



Independent Statistics & Analysis
U.S. Energy Information
Administration

Appendix D

Statistical Methodology of Estimating Petroleum Exports Using Data from U.S. Customs and Border Protection

Updated March 14, 2018



This report was prepared by the U.S. Energy Information Administration (EIA), the statistical and analytical agency within the U.S. Department of Energy. By law, EIA's data, analyses, and forecasts are independent of approval by any other officer or employee of the United States Government. The views in this report therefore should not be construed as representing those of the U.S. Department of Energy or other federal agencies.

Introduction

Beginning August 31, 2016, EIA took its first step to incorporate export data collected by the U.S. Customs Border Protection (CBP) through the Automated Commercial Environment (ACE) system for estimating U.S. petroleum exports used in the *Weekly Petroleum Status Report (WPSR)*. The updated WPSR methodology provided a more accurate estimate of U.S. petroleum exports to better inform EIA's weekly estimates of domestic quantity demanded for crude oil, petroleum products, and biofuels.

Exports are an important component for the calculation of EIA weekly estimates of quantity demanded for petroleum products. EIA calculates domestic quantity demanded (measured as product supplied) for petroleum products as:

$$\text{Domestic quantity demanded} = \text{Production} + \text{Imports} - \text{Stock change} - \text{Exports.}$$

EIA conducts weekly surveys to collect data that are used to estimate production, imports and stocks, but not exports. Export data are published on a monthly basis by the U.S. Census Bureau (USCB). To estimate weekly exports, EIA previously used forecasts based on an autoregressive integrated moving-average (ARIMA) procedure to model weekly export statistics using the latest USCB monthly data. EIA obtained official USCB export data for crude oil, petroleum products, and biofuels on a monthly basis approximately six weeks after the close of the reporting month and published them in the *Petroleum Supply Monthly (PSM)*. These data provided the inputs to the weekly export ARIMA model. However, as the volume and variability of U.S. exports increased, significant discrepancies were sometimes observed between modeled weekly estimates and final monthly exports data provided by the USCB. Because of the growing impact of exports on EIA's weekly estimates of U.S. petroleum quantity demanded, EIA recognized the need to improve the weekly export estimates used in the WPSR.

Beginning in 2013, EIA investigated data sources and methodologies to improve its weekly estimates of petroleum and biofuels exports. EIA found that administrative data collected by CBP through the ACE system, which cover U.S. exports except for most of those going to Canada, could be used to improve EIA's weekly export estimates. To obtain these data, EIA established a Memorandum of Understanding with CBP in September 2014 through the International Trade Data System (ITDS) process. In August 2016, EIA obtained a letter from Office of Management and Budget (OMB) granting EIA an exception to the provisions of OMB's *Statistical Policy Directive No. 3 on the Compilation, Release, and Evaluation of Principal Federal Economic Indicators*. This exception permits EIA to publish weekly estimates of export volumes of crude oil and six refined petroleum products and biofuel categories in advance of the release of official U.S. trade data from USCB.

Since August 31, 2016, EIA has monitored and continuously analyzed the estimates derived from the recently acquired near real-time CBP export data. After reviewing the results to date, EIA further revised its estimation methodology in three phases during 2017-2018. These changes were based on continuous monitoring and comparison of results with the *Petroleum Supply Monthly*. These changes are described in Attachment B. The remainder of this document describes the method deployed for the WPSR release on March 14, 2018 onward.

Trade Data Collection and Dissemination

To derive weekly estimates of domestic petroleum product supplies, the WPSR relies on estimates of weekly U.S. exports of crude oil, finished motor gasoline, kerosene-type jet fuel, distillate fuel oil, residual fuel oil, propane, and other oils including biofuels, asphalt, aviation gasoline, butanes, ethane, lubricants, gasoline blending components, miscellaneous products, non-biofuel oxygenates, petroleum coke, unfinished oils, and wax. The export data that EIA uses to develop these weekly estimates are collected by CBP.

In general, entities exporting products from the United States are required to file export transaction data with CBP. This information is then included in CBP's ACE system. As a result of an agreement between the U.S. Customs and Border Protection Agency, EIA now receives these unedited transactional ACE data files from CBP and uses them to develop aggregate estimates of weekly exports that are used in the WPSR. The data from CBP include an export date that makes it possible to group transactions into WPSR weekly reporting periods. EIA has found that data received from CBP require application of edits and imputation to avoid using questionable data values for analysis, modeling, and aggregation. Edit and imputation methods applied by EIA are described later in this document. EIA is unable to follow up with exporters to resolve questions about the CBP data. USCB may follow up with exporters to address questionable data and, if necessary, obtain corrections during their preparation of official U.S. trade statistics.

Per U.S.-Canadian agreements, entities exporting to Canada are not required to report to CBP, so these files generally exclude Canadian transactions. Products that require an export license are an exception, and export transactions involving licensed products must be reported to CBP regardless of the destination country. Crude oil was a licensed export product until December 2015. USCB has an agreement with Statistics Canada to obtain monthly Canadian imports data from the United States, which USCB then uses to produce monthly estimates of U.S. exports in official U.S. trade statistics. However, EIA does not currently have early access to Canadian import data. To address these issues, EIA developed methodologies to edit the raw CBP data and to estimate weekly exports to Canada.

The CBP data EIA uses to estimate weekly exports begin with the same data that the USCB uses as the initial input for preparation of the monthly *International Trade in Goods and Services* report (FT900), a principal economic indicator. The FT900 has a six-week publication lag, while WPSR data are released each Wednesday for the weekly report period ending the previous Friday. Because the monthly publication cycles are more closely aligned, the FT900 data are used by EIA to report monthly export volumes that are included in EIA's *Petroleum Supply Monthly* (PSM) and *Petroleum Supply Annual* (PSA).

Estimation Methodology

Prior to producing the published estimates for the WPSR, EIA performs the following edit steps after receipt of CBP data for all products:

- Remove duplicate records
- Exclude records with entry dates beyond the reporting period
- Exclude products that are out of scope of the WPSR
- Convert to barrels exported quantities reported in other units
- For each product, calculate the dollar value per barrel and compare against daily spot prices from Thomson-Reuters to identify extreme values. Use a procedure based on shipment or commodity weight to replace these values, if necessary. Shipment weight is the combined weight of the commodity being shipped and its packaging.

See Attachment A for a more complete list of edit rules that EIA applies to the CBP data.

Following editing, EIA's estimation method consists of two parts:

1. Estimates based on aggregations of edited CBP data

For those products that have adequate coverage in the CBP data, EIA reports the sum of the quantities exported using the edited data. These products include distillate fuel, propane, residual fuel oil, and finished motor gasoline. EIA will continue to monitor weekly export estimates for these products and may use some form of a post-editing statistical model if the weekly estimates consistently diverge from USCB's monthly exports data.

2. Estimates based on post-editing statistical models

For those products that appear to have coverage differences in the CBP data based on EIA's comparisons to the USCB monthly volumes that are published in the PSM, EIA uses a post-editing statistical model to correct for the coverage differences. These products include crude oil, kerosene-type jet fuel, and other oils. The post-editing statistical model is described below:

EIA uses a two-component approach to generate estimates. The first component is a linear regression model to estimate the exports to destinations other than Canada. The formula for the updated regression component is:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

where t denotes a month and:

- Y_t is the non-Canadian (from PSM) export of crude oil, kerosene-type jet fuel, or other oils in barrels/day as published in the PSM for month t

- x_t is non-Canadian edited CBP export data in barrels/day aggregated for month t to be used as the regressor
- β_0 is the intercept
- β_1 is the regression coefficient for the regressor x_t
- ε_t is the error term

The estimated regression coefficients are then applied to the non-Canadian edited CBP export data on a weekly basis to produce the published weekly exports data. This method assumes a similar linear relationship between the non-Canadian PSM and CBP data on a weekly basis for the weeks in a given month. EIA will continue to assess the results of this approach and identify alternative methods to develop weekly export volume estimates, if needed.

The second component is a three-month moving average of Canadian exports in thousand barrels per day as reported in the latest series of Petroleum Supply Monthly.

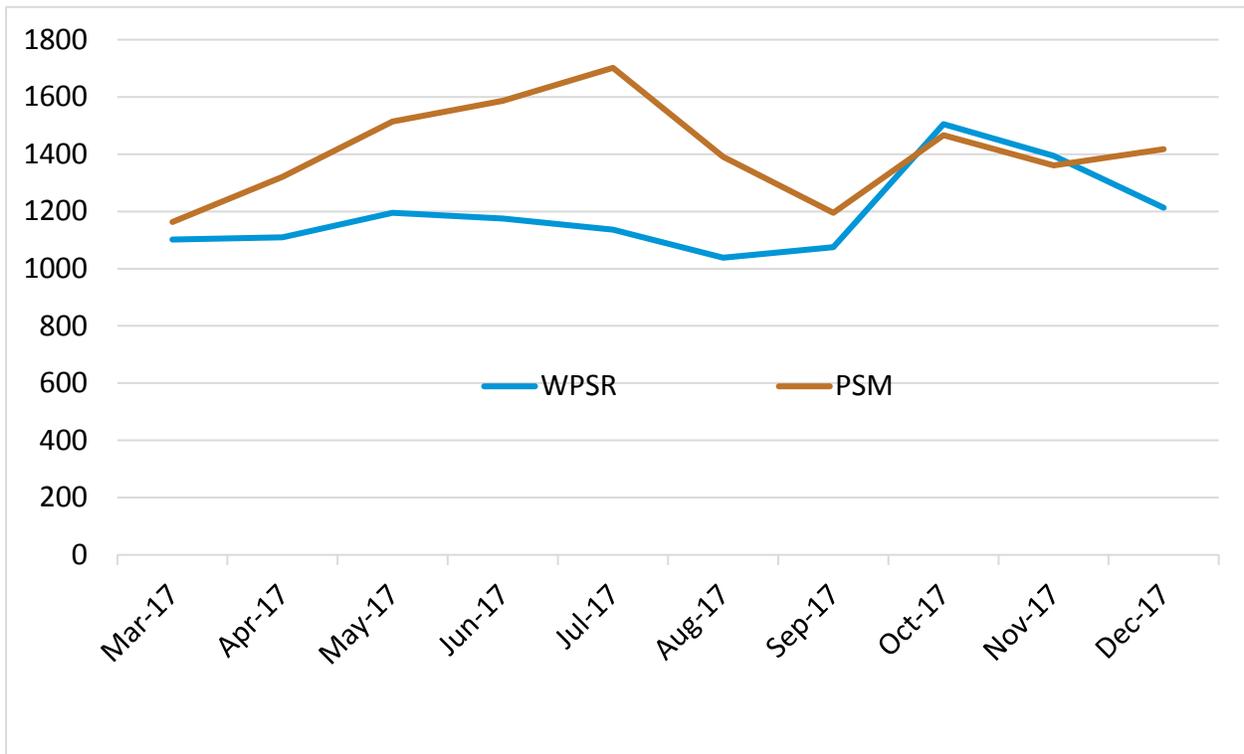
Results for Products

Estimates based on aggregations of edited CBP data

For the months from March 2017 to December 2017, Figures 1–2 show the currently published aggregated weekly estimates for distillate fuel oil and propane produced by using edited CBP data (blue line) compared to USCB/PSM monthly volumes (brown line). The estimation methodology for these products is based on aggregations of edited CBP data and is unchanged since August 2016.

Figure 1. Monthly distillate fuel oil exports for March 2017–December 2017

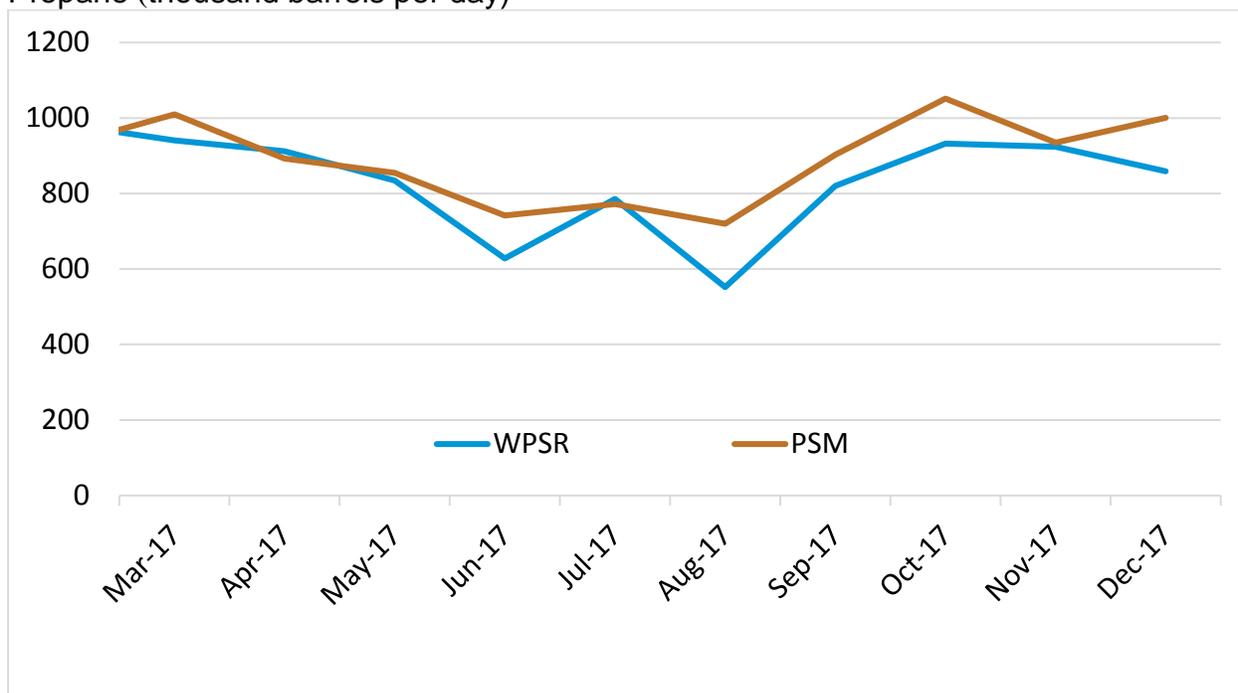
Distillate fuel oil (thousand barrels per day)



Source: U.S. Energy Information Administration

Figure 2. Monthly propane exports for March 2017–December 2017

Propane (thousand barrels per day)

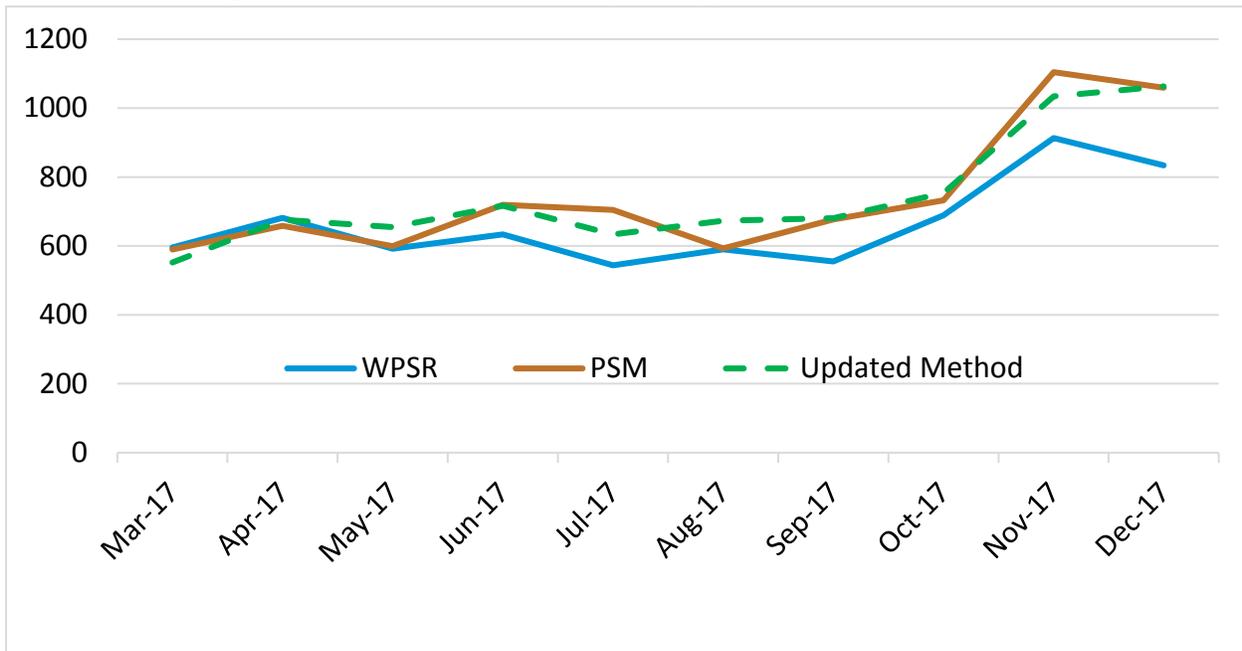


Source: U.S. Energy Information Administration

For the months from March 2017 to December 2017, Figures 3 and 4 show the aggregated weekly estimates for finished motor gasoline and residual fuel oil estimates using edited CBP (dotted green line) against USCB/PSM monthly volumes (brown line) and the currently published WPSR estimates based on the previous regression model (M2) (blue line). Beginning March 2018, the estimation methodology for these products are based on aggregations of edited CBP data, instead of the previous regression model used since March 2017.

Figure 3. Monthly finished motor gasoline exports for March 2017–December 2017

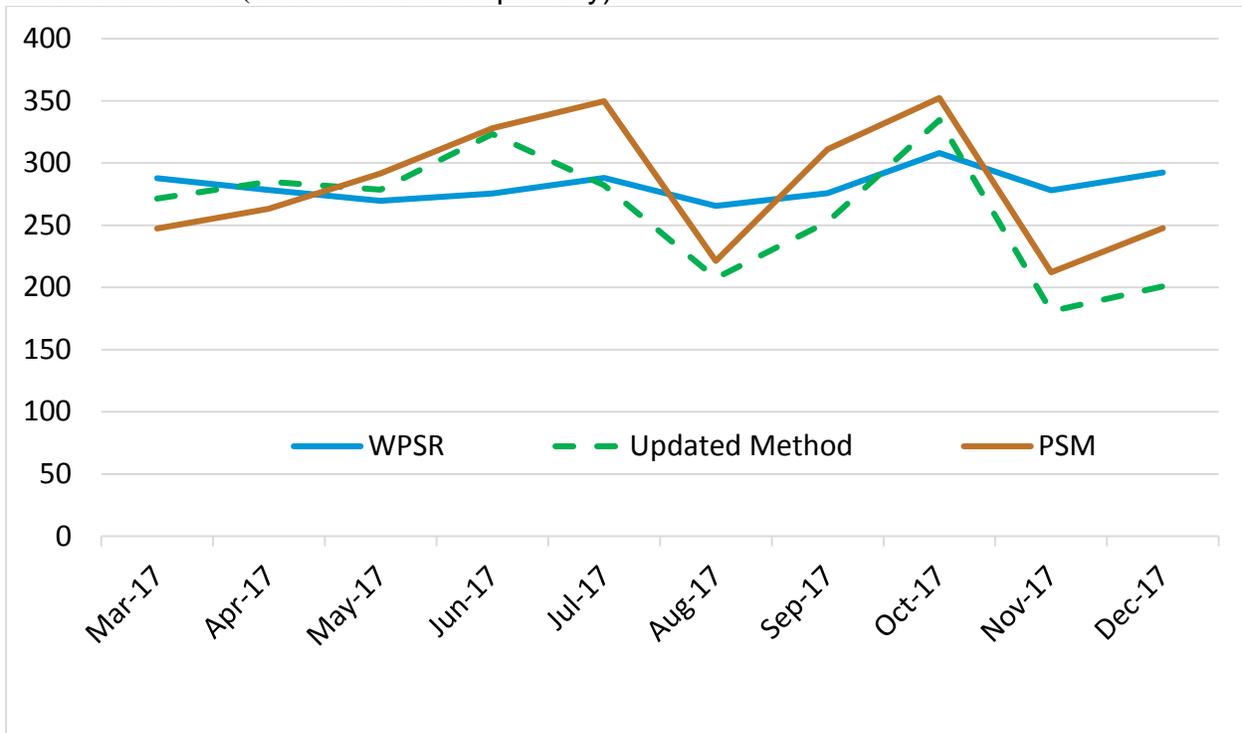
Finished motor gasoline (thousand barrels per day)



Source: U.S. Energy Information Administration

Figure 4. Monthly residual fuel oil exports for March 2017–December 2017

Residual fuel oil (thousand barrels per day)



Source: U.S. Energy Information Administration

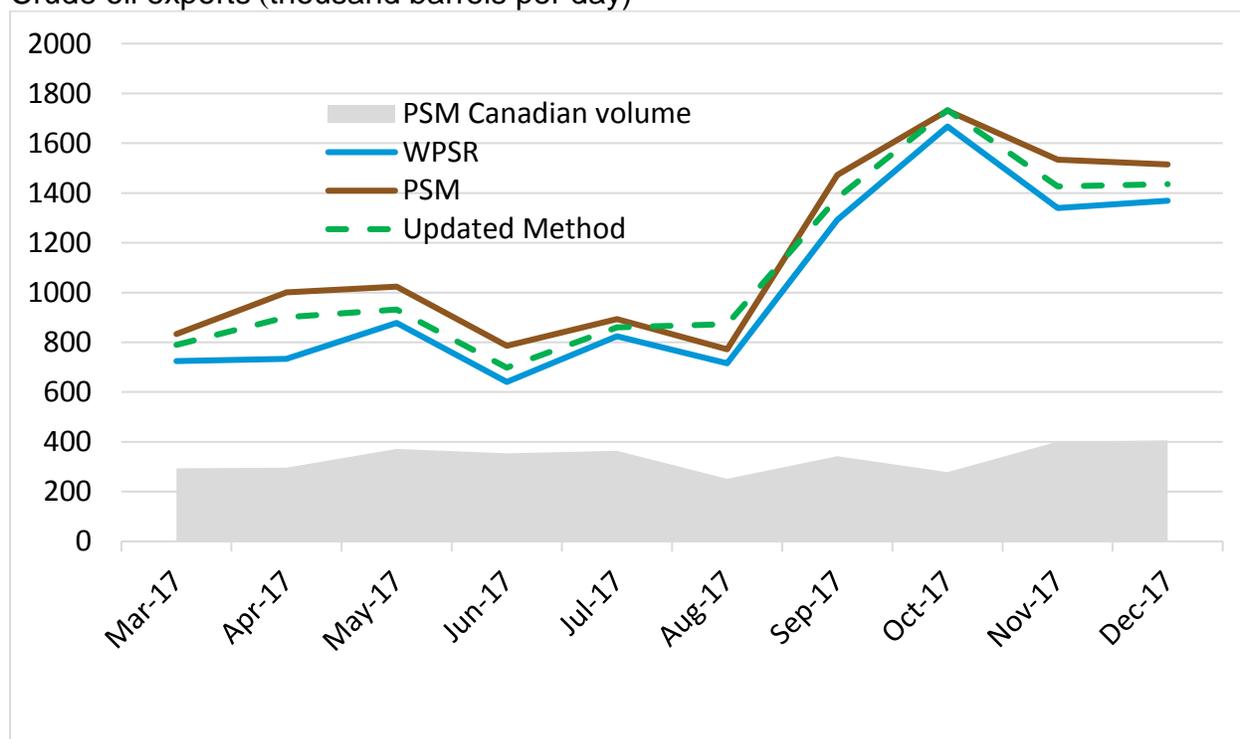
Estimates based on post-editing statistical models

Figures 5–7 describe results of the updated regression model for crude oil, kerosene-type jet fuel, and other oils. These figures show monthly estimates of crude oil, kerosene type jet fuel, and other oils produced by the regression models against USCB/PSM shown by the brown line. The published estimates used currently in the WPSR are shown by the blue line, and the dotted green line shows what the estimates would have been using the updated method. Beginning March 2018, the estimation methodology for crude oil and other oils is based on the updated two-component model instead of the previous regression model. For crude oil, the previous regression model was used since June 2017. For other oils, the previous regression model was used since March 2017. Beginning March 2018, the estimation methodology for kerosene-type jet fuel is based on the updated two-component model instead of aggregated edited CBP data, which was used since August 2016.

In Figure 5, the shaded gray area shows the portion of the PSM-published crude oil exports to Canada, which typically range between 200,000 b/d and 400,000 b/d. In Figure 6, the shaded gray area shows the portion of the PSM kerosene-type jet fuel exports that consists of exports to Canada, which typically range from 20,000 b/d to 50,000 b/d. In Figure 7, the shaded gray area shows the portion of the PSM other oils exports that consists of other oils exports to Canada, which typically is about 400,000 b/d.

Figure 5. Monthly crude oil exports for March 2017–December 2017

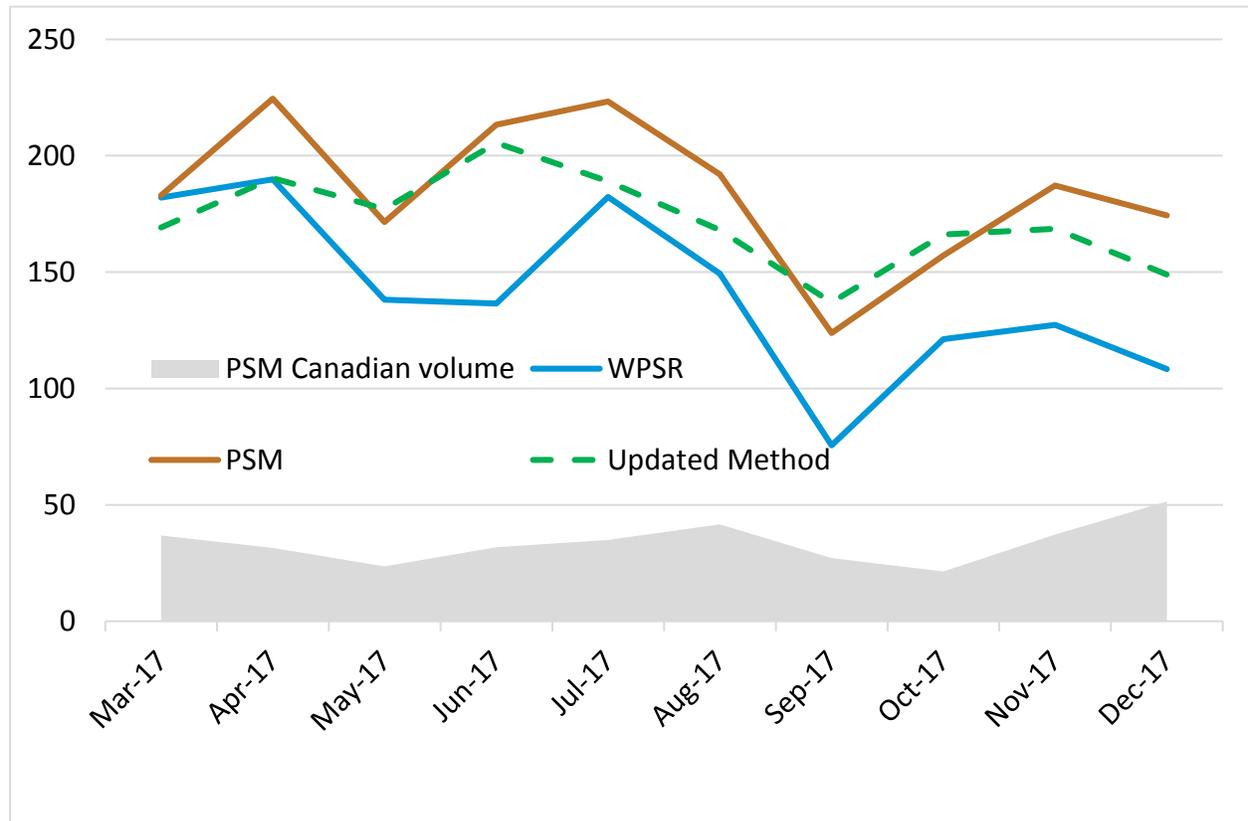
Crude oil exports (thousand barrels per day)



Source: U.S. Energy Information Administration

Figure 6. Monthly kerosene-type jet fuel exports for March 2017–December 2017

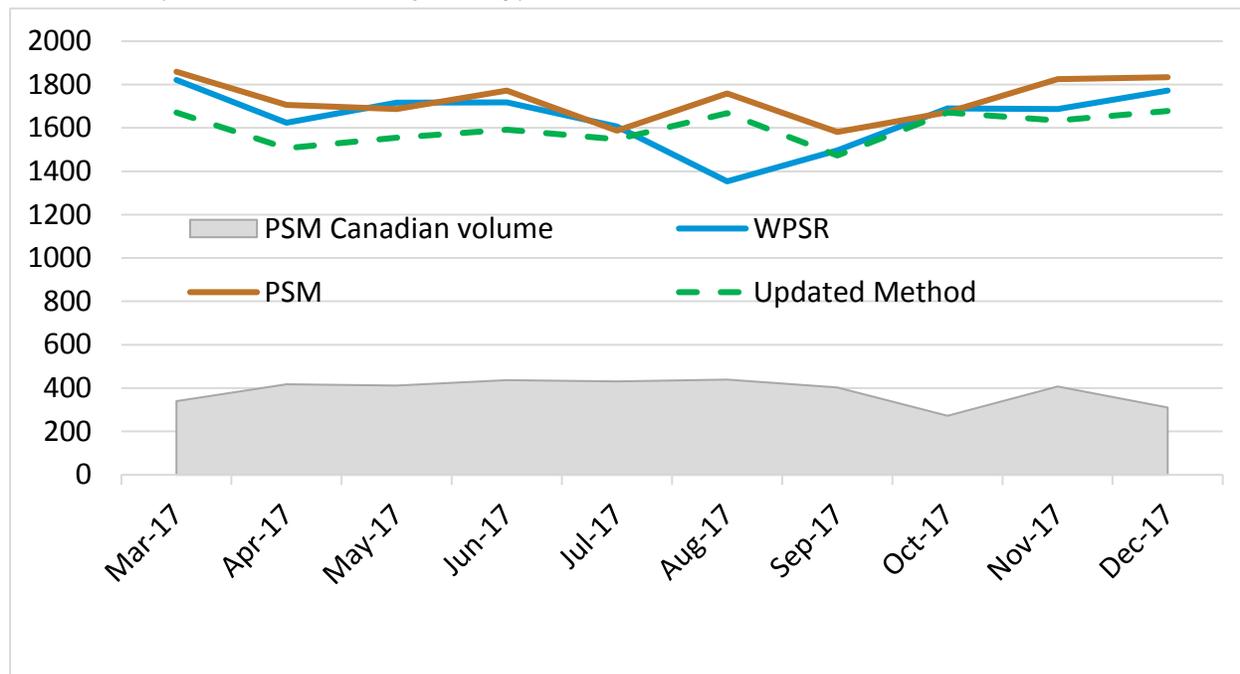
Kerosene-type jet fuel (thousand barrels per day)



Source: U.S. Energy Information Administration

Figure 7. Monthly other oils exports for March 2017–December 2017

Other oils (thousand barrels per day)

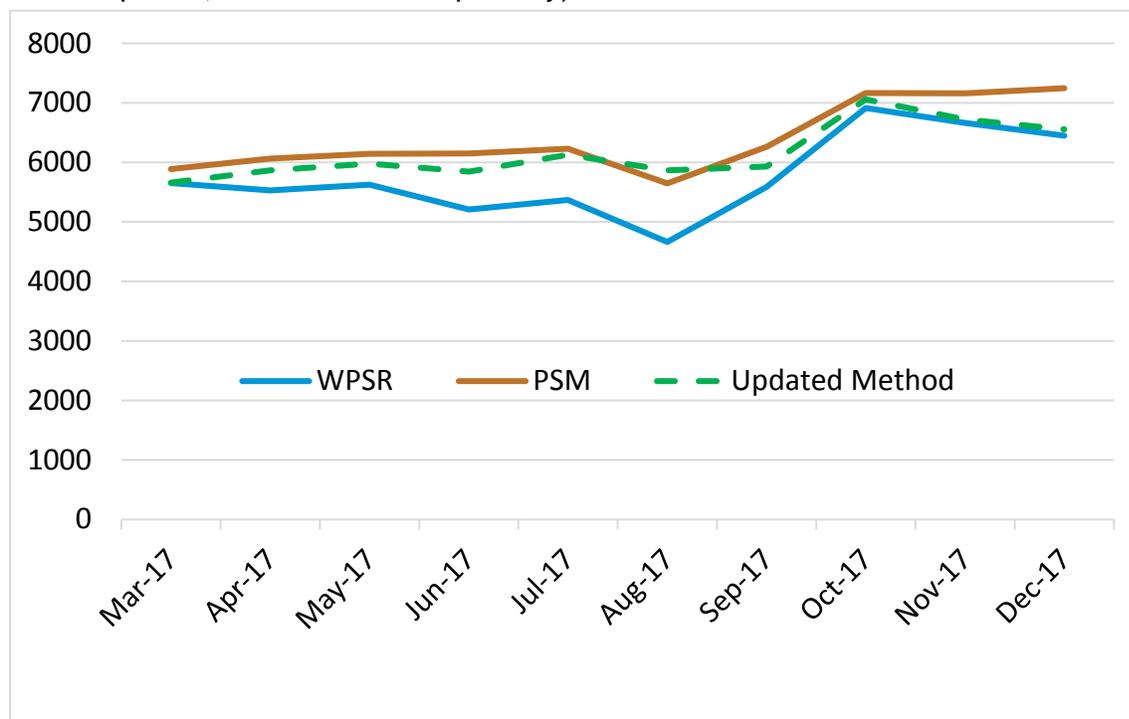


Source: U.S. Energy Information Administration

For March 2017–December 2017, Figure 8 shows total monthly estimates of exports used for the WPSR compared with USCB/PSM volumes and the estimates using the updated methodology. Though the differences between the WPSR and PSM estimates are typically within 10% of the PSM export volumes in absolute value, the estimates using the updated methodology appear to be closer to the PSM estimates.

Figure 8. Monthly crude oil, petroleum products and biofuels exports for March 2017–December 2017

Total exports (thousand barrels per day)



Source: U.S. Energy Information Administration

Summary

By making use of near real-time petroleum export data from CBP, published monthly trade data from the USCB, and revised estimation methodologies, EIA’s WPSR reports estimates of U.S. weekly exports with improved accuracy. EIA will continue to monitor the performance of these methods, and as with crude oil and finished gasoline, EIA will modify them as needed to further improve their accuracy.

Attachment A: Edit Rules

Below are the calculations and edit rules in place to prepare the CBP data for estimates:

1. Identify and remove records that are outside the reporting period.
2. Identify and remove duplicate records.
3. Exclude exports originating from U.S. territories (e.g., Puerto Rico and U.S. Virgin Islands).
4. Map Schedule B/Harmonized Tariff Schedule codes to WPSR products.
5. Keep only products that are within scope of WPSR products.
6. Convert units of quantity not already in barrels (e.g., kg, liters, and tons) to barrels.
7. Calculate a dollar value per barrel using spot prices from Thomson-Reuters.
8. Based on value (in dollars) divided by quantity of commodity exported, calculate a unit price (dollars/barrel).
9. Compare the unit price with spot prices to determine outliers based on an acceptable range around spot prices.
10. Flag extreme values. Then use the following procedure:
 - a. Recalculate weight of the commodity exported based on quantity (in barrels) and a conversion factor.
 - b. Compare the recalculated weight against a proxy quantity such as shipment or commodity weight based on an acceptable range around shipment or commodity weight.
 - c. If the recalculated weight is within an acceptable range around shipment or commodity weight, then use the original reported quantity in barrels.
 - d. If the recalculated weight is not within an acceptable range around the shipment or commodity weight, then convert the shipment or commodity weight to quantity in barrels using a conversion factor and use this as recalculated quantity.

Attachment B: Updates to WPSR's Methodology for Estimating Exports of Crude Oil, Finished Motor Gasoline, Residual Fuel Oil, and Kerosene-Type Jet Fuel

Phase 1: Implemented starting with the WPSR published on March 8, 2017

- Updated the regression model, Model 1, which was used to estimate total exports from edited CBP data. The regression model used CBP data as an input and accounted for U.S. exports to Canada that are missing from CBP data. The updated model (Model 2), included an intercept term and used daily CBP data that were edited to identify and impute questionable quantities and then aggregated to monthly totals as input. The Model 1, used monthly USCB data that were interpolated to a weekly series.
- Began using Model 2 to estimate finished motor gasoline exports. The prior method had been used to report finished motor gasoline exports equal to the aggregation of CBP data after editing and imputation.

Phase 2: Implemented starting with the WPSR published on June 7, 2017

- Model 2 was used to estimate crude oil exports, replacing the use of the Unobserved Components Model (UCM).

Phase 3: Implemented starting with the WPSR published on March 14, 2018

- Continued using same methodology using edited CBP data for estimating distillate fuel oil and propane (unchanged since August 2016).
- Updated the methodology to estimate exports of crude oil, kerosene-type jet fuel, residual fuel oil, finished motor gasoline, and other oils.
 - For residual fuel oil and finished motor gasoline, EIA is reporting edited CBP data without further modeling.
 - For crude oil, kerosene-type jet fuel, and other oils, EIA uses an updated methodology (termed as Model 3) as described below:

EIA uses a two-component approach to generate more accurate estimates. The first component is a linear regression model to estimate the exports to destinations other than Canada. The formula for the updated regression component is:

$$Y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

where t denotes a month and:

- Y_t is the non-Canadian (from PSM) export of crude oil, kerosene-type jet oil, or other oils in barrels/day as published in the PSM for month t

- x_t is non-Canadian edited CBP export data in barrels/day aggregated for month t to be used as the regressor
- β_0 is the intercept
- β_1 is the regression coefficient for the regressor x_t
- ε_t is the error term

The estimated regression coefficients are then applied to the non-Canadian edited CBP export data on a weekly basis to produce the published weekly exports data. This method assumes a similar linear relationship between the non-Canadian PSM and CBP data on a weekly basis for the weeks in a given month. EIA will continue to assess the results of this approach and continues to identify alternative methods to develop weekly export volume estimates, if needed.

The second component is a three-month moving average of Canadian exports in thousand barrels per day as reported in the latest series of Petroleum Supply Monthly.

Table 1 below summarizes the estimation methodology used in WPSR for a given product or its component since EIA began incorporating data collected by CBP through the ACE system.

Table 1. Summary of Estimation Methodology in WPSR by Product or Component

Product/Component	2017												2018		
	J	F	M	A	M	J	J	A	S	O	N	D	J	F	M
Crude oil															
Finished motor Gasoline															
Kerosene type jet fuel															
Distillate fuel oil															
Residual fuel oil															
Propane															
Other															

Estimation Method	
Edited CBP	
UCM	
Model 1 Regression	
Model 2 Regression	
Model 3 Regression	