Accuracy of Petroleum Supply Data

by Tammy G. Heppner and Carol L. French

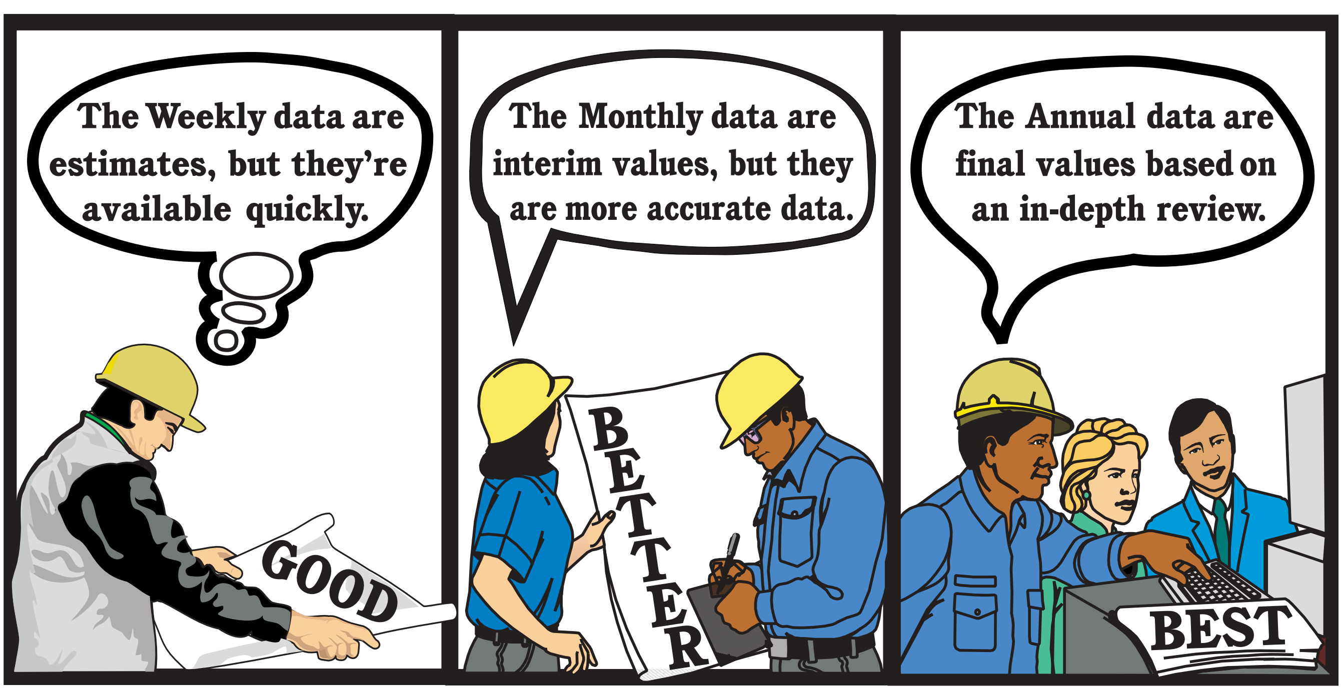
Overview

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 2001 data from good, to better, to best, for initial estimates to final values. These data were presented in a series of PD publications: the *Weekly Petroleum Status Report* (WPSR), the *Winter Fuels Report* (WFR), the *Petroleum Supply Monthly* (PSM), and the *Petroleum Supply Annual* (PSA). Weekly estimates in the *WPSR* and *WFR* were the first values available.

Figure FE1 illustrates that as reporting time increases from the weekly estimates to the interim monthly values to the final petroleum supply values, there is more in-depth review resulting in an improvement in the accuracy of the data. For the monthly-from-weekly (MFW) data, respondents have the shortest reporting time, and the data are least accurate but "good." For the *PSM* data, respondents have a longer reporting time than the weekly, and the data are more accurate or "better." For the *PSA* data, respondents have the longest reporting time, and the data are the most accurate or "best." For 2001, 66 petroleum supply data series were analyzed to determine how close the *PSM* values were to the final *PSA* values. For these series, 36 out of the 66 were within 1 percent of the *PSA* values in terms of mean absolute percent error as compared to 37 in 2000. Sixty-one petroleum supply data series were analyzed to see how close the MFW estimates were to the final *PSA* values. For these 61 series, 25 were within 2 percent of the *PSA* values in terms of mean absolute percent error and, of those, 11 were within 1 percent, compared to 24 and 8, respectively, for 2000.

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 later 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* is available electronically 5 days after and in hardcopy 7 days after the close of the reference week (excluding holiday weeks). WFR data are available

Figure FE1. Working Hard to Improve the Accuracy of 2001 Data



electronically and in the *WPSR*. About 5 months after the end of the reference year, final monthly values, reflecting resubmissions, are published in the *PSA*.

Historically, the weekly publications (*WPSR* and *WFR*) 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

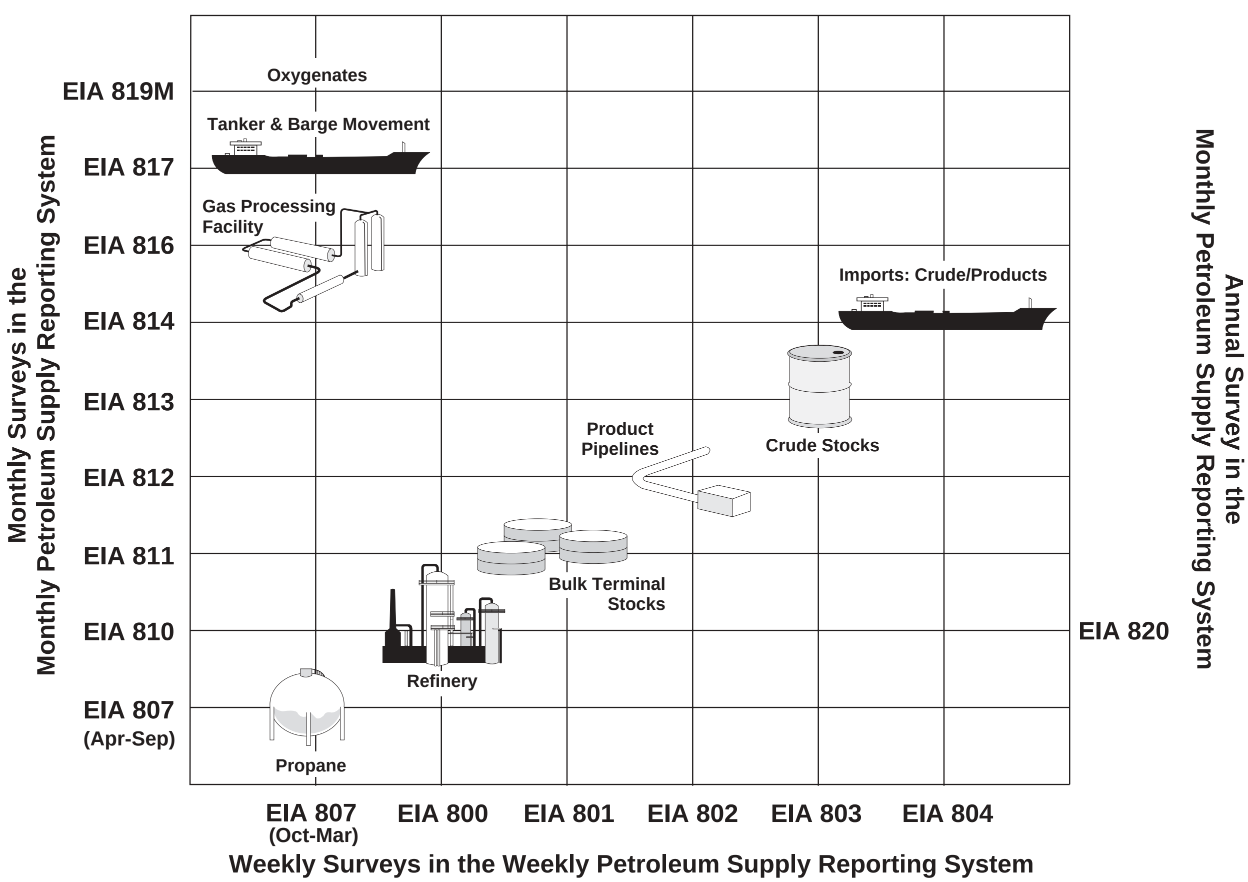
The 15 surveys in the Petroleum Supply Reporting System (PSRS) track the supply and disposition of crude oil, petroleum products, and natural gas liquids in the United States. To maintain a database with historically accurate observations and current estimates from the petroleum industry, EIA administers 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 five weekly surveys. In addition, the Form EIA-807, "Propane Telephone Survey," collects data weekly from October through March. The MPSRS includes eight monthly surveys, one annual survey, and the Form EIA-807 monthly data, which are collected from April through September.

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 five key points along the petroleum production and supply path: (1) refineries, (2) bulk terminals, (3) product pipelines, (4) crude oil stock holders, and (5) importers of crude oil and products.

Annual U.S. refinery capacity data are collected on the Form EIA-820, "Annual Refinery Report." These data were collected and published in Volume 1 of the *PSA* for 2001, available in hardcopy and electronically. Volume 2 of the *PSA* is only available electronically.

Figure FE2. Petroleum Supply Reporting System: Surveys and Subsystems



Source: Energy Information Administration, Petroleum Supply Reporting System.

The Weekly Petroleum Supply Reporting System

The WPSRS contains the data collected from the five 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 terminal stocks. 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 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, facsimile, electronic spreadsheets, 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 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, data are made available to the public on the EIA's internet web site (http://www.eia.doe.gov) and within 7 days in hardcopy (through the *WPSR*). Except when holidays delay data processing schedules, values for the weekly variables, with the exception of propane, are available via the internet at 9:00 a.m. Eastern Time on the Wednesday following the close of the reference week. A new weekly web product, This Week in Petroleum (TWIP), provides analysis, data, and charts of the latest weekly petroleum supply and price data, including propane data. TWIP is generally available at 1:00 p.m. on Wednesdays at http://tonto.eia.doe.gov/oog/info/twip/twip.asp.

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. Except for the Form EIA-819M, the deadline for filing monthly surveys is the 20th calendar day following the end of the report month. Data collection for the Form EIA-819M begins on the seventh working day of the month. Form EIA-819M data are solicited by telephone or received by facsimile or electronic mail. Data for the other monthly surveys are reported via mail, telephone, facsimile, electronic spreadsheets, 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* and on the internet. 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. The impact of resubmissions to previous months published data are presented in Appendix C of the *PSM*. Additionally, preliminary company-level imports data are released electronically between the 7th and 10th of each month.

Beginning with the February 1994 *PSM*, Table H1, "Petroleum Supply Summary" was included to show early estimates of monthly data. The current-month values in Table H1 are preliminary estimates based on weekly submissions. These monthly-from-weekly estimates are published in the *WPSR* and on the internet on the Wednesday following the first Friday of each month.

Within 5 months of the end of the calendar year, the final monthly values for the previous year are published in the *PSA*. 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

Table FE1. Average Coverage for Weekly Surveys, 2001 and 2000 (Percent of Final Monthly Volumes Included in Monthly-from-Weekly Sample)

Product	Stocks						Production		Imports	
	Refinery		Bulk Terminal		Pipeline					
	2001	2000	2001	2000	2001	2000	2001	2000	2001	2000
Total Motor Gasoline	98	97	93	94	97	97	99	99	91	97
Jet Fuel	98	98	92	92	100	99	99	99	91	64
Distillate Fuel Oil	96	95	87	88	98	98	97	97	96	94
Residual Fuel Oil	96	96	89	89	—	—	95	95	96	95
Crude Oil	96	96	—	—	—	—	—	—	94	95

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

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, the Form EIA-819M, “Monthly Oxygenate Telephone Report,” and the Form EIA-807, “Propane Telephone Survey,” 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) 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 subsection 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 five 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.

For the weekly surveys, better coverage will most likely reduce sampling error. As shown in Table FE1, 2001 coverage was comparable to 2000. Of the 21 product and supply type combinations, 19 had coverage of 90 percent or above in 2001. For 11 of the 21 combinations, 2001 coverage increased from 2000. Tabulations were done before rounding of the coverage values. Jet fuel imports display a large percentage increase from 2000 to 2001, from 64 to 91 percent, because of the resolution of noncompliance issues.

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, newspaper articles, and correspondence from respondents.

Table FE2. Average Response Rates for Monthly and Weekly Surveys, 2001

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	254	248	97.7	190	177	93.1
Bulk Terminal	256	245	95.5	65	60	92.9
Pipeline	82	81	98.4	42	40	95.1
Crude Oil Stocks	163	160	97.8	67	64	95.1

[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.

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 2001 response rates for the weekly surveys and their corresponding monthly surveys are enumerated in Table FE2. All of the 2001 response rates for each of the EIA weekly and monthly surveys decreased from 2000. The largest decrease in response rate was for the weekly refinery survey, decreasing from 96.5 percent in 2000 to 93.1 percent in 2001.

To mitigate the effect of nonresponse, imputed values are calculated for all nonreported values except monthly imports. Weekly imputed values are the exponentially smoothed mean of that respondent's historical values for that variable. Monthly imputed values are 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, zero is the value imputed for nonreported cells on the monthly imports survey. In addition, the monthly imports are collected and published at a much greater level of detail than the weekly imports, which makes imputation impractical.

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, except the Form EIA-819M and Form EIA-807, 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.

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 61 of them were published weekly throughout 2001. 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.

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

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

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 observation.

The average final absolute volumes and the mean absolute percent error for MFW estimates and *PSM* interim values for 2001 and 2000 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 2001 MFW mean absolute percent errors which were within 2 percent of their respective *PSA* values (25 of the 61 MFW series), and the 2001 *PSM* mean absolute percent errors which were within 1 percent of their *PSA* values (36 of the 66 *PSM* series), are distinguished by a single asterisk. Mean absolute percent errors that were greater than 10 percent are marked by a double asterisk. There were 16 such MFW series and 8 *PSM* series, compared to 16 and 9, respectively, for 2000.

For 2001, 7 of the 11 weekly production series decreased in mean absolute percent error from 2000. 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, 11 of the 14 *PSM* production series in 2001 show an increase in mean absolute percent error from 2000. Weekly fuel ethanol supply and disposition data are not available; therefore, the weekly oxygenated motor gasoline field production is based on the latest available monthly value.

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. A major exception is the double asterisk shown by the MFW percent error for oxygenated motor gasoline stocks. The increase is related to the average absolute volume. Fuel ethanol and methyl tertiary butyl ether stocks are not collected weekly, but are collected on the Form EIA-819M, "Monthly Oxygenate Telephone Report." The survey provides production data and preliminary stock data from a sample of respondents reporting on the monthly surveys and from the universe of oxygenate producers. These data are displayed in Appendix D of the *PSM.* Interim data are collected later on the monthly surveys and published in the *PSM*. Twelve of the 19 weekly and monthly stock series for 2001 decreased or stayed the same in mean absolute percent error from 2000.

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.

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 2001 and 2000

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	2001	2000	2001	2000	2001	2000
Crude Oil Production (thousand barrels/day)	5,801	5,822	* 0.64	1.20	* 0.92	0.42
Refinery Operations						
Refinery Crude Oil Inputs (thousand barrels/day)	15,128	15,067	* 0.57	0.39	* 0.06	0.08
Operating Utilization Rate (percent)	93	93	* 1.17	1.39	* 0.34	0.07
Production (thousand barrels/day)						
Total Production	19,537	19,531	—	—	* 0.35	0.10
Refinery Production	17,285	17,243	4.17	1.10	* 0.33	0.08
Finished Motor Gasoline	8,312	8,186	* 0.71	1.23	* 0.28	0.42
Reformulated Motor Gasoline	2,597	2,567	* 1.30	1.94	1.09	0.34
Oxygenated Motor Gasoline	762	774	** 24.64	15.99	** 13.92	6.15
Other Motor Gasoline	4,953	4,845	4.13	2.37	1.69	1.21
Jet Fuel	1,530	1,607	* 0.82	0.87	* 0.05	0.05
Distillate Fuel Oil	3,695	3,580	* 1.15	1.99	* 0.14	0.08
Low Sulfur Distillate Fuel Oil	2,617	2,473	* 1.26	1.80	* 0.42	0.16
High Sulfur Distillate Fuel Oil	1,077	1,107	2.69	3.80	* 0.78	0.34
Residual Fuel Oil	721	696	* 1.98	3.58	* 0.30	1.44
Other Products	5,281	5,463	—	—	* 0.90	0.81
Propane	1,095	1,122	—	—	* 0.59	0.25
Other Products Refinery Production	3,318	3,410	** 27.16	7.46	* 0.35	0.15
Stocks (thousand barrels)						
Total Stocks	1,543,493	1,503,920	* 0.78	0.32	* 0.20	0.29
Total Stocks, excl. SPR	999,365	938,942	* 1.20	0.51	* 0.31	0.47
Total Crude Stocks	854,131	852,743	* 0.49	0.53	* 0.27	0.31
Crude Oil Stocks, excl. SPR	310,003	287,765	* 1.33	1.37	* 0.73	0.93
SPR Stocks	544,128	568,498	* 0.00	0.04	* 0.00	0.00
Refined Products Stocks	689,362	651,177	* 1.24	0.83	* 0.15	0.30
Total Motor Gasoline Stocks	206,421	201,610	* 0.59	1.47	* 0.19	0.34
Reformulated Motor Gasoline Stocks	43,271	41,516	* 2.00	2.81	* 0.65	0.20
Oxygenated Motor Gasoline Stocks	658	891	** 30.84	21.12	** 14.12	20.56
Other Motor Gasoline Stocks	113,700	115,362	* 0.87	1.50	* 0.45	0.35
Jet Fuel Stocks	41,851	42,511	* 1.92	2.11	* 0.42	0.30
Distillate Fuel Oil Stocks	121,063	109,313	* 1.57	1.95	* 0.19	0.31
Low Sulfur Distillate Fuel Oil Stocks	70,535	67,175	* 1.92	2.06	* 0.46	0.17
High Sulfur Distillate Fuel Oil Stocks	50,528	42,138	* 1.39	2.21	* 0.58	0.67
Residual Fuel Oil Stocks	38,896	36,420	2.03	1.96	* 0.99	0.58
Other Products Stocks	281,131	261,322	2.37	1.59	* 0.42	0.44
Propane Stocks	50,330	42,965	2.17	3.42	1.07	0.61
Fuel Ethanol Stocks	3,568	4,299	7.06	4.37	4.19	2.59
Methyl Tertiary Butyl Ether Stocks	7,866	8,276	3.16	3.36	* 0.56	0.93
Stock Change (thousand barrels/day)						
Total Stock Change	485	469	**105.63	80.31	** 33.78	29.41
Crude Stock Change	310	220	**103.24	125.80	** 37.05	42.31
Refined Products Stock Change	453	490	**235.07	74.45	** 35.60	11.77
Imports (thousand barrels/day)						
Total Imports	11,871	11,459	3.77	4.17	2.11	3.21
Total Crude Imports	9,318	9,062	2.37	2.41	1.94	1.54
Crude Oil Imports, excl. SPR	9,328	9,070	2.38	2.41	1.94	1.54
SPR Imports	0	0	* 0.00	0.00	* 0.00	0.00
Refined Products Imports	2,543	2,389	8.63	12.29	2.55	9.50
Finished Motor Gasoline Imports	455	427	9.31	19.93	2.24	15.69
Reformulated Motor Gasoline Imports	217	197	8.72	16.11	2.49	4.91
Oxygenated Motor Gasoline Imports	1	1	** 22.60	23.29	* 0.00	8.47
Other Motor Gasoline Imports	236	229	** 14.71	27.65	2.89	24.97
Jet Fuel Imports	148	162	** 15.56	18.42	2.24	12.20

See footnotes at end of table.

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 2001 and 2000 (Continued)

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	2001	2000	2001	2000	2001	2000
Distillate Fuel Oil Imports	344	295	** 15.58	8.53	5.41	5.90
Low Sulfur Distillate Fuel Oil Imports	130	134	** 18.69	19.37	6.73	7.29
High Sulfur Distillate Fuel Oil Imports	215	161	** 14.98	8.58	8.07	5.70
Residual Fuel Oil Imports	295	352	** 28.74	28.72	** 33.54	24.78
Other Products Imports	1,301	1,153	** 16.04	5.32	** 12.60	4.07
Propane Imports	145	161	—	—	5.07	19.52
Exports (thousand barrels/day)						
Total Exports	971	1,040	7.68	10.76	2.03	0.00
Crude Oil Exports	20	50	** 540.90	824.82	5.66	0.00
Refined Products Exports	951	990	6.95	11.34	1.80	0.00
Total Net Imports (thousand barrels/day)	10,900	10,419	4.18	4.55	2.38	3.53
Products Supplied (thousand barrels/day)						
Total Products Supplied	19,649	19,701	* 1.63	1.26	* 0.44	1.15
Finished Motor Gasoline Supplied	8,610	8,472	* 0.82	1.73	* 0.37	1.28
Jet Fuel Supplied	1,655	1,725	2.58	2.49	* 0.60	1.27
Distillate Fuel Oil Supplied	3,847	3,722	2.67	1.42	* 0.84	0.86
Residual Fuel Oil Supplied	811	909	** 14.91	9.52	** 15.95	8.34
Other Products Supplied	4,725	4,873	6.36	3.61	2.62	0.80
Propane Supplied	1,142	1,235	—	—	1.58	2.52

— = 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 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.

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 2001 and 2000. Three out of the six stock change values for 2001 increased the number of months that differed from the direction of the *PSA* values, while the other three stayed the same number of months.

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 greater impact when the end of the month occurs in the middle of the week. Eleven of the 15 MFW import series in Table FE3 showed a decrease or stayed the same in mean absolute percent error from 2000 to 2001 compared to last

Table FE4. Number of Months In Which the Direction of NonFinal Stock Change Values Differed From *PSA*

	Number of Months	
	2001	2000
Total Stock Change		
MFW and *PSA* Values	1	1
PSM and *PSA* Values	1	1
Crude Stock Change		
MFW and *PSA* Values	2	1
PSM and *PSA* Values	1	0
Refined Products Stock Change		
MFW and *PSA* Values	2	2
PSM and *PSA* Values	1	0

Source: Energy Information Administration, Petroleum Supply Reporting System.

year's decrease of seven series from 1999 to 2000. For the *PSM*, 11 of the 16 import series decreased or stayed the same in mean absolute percent error compared to last year's decrease of seven import series.

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 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 horizontal line 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 1997 through 2001), 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 plots have the same scale.

The position of the box along the y-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.

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 data reported on Form EIA-182, "Domestic Crude Oil First Purchase Report," EIA estimates weekly and monthly production. Final values 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. The range (2.72) of the 2001 MFW percent errors for crude oil production was the smallest range over the 5-year period. The percent errors were distributed around the smallest median of 0.05 percent. In contrast, the 2001 *PSM* percent errors were distributed around the largest median (1.02) over the 5-year period. Most of the *PSM* interim values overestimated the final *PSA* values. The outlier in September was due to company misreporting.

For refinery crude oil inputs, the range (1.84) of the 2001 MFW percent errors was the smallest range of all other MFW plots analyzed for 2001. As in prior years, the 2001 *PSM* refinery crude oil inputs were extremely close to their final values except for two small outliers in September (0.33) and November (-0.22) due to company nonresponse.

Product Production

As expected, *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 2001 MFW motor gasoline production percent errors, displayed in Figure FE4, ranged from -1.71 to 1.04 percent. Similar to prior years, most of the 2001 *PSM* interim values for motor gasoline production underestimated the final *PSA* values but the percent errors were within 0.62 percent.

For 2001, the MFW percent errors for distillate fuel oil production ranged from -1.29 to 2.24 percent. Similar to 2000, the 2001 *PSM* percent errors were tightly distributed around the median of 0.02 percent. There was one outlier in November (-0.50) due to company nonresponse.

The box and whisker plots for residual fuel oil production and jet fuel production are shown in Figure FE5. The range (6.02) of the 2001 MFW percent errors for residual fuel oil production was the smallest range over the 5-year period, ranging from -3.52 to 2.50 percent. Similar to last year, the median was close to zero. The median for the 2001 *PSM* percent errors for residual fuel oil production was zero. The few months where revisions were sent in, resulted in outliers for those months.

The 2001 MFW percent errors for jet fuel production ranged from -2.56 to 1.69 percent and the median was close to zero. The range (0.28) of the 2001 *PSM* percent errors was the smallest over the 5 years studied and was the smallest range of all other *PSM* plots analyzed for 2001. The outlier in November (-0.21) was due to company nonresponse.

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 first and third 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, MFW vs *PSA* of Crude Oil Production for 1997.)

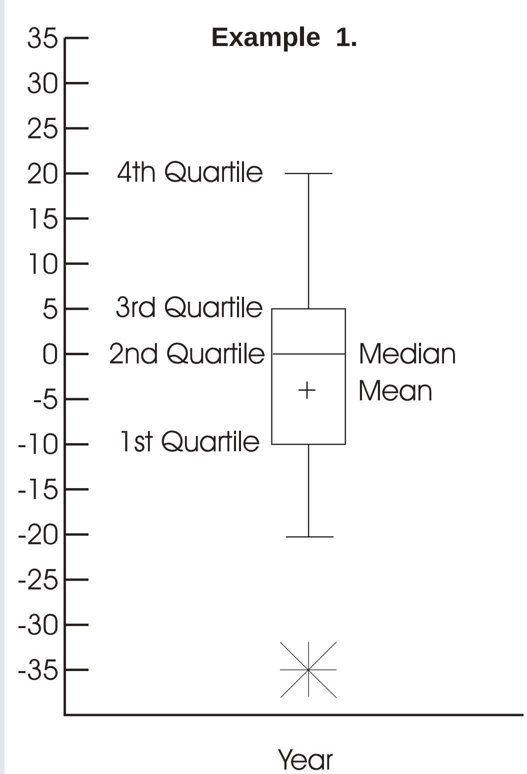
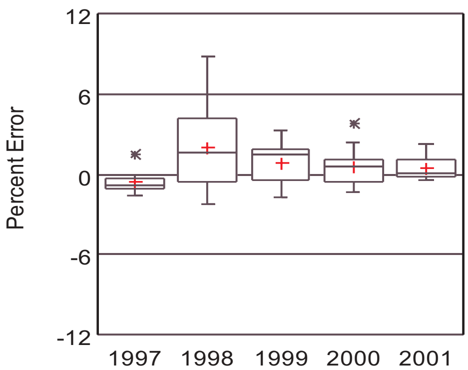


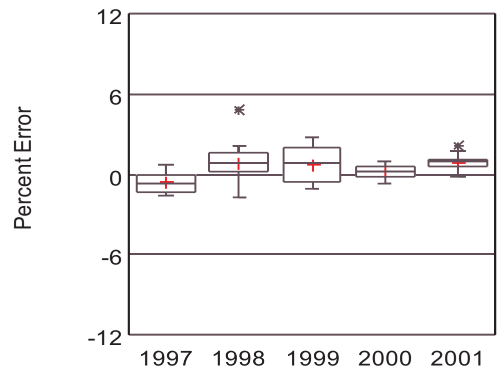
Figure FE3. Range of Percent Errors for MFW and *PSM* Crude Oil Production and Refinery Crude Oil Inputs Data, 1997 - 2001

Crude Oil Production

MFW vs. *PSA*

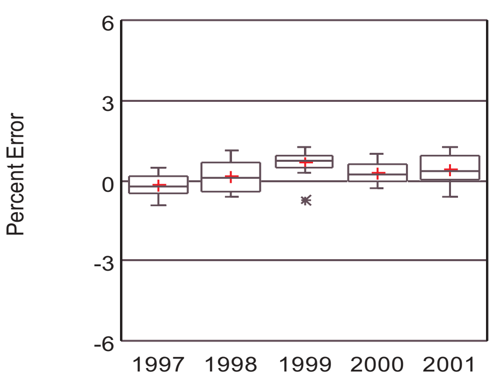


PSM* vs. *PSA

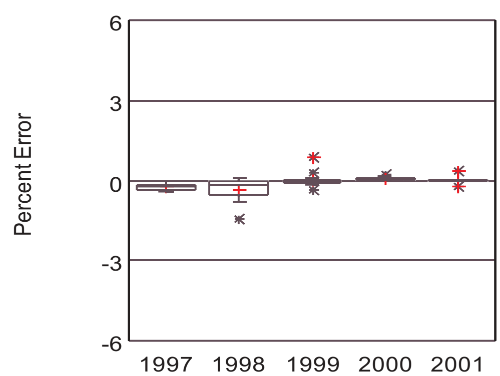


Refinery Crude Oil Inputs

MFW vs. *PSA*



PSM* vs. *PSA

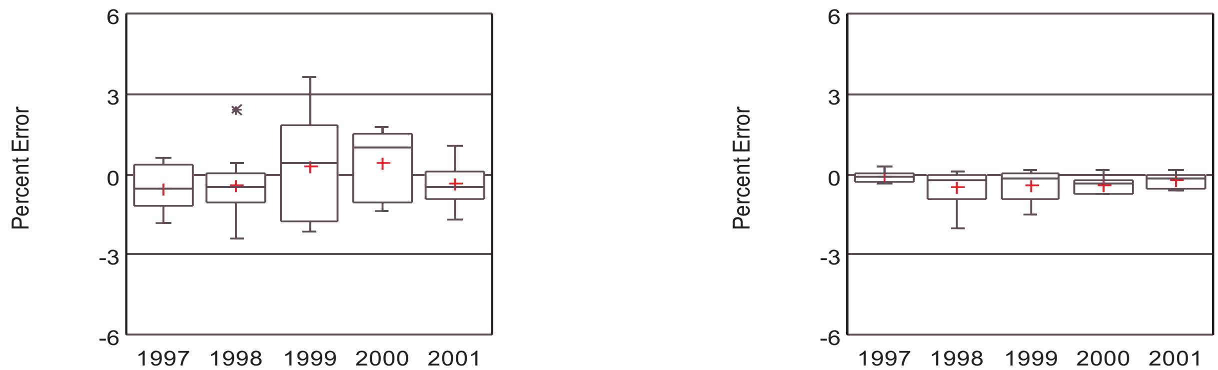


Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE4. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Production Data, 1997 - 2001

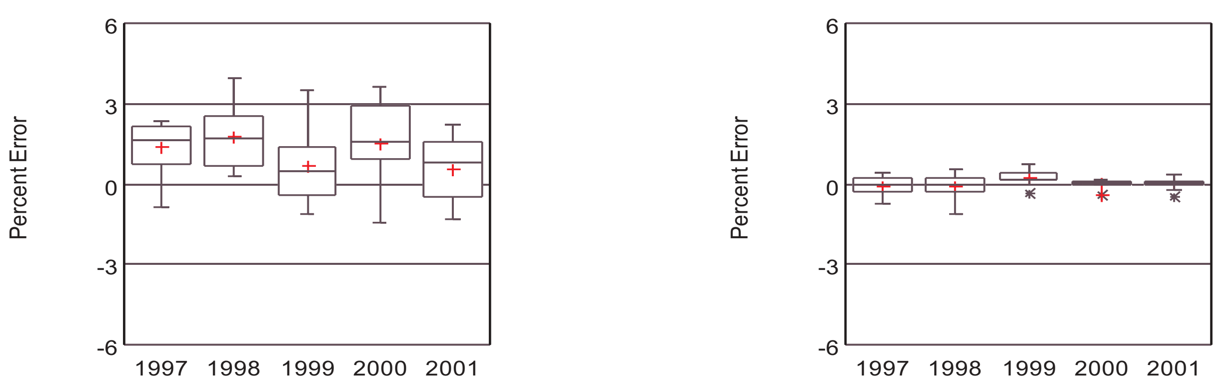
Motor Gasoline Production

MFW vs. *PSA* | *PSM* vs. *PSA*



Distillate Fuel Oil Production

MFW vs. *PSA* | *PSM* vs. *PSA*

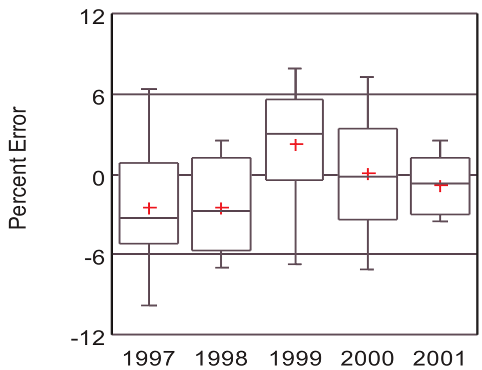


Source: Energy Information Administration, Petroleum Supply Reporting System.

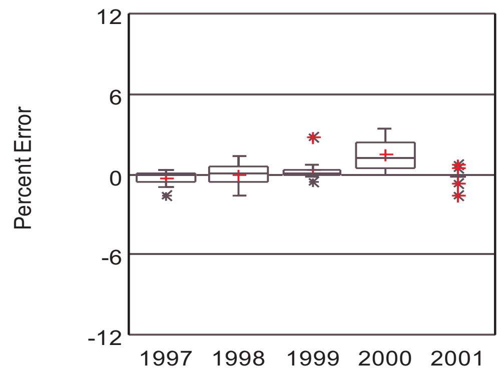
Figure FE5. Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Production Data, 1997 - 2001

Residual Fuel Oil Production

MFW vs. *PSA*

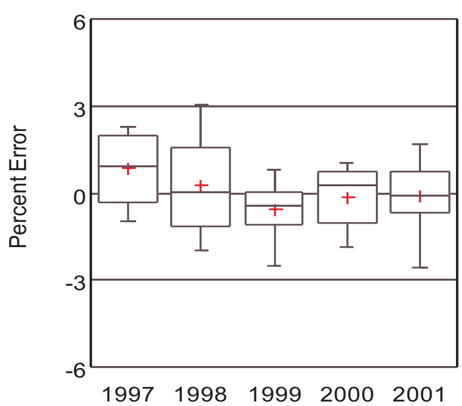


PSM* vs. *PSA

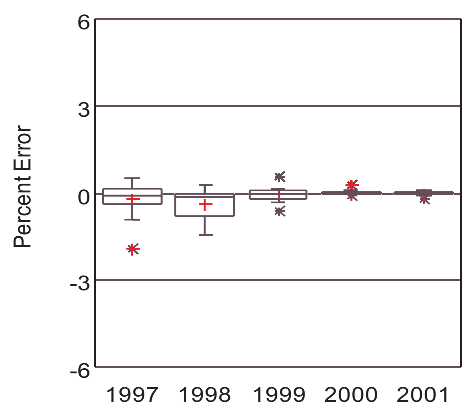


Jet Fuel Production

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. In contrast to last year, all but one of the 2001 MFW estimates for crude oil stocks underestimated the final *PSA* values. There was one outlier in April (-3.65) due to respondent reporting problems. Similarly, all of the 2001 *PSM* interim values for crude oil stocks underestimated the final values. The 2001 median of -0.73 percent had the largest absolute percent error over the 5-year period. The outliers in January, March, and April were due to respondent reporting problems.

Most of the 2001 MFW estimates for motor gasoline stocks were underestimated. The 2001 range of 2.21 percent was the smallest range over the 5-year period and the median was the closest to zero. Similarly, most of the 2001 *PSM* interim values were underestimates. The percent errors for 2001 were within 0.34 percent and were closely distributed around the median of -0.16 percent. The outlier in April (0.34) was due to company revisions.

Figure FE7 shows box and whisker plots for distillate and residual fuel oil stocks. As in prior years, most of the 2001 MFW estimates for distillate fuel oil stocks underestimated the final *PSA* values. The percent errors ranged from -2.93 to 0.92 percent. The 2001 *PSM* percent errors for distillate fuel oil stocks were tightly grouped around the median of -0.05 percent and were within 0.51 percent.

Residual fuel oil stocks typically have larger percent errors than other stock series. In contrast to prior years, most of the 2001 MFW values were overestimates compared to the final *PSA* values. The outliers in January (-4.97) and February (-2.70) were due to respondent reporting problems. The median (0.04) for the 2001 *PSM* percent errors for residual fuel oil stocks was the only positive median over the 5 years studied. The percent errors ranged from -0.71 to 3.22 percent.

The box and whisker plots for jet fuel stocks and propane stocks are shown in Figure FE8. Similar to last year, most of the 2001 MFW estimates for jet fuel stocks overestimated the final *PSA* values. The 2001 range of 7.14 percent was the largest range over the 5-year period, ranging from -3.15 to 3.99 percent. The median of the *PSM* percent errors was close to zero, similar to 2000. One outlier in March (-2.26) was due to company misreporting.

Most of the 2001 MFW estimates for propane stocks overestimated the final values. In contrast to 2000, most of the 2001 *PSM* interim values were underestimates. The 2001 range (3.87) was the largest range over the 5-year period and March (-3.50) had the largest absolute percent error over the past 60 months.

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 that the majority of the 2001 MFW estimates of crude oil imports underestimated the final *PSA* values. The 2001 range of 7.60 percent was the smallest range over the 5-year period. Similar to prior years, all of the 2001 *PSM* interim values underestimated the final *PSA* values.

The distributions of percent errors of the MFW estimates and *PSM* interim values for 1997 through 2001 of motor gasoline and distillate fuel oil imports are shown in Figure FE10. For the most part, the 2001 MFW and *PSM* percent errors for motor gasoline imports were smaller than the 2000 percent errors. The ranges of the MFW (27.84) and *PSM* (12.33) percent errors were the smallest ranges over the 5-year period.

Most of the 2001 MFW estimates for distillate fuel oil imports were underestimated. The 2001 range of 52.42 percent was the largest range over the 5-year period, ranging from -28.64 to 23.77 percent. Unlike prior years, most of the 2001 *PSM* interim values for distillate fuel oil imports overestimated the final *PSA* values. The 2001 median of 3.92 percent was the first positive median in the past three years.

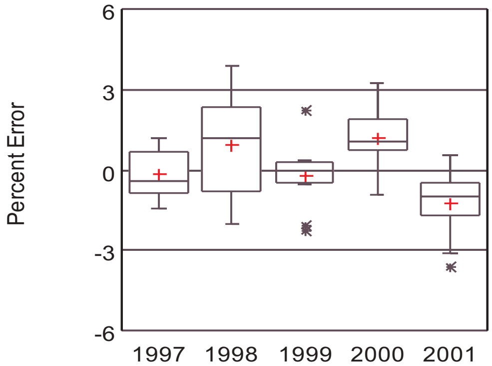
Figure FE11 shows the box and whisker plots for residual fuel oil imports and jet fuel imports. For September 2001, the MFW and *PSM* percent error for residual fuel oil imports was the largest over the 60 months studied, 89.11 and 69.80, respectively. The ranges of the 2001 MFW (91.08) and *PSM* (64.31) percent errors were the largest ranges over the 5-year period. These ranges were also the largest ranges of all other MFW and *PSM* plots analyzed for 2001. The overestimation of the 2001 MFW and *PSM* values was the result of some importers misclassifying unfinished oils as residual fuel oil.

For jet fuel imports, more of the 2001 MFW values overestimated the final *PSA* values than for 2000. The 2001 *PSM* percent errors were tightly grouped around the median of zero. Outliers occurred in May (3.43) and October (-15.87) due to company misreporting.

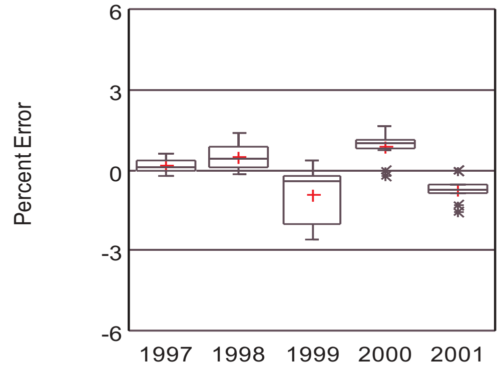
Figure FE6. Range of Percent Errors for MFW and *PSM* Crude Oil Stocks Excluding Strategic Petroleum Reserve (SPR) and Motor Gasoline Stocks Data, 1997 -2001

Crude Oil Stocks Excluding SPR

MFW vs. *PSA*

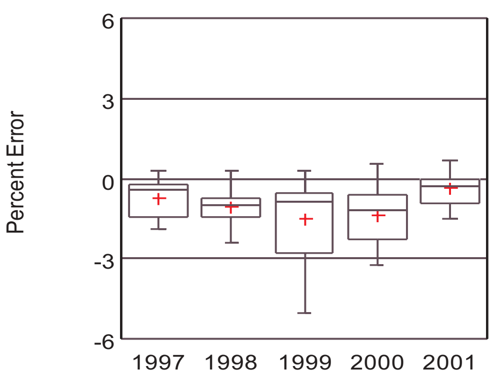


PSM* vs. *PSA

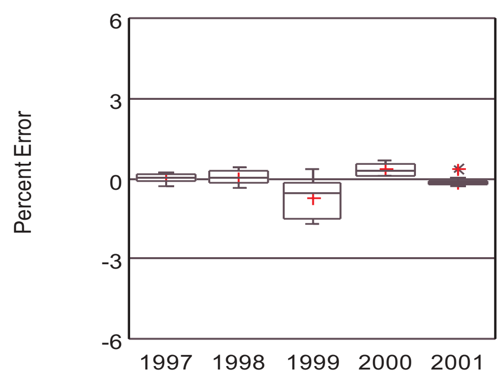


Motor Gasoline Stocks

MFW vs. *PSA*



PSM* vs. *PSA

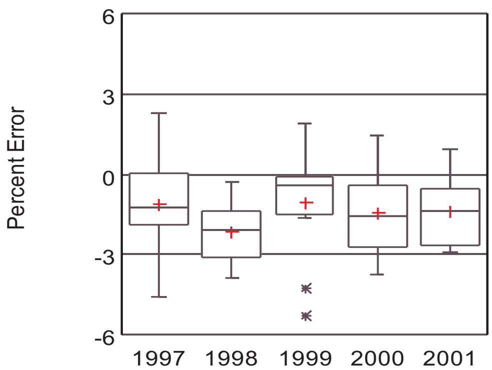


Source: Energy Information Administration, Petroleum Supply Reporting System.

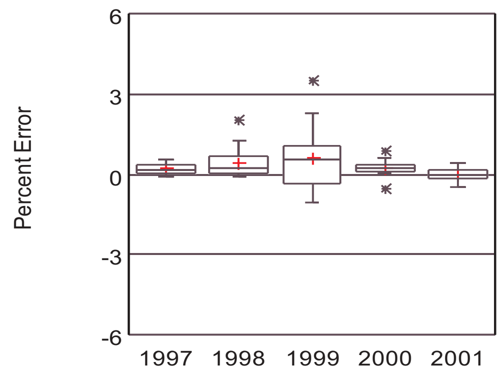
Figure FE7. Range of Percent Errors for MFW and *PSM* Distillate Fuel Oil and Residual Fuel Oil Stocks Data, 1997 - 2001

Distillate Fuel Oil Stocks

MFW vs. *PSA*

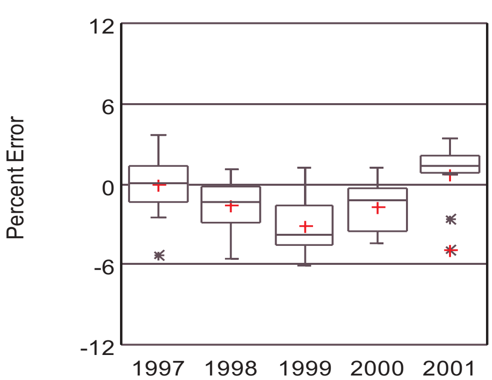


PSM vs. *PSA*

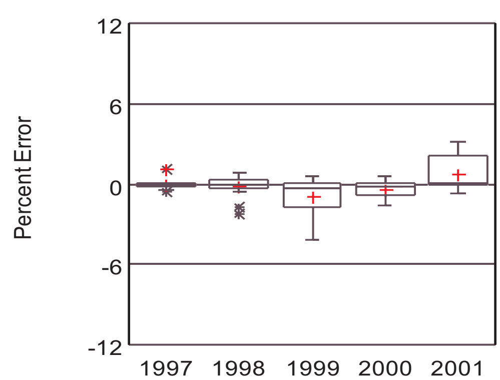


Residual Fuel Oil Stocks

MFW vs. *PSA*



PSM vs. *PSA*

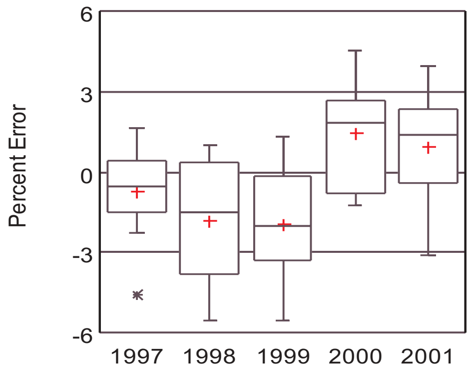


Source: Energy Information Administration, Petroleum Supply Reporting System.

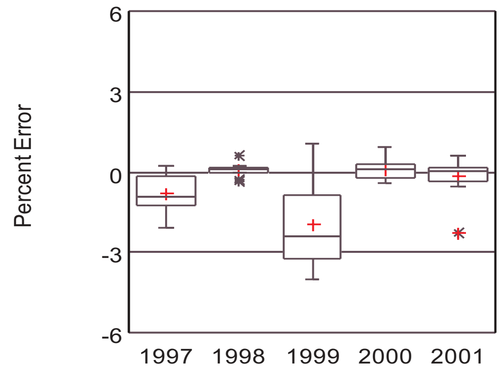
Figure FE8. **Range of Percent Errors for MFW and *PSM* Jet Fuel Stocks and Propane Stocks Data, 1997 - 2001**

Jet Fuel Stocks

MFW vs. *PSA*

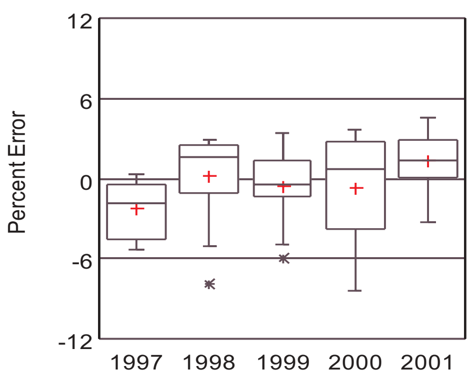


PSM* vs. *PSA

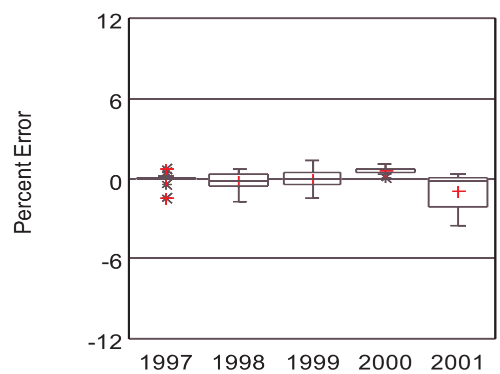


Propane Stocks

MFW vs. *PSA*

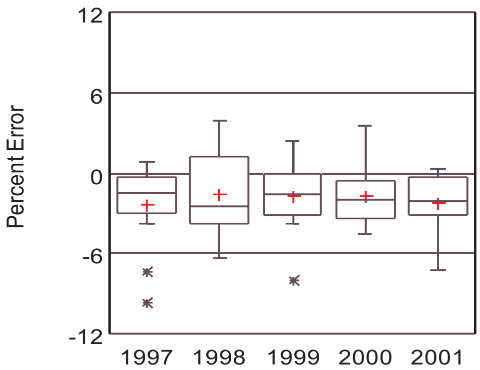


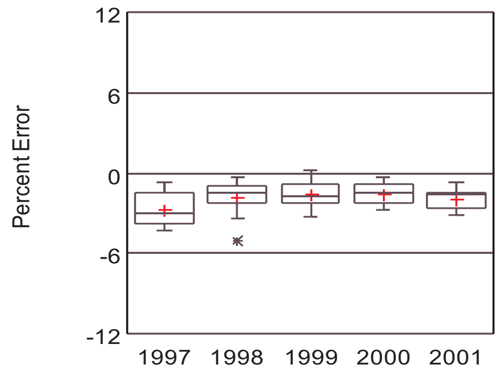
PSM* vs. *PSA



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE9. Range of Percent Errors for MFW and *PSM* Crude Oil Imports Excluding SPR Data, 1997 - 2001





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 monthly respondents' accounting systems.

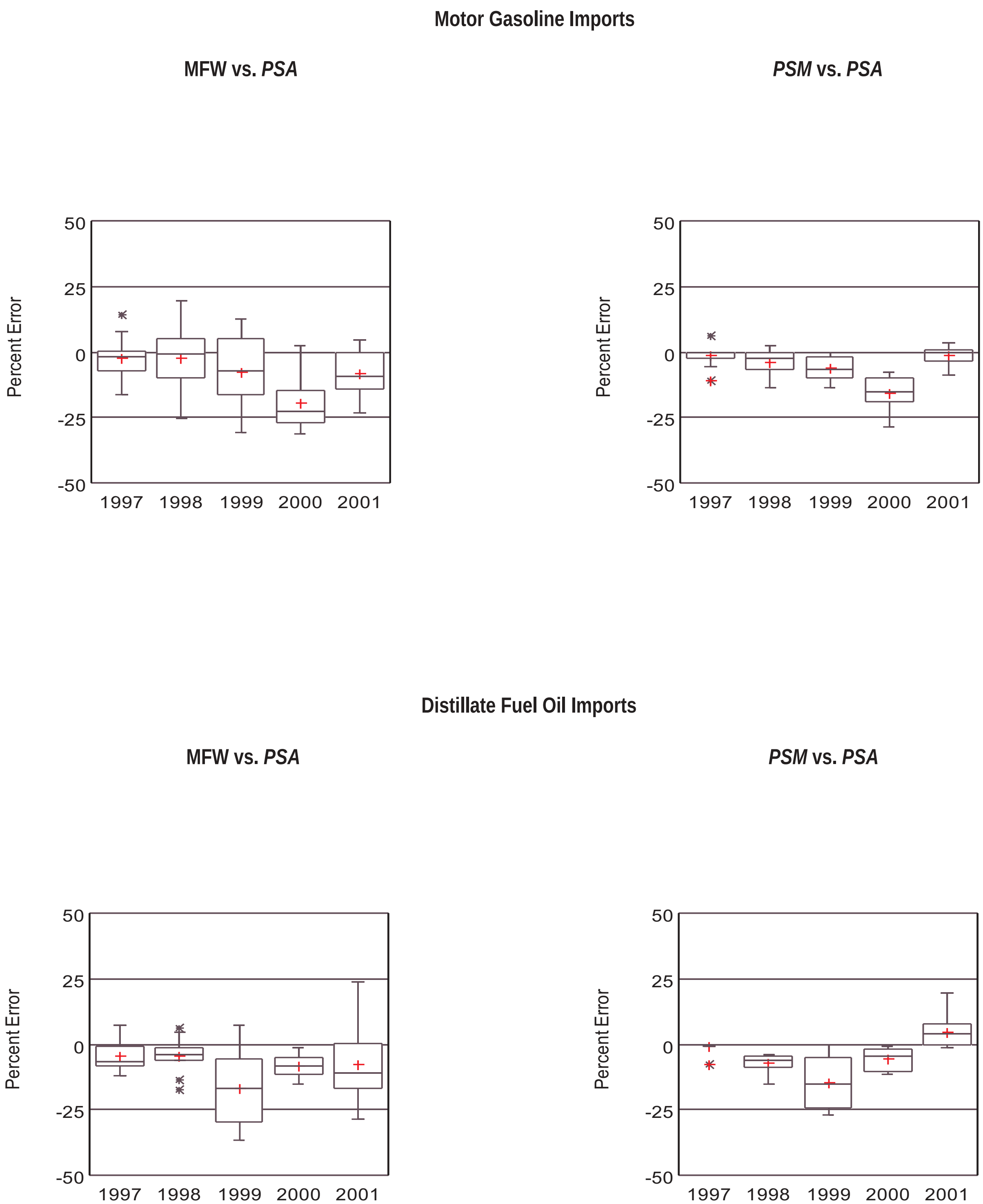
In 2001, 36 of 66 *PSM* interim values were within 1 percent (mean absolute percent error) of the final values; 25 of 61 MFW estimates were within 2 percent (mean absolute percent error) of the final values; and 11 of those 25 were within 1 percent. As in previous years, the accuracy of 2001 preliminary 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 2001, for 19 of the 21 categories, coverage was 90 percent or above. All of the 2001 response rates for the weekly and monthly surveys decreased from the 2000 response rates. Company mergers and budget constraints may have contributed to the decreased response rates.

To successfully maintain and improve the accuracy of these data, the Petroleum Division participated in several Office of Oil and Gas initiatives in the areas of nonresponse follow-up, reporting problems, data dissemination, survey processing systems, data retrieval and analysis systems, and frames maintenance. Some of the specific efforts during 2001 included the expansion and diligence of the nonresponse follow-up and reporting errors teams; increased efforts to insure compliance with reporting requirements; increased research of outside sources to identify new respondents; and the improvement of the petroleum information retrieval on the EIA web site, including many new user-friendly information retrieval options. With the goal of developing a new and improved survey processing system that will upgrade and unify legacy systems by incorporating state-of-the-art technology, the OOG is in the process of implementing a system from the U.S. Census Bureau called the Standard Economic Processing System (StEPS). The OOG system is called the Standard Energy Processing System (STEPS) which is designed to handle different surveys with different needs by using generalized programs and data structures to process surveys. It is written entirely in SAS® and operates in a UNIX environment. The system includes integrated code to perform various survey processing activities including edit/imputation, data review and correction, estimation, variance estimation, and other data collection activities. A team of programmers, analysts, and statisticians is working on the migration of surveys into STEPS.

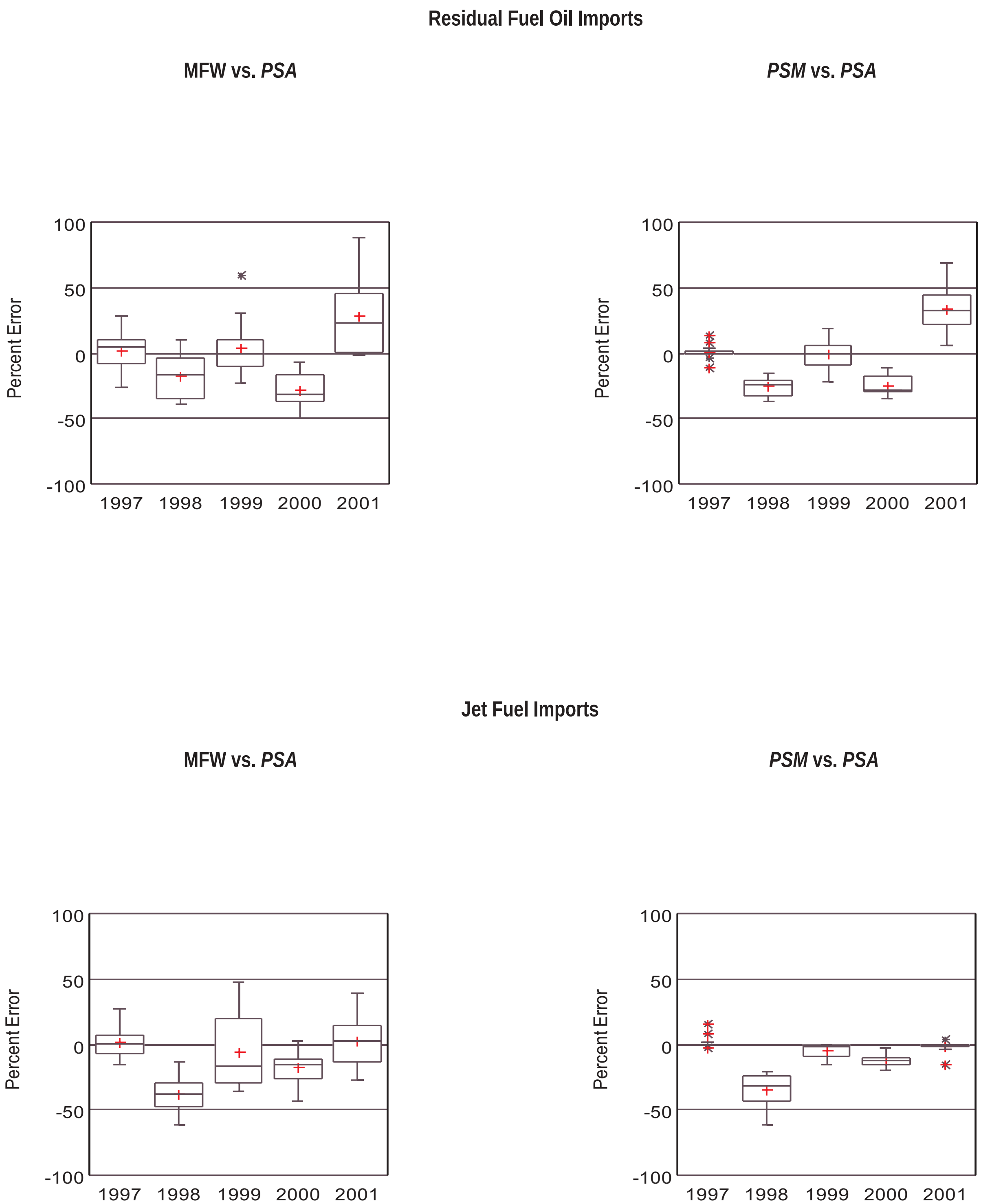
In January 2002, OOG released a new weekly web product, This Week in Petroleum that provides analysis, data, and charts of the latest weekly petroleum supply and price data. Additionally for 2002, OOG will continue to assess and improve PEDRO, the electronic data collection method, and continue to improve survey methodology, data collection forms and instructions, the transmission of encrypted data, the weekly survey processing system, graphical data validation, and the automated data retrieval and query system, Survey Information System (SIS). The SIS is currently being modified to include imputed values as well as reported data so that information will be readily available to analyze and improve the imputation methodologies. The results of these efforts should enable the PD and OOG to continue to provide accurate weekly and monthly data estimates.

Figure FE10. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Imports Data, 1997 - 2001



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE11. Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Imports Data, 1997 - 2001



Source: Energy Information Administration, Petroleum Supply Reporting System.