

# DATA MERGING: MAPPING OVERNIGHT COSTS TO INSTALLED CAPACITIES

ANDREW BLOHM

## 1. MERGING EIA OVERNIGHT COST ESTIMATES WITH FORM 860 GENERATOR LEVEL DATA

In this document we discuss how we merged information from the EIA overnight cost assumptions for electricity production plants with generator level data from the EIA Form-860 survey. We discuss the time dimension of the overnight costs in subsequent chapters. The result of this work is a dataset to be used in the multinomial logit.

**1.1. Overnight cost data.** The overnight cost data used in this analysis comes from the US Energy Information Administration (EIA) annual energy outlook (AEO) series, specifically the underlying assumptions on characteristics of new and existing power plants. Each annual energy outlook estimates the characteristics of new generating plants, including operational characteristics and cost assumptions. We use the overnight capital cost assumptions from each AEO to approximate the investment cost for each technology in each year of the analysis. The earliest AEO for which the overnight cost is available is the 1997 AEO; having been published in each subsequent year. However, as new technologies emerge and existing technologies evolve the categories within the overnight cost database have necessarily changed. Therefore, the database does not have consistent categories over time.

**1.2. EIA Form EIA-860 (plant level electricity sector data).** The EIA uses the survey form EIA-860 to collect generator level data on existing and planned power plants with an installed capacity greater than 1 MW. Form EIA-860 compiles information on the utility, plant, and generator for each of the generators, including information on environmental regulation compliance, ownership structure, etc. Of particular interest to the study is the plant level dataset, which amongst other information contains capacity, installation date, operating status, primary fuel, and technology use at a disaggregated scale not found elsewhere. However, the technology categories available in the Form EIA-860 do not readily map onto the overnight cost technology categories. In the next section, we discuss the mappings that we employ and the reasoning behind those.

The units of the installed summer capacity are not consistent across the years of the survey with some misspecified. For the EIA Form-860 years 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, and 2014 the summer capacity of each generator is in MW (and correctly identified as such). However, for EIA Form-860 years 1998, 1999, and 2000 the summer capacity is in kW despite the accompanying documentation stating the units as MW. Only EIA Form-860

for the years 1990, 1991, 1992, 1993, 1994, 1995, 1996, and 1997 has the summer capacity correctly listed as kW. To address the issue we convert all values of the installed summer capacity to MW.

The data also includes negative values in the summer capacity field. We change the negative values for utility code 13756, plant code 996, generator unit 9C and 9B from negative to positive, as it looks like it might have been an input error. We do the same for utility code 15263, plant code 6545, generator 2 for the same reason.

**1.3. Mappings.** There are 37 prime movers (i.e., power production technologies) and 78 primary fuels in the Form EIA-860 database, which together form 190 unique, pair-wise combinations that need to be mapped to the overnight cost database. The issue is that the overnight cost database has between 15 and 20 overnight cost categories from which to choose; the exact number of categories fluctuates year-to-year. In this section, we identify the mapping selections that we made and expound on the assumptions underlying them. We work from the more obvious to the less clear mappings.

The categorization below results in 595 unassigned generators with a total installed summer capacity of 0.0139 TW. The uncategorized generators are from two prime mover categories: turbines used in binary cycles (BT) and steam turbines (ST) with their primary fuel either purchased steam (PUR), steam (STM), waste wood liquids (WDL) or waste heat (WH).

It is important to remember that an individual generator might show up in each year of the EIA Form-860, which means that it could be counted 24 times. For this reason, the total installed summer capacity that we calculate for each prime mover (i.e., technology) could be extremely high.

**1.3.1. Missing or incomplete records.** There are a few categories that either don't have enough information for assignment (or would require too much effort to do so). First, we ignore the prime mover designation CG as it is undefined by the survey (3 natural gas fired CG plants with an installed summer capacity of 0.33 TW over the period of analysis).

Second, we ignore the prime mover designation other (OT) and primary fuel other (OTH) because the cost of determining how to allocate these resources is too high, as compared to the benefit (i.e., it would require reading the notes for approximately 575 generators with an installed capacity of 0.069 TW over the period of analysis).

Third, we ignore any generator without an identified primary fuel, except in the case of hydro power technology, where the choice of fuel is obvious. There are 6 non-hydro generators with an installed summer capacity of 0.096 TW that do not presently have a primary fuel listed.

Fourth, we ignore the primary fuel code WOC, which is not defined in any of the supporting literature. There are 12 generators with a total installed summer capacity of 0.30 TW with a primary fuel WOC.

Finally, we ignore the multi-fueled generators (MF) given that the database doesn't indicate the set of fuels that they are using. There are 21 generators with a total installed summer capacity of 0.635 TW.

In total we drop from the dataset 617 units, with an installed summer capacity of 1.43 TW or approximately 0.0001% of total installed summer capacity for all generators across the period of analysis.

1.3.2. *Hydropower.* The overnight cost dataset includes one cost estimate for conventional hydropower. We assume that all of the production technologies with water (WAT) listed as the primary fuel can be mapped to the hydropower investment cost. These include prime mover designations hydraulic turbines - conventional (HC), hydraulic turbine - pipeline (HL), hydraulic turbine - reversible (HR), hydraulic turbine - conventional (HY), and hydraulic turbine reversible - pumped storage (PS). We also map these production technologies to the hydro power investment cost even in cases where the primary fuel column entry is missing. There are 90,163 hydropower generators in the dataset with a total installed summer capacity of approximately 2.433 TW for the period of analysis.

1.3.3. *Geothermal.* The overnight cost dataset includes one cost estimate for geothermal power plants. We assign the geothermal overnight cost to any generator with geothermal (GEO) or geothermal steam (GST) as a primary fuel. This includes prime mover designations: turbines used in binary cycles (BT), other (OT), steam turbine - geothermal (GE), and steam turbine - boiler (ST). This categorization moves one plant type from other (OT) to geothermal. There are 3,269 geothermal generators in the dataset with a total installed summer capacity of approximately 0.0505 TW over the period of analysis.

1.3.4. *Photovoltaic.* The overnight cost dataset includes one cost estimate for photovoltaic power plants. We assign the photovoltaic overnight cost to both prime mover designations: photovoltaic (PV) and photovoltaic (SP) in the Form EIA-860 dataset. There are 3,460 photovoltaic generators in the dataset with a total installed summer capacity of approximately 0.0185 TW over the period of analysis.

1.3.5. *Solar thermal.* The overnight cost dataset includes one cost estimate for solar thermal power plants. We assign the solar thermal overnight cost to any generator with sun (SUN) as the primary fuel and a prime mover designation that is not photovoltaic (PV or SP). This includes other (OT), steam turbine - solar (SS), and concentrated solar power energy storage (CP) generators. This categorization moves one plant type from prime mover other (OT) to solar thermal. There are 170 solar thermal generators in the dataset with a total installed summer capacity of approximately 0.008 TW over the period of analysis.

1.3.6. *Wind.* The overnight cost dataset has two cost estimates for wind (i.e., onshore and offshore). However, we do not have this level of specificity in the Form EIA-860 dataset but we do know that most of the installed capacity in the US has been in onshore wind capacity. Therefore, we assign the onshore wind overnight cost to the wind turbine (WT) generator. There are 7,732 wind generators in the dataset with a total installed summer capacity of approximately 0.3845 TW over the period of analysis.

1.3.7. *Nuclear.* The overnight cost dataset includes one cost estimate for nuclear power plants. We assign the nuclear power overnight cost to any generator using nuclear (NUC), uranium (UR), plutonium (PT), or thorium (TH) as the primary fuel. This includes steam turbine - boiling water nuclear reactor (NB), steam turbine - graphite nuclear reactor (NG), steam turbine - high-temperature gas-cooled nuclear reactor (NH), and steam turbine - pressurized water nuclear reactor (NP). There are 2,450 nuclear generators in the dataset with a total installed summer capacity of approximately 2.2813 TW.

1.3.8. *Fuel cells.* The overnight cost dataset has one cost estimate for fuel cells. We assign the overnight cost estimate for fuel cells to fuel cell - electrochemical (FC). In the Form EIA-860 dataset the fuel cell plants use landfill gas (LFG), natural gas (NG), and other biomass gases (OBG) as their primary fuel. There are 273 fuel cell generators in the dataset with a total installed summer capacity of 0.000284 TW.

1.3.9. *Biomass.* The overnight cost dataset includes one cost estimate for biomass power plants. The DOE assumes that woody residues, round wood and woody energy crops, municipal solid wastes, wet herbaceous residues, and dry herbaceous residues and energy crops are all biomass feedstocks Source: <http://www.energy.gov/eere/bioenergy/biomass-feedstocks>. However, given that MSW is its own category in the overnight cost we assign that to its own overnight cost. We assign all non-combined cycle plants with the following primary fuels to the biomass overnight cost, biomass generic (BIO), other biomass solids (OBS); refuse, bagasse and all other non-wood waste (REF); wood and wood waste (WD), and wood waste solids (WDS). This categorization moves one plant from the internal combustion category to the biomass category and, two plant types from other to biomass (i.e., (OT, REF) and (OT, WDS)). There are 3,113 biomass generators in the dataset with a total installed summer capacity of 0.0521 TW.

1.3.10. *Municipal solid waste.* The overnight cost dataset includes one cost estimate for municipal solid waste power plants. We assign all plants with municipal solid waste (MSW) as the primary fuel to that overnight cost. There are 1,383 municipal solid waste generators in the dataset with a total installed summer capacity of 0.03196 TW.

1.3.11. *Distributed generation.* The overnight cost dataset includes two cost estimates for distributed generation: peak and non peak. We assign all generators with megawatt hour (MWH) as the primary fuel to distributed generation (i.e., battery energy storage (BA) and flywheel energy storage (FW)). In addition, we assign generators with compressed energy storage (CE) prime mover designation. There are 81 distributed generators in the dataset with a total installed summer capacity of 0.002875 TW.

1.3.12. *Steam turbines.* Gas-oil steam turbines exist as a category for the 1997 to 2000 AEOs. We assign all steam turbine plants that use either, any form of natural gas or fuel oil to this overnight cost estimate. On the natural gas side this includes landfill gas (LFG), synthetic natural gas (SNG), other gas (OG), natural gas (NG), liquified propane gas (LPG), blast furnace gas (BFG), black liquor (BLQ) (gasified before use), and other biomass gases (OBG). We assume that some data entry mistakes were made in the process of inputting the survey in that BL should have been BLQ. This change affects 7 generators. On the oil side this includes black liquor (BLQ), distillate fuel oil (DFO), fuel oil 1-6 (FO1 - FO6), residual fuel oil (RFO), oil-other and waste oil (WO), coal synfuel (SC), residual fuel oil (RFO), and sludge waste (SLW). There are a total of 25,341 steam turbine generators in the dataset with a total installed summer capacity of 3.2086 TW.

1.3.13. *Combined cycle generators.* The EIA has overnight cost estimates for conventional and advanced gas-oil combined cycle. However, the EIA Form-860 dataset does not include the information necessary to delineate between the two alternatives.<sup>1</sup> The prime mover designations in the Form-860 dataset include the combined cycle - total unit (CC), combined cycle combustion turbine part (CT), combined cycle single shaft - gas turbine and steam turbine (CS), combined cycle steam turbine - waste heat boiler (CW), and the combined cycle steam turbine with supplemental firing (CA). The cost estimates in the overnight cost database are for an oil and gas unit however, some of the CT, CC, CA, use purchased steam (PUR), bituminous coal (BIT), or wood and wood waste solids (WDS) as their primary fuel. We don't have enough information to assign these plants to anything but the combined cycle overnight cost. There are a total of 24,297 combined cycle generators in the dataset with a total installed summer capacity of 2.6571 TW.

1.3.14. *Combustion turbines.* The EIA overnight cost estimates for combustion turbines once again include both an advanced and conventional option. One option is to use the conventional turbine overnight cost for all of the combustion turbine plants in the Form EIA-860 dataset. This includes combustion (gas) turbine (GT), internal combustion (IC), and jet engine (JE). There are 156,670 combustion turbine generators in the dataset with a total installed summer capacity of 2.575 TW. For the time being we assume that all of them are conventional combustion turbines. At some future point (with more time) we might be able to use the average size of an advanced versus conventional combustion turbine to improve upon this assumption.

1.3.15. *Integrated coal gasification generators.* The EIA overnight cost dataset includes two overnight cost estimates for the integrated coal gasification plant; with and without carbon capture and storage (CCS). We assume that all IGCC plants built to date do not include the CCS option. Therefore, we assign the overnight cost for the IGCC without CCS to all IGCC plants in the Form EIA-860 dataset. There are a total of 13 IGCC generators in the dataset with a total installed summer capacity of 0.00203 TW.

1.3.16. *Coal generators (not IGCC).* The EIA overnight cost dataset includes a few coal power plant options. In the 1997 AEO there is one coal power plant option; pulverized coal power plants. In the 1999 AEO there is a scrubbed coal new option. In the 2001 AEO the coal option is the conventional pulverized coal power plant. Despite the changes in the name of the plant each of these plants has similar characteristics (i.e., cost, etc.). Therefore, we assign all coal fired plants bituminous coal (BIT), subbituminous coal (SUB), lignite (LIG), anthracite (ANT), coal generic (COL), refined coal (RC), and waste coal (WC), which are not IGCC (IG), combined cycle, or combustion turbine power plants to the coal power plant overnight cost (of varying name). Online resources suggest that petroleum coke is used primarily in steam turbine, coal-fired power plants so we assign the petroleum coke (PC) primary fuel plants to the coal plants as well. Online resources suggest that tire derived fuel (TDF) are primarily used as an additive in coal plants. There are a total of 33,126 coal generators in the dataset with a total installed summer capacity of 7.5244 TW.

---

<sup>1</sup>There might be an average size difference that could be used to assign the Form-860 combined cycle plants to the advanced and conventional categories.