Attenuation of Large-Scale Solar PV Production by Bushfire Smoke in South-East Australia

Geert Scholma (geert@amperon.co), Ydo Wexler, PhD (ydo@amperon.co)
Produced in collaboration with Australian Energy Market Operator (AEMO)

Summary

Australia is no stranger to bushfires, which were recorded for more than a century with various levels of severity. The fires in the summer of 2019-2020 were especially harsh and damaging, burning more than 180,000 square kilometers of forest and grassland, destroying more than 3,000 homes, and impacting wildlife, public health, and the local economy far more than in previous years[1,2]. Due to the smoke from these fires, it has also been reported that in Sydney and Canberra electricity production from rooftop solar PV has decreased by as much as 15-45% on some days[3]. One aspect of those damages still to be assessed is the magnitude of effect on large-scale solar generation.

In this study we examine the effect smoke plumes from bushfires had on electricity production of 20 large-scale photovoltaic solar farms in New South Wales (NSW), Victoria (VIC), the Australian Capital Territory (ACT) and South Australia (SA), within the grid operated by the Australian Energy Market Operator (AEMO)[4,5]. To facilitate such study, we employed SCADA electricity production data of the 20 solar farms under study, as well as estimations of the generation in each of the farms, calculated by a software package, named *pvlib,* for simulating the performance of photovoltaic energy systems[6]. Those simulated generation loads take into consideration the farms' specifications and solar irradiance information. To compute predictions that consider the prevailing amounts of smoke plumes we used both fine particulate matter measurements of up to 2.5 micrometers (PM2.5) and re-analysis estimates of solar irradiance with and without aerosols taken from NASA's Goddard Earth Observing System (GEOS-5)[7].

1 https://en.wikipedia.org/wiki/2019%E2%80%9320_Australian_bushfire_season
2 https://www.unenvironment.org/news-and-stories/story/ten-impacts-australian-bushfires
3 https://www.solaranalytics.com/au/blog/how-much-does-smoke-haze-affect-rooftop-solar-production
4 https://www.aemo.com.au
5 http://www.nemweb.com.au
6 https://pvlib-python.readthedocs.io
7 https://gmao.gsfc.nasa.gov/pubs/docs/Lucchesi617.pdf

For the solar farms under study, we found that during December 2019 and January 2020[8], a mean decrease in generation of 4.1% per plant (ACT: 10.3%, NSW: 4.1%, VIC: 3.5%, SA: 1.2%) could be attributed to smoke plumes. The biggest impact of smoke on electricity generation was found for the two solar farms near Canberra in the ACT, for which during December 2019 and January 2020 a decrease in generation of 8.6% and 12.0% respectively can be attributed to aerosols or PM2.5. We consider the analysis showing this significant impact particularly reliable, as the two farms lie within 10 km of an official PM2.5 monitoring station and SCADA data of the two farms was among the most reliable under study. We note that while bushfires started in NSW in August 2019, simulations show a significant decrease in expected solar generation when taking smoke into account only starting November 2019.

We believe this study provides a first glance into the effects of smoke from abundance of bushfires on large-scale solar electricity generation, as it occurred in the summer of 2019-2020, and hope the results provided herein can serve as a guide for future planning. Further development of large-scale solar farms in Australia and other parts of the world, as well as other participants in the electricity markets, can benefit from better understanding and/or forecasting of aerosol-corrected solar generation in areas regularly affected by large amounts of smoke.

Solar Farms

In this study we focused on 20 solar farms in NSW, VIC, the ACT and SA, for which we gathered publicly available SCADA data. For 9 locations, plant availability was variable and unknown, or generation was curtailed for extended periods, making the SCADA data practically unusable as a reference. Even for the remaining 11 locations, there is no reliable data source or timeseries stating the available capacity of the farms, which forms an uncertainty in the interpretation of the SCADA data. For convenience, from here on, all farms are specified by their DUID (Dispatchable Unit ID).

Data

For this study we used both publicly available data and proprietary weather forecasts. We briefly list the data sources and any processing performed on this data. Some of the main parameters used for each farm are shown in *Table 1.*

1. Plant Specifications**:** we used location, positioning, system type, and plant capacity. Any publicly available data, or otherwise the specifications for WRSF1 (the only farm for which we found a full specification) were used as a base/default for any missing data.
    a. Location: latitude, longitude, and altitude

[8] results based on data up to January 18th 2020

b. Positioning:
   i. Vertical tilt: Publicly available for 4 of the 6 untracked systems. For the two systems in ACT with unknown tilt, we used 25° as default (0 = horizontal)
   ii. Horizontal orientation: Orientation of the tracking axis, or otherwise of the panels. For 19 of the 20 farms obtained from Google maps, otherwise filled with 0 (North) as default.
   iii. Tracking: Single axis or no tracking. 14 of the 20 farms fully tracked.

c. System:
   i. Module and inverter brand / type, defaulted to specifications of WRSF1.
   ii. Farm layout: Number of panels per string and number of strings per inverter were defaulted to the specifications of WRSF1. The number of inverters was adjusted to match the total AC power of individual farms.

d. Capacity (MW):
   i. Max total module DC power and max total AC grid power: Mostly publicly available, but for NYNGAN1, MLSP1, ROYALLA1 and BROKENH1 the DC values have been estimated to best match the SCADA values.

	Name	DUID	Lat	Lon	Elevation	Tracking	Orientation	Tilt	DC Power	AC Grid Power	Region
0	Bannerton	BANN1	-34.69	142.78	72	1	0	NaN	110.00	88.0	VIC
1	Beryl	BERYLSF1	-32.35	149.46	408	1	8	NaN	110.00	87.0	NSW
2	Bungala One	BNGSF1	-32.43	137.84	32	1	358	NaN	138.00	110.0	SA
3	Bungala Two	BNGSF2	-32.43	137.84	32	1	358	NaN	138.00	110.0	SA
4	Broken Hill	BROKENH1	-31.99	141.39	286	0	0	25.0	53.00	53.0	NSW
5	Coleambally	COLEASF1	-34.75	145.93	120	1	352	NaN	188.00	150.0	NSW
6	Finley	FINLYSF1	-35.65	145.51	107	1	0	NaN	175.00	133.0	NSW
7	Gannawarra	GANNSF1	-35.73	143.78	83	1	0	NaN	60.00	50.0	VIC
8	Griffith	GRIFSF1	-34.32	146.12	127	1	0	NaN	36.00	27.0	NSW
9	Karadoc	KARSF1	-34.41	142.25	50	1	0	NaN	112.00	90.0	VIC
10	Manildra	MANSLR1	-33.17	148.72	468	1	7	NaN	55.90	46.7	NSW
11	Mugga Lane	MLSP1	-35.40	149.15	621	0	0	25.0	16.25	13.0	ACT
12	Moree	MOREESF1	-29.57	149.87	220	1	359	NaN	70.00	56.0	NSW
13	Numurkah	NUMURSF1	-36.16	145.48	108	1	0	NaN	128.00	100.0	VIC
14	Nyngan	NYNGAN1	-31.56	147.09	175	0	0	25.0	142.80	102.0	NSW
15	Parkes	PARSF1	-33.12	148.08	274	1	0	NaN	66.00	50.5	NSW
16	Royalla	ROYALLA1	-35.49	149.14	809	0	358	25.0	25.00	20.0	ACT
17	Tailem Bend	TBSF1	-35.28	139.49	27	0	345	17.0	127.00	95.0	SA
18	Wemen	WEMENSF1	-34.80	142.54	24	1	10	NaN	110.00	88.0	VIC
19	White Rock	WRSF1	-29.76	151.55	974	0	0	30.0	25.00	20.0	NSW

Table 1: Main specifications per solar farm, used for the analysis. The tracked farms have no fixed tilt, indicated with NaN.

2. SCADA solar production data: publicly available with a 5 minute resolution[4], resampled to hourly mean values.

3. Weather
    a. To model the total direct and diffuse solar irradiance incident on the solar panels, we used a solar irradiance re-analysis (best estimate of reality, combining observations and model simulations) with and without aerosols provided by NASA.
    b. To model the module/system temperature and spectral mismatch, we used weather data taken from a proprietary vendor feed, from which the following variables were used: air temperature, wind speed, air pressure, and relative humidity. This source consists of an optimized mix of global weather predictions of which we took an archived 0-1 hour ahead forecast for locations typically within 10 kilometers of the individual farms.

4. Air Quality
    a. Global Horizontal Irradiance (GHI) with and without aerosols was taken from NASA GEOS-5, from which we used the variables *surface incoming shortwave flux, surface net downward shortwave flux assuming no aerosol* [7] . The dataset provides model re-analysis estimates for a grid of size 0.312° longitude and 0.25° latitude. For this study we used the closest datapoint per solar farm.
    b. We downloaded data from four sources with measurements of the air quality[9, 10, 11, 12] from which we used concentrations of fine particulate matter with a diameter of less than 2.5 µm (PM2.5, in µg/m3). As those measurements contained small amounts of duplicates and gaps, we resampled them to hourly means and filled any gaps by interpolation. We used these measurements by converting them to a global radiation dimming factor between 0 and 1 according to *Equation 1* as suggested in a recent paper by *Peters et.al.* [13]. We note that alternative air quality sources like Aeronet[14], luftdaten.info[15] exist that might be added for future research or alternative locations.

---

9 https://www.dpie.nsw.gov.au/air-quality/search-for-and-download-air-quality-data
10 https://www.dpie.nsw.gov.au/air-quality/rural-air-quality-network-live-data
11 https://www.data.act.gov.au/Environment/Air-Quality-Monitoring-Data/94a5-zqnn/data
12 https://data.sa.gov.au/data/dataset
13 https://pubs.rsc.org/en/content/articlelanding/2018/EE/C8EE01100A
14 https://aeronet.gsfc.nasa.gov/cgi-bin/draw_map_display_aod_v3
15 https://luftdaten.info

Methodology

The core of our analysis revolves around the expected solar generation based on physical simulations. We analyze those estimates with respect to the difference in generation when simulating with and without the presence of aerosols or PM2.5 (smoke), and assessments of how well they fit the actual generation as it appears in the SCADA feeds. With these data sources and simulations, we are able to statistically decompose the generation data as the sum of the estimated generation without the effect of aerosols, and the estimated reduction of generation due to smoke in each of the solar farms.

Our simulations were performed using the open source package *pvlib*[6], a community supported tool that enables simulating the performance of photovoltaic energy systems. From there we employed an object called the *ModelChain* that allows for a high level of automation for the modeling process, and which calculates a system's AC output over time for a specific location and a given collection of modules and inverters.

The modelling process, as performed via the *ModelChain* object, first calculates the horizontal and vertical position of the sun (solpos) and the clear sky solar irradiance at the specified location for a given series of datetimes. Combining solpos with GHI – the sum of direct and diffuse solar irradiance on a horizontal surface, and air-pressure – it calculates the Direct Normal Irradiance (DNI) – direct solar irradiance on a plane facing the sun.

Then, the Diffuse Horizontal Irradiance (DHI) – diffuse solar irradiance on a horizontal surface – can be computed by subtracting from the GHI the fraction of DNI that's incident on a horizontal surface. Once direct and diffuse radiation are fully specified, the loss of efficiency due to spectral mismatch, reflection, and overheating, is calculated given the plant specifications, air temperature, wind speed, relative humidity and air pressure.

We simulated the hourly incident solar generation in each of the farms with and without smoke via *pvlib*, using both the NASA dataset and PM2.5 measurements as smoke inputs. Comparing the simulation predictions based on NASA GHI with aerosols, we validated the quality of the SCADA data per farm. We computed the correlation between the SCADA measurements and simulations, as well as the normalized Mean Absolute Error (nMAE) as given by *Equation 1:*

$$\textit{Equation 1.} \qquad \frac{\Sigma_t |s_t - p_t|}{\Sigma_t |s_t|}$$

where s and p are the vectors of SCADA and the simulated values respectively. In order to obtain a higher level of confidence and reduce noise present on both the hourly and daily levels that may stem from missing values and outliers, we performed the calculations on weekly mean values. For further use of the SCADA data in the analysis, we filtered out farms with high variability, accepting only the farms with an nMAE below 20% and a correlation with the simulated prediction over 70%, which left us with valid SCADA data for 11 of the 20 farms.

Denoting the weekly mean simulated loads with and without aerosols as S_{smoke} and S_{clear} , we define the absolute smoke impact I_{org} as in *Equation 2* to equal the total sum of the weekly mean differences S_{clear} - S_{smoke} divided by the sum over the weekly mean values S_{clear}, for the period of December 2019 and January 2020:

Equation 2. $$I_{org} = \frac{\Sigma_t (S_{clear} - S_{smoke})}{\Sigma_t (S_{clear})}$$

Since this impact indicator can be highly affected by factors other than smoke, we wanted to remove bias from our simulations. Such bias is more evident when no bushfire smoke is known to occur, and can be the result of background signals, arising from things like dust or sea-spray, or the result of incomplete farm specification, model inaccuracies or consistent errors in the weather data. To do that, we fitted a linear regression model for each farm, where the dependent variable is the actual SCADA generation and the two independent variables are the predictions without aerosols, S_{clear}, and the difference between the two predictions with and without aerosols, S_{clear} - S_{smoke}, all series first aggregated to weekly mean values, as given by *Equation 3:*

Equation 3. $$SCADA = \beta_1 (S_{clear}) + \beta_2 (S_{clear} - S_{smoke})$$

Next, we define an adjusted smoke impact, I_{adj}, per farm as $I_{adj} = I_{org} \frac{\beta_2}{\beta_1}$, which is an unbiased estimator of the decrease in generation due to smoke. For the 9 farms that didn't have valid SCADA data, we couldn't perform this statistical adjustment. Instead, we calculated the average adjustment factor $\frac{I_{adj}}{I_{org}} = 0.57 \pm 0.23$, by dividing the mean adjusted impact over the mean unadjusted impact for the 11 farms with valid SCADA data. We then applied this factor to the other farms to be able to present a possibly more accurate estimate of the real impact of smoke on power production.

Equation 4. $$\frac{I(PM2.5)}{I_0} = \exp\left(\frac{-PM2.5}{750 \pm 90}\right)$$

We performed this analysis both for NASA based GHI with aerosols, and for the PM2.5 observations as input for the simulations, as specified by *Equation 4*. On average both the adjusted and unadjusted smoke impact were significantly higher when using NASA GHI than when using PM2.5 measurements. We don't have an immediate explanation for this difference. The advantages of the NASA data are that it is available over a dense grid of 0.25° x 0.312° (latitude x longitude), the maximum distance of a closest grid point from any solar farm was 20 km, and the data is available as a forecast to use for daily forecasts. The disadvantage is that these values are weather model simulations, not measurements. For the PM2.5, however, we were dependent on a limited number of locations from which data is available. We used a total of 8 official PM2.5 measurement locations, as well as 9 additional locations, usually referred to as rural locations. Where possible, we computed a weighted average of pairs of locations to create additional virtual locations, in an attempt to reduce the effective distance to the solar farms. The average distance including those points was 58 km, ranging from 2 km to 139 km. The advantage of the PM2.5 values is that those are real measurements as opposed to the simulated values from NASA. The main drawback is the on average larger distance to the farm locations, compared to the NASA data.

Results

Weekly mean SCADA measurements and simulated weekly mean electricity generation loads with and without aerosols based on the NASA dataset are shown when available in *Figure 1* for each of the 20 solar farms. As evident from the figure, the quality of the SCADA data highly varies among the farms, whereas the simulations often contain some consistent bias, raising the need for a statistical adjustment. We note that for many farms the SCADA data values are lower than those from the simulations, indicating intermittent availability of the farm, unreported curtailment, or other unknown effects. For the rest of the study we have used only the data between Oct 1st 2019 and Jan 18th 2020, since for these dates all datasets had data and we could therefore consistently study the results among all different datasets.

To eliminate noisy results from farms with low quality SCADA data, where cleaning the data was not a viable option, we applied a filter that computes the fit quality between actual generation values and predictions when accounting for the NASA aerosols. In *Figure 2*, we show nMAE, the correlation, and correlation - nMAE (the last was mainly used for ranking). Using these scores, we've set a minimum requirement of a correlation over 70% and an nMAE below 20% for further use of the SCADA data. As can be seen in the figure, the SCADA data of the best (leftmost) 11 solar farms satisfy these conditions and therefore have been marked as reliable enough for further study. For these 11 farms, any weekly SCADA data more than 20% of full capacity lower than the simulated predictions, we marked as outliers and also excluded from further analysis.

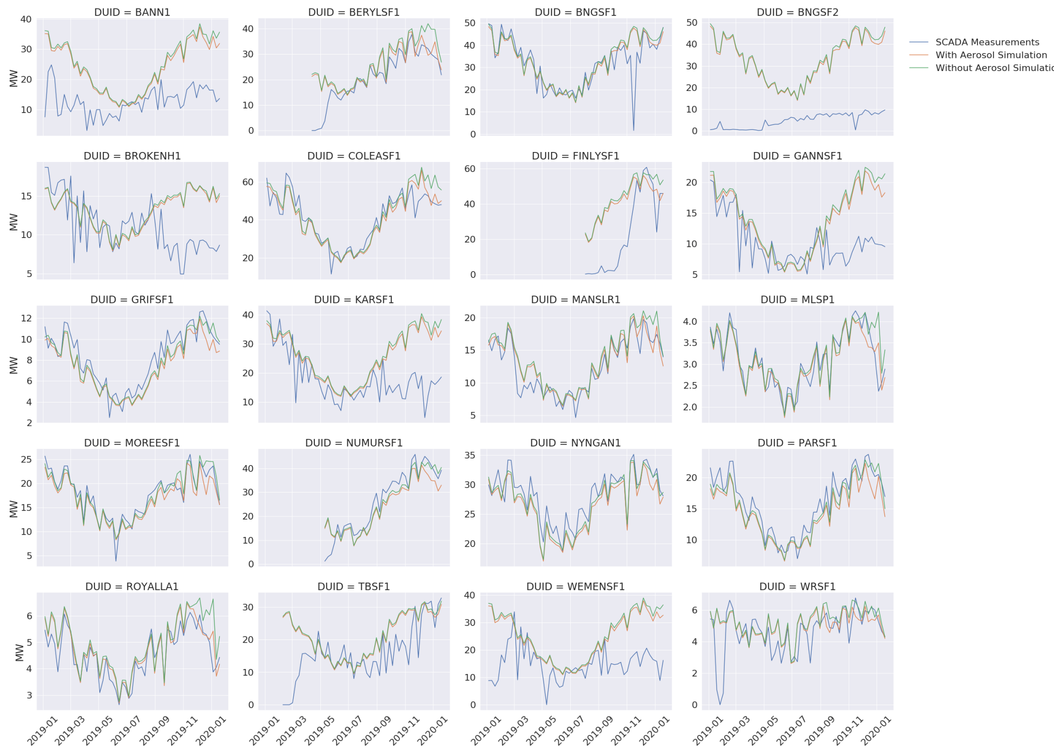

Figure 1. Weekly mean plots of SCADA Measurements (Blue) and pvlib's simulations based on NASA's GHI with (Orange) and without (Green) aerosols, all in MW.

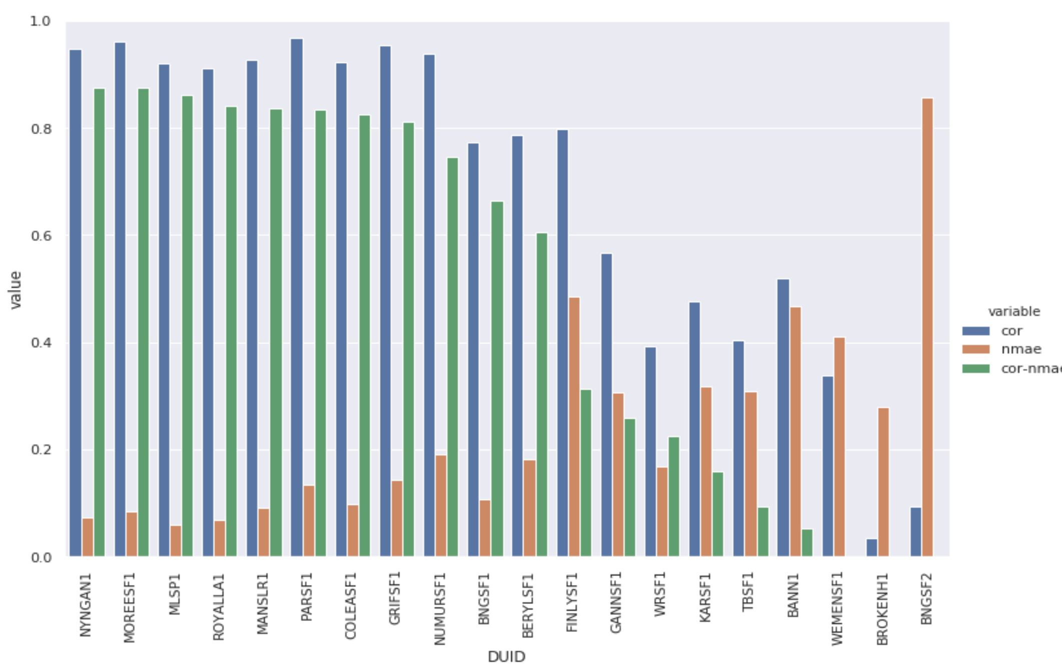

Figure 2. The nMAE and correlation quality for every farm when computed from weekly mean actual SCADA generation and simulations when accounting for NASA aerosols. DUID are shown when ranked by Correlation - nMAE. Only SCADA data of the leftmost 11 solar farms have been used for further study.

A main focus of this study has been to determine the approximate decrease in solar generation due to smoke, for the period of December 2019 - January 2020. *Figure 3* plots the solar farms by location and statistically unadjusted smoke impact as a percentage of estimated generation without aerosols. The outer circle displays the estimated impact when using PM2.5 measurements, and the inner circle the estimated impact when using simulations based on NASA data with aerosols. Unadjusted impacts above 10% were observed for 10 solar farms when simulated using NASA data, however, using PM2.5 measurements the impact was above that level for only 6 locations. We note that technically the locations for the two data sources aren't always identical, but their effect mostly matches. For MOREESF1 and BERYLSF1 in particular this was not the case, and we don't have any clear explanation for this finding. For the former we used PM2.5 data from a station located 10 km from the farm, indicating that the distance between PM2.5 measurements and the farm's location is probably not the reason for this outlier.

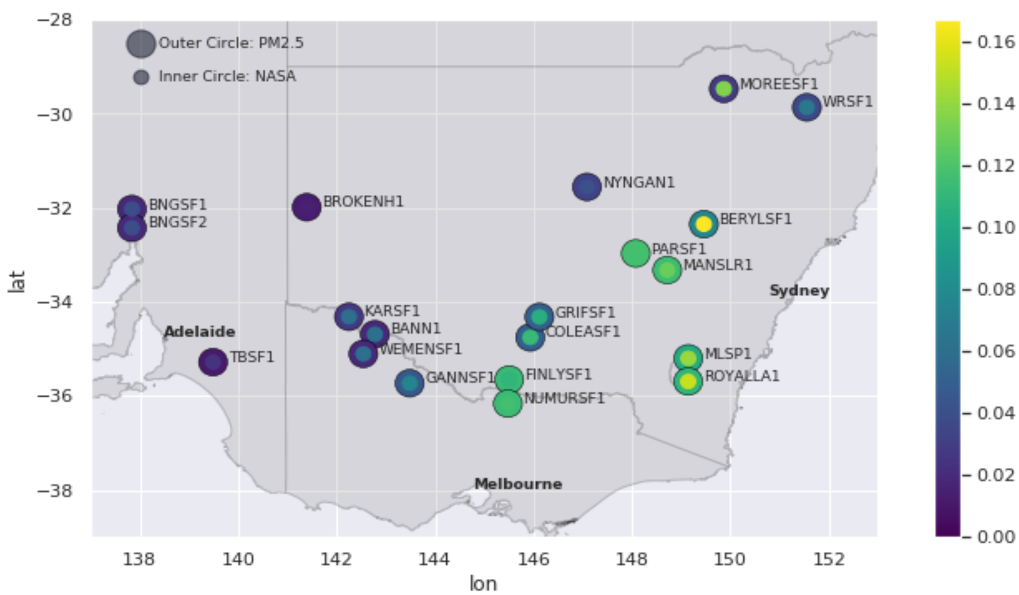

Figure 3. Unadjusted smoke impact on solar electricity generation, for inner circle based on NASA GHI and for outer circle based on PM2.5. Some nearby locations are slightly moved apart to prevent visual overlap. Though numbers for both sources generally align, especially for MOREESF1 and BERYLSF1 they are relatively far apart, for which we don't have any clear explanation.

The adjusted smoke impact we got after fitting the SCADA data with the solar irradiance based simulations is shown in *Figure 4 and Figure 5*, where the former is a geographical map and the latter a barplot. The plotted values are based on an average of both sources of NASA aerosols and PM2.5 measurements. We consider these values to be an unbiased estimator for the drop in generation due to smoke, but acknowledge that the quality of this estimation highly depends on the quality of all data sources.

It is clear from a comparison of *Figure 3* with *Figure 4* (and later from *Figure 7*) that the estimated values for smoke impact after the statistical adjustment are significantly lower than before the adjustment is applied, leaving only 6 solar farms with an estimated impact above 6% and only one (MLSP1) with an impact above 10%. This indicates the original simulations may have overestimated the smoke impact. Especially the relatively high unadjusted smoke impacts using NASA data have been adjusted to much lower values. *Figure 6* shows the differences between NASA and PM2.5 after the statistical adjustment.

The adjusted smoke impacts shown in *Figure 4* and *Figure 5* show a relatively low signal below 2% for 5 plants in the North/West of SA (BNGSF1/2, TBSF1), and NSW (BROKENH1, NYNGAN1). Consistent medium impacts in the range of 2%-6% have been found for 9 other farms in a zone between the coast and North/West for VIC (BANN1, GANNSF1, KARSF1, WEMENSF1), and NSW (COLEASF1, GRIFSF1, MOREESF1, PARSF1, WRSF1). Larger smoke effects of more than 6% have been found for 6 plants further towards the South/East in VIC (NUMURSF1), NSW (BERYLSF1, FINLYSF1, MANSLR1), and ACT (MLSP1, ROYALLA1). The difference in impact found for MLSP1 and ROYALLA1, which are within 10 km of each other, has remained unclear to us.

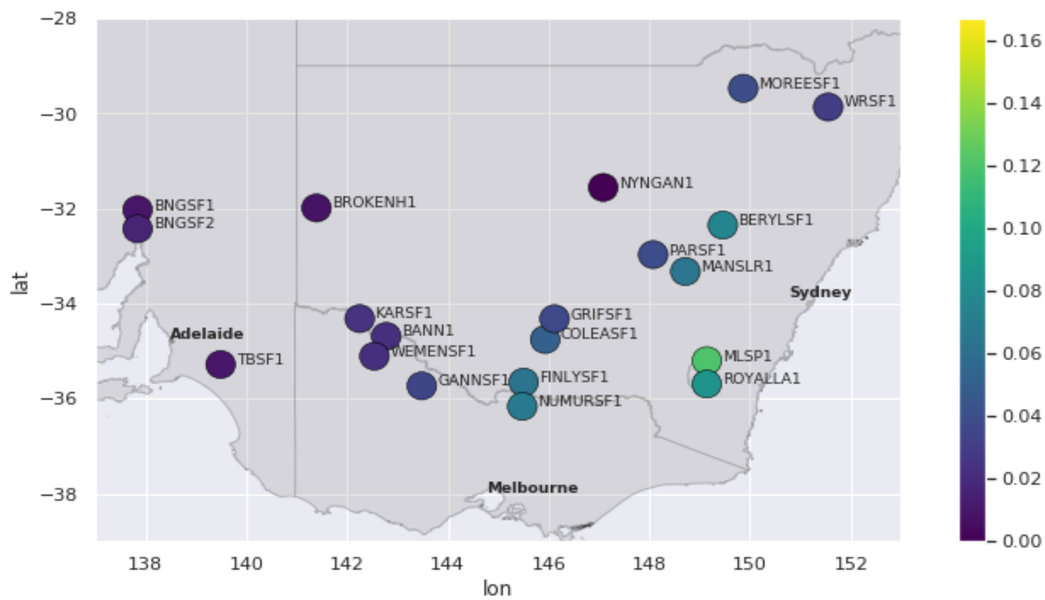

Figure 4. Geographical map of final adjusted smoke impacts for all solar farms, averaged over both smoke sources (NASA and PM2.5)

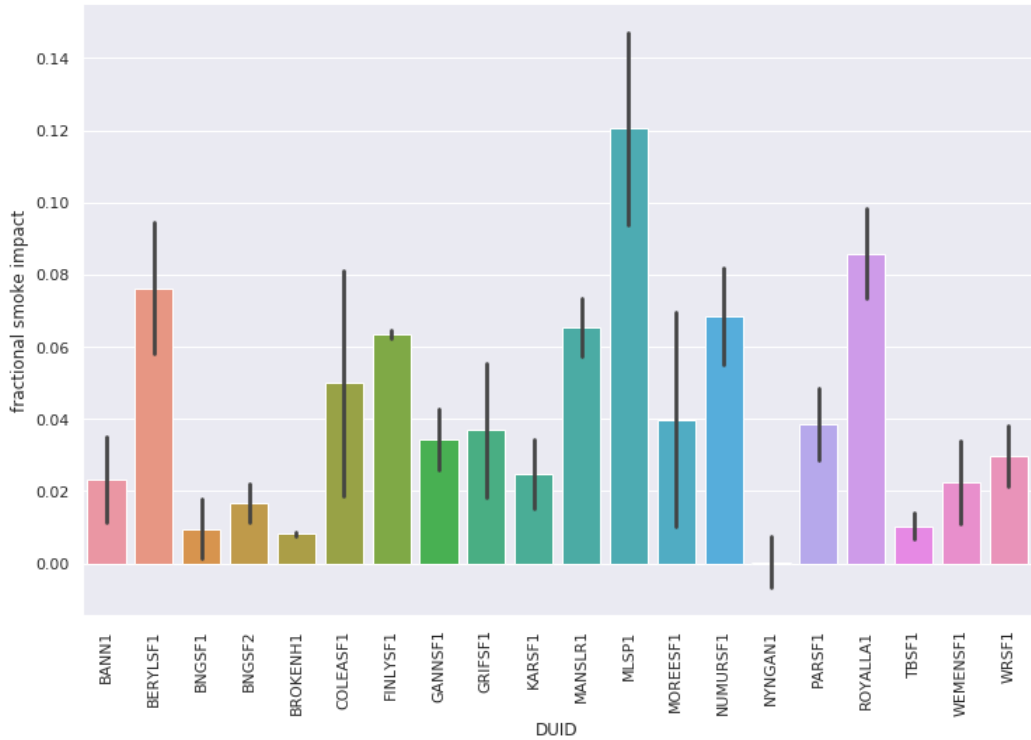

Figure 5. Barplot of final adjusted smoke impacts for all solar farms, averaged over both smoke sources (NASA and PM2.5). The black error bars in the plot indicates the difference between the two simulations per farm.

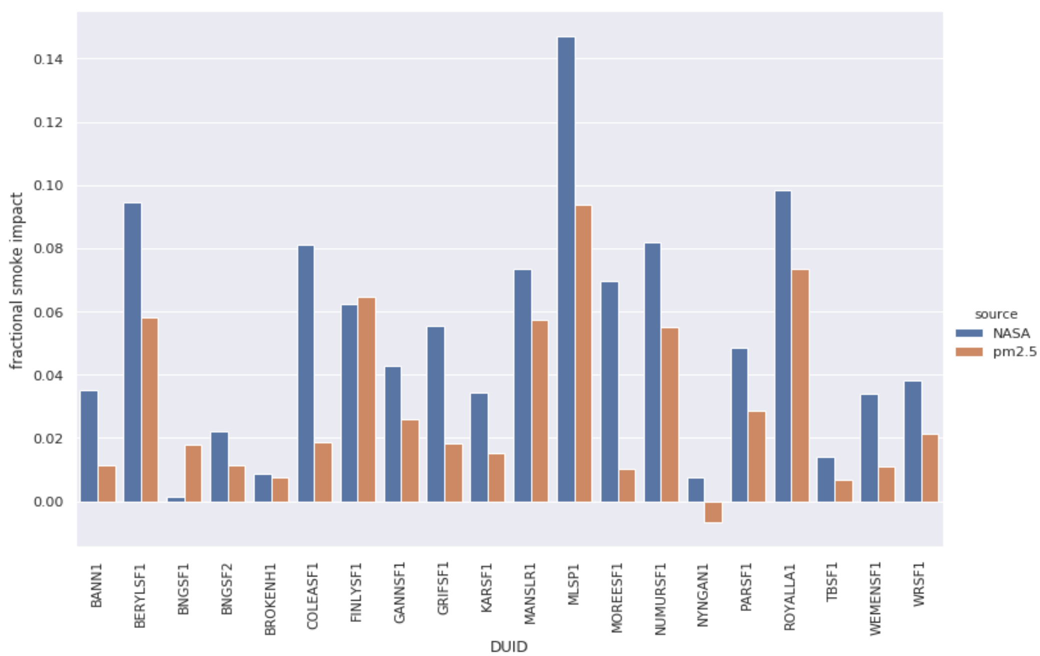

Figure 6. Barplot of the adjusted smoke impacts for all solar farms, compared for the two different smoke sources (NASA and PM2.5).

For some locations like BROKENH1 and FINLYSF1, the difference between the results based on NASA smoke and PM2.5 are relatively small, while for other locations such as MLSP1 and COLEASF1, the differences were more significant. In general, we can see that even after the statistical adjustment the smoke impact based on the NASA data was estimated to be much higher than based on the PM2.5 data. As both results have been obtained with the same statistical adjustment method and are thereby supposed to be unbiased estimators, it's difficult to determine which result is more accurate. Longer datasets of reliable SCADA and air quality data, as well as solar irradiance or cloud cover measurements taken at the solar farms, could help further verify this.

Figure 7 shows a barplot with the original and adjusted smoke impacts. Clearly the adjusted smoke impacts are much lower than the original ones. MLSP1 is the main outlier here, showing a similar smoke impact after adjustment, for unclear reasons. In general it's not clear why the adjusted smoke impact is significantly lower than estimated by the simulations. Lacking proper measurements of air quality and solar irradiance at the solar farms, and having SCADA data that might have been impacted by several factors other than smoke, like curtailment, technical unavailability or negative market prices, the difference between original and statistically adjusted) smoke impacts remains a topic for further study.

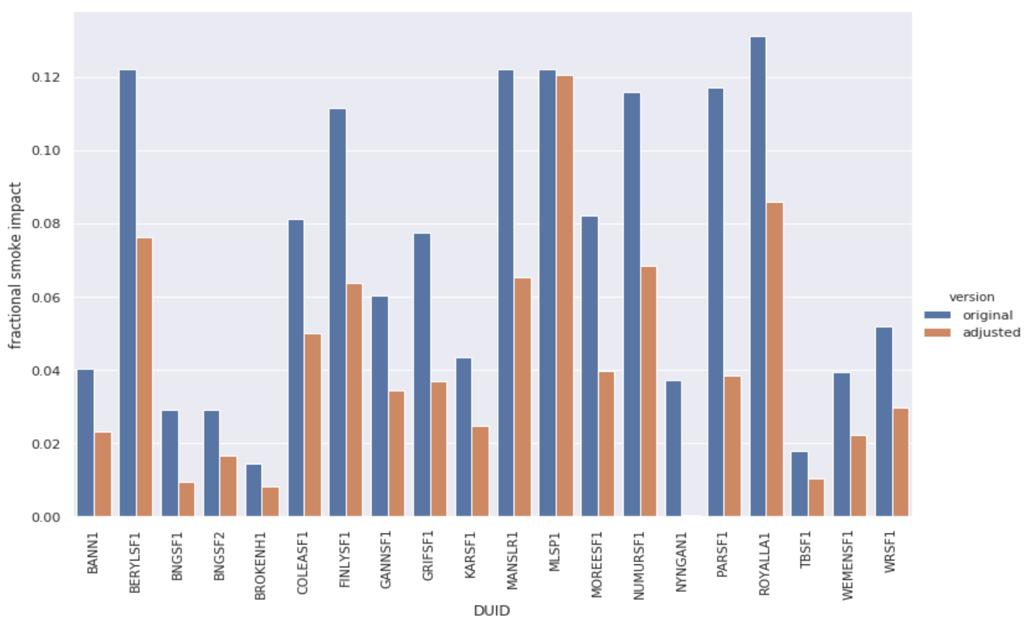

Figure 7. Barplot of the original vs adjusted smoke impacts for all solar farms, averaged over the two different smoke sources (NASA and PM2.5).

Finally, in *Figure 8* we show an example of the hourly shapes and the accuracy of the simulations with aerosols and without aerosols, for one of the farms near Canberra, with both NASA dataset and PM2.5. We show the predictions for ROYALLA1 for the first 5 days in 2020, towards the end of which some of the biggest bushfires for that summer were raging in

NSW, and the region saw a large increase in smoke plumes. In the upper plot with unadjusted values, it's visible that for this solar farm the original simulation using PM2.5 was over-forecasting the effect of smoke, while the one using NASA aerosols was a little closer, but under-forecasted the effect. The lower plot includes the final adjusted values averaged over both NASA and PM2.5. We can see the large drop in actual generation, and how, except in the last day, the adjusted simulations quite accurately account for much of it.

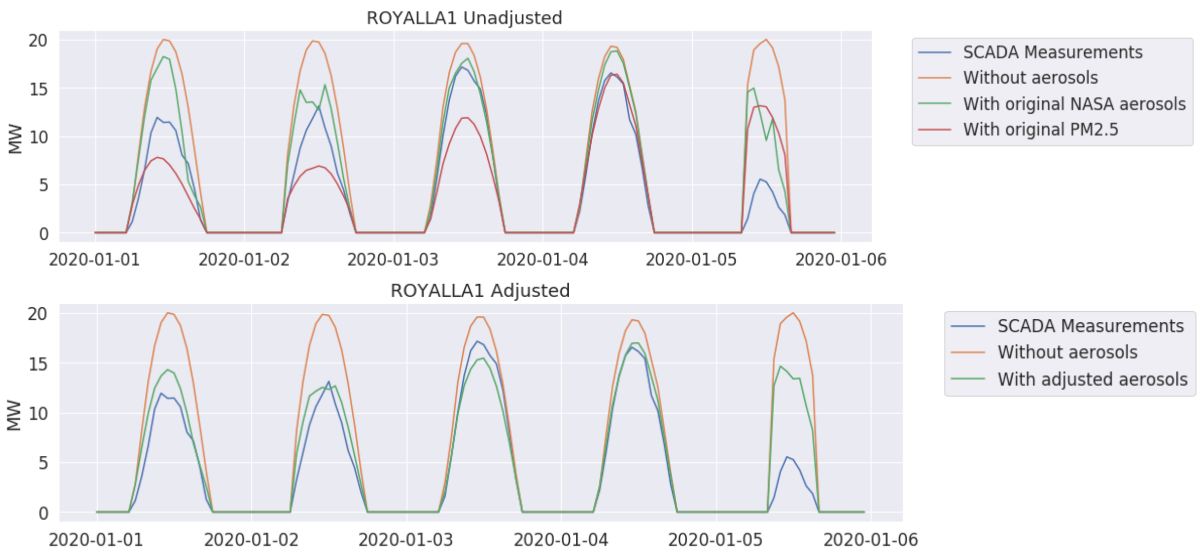

Figure 8. Solar generation and simulations before and after statistical adjustment for ROYALLA1 for the first five days in January 2020.