United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research and
Development Division
Washington DC 20250

RDD Research Report
RDD-07-05

August 2007

Comparison of Methods for Estimating Crop Yield at the County Level

Michael E. Bellow

EXECUTIVE SUMMARY

Estimation of agricultural commodities at the county and district levels is an important program of the USDA's National Agricultural Statistics Service (NASS). Such estimates are in heavy demand by users in government, the private sector and the academic community. In particular, county estimation of crop yields has received increasing attention over recent years. Yield estimation is more challenging than estimation of area planted or harvested in a crop due in part to the higher variability of yields from one year to the next. Ratio estimation has always been used by the Agency to derive yield numbers.

For some time, NASS has been interested in the potential of model-based small area estimation methods to improve upon the standard ratio estimator. Stasny, Goel et al. (1995), working under a cooperative agreement between NASS and the Ohio State University, developed a Bayesian county yield estimation method that takes into account spatial correlation among neighboring counties in a mixed effects model. Through a NASS cooperative agreement with Syracuse University, Griffith proposed an alternative method that involves an autoregressive model and employs Box-Cox and Box-Tidwell transformations. Both methods invoke an iterative algorithm and are capable of generating estimates for counties lacking positive survey data for a crop.

In this report, the Stasny-Goel (SG), Griffith (G) and simple ratio (R) methods are compared for a number of crops in ten geographically dispersed states using simulated data sets. The states in the project area were Colorado, Florida, Michigan, Mississippi, New York, North Dakota, Ohio, Oklahoma, Tennessee and Washington. 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 SG, G and R methods were compared for the 2002 and 2003 Quarterly Agricultural Survey (QAS) cycles. Efficiency measures used to evaluate the estimators included absolute bias, variance, mean square error and outlier metrics.

Results of the study indicated that the Stasny-Goel estimator was more efficient than the ratio estimator in all efficiency categories and superior to the Griffith estimator in most categories. Griffith's method showed lowest variance among the three most of the time, while both model-based methods were less outlier prone than the ratio method.

Reliability of convergence is a key issue if the Stasny-Goel method is to be adopted for use in operational NASS county estimation. The percentage of simulations where convergence occurred within the allowable number of iterations varied considerably, tending to be highest for the most prevalent crops in a state. Six crop/state/year cases where the algorithm failed to converge within the preset limit for a significant number of simulations were selected for further study. The SG estimates produced for non-convergent runs were compared with the corresponding ratio estimates and found to be superior despite the lack of convergence. An approach involving rerunning the algorithm to the point of highest log-likelihood instead of using the estimate from the maximum allowable iteration appeared to further improve estimation efficiency.

RECOMMENDATIONS

Based on the results documented in this report, the following recommendations are made:

1) Adopt the current version of the Stasny-Goel method for operational use by NASS Field Offices.

2) Investigate the convergence situation further to determine if there's a better solution than using estimates produced for non-convergent runs.

3) Explore the possibility of further enhancements to the Stasny-Goel algorithm. Perhaps some useful features of Griffith's procedure (such as missing data imputation) could be incorporated into the software.

These recommendations are discussed in more detail in Section 6.

Comparison of Methods for Estimating Crop Yield at the County Level

Michael E. Bellow[1]

Abstract

County level estimates of various agricultural commodities published by USDA's National Agricultural Statistics Service (NASS) are in heavy demand by users in government, the private sector and the academic community. In particular, accurate small area estimation of crop yields has become increasingly important over recent years. While NASS has traditionally used ratio estimation to derive yield numbers, model-based methods that make efficient use of available data sources hold the promise of significant improvement over the standard approach.

Stasny, Goel and other researchers at the Ohio State University developed a Bayesian mixed-effects county yield estimation algorithm with a spatial component involving correlations among neighboring counties. Griffith (at Syracuse University) proposed an alternative method involving Box-Cox and Box-Tidwell transformations in conjunction with an autoregressive model. This report documents a simulation study where the Stasny-Goel method, Griffith method and standard ratio estimation were compared for twelve crops in ten geographically dispersed states.

The Stasny-Goel method was found to be more efficient overall than either the ratio or Griffith method. The two model-based approaches and the simulation techniques used to compare them are described in some detail, followed by a discussion of results of the study. Convergence issues associated with the Stasny-Goel algorithm are also addressed, in particular the question of whether acceptable estimates can be produced in cases where the algorithm fails to converge within a preset upper limit on number of iterations.

Key Words: small area estimation; spatial modeling; simulation; convergence

[1] Michael E. Bellow is a Mathematical Statistician with the National Agricultural Statistics Service, Research & Development Division, 3251 Old Lee Highway, Room 305, Fairfax, VA 22030.

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 generated 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 used. Statisticians at the FOs use the outputs of this system to set final (official) county estimates.

When estimating area planted or harvested in a crop, the availability of reliable administrative data has been very important. Since planted area seldom varies dramatically from year to year, the estimation process is generally straightforward and repeatable. On the other hand, accurate estimation of crop yields at the county level has always been more difficult for the following reasons: 1) lack of reliable administrative data, 2) tendency of crop yields to fluctuate over time, and 3) lack of adequate survey data. County yield estimates are scrutinized heavily by crop insurance firms and other data users. Prior to 2002, NASS computed county estimates based on a non-probability sample of farms with little nonresponse follow-up and differential sampling/response rates for small vs. large farms. Since that sampling procedure precluded the use of standard small area estimation techniques based on known selection probabilities, NASS was motivated to pursue research into the potential application of model-based methodology to county level estimation.

NASS Field Offices conduct 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. The final county estimates must be consistent with district and state level figures published by NASS.

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 area from year to year. Furthermore, it does not make use of data from any county other than the one being estimated. Thus an estimate for a given county cannot be generated in the absence of 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 report.

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 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 among neighboring counties.

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

Griffith (1999), through a cooperative agreement between NASS and Syracuse University, proposed an alternative spatial county yield estimation method that predicts yield values using the published number of farms producing the crop of interest. Box-Cox and Box-Tidwell transformations are employed in conjunction with an autoregressive specification so as to optimize agreement with model assumptions. The sample data are used to project final estimates via back-transformed expected values. Estimates for counties with missing survey data can be computed via an imputation routine that utilizes the spatial correlation among neighboring counties as well as previous year county level data. Griffith (2001) identified this imputation capability (not tested in this study) as an advantage of his method over the one proposed by Stasny, Goel et al.

Both the Stasny-Goel (SG) and Griffith (G) algorithms are programmed in the SAS IML language. SG was coded originally in FORTRAN at Ohio State University and later converted to SAS. Some modifications to Griffith’s original program were necessary for the purposes of the study described in this report.

Crouse (2000) conducted the first evaluation of the Stasny-Goel method using simulated survey data, comparing it with the ratio method for estimation of county level corn and barley yields in Michigan. The SG method was found to produce more consistent estimates than the ratio (R) method across samples and performed better with respect to R for corn (a prevalent crop in Michigan) than for barley (much less common in that state). Crouse listed the following six tasks that needed to be done before this method could be considered for implementation in NASS FOs:

1) Perform additional testing to assess how well the method works for various crops in agriculturally diverse regions of the U.S.

2) Develop a scheme to identify problem survey records in the event that the SG algorithm fails to converge within a reasonable number of iterations.

3) Identify a method for obtaining previous

year county estimates to be used in the current year's estimation process.

4) Develop a method for integrating the SG algorithm into the NASS County Estimates System (CES) so that the computation of yield indications is transparent to the user.

5) Document the technical details of the algorithm for future reference by users.

6) Evaluate alternative methods or possible improvements to the SG method.

Items 1 and 6 represent the main impetus for the research documented in this report, while the convergence issue mentioned in item 2 is addressed in Section 6. Obtaining previous year county numbers (item 3) is no longer an issue due to the ready availability of online sources such as NASS's Data Warehouse and the Published Estimates Data Base. Items 4 and 5 are discussed in the final section of this report.

A ten state simulation study was conducted to compare the efficiency of the Stasny-Goel, Griffith and ratio estimators of county level yield. 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 simulation methodology is described in Section 3.

The states in the study area were selected for agricultural diversity. Each state falls in a different region from USDA's subdivision of the country: Colorado (Mountain), Florida (Southeast), Michigan (Lake), Mississippi (Delta), New York (Northeast), North Dakota (Northern Plains), Ohio (Corn Belt), Oklahoma (Southern Plains), Tennessee (Appalachia) and Washington (Pacific).

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. Appendix A provides official NASS state level estimates of production, harvested acreage and yield for all crops in the study in 2002 and 2003. Table 1 lists the specific crops tested in each state and also shows the prevalence class for all crop/state combinations.

2. DESCRIPTION OF METHODS

Table 2 shows the input data items required to compute county level yield estimates using the Stasny-Goel and Griffith methods, respectively. Data sources include the Quarterly Agricultural Survey (QAS), County Estimates Survey (CES), Census of Agriculture (COA) and Published Estimates Data Base (PEDB).

For a given state where either method is to be applied, an input file containing a two-column listing of pairs of counties in the state that share a common border is required. Both the Stasny-Goel and Griffith programs use this data set to form the *neighbor matrix*, an $nc \times nc$ array (where nc = number of counties in the state) with the entry in each row i, column j being 1 if the ith county (alphabetically within the state) is a neighbor of the jth county and 0 otherwise. Since each county is regarded as a neighbor of itself, all entries along the main diagonal are 1.

The Stasny-Goel method requires that post-stratification size groups be defined. Two criteria for defining the size groups were considered: 1) equal number of farms, and 2) equal land in farms. Calculating group boundaries so that the resulting groups had roughly equal land in farms turned out to more effective than the equal number of farms criterion in ensuring that each group contained at least one positive survey yield record for a given crop (an important consideration especially for less prevalent crops). The group definitions vary over states due to differences in average farm size. This fact is illustrated by Table 3, which compares the group boundaries for Colorado and Ohio in the project. The sizable discrepancy between the two states is attributable to farms in Colorado being much larger on average than farms in Ohio.

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

Table 2: Input Data for Stasny-Goel and Griffith Methods

Source	**Variable**	**Level**	**Year**	**Required For**	
				Stasny-Goel	Griffith
QAS	Production	Tract	Current	x	x
	Harvested Area	Tract	Current	x	x
	Total Land	Tract	Current	x	
CES	Production	County	Previous	x	x
		State	Current	x	
			Previous	x	
	Harvested Area	County	Previous	x	x
		State	Current	x	
			Previous	x	
COA	Final Nonresponse Weight	Record	Last Census	x	
	Total Land	Record	Last Census	x	
PEDB	Number of Farms	State	Current	x	
		State	Last Census	x	
	Total Land	State	Current	x	
		State	Last Census	x	
Other	Neighboring County Information	County	-	x	x
	County Area	County	-		x

Table 3: Example of Size Group Definitions for Two States

Group	State	
	Colorado	Ohio
1	FS < 3,180	FS < 320
2	3,180 <= FS < 11,000	320 <= FS < 1,020
3	FS >= 11,000	FS >= 1,020

(FS = farm size in acres)

For each county in a state, the Stasny-Goel 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 report.

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 ρ>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 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 algorithm uses an estimation and likelihood maximization process to adjust the estimates of group and county effects, variance and spatial correlation. 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 using the following formula:

$$\hat{y}_i = \sum_{j=1}^{G} w_{ij} (\hat{\mu} + \hat{\tau}_i + \hat{g}_j)$$

where:

G = number of size groups (2 or 3)

w_{ij} = post-stratification weight for ith county, jth size group

$\hat{\mu}$ = estimate of overall mean county yield

$\hat{\tau}_i$ = final (EM) estimate of random effect for ith county

$\hat{g}_j$ = final (EM) estimate of fixed effect for size group j

The Stasny-Goel program provides the user with an option to rescale the computed county yields to be consistent with official NASS state level yield estimates.

In the Griffith method (as adapted for this study), farm level QAS records of production and harvested acreage are first summed by county. The resulting production totals are then divided by the corresponding harvested acreage totals to obtain ratio estimates of county yield. Box-Cox transformations are applied to the three sets of data values for the purpose of stabilizing the variances. The Marquardt nonlinear estimation method (Marquardt, 1963) is then applied to a combined data set consisting of the Box-Cox transformed values and previous year production, harvested acreage and yield figures. The procedure involves the fitting of nonlinear models to derive relationships between the current and previous year variables. The fitted models define Box-Tidwell transformations of the form:

$$\begin{aligned} y_i &= (x_i + \delta)^{\gamma} \quad \text{if } \gamma > 0, \\ &= \log(x_i + \delta) \quad \text{if } \gamma = 0 \end{aligned}$$

where y is the dependent (current year) variable and x is the independent (previous year) variable. The program computes estimates of the parameters δ and γ.

The final step of the Griffith procedure is to estimate current year county yields via a spatial autoregressive model. The two independent variables are (Box-Tidwell) transformed previous year yield for a given county and the average transformed previous year yield for neighbors of the county. The model is once again fit using Marquardt iteration. As with the Stasny-Goel method, the final county yield estimates can be rescaled to agree with official state level figures.

3. SIMULATION METHODOLOGY

The simulation procedure used for the estimator comparison study was basically that employed by Crouse (2000), but with a few modifications. The NASS data sources needed to conduct the study were Quarterly Agricultural Survey data from 2002 and 2003, County Estimates Survey data from 2001-03 and Census of Agriculture data from 2002. 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. The CES data extracted from NASS's Data Warehouse provided previous year computed yields which served as initial values for the Stasny-Goel 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 in an effort to increase the spatial correlation of the simulated data so as to better model real survey data.

The regression equation used to generate replications of simulated data with three size groups has the following general form:

$$y_{ij}^{(k)} = \alpha + \beta_y Y_i + \beta_z Z_i + \beta_{s1}\psi_{1j} + \beta_{s2}\psi_{2j} + \varepsilon_{ij}^{(k)}$$

where:

$y_{ij}^{(k)}$ = yield value for simulation k, county i, survey record j

α, β's = regression parameters

Y_i = official NASS yield estimate for county i

Z_i = weighted average neighbor yield for county i

ψ_{gj} = 1 if record j in size group g (g=1, 2)
= 0 otherwise

$\varepsilon_{ij}^{(k)}$ = random error for simulation k, county i, record j

There was no need to include a size group indicator variable for group 3 since whether or not a given record belongs to it can be determined from the indicator variables for the first two groups. The random error term was assigned a normal distribution with mean zero and variance equal to the sample variance of QAS yield response values. For cases where two size groups were used instead of three, the following equation was applied:

$$y_{ij}^{(k)} = \alpha + \beta_y Y_i + \beta_z Z_i + \beta_{s1}\psi_{1j} + \varepsilon_{ij}^{(k)}$$

A very large number of simulated survey data sets (10,000) was generated 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.

As alluded to earlier, some revisions to the Griffith program were made to circumvent numerical problems that occasionally arose with the original code. Thus the method tested is a modified version of Griffith's procedure.

The maximum allowable number of iterations was set at 5,000 for both programs. 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 that prohibit 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. Due to this limitation, the capability of either SG or G to produce estimates in the absence of positive survey data for a county could not be tested in this study.

Three Census-based size groups were used for most crop/state/year combinations. 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 to get the SG program to run. Since the Griffith algorithm could not be run successfully for Ohio tobacco (2003), only the Stasny-Goel and ratio methods were compared for that crop/state/year combination.

An important aspect of the simulation process is ensuring that the simulated data sets accurately reflect the spatial correlation inherent in real survey data. Moran's I, a measure of spatial correlation (Moran, 1950), was computed for the original survey data sets and all simulated data sets for each test case. The tables in Appendix B compare the survey values with the average simulation values of Moran's I for all crop/state/year combinations in the study. The average simulation values were found to be within 0.1 of the survey values in nearly all cases and within 0.05 in most cases, so the simulation process appears to effectively model spatial correlation.

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 of the non-convergent Stasny-Goel simulation runs and all of the non-convergent Griffith runs. The reason for excluding such runs 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. Appendix F provides convergence statistics for the Stasny-Goel and Griffith algorithms.

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. 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. In other words, five percent of negative estimation errors are larger than the LTP and five percent of positive estimation errors exceed the UTP. High values of one or both of these measures suggest that the estimator in question is outlier prone.

Appendix C shows the overall pairwise results for all crops in the study. The entries in the SG column for the "SG vs. G" comparison for a given crop are the total number of counties (summed over states) for which SG had lower average absolute bias, variance, MSE, LTP or UTP than G in a given year, respectively (the remainder of each table is interpreted similarly). Combined totals and percentages for both years are also shown. While the pairwise comparisons are not rigorous statistical tests, they provide an indication as to which method may be best with respect to a given performance measure.

Tables 4 and 5 summarize the information from Appendix C by performance measure and crop, respectively. For each measure, Table 4 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 in an equal number of counties. Similarly, Table 5 displays for each crop the number of state/year/measure combinations favoring one method or the other.

From Table 4, both SG and G were appreciably better than R for all five efficiency 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 5 shows both SG and G to be better than R overall (five measures combined) for all twelve crops in the study. SG was superior to G for all crops with the exception of rye and soybeans, although there were too few cases for dry beans and tobacco to draw meaningful conclusions.

Table 6 summarizes crop/state/year cases (all measures combined) by prevalence class as defined in Section 1. Recall that class A contains the most prevalent crops, followed by B and C. Within each class, SG was superior to both G and R while G was better than R. This observation suggests that the relative performance of the three estimators is not strongly influenced by how common or rare a given crop may be in a state.

Table 4: 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 5: Summary of Pairwise Comparisons by Crop

Crop	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
Barley	29 (72.5%)	9 (22.5%)	2 (5%)	40 (100%)	0 (0%)	0 (0%)	26 (65%)	12 (30%)	2 (5%)
Corn	45 (56%)	32 (40%)	3 (4%)	80 (100%)	0 (0%)	0 (0%)	61 (76%)	17 (21%)	2 (3%)
Cotton (Upland)	11 (55%)	8 (40%)	1 (5%)	20 (100%)	0 (0%)	0 (0%)	18 (90%)	0 (0%)	2 (10%)
Dry Beans	6 (60%)	3 (30%)	1 (10%)	10 (100%)	0 (0%)	0 (0%)	9 (90%)	1 (10%)	0 (0%)
Oats	31 (56%)	21 (38%)	3 (5%)	55 (100%)	0 (0%)	0 (0%)	51 (93%)	3 (5%)	1 (2%)
Rye	6 (40%)	9 (60%)	0 (0%)	15 (100%)	0 (0%)	0 (0%)	15 (100%)	0 (0%)	0 (0%)
Sorghum	14 (56%)	11 (44%)	0 (0%)	23 (92%)	0 (0%)	2 (8%)	21 (84%)	3 (12%)	1 (4%)
Soybeans	24 (48%)	25 (50%)	1 (2%)	47 (94%)	3 (6%)	0 (0%)	44 (88%)	6 (12%)	0 (0%)
Sunflower	11 (55%)	7 (35%)	2 (10%)	20 (100%)	0 (0%)	0 (0%)	17 (85%)	2 (10%)	1 (5%)
Tobacco (Burley)	3 (60%)	2 (40%)	0 (0%)	10 (100%)	0 (0%)	0 (0%)	5 (100%)	0 (0%)	0 (0%)
Spring Wheat	14 (70%)	6 (30%)	0 (0%)	19 (95%)	1 (5%)	0 (0%)	13 (65%)	5 (25%)	2 (10%)
Winter Wheat	49 (65%)	25 (33%)	1 (1%)	72 (96%)	2 (3%)	1 (1%)	61 (81%)	13 (17%)	1 (1%)

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 'true' population values). This two-sample nonparametric procedure assesses whether the population medians of the two samples are significantly different from each other. The tests were performed on a pairwise basis at the ten percent significance level, with two one-sided tests done in each case. The null hypothesis for the one-sided tests was equality of median

Table 6: 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%)

absolute error (MAE) for the two methods. The alternative hypothesis for test A was the first method in the pair having a lower MAE than the second (vice versa for test B). The reason for using two one-sided tests instead of a single two-sided test is that the latter approach can only detect if one method has significantly different MAE than the other (not whether the MAE is lower or higher).

For each crop and pair of methods, Table 7 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 for all crops combined are also shown. Table 8 provides additional summary information, showing for each year the total number of crop/state cases for which the result favored one method (in each pair) for more counties than the other. The number of ties (i.e., cases where both methods were favored the same number of times) is also listed.

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 7 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 showed only a one percent advantage over R. Table 8 shows that in 95 percent of crop/state/year cases tested, the absolute bias of SG was significantly lower in more counties than that of R.

The Wilcoxon signed rank test, a one-sample nonparametric procedure that detects whether or not the median of a population is statistically different from zero, was run for each county where an estimate was produced. The objective was to assess whether the bias of the county estimators tended to be negative, zero or positive. Testing was performed on the simulated estimation errors, i.e., differences between the simulated estimates generated by each of the three methods and population 'true' county yields.

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 9 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. Appendix D provides a summary of the test results at the individual crop level.

Table 7: Summary of Wilcoxon Rank Sum Tests on Median Absolute Error by Crop

Crop	Year	**Stasny-Goel vs. Griffith**			**Stasny-Goel vs. Ratio**			**Griffith vs. Ratio**		
		No. Counties Favoring			No. Counties Favoring			No. Counties Favoring		
		SG	G	Neither	SG	R	Neither	G	R	Neither
Barley	2002	63	8	11	63	9	11	27	42	13
	2003	65	15	11	80	6	6	34	50	7
	Total	128 (74%)	23 (13%)	22 (13%)	143 (82%)	15 (9%)	17 (10%)	61 (35%)	92 (53%)	20 (12%)
Corn	2002	213	132	40	330	27	30	213	144	28
	2003	252	72	43	311	27	31	158	187	22
	Total	465 (62%)	204 (27%)	83 (11%)	641 (85%)	54 (7%)	61 (8%)	371 (49%)	331 (44%)	50 (7%)
Cotton (Upland)	2002	21	16	4	31	5	5	22	10	9
	2003	20	9	6	28	5	2	18	16	1
	Total	41 (54%)	25 (33%)	10 (13%)	59 (78%)	10 (13%)	7 (9%)	40 (53%)	26 (34%)	10 (13%)
Dry Beans	2002	16	5	4	22	2	2	15	8	2
	2003	21	7	2	25	2	3	9	16	5
	Total	37 (67%)	12 (22%)	6 (11%)	47 (84%)	4 (7%)	5 (9%)	24 (44%)	24 (44%)	7 (13%)
Oats	2002	103	62	17	131	29	26	96	71	15
	2003	78	27	12	98	5	14	50	58	9
	Total	181 (61%)	89 (30%)	29 (10%)	229 (76%)	34 (11%)	40 (13%)	146 (49%)	129 (43%)	24 (8%)
Rye	2002	4	5	2	5	2	4	4	5	2
	2003	8	7	4	14	1	4	12	6	1
	Total	12 (40%)	12 (40%)	6 (20%)	19 (63%)	3 (10%)	8 (27%)	16 (53%)	11 (37%)	3 (10%)
Sorghum	2002	30	19	3	36	6	10	28	17	7
	2003	5	3	3	5	2	4	5	3	3
	Total	35 (56%)	22 (35%)	6 (10%)	41 (65%)	8 (13%)	14 (22%)	33 (52%)	20 (32%)	10 (16%)
Soybeans	2002	135	79	21	175	47	15	119	94	22
	2003	140	67	13	191	6	24	91	113	16
	Total	275 (60%)	146 (32%)	34 (7%)	366 (80%)	53 (12%)	39 (9%)	210 (46%)	207 (45%)	38 (8%)
Sunflower	2002	33	16	2	46	3	3	23	25	3
	2003	40	12	7	48	3	8	26	29	4
	Total	73 (66%)	28 (25%)	9 (8%)	94 (85%)	6 (5%)	11 (10%)	49 (45%)	54 (49%)	7 (6%)
Tobacco (Burley)	2002	25	21	9	53	1	1	48	6	1
	2003	-	-	-	6	0	1	-	-	-
	Total	25 (45%)	21 (38%)	9 (16%)	59 (95%)	1 (2%)	2 (3%)	48 (87%)	6 (11%)	1 (2%)
Spring Wheat	2002	51	13	5	50	12	7	17	47	5
	2003	52	10	7	58	9	2	16	42	11
	Total	103 (75%)	23 (17%)	12 (9%)	108 (78%)	21 (15%)	9 (7%)	33 (24%)	89 (64%)	16 (12%)
Winter Wheat	2002	108	63	29	140	46	17	96	78	26
	2003	200	75	31	225	35	46	124	156	26
	Total	308 (61%)	138 (27%)	60 (12%)	365 (72%)	81 (16%)	63 (12%)	220 (43%)	234 (46%)	52 (10%)
All	2002	802	439	147	1082	189	131	708	547	133
	2003	881	304	139	1089	101	145	543	676	105
	Total	1683 (62%)	743 (27%)	286 (11%)	2171 (79%)	290 (11%)	276 (10%)	1251 (46%)	1223 (45%)	238 (9%)

Table 8: 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
Both	63 (76%)	18 (22%)	2 (2%)	80 (95%)	4 (5%)	0 (0%)	39 (47%)	42 (51%)	2 (2%)

Table 9: 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
Both	1607 (59%)	887 (32%)	243 (9%)	1456 (54%)	1174 (43%)	82 (3%)	292 (11%)	245 (9%)	2200 (80%)

Table 9 indicates that negative median error was concluded in 59 percent of all counties tested for SG and 54 percent for G. Zero median error was concluded by both one-sided tests in most counties (80 percent) for R, with the remaining 20 percent nearly evenly divided between negative and positive (agreeing with the fact that the ratio estimator is known is known from theory to be approximately unbiased for moderate or large sample sizes).

At the crop level (Appendix D), negative bias was concluded more often than positive bias for nine of the twelve crops with SG and eleven with G. The proportion of counties for which zero median error was concluded by both tests varied between 4 and 15 percent for SG and between 0 and 5 percent for G.

These findings suggest that the bias of both model-based estimators is generally negative. However, Table 10 shows why this observation 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 ten percent and less than twenty 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 all twelve crops. The G estimate was within 10 percent with higher proportion than R for all crops and within 20 percent with higher proportion for all but two crops.

Appendix E provides further insight into variability properties of the three estimators. Coefficients of variation computed over all usable simulation runs by crop, state and year are summarized in box plots. Counties with fewer than five positive survey records for a crop were not used in the computations in order to avoid ‘out of bounds’ CV values. For the Colorado winter wheat plots, CV values from Dolores County (2002) and Las Animas County (2003) were excluded due to low values of the Griffith estimates (which caused the CVs to exceed 100 percent in both cases). Note that the scale of the graphs varies, being tailored to the specific ranges of CV values.

The box plots illustrate the variance reduction achieved by using SG or G instead of R, as in general the CVs were highest for the latter. Furthermore, the interquartile and full ranges show more variability in the CVs for R than SG or G, suggesting that the two

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

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

model-based estimators are more stable with respect to variance than the ratio estimator.

A useful feature of the Stasny-Goel program is the computation of an estimate of root mean square error (RMSE) for each county having at least two positive records for a crop. In order to assess whether this analytic estimator is reasonable, it was compared with the square root of the simulation mean square error used in the pairwise comparisons discussed earlier. Table 11 shows the median (over states) correlations between the analytic and simulation RMSE values for each crop and year. The values were generally high enough to suggest that the RMSE estimates are valid, although there was one case (sunflower in 2002) where the correlation was very low.

The results 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. 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 specified 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. Occasionally, the SG program failed to produce an estimate due to numerical factors.

For each crop/state/year combination, the tables in Appendix F show the percentage of simulation runs for which SG converged and produced an estimate, respectively. Convergence percentages are also shown for the G algorithm, which always produced an estimate whether or not convergence occurred. Table 12 shows combined convergence and 'estimates produced' percentages by prevalence class and overall.

Note the discrepancy in convergence percentage of the Stasny-Goel algorithm between highly prevalent crops (class A) and less prevalent ones (B and C). From Appendix F, SG converged within 5,000 iterations 100 percent of the time in only 26 of 84 cases (31 percent). The three crops for

Table 11: Median Correlations Between Analytic and Simulation RMSE Values by Crop and Year

Crop	Median Correlation	
	2002	*2003*
Barley	0.64	0.78
Corn	0.65	0.7
Cotton (Upland)	0.86	0.83
Dry Beans	0.78	0.77
Oats	0.52	0.72
Rye	0.61	0.73
Sorghum	0.78	0.79
Soybeans	0.61	0.8
Sunflower	0.09	0.55
Tobacco (Burley)	0.31	0.5
Spring Wheat	0.41	0.33
Winter Wheat	0.5	0.63

Table 12: 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

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 for SG were tobacco, rye and spring wheat. However, SG was able to generate an estimate for most of the non-convergent simulation runs.

The most likely causes of convergence failure appear to be the presence of very few available yield reports for a given crop and the existence of one or more survey yield values that are much larger than the others in the same county. By removing two such problem records, Crouse (2000) was able to get the algorithm to converge quickly in a trial run for barley in Michigan that had previously gone through 50,000 iterations without convergence. He suggested that an automated procedure for detecting and removing problem records be developed. However, the same scheme that worked for Michigan barley was tried more recently for several other crops in states other than Michigan without achieving the same desirable result.

There may be no surefire way of getting the algorithm to converge other than perhaps allowing it to run for a nearly unlimited period of time (not feasible in operational practice). While weakening the convergence criteria may speed up convergence, that approach carries the risk of degrading the quality of the estimates.

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. In one case, the log-likelihood increased steadily over more than 20,000 additional iterations before the algorithm converged.

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 explore that possibility, code was added to the SG program to keep track of which iteration maximizes the log-likelihood and rerun the algorithm to that point when 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 (as discussed earlier). 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 Colorado barley) to 105 (Mississippi soybeans). Table 13 summarizes results of pairwise comparisons similar to those used in the full scale study, with absolute bias, variance, MSE, LTP and UTP of all three estimators compared for each county. For each of the six test cases, both SG(1) and SG(2) were clearly better than R in all five categories while SG(2) was superior to SG(1).

Pairwise Wilcoxon rank sum tests on absolute bias were also carried out, with the results shown in Table 14. The mean absolute error of SG(2) was found to be significantly lower than that of R more often than significantly higher for all six test cases, while SG(1) had significantly lower MAE than R more frequently than vice versa in five of the six cases. The comparison between SG(1) and SG(2) was favorable to the latter more often than the former, although in most cases neither method had a significant advantage in terms of MAE.

These findings suggest that estimates produced by the SG algorithm can improve upon ratio estimation even in cases where convergence does not occur within the maximum allowable number of iterations.

6. SUMMARY AND RECOMMENDATIONS

A ten state simulation study comparing the model-based Stasny-Goel and Griffith county crop yield estimators with the standard ratio estimator for various crops over two NASS estimation cycles was planned and carried out. The Stasny-Goel method was found to be best among the three in most efficiency categories. 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 better estimates than the ratio method even when convergence was not achieved within 5,000 iterations.

Table 13: Pairwise Comparison of Estimates for Non-Convergent Simulation Runs

State	Year	Measure	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	SG(2)	R	SG(1)	SG(2)
CO	2002	Absolute Bias	6	2	6	2	2	6
		Variance	6	2	7	1	0	8
		MSE	6	2	6	2	1	7
		LTP	7	1	7	1	2	6
		UTP	7	1	7	1	2	6
		All	32 (80%)	8 (20%)	33 (82.5%)	7 (17.5%)	7 (17.5%)	33 (82.5%)
ND	2002	Absolute Bias	23	3	23	3	5	21
		Variance	25	1	26	0	0	26
		MSE	25	1	25	1	3	23
		LTP	20	6	25	1	4	22
		UTP	20	6	25	1	2	24
		All	113 (87%)	17 (13%)	124 (95%)	6 (5%)	14 (11%)	116 (89%)
OH	2002	Absolute Bias	21	18	26	13	8	31
		Variance	35	4	39	0	0	39
		MSE	22	1	26	13	3	36
		LTP	22	17	31	8	4	35
		UTP	27	12	31	8	4	35
		All	127 (65%)	68 (35%)	153 (78%)	42 (22%)	19 (10%)	176 (90%)
OK	2003	Absolute Bias	11	2	11	2	4	9
		Variance	13	0	13	0	0	13
		MSE	11	2	11	2	2	11
		LTP	9	4	11	2	1	12
		UTP	12	1	12	1	2	11
		All	56 (86%)	9 (14%)	58 (89%)	7 (11%)	9 (14%)	56 (86%)
MS	2002	Absolute Bias	25	0	24	1	3	22
		Variance	25	0	25	0	0	25
		MSE	25	0	25	0	3	22
		LTP	22	3	23	2	3	22
		UTP	25	0	25	0	2	23
		All	122 (98%)	3 (2%)	122 (98%)	3 (2%)	11 (9%)	114 (91%)
NY	2002	Absolute Bias	18	4	16	6	8	14
		Variance	22	0	22	0	0	22
		MSE	19	3	18	4	8	14
		LTP	19	3	20	2	3	19
		UTP	21	1	20	2	8	14
		All	99 (90%)	11 (10%)	96 (87%)	14 (13%)	27 (25%)	83 (75%)

Table 14: 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
All			62 (47%)	14 (11%)	57 (43%)	86 (65%)	18 (14%)	29 (22%)	5 (4%)	43 (32%)	85 (64%)

Based on the findings documented in this report, the following recommendations are made:

1) Adopt the Stasny-Goel method for operational use by NASS Field Offices.

A previous version of the SG software was installed and tested in the Ohio, Michigan, Mississippi and Tennessee FOs during the 1999-2001 time period.

The simulation results suggest that the method could potentially improve upon ratio estimation for most crops in any region of the country.

Procedures will need to be developed for integrating the current version of SG into the County Estimates System, including a strategy for dealing with situations where the algorithm fails to produce a yield estimate for a county (if the problem cannot be resolved through modification of the program itself). Feedback from Field Office personnel involved with county estimation should be solicited. As mentioned earlier, the production of yield indications by the program should be transparent to the FO statistician using the code. To that end, a user document should be prepared that describes how to run the program and deal with various situations that may arise in practice.

2) Investigate the convergence issue further to determine if there's a better solution than using estimates produced for non-convergent runs.

Some further research into convergence properties of the SG algorithm may be called for to determine if a means of improving the convergence percentage can be found. However, the SG(1) and SG(2) schemes both outperformed the ratio estimator for non-convergent cases in the study described in Section 5.

The number of iterations that can be allowed in practice is an issue that will need to be worked out through consultation with FO staff. Time constraints may preclude very high limits. In general, the time required to run the algorithm on a desktop PC ranges from a few seconds to several minutes depending on the number of iterations required.

3) Explore the possibility of further enhancements to the Stasny-Goel method. Perhaps some aspects of Griffith's procedure (such as missing data imputation) could be incorporated into the software.

The modification discussed earlier, i.e., allowing the program to continue beyond the preset limit on number of iterations when the log-likelihood is maximized at that point, appears to be work well. In practice, a preset master limit on number of iterations allowable under any circumstances could be imposed. Griffith's method was competitive with the Stasny-Goel method for some crops tested in the study and for certain performance metrics (in particular, variance and outlier properties).

The G program includes a missing data imputation routine not tested in this study. SG uses the spatial aspect of its model to generate yield estimates for counties with no positive survey data without requiring imputation. If warranted, the two approaches to dealing with such counties could be compared via further simulations.

7. REFERENCES

Box, G.E.P. and Tidwell, P.W. (1962). "Transformations of the Independent Variables." Technometrics 4:531-550.

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.

Gibbons, J. (1985). Nonparametric Statistical Inference. Marcel Dekker, Inc., New York, NY.

Griffith, D. (1999). *A Methodology for Small Area Estimation with Special Reference to a One-Number Agricultural Census and Confidentiality: Results for Selected Major Crops and States*. Research Report No. RD-99-04, 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 Federal Committee on Statistical Methodology, Subcommittee on Small Area Estimation, Washington, DC 7.1-7.15.

Marquardt, D. (1963). "An Algorithm For Least-squares Estimation of Nonlinear Parameters." SIAM Journal of Applied Mathematics 11:431-441.

Moran, P.A.P. (1950), "Notes On Continuous Stochastic Phenomena." Biometrika 37:17-23.

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.

APPENDIX A. Official NASS State Level Statistics for Crops in Study Area

Table 1A: Barley

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Colorado	2002	7,488	72	104
	2003	8,938	82	109
Michigan	2002	663	13	51
	2003	784	14	56
North Dakota	2002	58,500	1,300	45
	2003	118,800	1,980	60
Washington	2002	19,040	340	56
	2003	14,570	310	47

Table 2A: Corn (For Grain)

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Colorado	2002	108,000	720	150
	2003	120,150	890	135
Michigan	2002	234,000	2,000	117
	2003	259,840	2,030	128
Mississippi	2002	63,600	530	120
	2003	71,550	530	135
New York	2002	44,620	460	97
	2003	53,240	440	121
North Dakota	2002	113,430	995	114
	2003	131,040	1,170	112
Ohio	2002	264,330	2,970	89
	2003	478,920	3,070	156
Oklahoma	2002	24,700	190	130
	2003	23,750	190	125
Tennessee	2002	65,270	610	107
	2003	81,220	620	131
Washington	2002	13,300	70	190
	2003	13,650	70	195

Table 3A: Cotton (Upland)

State	Year	Production (1000 bales*)	Harvested (1000 acres)	Yield (lbs/ac)
Florida	2002	96	105	439
	2003	117	92	610
Mississippi	2002	1,935	1,150	808
	2003	2,120	1,090	934
Tennessee	2002	818	530	741
	2003	890	530	806

* - one bale = 480 lbs

Table 4A: Dry Beans

State	Year	Production (1000 lbs)	Harvested (1000 acres)	Yield (lbs/ac)
North Dakota	2002	1,062,600	690	1,540
	2003	780,000	520	1,500

Table 5A: Rye

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
North Dakota	2002	210	7	30
	2003	750	15	50
Oklahoma	2002	1,300	65	20
	2003	1,540	70	22

Table 6A: Oats

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Colorado	2002	400	8	50
	2003	975	15	65
Michigan	2002	4,160	65	64
	2003	5,250	75	70
New York	2002	4,160	65	64
	2003	4,410	70	63
North Dakota	2002	12,600	300	42
	2003	21,240	360	59
Ohio	2002	3,355	55	61
	2003	3,960	60	66
Oklahoma	2002	740	20	37
	2003	900	25	36
Washington	2002	845	13	65
	2003	750	15	50

Table 7A: Sorghum (For Grain)

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Colorado	2002	1,800	90	20
	2003	4,320	160	27
Mississippi	2002	6,237	77	81
	2003	6,132	73	84
Oklahoma	2002	13,500	300	45
	2003	9,250	250	37
Tennessee	2002	2,080	26	80
	2003	3,280	40	82

Table 8A: Soybeans

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Michigan	2002	78,540	2,040	38.5
	2003	54,725	1,990	27.5
Mississippi	2002	43,840	1,370	32
	2003	55,770	1,430	39
New York	2002	4,608	144	32
	2003	4,830	138	35
Ohio	2002	151,040	4,720	32
	2003	164,780	4,280	38.5
Oklahoma	2002	6,760	260	26
	2003	6,370	245	26
Tennessee	2002	34,720	1,120	31
	2003	47,040	1,120	42

Table 9A: Sunflower (Oil and Non-Oil Varieties Combined)

State	Year	Production (1000 lbs)	Harvested (1000 acres)	Yield (lbs/ac)
Colorado	2002	49,860	70	712
	2003	118,330	118	1,003
North Dakota	2002	1,699,550	1,315	1,292
	2003	1,518,850	1,165	1,304

Table 10A: Tobacco (Air-Cured Light Burley)

State	Year	Production (1000 lbs)	Harvested (1000 acres)	Yield (lbs/ac)
Ohio	2002	9,625	5.5	1,750
	2003	8,745	5.3	1,650
Tennessee	2002	53,070	29	1,830
	2003	47,500	25	1,900

Table 11A: Spring Wheat

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
North Dakota	2002	165,200	5,900	28
	2003	252,800	6,400	39.5
Washington	2002	25,370	590	43
	2003	22,345	545	41

Table 12A: Winter Wheat

State	Year	Production (1000 bu)	Harvested (1000 acres)	Yield (bu/ac)
Colorado	2002	36,300	1,650	22
	2003	77,000	2,200	35
Michigan	2002	29,480	440	67
	2003	44,880	660	68
Mississippi	2002	7,200	180	40
	2003	6,125	125	49
New York	2002	6,844	118	58
	2003	6,360	120	53
North Dakota	2002	2,145	65	33
	2003	5,880	120	49
Ohio	2002	50,220	810	62
	2003	68,000	1,000	68
Oklahoma	2002	103,600	3,700	28
	2003	179,400	4,600	39
Tennessee	2002	14,100	300	47
	2003	13,500	270	50
Washington	2002	104,400	1,800	58
	2003	117,000	1,800	65

APPENDIX B. Moran's I Coefficient for Survey and Simulated Data

Table 1B: Barley

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.42	0.43
	2003	0.45	0.43
Michigan	2002	0.22	0.24
	2003	0.26	0.21
North Dakota	2002	0.76	0.71
	2003	0.59	0.55
Washington	2002	0.12	0.14
	2003	0.24	0.19

Table 2B: Corn

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.61	0.57
	2003	0.53	0.49
Michigan	2002	0.42	0.53
	2003	0.6	0.56
Mississippi	2002	0.45	0.47
	2003	0.46	0.41
New York	2002	0.43	0.36
North Dakota	2002	0.36	0.44
	2003	0.42	0.49
Ohio	2002	0.53	0.44
	2003	0.51	0.51
Oklahoma	2002	0.29	0.24
Tennessee	2002	0.2	0.23
	2003	0.31	0.27
Washington	2002	0.41	0.45
	2003	0.41	0.42

Table 3B: Cotton (Upland)

State	Year	Survey I	Avg. Sim. I
Florida	2002	0.32	0.33
Mississippi	2002	0.56	0.52
	2003	0.61	0.61
Tennessee	2002	0.66	0.64

Table 4B: Dry Beans

State	Year	Survey I	Avg. Sim. I
North Dakota	2002	0.58	0.48
	2003	0.49	0.54

Table 5B: Oats

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.31	0.25
	2003	0.21	0.22
Michigan	2002	0.41	0.38
New York	2002	0.43	0.37
North Dakota	2002	0.58	0.48
	2003	0.57	0.49
Ohio	2002	0.34	0.43
	2003	0.45	0.34
Oklahoma	2003	0.24	0.3
Washington	2002	0.21	0.22
	2003	0.14	0.15

Table 6B: Rye

State	Year	Survey I	Avg. Sim. I
North Dakota	2002	0.13	0.2
Oklahoma	2002	0.21	0.22
	2003	0.36	0.41

Table 7B: Sorghum

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.26	0.35
	2003	0.42	0.59
Mississippi	2002	0.48	0.47
Oklahoma	2002	0.31	0.26
Tennessee	2002	0.62	0.63

Table 8B: Soybeans

State	Year	Survey I	Avg. Sim. I
Michigan	2002	0.65	0.68
	2003	0.56	0.59
Mississippi	2002	0.48	0.43
	2002	0.31	0.35
New York	2002	0.37	0.36
Ohio	2002	0.4	0.38
	2003	0.41	0.37
Oklahoma	2002	0.26	0.23
Tennessee	2002	0.42	0.37
	2003	0.56	0.44

Table 9B: Sunflower

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.39	0.44
	2003	0.48	0.52
North Dakota	2002	0.41	0.46
	2003	0.58	0.53

Table 10B: Tobacco (Burley)

State	Year	Survey I	Avg. Sim. I
Ohio	2003	0.59	0.64
Tennessee	2002	0.69	0.62

Table 11B: Spring Wheat

State	Year	Survey I	Avg. Sim. I
North Dakota	2002	0.75	0.69
	2002	0.79	0.77
Washington	2002	0.59	0.59
	2003	0.44	0.4

Table 12B: Winter Wheat

State	Year	Survey I	Avg. Sim. I
Colorado	2002	0.41	0.35
	2003	0.32	0.32
Michigan	2002	0.63	0.64
Mississippi	2002	0.34	0.36
	2003	0.39	0.39
New York	2002	0.35	0.45
North Dakota	2002	0.18	0.2
	2003	0.37	0.4
Ohio	2002	0.54	0.48
	2003	0.62	0.55
Oklahoma	2003	0.25	0.31
Tennessee	2002	0.28	0.32
	2003	0.38	0.37
Washington	2002	0.54	0.54
	2003	0.4	0.45

APPENDIX C. Crop Level Pairwise Comparisons Between County Estimation Methods

Table 1C: Barley

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	72	10	72	11	33	49
	2003	70	21	86	6	42	49
	Total	142 (82%)	31 (18%)	158 (90%)	17 (10%)	75 (43%)	98 (57%)
Variance	2002	51	31	83	0	59	23
	2003	38	53	92	0	87	4
	Total	89 (51%)	84 (49%)	175 (100%)	0 (0%)	146 (84%)	27 (16%)
MSE	2002	69	13	75	8	33	49
	2003	64	27	86	6	43	48
	Total	133 (77%)	40 (23%)	161 (92%)	14 (8%)	76 (44%)	97 (56%)
LTP	2002	54	28	77	6	52	30
	2003	41	50	84	8	66	25
	Total	95 (55%)	78 (45%)	161 (92%)	14 (8%)	118 (68%)	55 (32%)
UTP	2002	52	30	72	11	54	28
	2003	54	37	90	2	67	24
	Total	106 (61%)	67 (39%)	162 (93%)	13 (7%)	121 (70%)	52 (30%)
All	2002	298	112	379	36	231	179
	2003	267	188	438	22	305	150
	Total	565 (65%)	300 (35%)	817 (93%)	58 (7%)	536 (62%)	329 (38%)

Table 2C: Corn

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	229	156	360	27	245	140
	2003	266	101	336	33	177	190
	Total	495 (66%)	257 (34%)	696 (92%)	60 (8%)	422 (56%)	330 (44%)
Variance	2002	137	248	387	0	378	7
	2003	109	258	368	1	345	22
	Total	246 (33%)	506 (67%)	755 (99.9%)	1 (0.1%)	723 (96%)	29 (4%)
MSE	2002	212	173	365	22	271	114
	2003	258	109	346	23	200	167
	Total	470 (62.5%)	282 (37.5%)	711 (94%)	45 (6%)	471 (63%)	281 (37%)
LTP	2002	138	247	366	21	315	70
	2003	174	193	335	34	261	106
	Total	312 (41%)	440 (59%)	701 (93%)	55 (7%)	576 (77%)	176 (23%)
UTP	2002	194	191	379	8	335	50
	2003	229	138	360	9	283	84
	Total	423 (56%)	329 (44%)	739 (98%)	17 (2%)	618 (82%)	134 (18%)
All	2002	910	1015	1857	78	1544	381
	2003	1036	799	1745	100	1266	569
	Total	1946 (52%)	1814 (48%)	3602 (95%)	178 (5%)	2810 (75%)	950 (25%)

Table 3C: Cotton (Upland)

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	22	19	35	6	28	13
	2003	22	13	30	5	20	15
	Total	44 (58%)	32 (42%)	65 (86%)	11 (14%)	48 (63%)	28 (37%)
Variance	2002	7	34	41	0	40	1
	2003	3	32	35	0	35	0
	Total	10 (13%)	66 (87%)	76 (100%)	0 (0%)	75 (99%)	1 (1%)
MSE	2002	22	19	36	5	31	10
	2003	20	15	32	3	23	12
	Total	42 (55%)	34 (45%)	68 (89%)	8 (11%)	54 (71%)	22 (29%)
LTP	2002	18	23	36	5	32	9
	2003	13	22	28	7	27	8
	Total	31 (41%)	45 (59%)	64 (84%)	12 (16%)	59 (78%)	17 (22%)
UTP	2002	23	18	40	1	36	5
	2003	17	18	34	1	33	2
	Total	40 (53%)	36 (47%)	74 (97%)	2 (3%)	69 (91%)	7 (9%)
All	2002	92	113	188	17	167	38
	2003	75	100	159	16	138	37
	Total	167 (44%)	213 (56%)	347 (91%)	33 (9%)	305 (80%)	75 (20%)

Table 4C: Dry Beans

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		**No. Counties Favoring**		**No. Counties Favoring**	
		SG	G	SG	R	G	R
Absolute Bias	2002	18	7	24	2	18	7
	2003	22	8	28	2	12	18
	Total	40 (73%)	15 (27%)	52 (93%)	4 (7%)	30 (55%)	25 (45%)
Variance	2002	2	23	26	0	25	0
	2003	9	21	30	0	30	0
	Total	11 (20%)	44 (80%)	56 (100%)	0 (0%)	55 (100%)	0 (0%)
MSE	2002	18	7	26	0	19	6
	2003	23	7	28	2	17	13
	Total	41 (75%)	14 (25%)	54 (96%)	2 (4%)	36 (65%)	19 (35%)
LTP	2002	14	11	25	1	23	2
	2003	21	9	29	1	21	9
	Total	35 (64%)	20 (36%)	54 (96%)	2 (4%)	44 (80%)	11 (20%)
UTP	2002	12	13	26	0	22	3
	2003	15	15	29	1	26	4
	Total	27 (49%)	28 (51%)	55 (98%)	1 (2%)	48 (87%)	7 (13%)
All	2002	64	61	127	3	107	18
	2003	90	60	144	6	106	44
	Total	154 (56%)	121 (44%)	271 (97%)	9 (3%)	213 (77%)	62 (23%)

Table 5C: Oats

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	107	75	156	30	115	67
	2003	81	36	111	6	60	57
	Total	188 (63%)	111 (37%)	267 (88%)	36 (12%)	175 (59%)	124 (41%)
Variance	2002	80	102	186	0	180	2
	2003	29	88	117	0	116	1
	Total	109 (36%)	190 (64%)	303 (100%)	0 (0%)	296 (99%)	3 (1%)
MSE	2002	104	78	161	25	128	54
	2003	77	40	112	5	63	54
	Total	181 (61%)	118 (39%)	273 (90%)	30 (10%)	191 (64%)	108 (36%)
LTP	2002	96	86	175	11	151	31
	2003	59	58	111	6	93	24
	Total	155 (52%)	144 (48%)	286 (94%)	17 (6%)	244 (82%)	55 (18%)
UTP	2002	69	113	170	16	166	16
	2003	61	56	109	8	95	22
	Total	130 (43%)	169 (57%)	279 (92%)	24 (8%)	261 (87%)	38 (13%)
All	2002	456	454	848	82	740	170
	2003	307	278	560	25	427	158
	Total	763 (51%)	732 (49%)	1408 (93%)	107 (7%)	1167 (78%)	328 (22%)

Table 6C: Rye

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	6	5	8	3	6	5
	2003	8	11	17	2	13	6
	Total	14 (47%)	16 (53%)	25 (83%)	5 (17%)	19 (63%)	11 (37%)
Variance	2002	4	7	11	0	10	1
	2003	19	0	18	1	19	0
	Total	23 (77%)	7 (23%)	29 (97%)	1 (3%)	29 (97%)	1 (3%)
MSE	2002	6	5	10	1	8	3
	2003	6	13	16	3	14	5
	Total	12 (40%)	18 (60%)	26 (87%)	4 (13%)	22 (73%)	8 (27%)
LTP	2002	8	3	11	0	10	1
	2003	4	15	16	3	16	3
	Total	12 (40%)	18 (60%)	27 (90%)	3 (10%)	26 (87%)	4 (13%)
UTP	2002	1	10	11	0	10	1
	2003	9	10	18	1	18	1
	Total	10 (33%)	20 (67%)	29 (97%)	1 (3%)	28 (93%)	2 (7%)
All	2002	25	30	51	4	44	11
	2003	46	49	85	10	80	15
	Total	71 (47%)	79 (53%)	136 (91%)	14 (9%)	124 (83%)	26 (17%)

Table 7C: Sorghum

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	31	21	44	8	34	18
	2003	6	5	9	2	8	3
	Total	37 (59%)	26 (41%)	53 (84%)	10 (16%)	42 (67%)	21 (33%)
Variance	2002	14	38	51	1	52	0
	2003	2	9	11	0	11	0
	Total	16 (25%)	47 (75%)	62 (98%)	1 (2%)	63 (100%)	0 (0%)
MSE	2002	29	23	45	7	36	16
	2003	3	8	8	3	9	2
	Total	32 (51%)	31 (49%)	53 (84%)	10 (16%)	45 (71%)	18 (29%)
LTP	2002	32	20	50	2	42	10
	2003	5	6	11	0	11	0
	Total	37 (59%)	26 (41%)	61 (97%)	2 (3%)	53 (84%)	10 (16%)
UTP	2002	18	34	47	5	49	3
	2003	2	9	6	5	8	3
	Total	20 (32%)	43 (68%)	53 (84%)	10 (16%)	57 (90%)	6 (10%)
All	2002	124	136	237	23	213	47
	2003	18	37	45	10	47	8
		142 (45%)	173 (55%)	282 (90%)	33 (10%)	260 (83%)	55 (17%)

Table 8C: Soybeans

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	138	97	188	49	145	90
	2003	144	76	213	8	109	111
	Total	282 (62%)	173 (38%)	401 (88%)	57 (12%)	254 (56%)	201 (44%)
Variance	2002	162	73	237	0	234	1
	2003	20	200	221	0	217	3
	Total	182 (40%)	273 (60%)	458 (100%)	0 (0%)	451 (99%)	4 (1%)
MSE	2002	127	108	191	46	158	77
	2003	131	89	215	6	119	101
	Total	258 (57%)	197 (43%)	406 (89%)	52 (11%)	277 (61%)	178 (39%)
LTP	2002	94	141	188	49	184	51
	2003	78	142	201	20	167	53
	Total	172 (38%)	283 (62%)	389 (85%)	69 (15%)	351 (77%)	104 (23%)
UTP	2002	114	121	232	5	204	31
	2003	127	93	221	0	171	49
	Total	241 (53%)	214 (47%)	453 (99%)	5 (1%)	375 (82%)	80 (18%)
All	2002	635	540	1036	149	925	250
	2003	500	600	1071	34	783	317
	Total	1135 (50%)	1140 (50%)	2107 (92%)	183 (8%)	1708 (75%)	567 (25%)

Table 9C: Sunflower

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	34	17	48	4	27	24
	2003	42	17	56	3	30	29
	Total	76 (69%)	34 (31%)	104 (94%)	7 (6%)	57 (52%)	53 (48%)
Variance	2002	21	30	52	0	50	1
	2003	41	18	59	0	53	6
	Total	62 (56%)	48 (44%)	111 (100%)	0 (0%)	103 (94%)	7 (6%)
MSE	2002	33	18	50	2	32	19
	2003	39	20	56	3	35	24
	Total	72 (65%)	38 (35%)	106 (95.5%)	5 (4.5%)	67 (61%)	43 (39%)
LTP	2002	25	26	50	2	41	10
	2003	37	22	57	2	46	13
	Total	62 (56%)	48 (44%)	107 (96%)	4 (4%)	87 (79%)	23 (21%)
UTP	2002	19	32	50	2	42	9
	2003	28	31	51	8	47	12
	Total	47 (43%)	63 (57%)	101 (91%)	10 (9%)	89 (81%)	21 (19%)
All	2002	132	123	250	10	192	63
	2003	187	108	279	16	211	84
	Total	319 (58%)	231 (42%)	529 (95%)	26 (5%)	403 (73%)	147 (27%)

Table 10C: Tobacco (Burley)

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	31	24	54	1	50	5
	2003	-	-	7	0	-	-
	Total	31 (56%)	24 (44%)	61 (98%)	1 (2%)	50 (91%)	5 (9%)
Variance	2002	27	28	55	0	54	1
	2003	-	-	7	0	-	-
	Total	27 (49%)	28 (51%)	62 (100%)	0 (0%)	54 (98%)	1 (2%)
MSE	2002	29	26	55	0	51	4
	2003	-	-	7	0	-	-
	Total	29 (53%)	26 (47%)	62 (100%)	0 (0%)	51 (93%)	4 (7%)
LTP	2002	17	38	55	0	53	2
	2003	-	-	7	0	-	-
	Total	17 (31%)	38 (69%)	62 (100%)	0 (0%)	53 (96%)	2 (4%)
UTP	2002	38	17	54	1	53	2
	2003	-	-	7	0	-	-
	Total	38 (69%)	17 (31%)	61 (98%)	1 (2%)	53 (96%)	2 (4%)
All	2002	142	133	273	2	261	14
	2003	-	-	35	0	-	-
	Total	142 (52%)	133 (48%)	308 (99%)	2 (1%)	261 (95%)	14 (5%)

Table 11C: Spring Wheat

Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	52	17	55	14	21	48
	2003	56	13	59	10	27	42
	Total	108 (78%)	30 (22%)	114 (83%)	24 (17%)	48 (35%)	90 (65%)
Variance	2002	48	21	69	0	44	25
	2003	37	32	69	0	65	4
	Total	85 (62%)	53 (38%)	138 (100%)	0 (0%)	109 (79%)	29 (21%)
MSE	2002	54	15	57	12	23	46
	2003	56	13	60	9	27	42
	Total	110 (80%)	28 (20%)	117 (85%)	21 (15%)	50 (36%)	88 (64%)
LTP	2002	57	12	68	1	45	24
	2003	38	31	69	0	43	26
	Total	95 (69%)	43 (31%)	137 (99%)	1 (1%)	88 (64%)	50 (36%)
UTP	2002	33	36	58	11	46	23
	2003	32	37	59	10	47	22
	Total	65 (47%)	73 (53%)	117 (85%)	21 (15%)	93 (67%)	45 (33%)
All	2002	244	101	307	38	179	166
	2003	219	126	316	29	209	136
	Total	463 (67%)	227 (33%)	623 (90%)	67 (10%)	388 (56%)	302 (44%)

Table 12C: Winter Wheat

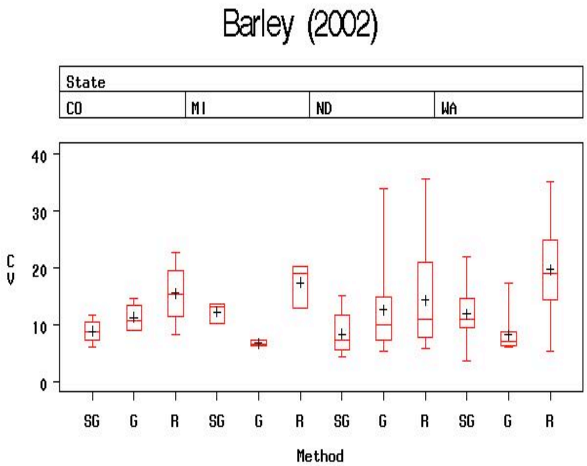
Measure	Year	SG vs. G		SG vs. R		G vs. R	
		No. Counties Favoring		No. Counties Favoring		No. Counties Favoring	
		SG	G	SG	R	G	R
Absolute Bias	2002	128	72	155	48	116	84
	2003	206	100	265	41	149	157
	Total	334 (66%)	172 (34%)	420 (83%)	89 (17%)	265 (52%)	241 (48%)
Variance	2002	76	124	203	0	190	10
	2003	141	165	306	0	302	4
	Total	217 (43%)	289 (57%)	509 (100%)	0 (0%)	492 (97%)	14 (3%)
MSE	2002	122	78	163	40	128	72
	2003	200	106	273	33	161	145
	Total	322 (64%)	184 (36%)	436 (86%)	73 (14%)	289 (57%)	217 (43%)
LTP	2002	102	98	184	19	156	44
	2003	150	156	268	38	226	80
	Total	252 (50%)	254 (50%)	452 (89%)	57 (11%)	382 (75%)	124 (25%)
UTP	2002	109	91	174	29	167	33
	2003	161	145	283	23	255	51
	Total	270 (53%)	236 (47%)	457 (90%)	52 (10%)	422 (83%)	84 (17%)
All	2002	537	463	879	136	757	243
	2003	858	672	1395	135	1093	437
	Total	1395 (55%)	1135 (45%)	2274 (89%)	271 (11%)	1850 (73%)	680 (27%)

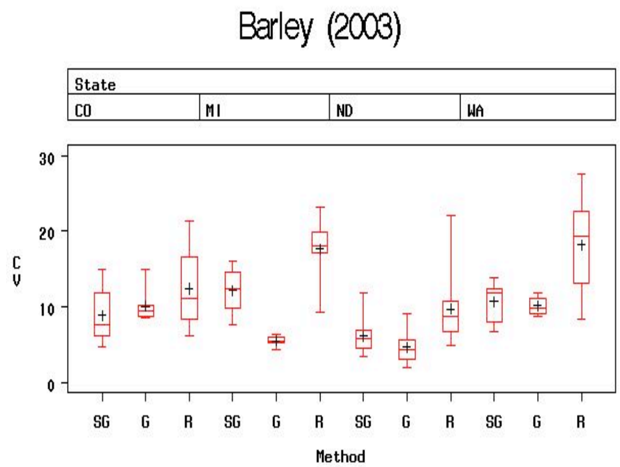
APPENDIX D. Wilcoxon Signed Rank Tests on Median Error (ME)

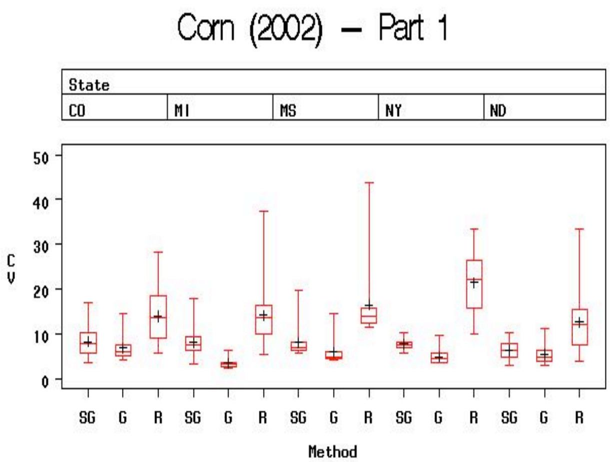
Table 1D. Summary of Test Results by Crop

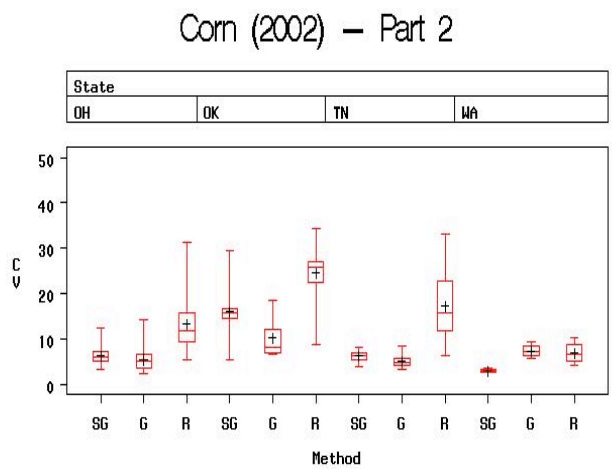
Crop	Year	Stasny-Goel			Griffith			Ratio		
		Counts of Results			Counts of Results			Counts of Results		
		ME<0	ME>0	ME=0	ME<0	ME>0	ME=0	ME<0	ME>0	ME=0
Barley	2002	38	38	7	39	36	7	7	10	66
	2003	52	29	11	45	44	2	12	5	75
	Total	90 (51%)	67 (38%)	18 (10%)	84 (49%)	80 (46%)	9 (5%)	19 (11%)	15 (9%)	141 (81%)
Corn	2002	228	129	30	199	172	14	48	43	296
	2003	255	79	35	208	147	12	43	33	293
	Total	483 (64%)	208 (28%)	65 (9%)	407 (54%)	319 (42%)	26 (3%)	91 (12%)	76 (10%)	589 (78%)
Cotton (Upland)	2002	34	4	3	26	14	1	2	6	33
	2003	27	8	0	23	11	1	1	3	31
	Total	61 (80%)	12 (16%)	3 (4%)	49 (64%)	25 (33%)	2 (3%)	3 (4%)	9 (12%)	64 (84%)
Dry Beans	2002	15	10	1	16	9	0	3	1	22
	2003	16	11	3	19	10	1	5	3	22
	Total	31 (55%)	21 (37.5%)	4 (7%)	35 (64%)	19 (35%)	1 (2%)	8 (14%)	4 (7%)	44 (79%)
Oats	2002	82	90	14	106	70	6	23	10	153
	2003	54	52	11	58	57	2	7	10	100
	Total	136 (45%)	142 (47%)	25 (8%)	164 (55%)	127 (42%)	8 (3%)	30 (10%)	20 (7%)	253 (83.5%)
Rye	2002	3	8	0	6	5	0	1	2	8
	2003	11	4	4	13	5	1	1	1	17
	Total	14 (47%)	12 (40%)	4 (13%)	19 (63%)	10 (33%)	1 (3%)	2 (7%)	3 (10%)	25 (83%)
Sorghum	2002	18	29	5	28	24	0	8	2	42
	2003	1	10	0	4	7	0	1	0	10
	Total	19 (30%)	39 (62%)	5 (8%)	32 (51%)	31 (49%)	0 (0%)	9 (14%)	2 (3%)	52 (83%)
Soybeans	2002	159	59	19	135	93	7	23	25	189
	2003	170	25	26	115	101	4	23	13	185
	Total	329 (72%)	84 (18%)	45 (10%)	250 (55%)	94 (43%)	11 (2%)	46 (10%)	38 (8%)	374 (82%)
Sunflower	2002	27	17	8	31	17	3	7	4	41
	2003	26	27	6	28	28	3	5	4	50
	Total	53 (48%)	44 (40%)	14 (13%)	59 (54%)	45 (41%)	6 (5%)	12 (11%)	8 (7%)	91 (82%)
Tobacco (Burley)	2002	37	14	4	28	24	3	7	3	45
	2003	3	0	4	-	-	-	0	0	7
	Total	40 (65%)	14 (23%)	8 (13%)	28 (51%)	24 (44%)	3 (5%)	7 (11%)	3 (5%)	52 (84%)
Spring Wheat	2002	22	36	11	34	30	5	10	5	54
	2003	23	36	10	32	37	0	5	7	57
	Total	45 (33%)	72 (52%)	21 (15%)	66 (48%)	67 (49%)	5 (100%)	15 (11%)	12 (9%)	111 (80%)
Winter Wheat	2002	107	82	14	103	91	6	26	22	155
	2003	199	90	17	160	142	4	24	33	249
	Total	306 (60%)	172 (34%)	31 (6%)	263 (52%)	233 (46%)	10 (2%)	50 (10%)	55 (11%)	404 (79%)

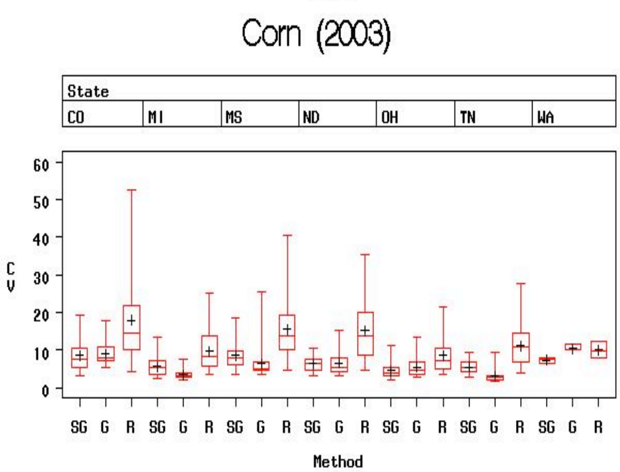
APPENDIX E. Box Plots of Coefficient of Variation



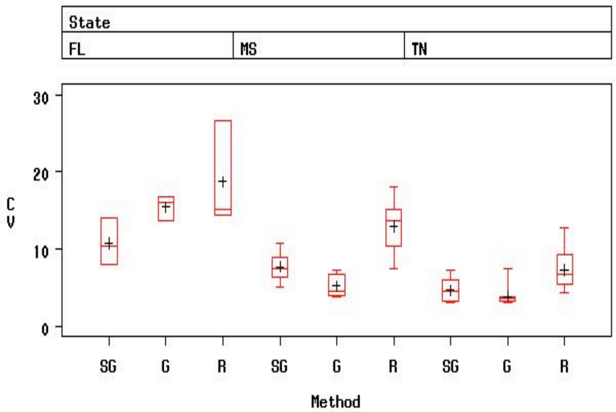




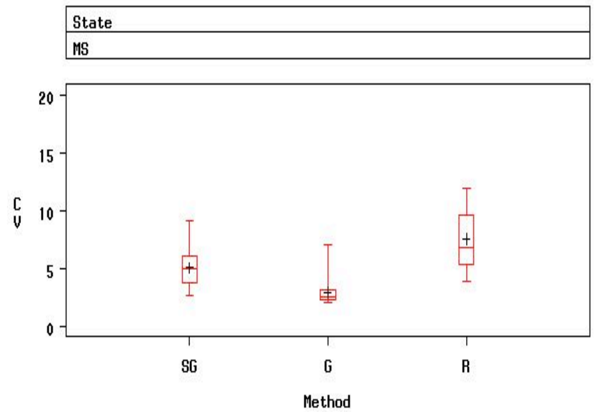




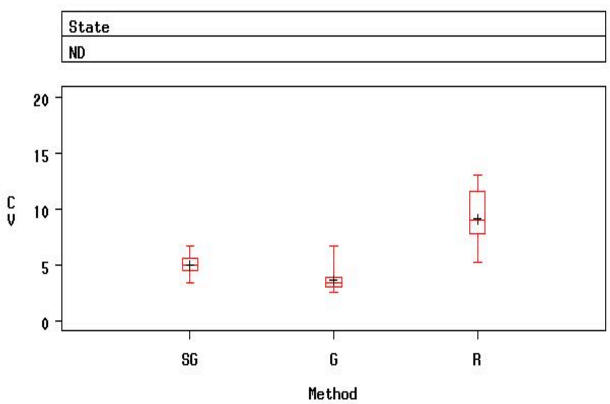
Cotton (2002)



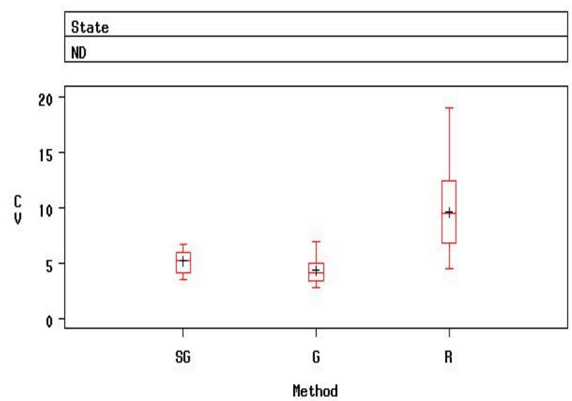
Cotton (2003)



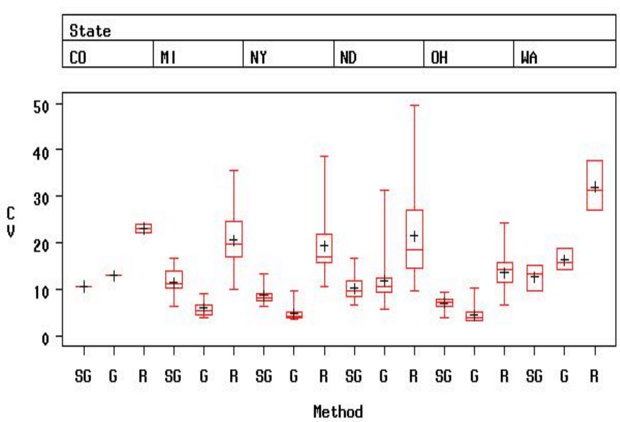
Dry Beans (2002)



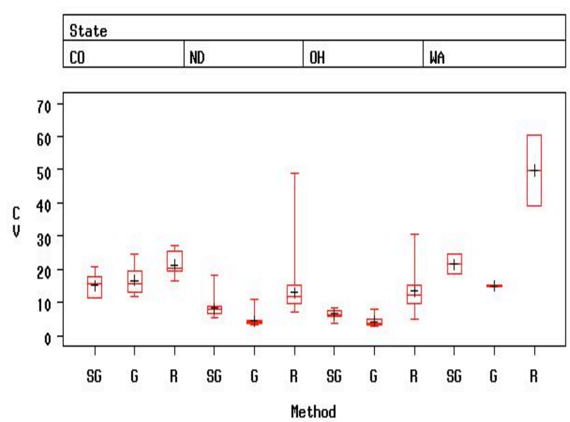
Dry Beans (2003)

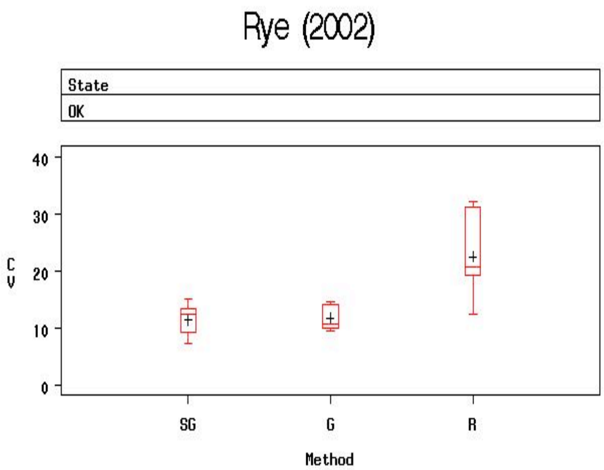


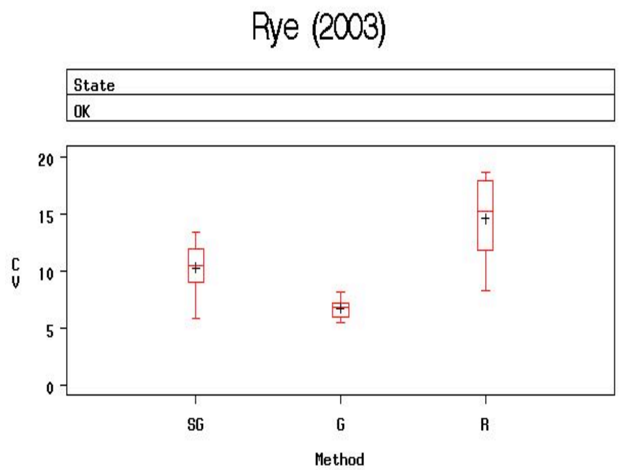
Oats (2002)

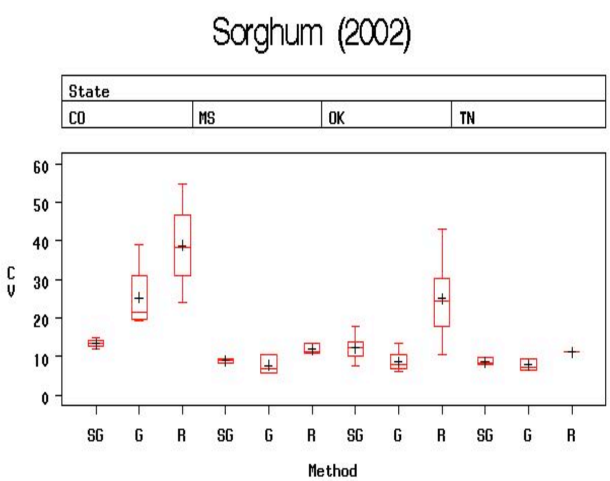


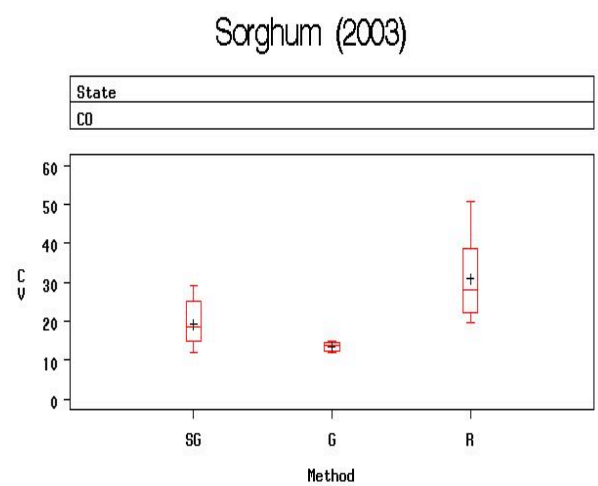
Oats (2003)

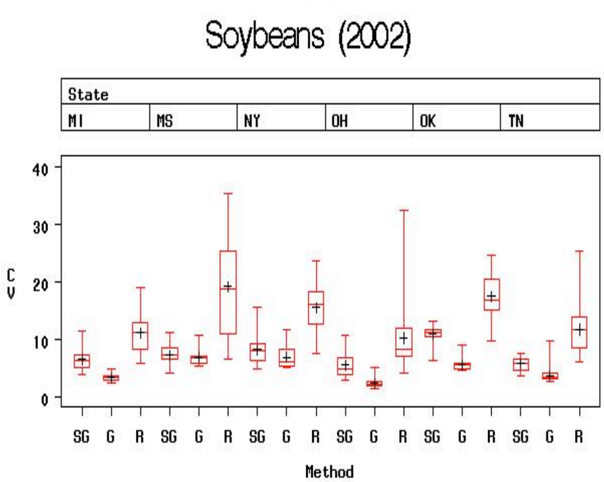


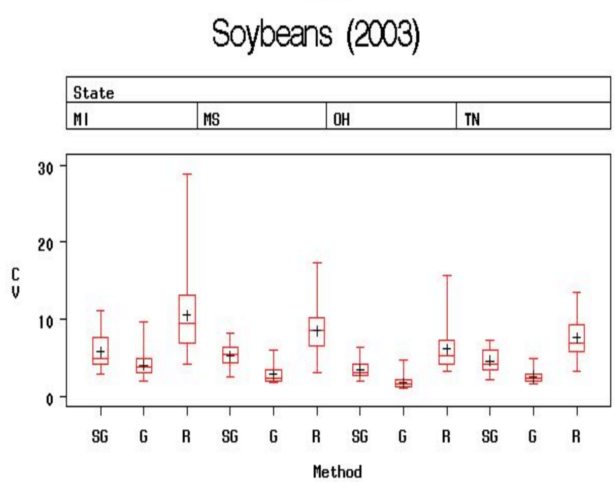

Rye (2002)
State
OK
CV
Method
SG
G
R


Rye (2003)
State
OK
CV
Method
SG
G
R

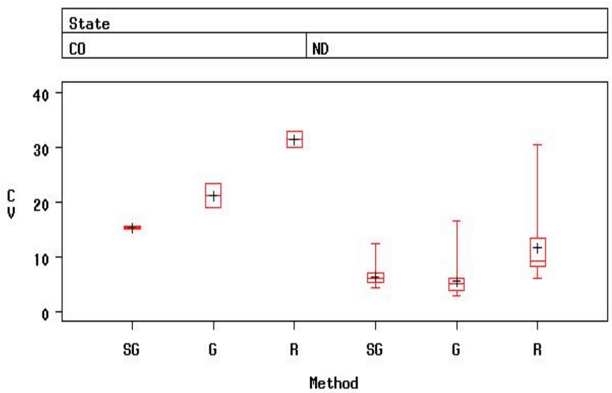

Sorghum (2002)
State
CO
MS
OK
TN
CV
Method


Sorghum (2003)
State
CO
CV
Method
SG
G
R

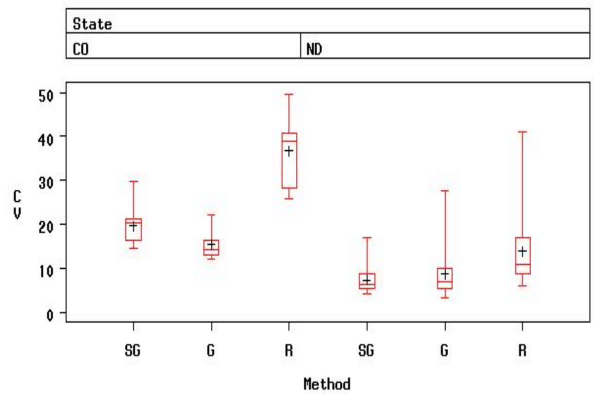

Soybeans (2002)
State
MI
MS
NY
OH
OK
TN
CV
Method


Soybeans (2003)
State
MI
MS
OH
TN
CV
Method

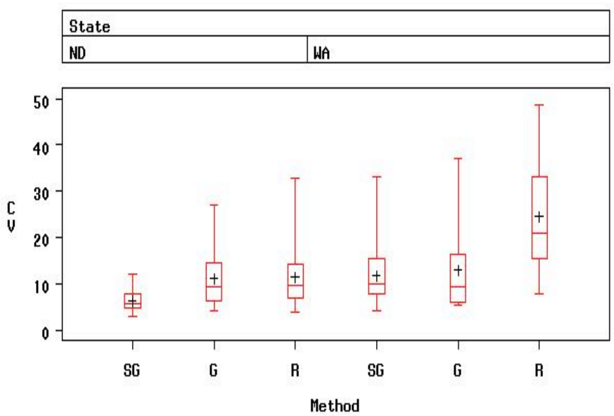
Sunflower (2002)



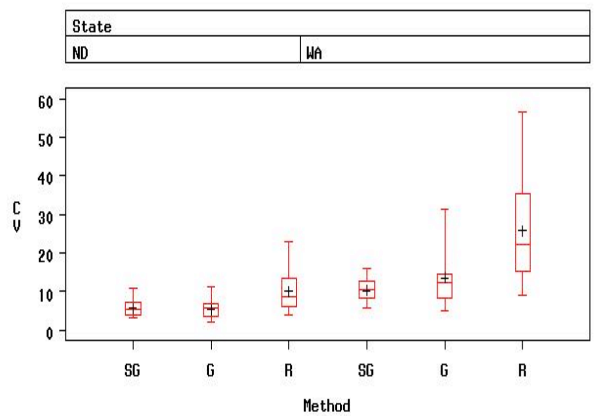
Sunflower (2003)



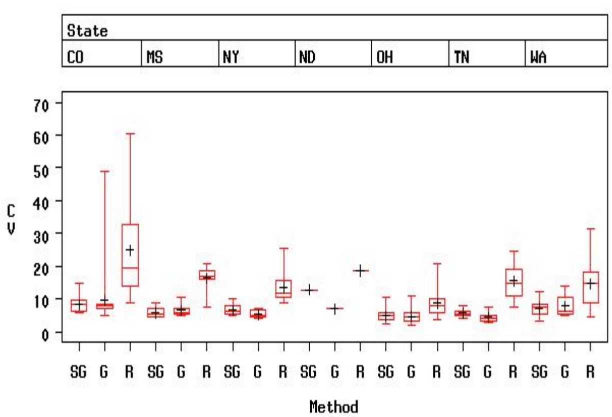
Spring Wheat (2002)



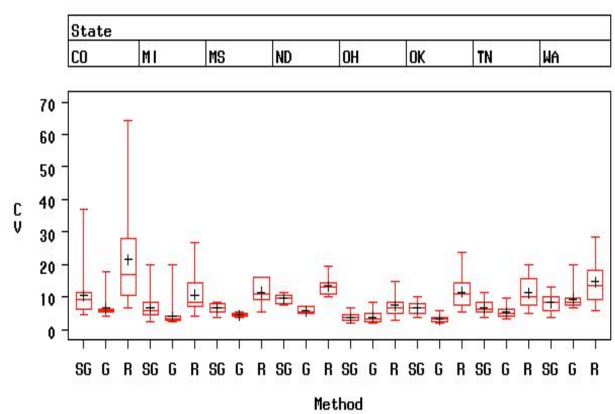
Spring Wheat (2003)



Winter Wheat (2002)



Winter Wheat (2003)



APPENDIX F. Algorithm Performance Statistics

Table 1F: Barley

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	82	98	76
	2003	100	100	92
Michigan	2002	84	100	68
	2003	96	100	82
North Dakota	2002	100	100	92
	2003	100	100	84
Washington	2002	96	100	26
	2003	85	94	24
All		93	99	68

Table 2F: Corn

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	99.6	100	74
	2003	12	100	87
Michigan	2002	100	100	95
	2003	100	100	95
Mississippi	2002	77	98	84
	2003	83	100	97
New York	2002	98	100	78
North Dakota	2002	96	100	83
	2003	100	100	66
Ohio	2002	100	100	33
	2003	100	100	83
Oklahoma	2002	84	100	80
Tennessee	2002	95	96	69
	2003	98	98	90
Washington	2002	70	90	53
	2003	78	98	64
All		87	99	77

Table 3F: Cotton (Upland)

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Florida	2002	74	77	70
Mississippi	2002	99.6	100	99
	2003	90	100	100
Tennessee	2002	58	58	86
All		81	84	89

Table 4F: Dry Beans

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
North Dakota	2002	84	100	69
	2003	94	100	80
All		89	100	75

Table 5F: Oats

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	93	100	83
	2003	90	99.6	86
Michigan	2002	14	46	73
New York	2002	70	100	86
North Dakota	2002	100	100	79
	2003	100	100	40
Ohio	2002	79	100	72
	2003	100	100	74
Oklahoma	2003	85	100	59
Washington	2002	66	96	66
	2003	81	99	61
All		80	95	71

Table 6F: Rye

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
North Dakota	2003	100	100	83
Oklahoma	2002	47	100	78
	2003	76	100	88
All		74	100	83

Table 7F: Sorghum

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	78	98	73
	2003	84	100	40
Mississippi	2002	78	91	66
Oklahoma	2002	99.6	100	67
Tennessee	2002	85	90	81
All		85	96	66

Table 8F: Soybeans

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Michigan	2002	100	100	80
	2003	100	100	62
Mississippi	2002	52	100	80
	2003	100	100	87
New York	2002	96	100	93
Ohio	2002	100	100	65
	2003	100	100	68
Oklahoma	2002	100	100	70
Tennessee	2002	100	100	50
	2003	82	100	79
All		93	100	73

Table 9F: Sunflower

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	79	99.6	74
	2003	83	99	79
North Dakota	2002	100	100	92
	2003	100	100	75
All		90.5	99.6	80

Table 10F: Tobacco (Burley)

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent (Estimates Produced)	Percent Converged
Ohio	2003	47	47	-
Tennessee	2002	34	100	52
All		41	74	52

Table 11F: Spring Wheat

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
North Dakota	2002	100	100	96
	2003	100	100	86
Washington	2002	15	100	23
	2003	38	100	5
All		63	100	52.5

Table 12F: Winter Wheat

State	Year	Stasny-Goel		Griffith
		Percent Converged	Percent Estimates Produced	Percent Converged
Colorado	2002	99	100	11
	2003	96	100	27
Michigan	2003	90	100	93
Mississippi	2002	77	96	64
	2003	77	99	67
New York	2002	83	100	86
North Dakota	2002	73	99.6	78
	2003	98	100	60
Ohio	2002	100	100	83
	2003	100	100	89
Oklahoma	2003	100	100	58
Tennessee	2002	63	100	93
	2003	84	100	90
Washington	2002	98	100	40
	2003	82	100	30
All		88	99.7	65