Understanding and Evaluating WAOB/USDA Soybean Yield Forecasts

Scott Irwin and **Darrel Good**

Department of Agricultural and Consumer Economics
University of Illinois

Dwight Sanders

Department of Agribusiness Economics
Southern Illinois University-Carbondale

In the *farmdoc daily* article of April 30, 2015, we examined the methodology and evaluated the accuracy of USDA's World Agricultural Outlook Board (WAOB) corn yield forecasts over the period 1993-2014. That analysis was provided for background as the 2015 WAOB forecast cycle begins with the WASDE report to be released on May 12. Here, we repeat that analysis for soybeans. As with corn, WAOB soybean yield forecasts have been used in the May, June, and July WASDE reports since 1993 to make supply, ending stocks, and price projections for the upcoming marketing year. Previous research has shown that the release of these early projections—particularly in May—has a significant impact on prices in the soybean futures market (Isengildina-Massa et al., 2008). While it is clear that the WAOB yield forecasts are perceived by market participants as containing important new information, these forecasts appear to be poorly understood by many and often confused with later forecasts released by the National Agricultural Statistics Service (NASS) of the USDA. In this article, we demonstrate how the WAOB methodology for soybeans has changed over time and evaluate the historical accuracy of the soybean yield forecasts. The analysis of forecast accuracy updates results published in an earlier research report on USDA forecasts and estimates (Irwin, Sanders, and Good, 2014). That evaluation was for the period 1993 through 2012. Here we extend the analysis for soybeans through the 2014 crop year.

Changes in WAOB Forecasting Procedures

WAOB forecasts of the national soybean yield are based on relatively simple trend analysis of historical yields. For the past two years, the trend analysis has been modified to include weather considerations. The procedure for the May projections, as presented in the May WASDE report each year over 1993-2014, is listed in Table 1. Most years reflect a relatively simple approach of trend regressions, either on a regional basis (1993-1996 and 2002-2009) or national basis (1997-2001 and 2010-2012). The 2013 trend analysis was supplemented with growing season weather considerations and the 2014 analysis made use of a more sophisticated crop weather model. There have also been changes to the sample period used in the estimation methodology. For 1993, trend estimates were based on the sample period from 1972-1992.

The sample period was changed to 1974-1993 in 1994 and expanded to 1974-1994 in 1995. For 1997-2001 the sample period was "since the mid-1980s." For 2002-2006, a switch was made to samples beginning in 1978 and for 2007-2012 a switch was made to samples beginning in 1989. The previous year's yield was omitted in the trend analysis for 2004, 2009, and 2012. Then, in 2013 the sample period was switched to 1988-2012.

Table 1. May WASDE Description of Soybean Yield Projection, 1993-2014	
Year	Description
1993	Projected Yield Based on 1972-92 regional trends, weighted by acres
1994	Projected Yield Based on 1974-93 regional trends, weighted by acres
1995	Projected yield based on 1974-1994 regional trends, weighted by acres
1996	Projected yield based on aggregated 1974-95 U.S.regional trends
1997	Projected yield based on U.S. trends since the mid-1980's
1998	Projected yield based on U.S. trends since the mid-1980's
1999	Projected yield based on U.S. trends since the mid-1980's
2000	Projected yield based on U.S. trends since the mid-1980's
2001	Projected yield based on U.S. trends since the mid-1980's
2002	Projected yield based on 1978-2001 regional trend analysis
2003	Projected yield based on 1978-2002 regional trend analysis
2004	Projected yield based on 1978-2002 regional trend analysis
2005	Projected yield based on 1978-2004 regional trend analysis
2006	Projected yield based on 1978-2005 regional trend analysis
2007	Projected yield based on 1989-2006 regional trend analysis
2008	Projected yield based on 1989-2007 regional trend analysis
2009	Projected yield based on 1989-2007 regional trend analysis
2010	Projected yield based on 1989-2009 trend analysis
2011	Projected yield based on 1989-2010 trend analysis
2012	Projected yields based on 1989-2010 trend analysis
2013	Projected yields based on 1988-2012 trend analysis adjusted for weather during the growing season
2014	The projected yield is based on a weather adjusted trend model and assumes normal summer weather
Note: The description for each year is exactly as listed in the footnotes to the U.S. soybeans and products balance table in the May WASDE report for that year.	

While some change in procedures over time is to be expected, similar to corn, WAOB soybean yield forecasting procedures have been surprisingly variable. The interesting empirical question is whether the

changes had much impact on the WAOB yield forecasts. We present evidence on this point in Figure 1, which shows the history of May U.S. soybean yield forecasts from the WAOB over 1993-2014. The dominance of using a linear trend of historical yields for making the projection for the current year is reflected in the generally steady increase in the yield projection over the entire time period. However, there are three distinct "steady state" periods in the yield forecasts. From 1994 through 1998, May yield forecasts increased about one bushel per year, but were in a very flat pattern from 1998 through 2005 when forecasts ranged from 39.5 to 40 bushels. For the period 2006 through 2014, the yield forecasts increased by an average of 0.52 bushels per year, with a linear trend explaining 99 percent of the annual variation in the forecasts. It is also interesting to contrast the relatively smooth pattern in May soybean yield forecasts over 2010-2014 with the variability in May corn yield forecasts over the same period. The May 2010 WAOB forecast of corn yield, 165.3 bushels, jumped 8.1 bushels over the 2009 forecast and this variability continued through 2014 (*farmdoc daily*, April 30, 2015).

Figure 1. WAOB/USDA May U.S. Average Soybean Yield Forecasts, 1993-2014
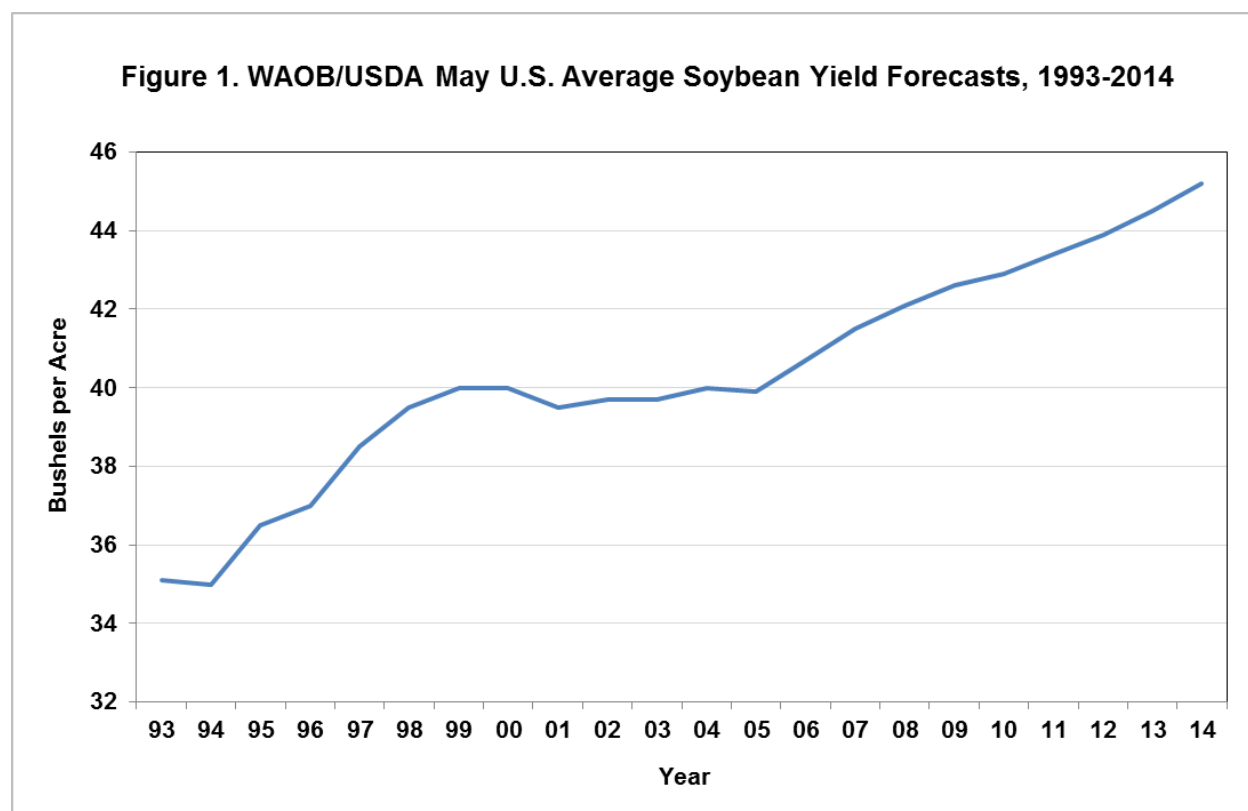
The changes through time in WAOB soybean yield forecasts for the U.S are not necessarily problematic if the methodology used to generate the forecasts is transparent and easily replicable. It would seem that a methodology limited to fitting a linear trend over average yields in a specific previous period should be transparent and easily replicated. Yet, the switching from regional to national trend analysis, the ambiguity of the sample period used for the analysis from 1997 through 2001, and the application of a weather model make forecasts less than fully transparent and more difficult to replicate. We do note that the WAOB published a report in 2013 that presented the crop weather model used to generate forecasts in 2014 (Westcott and Jewison, 2013). However, the WAOB has not published the data used to estimate the crop weather models nor the exact steps detailing how published forecasts are generated.

WAOB Forecast Accuracy

Since WAOB soybean yield forecasts until very recently have been based entirely on trend yields, expectations for the accuracy of those projections should be influenced by the historical relationship

between actual yields and trend yields. The findings in a recent *farmdoc daily* article (March 19, 2015) on soybean trend yield projections suggests the following expectations for WAOB forecasts relative to actual yields from 1993 through 2014: 1) the average difference between projected and actual yields will be close to zero (projections are unbiased), 2) the average yield will be under-estimated more often than over-estimated, 3) there will be years with large deviations between projected and actual yields, 4) the largest deviations between projected and actual yields will be in years when actual yields are well below trend, and 5) average absolute forecast errors should be relatively small since trend explains a high percentage of variation in annual yields.

To evaluate the historic accuracy of the WAOB soybean yield forecasts, the May, June, and July forecasts are compared to the "final" yield estimate released in January after harvest (we say "final" because January estimates are sometimes revised a year later or even later based on the *Agricultural Census* conducted every five years). The differences between the forecasts and the final estimates in percentages over 1993-2014 are presented in Figures 2 through 4. When interpreting the errors as indicated by the bars, note that a positive error implies an under-estimate on the part of WAOB and a negative error implies an over-estimate. The absolute size of the errors is also represented by the lines in each graph. As expected, forecast errors were occasionally very large and the magnitude of the most extreme over-estimate exceeded the magnitude of the most extreme under-estimate. Unexpectedly, the frequency of over-estimates exceeded the frequency of under-estimates in all three months. The two largest forecast errors occurred in 1994 and 2003 and are readily explained by unusually favorable growing conditions in 1994 and a widespread outbreak of soybean aphids in 2003.
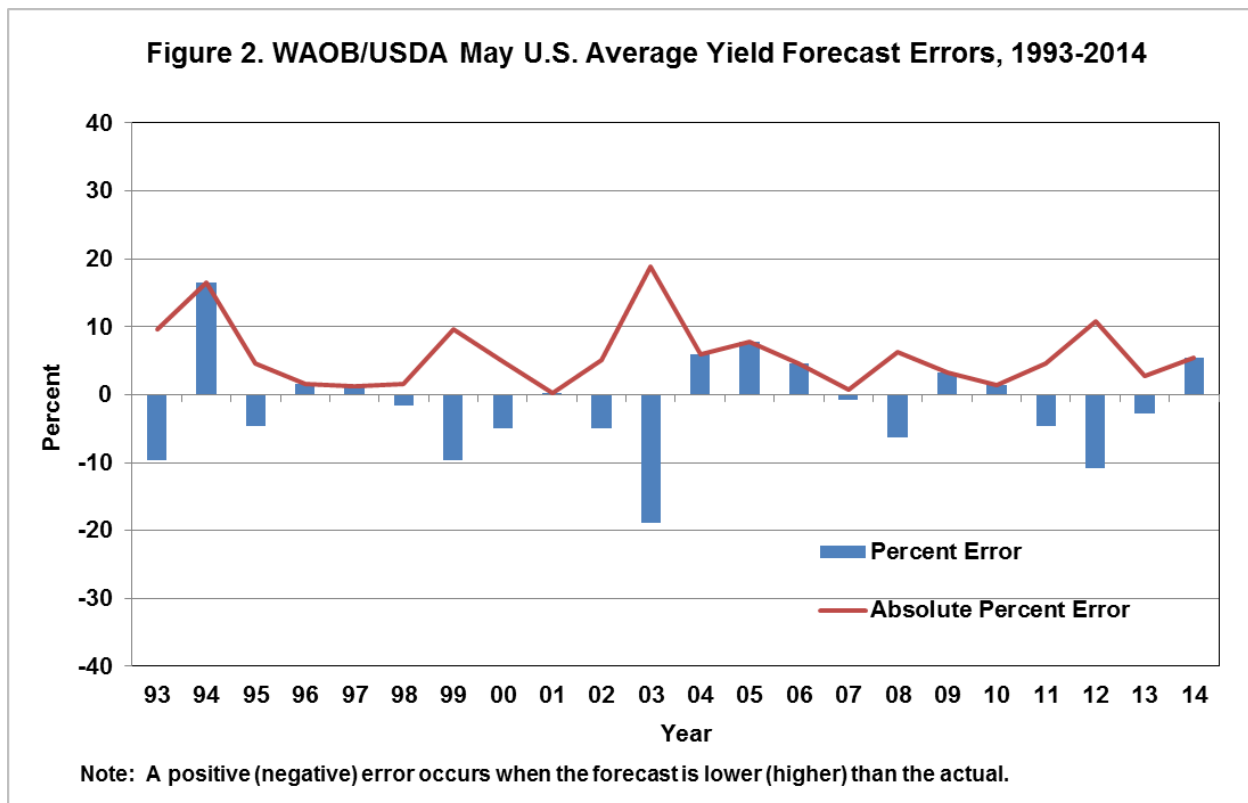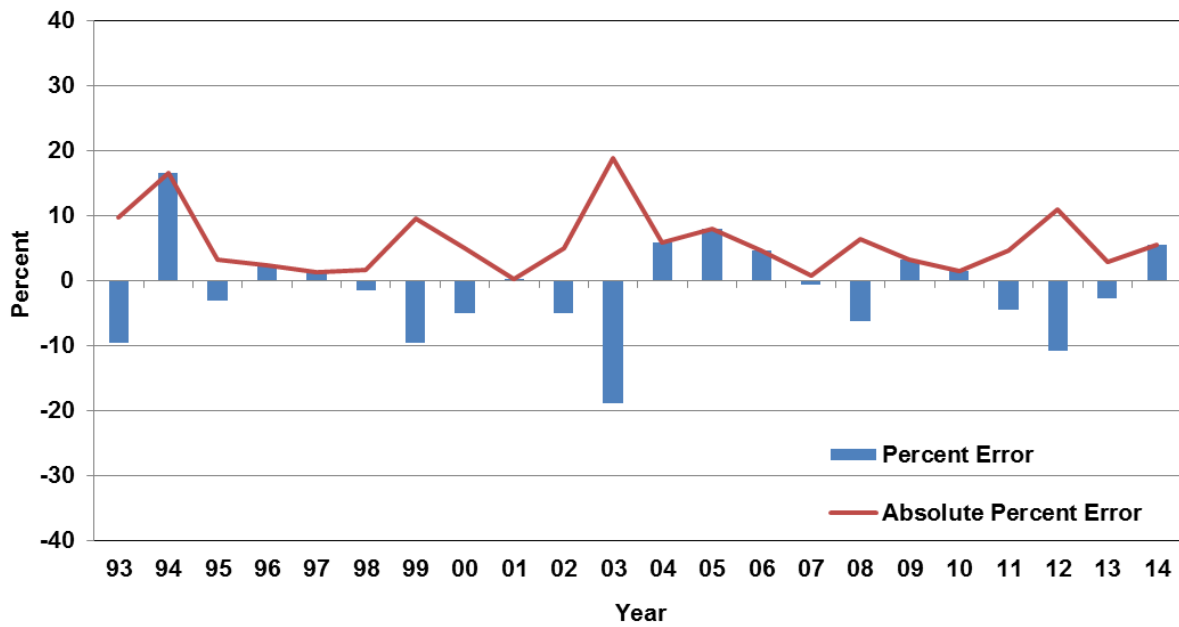

Figure 2. WAOB/USDA May U.S. Average Yield Forecast Errors, 1993-2014

Note: A positive (negative) error occurs when the forecast is lower (higher) than the actual.

Figure 3. WAOB/USDA June U.S. Average Soybean Yield Forecast Errors, 1993-2014

Note: A positive (negative) error occurs when the forecast is lower (higher) than the actual.
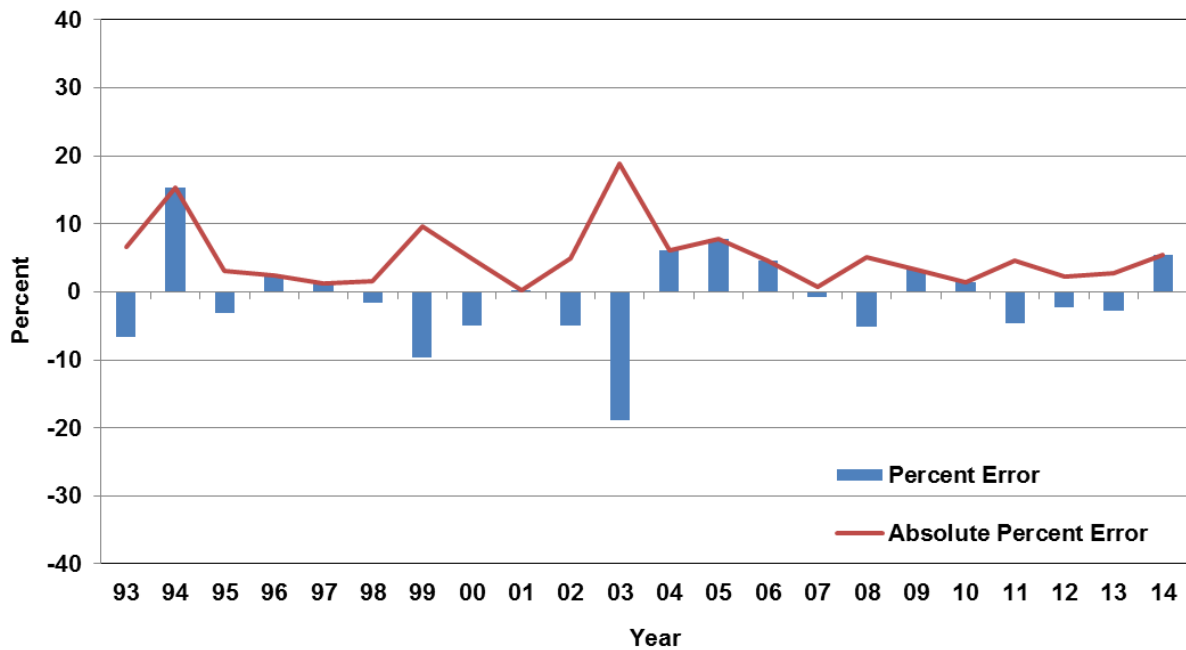

Figure 4. WAOB/USDA July U.S. Average Soybean Yield Forecast Errors, 1993-2014

Note: A positive (negative) error occurs when the forecast is lower (higher) than the actual.

The WAOB soybean yield forecasts are also examined for both bias and changes in accuracy through time. The average percent errors for each forecast month for the entire sample period and for sub-samples formed from 1993-2003 and 2004-2014 are presented in Table 2. The error calculations presented in Table 2 reveal that average errors for the earlier sub-sample and for the entire period were negative (that is on average WAOB forecasts were too high) for each forecast month. The average errors, however, were progressively less negative in June and July. For the latter sub-sample, average errors were positive for each forecast month. Surprisingly the average error in that sub-sample was larger in July than in May and June. The average errors ranged from -3.15 percent in May in the 1993-2003 period to 0.29 percent in May and June in the 2004-2014 period. For the entire period, the average errors ranged from -1.43 percent in May to -0.78 percent in July. The bias in all three months in the first sub-sample is equivalent to about -1 bushel per month, certainly a non-negligible amount. While the magnitude and direction of the average errors varied by month and sub-sample, the average differences were not statistically different from zero. As a result, no statistically significant bias in estimates was found across months or sub-samples. These results confirm the expectations that WAOB soybean yield forecasts are statistically unbiased and, with the exception of the first half of the sample, average errors are relatively small.

Table 2. Average Percent Errors for WAOB/USDA U.S. Average Soybean Yield Forecasts, 1993-2014

Marketing Years	May	June	July
1993-2003	-3.15	-2.95	-2.77
t-statistic	-1.18	-1.10	-1.09
p-value	0.26	0.30	0.30
2004-2014	0.29	0.29	1.20
t-statistic	0.16	0.16	0.88
p-value	0.87	0.87	0.40
1993-2014	-1.43	-1.33	-0.78
t-statistic	-0.89	-0.83	-0.53
p-value	0.38	0.42	0.60

Note: t-statistic tests for difference from zero.

Table 3 presents the average absolute percent errors in the soybean yield forecasts for each month and sample period. As expected, the average absolute errors declines modestly as the season progresses (absolute average errors were equal in May and June in the latter sub-sample). For the entire sample period, for example, the absolute error averaged 5.80 percent for the May forecast and 5.14 percent for the July forecast. Examining the two sub-samples, the absolute percent errors were smaller in 2004-2014 for each forecast month. The decreases in the average absolute forecast errors in the latter sub-sample, however, are not statistically significant. Changes in absolute average errors through time are also examined by regressing the absolute percent errors against a constant and a linear time trend:

$$|WAOB\ Percent\ Error_{m,t}| = \alpha + \beta Trend_t + e_t$$

where $Trend_t$ is a time trend variable for crop year t that takes a value of 1 in 1993, 2 in 1994, and so on and e_t is a standard, normal error term. The estimated trend coefficients are negative for all three forecast months, which is consistent with the smaller average errors in the latter sub-sample. The trend coefficients,

however, are not statistically different from zero. So, while the errors have gotten smaller, the decline may just be due to chance and random variation within the sample.

Table 3. Average Absolute Percent Errors for WAOB/USDA U.S. Average Soybean Yield Forecasts, 1993-2014

Marketing Years	May	June	July
1993-2014	5.80	5.77	5.14
1993-2003	6.72	6.66	6.27
2004-2014	4.88	4.88	4.01
t-statistic	-0.88	-0.85	-1.17
p-value	0.86	0.63	0.65
1993-2014 Trend	-0.12	-0.12	-0.16
t-statistic	-0.75	-0.71	-1.02
p-value	0.46	0.49	0.32

Note: The first t-statistic tests for difference in means between samples and the second t-statistic tests whether the trend is different from zero.

Comparing WAOB and NASS Forecast Accuracy

As a final step in the analysis, we compare the historic accuracy of WAOB and NASS soybean yield forecasts over the entire May through November forecasting cycle. This is an interesting exercise because WAOB and NASS soybean yield forecasts are based on entirely different procedures. As noted above, May, June, and July WAOB forecasts of national soybean yield are based on relatively simple trend analysis of historical yields, sometimes modified by planting progress and/or weather conditions. NASS forecasts of state and national soybean yield are based on large-scale farmer surveys and field measurement surveys. These forecasts are released in monthly crop production reports from August through November each year, with the final yield estimates released in January after harvest. The NASS soybean yield forecasts are released simultaneously with WASDE reports and the forecasts are used without adjustment in WASDE supply projections. The procedures used in making NASS forecasts and evaluations of the accuracy of those forecasts have been presented in earlier research reports (Good and Irwin, 2011; Irwin, Sanders, and Good, 2014) and *farmdoc daily* articles (August 19, 2011; September 1, 2011; August 28, 2013; September 4, 2014).

WAOB and NASS soybean yield forecast errors are summarized in Figure 5 using what is known as a "box-and-whisker" plot. For each month, the distance from the top of the upper whisker to the bottom of the lower whisker captures the entire range of forecasting errors over 1993-2014. The upper whisker reflects the range of forecasting errors for the largest 25 percent of under-estimates of yield and the lower whisker reflects the range of forecasting errors for the largest 25 percent of over-estimates of yield. The box captures the range of errors for the middle 50 percent of the errors. In May, for example, the WAOB forecasting errors for the U.S. average soybean yield ranged from an under-estimate of 15.5 percent to an over-estimate of 17.1 percent. The middle 50 percent of the errors ranged from an under-estimate of 3.2 percent to an over-estimate of 5.0 percent. It should come as no surprise that the range of USDA soybean forecasting errors gets progressively smaller after July. The closer that one moves toward harvest the better is information on crop development and yield prospects. That information should be captured in the NASS forecast methodology beginning in August.

Figure 5. USDA Errors in Forecasting U.S. Average Soybean Yield, May-November, 1993-2014
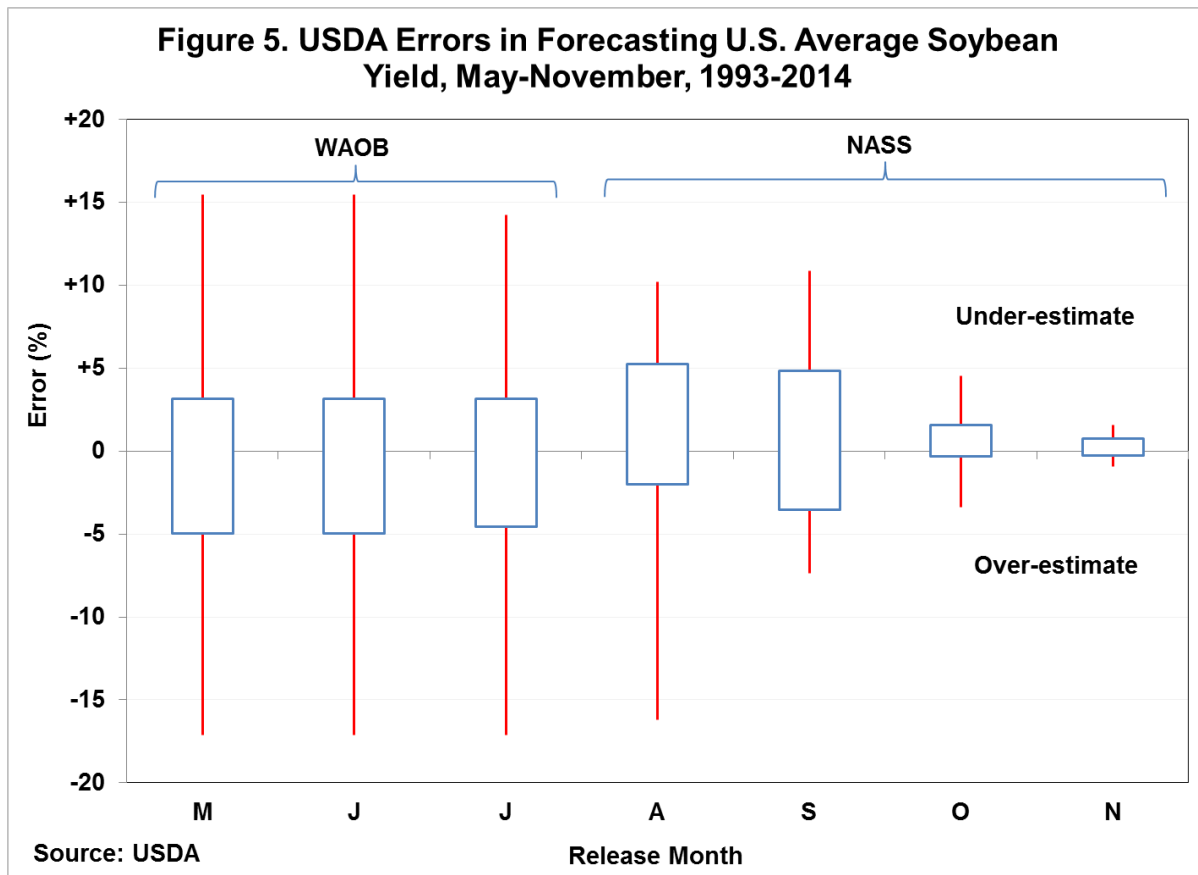
Several interesting observations can be made based on Figure 5. First, the distribution of errors is remarkably similar for May, June, and July. This reflects the fact that in most years over 1993-2014 only minimal adjustments were made to WAOB forecasts after May. Second, the boxes reveal that WAOB soybean yield forecasts tended towards over-estimation (consistent with the results presented earlier in Table 2), while NASS forecasts tended towards under-estimation, just the opposite directional bias. The downward bias in NASS forecasts is largest in August. We reported in this earlier *farmdoc daily* article (September 4, 2014) that the downward bias in NASS soybean yield forecasts over 1990-2013 was statistically significant in September, October, and November and was larger in the second half of the sample compared to the first half in all four release months. The magnitude of the bias in August and September over 2002-2013 was on the order of one bushel, again, a non-negligible amount. Third, the magnitude of the extreme over-estimate (bottom whisker) by the NASS August forecast was not smaller than the extreme over-estimate by the WAOB July forecast. Both extremes occurred in 2003. Fourth, the magnitude of the middle 50 percent of the range of WAOB and NASS forecast errors is almost equal from May through September. In fact, the middle 50 percent range for the September NASS forecast is slightly larger than the range for the other four months. This is surprising given the more advanced stage of the soybean crop.

Implications

The USDA's World Agricultural Outlook Board (WAOB) provides one of the key early assessments of prospects for the U.S. average soybean yield. WAOB forecasts of the national soybean yield are released in May, June, and July and are based on relatively simple trend analysis of historical yields (trend modified by weather considerations in 2013 and 2014). Despite the simplicity, WAOB soybean yield forecasting procedures have been surprisingly variable over time. The changing procedures generated three

distinctive periods of yield forecasts—relatively steep annual increases from 1994 to 1998, a very flat pattern from 1998 through 2005, and modest annual increases from 2006 forward. Our analysis of WAOB May, June, and July yield forecasts for soybeans over 1993-2014 showed that forecast errors were occasionally very large and the magnitude of the one extreme over-estimate exceeded the magnitude of the one extreme under-estimate. Surprisingly, the frequency of over-estimates exceeded the frequency of under-estimates in all three months. The two largest forecast errors occurred in 1994 and 2003 and are readily explained by the unusually favorable growing conditions in 1994 and a widespread outbreak of soybean aphids in 2003. Forecast error trend results suggest that observed improvements in forecast accuracy over time are not statistically significant.

In sum, while there are no glaring problems with the accuracy of WAOB soybean yield forecasts, the forecasts have been subject to criticism from time-to-time. Some of the criticism probably reflects a lack of understanding of the forecasting methodology. In particular, some market participants appear to be unaware of the difference between the WAOB and NASS forecasting methodologies. Additional criticism may stem from WAOB changes in methodology, changing periods for calculating trend, or lack of sensitivity to other potential yield indicators such as crop conditions. It is not unreasonable to anticipate that this changing menu of forecasting methods creates some confusion on the part of market participants. It would be very helpful if the WAOB made all data used in estimating yield models available to the public and produced a written document that outlines the exact process used to determine forecasts, including the roles of crop weather regression forecasts, subjective judgment, and any other inputs. This would go a long ways towards reducing confusion about WAOB yield forecasting methods and improving the transparency of these important forecasts to market participants.

References

Good, D. and S. Irwin. "USDA Corn and Soybean Production Forecasting Procedures Revisited With a Focus on Derived Ear Weights." *farmdoc daily* (3):164, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, August 28, 2013.

Good, D., and S. Irwin. "USDA Corn and Soybean Acreage Estimates and Yield Forecasts: Dispelling Myths and Misunderstandings." Marketing and Outlook Brief 2011-02, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, March 2011.

Irwin, S., and D. Good. "Forming Expectations for the 2015 U.S. Average Soybean Yield: What Does History Teach Us?" *farmdoc daily* (5):51, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, March 19, 2015.

Irwin, S., and D. Good. "USDA Corn and Soybean Yield Forecast Errors Across Report Release Months." *farmdoc daily* (1):143, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, September 1, 2011.

Irwin, S., and D. Good. "Have the Accuracy of USDA's August Corn and Soybean Production Forecasts Changed?" *farmdoc daily* (1):134, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, August 19, 2011.

Irwin, S., D. Good, and D. Sanders. "Understanding and Evaluating WAOB/USDA Corn Yield Forecasts." *farmdoc daily* (5):79, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, April 30, 2015.

Irwin, S., D. Good, and D. Sanders. "Are USDA Soybean Yield Forecasts Getting Better or Worse over Time?" *farmdoc daily* (4):169, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, September 4, 2014.

Irwin, S.H., D.R. Sanders, and D.L. Good. "Evaluation of Selected USDA WAOB and NASS Forecasts and Estimates in Corn and Soybeans." Marketing and Outlook Research Report 2014-01, Department of Agricultural and Consumer Economics, University of Illinois at Urbana-Champaign, January 2014.

Isengildina-Massa, O., S.H. Irwin, D.L. Good, and J.K. Gomez. "The Impact of Situation and Outlook Information in Corn and Soybean Futures Markets: Evidence from WASDE Reports." *Journal of Agricultural and Applied Economics* 40(2008):89-104. http://purl.umn.edu/45048

Westcott, P.C., and M. Jewison. "Weather Effects on Expected Corn and Soybean Yields." FDS-13g-01, Economic Research Service, U.S. Department of Agriculture, July 2013. http://www.ers.usda.gov/publications/fds-feed-outlook/fds-13g-01.aspx