Accuracy of Petroleum Supply Data

by Tammy G. Heppner and Carol L. French

Overview

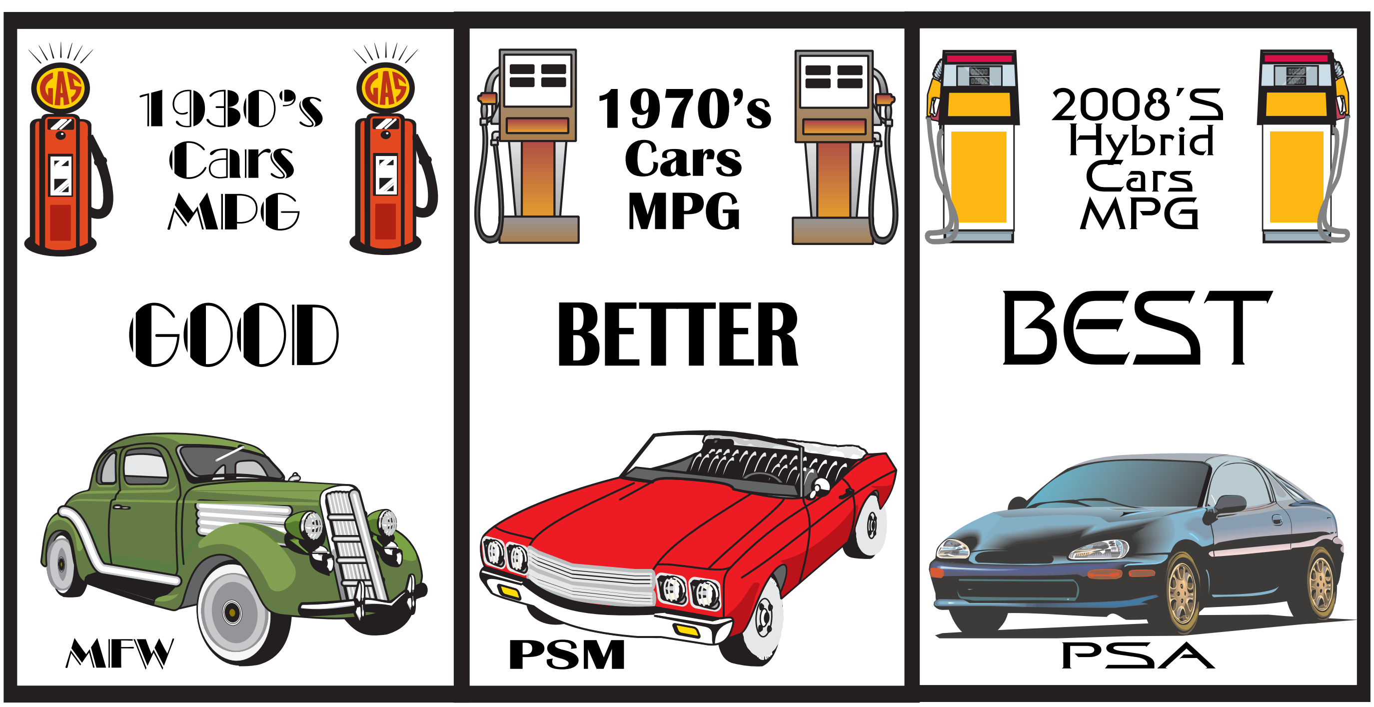
Petroleum supply data collected by the Petroleum Division (PD) in the Office of Oil and Gas (OOG) of the Energy Information Administration (EIA) showed an improvement in the accuracy of the 2006 data from initial estimates, to interim values, to final values. These data were presented in a series of PD products: the *Weekly Petroleum Status Report* (WPSR), *This Week in Petroleum* (TWIP), the *Petroleum Supply Monthly* (PSM), and the *Petroleum Supply Annual* (PSA). Weekly estimates in the *WPSR* and *TWIP* were the first values available.

Figure FE1 illustrates that just as there was an improvement in gas mileage over time, there was an improvement in petroleum supply data accuracy with increasing review time. For the monthly-from-weekly (MFW) data, respondents have the shortest reporting time, analysts have the shortest review time, and the data are least accurate. For the *PSM* data, respondents have a longer reporting time than the weekly, analysts have a longer review time, and the data are more accurate. For the *PSA* data, respondents have the longest reporting time, analysts have the longest review time, and the data are the most accurate.

For 2006, 66 petroleum supply data series were analyzed to determine how close the *PSM* values were to the final *PSA* values. For these series, 38 out of the 66 *PSM* values were within 1 percent of the *PSA* values in terms of mean absolute percent error as compared to 40 out of 66 in 2005. Sixty-two petroleum supply data series were analyzed to see how close the MFW estimates were to the final *PSA* values. For these 62 series, 27 MFW estimates were within 2 percent of the *PSA* values in terms of mean absolute percent error and, of those, 10 were within 1 percent, compared to 22 and 5, respectively, for 2005.

Two major factors that contribute to the *PSM* values being more accurate than the MFW estimates are: (1) the greater length of time between the close of the reference period and the publication date of the *PSM*; and, (2) most MFW values (weekly data converted to a monthly value) are based on company's operational records whereas *PSM* values are generally extracted from company's accounting systems, the latter being more accurate. The greater length of time allows more in-depth review of the data by the respondents and EIA. Within 2 months of the close of a reference month, interim values are published in the *PSM*. The weekly data are more quickly available. The *WPSR* and *TWIP* are available 5 days after the close of the reference week (excluding holiday

Figure FE1. Accuracy of Petroleum Data Improves Over Time



weeks). About 6 months after the end of the reference year, final monthly values, reflecting resubmissions, are published in the *PSA*.

Historically, the weekly publication (*WPSR*) and the monthly publication (*PSM)* provided volumes of crude oil and petroleum products data at relatively increasing levels of accuracy. This article provides petroleum analysts with a measure of the degree to which, on average, estimates and interim values vary from their final values.

The Petroleum Supply Reporting System

The 16 surveys in the Petroleum Supply Reporting System (PSRS) track the supply and disposition of crude oil, petroleum products, and natural gas liquids in the United States. To maintain a database with historically accurate observations and current estimates from the petroleum industry, EIA administers three survey series: weekly, monthly, and annual.

The PSRS is organized into two data collection subsystems, the Weekly Petroleum Supply Reporting System (WPSRS) and the Monthly Petroleum Supply Reporting System (MPSRS). The WPSRS processes data from the six weekly surveys. The MPSRS includes nine monthly surveys and one annual survey.

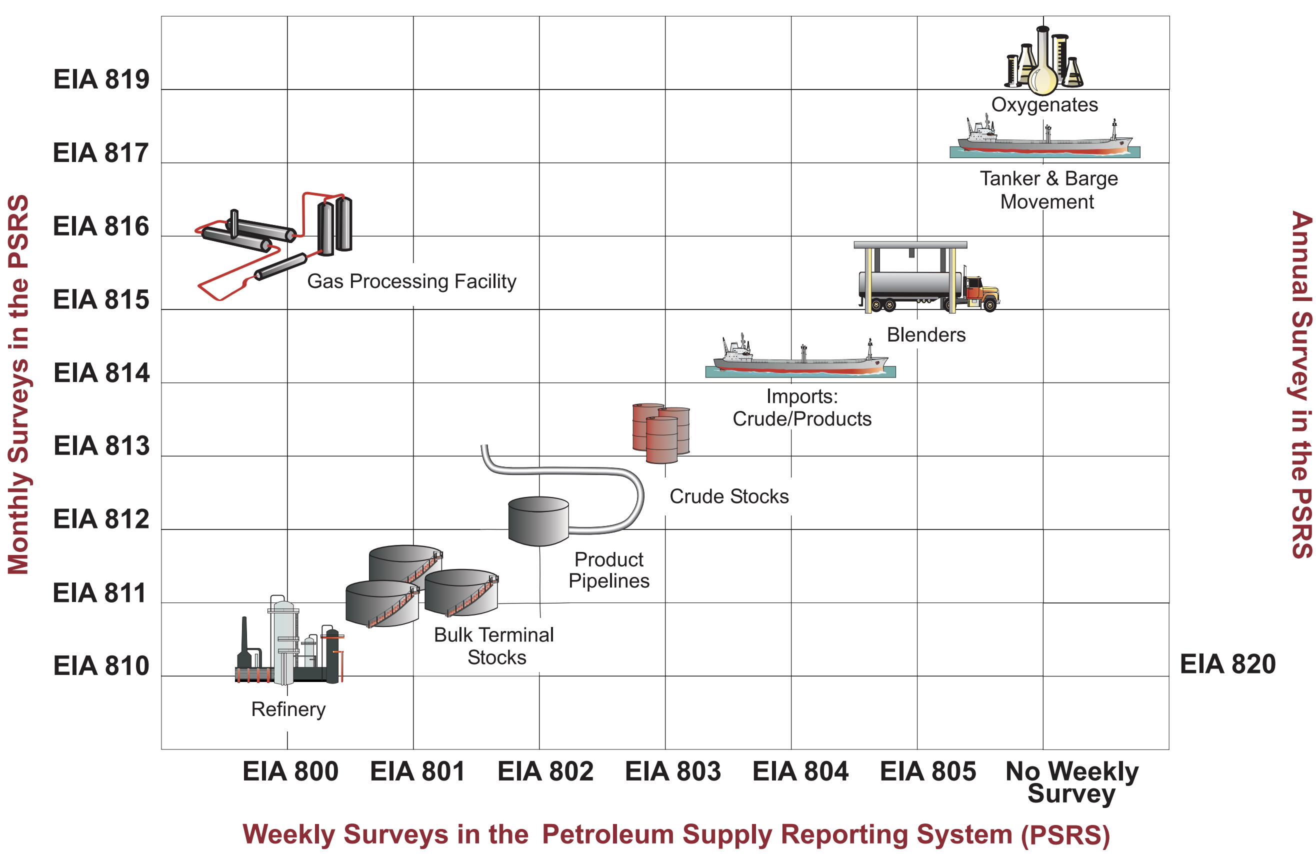
Figure FE2 displays the petroleum supply and distribution system and indicates the points at which petroleum supply data are collected. Both weekly and monthly surveys are administered at six key points along the petroleum production and supply path: (1) refineries, (2) bulk terminals, (3) product pipelines, (4) crude oil stock holders, (5) importers, and (6) blenders.

Annual U.S. refinery capacity data are collected on the Form EIA-820, "Annual Refinery Report." Beginning in 2006, these data are published in the *Refinery Capacity Report* as a separate product from the *PSA*.

The Weekly Petroleum Supply Reporting System

The WPSRS contains the data collected from the six weekly surveys. Each weekly survey is distributed to a sample of the corresponding monthly survey's universe. In Figure FE2, the icons represent the target population of the monthly and weekly surveys of the PSRS. For example, the target population for the survey Forms EIA-801 and EIA-811 is bulk terminals. Thus, the respondents to the Form EIA-801 are a sample of the respondents who report on Form EIA-811. For the weekly surveys, EIA aims for a minimum 90-percent multi-attribute-cutoff sample from the respondents to the corresponding monthly survey. In choosing the sample for

Figure FE2. Petroleum Supply Reporting System: Surveys and Subsystems



Source: Energy Information Administration, Petroleum Supply Reporting System.

each product, companies are ranked in descending order by volume. Respondents are chosen in order, down the list until the sample includes those companies contributing at least 90 percent of a variable's total volume. For example, for distillate fuel oil stocks, the weekly sample includes those respondents whose combined volumes of stocks for distillate fuel oil from refineries, bulk terminals, and pipelines constitute at least 90 percent of the total volume of distillate fuel oil stocks as reported in the corresponding monthly surveys.

These surveys enable EIA to provide timely, relatively accurate snapshots of the U.S. petroleum industry every week. The weekly surveys collect information on the supply and disposition of selected petroleum products and crude oil. The reference period for each weekly survey begins at 7:01 a.m. each Friday and ends at 7:00 a.m. the following Friday. Respondents report their data via telephone, fax, electronic spreadsheets (through email or secure file transfer), or EIA's electronic data collection software package, the Personal Computer Electronic Data Reporting Option (PEDRO). All respondents must submit their data by 5:00 p.m. on the Monday following the end of the reference period. During the next 2 working days, quality control procedures are executed. Cell values determined to be unusual or inconsistent with other cell values are flagged. The validity of the value of each flagged cell is investigated. Some flagged values are verified by the respondent to be correct; other flagged cells are corrected; and the remaining flagged values are referred to as unresolved. Nonrespondent and unresolved flagged data are imputed using an exponentially-smoothed mean of the respondents' historical data.

Within 5 days of the close of the reference week, weekly data are made available to the public on the EIA's internet web site (http://www.eia.doe.gov) through the *WPSR* and *TWIP* (This Week in Petroleum). Since 2002, *TWIP* has provided analysis, data, and charts of the latest weekly petroleum supply and price data. Except when holidays delay data processing schedules, weekly data are available via the internet at 10:30 a.m. Eastern Time on the Wednesday following the close of the reference week. *TWIP* is generally available at 1:00 p.m. on Wednesdays at twip.html. Additionally, early estimates of monthly data based on weekly submissions (monthly-from-weekly) are published in the *WPSR* on the Wednesday following the first Friday of each month.

The Monthly Petroleum Supply Reporting System

The reference period for the monthly surveys starts on the first day of the month at 12:01 a.m. and ends on the last day of the month at midnight. The deadline for filing monthly surveys is the 20th calendar day following the end of the report month. Data are reported via mail, telephone, fax, electronic spreadsheets (through email or secure file transfer), or PEDRO.

During the period of data editing, either the respondent or EIA staff may identify an error. If the respondent discovers an error, the EIA representative for a particular survey is notified and the value is corrected. If EIA's edits diagnose an unusual value, an EIA representative will determine if the value is correct or incorrect by calling the company and/or reviewing historical data.

Within 60 days of the close of the reference month, all of the interim monthly data are published in the *PSM* on the Internet at: psm.html. Additionally, preliminary company-level imports data are released electronically between the 7th and 10th of each monthly.

Throughout the year, EIA accepts data revisions of monthly data. If a revision is made after the *PSM* has been published, it is referred to as a resubmission. Generally, within 6 months of the end of the calendar year, the final monthly values for the previous year are published in the *PSA*, but may be delayed to ensure accuracy of the data. These values reflect all *PSM* resubmissions and other data corrections. The values contained in the *PSA* are EIA's most accurate measures of petroleum supply activity.

Factors Affecting Data Accuracy

Maintaining an accurate database is a major goal of EIA. The quality of the data drives the quality of all qualitative and quantitative analyses conducted using these data. Accuracy and timeliness are primary attributes of high quality data. Accuracy of survey data is measured as the closeness of the published values to the true values (i.e., those values that would be obtained if the entire target population had been surveyed and all the data had been precisely recorded).

Respondents to the monthly surveys have more time to file than the weekly respondents, enabling them to collect, review, and revise their data more carefully than the weekly respondents. Additionally, EIA has more time to edit the monthly data. Also, some weekly respondents report estimates while many monthly respondents extract actual data from accounting systems. Thus, the monthly data are typically more accurate.

Some sources of error, such as nonresponse, are not totally preventable. Other errors, such as sampling errors, are unique to a particular type of survey. One situation where sampling error occurs is if the group of sampled respondents is dissimilar to the full population. Within the PSRS, only weekly surveys are at risk of having sampling errors. However, all surveys in the PSRS are at risk for nonsampling errors, such as: (1) insufficient coverage of respondents (the survey frame does not include all members of the target population); (2)

Table FE1. Average Coverage for Weekly Surveys, 2006 and 2005
(Percent of Final Monthly Volumes Included in Monthly-from-Weekly Sample)

Product	Stocks						Production		Imports	
	Refinery		Bulk Terminal		Pipeline					
	2006	2005	2006	2005	2006	2005	2006	2005	2006	2005
Total Motor Gasoline	98	99	94	95	97	97	94	96	95	95
Jet Fuel	97	98	96	95	99	99	98	98	94	97
Distillate Fuel Oil	96	97	91	91	98	98	97	97	95	95
Residual Fuel Oil	94	95	95	93	—	—	92	94	81	77
Crude Oil	97	97	—	—	—	—	—	—	96	97

— = Not Applicable.
Source: Energy Information Administration, Petroleum Supply Reporting System.

nonresponse; (3) response error; and (4) errors due to lack of survey clarity. A detailed discussion of factors influencing data accuracy and how they are minimized in the PSRS follows.

Samples and Sampling Error

A sample is a subset of a universe identifying members of a target population. The weekly surveys are administered to samples of the monthly populations to reduce respondent burden and to expedite the turnaround of data from survey respondents to the public. As with any sample, the values obtained are different from those obtained if the full universe had been surveyed. Sampling error is the difference between a sample estimate and a population value.

There are six samples, one for each weekly petroleum supply survey, in the WPSRS. For these surveys, the sampling error is minimized by using a minimum 90-percent multi-attribute-cutoff sample from the corresponding monthly survey's frame. At the end of each month, updates are made to the samples and survey frames if a 90-percent coverage was not obtained and to account for births/deaths of companies. Coverage may be over 90 percent since companies report all of their data even though they were added to the sample based on a specific product or region or to achieve a higher level of accuracy.

For the weekly surveys, better coverage will most likely reduce sampling error. As shown in Table FE1, 2006 coverage was comparable to 2005. Of the 21 product and supply type combinations, 20 had coverage above 90 percent in 2006. For 16 of the 21 combinations, 2006 coverage decreased from 2005. Residual fuel imports had the largest percentage increase from 2005 to 2006, increasing by 3.4 percent. The largest percentage decrease from 2005 to 2006 was for jet fuel imports, decreasing by 2.9 percent. Tabulations were done before rounding of the coverage values. The 2005 total motor gasoline production percentage was revised to include production from blenders in addition to refiners.

Nonsampling Error

Unlike sampling errors, all survey data, even those from a census survey, are at risk of incurring nonsampling errors. There are two categories of nonsampling errors, random and systematic. With random error, on average, and over time, values will be overestimated by the same amount they are underestimated. Therefore, over time, random errors do not bias the data, but they will give an inaccurate portrayal at any point in time. On the other hand, systematic error is a source of bias in the data, since these patterns of errors are made repeatedly. The following is a discussion of how the four most frequently occurring types of nonsampling error are minimized within the PSRS.

Frame Updates

The list of all companies identified as members of the target population is called a frame. If members of the target population are not included in the frame, there is an undercount of the aggregate data. To diminish the chance of undercounting, the PSRS frames are continually updated. New companies are identified through continual review of petroleum industry periodicals, the Internet, newspaper articles, and correspondence from respondents.

Maintaining a Low Nonresponse

Survey respondents are required by law to report to EIA (see Explanatory Note 6 of the *PSM* for a description of action for chronic nonresponse). The 2006 response rates for the weekly surveys and their corresponding monthly surveys are enumerated in Table FE2. Compared to the 2005 response rates, the 2006 response rates for the weekly and monthly surveys were similar.

To mitigate the effect of nonresponse, imputed values are calculated for all nonreported values. Weekly imputed values are the exponentially smoothed mean of that respondent's historical values for that variable. Monthly imputed values are

Table FE2. Average Response Rates for Monthly and Weekly Surveys, 2006

Survey Site	Respondents to Monthly Surveys			Respondents to Weekly Surveys		
	Average Universe Size	Average Number of Respondents	Percent [1]	Average Weekly Sample Size	Average Number of Respondents	Percent [2]
Refinery	154	154	100.0	130	123	94.9
Bulk Terminal	222	222	100.0	88	85	95.9
Pipeline	74	74	100.0	44	43	98.0
Crude Oil Stocks	134	134	100.0	57	56	98.3
Importer	227	222	98.1	76	73	96.6
Blender	373	372	99.7	201	198	98.9

[1] The average response rates for monthly surveys are calculated by summing the individual monthly response rates and dividing by 12.
[2] The average response rates for weekly surveys are calculated by summing the individual weekly response rates and dividing by 52.
Note: Percents are calculated before rounding.
Source: Energy Information Administration, Petroleum Supply Reporting System.

the previous month's value for the particular respondent and variable. For imports, however, there is a great deal of fluctuation from one reference period to another, with respondents frequently having no imports of a particular product. As a result, the data for nonreported cells or incomplete reports are imputed based on contacts with the company and information from the U.S. Customs and Border Protection. Imputed values for monthly company-level imports are not published but are included in aggregate data.

Reducing Response Error

Improvements to the PSRS system are continuously being made to reduce response error. To satisfy customer needs and meet the particular requirements of some respondents, computerized spreadsheets that resemble the actual survey forms have been developed, and are available for respondent reporting. Another improvement has been the increased participation in the PEDRO system, which permits all weekly and monthly survey data, to be submitted to EIA electronically. A respondent entering values via PEDRO may execute edit routines prior to transmission of the survey responses. These routines include consistency and outlier (extreme value) checks of the data. Unusual or nonreported cells are flagged and, prior to transmission of the data, a representative of the company is able to review and verify or correct data in the flagged cells.

Even with sophisticated edit checks, response error (the difference between the reported value and the actual value) remains the most likely cause of data inaccuracy. The weekly surveys are more susceptible to response error since some of their values are estimates or based on operational records. Many monthly respondents abstract their monthly data from accounting systems and thus are generally more accurate.

Maintaining accurate accounting records, however, does not ensure against response error. For example, numbers can be transposed within the correct cell; an otherwise correct value may be entered in the wrong cell; a respondent may misinterpret the intent of a question; or the wrong units may be used.

Survey Clarity

The terms, layout, and definitions on all survey forms are periodically reviewed for completeness, clarity, and consistency across surveys. At regular intervals, survey intent, as well as what data are collected, are subject to industry and government review. To the extent possible, industry changes in terminology and practice are incorporated into the PSRS on an ongoing basis to ensure survey clarity.

Data Assessment

Each of the variables included in these analyses is of current and historical interest. Of the 66 variables for which both *PSM* and *PSA* values were published, only 62 of them were published weekly throughout 2006. For each variable, six measures of accuracy were calculated to compare the differences between the MFW and *PSM* values relative to the *PSA* values.

- **Error** is the difference between the estimate (MFW) or interim (*PSM*) value and the final (*PSA*) value for a given month. For inputs, production, stock change, imports, exports, and product supplied, values are expressed in units of thousands of barrels per day. For stocks, values are expressed in units of thousands of barrels.

MFW Error = MFW Volume - *PSA* Volume

PSM Error = *PSM* Volume - *PSA* Volume

- **Percent Error** is the error for a given month divided by the final value for a given month, and multiplied by 100.

$$\text{MFW Percent Error} = \frac{\text{MFW Error}}{PSA\text{ Volume}} \times 100$$

$$PSM\text{ Percent Error} = \frac{PSM\text{ Error}}{PSA\text{ Volume}} \times 100$$

- **Mean absolute error** is the weighted average over the 12 months of the year of the absolute values of the errors for each month. The mean absolute error measures the average magnitude of the revisions that took place over a year. Outliers increase the mean absolute error. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months.

- **Mean absolute percent error** is the weighted average over the 12 months of the year of the absolute values of the percent errors. It provides a measure of the average magnitude of the revisions relative to final values. The mean absolute percent error has an inverse relationship with data accuracy; i.e., the smaller the mean absolute error, the closer the interim data are to the final data; conversely, the larger the mean absolute percent error, the greater the difference in the interim value and the final value. Outliers inflate the mean absolute percent error.

- **Range** is the difference between the smallest and largest percent errors. The range shows the dispersion of the percent differences between interim and final values.

- **Median** of the percent errors is the point at which half the values are higher and half are lower. Unlike the mean, the median is not affected by an outlier. In these analyses, each distribution has 12 observations. The median is the average of the sixth and seventh ordered observations.

The average final absolute volumes and the mean absolute percent error for MFW estimates and *PSM* interim values for 2006 and 2005 are presented in Table FE3. The average final absolute volumes are presented to give the reader an idea of the magnitude of these volumes. Variables with very small volumes are prone to larger percent changes because a modest volume change is being compared to a small final volume. The mean absolute error and the size of the volumes involved must both be included in the interpretation of data accuracy.

The 2006 MFW mean absolute percent errors which were within 2 percent of their respective *PSA* values (27 of the 62 MFW series), and the 2006 *PSM* mean absolute percent errors which were within 1 percent of their *PSA* values (38 of the 66 *PSM* series), are distinguished by a single asterisk, compared to 22 and 40, respectively, for 2005. Mean absolute percent errors that were greater than 10 percent are marked by a double asterisk. There were 19 such MFW series and 6 *PSM* series, compared to 16 and 7, respectively, for 2005.

For 2006, 7 of the 12 weekly production series increased in mean absolute percent error from 2005. Eleven of the 14 production series have a single asterisk in the *PSM* column, indicating a mean absolute percent error of less than 1 percent from the *PSA*. Additionally, 9 of the 14 *PSM* production series in 2006 decreased in mean absolute percent error from 2005.

The single asterisks in Table FE3 by the stock series show that, as in prior years, the stock values for both MFW estimates and *PSM* interim values are very close to the final *PSA* values. Fuel ethanol and methyl tertiary butyl ether (MTBE) stocks are not collected weekly, but are collected on the Form EIA-819, "Monthly Oxygenate Telephone Report." Sixteen of the 17 weekly stock series and 12 of the 19 monthly stock series for 2006 decreased or stayed the same in mean absolute percent error from 2005.

Stock change is the difference between stocks at the beginning of the month and stocks at the end of the month. Since the monthly change in stock levels is small compared to the stock levels themselves, a large percent error in stock change can occur even when the percent errors in stock levels are small.

Crude oil stock change is one of the components in the calculation of unaccounted for crude oil (calculated disposition minus calculated supply of crude oil). For both the MFW and the *PSM* numbers, the volume of the unaccounted for crude oil may be increased by a combination of factors including an understatement of imports, an overstatement of exports, an understatement of crude oil production, an understatement of stock withdrawals, and an overstatement of crude oil inputs. The overstatement of crude oil inputs can be caused by injections along crude oil pipelines of natural gas liquids. When refiners receive this mixture, they process it as crude oil. As seen in Table FE3, the production, imports, and refinery inputs of crude oil have a small mean absolute percent error relative to crude oil stock change.

For petroleum products, stock change is a component in the calculation of product supplied (representing the consumption of petroleum products). Unlike the other variables, stock change values can be negative. Stock change thus has an added dimension by which to evaluate accuracy; this is the correctness of the direction of the change. Table FE4 provides a measure of accuracy of the direction of MFW and *PSM* stock change values for 2006 and 2005. The number of months that differed from the direction of the *PSA* values for the 2006 MFW total stock change decreased compared to 2005; while the 2006 MFW crude stock and refined products stock changes increased the number of months that differed from the direction of the *PSA* values from 2005. None of the 2006 *PSM* total and refined products stock change values differed in direction from the *PSA* values; whereas, one of the 2006 *PSM* crude stock change value differed in direction from the *PSA* value.

For imports, one reason for the large mean absolute percent errors in the MFW values is that shipments do not always arrive during the week in which they were expected. This has a

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 2006 and 2005

Variable	PSA Average Absolute Volumes 2006	PSA Average Absolute Volumes 2005	Monthly-from-Weekly Mean Absolute Percent Error 2006	Monthly-from-Weekly Mean Absolute Percent Error 2005	PSM Mean Absolute Percent Error 2006	PSM Mean Absolute Percent Error 2005
Crude Oil Production (thousand barrels/day)	5,102	5,178	* 1.51	1.85	1.36	1.38
Refinery Operations						
Refinery Crude Oil Inputs (thousand barrels/day)	15,242	15,220	* 0.44	0.60	* 0.05	0.17
Operating Utilization Rate (percent)	90	91	2.95	1.30	* 0.14	0.31
Production (thousand barrels/day)						
Total Production	20,209	19,954	—	—-	* 0.34	0.49
Refinery Production	17,975	17,800	* 1.57	1.42	* 0.40	0.51
Finished Motor Gasoline	8,865	8,672	* 1.15	1.46	* 0.24	0.55
Reformulated Motor Gasoline	3,014	2,865	2.61	2.06	* 0.52	0.85
Conventional Motor Gasoline	5,851	5,807	* 1.53	2.35	* 0.12	7.50
Jet Fuel	1,481	1,546	* 1.01	1.18	* 0.03	0.50
Distillate Fuel Oil	4,040	3,954	* 1.87	1.16	* 0.31	0.21
Ultra Low Sulfur Distillate Fuel Oil	1522	23	** 22.28	17.14	** 14.91	25.93
Low Sulfur Distillate Fuel Oil	1,577	2,933	3.21	1.24	2.28	0.49
High Sulfur Distillate Fuel Oil	941	1,022	3.30	3.30	* 0.43	0.74
Residual Fuel Oil	635	628	3.03	2.59	* 0.16	1.80
Other Products	5,187	5,154	—	—	1.21	0.94
Propane	1,044	1,040	2.03	2.66	* 0.26	0.24
Other Products Refinery Production	3,455	3,355	** 15.48	14.10	* 0.75	0.68
Stocks (thousand barrels)						
Total Stocks	1,733,336	1,704,002	* 0.39	1.15	* 0.10	0.22
Total Stocks, excluding SPR	1,046,006	1,013,947	* 0.64	1.89	* 0.17	0.37
Total Crude Stocks	1,022,069	1,007,757	* 0.38	0.38	* 0.10	0.18
Crude Oil Stocks, excluding SPR	334,740	317,702	* 1.13	1.15	* 0.31	0.58
SPR Stocks	687,329	690,055	* 0.02	0.05	* 0.00	0.00
Refined Products Stocks	711,267	696,245	* 0.79	2.62	* 0.32	0.42
Total Motor Gasoline Stocks	211,521	210,628	* 0.95	1.23	* 0.65	1.07
Reformulated Motor Gasoline Stocks	7,147	21,778	8.26	8.44	1.48	4.96
Conventional Motor Gasoline Stocks	114,348	114,649	* 1.56	1.95	* 0.59	0.79
Jet Fuel Stocks	40,781	40,013	* 1.54	2.07	* 0.90	0.63
Distillate Fuel Oil Stocks	135,415	123,050	2.65	3.16	* 0.29	0.53
Ultra Low Sulfur Distillate Fuel Oil Stocks	28,311	1262	9.27	15.74	3.19	1.69
Low Sulfur Distillate Fuel Oil Stocks	50,214	71,245	4.08	2.13	2.50	0.34
High Sulfur Distillate Fuel Oil Stocks	56,890	50,543	4.97	5.48	2.10	1.20
Residual Fuel Oil Stocks	41877	37,586	* 1.36	2.28	1.57	0.48
Other Products Stocks	281,672	284,968	* 1.17	4.43	* 0.12	0.32
Propane Stocks	53,138	52,483	* 1.27	1.52	* 0.11	0.78
Fuel Ethanol Stocks	8,309	6,022	—	—	* 0.71	0.68
Methyl Tertiary Butyl Ether Stocks	1,995	3,397	—	—	2.98	0.89
Stock Change (thousand barrels/day)						
Total Stocks Change	569	543	** 52.07	138.98	** 11.20	38.72
Crude Stock Change	229	297	**295.60	150.02	**127.01	74.02
Refined Products Stock Change	568	475	** 57.32	134.16	** 12.80	55.14
Imports (thousand barrels/day)						
Total Imports	13,707	13,714	* 1.55	2.18	* 0.70	1.37
Total Crude Imports	10,118	10,126	* 0.86	1.13	* 0.45	0.84
Crude Oil Imports, excluding SPR	10,110	10,074	* 0.89	1.17	* 0.48	0.70
SPR Imports	8	52	** 32.88	83.01	** 11.82	31.92
Refined Products Imports	3,589	3,588	4.48	7.31	2.00	3.44
Finished Motor Gasoline Imports	475	603	** 11.15	8.36	* 0.56	3.10
Reformulated Motor Gasoline Imports	53	239	9.02	11.38	* 0.32	2.14
Conventional Motor Gasoline Imports	422	364	** 12.60	14.00	* 0.57	4.16
Jet Fuel Imports	186	190	** 10.96	30.10	7.58	27.01

(Continued)

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 2006 and 2005 (Continued)

Variable	PSA Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		PSM Mean Absolute Percent Error	
	2006	2005	2006	2005	2006	2005
Distillate Fuel Oil Imports	365	329	** 12.26	10.81	1.61	0.59
Ultra Low Sulfur Distillate Fuel Oil Imports	112	4	** 28.84	223.97	* 0.96	4.25
Low Sulfur Distillate Fuel Oil Imports	77	151	** 30.21	17.09	** 15.01	0.67
High Sulfur Distillate Fuel Oil Imports	176	173	** 20.88	11.45	1.30	0.68
Residual Fuel Oil Imports	350	530	** 18.29	16.31	2.29	0.71
Other Products Imports	2,214	1,937	8.69	6.07	2.60	3.69
Propane Imports	228	233	** 15.06	11.79	5.79	3.38
Exports (thousand barrels/day)						
Total Exports	1,316	1,165	** 14.07	16.32	1.28	0.89
Crude Oil Exports	25	32	** 29.62	43.18	* 0.00	31.79
Refined Products Exports	1,292	1,133	** 14.15	16.03	1.30	0.02
Total Net Imports (thousand barrels/day)	12,390	12,549	* 1.50	1.71	* 0.89	1.58
Products Supplied (thousand barrels/day)						
Total Products Supplied	20,687	20,802	* 1.13	0.92	* 0.56	0.76
Finished Motor Gasoline Supplied	9,252	9,159	* 0.42	0.78	* 0.44	0.38
Jet Fuel Supplied	1,633	1,679	* 1.64	2.87	1.67	3.09
Distillate Fuel Oil Supplied	4,169	4,118	* 1.81	1.51	* 0.56	0.34
Residual Fuel Oil Supplied	689	920	** 13.30	7.99	2.57	1.83
Other Products Supplied	4,944	4,926	3.12	4.60	1.34	1.52
Propane Supplied	1,215	1,229	5.31	4.27	1.32	2.28

— = Not Applicable.
* = For MFW values, mean absolute percent error less than or equal to 2; for PSM values, mean absolute percent error less than or equal to 1.
** = Mean absolute percent error greater than or equal to 10.
SPR = Strategic Petroleum Reserve
Notes: Error is the difference between Monthly-from-Weekly estimates or interim monthly data published in the *Weekly Petroleum Status Report* or *Petroleum Supply Monthly* and the final value as published in the *Petroleum Supply Annual*. Percent error is the error multiplied by 100 and divided by the final published value. Mean absolute error is the weighted average of the absolute errors. Mean absolute percent error is the weighted average of the absolute percent errors. The number of days in the month is used for weighting all product categories except stocks. Stocks are weighted equally for each of the 12 months. Totals may not equal sum of components due to independent rounding.
Source: Energy Information Administration, Petroleum Supply Reporting System.

greater impact when the end of the month occurs in the middle of the week. Of the 16 import series, 9 MFW and 10 *PSM* series in Table FE3 showed a decrease in mean absolute percent error from 2005 to 2006.

With the exception of refinery receipts in the U.S. Territories, EIA does not collect export data. They are gathered by the U.S. Bureau of the Census on a monthly basis. They are received by EIA on a monthly basis approximately 7 weeks after the close of the reporting month. The weekly estimates for exports are projections based on past monthly data. Because the export data are highly variable, it is difficult to obtain estimates of comparable quality to domestic estimates.

Products supplied is the calculation of field production, plus refinery production, plus imports, plus unaccounted for crude

Table FE4. Number of Months In Which the Direction of Non-Final Stock Change Values Differed From PSA

Stock Change	Number of Months	
	2006	2005
Total Stock Change		
MFW and PSA Values	0	3
PSM and PSA Values	0	0
Crude Stock Change		
MFW and PSA Values	3	2
PSM and PSA Values	1	0
Refined Products Stock Change		
MFW and PSA Values	3	2
PSM and PSA Values	0	0

Source: Energy Information Administration, Petroleum Supply Reporting System.

oil, minus stock change, minus crude oil losses, minus refinery inputs, minus exports. Therefore, the accuracy of products supplied is affected by the individual components.

Box and Whisker Plots

Example 1 in the shaded box titled "Structure of Box and Whisker Plots," is a simplified illustration of the box and whisker plots that follow. The box and whisker plots map the 5-year trends in historical accuracy of weekly estimates and monthly interim values. The details provided by the box and whisker plots include: historical trends, the range of monthly percent errors, direction of the error (i.e., overestimation or underestimation), and the identification of unusual values.

Each box and whisker plot is placed on a graph, where the horizontal axis represents the year and the vertical axis represents the percent error. The center of the vertical axis for all the box and whisker plots is zero percent error. For each variable studied, a pair of charts, each containing five box and whisker plots (one for each year, from 2002 through 2006), are presented side-by-side; the chart on the left contains the percent errors for the MFW estimates, and the chart on the right contains the percent errors for the *PSM* values. To facilitate the comparison of MFW percent errors and the *PSM* percent errors, the corresponding plots have the same scale.

The position of the box along the vertical-axis denotes whether the MFW or *PSM* values are predominantly overestimates or underestimates of the *PSA* values. For example, if the majority of the MFW values were overestimates, more than half of the box would be above the zero percent error line.

The outliers, represented by an asterisk, are usually the result of resubmissions sent in throughout the year by respondents due to misreporting or reporting problems.

Crude Oil Production and Crude Oil Inputs

Crude oil production data are not collected through any of EIA's surveys. EIA's Dallas Field Office assembles data collected from State agencies responsible for measuring crude oil production. Based on historical trends and/or data reported on Form EIA-182, "Domestic Crude Oil First Purchase Report," EIA estimates weekly and monthly production. Final estimates based on revised Form EIA-182 data, State government agencies, and the U.S. Department of Interior's Minerals Management Service data are published in the *PSA*.

Figure FE3 presents errors of MFW and *PSM* values relative to *PSA* values for crude oil production and crude oil inputs. In contrast to 2005, more of the 2006 MFW estimates and *PSM* interim values for crude oil production overestimated the final *PSA* values. The 2006 median (1.00) of the *PSM* percent errors was the largest median over the 5-year period.

About one half of the 2006 MFW estimates for refinery crude oil inputs overestimated the final *PSA* values. The range (1.96) of the 2006 MFW percent errors was the smallest range of all other MFW plots analyzed for 2006, ranging from -0.86 to 1.10 percent. As in prior years, the 2006 *PSM* refinery crude oil inputs were extremely close to the final *PSA* values, with percent errors within 0.22 percent. The 2006 range (0.31) of *PSM* percent errors was the smallest range of all other *PSM* plots analyzed for 2006, ranging from -0.22 to 0.09 percent. There were two outliers in July (-0.22) and December (0.09).

Product Production

PSM interim values for production of each of the four major petroleum products were superior to their comparable MFW estimates. Figures FE4 and FE5 contain the box and whisker plots for motor gasoline and distillate fuel oil production, and residual fuel oil and jet fuel production, respectively.

The range (4.51) of the 2006 MFW motor gasoline production percent errors, displayed in Figure FE4, was similar to prior years, ranging from -2.32 to 2.19 percent. Most of the 2006 *PSM* interim values for motor gasoline production underestimated the final *PSA* values. The 2006 range (0.72) of the *PSM* percent errors was the smallest range over the past 5 years, ranging from -0.50 to 0.22 percent.

The 2006 range (6.49) of the MFW percent errors for distillate fuel oil production was the largest range over the past 5 years, ranging from -3.68 to 2.81 percent. Similarly, the range (2.83) of the 2006 *PSM* percent errors for distillate fuel production was the largest range over the 5-year period, ranging from -0.18 to 2.65 percent. The outlier in March 2006 (2.65) was the largest percent error over the 60 months studied.

The box and whisker plots for residual fuel oil production and jet fuel production are shown in Figure FE5. Most of the 2006 MFW estimates for residual fuel oil production underestimated the final *PSA* values. The range (8.10) of the 2006 MFW residual fuel oil production percent errors was the smallest range over the 5-year period, ranging from -6.72 to 1.38 percent. There were few revisions to the 2006 *PSM* residual fuel oil interim values. There was an outlier in January 2006 (-1.64).

For jet fuel production, the 2006 range (4.52) of MFW percent errors, ranging from -1.41 to 3.11 percent, was the largest range over the past 5 years. April 2006 (3.11) was the largest percent error the 60 months studied. The only two revisions in April (-0.07) and August (-0.34) for the 2006 *PSM* percent errors for jet fuel production were outliers.

Structure of Box and Whisker Plots

All box and whisker plots discussed in this article are the visual presentation of a variable's distribution of 12 values of percent errors for either MFW or *PSM* values relative to *PSA* values for a given year. In general, box and whisker plots group data, ordered from smallest to largest, into four areas of equal frequency, quartiles, and show the range and dispersion of data within the quartiles. Sometimes the values of quartiles must be interpolated, i.e., if there are two values that meet the criteria of a quartile, then the average of the two must be taken. Presented below is a discussion of components of box and whisker plots and how they apply to the 12-value distribution illustrated in Example 1: -35, -20, -11, -9, 0, 0, 0, 0, 4.5, 5.5, 15, and 20.

- **First Quartile**

Twenty-five percent of the values are equal to or below the first quartile. In Example 1, the first quartile is the average of the third and fourth ordered observations, i.e., (-11+(-9))/2=-10. The first quartile demarcates the lower boundary of the box.

- **Second Quartile**

The second quartile is the median, and it intersects the box. Fifty percent of the observations are equal to or below the median; in our example, the values of these six observations are: 0, 0, -9, -11, -20, and -35. Also, for this example, the median is the average of the sixth and seventh value, 0, i.e., (0+0)/2. The plot provides the value of the median (the second quartile) as well as information on how the median compares in magnitude to the rest of the observations. Outliers distort the magnitude of the mean, whereas a median is not distorted since it is the actual value that falls in the middle of the distribution. Since outliers have occurred in the distributions of values of PSRS variables, a median is preferred to a mean when assessing accuracy.

- **Third Quartile**

Seventy-five percent of the observations (9 in this case) have values equal to or below the third quartile. In Example 1, the third quartile is 5, i.e., (4.5+5.5)/2. The third quartile demarcates the upper boundary of the box.

- **Box**

The box contains half of all the values. In Example 1, as well as in each box found in Figures FE3-FE11, a minimum of six values are contained within the box. The interquartile range is the length of the box, the difference between the third and first quartiles. The interquartile range for Example 1 is 15, i.e., 5-(-10).

- **Whiskers**

Each whisker extends out from the box, one from the first quartile and the other from the third quartile, to the most extreme value that still falls within 1.5 times the interquartile range. In Example 1, a whisker extends from the third quartile, 5, to 20, which is the maximum value and is within 1.5 interquartile ranges of 5 (as it is less than 5+(1.5*15)=27.5). Also in Example 1, the lower whisker extends from the first quartile -10, to -20, which is the lowest value of the distribution within 1.5 interquartile ranges of the first quartile.

- **Fourth Quartile**

The fourth quartile is the maximum value of the distribution. In Example 1, the fourth quartile, 20, also demarcates the upper value of the top whisker as it is within 1.5 interquartile ranges of the third quartile.

- **Outlier**

An outlier, identified as an asterisk, is an observation that is more than 1.5 interquartile ranges greater than the third quartile, or more than 1.5 interquartile ranges less than the first quartile. In Example 1, there is one outlier, -35. It is less than the lower whisker's threshold value, which is -32.5 (-10-(1.5*15)). The importance of the occurrence of an outlier depends on the distribution of the variable. If the interquartile range is very tight and the outlier is in close proximity, then there is little concern about the occurrence of that outlier. (See Figure FE3, *PSM* vs *PSA* of Refinery Crude Oil Inputs for 2006.)

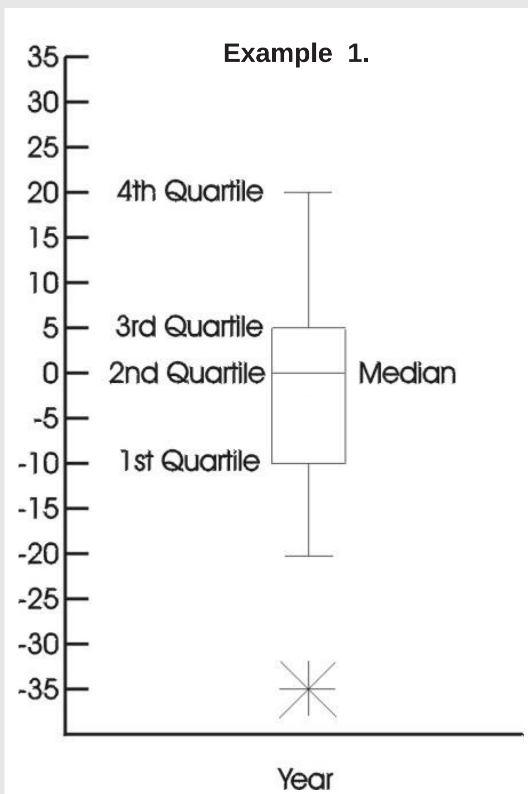
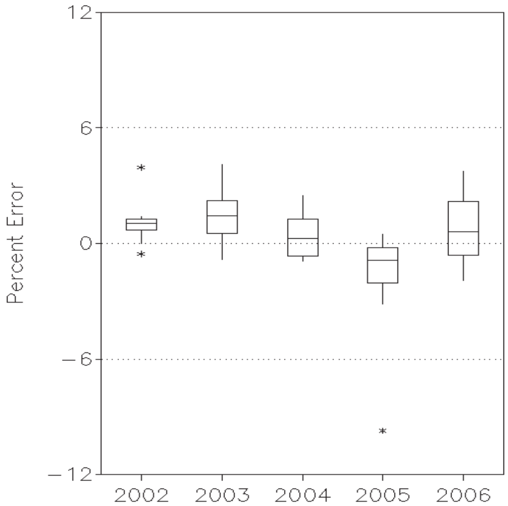


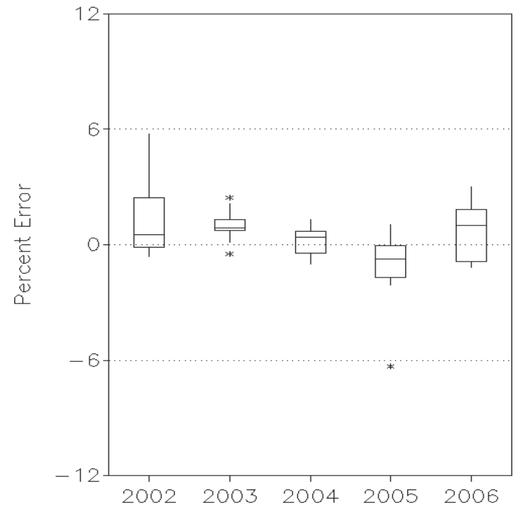
Figure FE3. Box Plots of Percent Errors for MFW and *PSM* Crude Oil Production and Refinery Crude Oil Inputs Data, 2002 - 2006

Crude Oil Production

MFW vs. *PSA*

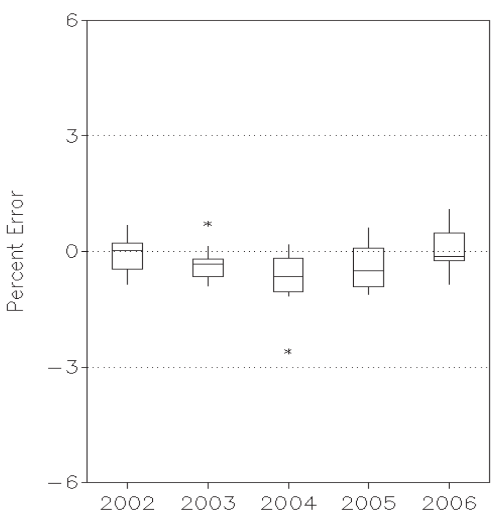


PSM* vs. *PSA

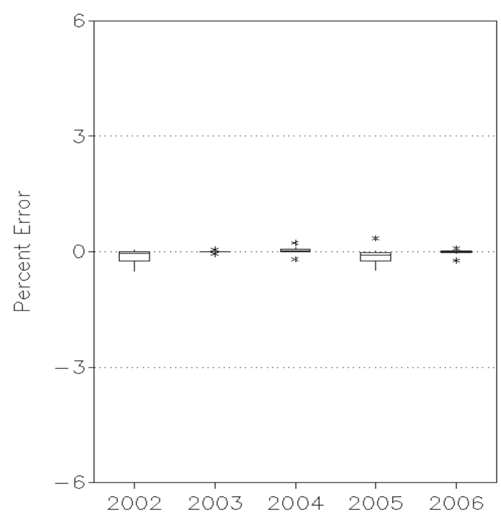


Refinery Crude Oil Inputs

MFW vs. *PSA*



PSM* vs. *PSA

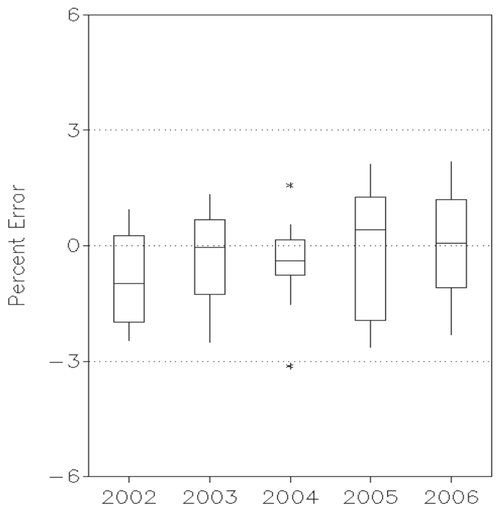


Source: Energy Information Administration, Petroleum Supply Reporting System.

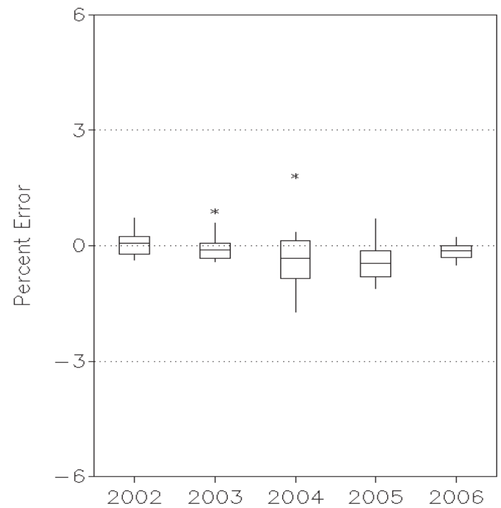
Figure FE4. Box Plots of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Production

Motor Gasoline Production

MFW vs. *PSA*

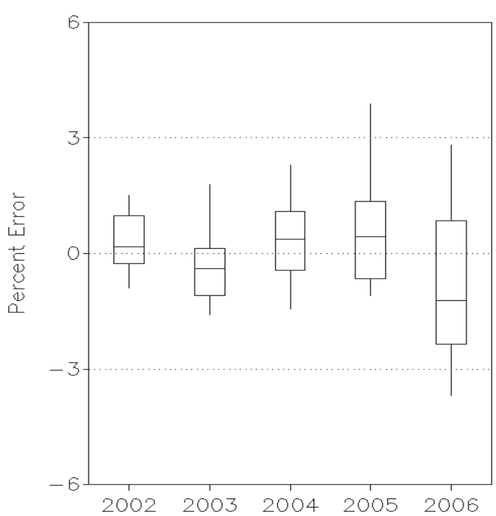


PSM* vs. *PSA

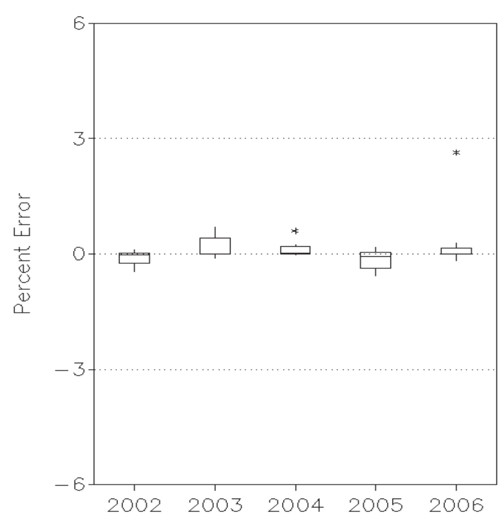


Distillate Fuel Oil Production

MFW vs. *PSA*



PSM* vs. *PSA

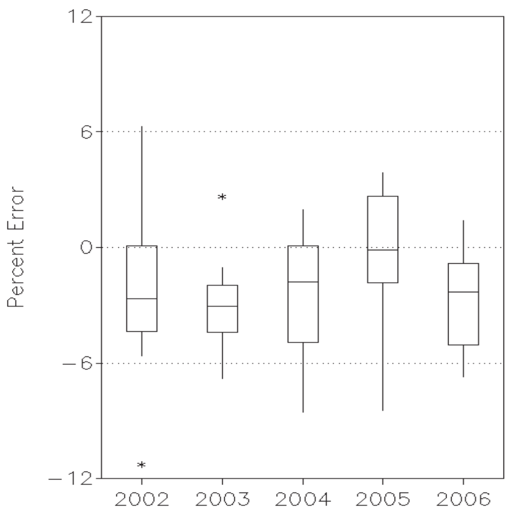


Source: Energy Information Administration, Petroleum Supply Reporting System.

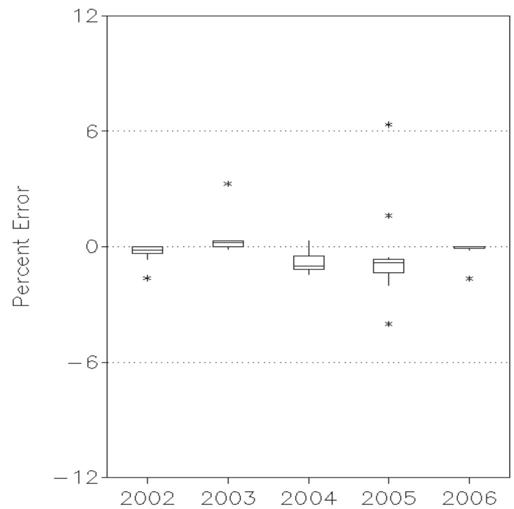
Figure FE5. Box Plots of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Production Data, 2002 - 2006

Residual Fuel Oil Production

MFW vs. *PSA*

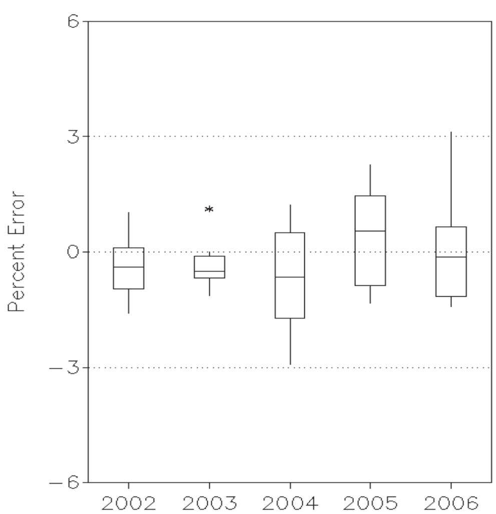


PSM* vs. *PSA

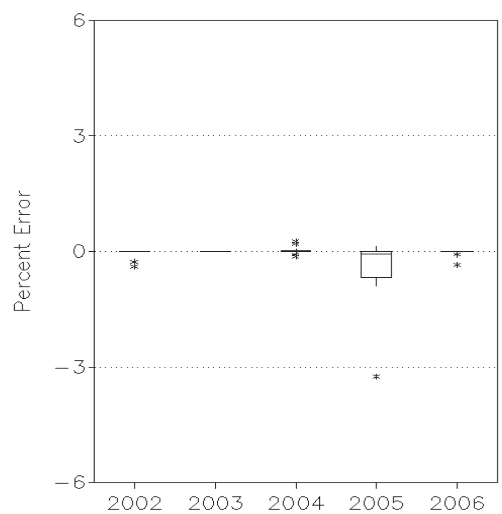


Jet Fuel Production

MFW vs. *PSA*



PSM* vs. *PSA



Source: Energy Information Administration, Petroleum Supply Reporting System.

Stocks

Figures FE6, FE7, and FE8 show the yearly distribution of percent errors for stocks of crude oil, motor gasoline, distillate fuel oil, residual fuel oil, jet fuel, and propane. Figure FE6 shows the box and whisker plots for crude oil stocks and motor gasoline stocks. One half of the 2006 MFW estimates for crude oil stocks overestimated the final *PSA* values. The 2006 median (-0.06) of MFW percent errors was the closest to zero over the 5-year period. Most of the 2006 *PSM* interim values for crude oil stocks underestimated the final *PSA* values.

One half of the 2006 MFW estimates for motor gasoline stocks underestimated the final *PSA* values. The 2006 median (-0.03) of MFW percent errors was the closest to zero over the past 5 years. In contrast to prior years, all of the *PSM* interim values for motor gasoline stocks overestimated the final *PSA* values. The 2006 *PSM* percent errors ranged from 0.05 to 1.60 percent.

Figure FE7 shows box and whisker plots for distillate and residual fuel oil stocks. Most of the 2006 MFW estimates for distillate fuel oil stocks underestimated the final *PSA* values. September 2006 (1.15) was an outlier. Most of the 2006 *PSM* interim values for distillate fuel oil stocks also underestimated the final *PSA* values. There was an outlier in July 2006 (0.92).

Residual fuel oil stocks typically have larger percent errors than other stock series. One half of the 2006 MFW estimates underestimated the final *PSA* values. The range (5.98) of the 2006 MFW percent errors was the smallest range for the 5 years analyzed, ranging from -2.86 to 3.12 percent. Unlike prior years, all of the 2006 *PSM* interim values for residual fuel oil stocks overestimated the final *PSA* values. The 2006 range (6.22) of the *PSM* percent errors was the largest range over the 5-year period, ranging from 0.03 to 6.25 percent. The outlier in February (6.25) was the largest percent error over the 60 months studied.

The box and whisker plots for jet fuel stocks and propane stocks are shown in Figure FE8. Similar to 2005, most of the 2006 MFW estimates for jet fuel stocks overestimated the final *PSA* values. The 2006 median (1.53) of MFW percent errors was the largest median over the 5-year period. The range (6.52) of the 2006 *PSM* percent errors for jet fuel stocks was the largest range over the 5 years studied, ranging from -0.91 to 5.61 percent. An outlier in August (5.61) was the largest percent error over the past 60 months. There was another outlier in January (2.13).

Most of the 2006 MFW estimates for propane stocks underestimated the final *PSA* values. Over the 5-year period, the 2006 range (4.03) of percent errors was the smallest range, ranging from -2.71 to 1.32 percent. Similarly, the range (0.63) of the 2006 *PSM* percent errors for propane stocks was the smallest range over the past 5 years, ranging from -0.32 to 0.31 percent. There were outliers in March (-0.32), April (-0.29), and May (0.31).

Imports

Figures FE9, FE10, and FE11 show the yearly distributions of percent errors for the imports of crude oil and four products: motor gasoline, distillate fuel oil, residual fuel oil, and jet fuel. Because of the irregularity of imports for crude oil and petroleum products, the magnitude and range of percent errors for both the MFW and the *PSM* imports numbers can be expected to be much larger and wider than for production and stocks.

Figure FE9 shows the box and whisker plots for crude oil imports. The 2006 range (2.76) of the MFW percent errors was the smallest range over the 5-year period, ranging from -1.48 to 1.28 percent. Most of the 2006 *PSM* interim values for crude oil imports underestimated the final *PSA* values. There was one outlier in March 2006 (1.13).

The distributions of percent errors of the MFW estimates and *PSM* interim values for 2002 through 2006 of motor gasoline and distillate fuel oil imports are shown in Figure FE10. The range (49.84) of the 2006 MFW percent errors for motor gasoline imports was the largest range for the 5-year period. The outlier in August 2006 (-25.36) was the largest absolute percent error over the past 60 months. The 2006 range (2.76) of the *PSM* percent errors was the smallest range over the 5-year period, ranging from -0.25 to 2.51 percent. The outliers were in July (2.51) and October (2.47).

The 2006 range (45.67) of the MFW percent errors for distillate fuel oil imports was the largest range over the past 5 years, ranging from -16.94 to 28.73 percent. All but one of the 2006 *PSM* interim values for distillate fuel oil imports underestimated the final *PSA* values.

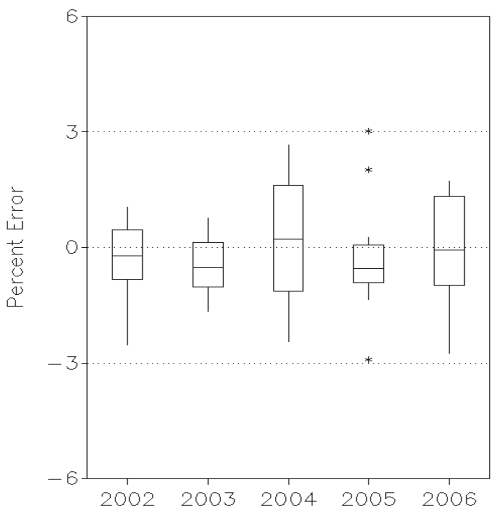
Figure FE11 shows the box and whisker plots for residual fuel oil imports and jet fuel imports. Most of the 2006 MFW estimates for residual fuel oil imports overestimated the final *PSA* values. The 2006 MFW range of percent errors, ranging from -25.52 to 45.13 percent, was the largest range (70.65) of all other MFW plots analyzed for 2006. Most of the 2006 *PSM* interim values for residual fuel oil imports underestimated the final *PSA* values. The 2006 *PSM* percent errors ranged from -5.57 to 2.49 percent.

Unlike 2005, one half of the 2006 MFW estimates for jet fuel imports overestimated the final *PSA* values. The 2006 MFW median (-0.78) was the closest to zero for the 5-year period. There was an outlier in October 2006 (40.94). The 2006 range (61.95) of the *PSM* percent errors was the largest range over the 5-year period and of all other *PSM* plots analyzed for 2006, ranging from -56.10 to 5.85 percent. The outlier in February 2006 (-56.10) had the largest absolute percent error over the past 60 months. Additional outliers were in January (-26.11) and October (5.85).

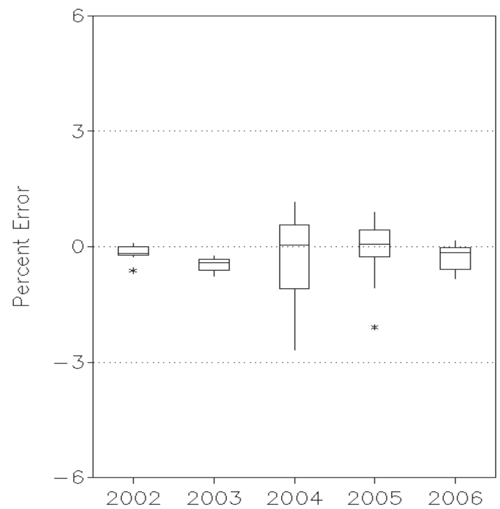
Figure FE6. Box Plots of Percent Errors for MFW and *PSM* Crude Oil Stocks Excluding Strategic Petroleum

Crude Oil Stocks Excluding SPR

MFW vs. *PSA*

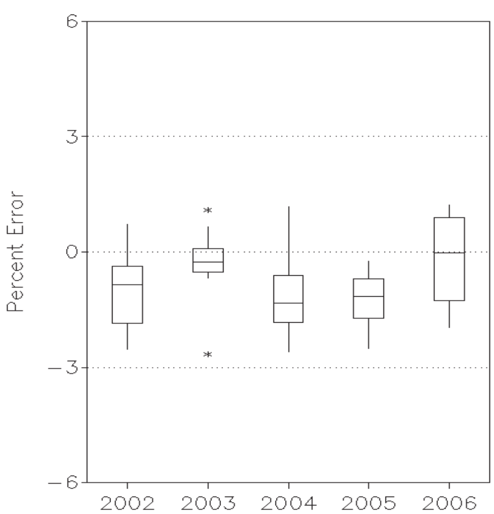


PSM vs. *PSA*

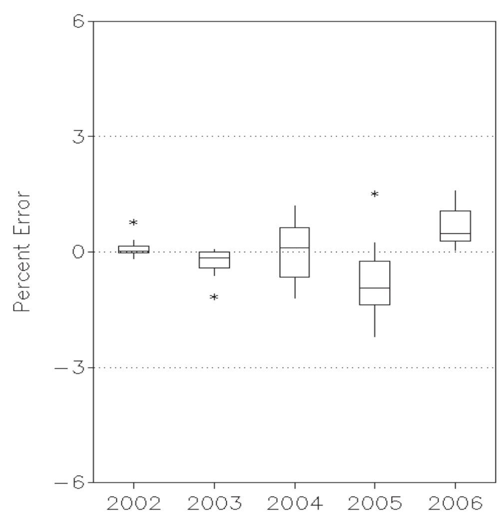


Motor Gasoline Stocks

MFW vs. *PSA*



PSM* vs. *PSA

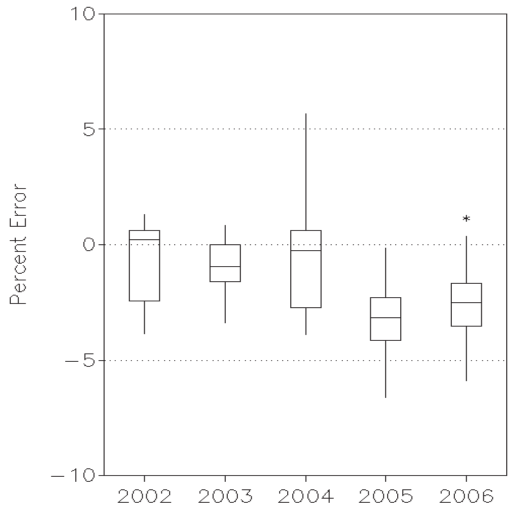


Source: Energy Information Administration, Petroleum Supply Reporting System.

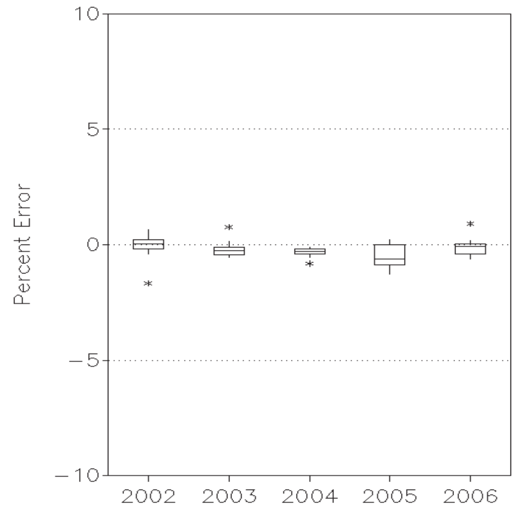
Figure FE7. Box Plots of Percent Errors for MFW and *PSM* Distillate Fuel Oil and Residual Fuel Oil Stocks Data, 2002 - 2006

Distillate Fuel Oil Stocks

MFW vs. *PSA*

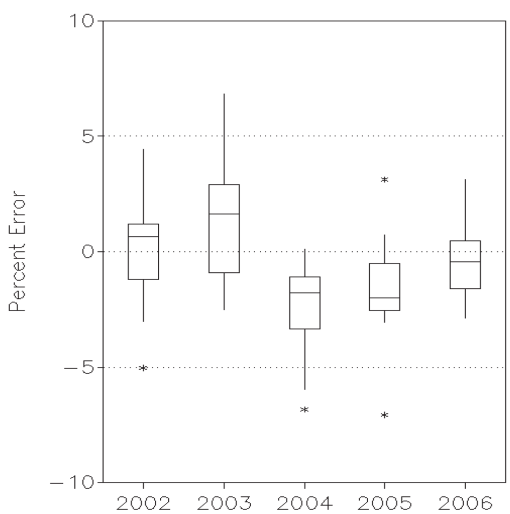


PSM vs. *PSA*

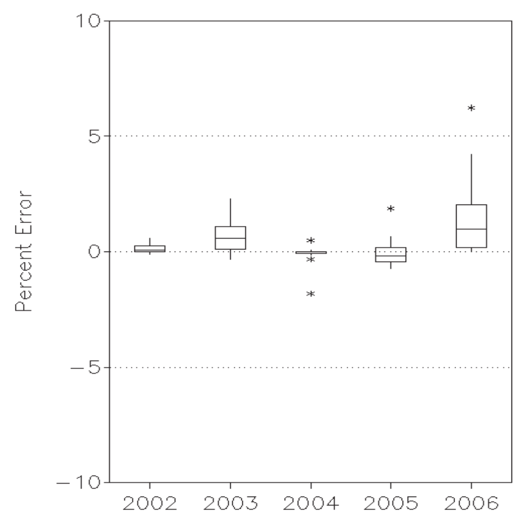


Residual Fuel Oil Stocks

MFW vs. *PSA*



PSM vs. *PSA*

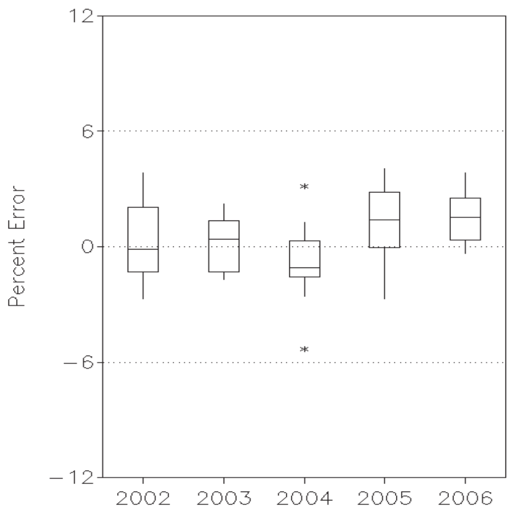


Source: Energy Information Administration, Petroleum Supply Reporting System.

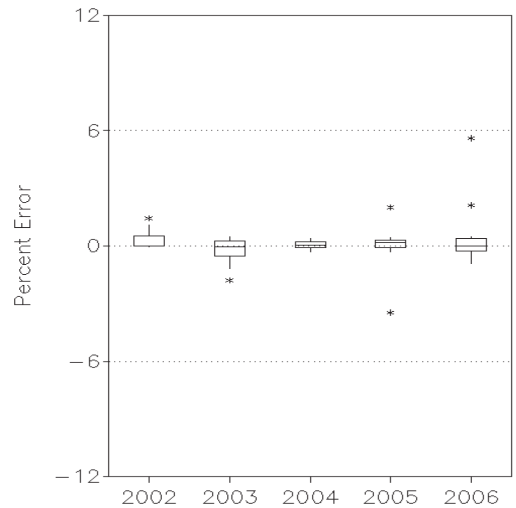
Figure FE8. Box Plots of Percent Errors for MFW and *PSM* Jet Fuel Stocks and Propane Stocks Data, 2002 - 2006

Jet Fuel Stocks

MFW vs. *PSA*

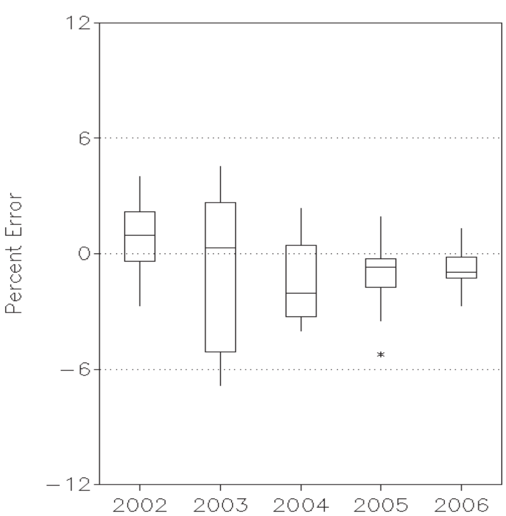


PSM* vs. *PSA

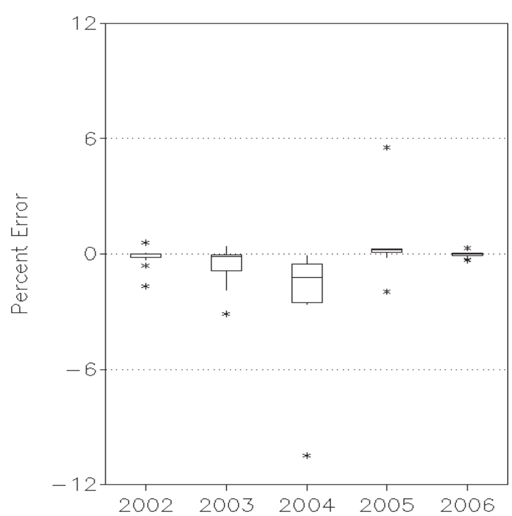


Propane Stocks

MFW vs. *PSA*

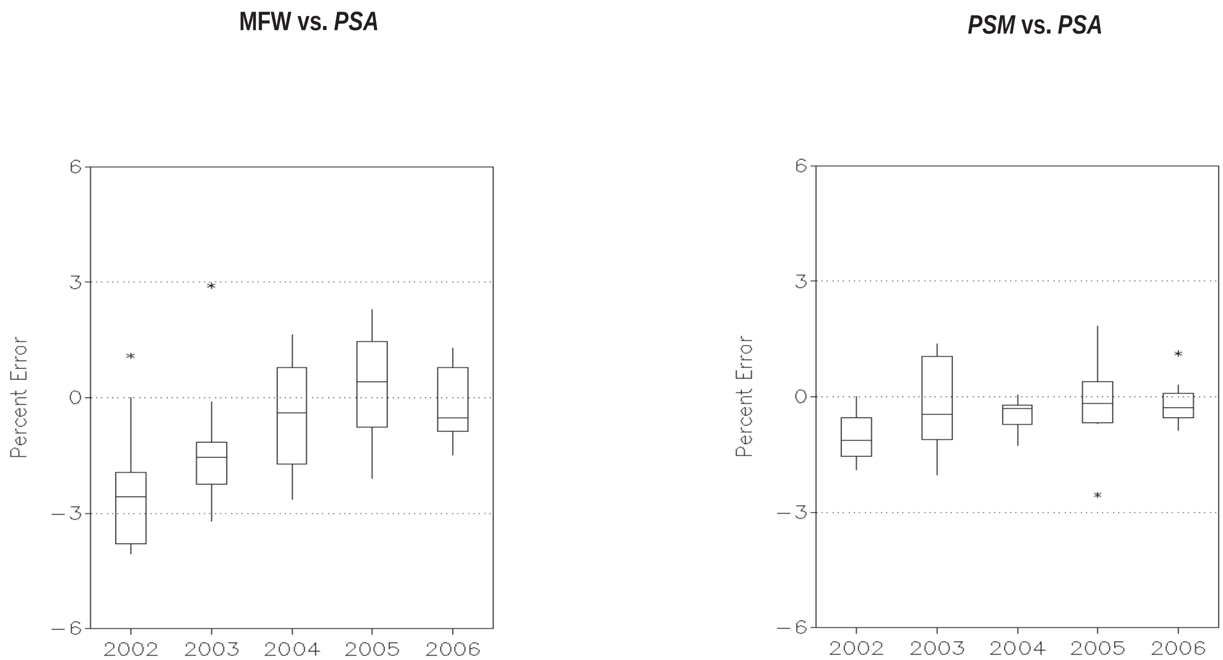


PSM* vs. *PSA



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE9. Box Plots of Percent Errors for MFW and *PSM* Crude Oil Imports Excluding SPR Data, 2002 - 2006



Source: Energy Information Administration, Petroleum Supply Reporting System.

Conclusion

In summary, similar to previous years, the interim *PSM* data were closer in value to the final *PSA* volumes than the MFW estimates. This is largely a result of the longer time period provided to process the monthly data and automated accounting systems of the monthly respondents.

In 2006, 38 of 66 *PSM* interim values were within 1 percent (mean absolute percent error) of the final values; 27 of 62 MFW estimates were within 2 percent (mean absolute percent error) of the final values; and 10 of those 27 were within 1 percent. As in previous years, the accuracy of 2006 preliminary and interim values varied by product and by petroleum supply type. As a group, stocks continued to have the most accurate MFW estimates and *PSM* interim values.

The good coverage for weekly surveys across petroleum supply type and product combinations has contributed to the accuracy of weekly estimates. In 2006, for 20 of the 21 categories, coverage was above 90 percent. The 2006 response rates for the weekly and monthly surveys were comparable to the 2005 response rates.

To successfully maintain and improve the accuracy of these data, the Petroleum Division (PD) is participating in several Office of Oil and Gas initiatives in the areas of survey data collection, survey processing, automation, data quality control, and data dissemination.

In the area of survey data collection and processing, the PD continued to perform a comprehensive review of current petroleum industry operations to ensure relevant data are collected for Federal, State, and private customers to analyze and assess the U.S. petroleum market. To improve survey clarity, PD added information to the "Frequently Asked Questions" or the FAQ section for each of the petroleum surveys on the EIA website at the following link: petroleum_survey_forms In addition, a tracking system for the *Petroleum Supply Annual* (PSA) was developed to highlight issues that were unresolved during monthly processing that need to be resolved for the *PSA*.

In the area of automation, the PD continued to make enhancements to the Data Collection Module (DCM) which allows data from numerous data collection forms to be transformed into an electronic form within a common system and make enhancements to the Standard Energy Processing System (STEPS) which is designed to handle different surveys with different needs using generalized programs and data structures to process survey data. In addition, the Electronic Data Extraction System (EDES) automatically extracts data from Excel spreadsheets submitted by some survey respondents through Secure File Transfer or email, and transforms the data into a format that can be sent to the DCM and then to STEPS. In 2006, EDES was developed for the Petroleum Supply Weekly Surveys. The data are transformed into a format that can be sent to the Weekly Petroleum Supply Processing System. The PD continues to explore Internet Data Collection and resources needed for implementation.

In the area of data quality control, the PD enhanced the edit and imputation functionality in STEPS and continued to expand the Survey Information System (SIS) which contains information needed for data validation and ad hoc queries. The system is a valuable link between the output from STEPS and data repository systems which produce the publications. A new version of the query system was released in 2007 to expand the data series, incorporate user requests, and improve functionality. The PD is currently working on developing automated imputation for the monthly surveys in STEPS, which will improve the quality of the data. In addition, a quality control tool was developed to facilitate monthly data validation by identifying data outliers. The tool allows analysts to select, using a drop down menu, the survey, product code, supply type, and geographic region. The analyst can then rank the data that meet the specified criteria according to data volume or change from prior month. Selected data are displayed in a spreadsheet as well as graphically. Two graphs are generated. The first graph displays 6 years of historical monthly data as well as standard deviations, monthly-from-weekly values, and system-generated imputation values. The second graph displays weekly submitted, weekly published, and monthly data.

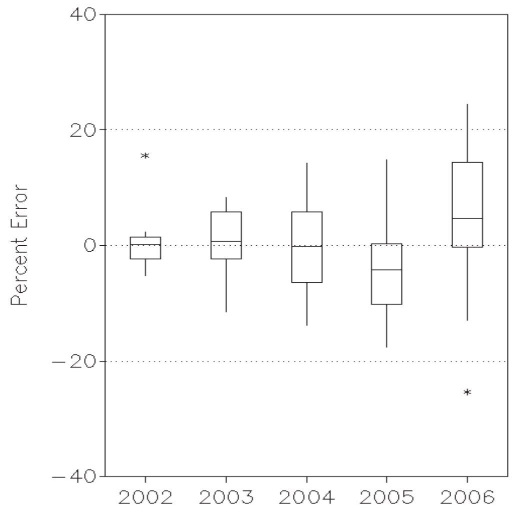
In the area of data dissemination, the web product, Petroleum Navigator, provides an integrated and consistent interface for accessing a comprehensive set of EIA's petroleum data. Features include: downloadable spreadsheets containing complete data history, data tables which "pivot" to present different perspectives, and selection boxes to easily change the product, area, process, period, and unit of measure. Petroleum Navigator can be accessed at the following website: petroleum_navigator. In 2006, Petroleum Navigator was enhanced to include additional data series and to extend current series further back in time. In 2007, data on drilling and wells were included. There are now more than 100,000 data series. In addition, a new design for the Petroleum Supply and Disposition table was implemented which is similar to the publication format. The new design provides a larger perspective of the data, displaying all the components of supply and disposition for all products on one page in a given period. This new table can be found at the following link: petroleum_supply_disposition.

Some other ongoing areas of improvement include the continuation of non-response follow-up and customer outreach and the continuation of efforts to insure compliance with reporting requirements. Results of these efforts should enable the PD to continue to provide accurate weekly and monthly data estimates.

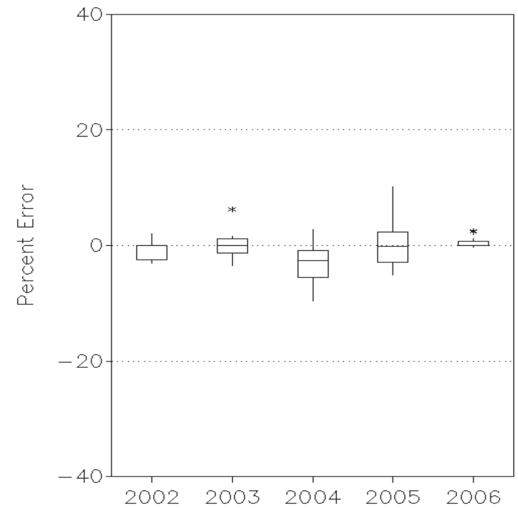
Figure FE10. Box Plots of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Imports Data, 2002 - 2006

Motor Gasoline Imports

MFW vs. *PSA*

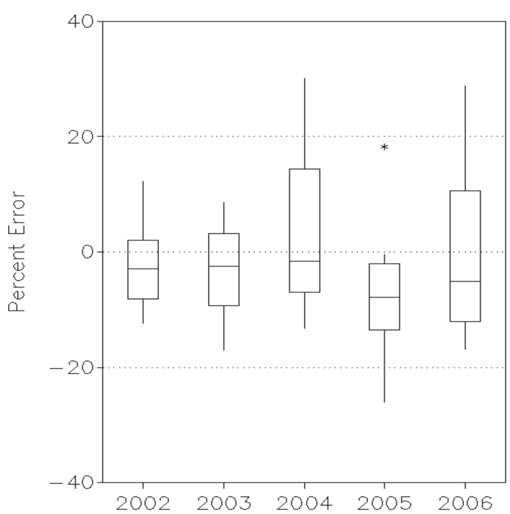


PSM* vs. *PSA

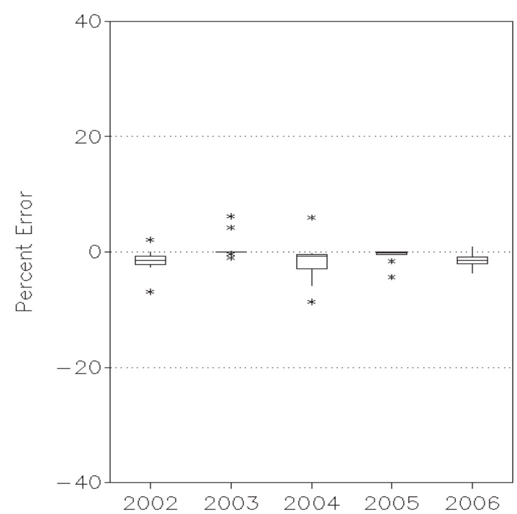


Distillate Fuel Oil Imports

MFW vs. *PSA*



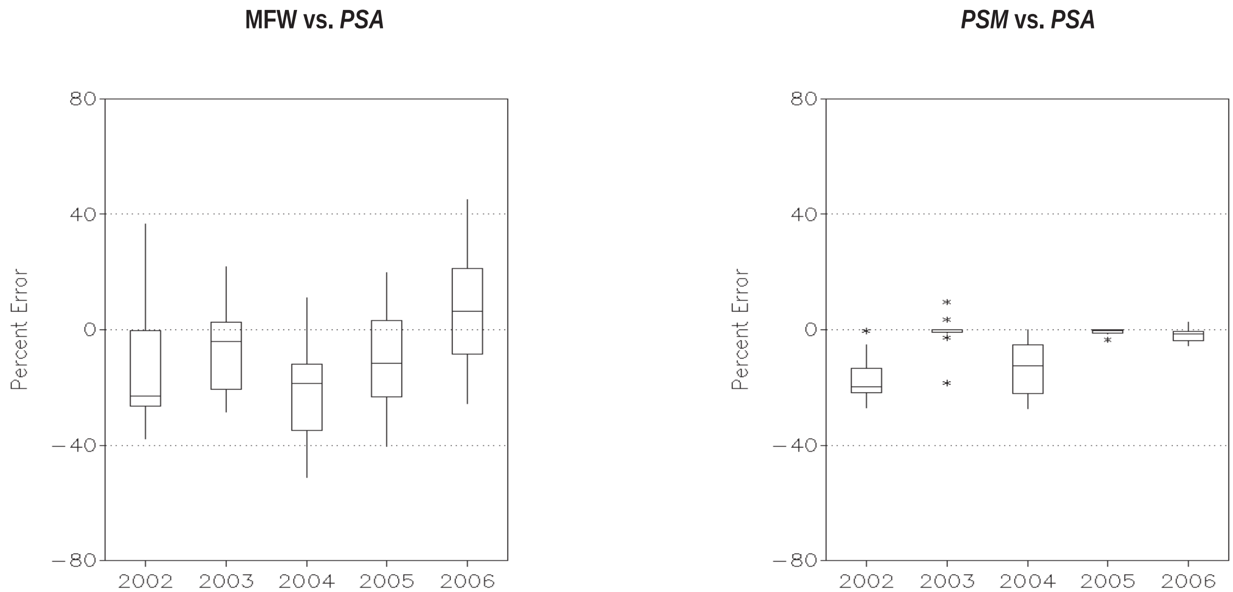
PSM* vs. *PSA



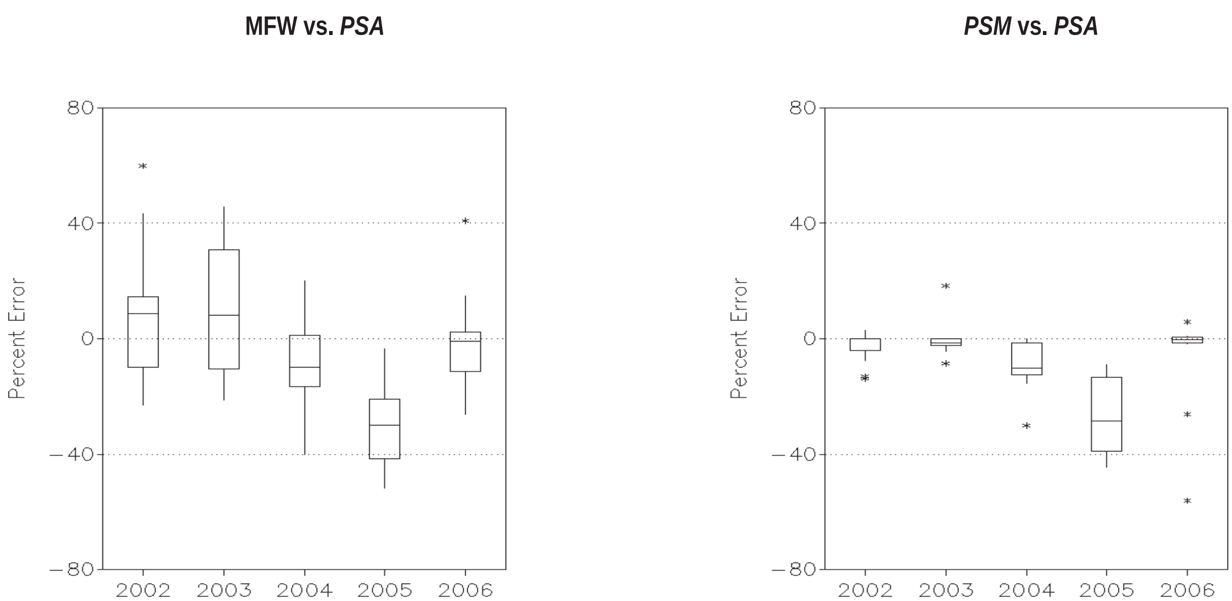
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE11. Box Plots of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Imports Data, 2002 - 2006

Residual Fuel Oil Imports



Jet Fuel Imports



Source: Energy Information Administration, Petroleum Supply Reporting System.