Statistics Canada Statistique Canada

Canada

Statistics Canada
www.statcan.gc.ca

Changes in Collecting Canadian Agricultural Data

Cindy St-Germain

Pacioli 23, Belgrade Serbia

September 30th 2015

Outline

- Changes in Collecting Agricultural Data
  - Integrated Business Statistics Process
  - Electronic Questionnaire (EQ)
  - Data collection beyond respondent surveys
    - Combining survey and administrative data
    - Remote sensing

Integrated Business Statistics Program (IBSP)

- IBSP provides a standardized framework for economic surveys conducted at Statistics Canada (SC)
- Uses common tools and approaches
- Harmonized conceptual framework

IBSP

- Major transformational project for business and agriculture surveys aimed at
  - enhancing quality assurance
  - achieving efficiencies
  - improving responsiveness in the delivery of new programs
- Covers all statistical processes from sampling to estimation
- 150 economic surveys to be integrated by 2019

IBSP

- Common sampling framework. All surveys use the Business Register (BR)
  - BR – list of all businesses operating in Canada and all foreign businesses that have links to Canadian companies. (approx. 5 million)
- Common collection tool (Electronic Questionnaire)
- Rolling estimate approach to process data in a more timely method.

Rolling Estimate – What does it mean?

- Currently data collection and processing is quite sequential
  - Complete one step, then start the next
- The Rolling Estimates (RE) method takes a different approach. At certain points during (and after) the collection period, the entire set of processing steps are undertaken
- An iterative approach, the RE runs multiple times during and after collection

Rolling Estimates

- The RE steps
  1. Take the data received to that point
  2. Clean it (resolve inconsistencies), impute for non-response and edit failures, weight and generate estimates
  3. Determine the quality of the estimates and which “key” estimates are of good quality. Focus collection follow-up on units in domains with poorer quality
  4. Allow the analysts to start editing the data
  5. Continue collection and do steps 1-4 again at a later date
- Steps 1-3 are automatically done by the system

The RE Process

- The RE process continues until collection is complete and the analysts' editing is complete
- At that time the final estimates are exported to the confidentiality step
- The number of RE runs will vary by survey
  - For a survey with a short collection period (ex. crops), it may only take place twice during collection (and additional runs during post-collection)
  - For those with longer collection periods (ex. FFS) there will be more runs (perhaps weekly)

Electronic Questionnaire (EQ)

- Primary mode of collection will be EQ for all agriculture surveys
- Presently, computer assisted telephone interview (CATI) is used for most of the 30 agricultural surveys collected by SC.
- EQ is r-EQ (respondent) and i-EQ (interviewer)
  - r-EQ must be Common Look and Feel (CLF) compliant
  - i-EQ: will be used to capture data collected via other modes –all surveys will have an i-EQ (even if r-EQ is not developed)

Electronic Questionnaire (EQ)

- Mail-out: EQ – Security Access Code (SAC) Letter (email as the main tool)
- Low cost non-response follow-up (NRFU)
  - High impact (MI) **eligible** records in a low quality domain (QI) will be prioritized for follow up; non eligible records will be flagged for subject matter analysis review

  Example for annual survey:
  - EQ - Email reminders sent weekly for 4 weeks, then every 3 weeks for a maximum of 3 more times (total 7 reminders max.)

Follow-up Process for EQ

- Only units that were identified as priority for follow-up using the quality indicators are sent back to the field (for telephone interviewing)
- This can lead to more follow-up taking place in one region than another
- The quality indicator and follow-up process can also be used to identify if we should re-contact respondents whose data fails important edits

Electronic Questionnaire (EQ)

- Presently, a survey such as the Farm Financial Survey (FFS) which is 16 pages of financial questions and a sample of 10,000 requires average of 15 calls for a response rate of 71%.
- Cost of interview time is expected to decrease overtime.
- Many FFS respondents already have the questionnaire completed by the time of interview

EQ test

- Frame update survey – results are promising
  - Invitation letter to complete the internet EQ yield a 30% response rate with no follow-up.
  - Test also showed a desire to complete the questionnaire using a hand held device (cell phone etc.)
- 2016 Census of Agriculture will implement EQ collection as a primary mode of collection
  - Expected response rate via EQ is 30%
  - Paper questionnaire option available (mail back)

Data collection beyond respondent survey questionnaires

- We are looking for opportunities to reduce response burdent.
  - Increase use of administrative data
    - Replace survey questions with administrative data
    - Evolution in the use of remote sensing

Combining survey data and administrative data

- 2016 Farm Financial Survey will replace the revenue and expenses questions with taxation data
- 18 questions will be replaced with tax data (approximately 20% of the content)
- Estimated to reduce the collection time by approximately 25%
- Replacing the most sensitive information

How to replace the survey data

- Need a common frame
- Need a common identifier between the administrative data file and survey data file
- Authority to link the two data sets
- Respondent consent to share record level data with survey sponsor

Administrative Data

- Processing taxation data using common tool
  - Agriculture business tax data will be processed using the same tools as other industry tax filers.
  - Need to estimate the agricultural business from other business owned by the same person.
    - Create a special agriculture NAICS code (classification system)
    - Estimate the value of revenue and expenses for the agriculture portion

Combine two data sets to make one

- Link the Agriculture tax data with tax data about off-farm income.
  - Value of the off-farm income
- Link the Agriculture tax data with personal income tax files to look at family farms.
  - Profile the families running the farms in Canada

Move towards more administrative data

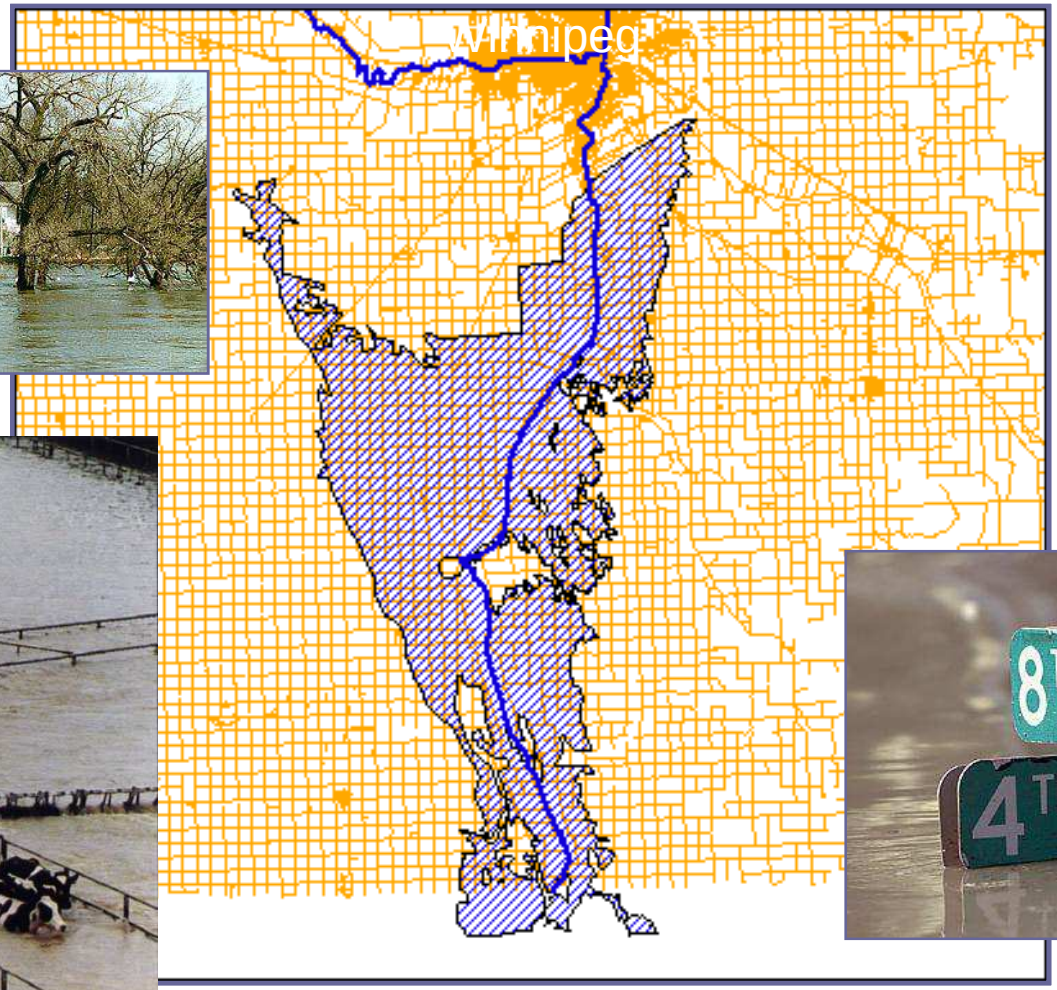
- Presently we use over 100 administrative files
- Working to access more:
  - Administrative data from disaster relief payments
  - SMART meter reading for electricity
  - Crop insurance data
  - Animal traceability data
  - Commodity marking boards
  - SMART Data (GIS data from tractors) – not there yet!

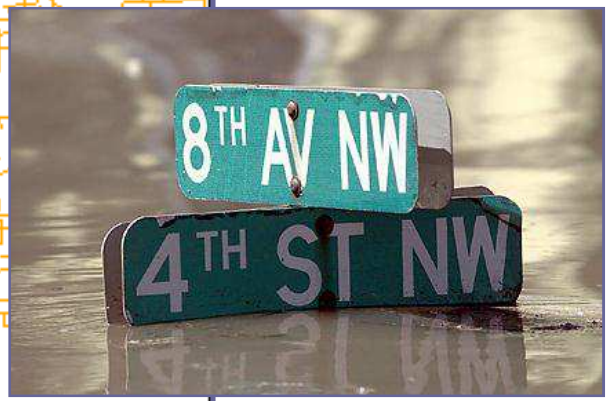
Traditional applications of remote sensing in agriculture

- Geomatics support for the Agriculture Division
- Crop Condition Assessment Program
- Land Cover Classification
- Disaster Monitoring: Red River Flood

Disaster Monitoring: Red River Flood (1997)

- Need to determine number of affected farms for survey purposes





Processing Steps:

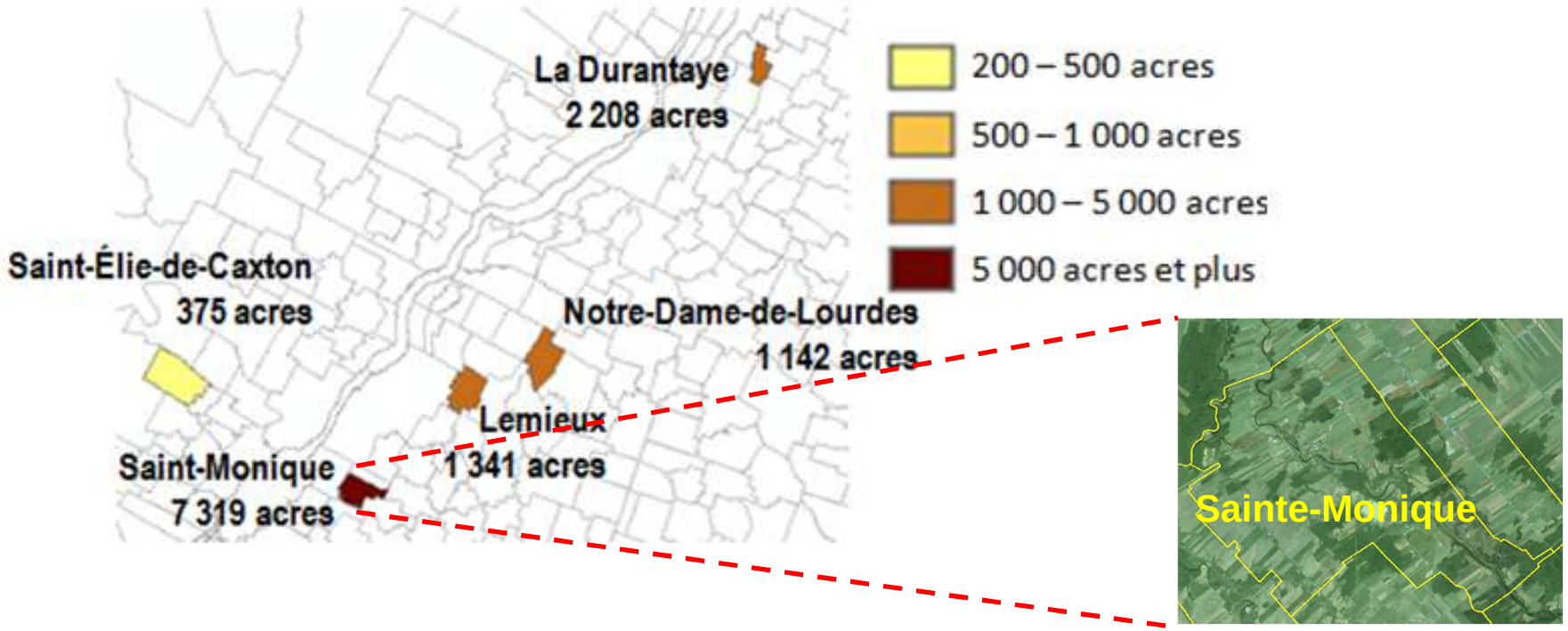
1. Imagery (Radar)
2. Classification
3. GIS integration
4. Determine total number of farms, crop area and livestock affected

Innovative applications of remote sensing in agriculture

- Census of Agriculture validation tool
- Greenhouse Area Estimation
- Crop Yield Modelling
- Crop Area Estimation

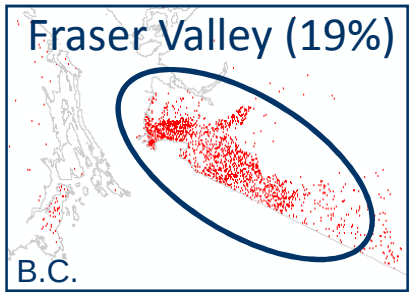
Detection of agricultural land redistribution discrepancies – CEAG 2016

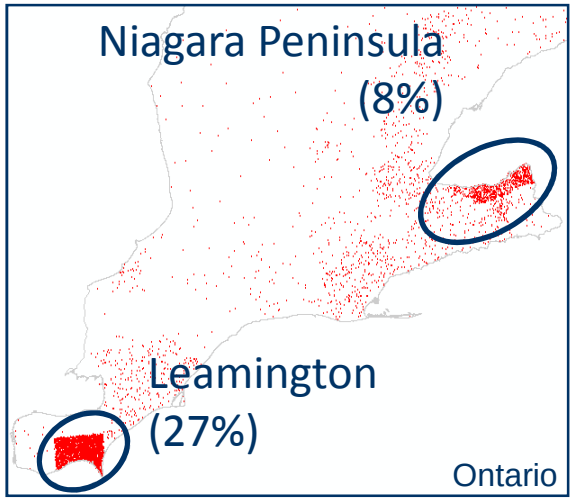
- Use satellite images to detect cropland and compare with the Census of Agriculture (2011)



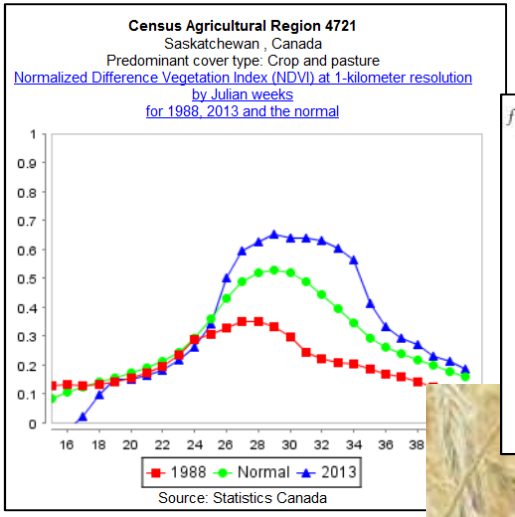
Greenhouse Area Estimation

- Objectives
  - Identify individual operations
  - Identify all greenhouses in a region
  - Validate greenhouse survey results
  - Adapt survey design
- Pilot project focused on 3 regions
  - 54% of Canada's greenhouse area

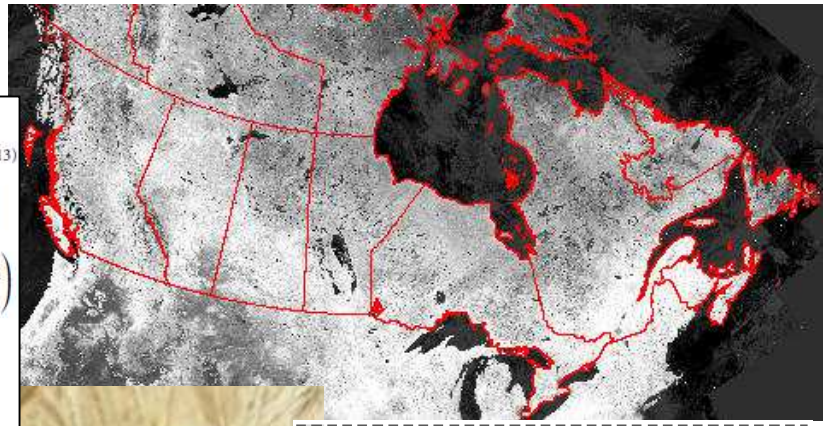


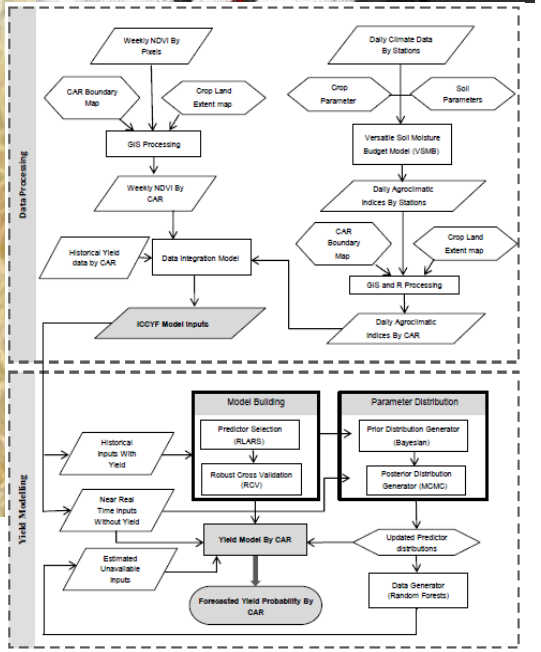


Crop Yield Modelling



$$f\left(y_2,\ldots,y_n|y_1,\Theta,\mathbf{D}_{C,T,(n_p+m^*)}\right)$$
$$=f\left(y_2|y_1\right)f\left(y_3|y_1,y_2\right)\cdots f\left(y_n|y_1,y_2,\ldots,y_{n-1}\right) \quad (13)$$
$$=N\left(\gamma_0+2\gamma_1+\sum_{l=1}^{n_p}\beta_2^{(l)}x_2^{(l)}+\sum_{l=n_{p+1}}^{n}\beta_2^{(l)}z_2^{(l)}+\alpha y_1,\sigma^2\right)$$
$$\times N\left(\gamma_0+3\gamma_1+\sum_{l=1}^{n_p}\beta_3^{(l)}x_3^{(l)}+\sum_{l=n_{p+1}}^{n}\beta_3^{(l)}z_3^{(l)}+\alpha y_2,\sigma^2\right)$$
$$\times\cdots\times N\left(\gamma_0+n\gamma_1+\sum_{l=1}^{n_p}\beta_n^{(l)}x_n^{(l)}\right.$$
$$\left.+\sum_{l=n_{p+1}}^{n}\beta_n^{(l)}z_n^{(l)}+\alpha y_{n-1},\sigma^2\right),$$





Lasso Model versus September survey

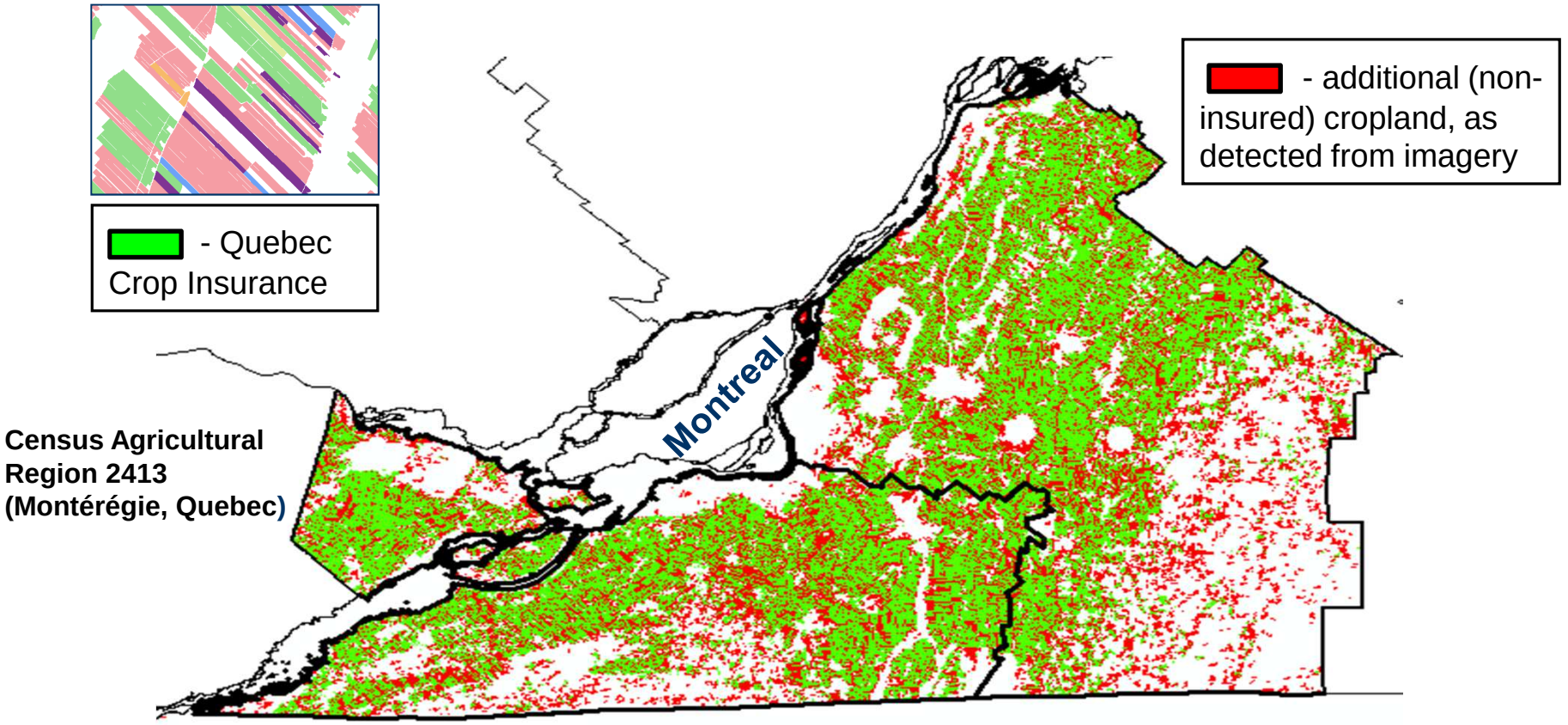
- 2014 yield model estimates – National level

Region	Model Yield (bu/ac)	Model Rel. diff.	Sept. survey Yield (bu/ac)	Sept. survey Rel. diff.	Nov. survey Yield (bu/ac)
Barley	61.8	-1.0%	62.1	-0.5%	62.4
Canola	34.1	-0.9%	32.2	-6.4%	34.4
Corn for grain	156.9	5.2%	147.4	-1.2%	149.2
Durum wheat	42.3	3.2%	37.9	-7.6%	41.0
Oats	80.9	-3.8%	80.1	-4.8%	84.1
Soybeans	43.8	6.3%	41.0	-0.5%	41.2
Spring wheat	44.3	-3.1%	43.5	-5.0%	45.8

(Shaded values show the smallest deviation with the November survey estimates)

Crop Area Estimation

- Crop Insurance Data – Coverage for Québec
  - Using satellite imagery to fill in gaps within the Crop Insurance data



Crop Area Estimation using crop insurance

- Crop insurance data & Statistics Canada survey data

Quebec (2012)			
Crop	**Insurance (acres)**	**Crops Survey (acres)**	**Coverage**
Wheat	115 902	119 100	97.3%
Total Corn	1 062 074	1 098 500	96.7%
Soybeans	625 032	691 900	90.3%
Canola	36 732	42 000	87.5%
Oats	200 998	234 800	85.6%
Barley	152 141	180 500	84.3%
Tame Hay	1 153 590	1 860 700	62.0%

- **For more information, please contact:**

- **Pour plus d'information, veuillez contacter :**

Cindy St-Germain, Chief, Canadian Agricultural Financial Statistics

Agriculture Division

cindy.st-germain@statcan.gc.ca

cindy.st-germain@canada.ca

Information about Remote Sensing

Gordon Reichert

Gordon.reichert@statcan.gc.ca

Gordon.reichert@Canada.gc.ca