Expanding the Inventory of Emissions in the NETL Power Plant Flexible Model

Matt Jamieson – KeyLogic Systems, Inc.
Greg Cooney - KeyLogic Systems, Inc.
Timothy Skone – National Energy Technology Laboratory
ISSST 2018 – Buffalo, NY
June 27, 2018
Disclaimer and Attribution

DISCLAIMER

"This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof."

Attribution

KeyLogic Systems, Inc.’s contributions to this work were funded by the National Energy Technology Laboratory under the Mission Execution and Strategic Analysis contract (DE-FE0025912) for support services.
About the Power Plant Flexible Model (PPFM)

- Spreadsheet model for
  - pulverized coal
  - circulating fluidized bed power plants
  - natural gas combined cycle
  - solid oxide fuel cell power plants

- Reduced-order model allowing changing of coal characteristics and pollution control equipment configuration

- Emissions limited to those available in NETL techno-economic assessments
  - CO₂, SO₂, Hg, NOₓ, Particulate

https://netl.doe.gov/research/energy-analysis/search-publications/vuedetails?id=785
Power Plant Flexible Model Uses
Co-fire biomass vs. CO₂ emissions and net plant power

• PPFM intended as a tool to quickly assess changes in equipment or feedstock

• Example: Can relatively quickly assess impacts of co-firing varying amounts of biomass while maintaining sulfur emissions
  • 98.0% to 97.6% removal rate for SO₂ (Wet FGD) at 0.327 kg SO₂/MWh net
Moving Beyond GHGs, CAPs, and Water Use

- Past focus of LCAs have been on greenhouse gases (GHGs), criteria air pollutants (CAPs), and water use
- Expanding inventory across all NETL models to support broader analyses
  - Impact analysis via EPA TRACI 2.1
- As an input to other models (i.e., CO$_2$-enhanced oil recovery (CO$_2$-EOR), PPFM emissions inventory needed to be expanded
Goals for PPFM Inventory Expansion

• Emissions that respond to changing plant configuration
  • Boiler type (for coal): sub-critical vs. supercritical
  • Air pollution control: equipment type and operation
  • Coal type: bituminous vs. sub-bituminous vs. lignite

• Use publicly available data
  • National Emissions Inventory (NEI), Energy Information Administration (EIA) power plant data, etc.
Method Overview

Combined Dataset

NEI
EIA 923
EIA 860
ABB Velocity

75% to Training
Regression Using LassoCV
Test Regression Model Against Test Set
Keep Models with $R^2 \geq 0.5$

25% to Testing

Repeat 1,000 Times for Each Emission
Source Data

- 2011NEI: Annual emission rate for each plant
- Combined with data below using crosswalk provided by Eastern Research Group (EPA Plant ID to NEI EIS Facility ID)

### Coal Plants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Plant Unit Capacity (MW)</td>
<td>EIA 860</td>
</tr>
<tr>
<td>Power Plant Unit Generation (MWh)</td>
<td>EIA 923</td>
</tr>
<tr>
<td>Control Equipment (Boolean)</td>
<td>ABB Velocity</td>
</tr>
<tr>
<td>Supercritical (Boolean)</td>
<td>ABB Velocity</td>
</tr>
<tr>
<td>Fraction of Each Coal Type (BIT/SUB/LIG)</td>
<td>EIA 923</td>
</tr>
<tr>
<td>Coal Quality (heat/sulfur/ash content)</td>
<td>EIA 923</td>
</tr>
<tr>
<td>Heat Rate (BTU/kWh)</td>
<td>EIA 923 (calculated)</td>
</tr>
</tbody>
</table>

### Natural Gas Plants

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power Plant Parameters</td>
<td>Data Source</td>
</tr>
<tr>
<td>Power Plant Capacity (MW)</td>
<td>EIA 923</td>
</tr>
<tr>
<td>Power Plant Generation (MWh)</td>
<td>EIA 923</td>
</tr>
<tr>
<td>Control Equipment (Boolean by type)</td>
<td>ABB Velocity</td>
</tr>
<tr>
<td>Heat Rate (BTU/kWh)</td>
<td>EIA 923 (calculated)</td>
</tr>
</tbody>
</table>

*Water/Steam, Catalytic, Ammonia, Overfire, and Low NOx
Parameters for Regression Analysis

- The parameters were chosen based on configuration data available
- Some consideration given for options available within PPFM

Coal – 14 available parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nameplate Capacity</td>
<td>Bituminous Coal</td>
</tr>
<tr>
<td>Capacity Factor</td>
<td>Sub-bituminous Coal</td>
</tr>
<tr>
<td>Heat Rate</td>
<td>Lignite Coal</td>
</tr>
<tr>
<td>Heat Content of Coal</td>
<td>SO₂ Control (Boolean)</td>
</tr>
<tr>
<td>Sulfur Content of Coal</td>
<td>NOₓ Control (Boolean)</td>
</tr>
<tr>
<td>Ash Content of Coal</td>
<td>PM Control (Boolean)</td>
</tr>
<tr>
<td>Supercritical Plant (Boolean)</td>
<td>Mercury Control (Boolean)</td>
</tr>
</tbody>
</table>

NGCC – 11 available parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>NOₓ Control Types:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nameplate Capacity</td>
<td>Water</td>
</tr>
<tr>
<td>Capacity Factor</td>
<td>Catalytic</td>
</tr>
<tr>
<td>Heat Rate</td>
<td>Ammonia</td>
</tr>
<tr>
<td>Percent Gas</td>
<td>Overfire</td>
</tr>
<tr>
<td>Percent NGCC</td>
<td>Low NOₓ Burners</td>
</tr>
<tr>
<td>NOₓ Control (Boolean)</td>
<td></td>
</tr>
</tbody>
</table>
• Plants removed based on heat rate – any plants outside of these heat rates were removed
  • Coal: 7,500-12,000 BTU/kWh (28-45% efficiency)
  • Natural gas: 6,000-15,000 BTU/kWh (23-57% efficiency)

• All negative values were removed

• Species filtering
  • NEI has 151 emissions species for coal plants and 124 for NGCC plants
  • Species with less than 10 facilities reporting emissions were omitted
  • Remaining species count – coal: 90, NGCC: 38
Examined different ways to filter outliers – noticed trend to the right among most species

- The goal was to use as much of the data as possible to train and test the regression model
- Graph shows the average as the highest values are incrementally removed
- Only a handful of values need to be removed before the change in average stabilizes – in this case 4 values

This process was implemented in Python and used to filter outliers prior to creating testing and training datasets

The remaining data was then split into training and test sets
The Regression Method

Machine-learning algorithm LassoCV

Lasso

Multiple Regression

Set of α to try

Measure MSE

Repeat for 3 sets

Average MSE across 3 sets and use α at lowest MSE

2/3 Train

1/3 CV

1/3 CV

1/3 CV

3-fold Cross-Validation

Mean Square Error (MSE) = \( \frac{1}{n} \sum_{i=1}^{n} (Y_i - \sum_j x_{ij} \beta_j)^2 \)

Sum of regression coefficients (\( \beta_j \)) = \( \sum_{j=1}^{p} |\beta_j| \)

Lasso regression minimizes

\[
MSE + \alpha \cdot \sum_{j=1}^{p} |\beta_j|
\]

In minimizing, coefficients can be set to 0 (parameter selection)
Testing the Regression Model

Getting to a final model

- Training data set: 75% of data
- Testing dataset 25%
- LassoCV
- Calculate $R^2$ using LassoCV results
- Models with $R^2 \geq 0.5$ Kept
- Arithmetic mean of coefficients for final model
- Calculate adjusted $R^2$ for final model
- Adjusted $R^2 \geq 0.5$ - model for that emissions is kept

Repeated 1,000 times with randomly chosen Training and Test Sets
The number of successful models generated for each species ranged from 0 to 932 (of 1,000 possible).

The final count of species with accepted emissions models (Adjusted R² ≥ 0.5):

- Coal: 13 (of 32 species that had regression models generated)
- NGCC: 0 (of 17 species that had regression models generated)
Contribution Analysis of Results

What are the most important parameters?

Analysis uses emissions calculated using average plant parameters, and the parameters associated with the minimum and maximum emissions calculated for all plants. These 5 parameters account for 90+ percent of the calculated emissions. These 5 parameters also translate to the amount of coal burned. So why not more successful regression models?

Nameplate – Nameplate capacity (MW), CF – Capacity Factor, Heat Rate (btu/kWh), Heat Content [of coal] (MJ/kg), BIT – bituminous coal, SUB – sub-bituminous coal, LIG – lignite, Supercritical_int – Boolean for supercritical boiler, SO2 control_int, NOx control_int, PM control_int, Mercury control_int – Booleans for presence of pollution control equipment.
Back to the Drawing Board

New Dataset!

- Update the data to use 2014 NEI and power plant data
  - Previous analysis used 2011 NEI and power plant data
- NEI emissions are reported at the boiler level (Tennessee Valley Authority Kingston Plant has 9 boilers)
- EPA 860 data is available at the boiler level and includes
  - Information on all pollution control equipment
    - 10 codes for FGD
    - 24 codes for NO\textsubscript{X} control
    - 16 codes for PM control
    - 21 codes for Hg control
  - Boiler-level fuel consumption with sulfur and ash specs
    - Also includes combustion of DFO and natural gas for auxiliary operations
A Deeper Look at the NEI Data

What do the reported emissions actually represent?

• Working directly with the NEI database has revealed some more detail in the metadata:
  • Of the 53,598 non-zero data points, over half use an EPA, no control emission factor – 119 of 119 species have an EPA emission factor
  • Most emission factors are based on coal throughput (skimming through comments for the calculations)
  • The majority of continuous emission monitoring system (CEMS) data is for NO$_x$, SO$_2$, VOCs, and PM

<table>
<thead>
<tr>
<th>Method of Emission Calculation</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous Emission Monitoring System</td>
<td>1,237</td>
<td>2.3%</td>
</tr>
<tr>
<td>Engineering Judgment</td>
<td>4,281</td>
<td>8.0%</td>
</tr>
<tr>
<td>Manufacturer Specification</td>
<td>66</td>
<td>0.1%</td>
</tr>
<tr>
<td>Material Balance</td>
<td>319</td>
<td>0.6%</td>
</tr>
<tr>
<td>Other Emission Factor (no Control Efficiency used)</td>
<td>833</td>
<td>1.6%</td>
</tr>
<tr>
<td>Other Emission Factor (pre-control) plus Control Efficiency</td>
<td>463</td>
<td>0.9%</td>
</tr>
<tr>
<td>S/L/T Emission Factor (no Control Efficiency used)</td>
<td>3,409</td>
<td>6.4%</td>
</tr>
<tr>
<td>S/L/T Emission Factor (pre-control) plus Control Efficiency</td>
<td>8</td>
<td>0.0%</td>
</tr>
<tr>
<td>S/L/T Speciation Profile</td>
<td>31</td>
<td>0.1%</td>
</tr>
<tr>
<td>Site-Specific Emission Factor (no Control Efficiency used)</td>
<td>906</td>
<td>1.7%</td>
</tr>
<tr>
<td>Site-Specific Emission Factor (pre-control) plus Control Efficiency</td>
<td>30</td>
<td>0.1%</td>
</tr>
<tr>
<td>Stack Test (no Control Efficiency used)</td>
<td>2,736</td>
<td>5.1%</td>
</tr>
<tr>
<td>Stack Test (pre-control) plus Control Efficiency</td>
<td>105</td>
<td>0.2%</td>
</tr>
<tr>
<td>Trade Group Emission Factor (no Control Efficiency used)</td>
<td>1,474</td>
<td>2.8%</td>
</tr>
<tr>
<td>Trade Group Emission Factor (pre-control) plus Control Efficiency</td>
<td>8</td>
<td>0.0%</td>
</tr>
<tr>
<td>USEPA Emission Factor (no Control Efficiency used)</td>
<td>29,135</td>
<td>54.4%</td>
</tr>
<tr>
<td>USEPA Emission Factor (pre-control) plus Control Efficiency</td>
<td>2,827</td>
<td>5.3%</td>
</tr>
<tr>
<td>USEPA Speciation Profile</td>
<td>5,719</td>
<td>10.7%</td>
</tr>
<tr>
<td>Vendor Emission Factor (no Control Efficiency used)</td>
<td>11</td>
<td>0.0%</td>
</tr>
</tbody>
</table>
Adjustment to Method
Changing LassoCV Parameter

• 5-fold cross-validation used for species with > 20 samples, 2-fold otherwise

• Example path for anthracene
Good News, Bad News

Example phenol results - species with the lowest passing final $R^2$ test

Good News

- Species count for successful models and score of those models went up considerably
  - 63 species now have “good” models from 13 previously
  - Notable omissions from the “good” model list due to bad final $R^2$ test (as opposed to too few emissions or no successful regressions):
    - $SO_2$, $NO_X$, ammonia, PM 10, PM 2.5
    - Suspect these score low because emissions are driven by regulation rather than any of the variables

Bad News

- We’re mostly regressing against emissions that are generated using emission factors rather than measured emissions
  - There’s still value in providing a weighted average without examining all of the metadata
Conclusions

• A large number of NEI emissions are the result of emissions factors applied to coal throughput
  • Depending on how you get NEI data this isn’t readily apparent
  • Despite this, using this method to provide a weighted factor used by the fleet is still useful

• The emissions that should be most responsive to plant configuration are not (SO$_2$, NO$_x$, CO, etc.)
  • Suspect this is because permits drive these emissions more than the existence of particular control equipment
  • Would like to re-do the analysis for these emissions using locale as a parameter

• More work to do
  • Revisit the analysis for natural gas plants to see if boiler- or turbine-level data results in successful models
  • Include facility-level emissions: species count from original analysis was 158 vs. 119 in the new approach, presumably omitting facility-level emissions from TRI
Acknowledgements

• Dave Morgan – National Energy Technology Laboratory
• Greg Schivley – Carnegie Mellon University
Contact Information

Timothy J. Skone, P.E.
Senior Environmental Engineer • U.S. DOE, NETL
(412) 386-4495 • timothy.skone@netl.doe.gov

Matt Jamieson
Staff Engineer • KeyLogic
matthew.jamieson@netl.doe.gov

Greg Cooney
Senior Engineer • KeyLogic
gregory.cooney@netl.doe.gov

netl.doe.gov/LCA  LCA@netl.doe.gov  @NETL_News