Probabilistic forecasting of crop yield across Canada under environmental uncertainty

Nathaniel K. Newlands

Canadian STC/AAFC Field Crops Remote Sensing Team

Science and Technology
Agriculture and Agri-Food Canada (Government of Canada)
SUMMERLAND, BRITISH COLUMBIA, CANADA

“Earth Observation for Development and Adaptation to a Changing World”
Beyond crop monitoring, from data to actionable knowledge within a changing world

37th International Symposium on Remote Sensing of Environment (ISRSE 37)
Tshwane, South Africa (May 8 - 12, 2017)
Canada

Population (2016): 36.3 million

Total land area (2015): 9,093,507 km$^2$

Total farmland (2016): 648,250 km$^2$ (~7.1 % of total land area)

10 Provinces and 3 Territories

82 Census of Agriculture Regions

32 field crops

Canada is a major wheat producer

70% of wheat production in Canada is exported (20 million tons annually)
Objective/Motivation

Develop a robust crop yield (probabilistic) model-based and operational forecasting system for principal field crops across Canada (wheat, barley, corn, canola, soybean)

- **Successful examples**
  - European MARS Crop Yield Forecasting System (MCYFS)
  - China CropWatch
  - Regional Yield Forecasting (Australia Ag. Prod. Sys. Res. Unit, APSRU)

- **Issues with traditional survey**
  - observer biases, short lead time, farmer questionnaire burden

- **Aid in assessing risks, with added lead-time**
  - provide early warning, near real time updates

- **Adapt, adjust and improve their crop production plans**
  - crop selection, planting date; promote adaptive decision making

- **Value-added efficiencies**: better use/timing of irrigation, fertilizer, insecticide/herbicide/pesticide application
  - improve sustainability - reduce GHGs, costs, pollution
Integrated Canadian Crop Yield Forecaster (ICCYF)

(1) historical data
Model selection, identification of neighbours for each spatial unit, generation of empirical priors for model parameters.

\[ y = X\beta + \epsilon \]

(2) in-season forecast update
Periodic update of crop yield forecast based on in-season data accumulated at the end of each month.

(3) future scenarios
Long-term yield curves incorporating climate model scenario output.

Growing season of current year:
- May
- June
- July
- August
- Sept

Long-term projection:
- 2020
- 2040
- 2060
Data sources

**Spatial Unit:** The Census Agricultural Region (CAR) is the smallest unit, then aggregated to the provincial and national levels.

**Observed Crop Yield:** Historical yield data from STC’s November Survey at CAR level are used to build the model.

**Climate Data:** Station based climate data (1987- present).

**Crop Health (Greenness) Data:** Satellite derived Normalized Difference Vegetation Index (NDVI) (1987- present).

**Crop Distribution Maps:** Used to filter out NDVI and climate stations out of crop area.

**Crop and Soil Parameters:** used to derive water stress index via a soil moisture budget model (VSMB).
Census Agriculture Regions of Canada, Climate Stations and Crop Land Distributions
Crop Yield Variations at Census Agriculture Regions

Coefficient of Variation

- 0% - 5%
- 5.01% - 10%
- 10.01% - 15%
- 15.01% - 20%
- 20.01% - 25%
- 25.01% - 30%
- 30.01% - 35%

--Coefficient of Variation, i.e. percentage of standard deviation over mean, is calculated from Statistics Canada’s crop yield data of 1987-2011

Spring Wheat

Barley

Canola
Figure 3. Observed distribution and variability of agro-climate and NDVI variables, through the growing season (May-September) based on historical data, 1987-2011: A) Average total monthly precipitation (mm/month), B) Average soil water availability (SWA) (as % AWHC)(no units), C) Average crop water deficit index (WDI) (no units), D) Sum of growing degree days (GDD) (degree-days), and, E) Three-week averaged AVHRR NDVI (no units).
EO-based (annual) crop inventory/maps

- 30 m resolution
- crop identification relies on image acquisitions during key growth stages
- crop classification accuracy increases with multi-temporal imagery (>85%)
- Ground-truth/validation: provincial crop insurance companies and point observations
- bias: some crops insured more than others

Optical Satellite Data
- Landsat-5, -7 and -8; AWiFS; DMC; SPOT-5
- current AAFC operations rely heavily on Landsat-8 (30m) from USGS
- images acquired 3 times during each growing season
- Data gaps can occur in certain regions because of sustained cloud cover
- NDVI/EVI on crop yield, (from NOAA/AVHRR, NASA/MODIS)

Synthetic Aperture Radar (SAR)
- RADARSAT-2 (Wide beam mode W2: 150km swath; ~25m resolution)
- integration of SAR increases optical-based mapping accuracy by 5-15%
- best separation of crops uses ascending mode dual-pol (VV, VH) data

Awifs: Advanced Wide Field Sensor (IRS/India), DMC: Disaster Monitoring Constellation

Credit: A. Davidson, H. McNairn and T. Fisette (AAFC) Operational Space-Based Crop Mapping Protocols at AAFC
Available at: http://www.agr.gc.ca/atlas/aci/
Crop Type Mapping in Canada (2012)
Cartographie des types de cultures au Canada (2012)

Earth Observation Service / Service d'observation de la terre
Science and Technology Branch (STB) / Direction générale des sciences et de la technologie (DGST)
Agriculture and Agri-Food Canada / Agriculture et Agroalimentaire Canada
Multivariate, spatio-temporal model for yield prediction (forecast calibration)

\[ y_t = \alpha_0 + \alpha_1 t + \alpha_2 y_{t-1} + \sum_{i=1}^{l} \beta_i x_{ti} + \sum_{i=1}^{l} \sum_{j=1}^{l} \beta_{ij} x_{ti} x_{tj} + \varepsilon_t \]

- **yield at time** \( t \)
- **de-trend the annual yield**
- **first order autoregressive term**
- **main effects of the top ranked index variables**
- **two way interaction terms between top ranked index variables**
- **error term**
Historical data and auxiliary information:

- Crop yield, phenology
- Agro-climate indices
- Remotely-sensed crop vegetation indices

Variable-selection using robust least-angle regression (LARS)

Model-selection and ranking of leading predictors via robust cross-validation

Identify neighbouring spatial regions, bootstrap re-sampling of residuals (spatial covariance)

Generate prior distributions

Calibrate multivariate yield regression model within each spatial region

Generate joint posterior distribution of model parameters

Markov Chain Monte Carlo (MCMC) sampling, across spatial regions

Integrate other model-based auxiliary indices

Regional-scale, tele-footsprints (e.g., ENSO, PDO, PNA)

Complex agro-ecosystem model (e.g., WOFOST, APSIM)

Down-scaled, regional climate model output (i.e., CCCMa-CanRCM4)

Forecasting with sequential-based updating

Select best predictors for unobserved variables (random-forests learning)

Bootstrap estimates of future values of predictors, superimposed on historical residual spatial covariance structure

Compute crop yield prediction and cross-validation statistics (e.g., CRM, RMSE, MRE, MEI)

Leading algorithm parameters:

- rank
- penalty weights
- CORR trim
- num.neighbours
- bootstrap.n
- boot.n
- init.beta
- beta.cov
- mcmc.n

Other parameters:

- ntree
- mtry
- nodesize
- maxnodes
- nPerm
- alpha
- burn.in
- acf.threshold
- fold
- CV trim
- kmax
- efine.tol
- rel.tol
- rel.tol
- max.it
- maxit.scale
August Forecasted VS. November Surveyed Yields at Provincial and National Scales During LOOCV Tests (1987-2012)

- Modelled Yield
- Surveyed Yield
Figure 5. Sensitivity of forecast error (i.e., Root-mean-squared error variance, RMSE) for different combinations of selected agroclimate variables and NDVI. A) forecasted yield in 2012, and B) cross-validated (LOOCV) (i.e., backcasted) forecast yield. The predictors sets considered in each run were: precipitation P; agroclimate variables of crop water deficit index (WDI), growing degree-days (GDD), soil water availability (SWA), where * denotes P is not included, and ** denotes P is included; NDVI; WDI, GDD, SWA and NDVI. Note: seeding date was assumed fixed, and year, t, was included as an additional input variable in all the sensitivity runs. The ID’s of outlier CARs are indicated.
Crop yield (spring wheat) and NDVI

- semi-arid – best agreement
- validation RMSE: 6-34%
- high degree of model sensitivity

Fig. 7. Relationship between spring wheat grain yield and NDVI in the sub humid, semi arid and arid agro-climatic zones of the Canadian Prairies.

Phenology (NDVI, EVI)

Normalized difference vegetation index (NDVI), Enhanced vegetation index (EVI)

- NDVI saturates at high leaf area
- sensitivity to background reflectance not well understood
- EVI is not always available, can be difficult to extract

\[
NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}
\]

\[
EVI = 2.5 \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + 6\rho_{RED} - 7.5\rho_{BLUE} + 1}
\]

\(\rho_{NIR}: 841-876\) nm, \(\rho_{RED}: 620-670\) nm, \(\rho_{BLUE}: 459-479\) nm

MODerate resolution Imaging Spectrodiometer (Terra-MODIS, 2000-)
Phenology (NDVI and EVI)

Figure 2. Distribution and variability of MODIS NDVI (A) and EVI (B) during the cropping season (May - August) based on historical data, 2000-2009 period. The ends of the boxplots indicate the upper and lower quantiles, the solid line indicates the median. The whiskers are 1.5 times of the box height towards upper and lower from the median. Asterisks are the outliers.
Integrated Canadian Crop Yield Forecaster (ICCYF)

'Outlooks" released in July, August and September

Model-based estimates replace Sept field survey, validated with November survey
Forecast uncertainty (cross-validated)

RMSE (Bushels/Acre)
- 0.00 - 3.00
- 3.01 - 6.00
- 6.01 - 9.00
- 9.01 - 12.00
- 12.01 - 15.00
- > 15.00

Spring Wheat

Bu/ac = 67.25 kg/ha

Barley

Bu/ac = 0.0673 t/ha

Canola

Bu/ac = 57.25 kg/ha

Bu/ac = 0.0673 t/ha
## 2016 forecast vs. survey yield (spring wheat)

<table>
<thead>
<tr>
<th>Source-Method (unit)</th>
<th>AAFC-Model (bu/ac)</th>
<th>STC-Survey (bu/ac)</th>
<th>% Difference (August) (AAFC-STC)/STC*100</th>
<th>AAFC-Model (bu/ac)</th>
<th>STC-Survey (bu/ac)</th>
<th>% Difference (Sept.) (AAFC-STC)/STC*100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Release Date</td>
<td>August 15</td>
<td>August 23</td>
<td></td>
<td>September 15</td>
<td>September 20</td>
<td></td>
</tr>
<tr>
<td>Prince Edward Island</td>
<td>47.8</td>
<td>50.3</td>
<td>-5.0</td>
<td>47.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nova Scotia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New Brunswick</td>
<td>48</td>
<td>46.7</td>
<td>2.8</td>
<td>47.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Québec</td>
<td>48.4</td>
<td>51.9</td>
<td>-6.7</td>
<td>47.2</td>
<td>47.9</td>
<td>-1.5</td>
</tr>
<tr>
<td>Ontario</td>
<td>55.9</td>
<td>57.8</td>
<td>-3.3</td>
<td>58</td>
<td>58</td>
<td>0.0</td>
</tr>
<tr>
<td>Manitoba</td>
<td>48.8</td>
<td>50.3</td>
<td>-3.0</td>
<td>50.8</td>
<td>52.3</td>
<td>-2.9</td>
</tr>
<tr>
<td>Saskatchewan</td>
<td>43.1</td>
<td>43</td>
<td>0.2</td>
<td>42.6</td>
<td>43.3</td>
<td>-1.6</td>
</tr>
<tr>
<td>Alberta</td>
<td>56.9</td>
<td>54.9</td>
<td>3.6</td>
<td>57.2</td>
<td>57</td>
<td>0.4</td>
</tr>
<tr>
<td>British Columbia</td>
<td>56.3</td>
<td>51.3</td>
<td>9.7</td>
<td>56.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada</td>
<td>49.1</td>
<td>48.7</td>
<td>0.8</td>
<td>49.4</td>
<td>49.9</td>
<td>-1.0</td>
</tr>
</tbody>
</table>
Summary

• Integrating satellite and agroclimate data increases forecast accuracy (comparable or better to existing StatsCan field survey data)

• for both grains and oilseed crops, with differing phenology

• Model achieves same/higher accuracy as ensemble forecasts with greater efficiency, less cost, lower complexity

• Hotspots – higher error in areas where water dominates backscatter (SE Manitoba, some years like 2012 had major flooding)

• Higher (ecodistrict) scale great improvement over coarser CAR scale

• EVI better early and late season, NDVI better mid-season, and they produce consistent error maps
In 2016, Statistics Canada first used new methodology replacing survey with a remote sensing yield *model-based approach* is a first for any statistical agency worldwide.

- Finer-scale (ecodistricts)
- Crop response: phenology-stages, cultivar differences
- Better capture yield losses:
  - ENSO inter-annual, extremes (crop heat stress)
  - Disease losses (wheat rust)
- Deep learning vs. random forest algorithm
Acknowledgements

Ian Jarvis, Director,
AgroClimate, Geomatics, and Earth Observations Division (ACGEO)

Gordon Reichert, Director
Agriculture Division, Statistics Canada

Andries Potgieter
Queensland Alliance for Ag. and Food Innovation, QUAFFI, University of Queensland

Aston Chipanshi, Nathaniel Newlands, Yinsuo Zhang, Andrew Davidson, Budong Qian, Bahram Daneshfar, Catherine Champagne
Science and Technology, Agriculture and Agri-Food Canada

Frédéric Bédard, Chris Mohl, Wesley Benjamin, Valéry Jiongo
Agricultural Division, Statistics Canada

New book on sustainability and integrated risk!

**Future Sustainable Ecosystems: Complexity, Risk, and Uncertainty**

Nathaniel Kenneth Newlands

**Hardback**

$89.95

August 16, 2016 *Forthcoming* by Chapman and Hall/CRC
Reference - 408 Pages - 49 B/W Illustrations
ISBN 9781466582569 - CAT# K18941
Series: Chapman & Hall/CRC Applied Environmental Statistics

**Features**

- Provides an interdisciplinary introduction to environmental problem-solving using statistics
- Includes the latest research in the modeling of ecosystems
- Covers many global issues, including sustainable development
- Features many real-world applications, highlighting the role statistics plays in solving environmental problems
- Presents information in a way that is accessible for statisticians and applied scientists

**Summary**

This book provides an interdisciplinary and integrative overview of environmental problem-solving using statistical methods. It shows how statistics can be used to solve problems related to food, water, energy scarcity, and climate change risks. It synthesizes the very latest knowledge to highlight the key outstanding challenges related to agro-ecosystems modeling. It includes many real-world examples and applications and has been written so as to be accessible to statisticians and applied scientists who need to use statistics in their daily work.