Chapter 3

Sub-selection of CMIP5 GCMs for downscaling over Singapore

Authors: Carol McSweeney[1], Raizan Rahmat[2], Grace Redmond[1], Charline Marzin[2], James Murphy[1], Richard Jones[1], Wee-Kiong Cheong[2], See-Yee Lim[2] and Xiangming Sun[2]

Met Office internal reviewers: Ben Booth[1], James Murphy[1], David Sexton[1] and Matt Palmer[1]

1 - Met Office, Exeter, UK

2 - Centre for Climate Research Singapore, Singapore

Contents

3.1 Introduction

Uncertainties in our knowledge of the response of the climate to increasing atmospheric concentrations of greenhouse gases are known to be large for some variables, due partly to differences in the formulation of climate models used to generate this information. In order to provide information about future climate that is consistent with current science and therefore relevant to policy applications, regional projections for use in impacts assessments and planning must be designed to account for this uncertainty.

The Coupled Model Inter-comparison Project 5 (CMIP5) brings together a co-ordinated set of climate model simulations from climate modelling centres around the world (Taylor et al, 2012). Like the preceding CMIP3 experiment, this has allowed a co-ordinated assessment of each model's relative strengths and weaknesses, as well as the quantification of the range of projections across the contributing models. In contrast to CMIP3, the suite of datasets distributed via CMIP5 includes 6-hourly instantaneous fields of prognostic variables from which lateral boundary conditions (LBCs), required for driving regional climate models (RCMs), can be derived. This provides an opportunity to generate higher-resolution baseline and future climates by downscaling with multiple combinations of global and regional climate models, allowing exploration of the implications of modelling uncertainty in ensembles of high-resolution projections for one or more regions of the world.

Sub-selecting a representative sub-set from available general circulation models (GCMs) provides an efficient approach to generating a set of higher-resolution regional climate projections which represent the range of future climates indicated by the full ensemble (e.g. McSweeney et al., 2012). In addition to providing a framework for a computationally efficient project, the provision of a sub-set of representative models provide a dataset which is more manageable in terms of storage, distribution and application to impacts assessments.

The sub-selection process also provides the challenge and opportunity to discount any models which we find unsatisfactory in their representation of key processes or features of climate. The down-weighting or exclusion of GCMs has been explored in a number of studies (e.g. Tebaldi et al., 2005; Greene et al., 2006; Tebaldi and Sanso, 2009; Watterson and Whetton, 2011; Sexton et al., 2012). However this is a challenging problem open to further investigation (see Knutti, 2010 for a discussion of the issues), hence IPCC has to date avoided attempting to weight individual models in its presentations of future projections. While it is clear from the global analyses of CMIP3 GCMs in Gleckler *et al* (2008) that some models perform substantially better and worse than others when compared via a range of global performance metrics, within such an assessment lies a range of relative merits of each different model across a wide range of processes and phenomena. Here, we focus on a regional performance assessment for SE Asia. Rather than attempting quantitative estimation of individual model weights, we capitalise on the differences in performance in order to exclude a subset of model projections which, for our more specific application, do not represent the most relevant processes realistically enough to generate plausible scenarios of future change.

This report describes the methodology employed to select a sub-set of 10 of the available CMIP5 GCMs which will be used to drive the Hadley Centre's latest Regional Climate Model (RCM), HadGEM3-RA (Moufouma-Okia and Jones, *2015*; Karmacharya et al, *In Review*), in order to generate higher-resolution scenarios of future climate for

Singapore under two scenarios of future atmospheric GHG concentrations (RCP4.5 and RCP8.5).

Section 3.2 describes the rationale for the sub-selection methodology. Section 3.3 summarises the results on the model evaluations and consequent elimination decisions (a full account of the evaluation of the models against all criteria can be found in Appendix 3.3). Section 3.4 describes the implementation of the sub-selection from remaining models, and results and their implications are summarised in Section 3.5.

3.2 Rationale for sub-selection methodology

McSweeney et al (2012) addressed ensemble sub-selection using a 2-stage process: The models were first assessed in their simulation of a realistic baseline climate, with unsatisfactory models being eliminated before; secondly, a subset of *n* models was selected to span the range of projected changes in temperature and precipitation.

As described in McSweeney et al (2012), a regional scale projection can be eliminated if the relevant model clearly fails to represent a feature or process that is key to representing the local large-scale-regional climate realistically. The criteria for elimination are underpinned by the following key guidance points from Knutti (2010):

- Metrics and criteria for evaluation should be demonstrated to relate to projection
- It may be less controversial to down-weight or eliminate models that are clearly unable to mimic important processes than to agree on the best model.
- Process understanding must complement 'broad brush metrics'.

For each available model, the McSweeney et al. (2012) approach requires us to assess for every model, whether the shortcomings in its simulation of baseline climatology are significant enough to render its projections as 'implausible', and thus should be eliminated. For this study, we use an amended version of the McSweeney et al (2012) approach where, by assessing the baseline realism and characteristics of the projection simultaneously, we only need to consider the 'plausibility' of a model if its projections out-lie the ensemble. This is because elimination of an outlier would restrict the range of future outcomes that we could capture with a subset. If Model X sits well within the range of future projections compared with other models, then we can easily avoid including this model in favour of others which give similar projections, but in which we have more confidence, avoiding the difficult question of whether the projection should be considered implausible. A more difficult decision arises if the projections from Model X lie outside the range of the rest of the ensemble; in this case we must make the difficult decision based on our best knowledge. However, by employing this approach we minimise the burden and impact of this decision-making process. This approach to the decision making is summarised as a matrix in Figure 3.1. Here the most important and difficult decision occurs in allocating a model its position on the performance scale ('y-axis' in Figure 3.1) between 'Implausible' and 'Significant Biases'. Our criterion in this situation is that if it is clear that a model fails to simulate a large-scale process that is a significant driver of the climate of a region, for example extra-tropical storm tracks or monsoonal circulations, then this model is unlikely to correctly capture how global climate change will manifest itself over the region. It will be unlikely to, for example, transport realistically any additional heat or moisture resulting from climate change into or out of the region. Where we find evidence of very significant shortcomings of this nature in a model then we feel it reasonable to class it as 'Implausible' and eliminate it.

In addition, we class a model as ‘Implausible’ if it shows ‘significant biases’ across multiple climatological features of interest, even if its simulation of any individual phenomenon is not quite poor enough to warrant exclusion on that basis alone. Here, the assumption is that widespread biases would be an indicator of serious shortcomings in the model’s general representation of climate system processes. Models which show shortcomings in a more restricted subset of emergent properties are not eliminated outright, but classed as having ‘Biases’ or ‘Significant Biases’.

		Model Projections	
		Outlier	**Other models predict similar outcomes too.**
Model Performance	**Model suffers shortcoming(s) sufficiently serious to significantly reduce our confidence in its projections (‘Implausible’)**	Exclude: we should carefully document justification for this, however, as exclusion will affect the range of outcomes.	Exclude: We can avoid using these models without much affecting the range of projected outcomes.
	Model suffers significant shortcomings which we cannot clearly link to confidence in its projections. (‘Biases/Significant Biases’)	Include: we do not have strong enough evidence to exclude these outcomes from the projections.	Exclude: We can avoid using these models without much affecting the range of projected outcomes.
	Model performance is satisfactory (‘Satisfactory’)	Include	Include

Figure 3.1: Decision making matrix for potential elimination of ensemble members.

The assessment of the credibility of CMIP5 historical simulations for the assessment of model realism combines analyses for this report with results from published literature, where available. A limitation on this study is that we cannot fully validate each CMIP5 model within the scope of this study – the CMIP5 archive is still relatively new and results will continue to appear in the literature over the coming years. While the cited published literature offers rigorous and well documented analyses of model behaviour by experts in their field, the studies are often limited to an incomplete set of CMIP5 models.

The validation criteria were agreed between scientists with knowledge of local meteorology and its large scale influences, and scientists with prior experience in GCM sub-selection. The proposed criteria aim to identify some key aspects of large scale model behaviour which are important for simulating the local meteorological characteristics realistically, in terms of both the climatology and variability, where possible. The large-scale aspects of climate that we assess in the GCMs will be passed to the regional model via the lateral boundary conditions. Smaller scale processes – for example, the local scale convective rainfall events that can lead to very extreme rainfall events of short duration, and processes relating strongly to interactions with the land surface will be determined by the regional model. These can only be assessed after downscaling.

Not all models in the CMIP5 archive are available for downscaling – a subset of 28 of around 40 models have the 6-hourly variable fields available that are required to drive regional models (see Appendix 3.2 for a list of all models). However, we present the results of assessments for all models for which historical simulations are available in order to give a better impression of the relative strengths and weaknesses. In several cases, particular variables were missing from the archive at the time of assessment.

3.3 Summary of performance and candidate models for elimination

All models were assessed for their ability to represent key aspects of climate in the region including; the south-west and north-east monsoon circulations, the migration of the Inter-Tropical Convergence Zone (ITCZ) and regional rainfall patterns, sea-surface and near-surface air temperatures and key modes of variability – the El Nino Southern Oscillation (ENSO) and The Madden Julian Oscillation (MJO). Tables 3.1 and 3.2 summarise the assessment carried out and the results for each model under each of the categories considered full details of the evaluation of each of these characteristics can be found in Appendix 3.3.

Notably, 2 models were found to have such poor representation of the south west monsoon circulation, that they were classed as 'implausible' – these models were MIROC-ESM and MIROC-ESM-CHEM (see Figure 3.2). In these models, regions of strong the 850hpa flow that characterises the southwest monsoon does not extend far enough east to reach south-east Asia. We therefore conclude that those models cannot provide a useful indication of how any changes to this important feature of the regional climate might affect South East Asia and Singapore.

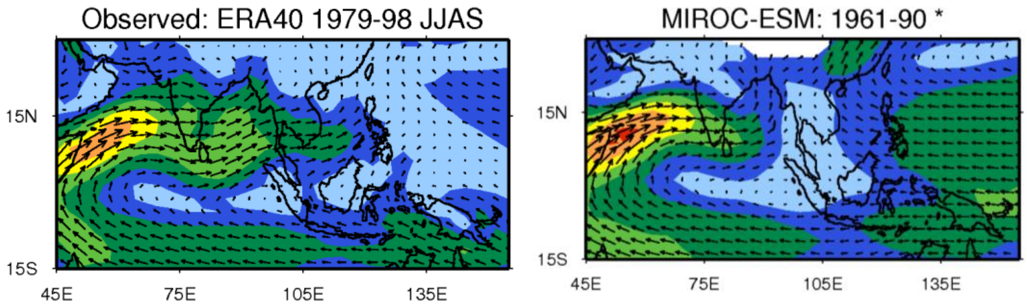

Figure 3.2: South west monsoon circulation in ERA40 as the 'best estimate' of observed climatology and in the MIROC-ESM GCM.

Based on the assessments for each of the climate features listed in Tables 3.1 and 3.2, each model is allocated an overall score based on the number of criteria which are flagged, and the severity of the flags. Models are allocated an overall category as follows:

- Overall 'Implausible': Any one category is scored 'implausible', or four or more 'Significant Biases'.
- Overall 'Significant Biases': Two 'Significant Biases' or three or more 'Biases'/ Significant Biases', where at least one is 'Significant Biases'.
- Overall 'Biases': One 'Significant Biases', or three or more 'Biases',
- Overall 'Satisfactory': Fewer than three 'Biases'.

Additionally to the criteria listed above, in cases where two models from the same modelling centre which have similar formulation are grouped into the same overall category but have different numbers of 'flags', we down-grade the model with the most 'flags'. This applies to the models *bcc-csm1-1*, *IPSL-CM5A-LR and HadGEM2-CC,* which are all overall 'biases' ratings despite having only two 'biases' flags in order to differentiate their lesser performance compared to models *bcc-csm1-1-m, IPSL-CM5A-MR* and *HadGEM2-ES*, respectively.

We compare the performance summary information in Table 3.2 with the projections for future change in mean temperature and rainfall in Figure 3.3 in order to assess which models are to be eliminated, according to the decision-making framework set out in Figure 3.1. Here we explore the areal mean change in both temperature and rainfall as well as exploring changes in the characteristics of the rainfall change (see McSweeney *et al*., 2012) by showing the root-mean-square of changes averaged over the region (an indicator of the overall magnitude of rainfall changes, which does not 'average out' increases and decreases) and the spatial correlation of the changes from each model with the ensemble mean (this gives an indication of whether the spatial pattern of rainfall change is 'typical' of the ensemble thus having higher correlation with the ensemble mean, or is 'atypical', having low or negative correlation). The characteristics of the projections of future climate are determined by comparing projections under RCP8.5 (see Moss *et al*., 2010 for a description of the RCP scenarios) with their corresponding historical simulations. This RCP scenario has the largest greenhouse gas forcing and therefore we can expect the signal-to-noise ratio (the relative size of the long-term trend compared with the natural variability) to be the largest, and providing the clearest information about each model's response. The outcomes of the following observations and decisions are summarised in Table 3.3.

Table 3.1: Summary of evaluation methods employed. Full results of evaluation for each criteria are available in Appendix 3.3.

Evaluation Criteria	Evaluation method
SW monsoon	Comparison by visual inspection of circulation at 850hpa in models with ERA40 (Uppala et al., 2003) for JJAS, looking for evidence that key features are captured. Misplaced flow or mis-directed flow are considered more serious biases than systematic errors in strength of flow.
Summer monsoon variability	Indices representing characteristics of the variability of the summer monsoon published in Sperber et al 2012. The indices include those representing the relationship between ENSO (Nino3.4) and SW Monsoon intensity (indicated by All-India Rainfall, AIR) , characteristics of the East Asian summer Monsoon, and indices describing the magnitude of variability, and characteristics of the life cycle of intraseasonal variability (Boreal Summer Intra-seasonal variability BSISV).
NE monsoon	Comparison by visual inspection of circulation at 850hpa with ERA40 in NDJ, looking for evidence that key features are captured. A key detail of this circulation for Singapore is the North-easterly flow over the South China Sea directing near-surface flow towards the Malaysian peninsula.
ITCZ migration	Comparison by visual inspection of Hovmoller plots of the seasonal migration of the latitude of peak rainfall in GCMs with GPCP2.2 (Adler et al, 2003) observations.
Regional annual cycles of temperature and precipitation	Annual cycles of average rainfall and temperature were calculated as areas averages over sub-regions and compared with multiple observed datasets. Poorest models were identified using the correlation of the seasonal cycles and root-mean-square errors (RMSE).
SST and mean air temperature	Mean and spatial patterns of average air temperature over land were compared with CRU (Mitchell et al., 2005) observations, mean biases in SST compared with HadISST observations (Rayner et al.,2003).
Cold tongue bias	Comparison of cold tongue bias features in models with HadISST (Rayner et al, 2003) by visual inspection and use of a cold-tongue bias index proposed by Hirota and Takayabu (2013).
ENSO	Use of metrics published in Guilyardi et al (2012) and, in greater detail, Bellenger *et al*. (2014). These assess in CMIP5 models a number of metrics of ENSO amplitude (NINO3 SST standard deviation), structure (Nino3 vs Nino4 amplitude), frequency (Root Mean Square Error, RMSE, of Nino3 SSTA spectra) and heating source (Nino4 precipitation standard deviation), and additionally several process-based metrics which reflect the role of the atmosphere response to ENSO, and therefore represent an assessment of whether the models capture the key processes required to represent ENSO realistically.
MJO indices	Use of published process-based indices from Kim et al (2014).

Table 3.2: Summary of model performance (Note that for clarity we only show results for models for which 6hourly fields are available for downscaling). Pink='Implausible', Orange='Significant Biases', yellow= 'Biases' and green ='Satisfactory'. Grey indicates cases where data were not available for analysis.
Overall performance scores in the final column are allocated based on the following criteria: Criteria for overall 'Implausible' (Pink): Any one category is scored 'implausible', or 4 or more categories scored 'Significant biases'. Overall Significant Biases: Two 'Significant Biases' or three 'Biases'/'Significant Biases', of which at least one is 'Significant biases'. Overall Biases (Yellow): One 'Significant Biases', or two or more 'Biases'*. Overall Satisfactory: Fewer than three 'Biases'*.
*** Indicates a model that has been downgraded from overall 'satisfactory' to 'biases' in order to differentiate between the performance of that model and another from the same centre with similar formulation.**

	SW Monsoon	Summer Monsoon variability Indices (Sperber et al)	NE Monsoon	ITCZ Migration	Regional Annual cycles of rainfall	SST and temperature mean bias	Cold tongue bias	ENSO	MJO Indices	Overall
ACCESS1-0										ACCESS1-0
ACCESS1-3										ACCESS1-3
bcc-csm1-1										bcc-csm1-1*
bcc-csm1-1-m										bcc-csm1-1-m
BNU-ESM										BNU-ESM
CanESM2										CanESM2
CCSM4										CCSM4
CMCC-CM										CMCC-CM
CNRM-CM5										CNRM-CM5
CSIRO-Mk3-6-0										CSIRO-Mk3-6-0
EC-EARTH										EC-EARTH
FGOALS-g2										FGOALS-g2
FGOALS-s2										FGOALS-s2
GFDL-CM3										GFDL-CM3
GFDL-ESM2G										GFDL-ESM2G
GFDL-ESM2M										GFDL-ESM2M
HadGEM2-CC										HadGEM2-CC*
HadGEM2-ES										HadGEM2-ES
inmcm4										inmcm4
IPSL-CM5A-LR										IPSL-CM5A-LR*
IPSL-CM5A-MR										IPSL-CM5A-MR
IPSL-CM5B-LR										IPSL-CM5B-LR
MIROC5										MIROC5
MIROC-ESM										MIROC-ESM
MIROC-ESM-CHEM										MIROC-ESM-CHEM
MPI-ESM-LR										MPI-ESM-LR
MPI-ESM-MR										MPI-ESM-MR
MRI-CGCM3										MRI-CGCM3
Nor-ESM1-M										Nor-ESM1-M

Firstly we note that models *MIROC-ESM* and *MIROC-ESM-CHEM* are flagged as 'implausible' due to their particularly unrealistic representation of the SW Monsoon, particularly over south-east Asia, and are therefore eliminated regardless of the characteristics of their projections. We note however, that their projections are not outliers in the mean changes in temperature and rainfall, but they do demonstrate spatial patterns of change that are 'atypical' of the ensemble in most seasons. The *inmcm4* model is also flagged as 'implausible', due to the identification of multiple significant shortcomings. This model stands out as an outlier in its future projections of mean

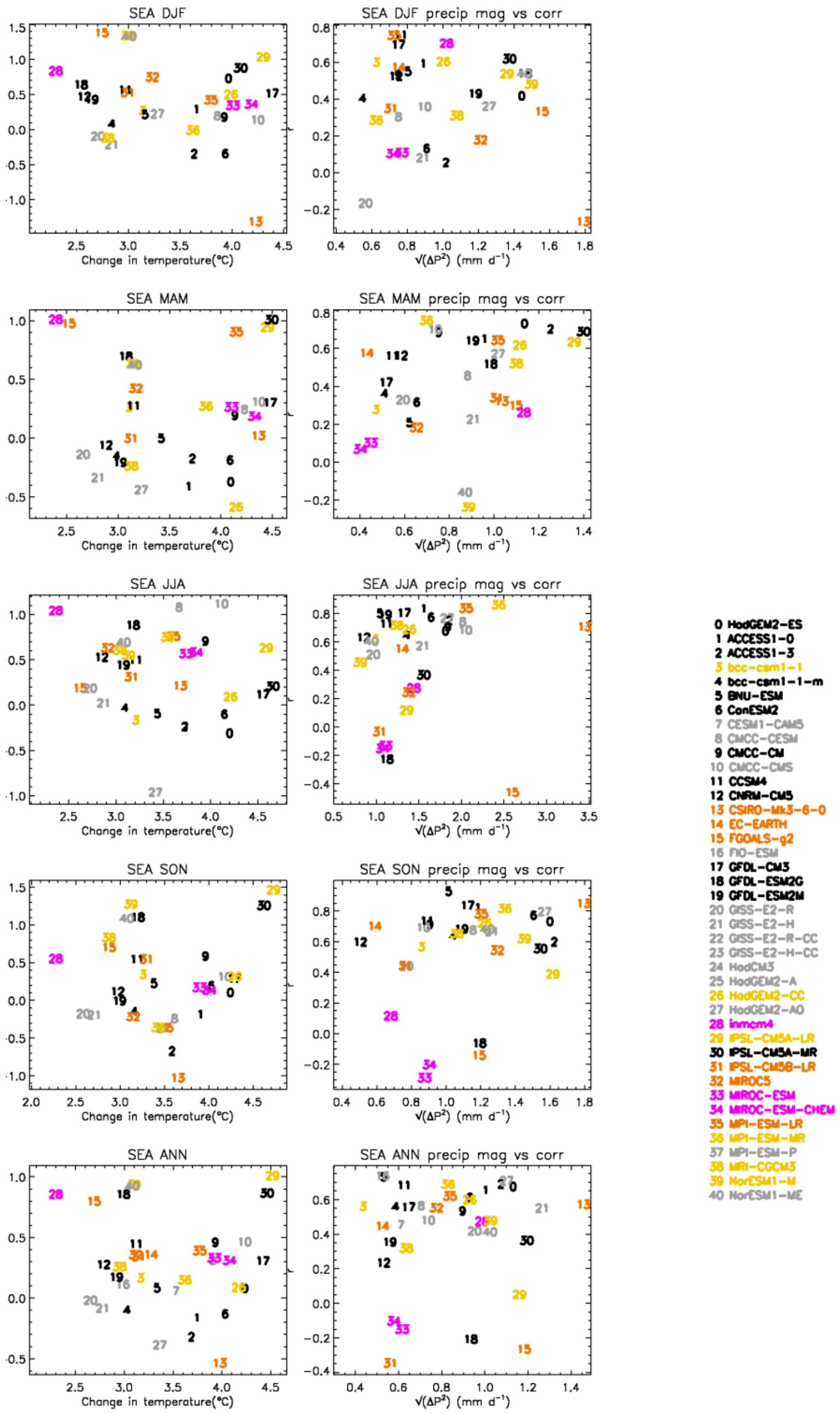

Figure 3.3: Characteristics of the projected changes by 2070-2100 from 1961-90 from each GCM compared with their overall performance rating. Left: Change in mean rainfall vs change in mean temperature over the SEA region. Right: Root-mean-square change in rainfall vs spatial correlation between each model and the ensemble mean. Pink=Models with projections deemed 'implausible', orange: Models rated 'Significant biases' overall; Yellow: Models rated 'Biases' overall; black: models rated 'satisfactory' and grey: models for which 6-hourly fields for downscaling are not available.

climate, as the model with the lowest increase in mean temperature, but also as the model which throughout all seasons occupies the 'cool and wet' region of the scatter plot. Omitting *inmcm4* therefore has significant implications for the range of large-scale-regional climate change characteristics sampled in subsequent regional climate model simulations. MSS climate research experts were consulted regarding the overall assessment of *inmcm4,* and the overall rating above is based on an agreed joint assessment.

Of those models with an overall 'Significant Biases' rating, *FGOALS-g2* and *IPSL-CM5B-LR* projections do not lie outside of the range of change in mean temperature, but display atypical spatial characteristics in rainfall change in some seasons. In the case of *FGOALS-g2*, we have additional concerns about the reliability of the data submitted to the CMIP5 archive[1], and the uncertainty in the data submitted leads us to exclude this model partially for practical reasons. *IPSL-CM5B-LR* is among several models for which the spatial patterns of rainfall changes are relatively 'atypical'. However, by retaining some of the other 'atypical' models with either 'Satisfactory' or 'Biases' ratings preferentially we can retain a good diversity in the spatial patterns of rainfall change amongst remaining models. *CSIRO-mk3-6-0* is a very clear outlier, projecting the driest climate in the region in DJF by some margin, as well as having the largest magnitude of rainfall response (evident in the root-mean-square changes), and an atypical response in the spatial patterns of rainfall change in DJF. The model is therefore retained. Neither MIROC5 nor EC-EARTH lie beyond the range of other models in the ensemble in the characteristics of change addressed.

Table 3.3: Completed decision-making matrix for model selection. Eliminated models are highlighted in shaded cells. * FGOALS-g2 is starred because there remain uncertainties in the dataset submissions which are part of the reason for their elimination.

	Outlier	**Other models predict similar outcomes.**
We are confident that the model is implausible.	MIROC-ESM MIROC ESM-CHEM Inmcm4	
Model demonstrates Biases/Significant biases in performance	CSIRO-Mk3-6-0 NorESM1-M	Bcc-csm1-1 EC-EARTH HadGEM2-CC IPSL-CM5A-LR IPSL-CM5B-LR MPI-ESM-LR FGOALS-g2 * MIROC5 MRI-CGCM3
Model is satisfactory	HadGEM2-ES ACCESS1-0 ACCESS1-3 Bcc-csm1-1-m BNU-ESM CanESM2 CMCC-CM	CCSM4 CNRM-CM5 GFDL-CM3 GFDL-ESM2G GFDL-ESM2M IPSL-CM5A-MR MPI-ESM-MR

[1] FGOALS-s2 historical runs were recently withdrawn completely from the CMIP5 archive due to suspected errors in the mean temperature. FGOALS-g2 data initially submitted were replaced due to inconsistencies and errors in the labelling of ensemble members. While this has been corrected, some further sources of uncertainty remain in the dataset version numbering.

Of the models with a 'Biases' overall rating *IPSL-CM5A-LR* is the model with the largest overall annual temperature and rainfall increase, but not does not lie significantly outside of the range of other models – its projections tend to be similar in characteristics to the high-resolution version of the same model, *IPSL-CM5A-MR* and so the model is eliminated. *NorESM1-M* is retained due to its 'atypical' rainfall pattern projections in order to maintain diversity in this aspect of the remaining models. All other models categorised with 'Biases' are eliminated.

3.4 Sub-selection from remaining models

We have determined which of the CMIP5 models should be eliminated based on their performance and we now select a subset which best spans the range of climate change responses in terms of the mean change in temperature and precipitation, based on the difference between RCP8.5 projections and the corresponding historical runs. From the remaining 16 models, we select a subset of 10 models for downscaling.

In addition to the changes in mean surface variables, a representative sub-set should also span the range of changes in various other aspects if climate change, including changes to key large scale process, such as the monsoon circulations, and changes to characteristics of the modes of variability, such as ENSO. However, there are a number of limitations to this. Firstly, characterising changes in a large scale process such as the monsoon may require us to account for a number of characteristics of changes which might include change in the strength of the flow, the timing of onset, peak and cessation, geographical changes to the location of the flow as well as changes in the precipitation associated with the monsoon. With a limited sub-set size we must restrict the number of 'dimensions' in which we attempt to span to key characteristics of change. Summarising changes in models of variability such as ENSO can again require us to address a number of characteristics of change (i.e. frequency, magnitude, relationship with mean SST changes, tele-connections), but identifying clear signal in a change in variability requires multiple ensemble members (Stevenson, 2012). The varying numbers of ensemble members from different models in the CMIP5 archive make it difficult to characterise the range of changes across the full ensemble. Many studies in the published literature address a subset of the ensemble for which appropriate data are available.

The approach we take is to select the subsets based primarily on capturing the range of mean temperature and rainfall changes throughout the SEA region (defined in figure 3.10 as 93E-127.5E and 11S-25N), based on both areal mean changes and sub-regional patterns. Secondly, we assess how well those subsets span the range of changes in additional key characteristics of the regional climate, where feasible, or where information exists in published literature.

3.4.1 Capturing the range of changes in mean temperature and precipitation

The preliminary subsets are identified by randomly selecting 500 unique sets of 10 models, from the 16 remaining, and then defining an 'optimal sample' which best spans the range of changes. The criteria for the 'optimal' sample is determined using the *fractional range coverage* (FRC), which is simply the proportion of the range of future changes that a subset of models captures compared with the fuller ensemble (in this case, all CMIP5 models which have 6-hourly data available and have not been eliminated on grounds of performance in Section 3.3), where 'changes' are the difference in mean temperature or precipitation in 2070-2100 relative to 1961-90 under RCP8.5.

$$Fraction\ Range\ Coverage\ (FRC) = \frac{Range\ of\ changes\ indicated\ by\ 10\ model\ subset}{Range\ of\ changes\ indicated\ by\ 16\ models}$$

For this calculation, all model data is first interpolated to a common 2.5 x 3.75 degree grid. In order to capture both the range of changes at locations within the region of Southeast Asia, FRC is then calculated for every grid-box, for each 3-month standard season, for changes in both mean surface air temperature and mean precipitation .The FRC values for each candidate subset are normalised by the mean and standard deviation of the FRC values across all 500 samples for each of the two variables. Due to the lesser probability of capturing the range of changes in precipitation compared with temperature noted above, the precipitation values were weighted x2 compared with those for temperature. The 'optimal' sample of 10 models is simply the sample with the largest average Normalised FRC across the two variables and all seasons.

The 'optimal' sample of 10 models is listed in Table 3.4 and the FRC values across the region for each of temperature and precipitation are shown in Figure 3.4. For the identified subset of 10, we achieve an average FRC of 0.9 for surface air temperature and 0.85 for precipitation across all grid-boxes and seasons.

Further, we assess how well the subset captures the average change across the whole SEA region in Figure 3.5. The sample of 10 models fails only to capture the largest rainfall increases in DJF and otherwise provide almost fully representation of the range of area-average changes.

Table 3.4: Model sub-sets recommended based on their ability to capture the range of mean temperature and precipitation change under RCP8.5 for the Southeast Asia region. *GFDL-ESM2G was later dropped due to apparent physical inconsistencies with the RCM.

Selected sub-set of 10 models	
Models to be downscaled at MOHC	**Models to be downscaled at MSS**
CNRM-CM5	ACCESS1-3
CSIRO-Mk3-6-0	Bcc_csm1-1-m
GFDL-CM3	CanESM2
GFDL-ESM2G*	CMCC-CM
HadGEM2-ES	IPSL-CM5A-MR

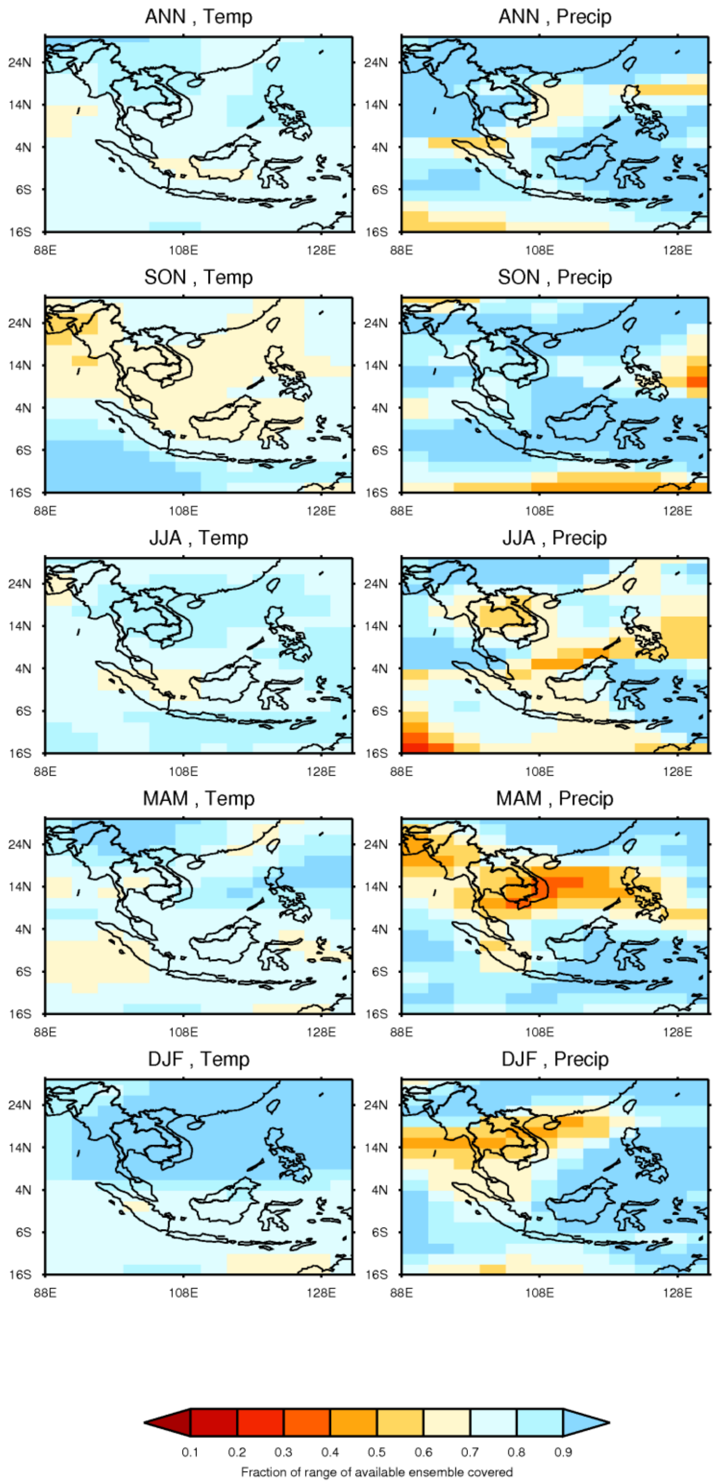

Figure 3.4: Fractional range coverage of subset of 10 models listed in Table 3.4 for both mean temperature and mean precipitation.

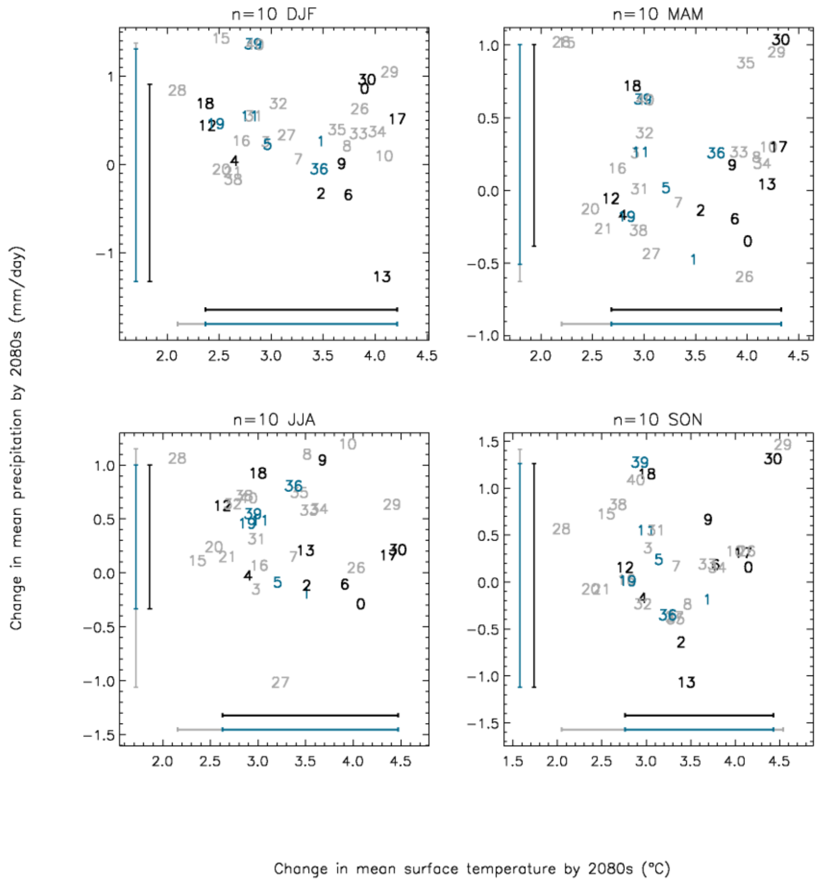

Figure 3.5: Summary of the seasonally-averaged area-average change in mean temperature and precipitation in CMIP5 projections under RCP8.5. Standard 3-month seasons are used in order to give equal weighting to each month of the year. Models in grey are those which either do not have 6-hourly variables available for downscaling, or have been eliminated based on performance. Models in black are those included in the 10 model subset, models in blue were not eliminated but not selected. Bars show the range spanned by the full ensemble (grey), models which are not eliminated but not selected (blue) and those which have been selected (black). For list of model names see Figure 3.3.

3.4.2 Capturing other characteristics of change in the projections sub-set.

3.4.2.1 Changes in monsoon circulation and intensity

Previous studies based on CMIP3 have indicated a weakening of the SW monsoon circulation but an increase of mean monsoon rainfall, which could be attributed to the projected intensification of the heat low over northwest India, the trough of low pressure over the Indo-Gangetic plains, and the land–ocean pressure gradient during the establishment phase of the monsoon (Kripalani et al., 2007). Jiang and Tiang (2013) have assessed changes in the East Asian winter (NE) monsoon, and found that as a whole the monsoon circulation changes little, but that a geographical response can be identified, with regions north of 25N experiences a weakening, and south of 25N strengthening.

We summarise the monsoon response by taking a simple difference between the average 850hpa wind speed and mean precipitation for the regions 5-30N,60-120E (SW monsoon) and 0-25N, 95-125E (NE Monsoon) (Figure 3.6). In the cases of both the SW and NE Monsoons, models span both increases and decreases in flow, but models consistently indicate increases in mean rainfall in the regions. In the case of the NE Monsoon, we see a stronger relationship between the two variables, and a tendency for more models to indicate a strengthening of the 850hpa flow. The subset captures the range of responses in the monsoon well, as defined here, including models demonstrating both increases and decreases in flow in both monsoon systems.

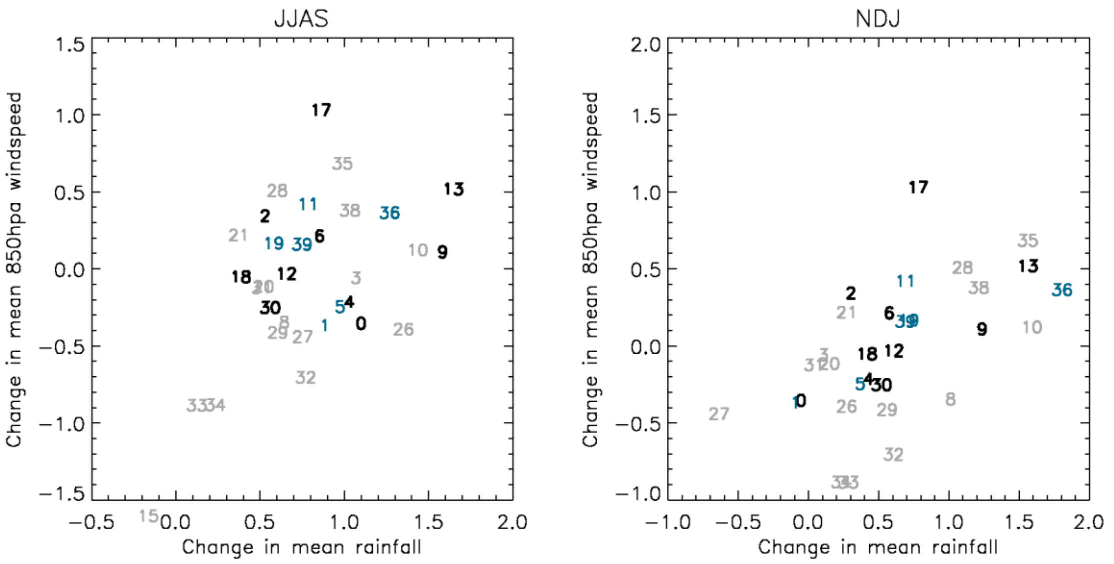

Figure 3.6: Area-average changes in 850hpa flow and mean precipitation, as spanned by subset of 10. For list of model names see Figure 3.3. Selected models in black, non-selected models in Blue, and eliminated or unavailable models in grey.

3.4.2.2 Changes in ITCZ position and intensity

Figure 3.7 summarises changes in the intensity of ITCZ rainfall and shifts in the latitude of the peak rainfall intensity by the end of the century under RCP8.5 in CMIP5 models. Projected changes during the months October-February are mainly characterised by changes to the intensity, with little change in the ITCZ position. During these months, the projected changes in intensity span both increases and decreases, and this range is reflected in the 10-member subset; for example, models CanESM2 and HadGEM2-ES demonstrate decreased rainfall intensity in these months, while IPSL-CM5A-MR demonstrates increases.

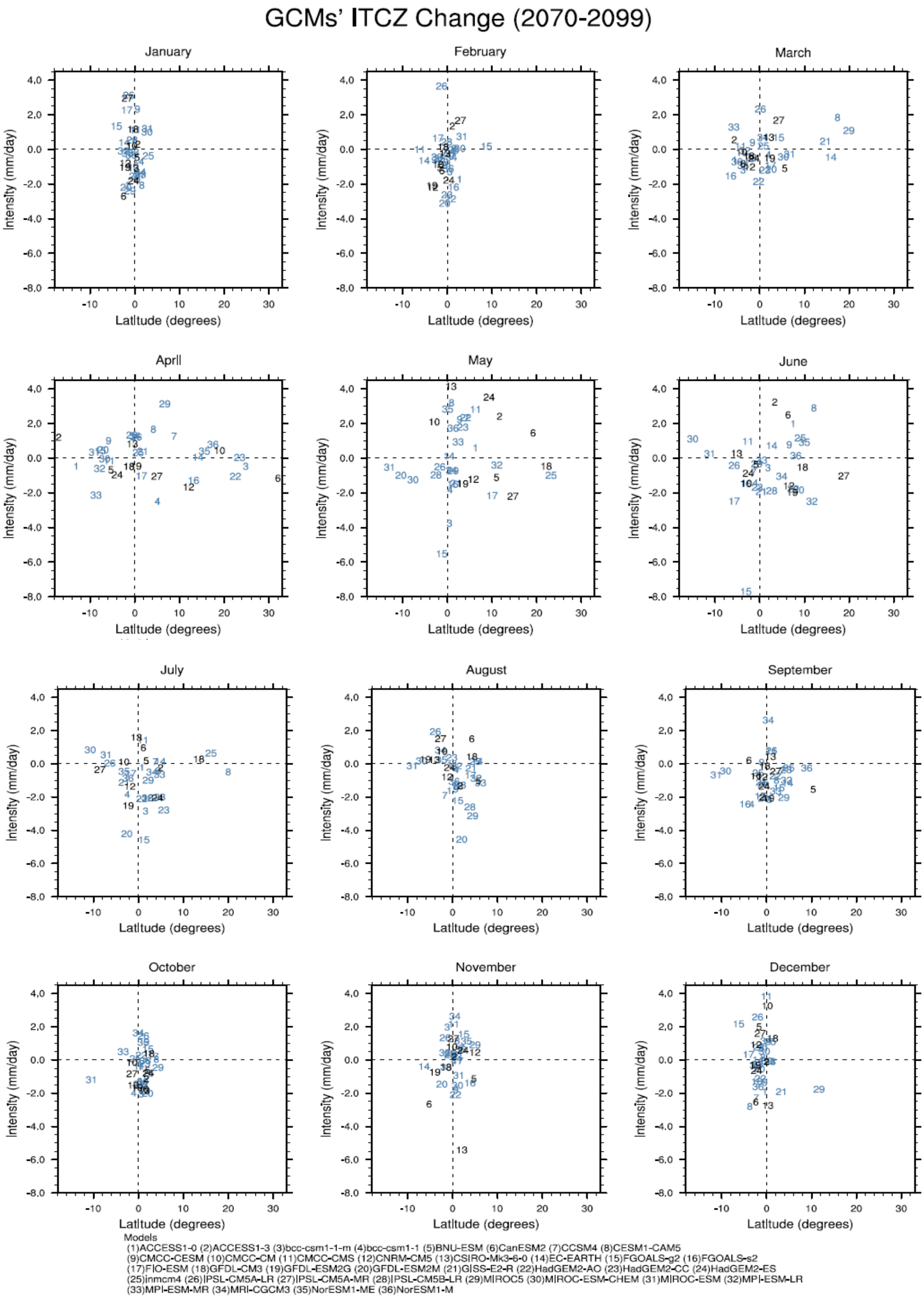

Figure 3.7: Change in mean intensity and latitude of zonal maximum ITCZ rainfall, where rainfall are averaged between 90-120E. Models selected for inclusion in the 10-model subset are highlighted in blue and the remaining models in the ensemble are in grey.

The projected changes in ITCZ intensity during the months March-August similarly span both increases and decreases, but shifts in latitude are also indicated. While the sub-set of 10 does not include the models projecting the largest decreases in July and August (e.g. *FGOALS-g2* (15 in Figure 3.7) and *GFDL-ESM2M* (20 in Figure 3.7)), they otherwise span a range a reasonable range of these increases and decreases in all months. Models ACCESS1-3 and CanESM2 indicate increases in intensity in May and June (2 and 6 in Figure 3.7) while GFDL-ESM2G (19 in Figure 3.7) characterises reduced intensity during these months. Model GFDL-CM3 (18 in Figure 3.7) characterises a shift towards the north of the peak rainfall intensity during May-July. Some models also indicate shifts to the south (e.g. IPSL-CM5A-MR in July (27 in Figure 3.7) but these are smaller changes and occur in fewer models than the northern shifts. The subset tends not to capture the models with the largest southern shifts in the ITCZ rainfall, but we note that some of the models with the largest shifts to the south are *MIROC-ESM* (30 in Figure 3.7) and *MIROC-ESM-CHEM* (31 in Fig 3.7) are those which were judged to be too unrealistic in their simulation of baseline climate to generate plausible projections in Section 3.3.

3.4.2.3 Changes in ENSO

Guilyardi et al. (2012) note in their assessment of ENSO in CMIP5 models that the models do not paint a consistent picture of the changed character of ENSO under warming scenarios. The study notes that models in the ensemble span both increases and decreases in ENSO variability, with results summarised in Table 3.5.

Although we do not have this information about the changes in ENSO for all selected models, we can assess whether or not the subsets include a representative range of outcomes. The 10-member subset spans the range of responses well, including models with significant decreases (CanESM2), increases (CNRM-CM5 and CSIRO-mk3-6-0) and no significant change (HadGEM2-ES) in ENSO variability.

Table 3.5: ENSO variability responses of CMIP5 models under warming scenarios, Guilyardi et al, 2012. NorESM1 indicates a decrease of the largest magnitude, but this is only just inside natural variability). Models in bold face are included in our 10-member subset.

Significant decreases	No Significant change	Significant Increases
IPSL-CM5A-LR **CanESM2** (Nor-ESM1-M)	Bcc-cms1-1 GFDL-ESM2M **HadGEM2-ES** Inmcm4 MIROC-ESM	MIROC5 MPI-ESM-LR MRI-CGCM3 **CNRM-CM5** **CSIRO-mk3-6-0**

3.4.2.4 Changes in Tropical Cyclones

The work in Chapters 8 and 9 of the report will address the vulnerability of Singapore to potential changes in mean and extreme sea level. The downscaled sub-set of projections will be used to drive surge and wave models for the region in order investigate both contemporary and future drivers of extreme sea levels.

While Tropical cyclones (TCs) currently only rarely affect Singapore directly due to its proximity to the equator, they are of interest in the context of sea-level variability in the wider region due to the associated swell that accompanies deep low pressure systems in the atmosphere. By addressing the potential changes in TC characteristics in the north-west pacific, we will be better able to explore the implications of those changes for the wave climate in the South China Sea and approaches to the Singapore Strait. It is therefore helpful to understand how well the selected subset spans any changes in TCs indicated by the full CMIP5 ensemble.

Global model resolution is generally too coarse to reproduce TCs and their structure realistically, but some models are capable of producing similar, high vorticity tropical depressions which can be tracked and studied as an indicator or TC behaviour. Generally, as model resolution increases, so does the number of TCs generated by global models (e.g. see Camargo, 2013; Bengtsson et al. 2007), although the choice of convective parameterisation can also have a significant impact (Hill and Lackmann 2009; Murakami et al. 2012). For many lower resolution models, however, the number of cyclone-like systems generated is too low to relate to observations, or calculate meaningful relative changes. For this reason, we can only assess the changes in TCs in a limited subset of the CMIP5 ensemble.

The method used for tracking TCs is described in Appendix 3.4. After eliminating those models which were unable to realistically reproduce key features of present day TCs in the North-West Pacific, 7 CMIP5 ensemble members remain; bcc-csm1-1-m, CanESM2, CMCC-CM, CNRM-CM5, HadGEM2-CC, HadGEM2-ES and MRI-CGCM3. All of these models (with the exception of HadGEM2-ES and HadGEM2-CC) produce too many TCs outside of the observed TC season, but overall their annual cycle is reasonable.

The 7 CMIP5 models in which TC changes are assessed project a wide range of possible changes for the North-West pacific in both TC number and TC intensity (Figure 3.8). Of these 7 models, 5 are included in the selected 10-member sub-set; of these we capture examples of models which indicate increases in intensity and increases in frequency (*bcc-csm1-1-m*), increases in intensity and decreases in frequency (*CanESM2*, *CNRM-CM5* and *CMCC-CM*) and , and decreases in both intensity and frequency (*HadGEM2-ES*).

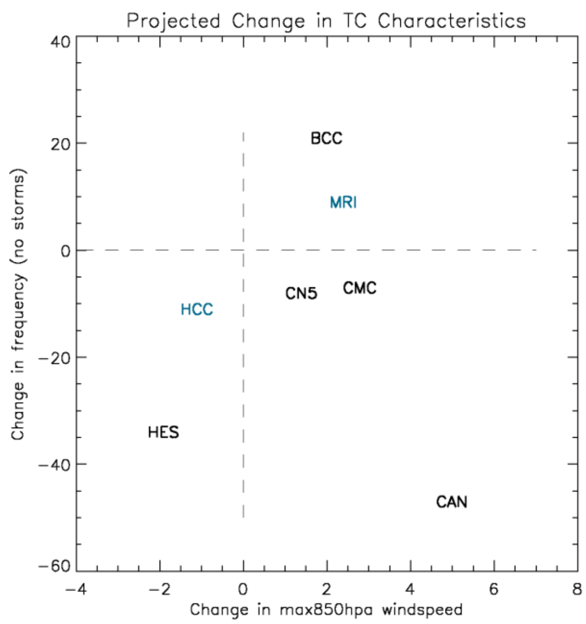

Figure 3.8: Percentage change in a) number of TCs and b) maximum 850hPa wind speed m/s 2064-2095 minus 1971-2002 from CMIP5 models which produce enough TCs to allow assessment of the change in behaviour (See Appendix 3.4 for method). 'BCC'=bcc-csm1-1-m, 'CAN'=CanESM2,

'CMC'=CMCC-CM, 'CN5'=CNRM-CM5, 'HCC'='HadGEM2-CC, 'HES'=HadGEM2-ES, 'MRI'=MRI-CGCM3. Selected models are in black, Non-selected models in Blue.

3.5 Summary and notes for use of the CMIP5 sub-set

3.5.1 Summary and remaining issues

The CMIP5 multi-model ensemble provides simulations from the latest generation of state-of-the-art global climate models. While the ensemble cannot be considered to span all plausible future climate outcomes, the range projections across the ensemble provides an indication of the magnitudes of uncertainties that relate to the formulation of GCMs. By sub-selecting from the CMIP5 models, we can capitalise on this information about projection uncertainty while moderating the resource requirements of downscaling projections to high resolution.

A subset of 10 models has been identified for use in downscaling experiments for the generation of high-resolution climate projections for Singapore. The subset of 10 models demonstrates a representative coverage of the range of changes, in mean temperature and precipitation across and within the Southeast Asia region and also specifically when we address changes in a broader range of climate characteristics, such as changes in monsoon circulation, tropical cyclone characteristics, ENSO amplitude and the position and latitude of the ITCZ.

The model sub-set was selected with a primary concern for maximising the range of plausible climate futures for the region. Three models (*MIROC-ESM* and *MIROC-ESM-CHEM* and *inmcm4*) were eliminated from our analysis, in the first two cases because their representation of the south-west Monsoon was insufficiently realistic to provide plausible projections of climate change in the region, and due to widespread occurrences of significant biases in the case of *inmcm4*. Results of evaluating the realism of models in a number of characteristics of their climatology and variability provided the basis for sub-selecting from the remaining models based on maximising the range of projections spanned by the sub-set whilst avoiding the models in which we have least confidence.

A further benefit of conducting a sub-selection exercise is that it provides some initial understanding of the strengths and weaknesses of the GCMs from the outset. This information provides very important context for:

1. understanding sources of errors when evaluating the downscaled projections
2. understanding the limitations of the downscaled projections

The assessment and sub-selection of the GCMs for use in this project has been undertaken to make best use of available resource by prioritising the assessment of aspects of model behaviour which are key to the generation of plausible projections. However, as with any study that uses model simulations, a number of limitations to this activity should be recognised.

There remain significant aspects of the climate, particularly relating to modes of variability (ENSO and MJO) as well as Tropical Cyclones and cold surges, which are not fully understood, and/or not well represented in coupled models. Studying the diagnosed changes in these features of the climate system provide one source of information about how they might change in the future. However, the shortcomings in their representation mean that it is difficult to attach confidence to those projections. The demonstrated shortcomings in representation of key driving processes that underpin these models, mean that these projections should be interpreted as highly uncertain. Projected changes in the characteristics of ENSO in CMIP5 models reflect this lack of confidence in the fidelity of the key processes, and the subsets that we have proposed broadly represent this uncertainty. We have not been able to assess changes in MJO similarly, and therefore note that the interpretation of any change in the MJO diagnosed from the subset, either before or after downscaling, should be interpreted with due consideration for these uncertainties.

Much of our analysis is dependent on assessment by visual inspection of plotted fields of variables. This assessment process can therefore be considered as essentially subjective. However, this allows us to look for a wide range of possible error characteristics, and judge their importance or magnitude on a case-by-case basis. While indices and metrics might provide a more objective measure of performance for very specific aspects of model behaviour, we are interested in too wide a range of behaviours and characteristics of the model behaviour to warrant the use of indices in many cases.
Where we have drawn on existing literature to inform our sub-selection, the studies do not always include all of the models in the ensemble that we have considered. Where we have undertaken our own assessments, we have been unable to assess some aspects of some models due to missing parts of datasets, or uncertainty in the accuracy of data cataloguing. This is a result of the evolving nature of the CMIP5 archive, and a necessary implication of using projections from state-of-the-art models, of which analyses in the literature are only just emerging.

3.5.2 Guidance on the interpretation of multi-model projections

The sub-selection methodology provided in this report provides an efficient way of sampling the CMIP5 range of outcomes for use in dynamical or statistical downscaling. Given the widespread use of CMIP5-based downscaling expected via the international CORDEX initiative (http://wcrp-cordex.ipsl.jussieu.fr/), this work will place users in Singapore in a strong position to compare their diagnosed climate change impacts consistently with future studies carried out in many other regions. However, it is also important to understand the nature and limitations of the information provided.

Firstly, the CMIP5 ensemble, despite its status as the latest available modelling technology, cannot be assumed to sample the full range of plausible outcomes consistent with current knowledge. This reflects its construction as an ensemble of opportunity, which was not explicitly designed to span some notional “space of possible climate models” (Knutti et al., 2010). Also, the range of outcomes is conditional upon the set of earth system feedbacks considered in the ensemble. For example, all the CMIP5 models considered in this report were driven by prescribed concentrations of CO_2. This is because regional climate model driving data were not saved from emissions-driven CMIP5 simulations made with earth system model configurations. This means that uncertainties in converting carbon emissions to atmospheric CO_2 concentrations are not accounted for in our subset of recommended models. Thirdly, whilst we make extensive use of evaluation against observations to inform our recommendations, we do not attempt to estimate the relative credibility of different CMIP5 models by formally applying observational constraints based on historical performance.

The sub-selection work is designed to capture the range of future projections for the region, but is not designed to support the estimation of relative likelihoods for any of the different future outcomes within or outside the range of projections defined by our subset of selected models. The probabilistic projection work in Chapter 7 will, however, provide some relevant context to this effect, by supplying probabilistic estimates of future change based on a more comprehensive strategy for sampling modelling uncertainties, making use of additional information about uncertainty from larger ensembles of global climate models.

The significance of systematic errors common to the set of available models should also be recognised. The presence of common model errors (e.g. Pennell and Reichler, 2011) may indicate that important processes are missing from the models, or that they share structural deficiencies in how some processes are represented. Such errors imply an unquantifiable element of uncertainty in how any set of model projections might relate to future climate change in the real world (e.g. Smith, 2002). In practice, common systematic biases may turn out to be negligible for some basic variables of interest (say multiyear spatial averages of surface temperature), but much more significant for others. This might apply, for example, to aspects of intraseasonal variability (the MJO results in Appendix 3.3 of this report are a good example) or extreme events difficult to capture at the typical spatial resolution of CMIP5 models. However, the process of thoroughly evaluating the driving GCMs before downscaling provides important understanding of weaknesses (as well as the strengths) of each model). This will provide useful contextual information for the evaluation of the downscaled projections and interpretation of the projections, thus forming the basis for robust guidance on their appropriate use. Further evaluation of the models after downscaling is detailed in Chapter 4 and Supplementary Information Report 1, and includes analysis of climate variability and extremes with direct relevance for impacts assessment.

Acknowledgements

We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We are also grateful to Jamie Kettleborough, Ian Edmond and Emma Hibling at the Met Office Hadley Centre for developing tools which have allowed us to download and access this data easily. Thanks to Daeyun Kim and Prince Xavier for sharing codes for the calculation of MJO indices.

References

Adler, R. F., and Coauthors, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). Journal of Hydrometeorology, 4, 1147-1167.

BENGTSSON, L., HODGES, K. I., ESCH, M., KEENLYSIDE, N., KORNBLUEH, L., LUO, J. J. & YAMAGATA, T. 2007. How may tropical cyclones change in a warmer climate? *Tellus Series a-Dynamic Meteorology and Oceanography,* 59, 539-561.

Bellenger. H., Guiliyardi, E, Leloup, J. Lengaigne, M and Vialard, J. (2014) ENSO representation in climate models: from CMIP3 to CMIP5. *Climate Dynamics 42(7-8), 1999-2018.*

Camargo, S.J (2013). Global and regional aspects of tropical cyclone activity in the CMIP5 models. Submitted to *Journal of Climate,* July (2012), revised (March 2013).

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *Journal of Geophysical Research-Atmospheres,* 113, 20.

Guilyardi, E., Bellenger, H., Collins, M., Ferrett, S. Cai, W and Wittenberg, A (2012) A first look at ENSO in CMIP5. *CLIVAR Exchanges* No. 58, Vol. 17, No.1.

Harris, G.R., D.M.H. Sexton, B.B.B. Booth, M. Collins and J.M. Murphy, 2013: Probabilistic projections of transient climate change. *Climate Dynamics* 40: 2937-2972.

HILL, K. A. & LACKMANN, G. M. 2009. Analysis of Idealized Tropical Cyclone Simulations Using the Weather Research and Forecasting Model: Sensitivity to Turbulence Parameterization and Grid Spacing. *Monthly Weather Review,* 137, 745-765.

Hirota N. And Takayabu Y.N (2013) Reproducibility of precipitation distribution over the tropical oceans in CMIP5 multi-climate models compared to CMIP3. Climate Dynamics 41(11-12), pp2909-2920.

Jiang, D. B., and Z. P. Tian, 2013: East Asian monsoon change for the 21st century: Results of CMIP3 and CMIP5 models. *Chinese Science Bulletin,* 58, 1427-1435.

Kim, Daehyun, P.Xavier, E.Maloney, M. Wheeler, D. Waliser, K.Sperber, H. Hendon, C. Zhang, R.Neale, D. Raymond and H. Liu (2014). Process-Oriented MJO Simulation Diagnostic: Moisture Sensitivity of Simulated Convection. J. Climate 27(14), 5379-5395.

Kripalani, R. H., J. H. Oh, A. Kulkarni, S. S. Sabade, and H. S. Chaudhari, 2007: South Asian summer monsoon precipitation variability: Coupled climate model simulations and projections under IPCC AR4. *Theoretical and Applied Climatology,* 90, 133-159.

Karmacharya, J., Jones, R and New, M (In Review) Realistic simulation of the South Asian monsoon in a regional climate model: the role of domain size and implications for the parent global model. *Climate Dynamics*.

Knutti, R., 2010: The end of model democracy? *Climatic Change*, 102, 395-404.

Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in Combining Projections from Multiple Climate Models. *Journal of Climate,* 23, 2739-2758.

Knutti, R. and Sedlacek, J. (2012) Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change* 1716.

McSweeney, C. F., R. G. Jones, and B. B. B. Booth, 2012: Selecting Ensemble Members to Provide Regional Climate Change Information. *Journal of Climate,* 25, 7100-7121.

Mitchell, T.D. and Jones, P.D., 2005: An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology

Moss, R. H., and Coauthors, 2010: The next generation of scenarios for climate change research and assessment. *Nature,* 463, 747-756.

Moufouma-Okia, W and Jones R (2015) Resolution dependence in simulating the hydro-climate of Africa using the HadGEM3 regional climate model. *Climate dynamics*.

MURAKAMI, H., MIZUTA, R. & SHINDO, E. 2012. Future changes in tropical cyclone activity projected by multi-physics and multi-SST ensemble experiments using the 60-km-mesh MRI-AGCM. Climate Dynamics, 39, 2569-2584.

Pennell C., and T. Reichler, 2011: On the effective number of climate models. *Journal of Climate, 24*: 2358-2367.

Rayner, N. A.; Parker, D. E.; Horton, E. B.; Folland, C. K.; Alexander, L. V.; Rowell, D. P.; Kent, E. C.; Kaplan, A. (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century *J. Geophys. Res.Vol.* 108, No. D14, 4407 10.1029/2002JD002670

Reichler, T., and J. Kim (2008): How Well do Coupled Models Simulate Today's Climate? *Bull. Amer. Meteor. Soc*, 89, 303-311.

Smith, L.A., 2002: What might we learn from climate forecasts? *Proc. Natl. Acad. Sci. U.S.A.* 99: 2487-2492.

Sperber KR, H Annamalai, I-S Kang , A Kitoh, A Moise, A Turner B Wang T Zhou (2012) The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late century. *Climate Dynamics.* Doi: *10.1007/s00382-012-1607-6*

Stevenson, SL (2012) Significant changes to ENSO strength and impacts in the twenty-first century: results from CMIP5. *Geophysical Research Letters* 39, L17703

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: AN OVERVIEW OF CMIP5 AND THE EXPERIMENT DESIGN. *Bulletin of the American Meteorological Society*, 93, 485-498.

Tebaldi, C., R.W. Smith, D. Nychka, and L.O. Mearns, 2005: Quantifying uncertainty in projections of regional climate change: A Bayesian approach to the analysis of multi-model ensembles. *Journal of Climate* 18: 1524-1540.

Tebaldi, C., J. M. Arblaster, and R. Knutti, 2011: Mapping model agreement on future climate projections. *Geophysical Research Letters,* 38.

Tkalich, P; Vethamony, P.; Babu, M. T.; Malanotte-Rizzoli, Paola. *(2013)* Storm surges in the Singapore Strait due to winds in the South China Sea. *Natural Hazards* vol. 66(3), p1345 – 1362

Watterson, I.G., 2008: Calculation of probability density functions for temperature and precipitation change under global warming. *J. Geophys. Res.* 113:D12106.