# 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** | 2017 |
| **Project Title:** | Constraining stochastic parametrisation schemes through coarse graining |
| **Computer Project Account:** | spgbtpcs |
| **Principal Investigator(s):** | Hannah Christensen<br>Tim Palmer |
| **Affiliation:** | University of Oxford |
| **Name of ECMWF scientist(s) collaborating to the project** (if applicable) | Andrew Dawson |
| **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) | 3,000,000 | | 4,000,000 | 2,136,405 |
| **Data storage capacity** | (Gbytes) | 3,600 | | 4,800 | |

## Summary of project objectives

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)

N/A

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

The software is now in place to perform the coarse graining and automate the running of the SCM over the coarse-grained domain, and has been thoroughly tested. The coarse graining has been carried out over the whole high resolution dataset and the IFS SCM has been run over the data. Before analysing the statistics of the optimal perturbation for SPPT, we have taken the opportunity to validate the proposal to use high-resolution model data to force a single column model.

Introduction

To justify our proposal to use high-resolution simulations to force a SCM, it is helpful to first consider the state of the art. A key challenge in using SCM is deriving the required forcing data, which includes the initial conditions and advected tendencies of the prognostic variables. The canonical technique is to use forcing datasets derived from observations, such as data from the Atmospheric Radiation Measurement (ARM) program. The evolution of the SCM can then be directly compared to the observed dataset. To derive the required forcing datasets, high quality observational data is required. An array of observation platforms is required to estimate the advective fluxes of prognostic variables. Balloon soundings are used to measure the vertical profiles of the prognostic variables, while surface instruments provide further measurements. There are only a handful of sites globally with this capability. In order to derive the observational forcing datasets, assumptions and approximations must be made. Many of these assumptions are subjective and can affect the final analysis. For example, two separate groups derived analyses for the Tropical Ocean and Global Atmosphere Coupled Ocean-Atmospheric Response Experiment (TOGA COARE), which ran from November 1992 to February 1993. The different derivation process led to large differences in the moisture budget: the average precipitation over the period was reported as 5.7–6.1 mm/day by Lin et al, (1996) and as 10.5–11.8 mm/day by Frank et al (1996).

An alternative is to use reanalysis products to derive the required forcing fields. Such products have the required temporal and spatial frequency, and can be produced for long periods at any global location, which is not possible with observationally derived data sets. However, the models used to produce the reanalysis products can exhibit large biases in the representation of clouds and precipitation (the operational ECMWF analysis has resolution of order 50km, so these processes are parametrised). Xie et al. (2003) demonstrate large differences between SCM forced by observations and those forced by reanalysis products, which they relate to errors in the ECMWF model's representation of clouds and precipitation.

Recent years have seen an increase in the production of large-domain, convection resolving (or at least convection permitting) atmospheric simulations, including both cloud resolving models (CRM) with resolutions of a few km (Holloway et al, 2012) and large eddy simulations (LES) with resolutions of order 100 m. These simulations cover large domains. For example, the ‘Cascade’ project produced high-resolution CRM simulations using the UK Met Office Unified Model (UM) at 1.5 km and 4 km resolution over a domain spanning 15,500 km by 4,500 km over the Indian Ocean, Warm Pool and tropical west Pacific.

The availability of these high-resolution datasets suggests a new alternative, namely the possibility to combine high-resolution data with the SCM framework to further parametrisation development. Unlike the traditional partnership between LES and SCM, whereby observations are used to drive both the SCM and a LES that covers the same region (Randall et al, 1996), we propose the use of large domain high-resolution simulations to derive the forcing datasets required by the SCM. This allows for perfect model studies to be carried out, whereby the high-resolution data takes the place of observations, and becomes the target which the parametrised SCM seeks to match.

This approach has several advantages over using observations. Firstly, the high-resolution datasets cover a large spatial area. This allows for parametrisation schemes to be tested across many climate and weather regimes and over a wide range of land/sea boundary conditions. By considering SCM simulations as a function of spatial location, spatial correlations in biases can be identified, which may point to missing processes. It is also increasingly the case that high resolution datasets have excellent temporal coverage, allowing such simulations to rival the observational data from IOP. Evaluating SCM in a perfect-model framework also simplifies the evaluation process, as all model fields are available for comparison (Bechtold et al, 2000). This includes quantities that are hard to measure, such as the vertical distribution of liquid water and ice and the in-cloud temperature (Noh et al 2013).

We consider the possibility of using high-resolution atmospheric simulations as an alternative to observations for driving a SCM. To illustrate the procedure, we use data from the Cascade project produced using the UK Met Office UM at 4km resolution. To demonstrate the generality of the procedure, we use this data to derive forcing data sets for the European Centre for Medium Range Weather Forecasting (ECMWF) Integrated Forecasting System (IFS) SCM.

## Method

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 openIFS SCM, CY40R1. 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 test the validity of driving a SCM using high resolution data by evaluating SCM simulations that span the whole ten-day high resolution simulation. We consider sub-regions in the Cascade dataset that highlight different climate regimes. Here, we describe results only from the West Pacific region (5S-5N, 155-178E), for the first 72 hours of simulation.

Results: Impact of Nudging

We run an independent SCM over each tile in the West Pacific region defined above (~2000 SCM tiles). We evaluate the difference between each SCM and the evolution of the Cascade model averaged over that tile. Figures indicate the time-series of bias averaged over the region, as a function of height.

We first consider the impact of nudging on the evolution of the SCM. We compare the biases from a free-running SCM with those from the SCM nudged towards the Cascade fields with a time scale of 3 hours or 24 hours respectively. As expected, nudging the SCM towards Cascade reduces the degree to which the SCM drifts away from Cascade, as shown in Fig 1 for the average temperature in the region, though biases in the troposphere are low for all simulations.

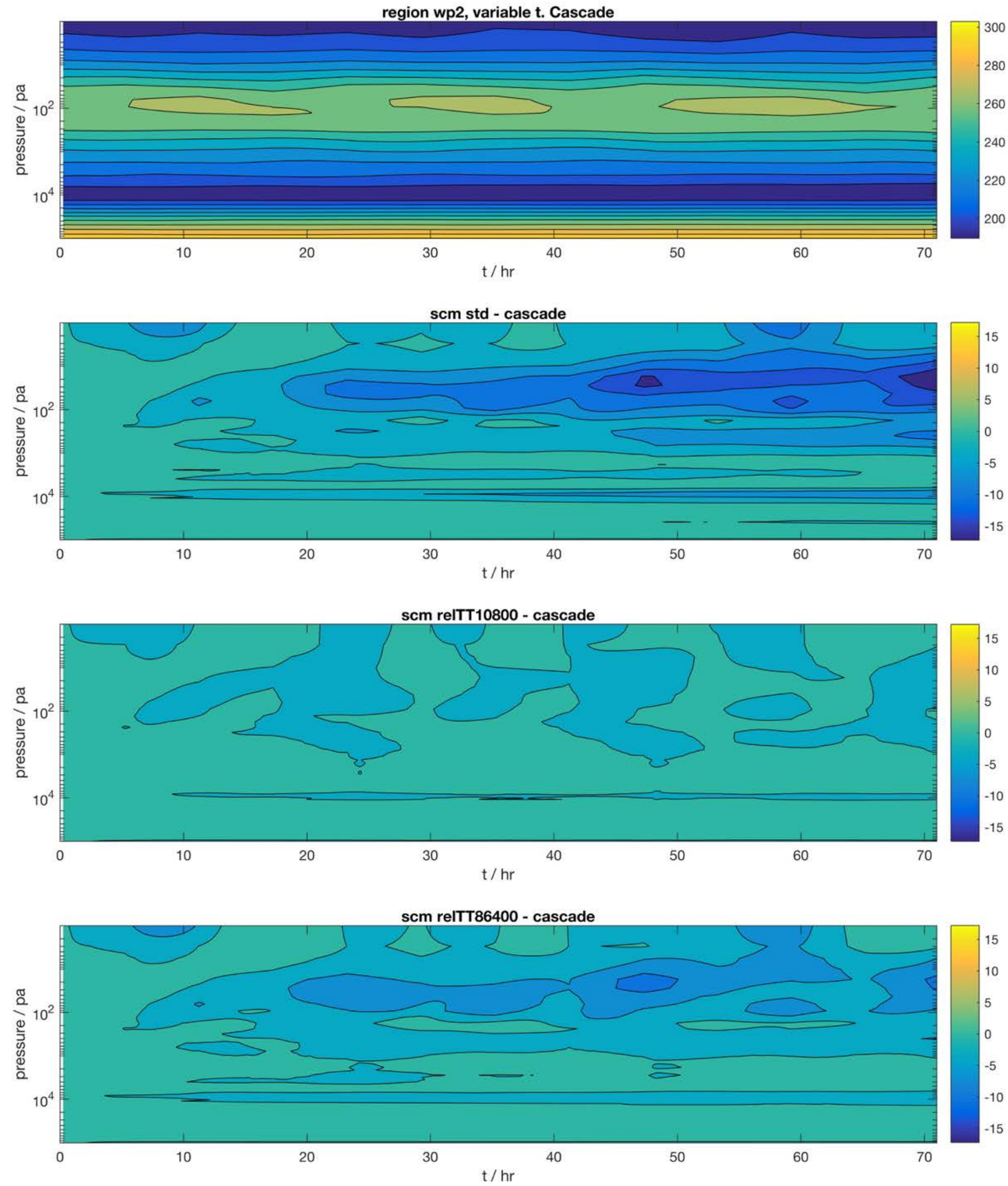


Fig 1: Impact of nudging on mean temperature for SCM over West Pacific region. (a) Mean evolution in Cascade. (b) Bias in free running SCM. (c) Bias in SCM nudged towards Cascade fields with time scale of 3 hours. (d) Bias in SCM nudged towards Cascade fields with time scale of 24 hours.

However, nudging degraded the mean precipitation in this region (Fig 2), possibly due to non-conservation of moisture from nudging the humidity fields towards Cascade. Considering the mean precipitation fields in Fig 2 reveals several interesting points. Firstly, the IFS SCM has a strong spin up period over approximately four time-steps (one hour). During this time, the precipitation flux is considerably higher than at other times, before reducing to a more reasonable value. Similarly, the Cascade simulation also appears to have a spin up period. Over the first 24 hours, the mean precipitation in Cascade increases from zero towards a value that more closely matches the SCM.

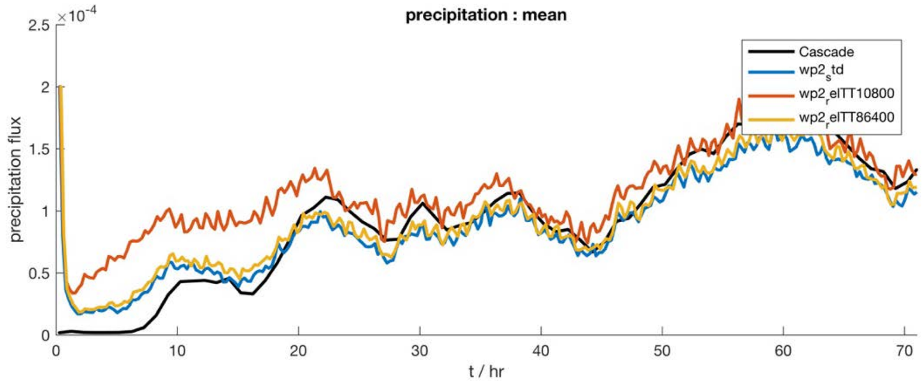


Fig 2. Mean precipitation over the West Pacific region as a function of lead time. Black: Cascade. Blue: free running SCM. Red: SCM nudged with 3-hour time scale. Yellow: SCM nudged with 24-hour time scale.

Consideration of the humidity field shows that this high precipitation flux is concurrent with the formation of a strong dry bias in the SCM compared to Cascade (Fig 3). This bias extends throughout the troposphere, though the lowest levels show a moistening at later times compared to Cascade. To verify that this is a true bias in the SCM, and not a bias in Cascade, we also validated the SCM against ERA interim data from the same period: this also indicates a dry bias in the SCM, with a moist bias close to the surface.

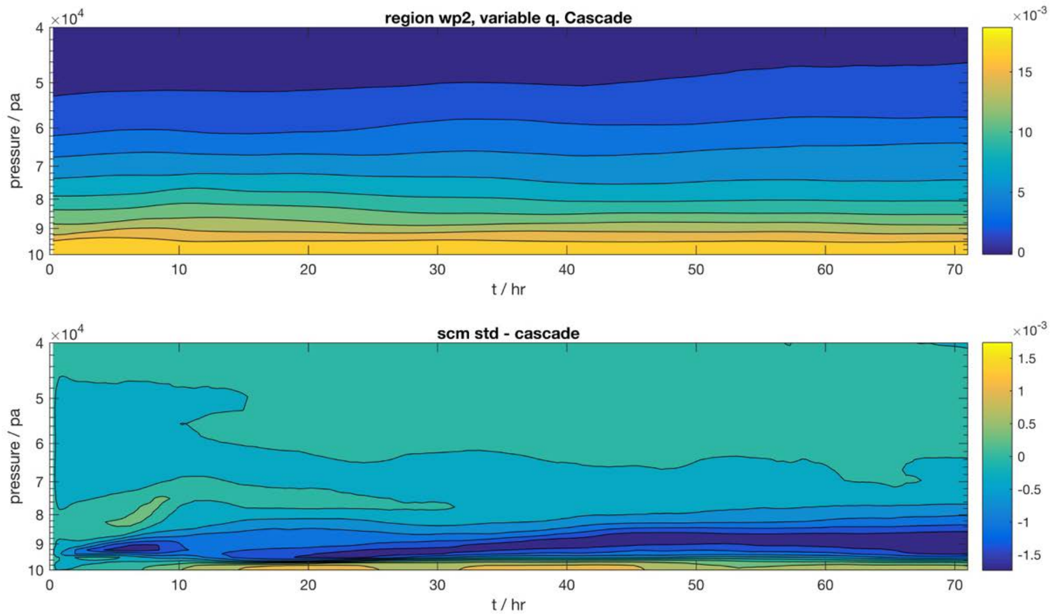


Fig 3: Impact of nudging on mean humidity for SCM over West Pacific region. (a) Mean evolution in Cascade. (b) Bias in free running SCM. (c) Bias in SCM nudged towards Cascade fields with time scale of 3 hours. (d) Bias in SCM nudged towards Cascade fields with time scale of 24 hours.

## Results: Impact of Geostrophic Wind Forcing

An important forcing field for the SCM is the geostrophic wind forcing. This field characterises the pressure gradient force acting on the column. This field required careful calculation from the Cascade simulation, as it is calculated as the small difference between two large terms. If this field is not available, it is not uncommon for SCM users to set this field either to zero, or to the total wind field. To demonstrate the importance of including the ‘correct’ geostrophic winds, we forced the SCM firstly with the calculated fields, before testing each of these alternatives. The results are shown in Fig 4.

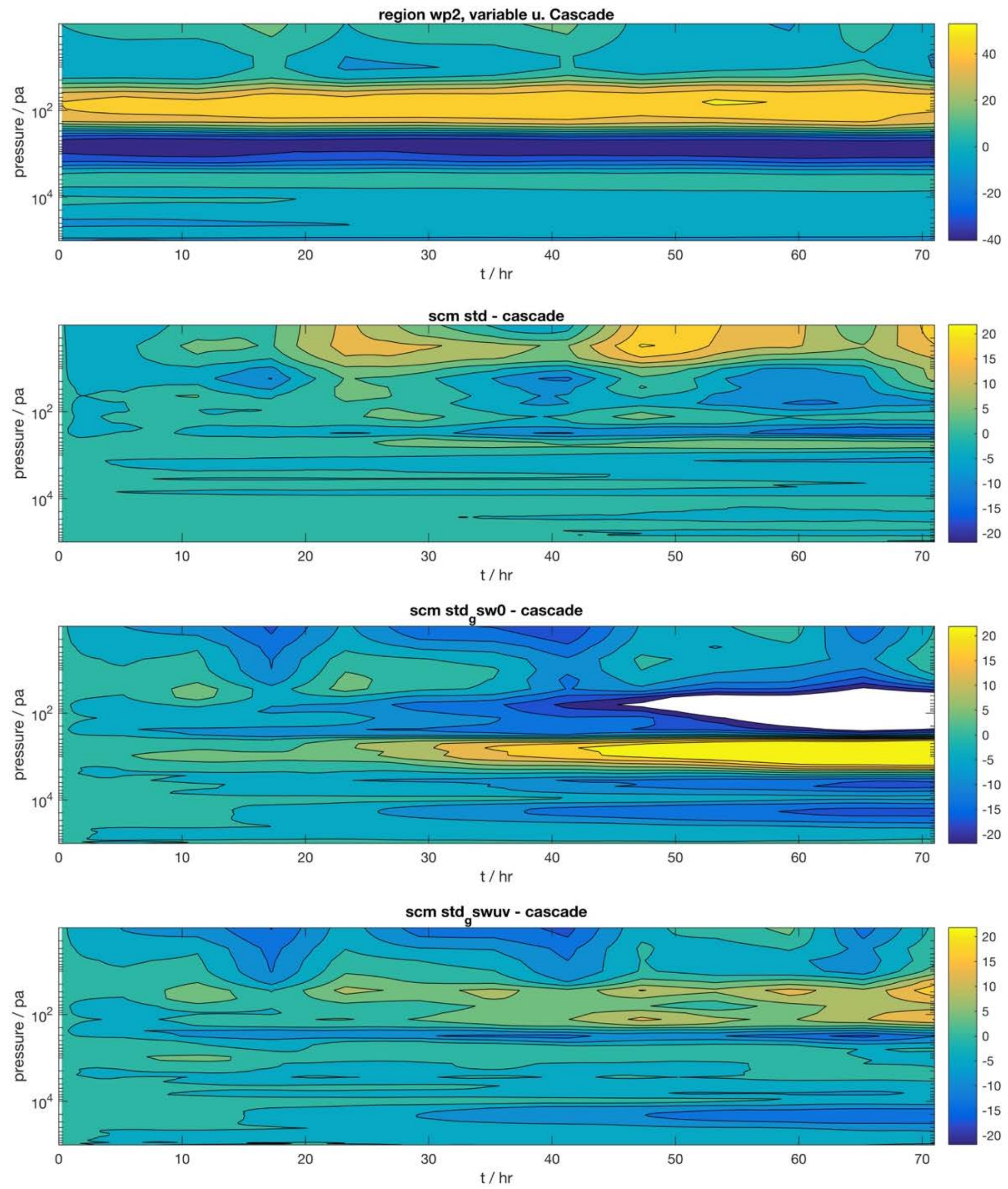


Fig 4: Impact of chosen geostrophic wind forcing on mean zonal wind for SCM over West Pacific region. (a) Mean evolution in Cascade. (b) Bias in SCM with geostrophic winds derived from Cascade. (c) Bias in SCM with geostrophic winds set to zero. (d) Bias in SCM with geostrophic winds set to the total wind field.

The biases change considerably if the incorrect geostrophic wind forcing is used. In the troposphere in particular, using the ‘correct’ geostrophic wind forcing leads to significantly smaller wind biases than if the geostrophic wind forcing is set to zero, or to the total wind field. Setting the geostrophic wind field to zero leads to the largest biases developing, particularly in the upper stratosphere.

## Conclusions

While analysis is ongoing, these results demonstrate the potential for using high-resolution model simulations to derive forcing fields for SCM. By running many SCM over neighbouring grid boxes with similar boundary conditions, robust statistics can be derived, and biases clearly identified.

Running the SCM for a range of different boundary conditions can highlight the state-dependence of biases. For example, running SCM over an ocean region to the West of Australia revealed no such dry bias, indicating the dry bias in the West Pacific is dependent on local processes, likely linked to convection. Updates to parametrisation schemes can therefore be evaluated over a range of boundary conditions in a controlled environment, in which the scientist can prescribe the dynamical tendencies as required.

## List of publications/reports from the project with complete references

As we near the end of this project, two publications are planned. The first will motivate and describe the coarse graining procedure, as well as outline results that demonstrate the potential for driving a SCM using these coarse grained fields, as described above. The second will describe the use of this technique to constrain the SPPT stochastic parametrisation scheme, as has been outlined in previous progress reports.

## Summary of plans for the continuation of the project

(10 lines max)

- Finish writing ‘implementation’ paper
- Complete analysis of Cascade data with a view to constraining and informing the SPPT scheme
- Evaluate performance of SPPT scheme with parameters constrained from Cascade dataset
- Write up into second paper.