Renewable Energy-Associated Emissions

Mark Barna (Econ 6375: Applied Econometrics)

Renewable energy has great potential to reduce pollutant emissions associated with the electricity generation sector, however solar and wind produce power based on the whims of the weather. Thus, their output fluctuates over time. Until large-scale energy storage that can shift excess supply to meet demand has increased across the country, flexible fossil fuel power plants must quickly power up and down to compensate for these variations in supply. In this paper, I create a regression model using solar and wind power generation data as well as fossil fuel plant emissions data from several mid-Atlantic and Midwest states to investigate the relationship between oscillations in intermittent renewable energy generation and power plant emissions. I find that there is a small inverse relationship, suggesting that sudden drops in wind and solar output may cause additional pollutant discharges.

I. Introduction

Over the last decade, the United States has seen rapid build-out of renewable energy generation, especially wind and solar power. This expansion is necessary to reduce pollution and greenhouse gas emissions that lead to climate change. In 2014, total U.S. carbon dioxide (CO_2) equaled 6,124 million tons, of which roughly 37 percent originated from electricity generation (EPA 2014). If the country plans to seriously address its contribution to climate change, it needs to shift the power generation fuel mix away from fossil sources of energy, namely coal and natural gas.

In addition to the greenhouse gas, CO_2, fossil power plants emit nitrogen oxides (NO_x) and sulfur dioxide (SO_2). Both chemicals contribute to acid rain. They also react with other airborne compounds to form particulate matter than can damage the human respiratory system as well as cause smog.[1] [2]

In order to address these environmental impacts, states and the federal government have implemented subsidies for renewable energy generation (Kaffine, McBeen, and Lieskovsky 2013). Two predominant methods of government support are federal tax credits and state renewable portfolio standards (RPS), the latter being a mandate to produce a certain percentage of electricity from renewable energy. As of February 2017, 29 states plus Washington, DC had some form of RPS directive.[3]

To quantify the expected reductions in emissions achieved by these policies, lawmakers must model the projected pollution avoided by renewable energy. If they simply assume that solar and wind power will directly offset all emissions on a one-to-one basis for each megawatt (MW) of electricity produced, they may overestimate renewable energy's benefits. As Katzenstein and Apt (2009) and Gonzalez-Salazar et al. (2017) show, solar and wind power may actually cause additional emissions by requiring fossil fuel plants to quickly ramp up when the wind suddenly dies down or a cloud obstructs a solar panel's view of the sun. These spikes in emissions should be considered in any policy proposals that examine the emissions reductions achieved by clean energy.

In this paper, I attempt to quantify the additional emissions triggered by variations in solar and wind generation in the region of the United States served by PJM Interconnection. PJM

[1] U.S. Environmental Protection Agency. "Basic Information about NO_2." https://www.epa.gov/no2-pollution/basic-information-about-no2#Effects (accessed December 10, 2017).

[2] U.S. Environmental Protection Agency. "Sulfur Dioxide Basics." https://www.epa.gov/so2-pollution/sulfur-dioxide-basics#effects (accessed December 10, 2017).

[3] DSIRE. "Renewable Portfolio Standard Policies." NC Clean Energy Technology Center. http://www.dsireusa.org/resources/detailed-summary-maps/ (accessed December 10, 2017).

manages the electricity transmissions system and wholesale electricity market in the Mid-Atlantic and parts of the Midwest (Figure 1). I model the relationship by regressing emissions on changes in renewable generation. I use the change rather than the total hourly amount of renewable energy generated because I am interested in the oscillations, themselves..

(Note: for ease, I will often refer to solar and wind generation simply as "renewable generation", understanding that these are not the only forms of renewable energy).

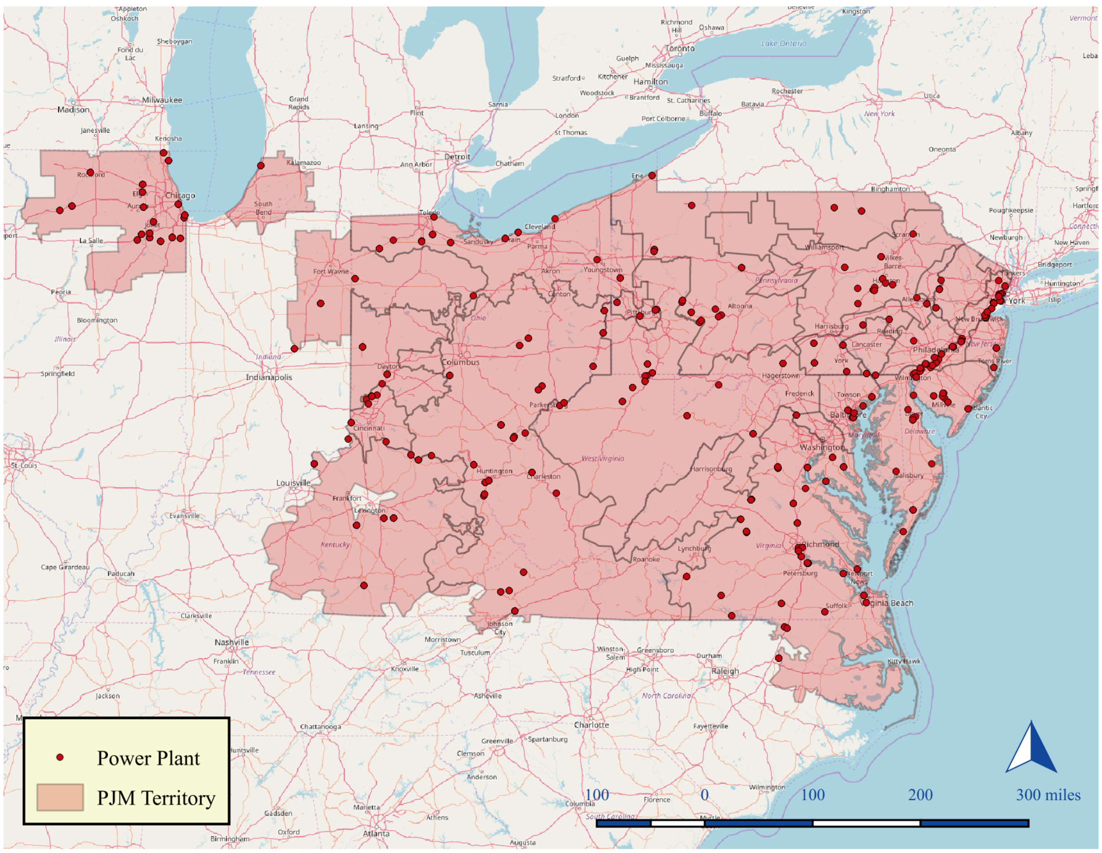

FIGURE 1: PJM INTERCONNECTION

Source: Author's calculations using GIS data supplies by PJM.

II. Literature Review

In reviewing other papers on this topic, I did not find any prior analysis that looks explicitly at the relationship between fossil fuel emissions and intermittent solar and wind output using real-world historical data from the PJM Interconnection region. Katzenstein and Apt model emissions from hypothetical power systems using factory specifications for fossil power generators (2009), and Gonzalez-Salazar et al. examine an abstract "typical" solar and wind energy generation curve to determine different fossil fuel plant types' ability to provide backup power (2017). Kaffine, McBeen, and Lieskovsky (2013) look at the fossil fuel power plant emissions avoided by wind power generation in Texas, and in doing so, note that the offset is reduced by plant emissions due to cycling.

In a 2009 paper, Katzenstein and Apt model a theoretical power system that is supplied from both renewables (wind and solar) and natural gas generators. The goal of this paper is to quantify emissions of CO_2 and NO_x that result from the cycling up and down of natural gas generators in response to intermittent solar and wind resources. To populate the model, the authors use actual generation data from four wind farms and one solar facility sampled at one-minute intervals. In addition, they sourced manufacturer emissions and heat rate data from two models of natural gas turbines. The authors find that as a result of plant cycling, the CO_2 emissions reductions achieved were about 75 to 80 percent of the amount expected without factoring in cycling emissions. Further, NO_x emission reductions were only 30 to 50 percent of the expected amount for one model of gas turbine, while they actually increased by 200 to 400 percent for the other model.

Gonzalez-Salazar et al. look at the operational characteristics of conventional power plants to determine their ability to provide backup generation to intermittent renewables. The authors propose a method to quantify these attributes in order to compare types of conventional power plants. They measure the hourly generation of wind and solar in Germany over a period of 48 hours, normalize it, and then scale the maximum output to match the maximum possible output of the conventional plant being analyzed. They then calculate a ratio of backup ability by taking the ratio of the modeled output of the plant plus the scaled renewables measured against the total capacity of the renewables times 48. To determine the output of the conventional power plant, the authors model the power output versus time curve of various technologies to see how long it takes to ramp up to full power. Using this model, they found natural gas plants to be the most flexible.

While looking at the power output curves for various technologies, the authors also estimate emissions output by technology based on whether the plant started up cold or was running in an idling mode, generating little or no power. They found that cold starts caused 2 to 16 times more emissions than hot starts. The authors also show that gas plants (the most flexible resource) produce more NO_x and CO (carbon monoxide) emissions than coal plants while idling, but less CO_2 and almost no SO_x (sulfur oxides).

Finally, while these prior two papers construct theoretical models for the relationship between renewables and fossil plant emissions, Kaffine, McBeen, and Lieskovsky use actual wind generation data from Texas from 2007 to 2009, along with average emissions data from 2000, to determine how emissions from fossil fuel power plants are offset by wind power installations. The authors show that most prior research in this area had used average power plant emissions data weighted by fuel mix to calculate the displaced CO_2, SO_2 and NO_x emissions. This type of analysis does not take into account the variation in fuel mix based on the marginal fuel cost. In other words, depending on the amount of load served at any particular point, if the marginal fuel is natural gas, wind power will displace natural gas plants before coal plants. Since natural gas is cleaner than coal, the emissions savings will be less than the prior models had suggested.

Kaffine, McBeen, and Lieskovsky also note that the data capture additional emissions as a result of fossil fuel plants having to cycle up and down to compensate for intermittent wind generation. While the authors make many references to this effect, it is not the primary focus of their paper.

III. Data

I combined hourly data for 2016 from multiple sources to create my dataset for this analysis. A total of 8,757 of the 8,784 hours in 2016 are used. Summary statistics for all variables are shown in Table 1.

TABLE 1: SUMMARY STATISTICS

Statistic	N	Mean	St. Dev.	Min	Max
Emissions (tons)	288,981	2,629.58	11,508.38	0	107,399.30
Generation (MW)	8,757	1,972.57	1,320.96	19.6	6,256.90
delta Generation (MW)	8,757	-0.072	246.498	-1,479.50	1,782.00
Load (MW)	8,757	90,318.17	17,711.38	57,750	152,178
Temperature (F)	8,757	56.027	18.972	-2.667	93.5

Source: Author's calculations

The emissions data originated from the EPA's air markets program, specifically the Cross-State Air Pollution NOx Annual Program (CSNOX), which regulates various pollutants. Through this program, the EPA tracks SO_2, NO_x, and CO_2 emissions from power plants with capacities of at least 25 megawatts (MW). The emissions data for Delaware was not available through this program, but were available via the Regional Greenhouse Gas Emissions Initiative (RGGI), which is a regional cap-and-trade program operated by several states in the Northeast. In both cases, the emissions data are reported hourly on a per-generator basis. Each state's data resides in its own table, which I had to separately download and combine. Since I am interested in aggregate emissions, I grouped the dataset by power plant, fuel type and emissions type. There are eleven fuel types and three emissions types, which are listed in Table 2. Thus, for each hour, the emissions data contains 33 readings.

Since the PJM territory is not perfectly aligned with state borders (Figure 1), I used a geospatial processing package in R to subset the emissions data to include only power plants within PJM, using their latitude and longitude coordinates.

The solar and wind power data comes from the from PJM Interconnection's Data Miner 2 web data tool and originate in a dataset of hourly generation by fuel type and utility. Since the electricity grid is essentially one system, it is reasonable to aggregate the generation data by hour without regard to the geographic location. Since all electrons are identical and the total generation must match the total load, it is not necessary to know which electrons are generated at fossil plants versus solar farms. What is important is the total amount and fuel type, in aggregate. After removing the utility groupings, I filtered for only solar and wind generation. Finally, I created a new variable that calculates hourly change in renewable generation.

The electricity load data also originates from PJM Interconnection's Data Miner 2 tool. Like the generation data, it is grouped by utility so I aggregated this by hour in the same way.

Hourly temperature readings come from the National Oceanic and Atmospheric Administration (NOAA) Local Climatological Data (LCD) dataset. Ideally, I would use local temperature data from many weather stations to give an exact reading by location, however given the scope of this paper, I simplified the process by using data from the weather station closest to the center of the PJM territory. I located the center using QGIS, an open source mapping software package, then found the closest weather station, which is Hart Field in Morgantown, WV.

IV. Model

A. Baseline Specification

The baseline model specification is a regression of emissions on the hourly change in renewable generation (first derivative), along with a vector of indicator variables for power plant type (***P***), a vector of emissions type indicator variables (***M***), and their interactions (see Table 2 for the full list of dummy variables).

TABLE 2: DUMMY VARIABLES

Fuel Types	*Emissions Types*
Coal	SO_2
Coal Refuse	NO_x
Diesel Oil	CO_2
Natural Gas - Boiler	
Natural Gas - Combined Cycle	
Natural Gas - Combustion Turbine	
Other Gas	
Other Oil	
Petroleum Coke	
Residual Oil	
Wood	

The formula for the baseline model is:

$$E_{it} = \beta_0 + \beta_1 \Delta G_t + \beta_2 \boldsymbol{P}_{it} + \beta_3 \boldsymbol{M}_{it} + \beta_4 \boldsymbol{P}_{it} \boldsymbol{M}_{it}^T + u_{it} \tag{1}$$

My null hypothesis, is $\beta_1 = 0$, meaning that there is no impact on emissions from variations in renewable energy generation.

A scatterplot of hourly change in generation against emissions appears in Figure 2. The plot shows a subset of 500 random samples drawn from the dataset. I represent the emissions on a log scale here because the values for CO_2 are significantly higher than for the other two pollutants. There are too many zero values in the emissions data to use a log scale in the regression.

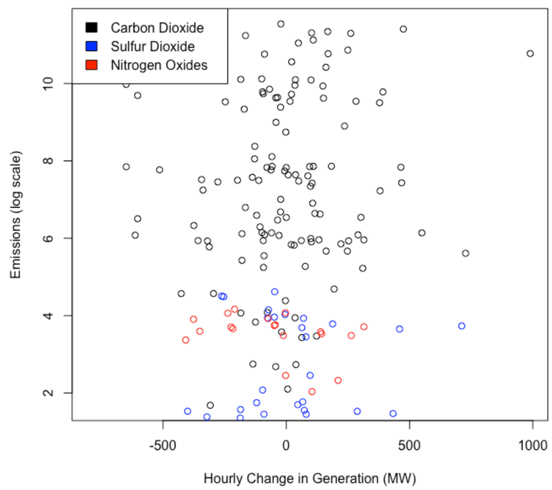

FIGURE 2: CHANGE IN GENERATION VS LOG OF EMISSIONS

Source: Author's calculations using random sample of n=500 from the dataset.

B. Robustness Specifications

The base model likely suffers from omitted variable bias. To examine robustness, I incrementally add in control variables (Equation 2). I include total electricity load (L), which is the amount of electricity consumed, because it follows a cyclical pattern throughout the day. As one might expect, load is higher during the daytime when people are awake and active. I would expect load to be positively correlated with both emissions and generation. As a result, it could be a source of positive bias in the base model.

Similarly, it is also possible that temperature (T) is correlated both with emissions and generation (since people need to use electricity for heat and air conditioning). In fact, it is likely that the relationship is nonlinear because electrically-powered heating and air conditioning are used most at extreme temperatures in both directions. I would expect the square of temperature to be positively correlated with emissions and with generation (at least, with solar generation). This suggests that the square of temperature could be causing positive bias in the base model.

In addition, since this dataset contains repeated measurements of the same entities over a year, I include entity fixed effects (α), where the entities are the power plant and emissions types. I cannot use time fixed effects because the renewable generation is assumed to be constant across entities for each time period. Including this would lead to perfect multicollinearity.

$$E_{it} = \alpha_i + \beta_1 \Delta G_t + \beta_2 \boldsymbol{P}_{it} + \beta_3 \boldsymbol{M}_{it} + \beta_4 \boldsymbol{P}_{it} \boldsymbol{M}_{it}^T + \beta_5 L_{it} + \beta_6 T_{it} + \beta_7 T_{it}^2 + u_{it} \tag{2}$$

Finally, as a further robustness check, I use the natural log of renewable generation as an alternative regressor to hourly change in generation (Equation 3). This allows the model to show how percent changes in generation affect emissions. Figure 3 shows a scatterplot of this relationship, again using a log scale for emissions.

$$E_{it} = \alpha_i + \beta_1 log(G_t) + \beta_2 \boldsymbol{P}_{it} + \beta_3 \boldsymbol{M} + \beta_4 \boldsymbol{P}_{it} \boldsymbol{M}_{it}^T + \beta_5 L_{it} + \beta_6 T_{it} + \beta_7 T_{it}^2 + u_{it} \tag{3}$$

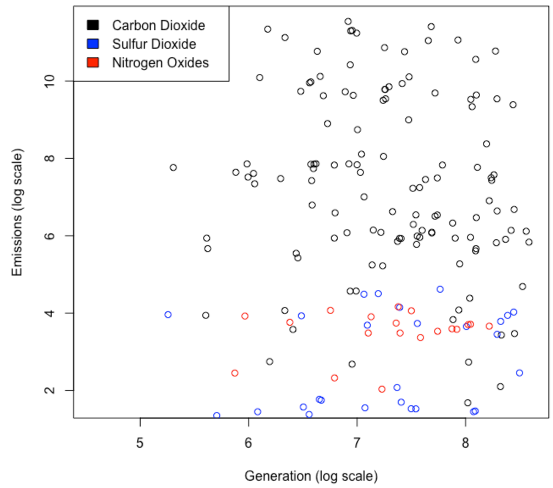

FIGURE 3: LOG OF GENERATION VS LOG OF EMISSIONS

Source: Author's calculations using random sample of n=500 from the dataset.

V. Model Results

A. Analysis of Regression Output

The results from the base specification regression indicate a positive relationship between oscillations in generation and emissions, which is not an expected result, but likely caused by omitted variable bias. This result is, however, statistically significant at the 1 percent confidence level. This surprising result may be particularly due to omission of electricity load (Table 2, column 1).

As expected, once we control for load, the sign of β_1 flips to negative (column 3). This result suggests that there are indeed additional emissions associated with fluctuations in renewable generation. After adding in controls for temperature, and its square, we see the inverse relationship remains, though is somewhat less. Column 4 indicates that a 1 MW per hour reduction in renewable generation is associated with an increase in emissions of 0.003 to 0.076

tons. For context, in 2016 the average U.S. home consumed about 0.9 megawatt hours (MWh) of electricity per month (1 MWh = 1 MW used for 1 hour) while the total of CO_2, NO_x, and SO_2

TABLE 3

	Dependent variable: Emissions (Tons)						
	OLS					*panel linear*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Hourly delta Gen. (MW)	0.079***		-0.050**	-0.039*		-0.039	
	(0.023)		(0.022)	(0.022)		(0.026)	
log(Gen)		-279.161***			-95.200***		-95.199
		(7.634)			(20.249)		(71.969)
Total Load (MW)			0.040***	0.036***	0.036***	0.036	0.036
			(0.001)	(0.001)	(0.001)	(0.025)	(0.024)
Temperature (F)				-17.218***	-16.660**	-17.219	-16.661
				(2.689)	(6.880)	(19.215)	(18.813)
Square of Temp. (F)				0.203***	0.186***	0.203	0.186
				(0.025)	(0.066)	(0.179)	(0.167)
Constant	322.548***	2,363.715***	-3,290.102***	-2,674.548***	29,398.590***		
	(1.460)	(55.905)	(46.794)	(106.817)	(383.858)		
Fuel / Emissions Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	288,981	288,981	288,981	288,981	288,981	288,981	288,981
R2	0.923	0.923	0.927	0.927	0.478	0.050	0.050
Adjusted R2	0.923	0.923	0.927	0.927	0.478	0.050	0.050
Residual Std. Error	3,195.025 (df = 288947)	3,186.871 (df = 288947)	3,115.661 (df = 288946)	3,114.565 (df = 288944)	8,314.404 (df = 288947)		
F Statistic	104,858.400*** (df = 33; 288947)	105,440.600*** (df = 33; 288947)	107,463.800*** (df = 34; 288946)	101,570.800*** (df = 36; 288944)	8,021.243*** (df = 33; 288947)	445.192*** (df = 34; 288914)	450.085*** (df = 34; 288914)
Note:	*p<0.1; **p<0.05; ***p<0.01						

Source: Author’s calculations.

emissions from electricity generation in the United States in 2016 was 2,129 million tons.[4] [5]

Thus, the observed effect is small. Put a different way, a one standard deviation decrease per hour in renewable generation is correlated with only a 0.006 to 0.16 percent standard deviation increase in emissions.

When adding in the entity fixed effects (column 6), we see no change in β_1, indicating robustness in the result, though including the fixed effects increases the standard error to the point of removing statistical significance at even the 10 percent level. We cannot reject the null hypothesis in this case.

Examining the alternate specification using the log of generation provides only slightly more powerful results. We see that the simple regression of emissions on $log(G)$ suffers from omitted variable bias (column 2). When adding in the full set of control variables (column 5) and then the entity fixed effects (column 7), we see stability in the results, though a noteworthy increase in the standard errors. The result without the fixed effects is significant at the one percent level, while the fixed effects model is not significant at any confidence level. Thus, when including entity fixed effects, we cannot reject the null hypothesis.

The regression in column 5 indicates that a 1 percent decrease in renewable generation is associated with an increase in emissions of 0.555 to 1.35 tons. In other words, a one standard deviation percent decrease in generation is associated with 1.19 to 2.89 percent of a standard deviation increase in emissions. While this result is larger than in the regression using change in hourly generation, it is still quite small in real-world terms.

[4] U.S. Energy Information Administration. 2017. "How much electricity does an American home use?." https://www.eia.gov/tools/faqs/faq.php?id=97&t=3 (accessed December 10, 2017).

[5] U.S. Energy Information Administration. "Table 9.1 Emissions from Energy Consumption at Conventional Power Plants and Combined Heat-and-Power Plants 2006 through 2016." https://www.eia.gov/electricity/annual/html/epa_09_01.html (accessed December 10, 2017).

B. Threats to Validity

While controlling for omitted variable bias clearly improved the model, there remain several threats to validity.

First, the emissions dataset only contains readings for power plants with output of at least 25 MW. While this covers the majority of power stations, it means that the sample selection contains bias. If smaller natural gas generators are the first to respond to sudden drops in renewable generation, the true β_1 might be more negative.

Second, the data contains hourly measurements, however solar and wind generation output fluctuate more frequently. Thus, the coarseness of the time resolution may mask some of the relationship between emissions and the oscillations in renewable generation. If I were able to include finer grain data readings in this analysis, I would expect β_1 to be larger in the negative direction.

Third, as discussed in section III, I made the assumption that renewable generation is constant at a given hour across the PJM region, and thus does not vary over entities in the regression. While I believe this is a reasonable assumption given the scope of this paper, in reality, transmission system congestion limits the ability for generation to uniformly serve all load within PJM, an effect known as Locational Constraint. In areas with excess load at a particular time, additional local fossil fuel plants may need to power up to meet the demand, even if there is surplus renewable energy generation elsewhere.[6] Therefore, my assumption of uniformity could introduce bias into the regression. This could be mitigated by weighting the renewable generation by the locational marginal price (LMP). LMP gives the actual wholesale

[6] PJM Interconnection. 2014. "PJM Manual 35: Definitions and Acronyms." http://pjm.com/~/media/documents/manuals/archive/m35/m35v23-definitions-and-acronyms-04-11-2014.ashx (accessed December 11, 2017)

electricity price at a particular geographic node, factoring in transmission constraints. This weighting would also cause the amount of renewable generation to vary across entities, allowing time fixed effects to be included in the regression.

VI. Conclusion

While I achieved a measure of consistency in the analysis results, and controlled for some omitted variable bias, the regressions indicate that the real impact of renewable generation fluctuations on emissions is small. This is good news for wind and solar energy's ability to reduce greenhouse gas and other pollutant discharges associated with fossil fuel plants. That said, lawmakers should account for these excess emissions when calculating the expected benefits of pro clean energy policies, like renewable portfolio standards. Going forward, a useful follow-up to this paper would be a comparison of the overall emissions offset by renewable energy with the additional emissions created by fluctuations in solar and wind.

Datasets

National Oceanic and Atmospheric Administration. 2016. Hourly Weather for Morgantown Hart Field, WV. Local Climatological Data (LCD) Dataset. https://www.ncdc.noaa.gov/cdo-web/datatools/lcd (accessed November 19, 2017).

PJM Interconnection. 2016. Hourly Electricity Generation by Fuel Type. PJM Data Miner 2. http://dataminer2.pjm.com/feed/hrl_load_estimated/definition (accessed November 18, 2017).

PJM Interconnection. 2016. Hourly Estimated Electricity Load. PJM Data Miner 2. http://dataminer2.pjm.com/feed/hrl_load_estimated/definition (accessed November 18, 2017).

PJM Interconnection. 2016. PJM Zones GIS shapefile. (Received by email January 31, 2017).

U.S. Environmental Protection Agency. 2016**.** Unit-level hourly NOx, SO2, CO2 emissions, gross load and unit attributes for 2016. Cross-State Air Pollution NOx Annual Program (CSNOX). https://ampd.epa.gov/ampd/ (accessed November 5, 2017).

U.S. Environmental Protection Agency. 2016. Unit-level hourly NOx, SO2, CO2 emissions, gross load and unit attributes for 2016. Regional Greenhouse Gas Emissions Initiative (RGGI). https://ampd.epa.gov/ampd/ (accessed November 9, 2017).

References

Gonzalez-Salazar, Miguel Angel, Trevor Kirsten, Lubos Prchlik. 2017. “Review of the operational flexibility and emissions of gas- and coal-fired power plants in a future with growing renewables.” Renewable and Sustainable Energy Reviews. (in press. Accessed proof online).

Kaffine, Daniel T., Brannin J. McBee, and Jozef Lieskovsky. 2013. “Emissions Savings from Wind Power Generation in Texas.” The Energy Journal. 2013, 34(1): 155-175.

Katzenstein, Warren and Jay Apt. 2009. “Air Emissions Due To Wind And Solar Power.” Environmental Science & Technology. 2009 43 (2), 253-258

U.S. Environmental Protection Agency. 2016. “Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990-2014.” Executive Summary. https://www.epa.gov/ghgemissions/inventory-us-greenhouse-gas-emissions-and-sinks-1990-2014