Supplement of Geosci. Model Dev., 12, 4571–4584, 2019
https://doi.org/10.5194/gmd-12-4571-2019-supplement

Geoscientific
Model Development
Open Access

Supplement of

Model evaluation of high-resolution urban climate simulations: using the WRF/Noah LSM/SLUCM model (Version 3.7.1) as a case study

Zhiqiang Li et al.

Correspondence to: Bo Huang (bohuang@cuhk.edu.hk)

S1 The Urban Climate Modelling System

We used the Advanced Research WRF (ARW) modelling system coupled with Noah LSM/UCM Model (Noah Land Surface Model/Urban Canopy Model) developed by National Center of Atmospheric Research in this study for the urban climate simulation. It is a limited-area, non-hydrostatic and meso-scale atmospheric modelling system with the terrain-following mass vertical coordinate, designed for atmospheric research applications (Skamarock et al., 2005, 2008; Lo et al., 2008). The ARW model is a typical atmospheric model integrating with a set of five interacting physical components (Microphysics, Cumulus Parameterization, Radiation, Planetary Boundary Layer/Vertical Diffusion and Surface) (Skamarock et al., 2005, 2008).

The Noah LSM model is coupled with the ARW model by the surface component. The in-homogeneity of the surface affects energy and mass redistribution in the atmosphere. The Noah LSM model utilizes the following parameters in the representation of the in-homogeneous texture of the surface to simulate the land surface process.

- Land use
- Land covers (vegetation)
- Soil texture
- Secondary parameters related to the above three primary parameters

S2 The 2010 PRD Urban Land Surface Dataset

The 2010 PRD Urban Land Surface Dataset includes the land cover data, vegetation coverage data, urban morphology data, and anthropogenic heat data, for which the spatial resolution is 1-km^2. The vegetation coverage includes 12 monthly vegetation coverage maps. The urban morphology data includes the urban fraction, area fraction of building plan area, mean building height area weighted, standard deviation of mean building height area weighted, building surface area to building plan area ratio, building height number mean, and frontal area index. The anthropogenic heat data includes the anthropogenic sensible heat and the anthropogenic latent heat.

S3 Design of the Four-days Simulation Segment

An atmosphere model initially needs to run for a period of time in order to stabilize its condition, which is called as the initial-model-run period. Modelling results during this period is frustrating. Usually, to reduce the negative effect of model instability in the initial-model-run period, the model result data of the initial-model-run period is discarded. This procedure is called model spin-up. However, there is no statistical report told the modellers how long the spin-up for running a model is the best (Kleczek et al. 2014).

Nevertheless, a minimum spin-up time of 12 hours is necessary for balancing the NWP model (Jankov et al. 2007; Skamarock and Klemp, 2008; Kleczek et al. 2014). Moreover, practically in the atmosphere modeller community that the longer a simulation period is, the longer of the spin-up time is required. Similarly, if the period of a weather simulation case is too long, the results of the last few days might be distorted. Therefore, the appropriate simulation and spin-up periods would improve the quality of the result in a weather simulation case. To sum up, the period of a simulation segment was set to 4 days and the first day was used for the spin-up period.

In this study, the one-year urban climate simulation case was divided into a series of sequent simulation segments. The first day of the next simulation segment overlaps with the last day of the previous simulation segment. The sequence of simulation segments for an urban climate simulation case is shown in Figure S1.

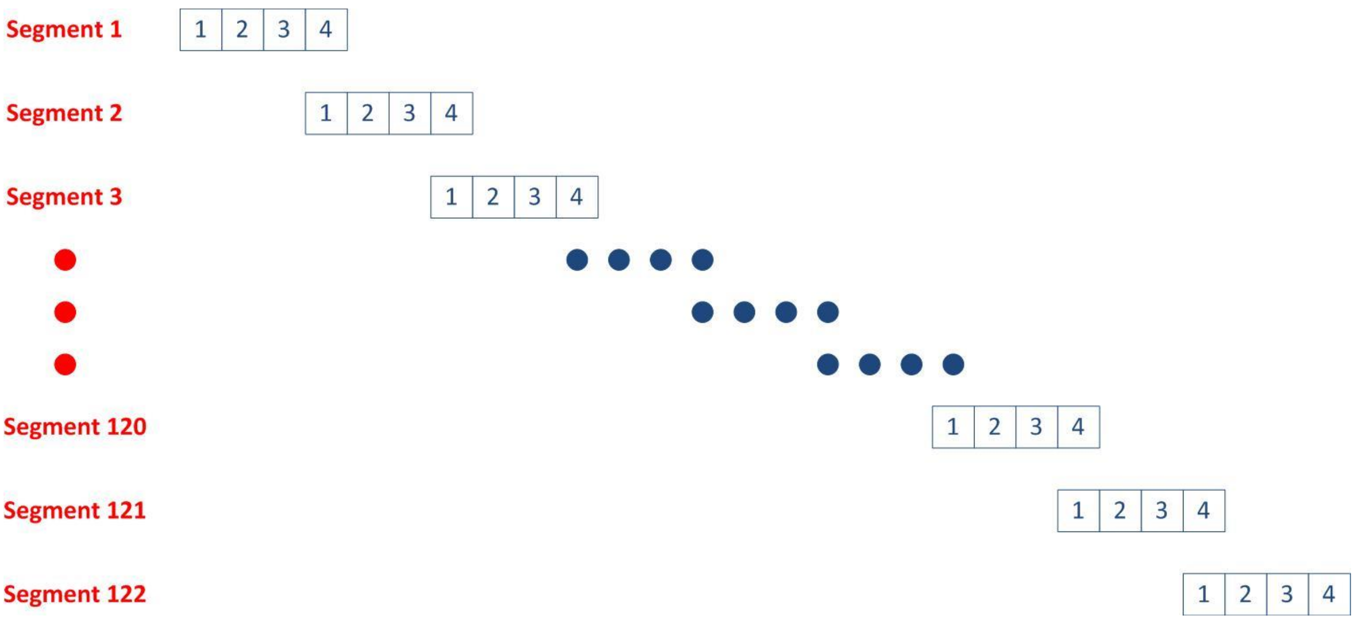


Figure S1: The Simulation Segments' Sequence.

S4 Schemes Choosing for Physics Components of WRF ARW/Noah LSM/SLUCM

There are 7 physical components in the model, and each component has different candidate schemes.

(1) Cumulus Parameterization

(2) Microphysics

(3) Radiation

(4) Planetary Boundary Layer

(5) Surface layer

(6) Land Surface Model

(7) Urban Canopy Model

First, Cumulus Parameterization directly outputs the effects of physics process rather than simulates the physics process itself (Chen, 2011). The New Simplified Arakawa-Schubert scheme was chosen for Cumulus Parameterization, which supports deep and shallow convection and momentum transport for the ARW core.

Second, the microphysics component is responsible for the processes of resolved water, cloud and precipitation (Skamarock et al., 2005, 2008). Based on the sophistication ranking of the schemes, the WDM5 scheme was chosen for all domains in this study.

Third, the Radiation component simulates the atmospheric radiation processes in a vertical column of a horizontal grid. It consists of longwave and shortwave modules and uses the precipitation, water vapor and cloud-related variables as the inputs. It also exchanges the radiation fluxes related variables with the surface component and updates the potential temperature-related variables. The RRTMG scheme also was chosen as the shortwave and longwave radiation scheme because it supports the climatological ozone and aerosol data input which interacts with microphysics and cumulus parameterization components by Qc, Qr, Qi, and Qs.

Fourth, the land-surface models (LSMs) requires the input of the atmospheric data from surface layer component, the downward radiative fluxes from radiation component, and the precipitation data from microphysics and cumulus parameterization components. It outputs the heat and moisture fluxes data to PBL component and the upward radiative fluxes to radiation component over land and sea-ice points (Skamarock et al., 2008). The Noah LSM was chosen as the scheme of land-surface model for the inner-most horizontal domain because the urban canopy model can only be coupled with Noah LSM.

Fifth, the Urban Canopy Model component is responsible for the physical processes of the land surface in urban environment. The Single-Layer Urban Canopy Model was chosen as the scheme of the Urban Canopy Model because of its sophistication.

Sixth, the Planetary Boundary Layer component simulates the vertical heat, moisture and momentum fluxes vertical diffusion which are caused by the turbulence exchanges in a whole column of a grid (Wang, 2014, 2015). The Bougeault–Lacarrere scheme was chosen for all domains because it supports the NUDAPT format data that used in this study for taking the urban morphology into account.

Seventh, in the WRF model, the Surface Layer component is responsible for providing the friction stress, the surface fluxes of heat and moisture to PBL (Skamarock et al., 2008). As a result of a comparison, the revised MM5 surface layer scheme was chosen for all horizontal domains in this study.

S5 Guideline for Model Evaluation

We used four statistic figures (Table S1) in the practice of model evaluation. We first conducted a data processing operation (run by a program) on each pair of the raw observation dataset, and its corresponding modelled result to produce an evaluation 3D-matrix which consisted of a one-year temporal series of 2D-matrixes. Each non-empty element of the 2D-matrixes geographically corresponded with a pair of observation data and its corresponding modelled one at a time point. A location map of meteorological observations was also produced if the raw observation dataset was a meteorological observation dataset. Secondly, we conducted the figure plotting operations (using the programs) on each evaluation 3D-matrix step by step to produce a series of statistic figures. Specially, we designed a guideline (Table S2) for specifying the intervals in the PDFD figure, which were used for measuring the accuracy. Finally, we conducted grading or checking operations on these statistic figures artificially based on the guidelines for grading (Table S3) or checking (Table S4).

Table S1: The Statistic Figures for Model Evaluation.

Statistical Perspectives	**Methods**	
	Statistic Figures	**Usage**
Descriptive Statistics	Temporal Comparison of Spatial Variation (TCSV)	It was used to temporally compare two variables' spatial variation ranges and median in the whole year.
	Diurnal Variation (DV)	It was used to temporally compare the diurnal variations of two variables' spatial variation ranges and median in the whole year.

Statistical Distributions	Probability Density Function of Difference (PDFD)	It was used to show the probability density of difference between the modelled variable and its corresponding observed one.
	Perkins Skill Score (PSS)	It was used for revealing quantifiably the extent of overlap between the observed and modelled variables' Probability Density Function (PDF). A value of 1 indicated a perfect modelling of the observation. On the contrary, a value of 0 meant the worst simulation.

Table S2: The Guideline for Specifying the Interval in a PDFD Figure.

No.	**The Range of Coefficient (A)**	**Intervals**
1	0.1 – 0.2	
2	0.2 – 0.4	[-Aσ, Aσ]
3	0.4 – 0.6	

Remark: σ is the annual mean value of the monthly standard deviation of the modelled variable

Table S3: The Guideline for Grading.

Range		**Grading**
PSS	**Accuracy(a)**	
$0.7 \leq PSS \leq 1$	$70\% \leq a \leq 100\%$	Good
$0.7 > PSS \geq 0.5$	$70\% > a \geq 50\%$	Acceptable
$PSS < 0.5$	$a < 50\%$	Unacceptable

Remark: Accuracy is the PDFD value of interval 2 or interval 3.

Table S4: The Guideline for Checking.

Statistic Figures	**Temporal Perspectives for Checking**		
	Annual	**Monthly**	**Daily**
TCSV	Annual Pattern	Monthly Pattern	
DV	Annual Pattern		Diurnal pattern

S6 Observation Datasets

A quality control had been applied to all meteorological observation datasets by the data provider. Table S1 shows the total number of observations and the numbers of observations in urban area and non-urban area. Moreover, Figures S2, S3, S4, and S5 show the locations of the meteorological observations.

Table S1: The Numbers of Meteorological Observations

Type	**Number in urban area**	**Number in non-urban area**	**Total number**
2-meters temperature	34	23	57
10-meters wind speed	37	26	62
Precipitation	31	33	64
Relative humidity	15	9	24

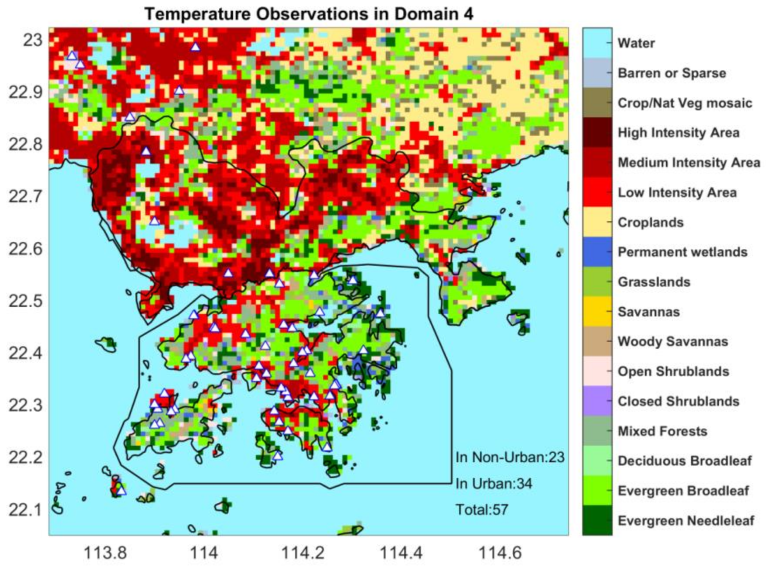


Figure S2: Temperature Observations in Domain 4.

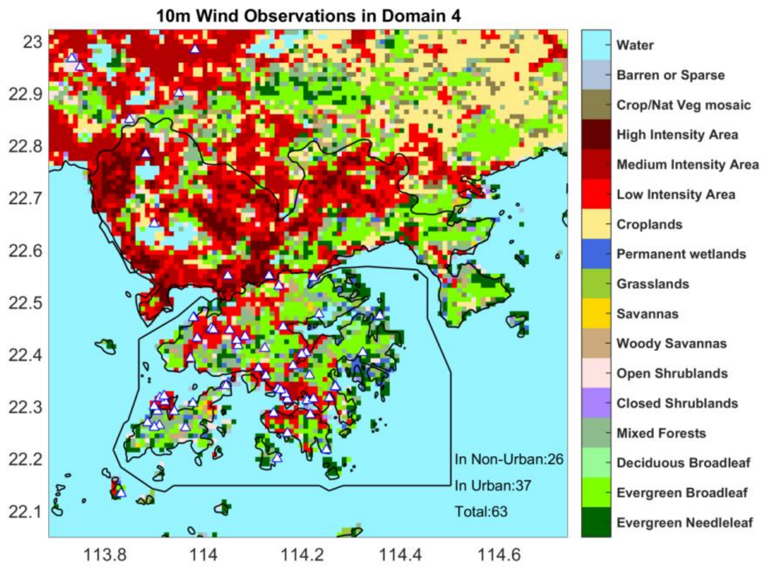


Figure S3: The 10-Meters Wind Observations in Domain 4.

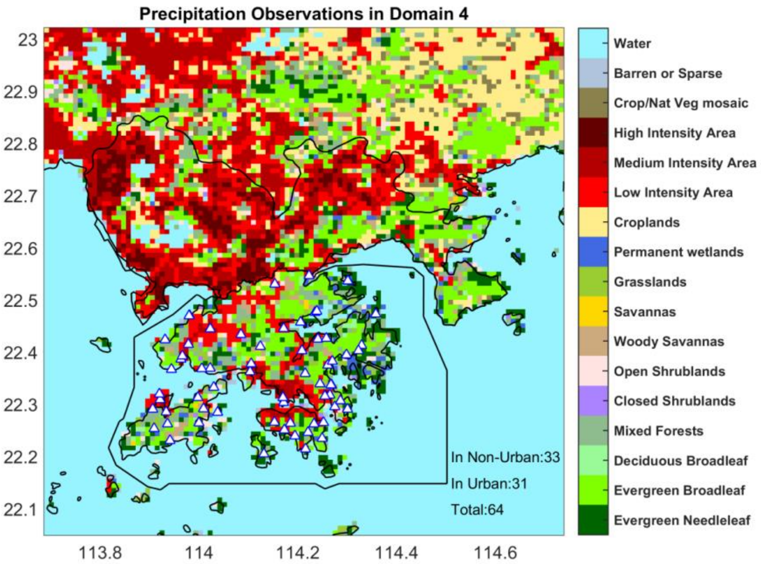


Figure S4: Precipitation Observations in domain 4.

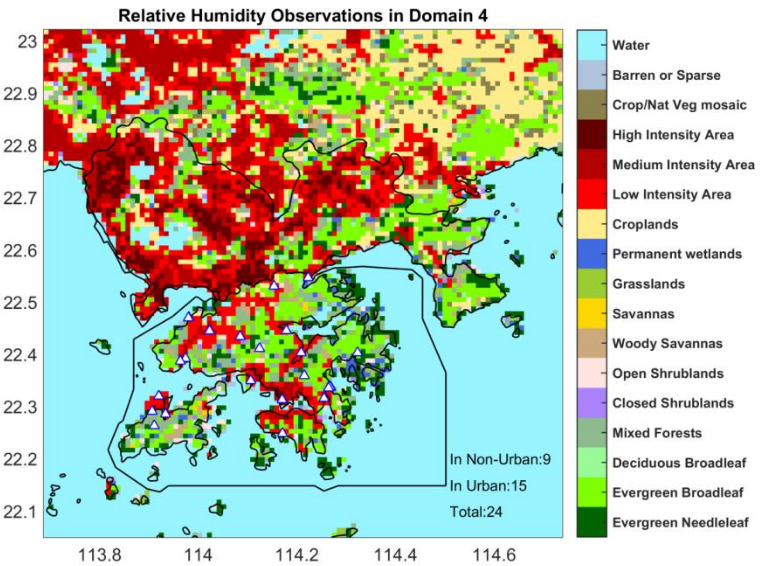


Figure S5: Relative Humidity Observations in domain 4.

The MODIS/Aqua Land Surface Temperature and Emissivity (LST/E) product (Short name: MYD11A1) provided by the U.S. Geological Survey (USGS) was used for the evaluation. This product includes a grid surface temperature with a 1-km horizontal resolution at around 2:00 and 14:00 (Beijing time) per day. It also has a quality control attribute for each surface temperature record to identify the level of data quality. Such quality control attribute was used for filtering the poor-quality records that were at least 5 degrees' different from the corresponding modelled value.

S7 Figures for Comparisons in Surface Temperature, 10-Meters Wind Speed, Precipitation and Relative Humidity

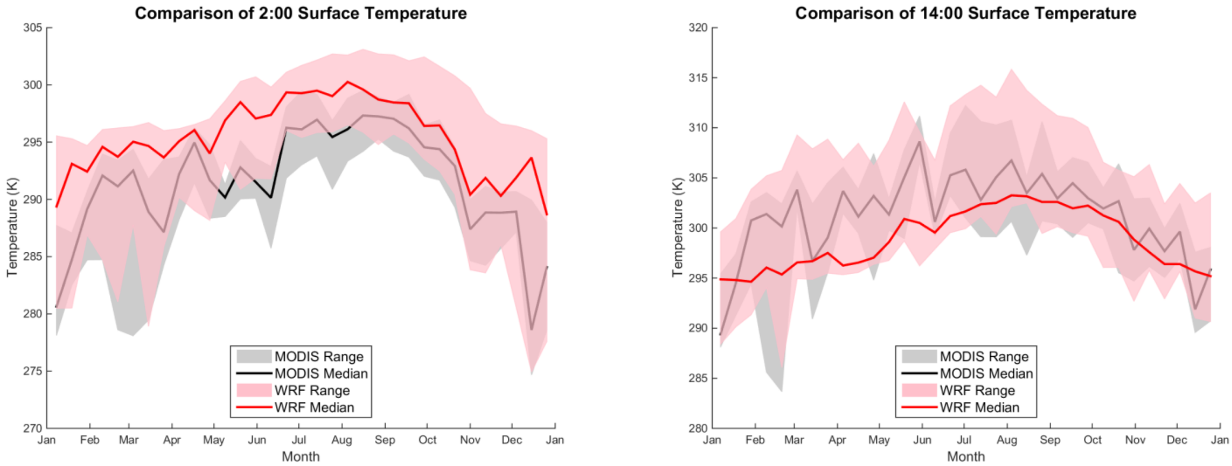


Figure S6: Comparison of Modelled and MODIS Surface Temperatures at 2:00 and 14:00.

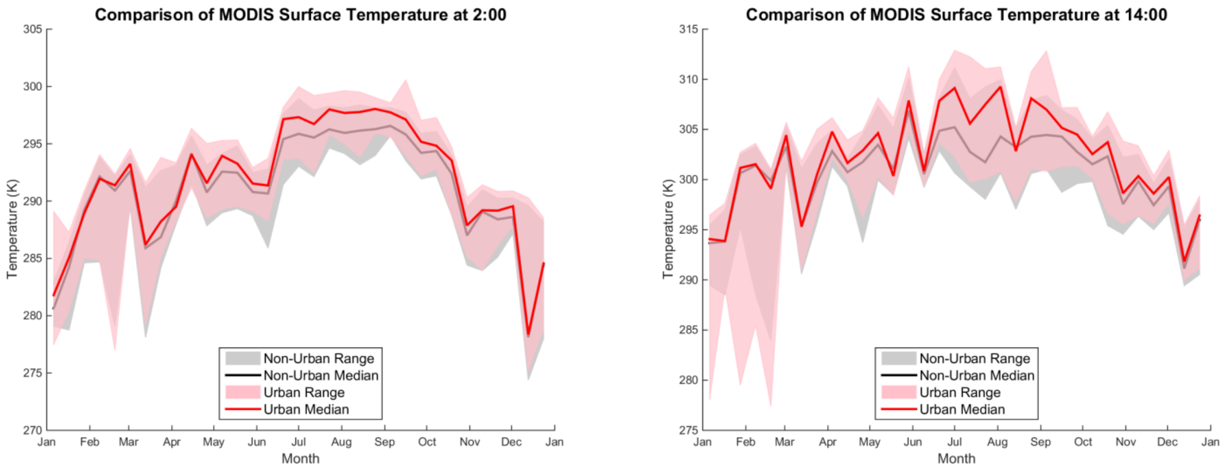


Figure S7: Comparison of MODIS Surface Temperatures (at 2:00 and 14:00) in Urban and Non-Urban Areas.

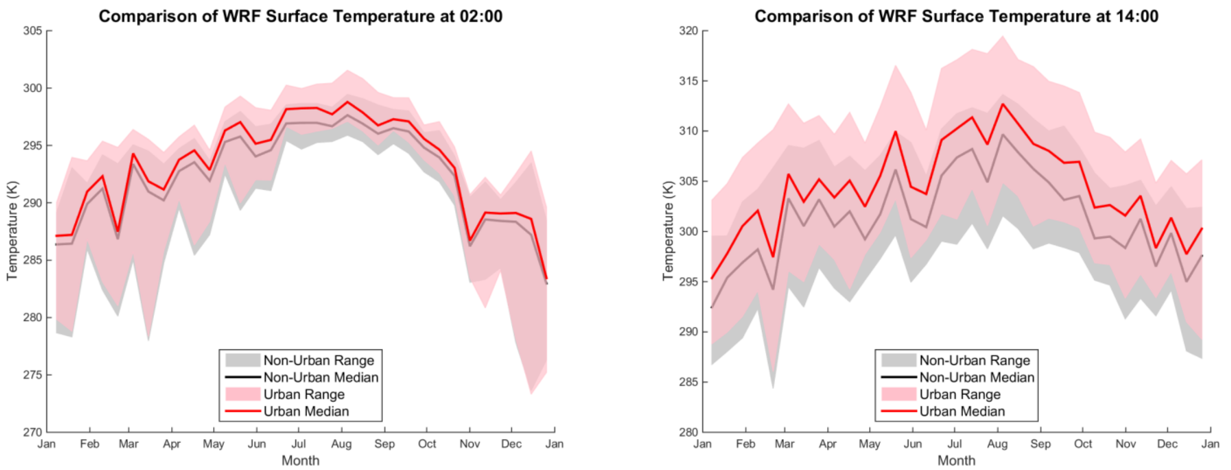


Figure S8: Comparison of Modelled and Observed Surface Temperatures (at 2:00 and 14:00) in Urban Area and the Non-Urban area.

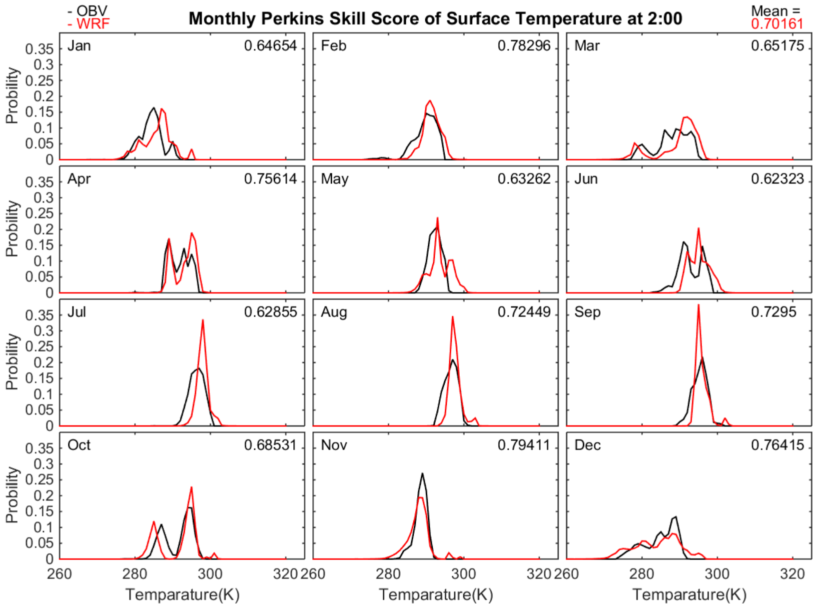


Figure S9: Monthly PSS of Surface Temperature at 2:00.

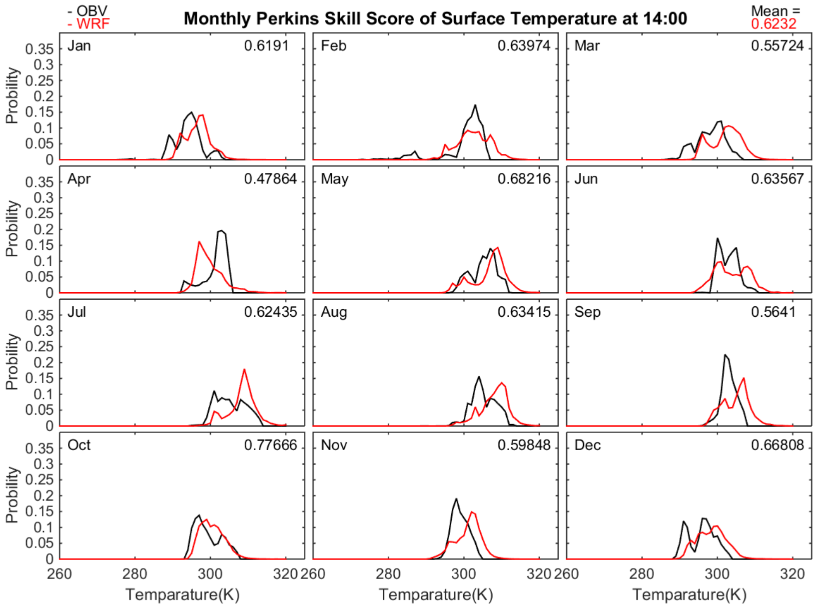


Figure S10: Monthly PSS of Surface Temperature at 14:00.

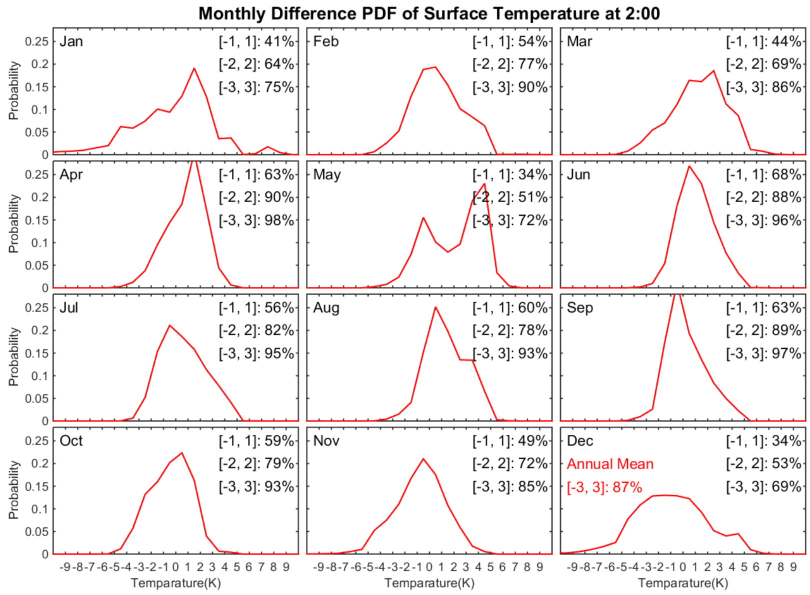


Figure S11: Monthly PDF of 2:00 Surface Temperature Difference.

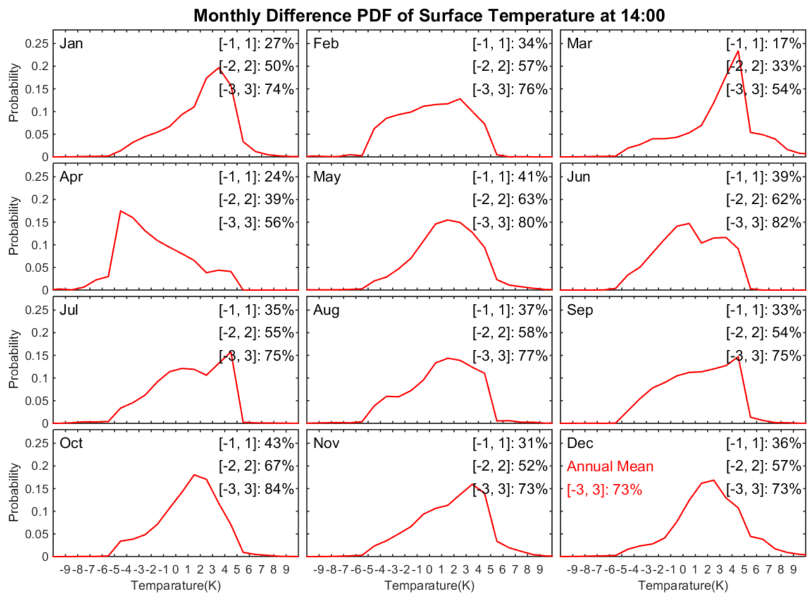
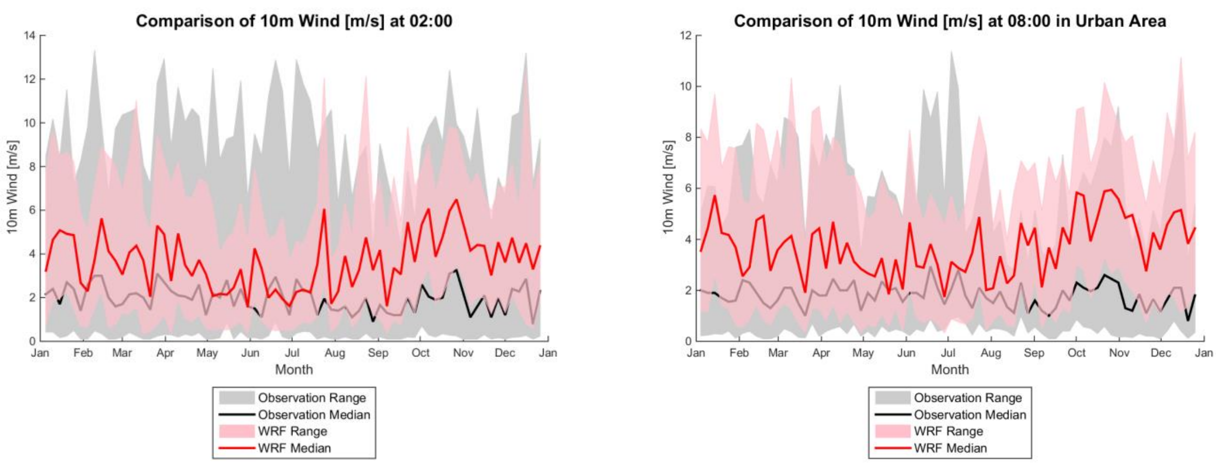


Figure S12: Monthly PDF of 14:00 Surface Temperature Difference.



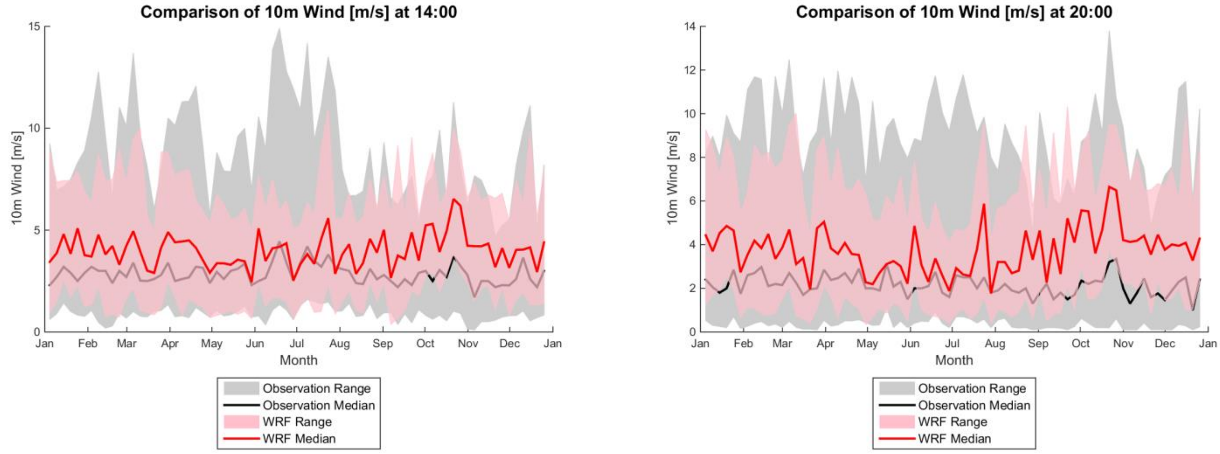


Figure S13: Comparisons of Modelled and Observed 10-Meters Wind Speed at 2:00, 8:00, 14:00 and 20:00.

Figure S14: Comparison of Observed 10-Meters Wind (at 2:00, 8:00, 14:00 and 20:00) in Urban Area and in the Non-Urban area.

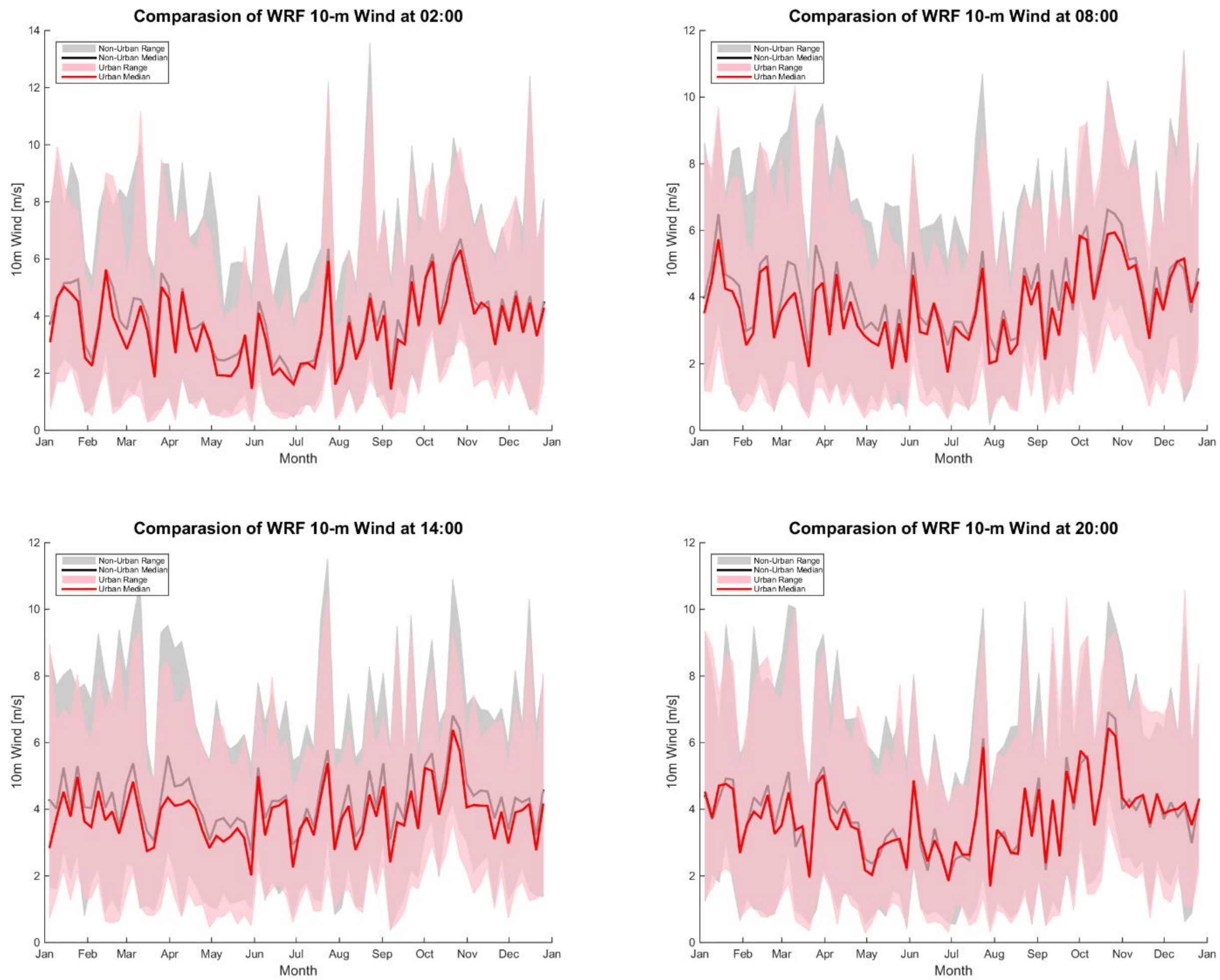


Figure S15: Comparison of WRF 10-meters Wind (at 2:00, 8:00, 14:00 and 20:00) in Urban Area and in the Non-Urban area.

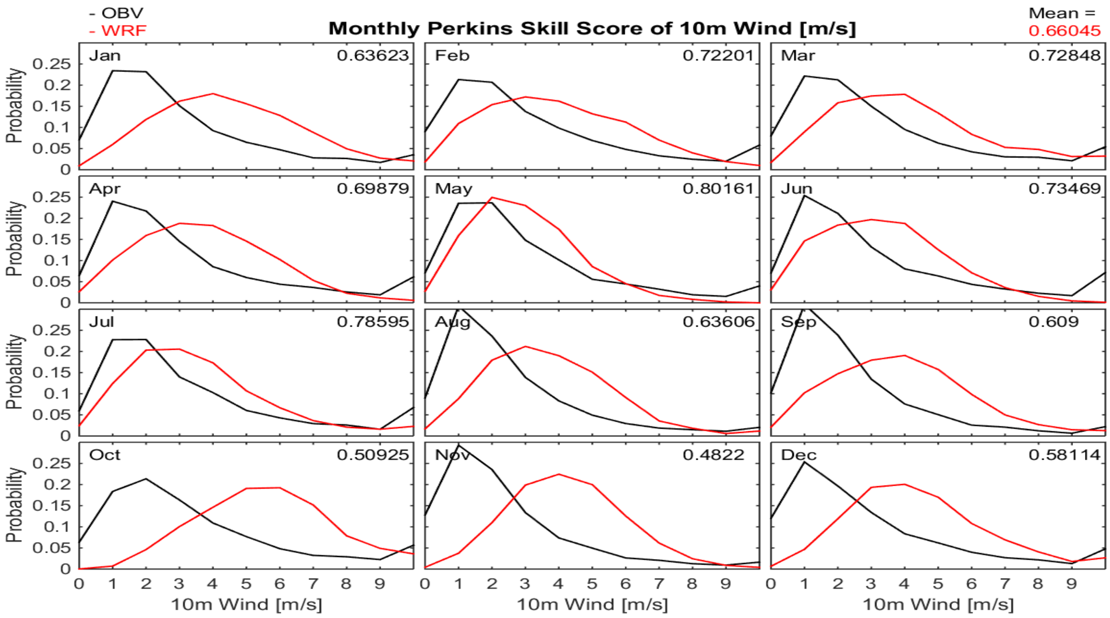


Figure S16: Monthly PSS of 10-Meters Wind Speed.

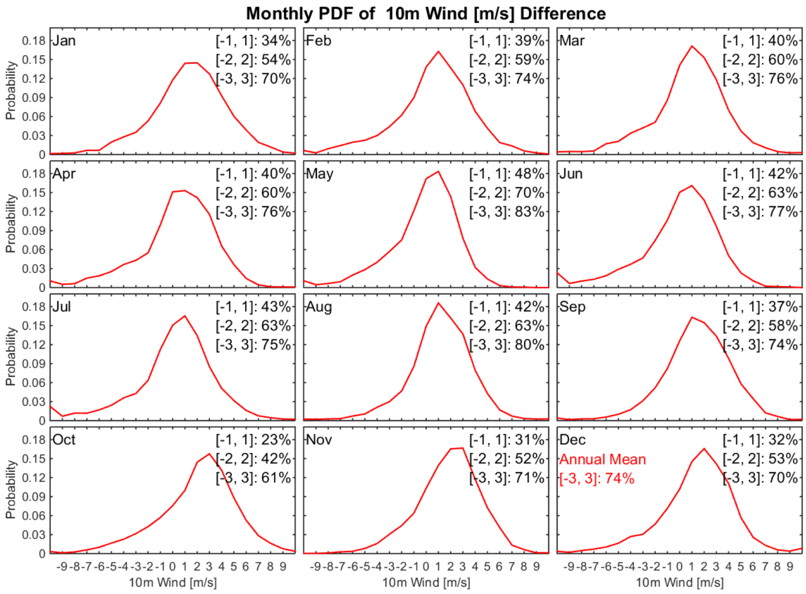


Figure S17: Monthly PDF of 10-Meters Wind Speed Difference.

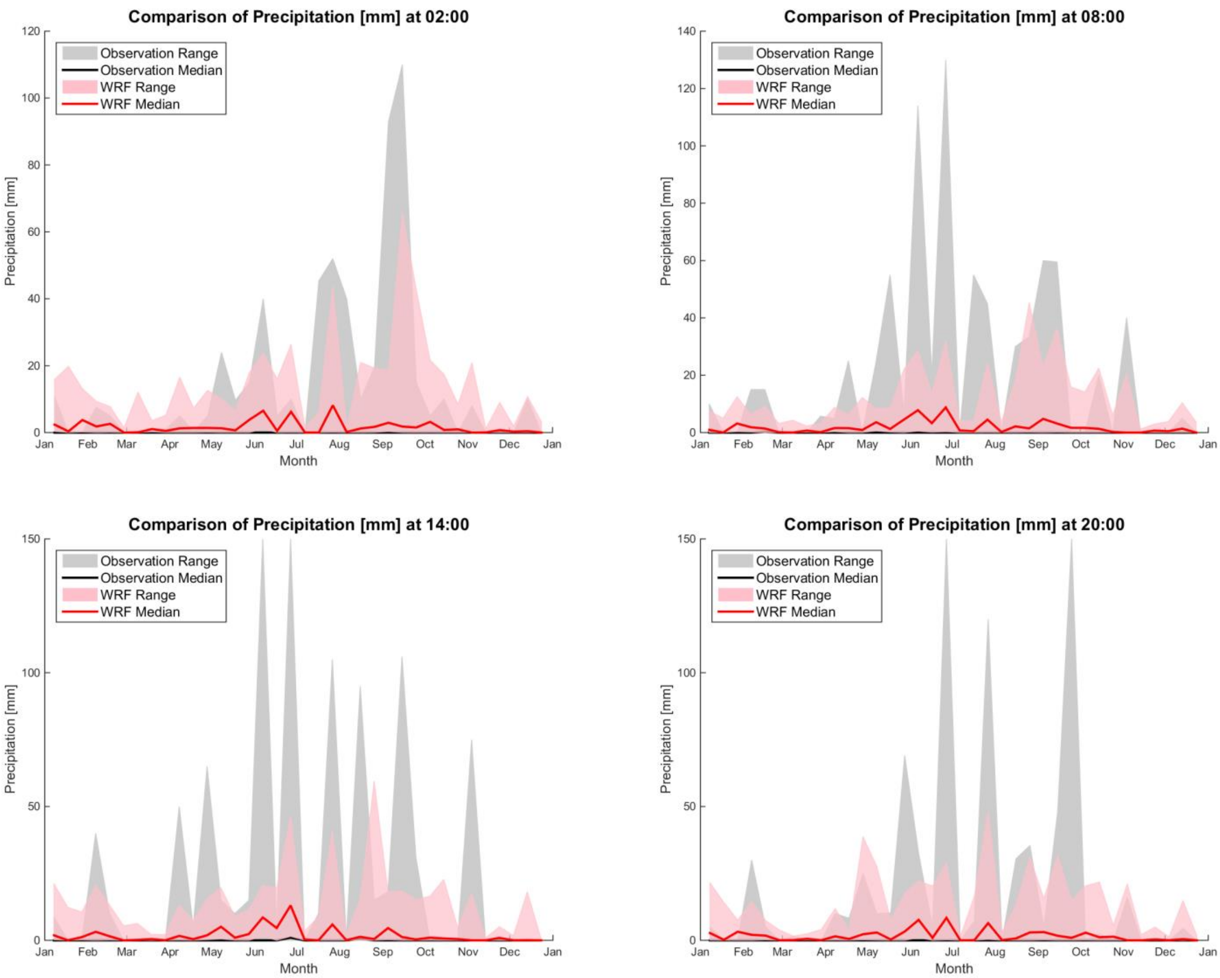


Figure S18: Comparison of Modelled and Observed Precipitations at 2:00, 8:00, 14:00, and 20:00.

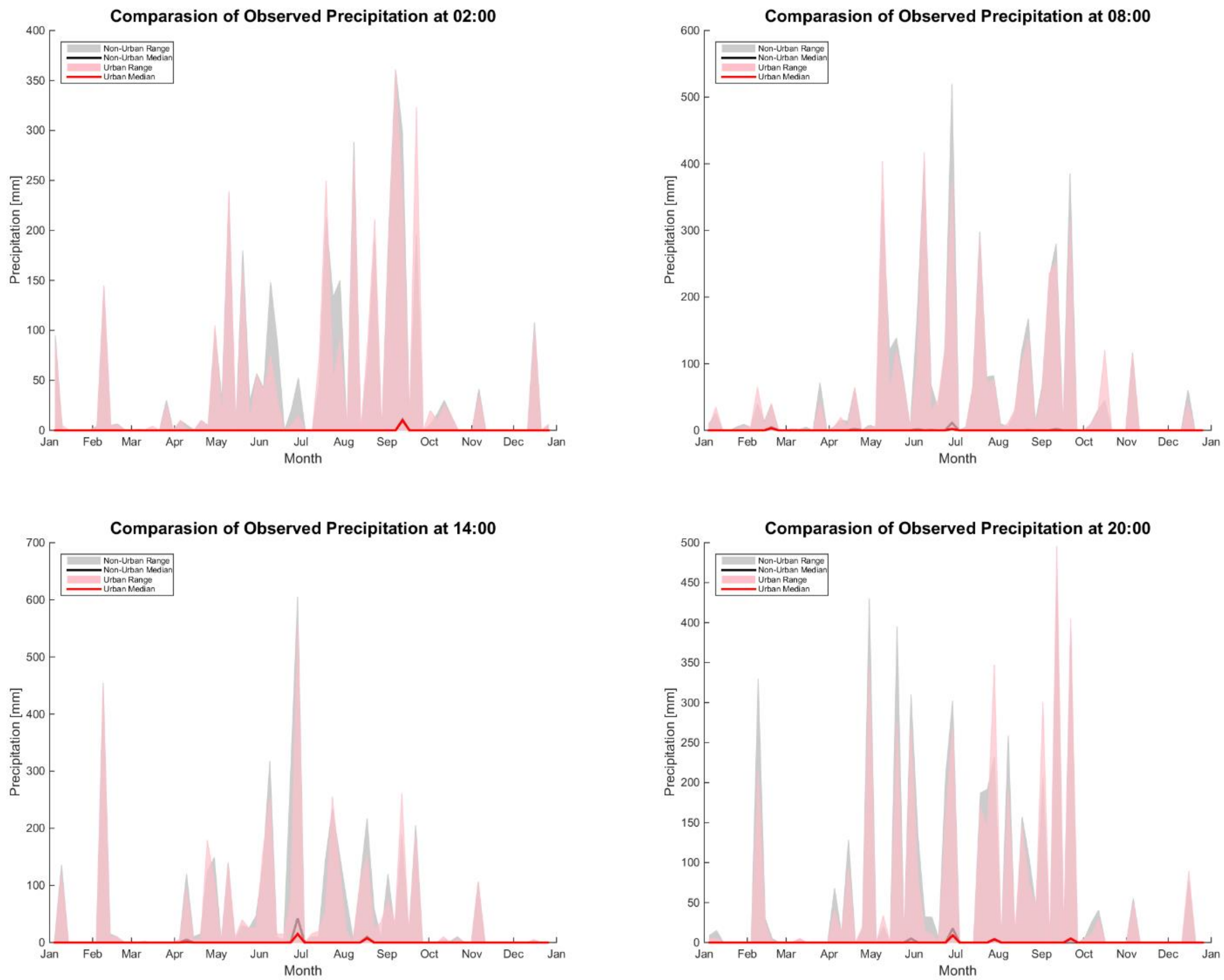
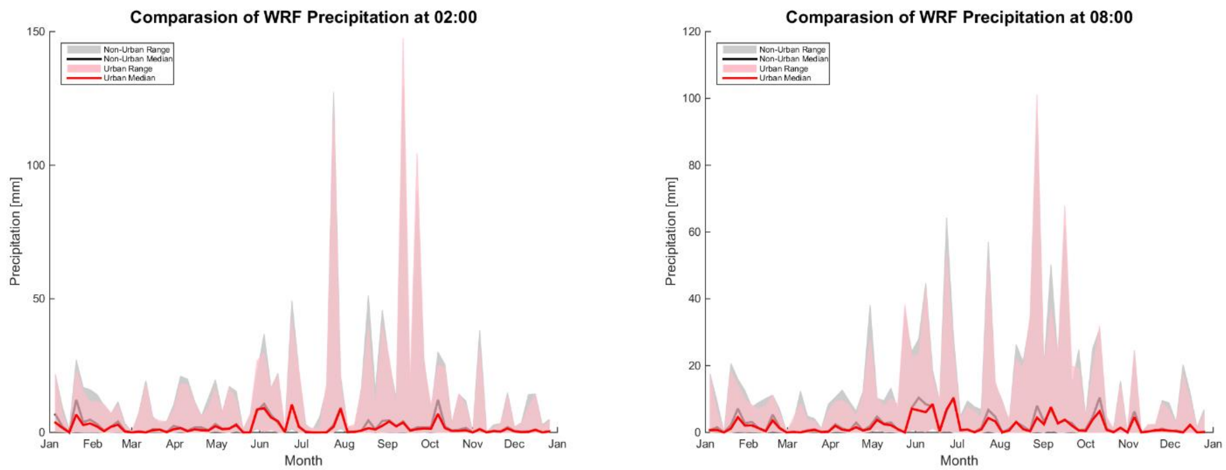


Figure S19: Comparison of Observed Precipitation (at 2:00, 8:00, 14:00 and 20:00) in Urban Area and in the Non-Urban area.



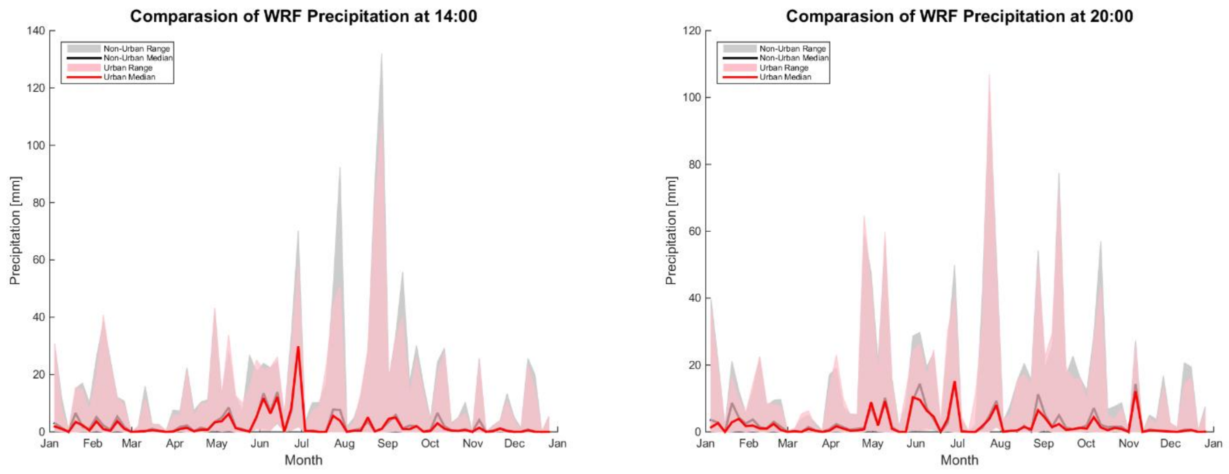


Figure S20: Comparison of WRF Precipitation (at 2:00, 8:00, 14:00 and 20:00) in Urban Area and in the Non-Urban area.

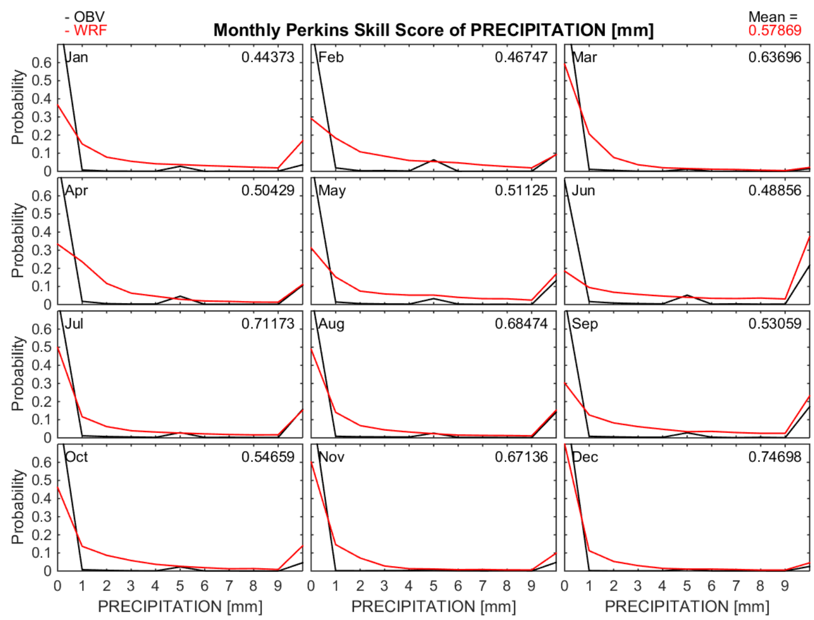


Figure S21: Monthly PSS of Precipitation.

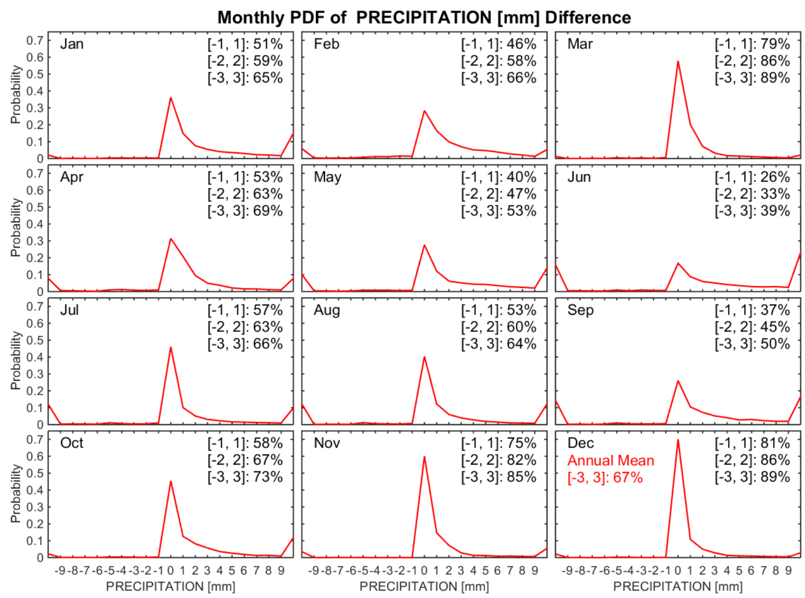


Figure S22: Monthly PDF of Precipitation Difference.

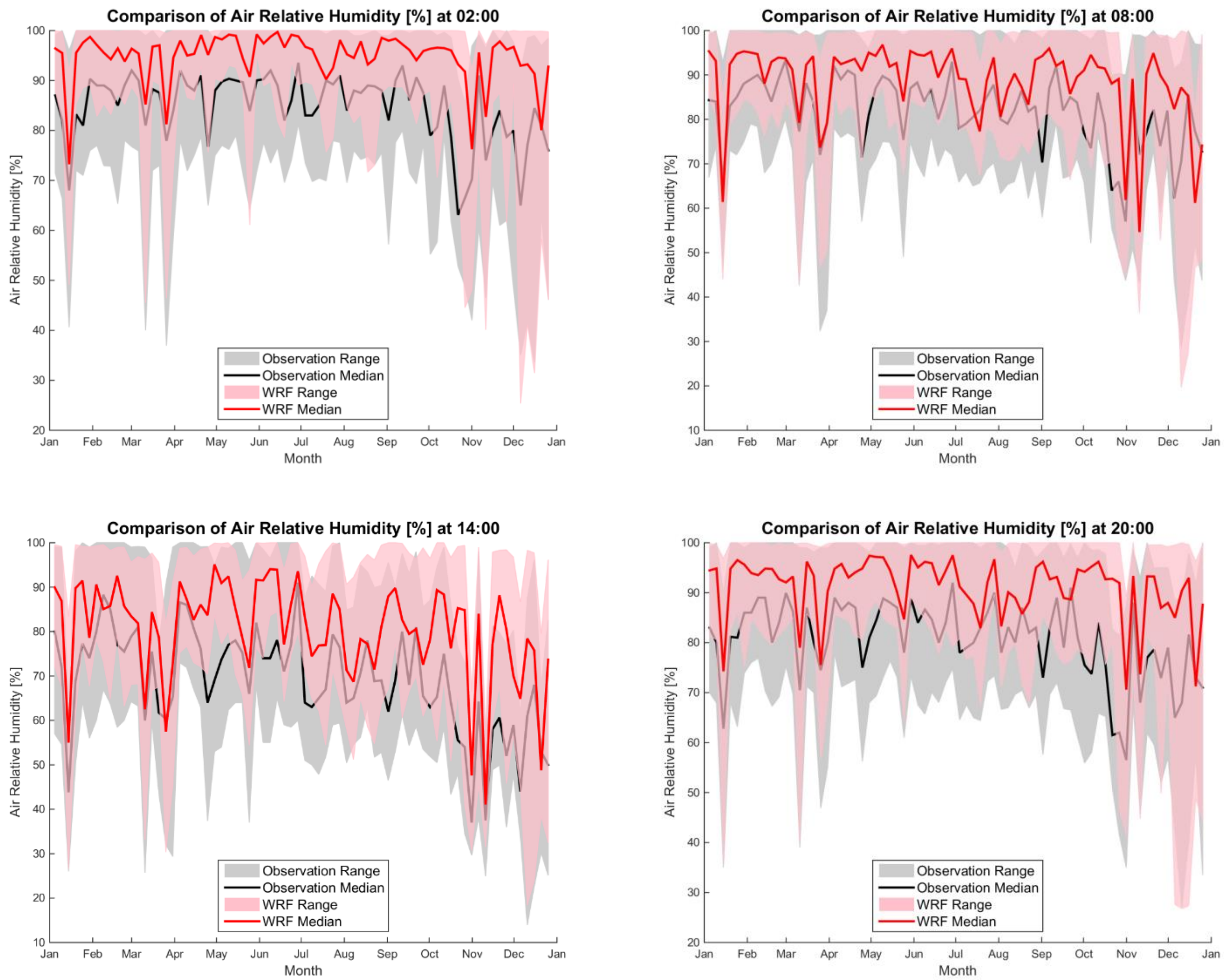
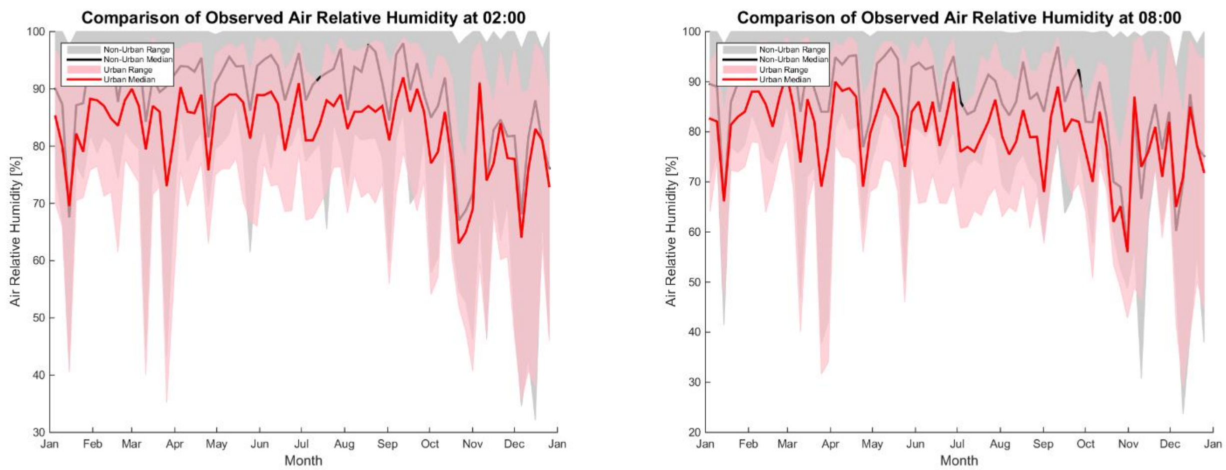


Figure S23: Comparison of Modelled and Observed Relative Humidity at 2:00, 8:00, 14:00, and 20:00.



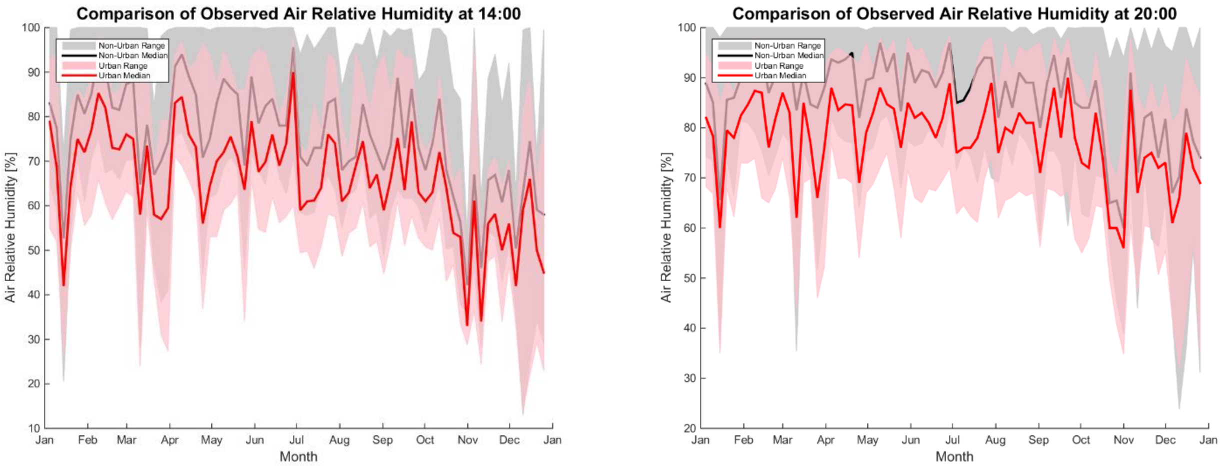


Figure S24: Comparison of Observed Humidity (at 2:00, 8:00, 14:00 and 20:00) in Urban and in Non-Urban Areas.

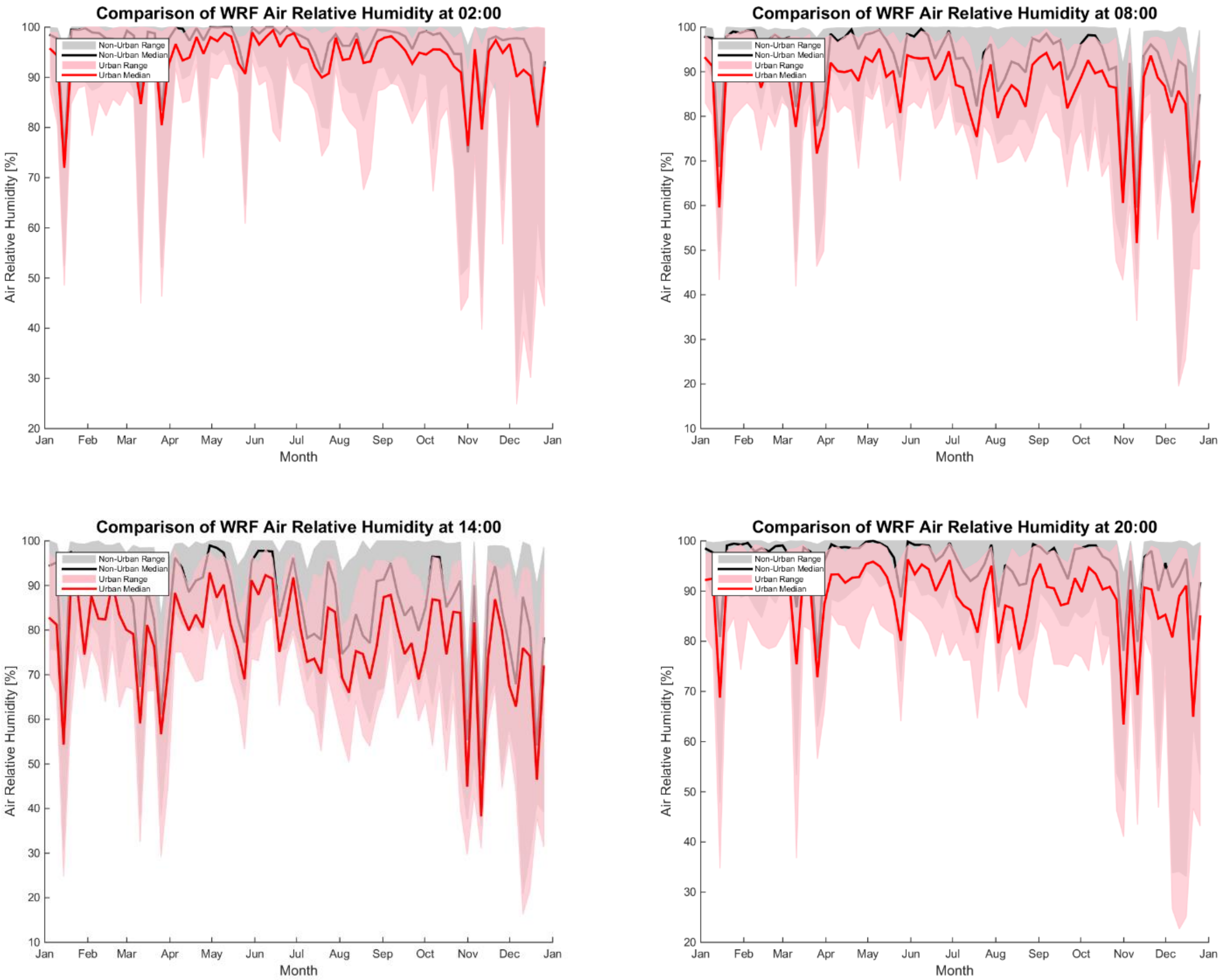


Figure S25: Comparison of Modelled Relative Humidity (at 2:00, 8:00, 14:00 and 20:00) in Urban and Non-Urban Areas.

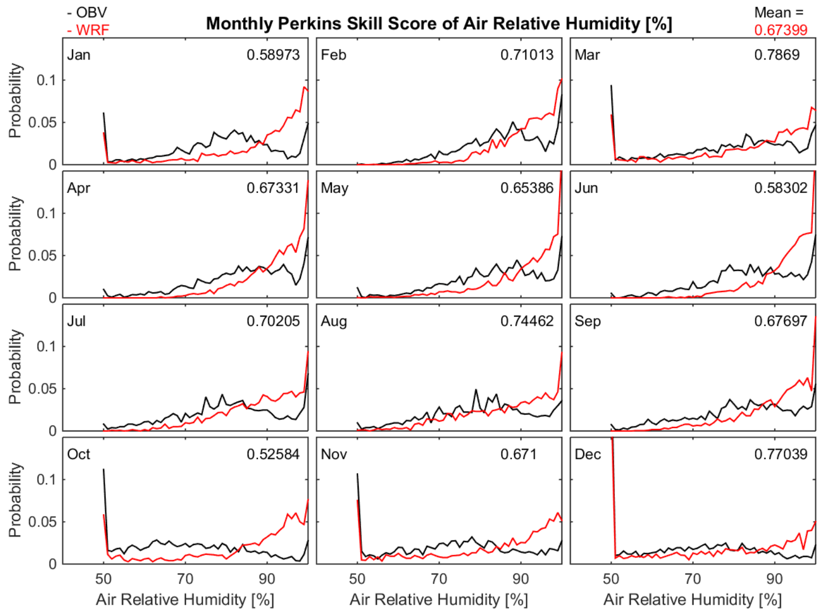


Figure S26: Monthly PSS of Relative Humidity.

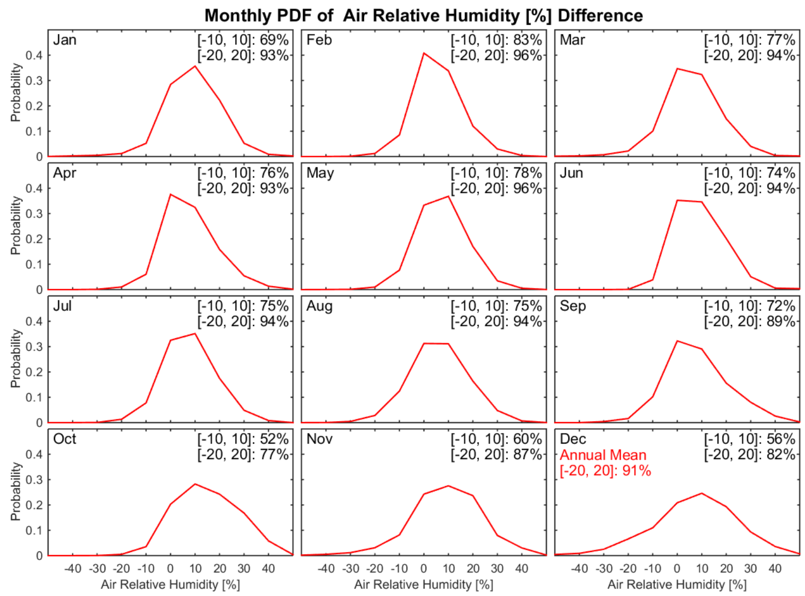
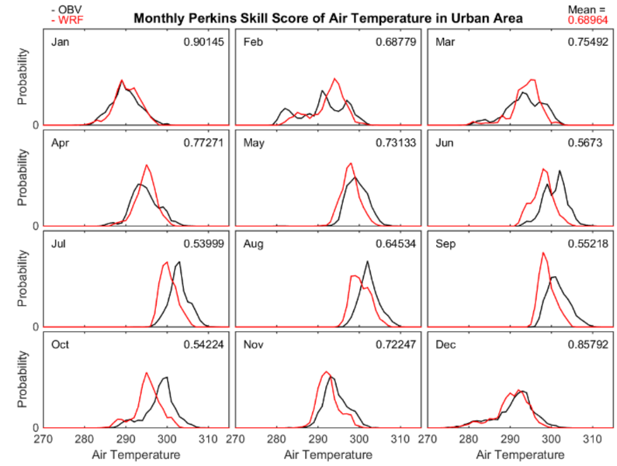


Figure S27: Monthly PDF of Humidity Difference.

S8 The PSSs in Urban and Non-Urban Areas

(a) **(b)**



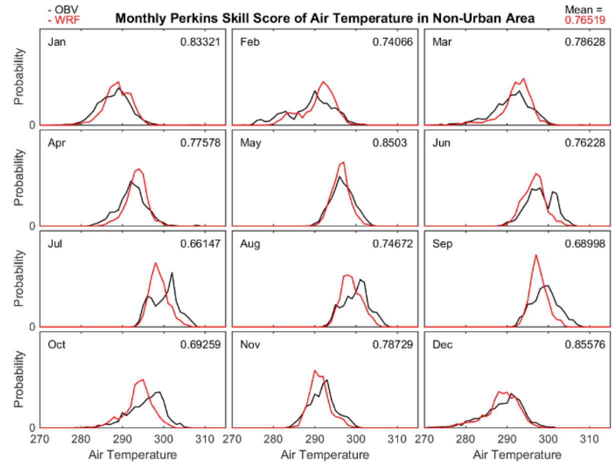
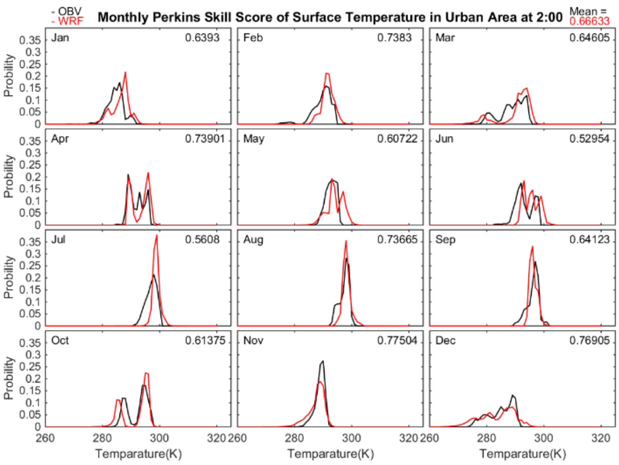


Figure S28: Monthly PSS of 2-Meters Air Temperature in Urban (a) and Non-Urban (b) Area.

(a) **(b)**



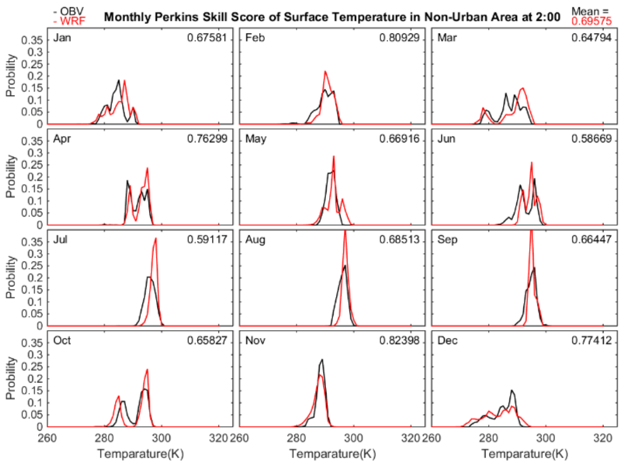
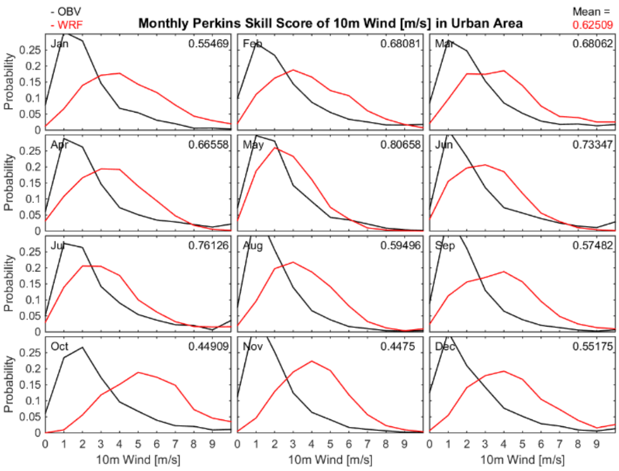


Figure S29: Monthly PSS of Surface Temperature in Urban (a) and Non-Urban (b) Area.

(a) **(b)**



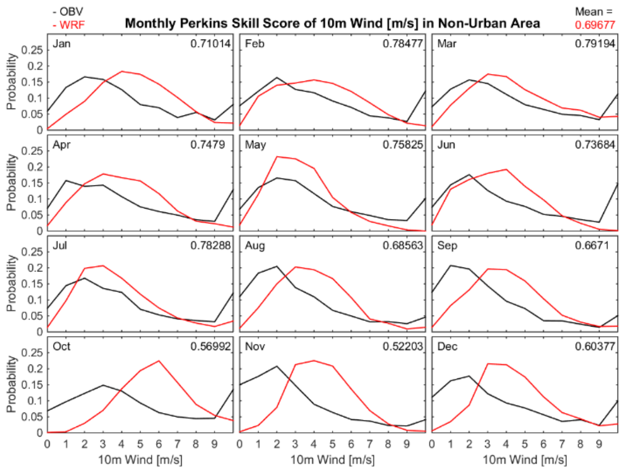


Figure S30: Monthly PSS of 10-Meters Wind Speed in Urban (a) and Non-Urban (b) Area.

(a) (b)

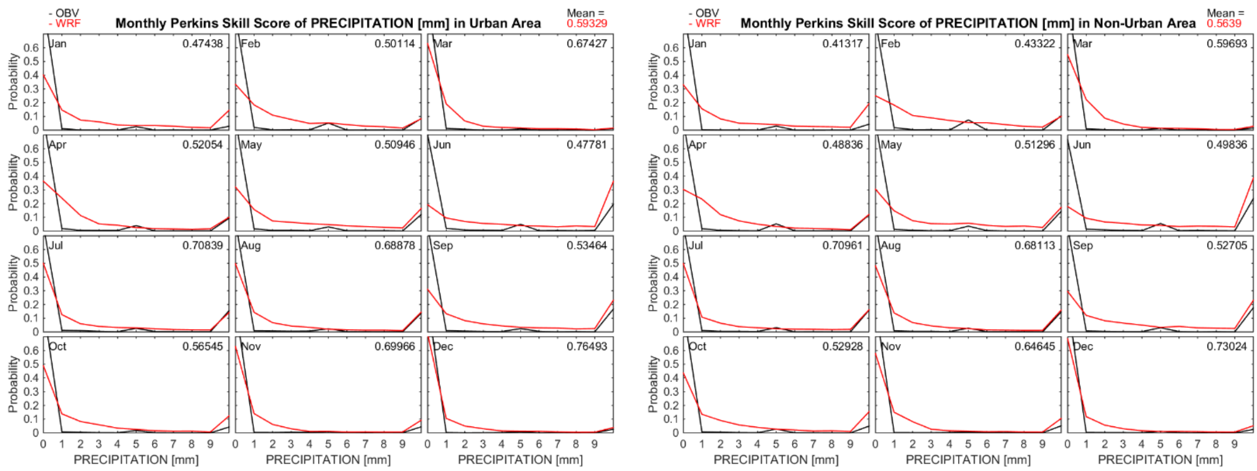


Figure S31: Monthly PSS of Precipitation in Urban (a) and Non-Urban (b) Area.

(a) (b)

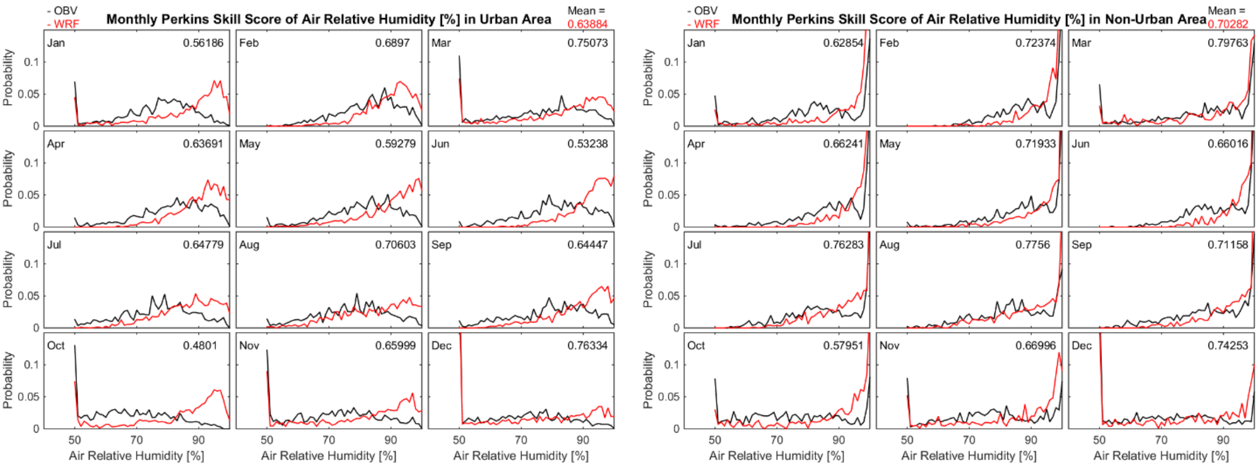


Figure S32: Monthly PSS of Relative Humidity in Urban (a) and Non-Urban (b) Area.

S9 The Capabilities of WRF ARW/Noah LSM/SLUCM in Simulating Meteorological Phenomena

This paper aims to present a standardized methodological framework that can be used in all regions to evaluate the modelling performance rather than to examine the capabilities of the model in simulating meteorological phenomena in a specified region. Therefore, we proposed model evaluation methods for the comparison between modelled variables and its corresponding observed ones rather than investigating the meteorological phenomena in the modelled results. Nevertheless, the capabilities of a model in simulating meteorological phenomena are an essential research direction but beyond the scope of this study. Therefore, we would like to contribute some opinions about the capabilities of WRF ARW/Noah LSM/SLUCM in simulating meteorological phenomena to urban climate modellers for their reference.

Firstly, a reasonable temporal variation of the height of PBL can be seen in the modelled results but it cannot be examined by the observation because of unavailability of its corresponding observed ones. Secondly, the land-sea breeze should be observed in our study area because it is a coastal area. However, we didn't find out this phenomenon in the modelled results due to the temporal resolution of 6 hours wasn't enough fine for supporting the investigation of it (only four modelled variables of 10-meters wind speed at 2:00, 8:00, 14:00 and 20:00 on each day). Thirdly, the annual climatological variation in our study area was associated with the monsoon flow, especially the annual variations of 2-meters air temperature, 10-meters wind speed, and precipitation. Figures 2, 13 and 16 demonstrated that the modelled 2-meters air temperature, 10-meters wind speed, and precipitation had the same annual variation behaviour as its corresponding observed ones, which indicated that the model could simulate these climatological features in a study area affected by the monsoon flows. However, the model cannot reach the extreme value of these variables, especially precipitation. Finally, the spatial distribution of temperature was strongly associated with the local land surface attributes. The model can simulate the temperature difference between in urban and non-urban area.

S10 A Discussion on the Standard for Considering a Given PSS Acceptable

A reliable standard of 'acceptable PSS values' cannot be fully dependent on one single study - it has to be a joint effort over time. This study was intended to make the first step in this effort, and the standard will likely improve as more researchers apply the PSS method to many quantities, time, and spatial scales. Compared to the case studies in Perkins et al. (2007), the lower bounds for PSS in our standard was lower, which is due to the increased resolution in our simulations. We are fully aware that, despite the sophisticated analysis we have conducted on different spatial and temporal scales, it is difficult to define 'acceptable' using results from one case study. In this vein, the proposed standard of 'acceptable' for high-resolution urban climate simulations based on our case study was meant to be the starting point and to be improved by future case studies using the proposed model evaluation framework.

We summarized the 72 monthly-analysis for 6 meteorological attributes in our case study. The PSS values generally followed a normal distribution ranging from 0.444 to 0.886, with an average of 0.660 and a standard deviation of 0.098 (Figure S33). Therefore, the PSS values are larger than 0.500 with a probability of 95%.

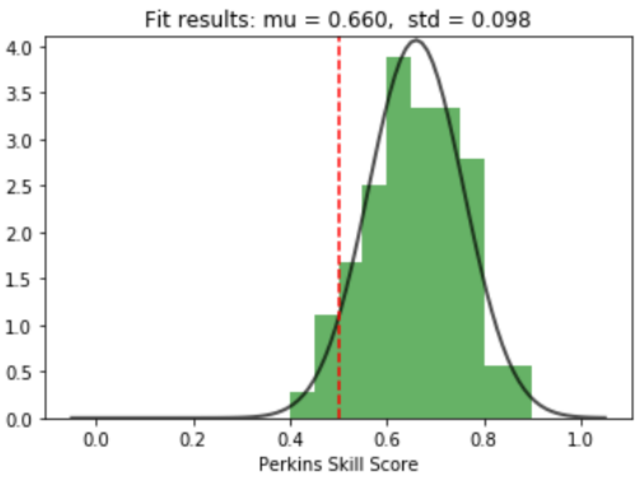


Figure S33: Histogram of the Perkins skill score for 72 monthly PDF analysis. Normal distribution was fit. The red dashed line indicate the lower bound for 95% confidence (PSS=0.500).

We checked the variations of PSS values in our study scope due to different quantities and time of day/year. Figure S34 shows the variations of PSS values due to different quantities of interest. We also checked the statistical significance of the between-group difference using the t-test. P-values among the groups of PSS values for different quantities show that no significant ($p<0.05$) difference was found among T-2, ST2, RH, and W10, while ST14 and Precip had significantly ($p<0.05$) lower PSS values compared to the other attributes but the difference between the average levels of the largest and the smallest group was below 0.2. Therefore, it is possible to have a unified standard of acceptable PSS values while highlighting the standard can be relaxed slightly for specific quantities known to have lower reliability.

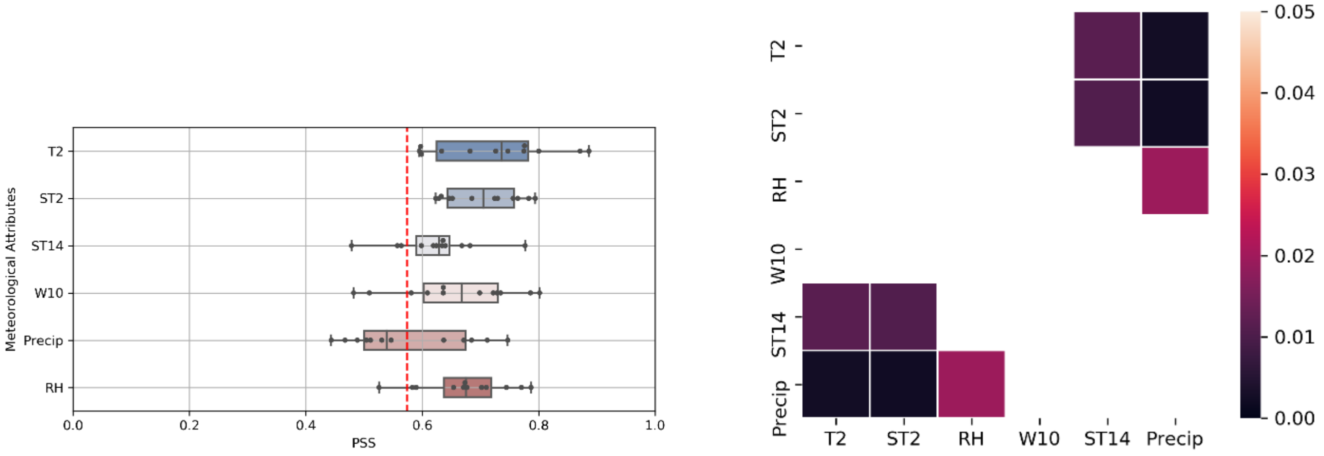


Figure S34: Variations in the PSS values due to different quantities (left) and the t-tests among the PSS values for different quantities (right). The red dashed line indicates the 75% quantile level among all PSS values.

We also checked the variations of PSS values over the time of year. PSS values in all months of the year had mean/median PSS values larger than the 75% threshold we proposed. No statistically significant ($p<0.05$) were observed among any monthly groups of PSS values. It is for future studies to check further how PSS values change over time, for example, ten years ago or later, with significant changes in meteorological contexts.

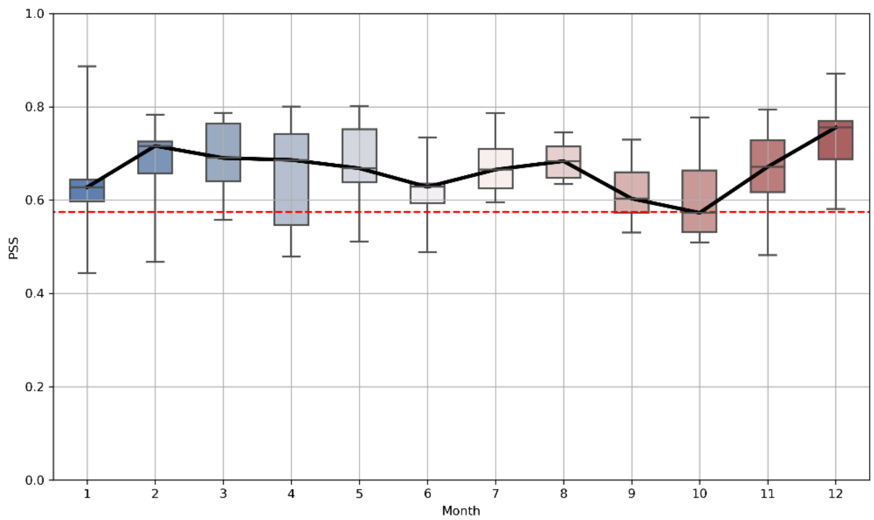


Figure S35: Variation of PSS values in different months of the year.

Since all simulations conducted in this paper use the same spatial extent and scale, we cannot thoroughly check how PSS values vary over different extent and scales. However, it is reasonable to claim that the proposed standard of acceptable PSS values in this study sets the minimum requirement since simulation accuracies were usually found higher for simulations having lower spatial resolutions due to the spatial-smoothing effects. Simulations using coarser spatial resolutions should at least meet our standard of acceptable PSS values, and the standard can be tightened in future studies using coarser spatial resolutions.

S11 The Guideline for Model Evaluation Software Package

We provided the readers a software package for model evaluation, which was developed in Matlab. Figure S36 shows its working folder's structure.

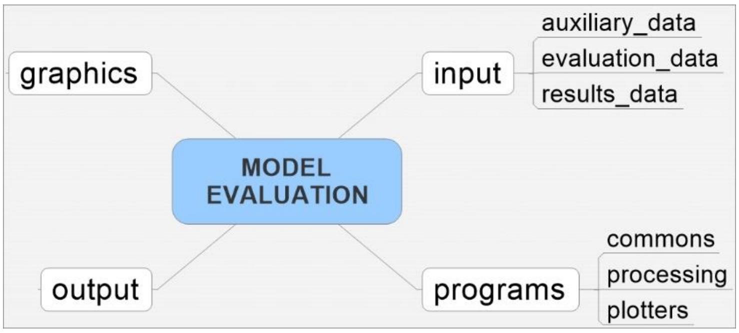


Figure S36: The Working Folder Structure of Model Evaluation Software Package.

The 'MODEL EVALUATION/input' folder is used for the storage of input data which includes auxiliary data (in 'MODEL EVALUATION /input/auxiliary_data' folder), evaluation data (observation data, in 'MODEL EVALUATION /input/evaluation_data' folder) , and modelling result data (WRF modelling variables for evaluation, in 'MODEL EVALUATION /input/result_data' folder). The 'MODEL EVALUATION /programs' folder is used for the storage of the programs which include 3 components: (1) the 'commons' component (in 'MODEL EVALUATION /programs/commons' folder) includes some utility modules called by the other components, (2) the 'processing' components (in 'MODEL EVALUATION /programs/processing' folder) is used for processing the observation and modelling results data, and must run step by step (get_basic_data → get_observation_data → get_modis_data → get_WRF_data), and (3) the 'plotters' component (in 'MODEL EVALUATION /programs/plotters' folder) provides the programs which are used for drawing the evaluation graphics. The 'MODEL EVALUATION/output' folder is used for the storage of output data produced by the 'processing' component. The 'MODEL EVALUATION/graphics' folder is used for the storage of the evaluation graphics produced by the 'plotter' component.