

Methodology for Power Generator Datasets

Power Transformation Lab, UCSD

Updated: July 2026

Overview

This document describes the methodology used to build the Power Transformation Lab's processed power generator datasets. The workflow is based primarily on Global Energy Monitor (GEM) power plant data and includes data collection, processing, labeling, and province name standardization. The resulting datasets provide standardized and consistently processed generator information for China, Indonesia, and Vietnam, supporting power system analysis, statistics, and energy system modeling.

Data Sources

- [Global Energy Monitor](#)

Global Energy Monitor (GEM) is a global, independent nonprofit that maps and tracks energy infrastructures worldwide, from power stations and coal mines to steel plants and shipping terminals.

- [Global Integrated Power Tracker](#)

Global Integrated Power Tracker is a project under GEM. It is a global dataset of power plants that tracks the energy transition across all major generation technologies.

- [Global Coal Plant Tracker](#)

Global Coal Plant Tracker is a project under GEM. It catalogues coal-fired power units worldwide. The tracker monitors every unit with a capacity of 30+ megawatts, including operating units, new units proposed since 2010, and units retired since 2000.

Data Collection

Collect the latest version of the Global Integrated Power Tracker data as a base, and supplement it with the latest version of the Global Coal Plant Tracker data.

Collected datasets:

- Global-Integrated-Power-March-2026-II.xlsx
- Global-Coal-Plant-Tracker-January-2026.xlsx

Data Processing

This dataset processing workflow has three steps and focuses on exactly three countries: China, Indonesia, and Vietnam. The sequence is: create a base generator dataset from source files, add CHP and captive labels to create a labeled dataset, then split the labeled dataset into three output files, and standardize province names for analysis and modeling.

Step 1 — Build base generator dataset

Read two source files (an integrated power plant generator file and a coal unit file), filter records to China, Indonesia, and Vietnam, merge records by unit identifier, and enrich coal units with key performance and emissions fields.

Step 2 — Add CHP and captive labels

Read the base generator dataset and a reference file used for CHP identification, preserve valid existing labels, add CHP labels using rule-based and name-matching logic, generate captive labels, and remove columns not relevant to the intended analysis and modeling.

Step 3 — Split into country datasets

Read the labeled generator dataset and split records by country. Province names are subsequently standardized to ensure geographic consistency. Finally, export the three country datasets (China, Indonesia, Vietnam) in text formats. These three country files are the final processed outputs used for province-level statistics, analysis, and modeling.

Data Coverage

The table below describes the source coverage of the underlying GEM trackers by technology. The Global Integrated Power Tracker used in this workflow (see Data Collection) already consolidates all eight technology groups shown below; the Global Coal Plant Tracker is used separately to supplement coal-unit detail.

Power Trackers	Latest release	Summary Tables include capacity above specified thresholds
Coal	Jan-26	≥30 megawatts (MW) or larger within the following parameters: all operating/mothballed units; all

Power Trackers	Latest release	Summary Tables include capacity above specified thresholds
		units retired since 2000; and all units proposed since 2010.
Oil and gas	Jan-26	≥50MW; ≥20MW in the European Union and the United Kingdom
Utility-scale solar	Feb-26	≥1MW for operating projects; announced, pre-construction, construction, and shelved projects with capacities greater than 20 MW. Note, the Global Solar Power Tracker also includes nationally aggregated distributed (<1 MW) solar capacities for 31 countries/areas, with expanded geographical and temporal coverage planned in future data releases. Data available in Global Solar Power Tracker summary tables.
Wind	Feb-26	≥10MW (≥1MW data are provided for some countries when the data were readily available.)
Nuclear	Sep-25	n.a. no threshold applied
Hydropower	Mar-26	≥30MW (≥10MW data are provided for some countries when the data were readily available.)
Bioenergy	Sep-25	≥20MW (≥10MW data are provided for some countries when the data were readily available.)
Geothermal	Mar-26	≥1MW

Data columns

The table below describes the column definitions of the processed datasets. Most column definitions are inherited from the Global Integrated Power Tracker, with two exceptions. First, the Captive column is derived by consolidating three related source columns into a single binary field (yes or blank). Second, Capacity factor, Heat rate (Btu/kWh), and Emission factor (kg CO₂/TJ) are sourced from the Global Coal Plant Tracker.

Columns not relevant to the intended analysis and modeling have been removed, and some definitions have been simplified for clarity. For the complete set of original columns and definitions, refer to the source datasets listed in the Data Sources section.

Column	Definition
Type	One of eight power technology groups (coal, oil/gas, utility-scale solar, wind, hydropower, nuclear, bioenergy, geothermal)
Country/area	The Country/area column presents energy data within various economic contexts, and the geographical unit does not strictly follow political boundaries
Subnational unit (state, province)	Subnational unit where project is located, usually state or province
Plant / Project name	Commonly used name
Unit / Phase name	Name or number of phase in plant/project
Plant / Project name (local)	Commonly used name in the locally used script
Plant / Project name (other)	Alternative names, including the name in local script
Capacity (MW)	Total nameplate capacity, or best available data, of the project, in megawatts
Status	One of 8 status categories
Start year	Year the project was or will be commissioned
Retired year	Year the project was or will be taken offline
Technology	The type of technology used in the project
Capacity rating (solar only)	The reported capacity rating, unknown selected if not provided
Associated storage	Value of 'yes' if a project phase has existing or planned co-located storage. Blank value otherwise. Currently data for solar and wind only.

Column	Definition
Fuel (combustion only)	The type of fuel used in the project
CHP	Value of 'yes' if a project phase has combined heat and power/cogeneration, 'no' if not, 'not found' or blank if unknown or no data found.
Captive	Value of 'yes' if a project phase includes an on-site captive power plant (primarily for internal consumption), blank if unknown or no data found.
Capacity factor	Country-level average based on 2025 GEM operating coal capacity and Jan 2026 Ember generation data. Retired/cancelled plants are set to 0. Indonesia includes grid plants only (excludes captive). Excludes units <30 MW.
Heat rate (Btu per kWh)	Heat rate is based on type of combustion technology, with performance penalties for older and smaller units.
Emission factor (kg of CO ₂ per TJ)	Based on type of coal using the 2006 IPCC Guidelines for National Greenhouse Inventories, Table 2.2, page 2.16.
Latitude	Latitude of the project, assumed WGS84 (Google Maps)
Longitude	Longitude of the project, assumed WGS84 (Google Maps)