Identifying Coal-Fired Power Plants for Early Retirement, Codebook amd Data Collection Protocol

Nada Maamoun [a, b]
Helmut-Schmidt University
University of Hamburg

Ryan Kennedy [c]
University of Houston

Xiaomeng Jin [d]
Columbia University

Johannes Urpelainen [e,*]
Johns Hopkins SAIS

April 21, 2020

General comments

This document contains the data collection protocol and codebook for the dataset used in creating the global retirement index for the coal-fired power plants. We use data on 2143 operating coal-fired plants worldwide to rank plants for retirement. We use data collected and aggregated by the authors in 2017 covering coal-fired plants and their units worldwide; the data is derived from five main sources that provide information on the coal plants globally: The Global Coal Plant Tracker, Climate Analytics, Enipedia, The Global Energy observatory, and Carbon Monitoring for Action (CARMA). Data on coal-fired plants includes plant capacity, emissions per plant and plant location. We also estimate where the air pollutants of the plant end up, estimated by the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model. Based on this data, we formulate the retirement index upon which plants are ranked by order of how much pollution damage they cause and thus the urgency of their retirement.

Data collection and Index formation

Data sources

The data on coal-fired plants covers essential and up-to-date information about global coal-fired plants and their units. It includes the plants and units' names and their exact locations (longitude and latitude), the plants and their units' capacity in MW, the current status of the plant and its units as of 2018 (ranging from announced till retired) and date of commission and decommission (in case of retirement or closure), as well as annual carbon dioxide emissions.

[a]Department of Economics, Helmut-Schmidt University, Holstenhofweg 85, H1 Rm.2175, Hamburg, 22043, Germany.

[b]Institute for Law and Economics, University of Hamburg, Johnsallee, 35, 20148, Hamburg, Germany.

[c]Department of Political Science, University of Houston, 3551 Cullen Blvd, Rm. 447 PGH, Houston, TX 77204-3011.

[d]Department of Earth and Environmental Sciences and Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, 10964

[e]Corresponding author. Address: Rome Building, 4th Floor. 1619 Massachusetts Avenue, NW. Washington, DC 20036, USA. Tel: +1-202-779-1352. Email: JohannesU@jhu.edu

- The data on coal-fired plants was collected by the authors from 6 main sources:
  - *The Global Coal Plant Tracker* (https://endcoal.org/global-coal-plant-tracker/) produced by CoalSwarm (http://coalswarm.org), a wiki that has more than 9,000 articles about coal housed on SourceWatch[a]. The Global Coal Plant Tracker, provides information on global coal plants that are larger than 30 MW of capacity [Endcoal.org, 2018].
  - *Climate Analytics* (https://climateanalytics.org) is a non profit institute for climate science and policy based in Berlin aiming at advancing scientific knowledge in the area of climate change and accordingly provide support and capacity building to stakeholders.
  - *Enipedia* was started by the Energy and Industry Group at the department of Technology, Policy and Management at the Delft University of Technology in the Netherlands. It explores the application of wikis in issues related to energy and industry [Davis et al., 2015].
  - *The Global Energy observatory* (http://globalenergyobservatory.org/) is a project providing a set of free interactive databases, aiming at promoting the global understanding of energy systems, their emissions impacts, as well as accelerating the transition to globally affordable carbon-neutral energy systems.
  - *Carbon Monitoring for Action (CARMA)* (http://carma.org/) is a database that includes information on more than 60,000 power plant and their carbon emission as well as 20,000 power companies globally. It is produced and financed by *Confronting Climate Change Initiative* at the Center for Global Development, an independent and objective think tank in Washington, DC.
  - *World resources institute (WRI)* provides a comprehensive dataset on the global power plants. The data is collected from governmental data as well as independent sources and provides information on technical and operation characteristics of all power plants [Byers et al., 2019]. We are mainly interested in the generation data for coal-fired plants. The data covers more than 2000 coal-fired power plant, only 1752 plants of those matched plants in our dataset; however, the missing plants do not exceed 8% of the total global coal capacity.

- *Center for International Earth Science Information Network (CIESEN)* provides gridded data on population count, we use the fourth version of the the gridded population of the world data (GPWv4) : https://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11. It is a globally integrated data on population count per grid cell (number of people living in a grid cell) with a resolution of 2.5 arc-min (around 5 km at the equator) based on the 2010 round of the Population and Housing Census that took place between 2005 and 2014 [CIESIN, 2017].

Population-weighted damage

We use the HYSPLIT model to estimate where the emissions of the coal-fired plants end up and how frequently they end up in one area in order to determine the degree of harm they cause to those areas and the damage attributed to each plant.

- We run a forward trajectory analysis using historical wind data for each air pollutant emitted from coal-fired plants worldwide at its corresponding latitude/longitude point.

[a]Sourcewatch is an open-source encyclopedia sponsored by the Center for Media and Democracy

- We set a standard number of simulated particles for all plants and track their expected dispersion over the course of a year. We do not have data on the exact emissions each plant produces,so we weigh the standard number of air pollutants assigned to each plant based on plant's capacity. Larger plants are expected to produce more emissions (assuming that plants are operating at capacity).

- Based on the location of the plant and the wind speed and direction, the model provides an estimation of where the pollutants of each plant end up and, tracking the sum of simulated particles in an area, the frequency of where pollutants end up. Running the HYSPLIT model, does not give directly the frequency. Each run of HYSPLIT model only gives one trajectory, then the frequency is calculated by combining hundreds of runs, summing how many times the trajectory passes through a given location. For this paper, we ran the models four times a day, every other 8 days for 2013. We assume the lifetime for major air pollutants is around 4 days, and therefore the trajectory was computed for 96 hours. Thus we have about 180 trajectories generated for each plant. In total, we have about $180 \times 3500 = 630,000$ trajectories generated. Given our focus is just on operating plants, for this paper the relevant trajectories generated are $180 \times 2143 = 385{,}740$ trajectories.

- We combine the data on the count of the simulated particles and their end location as estimated by the HYSPLIT with the gridded population count data to identify the areas (or population) likely to be affected by the emissions from each plant.

Using the Population data and the HYSPLIT model's out out we are able to compute the population exposed to air pollution from each power plant using the following equation.

$$\text{PWD}_{ij} = \frac{\text{capacity}_i}{\max_{1 \le i \le n} \text{capacity}_n} \times \sum_{j=1}^{N} \frac{\text{population}_{ij}}{\text{N}} \tag{1}$$

where i represents the coal plant emitting j polluting particles and PWD_{ij} measures the average population exposed to j pollutants emitted by plant i based on the HYSPLIT model's estimated end location of the j polluting particles and the population count from the GPWv4 data. $Capacity_i$ represents the capacity of plant i in MW, $capacity_n$ is the maximum capacity of a currently operating plant in MW. Thus, the first term in equation 1 measures the plant's weighted capacity relative to the largest plant and its value ranges from 0 to 1. $Population_{ij}$ represents the population exposed to j pollutants based on the GPWv4 data and the estimated location of plant i's pollutants by HYSPLIT model. N is the total number of polluting particles emitted per plant. So the population-weighted damage of plant i is the average population exposure to the j pollutants emitted by plant i multiplied by the weighted capacity of plant i. In cases where some of the air pollutants of plant i are estimated to end up in the oceans or in unpopulated areas (no human life), the corresponding PWD_{ij} for these pollutants is zero, as $population_{ij}$ would be zero.

Index formation

To formulate the index presented in the paper, the 3 relevant variables: Age, annual CO_2 emissions, and PWD were standardized and assigned equal weights.

Variable name	Description
plant	Plant name
country	country name
sum_capacity	capacity per plant measured in MW
age_std	Age of power plant as of 2018, standardized
co2_std	annual CO_2 emissions in million tons, standardized
pwd_std	Population weighted damage, standardized
gpwv4popula	average population exposed per plant
index	index value per plant

Codebook

References

Byers, L., Friedrich, J., Hennig, R., Kressig, A., Li, X., McCormick, C., and Malaguzzi Valeri, L. (2019). "a global database of power plants.". Technical Note by the WRI on " A Global Dataset of Power Plants", June 2019.

CIESIN (2017). Center for international earth science information network (ciesin), Columbia university. 2017. documentation for the gridded population of the world, version 4 (gpwv4), revision 10 data sets. palisades ny: Nasa socioeconomic data and applications center (sedac). Accessed on May 7,2018.

Davis, C., Chmieliauskas, A., Dijkema, G., and Nikolic, I. (2015). Enipedia, http://enipedia.tudelft.nl, energy & industry group, faculty of technology, policy and management, tu delft, delft, the netherlands.

Endcoal.org (2018). New coal plants by country, 2006-2017 (mw). Data on New coal plants (units 30 MW and larger) accessed from endcoal.org on June 4, 2018. Sources: CoalSwarm and Global Plant Tracker.