Integrating Survey Data with Auxiliary Sources of Information to Estimate Crop Yields

Nathan B. Cruze*

Abstract

The USDA's National Agricultural Statistics Service conducts multiple surveys for major crops, including winter wheat, during a growing season. These surveys are designed to capture the current status of crops at state, regional and national levels. Each of the surveys also provides an indication of end-or-season yields. A Bayesian hierarchical model gives improved yield forecasts by combining these indications from three different types of surveys together with auxiliary data. Modeled state forecasts are benchmarked against a regional forecast and rigorous measures of uncertainty are provided. Advantages of this model are the flexibility for inclusion of new sources of auxiliary information and the incorporation of expert judgment while retaining reproducibility and estimability of standard errors. The model for winter wheat is shown to perform well over a wide variety of conditions, as illustrated by data from the 2012 crop year.

Key Words: Bayesian hierarchical model; Composite estimation; Model-based estimation; Survey sampling.

1. Introduction

The National Agricultural Statistics Service (NASS) is a statistical agency within the United States Department of Agriculture (USDA) tasked with a mission to provide timely, accurate, and useful statistics in service to U.S. agriculture. Among the many valuable official statistics and reports published by NASS, the monthly Crop Production Report includes in-season forecasts and estimates of planted and harvested area, crop yield, and production for most major U.S. crops. Estimates contained in this report are valued by many end users from both the public and private sectors, and the state and national estimates contained in the Crop Production Report serve as inputs into other USDA reports including the World Agricultural Supply and Demand Estimates (WASDE) produced by the World Agricultural Outlook Board in the Office of the Chief Economist . Through their direct release to the public and their role in other USDA processes, NASS's state and national crop forecasts and estimates inform many vital economic decisions in domestic and international commodity markets.

The preparation and release of the Crop Production Report is conducted under the auspices of NASS's Agricultural Statistics Board (ASB). The ASB is a panel of commodity experts that convenes on a monthly basis to review current and historical survey outcomes, available weather data, and other supplemental information. Based on the *survey data and auxiliary information*, the commodity experts reach a consensus. Out of the ASB process, monthly in-season forecasts of planted

*USDA National Agricultural Statistics Service, Room 6412A–South Building, 1400 Independence Ave, SW, Washington, DC 20250

acres, harvested acres, and crop yield at state and national levels are derived for inclusion in the Crop Production Report. Season-end estimates of these quantities are prepared in summary reports released in September and December, which focus on small grains and row crops, respectively.

The Office of Management and Budget (2006) released OMB Standard 4.1 advising federal statistical agencies to "use accepted theory and methods when deriving...projections that use survey data" and that "error estimates must be calculated and disseminated to support assessment of the appropriateness of the uses of the estimates or projections." In response to the OMB standard, and after internal review, NASS began to explore model-based estimation of crop yields as a strategy for capturing the expert assessment of the ASB while making the process of combining available information sources more easily repeatable and providing pertinent measures of uncertainty. NASS research in this area began with a collaborative agreement with the National Institute of Statistical Science (NISS) in 2009. Since 2011, the ASB has received variants of models for corn and soybean yields first developed by Wang et al. (2012), and discussed in Nandram et al. (2014) and Adrian (2012) for consideration in their decision-making processes. At the request of the ASB, NASS has recently undertaken extensions to other commodities including winter wheat. Taking winter wheat as an example, this paper describes NASS methods for model-based estimation of crop yields. Section 2 describes the available surveys and production timelines for release of winter wheat forecasts and estimates. A general specification of the Bayesian hierarchical model is presented in Section 3, and a comparison of modeled winter wheat yields to published yield statistics is shown in Section 4. Finally some possible extensions are discussed in Section 5.

2. NASS Yield Survey Indications and Timelines

Crop yield is a measure of agricultural production per area harvested. In the United States, yield is typically measured in units of bushels per acre, where the weight of a bushel may differ by commodity; a bushel of wheat or soybeans weighs approximately 60 pounds per bushel whereas a bushel of corn weighs approximately 56 pounds, controlling for standard levels of moisture. Since yield is the ratio of output to acreage harvested, a natural materials balance identity arises when comparing regional yield to the yield of smaller constituent units. Letting p, h and μ denote production, harvested acres and yield at a regional level, and letting index $j \in \{1, 2, ..., J\}$ identify the corresponding quantities at smaller geographical units, e.g., states, yield for the region is a *weighted average* of state yields as shown in Equation 1

$$\mu = \frac{p}{h} = \sum_{j=1}^{J} \left(\frac{h_j}{h}\right) \frac{p_j}{h_j} = \sum_{j=1}^{J} w_j \mu_j \tag{1}$$

where weights w_j are proportional to the j^{th} state's harvested acreage. Since Equation 1 represents a physical balance, benchmarking of state yield estimates with regional estimates is desirable.

Production and yield are heavily dependent on many factors throughout the course of the growing season, ranging from weather conditions to individual farmer practices. NASS conducts multiple surveys throughout the growing season to better assess yield potential. In order to distinguish these survey outputs from published values, the direct expansions from yield surveys are referred to as survey 'indications' whereas the term 'estimate' is reserved for an officially published value. Over the course of a season, the ASB will have access to yield indications generated by three distinct surveys: the Objective Yield Survey (OYS), the Agricultural Yield Survey (AYS), and one quarterly Agricultural Production Survey (APS), conducted in September to capture activities related to small grains and commodities such as winter wheat. The surveys differ in scope and geographic coverage as noted below. (For additional information on each, see USDA National Agricultural Statistics Service (2012).)

- *Objective Yield Survey*–NASS objective yield surveys are panel surveys. They are objective in the sense that the survey indications are derived from measurements obtained in the field as opposed to an interview with farmers. The OYS is commodity-specific. Given the expense of conducting such a panel survey, the geographic coverage of the OYS is limited. The OYS sample is drawn only from those states that make up the *speculative region* for the commodity in question. A speculative region represents a collection of the top-producing states in the U.S. with respect to some commodity. For winter wheat, the speculative region is comprised of ten states.
- *Agricultural Yield Survey*–NASS AYS surveys are another panel survey based on *respondent interview*. Whereas OYS was commodity-specific, the AYS is designed for the larger needs of the NASS yield programs, and farmers may be asked to report their expected yields for a range of commodities that they may grow during the season. Although the scope of the survey is not limited to speculative region states, indications and associated standard errors pertaining to the speculative region are also produced during summarization.
- *Acreage, Production and Stocks Survey*–The quarterly crops APS survey is conducted at the conclusion of the harvest in September for winter wheat and small grains (and in December for activities related to row crops). The APS is another interview-based survey, and as the name suggests, the survey covers other season-end quantities in addition to yield. As with the AYS, data on several commodities may be collected simultaneously. Given that the survey is conducted at the conclusion of harvest with a significantly larger sample size, NASS typically views indications derived from this survey as a "gold-standard."

The timeline for NASS yield surveys and publications of winter wheat forecasts and estimates is summarized in Figure 1. Winter wheat is planted in late Autumn of the previous calendar year and enters a state of dormancy during winter months. Its continued development and yield potential can be captured by surveys in May through September. (Row crops such as corn and soybeans are supported by a similar cycle of surveys conducted from August through December.) In all, survey indications and standard errors are available at the state and speculative region level for a total of 5 months of OYS indications, 4 months of AYS indications, and the quarterly APS survey indication prior to the release of season-end estimates. The timeline presents an added challenge in that this typically leaves just a three- to four-day window between survey summarization and the release of the official report. Consequently, any auxiliary information must be brought to bear within this relatively short timeframe.

Example data from the three survey sources for Illinois are plotted (with coarse scale due to

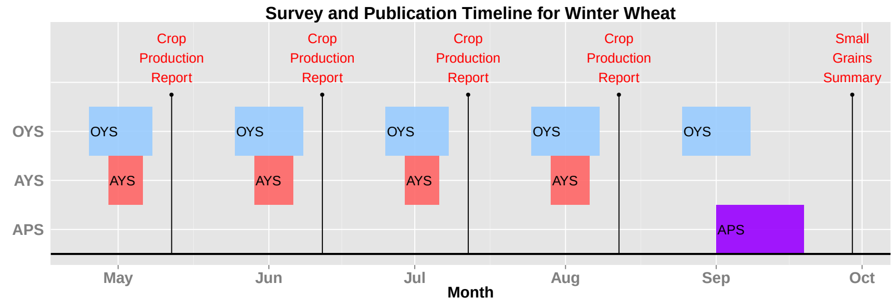

Figure 1: Survey and report production timeline for NASS winter wheat forecasts and estimates

disclosure guidelines) in Figure 2. The length of each time series represent the availability of survey indications at the time of writing. In black, the September APS represents yield at the conclusion of harvest, and may be thought of as an unbiased indication for the final published yield. The OYS survey indications (shown in green) are typically biased upward with respect to the APS survey indications. The AYS survey indications (in red) are typically biased downward relative to APS indications. The apparent biases may become smaller as events of the season unfold. These relationships between NASS yield survey indications tend to hold irrespective of state or commodity. Finally, some evidence of increasing trend may be apparent in Figure 2, reflecting possible innovations in yield over time. These observations help to inform the modeling strategy developed in Section 3.

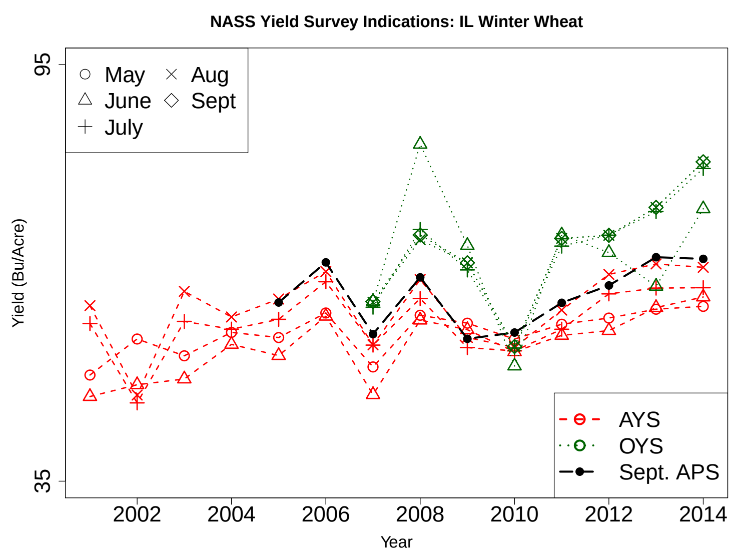

Figure 2: Illinois winter wheat survey indications

3. Model Specification

Just as ASB members must review survey indications and additional information in order to synthesize estimates for the speculative region, the initial emphasis of the NISS-NASS collaborative agreement focused on the synthesis of data available at the speculative region level, with initial application to estimation of corn and soybean yields in Wang et al. (2012). Nandram et al. (2014) details subsequent development of state level models for corn that include benchmarking to the regional yield. Adrian (2012) presented a more parsimonious version of this model applied to both corn and soybeans, omitting the AR(1) correlation structure reflecting the panel-nature of the OYS and AYS survey indications found in the other two. This specification more closely resembles the ASB's thought processes of reviewing only same-month indications while providing comparable performance in terms of root means squared error associated with forecasts. The parsimonious specification in Adrian (2012) has formed the basis for development of NASS winter wheat yield models.

3.1 Speculative Region Model

In the spirit of Wikle (2003) and others, a Bayesian hierarchical model is specified as a collection of conditional and marginal distributions describing the behavior of observed data, an underlying latent yield process, and additional model parameters. Let y_{ktm} denote observed yields from survey $k \in \{O, A, Q\}$ (for OYS, AYS, and quarterly APS, respectively), in year $t \in \{1, 2, ..., T\}$ and month m. Then conditional on regional yield, μ_t, data models for forecast month m are described by

$$y_{ktm}|\mu_t \sim \ indep\ N\left(\mu_t + b_{km}, s^2_{ktm} + \sigma^2_{km}\right), k = O, A \tag{2}$$

and

$$y_{Qt}|\mu_t \sim \ indep\ N\left(\mu_t, s^2_{Qt}\right) \tag{3}$$

where it is understood in Equation 3 that the survey is conducted in September for winter wheat, or December for corn and soybeans yields. Note that Equation 2 models the AYS and OYS data with month-specific biases b_{km}, whereas the APS data are assumed to be unbiased indications for true yield.

The process model describes variation of yield μ_t around a linear function of covariates, $\boldsymbol{z}_t$, observable in year t.

$$\mu_t \sim \ indep\ N\left(\boldsymbol{z}_t'\boldsymbol{\beta}, \sigma^2_\eta\right) \tag{4}$$

Finally, diffuse prior distributions complete the specification of model; The month-specific biases b_{km} and regression coefficients $\boldsymbol{\beta} \sim indep\ MVN(\mathbf{0}, 10^6 \times \boldsymbol{I})$. Prior distributions on variances are specified as σ^2_{km}, $\sigma^2_\eta \sim \ indep\ IG(.001, .001)$. For notational simplicity, it will be convenient to define the collection of data model parameters as $\boldsymbol{\Theta}_d \equiv \left(b_{km}, \sigma^2_{km}\right)$ and the collection of process model parameters $\boldsymbol{\Theta}_p \equiv \left(\boldsymbol{\beta}, \sigma^2_\eta\right)$.

Assuming conditional independence, the likelihood function has the following form

$$[y_O, y_A, y_Q | \mu_t, \mathbf{\Theta}_d] = \prod_{k \in \{O,A,Q\}} [y_k | \mu_t, \mathbf{\Theta}_d] \tag{5}$$

and it follows by Bayes' Rule that the posterior distribution takes the form shown in Equation 6

$$[\mu_t, \mathbf{\Theta}_d, \mathbf{\Theta}_p | y_O, y_A, y_Q] \propto \prod_{k \in \{O,A,Q\}} [y_k | \mu_t, \mathbf{\Theta}_d][\mu | \mathbf{\Theta}_p][\mathbf{\Theta}_d][\mathbf{\Theta}_p] \tag{6}$$

Full conditional distributions for model parameters may be obtained explicitly. A Gibbs sampling algorithm is employed to obtain estimates of model parameters Gelman et al. (2003). The reader is referred to (Adrian, 2012, Appendix A) for more details about the full conditional distributions of all model parameters, and the Gibbs sampling algorithm. For brevity, we discuss only the full conditional distribution for the main quantity of interest, regional yield μ_t,

$$[\mu_t | y_O, y_A, y_Q, \mathbf{\Theta}_d, \mathbf{\Theta}_p] \sim N\left(\frac{\Delta_{2t}}{\Delta_{1t}}, \frac{1}{\Delta_{1t}}\right) \tag{7}$$

where

$$\Delta_{1t} = \sum_{k=O,A} \frac{1}{\sigma_{km}^2 + s_{ktm}^2} + \frac{I_{\{Q\}}}{s_{Qt}^2} + \frac{1}{\sigma_\eta^2} \tag{8}$$

$$\Delta_{2t} = \sum_{k=O,A} \frac{y_{ktm} - b_{km}}{\sigma_{km}^2 + s_{ktm}^2} + \frac{I_{\{Q\}} y_{Qt}}{s_{Qt}^2} + \frac{\boldsymbol{z}_t' \boldsymbol{\beta}}{\sigma_\eta^2}. \tag{9}$$

Note that Equation 8 represents the sum of the precisions of each information source. Taken together, Equation 8 and Equation 9 indicate that the posterior mean of Equation 7 is *a composite estimator* which incorporates the following available information sources: bias corrected AYS and OYS indications, the quarterly APS indication (when it is available), and covariates information. This is a built-in rule for the combination of many sources of information which applies proportionally larger weight to more precise sources.

3.2 Unconstrained and Constrained State Models

Data and process models for the states resemble those of the speculative region with models for each state j given by:

$$y_{ktmj} | \mu_{tj} \sim indep\ N\left(\mu_{tj} + b_{kmj}, s_{ktmj}^2 + \sigma_{kmj}^2\right), k = O, A, \tag{10}$$

$$y_{Qtj} | \mu_{tj} \sim indep\ N\left(\mu_{tj}, s_{Qtj}^2\right), \tag{11}$$

$$\mu_{tj} \sim indep\ N\left(\boldsymbol{z}_{tj}' \boldsymbol{\beta}_j, \sigma_{\eta j}^2\right). \tag{12}$$

As before, diffuse prior distributions are specified on the data and process model parameters of each state. The full conditional distribution of yield in the j^{th} state, μ_{tj} resembles Equation 7.

Assuming independence of state yields and defining $\boldsymbol{\mu}_{t\cdot} \equiv (\mu_{t1}, \mu_{t2}, \ldots, \mu_{tJ})$, the full conditional distribution of the collection of state yields is a multivariate normal distribution:

$$[\boldsymbol{\mu}_{t\cdot}|\boldsymbol{y}_O, \boldsymbol{y}_A, \boldsymbol{y}_Q, \boldsymbol{\Theta}_d, \boldsymbol{\Theta}_p] \sim indep\, MVN\left(vec\left(\frac{\Delta_{2tj}}{\Delta_{1tj}}\right), diag\left(\frac{1}{\Delta_{1tj}}\right)\right) \quad (13)$$

Although parameters μ_{tj} respect Equation 1, estimates of these parameters derived under Equation 13 may not, therefore, it is desirable to enforce the balance constraint between the speculative region and member states. Using a strategy similar to Nandram and Sayit (2011), iterates of the speculative region MCMC simulation are fed into the MCMC simulation for a 'constrained' state level model. Conditioning Equation 13 on Equation 1, the collection of the first $j-1$ states will follow a multivariate normal distribution

$$\left(\mu_{t1}, \mu_{t2}, \ldots, \mu_{t(J-1)}\right) \sim MVN(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}). \quad (14)$$

(Complete expressions for $\bar{\boldsymbol{\mu}}$ and $\bar{\boldsymbol{\Sigma}}$ are given in Adrian (2012), Eqs. 10 and 11.) At each year t, the the yield for the J^{th} state is given by

$$\mu_{tJ} = \mu_t - \frac{1}{w_{tJ}}\sum_{j=1}^{J-1} w_{tj}\mu_{tj}. \quad (15)$$

3.3 Model Outputs and State Forecast Decomposition

Briefly summarized, the steps to producing model-based yield estimates are as follows.

1. Fit the speculative region model. Model-based estimates of speculative region yield may be obtained as the mean of the Monte Carlo sample for μ_t. Standard errors are obtained as the sample variances of the Monte Carlo samples.
2. Fit the constrained state model. Enter the μ_t iterates into the MCMC simulation for state yields. Benchmarked state yields and measures of uncertainty are obtained from the Monte Carlo samples under Equation 14 and Equation 15.

For the purposes of publication of yield over some aggregate region and benchmarked state yields, these two items suffice. An additional benefit in fitting an 'unconstrained' state model is that it provides for an approximate Bayesian decomposition of the state forecasts by information source. In the spirit of Equation 8 and Equation 9, the constrained state yield estimate may be thought of as a weighted average of various information sources, plus an additive benchmarking adjustment, d_j.

$$\begin{aligned} \hat{\mu}_{Tj} &= \sum_{k\in\{O,A,Q,Covariates\}} c_k(SOURCE)_{Tk} + d_j \quad (16) \\ c_k &\propto (variance(SOURCE_k))^{-1} \end{aligned}$$

The benchmarking adjustment represents the difference between yield under the constrained state model and yield under the unconstrained model. Due to disclosure guidelines, no survey indications will be published within this report. However, a collection of information that can be readily obtained for each supported commodity by the current NASS yield modeling strategy is summarized in Table 1. The first two rows represent benchmarked state and regional yield estimates and associated errors, e.g., standard error or RMSE. Subsequent rows represent bias-corrected OYS and AYS indications, as well as a fitted yield based on observed auxiliary data, and the September APS indication when available it is available. The final row represents the benchmarking adjustment, or the difference between the constrained and unconstrained state yield forecasts. In the decomposition by information source the posterior means and variances, i.e., components of Equation 16, are taken with respect to the posterior distribution under the *unconstrained* state model as justified by Kass and Steffey (1989).

		State 1	**State 2**	$\cdots$	**State J**	**SPEC**
Overall Forecast	$\hat{\mu}_{Tj}$	x	x	$\cdots$	x	x
Error		x	x	$\cdots$	x	x
OYS	$y_{OTmj} - \hat{b}_{Omj}$	x	x	$\cdots$	x	x
AYS	$y_{ATmj} - \hat{b}_{Amj}$	x	x	$\cdots$	x	x
Covariates	$\boldsymbol{z}'_{Tj}\hat{\boldsymbol{\beta}}_j$	x	x	$\cdots$	x	x
Sept. APS	y_{QTj}	x	x	$\cdots$	x	x
Benchmarking Adj.	d_j	x	x	$\cdots$	x	

Table 1: Overall forecasts (BU/acre), Error, and forecast decompositions

4. Winter Wheat Yield

4.1 Winter Wheat Speculative Region

The winter wheat speculative region is comprised of ten states: Washington, Montana, Colorado, Nebraska, Kansas, Oklahoma, Texas, Missouri, Illinois, and Ohio. This region represents a collection of some of the top producers of winter wheat in the United States, and the Objective Yield Survey for winter wheat is conducted exclusively within these states. The speculative region states are shown in Figure 3, along with each state's proportion of harvested acres (weights, w_j) in year 2012. In 2012, Kansas, Oklahoma, and Texas combined accounted for more than 64% of all acres of winter wheat harvested within the speculative region.

The breadth of this geographical region encompasses a diversity of planting and harvesting decisions. NASS measures production of four distinct classes of winter wheat: red hard, red soft, white hard, and white soft wheat. Figure 4 shows the typical share of each state's total *production* by class of winter wheat. In some sense, the states specialize with respect to each class of winter wheat. Higher yields tend to be associated with soft types of winter wheat. Consequently, yields in Washington, Missouri, Illinois, and Ohio are among the highest in the speculative region. To the extent that yields truly differ by type of wheat, these effects are confounded with state yield.

States begin and complete the harvest of winter wheat at different time points throughout the survey cycle. Figure 5 describes the usual initial harvest, active harvest, and final harvest dates as described in USDA National Agricultural Statistics Service (2010). In general, the progres-

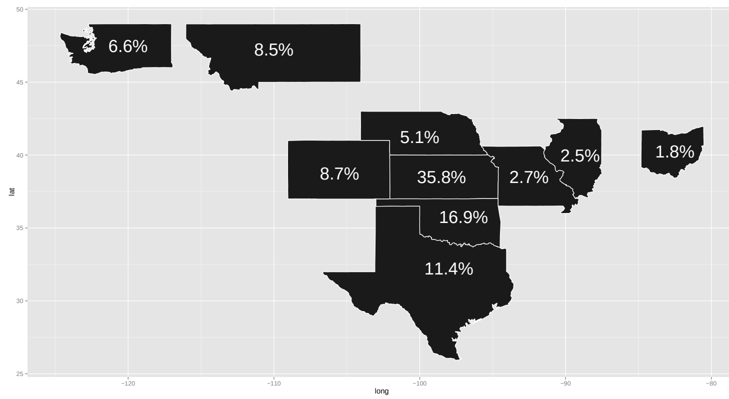

Figure 3: Winter wheat speculative region states. Percentages represent proportion of acres harvested (w_j) in each state in 2012.

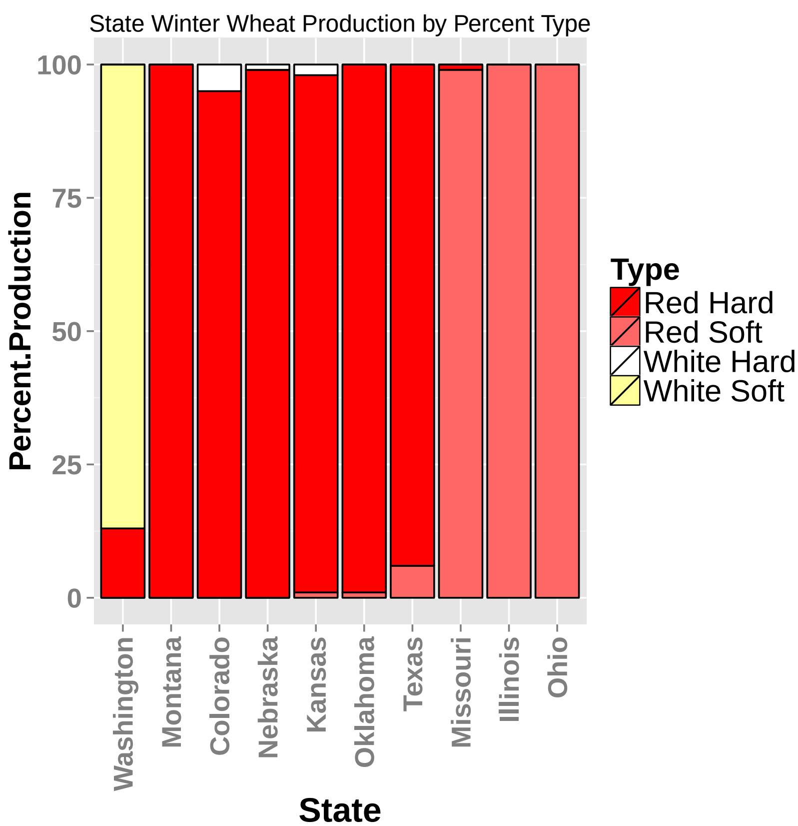

Figure 4: Share of production by percent type

sion of harvest dates follows a south-to-north pattern throughout the speculative region. Harvest is typically completed in Texas before it has even begun in Washington or Montana. This differential harvest influences the availability of some survey indications; May OYS indications are only available for Texas, Oklahoma, and Kansas. Furthermore, it informs the inclusion and timing of covariates within the model.

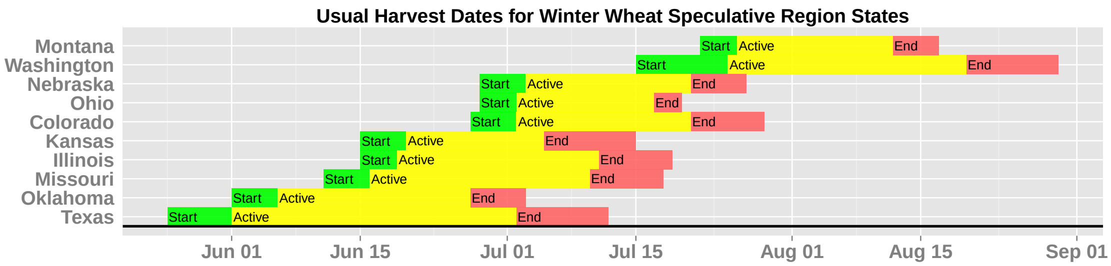

Figure 5: Usual harvest dates for winter wheat speculative region states

4.2 Covariates

The specification of state-level latent process models in Equation 12 asserted that the true yield of state j, μ_{tj}, varied around a linear function of observable covariates, that is $\tilde{\mu}_t = \boldsymbol{z}'_{tj}\boldsymbol{\beta}_j$. The vetted model for winter wheat utilizes a combination of covariates described by Equation 17

$$\tilde{\mu}_{tj} = \beta_{j1} + \beta_{j2}z_{tj2} + \beta_{j3}z_{tj3} + \beta_{j4}z_{tj4} + \beta_{j5}z_{tj5} \tag{17}$$

where

- z_{j2}–a linear time trend variable reflecting potential for innovations in yield over time
- z_{j3}–a monthly state average rainfall, in inches (NOAA National Climatic Data Center (2015))
- z_{j4}–a monthly state average temperature (NOAA National Climatic Data Center (2015))
- z_{j5}–percent of crop rated good or excellent in a particular week (USDA National Agricultural Statistics Service (2015b)). This variable is essentially produced by polling field agents and operators within each state, and it further represents the role of expert assessment within the model.

In the specification above, coefficients β are allowed to vary by state. Corresponding speculative region covariates are derived from state-level covariates by taking a weighted average based on weights w_j, each state's share of harvested acres within the spec region.

To accommodate the differential harvest discussed in Section 4.1, the schedule of updates described in Table 2 was adopted. In essence, the covariates have been selected to describe conditions leading up to initial and active harvest dates. To produce a May forecast, the model is initialized with crop and weather conditions as of April. Updates are made selectively in June and July, following the South-to-North pattern in harvest.

State/FIPS		May Covariates		June Covariates		July-September Covariates	
		Condition (Week #)	*Weather (Month)*	*Condition (Week #)*	*Weather (Month)*	*Condition (Week #)*	*Weather (Month)*
CO	8	15	April	21	May	21	May
IL	17	15	April	19	May	19	May
KS	20	15	April	19	May	19	May
MO	29	15	April	19	May	19	May
MT	30	15	April	19	May	24	June
NE	31	15	April	21	May	21	May
OH	39	15	April	21	May	21	May
OK	40	15	April	17	April	17	April
TX	48	15	April	17	April	17	April
WA	53	15	April	22	May	22	May

Table 2: Schedule of covariates updates by state. Final updates are made in July in Montana.

4.3 Year 2012 Model Results

One important measure of performance of the model is through a simple comparison to the officially published statistic. Results for the 2012 crop year are depicted in Figure 6. In most months and states, the model (shown in black) and the ASB (in red) show a high degree of agreement. The speculative region results are nearly coincident with those of the ASB. Even for those states such as Colorado, Nebraska, Montana and Washington where the model may deviate from the published value, the published yield is captured within the 95% credible interval of the model-based indications, with the exception of the August forecast in Colorado. As mentioned, the producers of soft varieties of winter wheat tend to have significantly higher yields. Figure 6 lends some visual evidence that Equation 1 has been enforced, since the speculative region yield is closer to yields in those states with the largest acreages harvested, including Kansas and Oklahoma. We report that all benchmarking adjustments associated with state estimates were positive, indicating that state winter wheat yields modeled independently would otherwise fail to account for some output indicated at the regional level. The winter wheat model has only been presented to the ASB for their deliberations as of the 2015 crop year, thus, the comparison shown for the 2012 crop year truly demonstrates whether the model can reflect the consensus estimate of NASS commodity experts.

One benefit of the model is that it can be described in terms of a decomposition of available data sources. The 'rule' for the combination of different data sources is described by Figure 7. Due to the uncertainty associated with early season survey indications and the limited OYS participation in seven states in May, the covariates portion of the model is most heavily emphasized in the weighted average decomposition of overall yield for the early season indications. As the season progresses, the precision of the survey indications improves, and covariates play a smaller role in the overall yield. The September APS survey is conducted after the completion of harvest, and has a much larger sample size than either of the other surveys. Therefore, it is not surprising that the September APS indications become a driving factor in overall yield at season's end.

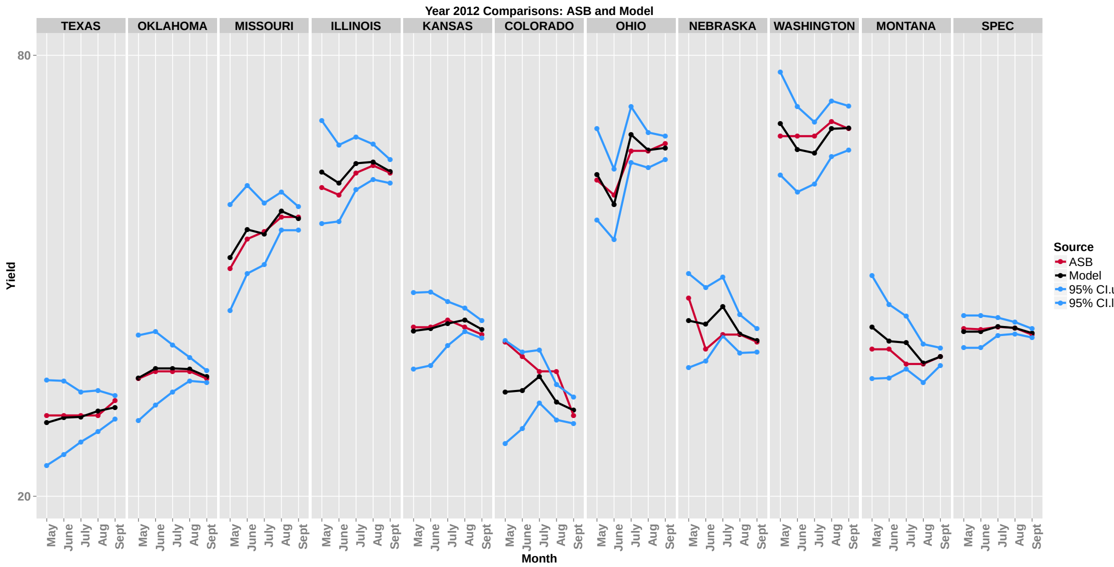

Figure 6: Comparison of ASB forecasts and estimates to model-based yield indications

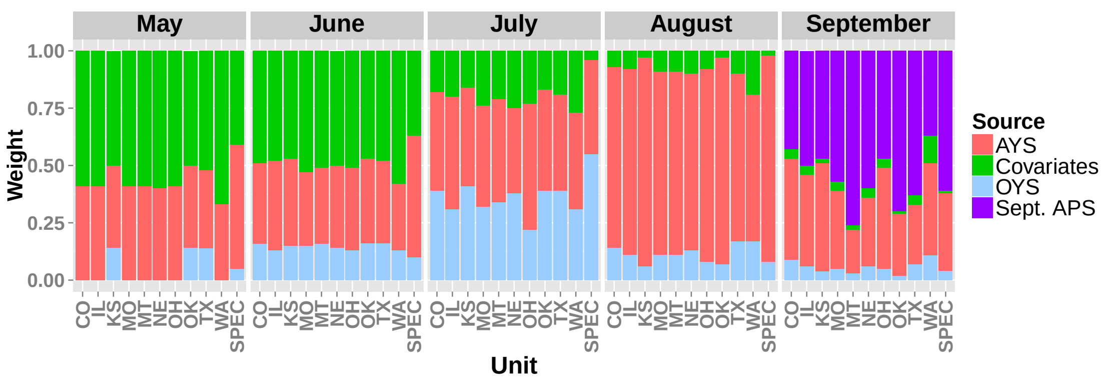

Figure 7: Emphasis placed on each information source throughout the season

5. Extensions and Conclusions

Currently the NASS model-based yield estimates program supports three commodities. In order to fulfill a larger role in the publication of the Crop Production Report and annual summaries, extensions and future research is required in the following areas:

1. inclusion of additional commodities,
2. support for states outside the speculative region,
3. accounting for new technologies such as soil moisture monitors.

In addition to corn, soybeans, and winter wheat, commodities including upland cotton and potatoes are presently targeted by their own Objective Yield Surveys as well. With the identification of suitable covariates, these commodities can be readily supported by the models presented in this paper. More generally, it may be desirable to publish yields for smaller commodities that are not supported by an Objective Yield Survey, and it *will* be desirable to publish yields at the *national* level. This entails estimation of yield for states that are not part of the speculative region. In the absence of OYS indications for these states, the role of auxiliary information becomes even more critical, particularly for producing early season forecasts of crop yield. The models perform well over a wide variety of conditions, but promising new technologies may provide data that help improve model performance in the most extreme conditions. The models presented in this paper permit flexibility to include these new sources of auxiliary information as they become available, while capturing expert judgment in a manner that is easily reproducible and gives rise to appropriate measures of uncertainty.

Acknowledgments

I would like to recognize my colleagues Edwin Anderson and Thomas Jacob of USDA NASS for their help in obtaining and processing historical winter wheat yield survey indications. I also gratefully acknowledge Nell Sedransk of the National Institute of Statistical Science for many helpful conversations during the development of the winter wheat yield models.

References

Adrian, D. (2012). A model-based approach to forecasting corn and soybean yields. Fourth International Conference on Establishment Surveys.

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (2003). *Bayesian Data Analysis (2nd ed.)*. Chapman & Hall/CRC.

Kass, R. and Steffey, D. (1989). Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association*, 84(407):717–726.

Nandram, B., Berg, E., and Barboza, W. (2014). A hierarchical Bayesian model for forecasting state-level corn yield. *Environmental and Ecological Statistics*, 21(3):507–530.

Nandram, B. and Sayit, H. (2011). A Bayesian analysis of small area probabilities under a constraint. *Survey Methodology*, 37:137–152.

NOAA National Climatic Data Center (2015). U.S. Climate Divisions: nClimDiv Data Set. `http://www.ncdc.noaa.gov/monitoring-references/maps/us-climate-divisions.php`. Accessed: 24 August 2015.

Office of Management and Budget (2006). Standards and guidelines for statistical surveys. `https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf`. Accessed: 24 August 2015.

USDA National Agricultural Statistics Service (2010). Field crops usual planting and harvest dates. http://usda.mannlib.cornell.edu/usda/current/planting/planting-10-29-2010.pdf. Accessed: 24 August 2015.

USDA National Agricultural Statistics Service (2012). The Yield Forecasting Program of NASS. http://www.nass.usda.gov/Education_and_Outreach/Understanding_Statistics/Yield_Forecasting_Program_may2012.pdf. Accessed: 24 August 2015.

USDA National Agricultural Statistics Service (2015a). Crop Production. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1046. Accessed: 24 August 2015.

USDA National Agricultural Statistics Service (2015b). State Crop Progress and Condition. http://www.nass.usda.gov/Publications/State_Crop_Progress_and_Condition/index.asp. Accessed: 24 August 2015.

USDA Office of the Chief Economist (2015). How the WASDE is Prepared. http://www.usda.gov/oce/commodity/wasde/prepared.htm. Accessed: 24 August 2015.

Wang, J. C., Holan, S. H., Nandram, B., Barboza, W., Toto, C., and Anderson, E. (2012). A Bayesian approach to estimating agricultural yield based on multiple repeated surveys. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(1):84–106.

Wikle, C. (2003). Hierarchical Bayesian Models for Predicting the Spread of Ecological Processes. *Ecology*, 84:1382–1394.