1.2

FORECASTING THE YIELD OF THREE GRAIN CROPS ACROSS CANADA WITH THE INTEGRATED CANADIAN CROP YIELD FORECASTER (ICCYF)

Yinsuo Zhang*, Aston Chipanshi, Nathaniel Newlands, Louis Kouadio and Andrew Davidson

Science and Technology Branch, Agriculture and Agri-Food Canada (AAFC), Ottawa, Canada

1. INTRODUCTION

Driven by the increasing societal awareness of the impacts of extreme weather events on crop yields and quality, and the increasing information demand from producers, grain traders, transporters, and government policy makers regarding market access and food security planning, many countries have developed crop monitoring and yield forecasting systems to provide regional, national and global outlooks (Nikolova, et al., 2012; Johnson, 2014; Wu et al. 2014). Traditional survey-based yield reporting method faces increasing challenges including restrictions in resources, demands to increase the lead time, lower responding rate from the farmers, and credibility concerns associated with sampling and non-sampling errors (Statistics Canada, 2014). As more and more regional Earth Observation (EO) datasets become available in Near Real Time (NRT), the EO based crop forecasting methods are thus received increasing attention as an alternative to the traditional survey method.

The Integrated Canadian Crop Yield Forecaster (ICCYF) is a modelling tool for crop yield forecasting and risk analysis based on the integration of geospatial earth observation data using statistics and a Geographic Information System (GIS). By integrating climate, remote sensing and other earth observation information (e.g., historical yields, soil and cropland maps), it generates crop yield outlooks at various spatial scales during and shortly after the growing season.

The objectives of this paper are: (1) to present the data processing and modelling methodology that leads to the operational application of ICCYF tool in forecasting the yields of three major crops across the agricultural landscape of Canada, and (2) evaluate the ICCYF performance using various forecasting quality and skill measures.

2. MATERIALS AND METHODS

The basic spatial modelling units of this study are the Census Agricultural Regions (CARs) that were delineated in the 2011 census of agricultural data collection and dissemination activities (Fig.1). Results are also aggregated to

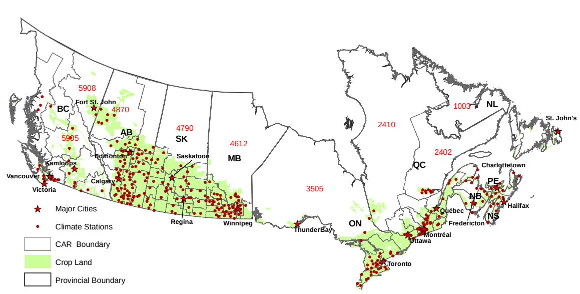

Figure 1: Census Agricultural Regions (CARs), crop land extent and selected climate stations across Canadian agricultural landscapes.

provincial and national levels to evaluate the model performance at larger scales. The station based climate data are provided by Environment

* *Corresponding author address:* Yinsuo Zhang, Agricultural and Agri-Food Canada, 960 Carling Ave., Ottawa, ON, Canada, K1A0C6; Phone: 613-715-5026; e-mail: Yinsuo.Zhang@agr.gc.ca

Canada and other partner institutions. In total, 330 climate stations are selected to represent the climate of the 82 CARs across Canadian agricultural landscape. In the ICCYF, daily series of air temperature and precipitation were fed into a simple process-based Versatile Soil Moisture Budget (VSMB) model (Baier et al. 2000) to generate the agro-climate indices including Growing Degree Days (GDD), precipitation (P), Soil Water Availability (SWA) expressed as the percentage of plant Available Water Holding Capacity (AWHC), Water Deficit Index (WDI) and Crop Seeding Date (CSD). The pixel based NDVI were obtained from the National Oceanographic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR) platform (NOAA, 2013). The daily agroclimatic indices were further aggregated into monthly sums (GDD and P), means (SWA and WDI) and their standard deviations (Std). The weekly NDVI were aggregated into 3-week running means to form the input matrix with the historical yield and the aggregated agroclimate indices.

The ICCYF use several statistical algorithms to forecast the yield. A Robust Least Angle Regression Scheme (RLARS) (Khan et al., 2007) and a Robust Cross Validation (RCV) scheme (Khan et al., 2010) were used for predictor selection and model building. A Bayesian statistical approach (Bornn and Zidek, 2012) and a Markov Chain Monte Carlo (MCMC) scheme (Dowd, 2006) were adopted to derive predictor distributions and a random forests algorithm (Liaw and Wiener, 2002) was applied to estimate the unavailable predictors at the time of forecast. The model derived and observed variables were then used as input into the selected yield model to forecast the yield probability for each CAR (Fig. 2). The geospatial processing is achieved using ArcGIS 10.1 and the coding of the statistical modelling is done using an open source software R. Detailed description of the modelling methodology of the ICCYF can be found in Newlands et al. (2014).

A leave-One-Out-Cross-Validation (LOOCV) scheme was employed to evaluate the model's strength in forecasting the crop yield at different time of growing season for spring wheat, barley and canola, and at different spatial scales. Model performances are evaluated by statistics between the forecasted and surveyed yields such as coefficient of determination (R^2), Model Efficiency Index (MEI) and Mean Absolute Percentage Error (MAPE).

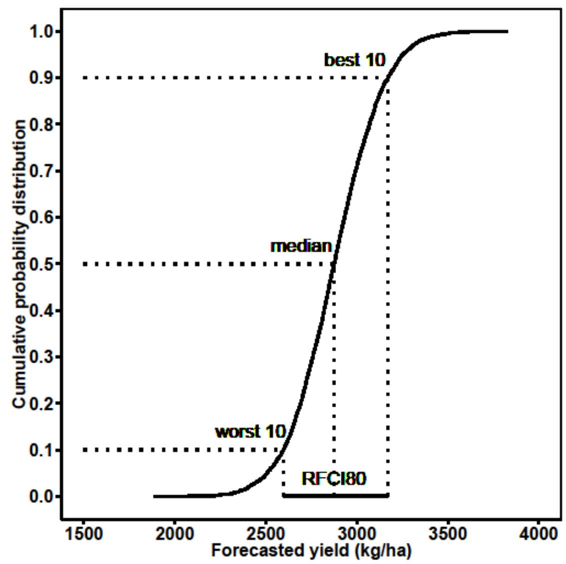

Figure 2: Example of forecasted yield distribution obtained through the Integrated Canadian Crop Forecaster. Dotted lines indicate the positions of forecasted 10th percentile (Worst 10), 50th percentile (Median) and 90th percentile (Best 10) of the cumulative yield distribution curve. The horizontal thick black line between the worst 10 and best 10 lines represents the range of forecasted 80% confidence interval (RFCI80).

3. RESULTS

3.1 Yield Correlation with Model Predictors

The coefficient of determination (R^2) at CAR level during model calibration showed very distinct regional patterns (Fig. 3). The regions with higher R^2 had a good coverage of climate stations and higher percentages of crop coverage in their agricultural land (Fig.1). The R^2 for all CARs ranged from 0.30-0.90, 0.11-0.88 and 0.34-0.86 and their median values were 0.66, 0.51 and 0.67 for spring wheat, barley and canola respectively.

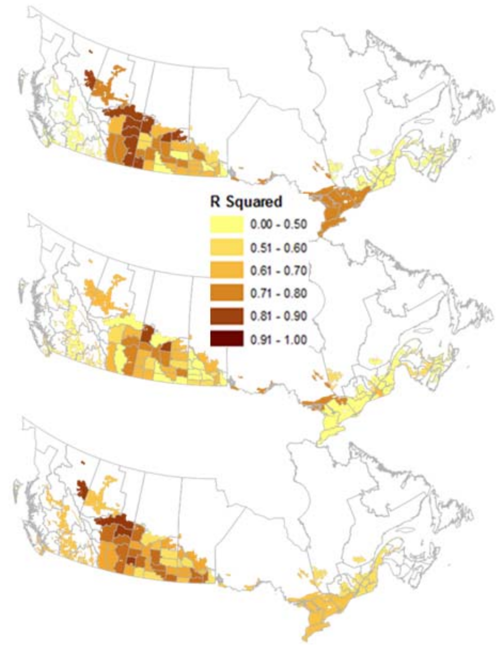

Figure 3: Distribution of Bravais and Pearson Coefficient of determination (R^2) during model calibration at Census Agricultural Region (CAR) level across Canada for the three tested crops.

3.2 Forecasting Skill vs. Forecasting Time

The overall trends of change in forecast error (MAPE) and credibility range (RFCI80) over the four tested forecasts (June, July, August and September) were consistent across the three crop types (Fig. 4). The boxplots shown were obtained from all the forecasts during LOOCV tests from 1987-2012 of all the CARs. At each forecast point, the observed data after the last day of previous month were replaced by the model generated data using random forecasts algorithm to mimic the near real time forecast situation. Both the MAPEs (left panels of Fig. 4) and the RFCI80s (right panels of Fig. 4) decreased significantly from July forecasts to August forecasts for all three crops, but the improvement between any other two consecutive forecasts was small. This indicates that the July predictors are critical to the yield of all the three crops and a skillful forecast is likely achievable at mid-August.

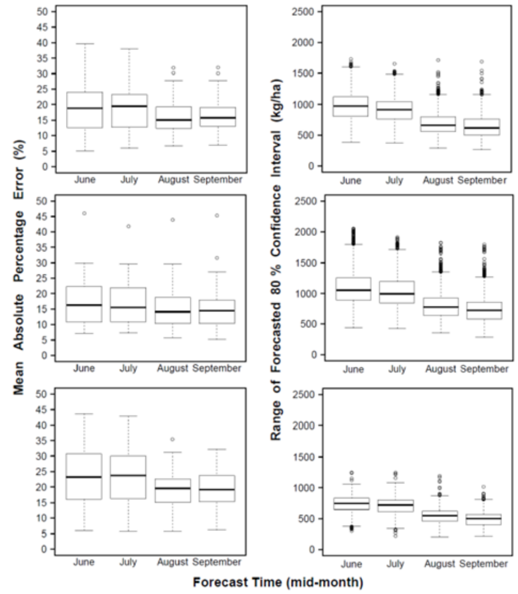

Figure 4: Boxplots of the mean absolute percentage error (MAPE, left) and the range of forecasted 80% confidence interval (right) for mid-June, mid-July, mid-August and mid-September forecasts at the Census Agricultural Region (CAR) scale.

3.3 Spatial Distribution of CAR level forecasting skill (MEI) and quality (MAPE)

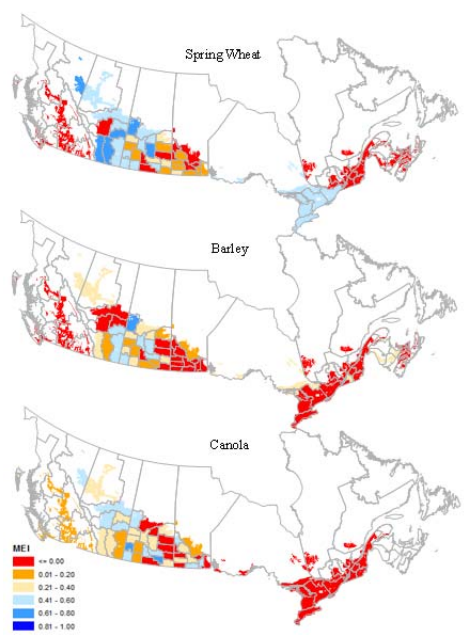

Figure 5: Distribution of MEI for the three tested crops. Areas in red (MEI < 0) indicates where mean yield is a better estimation than the model outputs. MEI towards 1 (blue) indicates a better model.

The spatial distribution of the model efficiency index (MEI) identified the regions where the current ICCYF models need significant improvements (regions in red in Fig. 5). The percentage of CARs that showed positive MEI was 70%, 43% and 70% for spring wheat, barley and canola respectively. The majority of CARs with negative MEI values were located outside the Prairie region, where the harvest area was small.

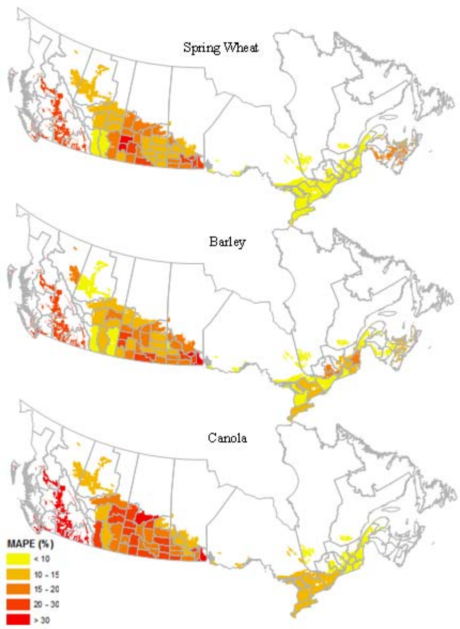

Figure 6: Distribution of MAPE for the three tested crops.

The spatial distribution of forecast quality index MAPE (Fig. 6) was different from the distribution of forecast skill index MEI (Fig. 5). Forecasted relative error was much larger in western Canada than in eastern Canada. This is because the historical yield variation is much larger in western Canada than in eastern Canada. While the MAPE will guide the forecast users in assessing their risk level as part of their decision making process, the MEI is more useful to the forecasters in terms of pointing out where model improvements are needed.

The average MAPE at CAR level for all CARs across Canada were 16%, 15% and 19% for spring wheat, barley and canola, respectively, which were below the mean historical Coefficient of Variation (CV) of 21%, 17% and 25% correspondingly.

3.4 Yield forecasts at provincial and national scales

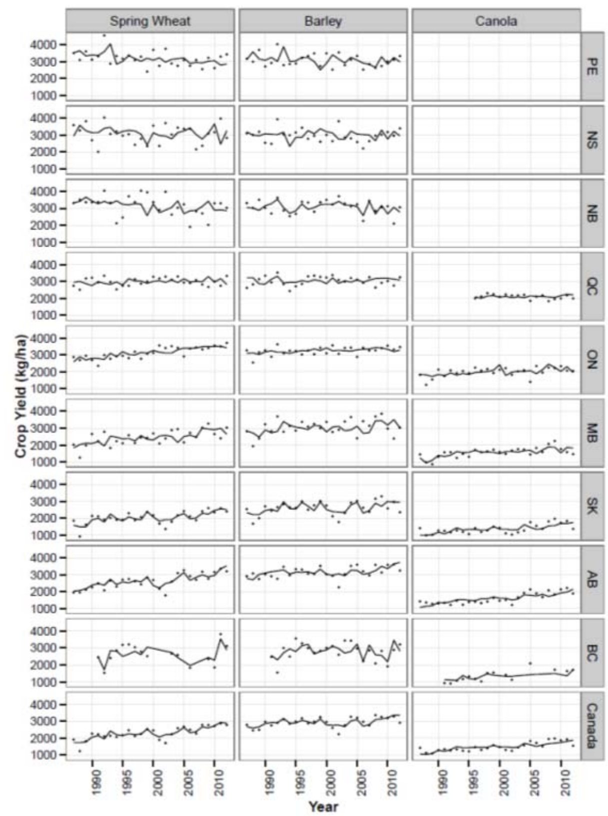

Figure 7: Forecasted provincial and national yields (Lines) vs. surveyed yields (closed circles) of spring wheat, barley and canola using ICCYF during a LOOCV test. Data are not available for canola yield in PE, NB and NS (blank panels).

Overall, the ICCYF forecasted yield followed the surveyed yield trend and variation reasonably well at the national level and at most provinces (Fig. 7), especially in those provinces with a high percentage of harvested area (e.g., AB, SK and MB). However, the current model seemed weak in forecasting extreme yields, e.g. the extremely low spring wheat and barley yields in 1988 at BC, SK, MB and Canada.

4. CONCLUSIONS

The Integrated Canadian Crop Yield Forecaster (ICCYF) was evaluated using multiple model performance measures with a leave-one-out-cross-validation procedure during 1987-2012 for three major field crops and at three regional scales across the Canadian agricultural landscape. The results showed that the ICCYF performance exhibited a strong spatial pattern at both CAR and provincial scales. The performances were better at regions with a good coverage of climate stations and a high percentage of cropped area. The forecast performance improved when aggregating the CAR level forecasts to provincial and national scales. At the national scale, forecasted MAPE values was 7.5%, 5.3% and 8.5% for spring wheat, barley and canola respectively, which were considerably smaller than the corresponding historical coefficients of variation of 16.9%, 9.6% and 17.3% for the three crops. Overall, the ICCYF performed better for spring wheat than for canola and barley at all the three spatial scales. Skillful forecasts were achieved at mid-August, giving a lead time of about one month before harvest and about three to four months before the official final release of survey results. As such, the ICCYF could be used as a complementary tool for the traditional survey method, especially in areas where it is not practical to conduct such surveys.

5. REFERENCES

Baier, W., Boisvert, J.B. and Dyer, J.A., 2000. The Versatile Soil Moisture Budget (VSMB) reference manual [Computer software], ECORC Contribution No. 001553. Agriculture and Agri-Food Canada, Eastern Cereal and Oilseed Research Centre, Ottawa, ON, Canada, pp. A1-D4.

Bornn, L. and Zidek, J.V., 2012. Efficient stabilization of crop yield prediction in the Canadian Prairies. *Agr. Forest Meteorol.,* 152(0): 223-232.

Dowd, M., 2006. A sequential Monte Carlo approach for marine ecological prediction. *Environmetrics*, 17(5): 435-455.

Johnson, D.M., 2014. An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States*. Remote Sens. Environ.,* 141(0): 116-128.

Khan, J., Aelst, S. V., and Zamar, R., 2010. Fast robust estimation of prediction error based on resampling. *Comput. Stat. Data An.,* 54: 3121–3130.

Khan, J.A., Van Aelst, S. and Zamar, R.H., 2007. Robust linear model selection based on least angle regression. *J. Am. Statist. Assoc.,* 102(480): 1289-1299.

Liaw, A. and Wiener, M., 2002. Classification and Regression by random Forest. *R news*, 2(3): 18-22.

Newlands, N.K,. Zamar, D., Kouadio, L. et al., 2014. An integrated model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. *Frontiers Environ. Sci.,* 2:17. doi: 10.3389/fenvs.2014.00017.

Nikolova, S., Bruce, S., Randall, L. et al., 2012. Using remote sensing data and crop modelling to improve crop production forecasting: a scoping study. Australian Bureau of Agricultural and Resource Economics and Sciences (ABARES). ABARES technical report 12.3, Canberra, Australia, pp. 23.

NOAA, 2013. GVI - Normalized Difference Vegetation Index. Office of Satellite and product operations, The National Oceanic and Atmospheric Administration, USA. URL: http://www.ospo.noaa.gov/Products/land/gvi/NDVI.html (Accessed May, 2014)

Wu, B.. Meng, J., Li, Q. et al., 2014. Remote sensing-based global crop monitoring: experiences with China's CropWatch system. *Int. J. Digital Earth*, 7(2): 113-137.

Statistics Canada, 2014. Definitions, data sources and methods of Field Crop Reporting Series, Record number: 3401. Statistics Canada, Agriculture Statistics Division, URL: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=3401 (Accessed May, 2014).