An Integrated Assessment Model for Exploring Potential Impacts of Global Change Scenarios on the Canadian Agricultural System

Nathaniel K. Newlands, David S. Zamar, Brian McConkey, Lawrence Townley-Smith, Agriculture and Agri-Food Canada
nathaniel.newlands@agr.gc.ca

O. Grant Clark
McGill University
grant.clark@mcgill.ca

Abstract
An integrated assessment model is being developed and tested to explore possible future trends of Canadian agriculture in response to projected global scenarios in the ecological, economic, and social dimensions. Although many studies suggest more favorable growing conditions for Canada due to rising temperature and CO₂ concentration, the agricultural sector will likely face challenges of water scarcity and water and land quality degradation, compounded by uncertainty concerning technological and socio-economic evolution at various temporal and spatial scales. To help inform agricultural policy development and investment decisions, integrated assessments from a broad set of scenarios are needed. Our methodology is able to integrate stakeholder/expert knowledge, empirical and process-based model algorithms using remote-sensing and national agri-environmental datasets. We showcase this tool by assessing potential risks and impacts in relation to land suitability and nitrogen loading of water for two primary agricultural regions of Canadian (Western Prairies and Southern Ontario).

1. Introduction
Agricultural systems (i.e., managed ecosystems) provide humans worldwide with sustainable and equitable supply of food, energy, fibre and other ecosystem services. Economic trends in global agricultural markets are defined by changes in supply and demand for the production of agricultural commodities, yet there are a myriad of inter-related environmental, social, technological, and resource-based considerations that underlie such economic trends [3, 17, 19]. Understanding interconnections, risks and beneficial trade-offs is key to ensuring greater efficiency and equitability of agricultural resource supply, alongside long-term sustainability and resilience of our agricultural systems [9]. There is increasing pressure also for multi-criteria based decision-making due to more stringent market-driven sustainability requirements and greater public awareness of environmental risks. Many food retailers and manufacturers are looking beyond their own operations to realize improvements in environmental performance as an estimated 90% of the food industry’s environmental footprint occurs in commodity production, an area outside their direct control. While there is increasing uncertainty and growing concern for sustainable development internationally, there is also increased availability of higher resolution data sets, together with progress in computation and statistics that continues to increase our ability to forecast ecosystem change. In addition to data and model complexity, informatics complexity also exists because of an ever increasing numbers of specialized tools, models and software platforms under development worldwide to measure sustainability [2]. Each of these methodologies is based on a different set of assumptions, sustainability measures (i.e., criteria/sets of indicators), and accuracy of input data sets. Currently, it is unclear which of these methodologies are best suited to accurately assess Canadian production and processing systems and enable compliance with emerging sustainability requirements.

Integrated assessment models (IAM’s) can help to address the above complexities, as they offer a system-level perspective of inter-relationships between different agricultural system components, such as: agricultural production/productivity (crop yield, phenology, commodity price fluctuations), land capability/use/change, soil and air quality, water management, energy use (on-farm use, transport prices, renewable subsidies, biofuel production), biodiversity (invasive species, habitat capacity), and socio-economic indicators (urban/rural population change, labour availability and consumer food prices). Also, IAM’s can use remote-sensing data alongside a broad range of data types, thereby enabling a more transparent and comprehensive way to link geospatial information efficiently and
accurately. Integrated models are coupled to institutional, computing and knowledge/informatics frameworks that support their operational maintenance and enhancement [8, 14, 20]. In this way, modeling and broader integration support contributes to ensuring potential global change-related risks and impacts on agricultural systems (and changes and introduction of new policy instruments) are readily communicated to the public using the best information and knowledge available as part of sustainable environmental management decision-making at regional, provincial/state and national levels [1, 5, 6, 11]. Decision-makers have an ever growing need for a common platform for multi-objective regional collaboration and decision-making: to enable more strategic thinking which addresses key gaps and uncertainties, to identify and rank priorities, and to generate useful interdisciplinary insights. Across jurisdictions and scales potential impact from global changes, there is critical need for reliable and consistent information [6]. Yet, there are still challenges faced in developing reliable IAM’s. For example, many IAM’s lack the ability to forecast, do not incorporate spatially-explicit information and trends, because high-resolution, geospatial information for assessing regional-scale impacts, vulnerabilities, risks and uncertainties is often sparse and difficult to obtain [15, 18]. Past efforts in developing an integrated assessment models include: 1) SEAMLESS-IF (System for Environmental and Agricultural Modeling: Linking European Science and Society) to enable quantitative analysis and scenario-based exploration of the impacts of future policies; funded by the European Union’s 6th Framework Programme for Research Technological Development and Demonstration [16], and, 2) RegIS (Regional Climate Change Impact and Response Studies in East Anglia and North West England) was developed as a stakeholder-driven methodology for regional climate change impact assessment has evaluated scales of impact, adaptation options, and cross-sector interactions (biodiversity, agriculture, water, land suitability) [10].

The development of IAM’s parallels similar advances in developing Earth System Models of Intermediate Complexity (EMIC’s) for exploring ways to bridge gaps between 3D comprehensive and simpler models. These models facilitate evaluation and comparison of historical and auxiliary/proxy observational data. These models need to be enhanced to enable exploration of regional aspects of climate change – especially to be relevant in addressing extreme event impacts (i.e., droughts/floods) on agricultural production. In Canada, the main economic model applied is the Canadian Regional Agriculture Model (CRAM) by Agriculture and Agri-Food Canada (AAFC). This model estimates changes in resource allocations for various crop and livestock activities that occur in response to changes in technology, government programs and policies or market conditions. It covers grains/oilseeds, forage, beef, hogs, dairy and poultry.

The Canadian Economic and Emission Model for Agriculture (CEEMA) links the CRAM model with a GHG emissions indicator to estimate the agriculture sector’s potential contribution to climate change mitigation policies. The Canadian Regional Agriculture Water Use Model (CRAWUM) is aimed at assessing the total agricultural demand for water by sub-sectors and regions. In the early 2000’s integrated economic-environmental modelling capacity was used to identify provincial environmental goals and targets. The study helped identify appropriate environmental goals by indicating the range of achievable outcomes based on various adoption rates for beneficial management practices (BMP’s) [7]. A resurgence of collaborative efforts is now underway to advance integrated modeling at a higher spatial resolution (ecological ‘eco-districts’) under Canada’s ecological framework. The idea is to enable ‘place-based’ risk assessment to improve the ability to forecast agricultural production, environmental impacts and enhance the long-term resilience and sustainability of agro-ecosystems. IAM models need to better utilize new data types – e.g., near-real time climate, remote-sensing and wireless sensor-based information. Lessons learned from past IAM modeling efforts help in re-engineering of compartmental models into a fully-integrated model for regional impact assessment. IAM models are able to track a broader set of agricultural system variables, take spatial variation and uncertainty into account at a finer-scale, and consider a broader set of economic-environmental scenarios via stochastic simulation. As a result, IAM’s are becoming increasingly important as objective ‘guide-posts’ in regional-level decision making by finding critical thresholds and trade-offs for rebalancing system dynamics – especially during regime shifts and periods of high economic volatility and environmental uncertainty.

The long-term objective is to develop a reliable analytical framework and integrated model for assessing the economic and environmental performance of the Canadian agricultural system. The short-term goal is to devise, implement and test a prototype (i.e., reduced) model based on the best available information/data to showcase the importance and future potential for regional IAM-based assessment. This prototype will be embedded
in a software decision-support tool. The operational objective is to help accelerate the synthesis of interdisciplinary knowledge and to utilize it for assessing vulnerabilities, risks, cumulative impacts, and uncertainties in a regional context and spatially-explicit way. To this end, the model is being embedded within a software decision-support tool to provide greater user flexibility, and adaptability to different levels of data availability. This helps to inform stakeholders, including the public, policymakers, natural resource specialists, and agricultural investors. The research objective is to investigate the ability of state-of-the-art statistical modeling techniques to capture spatial trends exhibited by complex sets of interacting variables, to investigate how spatial scale mediates variability, and to determine how to improve the accuracy of integrated forecasts – while keeping IAM models simple, transparent, accurate, and insightful.

2. Material and Methods

2.1. Data sources

Data is being integrated into the Canadian IAM in various phases. The first phase is geared towards the development of a prototype that forecasts at the 'eco-district' scale, and considers two major crops (i.e., spring wheat, and canola) within two regions of Canada (i.e., Western Prairies, South-Eastern Canada/Ontario Great Lakes Region). Ecodistricts are subdivisions of larger eco-regions that are characterized by having distinctive assemblages of relief, landforms, geology, soil, vegetation, water bodies and fauna. The National Ecological Framework for Canada was developed between 1991 and 1999 by the Ecosystems Science Directorate of Environment Canada (EC) and the Center for Land and Biological Resources Research, of AAFC. Over 100 federal and provincial agencies, non-governmental organizations, and private sector companies contributed to its development. The Canadian Ecological Land Classification System divides Canada's natural landscapes into 15 terrestrial ecozones, which in turn sub-divided into 45 ecoprovinces, 177 ecoregions and 5428 ecodistricts. Ecozones are the most generalized level in the classification system. Ecoprovinces, ecoregions and ecodistricts represent subdivisions at progressively more detailed levels of ecological resolution. At this fine spatial scale (delineated polygonal areas of a minimum size of 10,000 ha) historical crop area, production, and yield statistics from Statistics Canada is used. This data was supplemented with additional data from the Canola Council of Canada, Canadian Wheat Board, and provincial sources (Alberta Agriculture and Rural Development (AARD), Saskatchewan Ministry of Agriculture, Manitoba Agriculture, Food and Rural Initiatives (MAFRI)). National historical mean temperature and precipitation anomaly along with regional station-based data was obtained from Environment Canada [21]. Information on historical growing season length and its variation was obtained from Natural Resources Canada. Geospatial data for a selected set of national-level agri-environmental indicators (shown below for data up to the national agricultural census of 2006) was obtained from AAFC [7]. Figure 1 shows four ESRI™ ArcGIS geographical information system (GIS)-generated maps of the four primary agri-environmental sustainability indices input into the agricultural system prototype IAM model. These maps are colour-coded according to five risk classes for which descriptions are also provided. Agricultural future production scenario outlook data up to 2020 for Canada was obtained from AAFC and the Organization for Economic Cooperation and Development (OECD) [13]. Socio-economic indicators, such as urban/rural population change, labour availability and consumer food price affecting the supply/demand of agricultural commodities, farm size and livelihood of farming communities will be included as input variables.

![Figure 1. Agri-environmental indices input into the IAM model associated with five risk classes and their meaning and implication.](http://sis.agr.gc.ca/cansis/nsdb/ecostrat/index.html)
2.2. Statistical methodology

The prototype IAM uses a probabilistic (i.e., Bayesian statistical) approach to enable forecasting, ranking and the use of high-resolution geospatial data. The main strength of this approach is that it incorporates more comprehensively information on the full empirical/observed distribution of input variables (e.g., mean, variance, extremes) and related uncertainty, thereby enhancing robustness of its predictions. In addition, it is very flexible and highly automated, so that new data/additional variables can be integrated and output generated more efficiently than non-statistical approaches. Our approach also employs a number of state-of-the-art statistical algorithms/techniques for finding the best representation of data from a selected set of leading variables, and for conditioning and simulating a model (at the monthly time-scale and across a large number of ecodistrict polygons based on large complex geospatial input datasets. Its main weakness is its need for a sufficiently large amount of data for generating input variable distributions, and larger computational time/effort required to cross-validate. Furthermore, because the approach does not rely on fixed functional relationships and/or deterministic assumptions, understanding and interpreting model output can be more challenging. Whether one builds an IAM under a deterministic or statistical approach, it remains crucial to have an interdisciplinary group of experts assist in the interpretation of IAM model scenario output and associated validation results. Figure 2 provides a summary flowchart of the model and its major components. The general statistical equation for de-trending crop yield data within each spatial unit (ecodistrict polygon) is shown. An autoregressive term (time-lag) of 1 year is required in de-trending relates to the tendency for farmers to seed major crops to the same extent as they did in a previous year to hedge against climate risks and increase future/forward contract profitability. Robust least angle regression (R-LARS) is used to combine different data types, perform variable-selection and identify the best fitting regression model for each ecodistrict [12]. Model-selection involves reference to a subset of historical training data, and combines climate and other spatial information across neighbouring eco-districts to provide enhanced spatial covariance accuracy in using bootstrap to form a Bayesian prior distribution and posterior (i.e., forecasted) distribution for each ecodistrict. The IAM model was implemented in the R statistical language\(^2\) that provides open-source, peer-reviewed, well-validated statistical algorithms. Statistical equations involved in Bayesian forecasting are sufficiently complicated that they are not detailed here. The model uses a computational approach called Markov Chain Monte Carlo (MCMC) to sample from a joint posterior distribution of interest and incorporates statistical bootstrap sampling in building a prior for the model parameters based on data from neighbouring ecodistricts.

![Flowchart of model and its major components](image)

**Figure 2. Methodology for prototype model for regional agriculture impact assessment.**

The general statistical equation for de-trending historical crop yield data considers trend effects due to technology, productivity, interactions and random effects.

Historical distributions of selected indicators are typically non-normal and highly skewed/‘heavy-tailed’, such that normality cannot be assumed. For this reason, our methodology does not use the classical multivariate normal Bayesian model with conjugate prior. As a result, sequential importance sampling (SIR) with weighted re-sampling, which is often used in this context, is not employed.
Moreover, a non-parametric approach called Random Forests [4] is used in conjunction with residual sampling to predict the values of input variables that have not yet been observed, but required as input to the forecasting model. Our procedure uses the best-fitting model obtained from regression and variable-selection based on historical training data, and introduces new indicator variables as covariates to generate forecasted values for crop yield that are linked with scenario/indicator data.

If a fully independent data set is available (distinct from the data used to train or test an IAM model), then model results can be cross-validated against such data. However, a sufficient amount of fully independent data is rarely available, such that statistical sampling techniques for cross-validation need to be employed that partition available data. In this way, cross-validation of our model can be performed by ‘hindcasting’, involving the removal of one year of historical input data at a time (for all input variables of the IAM) (so called ‘leave-one-out’ scheme), and re-running the IAM model with all other years available to see the effect of missing data on the model results. Similarly, cross-validation testing can also be performed by removal of specific input variables, thereby providing an estimate of model structural uncertainty associated with the effect of not considering a specific variable on model output results. Statistical measures, such as root-mean-square error (RMSE) that track variance in model uncertainty, and mean absolute error (MAE) for tracking error in mean can be computed. In addition to the hindcasting scheme, following a recursive sequential procedure; such that when new data is obtained it is independently cross-validated with model scenario ‘forecast’ results before being incorporated as model input and used to train it further. The use of expert opinion provides additional model cross-validation by comparing the integrated model results against local observer knowledge.

An agricultural forecast scenario for 2020 at the national scale and also across two regions of Canada, each comprising a set of six eco-districts (Figure 3): Western Prairies (left) and South-eastern Great Lakes (right) were simulated. Ecodistrict polygons are grey shaded, and shown super-imposed on finer-scale soil landscape polygon units outlined in black. Green boundaries outline larger Canadian Census of Agriculture (CAR) statistical units. This figure shows that there are large portions of total cropland area in the Western region of class 3-5 land i.e., under moderately to severe soil nutrient and water limitations and capability for long-term sustainable agriculture. In the Eastern region, cropland consists of primarily class 1-3 land i.e., with negligible to moderate limitations. The 2020 scenario considers an ‘optimistic’ agriculture future, whereby cropland net greenhouse gas emissions are sufficiently well mitigated by best management practices, the 2020 Copenhagen GHG target of 17% reduction from 2005 levels is achieved, growing season length increases by 10 days per degree temperature warming, climate varies according to projected temperature and precipitation anomalies and minimal increases in water contamination, land degradation/soil erosion.

3. Results and Discussion

National and regional-scale forecast output of the IAM prototype model, based on the 2020 scenario, are summarized in Figures 4 and 5. A broad range of input variables was considered with varying degree of correlation (as profiled in Figure 4a). This inset plots correlation matrix ellipses, where correlation is positive (blue shaded) and negative (red shaded), and each ellipse represents a level curve of the density of a bivariate normal with the matching correlation (narrower the ellipse, the higher the correlation). Soil erosion, crop area show significant positive correlation across a wide range of inputs variables, while crop area (ha), yield (kg/ha) are strongly negatively correlated, as expected. The agricultural indicators of cropland net greenhouse gas emission (AirGHGIndex), cumulative growing season forcing temperatures for crop growth (i.e., summation of growing degree days, GSumGDD), growing-season precipitation (GSumPrecip), soil erosion index (soilErosionIndex) and water contamination index (WaterContamIndex) were all input. Positive correlation between airGHGIndex and crop yield, and waterContamIndex and soilErosionIndex is shown, consistent with expectations. From the input set of variables, a reduced set was selected by the IAM model’s variable-selection routine (inset 4b).

The relative importance or influence of the selected variables is represented visually by a radar plot, associated with the Western (involving spring wheat, shown in black) and Eastern test regions (canola, red). In the East (Great Lakes), precipitation, soil erosion, cropped area and GHG emission all have a strong influence on crop yield, while, in the semi-arid dryland region of the West (Prairies), soil erosion, temperature variation and crop area have the strongest influence on historical yield. Inset 4c shows the distribution and 95% percentiles associated with the historical input data used to train the IAM model (1976-2000, shown in red) is compared to the forecasted distribution generated by the IAM model (2001-2020, shown in blue). FPPI denotes Food Price Index (dimensionless) for a given crop (i.e.,
canola/oilseeds, wheat/grains), $T$ is mean annual ambient air temperature ($^\circ$C), $PcP$ is mean annual precipitation (mm). Open circles are outlier ecodistricts due to a stronger influence of environmental variability and having a higher associated risk in sustaining crop production out to 2020. There was a larger shift forecasted for the Eastern region, compared to the West, with large increases in canola yield forecasted relative to yield gains for spring wheat. Figure 5a provides scenario results of crop yield (kg/ha) (wheat, canola) for each of the test regions and ecodistricts, relative to the historical baseline. For some ecodistricts, specific years (i.e., 2002, 2004, and 2007) show marked yield deviation from the others in the same region, alongside significant shifts of some ecodistrict yield distributions forecasted by 2020 (i.e. ID’s 565, 567, 568 in East and 808 in West). Inset 5b profiles the forecasted yields and reveals that for both of the regions, while the IAM can identify significant differences between yield at the finest ecodistrict scale, overall, the yield range associated with the 2020 scenario is forecasted to be similar to yield variation already observed in the baseline period (i.e., 1990-2000). Trends and relative uncertainty associated with the agri-environmental indicators (spatial covariates), for each ecodistrict, are profiled in inset 5c.

4. Conclusions

Significant regional variability and uncertainty is associated with the forecasting yield (kg/ha) out to 2020. Under the optimistic agriculture future canola is anticipated to show larger yield increases than wheat (increasing from roughly 1400 kg/ha to 2150 kg/ha versus 4000 kg/ha to 4400 kg/ha, respectively). Significant differences in forecast uncertainty (with reference to 95% confidence intervals) are evident at the national, regional and ecodistrict spatial scales. Our work illustrates the potential of IAM’s to provide a significantly enhanced understanding of regional vulnerability, risk and impacts, by incorporating a broad set of agricultural system data and knowledge, simplify data and model complexity by identifying a reduced set of leading predictor variables, and integrate spatial dependence into probabilistic forecasts. At the ecodistrict scale, the crop response within a single or small set of ecodistricts may have a strong influence on forecasted ranges at higher scales. A hierarchical perspective enables more informed decision-making. Given that variation in the input indicators drives variation in the IAM forecasts, there is a clear benefit of sequentially updating input indicator data as well as integrating additional indicators to further reduce forecast uncertainty. Our research showcases the importance of a system level perspective in forecasting spatial impacts and risk, the benefits of linking models to geographic information systems (GIS)/geospatial data for enhancing regional decision making, and the application of ecological informatics to design and test innovative frameworks and models to help improve the sustainable development and resilience of agriculture systems.

5. Decision-support software tool

A decision support software tool that links the IAM model code to a graphical user interface is currently being designed and programmed using the Java NetBeans Version 7.1, Integrated Development Environment (IDE). The statistical code and library algorithms employed are written in the R Statistical Language. First release of this IAM tool is anticipated in 2013. Information on the spatial range and intensity of major Northern Hemisphere atmospheric climate teleconnections, (i.e., ‘telefootprints’) will be integrated into the IAM in an effort to further enhance its reliability. The International Panel on Climate Change (IPCC’s) 5th Assessment Report (AR5) initiates a new process for developing and testing global change scenarios using IAM’s. A broad set of global change scenarios will be generated using baseline climate change projections (RCP’s) to generate a range of forecasts with the Canadian IAM model. This will enable a more comprehensive exploration of future impacts adaptation risks and vulnerabilities of the agricultural sector. Future work will also integrate updated indicator data, and new livestock and renewable energy/biofuel components.

6. References


Figure 3. Top: Ecodistricts (grey shaded) selected as part of testing the IAM model on a 2020 scenario are shown for the Western Prairies (ID’s: 808, 815, 816, 826, 828, 833) and Eastern Great Lakes test regions (ID’s: 557, 565, 567, 568, 570, 572). These regions are located within the two major cropland regions of Canada for the purpose of generating the 2020 agricultural IAM model forecast regional and national-scale scenario output (West: spring wheat, East: canola). Bottom: Major regional differences in land capability for agriculture exist between the two cropland test regions with land under different agricultural production limitations (land classes 1-5).
Figure 4. Prototype IAM model output for national-scale 2020 agriculture (cropland) forecast: a) Correlation between input variables used to train the IAM model, b) importance rankings of reduced set of predictor variables, c) national-level 2020 forecasted change based on global change impacts on Canadian agriculture, based on a subset of representative ecodistricts.
Figure 5. Regional-scale crop outlook for spring wheat and canola out to 2020: a) forecasted crop yield distribution and b) mean trend by ecodistrict in relation to the historical baseline, c) spatial variation in agri-environmental indicators used to guide the IAM system model forecast.