SPECIAL PROJECT PROGRESS REPORT

Progress Reports should be 2 to 10 pages in length, depending on importance of the project. All the following mandatory information needs to be provided.

Reporting year	2016
Project Title:	Constraining stochastic parametrisation schemes through coarse graining
Computer Project Account:	spgbtpcs
Principal Investigator(s):	Prof Tim Palmer Dr Hannah Christensen
Affiliation:	University of Oxford
Name of ECMWF scientist(s) collaborating to the project (if applicable)	Dr Antje Weisheimer
Start date of the project:	Jan 2015
Expected end date:	Dec 2017

Computer resources allocated/used for the current year and the previous one

(if applicable)
Please answer for all project resources

		Previous year		Current year	
		Allocated	Used	Allocated	Used
High Performance Computing Facility	(units)	2,000,000		3,000,000	311,845
Data storage capacity	(Gbytes)	2,400		3,600	

Summary of project objectives
(10 lines max)

Stochastically Perturbed Parametrisation Tendencies (SPPT) is an attractive stochastic parametrisation scheme due to its ease of use and beneficial impact on ensemble forecast reliability. However, despite its popularity, the SPPT scheme remains ad hoc in its assumptions. For example, the imposed spatial and temporal correlations have not been derived from theory or observation, and have simply been tuned to give the best results. SPPT also does not distinguish between different parametrisation schemes, and assumes the errors from each scheme are perfectly correlated. This project seeks to address these shortcomings using coarse graining experiments: a high resolution data set will be coarse grained to the resolution of a NWP model, and the characteristics of the 'error' between high resolution data set and NWP tendencies will be calculated

Summary of problems encountered (if any)
(20 lines max)

Since the previous report, the new version of openIFS has been released, including the IFS SCM for CY41R1, which has allowed us to progress with the project. However, we have recently discovered an issue with one of the logical flags in the model.

For our setup, we would like to force the IFS SCM using surface sensible and latent heat fluxes instead of having an interactive land surface scheme. We have therefore set the flag LESURF = .F. to turn off the land surface scheme and LSFCFLX = .T. to indicate that we will supply the fluxes. However, the LSFCFLX flag is not routinely used by ECMWF, and we do not think it has been consistently maintained as a valid option when other changes have been made to the model since CY23.

If these two flags are specified, the model often crashes over land points. We suspect it is to do with the calculation of the skin temperature in the model. In fact, *the skin temperature output by the SCM is independent of whether these two flags are set to be true or false, and also independent of the input soil temperature fields, input skin temperature and supplied sensible and latent heat fluxes (i.e. inverting the sensible and latent heat fluxes has no impact on the skin temperature).*

We do not know what the problem is, but have implemented a crude hack whereby the skin temperature is read in from the SCM input file. This means we can specify a skin temperature that is consistent with our surface fluxes. We are also in contact with scientists at ECMWF to try to rectify this fault.

Summary of results of the current year (from July of previous year to June of current year)

Our high resolution data set of choice is the CASCADE dataset, provided to us by Chris Holloway (University of Reading). This is a high resolution integration carried out with the limited area version of the Met Office Unified Model. The dataset covers the Warm Pool region, spanning 40-180E, 20S-20N at a horizontal resolution of 4km for ten days in April 2009. The model uses semi-Lagrangian, non-hydrostatic dynamics and 3D Smagorinsky mixing. The convection parametrisation scheme is switched on in the run, but the closure is such that the convection scheme is only active in low CAPE environments, and so tends to only produce shallow/congestus cloud.

The CASCADE dataset is coarse grained to a T639 resolution to provide the input atmospheric data fields for the IFS SCM. Atmospheric forcing fields (advective tendencies, geostrophic winds) are derived from the data. Any missing fields or parts of fields, such as atmospheric variables in the upper stratosphere and surface parameters, are taken from the MARS archive.

We have put in place much of the computing framework to allow us to coarse grain the high resolution CASCADE dataset to provide input files for running the IFS SCM. The process has been tested over the first few time steps with promising results. We are at the stage where we are ready to start performing the analysis over the whole CASCADE data set.

Preliminary results:

The first eight time steps have been analysed so far to test the procedure.

Let us write the SPPT scheme as

$$T = D + (1+e)\sum_{i=1}^{5} P_i$$

where *T* is the total tendency for a particular prognostic variable (as a function of height), *D* is the dynamics tendency (as a function of height), P_i is the tendency from the *i*th parametrisation scheme (as a function of height)and *e* is the spatially and temporally correlated stochastic perturbation, (constant in height).

We can identify each of the terms in this equation as follows:

T – the 'true' total tendency as indicated by the CASCADE data set
D – the IFS dynamics tendency at T639. This is provided by the SCM.
P_i – the IFS physics tendency for the *i*th parametrisation scheme. This is provided by the SCM.

This leaves e, the instantaneous perturbation that would give us an optimal representation of the error in the SCM physics tendencies. This can be deduced from the CASCADE data and the SCM tendencies. We form the vector equation:

$$T - D - \sum_{i=1}^{5} P_i = e\sum_{i=1}^{5} P_i$$

where the left hand side represents the error between the IFS SCM tendency and the CASCADE 'true' tendency at a particular horizontal position. This vertical profile of error is available for each of the prognostic variables (T, U, V, q). The right hand side indicates that SPPT represents this vertical error profile as a constant multiplied by the sum of the physics tendencies. We solve this over-constrained vector equation simultaneously for all prognostic variables by minimising the squared residual. The solution is the optimal *e* as a function of horizontal position and time – this allows us to see what the *e* used in SPPT should look like in terms of its statistical properties.

Figure 1 shows this optimal stochastic perturbation field for the first time step. The pattern is smoothly varying. The day-night boundary (at approx. 90 E) is visible in the increased errors over land in the night regions. This is because the tendencies are smaller at night, so the fractional error (*e*) increases.

The statistics of this error field have been calculated, averaged over the first eight timesteps:

mean: -0.77
standard deviation: 0.44
temporal decorrelation: 6.87 hours
spatial decorrelation: 750 km

These are comparable to the characteristics of the fastest varying pattern (standard deviation 0.52; temporal decorrelation 6 hrs; spatial decorrelation 500km). The mean of -0.77 is concerning, as it indicates a systematic reduction in magnitude of the perturbations coming from the physics scheme. This will be investigated further once more data is available. One option is to de-bias the forecasts prior to analysis to ensure that SPPT is only accounting for the random component of the model error.

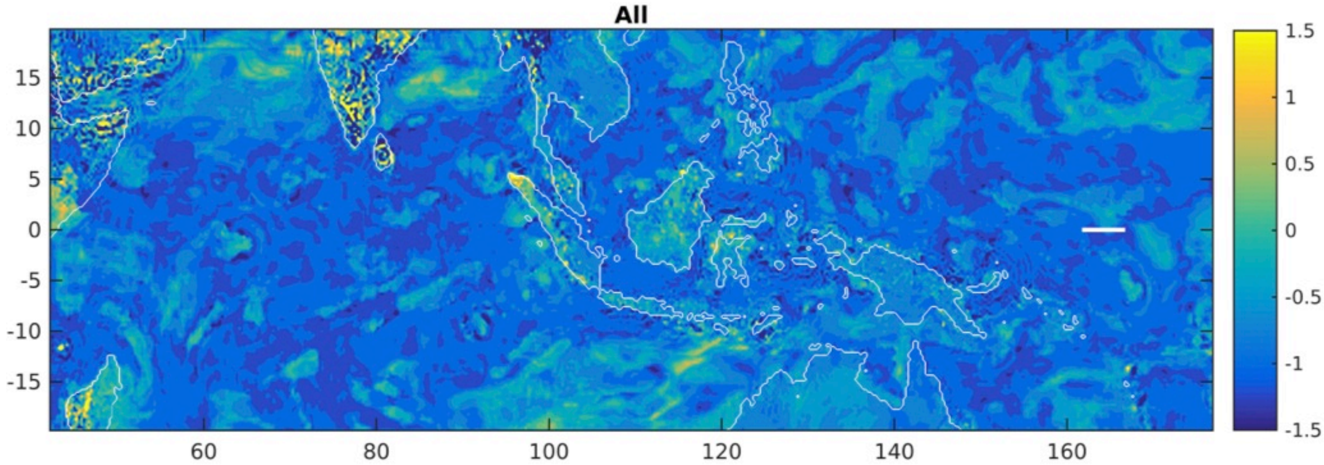

Fig 1: Optimal e perturbation for the first time step.

We have also repeated the analysis for the proposed independent SPPT whereby an independent random number is used to multiply the tendencies from each parametrisation scheme. The following equation is solved:

$$T - D - \sum_{i=1}^{5} P_i = \sum_{i=1}^{5} e_i P_i$$

to give a separate e_i for each parametrisation. The results for the first time step are shown in Figure 2. A number of interesting results are apparent. Firstly, different parametrisation schemes seem to have different error characteristics: for example, radiation appears to have large magnitude errors on smaller spatial scales than the other schemes. Secondly, in regions where the convection scheme did not trigger (white in panel three), the error in the ‘large scale water processes’ scheme tendencies is different to in regions where convection did trigger. This could indicate that the convection parametrisation scheme has uncertainties that are not well represented by (independent) SPPT, for example, errors in triggering of convective events. This has motivated investigation into a stochastic parametrisation scheme for triggering of convection, currently under development in our group at Oxford. From a practical point of view, it could highlight that we should treat the two schemes together in independent SPPT.

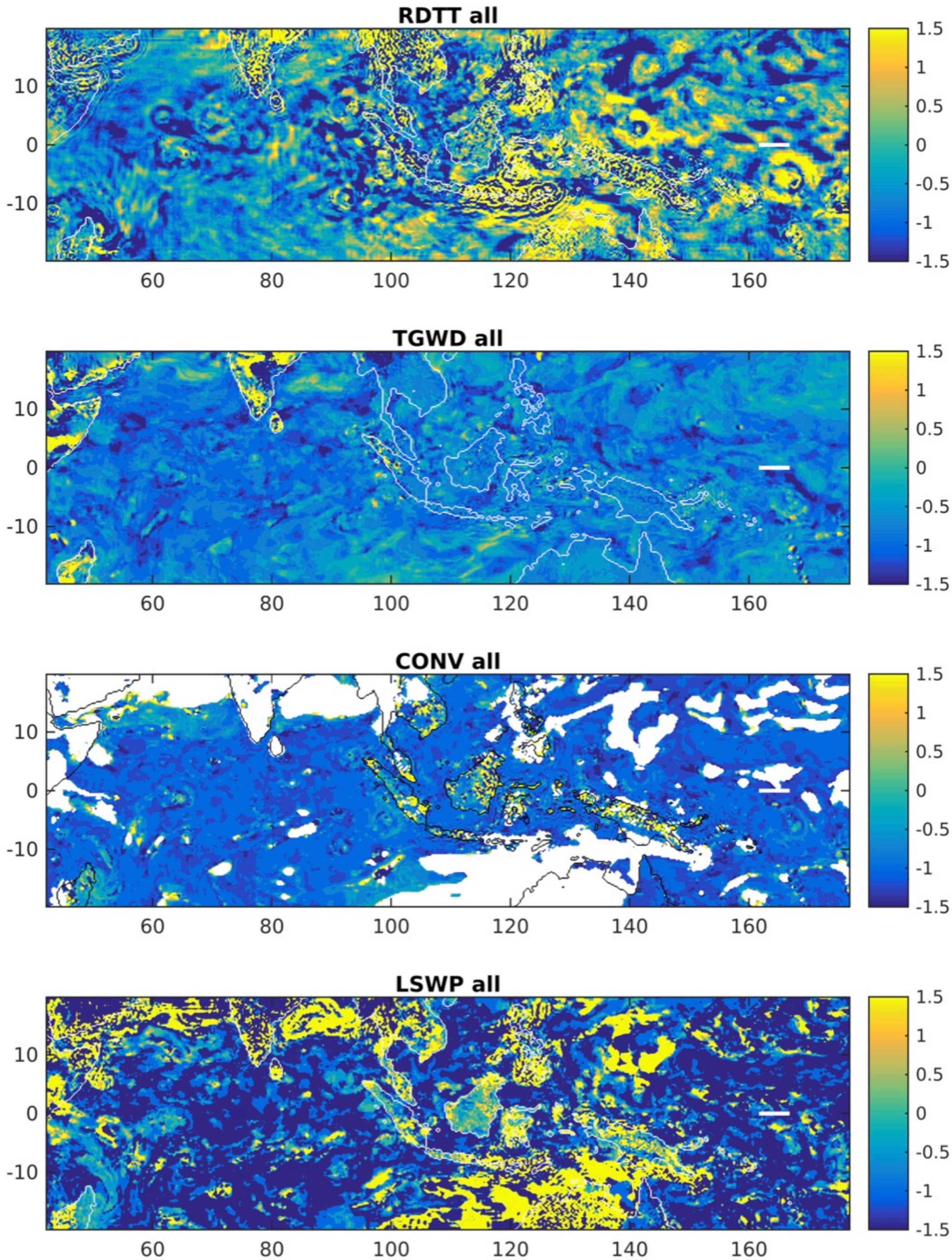

Figure 2: Optimal perturbations for independent SPPT.

List of publications/reports from the project with complete references

We have presented preliminary results from this project at a number of workshops:

HD(CP)^2 conference on Understanding Clouds and Precipitation, Berlin, Germany. Feb 15-19 2016
ECMWF/WWRP Workshop on Model Uncertainty, Reading, U.K. 11-15 April 2016
Workshop on Cloud and Boundary Layer Dynamics, Zurich, Switzerland. 13-15 June 2016

Summary of plans for the continuation of the project

(10 lines max)

Now that the software is in place to perform the coarse graining and automate the running of the SCM over the coarse-grained domain, and has been tested, we anticipate rapid progress will be made. Once the coarse-graining has been carried out over the whole dataset and the IFS SCM has been run over the data, we have the large task of analysing this data and extracting the statistics of the optimal perturbation. We seek to constrain the parameters in SPPT using this data, but will also consider whether the error statistics show a diurnal cycle or are different over land or ocean points, We will also consider whether the error is correlated to any prognostic variables (i.e. there exists a predictable component of the error, which could hint at a missing process). Finally, we will test SPPT and independent SPPT with the derived perturbation characteristics to test whether the forecast skill is improved.