In-season crop area estimation using satellite imagery

DScD: Blair Drummond, Saptarshi Dutta Gupta, Najeeb Qazi, Nick Denis

Agriculture Division: Sara Burns

Methodologist: Herbert Nkwimi Tchahou

Delivering insight through data for a better Canada

Outline

- DS and ML at Statistics Canada
- In-season crop area estimation: project scope
- Prior success
- Modelling approach
- Quality Assurance

VALUE PROPOSITION

DS will bring value to policymakers, federal, provincial and territorial client departments and Canadians in two ways:

1. DS enables the creation of high-value, high-quality, trusted data products that are used by decision makers to make evidence-based decisions

2. Leading-edge DS tools and platforms enable more effective service delivery with a new scale of knowledge work automation and cost-efficient operations

Deliver timelier products to Canadians

Reduce response burden on households and businesses

Produce more granular, complete statistics

Enhance tools for privacy and confidentiality

Implement a new scale of knowledge work automation and cost-efficient operations

New data collection tools

Agile, user-centric way of working

Produce new insights

Meet evolving, data-driven needs in an accountable, highly entrepreneurial manner

Types of data science applications

-Natural Language Processing (NLP for short)	-Image Processing	-Predictive Analytics	-Outlier and Error Detection
-Data Engineering and Pipeline	-PDF Extraction and Optical Character Recognition (OCR)	-Epidemiological/ Agent-based Modelling	-Privacy Preserving Technologies
-Web Information Retrieval	-Spatiotemporal Modelling	-Survival Analysis	-Network Analysis
-Responsible/ Explainable ML	-Chat bot systems	-Meta-data management	Research projects

Types of Data

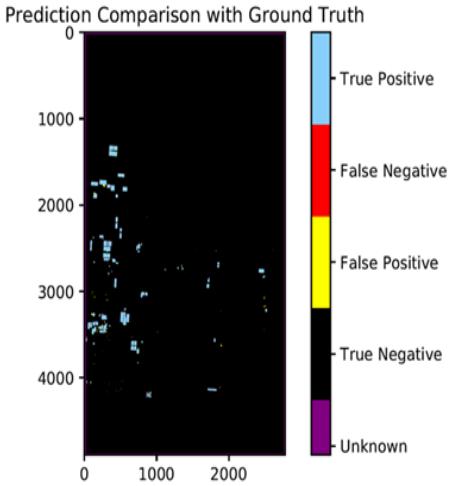
Greenhouse Detection

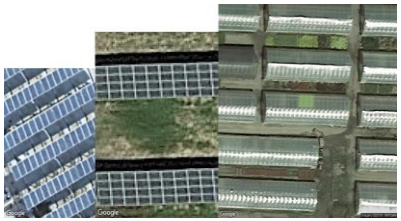
- **Objective:** The goal of this work is to detect, via satellite images or aerial images, the locations of greenhouses and area estimates for the land covered. Two Proof of Concept projects have been completed (one using satellite imagery and one using higher resolution aerial images). **A third project is being implemented with the objectives of** (i) improving model results developed from high resolution imagery, and (ii) determine if additional site features can be identified (e.g. cover type, active or in-active, vegetation type, etc.).

- **Data Used:** The original PoC used RapidEye satellite images at the 5 meter level of resolution. The most recent PoC used higher resolution aerial images provided by municipalities and Land Information Ontario, which was at the 7-15 cm resolution. For both projects, shape files labelled by the Agriculture Division were used as training data. The third project, currently under way, will continue with existing and newly obtained high resolution data.

- **Methods Used:** For the first phase a simple convolution neural network was used. The second phase used a U-Net based architecture with a pretrained residual network for the encoder. The third phase will continue with the U-Net approach but with additional data and further parameter tuning.

- **Results & Value:** The end goal is to leverage an AI solution so as to remove questions from the Census of Agriculture to reduce response burden and help maintain the survey frame for the Annual Greenhouse Survey. The first phase produced promising results for two test areas but poor results for the third. The second phase produced good results across multiple test areas that could be potentially further improved to "production level results" with further hyper-parameter tuning and more training data.





In-season Crop Area Estimation

Crop area estimation from satellite imagery

- STATCAN produces a sequence of surveys every crop season
  - Intention to seed
  - Seed area
  - Estimated yield
  - Yield
- Agriculture and Agri-Food Canada produce yearly land cover estimates
  - >200 models
  - Re-build the model every year, post-season
  - ~6-8 months to produce
- Ag-Zero Initiative:
  - "Zero contact" statistical products
  - Landsat-8 gets >98% field coverage

Wheat: 100k acres
Canola: 223k acres
….
alfalfa: 10 acres

Where we have been

(previous successes)

Where we have been

June 2, 2021

August 13, 2021

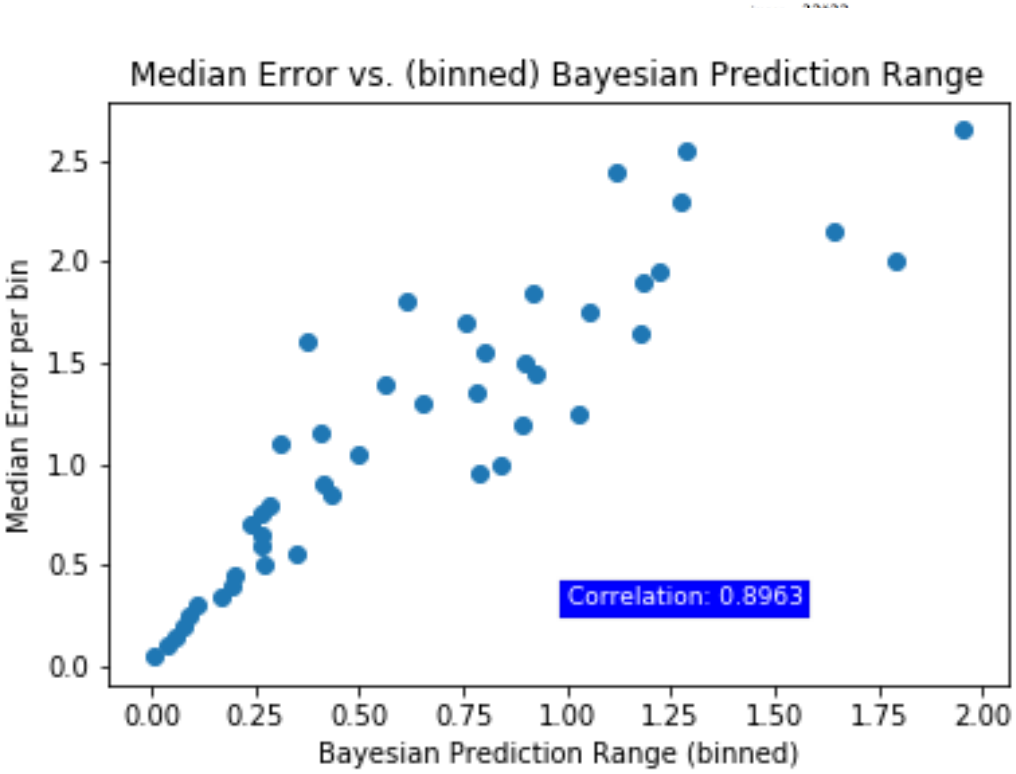
September 28, 2021

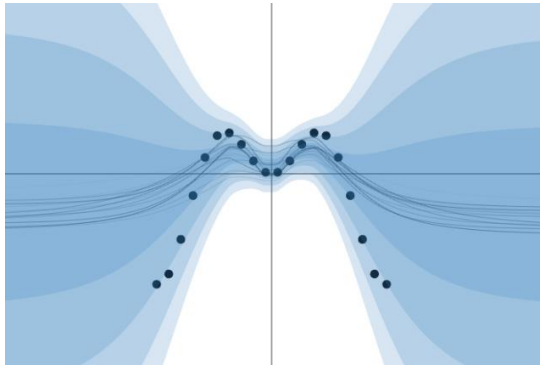
7 Channels: red, green, blue, near infrared, far infrared (x2), NDVI

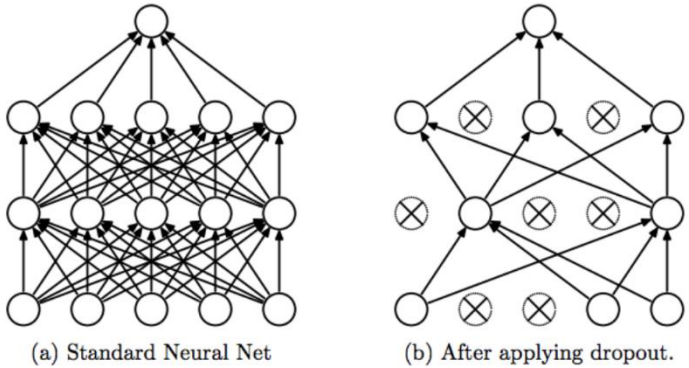
3 images concatenated into tensor of shape (21, 32, 32)

Where we have been

Bayesian ResNet-18 CNN







(a) Standard Neural Net (b) After applying dropout.

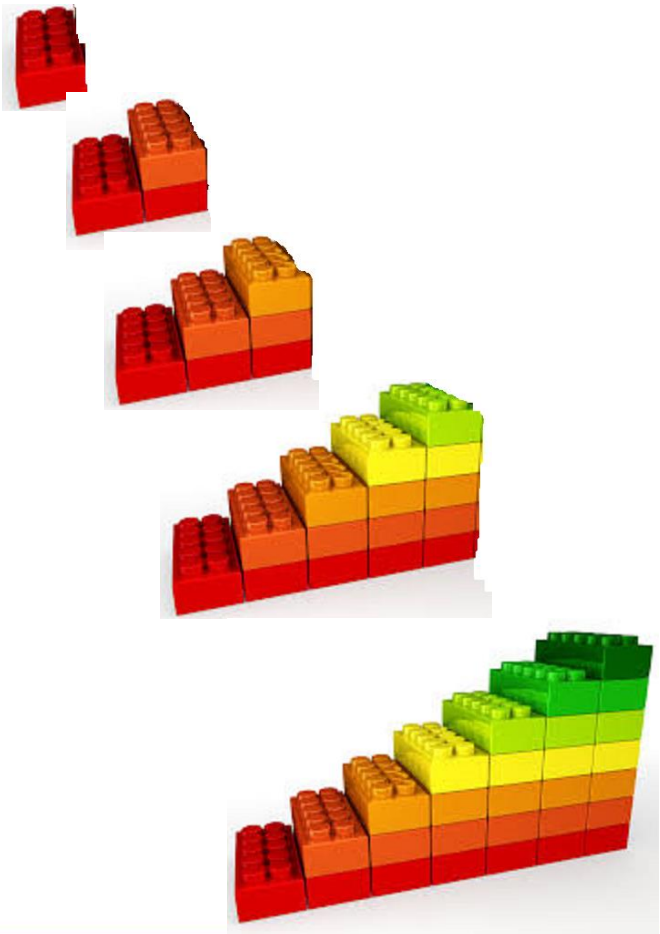
Where we are going

- 13 crop classes → 46 crop classes
- Inputs: 3 images of field over season → Inputs: variable sequence (1-13) of images of field over season
- Alberta, Manitoba → Alberta, Manitoba, Saskatchewan
- 2014-2017 data → 2014-2020 data
- (~ 1.5 TB of data)
- Automated pipeline

Modelling Approach

General deep learning training methodology

- Occam's razor
  - Want the smallest model to get the job done
- Dataset of 1 instance
  - Start with smallest version of model that can achieve near zero error
- Dataset of 10 instances
  - Gradually increase model complexity until achieve near zero error
- Dataset of 100 instances
  - repeat
- …
- Arrive at full dataset and architecture to begin regularization and hyper-parameter tuning



Assessing Model Performance

Quality Assurance

- Project methodologist
- Advised by Advisory Committee on Statistical Methods (ACSM)
- Full error analysis
- Peer review process
- Parallel production run - 2022
- Production run -2023

Merci!
Thank You!

Questions?

Nicholas.denis2@statcan.gc.ca