WRF Model Moisture Adjustment Method: A Case Study with Wintertime Cloudy Biases in Xinjiang, China

GUITING SONG,[a] ROBERT HUVA,[a] YU XING,[b] AND XIAOHUI ZHONG[a]

[a] *Envision Digital International Pte. Ltd., Singapore*
[b] *Shenzhen Institute of Artificial Intelligence and Robotics for Society, and Institute of Robotics and Intelligent Manufacturing, The Chinese University of Hong Kong, Shenzhen, China*

(Manuscript received 14 July 2020, in final form 6 November 2020)

ABSTRACT: For most locations on Earth the ability of a numerical weather prediction (NWP) model to accurately simulate surface irradiance relies heavily on the NWP model being able to resolve cloud coverage and thickness. At horizontal resolutions at or below a few kilometers NWP models begin to explicitly resolve convection and the clouds that arise from convective processes. However, even at high resolutions, biases may remain in the model and result in under- or overprediction of surface irradiance. In this study we explore the correction of such systematic biases using a moisture adjustment method in tandem with the Weather Research and Forecasting (WRF) Model for a location in Xinjiang, China. After extensive optimization of the configuration of the WRF Model we show that systematic biases still exist—in particular for wintertime in Xinjiang. We then demonstrate the moisture adjustment method with cloudy days for January 2019. Adjusting the relative humidity by 12% through the vertical led to a root-mean-square error (RMSE) improvement of 57.8% and a 90.5% reduction in bias for surface irradiance.

KEYWORDS: Moisture/moisture budget; Numerical weather prediction/forecasting; Operational forecasting; Clouds; Model errors; Renewable energy

1. Introduction

In 2015 a historic agreement was reached to limit the amount of global temperature increase, due to anthropogenic carbon dioxide emissions, to 2°C beyond preindustrial levels (UNFCCC 2015). However, emissions from the energy sector continue to grow. In their 2019 report the International Energy Agency found that carbon emissions from coal combustion alone were responsible for 0.3°C of the 1°C increase in global average temperature since preindustrial levels (IEA 2019). If the World plans to reach the Paris Agreement targets a vital area for reducing emissions is the energy, and thus electricity, sector. Mature technologies including wind and solar exist and can substitute fossil-fuel-intensive sources of electricity without compromising the supply of electricity. Numerous studies have assessed the ability of whole grids to rely either entirely, or almost entirely, on renewable sources of electricity (Blakers et al. 2017; Budischak et al. 2013; Meneguzzo et al. 2016; others).

In such high-penetration renewable scenarios a key consideration is the ability to forecast abnormally high or low output from the variable electricity sources. Anticipating peaks or troughs from the combination of renewable technologies not only allows grid operators to maintain voltage and frequency standards, it also enables individual operators to maximize their participation in the market. Modeling for a photovoltaic (PV) plant in the Spanish electricity market Antonanzas et al. 2017 showed that greater income was achieved with smaller forecast error. Klingler and Teichtmann 2017 also showed higher forecast error to be an inhibiting factor in the participation of grid-connected PV—even with the use of battery storage to provide a buffer. To reach greater penetrations of renewable electricity across the world there is thus a critical need for research that improves the forecast skill of renewable technologies such as wind and solar.

For solar, in particular, there is a growing body of research showing improvements in forecast accuracy using various methods depending on the forecast horizon and the location of interest (Ahmed et al. 2020). It is generally accepted that the importance of numerical weather prediction (NWP) techniques increases with time horizon and that NWP models are most beneficial beyond about 6 h (Perez et al. 2013). NWP models can assist at longer time horizons and for conditions that are more chaotic due to their physical modeling of the underlying processes in the atmosphere. For a location that is not persistently cloudy it is, for instance, currently beyond the scope of purely statistical or regression-based model to generate cloudy conditions a day in advance—without huge amounts of information regarding the general stability of the atmosphere and any upstream influences. NWP models, which capture the full state of the atmosphere, can simulate such conditions and their role in forecasting day-ahead solar irradiance is critical. It is for this reason that the most commonly used NWP model, the Weather and Research Forecasting (WRF) Model, has recently included solar-specific configurations (Jiménez et al. 2016a).

The region of interest for this study, Xinjiang in China, has one of the best climates for large-scale solar energy production (He and Kammen 2016). As at mid-2019 the combined installed capacity of solar and wind in Xinjiang reached 29.47 GW (China Daily 2019). The region is high elevation, and in the summer

Corresponding author: Robert Huva, robert.huva@envision-digital.com

DOI: 10.1175/WAF-D-20-0117.1

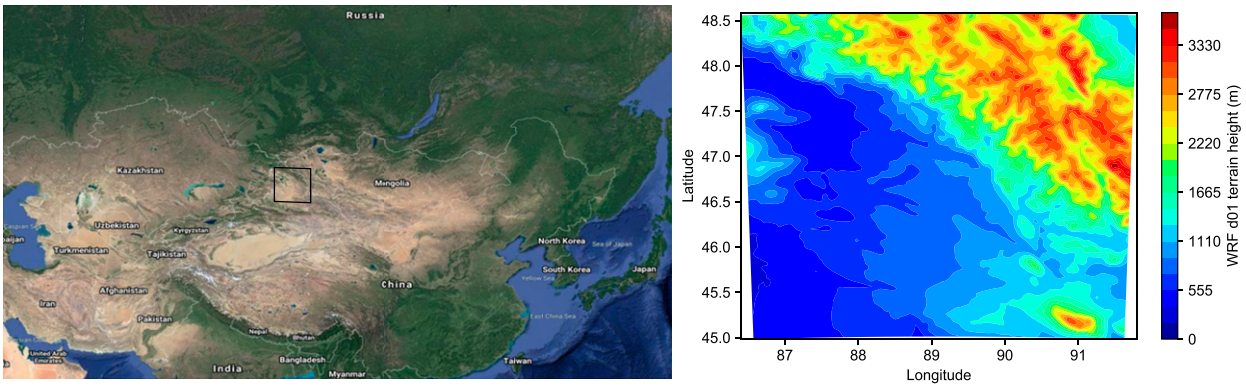

FIG. 1. (left) China area from Google Maps with the approximate WRF Model domain boundary overlaid for illustration purposes. (right) The exact WRF single domain boundary with terrain height above sea level as background.

Xinjiang experiences temperatures exceeding 30°C while in the wintertime temperatures can be as low as −20°C with low precipitation throughout the year. Li et al. (2018) studied the summertime meteorological conditions around Bosten Lake in Xinjiang using WRF. The authors found complex dynamics with diurnal variations relating to the cooler and moister surface of the lake during the daytime, along with flow that was influenced nearby mountains. Cai et al. 2020 also found that the ongoing urbanization of Xinjiang is affecting the land surface energy balance. In particular, the relatively vegetated area north of the Tianshan Mountains has seen significant changes in recent decades, and these changes were found to be altering the background circulations (Cai et al. 2020).

In general, the WRF Model has been shown to overestimate surface solar irradiance (Jiménez et al. 2016a). Indeed Ruiz-Arias et al. (2016) found a consistent overestimation of surface irradiance based on a 10-yr study in Spain. Much of the research surrounding the WRF-Solar advancements has hence focused on the allowing the user to reduce some of the common biases. For clear-sky conditions the inclusion of time varying aerosol input options, as well as the interaction of aerosols with the microphysics schemes have been added. Ruiz-Arias et al. (2012) showed improved WRF clear-sky simulation by adjusting aerosol values. Jiménez et al. (2016b) showed the benefit of using perturbed ensembles to improve the representation of unresolved cloud processes in WRF. In Australia Prasad and Kay (2020) illustrated the difficulty of the WRF-Solar model to appropriately capture location and timing of clouds under partly cloudy conditions, while in Southern California López-Coto et al. (2013) tested 72 different configurations of the WRF Model and found overestimation of irradiance, in general, with a tendency for underprediction of temperature and moisture through the vertical.

To reduce bias, or more broadly error, in a physical manner resulting from NWP simulations there are a few methods to be considered depend on to the source of the bias/error. Data assimilation (DA) is a commonly used tool for adjusting the inputs to NWP models. DA allows users to reduce error in the initial inputs to the model by the minimizing the distance between observations and input conditions. However, the benefits of DA tend to fade over time and would most likely be indistinguishable from a control run at the day-ahead time horizon [differences in surface irradiance can wash out by about 9 h; Huva et al. (2020)]. For clear-sky conditions the use of aerosol optical depth (AOD) values that are based on observation has been shown to improve global horizontal irradiance (GHI) biases (Ruiz-Arias et al. 2012); however, the role of aerosols in modulation the irradiance time series is most

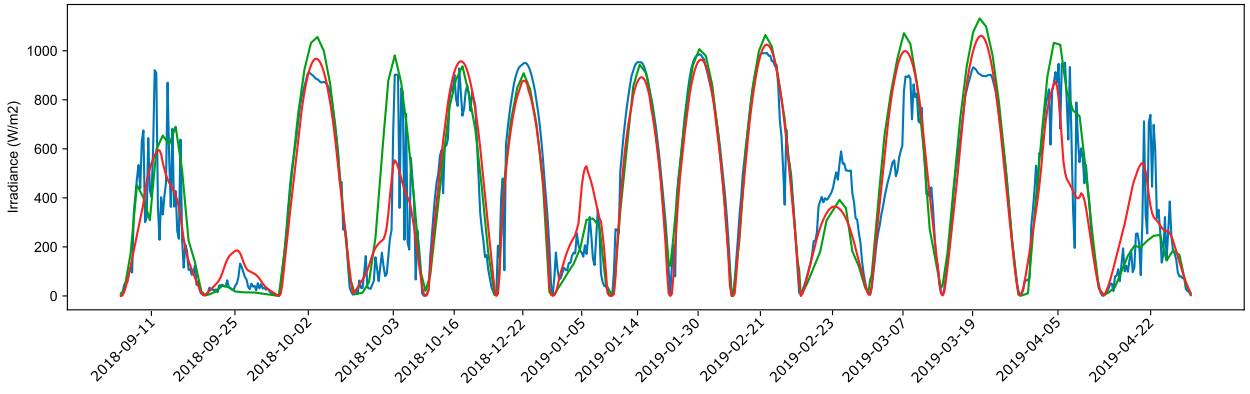

FIG. 2. Time series plot showing the best overall WRF configuration (green) compared to the observations (blue) and the EC model GHI (red). Nighttime values are not plotted for brevity and to enhance plot readability.

influential on the direct beam and for cloudy/overcast days where the direct beam approaches zero other mechanisms for bias correction need to be considered. Coleman et al. (2010) considered surface moisture biases in simulations of Southern California. By adding in missing anthropogenic moisture sources to the WRF inputs prior to running the model the authors were able to show improved representation of near-surface daytime moisture and temperature biases for some locations. Similarly, Byun et al. (2011) analyzed the impact of nudging soil moisture values with the Pleim and Xiu land surface model (Xiu and Pleim 2001). The authors also found reductions in 2-m temperature biases when nudging to observed soil moisture amounts when compared to the control simulation without nudging. Meanwhile, Yang and Kleissl (2016) analyzed the effect of initializing the WRF Model with cloud information from satellite observations. They found reductions in error and bias lasting up to 24 h for the simulations that were initialized with cloud information.

In this study we show that systematic biases beyond 24 h can be physically reduced by adjusting the moisture fields used to guide the WRF Model. We first optimize the WRF Model configuration for day-ahead solar irradiance forecasting by systematically testing combinations of the radiative transfer model (RTM) with microphysics and cumulus schemes for a location in Xinjiang, China. Following this we demonstrate that significant underprediction of irradiance is still present for cloudy days in the wintertime. We then demonstrate a method to alter the 3D moisture fields used to guide the WRF Model for January 2019 and show an almost complete reduction in bias. The rest of the paper is structured as follows: section 2 describes the observations and methods used, section 3 presents the results of the study (both in terms of optimizing the WRF Model configuration and the time period with the adjustment method applied), while section 4 concludes the work with suggestions on future work that could stem from the current study.

2. Experimental methods

a. Observations

This study utilizes various dates from the years 2018 and 2019 to optimize the WRF Model configuration and then test the performance of selected physics configurations. The observations are for a location in Xinjiang, China, which measures GHI for a solar farm. The observations are at 15-min frequency and have been filtered for erroneous data through a system of checks that examine various known errors including nonzero nighttime values, shading of the sensor leading to unrealistic daily maximum values, daytime values that are unchanging, daytime values of zero, and days with mostly missing values.

b. Numerical weather prediction

The numerical model used in this study was the WRF Model (Skamarock and Klemp 2008). We utilized version 3.9.1 and set up a limited area domain centered on the location of interest (Fig. 1). Initial and boundary conditions were provided by the European Centre for Medium-Range Weather Forecasts

TABLE 1. List of dates used in the physics optimization, their associated k_t value and the cutoff k_t value for each of the 15 bins. Dates are in local time.

Bin No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Maximum k_t	0.25	0.40	0.50	0.55	0.65	0.74	0.81	0.84	0.89	0.91	0.93	0.97	1.0	1.03	1.11
Date chosen	25 Sep 2018	5 Jan 2019	22 Apr 2019	3 Oct 2018	23 Feb 2019	11 Sep 2018	7 Mar 2019	5 Apr 2019	16 Oct 2018	19 Mar 2018	21 Feb 2019	2 Oct 2018	30 Jan 2019	14 Jan 2019	22 Dec 2018
Date k_t	0.06	0.22	0.32	0.40	0.49	0.59	0.71	0.80	0.85	0.92	0.95	0.99	1.04	1.07	1.11

TABLE 2. List of WRF schemes used in the physics optimization.

Role of scheme	WRF option	Scheme name
Shortwave RTM	1	Dudhia scheme (Dudhia 1989)
	3	CAM scheme (Collins et al. 2004)
	4	RRTMG shortwave (Iacono et al. 2008)
	7	Fu–Liou–Gu scheme (Gu et al. 2011)
Microphysics	2	Purdue Lin scheme (Chen et al. 2002)
	14	WRF double-moment 5-class scheme (Lim and Hong 2010)
	16	WRF double-moment 6-class scheme (Lim and Hong 2010)
	22	NSSL two-moment scheme (Mansell et al. 2010)
	28	Thompson aerosol-aware (Thompson and Eidhammer 2014)
Cumulus parameterization	0	No cumulus scheme
	0, shcu_physics = 2	Park and Bretherton shallow cumulus (Park and Bretherton 2009)
	0, shcu_physics = 3	GRIMS shallow cumulus (Hong and Jang 2018)
	1	Kain–Fritsch scheme (Kain and Kain 2004)
	3	Grell–Freitas (GF) scheme (Grell and Freitas 2014)
	5	Grell 3D (Grell and Dévényi 2002)
	10	Kain–Fritsch–cumulus potential scheme (Berg et al. 2013)
	11	Multiscale Kain–Fritsch scheme (Glotfelty et al. 2019)

(ECMWF, hereafter referred to as the EC model) at a 3-hourly interval and 0.1° resolution. We utilized the 1200 UTC EC model release and forecasted for 52 h where the 28–52-h forecast horizon represents the day of interest at midnight–midnight local time (+8 h from UTC). The WRF Model domain was set at 3-km resolution with 65 vertical levels.

c. *WRF physics optimization*

Previous research by the authors (Huva et al. 2021, manuscript submitted to *Meteor. Appl.*) has shown that the most important physical schemes to optimize for day-ahead surface irradiance forecasting involve the RTM, microphysics and cumulus physics schemes. It was also shown that the optimum physics configuration varied depending on the type of day, where the biggest differences were found on cloudy days (Huva et al. 2021, manuscript submitted to *Meteor. Appl.*). We explored this hypothesis for the Xinjiang site and conducted physics optimization using a set of 15 days made up of five sunny days, five partly cloudy days and five cloudy days from across various seasons (see the blue curve in Fig. 2).

The days selected were based on cloud amount using the k_t formula (section 2e) and with the intention of maximizing coverage from across the seasons to reduce the influence of seasonality, which could impact configuration choice but was not explored here. We sampled the full distribution of daily k_t values from March 2018 to July 2019 into 15 bins with equal samples in each bin. Based on a visual analysis of dates near the bin boundaries (not shown), we then used the first four bins to represent cloudy conditions, the next five bins to represent partly cloudy conditions and the final six bins to represent sunny conditions. The chosen days, their k_t value, and the k_t bins are outlined in Table 1. We ran day-ahead simulations for each date in the optimization set and then analyzed the results across all days, as well as by day type. The list of physics options tested are outlined in Table 2. In total 120 combinations successfully completed all 15 days of simulation. Configurations that were unstable and not able to complete all days in the optimization set were not analyzed. The best configuration was determined as the configuration with lowest RMSE—either for all days (overall) or by day type. A combined model was also tested. The combined model was a concatenation/splicing of the best configurations for each day type.

Regarding the RTMs, the list in Table 2 contains the shortwave schemes. The longwave version of each scheme was also used, with the exception of the Dudhia scheme, which has no longwave version and where the RRTMG longwave scheme was used instead. For the Dudhia and RRTMG-based RTM models there is an option to include cloud effects on optical depth (icloud in WRF), which was turned on. With respect to the cumulus schemes it should be noted that preference was given to schemes, or versions of schemes that have some scale awareness. There are also extra options that involve the interaction of the cumulus schemes and the RTM that were utilized in this study. They include the following: feedback from shallow cumulus to the RTM (shallow_cu_fored_ra in WRF), which is available for Kain–Fritsch-based cumulus schemes and was used with the default binning (21 bins) of potential temperature and mixing ratio, and subgrid-scale cloud interaction (cu_rad_feedback in WRF), which is available and was turned for Grell and Kain–Fritsch cumulus schemes.

d. *Moisture adjustment method*

Where the optimized configuration of the WRF Model still exhibits bias and the bias is systematic we propose a method for

TABLE 3. Performance of the combined model (best physics for each day type concatenated), the best overall model, and the EC model for the 15-day optimization set.

Model	RMSE ($W\,m^{-2}$)	Bias ($W\,m^{-2}$)
Combined WRF	134.23	2.42
Overall WRF	144.94	21.15
EC model	130.69	2.33

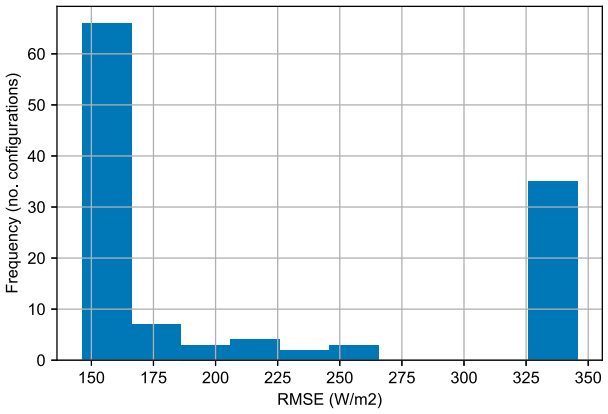

FIG. 3. Histogram of performance (RMSE) for all configurations and all 15 test days.

addressing the bias by way of adjusting the 3D moisture field within WRF. Moisture in the form vapor, cloud condensation nuclei, or other forms influences the surface shortwave irradiance by virtue of blocking/deflecting or absorbing and re-emitting incoming shortwave radiation from the sun. Under normal circumstances the WRF Model is used in such a way that the limited area domain is initialized by global/parent model and the only extra information at following time steps comes in the form of boundary conditions from the parent model. However, functionality exists whereby the WRF Model can be nudged toward the parent model values (grid analysis nudging, called grid fdda in WRF). We utilized this option in WRF. By using reasonable nudging coefficients, and where necessary by adjusting the parent moisture values, we show that surface irradiance biases can be effectively reduced.

The EC model conditions used to drive the WRF Model in this study come in GRIB data format and at 3-hourly intervals. From the GRIB data, the WPS module unpacks the data and creates a series of intermediate files from FORTRAN write statements. The intermediate files are then ingested by the metgrid program where the data are spatially interpolated onto the WRF Model grid. We propose a method to overwrite water vapor mixing ratio values contained in the indeterminate files for circumstances where the optimized WRF Model is systematically biased in surface irradiance. The method uses FORTRAN code to read in the water vapor mixing ratio Q_{vap}, pressure and temperature values on the EC model grid, calculate relative humidity (RH), adjust the RH values, and then write out altered water vapor mixing ratios to a new file. RH was utilized instead of directly altering the Q_{vap} due to RH being bounded from 0 to 100, which leads to a lower susceptibility for extreme alteration. RH is calculated at each grid point and EC model level as follows:

$$\text{RH}\,(\%) = 100.0 \times \frac{P_{\text{vap}}}{P_{\text{sat.vap}}}, \tag{1}$$

where the vapor pressure P_{vap} is

$$P\,(\text{Pa})_{\text{vap}} = \frac{Q_{\text{vap}} \times P}{\left(\frac{R_{\text{gas}}}{R_{\text{vap}}} + Q_{\text{vap}}\right)}. \tag{2}$$

The term R_{gas} is the gas constant for air (287 J kg^{-1} K^{-1}), R_{vap} is the specific gas constant of water vapor (461.6 J kg^{-1} K^{-1}), and $P_{\text{sat.vap}}$ is the saturation vapor pressure calculated following the data on page 113 of Curry and Webster (1999) as follows:

$$P\,(\text{Pa})_{\text{sat.vap}} = 100.0 \times 6.11 \times \exp\left[\frac{L}{R_{\text{vap}} \times \left(\frac{1}{273.15} - \frac{1}{T}\right)}\right] \in T < 223.15\text{K}, \tag{3}$$

$$P\,(\text{Pa})_{\text{sat.vap}} = 100.0 \times \left[a(1) + \sum_{i=2}^{i=7} a(i) \times (T - 273.15)^{i-1}\right] \in T \geq 223.15\text{K}, \tag{4}$$

where L is the latent heat of vaporization (2.5×10^6 J K^{-1}), and a = (6.111 767 5, 0.443 986 062, 0.143 053 301 $\times$ 10^{-1}, 0.265 027 242 $\times$ 10^{-3}, 0.302 246 994 $\times$ 10^{-5}, 0.203 886 313 $\times$ 10^{-7}, 0.638 780 966 $\times$ 10^{-10}) with indexing starting from 1.

The RH of each EC model level and at each location within the WRF Model domain is then altered by the desired amount, Eqs. (1)–(4) applied in reverse to calculate the new Q_{vap} and this value then written to a separate intermediate file in the same format as the other intermediate files. The intermediate file with the altered Q_{vap} values is then preferentially utilized by the metgrid program by listing the file prefix last in the namelist. Following these procedures allows the user to alter the Q_{vap} values in the metgrid output. After running metgrid the user is required to run the real program in WRF to provide the input and boundary conditions on the WRF vertical levels. The real program is also responsible for creating the nudging files when grid nudging is turned on. Based on the values

TABLE 4. Important details for the options that produced the lowest RMSE for all days (overall) and for each day type separately from the 15-day optimization set.

Scheme	mp_ phsyics	cu_ physics	cudt	shcu_ physics	ra_lw_ physics	ra_sw_ physics	icloud	cu_rad_ feedback	shallow_ forced_ra	nbins	cu_ diag	ishallow
Overall option	16	0	0	2	4	4	3	False	False	—	0	0
Sunny option	22	1	5	0	4	4	3	False	False	—	0	0
Partly cloudy option	16	0	0	2	4	4	3	False	False	—	0	0
Cloudy option	2	0	0	0	4	1	3	False	False	—	0	0

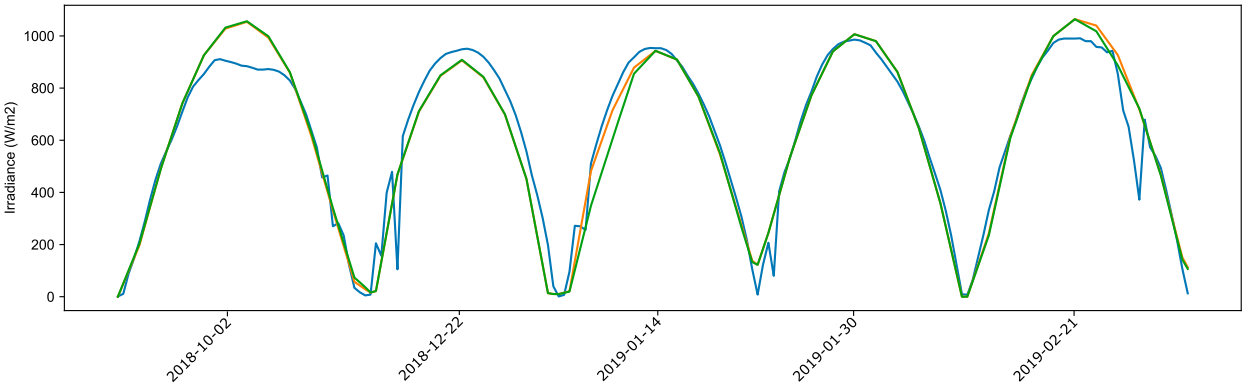

FIG. 4. Comparison of best overall configuration (green) with the best configuration based on RMSE for sunny days only (orange) and with observations plotted in blue. As in Fig. 2, the nighttime values are not plotted for brevity.

preset by the user the WRF Model can then be run and nudged toward the values from metgrid for the wind components U and V, as well as T and Q_{vap}. The strength of the nudging, the WRF Model levels on which to nudge and the duration of the nudging are all controllable by the user through the namelist. We utilized nudging for all of U, V, T and Q_{vap} and for all levels above model level 6. The nudging coefficients for U, V and T were set to 0.0002, while for Q_{vap} it was set at 0.000 12.

e. Sky condition characterization

When selecting the dates included in the physics optimization set, the sky cover amount k_t was used. Sky cover amount is the ratio between the observed irradiance and theoretical clear-sky irradiance for a given time. Values closer to 1 indicate sunny conditions and the values that approach 0 indicate cloudy condition. The sky cover amount k_t is calculated as follows:

$$k_t(t) = \frac{I_{\text{observed}}(t)}{I_{\text{clearsky}}(t)}, \tag{5}$$

where I_{observed} is the observed or modeled irradiance for the Xinjiang site, and I_{clearsky} is the theoretical clear sky irradiance. Modeled irradiance is used in the numerator of Eq. (5) to calculate the EC model cloudiness. The EC model k_t is used to identify the type of day ahead of time (see section 3c for its use). Ideally, a clear sky model that is optimized for the location of interest, based on observed clear sky conditions is utilized. We, however, did not have access to such observations for the Xinjiang site and instead use the Python programming language module, pvlib, and the default Ineichen/Perez model for this study.

f. Measuring forecast accuracy

In this study we utilize the root-mean-square error (RMSE) and bias statistics to measure performance [Eqs. (6) and (7)]. The WRF SWDOWN (shortwave GHI) or the EC model SSRD (surface solar radiation downward) values are compared against the observed GHI at the frequency of the observations (15 min):

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum\left(I_{\text{pred}} - I_{\text{obs}}\right)^2}, \tag{6}$$

$$\text{bias} = \frac{1}{N}\sum\left(I_{\text{pred}} - I_{\text{obs}}\right), \tag{7}$$

where I_{pred} is the WRF/EC predicted irradiance (GHI) and I_{obs} is the measured GHI at the Xinjiang site. For both calculations nighttime values were not included. Nighttime was determined by when the observations were equal to zero. Irradiance values from the WRF Model are derived by the RTM. We used a 3-min

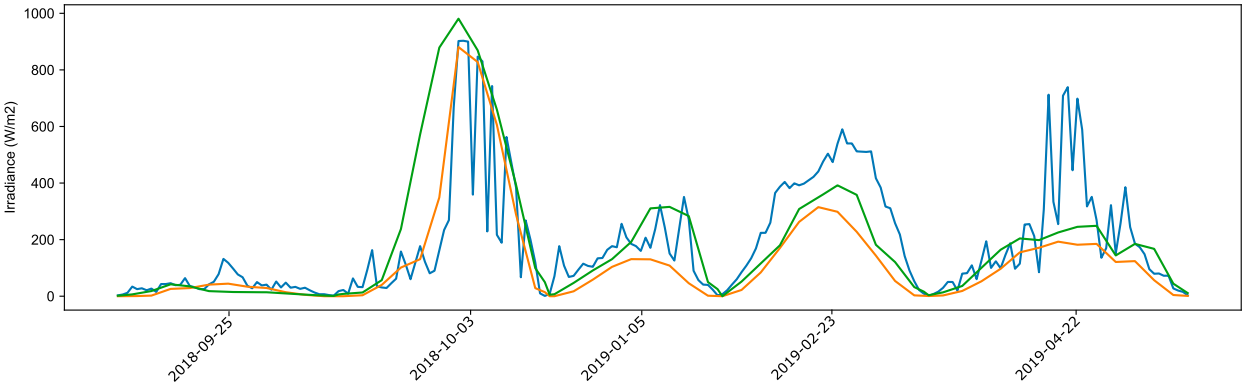

FIG. 5. As in Fig. 4, but for cloudy days from the optimization set.

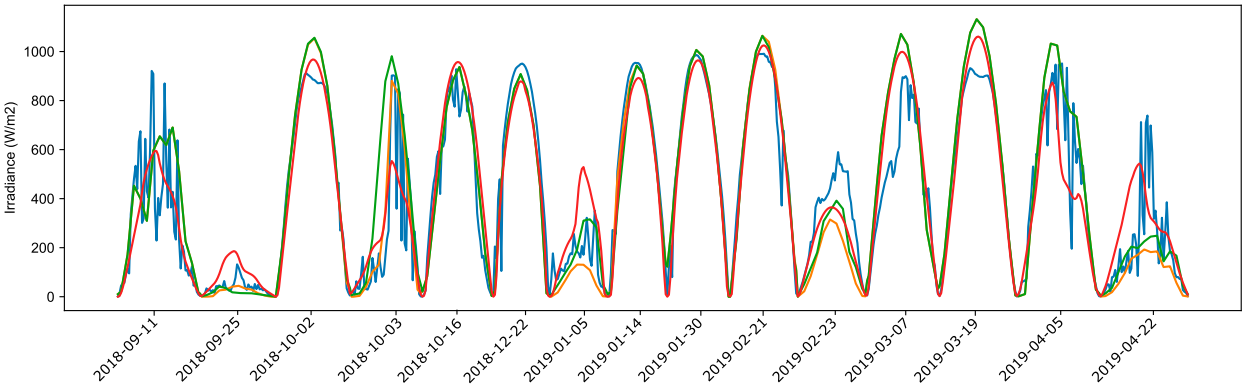

FIG. 6. Time series comparison of the combined model (orange) with the best overall model (green), the EC model (red), and with observations plotted in blue for the optimization set. As per previous plots, nighttime values are not plotted for brevity.

frequency for calling the RTM but then extract only the 15-min instantaneous values starting on the hour. To achieve 15-min frequency EC model GHI values we first calculated average values ($\mathrm{W\,m^{-2}}$) from each 3-h window. This was done by dividing the raw EC values, which are 3-hourly accumulated from model initialization (in units of $\mathrm{J\,m^{-2}}$), by the number of seconds in 3 h. This calculation was also done for 3-hourly accumulated clear-sky model values. EC model average k_t values were then calculated following Eq. (5) and a linear interpolation performed in k_t space to achieve 15-min frequency k_t values. Finally, 15-min frequency EC model values were extracted by multiplying the 15-min frequency k_t values by the 15-min frequency clear-sky model values.

3. Results

The following result sections first presents results from optimizing the WRF Model physics configuration for the site in Xinjiang, China. The WRF Model configuration is optimized for all types of days, as well for each type of day separately. Following this, an analysis of systematic biases in performance is conducted. The results of the RH adjustment method are then presented for forested cloudy days in January 2019.

a. RTM, microphysics, and cumulus optimization

In this section we show the results of optimizing the RTM, microphysics parameterization and cumulus scheme used in WRF. Previous work by the authors has shown that other options (including planetary boundary layer, surface layer physics, and land surface model) had almost no effect on the day-ahead forecast of irradiance for Qinghai in China (Huva et al. 2021, manuscript submitted to *Meteor. Appl.*) and were thus not considered here. The best overall configuration (the configuration with lowest RMSE for all day types, outlined in Table 3) is plotted in Fig. 2, while the histogram of RMSE for the various configurations is shown in Fig. 3.

From the histogram in Fig. 3 we can see that there was a cluster of similarly performing near-optimum configurations at approximately 150 $\mathrm{W\,m^{-2}}$ RMSE. Beyond this cluster the was set of configurations greater than 325 $\mathrm{W\,m^{-2}}$, which all involved the CAM RTM model.

b. Optimization based on type of day

In this section explore the hypothesis that the optimum configuration of the WRF Model has some dependency based on the type of day being forecasted. When the analysis of RMSE was split based on the type of day (sunny, partly cloudy and cloudy) the best performing configuration showed some changes (Table 4), and the overall RMSE was reduced when the configuration was tailored (Table 3). A combined model was also created. The combined model was created by concatenating the day-type configurations (5 days from each day type). In this way the combined model was a best-case scenario if the type of day were known in advance and the WRF Model configuration chosen according to the best performing for that day type.

Figure 4 outlines the time series of sunny days from the optimization set. Figure 4 compares the best sunny configuration with the best overall configuration. As can be seen from Fig. 4, the improvement by using a physics combination tailored to sunny days was marginal and for the five sunny days the difference in RMSE was 5.9% (83.8 compared to 79.9 $\mathrm{W\,m^{-2}}$).

For partly cloudy the best configuration was the same as the best overall configuration (Table 4). However, for cloudy days the difference between the tailored and best overall configurations became more obvious (Fig. 5). For cloudy days there

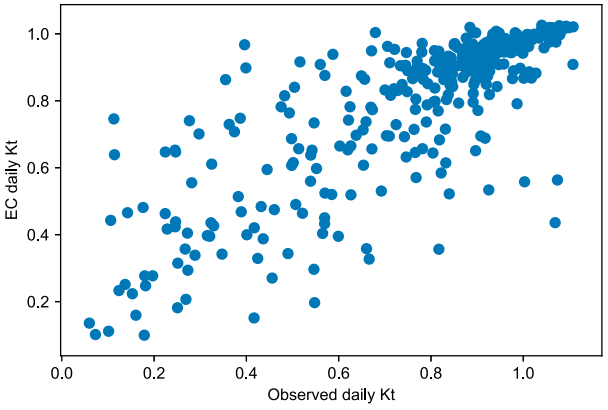

FIG. 7. Scatterplot comparison of EC model derived daily k_t values and the observed daily k_t values for the May 2018–July 2019 period.

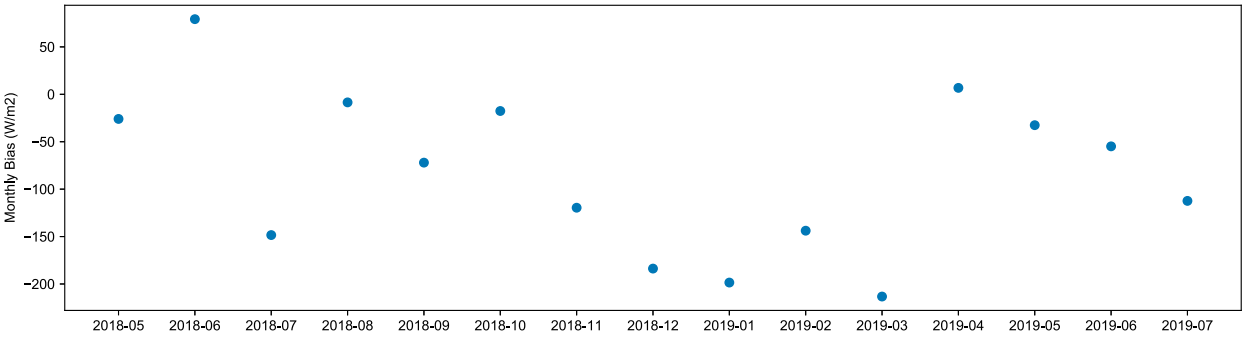

FIG. 8. Monthly averaged bias for the cloudy configuration on EC forecasted cloudy days in the testing set.

was a 15% improvement in RMSE when the best cloudy configuration was selected instead of the best overall configuration (171.5 compared to 145.8 $W m^{-2}$ for the cloudy configuration), with most of the improvement coming on the second day.

Comparing the configurations for the day types with the best overall (Table 4) it may be a surprising result that a cumulus scheme was preferred for sunny days and not for cloudy days. However, for sunny days the top 80 configurations were separated by just 3.8 $W m^{-2}$ RMSE and not all the configurations utilize cumulus physics, suggesting little dependence on this option for sunny days (not shown). For cloudy days there was a much larger spread in performance. By comparison the top 80 cloudy configurations were separated by 153 $W m^{-2}$ RMSE and the top 10 configurations were separated by 4.4 $W m^{-2}$ RMSE with 8 of the top 10 utilizing some form of cumulus physics (either shallow or not) (not shown). Comparing the combined model with the best overall configuration also saw some improvement, mostly represented by the cloudy days (Fig. 6). RMSE was reduced by 7.4% for the combined model when compared to the best overall model, with the combined model performing approximately as well as the EC model (Table 3).

c. *Cloudy day biases*

The cloudy days in the optimization set showed the best performance increase when the physics was tailored for that type of day. In this section we explore the performance of this configuration over a longer time period. We analyzed cloudy days from May 2018 to July 2019 (hereafter referred to as the testing set) and where cloudy days were defined using the EC model forecast. The EC model was used for operational considerations. In reality the type of day being forecasted is not known a priori. The correlation between the type of day being forecast by EC and the observed day during this period was quite high (0.81 based on daily averaged k_t values) and the scatterplot of daily averaged k_t shows good relationship (Fig. 7). Figure 7 suggests that little error is introduced when using the EC model to forecast the type of day.

To classify the days into cloudy, partly cloudy and sunny a similar process to the physics optimization set was used. The distribution of EC daily average k_t values were split into 15 equal weighted bins with increasing k_t value. As is maybe evident from Fig. 7 the EC model had a sunny bias for the testing set. We thus had to also adjust the number of bins in the cloudy versus other categories. The first three bins for EC were used to defined cloudy with a cutoff daily averaged k_t value of 0.6382. The cloudy configuration was run for all EC model forecasted cloudy days for the testing set and the monthly averaged bias is shown in Fig. 8.

From Fig. 8 we can see that forecasted cloudy days, even for the optimized cloudy configuration, exhibit very large negative biases during the wintertime (too cloudy). Months other than

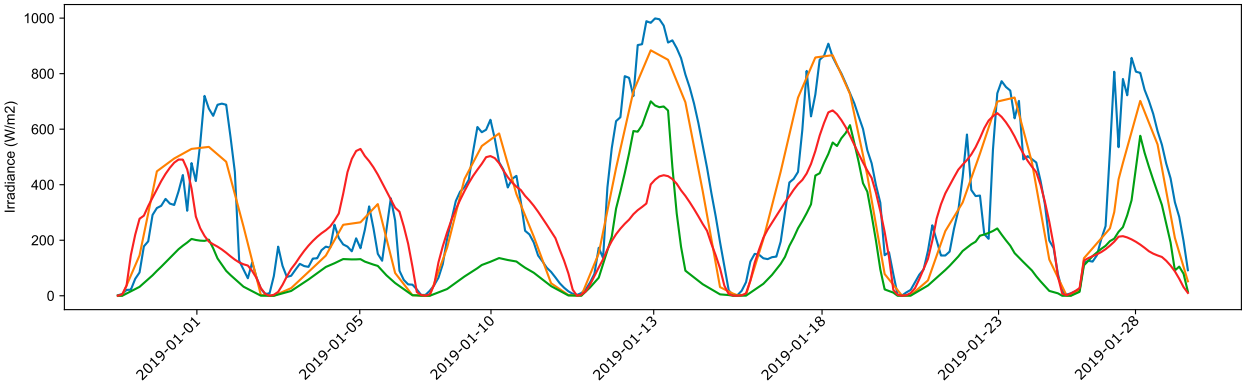

FIG. 9. EC model forecasted cloudy days with the optimized cloudy WRF configuration in green, the −12% RH adjusted WRF output in orange, the EC model in red, and the observations in blue.

TABLE 5. Statistics for January 2019 EC forecasted cloudy days where cloudy WRF represents the configuration optimized for cloudy days and the adjusted WRF the version with −12% RH adjusted nudging inputs.

Model	RMSE ($W m^{-2}$)	Bias ($W m^{-2}$)
Adjusted WRF	111.18	−18.94
Cloudy WRF	263.37	−198.5
EC model	239.49	−76.16

November through March also had a negative bias tendency but with much smaller magnitudes. With these results in mind we thus explored a method for addressing the wintertime bias, using January 2019 as an example.

d. *Bias reduction using moisture adjustment*

To physically address the bias, rather than statistically, we utilized the moisture adjustment method outlined in section 2d. We used January 2019, which was one of the worst performing months, as the example. The 3D moisture field from the EC model was adjusted by −12% in terms of RH, which we found in combination with the Q_{vap} nudging coefficient at 0.0012 worked well for January. However, in our experience the nudging coefficient can be made stronger (weaker) with a smaller (larger) RH % to achieve a similar outcome (not shown). The nudging coeffects were applied through the 54-h simulation and the Q_{vap} values from EC adjusted in the 1500–1500 UTC time period corresponding to forecast hours from +27 to +51. Adjusting through this time period effectively covered the midnight to midnight local time for the day of concern. It should also be noted that only the values from the EC model grid corresponding to the WRF Model grid were altered as it was deemed unnecessary to adjust the whole EC model domain—especially for points in space that would ultimately not be written to the metgrid output files. Figure 9 shows the time series of the optimized cloudy configuration with and without moisture adjustment, as well as the EC model and observed irradiance.

Table 5 also outlines the statistics for January 2019. Clearly, the combination of grid nudging and the adjustment of the moisture values being nudged toward had a profoundly positive effect on the irradiance profile for the Xinjiang site. The bias was almost completely removed (90.5% reduction) and despite the underestimation by the original WRF Model varying from day to day the adjustment of −12% RH appeared suitable for all days. In accordance with the reduction in bias the RMSE also improved by 57.8%.

4. Summary and conclusions

In this study we have investigated the performance of the WRF Model for day-ahead solar irradiance forecasting in the Xinjiang area of China. We utilized the EC model as input to the WRF Model and then conducted an optimization of WRF Model configuration, systematically testing combinations of RTM, microphysics and cumulus parameterization schemes. We showed improvements in performance when the WRF Model configuration was adjusted based on the type of day—most notably for cloudy days, which exhibited a 15% reduction in RMSE. We then explored the performance of the cloudy configuration over a 14-month period and showed systematic underprediction of surface irradiance for the wintertime. To explore the correction of this bias we proposed a method to adjust the moisture in the WRF Model by means of nudging the model toward adjusted EC model moisture fields. Using January 2019 as an example month we showed that by drying the EC model by 12% in terms of RH and nudging the WRF Model toward these new values produced a 90.5% reduction in bias and a 57.8% reduction in RMSE. It was also shown that the constant 12% drying across the cloudy days of January 2019 was a suitable value for most days despite varying amounts of underprediction across the month. Further work should analyze the possibility of tailoring the moisture adjustment for other locations and other months where there exists a systematic bias from WRF.

Acknowledgments. The authors acknowledge Envision Group Pte. Ltd. for funding the research, the European Centre for Medium-Range Weather Forecasts for the data used to drive the WRF Model, and Tianhe-2 supercomputer where the simulations were run.

Data availability statement. Where reasonable, the authors will provide data utilized in this study.

REFERENCES

Ahmed, R., V. Sreeram, Y. Mishra, and M. D. Arif, 2020: A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable Sustain. Energy Rev.*, **124**, 109792, https://doi.org/10.1016/j.rser.2020.109792.

Antonanzas, J., D. Pozo-Vázquez, L. A. Fernandez-Jimenez, and F. J. Martinez-de-Pison, 2017: The value of day-ahead forecasting for photovoltaics in the Spanish electricity market. *Sol. Energy*, **158**, 140–146, https://doi.org/10.1016/j.solener.2017.09.043.

Berg, L. K., W. I. Gustafson, E. I. Kassianov, and L. Deng, 2013: Evaluation of a modified scheme for shallow convection: Implementation of CuP and case studies. *Mon. Wea. Rev.*, **141**, 134–147, https://doi.org/10.1175/MWR-D-12-00136.1.

Blakers, A., B. Lu, and M. Stocks, 2017: 100% renewable electricity in Australia. *Energy*, **133**, 471–482, https://doi.org/10.1016/j.energy.2017.05.168.

Budischak, C., D. Sewell, H. Thomson, L. Mach, D. E. Veron, and W. Kempton, 2013: Cost-minimized combinations of wind power, solar power and electrochemical storage, powering the grid up to 99.9% of the time. *J. Power Sources*, **225**, 60–74, https://doi.org/10.1016/j.jpowsour.2012.09.054.

Byun, D., H.-C. Kim, and F. Ngan, 2011: Improvement of meteorological modeling by accurate prediction of soil moisture in the Weather Research and Forecasting (WRF) Model. NOAA, 46 pp.

Cai, P., and Coauthors, 2020: Numerical study of the interaction between oasis and urban areas within an arid mountains-desert system in Xinjiang, China. *Atmosphere*, **11**, 85, https://doi.org/10.3390/atmos11010085.

Chen, J., B. E. Carlson, and A. D. Del Genio, 2002: Evidence for strengthening of the tropical general circulation in the

1990s. *Science*, **295**, 838–841, https://doi.org/10.1126/science.1065835.

China Daily, 2019: Xinjiang sees steady growth in new energy power generation. *China Daily*, accessed 5 March 2021, http://www.chinadaily.com.cn/a/201907/29/WS5d3e6711a310d8305640181f.html.

Coleman, R. F., J. F. Drake, M. D. McAtee, and L. O. Belsma, 2010: Anthropogenic moisture effects on WRF summertime surface temperature and mixing ratio forecast skill in Southern California. *Wea. Forecasting*, **25**, 1522–1535, https://doi.org/10.1175/2010WAF2222384.1.

Collins, W. D., and Coauthors, 2004: Description of the NCAR Community Atmosphere Model (CAM 3.0). NCAR/TN-464+STR, 226 pp.

Curry, J. R., and P. J. Webster, 1999: *Thermodynamics of Atmospheres and Oceans*. Academic Press, 471 pp.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *J. Atmos. Sci.*, **46**, 3077–3107, https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2.

Glotfelty, T., K. Alapaty, J. He, P. Hawbecker, X. Song, and G. Zhang, 2019: The Weather Research and Forecasting Model with Aerosol–Cloud Interactions (WRF-ACI): Development, evaluation, and initial application. *Mon. Wea. Rev.*, **147**, 1491–1511, https://doi.org/10.1175/MWR-D-18-0267.1.

Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 10–13, https://doi.org/10.1029/2002GL015311.

——, and S. R. Freitas, 2014: A scale and aerosol aware stochastic convective parameterization for weather and air quality modeling. *Atmos. Chem. Phys.*, **14**, 5233–5250, https://doi.org/10.5194/acpd-13-23845-2013.

Gu, Y., K. N. Liou, S. C. Ou, and R. Fovell, 2011: Cirrus cloud simulations using WRF with improved radiation parameterization and increased vertical resolution. *J. Geophys. Res.*, **116**, D06119, https://doi.org/10.1029/2010JD014574.

He, G., and D. M. Kammen, 2016: Where, when and how much solar is available? A provincial-scale solar resource assessment for China. *Renewable Energy*, **85**, 74–82, https://doi.org/10.1016/j.renene.2015.06.027.

Hong, S. Y., and J. Jang, 2018: Impacts of shallow convection processes on a simulated boreal summer climatology in a global atmospheric model. *Asia-Pac. J. Atmos. Sci.*, **54**, 361–370, https://doi.org/10.1007/s13143-018-0013-3.

Huva, R., H. Verbois, and W. Walsh, 2020: Comparisons of next-day solar forecasting for Singapore using 3DVAR and 4DVAR data assimilation approaches with the WRF model. *Renewable Energy*, **147**, 663–671, https://doi.org/10.1016/j.renene.2019.09.011.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13103, https://doi.org/10.1029/2008JD009944.

IEA, 2019: Global Energy and C02 Status Report 2019. IEA, accessed 1 June 2020, https://www.iea.org/reports/global-energy-co2-status-report-2019.

Jiménez, P. A., and Coauthors, 2016a: WRF-SOLAR: Description and clear-sky assessment of an augmented NWP model for solar power prediction. *Bull. Amer. Meteor. Soc.*, **97**, 1249–1264, https://doi.org/10.1175/BAMS-D-14-00279.1.

——, S. Alessandrini, S. E. Haupt, A. Deng, B. Kosovic, J. A. Lee, and L. D. Monache, 2016b: The role of unresolved clouds on short-range global horizontal irradiance predictability. *Mon. Wea. Rev.*, **144**, 3099–3107, https://doi.org/10.1175/MWR-D-16-0104.1.

Kain, J. S., and J. Kain, 2004: The Kain–Fritsch convective parameterization: An update. *J. Appl. Meteor.*, **43**, 170–181, https://doi.org/10.1175/1520-0450(2004)043<0170:TKCPAU>2.0.CO;2.

Klingler, A. L., and L. Teichtmann, 2017: Impacts of a forecast-based operation strategy for grid-connected PV storage systems on profitability and the energy system. *Sol. Energy*, **158**, 861–868, https://doi.org/10.1016/j.solener.2017.10.052.

Li, Y., X. Sun, H. Ning, H. Qin, and J. Zhao, 2018: An analysis of the physical characteristics of the summer low atmosphere in the Gobi Desert adjacent to Bosten Lake, Xinjiang, China. *Adv. Meteor.*, **2018**, 3061282, https://doi.org/10.1155/2018/3061282.

Lim, K. S. S., and S. Y. Hong, 2010: Development of an effective double-moment cloud microphysics scheme with prognostic cloud condensation nuclei (CCN) for weather and climate models. *Mon. Wea. Rev.*, **138**, 1587–1612, https://doi.org/10.1175/2009MWR2968.1.

López-Coto, I., J. L. Bosch, P. Mathiensen, and J. Kleissl, 2013: Comparison between several parameterization schemes in WRF for solar forecasting in coastal zones. *42nd ASES Natl. Solar Conf. 2013 (Sol. 2013)*, 605–614, http://proceedings.ases.org/wp-content/uploads/2014/02/SOLAR2013_0133_final-paper.pdf.

Mansell, E. R., C. L. Ziegler, and E. C. Bruning, 2010: Simulated electrification of a small thunderstorm with two-moment bulk microphysics. *J. Atmos. Sci.*, **67**, 171–194, https://doi.org/10.1175/2009JAS2965.1.

Meneguzzo, F., R. Ciriminna, L. Albanese, and M. Pagliaro, 2016: Italy 100% renewable: A suitable energy transition roadmap. arXiv:1609.08380.

Park, S., and C. S. Bretherton, 2009: The University of Washington shallow convection and moist turbulence schemes and their impact on climate simulations with the Community Atmosphere Model. *J. Climate*, **22**, 3449–3469, https://doi.org/10.1175/2008JCLI2557.1.

Perez, R., and Coauthors, 2013: Comparison of numerical weather prediction solar irradiance forecasts in the U.S., Canada and Europe. *Sol. Energy*, **94**, 305–3C26, https://doi.org/10.1016/j.solener.2013.05.005.

Prasad, A. A., and M. Kay, 2020: Assessment of simulated solar irradiance on days of high intermittency using WRF-solar. *Energies*, **13**, 385, https://doi.org/10.3390/en13020385.

Ruiz-Arias, J. A., C. A. Gueymard, J. Dudhia, and D. Pozo-Vázquez, 2012: Improvement of the Weather Research and Forecasting (WRF) model for solar resource assessments and forecasts under clear skies. *World Renewable Energy Forum (WREF 2012)*, Denver, CO, American Solar Energy Society, Vol. 3, 2379–2386, http://toc.proceedings.com/15138webtoc.pdf.

——, C. Arbizu-Barrena, F. J. Santos-Alamillos, J. Tovar-Pescador, and D. Pozo-Vázquez, 2016: Assessing the surface solar radiation budget in the WRF model: A spatiotemporal analysis of the bias and its causes. *Mon. Wea. Rev.*, **144**, 703–711, https://doi.org/10.1175/MWR-D-15-0262.1.

Skamarock, W. C., and J. B. Klemp, 2008: A time-split nonhydrostatic atmospheric model for weather research and forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, https://doi.org/10.1016/j.jcp.2007.01.037.

Thompson, G., and T. Eidhammer, 2014: A study of aerosol impacts on clouds and precipitation development in a large winter cyclone. *J. Atmos. Sci.*, **71**, 3636–3658, https://doi.org/10.1175/JAS-D-13-0305.1.

UNFCCC, 2015: Paris Agreement: Report of the Conference of the Parties on its twenty-first session, held in Paris from 30 November to 13 December 2015. UNFCCC, 42 pp., https://unfccc.int/resource/docs/2015/cop21/eng/10.pdf.

Xiu, A., and J. E. Pleim, 2001: Development of a land surface model. Part I: Application in a mesoscale meteorological model. *J. Appl. Meteor.*, **40**, 192–209, https://doi.org/10.1175/1520-0450(2001)040<0192:DOALSM>2.0.CO;2.

Yang, H., and J. Kleissl, 2016: Preprocessing WRF initial conditions for coastal stratocumulus forecasting. *Sol. Energy*, **133**, 180–193, https://doi.org/10.1016/j.solener.2016.04.003.