Pacific Northwest
NATIONAL LABORATORY

PNNL- 29905

Energy Intensity Indicators

Detailed Methodology and Data Sources

May 2020

Prepared by:

S Ganguli
DB Belzer

U.S. DEPARTMENT OF ENERGY

Prepared for the U.S. Department of Energy
under Contract DE-AC05-76RL01830

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

Printed in the United States of America

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information,
P.O. Box 62, Oak Ridge, TN 37831-0062;
ph: (865) 576-8401
fax: (865) 576-5728
email: reports@adonis.osti.gov

Available to the public from the National Technical Information Service
5301 Shawnee Rd., Alexandria, VA 22312
ph: (800) 553-NTIS (6847)
email: orders@ntis.gov <https://www.ntis.gov/about>
Online ordering: http://www.ntis.gov

PNNL- 29905

Energy Intensity Indicators

Detailed Methodology and Data Sources

January 28-29, 2020
Battelle Washington Office,
901 D Street
Washington, D.C.

May 2020

Prepared by:

S Ganguli
DB Belzer

Prepared for
the U.S. Department of Energy
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory
Richland, Washington 99354

Contents

1.0 Introduction

This document was written as a guide to the researcher who will undertake the task of generating future updates to the system of energy intensity indicators developed by the Department of Energy (DOE) Office of Energy Efficiency and Renewable Energy. This system of intensity indicators was developed in the mid-2000's as a tool to help energy analysts and policymakers better understand the changes that have occurred with regard to energy efficiency across various sectors of the U.S. economy.

The most recent report that describes the overall methodology and data is the Pacific Northwest National Laboratory (PNNL) report, *A Comprehensive System of Energy Intensity Indicators for the U.S.:Method, Data, and Key Trends, Rev 3* (to be referred to below as the Comprehensive Report, Rev. 3). This report describes the overall trends in the intensity indicators through 2017. The document here is intended to supplement that report, by identifying in detail the various data sources and methodologies that have been used in the construction of the indicators. This report provides a high-level summary of key data sources and methods used in the updating indicators. It identifies the various supporting spreadsheets that have used in the development of the indicators, as well as the key relationships among those spreadsheets. It also provides a very brief overview of the methodology and assumptions that went in to these computations. The high-level summaries in Sections 2-7 are meant to be more informative as to how the overall system of indicators has been constructed.

The appendix of this document provides detailed descriptions of the steps taken to compute the indicators in the most recent update through 2017. Historically, the entire process has been conducted in a spreadsheet environment. MS Excel spreadsheets have a variety of functions that facilitate the process, including linking, pivot tables, built-in regression functions, non-linear optimization add-ins, and graphing. These capabilities have been employed within the overall methodology to generate the indicators. This document describes the sources of the major underlying data sources, which includes multiple government statistical agencies (primarily the Energy Information Administration and the Bureau of Census) and a variety of industry trade associations. When available, the exact web addresses (as of the 2017 update) for the sources is provided, along with the steps needed to convert the downloaded data into more usable formats. The further steps required to calculate either energy consumption or activity measures are described in detail, with specific references to the worksheets and cell locations of the data. In many cases the rationale for a particular methodological step is discussed.

1.1 Organization

Section 1.2 presents an overview of the commonalities that have been employed, regarding formatting and methodology, across two or more sectors.

Sections 2 through 7 present the more detailed sectoral overviews for each of the 5 sectors starting with the methodology for the economy wide index followed by Residential, Commercial, Transportation, Industrial and Electricity - in that order. The Appendix of this report includes detailed process steps and technical notes regarding energy intensity indicator calculations for each sector.

1.2 Commonalities across Multiple Sectors

The following data formatting and naming convention is applicable to all EII sectors:

Spreadsheet name typically include the term date in their name. Date pertains to the two digit "mo/day/yr" format. Following the convention of several data processing platforms (e.g., "R"), references to a specific worksheet within spreadsheet is denoted in the following manner: spreadsheet_name_date/worksheet_name.xlsx.

Regional Data and Weather Factors for the Residential and Commercial sectors

1. Estimation of "Fuel and Electricity Consumption by Census Region"

    - Data Source for Computations: EIA's State Energy Data System
    - Methodology:

        a) Energy consumption data by region is used to calculate approximate intensity time series.

        b) These series are employed to develop weather adjustment factors at the regional level.

        c) The regional energy consumption data are derived from the state-level estimates as part of EIA's State Energy Data System.

2. Estimation of "Weather Factors"

    - Data source: Annual heating and cooling degree day data from the EIA website. These times series are contained in Tables 1.9 and 1.10 in the Monthly Energy Review.
    - Methodology:

        (a) Estimate a simple regression model to fit the regional intensity to a linear function of time (included squared and cubed values of time) and degree days.

        (b) For electricity the regression model is in terms of a constant term, heating degree day (HDD), cooling degree day (CDD), time, time-squared, and time-cubed.

        (c) The regression-based weather factor is computed as the ratio of the predicted value of the intensity based on actual degree days divided by the predicted value of the intensity based on the long-term averages of the degree days.

        (d) The weather-normalized intensity measures are calculated as the actual intensity divided by the weather factor.

        (e) For fuels, the same general methodology was used. The set of independent variables for these regressions include: HDD, HDD*Time, Time, Time-squared, and a composite fuel price index. The composite fuel price index was developed as a weighted average of the national distillate fuel oil price index and a national average price for natural gas.

        (f) The calculational procedures were repeated for all the census regions.

        (g) The weather factors for delivered energy and source energy are computed implicitly. For delivered energy, they are calculated as the sum of reported electricity and fuels divided by the sum of the weather-adjusted electricity and weather-adjusted fuels. A similar procedure is followed for source energy.

2.0 Economy-Wide Energy Intensity Index

Figure 1 provides a graphical illustration of some of the key linkages that go into to the development of the intensity indexes and depicts the process steps to develop the economy-wide energy intensity index.

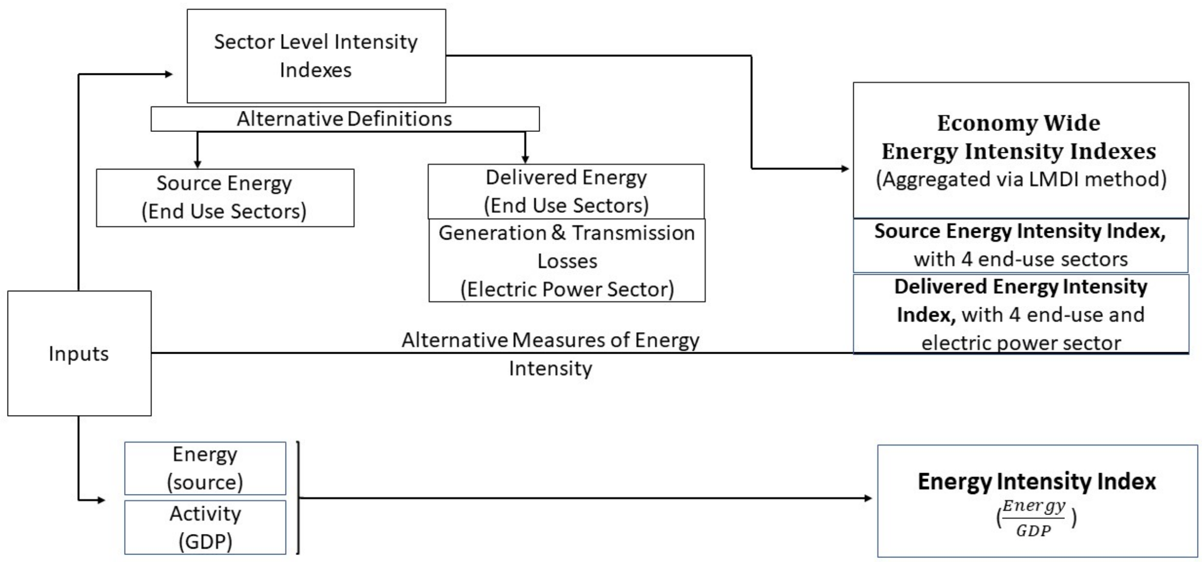

Figure 1. Economy-Wide Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing the economy-wide energy intensity index.

- **Location** (top level spreadsheet): economywide_indicators_date.xls
- **Data Source for computations:** End-use consumption data collected by EIA and end use sector spread sheets, total U.S. Gross Domestic Product (GDP) from the Bureau of Economic Analysis, and end-use sector spreadsheets containing the energy intensity indexes for both source energy and delivered energy.

Overview and Assumptions

- Two alternative measures of energy intensity indexes for the overall economy are developed.
- The first index is based upon the energy-to-GDP ratio.
- The second index (and the preferred one) is computed using Log Mean Division Index (LMDI) methodology to aggregate the source energy intensity indexes for the four major end-use sectors (residential, commercial, industrial, and transportation).
- A variant of the second index is based upon delivered energy intensities (with the four major end-use sectors plus the electric power sector.

3.0 Residential Sector

Figure 2 provides a graphical illustration of some of the key linkages that go into to the development of the intensity indexes and depicts the process steps to develop the Residential sector energy intensity index.

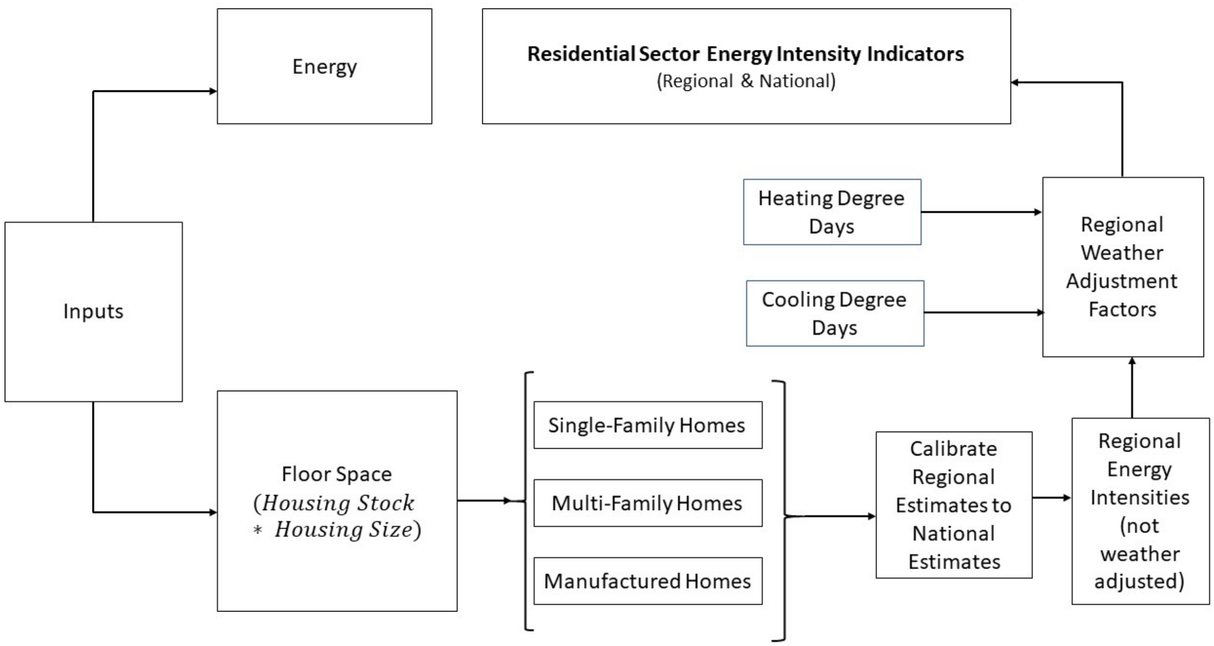

Figure 2. Residential Sector Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing Residential sector EII.

- **Location** (top level spreadsheet)**:** residential_indicators_*date*.xls
- **Data Source for computations:** end-use consumption data collected by EIA and estimates of residential floor space (location: AHS_summary_results.date.xlsx) by census region.

Overview and Assumptions:

A. Data on the number and average size of occupied housing units from the biennial American Housing Survey were employed to generate many of the activity metrics for this sector.

B. Three types of residential housing units are distinguished: single-family, multi-family, and manufactured homes.

C. Regional data from EIA's State Energy Data System (SEDS) are employed to develop regional intensity indicators.

D. Regression models at the regional level are used to adjust for year-to-year changes in weather.

E. Two separate data construction elements are required to generate the regional and national estimates of energy intensity indicators for this sector.

   1. Regional time series of floor space for residential housing units in the U.S (census level).

2. Weather adjustment for the four census regions.

3.1 Estimation of floor space: Occupied Housing Units at the Regional Level

Location: AHS_summary_results.date.xlsx – estimates of regional housing units and regional floor space.

- By housing type:
    - AHS_summary_results_date \Total_Stock_SF.xlsx (single family)
    - AHS_summary_results_date \Total_Stock_MF.xlsx (multifamily)
    - AHD_summary_results_date \Total_Stock_MH.xlsx (manufactured homes)

Data Source for computations: American Housing Survey (AHS) conducted by the Census Bureau to estimate aggregate floor space for three types of housing units: single-family (attached and detached), multi-family, and manufactured homes.

Methodology:

1. An estimated survival curve was first developed from vintage data over the 1999 through 2009 AHS surveys.
2. Curve was used along with reported new construction from the Characteristics of New Housing reports from the Census Bureau.
3. The "stock adjustment model" was used to arrive at estimates of "Occupied Housing Units" at the national level.

3.2 Estimation of floor space: Housing Unit Size at the National Level

Location:

- AHS_summary_results_date \Total_Stock_SF.xlsx (single family) – estimates for housing unit size are to be found in this worksheet to the right of the estimates for the number of "single family" units.
- AHS_summary_results_date \Total_Stock_MF.xlsx (multifamily) - estimates for housing unit size are to be found in this worksheet to the right of the estimates for the number of "multifamily" units.
- AHS_summary_results_date \Total_Stock_MH.xlsx (manufactured homes) - estimates for housing unit size are to be found in this worksheet to the right of the estimates for the number of "manufactured homes" units.

Data Source: American Housing Survey (AHS) conducted by the Census Bureau to estimate aggregate floor space for three types of housing units: single-family (attached and detached), multi-family, and manufactured homes.

Methodology:

- Single family

1. Estimate the average size for existing units after 1985.
2. Estimates of the stock for units constructed prior to 1985, and for 1985 and subsequent years, were made separately.

3. The average size of new single-family homes to the existing housing stock was based upon data from the Characteristics of New Housing (with a 15% upward adjustment to better match the AHS data).

- Multifamily

The same procedure was followed for multi-family units to estimate average national unit size.

- Manufactured Homes

1. The estimates for manufactured home size from the AHS were deemed unsuitable for inclusion in the time series estimates of residential floor space.
2. Instead, the size estimates for mobile homes from the various RECS were employed. While the RECS had inconsistent methods of estimating square footage for single- and multi-family housing units, that does not appear to be the case for mobile homes.

3.3 Estimation of floor space: Regional Shares of National Level Housing Units

Location: AHSa_Summary_Results.xlsx

Data Source: American Housing Survey (AHS) conducted by the Census Bureau to estimate aggregate floor space for three types of housing units: single-family (attached and detached), multi-family, and manufactured homes.

Methodology:

The regional shares for the non-AHS years are computed via a simple average of the preceding (odd) year and subsequent (odd) year.

3.4 Estimation of floor space: Final floor space estimates by housing type

Location: AHS_Summary_Results_date \Final Floorspace Estimates.xlsx

Data Source: American Housing Survey (AHS) conducted by the Census Bureau to estimate aggregate floor space for three types of housing units: single-family (attached and detached), multi-family, and manufactured homes. The estimates of floor space are calculated by multiplying the number of housing units times the average size per unit, as summarized in previous two sections. e regional based estimates of floor space (as explained in the sections above) as control totals to which the regional estimates are calibrated.

3.5 Estimation of Fuel and Electricity Consumption and Intensities (by census region)

The methodology for the estimation of "fuel and electricity consumption" is the same as the one set out Section 1.2 of this document. Guidance to estimate fuel and electricity intensities by census region is as follows:

Location:

- residential_indicators_date/Northeast.xlsx
- residential_indicators_date/Midwest.xlsx
- residential_indicators_date/South.xlsx

- residential_indicators_date/West.xlsx

Note: Each of the above regional worksheets have four sets of energy intensity indicators computed for fuels, electricity, delivered and source energy.

- residential_indicators_date/ Regional_intensity(aggregate).xlsx – contains the time series for fuels and electricity intensity for each of the census regions. These intensities are used in the procedure to compute weather adjustment factors described above.

Data Source for computations: residential_indicators_date/ SEDS_CensusRgn.xlsx

3.6 Estimation of Weather Factors

The methodology for the estimation of "weather factors" is the same as the one set out at the beginning of this summary write-up with one addition/modification. The weather factors are applied at the regional level to generate the weather-normalized intensity indexes for each of the four census regions.

4.0 Commercial Sector

Figure 3 provides a graphical illustration of some of the key linkages that go into to the development of the intensity indexes and depicts the process steps to develop the Commercial sector energy intensity index.

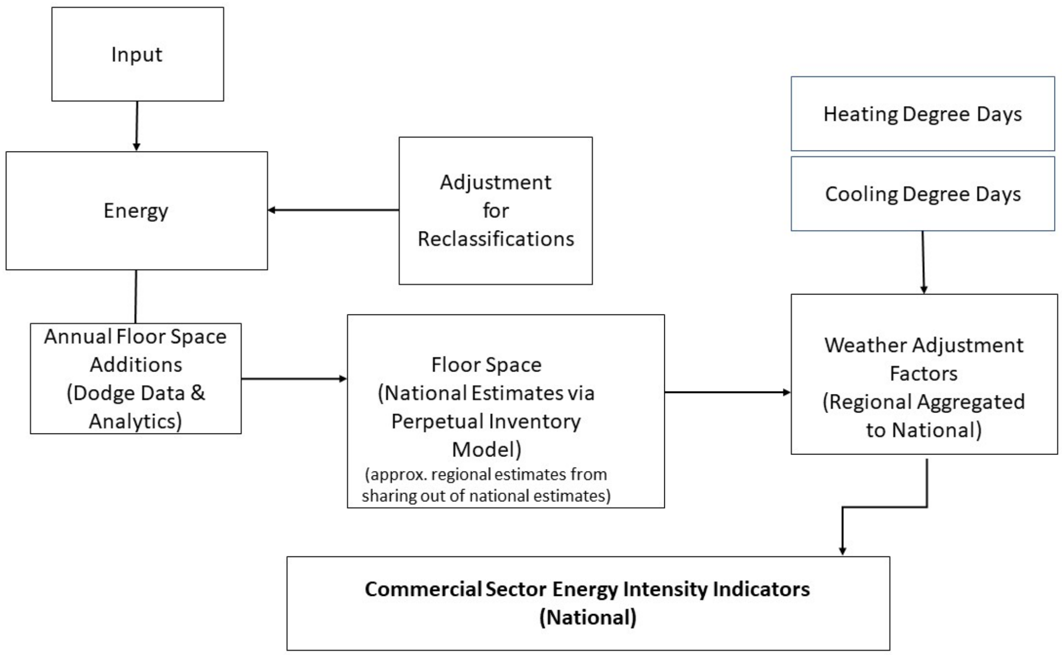

Figure 3. Commercial Sector Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing Commercial sector EII.

- **Location** (top level spreadsheet)**:** commercial_indicators_*date*.xls
- **Data Source for computations:** aggregate commercial sector consumption data collected by EIA and a national estimate of commercial floor space.

Overview and Assumptions:

A. A national time series of floor space for the commercial buildings in the US.

B. Weather adjustment for the four census regions.

C. Adjustment for major reclassifications of customers by individual utilities. The reclassifications generally involve customers whose electricity purchases were classified under an industrial rate structure being moved to commercial rate structure, and to a lesser degree, customers moving from a commercial rate to an industrial rate classification. These reclassifications can distort the short-term aggregate changes in commercial and industrial electricity consumption, as reported by EIA's Estimation of National Commercial floor space.

Location: historical_floorspace_*date*.xlsx

Data Source for computations: New construction is based on data from Dodge Data and Analytics, available from the published versions of the *Statistical Abstract of the United States* (SAUS).

1. commercial_indicators_date\worksheet\Dodge_Revised.xlsx - Values for individual building categories - commercial (principally retail, office, and warehouses), manufacturing (not included in the commercial floor space for the intensity indicators), education, health, public, hospitals, religious, social and recreational and miscellaneous.)
2. commercial_indicators_date\worksheet\ Dodge_to_CBECS.xlsx – Reallocation of building types following cross-walk that was developed between the building categories in the Doge data and the Commercial Building Energy Consumption Survey (CBECS) conducted by EIA.

Methodology: Perpetual inventory model, where estimates of new additions and removals are added to the previous year’s estimate of stock to update to the current year.

4.1 Estimation of Intensity Indexes for Census Region

Location: Regional_intensity.xlsx

Data Source for computations:

1. Fuel consumption and electricity consumption for each region is taken from commercial_indicators\worksheet\SEDS_CensusRgn.xlsx[1]
2. Shares of regional floor space reported in the various Commercial Building Energy Consumption Surveys (CBECS)
3. The regional intensity indexes are used only to produce weather adjustment factors. These indexes are not considered sufficiently robust to incorporate them in the intensity indicator hierarchy.

Methodology:

1. Step 1: Develop estimates of regional commercial floor space (location: Regional_Floorspace.xlsx) by applying regional shares (location: Regional_Shares.xlsx) to total national floor space (location: Commercial_Toal.xlsx.
   a) To arrive at the annual estimates of the regional shares, it was assumed that the commercial floor space in each region would follow the same trends as population or housing units.
   b) It was also assumed that residential housing units reflect these overall trends.
   c) For each region, a simple regression model was estimated between the regional housing unit share and the regional commercial building floor space share from the various historical CBECS.
   d) Based upon the estimated coefficients from this regression, the annual housing unit values were used to predict the share of commercial floor space in each region.
2. Step 2: Compute the intensity indexes as the ratio of energy (electricity and fuels) consumption to regional floor space.

[1] Details on how the fuel consumption estimates were arrived at are set out at the beginning of the document.

4.2 Estimation of Reclassification of Electricity Sector Sales

Location: Adjusted_Supplier_Data.xlsx

Data Source for computations: Commercial electricity sales from the EIA State Energy Data System.

Methodology:

1. It was assumed that the significant changes (same magnitude but opposite directions in the same year) in state-level electricity sales, typically showing up in the commercial and industrial sectors. These changes were presumed to reflect reclassification of some customers from the industrial rate class to the commercial rate class or vice versa.
2. The strategy used was to adjust the more recent data by adding or subtracting a constant value, a value determined from year in which the reclassification was judged to have occurred. These state-level adjustments were aggregated to a national level for each year. The national level adjustment was applied to the reported level of commercial electricity sales in each year.
3. In essence, the objective was to seek a more consistent time series of electricity sales that would have occurred without these major reclassifications.

5.0 Transportation Sector

Figure 4 provides a graphical illustration of some of the key linkages that go into to the development of the intensity indexes and depicts the process steps to develop the Transportation sector energy intensity index.

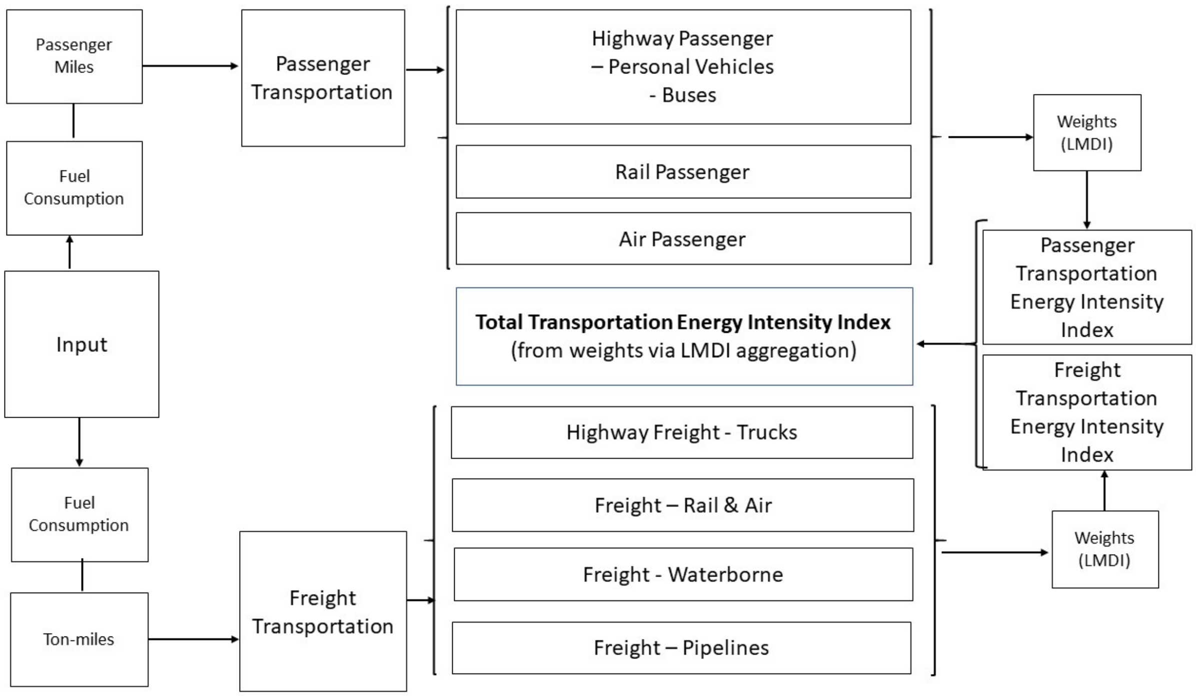

Figure 4. Transportation Sector Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing Transportation sector EII.

<u>Assumptions</u>

The transportation sector was disaggregated as follows:

A. Passenger Transportation
B. Freight Transportation

Passenger transportation is broken out into three major modes: 1) highways, 2) rail, and 3) air. Passenger *highway* transportation is split into segments for personal vehicles and buses, with further disaggregation for each mode.

A. Passenger Transportation:
   i. Highway Passenger – Personal Vehicles
      - Cars/short wheelbase vehicles
      - Light trucks/long wheelbase vehicles
      - Motorcycles
   ii. Highway Passenger – Buses

        - Transit
        - Intercity
        - School
    1. Passenger – Rail
    2. Passenger – Air
B. Freight Transportation
    1. Highway Freight – Trucks
        - Medium
        - Single-Unit
        - Heavy Combination
    2. Freight – Rail and Air
    3. Freight – Waterborne
    4. Freight – Pipelines (natural gas)

Both the activity and energy intensity indexes are developed as energy-weighted averages of the respective separate indexes for the total passenger and total freight segments, with the weights changing over time according to the LMDI methodology. In the current (2020) update to the intensity indicators, the methodology has now been modified such that the latest data from Table VM-1 in *Highway Statistics* is employed in a straightforward fashion. See the Comprehensive Report, Rev. 3.

Location (top level spreadsheet): transportation_indicators_*date*.xlsx.

(Constructed) Data Sources for computations:

1. transportation_indicators_*date* ****FuelConsump.xlsx **-** time series fuel consumption for the major transportation sub-sectors. Data are generally entered in (millions of) gallons or barrels of petroleum product.
2. transportation_indicators_*date* ****Fuel Heat Content.xlsx **-** assumed Btu content of the various types of petroleum products. Note: No user input is required for updating the indicators.
3. transportation_indicators_*date* ****Passenger-based Energy Use.xlsx **-** calculates the fuel consumption in Btu for passenger transportation. Note: No user input is required for updating the indicators (other than adding new rows for most recent years).
4. transportation_indicators_*date* ****Freight-based Energy Use.xlsx **-** calculates the fuel consumption in Btu for freight-based transportation.
5. transportation_indicators_*date* ****Passenger-based Activity.xlsx **-** time series for the activity measures for passenger transportation. Note: the user must input the latest data from various sources.
6. transportation_indicators_*date* \Freight-based Activity.xlsx - time series for the activity measures for freight transportation. Note: the user must input the latest data from various sources.
7. Please refer to the appendix tables in Rev3 for more detailed information on the data sources.

6.0 Industrial Sector

Figure 6 provides a graphical illustration of some of the key linkages that go into to the development of the intensity indexes and depicts the process steps to develop the Industrial sector energy intensity index.

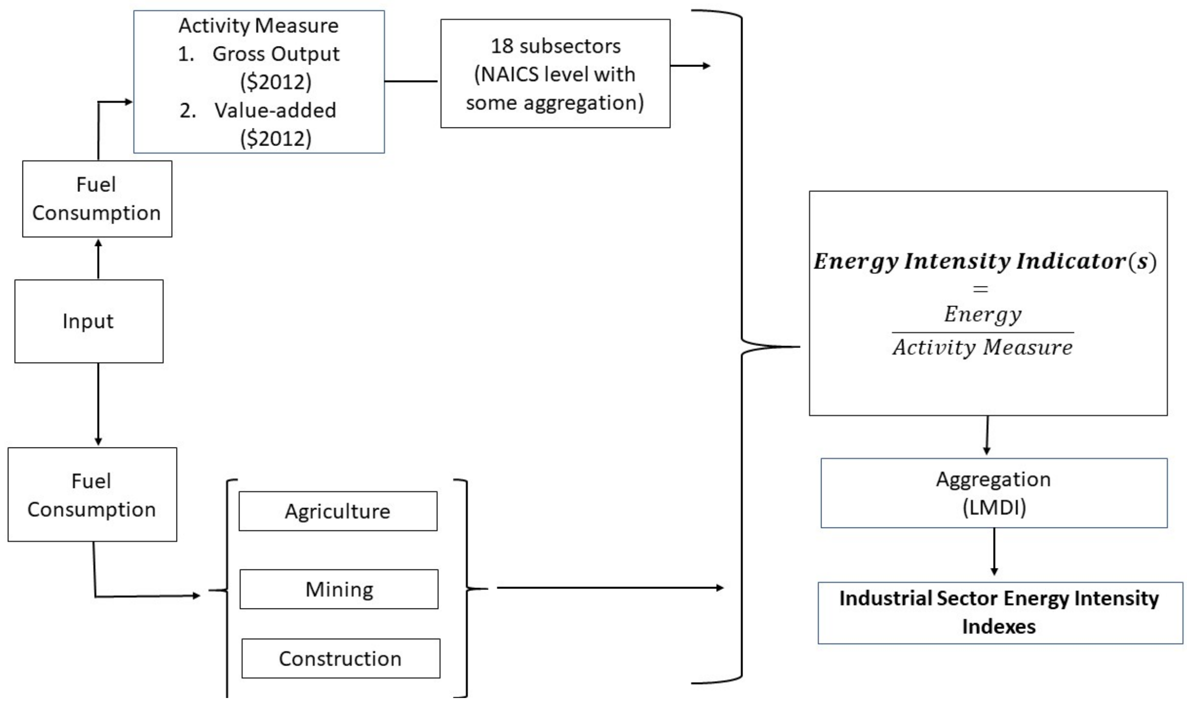

Figure 5. Industrial Sector Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing Industrial sector EII.

Location (top level spreadsheet): industrial_indicators_*date*.xlsx.

Data Source for computations: Output data (gross output and valued added, both in chained $2012, and energy consumption for electricity and total fuels. The key supporting spreadsheets are listed below:

a) indhap3_*date* (manufacturing outputs and energy consumption)
b) GrossOutput_1969-2017_PNNL_*date* (gross output for manufacturing sectors)
c) ValueAdded_1969-2017_PNNL_*date* (value added for manufacturing sectors)
d) NonMan_output_data_*date* (value added estimate for total nonmanufacturing)
e) nonmanufacturing_reconciliation_*date* (gross output and value added for nonmanufacturing sectors
f) Mining_energy_*date* (gross output and energy consumption for three subsectors in mining)

Note: In the most recent update of the intensity indicators in early 2020, the links to these spreadsheets have been retained. In previous updates of the intensity indicators, the industrial_indicators_date spreadsheet has been included as one of the publicly available *spreadsheets. In those cases, the active links to supporting spreadsheets were removed, and the contents of the relevant cells were converted to values only. Parenthetical remarks were included to show the source of the input values – primarily to aid further updates; the supporting spreadsheets were never made available for public distribution*

Overview and Assumptions:

A. This spreadsheet develops the energy intensity indexes for the total industrial sector (level 2 of the energy intensity indicator hierarchy).
B. Separate intensity indexes are computed for manufacturing and nonmanufacturing (level 3 of the intensity indicators hierarchy).
C. Within manufacturing, intensity indexes are computed for 18 individual sectors in manufacturing (generally at the 3-digit NAICS level).
D. Within nonmanufacturing, energy intensity indexes are computed three sectors: agriculture, mining and construction.
E. Within mining, three subsectors are identified: oil and gas extraction, other mining, and mining support.
F. All indexes are computed using the Log Mean Division Index method – type 2 (LMDI-II).

6.1 Manufacturing: Fuels and Electricity Consumption

Location: ASMdata_*date*.xlsx

Data Source for computations: For 3-digit NAICS manufacturing sectors: fuel consumption data collected by EIA (Manufacturing Energy Consumption Survey, MECS), cost of purchased fuels and quantity of electricity consumption from the Census Bureau, and annual estimates of fuel prices by 3-digit NAICS manufacturing sector.

Methodology:

A. Data on fuel consumption by 3-digit NAICS sector derived from EIA's Manufacturing Energy Consumption Survey (MECS). This survey has been conducted at 3- and 4-year intervals since 1985. (Prior to 1985, fuel consumption estimates are derived from the National Energy Accounts – a source no longer available for more recent data).
B. Two different estimates of fuel consumption are used from MECS: 1) Purchased fuels used for heat, power, and electricity generation from offsite-produced sources, and 2) Purchased fuels from offsite + onsite sources = *total* fuel consumption for heat, power, and electricity generation.
C. The first estimate from MECS, purchased fuels from offsite-produced sources is definitionally consistent with data collected by the Census Bureau on purchased fuels as part of its Annual Survey of Manufactures and the Economic Census (ASM/EC).
D. Annual data on cost of purchased fuels from ASM/EC are used to interpolate between MECS estimates of the cost of purchased fuels from offsite sources in MECS years. In MECS years, ratios are developed between the MECS estimate and the reported value from the ASM/EC. The ratios are interpolated between MECS years, and then used to generate calibrated (to the MECS) estimates of cost of purchased fuels for those years.

E. Quantities of fuel consumption for non-MECS years are derived by deflating the calibrated estimates (from D) of the cost of purchased fuels by a price deflator for the cost per MMBtu, averaged over all fuels. Price deflators are developed from price information, by major fuel type, available from each MECS.
F. The weights used to calculate overall cost per MMBtu of fuel are derived from quantities of individual fuels from the MECS. The weights are linearly interpolated between MECS years to generate the weights for non-MECS years.
G. Prices for individual fuels are required for non-MECS years. These prices are derived in separate spreadsheets for each type of fuel. The prices are based on regression models, using national-level fuel prices as independent variables, to predict annual prices for each fuel consistent with fuel prices published in the MECS.
H. The prices for individual fuels from (M) are multiplied by the weights (G) to compute an overall cost per MMBtu (i.e., fuel price deflator) for each year. The calibrated estimates of total cost of purchased fuels [C] are divided by these price deflators to yield estimates of the total quantity of fuel consumption (from offsite-produced sources) for each year. If the calculational process is correct, the deflated values for MECS years will match the reported quantity of total fuel consumption for those years
I. Ratios between the total fuel consumption (offsite-produced + onsite-produced) and fuel consumption (offsite-produced only) are computed for each MECS year. The ratios are interpolated between MECS years and held at the same value for the last available MECS year. The quantity values from (H) are multiplied by these ratios to produce estimates of total fuel consumption for each year.
J. Reported values of electricity consumption from the ASM/EC are collected and reformatted into array form. The electricity values are used directly in the process of intensity indicators for manufacturing. They are *not* calibrated to the MECS, as are the fuel estimates.
K. Estimates of annual fuel consumption and ASM/EM values of electricity consumption are transferred to a second spreadsheet (ind_hap3_*date*.xlsx) for final processing before use in spreadsheet that calculates the final energy intensity indicators for the industrial sector (industrial_indicators_*date*).

6.1.1 Annual Composite Fuel Prices (dollars/MMBtu) for 3-Digit NAICS Manufacturing Sectors

Location: MECS_EnergyPrices_*date*,xlsx

Data Sources for computations: Data for quantities of fuel consumption and prices for individual fuels, both at the 3-digit NAICS level, both values from the MECS available from 1985 through (currently) 2014. Regression-based annual estimates of fuel prices from individual spreadsheets for each fuel (see next section)

Methodology:

A. Construct arrays of fuel consumption (for offsite produced fuels) for each of the MECS from 1985 forward. For MECS prior to 1998, consumption for five major fuels are available (coal, natural gas, distillate fuel oil, residual fuel oil, and other fuels. For 1998 and later, LPG and Coke are added -- making seven fuels in all.

B. Construct arrays of fuel prices paid by each 3-digit NAICS manufacturing sector in each MECS. From 1985 through 1998, prices are taken (or imputed) from the MECS for 5 fuels. For 1998 and later, prices are developed for 7 fuels.
C. The quantities of consumption by fuel for each sector for each MECS year are converted to weights, the weights are the shares of the total fuel consumption attributable to each fuel.
D. The weights are applied to the fuel prices for individual fuels to calculate a composite price (or cost) per MMBtu of total fuel for each year. The prices for individual fuels for MECS years are from the MECS. The estimated prices for non-MECS years are based on regression models for each fuel and each 3-digit NAICS sector – see next section on "Annual Prices by Fuel by 3-Digit NAICS Manufacturing Sector."
E. These prices from MECS_EnergyPrices spreadsheet are linked to the ASM_data_*date* spreadsheet where the prices are used to deflate the calibrated estimates of the cost of purchased fuels from the ASM/EC to yield estimates of fuel consumption.
F. For MECS up through 1998, the MECS_EnergyPrices spreadsheet is also employed to compute the ratio of the total quantity of fuel consumption (offsite-produced + onsite produced) to the quantity of offsite-produced consumption. These ratios also linked to spreadsheet ASM_data_*date* for the estimation of final quantities of fuel consumption for each manufacturing sector.

6.1.2 Annual Prices by Fuel (dollars/MMBtu) by 3-Digit NAICS Manufacturing Sector

Location: Ind_Gas_Price_*date*, Ind_Coal_Price_*date*, Ind_Distillate_Price_*date*, Ind_Resideual_Price_*date*, Ind_LPG_Price_*date*, Ind_Coke_Price_*date*, and Ind_Other_Price_*date*

Data Sources for computations: MECS data on prices for individual fuels by 3-digit NAICS sector from spreadsheet ASM_EnergyPrices_*date*; Aggregate, annual, prices for individual fuels from spreadsheet ASMdata_*date*.

Methodology:

A. Annual prices for each major fuel type, and three-digit NAICS sector, are developed within a separate spreadsheet. (For example, for distillate fuel oil prices, the spreadsheet Ind_Distillate_Prices_*date*).
B. Prices for MECS years are linked from spreadsheet ASM_EnergyPrices_*date*.
C. Aggregate (or national-level) fuel prices for all years linked from spreadsheet ASMdata_*date*.
D. A separate worksheet is devoted to each 3-digit NAICS manufacturing sector. Within each worksheet a non-linear regression model is employed, where the NAICS-sector MECS prices are the dependent variable, and current and lagged prices of the aggregate fuel prices are the independent variables.
E. The nonlinear models must be manually re-estimated when new MECS prices become available. The procedure uses the Excel add-in procedure, Solver, to perform the estimation.
F. The regression residuals are added back to the *predicted* value of the MECS prices to yield the actual (reported) MECS price in MECS years.

G. In non-MECS years, the values of the residuals are interpolated linearly between the residuals in the MECS years. These residuals are added back to the predicted values of the sector-specific prices to yield a final set of prices.
H. The resulting estimates of annual prices by 3-digit NAICS sector are transferred to spreadsheet MECS_EnergyPrices_*date* to calculate composite level fuel prices by 3-digit NAICS sector (i.e., price per MMBtu across all fuels).

6.1.3 Fuels and Electricity Consumption – Final Estimates 1970 – 2017

Location: ind-hap3_*date*.xls

Data Source for computations: All relevant data for updating comes from spreadsheet ASMdata_*date*.

Methodology:

A. This spreadsheet has three functions: 1) performs a second interpolation procedure for estimating fuel consumption in non-MECS years, 2) allows the interpolation to be made with *either* fuel consumption estimates or electricity consumption estimates, and 3) links the pre-1985 estimates of fuels and electricity with estimates for 1985 and later.
B. The second interpolation is used to handle cases where the fuel consumption estimates made in spreadsheet ASMdata_*date* do not match exactly the reported MECS consumption values.
C. The spreadsheet permits the user to use either 1) the ASM-derived estimates of fuel consumption (after deflation of cost of purchased fuels by composite fuel prices), or 2) electricity consumption -- in the interpolation of fuel consumption between values reported in successive MECS.
D. The resulting series of fuel and electricity consumption estimates from 1985 and forward are combined with the 1970-1985 estimates derived from prior work.
E. The complete time series (currently 1970-2017) for each sector are arranged into arrays where years are shown down the rows and sectors across columns.
F. The electricity and fuel consumption final estimates are transferred (linked) to the spreadsheet that computes the energy intensity indexes by manufacturing sector, in spreadsheet industrial_indicators_*date.*

6.1.4 Gross Output and Value-Added by Industrial Sector

Location: GrossOutput_1967-2018_PNNL_*date*.xlsx. ValueAdded_1969-2018_PNNL_*date*.xlsx

Data Source for computations: Estimates of gross output and value added available from the Bureau of Economic Analysis (BEA) website.

Methodology:

A. The overall methodology will be outlined for gross output. Virtually identical methodology for value added.
B. From the BEA website, two spreadsheets with gross output estimates are downloaded: 1) the first with the most recent data, now beginning 1997, and 2) historical data through from either BEA's interactive data selection system, or the full 1947-1997 dataset.
C. The spreadsheet with the most recent data is used as the data processing platform. The pre-1998 data for the "Chain-type Quantity Indexes for Gross Output by Industry" are moved from the historical data spreadsheet to the spreadsheet with the most current data.

D. The quantity indexes for the historical and recent datasets are linked at the year 1977 to create a single array 1970-2018 (or more recent year). Only the rows with industries up through Wholesale Trade need be retained in this linked array.
E. A single column in the worksheet supplied the values of gross output for the base year of the indexes (currently 2012). The values in that column are multiplied by the values for the quantity indexes for all other years. If the quantity indexes are not 1.0 in the base year, the quantity measure are adjusted accordingly. This process converts the indexes to gross output values in chained 2012 dollars.
F. Two special time series of gross output must be created for the industrial indicators system. The first is one is a special construction for “Nonmanufacturing,” combining the BEA estimates for agriculture, mining, and construction. The second time series is for NAICS sector 36, Transportation Equipment. This series is a combination of the BEA estimates for i) Motor vehicles, bodies and trailers, and parts, and ii) Other transportation equipment.
G. For both special constructions, the first step is to construct a chained Fisher-type quantity index (this follows the index construction method used by BEA). The second step is to convert these indexes to constant (2012) dollars by the same multiplication (scaling) process in [E] above.
H. Transpose the values for industry gross output to order the years down the rows rather than across columns as originally supplied by BEA. The transposed values of gross output are linked to the spreadsheet used to compute the final industrial energy intensity indexes, industrial_indicators_*date*.xlsx.
I. The procedure for value added follows the same procedure for gross output. The final values of value added in 2012 (or later base year) dollars are linked to the industrial indicators spreadsheet (industrial_indicators_*date*).

6.2 Nonmanufacturing Sector

The following provides data sources and a brief overview of the methodological approach to developing nonmanufacturing EII estimates for the Industrial sector.

6.2.1 Gross Outputs and Value Added for Nonmanufacturing Sectors

Location: NonMan_output_data_*date*.xlsx

Data Source for computations: The separate spreadsheets for gross output and value added by sector, described above. They are: GrossOutput_1967-2018_PNNL_*date*.xlsx and ValueAdded_1969-2018_PNNL_*date*.xlsx.

A. The gross output and value-added measures for agriculture, mining, and construction are transferred from the two spreadsheets (cited above) derived from the Bureau of Economic Analysis data, as described in the above section “Gross Output and Value Added by Industry.”
B. This spreadsheet simply provides a convenient place from which the separate spreadsheets for agriculture, mining, and construction can link to transfer the appropriate output measures.

6.2.2 Agriculture

Location: Agricultural_energy_*date.*

Data Source for computations: Annual expenditures on fuels from the National Agricultural Statistical Service (NASS), and aggregate (annual) fuel prices from EIA (*Monthly Energy Review*). Output data from spreadsheet NonMan_output_data_*date.*

Methodology:

A. Two special studies provided estimates of fuels and electricity consumption for years 1970-2002, and 2005.
B. Data on farm fuel expenditures are taken from the annual survey of farm expenses conducted by NASS.
C. Aggregate (national-level) prices on distillate fuel oil, gasoline, natural gas, and LPG are taken from the *Monthly Energy Review.* In most cases, these prices are used to extrapolate previous estimates of prices from the *Annual Energy Review* beyond the year 2010. Adjustments to retail prices are made to remove the portion attributable to state and federal taxes, typically not paid by farmers.
D. Consumption estimates are derived by dividing the expenditure data by average fuel prices. These consumption estimates are combined with those from the two special studies mentioned in (A) above.
E. The gross output and value-added measures for agriculture are transferred from the spreadsheet NonMan_output_data_*date*, measures taken from the Bureau of Economic Analysis website.
F. The resulting output and energy consumption estimates for agriculture are transferred to spreadsheet nonmanfacturing_reconciliation_*date*.

6.2.3 Mining

Location: Mining_energy_*date*.

Data Source for computations: For updating estimates, cost of purchased fuels from the Economic Census and aggregate (annual) fuel prices from EIA (*Monthly Energy Review*). Output data (gross output and value added) derived from the Bureau of Economic Analysis (through spreadsheet NonMan_output_data_*date*, and gross output data from the Bureau of Labor Statistics (for detailed sub-sectors in mining).

Methodology:

Prior Economic Census (EC) data for Mining for census years up through 2007 supplied considerable data to estimate both cost and quantity of fuel consumption by fuel type. Data available for 29 individual 6-digit NAICS mining sectors.

A. Data on the cost of purchased fuels (by fuel) from the 2012 Economic Census, downloadable data for 6-digit NAICS sectors in mining (from a single Census Bureau data file identified for Mining, NAICS Industries 211 through 213).
B. Data on the cost of purchased fuels (total) from the 2017 Economic Census, downloadable data for 6-digit NAICS sectors in mining (a single Census Bureau file identified for Mining, for NAICS Industries 211 through 213).
C. Aggregate (national-level) prices on coal, distillate fuel oil, residual fuel oil, gasoline, natural gas, and LPG are taken from the *Monthly Energy Review,* either from that publication directly or via spreadsheet ASMdata_*date*. In some cases, these prices are used to

extrapolate previous estimates of prices from the *Annual Energy Review* beyond the year 2010.

D. Consumption estimates for non-EC years are linearly interpolated between EC years. (They would be extrapolated beyond the last census year at the same (fixed) value from that that year – not relevant in most recent update because last EC year is 2017).
E. For 2012, the consumption quantity estimates by fuel are extrapolated from 2007 based on the change in the purchased cost of the fuel between 2007 and 2012, adjusted for price change (prices from D. above).
F. The gross output and value-added measures for mining are transferred from the spreadsheet NonMan_output_data_*date*, derived from the Bureau of Economic Analysis data.
G. More detailed output data for mining subsectors is derived from data provided by the Bureau of Labor Statistics.
H. The output and energy consumption estimates, 1970-2017, for total mining are transferred (via links) to spreadsheet nonmanfacturing_reconciliation_*date*.
I. Gross output and energy consumption data for three mining sub-sectors – Oil and gas extraction, Other mining, and Mining Service—are transferred to the spreadsheet that computes energy intensity indexes for the industrial sector (industrial_indicators_*date*).

6.2.4 Construction

Location: Construction_energy_*date*.

Data Source for computations: For updating estimates, cost of purchased fuels by fuel type from the Economic Census and aggregate (annual) fuel prices from EIA (*Monthly Energy Review*). Output data from the Bureau of Economic Analysis via spreadsheet NonMan_output_data.*date*.

Methodology:

A. Cost of purchased fuels from the Economic Census beginning with 1977. Cost data provided for electricity, natural gas, gasoline and diesel-on road, gasoline and diesel-off road, and lubricating fuels [not used]. Values for total construction only (NAICS 23).
B. Aggregate (national-level) prices on distillate fuel oil, gasoline, natural gas, and electricity-commercial sector are taken from the *Monthly Energy Review*. In some cases, these prices are used to extrapolate previous estimates of prices from the *Annual Energy Review* beyond the year 2010.
C. Consumption quantities are estimated as straightforward division of the reported cost divided by the average price for that fuel type.
D. The gross output and value-added measures for agriculture are transferred from the spreadsheet NonMan_output_data_*date*, derived from the Bureau of Economic Analysis data.
E. Consumption estimates prior to 1977 are not available. Values for 1970-1976 are fixed at values consistent with computed energy intensity (fuels and electricity) for 1977.
F. The output and energy consumption estimates, 1970-2017, for construction are transferred (via links) to spreadsheet nonmanfacturing_reconciliation_*date*.

6.2.5 Assembly of Nonmanufacturing Energy Consumption and Output Estimates

Location: nonmanufacturing_reconciliation_*date*.

Data Source for computations: Energy consumption and output estimates for the three major nonmanufacturing (industrial) sectors: Agriculture (Agricultural_energy_*date*); Mining (Mining_energy_*date*); Construction (Construction_energy.*date*).

Methodology:

A. The spreadsheet nonmanufacturing_reconciliation_*date* is designed to bring together the separate estimates of consumption and output from the three major nonmanufacturing sectors: agriculture, mining, and construction.
B. The aggregation of the energy use from the three major nonmanufacturing provides a means of comparison with an alternative estimates of nonmanufacturing energy use based on the difference between EIA reported energy consumption for the industrial sector and the MECS (and ASM/EC) estimates of manufacturing energy use. This comparison step is not required in the current indicators methodology and was not undertaken in the most recent 2020 update).
C. The energy consumption estimates (electricity and total fuel) and output measures for agriculture and construction are transferred from this spreadsheet to the spreadsheet industrial_indicators_*date* where the computations for the final sets of industrial energy intensity indicators are located.
D. The energy and output measures for mining are *not* transferred to the industrial indicators spreadsheet from nonmanufacturing_reconciliation_*date*. Rather, the estimates for the three mining sub-sectors are transferred to the industrial indicators spreadsheet by direct link from Mining_energy_*date*.

7.0 Electricity Sector

Figure 6 provides a flow diagram depicting the process steps to develop the Electricity sector energy intensity index.

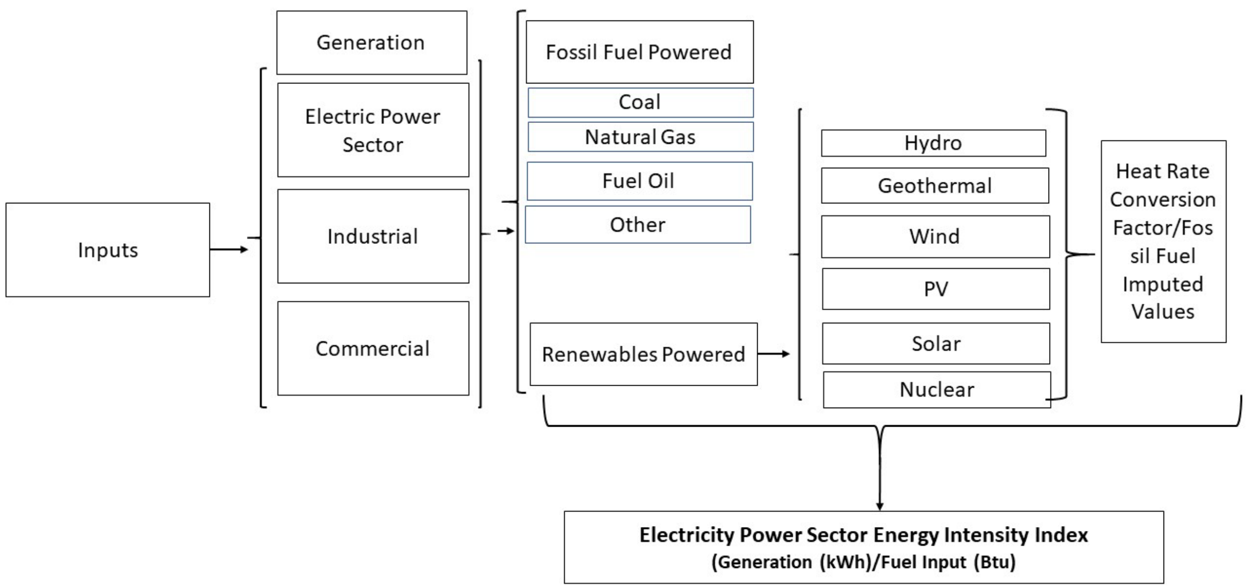

Figure 6. Electricity Sector Methodology Flow Diagram

The following provides data sources and a brief overview of the methodological approach to developing Electricity sector EII.

Location (top level spreadsheet): electricity_indicators_*date*.xls

Data Source for Computations: EIA. Prior to 2005 – exclusively from EIA's *Annual Energy Review* (AER), from 2006 through 2010, from AER and special processing of the EIA-923 survey. For most recent years, exclusive reliance on processing of the annual EIA-923 utility-level datasets available from EIA.

Methodology:

1. Prior to 2005, the data on energy consumption fuels to produce electricity) was generally supplied in physical units only (e.g., mcf of natural gas, tons of coal, etc.), as published in EIA's Annual Energy Review (discontinued after 2011)
2. The values needed to be converted to Btu, and also be consistent with aggregate energy consumption for this sector as published by EIA.
3. Each major fossil fuel (natural gas, coal, petroleum and other fuels) has a separate worksheet and the suffix "reconcile" – to reflect the adjustment necessary to be consistent with aggregate energy consumption for that sector. Ex: "NatGas Reconcile" seeks to produce an estimate of the Btu consumption of natural gas used to generate electricity

4. The data for fuel consumption and electricity generation for the years subsequent to 2005 were derived primarily from special processing of the utility-level micro datasets, as performed by EIA staff and transmitted to PNNL. For years 2015-2017, this processing was done by PNNL using the downloadable utility-level data sets from the EIA-923 survey.

8.0 Conclusion

This sectoral summary was written with the intent of providing the researcher with a very high-level overview of this project. A major challenge stemmed from changes in the methodology of the underlying data sources. So, going forward, the researcher needs to be cognizant of the underlying data sources that are to be used for updating these numbers – since changes in the underlying methodology employed in documenting these data sources may call for a change in the methodology used to derive the indicators.

One should note that a limitation of this analysis stems from the lag in the available data; generally, the experience has been that the indicators cannot be fully updated to include a particular year until approximately 18 months later. Thus, for example, the indicators through 2017 (reflected in the current methodology) could not have been completed until mid-2019. The release dates for major updates in EIA's end-use consumption surveys (RECS, CBECS, and MECS), and those for the key Census Bureau publications (e.g, Annual Survey of Manufactures) may also influence when the most robust measure of recent trends of energy intensities can be estimated.

Appendix-- Detailed Description of Process Steps and Data Sources

This appendix provides a detailed description of the data sources and how they are used to develop the energy consumption and economic activity metrics that are required to produce the various energy intensity indexes. The section describes the specific worksheets and cell locations of the various data elements as the process moves toward the final sets of consumption and activity measures.

One major caveat as the researcher may undertake the process of updating the indicators using the spreadsheet environment. The spreadsheets employed were developed at PNNL over a roughly a 15-year period, with many contributors and various revisions to the methodology over that time. In the last few iterations of the updating process, the focus was maintained on producing the most robust and historically consistent set of intensity indictors. This has resulted in a situation where some the supporting spreadsheets have many apparent duplicative worksheets or somewhat illogical linkages. (This deficiency does not carry over to the top-level spreadsheets for each major sector [as identified in Section I of this report]. Those spreadsheets have been made publicly available on EERE's website over a number of years.) The apparent duplicative elements in the spreadsheet reflect old methodologies that were retained, until a new methodology was deemed to provide a more accurate measure of the relevant energy intensity. In any event, the discussion in the sections below describe the steps required using the current versions of these spreadsheets in a hopefully straightforward manner. There has been no attempt to fully document the spreadsheet in every aspect. This includes no effort to document the complete history of the data and sources used to generate the indicators going all the way back to 1970. Rather, the aim has been to aid the researcher who wishes to continue using the spreadsheet environment to update the indicators subsequent to 2017.

I. Residential Sector: Process Steps and Technical Notes

A. Background

The energy intensity indicators for the residential sector are contained in a single spreadsheet, residential_indicators_date.xlsx, where date is expressed in mo/day/yr format. The indicators for the residential sector are derived entirely from the end-use consumption data collected by EIA and estimates of residential floor space by census region.

The various changes in how the residential end-use surveys were conducted by EIA (RECS) preclude the use of these surveys alone to develop time series of intensity indicators for this sector. As a result, the data on the number and average size of occupied housing units from the biennial American Housing Survey were employed to generate many of the activity metrics for this sector. Three types of residential housing units are distinguished: single-family, multi-family, and manufactured homes. Regional data from EIA's State Energy Data System (SEDS) are employed to develop regional intensity indicators. Regression models at the regional level are used to adjust for year-to-year changes in weather (as reflected by degree-days for both heating and cooling). Greater geographical disaggregation is not possible given that the AHS survey units are classified only at the census region level through 2013.

Two separate data construction elements are required to generate the regional and national estimates of energy intensity indicators for this sector. The first is to develop (census) regional time series of floor space for residential housing units in the U.S. The second element involves the weather adjustment for the four census regions.

The historical estimates for floor space are developed in a single spreadsheet, "AHS_summary_results.date.xlsx," where again date is in the two-digit "mo/day/yr" format. The "AHS" denotes that the underlying source of the data is from the American Housing Survey (AHS) conducted by the Census Bureau. (Prior to 1985, this survey was called the Annual Housing Survey). This current spreadsheet is the result of several years of data and methodological revisions. No substantive effort was made to streamline the spreadsheet to include only sections that reflect the latest methodology. The AHS data were used to estimate aggregate floor space for three types of housing units: single-family (attached and detached), multi-family, and manufactured homes.

Despite the large sample size in the AHS, over 60,000 housing units in recent years, there still appeared to be sufficient sampling variability to make some of bi-annual changes in the national estimates of the total number of units somewhat implausible. To smooth out these variations, the basic methodology to update the series of floor space is to develop a perpetual inventory model, where estimates of new additions and removals are added to the previous year's estimated stock to update to the current year. New construction, in terms of the number of new units completed, is based on data collected by the Census Bureau (and partially funded by the Department and Urban Development) from the Survey of New Construction. There was no effort to estimate separate models for the number of housing units at the regional level. Instead, the regional shares of housing units were independently estimated on the basis of simple trend models using the underlying data from the AHS.

To estimate floor space, separate estimates of average size per unit (in square feet) were developed for each of the three housing types at the national level. The average size of housing units at the regional level was based on simple models (or analyst judgment) of trends in the ratio of regional average size to the national level average size.

B. Update Data from the American Housing Survey

For the 2020 update of the EERE intensity indicators, the most recent versions of the AHS public use files were downloaded, with 2017 as the latest version released by the Census Bureau. Links to the available files for 2017 and earlier years can be found at: https://www.census.gov/programs-surveys/ahs/data.html.

The following discussion will relate primarily to the processing of the 2017 data. One can expect that future update work would be similar, given that the Census Bureau will release any future AHS in the same format. From the Census Bureau website given by the URL above, the 2017 data is downloaded as a folder (AHS 2017 National PUF v3.0 CSV) with four Excel comma-separated values files. Load the largest file, "household" into Excel. This flat file for 2017 extends to column AOX and to row 66,753. The user must be patient for the full file to load into Excel.

Create a new Excel spreadsheet to process the required variables. For 2020 update, this file was named AHS_2017_extract. From the Census public use file, copy seven columns of data into the extract file, using the following procedure. Set the cursor on the top row of the table and in the desired column. The entire column can be highlighted for copying with the keystroke SHIFT-CTRL-down arrow. This is followed up by the normal keystroke CTRL-C to copy, and then CTRL-V to paste the entire column of values into the extract file. The following variables (columns) should be copied into the extract file: JYRBUILT, WEIGHT, YRBUILT, DIVISION, BLD, UNITSIZE, and VACANCY. A hint to help this process is to first copy the row of variable names into a separate worksheet in the extract spreadsheet. One can then use the search function within that worksheet to locate the specific column in which the desired variable is located. In the spreadsheet AHS_2017_extract, this worksheet has been named "name search."

The next step is to create a pivot table from the worksheet that contains the seven variables (for the case here, the worksheet "extract_data." In the worksheet that is created by the instruction to create a pivot table (by moving to the "Insert" label on the top row of Excel instructions, and clicking on the leftmost item, "PivotTable" beneath). Check the boxes for the following variables: WEIGHT, DIVISION, BLD, UNITSIZE, and VACANCY. First drag the variable name VACANCY to be the top name in the Filters portion of the pivot table layout. Drag the variable BLD to be directly underneath in the Filters area. Next drag the variable DIVISION to the top of the "Columns" area of the pivot table. Drag the UNITSIZE variable under the "Rows" heading (upper right quadrant of pivot table). And finally drag the WEIGHT variable into the lower right portion of the table, under the [Sum of] Values heading. The WEIGHTS represent the number of housing units in the population for each particular observation in the sample.

As shown in the spreadsheet AHS_2017_extract, worksheet Pivot Tables, the pivot table processing generated six tables shown below the active pivot table in the first 16 rows of the worksheet. The six tables are for Single Family (SF) units (Total and Occupied), Multi-Family (MF) units (Total and Occupied), and Manufactured Home (Man.Homes) units (Total and Occupied). These tables were obtained by copying the values from the active pivot table at the top of the worksheet, each table reflecting a different choice for VACANCY and BLD. For the total number of units of each building type, the variable VACANCY is set to "All". For occupied units, set the VACANCY variable to '-6'. For the building type, single family units are obtained by selecting building code '02' and '03' (corresponding to detached and attached units). Multi-family units are selected by checking the boxes under BLD, '04' through '09'). Manufactured homes are selected with a check of the code '01.'

The rows of the pivot table correspond to various size categories. As discussed in the comprehensive documentation report, only categorical size data were included in the AHS surveys for 2015 and 2017. An effort to utilize the category data to estimate the average size change between 2015 and 2017 did not provide useful results. Processing of future AHS public use files may exclude any consideration of this variable. In the current tables, only the values for total units across all sizes were considered.

Using the data representing occupied units, an aggregation to total units by census region was undertaken. (Unfortunately, the public use files for 2015 and 2017 did not contain a variable for the census region, only a variable for the nine census divisions.) The formulas for this aggregation are shown in columns M through P to the right of the extracted pivot tables. The total units across all size categories, and for each of the major types of residential units, are shown in column T.

C. National estimates of occupied housing units

The large spreadsheet that develops the estimates of regional housing units and regional floor space is AHS_Summary_results_date. This spreadsheet incorporates a somewhat complex set of analytical steps with the goal of yielding robust estimate of floor space at the regional level. Much of the analysis involved various interpolation and data-fitting routines with the AHS data sets prior to 2015. As noted in the comprehensive documentation report, the AHS collected information on the actual size of housing units (as supplied by respondents) over the period 1999 through 2013. A significant portion of the spreadsheet is devoted to processing the information on unit size to develop aggregate measures of floor area. The documentation here will make only a cursory effort to explain the prior work to develop such estimates.

The national annual estimates of housing units by type are developed in three separate worksheets, corresponding to the three housing types in the AHS_Summary_results spreadsheet. The worksheets are Total_Stock_SF (single family), Total_Stock_MF for multi-family, and Total_Stock_MH. To remind the reader, the methodology here was designed to smooth out some of the implausible changes in the reported number of housing from one AHS to the next. The overall methodology is described more generally in the comprehensive documentation report, Section A.1.2.

As a means of smoothing out these time series, as explained in the comprehensive documentation, an estimated survival curve was first developed from vintage data over the 1999 through 2009 AHS surveys. This curve was used along with reported new construction from the Characteristics of New Housing reports from the Census Bureau.

Several models were estimated in an effort to produce robust estimates of the annual number of occupied single-family units. The data and calculations from some of the less successful models have not been removed from the spreadsheet. In this case, most of the data and calculations in columns A through AH are not used in the final estimation. The exceptions are columns J and X. In column J, the values for occupied single-family units are shown for the AHS years. As shown by notes for cells J45 and J47, these values have been derived from the pivot tables from the 2015 and 2017 AHS public use (micro) data files. Column X shows the data for all single-family units (occupied + vacant units). Again, the values for the last two years are derived from the pivot tables. For the previous years, the totals were inserted from the values in the published Census Bureau reports. For updates beyond 2017, the user will need to add rows to the portion of the worksheet and insert the values from pivot tables or from any published values from the American Housing Survey.

Column E shows the values for the number of new units, derived from the Census Bureau's webpage for New Residential Construction at https://www.census.gov/construction/nrc/historical_data/index.html. . The Excel table from this website, associated with the menu item "Housing Units Completed," is included as the worksheet "Comps Ann". The table shows new privately-owned housing units completed from 1968 through 2018. Data for single-family homes is in the first column of data in the table. The data from worksheet "Comps Ann" is linked to column E in the Total_Stock_SF worksheet.

The stock adjustment model is shown in the columns AJ through AL. These columns represent the existing (prior year's) stock, estimated retirements, and new units. The sum of these columns is the estimated end-of-year stock shown in column AM. The actual stock reported by the AHS is shown in column AN. (More detailed discussion of the actual model employed here is provided in Section A.1.2 of the comprehensive documentation report.)

The parameters to be adjusted to best fit the actual values for units (all units, occupied + vacant) are shown in cells AJ9:AJ11. The objective function is set to minimize the squared differences between the predicted stock and the actual stock.

The Excel solver routine is used to estimate the values of the parameters AJ9 and AJ11 (in this model, AJ10 is set equal to 1). To load this particular model, the user must first set the Excel task bar to Data (rather than "Home", "Insert", "Page Layout", "Formulas", "Review", "View" or "Help"). The Solver icon (name) will appear at right side of the task bar, under "Comments." Click on the Solver name. Load the saved model that is currently in cell range AV5 to AV9. The objective function at the top of Solver box should be listed as cell AJ11. Click on the button at the bottom of the outlined box, "Solve." With the current data from the 2020 update, there should be no change in the parameters, AJ9 and AJ11.

The predicted values for the number of units in column AM are repeated in column AU. To calculate the number of occupied units, the reported values from the AHS on vacant units is used. In early work, the vacant units were considered as only those for sale or for rent. Column AV shows the calculation of occupied units, by subtracting this category (only) of vacant unit from total units. This approach was subsequently dropped; thus, AV is NOT USED in the current methodology. Rather, a vacancy rate encompassing all vacant units was employed. That rate is calculated in column Z. An occupancy fraction is calculated as (1 – Z), and is multiplied by the regression-based predicted values of total stock from column AV. The resulting estimate of occupied units are shown in the yellow highlighted column AX. In the columns AP through AW, the predicted and actual values of both total stock and occupied are plotted. The graphs provide some visual support to the value of smoothing out some of the variability in the short-term changes in the number of units between successive AHS surveys.

For manufactured homes, the number of new units (placed) comes from a separate survey (Manufactured Housing Survey) conducted by the Census Bureau. Thus, the source for the number of units in column E of the Total_Stock_MH worksheet is different from that used for single-family and multi-family units. The data from this survey are downloaded and put in a separate worksheet, place_nsa_all. Annual data for the most current years were not found on the Census Bureau website. Monthly data were downloaded for both total units and single (wide) units from the Census Bureau (in worksheets CIDR-1 and CIDR-single). The monthly data were aggregated to an annual basis for the years 2014 through 2018 on these worksheets and the annual values appended to the existing data in the place_nsa_all worksheet. (Ideally, the place_nsa_all spreadsheet would be available from the Census Bureau for the most recent years but was not found as part of the 2020 update work.)

D. National estimates of housing unit size

The comprehensive documentation report explains in general fashion how the data from the AHS were used to develop estimates of the national average size of housing units within each of the three types. As stated earlier, the question requesting the specific size of a sample unit was discontinued in the two most recent surveys for 2015 and 2017.

In the documentation here, no attempt will be made to explain the modeling approach to estimate average sizes per housing unit up through 1985. The implementation of this approach follows the right of the columns used to estimate the number of occupied units (in worksheets Total_Stock_SF, Total_Stock_MF, and Total_Stock_MH. For single-family units, the model comprises the columns BM through BT. In general, an effort was made to estimate the average size for existing units after 1985. Estimates of the stock for units constructed prior to 1985, and for 1985 and subsequent years, were made separately. The average size of new single-family to the stock was based upon data from the Characteristics of New Housing (with a 15% upward adjustment to better match the AHS data). The average size for new units is displayed in column BE in the worksheet. The predicted values of the average size are shown in column BR. The adjoining column shows the actual sizes. Through 201l the actual average sizes are derived from the AHS. The average sizes are transferred from worksheet AHS_tables, where the relevant data from each edition of the AHS from 1985 through 2011 is shown.

The "actual" values for subsequent years--2013, 2015, and 2017--are based on estimates derived from the 2009 and 2015 Residential Energy Consumption Surveys. This information comes from worksheet RECS_4_Adj. This worksheet contains national-level estimates of floor space as reflected in the various RECS starting in 1980. Data from the two most recent RECS (for 2009 and 2015) were appended to the table that was posted by EIA in 2009 on its efficiency indicators website. A separate worksheet, RECS_2, was also derived from the table posted by EIA in 2009; this table displayed the national estimates of housing units by type from the historical RECS. Again, the data (on the number of units) from the two most recent RECS were appended to the data prepared by EIA in 2009.

The bottom panel in RECS_4_adj show the average sizes for each type of housing unit (Rows 38 through 52). The next to last column displays average size for the single-family units (attached and detached). Over the six-year period from the 2009 to 2015, the average size increased about 60 square feet per unit. Assuming a uniform change over this period, the change over each two-period is 20 square feet. A conservative estimate of the change was deemed to be 20 percent lower, or sixteen square feet in each two-year period, a value in the yellow highlighted cell W52.

Lacking other information to extrapolate the AHS estimates of average size, this incremental change was assumed to have occurred over each of the three two-year periods between 2011 and 2017. This assumption is reflected in the changes shown in cells BD43, BD45, and BD47 in worksheet Total_stock_SF. These values are assumed to reflect actual changes over this period and are included in the model estimation from 1985 and forward. Again, the predicted sizes from the model for all (odd) years are shown in column BR.

The model-estimated change in average square feet from 2009 to 2017 (in column BR) was considerably higher than the change reflected in the AHS estimates between 2009 and 2011 and the subsequent incremental changes suggested by the RECS data. To bring these series into more agreement, the predicted changes were adjusted downward between 2009 and 2013, by variable amounts up to 13 square feet per unit. The adjusted values are shown in orange highlighted cells in column BR. This makes the average size in the final estimates very comparable to the annual changes implied by comparison of the 2009 and 2015 RECS. The predicted (smoothed) values of average square feet are shown in the graphic in columns BZ through CF to the right of the numeric values.

<u>Multi-family</u>

In general, the same procedure was followed for multi-family units to estimate average national unit size. These calculations are located in worksheet Total_Stock_MF to the right of the estimates for the number of multi-family units. The AHS-derived average size for multi-family units for each edition of the AHS are transferred into several columns in the Total_Stock_MF worksheet. (The first column in the worksheet is BC.) The columns to the right, in the column range BD through CJ can be ignored; they represent a preliminary effort to estimate a model to predict the average size of the stock of multi-family units.

The final method is shown in the columns CK through CU. Two methods were used to predict average size. The one shown in the light blue column (CS) is used to back cast the estimates from 1997 back to 1985. The second model is used for the years 1999 through 2009. In the most recent years, there appears to be no real trend in the average size of multi-family. For the odd-numbered years from 2007 through 2013, the national average sizes from the AHS are: 1,021, 1,043, 1,030, and 1,035. Thus, it was deemed that to attempt to model these small changes would likely introduce spurious changes in the average size of these units. Accordingly, from 2009 and later years, a simple average of the values computed from the 2007 through 2013 AHS surveys was employed. Those values are shown in dark tan shaded cells, CR40 through CR48. The resulting series of national average size estimate from 1985 through 2018 are plotted in the area to the right of column CR. The graph provides some visual support that the smoothed (predicted) estimates of average size are in reasonable congruence with the data derived from the AHS over the 1999 through 2013.

<u>Manufactured Homes</u>

The estimates for manufactured home size from the AHS were deemed unsuitable for inclusion in the time series estimates of residential floor space. While the AHS showed a general increase in the size of manufactured homes between 1999 and 2013, the estimates appear to unrealistically high. They would imply that nearly all such homes were double-wide mobile homes in the most recent years of the AHS survey. Perhaps respondents included other types of homes that were considered factory-built, but not an actual mobile home that is transported to a (typically) permanent site. Moreover, the estimates over the last four AHS surveys with square footage data (2007 through 2013) show large fluctuations—on the order of 100 square feet in either direction—from one survey to the next.

Accordingly, the size estimates for mobile homes from the various RECS were employed. While the RECS had inconsistent methods of estimating square footage for single- and multi-family housing units, that does not appear to be the case for mobile homes.

The average square footage estimates are shown in the yellow-highlighted column CD in worksheet Total_Stock_MH. The estimates for non-RECS years were made by simple linear interpolation between years for which RECS was conducted. As a conservative estimate of the increase in national average size after the 2015 RECS, the values were increased about 5 square feet for both 2016 and 2017. Future updates will need to incorporate the results from the 2019 RECS.

E. Regional Shares of national-level housing units

From the pivot tables developed from the most recent public use files (2013, 2015, estimates of the number of housing units by census region and housing type were calculated (as explained in the previous section). The next step is to transfer (or link) these regional housing unit estimates to the appropriate worksheet and location in the AHS_Summary_results spreadsheet. The 2017 regional estimates for single-family homes are transferred to cell range CM500:CM503 in worksheet AHS_Summary_results. The multi-family estimates are in same column, in rows 558-561. Manufactured home estimates are put into rows 622-625.

Below each set of regional housing unit estimates, regional shares are computed. For example, the regional shares for 2017 are in cell range CM508:CM511. This set of values makes up the latest year of a table that shows the shares for the years from 1985 forward (beginning in column BW). This table (BW508:CM511) is transposed further to the right in the worksheet. See cell range DP497:EC514. The regional shares for the non-AHS years are computed via a simple average of the preceding (odd) year and subsequent (odd) year. The final set of regional shares of housing units for all years is shown in columns EG through EJ. Similar tables are computed for the multi-family units and manufactured homes.

These regional shares are linked into the worksheet that shows the final set of housing unit floor space estimates – Final Floorspace Estimates. The regional shares of housing units are shown in this worksheet in the light-green shaded columns I through K, corresponding to single-family, multi-family, and manufactured homes.

F. Regional estimates of average housing unit size

As stated above, regional estimates of average housing unit size were based on an examination of the ratios of regional average size to national average size. The estimates for these ratios are included in the worksheet AHS_tables.

The calculation of regional/national size ratios for single-family housing units are contained in rows 152 to 264. These calculations are performed only for the years in which size was included in the AHS, 1997 through 2013. All the data and formulas for the calculations begin in column BW.

An early attempt was made to analyze the size ratios by vintage for specific AHS surveys. Thus, the cell range BX157:CF161 includes these ratios for the 1997 AHS. These ratios are NOT used in the final methodology. Below that cell range are the size ratios for all single-family units, as reported in the various editions of the AHS. (In this worksheet, the summary results for all the AHS surveys are shown in the panels across the top of worksheet.) Thus, for example, the size of the average housing unit in the Northeast census region was 1.136 times the average national size of single-family units. The ratios for the Northeast census region are transposed in the cell range to the right CI165 to CI172.

A simple linear time trend regression was performed, using the ratios from the 1999 through 2013 AHS. (The 1997 data were dropped as that AHS contained significantly fewer observations with size information.) The light-green shaded area shows the application of the LINEST (matrix) regression function in Excel. In the LINEST formula, the values of dependent variable are designated with cell range above, CI165:CI172. The top row of the shaded area shows the constant term (12.902) and the predicted annual change in the ratio (-00588).

In the columns directly under the LINEST matrix function, the predicted and actual values are shown. The predicted values are in cells CI188 through CI221. Above the predicted values are two parameters show in purple-highlighted cells. The lower of these two cells has a value of 0.75. This parameter is used in the back casting of the ratio from 1999 to 1985. The 0.75 value used to lower the rate of change as compared to that over the estimation period, 1999 through 2013. The upper cell is a parameter that influences the rate of change in the ratios later than 2013. In this application the parameter controls a geometrically declining series. With a value of 0.5, the change from 2013 to 2014 was 0.5 times the annual change over the 1999-2013 period (as reflected in the trend coefficient from the regression fit). For the change between 2014 and 2015, the change is 0.5 applied once again to the 2013-2014 change. Clearly, the approach taken here is not to simply assume the rate of change continues unchanged beyond the regression period.

The graphic shown below the numerical values for the Northeast census region illustrates the result of the regression fitting, and the extensions in the periods before and after the data period in the regression fit (1999 – 2013). Clearly, in the years beyond 2013, the trend is moderated rather quickly in the years 2014 and later.

As one scrolls across the plots for the other census regions, it is clear that the projected trends (relative to trend for 1999 through 2013) for the years prior to 1999 are somewhat different for each region. This calls for some explanation. In seeking to back cast the size ratios back to 1985, there was some effort to try make those ratios yield growth rates in regional stock are comparable to the ratios implied by the various RECS. While the absolute values of floorspace reported across the full set of RECS were deemed sufficiently inconsistent to be used in estimating total floorspace over the time period between 1985 and 2015, the results from RECS conducted prior to 1999 seem to be reasonably consistent. The worksheet Compare_FS shows the values of regional floor space from the various RECS in a set of stacked panels. The data points for the specific years are contrasted against the final predicted floor space estimates derived from the AHS and the various estimation procedures described above. To be clear, the predicted annual values depend upon the pre-1999 trend in average national unit size, as well as the trend in the ratios of regional to national average size. This latter trend is dependent upon the parameters that are used to develop the pre-1999 trend, namely the values set in the purple-shaded cells along row 185 in worksheet AHS_tables (CI185, CN185, CT185, and CZ185). In setting these parameters, attention was given the general congruence between the predicted values of the regional stock from the method described here versus the RECS-based regional estimates of floor space. Thus, the parameters controlling the trends in regional/national average size prior to 1999 were adjusted within a limited range to better fit (on a visual basis) the RECS regional estimates from 1984 through 1993. To reiterate, these comparisons between RECS-based stocks and the stock estimated via the AHS are shown in worksheet Compare_FS.

Multi-family

The procedures to estimate the regional/national size ratios for multi-family are located in the area of the AHS_tables below the treatment of single-family units. Specifically, the estimation procedures are shown in rows 267 through 384. The layout for multi-family is identical to that used for single-family. The ratios of regional size to national size for 1999 through 2013 are in cell range are BY279:CF282.

Again, these ratios (for the Northeast census region) are transposed to cells CI278 to CI285. The trend regression formulation using LINEST is in the light green shaded block of cells.

Based upon the estimated coefficients from the trend regression over 1999 through 2013, and the procedure to extrapolate the predicted ratios outside the estimation period, the predicted ratios for all years are plotted within cell range CC339:CK357. The size ratios (regional/national) for multi-family unit have significantly greater variability than those for single-family units. As a result, the general approach was to be very conservative in projecting the trends outside the years in the estimation period. The graphs for each census region will illustrate the choice of parameters that follow this general approach. For the Northeast, the regression fit included only the data through the 2009 AHS, as the actual ratios for 2011 and 2013 seemed inconsistent with the earlier data. In contrast to the single-family homes, no attempt was made to check the congruence with the RECS estimates.

<u>Manufactured Homes</u>

The procedures to project the size ratios for manufactured home are located in rows 385 through 500. As remarked in the discussion about the national average size of this housing type, the AHS estimates display considerable variability from one survey to the next. As a result, the size information for manufactured homes is taken from the RECS (for mobile homes).

Nevertheless, some effort was made to determine if the AHS-derived size ratios were reasonably consistent over time, even if the absolute sizes of these units showed significant variability. Thus, the ratios of the regional to national average size were computed and put in the same format as the other two housing types. These ratios are shown the cell range BY401:CF404. The subsequent analytics are performed in same manner as that described above.

The plots of these ratios are shown in the rows immediately after row 461. For only one region, the Midwest, do the data suggest a clear trend over the 1999-2013 period. As a result, the decision was made to not try to extract regional trends when the data appear to reflect rather large sampling error. To support that position, one can point to what were likely rather uniform national trends in the manufacture and marketing of mobile homes. Thus, the size ratios by region were set at a fixed value over the entire time period. For each region, the average ratio was calculated over the eight separate AHS surveys. For the Northeast census region, this average was calculated in cell CJ460 (with a value of 0.9605). To reflect the uncertainty about the accuracy of these averages, the final fixed ratio was set midway between the calculated average and 1.0 (i.e., taking an average of these two values.) The resulting ratio is shown in the yellow highlighted cell CK460 (0.9802). The fixed ratios for the other three census regions are calculated in like manner, all shown in the yellow-highlighted cells across row 460.

G. Final floor space estimates by housing type

The final set of floor space estimates are put together in the worksheet Final Floorspace Estimates. The basic methodology is to employ the national based estimates of floor space (as explained in the sections above) as control totals to which the regional estimates are calibrated.

The worksheet consists of five panels of time series estimates, with the national estimates as the top panel (in the present version, in rows 10 through 43. The values derived in the other worksheets and used in the overall calculational procedure are shown in the light blue shaded cells. In the top panel, the values for the number of housing units by type are shown in the cell range D10:D43. To the right of these cells, in columns Q through S, are the estimated national average size of these units.

Moving lower in the worksheet for the first census region, Northeast, the regional shares for each type of housing unit are shown in range I49:K82. The regional shares for the other census regions are located in the successively lower panels in the same columns. The ratios of the sizes of housing units in the Northeast census region are shown in cell range AJ49:AL82.

For each census region a preliminary estimate of floor space by housing type is developed. Using the Northeast census region as an example, the number of units is computed as the regional share of the national total. The resulting values are shown in cell range D49:F82. An initial estimate of the average size for each of these types of units is shown in cells AF49 through AH82. These values are product of the relative size ratios (derived from worksheet AHS_tables) and the average national size of the corresponding type of housing unit. In the three columns to the immediate left, the initial regional estimates of floor space (billion square feet) are calculated as the number of regional units (columns D through F) times the average size (AF through AH). The initial estimates for floor space for the other census regions are calculated in the same manner.

The sum of the initial floor space estimates across the four census regions is calculated in the cell range at the top of worksheet, cells AB10 through AB44. These values can be compared against the total national floor space derived solely from the national-level data (i.e., in the cells U10 through W43. The ratio of the regional sum of floor space to the national-level estimate is calculated in each cell of the range AF10:AH43. These ratios are used as calibration factors, applied uniformly across the four census regions. The procedure is to adjust the initial size estimates by the appropriate ratio, applied as simple scaling factor that is the same for all census regions. The calibrated regional size estimates for the Northeast census region are show in cell range Q49:S82. Using these adjusted average sizes, the regional estimates of total floor space are derived by the product of the adjusted average sizes times the number of units. The final regional estimates of floor space (for the Northeast) are shown in the yellow highlighted cells, U49 through W82. The estimates for the other regions are shown in similar yellow-highlighted ranges in the same columns, but in the lower areas of the worksheet.

H. Transfer final floor space estimates to residential indicators spreadsheet

The final step in the process to employ floor space in the construction of residential intensity indicators is to copy the values from the AHS_Summary_results.xlsx file to the intensity indicators file for the residential sector, residential_indicators_date.xlsx. In the most recent update (2020), the links between the AHS_Summary_results.xlsx file and the indicators spreadsheet have been left in place for transparency.

In previous updates of the energy intensity indicators, the indicators spreadsheets were made available on the DOE/EERE website. In this case, the linked values were eliminated and just the values were retained.

If copying of values only is desired, five individual copy and paste operations must be undertaken, corresponding to the national set of values and sets of values for each of the census regions. To assist in this copy operation, the upper left-most cell to be copied should be the light purple colored cell, “1985,” in column C of the worksheet “Final Floorspace Estimates” in the spreadsheet, AHS_Summary_results. Highlight all the required rows and the columns out through column X. Then find the corresponding cell (again, year 1985 in purple) in the Final Floorspace Estimates worksheet in the residential_indicators spreadsheet. Then execute the Excel function to copy values, implemented by a right mouse click and selecting the “paste special” command to copy Values. As mentioned above, the copy and pasted operations must be performed for each census region, after the copy is made for the national-level values.

I. Fuels and electricity consumption by census region

Energy consumption data by region is used to calculate approximate intensity time series; these series are employed to develop weather adjustment factors at the national level. The regional energy consumption data are derived from the state-level estimates as part of EIA's State Energy Data System. The following steps have been followed to process this data. It should be noted that this process is the same as that described for the commercial sector. Generally, when updating the consumption data from the SEDS, the regional values for both sectors should be processed at the same time.

1. Download the state-level data file from the EIA website. The 1960-2017 SEDS data can be found at https://www.eia.gov/state/seds/seds-data-complete.php?sid=US. Download the consumption file, use_all_btu.xlsx. This is a single worksheet with the data by each state arranged stacked on top of each other and the time series data from 1960 through 2017 as individual columns. The first three columns provide the year of the release (in this latest iteration, 2017), the state label (two-letter abbreviation), and the fuel type (termed "MSN" by EIA). For the 2017 version of the file, there are 10,508 rows.

2. Relabel the file with an appended name. Typically, the suffix "_pnnl" has been added to the file name.

3. From a previous version of the processed file, copy the worksheet entitled "Regions." This worksheet contains a mapping between states and census regions, with the state abbreviation in column 1 and the corresponding census region in column 2.

4. From a previous version of the processed file, examine the instruction to append the appropriate census region (1, 2, 3, or 4) to each row in the large dataset. A new column is added to the right of the last year of data. Use the Excel function VLOOKUP to add these census region identifiers. The actual instruction copied down in each cell down this column (of over 10,000 values) is LOOKUP($B2,Regions!$A$2:$B$53,2,FALSE).

5. Create a pivot table in a separate worksheet. The fields selected in the pivot table report are MSN, Region, and all the years (1960 through 2017 in latest work).

6. In the construction of the pivot table, drag the entry MSN into the upper right portion of the pivot table menu, under the heading "columns." No entry is required in the upper left portion under "Filters." In the lower left portion of the pivot table layout ("rows") drag the variable name "region." In the lower right-hand portion of the pivot table layout, drag each year identifier. Ensure the value field setting is "sum." Thus, the first entry under Values is "Sum of 1960," followed by "Sum of 1961" and so forth.

7. In the upper left portion of the resulting pivot table, choose "select all" under the heading "Row Labels." This will typically in column A of the worksheet.

8. In the upper left portion of the resulting pivot table, select only two values under the heading "Column Labels." The first is "ESRCB," an entry that selects out the electricity sales (consumption) for the residential sector. The second selection under "Column Label" is "TNRCB". This choice will pull out total primary energy consumption. The entry contains the sum of fuel consumption plus electricity.

9. The resulting pivot table will thus consist of three columns. The leftmost column displays the year, labeled as "Sum of 1960," Sum of 1961" … The next two columns give the values for electricity and fuels, respectively. The time series 1960 through 2017 are stacked, with the national data in the first set of rows, followed by the time series for each of the regions.

10. To provide a better organized structure to the data, the values from the pivot table are transferred to a separate worksheet. For the most recent work in 2020, this separate worksheet for the residential sector data was labeled "PT results – Residential." This reorganization contains a single set of years down the column, with the energy data for each census region in adjoining columns. Two sets of data

are put together in this fashion, one for electricity and the second for total energy. For the 2020 work, in the spreadsheet “use_all_btu_2017_pnnl, the electricity totals are in columns F through I. Total energy is in columns L through O. Total fuel consumption is calculated as the differences between total energy and electricity sales. In this particular file, the derived fuel consumption estimates are shown in columns R through U.

11. A final minor rearrangement of the data is made in columns further to the right (X through AB for electricity) and (AE through AI for fuel consumption.) The data values from these two areas in the worksheets are copied to the residential_indicators spreadsheet, worksheet SEDS_CensusRgn.

J. Fuels and electricity intensities for census regions

In the residential sector, separate sets of intensity indicators are developed for each census region. Each set (for fuels, electricity, delivered, and source) of regional energy intensity indicators is computed in a separate worksheet: Northeast, Midwest, South, and West. The regional estimates of households by housing types are pulled into each worksheet from the worksheet Final Floorspace Estimates. For the most recent update (2020), the fuels and consumption data are linked from the worksheet SEDS_CensusRgn. The nominal energy intensity, kBut/sq.ft., is computed in column R of these regional worksheets. These intensities, for both fuels and electricity, are passed to worksheet Regional_intensity (aggregate), via active links. Both intensities are plotted on the same graph, one graph for each census region. These intensities are subsequently used to develop weather adjustment factors for both fuels and electricity, as explained in the following section.

K. Weather Factors

A set of regression models are used to compute weather factors that are intended to remove, at least approximately, the effects of weather on the intensity indexes. The following steps are followed to update the weather factors. This process is essentially identical to the weather adjustment methodology used for the commercial sector indicators. A major difference is that for the residential sector the weather factors are applied at the regional level to generate the weather-normalized intensity indexes for each of the four census regions.

1. Download the annual heating and cooling degree day data from the EIA website. These times series are contained in Tables 1.9 and 1.10 in the Monthly Energy Review. These tables are inserted as worksheets in the residential_indicators spreadsheets. In the 2020 update, these worksheets are labeled HDD_by_Division18 and CDD_by_Division18.

2. The weather factors are computed in a single worksheet, Weather_Factors. The weather factors resulting from this process are divided into the intensity indexes to account for fluctuations in heating and cooling degree days. The weather factors are computed for each region and energy type (electricity and fuels) and then combined to produce a single national level factor for each energy type.

3. The first step in updating the weather factors is to add rows at the bottom of the existing series to accommodate the additional set of years. For the 2020 update, rows 74 to 77 were added to the prior update. Row 77 should be considered as a placeholder for a subsequent update. The intensity indicators, based on the SEDS, only went up through 2017. (Note that some places in the worksheet there is value for 2018, but the intensity value is the same as 2017.)

4. The process for developing weather factors will be discussed in the detail for the northeast region. Columns R and S show heating degree days for the two census divisions in the northeast census region. The HDD values are computed for the census region, using fixed weights derived from the 1993 RECS. These weights are shown in the blue cells in row 83. The same calculation is performed for CDD, fixed weights are used to compute a value of CDD for the census region.

5. For each census region, an area of the spreadsheet is used to first compute the weather adjustment factor for electricity, followed by an area to the right that is used for an adjustment factor for fuels. In the case of the northeast census region, the area for the electricity adjustment factor is comprised of columns AA through AL. For fuels, the adjustment factor is computed in the columns AT through BF.

6. The values shown the columns shaded in tan (columns Z and AS) are not used in the current methodology. They are computed strictly as weighted functions of the HDD and/or CDD. These columns should be ignored in the update process.

7. The current methodology for developing weather adjustment factors relies on estimating a simple regression model to fit the regional intensity to a linear function of time (included squared and cubed values of time) and degree days. For electricity the regression model is in terms of a constant term, HDD, CDD, time, time-squared, and time-cubed. The variables are show in columns AB through AF. The dependent variable, electricity intensity, is shown in column AH. The regression coefficients are obtained through the use of the "LINEST" matrix function in Excel. Separate models are estimated for two sub-periods, 1970-1984, and 1985-2017. For the earlier sub-period, the LINEST function comprises the light-green highlighted range AB:AG8. In cell Z4, the following formula is entered {=LINEST(AH29:AF43,AC29:AG43,1,1)}. The first range shows the values of the dependent variable (rows 29 through 43 corresponding to 1970 through 1984). The second cell range shows the set of independent variables. The regression for the second subperiod is in range AB16:AG20 again highlighted in light green. The regression formula in each of these cells is {=LINEST(AF44:AF76,AA44:AE76,1,1)}. Note the values, AF76 and AE76. The '76' refers to the last row (year) for the regression. As additional years are added, one must go into this formula and update the row with the last year of data and reset the regression model with the SHIFT-CNTRL-ENTER keystroke. (See Excel documentation for executing matrix functions. Once the data range has been redefined, the regression coefficients will automatically update should the data values themselves be changed.

8. An inconvenient feature of the Excel LINEST regression function is that it reverses the order of the coefficients in the results, as compared to the order of the corresponding independent. The regression coefficients are in the first row of the LINEST matrix function range. Rows 11 and 23 display coefficients after re-ordering, such that the coefficient is in the same column as the corresponding independent variable.

9. The predicted values from the regression model are shown in column AG. In Column AI, the predicted value of the intensity is computed once again, only in this case using 30-year averages of the HDD and CDD. For the regression model covering 1970-1984, the 30-year average is taken over the period 1961-1990. For the later regression sub-period, the predicted values of intensity employ the HDD and CDD values for the period 1981-2010.

10. The regression-based weather factor is computed as the ratio of the predicted value of the intensity based on actual degree days divided by the predicted value of the intensity based on the long-term averages of the degree days. These values are shown in the yellow-highlighted column AL. The weather-normalized intensity measures, calculated as the actual intensity divided by the weather factor, are shown in column AJ.

11. The same general methodology was followed for developing a weather adjustment factor for fuels. For the Northeast census region, the data and formulas for these factors are shown in columns AT through BF. The set of independent variables for these regressions include: HDD, HDD*Time, Time, Time-squared, and a composite fuel price index. The composite fuel price index was developed as a weighted average of the national distillate fuel oil price index and a national average price for natural gas. The fuel price is actually entered as a three-year distributed lag of the prices, with weights 0.5,

0.33, and 0.17. See columns K and L in the worksheet. In the majority of cases (region and time period) the coefficient on the fuel price was negative and moderately statistically significant. (For the northeast fuels regression over the 1985-2017 time period, the coefficient was -2.71 with a standard error of 1.61, as shown in cells AU16 and AU17.)

12. The calculational procedures are performed identically for the other census regions, Midwest, South, and West. These procedures are contained in the portion of the worksheet to the right of those for the Northeast.

13. The weather factors for each census region are then linked to the appropriate cell range in each worksheet (i.e., Northeast, Midwest, South, and West) where the regional intensity indexes are computed. (Thus, in contrast to the commercial sector, there is no need to calculate a national-level weather-adjustment factor prior to the calculation of the national energy intensity indexes.) Thus, in this case, many of the columns in the Weather_Factors worksheet to the right of the calculations for the West census region can be ignored. The implicit national weather factors for fuels are shown in the yellow-highlighted cells in columns IK and IL. As implicit factors, they are derived from the aggregation procedures contained in the worksheet to compute the national-level indicators, worksheet National. Accordingly, they are shown here only for informational purposes.

II. Commercial Sector: Process Steps and Technical Notes

A. Background

The energy intensity indicators for the commercial sector are contained in a single spreadsheet, commercial_indicators_date.xlsx, where date is expressed in mo/day/yr format. The indicators for the commercial sector are derived from the end-use consumption data collected by EIA and a national estimate of commercial floor space.

The various changes in the end-use surveys conducted by EIA (NBECS, CBECS) have precluded the use of these surveys to develop consistent intensity indicators for this sector over time. The EERE indicators for this sector relate to national-level only data. However, regional data from EIA's State Energy Data System are employed to develop approximate regional intensity measures. These intensity measures are used in regression models to make weather adjustments at the regional level. However, the regional intensity estimates are not considered sufficiently robust to include as part of the final set of indicators for the commercial sector.

Three separate elements make up the process to generate the national estimates for the intensity indicators. The first is to develop a national time series of floor space for the commercial buildings in the U.S. The second element involves the weather adjustment for the four census regions. The third step involves an adjustment for major reclassifications of electricity customers that have occurred historically at the state level. These reclassifications have generally involved changes in utility rate structures that have moved industrial customers into a commercial category, and vice versa.

B. Updating National Commercial Floor Space

The estimates for commercial floor space are developed in a single spreadsheet, "historical_floorspace_date.xlsx," where date is in the two-digit "mo/day/yr" format. This spreadsheet is the result of many years of data and methodological revisions. No substantive effort was made to streamline the spreadsheet to include only sections that reflect the latest methodology. A few no-longer needed worksheets were deleted in early 2020 to maintain transparency.

The basic methodology to update the series of floor space is to employ a perpetual inventory model, where estimates of new additions and removals are added to the previous year's estimate of stock to update to the current year. New construction is based on data from Dodge Data and Analytics. Dodge Data and Analytics is the most recent name of a firm, originally termed F.W. Dodge, that collects new construction information from local jurisdictions across the United States.

The "Dodge data" on new floor space additions is available from the published versions of the Statistical Abstract of the United States (SAUS). (It is not available on the electronic versions of that volume). The most recent data is from the 2020 SAUS, Table 995, "Construction Contracts Started—Value of Construction and Floor Space of Buildings by Class of Construction: 2014 to 2018." In this edition, new floor space is reported in million square feet. For 2017, total nonresidential floor space additions were 1,190 million square feet. Individual building categories include: commercial (principally retail, office, and warehouses), manufacturing (not included in the commercial floor space for the intensity indicators), education, health, public, hospitals, religious, social and recreational and miscellaneous.) The values for each of these building types are included in worksheet "Dodge_Revised" in the commercial_indicators spreadsheet. The 2017 data is inserted into row 117 of this worksheet.

In the mid-2000's a cross-walk was developed between the building categories in the "Dodge data" and the Commercial Building Energy Consumption Survey (CBECS) conducted by EIA. This reallocation of building types occurs in a separate tab in the spreadsheet, "Dodge_to_CBECS."

The data from this latter worksheet are then linked to a portion of the primary worksheet in the spreadsheet, "NEMS_Logistic (current). The user must make room to add new data years at the left side of the worksheet. The latest data (2018) is now shown in row 397, columns G through Q.

Column D (light purple) shows the values of the survival function over a period of 150 years. At 150 years, row 247 shows that 0.05662 of the original floorspace would be left standing. The survival curve is based upon nonlinear regression model applied to the vintage data from the 1989 and 1999 CBECS.

Column T shows the estimated floor space additions for all building types from the Dodge data (after adjustments). In the adjoining column, a timing adjustment is made. The Dodge data show new additions that were started in each year. Based on some analysis of the value put in place estimates from the Census Bureau, a rough estimate is that 40% of the additions are completed in the same year, 60% in the following year. The lag weights are shown in cells W210 and W211. The floor space additions after application of this timing adjustment are shown in column U.

Column AD reflects an adjustment for underreporting by Dodge. The floorspace additions are adjusted by factors that depend upon the general vintage of the building. The adjustment factors are based on an analysis of the 1989 and 1999 CBECS data by vintage. For the data after 1999, the factors are selected by a switch in cell AA364. Currently the adjustment for the recent data is 20%; that is, the Dodge data are increased by a factor of 1.2 to reflect under reporting. This value is generally consistent with the factors used by the Census Bureau to account for underreporting by Dodge (although the Census Bureau is interested only in the value of construction.).

The adjusted values for new additions are repeated in column AL. The next column shows the stock of floorspace after applying the values from the survival curve. For 2017, this process resulted in an estimate of 87, 227 million square feet of commercial floor space.

Mechanically applying the survival curve to the additions data after 2008 results in a very low growth in the stock over the subsequent decade. Given the severe collapse in the new commercial construction in the aftermath of the 2009 recession, the decision was made to scale down the removals estimated from the survival curve. Unfortunately, this is a subjective judgement; no data exist on the level of removals (demolitions) for commercial buildings in the U.S. The fractional reductions in the level of removals are provided in column AQ. The values shown range from 30 to 40% (0.3 to 0.4). These fractions relate to fraction of building that would otherwise would have been removed but were not because of the recessionary conditions affecting commercial construction. The estimated floor space assumed to remain in the stock are shown in column AS. These estimates are added back to the stock estimate from column AM to produce revised estimates of the stock. The final stock estimates are shown in the yellow-highlighted column AT. These values are plotted in the worksheet “Floor Space Graph.” These values for national floor space are transferred to the commercial_indicators file to generate the national intensity indexes. The specific location is column F in worksheet Floorspace_estimates in the commercial_indicators spreadsheet.

C. Fuels and Electricity Consumption by Census Region

Energy consumption data by region is used to calculate approximate intensity time series; these series are employed to develop weather adjustment factors at the national level. The regional energy consumption data are derived from the state-level estimates as part of EIA’s State Energy Data System. The following steps have been followed to process this data.

1. Locate state-level energy consumption data file from the EIA website. The 1960-2017 SEDS data can be found at https://www.eia.gov/state/seds/seds-data-complete.php?sid=US. Download the consumption file, use_all_btu.xlsx. This is a single worksheet with the fuel-specific energy data for each state, each state stacked on top of each other and the time series data from 1960 through 2017 as individual rows. The first three columns in this spreadsheet provide: i) the year of the release (in this latest iteration, 2017), ii) the state label (two-letter abbreviation), and, (iii) the fuel type (termed “MSN” by EIA). For the 2017 version of the file, there are 10,508 rows.

2. It is suggested that the file be renamed, with an appended label. Typically, the suffix “_pnnl” has been added to the file name.

3. From a previous version of the processed file, copy the worksheet entitled “Regions.” This worksheet contains a mapping between states and census regions, with the state abbreviation in column 1 and the corresponding census region in column 2.

4. From a previous version of the processed file, examine the Excel (VLookup) instruction to append the appropriate census region (1, 2, 3, or 4) to each row in the large dataset. A new column is added to the right of the last year of data, where the use of the Excel function VLOOKUP adds these census region identifiers for each row of the data file. The actual instruction can be copied down in each cell down this column (of over 10,000 values) is LOOKUP($B2,Regions!$A$2:$B$53,2,FALSE).

5. Create a pivot table in a separate worksheet. The fields selected in the pivot table report are MSN, Region, and all the years (1960 through 2017 in latest work).

6. In the construction of the pivot table, drag the entry MSN into the upper right portion of the pivot table menu, under the heading “Columns.” No entry is required in the upper left portion under “Filters.” In the lower left portion of the pivot table layout (“Rows”) drag the variable name “region.” In the lower right-hand portion of the pivot table layout, drag each year identifier. Ensure the value field setting is “sum.” Thus, the first entry under Values is “Sum of 1960,” followed by “Sum of 1961” and so forth.

7. In the upper left portion of the resulting pivot table, choose “select all” under the heading “Row Labels.” This will typically be located in column A of the worksheet.

8. In the upper left portion of the resulting pivot table, select only two values under the heading “Column Labels.” The first is “ESCCB,” an entry that selects out the electricity sales (consumption) for the commercial sector. The second selection under “Column Label” is “TNCCB”. This choice will pull out total primary energy consumption. The entry contains the sum of fuel (and renewable) consumption plus electricity.

9. The resulting pivot table will thus consist of three columns. The leftmost column displays the year, labeled as “Sum of 1960,” Sum of 1961” … The next two columns give the values for electricity and fuels, respectively. The time series 1960 through 2017 are stacked in blocks, with the national data in the first set of rows, followed by the time series for each of the regions.

10. To provide a better organized structure to the data, the values from the pivot table are transferred to a separate worksheet. For the most recent work in 2020, this separate worksheet for the commercial sector data was labeled “PT results – Commercial.” This reorganization contains a single set of years down the column, with the energy data for each census region in adjoining columns. Two sets of data are put together in this fashion, one for electricity and the second for total energy. For the 2020 work, in the spreadsheet “use_all_btu_2017_pnnl, the electricity totals are in columns F through I. Total energy is in columns L through O. Total fuel consumption is calculated as the differences between total energy and electricity sales. In this particular file, the derived fuel consumption estimates are shown in columns R through U.

11. A final minor rearrangement of the data is made in columns further to the right (X through AB for electricity) and (AE through AI for fuel consumption.) The data values from these two areas in the worksheets are copied to the commercial_indicators spreadsheet, worksheet SEDS_CensusRgn.

D. Intensity Indexes for Census Regions

This process first involves developing estimates of regional commercial floor space. The intensity indexes are then readily calculated using the ratio of the energy consumption (electricity and fuels as the subject in the prior section, C.).

Regional estimates of floorspace are derived by applying regional shares to the total national floor space that were estimated within the spreadsheet historical_floorspace.xlsx (as explained in section A above.) Thus, the regional estimates are not derived from regional estimates of new additions to floorspace and estimated removals. Insufficient data are available to apply this approach at the regional level.

The time series of regional shares are estimated in the worksheet “Regional Shares.” The key data for this process are the shares of regional floor space reported in the various Commercial Building Energy Consumption Surveys (CBECS). (In the years prior to 1986, the survey was titled Nonresidential Energy Consumption Survey, NBECS.) These values are shown the upper left-hand portion of the worksheet. The implied shares are shown in the table defined from cell L9 to O17.

To provide annual estimates of the shares, the assumption was made that commercial floor space in each region would generally follow the same trends as population or housing units. Here residential housing units were used to reflect these overall trends. Based on the regional estimates of residential housing units, regional shares are readily computed. These shares are shown in columns X through AA in the worksheet.

For each region, a simple regression model was estimated between the regional housing unit share and the regional commercial building floor space share from the NBECS/CBECS. This regression employed both shares in log form; the regression data for the Northeast, for example, are shown in cells AC11:AD19.

Based upon the estimated coefficients from this regression, the annual housing unit values are used to predict the share of commercial floor space in each region. For the Northeast, these predicted shares are shown in column AG. The sum of the predicted shares is calculated in column BL. The next set of columns (BO to BR) display the shares after they have been normalized to add to 1.0.

The normalized shares of floor space by census region are transferred to the worksheet Regional_Floorspace, columns E through H. Columns K through N calculate the floor space levels (billion square feet) by multiplying the regional share times the national estimate of floor space. The national estimate is taken from the worksheet (Commercial_Total) where the final set of commercial sector intensity indexes are calculated.

In the worksheet Regional_intensity (aggregate) the intensities (for fuels and electricity) are computed for each census region. The floor space estimates are transferred to this worksheet from the Regional_Floorspace worksheet. Fuel consumption and electricity consumption for each region is taken from SEDS_CensusRgn worksheet, as documented above. The time series of absolute intensities (kBtu/sq. ft.) are shown the pairwise columns and also plotted for each region.

E. Weather Factors

A set of regression models are used to compute weather factors that are intended to remove, at least to an approximate degree, the effects of annual weather on the intensity indexes. The following steps are followed to update the weather factors.

1. Download the annual heating and cooling degree day data from the EIA website. These times series are contained in Tables 1.9 and 1.10 in the Monthly Energy Review. These tables are inserted as worksheets in the commercial_indicators spreadsheets. In the 2020 update, these worksheets are labeled HDD_by_Division18 and CDD_by_Division18.

2. The weather factors are computed in a single worksheet, Weather factors. The weather factors resulting from this process are divided into the intensity indexes to account for fluctuations in heating and cooling degree days. The weather factors are computed for each region and energy type (electricity and fuels) and then combined to produce a single national level factor for each energy type.

3. The first step in updating the weather factors is to add rows at the bottom of the existing series to accommodate the additional set of years. For the 2020 update, rows 74 to 77 were added to the prior update. Row 77 should be considered as a placeholder for a subsequent update. The intensity indicators, based on the SEDS, only went up through 2017. (Note that some places in the worksheet there is a value for 2018, but the intensity value is the same as 2017.)

4. The process for developing weather factors will be discussed in the detail for the Northeast region. (The identical procedure is applied to the other three census regions.) Columns P and Q show heating degree days for the two census divisions in the Northeast census region. The HDD values are computed for the census region, using fixed weights derived from the 1995 CBECS. These weights are shown in the blue cells in row 83. The same calculation is performed for CDD, fixed weights are used to compute a value of CDD for the census region.

5. For each census region, an area of the spreadsheet is used to first compute the weather adjustment factor for electricity, followed by an area to the right that is used to generate an adjustment factor for fuels. In the case of the Northeast census region, the area for the electricity adjustment factor

is comprised of columns Z through AJ. For fuels, the adjustment factor is computed in the columns AM through BE.

6. The values shown the columns shaded in tan (column X) and light blue (column AR) are not used in the current methodology. They are computed strictly as weighted functions of the HDD and/or CDD. These columns should be ignored in the update process.

7. The current methodology for developing weather adjustment factors relies on estimating a simple regression model to fit the regional intensity to a linear function of time (included squared and cubed values of time) and degree days. For electricity the regression model is in terms of a constant term, HDD, CDD, time, time-squared, and time-cubed. The variables are show in columns AA through AE. The dependent variable, electricity intensity, is shown in column AF. The regression coefficients are obtained by the use of the "LINEST" matrix function in Excel. Separate models are estimated for two sub-periods, 1970-1984, and 1985-2017. For the earlier sub-period, the LINEST function comprises the range Z4:AE8. In cell Z4, the following formula is entered {=LINEST(AF29:AF43,AA29:AE43,1,1)}. The first range shows the values of the dependent variable (rows 29 through 43 corresponding to 1970 through 1984). The second cell range shows the set of independent variables. The LINEST regression matrix formulation for the second subperiod is in range Z16:AE16. The regression formula in each of these cells is {=LINEST(AF44:AF76,AA44:AE76,1,1)}. Note the values, AF76 and AE76. The '76' refers to the last row (year) for the regression. As additional years are added, one must go into this formula and update the row with the last year of data and reset the regression model with the SHIFT-CNTRL-ENTER keystroke. (See Excel documentation for executing matrix functions. Once the data range has been redefined, the regression coefficients will automatically update should the data values themselves be changed.

8. An inconvenient feature of the Excel LINEST regression function is that it reverses the order of the coefficients in the results, as compared to the order of the corresponding independent variables. The regression coefficients are in the first row of the LINEST matrix function range. Rows 11 and 23 display coefficients after re-ordering, such that the coefficient is in the same column as the corresponding independent variable.

9. The predicted values from the regression model are shown in column AG. In Column AI, the predicted value of the intensity is computed once again, only in this case using 30-year averages of the HDD and CDD. For the regression model covering 1970-1984, the 30-year average is taken over the period 1961-1990. For the later regression sub-period, the predicted values of intensity employ the HDD and CDD values for the period 1981-2010.

10. The regression-based weather factor is computed as the ratio of the predicted value of the intensity based on actual degree days divided by the predicted value of the intensity based on the long-term averages of the degree days. These values are shown in the yellow-highlighted column AJ. The weather-normalized intensity measures, calculated as the actual intensity divided by the weather factor, are shown in column AH.

11. The same general methodology was followed for developing a weather adjustment factors for fuels. For the Northeast census region, the data and formulas for these factors are shown in columns AM through BE. The set of independent variables for these regressions include: HDD, HDD*Time, Time, Time-squared, and a composite fuel price index. The composite fuel price index was developed as a weighted average of the national distillate fuel oil price index and a national average price for natural gas. The fuel price is actually entered as a three-year distributed lag of the prices, with weights 0.5, 0.33, and 0.17. See columns J and K in the worksheet. In the majority of cases (region and time period) the coefficient on the fuel price was statistically significant. (For the Northeast fuels regression over the

1985-2017 time period, the coefficient was -20.3 with a standard error of 2.1, as shown in cells AT16 and AT17.)

12. The calculational procedures are performed identically for the other census regions, Midwest, South, and West. These procedures are contained in the portion of the worksheet to the right of those for the Northeast.

13. In the columns beginning with IN, the calculations are made to develop a single national weather adjustment factor for electricity, and subsequently for fuels. In columns IP through IS, electricity consumption by census region is shown, transferred from the SEDS_Census_Rgn worksheet. In columns IV through IY, the consumption values are adjusted to reflect the weather adjustment. That is, they are obtained by dividing the reported consumption by the appropriate. The national level weather adjustment factor is shown in column JB, derived as the ratio of the sum of the weather-adjusted regional values divided by the sum of the reported values. The identical procedure is followed for national-level weather factors for fuels. The resulting factors for fuels are shown in column JV. The weather factors are transferred to the worksheet that develops the final intensity estimates for the commercial sector, "Commercial_Total." The factors for fuels are transferred to the fuels section of the indicators in rows 257 through 325, column O. The factors for electricity are in rows 175 through 243, also in column O.

14. The weather factors for delivered energy and source energy are computed implicitly. For delivered energy, they are calculated as the sum of reported electricity and fuels divided by the sum of the weather-adjusted electricity and weather-adjusted fuels. A similar procedure is followed for source energy. As such, the implied weather factors are a result of the process, not an independent variable that influences the values of intensity indexes for delivered energy and source energy. All of these computation occur within the Commercial_Total worksheet.

F. Reclassification of Electricity Sales

The worksheet Adjusted_Supplier_Data shows adjustments made to the reported total commercial electricity sales from EIA. State-level electricity sales data were examined over the period 1990 through 2011. In some states, there occurred significant one-year changes in sales, typically showing up in the commercial and industrial sectors. The changes were often in the same magnitude in these two sectors, but in opposite directions. The general assumption was made that these abrupt changes reflected reclassification of some customers from the industrial rate class to a commercial rate class, or vice versa. Overall, the predominant changes appeared to be from industrial to a commercial classification. These adjustments were made in a separate spreadsheet named Sectoral_reclassification5.xlsx. The general strategy was to adjust the more recent data by removing a constant value of the reclassified level of sales, based on an estimate of such sales in the year in which the reclassification occurred. In essence, the objective was to seek a more consistent time series of electricity sales that would have occurred without these major reclassifications.

The net total of the reclassified electricity sales is shown in column C of worksheet Adjusted_Supplier_Data. These entries reflect the change in industrial electricity sales after the reclassification to the commercial sector. These (typically negative) entries are added to the published national commercial sector electricity sales. Thus, total electricity sales in the commercial sector used to compute the electricity intensity index is roughly 3% smaller in recent years than the sales published by EIA. The adjusted values of reclassified electricity sales are transferred to final intensity indicators worksheet, Commercial_Total (in the cells D175 through D243.

III. Transportation Sector: Process Steps and Technical Notes

Nearly all of the data and calculation procedures used to develop the set of transportation energy intensity indicators are included in the spreadsheet transportation_indicators_*date*, where date is put in mo/day/yr format. The primary data are inserted or processed in one of six major worksheets:

1. **FuelConsump.** This worksheet contains the time series fuel consumption for the major transportation sub-sectors. In this worksheet the data are generally entered in (millions of) gallons or barrels of petroleum. Detailed references for sources used are in the notes below the table.
2. **Fuel Heat Content**. This worksheet shows the assumed Btu content of the various types of petroleum products. No user input is required for updating the indicators.
3. **Passenger-based Energy Use**. This worksheet calculates the fuel consumption in Btu for passenger transportation. No user input is required for updating the indicators (other than adding new rows for most recent years)
4. **Freight-based Energy Use.** This worksheet has the same purpose as the worksheet for passenger-based energy use.
5. **Passenger-based Activity**. This worksheet contains the time series for the activity measures for passenger transportation. Based upon the sources in the notes below the table, the user must input the latest data from various sources.
6. **Freight-based Activity**. This worksheet contains the time series for the activity measures for freight transportation. Based upon the sources in the notes below the table, the user must input the latest data from various sources.

In the documentation of the data sources in the tables in remainder of this section, the objective is to present one or more available sources for each data item. The source or sources actually used in the most recent edition of the energy intensity indicators will typically be indicated by highlighted cells in the tables. Because the data sources often present information related to transportation activity along with energy consumption data, both activity and energy use will be cited together for each individual transportation segment. In general, the discussion below will focus on special issue areas and special adjustments. ***The tables documenting these sources, along with notes alluding to any special data adjustment procedures, are included together at the end of this (transportation) section.***

Passenger Transportation

Passenger transportation is broken out into three major modes: 1) highways, 2) rail, and 3) air. Passenger *highway* transportation is split into segments for personal vehicles and buses.

Highway Passenger – Personal Vehicles

Table A.17 (below) presents the data sources for the personal vehicles portion of the passenger highway segment, comprised of three types of vehicles: cars/short wheelbase vehicles, light trucks/long wheelbase vehicles, and motorcycles. The primary source for the most recent energy use and activity for personal vehicles and motorcycles is Table VM-1 in the publication *Highway Statistics* from Federal Highway Administration.

The comprehensive documentation report (Rev 3 of report) explains the change that was made in the data collection and estimation methodology that was undertaken by the Federal Highway Administration in 2008. This change affected all classifications of vehicles that were reported in previous versions of *Highway Statistics*.

In the current (2020) update to the intensity indicators, the methodology has now been modified such that the latest data from Table VM-1 in *Highway Statistics* is employed in a straightforward fashion. Thus, the user need only input the values shown in that table for total fuel consumption (thousands of gallons) and vehicle-miles (millions). The most recent data for fuel consumption are entered in the appropriate rows for additional years in the worksheet, FuelConsump. The values for vehicle-miles are inserted in columns AJ and AK in the worksheet Passenger-based Activity.

The time series for motorcycle travel and fuel consumption was also affected by the FHWA change in methodology in 2007. The resulting increase in vehicle-miles travelled for motorcycles was in the range of 60 to 70%, compared to the previous methodology. However, because the share of total highway passenger-miles from motorcycles is still estimated to be on the order of 0.5%, no adjustments have yet been made to eliminate the 2006-2007 discontinuity in this series. Fuel consumption for motorcycles from Table VM-1 is entered on the worksheet FuelConsump in column J (in millions, rather than thousands of gallons). Vehicle-mile estimates for motorcycles are entered in column AF.

A factor to be recognized in all segments of highway transportation is that the estimated passenger-miles should all be viewed with some degree of caution. The load factors used to convert from vehicle-miles to passenger-miles are based upon periodic surveys, as documented in Table A.17, and thus carry with them the sampling error associated with any survey. The motivation for using passenger-mile estimates in the system of energy intensity indicators is that the highway estimates can be compared to other travel modes. The detailed, publicly available spreadsheets contain data for vehicle-miles for those analyses for which vehicle-miles (and energy intensity per vehicle-mile, i.e., average miles per gallon) is more appropriate. The following columns in the worksheet Passenger-based Activity show the locations of the load factors for cars/SWB vehicles, light trucks/LWB vehicles, and motorcycles: columns AB and AE. A constant load factor of 1.1 is assumed for all years for motorcycles.

Highway Passenger – Buses

A detailed discussion of the data and methods used for buses was presented in Section, A.4.3 of overall documentation report. Accordingly, Table A.18 only summarizes the major elements from that discussion. To reiterate, buses are divided into three categories: 1) transit, 2) intercity, and 3) school. The historical estimates for fuel use and activity for transit buses are robust, relying on the compilation of data from across the U.S. the Federal Transit Administration and the American Public Transit Association (APTA). As evident in the discussion in the previous section, the data situation with the other two segments of bus transportation is much weaker. For the intercity bus segment, the estimates were generally based upon various "census" reports that have been issued by the American Bus Association. Unfortunately, these reports appear to reflect some inconsistent data collection methods.

For school buses estimates for activity and fuel use were derived from reports from the National Safety Council (in *Accident Facts*) and from estimates developed and published by the Eno Transportation Foundation For the most recent years, energy use for school buses has been extrapolated on the basis of national estimates of route miles traveled by school buses. For the most recent update (2020), no effort was made to undertake another state-by-state compilation of route miles. Thus, both energy use and vehicle-miles have been fixed at their levels for 2011.

Passenger – Rail

Table A.19 shows data source for passenger rail transportation. The American Public Transit Association (APTA) provides detailed data for the rail segment of commuter and transit transportation. As shown in the last column of the table, this data is very accessible from the APTA website, and is published as an appendix in their annual *Fact Book*. The APTA source is particularly good in that it provides a consistent source for both energy use and activity (i.e., estimated passenger-miles) in this segment.

The data for intercity rail (Amtrak) used in the EERE system of energy intensity indicators is taken from the TEDB (Edition 37). The TEDB conveniently shows both historical energy use and activity back to 1971 (the first year of operations for Amtrak).

The National Transportation Statistics data published by the Bureau of Transportation Statistics is an alternative source for some of the data elements. However, as shown in Table A.19, the fuel data shown in the BTS-NTS Table 4-15 combines both transit rail and bus components, and thus does not provide the necessary detail required for the energy intensity indicators system.

Passenger – Air

The sources for the energy use and activity for airlines and "general aviation" are shown in Table A.20. As shown by the highlighted cells, the primary data source is the TEDB. Several issues need to be noted with regard to the construction of energy intensity indicators for airlines. First, the data include both scheduled and nonscheduled activity of U.S. air carriers. Second, the system of energy intensity indicators has sought to develop separate intensity indexes for the passenger and freight segments of major carriers. The highlighted cell in the rightmost column of Table A.20 indicates that an allocation of total fuel use is made on the basis of reported and estimated ton-miles. Passenger-miles are converted to ton-miles by assuming an average 0.2 tons (400 pounds) per passenger-mile. This assumption was based on a comparison of load capacities for the cargo version and the passenger version of a commonly used aircraft in the U.S for both freight and passenger transportation.[1] Given the separate estimates of freight and passenger ton-miles, the fuel data are allocated proportionately between the two categories.

Finally, there is the issue of combining domestic and international operations. In various tables, Edition 37 of the TEDB shows either data associated with domestic use or data combining domestic and international data.[2] For the current set of energy intensity indicators, the intensities are based upon the combination of domestic and international operations (with subsequent allocation between passenger and freight modes). Future updates may attempt to separately characterize domestic and international categories. It should be noted that no consistent set of historical data for both domestic and international operations appears to be presently available for the 2020 update of the intensity indicators.

[1] The largest proportion of jet aircraft currently used by Federal Express is the cargo version of the Boeing 757. The maximum revenue cargo capacity for this plane is 87,700 pounds. A common seating configuration of the passenger version of this plane (757-200PF) is 224 seats. Thus, at full capacity, the load per (revenue) passenger, is about 400 pounds (~ 87,700/224). The value appears to be reasonable given the load associated with each passenger, in terms of the additional weight of the seat, baggage, flight attendants, and other amenities on a passenger aircraft.

[2] For example, Table 9.2 in TEDB-37 shows total energy use for domestic and international operations, while Table 2.13 shows the 2016 energy intensity based solely data for domestic operations. As cited in Table 9.2, fuel use for international operations includes fuel that is purchased outside the U.S.

The bottom three rows of Table A.20 relate to general aviation. As shown in the row with the sources for fuel use, data from TEDB-37 (and the FAA for 2017) were used for the update described in this report. Unfortunately, there is no current source of corresponding information for activity, as measured by passenger-miles. Estimates of passenger-miles up through 2001 were developed by the Eno Transportation Foundation (2002). After that point, the estimates of passenger-miles have been extrapolated by the fuel consumption data provided by TEDB (ultimately, from a survey conducted by the Federal Aviation Administration). This method, of course, results in no change in energy intensity after 2001 for this segment. For 2017, the estimates indicate that general aviation accounted for about 10% of total passenger air transportation energy use, but only about 2.3% of total passenger-miles. At this point, there is no independent corroboration of the relative magnitudes of the average energy intensities across these two categories, although one clearly expects energy intensities in general aviation to be much higher than those associated with the large aircraft used by the major air carriers. Until the FAA requests information in its survey that could be used to estimate travel activity in the general aviation segment, there does not appear to be an easy way to resolve this problem.[1]

Freight Transportation

Freight transportation is disaggregated into four major modes: 1) trucks, 2) rail, 3) air, and 4) water.

Highway Freight – Trucks

Table A.21 presents the data sources and methodology summary for trucks used for freight hauling. Two categories of trucks are presented in the table: medium, single-unit trucks, and heavy combination trucks.

As discussed above, the Federal Highway Administration (FWHA) recently changed the methodology by which it categorized and measured the fuel use and travel volume for all segments of highway travel. This change in methodology resulted in revised estimates for 2007 and later years for all segments of highway transportation. With regard to trucks, the approach to make a historically consistent set of vehicle-mile and fuel consumption estimates was to adjust the pre-2007 data to line up with the more recent data. These adjustments were summarized in Table A.16 in the comprehensive documentation report. In Table A.21 the highlighted cells in the rightmost column related to fuel use and vehicle-miles refer to these adjustments. This adjustment permits the recent and future data to be readily accessed from the various FHWA *Highway Statistics* reports, as shown in Table VM-1 of those publications (and available on-line).

The conversion of total fuel use in gallons to energy use in trillion Btu (TBtu) depends upon the mix of fuels consumed by trucks. As shown in the table, allocation of fuels among diesel, gasoline, and LPG was based on data shown in the TEDB (that was derived originally from various periodic surveys conducted by the Census Bureau.[2])

[1] Two potential solutions to this issue should be noted. The first is to try to reproduce the methodology used by the Eno Transportation Foundation (ETF) in estimating passenger-miles after 2001. Unfortunately, the ETF methodology is itself based upon extrapolating a very old estimate of passenger-miles, published in the 1974 edition of the Interstate Commerce Commission's *Transport Economics*. The second approach would be to estimate energy intensity from a sample of aircraft used in general aviation and try to find an appropriate surrogate measure from the FAA survey of general aviation to estimate changes over time.

[2] For brief description of these surveys, see: http://www.census.gov/econ/overview/se0501.html

The last element in the development of intensity indicators for trucking is to convert estimates of vehicle-miles into ton-miles. For single-unit trucks, no data source is available to that support any particular choice of an average load. For purposes of the energy intensity indicators system, the assumption has been to assign a constant average load of 3 tons. It should be noted that this assumption only has relevance for estimating the total share of total freight that can be attributed to single-unit trucks. The specific magnitude of the load (i.e., 2 tons versus 3 tons, or some other value) has no bearing on the overall energy intensity index for freight transportation, because the indexes across the various modes are weighted by shares of *energy, not activity*.

For combination trucks, there are independent estimates of intercity freight carried by trucks. As Table A.21 shows, these estimates were taken from the Bureau of Transportation Statistics' website containing the on-line version of the National Transportation Statistics (BTS-NTS). The ultimate source of these estimates is the most recent (20th) edition of *Transportation in America*, published in 2007 by the Eno Transportation Foundation (ETF 2007).

This most recent edition from ETF contains estimates of intercity freight ton-miles (here assigned to combination trucks) from 1990 through 2003. These estimates are somewhat higher than ETF's estimates shown in its previous (19th) edition of the report (approximately 16% higher in 1990). Accordingly, the previous estimates from the 19th edition of Transportation in America were used to extrapolate the 1990 estimate from the most recent edition back to 1970. For years beyond 2003, the ton-mile estimates were extrapolated on the basis of the (revised) vehicle-mile estimates from Table VM-1 in *Highway Statistics*. (The implied average load for these most recent years, based upon the vehicle-mile and ton-mile estimates, is just over 7 tons.)

Freight – Rail and Air

Table A.22 shows the data source and methodology summaries for rail and air freight transportation. The data for rail freight is taken from Edition 37 of the TEDB. The primary source for the data is the Association of American Railroads' (AAR) publication *Railroad Facts*. This source provides estimates for both total fuel use and ton-miles of freight carried by the nation's railroads. All fuel is assumed to be diesel in the calculation of energy in TBtu. (It should be noted that *Railroad Facts* must be purchased from AAR.)

The first element for air freight is to estimate the amount of fuel used by major carriers for freight. This estimate is based upon an allocation of total fuel between passenger transportation and freight transportation. As explained above, this allocation is based upon the shares of ton-miles between passenger and freight, where the passenger portion has been estimated on the basis of 0.2 tons per passenger-mile. The historical series of ton-miles and passenger miles relies on data from both the TEDB and downloaded data from the Bureau of Transportation website (as cited in the rightmost cell in the table related to ton-mile estimates.) After the allocation of energy use between passenger and freight transportation, the calculation of historical energy intensities for freight based on freight ton-miles is straightforward.

Freight – Waterborne

The cells in the top three rows of Table A.23 pertain to waterborne freight transportation. Waterborne freight in the system of energy intensity indicators covers the movement of freight within the U.S. by means of inland waterways (lakes, rivers, and canals), as well as along the coasts.

The following description of the situation with respect to waterborne commerce was made in Edition 32 of the TEDB with regard to Table 9.5, "Summary Statistics for Waterborne Commerce":

> The U.S. Army Corps of Engineers Navigation Data Center collects a wealth of waterborne commerce data. Energy use data, however, have never been collected as part of this effort. The energy use data collected by the Energy Information Administration (EIA) on vessel bunkering was formerly displayed on this table. The EIA data include different uses of fuel, not just fuel for domestic waterborne commerce; therefore, it was misleading to display those data together.

However, for the first time in any edition of the TEDB, a set of credible estimates of energy intensity for domestic waterborne commerce were published in Edition 32. These estimates were developed by a researcher at the University of Tennessee and cover the years 1997 through 2010. The specific citation of these estimates was provided in the TEDB as follows:[1]

> ***Energy use –*** Modeled by Chrisman A. Dager, University of Tennessee, Knoxville, using Waterborne Commerce Statistics Center detail records and annual IRS reports on the Inland Waterway Trust Fund tax on diesel fuel used on the inland waterway.
>
> ***Ton-miles –*** Based on detailed records from the U.S. Department of the Army, Army Corps of Engineers, Waterborne Commerce Statistics Center. Includes only ton-miles on taxable waterways

To employ these data in the energy intensity indicators, it was assumed that the sample of waterborne freight used by Dager was representative of all domestic waterborne. Thus, the energy intensities in the Dager analysis could be applied more broadly. As cited above, the Corps of Engineers gathers detailed data on the volume of such freight. Accordingly, an estimate of total energy associated with all domestic freight was calculated by multiplying total ton-miles times the Dager-generated energy intensities.

This methodology provides the appropriate estimates for the years 1997 through 2010, but still leaves the question of how to generate estimates for earlier years. As noted by ORNL above, the EIA data on vessel bunkering fuel cannot be directly associated with the Corps of Engineers data on waterborne freight transportation. However, some previous editions of the TEDB presented proxy estimates of waterborne fuel based upon estimating the amount of total bunkering fuel that is consumed for domestic use.[2] Domestic use was assumed to consume 77.5% of distillate bunkering fuel and 9.3% of residual bunkering fuel. Previous estimates of the energy intensity indicators used these percentages to estimate historical series of energy that was assumed to correspond to the ton-mile magnitudes from the Corps of Engineers.

The implied historical series of energy intensities derived from the adjusted (for domestic) bunkering fuel from EIA indicate that over the period 1970 through 1995, there had been a distinct downward trend in the intensities. However, the year-to-year and cyclical behavior of this series showed implausible changes in intensity that cannot be explained by changes in the energy efficiency of the vessels themselves. Nevertheless, there remains the strong presumption that over this period, some improvement in overall vessel or system energy efficiency warrants a method that in some manner reflects this change.

In this case, the methodology undertaken for the system of intensity indicators was to represent the changes in energy intensity in waterborne freight over the 1970-1997 time period as a constant percentage decline in each year. This average rate of decline is based on a simple logarithmic trend regression of the implied energy intensities. This trend regression yielded a coefficient of -0.013. The

[1] ORNL, TEDB-32, Appendix A, p. A-32.
[2] For example: ORNL, *Transportation Energy Data Book*, Edition 20, ORNL-6959, October 2000, p. A-15

associated t-statistic of 4.1 indicates the presence of a strong downward trend over this overall time frame, in spite of implausible behavior over short sub-periods within.

Clearly, there is considerable uncertainty, as reflected in the ORNL statement in the TEDB-32, that it is problematic to link the EIA fuel consumption data with the freight ton-mile statistics. One can argue however that the vast majority of the bunkering fuel not used for domestic freight was used for international waterborne freight. Moreover, it seems evident that the growth in international trade, and concomitant freight energy use, has been greater than the growth in the domestic economy and waterborne commerce within the U.S. Thus, any historical series that is based upon *constant percentages* of total bunkering fuel is likely to overstate the growth in fuel use for domestic freight activity, and conversely understate the decline in energy intensities for domestic freight transportation since 1970.

As a final step, the trend rate of reduction in the implied energy intensities was set at -1.0% per year, rather than the -1.3% estimated in the regression. This further step is taken to help ensure that the energy intensity indicators derived in the manner is likely to represent a conservative estimate of the overall decline over the 1970-1997 time period. Figure A.1 shows the historical series of energy intensities implied by employing the trend rate from 1970 through 1997, followed by the intensities by measured by Dager at the University of Tennessee.

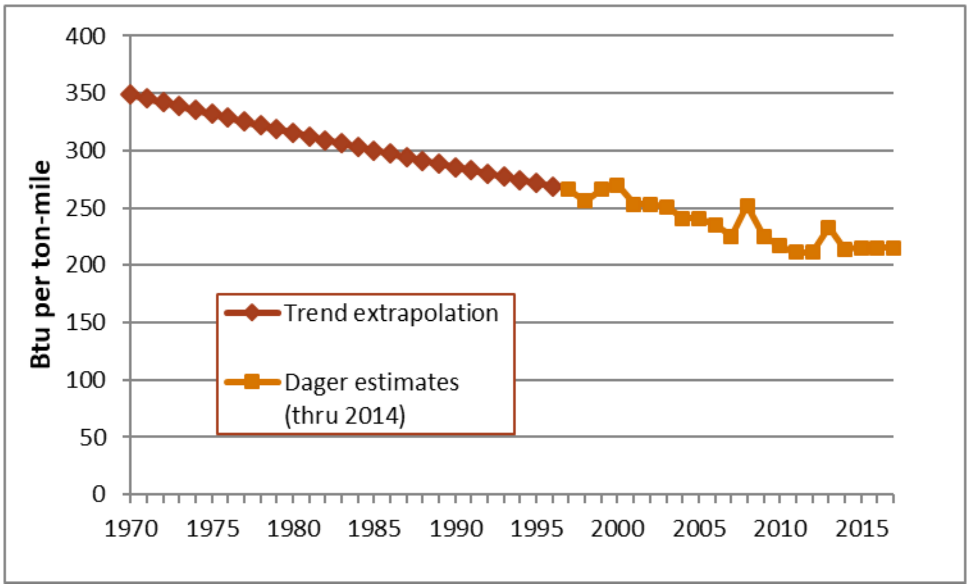

Figure A. 1. Estimated energy intensities for domestic waterborne transportation

One final observation is warranted with regard to the energy intensities benchmarked to the Dager estimates. For 2010 the energy intensity for waterborne transportation was 217 Btu/ton-mile, compared the estimate for rail freight of 289 Btu/ton-mile. Thus, at least qualitatively, this value reflects the supposition that the energy required to move freight by water is lower than that by land transportation. (Note: the Dager value is roughly one-half the magnitude of the previous estimate of waterborne intensity based on the (adjusted) EIA vessel bunkering fuel data.)

Intensity estimates are available through 2014 from the Transportation Energy Data Book, Ed. 37, Table 2.16. The values for 2013 and 214 are 233 and 214 (Btu per ton-mile). For the years 2015-2017, a fixed value of 215 was assumed. For updates, this series must be extended in the transportation indicators spreadsheet. The specific location is column in the worksheet Freight-based Energy Use.

Freight - Pipelines

The bottom portion of Table A.23 shows the sources and methodology for developing intensity indicators for pipelines. Currently, only natural gas pipelines are included in the system of energy intensity indicators. As shown in the first row devoted to gas pipelines, the method seeks to estimate both the natural gas and electricity used to transport natural gas through the nation's pipelines. The rightmost cell in this row displays the source for the natural gas consumption (EIA's *Annual Energy Review*, Table 6.5, *Monthly Energy Review,* Table 4.3). The electricity consumption is based upon engineering analysis that was conducted by ORNL and is reported in Table A.12 of TEDB-37. As shown in the following row (rightmost cell), electricity consumption is converted to (source) Btu with a constant conversion factor of 10,339 Btu/kWh (following the convention in TEDB-37).

The final row of the table presents the methodology used to calculate ton-miles for pipeline transportation of natural gas. The methodology here again depends upon an engineering analysis. Essentially, the conversion depends upon an estimate of the weight of a given quantity of natural gas (at a defined pressure) and the average distance of transport.

At this point, oil pipelines are excluded from the system of energy intensity indicators, because there are not suitable historical time series of energy consumption *and* quantities of oil transported. ORNL apparently included the energy (electricity) used by oil pipelines in their estimates of total pipeline use in Table 9.1 of the TEDB-32, but according to other documentation in the TEDB, the estimate has remained constant over time.[1] With regard to the quantity of oil transported by pipelines, the most recent data on the BTS-NTS website (Table 1-61) provides estimates of ton-miles for both crude petroleum and refined petroleum products (provided by the Association of Oil Pipelines).

Table A.1. Data sources for highway passenger transportation – personal vehicles

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Passenger cars, short wheelbase vehicles – total fuel (gallons)	TEDB-37, Table 4.1. Data through 2009 for cars.		Federal Highway Administration, Highway Statistics, Table VM-1; data for 2007 and later for short wheelbase vehicles
Passenger cars, short wheelbase vehicles – total energy (TBtu)	Total fuel allocated among gasoline, gasohol, and diesel. Source: TEDB-37, Table A.18		
Passenger cars short wheelbase vehicles - vehicle-miles	TEDB-37, Table 4.1 Data through 2009 for cars	BTS Table 1-35 (matches TEDB through 2006)	Federal Highway Administration, Highway Statistics, Table VM-1; data for 2007 and later for short wheelbase vehicles
Passenger cars – Passenger-miles	Combine vehicle-miles and load factors from TEDB-37, Table A.18	BTS Table 1-40 (Data does not match calculation using TEDB data prior to	Load factors from Nationwide Personal Transportation Surveys and

[1] Specifically, page A-18 of Appendix A in the TEDB-37 discusses the methodologies to estimate pipeline energy use.

		2007. After 2007 data is for short wheel base vehicles, not consistent with earlier data, see text discussion)	National Household Travel Survey (See TEDB-37 for sources)
Light trucks, long wheelbase vehicles –total fuel (gallons)	TEDB-37, Table 4.2		Federal Highway Administration, Highway Statistics, Table VM-1; data for 2007 and later for long wheelbase vehicles
Light trucks, long wheelbase vehicles – total energy (TBtu)	Total fuel allocated among gasoline, gasohol, and diesel. Source: TEDB-37, Table A.5		
Light trucks, long wheelbase vehicles – vehicle-miles	TEDB-37, Table 4.2 Data through 2009 for light trucks	BTS Table 1-35 (matches TEDB through 2006)	Federal Highway Administration, Highway Statistics, Table VM-1; data for 2007 and later for long wheelbase vehicles
Light trucks, long wheelbase vehicles – passenger-miles	Combine vehicle-miles and load factors from TEDB-37, Table A.19		Values interpolated between 2001 and 2008
Motorcycles – total fuel (gallons)	TEDB-37, Table A.2		Highway Statistics, Table VM-1
Motorcycles – total energy (TBtu)	Assume all motorcycle fuel is gasoline		Assume all motorcycle fuel is gasoline
Motorcycles – vehicle-miles	Table 2.13 for most current year, time series not available in TEDB-30 or TEDB-37	BTS Table 1-35 (with some interpolation prior to 1990)	Highway Statistics, Table VM-1
Motorcycles – passenger-miles	Table 2.13 for most current year (not used)	BTS Table 1-40 (not used, implied load factors are not stable/plausible in historical series	Assumed load factor of 1.1 (for all years) x vehicle-miles

Table A.2. Data sources for highway passenger transportation - buses

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Transit buses – Fuel use (gallons by type of fuel)	TEDB-37, Table A.3	BTS-NTS, Table 4-15 (Includes fuel and electricity use for all commuters and transit, buses + rail), Not used.	American Public Transit Association (APTA), Public Transportation Fact Book. https://www.apta.com/research-technical-resources/transit-statistics/public-transportation-fact-book/
Transit buses – Energy use	2016 estimate in TEDB-37, Table 2.13		Apply conversion factors (Btu/gallon). Note: APTA reports CNG in diesel-equivalent. Biodiesel conversion factor from AER 2008, p. 373
Transit buses – Passenger-miles	TEDB-37, Table 7.9	BTS-NTS Table 1-40	Same source as for fuel use, APTA website.

		(data for most years excludes ferries, no exact match with APTA data, not used.	
Intercity buses, Fuel use (gallons)	TEDB-37, Table A.4. Data for 2011 estimated based upon extrapolation of 2000 Eno estimate. Not used.		Primary source is American Bus Association (ABA) with earlier data in TEDB-19, and later data from ABA reports. Fuel use was estimated on basis of separate estimates of average MPG and vehicle-miles. See text.
Intercity buses – Energy use (TBtu)			All fuel for intercity buses assumed to be diesel.
Intercity Buses – Passenger-miles	Not estimated	BTS Annual Report, published in September 1993. Table 6 for estimates from 1970 through 1991. For 1991-1997, TEDB-23, Table 5.23.	For 1998-2006, estimates of vehicle-miles interpolated between 1997 value from TEDB-19 and ABA 2007 value. Average passenger load factors interpolated over same period. Passenger-miles computed as load factor x vehicle miles. Most recent estimates from ABA reports. See text.
School buses – Fuel use (gallons)	TEDB-37 Table A.4 for years 1970-1994, at typically five-year intervals. Data after 1994 extrapolated by vehicle-miles (not used)	Data (estimates) for school buses not separately identified	Estimates of fuel use interpolated for years other than ending in 0 or 5, based upon estimates of vehicle-miles and average fuel economy. The approach was used over the period 1970-1994. See text.
School buses – Energy use (TBtu)	TEDB-37, Table A.4. Fuel assumed to be 90% diesel, 10% gasoline		
School buses – Passenger-miles	1980-1994: TEDB, various issues for vehicle-miles. No estimates for recent years.	Data (estimates) for school buses not separately identified	1995-2011: 1994 estimate extrapolated from U.S. route mileage for public school transportation, used by permission from *School Bus Fleet* magazine, Fact Books. Passenger-miles calculated from assumed constant pupil load of 23. See text in Rev 3 report.

Table A.3. Data sources for passenger rail transportation

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Commuter rail – Fuel use (kWh, Gallons)	TEDB-37, Table A.14	BTS-NTS, Table 4-15 (Includes fuel and electricity use for all commuters and transit, buses + rail), Not used.	American Public Transit Association (APTA), Public Transportation Fact Book. https://www.apta.com/research-technical-resources/transit-statistics/public-transportation-fact-book. See links to data.
Commuter rail – Energy use – TBtu	TEDB-37, Tables 9.11 and 9.12. (Data from APTA) (matches direct calculation with APTA data)		Conversion to TBtu with electricity and diesel energy conversion factors (electricity conversion at 10,339 Btu/kWh from TEDB
Commuter rail – Passenger-miles	TEDB-37, Tables 9.11 and 9.12. (Data from APTA)	BTS-NTS, Table 1-40	APTA source as above for fuel use. Passenger-miles in Table 3 of Appendix A of APTA publication.
Transit rail – Fuel use (kWh, Gallons) (combines heavy and light rail)	TEDB-37, Table A.15	BTS-NTS, Table 4-15 (Includes fuel and electricity use for all transit, buses + rail), Not used.	Same source as commuter rail (APTA, Fact Book), Tables 38 and 39 in Appendix A.
Transit rail – Energy use – Tbtu	TEDB-37, Table 9.12 and 9.12. (Data from APTA) (matches direct calculation with APTA data)		Conversion to TBtu as for commuter rail
Commuter rail – Passenger-miles	TEDB-37, Table 9.12 and 9.12. (Data from APTA)	BTS-NTS, Table 1-40	APTA source as above for fuel use. Passenger-miles in Table 3 of Appendix
Intercity rail (Amtrak) – Fuel use (kWh, Gallons)	1994-2016: TEDB-37, Table A.16; prior data extrapolated from TEDB-30, Table 9.10	BTS-NTS, Table 4-18 From: www.bts.gov/content/amtrak-fuel-consumption-and-travel-1	Assume 1993 and 1994 have same energy use to account for discontinuity in 1994.
Intercity rail – Energy use (TBtu)	TEDB-37, Table 9.10		Conversion to TBtu as for commuter rail
Intercity rail – Passenger-miles	TEDB-37, Table 9.10	BTS-NTS, Table 1-40	Revenue passenger-miles (does not include contract commuter passengers)

Table A.4. Data sources and methodology for airline passenger transportation

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Air carriers – Fuel use (gallons)	TEDB-37, Table A.9	BTS-NTS, Table 4-8 (Another source cited by TEDB is: www.transtats.bts.gov/fuel.asp	1) All fuel for international flights for domestic carriers is included with fuel for domestic operations. This permits the more accurate calculation of energy intensity. Data for 2017 from BTS-NTS used to extrapolate 2016 value from the TEDB-37.
Air carriers – Energy use	TEDB-37 Table 9.2 includes all energy for domestic and international operations, (Table 2.12 includes only domestic air service)		1)All fuel is assumed to be jet fuel 2) Allocation between passenger and freight transportation based on ton-miles. 1 pass-mile = 0.1 ton-mile, as per BTS. Passenger percentage of ton-mile ~ 83% in 2017,
Air carriers – Passenger-miles	TEDB-37, Table 9.2, and previous TEDB's for years before 1985	BTS-NTS Table 1-40. Table 1-40 contains only data for domestic operations and does not match other sources. Not used.	
General aviation – Fuel use (gallons)	TEDB-37, Table A.10		Department of Transportation, Federal Aviation Administration, *General Aviation and Avionics Survey* *https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2017/*
General aviation – Energy use (TBtu)	TEDB-37, Table 9.3		Conversion to Btu with factors for aviation fuel and jet fuel (120,2 and 135.0 kBtu/gallon, respectively.
General aviation – Passenger-miles	Not available	Not available	1970-2001: Eno Transportation Foundation, *Transportation in America 2001*, 19th *Edition*, p.45 Passenger-miles after 2001 extrapolated by total energy use.

Table A.5. Data sources and methodology for highway freight transportation

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Single-unit trucks, Class 3 to 8 – fuel use (gallons)	TEDB-37, Table 5.1 Table shows discontinuity from FWHA methodology change in 2007, not used.		Pre-2007 revised to match current FHWA methodology, see text and Table A.13.
Single-unit trucks, energy use (TBtu)	Total fuel allocated among gasoline, gasohol, and diesel. Source: TEDB-37, Table A.6		Implausible estimate from 1982 Truck Inventory and Use survey (TIUS) showing diesel at only 60% of truck fuel. Changed to remain above 80% in years proximate to 1982.
Single-unit trucks, vehicle-miles	TEDB-37, Table 5.1 Table shows discontinuity from FHWA methodology change in 2007, not used.	BTS-NTS Table 1-35 (Carries discontinuity from FHWA methodology change in 2007), not used.	Pre-2007 revised to match current FHWA methodology, see text and Table A.13 in this report..
Single-unit trucks, ton-miles	Not estimated by ORNL in TEDB	Not estimated by BTS	Estimate based upon assumed average load of 3 tons per truck (x vehicle-miles)
Combination trucks, Class 7-8, fuel use (gallons)	TEDB-37, Table 5.2 Table shows discontinuity from FWHA methodology change in 2007, not used.		Pre-2007 revised to match current FHWA methodology, see text and Table A.13.
Combination trucks, Class 7-8, energy use (TBtu)	Total fuel allocated among gasoline, gasohol, and diesel. Source: TEDB-37, Table A.6		See note above for single-unit trucks
Combination trucks, Class 7-8, ton-miles	Not estimated by ORNL in TEDB	BTS-NTS Table 1-49 BTS publishes ton-miles for intercity truck (1990-2003) from most recent edition of *Transportation in America*, published by Eno Transportation Foundation in 2007. Column to right shows data used to extrapolate before 1990 and after 2003	1970-1989: Eno Transportation Foundation, *Transportation in America 2001*, 19th *Edition*, p.42 Used to extrapolate 1990 estimate back to 1970. Extrapolate 2003 ton-miles with vehicle-miles

Table A.6. Data sources and methodology for rail and air freight transportation

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Rail – Fuel use (gallons)	TEDB-37, Table A.13		Association of American Railroads, *Railroad Facts*, 2011 Edition
Rail – Energy use (TBtu)	TEDB-37, Table 9.8		Convert to TBtu with conversion factors for diesel fuel (matches TEDB-37 estimates)
Rail – Ton-miles	TEDB-37, Table 9.8	BTS-NTS, Table 4-25	Association of American Railroads, *Railroad Facts*, 2011 Edition, more recent issues will possibly require purchase
Air carriers – Fuel use (gallons)	TEDB-37, Table A.9, Reports fuel use for both domestic and international operations		See note in cell below – freight fuel use based on allocation of total fuel use allocated on basis of ton-miles for passengers and ton-miles for freight
Air carriers – Energy use for freight	TEDB-37 Table 9.2 includes all energy for international + domestic operations		1)All fuel is assumed to be jet fuel 2) Allocation between passenger and freight transportation based on ton-miles. 1 pass-mile = 0.1 ton-mile, as per BTS. Passenger percentage ~ 70% in 2011.
Air carriers – Ton-miles	Not reported, no intensities estimated for air freight. Table 9.2 contains historical data on revenue cargo ton-miles.	BTS-NTS, Table 1-46a. (Reports only freight from domestic operations, not used)	BTS, Airline Data and Statistics: revenue ton-miles, http://www.bts.gov/xml/air_traffic/src/index.xml#CustomizeTable Data for both domestic and international operations

Table A.7. Data sources and methodology for water and pipeline freight transportation

Transportation Segment	Transportation Energy Data Book (Ed. 37)	Bureau of Transportation Statistics (National Transportation Statistics)	Primary or Other Source/Methodological Note
Waterborne – Fuel use (gallons)	TEDB-37, Table A.10 (Used for pre-1997 regression, see text)		
Waterborne – Energy use (TBtu)	Not available, only intensities published in TEDB-37, Table 2.15		TEDB-32, Table 2.15 presents improved estimates of intensity (Btu/ton-mile). Intensities used with domestic ton-miles to estimate historical energy consumption. See text.
Waterborne – Ton-miles	TEDB-37, Table 9.5	BTS-NTS, Table 1-56, only tons shipped rather than ton-miles	
Natural gas pipelines – Fuel use (cubic feet and kWh)	TEDB-37, Table A.12, Reports natural gas and electricity. Gas from EIA, electricity is estimated, see note in TEDB		EIA, *Annual Energy Review 2011*, Table 6.5 (Natural gas used as fuel in delivery to customers) More recent data from *Monthly Energy Review*, Table 4.3
Natural gas pipelines – Energy (TBtu)	TEDB-37, Table 2.7 shows energy consumption for all pipelines for 2016, however electricity use for oil pipelines and coal slurry is an old estimate, and held constant		Conversion to energy units with factor of 1,031 Btu/cu.ft for gas, and 10,339 Btu/kWh for electricity
Natural gas pipelines – ton-miles		Not estimated or reported	Natural gas converted to tons using methane density of 0.0448 lb/cu.ft. Average length of travel for natural gas assumed to be 620 miles.
Oil pipelines	TEDB-37, p. A-18 includes discussion of energy use for oil pipelines	BTS-NTS, Table 1-61 contains statistics on ton-miles for both crude petroleum and refined oil products	No energy intensity indicator, as there are no historical series for both energy use and ton-miles of oil transported through pipelines

IV. Industrial Sector: Process Steps and Technical Notes

Background

The energy intensity indicators for industrial sector are developed separately for 18 separate industries in manufacturing sector and for three subsectors in nonmanufacturing: agriculture, mining, and construction.

The available data sources for historical energy consumption differ between manufacturing and nonmanufacturing. The principal data source for manufacturing is the periodic Manufacturing Energy Consumption Survey (MECS), conducted by EIA since 1985. This information is supplemented for non-MECS years with estimates derived from the Annual Survey of Manufactures (ASM) and the Economic Census (EC) conducted every five years. In the non-manufacturing sectors, the primary data sources include the Economic Census, and surveys of farms conducted by the U.S. Department of Agriculture. (In earlier years, the Economic Census was carried out under separate labels for broad sectors, e.g., Census of Manufactures, Census of Agriculture, and Census of Mining. For the discussion below, all Census efforts conducted at 5-years intervals will be referred to as falling under the more recent general term Economic Census, EC.)

For years prior to 1985, a primary data source is the National Energy Accounts (NEA), a project sponsored by the Department of Energy and the Department of Commerce until the mid-1980s. The energy consumption data in the manufacturing and nonmanufacturing sectors were based upon the NEA and other economic census data over the period up through 1985.

The energy intensity indicators for the industrial sector are constructed within a single spreadsheet, industrial_indicators_date.xls, where date is expressed in mo/day/yr format. The hierarchical system used in the EERE set of energy intensity indicators disaggregates the industrial sector into two major sub-sectors, manufacturing and nonmanufacturing. The manufacturing sector is comprised of 18 sectors, primarily corresponding to the 3-digit level of classification used in the North American Industrial Classification System (NAICS). The nonmanufacturing sector recognizes three sub-sectors: agriculture, mining, and construction. The indicators for the industrial sector are not derived from the supply-side consumption data collected by EIA. That data has no sectoral disaggregation and includes energy use for petrochemical feedstocks (and other non-fuel uses of energy). As mentioned above, the estimates for energy consumption since 1985 in manufacturing are derived from the periodic Manufacturing Energy Consumption Surveys (MECS), supplemented by energy purchase information collected by Census Bureau.

Manufacturing

This section will describe the various steps that required to update the energy and output data that are used to develop the energy intensity indicators in manufacturing. The overall process is complex, given the objective of producing annual intensity estimates without the availability of direct annual data on the quantities of fuels consumption by manufacturing sector.

Historically, two changes in the collection of statistics related to manufacturing energy consumption had severe consequences to the development of consistent energy intensity indicators for manufacturing over time. First, in the wake of the energy crises of the 1970s, the Census Bureau added questions to its Annual Survey of Manufactures (ASM) that requested very specific data on the cost and quantity of individual fuels. Unfortunately, the questions pertaining to the quantity of fuels were dropped from the survey after 1982. In response to this deficiency, the Energy Information Administration conducted its first Manufacturing Energy Consumption (MECS) in 1985. The MECS, however, did not have the same level of coverage as the ASM and Census, and it has been conducted only every three to four years.

Second, the U.S. government adopted the NAICS in 1997 to align the classification of statistics for businesses within Canada, Mexico, and the U.S. The NAICS also recognized the emergence of important new industries that were either small or non-existent when the previous Standard Industrial Classification (SIC) was first adopted in 1937 (and through later editions up through 1987). With regard to developing consistent time series of energy intensity indicators, the changeover from the SIC to the NAICS presented a particular challenge.

In part, with these developments in mind, the methodology to develop indicators, particularly for manufacturing, has undergone several revisions over time. With the framework of the spreadsheets used to construct these indicators, these revisions have resulted in many portions of within these spreadsheets that are no longer relevant (and sometimes duplicative) to the current methodology. In the previous two major updates to the indicators, in 2014 and 2017, there were insufficient resources to develop a more transparent flow of the various calculational elements used to produce the final indicators.

With this understanding in mind, the overall method to construct the indicators for the manufacturing sector involves three major separate spreadsheets. The principal spreadsheet used to collect input information from the MECS and the ASM/Census is ASMdata_date.xlsx, where, as before, date is expressed in two-digit format mo/day/yr. The most recent version of this spreadsheet ASMdata_010220.xlsx.) The other major spreadsheet is named ind_hap3_122219. This spreadsheet contains the pre-1985 data from the NEA and is used to link this earlier data to the more recent estimates from the ASMdata spreadsheet. The discussion in the following sections provide the detailed steps by which the historical energy and output data are constructed.

A. Obtain data from most recent MECS

The MECS provides necessary information on both the consumption of fuels as well as the average prices paid for fuels by establishment within each 3-digit NAICS sector. The process by which the fuel consumption data is obtained is discussed first.

A.1 Download and Process Energy Consumption Tables from MECS

With respect to energy consumption, several tables from the MECS must be downloaded. Table 4.2 is "Offsite-Produced Fuel Consumption," where the units are expressed in Trillion Btu. The second is Table 3.2 (or, alternatively, Table 3.1). The process by which the data from the most recent 2014 MECS is described below. In general, similar procedures were used for the prior MECS.

Download and process data from Table 4.2

These tables for the 2014 MECS can be found here: https://www.eia.gov/consumption/manufacturing/data/2014/. The consumption tables are located here: The downloaded file for Table 4.2 is simply Table4_2.xlsx. (The filename may have an additional identifier if it is the second or later time it is downloaded. Thus, an inadvertent third download may be named Table4_2 (3).xlsx. On most computers with Windows 10, the default system folder into which the downloaded file is put is labeled "Downloads."

It is suggested that a suffix be appended to each downloaded file to identify the year and differentiate it from the original, downloaded file. For the most recent update, the filename is table4_2_2014_pnnl.xlsx. Most of the tables from the MECS are processed in a similar manner; the description below pertains to Table 4.2. .Each row in the table shows a specific manufacturing sector, either the 3-digit NAICS classification level, or at the six-digit NAICS level (if the latter, as a sub-sector within the appropriate 3-digit sector.) Panels of industry data are shown with the national table at the top of the worksheet ("Table 4.2"), followed by panels for the census regions stacked in vertical order.

The user should create new rows in the worksheet that contain only the data at the 3-digit level. Two sets of rows where the 3-digit data will be processed should be created. The first of these sets contains the identical 3-digit data from the appropriate rows in the full table produced by EIA.

The 3-digit only NAICS extract tables will contain many entries with "*", EIA's designation whose values are less than 0.5. In some cases, though not typically at the 3-digit NAICS level, the data are withheld ("W") due to confidentiality concerns.

In order to utilize the MECS data for the indicators, these missing values must be imputed in some manner. A second table with the 3-digit entries should be created just below the first. This table contains special appended rows and columns to 1) aggregate the values by fuel for all manufacturing (as column totals), aggregate total fuel consumption by 3-digit sector (row totals). The special rows at the bottom include the published total consumption by fuel (from the original EIA table) and the calculated total after the missing entries are assigned a value. A similar set of columns is appended to the right of the table. (For this table table4_2_2014_pnnl.xlsx, a third column was added that summed the row totals with the "*" values (interpreted as 0) by Excel).

The process to fill in the missing values is one of analyst judgement. The objective to fill in the totals such that both column and row totals approximately match the EIA published totals, typically to within a range of 1 or 2 trillion Btu. When the missing data are known to be less than 0.5, then it becomes a judgment whether the size of the industry, at least with respect to the magnitude of published consumption for other fuels, is more likely to be near 0 or closer to 0.5 cut-off value. To identify those cells that have been manually imputed, they should be highlighted with some color. In the case of the data in Table 4.2 for 2014, the cells with imputed entries are highlighted in orange.

In general, the imputation process need not require great care. If the imputations are for entries less than 0.5 trillion Btu, errors in assigning a value for specific industries have negligible impact on the final indicators. One should note as well is that the intensity indicators consider only the aggregate of total fossil fuel consumption, not consumption for specific fuels. Consumption by fuel has only an impact in developing fuel price deflators for non-MECS year data. The small values that are imputed for specific fuels will have essentially no impact on the value of the deflator for total fossil fuel consumption.

Download and transfer MECS total fuel consumption – Tables 3.1 and 3.2

Tables 3.1 and 3.2 in the MECS shows the values for total fuel consumption, combining both off-site and on-site sources. The first part of the notes provided for these tables. (Either table can be used to develop the total fuel consumption.

Notes: Totals may not equal sum of components because of independent rounding. The estimates presented in this table are for the total consumption of energy (formerly total inputs of energy) for the production of heat, power, and electricity generation, regardless of where the energy was produced. Specifically, the estimates include the quantities of energy that were originally produced offsite and purchased by or transferred to the establishment, plus those that were produced onsite from other energy or input materials not classified as energy, or were extracted from captive (onsite) mines or wells.

Note that the fuel consumption from this table is for heat, power, and electricity generation, end uses consistent with the concept of energy efficiency as reflected in the intensity indicators. In addition, the fuel consumption from this table in the MECS was deemed to be more consistent with the measures of energy consumption reported by the EIA in their supply-side sources. (To be specific, Table 2.4, Industrial Energy Consumption, in the Monthly Energy Review.)

Some mention of the difference between Table 3.1 and Table 3.2 is warranted. Table 3.1 shows energy consumption by fuel in physical units, including the total across all fuels expressed in trillion Btu and electricity in kWh. From Table 3.1, total fuel consumption in Btu can be calculated as difference between total energy and electricity consumption after conversion to Btu. Table 3.2 only differs from Table 3.1 by showing all fuel types in Btu.

For 2014, the values for total energy consumption and electricity consumption, both defined in terms of trillion Btu, from Table 3.2 are transferred to spreadsheet ind_hap3. Worksheet MECS_Fuel in this spreadsheet has been used to collect the fuel consumption estimates for all the MECS dating back to the first MECS in 1985. The 2014 data are located in the cell range F218:F238.

The first six NAICS sectors are aggregated into three sectors (311-312, 313-314, and 315-316) as a part of the set of manufacturing indicators. The energy consumption data under this revised sectoring classification are shown in the columns to the right, columns O and P.

A.2 Download and Process Fuel Prices from MECS

Several aspects of this activity warrant discussion.

Download and Process Energy Price Tables from MECS

The process for downloading and processing the energy price information from the MECS is similar to that for consumption. The energy prices (computed as the total cost of purchased fuel divided by the quantity of fuels) are from MECS Table 7.2. As in the case of the consumption data, the user must select out the data at the 3-digit NAICS level from the more detailed sector classification in the table provided by EIA. The most recent processing of the price data for the 2014 MECS is shown in table 7_2_2014_pnnl.xlsx. The added rows that show only the 3-digit data are rows 105 through 126. Not all the fuel type information is required for the intensity indicators framework. The required fuel data is shown in the yellow-highlighted columns to the right of the data from EIA, columns AR through BD.

Transfer MECS data into MECS_Prices Spreadsheet

The information from Tables 4.2 and 7.2 are transferred into the spreadsheet MECS_prices_date. (For the 2020 update, the file with the date suffix is MECS_prices_122419. The first worksheet, MECS_data, contains data from the historical MECS starting in 1998. The consumption data from Table 4.2 from the 2014 MECS are linked to cell range D118:M140. The price data from Table 7.2 are linked to cells to the right, in range R118:Y138. Table 7.2 contains non-numeric entries for prices that have been withheld by EIA. These missing values have been imputed based on analyst judgment and are shown in cells shaded in an orange-tan color.

The quantity data from Table 4.2 are used to calculate fuel shares. These shares are used in subsequent calculations to develop of composite fuel price for each 3-digit NAICS sector (i.e., an overall cost per MMBtu for all fuels.) These shares are shown in the yellow-highlighted cells in columns AR through AX.

The prices are transferred to worksheet MECS_EnergyPrices in the same spreadsheet. This worksheet shows the historical prices for each fuel. The price values from the 2014 MECS are linked to the appropriate fuel in column K.

B. Develop estimates of fuels consumption for non-MECS years

The MECS produces several tables related to quantity of fuel consumption: a) total for all purposes, b) non-fuel use (e.g., feedstocks), c) fuel use (for heat, power, and electricity generation), and d) offsite-produced fuel consumption. For purposes of the energy intensity indicators work, the appropriate table involves the fuel use in item c), fuel consumption for heat, power, or electricity, irrespective of whether the fuel is purchased from offsite suppliers or is derived as a by-product of the manufacturing process. As an indicator of energy efficiency, fuels used as materials such as feedstocks or asphalt are not relevant. Beginning in 1998, the MECS reported this fuel use in its Table 3.2. (In prior MECS, the relevant table was Table A4).

The major issue with regard to fuel use is how to estimate fuel consumption for non-MECS years. As stated earlier, after 1981 the ASM and the EC report only the cost of purchased fuels, with no information about quantities. As a result, the following five-step methodology was followed in an effort to best characterize the fuel consumption in those years for which MECS data are not available.

B.1. Estimate total cost of purchased fuels (offsite-produced) in MECS years.

Over its 25-year history, the MECS has generally published estimates for several aggregates concerning total fuel use as listed in items a) through d) above. For the last item (d), EIA indicates that this value can be related to historical data collected by the U.S. Census Bureau as part of its Annual Survey of Manufactures and Census of Manufactures. As stated in Appendix A of the 1988 MECS, EIA describes "Energy consumed onsite as a fuel and produced offsite" as:

This derived value represents onsite consumption of fuels that were originally produced offsite. That is, they arrived at the establishment as the result of a purchase or were transferred to the establishment from outside sources. As such, this derived value is definitionally equivalent to "consumption of purchased" fuels reported by the Census Bureau for the years 1974-1981. The Census Bureau defines "purchased fuels" to include those actually purchased plus those transferred in from other establishments (EIA 1991, p. 139)

While dropping the request for quantity information on purchased fuels after 1981, the ASM and EC have continued to include the question related to total cost of such fuels. What the method here seeks to do is develop a MECS-based "cost of purchased fuels" that can be compared to the cost of purchased fuels obtained from the census data. From that point, the ASM purchased fuel data can be used to interpolate the definitionally-equivalent MECS values for non-MECS years and then finally apply a composite fuel cost ($/MMBtu) to derive quantity estimates.

The following paragraphs reiterate the general process that was described in Section A above. This first step is to obtain the MECS consumption data by fuel and 3-digit NAICS industry. Over time, the MECS table numbers providing this information have changed. In the most recent MECS, these data, converted to Btu, are in Table 4.2. Unfortunately, some small quantities of some fuel types are not published for some sectors. Using the row (NAICS sector) totals and quantity totals (total consumption by fuel), as well as any recent MECS where such data were not withheld, judgmental estimates of the missing fuels must be made. This process was explained in detail in Section A above. As mentioned there, this imputation process is not likely to affect the overall intensity estimates to any large degree because the withheld information usually appears to involve small quantities.

Next, the information on prices paid by fuel is obtained. In each of the historical MECS information on fuel prices by fuel and sector have been published. In the recent MECS, the prices in terms of dollar per MMBtu are in Table 7.2. As discussed above, this information is available from EIA on its website: https://www.eia.gov/consumption/manufacturing/data/2014/#r4

The prices are really the average costs paid, as they are the quotient of expenditures for purchased energy sources (Table 7.9 in the recent MECS) and total quantity of purchased energy sources (Table 7.6). Some of the purchased energy sources are not used for heat and power (i.e., feedstocks or material inputs). However, it is assumed that the average price paid is the same for all uses of the fuel.

With quantities and prices in hand, a synthetic estimate of the total cost of purchased fuels for “heat, power, and electricity production” is developed for each 3-digit NAICS sector. For most sectors, this estimated value is very close to the published figures for total cost of purchased energy sources (Table 7.9), because most, if not all, of the purchased energy sources are used as fuel.

A bridge between the NAICS sectors (1998 and later) and the 2-digit Standard Industrial Classification (used in the MECS from 1985 through 1994) was developed based on a detailed analysis of electricity quantity data. From this bridge, fuel quantities for all MECS on a SIC basis were assigned to a NAICS classification. In addition, prices based on the SIC were assigned to NAICS as appropriate. (Some weighting for the “new” NAICS sector, computer manufacturing, was performed to generate what might be expected to reasonably representative set of fuel prices. In general, the variance across 3-digit NAICS sectors in equipment manufacturing is not large).

B.2. Estimate total cost of offsite-produced fuels for non-MECS years

As a means of estimating the total expenditure on offsite-produced fuels on a consistent basis for non-MECS years, the ASM and EC cost of purchased fuels information was used for interpolation purposes and applied to each 3-digit NAICS sector. For each MECS year, the following ratio was computed:

Estimated cost of fuels from Step 1 (MECS)/Cost of purchased fuels (ASM, EC)

Between the successive ratios for MECS years (beginning in 1985), the ratios for intervening years were calculated by linear interpolation. The interpolated ratios were then applied to the ASM or EC cost data for the non-MECS years to provide an estimate of the MECS-defined costs for these years. Figure A.17 shows the cost ratios for the MECS years and the interpolated ratios for four different 3-digit NAICS sectors. The ratios for the years beyond 2010 will be retained at the 2010 values.

Download and process energy data from the Census Bureau for 3-digit NAICS sectors

The following discussion goes through some of the specific steps to obtain the cost of purchased fuels and the quantity of purchased electricity from the Annual Survey of Manufactures (ASM) or the Economic Census (EC). The Economic Census is conducted every 5 years, in years ending in “2” and “7.” For the 2020 update of the intensity indicators, the Economic Census for 2017 provided the most recent set of available data. Because the next effort to update the intensity indicators will involve the ASM (for 2018 through 2021), the explanation of the key steps will use the 2016 ASM as an example. It is most likely that the same set of steps would need to be followed for the 2018 and later surveys.

The ASM datafile contains a single worksheet with NAICS 3, 4, 5, and 6-digit sectors down through the rows. More than 90 variables in the various columns are included for each sector. For the 2016 ASM datafile, a worksheet EnergyData was added to pull out the needed variables. The variables are: 1) year.id, 2) NAICS.id, 3) CSTFU, 4) CSTELEC, 5) ELECPCH, 6) ELECSOLD, and 7) ELECSLD. These series are copied manually with the ctrl+shift_downarrow command to highlight, and then copy and paste to worksheet EnergyData.. The spreadsheet with the ASM 2016 data, after these revisions, is named “ASM_2016_31GS101_with_ann_pnnl.xlsx.”

The next step was to create a pivot table in a separate sheet (Excel does this automatically via request, typically naming the new worksheet “Sheet1”). The data for the pivot table is obtained from the EnergyData worksheet – highlight all data with “ctrl+shift_End” command. Then click on the “Insert” command along the top of Excel ribbon and select “PivotTable” as the left-most menu choice. It should be noted that the step of creating a separate worksheet, such as EnergyData, is not mandatory. The pivot table can be created directly from the full table (with all 90-plus variables). In this case all the data (included column labels) in that large worksheet is highlighted, before the PivotTable command is invoked. One then selects the four key variables from the “PivotTable Fields” portion of the instructions layout to create the pivot table (as explained below). The spreadsheet "ASM_2016_31GS101_with_ann_pnnl.xlsx." shows the results of both methods. The EnergyData worksheet was actually the result of pulling the appropriate variables from the 2015 ASM, with data for both 2014 and 2015. The pivot table produced from that previous work is name Pivot Table_2014-15. For the most recent data, the pivot table was produced directly from the full set of ASM variables, included as worksheet sheet "ASM_2016_31GS101_with_ann." The associated pivot table is in worksheet Pivot Table_2015-16. These worksheets have been renamed for transparency. The PivotTable process in Excel will simply create a new worksheet with the name "sheet1," or "sheet2" and the like in which to generate the actual pivot table. To repeat, these worksheets have been renamed after the data processing has been completed. The following paragraphs provide more detailed instruction on how to create the pivot tables.

In the worksheet (again to say typically, "sheet1") that is created by the instruction to create a pivot table (again, by moving to the "Insert" label on the top row of Excel instructions, and clicking on the leftmost item, "PivotTable" beneath). Check the boxes under the words "choose fields to add to report" for the following variables: year.id, NAICS.id, CSTFU, ELECPCH, and VACANCY. First drag the variable name year.id under the Columns quadrant of the pivot table layout. Next drag the NAICS.id under the Rows heading of the pivot. Finally, drag (one after another) the CSTFU and ELECPCH under the Values section. As for all variables, there will be downward-pointing arrow to the right in the box containing the variable name. Click on the arrow, and a menu will pop up. The bottom selection will be "Value field settings." Click on that item and choose "Sum" as the type of calculation. (For this pivot table, no variable is required under the Filters portion of the pivot table layout.) At this point there should be a box under the variable year.id in the Columns section that denotes a summing. (This should appear automatically as the pivot table is created; it is not a "variable" to be put manually into that position.)

The ASM tables provided by the Census Bureau typically contain data for two years. Typically, the Census data contain revisions for the previous year when the data for the most current year are released. Thus, the appropriate set of data to be used in the indicators is that for the prior year. In the case here, the 2016 ASM included both 2015 and 2016. The 2016 data was retained, because at this point, no revisions were yet available from the Census Bureau.

The pivot table will show the NAICS sector codes in consecutive order, with the lowest numbers for the codes listed first at the top of the pivot table. This convenient feature brings all of the 3-digit NAICS data together in one place. The user need only copy or link the values for the 3-digit NAICS sectors. In the case here for the processing of the ASM 2016, the pivot table results (in worksheet Pivot Table_2015-16) are subsequently to another worksheet labeled "Final Results – 3D NAICS."

The data for the most recent year, 2017, is found in the public use file for the Economic Census. This downloaded file, after inclusion of the data processing worksheets, is named "Manufacturing Census 2017_pnnl.xlsx." The general layout of this file is similar to the ASM, with NAICS sectors listed down the rows and various data elements listed across the columns. For the 2020 update, the process of creating a pivot table was not used. Rather, a table with 3-digit data only was created by simply choosing the appropriate rows at that level of aggregation. The final data are shown within columns BF and BK, rows 654 through 674. The key data items are highlighted in yellow.

The final step after the appropriate data from the ASM and EC have been generated is to transfer it to one of the supporting spreadsheets that are used to generate the intensity indicators for the manufacturing sector. In this case, the spreadsheet is ASMdata_010220.xlsx. The cost of purchased fuels and the quantity of purchased electricity for the most recent ASM years (are located in worksheet "Final Results – 3D NAICS" in cell range B31 through E51. The purchased electricity data is transferred via an active link to worksheet Electricity in ASM_data_010120, cells AE31:AF52. The purchased fuel data is linked to cells AE55 to AF75 in worksheet Purchased Fuels. The data from the 2017 Economic Census (in spreadsheet Manufacturing Census 2017_pnnl) are linked in the column to the immediate right (for 2017) in both worksheets, Electricity and Purchased Fuels in spreadsheet ASMdata.

B.3. Estimate cost of offsite-produced fuels for non-MECS years

As indicated in the previous section, ASM/EC data from the Census Bureau are transferred into two worksheets (Electricity and Purchased Fuels) in spreadsheet ASMdata. This section deals with the process of how the data from the Census Bureau is employed.

The purchased fuel values are linked from the worksheet Purchased Fuels to column J in worksheet “3D NAICS.” The most recent update covers the year 2015 through 2017. For example, for NAICS sector 311, the cost of purchased fuel is linked to cell J591. Columns A and B provide the year and sector code for the purchased fuel data.

The data are subsequently transferred to column F in worksheet “expenditure_ratios_revised.” Here is a case where the final overall methodology has not been streamlined, where seemingly redundant areas of calculations have not been removed. The relevant rows in this worksheet are from row 238 and larger, covering 1998 and subsequent years. The years in which a MECS was undertaken are highlighted in yellow.

Column E shows the estimates of the cost of fuels that are based on the MECS. The estimates are calculated as the product of the quantities of fuels purchased off-site and the prices paid for fuels by each 3-digit NAICS sector reported in the MECS. These estimates of expenditures for fuels will generally be smaller than the published values for expenditures in MECS Table 7.9. Table 7.9 includes expenditures for all fuels—including those used as materials—rather than those only for heat, power, and electricity generation.

The assumption underlying this element of the methodology is that the data from the MECS adequately represents the consumption and expenditures on fuels used in the manufacturing sector. Accordingly, the information from the Census Bureau for fuels is used as a means of estimating these values for years in which the MECS has not been undertaken by EIA. In the context here, the estimates of expenditures on fuels derived from the MECS is taken as the controlling set of data. The differences between these MECS estimates and the published cost of purchased fuels from the Census Bureau are taken to stem largely from the differences in the Census Bureau sample of establishments (ASM), or purported universe (EC) of establishments from that used in the MECS.

Under this framework, the ratios between the MECS and ASM/EC values are computed in each MECS year (MECS value divided by the ASM/EC value). Of course, in a MECS year, the MECS values can be reproduced when the ASM/EC values are, in turn, multiplied by these ratios. For other years, the ratios are assumed to be roughly in the same range as the most recent MECS. For the method here, the ratios for non-MECS years are generated by linear interpolation between the ratios calculated from the adjoining MECS years.

This process is straightforward in the worksheet Expenditures_ratios_revised. The ratios for MECS years are shown in column G (and highlighted in yellow for the 2010 and 2014 MECS). These ratios for the MECS years and the interpolated ratios are shown in column H. For the years beyond the 2014 MECS, the ratios have been left at their 2014 values. Column I shows the results of multiplying the ratios by the ASM/EC values for cost of purchased fuels in column F.

An estimate of the total quantity of fuels (consistent with the fuel used for heat, power, and electricity) is obtained by subsequently dividing these values by a composite price that represents the average cost per MMbtu of all (major) fuels. The following section of this documentation turns to the manner in which these composite prices are developed.

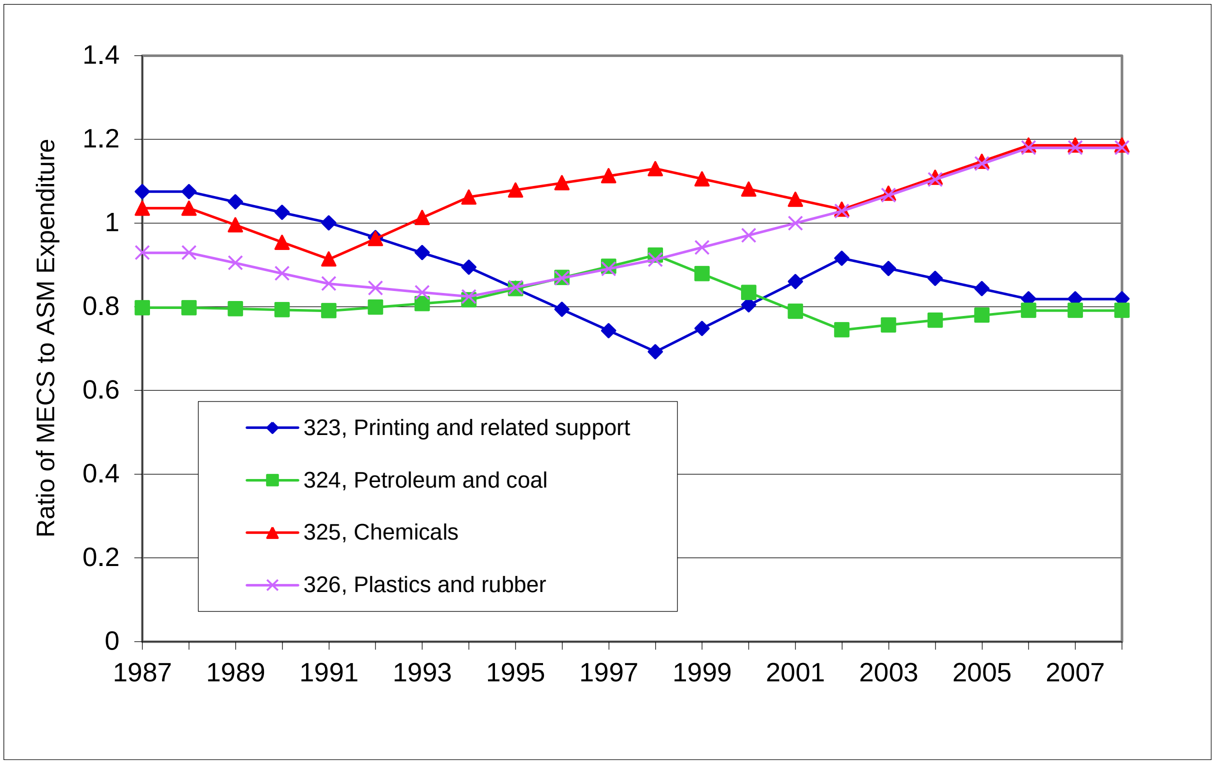

Figure A. 1. Ratios of MECS-based costs of purchased fuel to ASM/CM costs of purchased fuels for four selected sectors

B.3. Estimate annual costs per MMBtu for fuels for each NAICS sector

Estimates of cost per MMBtu by fuel for each sector were developed for all years. Individual fuel price models were estimated over the period 1988 through 2011 for each 3-digit NAICS sector and fuel type. For MECS beginning in 1998, the seven major fuels reported in the survey were used: 1) residual fuel oil, 2) distillate fuel oil (diesel), 3) natural gas, 4) LPG, 5) coal, 6) coke and breeze, and 7) other. For the earlier MECS, data for coke and other fuels were not published.

The purpose of these models was to try to ascertain any historical trends between the implied costs per unit from the periodic MECS and the various annual series of fuel prices published by EIA in the Annual Energy Review, Monthly Energy Review, (November 2019 issue for the 2020 update) or elsewhere available on the EIA website. These models must be judged as only very approximate in a process to impute how the published national fuel prices sector are correlated to the average fuel costs shown in the various MECS; the uncertainty in these models primarily stems from the presence of only seven (MECS) data points.

The source for the aggregate industrial price series include the last published Annual Energy Review 2010 (AER), a recent Monthly Energy Review (MER), or an alternative EIA table. These specific sources from EIA for fuel prices are:

- Natural gas - AER Table 6.8, MER Table 9.10; Natural gas prices by sector
- Distillate fuel oil – No.2 Distillate, Sales to end users, industrial, 2010 price extrapolated by prices for No. 2 fuel oil from Table 9.7 in the MER.
- Residual fuel oil – AER Table 5.23, Sales price to end users for residual fuel oil, 2010 price extrapolated by price for refiner price for sales to end users from Table 9.5 in the MER.
- Coal – Bituminous and anthracite prices from AER Table 7.8 through 2010. Prices for subsequent years from Table ES-4 in the Annual Coal Report 2018 (EIA).
- Liquified Petroleum Gas (LPG, Propane) – Table 9.7 from the MER.
- Wood/Waste – AER Table 3.4, Industrial biomass price.

The national-level fuel prices are collected in worksheet Prices, in spreadsheet ASMdata_date. (In the most recent update for 2020, the version of this spreadsheet was ASMdata_010220.xlsx.) The most recent prices for natural gas are inserted manually in column X from the November 2019 MER. The prices for distillate fuel oil are first entered in worksheet Distillate, and then subsequently linked to column L in worksheet Prices. The prices for residual fuel oil are first entered in worksheet Residual, and then subsequently linked to column M in worksheet Prices. The prices for LPG are shown in column AE in worksheet Prices.

The specification of the fuel price models was as follows.

$$\text{MECS\_Cost}(ym) = a\ \text{Price\_Ind}(ym) + b\ \text{Price-Ind}(ym-1) \qquad \text{(A.7)}$$

where MECS_Cost(ym) = published cost per MMBtu in MECS year ym (1985,1988,1991, 1994, 1998, 2002, 2006, 2010, and 2014)

Price_Ind(ym) = Aggregate industrial price for fuel in MECS year ym, as published by EIA in the Annual Energy Review, Monthly Energy Review or another publication on its website (see below).

The motivation for including both the current and lagged price was that the actual cost (price) paid by industrial customers could involve a long-term supply contract or be the result of fuel inventories held at the manufacturer's site.

Based upon the estimated parameters, a and b from Equation (A.6), prices were predicted for each year between 1983 and 2011. The regression residuals at the eight MECS years (actual – predicted) were computed. Between the MECS years, values were computed that were linear interpolations of these residuals between successive pairs of MECS years. These "interpolated" residuals for the non-MECS year, as well as the residuals for the MECS years were added back to raw predictions from the regression. The procedure ensures that the actual prices for all MECS years are the actual values corresponding to each MECS.

An illustration of the methodology is provided by Figure A.18 below. In this case, the regression model was applied to natural gas consumed by the chemicals sector. The top line in the figure shows the price for all industrial sales of natural gas. The lower graph shows the predicted, or perhaps more accurately, interpolated values of the natural gas price, consistent with the observed values in the MECS. The (red) squares in the background denote the MECS years and show how the regression errors have been adjusted to yield the MECS. Not surprisingly, the average price paid for natural gas is lower for the chemical sectors than it is for the entire industrial sector. The regression coefficients in this case were a = 0.0.85 and b = 0.0.23, suggesting only small support for the lag in the overall market price for natural gas to be fully reflected in prices paid by the chemicals sector. In the case of NAICS 322, Paper, (not graphed here), the impact from the lagged price is more evident, with the coefficients on the current and lagged as 0.81 and 0.19, respectively.

The annual estimates of the fuel prices for the 18 3-digit NAICS sectors developed in seven separate spreadsheets, one spreadsheet for each fuel type. The spreadsheets are Ind_Gas_Price_date, Ind_Coal_Price_date, Ind_Distillate_Price_date, Ind_Resideual_Price_date, Ind_LPG_Price_date, Ind_Coke_Price_date, and Ind_Other_Price_date, Ind_Gas_Price_date. The spreadsheets are identical with each other in the way the annual prices are estimated.

The spreadsheet for natural gas will be used to illustrate the detailed estimation procedure. The first worksheet in Ind_Gas_Prices.xlsx is labeled Gas_Prices. This worksheet is used as the transfer point of the set of historical natural gas prices derived from the various MECS conducted since 1985 (through 2014). These prices are linked to spreadsheet MECS_EnergyPrices, a spreadsheet that collects the historical set of MECS energy prices for each fuel. In worksheet Gas_Prices, the MECS prices are in the range D33 through L53. These prices are moved to a similar size array to the right (columns N through V), using a Copy Values instruction. The user must then transpose those values in the range Z7 through AT15. Finally, the prices are transferred into a table immediately below where the years run consecutively from 1983 through 2018 (Z19:AT50). The prices are located in the rows for the specific MECS years; the rows (years) without a MECS are left blank.

The next step is to update the fuel price models, as described above, for all of the 21 manufacturing sectors. The model for each sector is in a single worksheet, labeled with the 3-digit NAICS code. For the natural gas spreadsheet, the worksheets are labeled “Gas-311,” “Gas-312,” “Gas-313,” and so on.

The organization of each of these worksheets is identical, but with one exception. In the very first worksheet, for NAICS sector 311, the EIA aggregate annual fuel prices for the period 1983-2018 from the spreadsheet ASM_data, are transferred into column D. The worksheets for the other NAICS sectors then link to the sector “311” worksheet to fill in the same set of prices.

The other columns in these price model worksheets are reasonably straightforward. In column E are the MECS prices for that particular sector, linked back to the first worksheet Gas_Prices. Column F shows the predicted prices based on the coefficient of the price model (to be explained below). Columns G and H contain, respectively, the regression residuals (i.e., actual value minus predicted value), and their squared values. Column I repeats the regression residuals for the MECS years and interpolated (linearly) residuals for years between the MECS years. Column J then adds the actual and interpolated residuals back to the aggregate prices (independent variable) for all years. Of course, for the MECS years, this calculation will yield the published MECS price (dependent variable.) This column is labeled “Calibrated Prediction.”

The columns to the right (N, O, and P) reproduce selected column to facilitate the adjoining Excel plot. As constructed, the “calibrated predicted” and the actual MECS price are the same in MECS years. In the other years, there is deviance between the movements between the calibrated predicted price and the aggregate prices. These deviances of course will depend upon the different market conditions overall for the particular fuel, and other factors influencing the prices that might have been actually paid by firms in that particular sector.

The estimation of coefficients a and b in Equation A.7 is performed by the use of Excel’s Solver (add-in) function. After this add-in function has been installed in Excel, it can be invoked by first selecting the “Data” menu at the top of Excel window. “Solver” will be listed at the far right of the menu choices, under the “Comments” link. Next set the cursor on cell H3 and then move it over the “Solver” label and click to enable. Go to the bottom of the Solver pop-up area and click on the “Solve” button.

The current model in the worksheet is very simple. The estimated coefficients a and b are in cells H3 and H4. The objective function is to minimize the sum of the squared residuals, defined in the Excel formula in cell H4. When Solver is invoked, the optimization routines will work to find those values of the coefficients that will minimize the objective function.

Given the current set-up of the regression models to estimate the fuel prices for non-MEC years, the coefficients need not be re-estimated until data from the next MECS is available. However, the user may need to add additional rows at the bottom of the relevant columns in the sector-specific worksheets to incorporate more recent years. For the 2020 update, the models make predictions through 2018, given the available aggregate fuel price information through that year. When price data for the 2018 MECS become available, the models should be re-estimated with this additional observation.

One other comment is pertinent to the use of these models in the 2020 update to predict the most recent prices for the individual manufacturing sectors. In general, the residuals for the years 2015 and later are set equal to the residual for 2014 (the last MECS year). In some cases, this assumption appeared to yield implausible estimates of the predicted prices. In those cases, different patterns of the residuals were set, based on analyst judgement. When the default assumption for the setting the residual for the years 2015-2018 is overridden, those cells in column I are highlighted.

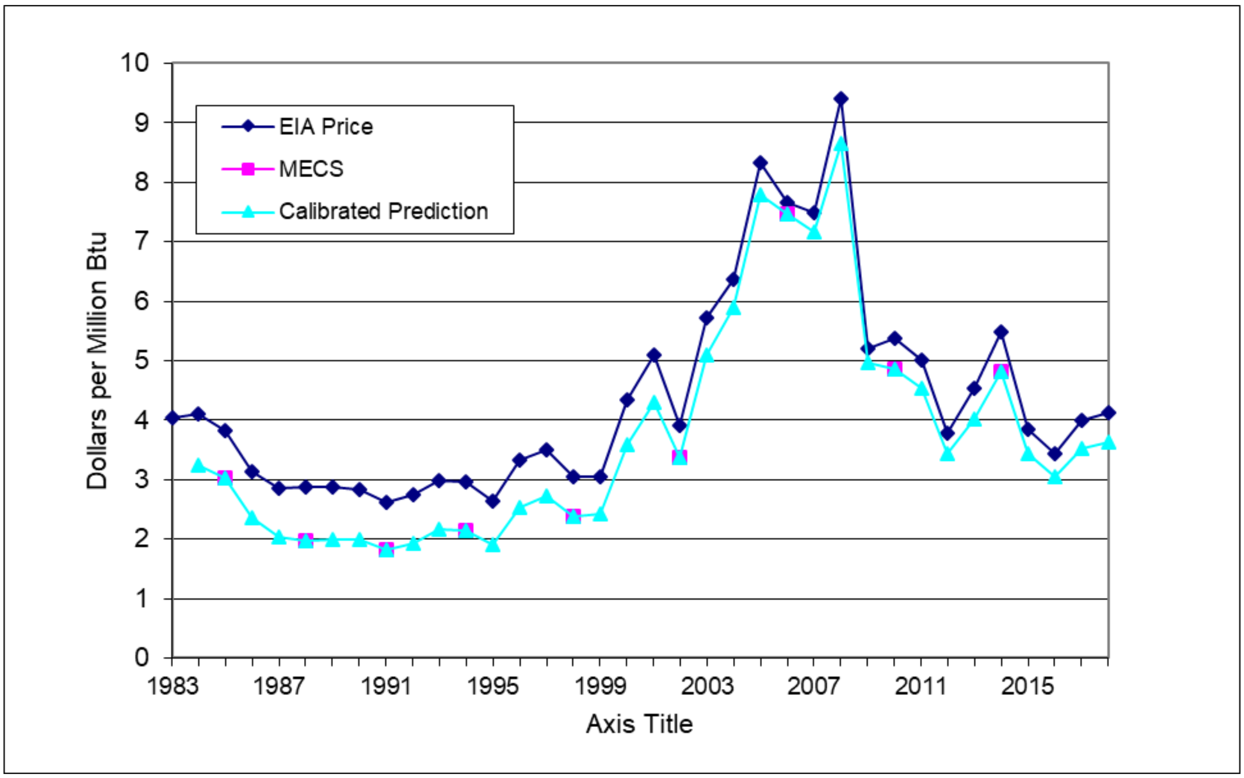

Figure A. 2. Predicted prices of natural gas consumed by the chemical sector, NAICS 325

Estimate composite fuel prices

Step 4 involves weighting the predicted fuel prices (actual prices in MECS years) to derive an overall price per MMBtu of fuel consumption. Weights were developed from fuel quantity shares derived from the MECS data on offsite-produced fuel consumption, and then linearly interpolated between MECS years. Until data from the next (2018) MECS is released, the fuel shares are expected to be maintained at the observed levels for 2014.

Composite fuel prices (i.e., $/MMBtu) were computed as the cross products of the weights (shares) in each year with the predicted MECS prices. For the period 1985 through 1998, the composite prices were based upon the five fuels for which sector-level prices were available in the MECS. For 1998 through 2018, the composite prices were developed from prices for each of the seven fuels. For 1998, composite prices were based on both five and seven fuels, as a way of bridging between the years just before and just after 1998.

The composite fuel prices are calculated in spreadsheet MECS_prices, worksheet QuantityShares_1998 forward. The quantity shares for the seven fuels are shown in columns Q through W. The highlighted rows are those in which MECS fuel consumption and prices are available. The quantities from the 2014 MECS are located in rows 345 to 365, these values come from the worksheet MECS_data (as described above). For the intervening years, the quantity shares are interpolated between the appropriate MECS years. The shares are held constant for the years after 2014.

The composite prices (dollars/MMBtu) are computed as simply a weighted average of the prices for the (seven) individual fuels. The fuel prices are transferred from the various spreadsheets that were used to develop annual estimates of prices by fuel and 3-digit NAICS sector (e.g., for natural gas from spreadsheet Ind_Gas_Price_date.xlsx). The composite prices are shown in the yellow-highlighted column AH for years 1998 through 2018.

These prices are subsequently transferred to spreadsheet ASMdata_date, worksheet "Quantity_shares_revised". They begin in cell CS277 (NAICS 311 for 1998) and extend further down the column shaded in light purple. The prices for earlier years, 1985 through 1998, based on four fuels, are in column CW.

B.4 Estimate annual offsite-produced fuel quantity consumption, consistent with the MECS.

With estimates of prices and total cost in hand, this step is straightforward. The estimates of the total cost of purchased (offsite-produced) fuels are divided by the corresponding composite price of energy for each NAICS sector. By construction, the total cost values and the composite prices match those that can be calculated directly from the MECS in the years the MECS has been published.

These calculations are performed in spreadsheet ASMdata_*date*, worksheet "Final_quantities_w_ASM_85). As noted above, column G shows the prices that have been linked from worksheet Quantity_shares_revised. Through row 301 (1998), these prices are based on data using only five fuels. The calculations are duplicated for the single year 1998, where the prices have been based on the separate estimates, one using four fuels, the second using seven fuels. The overlap occurs in rows 281 through 322. Rows 281 through 301, to repeat, use data from four fuels for 1998. Rows 302 through 322 are based on the composite prices using seven fuels.

As mentioned earlier, an estimate of the total quantity of fuels (consistent with the fuels used for heat, power, and electricity) is obtained by subsequently dividing estimated cost of fuels (in column F) by the composite price that represents the average cost per MMbtu of all (major) fuels (in column G). The calculated consumption values from this division are shown in column J. For MECS years, the published values from the MECS are shown in the highlighted rows in column K. If the prior calculations to estimate the composite fuel price and total expenditures of fuels are correct, then the values in column J and K should match.

B.5. Estimate on-site produced and total fuel consumption

The efforts in all the previous steps were employed to generate an estimate of fuel consumption that is consistent with ASM and EC data on "cost of purchased fuels." In terms of the MECS, the quantity of energy consistent with this definition, as mentioned above, is labeled offsite-produced energy consumption (from Table 4.2). The MECS data on *total* fuel consumption include both offsite-produced and onsite-produced energy. Sectors that have a large fraction of total fuel consumption from onsite sources include wood products, paper, petroleum refining, chemicals, and primary metals.

Typically, the onsite-generated fuel consumption shows up in the "other" column of the MECS tables for total fuel consumption (Table 3.1 or 3.2 in the most recent MECS). However, some of this energy also is purchased (or transferred) from offsite sources. For example, in 2010, the MECS reported "other" fuel use in Table 4.2 from offsite sources, for the Food industry (NAICS 311) to be 54 trillion Btu (TBtu). Total fuel consumption for the "other" category in Table 3.2, on the other hand, was 99 TBtu. The ratios of fuel consumption between offsite and onsite may be volatile from one MECS to the next, owing to the particular structure of the industry and underlying technologies. For purposes of the intensity indicators methodology, it was assumed that the ratios of *total* energy consumption to *total* offsite-produced energy consumption would be the most stable over time. Thus, for each MECS year, this ratio was calculated. For example, in the Food industry, the ratio in 2010 was calculated as 902/855 (TBtu) = 1.055. For non-MECS years, the ratios were interpolated linearly between successive MECS years. Again, taking Food as an example, the corresponding ratio was 1.071 in 2006. Thus, the ratio of total/offsite fuel consumption was gradually reduced from the 2006 value in each of the intervening non-MECS years, 2007 through 2009. These ratios were then applied to effectively "inflate" the offsite-produced estimates of fuel consumption as explained in the previous section.

The data and calculations to produce these ratios are included in worksheet 3DNAICS. Column AT shows the values for offsite-produced fuel consumption. Column AU shows the values for total fuel consumption, both offsite and on-site produced (MECS Tables 3.1 and 3.2). The values in column AU have been linked to column BT, where in prior work the values for individual fuels were used in the overall method. Those quantities for specific fuels are no longer employed. The values from the published totals for fuel consumption from MECS are shown in column AZ for the years 1998 and 2002; these are shown only as a data check on the values in column AU. While the worksheet contains other types of informal checks for later years, the basic method remains the same in those years; the values in column AT represent only offsite-produced fuel consumption and the values in column AU represent total fuel consumption.

The values in column AV show the ratios between total consumption and consumption from offsite sources. As implied above, the ratios reflect a large amount of onsite-produced fuel consumption for a few major sectors, including wood products, paper, petroleum refining, chemicals, and primary metals.

The application of these ratios occurs next in the worksheet Final_quantities_w_ASM_85. (Admittedly, an effort that started with a blank spreadsheet would have eliminated this unnecessary use of different worksheets; it results now from successive decisions to adjust, at least on its conceptual basis, the method to be more transparent and consistent with the available data sources.)

The ratios calculated in worksheet 3DNAICS are transferred to column L in Final_quantities_ASM_85. For example, the value for NAICS 311 (1.07) for 1998 is in cell L238. The ratios for 1998 and later MECS years are highlighted in yellow as one scrolls down column L. For the non-MECS years, the method is similar to that used to estimate the total expenditures on fuels used for heat, power, and electricity generation. The ratios between the two consumption categories (offsite-produced vs. total) are linearly interpolated for years between MECS years and held constant after the last MECS year (2014). These ratios are shown in the non-highlighted entries in column L. They are then applied to the onsite estimates (based in part on the ASM/EC data on cost of purchased fuels for non-MECS years) shown in column J. The resulting estimates (as a reminder, in trillion Btu) are shown in column M (and labeled as "Final Quantities" in rows 6 and 7 at the top of this column).

One additional reorganization of these estimates is made. The estimates in column M are transferred to an array defined by cells (X50 through BD70). In a second (green shaded) array, just below, the aggregation of the NAICS sectors (the first six sectors to three sectors) is made.

C. Develop estimates of annual electricity consumption in manufacturing

The electricity intensity indicators for manufacturing are based solely on the data provided by the Census Bureau in the ASM and EC. Because these census sources represent a larger sample (for both ASM and EC) of manufacturing establishments than the MECS, the electricity estimates (quantity in kWh purchased) are used directly in the system of energy intensity indicators. These census sources thus provide data for all years--from 1985 through 2017--for the 18 manufacturing sectors included in the system. The Census Bureau has collected information on the quantity of purchased electricity for many years. This data is judged to provide a more consistent reporting accuracy for the complete time series as compared to using the data from the periodic MECS and applying an interpolation procedure for non-MECS years.

The ASM and EC data also include a data item labeled "Generated less sold". This information is not used for the energy intensity, because it involves a "double counting" of total energy. Electricity generated by manufacturing is using purchased fuels and other fuels that are already included in total energy consumption.

This section provides a guide to how the ASM/EC was processed for use in the indicators. To begin, the basic data is downloaded from the either the ASM or EC from a Census Bureau website. As the electricity data were collected from the same sources as the cost of purchased fuels, the first portion of this process was described earlier in Section B.2. The resulting downloaded values at a 3-digit NAICS basis, are linked to the spreadsheet ASM_data, worksheet Electricity (tab highlighted in green as the third worksheet in the spreadsheet). Rows 30 through 52 show the primary data by NAICS sector. Column AB through AG show the data for 2012 and later years, each column representing a different year. The cells in these columns are linked to the specific spreadsheets that were used to process the downloaded data from either the ASM or the EC.

The electricity consumption data make their way through the remainder of the spreadsheet in a rather circuitous way, again a vestige of earlier and varying methods of processing this information. The purchased quantity of electricity (in thousand kWh) is first transferred (via links) to column E in worksheet 3DNAICS. The values are duplicated in column J. And finally, the values are converted to trillion Btu in column BV of the same worksheet.

The values of purchased electricity in worksheet 3DNAICS are subsequently transferred to column BV (the same column, by coincidence) in worksheet "Final_quant_elec_w_ASM_87." The data begin in 1987. Cell BV6 shows the value for electricity for NAICS sector 311 (after some previous adjustment to convert the SIC-based data from the ASM to the NAICS – not relevant to the process of future updating of the indicators.)

The values in the single column BV are subsequently transferred into a rectangular layout, with each column designating a different year. This array is shown in cell range CB6:DF26. (Note: the values for 1987-1997 were not taken from this array. Special analysis was conducted in the rows below to make a bridge between the NAICS and the earlier SIC-based data.) In the three rows immediately below (26-28) the aggregation for the first six NAICS sectors (311-316) is undertaken. After this aggregation, the resulting 18 sectors are shown in array CB95:DB112. As the heading to this table denotes (cell BZ92), these are the values transferred to spreadsheet ind_hap3.xlsx. The following section describes the data are handled in the ind_hap3 spreadsheet.

D. Perform final interpolations and adjustments

Some further adjustments to the estimates for fuel and electricity consumption were made prior to their use in generating energy intensity indexes in manufacturing. These adjustments were all made in spreadsheet ind_hap3. (For the 2020 update, the actual file name is ind_hap3.122219). The discussion begins how these adjustments were made to the fuel consumption estimates.

Generally, the derived total fuel consumption estimates from the calculations in section B.5 match the values from the MECS directly in the MECS years. In the years beginning with the 1998 MECS, this is particularly the case. Some small differences may result from slight adjustments made to one or more fuel consumption or expenditure data values in the course of estimating the composite fuel price. In the years prior to the 1998 MECS, there is the issue of data comparability from the change from the SIC to the NAICS. As mentioned at the outset of the discussion on manufacturing, an effort to try to rigorously adjust consumption and expenditures for each fuel type from its value under the SIC to its value under the NAICS was not performed. Thus, the interpolating series were based upon a rough cross-walk at the 2-digit SIC and 3-digit NAICS sector levels. The actual (preferred) estimates of NAICS-based total fuel consumption (i.e., including onsite as well as offsite-produced fuel) for the MECS for 1985, 1988, 1991, and 1994 were developed with a more detailed mapping procedure based upon EIA's published data from the 1998 MECS on both an SIC and NAICS sector basis.

One additional interpolation procedure is conducted to generate the final estimates for the non-MECS years. This final interpolation procedure involves imputing the deviations from a hypothetical linear interpolation between the total fuel quantities for successive MECS years.

These adjustments are made in worksheet MECS_Annual_Fuel1. The annual values from ASMdata are transferred to the top of this worksheet, in an array defined by cell range L6 through AR23. The rows below this array, 28 through 45, display the interpolation procedure, defined by what are termed "interpolation ratios." Values in this array are set equal to 1 in MECS years. For the non-MECS years, a value is calculated by assuming linear interpolation between the successive MECS years in the top array. The interpolation ratio is defined as the actual value of the previous non-MECS year estimate (in the top array) divided by the corresponding value in the linearly interpolated array. Thus, had the actual values of the non-MECS fallen along a linear path between the MECS years, the interpolation values would be 1.0.

The final estimates are shown in the third array, in the rows 51 through 68. In this array, the interpolation ratios are applied to a value determined by linear interpolation of the values from the relevant MECS years. Note that consumption values taken directly from the MECS (worksheet MECS_Fuel) are transferred into the columns for the MECS years (columns G, J, M, and so forth). If the data from the MECS have been accurately processed through the prior steps (in spreadsheet ASMdata), this procedure should essentially yield the same value for the non-MECS year as in the initial array. Thus, one may ask the value of what might appear to be a duplicative procedure. The procedure was added to cover those instances when slightly different values from the MECS appear in the initial array, resulting from possible adjustments made at earlier steps in the overall estimation framework.

The values in this array are then linked to the next worksheet, "MECS_Annual_Fuel2." They are found in range CB59:DH76. From this worksheet, the data are finally transferred to worksheet "ASM_Annual3_1970on." (Some additional discussion of MECS_Annual_Fuel2 is provided below in the section on electricity). The values are in array Z8:BF25. It is in this worksheet that the values from 1985 and later years are linked to estimates from 1970 through 1985. (These are years prior to the establishment of the MECS. The source for this data is generally the National Energy Accounts, as described in the more general documentation for the intensity indicators, Rev 3 of the comprehensive report.) The 1970-2017 estimates are located in the array defined by cells G30 through BB47.

One final transformation of the fuel consumption is performed in worksheet "ASM_Annual_Fuel3_1970on." The data from array G30:BB47 are transposed to array BQ10:CH57. The transposed values are highlighted in yellow. The values in this array are transferred to the spreadsheet that actually generates the manufacturing energy intensity indexes, industrial_indicators_*date*.xlsx. The worksheet Manufacturing_Energy_Data in the industrial indicators spreadsheet is the location to which the estimates of fuel consumption (from spreadsheet ind_hap3) are linked. The array with the fuel consumption estimates comprises cell range C66:U133 in worksheet Manufacturing_Energy_Data.

Electricity

As described in the previous section, the estimates of electricity consumption are primarily developed in the spreadsheet ASM_data. As described toward the end of the previous section, the final 1987-2017 estimates developed in this process are shown in worksheet "Final_quant_elec_w_ASM_87."

The array of electricity estimates is subsequently transferred (linked) to worksheet ASM_Annual_Elec_1970on in spreadsheet ind_hap3. The location of these transferred values is cell range AB8:BF25.

The worksheet ASM_Annual_Elec_1970on combines these more recent values of electricity consumption with values for earlier years, primarily derived from the National Energy Accounts. The estimated electricity consumption for the years 1970 through 1987 is in the array G8 through X25. In the range Z8:Z25 are found factors that are designed to scale the previous 1970-1987 set of estimates to be roughly consistent with the more recent estimates. The scaling calculation shows in the cell formulas in range G30:Q47. The transferred-in estimates are copied to the columns to the right of the scaled values for 1987 in column Q. The resulting 1970-2017 electricity array is thus defined over cell range G30 through BB47. In the area of worksheet below this array, plots of the resulting time series of consumption estimates are generated.

The values from the array G30:BB47 are transposed, as shown in the columns to the right. The cells with the transposed values, in arrayBR10:CI57, are highlighted in yellow. These values are transferred to worksheet Manufacturing_Energy_Data in spreadsheet industrial_indicators.*date*.xlsx to generate the final (electricity) intensity indicators for the 18 manufacturing sectors.

Choice of electricity consumption estimates for interpolation of non-Census year fuel estimates

The electricity estimates in worksheet ASM_Annual_elec_w_ASM_87 are also used in worksheet "MECS_Annual_Fuel2" in spreadsheet ind_hap3. The values from 1988 and later years are linked to cell range L6:AO23. In this worksheet, the electricity estimates are used to interpolate estimates of fuel consumption for non-MECS years. These interpolations are performed in rows 28 through 68. The interpolation method was described previously with regard to the use of the ASM-derived estimates of fuel consumption in non-MECS years. More specifically, this process was explained with respect to such interpolation in worksheet MECS_Annual_Fuel1. The procedure is identical in worksheet MECS_Annual_Fuel2, except that electricity values for non-MECS years are used to interpolated between MECS years. The resulting values of fuel consumption are shown in array CB26:DH53. Below that array are the results of using deflated cost of purchased fuels as the interpolating series between MECS years. These values were computed in worksheet MEC_Annual_Fuel1. Those values are transferred to MECS_Annual_Fuel2, in cell range CB59:DH76. Thus, two arrays of annual estimates for fuel consumption are stacked on top of each other, the top array derived from the use of electricity as an interpolator (between MECS years), and the lower one the result of using deflated cost of purchased as in interpolator.

The array in cell range CB80:DH97 provides a means of selecting between either of the upper two arrays to produce the final time series of fuel consumption for each sector. In the recent work, electricity was selected as the better interpolation series for Petroleum and Coal Products (NAICS 324), and Chemicals (NAICS 325). As shown in the blue highlighted rows, the series using electricity as an interpolating series was selected for the years 1987 through 2009. In the later years, the deflated cost of purchased fuels was used. This choice was influenced by comparison between these sources to how well aggregate manufacturing electricity consumption matched up to the EIA published value for total fuel consumption in the industrial sector. (See columns BJ through BN)

The following paragraph essential duplicates the discussion above with regard to fuels. The set of estimates after this selection of the interpolation method (array CB80:DH97 for 1985 through 2017) was transferred to worksheet ASM_Annual_Fuel3_1970on. As discussed above, these values are located in cell range Z7:BF25. After linkage to the estimates for earlier, and transposing the resulting array, the final estimates are in the yellow-highlighted array BQ10:CH57. These values are subsequently linked to worksheet Manufacturing_Energy_Data in spreadsheet industrial_indicators_*date*.xlsx to provide the fuels consumption measures needed to construct the energy intensity indicators in manufacturing.

E. Measures of Industrial Output

As described in the body of the report, the primary energy intensity indicators for the industrial sector are defined in terms of energy consumption per unit of industry gross output (GO). Linkages to more aggregate sector intensity are obtained by the use of value added (VA), where value added is the particular sector's contribution to gross domestic product (as defined below).

Spreadsheet files provided the Bureau of Economic Analysis (BEA) provide the required data to construct the time series of gross output and value added. These files can be obtained from the BEA website. https://apps.bea.gov/iTable/index_industry_gdpIndy.cfm. BEA has produced separate Excel tables for the 1947-1997 data and for 1997 and later years. They are accessed by clicking on the text links "Historical 1947-1997 data" and "Access Underlying Detail tables." The zip file under both of these categories should be downloaded. (It is also possible to select user customized tables with BEA's interactive data selection tool, but that method is a bit more cumbersome than simply downloading the zip files with data for all years.) Within each zip file, four possible Excel datasets are available: GrossOutput, IntermediateInputs,KLEMS, and Value-Added. For the intensity indicators work, the GrossOutput and Value-Added spreadsheets are the two files with the required data.

The following text is taken from the "ReadMe" tab of this spreadsheet (see below) used to disseminate the gross output and value added estimates from BEA. Formally, BEA defines value added and gross output as:

> Value added (VA) is the contribution of each private industry and of government to the gross domestic product (GDP) of the United States. VA is equal to an industry's gross output (sales or receipts and other operating income, commodity taxes, and inventory change) minus its intermediate inputs (consumption of goods and services purchased from other industries or imported). Current-dollar value added is calculated as the sum of distributions by an industry to its labor and capital, which are derived from the components of gross domestic income. [Gross Output capitalized here to improve clarity].

All of the value added and gross output data provided by BEA are in terms of current dollar and chained quantity indexes. For the most recent 2020 update of the intensity indictors, the chained quantity indexes are indexed to be 100.0 in 2012. For purposes of the indicators calculations, the real values of both value added and gross output are obtained by multiplying the 2012 estimates by the corresponding time series of quantity index values (and dividing all subsequent values by 100 to account for value of the index in 2012).

The following discussion provides some more detailed documentation on how the value added and gross output were treated in the relevant spreadsheets. The discussion will focus on the gross output data. The procedures undertaken for the value-added data are very similar.

The final spreadsheet used to compile the full time series of gross output data is labeled GrossOutput_1969-2018_PNNL_123119.xlsx. The key worksheet (as labeled by BEA) is ChainQtyIndexes. The data at the top of the worksheet covers the period 1997 through 2018 (array B7:X107). Cell C114 is the upper left corner of an array that was copied in from the BEA spreadsheet containing the 1947-1997 data. In this case, that data was obtained through BEA's query system where a customized spreadsheet with data from 1969 through 1997 was downloaded. As a single sheet it was copied into the worksheet labeled "Sheet0_Downloaded fr. BEA. As mentioned above, that same data could be obtained from a (downloaded) spreadsheet containing the full 1947-1997 data set.

The 1969-1997 and 1998-2018 data are put together in the same array covering rows 115 through 149 in worksheet ChainQtyIndexes. The number of rows is smaller in this array because it can ignore the sectors past those in manufacturing. The different sources of the data are shown by comparison of the formulas in columns AE and AF.

In column BB of ChainQtyIndexes, the actual values in dollars are shown for gross output. These were taken from worksheet GO. This worksheet again is found in the BEA spreadsheet GrossOutput with the most recent data from 1997 forward. The indexes of gross output are then converted to 2012 dollars by multiplying all the index values (C115:AZ149) by the constant 2012 value in column BB. The resulting set of chained 2012-dollar measures of gross output are found in array BD115 through DA149.

Ideally, the estimates by NAICS sectors provided by BEA would be sufficient to simply utilize in the system of industrial energy intensity indicators. Unfortunately, a number of aggregate measures of gross output (and value added) must be constructed from the BEA data. The particular aggregate measures are shown in the array, whose leftmost column is BA153:BS225.

Two aggregate measures are required: total nonmanufacturing (comprised of indexes for agriculture, mining aggregates), and industrial (manufacturing and nonmanufacturing). BEA does provide an aggregate index (later converted to 2012 dollars here) for manufacturing. Thus, the industrial index required for the indicators system is constructed as an aggregation of the *published* manufacturing measure output and a *constructed* measure for nonmanufacturing, both defined in terms of chained 2012-dollar gross output.

One additional aggregation must be performed for a single sub-sector in manufacturing. BEA produces separate estimates for "Motor vehicles, bodies, trailers, and parts" and "Other transportation equipment." These two sectors must be aggregated to form a single measure for transportation, NAICS sector 336.

Following the methodology used by BEA, the aggregate indexes are constructed by the use of the Fisher index formulation. A Fisher index is constructed at the geometric average of the Laspeyres and Paasche indexes from the underlying data. The geometric average is computed straightforwardly as the square root of the product of the two indexes.

The key data needed for the construction of both Laspeyres and Paasche indexes are gross output measures in current dollars and the corresponding quantity indexes. For agriculture, mining, and construction, these values are shown in rows 155-157, and 165-167. The construction of the Laspeyres and Paasche indexes are shown in rows below, 164 to 173. For any update, the user may simply extend the formulas to the more recent years in additional columns. (Thus, no attempt to explain the Laspeyres and Paasche indexes is undertaken here.) The final indexes for these two methods are in rows 168 and 174. The values in these rows are actually "index relatives," that is, the year-over-year change in the index. The index relatives for the Fisher index for nonmanufacturing are computed in row 176, as the square root of the Laspeyres index relative (row 168) and the Paasche index relative (row 174). The chained index itself (row 177) is calculated by multiplying the current year index relative by the cumulated value of the index for all previous years. Finally, that index can be normalized to any designated base year, in this case to 2012. The final index for nonmanufacturing is shown in row 178.

The construction of the two other required indexes are shown the rows below the development of the nonmanufacturing index. The calculational procedures are identical; again, values for the current dollar measure of gross output and the corresponding quantity index are the relative inputs. The final Fisher quantity index for the industrial sector is in row 202. The index for all transportation equipment (NAICS 336) is shown in row 225.

The full set of gross output data (chained 2012 dollars) for the required sectors is shown in the array BE228 through DA260. For most of the sectors, the values are simply copied from the BEA values in the rows above. Note the formulas to compute the quantity measures for the three special sectors considered above: total industrial (row 229), nonmanufacturing (row 203), and transportation equipment (row 239). The bottom of this array shows the required gross output estimates needed to compute the energy intensity indexes in nonmanufacturing (rows 251 through 260.)

For use in the industrial_indicators spreadsheet, the gross output (and value added) data must be transposed to put the years in row order, not columns. The required step to do the transportation is to generate an identical "values only" array. This array is shown in rows.264 through 296. The transposed data are shown in array DD301 through EJ349. This array is linked to worksheet Gross Output in spreadsheet industrial_indicators_*date*.xlsx.

As mentioned above, the overall process to compute the required measures of value added is identical to that used for gross output. The relevant data file, downloaded and subsequently modified, for value added in the 2020 update is ValueAdded_1969-2018_PNNL_010120.

Nonmanufacturing

Nonmanufacturing covers agriculture, mining, and construction. Prior to 2012, total nonmanufacturing energy consumption (electricity and fuels) was estimated as a residual between the supply-side estimates of industrial consumption published by EIA and the end-user estimates for manufacturing based upon the MECS (supplemented by census-based data, as described above). The residual-based method produced very unsatisfactory results; year-to-year changes in energy consumption were implausible in a large number of instances. A complicating factor for fuels is that industrial consumption estimates published by EIA include energy products used as chemical feedstocks and other nonfuel purposes. As a result, a preliminary effort was undertaken in mid-2012 to estimate energy consumption from the user side for these sectors. The following paragraphs briefly describe the data sources and methods used.

A. Agriculture

The principal data source for agriculture built on a set of estimates developed by John Miranowski, a professor at Iowa State University. Miranowski developed annual estimates of energy by fuel for the farm sector for the period 1965-2002.[1] These data were later employed by Randy Schnepf at the Congressional Research Service, who was kind enough to send the spreadsheet-based Miranowski consumption estimates to PNNL.[2]

Data beyond 2002 were developed from the survey of farm expenses conducted by the National Agricultural Statistics Service (NASS). Information on the expenses for fuels were derived from downloaded data from: http://www.nass.usda.gov/Statistics_by_Subject/index.php. This same website was also accessed for data for average prices paid. Estimates for fuel expenses were available for four fuel types: diesel, gasoline, LPG, and "other" (assumed to primarily consist of natural gas). The expenses data were downloaded for years 2000 through 2018. Data for average prices paid were available for diesel, gasoline, and LPG up through 2008. Estimates for later years were based upon extrapolating the series with fuel price data from the *Monthly Energy Review*. Some adjustment of the retail prices was made to better reflect prices more likely to be paid by farmers (e.g., remove highway use taxes for diesel and gasoline. Estimates of annual fuel consumption were derived by straightforward division of the expenditure data by the price estimates.

The "other" fuel category was assumed to consist primarily of natural gas. Expenditure data were converted to fuel quantities with the use of the natural gas prices for the commercial sector published by EIA.[3]

Diesel fuel represents the majority (~ 65%) of fuel used by farms in the U.S. The consumption estimates by fuel type for 2010 were: diesel - 449 TBtu, gasoline - 114 TBtu, LPG - 71 TBtu, and natural gas/other - 61 TBtu.

The following section provides some more detailed discussion of the sources of the primary energy data and subsequent processing for use in the intensity indicators. From the nass.usda.gov website cited above, one uses the interactive data selection feature to choose the fuel expenses data. The following steps should be followed: Under the "economics" in left-most box "Sector". Choose "expenses" in middle box "Group". Finally choose "fuels" in in right-most box, "Commodities." When menu "Choose data items" pops up, then select all entries where the entry ends with "measured in $." Then, press continue at the bottom of the screen. The data for the latest year will show on the screen, but then hit the "more" key. A full set of data will then fill the screen. Hit the menu item "spreadsheet" on the top right of the screen. This automatically downloads a spreadsheet with data for the latest year and a number of previous years. There are very few numbers to be selected for each year. They are perhaps most efficiently keyed manually into the relevant spreadsheet with all the historical data. The data should be entered or added to the existing array defined by C9 through H27 in worksheet "Farms" in spreadsheet "Agricultural_energy_*date*.xlsx (in 2020 update Agricultural_energy_010420.xlsx)

In columns L through V in worksheet Farms are the calculations to make an approximate estimate of the average fuel price paid by farmers. The first step was to calculate the prices for the years 2000-2002 implied by the expenses data from the NASS and the quantity of fuel consumption from the Miranowski survey. A later estimate for 2005 was based upon fuel consumption estimates from a greenhouse gas study for agriculture and forestry. These estimates are shown in the cell range E49:H51. These prices are then transferred to the cells with the same years in column Q. Columns N and O show the EIA data

[1] These estimates are displayed graphically in a PowerPoint presentation made by Miranowski to a Farm Foundation conference (*Agriculture as a Producer and Consumer of Energy*), June 24, 2004. This presentation file was accessed on the web through www.**farmfoundation.org**/news/articlefiles/**370-miranowski.ppt**, 10/11/2012.

[2] Personal communication with Randy Schnepf, Congressional Research Service, August 7, 2012.

[3] Table 3.4 in the 2011 *Annual Energy Review* (EIA 2012a)

that relate for bulk and retail prices for motor fuels (diesel and gasoline.). The approach is to develop a calibration procedure that will yield a calculated fuel price that would approximate the price estimate developed from the Miranowski and the greenhouse gas study. For motor fuels, a single factor, which may be interpreted as the share of fuels that are not taxed, is selected. These factors are shown in the green cells (P47 and P71). For LPG and natural gas, the factors are simply multiplicative factors that yield prices that are roughly in line with the 2000-2002 and 2005 estimates (cells P96 and P121).

The projected prices that result from this calibration procedure are shown in column P. These prices are then used to deflate the expenditures data to quantities. The quantity estimates are shown in range E96 through H113. The total quantity of fuels in TBtu is shown in cells J99:J113.

The fuel quantities are transferred to the adjacent worksheet Intensity_estimates, column E. Columns C and D show the data for electricity consumption, to the extent that such data are available. Below is some discussion related to the issue of electricity data.

Column B shows the gross output measures from which to calculate energy intensity measures. These data come from a spreadsheet specifically created to assemble the output data used in the nonmanufacturing (for the 2020 update: NonMan_output_data_010420.xlsx, worksheet Indicators_NonMan_2018_BEA). Most of the data in this spreadsheet were, in turn, linked to spreadsheets used to process the downloaded values (gross output and value added) from the BEA website. This process was described above in the general documentation of how the BEA output series were developed for the manufacturing sectors and other sectors used in the construction of industrial energy intensity indexes.

Electricity consumption data

Unfortunately, the NASS survey of farm expenses does not separately identify electricity expenditures or quantities. The Census of Agriculture includes electricity expenses as part of “utilities.” The following definition is supplied in an Appendix B to the 2007 Census of Agriculture (USDA 2009):

“These data show the farm share cost of electricity, telephone charges, internet fees, and water purchased in 2007. Included in the water cost is water purchased for irrigation purposes, livestock watering, etc. Household utility costs were excluded from these items.”

Given the likely increase in internet usage by 2007, there appears to be no reliable method of breaking out electricity expenditures, as either an absolute magnitude or as a stable share of the overall expenditures in this category. Accordingly, the computed intensity electricity in farms has been maintained at its 2002 level for the current edition of the energy intensity indicators. In 2002 the estimate developed by Miranowski was less than 120 TBtu, or less than 12% of total energy used by farms in that year. Thus, the assumption of a constant intensity for electricity since 2002 is deemed to have little impact the trend of overall energy intensity by farms, as total energy use is dominated by fuels.[1]

The Miranowski energy consumption estimates and subsequent updates all relate to farms, both livestock and crops. The other segment of agriculture in its broadest sense includes forestry and fisheries, and agricultural services. Based upon the National Energy Accounts data, these activities accounted for less than 4% of total agricultural fuel use, and about 6% of agricultural electricity use, in 1985. A review of NEA documentation suggests that these estimates are based upon a limited set of

[1] As yet, there has been no attempt to try to ascertain the source and methodology behind the Miranowski electricity estimates. Another potential source for more recent data would exploit the work done to estimate the greenhouse gas emissions from the agricultural sector. For example, see http://www.usda.gov/oce/climate_change/index.htm. These activities will be pursued for future editions of the energy intensity indicators as resources permit.

data sources, including one-time surveys. The energy estimates for the agricultural services sector appear to be based on Volume 3 of the 1978 Census of Agriculture, which covers the activities of this sector. A perusal of the USDA website with historical census data indicates that 1978 was the last year that agricultural services were covered by the census.

Given the relatively small amount of energy that is consumed in these other agricultural sectors, they are currently excluded as a part of the energy intensity indicators system. It should be noted, however, that the electricity and gas consumption in agricultural services is likely included in the commercial sector reporting of energy use by electric and gas utilities (and thus captured in the energy intensity indicators for the commercial sector). Thus, for example, a building that serves as a field office for employees of a large pesticide or fertilizer company would likely be classified by utilities as a commercial account.

Final data transfer

The final estimates for electricity and fuels consumption and output (gross output and value added) are shown in columns Y through A in worksheet Compute_intensities. These values are subsequently transferred to spreadsheet Nonmanufacturing_reconciliation.*date*.xlsx before their use in the final spreadsheet to compute energy intensity indexes (industrial_indicators.*date*.xlsx). This process is described in section D below.

B. Mining

The energy consumption estimates for mining depend entirely on the various editions of the periodic census (ending in years with ‘2” and “7” since 1967). Up through 1987, the information for mining was collected under the title “Census of Mineral Industries.” From 1992 forward, the same information is part of the mining segment of the Economic Census (which now is the broad term for all the census surveys in the census years).

Table A.10 shows the website data sources for the mining sector. For the most recent census in 2007, the data were selected from a flexible download procedure that allows the user to select key data elements for each specific NAICS sector. The specific data items were 1) “quantity of electricity purchased” and 2) fuels consumed by type: a) quantity, and b) delivered cost. For the previous years, the data were derived from downloaded industry series reports (or selected pages). In these reports, the cost and quantity of electricity is found in Table 3 (Detailed Statistics by Industry) and Table 7 (Selected Supplies, Minerals Received for Preparation, Purchased Machinery, and Fuels Consumed by Type).[1]

[1] In the 1992 and prior Censuses of Mineral Industries, the purchased fuels data was shown in T

able 7b.

Table A.8. Sources for mining data by census year

Year	Source
2017	Mining: Summary Statistics for the U.S., States, and Selected Geographies: 2017 https://www.census.gov/data/tables/2017/econ/economic-census/naics-sector-21.html
2012	https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk
2007	From printout of selected data items from table on Census Bureau (factfinder) website: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ECN_2007_US_21SG12&prodType=table
2002	Downloadable pdfs of industry reports: extracted Table 3 and Table 7. List of pdfs available at https://www.census.gov/econ/census02/guide/INDRPT21.HTM
1997	1997: Downloadable pdfs of industry reports: extracted Table 3 and Table 7. List of pdfs available at: http://www.census.gov/prod/www/abs/ec1997mining-ind.html
1992	Downloadable pdfs of industry reports: extracted Table 3 and Table 7. List of pdfs available at: http://www.census.gov/prod/1/manmin/92mmi/92minif.html

Since 1997, the mining industries have been classified under three major 3-digit NAICS sectors: 211, Oil and Gas Extraction; 212, Mining (except oil and gas); 213 Support Activities for Mining. Unfortunately, there are no aggregations of energy data from the more detailed industries to this level. Thus, an estimation of electricity and fuel consumption must begin with the more detailed mining sectors, essentially 6-digit NAICS since 1997 and 4-digit SIC in earlier years. At the NAICS level, there are 29 specific industries as shown in Table A.11 (the word "mining" has been omitted from most of the official NAICS titles in the table).

For the most part, there was a one-to-one translation between the 4-digit SIC sectors and the 6-digit NAICS sectors. That permits a reasonable transition between the pre- and post-1997 data. As shown in Table A.11, the mining support sectors were all collapsed into a broad (3-digit) NAICS sector, whereas under the SIC the support industries were classified with their production counterpart industries. The only other notable change was the reclassification of oil and gas well drilling to be a part of this overall support sector.

Table A.9. NAICS Detailed Mining Sectors

NAICS	Industry
211111	Crude Petroleum and Nat. Gas
211112	Natural Gas Liquid Extraction
212111	Bituminous, surface
212112	Bituminous, underground
212113	Anthracite
212210	Iron Ore
212221	Gold Ore
212222	Silver Ore
212231	Lead & Zinc Ore
212234	Copper & Nickel
212291	Uranium/Radium/Vanadium
212299	All Other Metal Ores
212311	Dimension Stone
212312	Crushed and Broken Limestone
212313	Crushed and Broken Granite
212319	Other Crushed and Broken Stone
212321	Construction Sand and Gravel
212322	Industrial Sand
212324	Kaolin & Ball Clay
212325	Clay
212391	Potash, Soda, & Borate
212392	Phosphate Rock
212393	Other Chem. & Fertilizer Minerals
212399	All Other Nonmetallic Minerals
213111	Drilling Oil and Gas Wells
213112	Support Activities for Oil and Gas
213113	Support Activities for Coal Mining
213114	Support Activities for Metal Mining
213115	Support Activities for Nonmetallic Minerals

There is no major difficulty in assembling the electricity quantity data, as quantities (kWh) were asked of respondents as part of the census in all years. The situation with fuels is quite another matter. The census forms ask for both cost and quantity information for specific fuels. The overall set of data throughout the years is plagued by missing values resulting from “withheld” information, i.e., data “withheld to avoid disclosing data of individual companies.” A variety of methods were employed to try to work around this problem. In some cases, an approximate value could be derived on the basis of subtracting the cost of those fuels for which data were published from total cost. (This residual approach yields a control total for the two or more fuels with missing data). Because of the cyclical nature of the mining industry and the fluctuations in the industry structure, estimates (or proportions) missing for one census year could sometimes be imputed on the basis of a previous census.

In addition, there are two categories that complicate the estimation of fuel quantities, both of which only present information on total cost. The first is "Other fuels – liquefied petroleum gas, coke, wood, etc." and the second is "Undistributed fuels." "Undistributed" as defined by the Census Bureau "represents cost for establishments that did not report detailed data, including establishments that were not mailed a form."

With regard to "Other fuels", the assumption was that the dominant fuel was propane. The cost estimates were converted to quantities by the use of the price of propane published by EIA.[1] For undistributed fuels, the assumption was that the average price of the unreported fuels was the same as the reported fuels. Operationally, this assumption was implemented as follows. The cost and quantity of reported fuels was estimated. Then the ratio of the total cost of all fuels with respect to the cost of reported fuels was calculated. This ratio (> 1.0) was then used as multiplicative adjustment factor applied to the quantity of all reported fuels.

Unfortunately, beginning in the 2007 census, generally only the *cost* of fuels was reported. The quantity estimates for 2007 were based upon extrapolating the 2002 quantity estimates by means of the 2007/2002 total cost ratio divided by a factor representing the percentage change in the appropriate fuel price between 2007 and 2002. As discussed below, the lack of quantity information for specific was generally the case in the subsequent 2012 census, and likely to be for the 2017 census.

The consumption estimates for years up through 1985 were based upon the National Energy Accounts. The NEA estimates were also based upon prior editions of the Census of Mineral Industries. For non-census years the NEA interpolated the energy intensities between the successive census years. This same approach is followed in the current work using the more recent census data.

Further work is required to refine the estimates of energy consumption during some time periods. The recent updating work has not included electricity data for the 1987 census (and thus interpolates intensities between 1985 and 1992). Another activity for future work is to use the 1982 census data directly to estimate energy consumption. The NEA data for years beyond 1982 did not utilize the 1987 information in its post-1982 extrapolation methods. Overall, however, the past efforts to construct estimates of total energy consumption for the mining sector from the detailed census information is deemed to be a significant improvement over the previous methodology, which included this energy as part of the difference between total industrial and total manufacturing.

Data Processing and Estimation Steps

The collection and processing of the census data related to mining is performed in spreadsheet Mining_energy_*date*.xlsx (for 2020 update, *date* = 031020). The principal worksheet in the spreadsheet is labeled "Data 1987-2012." The general layout of the spreadsheet is to gather all the data from the various census reports in sets of contiguous columns for each fuel. For each census year, the available reported data for the cost and quantity of purchased fuels are shown in pairs of columns. As mentioned above, there are many missing values for the data, primarily "withheld" data to not compromise confidentiality of reporting firms. Many different methods of imputing and extrapolating the data from one census year to the next were undertaken. For example, when a quantity of purchased fuel was reported in one census, but not the next, purchased cost (after adjustment for price change) was used to extrapolate the quantity value. (See note in cell AY4). In other cases, the quantity of fuel

[1] Energy prices to end users, as published in Table 5.23, *Annual Energy Review 2011*, accessed at http://www.eia.gov/totalenergy/data/annual/index.cfm

consumption was derived by simple division of the purchased cost by the average price of the fuel for that year. These steps were done at the six-digit level of NAICS for individual mining industries as shown in the table above.

To take distillate fuel consumption as an example, the values of cost and quantity are shown in columns AN through AY in worksheet Data 1987-2012. In general, an attempt was made to shade the cells in which the imputations and extrapolations occurred. Up through 2007, most of this information was entered manually from the census reports. For 2012, the availability of spreadsheets from the census bureau eased the process somewhat. The primary data from the census was processed by the use of a pivot table that selected out the required fuel information. In each of the census years, the census produced special tables that included detailed information on materials and specific fuels consumed for individual mining sectors. The results of this processing can be found in two places in the Mining_energy spreadsheet. In worksheet Pivot_Table_2012, the values for the cost of specific fuels are shown. For the 2020 update, the census data for 2012 were again downloaded and processed via a pivot table. The results from this process were placed in columns EX through FP in worksheet Data 1997-2007_NOT USED (the NOT USED refers to historical estimates of fuel consumption).

For the 2012 census, there were a very sparse number of entries for quantity information. Thus, for this year, all of the quantity estimates were derived by use of the cost information from 2012 and earlier years, again adjusted for price changes. The resulting extrapolations of distillate fuel consumption are shown in column AY. The only sizable entry for quantity of fuels was that for distillate consumption in iron ore mining. However, the reported value 3,888 thousand barrels extremely inconsistent with the cost data, and prior census reporting, and was not used.

Special procedure for 2017

For the 2017 economic census, the special tables for materials and fuels for the 2020 update had not yet been released by the Census Bureau. However, the data for the *total* cost of purchased fuels and the purchased quantity of electricity were available from the general statistics data that have been released. (Note: the data for electricity is only available from this general table.) For purchased fuels, the general approach was to follow the concept of a Passche quantity index. The idea is to compute (for each specific mining sector) the total cost of fuels for 2017 by holding the “known” quantities of specific fuels at their 2012 levels but using 2017 prices for the specific fuels. The *published* value for total fuels is essentially the product of the unknown levels of fuels in 2017 multiplied by the 2017 prices for fuels. The ratio of these two estimates (published 2017 value divided by the computed 2017 value with 2012 quantities) is a Paassche quantity index. This index can then be used to extrapolate (by simple multiplication) the total quantity of fuels in 2012 to provide an approximate quantity of fuels in 2017. Thus, for example, if the quantities of individual fuels were the same in both 2012 and 2017, the Paasche quantity index would compute to be 1.

The actual procedure is somewhat more complicated because we do not have the entire set of quantity and price information for 2012. As result, the procedure uses only the major fuels in the calculation. A “cost relative” value for each six-digit NAICS sector is computed for both 2012 and 2017, using the major fuels (coal, distillate fuel oil, residual fuel oil, natural gas, and gasoline) and prices by fuel for 2012 and 2017. The prices for 2017 are extrapolated from the 2012 prices by the change in national average prices between 2012 and 2017. Term this value CR1. A second cost relative (CR2) is computed, using the 2017 and 2012 *total published* cost of fuels for each sector (i.e., 2017 cost/2012 cost). Note that this cost relative will reflect both changes in prices and quantities between 2012 and 2017. A quantity multiplier for each NAICS is computed as the ratio of CR2/CR1 (in essence, to remove the overall effect of price change only.) This multiplier is then used to extrapolate the "known" quantity of total fuel consumption from 2012 to 2017.

All of these calculations are performed in separate worksheet Preliminary estimates 2017. The key elements of the process are as follows. In columns G through L, the estimated fuel 2012 consumption for each set is linked to prior work shown in worksheet Data 1987-2012. In columns O through S, the purchased cost is brought in from worksheet Data 1987-2012 for each of the five major fuels. In columns V and W, one can compare the total cost for the five major fuels as compared to the published total cost for all fuels. For most sectors, the total cost for these five fuels comprises a large fraction of the total published cost for all fuels.

In columns Y through AC are the extrapolated estimates of the cost in 2017, just using the change in prices between 2012 and 2017 to make the extrapolation. The 2012 and 2017 prices to make this calculation are shown in range BB3:BF4. The calculated total cost (5 fuels) for 2017 is in column AD. The cost multiplier associated with the 5 fuels, CR1, is in the adjoining column AE.

The total published costs of purchased fuels for 2012 and 2017 are shown in columns AH and AI. The cost ratio (2017/2012 = CR2) is in column AJ. As discussed above the quantity multiplier is calculated as the ratio of CR2/CR1. These values are in column AK. The final step is to apply the quantity multiplier to the estimate of total quantity of purchased fuels in 2012 (shown in column M). The estimated 2017 quantities are shown in the yellow-highlighted column AL.

There is no need to go through this extrapolation procedure for electricity since the values for purchased quantities of electricity were available in the initial data releases from 2017 economic census. These values are shown in the yellow-highlighted column AT. Column AU shows the reported values for 2012 for comparison.

The 2017 values for fuels and electricity consumption are transferred to the Data 1987-2012 worksheet. The estimates for total fuels are in column HL and the values for electricity in column HD. The values for the previous years of the economic census (1987 – 2012) are in the preceding columns.

Aggregate and review the energy and output data for mining

Below these areas in the worksheet Data 1987-2012, the six-digit NAICS sector fuels and electricity estimates are aggregated to 10 sectors, five of which are mining sectors (proper) and the other five are mining support services. Drilling oil and gas wells (NAICS 213111) is included in the NAICS three-digit sector 213, Support activities for mining. The aggregated data are show in arrays GY45:HD54 for electricity and HF45:HL54 for fuels.

The data are subsequently linked to identical arrays at the top of worksheet Sector_estimates. The arrays are transposed to show years down consecutive rows. These transposed arrays are E36:N42 (electricity after conversion to TBtu) and P25:Y31 (for fuels).

These transposed arrays are then linked to worksheet Compute_intensities, in rows 5 through 11. The data from the economic censuses from 1987 forward are then combined with estimates from the National Energy Accounts (NEA) for years up through 1985. A description of the National Energy Accounts is provided in the comprehensive documentation report (Rev. 3). The NEA data are shown in two arrays, L17:P44 for electricity and R17:V44 for fuels.

A large portion of the Compute_intensities worksheet is devoted to examining the consistency of the energy and output data for the above five mining sectors, and an aggregated sector encompassing all mining support activities. The calculations for Oil and gas extraction (Crude Petroleum and Natural Gas, old SIC name) are shown in columns Z through AS. Columns AB and AC show the energy data from the NEA, followed by the periodic values from the various censuses. The gross output estimates are derived from the BEA estimates, described earlier for the manufacturing sector. The output measures are linked from a separate worksheet, BEA&BLS_Output_data, to be discussed below. The output and energy data are plotted in the graph in columns AE through AK. The output data are scaled to be able to be shown on the same graph as the fuels and electricity data.

In the columns to the right of the graph, the energy intensities are computed for electricity and fuels. The intensities in this particular sector were set to a constant value between 1970 and 1977 (see note in cell AO24). For other years in which no consumption estimates were available the intensities were interpolated linearly between appropriate years. For those years, the consumption values were computed as the output times the intensity. The resulting series of electricity and fuels consumption are in columns AR and AS.

The same general layout of displaying the output and energy data for the other mining sectors was followed. As one moves across the columns, these data are shown successively for coal mining, metal mining, nonmetallic mineral mining and mining services. The values in columns DJ through DY are based on an aggregation of the first three of the above mining sectors – coal, metals, and nonmetallic minerals. This aggregate is labeled “Other Mining.” This aggregation to an “Other Mining” sector is one of three sectors used the energy intensity hierarchy. The rationale for this choice is that the BEA provides historical estimates for gross output and value added for this grouping of mining sectors (NAICS 312). The BEA also provides gross output and value-added estimates for Oil and Gas Extraction (NAICS 311), and Mining Services (NAICS 313). (Note: the values for coal mining, metal mining, and nonmetallic mineral mining spanning columns AV through DG are shown to help detect data anomalies in the energy data; Energy intensity indexes for these sectors on an individual basis are *not* used in the indicators hierarchy for mining.).

The final results from the calculations in “Compute_intensities” worksheet are shown in columns FF through FN. Annual estimates for electricity consumption, fuels consumption, and gross output for the final three mining sectors employed in the indicators framework: Oil and gas extraction, Other mining, and Mining services. These estimates are linked to worksheet Mining in the industrial_indicators spreadsheet: industrial_indicators_*date*.xlsx.

Output data for mining

The output measures used in the Compute_intensities worksheet were taken from the next worksheet, BLS&BEA_Output_data. The gross output measures from BEA are shown in columns M through O. The steps on how to download and process this data from the BEA website were explained previously (immediately after the section on manufacturing energy data).

The BLS data come from the database of sector-level historical estimates of output and employment maintained by BLS. This data is employed by BLS in their program to project employment and occupational demands by sector in the U.S. economy. The data file provided by BLS shows three values for each sector: gross output in constant 2012 dollars, a price index for output and the number of employees. These estimates are made for 205 sectors covering the entire economy (plus another set of estimates at a more aggregate level).

A zip file (folder labeled “industry”) can be downloaded from the BLS “Employment Projections” website: https://www.bls.gov/emp/data/industry-out-and-emp.htm. The download of the historical data is accomplished by clicking on the link "(ZIP)" at the end of the sentence "Industry output and employment data for the U.S. economy for historical years, and for the projected year 2028 (ZIP)." This folder contains the spreadsheet with the historical data as well as documentation files. In the most recent update, the data file with the [comma delimited) historical data is named simply "ind18." This file was loaded into Excel and labeled "BLS Output data 1972-2018.xlsx." This spreadsheet contains a single worksheet from the BLS, labeled "emp18."

Rows 5 through 9 in worksheet BEA&BLS output_data show the data extracted from worksheet "emp18" derived from the BLS file. As enumerated in column E, the (gross) output values were pulled for BLS sectors 7 through 11. The cell references in the array to the right indicate that the output series were taken from every third row of the "ind18" data, beginning with cell C21 (value for 1972 for sector 7). The BLS output data for these five sectors were transposed in worksheet BEA&BLS_output data, resulting in the array F15:J61.

Some casual comparison suggests that the underlying source for the BLS estimates for gross output is the BEA. For sectors defined by the same NAICS sector, the gross output measures are often identical or very close to each other. This is particularly true for manufacturing sectors. Outside manufacturing, the estimates can sometimes be very different, probably owing to somewhat different specific NAICS industries included in the sector. These differences show up in several of the mining sectors, particularly the older (pre-2006) estimates for oil and gas extraction and mining services. As stated above, the BLS output measure was used for oil and gas extraction for the full historical series. (As mentioned above, the BLS and BEA measures do converge over the period from 2006 and later.) For mining services as well, the choice was made to use the BLS series of gross output, rather than that from BEA. See note in cell FN16 in worksheet Compute_intensities.

The estimates developed in worksheet Compute_intensities are transferred to two other spreadsheets, Nonmanufacturing_reconciliation and industrial_indicators. This transfer process is explained in some detail Section D below.

C. Construction

Data for expenditures for various types of energy are available from the various editions of the Economic Census. For 1992 and earlier years, the census went under the title Census of Construction Industries, and the data were collected (and classified) under the SIC. Since 1997, the collection of data from the construction is subsumed under the general Economic Census with Construction as one of industry groups. As elsewhere, the construction sectors have been classified under the NAICS since 1997.

For the energy intensity indicators system, energy intensity is estimated only for total construction. Thus, the approach here is to aggregate across all of the detailed construction sectors when necessary (26 4-digit SIC sectors, and 29 5-digit NAICS sectors).

The data for energy expenditures is included in tables for "detailed statistics" for all years. Conveniently, Table 2 in the 1992 Census (Detailed Statistics for Establishments with Payroll: 1992 and Earlier Census Years) contained information for 1992 and for three previous census years: 1977, 1982, and 1987. These data were obtained from individual 4-digit SIC industry reports. The source of the 1992 information, as well as the census data for subsequent years, is shown in Table A.12.

Table A.10. Sources for construction data by census year

Year	Source
2017	Construction: Summary Statistics for the U.S., States, and Selected Geographies: 2017 https://www.census.gov/data/tables/2017/econ/economic-census/naics-sector-23.html
2012	Construction: Summary Statistics for the U.S., States, and Selected Geographies: 2012 https://www.census.gov/data/tables/2012/econ/census/construction.html
2007	2007: Construction: Industry Series: Preliminary Detailed Statistics for Establishments: 2007, downloadable spreadsheet from: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ECN_2007_US_23I1&prodType=table.
2002	2002: Construction Industry Series: Detailed Statistics for Establishments, downloadable spreadsheet from: http://factfinder2.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ECN_2002_US_23I04A&prodType=table
1997	1997: Industry Summary, 1997 Economic Census, *Construction*, Subject Series EC97C23S-IS, downloaded from: http://www.census.gov/epcd/www/97EC23.HTM
1992	1992: Industry Series (CC-92I), with downloadable pdf files for each 4-digit SIC construction industry. Downloadable from: http://www.census.gov/prod/www/abs/cciview.html

All of the census information for construction shows expenditures for four fuel types: 1) electricity, 2) natural and manufactured gas, 3) gasoline and diesel fuel, and 4) other, including lubricating oils and greases. Beginning with the 1987 census, category 3) was broken out into “on highway use” and “off highway use”).

Because “on highway use” is covered in the transportation sector, the focus in this work was on “off highway use.” For the years prior to 1987 (1977 and 1982), the share of total gasoline/diesel use for off-road activity was taken as the average across 1987 and 1992on. This imputation was done for each of the 26 4-digit SIC construction industries.

The energy expenditures estimates from the available census data are shown in Table A.13. Clearly, the increase in construction activity and overall increases in fuel prices lead to a large change between 2002 and 2007. An even greater jump occurs between 2012 and 2017; the 2017 figures apparently reflecting a robust recovery from the 2009 recession and the subsequent economic expansion.

Table A.11. Expenditures by fuel type for the U.S. construction, census years millions of dollars

			Gasoline and diesel		
	Electricity	Natural Gas	On-Road	Off-Road	Other/Lubricants
1977	468	227	2,573	779	472
1982	823	266	4,534	1,417	407
1987	1,089	304	4,125	1,602	519
1992	1,469	464	4,480	1,761	517
1997	1,741	515	5,336	2,117	548
2002	2,596	1,025	8,024	2,886	868
2007	5,749	1,709	17,977	5,678	1,188
2012	2,698	846	14,795	4,943	829
2017	7,926	2,555	29,038	7,692	1,999

Source: Census publications listed in Table A.12

Note: On-road and off-road values for 1977 and 1982 are imputed (see text).

The expenditures by fuel type were converted to quantities using fuel prices published by EIA. The sources for fuel prices are:

- **Electricity:** Table 3.4, Consumer Price Estimates for Energy by End-Use Sector, commercial sector, in the *Annual Energy Review 2011* (EIA 2012). Updated with (commercial) prices from Table 9.8 in the *Monthly Energy Review*.
- **Natural gas:** Same source as for electricity through 2007. Updated with (commercial) prices from Table 9.10 in the *Monthly Energy Review*.
- **Gasoline:** Table 5.23. All Sellers Sales Prices for Selected Petroleum Products, Sales Price to End-Users, Conventional motor gasoline, *Annual Energy Review 2011.* After 2010, updated with prices from Table 9.4 in *Monthly Energy Review*
- **Diesel:** Table 5.23. All Sellers Sales Prices for Selected Petroleum Products, Sales Price to End-Users, No. 2 Diesel Fuel, *Annual Energy Review 2011.* After 2010, updated with prices from Table 9.7 in *Monthly Energy Review*

For the off-road category of expenditures, it was assumed that the majority of equipment would be powered by diesel (road construction equipment, large compressors, etc.). Therefore, the composite price for off-road fuel was calculated with 30% weighting applied to gasoline and 70% weighting to diesel. There is no EIA published price for lubricants. Here the assumption was that lubricants were 3 times more expensive than the composite gasoline/diesel price.

Using these sources (and assumptions) for fuel prices, Table A.14 shows the estimated electricity and fuel consumption by census year.

Table A.12. Estimated energy consumption for the construction sector by fuel type and census year, trillion Btu

	Electricity	Natural Gas	Off-road Gasoline/Diesel	Other/ Lubricants	Total Fuels
1977	39.0	113.5	326.9	99.0	539.4
1982	40.9	56.5	259.9	37.3	353.7
1987	53.3	65.4	389.3	63.1	517.8
1992	66.3	97.7	360.3	52.9	510.9
1997	79.0	90.8	385.3	49.9	526.0
2002	113.8	157.9	451.2	67.9	677.0
2007	203.4	155.7	326.6	34.2	516.4
2012	91.3	107.9	203.0	17.0	327.8
2017	253.8	334.8	546.3	71.0	952.1

Source: Expenditures in Table A.13 converted to quantities via cost estimates described in text.

The values in Table A.14 were used to develop energy intensity estimates for electricity and total fuels for each of the census years. Similar to mining, the energy intensities were then interpolated for the non-census years.

For years prior to 1977, the consumption estimates were extrapolated back to 1970 by use of the National Energy Accounts series for construction sector electricity and fuel use.

Specific steps to download and process census data for construction

Some of the specific steps to download and process the census data on construction energy costs are explained in the following paragraphs. The top-level census bureau website for the Economic Census is: https://www.census.gov/programs-surveys/economic-census.html. Scroll down the page until the words "2017 Data Tables" are found. After clicking on that link, the user will end up at https://www.census.gov/programs-surveys/economic-census/news-updates/updates/2017-data-tables.html. The "2017 Data Table pages" now include direct links into data.census.gov and large ftp downloads. After clicking on pages, the webpage https://www.census.gov/programs-surveys/economic-census/data/tables.html comes up. Scroll down this page until the entry "Construction (NAICS Sector 23)" is found. After selecting this entry, the user is then automatically transferred to: https://www.census.gov/data/tables/2017/econ/economic-census/naics-sector-23.html.

Of the various choices, the user should select the link EC1723BASIC under "Datasets (Links to FTP)." Either "open" or "save;" the result will be a spreadsheet that shows various data items across the columns and the various NAICS sectors down the rows. The only sector of interest is the very first row of data – the data for all construction, NAICS 23. Only five variables are necessary to capture: CSTELEC (purchased electricity), CSTFUNG (cost of natural gas), CSTFUOF (cost of gasoline and diesel, off-road), CSTFUON (cost of gasoline and diesel, on-road), CSTFUOT (cost of other fuels). All values are in terms of thousands of dollars.

The spreadsheet to develop the preliminary energy intensity indicators for construction is entitled, "Construction_energy_*date*.xlsx. The cost of fuels (and electricity) data from the economic census are manually keyed into worksheet Construction, in cell range D9:H18. The layout of this worksheet is similar to that used for agriculture, with the exception that the prices for fuels are taken directly from EIA sources (i.e., without any calibration to special studies of fuel use). The prices for fuels ($/MMBtu) are shown in the array d34:H46. As shown in the cell notes for 2012 for electricity and natural gas, the prices from the earlier *Annual Energy Review* are updated with the relevant prices from *Monthly Energy Review* tables. The prices for motor fuels and "other" are developed in worksheet "Petroleum-prices" and linked into columns G and H.

The quantities of fuel consumption are derived from dividing the expenditures by the average price of fuels. The quantity estimates are shown in cell range D52:J60. The census year quantity estimates for electricity and fuels are then transferred to columns F and G in worksheet Intensity_estimates.

The output data for construction is taken from spreadsheet NonMan_output_data_*date*.xlsx (in 2020 update, *date* is 010420). The output values are gross outputs in 2012 chained dollars, downloaded from the BEA website as explained above. The values for gross output and fuel and electricity quantities are plotted in the top graph in columns I through P. In the graphic below, the intensities are plotted, with non-census year intensity values interpolated linearly between successive census years. In columns W through Y are the values that are transferred to the spreadsheet Nonmanufacturing_reconciliation_*date*.xlsx. From this spreadsheet, the energy and output estimates for construction are subsequently transferred to the industrial_indicators spreadsheet to construct the relevant intensity indexes. The process is explained in the following section.

D. Final steps for nonmanufacturing

The spreadsheet Nonmanufacturing_reconcilation.*date*.xlsx is used to assemble the energy consumption and output values for the three major nonmanufacturing sectors: agriculture, mining, and construction. Prior to making a focused effort to develop time series estimates of energy consumption for these sectors, total energy use in nonmanufacturing was calculated as a residual. Total industrial energy use from EIA (supply-side sources) was the starting point. The estimated energy consumption for manufacturing (and feedstock energy use) was subtracted from the industrial total. After the work to develop independent estimates of nonmanufacturing energy use was first initiated, there was some effort to compare the estimates of nonmanufacturing energy derived from these alternative methods (i.e., residual vs. end-use consumption information). This short explanation provides the rationale for the label "reconciliation" in the initial naming of this spreadsheet. This work to compare the estimates from the two methods was not attempted for the 2020 update of the indicators.

Thus, for purposes of an update, the key worksheet in the Nonmanufacturing_reconcilation spreadsheet is labeled "Final NonMan." This worksheet assembles the time series estimates from the three separate nonmanufacturing sectors. First, columns C through F show the energy consumption and output values for agriculture. All of these values are linked to the particular spreadsheet where these estimates were produced (and assembled), worksheet Intensity_estimates in spreadsheet Agricultural_energy_*date*.xlsx. Columns P through S show the energy and output time series for construction. These values are linked to the spreadsheet Construction_energy.*date*.xlsx (again, worksheet "Intensity_estimates).

The center block of columns (I through L) shows the energy and output values for mining. As for agriculture and construction, these values come from the special spreadsheet for mining, Mining_energy.*date*.xlsx. However, in the case of mining, these values are not subsequently used directly in the process to estimate an overall intensity measure for nonmanfacturing. They are shown here for comparison purposes only. Rather, the more detailed estimates for the individual mining sectors (Oil and gas extraction, Other mining, and Mining services) are employed. (See below).

The values in worksheet Final NonMan are linked to the spreadsheet that calculates the energy intensity indexes within the industrial sector indicators hierarchy (18 manufacturing sectors, two broad sectors in nonmanfacturing – agriculture and construction, and the three mining sub-sectors.) This final spreadsheet is industrial_indicators_*date*.xlsx. The values from worksheet NonMan in spreadsheet Nonmanufacturing_reconciliation are linked to worksheet Nonmanufacturing_Data in the industrial indicators spreadsheet. These values are put into columns C through S in worksheet NonManufacturing_Data. The values for mining are highlighted in blue; to repeat, these values are not used in the calculation of intensity indexes.

The estimates of energy consumption and output for the three mining subsectors are transferred (via active links or copied values) from spreadsheet Mining_energy.*date*.xlsx (columns FF through FN in worksheet "Compute_intensities) to the industrial indicators spreadsheet -   worksheet Mining. The gross output estimates are linked to cell range J36 through L83 in worksheet Mining. The electricity estimates are transferred from the Mining_energy spreadsheet to cell range D194:F241 in worksheet Mining. The estimated values for fuel consumption are just below, in cell range D276:F322. From these data inputs, the annual energy intensity indexes for fuels, electricity, delivered energy, and source energy are computed.

V. Electric Generation Sector: Process Steps and Technical Notes

A. Background

The energy intensity indicators for electricity generation are contained in a single spreadsheet, electricity_indicators_date.xlsx, where date is expressed in mo/day/yr format. The indicators for electricity are derived entirely from data collected by EIA. In recent years, the indicators are based entirely upon the EIA-923 survey.

Prior to 2012, the data for the indicators were taken directly from tables published (and downloaded in excel format) from EIA's Annual Energy Review (AER). The data for energy consumption to produce electricity were generally supplied in physical units only (e.g., mcf of natural gas, tons of coal, etc.) The values needed to be converted to Btu, and still be consistent with aggregate energy consumption for this sector as published by EIA. For each major fossil fuel, a separate worksheet was developed; these worksheets are identified with the suffix "reconcile." Thus, the worksheet "NatGas Reconcile" seeks to produce an estimate of the Btu consumption of natural gas used to generate electricity. Similar worksheets were developed for coal, petroleum, and other fuels.

In 2012, EIA discontinued publication of the Annual Energy Review. The expanded Monthly Energy Review did not contain all the tables to continue calculation of the intensity indicators at the level of disaggregation used in the EERE system of indicators. After contacting EIA, an EIA analyst, Channelle Wirman, graciously provided the required data needed to continue updating the tables used in the indicators. Wirman processed the various datafiles from the EIA-923 survey and sent PNNL a spreadsheet file with results for 2005 through 2012. These numbers were entered into the appropriate "_reconcile" worksheets in the electricity_indicators spreadsheet. For example, the values supplied by EIA for natural gas are shown in the columns AD and AE in the light-pink shaded area of the "NatGas Reconcile worksheet.". To the left of these numbers are the estimated consumption and generation values developed by PNNL. No effort was made to go back to years prior to 2005 to obtain estimates that would utilize the EIA estimates directly.

In the 2016-2017 update of the intensity indicators, the utility-level micro data from EIA were processed by PNNL to develop indicators for 2013 and 2014. This processing involved generation of several pivot tables from the detailed EIA worksheet file. This process is explained in detail in the following two sections.

B. Process data directly from EIA-923 utility-level data

From the EIA website, download the detailed EIA-923 micro data file. For 2017, the filename was EIA923_Schedules_2_3_4_5_M_12_2017_Final_Revision.xlsx. Save this file and append the filename with a label to keep separate from the download file. PNNL typically appended the label "_pnnl" to this file. Load the file into Excel. Go to worksheet labeled "Page 1 Generation and Fuel Data."

1. Insert a pivot table for electricity generation. Highlight (i.e., "select") the area in the worksheet that contains all the column titles and all data rows. Go to Insert tab in Excel and check box for Pivot Table. Check box to put the pivot table in a separate sheet.

2. Check boxes for variables required in the pivot table: a) Combined Heat and Power Plant; b) EIA Sector Number; c) AERFuel Type Code; and d) Net Generation (Megawatthours).

3. Every pivot table has four fields, in four separate quadrants that are used to define the table. These quadrants should be populated as follows: i) for "Filters," drag the Combined Heat and Power

Plant variable to this field (upper left-hand quadrant of pivot table). ii) Drag EIA sector number field to the “Columns” field (upper right quadrant), remove field for “values”; iii) Drag AERFuel Type Code to the “Rows” area (lower left quadrant of pivot table); iv). Drag Net Generation (Megawatthours) to lower right-hand quadrant of pivot table.). For this variable, make sure you have the “Sum of” type of calculation. This box will read “Sum of Net Generation (Megawatthours)

4. Label the sheet where Pivot Table is inserted, “Net Generation.” The pivot table will have 18 rows, labeled by type of fuel. There will be seven columns (B through H), and one additional column for the “Grand Total.”

5. Go back to the worksheet with the primary utility-level data from EIA. Ensure that the same portion of the worksheet is highlighted – column titles and all rows. Check menu item in Excel to add another pivot table in separate sheet. Repeat steps 3 through 6 above. Drag variable name for “Elec Fuel ConsumptionMMBtu” into the lower right quadrant of pivot table.

6. Label the sheet where this second pivot table is inserted as “Elec Btu Consumption.” This table will be identical in structure to the table for electricity generation.

C. Aggregate data from pivot tables.

1. Find the spreadsheet entitled “S923 Pivot_Tables.xlsx,” the table employed by PNNL to summarize the data from the pivot tables. Go to worksheet ElecGen. This worksheet contains the pivot tables developed from the EIA datafiles for the years 2014 through 2017. Add the next year’s pivot table for electricity generation below the last previous year. One can simply do a “copy values” or link the relevant data to the large extract file (with the micro data from the EIA-923 survey).

2. Copy down the formulas that will aggregate the data from the pivot table. These formulas are found in columns L through P. Column L (for electricity-only plants) sums the pivot table columns, B and C. Column L reproduces column D (CHP plants in the electric power sector). Column O (commercial) sums columns F and G from the pivot table. Finally, column H is transferred to column P – industrial.

3. In the columns further to the right are the formulas to aggregate by generation and fuel type. Copy these formulas down from the previous year. These formulas are in columns S through X. Make sure the headings align on the same row.

4. Further to the right on this worksheet are the aggregate data values that can be transferred directly to the electricity_indicators_date.xlsx spreadsheet. These are organized to be transferred to four separate worksheets in the electricity_indicators spreadsheet (however, see step 5 below). The first set is for the generation from fossil fuel plants in electricity-only plants (columns AH through AM). In columns AP to AS are the generation from various fuels from CHP plants in the electric power sector. In columns AV and AW are the entries for generation in CHP plants using renewable fuels (wood, waste).

At the bottom of this worksheet (ElecGen) are the areas where the data for the years above are put into successive rows. In columns BA through BG are the entries associated with hydroelectric generation for the three sectors: electricity-only plants, commercial, and industrial.

5. The final results are transferred to the separate worksheets for electricity-only plants and CHP plants in the electric power sector. These worksheets are titled “Electricity-only” and “CHP.” From these worksheets the results for the latest years can be transferred to the electricity_indicators_date.xlsx spreadsheet.

6. Repeat the above steps for the fuel use used in electricity generation. In the S923_Pivot_Tables spreadsheet, go to the Elec Btu Consumption worksheet. Again, the pivot tables from years 2014

through 2017 are arranged from top to bottom. At the right of each of these tables are the fomulas that aggregate these results, first by sector, and second by type (or fuel) of generation.

The results collected for all the years are collected at the bottom of the worksheet (now in rows 109 to 113 for electricity generation (ElecGen worksheet) and rows 105 to 109 for fuel consumption (Elec Btu Consumption).

The final four worksheets (rightmost along the bottom) are labeled Commercial, Industrial, Electricity-only and CHP. The data for consumption and generation are transferred into the appropriate worksheet of this set of worksheets. These worksheets rearrange the data to yield the same ordering as that used in the electricity_indicators spreadsheet. Thus, the worksheets facilitate copy and paste operations (with or without active links) to move the data into the electricity_indicators spreadsheet. In general, these worksheets contain the separate data categories for fossil fuel generation and renewable generation.

D. Load data from pivot tables into Electricity_indicators spreadsheet.

For the most recent update in 2020, the loading of the pivot table summarization of the EIA-923 data was accomplished with active links. The links for the 2014-2017 period are shown in the following worksheets of the Electricity_indicators spreadsheet (from left to right): Industrial_Sector_CHP>Renew, Industrial_Sector_CHP>Fossil, Comm_Sector_CHP>Renewable, Commercial_Sector_CHP>Fossil, Elec_Power_Sector>Renewable Elec_Power_Sector>Fossil, Electricity-Only>Renewable, Electricity-Only>Fossil, Electricity-Only>Total (for hydroelectric).

In previous editions of this spreadsheets that were available on the EERE website, no links to supporting spreadsheets were retained. In such cases, the data values only were shown in the spreadsheets, although in some cases notes were inserted to identify the supporting spreadsheet.

Pacific Northwest
NATIONAL LABORATORY

U.S. DEPARTMENT OF
ENERGY

Proudly Operated by **Battelle** *Since 1965*

902 Battelle Boulevard
P.O. Box 999
Richland, WA 99352
1-888-375-PNNL (7665)