4. Model evaluation and confidence

This chapter outlines the process of assessing the confidence in climate projections for Victoria, including the new high-resolution simulations. We cannot assess the climate projections against the future events as we do for weather forecasts, so we must assess the model in the current climate, compare projections from different models and assess our understanding of the relevant processes driving change to gauge the confidence in projections. The new CCAM climate modelling shows some inevitable biases compared to observations, as all climate models do, but is found to be appropriate for assessing regional climate change patterns with confidence.

4.1 Confidence

How to best use projections depends on the degree of confidence we have that they are reliable and complete. Projections with higher confidence can inform choices more definitely. In contrast, lower confidence projections can be used to inform scenario-based adaptive planning or risk-management approaches that can account for uncertainty. Confidence ratings are therefore a key tool when using projections.

VCP19 follows the conventions of the most recent national climate projections (CSIRO and Bureau of Meteorology 2015) and the Intergovernmental Panel on Climate Change (IPCC) assessment reports (Mastrandrea et al. 2010) in assigning confidence ratings to projections. Climate projections are not assessed in the same way as weather forecasts. Projections are made for a series of 'what if' emissions scenarios rather than a single set of inputs. They are estimates of the change in state rather than forecasts of the exact sequence of events. This means they are detailed scenarios of plausible future climates, which is a useful tool to inform decision-making, not a definitive set of results.

Confidence statements applied to a climate projection are determined through an expert elicitation process. This draws on multiple scientific experts' judgment of its reliability as a guide to the range of change for a given input scenario. For VCP19, we draw upon previous lines of evidence and expert judgments on confidence from previous studies including IPCC assessments, the national climate projections and VicCI. The project also draws on new lines of evidence, such as new model simulations, and also the expert judgement of the project team and technical reference group to refine and add to confidence statements. Model evaluation is one key line of evidence used to assess confidence in projections. The other lines include process understanding, theory, agreement with past trends that can be attributed to human influence, consistency between models and expert judgment. Confidence in a projected change is based on the type, amount, quality and consistency of evidence and the extent of agreement among the different lines of evidence (see Figure 7).

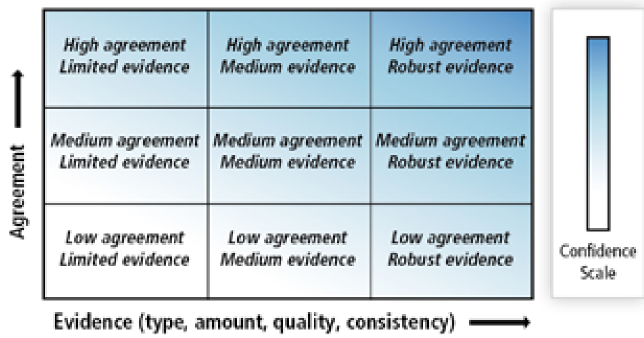

Figure 7. A depiction of evidence and agreement statements and their relationship to confidence. Confidence increases towards the top-right corner as suggested by the increasing strength of shading. Generally, evidence is most robust when there are multiple, consistent independent lines of high-quality evidence.

Confidence is high when:

- the processes involved in the change are well understood
- there is a well-established theoretical basis
- past trends due to human influence agree with the projected change
- the relevant Earth system processes that influence climate change are simulated by the models well
- the models largely agree on the projected change for an appropriately sized ensemble of climate model predictions.

Confidence can be assessed on both the direction of change, and the magnitude of change. For example, confidence may be high in the direction of change but lower in the magnitude of change. In general, projections of factors more directly related to the energy balance of the Earth and the effect of an enhanced greenhouse effect (e.g. ocean heat, temperature) have higher confidence than those that are primarily related to flow-on effects onto features such as atmospheric circulation (e.g. regional rainfall, frequency of storms).

For VCP19 we have assessed the confidence in the new high-resolution modelling. New insights into how climate change may vary across the regions of Victoria are potentially very valuable, so confidence in this detail needs to be carefully assessed.

For these projections we use all inputs that are available and have not been found to be in error or unacceptable. Rather than use only the new modelling and nothing else, we use all sources of information but put a special focus on the new insights generated by the new modelling. There are cases where other model simulations suggest plausible changes outside the range generated by the new VCP19 runs, and we recommend that these also be considered as they represent plausible projections of climate change. In doing this, we aim to reduce the risk of underestimating the range of projected change. Overreliance on a narrow range of change can lead to maladaptation or maladaptive decisions.

As well as the confidence in the nature of the effect of climate change on the regional climate, the other aspect informing the use of climate projections in decision-making is completeness. Here we will cover three main dimensions of completeness: emissions scenarios, climate response and downscaling methods. The first dimension of completeness is the emissions scenarios, where a range of plausible scenarios should be explored and if one scenario is not included then there needs to be a rationale given. VCP19 reports on a high scenario (RCP8.5) and a moderate scenario (RCP4.5) and includes some information about the ambitious mitigation scenario of meeting the Paris Agreement target of 2°C global warming since pre-industrial times. High-resolution modelling is available for RCP8.5 and RCP4.5. These were chosen as they are more relevant to managing higher risk scenarios through adaptation, but these data should always be placed in the context of their emissions scenario.

The next dimension of completeness is the range of plausible climate response to each scenario. The range of results from a set of GCMs provides our best estimate of the possible response to emissions (noting that this model range may not be a complete and reliable estimate of the response). The project uses the entire set of CMIP5 GCMs alongside the downscaled outputs. Also, the six models used for downscaling were chosen to be broadly representative of the CMIP5 ensemble in terms of temperature, rainfall and wind-speed change (changes to many other variables are then correlated with these).

The last dimension of completeness is choices of how to process and downscale data. Different methods give different results and a comprehensive intercomparison of global models, downscaling and processing methods is ideal. Currently the only such coordinated downscaling experiment for Australia is the 50 km resolution CORDEX Australasia experiment described in section 2.1. Comparison of the new CCAM 5 km resolution simulations against other downscaling methods available contributed to the assessment of confidence. The primary post-processing method of the CCAM 5 km simulations used was percentile-percentile scaling to form the application ready data sets described in section 2.5.

4.2 Model evaluation

Before using climate model output to contribute towards regional climate projections, it is important to evaluate a model's strengths and weaknesses. This evaluation informs the level of confidence in the CCAM projections provided in Chapter 5. In this section the performance of the CCAM regional climate model simulations at 5 km resolution for Victoria is evaluated, with a focus on:

- a combination of mean temperature and rainfall that is commonly used in climate impact studies
- extreme rainfall, as this is where the dynamical downscaling can add value to the GCM projections
- larger-scale features, including mean sea-level pressure and large-scale circulation patterns that reflects CCAM as a single modelling system.

When comparing CCAM's ability to represent regional features, the model evaluation relies on the *Australian Water Availability Project* (AWAP) data sets developed by the Australian Bureau of Meteorology and CSIRO. AWAP provides an approximately 5 km resolution gridded data set of daily maximum near-surface (2 m) air temperature, daily minimum near-surface (2 m) air temperature and daily rainfall, that is based on weather station measurements. AWAP is an important data set for evaluating high-resolution climate simulations, although it does have some limitations. For example, AWAP is based on land-based observations so that information over the ocean is interpolated. Also, some regions have a sparser density of weather stations, such as for mountain regions, which can lead to some local gaps in the measurements and potentially an underestimate of rainfall in some locations. Large-scale features of the simulated climate are evaluated using the European Centre for Medium-Range Weather Forecasting (ECMWF) ERA-Interim reanalysis data set. This reanalysis product has a resolution of approximately three-quarters of a degree and is based on the assimilation of various observation data sets, including satellite-based measurements, to build a consistent interpretation of the state of the atmosphere at that time. The ECMWF atmospheric model is used to address gaps in the observations when constructing the reanalysis data set. A new ERA-5 data set is being released by ECMWF that will replace ERA-Interim but this was not available at the time the CCAM simulations were conducted. Another data set is the Bureau of Meteorology Atmospheric Regional Reanalysis for Australia (BARRA) which is a regional reanalysis that assimilates observations into the ACCESS weather forecasting model and employs the model to fill in gaps in the observing network. However, we have not conducted an extensive evaluation using BARRA because at the time of writing the data set was incomplete with reduced number of simulation years and the final version of the data set had not been published. Nevertheless, we do comment on features represented by BARRA when relevant to the model evaluation.

As stated previously, the regional climate model output should be used in combination with the global climate model output. The following analysis compares the downscaled results with the global climate model results as appropriate. Additional information regarding the evaluation of global climate models can also be found in Chapter 5 of the CCIA technical report (CSIRO and Bureau of Meteorology 2015).

4.2.1 Temperature

The results for the daily maximum near-surface (2 m) air temperature (Figure 8) and the daily minimum near-surface (2 m) air temperature (Figure 9) show some of the improvements, as well as some limitations with the downscaled simulations compared to the global climate model results. Note that the AWAP data is based on land-based weather station measurements and can be less reliable over the ocean where the data is interpolated. A bias plot for the daily minimum and maximum temperatures can also be seen in the appendix of this report.

Daily maximum temperature results for CCAM show an improved representation of spatial detail, particularly when representing mountain ranges and, to a lesser extent, coastlines. However, there is also a simulated warm bias (i.e. CCAM compared to AWAP) of several degrees along the east coast of Victoria. This warm bias seems to correspond to forested regions with high vegetation, which may be related to a mismatch between CCAM's calculation of air temperature within the canopy and the observations which are made in clearings. Further investigation is needed to categorically identify the source of the bias in the maximum near-surface (2 m) air temperature. The projected changes in temperature under global warming discussed in Chapter 5 do not appear to be sensitive to the location of these biases, which suggests that the temperature bias does not directly affect the projected changes in temperature. Although all climate models have biases (e.g. see the discussion of minimum temperature below), it appears possible that the problem with the temperature bias could be reliably addressed in a post-processing procedure. If this is the case, then an updated temperature data set will be generated once the problem has been corrected.

Daily minimum near-surface (2 m) air temperatures are well represented by the CCAM model, which shows an improvement compared to the six host GCMs shown in Figure 9. In addition, the CCAM results show a realistic representation of the urban heat island, where daily minimum temperatures are typically 1°C warmer for urban areas than would be the case for natural vegetation. Urban heat islands are further discussed in section 4.2.2.

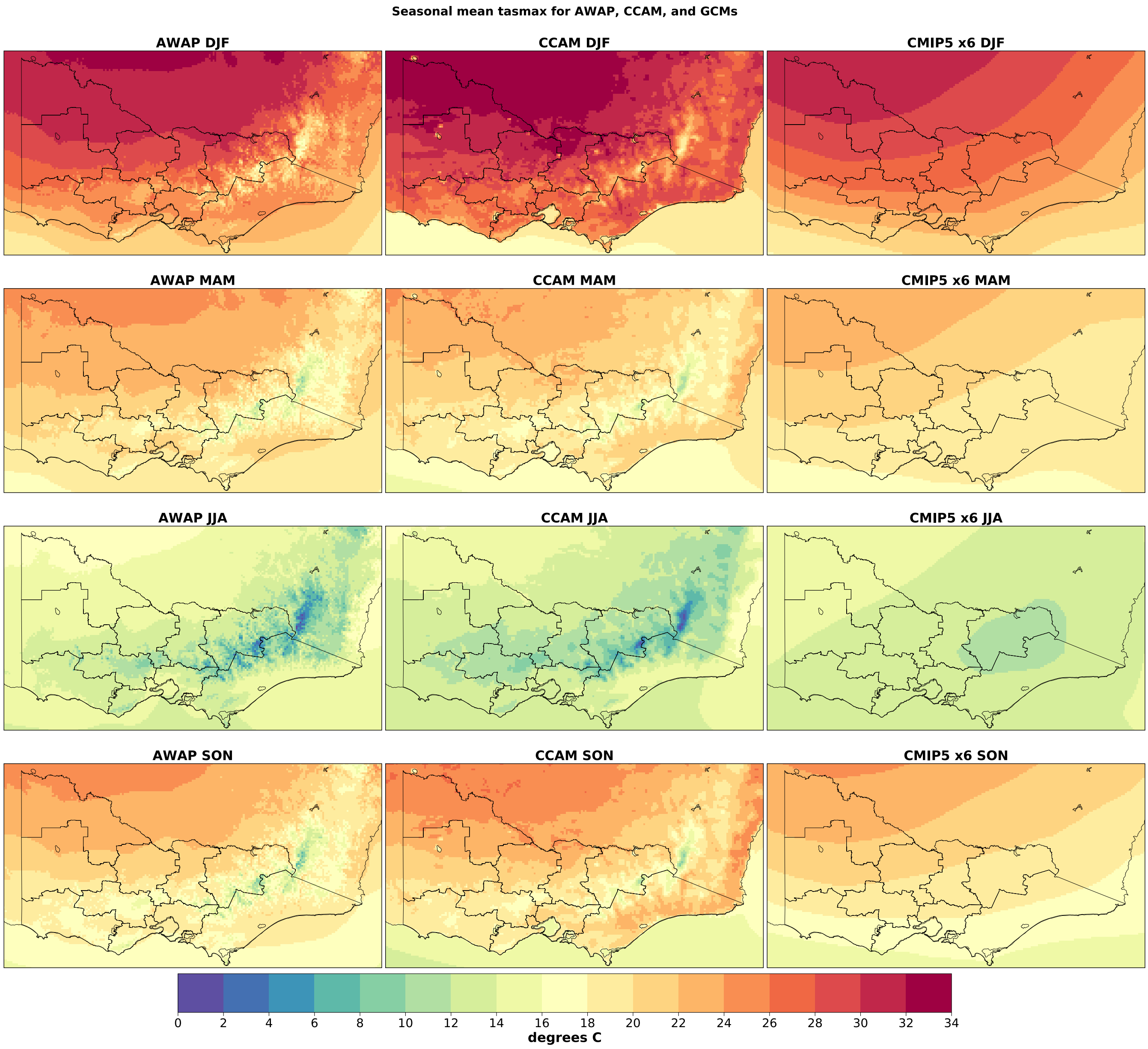

Figure 8. Average daily maximum near-surface (2 m) air temperature (°C) from the CCAM 5 km resolution simulations for Victoria for 1986–2005. The left column is the observed climate from the AWAP 5 km gridded climate data set, the middle column is the mean of six CCAM simulations, and the right column is the mean of the six host GCMs. Top row is December to February (DJF), the second row is March to May (MAM), the third row is June to August (JJA) and the fourth row is September to November (SON).

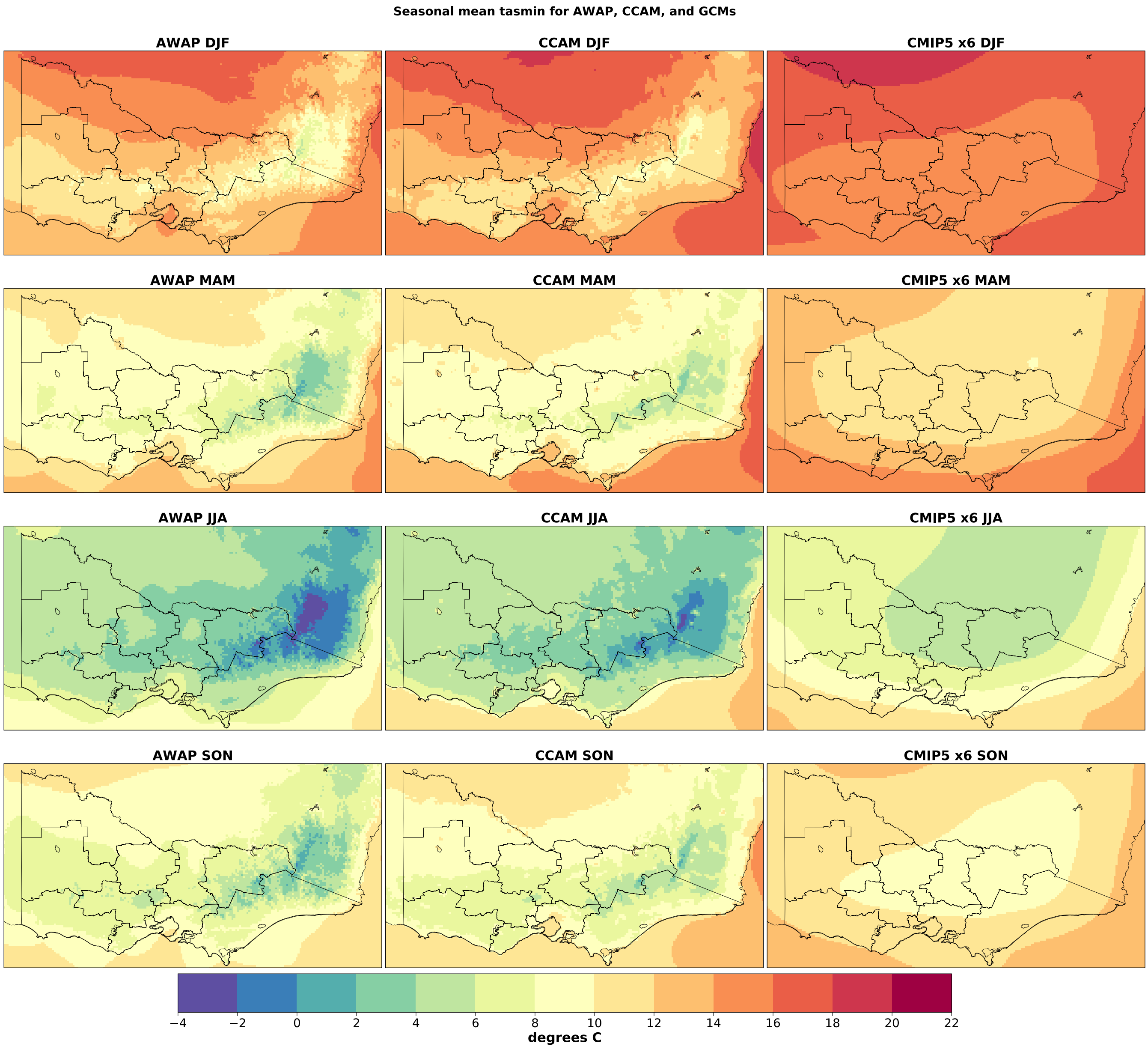

Figure 9. As for Figure 8, but showing average daily minimum near-surface (2 m) air temperature (°C)

4.2.2 Urban heat island

One of the potential advantages of using a regional climate model like CCAM is to better represent urban areas that are neither resolved nor parameterised in the host global climate models. The ability of a climate model to simulate urban areas in a realistic way can be assessed by its ability to model the urban heat island (UHI). The UHI refers to the increased daily minimum near-surface (2 m) air temperature in urban areas due to the storage of heat in buildings and roads. Most major Australian cities have an UHI of +1°C to +2°C, depending on the nature of the local built environment. The size of the UHI is usually estimated by comparing measurements of daily minimum temperature from the fringe of the city with that at the city centre, inferring the enhanced warming in daily minimum temperature due to the presence of the city. Although there are other factors which can influence the temperature difference, such as elevation and rainfall, the presence of the urban area is the main factor in determining the difference in daily minimum temperature.

The UHI is estimated by comparing inner-city temperature measurements to the temperatures measured at sites on the fringe of the city. The difference in daily minimum temperature is compared between the inner-city weather station and the outer-city site that approximates the natural vegetation. An example of this approach is shown in Figure 10, where we use the BOM regional office as an inner-city site indicated as a red dot. This inner-city site is then compared to three surrounding sites indicated by Laverton RAAF (blue dot), Coldstream (green dot) and Cranbourne Botanic Gardens (yellow dot). By comparing the differences in these temperatures between each of the three outer-city sites against the BOM regional office, we can estimate the temperature gradient arising due to the presence of the Melbourne urban area.

A comparison of the observations between the inner-city site and the three outer city sites is shown in Figure 11, between observations at weather stations, the simulated climate from CCAM and the simulated climate from the GCMs. This was done by calculating the difference in daily minimum temperature between the inner-city site and the three outer city sites for each day between 1986 and 2005. This difference in minimum temperatures is averaged over time and then shown in Figure 11. The blue bars indicate observed UHI of approximately 2°C warmer between the inner city and Laverton, over 4°C warmer between the inner city and Coldstream, as well as a 2°C difference between the inner city and Cranbourne. We note that the red bars show the CCAM results after downscaling the GCMs, indicating that CCAM correctly simulates the difference in daily minimum temperature between the inner city and Laverton, as well as

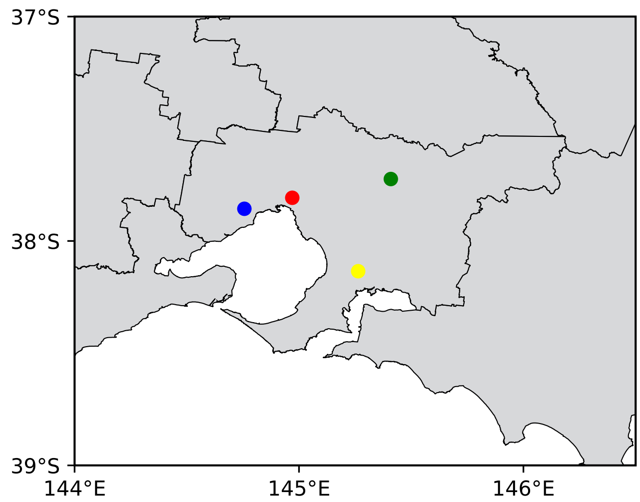

Figure 10. Plot of the locations used to estimate the urban heat island for Melbourne. Dots indicate the location of BOM weather stations: BOM Melbourne Regional Office (red), Laverton RAAF (blue), Coldstream (green) and Cranbourne Botanic Gardens (yellow).

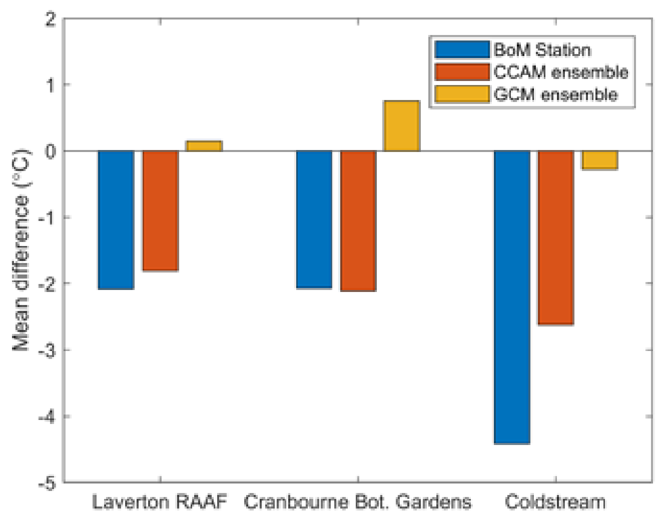

Figure 11. Estimated urban heat island (UHI) averaged over the time period 1986–2005. The UHI is measured as the difference in temperature between the inner-city BOM regional office weather station compared to the three outer-city sites of Laverton RAAF, Coldstream and Cranbourne Botanic Gardens (blue bars). Since the inner-city site is warmer than the outer-city sites due to urban development, we see a negative value for the difference in this plot. The observed data is shown as blue bars and can be compared to the CCAM simulation results (red bars) and the host GCM (orange bars). The results represent the average of the six CCAM simulations and the average of the corresponding six host GCMs that were downscaled by CCAM.

the difference between the inner city and Cranbourne. This is partially a consequence of the UCLEM urban parameterisation (see section 2.2) which represents urban areas in the simulation. CCAM underestimates the difference in minimum temperature between the inner city and Coldstream by effectively overestimating the minimum temperature at Coldstream, resulting in a smaller gradient in temperature between the inner city and Coldstream than was observed. In the case of the GCMs where the temperature data must be interpolated for the locations of the weather stations, the urban area is not resolved and not necessarily parameterised by the climate model (shown as orange bars). Consequently, it is difficult for GCMs to represent the temperature gradient between the inner-city site and the three surrounding outer-city sites. Overall, CCAM simulations show a substantially improved representation of the UHI than the host GCM.

4.2.3 Average rainfall

Climate models simulate precipitation, including rain and snow since the precipitation falls as snow when temperature and other atmospheric conditions are conducive. However, throughout this report precipitation is referred to as rainfall, only mentioning snow when relevant (e.g. section 5.3.4).

Spatial and seasonal characteristics of rainfall are particularly difficult for climate models to accurately represent. Notwithstanding, dynamical regional climate models have the potential to improve GCM simulations of rainfall. In part, this is due to better representation of topography such as mountain ranges and coastlines. Seasonal rainfall simulations from CCAM and GCMs compared to observations are shown in Figure 12. A bias plot for average rainfall can also be seen in the appendix to this report.

It is clear from Figure 12 that CCAM better represents rainfall along the Australian Alps compared to the host GCMs. This is due to better representation of the orography of the mountain range. In particular, the CCAM simulations show a rain shadow on the eastern slopes of the Alps, with corresponding enhanced rainfall on the western slopes. There is a tendency to show heavier rainfall over the mountains compared to the AWAP observations. However, AWAP is also known to underestimate daily extremes to some extent due to the lower network density in some alpine regions (Jones et al. 2009). Comparison with the preliminary results from the BARRA reanalysis data sets (not shown) also indicates higher rainfall in the alpine region, when compared to AWAP. Notwithstanding, the CCAM simulated extreme rainfall appears to be higher than the observed rainfall which is partly a limitation of the CCAM cloud microphysics parameterisations. In any event, the representation of rainfall is an improvement on the GCM simulations. This is meaningful for the projected rainfall change discussed in section 5.3. An interesting result shown in Figure 12 is that the larger-scale rainfall in the CCAM simulations is similar to that in the GCMs, but has additional detail for mountains and coastlines that was not represented in the GCM (explored further in section 5.3). This result can be explained by some of the similarities between the cloud microphysics parameterisations in CCAM and the host GCMs. We note that both CCAM and the host GCMs are slightly wetter over Victoria than the observed as depicted in the AWAP data set. For example, the CCAM and GCM simulations never show any regions where the seasonal average rainfall is less than 1 mm/day, although the AWAP observations show this occurs in the northwest part of the state.

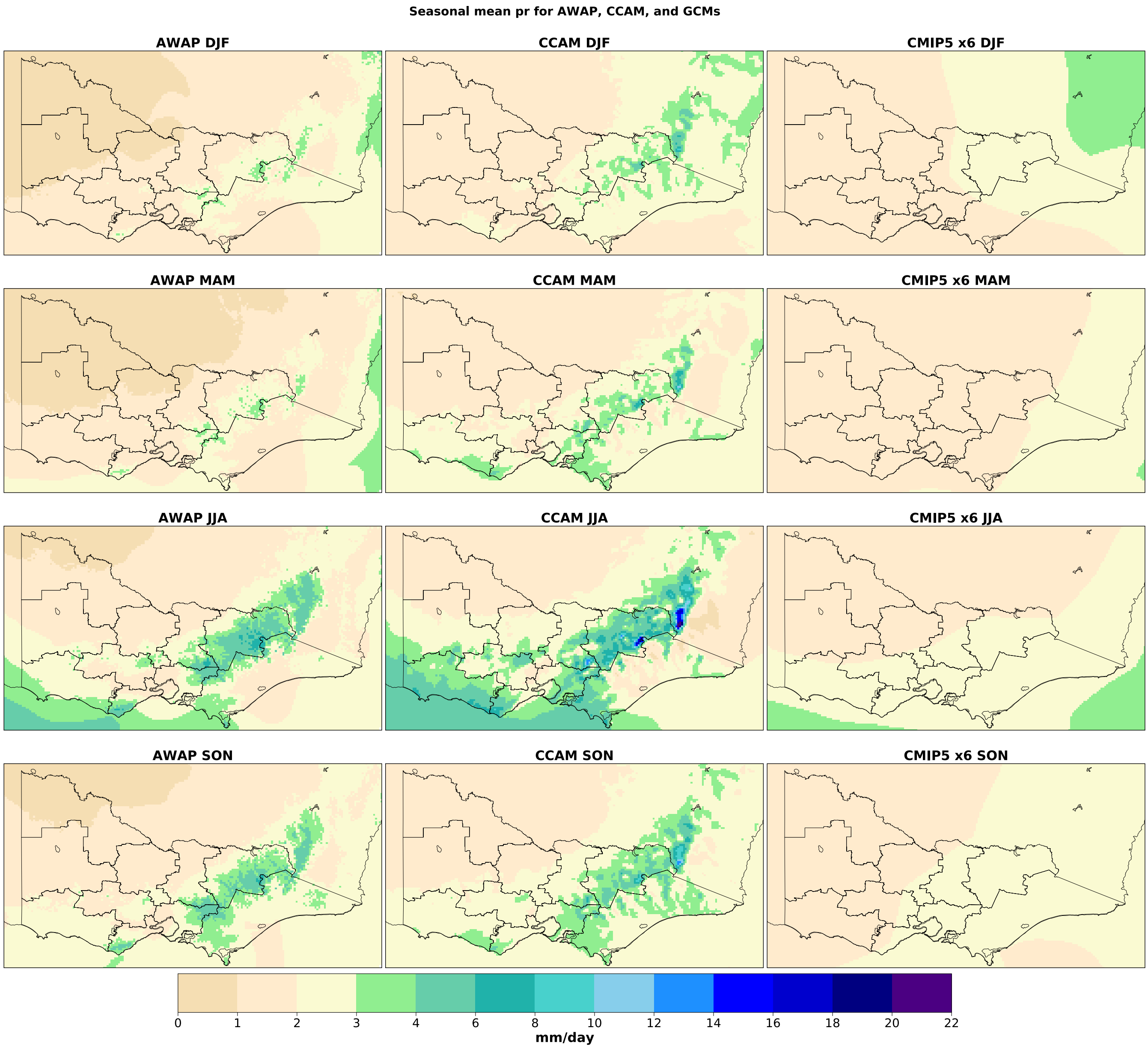

Figure 12. Annual and seasonal mean rainfall for 1986–2005 in the AWAP 5 km gridded climate data set, the mean of six CCAM 5 km simulations, and the mean of the six host GCMs. Top row is December to February (DJF), the second row is March to May (MAM), the third row is June to August (JJA) and the fourth row is September to November (SON).

4.2.4 Extreme rainfall

Extreme rainfall is a key area in which a regional climate model like CCAM has the potential to add important new information not provided by GCMs.

There are several different ways to characterise extreme rainfall, depending on the severity of the event. For simplicity, the 99th percentile of the 1986–2005 rainfall is used as an indicator of how extreme rainfall is simulated by CCAM and by the GCMs. Figure 13 shows the results of CCAM and the GCMs compared to observations (AWAP) for the 99th percentile of 1986–2005 rainfall. AWAP shows that the largest values for the 99th percentile of rainfall occur over the Australian Alps and the eastern coast of Victoria. This result is reflected in the CCAM downscaled simulations, although the values of rainfall are larger for the CCAM simulations compared to AWAP over the mountain ranges. It is probable that CCAM is overestimating these rainfall events; however, AWAP is known to underestimate extremes (Jones et al. 2009). Nevertheless, the CCAM simulations still have a noticeably better representation of the 99th percentile rainfall compared to the host GCM. The GCMs do not reproduce the higher 99th percentile rainfall values over the Australian Alps. The GCMs also fail to reproduce the higher 99th percentile rainfall values for the eastern coast. The difference in the CCAM and GCM results can be partially explained by the unresolved mountain ranges in the GCM, as well as the GCMs relying on their respective convective parameterisations. The extreme rainfall is also better resolved in the downscaled simulations. Although the CCAM simulations are not perfect in their ability to represent extreme rainfall, this is an example where the downscaled simulations have been able to add value compared to the existing GCM data.

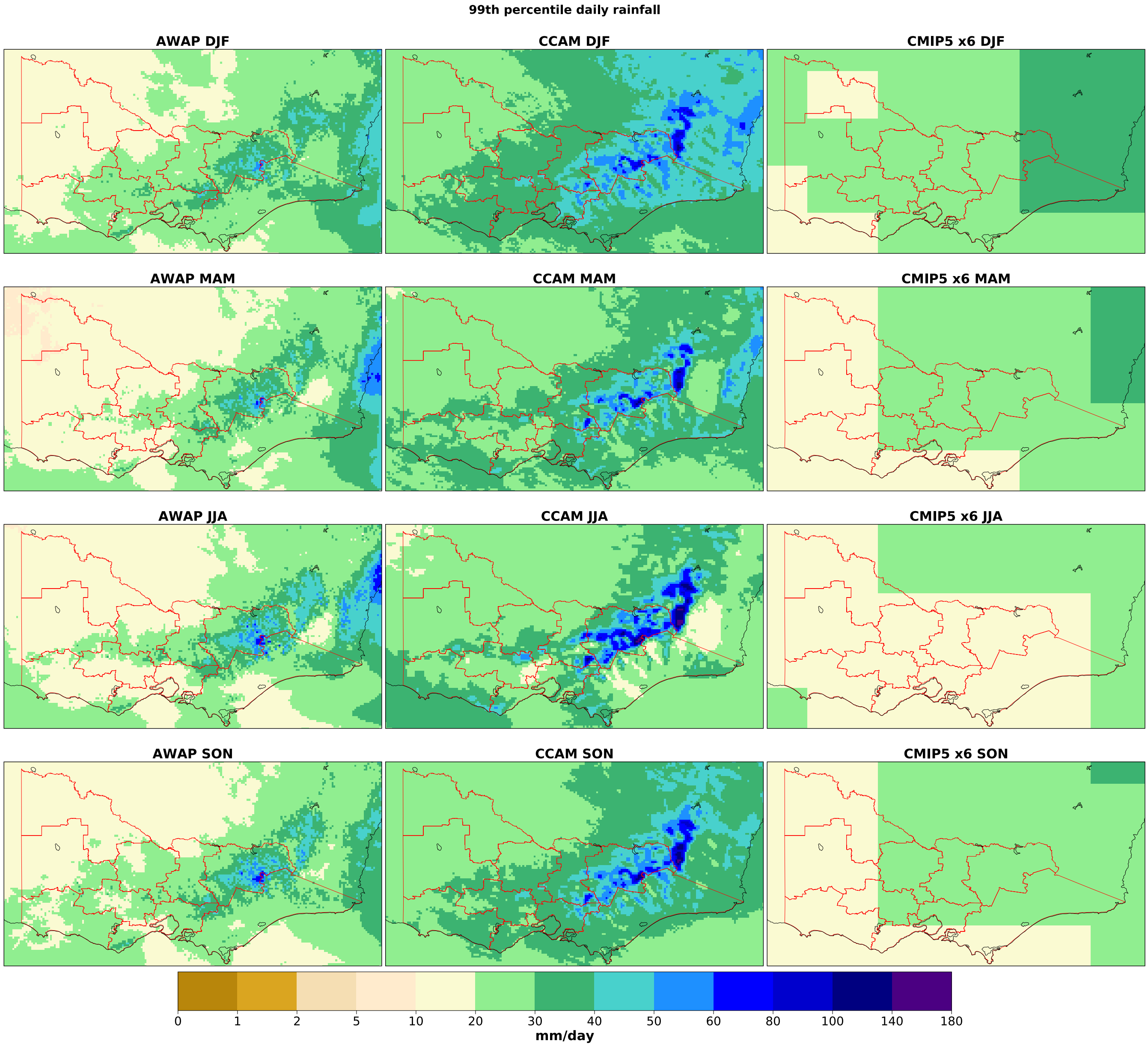

Figure 13. Extreme daily rainfall (99th percentile) for 1986–2005 from the CCAM higher-resolution simulations for Victoria. The left column is the observed climate from the AWAP 5 km gridded climate data set, the middle column is the mean of six CCAM 5 km simulations, and the right column is the mean of the six host GCMs. Top row is December to February (DJF), the second row is March to May (MAM), the third row is June to August (JJA) and the fourth row is September to November (SON).

4.2.5 Mean sea-level pressure

Evaluating the simulated mean sea-level pressure (MSLP) can provide an insight into a climate model's ability to represent the mean circulation. This can often indicate larger-scale issues with the simulation. This is important in the case of the CCAM simulations shown in this report, since CCAM employs the corrected sea surface temperatures (SSTs) from the host GCM (section 2.2). As a result, the simulated MSLP is not constrained by the host GCM and can deviate from the changes projected by the host GCM.

Figure 14 compares the MSLP results from ERA-Interim reanalyses, CCAM 50 km simulations and the host GCMs. As discussed in section 2.2, the CCAM 50 km simulations constrain the larger scale behaviour that is downscaled by the CCAM 5 km simulations and therefore influences the projections of the CCAM 5 km experiments. In this case the ERA-Interim reanalyses are a reasonable representation of the observed MSLP due to the reanalysis simulation being constrained by observations. The 50 km CCAM results are shown because they represent the larger-scale atmospheric circulation that is subsequently downscaled

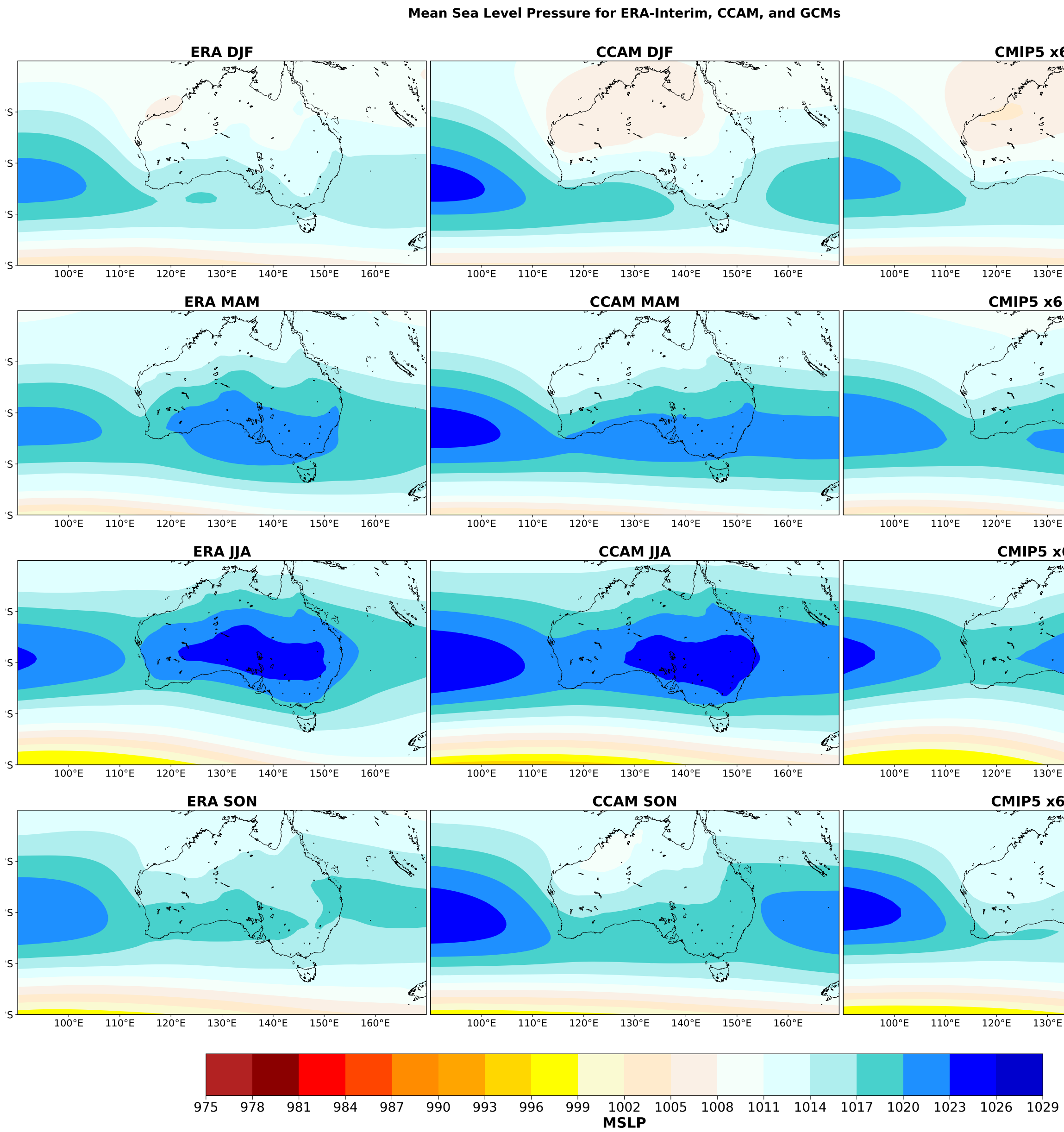

Figure 14. Average mean seasonal sea-level pressure for 1986–2005 in left: the ERA interim reanalysis (ERA); middle: the CCAM simulation averaged over the six downscaled GCMs; and right: average of the six host GCMs. Top row is December to February (DJF), the second row is March to May (MAM), the third row is June to August (JJA) and the fourth row is September to November (SON).

to 5 km resolution over Victoria. When compared to the ERA-Interim reanalyses, the CCAM results for MSLP are too zonal (east-west), with a stronger east-west component, weaker ridges and trough compared to observations. This is most noticeable in autumn (MAM) and winter (JJA) and is one possible cause of the model simulating too much rainfall in autumn. A stronger ridge to the east of Australia is also present in all seasons. In comparison, the GCM host models perform better than CCAM with respect to the simulated MSLP, since the zonal problem is less evident. The differences between the CCAM 50 km and GCMs arise because of the SST bias correction which required the global atmospheric circulation to be reconstructed consistent with the corrected SSTs (see section 2.2). The zonal problem with the CCAM MSLP can influence the dynamical response of the atmosphere under climate change, such as modifying the large-scale winds or large-scale changes to rainfall. This is taken into consideration in the interpretation of the projected changes presented in Chapter 5, and highlights the importance of being mindful of both CCAM and GCM results when looking at projections. This example illustrates some of the issues with using a single modelling system, as the large-scale features in the downscaled CCAM results may be reflective of CCAM as a single modelling system. However, when combined with other downscaling results (e.g. VicCI) and with GCM projections, a more comprehensive projection of the regional climate can be made.

4.2.6 Upper-level wind speed and direction at 850 hPa

Figure 15 shows the average wind speed and average wind vectors at 850 hPa (approximately 1 to 1.5 km above the surface) from ERA-interim, the CCAM 50 km resolution simulations and the six host GCMs interpolated to a common 1.5 x 1.5° lat/lon grid. ERA-Interim is a reasonably accurate depiction on the 850 hPa winds as it is constrained by observations. There is broad agreement among the reanalysis, CCAM and the host GCMs in terms of wind speed and direction for all seasons. However, the CCAM 850 hPa winds are too strong over Victoria in winter. This result is consistent with the mean sea-level pressure being too zonal as described in the section 4.2.5. The implications of this issue with the CCAM simulations are discussed when comparing the results to other models in Chapter 5.

4.2.7 Summary of CCAM evaluation

A common rule of thumb for using climate models is that the better they simulate the current climate then the greater the confidence in the future climate change simulation. Any difference between the modelled current climate and the observed current climate, known as bias, inevitably lowers confidence. However, the question of how much bias is acceptable is in fact complex and depends on the purpose of the model. An evaluation of CCAM downscaling found that it can contribute to the development of regional projections, although there are some deficiencies. The dynamical downscaling successfully captures regional influences on average temperature and rainfall. There is a temperature bias in the daily maximum temperature for eastern Victoria (e.g. Figure 8) which is being addressed by the CCAM developers. This bias is likely to represent an imperfection in how a particular feature of the climate is parameterised in the model, so lowers the confidence in temperature projections to some extent. However, the bias is smaller than many biases in GCMs, and the projected regional changes in temperature do not seem to be spatially correlated with this bias, suggesting the bias does not have a direct effect on the projection of temperature change. Therefore, the temperature results are presented with at least equal confidence as GCM projections.

The urban heat island is noticeably better represented in the CCAM output. Extreme rainfall is much better represented by the dynamical downscaling compared to the GCMs and should add value to the regional projections. Large-scale behaviour of the simulated climate in CCAM is plausible, but has some differences compared to the six host GCMs. Consequently, we should consider changes in the large-scale rainfall may differ from the predictions of the host CMIP5 GCMs. This result emphasises the importance of using the CCAM dynamically downscaled projections of large-scale temperature and rainfall change in conjunction with the GCM output until we have a compelling case to prefer one over the other.

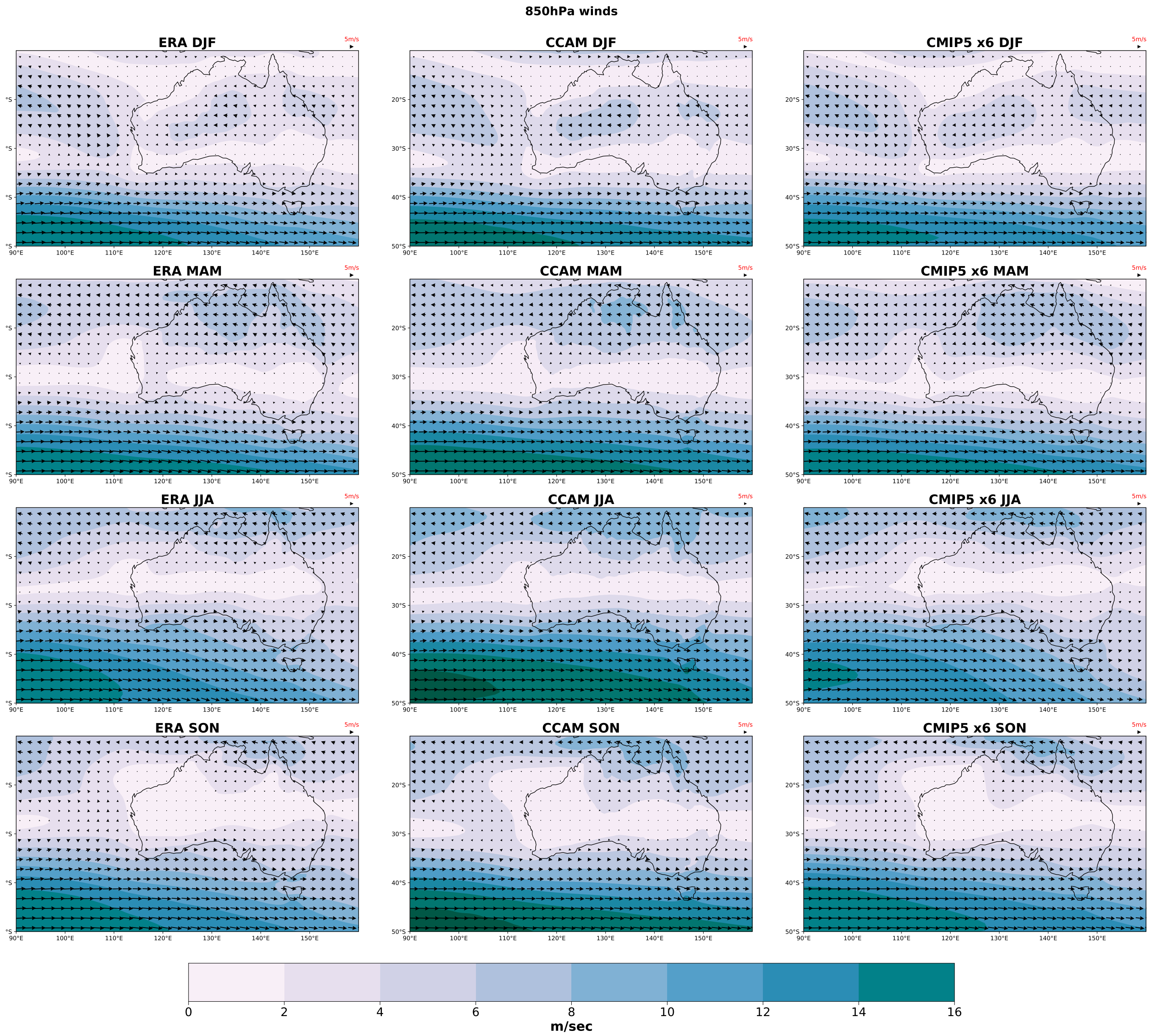

Figure 15. Average wind speed and average wind vectors at 850 hPa (approximately 1 to 1.5 km above the surface) from 1986–2005 for the different seasons, where the speed and direction are shown as vector arrows and the speed is also shown by the colour scale. Left shows ERA-Interim reanalysis (ERA), middle shows CCAM 50 km simulations averaged over the six downscaled GCMs, and right shows the average of the six host GCMs. Top row is December to February (DJF), the second row is March to May (MAM), the third row is June to August (JJA) and the fourth row is September to November (SON).