Accuracy of Petroleum Supply Data

by Tammy G. Heppner and Carol L. French

Overview

Petroleum supply data collected by the Petroleum Supply Division (PSD) of the Energy Information Administration (EIA) were on the right track in 1994. These data are tracked through a series of PSD publications: the *Weekly Petroleum Status Report* (*WPSR*), the *Winter Fuels Report* (*WFR*), the *Petroleum Supply Monthly* (*PSM*), and the *Petroleum Supply Annual* (*PSA*). For the major petroleum products, weekly estimates in the *WPSR* and *WFR* are the first values available.

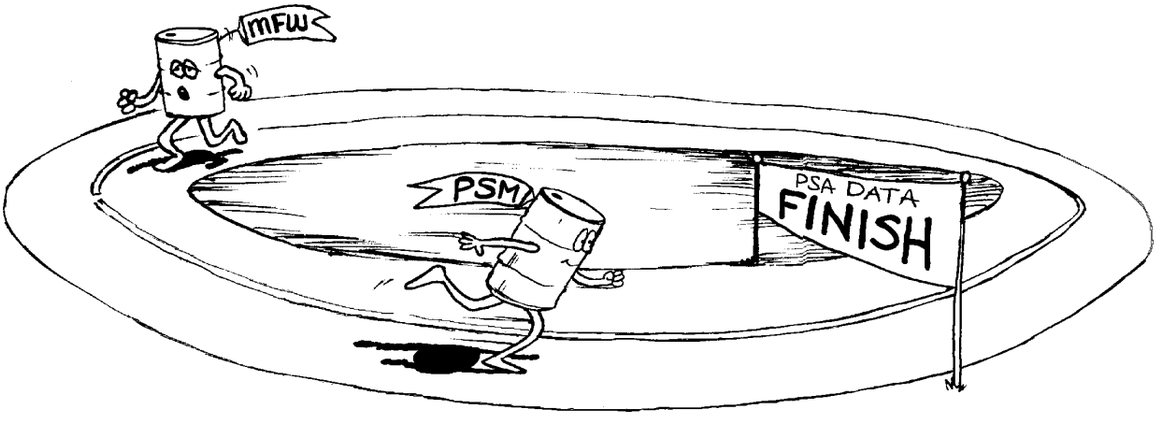
As illustrated in Figure FE1, the *PSA* data represents the "finish line" or "true" values. The *PSM* data are closer to the mark; whereas the monthly-from-weekly (MFW) data are close behind in terms of accuracy. For 1994, 59 petroleum supply data series were analyzed in order to determine how close the *PSM* values were to the final *PSA* values. For these series, 38 out of the 59 were within 1 percent of the *PSA* values in terms of mean absolute percent error. Forty-four petroleum supply series were analyzed in order to see how close the MFW estimates were to the final *PSA* values. For these 44 series, 16 were within 2 percent of the *PSA* values in terms of mean absolute percent error, and of those, 9 were within 1 percent.

One major factor that contributes to the *PSM* values being more accurate than the MFW estimates is the greater length of time between the close of the reference period and the publication date of the *PSM*, thus allowing more in-depth review of the data. 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 and the hardcopy 6 days after the close of the reference week (excluding holiday weeks). During the heating season, propane data are published in hardcopy in the *WFR* and electronically. MFW estimates are published in the *WPSR* and *PSM*. About 5 months after the end of the reference year, final monthly values, reflecting any resubmissions, are published in the *PSA*. Historically, the weekly publications (the *WPSR* and *WFR*), the monthly (*PSM)*, and the annual (*PSA)* provide volumes of crude oil and petroleum products data at relatively increasing levels of accuracy with decreasing timeliness. 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. In order to maintain a database with historically accurate observations and current estimates from the petroleum industry, EIA administers three survey series: weekly, monthly, and annual.

Figure FE1. 1994 PSD Data on the Right Track



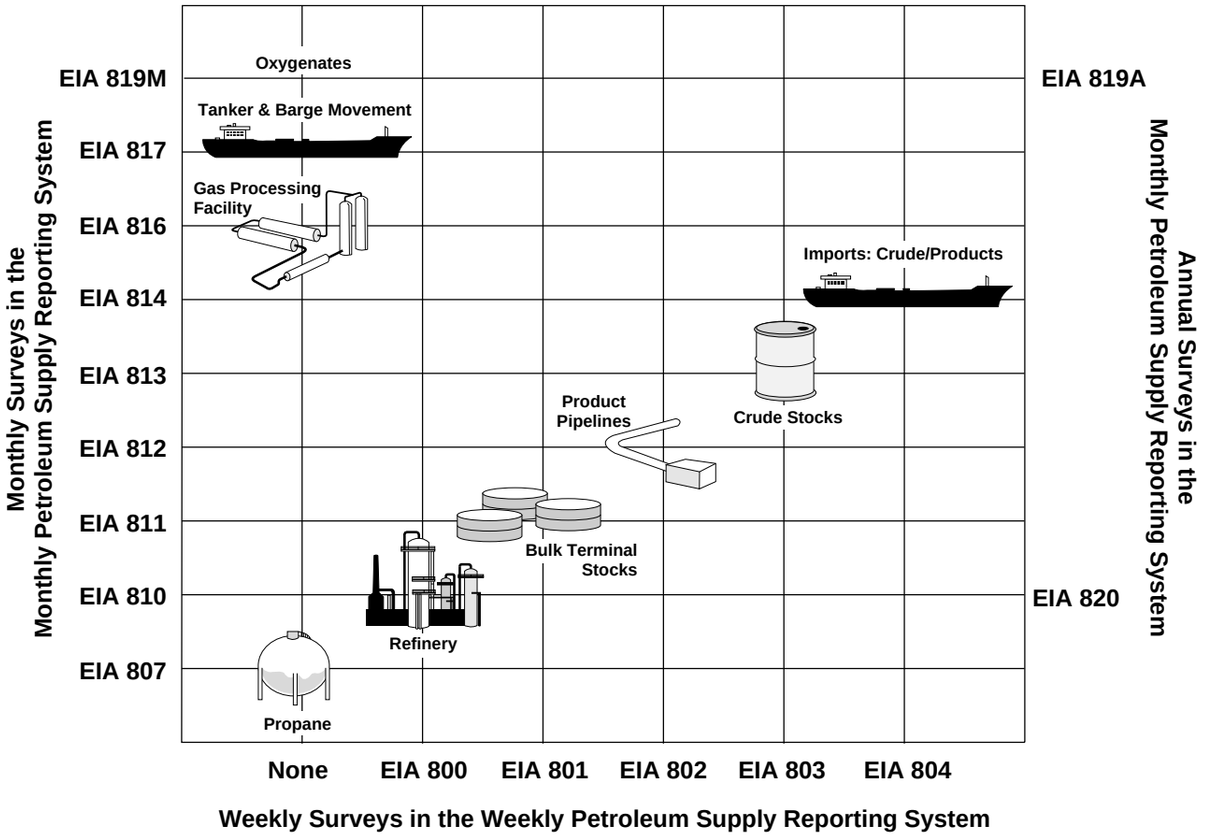
The PSRS is organized into two data collection subsystems, the Weekly Petroleum Supply Reporting System (WPSRS) and the Monthly Petroleum Supply Reporting System (MPSRS). The WPSRS processes data from the five weekly surveys. The MPSRS includes eight monthly surveys, two annual surveys, and the Form EIA-807, which collects propane data monthly from April through September and weekly from October through March. The two annual surveys are the Form EIA-819A, "Annual Oxygenate Capacity Report," used to collect data on current and projected production capacity of oxygenates and annual production and end-of-year inventories of fuel ethanol and the Form EIA-820, "Annual Refinery Report," used to collect data on refinery fuel use and consumption of steam and electricity, refinery receipts of crude oil by method of transportation, operable capacity for atmospheric crude oil distillation units and downstream units, as well as production capacity and storage capacity for petroleum products.

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.

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, weekly, and annual surveys of the PSRS. For example, the target population for the survey Forms EIA-801 and EIA-811 are bulk terminal stock holders. Thus, the respondents to the Form EIA-801 are a sample of the respondents who report to 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 to all published products. 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

Figure FE2. Petroleum Supply Reporting System: Surveys and Subsystems



Source: Energy Information Administration, Petroleum Supply Reporting System.

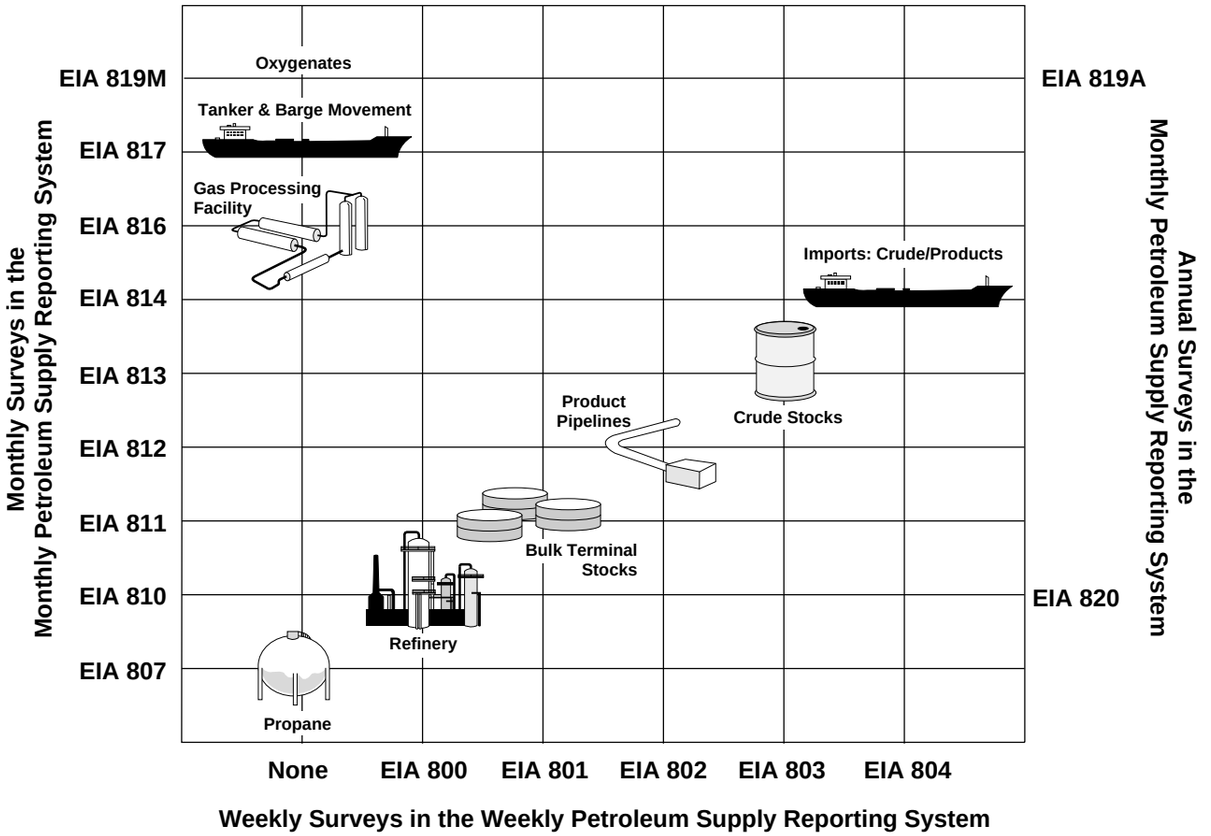
The PSRS is organized into two data collection subsystems, the Weekly Petroleum Supply Reporting System (WPSRS) and the Monthly Petroleum Supply Reporting System (MPSRS). The WPSRS processes data from the five weekly surveys. The MPSRS includes eight monthly surveys, two annual surveys, and the Form EIA-807, which collects propane data monthly from April through September and weekly from October through March. The two annual surveys are the Form EIA-819A, "Annual Oxygenate Capacity Report," used to collect data on current and projected production capacity of oxygenates and annual production and end-of-year inventories of fuel ethanol and the Form EIA-820, "Annual Refinery Report," used to collect data on refinery fuel use and consumption of steam and electricity, refinery receipts of crude oil by method of transportation, operable capacity for atmospheric crude oil distillation units and downstream units, as well as production capacity and storage capacity for petroleum products.

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.

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, weekly, and annual surveys of the PSRS. For example, the target population for the survey Forms EIA-801 and EIA-811 are bulk terminal stock holders. Thus, the respondents to the Form EIA-801 are a sample of the respondents who report to 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 to all published products. 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

Figure FE2. Petroleum Supply Reporting System: Surveys and Subsystems



Source: Energy Information Administration, Petroleum Supply Reporting System.

pipelines constitute at least 90 percent of the total volume of distillate fuel oil stocks as reported in the corresponding monthly surveys.

With these weekly surveys, EIA can provide timely, relatively accurate snapshots of the U.S. petroleum industry every week. The weekly surveys collect information on the supply and disposition of major 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, or EIA's electronic data communications 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 6 days of the close of the reference week, data are made available to the public in two forms: through the EIA electronic publishing system (EPUB) and hard copy (through the *WPSR*). Except when holidays delay data processing schedules, values for the weekly variables are available via EPUB at 9:00 a.m. on the Wednesday following the close of the reference week. The *WPSR* is distributed on the Thursday morning following the close of the reference week.

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 month.

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*. 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. Resubmissions for earlier months are published in Appendix C of the *PSM*, and are reflected in the *PSA*. Beginning with the February 1994 *PSM*, a new table (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 become available on the Wednesday following the first Friday of each month in the *WPSR* and EPUB as described above.

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 *PSA* also contains the results of the annual refinery capacity survey, Form EIA-820, and the annual oxygenate survey, Form EIA-819A. The values contained in the *PSA* are EIA's most accurate measures of petroleum industry 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 those 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 which would be obtained if the target population had been correctly 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 monthly data more carefully than they do their weekly data. Additionally, EIA has more time to edit the monthly data, and the weekly data come from a sample of the population. Thus, the monthly data are more accurate.

Some mechanisms introducing error, such as nonresponse, are not totally preventable. Other sources for errors, such as sampling errors, are unique to a particular type of survey. Sampling error occurs if the group of sampled respondents is dissimilar from the full population. Within the PSRS system, only weekly surveys, the monthly oxygenate survey, Form EIA-819M and the propane survey, Form EIA-807 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) internal processing errors such as incorrect data entry. 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 in order to reduce respondent burden and to expedite the turnaround of data from survey respondents to the public. As with any sample, the

Table FE1. Average Coverage for Weekly Surveys, 1994 and 1993 (Percent of Final Monthly Volumes Included in Monthly-from-Weekly Sample)

Product	Stocks						Production		Imports	
	Refinery		Bulk Terminal		Pipeline					
	1994	1993	1994	1993	1994	1993	1994	1993	1994	1993
Total Motor Gasoline.....	95	94	93	93	96	97	96	95	94	94
Jet Fuel	93	88	96	95	99	99	96	94	77	83
Distillate Fuel Oil	94	92	88	89	98	95	94	94	88	88
Residual Fuel Oil.....	91	86	92	93	--	--	90	89	97	97
Crude Oil.....	93	92	--	--	--	--	--	--	96	96

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

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 survey. EIA minimizes sampling error 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 *WPSR* survey frames. The weekly samples are updated when necessary to maintain the cutoff coverage.

For the weekly surveys, given better coverage, inaccurate data as a result of sampling error is less likely. As shown in Table FE1, 1994 coverage was comparable to 1993. Coverage was equal to or above 90 percent for 18 of the 21 product and supply type combinations in 1994. This is an improvement over 1993, where only 15 were over 90 percent. For 10 of the 21 combinations, 1994 coverage increased from 1993. There were two situations in which average coverage increased by 5 percentage points: coverage of jet fuel refinery stocks increased from 88 to 93 percent and coverage of residual fuel oil refinery stocks increased from 86 to 91 percent. However, there was one situation in which average 1994 coverage decreased by 6 percentage points: coverage of jet fuel imports dropped from 83 to 77 percent.

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 any 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. During the frames update, each frame is scrutinized to assure completeness.

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 1994 response rates for the weekly surveys and their complementary monthly surveys are enumerated in Table FE2. The 1994 average response rate for each of the EIA weekly surveys was over 96 percent. The corresponding monthly surveys had response rates of over 97 percent.

To mitigate the effect of nonresponse, imputed values are calculated for all unreported 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 survey.

Reducing Response Error

Over the past 5 years, many structural and procedural improvements to the PSRS system have been made in order to reduce the problem of nonsampling errors. One such improvement has been the PEDRO system, which permits all weekly and monthly survey data to be submitted to EIA electronically. A respondent entering values via PEDRO may

Table FE2. Average Response Rates for Monthly and Weekly Surveys, 1994

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	243	237	97.4	158	155	97.9
Bulk Terminal	332	331	99.6	82	80	96.8
Pipeline	85	85	99.6	47	46	98.5
Crude Oil Stocks	164	164	99.7	82	82	99.0

[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.
Source: Energy Information Administration, Petroleum Supply Reporting System.

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. Most monthly surveys are abstracted from accounting records 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 are of current and historical interest. Of the 59 variables for which both *PSM* and *PSA* values were published, only 44 of them were published weekly throughout 1994. 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 preliminary or interim value and the final value for a given month. For production, inputs, imports, and exports, values are expressed in units of thousands of barrels per day. For stocks, values are expressed in units of thousands of barrels.

  $$\text{MFW Error}_{var1} = \text{MFW Volume}_{var1} - \textit{PSA}\ \text{Volume}_{var1}$$

  $$\textit{PSM}\ \text{Error}_{var1} = \textit{PSM}\ \text{Volume}_{var1} - \textit{PSA}\ \text{Volume}_{var1}$$

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

  $$\text{MFW Percent Error}_{var1} = \frac{\text{MFW Error}_{var1}}{\textit{PSA}\ \text{Volume}_{var1}} \times 100$$

  $$\textit{PSM}\ \text{Percent Error}_{var1} = \frac{\textit{PSM}\ \text{Error}_{var1}}{\textit{PSA}\ \text{Volume}_{var1}} \times 100$$

- **Mean absolute error** is the 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. The mean absolute error is increased by an outlier (extreme value).

- **Mean absolute percent error** is the 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. An outlier makes the mean absolute percent error larger.

- **Range** of the percent errors 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 final absolute volumes and the mean absolute percent error for MFW estimates and *PSM* interim values for 1994 and 1993 are presented in Table FE3. The final absolute volumes are presented to give the reader an idea of the magnitude of the average 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 1994 MFW mean absolute percent errors which were within 2 percent of their respective *PSA* values (16 of the 44 MFW series), and the 1994 *PSM* mean absolute percent errors which were within 1 percent of their *PSA* values (38 of the 59 *PSM* series) are distinguished by a single asterisk. Mean absolute percent errors which were greater than 10.0 percent are marked by a double asterisk. There were 13 such MFW series and 3 *PSM* series.

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 interim estimates for the oxygenated motor gasoline stocks.

Oxygenated motor gasoline is a relatively new product (introduced in EIA's data series in January 1993), and there was still some confusion among respondents as to the definition. As a result, there were resubmissions, contributing to the large mean absolute percent error for this product in the stock series. This problem in reporting of oxygenated motor gasoline stocks did not affect the reporting of the aggregate. For total motor gasoline stocks, there is a single asterisk in both the MFW and *PSM* columns. In addition, reformulated gasoline was first reported on the weekly system for the week ending September 23, 1994; therefore, there is not a full year of data to analyze.

For distillate fuel oil, two relatively new subcategories are low sulfur (0.05 percent sulfur and under) and high sulfur (over 0.05 percent sulfur). For the first 3 months of 1994, respondents were still having trouble classifying distillate fuel oil into these subcategories. This did not affect the aggregate distillate values.

For 1994, 6 of the 11 production series have a single asterisk in the *PSM* column, indicating a mean absolute percent error of less than 1 percent from the *PSA*. Oxygenated motor gasoline production mean absolute percent error decreased from 19 to 9 percent from 1993 to 1994. In addition, the low sulfur distillate fuel oil production error decreased from 15 percent in 1993 to 1 percent in 1994.

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 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 (the arithmetic difference between the calculated supply and the calculated disposition 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 understatement of crude oil production, an understatment 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, the correctness of the direction of the change. Table FE4 provides a measure of accuracy of the direction of non-final stock change values for 1994 and 1993. In 1994, for each of the total stock change and refined products stock change categories, the direction of MFW stock change values was correct 100 percent of the time. In 1993, the direction of MFW stock change values was correct only 83 percent of the time. Except for 1 month in 1994, the direction of stock change of monthly interim values was correct. The 1 month for which the direction of crude stock change was inconsistent with the final direction was January 1994.

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.

EIA does not collect export data. They are gathered by the U.S. Customs Service on a monthly basis and are compiled by the U.S. Bureau of the Census. 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.

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 which follow. The box and whisker plots display

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 1994 and 1993

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	1994	1993	1994	1993	1994	1993
Crude Oil Production (thousand barrels/day)	6,662	6,847	* 0.51	0.40	* 0.65	0.32
Refinery Operations						
Refinery Crude Oil Inputs (thousand barrels/day)	13,866	13,613	* 0.34	0.47	* 0.05	0.21
Operating Utilization Rate (percent)	93	93	* 1.90	0.65	1.79	0.26
Production (thousand barrels/day)						
Total Production	17,774	17,776	** 10.88	11.54	* 0.09	0.49
Finished Motor Gasoline	7,312	7,360	* 0.89	1.05	* 0.17	0.20
Oxygenated Motor Gasoline	1,352	1,561	--	--	8.57	18.91
Other Motor Gasoline	5,617	5,799	--	--	1.36	4.26
Jet Fuel	1,448	1,422	2.37	2.10	* 0.25	0.08
Distillate Fuel Oil	3,205	3,132	2.11	1.65	* 0.12	0.38
Low Sulfur Distillate Fuel Oil	1823	898	--	--	1.01	14.84
High Sulfur Distillate Fuel Oil	1,382	2,233	--	--	1.46	3.50
Residual Fuel Oil	826	835	8.33	3.69	1.28	0.37
Other Products	4,984	5,028	** 39.03	40.63	* 0.17	1.58
Propane	969	963	--	--	* 0.25	0.66
Stocks (thousand barrels)						
Total Stocks	1,635,737	1,649,061	* 0.50	0.70	* 0.28	0.28
Total Stocks, excl. SPR	1,045,012	1,066,686	* 0.78	1.08	* 0.43	0.44
Total Crude Stocks	927,206	920,799	* 0.34	0.33	* 0.44	0.04
Crude Oil stocks, excl. SPR	336,481	338,424	* 0.92	0.94	1.22	0.10
SPR Stocks	590,725	582,375	* 0.02	0.02	* 0.00	0.00
Refined Products Stocks	708,530	728,262	* 1.08	1.82	* 0.12	0.63
Total Motor Gasoline Stocks	213,747	222,592	* 0.95	0.84	* 0.18	0.90
Oxygenated Motor Gasoline Stocks	11,111	18,976	--	--	** 34.31	10.88
Other Motor Gasoline Stocks	157,166	164,684	--	--	* 0.98	1.29
Jet Fuel Stocks	42,190	42,349	* 1.35	1.50	* 0.52	0.74
Distillate Fuel Oil Stocks	125,847	122,190	* 1.58	1.53	* 0.20	0.53
Low Sulfur Distillate Fuel Oil Stocks	62,241	31,061	--	--	* 0.96	12.39
High Sulfur Distillate Fuel Oil Stocks	63,606	91,129	--	--	* 0.95	2.32
Residual Fuel Oil Stocks	41,647	43,845	3.54	2.23	1.15	0.64
Other Products Stocks	285,100	297,295	2.64	3.18	* 0.29	1.12
Propane Stocks	44,342	44,429	2.86	3.64	* 0.38	1.12
Stock Change (thousand barrels/day)						
Total Stock Change	612	621	** 146.83	88.02	** 23.42	13.77
Crude Stock Change	174	248	** 101.88	129.74	** 44.00	123.90
Refined Products Stock Change	642	551	** 35.85	71.53	6.51	16.42
Imports (thousand barrels/day)						
Total Imports	8,996	8,620	2.36	3.89	* 0.79	1.19
Total Crude Imports	7,063	6,787	2.21	3.20	* 0.67	0.88
Crude Oil Imports, excl. SPR	7,051	6,772	2.25	3.20	* 0.67	0.88
SPR Imports	12	15	** 24.93	0.00	* 0.00	0.00
Refined Products Imports	1,933	1,833	9.59	6.67	1.53	2.37
Finished Motor Gasoline Imports	356	247	** 12.18	12.26	1.39	2.44
Oxygenated Motor Gasoline Imports	10	1	--	--	* 0.00	0.00
Other Motor Gasoline Imports	331	246	--	--	1.64	2.44
Jet Fuel Imports	117	100	** 19.01	20.02	2.17	7.56
Distillate Fuel Oil Imports	203	184	** 10.34	9.97	* 0.32	0.00
Low Sulfur Distillate Fuel Oil Imports	79	74	--	--	* 0.20	2.71
High Sulfur Distillate Fuel Oil Imports	124	110	--	--	1.13	0.00
Residual Fuel Oil Imports	314	373	** 11.41	9.79	3.39	2.67
Other Products Imports	943	929	** 10.13	7.88	1.81	2.52
Propane Imports	124	103	--	--	2.05	2.37

See footnotes at end of table.

Table FE3. Summary Statistics for Differences Between Interim and Final Data, 1994 and 1993 (Continued)

Variable	*PSA* Average Absolute Volumes		Monthly-from-Weekly Mean Absolute Percent Error		*PSM* Mean Absolute Percent Error	
	1994	1993	1994	1993	1994	1993
Exports (thousand barrels/day)						
Total Exports	942	1,003	5.64	11.47	* 0.00	5.94
Crude Oil Exports	99	98	** 31.82	44.71	* 0.00	0.00
Refined Products Exports	843	904	7.07	13.76	* 0.00	6.71
Total Net Imports (thousand barrels/day)	8,054	7,618	2.16	3.48	* 0.88	1.41
Products Supplied (thousand barrels/day)						
Total Products Supplied	17,718	17,237	* 1.18	1.91	* 0.25	0.63
Finished Motor Gasoline Supplied	7,601	7,476	* 1.27	1.52	* 0.32	0.37
Jet Fuel Supplied	1,527	1,469	* 1.69	4.03	* 0.25	0.93
Distillate Fuel Oil Supplied	3,162	3,041	5.04	5.94	* 0.28	0.67
Residual Fuel Oil Supplied	1,021	1,080	** 12.65	11.77	2.56	1.78
Other Products Supplied	4,407	4,170	4.21	5.17	* 0.55	1.26
Propane Supplied	1,082	1,006	--	--	* 0.66	1.46

-- = Not Applicable.

* = For MFW values, mean absolute percent error less than or equal to 2; for interim 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

Note: Error is the difference between Monthly-from-Weekly estimates or preliminary 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.

Source: Energy Information Administration, Petroleum Supply Reporting System.

historical accuracy of weekly estimates and monthly interim values for the last 5 years. The details provided by the box and whisker plots include: median error, 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 1990 through 1994) 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 a greater proportion of the MFW percent errors were positive (overestimate), the line approximately midway through the box (the median) would be above the x-axis.

Crude Oil Production and Crude Oil Inputs

Crude oil production data are not collected on any of EIA's surveys. EIA's Dallas Field Office assembles data collected

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

	Number of Months	
	1994	1993
Total Stock Change		
MFW and *PSA* Values	0	2
PSM and *PSA* Values	0	0
Crude Stock Change		
MFW and *PSA* Values	3	2
PSM and *PSA* Values	1	1
Refined Products Stock Change		
MFW and *PSA* Values	0	2
PSM and *PSA* Values	0	0

Note: MFW = Monthly-from-Weekly;
PSM = Petroleum Supply Monthly;
PSA = Petroleum Supply Annual

Source: Energy Information Administration, Petroleum Supply Reporting System.

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 beginning 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 plots provide 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., (5.5+4.5)/2.

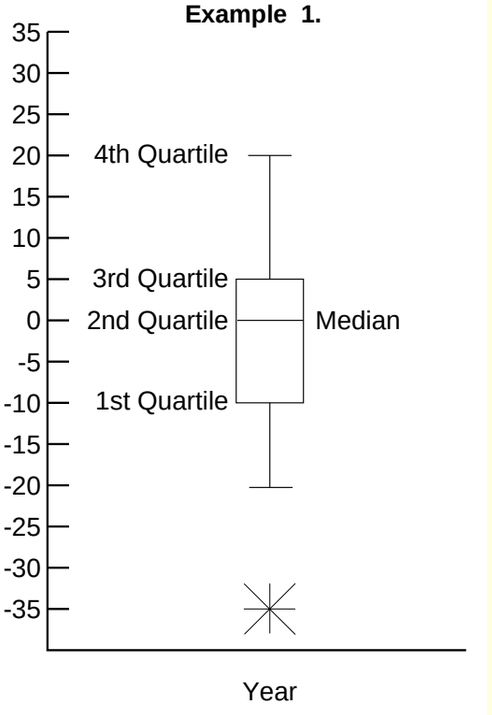
- **Box**

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.

Example 1.



- **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 1990)

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. Figure FE3 presents errors of MFW and *PSM* values relative to *PSA* values for crude oil production and inputs. Over the last 5 years, both MFW and *PSM* crude oil production values have been quite close to the *PSA* values. The 1994 MFW percent error distribution has values ranging from -1.36 to 1.14. Similarly, there is a tight distribution for the percent errors of crude oil *PSM* values, ranging between -1.30 and 0.42. The small percent errors of both MFW and *PSM* crude oil values demonstrate the consistency and precision of EIA's estimation procedures for weekly and monthly crude oil production.

For most months, the 1994 MFW values for refinery crude oil inputs were within 1 percent of the final values (Figure FE3). The MFW estimate for March (-1.38) was an outlier, which resulted from some shutdowns for maintenance that affected estimation. *PSM* refinery crude oil inputs have historically been extremely close to their final values. In 1994, for all months, the interim (*PSM)* values were within 0.1 percent of the final *PSA* values.

Product Production

As expected, *PSM* interim values for production of each of the four major petroleum products were superior to their comparable MFW preliminary 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 product production in Table FE3 and the box and whisker plots includes production from petroleum refineries, downstream motor gasoline blending facilities, and natural gas processing plants.

As illustrated in Figure FE4, MFW estimates of motor gasoline production have usually been underestimates of the final values. One month in 1994 (July) had the second largest percent error (-2.79) of the 60 months studied. Historically, the *PSM* interim motor gasoline production values have been excellent. In 1994, the distribution of percent errors ranged from -0.90 to 0.01 with one outlier (-0.90) for the month of May. This resulted from a revision in the *PSA* value of motor gasoline field production due to downstream blending of fuel ethanol.

For distillate fuel oil production, the median of the 1994 MFW estimates was close to 2 percent and MFW values were overestimates for all of the months except March. These overestimates were primarily the result of some respondents misclassifying residual fuel oil as distillate fuel oil on the weekly surveys. As in the box and whisker plots of motor gasoline production, those for distillate fuel oil production show that *PSM* interim values are close to final values. In 1994, for the *PSM* interim values, there was one outlier (0.63) for the month of July caused by a major resubmission. The remaining 11 months had a range of percent errors from -0.03 to 0.12. In 58 of the last 60 months, *PSM* distillate fuel oil production percent errors have been within 1 percent of the final values.

Figure FE5 shows the box and whisker plots for residual fuel oil production and jet fuel oil production. For all 12 months in 1994, the residual fuel oil production was understated by the MFW series, with a median error of -8.4 percent. These understatements resulted from some respondents misclassifying residual fuel oil as distillate fuel oil, in turn, overstating the distillate fuel oil as stated previously. The largest percent errors over the 60-month period were observed in 1994, -10.78 in July, and -12.34 in November. The 1994 distribution of *PSM* interim percent errors was not tightly clustered around zero as in previous years. The largest error over the 60-month period was -3.49, occurring in May 1994 which was due to a resubmission.

In 1994, the range of percent errors for the MFW and *PSM* estimates of jet fuel production was similar to the previous 4 years. The MFW estimates tend to overstate jet fuel production, with a median percent error of just over 2 percent. The *PSM* values had percent errors ranging from zero to 0.39.

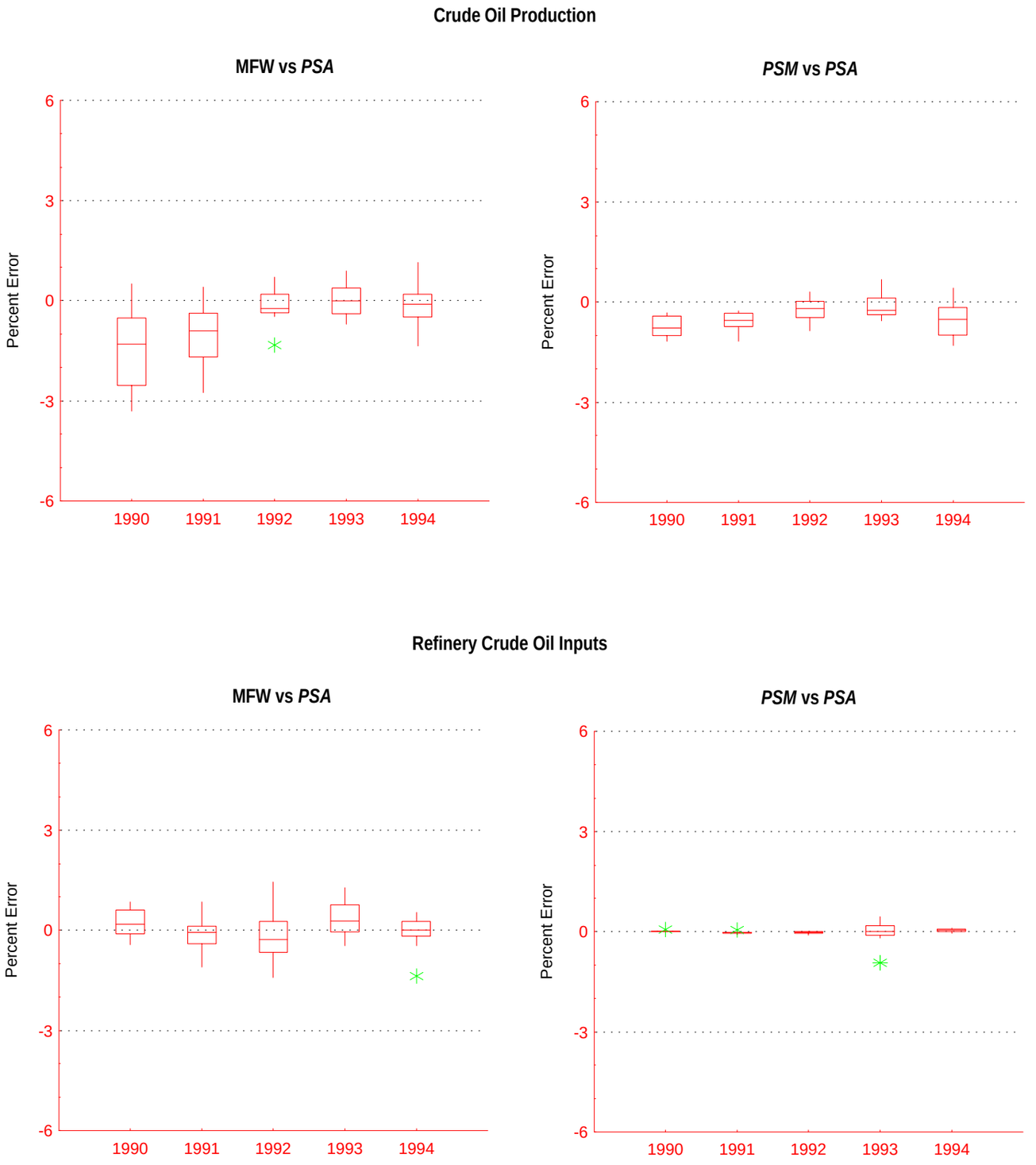
Stocks

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. The MFW percent errors for crude oil in 1994 differed from the previous 4 years. For the 1994 MFW estimates, 9 out of the 12 months understated the *PSA* values. For all of the months in 1994, the *PSM* values understated the *PSA* values. This is due to the reporting of volumes of Alaskan crude oil in transit on the *PSA* and not on the weekly and monthly surveys. In addition, a storage terminal was added to the crude oil survey late in 1994 that previously only reported stocks of petroleum products. The *PSM* interim estimate for June 1994 (-1.89) was the largest percent error over the past 60 months.

In 1994, the median percent error for MFW motor gasoline stocks was 0.36 percent, and the errors ranged from -2.05 percent to 1.57 percent. This is a change from 1993, when percent errors tended to be underestimates rather than overestimates. Similarly, the *PSM* percent errors were generally overestimates in 1994. They are tightly grouped around the median value which is 0.10 percent. There were 2 outliers in 1994, -0.37 in November and 0.58 in July, which were the result of resubmissions.

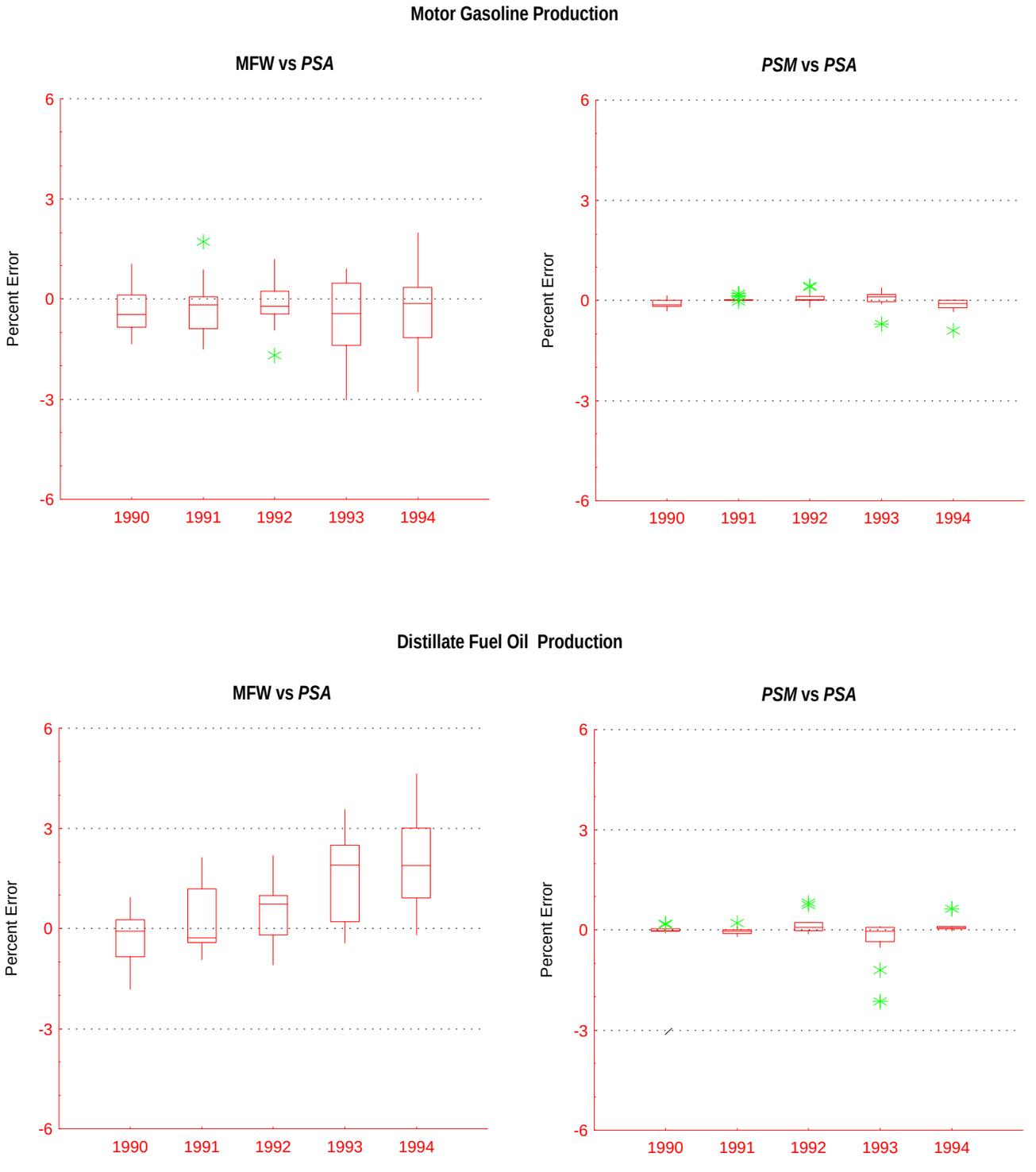
Figure FE7 shows box and whisker plots for distillate and residual fuel oil stocks. For distillate fuel oil stocks, the largest MFW percent error over the 60-month period was -4.77, in January 1994. The *PSM* interim estimates were clustered

Figure FE3. Range of Percent Errors for MFW and *PSM* Crude Oil Production and Refinery Crude Oil Inputs Data, 1990 - 1994



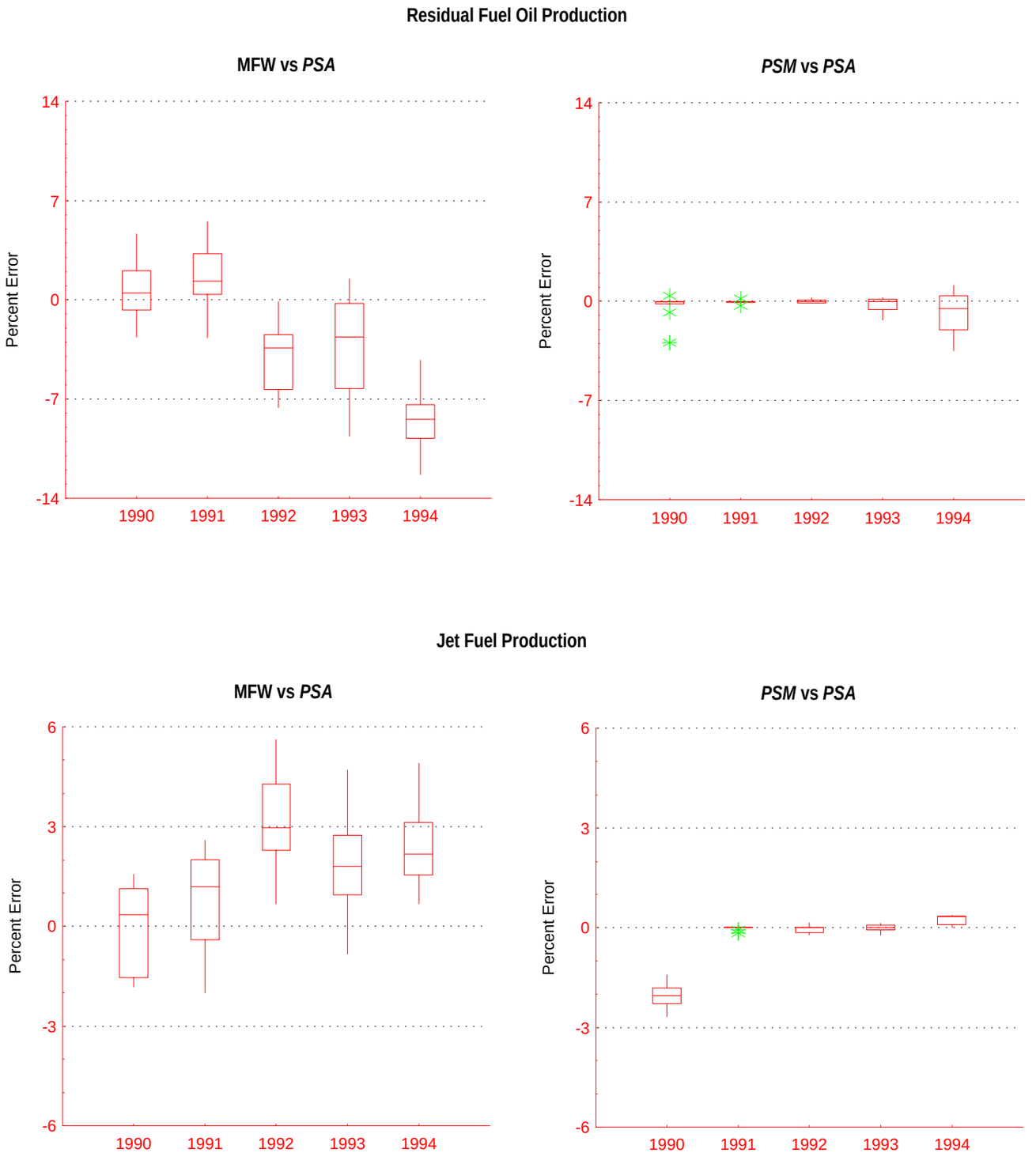
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE4. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Production Data, 1990 - 1994



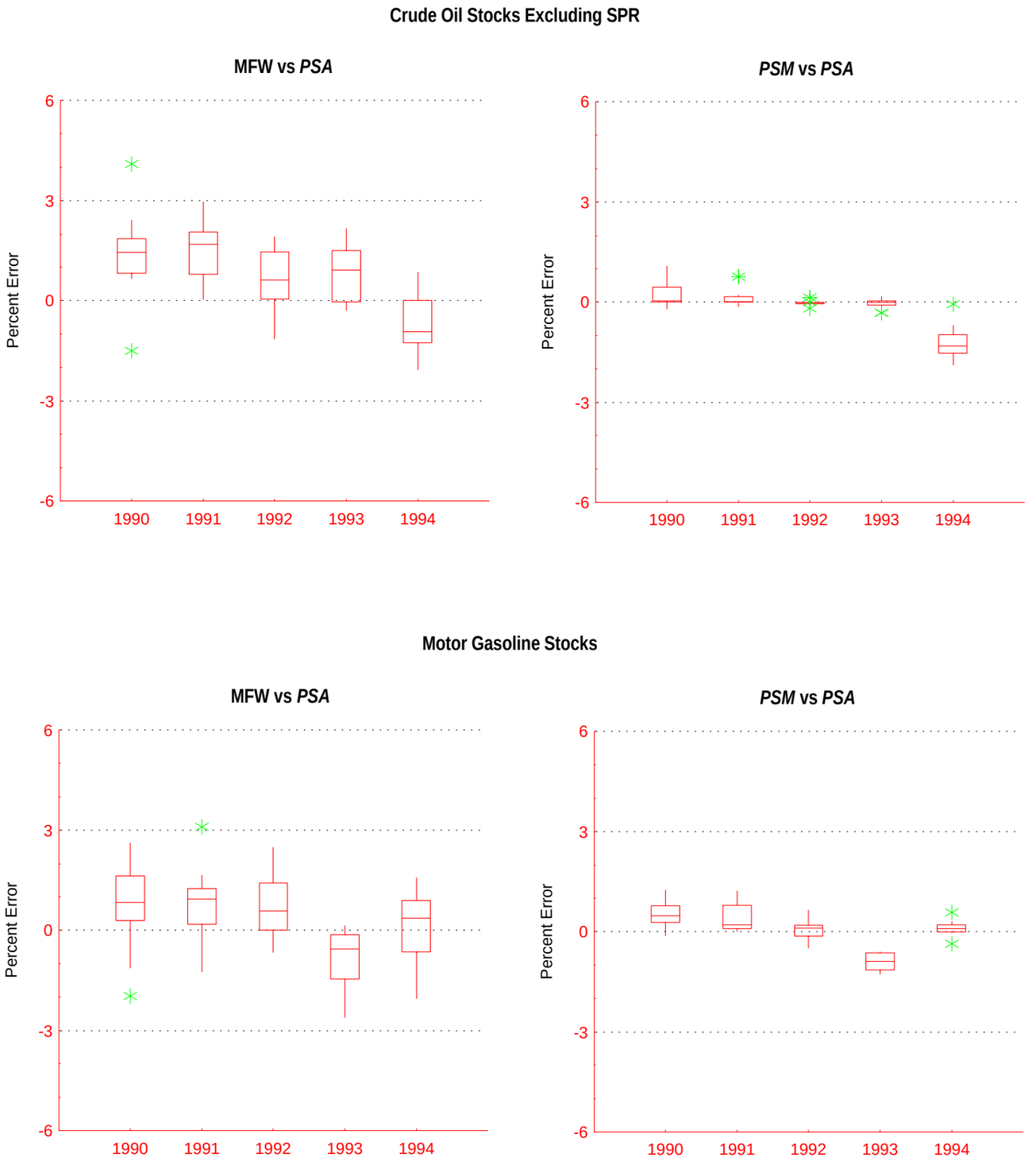
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE5. Range of Percent Errors for MFW and *PSM* Residual Fuel Oil and Jet Fuel Oil Production Data, 1990 -1994



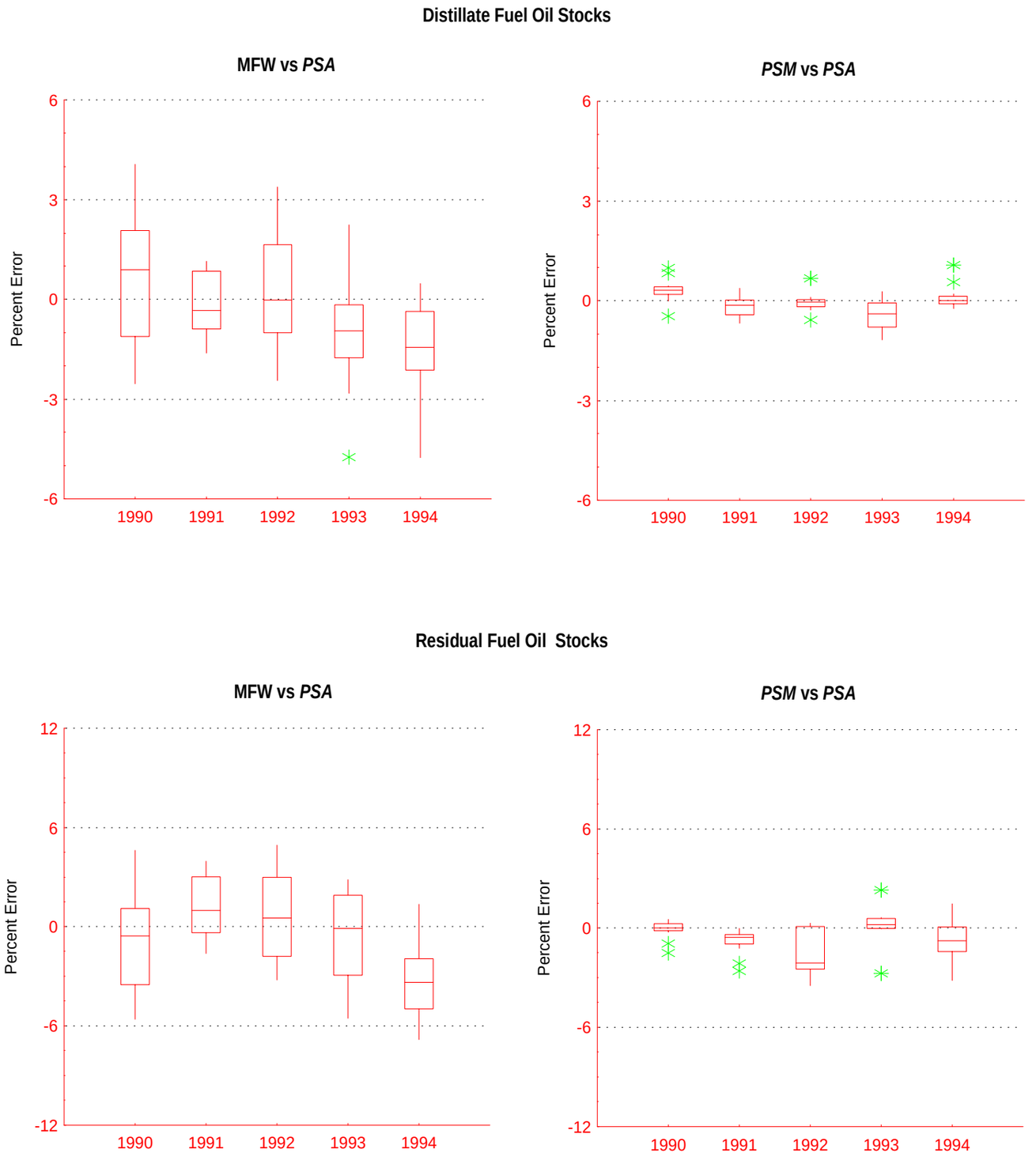
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE6. Range of Percent Errors for MFW and *PSM* Crude Oil Stocks Excluding SPR and Motor Gasoline Stocks Data, 1990 - 1994



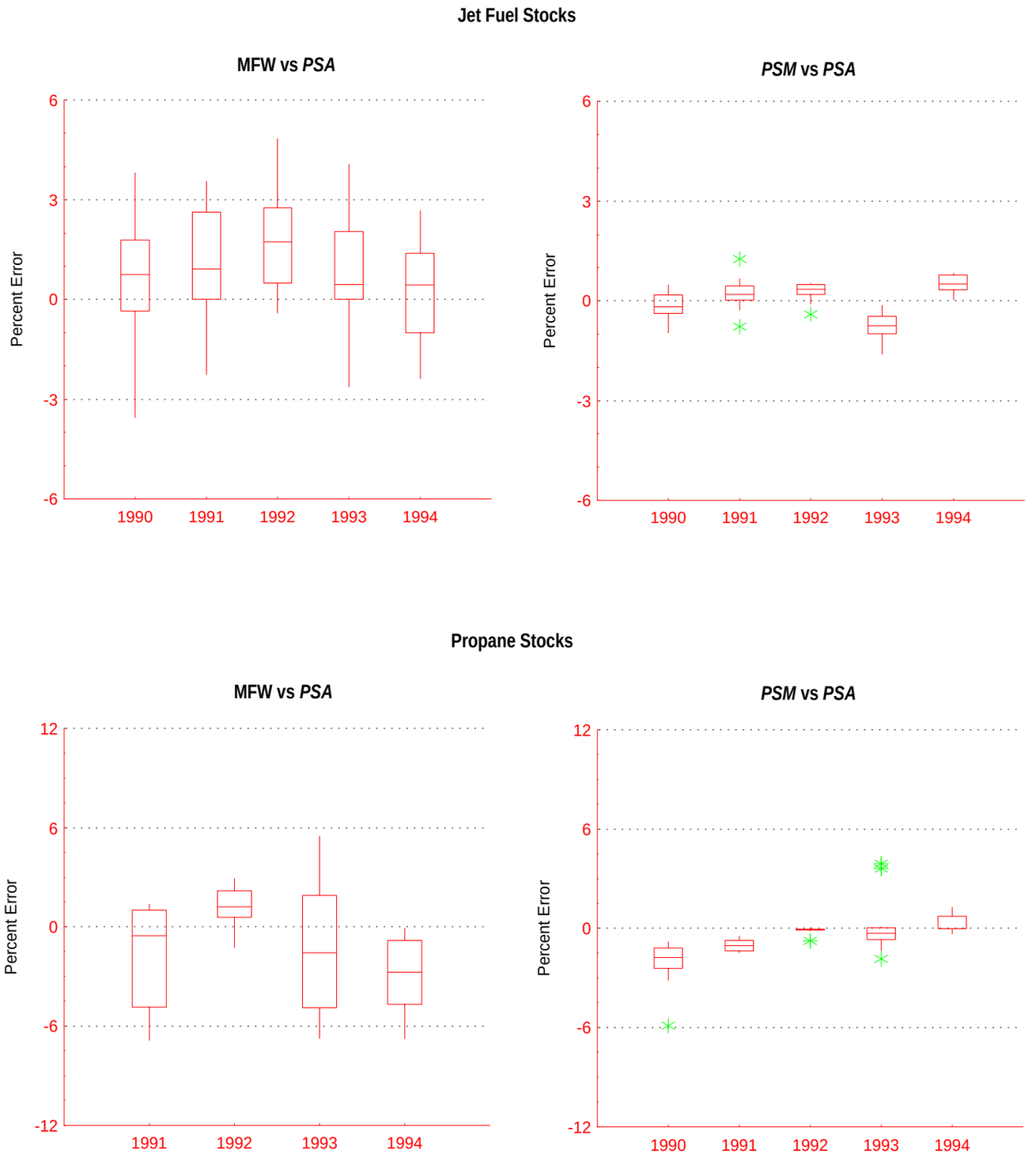
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE7. Range of Percent Errors for MFW and *PSM* Distillate Fuel Oil and Residual Oil Stocks Data, 1990 - 1994



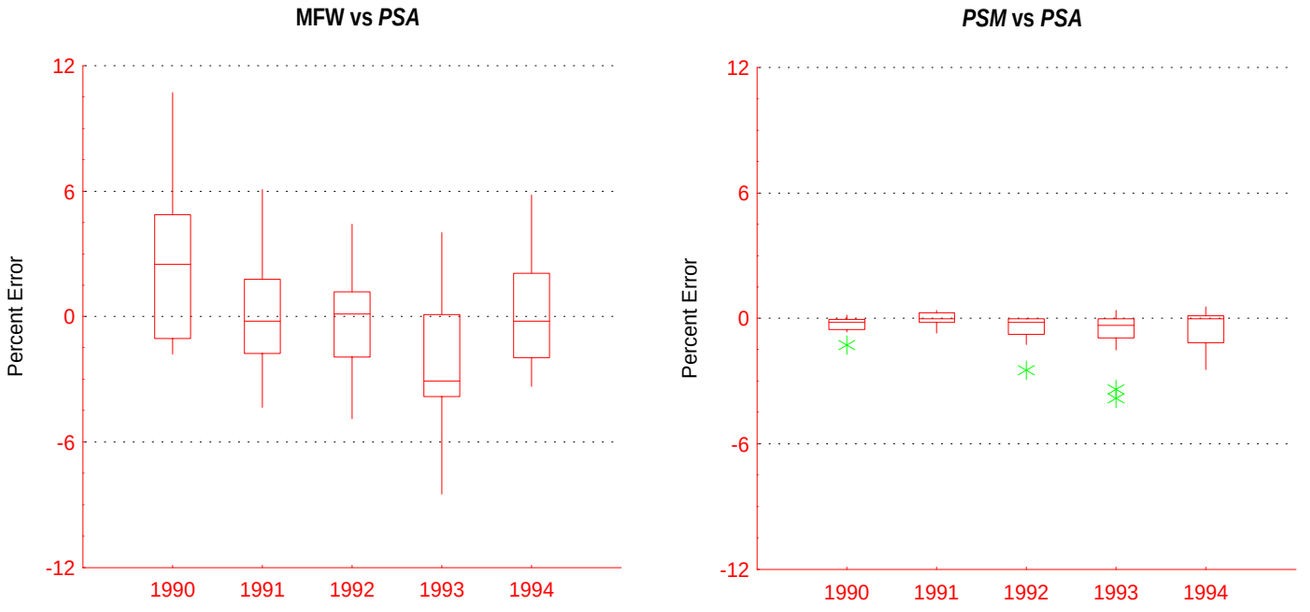
Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE8. Range of Percent Errors for MFW and *PSM* Jet Fuel Stocks and Propane Stocks Data, 1990- 1994



Source: Energy Information Administration, Petroleum Supply Reporting System.

Figure FE9. Range of Percent Errors for MFW and *PSM* Crude Oil Imports Excluding SPR Data, 1990 - 1994



Source: Energy Information Administration, Petroleum Supply Reporting System.

around zero percent, with two outliers in 1994, 0.56 in January and 1.07 in February.

Residual fuel oil typically has larger percent errors than other stock series. In 1994, the MFW percent errors ranged from -6.84 to 1.35. The two largest errors over the 60-month period occurred in July and September, -6.84 and -6.18, respectively. The *PSM* percent errors ranged from -3.16 to 1.45 percent.

The box and whisker plots for jet fuel stocks and propane stocks are shown in Figure FE8. In 1994, for jet fuel stocks, the median percent error for MFW estimates was the smallest in recent years (0.43 percent). The largest percent error was 2.67 percent in November. The *PSM* interim estimates were fairly tightly grouped about the median of 0.50 percent. In contrast with 1993, all *PSM* months in 1994 overestimated the *PSA*.

Propane stocks were first collected on the Form EIA-807 in 1991. For all of the months in 1994, the MFW percent errors underestimated the *PSA* with the largest error being -6.77 percent in February during the extreme weather conditions. The interim *PSM* estimates were tightly grouped around the median of 0.13, ranging from -0.34 to 1.27 percent.

Imports

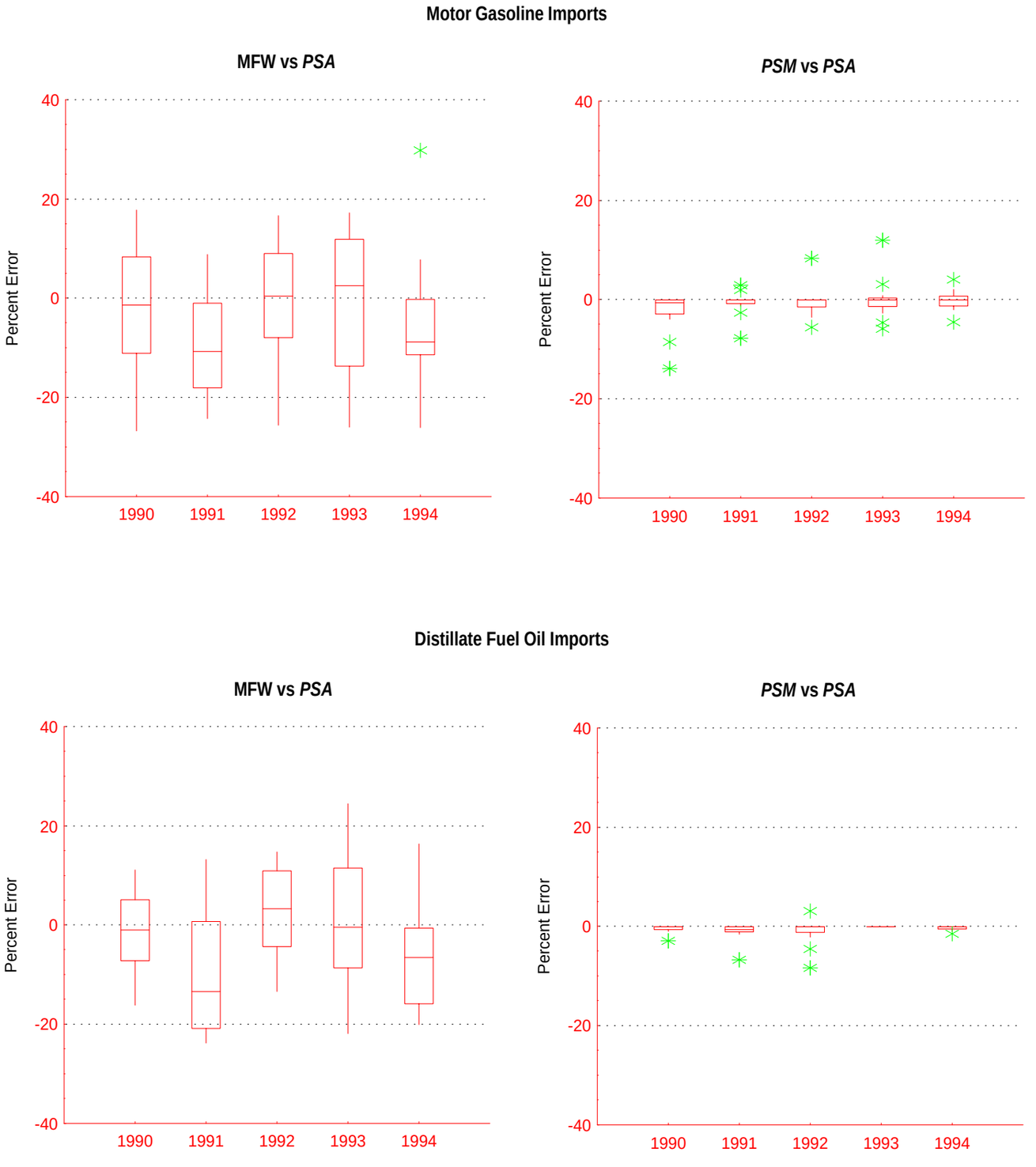
Figures FE9, FE10, and FE11 show the yearly distributions of percent errors for the imports of crude oil and four products: motor gasoline, distillate fuel oil, residual fuel oil, and jet fuel. Because of the irregularity of imports for crude oil and petroleum products, the magnitude and range of percent 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 1994 MFW values of imports of crude oil range from -3.34 to 5.79 with a median of -0.23. The *PSM* values have a similar median, zero percent, but have a tighter range, -2.43 to 0.53.

The distributions of percent errors of the MFW estimates and *PSM* interim values for 1990 through 1994 of motor gasoline and distillate fuel oil imports are shown in Figure FE10. The MFW estimates of motor gasoline imports tended to be underestimated in 1994, with a median of -8.82 percent error. There was one large outlier of 29.68 in November with the difficulty of product classification due to the Clean Air Act regulations. The median for the *PSM* percent errors for motor gasoline imports was zero percent, but there were two outliers, one underestimate and one overestimate (-4.57 and 4.04). The overestimate was due to misclassifying motor gasoline blending components as finished motor gasoline.

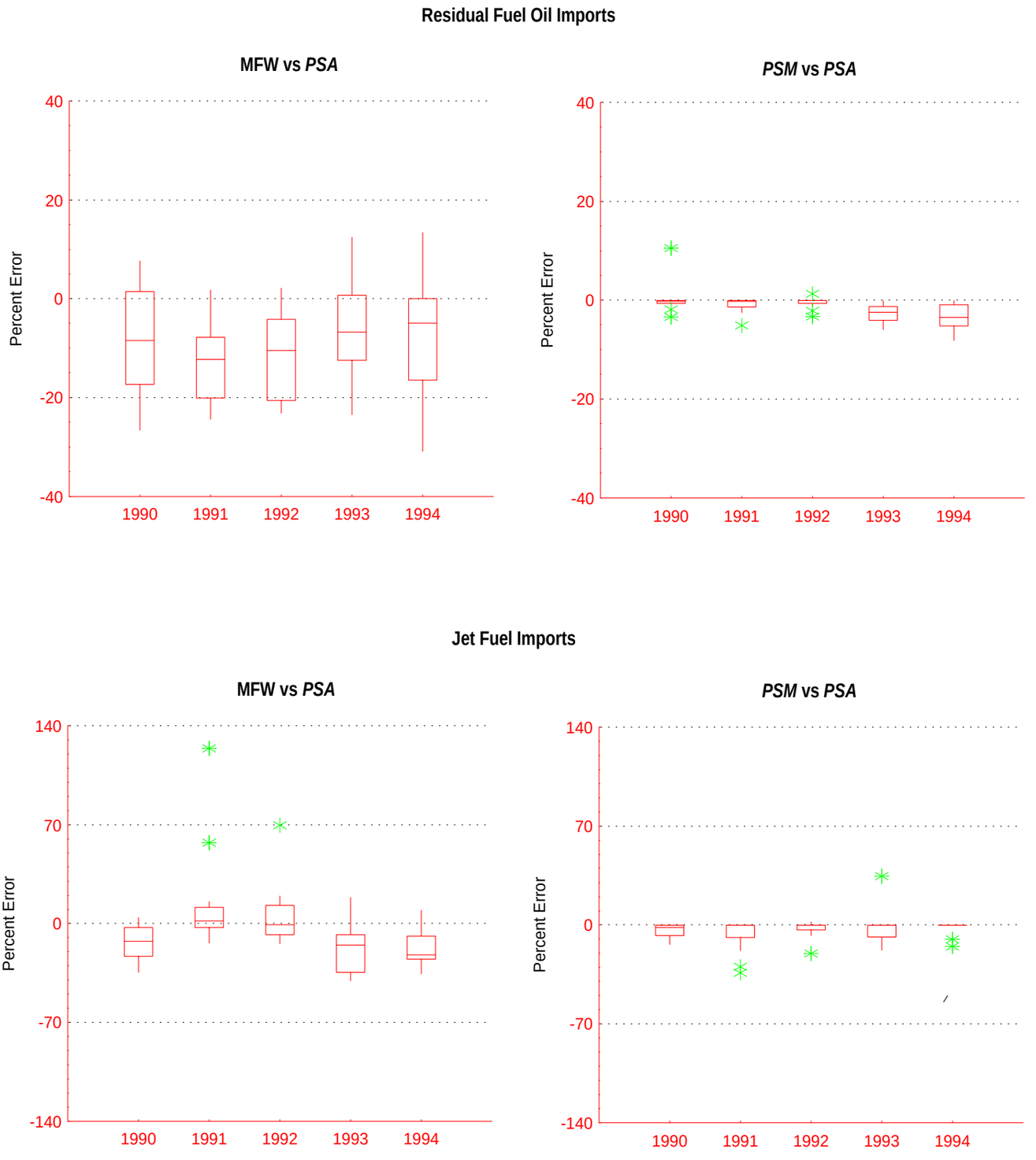
In 1994, February and November were the only months that the MFW estimates for distillate fuel oil imports were not underestimated. The range was from -20.13 to 16.30 percent error with a median of -6.57 percent error. In contrast to 1993, there were resubmissions for distillate fuel oil imports in 1994 resulting in understating of the *PSM*. The percent errors were

Figure FE10. Range of Percent Errors for MFW and *PSM* Motor Gasoline and Distillate Fuel Oil Imports Data, 1990 - 1994



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, 1990 - 1994



Source: Energy Information Administration, Petroleum Supply Reporting System.

tightly grouped around the median of zero with one outlier occurring in March (-1.57).

Figure FE11 shows the box and whisker plots for residual fuel oil imports and jet fuel imports. In 1994, the MFW estimates for residual fuel oil imports had the largest range in the 60-month period, ranging from -30.83 to 13.31 percent error. The median, -4.94 percent error, was also the highest in the 60-month period. Most of the percent errors for *PSM* estimates of residual fuel oil imports in 1994 were underestimated, with -8.10 in June being the lowest in the 60-month period.

As in previous years, the range of percent errors of MFW estimates of jet fuel imports was the widest of the import series, from -35.51 in February to 9.42 in April. July and August were the only months in 1994 that had resubmissions of jet fuel imports for *PSM* interim values, thus making them outliers, -10.20 and -15.38, respectively.

Conclusion

In summary, PSD data were on the right track in 1994. As illustrated in the box and whisker plots, 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 for both respondents to compile good numbers and EIA to process the data.

In 1994, 38 of 59 interim values were within 1 percent (mean absolute percent error) of the final values, 16 of 44 MFW estimates were within 2 percent (mean absolute percent error) of the final values, and 9 of those 16 were within 1 percent. As in previous years, the accuracy of 1994 preliminary and interim values varied by product and by petroleum supply type. As a group, stocks continued to have the most accurate preliminary MFW 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 1994, for 18 of the 21 categories, coverage was equal to or above 90 percent. The consistently high response rate, above 96 percent for weekly and 97 percent for monthly surveys, contributes to the high level of accuracy of these data.

In order to continuously maintain and improve the accuracy of these data, PSD is in the process of re-engineering its business functions. The four major functions include identify/design, collect, analyze, and disseminate. Techniques such as graphical data validation should provide enhanced editing procedures. Within 2 years, the outcome of these efforts should result in more timely and accurate petroleum supply data.