Improved County-Level Estimation of Crop Yield Using Model-Based Methodology with a Spatial Component

Michael E. Bellow
USDA/NASS

Abstract

Estimates of agricultural commodities at the county (small area) level published by the USDA's National Agricultural Statistics Service are scrutinized by users in government, the private sector and academia. For that reason, ensuring their accuracy and reliability is important. An interesting avenue of research along those lines has been the application of model-based methodology to estimation of crop yields. Such methods have the potential to provide substantial improvements in efficiency over standard ratio-based estimation. The Stasny-Goel method, developed through a cooperative agreement between NASS and The Ohio State University, assumes a mixed-effects model with a spatial component that takes into account correlations among yields from neighboring counties. Estimates are computed via an iterative Bayesian algorithm. I will describe a simulation study where the Stasny-Goel method was found to compare favorably with ratio estimation and another model-based approach for various crops in ten geographically dispersed states.

Key Words: small area estimation; simulation; convergence

1. INTRODUCTION

The National Agricultural Statistics Service (NASS) has been publishing estimates of crops, livestock and other commodities at the county level since 1917. The primary source of data for agricultural commodity estimation has always been surveys of farmers, ranchers and agribusiness managers who provide requested information on a voluntary, confidential basis. Since surveys designed and conducted at the national and state levels are seldom adequate for obtaining reliable county estimates, NASS has made extensive use of ancillary data sources such as list sampling frame control data, previous year estimates, earth observing satellite data and Census of Agriculture data. County level estimates are produced at NASS Field Offices (FOs) using the County Estimates System (Iwig, 1993), a set of computer programs that processes the combined input data from all internal and external sources. Statisticians at the FOs use the outputs of this system to set final (official) county estimates.

Each FO conducts a County Estimates Survey (CES) every year. Since 2002, multivariate probability proportional to size (MPPS) sampling has been used to select the samples of farms, with questionnaires mailed out to the operators and telephone follow-ups done where necessary. Data from other NASS surveys (such as the September and December Quarterly Agricultural Surveys (QAS) and January Cattle) are merged with the CES sample to form a combined data set which is then used to calculate various commodity estimates at the county level. Final published county estimates must be consistent with corresponding state and district level figures.

Ratio estimation is the standard method used by NASS to derive county level yields. The simple ratio estimator is computed as the sum of QAS reported crop production divided by the corresponding sum of reported harvested acreage. This estimator can produce unreliable yields due to fluctuations in harvested acreage from year to year. Furthermore, it does not use data from any county other than the one being estimated (so an estimate for a given county cannot be generated in the absence of positive survey records for that county). In NASS operational practice, a version of stratified sampling is used to generate ratio estimates that are weighted by the sampling rate. Although the weighting is difficult to replicate, Crouse (2000) found that non-weighted ratio estimates could be used for research purposes without loss of applicability. Therefore, the non-weighted approach was used for the study documented in this paper.

Stasny, Goel et al. (1995), working under a cooperative agreement between NASS and the Ohio State University, developed a Bayesian county yield estimation algorithm with a simple spatial component based on the notion that crop yields of counties in close geographic proximity tend to be more similar than those of counties further apart. This procedure, referred to as the Stasny-Goel (SG) method, assumes a mixed effects model with farms as the sample units, farm size (reduced to two or three size groups based on total land operated) as the fixed effect and county location as the random effect. The county effect is assumed to be multivariate normal with mean vector proportional to the previous year's county yields and variance-covariance matrix reflecting positive spatial correlation only between neighboring counties.

Survey records are post-stratified by farm size. The Stasny-Goel program attempts to fit the model using a version of the EM algorithm, with county level estimates computed as weighted averages of individual farm level estimates. The weights are derived from size group membership data obtained from the most recent Census of Agriculture.

Table 1: Crop/State Combinations Tested (Prevalence Class Denoted By A, B or C)

Crop	State									
	CO	FL	MI	MS	NY	ND	OH	OK	TN	WA
Barley	C		B			A				B
Corn	B		A	A	A	A	A	B	A	C
Cotton (Upland)		C		B					C	
Dry Beans						A				
Oats	C		A		A	A	A	B		B
Rye						C		B		
Sorghum	C			C				A	C	
Soybeans			B	A	B		A	A	A	
Sunflower	C					A				
Tobacco (Burley)							C		B	
Spring Wheat						A				B
Winter Wheat	B		A	B	B	B	A	A	A	B

A ten state simulation study was conducted to compare the efficiency of the Stasny-Goel method with both the ratio method and an alternative model-based approach developed by Griffith (2001). Griffith's method predicts yield values using the published number of farms producing the crop of interest, employing Box-Cox and Box-Tidwell transformations in conjunction with an autoregressive specification so as to optimize agreement with model assumptions. Both the Stasny-Goel and Griffith (G) algorithms are programmed in SAS IML. The crops tested were barley, corn (for grain), cotton (upland), dry beans, oats, rye, sorghum (for grain), soybeans, sunflower (oil and non-oil varieties combined), tobacco (air-cured light burley), spring wheat and winter wheat. The three estimators were compared for the 2002 and 2003 QAS cycles.

The ten states in the study area were Colorado, Florida, Michigan, Mississippi, New York, North Dakota, Ohio, Oklahoma, Tennessee and Washington. They were selected for agricultural diversity, with each representing a different region from USDA's subdivision of the country.

As an additional summary categorization by which the relative performance of the methods could be assessed, a measure of prevalence of a crop within a given state was computed as the percent of counties in the state for which positive harvested acreage for the crop was reported on the QAS. For crops tested in 2002 and 2003, the combined percentage over both years was used. For each state, crops were divided into the following three prevalence classes based on this measure: A (70 percent or higher), B (40 to 69 percent) and C (below 40 percent). The rationale for choosing these particular limits was to have intervals of roughly equal length and a sufficient number of crop/state combinations in each category. Table 1 lists the specific crops tested in each state and shows the prevalence class for all crop/state combinations.

2. STASNY-GOEL ESTIMATOR

Data sources for the Stasny-Goel method include the CES, QAS, Census of Agriculture and NASS's Published Estimates Data Base. Neighboring county information for a given state is provided by an input file containing a two-column listing of pairs of counties in the state that share a common border. The SG program uses this data set to form the *neighbor matrix*, an *nc* x *nc* array (where *nc* = number of counties in the state) with the entry in each row *i*, column *j* being 1 if the *i*th county (alphabetically within the state) is a neighbor of the *j*th county and 0 otherwise. Since each county is regarded as a neighbor of itself, all entries along the main diagonal are 1.

The SG method requires that post-stratification size groups be defined. The group definitions are based on total land in farms and vary over states due to differences in average farm size. For each county in a state, the program computes the percentages of Census total farm acreage operated within each size group. These percentages serve as post-stratification weights for the computation of county yield estimates. The program cannot run if one or more of the size groups contain no positive QAS records for the crop of interest. QAS tract level data for the current year are post-stratified by county and farm size based on the Census acreage data, with separate yield estimates computed for each size group in all counties.

For survey years not coinciding with a Census year, the post-stratification weights can be updated to the current year using: 1) ratios between official NASS state level estimates of total land for the current and Census years, and 2) ratios between official NASS state level estimates of number of farms for the current and Census years. This procedure was followed for the study described in this paper.

The Stasny-Goel method is based on the following mixed effects model:

$$y_{ijk} = \mu + \tau_i + g_j + \varepsilon_{ijk}$$

where:

y_{ijk} = yield for i^{th} county, j^{th} size group, k^{th} farm

μ = overall mean county yield

τ_i = random effect for i[th] county

g_j = fixed effect for size group j

ε_{ijk} = random error term

The random errors are assumed to be independent and normally distributed with zero mean and equal variance. The county effects are assumed to be multivariate normal with means proportional to the previous year's county yield estimates. The correlation (ρ) between county effects is assumed to be the same for all pairs of neighboring counties in the state and zero for all pairs of non-neighboring counties. This formulation gives the model a simple spatial component if $\rho>0$.

A version of the EM algorithm is used to fit the model, with the random county effects treated as missing data. Previous year county yields from the CES are used in conjunction with current year QAS farm level data to derive initial estimates of the size group effects, county effects and yield variances. If no previous year yield figure is available for a given county, the (agricultural statistics) district level yield is used instead. If the district figure is unavailable as well, then the state level yield is used. An initial estimate of the spatial correlation ρ is also generated.

At each iteration, the estimates of group and county effects, variance and spatial correlation are adjusted. Relative group and log-likelihood distances are computed based on ratios between measures computed at the current and previous iteration. The iterative process continues until either: 1) both distance metrics fall below preset limits, or 2) a preset maximum allowable number of iterations is reached.

Once the EM algorithm has terminated, the program computes final estimates of yield for each county as weighted sums over size groups of the estimated overall mean county yield, county effect and size group effect. The user is provided with an option to rescale the computed county yields for consistency with official NASS state level yield estimates.

3. SIMULATION METHODOLOGY

The NASS data sources needed for the study were QAS data from 2002 and 2003, CES data from 2001-03, Census data from 2002 and published county yield estimates for 2002 and 2003. QAS data obtained from the NASS Field Offices of the ten states in the study area included record level crop production, harvested acreage and yield. CES data provided previous year computed yields which served as initial values for the SG algorithm. Census data on number of farms and land in farms were used to define the post-stratification size groups.

Simulated populations of yield values were generated from which 'true' population parameters could be derived for later comparison with estimates computed over sampled subsets. For each crop of interest, multiple regression analysis was performed with the survey yield response values being the dependent variable. The four independent variables used were published county yield estimates for the current year, weighted average neighbor yield and two indicator variables pertaining to membership in size groups. The weighted average neighbor yield for a given county was computed as the weighted average of the official yield estimates of all neighboring counties. The weight assigned to each neighboring county was the ratio of harvested acreage (official estimate) for that county to the total harvested acreage of all the neighboring counties. This variable was included to try and increase the spatial correlation of the simulated data so as to better reflect real survey data.

A very large number of simulated survey data sets (10,000) was generated in order to ensure that the 'true' population parameters computed from these records would agree with the model. From this population, 250 data sets were selected using simple random sampling. The Stasny-Goel, Griffith and ratio methods were then applied to each of the sampled data sets. For each county, the sample based estimates for a given method were averaged and compared with the corresponding population values. The maximum allowable number of iterations was set at 5,000 for both algorithms. A provision for allowing SG to go further if the computed log-likelihood is maximized at the prespecified limit (continuing to either convergence or the next decrease in log-likelihood) was added to the program in an effort to increase the convergence percentage.

Occasionally, the regression equation generated negative yields which were rounded up to zero. Since the rounding process induces a minor bias into the simulated data, the intercept term needed to be adjusted. A pilot population of 10,000 simulated data sets was generated for this purpose. The adjustment term was selected so that the state level crop yield averaged over the simulated data sets equaled the official state yield estimate. The actual set of 10,000 simulated data sets used in the estimator comparison was generated via a different random number seed than the one used to create the pilot population. For internal consistency purposes, the same seed was used for all crops evaluated for a given year within a state. For both SG and G, the model-based simulated county yield estimates (not adjusted to agree with state level totals) were used in order to have a pure test of estimator efficiency.

Due to NASS data disclosure restrictions prohibiting publication of estimates for counties with fewer than three positive records for a given crop (although combined estimates for groups of counties ineligible for disclosure are often published), only those counties having at least three positive survey records were used in the estimator comparison.

For most crop/state/year combinations, three Census based size groups were defined. There were six cases for which one of the three groups contained no positive survey data for the crop being estimated so that two groups were used instead – Colorado barley (2002 and 2003), Colorado oats (2002 and 2003), Ohio tobacco (2003) and Washington oats (2003). With Florida cotton (2002) and Washington corn (2003), the two group setup resulted in one of the groups containing no positive survey data. For those two cases, alternative groups based on survey rather than Census data were used in order to get the SG program to run. For Ohio tobacco (2003), the Griffith algorithm could not be run successfully so only the SG and ratio estimates were used. Comparisons of Moran's I coefficient showed that the simulated data sets accurately reflected the spatial correlation inherent in real survey data.

4. RESULTS

Results of the estimator comparison tests for the ten state simulation study are discussed in this section. For both model-based methods, only those simulated data sets for which the algorithm converged within the maximum allowable number of iterations were used. Estimates were still produced for some non-convergent Stasny-Goel simulation runs and all non-convergent Griffith runs. The reason for excluding such cases from the comparison tests was to keep estimator efficiency issues separate from convergence issues (discussed in Section 5) so as not to cause results to be artificially biased in favor of one method or the other.

For all twelve crops tested, pairwise comparisons of the three estimators were done for the following five efficiency measures - absolute bias, variance, mean square error (MSE), lower tail proximity (LTP) and upper tail proximity (UTP). Absolute bias was computed as the average value (over simulations) of the absolute differences between the estimates produced by a given method and the population 'true' county yields. Variance was computed as the sample variance of simulated county yield estimates, while mean square error was calculated by averaging the squared deviations between estimates and 'true' county yields. The final two measures assess outlier properties of the estimators, i.e., the tendency to produce 'out of bounds' yield values. LTP is defined as the absolute difference between the 5^{th} percentile of the simulated yield estimates and the 'true' county yield, while UTP is defined similarly using the 95^{th} percentile. High values of one or both of these measures suggest that the estimator in question is outlier prone.

For each measure, Table 2 shows the total number of crop/state/year cases in the study where one method in a pair was better than the other in more counties than vice versa. The 'tied' column shows the number of cases where both methods were favored for an equal number of counties.

From Table 2, both SG and G were found to be appreciably better than R for all five performance measures. SG outperformed G by a wide margin for absolute bias and MSE and a narrow margin for the two outlier measures, while G was superior to SG for variance.

Table 3 summarizes crop/state/year cases by prevalence class (as defined in Section 1). The figures suggest that the relative performance of the three methods is not strongly influenced by how common or rare a given crop may be in a state.

In order to compare the three estimators for statistically significant differences with respect to absolute bias, one-sided Wilcoxon rank sum tests were run on absolute values of the residuals (differences between estimates and population 'true' values). This two-sample nonparametric procedure assesses whether the population medians of the two samples are significantly different. The tests were carried out on a pairwise basis at the 10 percent significance level for the three estimation methods, with two one-sided tests done in each case. The null hypothesis for each test was equality of median absolute error (MAE) for the two methods. The alternative hypothesis for test A was the first method (in the pair) having lower MAE than the second (vice versa for test B).

For each crop and pair of methods, Table 4 shows the number of counties (summed over states) for which: 1) test A detected lower MAE for the first method, 2) test B detected lower MAE for the second method, and 3) both tests concluded equal MAE for the two methods. Totals over all crops are also shown. Table 5 provides additional summary information, showing for each year the total number of crop/state cases for which one method in each pair was favored in more counties than the other as well as the number of ties (both methods favored the same number of times).

The results of the rank sum tests provide statistically defensible evidence that the Stasny-Goel method is better than the other two methods with respect to absolute bias. Table 4 shows SG having lower MAE than R for all 12 crops and lower MAE than G for 11 crops (rye being the exception) in most counties. Overall, SG was found to have lower MAE than R in 79 percent of counties tested while G had only a one percent advantage over R. Table 5 shows that in 95 percent of the crop/state/year cases tested, the absolute bias of SG was significantly lower in more counties than that of R.

Table 2: Summary of Pairwise Comparisons by Performance Measure

Measure	SG vs. G			SG vs. R			G vs. R		
	No. Cases Favoring			No. Cases Favoring			No. Cases Favoring		
	SG	G	Tied	SG	R	Tied	G	R	Tied
Absolute Bias	65 (78%)	16 (19%)	2 (2%)	80 (95%)	3 (4%)	1 (1%)	46 (55%)	33 (40%)	4 (5%)
Variance	35 (42%)	48 (58%)	0 (0%)	84 (100%)	0 (0%)	0 (0%)	81 (98%)	1 (1%)	1 (1%)
MSE	62 (75%)	19 (23%)	2 (2%)	81 (96%)	2 (2%)	1 (1%)	56 (67%)	22 (27%)	5 (6%)
LTP	43 (52%)	39 (47%)	1 (1%)	83 (99%)	1 (1%)	0 (0%)	78 (94%)	4 (5%)	1 (1%)
UTP	38 (46%)	36 (43%)	9 (11%)	83 (99%)	0 (0%)	1 (1%)	80 (96%)	2 (2%)	1 (1%)
All	243 (59%)	158 (38%)	14 (3%)	411 (98%)	6 (1%)	3 (1%)	341 (82%)	62 (15%)	12 (3%)

Table 3: Summary of Pairwise Comparisons by Prevalence Class

Class	SG vs. G			SG vs. R			G vs. R		
	No. Cases Favoring			No. Cases Favoring			No. Cases Favoring		
	SG	G	Tied	SG	R	Tied	G	R	Tied
A	106 (57%)	76 (41%)	3 (2%)	182 (98%)	3 (2%)	0 (0%)	155 (84%)	28 (15%)	2 (1%)
B	89 (57%)	60 (39%)	6 (4%)	151 (97%)	3 (2%)	1 (1%)	128 (83%)	24 (15%)	3 (2%)
C	48 (64%)	22 (29%)	5 (7%)	78 (97.5%)	0 (0%)	2 (2.5%)	58 (77%)	10 (13%)	7 (9%)

Two one-sided Wilcoxon signed rank tests (called A and B) were done at the ten percent significance level, with the null hypothesis being zero median error (ME) in both cases. The alternative hypothesis was negative median error for Test A and positive median error for Test B. For each method, Table 6 shows the total number of counties (summed over crops and states) for which: 1) test A detected negative median error, 2) test B detected positive median error, and 3) both tests concluded zero median error.

Table 6 indicates that negative median error was concluded in 59 percent of all counties tested for SG and 54 percent for G, suggesting that the bias of both estimators is generally negative. For the ratio estimator, zero median error was concluded by both one-sided tests in most counties (80 percent), with the remaining 20 percent nearly evenly divided between negative and positive. This observation agrees with the fact that R is known from theory to be approximately unbiased for moderate or large sample sizes.

Table 7 shows why the negative bias noted above should not be a major concern with regard to potential use of SG or G. For each crop, the percent of counties for which the average underestimate (over simulation runs) was less than 10 percent and less than 20 percent (respectively) of the true yield is shown for all three methods. The table shows that the SG estimate was within 10 percent and 20 percent of the true yield with higher proportion than R for each crop. The G estimate was within 10 percent with higher proportion than R for all twelve crops and within 20 percent with higher proportion for all but two crops.

The findings documented in this section provide strong evidence that for a variety of crops grown in the lower 48 states, the Stasny-Goel method is more efficient than the ratio method with respect to bias, variance and outlier properties. Furthermore, SG outperformed G in all efficiency categories tested with the exception of variance.

5. ALGORITHM PERFORMANCE ISSUES

The capability of a county yield estimation method to produce accurate numbers in a consistent manner is very important in evaluating its potential for operational use. As mentioned earlier, convergence of the Stasny-Goel algorithm within a fixed limit on number of iterations is not guaranteed. While estimates are generally produced when the limit is reached without convergence, their accuracy must be questioned until proven otherwise.

Table 8 shows for each crop the overall percentage of simulation runs (all states combined) for which SG converged and produced an estimate, respectively. Convergence percentages are also shown for the Griffith algorithm, which always produced an estimate whether or not

Table 4: Summary of Wilcoxon Rank Sum Tests on Absolute Bias by Crop

Crop	SG vs. G			SG vs. R			G vs. R		
	No. Counties Favoring			No. Counties Favoring			No. Counties Favoring		
	SG	G	Neither	SG	R	Neither	G	R	Neither
Barley	128 (74%)	23 (13%)	22 (13%)	143 (82%)	15 (9%)	17 (10%)	61 (35%)	92 (53%)	20 (12%)
Corn	465 (62%)	204 (27%)	83 (11%)	641 (85%)	54 (7%)	61 (8%)	371 (49%)	331 (44%)	50 (7%)
Cotton (Upland)	41 (54%)	25 (33%)	10 (13%)	59 (78%)	10 (13%)	7 (9%)	40 (53%)	26 (34%)	10 (13%)
Dry Beans	37 (67%)	12 (22%)	6 (11%)	47 (84%)	4 (7%)	5 (9%)	24 (44%)	24 (44%)	7 (13%)
Oats	181 (61%)	89 (30%)	29 (10%)	229 (76%)	34 (11%)	40 (13%)	146 (49%)	129 (43%)	24 (8%)
Rye	12 (40%)	12 (40%)	6 (20%)	19 (63%)	3 (10%)	8 (27%)	16 (53%)	11 (37%)	3 (10%)
Sorghum	35 (56%)	22 (35%)	6 (10%)	41 (65%)	8 (13%)	14 (22%)	33 (52%)	20 (32%)	10 (16%)
Soybeans	275 (60%)	146 (32%)	34 (7%)	366 (80%)	53 (12%)	39 (9%)	210 (46%)	207 (45%)	38 (8%)
Sunflower	73 (66%)	28 (25%)	9 (8%)	94 (85%)	6 (5%)	11 (10%)	49 (45%)	54 (49%)	7 (6%)
Tobacco (Burley)	25 (45%)	21 (38%)	9 (16%)	59 (95%)	1 (2%)	2 (3%)	48 (87%)	6 (11%)	1 (2%)
Spring Wheat	103 (75%)	23 (17%)	12 (9%)	108 (78%)	21 (15%)	9 (7%)	33 (24%)	89 (64%)	16 (12%)
Winter Wheat	308 (61%)	138 (27%)	60 (12%)	365 (72%)	81 (16%)	63 (12%)	220 (43%)	234 (46%)	52 (10%)
All	1683 (62%)	743 (27%)	286 (11%)	2171 (79%)	290 (11%)	276 (10%)	1251 (46%)	1223 (45%)	238 (9%)

Table 5: Summary of Wilcoxon Rank Sum Test Cases by Year

Year	SG vs. G			SG vs. R			G vs. R		
	No. Cases Favoring			No. Cases Favoring			No. Cases Favoring		
	SG	G	Tied	SG	R	Tied	G	R	Tied
2002	30	14	2	42	4	0	24	20	2
2003	33	4	0	38	0	0	15	22	0
Total	63 (76%)	18 (22%)	2 (2%)	80 (95%)	4 (5%)	0 (0%)	39 (47%)	42 (51%)	2 (2%)

Table 6: Summary of Wilcoxon Signed Rank Tests on Median Error (ME) By Year

Year	Stasny-Goel			Griffith			Ratio		
	Counts of Results			Counts of Results			Counts of Results		
	Neg. ME	Pos. ME	Zero ME	Neg. ME	Pos. ME	Zero ME	Neg. ME	Pos. ME	Zero ME
2002	770	516	116	751	585	52	165	133	1104
2003	837	371	127	705	589	30	127	112	1096
Total	1607 (59%)	887 (32%)	243 (9%)	1456 (54%)	1174 (43%)	82 (3%)	292 (11%)	245 (9%)	2200 (80%)

Table 7: Percent of Counties with Average Underestimate (AU) Less Than 10% and 20% of True Yield

Crop	AU < 10%			AU < 20%		
	SG	G	R	SG	G	R
Barley	81.0	62.2	46.3	97.7	94.2	84.6
Corn	82.9	71.4	41.9	98.1	94.0	82.5
Cotton (Upland)	78.95	78.4	64.5	100.0	95.95	96.05
Dry Beans	94.6	74.1	62.5	100.0	100	98.2
Oats	70.5	53.6	21.1	96.95	85.8	74.9
Rye	41.4	51.7	13.3	96.55	100.0	73.33
Sorghum	52.4	40.7	11.1	87.3	78.0	38.1
Soybeans	84.3	75.6	62.45	98.9	96.9	94.5
Sunflower	80.0	63.5	49.55	96.4	93.3	73.0
Tobacco (Burley)	92.7	98.1	27.3	100.0	100.0	92.7
Spring Wheat	93.9	54.8	53.6	99.2	87.1	88.4
Winter Wheat	85.8	74.7	51.5	97.8	94.2	90.0

convergence occurred. The three crops for which the combined convergence proportion over all test cases exceeded 90 percent were barley, soybeans and sunflower. The three crops showing lowest overall convergence percentage were tobacco, rye and spring wheat. However, SG was able to generate an estimate for most of the non-convergent simulation runs. Table 9 provides combined convergence and 'estimates produced' percentages by prevalence class. Note the discrepancy in convergence percentage of the SG algorithm between highly prevalent crops (class A) and less prevalent ones (B and C).

The enhancement to the algorithm mentioned earlier (allowing the program to continue beyond the maximum allowable number of iterations if the log-likelihood is highest at that point) did cause some previously non-convergent simulation runs to converge at a later iteration.

An interesting question that relates directly to the potential use of the Stasny-Goel program in operational county estimation is how the SG estimates produced in the absence of convergence compare with corresponding ratio estimates. If they could be shown to be equally or more efficient, the operational use of such numbers when convergence cannot be achieved might be justified. To that end, six crop/state/year combinations for which a sizable number of runs had failed to converge previously were selected for further simulation and evaluation.

While in theory the log-likelihood measure associated with the EM algorithm must increase with each successive iteration, numerical conditions can arise in actual practice that cause it to decrease from one iteration to the next. Such situations are often associated with non-convergence of the algorithm (as in the six SG cases just mentioned). Under those circumstances, it is reasonable to surmise that the iteration for which the computed log-likelihood is maximized will provide a better estimate than the final allowable iteration.

To investigate whether that's the case, code was added to the SG program to keep track of which iteration maximizes the log-likelihood and rerun the algorithm to that point if convergence is not achieved within the preset limit. If the iteration that maximizes the log-likelihood coincides with the maximum allowable one, the algorithm is allowed to continue until either convergence occurs or the log-likelihood decreases from one iteration to the next. In the latter situation, the estimate produced at the next-to-last iteration (highest log-likelihood) is used.

In the upcoming discussion, the estimate generated at the final allowable iteration (5,000) is referred to as SG(1) and the one computed at the iteration where the log-likelihood was highest as SG(2). Both types of estimate were compared with the corresponding ratio estimates. For each test case, the same number of simulations (250) was used as in the full scale study. The six test cases were Colorado barley (2002), North Dakota dry beans (2002), Ohio oats (2002), Oklahoma rye (2003), Mississippi soybeans (2002) and New York winter wheat (2002). The number of non-convergent simulation runs tested ranged from 37 (for barley in Colorado) to 105 (soybeans in Mississippi).

Table 10 shows the results of pairwise Wilcoxon rank sum tests on absolute bias. In five of six cases for SG(1) and all six for SG(2), the mean absolute error was found to be significantly lower than that of R more often than significantly higher. The comparison between SG(1) and SG(2) was favorable to the latter more often than the former, although in most cases neither method showed significantly lower MAE than the other. These findings suggest that the Stasny-Goel method can improve upon ratio estimation even in cases where convergence does not occur within a specified maximum allowable number of iterations.

Table 8: Algorithm Performance Statistics by Crop

Crop	**Stasny-Goel**		**Griffith**
	Percent Converged	Percent Estimates Produced	Percent Converged
Barley	93	99	68
Corn	87	99	77
Cotton (Upland)	81	84	89
Dry Beans	89	100	75
Oats	80	95	71
Rye	74	100	83
Sorghum	85	96	66
Soybeans	93	100	73
Sunflower	90.5	99.6	80
Tobacco (Burley)	41	74	52
Spring Wheat	63	100	52.5
Winter Wheat	88	99.7	65

Table 9: Algorithm Performance Statistics by Prevalence Class

Class	Stasny-Goel		Griffith
	Percent Converged	Percent Estimates Produced	Percent Converged
A	92	98	77
B	78	99.5	63
C	80	90	74
All	85	97	71

Table 10: Results of Pairwise Wilcoxon Rank Sum Tests on Absolute Bias for Non-Convergent Simulation Runs

Crop	State	Year	SG(1) vs. Ratio			SG(2) vs. Ratio			SG(1) vs. SG(2)		
			No. Counties Favoring			No. Counties Favoring			No. Counties Favoring		
			SG(1)	R	Neither	SG(2)	R	Neither	SG(1)	SG(2)	Neither
Barley	CO	2002	5	2	1	6	2	0	0	0	8
Dry Beans	ND	2002	9	0	17	17	1	8	0	10	16
Oats	OH	2002	7	10	22	17	11	11	1	18	20
Rye	OK	2003	7	0	6	9	0	4	0	1	12
Soybeans	MS	2002	23	0	2	24	0	1	0	13	12
W. Wheat	NY	2002	11	2	9	13	4	5	4	1	17
Total			62 (47%)	14 (11%)	57 (43%)	86 (65%)	18 (14%)	29 (22%)	5 (4%)	43 (32%)	85 (64%)

6. SUMMARY

A ten state simulation study comparing the efficiency of three county crop yield estimators for various crops was planned and carried out. The results showed the Stasny-Goel method to be superior to the standard ratio and Griffith methods for most efficiency categories, in particular absolute bias. Both model-based methods showed lower variance overall than the ratio method, with G usually having lower variance than SG. In a convergence study involving six test cases, SG was found to produce more efficient estimates than the ratio method even when convergence was not achieved within 5,000 iterations. For those reasons, the Stasny-Goel method has been selected for integration into NASS's operational County Estimates System.

REFERENCES

Crouse, C. (2000). *Evaluation of the Use of Spatial Modeling to Improve County Yield Estimation*. Research Report No. RDD-00-05, National Agricultural Statistics Service, U.S. Department of Agriculture.

Griffith, D. (2001). *Model-Based Small Geographic Area Estimation: A Comparison of Alternative Methodologies (Progress Report, Part II)*. Syracuse University Technical Report.

Iwig, W. (1993). "The National Agricultural Statistics Service County Estimates Program." In *Indirect Estimators in Federal Programs*, Statistical Policy Working Paper 21, Report of the FCSM, Subcommittee on Small Area Estimation, Washington, DC 7.1-7.15.

Stasny, E., Goel, P, Cooley, C. and Bohn, L. (1995), *Modeling County-Level Crop Yield with Spatial Correlations Among Neighboring Counties*. Technical Report No. 570, Department of Statistics, The Ohio State University.