# Spatial Yield Risk Issues: Comparing Yield Risk Across Region, Crop and Aggregation Method

Michael Popp
Associate Professor, University of Arkansas, Fayetteville, AR, 72701 (mpopp@uark.edu)
Margot Rudstrom
Assistant Professor, University of Minnesota, Morris, MN, 56267 (rudstrmv@mrs.umn.edu)
Patrick Manning
Research Specialist, University of Arkansas, Fayetteville, AR, 72701 (pmanning@uark.edu)

***Selected Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Denver, Colorado, August 1-4, 2004***

**Spatial Yield Risk Issues:
Comparing Yield Risk Across Region, Crop and Aggregation Method**

ABSTRACT

Crop yield risk analysis is difficult since historic field level yields are often not available. Spatially aggregated yield data are available, however, but aggregation distortion for farm level analysis may exist. This paper addresses how much aggregation distortion to expect and offers some adjustment solutions across crops and production regions.

Key words: spatial crop yield risk; data aggregation; spatial analysis

## Spatial Yield Risk Across Region, Crop And Aggregation Method

### INTRODUCTION

Economists interested in capturing yield risk have a range of data sources to choose from. For example, they can utilize experimental data from field research plots, GIS data collected from harvesting equipment to capture spatial variation within a field or even published data from various yield reporting agencies. A trend in these examples is an ever-increasing level of spatial aggregation. While the data source often dictates the types of decisions that can be analyzed with the data, researchers are often faced with a lack of farm level yield data and therefore choose, perhaps second best, spatially aggregated yield data for farm level analyses.

With an ever increasing emphasis on risk management in agriculture (Harwood et al., 1999) analysts attempting to capture an individual decision maker's production and price risk may therefore be introducing biased research results if aggregation distortions exist. Table 1 highlights the issue that decision makers face when using temporal yield risk measures obtained from aggregated yield data (AYV[1]) rather than a less biased average of yield risk (EYV) from individual fields. Use of AYV compared to EVY can lead to over- or under-estimation of yield variance estimates using aggregate data as shown by the ratios ($R_1$) presented in Table 1. Considerable research exists to test for the amount of distortion that is introduced when using aggregate data for farm level decisions (Bechtel and Young 1999; Debrah and Hall, 1989; Wang and Zhang 2002; Rudstrom et al., 2002). Some aggregation adjustments are therefore necessary to reflect differences between aggregate and farm level data to avoid biased results (Fulton, King and Fackler 1988; Popp, Dalsted and Skold 1997; Skees and Nutt in Mapp and Jeter 1988).

It is the lack of a data adjustment process that has prompted an extension of the research provided by Rudstrom et al. (2002). At the municipality level, the AYV for wheat underestimated the EYV for quarter-section data for each of the nineteen municipalities analyzed. Statistical clustering of FVEs into groups of similar FVE resulted in i) differences in $R_1$ values across clusters (with no relationship between the distortion and the level of risk); and ii) no distinctive spatial patterns in FVE measures. While distortions at the municipality level of aggregation always had $R_1$ values less than 1, at the smaller cluster level of aggregation, $R_1$ values were sometimes greater than 1. Aggregation bias for FVE thus appeared to favor underestimation of yield risk the greater the number of observations aggregated and/or the greater the range in individual variance estimates observed.

Given the above distortion results caused by aggregating from quarter section to municipality (or geographically disbursed clusters) for wheat, questions about the robustness of these findings across crops and other production regions arose. Since wheat is considered a relatively low risk crop in comparison to other crops like canola and flax (Popp and Rudstrom, 2000), one objective of this paper is to provide further empirical evidence of the type of distortion that can be expected between quarter section yield data and data that have been aggregated to some degree for crops other than wheat. A second question relates to the robustness of the aggregation bias across different production regions. Do production regions with different soil types and weather patterns exhibit similar aggregation bias? A final hypothesis is whether the metric used for clustering affects aggregation distortions and spatial patterns in yield risk.

## BACKGROUND

Row crop production is an important component of agriculture in Manitoba. In 2001 there were 3.7 million acres of red spring wheat, 436,000 acres of flax and 1.9 million acres of canola planted in the province (Statistics Canada, 2001). In terms of farm participation, 44 percent of farms produced spring wheat, 30 percent produced canola and 13 percent of farms produced flax.

Crop producers can use crop insurance to help management their risk. A crown corporation, Manitoba Crop Insurance Corporation (MCIC), is responsible for all crop risk insurance products. Payout to producers is based on either a producer's long-term yield history or the risk area average yield if the producer does not have a yield history. Adjustments are made to individual producer premiums based on his long-term average yield relative to the risk area long-term average yield, or the individual producer index.

Data aggregation is done at the farm level when determinations of insurance payouts are made. In addition to average crop yield, farmers are therefore also concerned with yield variability. Further, it is the temporal variance in farm level yields that is likely the basis for making crop acreage allocation decisions. That is, how does average annual yield on their fields vary and is the risk acceptable for the farm operation? The question related to data aggregation is how does the average of the temporal variation of fields (EYV) compare to the variance of the average annual yield across fields (AYV)? Since the later is most often used in the absence of producer data, the question of how to adjust for this distortion across crops and space is an important one.

Since growing conditions for crops vary across Manitoba, the province is divided into a number of risk areas. Each risk area has relatively similar growing conditions and soil types. The Red River Valley Risk Area 12 is a relatively low risk production area of heavier Osborne clay soils. It generally receives adequate precipitation at 20 inches annually with 15 inches from April to September. The region averages 1,750 - 1,800 growing degree days[2] annually (Manitoba Cooperator). Plant moisture stress is the difference between the amount of water a crop can potentially use and the amount of water it actually gets from planting to maturity. In Risk Area 12, the moisture stress for spring wheat ranges from -0.2 to -0.4 inches. Other parts of the province have more challenging growing conditions. Risk area 4 is located in Central Manitoba around the city of Brandon and compared to Risk area 12, is characterized by sandier soils. The average growing degree days in this region are 1,600 to 1,650. The moisture stress index for spring wheat is -0.6 to -0.8 inches.

Field level crop production data was obtained from MCIC. The data are reported on a filed level and a field is identified by a legal descriptor. The legal descriptor provides the location of the quarter section in which the field is located. Annual production information is recorded for each field that is insured by MCIC. It is possible that a municipality is not entirely contained within one risk area. For this analysis, two municipalities entirely within each of the two risk areas are used to test the hypotheses related to variance distortion across crops, production regions and clustering metric.

## METHODOLOGY

Similar to the procedure described in Rudstrom et al. (2002), it is possible to group or cluster quarter sections that have similar yield variation across time in order to be able to discern spatial patterns in yield risk across a production region. If spatial patterns are evident, clustering also allows for some insights on the range of aggregation distortion to expect across different areas of aggregation and or modifications to current risk areas considered to be similar in terms of yield risk.

Cluster analysis allows objects to be placed in groups, such that objects in the groups are similar. In this case, objects are quarter sections of land and the clustering statistics used to arrange the objects are the crop yield coefficient of variation (FCV) in addition to the FVE as used by Rudstrom et al. (2002). Since different crops are evaluated and since average yields vary across production regions, it is expected that variance alone is not appropriate for comparison of risk within the crop across regions or across crops and regions. A relative risk measure (FCV) is therefore chosen to cluster quarter sections into groups of like relative risk.

A number of clustering procedures are available and *k*-means, non-hierarchal clustering is used here (Johnson and Wichern, 1998). The quarter sections are partitioned into *k* clusters, where the number of clusters (*k*) is specified in advance. An initial set of *k* quarter sections is selected as seeds or starting points for the clusters. Using these seeds, a quarter section is assigned to a cluster whose centroid, or mean of the clustering statistic is nearest. Cluster centroids are recalculated after quarter section cluster reassignments to closer cluster centroids. The process is repeated until no more reassignments take place or the distance between cluster centroids and quarter sections assigned to different clusters is minimized. Since the technique is somewhat sensitive to the seeds used as starting points as well as the number of seeds to use, the pseudo F-statistic is used to provide some guidance (Milligan and Cooper 1985).

Once the clustering was performed, aggregation distortion statistics ($R_1$ = AYV / EYV and $R_2$ = ACV / ECV) were calculated similar to the procedure shown in Table 1 for each of the municipalities and clusters within the municipalities. To test for the third hypothesis of differences in clustering across clustering metric, the difference between cluster numbers, determined using FCV compared to FVE, were calculated and plotted for each quarter section. Note that cluster numbers using either clustering statistic were assigned in an increasing order of magnitude (i.e. a low cluster number had lowest risk observations whereas a high cluster number had highest risk observations). In cases where the two clustering statistics lead to a different number of clusters in a municipality, clusters with fewest observations were combined into a single cluster to make the number of clusters the same across clustering statistic. Spatial differences in cluster number assignments were subsequently plotted using GIS Arcview v 8.3.

## DATA MANAGEMENT

Crop yield data from MCIC was obtained for 1980-1990 for Manitoba. Annual per acre yields for wheat, flax and canola were recorded for each field that was insured by MCIC. Municipalities 510 and 561 are entirely contained in Risk Area 12 and municipalities 621 and 971 are within Risk Area 4 boundaries. Wheat and canola were clustered for all 4 municipalities. Flax clustering was not performed for Risk Area 4 due to the insufficient number of quarter section observations for flax in municipalities 621 and 971.

Long run averages of acreage allocated to wheat, flax and canola are approximately 40, 16 and 20 percent, respectively. In order to eliminate quarter sections were the crop is not typically grown, quarter sections with two low a frequency of production were eliminated. For wheat and flax this meant using quarter sections where the crop was grown for at least 4 of the 11 years and for canola this meant using quarter sections where the crop was grown for at least 3 of the 11 years. These restrictions fit with typical crop rotations in the area and remove yield risk bias that may be introduced if yield observations were included where the crop is typically not grown as a result of performing poorly or for other reasons.

While a quarter section is an area of 160 acres, fields could be less than 160 acres. It was possible to have, for example, two 80-acre fields on a quarter section. When there were multiple fields of a single crop on a quarter section in any given year, the simple average of the fields was calculated and reported for those quarter sections.

Using the annual field level harvested yield, the average annual yield in the municipality or cluster was calculated for each of the three crops. The statistics were calculated for both an entire municipality as well as individual clusters within the municipality. Table 2 presents aggregate and average field statistics for the municipalities analyzed. Counter to expectations, average yield for canola was one bushel less in the Risk Area 12 municipalities than the Risk Area 4 municipalities. Wheat, by contrast, exhibited higher yields in Risk Area 12 than 4. In terms of relative risk, flax was more risky than canola or wheat in Risk Area 12. For each crop, the AYV statistics are less than EYV statistics. This indicates again that yield risk is underestimated using aggregate data. In terms of relative yield risk (ACV), flax is more risky than canola or wheat using aggregate statistics. Using the ECV statistics in comparison to the ACV statistics, the relative risk position of the crops is less clear.

## RESULTS AND CONCLUSIONS

Aggregation bias for wheat led to questions about the robustness of Rudstrom et al.'s results. Specifically these questions were i) do crops other than wheat exhibit similar aggregation bias? ii) is the aggregation bias similar across production regions characterized by different resource conditions? and iii) does clustering by relative risk compared to absolute risk result in different spatial patterns and aggregation bias? Tables 3 to 6 and Figure 1 summarize the data to provide insights to the above questions.

Using non-clustered aggregate data for wheat tends to underestimate risk (note $R_{1/2} < 1$ in the third column in Table 5). The same is true for canola and flax using either of the two clustering metrics and production regions. The absence of distinct spatial patterns in risk measures in wheat was repeated for canola and flax regardless of production region or clustering metric. An example is shown in Figure 1 for wheat in municipality 561 in panels A and B. Similar observations (not shown here) were also found in other crops and municipalities and production regions. Aggregation bias thus exists across crops and to a similar extent judging by the $R_{1/2}$ values and lack of spatial pattern. There are differences in the number of clusters by crop in a given municipality, however, and the range in $R_{1/2}$ values across all clusters seems to be largest for canola, followed by flax and wheat. Rudstrom et al. reported a range of $R_1$ values of 0.27 to 1.53 in wheat across clusters, however. Across all crops and clusters a range of 0.36 to 1.86 is observed in this study. These findings suggest that Rudstrom et al.'s results are robust across crops.

The second question related to the robustness of aggregation bias across production regions. Similar results are seen, in terms of the tendency of aggregate data to underestimate yield risk. This is true for both wheat and canola. Overestimation of risk by aggregate data is more likely in risk area 4 for canola and in risk area 12 for wheat. Overall, there do not appear to be other differences across production regions, however. This further suggests that Rudstrom et al.'s findings may be relatively robust across production regions where similar production practices are used and similar crops are grown.

The third question related to using the relative risk measure rather than an absolute risk measure for clustering. For nearly all clusters, the $R_2$ values are closer to 1 or no aggregation bias compared to the $R_1$ values. The range in $R_2$ values across all clusters, crops and production regions is 0.59 to 1.52. In addition, the number of observations in clusters is more evenly spread across clusters when using the FCV metric compared to the FVE statistic. This suggests that yields are positively correlated with FVE. Overall, differences in spatial patterns as mapped in panel C of Figure 1, do not show patterns, however. Table 6 summarizes differences in cluster numbers across municipality 561 for canola and wheat. On average FCV cluster numbers are higher which is in line with the observation that yields and FVE are positively correlated.

In conclusion, findings in this study lend robustness to the observations reported by Rudstrom et al. (2002). The results suggest that risk is underestimated using aggregated data in most situations. Furthermore, aggregation distortions observed using absolute risk or FVE for the clustering statistic are greater than those observed for using relative risk or FCV. More rigorous statistical testing of comparisons across crops, regions or even across metric used for clustering would be preferred to make stronger conclusions but is left for further study. Decision makers using aggregate data should thus likely continue to entertain sensitivity analysis for their chosen risk measure. A range of distortion adjustment factors ($R_1$ or $R_2$) to arrive at farm level scenarios should be used when aggregate statistics are used as farm level risk may range from being nearly half to as much as nearly twice the aggregate statistics.

## NOTES

1. Calculation of acronyms is provided with an example in Table 1.
2. Growing degree days are a heat measure useful for the growth and development of plants. It is calculated by subtracting the minimum temperature for plant development from the daily mean temperature and summing that daily difference over the period of analysis.
3. *K*-means clustering was done using the FASTCLUS procedure in SAS Version 8 (SAS , Institute Inc., Cary NC) with the distance criterion being least squares and the maximum iterations being 15.

## REFERENCES

**Bechtel, A. I. and D. L. Young. 1999.** The Importance of Using Farm Level Risk Estimates in CRP Enrollment Decisions. Selected Paper at the Western Agricultural Economics Association Annual Meetings.

**Debrah, S. and H. H. Hall. 1989.** Data Aggregation and Farm Risk Analysis. *Agricultural Systems* 31:239-245.

**Fulton, J. R., R. P. King, and P. L. Fackler. 1988**. Combining Farm and County Data to Construct Farm Level Yield Distributions. Staff Paper P88-20. University of Minnesota, Department of Agricultural and Applied Economics.

**Harwood, J., R. Heifner, K. Coble, J. Perry and A. Somwaru. 1999.** Managing Risk in Farming: Concepts, Research, and Analysis. Economic Research Service, United States Department of Agriculture, Agricultural Economic Report No. 774.

**Johnson, R. A. and D. W. Wichern. 1998.** *Applied Multivariate Statistical Analysis.* Prentice-Hall Inc, New Jersey.

**Mapp, H. P. and K. J. Jeter. 1988.** Potential Impact of Participation in Commodity Programs and Multiple Peril Crop Insurance on a Southwestern Oklahoma Farm. Multiple Peril Crop Insurance: A Collection of Empirical Studies. Southern Cooperative Series Bulletin No. 334. Oklahoma State University.

**Manitoba Crop Insurance Corporation. 1980-90.** Unpublished field-level crop yield data, Risk Area 32. Portage la Prairie, Manitoba.

**Milligan, G.W. and M.C. Cooper. 1985.** An Examination of Procedures For Determining The Number of Clusters in a Data Set. *Psychometrika* 50:159-179.

**Popp, M. P., N. L. Dalsted and M. D. Skold. 1997.** Multiple Peril Crop Insurance: An Evaluation of Crop Insurance in Northeastern Colorado. *Journal of the American Society of Farm Managers and Rural Appraisers* 19:95-104.

**Popp, M. and M. Rudstrom. 2000.** Crop Enterprise Diversification and Specialty Crops. *Agricultural Finance Review* 60:85-98.

**Rudstrom, M., M. Popp, P. Manning and E. Gbur. 2002.** Data Aggregation Issues for Crop Yield Risk Analysis. *Canadian Journal of Agricultural Economics* 50:185-200.

**Statistics Canada. 2001**. 2001 Census of Agriculture. Catalogue No. 95F0301XE. Ottawa, Ontario.

**The Manitoba Co-operator. 2000**. *Yield Manitoba,* Supplement to *TheManitoba Co-operator*. Winnipeg, Manitoba.

**Wang, H. H. and H. Zhang. 2002**. Model-Based Clustering for Cross-Sectional Time Series.

**Table 1**. Hypothetical Examples of Data Aggregation Distortions in Variance Estimates.

| | Field Yields in bu/acre[a] | | | | | |
|---|---|---|---|---|---|---|
| Year | 1 | 2 | 3 | 4 | Avg. Yield (Fields 1 to 4) | Year |
| ***Scenario 1*** | | | | | | |
| 1 | 31 | - | 29 | 30 | 30.0 | 1 |
| 2 | 30 | 31 | 30 | 27 | 29.5 | 2 |
| 3 | 29 | 30 | 31 | - | 30.0 | 3 |
| 4 | 32 | - | - | 28 | 30.0 | 4 |
| 5 | - | 32 | 31 | 30 | 31.0 | 5 |
| 6 | 28 | 29 | 27 | - | 28.0 | 6 |
| 7 | - | 28 | 28 | 29 | 28.3 | 7 |
| 8 | 29 | 31 | 30 | 31 | 30.3 | 8 |
| Field Variance Estimates (**FVE**) | 2.17 | 2.17 | 2.29 | 2.17 | **1.00** | Agg. Yield Var. Estimate (**AYV**) |
| Mean Yield | 29.83 | 30.17 | 29.43 | 29.17 | 29.64 | Agg. Mean Yield |
| **FCV** | 4.93% | 4.88% | 5.14% | 5.05% | 3.38% | Agg. CV (**ACV**) |
| Avg. of FVE (**EYV**) | | **2.20** | | | *0.46* | AYV/EYV (**$R_1$**) |
| Avg. of CV (**ECV**) | | **5.00%** | | | *0.68* | ACV/ECV (**$R_2$**) |
| ***Scenario 2*** | | | | | | |
| 1 | 29 | - | 29 | 27 | 28.3 | 1 |
| 2 | 29 | 31 | 31 | 28 | 29.8 | 2 |
| 3 | 30 | 32 | 31 | - | 31.0 | 3 |
| 4 | 27 | - | | 25 | 26.0 | 4 |
| 5 | - | 31 | 31 | 28 | 30.0 | 5 |
| 6 | 30 | 31 | 30 | - | 30.3 | 6 |
| 7 | - | 31 | 30 | 28 | 29.7 | 7 |
| 8 | 28 | 29 | 28 | 27 | 28.0 | 8 |
| Field Variance Estimates (**FVE**) | 1.37 | 0.97 | 1.33 | 1.37 | **2.58** | Agg. Yield Var. Estimate (**AYV**) |
| Mean Yield | 28.83 | 30.83 | 30.00 | 27.17 | 29.14 | Agg. Mean Yield |
| **FCV** | 4.05% | 3.19% | 3.85% | 4.30% | 5.52% | Agg. CV |
| Avg. of FVE (**EYV**) | | **1.26** | | | *2.05* | AYV/EYV (**$R_1$**) |
| Avg. of CV (**ECV**) | | **3.85%** | | | *1.43* | ACV/ECV (**$R_2$**) |

Note:

[a] Field yield time series may not have observations each year due to crop rotations and other considerations.

**Table 2**. Summary statistics.

| | Risk Area 12 | | | | Risk Area 4 | | | |
|---|---|---|---|---|---|---|---|---|
| | Aggregate Statistics | | Average of Field Statistics | | Aggregate Statistics | | Average of Field Statistics | |
| **Municipality ID** | **510** | **561** | **510** | **561** | **621** | **971** | **621** | **971** |
| | | | | **Canola** | | | | |
| # of Obs.[a] | 11 | 11 | 115 | 178 | 11 | 11 | 50 | 85 |
| Mean Yield[b] | 22.8 | 21.3 | 22.8 | 21.3 | 23.0 | 22.1 | 23.0 | 22.1 |
| AYV/EYV ($R_1$)[b] | 59.2 | 39.8 | 95.0 | 82.5 | 56.1 | 39.3 | 81.3 | 87.6 |
| ACV/ECV($R_2$)[b] | 34% | 30% | 45% | 42% | 33% | 28% | 38% | 43% |
| | | | | **Flax** | | | | |
| # of Obs. | 11 | 11 | 228 | 116 | Not analyzed | | | |
| Mean Yield | 18.4 | 17.3 | 18.4 | 17.3 | | | | |
| AYV/EYV ($R_1$)[b] | 45.8 | 43.7 | 71.5 | 61.4 | | | | |
| ACV/ECV($R_2$)[b] | 37% | 38% | 44% | 45% | | | | |
| | | | | **Wheat** | | | | |
| # of Obs. | 11 | 11 | 719 | 583 | 11 | 11 | 279 | 265 |
| Mean Yield | 33.8 | 33.1 | 33.8 | 33.1 | 31.3 | 27.3 | 31.3 | 27.3 |
| AYV/EYV ($R_1$)[b] | 124.7 | 114.0 | 188.0 | 154.5 | 53.0 | 67.9 | 107.2 | 130.3 |
| ACV/ECV($R_2$)[b] | 33% | 32% | 39% | 36% | 23% | 30% | 32% | 41% |

Notes:

[a] Number of observations represents the number of years of aggregate data used in the Aggregate Statistics columns and the number of field or quarter section observations in the municipality in the Average of Field Statistics columns.

[b] Mean Yield, variance and coefficient of variation are calculated as shown in Table 1.

**Table 3**. Canola Cluster Analysis Results for Risk Areas 12 and 4.

| **Municipality ID** | **Item**[a] | **Not Clustered** Municipality | **Clustered**[b] 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|---|
| | | | | | **Risk Area 12** | | | |
| 510 | EYV | 95.0 | 21.2 | 67.2 | 118.2 | 167.6 | 205.9 | 364.7 |
| | ECV | 0.45 | 0.15 | 0.32 | 0.51 | 0.69 | 0.85 | |
| | $R_1$ | *0.62* | *1.71* | *0.79* | *0.61* | *0.53* | *0.56* | *0.77* |
| | $R_2$ | *0.75* | *1.52* | *0.84* | *0.74* | *0.59* | *0.59* | |
| | # of FVE Obs. | 115 | 35 | 28 | 21 | 23 | 6 | 2 |
| | # of FCV Obs. | 115 | 26 | 27 | 31 | 19 | 12 | |
| 561 | EYV | 82.5 | 33.5 | 104.9 | 178.4 | 326.0 | | |
| | ECV | 0.42 | 0.28 | 0.62 | | | | |
| | $R_1$ | *0.48* | *0.74* | *0.48* | *0.45* | *0.77* | | |
| | $R_2$ | *0.70* | *0.86* | *0.60* | | | | |
| | # of FVE Obs. | 178 | 95 | 53 | 26 | 4 | | |
| | # of FCV Obs. | 178 | 104 | 74 | | | | |
| | | | | | **Risk Area 4** | | | |
| 621 | EYV | 81.3 | 33.8 | 105.4 | 203.1 | | | |
| | ECV | 0.38 | 0.15 | 0.29 | 0.43 | 0.62 | | |
| | $R_1$ | *0.69* | *1.42* | *1.10* | *0.74* | | | |
| | $R_2$ | *0.87* | *1.30* | *1.40* | *0.86* | *0.90* | | |
| | # of FVE Obs. | 50 | 25 | 19 | 6 | | | |
| | # of FCV Obs. | 50 | 11 | 12 | 16 | 11 | | |
| 971 | EYV | 87.6 | 19.7 | 73.3 | 123.4 | 169.8 | 248.0 | 357.0 |
| | ECV | 0.43 | 0.25 | 0.70 | | | | |
| | $R_1$ | *0.45* | *1.02* | *0.60* | *0.52* | *0.36* | *0.63* | *0.81* |
| | $R_2$ | *0.67* | *0.88* | *0.62* | | | | |
| | # of FVE Obs. | 85 | 35 | 22 | 8 | 13 | 5 | 2 |
| | # of FCV Obs. | 85 | 52 | 33 | | | | |

Notes:

[a] See description of calculations for EYV, ECV and $R_{1/2}$ in Table 1. AYV and ACV value can be calculated from the information in the table.

[b] Data are reported for the municipality and then for the clusters in order of magnitude of FCV/FVE – i.e. in municipality 510, the first cluster with 26 FCV observations had the lowest FCVs up to the last cluster with 12 observations that had the highest FCVs.

**Table 4.** Flax Cluster Analysis Results for Risk Area 12**.**

| **Municipality ID** | **Item**[a] | **Not Clustered** | **Clustered**[b] | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Municipality | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 510 | EYV | 71.5 | 24.0 | 65.0 | 103.6 | 145.5 | 200.0 | | |
| | ECV | 0.45 | 0.16 | 0.30 | 0.43 | 0.54 | 0.67 | 0.93 | |
| | $R_1$ | *0.64* | *1.21* | *0.64* | *0.56* | *0.60* | *0.70* | | |
| | $R_2$ | *0.82* | *1.34* | *0.89* | *0.83* | *0.73* | *0.70* | *0.72* | |
| | # of FVE Obs. | 228 | 69 | 85 | 47 | 21 | 6 | | |
| | # of FCV Obs. | 228 | 29 | 44 | 52 | 63 | 32 | 8 | |
| 561 | EYV | 61.4 | 24.6 | 61.8 | 109.6 | 178.6 | | | |
| | ECV | 0.44 | 0.13 | 0.27 | 0.40 | 0.53 | 0.65 | 0.84 | 1.11 |
| | $R_1$ | *0.71* | *1.86* | *0.59* | *0.64* | *0.54* | | | |
| | $R_2$ | *0.88* | *0.91* | *0.78* | *0.83* | *0.76* | *0.75* | *0.73* | *1.06* |
| | # of FVE Obs. | 116 | 45 | 47 | 17 | 7 | | | |
| | # of FCV Obs. | 116 | 16 | 22 | 26 | 30 | 15 | 5 | 2 |

Notes:

[a] See description of calculations for EYV, ECV and $R_{1/2}$ in Table 1. AYV and ACV value can be calculated from the information in the table.

[b] Data are reported for the municipality and then for the clusters in order of magnitude of FCV/FVE – i.e. in municipality 510, the first cluster with 29 FCV observations had the lowest FCVs up to the last cluster with 8 observations that had the FCVs.

**Table 5**. Wheat Cluster Analysis Results for Risk Areas 12 and 4.

| Municipality ID | Item[a] | Not Clustered | Clustered[b] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Municipality | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | | | | | | **Risk Area 12** | | | | |
| 510 | EYV | 188.0 | 42.1 | 109.9 | 176.6 | 239.3 | 316.9 | 399.2 | 500.8 | 615.9 |
| | ECV | 0.39 | 0.15 | 0.29 | 0.39 | 0.48 | 0.61 | 0.78 | | |
| | $R_1$ | *0.66* | *1.26* | *0.72* | *0.70* | *0.66* | *0.61* | *0.51* | *0.49* | *0.66* |
| | $R_2$ | *0.85* | *1.06* | *0.90* | *0.81* | *0.78* | *0.70* | *0.73* | | |
| | # of FVE Obs. | 717 | 122 | 174 | 153 | 132 | 73 | 36 | 19 | 8 |
| | # of FCV Obs. | 719 | 108 | 155 | 198 | 145 | 90 | 23 | | |
| 561 | EYV | 154.5 | 62.2 | 168.2 | 280.6 | 488.3 | | | | |
| | ECV | 0.36 | 0.16 | 0.33 | 0.46 | 0.62 | | | | |
| | $R_1$ | *0.74* | *1.27* | *0.67* | *0.65* | *0.56* | | | | |
| | $R_2$ | *0.90* | *1.32* | *0.88* | *0.78* | *0.76* | | | | |
| | # of FVE Obs. | 583 | 241 | 219 | 105 | 18 | | | | |
| | # of FCV Obs. | 583 | 145 | 192 | 185 | 61 | | | | |
| | | | | | | **Risk Area 4** | | | | |
| 621 | EYV | 107.2 | 52.4 | 128.1 | 227.1 | 563.8 | | | | |
| | ECV | 0.32 | 0.16 | 0.25 | 0.34 | 0.42 | 0.54 | 0.67 | | |
| | $R_1$ | *0.49* | *0.59* | *0.55* | *0.59* | *1.03* | | | | |
| | $R_2$ | *0.72* | *0.83* | *0.76* | *0.77* | *0.70* | *0.75* | *0.72* | | |
| | # of FVE Obs. | 279 | 141 | 96 | 40 | 2 | | | | |
| | # of FCV Obs. | 279 | 45 | 78 | 75 | 57 | 17 | 7 | | |
| 971 | EYV | 130.3 | 32.9 | 73.9 | 125.9 | 181.0 | 250.9 | 342.5 | 462.1 | |
| | ECV | 0.41 | 0.23 | 0.37 | 0.50 | 0.65 | 0.96 | | | |
| | $R_1$ | *0.52* | *0.58* | *0.62* | *0.52* | *0.56* | *0.63* | *0.91* | *0.97* | |
| | $R_2$ | *0.74* | *0.82* | *0.76* | *0.75* | *0.72* | *0.81* | | | |
| | # of FVE Obs. | 265 | 44 | 65 | 67 | 58 | 19 | 8 | 4 | |
| | # of FCV Obs. | 265 | 73 | 85 | 70 | 33 | 4 | | | |

Notes:

[a] See description of calculations for EYV, ECV and $R_{1/2}$ in Table 1. AYV and ACV value can be calculated from the information in the table.

[b] Data are reported for the municipality and then for the clusters in order of magnitude of FCV/FVE – i.e. in municipality 510, the first cluster with 108 FCV observations had the lowest FCVs up to the last cluster with 23 observations that had the highest FCVs.

**Table 6**. Comparison of Clustering Using FCV vs. FVE for Wheat, Canola and Flax for Risk Areas 12 and 4.

| | | Risk Area 12 | | Risk Area 4 | |
|---|---|---|---|---|---|
| | **Municipality ID** | 510 | 561 | 621 | 971 |
| *Crop* | *Item*[a] | | | | |
| Canola | Average | 0.200 | -0.051 | 0.680 | [b] |
| | Std. Dev. | 0.829 | 0.415 | 0.621 | |
| | # of Obs. | 115 | 178 | 50 | |
| Flax | Average | 1.013 | [b] | Not analyzed | |
| | Std. Dev. | 0.847 | | | |
| | # of Obs. | 228 | | | |
| Wheat | Average | -0.032 | 0.446 | 1.036 | -0.589 |
| | Std. Dev. | 0.858 | 0.634 | 0.724 | 1.329 |
| | # of Obs. | 719 | 583 | 279 | 265 |

Notes:

[a] The average and standard deviation of the difference between cluster numbers was determined for each municipality by subtracting a quarter section's FVE cluster number from the FCV cluster number. Note that cluster numbers using either clustering statistic were assigned in an increasing order of magnitude (i.e. a low cluster number had lowest risk observations whereas a high cluster number had highest risk observations). In all cases except for municipality 561 for wheat, the two clustering statistics lead to a different number of clusters in a municipality. For these cases, the clusters with fewest observations were combined into a single cluster to make the number of clusters the same across clustering statistic. For example, in municipality 510 for canola, this meant combining FVE clusters 5 and 6 into one cluster so that the number of FVE and FCV clusters was the same and the average of the difference in cluster numbers could be calculated.

[b] For 971 Canola and 561 Flax, the difference in the number of clusters using FVE vs. FCV was deemed too large to combine clusters and arrive at a relatively unbiased difference statistic.

**Figure 1**. Comparison of Spatial Risk Patterns Using FCV vs. FVE Statistics for Clustering Wheat in Municipality 561.

**Cluster Information**

| *FCV clusters* | | | *FVE clusters* | | |
|---|---|---|---|---|---|
| Mean ECV | Range of FCV | # of Obs. | Mean EYV | Range of FVE | # of Obs. |
| 0.16 | 0.059 – 0.245 | 145 | 62.2 | 4.6 – 116.7 | 241 |
| 0.33 | 0.247 – 0.393 | 192 | 168.2 | 117.5 – 226.2 | 219 |
| 0.46 | 0.396 – 0.541 | 185 | 280.6 | 226.9 – 374.1 | 105 |
| 0.62 | 0.547 – 0.865 | 61 | 488.3 | 403.0 – 651.0 | 18 |

**Cluster Difference Information**

| FCV cluster # - FVE cluster # | # of Obs. |
|---|---|
| -1 | 8 |
| 0 | 340 |
| 1 | 206 |
| 2 | 25 |
| 3 | 4 |