Nuku'alofa, Tonga

Chapter 5
Climate Model Reliability

Summary

- Global climate models can simulate many aspects of climate, and generally give a reasonable representation of seasonal and interannual climate in the PCCSP region. Most models, however, show biases, such as a tendency for sea-surface temperatures along the equator to be too cold. Such biases must be taken into consideration when using the models for climate projections.
- The representation of the El Niño-Southern Oscillation (ENSO) in climate models has improved over the years but remains a challenge at the regional scale. For example, sea-surface temperature variability associated with ENSO tends to be too narrowly focused on the equator, too far to the west and variability away from the equator in the sub-tropical Pacific tends to be too small.
- Global climate models reproduce the observed pattern of the regional distribution of sea level reasonably well. Global climate models that include all factors that influence the climate generate similar temporal variability in global-averaged ocean thermal expansion to the observations, but a slightly smaller rate of rise over recent decades. Observations indicate sea level is currently rising at near the upper end of the projected range.
- The CSIRO Direct Detection, Genesis Potential Index and Curvature Vorticity Parameter projection methods generally performed well in reproducing the observed tropical cyclone climatology when applied to global climate models.
- An evaluation of the performance of 24 global climate models has identified a set of 18 models which provide a reasonable representation of climate over the PCCSP region. These models are used for constructing projections of future climate for the PCCSP region and for individual Partner Countries in Chapters 6 and 7.
- Dynamical downscaling models have finer resolution and can be used to obtain information at regional scales, including over topography or where coastal effects are important. Since this is very computer-intensive, the output from only six global models has been downscaled and used to provide additional projections information in Chapter 7.
- Bias-adjustment of the sea-surface temperatures and increased atmospheric model resolution improves the representation of some aspects of the current climate in the downscaled simulations. Moreover, downscaled projections can provide useful complementary information to projections from coarser global climate models. However, there is no guarantee that the downscaled projections are actually more reliable than the projections from the coarser global models.

5.1 Introduction

To make scientifically robust and confident projections of the future climate, it first has to be demonstrated that global or downscaled climate models are sufficiently realistic in simulating the present climate. Whether a model is skilful in simulating the present climate will depend on its ability to represent the long-term average and seasonal cycle of various atmosphere and ocean fields (e.g. temperature, rainfall, wind, salinity and sea level); important regional large-scale climate features (e.g. Intertropical Convergence Zone (ITCZ), South Pacific Convergence Zone (SPCZ) and West Pacific Monsoon); the major components of climate variability on various timescales (e.g. the El Niño-Southern Oscillation (ENSO), Pacific decadal variability); extreme weather events (e.g. heat waves, heavy rain, tropical cyclones); and long-term trends. In addition, climate models should be stable and not exhibit substantial model drift (model drift refers to spurious trends in climate simulations that occur in the absence of any change in factors that might be expected to induce change).

The level of agreement between model simulations and observations of the present climate is used as a method of assessing model reliability. It is assumed that a model which adequately simulates the present climate will provide more reliable projections of the future. As shown in this chapter, no one model is the best in representing all aspects of climate, and a range of models should be considered when making projections of future climate.

Because of the complexity of climate systems, and the wide range of climate change factors of interest to stakeholders, it is important that model evaluation considers a broad range of features, fields and processes – with a corresponding requirement for high quality observations. For some features (e.g. global gridded rainfall) more than one dataset is available, which can provide a useful measure of observational uncertainty. The various observational datasets used in this chapter (and publication) are described in Chapter 2 and listed (along with their abbreviations) in Table 2.2 for the atmosphere variables and Table 2.3 for the ocean variables.

The global climate models (CMIP3 models) assessed in this chapter are listed in Table 4.1, along with model numbers and names used throughout this publication. Features of global climate models, such as resolution and representation of physical processes, are discussed in general in Section 4.3, along with details of the CMIP3 suite of experiments. Most analysis in this chapter is carried out with respect to the climate of the 20th century experiments described in Section 4.3.1. Also described in Chapter 4 are the techniques used for dynamical and statistical downscaling and for the calculation of tropical cyclones and sea-level changes based on climate model output.

The performance of global climate models is evaluated for atmosphere variables, ocean variables, major climate features and patterns of variability and extremes in Section 5.2. The performance of dynamical and statistical downscaling is evaluated in Section 5.3. The simulation of tropical cyclones in climate models is evaluated in Section 5.4. A summary of the suitability of models for use in climate projections in the PCCSP region is provided in Section 5.5.

Box 5.1: Summary of Intergovernmental Panel on Climate Change Comments on Model Strengths and Weaknesses

Climate models are based on the laws of physics and can reproduce many observed features of the current climate and past climate changes. Confidence in model simulations is higher for some climate variables (e.g. temperature) than for others (e.g. rainfall). Confidence is also generally higher for changes on continental or larger scales than for sub-continental and island scales.

Confidence in the reliability of global climate models for climate projections has improved, based on tests of their ability to represent:

- Present day average climate and year-to-year variability.
- Observed climate trends in the recent past.
- Extreme events, such as storms and heatwaves.
- Climates from thousands of years ago.

Models show significant and increasing skill in representing many important mean climate features, such as the large-scale distributions of atmospheric temperature, rainfall, radiation and wind, and of oceanic temperatures, currents and sea-ice cover. Patterns of climate variability that are generally well-simulated include the advance and retreat of the major monsoon systems, the seasonal shifts of temperatures, storm tracks and rain belts. Simulations that include estimates of natural and human influences can reproduce the observed large-scale changes in surface temperature over the 20th century, including the global warming that has occurred during the past 50 years.

However, there remain significant deficiencies in models, stemming largely from the need to represent in approximate form (parameterise) some small-scale but key physical processes, such as cloud physics, radiation and rainfall processes. Importantly, errors or biases remain in a number of aspects of the simulation of tropical rainfall, ENSO and the Madden-Julian Oscillation, and tropical ocean temperatures. Relatively small-scale events such as tropical cyclones and thunderstorms are less skilfully reproduced. Finer resolution models tend to produce more realistic simulations, although increased resolution alone does not reduce some important biases.

In summary, climate models provide credible quantitative estimates of future climate change, particularly at larger scales. Some deficiencies remain at smaller scales. There will always be a range of uncertainty in climate projections. People doing impact assessments based on climate model projections need to understand and incorporate this uncertainty.

(Summary based on information in Chapter 8 of the Working Group 1 contribution to the IPCC Fourth Assessment Report (2007)).

5.2 Evaluation of CMIP3 Climate Models

In this section, the seasonal and annual mean climatology of the 24 CMIP3 models is compared with observations. First, key atmospheric variables are examined, then model ocean climatology is compared with observed ocean variables. The key features of mean climate and climate variability in the models are compared with observations, and finally the simulation of climate extremes is compared with the observed distribution of extremes.

5.2.1 Atmosphere Variables

In simulating the current climate, the CMIP3 models should ideally capture the observed mean state of the climate, the timing and strength of the seasonal cycle, and the location and magnitude of any spatial features. A number of statistics relevant to these aspects of a model simulation have been calculated (Table 5.1). Unless otherwise stated, the results presented for these statistics refer to the multi-model mean (plus/minus the inter-model standard deviation) of the all-observation test scores (i.e. the value obtained by averaging across the scores from all relevant observational datasets). These observational datasets include CMAP and GPCP for the rainfall analysis, and the ERA-40, NCEP/DOE R-2 and JRA25 datasets for surface temperature and wind (Table 2.2). A complete listing of the CMIP3 models that were assessed for each variable (i.e. the models that have monthly time scale data available for the climate of the 20th century simulation) is given in Appendix 1.

5.2.1.1 Surface Air Temperature

The vast majority of the CMIP3 models show a cold bias throughout much the PCCSP region (Figure 5.1; E_{bias}= -0.8 ± 0.8°C), which is largely responsible for the multi-model mean grid-point error magnitude of E_{abs} = 1.0 ± 0.5°C. The models are able to capture the general surface air temperature spatial pattern, however they tend to exaggerate the intensity and westward extent of the cool equatorial surface air associated with the oceanic cold tongue (r_p = 0.9 ± 0.1; $\sigma_{ratio,x}$ = 1.1 ± 0.1). The models also tend to slightly overestimate the amplitude of the seasonal cycle ($\sigma_{ratio,t}$ = 1.2 ± 0.3), however the phase of this cycle is relatively well represented (r_t = 0.8 ± 0.1).

5.2.1.2 Rainfall

The main large-scale features of the climatological rainfall in the PCCSP region, including the high rainfall areas associated with the tropical convergence zones and West Pacific Monsoon, are clearly present in most CMIP3 model simulations (Figure 5.2 and Section 5.2.3). However, discrepancies in the precise location of these simulated features generally lead to relatively large grid-point errors in the 1980–1999 annual mean rainfall field (E_{abs} = 1.8 ± 0.5 mm per day) and are reflected in the spatial correlation values obtained for each month (r_p = 0.7 ± 0.2). Given the pronounced differences between the CMAP and GPCP datasets with respect to the intensity of tropical rainfall (Yin et al., 2004; also see Figure 5.2 and discussion in Section 2.2.1), it is interesting to note that the amplitude of both the modelled seasonal cycle and spatial distribution of rainfall in the PCCSP region tends to compare more favourably with CMAP ($\sigma_{ratio,t}$ = 1.1 ± 0.3 and $\sigma_{ratio,x}$ = 1.1 ± 0.2) than GPCP ($\sigma_{ratio,t}$ = 1.5 ± 0.4 and $\sigma_{ratio,x}$ = 1.5 ± 0.3). The models compare similarly to CMAP and GPCP with respect to their ability to capture the phase of the seasonal cycle (r_t = 0.6 ± 0.1).

Table 5.1: Definition of the statistics used to assess the ability of the CMIP3 models to simulate surface air temperature, rainfall and the surface wind. Each statistic was calculated over the PCCSP region (as defined in Section 1.3) after interpolation to a common 2.5° latitude/longitude grid.

Aspect assessed	Statistical test	Data required
Mean state (bias)	Bias (or difference - model minus observed) in the spatial average value (E_{bias})	1980–1999 annual mean field
Mean state (error)	Grid point average of the magnitude of the difference between the model and observed field (E_{abs})	
Seasonal cycle (phase)	Grid point average temporal correlation (r_t)	Spatial field containing 12-step time series of the1980–1999 mean value for each month
Seasonal cycle (amplitude)	Grid point average temporal standard deviation ratio (model/observed; $\sigma_{ratio,t}$)	
Spatial features (location)	Monthly time step average spatial (or pattern) correlation (r_p)	1980–1999 mean field for each month
Spatial features (amplitude)	Monthly time step average spatial standard deviation ratio (model/observed; $\sigma_{ratio,x}$)	

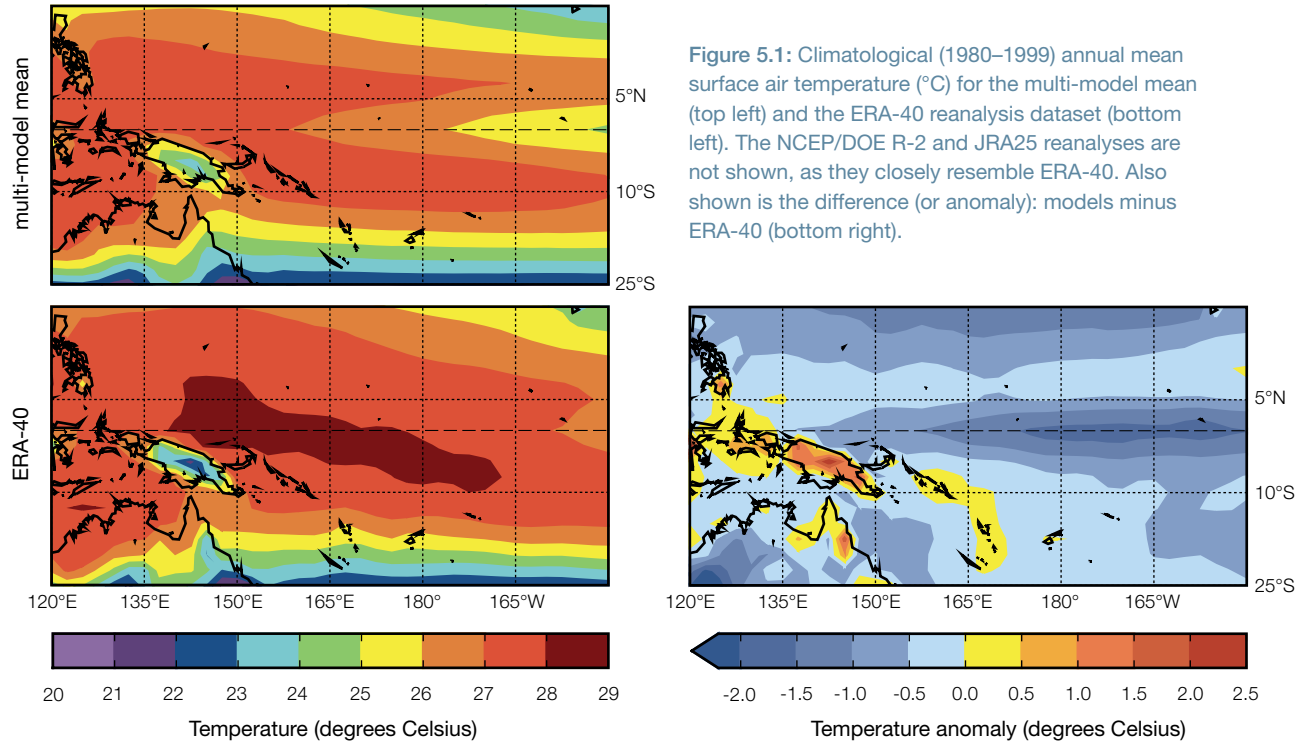

Figure 5.1: Climatological (1980–1999) annual mean surface air temperature (°C) for the multi-model mean (top left) and the ERA-40 reanalysis dataset (bottom left). The NCEP/DOE R-2 and JRA25 reanalyses are not shown, as they closely resemble ERA-40. Also shown is the difference (or anomaly): models minus ERA-40 (bottom right).

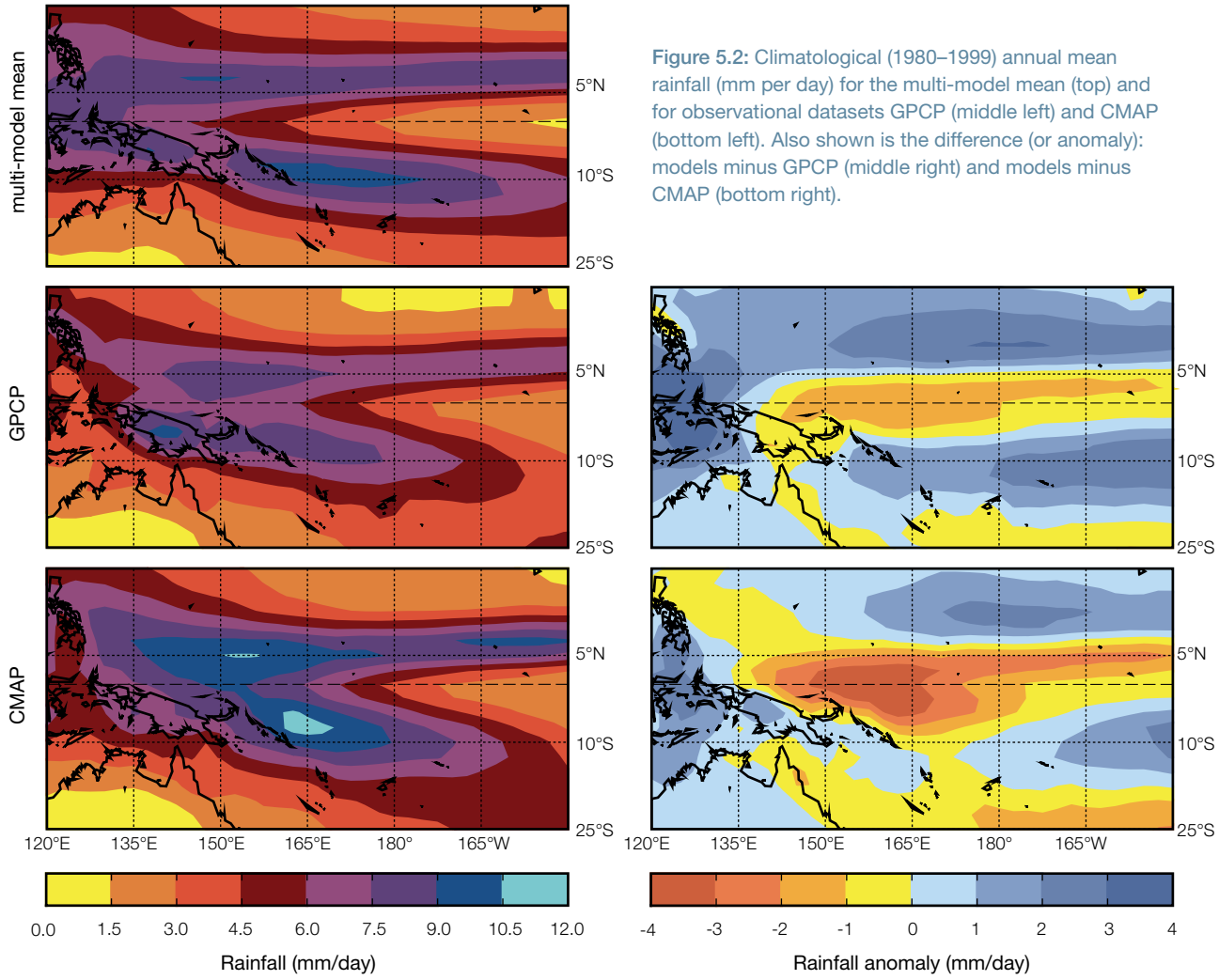

Figure 5.2: Climatological (1980–1999) annual mean rainfall (mm per day) for the multi-model mean (top) and for observational datasets GPCP (middle left) and CMAP (bottom left). Also shown is the difference (or anomaly): models minus GPCP (middle right) and models minus CMAP (bottom right).

5.2.1.3 Near-Surface Wind

The CMIP3 models tend to closely agree with observed climatological wind surface speed data (at 10 m above the ground) throughout most of the PCCSP region, except for the area immediately to the north and east of Papua New Guinea and the Solomon Islands (Figure 5.3). Over this area, the model tendency to overestimate the annual mean wind speed relates to model deficiencies in capturing the seasonal wind reversal associated with the West Pacific Monsoon (Section 5.2.3). Further, these sub-regional errors contribute greatly to the overall PCCSP region multi-model mean grid point error magnitude of E_{abs} = 0.9 ± 0.3 metres per second. The models have some difficulty in capturing the phase (r_t = 0.7 ± 0.1) and spatial pattern (r_p = 0.8 ± 0.1) of the seasonal cycle in wind speed.

The multi-model grid point error magnitude for wind direction is E_{abs} = 14.7 ± 4.3°, however on a sub-regional scale errors larger than 10° are typically confined to the region in and around Indonesia and Papua New Guinea. These directional errors may also arise from deficiencies in the simulation of the West Pacific Monsoon, however any mismatch between the model and reanalysis topography in the region would also be a factor. The models tend to underestimate the magnitude of the seasonal cycle in wind direction across the entire PCCSP region ($\sigma_{ratio,t}$ = 0.9 ± 0.2) and also show some difficulty in capturing the phase (r_t = 0.7 ± 0.1) and spatial pattern (r_p = 0.5 ± 0.1) of the seasonal cycle in wind direction.

5.2.1.4 Summary: Atmosphere Variables

In summary, the CMIP3 models are able to capture the broad-scale characteristics of the 1980–1999 average surface air temperature, rainfall and surface wind climatology. However they display some important deficiencies in simulating the finer details. In particular, the cold tongue bias influences the model simulated climate of Kiribati and Nauru, while many other Partner Countries are influenced by biases in the location of the tropical convergence zones (SPCZ and ITCZ) and the representation of seasonal wind reversals associated with the West Pacific Monsoon. The INM-CM3.0 and PCM models provide consistently poor simulations of the average climate across all atmospheric variables examined (Irving et al., in press). It is recommended that these models are not used for projections of future climate in Chapters 6 and 7 (Section 5.5).

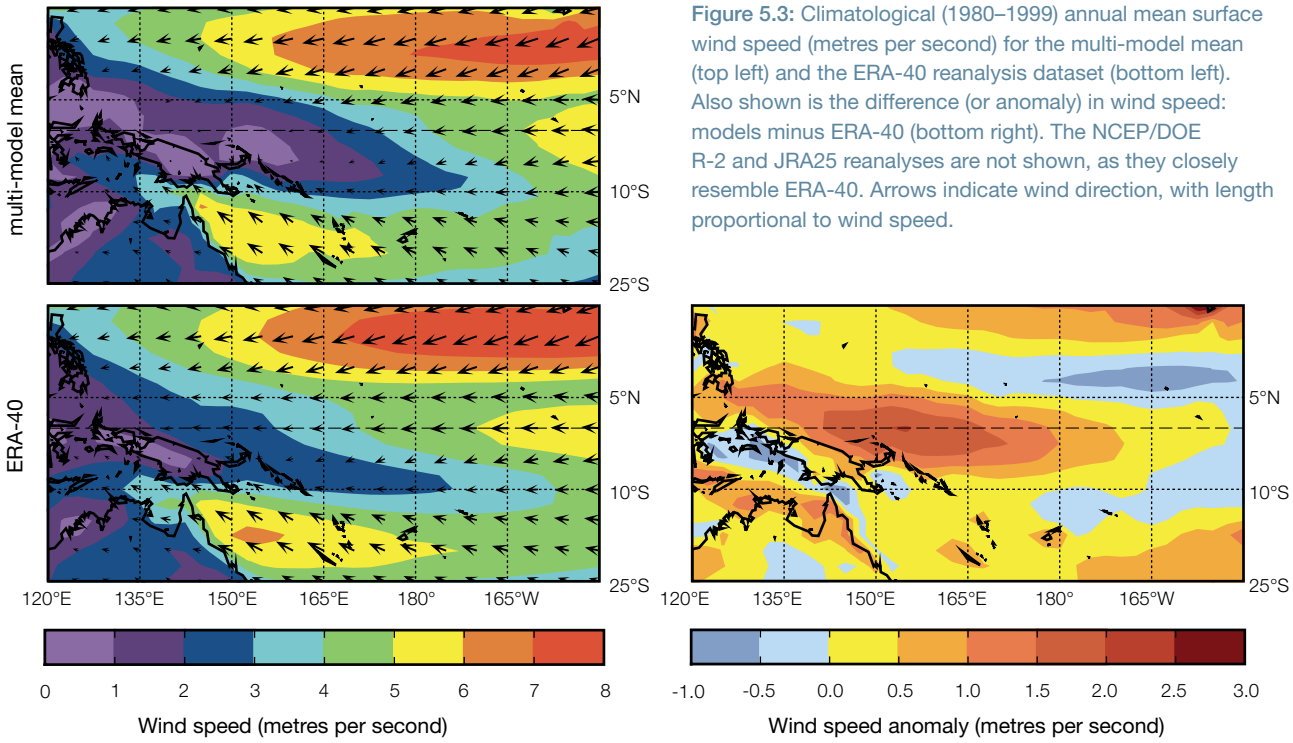

Figure 5.3: Climatological (1980–1999) annual mean surface wind speed (metres per second) for the multi-model mean (top left) and the ERA-40 reanalysis dataset (bottom left). Also shown is the difference (or anomaly) in wind speed: models minus ERA-40 (bottom right). The NCEP/DOE R-2 and JRA25 reanalyses are not shown, as they closely resemble ERA-40. Arrows indicate wind direction, with length proportional to wind speed.

5.2.2 Ocean Variables

5.2.2.1 Sea-Surface Temperature

Comparison of observed and multi-model mean sea-surface temperature (Figure 5.4) shows that important large-scale features are generally captured with some degree of fidelity in the simulations. These include: (1) the West Pacific Warm Pool, (2) the warm bands straddling the equator and (3) the cold tongue in the eastern Pacific. However, systematic errors are also evident. The difference between the multi-model mean and observations of sea-surface temperature is ~0.7°C on average at any given location (root mean square error is 0.73°C). Across the equatorial region, this error is considerably larger, as the sea-surface temperature is too cool in the majority of models. Despite a zonal (east-west) sea-surface temperature difference that is close to observations in many of the models, the cold tongue extends too far into the western Pacific, thereby reducing the extent of the Warm Pool and making it generally too cold. This bias is also manifest in air temperatures and has implications for the wind distribution, the location of the atmospheric convergence, and hence rainfall.

Considerable differences in sea-surface temperature distribution are evident across the models (Figure 5.5). Two models (GISS-ER and GISS-AOM) show little evidence of an equatorial cold tongue and consequently a lack of distinct convergence zones. This indicates a significant failing in their ability to properly capture tropical processes. In a few models (in particular INM-CM3.0 and BCCR-BCM2.0) the cold tongue reaches well into the western basin – a bias that adversely affects mean rainfall in the western Pacific region. Average sea-surface temperatures over the Pacific region are systematically too low, which also affects surface air temperatures (Figure 5.1). While two models are about 0.5°C too warm, the majority (including three of the flux-adjusted models) are more than 0.5°C too cold. Both the BCCR-BCM2.0 and CNRM-CM3.0 models are more than 1.8°C too cold.

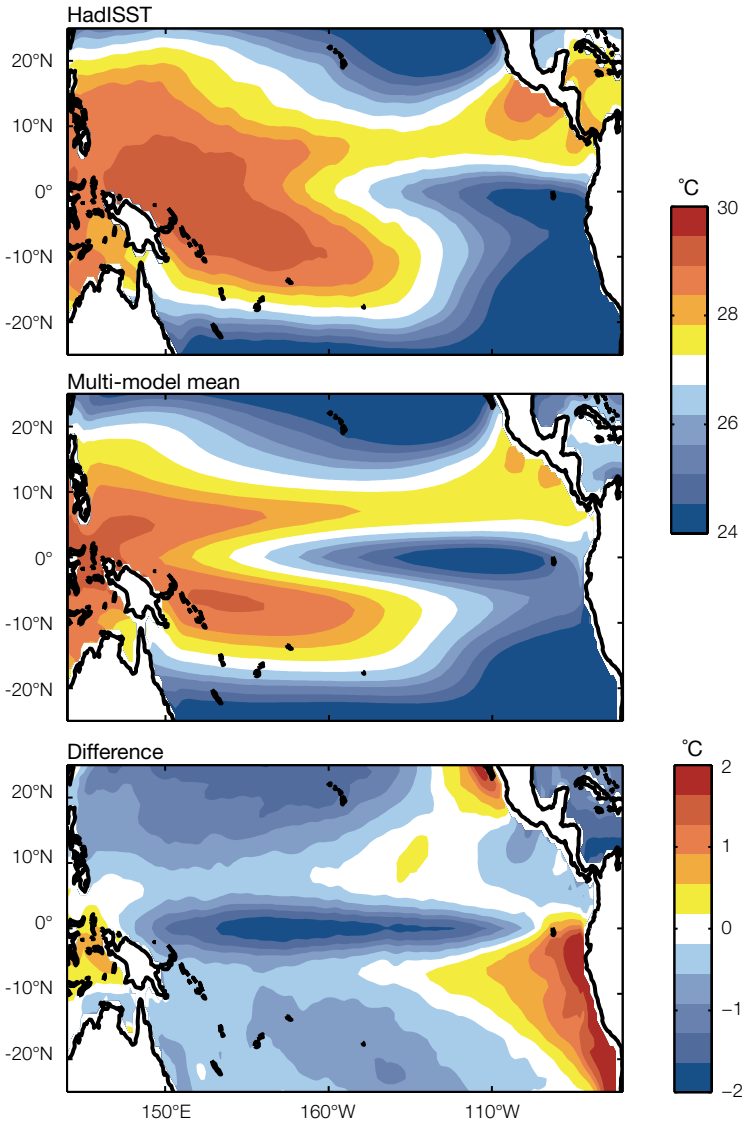

Figure 5.4: Climatological (1980–1999) annual mean sea-surface temperature (°C) for the observed HadISST dataset (top) and the multi-model mean (middle). Also shown is the difference, i.e. models minus observations (bottom).

5.2.2.2 Sea-Surface Salinity

The large-scale observed features in sea-surface salinity are captured in the multi-model mean of the PCCSP region (Figure 5.6), in particular, the low salinity under the convergence zones, where high rainfall levels dilute the salinity concentration, and the high salinity in the south-eastern Pacific where there is low rainfall and high rates of evaporation. Examination of individual models shows very different sea-surface salinity patterns and a large spread in regional averages (Figure 5.5). While the majority of models have average salinities between 34.3 and 35.2 practical salinity units (psu) in the PCCSP region, three models (CSIRO-Mk3.5, UKMO-HadCM3 and PCM) have average salinities below 33.9 psu (Figure 5.5). The multi-model root mean square error for sea-surface salinity in the region is ~0.48 psu.

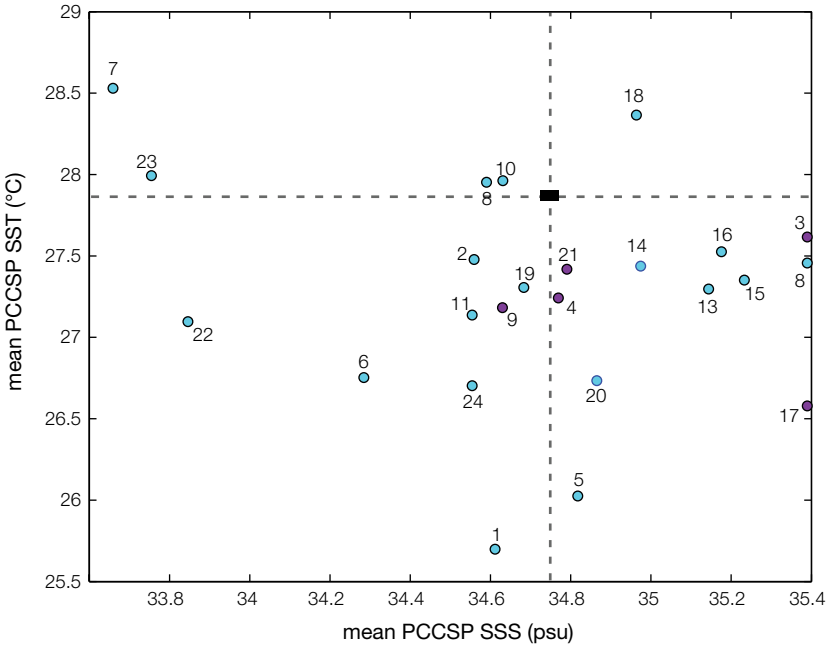

Figure 5.5: Average (1980–1999) PCCSP region sea-surface temperature and sea-surface salinity for individual CMIP3 models (numbered circles, see Table 4.1 for model numbers). The observed range (black box) spans two sea-surface temperature (SST) (HadISST and ERSSTv3) and two sea-surface salinity (SSS) (CARS06 and WOA05) observational datasets. Purple circles indicate flux-adjusted models (see the Glossary).

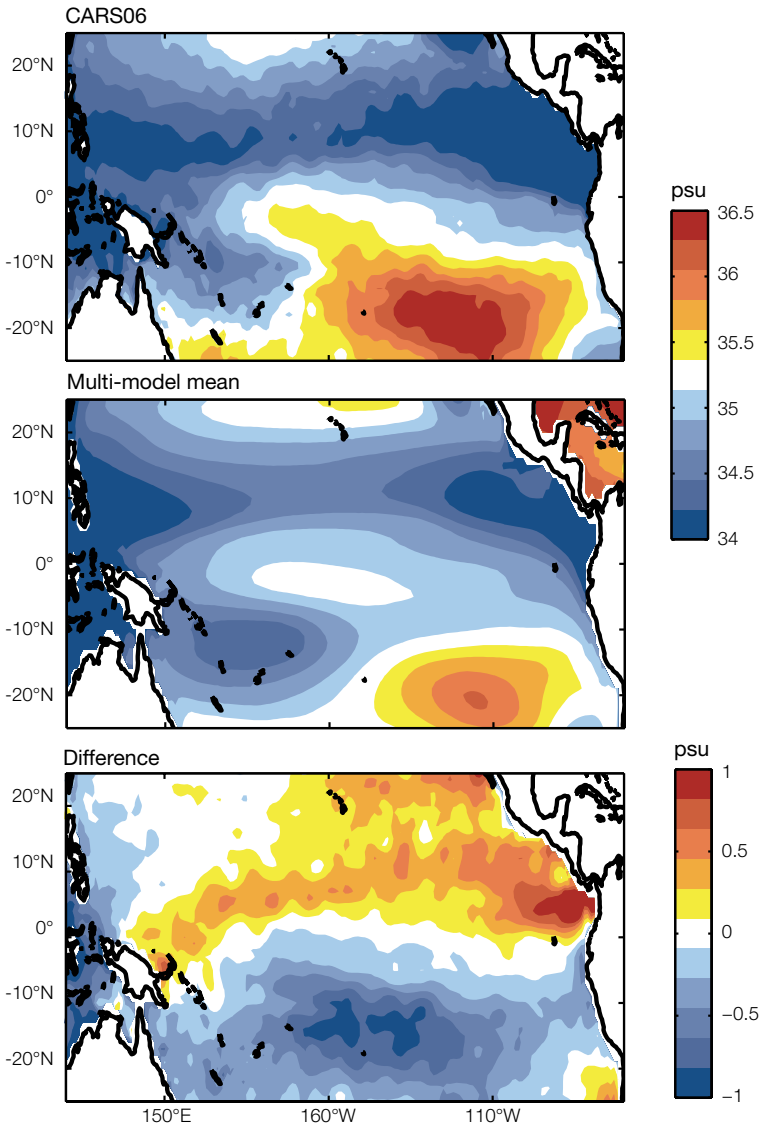

Figure 5.6: Climatological (1980–1999) annual mean sea-surface salinity (psu) for observed CARS06 dataset (top) and multi-model mean (middle). Also shown is the difference, i.e. models minus observations (bottom).

5.2.2.3 Trends and Variability

An important test for climate models is how well they are able to simulate known climate variability on interannual time scales, and how well they can reproduce historical trends. Such an assessment is complicated by the fact that observations are often sparse (particularly prior to the mid-20th century) and often come from multiple sources, making it difficult to combine observations into a consistent record.

A comparison of observed PCCSP region sea-surface temperature trends with individual climate models shows that all the models simulate a warming over the latter half of the 20th century (Figure 5.7). However, there is considerable spread, with some models producing more than double the observed warming, and others showing a very weak warming. Despite this, the multi-model mean compares well with observations. This does not imply that the multi-model mean is necessarily the best predictor of future climate.

The Pacific Decadal Oscillation is important (Chapter 3) as it can enhance or suppress ENSO impacts on decadal time scales. While most models simulate a Pacific Decadal Oscillation reasonably well, a number of models underestimate its strength, and three models (BCCR-BCM2.0, GISS-AOM and GISS-ER) give a sea-surface temperature response to the Pacific Decadal Oscillation that is of the wrong sign (Figure 5.7).

In addition to appropriately simulating natural variability such as ENSO and the Pacific Decadal Oscillation, models should be able to respond realistically to changes in insolation and to major volcanic eruptions. All models (except MIROC3.2 (medres)) that include the effect of volcanoes correctly simulate a cooling in the PCCSP region associated with increased volcanic activity, although the response is generally considerably larger than observed (Figure 5.7). In addition, all models, (except for GFDL-CM2.1) that include changes in insolation, simulate a warming in the PCCSP region associated with increased insolation.

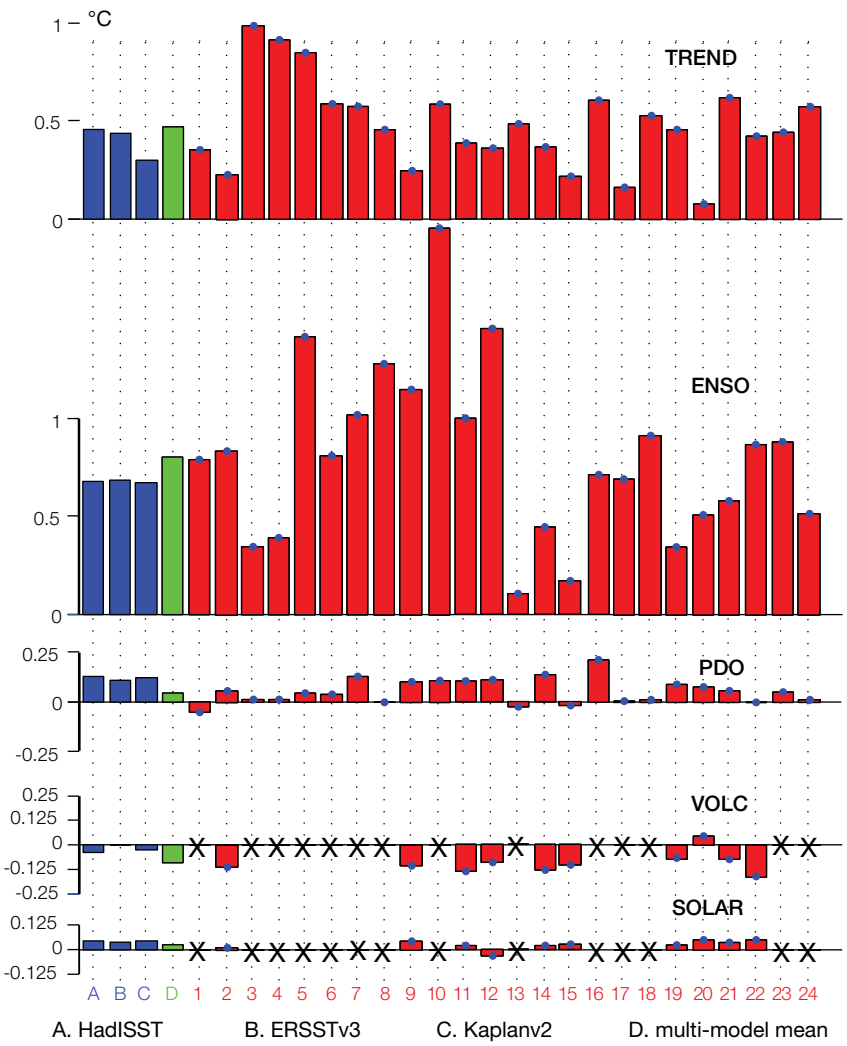

Figure 5.7: Sea-surface temperature trend (1950–2000) and the strength of sea-surface temperature variability associated with ENSO, Pacific Decadal Oscillation, volcanic and solar forcing for three observational datasets (HadISST, ERSSTv3b and Kaplanv2), the multi-model mean, and the CMIP3 models. ENSO and Pacific Decadal Oscillation strength relate to the sea-surface temperature response in the Niño3.4 region, 5°N–5°S, 120°W–170°W (as both of these phenomena have very different regional responses) while other forcing and the sea-surface temperature trend relates to the full PCCSP region (as changes are regionally homogeneous). X indicates models that do not incorporate solar and/or volcanic forcing in the 20th century simulation. See Table 4.1 for model numbers.

5.2.2.4 Sub-Surface Trends

Understanding what goes on beneath the ocean surface is fundamental to fully understanding climate variability and change. The ocean has absorbed about 90% of the total additional heat from global warming and about half of the CO_2 emitted since industrialisation (Sabine et al., 2004). Its huge inertia means that changes are less subject to the short-term fluctuations that are ubiquitous for atmospheric properties. Hence ocean heat content and sea-level change is often a more reliable indicator of climate change. The ocean plays a central role in ENSO and longer-term variability and takes part in important feedback processes that modulate the rate of long-term atmospheric warming. Physical ocean processes such as mixing are also vital for bringing essential nutrients into the surface ocean. These nutrients maintain biological productivity and ultimately the fisheries that are so important for the region.

The sub-surface temperature trend (1950–2000) shows a surface-intensified warming in both the models and observations consistent with increased greenhouse gas concentrations (Figure 5.8). Counter-intuitively, there is also a region of cooling centred near the equator at about 200 m depth. This cooling is consistent with a weakening in the equatorial trade winds that (through ocean processes) cause the thermocline to become shallower (Han et al., 2006) lifting the deeper cold water upwards. While the observed warming-cooling pattern could be caused by natural variability (McPhaden and Zhang, 2004), the fact that it is reproduced in the climate models (albeit with reduced intensity) suggests that the pattern is, in part, related to increased greenhouse gas concentrations. This is consistent with atmospheric studies that attribute the weakening of the Walker Circulation and the associated Pacific trade winds to both natural variability and the influence of greenhouse gases (Power and Kociuba, in press).

5.2.2.5 Climate Model Drift

Climate models that include historical or projected increases in atmospheric greenhouse gas concentrations would be expected to exhibit a global warming trend, with associated changes in other climate characteristics (e.g. rainfall, circulation). Conversely, climate simulations with greenhouse gas concentrations held fixed (i.e. control simulations) would be expected not to exhibit long-term trends. Many models do, however, have trends in their control simulations as a result of imperfections in the representation of physical processes in the models. Such spurious trends are referred to as climate model drift (Power, 1995). This is problematic, as a model that exhibits drift within the control simulation will also exhibit similar drift in the historical and projection simulations. The model drift therefore erroneously affects any global warming signal in the model output. Drift must therefore be accounted for when it is large.

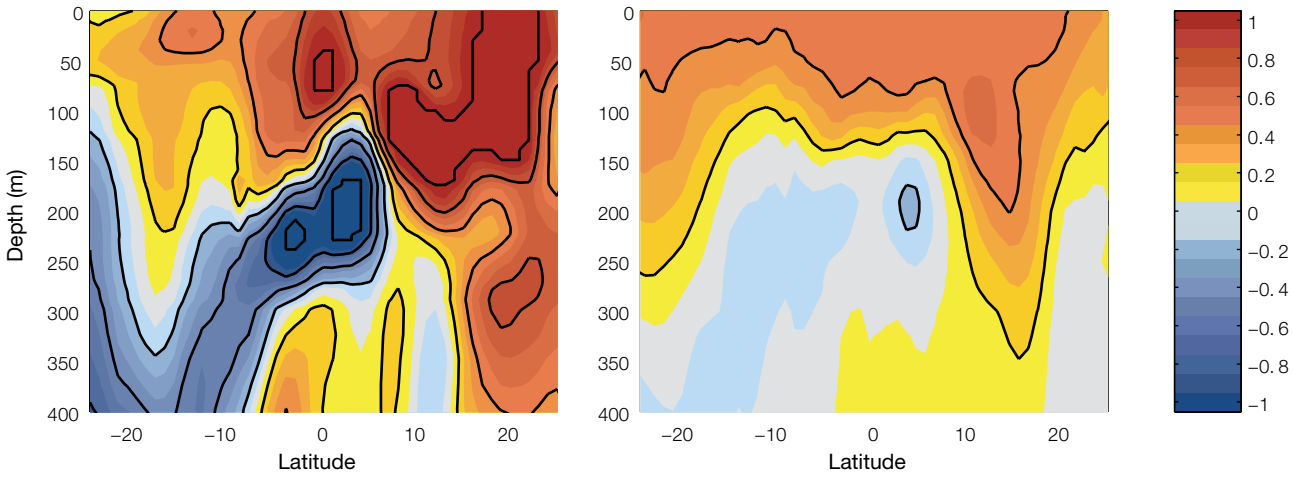

Figure 5.8: Linear trend across the equator at 165°E in sub-surface ocean temperatures based on the multi-model mean (right) and an observational reconstruction (left; Durack and Wijffels, 2010). Units are °C per 50 years; black contours every 0.2°C per 50 years.

The typical error introduced by model drift (if not accounted for) in the simulated trend for the PCCSP region is shown in Figure 5.9. For most models, the trend associated with model drift is less than 20% of the simulated 1950–2000 sea-surface temperature trend. In one model (INGV-SXG) the drift error is particularly pronounced, exceeding 40%. Model drift introduces a greater error in rainfall, with most models showing an error of 10 to 40%. Two models, MRI-CGCM2.3.2 and INGV-SXG, have rainfall drift errors that exceed 60%. In general, INGV-SXG shows high levels of drift in all climate variables. In addition, it is not possible to correct for model drift for this model as a concurrent control simulation is unavailable.

Model drift becomes relatively less important for projections as the anthropogenic climate signal become larger. In addition, model drift tends to be randomly distributed across models (i.e. it will make a trend larger in some models and smaller in others). As a result, model drift tends to be negligible when considering multi-model means. Model drift is often most important in the ocean, particularly at greater depth. As a result, ocean model fields used in projections of ocean variables (other than sea level) have been corrected for model drift by subtracting the drift in the control experiment from corresponding projection simulations.

5.2.2.6 Sea Level

Mean Sea-Level Pattern

The observed sea-level pattern was presented in Section 2.6.4 (see Chapter 4 and Gregory et al. (2001) for a description of techniques to calculate sea level from the models).

Globally, the modelled sea level reproduces the large-scale features of the observed pattern but some models have a higher correlation with the observations and a smaller root mean square error than others (not shown). For the tropical and sub-tropical Pacific basin, there is a similar level of agreement but some models have closer agreement with the observations (Figure 5.10). There is a smaller root mean square error and larger spatial correlation for the multi-model mean sea-level distribution (for the 17 models available) than the individual models (Table 5.2, Figure 5.10 and Figure 5.11), where larger inter-model differences of regional sea-level patterns can be seen (Figure 5.10). The root mean square error between the mult-model mean and the observations is 0.09 m (Table 5.2 and Figure 5.11) over the region (120°E–70°W, 45°S–45°N) and the spatial correlation is very high at 0.97 (Table 5.2).

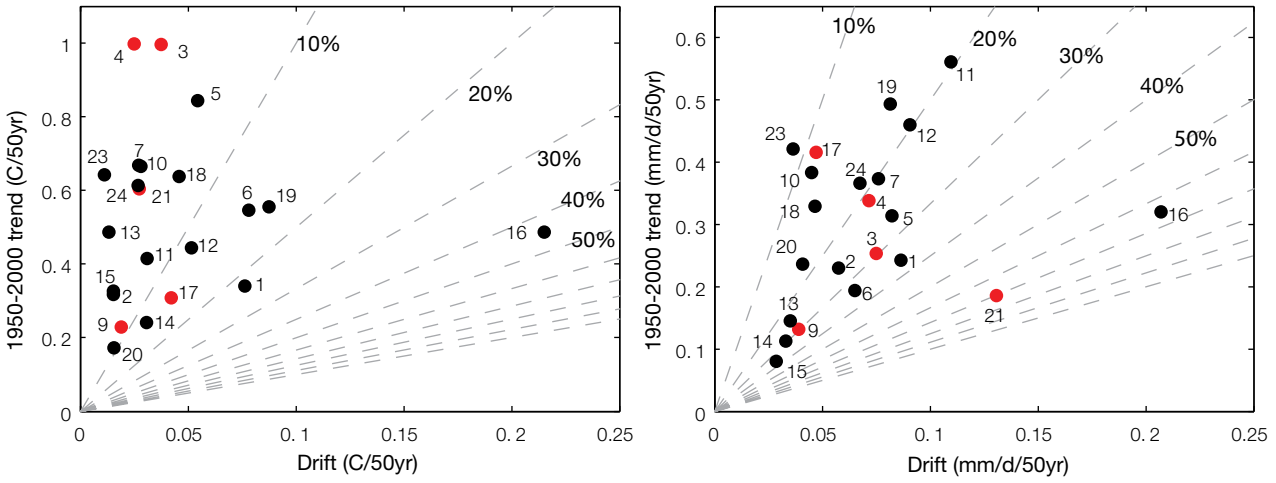

Figure 5.9: Relative size of climate drift in the CMIP3 models (based on their respective simulation with fixed greenhouse gases concentrations: the control run) versus the simulated 1950–2000 trend (where greenhouse gases concentrations follow historical levels) for surface air temperature (left) and rainfall (right). Model numbers are defined in Table 4.1 (Model 16 is INGV-SXG). Red dots indicate models that use flux adjustment (a correction made to some models to reduce the effect of model drift, see Glossary).

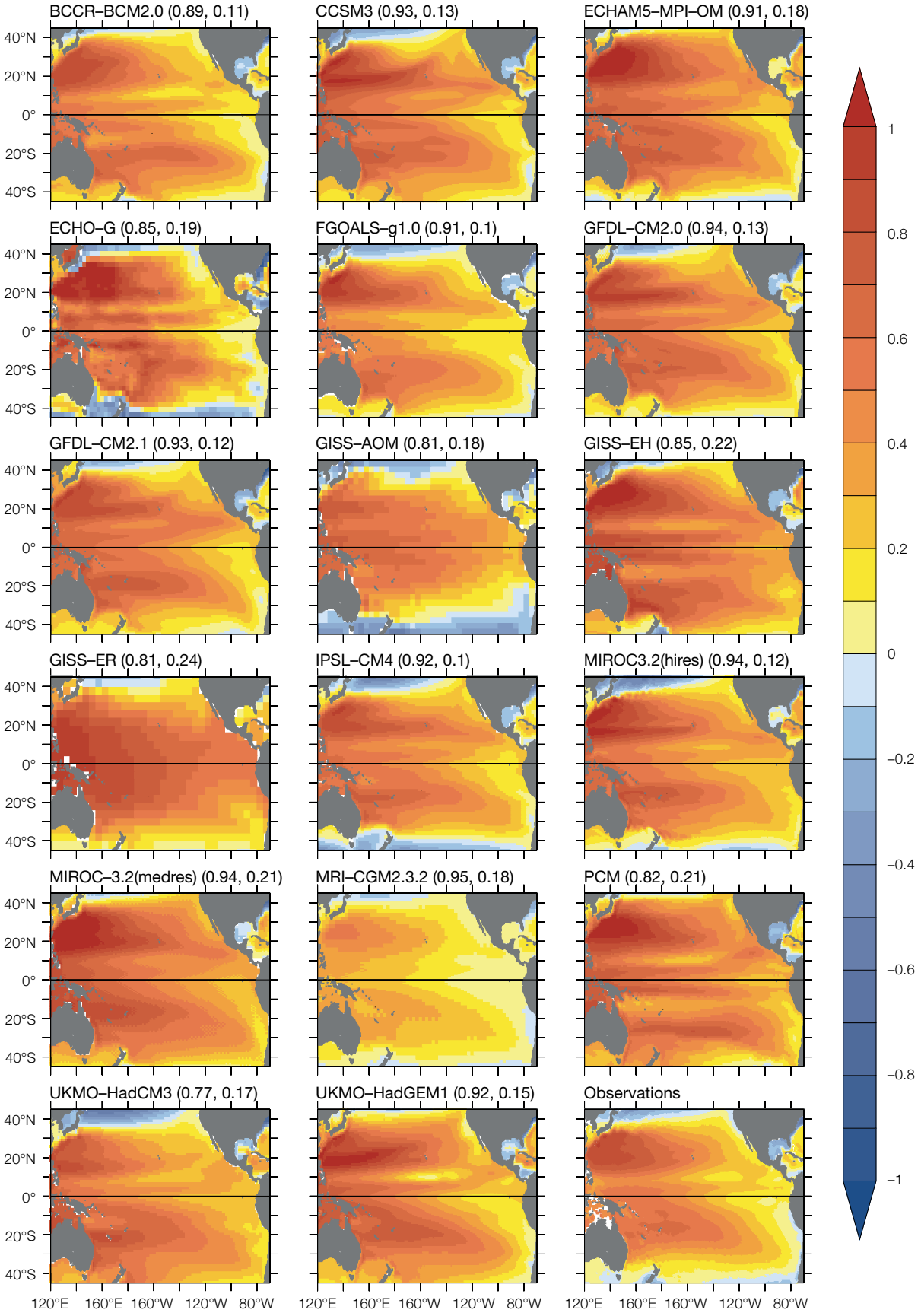

Figure 5.10: Tropical and sub-tropical Pacific basin mean sea level (in metres) from each climate model and observations for the period 1993–2000. The first number in brackets above each panel is the spatial correlation of mean sea level between each climate model and the observations for the tropical and sub-tropical Pacific basin (120°E–70°W, 45°S–45°N), while the second number is root mean square error. The global mean, averaged over the same period, is removed from each panel.

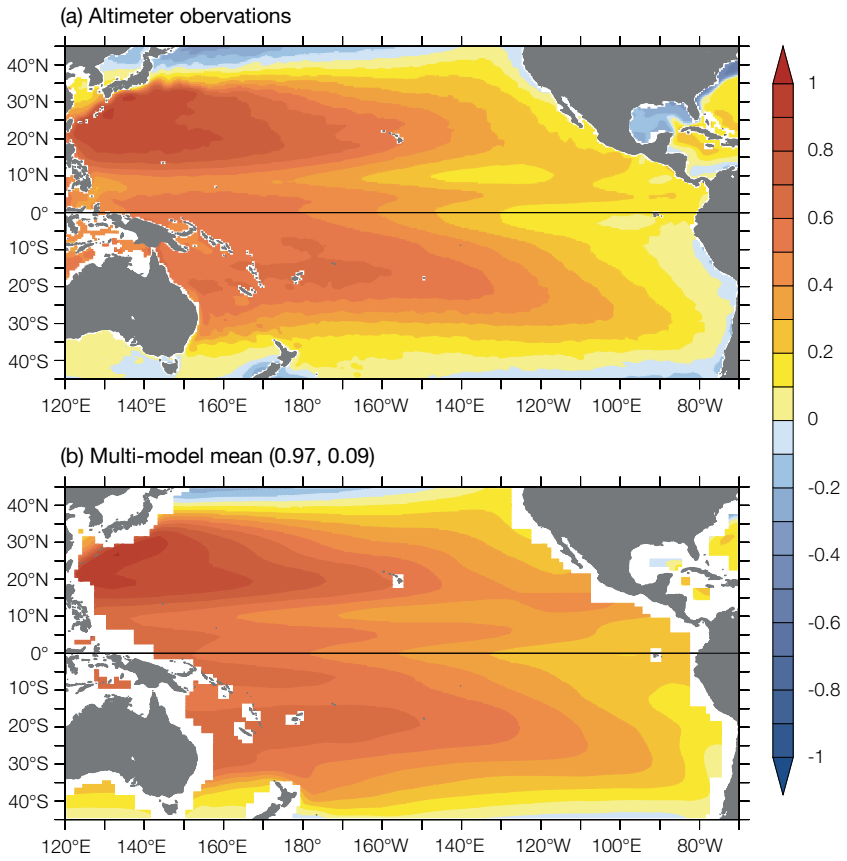

Figure 5.11: Tropical and sub-tropical Pacific basin mean sea level (in metres) based on (a) observations and (b) the multi-model mean for the period 1993–2000. The spatial correlation between (a) and (b) is 0.97, while the root mean square error is 0.09 m. The global mean is removed from each panel.

Table 5.2: Spatial correlation and root mean square (RMS) error of mean sea level (in metres) between each climate model and observation for the period 1993–2000 over the tropical and sub-tropical Pacific basin (120°E–70°W, 45°S–45°N).

Climate model	Spatial correlation	RMS error (m)
BCCR-BCM2.0	0.89	0.11
CCSM3	0.93	0.13
ECHAM5/MPI-OM	0.91	0.18
ECHO-G	0.85	0.19
FGOALS-g1.0	0.91	0.10
GFDL-CM2.0	0.94	0.13
GFDL-CM2.1	0.93	0.12
GISS-AOM	0.81	0.18
GISS-EH	0.85	0.22
GISS-ER	0.81	0.24
IPSL-CM4	0.92	0.10
MIROC3.2(hires)	0.94	0.12
MIROC3.2(medres)	0.94	0.21
MRI-CGCM2.3.2	0.95	0.18
PCM	0.82	0.21
UKMO-HadCM3	0.77	0.17
UKMO-HadGEM1	0.92	0.15
Multi-model mean	0.97	0.09

Sea-Level Trends

Robust projections of sea-level rise depend critically on understanding the ability to adequately represent past sea-level changes in model simulations. In the IPCC Fourth Assessment Report (Bindoff et al., 2007) it was pointed out that the inability to satisfactorily explain the observed sea-level rise over decades has been a significant limitation in all of the IPCC assessments to date and a barrier to narrowing projections of sea-level rise.

Consistent with these results, Rahmstorf et al. (2007) showed that the observed sea level from tide gauges (1990–2001) and from satellite altimeter data (1993–2006) was rising at the upper limit of the projections from the IPCC (2001). Church et al. (2011) have also compared the observed sea-level rise derived from an improved sea-level reconstruction (Church and White, in press) and a longer altimeter time series, with the projected sea-level rise based on the IPCC projections (2007). By the end of the observational time series, both the reconstructed and altimeter sea-level datasets are close to the top of the projections.

Climate model simulations that only included the impact of greenhouse gases (and not the natural climate-influencing factors such as solar variability and stratospheric aerosols from large volcanic eruptions) have significantly larger trends and significantly less variability (Domingues et al., 2008) than simulations with improved upper-ocean heat content and steric sea-level trends (Domingues et al., 2008; Ishii and Kimoto, 2009; Levitus et al., 2009). In contrast, the variability in models that do include these natural climate-influencing factors is similar to the observed ocean variability but the trend in both ocean heat content and thermosteric sea-level rise is slightly smaller than the observations (Domingues et al., 2008). Model simulations which include natural factors also suggest that steric sea level would have fallen by several

millimetres following the eruption of Mount Pinatubo in 1991 and that the ocean recovery from this cooling could add about 0.5 mm per year to the rate of steric sea-level rise over the decade from 1993, coincident with the first decade of high quality satellite altimeter observations (Church et al., 2005; Gregory et al., 2006).

There have been updated glacier and ice cap mass inventories (Radic and Hock, 2010) resulting in a larger ice volume estimate, and permitting the mass changes of Greenland and Antarctica to be explicitly included in projections. The dataset of glacier mass balance has been expanded by using the 'geodetic' observations of glacier volume change in addition to the surface mass-balance observations (Cogley, 2009). Radic and Hock (2011) have allowed for glacier and ice cap hypsometry (changing area with altitude) in recent projections. Most recently, LeClercq et al. (2011) have used glacier length observations together with the surface mass balance observations compiled by Cogley (2009) to estimate glacier and ice cap contributions to sea level since 1800. Current methods reproduce these observations but significant uncertainty in parameters used for projections remain and the methods used in the IPCC (2007) report do not allow glacier and ice cap contributions to come to a new equilibrium in a warmer climate.

In the IPCC (2007) report, the ice sheet projections were completed using empirically calibrated schemes for melting as a function of temperature change. More recent work employs climate models at high resolution in the ice sheet regions and incorporates detailed physical models of the surface energy balance, melting and runoff, and snow accumulation, both for Greenland (Fettweis, 2007) and Antarctica (Krinner, 2007). In the IPCC (2007) report, it was recognised that the current suite of ice sheet models was not capable of adequately simulating a potential rapid response of the ice sheets to global warming.

Much attention has been directed in recent years to developing improved models of ice sheet dynamics, and some first results are now available (Joughin et al., 2010). More are expected as an outcome of two large ongoing activities, the SeaRISE project supported by NASA, and the EU-funded ice2sea project.

5.2.2.7 Summary: Ocean Variables

In summary, while climate models are able to reproduce most of the important oceanic features, natural variability in the ocean, and the direction of historical changes, there are still significant biases that must be considered when applying these models to projections. The INGV-SXG model displays a large climate drift in all variables, and does not provide a corresponding control simulation to correct for climate drift. It is recommended that this model is not used for projections of future climate in Chapters 6 and 7 (Section 5.5).

Global climate models reproduce the observed pattern of the regional distribution of sea level with a fair degree of realism. Models that include natural and human emissions as climatic factors have similar temporal variability in global-averaged ocean thermal expansion as the observations, but a slightly smaller rate of rise over recent decades.

5.2.3 Climate Features and Patterns of Variability

In this section, the ability of models to simulate the major features and patterns of variability of climate in the PCCSP region is evaluated. A detailed description of the observed climate features (the SPCZ, ITCZ and West Pacific Monsoon) and components of climate variability (ENSO and the Indian Ocean Dipole) is given in Chapters 2 and 3.

5.2.3.1 El Niño-Southern Oscillation

As ENSO is a major component of climate variability in the PCCSP region, it is essential that it is well simulated in climate models. To simulate ENSO, a climate model needs to accurately simulate atmosphere-ocean interactions on all relevant time and space scales. In particular, a model requires a realistic mean climate, including the Walker Circulation and zonal and vertical ocean temperature gradients; realistic daily to weekly wind variability that helps trigger ENSO events; variations in the thermocline depth and slope; and appropriate ocean-atmosphere interactions. A key factor in successfully simulating all these features is sufficient spatial and temporal resolution to represent small-scale processes, such as ocean eddies. Other important model characteristics include realistic convection processes, representation of clouds and the feedbacks between clouds and radiation.

The adequate simulation of all these features contributes to a more realistic representation of the spatial pattern and strength of ENSO events, the frequency of events, the characteristics of different types of events, and the link between ENSO and climate variables such as rainfall. Reproducing ENSO-like behaviour in climate models is a very complex task and it is a significant achievement that, unlike two decades ago, most climate models simulate an ENSO-like phenomenon. The climate models involved in the CMIP3 comparisons (Section 4.3.1) perform much more realistically than previous generations of models because resolution, model formulation and representation of sub-grid scale processes (parameterisations) have been improved. Thus, many models are extremely useful tools for studying ENSO and its impacts (Guilyardi et al., 2009; Collins et al., 2010).

Challenges in simulating ENSO behaviour still remain. Of the CMIP3 models, a small minority fail to adequately reproduce ENSO variability, probably due to coarse

ocean resolution (van Oldenborgh et al., 2005). Two models in particular (GISS-AOM and GISS-ER) show virtually no ENSO-like variability. The remaining models have a pattern of El Niño warming that extends too far into the western Pacific (Figure 5.12). This means that in some models, some Pacific Islands experience sea-surface temperature anomalies of the wrong sign and therefore the opposite impact of the El Niño event to what is observed. This bias also explains most of the error in the sea-surface temperature pattern correlations for canonical ENSO.

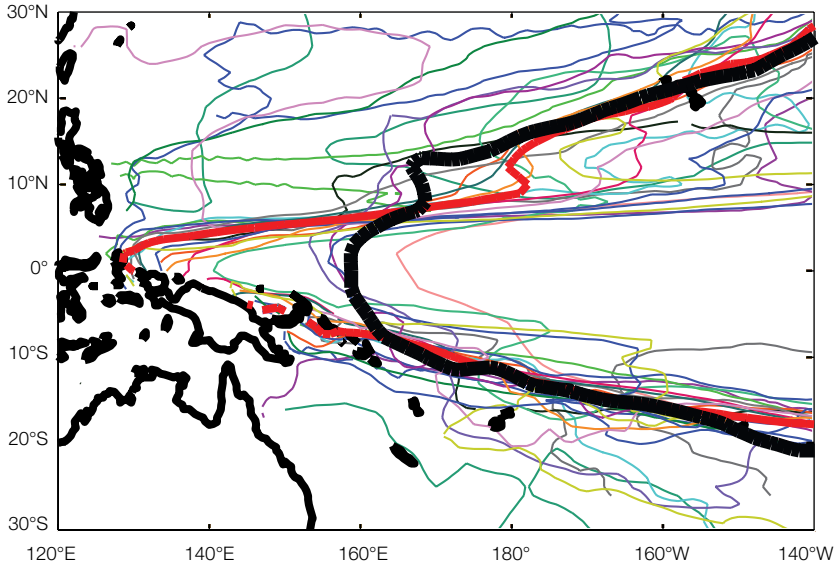

Figure 5.12: Zero-line for sea-surface temperature response to ENSO for individual model (thin lines), multi-model mean (thick red) and observed (HadISST, thick black). During an El Niño event, warming (cooling) occurs to the east (west) of this line. Clearly evident is the large systematic bias in the model sea-surface temperature response over the equatorial western Pacific. This has important implications for the rainfall response to ENSO, which tends to increase (decrease) over warmer (cooler) waters.

Every El Niño and La Niña event is different, but they have been observed to fall into two types based on the spatial patterns of sea-surface temperature anomalies that occur: the canonical ENSO and the El Niño Modoki (Section 3.4.1). The El Niño Modoki is thought to have become more prevalent in recent decades. It remains an ongoing topic of research whether this is due to natural variability or anthropogenic factors (Yeh et al., 2009).

CMIP3 models show varying degrees of skill in reproducing the strength and frequency of ENSO events. Observed ENSO events undergo a highly irregular two to seven year oscillation, but in some models ENSO events occur too often and too regularly, e.g. an El Niño or La Niña every two years (Randall et al., 2007), or too seldom (Joseph and Nigam, 2006). The variability of the strength of canonical and Modoki events (measured by standard deviation of Niño3.4 and El Niño Modoki Index respectively, Figure 5.13 left), vary substantially from very weak to overly strong. In general, models with excessively strong canonical ENSO variability also have excessively strong Modoki variability, or too weak for both, reflecting general sea-surface temperature variability in the tropics.

The Modoki pattern was identified in observations as the second most important pattern in explaining monthly variability of sea-surface temperature in the tropical Pacific region (Ashok et al., 2007). The first pattern is the canonical El Niño pattern of sea-surface temperature variability, which accounts for 45% of tropical sea-surface temperature variability, compared to 12% for El Niño Modoki. The observed spatial patterns of the two modes are very similar to the patterns in Figure 3.4. In order to determine whether the spatial patterns of these forms of variability in the CMIP3 models correspond to those observed, they have been calculated for all models and for observations (HadISST) from their linearly de-trended monthly sea-surface temperatures from 1975–1999, using Empirical Orthogonal Function analysis (von Storch and Zwiers, 2000). Pattern correlations between the observed and model patterns (Figure 5.13 right) show that most of the CMIP3 models do not successfully distinguish the canonical and Modoki types of ENSO variability. The canonical ENSO pattern is reasonably well reproduced by most models (pattern correlation, r_p is above 0.6 for most models), but in general they do not reproduce the Modoki pattern well (r_p is below 0.6 for most models). Only a small number of models produce a second mode of variability that is recognisably similar to the observed Modoki pattern. From the results shown here it also is clear that several models do not reproduce realistic variability for either canonical or Modoki ENSO.

The overly westward extension of sea-surface warming during El Niño events in most models has important implications for the rainfall response to ENSO, which tends to increase over warmer waters and decrease over cooler waters. This is evident from model rainfall maps (Figure 5.14, models 2 and 3), particularly in the region east of Papua New Guinea. Almost all models have an ENSO-related rainfall response in the wrong direction in parts of the western Pacific. Also, those models with weak sea-surface temperature variability tend to have a weak rainfall response (Figure 5.14, model 1), and those with excessively strong sea-surface temperature variability have a strong rainfall response (model 2). Models with realistic ENSO sea-surface temperature variability show a more realistic rainfall response (model 3).

Many models simulate errors in the ENSO related changes in temperature or rainfall that influence particular Partner Countries. As the line between regions of warmer and cooler water during ENSO events is shifted too far west in many models (Figure 5.12), countries may have the wrong sea-surface temperature anomaly during El Niño and La Niña events. The pattern of rainfall response to ENSO events is also shifted too far west in many models, and is too strong or too weak in some models (Figure 5.14). Therefore, models may not simulate the correct changes in rainfall during El Niño and La Niña events for some Partner Countries.

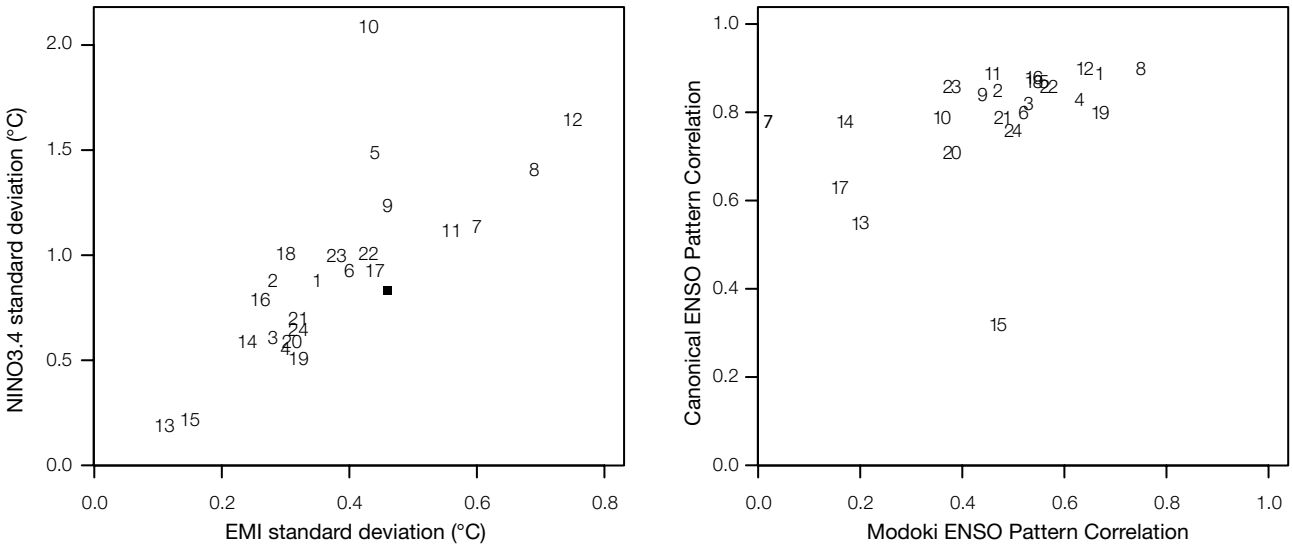

Figure 5.13: (Left) Variability (standard deviations) of the Niño3.4 Index and El Niño Modoki Index in the 24 CMIP3 climate model simulations (model numbers given in Table 4.1). The black square shows the observed standard deviations. (Right) Pattern correlations of canonical (Mode 1) and Modoki ENSO (Mode 2) spatial patterns of sea-surface temperature variability for each model with the observed patterns.

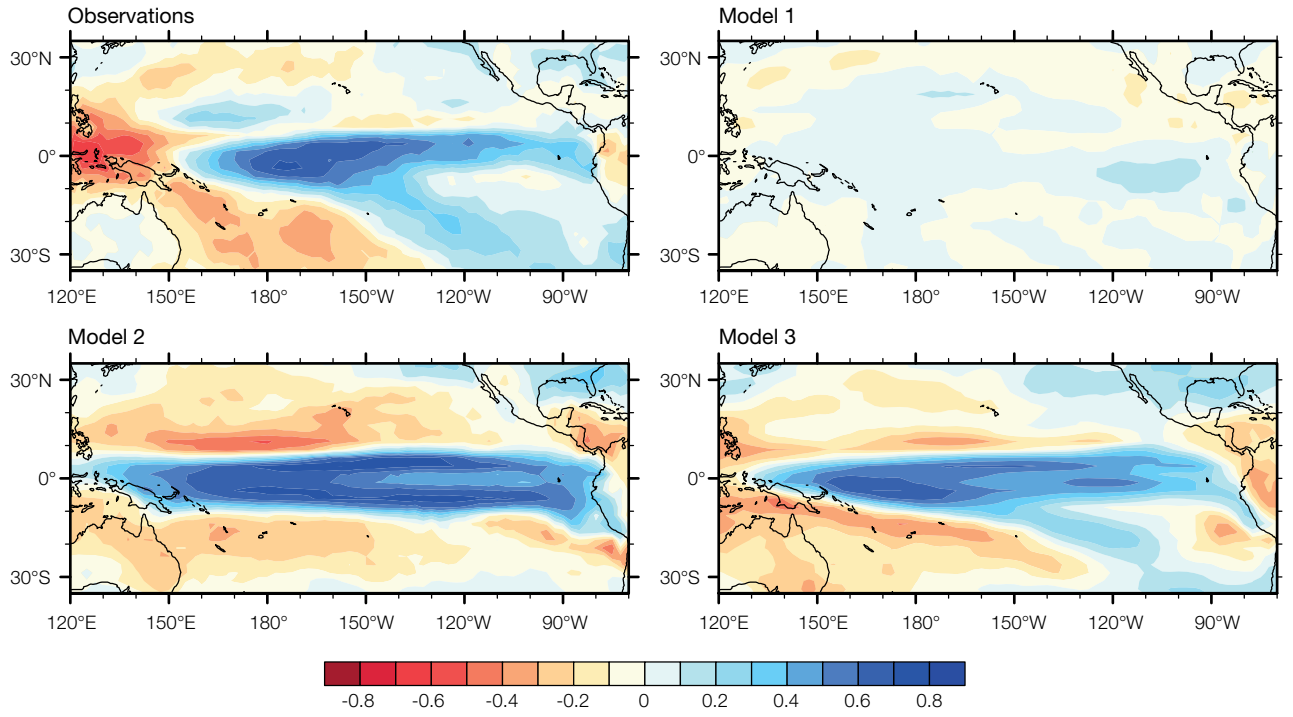

Figure 5.14: Observed (top left) and model (top right and bottom panels) rainfall response to a canonical El Niño event for three models. Contours represent the correlation coefficient between monthly Niño3.4 and rainfall. Positive correlations (blue) show where rainfall typically increases during El Niño events, and negative correlations (red) show where rainfall typically decreases during El Niño events.

5.2.3.2 Indian Ocean Dipole

The simulation of the Indian Ocean Dipole (IOD, Section 3.4.7) in CMIP3 models may be relevant to climate projection for regions affected by the IOD, including East Timor. Most CMIP3 models show some skill in reproducing the spatial pattern of surface temperatures that corresponds to the observed IOD. However the multi-model mean IOD pattern has a stronger east-west gradient than observed (Cai et al., 2009c). In five CMIP3 models (ECHO-G, IPSL-CM4, GISS-AOM, GISS-ER and INM-CM3.0), the IOD sea-surface temperature pattern is unrealistic, with a weak or negative correlation between the observed pattern and the model patterns (Cai et al., 2009c).

5.2.3.3 South Pacific Convergence Zone

The observed SPCZ is described in detail in Chapter 2. The simulation of the SPCZ in the CMIP3 global climate models is examined using December-February and June-August mean rainfall from the last two decades of the 20th century climate simulations compared with CMAP observed rainfall (Table 2.2). The SPCZ position is identified from the line of maximum rainfall in the South Pacific, extending from around 150°E towards the south-east.

The observed CMAP and multi-model mean rainfall for the December-February and June-August seasons (Figure 5.15) show that the observed SPCZ is more intense during December-February, and extends further into the south-east Pacific. During June-August, the ITCZ is more intense than the SPCZ, and the SPCZ breaks into distinct tropical and sub-tropical components. The multi-model mean rainfall also shows a distinct SPCZ band, which is most intense during December-February. The multi-model mean SPCZ has an orientation that is too zonal (east-west), rather than sloping to the south-east, and does not extend far enough into the sub-tropics in the eastern Pacific. In June-August, the multi-model mean rainfall does not show the sub-tropical section of the SPCZ located at around 30°S.

While most models are able to simulate a separate ITCZ and SPCZ, a small number of models do not show a distinct SPCZ (for individual model performance, see Brown et al., in press). GISS-AOM and GISS-ER simulate a single rainfall band over the equator, while MIROC3.2 (medres) and MIROC3.2 (hires) simulate an SPCZ that does not extend east of the International Date Line.

Observations show that the SPCZ tends to move north-east during El Niño years and south-west during La Niña years (Folland et al., 2002; Vincent et al., 2011). The majority of models reproduce this movement of the SPCZ in response to ENSO. This underpins the ability of the models to simulate the influence of ENSO on rainfall and atmospheric circulation in the South Pacific (Brown et al., in press) because ENSO impacts on rainfall and circulation are largely linked to shifts in the SPCZ in this region.

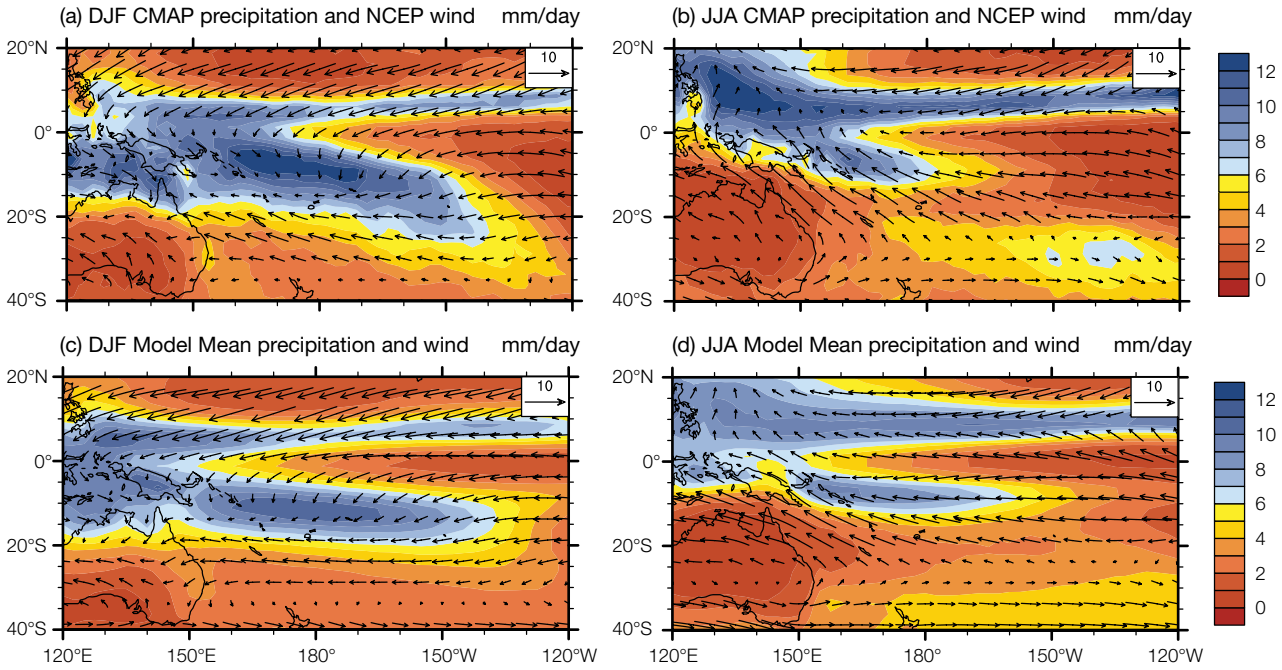

Figure 5.15: (Top left): Observed December-February seasonal mean rainfall (from CMAP; see Table 2.2) and near-surface wind (from NCEP). (Top right): As for top left panel but for June-August. (Bottom left): Multi-model mean December-February seasonal mean rainfall and near-surface wind. (Bottom right): As for bottom left panel but for June-August.

5.2.3.4 Intertropical Convergence Zone

Climate models typically simulate a narrow, zonal rainfall maximum corresponding to the ITCZ, with the exception of three models (GISS-AOM, GISS-ER and MIROC3.2(hires)) which show no Northern Hemisphere ITCZ branch in the western Pacific. These three models should therefore not be used for regional projections in that area. In the central and eastern Pacific, many models display a double ITCZ, with a Southern Hemisphere branch that is not found in the observations. In the multi-model mean there is too much rain in the south-eastern Pacific, with associated December-February northerly wind anomalies over the equator (in the observations this is a region of persistent southerlies).

The observed seasonal variation of the position of the ITCZ has an amplitude of approximately 2° of latitude, with maximum southward extent (although still in the Northern Hemisphere) in February, and a broad maximum of the northern extent between July and September (Waliser and Gautier, 1993). Models represent the December-February location well on average, but show an excessive northward displacement in July-September (Figure 5.16); a bias that occurs in almost all models. Models on average reproduce the seasonal cycle of ITCZ rainfall amounts reasonably well, although uncertainties in the observations are large in this region, precluding stronger conclusions. On interannual time scales, most models are able to reproduce the equatorward shift under El Niño conditions and the poleward shift under La Niña conditions found in the observations (Chapter 3), except for those models not showing significant ENSO variability.

5.2.3.5 West Pacific Monsoon

In the Southern Hemisphere the West Pacific Monsoon consists of the seasonal reversal of wind direction over the region to the north of Australia, extending from East Timor to the Solomon Islands. The observed December-February monsoon winds are north-westerly, and bring convection and heavy rainfall predominantly to East Timor, Indonesia, Papua New Guinea and the Solomon Islands (Figure 5.15 (a) and (b)). In the Northern Hemisphere, corresponding seasonal reversals affect Palau and the Federated States of Micronesia. The spatial pattern of rainfall and winds associated with the monsoon is reasonably well simulated by the multi-model mean (Figure 5.15 (c) and (d)).

Over the Southern Hemisphere box-1 region (Figure 2.10), all models simulate easterly winds that prevail during June-August and most models also simulate either a weakening or a reversal of the zonal wind component during December-February (Figure 5.17). Four models (GISS-AOM, GISS-EH, IPSL-CM4 and MIROC3.2(medres)) fail to simulate a reversal in wind direction.

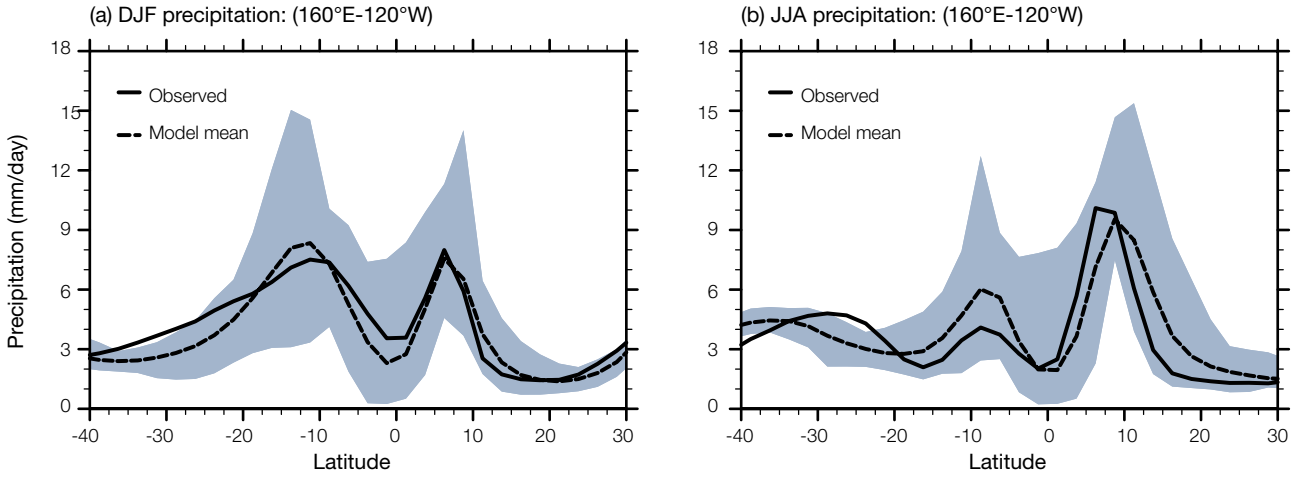

Figure 5.16: Zonal average rainfall for (a) December-February, and (b) June-August, across the Pacific (160°E–120°W). CMAP (Table 2.2) observations and multi-model means are shown as solid and dashed lines, with the inter-model range as blue shading.

When the seasonal rainfall contrast is expressed as the ratio of the December-February rainfall total divided by the June-August rainfall total over the Southern Hemisphere box-1 region (Figure 2.10), the observed ratio is close to two according to the CMAP data (Figure 5.18). Almost all the models simulate a ratio greater than one, with the exception of GISS-EH, which fails to simulate an enhancement of rainfall during the monsoon season.

Overall, with only a few exceptions, the CMIP3 climate models capture the major climatic features of the monsoon, including the seasonal reversal of the surface winds and the dominance of summer rainfall over winter rainfall. Of the four models without wind reversal, one, IPSL-CM4, has a reasonable rainfall seasonality, and the wind reversal failure is marginal. Consequently, three models (GISS-AOM, GISS-ER and MIROC3.2(medres)) do not adequately simulate the West Pacific Monsoon wind reversal and rainfall seasonality.

5.2.3.6 Summary: Climate Features and Patterns of Variability

Most of the 24 CMIP3 global climate models are able to reproduce the major climate features (SPCZ, ITCZ and West Pacific Monsoon) and modes of variability (ENSO, Interdecadal Pacific Oscillation and Pacific Decadal Oscillation). Two models (GISS-AOM and GISS-ER) show virtually no ENSO-like variability. These two models also fail to simulate a distinct SPCZ and ITCZ. In addition, MIROC3.2(medres) and MIROC3.2(hires) simulate an SPCZ that does not extend to the east of the International Date Line. Three models fail to simulate the West Pacific Monsoon wind reversal and rainfall seasonality (GISS-AOM, GISS-ER and MIROC3.2(medres)).

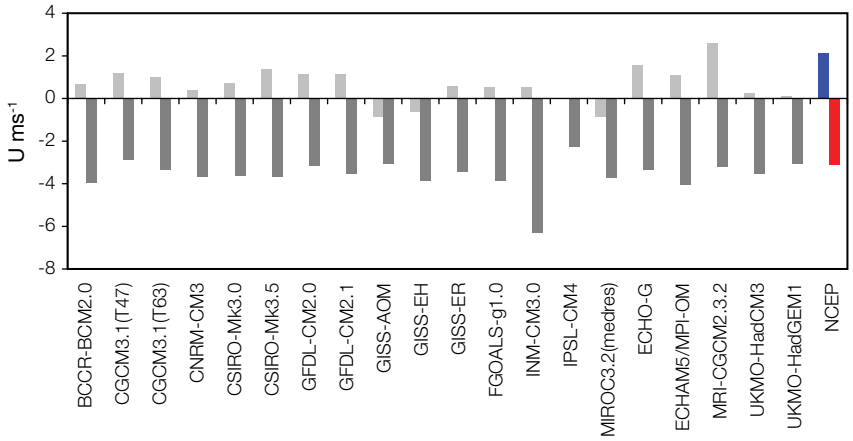

Figure 5.17: Seasonal surface zonal (easterly) winds as simulated by CMIP3 models for the Southern Hemisphere box-1 region (Figure 2.10) in December-February (light grey) and June-August (dark grey). Observed (NCEP, Table 2.2) zonal winds for December-February (blue) and June-August (red). Note that surface winds are only available for 20 CMIP3 models.

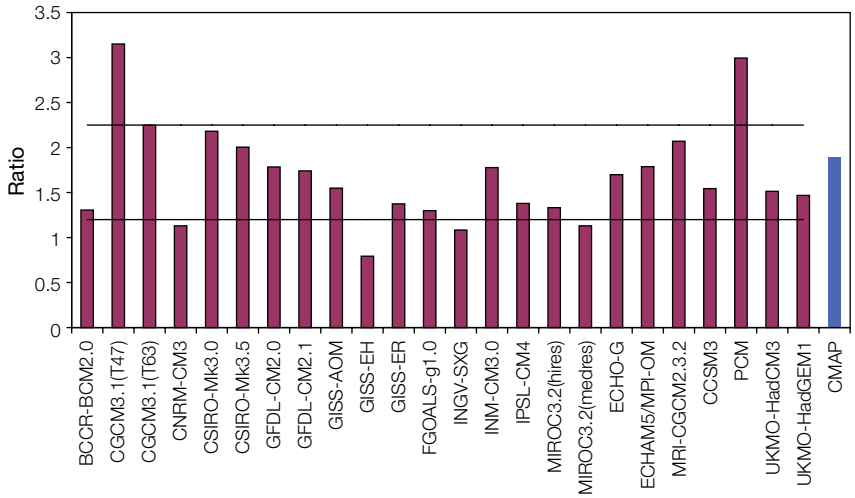

Figure 5.18: The ratio of December-February to June-August rainfall as simulated by CMIP3 models for the Southern Hemisphere box-1 region (Figure 2.10) (purple). The observed rainfall ratio (CMAP, Table 2.2) is indicated in blue. The black lines encompass the multi-model mean plus and minus one standard deviation.

5.2.4 Extremes

As a first step in assessing extreme events, an overview of the ability of the models to simulate the full spectrum of daily maximum temperature, minimum temperature and rainfall over the period 1980–1999 was obtained by calculating the Probability Density Function (PDF) overlap skill score (Perkins et al., 2007; Figure 5.19). In the context of daily maximum surface air temperature, for example, the PDF overlap skill score determines the degree to which a model simulates the correct number of cold, mild and hot days. Scores can range from 0.0 (no skill) to 1.0 (highest possible skill). The test is performed on the empirical distributions of the observed and model data, rather than a parametric distribution fitted to each dataset.

While the PDF overlap skill score is useful in gaining an overview of the ability of climate models to simulate the full spectrum of daily climate, it is not very useful for capturing rare and extreme events (e.g. the hottest day of the year has a relative frequency of 0.003, and is therefore barely visible on a PDF of daily maximum temperature). For this reason, the model bias was calculated for a number of commonly used statistics that capture the magnitude and frequency of extreme daily maximum temperature, minimum temperature and rainfall events (Table 5.3).

Unless otherwise stated, the results presented for all bias statistics refer to the multi-model mean (plus/minus the inter-model standard deviation) of the all observation test scores (the value obtained by averaging across the scores from the ERA-40, NCEP/DOE R-2 and JRA25 datasets). Only limited amounts of daily data are available for CMAP and GPCP, so these datasets could not be used for the rainfall analysis. A list of the CMIP3 models for which daily data are available is given in Appendix 1. The only exception to this listing is for the rainfall statistics, where GISS-ER and GISS-EH were excluded due to their outlying poor performance.

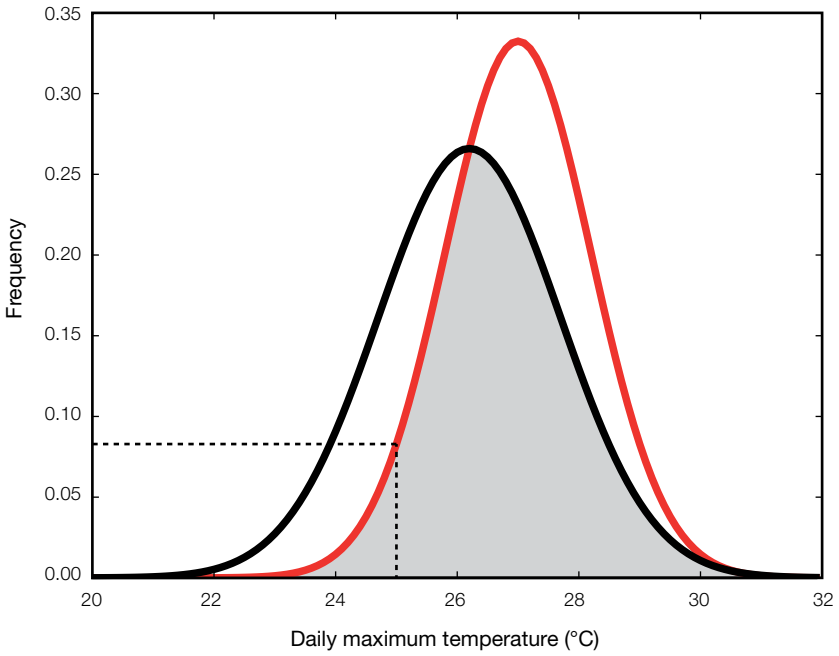

Figure 5.19: Schematic diagram illustrating the information contained in a Probability Density Function (PDF) and the PDF overlap skill score. A PDF provides information on the frequency of an event in a given sample. For example, for the red PDF a daily maximum temperature of 25°C occurs on 8% of days (a relative frequency of 0.08). As indicated by the grey shading, the PDF overlap skill score measures the degree of commonality (or overlap) between modelled (e.g. red PDF) and observed (e.g. black PDF) daily data.

Table 5.3: Definition of the statistics used to assess the ability of the models to capture the magnitude and frequency of extreme events.

Test	Description	Variables
1-in-20-year return value bias (B_{RV20})[a]	Difference (model minus observed) in the magnitude of the event that occurs on average only once every 20 years	Maximum temperature Minimum temperature Rainfall
90th percentile bias (B_{90pctl})	Difference in the value that is exceeded on only 10% of days	Maximum temperature Minimum temperature Rainfall
10th percentile bias (B_{10pctl})	Difference in the value that is exceeded on all but 10% of days	Maximum temperature Minimum temperature Rainfall
Heat wave duration index bias (B_{HWDI})[b,c]	Difference in the average length of a heat wave (defined as five or more consecutive days with a maximum temperature above ($\mu + 2\sigma$)°C of the observational data)	Maximum temperature
Warm nights (WN)[b]	Percentage of days where the minimum temperature is ≥ 90th percentile of the observational data (i.e. a perfect score is 10%)	Minimum temperature
Highest 5-day rainfall total bias (B_{RX5})[b]	Difference in the magnitude of the highest 5-day rainfall total	Rainfall
Extreme rainfall contribution index bias (B_{R95pT})[b]	Difference in the percentage of total annual rainfall that comes from intense rainfall events (defined as more intense than the 95th percentile)	Rainfall

μ = mean; σ = standard deviation

[a]Calculated using the Generalised Extreme Value distribution (Coles, 2001; Kharin et al., 2005)

[b]See http://cccma.seos.uvic.ca/ETCCDMI/ for details

[c]Threshold was modified from the original ETCCDMI definition, to better reflect the Indo-Pacific climate.

5.2.4.1 Surface Air Temperature

The ability of the CMIP3 models to reproduce the full spectrum of daily maximum and daily minimum surface air temperature is very similar. Both variables are particularly well simulated in the South Pacific, with ensemble average PDF overlap skill scores of up to 0.8 achieved in the vicinity of Vanuatu, Fiji and Tonga (Figure 5.20, inset). However, the equatorial Pacific is less well simulated, with scores as low as 0.3 near Kiribati and Nauru. These low equatorial scores can be explained by a general underestimation of the mean and overestimation of the variance in daily temperature (Figure 5.20) which may be related to the aforementioned cold bias and tendency of the models to overestimate the amplitude of the seasonal temperature cycle (Section 5.2.1.1).

With respect to the magnitude of rare and extreme temperature events, the 1-in-20-year maximum daily temperature tends to be slightly underestimated by the models throughout most of the PCCSP region (B_{RV20} = -1.0 ± 0.8°C), while the 1-in-20-year minimum daily temperature is either slightly over- or underestimated, depending on the location and particular model. The models tend to overestimate the maximum heat wave duration throughout the PCCSP region (B_{HWDI} = 6.4 ± 4.7 days) and the number of warm nights (WN = 17 ± 12%).

5.2.4.2 Rainfall

The CMIP3 models have a similar ability in simulating the full spectrum of daily rainfall throughout the PCCSP region, with multi-model average spatial PDF overlap skill scores all within the narrow range of 0.6 to 0.8 (Figure 5.21, inset). The nature of the differences between the model and observed PDF, however, differs depending on location (Figure 5.21). Throughout most of the PCCSP region, the models generally underestimate the number of days at both ends of the observed daily rainfall distribution (i.e. the number of days of both heavy rain and little to no rain). In contrast, the models tend to overestimate the number of days of little or no rain in the vicinity of Kiribati and Nauru. This contrasting performance may be related to the influence of the cold tongue bias (Section 5.2.2.1 and Box 5.2).

Consistent with this underestimation of the frequency of extreme rainfall events, the CMIP3 models also generally underestimate the intensity. This is evident from a tendency to underestimate the intensity of the 1-in-20-year event (B_{20RV} = -54 ± 28 mm per day), the contribution of heavy rainfall events to the annual rainfall total (B_{R95pT} = -3.1 ± 2.2%) and the highest 5-day rainfall total (B_{RX5} = -109 ± 67 mm). These model deficiencies in representing the frequency and intensity of extreme rainfall events may be partly explained by the failure of coarse resolution climate models to fully represent intense, localised rainfall features.

5.2.4.3 Summary: Extremes

In summary, although the CMIP3 models show some skill in representing the broad PDF of temperature and rainfall (as evidenced by the overlap statistic) they tend to underestimate the frequency and intensity of present day extreme (high) temperature and rainfall events. This finding may be related in part to the relatively coarse resolution of these models. Biases tend to be most pronounced in the vicinity of Kiribati and Nauru, which is likely due to the influence of the cold tongue bias. The PCM model provides a relatively poor simulation of present day temperature extremes, while the GISS-ER and GISS-EH models provide a poor simulation of present day rainfall extremes (see Perkins, in press for details). It is recommended that these models are not used for projections of future climate in Chapters 6 and 7 (Section 5.5).

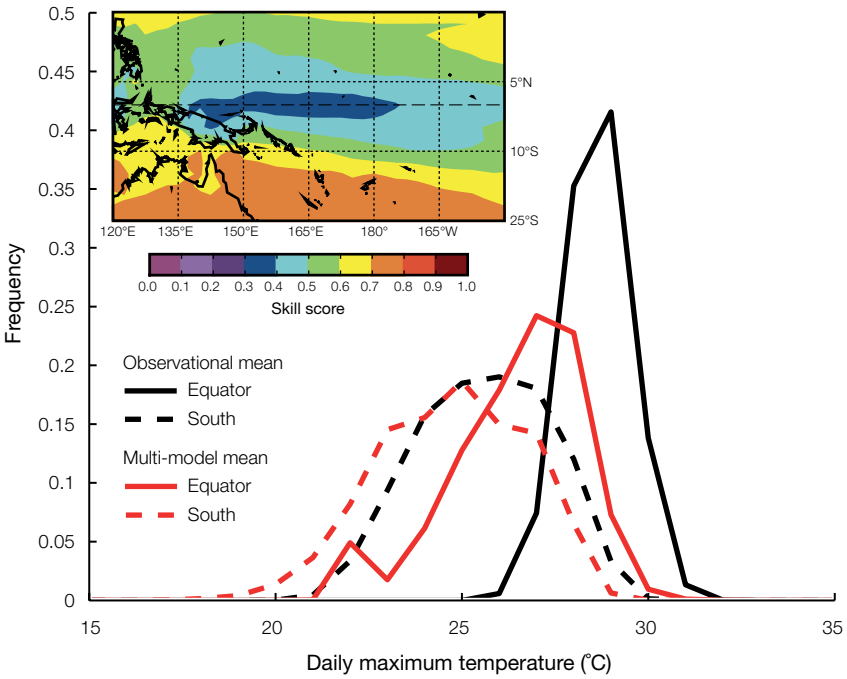

Figure 5.20: Multi-model mean and observed mean (ERA-40, NCEP/DOE R-2 and JRA25 average) 1980–1999 maximum daily surface air temperature PDFs for two locations: south (170°W, 20°S) and equator (175°E, 0°). Inset: Multi-model mean PDF overlap skill score for the 1980–1999 daily maximum surface air temperature, averaged over the observational datasets.

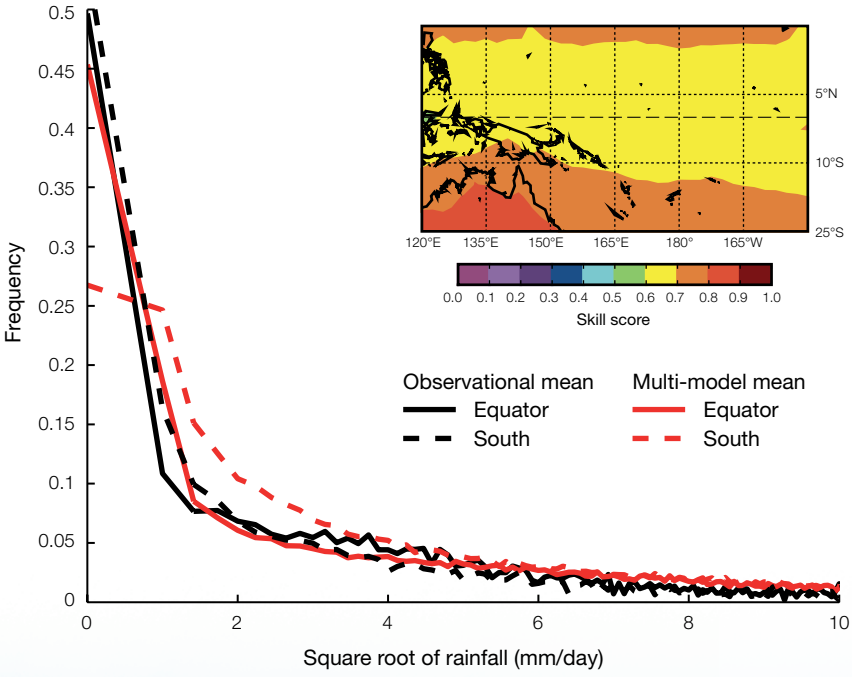

Figure 5.21: Multi-model mean (GISS-EH and GISS-ER excluded) and observed mean (ERA-40, NCEP/DOE R-2 and JRA25) 1980–1999 daily rainfall PDFs for two locations: south (170°W, 20°S) and equator (175°E, 0°). Inset: Multi-model mean PDF overlap skill score for the 1980–1999 daily rainfall, averaged over the observational datasets.

Box 5.2: CMIP3 Model Biases in the PCCSP Region

In general, the CMIP3 global climate models are able to represent essential aspects of the most important large-scale climate features of the Pacific region. These include representing the geographic and temporal patterns of sea-surface temperature and wind speed and direction, and the approximate shape, location and seasonality of the major convergence regions, e.g. SPCZ, ITCZ, and their associated rainfall. This provides confidence in the use of models for regional climate projections.

However, a number of common model biases and errors are apparent which lead to important limits in this confidence. Perhaps the most significant of these arise from a tendency for models to extend the Pacific equatorial cold tongue too far to the west, resulting in western and central equatorial Pacific sea-surface temperatures which are too cold (Figure 5.4). Maximum temperature biases occur in the vicinity of Nauru and Kiribati, however other countries are also affected, as this bias affects large-scale patterns of wind and rainfall, and thereby many aspects of broader Pacific climate and climate variability.

Both model sea-surface temperatures and surface air temperatures are lower than observed in this region, with resultant rainfall totals significantly too low, particularly near the equator. This manifests as an artificial split between the ITCZ and SPCZ, the latter of which is too east-west orientated (Figure 5.15). The ITCZ is also located too far off the equator in December-February. In addition, winds are consistently too strong here, with direction errors (at over 10°) being largest in the Pacific basin (Figure 5.3). The cold tongue bias also affects ENSO variability, in particular pushing the pattern of response too far to the west, with consequences for the interannual variation of surface temperatures and rainfall at low latitudes in the western Pacific (Figure 5.12).

It is critical that such biases and shortcomings are borne in mind when interpreting model output for practical applications within the region. They also provide some of the most significant challenges for ongoing research (Chapter 8).

Flux adjustment in global climate models may reduce some of these biases, but leads to uncertainty over their ability to simulate climate variability and the response to climate forcing.

5.3 Evaluation of Dynamical and Statistical Downscaling

In evaluating the dynamical downscaling performed with the CCAM model (Section 4.5.2), a number of comparisons have been made:

- A comparison of the large-scale PCCSP region present day climate simulated by CCAM (i.e. the 60 km resolution simulations) with observations and the six host global climate models (Section 5.3.1.1). Note that the six models chosen for downscaling here were not the top six performing members in the previous global climate model evaluation. They were selected based upon the research of Smith and Chandler (2010) because they generally scored high in this study and tended to have good interannual variability, such as ENSO. However, as the sea-surface temperature biases were corrected before being used by CCAM, and no atmospheric data from the global model was used, the performance of the global climate model does not directly relate to the accuracy of the downscaled simulations.
- A comparison of the small-scale present day climate simulated by CCAM (i.e. the 8 km resolution simulations) with relevant high resolution observations (Section 5.3.1.2). This analysis was conducted for two regions: Papua New Guinea and Fiji.
- A comparison of the performance of CCAM with additional dynamical downscaling simulations (Section 5.3.3).

5.3.1 CCAM Large-Scale (60 km) Climate

Using the same statistics and observational datasets as in Section 5.2.1, the CCAM 60 km simulations were upscaled to a common 2.5° latitude/longitude grid. This allowed for large-scale temperature, rainfall and extreme weather comparisons to be made against observations and the six host global climate models.

5.3.1.1 Surface Air Temperature

The evaluation of present day surface air temperature (Table 5.4) shows general improvement in the downscaled simulations versus the host global models. In particular, the downscaled simulations provide an:

- Improved mean state (E_{abs} = 0.55 vs. 1.03°C).
- Improved phase (r_t = 0.93 vs. 0.77) and amplitude ($\sigma_{ratio,t}$ = 1.07 vs. 1.21) of the seasonal cycle.
- Improved location (r_p = 0.97 vs. 0.90) and amplitude ($\sigma_{ratio,x}$ = 1.05 vs. 1.17) of spatial features.

The absence of any pronounced cold tongue bias appears to be the main factor behind improvements in the mean state and spatial pattern (Figure 5.22). This is due to the sea-surface temperature bias adjustment in the CCAM method. Also note the improved representation of the temperature over the land masses due to better resolved topography in the CCAM 60 km simulations, as for

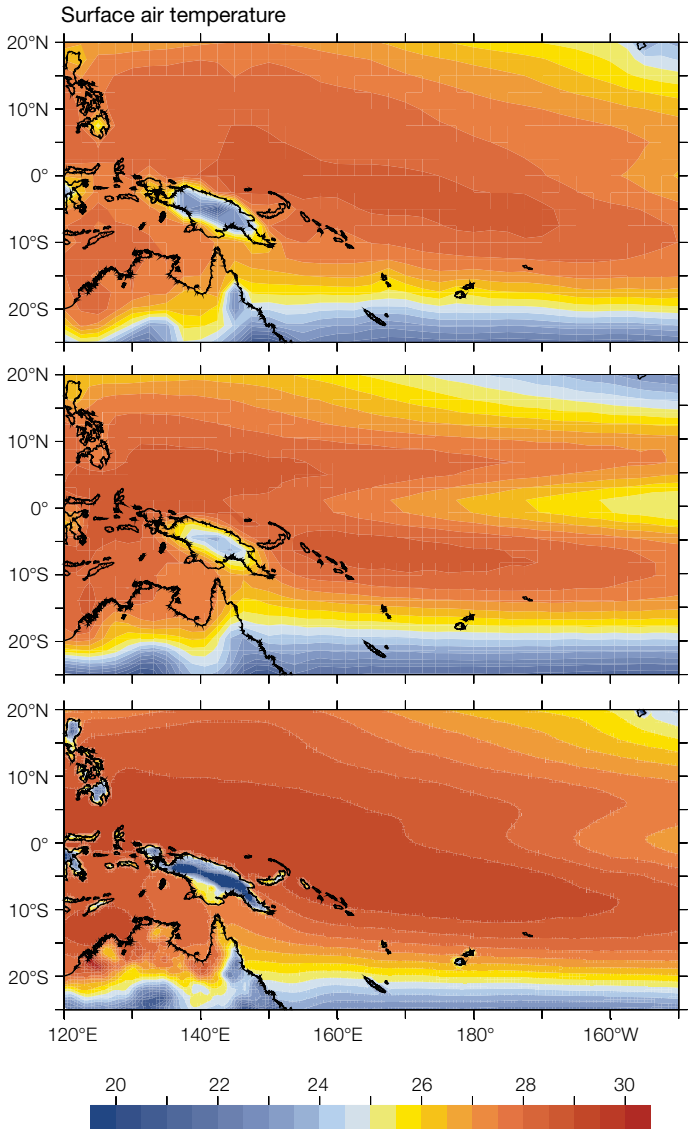

Figure 5.22: Annual mean surface air temperature for 1980–1999 ERA-40 reanalysis dataset (top), for the global climate model multi-model mean (middle) and CCAM 60 km six-model mean (bottom).

Papua New Guinea and Fiji, for example. The warm bias in the CCAM simulations is the result of a parameterisation of the sea-surface temperature enhancement when there are light winds and sunny skies. It appears as though this effect may be too large in these simulations. This can be seen in the evaluation statistics of the individual downscaled simulations (Table 5.4).

5.3.1.2 Rainfall

The evaluation of recent historical rainfall (Table 5.5) shows general improvement in the downscaled simulations versus the host global climate model. In particular, the downscaled simulations provide:

- Improved mean state (E_{abs} = 1.33 vs. 1.88 mm per day).
- Improved spatial pattern (r_p = 0.76 vs. 0.68).
- Slightly poorer seasonal cycle (r_t = 0.51 vs. 0.59).

It is difficult to comment on the amplitude of the seasonal cycle or the spatial features of the region, due the large differences between the amplitude of CMAP and GPCP (Yin et al., 2004; also see Figure 5.2 and discussion in Section 2.2.1.4). However, it is interesting to note that the outputs of global climate models compared more favourably with CMAP, while the downscaled runs compared more favourably with GPCP. An improved orientation of the SPCZ and the absence of excessive rainfall in the far west of the PCCSP region (i.e. the region influenced by the West Pacific Monsoon) appear to be the main factors behind improvements in the mean state and spatial pattern in CCAM (Figure 5.23). Note that for the CCAM simulations, the statistics for the various simulations are fairly similar, while the statistics for the global climate models vary significantly. Although the biases in the CCAM simulations (around 1 mm per day) are worse than the outputs of global climate models, the mean absolute errors are less. Evidently, the global climate models have both positive and negative errors in various regions which cancel out to give a smaller bias, but larger mean absolute errors.

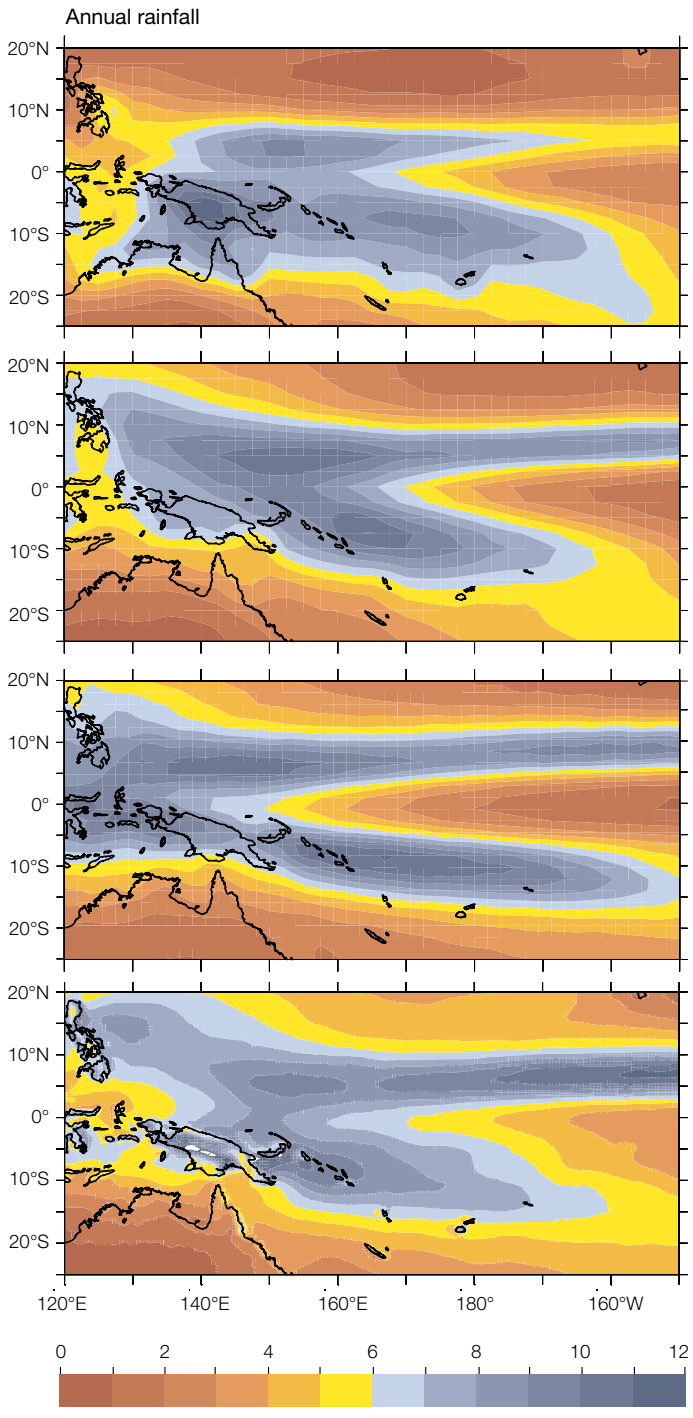

Figure 5.23: 1980–1999 annual mean rainfall (mm per day) for GPCP data (top), CMAP data (middle top), multi-model mean of the six global climate models that were downscaled (middle bottom) and multi-model mean of six CCAM 60 km models (bottom).

Table 5.4: Combined statistics for global climate multi-model mean and CCAM 60 km multi-model mean simulations for surface air temperature against ERA-40, NCEP2 and JRA25 reanalysis. Bold values show better results for multi-model means (top) and also for pairs of global climate model-CCAM simulations (middle and bottom sections). For comparison, statistics for the multi-model means of the 23 global climate models and for the six individual host global climate models and CCAM 60 km simulations are also given (top). Statistics are defined in Table 5.1.

Surface air temperature (entire PCCSP region)	Bias (°C)	E_{abs} (°C)	r_t	r_p	$\sigma_{ratio,t}$	$\sigma_{ratio\,x}$
Multi-model means						
Average (all 23 global climate models)	-0.76	1.02	0.74	0.89	1.18	1.08
Average of 6 global climate models	**-0.40**	1.03	0.77	0.90	1.21	1.17
Average of 6 downscaled models	0.50	**0.55**	**0.93**	**0.97**	**1.07**	**1.05**
Individual global model simulations						
CSIRO-Mk3.5	1.01	1.04	0.80	0.89	1.13	1.18
GFDL-CM2.0	-1.33	1.35	0.77	0.93	1.44	1.28
GFDL-CM2.1	-0.94	0.98	0.73	0.93	1.44	1.22
ECHAM5/MPI-OM	**0.16**	**0.46**	0.81	0.93	**1.05**	**1.04**
MIROC3.2 (medres)	-1.46	1.51	0.71	0.89	1.19	1.04
UKMO-HadCM3	**0.14**	0.86	0.83	0.85	**1.00**	1.25
Individual CCAM simulations						
CCAM (CSIRO-Mk3.5)	**0.54**	**0.58**	**0.94**	**0.97**	**1.06**	**1.06**
CCAM (GFDL-CM2.0)	**0.49**	**0.54**	**0.93**	**0.98**	**1.07**	**1.05**
CCAM (GFDL-CM2.1)	**0.52**	**0.57**	**0.93**	**0.97**	**1.08**	**1.05**
CCAM (ECHAM5/MPI-OM)	0.52	0.56	**0.94**	**0.98**	1.08	1.06
CCAM (MIROC3.2 (medres))	**0.43**	**0.49**	**0.93**	**0.97**	**1.09**	1.04
CCAM (UKMO-HadCM3)	0.53	**0.57**	**0.94**	**0.97**	1.06	**1.05**

Table 5.5: As for Table 5.4, but for rainfall. Observational data is from CMAP and GPCP.

Rainfall (entire PCCSP region)	Bias (mm/day)	E_{abs} (mm/day)	r_t	r_p	$\sigma_{ratio,t}$	$\sigma_{ratio\,x}$
Multi-model means						
Average (all 23 global climate models)	0.39	1.77	0.55	0.63	1.25	1.27
Average of 6 global climate models	**0.62**	1.88	**0.59**	0.68	1.31	1.40
Average of 6 downscaled models	1.00	**1.33**	0.51	**0.76**	**0.92**	**0.90**
Individual global model simulations						
CSIRO-Mk3.5	**0.75**	1.86	**0.66**	0.73	1.37	1.41
GFDL-CM2.0	**0.55**	1.60	**0.67**	0.76	1.49	1.50
GFDL-CM2.1	**1.04**	1.75	**0.62**	0.67	1.40	1.27
ECHAM5/MPI-OM	**0.95**	2.46	0.50	0.57	1.35	1.45
MIROC3.2 (medres)	**-0.13**	1.23	**0.53**	0.70	**1.02**	1.10
UKMO-HadCM3	**0.56**	2.39	0.55	0.67	1.23	1.68
Individual CCAM simulations						
CCAM (CSIRO-Mk3.5)	1.10	**1.37**	0.52	**0.78**	**0.97**	**0.97**
CCAM (GFDL-CM2.0)	0.94	**1.31**	0.54	0.76	**0.91**	**0.85**
CCAM (GFDL-CM2.1)	1.11	**1.44**	0.50	**0.75**	**0.91**	**0.89**
CCAM (ECHAM5/MPI-OM)	1.04	**1.31**	**0.55**	**0.78**	0.93	**0.93**
CCAM (MIROC3.2 (medres)	0.82	**1.20**	0.43	**0.73**	**0.89**	0.88
CCAM (UKMO-HadCM3)	0.97	**1.35**	0.55	**0.76**	**0.92**	**0.87**

5.3.1.3 Extremes

Evaluation of the 1-in-20-year maximum daily air temperature (Figure 5.24) shows large values of over 44°C in the reanalyses over Australia and fairly uniform values of around 30°C over the Pacific Ocean, with slightly lower values in the north-eastern and southern regions. The global climate models capture this pattern, but have higher values than the reanalyses. The upscaled CCAM 60 km simulations have even larger values over the ocean, but are closer to the reanalyses over Australia. The raw CCAM 60 km simulations have slightly higher values than the upscaled results, indicating that upscaling tends to slightly dampen the extremes.

The CCAM 8 km results for Fiji are fairly uniform, with lower temperatures to the south, similar to the CCAM 60 km results. In addition, some variation is seen across the islands, with lower values in the mountains and higher values in the lowlands for the southern island and generally lower values for the northern island. For Papua New Guinea, the 8 km variations are much more pronounced, with lower extremes clearly evident over the high mountains and higher values over the lowlands. This pattern was also partially captured in the CCAM 60 km simulations (not upscaled).

Evaluation of the percentage of days with heavy rainfall (20–50 mm) shows agreement between the mean of the five global climate models for which daily data were available and that were used to downscale, and the mean of the CCAM 60 km simulations (Figure 5.25), while the combined global climate model values appear to be too large, i.e. many of the global models appear to be simulating too many heavy rain days.

The CCAM 8 km percentage of heavy rain days for Fiji and Papua New Guinea (bottom row of Figure 5.25) shows a complex pattern, with higher values (greater than 20% of days with heavy rainfall) along the eastern edges of the Fiji Islands and lower values (around 2% of days with heavy rainfall) over western portions of the islands. Over Papua New Guinea, the largest percentage of heavy rain days is along mountain slopes and over the Solomon Sea, with fewer days over the Coral Sea and most inland areas. The lack of available quality-controlled, high-resolution gridded daily rainfall observational data prevents validation of these values.

The PDF overlap statistic used in the global climate models evaluation (Section 5.2.4) is also applied here to evaluate the extremes in the CCAM 60 km simulations. Evaluation scores in Figure 5.26 are for multi-model means for all CMIP3 models with available daily data (top), the five global climate models with daily data used to drive the CCAM (middle), and the six CCAM 60 km downscaled simulations (bottom). The PDF statistic for daily rainfall shows general agreement for all three, with some improvement in using CCAM compared to the global climate models. For daily maximum temperature, improvement is not evident. This is potentially related to the CCAM simulations having a greater temperature spread (wider PDF) than the reanalyses, giving a smaller PDF statistic score.

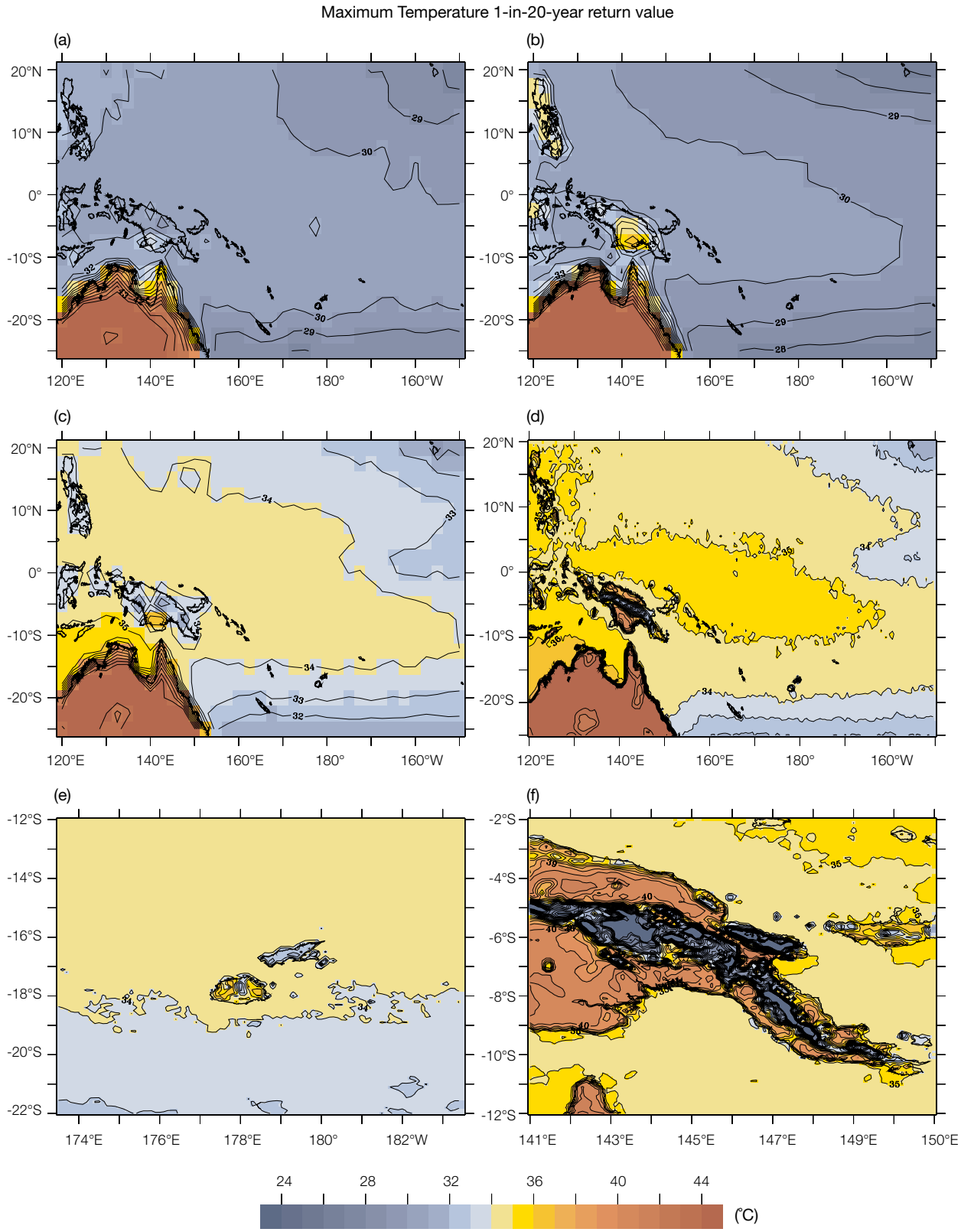

Figure 5.24: 1980–1999 1-in-20-year return maximum daily air temperature. Multi-model mean of (a) three reanalyses (NCEP2, ERA-40 and JMA25), (b) five global climate models, (c) six CCAM 60 km simulations upscaled to a 2.5 degree resolution grid, (d) six CCAM 60 km simulations on a 0.5 degree grid, (e) three CCAM 8 km simulations for Fiji, (f) three CCAM 8 km simulations for Papua New Guinea.

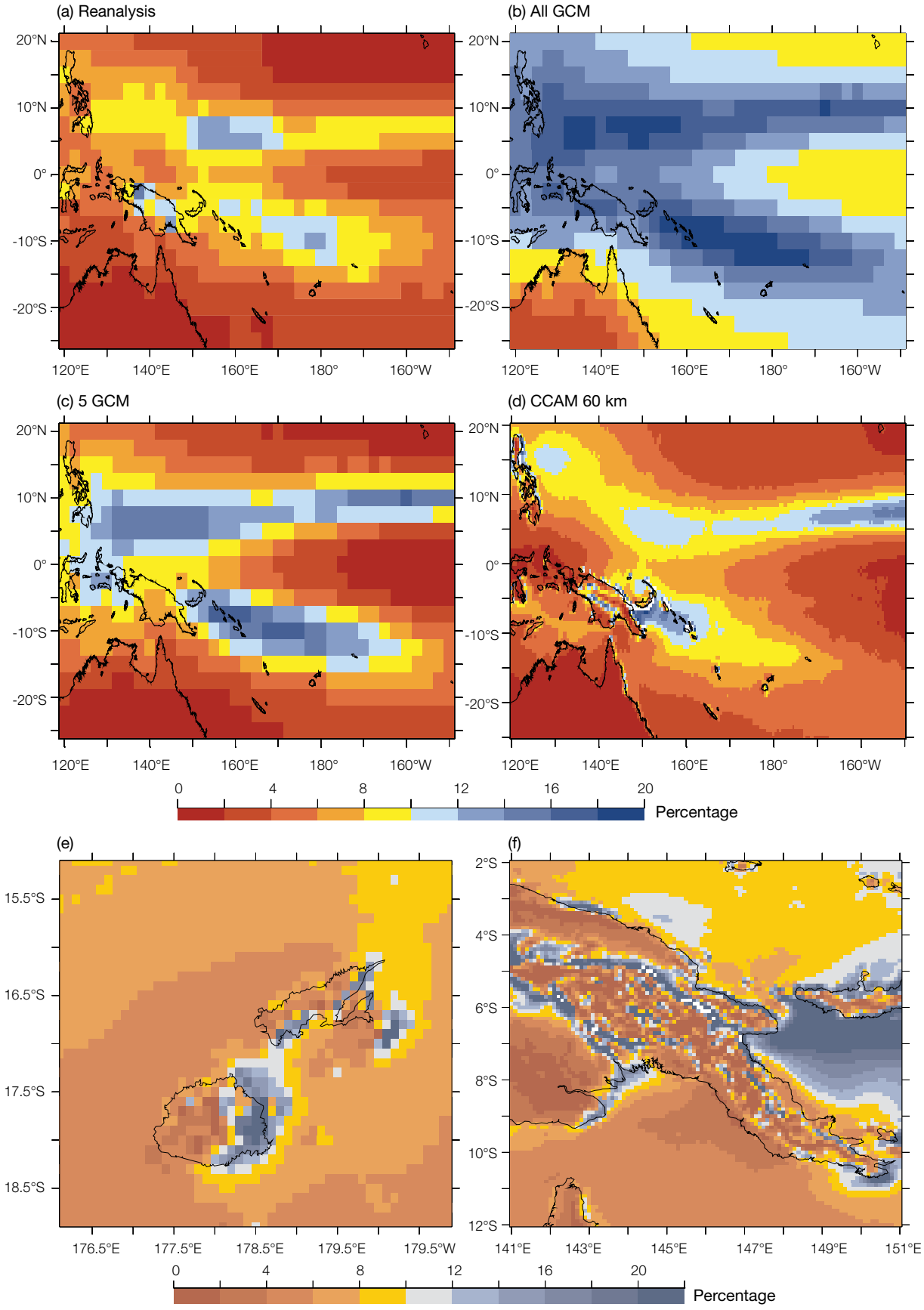

Figure 5.25: 1980–1999 percentage days with heavy rainfall (20–50 mm). Multi-model mean of three reanalyses (NCEP2, ERA40 and JMA25) (top left), average of all global climate models (top right), average of five global climate models (middle left), average of six CCAM 60 km simulations (middle right), average of three CCAM 8 km simulations for Fiji (bottom left) and average of three CCAM 8 km simulations for Papua New Guinea (bottom right).

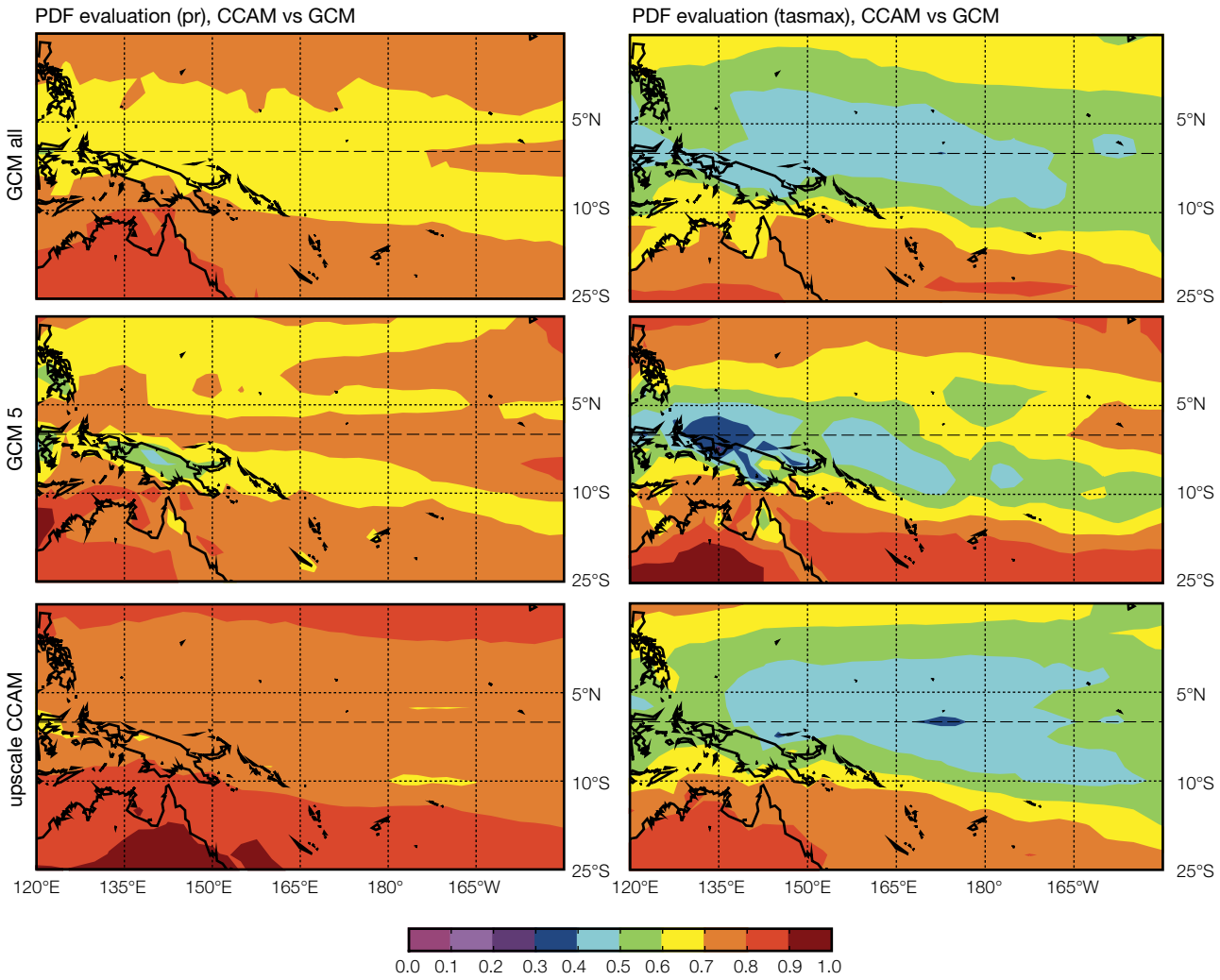

Figure 5.26: PDF statistic for annual rainfall (left) and maximum temperature (right). Multi-model mean for all global climate models (top), for five global climate models (middle), and six CCAM 60 km simulations averaged to 2.5 degree grid for validation (bottom).

5.3.2 CCAM Small-Scale (8 km) Climate

As demonstrated in the previous section, the CCAM 60 km multi-model mean rainfall simulations show a general statistical improvement over the mean values for all global climate models. By further downscaling to 8 km, the rainfall pattern can be better resolved and the effects of topographical and other local features can be incorporated. See, for example, the study by Lal et al. (2008) which showed the benefits of high resolution by using CCAM with 8 km resolution to downscale NCEP reanalyses over Fiji for 10 years. As explained previously, due to computational and time constraints, only three of the six CCAM 60 km simulations were further downscaled to 8 km. In this section, validation for Papua New Guinea and Fiji is presented. A more complete validation for these and other 8 km downscaled regions will be presented in a future technical report.

One of the issues in evaluating rainfall for the 8 km CCAM simulations is the lack of high-resolution gridded datasets to compare against. The TRMM satellite-based dataset chosen is available at 25 km resolution for the period 1998–2010 only, while the CRU dataset is at 50 km and is based upon station data (hence land only (Section 2.2.1)).

Evaluation of the downscaled annual rainfall for Papua New Guinea is presented in Figure 5.27. The top row in Figure 5.27 shows the rainfall pattern at 200 km resolution, approximately the resolution of the global climate models. Although the overall pattern in the various models is similar to the TRMM data, the magnitude is slightly better captured by the 8 km simulation. However, the key reason for the higher resolution simulations is to give more detail than is available from the coarse resolution global climate models. In particular, the pattern of higher rainfall along the slopes of the mountains is better captured by the 60 km simulations, and in even more detail in the 8 km simulations, though the observational dataset is not fine enough to evaluate if all the added detail is realistic.

Comparison of the upper row with the lower row demonstrates the added information at the higher resolutions. The validity of the technique has been shown in a previous study by Lal et al. (2008) using the same model for regional climate simulations over Fiji.

A similar comparison of rainfall for Fiji is shown in Figure 5.28. Fiji is made up of smaller islands than Papua New Guinea, with less complex topography. Again, the top row shows that all models capture the large-scale pattern reasonably well, with more rainfall to the north and less to the south. At the full resolution of the various datasets (bottom row), the topographic effect on the rainfall begins to show, with more rainfall on the eastern side of the main islands and less on the western sides (Lal et al., 2008). This is mainly a result of the easterly trade winds flowing over mountains on the island, rising on the eastern side, resulting in more rainfall, and descending on the downwind side, causing less rainfall.

5.3.2.1 Rainfall and Temperature Validation for Nadi, Fiji

As indicated in the previous section, as finer resolution downscaled simulations are run, more detail in the simulations is provided than is available in gridded observational datasets. An alternative method of validation is to compare model grid point data with station observations. In this section, the simulations from global climate models and CCAM simulations are evaluated for Nadi, Fiji. The observed rainfall climatology from various datasets for 1980–1999 is shown in Figure 5.29 (left). There is greater rainfall in November-April, with less rainfall in the other months. All datasets have generally similar seasonal cycles, though some differences are evident, especially in February, March and December.

Comparisons of rainfall for the global climate models, CCAM 60 km simulations and CCAM 8 km simulations are shown in Figure 5.29 (right). The larger spread of the global climate models (shown by the green dashed lines) is evident, with a tendency to underestimate the seasonal cycle of rainfall. The CCAM 60 km simulations of the seasonal rainfall cycle are closer to the station observations than those of the global climate model. The 8 km simulations tend to best capture the rainfall cycle, though they tend to overestimate rainfall amounts.

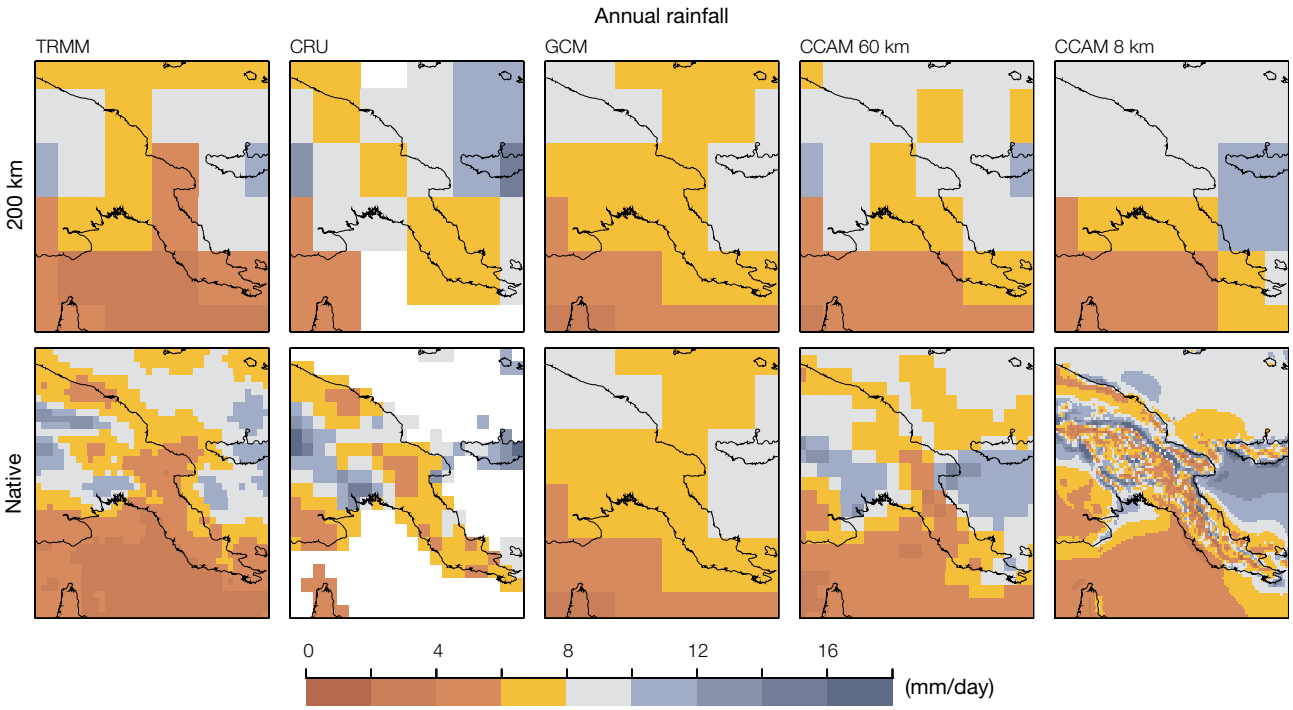

Figure 5.27: Papua New Guinea 1980–1999 annual rainfall climatology (mm per day) for TRMM satellite data (left); CRU data set (middle left); six global climate models multi-model mean (middle); six CCAM 60 km multi-model mean (middle right); and three CCAM 8 km multi-model mean (right). Top row all re-gridded to 200 km grid, bottom row all on the original grid for each dataset.

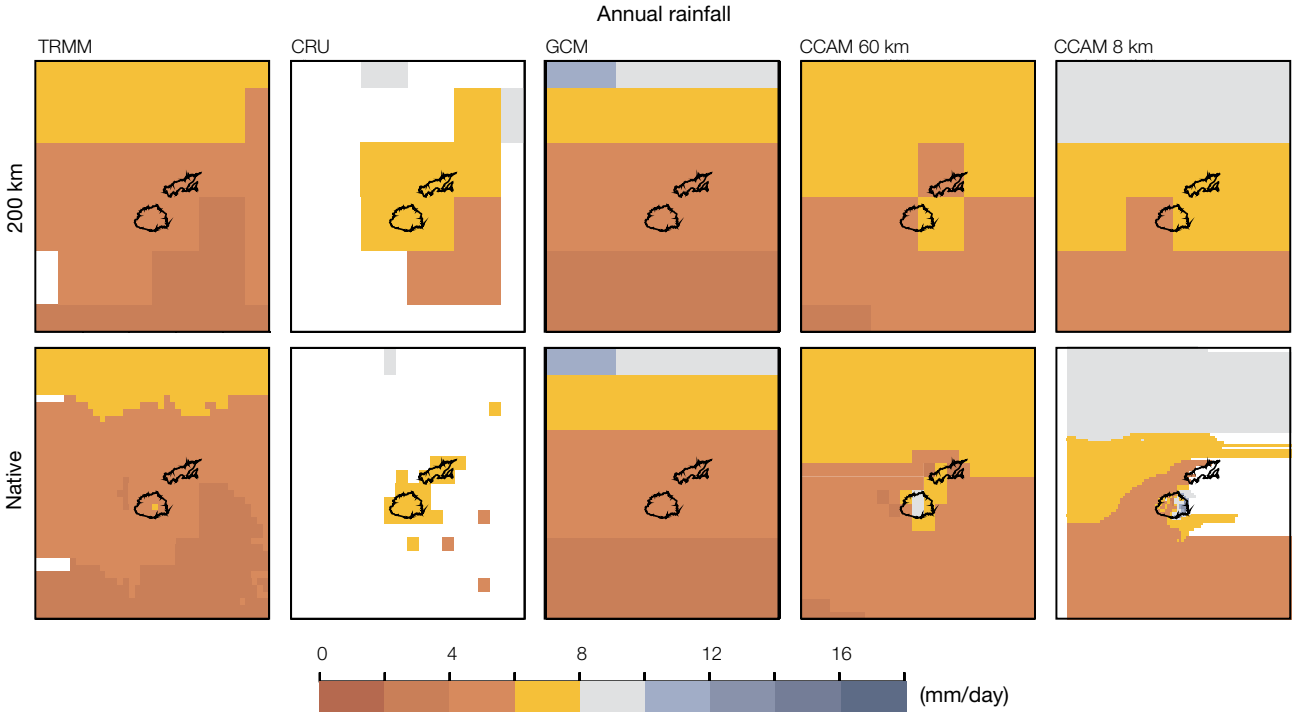

Figure 5.28: Fiji 1980–1999 annual rainfall climatology (mm per day) for TRMM satellite data (left); CRU (middle left); six global climate models model mean (middle); six CCAM 60 km multi-model mean (middle right); and three CCAM 8 km multi-model mean (right). Top row all re-gridded to 200 km grid, bottom row all on the original grid for each dataset.

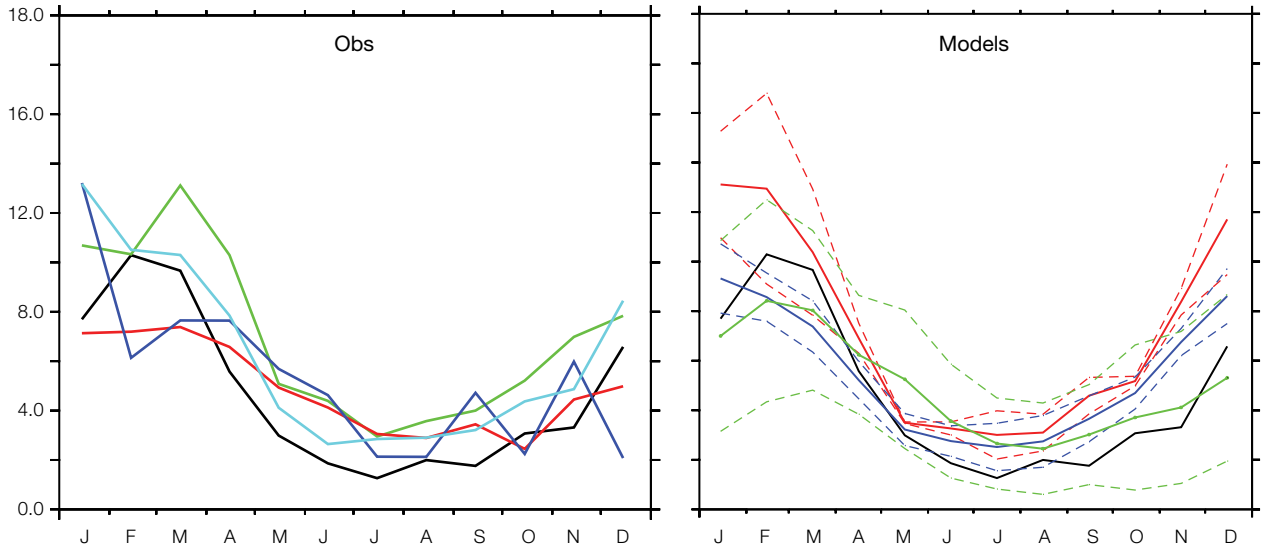

Figure 5.29: Evaluation of 1980–1999 rainfall (mm per day) for Nadi, Fiji. (Left) Monthly plot from various observational data sources: Nadi observations (black line), CMAP (red line), CRU (blue line), GPCP (green line) and TRMM (cyan line). (Right) Monthly plots for Nadi observations (black line), average of the global climate models (solid green line with +/- two standard deviations dashed green lines), CCAM 60 km (solid blue line with +/- two standard deviations dashed blue lines) and CCAM 8 km (solid red line with +/- two standard deviations dashed red lines).

The validation of daily temperature for Nadi, Fiji is presented in Figure 5.30. The multi-model means for the global climate models and the CCAM 60 km simulations show quite good agreement with the station observations, although the global climate models tend to be warmer by around 1°C from February to July. The CCAM 60 km simulations also have a cold bias of about 1°C for most of the year. The CCAM 8 km simulations are closest to the observed seasonal cycle of temperature. The seasonal cycle of maximum and minimum surface air temperature (Figure 5.31) indicates a cold bias in maximum temperatures and warm bias for minimum temperatures in the CCAM 60 km simulations. The CCAM 8 km simulations are very close to the observed maximum temperatures, with a 1°C cold bias in minimum temperatures.

Probability plots of the daily temperature for the various runs are also shown in Figure 5.30 (right). In general, all simulations capture the observed distribution reasonably well, though the global climate models tend to have a flatter peak to the distribution (green line), while the CCAM 60 km (blue line) and 8 km (blue line) simulations capture the peak more accurately, though with a probability slightly too high. In general, the global climate models have too broad a distribution, while the CCAM simulations have too narrow a distribution.

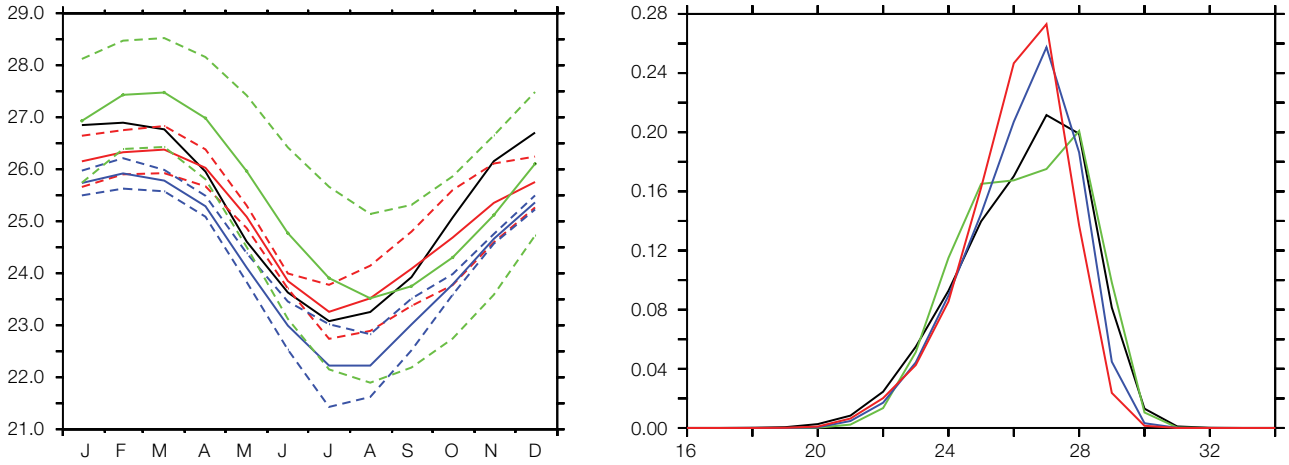

Figure 5.30: Evaluation of 1980–1999 monthly average temperature (mm per day) for Nadi, Fiji. (Left) Monthly plots for Nadi observations (black line), global climate models (solid green line with +/- two standard deviations dashed), CCAM 60 km (solid blue line with +/- two standard deviations dashed) and CCAM 8 km (solid red line with +/- two standard deviations). (Right) Frequency distribution plots for surface air temperature using same colour scheme as in the left hand figure.

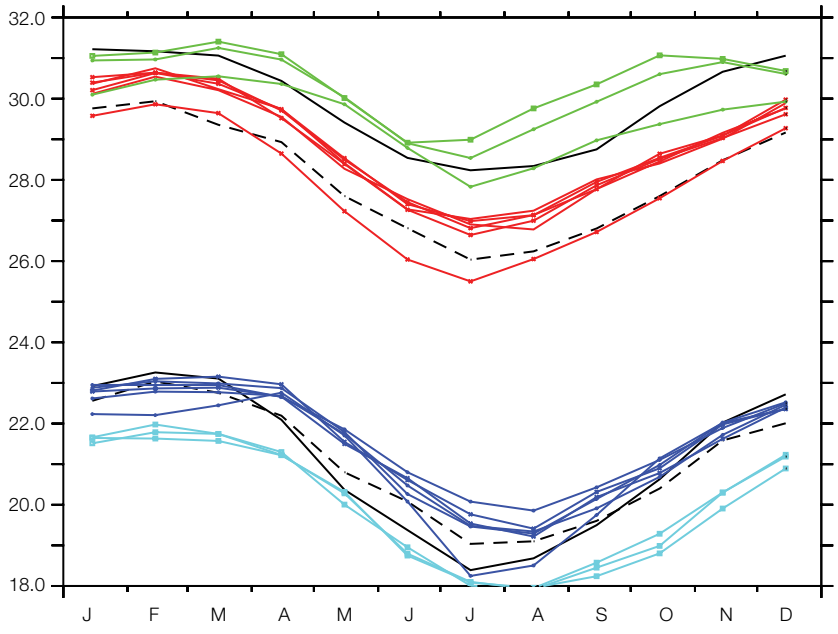

Figure 5.31: Monthly plots of surface air maximum (Tmax) and minimum (Tmin) temperatures (°C) for the grid point representing Nadi, Fiji for the period 1980–1999. Nadi observed Tmax and Tmin are thick black line; CRU observed values at nearest grid point are dashed black line, individual CCAM 60 km simulated Tmax are red and Tmin are blue; while individual CCAM 8 km simulated Tmax are green and Tmin are cyan.

5.3.3 Additional Regional Climate Model Simulations

In this section, validation of the various additional regional climate model simulations which were completed for the PCCSP is summarised briefly. As was noted in Chapter 4, there were three sets of runs completed: one set nested directly within NCEP2 for 1980–2000, and two sets nested within the CCAM 60 km simulations driven by GFDL-CM2.1 GCM sea-surface temperatures, one for the period 1980–2000 and the other for the period 2045–2065 (Figure 4.4). Results for the first two sets of regional climate simulations are presented in this section, while those of the third are presented in Chapter 7.

Comparison of the annual rainfall for the various regional climate models nested in the NCEP2 reanalyses (Figure 5.32) shows a large range of results. A key element of regional climate modelling is the level of impact provided by the lateral boundary conditions. Some of the models used some large-scale forcing (CCAM - top right centre, WRF - middle left, and Zetac - middle right) which provides an additional constraint on the simulation in the interior of the domain. However, this does not seem to have made significant impact on the quality of the simulations. CCAM has too strong an ITCZ, while most other models are too weak, apart from PRECIS, which is the best at capturing the ITCZ. For the southern portion of the domain, all models have difficulty correctly capturing the position and intensity of the SPCZ, with WRF probably the best. The impact on the simulation of using different convection parameterisation schemes can be seen by comparing the four examples for RegCM in the bottom row of Figure 5.32. Significant variation can be seen between the various runs. The way the models treat convection can have a very large impact on the quality of the simulation results, especially in this region.

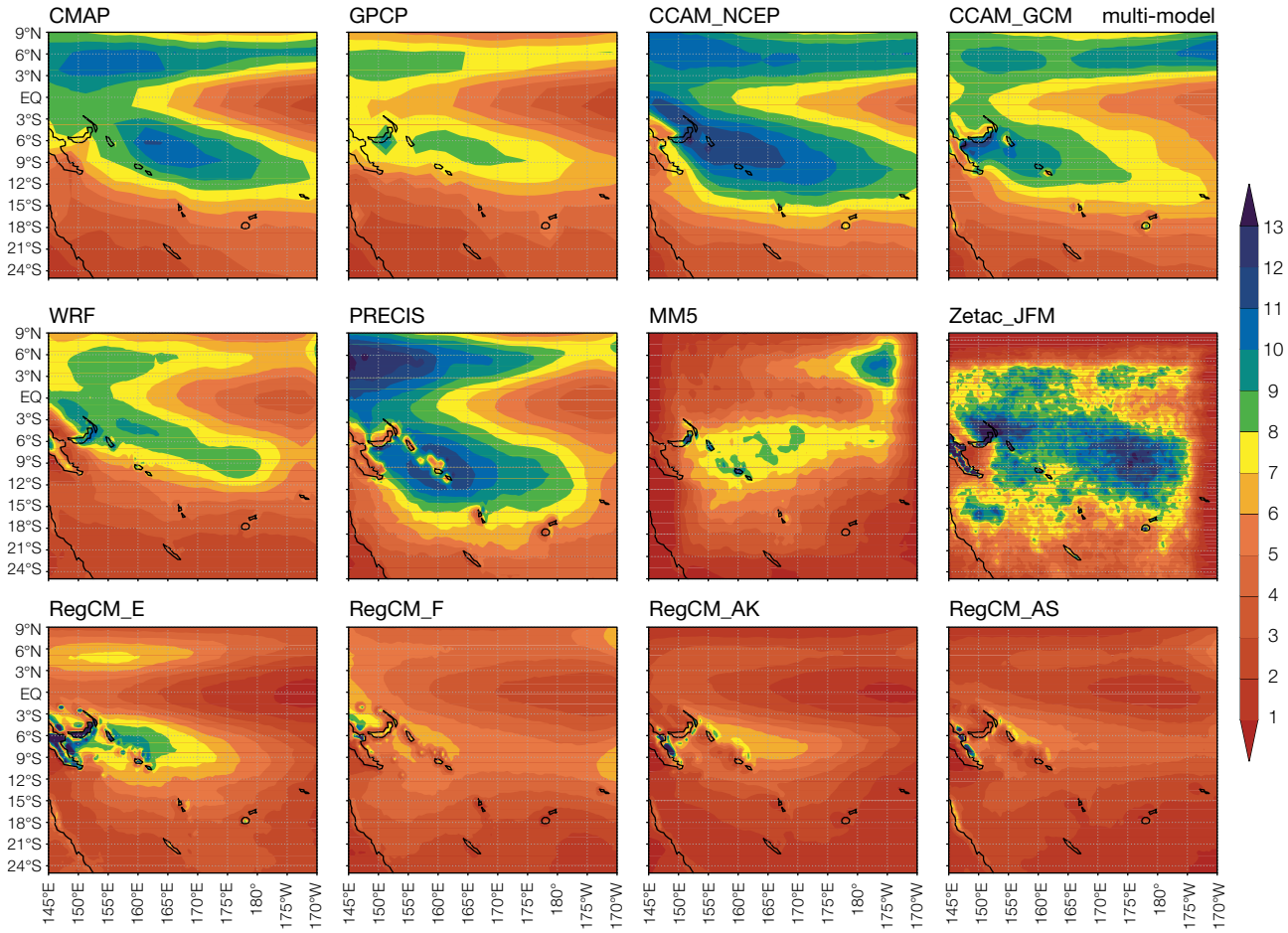

Figure 5.32: Annual rainfall (mm per day) for observed datasets and additional regional climate model simulations nested within the NCEP2 reanalyses. (Top): Observed datasets CMAP (left) and GPCP (centre left). CCAM at 60 km run with spectral nudging of large-scale fields of temperature, winds and surface pressure (centre right) and CCAM 60 km multi-model mean of runs using GCM sea-surface temperatures (right). (Middle): Various regional climate models nested directly within 6-hourly NCEP2 reanalyses. Note that the Zetac model is only for January-March, not annual. (Bottom): RegCM model run with four different convection schemes: Emanuel (left), Grell with Fritsch-Chappell (centre left), Anthes-Kuo (centre right) and Grell with Arakawa-Schubert (right).

Validation statistics, similar to those presented for global climate models and CCAM 60 km simulations (Table 5.5) are presented in Table 5.6 for runs nested within NCEP2 for the period 1980–1999. The statistics are computed over the simulation domain, excluding some boundary rows where the limited-area models have problems (Section 4.5.1). The PRECIS model has the lowest bias and mean absolute error, while CCAM has the lowest root mean square error and correlation, i.e. it captures the pattern best. The results support the discussion of Figure 5.32.

For the second set of climate change simulations, the regional models were nested within a global climate model. Here, the data from the CCAM 60 km simulations with bias-corrected sea-surface temperatures from the GFDLCM2.1 global climate model were used to drive the lateral boundary conditions of the limited-area models. In order to assess the impact of this technique, the validation statistics were recomputed for 1980–2000 from these simulations (Table 5.7). Again, the PRECIS model bias is the lowest, though the CCAM 60 km simulation has the lowest mean absolute error and root mean square error. Interestingly, although there is some decrease in quality of the simulations, the statistics do not show large changes when compared with the runs nested within NCEP. This suggests that the lateral boundary data were not causing significant negative impact on the limited area models' climatology.

In summary, the MM5 regional model did not perform satisfactorily, with large error statistics and poor pattern correlation, and will not be used in the downscaled projections in Chapter 7.

Table 5.6: Validation of annual rainfall for various regional climate models nested within the NCEP2 reanalyses for the period 1980–1999. The verifying dataset is the CMAP dataset. Equivalent comparison of GPCP rainfall analysis with CMAP is shown in last row. Bold numbers show best performance of the models.

Annual rainfall (mm per day) (Nested within NCEP2)	Bias	E_{abs}	RMS error	r_p
CCAM60	1.05	1.13	**1.38**	**0.93**
WRF	-1.14	1.26	1.49	0.92
PRECIS	**0.46**	**1.05**	1.41	0.86
MM5	-2.67	2.95	3.64	0.46
REGCM_E	-2.27	2.58	3.14	0.57
REGCM_FC	-2.58	2.65	3.06	0.75
REGCM_AS	-3.78	3.84	4.30	0.51
REGCM_AK	-3.88	3.93	4.35	0.57
GPCP	-1.10	1.21	1.38	0.96

Table 5.7: Validation of annual rainfall for various regional climate models nested within the CCAM 60 km simulations using GFDL-CM2.1 sea-surface temperatures for the period 1980–1999. Verifying dataset is the CMAP dataset. Bold numbers show best performance of the models.

Annual rainfall (mm per day) (Nested within CCAM 60 km GFDLCM2.1)	Bias	E_{abs}	RMS error	r_p
CCAM60	0.29	**0.91**	**1.30**	**0.85**
WRF	-1.42	1.68	1.90	**0.85**
PRECIS	**0.23**	1.41	1.72	0.84
MM5	1.27	5.52	7.20	0.58
REGCM_E	-0.68	1.60	2.34	0.67

5.3.4 Evaluation of Statistical Downscaling

To evaluate the statistical downscaling methodology used in this project (Section 4.7), an example comparison between the simulated temperature and rainfall data and the observed data at Nadi, Fiji is presented (Figure 5.33).

This figure shows that the statistical downscaling and bootstrap simulation procedure is capable of representing the year on year variation of standard deviations of temperature (top panel), and also is able to accurately represent the short-term autocorrelation structure of the temperature and rainfall data (four subplots on the lower panels). Further validation at this specific location is shown in Figure 5.34. The Linear Mixed Effect State-Space Model (Section 4.6) predicts minimum and maximum temperature and proportion of rain days well, but median rainfall on rain days is less well predicted. This is not unexpected and is similar to experience from dynamic models. Similar plots were obtained for all other locations for which statistical downscaling was performed. These also indicated that temperature was predicted quite accurately by the model and that rainfall was less well predicted. Further out of sample validation results of the Linear Mixed Effect State-Space Model can be found in Kokic et al. (2011).

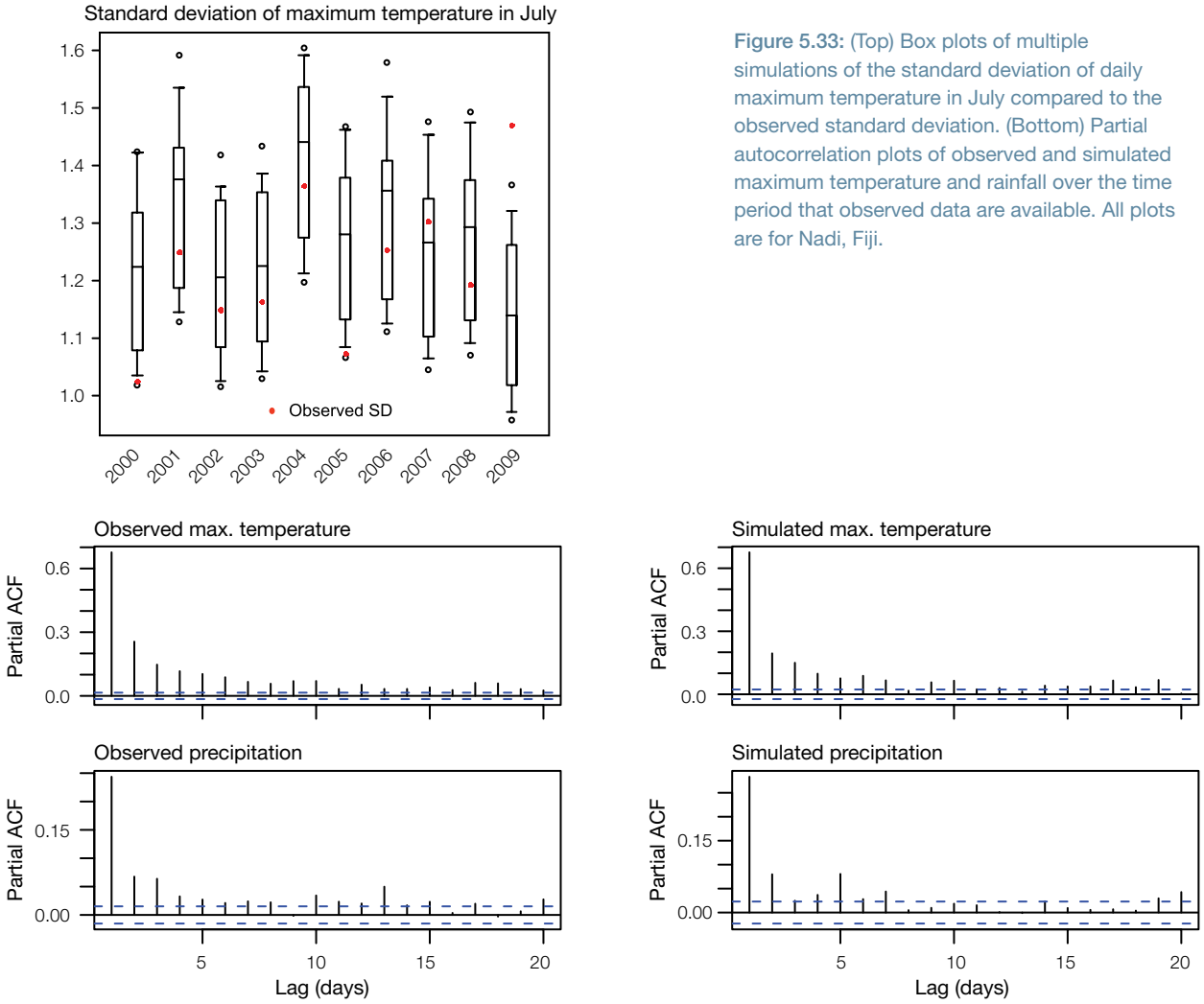

Figure 5.33: (Top) Box plots of multiple simulations of the standard deviation of daily maximum temperature in July compared to the observed standard deviation. (Bottom) Partial autocorrelation plots of observed and simulated maximum temperature and rainfall over the time period that observed data are available. All plots are for Nadi, Fiji.

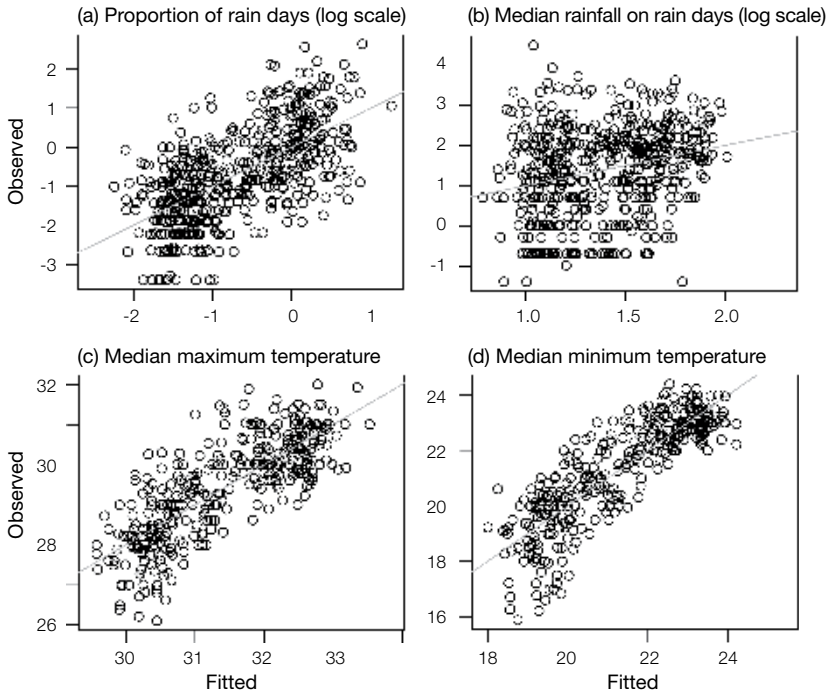

Figure 5.34: Observed monthly data versus predicted values from the Linear Mixed Effect State-Space Model. All plots are for Nadi, Fiji.

5.3.5 Summary: Dynamical and Statistical Downscaling

The dynamically downscaled simulations provide an improved, more detailed representation of most aspects of the current climate. At higher resolutions, the simulations are able to capture more detailed topographic and coastal effects of the islands on the atmospheric flow. One issue is the lack of suitable observational datasets to validate the downscaled simulations. Gridded datasets are of generally coarser resolution than the simulations. Comparing downscaled results against stations is also not totally appropriate, since the model simulations give a grid box average rather than the point value of a station.

The utility of validating the statistical downscaled results is less clear since the model is trained on the observational data. Care must be used not to interpret the accuracy of the statistical downscaled results for the current climate as an indication of the reliability of the technique for future climates.

In summary, although better performance of the downscaled results for most aspects of the current climate does build confidence in their use in projecting future climates, there is still some uncertainty that the downscaled models will perform accurately in a future climate, and the results must be used with a degree of caution, as with global climate model simulations.

5.4 Tropical Cyclones in Global and Downscaled Models

5.4.1 Introduction

It is difficult to make climate change projections of tropical cyclone activity. There are two fundamental reasons for this. The first is that the features of a tropical cyclone that cause damage occur at a spatial scale smaller than can be resolved by the climate models, particularly global climate models. These include the eye wall, the zone of high winds, heavy rainfall, wave action and storm surge. A consequence of this is that the models do not adequately simulate tropical cyclone behaviour when applied to current climate. Conventional wisdom suggests that without developing appropriate scale reduction or downscaling techniques, the models cannot be used to project future behaviour of tropical cyclones with any confidence (Knutson et al., 2010). However, the direct application to climate models of the Curvature Vorticity Parameter (CVP) method (Section 5.4.3.1) and a modified CSIRO Tropical Cyclone detector and tracker (modified using insight developed in this project) has yielded promising results.

The second reason for the difficulty in making projections of tropical cyclone behaviour is that the features of current climate that strongly influence tropical cyclone numbers and intensity have some systematic biases in climate models. These include the regional patterns of sea-surface temperature as well as the major climate features and patterns of variability in the PCCSP region, including ENSO, the ITCZ, West Pacific Monsoon and the SPCZ. The tropical cyclone research carried out for the PCCSP mainly addressed the first of these issues: the horizontal resolution or downscaling problem. The methods used are those described in Section 4.8 of this publication.

5.4.2 Tropical Cyclones in Downscaled Models

The CCAM modelling system used in this study includes bias-corrected sea-surface temperature as a boundary condition inherited from the host global climate model, and thus is expected to have significantly different regional circulation features to the host. Theoretically, the 60 km horizontal resolution enables more realistic cyclone circulations to develop than in the host global climate model. These circulations are detected using the CSIRO Direct Detection (CDD) method described in Section 4.8.

The spatial distribution of tropical cyclone genesis locations, detected from CCAM using the six different host global climate models for the late 20th century (1981–2000) compare well with each other and with the observed climatology derived from the IBTrACS data (Figure 5.35). Overall, the late 20th century cyclone climatology is well reproduced in the Southern Hemisphere. However, it underestimates cyclone frequency in the Northern Hemisphere in the eastern North Pacific and the North Atlantic. The performance is more apparent in Figure 5.36 which shows the observed and downscaled annual cyclone numbers for four of the six models, globally, by hemisphere and in three PCCSP sub-basins.

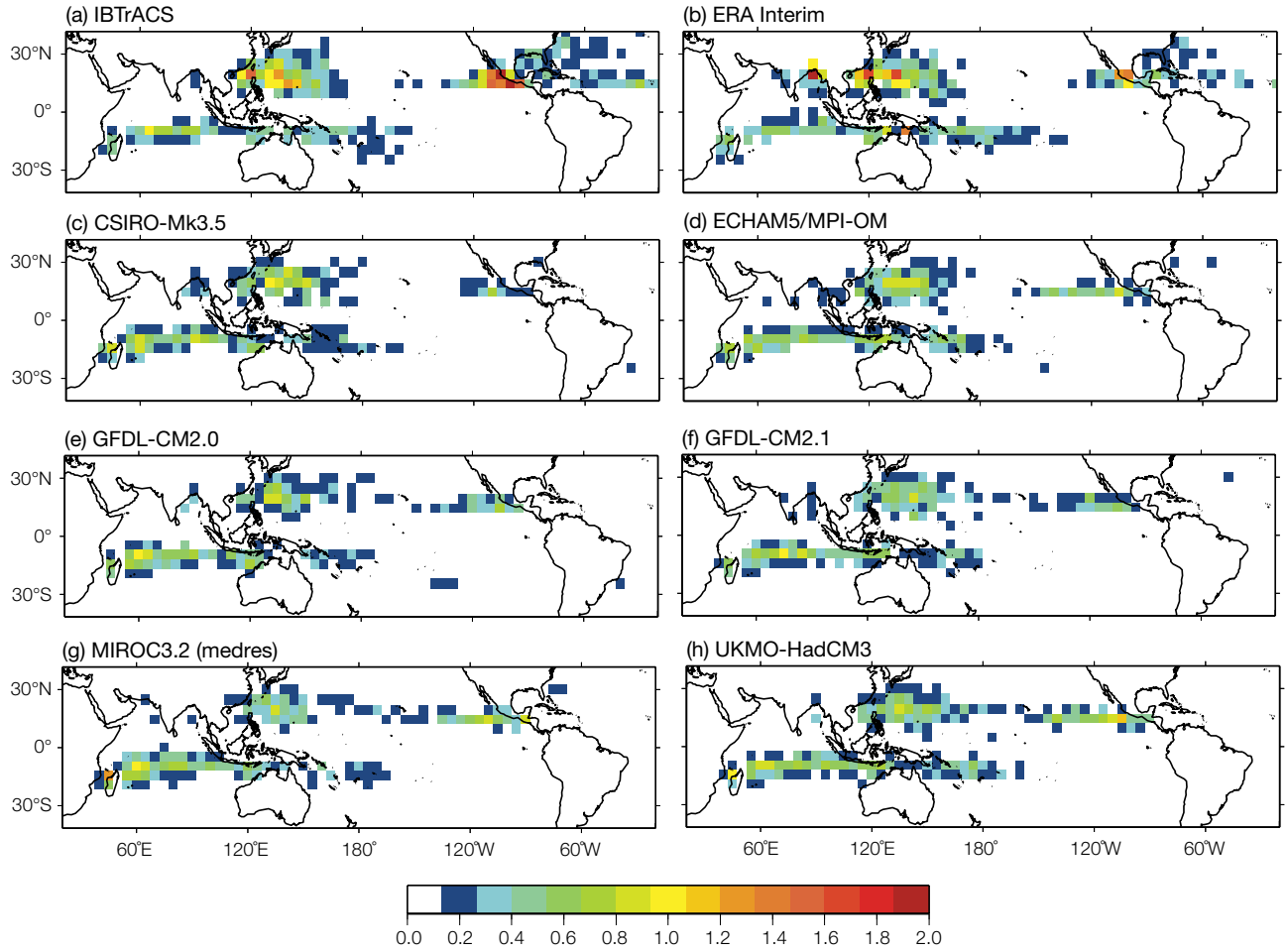

Figure 5.35: Spatial distribution of annual tropical cyclone genesis (a) in IBTrACS data, (b) detected in ERA-Interim using the CSIRO Direct Detection method and (c-h) detected in CCAM simulations for different host global climate models. Occurrence is expressed as the number of cyclone formations per year within a 5 x 5 degree grid cell.

Over the PCCSP region, results show that CCAM is able to reproduce the climatology of tropical cyclone numbers well in the south-west Pacific sub-basin (0–35ºS; 130–170ºE; Figure 5.36d) and the north Pacific sub-basin (0–15ºN; 130º–180ºE; Figure 5.36f). However, the spatial distribution of tropical cyclones detected in the South Pacific shows fewer tropical cyclones east of 180º longitude, which explains the lower number of cyclones detected in the south-east Pacific sub-basin (0–35ºS; 170ºE–130ºW) compared to the observed climatology (Figure 5.36e). One needs to be aware of the greater uncertainty in cyclone frequency projections in this sub-basin due to the weaker performance of the downscaled models, which has implications for country specific projections for the PCCSP Partner Countries that reside in that sub-basin (which include Fiji, Tuvalu, Samoa, Tonga, Niue and Cook Islands, as well as eastern parts of Vanuatu and the Solomon Islands).

5.4.3 Tropical Cyclones in Global Climate Models

Three analysis methods have been applied to the outputs from global climate models. These methods are (1) the Curvature Vorticity Parameter method, (2) the CSIRO Direct Detection method and (3) the Genesis Potential Index. These are described in detail in Section 4.8. Assessment of the late 20th century tropical cyclones in global models based on these methods is discussed in this section, and the results are compared with the observed tropical cyclone climatology.

5.4.3.1 Curvature Vorticity Parameter Method

The performance of the objective Curvature Vorticity Parameter (CVP) detection technique (Section 4.8.3) developed in this study has been assessed and applied directly to global climate models. This approach complements the CCAM-based downscaling method and provides an independent assessment of the ability of global models to generate tropical cyclone-like structures. A comparison of Figure 5.37(a) and (b) shows the detection method reproduces the observed spatial distribution of tropical cyclones and tropical cyclone numbers well. Figure 5.38 also shows that the observed annual tropical cyclone numbers (red bars) and those detected using the CVP approach (green bars) also compare well globally, in the two hemispheres, and in the three PCCSP sub-basins.

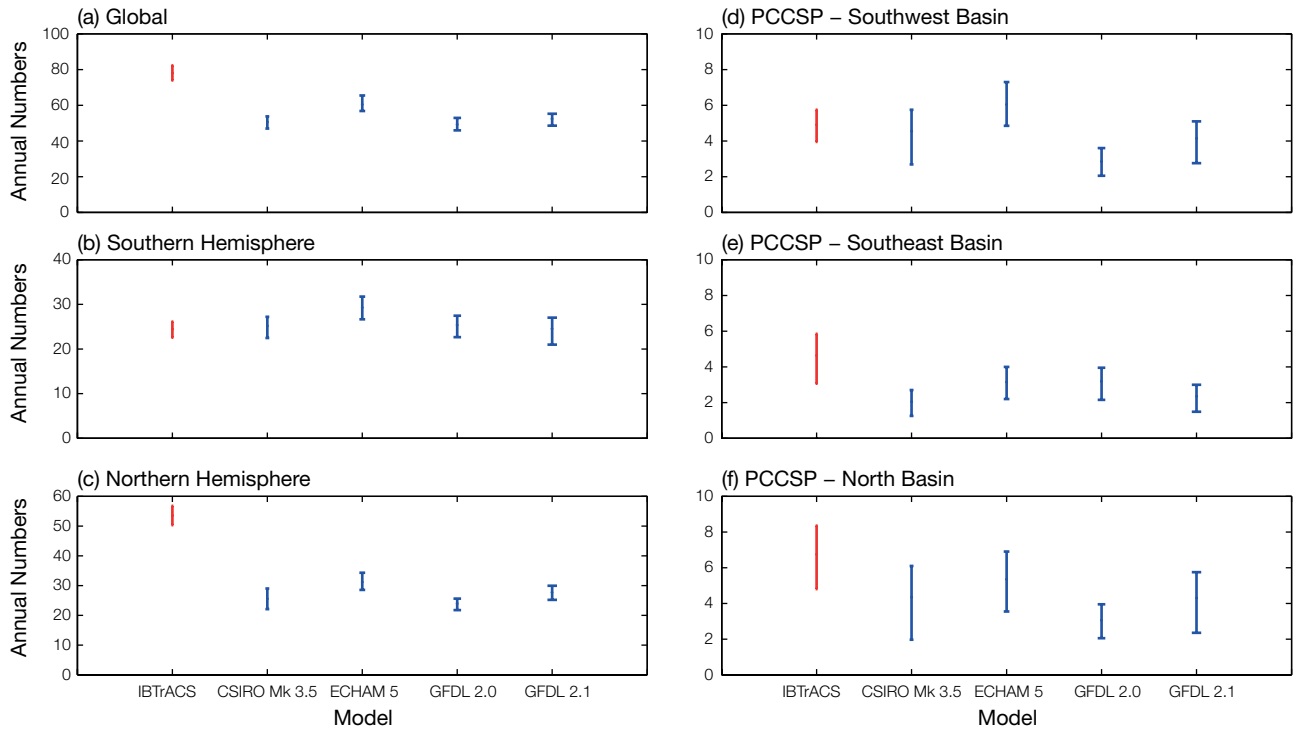

Figure 5.36: Annual number of tropical cyclones simulated using the CCAM downscaling methodology for different host models, and that observed in the IBTrACS data, in different hemispheres (left) and in the PCCSP region (right). Bars indicate distribution in annual mean number of tropical cyclones (at the 95% significance level) obtained via a statistical technique known as bootstrap sampling (i.e. repeat sampling).

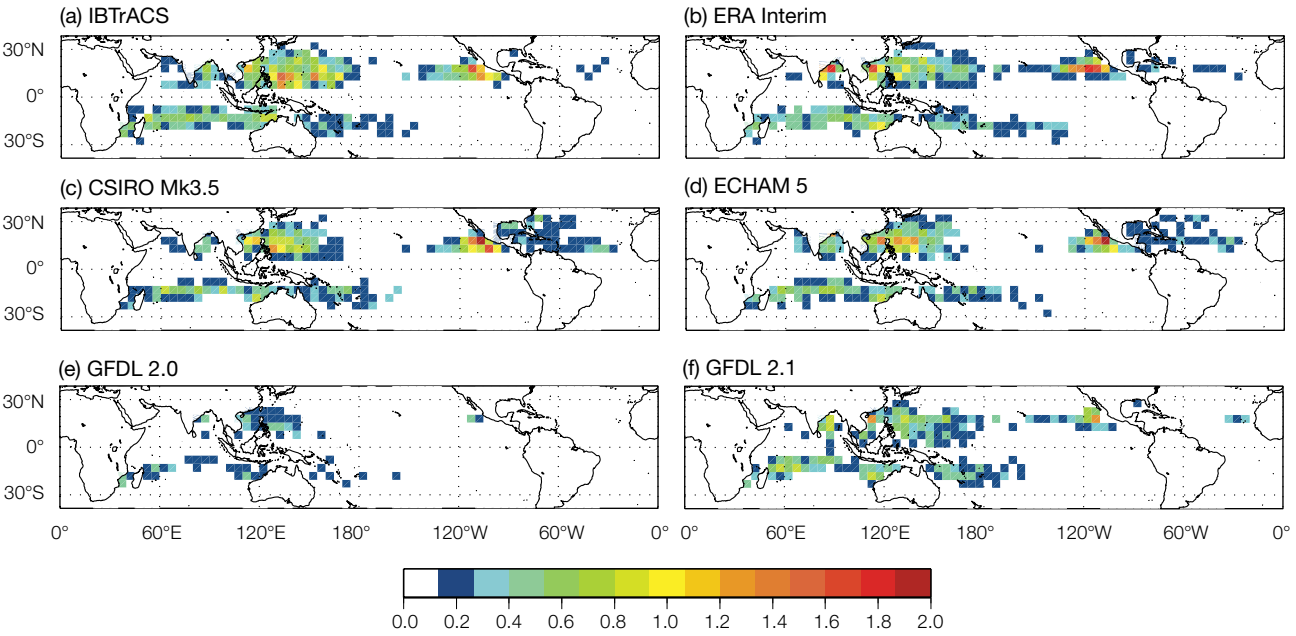

Figure 5.37: Spatial distribution of annual tropical cyclone genesis (a) in IBTrACS data, (b) detected in ERA-Interim and (c-f) detected in different global climate models using the Curvature Vorticity Parameter method. Occurrence is expressed as the number of cyclone formations per year within a 5 x 5 degree grid cell.

The global climate models (excluding GFDL-CM2.0[1]) reproduce the observed spatial distribution of tropical cyclones well (Figure 5.37), although the mean annual tropical cyclone numbers show some variation between the models (blue bars in Figure 5.38). Like the downscaled results, too few cyclones are detected in the North Atlantic. The healthier cyclone numbers in the north-east Pacific compared to the downscaled results (excluding GFDL-CM2.0), contribute to the overall better reproduction of the Northern Hemisphere cyclone climatology (compare Figure 5.36c and Figure 5.38c).

In the PCCSP region using the CVP technique, two global climate models (ECHAM5/MPI-OM and GFDL-CM2.1) reproduce the total number of tropical cyclones very well (i.e. errors of only 5% and 7% respectively, not shown). However, when divided into different basins, differences between the observed and detected tropical cyclone numbers become more apparent (Figure 5.38d-e). For example, ECHAM5/MPI-OM and CSIRO-Mk3.5, in the south-east (Figure 5.38e) and northern sub-basin (Figure 5.38f), respectively, deviate substantially from climatology. The overestimation in ECHAM5/MPI-OM is due to more tropical cyclones detected east of 180° (Figure 5.37d) than those observed.

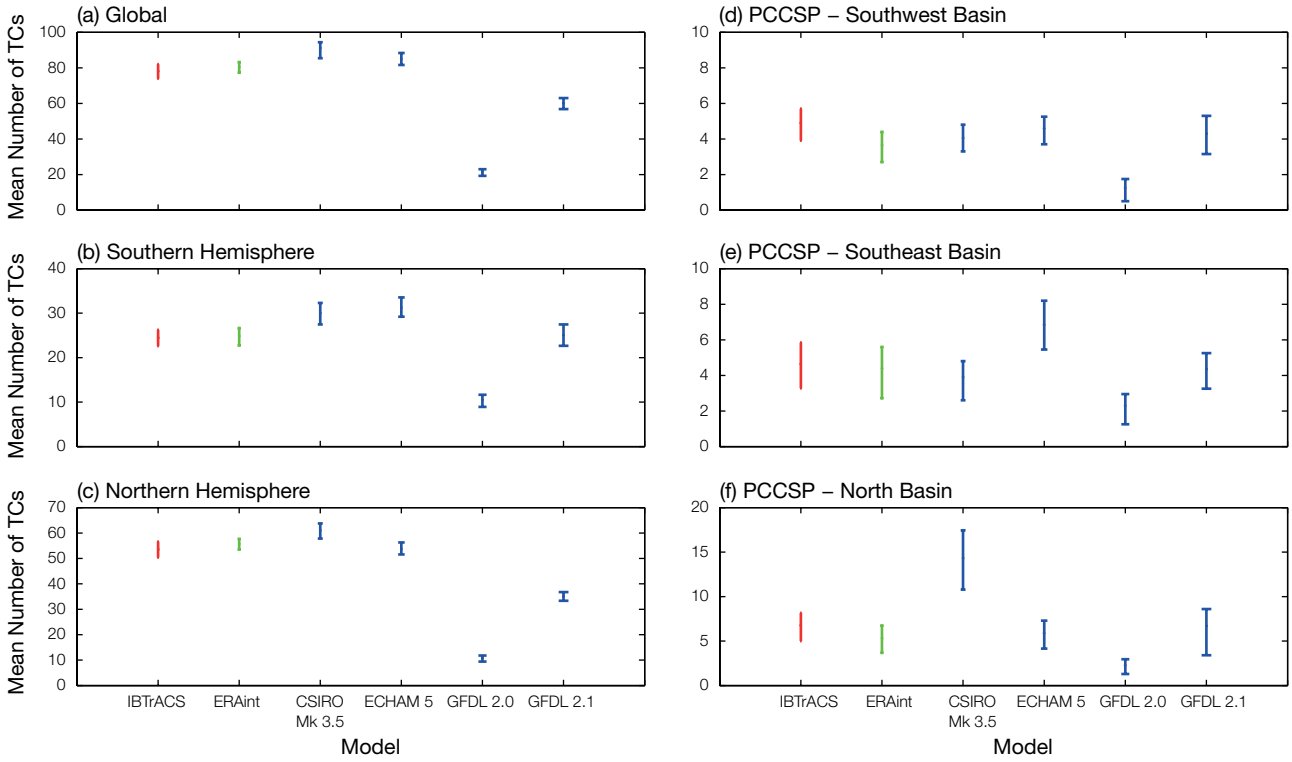

Figure 5.38: Annual number of tropical cyclones detected in different global climate models using the Curvature Vorticity Parameter scheme, and that observed in the IBTrACS data, in different hemispheres (left) and in the PCCSP region (right). Bars indicate distribution in annual mean number of tropical cyclones (at the 95% significance level) obtained via a statistical technique known as bootstrap sampling (i.e. repeat sampling).

[1] The Curvature Vorticity Parameter technique uses humidity as a detection criterion. Due to a dry bias in the GFDL-CM2.0 model many circulations were not detected. The CSIRO Direct Detection technique does not use humidity and is thus not negatively affected by the dry bias.

5.4.3.2 CSIRO Direct Detection Method

The CSIRO Direct Detection (CDD) method (Section 4.8.2) has been applied to nine global climate models for which appropriate data are available (Figure 5.39). The average spatial distribution of tropical cyclones compares favourably with the observed climatology although there is significant variation between the models in the details of this distribution. For example, there are no tropical cyclone-like vortices detected in the PCCSP region in the MIROC3.2 (medres) model. Thus the MIROC3.2 (medres) model was not used in further analysis using the CDD and results are based upon the eight models that produce realistic cyclone climatologies over the study region. In common with the methods described above, the CDD method applied to global models detects few cyclones in the North Atlantic.

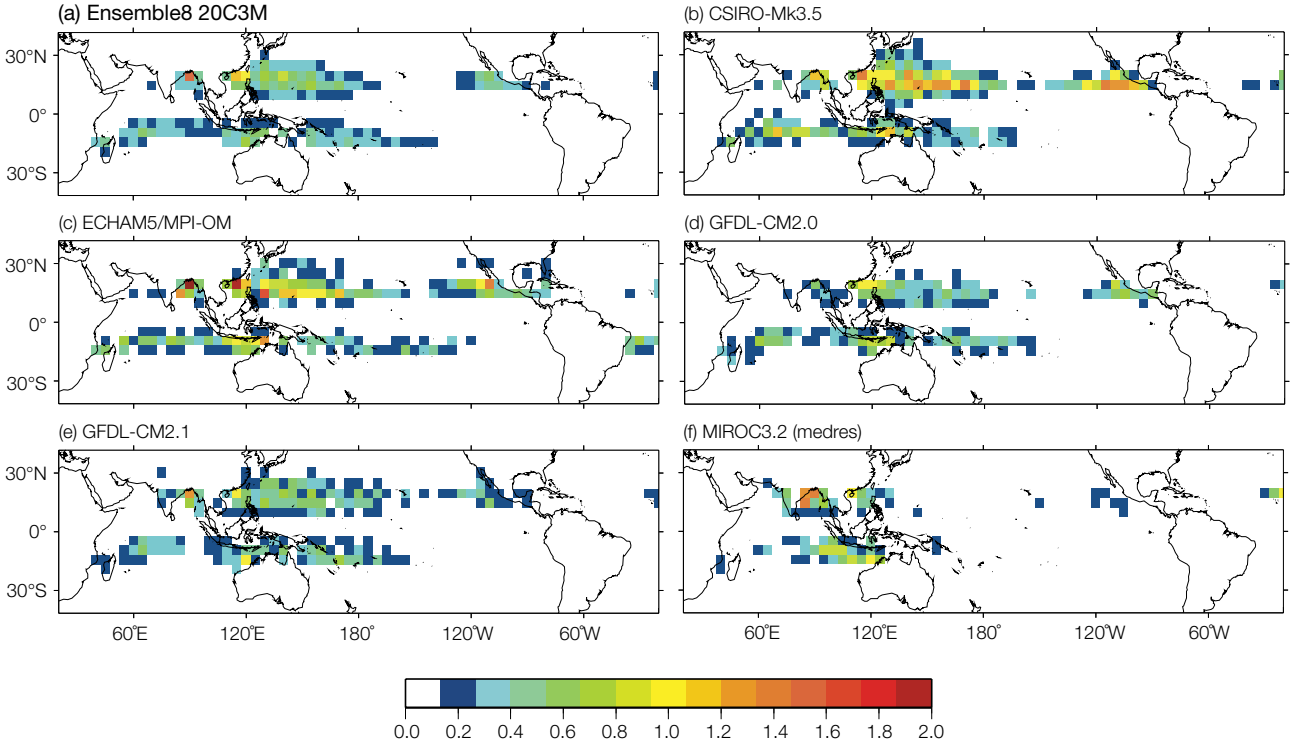

Figure 5.39: Spatial distribution of annual tropical cyclone genesis numbers using the CDD method applied to the CMIP3 global models. (a) is the average climatology based on the results from 8 global models with realistic cyclone climatology and (b-f) are individual results for a subset of the global models considered. Occurrence is expressed as the number of cyclone formations per year within a 5 x 5 degree grid cell.

5.4.3.3 Genesis Potential Index Method

The Genesis Potential Index (GPI, Section 4.8.1) of Emanuel and Nolan (2004) was calculated for the 17 CMIP3 global models for which daily data are available. A subset of 14 models[2] that were found to satisfactorily represent the large-scale climate features of the region is used in the analysis shown in Figure 5.40. The GPI method, a measure of the large-scale environmental conditions conducive to cyclone formation, was developed and calibrated using the NCEP reanalyses. When applied to the outputs of global models or other reanalyses, there can be large differences in results, as illustrated in Figure 5.40(b) and (c). When averaged over the 14 models, the more intense values of the index represent the observed genesis regions reasonably well (Figure 5.40(a)), although the spatial extent of the main genesis regions is too large.

5.4.4 Tropical Cyclone Wind Risk Hazard

Tropical cyclones present a significant hazard to countries situated in the warm tropical waters of the western Pacific. The hazards posed by these severe storms include extreme winds, storm surge inundation, salt water intrusion into ground water supplies, and flooding and landslides caused by the intense rainfall. Despite high exposure to tropical cyclones, there have been few studies attempting to quantify the hazard posed by these severe storms to this region. An exception is a limited number of detailed case studies that have been performed in support of developing a regional disaster insurance scheme similar to that implemented in the Caribbean (Shorten et al., 2003; Shorten et al., 2005). This study aims to address the limited understanding of the extreme wind hazard in this region by evaluating the wind hazard from tropical cyclones using a combination of historical tracks and downscaled climate models with Geoscience Australia's Tropical Cyclone Risk Model (TCRM, Section 4.8.4).

Historical track data from the IBTrACS tropical cyclone database for the period 1981–2008 were fed into TCRM to generate estimates of the maximum 3-second gust wind speed from tropical cyclones for a given return period. The 500-year return period wind speed is used as the primary measure of wind hazard for the following analysis since it is suitable for considering the design loads on residential buildings (AS/NZS 1170.0:2002). It should be noted that these estimates are of regional wind speed and do not account for local factors such as terrain roughness, wind shielding effects and topographic acceleration.

The results are presented in Table 5.8 for each of the 15 PCCSP Partner Countries. Where available, the 500-year return period wind

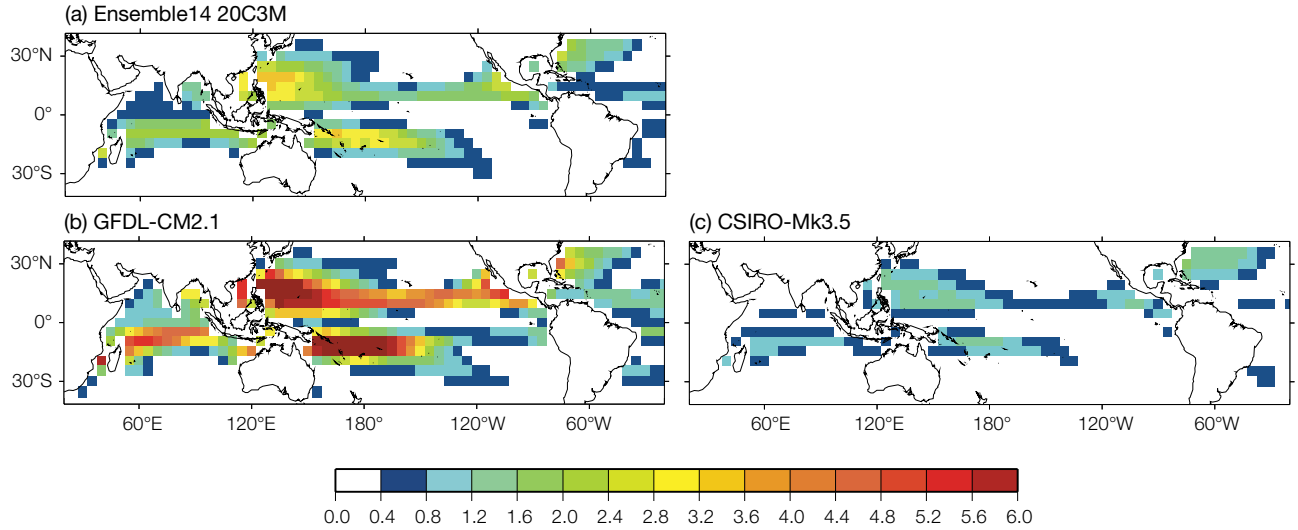

Figure 5.40: Spatial distribution of annual tropical cyclone genesis numbers using the GPI method applied to the CMIP3 global climate models, (a) is the average climatology based on the results from 14 global models and (b) and (c) are results for the GFDL-CM2.1 and CSIRO-Mk3.5 models respectively. Occurrence is expressed as the number of cyclone formations per year within a 5 x 5 degree grid cell. Note that the scale in this figure differs from Figure 5.35.

[2] The unsatisfactory global climate models that were removed from this analysis are those identified in Section 5.5.1

loading standard for each of the countries (HB 212-2002) is provided as a reference. The 500-year return period wind speed estimated from TCRM is found to exceed the design wind speeds by between 15% and 30%. These results may not be unreasonable, however, given that there is evidence that the current wind loading standards may underestimate the cyclonic wind hazard (Rattan and Sharma, 2005). Spatial maps of the 500-year return period cyclonic winds are shown in Figure 5.41.

5.4.5 Summary: Simulation of Tropical Cyclones

The CDD and CVP projection methods have provided realistic late 20th century tropical-cyclone climatologies and are suitable for cyclone detection of the late 21st century climate simulations introduced in Chapter 7. These are subsequently used to produce projected changes in tropical cyclone frequency and wind hazard by the late 21st century for the PCCSP region.

Table 5.8: Cyclonic wind hazard (metres per second) for the PCCSP Partner Countries. Values are taken as the median wind hazard found in a 2 degree x 2 degree region centred on each country's capital city.

Country	Return period wind hazard				Standard
	25yr	50yr	100yr	500yr	500yr
Cook Islands	68	77	84	95	-
East Timor	44	55	62	75	-
Federated States of Micronesia	50	58	64	74	-
Fiji	58	64	69	76	66
Kiribati	-	-	-	-	-
Marshall Islands	54	64	71	82	-
Nauru	-	-	-	-	-
Niue	63	71	77	86	-
Palau	57	65	71	80	-
Papua New Guinea	33	42	48	58	45
Samoa	62	69	75	84	66
Solomon Islands	34	41	46	53	45
Tonga	64	70	75	82	66
Tuvalu	35	41	46	53	-
Vanuatu	69	75	79	86	66

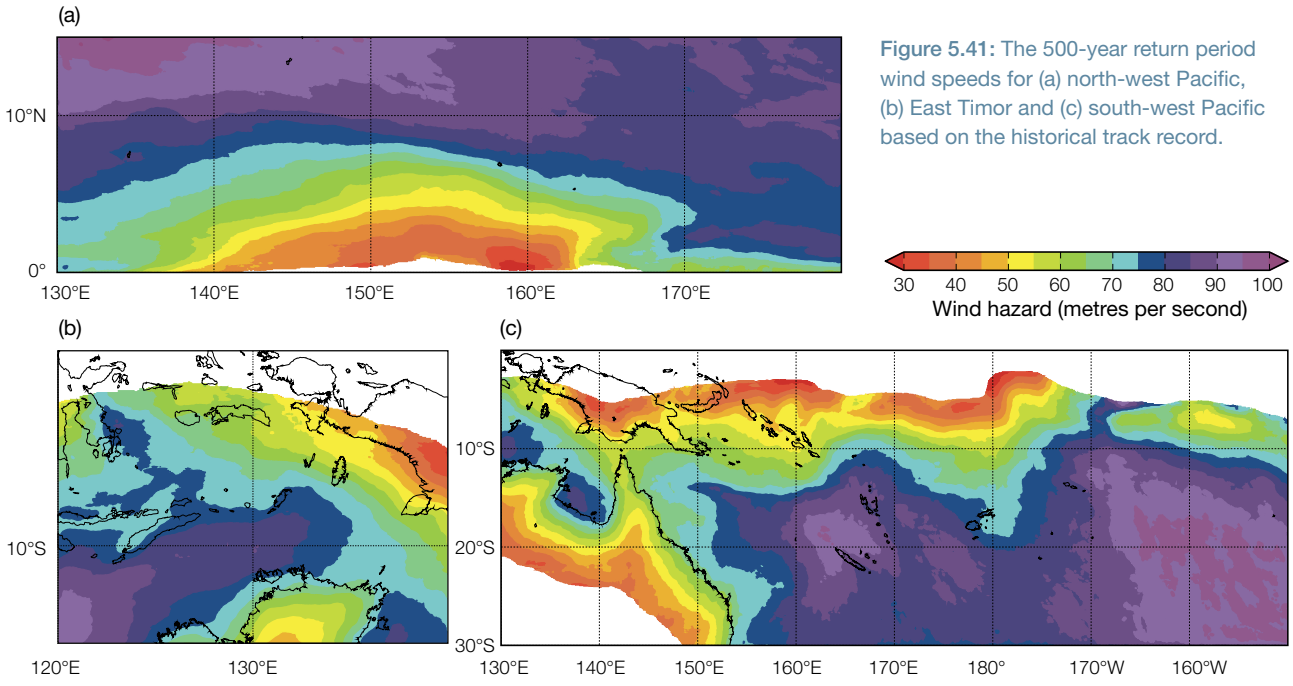

Figure 5.41: The 500-year return period wind speeds for (a) north-west Pacific, (b) East Timor and (c) south-west Pacific based on the historical track record.

5.5 Model Reliability and Implications for Projections

5.5.1 Use of Global Models for Climate Projections

As demonstrated throughout this chapter, the ability of individual CMIP3 models to simulate the western Pacific climate can vary depending on which aspect of a model simulation is considered. While this makes it difficult to identify a group of best performing models, it is possible to identify a small subset of models that perform consistently poorly across many aspects of a climate model simulation, or that perform poorly on critical aspects of a simulation. The approach adopted for determining climate change projections for the PCCSP region has been to equally weight all participating CMIP3 models, with the option of eliminating any models that display unsatisfactory performance in simulating key aspects of the Pacific climate (see Section 4.4.1 for justification of this approach).

To assist in the identification of poor performing models, much of the model evaluation summarised in Sections 5.2.1, 5.2.2 and 5.2.3 (or slight variations thereof) was combined in order to calculate a normalised skill score for each model, with respect to the simulation of surface air temperature, rainfall, surface wind, SPCZ, ITCZ, West Pacific Monsoon, ENSO, model drift and long-term sea-surface temperature trends over the PCCSP region (Irving et al., in press; Figure 5.42).

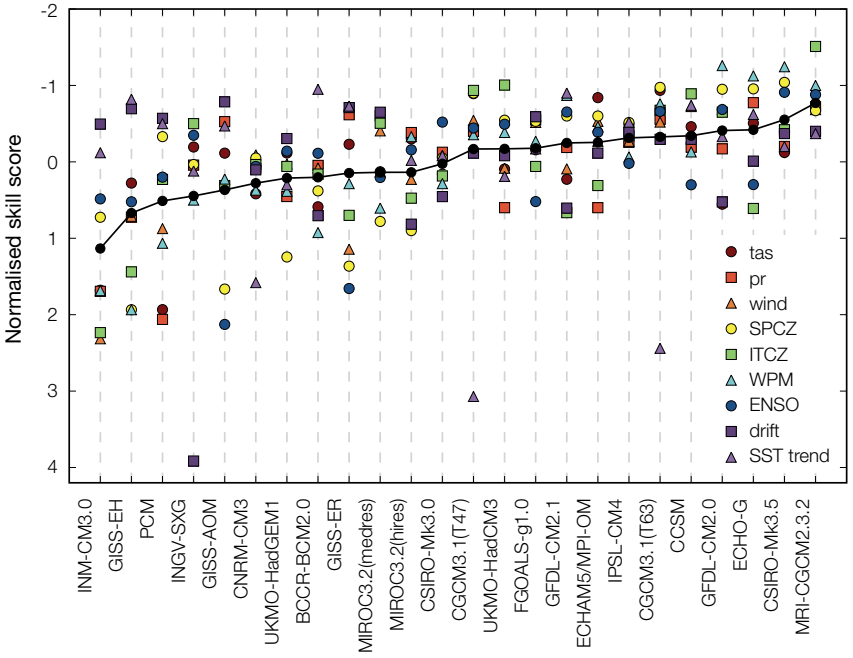

Figure 5.42: Summary of the normalised CMIP3 model skill scores for surface air temperature (tas), rainfall (pr), surface wind (speed and direction combined), SPCZ, ITCZ, West Pacific Monsoon, ENSO, drift and long-term sea-surface temperature trends (see Irving et al., in press for details). The connected solid black dots represent the average skill score across all categories. Increasingly negative scores indicate better model performance.

From the results of Irving et al. (in press) and additional analysis presented in the individual sections of this chapter, the following models were eliminated in calculating all PCCSP climate projections, for the reasons outlined (see also Table 5.9):

- GISS-EH, INM-CM3.0 and PCM: These models perform particularly poorly with respect to the simulation of many aspects of the present day climate over the PCCSP region (Sections 5.2.1 and 5.2.4).
- INGV-SXG: This model has a large climate drift (Section 5.2.2.5) and does not provide the required control simulation data to remove this drift from projected changes.
- GISS-AOM and GISS-ER: These models perform particularly poorly with respect to their simulation of the present day ENSO (Section 5.2.3.1), which was considered to be a critical aspect of the PCCSP region climate.

In addition, the following models were eliminated for specific projections, due to critical deficiencies relating to isolated climate features:

- MIROC3.2(hires) and MIROC3.2(medres): These models were eliminated in determining projections of future SPCZ activity, as they perform particularly poorly in simulating the present day characteristics of the SPCZ (Section 5.2.3.3; Brown et al., in press).
- MIROC3.2(hires): This model was eliminated in determining projections of future ITCZ activity, as it performs particularly poorly in simulating the present day characteristics of the ITCZ (Section 5.2.3.4).
- MIROC3.2(medres): This model was eliminated in determining projections of future West Pacific Monsoon activity, as it performs particularly poorly in simulating the present day characteristics of the monsoon (Section 5.2.3.5).

Table 5.9: CMIP3 models eliminated for the purposes of determining climate projections, based on an evaluation of their present day climate simulation.

Type of projection	Eliminated models	Reason for elimination
All projections	GISS-AOM GISS-ER	Very weak or no ENSO (both models) and a poor simulation of extreme events (GISS-ER)
	INGV-SXG	Model drift too large
	GISS-EH INM-CM3.0 PCM	Overall poor performance (all three models) and a poor simulation of extreme events (GISS-EH, PCM)
SPCZ projections	MIROC3.2(hires) MIROC3.2(medres)	SPCZ poorly simulated
ITCZ projections	MIROC3.2(hires)	ITCZ poorly simulated
West Pacific Monsoon projections	MIROC3.2(medres)	West Pacific Monsoon poorly simulated

5.5.2 Use of Downscaled Models for Climate Projections

The large-scale climate of the CCAM global simulations was found to be closer to observations than the corresponding CMIP3 climate models in some aspects but less in other respects. These simulations do not fully incorporate atmosphere-ocean feedbacks which have been shown to be important for capturing some features, such as monsoon processes (cf. Wang et al., 2005). For these reasons, downscaled projections of changes in these climate features need to be treated with caution.

Further, due to the computational cost, only a limited number of the global climate models and emissions scenarios have been downscaled, so the full range of possible projections has not been sampled. While not all aspects of the large-scale CCAM climate represent an improvement over the CMIP3 global climate models, the 8 km resolution simulations substantially improve the simulation of local climate influences associated with factors such as topography and coastline. As with all projections, it is necessary to assess and understand the physical mechanisms associated with the various changes. In light of these findings, the level of confidence associated with downscaled climate projections is not uniform across the PCCSP region.