


An evaluation of the performance of a WRF multi-physics ensemble for heatwave events over the city of Melbourne in southeast Australia

H. M. Imran¹  · J. Kala² · A. W. M. Ng¹ · S. Muthukumaran¹

Received: 1 July 2016 / Accepted: 9 June 2017 / Published online: 19 June 2017
© Springer-Verlag GmbH Germany 2017

Abstract Appropriate choice of physics options among many physics parameterizations is important when using the Weather Research and Forecasting (WRF) model. The responses of different physics parameterizations of the WRF model may vary due to geographical locations, the application of interest, and the temporal and spatial scales being investigated. Several studies have evaluated the performance of the WRF model in simulating the mean climate and extreme rainfall events for various regions in Australia. However, no study has explicitly evaluated the sensitivity of the WRF model in simulating heatwaves. Therefore, this study evaluates the performance of a WRF multi-physics ensemble that comprises 27 model configurations for a series of heatwave events in Melbourne, Australia. Unlike most previous studies, we not only evaluate temperature, but also wind speed and relative humidity, which are key factors influencing heatwave dynamics. No specific ensemble member for all events explicitly showed the best performance, for all the variables, considering all evaluation metrics. This study also found that the choice of planetary boundary layer (PBL) scheme had largest influence, the radiation scheme had moderate influence, and the microphysics scheme had the least influence on temperature simulations. The PBL and microphysics schemes were found to be more sensitive than the radiation scheme for wind speed and relative humidity. Additionally, the study

tested the role of Urban Canopy Model (UCM) and three Land Surface Models (LSMs). Although the UCM did not play significant role, the Noah-LSM showed better performance than the CLM4 and NOAH-MP LSMs in simulating the heatwave events. The study finally identifies an optimal configuration of WRF that will be a useful modelling tool for further investigations of heatwaves in Melbourne. Although our results are invariably region-specific, our results will be useful to WRF users investigating heatwave dynamics elsewhere.

Keywords WRF · Multi-physics ensemble · Heatwaves · Melbourne

1 Introduction

The duration, frequency and intensity of heatwaves in Australia are increasing and expected to increase into the future (Cowan et al. 2014). The definition of a heatwave varies depending on the application, but meteorological definitions are usually based on percentiles (Perkins and Alexander 2013). For example, Nairn and Fawcett (2013) define a heatwave as a period of at least 3 consecutive days where the average of maximum and minimum temperatures exceeds the climatological 95th percentile. Furthermore, a heatwave defined by Schoetter et al. (2015) is a period at least 3 days above the 98th percentile of maximum temperature where the combined effect of excess heat and heat stress is unusual with respect to the local climate. A considerable number of studies have examined the processes behind heatwave events in Australia. The formation of heatwaves in Australia has been linked to various physical drivers including the synoptic drivers (e.g., Perkins 2015) as well as teleconnections which influence heatwaves

✉ H. M. Imran
ihosen83@gmail.com

¹ College of Engineering and Science, Victoria University, PO Box 14428, Melbourne, Victoria, Australia

² Environmental and Conservation Sciences, School of Veterinary and Life Sciences, Murdoch University, Perth, Western Australia

dynamics (e.g., Parker et al. 2013). In general, heatwaves are driven by persistent anticyclonic conditions (commonly referred to as blocking highs). Conventional blocking highs form when higher-level atmospheric winds fragmented due to the meandering of the jet stream, form an area to be blocked from the zonal jet stream flow for several days (Pezza et al. 2012). The blocking highs (anticyclonic systems) commonly form over the Tasman Sea, and are the main synoptic driver for heatwaves over southeastern Australia. Further persistent highs occur at 10°S equatorward, where the subtropical ridge forms during the summer season associated with Rossby wave trains (Marshall et al. 2014; Perkins 2015; Boschat et al. 2015). These blocking highs have been responsible for numerous heatwaves in Australia (Marshall et al. 2014; Boschat et al. 2015). Over Australia, large-scale teleconnections and climate variability have also been shown to influence heatwave dynamics. For example, the El-Nino phase of the El Nino–Southern Oscillation and positive phases of the Indian Ocean Dipole generally result in lower rainfall over Eastern Australia, which has been shown to result in dry soil conditions, which can enhance seasonal extreme temperatures (Jones and Trewin 2000; Cai et al. 2009; Min et al. 2013). Other studies (e.g., Perkins et al. 2015; Herold et al. 2016) have explicitly examined the role of soil moisture deficits on heatwaves in Australia, and have shown that generally, low antecedent soil moisture generally leads to higher heatwave temperatures, but not necessarily more heatwave days in the eastern and Northern parts of Australia (Herold et al. 2016). Therefore, the combinations of anticyclonic conditions as well as soil dryness are important driving factors of heatwaves in southeast Australia.

Cities are particularly prone to the adverse impacts of heatwaves. Increased urbanization is one of the main causes of loss of vegetation, and this has resulted in important changes in land surface properties. As a consequence, pervious surfaces are replaced by impervious built surfaces (e.g. buildings, roads, driveways and sidewalks). These built surfaces are made of high thermal conductivity materials such as concrete, bricks, stones and asphalt. These materials absorb and store heat during the day from the sunlight due to their lower albedo and higher thermal conductivity, and then emit this excess heat at night, which has been shown to result in an increase in night-time temperatures (Arugueso et al. 2014). The well-documented urban heat island (UHI), characterized by the higher temperatures within the urban areas compared to surrounding rural areas, is one of the prominent urban effects. The most severe impacts (e.g. heat-related mortality, energy consumption, air pollution) of UHI are pronounced during heatwaves (Kunkel et al. 1996; Rosenzweig et al. 2005).

The UHI, together with summer time heatwaves, fosters biophysical hazards (Chow et al. 2012; Fischer et al. 2012),

influences air pollution (Rosenfeld et al. 1998), increases energy consumption (Konopac and Akbari 2002), affects ecosystem cycles (Imhoff et al. 2010) and influences local weather and exacerbates warming from climate change (Emmanuel and Krüger 2012). Therefore, it is necessary to study the characteristics and effects of the UHI, to help design mitigation strategies. For these purposes, numerical Regional Climate Models (RCMs) are very effective tools to simulate the UHI. RCMs dynamically downscale Global Climate Models and/or re-analysis products from the large scale (100–250 km) to simulate current and future climate change, and conduct climate and weather research at the regional scale (1–10 km) (Beniston et al. 2007). RCMs can be used to assess the major factors driving the UHI and therefore help design possible mitigation strategies, while taking land-use into consideration (e.g., the effect of vegetation, water bodies, etc). One of the most widely adopted RCMs is the Weather Research and Forecasting (WRF) model (Skamarock et al. 2008), which has been used for several applications, including studies focusing on the UHI (e.g., Giannaros et al. 2013; Hu et al. 2013; Chen et al. 2014; Fallmann et al. 2014). The WRF model has suite of sophisticated physics parameterizations for the land surface, planetary boundary layer (PBL), cumulus (CU), shortwave (SW) and longwave (LW) radiation and cloud microphysical processes. While this offers flexibility to the user, the large number of physical parameterizations can make it difficult to find the optimal setting for a particular application, given a certain spatial and temporal scale of interest.

Several studies have evaluated the performance of different WRF physical parameterizations in simulating high rainfall events (Evans et al. 2012) and sea surface temperature effects on extreme rainfall over south-east Australia (Evans and Boyer-Souchet 2012). Further similar studies have focused on seasonal time scales over south-west Western Australia (Kala et al. 2015) and the effects of land use change on temperature extremes over the Australian continent (Hirsch et al. 2014). However, the evaluation of the WRF physics options for heatwave conditions within urban areas in Australia has not been explicitly carried out. The study conducted by Evans et al. (2012) found that no single ensemble member showed best performance for all heavy rainfall events and all variables. By using a standardized super-metric to quantify the influence of one parameterization over another, their study suggested using the Mellor–Yamada–Janjic (MYJ) PBL scheme, the Betts Miller Janjic (BMJ) cumulus scheme (CU), the Dudhia short wave radiation (SW) scheme and the Rapid Radiative Transfer Model (RRTM) long wave radiation (LW) scheme as the most robust combination. They also found that the simulation of precipitation was more sensitive to the choice of CU scheme. Both maximum and minimum temperatures

were sensitive to the selection of LW and SW scheme while mean sea level pressure (MSLP) and wind speed were sensitive to both PBL and CU schemes. The overall conclusion of Evans et al. (2012) was that no single WRF model configuration performed best for all case studies for all variables and this is in agreement with previous studies carried out by Jankov et al. (2005) in the USA.

Jankov et al. (2005) investigated a WRF ensemble that consisted of 18 configurations including three two PBL schemes, three microphysics (MP) schemes and three CU schemes to simulate mesoscale convective system rainfall. They found that CU scheme was the most sensitive, PBL scheme was less sensitive and the MP scheme was the least sensitive. Kala et al. (2015) carried out a sensitivity study of WRF for the southwest of Western Australia over a seasonal timescale and found that both precipitation and temperature simulation were sensitive to the choice of LW and SW radiation scheme while PBL scheme had a stronger influence on minimum temperatures. A multi-physics WRF ensemble was investigated by Stegehuis et al. (2015) for simulating mega heatwaves in Europe. Their study found that precipitation was overestimated and temperature was underestimated by the WRF model. It was also concluded that the choice of CU scheme had the most significant impact in simulating temperature. Their study found that the WRF-single moment class 6 (WSM6) and Morrison MP, Yonsei University (YSU), Asymmetric Convective Model version 2 (ACM2) and Mellor–Yamada–Nakanishi–Niino (MYNN) PBL, RRTMG SW and LW radiation, Tiedtke and Grell–Devenyi CU schemes showed best performance in simulating heatwaves in Europe. On the other hand, the Quasi-Normal Scale Elimination (QNSE) and MYJ PBL, Community Atmospheric Model (CAM) SW and LW radiation, Kain–Fritsch CU schemes were the worst performing physics options for their study area. Furthermore, their results confirmed that temperature simulations were sensitive to soil moisture, which was, in turn, controlled by land surface scheme. Hu et al. (2010) evaluated three PBL schemes namely, YSU, MYJ and ACM2 for air quality simulations over Texas. The WRF simulations resulted in largest negative bias (underestimation) when using MYJ scheme. Both the YSU and ACM2 schemes resulted in smaller biases, and led to simulations of lower moisture and higher temperature in the lower boundary layer due to their strong vertical mixing.

In summary, the WRF model has numerous physics options and can be operated using different configurations which, in turn, influences the model results based on different climatic conditions and geographic locations (Bukovsky and Karoly 2009; Evans et al. 2012; Kala et al. 2015). While several multi-physics WRF studies have been conducted over Australia (Evans et al. 2012; Kala et al. 2015), there is currently a lack of information on the sensitivity of

WRF to different physics options during heatwave events, and additionally, the influence of the UHI on heatwave events is not well documented. This study uses a multi-physics ensemble consisting of 27 model configurations including three PBL schemes, three microphysics schemes (MP), and three short (SW) and long wave (LW) radiation schemes, to investigate the sensitivity of WRF parameterizations during heatwave events. The study focuses on four mega heatwave events in four summer seasons during the 2000–2009 period in the city of Melbourne, in southeast Australia. The aim of this study is to evaluate the WRF model over the Melbourne region and identifying systematic biases and areas of uncertainties and relate them to the underlying physical processes. Additionally, simulations are carried out with and without the inclusion of the effects of the Urban Canopy Model (UCM), as well as different land surface models, to better understand the role of the urban canopy and land surface processes during heatwaves. This study will help better inform WRF users when using the model to investigate heatwave dynamics, and it is the first initiative of a larger research project, which aims at assessing the effectiveness of different UHI mitigation strategies in reducing heatwave intensity for the Melbourne metropolitan region.

2 Selection of heatwave events

This study defines a heatwave as an event of extreme hot temperature lasting at least 3 consecutive days above the 95th percentile, following the definition by Nairn and Fawcett (2013). A case-study approach is used as it is a method that has been adopted by several studies focusing on European heatwaves (e.g., Stegehuis et al. 2013; Miralles et al. 2014). The present study selected four most severe heatwave events from different summer seasons during the time period of 2000–2009, similar to Evans et al. (2012), who focused on mega rainfall events. The four heatwaves were chosen from the years 2000, 2006, 2007 and 2009 following the definition by Nairn and Fawcett (2013). During these events, minimum and maximum temperatures of 37 and 45 °C were recorded respectively. The heatwave event in 2009 was exceptional and showed the strongest intensity among these four heatwave events. During mid-January to early February in 2009, the daily maximum temperatures were higher by up to 1–3 °C after droughts in the state of Victoria in southeast Australia, when winds were northerly (Nicholls and Larsen 2011). From the 26th to the 30th of January 2009, a dominant surface ridge formed extending from the Indian Ocean into the Tasman Sea, with a large heat low and trough covering the western half of the continent where the depth and dominance of the long wave ridge signified a stationary Rossby wave (Nairn and

Fawcett 2013). Furthermore, the dry antecedent conditions and resulting low soil moisture content across southern Australia were strong contributors to the intensity of day and night extreme temperatures during this extreme heat-wave event.

3 Description of WRF model configurations

3.1 Model domains and initialization

The Advanced Research WRF-ARW model (version 3.6.1) was used for this study. Three 2-way nested model domains with spatial grid resolutions of 18, 6 and 2 km were used as illustrated in Fig. 1a. Furthermore, Fig. 1b, c show the location of weather stations used for model evaluation, and the land use categories in the WRF model, respectively. The first nested domain at 6 km resolution covers whole Victorian state while the 2 km domain covers the Melbourne metropolitan area, surrounding suburbs and as well as rural areas. All the domains were centered at -37.81°S and 144.96°E consisting of 38 vertical levels spaced closer together close to the surface to ensure higher vertical resolution within the PBL. The United States Geological Survey (USGS) global topography, land use and land–water masks datasets were used with the spatial resolution $5'$, $2'$ and $30''$ arcsec for domains D01, D02 and D03, respectively. Initial and boundary conditions were derived from ERA-interim re-analysis with $0.75 \times 0.75^{\circ}$ spatial resolution and temporal resolution of 6 h available from 1970 onwards (Dee et al. 2011). ERA-interim re-analysis (product from European Centre for Medium-Range Weather

Forecasts, ECMWF) was chosen over other products as it has shown better performance with comparatively smaller biases than NCEP-FNL (National Centers for Environmental Prediction-Final) and NNRP (product from National Centre for Atmospheric Research, NCAR) in simulating rainfall and temperature over southwest Western Australia (Kala et al. 2015). Finally, the simulations were initialized from 12:00 UTC 1st February to 12:00 UTC 4th February 2000; 12:00 UTC 19th January to 12:00 22nd January 2006; 12:00 UTC 15th February to 12:00 UTC 18th February 2007; and 12:00 UTC 27th January to 12:00 UTC 30th January 2009.

3.2 WRF ensemble design

The WRF model has multi-physics options ranging from simple to sophisticated, newly developed to well-tested, and more computational cost to less computational cost, which can be used to design different model configurations. This model offers various physics options under each physics scheme to obtain the optimal model configuration for different study purposes in different geographical locations. The common schemes for different physics parameterization that are commonly tested include the: (1) PBL scheme, (2) SW and LW radiation schemes, (3) MP scheme, (4) land surface scheme and (5) CU scheme. The present study tests 27 WRF model configurations (Table 1) including three PBL schemes (MYJ, ACM2, QNSE), three MP schemes (WDM5, WSM6, Thompson), three SW radiation schemes (Dudhia, Goddard, RRTMG) and three LW radiation schemes (RRTM, Goddard, RRTMG). All these simulations use the Noah land surface scheme coupled with

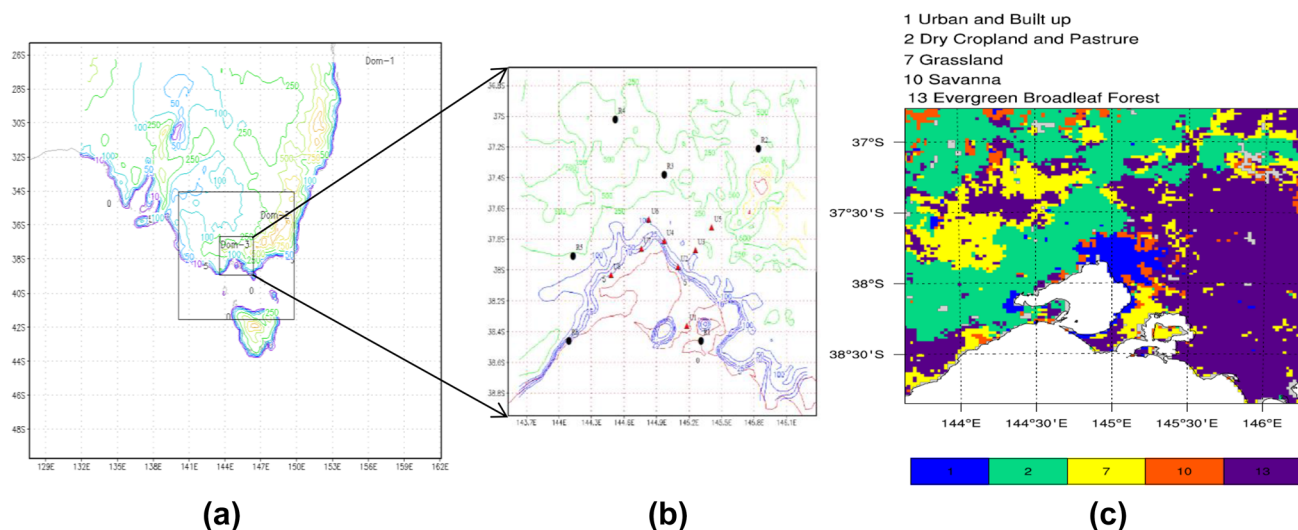


Fig. 1 a WRF model domains, b the locations of weather stations in the urban (red triangles) and rural (black circles) areas. c Land use in domain (D03)

Table 1 WRF ensemble design

Ensemble no.	Micro-physics	Ra_Sw/Ra_Lw physics	Surface layer physics	PBL physics	Land surface physics
1	WDM 5	Dudhia/RRTM	Eta similarity	MYJ	Coupled Noah-LSM-UCM
2	WSM 6				
3	Thompson				
4	WDM 5	Goddard/Goddard			
5	WSM 6				
6	Thompson				
7	WDM 5	RRTMG/RRTMG			
8	WSM 6				
9	Thompson				
10	WDM 5	Dudhia/RRTM	MM5 similarity	ACM2	
11	WSM 6				
12	Thompson				
13	WDM 5	Goddard/Goddard			
14	WSM 6				
15	Thompson				
16	WDM 5	RRTMG/RRTMG			
17	WSM 6				
18	Thompson				
19	WDM 5	Dudhia/RRTM	QNSE similarity	QNSE	
20	WSM 6				
21	Thompson				
22	WDM 5	Goddard/Goddard			
23	WSM 6				
24	Thompson				
25	WDM 5	RRTMG/RRTMG			
26	WSM 6				
27	Thompson				

Urban Canopy Model (UCM). These physics option were selected based on the recommendation of previous studies (Hall et al. 2005; Evans et al. 2012) and the WRF-ARW user's guide (User's Guide WRF-ARW 2015). A total of 108 simulations (27 simulations for each event following the configurations shown in Table 1) were carried out for the four case studies. The first 12 h of the simulations were considered as spin-up time, following previous studies (Giannaros et al. 2013; Hu et al. 2013; Salamanca et al. 2011) and the remaining 60 h were used for detailed statistical analyses. The following section describes the rationale for the choice of different parameterizations tested in this study.

3.2.1 Planetary boundary layer (PBL) scheme

The PBL scheme simulates the tendencies of temperature, moisture and horizontal momentum in the whole atmospheric column by parameterizing the unresolved turbulent vertical flux profiles within the boundary layer. The PBL

scheme computes the PBL fluxes at the sub-grid scale due to eddy transport in the atmosphere. The different PBL schemes consider different assumptions for energy, moisture and mass transformation which may influence model performance within the boundary layer and consequently, overall model performance. The YSU (non-local) and MYJ (local) PBL schemes have been extensively investigated specially for rainfall simulations over the south-east Australia by Evans and Boyer-Souchet (2012). The evaluation of the YSU and ACM2 PBL schemes for temperature and rainfall simulations was carried out by Kala et al. (2015) over the south-west of Western Australia. As per authors' knowledge, there is no comprehensive study that evaluates new and old PBL schemes to simulate the extreme hot temperature weather events with high spatial resolution (<5 km). Furthermore, PBL schemes evolution results will provide helpful information on future regional climate research. Previous studies suggest using the MYJ PBL scheme for the region of south-east Australia (Evans and Boyer-Souchet 2012) and the YSU scheme for south-west

of Western Australia (Kala et al. 2015). Therefore, the present study will examine the MYJ PBL scheme and two additional schemes such as the new QNSE scheme (Sukoriansky et al. 2005), a local-closure model; and the ACM2 scheme (Pleim 2007), which is a hybrid local/non-local scheme.

3.2.2 Radiation scheme

Solar radiation is one of the primary drivers of PBL dynamics, and radiation schemes provide the temperature tendencies in the entire atmosphere by resolving the radiative heat fluxes. Radiation schemes determine total radiative fluxes at any given location due to SW and LW radiative flux divergence. Radiation schemes have been shown to have a strong influence on temperature and precipitation simulations (Borge et al. 2008; Kala et al. 2015). The current study tests three SW radiation schemes, namely, the Dudhia (1989), Goddard (Chou and Suarez 1999) and RRTM for application to GCMs (RRTMG) (Iacono et al. 2008). The other three LW schemes are RRTM (Mlawer et al. 1997), Goddard, and RRTMG. These pairs of short and long wave schemes are set up as Dudhia+RRTM, Goddard+Goddard and RRTMG+RRTMG. The RRTMG long wave scheme uses the correlated-k approach to estimate heating rates and long wave fluxes for application to GCMs (Mlawer et al. 1997). When the WRF model runs using the RRTMG scheme for finer resolution, it takes up the entire grid space rather than using the Monte Carlo independent column approximation method of random cloud overlap to resolve sub-grid scale cloud variability. The Goddard scheme resolves explicit interactions with microphysical processes, which is important for high-resolution WRF simulations (Chou and Suarez 2001).

3.2.3 Micro-physics scheme

The WRF model has a range of micro-physics options ranging from simple to the more complex 3–6 class schemes. The micro-physics scheme explicitly resolves water, water vapor, cloud, and precipitation processes. The present study evaluates three microphysics schemes: WRF Double Moment 5-class (WDM5), WRF Single Moment 6-class (WSM6) and Thompson scheme. The WDM5 scheme was recommended by Evans and Boyer-Souchet (2012) over south-east Australia and the remaining two complex schemes were chosen according to the rule of thumb that higher resolution domains require more complex micro-physics (User's Guide WRF-ARW 2015). The WDM5 is a relatively sophisticated scheme that allows double-moment cloud, rain and cloud condensation nuclei for warm process. The WSM6 scheme is more complex scheme developed based on revised ice-microphysics. This scheme

behaves more realistically in response to the appropriate grid-resolvable force (vertical velocity) that increases as the effective grid size decreases (Hong and Lim 2006). The Thompson MP scheme has recently been improved by including a bulk microphysical parameterization and as well as new dependence on aerosol concentration (Skamarock et al. 2008). This scheme also includes parameterization for calculating the direct radiative effects from aerosols considering urban, continental, maritime, mineral dust and sea salt. The Thompson scheme has been extensively tested for mid-latitudes and suggested to use for better performance in simulating various climate variables (Hall et al. 2005).

3.2.4 Land surface scheme

The land surface scheme plays a vital role in climate modeling as it can influence climate from regional to global scale, and days to millennia time scale (Pitman 2003). One of the main roles of land surface scheme is to partition available energy into sensible and latent heat flux. The land surface scheme also combines atmospheric information from surface layer physics with land surface properties to evaluate the vertical transport in the PBL scheme, which has a direct influence on the PBL height estimation (Han et al. 2008). Several land surface schemes are supported by the WRF model, however, primarily this study uses only the Noah land surface model (Noah-LSM) coupled with a single layer UCM to test the influence of the difference PBL, radiation and microphysics options as described in the previous sections. The Noah-LSM is the most commonly used LSM in WRF due to its notably good performance (Stegehuis et al. 2015). The UCM considers two dimensional and symmetrical street canyons with simplified building geometry that helps for better representation of surface energy balance in urban areas. The UCM is more suitable for estimating not only fluxes from roof, wall, and road surfaces, but also temperatures (Tewari et al. 2007). The single layer UCM explicitly parameterizes in-canyon radiation exchange, turbulence exchanges of heat, moisture and momentum between atmosphere and urban face and substrate heat conduction (Chen et al. 2011). The radiation effect is parameterized by albedo, sky view factor and emissivity at the artificial urban surfaces (e.g. walls, road). Temperatures of these artificial surfaces are calculated by solving the thermal conduction equations (Lee et al. 2011). Both surface and canopy air temperature used to estimate the surface sensible flux at each facet. The urban canopy air temperature is calculated based on a local thermal equilibrium assumption. Furthermore, the canyon wind speed used to calculate the sensible heat fluxes is estimated by a combination of logarithmic profile (above mean building height) and exponential profile within the canyon (Lee et al.

2011). The radiation exchanges and turbulent momentum are calculated by using the Monin–Obukhov similarity theory. The Noah-LSM calculates surface fluxes and temperatures from natural vegetated surfaces while the UCM calculates fluxes and temperatures from artificial surfaces (e.g. road, concrete) in each grid cell of a domain area. Surface temperature is calculated based on the mean values of natural and artificial surfaces temperature and weighted based on their areal coverage (Chen et al. 2011). Several other land surface schemes are available in WRF, but these are generally not recommended when compared to the Noah-LSM. For example, Mooney et al. (2013) found that the Rapid Update Cycle LSM (RUC-LSM) showed poorer performance and higher temperature bias than the Noah-LSM. The Pleim–Xiu scheme is more suitable for air quality simulation (Gilliam and Pleim 2010) while the five-layer thermal diffusion scheme is not appropriate in situations where land–atmosphere feedbacks might be important, as it does not explicitly solve for soil moisture (Stegehuis et al. 2015). Another LSM option is the RUC LSM, however, Kala et al. (2015) found large biases with the RUC LSM over southwest Western Australia.

However, recent advances in WRF now include the more complex Noah-Multi Physics (Noah-MP) and Community Land Model version 4 (CLM4) LSM, which have not been extensively evaluated in the literature as some of the older LSM options as described in the previous paragraph. Hence, after testing for the best WRF configuration in terms of PBL, radiation and micro-physics schemes, two additional WRF configurations were tested with the Noah-MP and CLM4 LSMs. The Noah-MP permits multiple choices to parameterize different land, environmental, and hydrological processes. This scheme has four soil, three snow and one canopy layers including sub-grid option to allow for gaps in the vegetation canopy. Additionally, the Noah-MP considers the soil moisture-groundwater interactions, runoff and vegetation phenology. The default Noah-MP options from the WRF user guide were picked for this study. The CLM4 consists of five sub-grid land cover types including lake, wetland, glacier, urban, and vegetation where vegetated sub-grid consists of up to four plant functional types. Each type has a specific canopy height and index of leaf and stem area. The CLM4 vertical structure consists of a single-layer vegetation canopy, a five-layer snowpack and a ten layer soil-column unevenly spaced between the top layer (0.0–1.8 cm) and bottom layer (229.6–380.2 cm).

3.2.5 Cumulus scheme

Cumulus schemes are necessary to parameterize convection for grid resolutions between 5 and 10 km or coarser resolutions (Skamarock et al. 2005). The current study considered

only the innermost domain (D03) for detailed sensitivity analysis of the WRF model, which had 2 km horizontal grid resolution. The cumulus scheme was not used for the innermost domain as convection can be explicitly resolved at this resolution. The Grell3D scheme was used for the outer two domains. According to the recommendation by the WRF-ARW users' guide, the Grell3D is more suitable using for high-resolution simulation. The advantage of the scheme is that it spreads subsidence to neighboring columns, which makes it more suitable for resolutions less than 10 km.

4 Model evaluation

4.1 Observed data

Climate data for 14 weather stations were obtained from Bureau of Meteorology (BoM) of Australia. The locations of all 14 weather stations are shown in Fig. 1b. The weather monitoring stations belonged to two different networks: (1) eight stations in the urban areas (red triangles), and (2) six stations in rural areas (black circles). The weather stations in the urban area mainly covered the Melbourne metropolitan area including the central business district (CBD), major suburbs and airports. The rural stations are situated more than 50 km from the CBD of Melbourne, mainly in rural and forest areas. Atmospheric sounding data at 0000 and 1200 UTC from the Melbourne international airport was obtained from the website of department of atmospheric science of Wyoming University (<http://weather.uwyo.edu/upperair/sounding.html>). Gridded observations of daily maximum temperature at a 1 by 1 km resolution were obtained from the ANUClimate data-set (Hutchinson et al. 2014), available online at: <http://dapds00.nci.org.au/thredds/catalog.html>. This dataset is an interpolation of station observations across the Australian continent and details of the algorithm can be found in Hutchinson et al. (2009). This observed gridded temperatures data was used to compute biases in WRF across the model domain. All the model configurations were evaluated against three variables: temperature at 2 m, relative humidity at 2 m and wind speed at 10 m. Temperature is the key aspect for characterizing heatwave event. Additionally, wind speed (10 m) and relative humidity (2 m) were also considered to evaluate the performance of the WRF model. The rationale for evaluating wind speed and humidity is that small-scale variations can have a large impact on UHI intensity and its effects (Li and Bou-Zeid 2013). Moreover, the performance of the WRF ensemble was tested in both urban and rural areas. The weather stations were selected in both urban and rural areas based on hourly observed data (temperature,

wind speed, relative humidity) availability and geophysical conditions.

4.2 Statistics for model evaluation

Model performance was evaluated against the field observation data using mean bias (MB), mean absolute error (MAE), root mean squared error (RMSE) and correlation coefficient (CC) as shown in Eqs. 1–4. These statistical measures have been used in many studies (Borge et al. 2008; Evans and Boyer-Souchet 2012; Kala et al. 2015) and also recommended by a number of studies for quantifying model performance (Emery et al. 2001; Gilliam et al. 2006; Russell and Dennis 2000). Studies by Willmott et al. (2009), Chatterjee et al. (2013) and Jerez et al. (2012) have suggested that MAE is better than RMSE to test the quantitative performance of models. The reason behind this argument is that the RMSE is a function of number of errors, which changes the distribution of errors. The RMSE also represents the magnitude of an average error that creates more complexity for the interpretation of model performance. A combination of various statistical metrics has gained more acceptability for assessing the climate model performance rather than single performance measures (Chai and Draxler 2014). The lower magnitude of MB, MAE and RMSE values indicate better performance of the model. Linear correlation between model and observation is quantified by CC. These CC values range between -1 and $+1$, where zero value indicates no correlation. The MB value measures the average error between model and observations while positive (+Ve) and negative (–Ve) values show the tendency for over-estimation and under-estimation by the model, respectively. The MAE measures the gross error of model-simulated results.

$$MB = \frac{\sum_{i=1}^N (x_{sim,i} - x_{obs,i})}{N} \quad (1)$$

$$MAE = \frac{\sum_{i=1}^N |x_{sim,i} - x_{obs,i}|}{N} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (x_{sim,i} - x_{obs,i})^2}{N}} \quad (3)$$

$$PCC = \frac{\sum_{i=1}^N (x_{obs,i} - \overline{x_{obs}}) (x_{sim,i} - \overline{x_{sim}})}{\sqrt{\sum_{i=1}^N (x_{obs,i} - \overline{x_{obs}})^2} \sqrt{\sum_{i=1}^N (x_{sim,i} - \overline{x_{sim}})^2}} \quad (4)$$

where N indicates the total number of comparisons. The X_{sim} and X_{obs} are the observation and simulated/modelled values, respectively.

5 Results and discussion

5.1 Temperature

Taylor diagrams for temperature are illustrated in Fig. 2 for both urban and rural areas. The PBL schemes are represented by color: MYJ (red), ACM2 (blue), QNSE (green); radiation schemes combinations are indicated by shapes: Dudhia+RRTM (square), Goddard+Goddard (circle), RRTMG+RRTMG (triangle); microphysics schemes are illustrated by different filling symbols: WDM5 (smaller hollow), WSM6 (filled), Thompson (larger hollow). Taylor diagrams represent a way of graphically summarizing three different statistics about how closely model results match with observations. The similarity between model and observation is quantified by using their standard deviations (relative variance), centered root mean square (RMS) differences, and correlation coefficients (CC) (Taylor 2001). The horizontal and vertical axes indicate the standard deviation of the model, which is proportional to the radial distance from origin while the centered RMS difference between model and observation is proportional to the distance from a point (REF) on X-axis. The arc of the diagram indicates the temporal correlation coefficient (CC) pattern between model and observation. Therefore, a perfect model should lie near the X-axis and close to the observation arc.

Figure 2 shows that all simulations had a high pattern correlation (0.90–0.99) with observations for hourly temperatures in both the urban and rural areas. The amplitudes of relative variability (normalized standard deviation) were less than one in both urban and rural areas except for the event-1 in the urban areas. The PBL (different colors) and radiation schemes (different shapes) showed a larger influence in simulating temperature than the microphysics schemes (different filling). The ACM2 PBL scheme (blue) showed comparatively good performance in terms of correlation, RMS difference and relative variance in both urban and rural locations for the event-1 (Fig. 2a, e). The MYJ scheme (red) showed better performance for event-2 in the urban areas (Fig. 2b) and for events-2, 3 and 4 (Fig. 2f–h) in rural areas. The QNSE PBL scheme (green) showed good performance in urban area for the event-4 (Fig. 2d). In addition, the MYJ and QNSE schemes showed mixed performance for event-3 in the urban area (Fig. 2c). The MYJ scheme (red) showed comparatively lower variances in most cases while the ACM2 scheme (blue) showed higher RMS differences. The RRTMG SW and LW radiation schemes (rectangle) showed comparatively better performance in both urban and rural areas in four case studies while the Goddard scheme (circle) showed the poorest performance. Finally, all three microphysics schemes (different filling) showed mixed performance for all case studies in both urban and rural areas.

Fig. 2 Taylor diagrams for temperature for the four heatwave events. All 27 configurations are shown from **a–g**. PBL schemes represented by *color*: MYJ (*red*), ACM2 (*blue*), QNSE (*green*); radiation schemes represented by *shapes*: Dudhia + RRTM (*square*), Goddard + Goddard (*circle*), RRTMG + RRTMG (*triangle*); MP schemes represented by *filling symbols*: WDM5 (*small hollow*), WSM6 (*filled*), Thompson (*bigger hollow*)

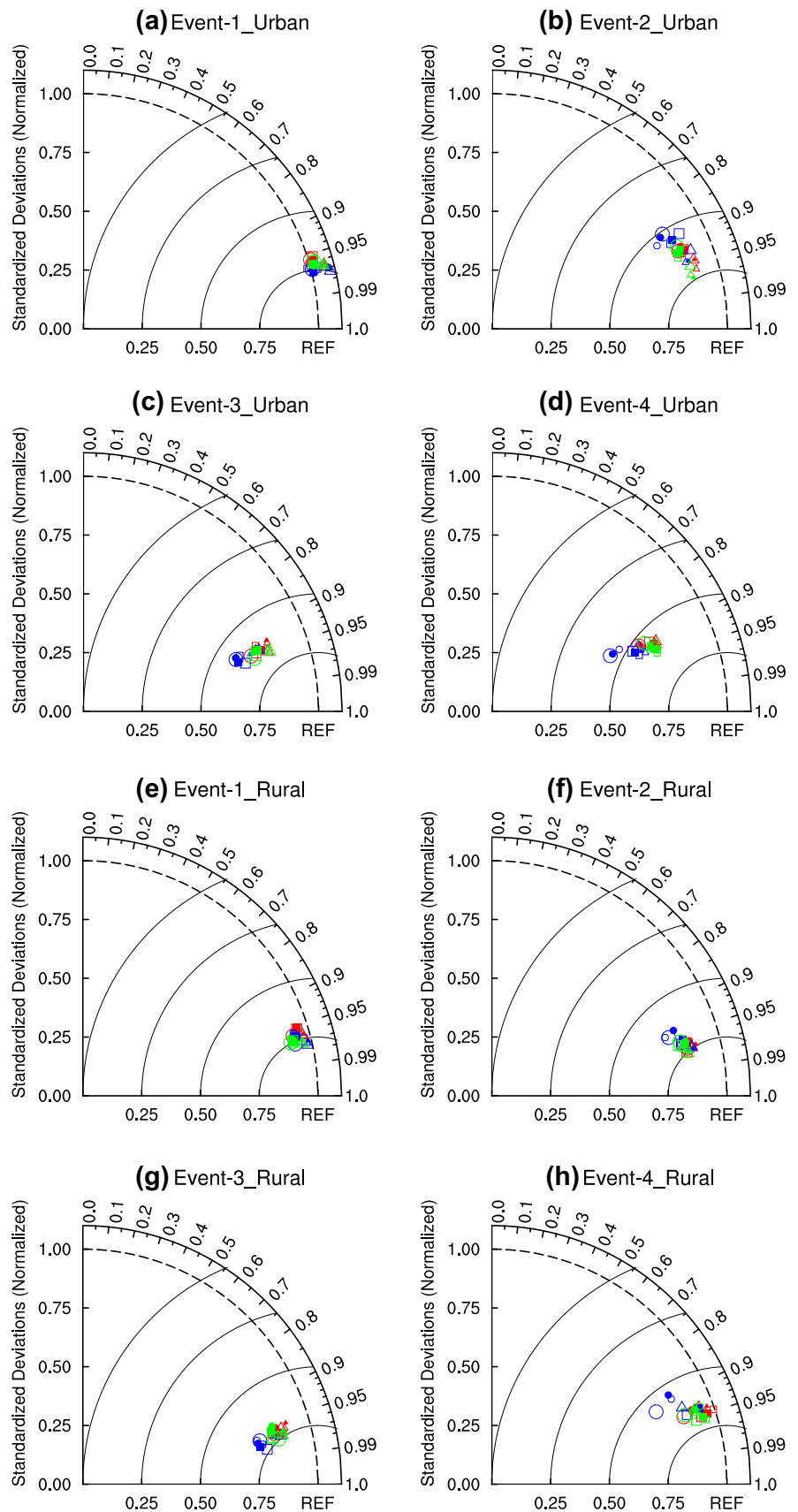


Figure 3 shows the MB, MAE and RMSE for temperature for all 27 ensemble simulations for the four events. There were large differences in performance between the WRF ensemble members, especially for event-4. Although some ensemble members showed large differences in terms of MB, MAE and RMSE, the WRF model showed an acceptable behavior overall. This finding supports the view that all the WRF model configurations represented major processes governing the near surface temperature reasonably well. The MB results indicated that the MYJ scheme (Ensemble ID 1–9) underestimated temperatures by 1.5–2.5 °C for events 3 and 4 in rural areas and 1–1.7 °C for event 3 in the urban areas. The MYJ scheme showed a tendency for overestimation (0.1–0.7 °C) in urban the areas and underestimation (<0.5 °C) in rural areas for the remaining events. The ACM2 PBL scheme (Ensemble ID 10–18) showed a tendency of overestimation (0.5–1.5 °C) except event 3 in the urban areas. For rural areas, this PBL scheme underestimated temperatures (0.5–2.5 °C) for events 3 and 4 while it overestimated temperatures with a relatively small bias (~0.5 °C) for events 1 and 2. Furthermore, the QNSE PBL scheme (Ensemble ID 19–27) overestimated

temperatures (~0.5 °C) for events-1 and 2 in urban areas and underestimated for remaining events with larger biases (0.5–3.5 °C) for the remaining cases, in both urban and rural locations. Both the ACM2 and QNSE schemes showed higher MB than the MYJ scheme. The WRF ensembles showed both the overestimation and underestimation tendency in the urban areas while they showed underestimation in rural areas. However, the MYJ scheme showed comparatively better performance especially for the urban areas. In summary, the WRF model showed a tendency of overestimation in the urban areas and underestimation in rural areas.

When considering the MB, MAE and RMSE for all case studies, the MYJ scheme, especially ensemble member 7 and 9, and the ACM2 scheme, particularly ensemble member 18, showed distinctively smaller errors in both urban and areas. The SW and LW radiation schemes combination RRTMG+RRTMG (Ensemble ID 7–9, 16–18 and 25–27) indicated better performance in terms of MB, MAE and RMSE in both urban and rural areas in the most case studies except event-1 in the urban areas. On the other hand, the Dudhia+RRTM combination (Ensemble ID 1–3,

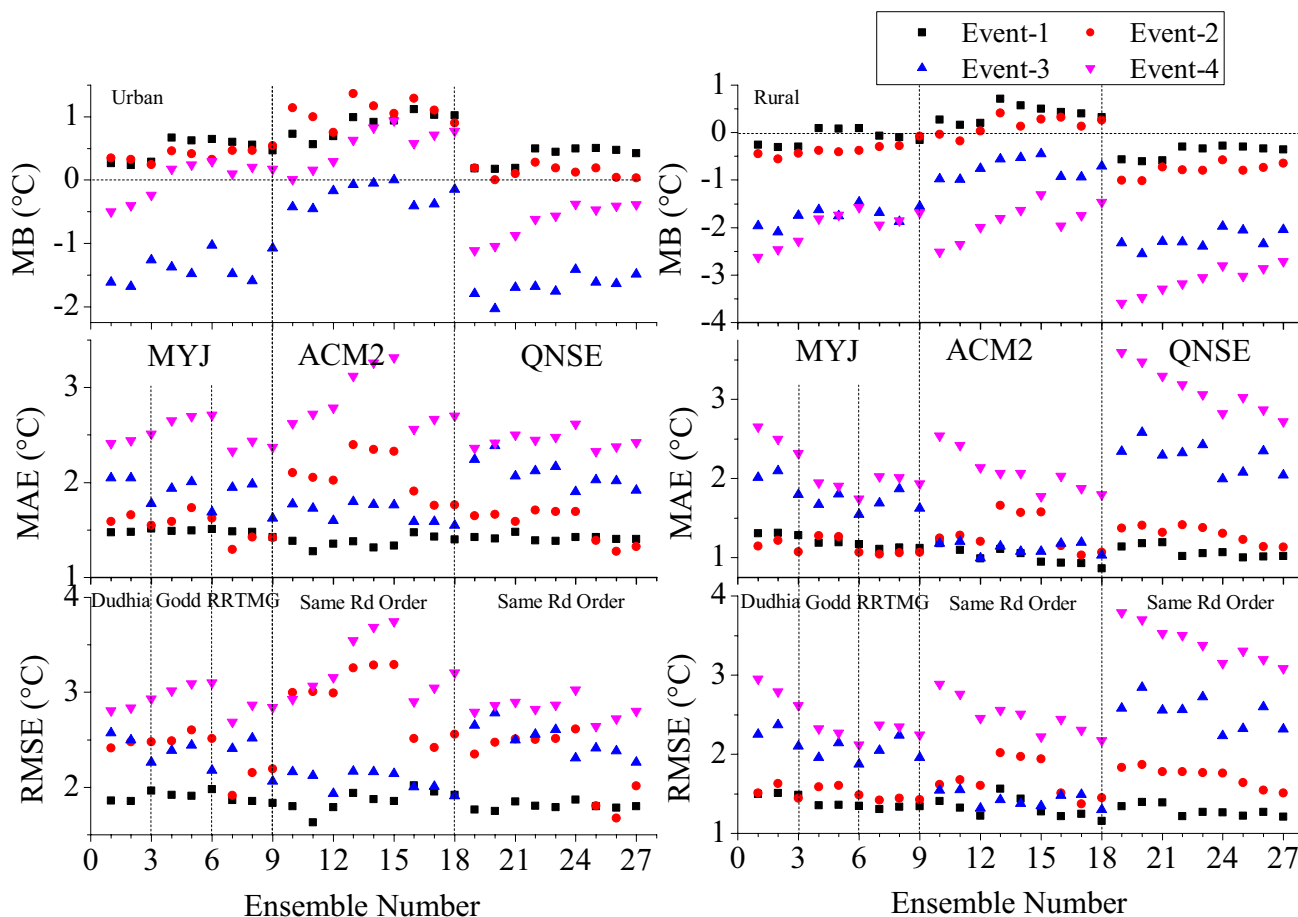


Fig. 3 MB, MAE and RMSE for temperature for the WRF ensemble members for both the urban and rural areas for the four events

10–12 and 19–21) and Goddard+Goddard combination (Ensemble ID 4–6, 13–15 and 22–24), showed poor performance in both urban and rural areas. The MB results indicated that the Dudhia+RRTM combination showed a tendency towards cooler bias, and Goddard+Goddard combination showed a tendency of warmer bias. The RRTMG+RRTMG combination resulted in a warmer bias than Dudhia+RRTM and cooler bias than Goddard+Goddard. Similar results have been found by Zempila et al. (2016). They attributed this to the dependence of horizontal irradiation with solar zenith angle, with larger solar zenith angles leading to an overestimation of global horizontal irradiation. Zempila et al. (2016) showed that this dependence is smaller for the Dudhia scheme and increases up to 30% for the RRTMG and Goddard schemes during clear sky conditions, which could explain the results here. Interestingly, the RRTMG SW and LW radiation schemes when used with the Thompson MP scheme (Ensemble ID 9, 18 and 27) showed much better performance in most cases in terms of MB, MAE and RMSE. The influence of microphysics on temperature was not very clear in most cases. However, the ensemble members 3, 6, 9, 12, 15, 18, 21, 24 and 27 showed comparatively lower MB, MAE and RMSE in the most cases for both urban and rural areas when the Thompson MP scheme was used. This finding is similar to previous studies which have found that the Thompson scheme is better at representing warmer weather conditions (Jankov et al. 2011). This scheme has also been suggested for use in mid-latitudes (Hall et al. 2005). Finally, results show that temperature simulations show the highest sensitivity to the choice of PBL options, lower sensitivity to LW and SW radiation schemes and the least sensitivity to the MP schemes.

Since the PBL schemes played an important role in simulating the heatwave events and ensemble member 9 (MYJ) and 18 (ACM2) showed comparatively better performance, further analysis on model performance in simulating maximum temperatures at 2 m (T2max) by the those WRF configurations is illustrated in Fig. 4 showing biases across the domain when compared with the ANUClimate gridded observational dataset. Additionally, ensemble number 27 is also included to allow comparisons with the QNSE PBL scheme. The ensemble number 18 (ACM2) showed slightly warmer biases (Fig. 4b) than the ensemble member 9 (Fig. 4a) for the events-2 and 3. On the other hand, the ACM2 scheme (ensemble member 18) showed cooler biases for events- 1 and 4 (Fig. 4b). Ensemble number 9 showed the lowest biases in simulating T2max for events-1 and 4 in both urban and rural areas and for event-2 near the coastal as compared to other ensemble members. Ensemble member 27 showed larger cooler biases than the ensemble numbers 9 and 18 except for event-4 (Fig. 4c). Although all three ensemble members showed slightly cooler biases

in most cases, the ensemble members 9 and 18 showed warmer biases near the coastal areas, especially for events-1 and 2. Overall, ensemble number 9 (MYJ+RRTMG/RRTMG+Thompson) showed better performance, especially for the most severe heatwave event (event-4).

The PBL height (PBLH) is an important atmospheric diagnostic, as this height indicates the strength of the turbulent mixing. Although WRF outputs PBL heights, these are not directly comparable between WRF experiments using different PBL schemes, as the latter are based on different definitions, which make them difficult to compare. Therefore, this study uses a generic calculation of PBL heights for all three PBL schemes following the method suggested by Nielsen-Gammon et al. (2008) and Garcia-Diez et al. (2013). According to this method, the PBL height is the first level where potential temperature exceeds minimum potential temperature within the mixed layer by more than 1.5 K. Figure 5 shows the hourly variations of PBLH using this method for the MYJ, ACM2 and QNSE PBL schemes for ensemble members 9, 18 and 27, respectively, in both urban and rural areas. Among the three PBL schemes, the MYJ scheme showed the lowest PBLH and the ACM2 showed the deepest PBLH during the daytime. The lower PBLH simulated by the MYJ schemes suggests less entrainment of free-tropospheric air into the PBL. The ACM2 scheme resulted in consistently deeper PBLH especially during the daytime in all events over both urban and rural areas. Although the MYJ and QNSE schemes produced a similar trend in PBL heights, the QNSE scheme produced slightly deeper PBLH than the MYJ scheme, especially in the urban areas in few cases. In most cases, the PBL heights reached their peak between 1400 and 1900 LST. Figure 5 also illustrates that PBLH sharply raised and collapsed after 1200 and 2000 LST, respectively. The lower prediction of PBLH by the MYJ scheme has been reported to be due to less entrainment of free tropospheric air into the PBL (Hu et al. 2010). In contrast, the ACM2 scheme showed deeper PBLH than the MYJ and QNSE schemes for both urban and rural areas. This finding indicates that the ACM2 has higher strength of entrainment and turbulent mixing.

One approach to investigate the entrainment process in the PBL is the inspection of the potential temperature and moisture profiles (Hu et al. 2010). Figures 6 and 7 illustrate instantaneous temperature and moisture profiles for the 1st day at 2300 LST and the 2nd day at 1100 LST within 3 days simulation period at Melbourne international airport, since the observed atmospheric sounding data (12 h interval) was available for this specific location within the study domain. All three PBL schemes produced slightly lower potential temperature than observed during night-time (2300 LST) in the lower to middle troposphere (Fig. 6). On the other hand, the PBL schemes simulated higher potential temperature

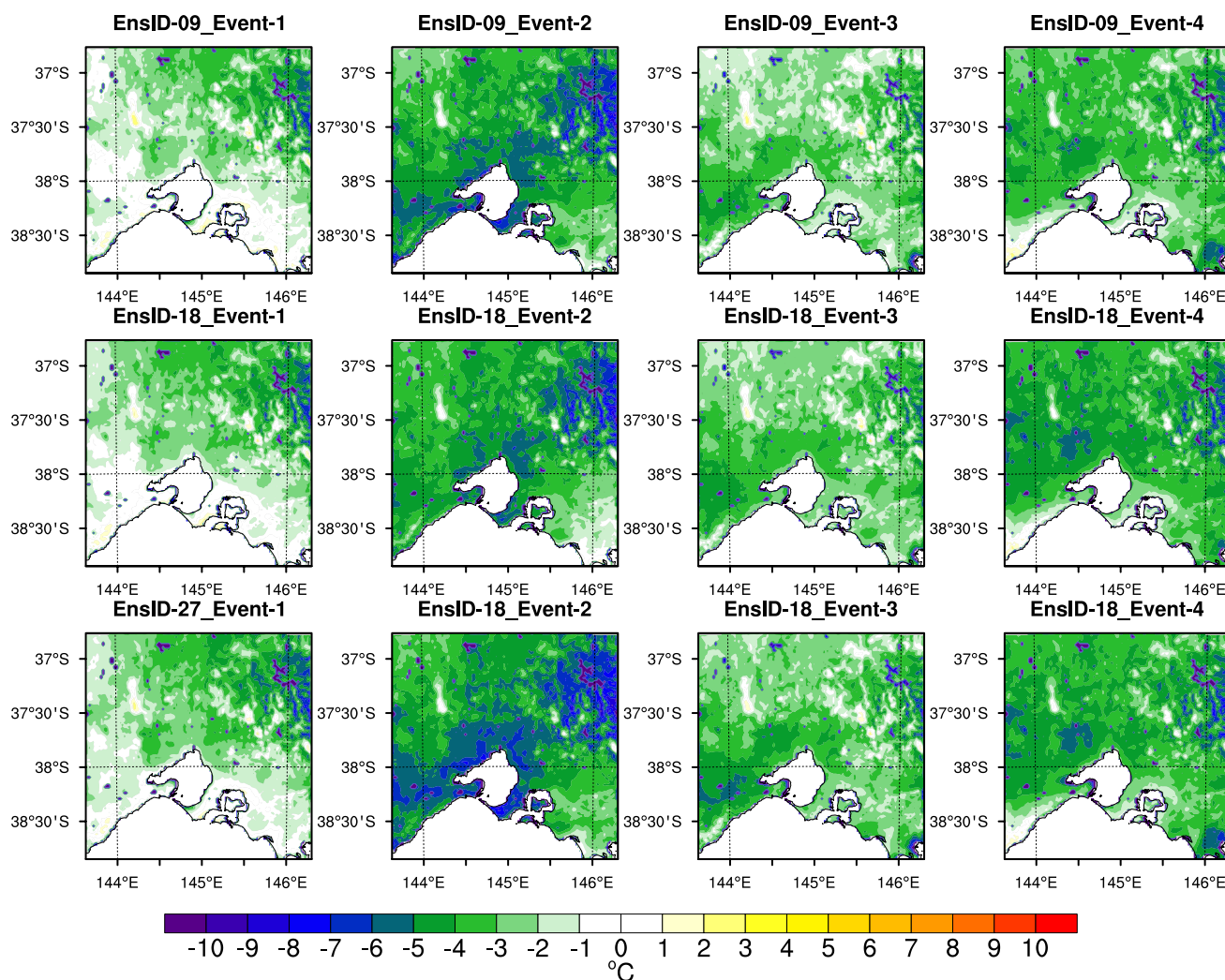


Fig. 4 T2max biases for simulations using the MYJ (ID-09), ACM2 (ID-18) and QNSE (ID-27) PBL schemes

than observed during day-time (1100 LST) especially in the lower troposphere. The moisture profiles showed that all the PBL schemes simulated higher atmospheric moisture during both night-time and day-time except event-2 during day-time (Fig. 7). Although, the PBL schemes showed similar trends in the potential temperature and moisture profiles, important differences can be seen in the vertical structure. For instance, the MYJ scheme simulated lower temperature and higher moisture than the ACM2 scheme in the lower troposphere (below 500 m) for all the four events. This finding suggests that when similar amounts of moisture and heat enter the atmosphere from the land surface, the MYJ scheme lacks sufficient vertical mixing to transport this moisture and heat away from the surface to the top of the PBL as compared to the ACM2, which is consistent with Hu et al. (2010). To obtain a better picture of differences between the different PBL schemes, the potential temperature and moisture profiles are plotted at 1700 LST

of the 3rd day of each event in Fig. 8, to better capture the peak of PBL development, which is well after 1100 LST in this region. These temperature and moisture profiles at 1700 LST show a strong relationship between temperature and moisture in the upper troposphere. Importantly, Fig. 8 shows that the MYJ schemes simulated lower moisture content and higher temperature than the ACM2 scheme in the upper troposphere (above 2700 m) especially for the events-1, 2 and 3.

Additionally, the vertical profiles of the vertical wind component is shown in Fig. 9 for representing the vertical mixing strength of the three PBL schemes for the 2nd day at 1100 LST and the 3rd day at 1700 LST within 3 days simulation period. The positive and negative vertical wind speed indicates upward and downward flow direction, respectively. Figure 9 illustrates that the MYJ and QNSE schemes showed a downward wind flow tendency in most cases, which indicates weaker vertical mixing by these two

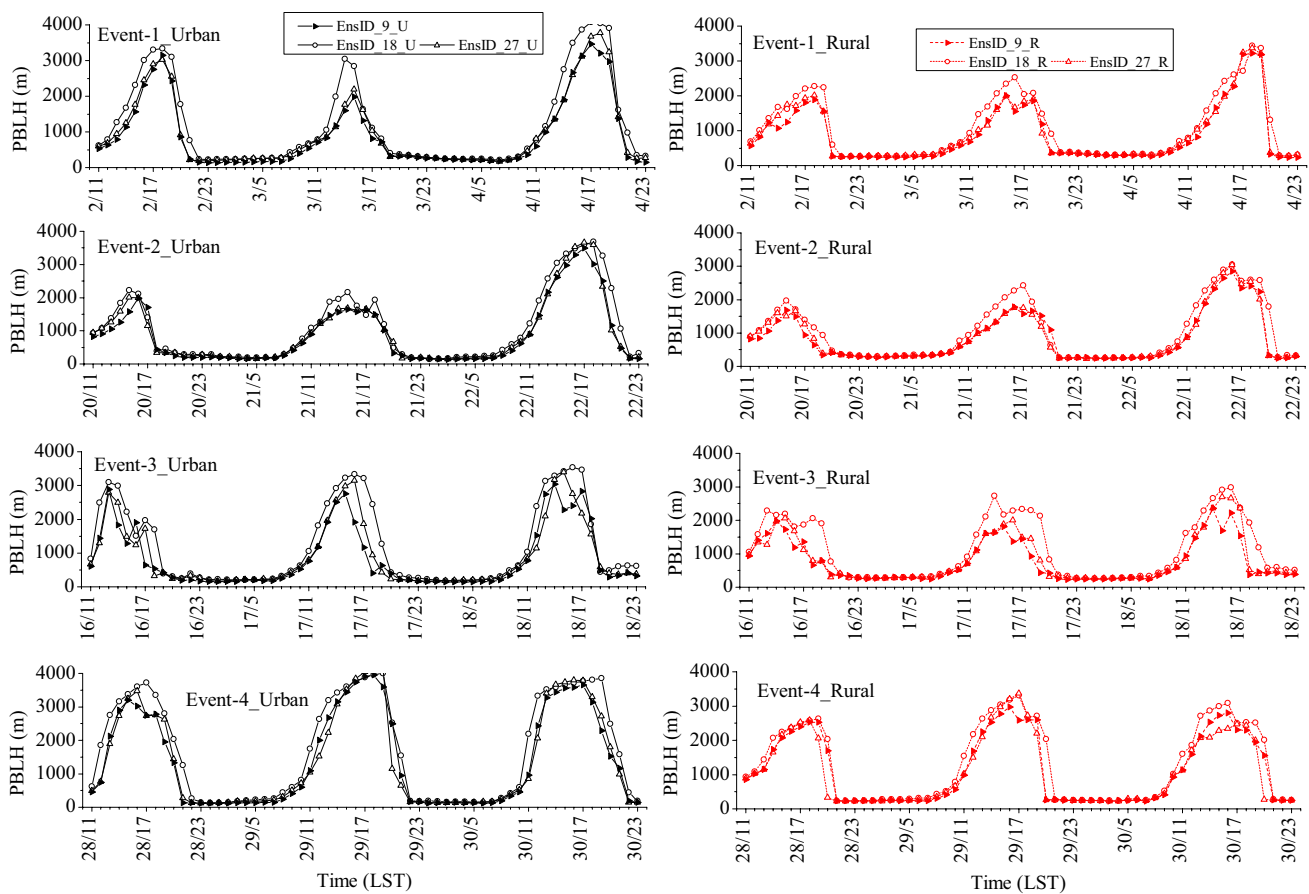


Fig. 5 Temporal variations of PBL heights for the MYJ (EnsID-9), ACM2 (EnsID-18) and QNSE (EnsID-27) PBL schemes

schemes. The ACM2 scheme showed an upward wind flow tendency for the most events during both morning (1100 LST) and afternoon (1700 LST), which indicates stronger vertical mixing by this scheme. This provides a plausible mechanism for the warmer and deeper PBL by the ACM2 PBL scheme, consistent with the previous study conducted by Hu et al. (2010), who obtained stronger vertical mixing with the ACM2 scheme, which encourages stronger entrainment at the top of PBL, and consequently, results in a warmer and dryer PBL. A similar finding has been documented by Srinivas et al. (2007), who found nonlocal scheme transports more moisture away from lower PBL to the top of the PBL. Finally, all three PBL schemes showed less fluctuation among them in simulating temperature and moisture at night in most case studies. The most likely reason is that non-local transport is shutdown in the ACM2 scheme (act as local scheme at night when conditions are stable) and vertical mixing is caused due to eddy diffusion as in the local MYJ and QNSE PBL schemes (Hu et al. 2010).

Based on the statistical analyses for temperature (Figs. 2, 3), the RRTMG SW and LW radiation schemes

showed the best performance in simulating temperatures in this study area. To explore this further, the differences in incoming shortwave radiation (SWDOWN) between the combinations of radiation schemes RRTMG+RRTMG and Goddard+Goddard, and Dudhia+RRTM and RRTMG+RRTMG were calculated for the ensemble members 12, 15 and 18. The differences ranged between -150 and 200 Wm^{-2} for events-1 and 2 in both urban and rural areas, and event-3 in urban areas (Fig. 10). The remaining events showed smaller differences around -50 to 50 Wm^{-2} . The incoming shortwave differences showed that the RRTMG+RRTMG combination lead to higher incoming SW radiation than the Dudhia+RRTM combination, and lower than Goddard+Goddard combination. A similar result was also found from the MB temperature analysis especially for the ensemble members 12, 15 and 18 in Fig. 3. These MB values also indicated that Goddard+Goddard combination showed the warmest bias. The Goddard SW radiation scheme coupled with the Goddard global aerosol transport model includes aerosol (sulfate, dust, organic carbon and black carbon) effects and is known to simulate higher magnitude of SW radiation

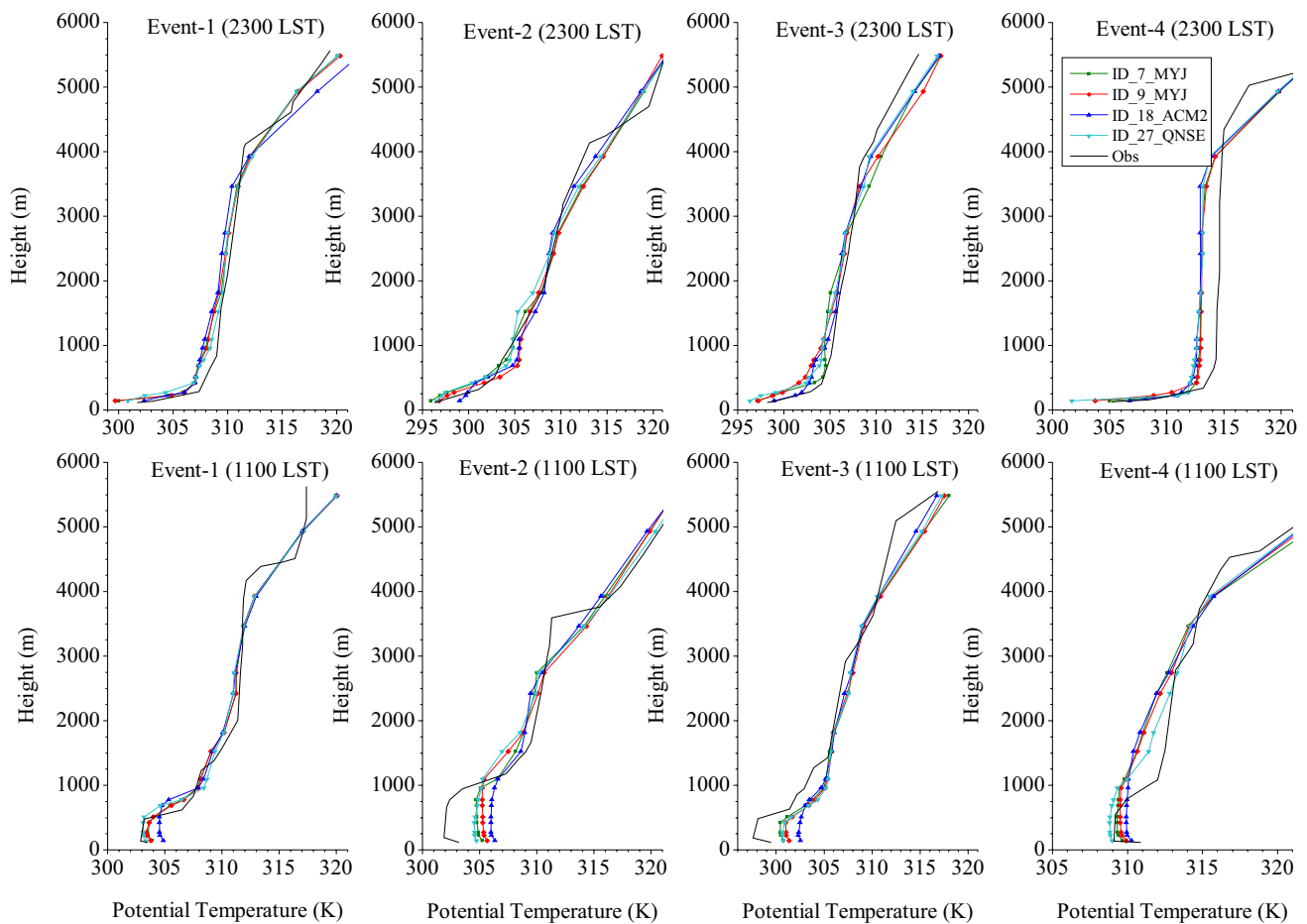


Fig. 6 Simulated and observed temperature profiles at 2300 and 1100 LST for the four events

(Shi et al. 2014). The differences in incoming SW radiation were slightly higher in urban areas than rural areas for the events- 1 and 3 (Fig. 10). Furthermore, both combinations of the Dudhia+RRTM and Goddard+Goddard schemes showed higher incoming SW radiation than the RRTMG+RRTMG combination during the afternoon. Therefore, it can be concluded that the shortwave incoming radiation (SWDOWN) results from the different radiation schemes combination are consistent with the results found from statistical analyses for temperature illustrated in Figs. 2 and 3.

5.2 Temporal biases of near surface temperature during hot weather events

The simulated hourly near surface (2 m) temperature biases in urban and rural areas are shown in Fig. 11. This section mainly focuses on the temporal variability of temperature biases for all four events. In addition, it highlights the temporal variations of the best performing ensemble members 9 (MYJ+RRTMG/RRTMG+Thompson),

18 (ACM2+RRTMG/RRTMG+Thompson) and 27 (QNSE+RRTMG+Thompson) discussed earlier in Sect. 5.1. In Fig. 11, the red, blue and green lines represent the ensemble member 9, 18 and 27, respectively while the grey shading represents the remaining 24 ensemble members. The majority of simulations showed negative biases (underestimation) during the daytime especially in the afternoon and positive biases (overestimation) at nighttime. The biases for event-1 showed that the WRF model simulated 2-m air temperature reasonably well in both urban and rural areas, although some ensemble members (e.g. 10, 11, 20, and 21) showed large discrepancies, especially in the urban areas. There were considerably lower biases and variation in both cases as compared to the other events. However, there was a lower bias and underestimation tendency for the first two days in both urban and rural areas. Interestingly, most ensemble members showed overestimation with higher bias in the 3rd day. The maximum bias occurred when the sign of the gradient changed. The comparison among the most influential ensemble members 9 (MYJ), 18 (ACM2) and 27 (QNSE) shows that the ensemble number

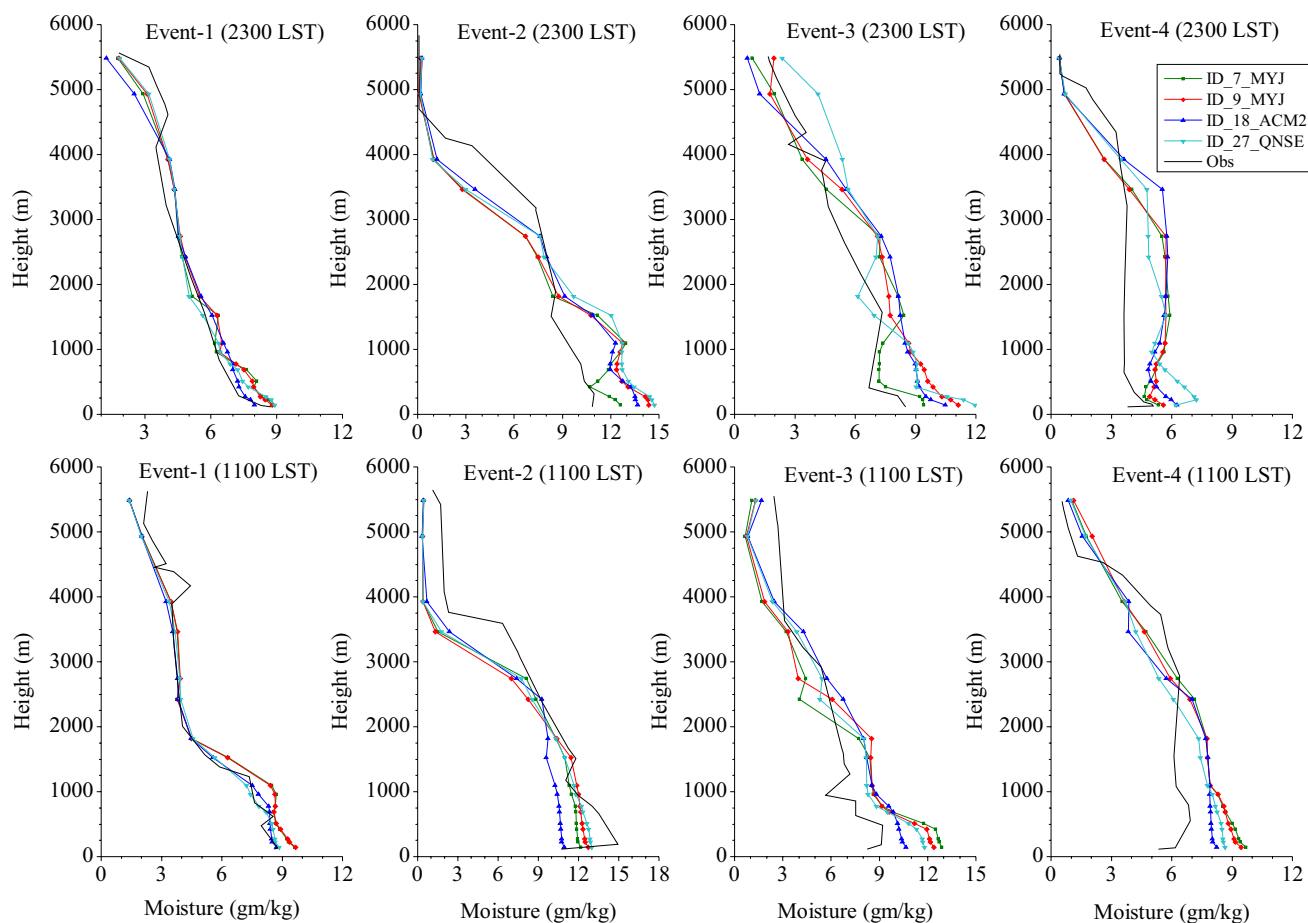


Fig. 7 Same as Fig. 6 except for atmospheric moisture

18 results in lower biases in both urban and rural areas. For event-2, most simulations showed a tendency to overestimate temperatures in urban areas. In rural areas, the model showed a slight over-estimation tendency at night time and under-estimation when temperatures rise to the daily maximum. However, there was an opposite trend when compared to event-1, especially for the urban areas. For this event, no specific ensemble member showed outstanding performance. However, ensemble number 9 (MYJ) resulted in less variability than other ensemble members.

For event-3, the majority of simulations showed that the WRF model underestimated the temperatures for both urban and rural areas except early morning (5.00 A.M.) on 18th February in urban areas. All simulations showed higher bias fluctuations in the urban areas. Moreover, ensemble member 18 produced the least variation in both urban area and rural areas. Finally, the ensemble number 18 (ACM2) showed better performance than ensemble number 9 (MYJ) and 27 (QNSE) in both urban and rural areas. For event-4, all simulations for urban areas showed underestimation during day-time and overestimation during night-time, which was similar to the trend of the simulations

for event-3. For rural areas, all the simulations indicated a general underestimation during both day and night-time. As event-4 was the most severe heatwave event, the temperature fluctuations were also higher than the other events. All the model configurations showed the highest bias fluctuations for the event-4, especially for the urban areas. The ensemble number 9 (MYJ) showed better performance over ensemble number 18 (ACM2) and 27 (QNSE) in both urban and rural areas.

5.3 Wind speed

This section describes the results for wind speed following the same statistical approach described in the previous section for temperature. The Taylor diagrams for wind speed show good temporal pattern correlation ranging from 0.70 to 0.90, except event-3 (Fig. 12c, g). The relative variance was generally greater than 1, indicating that the model simulated higher variability as compared to the observations in both urban and rural areas in most cases. Figure 12a shows that the MYJ and ACM2 schemes showed mixed performance, with the MYJ showing lower RMSD and higher

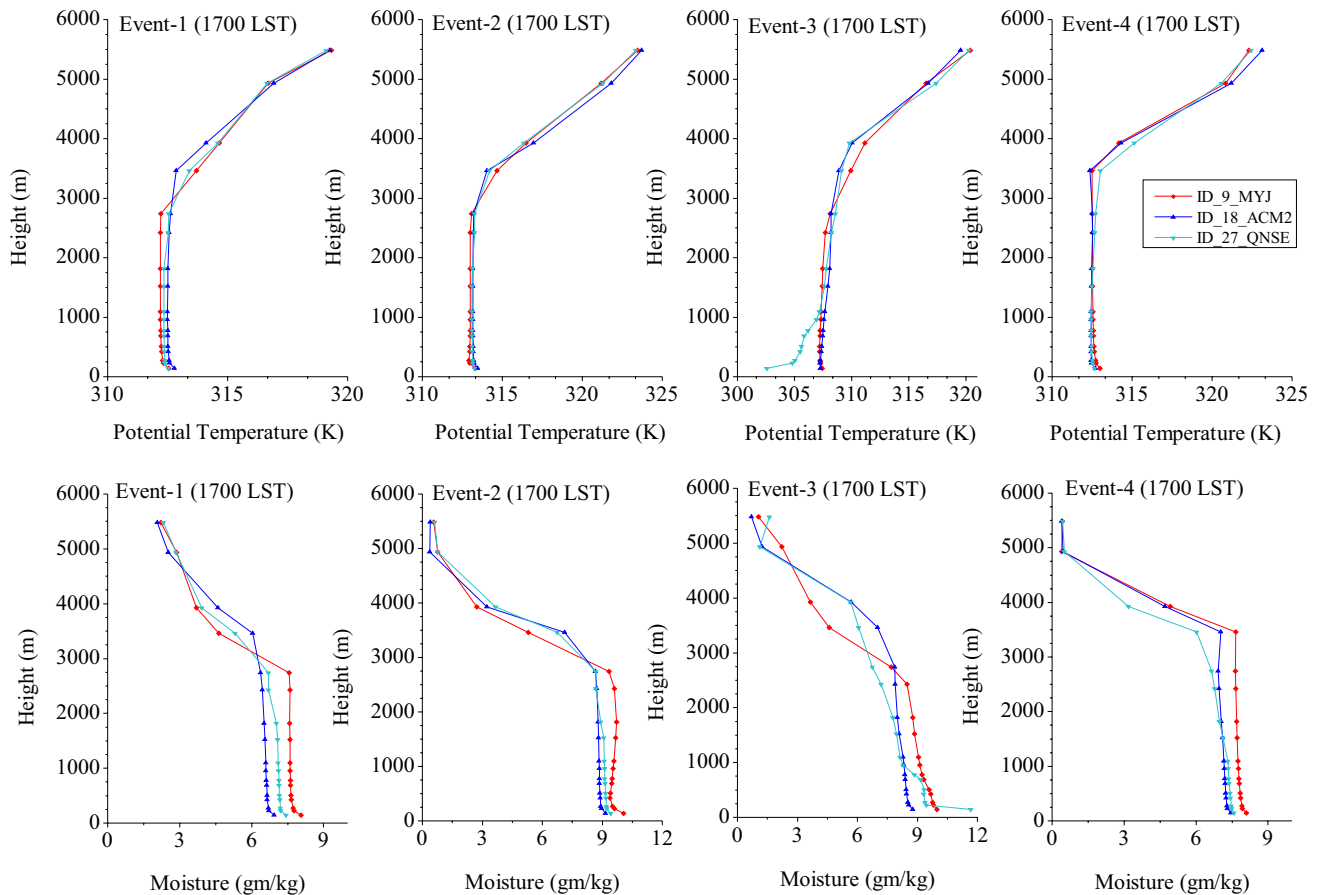


Fig. 8 Simulated temperature and moisture profiles at 17.00 LST for the four events

pattern correlation and standard deviation than the ACM2. The MYJ and ACM2 schemes also showed mixed performance for events-2 and 4 in the urban areas (Fig. 12b, d). The ACM2 performed well for events-1, 3 and 4 in rural areas (Fig. 12e, g, h). Furthermore, the MYJ and QNSE showed better performance for event-3 in the urban areas and event-2 in the rural areas, respectively. Finally, the ACM2 scheme showed slightly better performance than the MYJ and QNSE schemes for both urban and rural areas for four case studies. Although the Goddard radiation scheme showed better performance in simulating wind speed for event-4 (Fig. 12d, h) in both urban and rural areas, and for event-3 (Fig. 12g) in rural areas, the RRTMG showed comparatively better performance in both urban and rural areas for the remaining cases. Overall, the Thompson microphysics showed good performance for most cases in both urban and rural areas while the WDM5 scheme showed better performance only for event-2 in both urban and rural areas, and for event-4 in the urban areas.

Figure 13 shows the MB results, indicating that all the simulations overestimated the wind speed in the urban areas and underestimated in rural areas in most cases. The

ensemble members 7, 9, 12, and 18 illustrated better performance for wind speed simulation in the urban areas and ensemble members 7 and 14 in rural areas in terms of MAE and RMSE. Therefore, the MYJ and ACM2 PBL scheme performs better for wind speed simulations in both urban and rural areas. The hybrid (local and non-local) ACM2 PBL scheme transitions from non-local to local closure under very stable condition (Hu et al. 2010). When this scheme acts as a non-local scheme, it considers non-local momentum mixing that is an advantage for calculating the rapid increase of wind speed more accurately in the early stage of mixed layer development. On the other hand, no single radiation and microphysics scheme showed distinctive performance in terms of MAE and RMSE for wind speed simulation in both urban and rural areas for all case studies. However, the combination of RRTMG+RRTMG (for events-1 and 2 in urban, and events-2 and 3 in rural areas) and Goddard+Goddard (for events-1 and 4 in rural, and event-4 in urban areas) radiation schemes showed better performance in simulating wind speed, whereas the Dudhia+RRTM schemes combination showed poor performance. The Thompson scheme produced lower MAE

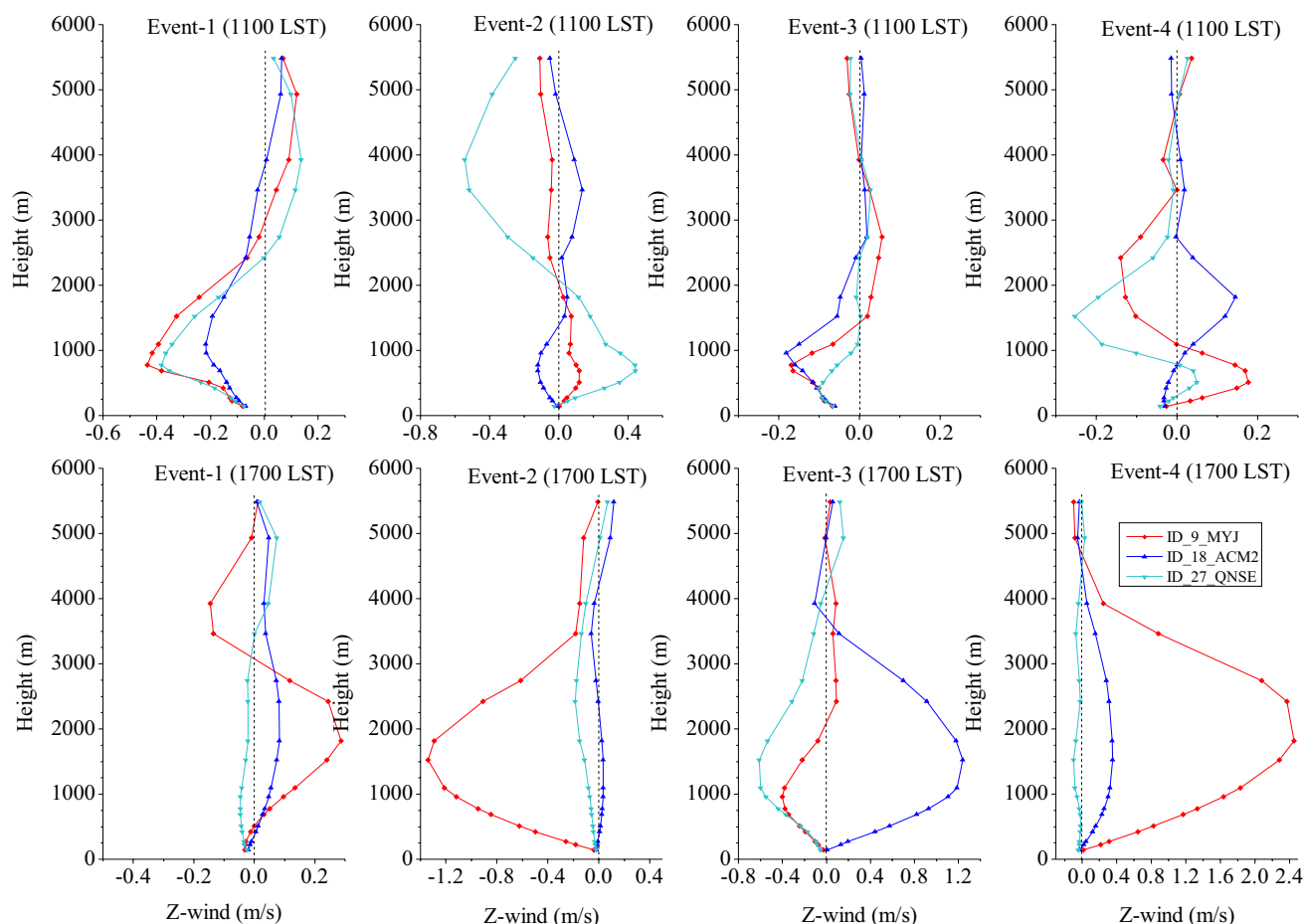


Fig. 9 Simulated vertical profile of vertical wind component at 1100 and 1700 LST for the four events

and RMSE for the events-1, 2 and 3 in the urban areas and for events-1 and 4 for rural areas. For the remaining events, the WDM5 scheme showed better performance in terms of MAE and RMSE analyses while the WSM6 scheme showed higher MAE and RMSE. All simulations showed larger errors in the urban areas compared to rural areas. This may occur due to an inaccurate representation of urban roughness. In summary, wind speed is more sensitive to the choice of PBL and microphysics schemes rather than SW and LW radiation schemes.

Figure 14 shows the wind profiles simulated by the three PBL schemes as compared to sounding observations taken at Melbourne airport at 2300 and 1100 LST for the 1st day and 2nd day of 3 days simulation period, respectively, similar to the temperature and moisture profiles shown in Figs. 6 and 7. The MYJ and QNSE PBL schemes showed higher wind speed than the ACM2 scheme within the lower PBL ($PBLH < 1000$ m) for the events-3 and 4 during night-time and for the events-1 and 3 during day-time. The MYJ and ACM2 schemes showed similar wind speed for the events-1 and 2 during night-time while the ACM2

simulated higher wind speed for the events-2 and 4 during day-time in the lower PBL ($PBLH < 1000$ m). All the PBL schemes underestimated the wind speed during day and overestimated at night for all the events except event-4 at night-time especially in the lower PBL. Overall, the MYJ scheme showed better performance (compared to observed wind profile) in most cases. During both the day- and night-time, the simulated wind speed showed large deviations from the observations. A possible reason is that the 2 km grid spacing does not properly resolve the eddies in turbulent boundary layer. Overestimation of wind speed is a common issue with the WRF model related to the low-level flow field, which is strongly influenced by nearby topography (Srikanth et al. 2015; Hariprasad et al. 2014).

5.4 Relative humidity

Figure 15 shows Taylor diagrams for relative humidity, showing high temporal pattern correlation (0.87–0.97) between model results and observations. All the three PBL schemes showed mixed performance in simulating relative

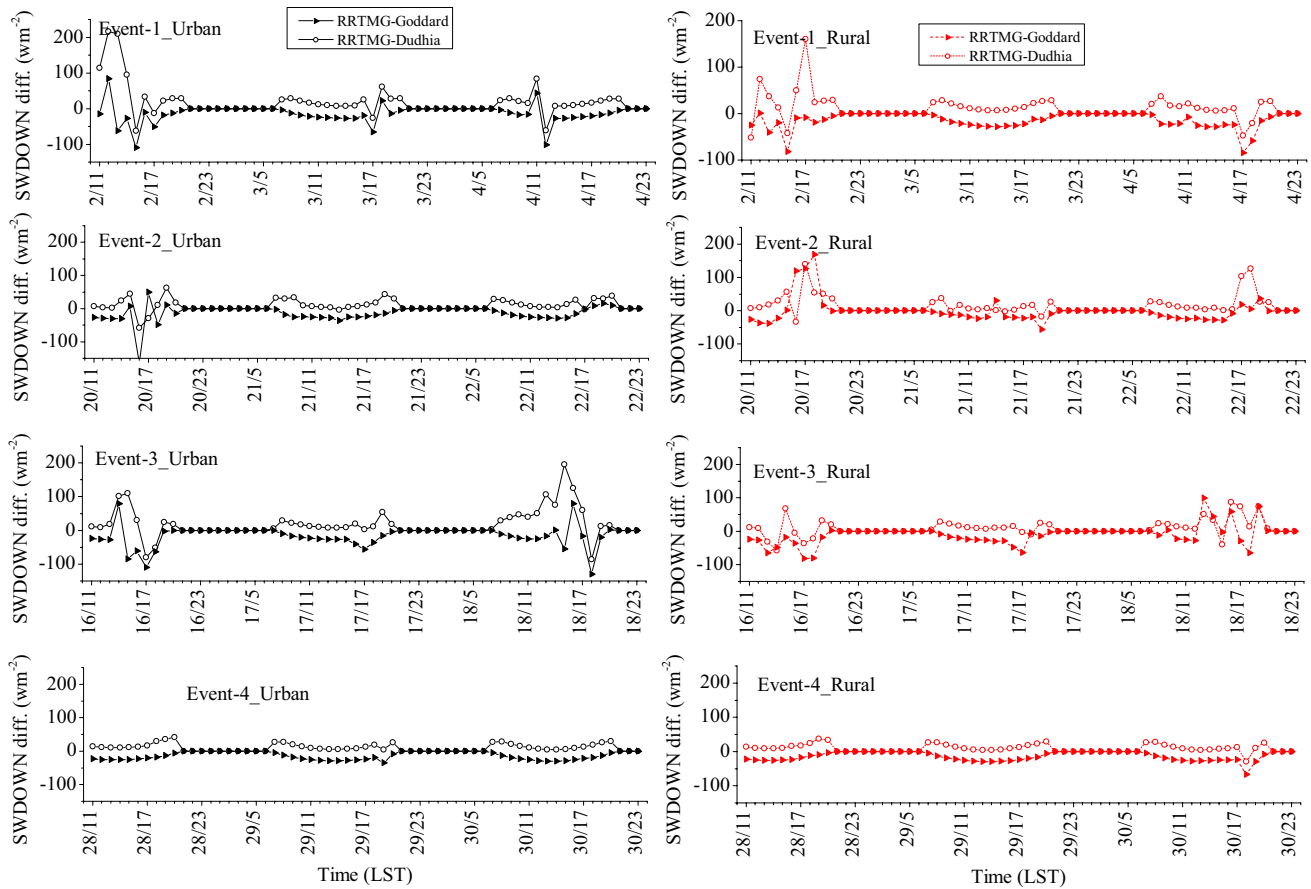


Fig. 10 Incoming SW radiation (SWDOWN) differences for different combinations of SW and LW radiation schemes

humidity in both urban and rural areas. The MYJ scheme showed better performance for event-2 in both urban and rural areas (Fig. 15b, f), and the QNSE showed mixed performance for event-3 in the urban areas (Fig. 15c) and for event-4 in rural areas (Fig. 15h), respectively. The ACM2 scheme performed better for event-1 in both urban and rural areas (Fig. 15a, e) and for event-3 in rural areas (Fig. 15g). The QNSE scheme performed better in two out of eight comparisons for event-4 in both urban and rural areas (Fig. 15d, h). The RRTMG+RRTMG combination performed very well in both urban and rural areas for all case studies except event-4 where the Dudhia+RRTM combination performed better. The Goddard+Goddard combination showed very poor performance in simulating relative humidity in both urban and rural areas. Interestingly, the Thompson microphysics scheme showed better performance in representing relative humidity for all events in both urban and rural areas except event-3 in rural areas.

Figure 16 shows the MB for relative humidity, indicating that all the PBL schemes had a tendency to underestimate relative humidity in the urban areas except event-3, and overestimate in rural areas. The ACM2 scheme showed an underestimation tendency as compared to

the MYJ and QNSE schemes in both urban and rural areas in most events. The ACM2 scheme also showed much higher biases in the urban areas especially for the events-1, 2 and 4, although it showed lower biases in rural areas. Overall, the MYJ scheme showed better performance in both urban and rural areas in terms of MB. Based on the MAE and RMSE results, it was clear that the ACM2 PBL scheme (especially ensemble member 18) showed better performance in rural areas, although the QNSE scheme indicated slightly lower MAE and RMSE in the urban areas except event-3. Although, the ACM2 showed better performance in terms of MB, MAE and RMSE in rural areas, this scheme showed large errors for the urban areas, except event-3. The QNSE scheme showed larger errors in rural areas and lower errors in urban areas, except event-3 as compared to the ACM2 scheme. Overall, the MYJ scheme showed more consistent and better performance in terms of MB, MAE and RMSE for both urban and rural areas, except event-3. The RRTMG+RRTMG combination clearly indicated good performance in both urban and rural areas except event-3 (where the Goddard+Goddard combination showed better performance) in rural areas in terms of MAE and

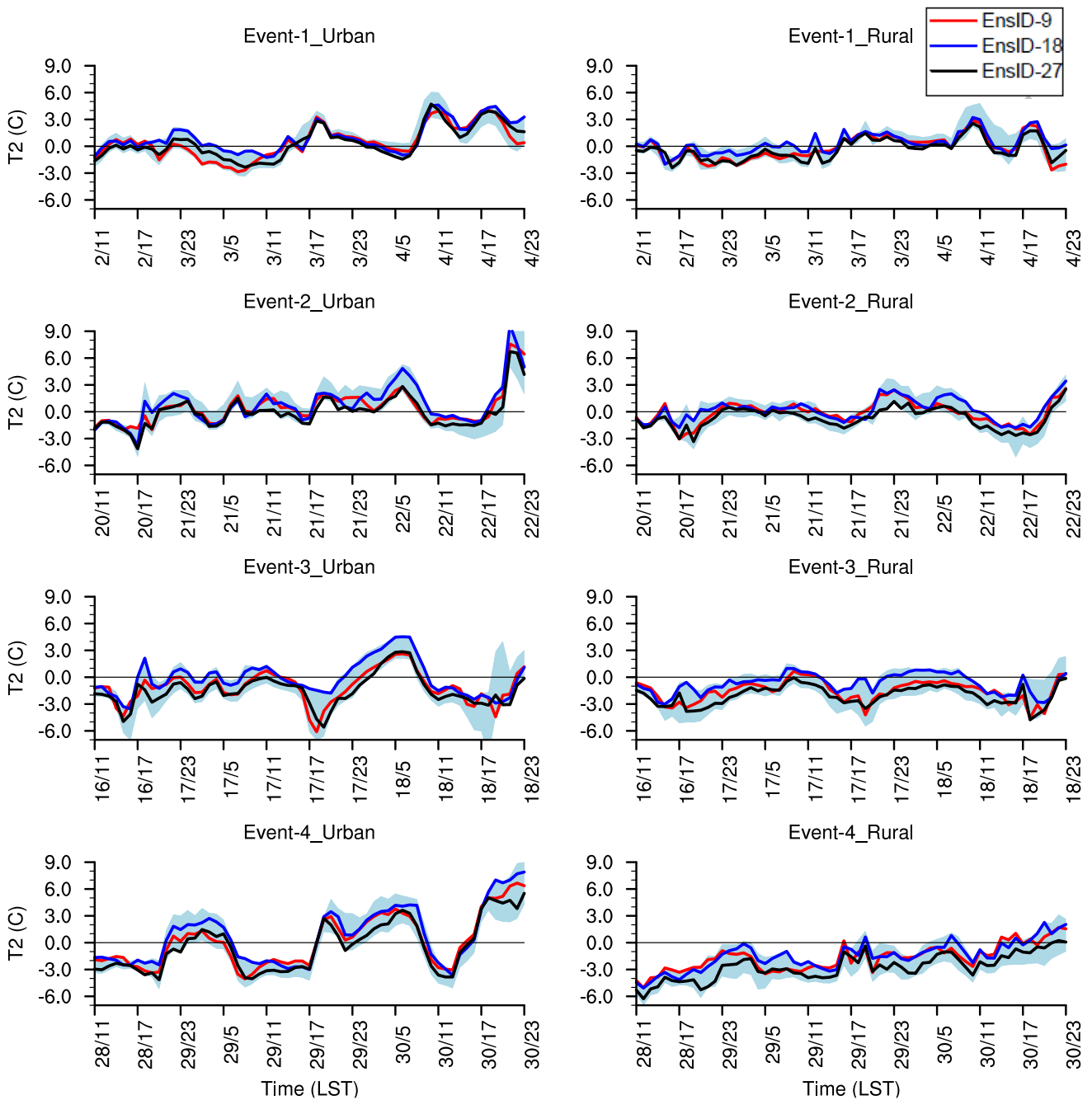


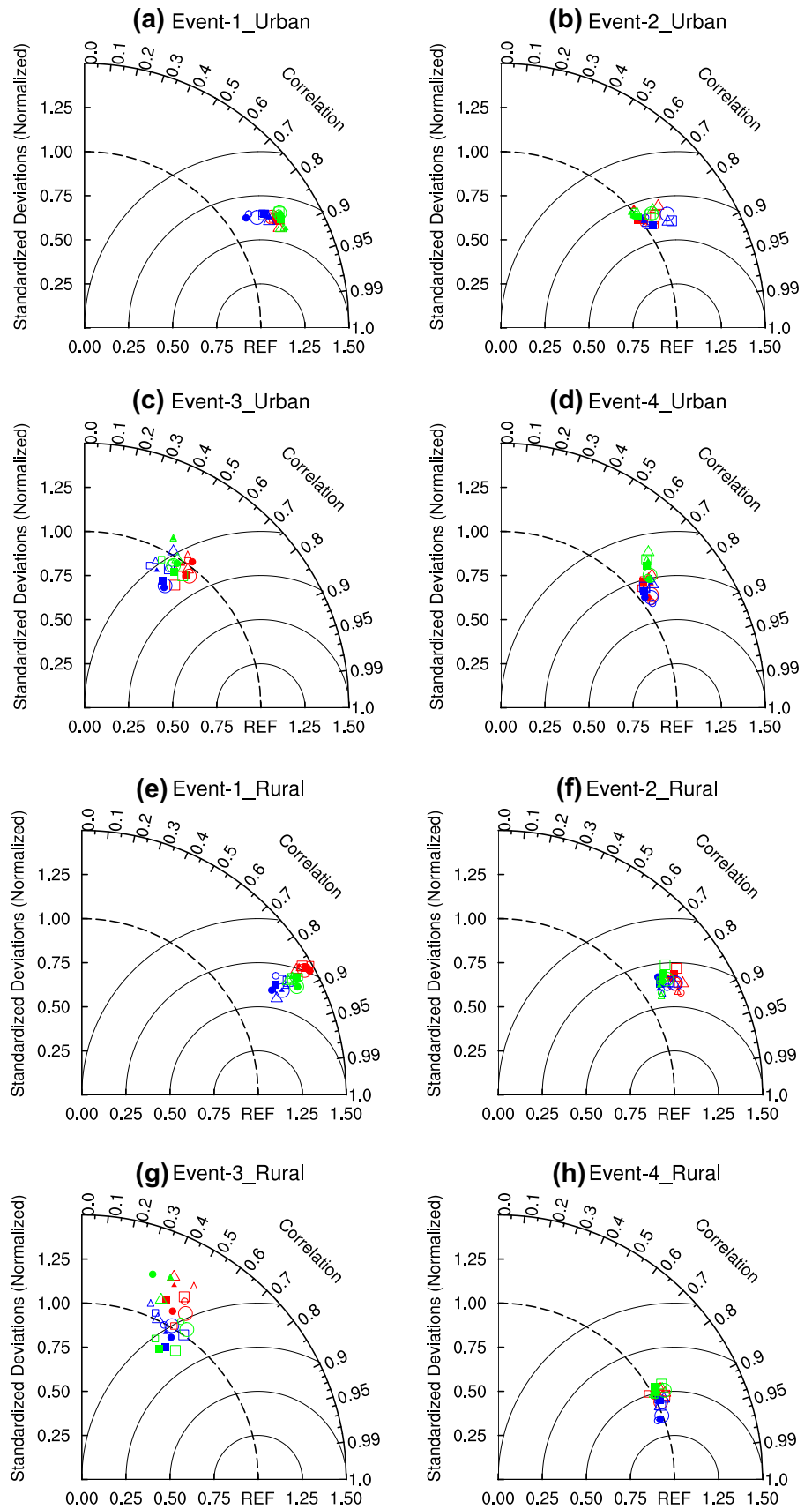
Fig. 11 Hourly temperature biases for the 27 WRF ensemble members: urban stations (*left*) and rural stations (*right*). The *gray shades* represent all ensemble members except ensemble members 9, 18 and 27

RMSE. Furthermore, the Thompson scheme showed best performance when it was used with the ACM2 PBL scheme, while the WDM5 scheme performed better with the QNSE PBL scheme. Therefore, the physics schemes showed non-linear interactions among them in simulating different variables.

6 Ranking of WRF ensemble members

To effectively summarize the results, the top five ensemble members in terms of temperature, wind speed and relative humidity simulations were identified based on the overall best performance for the four case studies. To achieve an overall ranking, the MAE metrics from all four case studies

Fig. 12 Same as Fig. 2 except for wind speed



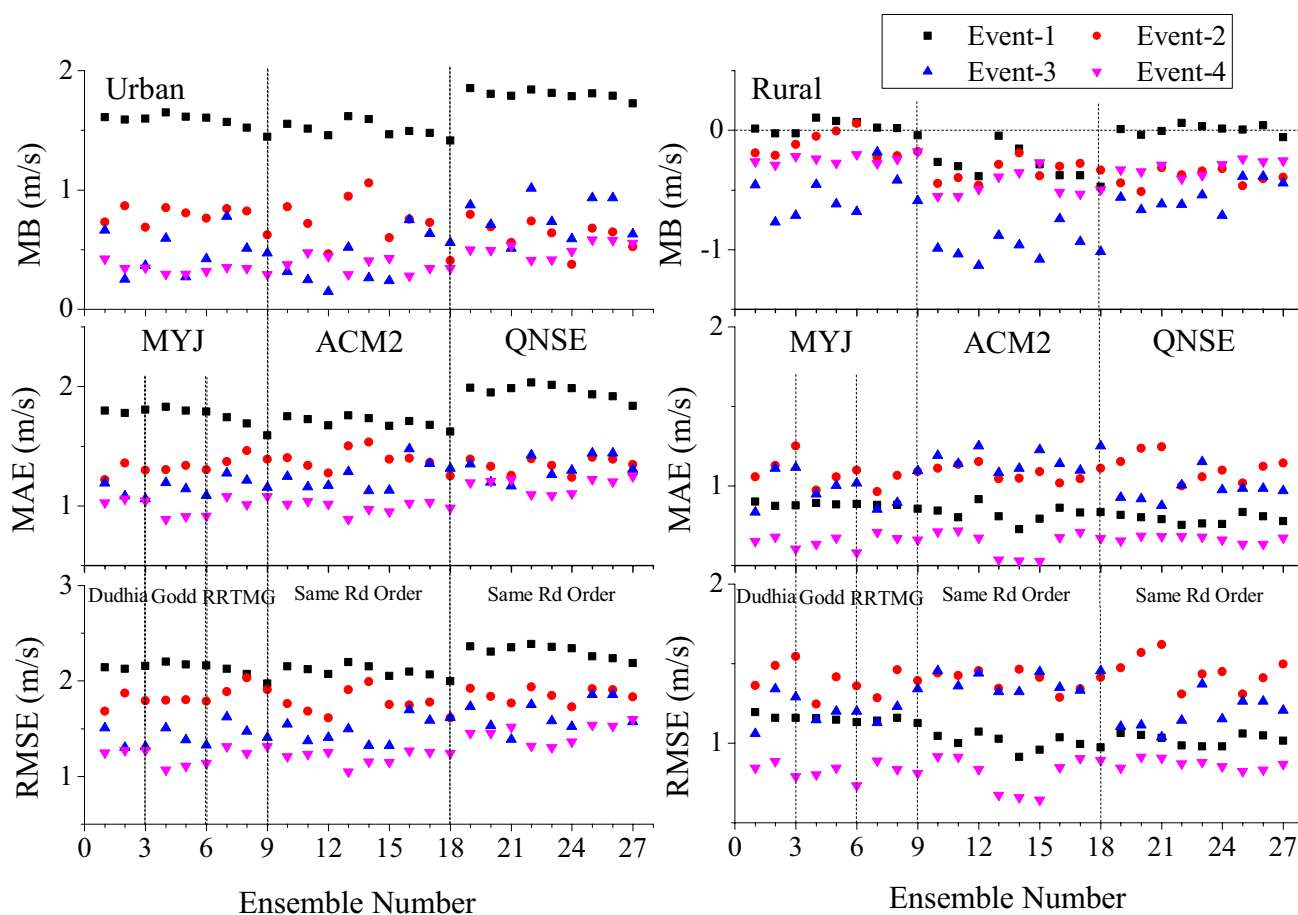


Fig. 13 Same as Fig. 3 except for wind speed

were first summed-up and then the mean metric was calculated for the three variables. The ranking of different ensemble members was made considering only the MAE metric for each variable separately in both urban and rural areas, since previous studies (Willmott et al. 2009; Chatterjee et al. 2013; Jerez et al. 2012) have emphasized the MAE metric rather than RMSE metric for evaluating the performance of a climate model. In this section, only the top five ranked ensemble members are presented and a brief discussion is presented for identifying the best WRF configuration. The ensemble member that showed the lowest mean metric was considered as the best WRF configuration.

Table 2 shows the top five model configurations based on the four case studies results in urban and rural areas. No unique ensemble member showed the best performance for all variables. However, the MYJ and ACM2 schemes showed better performance than the QNSE scheme in simulating temperature, wind and relative humidity in both urban and rural areas. The MYJ scheme had the highest ranking for the simulation of temperature in the urban areas, and wind speed simulation in both urban and rural areas. Conversely, the ACM2 scheme showed better

performance in simulating temperature in rural areas, and relative humidity in both urban and rural areas. Noticeably, the ACM2 scheme showed higher frequency as a better performing PBL scheme in the ranking Table 2. It is important to note that the results for event-3 were markedly different than other events (Figs. 3, 14), while the ACM2 PBL scheme showed consistently lower MAE for simulation of temperature and relative humidity in both urban and rural areas. This finding shows that evaluation of the WRF model using a single event could potentially be misleading. Therefore, the ACM2 scheme showed higher frequency as a better performing PBL scheme when considering overall MAE ranking in Table 2. The combination of RRTMG+RRTMG schemes performed better in most cases for temperature and relative humidity simulations, but this combination ranked in second position for wind speed simulation. The combination of Goddard+Goddard schemes showed better results for wind speed simulation. The Thompson microphysics scheme was a comparatively better option while the WDM5 scheme was the second best option. Based on overall performance, the best physics schemes were found to be the MYJ for PBL scheme, the

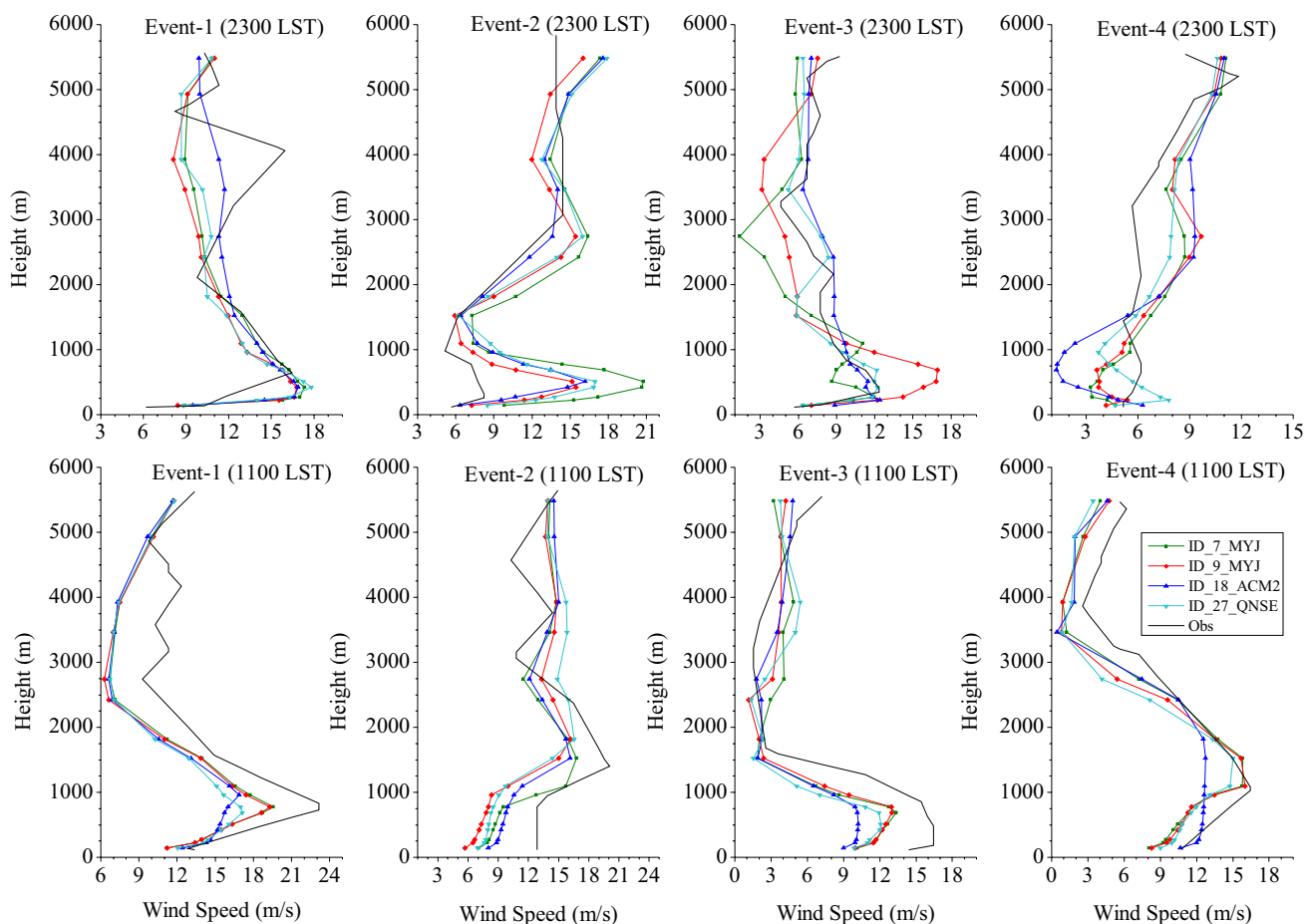


Fig. 14 Same as Fig. 6 except for wind speed

RRTMG+RRTMG combination for SW and LW radiation schemes and the Thompson for MP scheme for this study area.

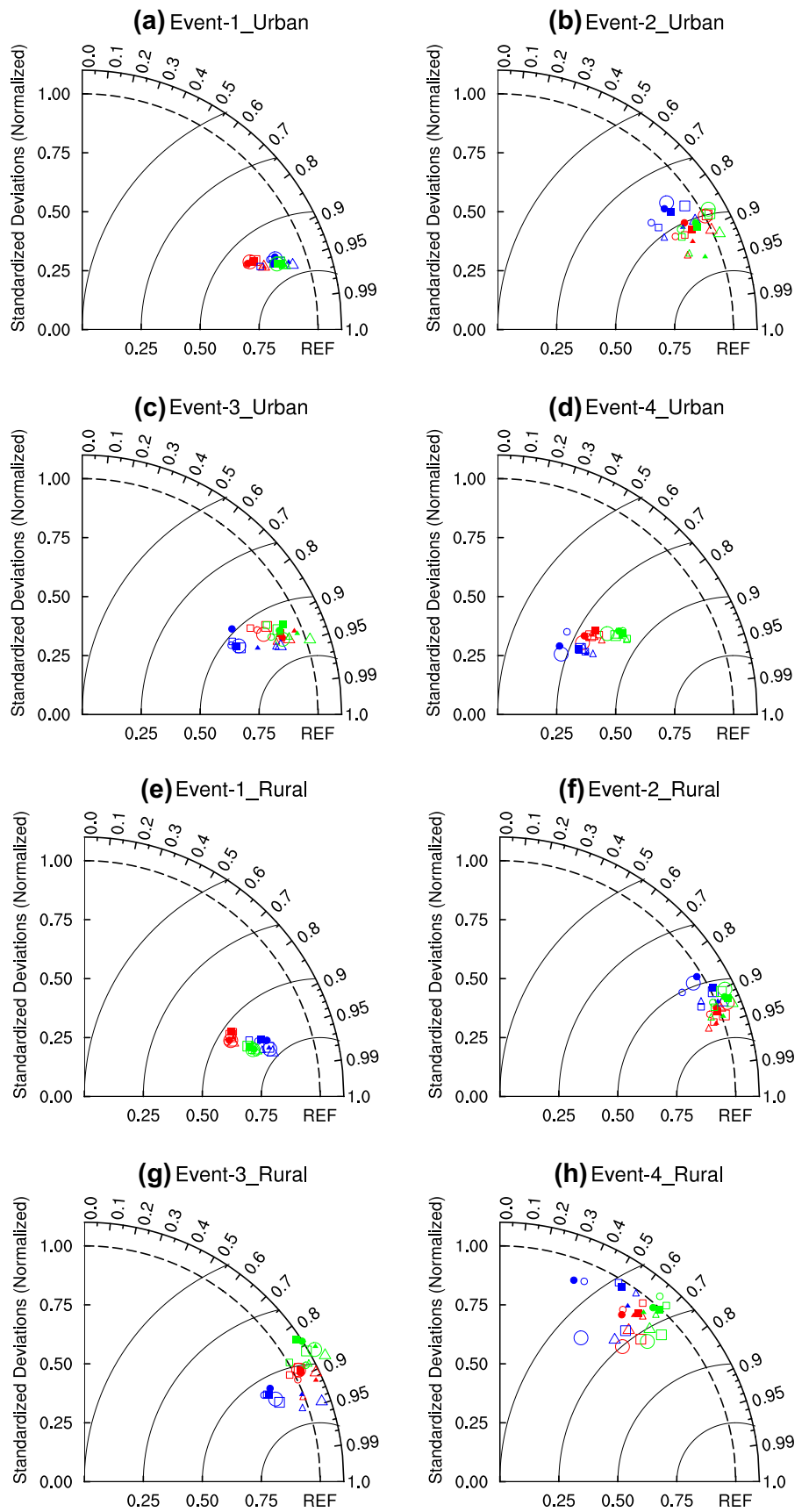
7 Evolution of synoptic heatwaves dynamics

This section describes the synoptic heatwave dynamics of event-4 during February 2009 as it was the most severe and well documented. Based on the statistical analysis in the previous section, the ensemble numbers 9 (MYJ+RRTMG/RRTMG+Thompson) and 18 (ACM2+RRTMG/RRTMG+Thompson) showed comparatively better performance. Additionally, the ensemble number 27 (QNSE+RRTMG/RRTMG+Thompson) was also tested for comparing the performance of the QNSE PBL scheme. The main change in these three WRF configurations is the PBL schemes where the MYJ, ACM2 and QNSE PBL schemes used for ensemble member 9, 18, and 27, respectively. The statistical analysis showed that the PBL schemes played most significant role and radiation schemes played

a moderate role in simulating heatwaves. Therefore, the performance of ensemble member 6 (MYJ+Goddard/Goddard+Thompson) was compared with the performance of ensemble number 9 (MYJ+RRTMG/RRTMG+Thompson) to better understand the effect of using different radiation schemes.

The WRF model simulations for mean sea level pressure (MSLP), wind speed, temperature, and geopotential heights at 850 and 500 hpa were considered for the analysis of heatwave dynamics. Daily (for 29th and 30th January 2009) means of MSLP for the event-4 is shown in Fig. 17a. This figure illustrates that all four ensemble members (EnsID-6, EnsID-9, EnsID-18 and EnsID-27) produced similar pattern and magnitude of MSLP over the southern ocean of the continent. However, the EnsID-9 shifts the center of the high slightly further southern ocean part as compared to other ensemble members. The reproduction of heatwave dynamics by the EnsID-9 is consistent with Engel et al. (2013), who showed that this heatwave event was largely driven by an anticyclone (high) over the south-east in the Tasman Sea. These highs are the key driver for

Fig. 15 Same as Fig. 2 except for relative humidity



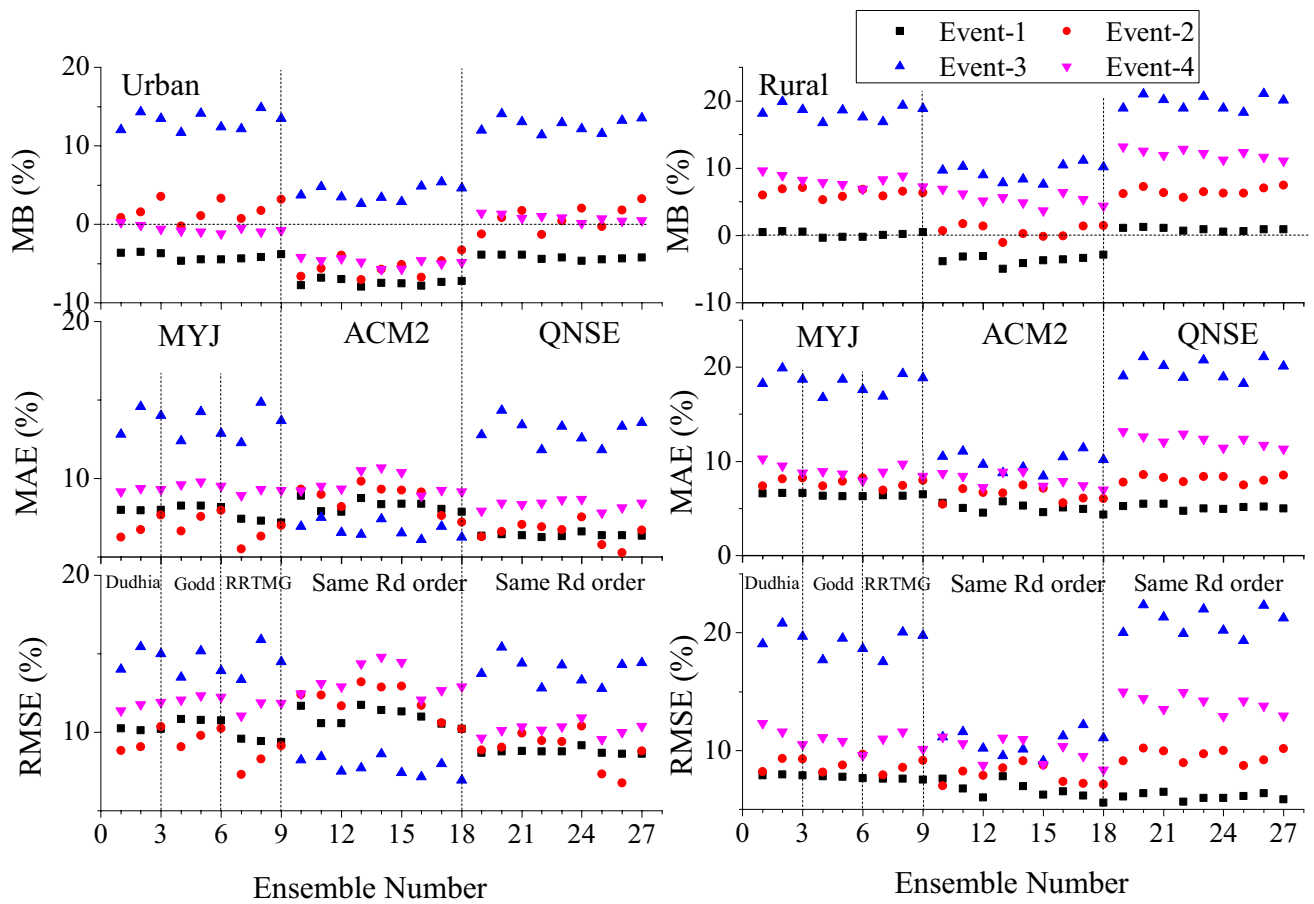


Fig. 16 Same as Fig. 3 except for relative humidity

Table 2 Ranking of WRF ensemble members based on MAE

Rank	Urban					Rural				
	Ens ID	PBL	ra_sw	ra_lw	mp	Ens ID	PBL	ra_sw	ra_lw	mp
Ranking in terms of temperature simulation										
1	7	MYJ	RRTMG	RRTMG	WDM5	18	ACM2	RRTMG	RRTMG	Thompson
2	9	MYJ	RRTMG	RRTMG	Thompson	17	ACM2	RRTMG	RRTMG	WSM6
3	18	ACM2	RRTMG	RRTMG	Thompson	16	ACM2	RRTMG	RRTMG	WDM5
4	17	ACM2	RRTMG	RRTMG	WSM6	12	ACM2	Dudhia	Dudhia	Thompson
5	16	ACM2	RRTMG	RRTMG	WDM5	15	ACM2	Goddard	Goddard	Thompson
Ranking in terms of wind speed simulation										
1	6	MYJ	Goddard	Goddard	Thompson	7	MYJ	RRTMG	RRTMG	WDM5
2	12	ACM2	Dudhia	Dudhia	Thompson	14	ACM2	Goddard	Goddard	WSM6
3	15	ACM2	Goddard	Goddard	Thompson	22	ACM2	Goddard	Goddard	WDM5
4	18	ACM2	RRTMG	RRTMG	Thompson	4	MYJ	Goddard	Goddard	WDM5
5	5	MYJ	Goddard	Goddard	WSM6	1	MYJ	Dudhia	Dudhia	WDM5
Ranking in terms of relative humidity simulation										
1	18	ACM2	RRTMG	RRTMG	Thompson	15	ACM2	Goddard	Goddard	Thompson
2	25	QNSE	RRTMG	RRTMG	WDM5	18	ACM2	RRTMG	RRTMG	Thompson
3	17	ACM2	RRTMG	RRTMG	WSM6	12	ACM2	Dudhia	Dudhia	Thompson
4	12	ACM2	Dudhia	Dudhia	Thompson	16	ACM2	RRTMG	RRTMG	WDM5
5	16	ACM2	RRTMG	RRTMG	WDM5	17	ACM2	RRTMG	RRTMG	WSM6

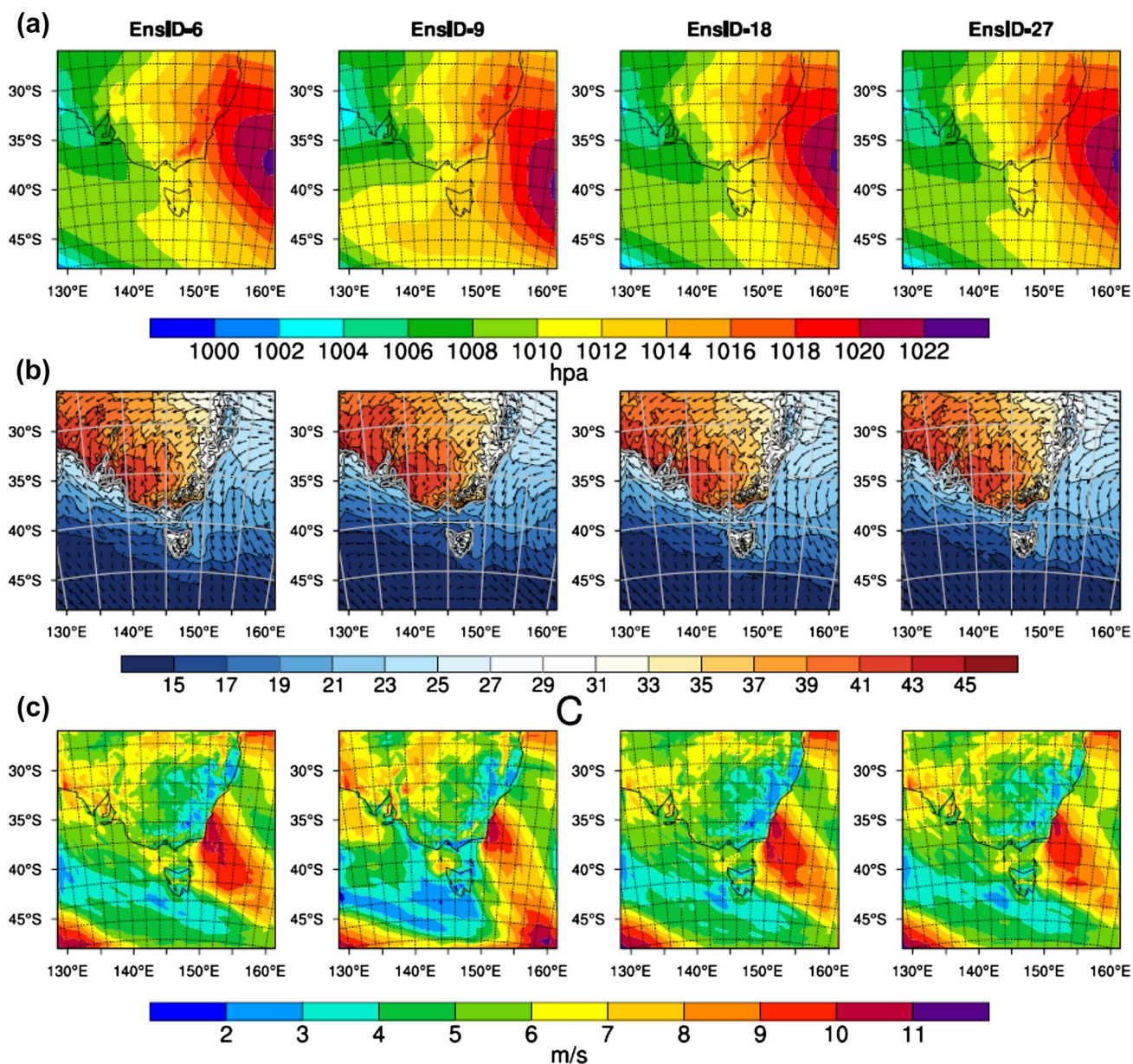


Fig. 17 Analyses of daily **a** MSLP, **b** overlaying temperature and wind rotated to earth coordinates and **c** wind speed

developing heatwaves in southeast Australia. Heatwaves in southeastern Australia are mainly driven by anticyclone systems over the Tasman Sea in line with the subtropical ridge (Marshall et al. 2014) and Rossby waves (Parker et al. 2014). According to the findings from those previous studies, the ensemble number 9 (MYJ) captured the extended anticyclone systems (ridge) over Tasman Sea and surrounding areas better than other ensemble members. Furthermore, Fig. 17b shows temperature and wind rotated to earth coordinates over the outermost domain (D01). The ensemble members 6 (MYJ) and 9 (MYJ) simulated a stronger temperature gradient over the southern part of the continent

and a weaker gradient over the southern ocean as compared to the ensemble members 18 (ACM2) and 27 (QNSE). Heatwaves in southeast Australia are also related to strong summertime frontogenesis over southern part of the continent (Berry et al. 2011), and the deformation of frontogenesis strengthens the temperature difference between the continent and the ocean when cyclone approaches heated continent (Engel et al. 2013). Finally, Fig. 17b, c shows that no significant differences of wind direction and wind speed among the four ensemble members. However, the MYJ scheme showed slightly lower wind speed as compared to the ACM2 and QNSE schemes over some parts of southern

ocean. Although, this study analyzed the geopotential heights at 850 and 500 hpa, there were no significant differences between geopotential heights simulated by different ensemble members and hence these results are not shown.

8 Sensitivity to different LSMs and role of the UCM

Finally, further sensitivity analysis were carried out to better understand the role of the UCM urban parameterization scheme as well as the use of different LSMs (Noah-MP and CLM4), and their effects on the heatwaves simulation. In this case, the best model configuration ensemble member 9

(MYJ+RRTMG/RRTMG+Thompson) from the previous analysis was used to test the influence of using the UCM as well as the CLM4 and NOAHMP LSMs. The CLM4 LSM was tested using same model configuration except that the MYJ PBL scheme was replaced with the AC2 PBL scheme as the MYJ scheme is not compatible with CLM4 in WRFv3.6.1. The additional model configurations for the sensitivity test are summarized in Table 3.

8.1 Statistical evaluation

The performance of the Noah-UCM, Noah-NoUCM, CLM4 and Noah-MP LSM for the four case studies is illustrated in Fig. 18 based on the urban and rural observed

Table 3 Additional WRF experiment design

Experiment name	Micro-physics	Ra_Sw/Ra_Lw physics	Surface layer physics	PBL physics	Land surface physics
Noah-UCM	Thompson	RRTMG/RRTMG	Eta similarity	MYJ	Coupled Noah-UCM LSM
Noah-NoUCM	Thompson	RRTMG/RRTMG	Eta similarity	MYJ	Unified Noah- LSM
CLM4	Thompson	RRTMG/RRTMG	MM5 similarity	ACM2	CLM4
Noah-MP	Thompson	RRTMG/RRTMG	Eta similarity	MYJ	Noah-MP

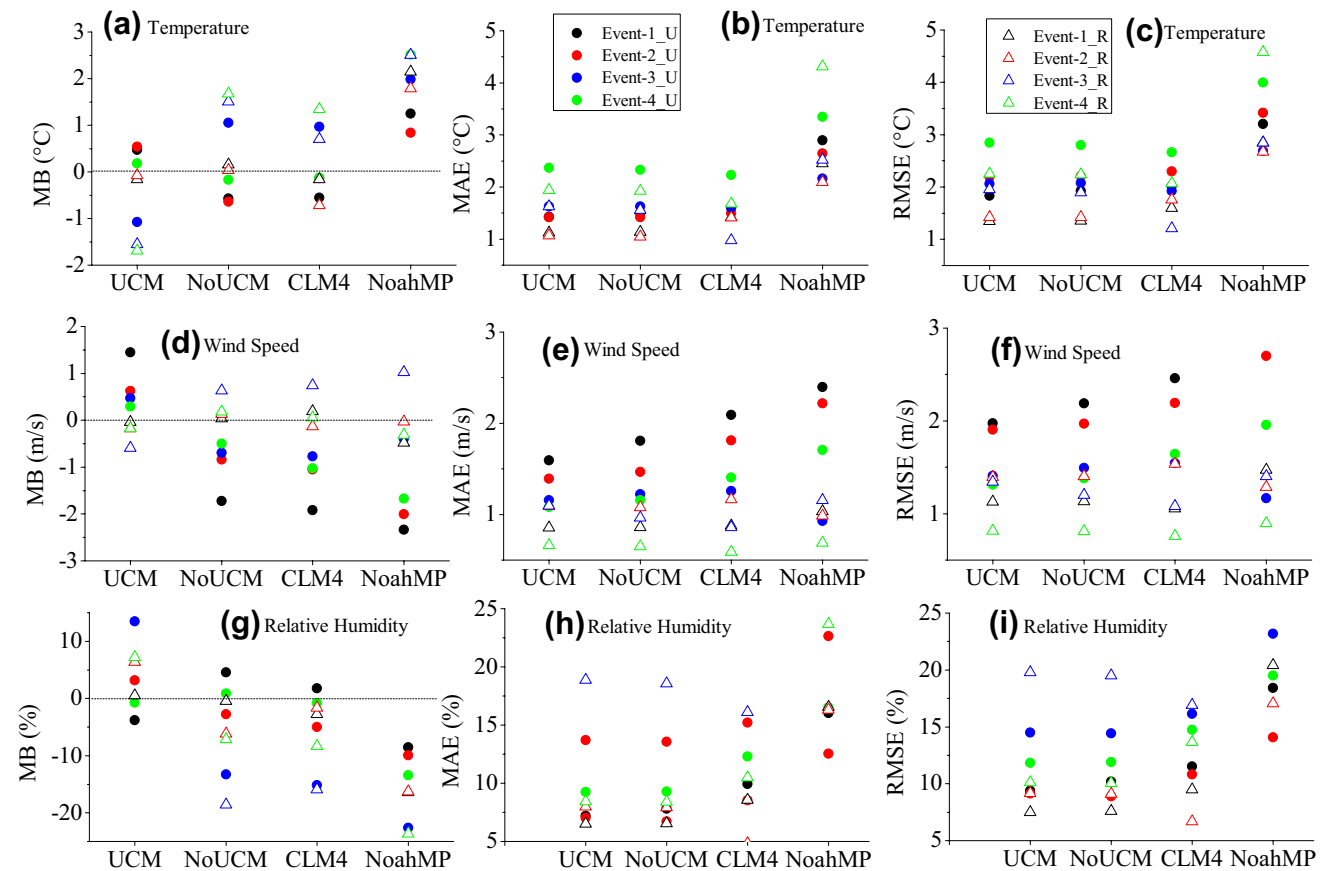


Fig. 18 MB, MAE and RMSE for T2max, wind and Rh2 for the Noah-UCM, Noah-NoUCM, Noah-MP and CLM4 LSMs

weather stations (similar to Fig. 3). For MB of temperature, the Noah-UCM showed a slight underestimation tendency in rural areas and overestimation in urban areas while the Noah-NoUCM showed the opposite trend. The CLM4 LSM leads to both over and underestimations for temperature. Interestingly, the Noah-MP showed consistent overestimation of temperature for both urban and rural areas for all four case studies. For wind speed and relative humidity simulations, the Noah-UCM showed a slight overestimation tendency while the CLM4, Noah-MP and Noah-NoUCM showed an underestimation tendency. Based on MAE and RMSE analyses, the Noah-UCM and Noah-NoUCM showed very similar performance while the CLM4 showed slightly improved performance for temperature simulations. The Noah-MP showed the worst performance for both urban and rural areas in all four cases. Finally, the Noah-UCM performed better for wind speed and relative humidity simulations for both urban and rural areas in all four cases. Overall, the Noah-UCM showed better performance.

8.2 Role of the urban canopy model

The differences of T2max bias are not significant between the simulations using the Noah-UCM and Noah-NoUCM in all case studies. All simulations (using the UCM and without the UCM) showed underestimation (mostly 2–5 °C) of maximum temperature in most case studies. Therefore, the UCM does not play significant role in reducing T2max bias during the heatwave events. The impacts of the UCM

on the surface fluxes (sensible and latent) are presented in Fig. 19. The significant differences of flux have been found only in the urban areas. The turbulent energy partitioning analysis showed that the Noah-NoUCM simulated higher sensible heat flux by 40–80 Wm⁻² than the coupled Noah-UCM especially over the urban areas. On the other hand, the coupled Noah-UCM simulated higher latent heat flux by 30–70 Wm⁻² than the Noah-NoUCM in the urban areas. The Noah-UCM LSM suppressed the latent heat flux by minimizing the effects of urban vegetation and enhanced sensible heat flux especially in the urban areas.

8.3 Role of the land surface model

The effect of different LSMs during heatwaves is evaluated in this section. The biases produced over the innermost domain by different LSMs for maximum temperature at 2 m (T2max) are shown in Fig. 20. All the analysis showed daily average biases of T2max for the 2nd and 3rd day of the 3 days simulation period. The Noah LSM without the UCM illustrates the major T2max bias ranging between 2 and 5 °C. When comparing the T2max bias between the Noah and Noah-MP, it was found that the Noah-MP showed maximum cooler bias (>7 °C) except for event-3. This finding indicates that land-surface processes strongly affect near-surface temperature during heatwave events. Although, the both unified Noah LSM with the MYJ PBL scheme and the CLM4 LSM with the ACM2 PBL scheme showed cooler biases, the Noah LSM with the MYJ scheme

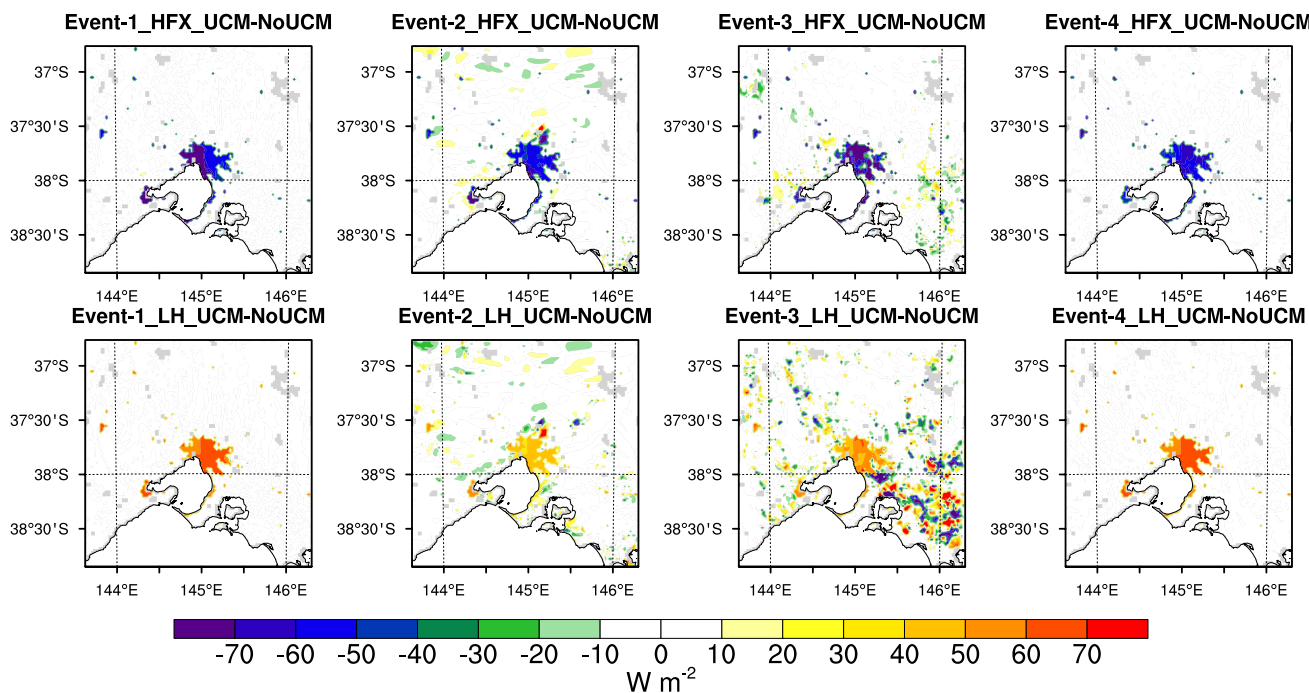


Fig. 19 The differences of sensible (upper) and latent (bottom) heat fluxes using the Noah-UCM and Noah-NoUCM

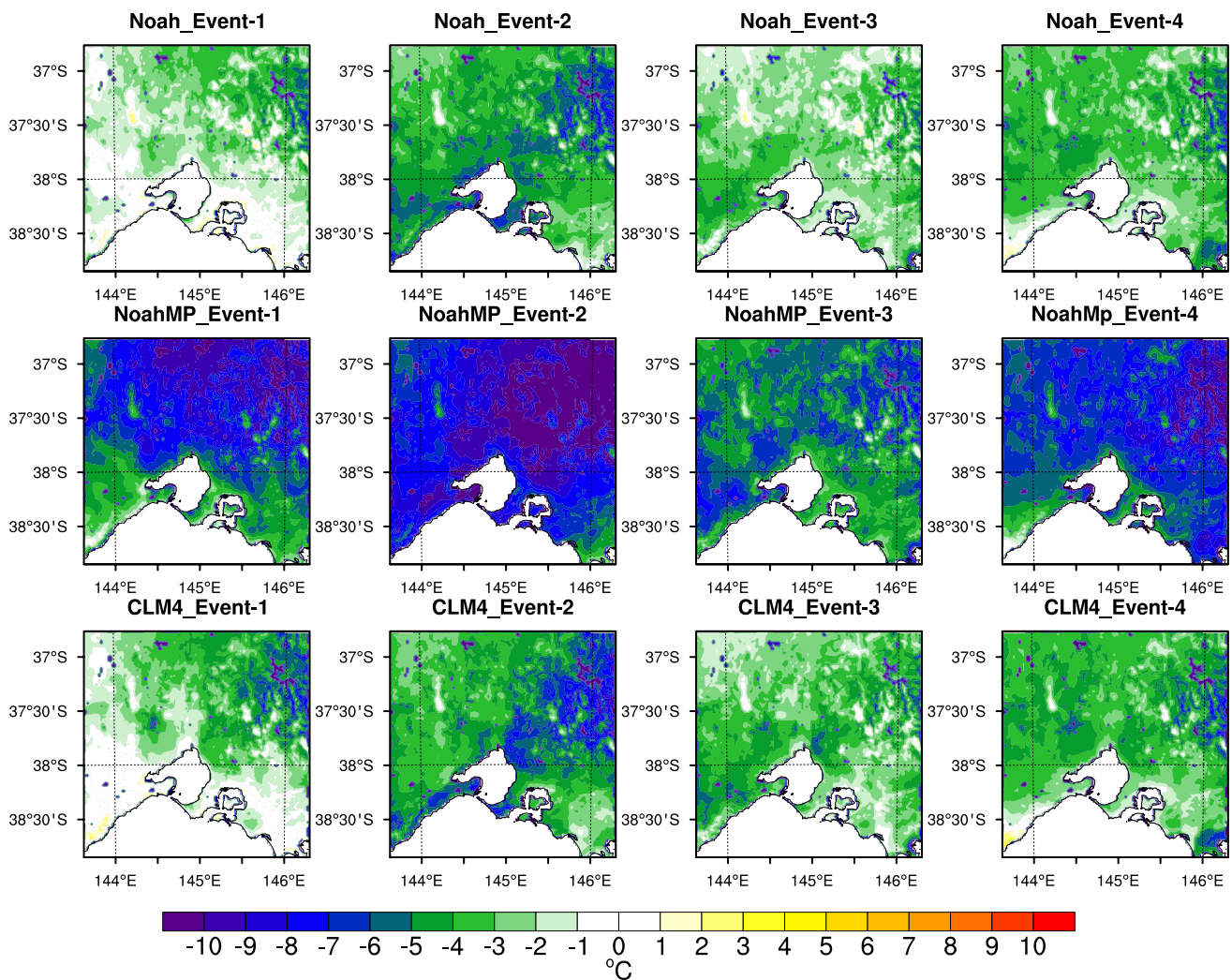


Fig. 20 T2max biases for simulations using the the Noah, Noah-MP and CLM4 LSMs

showed lower cooler bias than the CLM4 LSM with the ACM2 scheme. It is also important to note that the CLM4 LSM with the ACM2 PBL scheme showed slightly higher cooler bias (Fig. 20) than the Noah LSM with the ACM2 scheme (Fig. 4). Overall, all three LSM showed an underestimation tendency of T2max for all heatwave events. Finally, the Noah LSM showed slightly better performance (lower bias) than the Noah-MP and CLM4 LSMs.

The differences of surface energy balance fluxes (sensible and latent heat fluxes) for the three LSMs are shown in Fig. 21. The Noah-MP simulated higher sensible and latent heat fluxes than the Noah and CLM4 LSMs in the major areas of the domain in most cases. The Noah LSM shows higher sensible flux except for event-4 and lower latent heat flux except for event-3 as compared to the CLM4 in the major areas of the domain. Interestingly, the CLM4 shows higher latent heat flux over the highly dense (central of Melbourne city) urban areas while the Noah and

Noah-MP LSMs simulated higher latent heat flux over the rural areas that is main covered by evergreen broadleaf forest. These results show that the sensible and latent heat fluxes had considerable effect on the T2max bias (Fig. 20). For instance, the Noah LSM showed lower latent heat flux than the Noah-MP and CLM4 LSMs in major areas of the domain in most cases, and consequently, the Noah LSM produced lower T2max bias as compared to other the Noah-MP and CLM4 LSMs.

Figure 22 shows the differences of T2max simulated by the Noah, Noah-MP and CLM4 LSMs. The differences of T2max between the Noah and CLM4 are not significant. On the other hand, the Noah-MP simulates lower T2max ($\sim 7^\circ\text{C}$) for the events-1 and 2, and $2\text{--}4^\circ\text{C}$ for the events-3 and 4 in most areas of the domain as compared to the Noah and CLM4 LSMs. The Noah-MP experiment simulated lower T2max due to higher latent heat flux. Figure 21b illustrates that the Noah-MP simulates lower

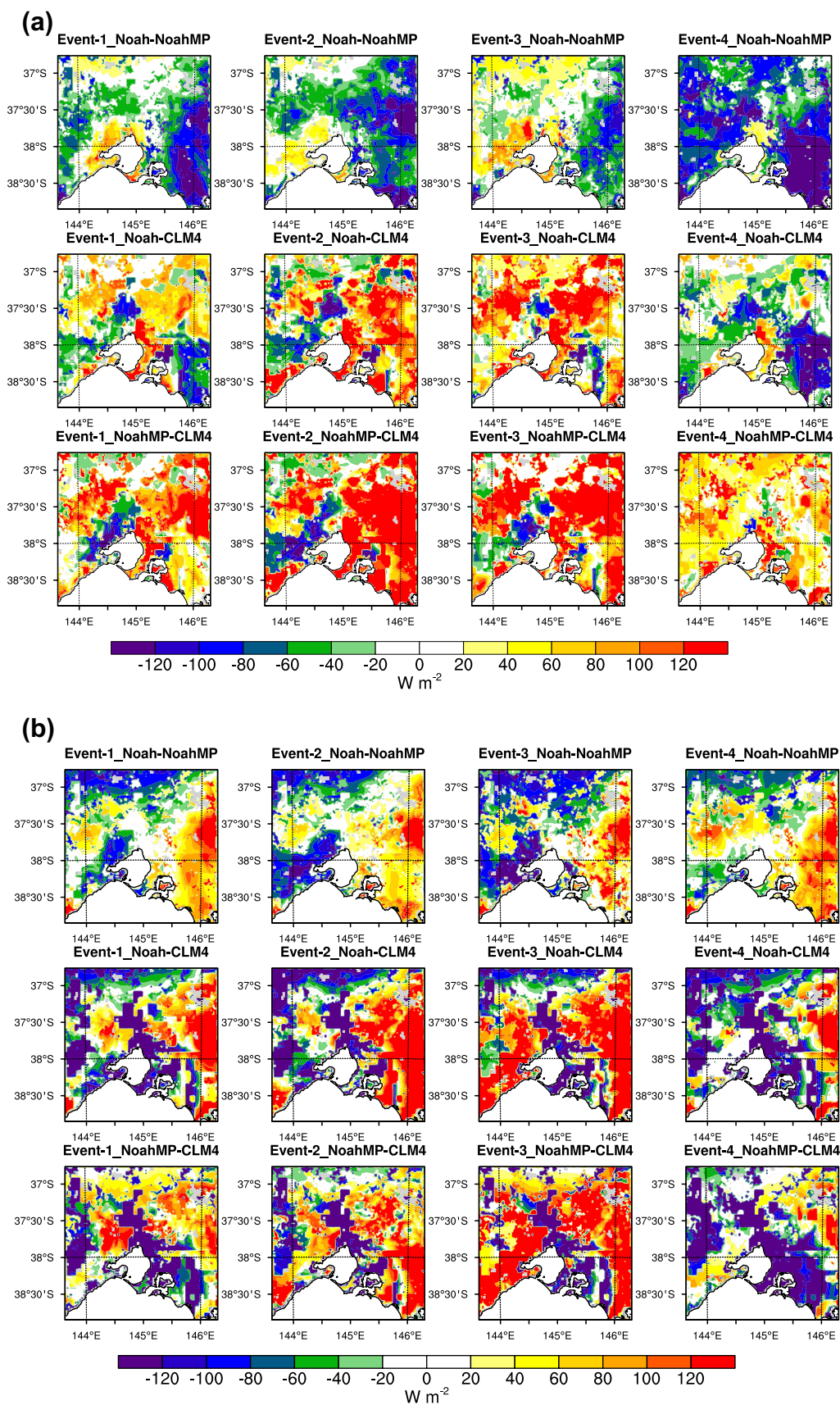


Fig. 21 Simulated daily **a** sensible heat flux and **b** latent heat flux differences for the Noah, Noah-MP and CLM4 LSMs

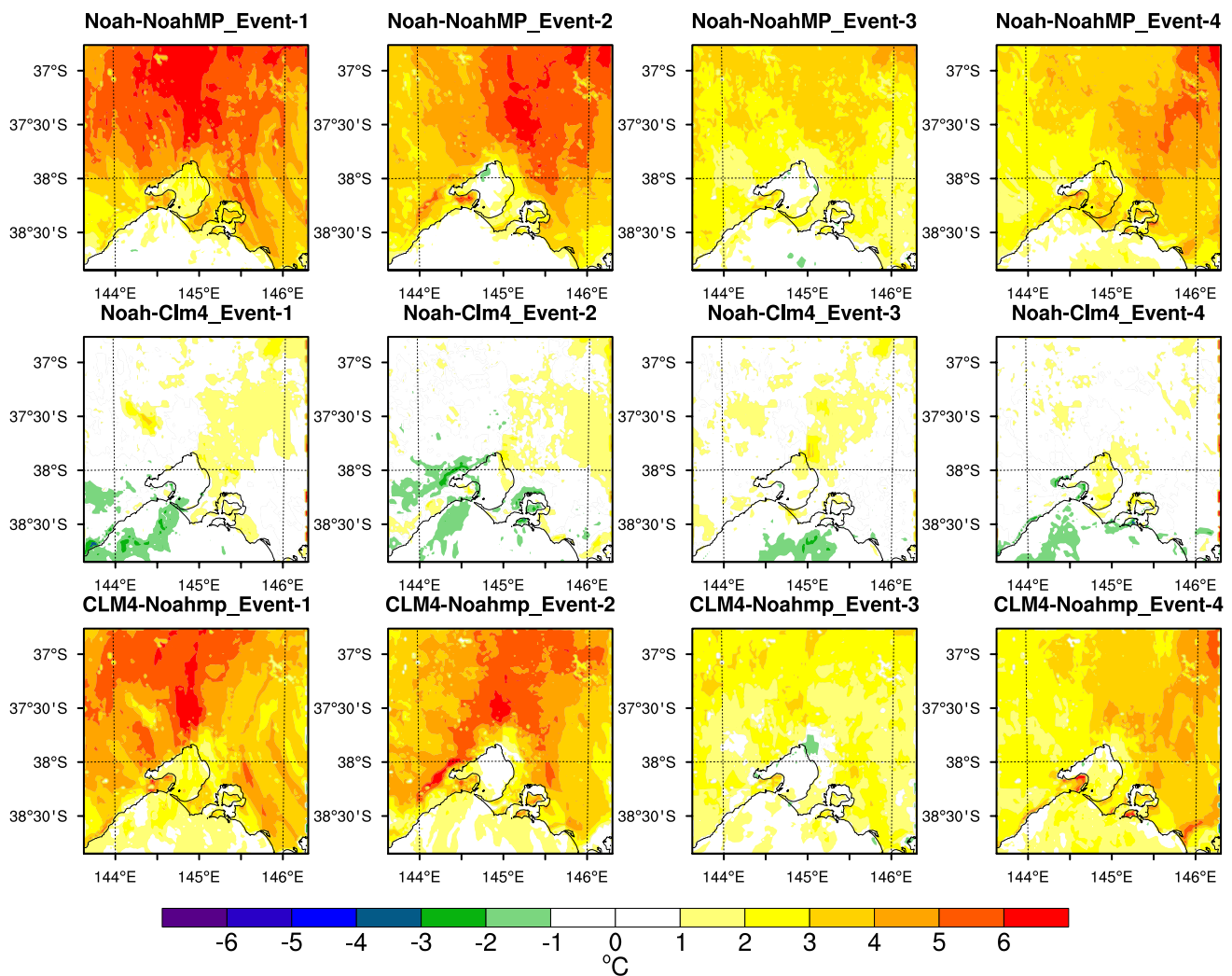


Fig. 22 Simulated daily temperature differences for the Noah, Noah-MP and CLM4 LSMs

T2max in those areas where this scheme simulated higher latent heat flux than the Noah and CLM4 LSMs. Finally, the Noah LSM shows higher T2max than the CLM4 due to higher sensible heat flux and lower latent heat flux. In some rural areas (evergreen broadleaf forest), the Noah-MP simulates higher sensible flux than the Noah and CLM4 LSMs (Fig. 21a), however, the T2max is still lower than the Noah and CLM4. Therefore, the surface fluxes (sensible and latent heat fluxes) only partly control the T2max. This shows that further work is required to fine-tune the different options for Noah-MP to reduce the T2max bias, rather than using the default options. This is however, outside the scope of this study. Another factor which can have a strong influence on T2max is soil moisture; however, there were very small differences in soil moisture between the different LSMs (not shown).

9 Conclusions

A series of simulations were conducted using a WRF multi-physics ensemble for assessing the sensitivity of the simulations to various physical parameterizations. A total 27 WRF configurations were generated including three PBL schemes (MYJ, ACM2, and QNSE), three MP schemes (WDM5, WSM6 and Thompson), and three SW (Dudhia, RRTMG, Goodard) and LW (RRTM, RRTMG, Goddard) radiation schemes. This study aimed at identifying the best WRF model configuration for reproducing heatwave events over Melbourne, Australia. The model outputs were compared with the observations for the innermost domain (D03) with 2 km grid resolution while the output of outermost domain (D01) was used to examine the heatwave dynamics. After obtaining the best WRF model

configuration, further experiments were carried out to investigate the role of the UCM and other LSMs (Noah-MP and CLM4) in simulating the heatwave events.

The evaluation metrics revealed that a particular WRF configuration rarely performed the best for all case studies for all variables and locations in terms of all evaluation metrics. This finding is consistent with the previous studies (Stegehuis et al. 2015; Evans et al. 2012) that no single configuration of the multi-physics ensemble reveals the best performance for all variables for all cases. Based on statistical measures, it was very difficult to identify a consistently best performing WRF ensemble member. Different metrics showed preferences for some particular physics options across all four events. In the current study, the best selection was made based on the aggregated performance. The results from this study confirmed that, overall, the configuration that consists of the MYJ PBL scheme, RRTMG+RRTMG SW and LW radiation schemes and Thompson MP scheme demonstrated better performance than any other configuration. Furthermore, any combination consisting of the QNSE PBL scheme, Dudhia+RRTM radiation schemes and WSM6 MP scheme should be avoided for this region.

The results have showed that the ACM2 scheme leads to simulations of higher temperature and lower moisture because of their stronger vertical mixing and entrainment at the top PBL due to non-local effects which help to develop a warmer and drier PBL. Meanwhile, the local-closure schemes MYJ and QNSE consider only local mixing for entrainment without considering the effect of large eddies. These two local-closure schemes simulate higher moisture in the lower PBL where the surface layer physics plays important role. Therefore, the MYJ and QNSE schemes produce a comparatively cooler PBL than the ACM2 scheme. All simulations have illustrated a reasonable degree of fidelity of the WRF model in simulating temperature, wind speed and relative humidity over the study area. The model shows a higher skill in the simulation of temperature and wind speed in rural areas as compared to urban areas. Furthermore, it shows the opposite skill for relative humidity.

This study quantitatively analyzed the role of land surface processes in the WRF model using the Noah, Noah-MP and CLM4 LSMs. The results of T2max bias for the LSMs show that near-surface temperature is significantly affected by the surface fluxes and land-surface processes, with use of the Noah-MP LSM resulting in large negative biases in T2max, and Noah and CLM4 LSM resulting in similar biases, with the Noah LSM providing the best performance overall. Using the UCM had an influence on the sensible and latent heat fluxes in the urban area, but this did not translate to a significant difference in T2max when compared with observations. The limitations of this

study are the use of default land use categories in the WRF model, which may not be an accurate representation of actual land-use. Therefore, further sensitivity analysis of WRF to more accurate land-use information will be carried out in future studies. Finally, the optimal WRF model configuration identified in this study can be used for further assessment of urban climate and the impacts of extreme hot weather on urban environment and human health.

Acknowledgements Data support by the Bureau of Meteorology (BoM) Australia, ECMWF (ERA-interim) data server (<http://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=ml/>) and University of Wyoming (atmospheric sounding data) data server (<http://weather.uwyo.edu/upperair/sounding.html>) are gratefully acknowledged. Authors also acknowledge Dr. Sachindra Dhanapala Arachchige for his cooperation in this study. The creation of the ANUCLim data was funded by the Terrestrial Ecosystem Research Network (TERN) Ecosystem Modelling and Scaling Infrastructure (eMAST) Facility under the National Collaborative Research Infrastructure Strategy (NCRIS) 2013–2014 budget initiative of the Australian Government Department of Industry, the Australian Government Department of Environment in support of the National Carbon Accounting System, and the Australian National University. The comments of an anonymous reviewer helped to improve the manuscript. All this assistance is gratefully acknowledged.

References

- Arugueso D, Evans JP, Fita L, Bormann KJ (2014) Temperature response to future urbanization and climate change. *Clim Dyn* 42:2183–2199. doi:10.1007/s00382-013-1789-6
- Beniston M et al (2007) Future extreme events in European climate: an exploration of regional climate model projections. *Clim Change* 81:71–95. doi:10.1007/s10584-006-9226-z
- Berry G, Reeder MJ, Jakob C (2011) A global climatology of atmospheric fronts. *Geophys Res Lett* 38:L04809. doi:10.1029/2010GL046451
- Borge R, Alexandrov V, José del Vas J, Lumbreras J, Rodríguez E (2008) A comprehensive sensitivity analysis of the WRF model for air quality applications over the Iberian Peninsula. *Atmos Environ* 42:8560–8574. doi:10.1016/j.atmosenv.2008.08.032
- Boschat G, Pezza A, Simmonds S, Perkins S, Cowan T, Purich A (2015) Large scale and sub-regional connections in the lead up to summer heat wave and extreme rainfall events in eastern Australia. *Clim Dyn* 44:1823–1840
- Bukovsky MS, Karoly DJ (2009) Precipitation simulations using WRF as a nested regional climate model. *J Appl Meteorol Climatol* 48:2152–2159. doi:10.1175/2009JAMC2186.1
- Cai W, Cowan T, Raupach M (2009) Positive Indian Ocean Dipole events precondition southeast Australia bushfires. *Geophys Res Lett* 36:L19710. doi:10.1029/2009GL039902
- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geosci Model Dev* 7:1247–1250. doi:10.5194/gmd-7-1247-2014
- Chatterjee A, Engelen RJ, Kawa SR, Sweeney C, Michalak AM (2013) Background error covariance estimation for atmospheric CO₂ data assimilation. *J Geophys Res Atmos* 118:10,140–110,154. doi:10.1002/jgrd.50654
- Chen F, Kusaka H, Bornstein R, Ching J, Grimmond CSB, Grossman-Clarke S, Loridan T, Manning KW, Martilli A, Miao S, Sailor D, Salamanca FP, Taha H, Tewari M, Wang X, Wyszogrodzki AA,

- Zhang C (2011) The integrated WRF/urban modelling system: development, evaluation, and applications to urban environmental problems. *Int J Climatol* 31:273–288. doi:10.1002/joc.2158
- Chen F, Yang X, Zhu W (2014) WRF simulations of urban heat island under hot-weather synoptic conditions: The case study of Hangzhou City, China. *Atmos Res* 138:364–377. doi:10.1016/j.atmosres.2013.12.005
- Chou M-D, Suarez MJ (1999) A solar radiation parameterization for atmospheric studies. NASA Tech Memo 104606:40
- Chou M. D., Suarez M. J. (2001) A thermal infrared radiation parameterization for atmospheric studies. NASA/TM-2001-104606 19:55
- Chow WTL, Brennan D, Brazel AJ (2012) Urban heat island research in Phoenix, Arizona: theoretical contributions and policy applications. *Bull Am Meteorol Soc* 93:517–530. doi:10.1175/BAMS-D-11-00011.1
- Cowan T, Purich A, Perkins S, Pezza A, Boschat G, Sadler K (2014) More frequent, longer, and hotter heat waves for Australia in the twenty-first century. *J Clim* 27:5851–5871. doi:10.1175/JCLI-D-14-00092.1
- Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Ho'lm EV, Isaksen I, Kallberg P, Kohler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park BK, Peubey C, de Rosnay P, Tavolato C, The'paut J-N, Vitart F (2011) The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q J R Meteorol Soc* 137:553–597. doi:10.1002/qj.828
- Dudhia J (1989) Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. *J Atmos Sci* 46:3077–3107
- Emery C, Tai E, Yarwood G (2001) Enhanced meteorological modeling and performance evaluation for two Texas ozone episodes. Final report. The Texas Natural Resource Conservation Commission, Austin. <https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/mm/EnhancedMetModelingAndPerformanceEvaluation.pdf>. Accessed 14 Nov 2016
- Emmanuel R, Krüger E (2012) Urban heat island and its impact on climate change resilience in a shrinking city: the case of Glasgow, UK. *Build Environ* 53:137–149. doi:10.1016/j.buildenv.2012.01.020
- Engel CB, Lane TP, Reeder MJ, Reznay M (2013) The meteorology of Black Saturday. *Q J R Meteorol Soc* 139:585–599
- Evans JP, Boyer-Souchet I (2012) Local sea surface temperatures add to extreme precipitation in northeast Australia during La Niña. *Geophys Res Lett*. doi:10.1029/2012GL052014
- Evans JP, Ekström M, Ji F (2012) Evaluating the performance of a WRF physics ensemble over South-East Australia. *Clim Dyn* 39:1241–1258. doi:10.1007/s00382-011-1244-5
- Fallmann J, Emeis S, Suppan P (2014) Mitigation of urban heat stress—a modelling case study for the area of Stuttgart. *DIE ERDE* 144(3–4):202–216. doi:10.12854/erde14415
- Fischer EM, Oleson KW, Lawrence DM (2012) Contrasting urban and rural heat stress responses to climate change. *Geophys Res Lett*. doi:10.1029/2011GL050576
- García-Díez M, Fernández J, Fita L, Yagüe C (2013) Seasonal dependence of WRF model biases and sensitivity to PBL schemes over Europe. *Q J R Meteorol Soc* 139:501–514
- Giannaros TM, Melas D, Daglis IA, Keramitsoglou I, Kourtidis K (2013) Numerical study of the urban heat island over Athens (Greece) with the WRF model. *Atmos Environ* 73:103–111. doi:10.1016/j.atmosenv.2013.02.055
- Gilliam RC, Pleim JE (2010) Performance assessment of new land surface and planetary boundary layer physics in the WRF-ARW. *J Appl Meteorol Climatol* 49:760–774. doi:10.1175/2009JAMC2126.1
- Gilliam RC, Hogrefe C, Rao ST (2006) New methods for evaluating meteorological models used in air quality applications. *Atmos Environ* 40:5073–5086. doi:10.1016/j.atmosenv.2006.01.023
- Hall WD, Rasmussen RM, Thompson G (2005) The new Thompson microphysical scheme in WRF. Preprints, 2005 WRF/MM5 user's workshop, Boulder, CO. NCAR, 6.1
- Han Z, Ueda H, An J (2008) Evaluation and intercomparison of meteorological predictions by five MM5-PBL parameterizations in combination with three land-surface models. *Atmos Environ* 42:233–249. doi:10.1016/j.atmosenv.2007.09.053
- Hariprasad KBRR, Srinivas CV, Singh AB, Rao SVB, Baskaran R, Venkatraman B (2014) Numerical simulation and intercomparison of boundary layer structure with different PBL schemes in WRF using experimental observations at a tropical site. *Atmos Res* 145:27–44
- Herold N, Kala J, Alexander L (2016) The influence of soil moisture deficits on Australian heat waves. *Environ Res Lett* 11:064003. doi:10.1088/1748-9326/11/6/064003
- Hirsch AL, Pitman AJ, Kala J (2014) The role of land cover change in modulating the soil moisture-temperature land-atmosphere coupling strength over Australia. *Geophys Res Lett* 41:5883–5890. doi:10.1002/2014GL061179
- Hong S-Y, Lim J-O (2006) The WRF single-moment 6-class microphysics scheme (WSM6). *J Korean Meteorol Soc* 42:129–151
- Hu X-M, Nielsen-Gammon JW, Zhang F (2010) Evaluation of three planetary boundary layer schemes in the WRF model. *J Appl Meteorol Climatol* 49:1831–1844. doi:10.1175/2010JAMC2432.1
- Hu X-M, Klein PM, Xue M, Lundquist JK, Zhang F, Qi Y (2013) Impact of low-level jets on the nocturnal urban heat island intensity in Oklahoma city. *J Appl Meteorol Climatol* 52:1779–1802. doi:10.1175/JAMC-D-12-0256.1
- Hutchinson MF, Mckenney DW, Lawrence K, Pedlar J, Hopkinson R, Milewska E, Papadopol P (2009) Development and testing of Canada-wide interpolated spatial models of daily minimum/maximum temperature and precipitation for 1961–2003. *J Appl Meteorol Climatol* 48:725–741. doi:10.1175/2008JAMC1979.1
- Hutchinson MF, Kesteven J, Xu T (2014) Daily maximum temperature: ANUClimate 1. 1, 0.01 degree, Australian coverage, 1970–2014. Australian National University, Canberra. <http://dap.nci.org.au>, made available by the Ecosystem Modelling and Scaling Infrastructure (eMAST, <http://www.emast.org.au>) of the Terrestrial Ecosystem Research Network (TERN, <http://www.tern.org.au>)
- Iacono MJ, Delamere JS, Mlawer EJ, Shephard MW, Clough SA, Collins WD (2008) Radiative forcing by long-lived greenhouse gases: calculations with the AER radiative transfer models. *J Geophys Res* 113:D13103
- Imhoff ML, Zhang P, Wolfe RE, Bounoua L (2010) Remote sensing of the urban heat island effect across biomes in the continental USA. *Remote Sens Environ* 114:504–513. doi:10.1016/j.rse.2009.10.008
- Jankov I, Jr, WAG, Segal M, Shaw B, Koch SE (2005) The impact of different WRF model physical parameterizations and their interactions on warm season MCS rainfall. *Weather Forecast* 20:1048–1060. doi:10.1175/WAF888.1
- Jankov I et al (2011) An evaluation of five ARW-WRF microphysics schemes using synthetic GOES imagery for an atmospheric river event affecting the California coast. *J Hydrometeorol* 12:618–633. doi:10.1175/2010JHM1282.1
- Jerez S, Montavez JP, Jimenez-Guerrero P, Gomez-Navarro JJ, Lorrente-Plazas R, Zorita E (2012) A multi-physics ensemble of present-day climate regional simulations over the Iberian Peninsula. *Clim Dyn* 40:3023–3046. doi:10.1007/s00382-012-1539-1

- Jones DA, Trewin BC (2000) On the relationships between the El Niño–Southern oscillation and Australian land surface temperature. *Int J Climatol* 20:697–719
- Kala J, Andrys J, Lyons TJ, Foster IJ, Evans BJ (2015) Sensitivity of WRF to driving data and physics options on a seasonal timescale for the southwest of Western Australia. *Clim Dyn* 44:633–659. doi:10.1007/s00382-014-2160-2
- Konopac KS, Akbari H (2002) Energy savings for heat island reduction strategies in Chicago and Houston (including updates for baton rouge, Sacramento, and Salt Lake city). Draft final report, LBNL-49638. University of California, Berkeley
- Kunkel KE, Changnon SA, Reinke BC, Arritt RW (1996) The July 1995 heat wave in the midwest: a climatic perspective and critical weather factors. *Bull Am Meteorol Soc* 77:1507–1518. doi:10.1175/1520-0477(1996)
- Lee SH, Kim SW, Angevine WM, Bianco L, McKeen SA, Senff CJ, Trainer M, Tucker SC, Zamora RJ (2011) Evaluation of urban surface parameterizations in the WRF model using measurements during the Texas Air Quality Study 2006 field campaign. *Atmos Chem Phys* 11:2127–2143
- Li D, Bou-Zeid E (2013) Synergistic interactions between urban heat islands and heat waves: the impact in cities is larger than the sum of its parts. *J Appl Meteorol Climatol* 52:2051–2064
- Marshall AG, Hudson D, Wheeler MC, Alves O, Hendon HH, Pook MJ, Risbey JS (2014) Intra-seasonal drivers of extreme heat over Australia in observations and POAMA-2. *Clim Dyn* 43(7–8):1915–1937. doi:10.1007/s00382-013-2016-1
- Min SK, Cai W, Whetton P (2013) Influence of climate variability on seasonal extremes over Australia. *J Geophys Res Atmos* 118(2):643–654
- Miralles DG, Teuling AJ, van Heerwaarden CC, Vilà-Guerau de Arellano J (2014) Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat Geosci* 7:345–349. doi:10.1038/ngeo2141
- Mlawer EJ, Taubman SJ, Brown PD, Iacono MJ, Clough SA (1997) Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *J Geophys Res Atmos* 102:16663–16682. doi:10.1029/97JD00237
- Mooney PA, Mulligan FJ, Fealy R (2013) Evaluation of the sensitivity of the weather research and forecasting model to parameterization schemes for regional climates of Europe over the period 1990–95. *J Clim* 26:1002–1017. doi:10.1175/jcli-d-11-00676.1
- Nairn J, Fawcett R (2013) Defining heatwaves: heatwave defined as a heat impact event servicing all community and business sectors in Australia. The Centre for Australian Weather and Climate Research. http://www.cawcr.gov.au/technical-reports/CTR_060.pdf. Accessed 22 Nov 2016
- Nicholls N, Larsen S (2011) Impact of drought on temperature extremes in Melbourne, Australia. *Aust Meteorol Oceanogr J* 61:113–6
- Nielsen-Gammon J, Powell C, Mahoney M, Angevine W, Senff C, White A, Berkowitz C, Doran C, Knupp K (2008) Multisensor estimation of mixing heights over a coastal city. *J Appl Meteorol Climatol* 47:27–43
- Parker TJ, Berry GJ, Reeder MJ (2013) The influence of tropical cyclones on heat waves in Southeastern Australia. *Geophys Res Lett* 40:6264–6270
- Parker TJ, Berry GJ, Reeder MJ (2014) The structure and evolution of heat waves in Southeastern Australia. *J Clim* 27:5768–5785. doi:10.1175/JCLI-D-13-00740.1
- Perkins SE (2015) A review on the scientific understanding of heatwaves—their measurement, driving mechanisms, and changes at the global scale. *Atmos Res* 164(165):242–267
- Perkins SE, Alexander LV (2013) On the measurement of heat waves. *J Clim* 26:4500–4517
- Perkins SE, Argüeso D, White CJ (2015) Relationships between climate variability, soil moisture, and Australian heatwaves. *J Geophys Res Atmos* 120(81):44–64
- Pezza AB, Van Rensch P, Cai W (2012) Severe heat waves in Southern Australia: synoptic climatology and large scale connections. *Clim Dyn* 38(1–2):209–224
- Pitman AJ (2003) The evolution of, and revolution in, land surface schemes designed for climate models. *Int J Climatol* 23:479–510. doi:10.1002/joc.893
- Pleim JE (2007) A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: application and evaluation in a mesoscale meteorological model. *J Appl Meteorol Climatol* 46:1396–1409. doi:10.1175/JAM2534.1
- Rosenfeld AH, Akbari H, Romn JJ, Pomerantz M (1998) Cool communities: strategies for heat island mitigation and smog reduction. *Energy Build* 28:51–62
- Rosenzweig C, Solecki WD, Parshall L, Chopping M, Pope G, Goldberg R (2005) Characterizing the urban heat island in current and future climates in New Jersey. *Glob Environ Change Part B Environ Hazards* 6:51–62. doi:10.1016/j.hazards.2004.12.001
- Russell A, Dennis R (2000) NARSTO critical review of photochemical models and modeling. *Atmos Environ* 34:2283–2324. doi:10.1016/S1352-2310(99)00468-9
- Salamanca F, Martilli A, Yague C (2011) A numerical study of the urban heat island over Madrid during the DESIREX (2008) campaign with WRF and an evaluation of simple mitigation strategies. *Int J Climatol* 32:2372–2386
- Schoetter R, Cattiaux J, Douville H (2015) Changes of western European heat wave characteristics projected by the CMIP5 ensemble. *Clim Dyn* 45:1601–1616. doi:10.1007/s00382-014-2434-8
- Shi JJ, Matsui T, Tao WK, Lidard CP, Chin M, Tan Q, Kemp E (2014) Implementation of an aerosol cloud microphysics radiation coupling into the NASA unified WRF: simulation results for the 67 August 2006 AMMA special observing period. *Q J R Meteorol Soc* 140:2158–2175. doi:10.1002/qj.2286
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Wang W, Power JG (2005) A description of the advanced research WRF version 2. NCAR technical note, NCAR/TND468 + STR. National Center for Atmospheric Research, Boulder
- Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang X-Y, Wang W, Powers JG (2008) A description of the advanced research WRF version 3. NCAR technical note NCAR/TN-475 + STR, National Center for Atmospheric Research, Boulder
- Srikanth M, Satyanarayana ANV, Srinivas CV, Kumar MK (2015) Mesoscale atmospheric flow-field simulations for air quality modelling over complex terrain region of Ranchi in eastern India using WRF. *Atmos Environ* 107:315–328
- Srinivas CV, Venkatesan R, Bagavath Singh A (2007) Sensitivity of mesoscale simulations of land–sea breeze to boundary layer turbulence parameterization. *Atmos Environ* 41:2534–2548. doi:10.1016/j.atmosenv.2006.11.027
- Stegehuis AI, Vautard R, Ciais P, Teuling AJ, Jung M, Yiou P (2013) Summer temperatures in Europe and land heat fluxes in observation-based data and regional climate model simulations. *Clim Dyn* 41:455–477. doi:10.1007/s00382-012-1559-x
- Stegehuis AI, Vautard R, Ciais P, Teuling AJ, Miralles DG, Wild M (2015) An observation-constrained multi-physics WRF ensemble for simulating European mega heat waves. *Geosci Model Dev* 8:2285–2298. doi:10.5194/gmd-8-2285-2015
- Sukoriansky S, Galperin B, Perov V (2005) Application of a new spectral theory of stably stratified turbulence to the atmospheric boundary layer over ice. *Bound Layer Meteorol* 117:231–257
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res* 106:7183–7192

- Tewari MF, Chen F, Kusaka H, Miao S (2007) Coupled WRF/unified Noah/urban-canopy modeling system. NCAR WRF documentation. NCAR, Boulder, pp 1–20. <https://ral.ucar.edu/sites/default/files/public/product-tool/WRF-LSM-Urban.pdf>. Accessed 13 Dec 2016
- Users' s Guide for the Advanced Research WRF (ARW) (2015) Modeling system version 3.7 (April 2015). WRF user's webpage: http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3.7/ARWUsersGuideV3.7.pdf. Accessed 11 Oct 2016
- Willmott CJ, Matsuura K, Robeson SM (2009) Ambiguities inherent in sums-of-squares-based error statistics. *Atmos Environ* 43:749–752. doi:10.1016/j.atmosenv.2008.10.005
- Zempila MM, Giannaros T, Bais A, Melas D, Kazantzidis A (2016) Evaluation of WRF shortwave radiation parameterizations in predicting global horizontal irradiance in Greece. *Renew Energy* 86:831–840