Skill-Testing Chemical Transport Models across Contrasting Atmospheric Mixing States Using Radon-222

Scott D. Chambers [1,2,*], Elise-Andree Guérette [2], Khalia Monk [3], Alan D. Griffiths [1,2], Yang Zhang [4], Hiep Duc [3], Martin Cope [5], Kathryn M. Emmerson [5], Lisa T. Chang [3], Jeremy D. Silver [6], Steven Utembe [6], Jagoda Crawford [1], Alastair G. Williams [1,2] and Melita Keywood [5]

[1] ANSTO, Environmental Research, Locked Bag 2001, Kirrawee DC, NSW 2232, Australia; agf@ansto.gov.au (A.D.G.); jc@ansto.gov.au (J.C.); agw@ansto.gov.au (A.G.W.)
[2] Centre for Atmospheric Chemistry, Chemistry, Science Medicine and Health, University of Wollongong, Wollongong, NSW 2500, Australia; eag873@uowmail.edu.au
[3] New South Wales Office of Environment and Heritage, Sydney, NSW 2000, Australia; Khalia.Monk@environment.nsw.gov.au (K.M.); Hiep.Duc@environment.nsw.gov.au (H.D.); LisaTzu-Chi.Chang@environment.nsw.gov.au (L.T.C.)
[4] Department of Marine, Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695, USA; yzhang9@ncsu.edu
[5] Climate Science Centre, CSIRO Oceans and Atmosphere, Aspendale, VIC 3195, Australia; Martin.Cope@csiro.au (M.C.); Kathryn.Emmerson@csiro.au (K.M.E.); Melita.Keywood@csiro.au (M.K.)
[6] School of Earth Sciences, University of Melbourne, Melbourne, VIC 3052, Australia; jeremy.silver@unimelb.edu.au (J.D.S.); sutembe@unimelb.edu.au (S.U.)
* Correspondence: szc@ansto.gov.au; Tel.: +61-2-9717-3058

Abstract: We propose a new technique to prepare statistically-robust benchmarking data for evaluating chemical transport model meteorology and air quality parameters within the urban boundary layer. The approach employs atmospheric class-typing, using nocturnal radon measurements to assign atmospheric mixing classes, and can be applied temporally (across the diurnal cycle), or spatially (to create angular distributions of pollutants as a top-down constraint on emissions inventories). In this study only a short (<1-month) campaign is used, but grouping of the relative mixing classes based on nocturnal mean radon concentrations can be adjusted according to dataset length (i.e., number of days per category), or desired range of within-class variability. Calculating hourly distributions of observed and simulated values across diurnal composites of each class-type helps to: (i) bridge the gap between scales of simulation and observation, (ii) represent the variability associated with spatial and temporal heterogeneity of sources and meteorology without being confused by it, and (iii) provide an objective way to group results over whole diurnal cycles that separates 'natural complicating factors' (synoptic non-stationarity, rainfall, mesoscale motions, extreme stability, etc.) from problems related to parameterizations, or between-model differences. We demonstrate the utility of this technique using output from a suite of seven contemporary regional forecast and chemical transport models. Meteorological model skill varied across the diurnal cycle for all models, with an additional dependence on the atmospheric mixing class that varied between models. From an air quality perspective, model skill regarding the duration and magnitude of morning and evening "rush hour" pollution events varied strongly as a function of mixing class. Model skill was typically the lowest when public exposure would have been the highest, which has important implications for assessing potential health risks in new and rapidly evolving urban regions, and also for prioritizing the areas of model improvement for future applications.

Keywords: radon; atmospheric stability; air quality; model evaluation; modelling; SNBL; ABL; urban pollution

1. Introduction

The population density of urban centers is rising globally and source strengths of fine particles to which residents are exposed scales directly with population [1]. Given the established links between aerosols and acute or chronic health effects (e.g., [2–5]), and the multitude of direct sources and pathways for secondary production in complex urban environments [6,7], there is growing scientific and political interest to: (i) provide improved health risk assessments, (ii) develop and evaluate emission mitigation measures, and (iii) provide realistic forecasting of air quality for future urban development scenarios [8,9].

Considering the scale and complexity of the problem, state-of-the-art chemical transport models (CTMs) linked to representative emissions inventories and driven by reliable prognostic meteorological models, potentially offer the most cost-effective means of providing valuable (if approximate) inputs to these problems in a timely manner. To this end, as well as updating emissions inventories, CTMs need to be continually improved and evaluated to reflect advancements in understanding of contributing physical and chemical processes, as well as computational ability. Targeted, highly collaborative measurement campaigns are a crucial part of this process, since they help drive advancements in understanding, provide the means by which to evaluate model performance, and enable top down 'sanity checks' of emissions inventories.

Skill-testing CTMs serves a dual purpose: on one hand it provides assurance (or otherwise) of the efficacy of incremental modifications, while on the other hand it provides valuable information on the level of confidence that can be placed in various representations and timescales of CTM output. As much as it is true for CTMs themselves, the ways in which their performance is evaluated should also be subjected to periodic review. Mismatches in spatial and vertical scales between observations and model grid-cell dimensions, along with spatial and temporal heterogeneity of urban pollution sources, can lead to vastly-different assessments of model skill depending on the timescale or particular approach adopted for the evaluation [10–12].

The aim of this investigation is to introduce a new approach for skill-testing CTMs regarding: (i) reproducing meteorological conditions, (ii) transporting and mixing primary emissions in the atmospheric boundary layer (ABL), and (iii) forming secondary gaseous and aerosol products. Since understanding of the physical processes associated with transport and mixing under fair-weather turbulent daytime conditions is more advanced than processes dominating near quiescent nocturnal conditions near the surface [12–16], model skill will also be dependent on atmospheric mixing state. Furthermore, the formation of secondary pollutants is also dependent on factors related to the atmospheric mixing state (e.g., mixing volume, temperature, sunlight intensity, humidity, etc.).

Continuous monitoring of atmospheric Radon-222 (radon) is known to be a convenient, economical, and highly-effective alternative to conventional meteorological approaches (e.g., Pasquill-Gifford typing, the Bulk Richardson method, Monin-Obukhov Similarity Theory) for classifying the nocturnal-mean (as opposed to hourly) atmospheric mixing state (e.g., [13,17–20]). Furthermore, with the exception of highly non-stationary synoptic conditions (i.e., ~80% of the time), the average mixing state of the daytime period following each classified night can usually be inferred, based on an assumption of short-term atmospheric persistence [19]. This enables objective meteorological class-typing to be performed simply and efficiently for large datasets over whole 24-h periods.

To a good approximation, radon is emitted only from unsaturated, unfrozen terrestrial surfaces, and this surface source function can be considered well-distributed and consistent (on greater than synoptic timescales). Since it is unreactive and poorly soluble, radon's sole atmospheric sink is radioactive decay. Its half-life (3.82 days) is substantially longer than mixing timescales in the ABL,

long enough for it to be considered conservative over a single night, yet short enough that it does not accumulate in the atmosphere on timescales of more than two weeks. This combination of physical characteristics ensures that near-surface concentration gradients are directly representative of the combined physical atmospheric mixing processes, making radon an effective and unambiguous tracer of transport and vertical mixing (e.g., [13,17,20–22]). Furthermore, atmospheric class-typing by radon-based methods yields a more even distribution of events between mixing categories than the categorical Pasquill-Gifford approaches, and is also more selective and representative of extremely stable conditions (e.g., [18,23]).

For sufficiently long measurement campaigns (e.g., >2 weeks) the radon-based technique enables the definition of 3–5 class types within which in-group variability of the mixing state hour by hour within the diurnal cycle is minimized. This enables comparison of observed and simulated quantities over complete diurnal cycles in each class-type, for which averaging within hourly bins reduces the observed short-term variability associated with spatial/temporal source heterogeneity and mesoscale atmospheric variability, thereby helping to 'bridge the scale gap'. Assessing model skill in this way promises to provide a more targeted approach to continual CTM improvement.

The evaluation dataset for this study was selected from observations made as part of the second Sydney Particle Study (SPS; [8,9,24]). Specifically, meteorology, particle precursor trace gases, and fine particles measured at Westmead (NSW, Australia) in autumn 2012 are compared with corresponding values simulated by seven models operated by ANSTO, CSIRO, the NSW Office of Environment and Heritage (NSW OEH), the University of Melbourne, and North Carolina State University. The model runs from which the results for this study have been selected formed part of a larger, more comprehensive inter-comparison/evaluation exercise (see [25–27]).

The contrasting impacts of sub-grid scale surface features and meteorological processes often result in the perceived "success" of model-measurement inter-comparisons being highly site dependent. Consequently, the potential utility of our evaluation approach, rather than the specific performance of any one of the models at the chosen site, is the main focus. We note that the Westmead site does not conform to the standards of Bureau of Meteorology observational sites (regarding exposure, fetch homogeneity, etc.; [28]). This will result in less than optimum comparisons between observed and simulated meteorological quantities. Detailed evaluations at more appropriate sites are the subject of separate investigations (e.g., [25,26,29–31]).

2. Experiment and Methods

2.1. Site and Measurement Campaign

The Sydney Particle Study was a comprehensive observation and modelling study of fine particles and selected gaseous urban pollutants in the Sydney Basin, Australia [8,9,24]. While SPS observations were made at numerous sites across the Sydney Basin Region (Figure 1), the focal point was the Westmead Air Quality Station, 26 km northwest of the Sydney Central Business District (CBD). Measurements were conducted over two 1-month periods of contrasting air quality characteristics: summer and autumn. The first stage of observations (SPS-I) spanned the period 5 February to 7 March 2011, and the second stage (SPS-II) from 16 April to 14 May 2012. A collaboration between seven partner organizations (CSIRO, NSW OEH, NSW EPA, ANSTO, QUT, BOM, SINAP) ensured that a comprehensive suite of air quality measurements were made to better understand the sources, chemical composition, and size distribution of the aerosol and gas-phase secondary aerosol precursors in the Sydney region.

The aim of SPS was to furnish NSW OEH and NSW EPA with an improved understanding and a quantitative model of gaseous and particulate pollutants that contribute to public exposure in the Sydney Basin Region. Information of this kind is crucial to the development of State and National policy for reducing levels of public exposure to urban pollutants [9]. An overview of SPS-I and its outcomes has been given by Keywood et al. [8]. Here we focus on observations and simulations from

only SPS-II, during which peak pollutant concentrations were often 1.5–3 times higher than observed during SPS-I [9].

Cope et al. [9] provide a comprehensive description of the meteorological, trace gas, particulate, and atmospheric radon measurements made during SPS-II. Wind speed and direction were measured at 10 m above ground level (a.g.l.), temperature and relative humidity were measured at 2 m a.g.l., whereas air quality observations were made ~5 m a.g.l. Simulated meteorological quantities have been calculated to match the observation heights, whereas simulated air quality parameters are averages over the lowest layer of each model (which varied from z_1 = 20 to 56 m; [25]).

In addition to the Westmead observations, radon was also measured at the University of Western Sydney Richmond campus (~30 km NW of Westmead; Figure 1), adjacent to the Richmond NSW OEH air quality monitoring site (see [17] for details). Richmond (both the NSW OEH site and RAAF Base site ~2.5 km to the northeast), better satisfies BoM observational requirements. Consequently, selected comparisons will also be made using data from this site. All findings presented here are based on radon-based classification of atmospheric stability (see Section 2.2) made separately at the Westmead and Richmond sites using the respective radon observations.
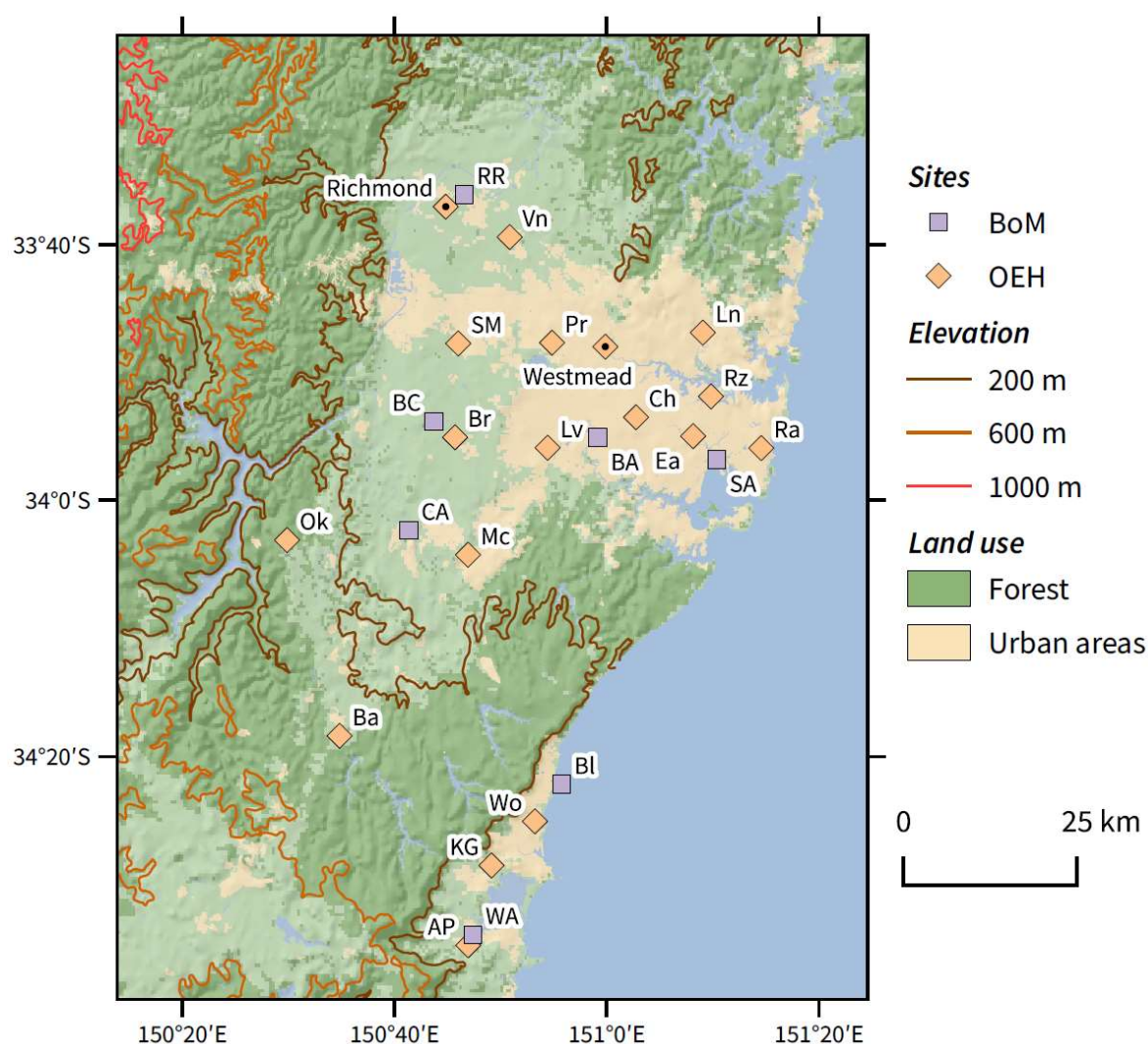
Figure 1. Location of monitoring sites used for the SPS campaigns in the Greater Sydney Region (including those operated by BoM and NSW OEH). A key for site abbreviations is provided at the end of this document. Land cover is derived from the MODIS satellite data [32], topography from Geoscience Australia [33], and basemap © OpenStreetMap contributors.

To assist with the evaluation of simulated ABL depths, a lidar was situated at Westmead during SPS-II. The instrument deployed (Leosphere, model ALS-300) operated in the ultraviolet (355 nm) with a repetition rate of 20 Hz and a single detection channel. The derivation of lidar mixing heights is explained in Griffiths et al. [34] and Cope et al. [9]. The lowest resolvable mixing height is determined by the overlap between the detector and transmitter fields of view. Typically, the most reliable lidar-based mixing depth estimates were available each day between 0900 and 1900 h LST. Outside these times, when mixing depths were often below 200 m a.g.l., we estimated the "equivalent nocturnal mixing height" using a radon-based method described in Griffiths et al. [34]; assuming a local radon flux of 20 mBq m^{-2} s^{-1} [35]. Since this approach assumes radon to be uniformly mixed within the stable nocturnal boundary layer (SNBL), which is rarely the case, interpretation of nocturnal mixing depths should be performed with due caution.

2.2. Radon-Based Stability Classification

Although radon does not accumulate in the atmosphere on greater than synoptic timescales, its half-life provides air masses with a ~2-week 'memory' of fetch influences. Consequently, near-surface radon observations are primarily a combination of two contributing factors: (i) an advection (or fetch-related) component, responding to recent changes in air mass fetch history and the related radon source function distribution, and (ii) a diurnal dilution component, driven by a near-constant local terrestrial source function emitting into an ABL of diurnally changing depth.

When the fetch-related component of a radon time series has been separated from the diurnally-varying component, averages of the diurnal radon signal over a 10–11 h nocturnal window can be related to the mean-nocturnal mixing state [17,19,20,23,36,37]. This separation can either be made directly, using vertical radon gradient measurements [13,22], or be approximated using pseudo-gradient estimates based on single-height observations ([19] and references therein). For the sake of brevity, a full description of the stability-classification technique will not be repeated here.

The final step of the stability classification process requires 19:00–05:00 h nocturnal-means of the diurnal radon component to be grouped into mixing classes that, in this case, were based on quartiles of the 1-month Westmead and 6-year Richmond autumn datasets. As described by Chambers et al. [17], the arbitrarily-assigned quartile groups corresponded to four separate relative nocturnal mixing states (most well-mixed; weakly stable; moderately stable; and most stable), also listed in Table 1. At Westmead the assigned mixing states are relative to the range of nocturnal mixing states experienced in just that 1-month of observations. At Richmond, however, the mixing states are relative to the range of nocturnal conditions experienced over 6-years of autumn measurements.

Table 1. Radon-based definition of 24-h relative atmospheric mixing categories.

Quartile	Mixing Category	Assigned Nocturnal Mixing State	Inferred Daytime Mixing State	Expected Meteorological Conditions
Q1	#1	Most well-mixed (near neutral)	Near neutral	Windiest nights, includes periods of severe synoptic non-stationarity. Frequent periods of extensive low cloud cover, rainfall common.
Q2	#2	Weakly stable	Weakly unstable	Moderate winds at night, periods of low clouds and occasional rainfall
Q3	#3	Moderately stable	Moderately unstable	Light winds at night, little-to-no low cloud, usually no rainfall.
Q4	#4	Most stable	Most unstable	Typically anticyclonic conditions. Calm-to-light nocturnal winds, very little low cloud at night (though fog possible). No rain at night, possible daytime convective showers.

Based on the assigned nocturnal mixing categories, and an assumption of short-term atmospheric persistence, daytime mixing states were then inferred and assigned to the remainder of each 24-h period (defined from afternoon to afternoon; 1500 h to 1400 h). As shown in Table 1 each 24-h period was assigned a mixing category identifier (#1 through #4), associated with a pair of nocturnal/daytime mixing categories.

At Westmead the 1-month of 24-h mixing categories was assigned fairly evenly (category #1 to #4: 26%, 26%, 26% and 22%, respectively). At Richmond, however, since mixing categories were defined using the range of conditions from 18 months of autumn measurements, assignment of events between categories for the 1-month of SPS-II observations was not as even (cat #1 to #4: 19%, 26%, 33% and 22%, respectively). These figures represent the fraction of the measurement campaign (in complete 24-h periods) of observed and simulated data contributing to the composite plots presented in Section 3.

The average time over which public exposure to fine particles and other pollutants takes place is critical to developing new health legislation. Skill-testing models using the class-typing method not only provides a clear indication of model performance for a range of air quality conditions, but also clearly indicates the relative potential exposure time represented by each category.

2.3. Summary of Chemical Transport Models

The models compared in this study will henceforth be referred to as: W-A11, W-UM1, W-UM2, C-CTM, O-CTM, W-NC1 and W-NC2. Brief descriptions of each model are provided below. Further information including model setup, descriptions and relevant references for specific parameterizations can be found in Monk et al. [25] and Guérette et al. [26].

Each of the models was run over four nested domains, downscaling from the regional (80–81 km resolution, encompassing all of Australia), to 3 km for the domain centered on the Greater Sydney Region [25,26]. Although the models have consistent horizontal grids, they differ in their number of vertical layers and in the height of their first layer.

All models used the NSW EPA emission inventory for 2008 [38], which may not reflect the actual source distributions and strengths in 2012. Emissions were provided as area sources (i.e., a surface-based emission rate from a 1 km × 1 km area) or point sources (i.e., a specific location with stack properties including the height and radius of the stack, the temperature and velocity of the effluent). The NSW EPA inventory, however, refers to an area that is smaller than the larger modelling domains. Consequently, in addition to this scheme, as discussed by Guérette et al. [26], W-UM1, W-UM2, W-NC1 and W-NC2 used additional emissions information from EDGAR for parts of the domains not covered by the NSW EPA inventory. Furthermore, C-CTM and O-CTM contain a parameterization for wood smoke emission based on heating degree days, and C-CTM alone included wildfire emissions.

W-A11 is ANSTO's configuration of the Weather Research and Forecasting model coupled to Chemistry (WRF-Chem) version 3.7.1. In this configuration, a custom radon tracer module runs instead of a full chemistry scheme. The radon tracer module computes radon concentration online, based on emissions from Griffiths et al. [35], mixing and advection from the standard WRF-Chem tracer advection scheme (which includes vertical mixing from moist convection), and radioactive decay. The meteorological part of the model uses the same options of W-UM1, described next, except for two things. First, W-A11 uses spectral nudging in the outer domain, which only constrains the large-scale features of the simulation. Second, W-A11 has 50 vertical levels, the lowest of which is 20 m thick.

W-UM1 represents the Community Multiscale Air Quality (CMAQ) model run in an offline mode, forced with meteorology from WRF (v3.6.1; [39]) at the University of Melbourne. CMAQ is a multi-scale, limited-area, open-source, chemistry-transport model developed and maintained by the US EPA [40]; version 5.0.2 was used for these simulations [41]. Both models used a vertical grid with 32 levels (z_1 = 33 m). WRF dynamical simulations used ERA-Interim reanalyses [42] and high resolution "Real-Time Global" sea surface temperatures [43]. WRF was run as a series of 36-h forecasts, with the first 12 h of spin-up discarded, during which simulations were nudged towards the global

analysis above the boundary layer in the outermost domain every 6 h. Physics parameterizations used included the Morrison 2–moment microphysics scheme [44], the Rapid Radiative Transfer (RRTMG) shortwave and longwave radiation schemes [45], the unified NOAH land surface model [46], the Mellor–Yamada–Janjic turbulence scheme [47], the Grell 3D Ensemble convection scheme for domains 1–3 [48,49] and no convection parameterization for the 3 km domain. CMAQ used initial and boundary concentrations based on global simulations from the Model for Ozone and Related chemical Tracers version 4 (MOZART) model [50]. Anthropogenic emissions beyond the greater Sydney region were based on EDGAR [51]. Natural emissions were parameterized, using the Model of Emissions of Gases and Aerosols from Nature (MEGAN) biogenic emissions scheme [52], CMAQ's online schemes for sea salt [53] and wind-blown dust [54]. Fire emissions were remapped from estimates from the ECMWF/CAMS GFAS [55]. For gas-phase chemistry, the Carbon Bond 5 (CB05) mechanism with active chlorine chemistry was used, with the updated toluene mechanism [56,57]. The aerosol module selected was the sixth-generation modal CMAQ aerosol model ('Aero6') with extensions for sea salt emissions and thermodynamics, with an updated formulation for secondary organic aerosol yields [54]. Clear-sky photolysis rates were calculated offline and stored in look-up tables [58]. Furthermore, radon was added as a non-reactive tracer subject to radioactive decay, with emissions calculated based on geology and surface moisture [35].

W-UM2 is based on version 3.7.1 of WRF-Chem operated by the University of Melbourne. As discussed in Utembe et al. [29], WRF-Chem is an online coupled model, here run in nudged mode, using the Advanced Research WRF (ARW) dynamic core. The model has 33 levels, with increased vertical resolution in the boundary layer. Meteorological fields, initial and lateral boundary conditions were taken from ERA Interim re-analyses. Sea surface temperatures were taken from NOAA real-time global sea surface temperature analyses. Initial and lateral boundary conditions for the mixing ratio fields were obtained from the global-scale MOZART. The WRF-Chem implementation for this study used the NOAH land surface model, the Yonsei University (YSU) ABL scheme, the WRF single-moment 5-class Lin microphysics scheme and the RRTMG radiation scheme. The Regional Atmospheric Chemistry Mechanism (RACM) of Stockwell et al. [59] is used for the gas phase chemistry scheme (with F-TUV photolysis scheme) and for the aerosol scheme it uses the Modal Aerosol Dynamics Model for Europe (MADE) which is coupled to a Secondary Organic Aerosol Module (SORGAM). Outside the Greater Sydney area EDGAR-HTAP (2008) emissions were used [60]. Dust and sea-salt emissions are proportional to wind speed at 10 m as parameterized from the GOCART model. Biogenic emissions are generated from MEGAN v2.1 using the default 24 category land use data from U.S. Geological Survey (USGS).

C-CTM is the three-dimensional Eulerian CSIRO CTM, developed for Australian regional air quality studies [61,62]. The model has 35 levels in the vertical to a height of 40 km. Meteorology derives from the Conformal Cubic Atmospheric Model (CCAM; [63]), a global stretched grid dynamical model, predicting wind velocity, temperature, water vapor mixing ratio (including clouds), radiation and turbulence. CCAM uses the Australian land surface scheme, CABLE [64]. Biogenic emissions are provided by the Australian Biogenic Canopy and Grass Emissions Model (ABCGEM; [65]). Gas-phase chemistry is modelled using an extended version of the CB05 mechanism with updated toluene chemistry. Aerosol concentrations are calculated using a two-bin sectional scheme, using the volatility basis set for the secondary organic species partitioning, and ISORROPIA II for the inorganic partitioning.

O-CTM here represents a NSW OEH modelling system consisting of a meteorology module (CCAM), emission module, and a CTM [30]. CCAM is a semi-implicit, semi-Lagrangian atmospheric climate model based on the conformal cubic grid [63,66]. In this study, the European Reanalysis Interim (ERA-Interim) reanalysis was the host global climate model (GCM) and data was fed into CCAM to downscale into the four Australian grid domains mentioned above. A configuration of 35 vertical levels in the CCAM was also employed. Emissions were considered from natural sources, including: wind-blown dust, and marine aerosols, influenced by the prevailing meteorology. Emissions of VOC

from vegetation were also considered, calculated in-line within the CTM. Anthropogenic emissions were taken from the NSW GMR Air Emissions Inventory for calendar year 2008 [38]. This inventory comprises detailed source and emissions data for over a hundred industrial, commercial, transport, agricultural and residential activities and over a thousand pollutants. The inventory is updated every five years, with the 2013 calendar year emissions inventory not released at the time the modelling study was undertaken. The CSIRO CTM [61] is a three-dimensional Eulerian chemical transport model with the capability of modelling the emission, transport, chemical transformation, wet and dry deposition of a coupled gas and aerosol phase atmospheric system. This model has 16 vertical layers, the lowest of which has a thickness $z_1 = 20$ m.

W-NC1 is based on the North Carolina State University (NCSU)'s version of WRF-Chem v3.7.1. The vertical resolution includes 34 layers from the surface ($z_1 = 35$ m) to approximately 100 hPa (~15 km). The physics options selected for the WRF-Chem simulation include the RRTMG for both shortwave and longwave radiation, the YSU ABL scheme, the Morrison double moment microphysics scheme, as well as the Multi-Scale Kain-Fritsch cumulus parameterization. The chemistry and aerosol option selected is the coupled CB05 gas mechanism with the chlorine chemistry and the Modal for Aerosol Dynamics in Europe/Volatility Basis Set (MADE/VBS) aerosol module, which allows the simulation of aerosol direct, semi-direct, and indirect effect as described in Wang et al. [67].

W-NC2 is based on the NCSU's version of WRF-Chem v3.7.1 coupled with the Regional Ocean Modeling System (ROMS) (WRF-Chem-ROMS) [68]. The WRF-Chem-ROMS simulation uses the same physics and chemistry options and has the same configuration as WRF-Chem with only one difference in the sea surface temperature (SST). SST is prescribed from reanalysis data for the WRF-Chem simulation but simulated by ROMS for the WRF-Chem-ROMS simulation. By explicitly simulating air-sea interactions, WRF-Chem-ROMS can improve the predictions of most cloud and radiative variables compared to WRF-Chem that does not simulate such interactions. More detailed descriptions of both W-NC1 and W-NC2 and their simulation results can be found in Zhang et al. [69].

3. Results and Discussion

3.1. Campaign Overview

For context, a campaign overview of meteorology and air quality is presented in the supplementary information (Figures S1 and S2). Meteorological conditions were variable over the campaign, and at times complex. On 17–19 April a low-pressure trough caused heavy rainfall (temporarily reducing the local radon flux), whereas May was very dry (see also [9]).

Afternoon minimum radon concentrations of <1 Bq m^{-3} indicate limited recent terrestrial influence. Low radon on 18–19 April (Figure S1) coincided with humid SE winds (oceanic fetch) associated with the low-pressure trough, whereas low radon on 25 April was associated with dry fast-moving air from the NNW, with high ozone but low concentrations of other pollutants (Figure S2); consistent with downward mixing of tropospheric air to the ABL. Days with the largest radon diurnal cycles were associated with the most stable nocturnal conditions. The increasing afternoon-minimum radon concentrations for the last week are indicative of increasing terrestrial fetch over eastern Australia.

Carbon monoxide (CO), nitric oxide (NO), sulphur dioxide (SO$_2$) and particulate matter (diameter $\leq$ 10 μm; PM$_{10}$) values (Figure S2) were all lower and less variable during the first half of the campaign, when meteorological conditions were more changeable and nocturnal mixing was stronger. All peak concentrations were well below National Environmental Protection Measure (NEPM) levels (CO 8-h 9000 ppb; NO$_2$ 1-h 120 ppb; SO$_2$ 1-h 200 ppb; O$_3$ 1-h 100 ppb; PM$_{10}$ 1-day 50 μg m^{-3}; [9]).

BoM [28] outlines instrument siting requirements (regarding terrain, obstacles, etc.) for meteorological observations. While Westmead does not meet these requirements, since it was the "super site" for SPS-I and II air quality observations, it is of interest to know how representative simulations are at this location. Cope et al. [9] evaluated C-CTM meteorology at more suitable sites

(i.e., Randwick, Prospect and St Mary's). Generally, good model skill was reported with the exception of transitional conditions (periods of substantial synoptic non-stationarity or in the presence of mesoscale motions). Likewise, Utembe et al. [29] evaluated W-UM2 performance in January 2013 using eight BoM-approved sites and reported better model skill than shown here in Sections 3.2–3.4. Details regarding the more extensive meteorological evaluations of the models discussed here are given in Monk et al. [25].

3.2. Evaluating Simulated Air Temperature

The fidelity with which a model reproduces near-surface temperature is important for meteorological prediction, mixing parameterizations, and kinetics of chemical reactions [70]. It is therefore important to evaluate model meteorological skill prior to investigating air quality parameters. Campaign-mean diurnal composites of observed and simulated temperature, as well as composites of their hourly differences, are shown in Figure 2. With the exception of O-CTM, model skill was typically best in the early afternoons when the ABL was deepest and most uniformly mixed.
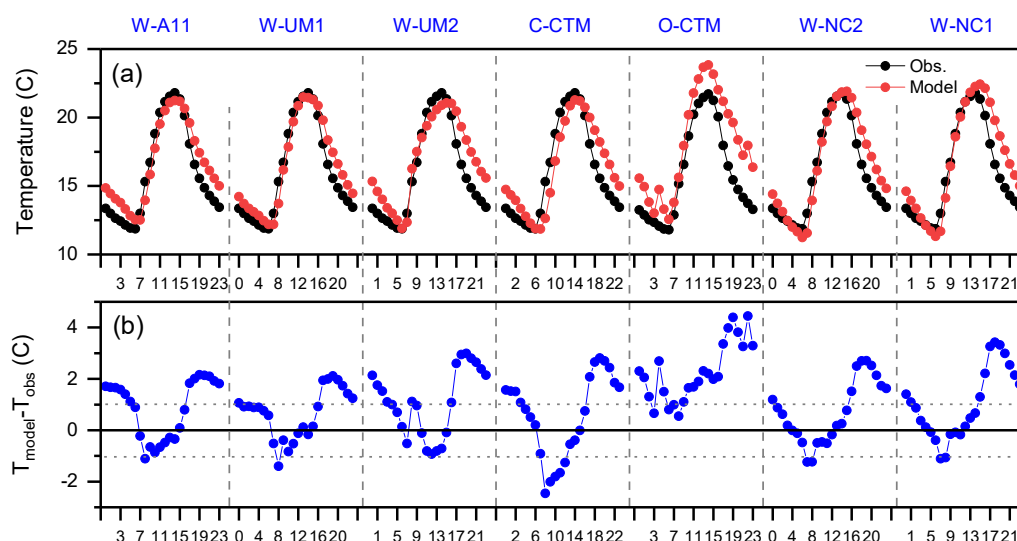
Figure 2. Diurnal composite (**a**) observed and simulated air temperature, and (**b**) hourly differences, over the 27-day SPS-II campaign at Westmead in autumn 2012.

For O-CTM, simulated temperatures were occasionally noisy and the composite diurnal mean bias (T_{MOD}-T_{OBS}) was consistently positive (Figure 2b). In addition to possible differences in configuration options for CCAM between C-CTM and O-CTM, another factor influencing the temperature estimates of these two models is that C-CTM uses the CABLE land-surface model whereas O-CTM used a MODIS-derived scheme [25].

The mean bias (μ_{MB} = <T_{MOD}-T_{OBS}>) and its standard deviation (σ_{MB}) provide a convenient pair of metrics by which to compare model skill for a given parameter [25,71]. To this end, the temperature μ_{MB} and σ_{MB} for each model are shown in Figure 3a. The importance of treating these metrics as a pair is clear in the case of C-CTM, which exhibits the smallest μ_{MB}, but one of the largest σ_{MB} values. The reason for the large σ_{MB} is clear in Figure 2b, where C-CTM exhibits absolute diurnal mean biases in excess of 1 °C at night as well as in the middle of the day. According to these metrics W-UM1 performed the best for SPS-II temperature simulations. The largest reported mean bias of any contributing model here was <1.1 °C (W-UM2; Figure 3a). Given the complexity of the Westmead site, and the benchmark μ_{MB} for temperature of around ±1 °C provided by Monk et al. [25], all models performed acceptably. By comparison, for January 2013 Utembe et al. [29] reported a temperature μ_{MB} of 0.2 °C for W-UM2 averaging over 17 sites in the Sydney region.
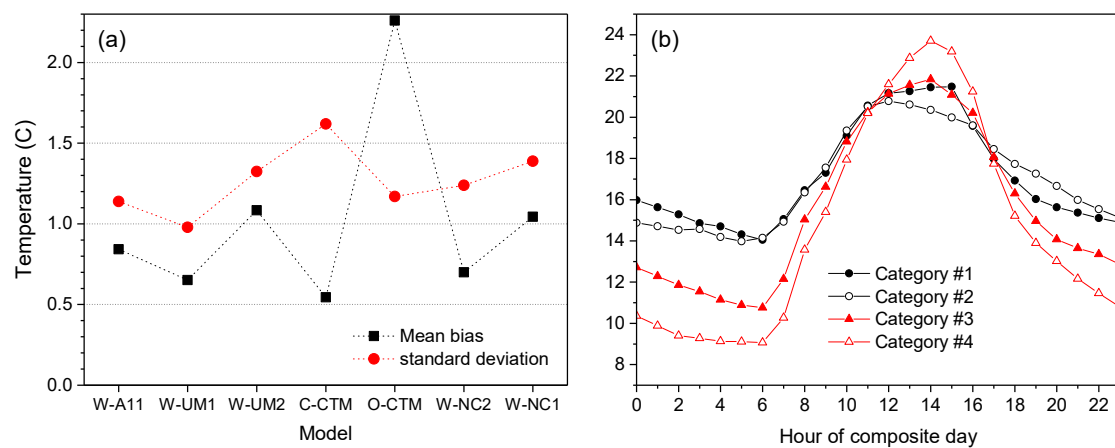
Figure 3. (**a**) Mean bias, and its standard deviation, of hourly observed and simulated 2 m temperature, and (**b**) diurnal composite observed 2 m temperature at Westmead for days in each of the radon-based atmospheric mixing categories (Table 1).

While broadly indicative of model skill generally, the whole campaign μ_{MB} and σ_{MB} values in Figure 3a provide little insight as to the causes of specific differences in model skill, such as why the diurnal mean bias values of C-CTM and O-CTM in Figure 2b are so different to the other contributing models.

It is well established that model performance is not independent of the atmospheric mixing state (or "stability") [12,72–74]. In Figure 2b, for example, across the diurnal cycle agreement between simulated and observed temperatures was usually the best when the ABL was the deepest and most uniformly mixed. Since the behavior of scalars in the atmosphere also varies strongly with stability (e.g., Figure 3b), we sought to evaluate model skill separately for a range of mixing states as summarized in Table 1. To this end, the comparisons of Figure 2a are reproduced in Figure 4 separately for each mixing category. Corresponding μ_{MB} and σ_{MB} values are presented in Figure 5.
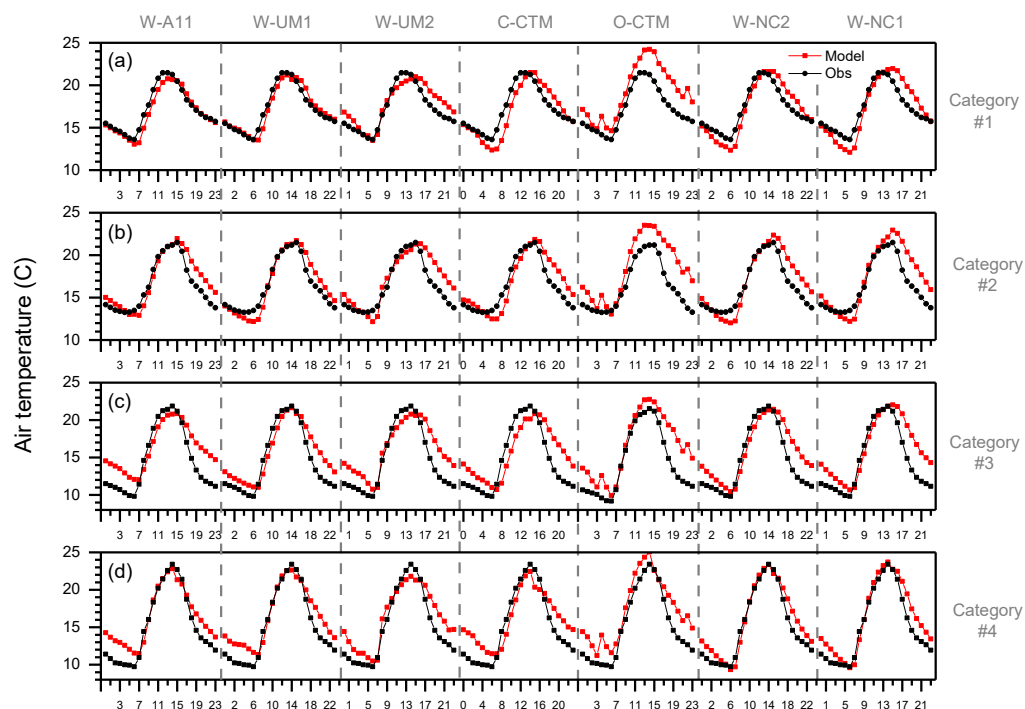
Figure 4. Diurnal composite observed (black) and modelled (red) 2 m temperature at Westmead during SPS-II. Panels a–d represent each of the four radon-based atmospheric mixing categories (Table 1).

Since the curves in Figure 4 are composites of 6–8 separate days under 'similar' conditions (i.e., within defined mixing classes), much of the hourly variability associated with meteorological and local-surface heterogeneity has been averaged out. If required, a measure of the variability within each mixing state can be retrieved by analyzing the hourly distributions within each diurnal composite.

Figure 5 shows there was a tendency in all models for temperature μ_{MB} to increase in absolute magnitude from category #1 to #4 mixing states. However, C-CTM, O-CTM, W-NC1 and W-NC2 show the least sensitivity in μ_{MB} from category #2 to #4 days (note that a proportion of category #1 days represent periods of significant synoptic non-stationarity, whereas conditions within category #2 to #4 days are more internally consistent). There was also a tendency for σ_{MB} to increase from category #1 to #4 conditions, most prominently in W-A11 and C-CTM. Between-model differences in parameterizations that may be contributing to these differences are discussed in Monk et al. [25].
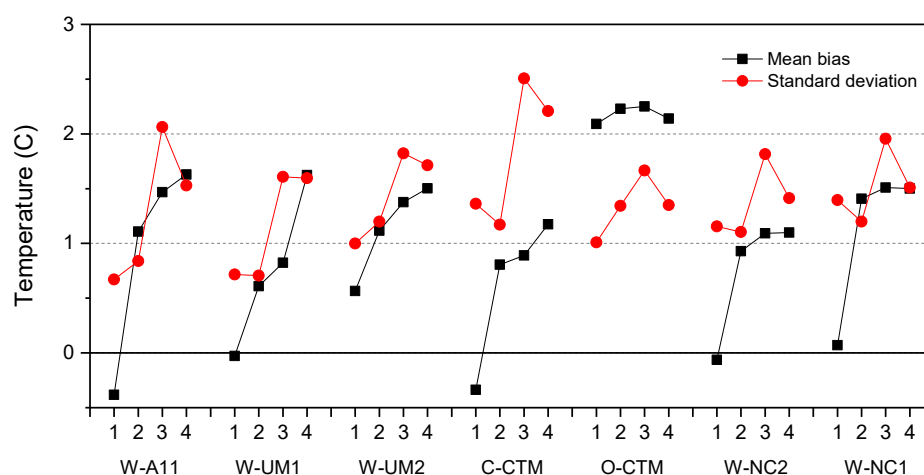
Figure 5. μ_{MB} and σ_{MB} of 2 m temperature at Westmead for each model. Numbers on the abscissa for each model represent the radon-based mixing states (categories #1 to #4) in Table 1.

Based on Figure 4, most of the change in each model's skill is attributable to the simulation of nocturnal temperatures as the nocturnal mixing state changes from well-mixed to stable (Table 1). Here, specific differences in model performance become apparent: (i) W-A11, W-UM1, W-UM2, C-CTM and O-CTM reproduce nocturnal temperatures the best under well mixed conditions, and least well under stable nocturnal conditions; (ii) W-NC1 and W-NC2 on the other hand, underestimate early morning temperatures under well mixed nocturnal conditions but perform better in the mornings under stable conditions. O-CTM is the only model whose skill for temperature prediction in the early afternoon significantly improves from category #1 to category #4 conditions.

3.3. Evaluating Simulated Relative Humidity

A model's ability to accurately reproduce relative humidity is dependent upon the fidelity of its temperature and water vapor predictions, as well as near-surface mixing, and will impact other processes such as convection, chemistry, and secondary aerosol formation.

Diurnal composite relative humidity values are compared in Figure 6 for each mixing category. Corresponding μ_{MB} and σ_{MB} values are shown in Figure 7. As for temperature, there was a tendency for the absolute magnitude of both μ_{MB} and σ_{MB} to increase in all models from category #1 to #4 conditions. In all cases model skill was best across the whole diurnal cycle under the windiest conditions (category #1; Figure 6a). Again, the largest discrepancies were typically observed at night, under increasingly stable nocturnal conditions. Overall, the best performance was exhibited by W-UM1, and the worst by W-UM2 (Figures 6 and 7). The low mean biases of C-CTM in this case for all mixing categories resulted from a combination of under-prediction at night and over-prediction during the day. W-NC2 appeared to perform the best at night for each of the mixing categories, and C-CTM exhibited the highest daytime

humidity values for all mixing categories. Between-model differences in parameterizations that may be contributing to these differences are discussed in Monk et al. [25].
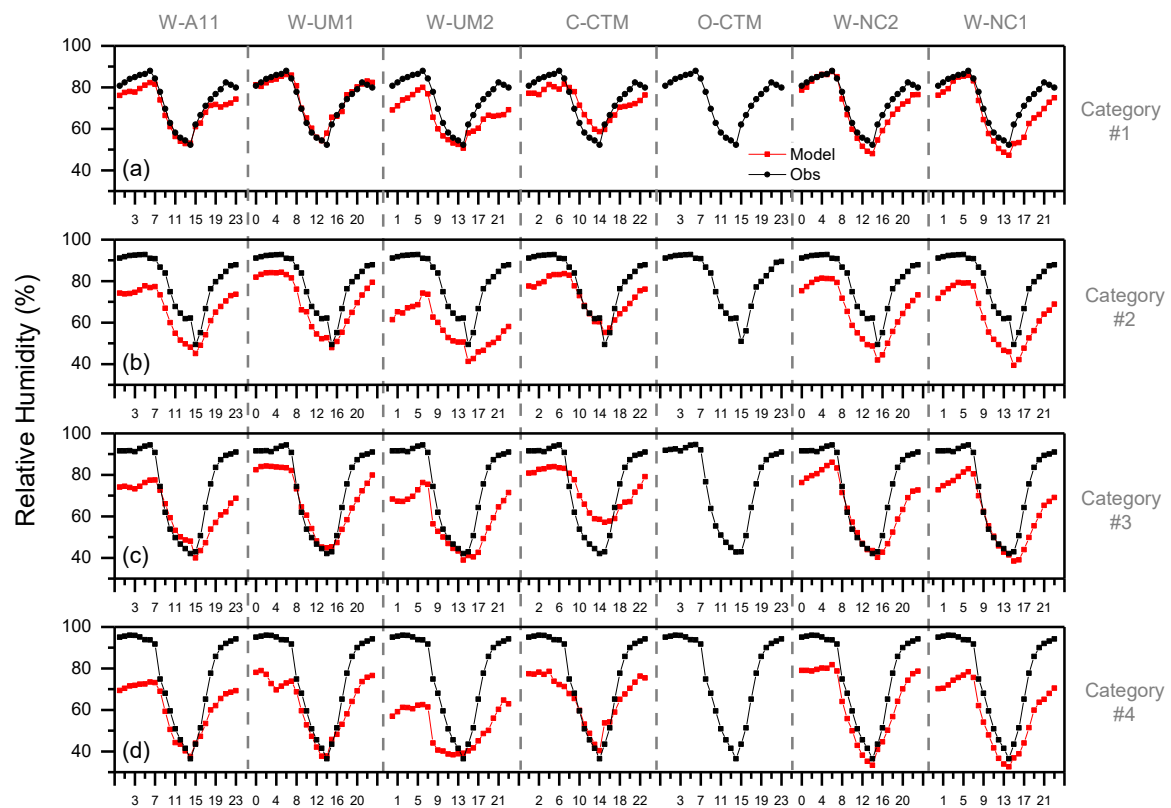
Figure 6. Diurnal composite observed and modelled 2 m relative humidity at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).
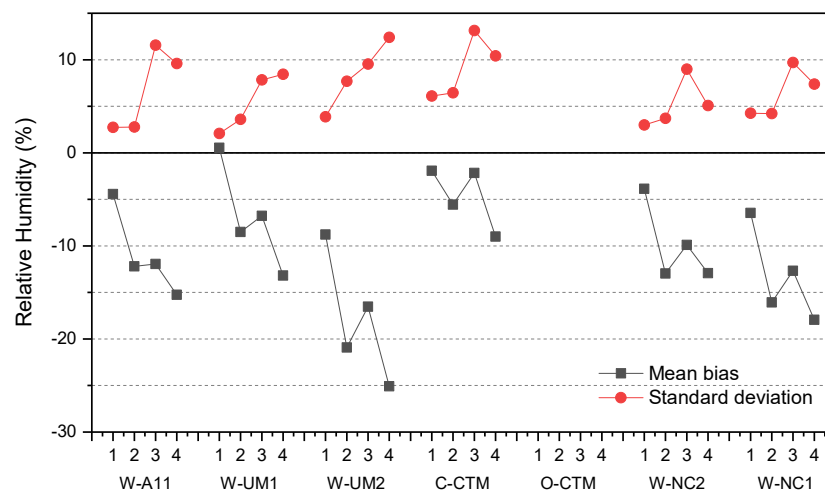
Figure 7. μ_{MB} and σ_{MB} of 2 m relative humidity at Westmead for each model. Numbers on the abscissa for each model represent the radon-based mixing states (categories #1 to #4) in Table 1.

Although W-A11, W-UM1, W-UM2, O-CTM and C-CTM all over-predict pre-dawn temperatures on the most stable nights, whereas W-NC1 and W-NC2 have more accurate simulated temperatures at this time, all models underestimate nocturnal relative humidity values under stable conditions by 10–25% (Figures 6 and 7). Consequently, factors other than the fidelity of near-surface temperature simulations are contributing to differences in simulated relative humidity. Other factors likely

to influence the fidelity of nocturnal relative humidity simulations are: (i) available moisture sources/sinks, and (ii) near-surface wind speed (see next section), particularly with regard to its relationship to the depth of the nocturnal boundary layer (Section 3.5); see Monk et al. [25] for further details.

3.4. Wind Speed

Appropriate simulation of physical processes within a model, such as momentum, heat and water vapor fluxes, as well as the advection and vertical mixing of gaseous and aerosol pollutants, depends critically upon representative simulation of wind speed, particularly in the model layers within the ABL. As noted by Cope et al. [9], variations in land use around Westmead, and associated impacts on surface roughness, are expected to impact the vertical wind profile. Furthermore, almost half of SPS-II was associated with periods of relatively weak synoptic forcing, resulting in less organized flow patterns that are more challenging to model.

10 m wind speeds were overestimated across the diurnal cycle by all models (Figure 8). The lowest campaign-average mean bias values (~1.3 m s^{-1}) were reported by W-NC1 and W-NC2 (Figure 9a), and the highest (~2.1 m s^{-1}) by W-A11. Some of this difference is likely attributable to the non-ideal nature of the site. The lower μ_{MB} of W-NC1 and W-NC2 is due to the use of a surface roughness correction algorithm available in WRF-Chem [69]. According to Monk et al. [25] the benchmark for wind speed bias in complex environments is around ±1.5 m s^{-1}. While this benchmark was not met by several of the contributing models at Westmead (Figure 9a), it was satisfied for model evaluations at more representative BoM sites [9,25,29].
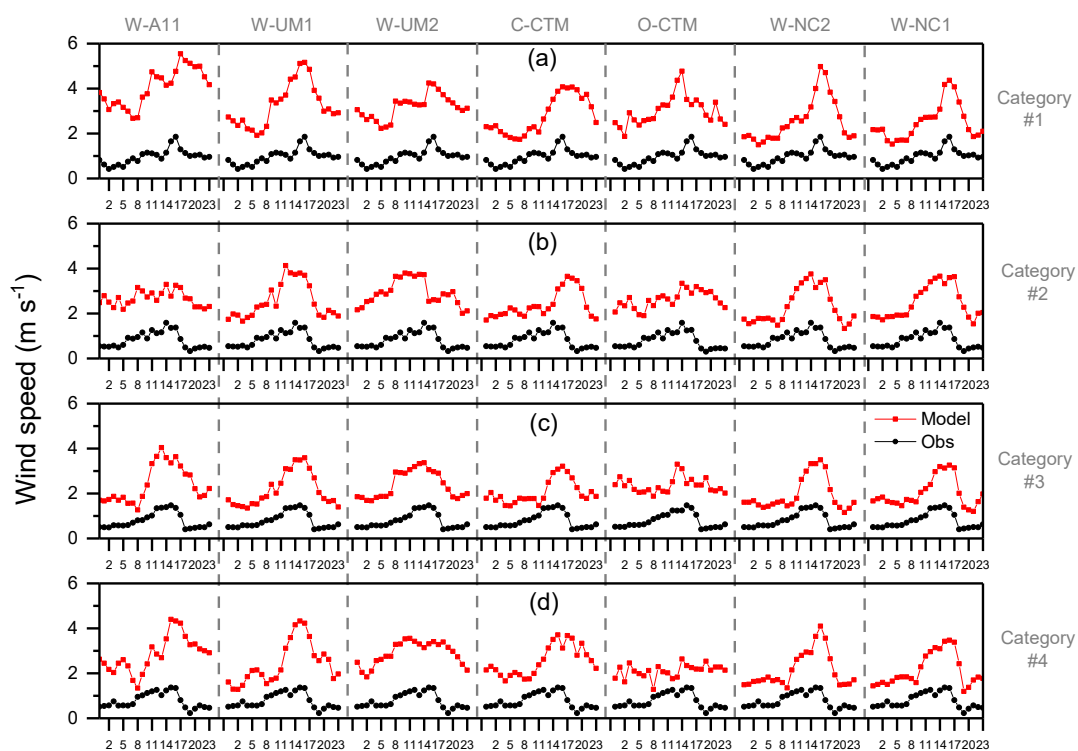
Figure 8. Diurnal composite observed and modelled 10 m wind speed at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).

There was a tendency for μ_{MB} to reduce from category #1 to #4 mixing states (Figure 9b), which was often attributable to slight improvements in morning or daytime simulations (Figure 8). By comparison, σ_{MB} was quite consistent between models and relatively invariant with mixing category.

The models using non-local closure schemes (W-UM2, C-CTM, O-CTM, W-NC1 and W-NC2; [25]) showed the least sensitivity of μ_{MB} to changing mixing state (Figure 9b), and generally had lower σ_{MB} values (Figure 9a).
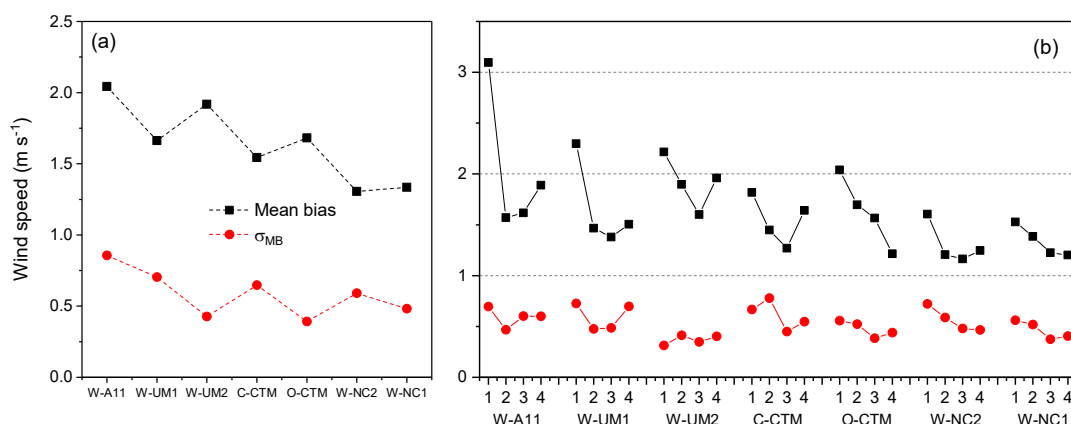
Figure 9. μ_{MB} and σ_{MB} of 10 m wind speed at Westmead for (**a**) the entire SPS-II campaign, and (**b**) each model and radon-based mixing state (categories #1 to #4, Table 1).

Overestimated nocturnal wind speeds would be expected to result in overestimated nocturnal mixing depths, but this was not consistently the case (Section 3.5). Of all the models, W-UM2 overestimated nocturnal wind speeds the most for all stability categories (Figure 8). This model also consistently predicted the deepest nocturnal mixing layers (Figure 10b). Conversely, nocturnal overestimates of wind speed by C-CTM and W-NC2 were similar for all mixing categories, but C-CTM consistently predicted the lowest nocturnal mixing heights.
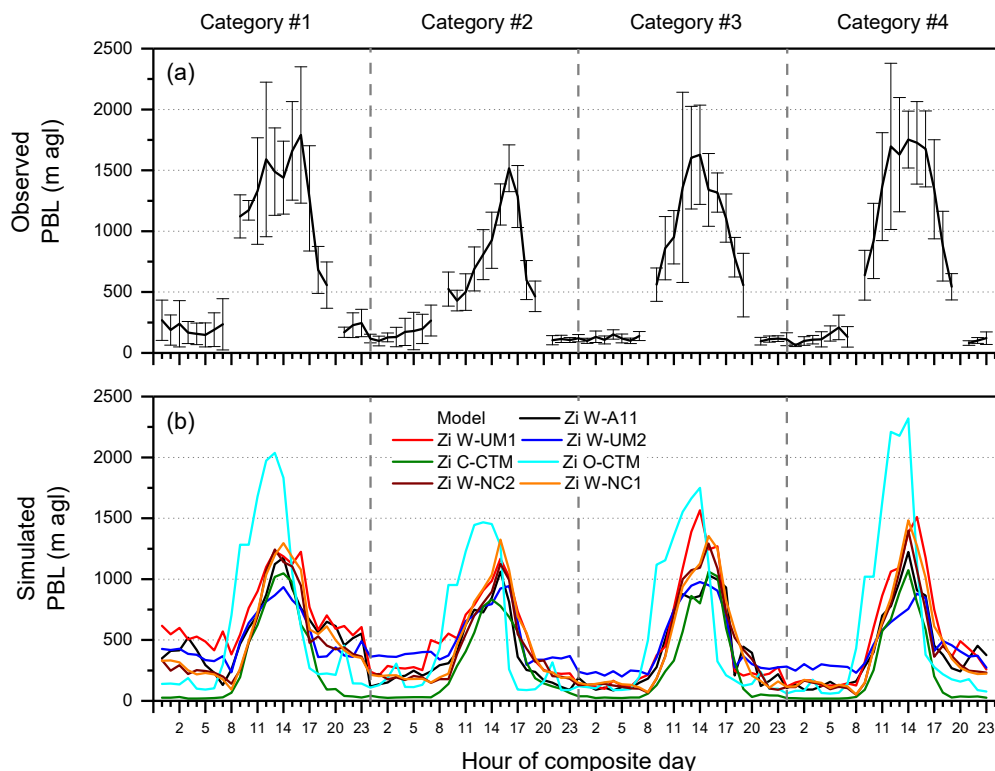
Figure 10. Diurnal composite (**a**) observed (lidar) and estimated (h_e—radon) mixing depths, and (**b**) simulated ABL depths from each model, for each of the radon-derived mixing categories at Westmead for the whole SPS-II campaign.

Under stable nocturnal conditions (category #4) W-UM2 exhibited the largest relative humidity bias (likely attributable to over-predicted nocturnal temperatures and wind speeds (i.e., mixing depth)). C-CTM, on the other hand, demonstrated one of the lowest nocturnal wind speed biases (Figure 8), associated with low nocturnal mixing heights (Figure 10), but still had a comparatively large nocturnal relative humidity bias (Figure 6d). This could, however, be related to the large nocturnal temperature bias of C-CTM compared with W-UM2 under stable conditions (Figure 4d). Under stable conditions, W-NC1 and W-NC2 exhibit the smallest nocturnal wind speed biases, have fairly representative nocturnal mixing depth estimates, and the smallest pre-dawn temperature biases. However, their nocturnal relative humidity bias is comparable to that of W-UM1 and C-CTM.

3.5. Atmospheric Boundary Layer Depth

The depth of the ABL is a crucial parameter as it dictates the volume into which primary emissions are diluted. Resulting concentrations of precursor species, along with various meteorological controls, then dictate reaction rates. Monk et al. [25] use twice daily sonde data (0600 h and 1500 h) from Sydney airport and corresponding profiles determined from each model to compare ABL depths calculated using a bulk Richardson Number method [75]. On average for SPS-II, model ABLs were too high at 6 a.m. (by 6–153 m) and too low at 3 p.m. (by 13–267 m; except W-NC1, +110 m).

Here we compare simulated ABL depths with values determined using a combination of lidar and radon-based techniques (Section 2.1). For details regarding different parameterizations that feed into each model's simulated ABL values, see Monk et al. [25].

The retrieval of hourly lidar mixing depths was not always consistent. Within the diurnal window 0900–1900 h each day, reliable mixing depth estimates were only obtained around 40% of the time. Outside this diurnal window the lidar frequently yielded unrepresentative results, so nocturnal effective mixing heights (h_e) derived from the near-surface radon observations (Section 2.1) were substituted. Diurnal composite mixing depths representative of each atmospheric mixing state are presented in Figure 10a. Corresponding composites of simulated ABL depths are presented in Figure 10b. Nocturnal h_e values are least representative for category #1 days (Figure 10a), since radon variability on these nights is driven more by advective effects than changes in mixing depth. Furthermore, in the predawn hours of category #4 days, h_e values begin to rise from a minimum value around 0000–0200 h. This behavior is an artefact of near-surface radon being "lost" to katabatic drainage flow, as discussed in Section 3.6.1.

The W-A11 and W-UM1 models employed the local closure MYJ ABL scheme [47], whereas the others used the non-local closure scheme of YSU [76]. Given that only two ABL schemes were used between the five WRF-based models, with a separate scheme for the CCAM models [25], considerable variability in both nocturnal and daytime ABL depths is evident for each of the mixing states (Figure 10b). Other factors contributing to these differences are discussed in Monk et al. [25].

For the synoptically non-stationary or most windy (category #1 and #2) conditions, models exhibited substantial variability in nocturnal ABL depths. For category #3 and #4 conditions, on the other hand, with the exception of W-UM2 and C-CTM, nocturnal ABL estimates were quite similar. Overall, C-CTM tended to consistently underestimate nocturnal ABL depths, whereas W-UM2 tended to overestimate them. However, C-CTM and W-UM2 consistently reported the lowest daytime ABL depths for all mixing categories. In C-CTM the ABL depth is estimated according to air mass buoyancy. Consequently, at night when no heat is input to the boundary layer, the mixing depth drops to the depth of the lowest model layer.

With the exception of O-CTM, which led observations by around 2 h, overestimated temperatures (Figure 5), and overestimated daytime ABL depths, all other models underestimated daytime ABL depths across all mixing categories. Typically these underestimates were at the upper end of, or larger than, those reported at the Sydney airport by Monk et al. [25]. In contrast to the Westmead site (Figures 8 and 9), however, all models tended to underestimate surface wind speeds at the airport.

3.6. Gaseous and Particulate Atmospheric Constituents

3.6.1. Passive Atmospheric Tracer (Radon-222)

Before introducing the complexities of large spatial and temporal changes in source strengths and variable atmospheric lifetimes, it is instructive to see the impact that each model's meteorological parameterizations have on concentrations of a well-distributed gaseous passive tracer species with a surface source, such as radon. Only two of the contributing models (W-A11 and W-UM1) simulated radon for SPS-II.

Results here focus on two radon simulations by W-A11 (one using the Griffiths et al. [35] Australian radon flux map (A11-FM), the other assuming a constant radon source function of 20 mBq m^{-2} s^{-1} (A11-CS)), and one simulation by W-UM1, also using the Griffiths et al. [35] flux map (UM1-FM). In each case, reported radon concentrations are averages across the model's lowest model layer (W-A11, 32 m; W-UM1, 33.5 m).

Section 2.2 (and references therein) discuss an approximate method to separate an observed or simulated radon time series into advective (fetch-related) and diurnal (mixing-related) components. To better evaluate aspects of the relevant model parameterizations separately, observed/simulated radon time series, along with their respective fetch-related and mixing-related components, are shown in Figure 11.
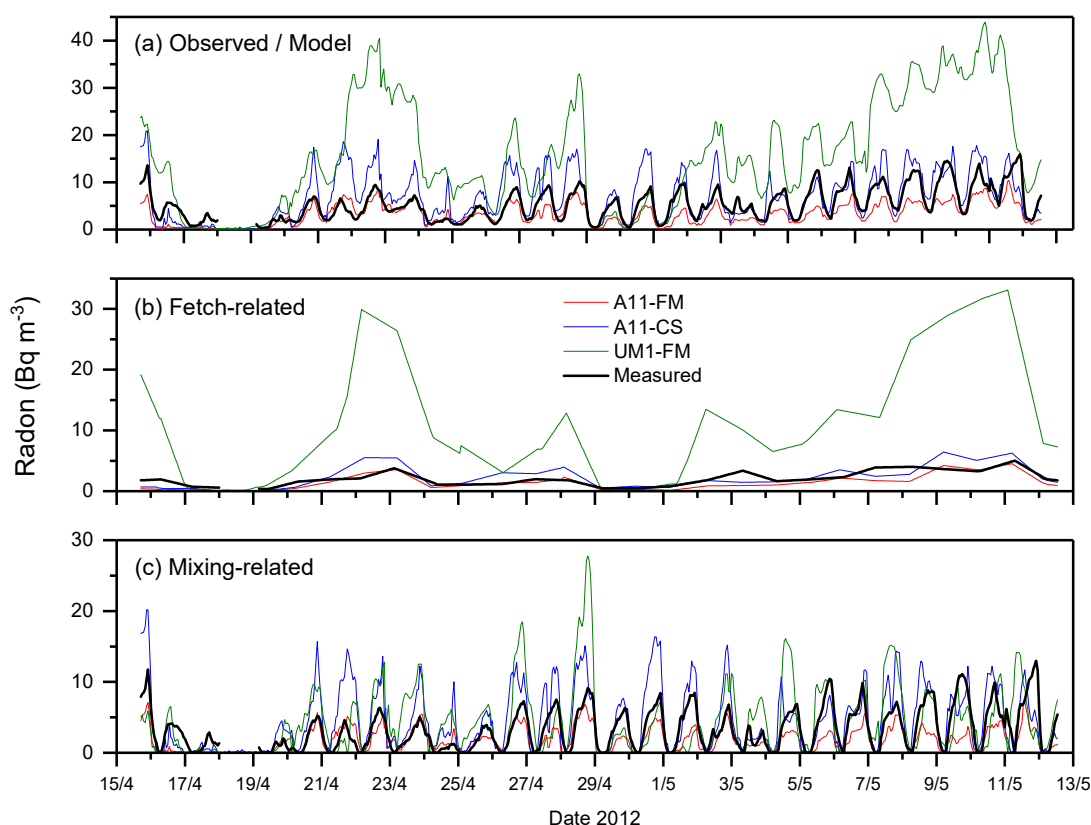
Figure 11. (**a**) Simulated and observed hourly radon time series at Westmead, (**b**) fetch-related component of the simulated & observed concentrations, and (**c**) mixing-related component of simulated & observed radon concentrations.

In some cases, the original time series indicate substantial differences between simulated and observed radon (Figure 11a), although much of the differences could be attributed to fetch contributions (radon advection in the models; Figure 11b). For the UM1-FM simulations, non-zero radon lateral boundary conditions were used, which is likely responsible for the large fetch effect.

The mean bias (Rn_{MOD}-Rn_{OBS}) and mean bias standard deviation, for the radon fetch components are: μ_{MB} −0.56, σ_{MB} 0.7 for A11-FM, μ_{MB} 0.34, σ_{MB} 1.11 for A11-CS, and μ_{MB} 9.37, σ_{MB} 8.62 for UM1-FM. The factor-of-two reduction in σ_{MB} for fetch effects between A11-CS and A11-FM demonstrates clear benefits in attempting to represent spatial variability in the radon source function. However, spatial variability in the radon flux map (driven by topsoil Radium-226 content), is better represented than short-term temporal variability caused by saturating rainfall events.

Looking at the diurnal radon component (Figure 11c), A11-CS and UM1-FM both have a tendency to overestimate nocturnal radon concentrations, despite these models also generally overestimating nocturnal mixing depths (Figure 10). A11-FM, on the other hand, appears to perform much better in the first half of the campaign than the second. Over the course of the month both rainfall (soil moisture) and meteorological conditions changed considerably (Figure S1; and [9]). To investigate which of these had the bigger influence on the observed differences we plotted simulated and observed diurnal radon as a function of the atmospheric mixing state (Figure 12a) and corresponding diurnal mean bias values (Figure 12b). It should be noted that only an approximate response time correction (data shift by 30 min to reduce instrument's 45-min response time) was applied to the radon observations in this study for the purpose of atmospheric stability analysis. Until a more accurate response time correction is applied (i.e., [77]), the observed diurnal radon cycles reported in Figure 12a will not be properly representative of the scalar dilution rate within the developing convective boundary layer.
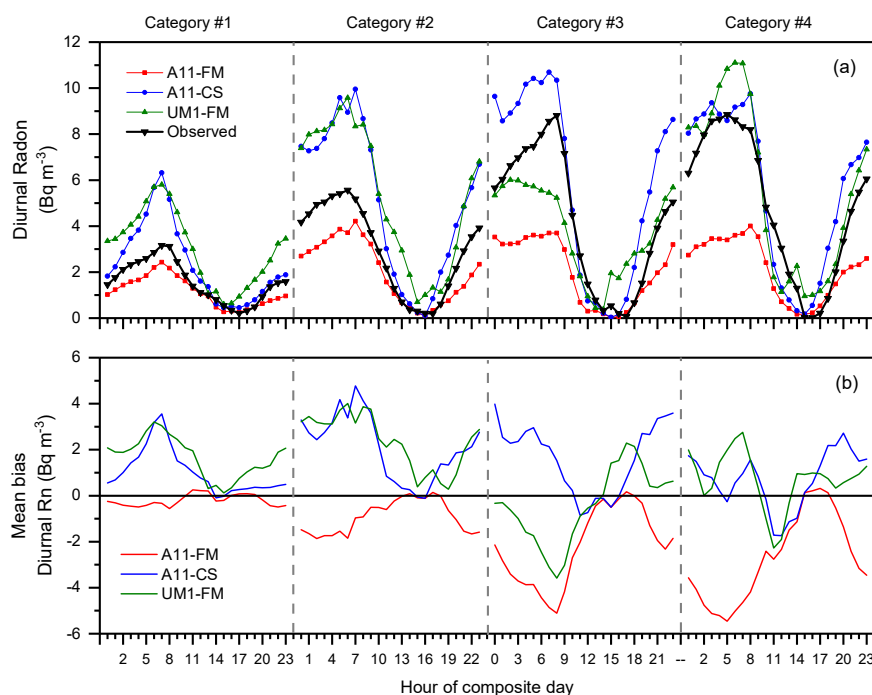
Figure 12. (**a**) Diurnal composite observed and simulated mixing-component of radon, and (**b**) diurnal mean biases of the mixing-component of radon, for each mixing category.

The rapidly increasing μ_{MB} and σ_{MB} (as implied by the amplitudes of the diurnal μ_{MB} curves) for A11-FM from category #1 to #4 conditions (i.e., as nocturnal conditions become progressively more stable and the contributing fetch region shrinks), may indicate that the Griffiths et al. [35] flux map underestimates the radon source function in the vicinity of Westmead (~10–20 km radius of the site). However, the generally positive mean bias of A11-CS seems to indicate that the assumed constant source value it is a little too high. The majority of SPS-II rainfall events occurred under category #1 and #2 conditions. Since these conditions were also the windiest, a larger fetch region was contributing to Westmead radon concentrations. Consequently, fetch regions in the vicinity of Westmead most likely influenced by the saturating rains would have a reduced influence on observed and simulated

concentrations. The comparatively low A11-FM μ_{MB} for category #1 and #2 days indicates that the flux map is more representative for fetch regions further afield from Westmead.

A conspicuous feature in the observed radon mixing-component signal (Figure 12a), is the lack of a pronounced pre-dawn peak under category #4 conditions of the kind seen on category #3 days. An investigation of nocturnal wind direction and wind speed indicated the onset of katabatic drainage flow conditions at Westmead between 0300 and 0700 h on category #4 days. This was characterized by a consistent mean wind direction of ~340° and an increase in mean wind speed (from ~0.4 m s^{-1}, 1900–0200 h to ~0.6 m s^{-1}, 0300–0700 h). Simulated wind directions (not shown) could not reproduce this feature, with typical differences of around 40–50° (including higher simulated variability). Overall, W-NC1 and W-NC2 performed the best, with average wind directions coming to within 10–20° of observed values in the hours just before dawn.

Disregarding the advective loss of near-surface radon on category #4 nights, there was a progressive increase in peak nocturnal radon concentrations from category #1 to #4 nights. This behavior was not reflected in the radon mixing-related components of any of the three simulations. The inability of these models to reproduce the near-surface accumulation of a passive tracer with a surface-only source in conditions of increasing nocturnal stability casts doubt on their ability to do the same for primary anthropogenic pollutants. In future studies, if all contributing models were encouraged to include a radon simulation (with each using the same radon source function), this might shed some light upon mean biases observed with other pollutant concentrations.

3.6.2. Locally-Sourced Primary Pollutants (NO, CO)

In the central western Sydney suburbs concentrations of nitric oxide (NO) and carbon monoxide (CO) are primarily controlled by the number of vehicles [9], the depth of the ABL, advection or venting from the ABL, and for NO in particular, its short atmospheric lifetime. Regarding primary emissions such as NO and CO, as discussed in Section 2.3, all models used approximately the same emissions (at least for the Sydney metropolitan region). Despite the similarity of urban Sydney emission schemes used between models, simulated diurnal cycles of these emissions showed marked differences between models for all mixing states (Figures 13 and 14).
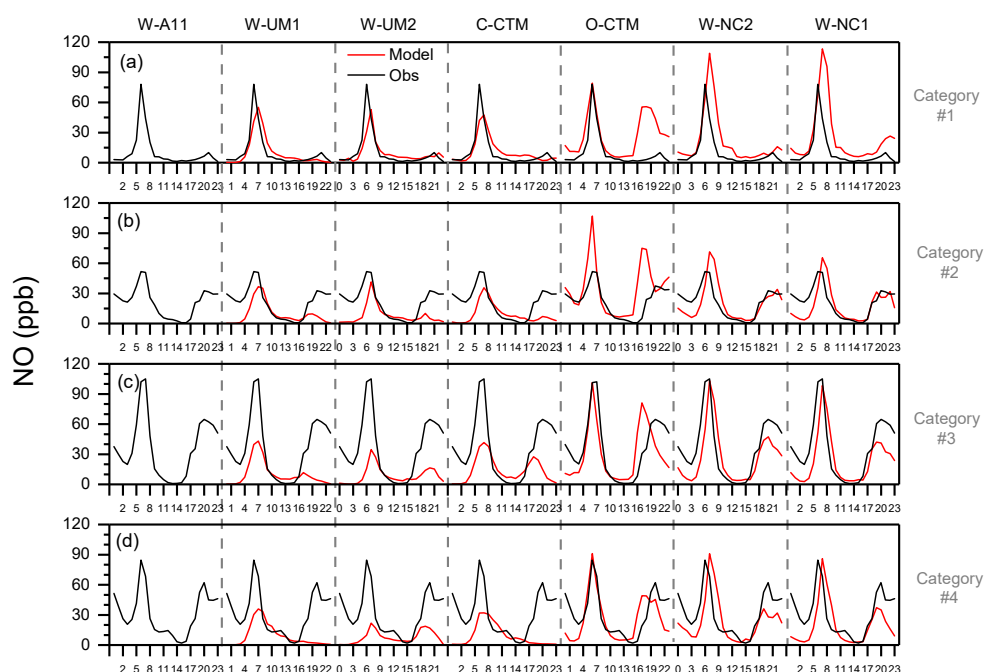
Figure 13. Diurnal composite observed (5 m) and modelled (z_1 average) NO at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).

There was a tendency in all models for NO μ_{MB} to become more negative from category #1 to #4 mixing states (Section 3.6.5). In the case of W-UM1, W-UM2 and C-CTM, there was also a tendency for σ_{MB} to increase from category #1 to #4 conditions. Overall, NO simulated by O-CTM, W-NC1 and W-NC2 was higher than for W-UM1, W-UM2 and C-CTM (Figure 13). Of particular interest is that none of the contributing models showed a consistent, progressive increase in peak morning NO with increasing nocturnal stability. In some cases, however, evening peak concentrations showed some dependence on the nocturnal mixing state. Of all the models, W-NC1 and W-NC2 exhibited the lowest NO μ_{MB} and σ_{MB} values (Section 3.6.5). A comparison and evaluation of the chemistry schemes of models contributing to this study is provided by Guérette et al. [26].

Some of the category #1 days represent non-stationary synoptic conditions; days on which large and rapid changes in weather conditions occurred. Reported morning and evening peak pollutant concentrations on individual category #1 days will therefore depend strongly upon the timing of these changes on these days. From category #2 to category #3 conditions, observed peak NO concentrations increased markedly. The fact that a similarly large increase was not observed from category #3 to category #4 days is related to the katabatic drainage flow discussed in Section 3.6.1. Only W-NC1 and W-NC2 showed consistent peak increases in morning NO concentrations from category #2 to #3 conditions; peak magnitudes were also simulated quite well in these cases.

As an example of observed and simulated primary pollutant concentrations at a flat site, without katabatic drainage effects, Figure S3 shows the corresponding SPS-II NO concentrations observed and simulated at the Richmond OEH site. The topography of this site is relatively flat within a 5 km radius of the observations [17]. In contrast to Westmead, observed morning and evening peak concentrations at Richmond increase consistently with radon-derived mixing category. Of all the models, only W-UM1 and C-CTM at this site show morning NO peaks increasing with mixing category, however the same is not true of the simulated afternoon peak concentrations. In the cases of C-CTM, O-CTM, W-NC1 and W-NC2, the comparison of simulated and observed NO between the Westmead and Richmond sites highlights how site-specific perceived model skill can be.

For each model in Figures 13 and 15, prescribed emissions are mixed each hour within the lowest model grid-cell (3 km × 3 km × z_1 m). Forming diurnal composites over, in this case, 8 days per mixing category, helps to reduce some of the differences that might otherwise arise between instantaneous hourly observations and simulated concentrations as a result of spatial and temporal source variability. Figure S2, for example, shows occasional peak hourly NO concentrations of up to 200 ppb, whereas average peak concentrations for stable nocturnal conditions are 80–100 ppb (Figure 13c,d). For longer campaigns, either the number of days per category can be increased (increasing the statistical robustness of results), or the number of mixing categories increased.

The between-model differences evident in Figures 13 and 14 are large, considering that each of the contributing models were essentially using the same NO and CO emission schemes for the Sydney metropolitan region. Further insight to the root causes of these differences can be found in Monk et al. [25] and Guérette et al. [26].

Individual model ABL schemes and cumulus parameterizations are likely significant for simulated air quality. W-UM1, for example, has a slowly-reducing evening ABL height (Figure 10), and shows little evening build-up of NO, whereas W-UM2 has a more rapid evening ABL decrease and evening NO concentrations increase more with increasing nocturnal stability. C-CTM, on the other hand, shows little change in nocturnal ABL depth with mixing category (Figure 10), and also shows little change in peak nocturnal NO concentrations (Figure 13). O-CTM has the earliest reduction in ABL depth in the evening, which is reflected by an early large NO evening peak, but this feature is much less pronounced for CO.
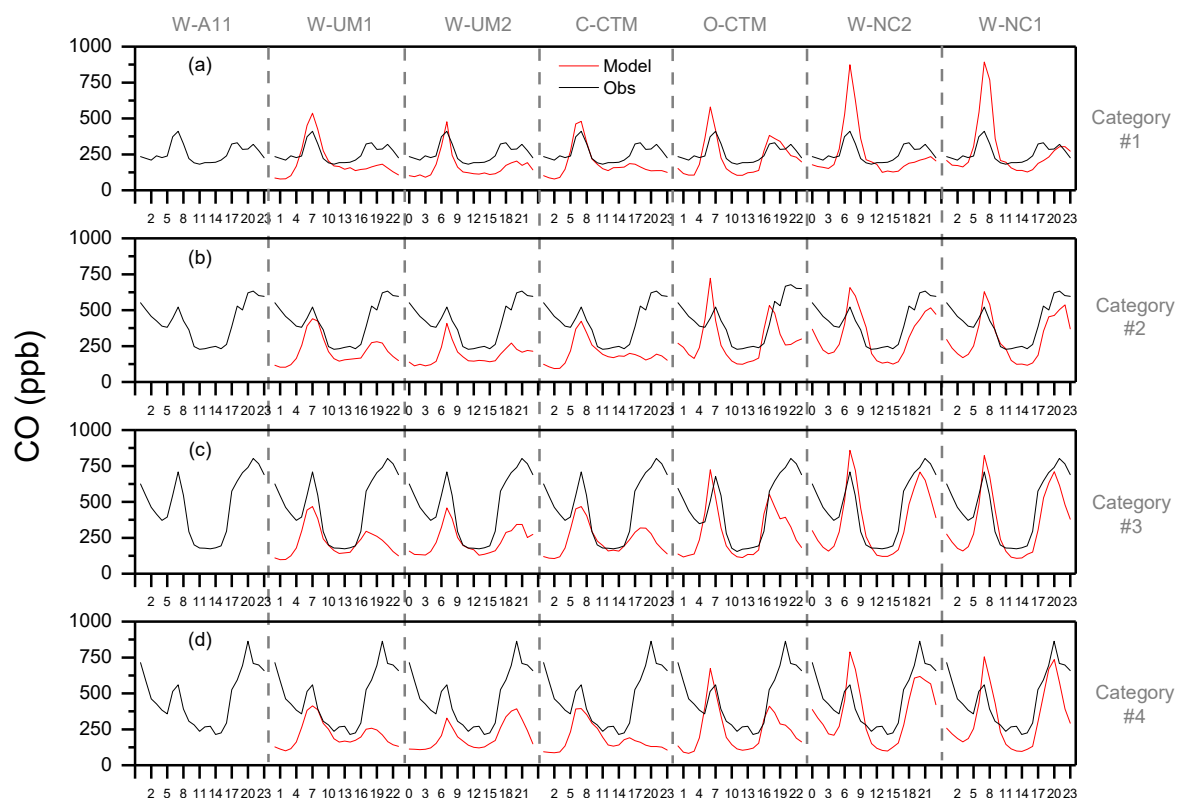
Figure 14. Diurnal composite observed (5 m) and modelled (z_1 average) CO at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).

For category #2 to #4 conditions W-UM1, W-UM2 and C-CTM underestimate CO the most. It is unlikely that these inter-model differences are solely attributable to different ABL calculations since W-UM2 and C-CTM represent opposite extremes of nocturnal ABL estimates (Figure 10), but similarly underestimate CO despite using the same emission scheme. Also curious is that O-CTM underestimates evening peak CO concentrations compared with W-NC1 and W-NC2 despite its evening ABL height dropping much sooner. While W-NC1 and W-NC2 generally do the best job at reproducing peak morning and evening CO concentrations, observed and simulated values diverge in the middle of the night. It is possible that the higher simulated wind speeds advect urban pollution from the city too quickly.

Of all the models, C-CTM and O-CTM include an anthropogenic wood burning source, with a parameterization that is dependent on heating degree days, otherwise anthropogenic emission data is the same in all cases. However, C-CTM exhibits a significantly greater CO mean bias than O-CTM (Section 3.6.5; Figure 14).

A corollary of Figure 13c,d and Figure 14c,d (along with their respective μ_{MB} and σ_{MB} values; Section 3.6.5), regarding each model's ability to reproduce a representative concentration of a primary pollutant with a surface source, is that at times when urban primary pollutant concentrations are most likely to pose a public health concern, the majority of models either under-predict the magnitude of the pollution event, the duration of the pollution event, or both. Caution should therefore be exercised when using such models in their current form for public health applications.

Comparing Figures 13 and 14, it is evident that over the diurnal cycle in each mixing category simulated NO and CO values are better correlated than the observed values. Figure 15 compares the relationship between NOx and CO for observed and simulated values at Westmead for only category #3 and #4 days (coldest, most stable nights). The observations show different relationships at different times of the day. During the morning peak, specifically 5–7 a.m., the slope of the NOx vs. CO curve is steepest. The period around noon, and the evening peak hour period, have similar slopes, but

the curves are offset due to accumulation of CO throughout the day. If the NOx and CO sources aren't substantially different between the early hours of the morning peak hour period to the rest of the day, then this modified NOx vs. CO relationship maybe related to a lack of ozone to convert NO to NO_2 (for subsequent reactions). As shown in Section 3.6.4 on category 3 and 4 days there is essentially no ozone present in the morning until around 8 a.m. However, since the experiment was conducted in autumn, changing contributions of wood smoke throughout the day from domestic heating cannot be ruled out. Generally, the slope of the CO vs. NOx relationship was quite similar between models, indicating that net differences between emissions were not very large and should not contribute significantly to the differences observed in Figures 13 and 14. Furthermore, the slope of the simulated NOx vs. CO curve generally lies between the morning peak and evening regimes indicated by the observations.
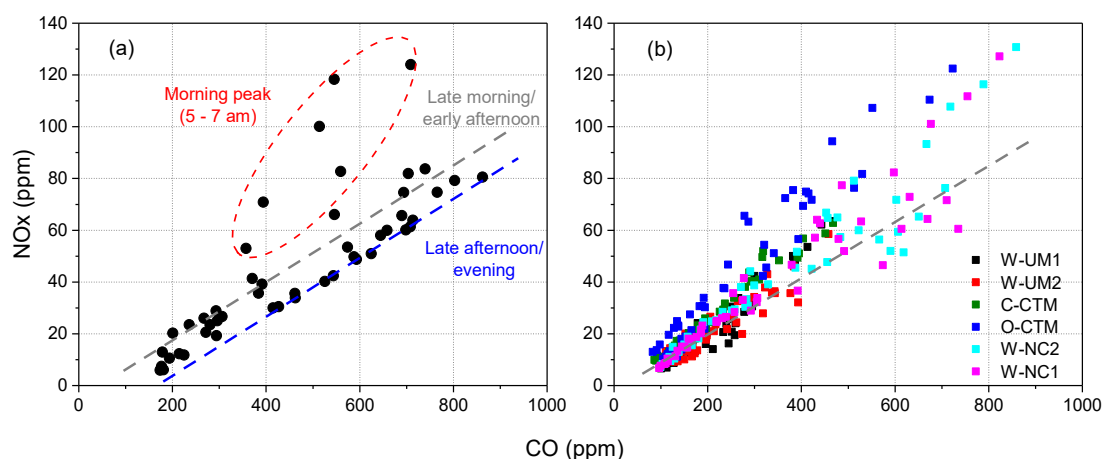
Figure 15. Relationship between NOx and CO at Westmead, category #3 and #4 days only, for (**a**) observed, and (**b**) simulated values. The grey and blue dashed lines in (**a**) are only to guide the eye, not regression fits; the grey line has been transferred to (**b**) for comparison purposes only.

3.6.3. Local and Non-Local, Primary and Secondary Pollutants (NO_2, PM_{10}, SO_2)

Unlike for NO and CO, there appeared to be no tendency for either the NO_2 μ_{MB} or σ_{MB} to increase with changing mixing state (Figure 16 and Section 3.6.5). Notably, however, models that exhibited the largest NO μ_{MB} (W-UM1, W-UM2) had the lowest NO_2 μ_{MB}, and the models that had the smallest NO μ_{MB} (W-NC1, W-NC2) had the largest NO_2 μ_{MB} (Section 3.6.5). While model photochemistry maybe a contributing factor the large comparative difference in mean bias of NO and NO_2 suggests this was not the only cause.

Taking category #3 days as an example (relatively stable evenings, but no significant katabatic drainage), the observed morning NO_2 peak appears to lag the NO peak by 1 h, whereas for the simulated values, the morning NO_2 peak either matches the timing of the NO peak, or precedes it by 1 h. Given that emissions are similar in all models, this feature is most likely attributable to each model's photochemistry.

In the case of PM_{10} aerosols there was a tendency for the μ_{MB} magnitude and σ_{MB} to increase with changing mixing state (Figure S4 and Section 3.6.5). Overall W-NC1 and W-NC2 exhibited the lowest PM_{10} μ_{MB} but their σ_{MB} was most sensitive to changing mixing state. This was mostly related to the magnitude of evening peak concentrations. Regardless of mixing category W-UM2 over-predicted PM_{10} concentrations across the diurnal cycle (Figure S4). As discussed by Guérette et al. [26], W-UM1, W-UM2, W-NC1 and W-NC2 all use a modal representation of the particle size distribution. Unique to W-UM2, however, is the use of the Secondary Organic Aerosol Module (SORGAM; [78]).
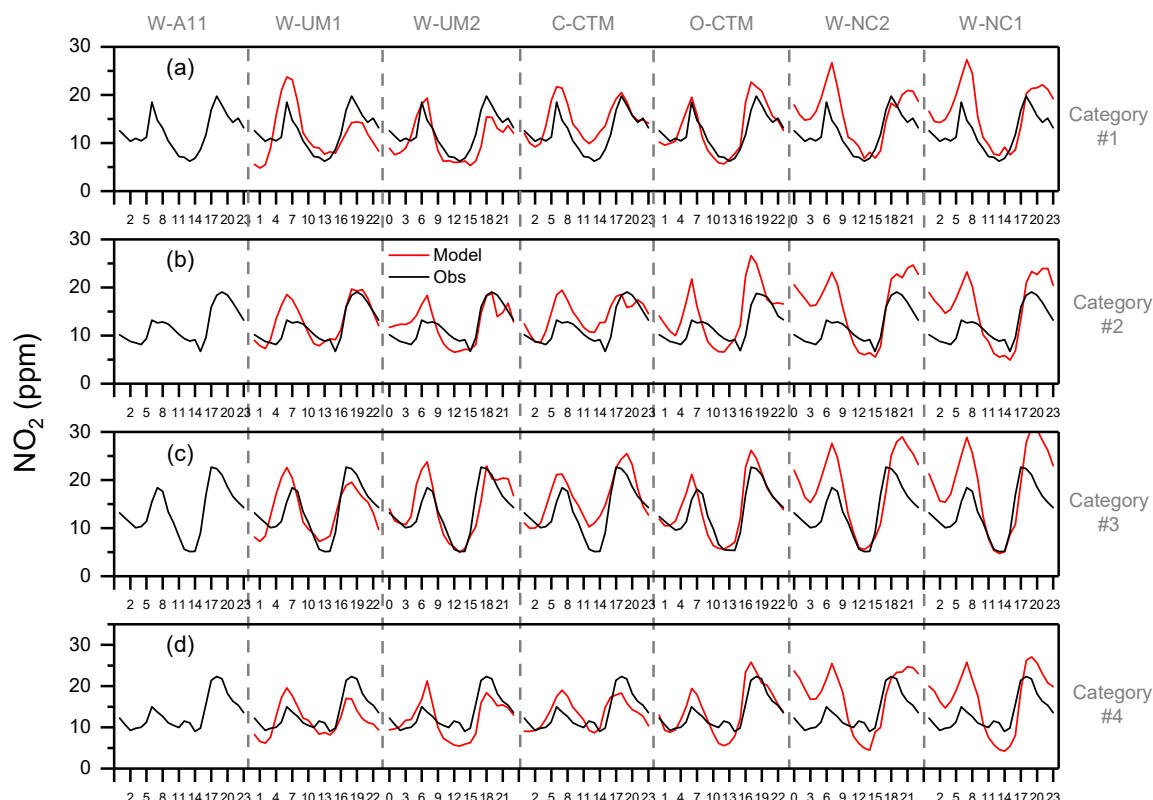
Figure 16. Diurnal composite observed (5 m) and modelled (z_1 average) NO_2 at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).

Each of the models used in this study have a wind speed dependent parameterization of dust particles and sea salt. Over-prediction of wind speed would therefore contribute to the over-predictions of these emissions and thus the PM_{10} concentrations. However, PM_{10} model skill at this single site may not represent the overall performance of models across all sites within the domain. For example, the normalized mean biases across all sites during SPS-II reported by Zhang et al. [69] were: -7% for W-NC1, -1.9% for W-NC2, -86% for W-UM1, 78% for W-UM2, -38.8% for O-CTM and -37% for C-CTM.

Although Figure S2 indicated some isolated spikes in hourly SO_2 concentration to ~9 ppb late in the campaign, mixing category composites showed no clear SO_2 diurnal cycle for category #1 or #2 days (Figure S5). Category #3 days, on the other hand, were characterized by morning and evening SO_2 peak concentrations, while category #4 days were characterized by peak concentrations during the day. The afternoon wind direction on category #3 days (not shown) was south-easterly, so local Sydney traffic sources are likely to dominate. On category #4 days, however, afternoon wind directions were southerly, so SO_2 originated from localized suburban sources (see Section 3.7). Peak morning and evening SO_2 concentrations on category #3 days were almost an order of magnitude below the NEPM hourly SO_2 guideline value of 200 ppb, indicating that there are few significant SO_2 sources in the Sydney CBD apart from those related to traffic emissions.

For W-UM1, W-UM2, C-CTM and O-CTM there was a tendency for both SO_2 μ_{MB} magnitude and σ_{MB} to increase with changing mixing state (Section 3.6.5). This tendency was not observed with the W-NC1 and W-NC2 models for which the SO_2 μ_{MB} was typically of the opposite sign to that of the other models. Zhang et al. [69] show that W-NC1 and W-NC2 largely under-predict precipitation, by 47–53% and 31–39%, respectively, compared to observations, which may explain to a large extent the under-predicted wet removal rates of SO_2 from the atmosphere and possible under-predicted conversion rates of SO_2 to sulfate in the cloud and rain droplets, or that the removal to the particulate phase is not acting quickly enough.

3.6.4. Local and Non-Local Secondary Pollutants (Photochemical Oxidants as O_3)

SPS-II ozone concentrations were characterized by nocturnal minimum values, and peak daytime values that were largely insensitive to atmospheric mixing state (Figure 17) on account of the relatively low autumn sunlight intensity. Nocturnal ozone concentrations, on the other hand, were a strong function of mixing state. Under windy, well-mixed conditions nocturnal mean ozone concentrations were typically around 5–7 ppb, prior to the onset of morning peak-hour NO emissions. Under these conditions, ozone is readily mixed down to the surface from above. By comparison, on the most stable nights (Figure 17d) ozone concentrations dropped essentially to zero between 2000 h and 0500 h, due to isolation of the near-surface air from the free-atmosphere by the nocturnal inversion, and ozone titration by NO.

There was a tendency for all models (leastwise for W-NC1) for the ozone μ_{MB} to increase (or at least become more positive) from category #1 to #4 conditions (Figure 18c). With the exception of W-NC1 and W-NC2, this was accompanied by an increase in σ_{MB}. W-NC1 and W-NC2 generally provided the most accurate ozone simulations across the diurnal cycle for all mixing states (Figure 17). The least representative peak daytime ozone concentrations for all mixing states were from O-CTM. All models (worst for W-UM1, W-UM2 and C-CTM) over-predicted nocturnal ozone concentrations, which may be due to higher simulated wind speeds mixing down too much ozone from above. However, these three models also "ran out" of NO in the middle of the night, which would remove the possibility of any nocturnal ozone titration. W-NC1, W-NC2 and O-CTM generally exhibited reasonable nocturnal ABL depths (Figure 10), which may be related to their superior nocturnal ozone simulations. However, C-CTM, which had the lowest of all nocturnal mixing depths, still overestimated nocturnal ozone concentrations. Future studies will need to investigate differences in ozone deposition between these models.
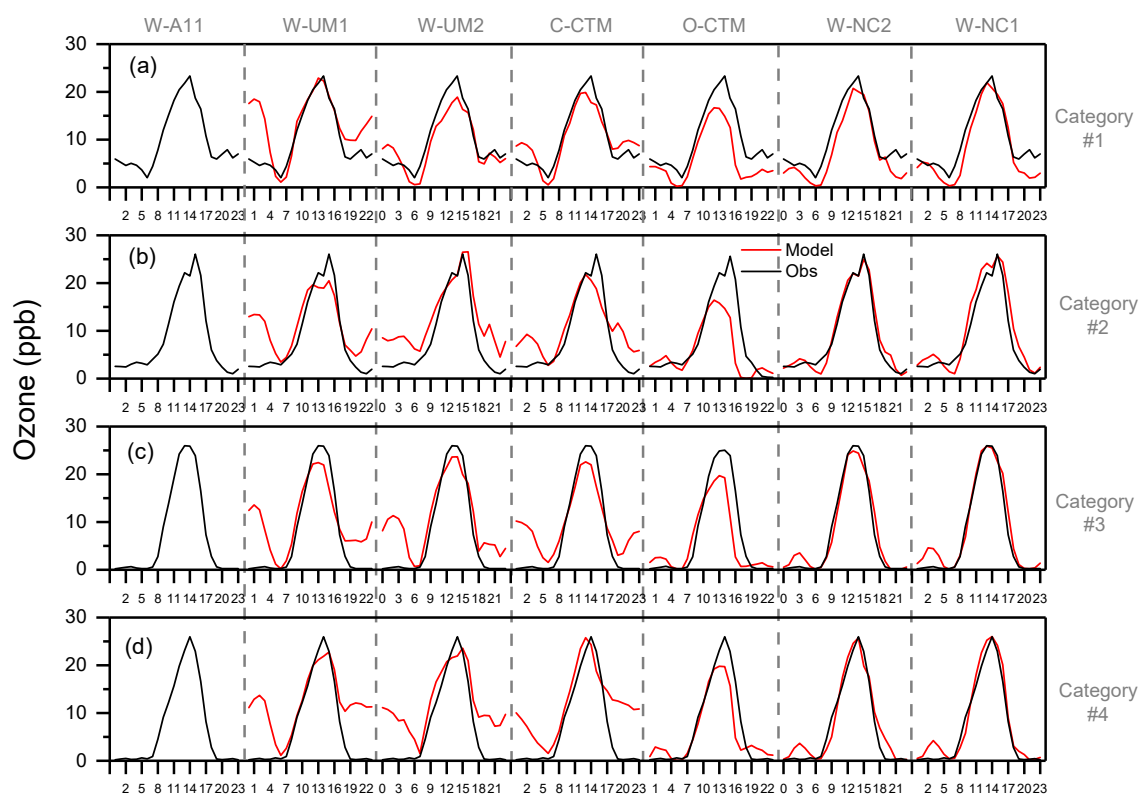
Figure 17. Diurnal composite observed (5 m) and modelled (z_1 average) O_3 at Westmead. Panels a—d represent each of the four radon-based mixing categories (Table 1).
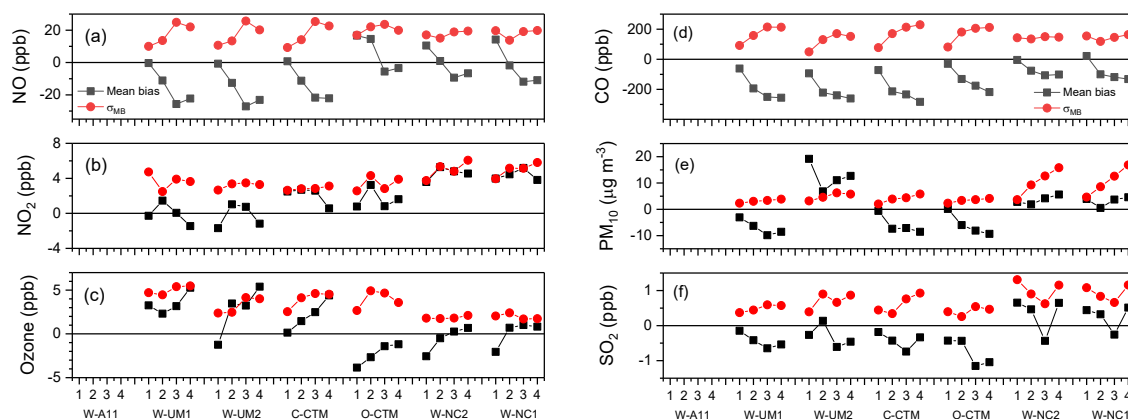
Figure 18. Mean bias (μ_{MB}) and σ_{MB} for (**a**) NO, (**b**) NO$_2$, (**c**) O$_3$, (**d**) CO, (**e**) PM$_{10}$, and (**f**) SO$_2$, at Westmead for all models and atmospheric mixing states.

3.6.5. Summary of Model Performance for Selected Air Pollutants

A summary of μ_{MB} and σ_{MB} values for all chemical species discussed in Sections 3.6.2–3.6.4, separated by mixing state, is given in Figure 18. In this format, between-model differences are clear for each species and mixing state, making it easier to infer which mixing states have the largest influence on perceived model skill. Furthermore, where similarities in behavior between species exist, as is the case for NO and CO in Figure 18a,d, these are also clearly evident.

For the locally-sourced primary pollutants (NO and CO) there was a tendency for models to underestimate concentrations on category #3 and #4 days averaged across the diurnal cycle. Most of this discrepancy was attributable to underestimated morning and evening peak concentrations. In the case of NO$_2$, which has both primary and secondary sources, no comparable dependence of μ_{MB} on mixing state was evident (Figure 18b). The absolute magnitude of PM$_{10}$ μ_{MB} as well as its σ_{MB} typically increased from category #1 to #4 conditions. In the case of ozone, with purely secondary sources, the μ_{MB} increased with mixing state for all models, mainly as a result of excessive nocturnal concentrations. No clear dependence of SO$_2$ μ_{MB} or σ_{MB} was evident with mixing category, perhaps on account of local and remote sources making comparable contributions.

Overall, these results demonstrate that contemporary model performance is typically least accurate on category #4 days, which are associated with the most stable nocturnal conditions and warm, clear-sky days. Unfortunately, these are also the conditions under which primary and secondary pollutant concentrations are most likely to exceed NEPM guideline values. Consequently, future studies should seek to improve the understanding, and parameterizations, of weather conditions associated with category #4 days.

3.7. Source Distribution and Magnitude

The spatial heterogeneity of sources in an urban region can have a large influence on urban air quality as determined from spot measurements. Depending on the chosen timescale or method of evaluation, it can also considerably influence perceived model skill. From the modelling perspective, the accuracy of the emission scheme (regarding spatial/temporal variability and source strengths) is critical. On this matter, Cope et al. [9] recommended investigating new 'top down' methods by which to check or constrain 'bottom up' inventories.

One way to assess the net effect of between-model differences in meteorology and chemistry schemes is to analyze the angular distribution of air quality parameters when all models are using the same emissions scheme. Supplementing such an analysis with corresponding observational data can help to constrain the emission inventory provided information regarding mixing depth is also available. To this end, class-typing by atmospheric mixing category is one way of minimizing variability in mixing depth within a sample set. By using the radon-based stability classification method to exclude periods

of strongest winds or highly variable synoptic conditions (i.e., category #1 and #2 days), day/night mixing depths can be well constrained for a given season (e.g., 1500–1800 m/70–110 m; Figure 10). Furthermore, excluding category #1 and #2 conditions also enables a more detailed representation of the positions and relative strengths of local pollutant sources to be obtained.

To demonstrate, Figure 19 compares directional distributions of pollutants relative to the Westmead site from: (i) all weather conditions (red open circles), and (ii) only category #3 & #4 mixing days (black filled circles). Using Figure 1 as a reference, the following features influencing Westmead air quality can be identified in Figure 19: (i) the emission plume from the Sydney CBD (E-SE); (ii) a separate broad traffic emission plume from the Blacktown to Castle Hill region (W-NW); (iii) a localized pollutant source due south of the site, and (iv) a high O_3 contribution from the NE, a more densely forested region with lower NOx contributions, minimizing O_3 titration.

Features like the point-source south of Westmead only appear when the non-stationary and most windy conditions are removed. According to the National Pollution Inventory (http://www.npi.gov.au/npidata/action/load/map-search), there is a Polymer Production Manufacturer (PPM) 7.5 km south of Westmead that, among other things, is a significant source of CO, NOx, PM and SO_2. The PPM SO_2 emissions for 2012 were reported to be ~135 kg yr^{-1}. Assuming that the PPM operates ~260 days per year (i.e., no weekends), for 8-h days, a mixing depth of approximately 1600 m (Figure 10a; category #3 & #4), and averaging over a 10° arc (as in Figure 19), daytime SO_2 concentrations at Westmead that have come directly from the PPM should be around 11 ppb. SO_2 concentrations from 190° in Figure 19f were around 4.7 ppb, but these were averaged over whole days (24-h), not an 8-h day (8-h equivalent mean concentration of 14.1 ppb, whereas peak hourly values in Figure S2 were 9 ppb). Given the potential variability in daytime ABL depth (1500–2000 m; Figure 10a), observed and bottom-up values are in good agreement but, on average, assuming no seasonality in production, actual annual emissions of SO_2 from the PPM in 2012 may have actually been in the range 180–200 kg yr^{-1}.
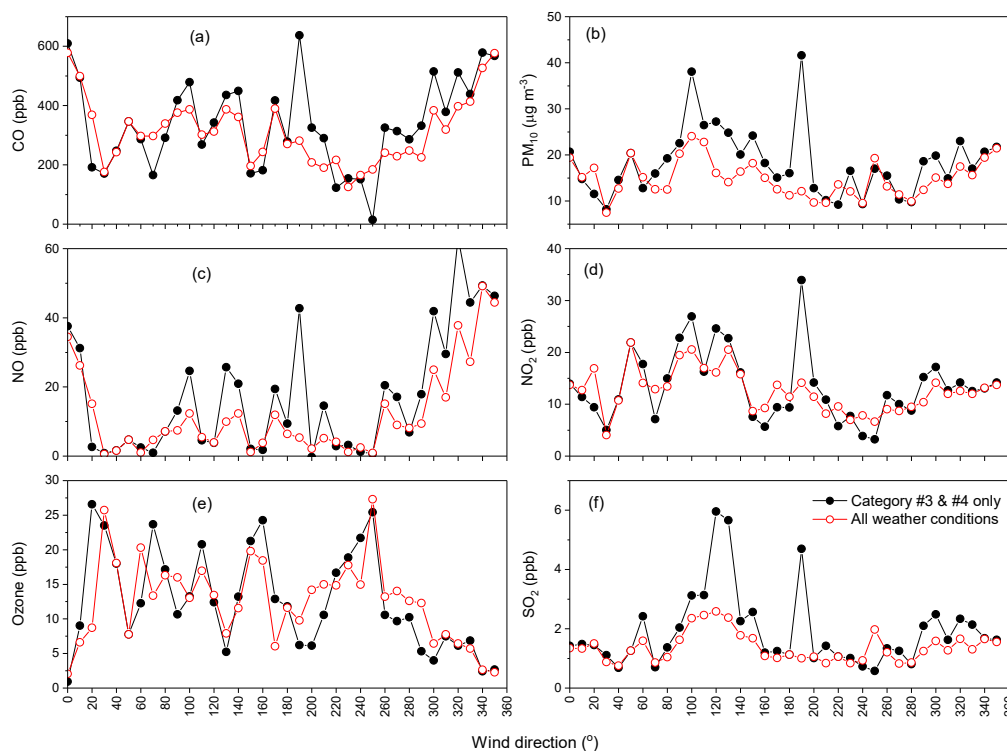
Figure 19. Angular distributions of (a) CO, (b) PM$_{10}$, (c) NO, (d) NO$_2$, (e) O$_3$ and (f) SO$_2$ at Westmead during SPS-II for all weather conditions (red) and category #3 & #4 days only (black).

To better understand some of the discrepancies between simulated and observed air quality reported in Sections 3.6.2–3.6.4, in Figure 20 we compare the observed and simulated source distributions under category #3 & #4 conditions only. Note that, for this example, day and night conditions have not been separated. To facilitate reading of the panels in Figure 20 a 3-point running mean smoothing has been run over the 10° angular distributions.

For the pollutants most directly influenced by primary emissions which have localized sources (CO, NO, SO_2 and PM_{10}), Figure 20 exhibits a remarkable degree of variability between models given that the same emissions were used for all. While much of the variability between simulated values of these quantities is undoubtedly attributable to dilution of the 1×1 km inventory data in the model grid cells, and differences in model meteorology (advection, dilution and dispersion; Monk et al. [25]), the low $NO:NO_2$ ratio in the Sydney CBD plume (100–150°) indicates that sufficient time-of-flight for air masses is involved for differences in model chemistry and process parameterizations to also play a role (Guérette et al. [26]). By comparison, there is considerably less variability between simulations of NO_2 and O_3, predominantly secondary species. These species, which have a more spatially dispersed source, are simulated much more effectively.

In a future study, ideally with a longer dataset, an equivalent analysis to Figure 20 could be performed separately for day and night conditions to better target problems with specific chemical processes. Utilizing the improved day/night mixing depth information (Figure 10), and spatial dimensions of the urban region (e.g., Figure 1) or detailed point source information, better constraints (or at least "sanity checks") could be performed on existing emission inventory information.
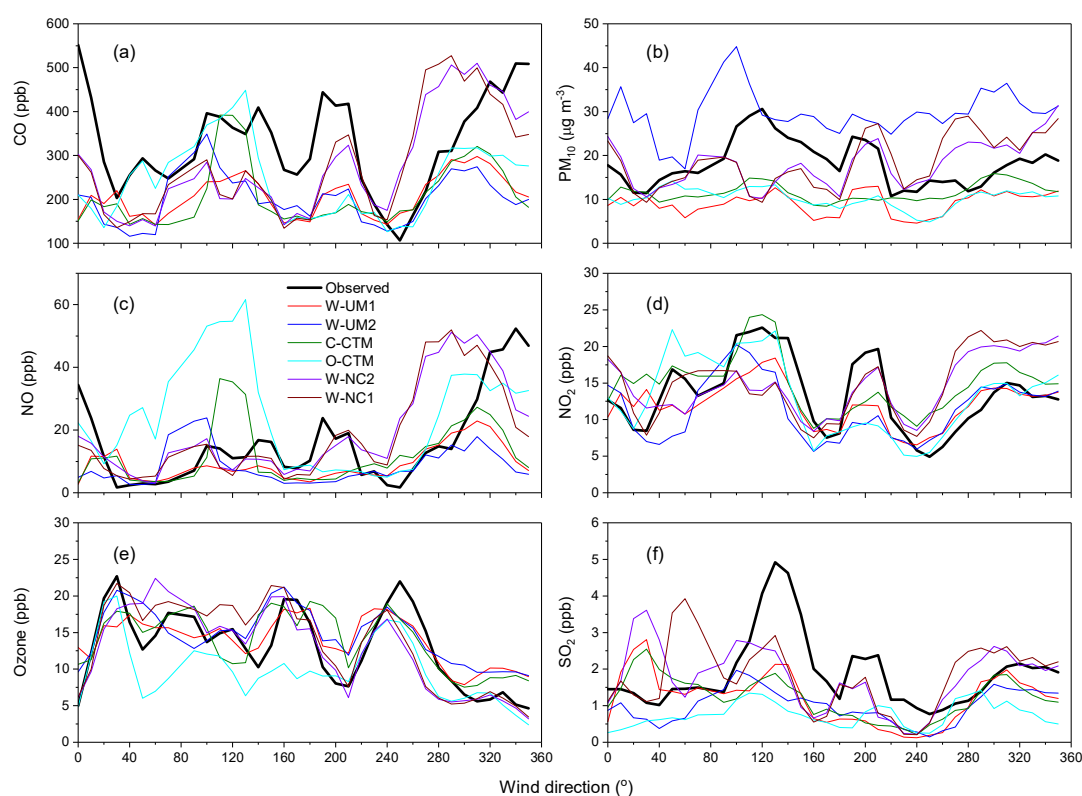
Figure 20. Comparison of observed and simulated angular distributions of (**a**) CO, (**b**) PM_{10}, (**c**) NO, (**d**) NO_2, (**e**) O_3 and (**f**) SO_2 at Westmead during the autumn 2012 SPS-II campaign.

4. Conclusions

A radon-based atmospheric stability classification technique, originally developed for characterizing the variability of urban climate and urban air quality, is adapted to prepare statistically-robust benchmarking datasets for evaluating chemical transport model performance

in the urban boundary layer as a function of the atmospheric mixing state. One month of hourly meteorological and air quality observations and simulated results from Westmead (NSW, Australia) are compared during autumn 2012. The stability classification (class-typing) approach enabled the identification/isolation of synoptic non-stationary conditions so that simulated and observed results could be compared at night, under: well mixed, weakly mixed and most-stable conditions; and during the day: under near-neutral, moderately unstable and most unstable conditions. Model skill (as inferred from the mean bias and its standard deviation) for each of the seven models in this study varied considerably as a function of the mixing state (or "class-type"). This was most notable for the primary pollutants (e.g., NO, CO), as well as for the photo-oxidant O_3. Results from the two models that included radon simulations did not indicate a consistent relationship between nocturnal accumulation of this passive surface-based tracer and the atmospheric mixing state. A corollary of this observation is that the significant between-model differences noted in simulated air quality parameters (given that the same emissions scheme was used throughout) is more attributable to the model physics in the ABL than the chemistry parameterizations. Importantly, all models failed to reproduce either (i) a representative magnitude, or (ii) the correct duration, of the morning/evening "peak-hour" concentrations of primary pollutants. These are the times of day most likely, and most often, to result in excessive public exposure to harmful anthropogenic emissions. Consequently, if CTMs are to continue to play a key role in the development of air quality guidelines and policy, considerable further research effort is required to improve existing parameterizations of ABL physical processes specifically under stable nocturnal conditions.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4433/10/1/25/s1, Figure S1: Summary of hourly meteorological and radon observations at Westmead (NSW, Australia), over the 1-month SPS-II campaign (autumn 2012). Wind information was recorded at 10 m, all other observations at "screen height" (2 m), Figure S2: Summary of hourly integrated values of selected air quality parameters at Westmead (NSW, Australia), over the 1-month SPS-II campaign (autumn 2012). All air quality sampling intakes located ~5 m a.g.l., Figure S3: Diurnal composite observed and simulated NO for each of the radon-based mixing categories (Table 1) at the Richmond OEH (NSW, Australia) site during the 1-month SPS-II campaign (autumn 2012), Figure S4: Diurnal composite observed and simulated PM_{10} concentration for each of the radon-based mixing categories (Table 1) at Westmead (NSW, Australia) for the 1-month SPS-II campaign (autumn 2012), Figure S5: Diurnal composite observed and simulated SO_2 for each of the radon-based mixing categories (Table 1) at Westmead (NSW, Australia) for the 1-month SPS-II campaign (autumn 2012).

Author Contributions: Conceptualization, S.D.C.; Data curation, E.-A.G., K.M., A.D.G., M.C. and J.C.; Formal analysis, S.D.C.; Funding acquisition, Y.Z.; Investigation, S.D.C.; Methodology, S.D.C.; Project administration, M.C., A.G.W. and M.K.; Resources, E.-A.G., K.M., A.D.G., Y.Z., H.D., M.C., K.M.E., L.T.C., J.D.S. and S.U.; Software, A.D.G. and J.C.; Visualization, S.D.C.; Writing–original draft, S.D.C.; Writing–review & editing, E.-A.G., K.M., A.D.G., Y.Z., H.D., M.C., K.M.E., L.T.C., J.D.S., S.U., A.G.W. and M.K.

Funding: Y.Z. acknowledges the support by the University of Wollongong (UOW) Vice—Chancellors Visiting International Scholar Award (VISA), the University Global Partnership Network (UGPN), and the NC State Internationalization Seed Grant. Simulations using W-NC1 and W-NC2 were performed on Stampede and Stampede 2, provided as an Extreme Science and Engineering Discovery Environment (XSEDE) digital service by the Texas Advanced Computing Center (TACC), and on Yellowstone (ark:/85065/d7wd3xhc) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. We also acknowledge the Clean Air and Urban Landscapes Hub of Australia's National Environmental Science Program for funding.

Acknowledgments: The authors would like to thank Ot Sisoutham and Sylvester Werczynski at ANSTO for their support of the radon measurement programs at Richmond and Westmead, as well as Jack Simmons for processing the SPS-II lidar observations.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following place name abbreviations are used in Figure 1:

AP	Albion Park Sth
BA	Bankstown Airport
BC	Badgerys Creek
Ba	Bargo
Bl	Bellambi
Br	Bringelly
CA	Camden Airport
Ch	Chullora
Ea	Earlwood
KG	Kembla Grange
Ln	Lindfield
Lv	Liverpool
Mc	Macarthur
Ok	Oakdale
Pr	Prospect
RR	Richmond RAAF
Ra	Randwick
Rz	Rozelle
SA	Sydney Airport
SM	St Marys
Vn	Vineyard
WA	Wollongong Airport
WR	Williamtown RAAF
Wo	Wollongong

References

1. Ayers, G.P.; Bigg, E.K.; Turvey, D.E.; Manton, M.J. Urban Influence on Condensation Nuclei over a Continent. *Atmos. Environ.* **1982**, *16*, 951–954. [CrossRef]

2. Pope, C.A.; Dockery, D.W. Health effects of fine particulate air pollution: Lines that connect. *J. Air Waste Manag.* **2006**, *56*, 709–742. [CrossRef]

3. Pope, C.A.; Ezzati, M.; Dockery, D.W. Fine-Particulate Air Pollution and Life Expectancy in the United States. *New Engl. J. Med.* **2009**, *360*, 376–386. [CrossRef] [PubMed]

4. Kim, K.-H.; Kabir, E.; Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015**, *74*, 136–143. [CrossRef] [PubMed]

5. Chen, R.; Hu, B.; Liu, Y.; Xu, J.; Yang, G.; Xu, D.; Chen, C. Beyond $PM_{2.5}$: The role of ultrafine particles on adverse health effects of air pollution. *Biochem. Biophys. Acta* **2016**, *1860*, 2844–2855. [CrossRef]

6. Apte, J.S.; Messier, K.P.; Gani, S.; Brauser, M.; Kirchstetter, T.W.; Lunden, M.M.; Marshall, J.D.; Portier, C.J.; Vermeulen, R.C.H.; Hamburg, S.P. High-Resolution Air Pollution Mapping with Google Street View Cars: Exploiting Big data. *Environ. Sci. Technol.* **2017**, *51*, 6999–7008. [CrossRef] [PubMed]

7. Zhang, R.; Liu, C.; Zhou, G.; Sun, J.; Liu, N.; Hsu, P.-C.; Wang, H.; Qiu, Y.; Zhao, J.; Wu, T.; et al. Morphology and property investigation of primary particulate matter particles from different sources. *Nano Res.* **2018**, *11*, 3182–3192. [CrossRef]

8. Keywood, M.D.; Galbally, I.; Crumeyrolle, S.; Miljevic, B.; Boast, K.; Chambers, S.D.; Cheng, M.; Dunne, E.; Fedele, R.; Gillett, R.; et al. *Sydney Particle Study-Stage-I: Executive Summary*; CSIRO: Melbourne, Australia, 2012; Available online: https://publications.csiro.au/rpr/pub?list=SEA&pid=csiro:EP19078&sb=RECENT& expert=false&n=1&rpp=25&page=2&tr=38&q=Keywood&dc4.publicationType=Report&dr=all (accessed on 11 January 2019).

9. Cope, M.; Keywood, M.; Emmerson, K.; Galbally, I.; Boast, K.; Chambers, S.; Cheng, M.; Crumeyrolle, S.; Dunne, E.; Fedele, R.; et al. *Sydney Particle Study—Stage-II*; The Centre for Australian Weather and Climate Research: Melbourne, VIC, Australia, June 2014; ISBN 978-1-4863-0359-5.

10. Gibson, M.D.; Kundu, S.; Satish, A. Dispersion model evaluation of $PM_{2.5}$, NOx and SO_2 from point and major line sources in Nova Scotia, Canada using AERMOD Gaussian plume air dispersion model. *Atmos. Poll. Res.* **2013**, *4*, 157–167. [CrossRef]

11. Irwin, J.S. A suggested method for dispersion model evaluation. *J. Air Waste Manag. Assoc.* **2014**, *64*, 255–264. [CrossRef] [PubMed]

12. Herring, S.; Huq, P. A Review of Methodology for Evaluating the Performance of Atmospheric Transport and Dispersion Models and Suggested Protocol for Providing More Informative Results. *Fluids* **2018**, *3*, 20. [CrossRef]

13. Williams, A.G.; Chambers, S.D.; Griffiths, A.D. Bulk mixing and decoupling of the nocturnal stable boundary layer characterized using a ubiquitous natural tracer. *Bound.-Lay. Meteorol.* **2013**, *20*, 381402. [CrossRef]

14. Aust, N.; Watkiss, P.; Boulter, P.; Bawden, K. *Methodology for Valuing the Health Impacts of Changes in Particle Emissions—Final Report*; NSW Environment Protection Authority (EPA): Sydney, NSW, Australia, 2013. Available online: http://www.nepc.gov.au/system/files/pages/18ae5913-2e17-4746-a5d6-ffa972cf4fdb/files/methodology-valuing-health-impacts-changes-particle-emissions.pdf (accessed on 30 November 2018).

15. Holtslag, A.A.M. Introduction to the third GEWEX atmospheric boundary layer study (GABLS3). *Bound.-Lay. Meteorol.* **2014**, *152*, 127132. [CrossRef]

16. Koffi, E.N.; Bergamaschi, P.; Karstens, U.; Krol, M.; Segers, A.; Schmidt, M.; Levin, I.; Vermeulen, A.T.; Fisher, R.E.; Kazan, V.; et al. Evaluation of the boundary layer dynamics of the TM5 model over Europe. *Geosci. Model Dev.* **2016**, *9*, 3137–3160. [CrossRef]

17. Chambers, S.D.; Williams, A.G.; Crawford, J.; Griffiths, A.D. On the use of radon for quantifying the effects of atmospheric stability on urban emissions. *Atmos. Chem. Phys.* **2015**, *15*, 1175–1190. [CrossRef]

18. Chambers, S.D.; Galeriu, D.; Williams, A.G.; Melintescu, A.; Griffiths, A.D.; Crawford, J.; Dyer, L.; Duma, M.; Zorila, B. Atmospheric stability effects on potential radiological releases at a nuclear research facility in Romania: Characterising the atmospheric mixing state. *J. Environ. Radioact.* **2016**, *154*, 68–82. [CrossRef]

19. Chambers, S.D.; Podstawczyńska, A.; Pawlak, W.; Fortuniak, K.; Williams, A.G.; Griffiths, A.D. Characterizing the state of the Urban Surface Layer using Radon-222. *J. Geophys. Res. Atmos.* **2018**. submitted. [CrossRef]

20. Williams, A.G.; Chambers, S.D.; Conen, F.; Reimann, S.; Hill, M.; Griffiths, A.D.; Crawford, J. Radon as a tracer of atmospheric influences on traffic-related air pollution in a small inland city. *Tellus B* **2016**, *68*. [CrossRef]

21. Perrino, C.; Pietrodangelo, A.; Febo, A. An atmospheric stability index based on radon progeny measurements for the evaluation of primary urban pollution. *Atmos. Environ.* **2001**, *35*, 5235–5244. [CrossRef]

22. Chambers, S.D.; Williams, A.G.; Zahorowski, W.; Griffiths, A.D.; Crawford, J. Separating remote fetch and local mixing influences on vertical radon measurements in the lower atmosphere. *Tellus B* **2011**, *63*, 843–859. [CrossRef]

23. Wang, F.; Chambers, S.D.; Zhang, Z.; Williams, A.G.; Deng, X.; Zhang, H.; Lonati, G.; Crawford, J.; Griffiths, A.D.; Ianniello, A.; et al. Quantifying stability influences on air pollution in Lanzhou, China, using a radon-based "stability monitor": Seasonality and extreme events. *Atmos. Environ.* **2016**, *145*, 376–391. [CrossRef]

24. Keywood, M.; Selleck, P.; Reisen, F.; Cohen, D.; Chambers, S.; Cheng, M.; Cope, M.; Crumeyrolle, S.; Dunne, E.; Emmerson, K.; et al. Comprehensive aerosol and gas data set from the Sydney Particle Study. *Earth Syst. Sci. Data* **2018**, in prep.

25. Monk, K.; Guérette, E.A.; Utembe, S.; Silver, J.D.; Emmerson, K.; Griffiths, A.; Duc, H.; Chang, L.T.C.; Trieu, T.; Jiang, N.; et al. Evaluation of regional air quality models over Sydney, Australia: Part 1 Meteorological model comparison. *Atmosphere* **2018**, submitted.

26. Guérette, E.-A.; Monk, K.; Emmerson, K.; Utembe, S.; Zhang, Y.; Silver, J.; Duc, H.N.; Chang, L.T.-C.; Trieu, T.; Griffiths, A.; et al. Evaluation of regional air quality models over Sydney, Australia: Part 2 Model performance for surface ozone and $PM_{2.5}$. *Atmosphere* **2018**, submitted.

27. Paton-Walsh, C.; Guerette, E.; Kubistin, D.; Humphries, R.; Wilson, S.R.; Dominick, D.; Galbally, I.; Buchholz, R.; Bhujel, M.; Chambers, S.; et al. The MUMBA campaign: Measurements of urban, marine and biogenic air. *Earth Syst. Sci. Data* **2017**, *9*, 349–362. [CrossRef]

28. BoM. *Meteorological Observations and Reports: Instrument Siting Requirements*; Version 1.0.; Bureau of Meteorology: Melbourne, Australia, 2014.

29. Utembe, S.; Rayner, P.; Silver, J.; Guerette, E.-A.; Fisher, J.A.; Emmerson, K.; Cope, M.; Paton-Walsh, C.; Griffiths, A.D.; Duc, H.; et al. Hot summers: Effect of extreme temperatures on ozone in Sydney, Australia. *Atmosphere* **2018**, submitted. [CrossRef]

30. Chang, L.T.-C.; Duc, H.N.; Scorgie, Y.; Trieu, T.; Monk, K.; Jiang, N. Performance evaluation of CCAM-CTM regional airshed modelling for the New South Wales Greater Metropolitan Region. *Atmosphere* **2018**. submitted. [CrossRef]

31. Duc, H.N.; Chang, L.T.-C.; Trieu, T.; Salter, D.; Scorgie, Y. Source Contributions to Ozone Formation in the New South Wales Greater Metropolitan Region, Australia. *Atmosphere* **2018**, *9*, 443. [CrossRef]

32. King, E.; Paget, M. Land Cover Type—MODIS, LPDAAC MCD12Q1 mosaic, Australia Coverage, [online]. Available online: http://portal.tern.org.au/land-cover-type-australia-coverage (accessed on 20 October 2018).

33. Geoscience Australia: GEODATA 9 Second DEM and D8: Digital Elevation Model Version 3 and Flow Direction Grid 2008, [online]. Available online: https://data.gov.au/dataset/0fc3357c-5852-4e6b-992c-c78bd10e9234 (accessed on 26 October 2018).

34. Griffiths, A.D.; Parkes, S.D.; Chambers, S.D.; McCabe, M.F.; Williams, A.G. Improved mixing height monitoring through a combination of lidar and radon measurements. *Atmos. Meas. Tech.* **2013**, *6*, 207–218. [CrossRef]

35. Griffiths, A.D.; Zahorowski, W.; Element, A.; Werczynski, S. A map of radon flux at the Australian land surface. *Atmos. Chem. Phys.* **2010**, *10*, 8969–8982. [CrossRef]

36. Pal, S.; Lopez, M.; Schmidt, M.; Ramonet, M.; Gibert, F.; Xueref-Remy, I.; Ciais, P. Investigation of the atmospheric boundary layer depth variability and its impact on the ^{222}Rn concentration at a rural site in France. *J. Geophys. Res. Atmos.* **2015**, *120*, 623–643. [CrossRef]

37. Podstawczyńska, A. Differences of near-ground atmospheric Rn-222 concentration between urban and rural area with reference to microclimate diversity. *Atmos. Environ.* **2016**, *126*, 225–234. [CrossRef]

38. *NSW-EPA 2008 Calendar Year Air Emissions Inventory for the Greater Metropolitan Region in NSW*; Environment Protection Authority: Sydney, Australia, 2012.

39. Skamarock, W.C.; Klemp, J.B.; Dudhia, J.; Gill, D.O.; Barker, D.M.; Wang, W.; Powers, J.G. *A Description of the Advanced Research WRF Version 2 (No. NCAR/TN-468+ STR)*; National Center for Atmospheric Research: Boulder, CO, USA, 2005.

40. Byun, D.; Schere, K.L. Review of the governing equations, computational algorithms, and other components of the Models-3 Community Multiscale Air Quality (CMAQ) modeling system. *Appl. Mech. Rev.* **2006**, *59*, 51–77. [CrossRef]

41. US EPA Office of Research and Development. *CMAQv5.0.2 (Version 5.0.2)*. Zenodo. 30 May 2014. Available online: https://zenodo.org/record/1079898#.XDLv9s0RV7M (accessed on 7 January 2019). [CrossRef]

42. Dee, D.P.; Uppala, S.M.; Simmons, A.J.; Berrisford, P.; Poli, P.; Kobayashi, S.; Andrae, U.; Balmaseda, M.A.; Balsamo, G.; Bauer, D.P.; et al. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **2011**, *137*, 553–597. [CrossRef]

43. Thiébaux, J.; Rogers, E.; Wang, W.; Katz, B. A new high-resolution blended real-time global sea surface temperature analysis. *Bull. Am. Meteorol. Soc.* **2003**, *84*, 645–656. [CrossRef]

44. Morrison, H.; Thompson, G.; Tatarskii, V. Impact of Cloud Microphysics on the Development of Trailing Stratiform Precipitation in a Simulated Squall Line: Comparison of One- and Two-Moment Schemes. *Mon. Weather Rev.* **2009**, *137*, 991–1007. [CrossRef]

45. Iacono, M.J.; Delamere, J.S.; Mlawer, E.J.; Shephard, M.W.; Clough, S.A.; Collins, W.D. Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.* **2008**, *113*, D13103. [CrossRef]

46. Tewari, M.; Chen, F.; Wang, W.; Dudhia, J.; LeMone, M.A.; Mitchell, K.; Ek, M.; Gayno, G.; Wegiel, J.; Cuenca, R.H. Implementation and verification of the unified NOAH land surface model in the WRF model. In Proceedings of the 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, Washington, USA, 12–15 January 2004; pp. 11–15.

47. Janjic, Z.I. The Step—Mountain Eta Coordinate Model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Weather Rev.* **1994**, *122*, 927–945. [CrossRef]

48. Grell, G.A. Prognostic Evaluation of Assumptions Used by Cumulus Parameterizations. *Mon. Weather Rev.* **1993**, *121*, 764–787. [CrossRef]

49. Grell, G.A.; Devenyi, D. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.* **2002**, *29*, 1693. [CrossRef]

50. Emmons, L.K.; Walters, S.; Hess, P.G.; Lamarque, J.-F.; Pfister, G.G.; Fillmore, D.; Granier, C.; Guenther, A.; Kinnison, D.; Laepple, T.; et al. Description and evaluation of the Model for Ozone and Related chemical Tracers, version 4 (MOZART-4). *Geosci. Model Dev.* **2010**, *3*, 43–67. [CrossRef]

51. Crippa, M.; Guizzardi, D.; Muntean, M.; Schaaf, E.; Dentener, F.; van Aardenne, J.A.; Monni, S.; Doering, U.; Olivier, J.G.J.; Pagliari, V.; et al. Gridded Emissions of Air Pollutants for the period 1970–2012 within EDGAR v4.3.2. *Earth Syst. Sci. Data Discuss.* **2018**. [CrossRef]

52. Guenther, A.; Karl, T.; Harley, P.; Wiedinmyer, C.; Palmer, P.I.; Geron, C. Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature). *Atmos. Chem. Phys.* **2006**, *6*, 3181–3210. [CrossRef]

53. Kelly, J.T.; Bhave, P.V.; Nolte, C.G.; Shankar, U.; Foley, K.M. Simulating emission and chemical evolution of coarse sea-salt particles in the Community Multiscale Air Quality (CMAQ) model. *Geosci. Model Dev.* **2010**, *3*, 257–273. [CrossRef]

54. Appel, K.W.; Pouliot, G.A.; Simon, H.; Sarwar, G.; Pye, H.O.T.; Napelenok, S.L.; Akhtar, F.; Roselle, S.J. Evaluation of dust and trace metal estimates from the Community Multiscale Air Quality (CMAQ) model version 5.0. *Geosci. Model Dev.* **2013**, *6*, 883–899. [CrossRef]

55. Kaiser, J.W.; Heil, A.; Andreae, M.O.; Benedetti, A.; Chubarova, N.; Jones, L.; Morcrette, J.-J.; Razinger, M.; Schultz, M.G.; Suttie, M.; et al. Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences* **2012**, *9*, 527–554. [CrossRef]

56. Sarwar, G.; Luecken, D.; Yarwood, G.; Whitten, G.; Carter, B. Impact of an updated Carbon Bond mechanism on air quality using the Community Multiscale Air Quality modeling system: Preliminary assessment. *J. Appl. Meteorol.* **2008**, *1151*, 3–14. [CrossRef]

57. Whitten, G.Z.; Heo, G.; Kimura, Y.; McDonald-Buller, E.; Allen, D.T.; Carter, W.P.L.; Yarwood, G. A new condensed toluene mechanism for Carbon Bond: CB05-TU. *Atmos. Environ.* **2010**, *44*, 5346–5355. [CrossRef]

58. Roselle, S.J.; Schere, K.L.; Pleim, J.E.; Hanna, A.F. Photolysis rates for CMAQ. Science Algorithms of the EPA Models-3 Community Multiscale Air Quality (CMAQ) Modeling System, EPA/600/R-99/030. March 1999.

59. Stockwell, W.R.; Kirchner, F.; Kuhn, M.; Seefeld, S. A new mechanism for regional atmospheric chemistry modeling. *J. Geophys. Res. Atmos.* **1997**, *102*, 25847–25879. [CrossRef]

60. Janssens-Maenhout, G.; Dentener, F.; van Aardenne, J.; Monni, S.; Pagliari, V.; Orlandini, L.; Klimont, Z.; Kurokawa, J.; Akimoto, H.; O'Hara, T.; et al. EDGAR-HTAP: A harmonized gridded air pollution emission dataset based on national inventories. In *JRC Scientific and Technical Reports, EUR 25229 EN—2012*; European Union: Luxembourg, 2012.

61. Cope, M.E.; Hess, G.D.; Lee, S.; Tory, K.; Azzi, M.; Carras, J.; Lilley, W.; Manins, P.C.; Nelson, P.; Ng, L.; et al. The Australian Air Quality Forecasting System. Part I: Project description and early outcomes. *J. Appl. Meteorol.* **2004**, *43*, 649–662. [CrossRef]

62. Emmerson, K.M.; Galbally, I.E.; Guenther, A.B.; Paton-Walsh, C.; Guerette, E.A.; Cope, M.E.; Keywood, M.D.; Lawson, S.J.; Molloy, S.B.; Dunne, E.; et al. Current estimates of biogenic emissions from eucalypts uncertain for southeast Australia. *Atmos. Chem. Phys.* **2016**, *16*, 6997–7011. [CrossRef]

63. McGregor, J.L.; Dix, M.R. An updated description of the Conformal-Cubic atmospheric model. In *High Resolution Numerical Modelling of the Atmosphere and Ocean*; Ohfuchi, K.H.a.W., Ed.; Springer: New York, NY, USA, 2008; pp. 51–75.

64. Kowalczyk, E.A.; Stevens, L.; Law, R.M.; Dix, M.; Wang, Y.P.; Harman, I.N.; Haynes, K.; Srbinovsky, J.; Pak, B.; Ziehn, T. The land surface model component of ACCESS: Description and impact on the simulated surface climatology. *Aust. Meteorol. Ocean* **2013**, *63*, 65–82. [CrossRef]

65. Emmerson, K.M.; Cope, M.E.; Galbally, I.E.; Lee, S.; Nelson, P.F. Isoprene and monoterpene emissions in south-east Australia: Comparison of a multi-layer canopy model with MEGAN and with atmospheric observations. *Atmos. Chem. Phys.* **2018**, *18*, 7539–7556. [CrossRef]

66. McGregor, J.L. *C-CAM Geometric Aspects and Dynamical Formulation*; CSIRO: Aspendale, Australia, 2005; p. 43.

67. Wang, K.; Zhang, Y.; Yahya, K.; Wu, S.-Y.; Grell, G. Implementation and initial application of new chemistry-aerosol options in WRF/Chem for simulating secondary organic aerosols and aerosol indirect effects for regional air quality. *Atmos. Environ.* **2015**. [CrossRef]

68. He, J.; He, R.; Zhang, Y. Impacts of Air-Sea Interactions on Regional Air Quality Predictions Using a Coupled Atmosphere-Ocean Model in Southeastern U.S. *Aerosol. Air Qual. Res.* **2018**, *18*, 1044–1067. [CrossRef]

69. Zhang, Y.; Jena, C.; Wang, K.; Paton-Walsh, C.; Guerette, E.-A.; Utembe, S.; Silver, J.D.; Keywood, M. Multiscale Applications of Two Online-Coupled Meteorology-Chemistry Models during Recent Field Campaigns in Australia: Evaluation, Intercomparison, and Sensitivity Simulations. *Atmosphere* **2018**, submitted.

70. Dameris, M.; Jöckel, P. Numerical Modeling of Climate-Chemistry Connections: Recent Developments and Future Challenges. *Atmosphere* **2013**, *4*, 132–156. [CrossRef]

71. Zhang, Y.; Bocquet, M.; Mallet, V.; Seigneur, C.; Baklanov, A. Real-time air quality forecasting, part I: History, techniques, and current status. *Atmos. Environ.* **2012**, *60*, 632–655. [CrossRef]

72. Mahrt, L. Stratified atmospheric boundary layers and breakdown of models. *Theoret. Comput. Fluid Dyn.* **1998**, *11*, 263–279. [CrossRef]

73. Cuxart, J.; Holtslag, A.A.M.; Beare, R.J.; Bazile, E.; Beljaars, A.; Cheng, A.; Conangla, L.; Ek, M.; Freedman, F.; Hamdi, R.; et al. Single-column model intercomparison for a stably stratified atmospheric boundary layer. *Bound.-Lay. Meteorol.* **2006**, *118*, 273–303. [CrossRef]

74. Price, J.D.; Vosper, S.; Brown, A.; Ross, A.; Clark, P.; Davies, F.; Horlacher, V.; Claxton, B.; McGregor, J.R.; Hoare, J.S.; et al. COLPEX—Field and numerical studies over a region of small hills. *Bull. Am. Meteorol. Soc.* **2011**, *92*, 1636–1650. [CrossRef]

75. Seidel, D.J.; Zhang, Y.; Beljaars, A.; Golaz, J.-C.; Jacobson, A.R.; Medeiros, B. Climatology of the planetary boundary layer over the continental United States and Europe. *J. Geophys. Res. Atmos.* **2012**, *117*. [CrossRef]

76. Hong, S.Y.; Noh, Y.; Dudhia, J. A New Vertical Diffusion Package with an Explicit Treatment of Entrainment Processes. *Mon. Weather Rev.* **2006**, *134*, 2318–2341. [CrossRef]

77. Griffiths, A.D.; Chambers, S.D.; Williams, A.G.; Werczynski, S. Increasing the accuracy and temporal resolution of two-filter radon–222 measurements by correcting for the instrument response. *Atmos. Meas. Tech.* **2016**, *9*, 2689–2707. [CrossRef]

78. Schell, B.; Ackermann, I.J.; Hass, H.; Binkowski, F.S.; Ebel, A. Modeling the formation of secondary organic aerosol within a comprehensive air quality model system. *J. G. R. Atmos.* **2001**, *106*, 28275–28293. [CrossRef]