rain in the intertropical and South Pacific convergence zones that were higher than observed. The results were indistinguishable by eye — the difference between forecast and observations is far greater than the differences between different forecasts.

5.6.2.2 Total Column Water Vapour

The impact of the different model uncertainty schemes on forecasts of total column water vapour (TCWV) is also a good indicator of improvement in the convection scheme. Figure 5.12 shows the RMS error-spread diagnostic for TCWV. The forecasts for this variable are poorly calibrated when compared to convective precipitation. The RMSE is systematically larger than the spread, and the slope of the scattered points is too shallow. This shallow slope indicates that the forecasting system is unable to distinguish between cases with low and high predictability for this variable — the expected error in the ensemble mean is poorly predicted by the ensemble spread. The different forecast schemes show a larger impact than for forecasts of precipitation — the TSCS model produces forecasts which are under-dispersive compared to the other forecasts. Figure 5.13 shows (a) the bias, (b) the RMSE and spread as a function of time, and (c)–(e) the forecast skill scores for each experiment. Figure (b) shows that the TSCPv forecasts have significantly the largest spread at lead times of 24 hours and greater. The TSCS forecasts have significantly the smallest spread at later lead times, but also significantly the largest error at all lead times. Figure 5.13 (a) shows the bias is also largest for the TSCS forecasts, and Figure 5.13 (c–e) indicates the skill is the lowest.

An early version of SPPT was found to dry out the tropics, and resulted in a decrease in TCWV of approximately 10% (Martin Leutbecher, pers. comm., 2013). This was corrected in a later version. It is possible that TCWV could be sensitive to the proposed perturbed parameter representations of model uncertainty. The average TCWV between 20°N and 20°S is averaged over all start dates separately for each ensemble member, and is shown in Figure 5.14. Initially, all experiments show a drying of the tropics of approximately $0.5\,\mathrm{kgm}^{-2}$ over the first 12 hours, indicating a spin-up period in the model. The TSCZ, TSCS and TSCPv forecasts then stabilise. *However, each ensemble member in the TSCP model has vastly different behaviour, with some showing systematic drying, and others showing systematic moistening over the ten day forecast.*

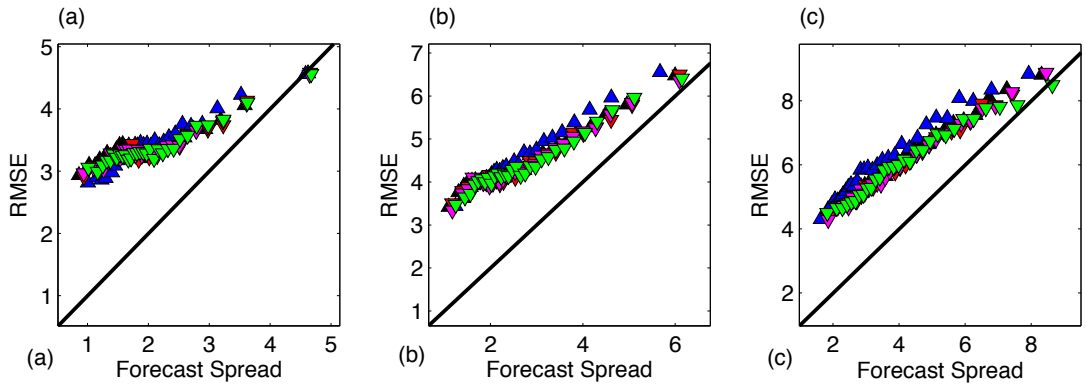

Figure 5.12: RMS error-spread diagnostic for total column water vapour for lead times of (a) 1 day, (b) 3 days and (c) 10 days. The diagnostic is calculated for tropical regions with **significant convection**. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. For a well calibrated ensemble, the scattered points should lie on the one-to-one diagonal shown in black.

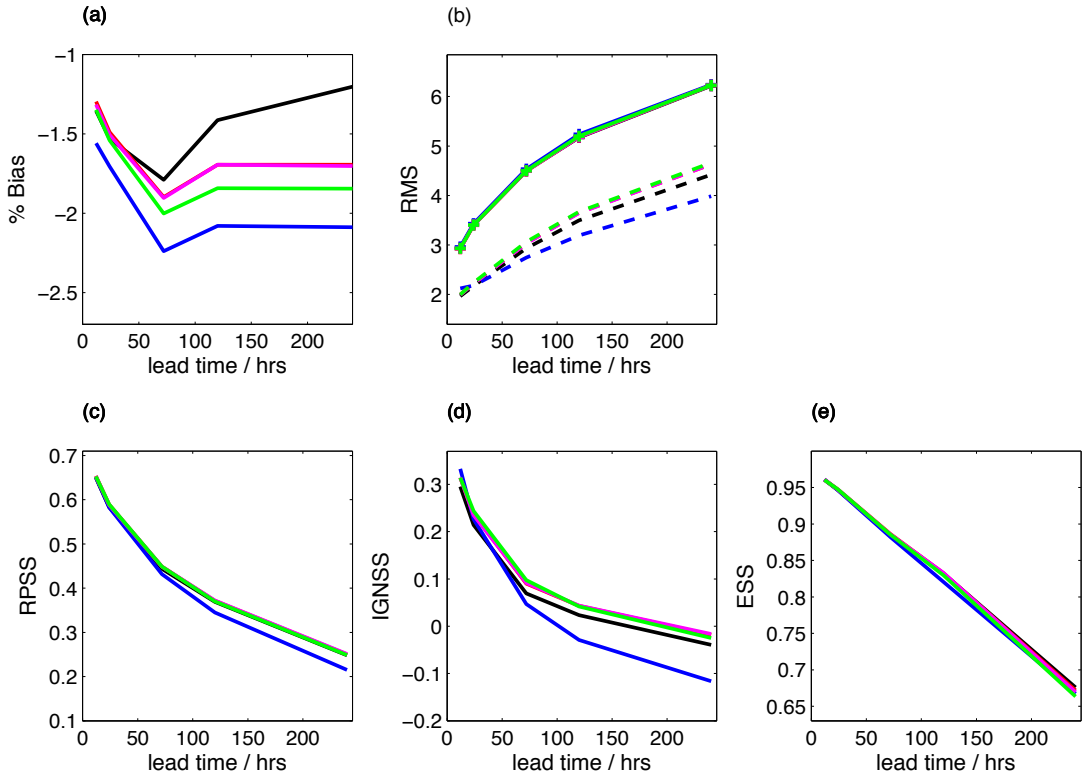

Figure 5.13: Summary forecast diagnostics for total column water vapour in tropical regions with **significant convection**. (a) Percentage bias. (b) Temporal evolution of RMS ensemble spread (dashed lines) and error (solid lines) averaged over the region. (c) Ranked Probability Skill Score (d) Ignorance Skill Score. (e) Error-spread Skill Score. Results are shown for the five experiments: black — TSCZ; blue — TSCS; red — TSCP; magenta — TSCPr; green — TSCPv. The red line is obscured by the magenta line in each figure.

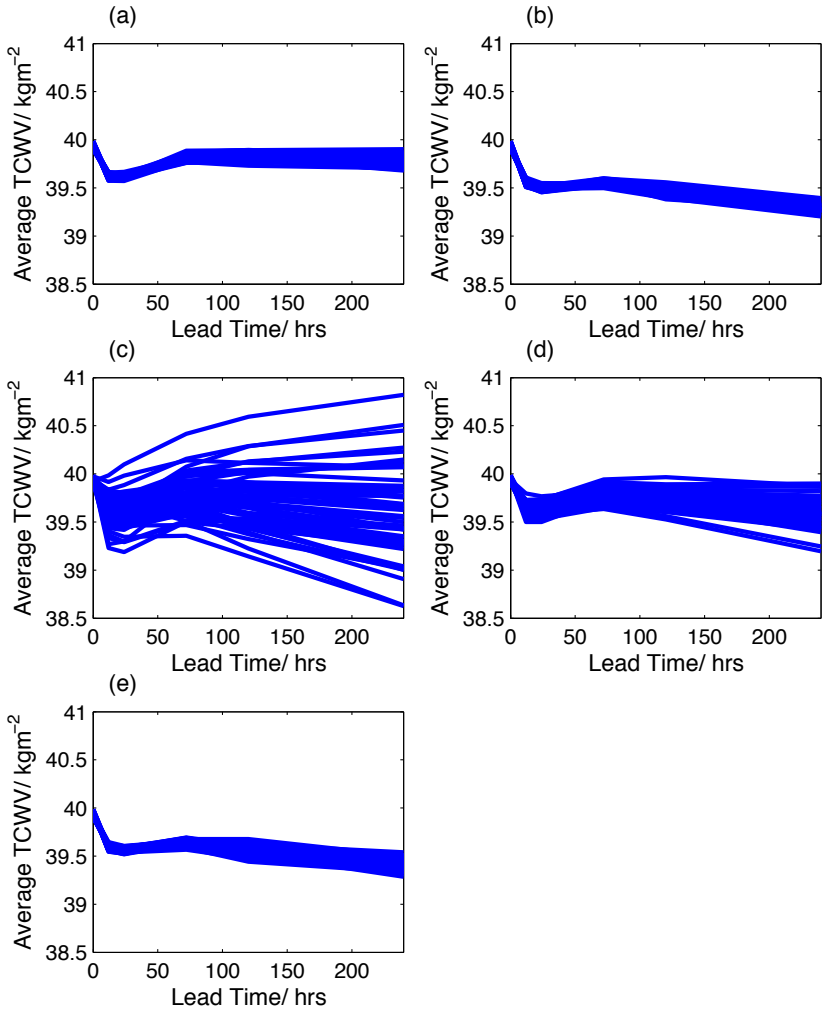

Figure 5.14: Average total column water vapour (TCWV) between 20°S and 20°N as a function of time. The spatial average is calculated for each ensemble member averaged over all start dates, and the averages for each of the fifty ensemble members are shown. Results are shown for the five experiments: (a) TSCZ, (b) TSCS, (c) TSCP, (d) TSCPr and (e) TSCPv.

The TSCPr model does not show this behaviour to the same extent. Figure 5.15 shows an alternative diagnostic. The TCWV is averaged over the region. The average and standard deviation of this diagnostic is calculated over all ensemble members and start dates. The average TCWV is similar for all experiments. The standard deviation initially decreases for all experiments. However, at longer lead times, the standard deviation increases for both TSCP and TSCPr, indicating differing trends in TCWV for different ensemble members for both experiments.

5.7 Discussion and Conclusion

The results presented above show that the perturbed parameter schemes have a positive impact on the IFS, though the impact is relatively small. Introducing the TSCP/TSCPr schemes

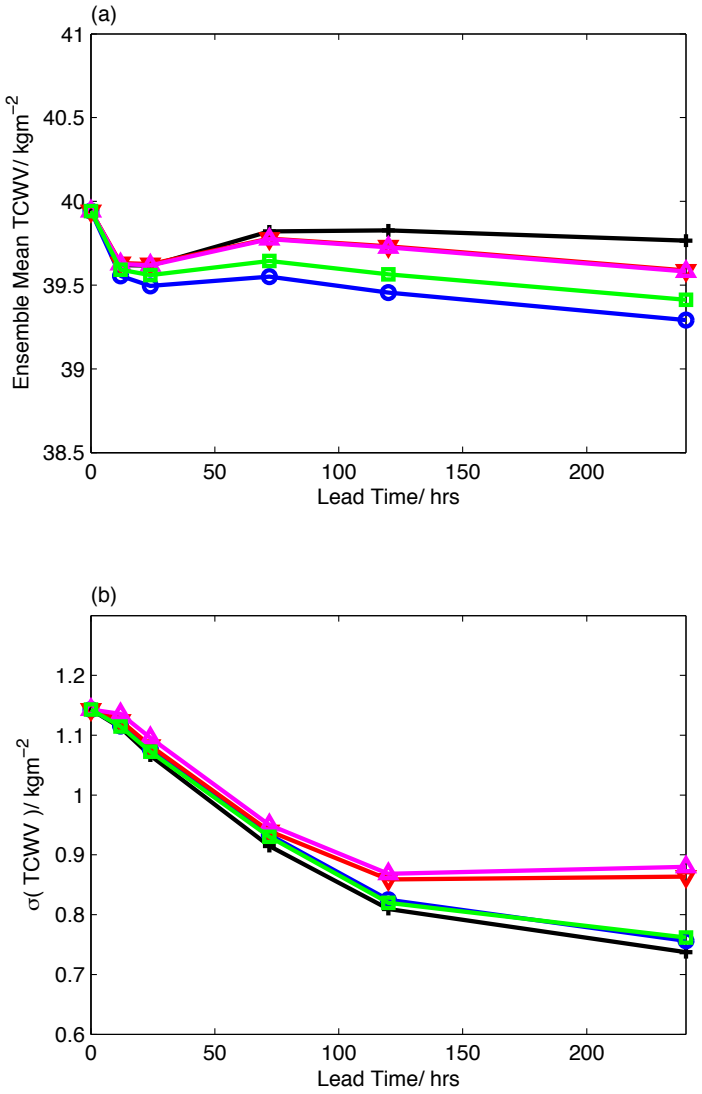

Figure 5.15: The total column water vapour (TCWV) is averaged between 20°S and 20°N as a function of time. The (a) mean and (b) standard deviation are then calculated over all ensemble members and start dates. Results are shown for the five experiments: black — TSCZ, blue — TSCS, red — TSCP, magenta — TSCPr and green — TSCV.

(defined in Table 5.3) does not lead to increased bias in T850, U850, Z500 or U200, indicating that systematic errors in these fields have not increased. An increase in ensemble spread is observed when the perturbed parameter schemes are used to represent uncertainty in convection instead of SPPT, and the TSCP/TSCPr forecasts have significantly the largest spread for T850, U850 and Z500 forecasts, which Figure 5.7 indicates is flow-dependent. The perturbed parameter schemes produce significantly the most skilful forecasts of T850, U850 and Z500 as ranked by the RPSS, IGNSS and ESS.

These results indicate that using a fixed perturbed parameter ensemble instead of SPPT improves the representation of uncertainty in convection. However, the fixed perturbed parameter ensembles remain under-dispersive. While an increase in spread is observed when the perturbed parameter schemes are used to represent uncertainty in convection compared to SPPT, a substantial proportion of this increase is also observed in Figure 5.6 when SPPT is switched off for the convection scheme (this counter-intuitive result is analysed in Chapter 6). Since SPPT is switched off for convection in TSCP and TSCPr, this indicates that *the parameter perturbations are contributing only slightly to the spread of the ensemble, and much of the spread increase can be attributed to this decoupling of the convection scheme from SPPT* (see Section 6.5 for further experiments which confirm this "decoupling" hypothesis). The small impact of the perturbed parameter scheme indicates that such schemes are not fully capturing the uncertainty in the convection scheme at weather forecasting time scales. This is surprising as the parameter uncertainty has been explicitly measured and used to develop the scheme.

The TSCPv scheme had a positive impact on the skill of the weather forecasts, and significantly improved over the TSCZ and TSCS forecasts for many diagnostics. The impact on spread and skill was smaller than the static perturbed parameter schemes. It is possible that the parameter perturbations vary on too fast a time scale for a significant impact to be observed — if the parameters varied more slowly, a larger, cumulative effect could be observed in the forecasts. It would be interesting to test the TSCPv scheme using a longer correlation time scale to test this hypothesis.

The two types of perturbed parameter scheme presented here represent fundamentally different error models. Fixed perturbed parameter schemes are based on the ansatz that there exists some optimal (or "correct") value of the parameters in the deterministic parametrisation scheme. Even using EPPES, the optimal parameters cannot be known with certainty,

so a perturbed parameter ensemble samples from a set of likely parameter values. The fixed perturbed parameter ensembles tested in this chapter were under-dispersive, and did not fully capture the uncertainty in the forecasts. *This indicates that fixed parameter uncertainty is not the only source of model uncertainty, and that fixed perturbed parameter ensembles cannot be used alone to represent model uncertainty in an atmospheric simulation.* While parameter uncertainty could account for systematic errors in the forecast, the results indicate that some component of the error cannot be captured by a deterministic uncertainty scheme. In particular, perturbed parameter ensembles are unable to represent structural uncertainty due to the choices made when developing the parametrisation scheme, and a different approach is required to represent uncertainties due to the bulk formula assumption.

The second error model recognises that in atmospheric modelling there is not necessarily a "correct" value for the parameters in the physics parametrisation schemes. Instead there exists some optimal distribution of the parameters in a physical scheme. Since in many cases the parameters in the physics schemes have no direct physical interpretation, but represent a group of interacting processes, it is likely that their optimal value may vary from day to day, or from grid box to grid box, or on larger scales they may be seasonally or latitudinally dependent. A stochastically perturbed parameter ensemble represents this parameter uncertainty. The stochastically perturbed parameter scheme also underestimated the error in the forecasts. *Even generalised to allow varying parameters, parameter uncertainty is not the only source of model uncertainty in weather forecasts.* Not all sub-grid scale processes can be accurately represented using a statistical parametrisation scheme, and some forecast errors cannot be represented using the phase space of the parametrised tendencies.

The EPPES indicated that the uncertainty in the convection parameters was moderate, and smaller than expected (Heikki Järvinen, pers. comm., 2013). The results presented here also indicate larger parameter perturbations could be necessary to capture the uncertainty in the forecast from the convection scheme. However, the average tropical total column water vapour indicates that even these moderate perturbations are sufficient for biases to develop in this field over the ten day forecast period. The ensemble members with different sets of parameters have vastly different behaviours, with some showing a systematic drying and others a systematic moistening in this region. This is very concerning. The second diagnostic presented indicates that TSCPr also has the problem of systematic drying or moistening for individual ensemble

members depending on the model parameters, and suggests that this is a fundamental problem with using a fixed perturbed parameter ensemble. The fact that this problem develops noticeably over a ten day window indicates that this could be a serious problem in climate prediction, where longer forecasts could result in even larger biases developing. This result supports the conclusions made in Chapter 3 in the context of L96, where individual perturbed parameter ensemble members were observed to have vastly different regime behaviour. The TSCPv forecasts did not develop biases in this way, as the parameter sets for each ensemble member varied over the course of the forecast, which did not allow these biases to develop. Therefore, stochastically varying perturbed parameter ensembles could be an attractive way of including parameter uncertainty into weather and climate forecasts.

A particularly interesting and counter-intuitive result is that *removing the stochastic perturbations from the convection tendency resulted in an increase in forecast spread for some variables*. This is observed for T850, U850 and TCWV in both regions considered, and for Z500 and U200 in non-convecting regions. SPPT perturbs the sum of the physics tendencies. It does not represent uncertainty in individual tendencies, but assumes uncertainty is proportional to the total tendency. The increase in spread for TSCZ forecasts compared to TSCS forecasts suggests that convection could act to reduce the sum of the tendencies, resulting in a smaller SPPT perturbation. This is as expected from the formulation of the parametrisation schemes in the IFS, and will be discussed further in Section 6.1. Perturbing each physics tendency independently would allow for an estimation of the uncertainty in each physics scheme, potentially improving the overall representation of model uncertainty. This is the subject of the next chapter.

Despite the reduced spread, the TSCS scheme outperforms the TSCZ scheme according to other forecast diagnostics. The error in T850 forecasts is reduced using the TSCS scheme, reflected by higher skill scores for this variable, and TSCS is significantly more skilful than TSCZ at lead times of up to 3 days for U850 and Z500. Additionally, TSCS results in an *increase* of spread for U200 and convective precipitation compared to TSCZ. At this point, it is important to remember that the parametrisation tendencies are not scalar quantities, but are vectors of values corresponding to the tendency at different vertical levels, and that SPPT uses the same stochastic perturbation field at each vertical level[3]. The convection parametrisation

[3]The perturbation is constant vertically except for tapering in the boundary layer and the stratosphere.

scheme is sensitive to the vertical distribution of temperature and humidity, and it is possible that the tendencies output by the convection parametrisation scheme act to damp or excite the scheme at subsequent time steps. Therefore perturbing the convective (vector) tendency using SPPT could lead to an increased variability in convective activity between ensemble members through amplification of this excitation process. Since both U200 and convective precipitation are directly sensitive to the convection parametrisation scheme, these variables are able to detect this increased variability, and show an increased ensemble spread as a result. In fact, TSCS has significantly the most skilful forecasts out of all five experiments between days three and ten for U200, and between days three and five for convective precipitation. T850, U850 and Z500 are less sensitive to convection than U200 and precipitation. Since in general the total perturbed tendency is reduced for TSCS compared to TSCZ, this could lead to the reduction in ensemble spread observed for these variables.

The experiments presented in this chapter have used the IFS at a resolution of T159. This is significantly lower than the operational resolution of T639. Nevertheless, the experiments give a good indication of what the impact of the different schemes would be on the skill of the operational resolution IFS. In Chapter 6, results are presented for T159 experiments which were repeated at T639; the same trends can be observed at the higher resolution. Therefore, the low resolution runs presented in this chapter can be used to indicate the expected results of the models at T639, and can suggest whether it would be interesting to run further experiments at higher resolution.

6

Experiments in the IFS: Independent SPPT

The only relevant test of the validity of a hypothesis is comparison of prediction with experience.

– Milton Friedman, 1953

6.1 Motivation

The generalised stochastically perturbed parametrisation tendencies (SPPT) scheme developed in the previous chapter allowed the SPPT perturbation to be switched off for the convection tendency and replaced with a perturbed parameter scheme. However, it also enables one to perturb the five IFS physics schemes with independent random fields. In the operational SPPT, the uncertainty is assumed to be proportional to the total net tendency, whereas this generalisation to SPPT assumes that the errors from the different parametrisation schemes are uncorrelated, and that the uncertainty in the forecast is proportional to the individual tendencies.

The standard deviation of the perturbed tendency, σ_{tend}, using operational SPPT is given by

$$\sigma_{tend}^2 = \sigma_n^2 \left(\sum_{i=1}^{5} P_i \right)^2 , \tag{6.1}$$

where σ_n is the standard deviation of the noise perturbation and P_i is the parametrised tendency from the ith physics scheme. This can be compared to the standard deviation using

independent SPPT (SPPTi):

$$\sigma^2_{tend} = \sum_{i=1}^{5} \left(\sigma_i^2 P_i^2 \right). \tag{6.2}$$

If the physics tendencies tend to act in opposite directions, SPPTi will acknowledge the large uncertainty in the individual tendencies and will increase the forecast ensemble spread.

A priori, it is not known whether SPPT or SPPTi is the more physically plausible error model for the IFS, though it is very unlikely that uncertainties in the different processes are precisely correlated, as modelled by SPPT. However, the different physics schemes in the IFS have been developed in tandem, and are called sequentially in the IFS to maintain a balance. For example, the cloud and convection schemes model two halves of the same set of atmospheric processes, as described in Section 5.2.1. The convection parametrisation scheme represents the warming due to moist convection, but the cloud scheme calculates the cooling due to evaporation of cloudy air that has been detrained from the convective plume. This means that the net tendency from the two schemes is smaller than each individual tendency, and SPPT represents a correspondingly small level of uncertainty. SPPTi could be beneficial in this case, as it is able to represent the potentially large errors in each individual tendency. On the other hand, if the two schemes have been closely tuned to each other, potentially with compensating errors, decoupling the two schemes by using independent perturbations could reduce the forecast skill and introduce errors into the forecasts.

The impact of using SPPTi in the IFS will be tested in this chapter. As a first attempt, each independent random number field will have the same characteristics as the field used operationally in SPPT, i.e. each of the five independent fields is itself a composite of three independent fields with differing temporal and spatial correlations and magnitudes (see Section 5.3). A series of experiments is carried out in the IFS. Four experiments are considered to investigate the impact of the SPPTi scheme, and are detailed in Table 6.1. The impact of the operational SKEB scheme is also considered as a benchmark. The same resolution, start dates and lead time were used as in Chapter 5.

Firstly, the impact of SPPTi on global diagnostics is presented in Section 6.2. In Section 6.3, the impact of the SPPTi scheme in the tropics is considered, including differences between behaviour of the scheme in the tropical regions with significant and little convection which were defined in Chapter 5. The impact of the scheme on the skill of convection diagnostics is presented in Section 6.4. A series of experiments are described in Section 6.5, which

Experiment Abbreviation	SPPT	SPPTi	SKEB
TSCS	ON	OFF	OFF
TSCS + SKEB	ON	OFF	ON
TSCSi	OFF	ON	OFF
TSCSi + SKEB	OFF	ON	ON

Table 6.1: The four experiments and their abbreviations considered in this chapter to investigate the impact of independent SPPT over operational SPPT. The impact of the SKEB scheme is also considered for comparison. See Table 5.3 for an explanation of the abbreviations.

aim to increase understanding of the mechanisms by which SPPTi impacts the ensemble. In Section 6.6, the results from experiments at operational T639 resolution are presented. In Section 6.7, the results are discussed and some conclusions are drawn.

6.2 Global Diagnostics

The representations of model uncertainty considered in Chapter 5 focused on the convection scheme only, so the different schemes were verified in the tropics where convection is most active. The SPPTi scheme discussed in this chapter affects all parametrisation schemes, so has a global impact. It is therefore important to evaluate the impact of the scheme on global forecasts. The global bias, calculated following (1.24) is shown as a percentage in Figure 6.1. Globally, the bias is small for each variable considered. However, for all variables, implementing the SPPTi scheme results in an increase in global bias. The impact is particularly large for U200, where the global bias has more than doubled in magnitude. The impact of SKEB (comparing blue with cyan, and red with magenta) is considerably smaller than the impact of the SPPTi scheme (comparing the warm and cool pairs of lines).

Figure 6.2 shows the temporal evolution of the RMS ensemble spread and error respectively, for each experiment considered, averaged over each standard ECMWF region: the northern extra-tropics is defined as north of $25°N$, the southern extra-tropics is defined as south of $25°S$, and the tropics is defined as $25°S - 25°N$. In the northern and southern extra-tropics (first and third columns respectively), the impact of SPPTi on the ensemble spread is comparable to, but slightly smaller than, the impact of SKEB, and all experiments have under-dispersive forecasts. The different schemes have little impact on the RMSE. However in the tropics (centre column), *SPPTi has a significant positive impact on the spread of the ensemble forecasts.* The impact is significantly larger than that of SKEB, and corrects the under-dispersive forecasts for

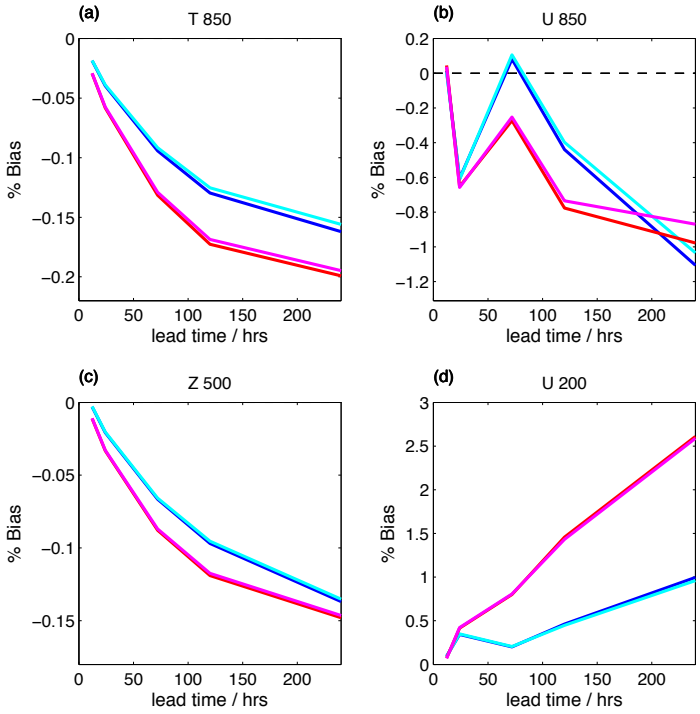

Figure 6.1: Global forecast bias as a function of time for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB. The bias is calculated as described in the text, and given as a percentage of the root mean square of the verification in the region of interest.

T850, U850 and U200. While SPPTi has a larger impact on the spread of Z500 forecasts than SKEB, the ensembles remain under-dispersive. A small impact on the RMSE is observed — the T850 and Z500 errors are slightly increased and the U850 and U200 errors slightly reduced by the SPPTi scheme. These results are very positive, and indicate the potential of the SPPTi scheme.

Figure 6.2 indicates that SPPTi has the largest impact in the tropics. Figure 6.3 shows the skill of the forecasts in this region evaluated using the RPSS, IGNSS and ESS for the four variables of interest. IGNSS indicates an improvement of skill for all variables when SPPTi is implemented. This is as expected from Figure 6.2: IGNSS strongly penalises under-dispersive ensemble forecasts, so reducing the degree of under-dispersion results in an improved score. RPSS and ESS indicate a slight improvement in skill for the U850 and U200 forecasts, but a slight reduction in skill for the T850 and Z500 forecasts when the SPPTi scheme is used. This could be due to the increase in root mean square error observed for these variables, linked to the increase in bias observed in Figure 6.1. The IFS is such a highly tuned forecast model that it would be very surprising if a newly proposed scheme resulted in an improvement in skill for all variables in all areas. Before operationally implementing a new scheme, the scheme would

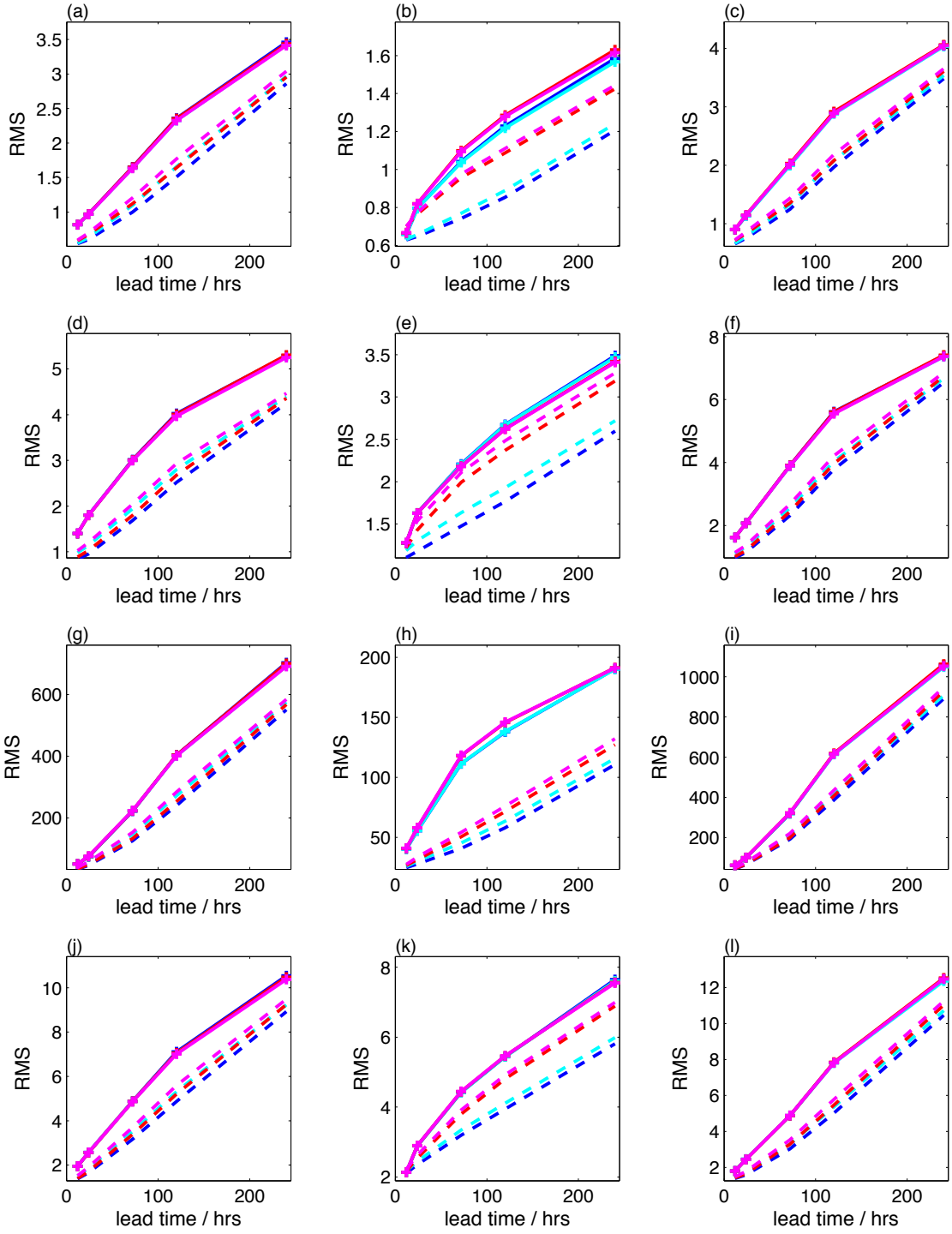

Figure 6.2: Temporal evolution of the RMS ensemble spread (dashed lines) and RMSE (solid lines) for each standard ECMWF region. First column: northern extra-tropics, north of 25N. Second column: tropics, 25S–25N. Third column: southern extra-tropics, south of 25S. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB.

need to be tuned, and the model re-calibrated to account for the effects of the new scheme. However, the significant improvement in spread observed in the tropics is sufficient to merit further investigation.

6.3 Effect of Independent SPPT in Tropical Areas

What is the cause of the improved spread in the tropics when the SPPTi scheme is implemented? As in Chapter 5, let us consider areas in the tropics where there is little convection, and areas where convection is the dominant process. The regions defined in Section 5.5.1 will be used as before.

The percentage bias as a function of lead time is shown for areas of little convection in Figure 6.4. The SPPTi forecasts have a larger bias at lead times greater than 24 hrs for T850 than for operational SPPT, and similar bias characteristics for Z500. However, U850 and U200 both show a reduction in the forecast bias when the SPPTi scheme is used compared to the operational SPPT scheme. Figure 6.5 shows the same diagnostic for regions with significant convection. The results are similar for T850 and Z500. U850 shows a slight improvement, but U200 indicates that SPPTi results in a small increase in bias. For operational SPPT, the negative bias in non-convecting regions cancels the positive bias in convecting regions to produce a small globally averaged bias. The SPPTi scheme has effectively reduced the negative bias in non-convecting regions, but has not had a large impact on the bias in convecting regions, resulting in the large increase in magnitude of the bias observed for global U200 in Figure 6.1(d). Considering regionally averaged bias can be misleading due to compensating errors.

Figures 6.6 and 6.7 show the evolution of the RMSE (solid lines) and RMS spread (dashed lines) for the tropical regions with little convection and with significant convection respectively. The operational SPPT ensembles (blue lines) are under-dispersive at all times, for all variables, in both regions. The under-dispersion is greater in regions with significant convection. Including SKEB (cyan lines) does not significantly increase the spread of the ensemble.

In regions with little convection, using SPPTi results in a moderately large correction to this under-dispersion, approximately halving the difference between spread and error for T850, U850 and U200 when compared to the operational runs. The impact is larger than the impact of including SKEB. The impact on spread is smaller for Z500, but still positive.

For regions with significant convection, the improvement in spread is greater than in regions

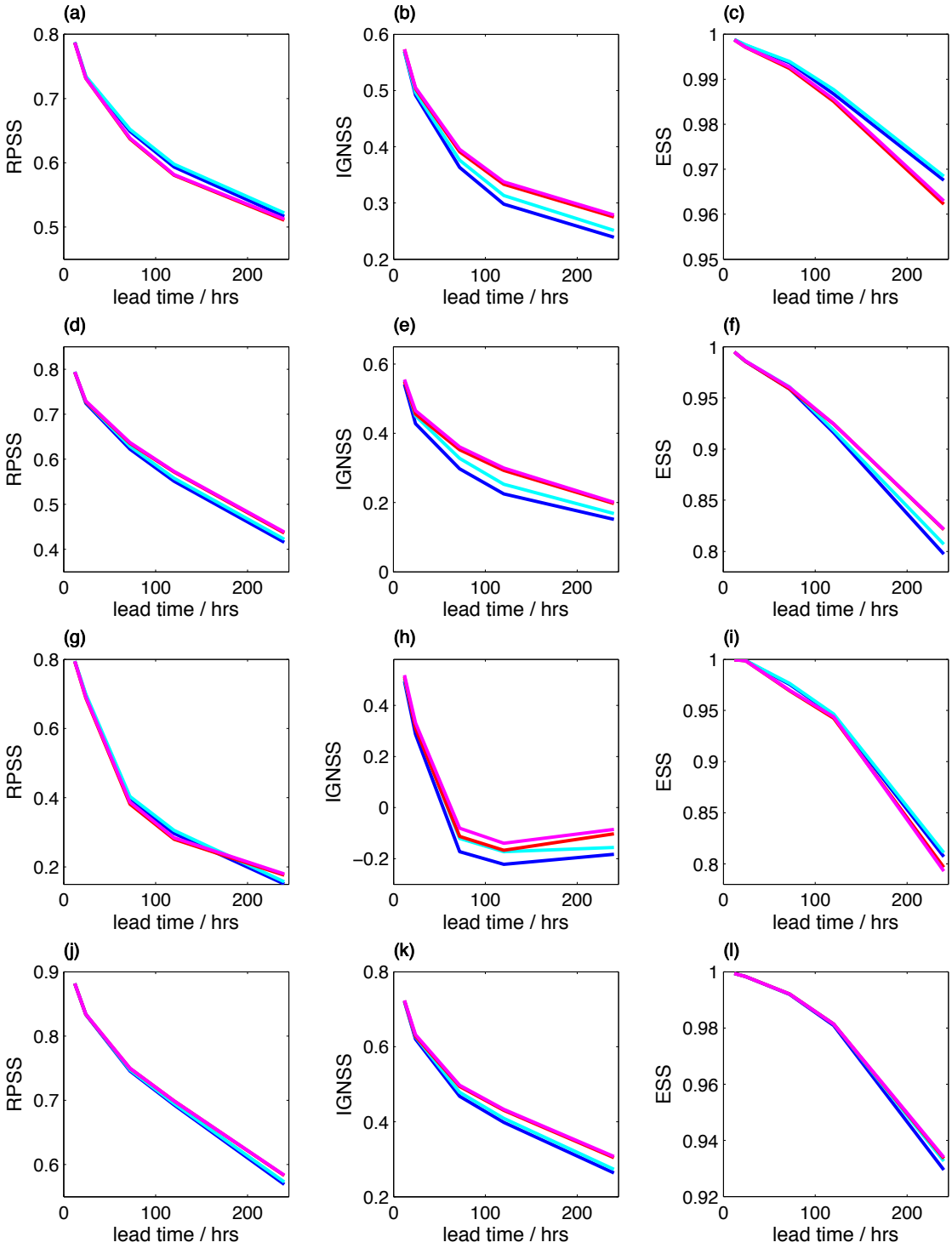

Figure 6.3: Ensemble forecast skill scores calculated for the **tropics** (25S–25N). First column: Ranked Probability Skill Score. Second column: Ignorance Skill Score. Third column: Error-spread Skill Score. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB.

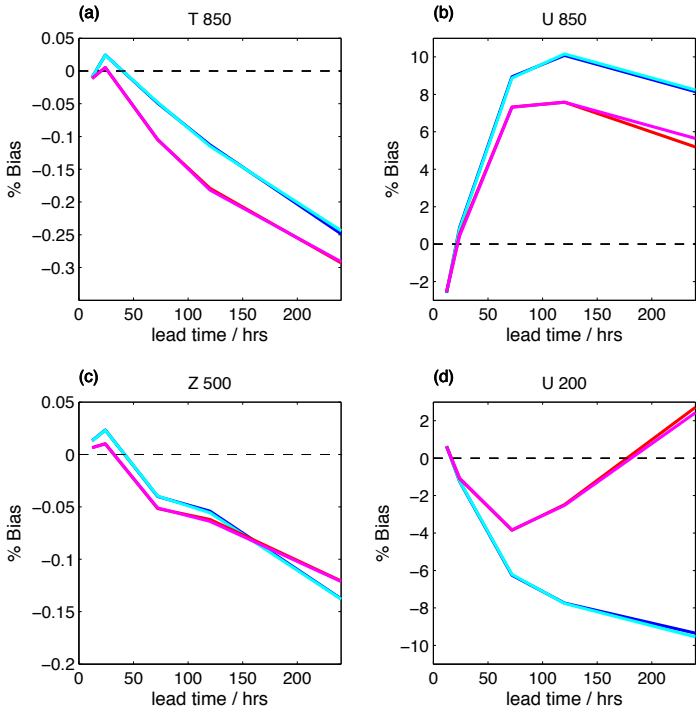

Figure 6.4: Percentage forecast bias in tropical regions with **little convection** as a function of time for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB. The bias is calculated as described in the text, and given as a percentage of the root mean square of the verification in the region of interest.

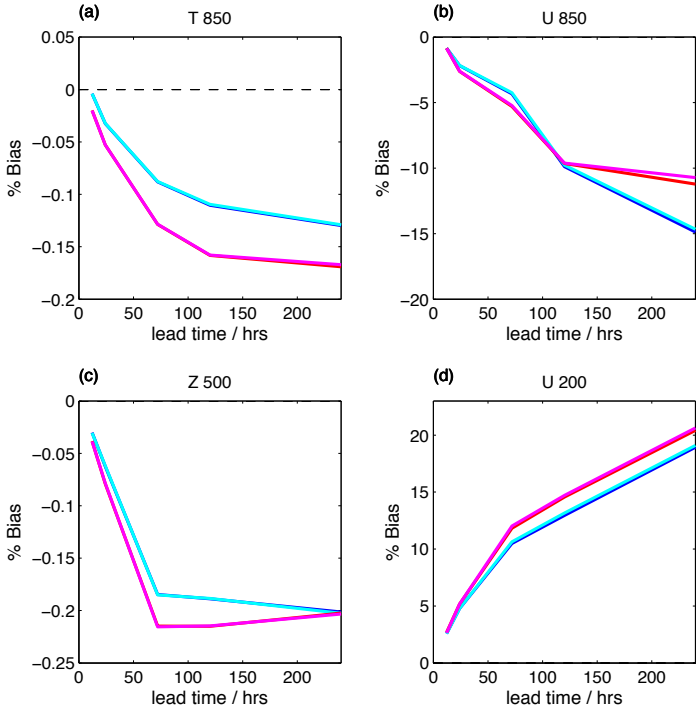

Figure 6.5: As for Figure 6.4, except for tropical regions with **significant convection**.

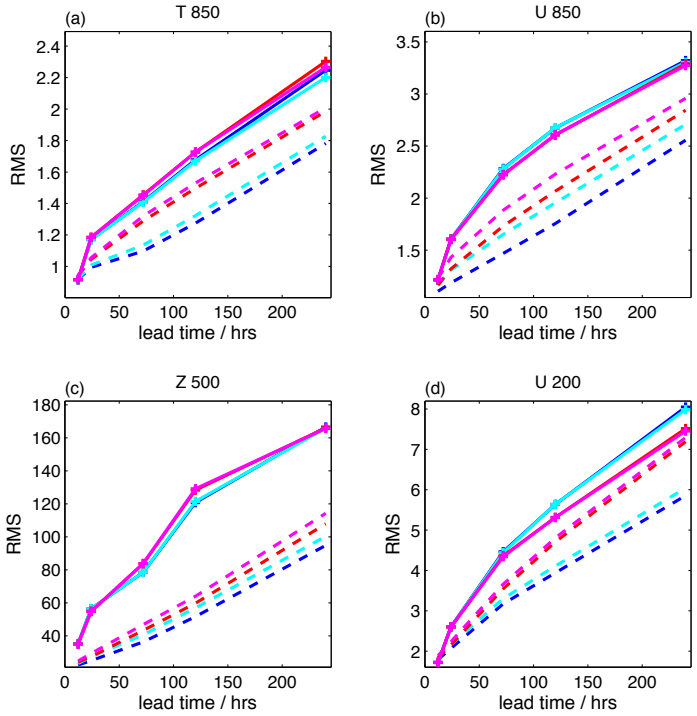

Figure 6.6: Temporal evolution of RMS ensemble spread (dashed lines) and RMSE (solid lines) for tropical regions with **little convection** for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB.

of little convection, whereas the improvement due to SKEB remains small. The spread of the ensembles closely matches the RMSE, and is slightly over-dispersive for U850. Moreover, *the temporal evolution of the spread has an improved profile for T850, U850 and U200.* For operational SPPT, the increase in spread is a fairly linear function of time, whereas for SPPTi, there is an initial period of rapid spread increase, followed by a reduction in rate, which closely matches the observed error growth. Figures 6.6 and 6.7 also indicate that *it is the convectively active regions that are primarily responsible for the observed increase in RMSE for T850 and Z500 for the SPPTi experiments.* This increase in error is a concern.

The RMS error-spread graphical diagnostic gives more information about the calibration of the forecast, testing whether the ensemble is able to skilfully indicate flow dependent uncertainty. Figures 6.8 and 6.9 show this diagnostic for tropical regions with little and significant convection respectively, for each variable of interest, at lead times of 1, 3 and 10 days. In both regions, Z500 is comparatively poorly forecast by the model. The error-spread relationship is weakly captured, and the ensemble spread is a poor predictor of the expected error in the ensemble mean. For the other variables, ensemble spread is a good predictor of RMSE in both regions.

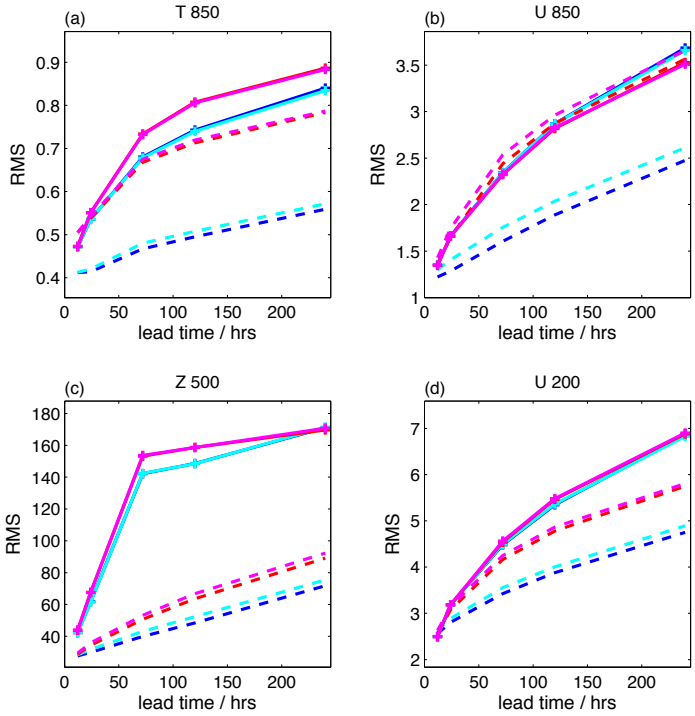

Figure 6.7: As for Figure 6.6, except for tropical regions with **significant convection**.

For the regions with little convection in Figure 6.8, the ensembles appear fairly well calibrated, and SPPTi has a small positive effect. The largest effect is seen at a lead time of 10 days for U200, where the operational SPPT ensemble was most under-dispersive. For the regions with significant convection shown in Figure 6.9, forecasts show a large improvement. *The increase in spread of the ensembles is state dependent: for most cases, the increase in spread results in an improved one-to-one relationship.* One exception is Figure 6.9(e), where the slope of the SPPTi scattered points is too shallow. It is interesting to note that the increase in RMSE for the T850 forecasts appears to occur predominantly for small error forecasts. This results in a flat tail to Figures 6.9(a)–(c) at small RMSE and spread, instead of a uniform increase in error across all forecasts. This tail appears to be unique to T850 out of the variables considered, visible at all lead times, and only in regions where convection dominates.

Figures 6.10 and 6.11 show the skill of the forecasts in regions of little and significant convection respectively, according to the RPSS, IGNSS and ESS. The skill scores for these two regions effectively summarise Figures 6.8 and 6.9, and provide more information as to the source of skill observed in 6.3. In particular:

- The poorer RPSS and ESS for SPPTi in the tropics for T850 (Figure 6.3) is mainly due to poorer forecast skill in convecting regions, though a small reduction in skill in non-convecting regions is also observed. The significant improvement in IGNSS in this region

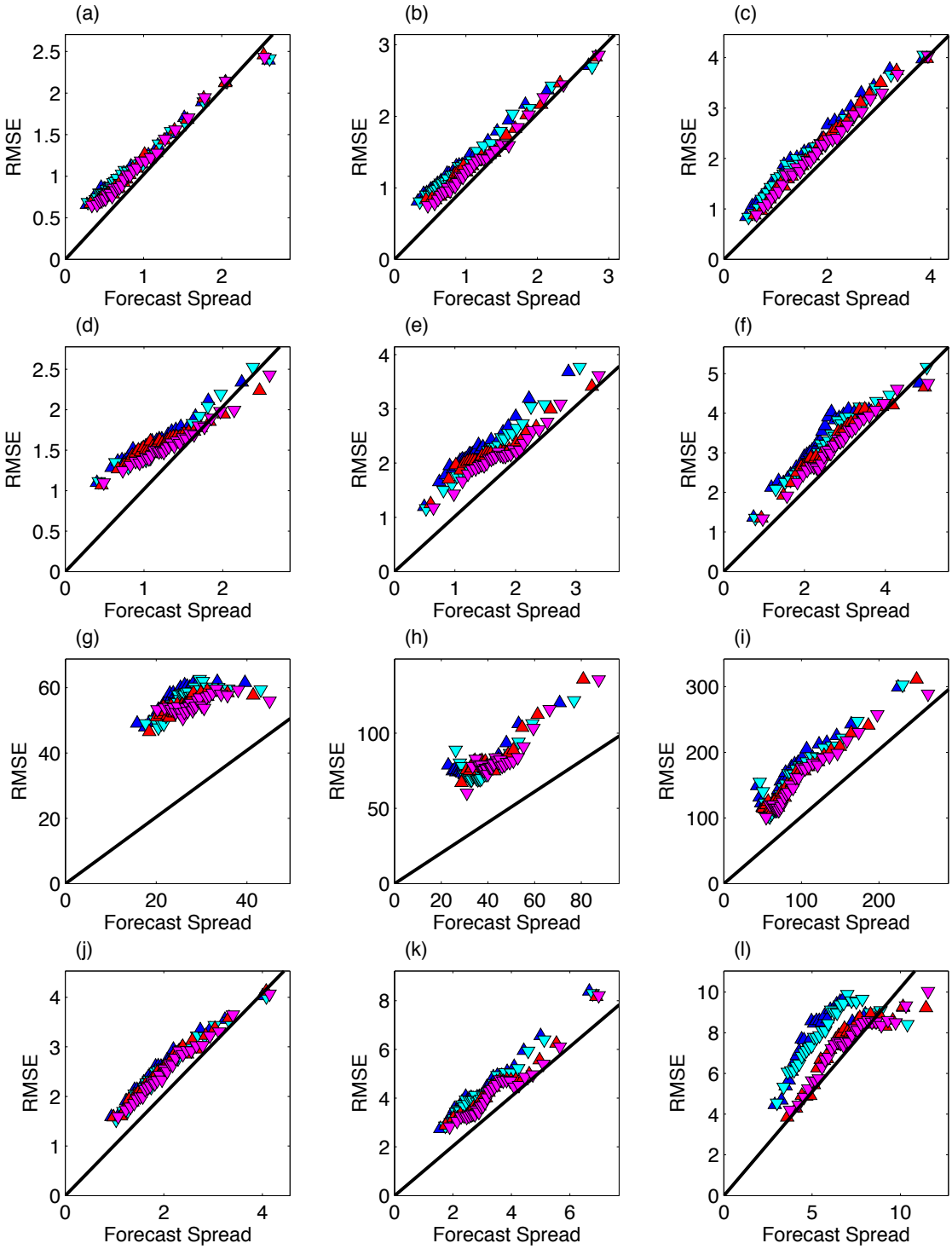

Figure 6.8: RMS error-spread diagnostic for tropical regions with **little convection** for (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200, at lead times of 1, 3 and 10 days for each variable (first, second and third columns respectively). Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB. The one-to-one diagonal is shown in black.

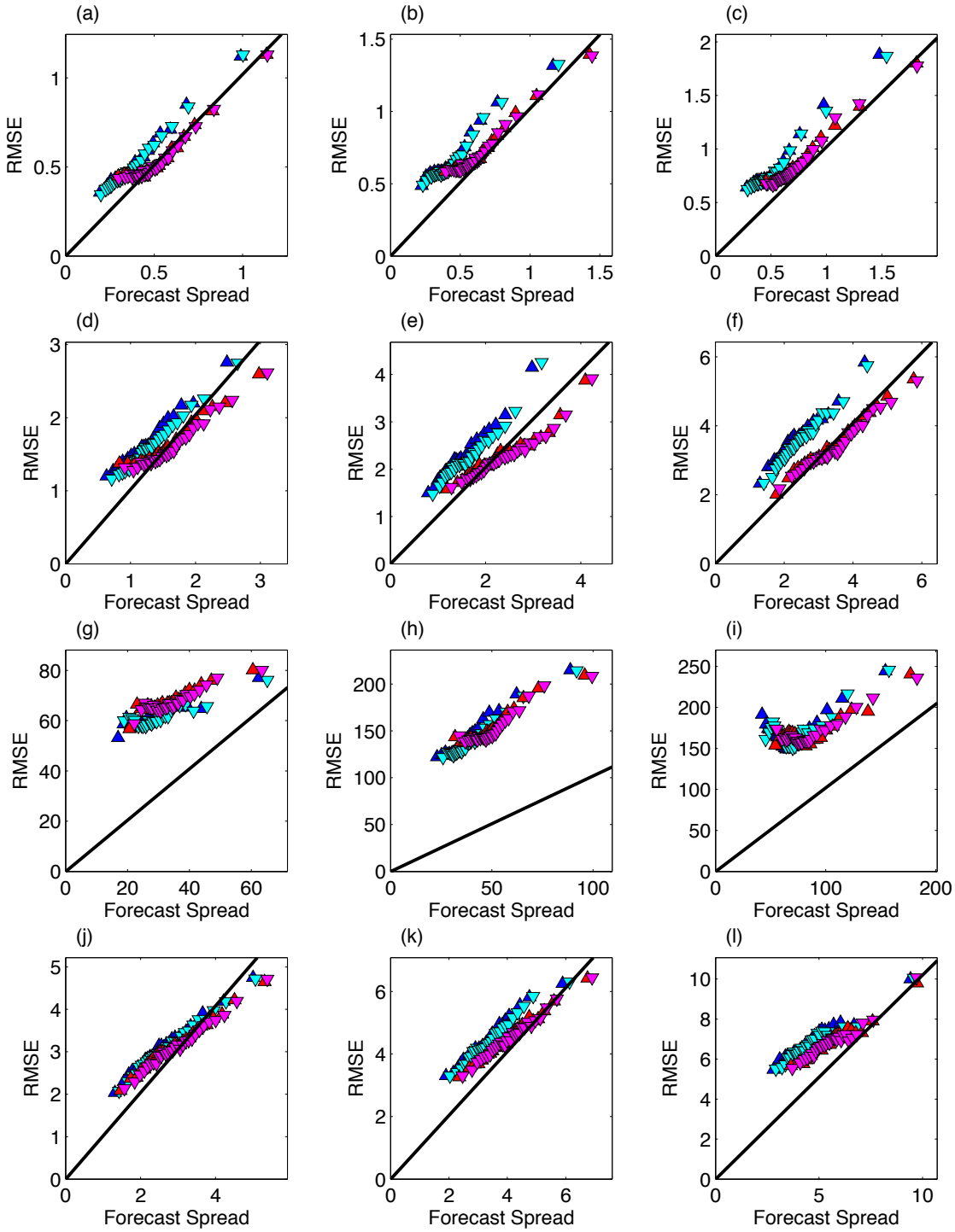

Figure 6.9: As for Figure 6.8, except for tropical regions with **significant convection**.

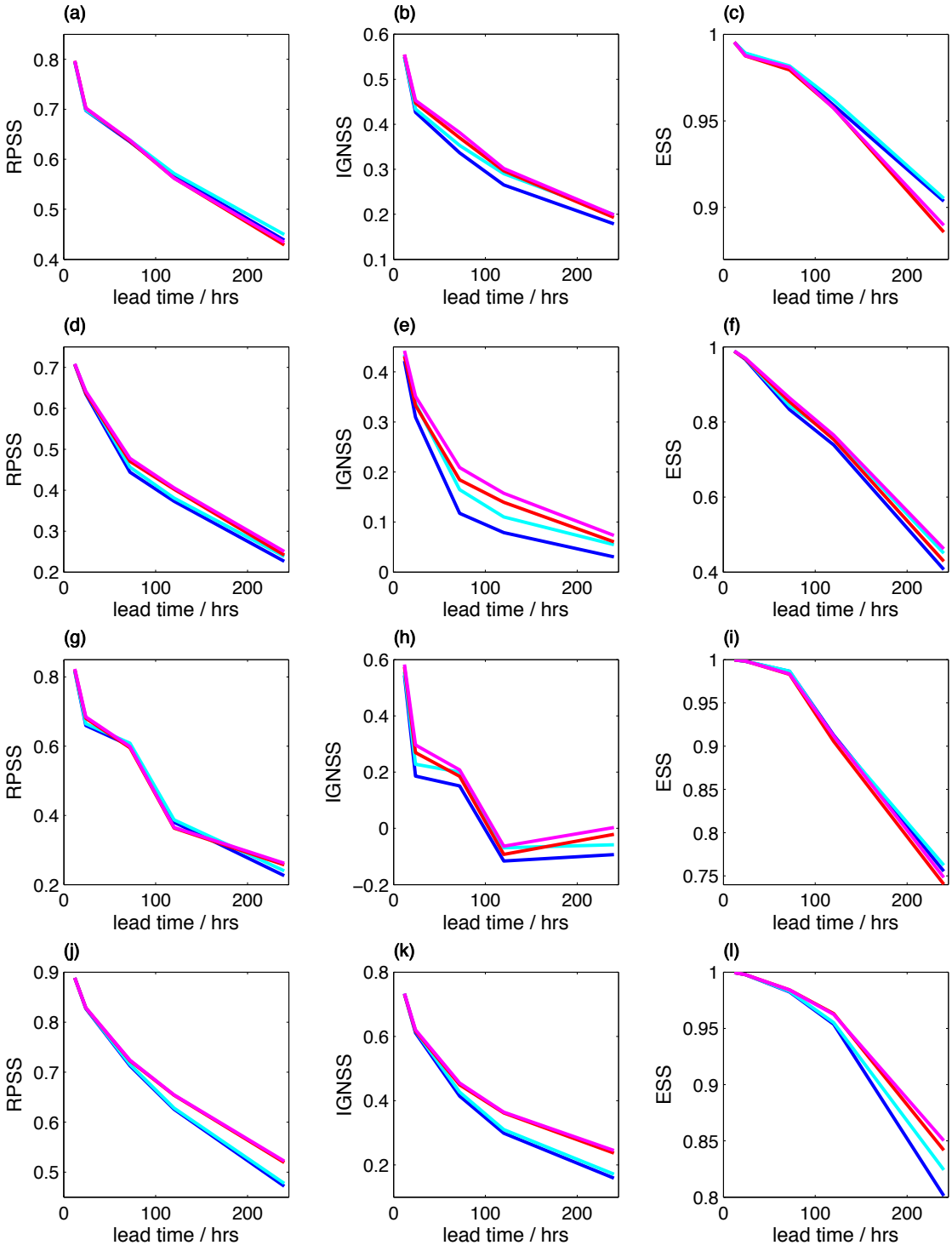

Figure 6.10: Ensemble forecast skill scores calculated for tropical regions with **little convection**. First column: Ranked Probability Skill Score. Second column: Ignorance Skill Score. Third column: Error-spread Skill Score. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB.

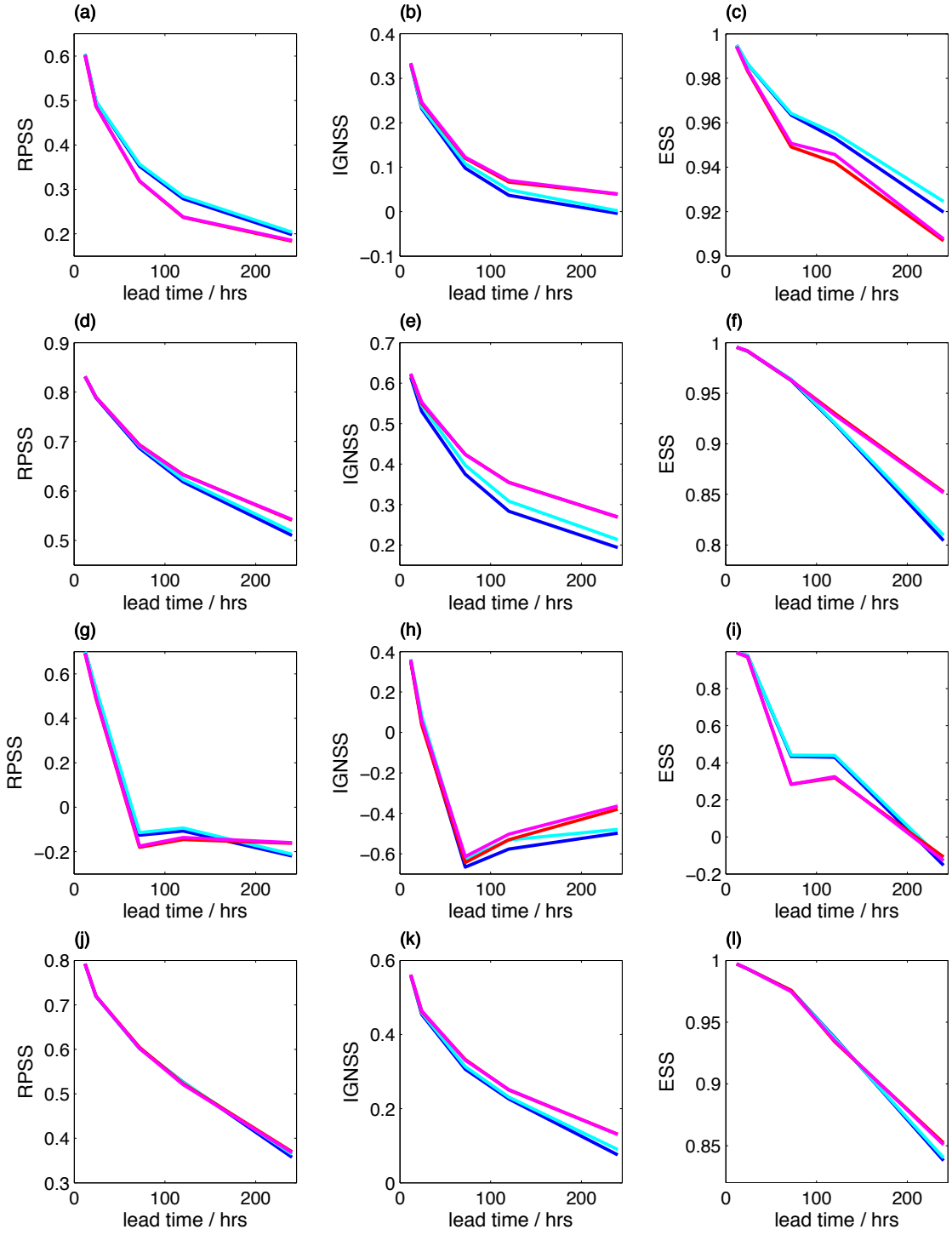

Figure 6.11: As for Figure 6.10, except for tropical regions with **significant convection**.

indicates an improvement in ensemble spread as observed, so it is likely this reduction in RPSS and ESS is due to the increased RMSE and the flat tail observed in the RMSE-spread diagnostic plots.

- The significantly improved RPSS and ESS for SPPTi for U850 is due to improved skill in convecting regions. The improvements in IGNSS involve contributions from both regions.

- Z500 is not a particularly informative field to study in the tropics as it is very flat and featureless. IGNSS indicates negative skill for lead times greater than 3 days, and RPSS and ESS also indicate little skill for this variable.

- The improvement of RPSS and IGNSS for U200 in the tropics is mostly due to an improved forecast skill in the regions with little convection. This improvement was clearly visible in the RMSE-spread scatter diagrams. A small but significant improvement in skill in convecting regions is also observed, especially at later lead times.

6.4 Convection Diagnostics

The impact of SPPTi is largest in regions with significant convection. To investigate if this is an indication that convection is modelled better by this scheme, the convection diagnostics discussed in Chapter 5 will be considered here, evaluated for tropical regions with significant convection.

6.4.1 Precipitation

Firstly, the skill of forecasting convective precipitation is considered. Convective precipitation is calculated by the convection scheme, and is not directly perturbed by SPPT. Therefore any impact of SPPTi on forecasting this variable indicates a feedback mechanism: the convection physics scheme is responding to the altered atmospheric state. As in Chapter 5, the GPCP data set is used for verification of precipitation. Figure 6.12 shows the RMS error-spread diagnostic for convective precipitation. It indicates that operational SPPT is under-dispersive for this variable, and that SPPTi results in an improved spread in forecast convective precipitation at all lead times. Figure 6.13 shows (a) the bias, (b) the RMSE and spread as a function of time, and (c)–(e) the forecast skill scores for convective precipitation. The bias in the convective

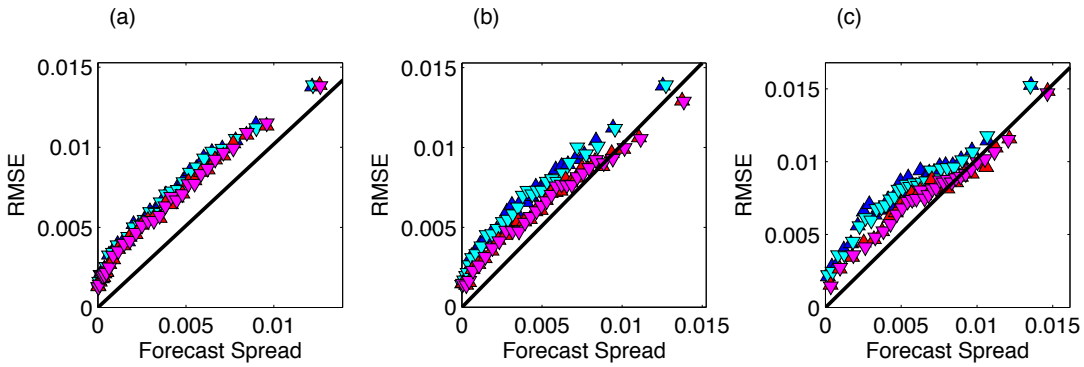

Figure 6.12: RMS error-spread diagnostic for cumulative convective precipitation for the 24 hour window before a lead time of (a) 1 day, (b) 3 days and (c) 10 days. The diagnostic is calculated for tropical regions with **significant convection**. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB. The one-to-one diagonal is shown in black.

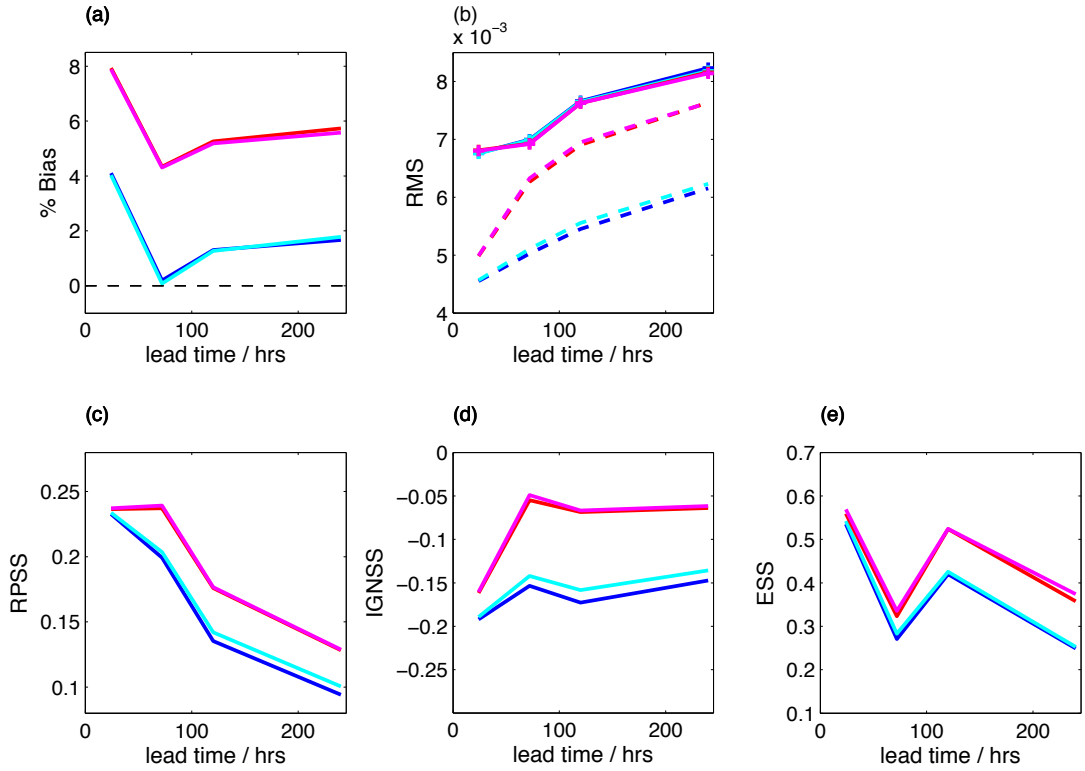

Figure 6.13: Summary forecast diagnostics for 24 hour cumulative convective precipitation in tropical regions with **significant convection**. (a) Percentage bias. (b) Temporal evolution of RMS ensemble spread (dashed lines) and error (solid lines) averaged over the region. (c) Ranked Probability Skill Score (d) Ignorance Skill Score. (e) Error-spread Skill Score. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB

precipitation forecasts is higher for SPPTi than for SPPT, but all other diagnostics indicate that the SPPTi forecasts are more skilful than the operational SPPT forecasts.

The skill of the precipitation forecasts can also be evaluated by considering the spatial distribution of cumulative precipitation (convective plus large-scale) for the different forecast models. The average 24-hour cumulative precipitation is shown for the GPCP data set in Figure 6.14. The difference between the forecast and GPCP fields is shown in Figure 6.15 for each of the four experiments in Table 6.1. Blue indicates the forecast has too little precipitation whereas red indicates too much precipitation. Figures (a) and (b) show the results for the operational SPPT scheme, with and without SKEB respectively. The results are very similar. Both show too much precipitation across the oceans. Figures (c) and (d) show the results for the SPPTi scheme, with and without SKEB respectively. Again, including SKEB has little impact, but including the SPPTi scheme has slightly increased the amount of precipitation over the oceans, as indicated earlier by the increase in bias in Figure 6.13(a). Using SPPTi does not result in a significant change in the spatial distribution of rain.

The skill of the precipitation forecasts can also be evaluated by considering the global frequency distribution of rain rate for the different forecast models and comparing to the observed rain rate distribution in the GPCP dataset. This is shown in Figure 6.16. All four forecast models perform well, though all underestimate the proportion of low rain rates and overestimate the proportion of high rain rates. The operational SPPT scheme is closer to the observations at mid to high rain rates, and the SPPTi scheme is marginally better at low rain rates. Importantly, the diagnostic does not flag up any major concerns for the new SPPTi: when SPPT was originally being developed, such a diagnostic showed that very high rain rates occurred at a significantly inflated frequency, which led to alterations to the SPPT scheme (Martin Leutbecher, pers. comm., 2013).

6.4.2 Total Column Water Vapour

Secondly, the skill of forecasting total column water vapour (TCWV) is considered, to which the convection scheme is sensitive. Figure 6.17 shows the RMS error-spread diagnostic for TCWV. As observed in Section 5.6.2.2, forecasts for this variable are poorly calibrated for all experiments. They have a systematically too large RMSE, and the RMS error-spread diagnostic has a shallow slope. Nevertheless, the SPPTi scheme increases the spread of the

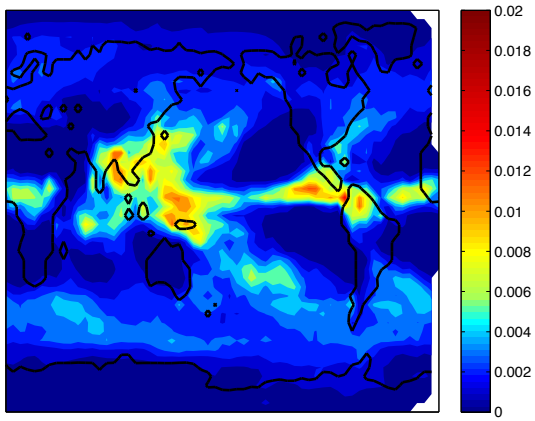

Figure 6.14: Distribution of 24-hour cumulative precipitation (metres) in the GPCP dataset, averaged for each successive 24-hour window between 14th April and 9th September 2012.

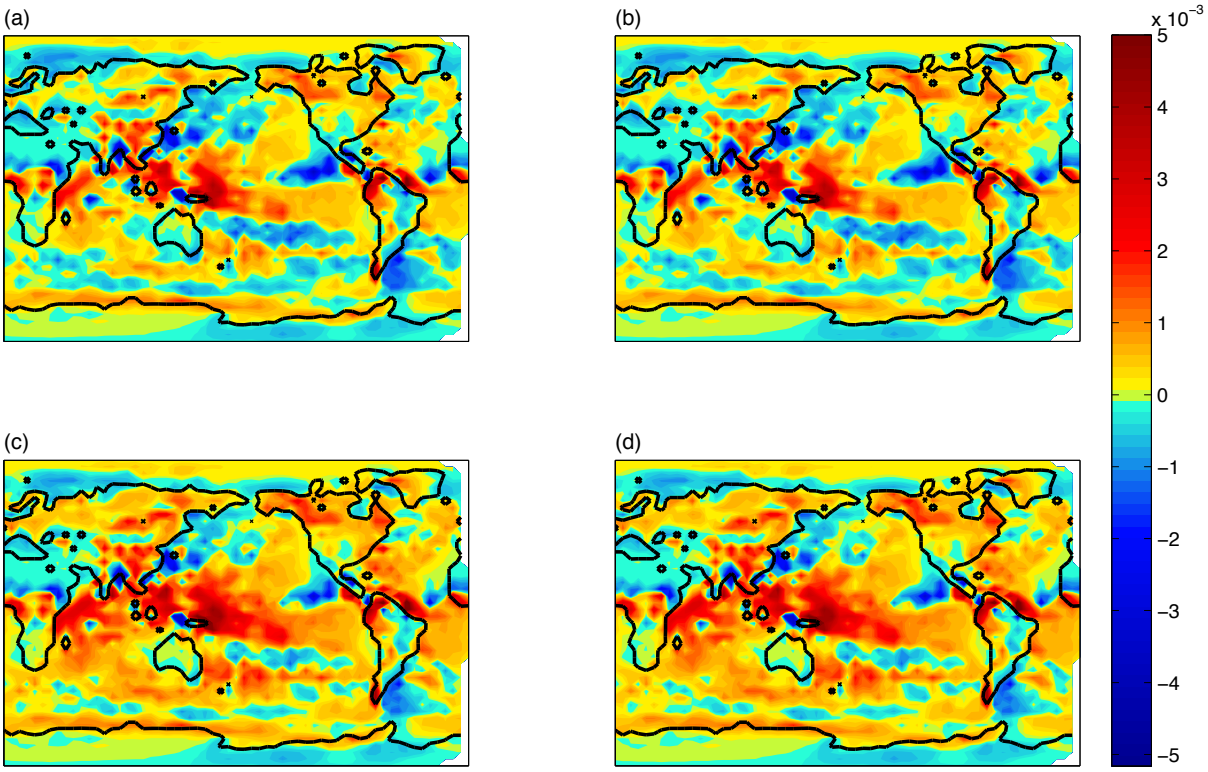

Figure 6.15: Difference between forecast and GPCP 24-hour cumulative precipitation (m). Blue indicates too little precipitation in the forecast, red indicates too much. The colour bar corresponds to all figures. Results are shown for the four experiments: (a) TSCS, (b) TSCS + SKEB, (c) TSCSi and (d) TSCSi + SKEB

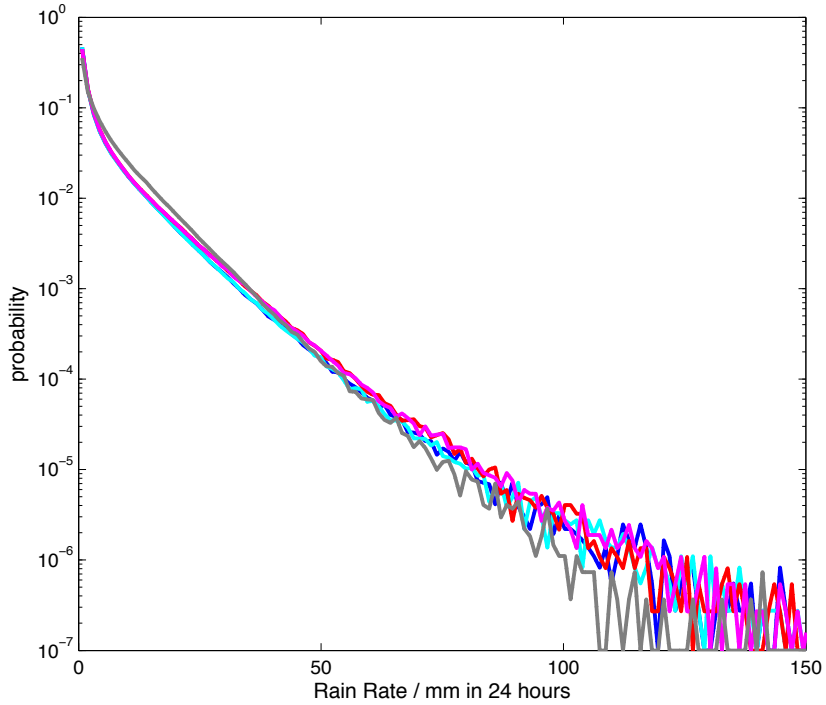

Figure 6.16: Probability distribution of rain rate (mm/12hrs) evaluated globally. The distribution has been normalised to 1, given that rain is observed in each 12 hour window. The observed result from the GPCP dataset (grey) are compared to TSCS (blue), TSCS + SKEB (cyan), TSCSi (red) and TSCSi + SKEB (magenta) forecasts.

ensemble compared to the operational scheme, improving the calibration. Figure 6.18 shows (a) the bias, (b) the RMSE and spread as a function of time, and (c)–(e) the forecast skill scores for TCWV. These diagnostics indicate a significant improvement in forecast skill of TCWV when SPPTi is used. The forecast bias is reduced, the RMS spread is increased without increasing the RMSE, and the RPSS, IGNSS and ESS all indicate higher skill.

It is possible that SPPTi could result in significant changes of TCWV in the tropics as was observed for the perturbed parameter experiments in Chapter 5, so this must be checked. Figure 6.19 shows the average TCWV between 20°S and 20°N (calculated as described in Section 5.6.2.2). This will diagnose if using SPPTi results in a systematic drying or moistening of the tropics. All experiments show an initial spin-down period where the tropics dry by $0.5\,\mathrm{kgm}^{-2}$ over the first 12 hours, before stabilising. The operational SPPT forecasts in figures (a) and (b) show a slight drying over the 240 hour forecast window, whereas the SPPTi forecasts in figures (c) and (d) have a more constant average TCWV. All four experiments show stable results.

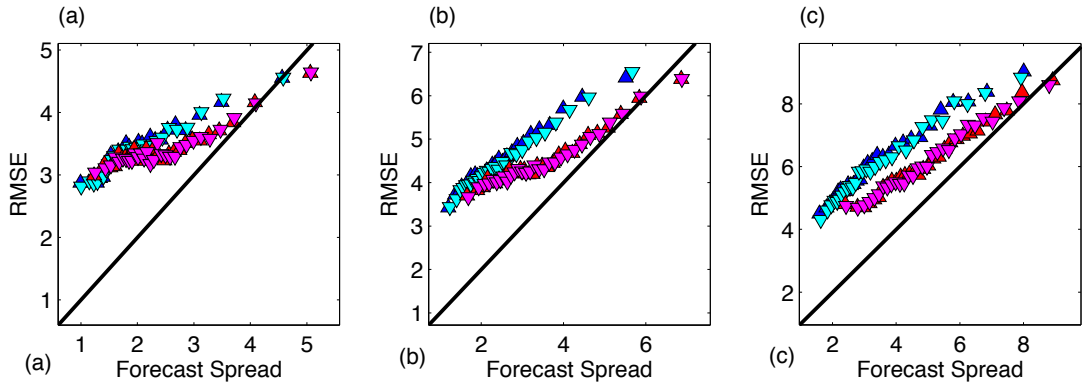

Figure 6.17: RMS error-spread diagnostic for TCWV for lead times of (a) 1 day, (b) 3 days and (c) 10 days. The diagnostic is calculated for tropical regions with **significant convection**. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB. The one-to-one diagonal is shown in black.

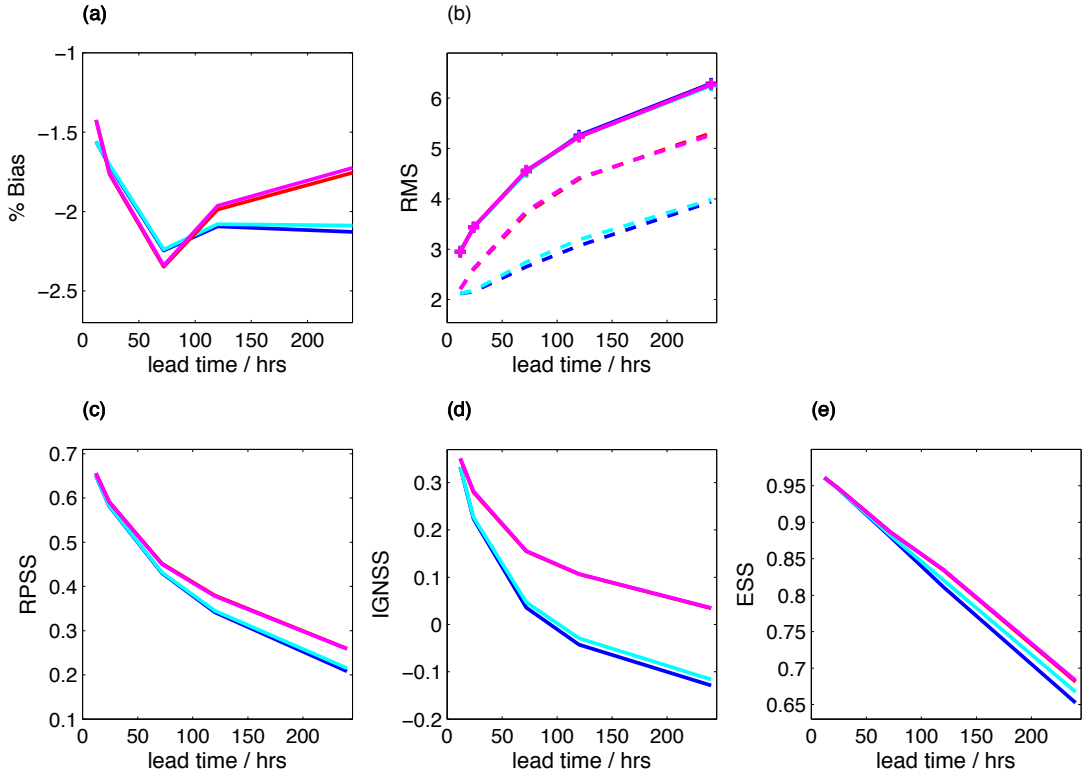

Figure 6.18: Summary forecast diagnostics for TCWV in tropical regions with **significant convection**. (a) Percentage bias. (b) Temporal evolution of RMS ensemble spread (dashed lines) and error (solid lines) averaged over the region. (c) Ranked Probability Skill Score (d) Ignorance Skill Score. (e) Error-spread Skill Score. Results are shown for the four experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi; magenta — TSCSi + SKEB

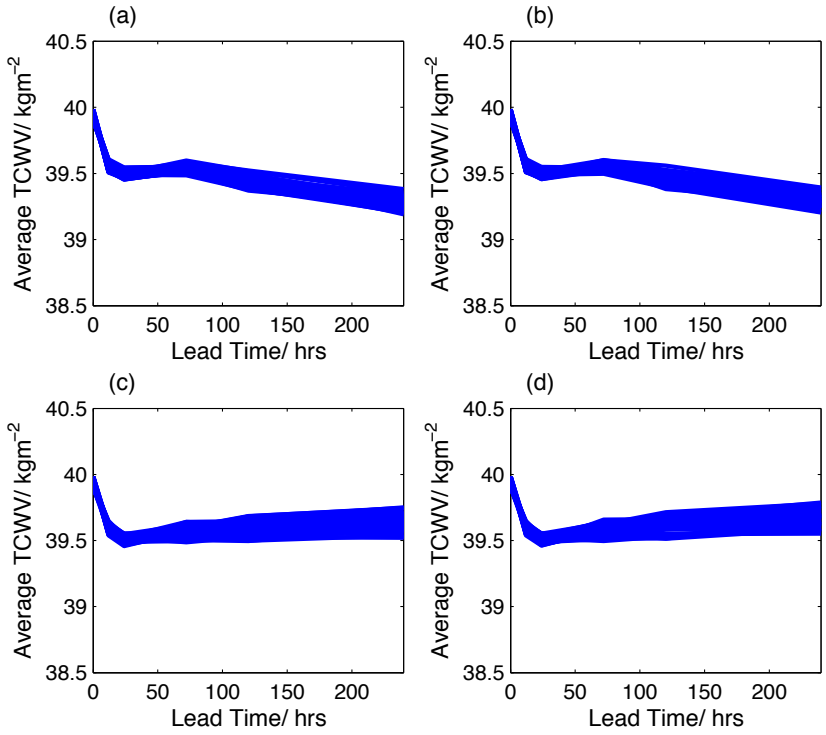

Figure 6.19: Average TCWV between 20°S and 20°N as a function of time. The spatial average is calculated for each ensemble member averaged over all start dates, and the averages for each of the fifty ensemble members are shown. Results are shown for the four experiments: (a) TSCS, (b) TSCS + SKEB, (c) TSCSi, and (d) TSCSi + SKEB.

6.5 Individually Independent SPPT

Independent SPPT assumes that the errors associated with different physics schemes are uncorrelated. It also has the effect of decoupling the physics schemes in the IFS: the random patterns are introduced after all calculations have been made so each physics scheme does not have the opportunity to react to the modified tendencies from the other schemes. The results presented in this chapter show that this assumption results in a large increase of spread, particularly in convecting regions, and for U200 in non-convecting regions. To probe further into the mechanisms of SPPTi, a series of five experiments was carried out. In each experiment, just one of the five physics schemes was perturbed with an independent random number field to the other four (Table 6.2). These "individually independent SPPT" experiments should indicate the degree to which a particular physics scheme should have an independent error distribution from the others. In particular, these experiments aim to answer the following questions:

1. Is it decoupling one particular scheme from the others that results in the large increase in spread, or is it important that all schemes are treated independently?

2. Does decoupling one particular scheme result in the increased error observed for T850?

Physics Scheme	Experiment Abbreviation if Independently Perturbed
Radiation	RDTTi
Turbulence and Gravity Wave Drag	TGWDi
Non-Orographic Gravity Wave Drag	NOGWi
Convection	CONVi
Large Scale Water Processes	LSWPi

Table 6.2: The experiment abbreviations for the individually independent SPPT experiments, in which each physics scheme in turn is perturbed with a different pattern to the other four schemes, which are perturbed together.

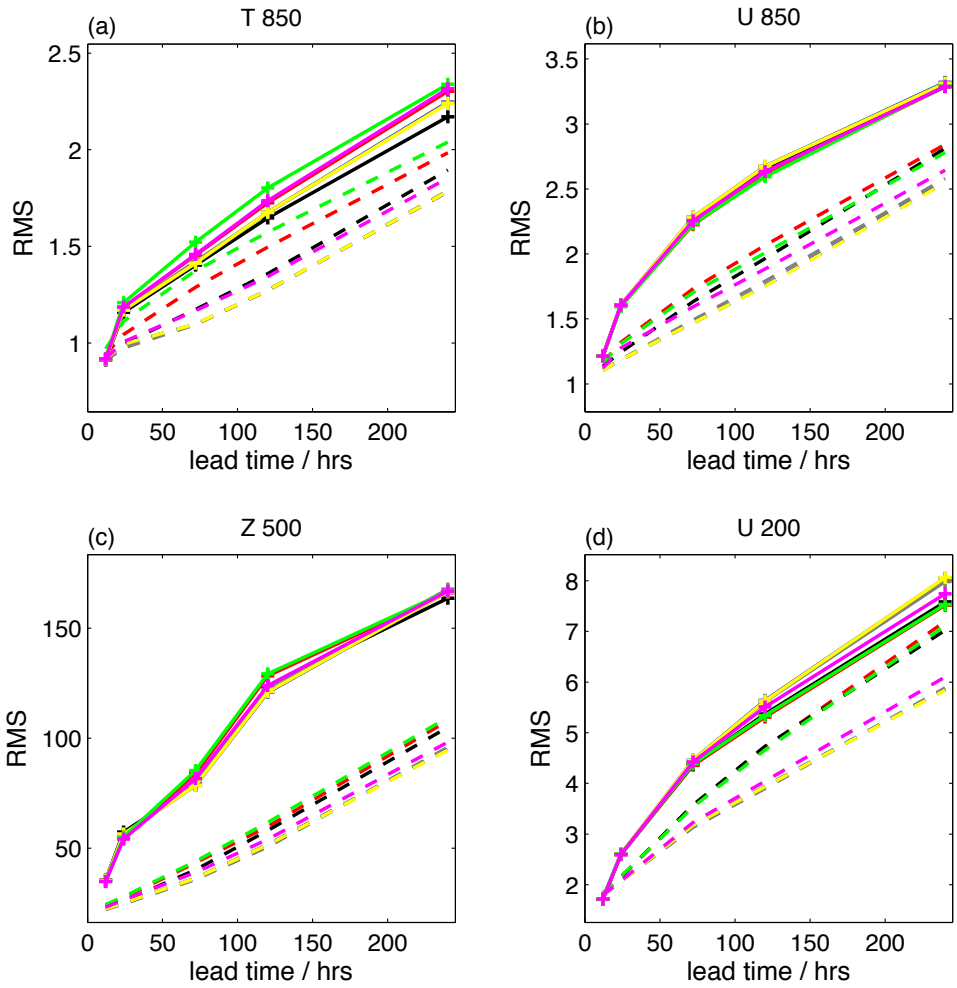

Figure 6.20: RMS error (solid lines) and spread (dashed lines) as a function of time for forecasts in tropical regions with **little convection**. (a) T850, (b) U850, (c) Z500 and (d) U200. The five individually independent SPPT experiments are shown. Black — RDTTi. Grey — TGWDi. Yellow — NOGWi. Green — CONVi. Magenta — LSWPi. Blue — operational SPPT, and Red — SPPTi are included for comparison. The blue lines are obscured by the yellow and grey lines in each figure.

Figure 6.20 shows the RMSE in the ensemble mean and RMS ensemble spread as a function of time for each of the five individually independent SPPT experiments in regions with little convection. The results for SPPT and SPPTi are also shown for comparison. The largest impact is observed for U200, where both CONVi and RDTTi show the same spread increase as SPPTi. This indicates that *it is decoupling these two schemes which results in the large spread increase observed for U200 when SPPTi is used instead of SPPT*. For the other variables considered, the impact of each individually independent scheme is more moderate, though in each case CONVi and RDTTi result in the largest increase in spread. For T850, RDTTi also results in a reduction in RMSE when compared to forecasts which use the operational SPPT scheme.

Apart from for U200, SPPTi has the largest impact in regions with significant convection. Figure 6.21 shows the RMSE and RMS spread as a function of time in these regions. CONVi has the largest impact for each variable — *perturbing the convection tendencies independently from the other schemes results in an increase of ensemble spread equal to or greater than independently perturbing all physics schemes*. This supports the results from Chapter 5, in which it was observed that decoupling the convection scheme by not perturbing its tendencies resulted in an increase in spread. The next most influential scheme is radiation. For Z500 and U200, perturbing this scheme independently also results in an increase of ensemble spread equal to SPPTi. A large impact is also seen for U850 and T850. For the variables at 850 hPa, LSWPi has a large impact. This is especially true at short lead times, when the impact is greater than that of radiation. Using independent random fields for TGWDi (grey) and NOGWi (yellow) has little impact on the ensemble spread — their RMSE and RMS spread are almost identical to those from the operational SPPT forecasts. This is probably because these two schemes act mainly in the boundary layer (TGWD) or in the middle atmosphere (NOGW), away from the variables of interest. Additionally, the stochastic perturbations to these schemes will be tapered, which will further reduce the impact of SPPTi.

Figure 6.21 also indicates which schemes contribute to the observed increase/decrease in RMSE for SPPTi in regions with significant convection. For T850, SPPTi resulted in an increase in RMSE. This same increase is observed for CONVi, and to a lesser extent for the LSWPi forecasts. For the other variables, CONVi shows a similar RMSE to operational SPPT. It is interesting to note that the RDTTi experiment does not result in an increase in error for

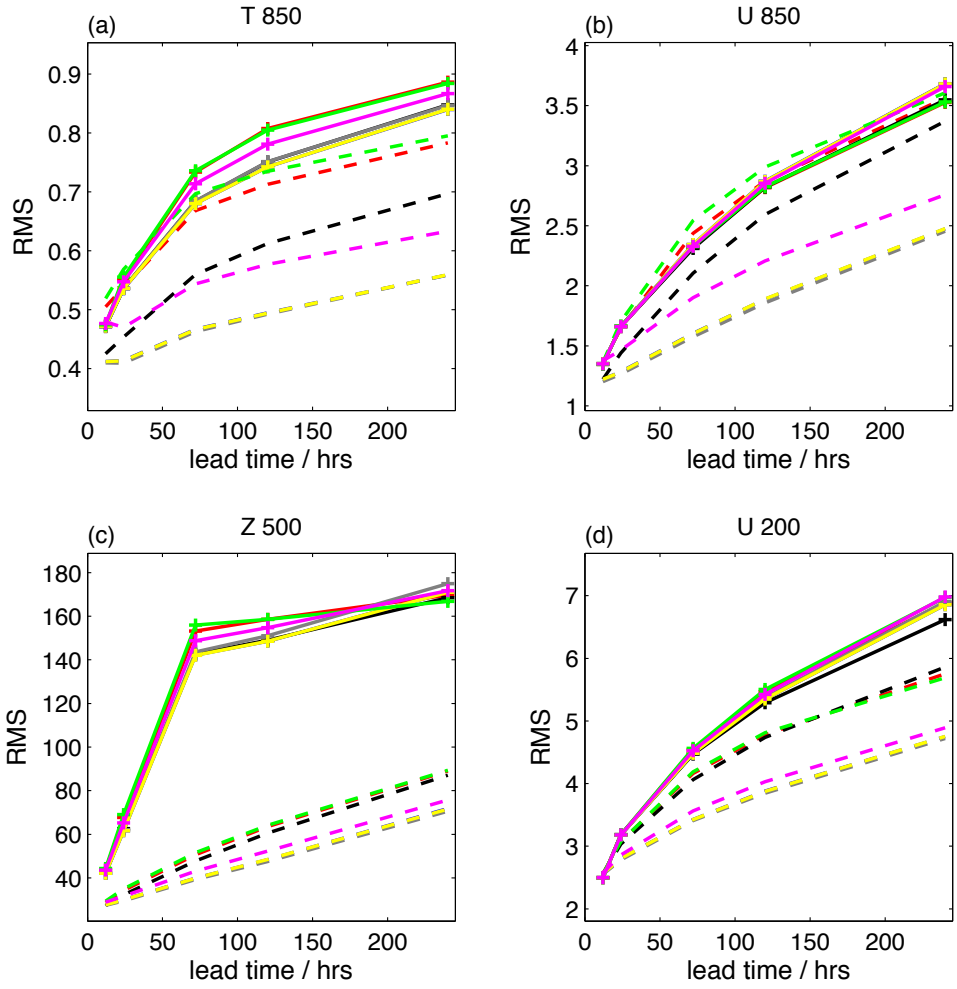

Figure 6.21: RMS error (solid lines) and spread (dashed lines) as a function of time for forecasts in tropical regions with **significant convection**. (a) T850, (b) U850, (c) Z500 and (d) U200. The five individually independent SPPT experiments are shown. Black — RDTTi. Grey — TGWDi. Yellow — NOGWi. Green — CONVi. Magenta — LSWPi. Blue — operational SPPT, and Red — SPPTi are included for comparison. The blue lines are obscured by the yellow and grey lines in each figure. The black solid line in (a) is obscured by the grey solid line.

T850, but does give a substantial increase in spread. The RDTTi experiment also performs well for U850 and Z500, with an increase in spread and no increase in error observed for both variables. For U200, the RDTTi scheme results in a decrease in error. These results imply that *much of the spread increase observed with SPPTi could be achieved by perturbing radiation independently from the other physics schemes, which will not result in the increase in RMSE for T850.*

Figure 6.22 shows the RMS error-spread diagnostic at a lead time of ten days for the individually independent experiments in regions with significant convection. This diagnostic confirms that both CONVi (green) and RDTTi (black) produce forecasts with a similar degree of spread to SPPTi (red). Furthermore, these individually independent schemes improve the one-to-one relationship between RMSE and RMS spread. Figure 6.22(a) shows the results for T850, including the increased error for predictable situations. The inset figure shows the region of interest in more detail, indicated by the grey rectangle. *LSWPi results in a significant increase of error and a flatter 'tail'. CONVi also results in an increase of error for the smallest forecast spread cases*, giving an upward 'hook' in the scatter diagnostic at smallest spreads. This indicates poorly calibrated forecasts: the forecast spread does not correctly indicate the error in the ensemble mean, and forecasts with the smallest spreads of between 0.4 and 0.5°C consistently have a higher error than those with spreads between 0.5 and 0.6°C. The results for RDTTi are positive, showing an increase in spread but no associated increase in error.

Figure 6.23 shows the skill of the individually independent SPPT forecasts in regions with significant convection, as indicated by the RPSS, IGNSS and ESS. Overall, the RDTTi forecasts are more skilful than forecasts from any other scheme. In fact, RDTTi is more skilful than SPPT for T850, whereas SPPTi was less skilful than SPPT for this variable. RDTTi also performs well for the other variables considered, and has skill equal to or better than the SPPTi scheme in most cases.

6.6 High Resolution Experiments

Due to limitations in computer resources, the experiments presented above ran the IFS at a relatively low resolution of T159. The question then arises: does SPPTi have the same impact when the model is run at the operational resolution of T639? I am grateful to Sarah-Jane Lock (ECMWF), who ran two experiments on my behalf to test SPPTi at operational resolution.

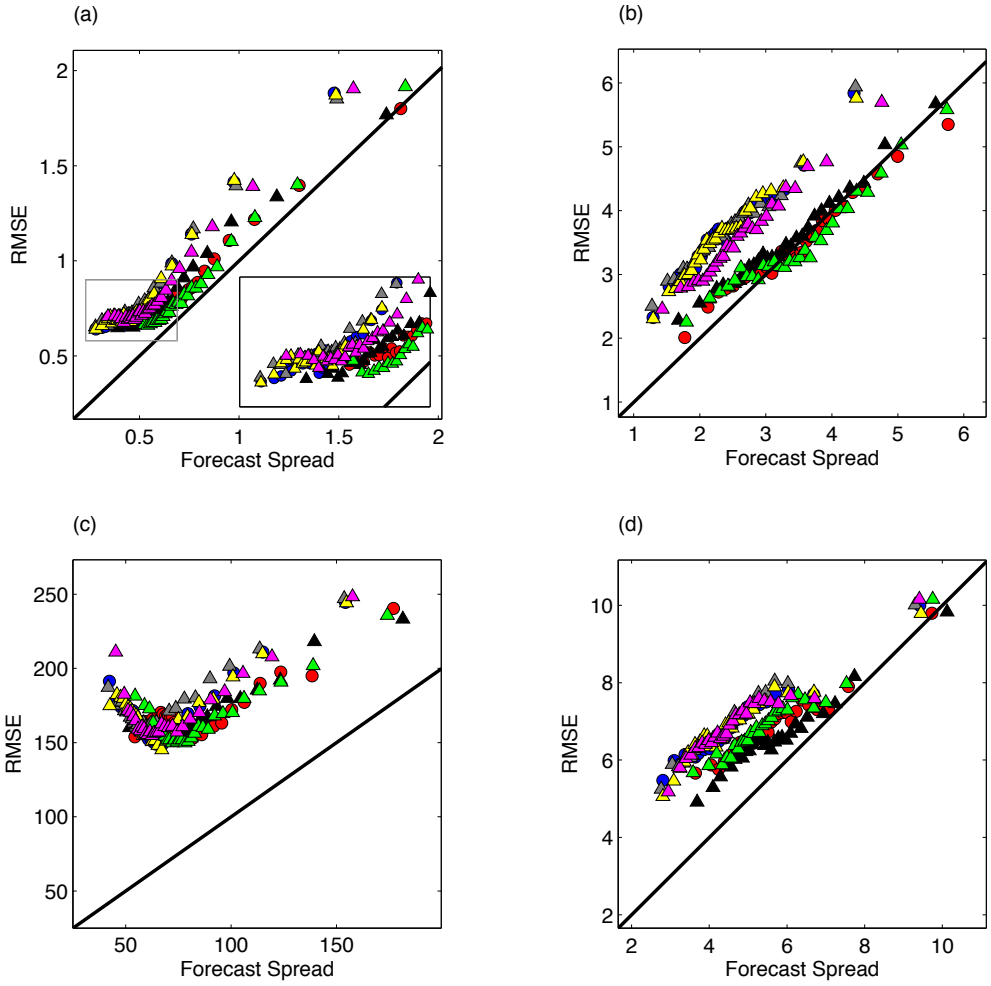

Figure 6.22: RMS error-spread diagnostic for tropical regions with **significant convection** for (a) T850, (b) U850, (c) Z500 and (d) U200, at a lead times of 10 days. The five individually independent SPPT experiments are shown (triangles): Black — RDTTi, Grey — TGWDi, Yellow — NOGWi, Green — CONVi, and Magenta — LSWPi. For comparison, the operational SPPT (blue circles) and SPPTi (red circles) are also shown. The one-to-one diagonal is shown in black. The tiled figure in (a) is a close up of the region indicated by the grey rectangle.

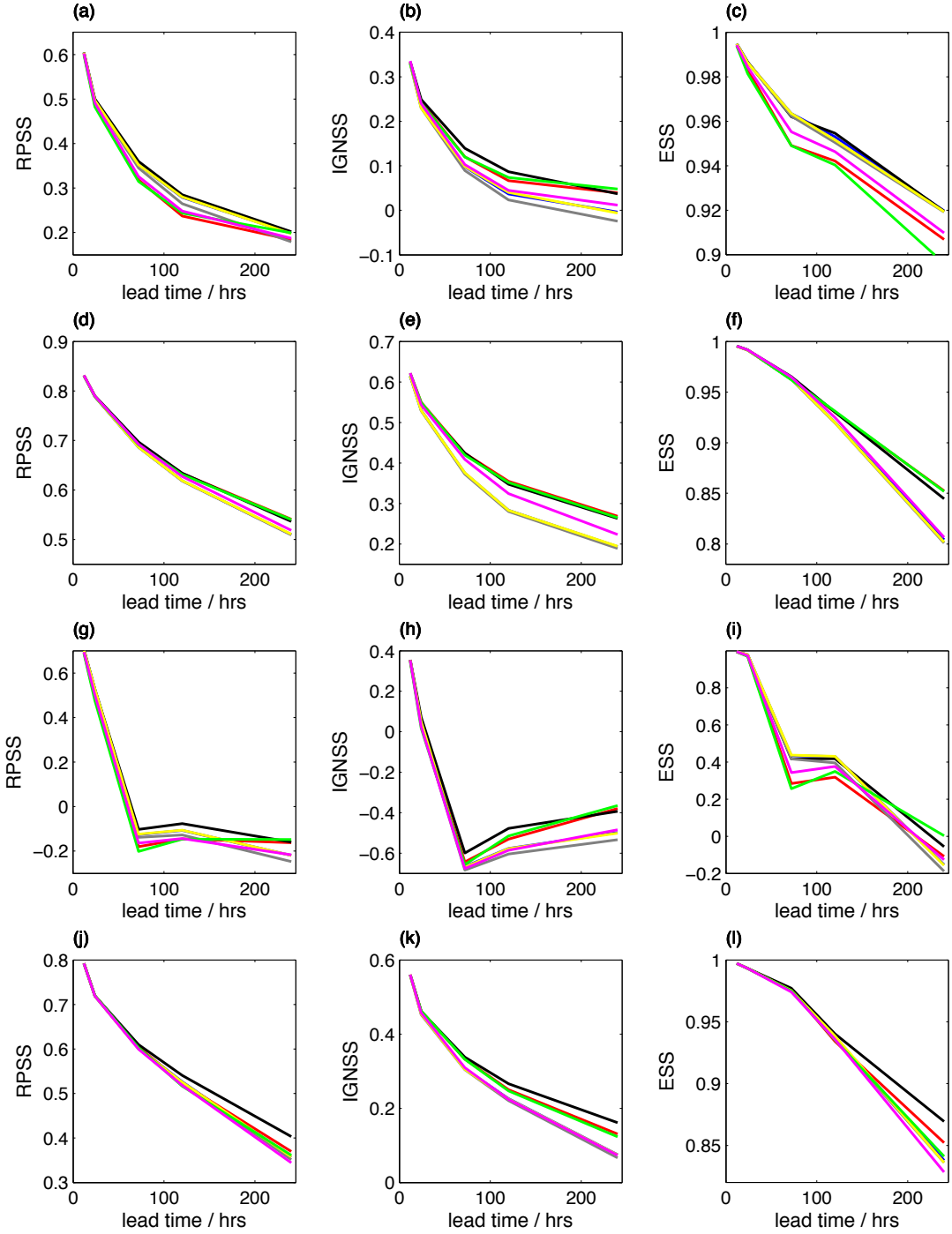

Figure 6.23: Ensemble forecast skill scores calculated for tropical regions with **significant convection**. First column: Ranked Probability Skill Score. Second column: Ignorance Skill Score. Third column: Error-spread Skill Score. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the five individually independent SPPT experiments: Black — RDTTi, Grey — TGWDi, Yellow — NOGWi, Green — CONVi, and Magenta — LSWPi. Blue — operational SPPT, and Red — SPPTi are included for comparison. The blue lines are obscured by the yellow lines in each figure. Additionally, in (j) the grey line is obscured by the magenta line; in (k) the yellow and grey lines are obscured by the magenta line; in (l) the grey line is obscured by the yellow line.

Ten-day ensemble hindcasts were initialised every five days between 14 April and 18 June 2012 (14 dates in total). The ensembles have 20 members instead of the operational 50. The two experiments repeated at T639 were "SPPT" and "SPPTi". A subset of twenty ensemble members is taken from the operational forecasts for these dates to produce equivalent "SPPT + SKEB" forecasts for comparison.

6.6.1 Global Diagnostics

Figure 6.24 shows the RMSE and RMS spread for each of the standard ECMWF global regions as a function of time for the variables of interest. At this higher resolution, the spread of the forecasts is well calibrated in the extra-tropics (first and third column). *SPPTi has little impact here, so the ensembles remain well calibrated.* In the tropics, the T639 forecasts are under-dispersive. *Here, SPPTi results in a significant increase in spread, and has a larger impact on ensemble spread than the operational SKEB scheme.* These results are similar to those at T159, shown in Figure 6.2. The key difference between T159 and T639 is that the operational T639 forecasts are better calibrated than the equivalent T159 forecasts. This means that when SPPTi is implemented at T639, the ensemble forecasts become over-dispersive for some variables (e.g. U850). *There is also a significant increase in RMSE for T850 and Z500 forecasts at T639.* The impact of SPPTi on U200 forecasts is an good match between the ensemble spread and RMS error in the ensemble mean.

It is important to note that for T850, U850 and U200, the ensemble spread is greater than the RMSE at a lead time of 12 hours for all experiments. This is indicative of inflation of initial condition uncertainty to compensate for an incomplete representation of model uncertainty. Because the ensembles are under-dispersive at longer lead times, the initial condition perturbations have been artificially inflated to increase the ensemble spread. In fact, in the IFS, the initial condition perturbations calculated by the EDA system are combined with singular vector perturbations before they are used. *If SPPT is replaced by SPPTi, this artificial inflation could be removed, and the raw initial condition uncertainty estimated by the EDA system used instead.* The temporal evolution of the ensemble spread for forecasts of U850 closely matches the evolution of the RMSE (Figure 6.24(e)), and if the initial condition perturbations were to be reduced to the raw EDA output, the results here indicate that SPPTi could produce a forecast that is well calibrated at all lead times for this variable. It would also be interesting

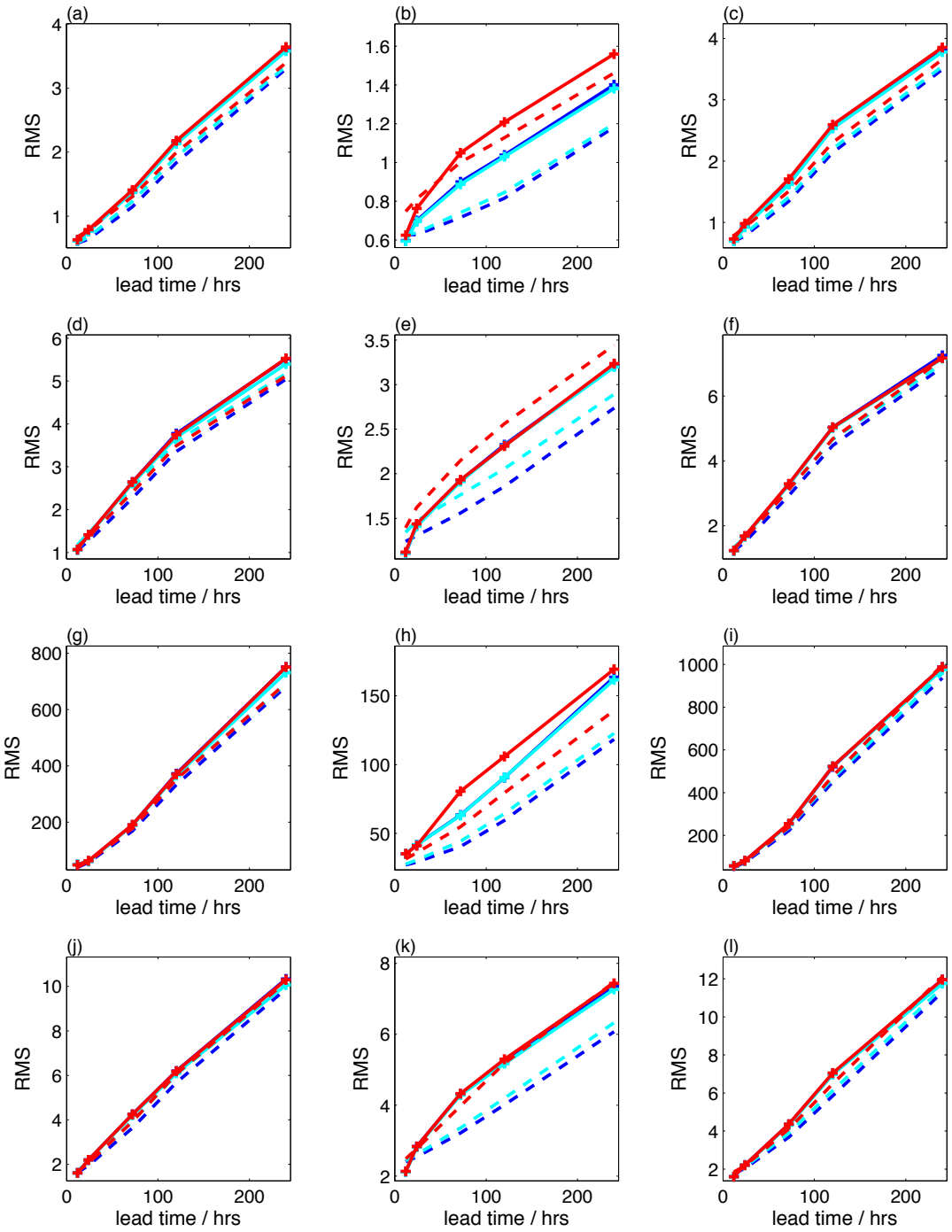

Figure 6.24: Temporal evolution of RMS ensemble spread (dashed lines) and RMSE (solid lines) for each standard ECMWF region. First column: northern extra-tropics, north of 25N. Second column: tropics, 25S–25N. Third column: southern extra-tropics, south of 25S. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the three T639 experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi.

to test using SPPTi in the EDA system, as it is possible that using SPPTi will impact the initial condition uncertainty estimated using the EDA.

6.6.2 Verification in the Tropics

As at T159, SPPTi has the largest impact in the tropics. To investigate the source of the increased spread, the forecasts will be verified in areas in the tropics where there is little convection, and in areas where convection is the dominant process. The areas considered will be those defined in Section 5.5.1. Results are shown in Figures 6.25 and 6.26 for regions with little and significant convection respectively. The operational forecasts are under-dispersive in both regions. As at T159, the under-dispersion is more severe in regions with significant convection. For both regions, SPPTi has the effect of significantly increasing the ensemble spread, whereas the impact of SKEB is more moderate. However, SPPTi also increases the RMSE for T850 and Z500 forecasts in both regions, and results in a slight increase of RMSE for U850 and U200 forecasts in convecting regions. In non-convecting regions, SPPTi results in a slight reduction in RMSE for U850 and U200. The improved temporal evolution of the ensemble spread identified above is observed in convecting regions, but not in non-convecting regions. *As at T159, the difference in behaviour between convecting and non-convecting regions indicates that it is convection, and its interactions with other physical processes, that is the key mechanism by which SPPTi affects the ensemble.*

Figures 6.27 and 6.28 show the RMS error-spread graphical diagnostic at a lead time of ten days for the three T639 experiments for regions with little and significant convection respectively. The forecasts have been binned into 14 bins instead of 30 to ensure the population of each bin is the same as before, and is a sufficiently large sample to estimate the statistics. The impact of SPPTi is small in regions with little convection, though the spread error relationship is improved slightly for U850. The average spread of forecasts is also improved for U200, but the flow dependent calibration is poor — the scattered points do not follow the one-to-one line. In regions of significant convection, the impact is greater. For T850 and Z500 there is an improved error-spread relationship when SPPTi is used instead of SPPT. The spread of the ensemble forecasts has increased, and the forecasts continue to give a flow-dependent indication of uncertainty in the forecast. However, for T850, an increase in RMSE is observed. Unlike at T159, this increase in RMSE occurs for all forecast spreads, not just for the small spread and

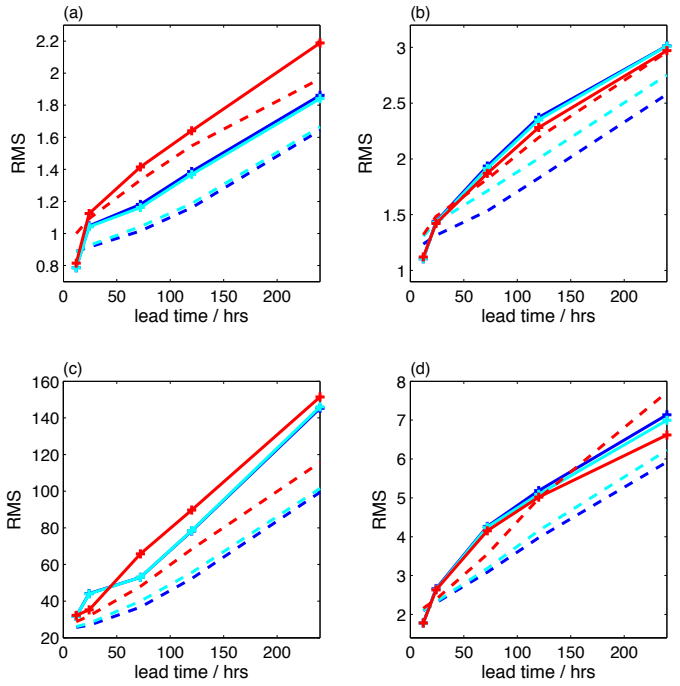

Figure 6.25: Temporal evolution of RMS ensemble spread (dashed lines) and RMSE (solid lines) for tropical regions with **little convection** for (a) T850, (b) U850, (c) Z500 and (d) U200. Results are shown for the three T639 experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi.

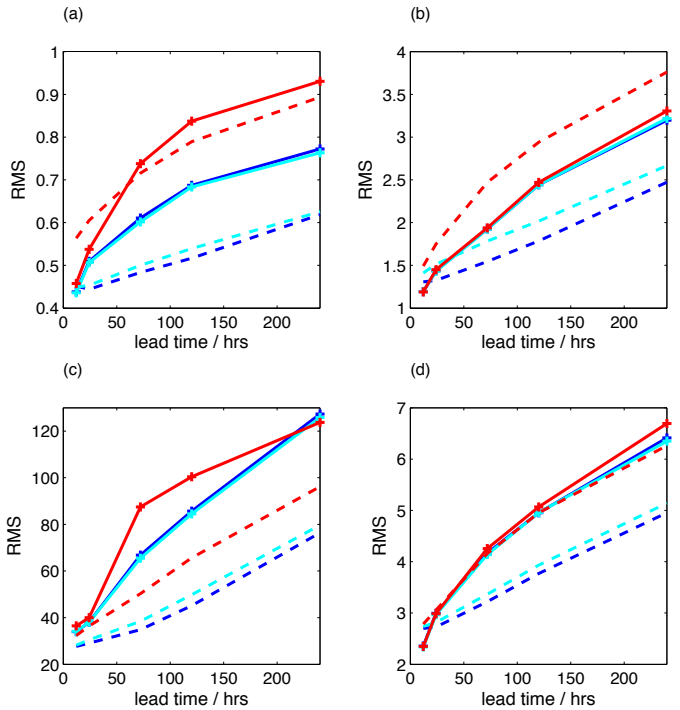

Figure 6.26: As for Figure 6.25, except for tropical regions with **significant convection**.

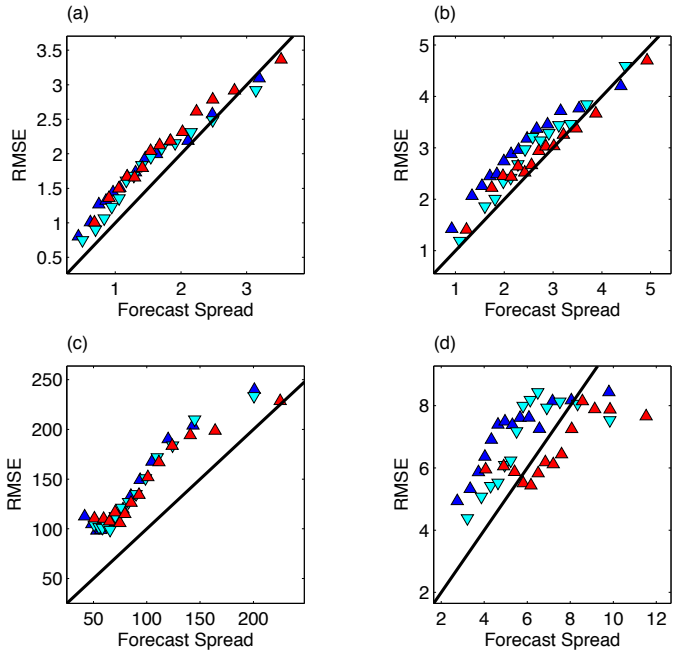

Figure 6.27: RMS error-spread diagnostic for tropical regions with **little convection** for (a) T850, (b) U850, (c) Z500 and (d) U200, at a lead times of 10 days for each variable. Results are shown for the three T639 experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi. The one-to-one diagonal is shown in black.

error cases. For U850, SPPTi also increases the spread of the forecasts, but by too great an amount. For U200, the SPPTi results follow a shallower slope than the SPPT forecast results. This indicates a reduced degree of flow-dependent predictability, though this is not a problem at earlier lead times (not shown).

Figures 6.29 and 6.30 show the skill of the ensemble forecasts in regions with little and significant convection respectively. Despite the improvement in spread, the SPPTi forecasts tend to score poorly due to an associated increase in RMSE. The SPPTi forecasts are more skilful than the SPPT forecasts for U850 and U200 in non-convecting regions, and for Z500 at long lead times in convecting regions according to all skill scores.

6.7 Discussion and Conclusion

SPPTi results in a significant increase of spread for all variables at all lead times. In the extra-tropics, the ensemble forecasts are well calibrated at T639, and moderately under-dispersive at T159. The impact of SPPTi is small in these regions. At T159, a small increase in ensemble spread is observed correcting for the under-dispersion, and at T639 the impact is smaller, and the ensemble forecasts remain well calibrated. The impact of SPPTi is similar to SKEB in these regions. In the tropics, forecasts made with SPPT are significantly under-dispersive at

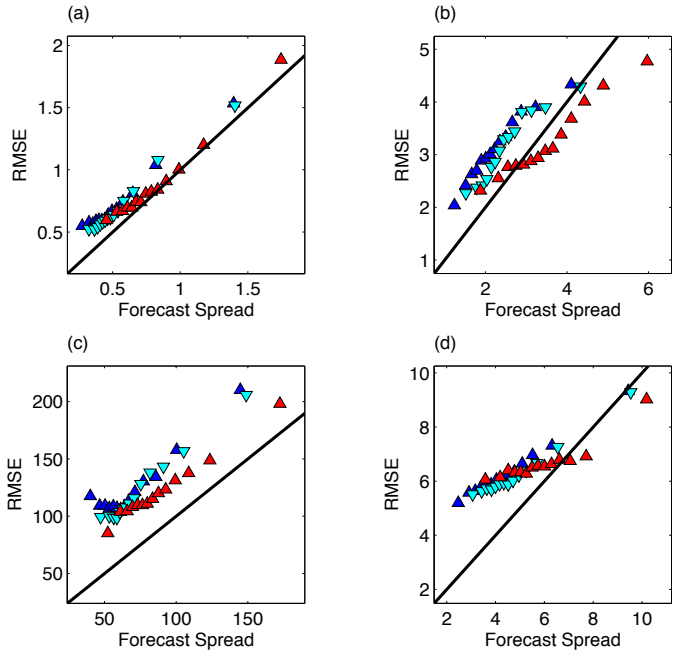

Figure 6.28: As for Figure 6.27, except for tropical regions with **significant convection**.

both T159 and T639. SPPTi has a large beneficial impact in these regions. The forecast spread is significantly larger than when SPPT is used, and the impact is considerably larger than the impact of SKEB. This is observed at both T159 and T639. *SPPTi produces skilful, flow-dependent estimates of forecast uncertainty, having a larger impact on forecasts that were more under-dispersive when using SPPT.*

The impact of SPPTi in tropical regions with significant convection (Figure 6.7) is considerably greater than in tropical regions with little convection (Figure 6.6) for T850 and U850, and to a lesser extent, for Z500. *This indicates that convection, together with its interactions with other physics schemes, is a key process by which SPPTi impacts the ensemble.* Equation (6.2) indicates that the forecast uncertainty represented by SPPTi will only be greater than SPPT for regions where the model tendencies act in opposite directions, i.e., where the individual tendencies are large but the net tendency is small. In tropical regions with significant convection, this is indeed the case for the IFS. The convection scheme parametrises the effect of convective latent heating on the atmosphere. This scheme interacts directly with the large scale water processes (clouds) scheme: water detrained from the convective plume acts as a source of water for clouds in the LSWP scheme, which then calculates the effect of evaporative cooling on the atmosphere (ECMWF, 2012). This interaction means that a warming due to convection tends to be associated with a cooling from the cloud scheme. The opposing nature of these tendencies results in the significant increase in ensemble spread associated with SPPTi

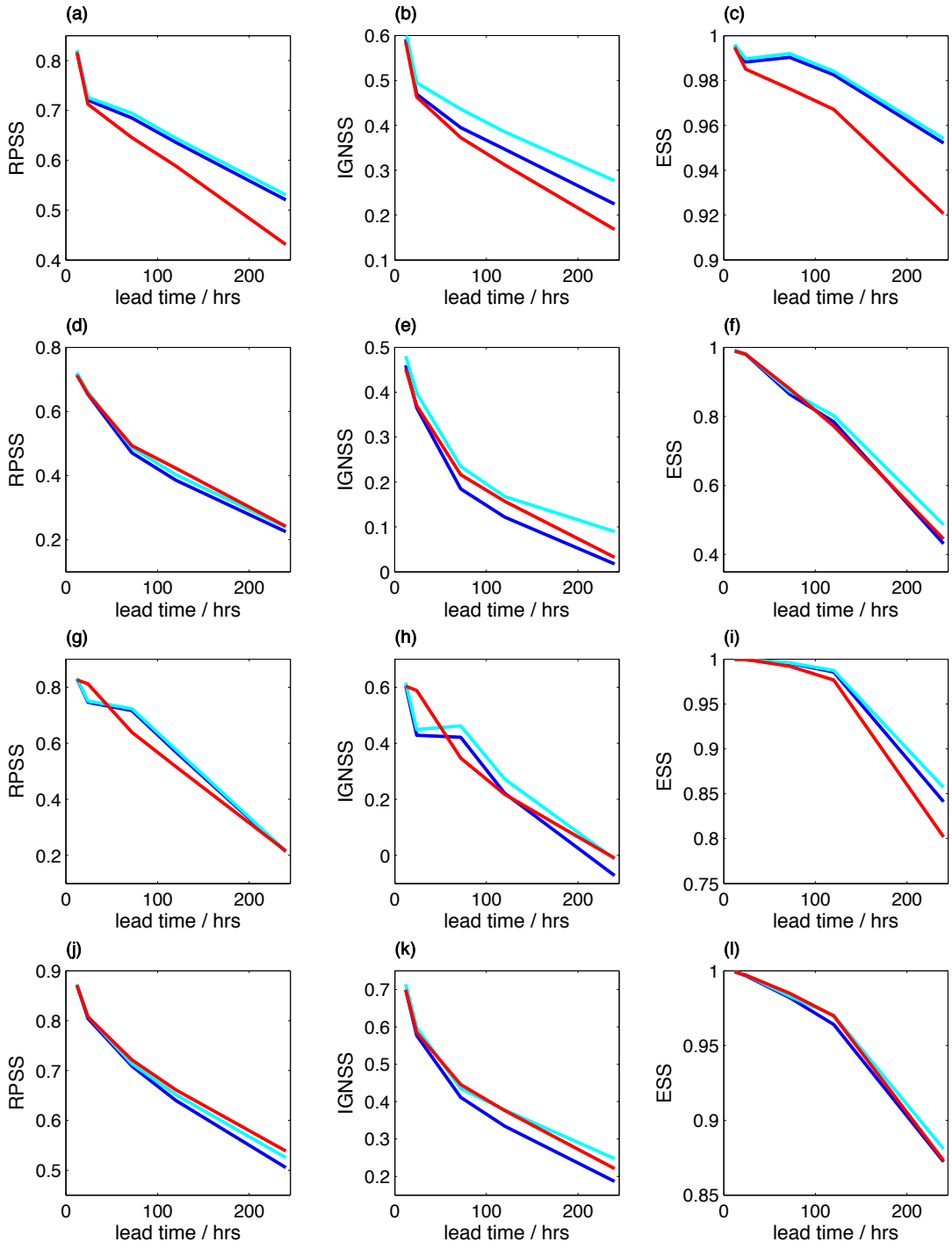

Figure 6.29: Ensemble forecast skill scores calculated for tropical regions with **little convection**. First column: Ranked Probability Skill Score. Second column: Ignorance Skill Score. Third column: Error-spread Skill Score. (a)–(c) T850, (d)–(f) U850, (g)–(i) Z500 and (j)–(l) U200. Results are shown for the three T639 experiments: blue — TSCS; cyan — TSCS + SKEB; red — TSCSi.

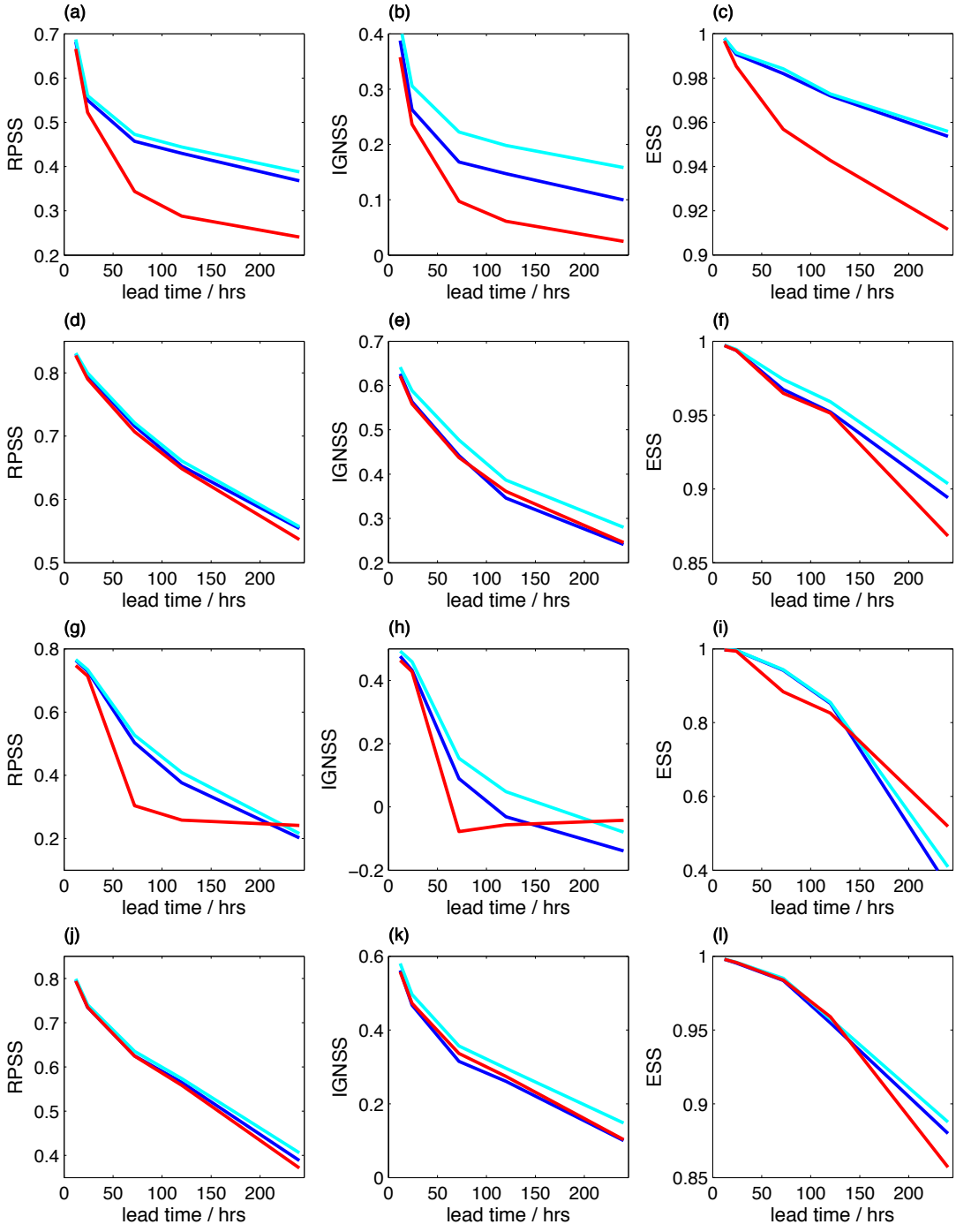

Figure 6.30: As for Figure 6.29, except for tropical regions with **significant convection**.

in these regions. The individually independent SPPT experiments also suggest *it is decoupling clouds and convection from each other that results in this large increase in spread for T850 and U850*, as both the CONVi and LSWPi experiments showed increases in spread compared to SPPT (Figure 6.21).

The impact of the convection scheme on clouds also impacts the radiation parametrisation scheme. As noted in Section 5.2.1, the radiation scheme interacts with the cloud scheme since both short- and long-wave radiative transfer are sensitive to the cloud fraction predicted by the cloud scheme. In particular, low level cloud is often associated with cooling from the radiation scheme (Morcrette, 2012), which opposes the warming from convection. This interaction between radiation and convection could contribute to the increase in spread for the SPPTi forecasts in regions with significant convection. The RDTTi experiment showed that decoupling RDTT from the other parametrisation schemes results in a large increase in spread for T850 and U850 (Figure 6.21), supporting this hypothesis.

For U200, the largest increase in spread was also observed in the tropics (Figure 6.2). However, unlike for the other variables, this spread increase is predominantly from forecasts for tropical regions with little convection (Figure 6.6). The individually independent SPPT experiments shown in Figure 6.20 show that independently perturbing just RDTT or CONV gives the same increase in ensemble spread as SPPTi: *for U200, it is decoupling RDTT from CONV that results in the large increase in forecast spread.* The variable U200 is sensitive to convection as it is located close to the level at which maximum convective outflow occurs. In regions with significant convection, it is expected that there will be thick cloud at this level due to the spreading out of the convective anvil. If the 200 hPa level falls at the top of an anvil cloud, significant radiative cooling will be observed as the cloud efficiently emits longwave radiation to space (Gray and Jacobson, 1977). However if the 200 hPa level falls below an anvil, a radiative warming would be observed due to an increase in trapped longwave radiation. This characteristic warming profile is shown schematically in Figure 6.31, taken from Gray and Jacobson (1977). For this reason, in regions of significant convection the radiation scheme will produce tendencies at 200 hPa which either oppose or enhance the convective warming, reducing the impact of SPPTi when averaged over many cases. In contrast, regions with little convection have reduced amounts of optically thick high level cloud, so the radiation scheme will tend to cool the atmosphere at this level. A large impact is observed in these regions

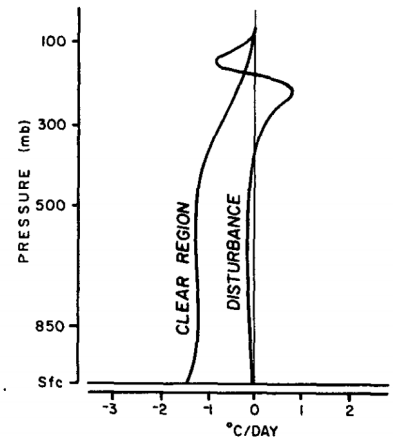

Figure 6.31: Typical radiation induced temperature changes in the tropics for a clear sky region compared to a 'disturbance' region with thick high cloud cover. Taken from Gray and Jacobson (1977).

when the opposing CONV and RDTT tendencies are decoupled using SPPTi. Decoupling the convective and radiative temperature tendencies affects the horizontal temperature gradients at 200 hPa, which could then affect the zonal wind speed at that level[1].

Considering the YOTC tendencies at 200 hPa provides support for this hypothesised mechanism. Figure 6.32 shows the 24 hr temperature tendencies of RDTT, CONV and LSWP at 200 hPa, averaged over 30 start dates between 14 April and 6 September 2009. Each scattered point represents the tendency from two schemes at a particular location. Figure (a) shows that in regions of significant convection, while the temperature tendencies are consistently positive, radiation tendencies can be either positive or negative. In regions with little convection, the convective tendencies remain positive on average, but the radiative tendencies are negative on average.

The impact of SPPTi on U200 could also be indicative of an improved variability of convection in the IFS. The upper level wind field is sensitive to waves generated remotely by convective systems (Stensrud, 2013). The improvement in ensemble spread and reduction in error is most apparent at a lead time of ten days (Figure 6.8), which could suggest that this improvement is due to a remote source. Other diagnostics indicate that convection is better represented when SPPTi is used instead of SPPT. The skill of convective precipitation forecasts at T159 improve at all lead times when SPPTi is used; forecasts show an improved ensemble spread and a slight reduction in RMSE, though the wet bias of the model is also increased. Forecasts of TCWV show a significant improvement when SPPTi is used. The bias is reduced,

[1]The RDTT parametrisation scheme directly affects the atmospheric temperature only, so the observed impact of RDTTi on U indicates a feedback mechanism

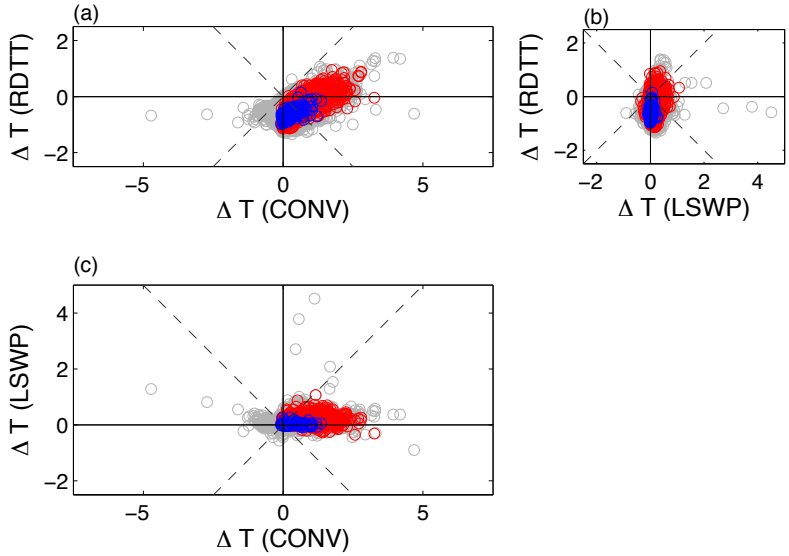

Figure 6.32: The 24-hour cumulative temperature tendencies at 200 hPa taken from the YOTC data set for the RDTT, CONV and LSWP parametrisation schemes. The tendencies have been averaged over 30 dates between 14 April and 6 September 2009, with subsequent start dates separated by five days. The scattered points represent pairs of parametrised tendencies from different spatial locations sampled over: the entire globe (grey); regions with significant convection (red); regions with little convection (blue). The regions are those defined in Section 5.5.1. Figure (a) compares the RDTT and CONV tendencies; (b) compares the RDTT and LSWP tendencies; (c) compares the LSWP and CONV tendencies.

and the spread of the forecasts improves significantly.

The increase in ensemble spread is certainly beneficial. However, is it physically reasonable to decouple the parametrisation schemes in this way? As described in Section 5.2.1, the IFS calls the parametrisation schemes sequentially to ensure balance between the different physics schemes. Additionally, the different schemes in the IFS have been tuned to each other, possibly with compensating errors. An increase of forecast error could be expected on decoupling the schemes. This is indeed the case when SPPTi is implemented, and an increase in forecast bias is observed for many variables. Maintaining balance is particularly important for the cloud and convection parametrisation schemes as they represent two halves of the same process, and because the division of labour between the two schemes is primarily dependent on model resolution (ECMWF, 2012). The convection scheme is in balance with the cloud scheme: the net deterministic tendency for the two schemes is close to zero. *It is plausible that the increase in RMSE in the SPPTi, CONVi and LSWPi experiments, most noticeably for T850 and Z500, could be attributed to the decoupling of the cloud and convection schemes, which could remove this balance.* At T159, the increase in RMSE for T850 only occurs for predictable forecast situations with small forecast spread. It is possible that the CONV and LSWP schemes

have been tuned to be very accurate for these specific cases, so decoupling the two schemes by introducing SPPTi has a detrimental effect on the forecast accuracy. This explanation is supported by the results of the CONVi experiment — the RMS error-spread diagnostic for this experiment had an upward 'hook' at small spreads, indicating a significant increase in error for previously accurate forecast situations. At T639, Figure 6.28(a) indicates that the increase in RMSE for T850 forecasts occurs uniformly across forecast cases. It is interesting that the high resolution model behaves differently to the low resolution model, but further experiments are required to diagnose the cause.

Despite the increase in forecast error, SPPTi clearly has some merit. It skilfully increases the ensemble spread in under-dispersive regions, which tend to be those with significant convection. The resultant ensemble spread evolves in a very similar way to the RMSE — the scheme appears to represent the uncertainty in the forecast very accurately. In the extra-tropics, forecasts remain well calibrated and SPPTi has little effect. Hermanson (2006) considers the difference between parametrised IFS tendencies at T799 and at T95, using the T799 integration as a proxy for "truth". He calculates a histogram for the sum of the temperature tendencies from clouds and convection for each model. Both peaks were centred at between -1 and 0K/day. However, the T95 model was observed to have a narrower, taller peak in the histogram than the T799 model. This indicates that the lower resolution version of the IFS is underestimating the variability of the sum of the cloud and convective tendencies — the low resolution model is too balanced. It would be interesting to perform similar experiments using the current model to test this balance hypothesis. If it is the case that the convection and cloud schemes are too balanced, this could justify the use of SPPTi, and explain the source of the increased skill, especially for the convection diagnostics. It would also be interesting to perform further experiments using SPPTi — it is possible that using a common base pattern for the CONV and LSWP perturbations and applying independent patterns to a smaller degree would be beneficial, retaining a large degree of the balance between convection and clouds, but allowing ensemble members to explore the slightly off-balance scenarios observed by Hermanson (2006) in the higher resolution model.

The RDTTi experiment also merits further investigation as it resulted in a significant improvement in ensemble spread when tested at T159, but with no associated increase in RMSE. In fact, RDDTi resulted in a reduction of error for U200 in regions with little convection, which

is also observed for forecasts using SPPTi in this region — the bias for U200 is reduced in regions with little convection. *The reduction in RMSE indicates that the stochastic perturbations to CONV and RDTT should be uncorrelated.* The radiation scheme is affected by the convection scheme through the cloud parametrisation. However, this coupling between radiation and clouds is through a Monte-Carlo calculation (the McICA scheme), so unless the cloud fractions are systematically wrong, the correlation between the error from the cloud scheme and the error from the radiation scheme should be reduced. The independent pattern approach could therefore be a physically reasonable model for the errors between the convection and radiation tendencies. It would be interesting to test the RDTTi scheme at T639 to see if a reduction in RMSE is observed and if the forecast skill is improved at operational resolution.

In conclusion, modifying the SPPT scheme to allow independent perturbations to the parametrised tendencies has a significant positive impact on the spread of the ensemble forecasts at T159 resolution. The improvement in ensemble spread was also observed at T639, though this was accompanied by an increase in RMSE for some variables. Nevertheless, ECMWF is very interested in performing further experiments using the SPPTi scheme to test its potential for use in the operational EPS.

Bibliography

E. Anderson and A. Persson. *User guide to ECMWF forecast products.* ECMWF, Shinfield Park, Reading, RG2 9AX, U.K., 1.1 edition, July 2013.

J. L. Anderson. The impact of dynamical constraints on the selection of initial conditions for ensemble predictions: Low-order perfect model results. *Mon. Weather Rev.*, 125(11): 2969–2983, 1997.

A Arakawa and W. H. Schubert. Interaction of a cumulus cloud ensemble with the large scale environment, part I. *J. Atmos. Sci.*, 31(3):674–701, 1974.

H. M. Arnold, I. M. Moroz, and T. N. Palmer. Stochastic parameterizations and model uncertainty in the Lorenz'96 system. *Phil. Trans. R. Soc. A*, 371(1991), 2013.

J. V. Beck and K. J. Arnold. *Parameter estimation in engineering and science.* Wiley, New York, USA, 1977.

L. Bengtsson, M. Steinheimer, P. Bechtold, and J.-F. Geleyn. A stochastic parametrization for deep convection using cellular automata. *Q. J. Roy. Meteor. Soc.*, 139(675):1533–1543, 2013.

J. Berner, G. J. Shutts, M. Leutbecher, and T. N. Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, 66(3):603–626, 2009.

J. Berner, T. Jung, and T. N. Palmer. Systematic model error: The impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, 25(14):4946–4962, 2012.

N. E. Bowler, A. Arribas, K. R. Mylne, K. B. Robertson, and S. E. Beare. The MOGREPS short-range ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 134(632):703–722, 2008.

G. W. Brier. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.*, 78(1):1–3, 1950.

J. Bröcker. Reliability, sufficiency, and the decomposition of proper scores. *Q. J. Roy. Meteor. Soc.*, 135(643):1512–1519, 2009.

J. Bröcker, D. Engster, and U. Parlitz. Probabilistic evaluation of time series models: A comparison of several approaches. *Chaos*, 19(4), 2009.

T. A. Brown. Probabilistic forecasts and reproducing scoring systems. Technical report, RAND Corporation, Santa Monica, California, June 1970.

R. A. Bryson. The paradigm of climatology: An essay. *B. Am. Meteorol. Soc.*, 78(3):449–455, 1997.

R. Buizza. Horizontal resolution impact on short- and long-range forecast error. *Q. J. Roy. Meteor. Soc.*, 136(649):1020–1035, 2010.

R. Buizza and T. N. Palmer. The singular-vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, 52(9):1434–1456, 1995.

R. Buizza, M. Miller, and T. N. Palmer. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 125(560):2887–2908, 1999.

R. Buizza, M. Leutbecher, and L. Isaksen. Potential use of an ensemble of analyses in the ecmwf ensemble prediction system. *Q. J. Roy. Meteor. Soc.*, 134(637):2051–2066, 2008.

B. G. Cohen and G. C. Craig. Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. *J. Atmos. Sci.*, 63(8):2005–2015, 2006.

S. Corti, F. Molteni, and T. N. Palmer. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature*, 398(6730):799–802, 1999.

G. C. Craig and B. G. Cohen. Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. *J. Atmos. Sci.*, 63(8):1996–2004, 2006.

D. Crommelin and E. Vanden-Eijnden. Subgrid-scale parametrisation with conditional markov chains. *J. Atmos. Sci.*, 65(8):2661–2675, 2008.

A. Dawson, T. N. Palmer, and S. Corti. Simulating regime structures in weather and climate prediction models. *Geophys. Res. Let.*, 39(21):L21805, 2012.

T. DelSole. Predictability and information theory. part I: Measures of predictability. *J. Atmos. Sci.*, 61(20):2425–2440, 2004.

F. J. Doblas-Reyes, A. Weisheimer, N. Keenlyside, M. McVean, J. M. Murphy, P. Rogel, D. Smith, and T. N. Palmer. Addressing model uncertainty in seasonal and annual dynamical ensemble forecasts. *Q. J. Roy. Meteor. Soc.*, 135(644):1538–1559, 2009.

O. Donati, G. F. Missiroli, and G. Pozzi. An experiment on electron interference. *Am. J. Phys.*, 41(5):639–644, 1973.

J. Dorrestijn, D. T. Crommelin, A. P. Siebesma, and H. J. J. Jonker. Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theor. Comp. Fluid Dyn.*, 27(1–2):133–148, 2012.

J. Dorrestijn, D. T. Crommelin, J. A. Biello, and S. J. Böing. A data-driven multi-cloud model for stochastic parametrization of deep convection. *Phil. Trans. R. Soc. A*, 371(1991), 2013.

ECMWF. *IFS documentation CY37r2.* ECMWF, Shinfield Park, Reading, RG2 9AX, U.K., 2012. http://www.ecmwf.int/research/ifsdocs/CY37r2/.

M. Ehrendorfer. Predicting the uncertainty of numerical weather forecasts: a review. *Meteorol. Z.*, 6(4):147–183, 1997.

T. H. A. Frame, J. Methven, S. L. Gray, and M. H. P. Ambaum. Flow-dependent predictability of the North-Atlantic jet. *Geophys. Res. Let.*, 40(10):2411–2416, 2013.

Y. Frenkel, A. J. Majda, and B. Khouider. Using the stochastic multicloud model to improve tropical convective parametrisation: A paradigm example. *J. Atmos. Sci.*, 69(3):1080–1105, 2012.

N. Gershenfeld, B. Schoner, and E. Metois. Cluster-weighted modelling for time-series analysis. *Nature*, 397(6717):329–332, 1999.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.*, 102(477):359–378, 2007.

M. Goldstein and D. Wooff. *Bayes Linear Statistics, Theory and Methods*. Wiley, Chichester, UK, 2007.

W. M. Gray and R. W. Jr. Jacobson. Diurnal variation of deep cumulus convection. *Mon. Weather Rev.*, 105(9):1171–1188, 1977.

E. Halley. An historical account of the trade winds, and monsoons, observable in the seas between and near the tropicks, with an attempt to assign the phisical cause of the said winds. *Phil. Trans.*, 16(183):153–168, 1686.

J. A. Hansen and C. Penland. Efficient approximation techniques for integrating stochastic differential equations. *Mon. Weather Rev.*, 134(10):3006–3014, 2006.

J. A. Hansen and C. Penland. On stochastic parameter estimation using data assimilation. *Physica D*, 230(1–2):88–98, 2007.

K. Hasselmann. Climate change — linear and nonlinear signatures. *Nature*, 398(6730):755–756, 1999.

L. Hermanson. *Stochastic Physics: A Comparative Study of Parametrized Temperature Tendencies in a Global Atmospheric Model*. PhD thesis, University of Reading, 2006.

H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather Forecast.*, 15(6):559–570, 2000.

P. Hess and H. Brezowsky. Katalog der grosswetterlagen Europas. *Berichte des Deutschen Wetterdienstes in der US-Zone*, 33:39, 1952.

P. Houtekamer, M. Charron, H. Mitchell, and G. Pellerin. Status of the global eps at environment canada. In *Workshop on Ensemble Prediction, 7–9 November 2007*, pages 57–68, Shinfield Park, Reading, 2007. ECMWF.

P. L Houtekamer, L. Lefaivre, and J. Derome. A system simulation approach to ensemble prediction. *Mon. Weather Rev.*, 124(6):1225–1242, 1996.

G. J. Huffman, R. F. Adler, M. M. Morrissey, D. T. Bolvin, S. Curtis, R. Joyce, B. McGavock, and J. and Susskind. Global precipitation at one-degree daily resolution from multisatellite observations. *J. Hydrometeor.*, 2(1):36–50, 2001.

L. Isaksen, M. Bonavita, R. Buizza, M. Fisher, J. Haseler, M. Leutbecher, and L. Raynaud. Ensemble of data assimilations at ECMWF. Technical Report 636, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2010.

C. Jakob. Accelerating progress in global atmospheric model development through improved parameterizations. *B. Am. Meteorol. Soc.*, 91(7):869–875, 2010.

H. Järvinen, M. Laine, A. Solonen, and H. Haario. Ensemble prediction and parameter estimation system: the concept. *Q. J. Roy. Meteor. Soc.*, 138(663):281–288, 2012.

P. Kaallberg. Forecast drift in ERA-Interim. ERA Report Series 10, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2011.

R. J. Keane and R. S. Plant. Large-scale length and time-scales for use with stochastic convective parametrization. *Q. J. Roy. Meteor. Soc.*, 138(666):1150–1164, 2012.

B. Khouider and A. J. Majda. A simple multicloud parameterization for convectively coupled tropical waves. part I: Linear analysis. *J. Atmos. Sci.*, 63(4):1308–1323, 2006.

B. Khouider and A. J. Majda. A simple multicloud parameterization for convectively coupled tropical waves. part II: Nonlinear simulations. *J. Atmos. Sci.*, 64(2):381–400, 2007.

B. Khouider, A. J. Majda, and M. A. Katsoulakis. Coarse-grained stochastic models for tropical convection and climate. *P. Natl. Acad. Sci. U.S.A.*, 100(21):11941–11946, 2003.

B. Khouider, J. Biello, and A. J. Majda. A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, 8(1):187–216, 2010.

C. G. Knight, S. H. E. Knight, N. Massey, T. Aina, C Christensen, D. J. Frame, J. A. Kettleborough, A. Martin, S. Pascoe, B. Sanderson, D. A. Stainforth, and M. R. Allen. Association of parameter, software, and hardware variation with large-scale behavior across 57,000 climate models. *P. Natl. Acad. Sci. U.S.A.*, 104(30):12259–12264, 2007.

R. H. Kraichnan and D. Montgomery. Two-dimensional turbulence. *Rep. Prog. Phys.*, 43: 547–619, 1980.

S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Stat.*, 22(1):79–86, 1951.

F. Kwasniok. Data-based stochastic subgrid-scale parametrisation: an approach using cluster-weighted modelling. *Phil. Trans. R. Soc. A*, 370(1962):1061–1086, 2012.

M. Laine, A. Solonen, H. Haario, and H. Järvinen. Ensemble prediction and parameter estimation system: the method. *Q. J. Roy. Meteor. Soc.*, 138(663):289–297, 2012.

L. A. Lee, K. S. Carslaw, K. J. Pringle, and G. W. Mann. Mapping the uncertainty in global CCN using emulation. *Atmos. Chem. Phys.*, 12(20):9739–9751, 2012.

M. Leutbecher. Diagnosis of ensemble forecasting systems. In *Seminar on Diagnosis of Forecasting and Data Assimilation Systems, 7 - 10 September 2009*, pages 235–266, Shinfield Park, Reading, 2010. ECMWF.

M. Leutbecher and T. N. Palmer. Ensemble forecasting. *J. Comput. Phys.*, 227(7):3515–3539, 2008.

J. W.-B. Lin and J. D. Neelin. Influence of a stochastic moist convective parametrisation on tropical climate variability. *Geophys. Res. Let.*, 27(22):3691–3694, 2000.

J. W.-B. Lin and J. D. Neelin. Considerations for stochastic convective parameterization. *J. Atmos. Sci.*, 59(5):959–975, 2002.

J. W.-B. Lin and J. D. Neelin. Towards stochastic deep convective parameterization in general circulation models. *Geophys. Res. Let.*, 30(4), 2003.

E. N. Lorenz. Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20(2):130–141, 1963.

E. N. Lorenz. Predictability; does the flap of a butterfly's wings in Brazil set off a tornado in Texas? In *American Association for the Advancement of Science, 139th Meeting*, December 1972.

E. N. Lorenz. Predictability Ű– a problem partly solved. In *Proceedings, Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1996. ECMWF.

E. N. Lorenz. Climate is what you expect. eaps4.mit.edu/research/Lorenz/publications, 1997. 52pp.

E. N. Lorenz. Regimes in simple systems. *J. Atmos. Sci.*, 63(8):2056–2073, 2006.

P. Lynch. Richardson's barotropic forecast: A reappraisal. *B. Am. Meteorol. Soc.*, 73(1):35–47, 1992.

A. J. Majda and B. Khouider. Stochastic and mesoscopic models for tropical convection. *P. Natl. Acad. Sci. U.S.A.*, 99(3):1123–1128, 2002.

G. M. Martin, S. F. Milton, C. A. Senior, M. E. Brooks, and S. Ineson. Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *J. Climate*, 23(22):5933–5957, 2010.

D. Masson and R. Knutti. Climate model genealogy. *Geophys. Res. Let.*, 38, 2011.

J.-J. Morcrette. Radiation and cloud radiative properties in the European Centre for Medium Range Weather Forecasts forecasting system. *J. Geophys. Res.-Atmos.*, 96(D5):9121–9132, 2012.

A. H. Murphy. A note on the utility of probabilistic predictions and the probability score in the cost-loss ratio decision situation. *J. Appl. Meteorol.*, 5(4):534–537, 1966.

A. H. Murphy. A new vector partition of the probability score. *J. Appl. Meteorol.*, 12(4): 595–600, 1973.

A. H. Murphy. The value of climatological, categorical and probabilistic forecasts in the cost-loss ratio situatiuon. *Mon. Weather Rev.*, 105(7):803–816, 1977.

A. H. Murphy. A new decomposition of the Brier score: Formulation and interpretation. *Mon. Weather Rev.*, 114(12):2671–2673, 1986.

A. H. Murphy and M. Ehrendorfer. On the relationship between the accuracy and value of forecasts in the cost-loss ratio situation. *Weather Forecast.*, 2(3):243–251, 1987.

J. M. Murphy, D. M. H. Sexton, D. N. Barnett, G. S. Jones, M. J. Webb, M. Collins, and D. A. Stainforth. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430(7001):768–772, 2004.

G. D. Nastrom and K. S. Gage. A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *J. Atmos. Sci.*, 42(9):950–960, 1985.

F. Nebeker. *Calculating the Weather: Meteorology in the 20th Century.* Academic Press, Inc., San Diego, California, U.S.A., 1995.

A. Oort and J. Yienger. Observed interannual variability in the Hadley circulation and its connection to ENSO. *J. Climate*, 9(11):2751–2767, 1996.

T. N. Palmer. A nonlinear dynamical perpective on climate change. *Weather*, 48(10):314–326, 1993.

T. N. Palmer. A nonlinear dynamical perpective on climate prediction. *J. Climate*, 12(2): 575–591, 1999.

T. N Palmer. A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrisation in weather and climate prediction models. *Q. J. Roy. Meteor. Soc.*, 127(572):279–304, 2001.

T. N Palmer. The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Q. J. Roy. Meteor. Soc.*, 128(581):747–774, 2002.

T. N. Palmer. Towards the probabilistic earth-system simulator: A vision for the future of climate and weather prediction. *Q. J. Roy. Meteor. Soc.*, 138(665):841–861, 2012.

T. N. Palmer, A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P Rogel, J.-M. Terres, and M. C. Thomson. Development of a European multimodel ensemble system for seasonal-to-interannual prediction (DEMETER). *B. Am. Meteorol. Soc.*, 85(6):853–872, 2004.

T. N. Palmer, F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell. Towards seamless prediction: Calibration of climate change projections using seasonal forecasts. *B. Am. Meteorol. Soc.*, 89(4):459–470, 2008.

T. N. Palmer, R. Buizza, F. Doblas-Reyes, T. Jung, M. Leutbecher, G. J. Shutts, M. Steinheimer, and A. Weisheimer. Stochastic parametrization and model uncertainty. Technical Report 598, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 2009.

C. Pennell and T. Reichler. On the effective number of climate models. *J. Climate*, 24(9): 2358–2367, 2011.

K. Peters, C. Jakob, L. Davies, B. Khouider, and A. J. Majda. Stochastic behavior of tropical convection in observations and a multicloud model. *J. Atmos. Sci.*, 2013. In press.

R. S. Plant and G. C. Craig. A stochastic parameterization for deep convection based on equilibrium statistics. *J. Atmos. Sci.*, 65(1):87–104, 2008.

B. Pohl and N. Fauchereau. The southern annular mode seen through weather regimes. *J. Climate*, 25(9):3336–3354, 2012.

D. Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge University Press, Cambridge, U.K. and New York, NY, U.S.A., 2002.

L. Ricciardulli and R. R. Garcia. The excitation of equatorial waves by deep convection in the NCAR community climate model (CCM3). *J. Atmos. Sci.*, 57(21):3461–3487, 2000.

D. S. Richardson. Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. Roy. Meteor. Soc.*, 127(577):2473–2489, 2001.

L. F. Richardson. *Weather Prediction by Numerical Process*. Cambridge University Press, The Edinburgh Building, Cambridge, CB2 8RU, England, 2nd edition, 2007.

M. J. Rodwell and T. N. Palmer. Using numerical weather prediction to assess climate models. *Q. J. Roy. Meteor. Soc.*, 133(622):129–146, 2007.

J. Rougier, D. M. H. Sexton, J. M. Murphy, and D. Stainforth. Analyzing the climate sensitivity of the HadSM3 climate model using ensembles from different but related experiments. *J. Climate*, 22:3540–3557, 2009.

M. S. Roulston and L. A. Smith. Evaluating probabilistic forecasts using information theory. *Mon. Weather Rev.*, 130(6):1653–1660, 2002.

M. S. Roulston and L. A. Smith. The boy who cried wolf revisited: The impact of false alarm intolerance on cost-loss scenarios. *Weather Forecast.*, 19(2):391–397, 2004.

F. Sanders. On subjective probability forecasting. *J. Appl. Meteorol.*, 2(2):191–201, 1963.

B. M. Sanderson. A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. *J. Climate*, 24(5):1362–1377, 2011.

B. M. Sanderson, C. Piani, W. J. Ingram, D. A. Stone, and M. R. Allen. Towards constraining climate sensitivity by linear analysis of feedback patterns in thousands of perturbed-physics GCM simulations. *Clim. Dynam.*, 30(2–3):175–190, 2008.

G. Shutts. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Q. J. Roy. Meteor. Soc.*, 131(612):3079–3102, 2005.

G. J. Shutts. Stochastic backscatter in the unified model. Met Office Scientific Advisory Committee Paper 14.5, U.K. Met Office, FitzRoy Road, Exeter, 2009.

G. J. Shutts and M. E. B. Gray. A numerical modelling study of the geostrophic adjustment process following deep convection. *Q. J. Roy. Meteor. Soc.*, 120(519):1145–1178, 1994.

G. J. Shutts and T. N. Palmer. Convective forcing fluctuations in a cloud-resolving model: Relevance to the stochastic parameterization problem. *J. Climate*, 20(2):187–202, 2007.

J. Slingo and T. N. Palmer. Uncertainty in weather and climate prediction. *Phil. Trans. R. Soc. A*, 369(1956), 2011.

S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, Tignor M., and H. L. Miller. Summary for policymakers. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge, United Kingdom and New York, NY, USA, 2007. Cambridge University Press.

D. A Stainforth, T. Aina, C. Christensen, M. Collins, N. Faull, D. J. Frame, J. A. Kettleborough, S. Knight, A. Martin, J. M. Murphy, C. Piani, D. Sexton, L. A. Smith, R. A Spicer, A. J. Thorpe, and M. R Allen. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024):403–406, 2005.

D. J. Stensrud. Upscale effects of deep convection during the North American monsoon. *J. Atmos. Sci.*, 70(9):2681–2695, 2013.

D. J. Stensrud, J.-W. Bao, and T. T. Warner. Using initial condition and model physics perturbations in short-range ensemble simulations of mesoscale convective systems. *Mon. Weather Rev.*, 128(7):2077–2107, 2000.

E. M. Stephens, T. L. Edwards, and D. Demeritt. Communicating probabilistic information from climate model ensembles — lessons from numerical weather prediction. *WIREs: Clim. Change*, 3(5):409–426, 2012.

D. B. Stephenson, A. Hannachi, and A. O'Neill. On the existence of multiple climate regimes. *Q. J. Roy. Meteor. Soc.*, 130(597):583–605, 2004.

D. M. Straus, S. Corti, and F. Molteni. Circulation regimes: Chaotic variability versus sst-forced predictability. *J. Climate*, 20(10):2251–2272, 2007.

K. E. Taylor, R. J. Stouffer, and G. A. Meehl. An overview of CMIP5 and the experiment design. *B. Am. Meteorol. Soc.*, 93(4):485–498, 2012.

M. Tiedtke. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Weather Rev.*, 117(8):1779–1800, 1989.

M. Tiedtke. Representation of clouds in large-scale models. *Mon. Weather Rev.*, 121(11): 3040–3061, 1993.

J. Tödter and B. Ahrens. Generalisation of the ignorance score: Continuous ranked version and its decomposition. *Mon. Weather Rev.*, 140(6):2005–2017, 2012.

N. P. Wedi. The numerical coupling of the physical parametrizations to the "dynamical" equations in a forecast model. Technical Report 274, European Centre for Medium-Range Weather Forecasts, Shinfield park, Reading, 1999.

A. P. Weigel, M. A. Liniger, and C. Appenzeller. Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Q. J. Roy. Meteor. Soc.*, 134(630): 241–260, 2008.

A. Weisheimer, T. N. Palmer, and F. J. Doblas-Reyes. Assessment of representations of model uncertainty. *Geophys. Res. Let.*, 38, 2011.

D. S. Wilks. Effects of stochastic parametrizations in the Lorenz '96 system. *Q. J. Roy. Meteor. Soc.*, 131(606):389–407, 2005.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*, volume 91 of *International Geophysics Series*. Elsevier, second edition, 2006.

K.-M. Xu, A. Arakawa, and S. K. Krueger. The macroscopic behavior of cumulus ensembles simulated by a cumulus ensemble model. *J. Atmos. Sci.*, 49(24):2402–2420, 1992.

T. Yokohata, M. J. Webb, M. Collins, K. D. Williams, M. Yoshimori, J. C. Hargreaves, and J. D. Annan. Structural similarities and differences in climate responses to CO2 increase between two perturbed physics ensembles. *J. Climate*, 23(6):1392–1410, 2010.

Y. Zhu, Z. Toth, R. Wobus, D. Richardson, and K. Mylne. The economic value of ensemble-based weather forecasts. *B. Am. Meteorol. Soc.*, 83(1):73–83, 2002.