Towards a New Generation of Agricultural System Models, Data, and Knowledge Products

January 31, 2015

Bill & Melinda Gates foundation

Columbia University in the City of New York

UF University of Florida

Oregon State University OSU

AgMIP The Agricultural Model Intercomparison and Improvement Project

Citation Guidance

Each paper of the series '*Towards a New Generation of Agricultural System Models, Data, and Knowledge Products*', is to be cited individually as follows.

Antle, J. M., J. W. Jones and C. Rosenzweig, 2015. *Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Introduction.* AgMIP, <http://goo.gl/MjNjHy>.

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler, 2015. *Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science.* AgMIP, <http://goo.gl/f4eVl4>.

Antle, J. M., B. O. Basso, R. T. Conant, C. Godfray, J. W. Jones, M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig, P. Tittonell, and T. R. Wheeler, 2015. *Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation.* AgMIP, <http://goo.gl/S9gpUH>.

Janssen, S., C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, and J. M. Antle, 2015. *Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support.* AgMIP, <http://goo.gl/kg81jf>.

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products

January 31, 2015

Contents

[1] Oregon State University
[2] University of Florida
[3] Columbia University
[4] NASA Goddard Institute for Space Studies
[5] Michigan State University
[6] Colorado State University
[7] University of Chicago
[8] Argonne National Laboratory
[9] University of Oxford
[10] Commonwealth Scientific and Industrial Research Organisation
[11] University of California, Davis
[12] Alterra Wageningen UR
[13] University of Reading
[14] Wageningen UR
[15] Democritus University of Thrace

Acknowledgments

This series of papers was possible owing to the generous support of the Bill and Melinda Gates Foundation Contract Number 24960 to Columbia University (C. Rosenzweig, Principal Investigator), Oregon State University (J. Antle, Principal Investigator), and the University of Florida, Gainesville (J. Jones, Principal Investigator). The views presented are those of the investigators and authors, not of the Foundation. S. Wood and K. Schneider provided generous and insightful guidance throughout the convening process. S. Price captured ideas and documented results of meetings, workshops and teleconferences. The AgMIP Coordination Team (A. Ruane, C. Mutter, S. Lifson, E. Mencos) supported the process, provided edits to text and graphics, and contributed the final content layout. The Bill and Melinda Gates Foundation kindly hosted the pivotal workshop. B. Basso kindly provided the cover photo of the drone, an emerging tool in precision agriculture. A. Ruane kindly provided the cover photo of the Tanzanian farmers and oxen.

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Introduction

J. M. Antle, J. W. Jones and C. Rosenzweig

Outline

1. Introduction

Agricultural system models have become important tools to provide predictive and assessment capability to a growing array of decision-makers in the private and public sectors. Despite ongoing research and model improvements, many of the agricultural models today are direct descendants of research investments initially made 30-40 years ago, and many of the major advances in data, information and communication technology (ICT) of the past decade have not been fully exploited. This state of science is explained in part by the inevitable lag between invention of new ICT tools and their application, but also by an underinvestment in agricultural research, particularly in non-proprietary public good research, and in research aiming to improve the well-being of poor, smallholder farm households in the developing world. At the same time, the private sector continues to utilize ICT developments – such as the recent advances in site-specific management and in the use of "big data" – to improve productivity in large-scale commercial agriculture. The private sector is also making increasing proprietary use of agricultural systems models, taking advantage of public sector investments made in model development. Even in commercial applications, advances in data are rapidly exceeding analytical capability. Moreover, these proprietary developments are not in turn contributing to the publicly available models, data or ICT tools for agricultural systems analysis. The result is that there is a large and growing gap between the potential uses of agricultural system models, particularly in the developing world, and their actual use.

This gap between actual and potential model developments and uses presents an opportunity to invest in a new generation of agricultural systems models that could dramatically improve the quality of information available to agricultural decision-makers on the farm, as well as for those making private and public investment and policy decisions. A key innovation envisaged for this new generation of models is their linkage to a suite of knowledge products – which could take the form of mobile technology "apps" as well as online analytical tools – that would enable the use of the models by a much more diverse set of stakeholders than is now possible. Because this new generation of models and their applications would represent a major departure from the current models that are largely based on an earlier wave of research investments, we call these new agricultural systems models and knowledge products "second generation" or NextGen.

With support from the Bill and Melinda Gates Foundation, leaders of the Agricultural Model Intercomparison and Improvement Project (AgMIP) organized a scoping study to create a roadmap towards this new generation of agricultural systems models. In this introduction, we describe the scoping study approach and introduce the Use Cases that were the basis for the study design and a stakeholder workshop. Accompanying this introduction are four papers and a workshop summary.

The NextGen scoping study brought together experts to review the current state of agricultural systems model science and to explore possibilities for advancing developments in models, data and information technology. The scope of the study was limited to field, farm and landscape-scale analysis of agricultural production systems, and did not include other aspects of the food system such as distribution. A key element in the study design was to employ a set of Use Cases to motivate and guide each component of the study. A Use Case is a description of a 'persona' and their activities that motivate the generation of knowledge products. Three background papers were developed: a review of current state of agricultural systems models; a vision for NextGen models and their potential uses; and a plan for how ICT can be used to develop new knowledge products based on NextGen models over a 5-10 year time horizon. The goal of the stakeholder workshop was to vet the background papers, obtain additional ideas for Use Cases, and strategize how to bring together the growing community of actual and potential users of new agricultural systems knowledge products with model and knowledge-product developers. The workshop was held in August 2014 at the Bill and Melinda Gates Foundation and brought together a diverse group of scientists and thought leaders from the private and public sectors.[1]

---

1 See also Appendix - Next Generation Farming Systems Models Convening

2. Towards a Computational Agricultural Science: Using Models to Accelerate Innovation

Our vision is for the new generation of agricultural systems models and knowledge products to accelerate progress towards sustainable food production and food security. NextGen models can accomplish this goal by accelerating the rate of agricultural innovation that has increasingly been the source of productivity growth in agriculture. But given the stresses now being placed on the air, land, water and genetic resources on which human life depends, these innovations must also reduce environmental impacts and enhance the resilience of food systems under changing climate conditions. We foresee the use of NextGen models leading to "virtual" and "computational" agricultural research and development that can complement and substitute to some degree for conventional real-time, on-the-ground methods. Likewise, significantly improved data and models can contribute to development of advanced farm-management systems, and by making better information available about new systems, could accelerate the adoption and efficient use of more productive and more sustainable technologies. Such data and models are also essential tools for assessing the landscape scale impacts of technologies, evaluating policies to improve resource management, and projecting the performance of technologies under changing climatic and other environmental conditions.

This vision for NextGen is consistent with research on the past and likely future sources of productivity growth and increases in food commodity production. Research shows that since the agricultural revolution of the mid-20th century, the rate of productivity growth in agriculture has averaged about 2 percent per year, but this average masks large differences between the high and low-income countries, with productivity levels and growth particularly low in Africa. There is also evidence suggesting that cereal yield growth has been declining over the past several decades. Moreover, the evidence indicates that recent productivity growth has come increasingly from more productive use of inputs rather than from increasing the intensity of input use (where intensity means the amount of non-land inputs per unit of land) (Fuglie and Wang 2014). Much of the growth in food supplies will have to come from increases in productivity, not from increasing the amount of land in agriculture in order to protect the dwindling areas of natural ecosystems. Thus, the challenge is to foster innovations that will continue to be the key source of productivity growth as well as the basis for increases in total food production, even as they deliver sustainable agricultural systems under changing climate conditions.

The current method for developing innovations in crops, livestock, and agricultural management is based almost entirely on conventional, time- and labor-intensive experimental methods in which new varieties and management practices are evaluated using field-scale experiments that may last for years. On-farm management decisions still depend largely on individual farmer knowledge acquired through personal experience, supported in some cases by "expert" or more formalized decision support. These processes are slow to improve, even with advances in genetic techniques and information technology. Taken together, these conditions suggest that, with appropriate investments, it may be possible to use simulation experiments carried out with NextGen models to greatly reduce the need for conventional field experimentation and trial-and-error learning. These advances could increase the rate of agricultural innovation, and also increase the rate at which these innovations are successfully adopted and implemented on farms.

To meet the food security challenge of this century, we need not only to continue to innovate, but to do it wisely – meaning, we need systems resilient to shocks and disruptions (including climate change) and that have smaller environmental footprints. Agricultural system models already play an essential role in assessing the broader environmental consequences of agricultural technologies. These consequences include long-term on-farm impacts on soil productivity, as well as off-farm impacts on air and water quality and biodiversity. However, the scientific challenge in making these assessments is great, and it is our judgment that substantial improvements are needed so that agricultural systems models, and the associated ICT platforms and knowledge products, will be able to support the goal of sustainable agricultural innovation. For example, NextGen models have the potential to improve understanding of system resilience to extreme weather events and economic shocks, and more research and development is needed to realize this potential.

3. Meeting the Challenge: The Role of Use Cases

The next generation of agricultural systems models must be driven by the information needs of a wide array of stakeholders. To address this challenge, the author team developed a set of Use Cases to guide the work. The stakeholder workshop validated the authors' Use Cases and developed additional ones[2], since there is a diverse array of potential users of knowledge products supported by agricultural systems models. The five Use Cases proposed by the background paper author teams help guide the critical assessment of the current state of agricultural systems models, and stimulate planning for new developments in models and knowledge products over a 5-10 year period of development.

The Use Cases were created to represent the array of likely users of knowledge products that are linked to NextGen models and data. In each of the scoping study papers, the use cases illustrate the wide range of plausible situations for which models and knowledge products can be used, highlight limitations of existing tools and data, and define the capabilities needed in NextGen models and knowledge products. The five use cases represent two types of farming systems:

Small-holder Farms: small-scale semi-subsistence farms typical of much of Africa and South Asia and other developing regions, many of which produce a mix of subsistence crops, cash crops, livestock, and, in some areas, aquaculture.

Commercial Crop Enterprises: large-scale commercially-oriented crop farms typical of the industrialized countries including the United States.

The Use Cases are designed according to the four criteria indicated in Table 1. Narratives further defining the Use Cases are presented in Section 7 of this paper.

2 See also Appendix - Next Generation Farming Systems Models Convening

4. Background Papers: Implications

The background papers provide a number of insights into the state of agricultural systems science and how the agricultural science community could advance towards a new general of models and knowledge products.

First, it is clear that agricultural system models have become important tools to provide predictive and assessment capability to a growing array of decision-makers in the private and public sectors, including farmers, researchers, agribusiness, and policy-makers. However, the Use Cases demonstrate clearly that in most situations, the decision-makers need to access model outputs through knowledge management tools like mobile applications or personal computer dashboards. Users do not need or want direct interaction with models or model outputs; rather they need to be able to access information through intuitive interfaces that provide decision-relevant information rapidly at low cost. Yet, the review of the agricultural systems models in the background papers shows that few, if any, current agricultural systems models are designed to work with such applications. This insight from the Use Cases shows that there is a major gap between current model and knowledge product capability and what is needed to realize the NextGen vision.

Second, the review of the current status of agricultural system models also shows that while many agricultural system components are represented in simulation tools, these models lack capabilities to address some of the key biophysical factors that limit yields as well as to represent management options and socioeconomic conditions that need to be considered at field and farming systems scales, particularly in developing countries. Furthermore, models have mostly been developed along biophysical and socioeconomic disciplinary lines without considering the need for integrating these component models to address context-specific farming systems and assessment goals.

Third, many of the major advances in data and in ICT of the past decade have not yet been fully exploited by available agricultural decision tools. This state of the science is explained in part by the inevitable lag between invention of new ICT and their applications, but also by an underinvestment in agricultural systems research, particularly in non-proprietary public good research, and in research aiming to utilize ICT to improve the well-being of poor, smallholder farm households in the developing world. This gap between the current models and those that are both needed and possible due to recent advances in ICT presents an opportunity to invest in a new generation of agricultural systems models that could dramatically improve the quality of information available to agricultural decision-makers on the farm, as well as for those making private and public investment and policy decisions.

A key innovation envisaged for this new generation of models would be their linkage to a suite of knowledge products – which could take the form of mobile technology "apps," personal computer-based dashboards, and online analytical and communication tools – that would enable the use of the models by a much more diverse set of stakeholders and for a wider range of purposes than is now possible.

5. Towards a Next Generation of Agricultural Systems Models and Knowledge Products

The stakeholder workshop confirmed the need for a new generation of knowledge products serving the needs of decision makers that exploit recent advances in data, ICT, analytics, and data visualization. The development of these knowledge products should respond to the needs of the rapidly growing user community, as exemplified by the study's Use Cases. The stakeholders also encouraged the science community to build on the momentum from this scoping project and related current initiatives, to develop a focused, near-term strategy to demonstrate the value of potential advances in models and knowledge products.

Based on the background papers and stakeholder workshop, the authors have devised a strategy for NextGen agricultural systems models and knowledge products. Following the background paper on the NextGen vision, as well as the discussions at the stakeholder convening, we envisage a two-pronged strategy that would facilitate the growth of two emerging and closely connected communities of science and practice:

- *Knowledge Product Community:* We see interested public and private organizations leading the development of a community of practice involving knowledge product developers and users, along with establishment of funding mechanisms that involve public institutions, private donors and the private-sector technology and agri-business communities.

Table 1. *Characteristics of Five Use-Cases*

Use cases					
	1	**2**	**3**	**4**	**5**
	Farm Extension in Africa	Developing and evaluation technologies for sustainable intensification.	Investing in agricultural development projects that support sustainable intensification.	Management support for precision agriculture.	Supplying for products that meet corporate sustainability goals.
Farming System	small-holder	small-holder	small-holder	commercial corp	commercial corp
Information User	Farm advisor	Agricultural research team/program	Analyst/adviser	Management consultant	Corporate analyst
Beneficiaries	Farm family	Research institution/ farm population	NGO & clients	Farm business	Agri-business firm
Outcomes	Improved livelihood (income,nutrition, food security)	Improved technology	Sustainable technology	Income, soil conservation & water quality	Profit, risk management, sustainability objectives

- *Science and Data Integration Community:* as an established leader in the agricultural modeling community of science, we see AgMIP and similar partner organizations playing a leadership role as conveners for the science community, to advance the agricultural systems science and data integration capabilities in close collaboration with the knowledge product community. AgMIP and other organizations could work with public and private sector organizations to develop funding mechanisms to support this component.

We believe that the best way to move these concepts into implementation would be to create a NextGen pilot program. A key goal of this pilot program would be to explore how best to develop and bring these two communities of science and practice together. The largest unmet need is for better knowledge products that can advance investments in smallholder farming systems, particularly in Sub-Saharan Africa. Based on the Use Cases, we see the need for pilot activities at both farm and landscape scales. Each of the focus areas could be linked to current projects already in operation to provide grounding in real, on-the-ground development activities and to contribute to their uptake.

6. The Background Papers

The three background papers and their authors are:

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science

Authors: *J. W. Jones, J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, T. R. Wheeler*

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation

Authors: J. *M. Antle, B. O. Basso, R. T. Conant, H. C. J. Godfray, J. W. Jones, M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig, P. Tittonell, T. R. Wheeler*

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support

Authors: *S. Janssen, C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, J. M. Antle*

7. The Use Case Narratives

The following Use Case narratives were developed collectively by the Background Paper authors.

Case 1. Farm Extension in Africa

Jan is working as a farm extension officer in an area in Southern Africa where many farms are very small, incomes are very low, and farmers typically grow maize and beans as staple crops for their family's subsistence and to sell for cash. Some households may have livestock and/or grow vegetables. The aim of the extension service is to help farmers achieve higher and more stable yields of maize and also to advise them on improving their nutrition so that they obtain sufficient protein and micronutrients for healthy families. Jan obtains information on new varieties of maize and beans that are now available to farmers in the area. These new varieties are more drought and heat-tolerant and the bean varieties are more resistant to a common foliar disease. Jan also has information on how to improve nutrient management of these crops using small doses of inorganic fertilizer along with animal manure and crop residues. He also has information on a new technique developed by the CGIAR to partially harvest rainfall to increase water availability to the field and vegetable crops. Because farms vary in size, labor availability, soils, and other characteristics, Jan wants to use the NextGen tools to help tailor advice to each farm family that is practical, likely to be adopted, and provide the best outcome in terms of more stable production, higher income, and better nutrition. Jan obtains information from the farmer to input into his smart phone, which has NextGen On-Farm Information apps, developed for the farming systems of his region, that help him determine combinations of system components that best fit specific-farm situations. He also has extension information sheets written in the local language that describe the components of crop and farming systems that are likely to succeed with the farm family. In turn, data from the individual farms that Jan works with are returned to the NextGen model platform to allow for continuous improvement of the tools.

Case 2. Developing and Evaluating Improved Crop and Livestock Systems for Sustainable Intensification

Debora is a plant breeder/geneticist working on developing a drought- and heat-tolerant hybrid of maize. She would like to be able to evaluate the potential adoption and impact of maize varieties with particular characteristics across the widely varying conditions in Africa. She realizes, however, that maize is only one part of the complex farming systems used by most farmers, which typically involve multiple crops and livestock. She would like to be able to evaluate the potential of new varieties in these complex systems, rather than evaluating maize by itself as had been typically done by most research programs. Moreover, she would like to know whether the new varieties meet the goals being set up for sustainable intensification, such as improving productivity not just in the short term, but taking longer-term impacts on soils, water, and greenhouse gases into account. Working with a team of colleagues at her research institution, she uses the NextGen Technology Adoption and Impact Assessment Tool for this purpose. This tool integrates the genetic characteristics of the maize varieties with soil, weather, economic and social data representing the farm populations where the new varieties could be used. The research team then simulates the potential for adoption and impacts of the new varieties, providing Debora with guidance for the kinds of genetic modifications that would be most valuable to farmers, and also provide an assessment of the long-term sustainability of the systems.

Case 3. Investment in Agricultural Development to Support Sustainable Intensification

Stanley is an investment manager for a prominent Foundation, and he needs to evaluate a project for small farms in Kenya that will increase the intensity of production by increasing fertilizer use per hectare on cash crops while maintaining the current sustainable nutrient balance between pasture grasses, crop residues and animal manure.

Before authorizing a project that combines extension information and fertilizer subsidies, Stanley wants to evaluate whether the higher crop yields would induce a non-sustainable system once the initial period of fertilizer subsidies and extension was completed. Initially he uses the NextGen data and crop and livestock model components to assess the yield and labor impacts of increased yields. An economic assessment model is used to estimate if the current cropping balance will change under the new fertilizer program and if increased fertilizer costs can be more than compensated by increase in cash crop yields in the long run. A long-term farm-level nutrient balance under increased intensification will show whether the new system is sustainable. Stanley would like to evaluate these results under a range of assumptions, and present these to local decision makers so that they share common expectations and uncertainties. For this he uses the NextGen Project Assessor, which opens as a webpage on his computer, and he sets up a new assessment, enters data supplied with the project proposal, and links this to general data layers available in the tool. The Project Assessor then uses the configuration of NextGen model components (both biophysical and socioeconomic) needed to conduct the specific assessment.

Case 4. Management Support for Precision Agriculture in the US for Profitability, Soil Conservation and Water Quality Protection

Greg is a farmer in the US, with a large corn/soybean-based operation and a high level of mechanization fully equipped with auto-tracking system and high-resolution differential GPS. His tractors are equipped with on-the-go sensors for variable applications of seeding, fertilizer, pesticide, and herbicide. Harold is Greg's precision agriculture consultant. Greg receives weekly updates on his smart-phone and tablet from Harold's Precision Agriculture Company about the status of his crops obtained from drone flights and crop model predictions using a combination of observed and forecasted weather. Harold's analysis relies on the NextGen models that are able to deliver strategic and tactical crop management strategy recommendations, process-based variable rate prescriptions for fertilizer/pesticide/herbicide application, and accurate harvest recommendations. The variable-rate prescription map created by Harold's company is cloud-based and is automatically integrated in Greg's tractor's automated system for variable rate application of inputs. This system allows Greg to track all the activities performed in the field and link them to the harvested product.

Case 5. Supplying Food Products that Meet Corporate Sustainability Goals

Jennifer is an economic analyst in a corporate sustainability group. This group has embarked on efforts to make sustainability the core of their mission: marketing food while conserving resources. She is assessing the life-cycle of food products to find ways to conserve energy, save water, minimize waste and reduce greenhouse gas emissions in an effort to make these products more sustainable from farm to fork. Using a web service, Jennifer works with her analysis team to access the NextGen Supply System Assessment Tool. This tool uses real-time weather and historical climate conditions to identify strategies that will optimize the amount of fertilizer to be used across many locations by the corporations enter into contracts with farmers with the goal of increasing yield and reducing greenhouse gas emissions. Using remote-monitoring solutions, as an integral part of the Next-Gen model platform, along with advanced cloud services, Jennifer can help the corporation's contract farmers with decisions regarding when to plant, when to irrigate and when and how much fertilizer to apply.

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science

J. W. Jones, J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler

Outline

Executive Summary

The goal of this paper is to summarize the background and current state of agricultural system models, methods and data that are used for a range of purposes. It summarizes a history of events that contributed to the evolution of agricultural system modeling. It includes process-based bio-physical models of crops and livestock, statistical models based on historical observations, as well as economic optimization and simulation models at household and regional to global scales. This history is followed by an overview of the characteristics of agricultural systems models and the wide range of purposes that various researchers in different disciplines had when developing and using them. These purposes have led to systems being defined, modeled and studied at a wide range of space and time scales. We also summarize the capabilities and limitations associated with these models, data, and approaches relative to what may be needed for next generation models. This is done for different “Use Cases” that cover a range of purposes and scales and that are illustrative of those needed for future applications in developing and developed countries. These Use Cases include models at field, community/landscape, and national scales for use in improving policies and decisions aimed at increasing productivity and improving food and nutrition security at local to national and global scales under changing climate conditions.

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler

1. Introduction

The world has become more complex in recent years due to many factors, including our growing population and its demands for more food, water, and energy, the limited arable land for expanding food production, and increasing pressures on natural resources. All of these factors are further compounded by climate change that will lead to many changes in the world as we have known it (e.g., Wheeler and von Braun, 2013). How can science help address these complexities? On the one hand, there is a continuing explosion in the amount of published information and data contributions from every field of science. On the other hand, the problem of managing it all becomes more difficult and leads to information overload. The information explosion is leading to greater recognition of the interconnectedness of what may have been treated earlier as independent components and processes. We now know that interactions among components can have major influences on responses of systems, hence it is not sufficient to draw conclusions about an overall system by studying components in isolation (Hieronymi 2013). These interactions transcend traditional disciplinary boundaries. Although there continues to be a strong emphasis on disciplinary science that leads to greater understanding of components and individual processes, there is also an increasing emphasis on systems science.

Systems science is the study of real world “systems” that consist of components defined by specialists. These components interact with one another and with their environment to determine overall system behavior (e.g., see Wallach et al. 2014). These interacting components are exposed to an external environment that may influence the behavior of system components but the environment itself may not be affected by the changes that take place within the system boundary. Although systems are abstractions of the real world defined for specific purposes, they are highly useful in science and engineering across all fields, including agriculture. An agricultural system, or agro-ecosystem, is a collection of components that has as its overall purpose the production of crops and raising livestock to produce food, fiber, and energy from the Earth’s natural resources. Such systems may also cause undesired effects on the environment.

Agricultural systems science is an interdisciplinary field that studies the behavior of agricultural systems. Although it is useful to study agricultural systems in nature using data collected that characterize how a particular system behaves under specific circumstances, it is impossible or impractical to do this in many situations. Scientific study of an agro-ecosystem requires a system model of components and their interactions considering agricultural production, natural resources, and human factors. Thus, models are necessary for studying overall agro-ecosystem performance for specific purposes. Data are needed to develop, evaluate, and run models so that when a system is studied, inferences about the real system can be stimulated by conducting model-based “experiments”. When we consider the “state of agricultural systems science”, it is thus important to consider the state of agricultural system models, the data needed to develop and use them, and all of the supporting tools and information used to interpret and communicate results of agricultural systems analyses for guiding decisions and policies.

Agricultural system models play increasingly important roles in the development of sustainable land management across diverse agro-ecological and socioeconomic conditions because field and farm experiments require large amounts of resources and may still not provide sufficient information in space and time to identify appropriate and effective management practices. Models can help identify management options for maximizing sustainability goals to land managers and policymakers across space and time as long as the needed soil, management, climate, and socioeconomic information is available. They can help screen for potential risk areas where more detailed field studies can be carried out. Decision Support Systems (DSS) are computer software programs that make use of models and other information to make site-specific recommendations for pest management (Michalski et al. 1983; Beck et al. 1989), farm financial planning (Boggess et al. 1989; Herrero et al. 1999), management of livestock enterprises (e.g., Herrero et al. 1998; Stuth and Stafford-Smith 1993), and general crop and land management (Plant 1989, Basso et al. 2013). DSS software packages have mainly been used by farm advisors and other specialists who work with farmers and policymakers, although some are used directly by farmers.

In this paper, we provide a critical overview of past agricultural systems science followed by a look toward next generation models and approaches. In this first white paper, we discuss the state of agricultural system science relative to current and future needs for models, methods and data that are required across a range of public and private stakeholders. We start this paper with an overview of major events that happened during the last 50+ years that led to an increased emphasis on agricultural systems science. This timeline identifies key drivers that led to the increasing interest and investments in agricultural system models, demonstrates the complexities of most of the issues, and illustrates a range of purposes. This is followed by an

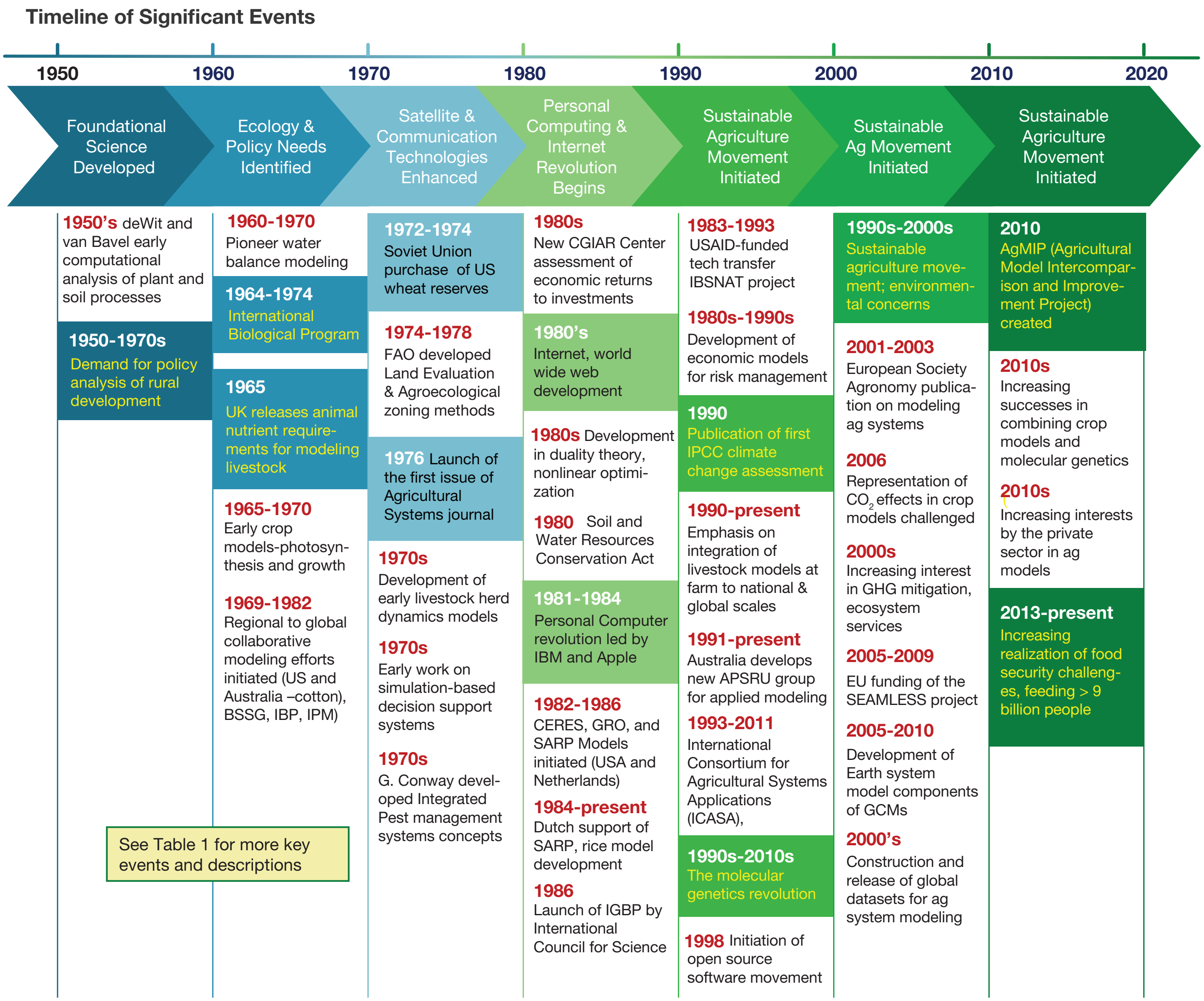

Figure 1. Summary timeline of selected key events and drivers that influenced the development of agricultural system models. Additional details and key events are provided in Table 1 and in the text.

overview of the characteristics of agricultural systems models and the wide range of purposes that various researchers in different disciplines have had when developing and using them. These purposes have led to systems being defined, modeled and studied at a wide range of space and time scales. We also summarize capabilities and limitations of these models and the data and approaches needed for applying them to address important user goals. This is done using "Use Cases" introduced in the paper by Antle et al. (2015 this volume) that cover a range of purposes and scales and are illustrative of those needed for future applications in developing and developed countries. These Use Cases include the ways actual practitioners employ agricultural system models at field, community/landscape, and national scales for improving policies and decisions that are aimed at increasing productivity (e.g., via sustainable intensification approaches) and improving food and nutrition security at local to national and global scales.

2. Brief History of Agricultural System Modeling

The history of agricultural system modeling is characterized by a number of key events and drivers that led scientists from different disciplines to develop and use models for different purposes; Figure 1 summarizes these major events. Some of the earliest agricultural systems modeling (Table 1) was done by E. Heady and his students to optimize decisions at a farm scale and evaluate the effects of policies on the economic benefits of rural development (Heady 1957; Heady and Dillon 1964). This early work by economists during the 1950s through the 1970s helped to inspire additional economic modeling since that time. The work of Dent and Blackie (1979) included models of farming systems with economic and biological components; their book provided an important source for various disciplines to learn about agricultural systems modeling. Soon after agricultural economists started modeling farm systems, the International Biological Program (IBP) was created. This led to the development of various ecological models, including models of grasslands during the late 1960s and early 1970s, which were also used for studying grazing of livestock. The IBP was inspired by forward-looking ecological scientists to create research tools that would allow them to study the complex behavior of ecosystems as affected by various environmental drivers (Worthington 1975; Van Dyne and Anway 1976).

Models of agricultural production systems were first conceived of in the 1960s. One of the pioneers of agricultural system modeling was a physicist, C. T. de Wit of Wageningen University, who, in the mid-1960s, believed that agricultural systems could be modeled by combining physical and biological principles. Another pioneer was a chemical engineer, W. G. Duncan, who had made a fortune in the fertilizer industry and returned to graduate school at age 58 to obtain his PhD degree in Agronomy, after which he began creating some of the first crop-specific simulation models (for corn, cotton, and peanut). Work by de Wit and Duncan intrigued a number of other scientists and engineers who started developing and using crop models. In 1969, a regional research project was developed in the USA to develop and use production system models for improving cotton production, building on the ideas of de Wit, Duncan, and Herb Stapleton, an agricultural engineer. Thus, some of the first crop models were curiosity-driven with scientists and engineers from different disciplines thinking out of the box, so to speak, and inspiring others to get involved in a new, risky research approach. During this early time period, most agricultural scientists were highly skeptical of the value of quantitative, systems approaches.

In 1972, the development of crop models received a major boost after the US government was surprised by very large purchases of wheat by the Soviet Union, causing major price increases and global wheat shortages (Pinter et al. 2003). New research programs were funded to create crop models that would allow the USA to use them with newly-available remote sensing information to predict the production of major crops that were grown anywhere in the world and traded internationally. This led to the development of the CERES-Wheat and CERES-Maize crop models by Joe Ritchie and his colleagues in Texas (Ritchie and Otter 1984; Jones and Kiniry 1986). These two models have continually evolved and are now contained in the DSSAT suite of crop models (Jones et al. 2003; Hoogenboom et al. 2012).

During much of the time since the 1960s, very small fractions of agricultural research funding were used to support agricultural system models, although the Dutch modeling group of C. T. de Wit was a notable exception (Bouman et al., 1996). Thus, most of those who were modeling cropping systems, for example, struggled to obtain financial support for the experimental and modeling research needed to develop new models or to evaluate and improve existing ones. Instead, there were other "crisis" events or realizations of key needs fueling model development (Table 1), each typically leading to infusion of additional financial support over short durations of time for model development or use.

The concept of Integrated Pest Management emerged in the 1970s, in particular from the work of Gordon Conway on the pests and diseases of plantation crops in Malaysia (see Conway 1987). In 1972, the so-called Huffaker Integrated Pest Management (IPM) project was funded in the USA to address the major problems associated with increasing pesticide use and development of resistance to pesticides by many of the target insects and diseases (Pimentel and Peshin 2014). Mathematical models of insect pests and crop and livestock diseases had been developed in the first half of the 20th century, though the success of synthetic agrichemicals led to a shift in attention from other control measures in the years after the Second World War. The Huffaker project infused funds for developing insect and disease models of several crops, combined with experimental efforts aimed at reducing pesticide use and more effective use of all measures to prevent economic damage to major crops in the USA. This project continued until 1985 (as the Consortium for IPM after 1978). Coincident with this project was a major increase in the sophistication of population dynamic models in ecology and a growing appreciation of the importance of non-linearities and the problems for forecasting they imply (May, 1982). Lively debate about the appropriate way to model ecological interactions in agricultural settings characterized these decades.

Globally, the FAO and various countries were also promoting IPM, with modeling as one of the approaches used to understand how to manage pests and diseases with minimal pesticide use. During this time period, a number of insect and disease dynamic models were developed, and some were coupled with cotton and soybean crop models (Wilkerson et al., 1983; Batchelor et al. 1993), including the SOYGRO model that is now in DSSAT (Jones et al. 2003). This period of time also led to the development of a general framework for coupling crop models with insect and disease information to estimate impacts on growth and yield (Boote et al. 1983).

Due to increased emphasis on interdisciplinary research of agro-ecosystems and the need to publish scientific advances in this area, a new journal was launched in 1976 – Agricultural Systems, edited by C. R. W. Spedding (1976). This journal helped legitimize agricultural system modeling, providing a place for scientists to publish their agricultural systems modeling and analyses, creating a collection of scholarly work in this area. Through its publication examples, it has provided encouragement for agricultural systems research to authors across all agricultural science disciplines.

The work started by the early pioneers has continued to evolve throughout the years. Notably, the Wageningen University has carried on the legacy of C. T. de Wit by training many agricultural system modelers and by developing a number of crop models that are still in use today (Bouman et al. 1996; van Ittersum et al. 2003). Similarly, some of the early work of Duncan, Ritchie, and others has evolved and contributed to the widely-used DSSAT suite of crop models through collaborative efforts among the University of Hawaii, University of Florida, Michigan State University, the International Fertilizer Development Institute, Washington State University, and others (Tsuji et al. 1998; Jones et al. 2003; Hoogenboom et al. 2012).

There were other notable government-funded initiatives in the U.S., Netherlands, and Australia that led to major developments of crop, livestock, and economic models. This includes the 1980 US Soil and Water Conservation Act that led to development of the EPIC model that is still in use today (Williams et al. 1983, 1989), the USAID-funded IBSNAT project that led to the creation of the DSSAT suite of crop models that incorporated the CERES and CROPGRO models (Jones 1993; Boote et al. 1998; Jones et al. 2003;

Hoogenboom et al. 2012), and the Systems Analysis of Rice Production (SARP) project funded by the Dutch government starting in 1984 that led to the development of the ORYZA rice crop model, now widely used globally (Bouman et al. 2001). Another major event was the development of the SEAMLESS project, funded in 2005 and operating for 5 years. This effort led to major collaboration among agricultural systems modelers and scientists across Europe for development of new data interfaces and models, and to development and integration of models at field, farm, and broader spatial scales, including cropping system and socioeconomic models (van Ittersum et al. 2008).

The evolution of economic models for different scales and purposes progressed steadily during the last five decades (Table 1). These developments were fueled by various needs at national and international levels as well as innovations in modeling approaches by the agricultural economics community. The needs included mandates of CGIAR Centers to evaluate returns on investments in research for development, the increased interest in liberalizing global agricultural trade, the evaluation of ecosystem services, and impacts of climate change and adaptation (Rosenzweig and Parry 1994; Curry et al. 1990; Thornton et al. 2006; Nelson et al. 2009). This steady progress included the development of agricultural risk management analyses, evaluation of national, regional and global policies, and integration of other models with economic models for more holistic assessments, including crop, livestock, grassland, and hydrology models (Havlik et al. 2014; Nelson et al. 2013; Rosegrant et al. 2009; de Fraiture et al. 2007).

In parallel with these events that brought significant funding into development and use of agricultural system models, other events also contributed significantly to this evolution. The introduction of the first IBM personal computer (PC) in 1981 and the Apple Mac computer in 1984 led to widespread availability of computers during the 1980s. Afterward, individual researchers could work with agricultural system models that were being made available on personal computers or develop their own models. This PC revolution led to many innovations in other fields that have contributed to modeling of agricultural systems, such as computer graphics, statistical analysis, GIS, and other software being made available on desktops, notebooks, and smart phones. In addition, the development of the internet and world-wide web in the 1980s ushered in a new era of communication and technologies that led to greater collaboration among scientists and more rapid development of agricultural models.

Another innovation in computer software development is noteworthy. In 1998, the concept of open source software was developed. As the agricultural systems science community is evolving, we are finding considerable interest in creating open-source agricultural system models, with modular components and with interfaces to common databases. Already, at least one cropping system model is being offered as open source by DSSAT, which allows free access to the model source code for anyone to make changes and submit them for possible inclusion in the official model version (i.e., Cropping System Model (CSM) in DSSAT; www.dssat.net).

In parallel to funded initiatives, scientists started creating consortia and networks to enhance collaboration for specific purposes. One example was the International Consortium for Agricultural Systems Applications (ICASA; Bouma and Jones 2001), which was formed in 1993 and developed data standards for use with crop models (Hunt et al. 1994; White et al. 2013). Another key development was in the construction and release of global datasets of cropping areas, sowing dates, yields, and other management inputs (Ramankuty and Foley 1999; You et al. 2006; Monfredda et al. 2008; Ramankutty et al. 2008; Fritz et al. 2013). A milestone was reached when these global cropland cover products were developed and used for regional and global analyses of agricultural systems. Without access to data for developing, testing, and applying the agricultural system models, they are not effective.

Several events in Table 1 are associated with climate change in various ways that individually and collectively contributed strongly to advances in agricultural system models. An early contributor to modeling climate change impacts was the International Geosphere-Biosphere Program (IGBP), formed in 1986. This global project led to increasing interest in climate change and the use of models to assess what likely impacts might be under future climate conditions. Included in this work was a project on agriculture (Global Change and Terrestrial Ecosystems, or GCTE; Steffen and Ingram 1995). This project led to collaboration among crop modelers, who were beginning to see the need for comparing different models (e.g., Jamieson et al. 1998). An early motivation for model use in climate change research was the publication of the first IPCC assessment report on climate change (IPCC 1990). This led to the use of crop, livestock, and economic models to assess climate change impacts on agriculture as well as agricultural adaptation and mitigation options (Rosenzweig and Parry 1994). This then prompted crop modelers to incorporate CO_2 effects on crop growth and yield if this effect was missing, and to use the models to perform simulation experiments using current and future projected climate conditions (e.g., Curry et al. 1990; Tubiello et al. 2002; Waha et al. 2013). These changes in crop productivity were used in socioeconomic models to evaluate impacts on agricultural trade, food prices, and distribution of impacts (e.g., Rosenzweig and Parry 1994; Adams et al. 1990; Fischer et al. 1995). Many studies have been conducted since the first work that was led by Rosenzweig, Parry, and others, in particular to provide information for subsequent IPCC assessments as well as various national and regional assessments (Fischer et al. 1995; Rosenzweig and Parry 1994; Parry et al. 2004).

Unfortunately, these assessments used existing models, and funding did not provide support for improving and evaluating the models. Long et al. (2006) challenged the findings of crop models that had been developed using older data, particularly results suggesting that the positive fertilization effects of CO_2 would offset the negative effects of rising temperature and lower soil moisture. Much more data are now available from FACE (Free Air CO_2 Experiments) and T-FACE (Temperature FACE) experiments to more comprehensively evaluate and improve the interactive effects of temperature, soil moisture, and CO_2 in current models (Kimball, B. A. 2005; Boote et al., 2010). This is one of the goals of the Agricultural Model Intercomparison and Improvement Project (AgMIP; see www.agmip.org; Rosenzweig et al. 2013a).

The creation of AgMIP in 2010 is another major milestone in the evolution of agricultural models. This initiative created a global community of agricultural system modelers with the goals of intercomparing and improving crop, livestock, and socioeconomic models, and using the improved models for assessing impacts and adaptation to climate change and climate variability at local to global scales, including assessing uncertainties of those assessments (e.g., see Asseng et al. 2013; Bassu et al. 2014; Rosenzweig et al. 2013b). Since its start, AgMIP has created collaboration among virtually all agricultural modeling groups globally, creating new opportunities for substantially improving abilities to understand and predict agricultural responses to climate, including interacting effects of CO_2, temperature, and water.

Finally, the increasing interest in improving the representation of the Earth's land area in regional and global climate models has led to new approaches for modeling agricultural systems (Osborn et al. 2007). This has led various modeling groups to develop models that represent CO_2, water, and GHG fluxes and also crop growth and yield of grid-cell areas (e.g., Fischer et al. 1995; Rosenzweig et al. 2013b; Elliott et al. 2014). On the livestock side, global gridded models for feed consumption, productivity, manure production, and greenhouse gas emissions for dairy, beef, small ruminants and pork and poultry are now available (Herrero et al. 2013; FAO 2013).

Two recent events shown in Table 1 have the potential for major advances in modeling agricultural systems, but these impacts have yet to be realized. The first one is the molecular genetics revolution. During the last 20 years, the progress in mapping genomes of major crops has been impressive, and the technological advances in performing DNA analyses on plants and animals have led to rapid and inexpensive genotyping that resulted in major changes in plant breeding. The potential value of this molecular genetics information includes the abilities of crop and livestock models to predict performance of crop varieties and animal breeds in specific climate and management conditions. Early work on this has shown that it is promising, yet considerably more work is needed to quantitatively link genes to physiological performance (e.g., see White and Hoogenboom 1996; Hoogenboom and White 2003; Messina et al. 2006; Hammer et al. 2006). The molecular biology revolution is also leading to the development of new genetic strategies for pest and disease control that are likely to be ready for regulatory study in the next decade, and this may lead to new demands for systems models to explore their efficacy and safety.

The second entry in Table 1 that holds unrealized promise is greater collaboration among public and private researchers. For example, the private sector invests heavily in data collection as part of their plant breeding process. Some companies have shown interest in providing some of those data for use in evaluating and improving models in the public sector (e.g., Gustafson et al. 2014; Kumudini et al. 2014). In addition, private companies realize that agricultural system models are becoming more widely used in assessing sustainability of new technologies. This is seen through the creation of CIMSANS (see www.ilsi.org/ResearchFoundation/CIMSANS), a new public-private partnership in the International Life Sciences Institute to address sustainable agriculture and nutrition security using agricultural models). Finally, the private sector is heavily invested in molecular genetics usage in plant breeding, and some companies are making use of crop models in their plant breeding efforts (e.g., Messina et al. 2011). This provides an opportunity for public and private researchers to work together to produce more reliable models of crops and breeds for greater use of these methods in the future.

Other events have contributed to development of specific agricultural models in different countries. We do not attempt to create a comprehensive list of all such events, but instead to highlight those that played major roles in getting this work started in addition to those that had major implications globally. Between events in Table 1, model development and use proceeded, but overall progress has been slow at times. The continued dedication to develop reliable models has been one of the main features of several agricultural modeling efforts for cropping systems, livestock, and economics (e.g., DSSAT, EPIC, APSIM, STICS, WOFOST, ORYZA, CROPSYST, RZWQM, TOA, IMPACT, SWAP, and GTAP).

Table 1. *Timeline of key events that shaped the development and use of agricultural system models.*

Year	Event	Impacts
1940s - 1950's	deWit (1958) and van Bavel (1953) develop early computational analyses of plant and soil processes; Development of nutritional requirement tables for cattle (NRC 1945)	Foundation established for the application of simulation and operations research optimization in plant-soil systems research and for modeling farm animal responses to nutrients
1950 – 1970s	Demand for policy analysis of rural development	Representative farm optimization models were developed and applied by Heady and students at Iowa State University, thus establishing use of linear programming methods for agricultural production
1960-1970	Pioneers in soil water balance modeling (WATBAL) [Slatyer (1960, 1964), Keig and McAlpine (1969); Ritchie (1972); McCown (1973)]	Water balance models proved to be useful in the evaluation of climatic constraints to agricultural development. Foundations for linking soil and plant models established.
1964-1974	International Biological Program	Strong emphasis on large scale ecological and environmental studies led to development of grassland ecosystem models; laid foundation for ongoing work today
1965	UK releases nutrient requirement tables for ruminants (ARC 1965, first work since the 50s)	Very influential publication; subsequent development of feeding systems models throughout Europe.

1965-70	Early crop modeling pioneers develop photosynthesis and growth models (C. T. de Wit, W. G. Duncan, R. Loomis)	Captured imagination of many crop and soil scientists. Prompted many to follow in their steps.
1969-75	S-69 Cotton Systems Analysis Project (Bowen et al. 1973; Stapleton et al. 1973; Jones et al.1974; Baker et al. 1983)	Prompted development of several cotton models (W. G. Duncan, J. D. Hesketh, D. Baker, J. Jones, J. McKinion)
1971	Creation of the Biological System Simulation Group (BSSG)	Led to self-supported annual workshops aimed at advancing cropping system and other biological system models, continuing through 2014
1970s and early 80s	Development of early herd dynamics simulation models (Freer et al. 1970; IADB 1975; Davis et al. 1976; ILCA 1978; Sanders and Cartwright 1979, Konandreas and Anderson 1982)	Established in the developed world but some early examples in the developing world. Crucial for the advancement of whole livestock farm modeling and for representing disease and reproductive impacts
1970s	Gordon Conway develops concept of IPM in Malaysia. Huffaker Integrated Pest Management (IPM) Project begins in USA, evolves into the Consortium for IPM, ending in 1985. Global emphasis on reducing pesticide use, due to major increases in pesticide use globally and resistance in target pest populations.	Insect and disease models developed and used to help establish economic thresholds and to predict timing of threshold exceedance; some pest models were linked with crop models
Mid 1970s	Discovery of chaos in ecological systems (May, 1976)	Led to new approaches to modeling predator-prey, host-disease interactions
1972-74	Soviet Union purchase of US wheat reserves, causing major price spike (see Pinter et al. 2003)	US Government created LACIE, AGRISTARS projects to develop and use crop models with remote sensing to obtain strategic crop forecasts. Led to development of CERES-Wheat and CERES-Maize models (first published in 1986)
1974-1978	FAO development of Land Evaluation Framework in 1974 and an automated Agro-Ecological Zoning (AEZ) in 1978. (FAO 1976; 1978-81)	Provided first methodology for land evaluation on a global basis, integrating soil, climate, vegetation, and socio-economic factors, leading to many applications and efforts to improve integrated assessment approaches
1975-1982	Early pioneers in computer simulation based decision support - SIROTAC and Australian Cotton Industry (CSIRO 1980)	This was the first major initiative to put crop and pest models in the hands of farmers for decision support

1976	Launch of the first issue of Agricultural Systems, edited by C. R. W. Spedding (Spedding 1976)	This journal helped legitimize agricultural system modeling, providing a place for scientists to publish their agricultural systems modeling and analyses as well as a collection of scholarly work in this area. This journal continues today with impact factor of about 2.5
1979	E.R. Orskov establishes the 'Dacron bag technique' for measuring the degradability of feed in the rumen (Orskov and McDonald 1979)	Very influential method developed for characterizing the nutritional value of feeds, opening possibilities of new types of models; a new era of dynamic feed characterization started, leading to better animal models
1980	Soil and Water Resources Conservation Act analysis for 1980, mandate to develop a model to predict impacts of soil erosion on crop productivity	The comprehensive soil-cropping system model, (EPIC, the Environmental Policy Integrated Climate model), was developed to estimate soil productivity as affected by erosion
1980s	Growth of CGIAR Centers creates demand for assessment of economic returns to investments in agricultural research	Market surplus methods developed for estimating economic returns to investments, demonstrating high returns to agriculture research investment
1981-1984	Personal Computer (PC) revolution led by IBM introduction of its Model 5150 personal computer and the first Apple Mac computer in 1984	These new PCs led to major increases in individual access to computer power; many agricultural models began appearing on PCs
1981	Development of the first soil nitrogen (N) model for predicting crop responses under both water and N limiting conditions (Seligman and van Keulen 1981)	This model was the foundation for future soil N models in APSIM, DSSAT, and other suites of crop models
1980s through early 1990s	Development and growth of the Internet that began to connect computers globally	Ushered in new era of global communication and information technologies that has affected all areas of our lives, including agricultural system model development and use
1982 to 1986	CERES Models (Maize and Wheat) and GRO (SOYGRO and PNUTGRO) models were developed (Wilkerson et al. 1983; Boote et al. 1986)	The CERES models linked soil water, soil nitrogen and crop growth and yield together in a comprehensive fashion for the first time. They stimulated interest and activity in crop modeling in many parts of the world..
1980s	Development of duality theory and advances in nonlinear optimization via development of GAMS by World Bank	Led to advances in applications of econometric methods for production model estimation and to national and regional policy analysis models; use of new entropy methods reduced data requirements for the models

1980-1990	Influential developments in pasture modeling (Hurley pasture model – Johnson and Thornley 1983) and the SAVANNA model (Cougenhour et al. 1984)	Led to a proliferation of pasture models for intensive temperate and tropical grasslands and savanna systems. These models simulated herbage mass and accounted for sward components, which led to a more sophisticated representation of grazing processes.
1983-1993; DSSAT continuing today	USAID funded international IBSNAT project for facilitating technology transfer using systems approaches and crop and soil models	This led to the creation of the DSSAT suite of crop models that combined the CERES family of models with the SOYGRO and PNUTGRO models. The availability of the IBSNAT guidelines for data collection for crop modeling strengthened the crop model testing effort around the world.
1984 – continuing today	Dutch Government funding of the SARP (Systems Analysis of Rice Production) project at IRRI in the Philippines.	Development of a dynamic rice model that later was named ORYZA, which is still widely used today (Penning de Vries et al. 1991)
1985-1992	Earliest application of crop-soil systems models in a developing country “research for development” context – Kenya-Australia Dryland Farming Systems Project (McCown et al. 1992, Keating et al. 1991)	First PC used in agricultural research in Kenya running CERES Maize (influenced strongly by the IBSNAT minimum data set guidance) in 1985. Formed the foundation for modeling low input subsistence agricultural systems and exploring development opportunities. This experience went on to strongly influence the evolution of the APSIM farming systems simulator.
1986	Launch of the IGBP (International Geosphere-Biosphere Program) by the International Council for Science (ICSU)	Brought attention to the planet under pressure, including climate change, and helped coordinate research at regional and global scales on interactions of Earth’s biological, chemical, physical, and human systems, including influence on ecosystem modeling
1970s-1980s	Development of optimization and econometric methods for application to production risks	Broadened analysis of production to include risk management behavior
1980s until now	Modelling herd replacement decisions with dynamic programming (van Arendonk and Dijkhuizen 1985)	As computer power increased, more complex applications attempting to optimize intensive and industrial livestock production occurred.
1990	Publication of the first Intergovernmental Panel on Climate Change (IPCC) Assessment Report	Led to first use of crop and economic models for climate change impact assessments on crops at field to global-scales (e.g., Curry et al. 1990; Rosenzweig and Parry 1994); led to broad use of agricultural and ecological models that estimate GHG emissions and carbon dynamics and economic models for assessing impacts of climate change on agriculture

1990s until now	The era of livestock systems model integration (Herrero et al 1996, 1999, Freer et al. 1997)	Many soft 'modular' couplings of simulation models of individual animal performance, herds dynamics, pasture and crop models happened at this time.
1990-1994	First studies on global impacts of potential climate change on agricultural systems (Rosenzweig and Parry 1994)	These were the first studies making broad use of crop and economic models for global impacts. These studies paved the way for many other national and global impact studies of climate change impacts and adaptation.
1991-continuing today	Australian governments develop a new APSRU group for modeling agricultural systems for practical uses	This led to the now widely used APSIM (McCown et al. 1996; Keating et al. 2003) suite of cropping system models which drew on early experience with CERES, EPIC and PERFECT models but re-engineered the "farming systems" foundations.
1992	The Cornell Net Carbohydrate and Protein System is launched (Russell et al. 1992)	The CNCPS became the first commercially available dynamic model of digestion in ruminants. Its development influenced the current livestock performance models in many parts of the world.
1993-2011	International Consortium for Agricultural Systems Applications (ICASA), formed in 1993, ended in 2011	Helped crop modelers collaborate to develop standards for input data for crop models (Hunt et al. 1994), leading later to the ICASA data dictionary and data standards (White et al. 2013), now used in harmonizing model inputs in AgMIP project (White et al. 2013).
1998	Initiation of open source software movement, leading to more collaborative software development	Led to interest in providing open-source versions of widely-used crop simulation models; now being done by some ag system modelers (e.g., APSIM, DSSAT).
1999	The Livestock Revolution study (Delgado et al. 1999)	Key analysis explaining projected growth of livestock sector showing that 'as people get richer and societies urbanize they consume more livestock'. Led to acknowledgement of need for increased understanding of livestock sector for agricultural development.
1980s-1990s	Interest in trade liberalization	Led to quantitative analysis of trade policies and development of national and global agricultural trade policy models.
1990s-2010s	The molecular genetics revolution: Genome sequencing technological advances and advances in understanding of the functions of crop and animal genes; ability to genotype new lines and breeds	Led to still evolving efforts by various public crop modeling groups and by seed companies to connect ecophysiological crop models for plant breeding and management purposes (e.g., see White and Hoogenboom 1996, Hoogenboom and White 2003; Hammer et al. 2006; Messina et al. 2006).
1990s-2000s	Sustainable agriculture movement; greater concern on environmental consequences of agriculture	Led to incorporation of biophysical processes into farm household, econometric and programming approaches; also led to development of "tradeoff analysis" approach; spatial data and tools increasingly used to develop spatially explicit biophysical and economic models

Late 1990s - 2000s	Construction and release of global datasets of cropping areas, sowing dates and yields (Ramankutty and Foley 1999, Ramankutty et al. 2008)	Allowed researchers to run simulations at finer resolution over greater model domains with more clearly documented assumptions and inputs.
2000s	Increasing interest in greenhouse gas (GHG) mitigation and the importance of ecosystem services	Led to models for analysis of mitigation of GHG in agriculture via soil C sequestration, afforestation, reduced livestock emissions; also led to linkages of economic models with crop, livestock, hydrology, and ecosystem models.
2001-2003	European Society Agronomy meeting hosts special session on modeling cropping systems. Published as Volume 18 European Journal Agronomy	This meeting led to a special issue of European Journal of Agronomy (vol 18) in which comprehensive papers on the major modeling systems, namely DSSAT, APSIM, CROPSYST, STICS, Wageningen models. Over 2000 citations for models in this publication.
2006	Representation of CO_2 effects in crop model simulations challenged by Long et al. (2006)	Opened a debate between plant experimenters and modelers on the skill of crop models for yield prediction in future climates; prompted interest in more evaluations of CO_2 effects interacting with temperature, other factors
2005-2009	European Union funding of the System for Environmental and Agricultural Modeling: Linking European Science and Society (SEAMLESS)	This led to major collaboration across Europe for developing models for use across scales, from field to farm, country, and EU levels.
2005-2010	Development of Earth system models, components of general circulation models (GCMs)	Led to new methods for coupling crop simulation models to land surface schemes of numerical climate models; Challinor et al. 2004.
2006	FAO Livestock's Long Shadow report (Steinfeld et al. 2006)	Demonstrated the large environmental footprint of livestock leading to programs for assessing and reducing the environmental impacts of livestock. Most of this work was done through modeling.
mid 2005s onwards	Development of global livestock models (Bouwman et al. 2005; FAO 2013, Herrero et al. 2013)	Global integrated assessment of livestock systems now possible at high resolution including land use, emissions, economics, biomass use and others (Havlik et al. 2014; Cohn et al. 2014, PBL 2013, Bouwman et al. 2013 and others) and their links to other sectors (crops, forestry, energy, etc.).
2010	Creation of the Agricultural Model Intercomparison and Improvement Project (AgMIP), a global program and community of agricultural scientists	This initiative led to model comparisons and initiatives for improving models, capturing the imagination and interest of agricultural modelers worldwide (Rosenzweig et al. 2013a,b; Asseng et al. 2014).

2010s	Increasing interests by the private sector in agricultural system models	Some companies create their own crop modeling teams, others start working in public-private collaborations.
2013 –	Increasing realization of the need to increase food production to meet needs of > 9 billion by 2050, including challenges of climate change and sustainability of natural resources	This realization is leading to greater interest in use of agricultural system models to help guide investments and development and to greater interest by the private sector.

3. Characteristics of Agricultural System Models

3.1 Purposes for Model Development

There are two important motivations for agricultural model development; scientific understanding, and decision/policy support (e.g., Boote et al. 1996; Bouman et al. 1996; van Ittersum et al. 2003; Ritchie 1991; McCown 1996). The first of these motivations is to increase basic scientific understanding of components of agricultural systems or understanding of interactions that lead to overall responses of those systems. Van Ittersum et al. (1998) referred to models with this purpose as explanatory. Models developed to increase scientific understanding tend to be mechanistic models as they are usually based on known or hypothesized control of physical, chemical, and biological processes occurring in crop or animal production systems. Examples are mechanistic models of photosynthesis (e.g., Farquhar et al. 1980) and water movement in soils (e.g., model implementation of the Richards (1931) equation).

At the basic science level, models developed to increase understanding are used as tools to address research questions about control of processes, magnitudes of responses, and interactions. Modeled outputs are compared with those that are measured in the laboratory or field for testing the understanding that is embedded in the model. For example, transport of water or mineral N through a soil involves many processes that affect the correct balance of water or N. Likewise, the flux of carbon dioxide in a field can be measured instantaneously in flux-site experiments. There are, however, many contributors to CO_2 flux including photosynthesis, aerial crop respiration, root respiration, and soil organism respiration, all of which are affected by the aerial and soil environment as well as by crop type, age, and condition. For livestock, the partitioning of nutrients for different physiological functions (growth, lactation, pregnancy and others) and the control of voluntary feed intake as well as their interactions and feedbacks have received considerable attention (Forbes, 1986; Illius and Allen, 1994). Such models, developed to increase scientific understanding typically describe processes at fine time scales (e.g., instantaneous photosynthesis and transpiration processes, hourly nutrient supply in animals).

Such explanatory models of agricultural systems typically include a large number of parameters, some of which may be unknown or only known with relatively large uncertainties. And they may require other explanatory input information that may not be readily available for general applications, such as the spatial variations in the relationship between soil water and water potential. Also, uncertainties in some of the hypotheses and assumptions used in developing mechanistic models make outputs uncertain and often less useful to those outside of the model development group. Functional models (Addiscott and Wagenett 1985; Ritchie and Alagarswamy 2002), which may also be referred to as phenomenological models (developed by using data to model relationships), are based on empirical functions that approximate complex processes, such as a crop's interception of energy using plant leaf area (as an indicator of biomass) and radiation use efficiency (RUE – a measure of biomass produced per unit of radiation intercepted). This type of function requires field data to estimate RUE and usually produces reasonable results when compared to field measurements. Another example of an empirical approach is the simulation of potential evapotranspiration using the well-known functional Penman-Monteith or Priestley-Taylor equations (Allen et al. 1998), which have been used successfully for decades even though they are highly simplified compared to more mechanistic evapotranspiration models.

Explanatory models may include various combinations of mechanistic and functional model components. The choice of relationships used by different modeling groups to represent processes and components is one of the main reasons that there have been multiple models developed of the same crop, livestock, and farming systems. For these reasons, currently developed agricultural system models have different levels of complexity, different parameters and input requirements, and vary in their abilities to predict system performance. This has been demonstrated recently by the AgMIP wheat and maize model intercomparison studies that found large variations among multiple wheat and maize model yield predictions. The median of multiple models was a better predictor of crop yield across multiple sites than any single model in these studies (Asseng et al. 2013; Bassu et al. 2014).

The second overall purpose for developing models, to provide information for supporting decisions and policies, requires models that describe how the agricultural system responds to the external environmental drivers as well as decisions or policies under consideration (referred to as descriptive models by van Ittersum et al. 1998). Users of such models may be interested in prediction of important responses that would help them make a decision, or they may be interested in how the system would respond if a particular decision was made. They may want to analyze alternative designs of agricultural systems or explore responses to different policies at crop, livestock, farm, or regional scales (Thornton and Herrero 2001; van Ittersum et al. 1998). Such models may or may not increase scientific understanding and they may have varying degrees of explanatory mechanisms, but the key requirement is that the models provide reliable system response information that decision and policy makers need.

Models for increasing scientific understanding of agricultural systems will continue to be pursued using various scales and approaches. While these models form the basis for the decision-enabling modeling, our focus is on next generation agricultural system models for use in planning and strategic decision analyses. A key task is the evaluation of tradeoffs among possibly conflicting objectives of decision/policy makers at various levels, from field and pasture to farm, landscape, and regional scales, and for smallholder to large industrial scale farmers.

3.2 Approaches for Modeling Agricultural Systems

Several dimensions are needed to describe the types of models that have been developed in the past for use in improving decisions and policies. Here we discuss the major types of models that produce response outputs that are of interest to decision/policy makers. First, statistical models have been developed using historical data sets on system responses, such as crop yield, milk production, and prices of commodities. For example, statistical models — fitting a function to predict crop yield using observed weather variables and crop regional yield statistics over multiple years — were the first crop models used for large-scale yield estimations. Average regional yields were regressed on time to reveal a general trend in crop yields (Thompson 1969). It is assumed that the data used to create statistical models are samples of a population such that the model can be used to predict regional yields in a new years with different weather patterns. In general, results of statistical models cannot be extrapolated to other places because of variations in soils, landscapes, management, and weather not included in the population of information from which the statistical relationship was derived. Furthermore, they are not well suited to estimate climate change impacts in the future because they cannot capture changes in management (adaptation), soil properties, pests and diseases, and the influence of increasing atmospheric CO_2 concentrations (beyond the range of historical data). Despite these limitations, statistical models are still very useful. When sufficient data are available to develop such models; they can provide many insights about historical influences on past yields and can inform other kinds of models (Lobell et al. 2011; Schlenker et al. 2013).

A widely used approach for modeling agricultural systems can be classified as dynamic system simulation models. In contrast to the statistical approach, these models have functions that describe the changes in systems states in response to external drivers (e.g., weather and management practices), and how those changes are affected by other components in the system (see Wallach et al. 2014). This approach is used for all types of models, including crop, livestock, and farming system models, with model outputs being the values of model state variables over time (e.g., typically daily outputs for crop and pasture models). These dynamic models can be used to simulate multiple responses for the specific time and variables as needed (Wallach et al. 2014), and thus can compare effects of alternative decisions or policies on tradeoffs among those various responses. These dynamic system models may have mechanistic and functional components. Examples of dynamic models for cropping systems are those in the DSSAT suite of models (Jones et al., 2003), and APSIM (Keating et al. 2003), CROPSYST (Stockle et al. 2003), and EPIC (Williams et al. 1983, 1989). However, because some of these models are very complex, containing many descriptive variables and parameters thus requiring many inputs and long run times, some authors have shown that reduced form or summary models can be derived from much more complex models for specific purposes (e.g., Jones et al. 1999; Chikowo et al. 2008; Dzotsi et al. 2013). This approach is particularly useful when one wants to integrate crop models, for example, into models of more comprehensive agricultural systems such as economic analyses at farm, national, or global scales.

Similarly, dynamic livestock models include Ruminant (Herrero et al 2013, 1996); LiveSim (Rufino et al 2009); CNCPS (Ruitz et al. 2002), Grazplan (Freer et al. 1997), GLEAM (Gerber et al. 2014), amongst others, and farming system models include IMPACT-HHM (Herrero et al 2007), Gamede (Vayssieres et al 2011); IAT (Lisson et al. 2010); APSFARM (Rodriguez et al. 2014), and FARMSIM (Van Wijk et al. 2009). For a detailed review see van Wijk et al. (2014).

One other point to make about the use of models for decision-making is the type of decision being considered. To date, many models have been developed to help inform tactical decisions, such as when to apply a pesticide, when to irrigate, or when to sell livestock. However, the models that are most useful for those kinds of decisions are narrow in scope. They are not about how to best manage a crop for multiple inputs over a full growing system altogether, but simply when to perform those predetermined management operations. They only predict when a particular threshold is reached that has previously been shown to provide effective management. To address these broader decisions, a cropping system model might be used to develop Apps in the future for use on smart phones or other hand held devices (e.g., see www.agroclimate.org; Fraisse et al. 2006; Janssen et al. 2015 – paper 3 in this series).

For planning and strategic decisions, it is clear that multiple responses and tradeoffs need to be tested. Dynamic models of component subsystems (e.g., simulating daily growth and partitioning of biomass) can be used to represent functional responses (e.g., end of season grain, biomass yield, or residues in response to a range of nitrogen fertilizer use). Ideally, such model-simulated responses can be used to infer responses by real systems. Virtual experiments (simulations) using the models can thus complement real experiments, but there is a need to evaluate model responses relative to real system responses for a range of conditions to establish confidence in the model and also provide a measure of uncertainty. In the past, very little has been done to establish uncertainty of agricultural systems models until recently (e.g., Rosenzweig et al. 2013; Asseng et al. 2013).

3.3 Spatial and Temporal Scales of Agricultural System Models

Users of models or information derived from them and the models themselves vary considerably across scales as indicated in Figure 2. Similarly, the scope of the system being modeled and managed varies depending on the questions being asked and the decisions and policies that are being studied. Users in Figure 2 are not necessarily those who run the models; instead, they are those who want information about responses of the systems to different ways of managing them in whatever physical, biological, and socioeconomic climate conditions are involved. Thus, for example, a set of simulation experiments would be conducted by a researcher or advisor to address specific questions about alternative decisions or policies to help them make more informed decisions. Results from the simulation experiments could be summarized into advisory fact sheets or policy briefs for users. Or, results could be summarized in decision-support systems that are designed to provide information for key decisions of users (e.g., see www.agroclimate.org that targets extension agent and farmer users). Participatory modeling, where the development of a model is accomplished by model developers and stakeholders working together and discussing model results to refine simulations and better represent stakeholders' objectives, has also been used successfully with many farming communities (Thornton and Herrero 2001; McCown 2002).

3.3.1 Field Scale

The scope of the system is important in determining what type of model is needed and what users are being targeted. Agricultural system models have been developed at all of the scales shown in Figure 2. The system could be at the field scale where one wants to know the best management practices that meet production, profitability, and environmental protection goals of a farmer who is producing the crop in an area with strict environmental regulations (e.g., He et al. 2012). Cropping system models are used to predict how much economic yield the crop will produce and how much nutrient leaching would occur under different combinations of management practices and crop seasons. Similarly, the system could be a livestock-system managed by a rancher or dairyman. Livestock models are designed to predict herd or animal performance under different combinations of breeds and management. They may predict the number of livestock of different ages by sex and the body mass, or milk production per day of each lactating cow, all influenced by herd management and marketing of meat, calves, and/or milk. Thus, at the field or enterprise level, biophysical models are used to analyze responses similar to the way that experiments on the real systems would be analyzed. In many cases, these models are used to perform simulation experiments in combination with limited treatments in real experiments to help provide confidence in simulated experiment results when extending them to a wider range of options than could be tested in the real world.

Field scale models usually assume homogeneity of conditions horizontally across the field, but may consider that the soil properties vary vertically with depth. Spatially homogeneous models are also referred to as "point" models implying that all points in a field area have the same properties. These point or field models (e.g., crop models) are also used to simulate responses at more aggregate scales by providing them with spatially-varying inputs (e.g., spatially-variable vertical soil properties, daily weather data, and/or management). For example, a crop model may be used to simulate multiple homogeneous fields across a farm, each with its own set of input conditions. As such, models of any particular scale in Figure 2 may be represented by multiple instances of models of smaller areas, thereby serving as building blocks in a hierarchical sense.

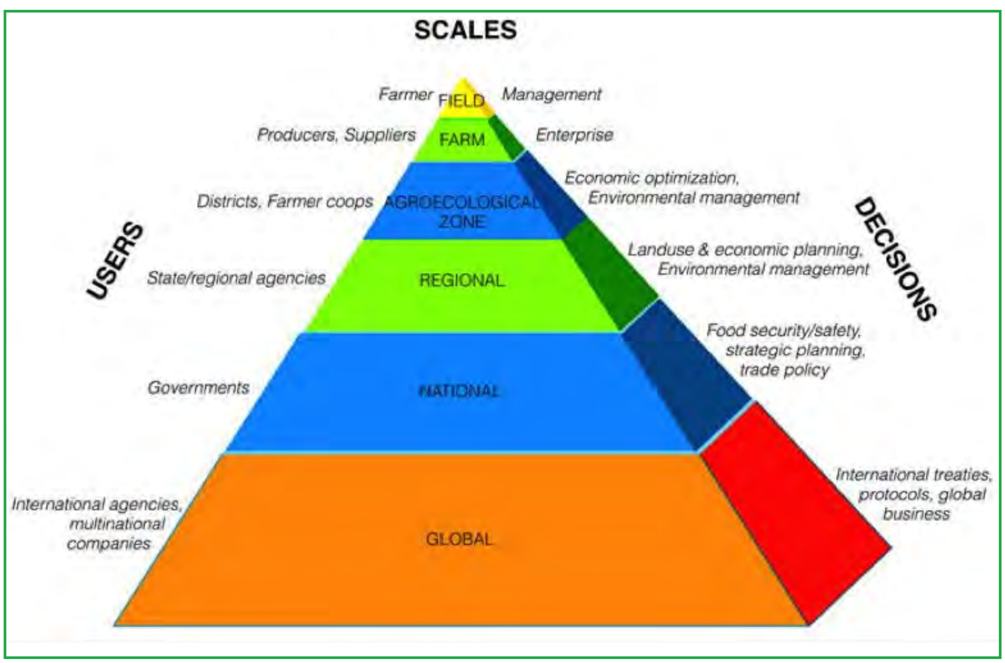

Figure 2. *Scales at which agricultural system models are developed along with types of users and decisions and policies of interest.*

3.3.2 Farm and Broader Scales

An agricultural system could also be defined as a farm with land area on which different crops and livestock are produced, each of which is managed by a farm family or business entity. In this case, the enterprises of a farm interact in various ways, which will be described later. At a broader spatial scale, one may define an agricultural system as the land area in a region, district, or landscape that produces a particular commodity or various crops. The system model for that set of users predicts total production of the crop or crops in that area as affected by weather, soil, management and socioeconomic conditions, including a capability to evaluate decision and policy options. This landscape or regional model may also predict the amount of nutrient leaching or soil erosion for particular practices and policies being analyzed. Depending on the goals of the users, different approaches are used to develop the system model. But, typically at this scale, the models should include aspects of biophysical responses of crops and livestock as well as socioeconomic, environmental, policy, and business issues. These same characteristics of system models are important at national and global scales, in that biophysical, socioeconomic, and policy components are needed to model the important interactions and production, environmental, and economic responses to different decision and policy options. Recent years have seen increased interest in studying the interaction of agro-ecosystems with other managed and unmanaged ecosystems. This has several motivations, including understanding the importance of ecosystems services such as pollination and biological pest control provided to agriculture by natural habitats as well as issues of managing biodiversity in landscape mosaics.

Figure 2 shows users that range from farmers to policy makers and businesses that are interested in improving decisions and policies ranging from field, landscape, regional, national, and global scales. The delineation of the land area over which decisions and policies are made vary considerably, depending on the stakeholder/user and his/her interests. At each scale, the landscape can be decomposed into areas delineated by agro-ecological boundaries (such as a watershed) or into areas delineated by socioeconomic boundaries, such as the political boundaries of a district or country. Models at each of these scales may be developed by using component models of smaller areas. For example, a national model may make use of field scale crop models to simulate production across many districts then aggregated to the national scale for use in an economic model of the policy impacts on the aggregate production or its variations across districts. An alternative to this approach would be to use an aggregate national production model.

Agricultural system models at each of the scales in Figure 2 are imperfect predictors of real system performance. To quote a famous statistician, "All models are wrong, but some are useful" (Box and Draper, 1987, p. 424). Model developers make assumptions about what components to include in the system, how these components interact, and how they respond to the environment and to management practices and policies. The models themselves and their performances also depend heavily on the data used to develop and evaluate them. In the next section, we present examples of several of the most widely used types of agricultural system models, focusing on their general capabilities and limitations relative to their applications and user needs. Then, we present five specific use cases, explain how existing models have addressed these types of uses, and describe limitations of existing system models for these cases.

4. Levels of Detail, Comprehensiveness, and Complexity

Here, we identify and summarize five types of agricultural systems that are modeled as subsystems for a wide range of applications by users at different scales. For each type, we identify the responses that the models are generally expected to produce and the factors to which they respond. We also summarize the main approaches used in the models and the types of data needed to develop and use them. Example applications are cited and general limitations of current models described.

4.1 Cropping System and Grassland Models

The basic characteristics of cropping and grassland systems are fundamentally the same in that they describe crop or grassland agro-ecosystem growth and yield responses to climate, soil, species characteristics, and management. However, there are several aspects of grassland/rangeland modeling that present unique challenges relative to modeling cropping systems. Many of these challenges stem from the requirement that grassland models represent several interacting species, including perennial and woody species of grasses. Persistence of plants over multiple years forces the models to consider residual effects over time. Dependency on soil-derived nutrients or human-induced disturbances like fire reinforce the longer-term perspective needed for grassland modeling. Thus, although most biophysical processes are similar relative to photosynthesis, growth, water and nutrient uptake from soil, etc., additional considerations for modeling grasslands are presented below.

4.1.1 Model-Simulated Responses of Interest to Users

The most common response variable modeled for cropping systems is yield, which may be grain, tuber or forage biomass yield. Although statistical models may be useful for predicting these biological yields in response to some combination of weather conditions, nutrient levels, irrigation amounts, etc. (e.g., Schlenker and Lobell 2010; Lobell et al. 2011), such models are only able to predict responses that have been measured under specific conditions in the past. In contrast, dynamic cropping and grassland system models may simulate these biological yields in addition to other responses that may be important to analysts, such as crop water use, nitrogen uptake, nitrate leaching, soil erosion, soil carbon, greenhouse gas emissions, and residual soil nutrients. In addition, these dynamic models can be used to estimate responses in places and time periods in which there are no prior experiments. They can be used to simulate experiments and estimate responses that allow users to evaluate economic and environmental tradeoffs among alternative systems. The simulation experiments are able to produce responses to the various climate and soil conditions, genetics, and management factors that are represented in the model.

4.1.2. Factors to Which Cropping and Grassland Systems Respond

Many factors affect crop growth and yield in agricultural fields. One innovation of the early pioneers in crop modeling was to categorize the crop production situation being modeled so that one could narrow down the many factors that need to be included in the crop model (Bouman et al. 1996; van Ittersum et al. 2003). Figure 3 summarizes three overall crop production levels and shows the factors that influence each. The potential production level is defined as crop production that is determined completely by the defining factors of CO_2, radiation, temperature, and crop characteristics (e.g., genetic control of physiology and phenology and the canopy architecture). This potential production level is rarely achieved in real production situations, but under highly intensive management (supply of adequate water and nutrients and control of insects, weeds, and diseases), production approximates the potential level for the specific CO_2, temperature, radiation, genetics, and canopy architecture used). For example

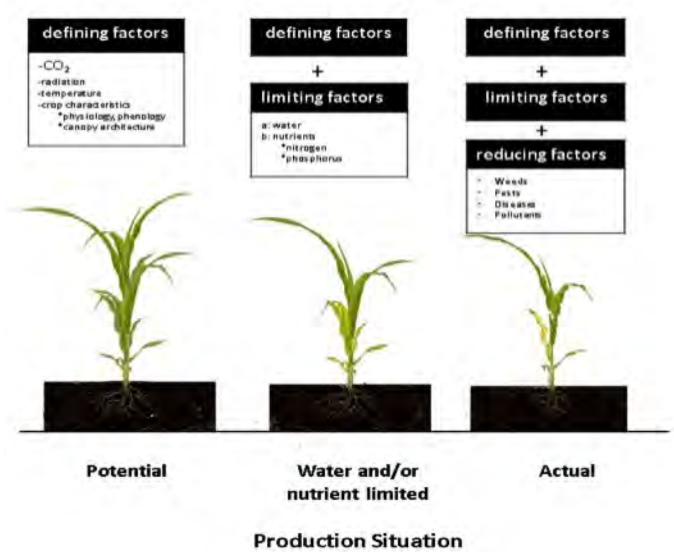

Figure 3. *Diagram of production situation used to characterize factors included and excluded from cropping system models to help guide their development and inform users of their applicability to address different questions. Adapted from van Ittersum et al. (2003).*

crops grown in greenhouses or in intensively managed fields in some regions produce yields that are at or near the potential level. Models of cropping systems at the potential level incorporate crop phenology and growth processes, including partitioning of biomass growth into grain and other plant parts, with the defining factors modeled to affect these processes.

The next production situation is referred to as water-limited and/or nutrient-limited production (Figure 3). At this level, the defining factors are still important, but also there may be limitations of water and/or nutrients needed to achieve the growth potential. Crop models that simulate water and/or nutrient-limitations must include soil water and nutrient component models that simulate the time-varying availability of water and nutrients, the uptake of these resources, and reductions in growth and development if they are not adequate to meet demands. Most cropping and grassland system models contain component modules that simulate soil water, nitrogen, and carbon dynamics because of the critical importance of these resources in determining yield globally. Although some models also simulate soil phosphorus, most current cropping and grassland systems models have limited or no capabilities to simulate responses to phosphorus, potassium, and micro-nutrients. Models that include water, soil N and soil C dynamics are complicated not only because of the physical and chemical processes that occur in the soil, but also because of the complexities in managing these resources (including water-harvesting, drip irrigation, types of inorganic or organic fertilizer applied, micro-dosing, etc.).

Finally, the actual production level (Figure 3) includes additional factors that reduce growth and yield (insects, diseases, weeds, and pollutants). Whereas some crop models have the capabilities to introduce damage by diseases and insects (e.g., Boote et al. 1983; Pinnschmidt et al. 1995), modeling these reducing factors has not kept up with advances in crop modeling. Most groups modeling cropping and grassland systems do not include these factors. Thus, most current models are not capable of simulating responses to pest and disease damage or to management of these factors using resistant varieties, agro-chemicals, or other approaches. This is a major limitation of current models for some applications.

4.1.3 Components of cropping system models – crop, soil, atmosphere, management

Generally, dynamic crop models include those factors at the potential yield level (Figure 3) in addition to water- and nitrogen-limited production level. However, the ways that different models include those factors vary. Figure 4 shows a schematic of the components in the Cropping System Model (CSM) that incorporate the CERES and CROPGRO models in DSSAT (adapted from Jones et al. 2003). This model has the capability to include soil water, nitrogen, carbon, and phosphorus dynamics as well as to introduce pest and disease damage into some of the crop models using the approach described by Boote et al. (1983). It also has the capability to simulate multiple seasons so that the carry-over changes in soil water, N, and P are simulated to represent longer-term changes in soil resources in response to different management systems (Porter et al. 2010; Basso et al. 2011)

A number of other cropping and grassland system models have similar components and capabilities (e.g., APSIM, CROPSYST, EPIC, and SALUS), although those models do not have components to simulate pest and disease limitations. Some models (in particular APSIM) have the ability to simulate intercropping of some crops (Thorburn et al. 2014). The current situation is that modules from one set of models are not compatible with other systems. For example, a major effort would be needed to incorporate the capabilities to simulate intercropping into the DSSAT CSM or to incorporate modules from the DSSAT CSM for causing pest and disease damage to crops in the APSIM system.

4.1.4 Approaches

Most current cropping system models are modeled using a similar structure as shown in Figure 4. Most operate on a daily time-step such that photosynthesis, biomass growth, phenological development, and partitioning of biomass into grain, leaves, stems, and roots are computed daily to update the state variables of the crop. In addition, soil-water processes including rainfall, infiltration, runoff, percolation, redistribution, and plant uptake are computed, and changes in soil nitrogen are calculated in order to simulate the time-changes in soil water, soil N, and crop biomass on a daily basis. However, the details of how different growth, hydrology, and soil nutrient processes are represented vary among models.

The choice of whether to use functional vs. mechanistic approaches to model processes depends on the modeling team's knowledge of the system, data that they have for parameterization, and their experience in developing and evaluating models. This is one of the main reasons that different models produce different responses when used to simulate the same experiment (e.g., see Asseng et al. 2013; Bassu et al. 2014). Most models use simplified functional equations and logic to partition simulated biomass into various plant organs. The functional models also primarily use "capacity" concepts to describe the amount of water stored in a soil and available to plants as compared to using potential energy of soil water and "instantaneous rate" concepts from soil physics. The difference between the upper and lower limits of soil water-holding capacity determines the amount of water available to plants. In this type of soil water model, water movement and its availability for crop growth are represented by functional equations on a daily time step, even though infiltration and runoff processes may occur much faster.

4.1.5 Additional considerations for modeling grasslands

Grass stands, whether in planted pastures or grazing-lands with native species, are usually mixed stands comprised of a variety of grasses and forbs, including legumes and sometimes woody species (Allen et al. 2011). Unlike croplands, the diversity of species generally precludes use of a single-species parameterization, since species vary in their ability to compete for space, water, nutrients (most commonly nitrogen), and light. Grassland models generally represent plant behavior and competition between herbaceous plants using (1) a set of species, each independently parameterized; (2) amalgamations of plants into parameters for plant functional types (e.g., warm-season grasses, legumes, etc.); or (3) community-averaged parameterizations (Taubert et al. 2012). While requiring more effort for parameterization, these amalgamated approaches enable representation of changes in plant community composition over time in response to climate change, for example, as a function of competition among plant populations driven by plant establishment, growth, growth form, senescence, and mortality.

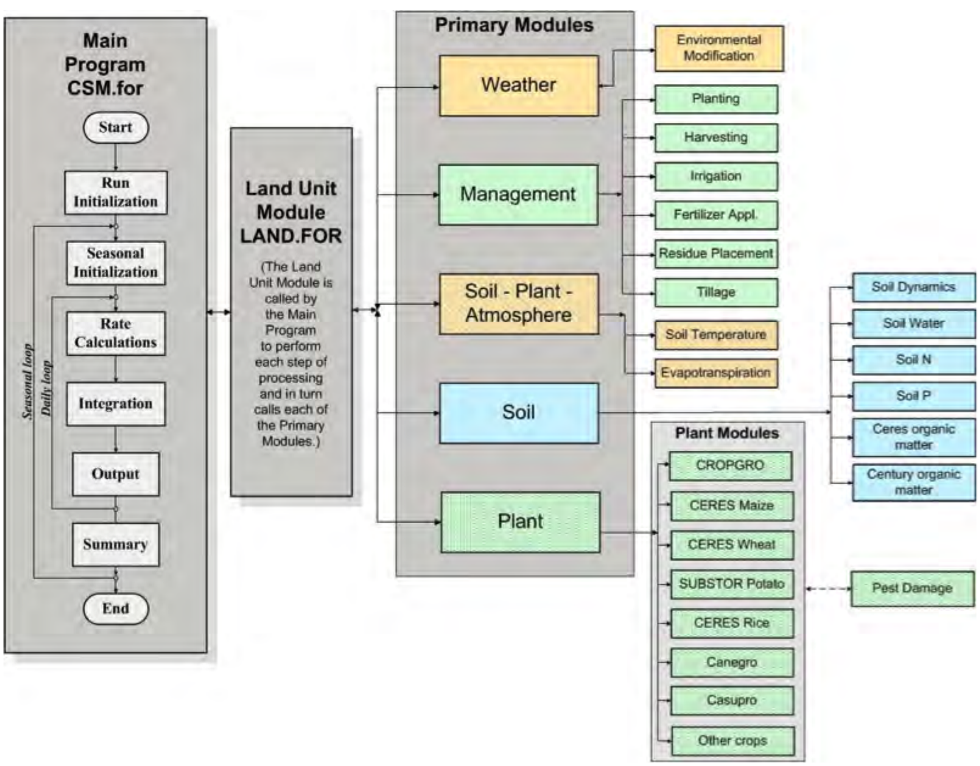

Figure 4. *Land, soil, crop, climate, and management components in the DSSAT Cropping System Model (CSM, adapted from Jones et al. 2003).*

Trees, dynamic components in much of the world's native grazing land area, can have significant impacts on ecosystem function (Schlesinger et al. 1990). Representing tree/grass competition is challenging because trees respond differently to a variety of drivers (such as fire, grazing, and CO_2 concentration), and depend on plant population characteristics (e.g., seed banks). Shifts in plant community composition can be self-reinforcing due to co-occurring population and biophysical changes (D'Odorico et al. 2012). Dynamic vegetation modeling approaches are used to represent competition between herbaceous and woody types for water, nitrogen, light, and space. Dynamic rangeland vegetation models and state-and-transition models identify a set of plant communities that tend to resist change due to disturbance, but describe drivers (e.g., fire, grazing, climate change) that can lead to a transition to another quasi-stable plant community (Stringham et al. 2003). Expansion of woody species and increases in woody cover are widespread phenomena that under many but not all environmental conditions lead to the transition of early successional communities dominated by grasses and forbs to forests (Van Auken 2000). Studying woody encroachment and understanding the importance of competing drivers has been challenging, in part because of the slow processes driving changes (e.g., Morgan et al. 2007). These slow changes are reflected in the drivers of transitions in state-and-transition models and contribute to uncertainty in our ability to represent longer-term changes in the tree-grass balance. Ecological succession has been studied by plant ecologists since the pioneering work of Clements (1874-1945) and greater interaction between agricultural system and ecological modelers is likely to be mutually beneficial.

Grazing animals of all kinds have an impact on plant productivity by removing photosynthesizing tissues, altering light transmission through the canopy, and affecting plant allocation patterns and differentially influencing species mortality and recruitment rates in grasslands (Diaz et al. 2007). Such changes to groups of plants (species, functional groups, etc.) can drive changes in the competitive balance and thus plant community composition. Whereas grassland models incorporating species or plant functional types are capable of representing grazing-induced changes in the competitive balance, such models that represent plants with a set of community-wide parameters usually rely on some combination of LAI (Leaf Area Index)-driven reduction in production potential along with grazing response curves. In grasslands/rangelands, grazing (or cutting for hay) removes some of the productive capacity of the plants, meaning that models cannot rely upon deterministic growth curves, but must have the capacity to forecast growth for plants with an amount of biomass or leaf area that varies independent of the time-of-year or climate. Furthermore, there can be significant differences in growth rates after a grazing event between and even within species (Milchunas and Lauenroth 1993; Vesk and Westoby 2001).

4.2 Reduced form summary crop models

The types of crop models discussed above respond to soil, weather, genetics, and management inputs. As noted, the number of factors to which the models respond vary among models and evolve as modelers attempt to make them more comprehensive and universally applicable. In contrast, some researchers who want to apply crop models do not have all of the inputs needed or they may want to imbed the crop model in an economic or other model for analyzing responses across scales. There are good examples where researchers have used more comprehensive crop models to create reduced forms of models that have much fewer input requirements, run much faster, and produce the responses needed for specific applications. One good example is described by Chikowo et al. (2008) in which the APSIM cropping system model was used to generate parameters and variables needed to operate a much simpler field scale crop model (NUANCES-FIELD, Titonell et al. 2007). The reduced form summary model responds to nitrogen and phosphorus levels for different soil characteristics and management inputs, which is what the farm scale model required. A similar approach was used by Dzotsi et al. (2013), demonstrating that reduced forms of maize, peanut, and cotton models parameterized using the DSSAT cropping system model accurately reproduced data and DSSAT simulations across a range of locations. This approach allows researchers to produce situation-specific summary models that approximate the responses of a more comprehensive model for use in broader scale analyses that may involve socioeconomic, livestock, and environmental sustainability components as well. Other advantages of such reduced form models is their ability to be more easily understood by those not involved in agronomy, crop or other discipline-specific modeling, as well as more rapid run times and smaller memory requirements.

4.3 Livestock systems

Livestock systems are complex and require modeling at several levels: the animal, the herd, and its interaction with its environment via consumption of feed, use of land and water, and other resources. Several types of models have been used in the past to describe the different components of livestock systems. The most commonly used are summarized in Table 2.

Animal performance models: A central element driving production, profitability, and efficiency in livestock systems is animal performance. Hence models that predict animal productivity in terms of meat and milk are the most commonly used type of livestock model. Precursors to performance models have existed since the 1940's when the first feed requirements for livestock were developed (NRC 1945). Since then, many have been built and refined regularly across the US and Europe (AFRC 1993, NRC 2001, SCA 2007). Nutrient requirements models are the workhorse of the feed industry for ration formulation (linked to linear programming models for least-cost ration formulation) and for recommending changes in feed management to farm advisors. They are often based on a mixture of statistical regressions derived from experimental data and mechanistic principles of the energetics and protein metabolism of mammals.

Animal performance models usually require information on the animal (i.e., bodyweight, target milk production, milk composition, breed, days pregnant) and the feeds (digestibility and crude protein at the minimum, but increasingly several parameters related to the fiber, mineral and/or amino acid composition of feeds are also used). They also need an estimate of feed-intake, perhaps the most important parameter. While these models are good for calculating feed requirements, dynamic models of digestion are more accurate at predicting the nutrient supply to animals under a wide range of conditions (from the high-yielding dairy cow to the smallholder goat) (Tedeschi et al. 2014; Herrero et al. 2013; Illius and Allen 1994; Fox et al. 1992), because they predict intake more accurately, and also because they can deal with more complex diets and their interactions. Additionally some of these models predict methane production in ruminants and manure quantity and quality, which are important for estimating GHG emissions and the role of livestock in nutrient cycles. Typical questions solved with these models are 'what if' questions around the impacts of different feeding practices (different feeds and/or different quantities) or changes of animal types (breeds, different production potential) on animal performance (meat and/or milk output, GHG emissions, manure output).

Herd dynamics models: The objective of herd dynamics models is to follow the evolution of the herd over time in terms of animal numbers and herd structure. Herd dynamics simulation usually starts by splitting a herd of animals into cohorts of different ages or weight, and sex. These cohorts are specified with different mortality, reproductive, selling and replacement rates. Adult females will produce offspring at specified reproductive rates, these grow or die, become part of the next cohort, and get sold, and the cycle continues. The better examples of these models include interactions between animal nutrition and reproduction to drive reproductive and mortality parameters stochastically (Konandreas and Anderson, 1982; Rufino et al., 2009). This is important as feed availability or supplementation strategies have significant impacts on herd reproduction and performance. Some applications of herd dynamics models include estimating optimal stocking rates and carrying capacities, assessing the impacts of reproductive technologies and/or reductions in mortality, and predicting removal of biomass from crop or pasture systems. These models are also widely used by livestock epidemiologists for estimating impacts of diseases on herd mortality and morbidity. They have also been used with dynamic programming for optimizing replacement decisions in commercial dairy herds (van Arendonk and Dijkhuizen 1985), or in linear programming applications for studying optimal sales policies, herd sizes, etc.

Integrated livestock systems models: These models represent whole livestock farms and their key compo-

Table 2. Livestock models and some types of questions they can help answer

Type of model	Simulation	Outputs
Individual animal performance	- Prediction of performance - Assessing impacts of alternative feeding practices on yields, GHG emissions - Assessing sustainable intensification strategies - Impacts of changes in breeds or types of animals - Yield gap studies - Assessment of manure quantity and quality - Establishment of substitution rates between feeds - impacts of feed scarcity	- Least-cost diet formulation - Optimization of supplementation practices - Nutrient synchrony studies - Amino acid adequacy for high yielding dairy cows - Optimal feed management for different types of animals in a herd
Herd dynamics	- Impacts of reproductive management - Stocking rate decisions - Impacts of climate variability on herd dynamics - Epidemiological studies of disease spread and impacts on herd numbers, profitability - Value chain studies	- Optimal replacement - Optimal times to sell animals - Impacts of climate variability on herd dynamics
Integrated livestock systems	- Assessment of the feasibility of new management strategies - Land use management strategies - Best grazing practices - Feed conservation strategies - Trade-offs in the use of resources technology targeting - Identification of key constraints (labour, etc) - Gender sensitive strategies - Selling and replacement strategies	- Optimal herd sizes - Land-use management in livestock farms - Intensification potentials - Trade-offs in the use of resources - Impacts of input and output prices - Impacts of intensification or environmental policies - Optimal stocking rates and carrying capacities - Selling and replacement strategies - Matching seasonal feed resources to herd dynamics

nents (Figure 5). The complexity of some livestock systems justifies the need to build whole-system models using simulation and optimization techniques to represent different components and their interactions (Herrero et al. 1996). For example, grazing management strategies cannot be made without considering herd and nutritional management, since herd dynamics or feed supplementation practices will determine the grazing intensity, use of forage, and subsequently animal performance. Thus, the simulations representing the biology of livestock enterprises includes flexible models representing pasture growth, structure and quality; individual animal performance to test nutritional strategies; and population dynamics describing management practices at herd or flock level (i.e., stocking rates; sales of animals; mortality or replacement rates; calving intervals; etc.), which subsequently determine animal numbers and their age or physiological state (lactating vs. pregnant cows, heifers, calves, etc.) classes (Freer et al. 1997; Loewer 1998; Johnson 2002).

In some instances, biological simulation systems are used as input-output coefficient generators for linear programming models to aid in the selection of management strategies in livestock systems (Woodward 1998, Nicholson et al. 1994; Herrero et al. 1999).

4.4 Modeling pests and diseases of crops and livestock

Biologists have been building mathematical models to describe the population dynamics of agricultural weeds, pests and diseases for over a hundred years. The diversity of modeling approaches that constitute the current state of science can be categorized in different ways. The first and most obvious is by production type and threat. Thus there are models that describe the dynamics of weeds, diseases and pests that are threats to arable crops, the diseases of livestock, and the diseases of fish used in aquaculture. While threats such as pests and diseases have been recognized since pre-history, the complexities of the microbial communities on the crop surface and in the soil around plants, and in the gut and rumen, are only just becoming fully understood. Models of the mixture of beneficial, commensal (those that neither benefit nor harm their hosts) and pathogenic organisms these communities contain have not yet been developed.

A broad distinction can be made between mechanistic (or process-based) and non-process-based pest and disease models. The former include explicit biology while the latter use a purely statistical approach.

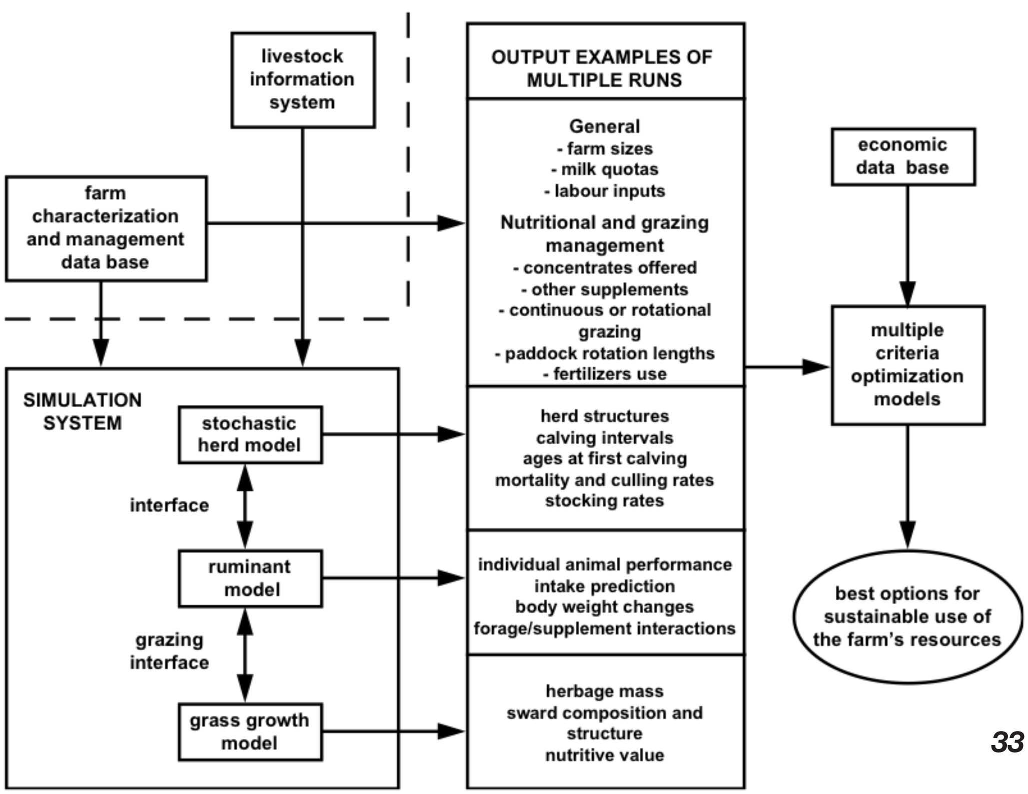

Figure 5. *Integrated livestock modeling framework (adapted from Herrero et al 1996).*

For example, a farm manager may want to know when to apply a prophylactic insecticide against a common insect pest. A statistical model containing biological variables such as crop stage, and weather variables such as temperature and rainfall, may best predict future insect population density. Future insect population density may be best predicted by a statistical model containing independent biological variables such as crop stage, and dependent weather variables such as temperature and rainfall. In some cases, information about the pest itself may be included in the model, for example from pheromone or other traps monitored by the farmer or in the case of mobile insects from publicly-operated monitoring networks. A different statistical application is the use of climate-matching models to predict future pest problems. The current distribution of an organism is modeled using a set of predictors including climate. The distribution of the organism after climate change is then estimated by mapping the "climate envelope" using scenarios developed from global climate models. There is now broad literature on the strengths and weaknesses of this approach, particularly challenging the assumption that organisms are able to move easily to track climate. Important recent advances in statistical models of pest dynamics have included the application of modern spline and neural net estimation techniques, and in the use of personal computers and increasingly mobile devices.

Mechanistic models incorporate at least some information about the biology of the crop and pest species concerned. The models may be highly abstract summarizing, for example, a pest population by a single state variable such as density, or highly complex with individual pests (or crop plants, or farm animals) each represented by numerous attributes. The simplest models sacrifice realism to give mathematical tractability and general insights, while models of intermediate complexity include more biological detail but are constructed in such a way that simpler analytical models can be recovered as limiting cases in order to help interpretation. Pest and disease models also vary in the degree to which they explicitly incorporate stochastic processes (often critical in epidemiological models) and in whether they treat a population as homogeneous or spatially variable.

4.4.1 Near-future pest and disease threats

Mechanistic models can be used to predict near-future pest and disease threats in ways that are similar to the statistical models discussed above. As was discussed with crop models, they may be more successful than statistical models if biological insights can substitute for missing data or if they can aid prediction by suggesting a model structure that simple statistical fitting would miss. Consider, e.g., the response of an insect to daily temperature. Higher temperatures may elevate growth leading to more pests, a pattern that could be derived with sufficient weather and population data. Alternatively, the physiological response of the insect could be modeled, which might improve the model's predictive power or allow insect dynamics to be predicted in data-poor systems (or under future climate). Several schools of physiological modeling exist and have produced parameterized models of insect pests, but we are aware of no formal comparison of different process and statistical approaches to the same problem.

An area where biological insights have proven particular fruitful has been in disease spread through commercial livestock populations. An understanding of how animals interact, but more importantly how animals are moved around, can provide critical advice to policy-makers. Current state-of-the-art livestock models incorporate data on the movements of animals between individual farms coupled with modern Bayesian parameter estimation. However, the type of data needed for such approaches is prohibitively expensive to obtain or politically unacceptable for many countries to collect (Brooks-Pollock et al. 2014).

Modeling has also proved valuable in assessing possible pest risks and in guiding general policy development. The basic epidemiological number (R_0) is the number of secondary cases of a disease that are expected to happen when a primary case occurs in a susceptible population. Calculation of R_0 for prevalent human diseases has proved useful in prioritizing investment in control strategies and vaccine development. Today, sophisticated mathematical tools are available for calculating R_0 for complex structured populations, for spatially extended populations, and in the presence of stochastic effects (Diekman and Heesterbeck 2012).

The discussion so far has explored only weed, pest and disease models where the state variables are population densities. Models have also been constructed to explore evolutionary and economic processes. Evolutionary models can be broadly categorized as genetic or phenotypic, the former explicitly simulates genetic processes while the later only simulates trait dynamics. Phenotypic models have been explored to a certain extent in agriculture (for example, under the rubric of Darwinian agriculture (Denisen 2012) but the vast majority of models have been genetic. Probably the most sophisticated areas of population genetics applied to weed, pest and disease issues in agriculture are models of the evolution of resistance to pesticides, and of the dynamics of plant diseases. Based on theoretical analyses, areas of fields have been set aside unsprayed or not planted with modified crops that express an insecticide in order to slow the rate of spread of resistance (Bates et al. 2005). The genetic basis of plant-pathogen interactions have been resolved for a number of major systems, which has allowed detailed analysis of strain dynamics and how disease spread may be slowed by judicious use of a range of different crop varieties. State-of-the-art in genetic models of weeds, pests and diseases includes utilizing the avalanche of data that modern high-throughput DNA measurement technologies are providing, and models of how novel genetic interventions may be used to suppress pest populations. Some of the most sophisticated pest monitoring software (typically based on statistical rather than on process models) now includes specific economic variables with parameters such as commodity prices that can be updated dynamically. The farmer may make different decisions about pest management depending on current market conditions.

More generally, a goal of many people working to increase the sustainability of agriculture is to reduce chemical inputs by practicing “integrated pest (or disease) management”. Such models are challenging to construct but some of the most advanced incorporate economic elements as well as various biological processes.

4.5 Economic Models

A number of approaches have been developed to model the economic implications of decisions and policies for a range of scales and purposes. Here, we summarize the most important approaches that have been used, most of which are still in use today, and indicate the purposes for which they are typically used. Also, important limitations of each approach are presented.

Farm management linear programming models. Linear economic optimization models of farm systems, developed in the 1950-60s, provide a basis for prescriptive farm management advice (see Heady and Dillon 1964). These models are characterized by a complex set of linear inequality constraints that represent the production possibilities available to a farmer. The simplex optimization algorithm is used to select the optimum production possibilities. One disadvantage of this approach is that the solutions are restricted to extreme points in the multidimensional decision variable space and thus it is unable to explore intermediate solutions. A major problem with these linear programming models is that they need complex constraint structures to achieve some degree of calibration to base data, but those constraint structures restrict alternative solutions.

Econometric production models. Econometric methods have been developed and used for single crop production function models as well as single-equation and simultaneous system models that represent input demand and output supply behavior. Early work focused on primal representations (Mundlak 1961), but much effort shifted to dual representations in the 1970s and later (Lau and Yotopolous 1971). Both static and dynamic models have been developed. Single crop production functions are estimated directly from data on the physical quantities of inputs and outputs observed from experimental plots, or, in later stages, from comprehensive farm production surveys. E.O Heady (1957) was an early proponent and researcher in this area. In many cases the functional form for the production functions is a quadratic or Cobb Douglas specification, both of which have implicit restrictive assumptions on the production technology. Later work emphasized various more flexible technology representations (Carter 1984).

Econometric estimation of agricultural systems was expanded to represent both multi-crop production with its associated interdependencies, the endogenous nature of agricultural supply response, and the imputed value of some key agricultural inputs that are often incompletely priced. A landmark article in this literature is the paper by Just et al. (1983). They noted that multi-crop farm businesses responded to changes in prices or technology by adjusting both the intensity of input use per acre, termed the intensive margin (i.e., fertilizer amount per land area); and also the allocation of land to crops termed the extensive margin. This distinction is important for modeling optimal input allocation in multi-crop farming systems. The importance of the interaction of multi-crops in a farm unit was a significant step forward in realistic economic models of farming systems. However, formal linkage to biophysical models of agricultural processes was not included in the approach. The econometric approach has limitations in its ability to extrapolate responses that are outside the estimation sample, or those that employ systems that are not present in the data sample. These limitations were emphasized by Antle and Capalbo (2001) in their development of economic simulation models that combine econometric and other disciplinary simulation models into an integrated assessment framework.

Risk behavior models. The importance of risk on farm decisions was recognized early in the development of linear optimization models of farming systems. Early articles on this linear approach to risk analysis are by Lin et al. (1974) and Hazell and Scandizzo (1975). As improved algorithms to solve quadratic optimization problems were developed, specification of risk in optimization models of farm systems expanded to a mean-variance measure of risk and imputed a risk-aversion value based on observed farmer actions or primary surveys. The book by Hazell and Norton (1986) shows the initial development of this approach. Just and Pope (1978) introduced a widely-used econometric risk model. Antle (1983) introduced a general moment-based representation of output distributions that has been widely used to study production risk behavior, including downside risk. Recent research has extended this approach to investigate impacts of climate change (Tack et al. 2012).

Spatial equilibrium models. The importance of space in agricultural production and modeling agricultural systems was first introduced in terms of trade between regions of different comparative advantage. Takayama and Judge (1964) showed that spatial equilibrium conditions and transport cost between different production locations could be characterized as a quadratic optimization problem. Spatial econometrics advanced to include rates of development and specialization of production (Anselin 1988). Only recently has the availability of remotely sensed measures of agricultural land and water use led to the use of spatial econometrics methods to address spatially varying farm production. (Anselin et al. 2004 and Staal et al. 2002). Techniques are emerging that use both remotely sensed data and spatial econometrics to draw conclusions about resource use or the effect of spatial variation on agricultural supply response.

Structural simulation models. Complex simulation models have been used for the past 45 years to describe dynamic agricultural systems. Early examples were often based on Forrester's (1968) concept of system dynamics that uses storage and flow variables to describe the system. However the underlying philosophy that a comprehensive and complex feedback system is, in of itself, stable and reproducible has never been convincingly demonstrated. Structural simulation models can be very useful for representing a combination of consistent behavioral relationships (i.e., that the quantities of product supplied by farmers is able to be sold at a price that recovers the costs of the inputs used to produce) based on theory and empirical measurement. They are however, subject to interpretation by the researcher in the absence of robustly estimated relationships describing the dynamic behavior of the system.

More recently, agent-based modeling (Billari et al. 2006, Troost and Berger 2014) has been widely used as a way of modeling interactive human behavior and natural systems. Some agent-based models have a more formal dynamic and calibration structure and use mixed-integer optimization approaches for solutions. However, the generality of the approach makes it susceptible to the same difficulties of empirical verification and reproducibility that earlier complex structural simulation models had.

Calibrating optimization models. Along with more complex constrained models, researchers have developed optimization models that utilize the shadow values on resource and calibration constraints to derive nonlinear calibrating functions, which are termed positive mathematical programming (PMP) (Howitt, 1995). In the past 10 years PMP has developed from formal calibration methods that reproduce the observed cropping pattern, termed first-order calibration, to approaches that calibrate crop supplies to prior estimates of supply elasticities (second-order calibration), and more complex production functions that calibrate against elasticities of substitution and returns to scale. In addition, PMP models are now being formally linked with biophysical models. A comprehensive survey of the more recent developments in optimization of economic models is presented in Merel and Howitt (2014).

Computable general equilibrium models. These macroeconomic models (e.g., computable general equilibrium, or CGE models) spawned a series of smaller-scale models which are usually called village or household models. General equilibrium village models account for all the flows in the village economy and remittances within the village to different workers and landowners. In addition, they include the flows of revenue in and out of the village boundary. This is particularly useful in developing country farm economies where much of the labor is supplied by family members with little or no pay. Another advantage of village-level equilibrium models is that they account for the utility gained from subsistence food grown in the village. These equilibrium models are anchored by a social accounting matrix that accounts for flows within and outside the economy. Moreover, it is common practice to fit the standard functional form such as a constant elasticity of substitution production, supply, or transformation function that is calibrated against exogenously estimated elasticities (see Taylor and Adelman 2006). CGE models have the disadvantage of being data and computationally intensive due to their more general specification. Compared with the more detailed partial equilibrium models above, general equilibrium models are harder to interact with process models, as they are less detailed in their production specification.

Dynamic optimization models. Some forms of agricultural production are anchored by a stock of resources that may be renewable or depletable, and whose dynamic relationships determine production over time. In this case, farm system models relying on this type of resource have to be analyzed in a formal dynamic economic context. The stocks of resources are the state variables for the problem, and the dynamic equations for the state variables are the biophysical equations that determine the changes in state variables over time. Solution approaches to dynamic economic problems are often categorized either into dynamic programming problems or optimal control problems. Solutions to problems formulated in this way are based on developments principle of optimality and the recursive Bellman (1957) equation. Dynamic programming, as are all formal dynamic methods, is plagued by what Bellman termed the curse of dimensionality, which restricts the number of state variables and thus the realism of the dynamic models. This problem of dimensionality is somewhat relaxed by using an iterative approach to solve for the optimal long-run equilibrium solution. Publications on these methods can be found in Judd (1998) or Miranda and Fackler (2002).

The optimal control specification of dynamic economic problems is a broader definition than dynamic programming and is most often associated with analytical solutions to relatively small theoretical problems where the optimal policy can be found by analytical solution of the differential equations that characterize the dynamic economic and biophysical processes. While these optimal control solutions are widely found throughout the resource economics literature, they are rarely applied to farm system models due to the problems of functional complexity and dimensionality that tend to dictate a numerical solution. Solution approaches to optimal control problems in the 1970s were often variants around the familiar linear quadratic Gaussian specification.

Agricultural system models. Flichman (2012) contains a number of recent studies on the application of models that combine bio-physical and economic models to represent agricultural systems. Flichman and Allen (2012) and van Wijk (2014) have also published surveys of economic agricultural system models. They characterize the scale of bio-economic models into farm models, landscape models, regional, and national models. Systems in each of these scales have components that include crops, livestock, and socioeconomics that interact with one another in complex ways. For example, the diagram of a farming system shown in Figure 6 shows important components that need to be included in models at that scale. It shows the crop and livestock production enterprises of a farm, the household decision and production processes, and the interactions among the household and production systems of the farm.

Within these scale levels they address both static and dynamic specifications. In his introduction Flichman attributes the growth of bio-economic modeling to two developments: The first development being the improvement of biophysical agricultural simulation models, and the second being the evolution of agricultural policies that demand integrated assessments that conventional economic models do not provide. We will briefly address three areas of application of integrated bio-economic models that are prominent in the literature.

i) *Climate Change Impact Assessment Models.* Economic modeling approaches to the impact of climate change on agricultural systems have been addressed in two very different ways. The first method links optimization models of agricultural production to climate models using agronomic models which map climate change variables into crop yields effects. The best-known example of this type modeling on a national basis for the United States is by Adams et al (1990) and the results from this type of model were widely used in the 1990 IPCC report. Since then, this same approach has been used in different types of optimization models over smaller geographic areas and driven by downscaled climate data. One advantage of using formal economic models to estimate the

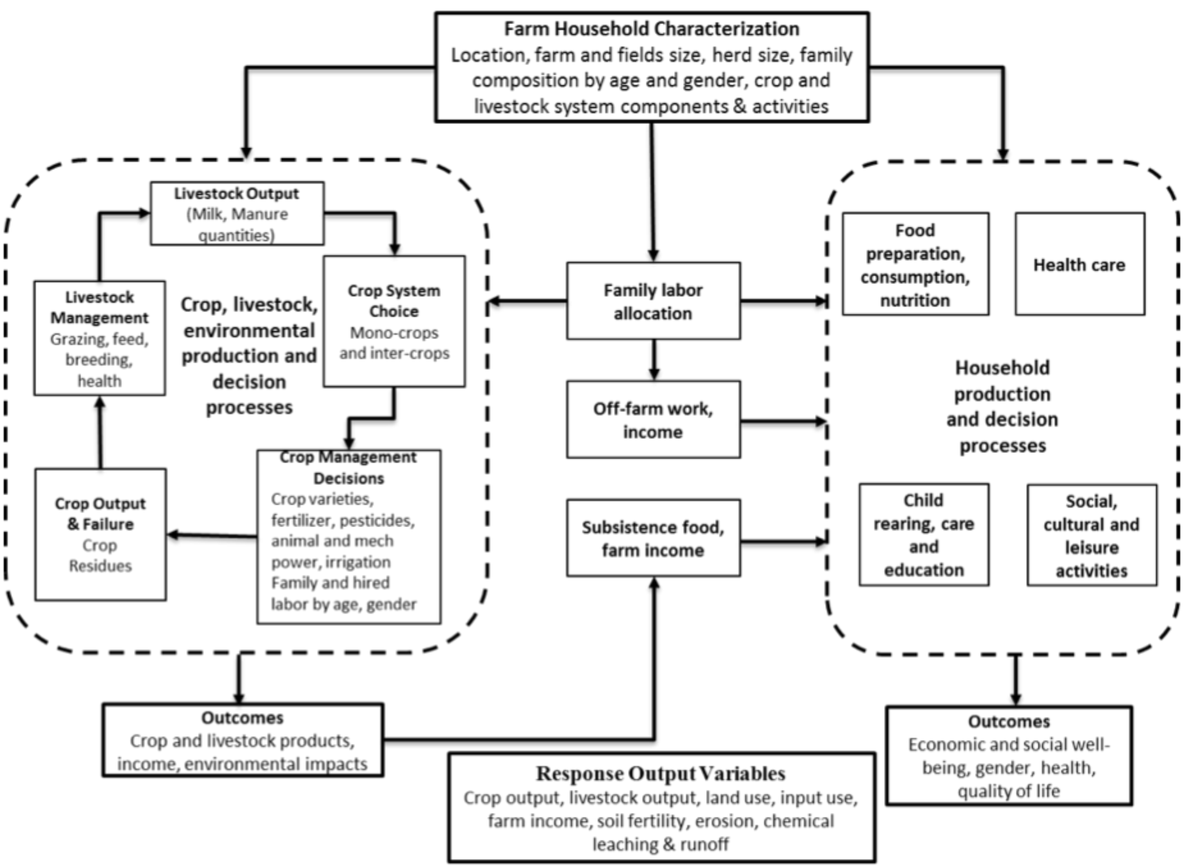

Figure 6. *Diagram of a farming system showing the household and production system components and interactions that need to be included in models.*

impacts of climate change is that they are able to incorporate the effects of both adaptation and mitigation that results from economically optimal adjustments to agricultural production under both changed growing conditions and a carbon tax.

An alternative econometric approach to measuring the impact of climate change on agricultural crop yields as well as economic variables such as land values and economic returns is to estimate statistical models based on observed behavior. These statistical models are then simulated with data from future climate projections. A justification for this approach is that it embeds realistic adaptive behavior into the model (Mendelsohn et al.1994). However, this type of model also has significant weaknesses, for example, it does not incorporate the effects of CO_2 fertilization on crop productivity, and cannot be used to identify the type of adaptive behavior or technological adaptations. Various researchers have used econometric methods to model the effects of climate on yields and other variables (Schlenker et al. 2013).

ii) *Hydro-Agricultural Economic System Models.* There is a long history of applying hydro-economics of agricultural systems, since irrigated agriculture is the largest user of water in many parts of the world. The tradition of integrating hydrology models with economic models stretches back 35 years since it was recognized early that the motivation for water use is strictly economic in agriculture, but that the equations associated with water use had to be modeled by physical hydrologic models. Accordingly, hydro-economic models were developed as coupled individual modeling systems. Two approaches are used. In the first approach the economic models provide benefit-response functions, which are then embedded in a hydrologic policy model. This approach is found in the Calvin water allocation model (e.g., Jenkins et al. 2004). An alternative hydro-economic modeling approach is to characterize the response in the hydrologic models by statistically fitting a more simplified function to results from complex simulations over a range of hydrologic and climatological parameters. These response functions can then be included in an economic policy model. This approach has often been used to analyze the optimal management of common property resources used in agriculture, such as groundwater (see Knapp et al., 1996 for an example). A review of concepts and applications of hydro-economic models from an economic perspective can be found in Booker et al. (2012), while Harou et al. (2009) published a similar survey from a hydrologic perspective.

iii) *Integrated Economic Livestock Models.* These models usually fall into one of two combinations; namely biological process models with an economic component, or an economic model with livestock equations and response functions. Several models in developing countries expand their specification to take into account household linkages and village-level interactions where there is some degree of subsistence consumption of livestock products, most commonly milk. A recent publication provides a good overview of integrated livestock modeling and its effects in mitigating climate change (Havlik et al. 2014). Their analysis is driven by a large-scale economic optimization model that assesses crop bioenergy production, land-use changes, water requirements, and greenhouse gas emissions. Their results show that improvements in the livestock production system can significantly reduce impacts on fragile land use, as well as improve the effectiveness of climate mitigation policies. In another approach, Kobayashi (2007) analyzed stocking density impacts on Kazakhstan's extensive rangelands by estimating a stochastic dynamic programming model for multiple livestock systems with stochastic forage production. This model showed that cost of capital strongly affects flock size and productivity.

4.6 Landscape/Watershed; Water and Environmental Quality

There are at least two different perspectives about modeling across space, including the interconnectedness of agricultural and ecological systems across the landscape. The first perspective is that human systems, including the farm, communities, and administrative and political areas in which agricultural systems interact through decisions and policies, affect production systems, markets, and trade. The other perspective is that the interconnectedness among hydrological and biophysical processes establishes the underlying behavior of agricultural systems over the landscape, with a particular emphasis on understanding physical, chemical, and biological processes that occur in watersheds. Both are important, yet agricultural models rarely consider both in the same assessments.

Figure 7 shows the regional integrated assessment approach developed by AgMIP that emphasizes linkages of agricultural systems across space (A. farm household; B. heterogeneous farms in one or more communities across the landscape; C. heterogeneity of administrative districts; and D. national/global scale). In this perspective, based in part on the impact assessment approach developed by Antle (2011), the focus is on the economic, environmental, and social impacts of alternative systems within heterogeneous household populations. However, this framework also illustrates the feedbacks from farms to agro-ecological regions to national and global scales.

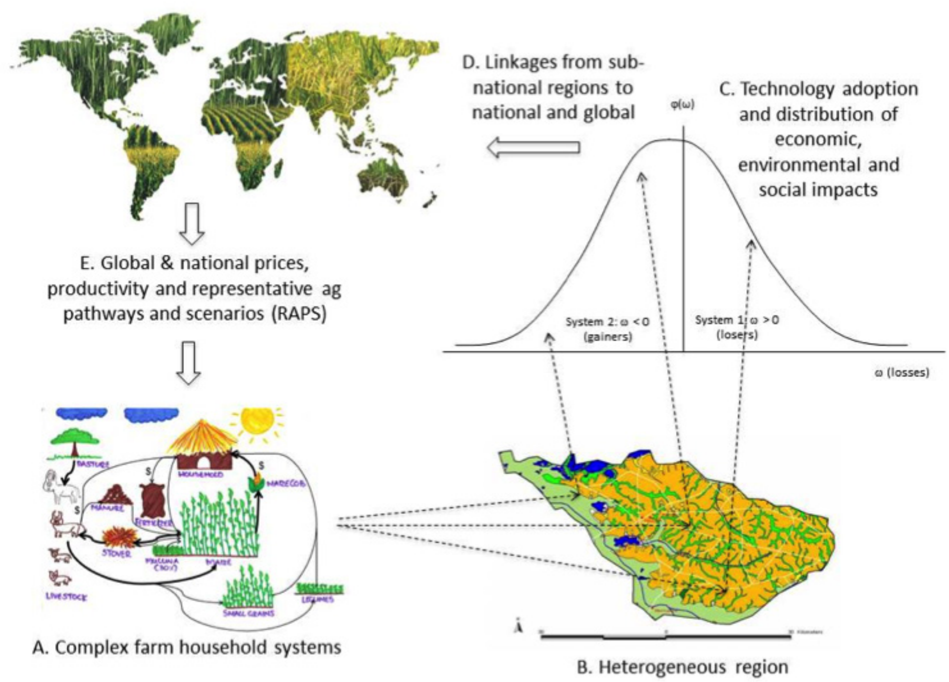

Figure 7. *The AgMIP Regional Integrated Assessment framework emphasizing linkages across scales and analysis of distributional impacts in heterogeneous populations of farm households (Antle et al. 2015, this volume).*

We often use the term "scaling up" of model results to refer to the aggregation of model results from finer spatial resolutions to a larger area. If the areas of interest are defined by hydrologists, they tend to be watersheds. In contrast, if the areas are defined by economists, they tend to be administrative and political units (e.g., urban areas, districts, countries). These two perspectives are not mutually exclusive, however. In fact, they lend themselves to include both human and biophysical/hydrological processes. The challenge for next generation agricultural models includes the technical aspects of integrated modeling and a transdisciplinary approach in which scientists recognize the need for collaboration, not only on specific projects, but also in designing models and decision support tools to achieve their goals.

Many current agricultural system models have been developed to evaluate practices and policies associated with environmental quality. Biophysical models (e.g., crop or nutrient models) typically operate at the point/field scales (Figure 8a) with an emphasis on vertical fluxes of energy, water, C, N and nutrients throughout the atmosphere, plant, and soil root-zone continuum. Upscaling from point to the landscape scale (Figure 8b) requires estimation of surface and subsurface fluxes and ecological transitions along the lateral scale. Coupling with landscape microclimate models provides not only the vertical inputs used by agricultural models, but also gradients (precipitation, temperature, wind, vapor pressure deficit) along the landscape. Coupling with hydrological models provides water flow paths such as surface runoff, vertical and lateral groundwater flow, and interactions between shallow soil and groundwater zones and with adjacent surface water bodies (channels, rivers, lakes and coastal waters). Water quality models provide sediment and solute transport along the landscape controlled by water flows (Figure 8b), and other effects such as wind erosion.

Integration and upscaling of landscapes into the watershed scale (Figure 8c) requires 3-dimensional coupling of the surface and subsurface water, energy and mass transfers (see more details on coupled versus integrated models in Condon and Maxwell, 2013, and in Maxwell et al. 2014). At this scale, the groundwater aquifer system typically transcends the boundaries of the watershed and necessitates analysis at a regional scale to evaluate not only the impacts of cropping and animal production systems on water quantity and quality, but also feedbacks from the hydrological system into the agricultural system (shallow water table effects, drought or low water availability for irrigation) (Muñoz-Carpena et al. 2006). Further, meso-scale rainfall and evapotranspiration distribution models control the local surface and subsurface flow intensities, pollution and abatement (Shrivastava 2014). At this scale, human effects through land use changes, as well as ecological (vegetation, wildlife) dynamics and transitions on natural or protected lands (riparian zones, conservation areas, etc.) are also important components needed to evaluate the overall sustainability of agricultural systems (Matson et al. 1997). Although some efforts have gone into integrating biophysical models (e.g., crop, hydrology, livestock, ecological, and economic), more is needed to enable more comprehensive assessments of agricultural systems across scales and adequately address environmental and economic responses to decisions and policies.

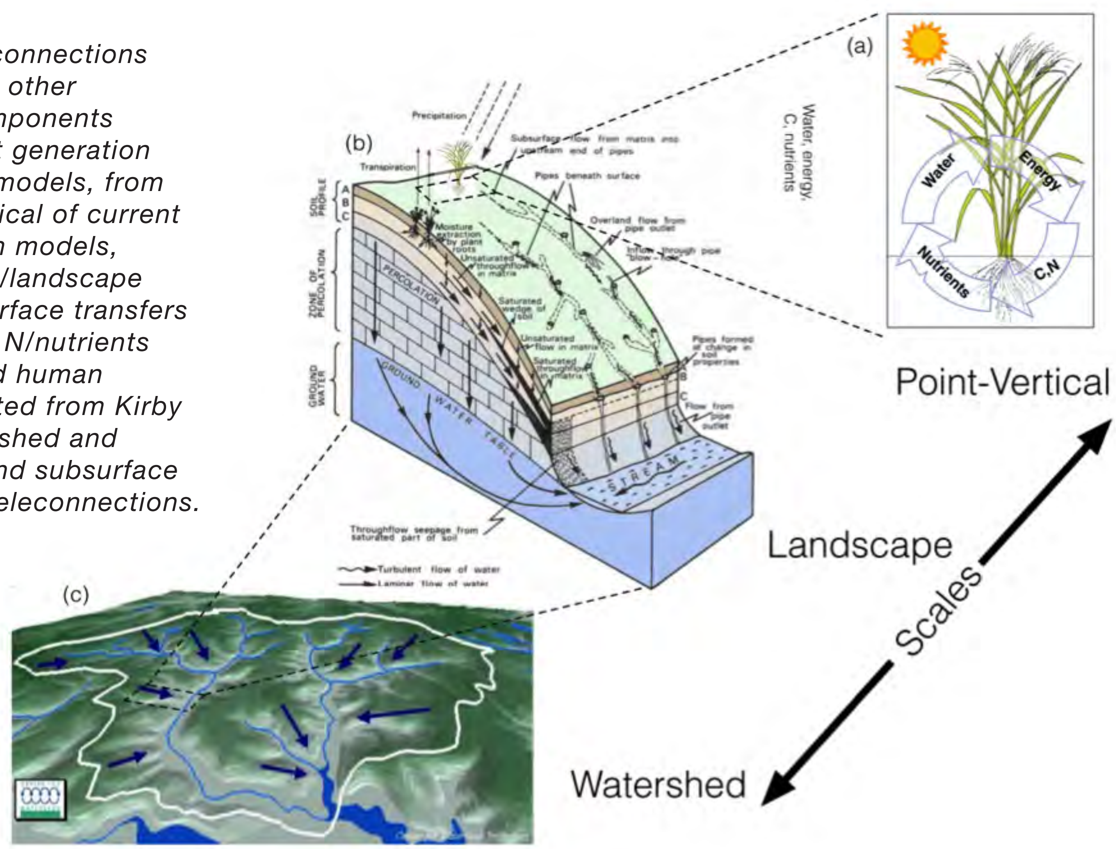

Figure 8. *Lateral connections across scales with other environmental components needed in the next generation agro-ecosystems models, from (a) pointscales typical of current agricultural system models, (b) lateral hillslope/landscape surface and subsurface transfers of energy/water/C,N/nutrients and ecological and human interactions (adapted from Kirby 1978), to (c) watershed and regional surface and subsurface connections and teleconnections.*

4.6 Aggregate agricultural system models (district, country)

In the last section, we discussed efforts where models developed separately are sometimes linked or combined to scale up from points or fields to watersheds and larger scales. Generally, there are many agricultural system models that address decisions and policies at scales where there may be many individual production systems, each with its own crop and livestock enterprises. Resolving the time and space scale differences among components of system models is a major issue brought about by independent development of component models for different purposes. This is a general problem that one encounters when attempting to create a model that combines crop and hydrology models crop and economic models or crop and climate models (Osborne et al. 2007; Elliott et al. 2014). There have also been efforts in which dynamic models have been developed to provide forecasts over aggregated areas (e.g., to provide aggregate crop forecasts) starting in the early 2000s. Traditionally, climate model output for a grid cell is downscaled to produce weather data time series for points that are then fed into crop models. However, the land surface also influences climate; processes within the atmosphere, oceans and on the land are coupled and dynamically interact over space on timescales from fractions of seconds to thousands of years. Crops are a major component of the land surface of the globe, occupying about a quarter of the land area. Regional climate can be sensitive to large-scale changes in cropped areas that can result from changes in economic or climate conditions (Osborne et al 2004). Therefore, another direction for agricultural impacts assessments at a large-scale is to dynamically couple the crop simulation with models of land and atmosphere processes.

Five research groups succeeded in coupling aggregate crop into climate models (Bondeau et al 2007; Gervois et al 2004; Kucharik 2003; Osborne et al 2007 and Stehfest et al 2007). One of these has shown that in some parts of the world the impact of changes in cropped area on regional surface temperature can be of the same magnitude as regional human-induced climate change (Osborne et al 2009). This raises the question of whether or not new fully-coupled climate change impacts studies will completely revise our previous estimates of food security impacts. It is clear that the full coupling of crop simulation within global climate models is opening up new possibilities for studies of the impact of climate change on agricultural production that capture some of the complex and important feedbacks within the Earth system at a large scale.

Restriction of the skill of large-area modeling of crop production and yield is dominated by the density of data used in the simulation. More data should equate to better skill. However, the skill of large-scale modeling is determined by the smallest data set, whether this is the grid cell with the shortest run of observed yields, or the data grid with the largest resolution (climate, crop, soils or otherwise). We have seen recent increases in the resolution of climate input data and global grids of crop management and soil information. In this field of agricultural modeling, any future increase in data resolution across the range of data needed for modeling should produce a more skillful model simulation.

5. Current Agricultural System Models in Context of Selected Use Cases

Here, we discuss the state of current agricultural system science relative to its use in providing information to assist a wide range of decision makers represented by five Use Cases, which were described in an earlier paper in this series. The paper by Antle et al. (introduction paper in this series) indicated that these Use Cases need crop, livestock, and farming system models. The question that we address here is whether current agricultural system models, existing sources of data, and existing decision support tools are adequate for providing information needed for these Use Cases. Here we summarize current capabilities and limitations of existing models and DSSs relative to those needs.

5.1 Farm Extension in Africa

In this Use Case, the user is Jan, an extension officer who is providing advice to a small farmer in Southern Africa to help her and her family produce more food and income on a consistent basis.

Capabilities and Limitations

<u>*Models.*</u> Can existing crop, livestock, and farming system models, data, and ICT tools provide the information that Jan needs to advise the small farmer? The short answer is "No", but there are models that partially meet these needs. There are no Apps that currently summarize output results from models that are appropriate for this particular farmer or to connect with models in the "cloud" to make runs as requested by an App that Jan activates. However, cropping and farming system models can analyze some of these changes that Jan needs to evaluate for the farmer. Current models can simulate responses of crops to water and N fertilizer input, but most crop models do not include pest and disease components. Also, simulation of new or future varieties with drought and high temperature tolerance are being conducted for a few crops (Singh et al. 2012, 2013, 2014), although data for testing simulations are still lacking in most cases. And, although models can simulate crop growth and yield under various soil and management conditions, very little has been done to simulate the effects of rainfall harvesting on productivity, even though field experiments have shown considerable increases in some situations (e.g., Fatondji et al. 2010).

One of the most serious limitations of existing cropping system models is their inabilities to include losses associated with the huge range of pest, disease, and weed species that damage crops. In many intensive production systems, weeds, pests, and diseases are controlled so that responses in those areas can be represented by the costs of control and the production responses to climate, water, and nutrient management. Typically, cropping system models simulate yields that are higher than actual yields in farmers' fields, and much of this is due to damage caused by weeds, pests and diseases. As indicated, most crop models can operate at the water-limited and nitrogen-limited production situation (Figure 3), but do not simulate actual yield in production situations where these factors are not controlled. In addition, farmer's fields sometimes are not homogeneous; for example, spacing between plants may vary considerably, even though the models assume homogeneity. However, if pest and disease data are observed and available, these can be input to some existing crop models to compute yield loss associated with specific pests and to diagnose the reasons for the gap between potential and actual yield (including the gap associated with water and some nutrients, especially N and P; e.g., Boote et al. 1983; Naab et al. 2004).

Generally, farming system models currently in use have capabilities needed to analyze this use-case. However, most of the farming system models are not developed to be easily implemented in farms where they were not developed. (An exception to this is the TOA-MD farming system model that is set up for a population of farms in contrast to a particular farm). And most do not consider risks associated with weather variability and extremes, the variations in management that exist within and across farms, or all of the interactions among enterprises.

<u>*Input Data.*</u> Data are needed to describe a range of farming systems so that Mark could select the combination of biophysical, farming system, and household characteristics from data that have been collected and made available. This would include information to allow Mark to tailor the inputs to most closely match the conditions of specific farms, which would also need to be made available to him as he makes use of the NextGen models and DSS tools. This includes climate, soil, management practice, labor and other inputs available for production and marketing of outputs, typical pest and disease pressures, availability and prices for farming inputs, and other farm, economic, and environmental information. This could also benefit from use of reduced form summary models to facilitate quick access to information needed for specific questions in target farming communities. Generally, sufficient data on the biophysical, environmental, and socio-economic conditions of each farm or for a range of farm typologies in the regions are not available. Although some data, such as climate and soil data, are available from databases that are georeferenced to provide some of the minimum data needed, these are generally not organized such that agricultural systems models can readily access them for analysis of specific farms. In other cases, such as information on the weed, pest and disease pressures and on crop varieties' and livestock breeds' tolerance levels, is not available. Although there has been research that shows that some analyses need-

ed to advise the farmer can be done, the availability of accurate input data for agricultural systems models is a major limitation at the current time. Furthermore, there are studies that collect much of the data needed for analyses of some of the issues that the smallholder farmer needs, including the integrated assessments of climate change impacts and adaptation in AgMIP and other projects.

Decision Support Tools. Most existing DSS tools that are available in Apps are focused on relatively narrow issues, such as when to apply a fungicide to a particular crop (e.g., see www.agroclimate.org), when to apply the next irrigation, or how much N fertilizer to apply to a particular crop that will be grown on a particular type of soil in a specific setting. The authors are not aware of DSS tools that make use of more integrated models that help farmers decide on overall farming system design. Here, we envision a DSS platform that will connect various models, databases, analysis tools, and information synthesis tools to provide an easy-to-use interface for Mark to set up the analyses and to select outputs that provide information to answer questions he has concerning the management of his farming system that considers specifics of that particular farms' biophysical and socioeconomic situation. In addition, some information would include an estimate of the confidence DSS developers have in any recommendation given to users. Such DSS platforms are possible, but not yet constructed.

Additional development in the models, databases, and DSS tools are needed to realize this potential, for cropping system and farming system models. In particular, the Apps and associated databases needed for Mark to make use of the models and farm information are mostly missing currently. It is not practical for Mark to collect the information on a particular farm, go back to his office and work with an analyst to evaluate options that the farmer is interested in.

5.2 Developing and Evaluating Improved Crop and Livestock Systems for Sustainable Intensification

Debora, a plant breeder who is developing a drought- and heat-tolerant hybrid of maize, would like to be able to evaluate the potential adoption and impacts of her new maize varieties across the widely varying conditions in Africa. She would like to evaluate the potential of new varieties in complex mixed crop-livestock farming systems relative to meeting sustainable intensification goals, such as improving productivity, taking into account long term impacts on soils, water and greenhouse gases.

5.2.1 Capabilities and Limitations

Models of maize and other crops, livestock, and the farm household are also needed for this Use Case. These models are available for at least partially performing this type of analysis. The Global Futures, Harvest Choice, and other projects being led by IFPRI have used crop and economic models to evaluate the potential benefits of developing new technologies, including new crop varieties (e.g., see Singh et al. 2012, 2013, 2014; Rosegrant et al. 2014). For example, Singh et al. (2012) used the DSSAT CROPGRO groundnut model with climate and soil condition inputs at six locations in India to evaluate different target traits that are being selected by plant breeders in the CGIAR. They found that the effect of combining various traits was beneficial, with estimated yield gains varying, depending on location and climate change conditions. Starting in the 1980s, several groups began using crop simulation models to evaluate alternative management systems in developing countries (Keating et al. 1991; Uehara and Tsuji 1998; Penning de Vries et al. 1991). The models used in those efforts were generally based on the CERES and other crop models in DSSAT and on the ORYZA rice model developed by the Dutch modeling group.

Similarly, considerable work has been done on use of farming system models to evaluate various options for improving the livelihoods of farmers. These include farm simulation models (e.g., Baudron et al. 2014), optimization models that attempt to select the best combination of enterprises and their management to achieve one or multiple goals of the farmer (usually, maximizing profit, for example, or maximizing utility taking into account attitudes toward risk; e.g., see Nicholson et al. 1994; Herrero et al. 1999; Castelan et al. 2003; Waithaka et al. 2006; Gonzalez-Estrada et al. 2008). Also, the Tradeoff Analysis (TOA) farming system model (Stoorvogel et al., 2004; Antle, 2011) is currently being used as the basis for model-based assessments in the AgMIP project (Rosenzweig et al., 2013) and other projects (e.g., see Claessens et al. 2009). Furthermore, this farming system approach can incorporate results from crop and livestock models, and it has been developed such that it can be adapted for smallholder, as well as large commercial farming systems.

However, there are important limitations in the capabilities of these models. In particular, the crop models mostly have capabilities for simulating productivity at the water and nitrogen limitation levels (van Ittersum et al. 2003). There may be a number of other yield limiting soil nutrients, soil physical constraints, and various pest, disease, weed, and other yield reducing factors that lead to actual yields in farmers' fields that are less than model prediction. This is particularly true in developing countries where agrochemicals are not widely used to prevent these factors from reducing yields. Thus, there are large yield gaps between actual yields in farmers' fields and the potential productivity that a particular variety of a crop is capable of producing in those fields (e.g. van Ittersum et al., 2013; Gustafson et al. 2014). Crop models are useful for predicting potential productivity as well as water-limited and nitrogen-limited yields of major food crops grown worldwide. Water or nitrogen or climate conditions are, of course, major factors that determine yields in developing and developed countries. When those are the major limitations, current crop models are highly useful, assuming that soil, weather, cultivar, and management conditions input data are available for the analyses. But, for this use case, it is likely that there are other factors, including other soil nutrients, pests, diseases, and weeds, that need to be taken into account. The challenge for next generation models includes not only modeling those factors but also collecting data that describe the production situation with all of the important yield limiting and reducing factors.

Another major question for this Use Case is whether existing biophysical models can predict performance of the wide range intensification options that may be used by farmers. Some intensification practices would include increased fertilizer, use of some agro-chemicals, modified tillage, increased plant population, soil additives, changing row geometries, and more precise timing and placement of fertilizers. They also could include water harvesting methods (in the field or on the farm), use of drip irrigation, and the use of mulches to reduce soil evaporation and erosion. Although some of these intensification options are in use, most models have not been developed or tested for the wide range of potential changes to production systems. Furthermore, most biophysical models lack components that compute sustainability metrics needed to assess sustainable intensification.

Livestock models also have various limitations that need to be addressed in the next generation of tools. While the understanding of animal feed requirements is relatively robust, there are still large errors associated with the prediction of feed intake in ruminants. Additionally, few of the models have the capabilities of predicting methane production as part of their outputs, and this is becoming more important as assessments of environmental impacts of livestock gain prominence. Gathering the much needed experimental information is required for validation purposes, especially in the tropics. Another area that merits more work is understanding how climate change is likely to affect livestock systems. We still cannot model the animal and herd responses to increased temperatures, climate variability, more severe feed fluctuations. Thus, we have limited ability in designing adaptation strategies in livestock systems. There are also few assessments of how feed quality and intake are likely to evolve as a result of climate change. The decision-making process of livestock farmers needs also to be better incorporated in whole farm models, with better rules for governing behavioral changes as systems intensify.

The demand for livestock products is relatively inelastic, which works in both directions in that a given reduction in the price of meat does not lead to the same proportional increase in demand. Similarily, if there is a substantial increase in the quantity of meat offered for sale, prices will fall by a larger proportion than the quantities offered. The same phenomenon tends to occur in the demand and supply for inputs for livestock production, in particular those deriving from rangeland. The net effect of these price responses to quantity changes is to magnify the effect of variations in the supply of livestock products. A classic example is the situation at the beginning of a drought when a shortage of pasture dry matter forces herdsmen to sell more than usual to reduce numbers. The inelastic demand for their product stimulates a disproportionate fall in prices, which further penalizes the herdsmen in terms of loss of income and access to money to rebuild their herds after the drought. Livestock system models need to integrate both the biophysical fluctuations in productivity and the likely economic responses to these fluctuations in order to get an accurate measure of the impact on the well-being of families who rely on livestock for a substantial part of their income.

There are also limitations in the socioeconomic models used for evaluating benefits and tradeoffs among different technologies and management of crops and livestock and in managing the farm and its resources. For example, most available economic models solve for or simulate average responses. In this use case, it is very important that uncertainty and economic risks be taken into account. Many of those risks are associated with variations in weather (particularly rainfall) from season to season and in the market place where farmers purchase supplies and sell their products.

Input Data. Some of the issues lie in the models themselves, but also in the development of comprehensive datasets for running the models. This is true for production models of crops and animals as well as economic models across the first three Use Cases that address issues in data-poor areas in Sub Saharan Africa. Thus, the same limitations discussed in Use Case 1 are also relevant to this Use Case. In addition to those data limitations, there are various types of input data needs for livestock that are difficult to obtain, including species composition in rangelands, diet selection by animals, better spatial representation of feeding practices, adequate parameterization of the feed quality parameters and how they change in space and time, improving productions systems descriptions, and others.

To operate models that assess the performance of sustainable intensification options, data are needed first of all to develop and evaluate those models. When models are used to address Deborah's needs, input data will be needed to characterize the intensification technologies for use in the biophysical models as well as to characterize the fields, farm, landscape, hydrology, and ecological components.

Decision Support Tools. Some progress has been made on information systems that allow one to compute sustainability metrics for specific farming practices. Much is being done by the private sector, and more public-private collaboration in defining and developing improved metrics of sustainability should be considered. However, little has been done to date to produce the types of decision information systems needed to help Deborah advise farmers on sustainable intensification options that are tailored for specific regions and farming systems.

5.3 Investment in Agricultural Development to Support Sustainable Intensification

Stanley is an investment manager for a prominent Foundation, and he needs to evaluate a project for small farms in Kenya that will increase the intensity of production by increasing fertilizer use per hectare on cash crops while maintaining the current sustainable nutrient balance between pasture grasses, crop residues and animal manure.

5.3.1 Capabilities and Limitations

While the capacity to solve each of these individual component problems has been shown in journal articles, and in some cases implemented in practice, their integration into a decision support system for landscape and regional decision has not yet been demonstrated. In particular, the feedback loop between the data sensing system and the crop and livestock models based on primary experimental data is an important and untested component for assessing the inevitable gap between experimental results and field implementation. Using remote sensing systems from satellites and/or drones, we envisage a much more rapid and readily quantified flow of data that will allow updating and assessment of the project as it proceeds. We see the principal limitations for both crop growth, livestock, and farming system models as being due to sparse data rather than gaps in the conceptual theory. It is the formal integration of the sample points using landscape scale GIS systems to provide the distribution of impacts, coupled with the rapid sequential and consistent updating of estimates of key parameters such as crop yield and water use that is an innovative approach incorporated in NexGen models. However, throughout the landscape scale analysis, the tradeoffs between the number and complexity of sampling points to represent the distribution of impacts over the landscape must be approached in a formal manner.

5.4 Management Support for Precision Agriculture in the US for Profitability, Soil Conservation, and Water Quality Protection

Greg is a farmer in the US, with a large corn/soybean-based operation and a high level of mechanization fully equipped with auto-tracking system and high-resolution differential GPS. He wants to manage his fields using precision management of input resources to increase efficiency and profits and to reduce environmental risks.

5.4.1. Capabilities and Limitations

Strategies to overcome spatial resolution in point-based crop models were first addressed by Basso et al. (2001) and Batchelor et al. (2002). Such strategies include running point-based models at small scales within a field; geospatial technologies (remote sensing, electrical resistivity tomography, yield mapping) to target the application of models to areas with similar plant responses; and linking point-based models to three-dimensional water flow models to better represent water transport across the landscape (Basso et al. 2001).

The application of point-based models on small homogenous areas within a field has had limited success due to difficulties in obtaining critical fine scale soil and management information (soil physical characteristics, including rooting depth, plant population and effective tile drain spacing) necessary to run the models. A current limitation in most crop models is the assumption of uniform plant distribution. Visual observations as well as measurements commonly indicate that crops are not uniformly distributed, and therefore assuming they can be an unrealistic assumption and a significant source of uncertainty in yield simulations (Ritchie and Basso 2008; Basso and Ritchie 2012). A correction procedure based on the extent of variation in plant stand uniformity or dominant plant density may be necessary. Correction also is required to compensate for yield loss from plants missing in a population; to some extent neighboring plants can compensate for missing plants as they have more space to intercept light.

Recent advances on the resolution and availability of remote sensing imagery (satellite, airborne, and Unmanned Arial Vehicles – UAVs) coupled with a decrease in their associated costs, allow for the collection of timely information on soil and crop variability by examining spatial and temporal patterns of vegetation indices (Ehmke 2013). Such information can be used to derive inputs for crop models in conjunction with yield mapping analysis to identify areas in the field that are stable over space and time. Crop models can be executed on those areas to provide insights on the reason of variability as well as estimates of potential economic return of variable-rate input prescriptions.

The assessment of spatial soil water availability is crucial for understanding the interaction of water stress and crop yield variability in agricultural fields, especially now with increased climate variability and extended drought periods. Spatial variation in soil water is often the cause of crop yield spatial variability due to its influence on the uniformity of the plant stand at emergence and in-season water stress. Soil water content is highly variable within a field due to spatial variation in rainfall, topography, soil properties, and vegetation. The ability to simulate spatial soil water content over time is important for models used for agricultural and hydrological systems assessment (i.e. nitrate and pesticide leaching to groundwater, erosion modeling, water logging, and Precision Agriculture applications).

Soil-plant-atmosphere models have proven to be effective in simulating the water balance of soils when the drainage is assumed to be vertical. However, this assumption is incorrect in many fields. For instance, runoff computed by one-dimensional models is not distributed over space, and thus results in inaccurate predictions of surface soil water balance in neighboring areas within a field. The automation of terrain analysis and the use of Digital Terrain Models (DTMs) have made it possible to quantify the topographic attributes of the landscape and to use topography as one of the major driving variables for many hydrological models (Western et al. 1999; Wilson and Gallant, 2000). Basso et al. (2001) developed a spatial soil water balance model that simulates three-dimensional surface and subsurface water flow. The model requires a digital elevation model for partitioning the landscape into a series of interconnected irregular elements, daily weather data, and spatial soil information for the soil water balance simulation. These aspects are considered a serious limitation in crop models and despite their importance have hitherto received limited attention, thus warranting additional improvements and testing.

An example that combines strategic and tactical application of a crop model in a spatial context is described by Basso et al. (2011). A dual-criteria optimization through a tested model could determine the N rate that minimized nitrate leaching and increased net revenues for the farmer for three zones within the same field characterized by different yield potentials (Figure 9).

5.5 Supplying Food Products that Meet Corporate Sustainability Goals

Jennifer, an economic analyst in a corporate sustainability group, embraces sustainability as the core of their mission: marketing food while conserving resources. She needs to help the corporation's contract farmers with decisions regarding when to plant, when to irrigate and when and how much fertilizer to apply to conserve energy, save water, minimize waste and reduce greenhouse gas emissions in an effort to make these products more sustainable from farm to fork.

5.5.1 Capabilities and Limitations

An example of the application of crop models to illustrate how reduced N fertilizer rates result in reductions of greenhouse gas emissions (expressed in CO_2equivalents) at the field scale are described in Basso et al. 2013 (see Figure 10).

The next-gen crop models with capability of using real time weather and historical climate conditions will be able to identify strategies that are able to optimize the amount of fertilizer used at a particular location, soil and weather conditions with the goal of increasing

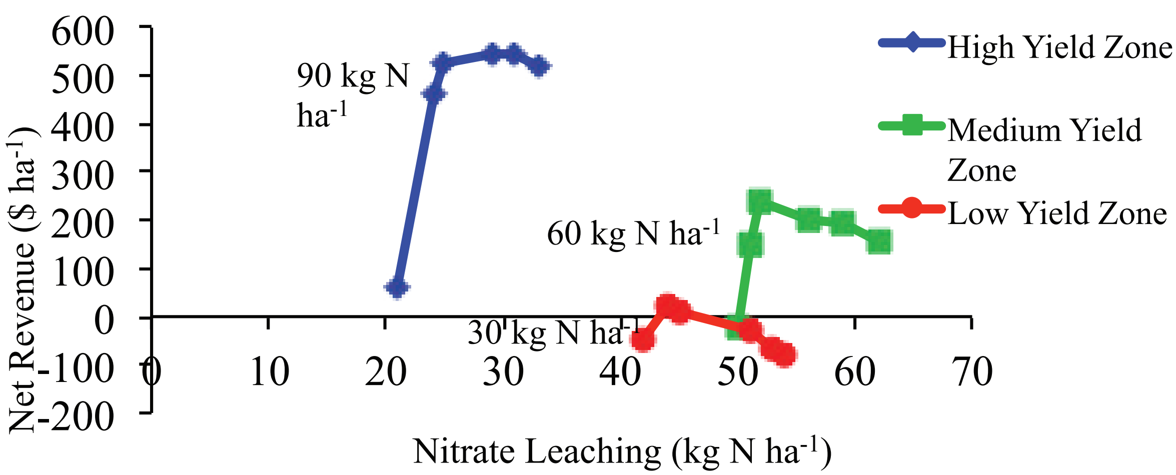

Figure 9. *Net revenue simulated for three different management zones showing the differences in nitrate leaching associated with each zone.*

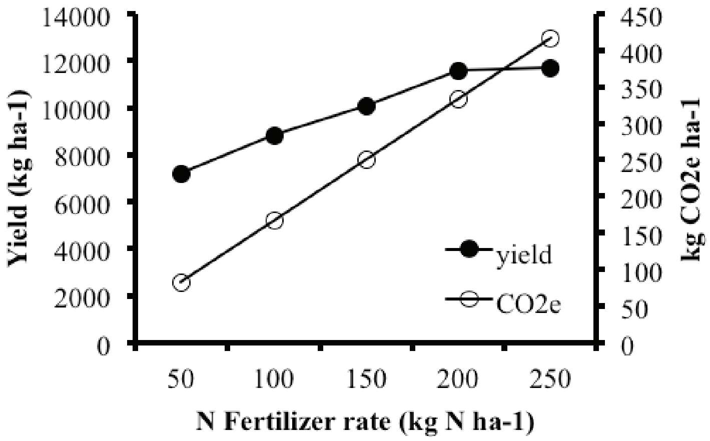

Figure 10. *Simulation of greenhouse gas emissions and maize yield for different N fertilizer application rates, showing tradeoffs between emissions and yield.*

yield and reducing greenhouse gas emissions. Crop models can evaluate the effects of unknown weather conditions and help decide the optimal nitrogen to apply to crops with different amounts within the field using precision agriculture prescription maps. Communication companies have partnered with different high-tech companies to deliver solutions for the meteorological, geo-spatial and operational challenges facing the agriculture industry. Remote monitoring solutions, as an integral part of the Next-Gen model platform, along with advanced cloud services, will help farmers with decisions regarding when to plant, when to irrigate and when and how much N fertilizer to apply.

Some of the large corporate supply chains companies have recently set a goal to improve fertilizer-application efficiency of U.S. row crop farmers in its food supply chain by 30% by 2020.

System models can further help these companies by setting emission reduction protocols, benchmarks and baselines to compare emissions between different management strategies, and by incorporating sustainable agricultural criteria into their future plans validating mechanism, including certification to verify that the farmers are meeting the sustainability criteria.

An important aspect to consider in simulating the crop N uptake and soil N balance is the initialization of the soil carbon pool in order to properly simulate the soil carbon and nitrogen dynamics as pointed out by Basso et al. (2011).

Discussion

The history of agricultural systems modeling shows that major contributions have been made by different disciplines, addressing different production systems from field to farms, landscapes, and beyond. In addition, there are excellent examples in which component models from different disciplines have been combined in different ways to produce more comprehensive system models that consider biophysical, socioeconomic, and environmental responses. For example, there are examples where crop, livestock, and economic models have been combined to study farming systems as well as to analyze national and global impacts of climate change, policies, or alternative technologies for different purposes. This history also shows that the development of agricultural system models is still evolving through efforts of an increasing number of research organizations worldwide and through various global efforts demonstrating that researchers in these groups are increasingly interested in contributing to communities of science (e.g., via the global AgMIP effort, various CGIAR–led programs, e.g., like the IFPRI-led Global Futures and Harvest Choice projects and the CIAT-led CCAFS project), the new CIMSANS Center being led by the Public-Private ILSI initiative, and the various global initiatives that aim to provide more harmonized and open databases for agriculture (such as GEOGLAM, GEOSHARE, and others). However, through the review of existing initiatives and discussions among the experts involved in this NextGen study, it is clear that there is a need for a more focused effort to connect these various agricultural systems modeling, database, harmonization and open-access data, and DSS efforts together, so that the scientific resources being invested in these different initiatives will contribute to compatible set of models, data, and platforms to ensure global public goods. This is critically important, considering that these tools are increasingly needed to ensure that agriculture will meet the food demands of the next 50 to 100 years and will be sustainable environmentally and economically.

These papers address several questions regarding the status of existing models, data, and DSS as we consider what is needed for NextGen agricultural system models. The Use Cases described in the introductory paper and discussed above demonstrate that a minimum set of component models are needed to develop agricultural system models that are more or less common across the Use Cases. These include crop models that combine weather, soil, genetic, and management components to simulate yield, resource use, and outputs of nutrients and chemicals to surrounding water, air, and ecological systems, taking into account weed, pest and disease pressures, and predict performance to a range of inputs and practices that represent subsistence to highly controlled, intensive production technologies and new varieties. Similarly, livestock models are needed that take into account climate, herd management, feed sources, and breeds. Farming system models are needed that integrate the various livestock and cropping systems, including their interactions, taking into account the socioeconomic and landscape characteristics of specific farms and a population of farms to address questions by individual farmers to policy makers at community to subnational, national, and global scales. The commonality of needs across these use cases should provide incentives for having the current agricultural system models and their components evolve to address these needs. Similarly, this commonality should provide incentive for the efforts at creating harmonized and open databases to ensure that these basic needs for data will address the needs of the Use Cases and models. The current status also led us to conclude that different platforms for combining models and data for specific purposes will be necessary, and that the design of NextGen models and data should take into account this need for a range of platforms for applying the models and providing outputs needed for DSSs.

Finally, based on the current status of models, data, and DSS, a strategy should include the appropriate modification and in some cases re-programming of existing component models that already include many of the capabilities needed for NextGen, such that components can be extended to respond to factors that are not currently considered by the models, to facilitate the use of a range of methods, including statistical models and reduced form models, making use of extended databases and component models. Recent experience in AgMIP demonstrating the value of using multiple models should be considered so that time is not wasted in pursuing an unattainable goal of producing a perfect model for crops, livestock, and farming systems. The applications, as discussed herein via the Use Cases, should serve as strong incentives for setting strategies for continuing investments in agricultural models, data, and DSSs.

References

Adams, R. M., C. Rosenzweig, R. M. Peart, J. T. Ritchie, B. M. McCarl, J. D. Glyer, R. B. Curry, J. W. Jones, K. J. Boote, and L. H. Allen, Jr. 1990. "Global climate change and U.S. agriculture: An interdisciplinary assessment." *Nature* 345:219-224.

Addiscott, T. M., and R. J. Wagenet. 1985. "Concepts of solute leaching in soils: A review of modeling approaches." *Journal of Soil Science* 36:411-424.

AFRC 1993. Energy and protein requirements of ruminants. An Advisory manual prepared by the AFRC Technical Committee on Response to Nutrients. CABI International, Wallingford, UK; 1993.

AgMIP. 2014. "Regional Integrated Assessments Handbook." http://www.agmip.org/regional-integrated-assessments-handbook/.

Allen, R. G., L. S. Pereira, D. Raes and M. Smith. 1998. "Crop Evapotranspiration – Guidelines for Computing Crop Water Requirements. FAO Irrigation and drainage paper 56." Rome, Italy: Food and Agriculture Organization of the United Nations. ISBN 92-5-104219-5.

Allen, V. G., C. Batello, E. J. Berretta, J. Hodgson, M. Kothmann, X. D. Li, J. McIvoer, J. Milne, C. Morris, A. Peeters, and M. Sanderson. 2011. "An international terminology for grazing lands and grazing animals." *Grass and Forage Science* 66:2-28.

Anselin, L. 1988. "Spatial Econometrics: Methods and Models." *Studies in Operational Regional Science*. Norwell, MA: Kluwer Academic Press Netherlands.

Anselin, L., R. Bongiovanni, and J. Lowenberg-DeBoer 2004. "A Spatial Econometric Approach to the Economics of Site-Specific Nitrogen Management in Corn Production." *American Journal of Agricultural Economics. 86 (3):675-687.*

Antle, J.M. 1983. "Testing the Stochastic Structure of Production: A Flexible Moment-Based Approach." *Journal of Business and Economic Statistics* 1(3):192-201.

Antle, J.M. 2011. "Parsimonious Multi-Dimensional Impact Assessment." *American Journal of Agricultural Economics 93* (5):1292-1311.

Antle, J.M. and S.M. Capalbo. 2001. "Econometric-Process Models for Integrated Assessment of Agricultural Production Systems." *American Journal of Agricultural Economics* 83(2):389401.

Antle, J. M., B. O. Basso, R. T. Conant, C. Godfray, J. W. Jones, M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig, P. Tittonell, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation*.*" AgMIP, < url tbd >.

Antle, J. M., J. W. Jones and C. Rosenzweig. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Introduction." AgMIP, < url tbd >.

Antle, J. M., R.O. Valdivia, K.J. Boote, S. Janssen, J.W. Jones, C.H. Porter, C. Rosenzweig, A.C. Ruane, and P.J. Thorburn. 2015. AgMIP's Trans-disciplinary Agricultural Systems Approach to Regional Integrated Assessment of Climate Impact, Vulnerability and Adaptation. C. Rosenzweig and D. Hillel, eds. *Handbook of Climate Change and Agroecosystems: The Agricultural Model Intercomparison and Improvement Project Integrated Crop and Economic Assessments, Part 1*. London: Imperial College Press.

Asseng, S., F. Ewert, C. Rosenzweig, J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P.J. Thorburn, R.P. Rotter, D. Cammarano, N. Brisson, B. Basso, P. Martre, P.K. Aggarwal, C. Angulo, P. Bertuzzi, C. Beirnath, A.J. Challinor, J. Doltra, S. Gayler, R. Goldberg, R. Grant, L. Heng, J. Hooker, L.A. Hunt, J. Ingwersen, R.C. Izaurralde, K.C. Kersebaum, C. Müller, S. Naresh Kumar, C. Nendel, G. O'Leary, J.E. Olesen, T.M. Osborne, T. Palosuo, E. Priesack, D. Ripoche, M.A. Semenov, I. Shcherbak, P. Steduto, C. Stöckle, P. Stratonovitch, T. Streck, I. Supit, F. Tao, M. Travasso, K. Waha, D. Wallach, J.W. White, J.R. Williams, and J. Wolf. 2013. "Uncertainty in simulating wheat yields under climate change." *Nature of Climate Change* 3:827-832 doi:10.1038/nclimate1916.

ARC (1965). The Nutrient Requirements of Farm Livestock. No. 2, Ruminants. London; HMSO.

Baker, D.N., J.R. Lambert, and J.M. McKinion, 1983. "GOSSYM: A Simulator of Cotton Crop Growth and Yield. Tech. Bull. 1089, S. Carolina Agric. Expt. Stat., Clemson, SC., 134 pp.

Basso, B., O. Gargiulo, K. Paustian, P. G. Robertson, C. Porter, P. R. Grace, and J. W. Jones. 2011. "Procedures for initializing soil organic carbon pools in DSSAT-Century model for agricultural systems." *Soil Science Society of America Journal* 75(1):69-78.

Basso, B., D. Cammarano, D. Fiorentino, J.T. Ritchie, 2013. "Wheat yield response to spatially variable nitrogen fertilizer in Mediterranean environment." *European Journal of Agronomy* 51:65-70

Basso, B., J.T. Ritchie, F.J. Pierce, R.P. Braga, and J.W. Jones. 2001. "Spatial validation of crop models for precision agriculture." *Agricultural Systems* 68:97–112.

Basso, B., and J.T. Ritchie. 2012. "Assessing the impact of management strategies on water use efficiency using soil-plant-atmosphere models." *Vadose Zone Journal* 11(3): doi:10.2136/vzj2011.017

Bassu, S., N. Brisson, J.L. Durand, K.J. Boote, J. Lizaso, J.W. Jones, C. Rosenzweig, A.C. Ruane, M. Adam, C. Baron, B. Basso, C. Biernath, H. Boogaard, S. Conijn, M. Corbeels, D. Deryng, G. De Sanctis, S. Gayler, P. Grassini, J. Hatfield, S. Hoek, C. Izaurralde, R. Jongschaap, A. Kemanian, K. Kersebaum, C. Müller, C. Nendel, E.Priesack, M. Virginia, P. Nin, F. Sau, I. Shcherbak, F. Tao, E.Teixeira, D. Makowski, D. Timlin, and K. Waha. 2014. "How do various maize crop models vary in their responses to climate change factors?" *Global Change Biology 20*(7): 2301-2320.

Bates, S.L., J.Z. Zhao, R.T. Roush, and A.M. Shelton. 2005. "Insect resistance management in GM crops: past, present and future." *Nature Biotechnology* 23:57-62.

Batchelor, W. D., J. W. Jones, and K. J. Boote. 1993. "Extending the Use of Crop Models to Study Pest Damage." *Transactions of the American Society of Agricultural Engineers* 36(2):551-558.

Baudron, F., S. Delmotte, M. Corbeels, J.M. Herrera, and P. Tittonell. 2014 (in press). "Multi-scale trade-off analysis of cereal residue use for livestock feeding vs. soil mulching in the Mid-Zambezi Valley, Zimbabwe." *Agricultural Systems* http://dx.doi.org/10.1016/j.agsy.2014.03.002.

Beck, H.W., P. Jones and J.W. Jones. 1989. "SOYBUG: An expert system for soybean insect pest management." *Agricultural Systems* 30(3):269-286.

Bellman, R.E. 1957. Dynamic Programming. Princeton University Press, Princeton, NJ. Republished 2003: Dover, ISBN 0-486-42809-5.

Berger, T., and C. Troost. 2014. "Agent-based Modelling of Climate Adaptation and Mitigation Options in Agriculture." *Journal of Agricultural Economics* 65(2):323-348.

Billari, F. C., T. Fent, A. Prskawetz, and J. Scheffran. 2006. Agent-Based Computational Modelling: An Introduction. Agent-Based Computational Modelling. Physica-Verlag HD: 1-16.

Boggess, W.G., P.J. van Blokland, and S.D. Moss. 1989. "FinARS: A financial analysis review expert system." *Agricultural Systems* 31(1):19-34.

Booker, J. F, R.E. Howitt, A. Michelsen, and R. Young. 2012. "Economics and the Modeling of Water resources and Policies." *Natural Resource Modeling*. 25(1).

Bondeau, A., P.C. Smith, S. Zaehle, S. Schaphoff, W. Lucht, W. Cramer, D. Gerten, H. Lotze-Campen, C. Muller, M. Reichstein, and B. Smith. 2007. "Modelling the role of agriculture for the 20th century global terrestrial carbon balance." *Global Change Biology* 13:679-706. Doi:10.1111/j.1365-2486.2006.01305.x.

Boote, K. J., L. H. Allen, Jr., P. V. Vara Prasad, and J. W. Jones. 2010. "Testing effects of climate change in crop models." In: D. Hillel and C. Rosenzweig (eds.). *Handbook of Climate Change and Agroecosystems.* Imperial College Press, London UK.

Boote, K. J., J. W. Jones, G. Hoogenboom, and N. B. Pickering. 1998. "The CROPGRO Model for Grain Legumes". In G. Y Tsuji, G. Hoogenboom, and P. K. Thornton (eds.). *Understanding Options for Agricultural Production*. Kluwer Academic Publishers, Dordrecht. 99-128.

Boote, K. J., J. W. Jones, J. W. Mishoe, and R. D. Berger. 1983. "Coupling pests to crop growth simulators to predict yield reductions." *Phytopathology* 73:1581-1587.

Boote, K.J., J.W. Jones, J. W. Mishoe, and G. G. Wilkerson. 1986. "Modeling growth and yield of groundnut." In Agrometeorology of Groundnut: Proceedings of an International Symposium, 21-26 Aug 1985, ICRISAT Sahelian Center, Niamey, Niger. ICRISAT, Patancheru, A.P. 502 324, India, 243—254

Boote, K. J., J. W. Jones, and N. B. Pickering. 1996. "Potential uses and limitations of crop models." *Agronomy Journal* 88:704-716.

Bouma, J., and J. W. Jones. 2001. "An international collaborative network for agricultural systems applications (ICASA)." *Agricultural Systems* 70(2):355-368.

Bouman, B. A. M., H. van Keulen, H. H. van Laar, and R. Rabbinge. 1996. "The 'School of de Wit' Crop Growth Simulation Models: A Pedigree and Historical Overview." *Agricultural Systems* 52(2-3):171-198.

Bouman, B. A. M., M. J. Kropff, T. P. Tuong, M. C. S. Woperies, and H. H. van Laar. 2001. "ORYZA 2000: Modeling lowland rice." International Rice Research Institute and Wageningen University and Research Center. p 235.

Bouwman, A.F., A.H.W. Beusen, J. Griffioen, J.W. Van Groenigen, M.M. Hefting, O. Oenema, P.J.T.M. Van Puijenbroek, S. Seitzinger, C.P. Slomp, and E. Stehfest. 2013. "Global trends and uncertainties in terrestrial denitrification and N_2O emissions." *Philosophical Transactions of the Royal Society* 368(1621):20130112 doi:10.1098/rstb.2013.0112.

Bouwman, A.F., K.W. Van der Hoek, B. Eickhout, and I. Soenario. 2005. "Exploring changes in world ruminant production systems." *Agricultural Systems* 84(2):121-153.

Bowen, H.D., R.F. Colwick and D.G. Batchelder. 1973. "Computer Simulation of Crop Production-Potential and Hazards." *Journal of Agricultural Engineering Research* 54(10):42-45.

Box, G.E.P., and N. R. Draper. 1987. Empirical Model Building and Response Surfaces. John Wiley & Sons, New York, NY. p 424.

Brooks-Pollock, E., Roberts, G.O. and Keeling, M.J. 2014. "A dynamic model of bovine tuberculosis spread and control in Great Britain." *Nature* 511:228-231.

Burt. O.R, and J.R. Allison. 1963. "Farm Management Decisions with Dynamic Programming." *Journal of Farm Economics* 45(1):121-136.

Carter. M,R. 1984. "Identification of the Inverse Relationship between Farm Size and Productivity: An Empirical Analysis of Peasant Agricultural Production." *Oxford Economic Papers New Series* 36(1):131-145.

Castelan-Ortega, O., R.H. Fawcett, C. Arriaga, M. and Herrero. 2003. "A decision support system for smallholder campesino maize-cattle production systems of the Toluca Valley in Central Mexico. 1. Integrating biological and socio-economic models into a holistic system." *Agricultural Systems* 75:1-21.

Challinor, A. J., T. R. Wheeler, J. M. Slingo, P. Q. Craufurd, and D. I. F. Grimes. 2004. "Design and optimization of a large-area process-based model for annual crops." *Agricultural and Forest Meteorology Journal* 124:99–120.

Chikowo, R., M. Corbeels, P. Tittonell, B. Vanlauwe, A. Whitbread, and K.E. Giller. 2008. "Aggregating field-scale knowledge into farm-scale models of African smallholder systems: summary functions to simulate crop production using APSIM." *Agricultural Systems* 97:151–166.

Claessens, L., J.J. Stoorvogel, and J.M. Antle. 2009. "Economic viability of adopting dual-purpose sweetpotato in Vihiga district, Western Kenya: a minimum data approach." *Agricultural Systems* 99:13-22.

Cohn, A.S., A. Mosnier, P. Havlik, H. Valin, M. Herrero, E. Schmid, M. O'Hare, and M. Obsersteiner. 2014. " Cattle ranching intensification in Brazil can reduce global greenhouse gas emissions by sparing land from deforestation." *Proceedings of the National Academy of Sciences* 111(20):7236-7241.

Condon, L.E. and R.M. Maxwell. 2013. "Implementation of a linear optimization water allocation algorithm into a fully integrated physical hydrology model." *Advances in Water Resources* 60:135–147.

Conway, G. R. 1987. "The properties of agroecosystems." *Agricultural Systems* 24(2):95-117.

Coughenour, M.B., S.J. McMaughton, and L.L. Wallace. 1984."Modelling primary production of perennial graminoids – uniting physiological processes and morphometric traits." *Ecological Modelling* 23:101-134.

CSIRO. 1980. SIROTAC - Cotton by computer. http://www.scienceimage.csiro.au/library/plants/v/12312/cotton-by-computer/

Curry, R. B., R. M. Peart, J. W. Jones, K. J. Boote, and L. H. Allen, Jr. 1990. "Response of crop yield to predicted changes in climate and atmospheric CO2 using simulation." *Transactions of the American Society of Agricultural Engineers* 33:1383-1390.

Davis, J.M., T.C. Cartwright, and J.O. Sanders. 1976. "Alternative beef production systems for Guyana." *Journal of Animal Science* 43: 235.

De Wit, C. T. 1958. Transpiration and crop yields. *Volume 64 of Agricultural research report / Netherlands Volume 59 of Mededeling (Instituut voor Biologisch en Scheikundig Onderzoek va Landbouwgewasses) Verslagen van landbouwkundige onderzoekingen*. Institute of Biological and Chemical Research on Field Crops and Herbage, 1958.

Delgado, C., M. Rosegrant, H. Steinfeld, S. Ehui, C. Courbois. 1999. "Livestock to 2020: The Next Food Revolution." *Outlook on Agriculture* 30(1):27-29.

Denison, R.F. 2012. Darwinian Agriculture: How Understanding Evolution Can Improve Agriculture. Princeton University Press, Princeton.

Dent, J. B., and M. J. Blackie. 1979. Systems Simulation in Agriculture. Applied Science Publishers. London, 180.

De Fraiture, C., V. Smakhtin, D. Bossio, P. McCornick, C.T. Hoanh, A. Noble, D. Molden, F. Gichuki, M. Giordano, M. Finlayson, and H. Turral. 2007. "Facing climate change by securing water for food, livelihoods and ecosystems." *Journal of Semi-Arid Tropical Agricultural Research* 4(1):12.

Diaz, S., S. Lavorel, S. McIntyre, V. Falczuk, F. Cadanoves, D. G. Milchunas, C. Skarpe, G. Rusch, M. Sternberg, I. Noy-Meir, J. J. Landsberg, W. Zhang, H. Clark, and B. D. Campbell. 2007. "Plant trait responses to grazing – a global synthesis." *Global Change Biology* 13:313-341.

Diekmann, O. and H. Heesterbeck. 2012. Mathematical Tools for Understanding Infectious Disease Dynamics. Princeton University Press, Princeton.

D'Odorico, P., G. S. Okin, and B. T. Bestelmeyer. 2012. "A synthetic review of feedbacks and drivers of shrub encroachment in arid grasslands." *Ecohydrology* 5:520-530.

Duncan, W.G. 1972. "SIMCOT: A simulation of cotton growth and yield." In: C.M.Murphy (Ed.) *Proceedings of a Workshop for Modeling Tree Growth*. Duke University, Durham, North Carolina pp 115-118.

Dzotsi, K. A., B. Basso, and J.W. Jones. 2013. "Development, uncertainty and sensitivity analysis of the simple SALUS crop model in DSSAT." *Ecological Modelling* 260:62–76.

Ehmke, T. (2013). "Unmanned Aerial Systems for Field Scouting and Spraying." CSA News 58(12): 4-9.

Elliott, J., D. Kelly, J. Chryssanthacopoulos, M. Glotter, K. Jhunjhnuwala, N. Best, M. Wilde, and I. Foster. 2014. "The parallel system for integrating impact models and sectors (pSIMS)." *Environmental Modelling & Software* DOI: 10.1016/j.envsoft.2014.04.008.

FAO, 1976. A framework for land evaluation. FAO Soils Bulletin 32.

FAO, 1978-81. Report on the Agro-ecological Zones project. Vol. 1. Methodology and results for Africa. Vol.2. Results for southwest Asia; Vol.3. Methodology and results for South and Central America; Vol. 4, Results for Southeast Asia. FAO World Soil Resources Report, 48/1, 4.

FAO. 2006. *Livestock's long shadow: Environmental issues and options*. FAO, Rome, Italy, p. 390.

FAO. 2013. Greenhouse gas emissions from ruminant supply chains. A global life cycle assessment. Food and Agriculture Organisation of the United Nations. Rome, Italy.

Farquhar, G.D., S. von Caemmerer, and J.A. Berry. 1980." A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species." *Planta* 149:78-90.

Fatondji, D., R. Tabo, J. W. Jones, A. Adamou, and O. Hassane. 2010. "Water use and yield of millet under the zai system: Understanding the processes using simulation." In: Increasing the Productivity & Sustainability of Rainfed Cropping Systems of Poor Smallholder Farmers, D. Humphreys and R. S. Bayot (eds.). Proceedings of the CGIAR Challenge Program on Water and Food International Workshop on Rainfed Cropping Systems, Colombo, Sri Lanka. pp. 125-146.

Fischer, G., K. Frohberg, M.L. Parry, and C. Rosenzweig. 1995. "Climate change and world food supply, demand, and trade." In: Rosenzweig et al. (eds.), *Climate Change and Agriculture: Analysis of Potential International Impacts, ASA special publication* 59. pp. 341-382. American Society of Agronomy, Madison, WI, USA.

Flichman, G. and T. Allen 2012. "Bio-economic modeling: State-of-the-art and key priorities" Centre International de Hautes Etudes Agronomiques Méditerranéennes - Institut Agronomique Méditerranéen de Montpellier, France.

Flichman, G. (Ed). 2012. Bio-Economic Models Applied to Agricultural Systems. Springer

Forbes, J.M. The Voluntary Food Intake of Farm Animals. Butterworth-Heineman 1986.

Forrester, J.W. 1968. Principles of Systems. (2nd ed.). Waltham, M. Pegasus Communications. 391 pp.

Fraisse, C.W., N. Breuer, JG. Bellow, V. Cabrera, U. Hatch, G. Hoogenboom, K. Ingram, J.W. Jones, J. O'Brien, J. Paz, and D. Zierden. 2006. "AgClimate: A climate forecast information system for agricultural risk management in the southeastern USA." *Computers & Electronics in Agriculture* 53(1):13-27.

Freer,M., J.L. Davidson, J.S. Armstrong, and J.R. Donnelly. 1970. "Simulation of grazing systems." In: *Proceedings of the XI International Grassland Congress*, University of Queensland Press, St. Lucia, Queensland, Australia, pp.913-917.

Freer, M. A.D. Moore, and J.R. Donnelly. 1997. "GRAZPLAN: Decision support systems for Australian grazing enterprises. I. Overview of the GRAZPLAN project and a description of the MetAccess and LambAlive DSS." *Agricultural Systems* 54:57-76.

Fritz, S., L. You, A. Bun, L. See, I. McCallum, C. Schill, C. Perger, J. Liu, M. Hansen, and M. Obersteiner. 2011. "Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets." *Geophysical Research Letters 38*(4):L04404, doi:10.1029/2010GL046213.

Fritz, S., L. See, C. Justice, I. Becker-Reshef, L. Bydekerke, R. Cumani, and P. Defourney. 2013. "The need for improvement maps of global cropland." *Eos Transcations, American Geophysical Union* 94(3):31 doi:10.1002/2013EO030005.

Gervois, S., N. de Noblet-Ducoudre, N. Viovy, and P. Ciais. 2008. "Including croplands in a global biosphere model: methodology and evaluation at specific sites." *Earth Interactions* 8(16):1-25.

Gonzalez-Estrada, E., L.C. Rodriguez, V.K. Walen, J.B. Naab, J. Koo, J.W. Jones, M. Herrero, and P. K. Thornton. 2008. "Carbon sequestration and farm income in West Africa: Identifying best management practices for smallholder agricultural systems in northern Ghana." *Ecological Economics* 67(3):492-502.

Gustafson, D., J. W. Jones, C. H. Porter, G. Hyman, M. Edgerton, T. Gocken, J. Shryock, M. Doane, and N. Ramsey. 2014. "Climate adaptation imperatives: Untapped global maize yield opportunities." *International Journal of Agricultural Sustainability* DOI:10.1080/14735903.2013.867694.

Hammer, G., M. Cooper, F.Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, and D. Podlich. 2006. "Models for navigating biological complexity in breeding improved crop plants." *TRENDS in Plant Science* 11(12):587-593.

Havlik, P., H. Valin, M. Herrero, M. Obersteiner, E. Schmid, M.C. Rufino, A. Mosnier, P.K. Thornton, H. Bottcher, R.T. Conant, S. Frank, S. Fritz, S. Fuss, F. Kraxner, and A. Notenbaert. 2014. "Climate change mitigation through livestock system transitions." *Proceedings of the National Academy of Sciences* 111:3709-3714.

Harou, J, M. Pulido-Velazquez, D. E. Rosenberg, J. Medellin-Azuara, J. R. Lund, and R. E. Howitt. 2009. "Hydro-economic models: Concepts, design, applications, and future prospects." *Journal of Hydrology* 375:627-643.

Hazell, P.B.R, and R.D. Norton. 1986. Mathematical Programming for Economic Analysis in Agriculture. Macmillan Publishing Company, London.

Hazell, P.B.R, and P. L. Scandizzo. 1975. "Market Intervention Policies when Production is Risky" *Volume 24 of World Bank reprint series*. World Bank. Washington. DC.

He, J., M. D. Dukes, G. J. Hochmuth, J. W. Jones, and W.D. Graham. 2012. "Identifying irrigation and nitrogen best management practices for sweet corn production on sandy soils using CERES-Maize model." *Agricultural Water Management* 109:61-70.

Heady. E.O. 1957 "An Econometric Investigation of Agricultural Production Functions" *Econometrica* 25(2):249-268.

Heady. E.O. and J.L. Dillon. 1964 Agricultural production functions. Ames Iowa State University Press

Herrero, M., R.H. Fawcett, and J.B. Dent. 1996. Integrating simulation models to optimise nutrition and management for dairy farms: a methodology. In: Livestock farming Systems: Research, Development, Socio-Economics and the Land Manager. J.B. Dent et al. (eds.) European Association for Animal Production Publication No. 79. Wageiningen Pers., The Netherlands. pp. 322-326.

Herrero, M., R. H. Fawcett, J.B. Dent. 1998. The plant/animal interface in models of grazing systems. In: Agricultural Systems Modelling and Simulation. M. Peart and R.B. Curry (eds.) Marcel Dekker, New York, pp. 495-542.

Herrero, M., R.H. Fawcett, and J.B. Dent. 1999. "Bio-economic evaluation of dairy farm management scenarios using integrated simulation and multiple-criteria models." *Agricultural Systems* 62:149-168.

Herrero, M., E. Gonzalez-Estrada, P.K. Thornton, C. Quiros, M.M. Waithaka, R. Ruiz, and G. Hoogenboom. 2007. "IMPACT: generic household-level databases and diagnostics tools for integrated crop–livestock systems analysis." *Agricultural Systems* 92:240–265.

Herrero, M., P. Havlík, H. Valin, A. Notenbaert, M.C. Rufino, P.K. Thornton, M. Blümmel, F. Weiss, D. Grace, and M. Obersteiner. 2013. "Biomass use, production, feed efficiencies, and greenhouse gas emissions from global livestock systems." *Proceedings of the National Academy of Sciences* 110(52):20888-20893.

Herrero, M., P.K. Thornton, A. Bernués, I. Baltenweck, J. Vervoort, J. van de Steeg, S. Makokha, M.T. van Wijk, S. Karanja, and S. Staal. 2014. "Exploring future changes in smallholder farming systems by linking socio-economic scenarios with regional and household models." *Global Environmental Change* 24:165–182.

Hieronymi, A. 2013. "Understanding Systems Science: A Visual and Integrative Approach." *Systems Research* doi:10.1002/sres.2215.

Hillel, D. and C. Rosenzweig. 2014. Handbook of Climate Change and Agroecosystems: The Agricultural Model Intercomparison and Improvement Project Integrated Crop and Economic Assessments (Vols. 3 & 4). ICP Series on Climate Change Impacts, Adaptation, and Mitigation. Imperial College Press. In Press.

Hoogenboom, G., J.W. Jones, P.W. Wilkens, C.H. Porter, K.J. Boote, L.A. Hunt, U. Singh, J.L. Lizaso, J.W. White, O. Uryasev, F.S. Royce, R. Ogoshi, A.J. Gijsman, G.Y. Tsuji, and J. Koo. 2012. Decision Support System for Agrotechnology Transfer (DSSAT) Version 4.5 [CD-ROM]. University of Hawaii, Honolulu, Hawaii.

Hoogenboom, G., and J.W. White. 2003. "Improving Physiological Assumptions Of Simulation Models By Using Gene-Based Approaches." *Agronomy Journal* 95(1):82-89.

Howitt, R.E. 1995. "Positive Mathematical Programming" *American Journal of Agricultural Economics* 77: 329-342.

Hunt, L. A., J. W. Jones, G. Hoogenboom, D. C. Godwin, U. Singh, N. Pickering, P. K. Thornton, K. J. Boote, and J. T. Ritchie. 1994. "General Input and Output File Structures for Crop Simulation Models." In: *Application of Modeling in the Semi-arid Tropics*. Published by CO-DATA, International Council of Scientific Unions. pp. 35-72.

IADB 1975. HerdSIM simulation model: user manual. Project Analysis Paper No. 2. Economic and Social Development Department, Inter-American Development Bank. Washington, D.C.

ILCA 1978. Mathematical modeling of livestock production systems: application of the Texas A&M University beef cattle production model to Botswana. International Livestock Centre for Africa. Systems Study No.1 , Addis Ababa, Ethiopia.

Illius, A.W., and M.S. Allen. 1994. "Assessing forage quality using integrated models of intake and digestion by ruminants." In: Fahey, G.C. (Ed), *Forage Quality, Evaluation, and Utilization*. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, Madison, Wisconsin, USA, pp. 869–890

IBSNAT. 1984. Experimental design and data collection procedures for IBSNAT. The minimum data set for systems analysis and crop simulation. 1st Edition, IBSNAT Technical Report No 1. http://pdf.usaid.gov/pdf_docs/PNABK919.pdf

IPCC. 1990. Climate Change: The IPCC Scientific Assessment. J.T. Houghton, G.J. Jenkins and J.J. Ephraums (eds.). Cambridge University Press, Cambridge, Great Britain, New York, NY, USA pp. 410

Jamieson, P. D., J. R. Porter, J. Goudriaan, J. T. Ritchie, H. van Keulen, and W. Stoll. 1998. "A comparison of the models AFRCWHEAT2, CERES-Wheat, Sirius, SUCROS2 and SWHEAT with measurements from wheat grown under drought." *Field Crops Research* 55(1):23-44.

Janssen, S., C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, and J. M. Antle. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support." AgMIP, < url tbd >.

Jenkins, M.W., J.R. Lund, R.E. Howitt, A.J. Draper, S.M. Msangi, S.K. Tanaka, R.S. Ritzema, and G.F. Marques. 2004. "Optimization of California's Water System: Results and Insights." *Journal of Water Resources Planning and Management* 130(4):271-280.

Jones, C. A., and J. R. Kiniry (eds.). 1986. "CERES-Maize: A Simulation Model of Maize Growth and Development." *Texas A&M University Press*. College Station, Texas. p 208.

Jones, J.W., 1993. "Decision support systems for agricultural development." In: Penning de Vries, F., Teng, P., Metselaar, K. (eds.). *Systems Approaches for Agricultural Development.* Kluwer Academic Press, Boston. p. 459-471.

Jones, J.W., L.G. Brown, and J. Hesketh. 1980. "COTCROP: A computer model for cotton growth and yield." In: J.D. Hesketh, J.W. Jones (Eds.) *Predicting Photosynthesis for Ecosystem Models*, CRC Press, Boca Raton, Florida, pp. 209-241.

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science." AgMIP, < url tbd >.

Jones, J.W., J.D. Hesketh, E.J. Kamprath, and H.D. Bowen, 1974. "Development of a Nitrogen Balance for Cotton Growth Models: A First Approximation." *Crop Science* 14(4):541-546.

Jones, J. W., A. Kenig, and C. E. Vallejos. 1999. Reduced state-variable tomato growth model. Trans. ASAE 42(1):255-265.

Jones, J. W. G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. "The DSSAT cropping system model." *European Journal of Agronomy* 18(3-4):235-265.

Johnson, I, and J.H.M. Thornley. 1983. "Vegetative crop growth model incorporating leaf area expansion and senescence and applied to grass." *Plant Cell and Environment* 6:721-729.

Johnson, I. 2002. The SGS pasture model: documentation. www.imj.com.au/sgs

Judd. K.L. 1998. "Numerical Methods in Economics" *MIT Press*. Cambridge. MA

Just.R.E, D.Zilberman, and E Hochman. 1983 "Estimation of Multicrop production Functions." *American Journal of Agricultural Economics* 65(4).

Just, R. E., and R. D. Pope. 1978. "Stochastic Specification of Production Function and Economic Implications." *Journal of Econometrics* 7:67-76.

Knapp, K.C., and L.J. Olson. 1996. "Dynamic resource management: intertemporal substitution and risk aversion." *American Journal of Agricultural Economics* 78(4):1004-1014.

Keig, G., and J.R. McAlpine. 1969. "Watbal: A computer system for the estimation of soil moisture regimes from simple climatic data." CSIRO Australia, Division of Land Research, Tech Memo 69/9.

Keating, B.A., D.C. Godwin, and J.M. Watiki. 1991. "Optimization of nitrogen inputs under climatic risk." In: R.C. Muchow and J.A. Bellamy (eds.) *Climatic Risk in Crop Production - models and management for the semi-arid tropics and sub-tropics.* CAB International, Wallingford, UK. p 329-357.

Keating, B.A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburn, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairn, and C.J. Smith. 2003. "An overview of APSIM, a model designed for farming systems simulation." *European Journal of Agronomy* 18:267-288.

Kirkby, M.J. (Ed.). Hillslope Hydrology, Chichester: John Wiley & Sons. 1978.

Kimball, B. A. 2005. "Theory and performance of an infrared heater for ecosystem warming." *Global Change Biology* 11:2041-2056.

Kobayashi,M, R.E. Howitt, L.S. Jarvis, and E.A. Laca. 2007 "Stochastic Rangeland Use Under Capital Constraints." *American Journal of Agricultural Economics* 89(3):805-817.

Konandreas, P.A. and F.M. Anderson. 1982. "Cattle herd dynamics: an integer and stochastic model for evaluating production alternatives." *ILCA Research Report No. 2.* International Livestock Centre for Africa, Addis Ababa, Ethiopia.

Kucharik, C.J. 2003. "Evaluation of a process-based agro-ecosystem model (Agro-IBIS) across the U.S. cornbelt: simulations of the inter-annual variability in maize yield." *Earth Interactions* 7:1-33.

Kumudini, S., F.H. Andrade, K.J. Boote, G.A. Brown, K.A. Dzotsi, G.O. Edmeades, T. Gocken, M. Goodwin, A.L. Halter, G.L. Hammer, J.L. Hatfield, J.W. Jones, A.R. Kemanian, S.-H. Kim, J. Kiniry, J.I. Lizaso, C. Nendel, R.L. Nielsen, B. Parent, C. O. Stöckle, F. Tardieu, P.R. Thomison, D.J. Timlin, T.J. Vyn, D. Wallach, H.S. Yang, and M. Tollenaar. 2014. "Predicting maize phenology: Intercomparison of functions for developmental response to temperature." *Agronomy Journal* 106(6): (in press).

Lau, L.J. and P.A. Yotopolous. 1971. "A Test for Relative Efficiency and Application to Indian Agriculture." *American Economic Review* 61(1):94-109

Lin. W, G. W. Dean and C. V. Moore 1974. "An Empirical Test of Utility vs. Profit Maximization in Agricultural Production" *American Journal of Agricultural Economics* 56(3):497-508.

Lisson, S., N. MacLeod, C. McDonald, J. Corfield, B. Pengelly, L. Wirajaswadi, R. Rahman, S. Bahar, R. Padjung, N. Razak, K. Puspadi, Dahlanuddin, Y. Sutaryono, S. Saenong, T. Panjaitan, L. Hadiawati, A. Ash, and L. Brennan. 2010. "A participatory, farming systems approach to improving Bali cattle production in the smallholder crop-livestock systems of Eastern Indonesia." *Agricultural Systems* 103:486–497.

Lobell, D.B., W. Schlenker, and J. Costa-Roberts. 2011. "Climate Trends and Global Crop Production Since 1980." *Science* 333(6042):616-620.

Long, S.P., E.A. Ainsworth, A.D.B. Leakey, J. Nösberger, and D.R. Ort. 2006. "Food for thought: lower-than-expected crop yield stimulation with rising CO2 concentrations." *Science* 312:1918-1921.

Loewer, O. 1998. "GRAZE: A beef-forage model of selective grazing." In: M. Peart and R.B. Curry (eds.). *Agricultural Systems Modelling and Simulation*. Marcel Dekker, New York, p. 301-417.

Matson, P.A., W.J. Parton, A.G. Power, and M.J. Swift. 1997. "Agricultural Intensification and Ecosystem Properties." *Science* 277(504): DOI: 10.1126/science.277.5325.504.

Maxwell, R.M., M. Putti, S.B. Meyerhoff, J.O. Delfs, I.M. Ferguson, V. Ivanov, J. Kim, O. Kolditz, S.J. Kollet, M. Kumar, S. Lopez, J. Niu, C. Paniconi, Y.-J. Park, M.S. Phanikumar, C. Shen, E.A. Sudicky, and M. Sulis. 2014. "Surface–subsurface model intercomparison: a first set of benchmark results to diagnose integrated hydrology and feedbacks." *Water Resources Research* 50(2):1531–1549.

May, R.M. 1976. "Simple mathematical models with very complicated dynamics." *Nature* 261:459-467.

McCown, R. L. 1973. "An evaluation of the influence of available soil water storage capacity on growing season length and yield of tropical pastures using simple water balance models." *Agricultural and Forest Meteorology* 11:53-63.

McCown, R.L. 2002. "Changing systems for supporting farmers' decisions: problems, paradigms, and prospects." *Agricultural Systems* 74:179–220.

McCown, R.L., B.A. Keating, M.E. Probert, and R.K. Jones. 1992. "Strategies for Sustainable Crop Production in Semi-arid Africa." *Outlook on Agriculture* 21:21-31.

McCown, R.L., G.L. Hammer, J.N.G. Hargreaves, D.P. Holzworth, and D.M. Freebairn. 1996. "APSIM: a novel software system for model development, model testing and simulation in agricultural systems research." *Agricultural Systems* 50:255-271.

Mérel. P, F. Yi, S. Bucaram, J. Lee, R. Howitt and J. Six. "Incorporating yield information from a biogeochemical model into an agricultural production model to infer adoption of a new bioenergy crop." In G Flichman 2012 (ed.) *Bioeconomic Models Applied Agricultural Systems*. Springer

Mérel. P, and R.E. Howitt. 2014 "Theory and Application of Positive Mathematical Programming in Agriculture and the Environment." *Annual Review of Resource Economics* 6.

Messina, C. D., J. W. Jones, K. J. Boote, and C. E. Vallejos. 2006. "A gene-based model to simulate soybean development and yield responses to environment." *Crop Science* 46:456-466.

Messina, C.D., D. Podlich, Z. Dong, M. Samples, and M. Cooper. 2011. "Yield-trait performance landscapes: from theory to application in breeding maize for drought tolerance." *Journal of Experimental Botany* 62(3):855-868.

Mendelsohn, R., W.D. Nordhaus and D. Shaw. 1994. "The Impact of Global Warming on Agriculture: A Ricardian Approach." *American Economic Review* 84(4):753-771.

Michalski, R.S., J.G. Carbonell, T.M. Mitchell, and M. Kaufmann. Machine Learning: An Artificial Intelligence Approach, Volume 1. Kaufmann, 1985.

Milchunas, D. G. and W. K. Lauenroth. 1993. "Quantitative effects of grazing on vegetation and soils over a global range of environments." *Ecological Monographs* 63:327-366.

Miranda.M.J, and P.J.Fackler. 2002. "Applied Computational Economics and Finance." *MIT Press*. Cambridge, MA

Monfreda, C., N. Ramankutty, and J.A. Foley. 2008. "Farming the planet 2: Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000." *Global Biogeochemical Cycles* 22(1):GB1022 doi:10.1029/2007GB002947.

Morgan, J. A., D. G. Milchunas, D. R. LeCain, M. S. West, and A. Mosier. 2007. "Carbon dioxide enrichment alters plant community structure and accelerates shrub growth in the shortgrass steppe." *Proceedings of the National Academy of Sciences* 104:14724-14729.

Mundlak, Y. 1961. "Empirical production function free of management bias." *American Journal of Agricultural Economics* 43(1):44-56.

Muñoz-Carpena, R., G. Vellidis, A. Shirmohammadi and W.W. Wallender. 2006. "Evaluation of Modeling Tools for TMDL Development and Implementation 392KB]." *Transactions of the American Society of Agricultural Engineers* 49(4):961-965.

Naab, J. B., P. Singh, K. J. Boote, J. W. Jones, and K. O. Marfo. 2004. "Using the CROPGRO-peanut model to quantify yield gaps of peanut in the Guinean savanna zone of Ghana." *Agronomy Journal* 96:1231-1242.

Nelson, G.C., M.W. Rosegrant, J. Koo, R.Robertson, T. Sulser, T. Zhu, C. Ringler, S. Msangi, A. Palazzo, M. Batka, M. Magalhaes, R. Valmonte-Santos, M. Ewing, and D. Lee. 2009. "Climate Change. Impact on Agriculture and Costs of Adaptation." *International Food Policy Research Institute.* International Food Policy Research Institute, Washington, D.C.

Nelson, G.C., H. Valin, R.D. Sands, P. Havlik, H. Ahammad, D. Deryng, J. Elliott, S. Fujimori, T. Hasegawa, E. Heyhoe, P. Kyle, M. Von Lampe, H. Lotze-Campen, D.M. d'Croz, H. van Meijl, D. van der Mensbrugghe, C. Muller, A. Popp, R. Roberston, S. Robinson, E. Schmid, C. Schmitz, A. Tabeu, D. Willenbockel. 2013. "Climate change effects on agriculture: Economic responses to biophysical shocks." *Proceedings of the National Academy of Sciences* 111(9):3274-3279 doi:10.1073/pnas.1222465110.

Nicholson, C.F., R.W. Blake, C.I. Urbina, D.R. Lee, D.G. Fox, and P.J. Van Soest. 1994. "Economic comparison of nutritional management strategies for Venezuelan dual-purpose cattle systems." *Journal of Animal Science* 72:1680-1696.

NRC. 2001. "Nutrient requirements of dairy cattle." *Seventh Revised Edition*. National Research Council, Washington DC, US.

NRC 1945. Nutrient requirements of dairy cattle. Committee on Animal Nutrition, National Research Council. Washington, DC.

Orskov, E.R., and I. McDonald. 1979. "The estimation of protein degradability in the rumen from incubation measurements weighted according to rate of passage." *Journal of Agricultural Science* 92(2):499-503.

Osborne, T.M., D.M. Lawrence, J.M. Slingo, A.J. Challinor, and T.R. Wheeler. 2007. "Influence of Vegetation on Local Climate and Hydrology in the Tropics: Sensitivity to Soil Parameters." *Climate Dynamics* 23:45-61.

Osborne, T. M., D.M. Lawrence, A.J. Challinor, J.M. Slingo, and T.R. Wheeler. 2007. "Development and assessment of a coupled crop–climate model." *Global Change Biology* 13(1):169-183.

Osborne, T.M., J.M. Slingo, D.M. Lawrence, and T.R. Wheeler. 2009. "Examining the Interaction of Growing Crops with Local Climate using a Coupled Crop-Climate Model." *Journal of Climate* 22:1393-1411.

Parry, M. L., C. Rosenzweig, A. Iglesias, M. Livermore, and G. Fischer. 2004. "Effects of climate change on global food production under SRES emissions and socio-economic scenarios." *Global Environmental Change* 14:53-67.

PBL Netherlands Environmental Assessment Agency. "Trends in global CO_2 emissions: 2013 Report." PBL Netherlands Environmental Assessment Agency Institute for Environment and Sustainability (IES) of the European Commission's Joint Research Centre (JRC). The Hauge, 2013.

Penning de Vries, F. W. T., H. H. van Laar and M. J. Kropff (eds.). 1991. "Simulation and systems analysis for rice production (SARP)." *PUDOC*, Waneningen, The Netherlands, 1991. 369.

Pimentel, D., and R. Peshin (Eds). 2014. Use and benefit of pesticides in agricultural pest control. Integrated Pest Management, Pesticide Problems Vol 3. Springer 2014.

Pinnschmidt, H.O., W.D. Batchelor, and P.S. Teng. 1995. "Simulation of multiple species pest damage in rice using CERES-rice." *Agricultural Systems* 48(2):193-222.

Pinter, P. J. Jr., J. C. Ritchie, J. L. Hatfield, and G. F. Hart. 2003. "The agricultural research service's remote sensing program: an example of interagency collaboration". *Photogrammetric Engineering & Remote Sensing* 69(6):615–618.

Plant, R.E. 1989. "An integrated expert decision support system for agricultural management." *Agricultural Systems* 29:49-66.

Porter, C.H., J.W. Jones, S. Adiku, A.J. Gijsman, O. Gargiulo, and J.B. Naab. 2010. "Modeling organic carbon and carbon-mediated soil processes in DSSAT v4.5." *International Journal of Operational Research* 10(3):247-278.

Ramankutty, N., and J.A. Foley. 1999. "Estimating historical changes in global land cover: Croplands from 1700 to 1992." *Global Biogeochemical Cycles* 13:997-1027.

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley. 2008. "Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000." *Global Biogeochemical Cycles* 22:GB1003, doi:1010.1029/2007GB002952.

Richards, L.A. 1931. "Capillary conduction of liquids through porous mediums." *Physics* 1(5): 318–333.

Ritchie, J. T., 1972, "Model for predicting evaporation from a row crop with incomplete cover." *Water Resources Research* 8:1204.

Ritchie, J.T. 1991. "Specifications of the ideal model for predicting crop yields." In R.C. Muchow and J.A. Bellamy (eds.) *Climatic risk in crop production: Models and management for the semi-arid tropics and subtropics. Proc. Intnl. Symposium*, St. Lucia, Brisbane, Queensland, Australia. July 2-6, 1990. C.A.B. International, Wallingford, U.K. p. 97-122.

Ritchie, J.T., and B. Basso. 2008. "Water use efficiency is not constant when crop water supply is adequate or fixed: The role of agronomic management." *European Journal of Agronomy* 28(3):273-281.

Ritchie, J.T., and G. Alagarswamy. 2002. "Overview of crop models for assessment of crop production." In: O.C. Doering III, J.C. Randolph, J. Southworth, and R.A. Pfeifer (eds). *Effects of climate change and variability on agricultural production systems.* Kluwer Academic Publishing, Dordrecht, The Netherlands.

Ritchie, J. T., and S. Otter. 1984. "Description and performance of CERES-Wheat: a user-oriented wheat yield model." In: ARS Wheat Yield Project. ARS-38. National Technical Information Service, Springfield, Missouri, pp. 159-175.

Rodriguez, D., H. Cox, P. deVoil, and B. Power. 2014. "A participatory whole farm modeling approach to understand impacts and increase preparedness to climate change in Australia." *Agricultural Systems* 126:50-61.

Rosegrant, M. W., J. Koo, N. Cenacchi, C. Ringler, R. Robertson, M. Fisher, C. Cox, K. Garrett, N.D. Perez, and P. Sabbagh. 2014. "Food Security in a World of Natural Resource Scarcity: The Role of Agricultural Technologies." *International Food Policy Research Institute*. International Food Policy Research Institute, Washington, DC. DOI:http://dx.doi.org/10.2499/9780896298477

Rosegrant, M. W., M. Fernandez, A. Sinha, J. Alder, H. Ahammad, C. de Fraiture, B. Eickhour, J. Fonseca, J. Huang, O. Koyama, A.M. Omezzine, P. Pingali, R. Ramirez, C. Ringler, S. Robinson, P.Thornton, D. van Vuuren, and H. Yana-Shapiro. 2009. "Looking into the future for agriculture and AKST." In McIntyre, B. D.; Herren, H. R.; Wakhungu, J.; Watson, R. T. (Eds.). International Assessment of Agricultural Knowledge, Science and Technology for Development (IAASTD): Agriculture at a Crossroads, global report. Washington, DC, USA: Island Press. pp.307-376

Rosenzweig, C. and M. L. Parry. 1994. "Potential impact of climate change on world food supply." *Nature* 367:133-138.

Rosenzweig, C., J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P. Thorburn, J.M. Antle, G.C. Nelson, C. Porter, S. Janssen, S. Asseng, B. Basso, F. Ewert, D. Wallach, G. Baigorria, and J.M. Winter. 2013a. "The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies." *Agricultural and Forest Meteorology Journal* 170:166-182. doi:10.1016/j.agrformet.2012.09.011.

Rosenzweig, C., J. Elliott, D. Deryng, A. C. Ruane, A. Arneth, K. J. Boote, C. Folberth, M. Glotter, C. Müller, K. Neumann, F. Piontek, T. Pugh, E. Schmid, E. Stehfest, and J. W. Jones. 2013b. "Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison." *Proceedings of the National Academy of Sciences* doi:10.1073/pnas.1222463110.

Rufino, M., M. Herrero, M. van Wijk, R. Hemerink, N. de Ridder, and K. Giller. 2009. "Lifetime productivity of dairy cows in smallholder systems in the highlands of Kenya." *Journal of Animal Science* 3:1044-1056.

Ruiz, R., L. O. Tedeschi, J. C. Marini, D. G. Fox, A. N. Pell, G. Jarvis, and J. B. Russell. 2002. "The effect of a ruminal nitrogen (N) deficiency in dairy cows: evaluation of the Cornell net carbohydrate and protein system ruminal N deficiency adjustment." *Journal of Dairy Science* 85:2986-2999.

Russell, J. B., J. D. O'Connor, D. G. Fox, P. J. Van Soest, and C. J. Sniffen. 1992. "A net carbohydrate and protein system for evaluating cattle diets: I. Ruminal fermentation." *Journal of Animal Science* 70:3551-3561.

Sanders, J.O., and T.C. Cartwright. 1979. "A general cattle production systems model. 2. Procedures used for simulating animal performance." *Agricultural Systems* 4:289-309.

Schlenker, W., and D.B. Lobell. 2010. "Robust negative impacts of climate change on African agriculture." *Environmental Research Letters* 5:8 doi:10.1088/1748-9326/5/1/014010.

Schlenker .W, M. Auffhammer, S. Hsiang, and A. Sobel. 2013 "Using weather data and climate model output in economic analyses of climate change." *Review of Environmental Economics and Policy* 7(2):181-198.

Schlesinger, W.H., J.F. Reynolds, G.L. Cunningham, L. F. Huenneke, W.M. Jarrell, R.A. Virginia, and W.G. Whitford. 1990. "Biological feedbacks in global desertification." *Science* 247:1043-1048.

Seligman, N.G., and H. Van Keulen. 1981. "PAPRAN: a simulation model of annual pasture production limited by rainfall and nitrogen." In: M.J Frissel, J.A Van Veen (Eds.). *Simulation of Nitrogen Behaviour of Soil-Plant Systems, Pudoc*, Wageningen, The Netherlands (1981), pp. 192–220.

Singh, P., K. J. Boote, U. Kumar, K. Srinivas, S. N. Nigam and J. W. Jones. 2012. "Evaluation of genetic traits for improving productivity and adaptation of groundnut to climate change in India." *Journal of Agronomy and Crop Science* doi:10.1111/j.1439-037X.2012.00522.x.

Singh, P., S. Nedumaran, B. R. Ntare, K. J. Boote, N. P. Singh, K. Srinivas, and M. C. S. Bantilan. 2013. "Potential benefits of drought and heat tolerance in groundnut for adaptation to climate change in India and West Africa" *Mitigation and Adaptation Strategies of Global Change* 18:(doi 10.1007/s1027-012-9446-7).

Singh, P., S. Nedumaran, P. C. S. Traore, K. J. Boote, H. F. W. Rattunde, P. V. Vara Prasad, N. P. Singh, K. Srinivas, and M. C. S. Bantilan. 2014. "Quantifying potential benefits of drought and heat tolerance in rainy season sorghum for adapting to climate change. *Agricultural and Forest Meteorology Journal* 185:37-48.

Slatyer, R. O. 1960. "Agricultural climatology of the Yass valley." CSIRO Aust. Div. Land Res. Reg. Surv. Tech. Paper No. 13.

Slatyer, R. O. 1964. "Climate of the Leichhardt-Gilbert area." CSIRO Aust. Land Res. Ser. No. 11.

Spedding, C. R. W. 1976. Editorial. *Agricultural Systems* 1:1-3.

Shrivastava, V., W.D. Graham, R. Muñoz-Carpena and R. Maxwell. 2014. "Insights on geologic and vegetative controls over hydrologic behavior of a large complex basin - Global Sensitivity Analysis of an Integrated Parallel Hydrologic Model." *Journal of Hydrology* 519(B):2238–2257. doi: 10.1016/j.jhydrol.2014.10.020.

Staal S. J, I. Baltenweck, M.M. Waithaka, T. DeWolff, and L. Njoroge. 2002. "Location and uptake: integrated household and GIS analysis of technology adoption and land use, with application to smallholder dairy farms in Kenya." *Agricultural Economics* 27(3):295-315.

Stapleton, H.N., D.R. Buxton, F.L. Watson, D.J. Notling, and D.N. Baker. 1973. "COTTON: a computer simulation of cotton growth." University of Arizona, Agricultural Experiment Station. *Technical Bulletin* 206.

Steffen, W., and J. S. I. Ingram. 1995. "Global change and terrestrial ecosystems: An initial integration." *Journal of Biogeography* 22(2/3):165-174.

Stehfest, E., L. Bouwman, D. Vuuren, M.J. Elzen, B. Eickhout, and P. Kabat. 2009. "Climate benefits of changing diet." *Climatic Change* 95:83-102.

Steinfeld, H., P. Gerber, T. Wassenaar, V. Castel, M. Rosales, and C. de Haan. 2006. "Livestock's long shadow." FAO, Rome 2006

Stöckle, C.O., M. Donatelli, and R. Nelson. 2003. "CropSyst, a cropping systems simulation model." *European Journal of Agronomy* 18:289-307.

Stoorvogel, J.J., J.M. Antle, C.C. Crissman, and W. Bowen. 2004. "The Tradeoff Analysis Model: Integrated Bio-physical and Economic Modeling of Agricultural Production Systems." *Agricultural Systems* 80(1):43-66.

Stringham, T. K., W. C. Krueger, and P. L. Shaver. 2003. "State and transition modeling: An ecological process approach." *Journal of Range Management* 56:106-113.

Stuth, J. W., and M. Stafford-Smith. 1993. "Decision support for grazing lands: An overview." In: *Decision Support for the Management of Grazing Lands, Emerging Issues.*

Stuth, J.W., and B.G. Lyons (Eds.). Man Biosphere Ser. No. 11. Paris France: UNESCO. Pp. 1-35.

Tack, J., A. Harri and K. Coble. 2012. "More than mean effects: Modeling the effect of climate on the higher order moments of crop yields." *American Journal of Agricultural Economics* 94(5):1037-1054.

Takayama, T, and G.G. Judge. 1964. "An interregional activity analysis model of the agricultural sector". *Journal of Farm Economics* 349–365.

Taubert, F., K. Frank, and A. Huth. 2012. "A review of grassland models in the biofuel context." *Ecological Modelling* 245:84-93.

Taylor. E, and I Adelman. (2006) "Village Economies. The Design, Estimation and Use of Village wide Economic Models". Cambridge: *Cambridge University Press*, 2006.

Tedeschi,L., L. Cavalcanti, M. Fonseca, M. Herrero, and P.K. Thornton. 2014. "The evolution of dairy cattle models for predicting milk production: an agricultural model intercomparison and improvement project (AgMIP) for livestock." *Animal Production Science* (in press).

Thompson, L.M. 1969. "Weather and Technology in the Production of Corn in the US Corn Belt." *Agronomy Journal* 61(3):453-456.

Thorburn, P., K. Boote, J. Hargreaves, P. Poulton, and J. Jones. 2014. "Crop Systems Modeling in AgMIP: A New Protocol-driven Approach for Regional Integrated Assessments." In D. Hillel and C. Rosenzweig (Ed.), *Handbook of Climate Change and Agroecosystems: Agricultural Model Intercomparison and Improvement Project Integrated Crop and Economic Assessments (Vol. 3 & 4).* ICP Series on Climate Change Impacts, Adaptation, and Mitigation. Imperial College Press. *In Press.*

Thornton P.K., P.G. Jones, T. Owiyo, R.L. Kruska, M. Herrero, P. Kristjanson, A. Notenbaert, N. Bekele, and A. Omolo, with contributions from V. Orindi, A. Adwerah, B. Otiende, S. Bhadwal, K. Anantram, S. Nair, and V. Kumar. 2006. "Mapping climate vulnerability and poverty in Africa." *Report to the Department for International Development, ILRI*, Nairobi, Kenya, 198 pp.

Thornton, P.K., and M. Herrero, M. 2001. "Integrated crop-livestock simulation models for scenario analysis and impact assessment." *Agricultural Systems* 70:581-602.

Tittonell, P., S. Zingore, M.T. van Wijk, M. Corbeels, and K.E. Giller. 2007. "Nutrient use efficiencies and crop responses to N, P and manure applications in Zimbabwean soils: exploring management strategies across fertility gradients." *Field Crops Research* 100:348–368.

Tsuji, G. Y., G. Hoogenboom, and P.K. Thornton. 1998. Understanding options for agricultural production. Springer. 400 pp.

Tubiello, F.N., C. Rosenzweig, R. A. Goldberg, S. Jagtap, and J.W. Jones. 2002. "Effects of Climate Change on U.S. Crop Production: Simulation Results Using Two Different GCM Scenarios. Part I: Wheat, Potato, Maize, and Citrus." *Climate Research* 20(3):259-270.

Uehara, G., and G. Y. Tsuji. 1998. "Overview of IBSNAT." In: G.Y. Tsuji, G. Hoogenboom, and P.K. Thornton (eds), Chapter 1, pp. 1-7. Understanding Options for Agricultural Production. Springer, The Netherlands.

Van Arendonk, J.A.M., and A.A. Dijkhuizen. 1985 "Studies on the replacement policies in dairy cattle. III. Influence of variation in reproduction and production." *Livestock Production Science* 13:333–349.

Van Auken, O.W. 2000. "Shrub invasions of North American semiarid grasslands." *Annual Review of Ecology and Systematics* 31:197-205.

Van Bavel, C.H.M. 1953. "A drought criterion and its application in evaluating drought incidence and hazard." *Journal of Agronomy* 4:167-172.

Van Dyne, G.M. and J C. Anway. 1976. "Research –program for and process of building and testing grassland ecosystem models." *Journal of Range Management* 29:114 -122.

Van Ittersum, M. K., K. G. Cassman, P. Grassini, J. Wolf, P. Tittonell, and Z. Hochman. 2013. "Yield gap analysis with local to global relevance—A review." *Field Crops Research* 143:4-17.

van Ittersum, M.K., F. Ewert, T. Heckelei, J. Wery, J. Alkan Olsson, E. Andersen, I. Bezlepkina, F. Brouwer, M. Donatelli, G. Flichman, L. Olsson, A. Rizzoli, T. Van der Wal, J.E. Wien, and J.Wolf. 2008. "Integrated assessment of agricultural systems – A component-based framework for the European Union (SEAMLESS)." *Agricultural Systems* 96:150-165.

van Ittersum, M. K., P. A. Leffelaar, H. van Keulen, M. J. Kropff, L. Bastiaans, and J. Goudriaan. 2003. "On approaches and applications of the Wageningen crop models." *European Journal of Agronomy* 18:201-234.

Van Ittersum, M.K., R. Rabbinge, and H.C. Van Latesteijn. 1998. "Exploratory land use studies and their role in strategic policy making." *Agricultural Systems* 58:309–330.

van Wijk, M.T., P. Tittonell, M.C. Rufino, M.Herrero, C. Pacini, N. Ridder, and K.E. de, Giller. 2009. "Identifying key entry-points for strategic management of smallholder farming systems in sub-Saharan Africa using the dynamic farm-scale simulation model NUANCES-FARMSIM." *Agricultural Systems* 102:89-101.

van Wijk, M., M. Rufino, D. Enahoro, D. Parsons, S. Silvestri, R.O. Valdivia, R.O. and M. Herrero. 2014. "Farm household models to analyse food security in a changing climate. A review." *Global Food Security* (in press).

van Wijk, M., M. Rufino, D. Enahoro, D. Parsons, S. Silvestri, R.O. Valdivia, R. and M. Herrero. 2014. "Farm household models to analyse food security in a changing climate: A review." *Global Food Security* 3:77-84.

Vayssières, J., F. Guerrin, J.M. Paillat, P. Lecomte. 2009. "GAMEDE: A global activity model for evaluating the sustainability of dairy enterprises Part I - Whole-farm dynamic model." *Agricultural Systems* 101:128-138.

Vesk, P.A. and M. Westoby. 2001. "Predicting plant species' responses to grazing." *Journal of Applied Ecology* 38:897-909.

Waithaka, M.M., P.K. Thornton, M. Herrero, and K.D. Shepherd. 2006. "Bio-economic evaluation of farmers' perceptions of viable farms in western Kenya." *Agricultural Systems* 90:243-271.

Wallach, D., D. Makowski, J. W. Jones, and F. Brun. 2014. "Working with Dynamic Crop Models: Methods, Tools and Examples for Agriculture and Environment (Second Edition)." *Academic Press.* Waltham, MA.

Western, A.W., R.B. Grayson, G. Bloschl, G.R. Willgoose, and T.A. McMahon. 1999. "Observed spatial organization of soil moisture and its relation to terrain indices." *Water Resources Research* 35(3):797-810.

Wheeler, T., and J. von Braun. 2013. "Climate Change Impacts on Global Food Security." *Science* 341(6145):508-513.

White, J.W., and G. Hoogenboom. 1996. "Simulating Effects of Genes for Physiological Traits in a Process-Oriented Crop Model." *Agronomy Journal* 88:416-422.

White, J. W., L.A. Hunt, K. J. Boote, J. W. Jones, J. Koo, S. Kim, C. H. Porter, P. W. Wilkens, and G. Hoogenboom. 2013. "Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards." *Computers and Electronics in Agriculture* 96:1-12.

Wilkerson, G. G., J.W. Jones, K. J. Boote, K.T. Ingram, and J.W. Mishoe. 1983. "Modeling soybean growth for crop management." *Transactions of the American Society of Agricultural Engineers* 26:63-73

Williams, J.R., C.A. Jones, J.R. Kiniry, and D.A. Spanel. 1989. "he EPIC crop growth model." *Transactions of the American Society of Agricultural Engineers* 32(2): 497-511.

Williams, J.R., K.G. Renard, and P.T. Dyke. 1983. "EPIC: a new method for assessing erosion's effect on soil productivity." *Journal of Soil and Water Conservation* 38(5): 381-383.

Williams, J.R., C.A. Jones, J.R. Kiniry, and D.A. Spanel. 1989. "The EPIC crop growth model." *Transactions of the American Society of Agricultural Engineers* 32(2): 497-511.

Wilson, J.P., and J.C. Gallant (Eds). Terrain Analysis: Principles and Applications. Wiley. 2000.

Woodward, S.J.R. 1998. "Dynamical systems models and their application to optimizing grazing management." In: Peart RM, Curry RB (eds) *Agricultural Systems Modeling and Simulation*. Marcel Dekker Inc., New York, pp. 419-473

Worthington, E. B. (ed.). 1975. "The evolution of IBP." *Cambridge University Press*, Cambridge.

You, L., S. Wood, and U. Wood-Sichra. 2006. "Generating global crop maps: from census to grid." Selected paper, IAAE (International Association of Agricultural Economists) Annual Conference, Gold Coast, Australia.

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation

J. M. Antle, B. O. Basso, R. T. Conant, H. C. J. Godfray, J. W. Jones,
M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig,
P. Tittonell, T. R. Wheeler

Outline

Executive Summary

This paper presents ideas for a new generation of agricultural system models and data that could meet the needs of a growing community of end-users exemplified by a set of Use Cases. We envision new models and knowledge products that could accelerate the innovation process that is needed to achieve the goal of achieving sustainable local, regional and global food security. We identify desirable features for models, and describe some of the potential advances that we envisage for model components and their integration. We also discuss possible advances in model evaluation and strategies for model improvement, an important part of achieving our vision. We conclude with a multi-pronged implementation strategy that includes more thorough testing and evaluation of existing models, the development and testing of modular model components and integration, improvements in data management and visualization tools, and development of knowledge-products for end users.

1. Introduction

The idea of creating a new generation of agricultural system models and knowledge products is motived by the convergence of several powerful forces. First, there is an emerging consensus that a sustainable and more productive agriculture is needed that can meet the local, regional and global food security challenges of the 21st Century. This consensus implies there would be value in new and improved tools that can be used to assess the sustainability of current and prospective systems, design more sustainable systems, and manage systems sustainably. These distinct but inter-related challenges in turn create a demand for advances in analytical capabilities and data. Second, as discussed in the companion paper on *The State of Agricultural System Science*, we now have a large and growing foundation of knowledge about the processes driving agricultural systems. Third, rapid advances in data acquisition and management, modeling, computation power, and information technology provide the opportunity to harness this knowledge in new and powerful ways to achieve more productive and sustainable agricultural systems, as discussed in the companion paper on *Building an Open, Web-Based Approach*.

Our vision for the new generation of agricultural systems models is to accelerate progress towards the goal of meeting global food security challenges sustainably. In this paper and the companion paper on information technology and data systems, we employ the Use Cases presented in the Introductory paper, and our collective experiences with agricultural systems, data, and modeling, to describe the features that we think the new generation of models, data and knowledge products need to fulfill this vision. A key innovation of the new generation that we foresee would be their linkage to a suite of knowledge products – which could take the form of new, user-friendly analytical tools and mobile technology "apps" – that would enable the use of the models and their outputs by a much more diverse set of stakeholders than is now possible. Because this new generation of agricultural models would represent a major departure from the current generation of models, we call these new models and knowledge products "second generation" or NextGen.

We organize this paper as follows. First, we return to the Use Cases and identify key features that NextGen models require to meet the needs of the Use Case personas. Second, we discuss new approaches that could be used to advance model development that go beyond the ways that first generation models were developed, and in particular, the idea of creating a more collaborative "pre-competitive space" for model development and improvement, as well as a "competitive space" for knowledge product development. Then we describe some of the potential advances that we envisage for the components of NextGen models and their integration. We also discuss possible advances in model evaluation and strategies for model improvement, an important part of the approach. Finally, we consider both near-term and longer-term strategies for implementation.

2. Use Cases: Implications for Next Generation Models

We now discuss the implications of the five Use Cases for the development of second generation models and knowledge products. Table 1 summarizes their characteristics.

2.1 Farm Extension in Africa

Sizani is working as a farm extension officer in an area in Southern Africa where many farms are very small, incomes are very low, and farmers typically grow maize and beans as staple crops for their family's subsistence and to sell for cash. She needs to have access to analyses that help her tailor advice to specific farmers whose land and other endowments may vary considerably. Cropping system models are needed in this use case, either for Sizani to run through an application on her smart phone after keying in the location of the farm (to access soil, weather, market and other databases for the simulation analyses) or to access results that have been pre-run to provide best management options for the particular farming system. In either case, models of cropping systems and of the farm household are essential, in addition to databases that contain information on soil, weather, markets, diets, nutrition, and crop varieties that are available for the farmer to use. Ideally, the models would have been run ahead of time so that information could be provided to the farmer relative to costs of the new heat, drought, and disease-tolerant maize varieties and their performance in terms of grain yield and fodder production using current management practices. Also, the app would provide information on the benefits of rainfall harvesting in terms of increasing productivity and stability of yield across years when rainfall is limited. Sizani also has access to information from farm-scale analyses that take into account the labor available to the farmer and needed to implement and manage the rainfall harvesting approach, and the overall costs and benefits to the farmer.

Cropping and farming systems models and data are needed to produce the results for the smart phone application, and thus help Sizani deliver farm-specific advice to increase maize productivity and stability and to increase the economic and nutrition well-being of the farm family. The cropping system models are needed to simulate maize, beans, and vegetables that are produced by the farmer. In addition, the models need to take into account the benefits of using new varieties of maize and beans that are tolerant to high temperature and drought, since these are projected to increase under changing climate conditions. Furthermore, the crop models need to be able to simulate the effects of small increases in inorganic fertilizer as well as organic matter, and to simulate the effects of partially harvesting rainfall. A crop disease module is needed to simulate the effects of foliar diseases for susceptible and tolerant varieties. Together, these modules need to provide output information on grain, vegetable, and fodder production under alternative management systems. The farming system model will take into account the labor requirements, costs of adopting and managing new systems, and markets, providing information to the farmer on average benefits and risks of losses in particular years associated with climate variability and variability in disease pressure.

After the information is accessed by Sizani, she is able to inform the farmer of the economic benefits and risks associated with adoption of the new varieties, fertility management, and water-harvesting approaches on his farm. Sizani also informs the farmer of an app in the local language that can be accessed to learn more about the varieties, water harvesting, and nutrient management, and she leaves the farmer an Extension Fact Sheet that provides more general information, also in the local language, about these technologies and where to obtain them.

2.2 Developing and Evaluating Improved Crop and Livestock Systems for Sustainable Intensification

Xiaoming is a plant breeder/geneticist working on developing a drought- and heat-tolerant hybrid of maize.

Table 1. *Characteristics of Five Use-Cases*

	Use cases				
	1	**2**	**3**	**4**	**5**
	Farm Extension in Africa	Developing and evaluation technologies for sustainable intensification.	Investing in agricultural development projects that support sustainable intensification.	Management support for precision agriculture.	Supplying for products that meet corporate sustainability goals.
Farming System	small-holder	small-holder	small-holder	commercial corp	commercial corp
Information User	Farm advisor	Agricultural research team/program	Analyst/adviser	Management consultant	Corporate analyst
Beneficiaries	Farm family	Research institution/ farm population	NGO & clients	Farm business	Agri-business firm
Outcomes	Improved livelihood (income,nutrition, food security)	Improved technology	Sustainable technology	Income, soil conservation & water quality	Profit, risk management, sustainability objectives

She would like to be able to evaluate the potential adoption and impact of maize varieties with particular characteristics across the widely varying conditions in Africa.

A maize cropping system model is needed that has the capability to predict the benefits of the new drought and heat-tolerant maize varieties under the range of soils, weather, and management conditions across the regions of interest in Africa. Furthermore, a household economic model is needed to evaluate the adoption of the new maize varieties, resources (e.g., access to credit, labor, and fertilizer inputs) needed to produce the new variety. One question would be about the costs of purchasing the new variety, as well as the benefits and risks of growing it relative to traditional varieties. Therefore, information is needed on the household resources and constraints as well as information on the yield gains expected by switching to the new variety and the overall impacts on the economic livelihood of the farm family. Costs of inputs and likely prices of grain are needed for the economic model. Also, soil, weather, and management information are needed as inputs to run the crop and household models to evaluate the switch to the new variety.

The model-based analysis needs to take into account the risks associated with weather variability in the short term as well as responses to changes in climate that are projected for the longer term. Assuming that the farmer grows other crops for food/fodder and for sale and has livestock, models for these other enterprises are also needed. Ideally, the crop and household economic model would be used to perform simulation experiments, similar to how a randomized controlled trial might be performed if that were possible. Results from these simulation or optimization experiments would allow Debora to evaluate multiple factors, such as variability in maize grain and fodder yield, income, return on investment, and nutrition.

2.3 Investment in Agricultural Development to Support Sustainable Intensification

Carlos is an investment manager for a prominent Foundation, and he needs to evaluate a project for small farms in Kenya that will increase the intensity of production by increasing fertilizer use per hectare on cash crops while maintaining the current sustainable nutrient balance between pasture grasses, crop residues and animal manure. Carlos wants to evaluate whether the higher crop yields would induce a non-sustainable system once the initial period of fertilizer subsidies and extension was completed.

Given that the proposed project extends over a substantial area of many thousands of hectares, any analysis will have to be presented on the landscape basis. However it's equally important that the heterogeneity of the agricultural resource base, and thus the differing yields and potential fertilizer response, is also represented. To achieve this, an integrated whole-farm system model is needed with crop, livestock, economic and environmental components.

Sampling sets of regional parameters that can be representative of the landscape as a whole is necessary before implementing crop or livestock production models. The analyst is faced with balancing the accuracy of representation of the landscape against the proliferation of model runs and their associated expenses. This first step in project design requires careful summaries of the range of soils, altitudes, microclimates, and water resources systems in the whole area.

Since animal production is an integral part of the farming system, the livestock model should be integrated with the smallholder crop models. Ideally both livestock and crop models can be run simultaneously thus showing the nutrient flows between different production sectors and the sustainability of the system as a whole. The type of cropping system model used, will have year-after year carry-over of soil carbon, soil fertility, residue return, and use of both animal manures and inorganic fertilizer.

Samples of crop yields, input changes, and responses to fertilizer policy are generated by the integrated plant growth and livestock models. They are then used to populate the distribution of productivity and economic and social impacts from changes in fertilizer extension policy in the different regions sampled on the landscape. For this stage of analysis, Carlos can use an economic impact assessment model driven by an empirical distribution of characteristics across the landscape. The change in risks to different metrics such as income level, nutritional balance, and distribution of benefits across farm sizes is essential for assessment of the project.

To integrate the information from these three stages into the decision model Carlos needs a dashboard application that he can access from his laptop computer. Using this application, he can set up an assessment, enter data supplied with the project proposal, and link to general data layers available for the region. The dashboard provides a variety of ways for Carlos to visualize the model outputs and prepare them for presentations to his organization.

2.4 Management Support for Precision Agriculture in the US for Profitability, Soil Conservation and Water Quality Protection

Greg is among an increasing number of commercial growers with interests in Precision Agriculture. Despite the rapid advancements of sophistication and automation of farm equipment in recent decades, there is still a vital part of the equation that remains incomplete – the analysis of the vast amount of available data that gives farmers like Greg a map of what action to take where and when. Most of the variable rate application he and others currently rely on is based on rules of thumb and empirical approaches, as opposed to a systems approach that accounts for the interaction of soil, crop, management, and weather.

Process-oriented crop growth models simulate the effects of genetics, management, weather and stresses on the daily growth of crops using carbon, nitrogen and water balance principles. The strength of these models is their ability to account for stress by simulating the temporal interaction of stress on plant growth each day during the season. Thus, they tend to be sensitive to temporal patterns of stress. However, these models were designed for homogeneous areas, and as a result, inputs that are spatial in nature must be assumed to be uniform. Furthermore, spatial characteristics are often unknown or difficult and expensive to measure. The advent of Precision Agriculture has resulted in the need to extend the use of point-based crop models to account for spatial processes. Crop models can provide useful estimates of potential economic return for management recommendations, along with the sensitivity of a recommended management action in response to weather variability. The next generation of crop models for Precision Agriculture will account for spatially connected processes and use publicly available data on soil type, weather forecasts, along with location specific data from farmers' yield maps, to provide a prescriptive crop management plan on a very high spatial resolution.

2.5 Supplying Food Products that Meet Corporate Sustainability Goals

Sophia is an economic analyst in a corporate sustainability group. This group has embarked on efforts to make sustainability the core of their mission: marketing food while conserving resources. She is assessing the life cycle of food products to find ways to conserve energy, save water, minimize waste and reduce greenhouse gas emissions in an effort to make these products more sustainable from the farm to fork.

The system models needed to support supply chains in their pledge for sustainability are the same system models described in the precision agriculture user case. Crop system models are able to simulate the annual fluxes of N2O from soils under different pedo-climatic and management conditions rather well, but their performance requires improvements when simulating the daily fluxes of N2O. As N2O is directly linked to the amount of fertilizer used, the next-gen models will play a crucial role in identifying the optimal N rate that maximizes profits and reduces nitrous oxide emissions and nitrate leaching.

2.6 Implications for Second Generation Models and Data

Table 2 summarizes a number of agricultural system model features that are suggested by the Use Cases. These have important implications for the design of new-generation models and knowledge products.

- All of the small-holder use cases (1-3) require whole-farm models, and decision-makers in the commercial crop use cases (4 and 5) are likely to

Table 2. *Model and Data Features Implied by the Use Cases Defined in Table 1*

	Use Cases				
	Farm Extension	Improved Systems	Investment in Sustainable Intensification	Precision Ag	Sustainable Value Chains
System Features					
- single production activity	?	?		x	x
- multiple production activities	x	x			
- interacting activities	?	?		?	
- whole farm	x	x	x	?	?
Data (spatially referenced)					
- single activity	?			x	
- individual farm	x			x	
- representative sample		x	x		x
Outputs					
- bio-physical production (yield)	x	x	x	x	x
- economic (profit, income)	x	x	x	x	x
- environmental		x	x	x	x
- social	x	x	x		
Output Access					
- model		?			
- mobile app	x			x	
- computer dashboard	?	x	x		x
Spatial Scale					
- field	x			x	
- farm	x			x	
- region (many farms)		x	x		x
Temporal Scale					
- within-season	x	?		x	
- season	x	x	x	x	x
- multiple seasons	x		x		x

want whole-farm information as well, even if the specific use case (e.g., precision nitrogen application) does not require it.

- All cases need spatially referenced data, but the type and resolution of data required varies across the Use Cases.

- All of the Use Cases need biophysical production outputs and economic outputs. The need for environmental and social outputs is case-specific.

- Most, if not all, of the personas in the Use Cases would want to access model outputs via a dashboard application that would probably run on a laptop or larger tablet to facilitate visualization and integration of outputs with other applications and data, although some farm decision-makers or farm advisers might only want mobile applications.

- Only one of the Use Cases might want direct access to model output (the scientist Use Case 2).

- The spatial scale of the data needed varies by case, but all cases need season-specific data. Some farm-level uses will need within season data (e.g., for pest management or precision nutrient application).

There are several striking findings: All the Use Cases require whole-farm systems models; all require both biophysical production and economic outputs: all would need at least season-specific, spatially referenced data; and the farm-level decision makers are likely to need within-season data. Also striking is the fact that only one of the personas (the scientist) might want direct access to model output. An important group of model users not highlighted here is the science community itself that is developing and using models for research. Even among the scientists themselves, it is often the case that one user (say, an economist) does not require the output of another model (say, a crop model) in the form it comes out of the model, but would rather have the output put into a format suitable for further manipulation. As these use cases illustrate, this is even more so for non-scientist users: there are few if any users that require direct access to the model output.

From the companion paper on *The State of Agricultural Systems Science*, we know that few if any agricultural models currently available meet the needs of the five Use Cases. Few provide whole-farm analysis capabilities, for example, or make model outputs accessible through user-friendly web-based dashboard applications. Thus, we can conclude that there is a substantial gap to be bridged between current models and the capabilities needed to provide information that would be useful to most potential users.

3. Designing Next Generation Models

Given the gap between the current state of agricultural systems models and the needs of actual and potential users, this section discusses how the new generation of models can be created to bridge this gap and realize the vision for next generation models presented in the Introduction.

3.1 A Demand-Driven, Forward-Looking Approach

A first step towards realizing the potential for agricultural systems models is to recognize that until now, most model development has been motivated by research and academic considerations, not by user needs. This means that the model development community needs to turn the model development process “on its head” by starting with outcomes and working back to the models and data needed to quantify relevant model outputs. For example, the Use Cases show that in most cases whole-farm models are needed, and particularly for small-holder farms, models are needed that take into account interactions among multiple crops and often livestock. Yet, many agricultural systems models represent only single crops and have limited capability to simulate inter-cropping or crop-livestock interactions.

Another feature is that many models produce only estimates of biophysical quantities of crop or livestock production (e.g., crop yield) or basic economic variables such as net returns to a crop activity. Why? Models of single crops are easier to create, require less computational resources, and are driven by a smaller set of data than models of crop rotations, inter-crops or crop-livestock systems. But perhaps more importantly, researchers are typically responding to the incentives of scientific institutions that reward advances in science above all else. As a result, research on component processes within single crops may be more researchable within a laboratory or institutional setting, and may result in more publishable findings. The reward for producing useful decision tools for farmers or policy decision-makers may be minimal in academic settings.

While it is clear that model development should be much more driven by user needs, it is also important to recognize that science informs stakeholders, about what may be important, and about what may be possible. People know, in a general way, what they want, but they may have no idea what kind of science is needed to meet that need. Henry Ford famously said, "If you had asked people what they wanted, they would have said 'faster horses'." We can safely assume that potential users of agricultural systems models want a secure and sustainable supply of food. Who imagined even a few years ago that we might accomplish that goal, in part, by using data collected by aerial drones linked to agricultural systems simulation models?

So while model and data development need to be driven by user-defined needs, they must also be forward-looking, using both the best science and the imagination.

3.2 A Systems Approach

The Use Cases show clearly the need for whole-farm systems approaches. Agricultural systems are managed ecosystems (or agro-ecosystems) comprised of biological, physical and human components operating at various scales (e.g., cell, organism, field, farm). Farms are embedded within larger ecological and human systems operating at regional scales (e.g., watershed, population), as well as larger (continental, national, global) scales. It is typically important to consider many different interactions within and among these systems if we are to meet stakeholder needs for actionable outcomes.

The systems approach has several important implications for second generation models. Within each system level, a set of interacting sub-systems is involved. This suggests the possibility of constructing models of large, complex systems by combining models of modular sub-systems. The level at which modularization may be possible remains an important question, and this in turn has implications for software engineering. For example, as discussed in the companion State of Science paper, many crops are now modeled individually and separate from livestock. Systems with multiple interacting crops (e.g., through rotations or inter-crops), livestock, and crop-livestock interactions, are needed for various Use Cases, raising the question whether these interacting components can be incorporated in a modular "plug and play" system. Also, these biophysical production system components interact with economic-behavioral components and environmental components. These interactions among sub-systems imply the need for standard ways to link inputs and outputs among sub-systems.

Another important issue raised by the systems approach is the appropriate level of complexity for Use Cases, an issue discussed further in section 4.7. Research in environmental modeling indicates there are often diminishing returns to complexity. Similarly, experience with economic modeling has shown the value to "minimum data" or "parsimonious" approaches (Antle, Stoorvogel and Valdivia 2014). These ideas also relate to the need for a more generic approach (section 3.4 below).

The small-holder Use Cases illustrate the need for a systems approach at the farm level. In order to assess the well-being of the farm family in terms of income and nutrition, all relevant economic activities of the farm household need to be taken into consideration, including the income generated by the farming activities as well as other non-agricultural activities of the household members (e.g., off-farm work). Additionally, because the farm often involves multiple production activities, including crops and livestock, all of these activities and their key interactions need to be represented, as illustrated by the circular flow of nutrients from crops to livestock in the form of crops, crop residues and household waste fed to livestock, and then back to crops in the form of manure and composted materials.

The commercial-crop Use Cases also illustrate the need for a systems approach. Crop rotations are important to the management of soil fertility and soil pests, and thus play a key role in achieving more sustainable management of input-intensive systems. It is also likely that to improve the sustainability of large-scale systems, it will be necessary to move towards more diversified systems that use crop rotations and integrate crops with livestock. The commercial-crop Use Cases also illustrate the need for assessments of landscape-scale impacts, including water quality (through soil and chemical runoff and chemical leaching), biodiversity (through impacts of fish and other wildlife), and greenhouse gas emissions (e.g., through soil management and fertilizer use). Similar types of assessment are needed to design and evaluate systems that meet the goals of "sustainable intensification" and "climate-smart agriculture."

3.3 An Open, Pre-Competitive Space for Model and Platform Development Linked to a Competitive Space for Knowledge Product Development

Figure 1 presents a diagram of the linkages between the "pre-competitive space" of basic science and model development, and the "competitive space" of knowledge product development. The arrows between these two "spaces" point both ways to represent the inevitable and important give-and-take. There is a need for a demand-driven but forward-looking process that enhances interactions between these two realms. The concept of "pre-competitive space" grew out of the efforts of the pharmaceutical industry to collaborate on basic research while competing in product development. We think this distinction is also useful for thinking about how we might develop and apply agricultural systems models, while recognizing that there is also a competitive element among the researchers in the model development arena.

Facilitating a pre-competitive environment is likely to require innovations in the way research organizations operate, and may need to involve public-private partnerships (PPPs). PPPs are one way that science and industry can collaborate to generate new applied knowledge that can feed into the creation of new business and services. In PPPs it is common that both private and public partners provide funding and jointly formulate the research questions that can subsequently be tackled by research institutes and universities. There are a number of challenges in structuring PPPs. For example, in the European Union PPPs have been regulated to avoid unfair competition. The EU regulations stipulate that there always has to be more than one private partner involved and intellectual property rights of the knowledge developed (e.g., tools, models, articles, methods) belong to the research partner, which can then license the use to private partners for commercial purposes.

An important aspect for a NextGen community of practice is openness. Open here means: first, inviting and engaging others to join and become involved; second, being ready to jointly set priorities with a broader stakeholder community (i.e. research programming, private partners, policy partners, non-governmental organizations); and third, being transparent for scientific and public scrutiny of methods, tools and results through not-solely scientific venues. Only a few of the

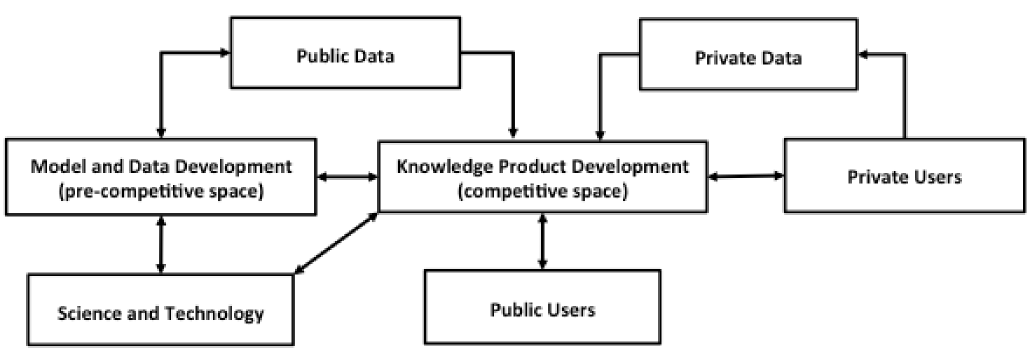

Fig. 1. *Possible Linkages between the Pre-Competitive Space of Model and Data Development and the Competitive Space of Knowledge Product Development*

agricultural systems models now in use can be said to be "open" in the sense that both the model equations and programming code are fully documented and freely available to the community of science. Establishing an open approach consistent with the principles of good science, including sufficient documentation and sharing of code to allow replication of results with reasonable effort, should be a priority of the practitioner community. Such an approach would facilitate model improvement through peer review, model inter-comparison and more extensive testing, new modes of model improvement and development such as crowd sourcing, and education of the next generation of model developers and users. Creating this open approach would also raise challenges related to incentives and intellectual property that would need to be addressed. The recent positive experience with the Agricultural Model Inter-comparison and Improvement Project (AgMIP; Rosenzweig et al. 2013), a new community of science dedicated to an open approach, suggests that researchers are now ready and willing to participate.

An open approach will also encourage the emergence of competing models and modeling approaches, rather than a single "super model." One dominant "super-model" could eventually emerge. But the only way to know that such a model is desirable is to allow a multi-model environment to flourish. We also expect to see alternative approaches emerge as modelers tackle challenging features such as representation of heterogeneity and dynamics and linkages across scales. For models to be tractable, tradeoffs have to be made, and an open approach is needed to facilitate the testing of alternative solutions.

There are important examples of recent efforts at creating a more open approach to agricultural model development. The bio-economic farm model FSSIM (Janssen et al. 2010) was made available as open source in 2010 after completion of its main project-related development and published with a license that allowed further use and extension. The open sourcing of the model was combined with training sessions, but this did not lead to spontaneous community uptake and large-scale development. The DSSAT crop modeling community is undertaking an effort to make its code open-source with the participation of more than 20 developers. The Global Trade and Analysis Project has provided extensive documentation of its model and data and allows user-modification of its standard model (Global Trade Analysis Project, 2014), and there is a large number of users of the model globally. The IMPACT model developed by the International Food Policy Research Center is publicly documented and available to other researchers (Rosegrant 2012). The TOA-MD regional model for technology adoption and sustainability assessment of agricultural systems is fully documented along with a self-guided learning course. Model code is available to "registered users" who have signed an end-user license agreement that requires acknowledgment of the developers and allows only research and other public-good uses (Antle, Stoorvogel and Valdivia 2014). There are now more than 500 registered users of the model, but only a few have shown interest in modifying or further developing the model independently, possibly because it is programmed in a language that relatively few researchers use.

To achieve the goal of demand-driven model development, it will be necessary to strengthen the linkages between the pre-competitive space of model development and the competitive space of knowledge product development. The current state of affairs appears to be that, on the one hand, the modeling community is strong on analytical capability but weak on linkage to user demand; while on the other hand, the developers of user-related farm-level products (e.g., providing data from mobile devices) are weak on analytics. Thus, there appears to be the opportunity for "gains from trade" by facilitating more interaction between the two communities. An important part of this interaction has to be to identify the key research that could enable better service delivery to knowledge-product users. Additionally, as emphasized in the Use Cases, there is a public good value to enhancing a broader community that can provide both data and analytics for public investment and policy decision-making. These issues are further developed in the companion paper on *Building an Open Web-Based Approach*.

3.4 New Approaches to Data Acquisition, Management and Use

The explosion in the availability of many kinds of data and the capability to manage and use it creates new opportunities for systems modeling at farm and landscape scales. Figure 2 presents an example of the possible types of private and public data that could be generated and used for both farm-level management (as in Use Cases 1, 4 and 5) and landscape-scale investment and policy analysis (Use Cases 2, 3 and 5). Some of these data would be generated and used at the farm-level, others would be generated and used for landscape-scale analysis to support investment decision-making and science-based policy-making. While farm-level decision making and landscape-scale analysis have different purposes, they both depend on two kinds of data:

- Private data: site- and farm-specific characteristics of the land and the farm operation, and the site- and farm-specific management decisions that are made. These data can be used to evaluate the farm's biophysical, economic, and environmental performance.
- Public data: weather, climate, soils, and other physical data describing a specific location, as well as prices and other publicly available economic data (note that not all public data is accessible).

A key question for the design of the agricultural knowledge infrastructure is how both types of data can be collected, managed and utilized efficiently and securely. Figure 2 is a design envisaged for a setting where farm decision-makers are able to utilize advanced decision tools that would be integrated with cloud-based data and computing resources. Although such tools may currently only be feasible in high-income countries, we expect they will become increasingly available throughout the world.

3.5. Credibility, Uncertainty and Model Improvement

A clear message from the NextGen Stakeholder Workshop was that model credibility is a key issue limiting the use of models for decision-making. In some areas of commerce where long-term projections are important, for example the insurance industry, there has been growing acceptance and use of quantitative climate models and impact assessment models. But for many decision-makers, ranging from farmers and agribusiness, to the development donor community and government, quantitative models remain an arcane and poorly understood part of science.

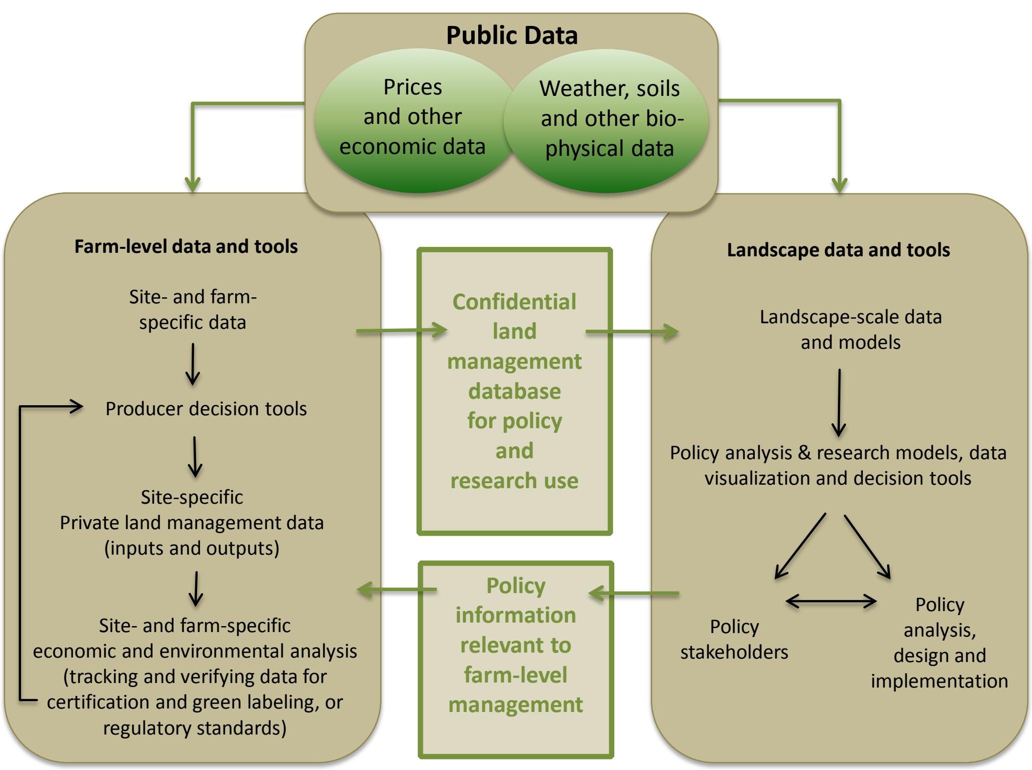

Fig. 2. *Possible linkages between data and decision tools at farm and landscape scales (source: Antle, Capalbo and Houston 2014).*

There are many aspects to establishing, maintaining and improving model credibility. First and foremost, models must be relevant to user's information needs. In addition, the participants in the Stakeholder Workshop emphasized the need to communicate what models are, what they can and cannot do, and to quantify and communicate model uncertainty effectively so that users understand how to use model outputs. But besides being relevant to users' needs, models must perform well enough to be judged credible and useful. As the companion paper on *The State of Agricultural Systems Science* shows, there are many short-comings of current models' capabilities that limit their relevance and usefulness for the Use Cases described here and the others discussed in the NextGen Stakeholder Workshop. Thus, achieving NextGen goals will involve developing better data and methods to evaluate model performance, both to help developers improve them and to help inform end-users about their validity and reliability.

There are potentially many different uses of models, from basic science to on-farm management to policy decision-making. The criteria for a "useful" model differ among these. For some science purposes, a high level of precision may be needed, whereas for policy analysis, the timeliness of the information produced may be much more important than its precision or accuracy. Thus model evaluation involves devising appropriate performance criteria, including overall model performance in providing outputs desired by end-users, as well as criteria for modules that can be used for component improvement.

Several types of formal model evaluation techniques have been developed to assess complex systems model performance under current as well as future conditions. Evaluation under current conditions can be based on comparison with observed data through numerical, graphical, and qualitative methods. An extensive survey of general classes of direct value comparison, coupling real and modeled values, preserving data patterns, indirect metrics based on parameter values, and data transformations is presented in Bennet et al. (2013). As explained by the authors, systems modeling requires the use and implementation of workflows that combine several methods, tailored to the model purpose and dependent upon the data and information available. A five-step procedure for performance evaluation of models is suggested, with the key elements including: (i) (re)assessment of the model's aim, scale and scope; (ii) characterization of the data for calibration and testing; (iii) visual and other analysis to detect poorly or non-modeled behavior and to gain an overview of overall performance; (iv) selection of basic performance criteria; and (v) consideration of more advanced methods to handle problems such as systematic divergence between modeled and observed values.

The evaluation of integrated models under future conditions cannot be directly assessed as available data may not be representative; this is particularly the case where the model includes an intervention that will change the behavior of the system. Instead, conceptual understanding of the system weighed against future projections can provide complementary lines of evidence in the assessment of the model. Global sensitivity and uncertainty analysis (GSA/UA) of future projections based on tailored scenarios provides a rich platform in the conceptual analysis of the models (Saltelli et al. 2004). GSA/UA provides a detailed understanding of the important factors and underlying processes driving the numerical model output variance under particular scenarios that can be compared with conceptual models of the system. The statistical techniques used also offer the opportunity to identify surprises in the future system behavior, as well as important feedbacks and non-linearities. GSA/UA also allows the modeler to: (1) examine model behavior (model check); (2) simplify the model based only on its important components; (3) identify important input factors and interactions to guide the calibration of the model; (4) identify input data or parameters that should be measured or estimated more accurately to reduce the uncertainty of the model outputs; (5) identify optimal locations where additional data should be measured to reduce the uncertainty of the model; and (6) quantify the uncertainty of the modeling results (Saltelli et al. 2005).

Another approach to model improvement that has been pioneered in the climate modeling field is intercomparison of models, implemented through the establishment of the Coupled Model Intercomparison Project (Taylor et al. 2012. By establishing protocols for the use of "reference scenarios" it is possible to inter-compare model results, identify important differences in model outputs, and through this process ultimately improve the models and their performance relative to the criteria described above. The use of model "ensembles" is also considered by some researchers as a way to characterize model uncertainty, although this interpretation is controversial. Model ensembles have been shown to perform better in some respects than individual models, suggesting the use of ensembles as a way to improve performance (Martre et al. 2014). A limitation of the ensemble approach is that it requires a relatively large number of alternative, independently developed models. In many cases, there are not enough distinct models to make model inter-comparison or ensemble approaches useful.

4. Potential Advances in Model Components

We next present examples of potential improvements that are important and may be achievable in the disciplinary components of agricultural systems models. We begin with a set of cross-cutting issues that are common to all of the model components, and then focus on disciplinary themes.

4.1 Cross-Cutting Issues

4.1.1 Representing and Incorporating Human Behavior into Agricultural Systems Models

Agricultural systems are managed by people for people. The objectives of the people using the information generated by models, and the behavior of decision makers whose behavior is represented in models, must influence model design. Most existing models have a limited capability to represent economic or other behavioral motivations of decision makers. This is a cross-cutting theme in modeling because the management decisions made by farmers related to crop and livestock productivity as well as to economic costs and returns as well as environmental and social outcomes. There are several ways that behavior needs to be incorporated into NextGen models.

First, a better understanding of decision maker objectives is needed if we are to develop models that provide information to farm managers to improve decision making. For example, if production risk management is an important objective of decision makers, then they will need different kinds of information than if production risk is not a major issue. Thus, modelers need to know what managers think are the major production risks. Note, however, that in this case the actual behavior of the decision makers does not need to be modeled. The goal is to inform decision making, not to make decisions for farmers.

Second, for models that are designed to predict or project plausible outcomes or impacts of decisions made by farmers, the behavior of the decision makers must also be modeled. This need adds a large amount of additional complexity above and beyond the capability of modeling bio-physical production processes. Knowing what behavioral models are most useful for the Use Cases (e.g., profit maximization, risk management, achieving social status, other social or environmental objectives) is a key issue that needs to be addressed in NextGen model development.

Third, the social dimension of farmer decision making needs to be better understood and represented in models, including how social interactions influence decision making. Agent-based models incorporate interactions among "agents", i.e. farmers, but lack a rigorous foundation for the rules that govern behavior. Modeling social interactions is an active area of economic research, but data demands are high and as yet empirical generalizations that could be used to structure models are not available. Other social scientists also study social interactions, but typically using qualitative methods that also are difficult to translate into quantitative models.

4.1.2 Representing Heterogeneity

A key fact that has emerged from the increasing availability of field- and farm-level data is the high degree of biological, physical, economic and social heterogeneity of agricultural systems, in both space and time. The farms represented by the use cases demonstrate this point: among smallholder maize-based farms in Kenya, for example, coefficients of variation of key characteristics like farm size are on the order of 100% or more; for commercial crop farms in the United States, they are also large, ranging from 50-150%. This heterogeneity has several important implications for how we represent agricultural systems in models:

- Accurate representation of bio-physical processes (e.g., crop growth, chemical leaching, erosion, chemical runoff) requires site-specific data (i.e., soils, slope, weather, management).

- Accurate representation of economic and social processes and outcomes (e.g., income and nutritional and food security) requires person- or household-specific data.

- Modeling environmental outcomes requires an integrated treatment of bio-physical processes and farmer decision making processes, and thus consistency between the spatial and temporal units of both types of processes.

- Representation of important economic, social and environmental outcomes at scales relevant to investment or policy decision making requires population-level outcomes that can be expressed not only as means or averages but more generally as distributions of outcomes. Only then is it possible to use indicators based on threshold concepts to represent vulnerability (e.g., poverty rates, risk of food insecurity, environmental risk, etc.).

- Behavioral heterogeneity is recognized in economics as one of the most important but also methodologically challenging aspects of modeling, because decision maker characteristics (e.g., experience, capability, motivations) are difficult to quantify and typically unobserved by the analyst.

4.1.3 Representing Dynamics

Agricultural systems are inherently dynamic. For example, crop growth occurs over time within the growing season, and crop productivity across growing seasons depends on crop rotations and other dynamics of the system. Most bio-physical system component models (crop growth, livestock growth, environmental processes) are inherently dynamic, but can only represent heterogeneity to limited degrees. Economic behavior depends on expectations of future outcomes, and decisions are made sequentially, with information being acquired as decisions are made and realizations are observed. Some management decisions like fertilization rates are based on intra-seasonal processes (getting the highest profit that season); other longer-term decisions span multiple growing seasons (multi-season crop rotations; machinery investments; livestock purchases and sales, and perennial crop planting and management decisions). Similarly, it is challenging to represent both dynamics and heterogeneity in economic models, and most dynamic models are highly simplified or stylized. The challenge is even greater when multiple dynamic model components are linked, due to differences in spatial and temporal units and overall model complexity.

Dynamics are often critically important at the regional scale. Witness, for example, the impact of weather shocks on regional food prices in some parts of the world, particularly in less-developed regions. Progress in modeling system dynamics is thus essential. How to achieve this progress in a tractable and useful way should be a priority for NextGen research.

4.1.4 Pathway and Scenario Design

Everything that influences an agricultural system, whether at the field, farm or regional scale, cannot be modeled. Consequently, most modeling is based on a logical structure in which some factors (“drivers”, or exogenous variables) take on values specified by the modeler or the model user. How these drivers are set or modified to represent the conditions under which the analysis is being carried out is a key aspect of modeling that has been under-studied. The issue is now receiving more attention in climate research (cite Moss, SSPs), but needs to receive more attention from the model development community. In particular, if models are to be linked to end-users through knowledge products, the user needs to understand the context in which the analysis or “simulation experiment” is being conducted. There has been little attention paid to how end-users could define or select those conditions or assumptions in which the modeling is carried out. These issues relate directly to the considerations of relevance and credibility discussed above.

For agricultural systems modeling AgMIP has been developing more systematic approaches to development of "pathways" (plausible future conditions) and "scenarios" (specific parametric representations of a system consistent with a pathway), using the concept of Representative Agricultural Pathways (Valdivia et al. 2014). Further work is needed to better develop these methods for use at farm, regional and global scales.

4.2 Crop Systems

Next steps in developing next-generation crop models fall into several categories: significant improvements in simulation of important crop processes and responses to stress; extension from simplified crop models to complex cropping systems models; and scaling up from site-based models to landscape, national, continental, and global scales.

4.2.1 Key crop processes that require quantum leaps in improvement

Several crop processes require major advances in understanding and simulation capability in order to narrow uncertainties around how crops will respond to changing atmospheric conditions. Experimentalists and modelers need to work together from the outset to ensure that the right research questions are posed as experiments are planned, critical field data are gathered at appropriate times, and process-based understanding is captured so as to transfer new insights from the field to the crop models directly and expeditiously.

Genetics. Developing predictive capacity that scales from genotype to phenotype is challenging due to biological complexities associated with genetic controls, environmental effects, and interactions among plant growth and development processes. Crop model improvements are needed to link complex traits at gene network, organ, and whole plant levels. Phenotypes are linked to changes in genomic regions via associations with model coefficients (Hammer et al., 2006).

Carbon, temperature, water, and nitrogen. Crops are already experiencing higher levels of carbon dioxide (CO_2) and temperature in agricultural regions around the world. Understanding of how accelerated rates of CO_2 and temperature rise will interact to affect crop growth and productivity is growing, but this improved understanding needs to be incorporated into crop models (Leakey et al., 2009). Water relations of soils and crops are also of perennial importance and carbon-nitrogen cycling plays a crucial role in sustainable intensification. The simulation of all of these processes and their interactions and management, especially under conditions of stress, needs to be radically improved.

Ozone. The magnitude of ozone damage is expected to be comparable to climate change in the next several decades, but ozone damage is rarely considered in crop modeling studies (Leisner and Ainsworth, 2012). Information about ozone impacts on crop yields is available, but damage processes and functions need to be developed. Model improvements in regard to ozone effects on crops include inclusion of ozone response functions and comparison of response functions with process-based approaches such as leaf conduction, aerodynamic boundary-layer resistance, and whole canopy conductance parameterizations. In order to learn much more about the different responses of different crop species and varieties, ozone data collection should be incorporated into the AgMIP protocols for sentinel sites experimental design.

Nutrition. Crop modelers, breeders, physiologists, and human health and nutrition researchers need to broaden the scope of modeling to include key nutritional processes and future risk of hunger. This requires moving from a yield-only perspective to one that includes processes that affect nutritional quality. Non-staple crops, for which crop models have not been developed, are likely to become increasingly important (Müller et al. 2014).

4.2.2 Extension from 'crop models' to 'cropping system models'

The field of crop modeling has been built on a single crop-by-crop approach. It is now time to create a new paradigm, moving from 'crop' to 'cropping system.'

Intercrops and complex rotations. A first step is to set up the simulation technology so that modelers can rapidly incorporate multiple crops within fields, and multiple crops over time. Then the response of these more complex cropping systems can be tested under different sustainable intensification management strategies utilizing the updated simulation environments. Similarly, studies can be performed to determine optimal cropping systems and management strategies for particular desired outcomes.

Pests, diseases, and weeds and their management. Diseases, pests, and weeds (DPW) are important yield-reducing factors in terms of food production and economic impact, and pose significant simulation challenges due to complex processes that occur over fine temporal but broad spatial scales. For each crop species, there is a portfolio of diseases, pests, and weeds, interacting over a range of time and space scales. Model improvements for DPWs include developing process-based models for important diseases and vectors, frameworks for coupling air-borne diseases to crop models, gathering significantly more data on crop impacts, and enabling the evaluation of pest management strategies.

Linkages to livestock production. Most smallholder farming in the world involves integrated crop-livestock systems that cannot be represented by crop modeling alone. Thus, next-generation cropping system models need to include key linkages to livestock. Livestock linkages to be incorporated include growth and productivity models for grasslands and rangelands as well as the usual annual crops. Information from local experiments (such as the AgMIP sentinel sites) will be required to develop and test the grassland and rangeland models. These models will then be capable of deployment with livestock models, regional farm data, and inputs on management and climate. On the management side, the effects of animal labor need to be included as well.

4.2.3 Scaling up from field scale to landscape scale

Cropping system models need to be able to simulate easily a diverse set of farms rather than just one representative farm, as has been common practice in the past. There are several approaches for scaling up, including use of gridded models and development of simpler quasi-empirical models for landscape-scale analysis (Lobell and Burke 2010). Large-scale computation can allow for much more extensive use of gridded models than in the past (Elliott, Kelly, et al. 2014). Soils and climate input datasets become important as simulation goes from field to landscape scale. There are several types of dynamic process gridded crop models: those developed from the site-based models such as DSSAT and APSIM; ecosystem-based models; and dynamic land-surface models. An example of a more statistical model is the agroecological zone (AEZ) approach developed by IIASA and the FAO (Fischer et al. 2002).

4.2.4. Crop Model Interoperability and Improvement

A key question for the next generation of cropping system models is the degree of interoperability. Historically, scientists (as individuals or groups) tended to have exposure to, and in-depth knowledge of, a single crop model (Thorburn et al. 2014). The Agricultural Model Intercomparison and Improvement Project (AgMIP) aims to increase efficiency of model improvement and application by sharing information between different models and encouraging the use of multiple models in impact assessment (Rosenzweig et al., 2013). Ideally, parameters from one crop model can be uploaded into databases and then downloaded, reformatted for use in another model. However, AgMIP has found that this sharing of parameter values between models is not necessarily straightforward.

The AgMIP Program is bringing different modeling groups together to compare and thus improve their models. The aims are to develop a better understanding of different crop models across the agricultural modeling community; improve both individual crop models and the entire group of models for a particular crop; and improve the efficiency and effectiveness of multi-model applications in agriculture.

4.3 Soils and Precision Management

Integrated agricultural technologies, defined as the integration of improved genetics, agronomic input, information technology, sensors, and intelligent machinery, will play a pivotal role in agriculture in the years to come. These innovations will be driven by economic forces, by the need to produce more food with limited land and water for the increasing population, and at the same time by the push to save resources to reduce the environmental impact associated with food production. While these changes are occurring now in the commercial-scale industrialized agricultures of the world, many of these technologies have the capability to be adapted to conditions in other parts of the world. The cell phone now allows farmers in rural areas almost everywhere in the world to have low-cost information about prices, for example. Similarly, it is likely that unmanned aerial vehicles will rapidly be adapted to conditions around the world and used to carry out activities such as monitoring crop growth and pest occurrence, and improve management decisions. In large-scale, capital-intensive agricultural systems, these technologies are rapidly leading to the automation of many production activities, particularly machinery operation and decisions about input application rates.

The automation of agriculture began in the mid-nineties, resulting in large amounts of data available to farmers and agribusiness companies. Farm machinery and tools sold today are largely equipped with high precision global positioning system (GPS) driven controllers, which allow all activity on the farm to be recorded, geo-referenced, and stored on remote computers: "in the cloud." All modern tractors collect data on a continuous basis and are equipped with wireless connectivity for data transmission. Harvesters record the yield at a particular location, planters can vary the plant spacing or type of seed by location, and sprayers can adjust quantity and type of fertilizer, fungicide or pesticide by location; all to a granularity of just a few square meters. Yield monitoring can now be linked to UAV imagery to produce a prescription map for the farmer to implement. These private data could also provide tremendous benefit to the researcher community, should access be increased.

Producers in the developed world now have historical crop yield data for their fields, at a few square meter resolution, for the last twenty years. Combined with advanced satellite-based imagery, high-resolution spectral and thermal data obtained from unmanned aerial vehicles (UAVs), and weather forecasts, growers have most of the critical inputs required to convert this "big data" into a proper actionable management plan that allows for the application of inputs to vary spatially within the same field. Despite these rapid advances in the sophistication and automation of farm equipment, a vital piece of the equation is still lacking: the analysis of the vast amount of newly available data in order to provide the farmer with a map of what action to take where and when. Most variable rate application is currently managed by farmers, using rule-of-thumb and empirical approaches, and not by using a systems approach that accounts for the interaction of soil, crop, management, and weather. Thus much of the power of automation remains unexploited.

In order to realize the full potential of more sophisticated equipment, new modeling systems for precision agriculture are needed. These systems could be based on comprehensive predictive crop yield models that combine publicly available data, such as soil type, weather, and others, along with location-specific data from farmers' yield maps of their fields, to provide a prescriptive crop management plan at high spatial resolution, as in Figure 2. This type of system could deliver automated crop simulations, crop management strategy recommendations, process-based variable rate prescriptions, risk assessments, continual in-season simulations, integration of in-season crop scouting UAVs flight information, pesticide/fungicide/herbicide prescriptions and accurate harvest recommendations via simple-to-use apps, websites, and/or smart phone texting.

The NextGen system will help farmers in two primary ways: better yields and higher profit margins. The ancillary benefits of improved compliance with environmental mandates and better stewardship of natural resources are also important motivators.

In this scoping study we have chosen purposefully to limit our scope to the farm and landscape scale. Increasingly, there will be demand for agricultural systems models to simulate and integrate the different components of the agricultural value chain (Fig. 3). Genetics, agronomic management (production input), weather, soil, information technology and machinery need to be linked in a system approach.

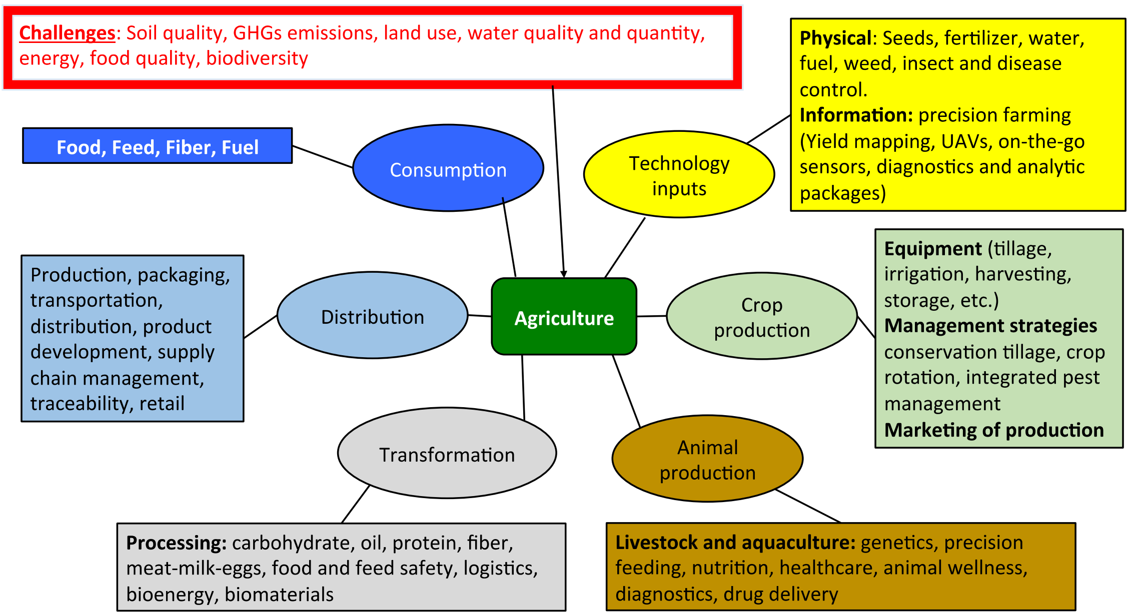

Fig. 3. *Schematic representation of the agricultural sector value chain.*

4.4 Pests and Diseases for Crops and Livestock

As noted above, a major limitation of existing models is how they represent pests and diseases. We expand here upon some of the important areas that must be improved in NextGen models.

Improved statistical modeling of within-season pest and disease threats using automated data collection and cloud computing. It is now possible to collect weather data continuously from ground-based sensors and to merge these data with medium-term weather forecasts and remote sensing data on crop growth and pest and disease damage. (Both growth and damage can be detected by satellite or drone by monitoring the crop's spectral properties.) Then, using sophisticated statistical modeling done centrally, real-time advice can be distributed to farmers through the web or through mobile phones enabling them to take precautionary actions.

Understanding the consequences of climate change for weed, pest and disease threats. The IPCC has reviewed the existing evidence for how climate change may affect weeds, pests, and diseases. One issue with this evidence base is that there is a clear publication bias towards reports of increased threats – people often do not bother to write up no-effect results. There is a general recognition that we need good models to help tease out the different effects that changing weather will have simultaneously on both crops and the organisms that compete with or attack them. There has already been some work applying crop physiology-type models to weeds, and developing more mechanistic models of the effect of temperature on insect pests. There is an opportunity and need for more integrated models that include interactions between organisms, for example between weeds and crops, and between pests and the predators and parasites that attack them. A variety of different approaches are possible, and there is a need for an AGMIP-type approach to help the community decide how best to move forward.

Livestock disease. Highly contagious diseases of livestock present a major threat to agriculture, both in the developed and developing worlds. Diseases may be chronic in livestock populations, emerge from wildlife reservoirs, or possibly be introduced deliberately by man as an act of bioterrorism. Models are required to help understand how a disease will spread, and to help policymakers design optimal interventions. These models must encompass not only the epidemiology of the disease but also how it is affected by agricultural practices and in particular the movement of livestock by farmers. There have been significant recent advances in this area, often building on work on human diseases. For example, it is now possible to take livestock movement data and use it to parameterize an epidemiological model (Kobayashi, Carpenter, Dickey and Howitt 2007; Brooks-Pollock, Roberts, and Keeling 2014). There are the beginnings of a model comparison movement in human epidemiology; livestock disease epidemiology would also benefit from this approach.

Novel genetic control methods. There is intense current research activity into novel genetic methods of insect control. Most of this work, much funded by the Gates Foundation, is currently directed at the insect vectors of human diseases such as malaria, though the same methodology can be applied to insect pests of crops and of course the vectors of livestock diseases. The greatest advantage of these approaches is that they involve self-sustaining interventions that spread naturally through a pest population, although because they are nearly all classified as genetically modified, the regulatory issues surrounding them are complex. Cutting-edge modeling work in this field involves joint population and genetic dynamic models, many of which are explicitly spatial. This topic is likely to be one of the most important and exciting areas of modeling as applied to agriculture over the next few decades.

4.5 Livestock Production

There are a number of areas in which advances in livestock modeling could improve the information needed to support the Use Cases identified in Box 1, for farm-level and landscape-scale decisions.

For farm-level decision support:

More comprehensive livestock models covering a wide diversity of ruminant species, adequately pre-parameterized for most common situations and with default values for users to parameterize models to their conditions.

Summary models from comprehensive, dynamic models for on-farm support. This work includes summary models for intake, production and greenhouse gas emissions calculations. Some of these summary models could be developed as mobile phone technologies.

Development of extensive, standardized feed libraries linked to a GEO-WIKI for improving our mapping of feeds globally, but also to build a library that then can be used for deriving functions of feed quality for different agroecological conditions. One way this could be accomplished would be to expand existing household data collection protocols to include suitable data for livestock.

For regional investment and policy analysis:

Development of high resolution improved crop and livestock production systems typologies. These typologies could be derived from existing farm household, agro-ecology, farm, rangeland, population, markets and other spatial data. NextGen production systems mapping needs to include intensification, gender dimensions of family labor and control over assets and income, and operation size indicators.

Spatially explicit standardized feeds and productivity data. Ideally these data would be linked to crowd-sourcing and large data rescue initiatives.

Standardized linkages to global integrated assessment and economic models of different types (from Globiom, IMPACT to GTAP and others).

Improved spatially explicit farm and regional data on production costs for different livestock technologies. This information is seldom available and is crucial for both regional and global analyses.

Livestock yield gap analysis. A much deeper and better quantified bio-spatial analysis of livestock yield gaps is needed to guide investments and to identify opportunities to use livestock as a vehicle for agricultural development, poverty reduction, and environmental protection.

Livestock scenarios. Improved and consistent story-lines are required for the livestock sector in all scenarios. These story-lines can be produced as part of global and regional "representative agricultural pathways" being developed by AgMIP and other research teams. (Currently, such story-lines exist only for the global "shared socio-economic pathways" used in climate impact assessments; see Havlik et al 2014; Herrero 2014.)

4.6 Pastures and Rangelands

Pastures and rangelands are integral to all livestock production systems and are often closely integrated with crop production systems (e.g., pasture in rotation). The biophysical components of these systems and driving data required to model them are largely similar to those of crop production systems (see first chapter), but management data tend to be sparsely available and representing continuity of plant populations is challenging. Advancing our ability to understand how grasslands are managed – to understand, for example, what species are planted, what inputs (irrigation, fertilization, etc.) are provided, what grazing management (timing, intensity) is applied – is centrally important for improving our ability to model pasture and rangeland systems. At the same time, we have identified several features of next generation models necessary to improve the utility of models for pasture and rangeland systems, as we now discuss.

Planted pastures and native grazing lands both contain a variety of species, some of which are more palatable, nutritious, grazing-resistant, or fire-resilient than others. A more open, data-rich environment could facilitate evaluation of a variety of approaches for representing long-term dynamics, which could address several important grassland management/assessment issues. Managing grass swards (and desirable forb and species) to maintain desirable plants is a primary goal of grassland management, but one for which modeling tools have offered limited assistance. Models that represent vegetation dynamics are also desirable for understanding longer-term changes in species that can impact productive capacity, sensitivity to degradation, and carbon dynamics (particularly woody encroachment). Year-to-year variability is a key component for understanding potential utility and risk of relying on grassland forage resources. Next generation models that enhance our ability to forecast this risk would mark a substantial and meaningful advance.

The primary use of forage resources is for grazing animals, yet most grassland models are only loosely coupled with grazers (livestock or wildlife). Better integration between grassland and livestock models – through grazing effects on grasslands, grazer distributions across landscapes, forage demand/consumption, livestock/wildlife movement, etc. – would enhance the ability of models to contribute to important emerging issues. For example, holistic grazing management, in which several aspects of management vary in response to a variety of different cues from the land and expectations about future conditions, can be impossible to evaluate with current modeling frameworks. A system that integrated user demand into the model development process could lead to implementation of new data-management feedback loops within models. Such interactions between users and producers of information could direct data collection (e.g., by drone or remote sensing) to facilitate model use. Models that better represent grazer-grassland interaction are also crucial for understanding how efficiently livestock use forage resources , what is necessary to sustain wildlife populations, and how much grassland output might be available for other uses (e.g., biofuels).

4.7 Economics

Areas in which advances in economic modeling could improve the information needed to support the Use Cases identified in Box 1 also correspond to farm-level and regional decision support.

4.7.1. Farm-level decision support

Advanced analytics need to be coupled with the data on management decisions that are becoming available through mobile technologies (e.g., tracking soil conditions, seeding and fertilizer application rates, pesticide applications) and their results (e.g., crop growth, yield). An example of this analytical capability is the AgTools software developed by several university extension programs, which allows managers to calculate short-term profitability and rates of return on long-term investments (www.agtools.org). Similar proprietary software tools are being developed and used. These analytical tools could be linked with modules that track or predict environmental outcomes such as soil erosion and net greenhouse gas emissions (e.g., AgBalance by BASF). Low-bandwidth versions of these tools need to be developed for use in areas where mobile phone technology is a limiting factor. Analytical tools need to be adapted to fit small-holder systems.

Dynamic Estimation and Learning. The flood of data on physical land-use, water availability and use, and yields coming from mobile devices and remote sensing systems suggest that both the biophysical and behavioral aspects of farm production at specific locations can be estimated by sequential learning processes. Recent advances in numerical approximation to dynamic estimators have reduced the dimensional and computational restrictions on their use. Two of several approaches that seem practical for remotely sensed data sets are Ensemble Kalman filters which use numerical sampling approaches to avoid inverting large matrices, and Cross Entropy filters that use the Kullback formulation to reduce the Bayesian solution to a nonlinear finite optimization problem. These recent advances in remote sensing are evident in analysis of the impact of the 2009 and 2014 droughts on California agriculture, which demonstrated the advantages of better data (Howitt, Medellin-Azuara, MacEwan, Lund, and Sumner 2014).

4.6.2 Regional investment and policy analysis

Modularization and input standardization. Models need to be incorporated into modules with standardized inputs and outputs, including farm-level optimization models, regional positive quadratic programming models, econometric land-use models, and regional impact assessment models. With this investment, these models could then be coupled more effectively for landscape-scale and population-level analysis of technology investments and other policy analysis.

Model linkages across scales. Methods and protocols are required to link regional economic models (price-taking land use and impact assessment models) with market equilibrium models (e.g., regional partial or general equilibrium models). Some progress has been made on this front but much more development is needed (Antle, Stoorvogel and Valdivia 2014).

Richer characterization of behavior. Generalization of behavioral assumptions and investigation of their effects on investment and policy analysis. Most economic models make simple profit maximization assumptions. There is a rich literature on risk modeling which could be incorporated. Recent advances in the expectations formation literature and the behavioral economics literature could be investigated for use in agricultural systems models.

4.7 Environment and System Complexity

Current agricultural system models typically operate at the point/field scales (Fig. 4a) with an emphasis on vertical fluxes of energy, water, C, N and nutrients between the atmosphere, plant and soil root zone continuum. A holistic upscaling from the point source to the landscape scale (Fig. 4b) requires incorporation of several interacting, complex components, adding substantial complexity above and beyond the agricultural system itself. Thus, a major consideration in environmental modeling is how to best capture essential interactions while maintaining models that are feasible to implement with available data and computational resources.

Figure 4 illustrates the various components linking point to landscape scales. A first element for the linkage from point to landscape is estimation of surface and subsurface fluxes and ecological transitions along the lateral scale. Coupling with landscape microclimate models provides the vertical inputs used by the agricultural systems models, as well as gradients (precipitation, temperature, wind, vapor pressure deficit) along the landscape. Coupling with hydrological models provides water flow paths like surface runoff, vertical and lateral groundwater flow, and interactions between vadose and groundwater zones and with adjacent surface water bodies (channels, rivers, lakes and coastal waters). Water quality models provides sediment and solute transport along the landscape controlled by water flows (Fig. 4b), and other effects like wind erosion. Integration and upscaling of landscapes into the watershed scale (Fig. 4c) requires 3-dimensional coupling of the surface and subsur-

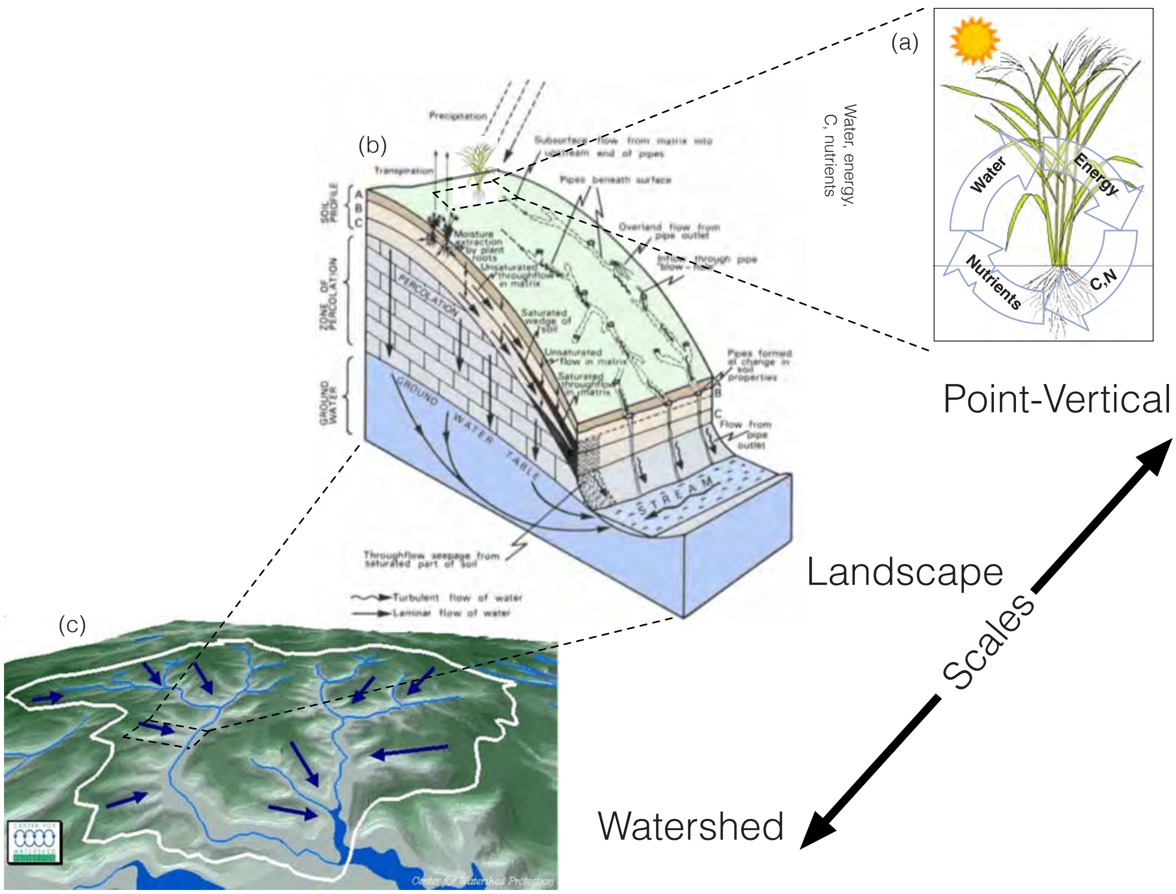

Fig 4. *Lateral connections across scales with other environmental components needed in the next generation agrisystems models, from (a) point vertical scales typical of current agrisystem models, (b) lateral hillslope/ landscape surface and subsurface energy/water/C,N/nutrients transfers and ecological and human interactions (adapted from Kirby, 1976), to (c) watershed and regional surface and subsurface connections and teleconnections.*

face water, energy and mass transfers. At this scale, the groundwater aquifer system typically transcends the boundaries of the watershed and necessitates analysis at the regional scale to evaluate not only the impacts of the cropping and animal production systems on water quantity and quality, but also feedbacks from the hydrological system in the agricultural system (shallow water table effects, drought or low water availability for irrigation). Further, mesoscale rainfall and evapotranspiration distribution models control the local surface and subsurface flow intensities, pollution and abatement. At this scale, human effects through land-use changes as well as ecological (vegetation, wildlife) dynamics and transitions on natural or protected lands (riparian zones, conservation areas, water resource management infrastructure etc.) are also an important and critical component to evaluate the overall sustainability of the agricultural system.

It is important to recognize that although current crop modeling upscaling approaches based on land use maps are an efficient first approximation, the next generation models should consider the lateral connections through the landscape and regional scales to evaluate the sustainability of the integrated system, including effects on water and soil resources quality and quantity and ecological value.

The complexity resulting from the proposed integration at a landscape scale cannot be understated. In particular, additional emphasis is urgently needed on rational approaches to guide decision making through uncertainties surrounding the integrated agricultural system across all scales. As with all models (Raick et al. 2006; Kotz and Dorp 2004), those predicting agricultural production changes and interactions with the coupled natural and human components produce unavoidable uncertainty around the predicted responses. These two issues – the need for coupled models that can answer the pertinent questions and the need for models that do so with sufficient certainty – are the key indicators of a model's *relevance*. Model relevance is inextricably linked with model complexity.

Although model complexity has advanced greatly in recent years and is a natural outcome of the proposed next generation integrated modeling, there has been little work to rigorously characterize the threshold of relevance in integrated and complex models. Formally assessing the relevance of the model in the face of increasing complexity would be valuable because there is growing unease among developers and users of complex models about the cumulative effects of various sources of uncertainty on model outputs (McDonald and Harbaugh 1983; Manson 2007; Cressie et al. 2009; Morris 1991). New approaches have been proposed recently to evaluate the *uncertainty-complexity-relevance modeling trilemma* (Muller, Muñoz-Carpena and Kiker 2011).

Due to the complexity of the coupling process needed in the upscaling and integration processes, innovative approaches to simplify model outcomes to make them relevant in decision-making will be central to the next generation modeling efforts. New methods for evaluating uncertainty also can be used to devise model simplification strategies. For example, the identification of non-important processes for particular scenarios might lead to their removal or fixing (variance cutting) without affecting the overall results while reducing the overall output uncertainty. The identification of the important model factors and the output response surfaces obtained from the analysis for particular scenarios can inform meta-modeling efforts, were simplified functions or databases of the model outputs are used in place of the full model for decision analytics (Ratto et al. 2007; Villa-Vialaneixa, N. et al. 2011; Ruane et al. 2014).

4.8 Social Dimensions

As noted in section 3, a demand-driven approach is needed that begins with user-selected outcomes. Various outcomes are of interest in the context of sustainability. Here we identify some key outcomes that need to be incorporated into modeling approaches.

Income distribution and poverty. Most economic models provide an estimate of some components of income, but a complete characterization of income sources is needed to evaluate income distribution and poverty. Population-level outcomes are needed, not only means or averages.

Food and nutritional security. Existing models represent food production, but no existing model characterizes all factors that affect food security (availability, access, stability, utilization) at the household or regional levels. A major limitation is data on food consumption at the household and personal levels over time. New methods of collecting these data using mobile devices are being developed. Additionally, it is necessary to express these data in other nutrient currencies beyond kilocalories, in order to explore nutritional diversity issues, as well as sustainable diets (Müller et al. 2014).

Health. Earlier work on health impacts of pesticide use on farm workers and other occupational risks could be used to construct health impact modules (Antle and Pingali 1994). As elsewhere, big data (e.g., in this case, data from medical records or insurance claims) can be used to improve understanding of impacts (Rzhetsky et al. 2014).

Age, Gender and Health Status. Research on various aspects of gender impacts and outcomes has advanced, primarily in terms of relevant measures. With better data, analysis of gender impacts associated with new technologies could be incorporated into existing farm household models and impact assessment models. A similar situation exists for analysis of impacts by age and health status.

Vulnerability and equity. The application of different farm improvement methods has explicit winners but also unintended 'casualties' and perverse incentives. From a development standpoint, it is essential to understand these dynamics to ensure that appropriate policies are developed to maintain equal opportunities for all sectors of society. For example, in many cases, rich farmers are the ones who adopt technologies early. This factor could potentially disrupt power relationships in markets, thus affecting poorer farmers. In this case it is essential to design alternative options and safety nets for poorer farmers to prevent widening the gap and making them more vulnerable. New models should improve our understanding of these processes, as we move from single farm models to multi-farm and regional models.

Understanding structural change and rural development. When is rural development really about agriculture? New models should help us to target this question more effectively, and to find out when interventions in the agricultural sector will not be efficient in lifting the livelihood status of farmers or a region. Identification of thresholds in farm sizes, farm-derived incomes and others, will be a necessary feature of some NextGen models.

5. Towards Implementation

A long-term strategy for implementation of NextGen models could be to encourage developments on both the demand side and the supply side of the "market" for agricultural system models and knowledge products. On the demand side, we see a need for knowledge product developments to be linked with improved engagement of traditional end-users including both small-holders in the developing world and larger-scale commercial agriculture, as well as new potential end-users such as the crop insurance and reinsurance industries. On the supply side, we see a role for private-public partnerships to facilitate data and collection and sharing, as well as collaborative model development and testing, combined with better communication with the demand side to help guide the researchers in the "pre-competitive space" towards the model developments that could be useful in the "competitive space" of knowledge products. One such initiative has already been started through collaboration between AgMIP and CIMSANS. (see President's Climate Data Initiative, https://www.whitehouse.gov/the-press-office/2014/07/29/fact-sheet-empowering-america-s-agricultural-sector-and-strengthening-fo)

The concept of "competitive space" is typically conceived as the development of knowledge products that are provided through commercial markets – i.e., as "private goods." There is also an important public good aspect to these knowledge products. Some of these public goods are for public policy and investment decision making. In addition, it is important to consider the possibility that there could be obstacles, in the form of up-front fixed costs, to the development of decision support tools needed by small-holder farmers in the developing world, even though these tools could ultimately have substantial private and social value. Thus, there is arguably a role for some form of public or private charitable support for the development of these tools.

In order to facilitate the development of NextGen models, we see value in a multi-pronged approach.

First, we see a need for better testing and inter-comparison of existing models, extending the model-inter-comparison work already pioneered by AgMIP. In addition to the work begun to inter-compare and improve process-based crop models and global economic models, there is a need for similar work with livestock models and farm-level and regional economic models. This type of work could be facilitated by the identification of some "test areas" where high-quality data are available for important types of agricultural systems. Using these test areas, various types model inter-comparisons and model testing and validation exercises could be carried out using standard evaluations protocols.

Second, in parallel with this testing, we see great need for investments in the design and testing of modular open-source model components and in the testing of alternative model integration strategies. AgMIP has begun work in this area, and is encouraging participation across the modeling community.

Third, as discussed further in the companion paper on *Building an Open Web-Based Approach,* there is a need for parallel development of ITC tools to support the software engineering, data input and output, and data visualization needed to make NextGen models useful to knowledge product developers and end-users.

References

Antle, J.M. and P. Pingali. 1994. "Pesticides, Productivity, and Farmer Health: A Philippine Case Study." 76(3):418-430.

Antle, J.M., S.M. Capalbo and L. Houston. 2014. "Towards a Knowledge Infrastructure for Science-Based Policy and the Sustainable Management of Agricultural Landscapes." http://www.foodandagpolicy.org/sites/default/files/Antle,%20Capalbo,%20and%20Houston%20Sept2014_0.pdf

Antle, J.M., J. Stoorvogel and R. Valdivia. 2014. "New Parsimonious Simulation Methods and Tools to Assess Future Food and Environmental Security of Farm Populations." 369:20120280.

Antle, J. M., J. W. Jones and C. Rosenzweig. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Introduction." AgMIP, < url tbd >.

Antle, J. M., B. O. Basso, R. T. Conant, C. Godfray, J. W. Jones, M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig, P. Tittonell, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation." AgMIP, < url tbd >.

Asseng, S., F. Ewert, C. Rosenzweig, J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P.J. Thorburn, R.P. Rotter, D. Cammarano, N. Brisson, B. Basso, P. Martre, P.K. Aggarwal, C. Angulo, P. Bertuzzi, C. Beirnath, A.J. Challinor, J. Doltra, S. Gayler, R. Goldberg, R. Grant, L. Heng, J. Hooker, L.A. Hunt, J. Ingwersen, R.C. Izaurralde, K.C. Kersebaum, C. Müller, S. Naresh Kumar, C. Nendel, G. O'Leary, J.E. Olesen, T.M. Osborne, T. Palosuo, E. Priesack, D. Ripoche, M.A. Semenov, I. Shcherbak, P. Steduto, C. Stöckle, P. Stratonovitch, T. Streck, I. Supit, F. Tao, M. Travasso, K. Waha, D. Wallach, J.W. White, J.R. Williams, and J. Wolf. 2013. "Uncertainty in simulating wheat yields under climate change." 3:827-832 doi:10.1038/nclimate1916.

Bender, J., and H.J. Weigel. 2010. "Changes in atmospheric chemistry and crop health: A review." 31:81-89.

Basso, B., M. Bertocco, L. Sartori, and E.C. Martin. 2007. "Analyzing the effects of climate variability on spatial pattern of yield in a maize–wheat–soybean rotation." 26(2): 82-91.

Bassu, S., N. Brisson, J.L. Durand, K.J. Boote, J. Lizaso, J.W. Jones, C. Rosenzweig, A.C. Ruane, M. Adam, C. Baron, B. Basso, C. Biernath, H. Boogaard, S. Conijn, M. Corbeels, D. Deryng, G. De Sanctis, S. Gayler, P. Grassini, J. Hatfield, S. Hoek, C. Izaurralde, R. Jongschaap, A. Kemanian, K. Kersebaum, C. Müller, C. Nendel, E.Priesack, M. Virginia, P. Nin, F. Sau, I. Shcherbak, F. Tao, E.Teixeira, D. Makowski, D. Timlin, and K. Waha. 2014. "How do various maize crop models vary in their responses to climate change factors?" (7): 2301-2320.

Bennett, N.D., B.F.W. Croke, G. Guariso, J.H.A. Guillaume, S.H. Hamilton, A.J. Jakeman, S. Marsili-Libelli, L.T.H. Newham, J.P. Norton, C. Perrin, S.A. Pierce, B. Robson, R. Seppelt, A.A. Voinov, B.D. Fath, and V. Andreassian. 2013. "Characterising performance of environmental models." 40:1-20.

Brooks-Pollock, E., G.O. Roberts, and M.J. Keeling. 2014. "A dynamic model of bovine tuberculosis spread and control in Great Britain." Nature 511:228–231.

Colbach, N., A. Collard, S.H. Guyot, D. Mézière, and N. Munier-Jolain. 2014. "Assessing innovative sowing patterns for integrated weed management with a 3D crop: weed competition model." :74-89.

Coupled Model Inter-comparison Project. (2012) http://cmip-pcmdi.llnl.gov/. Accessed December 23, 2012.

Cressie, N., C.A. Calder, J.S. Clark, J.M. Ver, and C.K. Wikle. 2009. "Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling." 19(3):553-570.

Elliott, J., D. Deryng, C. Müller, K. Frieler, M. Konzmann, D. Gerten, M. Glotter, M. Florke, Y. Wada, N. Best, S. Eisner, B.M. Fekete, C. Folberth, I. Foster, S.N. Gosling, I. Haddeland, N. Khabarov, F. Ludwig, Y. Masaki, S. Olin, C. Rosenzweig, A.C. Ruane, Y. Satoh, E. Schmid, T. Stacke, Q. Tang and D. Wisser. 2014. "Constraints and potentials of future irrigation water availability on agricultural production under climate change." (9):3239-3244.

Elliott, J., D. Kelly, J. Chryssanthacopoulos, M. Glotter, K. Jhunjhnuwala, N. Best, M. Wilde and I. Foster. 2014. "The parallel system for integrating impact models and sectors (pSIMS)." 62:509-516.

Estay, S. A., M. Lima, & F.A. Labra. 2009. "Predicting insect pest status under climate change scenarios: combining experimental data and population dynamics modelling." (7):491-499.

Fischer, G., M. Shah, H. van Velthuizen, and F.O. Nachtergaele. 2001. "Global agro-ecological assessment for agriculture in the 21st century: Methodology and Results." International Institute for Applied Systems Analysis. RR-02-002.

Fischer, G., M. Shah, and H. van Velthuzien. 2002. "Climate Change and Agricultural Vulnerability." A special report, prepared by the International Institute for Applied Systems Analysis under United Nations Institutional Contract Agreement No. 1113 on "Climate Change and Agricultural Vulnerability" as a contribution to the World Summit on Sustainable Development, Johannesburg 2002.

Fischer, G., H. Van Velthuizen, F.O. Nachtergaele, and A. Jernelöv. 2000. Global agro-ecological zones assessment: methodology and results. International Institute for Applied Systems Analysis. IR-00-064.

Fuglie, K., and S.L. Wang. 2014. "New Evidence Points to Robust But Uneven Productivity Growth in Global Agriculture." United States Department of Agriculture. http://www.ers.usda.gov/amber-waves/2012-september/global-agriculture.aspx#.U9U2GKjc0vA. Accessed July 25 2014.

Trade Analysis Project. 2014. Current GTAP Model. https://www.gtap.agecon.purdue.edu/models/current.asp. Accessed July 30, 2014.

Hammer, G., M. Cooper, F. Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, D. Podlich. 2006. "Models for navigating biological complexity in breeding improved crop plants." (12):587-593.

Havlik,P, H. Valin, M. Herrero, M. Obersteiner, E. Schmid, M.C. Rufino, A. Mosnier, P.K. Thornton, H. Bottcher, R.T. Conant, S. Frank, S. Fritz, S. Fuss, F. Kraxner, and A. Notenbaert. 2014. "Climate Change Mitigation through Livestock Transitions." 111(10):3709-3714; doi:10.1073/pnas.1308044111.

Herrero, M., P. Havlik, J. McIntire, A. Palazzo, and H. Valin. 2014. "African Livestock Futures: Realising the Potential of Livestock for Food Security, Poverty Reduction and the environment in sub-Saharan Africa." Office of the Special Representative of the UN Secretary General for Food Security and nutrition and the United Nations Systems Influenza Coordination (UNSIC), Geneva, Switzercland, 118p.

Howitt, R.E., J. Medellin-Azuara, D. MacEwan, J.R. Lund, and D.A. Sumner. 2014. "Economic Analysis of the 2014 Drought for California Agriculture." Center for Watershed Sciences, University of California, Davis, California. 20p. Available at <http://watershed.ucdavis.edu>.

Janssen, S., K. Louhichi, A. Kanellopoulos, P. Zander, G. Flichman, H. Hengsdijk, E. Meuter, E. Andersen, H. Belhouchette, M. Blanco, N. Borkowski, T. Heckelei, M. Hecker, H. Li, A. Oude Lansink, G. Stokstad, P. Thorne, H. van Keulen, and M. van Ittersum. 2010. "A Generic Bio-Economic Farm Model for Environmental and Economic Assessment of Agricultural Systems." 46:862-877.

Janssen, S., C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, and J. M. Antle. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support." AgMIP, < url tbd >.

Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science." AgMIP, < url tbd >.

Kobayashi, M. T. E. Carpenter, B. F. Dickey and R. E. Howitt. 2007. "A dynamic, optimal disease control model for foot-and-mouth disease." 79:257- 273.

Kotz, S., and J.R. van Dorp. Beyond Beta: Other Continuous Families Of Distributions With Bounded Support and Applications. World Scientific Publishing Co. Pte. Ltd. Singapore. 2004.

Leakey, A. D., E.A. Ainsworth, C.J. Bernacchi, A.Rogers, S.P. Long, and D.R. Ort. 2009. "Elevated CO2 effects on plant carbon, nitrogen, and water relations: six important lessons from FACE." (10):2859-2876.

Leisner, C. P., and E.A. Ainsworth, E.A. 2012. "Quantifying the effects of ozone on plant reproductive growth and development." (2):606-616.

Manson, S.M. 2007. "Challenges in evaluating models of geographic complexity." 34: 245-260.

Marin, F. R., J. W. J., A. Singels, F. Royce, E.D. Assad, Giampaolo, Q. Pellegrino & F. Justino. 2013. "Climate change impacts on sugarcane attainable yield in southern Brazil." (2):1–13.

Martre, P., D. Wallach, S. Asseng, F. Ewert, J.W. Jones, R.P. Rotter, K.J. Boote, A.C. Ruane, P.J. Thorburn, D. Cammarano, J.L. Hatfield, C. Rosenzweig, P.K. Aggarwal. C. Angulo, B. Basso, P. Bertuzzi, C. Biernath, N. Brisson, A.J. Challinor, J. Doltra, S. Gayler, R. Goldberg, R.F. Grant, L. Heng, J. Hooker, L.A. Hunt, J. Ingwersen, R.C. Izaurralde, K.C. Kersebaum, C. Muller, S.N. Kumar, C. Nendel, G. O'Leary, J.E. Olsen, T.M. Osborne, T. Palosuo, E. Priesack, D. Ripoche, M.A. Semenov, I. Shcherbak, P. Steduto, C.O. Stockel, P. Stratonovitch, T. Streck, I. Supit, F. Tao, M. Travasso, K. Waha, J.W. White, and J. Wolf. 2014. "Mutlimodel ensembles of wheat growth: many models are better than one." doi: 10.1111/gcb.12768

McDonald, M.G., and A.W. Harbaugh. 1983. "A modular three dimensional finite-difference ground-water flow model." US Geological Survey Open-File Report 83-875, 528.

Morris M.D. 1991. "Factorial sampling plans for preliminary computational experiments." 33(2):161-174.

Müller, C., J. Elliott, and A. Levermann. 2014. "Food security: Fertilizing hidden hunger." (7):540-541.

Muller, S., R. Munoz-Carpena, and G. Kiker. 2011. "Model Relevance: Frameworks for Exploring the Complexity-Sensitivity-Uncertainty Trilemma." In: I. Linkov and T.S.S. Bridges (Eds). . NATO Science for Peace and Security Series C: Environmental Security. Pp. 35-65. Springer: Boston doi:10.007/978-94-007-1770-1_4.

Parsons, D., C.F. Nicholson, R.W. Blake, Q.M. Ketterings, L. Ramírez-Aviles, D.G. Fox, L.O. Tedeschi, and J.H. Cherney. 2011. "Development and evaluation of an integrated simulation model for assessing smallholder crop–livestock production in Yucatán, Mexico." (1):1-12.

Raick, C., K. Soetaert and M. Grégoire (2006). "Model complexity and performance: How far can we simplify?" 70(1): 27-57.

Ratto, M., P.C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano. 2007. "Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology." , Copernicus Publications 11(4):1249-1266.

Rosegrant, M.W. and the IMPACT Development Team. 2012. International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT): Model Description. International Food Policy Research Institute (IFPRI), Washington, D.C. (2)

Rosenzweig, C., J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P.J. Thorburn, J.M. Antle, G.C. Nelson, C.H. Porter, S. Janssen, S. Asseng, B. Basso, F. Ewert, D. Wallach, G.A. Baigorria, and J.M. Winter. 2013. "The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies." :166–182.

Ruane, A. C., S. McDermid, C. Rosenzweig, G.A. Baigorria, J.W. Jones, C.C. Romero, and C.L. DeWayne. 2014. "Carbon–Temperature–Water change analysis for peanut production under climate change: a prototype for the AgMIP Coordinated Climate□Crop Modeling Project (C3MP)." (2):394-407.

Rzhetsky, A., S.C. Bagley, K. Wang, C.S. Lyttle, E.H. Cook Jr, R.B. Altman, and R.D. Gibbons. 2014. "Environmental and state-level regulatory factors affect the incidence of autism and intellectual disability." (3):e1003518.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Rattio. 2004. Sensitivity analysis in practice: a guide to assessing scientific models. Wiley, Hoboken, NJ.

Saltelli, A., M. Ratto, S. Tarantola, and F. Campolongo. 2005. "Sensitivity Analysis for Chemical Models." 105(7):2811-2828.

Taylor, K.E., R.J. Stouffer, and G.A. Meehl. 2012. "An Overview of CMIP5 and the Experiment Design." 93(4):485-498

Thorburn, P., K. Boote, J. Hargreaves, P. Poulton, and J. Jones. 2014. "Crop Systems Modeling in AgMIP: A New Protocol-driven Approach for Regional Integrated Assessments."In D. Hillel and C. Rosenzweig (Ed.), ICP Series on Climate Change Impacts, Adaptation, and Mitigation. Imperial College Press. Thornton, P. K., & M. Herrero. 2001. "Integrated crop–livestock simulation models for scenario analysis and impact assessment." (2):581-602.

Valdivia, R.O., J.M. Antle, C. Rosenzweig, A.C. Ruane, J. Vervoot, M. Ashfaq, I. Hathie, S.H.K. Tui, R. Mulwa, C. Nhemachena, P. Ponnusamy, H. Rasnayaka, and H. Singh. 2014. "Representative Agricultural Pathways and Scenarios for Regional Integrated Assessment of Climate Change Impact, Vulnerability and Adaptation." In , ed. D. Hillel and C. Rosenzweig. ICP Series on Climate Change Impacts, Adaptation, and Mitigation Vol. 3. Imperial College Press, 101-156.

Villa-Vialaneix, N., M. Follador, M. Ratto, and A. Leip. 2011. "A comparison of eight metamodeling techniques for the simulation of N_2O fluxes and N leaching from corn crops." 43:80-87 doi:10.1016/j.envsoft.2011.05.003.

White, J. W., G. Hoogenboom, B.A. Kimball, and G.W. Wall. 2011. "Methodologies for simulating impacts of climate change on crop production." (3):357-368.

Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support

S. Janssen, C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, J. M. Antle

Outline

Executive Summary

Agricultural modeling has long suffered from fragmentation in model implementation. Many models are developed, there is much redundancy, models are often poorly coupled, and it is often difficult to apply models to generate real solutions for the agricultural sector. To change this situation, we argue that an open, self-sustained, and committed community is required that co-develops agricultural models and associated data and tools as a common pool resource. Such a community can benefit from recent developments in information and communications technology (ICT). In this paper, we examine how such developments can be leveraged to design and implement the next generation of data, models, and decision support tools for agricultural production systems. We review relevant technologies from the perspectives of maturity, expected development, and potential to benefit the agricultural modeling community. The technologies considered go beyond mere hardware and software to encompass methods for collaborative development and for involving stakeholders and users in development in a transdisciplinary manner. Our qualitative evaluation suggests that ICT can bring many relevant developments to the agricultural modeling community, such as: (1) development of a modular "plug-and-play" approach to model components, rather than stand-alone, larger, and increasingly complex models, that can be combined flexibly to represent the wide array of systems that are now or could be in use in the future; (2) a collaborative and open approach to software development designed to meet end-user needs; (3) use of cloud and web-based computing technologies, to reduce costs of operating and delivering models and to broaden access to models; (4) better utilization of sensor and data collection methods, including remote sensing, crowdsourcing, and mobile technology; and (5) creation of new, and exploitation of existing, tools to generate, archive, access, analyze, visualize, and interpret model inputs and outputs, building on the tools that AgMIP and other organizations are now developing. New ways to collect, archive and supply experimental and observational data are needed to better manage the data side of modeling. Data standardization, archiving, and access methods are required that use web-based cloud architecture to simplify and broaden access.

1. Introduction

Information and computer technology is changing at a rapid pace. Digital technologies allow people to connect across the globe at high speeds at any time. Even people in remote, developing regions increasingly have the ability to connect online via telephone and internet providers. Satellite and drone capabilities allow us to obtain remotely sensed data in real-time regarding in-season crop growth and development, soil moisture, and other dynamic variables. High performance computing allows us to process and make sense of big data (i.e. large quantities of structured and unstructured data that require special processing and visualization power) collected using new sensing technologies, and to scale and validate models in ways not previously possible. As a result, we expect more and higher-quality information to be available in support of daily decision-making. However, modeling and decision support systems in the agricultural sciences have not kept up with these advances in technology. Many of the frameworks used in these systems originated in the 1970s through 1990s, prior to the availability of advanced data collection, computing, storage, access, and processing technologies.

Undoubtedly, agricultural systems modeling could benefit from the advancements in ICT with big data, crowdsourcing, remote sensing, and high computational abilities, catching up with the relative slow developments over the last two decades. Against this background the practice of agricultural systems modeling could change and a next generation of tools could emerge, also with respect to ICT implementation. This would similarly hold then for the community developing, supporting and applying these tools. The vision for the next generation modeling community as described in the companion Paper 2 (Antle et al. 2014) from the modeling point of view, includes modelers and model developers working across disciplines, spatial scales and temporal scales, as well as software developers for the modeling framework, data processing applications, and visualization tools. This paper approaches the envisioned next generation modeling community for agricultural systems from the ICT perspective. It describes relevant developments in ICT in recent years, and their links to agricultural applications or systems modeling, thereby qualitatively assessing the maturity, expected development, and potential to benefit the agricultural modeling community. The technologies considered go beyond mere hardware and software to encompass methods for collaborative development and for involving stakeholders and users in development in a transdisciplinary manner. Central to our vision for the next generation modeling community to emerge and benefit from ICT developments are the principles that any development must (1) be open and transparent, so that all can contribute and understand the steps taken; (2) focus on community building as an integral part; and (3) include distributed, web-based components such as cloud computing and linked data. The recent developments in ICT will be assessed against these criteria, and also against five stylized use cases that have been formulated to support the vision for a next generation modeling community (see also the introductory overview by Antle et al. and accompanying papers by Jones et al. and Antle et al. 2015).

Background and theoretical considerations

A next generation of data-driven agricultural modeling and decision-making applications can help companies, governments and farmers in the food chain to make informed decisions. Two different concepts provide complementary perspectives on the value of data in this context.

The first perspective is based on the concept of a **knowledge chain**, as shown in Figure 1. This perspective postulates that data comprise a raw material and that when combined with description and quality attributes leads to information. Information can be linked to other information sources and placed in causal chains leading to knowledge. Ultimately, knowledge serves as an input for decisions based on wisdom, which cannot be digitized and which occurs in the mind of the decision-maker. Information also flows in the opposite direction, as when integration of data from different sources, and comparison with models, suggests deficiencies in data, and when stakeholder requests identify needs for new data.

A second perspective focuses on **application chains**. Data-driven applications are realized along a chain of user needs and requirements, implementation in infor-

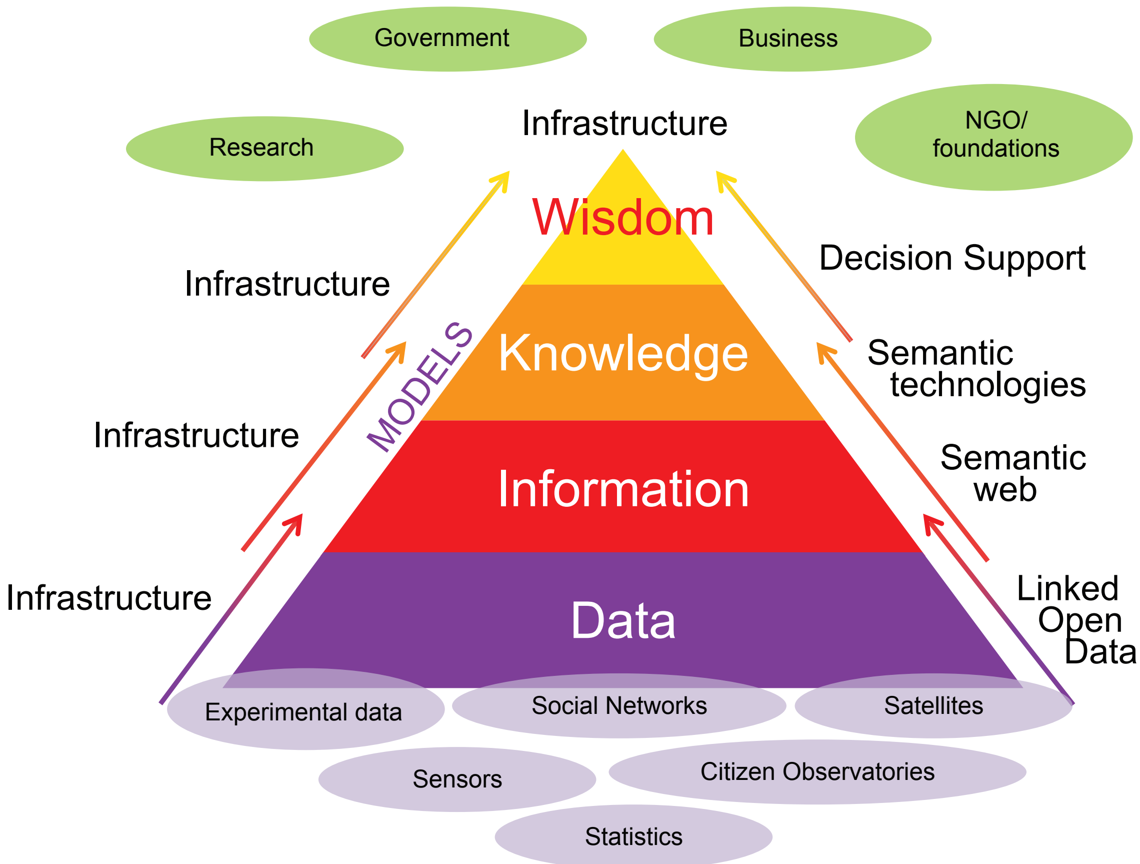

Figure 1. *Modeling system knowledge chain. Infrastructure includes technical, institutional, and organizational aspects (adapted from Lokers & Janssen 2014).*

mation systems (i.e., web sites, apps, text messaging services, computer programs), operated within a technical infrastructure. This technical infrastructure includes methods for data assimilation and synthesis, of which models and model outputs form a part. These methods of data assimilation and combination use in their operations raw data collected by governments, private organizations, scientists, or networks, in addition to information sources such as satellite data repositories and model simulations.

Data alone are not enough to meet the requirements of the NextGen agricultural modeling community, but must be engaged in an infrastructure consisting of both software and hardware (i.e., servers, computing capacity, and storage) as depicted in Figure 2. Based on the data in the infrastructure, applications targeted at end-users serve information and knowledge. Application chains may be simple or complex, and may include, for example, data access, extraction, transformation (e.g., format, grid), and integration operations; one or multiple models; integration of output from different models; and model output transformation, analysis, and visualization steps. Design of the knowledge framework must consider not only the end-users, but the many people who are needed to contribute data, software and scientific expertise to the components and operation of the system. Figure 2 illustrates the full spectrum of users of next generation ICT infrastructure including primary data collectors, database professionals, software developers, modelers and the end-users of knowledge and information.

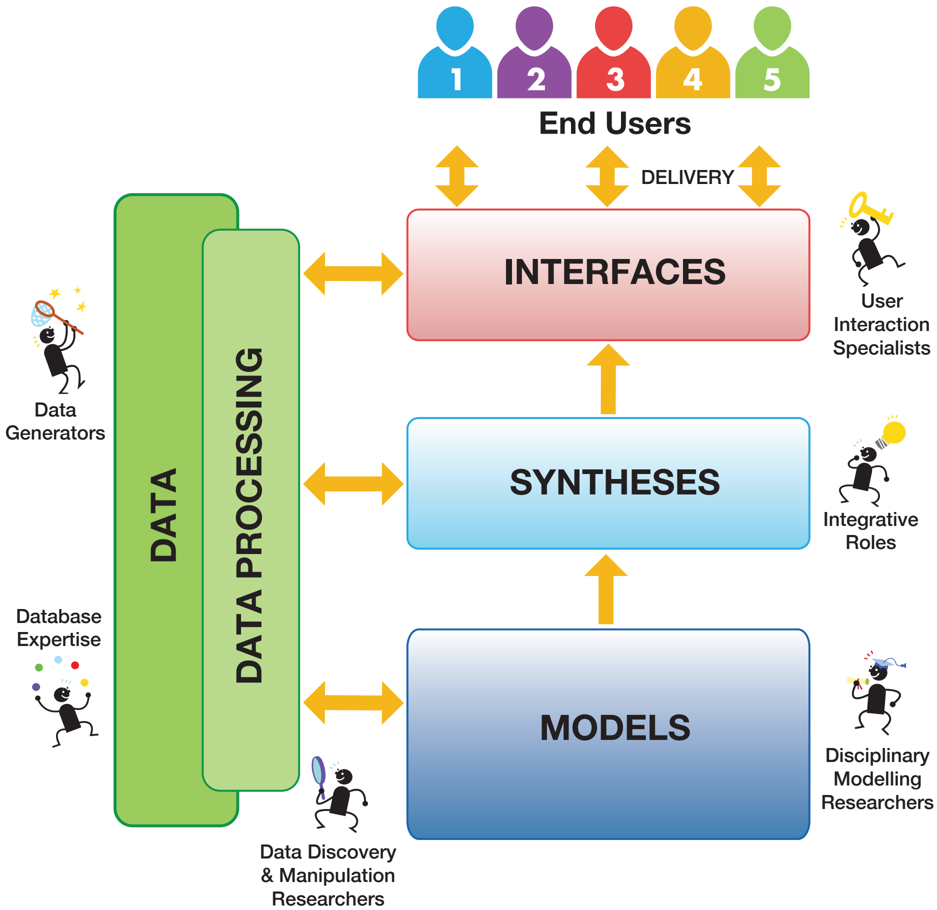

Figure 2. *Agricultural knowledge systems framework schematic. IT infrastructures are designed to serve the entire knowledge systems process, including data collection, model development, integration of model outputs, and delivery of products to end users.*

Information and knowledge provided by next-generation applications can not only improve understanding but also change the balance of power by allowing end users to better understand both the biological systems that they manage through their farming practices and each other's modes of operation. For example, farmers in Africa can now receive text messages regarding current crop prices, seed and fertilizer locations, and crop insurance, thus allowing them to make informed decisions based on up-to-date information (Rojas-Ruiz and Diofasi 2014).

The rapidly increasing digitization of society along with the increasing availability of internet and mobile technologies in agricultural communities provides massive opportunities for the hitherto underserved (Danes et al. 2014). In addition to ICT infrastructure, applications must be developed in a community of stakeholders including businesses, farmers, citizens, government, Non-Governmental Organizations (NGOs), and research institutions.

Data assimilation techniques have developed significantly in recent decades, and in some cases have reached maturity, thereby providing companies or policy-makers with relevant evidence for their decision-making. Data assimilation refers to the analysis of observations according to standard procedures and processes that integrate a mix of data from different sources as implemented in a computer. This requires specific investments and visionary analysts, researchers and problem-owners to develop new applications in such a way that usability increases. Examples include Climate Corp. in the U.S., recently acquired for

$1 billion by Monsanto, Monitoring Agricultural ResourceS (JRC, *www.mars.info*), and the Famine Early Warning Systems Network (USAID, *www.fews.net*).

2. Envisaged end users and user definition

As noted in the introductory paper (Antle et al.), next-generation model development starts with an understanding of information required by various stakeholders and then works back from those requirements to determine the models and data needed to deliver that information in the form that users want. ICT offers various techniques for scoping user requirements, from more traditional methods of user requirements analysis to modern techniques of user-centered design, in which software is built in direct contact with the end-user in short iterations. In the latter approach, user needs and requirements guide and modify the development in each iteration (Cockburn 2006).

At the same time, opportunities to be in touch with end-users have become more numerous with the advent of mobile technology, social media, text messaging, radio and TV shows, and apps designed for tablets and mobile phones. At the moment there are still geographic areas where smallholder farmers lack the mobile networks for sharing data; however, access to SMS and voice message services is increasing rapidly and it is expected that end-users in rural areas in developing countries will skip the step of personal computers and make direct use of mobile phones and possibly tablets (Danes et al. 2014). This trend suggests that there is a huge untapped potential to boost the amount of information provided to farmers and processors in the chain, both because their needs are not yet defined and because services specifically focused on their use are not yet developed.

Numerous use cases can be developed to represent the stakeholders who define the outputs and characteristics of next generation agricultural modeling. Box 1 summarizes five use cases, chosen to represent a wide range of farming systems, beneficiaries, and requirements for data and modeling components. Smallholder farming systems are featured in use cases 1, 2 and 3, as addressing the needs of these systems is considered to be essential to achieving food security in developing regions where smallholder farms account for most of the food production (Dixon et al. 2004). The end-users in these cases are (1) a farm extension professional evaluating the potential of improving local yields and nutrition using a combination of new cultivars, better field management, and improved water-harvesting methods; (2) a plant breeder evaluating adoption of drought-resistant varieties of maize for a sustainable intensification application in Africa; and (3) an investment manager for an NGO wishing to evaluate the potential yield gains and environmental sustainability of a fertilizer supply project in Kenya. Use cases 4 and 5 consider agri-businesses in developed regions with end-users being (4) a farmer in the US who wants to improve his yields while minimizing inputs by using model-based decision support in his precision agricultural management and (5) an economic analyst in the sustainability group of a large food supply corporation, who wants to use NextGen models to improve sustainability in their supply chain.

These use cases have been described more fully in the introductory paper (Antle et al.). In the following sections of this paper we present overviews of the various facets of a next generation modeling community and the necessary ICT infrastructure to support the community with a focus on the capabilities of existing technologies, current trends and the suitability of each to meet the requirements of end users.

In all five use cases, the solution requires integrated modeling capabilities using high quality, up-to-date data products. The end users are neither modelers nor data collectors and therefore solutions must focus on techniques that provide synthesized modeling results in formats that allow stakeholders to make informed decisions based on the current data and technologies. We return to these five use cases in Section 9, where we discuss the data, modeling components, and ICT infrastructure required by each. In each case, it is shown that the existing agricultural modeling frameworks are inadequate to provide rapid, reliable answers that the envisioned end users require for their livelihoods, health, and nutritional security. Components of an integrated system that could provide the needed data, information and knowledge to the specified end users are then put forward as straw-man proposals for each use case, to provide a road-map for establishing next

Box 1. Use Cases

The use cases were created to represent a range of plausible users of knowledge products that are linked to next generation models and data. The five use cases, which are designed according to four components indicated in Table 1, represent two types of farming systems:

Small-holders: small-scale semi-subsistence farms typical of Africa and much of the developing world, many of which produce a mix of subsistence crops, cash crops and some livestock;

Commercial crop farms: large-scale commercially-oriented crop farms typical of the industrialized countries including the United States.

	Use cases				
	1	2	3	4	5
	Farm Extension in Africa	Developing and evaluation technologies for sustainable intensification.	Investing in agricultural development projects that support sustainable intensification.	Management support for precision agriculture.	Supplying for products that meet corporate sustainability goals.
Farming System	small-holder	small-holder	small-holder	commercial corp	commercial corp
Information User	Farm advisor	Agricultural research team/program	Analyst/adviser	Management consultant	Corporate analyst
Beneficiaries	Farm family	Research institution/ farm population	NGO & clients	Farm business	Agri-business firm
Outcomes	Improved livelihood (income,nutrition, food security)	Improved technology	Sustainable technology	Income, soil conservation & water quality	Profit, risk management, sustainability objectives

generation models, modelers and the associated ICT infrastructure.

3. Envisaged knowledge chain users

Just as important as the end users in the NextGen agricultural modeling infrastructure are the data collectors, software and model developers, database experts and user interaction specialists who are needed to contribute new capabilities to the envisaged modeling framework and benefit from its capabilities. The successful implementation of the NextGen system must be designed in collaboration with the professional communities of these contributor groups. Existing model development teams will play a key role in defining the capabilities and bounds of the proposed NextGen modeling infrastructure. This definition can be done initially through a series of planning workshops, which can take advantage of existing agricultural modeling communities such as AgMIP (Rosenzweig et al. 2013) and MACSUR (*www.macsur.eu*) and by engagement with projects such as FACE-IT (*www.faceit-portal.org*), GeoShare (*geoshareproject.org*) and BIOMA (Donnatelli and Rizolli 2008) that are exploring new approaches to model development, integration, and workflow management. Continued model development partnerships using collaborative design methodologies are a necessary component of successful development infrastructure.

A long-term strategy of the NextGen modeling infrastructure will be to develop a means of entraining new model developers and other knowledge system specialists from the user community, especially in developing regions of the world. Users and developers from emerging economies can bring unique perspectives that can guide the model development process to include key relevant components critical to their user communities. For example, a modeler in West Africa might emphasize the importance of soil phosphorus in yield limitations of that region – a modeling component that has been lacking in many existing agricultural production models that were generated in regions where phosphorus is not typically a limiting factor. In another example, the role of individual actors making ranges of decisions about farm management is rarely adequately represented in current models which generally assume uniform management across the study area. Participants with deep knowledge of the decision-making requirements of such actors, and the ICT technologies available to them (e.g., mobile phones) can contribute to components that meet the unique needs of these actors.

Representation of the many factors used by human actors in decision-making will require a change in current modeling approaches. These examples underscore the need for a flexible infrastructure allowing compatible components to be developed independently by various user communities and then combined in multiple, possibly unforeseen ways for agricultural model application, intercomparison, and evaluation. Existing modeling communities are adapting and new model development teams are emerging to contribute to this effort.

The development of the NextGen modeling infrastructure need not imply that existing models be discarded. On the contrary, while we certainly hope to spur development of new model and data components, we also expect to “retrofit” existing complete models to allow enhanced capabilities of model linkages, connections with new data sources, and visualization and archiving of model outputs. The NextGen infrastructure should facilitate the comparison of new models and model components with widely used and trusted existing models. This parallel path of generating new models while maintaining existing models optimizes the extensive and irreplaceable knowledge that the current generation of modelers bring to the community.

In addition to the requirements for model component development, the NextGen framework requires a substantial commitment in the development of a software infrastructure to allow the data and modeling components to be linked through compatible interfaces into cohesive workflows. Data products from the NextGen system will include both model outputs and data products synthesized from the model outputs or from raw data. Synthesis applications include tools for data discovery and visualization, dynamic mapping, and statistical analyses. It is anticipated that these types of infrastructure development projects will be realized through grant funding for collaborations between research organizations similar to research projects such as FACE-IT (*www.faceit-portal.org*), GeoShare (*geoshareproject.org*) and iPlant (*www.iplantcollaborative.org*).

Finally, software development for access of NextGen data products and delivery of the final products to users is required. This top layer, represented in Figure 2, is likely to include both proprietary products developed by private industry and non-proprietary products developed in the public sector. It is envisioned that with the proper infrastructure, enabling rapid data discovery and use, the delivery of agricultural data products may become the realm of many small and medium-sized, local enterprises that can profit from the opportunities provided by the data and products through the development of mobile and web service applications for use directly by farmers and NGOs.

4. Agricultural data

The availability of data at the level of farming households and communities is low in the developing world compared to the developed world. We identify three main traditional methods of data collection:

1. Governments collect data for monitoring purposes, management of information and administrative procedures. These data, which include national statistics, monitoring data for subsidies and taxes, and data to monitor environmental performance, are generally uniform in format and are usually collected on a regularly scheduled basis for as long as they are relevant for policies.

2. Research projects collect data (e.g., field and household surveys, multi-dimensional panel data, soil sampling, measurements in laboratories) to meet specific project needs. These data are often incidental (i.e., collected on an irregular schedule) and not structured (i.e., non-uniform in format).

3. Industries (including farmers and business-to-business service operators) collect data for their own operations. They do not usually share data due to competitive concerns.

These sources of data have led to masses of data being potentially available for research; however, often these data are closed as they are only used for specific purposes by institutions or not well managed for future accessibility. These traditional sources are becoming more available as open data as evidenced by the G8 International Conference on Open Data for Agriculture (*feedthefuture.gov/event/g8-international-conference-open-data-agriculture*), several open data projects in Africa, the iPlant Collaborative for plant genetics data (*www.iplantcollaborative.org*), and

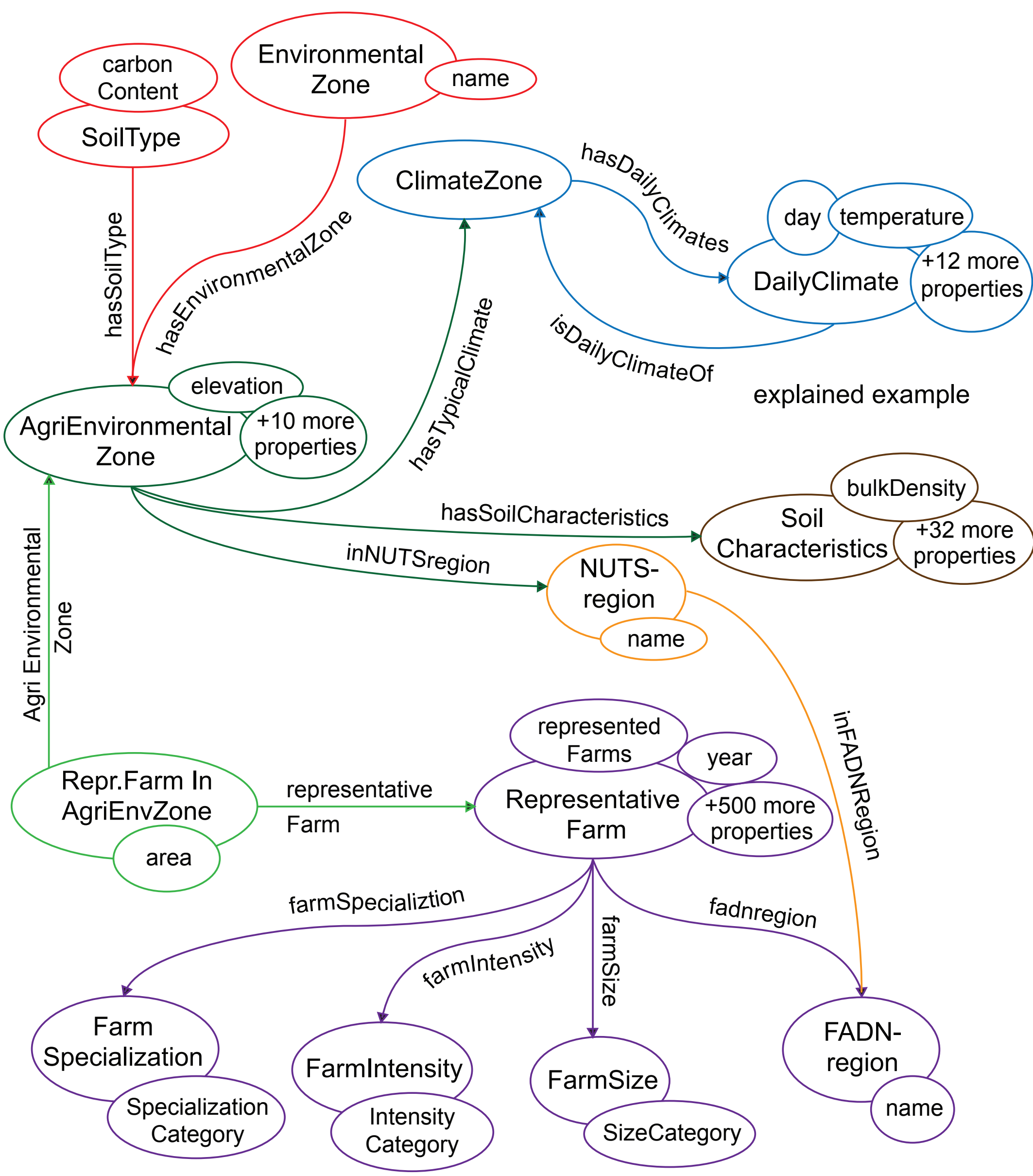

Figure 3. *Example of a common vocabulary snippet available for European farming systems as described in Janssen et al. (2009).*

numerous other examples. The open data movement can help to increase the availability of data. Within this open data movement, governments, international organizations, research institutions, and businesses work to offer open access to their data sets to make re-use easier. This also requires infrastructure to serve the data, for example, data.gov, data.gov.uk, data.overheid.nl or data.fao.org. Global Open Data for Agricultural and Nutrition (GODAN, *www.godan.info*) is a particularly relevant initiative for open data in the area of food security, as a multi-party discussion and advocacy forum initiated by the U.S. and UK governments and supported by many different parties. In science, several specialized journals to publish data files are appearing, for example the Open Data Journal for Agricultural Research (*www.odjar.org*) that originated from AgMIP.

With the increased availability of data, a greater need is created to ensure interoperability of data by aligning both syntax (formats) and semantics (definitions). Improved data interoperability creates new opportunities for all types of analysis and the development of new products. However, the necessary standardization has not yet been reached. There are technical standards that are maintained by International Organization for Standardization (ISO), W3C (World Wide Web Consortium), and Open Geospatial Consortium (OGC), which, however, do not cover connection to the content level of the significance and usefulness of the data. To this end, there are all kinds of developments around semantics, which leads to better descriptions of concepts (i.e., variables) in data sources. Major efforts for agricultural data harmonization include Agrovoc of FAO, CABI's Thesaurus, the CGIAR crop ontology, and AgMIP data interoperability tools (Porter et al. 2014), as examples. However, this effort is still largely in its infancy, while mostly lacking in common vocabularies over different domains, as achieved in Janssen et al. (2009) (Fig. 3).

The emergence of mobile technology capabilities significantly supports the advancement of crowdsourcing (sometimes called citizen science or civic science). Mobile phones, GPS, and tablet devices act as sensors or instruments which directly place data online, with accurate location and timing information. These techniques are often seen as an opportunity for near-sensing, i.e. using sensor-equipped tools in the field for capturing observations, e.g., temperature measurements on the basis of data derived from a mobile device. There are also special tools such as leaf area index sensors and unmanned aerial vehicles (UAV), which obtain more location-specific data. These crowdsourcing technologies offer the opportunity to gather more data and at lower cost. In these cases, citizens help to collect data through voluntary efforts, for example biodiversity measurements, mapping, and early warning. Smallholders, citizens, and organizations can thus manage their own data as well as contribute to public data. This offers many possibilities (especially as technology is still in its infancy), with some successful applications such as IIASA's Geowiki (*www.geo-wiki.org*). Sometimes crowdsourcing is organized as public events, for example, air quality measurements in the Netherlands on the same day at the same time (Zeegers 2012).

Earth observation through satellites now provides a continuous record since the early 1980s, forming a data source for time-series observations at any location. More detailed satellite data are coming online through NASA and EU space programs. This leads to an increasing demand for satellite applications and analyses based on satellite information.

Making data available to models requires processing and transformation of the data to ensure quality, consistency and compatibility. Data integration requires reconciliation of semantic properties of the data (i.e., the definitions of the variables in the data, their units, and the relationships between the variables) and the modeling components that use the data. The data may require statistical or geo-processing steps. Data-quality control requires manipulation to provide accurate and complete data needed by models, including synthesizing data to fill gaps and deficiencies. Data provenance, including data source and how data were cleaned, manipulated, linked and combined must be included in metadata (i.e., the data providing information on other data such as ownership, units, resolution) and maintained at every step in the process. Similarly, data quality over time is important to the data analyst and must become part of the provenance metadata. It is important that a user know that older data may not be as accurate or relevant as more recent data due to improvements in methods of data collection or manipulation. For most current integrated modeling approaches, these steps are done in a semi-automated way and on an ***ad hoc*** basis for each modeling project. Laniak et al. (2013) report that solutions to this issue are beginning to emerge including the GeoSciences Network (GEON, Ludäscher et al. 2003) and Data for Environmental Modeling (D4EM, Johnston et al. 2011), CUAHSI (Maidment et al. 2009).

Data relevant for the end-user include both near-term and long-range decision support information including (1) fundamental information on farming techniques; (2) contextual information about the current weather and types of crops that work well in the local area; and (3) market information such as prices of inputs and commodities, demand information and transport and logistics information (Mittal et al. 2010 and Steinfield et al. 2013). The types of data required vary during the agricultural life-cycle from planning the crop season through marketing the products. Chapman and Slaymaker (2002) further delineate information categories into Type A, representing information for long-term capacity building involving education, training and technical support; and Type B, representing information for short-term decision-making used to maximize the potential of a particular asset at any one time, reduce vulnerability to shocks and respond to immediate needs. Type A information is the traditional focus of agricultural extension. Type B information could include information about markets and news relating to weather or rural services that require frequent updating.

5. Confidentiality, privacy and intellectual property issues

A major motivation for the NextGen modeling framework is to open up access to data and software that have previously been inaccessible for various reasons, in ways that facilitate discovery, composition, and application by a wide variety of researchers and disciplines. However, while NextGen will certainly benefit from a growing interest in open data and open source software, confidentiality, privacy, and intellectual property issues remain important. While interconnected, they are different issues.

The first issue involves allowing access to data and software. This entails appropriate licensing schemes being endorsed to allow access to information. While reproducibility of science has always been advocated, the typical interpretation in natural sciences does not include allowing access to data sources and software. Many reasons are behind this, which vary in different parts of the world, most notably, lack of legislation that obliges public access to environmental data, limited credit that academics receive for releasing datasets or software, and a lack of sustainable funding mechanisms for long-term collection and curation of important classes of data. To change the current practice, the NextGen community needs to move rapidly to the paradigm of sharing open data.

There has been significant progress in other disciplines as evidenced by such open data collections as the iPlant Collaborative for genetics data (*www.iplantcollaborative.org*) and the National Ecological Observatory Network (NEON, *www.neoninc.org*) for environmental science observations. One of the most important aspects is that open-access licenses should facilitate ease of access. Thus, simple licenses that give clear rights and obligations to the users need to be endorsed. In other disciplines, scientists have struggled to cope with complex licenses that ultimately are introduced only for enforcing certain wishes of the data owners, and introduce obstacles for the one who tries to reuse the dataset. This ends up being complex and cumbersome for the end-user. A simple licensing scheme, that consists of licenses that are compatible with each other (as a creative commons) should be considered as the way forward. An important activity of the NextGen community is to thoroughly examine its data access needs and pre-select a set of licenses for this end. Effective open access to information needs to be seen as an enabling mechanism for the community. Many kinds of data that are critical to NextGen (for example weather records) are curated on behalf of many users by distinct communities of practice; in these situations, NextGen will need to have an influencing rather than a leading role in structuring data-licensing arrangements.

Second are issues of privacy. In several cases, agricultural data may come packaged with sensitive or confidential information. Examples are nominal records, geographic location, economic information, agronomic practices, etc. that reveal a subject's identity, location or entrepreneurial knowledge. Such information can be protected by not disclosing it, but sometimes effective mechanisms for data anonymization or obfuscation are needed to protect privacy (CAIDA 2010). There are also techniques for ensuring privacy preservation during computations (e.g., see Drosatos et al. 2014). The NextGen community needs to invest in protocols for protecting privacy and confidentiality.

While open access is a fundamental pillar of the NextGen community, we need to acknowledge that some resources (data or software) will always be proprietary for a range of reasons, and thus not shareable with others. For the NextGen platform to be useful, it requires the endorsement of non-invasive licenses (i.e., ones that do not oblige users to replicate the same license), while it should support some kind of permit control (i.e., enforcing attribution, correct use) over access to data and software. Again, there is a lot of experience in other sectors on these approaches, and NextGen needs to adopt best licensing practices.

6. Visualizing and interpreting data and model outputs

Tools to enable visualization of agricultural source data, model outputs and synthesized data products are needed to enhance the discovery and understanding of information for the entire spectrum of NextGen users, including data collectors, model developers, model users, integrative research, application developers, and end-users. To make sense of large amounts of unfamiliar or complex data, humans need overviews, summaries, and the capability to identify patterns and discover emergent phenomena, empirical models, and theories related to the data (Fekete 2013). One of the big challenges for the NextGen infrastructure is to include data visualization tools that support the exploration of and interaction with big data.

Currently most visualization in agricultural systems modeling is organized in an ad-hoc way, by adding some visualization modules to models to produce

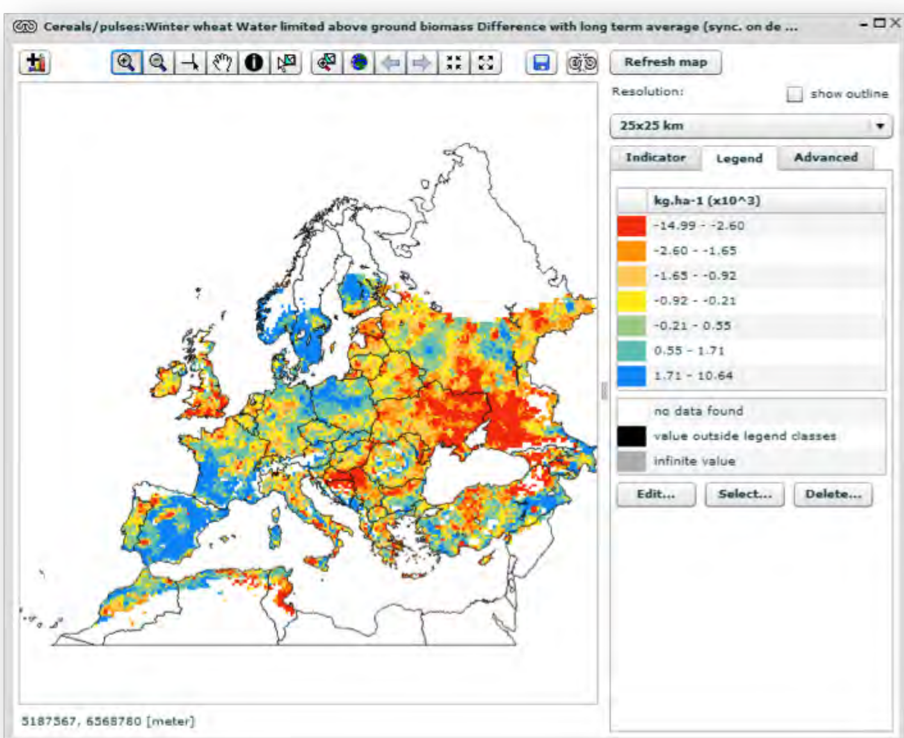

Figure 4. *Visualization of winter wheat water limited above ground biomass in comparison to long term-average in dedicated visualization software from the operational yield forecasting service of the European Commission (copyright: EC-JRC, www.marsop.info)*

graphs, tables and maps, or alternatively, by inputting and analyzing model outputs in other packages, for example spreadsheet or statistical programs. In these types of packages the key visualizations as messages for scientific papers are prepared. In cases where models are applied on a more regular basis for a specific purpose, more elaborated visualizations have been built, for example see Figure 4 below for a visualization from the Monitoring Agricultural Resources (MARS) Unit of the European Union.

6.1 Existing visualization packages

A whole industry has been built around the ability to easily visualize data, including interactive graphics driven by data that are collected, cleaned, and linked in real-time. There has been a shift away from batch-based processing toward near-real-time or stream processing, with real-time query engines integrated, driven by the increasing desire to visualize data and see results as they occur. Visualization tools have been getting easier to use and often end-users are able to generate their own visualizations using these tools. There are numerous visualization packages available; a few of the more popular are:

- Tableau (*www.tableausoftware.com*) – Visualization app / service for Windows
- Exhibit (*www.simile-widgets.org/exhibit*) – Web-based library of visualization tools
- Many Eyes (*www-958.ibm.com/software/analytics/manyeyes*) – Web-based visualization app service for browser
- D3.js (*d3js.org*) – JavaScript library of visualization tools
- InfoVis (*philogb.github.io/jit*) – JavaScript library of visualization tools
- Leaflet (*leafletjs.com*) – Library of GIS tools
- OpenLayers (*openlayers.org*) – Library of GIS/mapping tools

6.2 Visualization for mobile technologies

One of the biggest trends in visualization of data involves applications for mobile devices. The application of mobile technologies to agricultural data dissemination has enabled agricultural stakeholders and decision-makers to consume data wherever they are. The incorporation of mapping and geolocation capabilities into smart device applications allows customized visualization applications to be sent to users via mobile devices. The ability to handle maps, drill down to information based on a specific region, or use data that is automatically refreshed according to a specific location can lead to a richer data visualization experience for the user. Mobile technology offers a different user experience, as the screens are smaller and handled under different circumstances. This leads to mobile visualizations having to be very simple, yet at the same time very powerful for a user to come back to the application. Often it only works with very few data points communicated to the end-user with stronger visuals on colors and readability.

6.3 Visual analytics

Visual analytics is a branch of computer science that blends analytical algorithms with data management, visualization, and interactive visual interfaces. Visual analytics tools and techniques are used to synthesize information from massive, dynamic ambiguous and often conflicting data; to detect the expected and discover the unexpected; provide timely, defensible and understandable assessments; and to communicate the assessment effectively (Thomas & Cook 2005). Visual analytics consists of algorithms, representations, and big data management. Current state-of-the-art analytics and data management do not yet meet the requirements for big data visual analytics (Fekete 2013), as they cannot yet process enough data rapidly and a suitable way to process them and help in the understanding. Existing software and hardware architectures in visualization, analytics, and data management are not yet suited to integration into visual analytics applications, primarily because big data exploration is relatively new. Eventually, visual analytics could become easier to implement and more widespread, but until then providing the mechanisms for exploration in databases and analysis systems will benefit any situation in which users are willing to trade accuracy for time (Fekete 2013).

Data management and analysis tools have begun to converge through the use of multiple technologies, including grids, cloud computing, and general-purpose graphics processing units. However, data discovery and exploration has not adequately been taken into account in these new infrastructures. New developments in both data management and analysis computation will be required to incorporate visual analytic tools into these infrastructures (Fekete 2013 and Fekete and Silva 2012). These include: (1) Databases will need to be designed to include visual structures within the database, in addition to storing data. These visualization data structures are likely to become a standard functionality in database systems. (2) Standards for specifying multi-dimensional geometric schemas (e.g. specifying location and time) within databases are needed. (3) Databases must support geometry at multiple scales. (4) Extensive caching (i.e. loading chunks of data from the storing machine to visualizing machine) must be supported to maintaining the required computation and rendering speed for visualization structures. And, (5) asynchronous operations are required to analyze large datasets while providing continuous feedback to the user.

The traditional workflow process of loading a file, processing it and computing some analytical quantities may not work well for exploration of large datasets. The analyst may need to try several process and methods in order to find relevant results. With big data, the loose coupling between visualization and analysis presents problems, as data transfer time exceeds the time of time used for analysis and visualization. High Performance Computers (HPCs) have been used to accelerate analytics of big data, but for data exploration, the speed of data throughput may limit the usefulness. Implementations of existing algorithms such as hierarchical clustering and principal component analysis computation have been used to pre-process data. New types of workflows are being developed for use in visual analytics, including reactive workflows (e.g., EdiFlow, Manolescu et al. 2009), which specify that a set of operations occur each time the data change; and interactive workflows (e.g., VisTrails, Callahan et al. 2006), which interactively build and run a sequence including visualizations. VisTrails tracks changes to the workflow so that a provenance for visual outputs is created.

7. Modeling concepts and methods of model development

Building a model is not as straightforward as it sounds. There are many aspects to consider, especially if the model is moving beyond a single scientific discipline. Over the past decades there have been many advances in model development in components, model linking in frameworks, and model interoperability.

7.1 Model creation, composition and reuse

Modeling of agricultural systems is influenced simultaneously by the creators' scientific viewpoints, institutional settings, and by differing views on the relationship between models and software. Alternative perspectives in each of these domains emerged in the early days of the discipline and persist to this day. The physiologically-driven, 'bottom-up' scientific strategy of orderly generalization that is discernable in the Wageningen group of crop models (van Ittersum et al. 2003) contrasts with a more 'top-down', ecosystem-oriented perspective exemplified by the SPUR rangeland model (Wight and Skiles 1987). These different perspectives result in different choices about the detail with which biophysical processes are represented. Even when working from a similar scientific perspective, a scientist who constructs a model as a single individual (for example in a PhD dissertation, e.g., Noble 1975) will follow a different process of model specification and implementation compared to a large team working in a formally managed project (e.g., as in creating ELM, Innis 1978). Researchers who viewed a model as primarily a mathematical system tend to implement them within generic computational packages such as ACSL (Mitchell and Gauthier 1976), CSMP (the Continuous System Modelling Program or Simile (Muetzelfeldt and Massheder 2003), in which the model proper is a document. In contrast, researchers for whom models were engineering artifacts tend to implement them as stand-alone programs (e.g., Ceres-Maize, Ritchie et al. 1991), or as part of ***modeling frameworks***.

7.2 Modularity, components and "plug-and-play" approaches

Over time, the "bottom-up" biophysical models have expanded their scope and the "top-down" models have included greater detail. One consequence has been a clear trend toward modularization of models, both in terms of concepts (Reynolds & Acock 1997) but also in the way they are coded. In a continuation of this trend, several modeling groups have adopted a modular approach to constructing particular simulations as well as the models on which simulations are based, i.e., modularity in the configuration of simulations (for example, Jones et al. 2003 and Keating et al. 2003; Donatelli, et al. 2014). The rationale for this approach to model development is threefold:

i to allow model users to configure simulations containing alternative formulations of biophysical processes, based on the need for a particular level of detail or else to compare alternative sub-models;

ii to permit specialists in particular scientific disciplines to take on custodianship of sub-models, while ensuring that the larger systems model remains coherent; and

iii to facilitate minimization and easier diagnosis of unexpected consequences when a sub-model is changed.

In practice, encapsulation of sub-model logic in components needs to be accompanied by transparency through adequate documentation if the confidence of model users is to be maintained; "black box" sub-models are less likely to be trusted.

As the number of components has increased, it has become natural to assemble them together in order to address more complex problems than the original ones for which the models were developed. While composing large models this way seems both natural and trivial, this is not the case. New limitations are introduced when a model is encoded in a programming language, and seldom are these assumptions represented in the model design or implementation (Athanasiadis and Villa 2013). Models, as implemented in software, do not declare their dependencies or assumptions and leave the burden of integration to the modelers. This situation has been the driving force behind several efforts that have focused on the mechanics of integration, through computerized e-science tools for managing data and software to assist scientists with the technical linking of models to create scientific workflows, for example Galaxy (Goecks et al. 2010), Kepler (Altintas, 2004), OpenMI (Gregersen et al. 2007), OMS (David et al. 2013), Swift (Wilde et al. 2011), and Taverna (Wolstencroft 2013). Integrated modeling so far has been focused on organizational issues (e.g., Laniak et al. 2013) regarding stakeholder involvement, adaptive decision-making, and community engagement, employing mostly training, peer review, and at the same time, reuse of existing models. While holistic thinking and interdisciplinary modeling thinking are still evolving, a methodological and conceptual challenge is how to translate them into ***modeling frameworks*** that produce robust and defensible results, are calibrated with observations, are transparent in methods and calculations, and are useful for answering scientific or policy questions (Janssen et al. 2011). Arguably, currently the science of integrated modeling is not advanced enough to produce the rigor of the results and methods required.

7.3 Modeling frameworks, integration and reuse

A modeling framework is a set of software libraries, classes, and components, which can be assembled by a software developer to deliver a range of applications that use mathematical models to perform complex analysis and prognosis tasks (Rizzoli et al. 2008). A modeling framework enables the "models" to be seen as a single, structured entity rather than a collection of disparate computations. In the case of agro-ecosystem modeling, there are several frameworks that have been developed and used by different research groups, such as ModCom (Hillyer et al. 2003), the Common Modeling Protocol, (Moore et al. 2007), and BIOMA (Donatelli & Rizzoli 2008). Modeling frameworks aim to be domain-independent; however, many of them originate from a certain discipline that drives several of their requirements. For example MMS (Leavesley 2002), OpenMI (Gregersen et al. 2007), and OMS (David et al. 2013) originate from the hydrology domain, and thus allow for component interaction over space at watershed scales.

An important aspect of model integration is information exchange/sharing. State-of-the-art agricultural modeling systems typically employ a blackboard solution, in which common data structures (either classes or components) are shared among models and each model is allowed to read and write values on the shared “blackboard”. The only commonly used contract is that of using the same name for the variable on the shared “blackboard” (see Holzworth et al. 2010 and Donatelli & Rizolli, 2008). A similar situation applies in the case of generic modeling frameworks such as OMS or OpenMI. NextGen models will need more sophisticated semantics for robust coupling of models. Villa et al. (2009) demonstrated that the use of formal semantics can be the key to automatic integration of datasets, models, and analytical pipelines. Athanasiadis and Villa (2013), previewed rich semantics for domain-specific programming languages, which could be considered for NextGen developments.

No consensus on how to implement component-level modularity in agro-ecosystem models can be expected in the near future. While the various frameworks show a strong family resemblance, the differences between them, which reflect different points of departure on the mathematics-to-engineering spectrum and also different views on the trade-offs involved in decentralizing model development, mean that the technical barriers to linking them together are quite high.

Lloyd et al. (2011) compared four modeling frameworks[1] for implementing the same model. They investigated modeling framework invasiveness (i.e., amount of changes required in model code to accommodate a framework), and observed (a) a five-fold variation in model size and (b) a three-fold variation in framework-specific function usage compared to total code size. These findings imply that there is a large impact of the framework-specific requirements on model implementation, and that lightweight frameworks have indeed a smaller footprint. Despite the advantages that modeling frameworks were supposed to deliver in easing software development, they are mostly used within the groups that originally developed them, with very little reuse of models developed by other researchers (Rizzoli et al. 2008). At the same time, modeling software reuse is hindered by other issues such as model granularity.

7.4 Model granularity

The goal of software for integrated modeling is to ensure soundness of results and to maximize model reuse. This can be achieved by finding the right balance between the invasiveness of the modeling framework, as measured by the amount of code change to a model component required to include it in a framework, and the expected benefit of component reuse. A key factor in this balance is the granularity (i.e. break down of a bigger entity into the smaller parts of which it is composed) of the model components. This choice of module granularity involves setting the boundary between one model or sub-model and the next which can be a subjective and subtle process (Holzworth et al. 2010). In a well-designed modeling framework, each state variable and process must be represented in a single granular component class, but it is not always obvious where the boundaries between sub-models should be drawn. Consider, for example, detached plant residues on the soil surface: should they be regarded primarily as part of the organic matter cycle (as in DayCent and DSSAT), as part of the forage available to livestock (as in the GRAZPLAN model) or as a separate part of the system (as in APSIM). Also, if a modeling framework is to support a range of different process representations of differing complexity (for example sub-models for multi-species radiation interception), then the interfaces between the component classes must be carefully designed to be highly generic in the way they describe the relevant features of the system, and also to have unambiguous semantics. This design work, which is essentially a form of conceptual modeling, can improve the clarity of scientific understanding of ecosystems, but it is unavoidably time-consuming and has been considered an overhead by most modelers.

1 The four frameworks were the Earth System Modeling Framework (ESMF), the Common Component Architecture (CCA), the Open Modeling Interface (OpenMI), and the Object Modeling System (OMS).

State-of-the-art crop models tend to have their subcomponents tightly coupled (possibly for better performance), which makes component substitution a laborious task that needs heavy code disaggregation and restructuring, while additional calibration may be needed. Holzworth et al. (2010) present a variety of model-reuse strategies that have been applied in APSIM, from components that were re-implemented, to components that have been wrapped with APSIM-specific code, and to components that were linked via some protocol at a binary level (as black boxes). The variety of solutions was dependent on complexity, implications for other models, and amount of work needed.

7.5 Model linking and workflows

Linking of models and other computations together into larger model chains or workflows is less well developed in the agricultural domain, than are the methods for linking together sub-models of the same kind into a single model. Reflecting on SEAMLESS, the most ambitious such effort to date, Janssen et al. (2011) concluded that integration of different types of models requires (1) conceptual integration, i.e., clear specification of how outputs from one model in a computation should be used to correctly derive inputs to another model; (2) semantic integration, which is a necessary first step toward conceptual integration; and (3) technical integration in software. The most challenging aspect of integrating components across models is that of conceptual modeling. Formal ontologies provide a technology by which semantic integration can be achieved (and linked to datasets) (Athanasiadis et al. 2006).

Generic software tools for linking chains of models into larger computations ("workflows") are now becoming readily available. For example, the bioinformatics community makes extensive use of Galaxy (Goecks et al. 2010), a Web-based tool that provides an interface for composing, executing and storing workflows, together with a repository of thousands of high-level components that can be employed within workflows for tasks such as reading data, converting data between formats or visualizing outputs. To incorporate a model as a computation node, it is necessary to implement it in code that will run without user intervention – which may be a non-trivial task – and then to specify its inputs and outputs in a package-specific configuration file. Some applications of the Galaxy software package have commenced in the agricultural modeling domain. For example, the FACE-IT project (*www.faceit-portal.org*) is applying it to problems in agricultural modeling of climate change impacts.

Experience from SEAMLESS (Janssen et al. 2011), AgMIP model intercomparisons (Asseng et al. 2013; Bassu et al. 2014), and pSIMS (Elliott et al. 2014) suggests that even with these underpinning technologies, a considerable amount of software for converting and translating data between different units, formats, grids, and resolutions will need to be written. In some cases, translation tools will be required in order to integrate data sources and models that do not adhere to community standards. In other cases, translation tools will be required because different communities adhere to different standards.

7.6 Collaborative development

Most of the collaborative development methodologies have been developed by the open-source movement. Open-source is not a prerequisite for collaborative development, however; many closed-source products follow similar methods for project management. The seminal work of Raymond (1999) introduced two major project governance models, the Cathedral and the Bazaar, that still dominate software development in various ways. In the Cathedral model, code is shared only among the members of the development group, and decisions are taken through a strict hierarchy of roles. In contrast, the Bazaar model allows a large pool of developers to contribute with changes over the Internet. In the development of agricultural models, both modes of work have observed. Most projects started with the Cathedral model, and few have changed to the Bazaar.

Where a single organization or a small group of individuals takes responsibility for specifying the design of a modeling system, then the simplest method of collaborative development is a Cathedral approach. The custodian organization arranges with its collaborators for new model elements to be coded in accordance with its internal requirements (often by building a "wrapper" if the new element is pre-existing), and a single code base is maintained by the custodian. Successful examples of such collaboration include the CENTURY and STICS models. The main benefit of the approach is that there is a clear definition of what constitutes a given model at a given time.

Cathedral approaches to collaboration are unlikely to be workable for many elements of a NextGen agricultural modeling system; there will simply be too many peer stakeholders. The 'Bazaar' alternative approach to collaboration introduces the use of a common code repository with a version control system together with social technologies to manage modifications. For example, the APSIM Initiative follows an 'open-source, open-development' philosophy of collaboration. The APSIM source code is managed in a single repository that is visible to the public. Any APSIM user can propose an improvement to any part of APSIM, but the collaboration model is not a pure Bazaar because such proposals are subjected to acceptance by a reference panel (Holzworth et al. in press; www.apsim.info). The DSSAT effort, which has long been a distributed, collaborative effort of scientists and engineers from multiple organizations, is currently moving towards a more open approach with its source code now hosted on GitHub.com. Source code additions and modifications can be contributed, but are subject to rigorous testing and review by a technical board (*dssat.net*).

Collaborative approaches to model development are a consequence of open-source development and carry transaction costs (i.e. meetings, increased communication and increased effort in documentation). The costs to the modeling community of maintaining software quality assurance technologies and governance mechanisms are non-trivial; but the costs to a model developer of joining an open-source community (to translate existing code or adjust to a different conceptual framework) can also be significant. Most important is the cost of paradigm shift. With respect to NextGen Models, creating an open-source project will not be as straightforward as open sourcing an existing individual model or starting a single open source software project from scratch. In merging different communities and scientific approaches, a medium-to-long-term investment by a core group of adopters is vital to achieve the critical mass of benefits that are required to make participation attractive.

7.7 Linking datasets into the model development process

Generally speaking there are four pathways to link models to data: (1) Spatial databases of weather, soil, climate, and other model inputs; (2) Measured data from experimental sources or farm surveys; (3) Global and local datasets of forest cover and land use; and (4) Approaches based on estimating empirical relationships.

In the first case, spatial databases are used for analyses spanning larger spatial scales (for example in Use Case 2), for which there will be a need to access geographic databases of climate, soil and vegetation attributes and of management systems in order to initialize model computations. Systematic methods for estimating local weather and primary soil attributes (e.g., Jeffrey et al. 2001; Sanchez et al. 2009) and land use (e.g., Ramankutty et al. 2008) exist and are undergoing improvement; the main need for new R&D lies in documenting management practices. There is an opportunity to further improve the estimates of spatial biophysical and management information by inference from other data sets, for example using inverse modeling of soil moisture datasets to improve spatial estimates of plant-available soil water content or by inferring the rules that landholders follow to schedule sowing of crops from historical survey data and local weather.

The second case involves measured data, which can also be used to drive simulations of the historical trajectory of agricultural systems for monitoring purposes. Weather data are routinely used in this way; indices of green area derived from remotely-sensed data are another data source that is used in forestry (e.g., Running et al. 1989). Regardless of the kind of input data, the main technical need is to develop readily-accessible repositories that can be accessed seamlessly within modeling workflows. When forcing a dynamic model with measured data, there is an important distinction between running an emulation of a historical trajectory and carrying out simulations of hypothetical or future situations. To validly use measured data as driving variables in the latter two cases, it is necessary to assume that they can be regarded as realizations drawn from the distribution of possible future values of the driving variables. This stationarity assumption is routinely made for weather data, except in climate change studies, but it unlikely to hold for measured soil water or leaf area index data once land management is varied (Wallach et al. 2012).

The third case of land use and cover data is an essential linkage between the development of agricultural models and datasets and is used in evaluation of the predictive capacity of the former and in calibration of their parameters. At the global scale there are substantial databases of permanent forest plots (e.g. Lopez-Gonzalez et al. 2011), but experimental data for other land uses are widely scattered. The main needs for model evaluation purposes are to organize these data sets for other land uses, particularly grasslands and horticulture in subtropical and tropical environments; to collate data sets that allow evaluation of modeled soil water and nutrient dynamics over time-scales beyond single growing seasons and crops, including land use transitions; and to improve the discoverability and accessibility of existing experimental data. The Long Term Ecological Research Network (www.lternet.edu) is a good example of what is possible with existing technology. Over time, we should expect that automated estimation of key model parameters using standard data archives will become a standard feature of modeling frameworks, as will recalibration of models as new data become available. Techniques for automated calibration of complex models have been developed for the hydrologic modeling community (see, for example, Eckhardt et al. 2005 and Duan et al. 1992).

The fourth case, empirical relationships derived from datasets, will also find a place in NextGen workflows. Farm economic modeling and land use/land cover modeling place strong emphasis on data-driven approaches; there may well be scope to employ them in modeling agro-ecosystems, particularly in domains such as pathogen impacts or managers' responses to price variations where process-driven representations are not yet well integrated. The use of data-based sub-models can be thought of as a limiting case of model calibration, where the model structure and parameters are inferred from the data. There is a need for 'generic' model components that can be used to couple such empirical relationships into larger models; at present management-rule languages are used in an ***ad hoc*** way to implement them (Moore et al. 2014).

7.8 The way forward - requirements for the future

There is a real opportunity to realize the challenging aims of NextGen more efficiently through the use of generic workflow management software and a common ontology; i.e. a conceptualization of a system based on concepts and relationships between concepts (Antoniou and Van Harmelen, 2004). For this opportunity to be grasped, there are several prerequisites. First, the agricultural modeling community needs rapidly to reach consensus on which of many competing packages (Galaxy, Kepler, etc.) should be ***the workflow tool*** of choice. If this is not done early, and in a way that most current and future needs are taken into account, experience has shown that different R&D groups are likely to dissipate their effort over a range of platforms with limited interchangeability. Second, a common ontology needs to be selected or developed for the quantities that are input to, and output from models of agricultural systems and landscapes, so that when models are implemented in workflows the values transferred between them are clearly understood. While the ontologies developed in SEAMLESS will be a useful starting point, further conceptual modeling will be required. While an ontological framework may seem an overhead for many researchers, still there is an emergent need for standardization actions for both terminology and data formats (i.e., see Porter et al. in press). Third, it is likely to be efficient to develop a suite of output-presentation components that are in the public domain, well-designed for agricultural problems, and suitable for use across many different NextGen applications. Fourth, clearly separating code that implements model equations from their user interfaces will make constructing computational chains easier.

At the same time, coding a model needs to become simpler. We believe that by providing appropriate domain-specific structures and functions as libraries, we can enable NextGen model implementations that are significantly smaller, in terms of lines of model-specific code, than today's models. As programming and maintenance costs tend to scale with lines of code (Boehm 1987), such developments will have a considerable positive impact.

8. Infrastructure and interfaces: State-of-the-art of IT

Ultimately, any end user needs to interact with models and tools in a stable and robust way. For stability and robustness of tools and models an infrastructure is required, while the point of access for the user is user interface, which can take many different forms.

8.1 Interfaces for end users

Much consumer and business software today is not installed on PCs but is instead delivered by cloud-hosted services accessed over the network from a Web browser, often running on a mobile phone or tablet. Intuitive Web 2.0 interfaces make user manuals largely a thing of the past. The result of these developments is a considerable reduction in cost, often an increase in usability, and above all a dramatic increase in the number of people who make use of advanced software. For example, 20 years ago, few people had any sort of mapping or navigation software on their computer; now everyone has access to Google and Apple Maps on their computers and phones, and many use those services regularly.

Steinfield and Wyche (2013) present a summary of mobile phone-based agricultural services currently available, with a focus on developing countries. These applications were sorted into four categories: (1) farmer advisory and information services, which provide extension-type agricultural information to farmers (example: *www.m-kilimo.com*); (2) market information services, which provide information via text message on current market prices of crops upon request (examples: M-Farm, *mfarm.co.ke*; Esoko, *www.esoko.com*); (3) financial services, which allow people to make and receive payments over their phones (example: M-PESA, *www.safaricom.co.ke*); and (4) decision support services, which include a range of services for mobile devices to collect information from farmers then provide prescriptive information to support decision-making (example: iCow, *icow.co.ke*). Apps available to farmers in developed countries are even more numerous. These fall into the same general four categories outlined above, but in many cases include far richer data stores for weather, soils, chemical applications, pest and diseases, and precision farming applications.

There are numerous examples of how applications for mobile devices and desktop computers are already helping farmers to gain information to improve their ability to generate farm income in sustainable ways. Many of these applications are being developed in the private sector for use directly by farmers. Social network feedback from users often provides assurance that the apps are useful and reliable.

The technology for delivering information through mobile and other devices has rapidly developed in a short time. It is likely that when the appropriate knowledge and information products become available, private and public software developers will already have the appropriate tools to implement the required applications to serve those data products to stakeholders.

8.2 Data and model discovery

Use of big data, component-based models, synthesized information products, and apps for delivering knowledge through mobile devices, as proposed components of the NextGen ICT infrastructure, can all be easily envisioned using technologies that exist today. One of the grand challenges of the NextGen ICT framework will be to provide common protocols for making these numerous databases, models and software applications discoverable and available to users and developers through web services and distributed modeling frameworks. By using common semantic and ontological properties in Web 3.0 interfaces, the data and modeling components can be made available for coherent use in a proposed NextGen platform. Web 3.0, also called the Semantic Web, refers to a standard being developed by the World Wide Web Consortium (W3C) that allows data to be discovered, shared and reused across applications, enterprises and community boundaries (*www.w3c.org*). Linked data refers to connections between the contents of datasets to build a “web of data”. This technology is relatively new and as yet unproven for practical use in the scientific and big data realms. Many claims about the potential of linked open data (LOD) using W3C protocols do not accurately portray the complexity of designing these systems. Tools for working with linked data are not yet easy to use and few people have access to the technology and skills to publish linked datasets (Powell et al. 2012). However, despite the lack of maturity of this technology, it holds great promise for use in a distributed ICT modeling framework, provided that the Web 3.0 protocols continue to develop and coalesce around common standards and that tools are introduced that allow more rapid development of customized and complementary ontologies.

Other elements of a distributed modeling framework, such as cloud and web-based computing, movement of big data across the web, and software-as-a-service (SaaS) are already in wide use. Standards have been established for many aspects of cloud computing (e.g., Open Cloud Computing Interface and Cloud Data Management Interface). However, standardization gaps still exist for many other areas as delineated by the National Institute of Standards and Technology (NIST, 2013). These standardization gaps include SaaS interfaces for data and metadata formats to support interoperability and portability and standards for resource description and discovery. The NIST report lists 15 groups that are actively working on development of standards for all aspects of cloud-based computing.

9. Analysis of Use Cases

In this section, we view the components of a next generation modeling infrastructure through the lens of the five use cases. In each case, deficiencies in the current modeling systems are recognized and a straw man NextGen solution is proposed.

9.1 Use Case 1 - Farm Extension in Africa

9.1.1 Problem statement

Jan is working as a farm extension officer in an area in Southern Africa where many farms are very small, incomes are very low, and farmers typically grow maize and beans as staple crops for their family's subsistence and to sell for cash. Some households may have livestock and/or grow vegetables. The aim of the extension service is to help farmers achieve higher and more stable yields of maize and also to advise them on improving their nutrition so that they obtain sufficient protein and micronutrients for healthy families. Jan obtains information on new varieties of maize and beans that are now available to farmers in the area. These new varieties are more drought and heat-tolerant, and the bean varieties are more resistant to a common foliar disease. Jan also has information on how to improve nutrient management of these crops using small doses of inorganic fertilizer along with animal manure and crop residues. He also has information on a new technique developed by the CGIAR to partially harvest rainfall to increase water availability to the field and vegetable crops.

9.1.2 Current deficiencies

Jan is not a modeler, but he can benefit from the outputs of agricultural production models and farm-scale economic models. In this case, the modeling system must be able to combine existing data about localized conditions (soils, weather, genetics, household economics, local markets, etc.) with farm-scale models to predict the viability of using the new varieties. These data are often very difficult to access if they exist at all. In many areas, weather data are considered to be proprietary and are not distributed freely. Good quality, localized soil data suitable for crop production modeling are usually non-existent or available only at a scale that is not practical at a field level. Information about household demographics and economics is rarely available except in cases where a research survey has been conducted recently in the area. In any case, these data may contain sensitive information, which should not be made publicly available until the data are anonymized. Pre-configured models appropriate to the smallholder systems of this region are needed, including components for mixed livestock/cropping systems driven by data relevant to management practices in common in the region (e.g., planting date "rules", cropping densities, varieties cultivated, etc.).

The current infrastructure would not allow an easy answer for Jan. Existing models can simulate such systems, but the required data collection and model configuration would require a considerable effort and collaboration with modelers and primary data collectors. Studies such as this, using current technology, would typically be performed to give generic suggested management for a region, with results not tailored to individual farms.

9.1.3 Straw man proposal

Because farms vary in size, labor availability, soils, and other characteristics, Jan wants to use the NextGen tools to help tailor advice to each farm family that is practical, likely to be adopted, and provide the best outcome in terms of more stable production, higher income, and better nutrition. Jan obtains information from the farmer to input into his smart phone, which has NextGen apps that were developed for the farming systems of his region and that help him determine combinations of system components that might best fit specific farm situations. This software also provides print files for extension information sheets written in the local language that describe the components of crop and farming systems that are likely to succeed with the farm family. The following design-time narrative and Figure 5 describe the components of the data, modeling and delivery infrastructure that could allow Jan to deliver the necessary information to the smallholder farmers that he serves.

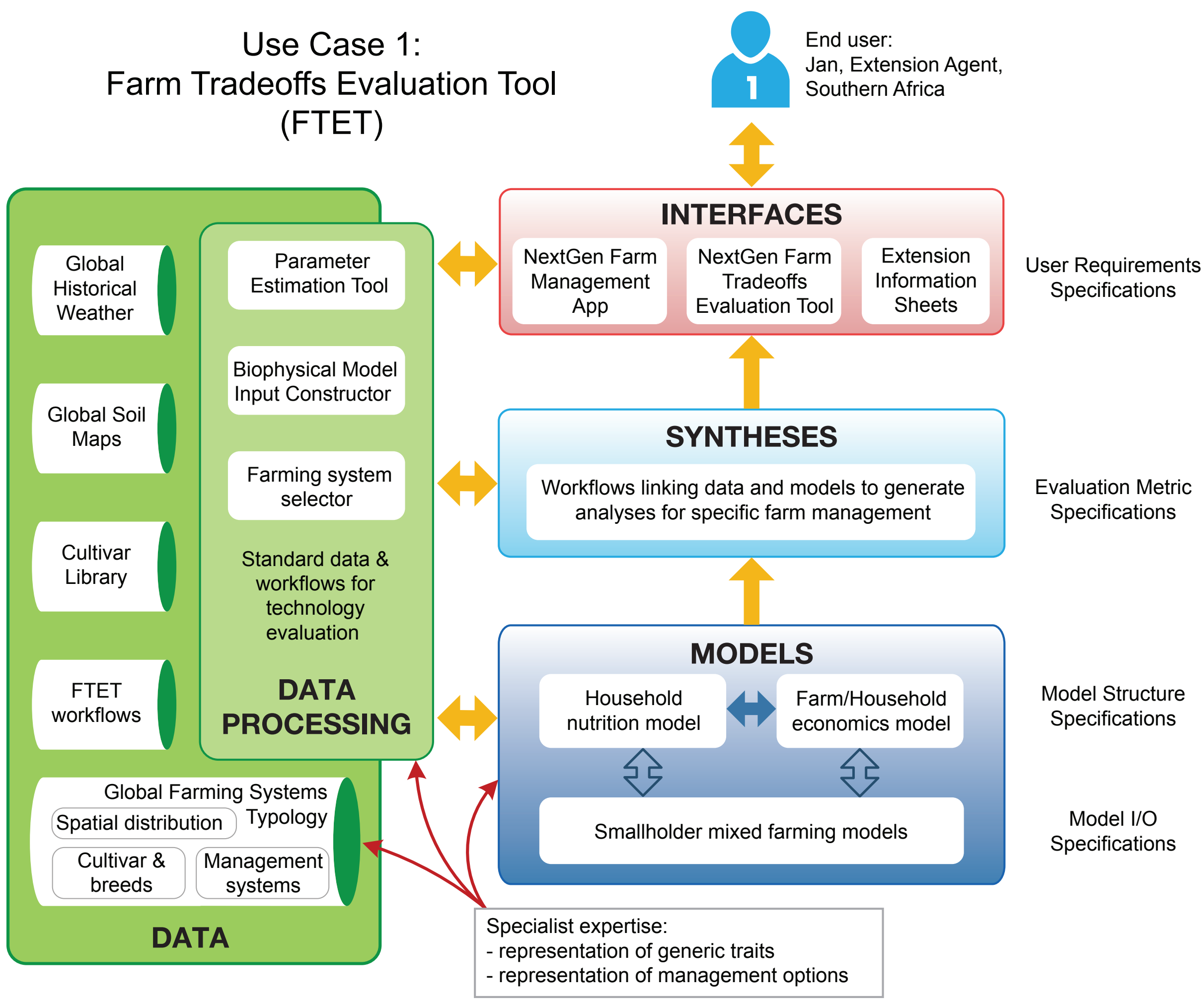

Figure 5. *Straw man proposal for Use Case 1: Farm extension in Africa*

1. Jan has used the NextGen apps previously for evaluating improvements to cropping system management and so he is already familiar with the user interfaces and options available. He uses the **NextGen Farm Tradeoffs Evaluation Tool (FTET)** for use in evaluating the efficacy of the new varieties.

2. The improved varieties of maize and beans have been developed by scientists at the CGIAR centers, who work closely with the **NextGen cultivar library** and have used the **NextGen parameter estimation tool** to develop crop model parameters for a **suite of NextGen models** for the new cultivars. These cultivar parameters are now stored in the cultivar library and are available for use in the NextGen suite of applications.

3. Jan obtains information from the farmer and inputs these data into the **NextGen Farm Management App** on his smartphone, which has an interface developed specifically for the farming systems of his region. The app will help him determine combinations of system components that might best fit specific farm situations and register these management systems within the **Global Farming Systems Typology Database.**

4. **Soils and weather records** specific to the farm locations in Jan's region are already available in the NextGen database for use with the FTET.

5. The FTET is a workflow that was generated for evaluating tradeoffs between management decisions and overall farm/household level profit and nutrition. Components of the tool include farm production using biophysical models, a nutritional analysis based on inputs and outputs to the farm, and prediction of household income under each scenario. Jan's input data from each household and the proposed improved varieties can be added to the workflow using the FTET user interface.

6. Based on outputs from the FTET, Jan distributes and discusses **extension information sheets** written in the local language that describe the components of crop and farming systems that are likely to succeed with the farm family.

9.2 Use Case 2 - Developing and evaluating technologies for sustainable intensification

9.2.1 Problem statement

Debora is a plant breeder/geneticist at CIMMYT, the International Center for Maize and Wheat Improvement, working on developing a drought- and heat-tolerant hybrid of maize for. She would like to be able to evaluate the potential adoption and impact of maize varieties with particular characteristics across the widely varying conditions in Africa where maize is an important crop. She realizes, however, that maize is only one part of the complex farming systems used by most farmers, which typically involve multiple crops and livestock. She would like to be able to evaluate the potential of new varieties in these complex systems, rather than evaluating maize by itself as had been typically done by most research programs. Moreover, she would like to know whether the new varieties meet goals for sustainable intensification, such as improving productivity not just in the short term, but taking longer-term impacts on soils, water, and greenhouse gases into account.

9.2.2 Current deficiencies

Rapid analysis of this type would require the use of a database of farming systems typologies that would contain information about the types, management regimens, and frequencies of occurrence of crop and livestock production systems in the region. A crop variety library linked to both genetic markers and crop modeling parameters for multiple models would allow Debora to create model parameters using the genetic characteristics of the new varieties. Regionalized soil and weather databases, with data suitable to drive these models would be needed. Mixed cropping-systems models, integrated with livestock systems and household economic models would be required to generate predictions of production and income over a region or country. If Debora wants to estimate the uncertainty of the model predictions, she may choose to generate the production estimates with simulations by multiple models.

Again, current infrastructures are inadequate to rapidly and accurately address this use case. Gene-based modeling is in its infancy (White and Hoogenboom 1996; Hoogenboom and White 2003; Messina et al. 2006; Hammer et al. 2006), but is progressing rapidly. The modeling technology may be adequate already. However the farming systems topology database and cultivar libraries described in the use case do not currently exist. Existing soils and weather data suitable for regional analyses may be available, but are often not usable directly by crop models due to missing data and incompatible formats.

9.2.3 Straw man proposal

Working with a team of colleagues at her research institution, she uses the NextGen Technology Adoption and Impact Tool (TA&IT) for this purpose. This tool integrates the genetic characteristics of the maize varieties with soil, weather, economic and social data representing the farm populations where the new varieties could be used. Jan's recent contributions to these databases have been used to update them seamlessly. The research team then simulates the potential for adoption and impacts of the new varieties, providing Debora with guidance for the kinds of genetic modifications that would be most valuable to farmers, and also provide an assessment of the long-term sustainability of the systems.

Figure 5 and the following design-time narrative explore the possible components of a solution for Deborah's use case.

1. Debora sees herself as a "crop breeder, not a modeler" so she engages with a colleague (Eduardo) who has previous experience with the **Technology Adoption and Impact Tool** (TA&IT). Like the models it uses, the TA&IT is more a collection of workflows and components than a single piece of software.
2. Eduardo scopes out the analysis requirements with Deborah. By using the query tool associated with the **Global Farming Systems Typology** (GFST) database, they confirm Eduardo's initial judgment that a farming systems typology at an intermediate level of resolution will provide an appropriate balance between detail and computational effort. Consistent descriptions of approximately 200 **representative farming systems** (RFSs) across sub-Saharan Africa that depend on maize production are selected, along with stored estimates of the populations they support
3. Fortunately for Debora & Eduardo, a previous project on conservation farming practices has developed and **recorded templates** (instructions) for translating GFST descriptions into formats needed by several **biophysical models** of smallholder farming systems. The biophysical model configurations include multiple crops, livestock, soil C, N & P cycling and representative diversity of soil types, plus responsive land use and tactical management systems that reflect households' human capacity, and household consumption of product. They are parameterized from the GFST.
4. Eduardo adjusts the templates to place more emphasis on in-crop management of maize, and swaps in a livestock production model that will respond to quality changes in corn residues fed to animals. He chooses the latter from a **library of several livestock production models** that share a common within-simulation interface.
5. Debora has seen the results of model intercomparisons and wants to be reasonably sure the analysis isn't dependent on the choice of maize model. Eduardo therefore builds three versions of the mixed farming model, using **3 different maize growth models**. To include the third maize model, he has to write a small **OpenMI translator widget** but, working from the example already available for maize model #1, this only takes a day or two. Eduardo adds this work to the ever-growing interface repository.
6. Meanwhile, Debora and the crop physiologist in her team have been describing the target traits that are to be evaluated in terms of their physiological function. Eduardo then arranges teleconferences with Fritz, Gemma, and Hortense, who are expert users/developers of each of the maize models. At these meetings the best ways to describe each trait in terms of parameters of each model are decided.
7. Standard **translator Web services** are built into the **modeling workflow**, so that simulations can

access weather and soils data for each RFS model from consistent global databases and convert them into formats that are useable by the models.

8. A set of preliminary runs are carried out using current maize genetics and are checked (using standardized reports) to ensure that inter-annual distributions of maize production, other crop and livestock production, cash and labor budgets are sensible. This step goes smoothly because these "base" models are being re-used.

9. Output from the biophysical models, along with RFS-specific context, are piped to **cash and labor budgeting modules**, an **assessment of** changes in the adequacy of **household diets and behavioral information about adoption**, and a semi-quantitative tool that accounts for factors (cost, complexity, etc.) that will influence the **adoption** of new genotypes.

10. Eduardo advises Debora on the design of a simulation experiment that varies each target genetic trait against the existing genetic background and management system of each RFS. They also decide to include **systematic variations in fertilizer input intensity** and the **proportion of land devoted to maize**, in case Genetics x Management x Environment (GxMxE) interactions are important.

11. Because the standard TA&IT "trait evaluation" workflow is being adapted to this analysis, Eduardo has a straightforward job of **synthesis** once the analysis is run. A **collection of useful statistics and presentations** has already been prepared (adoption rates, maps of where the net benefits are highest, differentials in effect on richer and poorer household types, income-risk-natural resource management tradeoffs) and the TA&IT-specific visualization tool allows him to show the results to Debora for her interpretation.

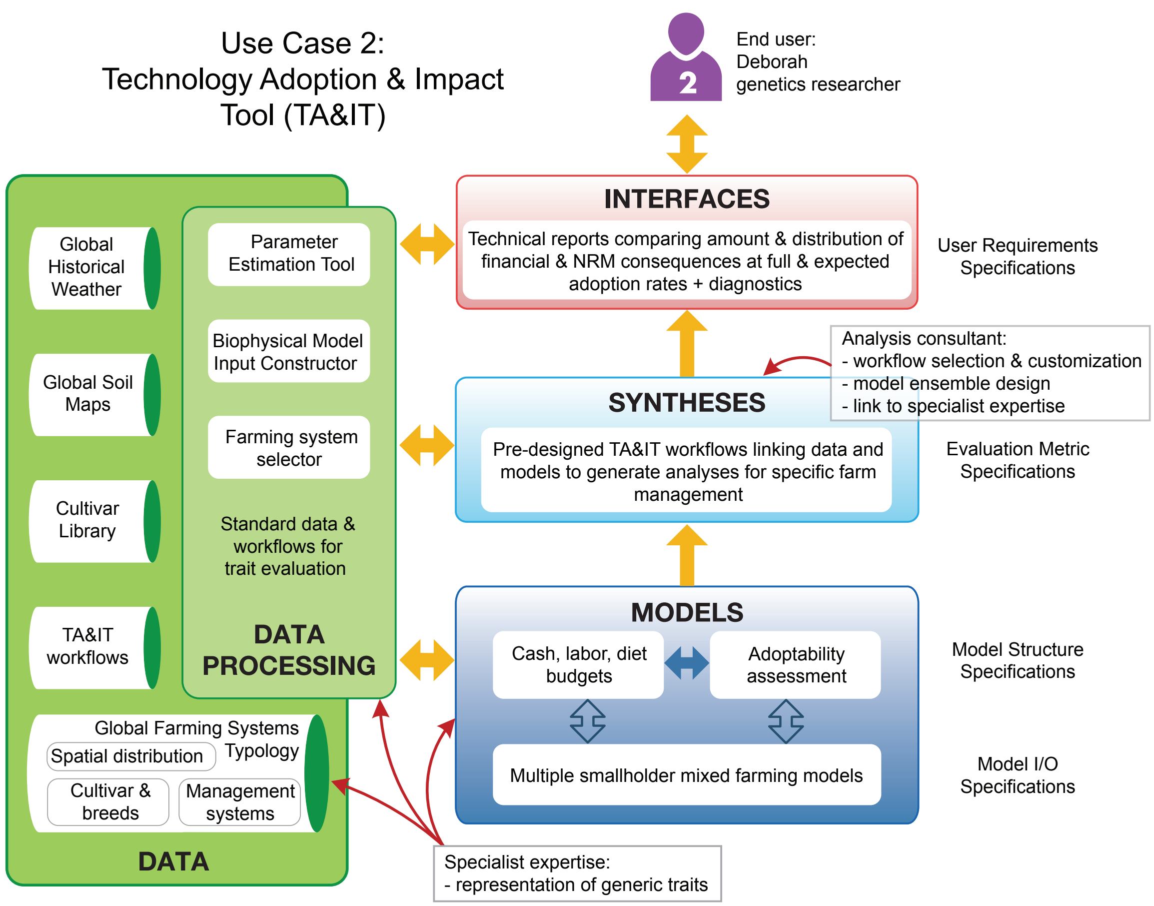

Figure 6.
Straw man proposal for Use Case 2: Developing and evaluating technologies for sustainable intensification

9.3 Use Case 3 - Investing in agricultural development projects that support sustainable intensification

9.3.1. Problem statement

Stanley is an investment manager for a prominent foundation, and he needs to evaluate a project proposal for small farms in Kenya that will increase the intensity of production by increasing fertilizer use per hectare on cash crops while maintaining the current sustainable nutrient balance between pasture grasses, crop residues and animal manure. Before authorizing a project that combines extension information and fertilizer subsidies, Stanley wants to evaluate whether the higher crop yields would induce a non-sustainable system once the initial period of fertilizer subsidies and extension was completed.

9.3.2 Current deficiencies

In current agricultural systems modeling, many models (i.e., livestock, crop production, economics, soil, water) would need to be linked and different pieces of information extracted from them, and then combined to produce relevant indicators, mostly done by the researcher or analyst personally, or even a combination of researchers from different domains, with fragmented access to data. Data sources used, their aggregations, and their availability would be managed by each modeler for each model, which creates a confused situation in that the same data are used for different models, but with different formats. A significant coordination effort is required to succeed at all, which focuses on the different modelers applying the models, getting all the relevant data, and ensuring some consistency and interchangeability across model applications.

9.3.3 Straw man proposal

Stanley implements the NextGen Project Assessor Tool (PAT) to access data and crop and livestock model components to assess the yield and labor impacts of increased yields. An economic assessment model is used to estimate if the current cropping balance will change under the new fertilizer program and if increased fertilizer costs can be more than compensated by increase in cash crop yields in the long run. A long-term farm level nutrient balance under increased intensification will show whether the new system is sustainable. Stanley would like to evaluate these results under a range of assumptions, and present these to local decision makers so that they share common expectations and uncertainties. For this he uses the Next Gen Project Assessor, which opens as a web page on his computer, and he sets up a new assessment, enters data supplied with the project proposal, and links this to general data layers available in the tool. The Next Gen Project Assessor uses multiple Next Gen Models and the Global Farming Systems Typology as tools for impact assessments. Figure 7 illustrates the components of the proposed solution, which includes the steps listed below.

1. Ideally in the next generation modeling framework the necessary data would be available on a common platform in the NextGen Project Assessor Tool (PAT). Data (all formatted to a common data dictionary/ontology) include traditional sources like **household surveys**, **field experiments**, **regional statistics**, but also **crowd sources** of recent estimates of biomass growth and disease spread and **remotely-sensed images** of field distribution and water availability.

2. **Data assimilation** through models could subsequently be executed with summary processes described in the NextGen Project Assessor library, or by an export to more comprehensive modeling tools following standardized export formats.

3. Ultimately, in the Next Gen Project Assessor, a library of tools can be used to present the best combination of **integrated indicators**, describing the likely impact of the measures proposed in the project in an appealing way for external stakeholders building on state-of-the-art visualization software.

4. Here it is unlikely that Stanley does all these analyses himself, but instead he invites an analyst to do this for him, once he has formulated his question in the start page of the Next Gen Project Assessor.

Crucial to success of the NextGen Project Assessor are: (1) ample availability of good quality data that can be used freely for analysis; (2) easy integration across domains of data and analysis tools without excess in details of any particular domain; and (3) flexibility in import and export of data according to standardized formats to facilitate sharing and visualization. From an ICT point of view, this would require innovations in: (a) data discovery with a necessity to make data searchable and easily transformable, even if residing physically at many different locations across the globe; and (b) easy mash-up of data from different sources using standardized analysis tools in a web-based platform accessible to many different users with different roles with a strong presentation layer.

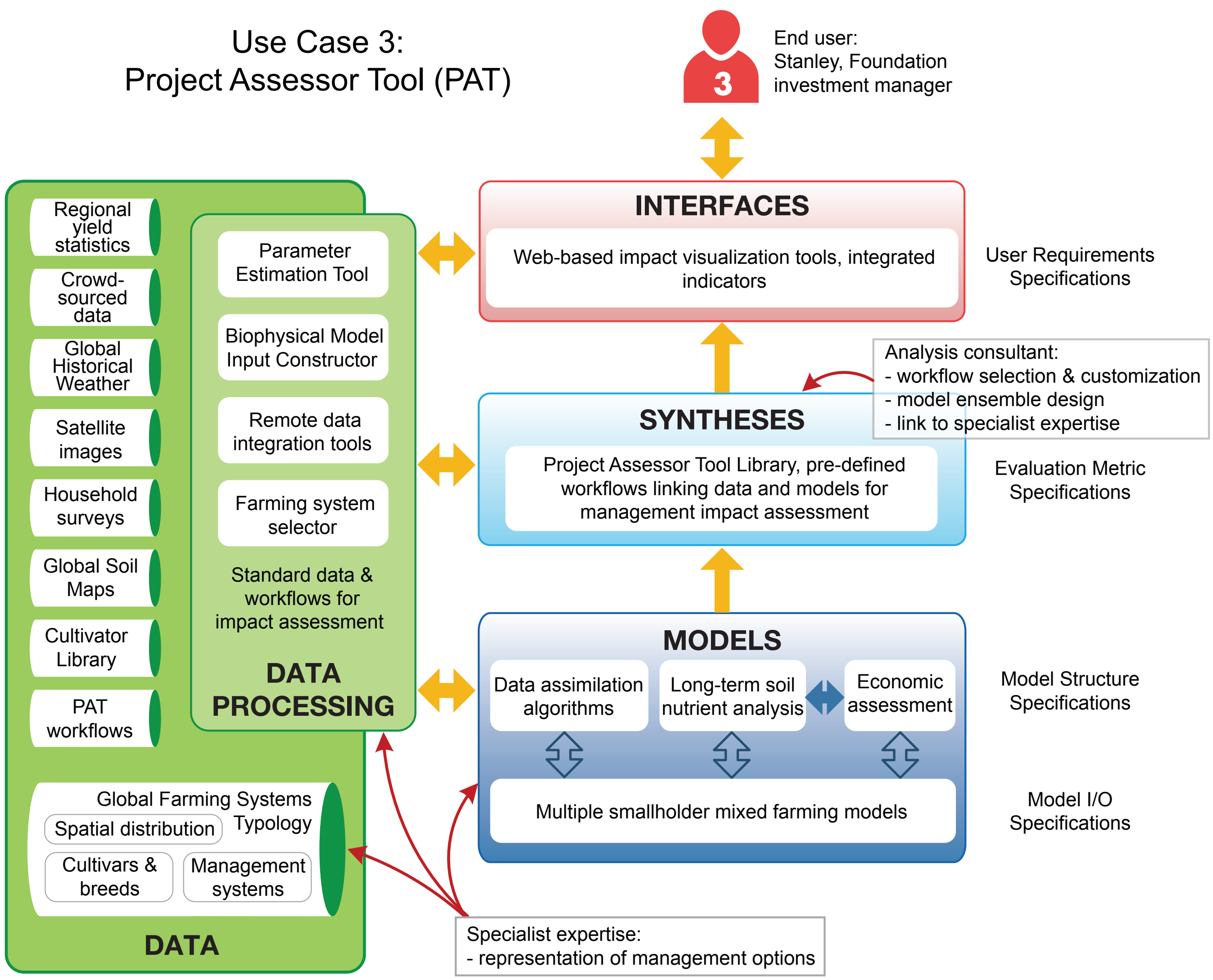

Figure 7. *Straw man proposal for Use Case 3: Investment in Agricultural Development to Support Sustainable Intensification*

9.4 Use Case 4 - Management support for precision agriculture

9.4.1 Problem statement

Greg is a farmer in the US, with a large corn/soybean-based farm and a high level of mechanization fully equipped with auto-tracking system and high-resolution differential GPS. Greg has a historical archive with more than 15 years of data on crop yield spatial variation at five-square-meter resolution. His tractors are equipped with on-the-go sensors for variable applications of seeding, fertilizer, pesticide, and herbicide. Greg consults with a precision agriculture consultant, Harold, who provides him with up-to-date management prescriptions, tailored to the current crop stage and soil variations in his fields. These strategic and tactical crop management recommendations include variable rate prescriptions for fertilizer/pesticide/herbicide application and accurate harvest recommendations that are automatically integrated in Greg's Controller Area Network (CAN)-bus enabled tractor for variable rate application of inputs.

9.4.2 Current deficiencies

Much of the information that Greg and Harold need to realize this use case is available with current technologies, including remote sensing through high-resolution satellites, airborne imagery and drones; farm equipment designed for precision agriculture; and existing crop modeling and data analysis software. What is currently lacking is the ability to link remotely sensed data with a decision support system that allows a farmer to make informed decisions regarding management required for specific locations in a field or farm, translated to the on-farm GIS-equipped machinery. Crop models allow useful extrapolation and prediction for prescriptive management, but most current crop models lack the ability to handle spatially connected processes (i.e., water flow, weeds, and pest dynamics) within a field or landscape. Use of the models with real-time, remotely sensed data is not currently available to farmers or farm advisors.

9.4.3 Straw man proposal

Greg receives weekly updates on his smart-phone and tablet from Harold's Precision Agriculture Company about the status of his crops. Information contained in these updates is obtained from drone flights and crop model predictions using a combination of observed and forecasted weather. Harold's analysis relies on the next generation models that are able to deliver strategic and tactical crop management recommendations, process-based variable-rate prescriptions for fertilizer/pesticide/herbicide applications, and accurate recommendations on harvest management. The variable-rate prescription map created by Harold's company is cloud-based and is automatically integrated in Greg's CAN-bus enabled tractor for variable rate application of inputs. Greg's farm technologies allow him to trace back all the activities performed in the field and the harvested product.

The following design-time narrative and Figure 8 describe the components of the data, modeling, and delivery infrastructure that could allow Greg to implement GIS-based precision farm management.

1. Harold's Precision Agriculture (HPA) consulting business maintains **high-quality, high-resolution soil attribute maps** for Greg's fields. These data are considered to be **proprietary**, as Harold and Greg have invested in the collection of the data specific to these fields. NextGen **soil data harmonization tools** were used to prepare the data in a format that can be used by multiple crop models. Soils are tested annually at several locations in the field and the files are updated as newer or corrected data become available and data are stored on a cloud-based server.

2. HPA buys **high-quality observed weather data** from a service that also prepares **ensembles of seasonal forecast weather data**. These data are combined with **rainfall data** recorded in Greg's fields. These cleaned and combined data are also provided in a **harmonized format** ready for use by crop models. HPA serves the observed weather data on public servers as **linked open data**, which are then available for discovery and use by other NextGen models. The ensembles of weather

forecasts are proprietary, and although stored in the same formats, are not made available as open data.

3. HPA owns several drones, which are flown over Greg's fields bi-weekly to generate precision maps of leaf area index, biomass and chlorophyll content, which can be converted into site-specific nitrogen uptake by crops. A **history of** these **aerial field maps** can be retrieved to generate time series for the crop growth at any point in the field.

4. Greg's CAN-bus enabled tractor allows Greg and Harold to archive all precision management data for the variable rate application of inputs to Greg's fields. HPA maintains **proprietary software** that converts these management data to crop model-ready formats.

5. HPA's crop modeling staff have pre-configured the NextGen **Precision Agricultural Management Tool** (PAM-Tool) to generate an ensemble of crop growth simulations for Greg's fields using the detailed soil maps, observed weather data, ensembles of seasonal weather forecasts, cultivars planted and management history. A **data assimilation package** is included in the crop model used for simulation so that estimations of in-season biomass and LAI can be improved using Bayesian filtering techniques.

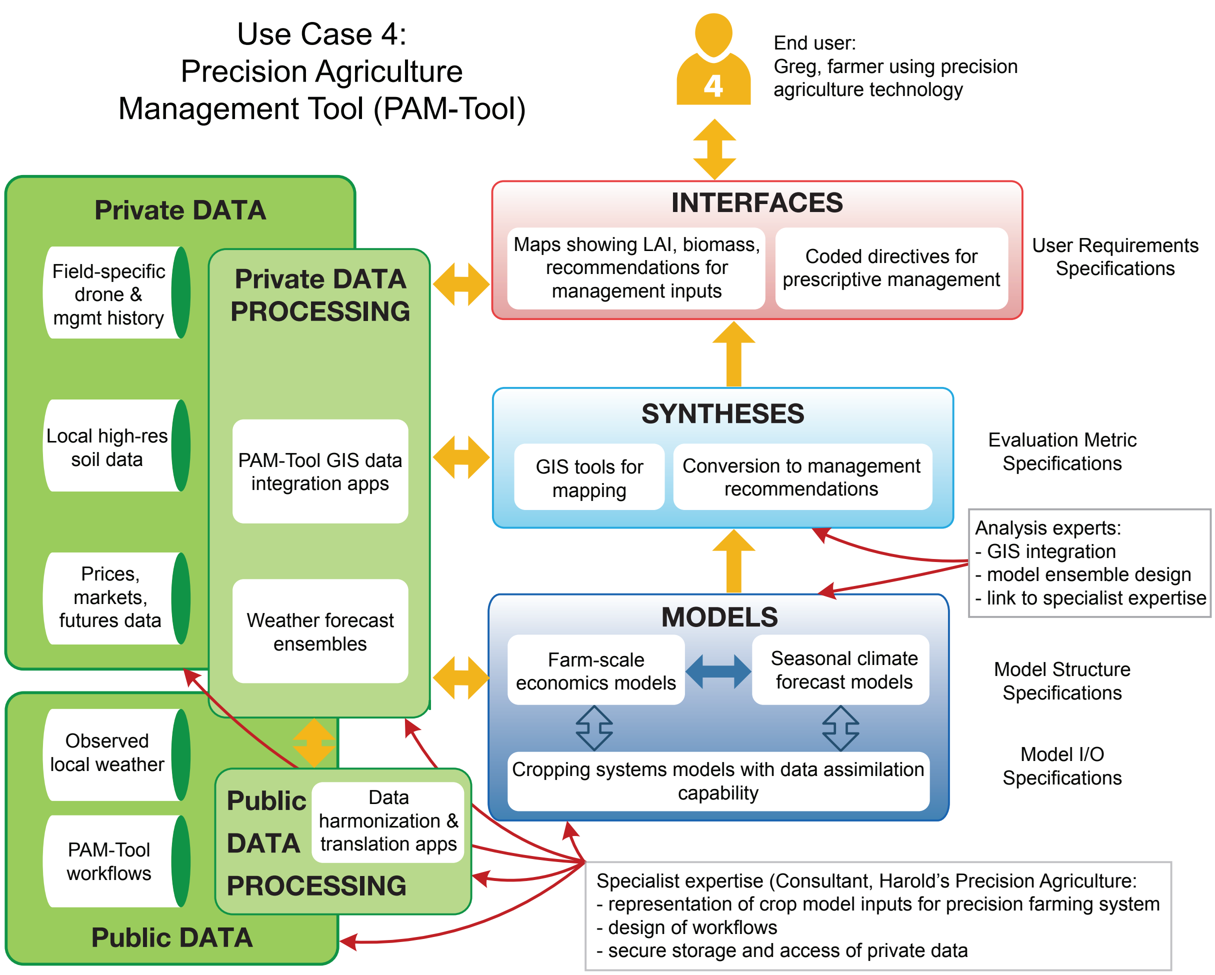

Figure 8. *Straw man proposal for Use case 4: Precision agriculture*

6. The PAM-Tool working on a **high-performance cluster** quickly identifies optimum irrigation scheduling; fertilizer, pesticide and herbicide applications; and harvesting schedules. These recommendations are sent via smartphone connection for review by Greg and Harold. If approved, the recommendations are sent to the tractor's computer for implementation.

7. Greg's use of the PAM-Tool relies heavily on his soil attribute maps. These are derived from his proprietary data using inference algorithms that were created by publically-funded research and are made freely available. As matter of enlightened self-interest and of citizenship, Greg agrees to a request to make his soil attribute data available (via a standard soils description Web service) for the quarterly re-estimation of the empirical parameters used by these algorithms.

9.5 Use Case 5 - Supplying food products that meet corporate sustainability goals

9.5.1 Problem statement

Jennifer is an economic analyst with in a corporate sustainability group. This group has embarked on efforts to make sustainability the core of their mission: marketing food while conserving resources. She is assessing the lifecycle of food products to find ways to conserve energy, save water, minimize waste, and reduce greenhouse gas emissions in an effort to make these products more sustainable from the farm to fork. Specifically she is looking at the potato chips division as a case, having in mind that the corporation set an ambitious reduction target for greenhouse gas (GHGs) emissions by 2020. She wants a monitoring and evaluation system, in which she can track the different sources of emissions, synthetically test interventions, and follow the year-to-year variability in the emissions.

9.5.2 Current deficiencies

Current modeling of supply chains mostly uses life cycle analyses that have little connection to landscape or field level modeling. Innovations in modeling are required to bring new algorithms forward; that can also parse data near-real time along the supply chain and that can compute implications in nutritional content of the different food stuffs. Data from remote sensing, climate scenarios and weather forecasts are not extensively used to predict supply chain implications using state of the art models. Also, the combination of data and information from private large corporations and public sources only occurs on the premises of the private corporation.

9.5.3 Straw man proposal

Figure 9 and the following design-time narrative present a proposed solution for Use Case 5 that uses the NextGen **Supply System Assessment Tool** (SSAT). The SSAT is initiated and implemented by a consortium of large agribusiness companies. In recognition of their common interest in the transparency of their corporate sustainability policies, the consortium place the SSAT code in a public software repository and invite proposals for improvement to it

1. Using a Web service, Jennifer works with her analysis team to access the SSAT. This tool monitors and visualizes the GHG emissions in the supply chain at the different steps in the supply chain. To use the tool, Jennifer and her team first configure it with information for their chips **supply chain** with factory locations, approximate location of farmers delivering input for the chips, and transportation moves to and from the different locations in the supply chain.

2. The tool offers **real-time weather** and **historical climate** conditions around the globe as standard information, together with **exchange rates, trade flows, soils, population densities, and GDP**.

3. Jennifer discovers that most GHG emissions occur in crop production, so she sets out to identify strategies that will optimize the amount of fertilizer to be used at a particular location with the goal of increasing yield and reducing greenhouse gas emissions.

4. Through a web service with YieldGap.org, she imports data on **yield gaps** for crucial crops in

the most promising production regions for her corporation, and estimates the room for improvement in yield, while at the same time getting information on **nitrogen application rates** and **irrigation** techniques from the local supply chain contacts and the Global Farming Systems Typology.

5. With information from remote sensing and seasonal weather forecasts, that produce **yield forecasts**, she designs different management strategies for the current season, with estimates of the expected yields and GHG emissions, also with timing for harvest to optimize transport movements.

6. For Jennifer, the NextGen Supply System Assessment Tool also computes a total of **GHG emissions saved** over the supply chain as an estimate. Proposals for management practices to be included in supply contracts with farmers are presented and discussed with the local supply chain managers of the corporations; the resulting local knowledge is incorporated as changes in the SSAT inputs data and new simulations are carried out before the final supply contracts are prepared.

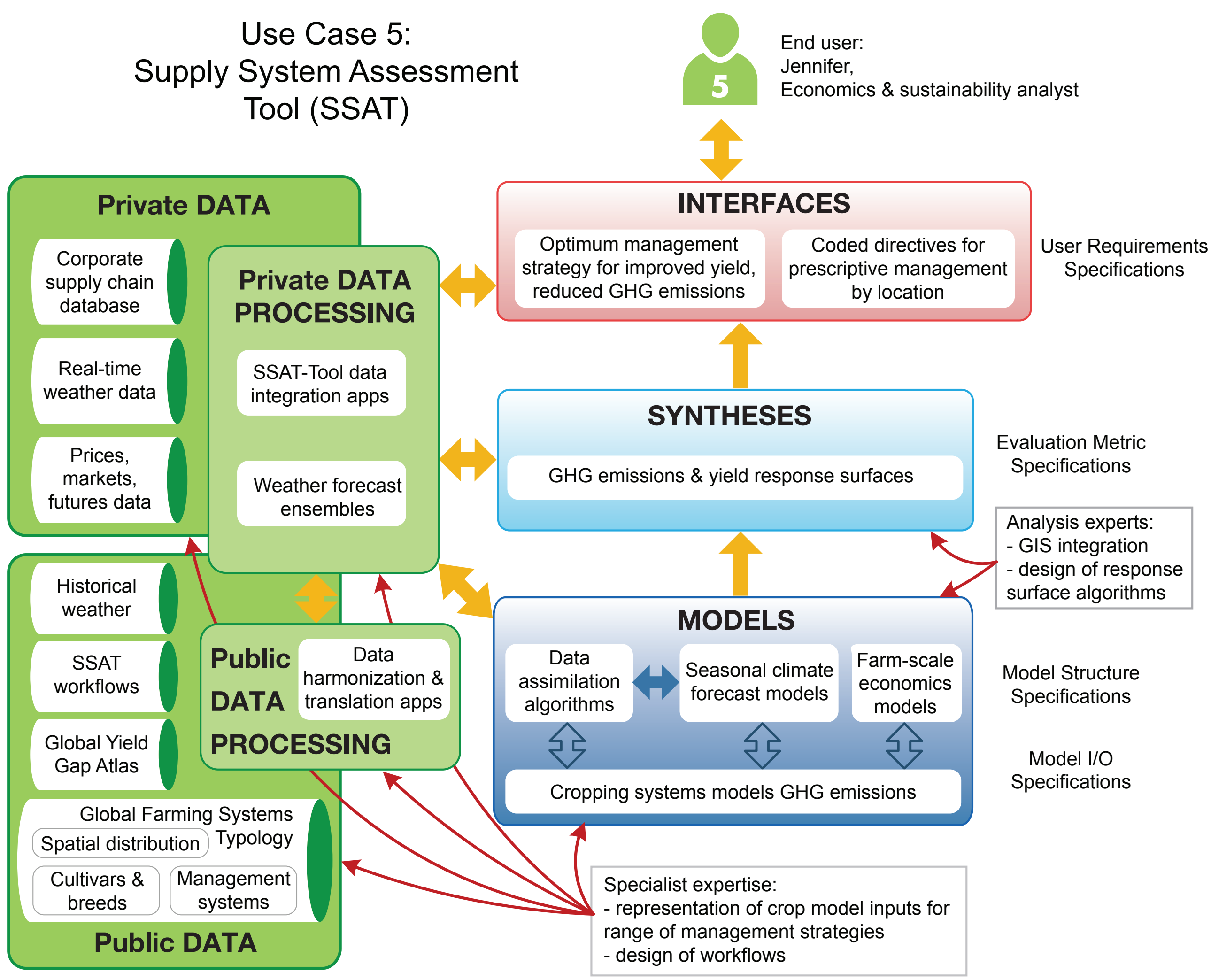

Figure 9. *Straw man proposal for Use case 5: Supplying Food Products that Meet Corporate Sustainability Goals*

10. Conclusion and recommendations for future development

10.1 Overall analysis of use cases

An overall analysis of the use cases demonstrates that these require a mix of recent innovations in technology and data (Table 1) to be realized. This overall analysis is necessarily qualitative, and dependent on the straw man proposals as these have been developed (which could have been done in many different ways). In general there is an emphasis on data integration, in the broadest sense with data coming from different domains, from different sources, and by combining private and public data. All of the use cases thus require a more intensive use and combination of data, which so far has not frequently occurred. Big data, and with it visual analytics, become important in only one of the use cases. This could also be due to a lack of familiarity with big data to foresee the possibilities.

Most uses cases depend on the availability of good quality, openly available data, suitable for use in modeling applications. Soil data, for example, are needed that have relevance to localized agricultural fields, are complete, and are suitable for use in crop models. Similarly, good quality, openly available, up-to-date weather data are needed, which are complete and ready for use by models without the intervention of a data preparation specialist. These data requirements would appear to be the low-hanging fruit of any modeling system, but they are often surprisingly difficult to obtain in today's modeling world. The Global Soil Map (Grundy et al. 2104; globalsoilmap.org) will help relieve this constraint as it is completed.

With respect to data, semantics and linked data can be helpful to realize these use cases more easily and to benefit from resources across use cases, but it is never crucial. Standardized data protocols are needed that allow these data to be shared, discovered, combined with other data from different sources and used in multiple applications and analyses in various modeling domains. The use of Linked Open Data protocols will ultimately be a central component of the Next Gen framework. Given their current state of development, however, NextGen will have to evolve toward using these protocols.

In terms of users and usability, in most use cases the users are already quite clearly identified and mainly developments are required to more explicitly define what these users really need through state-of-the-art requirements analysis techniques. Each of the use cases formulates some general ideas and directions, but clearly much more information is required to elaborate the real applications.

Targeted visualization is needed to bring the message across to the users as specific user interfaces and applications are proposed. Only Use Case 2 could probably work with already available visualization techniques. This use case requires highly variable visualizations in tables that generated and analyzed for each realization of the use case. The other use cases require much more standardized visualizations that need to be integrated into specific interfaces with a clear link to underlying data and assumptions. Mostly these visualizations are not available, and interestingly in some cases there are clear benefits seen from apps instead of desktop based solutions.

Finally in terms of IT infrastructure and modeling, the individual models clearly need to move toward availability as stable, robust, granular and well-defined components instead of the current situation of large containers of analytical steps that are not flexibly steered through external programs. Models thus need to become advanced algorithms that can be robustly called in a service-oriented set up. From an infrastructure point, the availability of services for data and models (analytics) is as crucial for realizing the service oriented infrastructures underlying the applications.

Model linking and frameworks play a role in some of the use cases, but not as much as flexible environment in which a user can play around with models (with the exception of Use Case 2). Instead, once a linking has been realized, this should be repetitive and stable in rather specific environment. Arguably, to prepare such 'stable linking' of models a flexible modeling and workflow environment could provide some value. This framework that allows researchers to generate and share these workflows, including connections to multiple data sources, could be a key element of the NextGen modeling infrastructure.

Table 1. *Overall analysis of the five NextGen Use cases as analyzed in this paper for their relevant IT and data aspects as discussed across sections of the paper. Scores are from 'no score'; to + = element, but not crucial; to ++ = important innovation required; to +++ = crucial innovation required. ? = unknown importance. (1) User identification refers to activities in which it is relatively unsure who the user really is, and this needs to be further investigated; (2) complexity is a subjective assessment of the overall complexity of the use case as judged from the number of data sources, ICT innovations and visualization techniques; (3) user requirements refers to the extent to which additional user requirements analysis is needed to progress. Use cases are: Use case definitions: 1 = farm extension in Africa; 2 = developing technologies for sustainable intensification; 3 = investing in projects for sustainable intensification; 4 = management support for precision agriculture; 5 = supplying food products that meet corporate sustainability goals.*

Characteristics			**Use Cases**		
	1	2	3	4	5
Users and usability					
User identification (1)	?	++	+	+	++
Complexity (2)	++	+++	+++	++	++
User requirements (3)	++	++	++	+++	+++
Data and IPR					
Open data	+	+++	+++	+	+++
Private data	+++	+		+++	+++
Data integration	++	+++	+++	++	+++
'Novel' data sources (i.e. social media, remote sensing, crowd sourcing)	+++		+++	+++	
Big data			++	+	
Linked data and semantics		++	++	++	++
Visualization					
Targeted visualization required	+++	+	+++	+++	+++
Visual Analytics required			++	++	
Apps	+++		++	+++	+
Model development					
Model as components	+++	+++	+++	+++	+++
Model linking	+	+	++	++	++
Flexible workflow frameworks	++	+++	++	+	++
Collaborative development	++	++	+++	++	++
IT infrastructure					
Service-oriented architecture	+++	++	+++	+++	+++
Desktop based	partly	yes	no	no	yes
Application (app) based	yes	no	yes	yes	partly

10.2 Recommendations

Experience in other fields of science and industry suggests that the key to success in both areas is to reduce impedance to adoption, so that the discovery, use, and contribution of data and software all become trivial tasks – or at least as close to trivial as may be possible, given the complexities involved in agricultural systems. Experience also suggests that modern Web 2.0 and cloud technologies can play an important role in reducing such friction, by providing intuitive interfaces and eliminating the need for users to install and maintain software. We propose the following principles for the design of next generation agricultural framework modeling:

All elements of the system should be linked via **intuitive, Web 2.0 interfaces**, with associated REST APIs (Representational State Transfer Application Programming Interface) for programmatic access. This will allow for self-documenting applications and interfaces and will facilitate navigation and integration of the various modeling and data components.

Software and data should be **cloud-hosted** to permit access by any authorized user, without the need to install local software. This does not necessarily require that access be free of charge: there may be a need for some payment mechanisms for resource-intensive activities in order to achieve sustainability. Ideally most data and modeling components would be free in both senses of the word, little to no cost and few to no restrictions on use, thus encouraging an active community of application developers to provide value-added products to end users.

Software should be **modular** to facilitate adding new components and combining components to create new applications. Documentation of components will facilitate component consistency and coherence with respect to complexity, data requirements and uncertainty.

The system should be **populated with an initially compelling set of components** and example workflows so that users can, for example, upload data and get immediate value in terms of analyses, comparisons with other similar data, and visualizations. Include a core set of utilities available for commonly-required tasks such as spatial and temporal data aggregation, format conversions, visualization of commonly used data and model products.

Integrated **user access control** must be provided for all contributions, so that users can feel confident uploading private, confidential and restricted-access data and software.

The system should make it **easy for users to upload and publish** new data, software, and workflows. Following the principle of "publish then filter", we want to encourage sharing of data and software, and use feedback mechanisms (such as ratings and post-publication review) to identify what is good – rather than interposing onerous curation processes. Mechanisms for citing specific contributions of data and software (e.g., DOIs) and for tracking accesses to contributions, in order to provide positive and quantitative feedback to contributors should be integrated.

Common vocabularies and ontologies should allow interchange of data among and between disciplines. Development of these vocabularies has the important side effect of requiring that the disciplines work in a coordinated way, thus breaching the disciplinary silos that currently impede progress in integrated modeling. Use of linked data protocols will allow interpretation of data from multiple, distributed sources.

Openness, transparency, and collaboration should be encouraged. As much as possible, the data, models, model components, analytical tools, visualization utilities, and model syntheses should be openly available for scrutiny, improvement, and re-use. Some components of the system will be proprietary, as public-private partnerships are encouraged to flourish.

Communities of users must be allowed to guide the development of the NextGen system. These communities are the most important element of the AgGen2 infrastructure for integrated agricultural modeling. They include not only end users and stakeholders, but data collectors, model developers, model users, and ICT professionals.

We believe that a next generation agricultural modeling community that follows these principles is likely to gain a substantial following and encourage increased collaboration within and between communities and provide significantly enhanced tools to help deliver sustainable food production under changing climate conditions.

References

Altintas, I., C. Berkley, E. Jaeger, M. Jones, B. Ludäscher, and S. Mock. 2004. "Kepler: An Extensible System for Design and Execution of Scientific Workflows." In *Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004)*, pages 423-424. IEEE Computer Society, 2004.

Antle, J. et al. 2014. Companion Paper 2. Antle, J. M., B. O. Basso, R. T. Conant, C. Godfray, J. W. Jones, M. Herrero, R. E. Howitt, B. A. Keating, R. Munoz-Carpena, C. Rosenzweig, P. Tittonell, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Model Design, Improvement and Implementation." AgMIP, < url tbd >.

Antle, J. M., J. W. Jones and C. Rosenzweig. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Introduction." AgMIP, < url tbd >.

Asseng, S., F. Ewert, C. Rosenzweig, J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P.J. Thorburn, R.P. Rotter, D. Cammarano, N. Brisson, B. Basso, P. Martre, P.K. Aggarwal, C. Angulo, P. Bertuzzi, C. Beirnath, A.J. Challinor, J. Doltra, S. Gayler, R. Goldberg, R. Grant, L. Heng, J. Hooker, L.A. Hunt, J. Ingwersen, R.C. Izaurralde, K.C. Kersebaum, C. Müller, S. Naresh Kumar, C. Nendel, G. O'Leary, J.E. Olesen, T.M. Osborne, T. Palosuo, E. Priesack, D. Ripoche, M.A. Semenov, I. Shcherbak, P. Steduto, C. Stöckle, P. Stratonovitch, T. Streck, I. Supit, F. Tao, M. Travasso, K. Waha, D. Wallach, J.W. White, J.R. Williams, and J. Wolf. 2013. "Uncertainty in simulating wheat yields under climate change." *Nature of Climate Change* 3:827-832 doi:10.1038/nclimate1916.

Athanasiadis, I., A.E. Rizzoli, M. Donatelli, and L. Carlini. 2006. "Enriching software model interfaces using ontology-based tools." In: Voinov, A., Jakeman, A., Rizzoli, A. (Eds.), The iEMSs Third Biannual Meeting "Summit on Environmental Modelling and Software Third Biennial Meeting". International Environmental Modelling and Software Society, Burlington, USA.

Athanasiadis, I., and F. Villa. 2013. "A roadmap to domain specific programming languages for environmental modeling: key requirements and concepts." *Proceedings of the 2013 ACM workshop on Domain-specific modeling* 27-32. New York, NY, USA ©2013. doi>10.1145/2541928.2541934.

Antoniou, G. and F. van Harmelen. 2004. "A semantic web primer." *The MIT Press*, Cambridge, Massachusetts; London, England.

Arrouays, D., M.G. Grundy, A.E. Hartemink, J.W. Hempel, G.B.M. Heuvelink, S.Y. Hong, P. Lagacherie, G. Lelyk, A.B. McBratney, N.J. McKenzie, M.D.L. Mendonca-Santos, B. Minasny, L. Montanarella, I.O.A. Odeh, P.A. Sanchex, J.A. Thompson, and G.L. Zhang. "Chapter Three – GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties." *Advances in Agronomy* 125:93-134.

Bassu, S., N. Brisson, J.L. Durand, K.J. Boote, J.I. Lizaso, J.W. Jones, C. Rosenzweig, A.C. Ruane, M. Adam, C. Baron, B. Basso, C. Biernath, H. Boogaard, S. Conijn, M. Corbeels, D. Deryng, G. De Sanctis, S. Gayler, P. Grassini, J. Hatfield, S. Hoek, C. Izaurralde, R. Jongschaap, A.R. Kemanian, K.C. Kersebaum, S.H. Kim, N.S. Kumar, D. Makowski, C. Müller, C. Nendel, E. Priesack, M.V. Pravia, F. Sau, I. Shcherbak, F. Tao, E. Teixeira, D. Timlin, and K. Waha. 2014. "How do various maize crop models vary in their responses to climate change factors?" *Global Change Biology* 20: 2301–2320. doi:10.1111/gcb.12520

Boehm, B. W. 1987. "Improving software productivity." Computer, 20(9):43-57.

CAIDA, Center for Applied Internet Data Analysis. 2010. "Summary of Anonymization Best Practice Techniques." www.caida.org/projects/predict/anonymization. Accessed Sept 5, 2014.

Callahan, S.P., J. Freire, E. Santos, C.E. Scheidegger, C.T. Silva, and H.T. Vo. 2006. VisTrails: Visualization Meets Data Management. Proc. SIGMOD Int'l Conf. Management of Data (SIGMOD 06), ACM. 745-747.

Chapman, R., and T. Slaymaker. 2002. "ICTs and Rural Development: Review of the Literature, Current Interventions and Opportunities for Action." Working Paper 192 Results of ODI research presented in preliminary form for discussion and critical comment. Overseas Development Institute.

Cockburn, A. 2006. Agile Software Development: The Cooperative Game, 2nd edition. Addison-Wesley Professional. ISBN 0-321-48275-1, ISBN 978-0-321-48275-4.

Danes, M., A. Jellema, and S. Janssen. 2014. "Mobiles for agricultural development : exploring trends, challenges and policy options for the Dutch government." *Alterra report 2501*, Alterra, Wageningen-UR, Wageningen. P 25: http://edepot.wur.nl/297683

David, O., J.C. Ascough II, W. Lloyd, T.R. Green, K.W. Rojas, G.H. Leavesley, and L.R. Ahuja. 2013. "A software engineering perspective on environmental modeling framework design: The Object Modeling System." *Environmental Modelling & Software* 39:201-213.

Dixon, J., K. Taniguchi, H. Wattenback, and A. Tanyeri-Arbur. 2004. Smallholders, globalization and policy analysis. Agricultural Management, Marketing and Finance Service (AGSF), Agricultural Support Systems Division, Rome, FAO.

Donatelli M., and A. Rizzoli. 2008. "A design for framework-independent model components of biophysical systems." In: 'International Congress on Environmental Modelling and Software iEMSs 2008'. *Proceedings of the iEMSs Fourth Biennial Meeting*, Barcelona, Catalonia 7-10 July 2008 pp. 727-734.

Donatelli, M., S. Bregaglio, R. Confalonieri, R. De Mascellis, and M. Acutis. 2014. "A generic framework for evaluating hybrid models by reuse and composition – A case study on soil temperature simulation." *Environmental Modelling & Software* 62:478-486.

Drosatos, G., P.S. Efraimidis, I.N. Athanasiadis, M. Stevens and Ellie D'Hondt. "Privacy-Preserving Computation of Participatory Noise Maps in the Cloud." *Journal of Systems and Software*, Elsevier, 2014.

Duan, Q., S. Sorooshian, and V.K. Gupta. 1992. "Effective and efficient global optimization for conceptual rainfall– runoff models." *Water Resources Research* 28: 1015– 1031.

Eckhardt, K., N. Fohrer, and H.G. Frede. 2005. "Automatic model calibration." *Hydrological Processes* 19: 651–658. doi: 10.1002/hyp.5613.

Elliott, J., D. Deryng, C. Müller, K. Frieler, M. Konzmann, D. Gerten, M. Glotter, M. Florke, Y. Wada, N. Best, S. Eisner, B.M. Fekete, C. Folberth, I. Foster, S.N. Gosling, I. Haddeland, N. Khabarov, F. Ludwig, Y. Masaki, S. Olin, C. Rosenzweig, A.C. Ruane, Y. Satoh, E. Schmid, T. Stacke, Q. Tang, and D. Wisser. 2014. "Constraints and potentials of future irrigation water availability on agricultural production under climate change." *Proceedings of the National Academy of Sciences 111*(9):3239-3244.

Elliott, J., D. Kelly, J. Chryssanthacopoulos, M. Glotter, K. Jhunjhnuwala, N. Best, ... and I. Foster. 2014. "The parallel system for integrating impact models and sectors (pSIMS)." *Environmental Modelling & Software*.

Fekete, J.D. 2013. "Visual Analytics Infrastructures: From Data Management to Exploration." *Computer*, Institute of Electrical and Electronics Engineers (IEEE), Visual Analytics: Seeking the Unknown, 46(7):22-29.

Fekete, J.D. and C. Silva. 2012. "Managing Data for Visual Analytics: Opportunities and Challenges." IEEE Data Engineering Bulletin 35(3):27-36.

Fielding, R. 2000. Architectural Styles and the Design of Network-based Software Architectures, PhD thesis, Information and Computer Science, University of California Irvine.

Goecks, J., A. Nekrutenko, and J. Taylor. 2010. "Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences." *Genome Biology 11*(8): R86.

Gregersen, J.B., P.J.A. Gijsbers, and S.J.P. Westen. 2007. "OpenMI: open modelling interface." *Journal of Hydroinformatics* 9(3):175–191.

Hammer, G., M. Cooper, F.Tardieu, S. Welch, B. Walsh, F. van Eeuwijk, S. Chapman, and D. Podlich. 2006. "Models for navigating biological complexity in breeding improved crop plants." *TRENDS in Plant Science* 11(12):587-593.

Hillyer, C., J. Bolte, F. van Evert, A. Lamaker. 2003. "The ModCom modular simulation system." *European Journal of Agronomy* 18:333-343.

Holzworth, D.P., N.I. Huth, and P.G. deVoil. 2010. "Simplifying environmental model reuse." *Environmental Modleing and Software* 25(2):269-275. DOI: 10.1016/j.envsoft.2008.10.018.

Holzworth, D.P., N.I. Huth, P.G. deVoil, E.J. Zurcher, N.I. Herrmann, G. McLean, K. Chenu, E. J. van Oosterom, V. Snow, C. Murphy, A.D. Moore, H. Brown, J.P.M. Whish, S. Verrall, J. Fainges, L.W. Bell, A.S. Peake, P.L. Poulton, Z. Hochman, P.J. Thorburn, D.S. Gaydon, N.P. Dallgliesh, D. Rodrigues, H. Cox, S. Chapman, A. Doherty, E. Teixeira, J. Sharp, R. Cichota, I. Vogeler, F.Y. Li, E. Wang, G.L. Hammer, M.J. Robertson, J.P. Dimes, A.M. Whitbread, J. Hunt, H. van Rees, T. McClelland, P.S. Carberry, J.N.G. Hargreaves, N. MacLeod, C. McDonald, J. Harsdorf, S. Wedgwood, and B.A. Keating. 2014. "APSIM: Evolution towards a new generation of agricultural systems simulation." *Environmental Modeling & Software* 62:327-360 doi:10.1016/j.envsoft.2014.07.009.

Hoogenboom, G., and J.W. White. 2003. "Improving Physiological Assumptions Of Simulation Models By Using Gene-Based Approaches." *Agronomy Journal* 95(1):82-89.

Innis, G.S. 1978. "Objectives and structure for a grassland simulation model." In: Innis, G.S. (Ed.), Grassland simulation model. Springer-Verlag, New York, pp. 1-21.

Janssen, S., I.A. Athanasiadis, I. Bezlepkina, R. Li HT Knapen, I.P. Domínguez, A.E. Rizzoli, and M.K. van Ittersum. 2011. "Linking models for assessing agricultural land use change." *Computers and Electronics in Agriculture* 76:148-160.

Janssen, S., E. Andersen, I.N. Athanasiadis, and M.K. Van Ittersum. 2009. "A database for integrated assessment of European agricultural systems." *Environmental Science & Policy* 12:573-587.

Janssen, S., C. H. Porter, A. D. Moore, I. N. Athanasiadis, I. Foster, J. W. Jones, and J. M. Antle. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support." AgMIP, < url tbd >.

Jeffrey, S.J., J.O. Carter, K.B. Moodie, and A.R. Beswick. 2001. "Using spatial interpolation to construct a comprehensive archive of Australian climate data." *Environmental Modelling & Software* 16:309-330.

Johnston, J.M., D.J. McGarvey, M.C. Barber, G. Laniak, J.E. Babendreier, R. Parmar, K. Wolfe, S.R. Kraemer, M. Cyterski, C. Knightes, B. Rashleigh, L. Suarez, and R. Ambrose. 2011. "An integrated modeling framework for performing environmental assessments: application to ecosystem services in the Albemarlee Pamlico basins (NC and VA, USA)." *Ecological Modelling* 222(14):2471-2484.

Jones, J.W. et al. 2014. Companion Paper 1. Jones, J. W., J. M. Antle, B. O. Basso, K. J. Boote, R. T. Conant, I. Foster, H. C. J. Godfray, M. Herrero, R. E. Howitt, S. Janssen, B. A. Keating, R. Munoz-Carpena, C. H. Porter, C. Rosenzweig, and T. R. Wheeler. 2015. "Towards a New Generation of Agricultural System Models, Data, and Knowledge Products: State of Agricultural Systems Science." AgMIP, < url tbd >.

Jones, J. W., G. Hoogenboom, C. H. Porter, K. J. Boote, W. D. Batchelor, L. A. Hunt, P. W. Wilkens, U. Singh, A. J. Gijsman, and J. T. Ritchie. 2003. "The DSSAT cropping system model." *European Journal of Agronomy* 18(3-4):235-265.

Keating, B.A., P.S. Carberry, G.L. Hammer, M.E. Probert, M.J. Robertson, D. Holzworth, N.I. Huth, J.N.G. Hargreaves, H. Meinke, Z. Hochman, G. McLean, K. Verburg, V. Snow, J.P. Dimes, M. Silburne, E. Wang, S. Brown, K.L. Bristow, S. Asseng, S. Chapman, R.L. McCown, D.M. Freebairne, and C.J. Smith. 2003. "An overview of APSIM, a model designed for farming systems simulation*." European Journal of Agronomy* 18:267-288.

Laniak, G.F., G. Olchin, J. Goodall, A. Voinov, M. Hill, P. Glynn, G. Whelan, G. Geller, N. Quinn, M. Blind, S. Peckham, S. Reaney, N. Gaber, R. Kennedy, and A. Hughes. 2013. "Integrated environmental modeling: A vision and roadmap for the future." *Environmental Modelling and Software* 39:3-23.

Leavesley, G.H., S.L. Markstrom, P.J. Restrepo, and R.J. Viger. 2002. "A modular approach to addressing model design, scale, and parameter estimation issues in distributed hydrological modelling." *Hydrological Processes* 16:173-187.

Lloyd, W., O. David, J.C. Ascough II, K.W. Rojas, J.R. Carlson, G.H. Leavesley, P. Krause, T.R. Green, and L.R. Ahuja. 2011. "Environmental modeling framework invasiveness: Analysis and implications." *Environmental Modelling & Software* 25(10):1240-1250. , http://dx.doi.org/10.1016/j.envsoft.2011.03.011

Lopez-Gonzalez, G., S.L. Lewis, M. Burkitt, M., and O.L. Phillips. 2011. "ForestPlots.net: a web application and research tool to manage and analyse tropical forest plot data." *Journal of Vegetation Science* 22:610-613.

Lokers, R., and S. Janssen. 2014. Presentation "Putting Food Security Research in Practice at Wageningen UR: Exploiting global knowledge for food-security through cooperation, co-creation and sharing of information" at ICRI 2014 (www.icri.eu), 3 April 2014.

Ludäscher, B., K. Lin, B. Brodaric, and C. Baru. 2003. "GEON: Toward a Cyberinfrastructure for the Geosciences: A Prototype for Geologic Map Integration via Domain Ontologies." U.S. Geological Survey Open-File Report 03-071. pubs.usgs.gov/of/2003/of03-471/ludascher

Maidment, D.R., R.P. Hooper, D.G. Tarboton, and I. Zaslaksky. 2009. "Accessing and sharing data using CUAHSI water data services." In: *Hydroinformatics in Hydrology, Hydrogeology and Water Resources* (Proc. of Symposium JS.4 at the Joint IAHS & IAH Convention, Hyderabad, India, September 2009). IAHS Publ. 331, pp. 213e223.

Manolescu, I., W. Khemiri, V. Benzaken, and J.D. Fekete. 2009. "Reactive workflows for visual analytics." In Data Management & Visual Analytics workshop, Berlin.

Messina, C. D., J. W. Jones, K. J. Boote, and C. E. Vallejos. 2006. "A gene-based model to simulate soybean development and yield responses to environment." *Crop Science* 46:456-466.

Mitchell, E.E.L., and J.S. Gauthier. 1976. "Advanced Continuous Simulation Language (ACSL)." Simulation 1976 26(3):72-78. DOI: 10.1177/003754977602600302.

Mittal, S., S. Gandhi, and G. Tripathi. 2010. "Socio-economic impact of mobile phones on Indian agriculture." Indian Council for Research on International Economic Relations. New Delhi.

Moore, A.D., D.P. Holzworth, N.I. Herrmann, N.I. Huth, and M.J. Robertson. 2007. "The Common Modelling Protocol: A hierarchical framework for simulation of agricultural and environmental systems." *Environmental Modelling & Software* 95: 37-48.

Moore, A.D., D.P. Holzworth, N.I. Herrmann, H.E. Brown, P.G. de Voil, V.O. Snow, E.J. Zurcher, and N.I. Huth. In press. "Modelling the manager: representing rule-based management in farming systems simulation models." *Environmental Modelling & Software* DOI:10.1016/j.envsoft.2014.09.001

Muetzelfeldt, R., and J. Massheder. 2003. "The Simile visual modelling environment." *European Journal of Agronomy* 18:345-358.

NIST, National Institute of Standards and Technology. 2013. "NIST Cloud Computing Standards Roadmap." US Department of Commerce. Special Publication 500-291, Version 2. www.nist.gov/itl/cloud/upload/NIST_SP-500-291_Version-2_2013_June18_FINAL.pdf

Noble, I.R. 1975. Computer simulations of sheep grazing in the arid zone. Ph.D. thesis, University of Adelaide. 209 pp.

Porter, C.H., C. Villalobos, D. Holzworth, R. Nelson, J.W. White, I.N. Athanasiadis, S. Janssen, D. Ripoche, J. Cufi, D. Raes, M. Zhang, R. Knapen, R. Sahajpal, K.J. Boote, and J.W. Jones. 2014. "Harmonization and translation of crop modeling data to ensure interoperability." *Environmental Modeling and Software*. In press. DOI: 10.1016/j.envsoft.2014.09.004.

Powell, M., T. Davies, and K.C. Taylor. 2012. "ICT for or against development? An introduction to the ongoing case of Web 3.0." IKM Working Paper No. 16, March 2012. http://wiki.ikmemergent.net/index.php/Documents#IKM.C2.A0Working_Papers_with_Summaries. Accessed 24 July, 2014.

Ramankutty, N., A.T. Evan, C. Monfreda, and J.A. Foley. 2008. "Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000." *Global Biogeochemal Cycles* 22:GB1003

Raymond, E. S. The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media, October 1999, p.279.

Reynolds, J.F., and B. Acock. 1997. "Modularity and genericness in plant and ecosystem models." *Ecological Modelling* 94:7-16.

Ritchie, J.R., U. Singh, D. Godwin, and L.A. Hunt. 1991. "A user's guide to CERES Maize - V2.10." International Fertilizer Development Center, Muscle Shoals, Alabama.

Rizzoli, A.E., M. Donatelli, I.N. Athanasiadis, F. Villa, and D. Huber. 2008. "Semantic links in integrated modelling frameworks." *Mathematics and Computers in Simulation* 78(2):412-423.

Rojas-Ruiz, J. and A. Diofasi. 2014. "Upwardly mobile: how cell phones are improving food and nutrition security." *Hunger and Undernutrition Blog*. http://www.hunger-undernutrition.org/blog/2014/04/upwardly-mobile-how-cell-phones-are-improving-food-and-nutrition-security.html

Rosenzweig, C., J.W. Jones, J.L. Hatfield, A.C. Ruane, K.J. Boote, P.J. Thorburn, J.M. Antle, G.C. Nelson, C.H. Porter, S. Janssen, S. Asseng, B. Basso, F. Ewert, D. Wallach, G.A. Baigorria, and J.M. Winter. 2013. "The Agricultural Model Intercomparison and Improvement Project (AgMIP): Protocols and pilot studies." *Agricultural and Forest Meteorology* 170:166-182.

Running, S.W., R.R. Nemani, D.L. Peterson, L.E. Band, D.F. Potts, L.L. Pierce, and M.A. Spanner. 1989. "Mapping regional forest evapotranspiration and photosynthesis by coupling satellite data with ecosystem simulation." *Ecology* 70:1090-1101.

Sanchez, P. A., S. Ahamed, F. Carre, A. E. Hartemink, J. Hempel, J. Huising, P. Lagacherie, A. B. McBratney, N. J. McKenzie, M L. de Mendonca-Santos, B. Minasny, L. Montanarella, P. Okoth, C.A. Palm, J.D. Sachs, K.D. Shepherd, T. Vagen, B. Vanlauwe, M.G. Walsh, L.A. Winowiecki, and G. Zhang. 2009. "Digital soil map of the world." *Science* 325:680-681.

Shirky, C. 2009. Here Comes Everybody: The Power of Organizing Without Organizations. Penguin Books. 352 pp.

Steinfield, C. and S. Wyche. 2013. "Assessing the Role of Information and Communication Technologies to Enhance Food Systems in Developing Countries." Global Center for Food Systems Innovation, White Paper Series, Michigan State University, East Lansing, Michigan, USA 39 pp.

Thomas, J. and K. Cook (Eds.) 2005. "Illuminating the Path: Research and Development Agenda for Visual Analytics." National Visualization and Analytics Center, IEEE. http://vis.pnnl.gov/pdf/RD_Agenda_VisualAnalytics.pdf

van Ittersum, M.K., P.A. Leffelaar, H. van Keulen, M.J. Kropff, L. Bastiaans, and J. Goudriaan. 2003. "On approaches and applications of the Wageningen crop models." *European Journal of Agronomy* 18:201-234.

Villa, F., M. Ceroni, K. Bagstad, G. Johnson, and S. Krivov. 2009. "ARIES (Artificial Intelligence for Ecosystem Services): a new tool for ecosystem services assessment, planning and valuation." *BioEconomics* Ecoinformatics Collaboratory, Gund Institute for Ecological Economics, University of Vermont.

Wallach, D., F. Brun, N. Keussayan, B. Lacroix, and J.E. Bergez. 2012. "Assessing the uncertainty when using a model to compare irrigation strategies." *Agronomy Journal* 104:1274-1283.

Weinberger, D. Small pieces loosely joined: A unified theory of the Web. Basic Books, Perseus Book Group, New York, 2008.

White, J.W., and G. Hoogenboom. 1996. "Simulating Effects of Genes for Physiological Traits in a Process-Oriented Crop Model." *Agronomy Journal* 88:416-422.

Wight, J. Ross & J.W. Skiles, eds. 1987. "SPUR: Simulation of Production and Utilization of Rangelands. Documentation and User Guide." U.S. Department of Agriculture, Agricultural Research Service, ARS 63, 372 pp

Wilde, M., M. Hategan, J.M. Wozniak, B. Clifford, D.S. Katz, and I.S. Foster. 2011. "A language for distributed parallel scripting." *Parallel Computing* 37(9):633-652.

Wolstencroft, K., R. Haines, D. Fellows, A. Williams, D. Withers, S. Owen, S. Soiland-Reyes, I. Dunlop, A. Nenadic, P. Fisher, J. Bhagat, K. Belhajjame, F. Bacall, A. Hardisty, A. Nieva de la Hidalga, M.P. Balcazar Vargas, S. Sufi, and C. Goble. 2013. "The Taverna workflow suite: designing and executing workflows of Web Services on the desktop, web or in the cloud." *Nucleic Acids Research* 41(W1): W557-W561.doi:10.1093/nar/gkt328.

Zeegers, E. 2012. "Thousands of Dutch citizens take part in the first national iSPEX-measure-day." http://ispex.nl/en/duizenden-nederlanders-doen-mee-aan-eerste-nationale-ispex-meetdag/. Accessed 9/4/2014.

Next Generation Agricultural Systems Models

Report on the Convening Held at the Bill & Melinda Gates Foundation August 12-14, 2014

Prepared by John Antle and Stephanie Price

September 15, 2014

Introduction and Summary

A scoping project on *Next Generation Agricultural Systems Models* was commissioned by the Foundation with the goal of designing a roadmap for investments that would improve decision support tools for evaluating farm-level sustainable intensification options for smallholder farmers in the developing world. This scoping project was coordinated by the Agricultural Model Inter-comparison and Improvement Project (AgMIP), and led by John Antle (Oregon State University).

The scoping project began with the drafting of three background papers by an expert author team (provided to all participants of the workshop). A stakeholder workshop (this convening) was planned to obtain broader input from the various communities of science and practice that are now or potentially could be engaged in the pursuit of global food security and poverty reduction for the world's smallholder farmers.

This report summarizes the proceedings of the convening, including panel presentations and discussions. Background papers and PowerPoint presentations are available in the Convening Dropbox (access available on request). The Appendix to this report contains the convening agenda, a list of participants, and summaries of the Use Cases developed by the expert author team and by the participants in the convening.

Based on the discussions at the convening, and subsequent discussions among the scoping study leaders, participants at the convening and Foundation staff, the following next steps were identified:

- There was general agreement to build on the momentum achieved by the scoping study and the convening, through follow-up meetings (perhaps to be held on-line) before the end of 2014. These meetings would plan preliminary NextGen activities that utilize selected Use Cases to further engage the relevant communities of science and practice, based on suggestions made during the convening. It was agreed that the Foundation staff would continue to play a coordination role for these meetings.
- The background papers would be re-organized with an overall introduction to the scoping study and the Use Cases, followed by the three papers. In addition to the background papers and roadmap document being prepared for the Foundation, the goal of the author team is to publish the papers in a peer-reviewed journal, and also to pursue publication of a high-impact "policy" paper summarizing the scoping study approach and findings.

The leaders of AgMIP have also decided to provide a forum for further discussion and planning of activities in line with the NextGen vision at their annual meeting to be held February 28-30, 2015, at University of Florida. This meeting is open to all interested in attending.**Convening Organization and Development**

The convening agenda was developed by Stan Wood and Kate Schneider (Gates Foundation) in collaboration with the John Antle and other AgMIP leaders. Peter Craufurd (International Maize and Wheat Improvement Center) served as facilitator. Participants were invited to represent private and public sector interests, including agribusiness and the information and computer technology sector, and the broader science community. A cohort of early career scientists was also invited.

Objectives, Agenda and Outcomes

The background papers (see below) developed five Use Cases to motivate the review of the relevant science, and to discuss a vision for the Next Generation of models (referred to here as *NextGen*). The goal of the convening was to have a broader conversation around use cases, technical challenges, data, and how practitioner and user communities could be strengthened.

The convening took place over two and a half days (see the attached agenda, some modifications made during the workshop are noted below). The first day was focused around developing and discussing Use Cases that can respond to the information demands set forth by the global community addressing agriculture and food security. The second day addressed technical opportunities and challenges, identified potential partnerships, and considered how a community could be fostered to advance this work. The third (half) day reviewed the draft papers and gathered the community's input and feedback to guide their completion and to formulate a roadmap for advancing the development of NextGen models, data and information technology tools. After authors presented and collected more feedback, attendees engaged in a creative networking session to brainstorm possible NextGen modelling applications.

The expected outcomes of this convening were to:

- Generate realistic Use Cases that meet actual information needs of stakeholders, a roadmap of possible projects or approaches, and partnerships that could implement them and address related technical challenges.
- Begin to develop a multi-sectoral and inter-disciplinary community dedicated to applying their diverse skills to delivering data, analysis, and information that serves global agriculture, food security, and poverty reduction decision-making.
- Solicit comments on the background papers, and suggestions for the roadmap towards NextGen.

Background Papers

The three background papers prepared before the convening (available in the convening dropbox):

Paper 1. *Next Generation Agricultural Systems Models, Data and Knowledge Products: State of Agricultural Systems Science*

Authors: J. W. Jones, John Antle, Bruno Basso, Ken Boote, Richard Conant, Ian Foster, Charles Godfray, Mario Herrero, Richard Howitt, Sander Jansen, Brian Keating, Rafa Munoz-Carpena, Cheryl Porter, Cynthia Rosenzweig, Tim Wheeler

Paper 2. *Next Generation Agricultural Systems Models, Data and Knowledge Products: New Approaches to Model Development, Improvement and Use*

Authors: J. Antle, B. Basso, R. Connant, C. Godfray, J. Jones, M. Herrero, R. Howitt, B. Keating, Rafael Munoz-Carpena, C. Rosenzweig, P. Tittonell, T. Wheeler

Paper 3. *Next Generation Agricultural Systems Models, Data and Knowledge Products: Building an Open Web-Based Approach to Agricultural Data, System Modeling and Decision Support*

Authors: S. Janssen, C.H. Porter, A. Moore, I.N. Athanasiadis, I. Foster, J.W. Jones, J. Antle

Day 1

Opening Comments

The convening opened with Peter Craufurd welcoming the attendees and motivating the group to do some out-of-the-box thinking. Discussants were asked to participate genuinely as individuals and to remain in the "pre-competitive" sphere. Participants were asked to keep focus on the main meeting objectives.

Stan Wood introduced the Agricultural Development program at the Foundation. He explained the movement towards evidence-based decision making, with a focus on data for analytics. Stan outlined the main goals of meeting as trying to answer the following questions:

- What are the questions that need to be addressed to improve livelihoods of smallholders?
- How can we use our experience and insights, without the baggage of institutional and other vested interests, to think of analytical inputs that address challenges?

John Antle further elaborated on the questions to be considered during the convening:

- How to deal (sustainably) with food security?
- How to accelerate productivity growth to meet food security needs this century?
- How to move NextGen using the Use Cases from concept to implementation?
- How to incorporate the science and user communities into this conversation?

Expert Panel: "What do Next Generation Models need to deliver?"

The opening panel spoke on what the "Next Generation" models need to deliver. Panelists included Marc Sadler (World Bank), Hilary Parsons (Nestle), Rob Bertram (USAID), Lystra Antoine (DuPont/Pioneer), Wendy-Lin Bartels (University of Florida).

The panelists focused on themes of flexibility, simplicity, transparency, integration of data and research disciplines, and engagement with the ultimate problems and beneficiaries addressed by NextGen models. Marc opened the panel by stressing that policy makers make hard decisions based on these models, but the models and assumptions that inform these individuals are often outdated or inappropriate for the question. Furthermore, there is a proliferation of models, interests and techniques that may overwhelm decision makers. Thus, in order to progress to the NextGen, we will need flexible and simple tools that communicate results that people understand. Hilary echoed Marc's observations of model proliferation (through isolation), and explained Nestle's approach to integrated modelling of driving concepts to achieve results. Hilary mentioned the example of the "Livewell" project in France that links food, nutrition and GHG emissions. Rob Bertram spoke on the gap between modelers and empiricists and the general lack of trust in models by decision makers. Lystra posed a number of challenges, notably how to measure success and link data providers with users. Wendy-Lin, the sole anthropologist in attendance, echoed Lystra's point on identifying the actual beneficiaries of NextGen models.

Questions and comments focused on how to engage users and how to feasibly and sustainably use NextGen models. A recurring theme of "credibility" in models and applications was raised by Stan Wood, which led to some discussion on user engagement. Brian Keating, the moderator, concluded by posing the question: how do we develop useful tools instead of proliferating data? The panelists closed with their perspectives, which included making sure the right questions are being asked, and that channels of communication and local involvement need to be seriously planned.

Developing NextGen Use Cases

The NextGen Use Case development session began with an introduction to Use Case concepts and purposes by Sander Janssen. Sander's presentation was followed by groups of attendees discussing and developing new Use Cases. The supplemental materials provided with this report summarize the Use Cases developed in this breakout section, along with the five Use Cases developed by the author team for the background papers.

Following the breakout session, participants commented and voted on their two favorite Use Cases. Rather than selecting the "winning" use cases as described in the agenda, the Use Cases were grouped according to common themes (see below, breakout reports) and selected to be discussed in the subsequent breakouts. These breakouts were asked to address the following questions:

- Who are the users? What are their decisions?
- Who will benefit, and how?
- What are the technical challenges involved?
- What input data are needed, and at what temporal and spatial resolution?
- How does the information need to be delivered?
- What are the outstanding or new challenges that need to be addressed to implement this case (technical, partnerships, etc.)?

After the breakouts, Rapporteurs presented the results of the discussions. PowerPoint presentations are available in a dropbox, a link to which is available by request. Each presented Use Case is summarized below:

Farm Extension Africa: Sander presented a flow chart of smallholder farmers and extension agents. The flow chart explored the different relationships between stakeholders, tools, and data. This Use Case considers the temporal scale of farming decisions: some support is needed pre-season, in season, and post-season. Examples of this include a farmer deciding what crop or variety to grow or when to apply a pesticide.

Early Warning Systems: This Use Case considers the perspective of an extension agent in a remote district when dealing with a threat, such as pests or a drought. The system advises on things farmers can do to improve their odds given an adverse event will take place. Tools to advise include maps that display visually the spatial impact of a threat, community communications, and mobile technologies. A system such as this is reliant on quality monitoring data to forecast impacts and integrate feedback. A major challenge for this type of NextGen model is finding skilled individuals to make decisions based on the forecasts given and provide constructive feedback.

Precision Agriculture: The precision agriculture case sets out a vision of increased profitability from better data inputs and integrated models. Implementing within season changes to adapt to environmental conditions (such as response to climate variability) can be facilitated by precision agriculture.

Government: The government Use Case works through investment and policy interventions, tradeoffs, intensification and sustainability in mixed smallholder and commercial farming systems. The case considers investment policy decisions based on objectives such as production, profit, and health. In the NextGen, this Use Case envisions decision support tools for regional governments modeling farm system up to watershed, regional, national scales. The group identified challenges in this Use Case as cloud infrastructure, feedback systems, data collection, transparency, and tradeoffs.

Nutrition: This Use Case traces national governments and international agencies paths to achieve objectives of nutrition security and sustainability at global and national scales. The case considers factors such as climate change, food and agriculture, natural resource conservation, trade, and international agreements. Challenges identified in this Use Case include successful cross-disciplinary work and formation of public-private partnerships.

Agribusiness, service providers, and financial services: This Use Case examines NextGen models' use for financial service providers and agribusinesses serving farmers. The case considers how to target inputs (fertilizer, seeds, pesticides, equipment, labor, financial products, insurance, crop protection) and investments (market prices or selling advice, whether to sell forward).

Day 2

Expert Panel: "What can the tech industry deliver to ag and food systems modeling today? What needs to be developed for tomorrow? How do we better deliver information to users?"

The panel spanned representatives from the tech industry, with each speaking on what the industry can provide to agriculture and food systems modeling today. Each panelist presented an issue they perceived existed in agriculture and food modeling today, a tool, and an example of how these tools have been applied to similar problems. Panelists included Ashish Kapoor (Microsoft Research), Jason Cawley (Wolfram), Nancy Harvey (University of Chicago), and Dan Halprin (University of Washington).

Panelists presented examples that included machine learning concepts, database solutions, open and sharable development environments, and the financial gains from using the tools at hand. The panelists also recognized the relationship between human and user intelligence and technological resources. Ashish opened the panel describing how machine learning and indirect or auxiliary data collection can address sparse data issues. He also mentioned additional concepts around value of information analysis, transfer and meta learning. Jason spoke on addressing issues with collaboration, process design and large sunk start-up costs. He outlined how Wolfram Alpha™ addresses user queries as well as employs flexible user interfaces with pre-loaded data. Nancy shared her experiences helping academics and providers of IP monetize and deploy intellectual property. She stated that, as researchers, we should try to not let "the perfect be the enemy of the good." She stressed that the modeling community should maximize the impact of what is currently available and take this technology to the last mile. Dan, of the eScience Institute at UW, described his software development process and shared how he uses the tools at his disposal efficiently. Particularly, Dan described how to build scalable tools and leverage human intelligence and computational knowledge to create self-sufficiency.

Questions to the panel again focused on engaging users, but also included comments on seeing users and developers as an "eco-system" with multiple feedback channels. Modelling should account for these channels in development. Some questions alluded to the relationship between problem identification and tools. Particularly, technology has changed how science is done, so what are the questions we want really to ask? Are the tools we want to use really appropriate? Are we designing new or better tools to answer the *same* questions? In closing, the panel recommended that the community needed to really think about what the problem we are trying to solve is and do it.

Breakouts on Technical and Institutional Opportunities and Challenges

Based on the previous day's discussions, breakouts were organized around opportunities and challenges identified by the participants. Themes were:

- Big Data Collection & Analytics
- Community Building
- Data Access & Interoperability
- Model Assessment & Credibility
- Model Components & Connectivity
- Model Improvement & Tool Development
- Private Public Partnerships
- User Engagement & Experience

Breakout PowerPoint reports from the session Rapporteurs are available. Major points were:

Big Data Collection & Analytics: Using information from tools and models supported by big data will enhance resource efficiency (natural, financial, labor) in real time. We need to get incentives right for private and public sectors, while utilizing mobile technologies and sharing data openly.

Community Building: The NextGen should take the Agmip structure ("lean centre and coalition of the willing") and broaden engagement. There are several challenges associated with this, including lack of incentives, transaction costs, inconsistent reward metrics, and lack of transparency in model development (ex. standardized documentation).

Data Access & Interoperability: The NextGen of models needs to address the entire data life cycle: data collection, manipulation, analysis and visualization, storage, archiving and sharing. The lifecycle needs to also incorporate government and donor initiatives and interventions as well as utilizing common standards. Appropriate incentives should be in place for users and providers of data.

Model Assessment & Credibility: Model improvement should be done through focused research on identified limitations. A major challenge will be communicating results and sensitivities in a credible manner: this will require better visualization of results and increasing skills and understanding of users.

Model Components & Connectivity: We should take notes from the software development industry's practices to employ modularly structured systems: examples of tools and techniques include wrappers and data source compatibility. A major challenge will be whether the modeling community can agree on a common wrapping protocol that is able to perform the necessary functions.

Model Improvement & Tool Development: The NextGen of models should be more generic and easily adaptable. They should also be expandable to include important soil, nutrition, water, and disease impacts.

Private Public Partnerships: NextGen models should be based on a parsimonious approach that may or may not come from the private sector. These models should use real economic costs to produce rate of returns and other outputs, as well as serve as a precursor to business models. In five-years, the ideal is an operational set of models that are driven by actionable questions and benefit multiple stakeholders.

User Engagement & Experience: We need to present the same data in many ways, depending on the intended users and Use Case. To achieve this goal, we should use lessons from the technology industry, see Pasteur's Quadrant.

Expert Panel: "Opportunities for Innovation in Open Environments"

The final panel on opportunities for innovation in open environments provided a diverse array of perspectives. Panelists included Jamie Kinney (Amazon), Liz Carolan (Open Data Institute), Dave Gustafson (ILSI Research Foundation), Caroline Figueres (IICD), and Scott Malcolm (USDA).

Perspectives spanned cutting edge tech development to other more traditional development practitioners and modelers. The panel focused on the themes of partnerships, understanding and enabling data use ecosystems, and bringing accountability and exploration to the fore. Dave opened the panel speaking on sustainable food and nutrition security and stressed the importance of partnerships with scientists from private, academic, and government sectors. Liz shared her unique experiences working for the Open Data Institute. The ODI focuses on incubating startups using open data as well as supporting governmental open data initiatives in developing countries. Liz explained how supporting open data is about enabling an ecosystem of users to change their culture, management systems, and institutions' way of doing business. Scott's position at the ERS in the USDA enabled him to share some thoughts on adoption of technologies (and timing) and thinking critically about data collection. Jamie spoke on his experiences at Amazon collaborating with the public sector to develop resources that enable users to focus on science instead of infrastructure. Jamie also spoke on the benefits of open innovation for accountability and exploration. Caroline shared her wide array of development experiences and the importance of quality data, as well as the importance of considering different perspectives and motivations of beneficiaries.

Questions and comments from the audience focused on how to deploy transparency while preserving the quality and security of data and methods. One question posed wondered how we can we get to the point where users can replicate results from models developed by experts. This brought up the need for proper documentation, but also understanding that data with privacy protections (that is, not completely open) will need to be handled before dissemination.

Day 3

Background Paper Summaries and Reflections of Lead Authors

The first session of Day 3 had brief summaries of the background papers and reflections by the lead authors based on the discussions from the previous two days. These presentation PowerPoints are available in the Convening Dropbox.

Key reflections from the presentations:

Paper 1:

- Lack of trust, credibility of models
- Cloud, mobile phone, apps – opportunities!
- Need to take global science and make it relevant locally (major variability across the landscape)
- Tradeoffs – need to be able to address in terms familiar to stakeholders
- Communication is critical, particularly visualization
- Need to identify and use best practices
- Community is important

- Public-private partnerships
- Don't let the "perfect" be the enemy of the "good"
- Open source is powerful and should be embraced

Paper 2:

- Validated Use Case approach and vision
  - NextGen model design: insights from tech and user community: we have much to learn from them: need collaboration!
- Potential Advances in Model Components: our job
- Evaluating Model Performance for Validation and Improvement
  - Usability and credibility both important
  - What it means for scientists and end users to build and maintain credibility
- Towards Implementation
  - Identify critical problems with well-defined needs: proof of concept for the potential to accelerate innovation
  - Build on and improve existing models, taking advantage of ITC collaborations, linkages to users and "competitive space"

Paper 3:

- Multiple open data standards are emerging (OADA, AgGateway, cropontology.org, AgMIP, ICASA, GODAN, LOD)
- Mobile technologies needed for both gathering and delivering data
- Data quality and validation (Ground-truthing, hind casting)
- Modularity in model components, data components
- Standard workflows for farming system typologies
- Metrics of credibility: End user feedback, reference data sets, reproducibility of results - provenance
- Collaboration with tech companies: tool development, modern software techniques, and self-learning algorithms (self-calibrating models).
- Community needed for transition to NextGen
  - Standards and protocols
  - Open source software
  - Open data
  - Incentives for compliance
- Costs of transition and development of commonalities (not always project-related)
- Need to design a system which will maximize openness and accessibility
  - Minimize elite capture of data and models
  - Preferential options (nudges) for accessibility
  - Create incentives for open source and open data
    - Publication and citation
    - Government and corporate policy
    - Community expectation
    - Donor requirements

The following discussion covered a wide variety of topics, but the most pertinent questions and comments are as follows:

- How do you integrate economic and biophysical models, but continue to add people and landscape structure? Where do forest, bio-diversity, and wildlife characteristics fit in?
- How can we actually get good models in the wild: the private sector is looking for models and research to commercialize and deploy.
- Models will need to adapt to specific needs. Thus we will need to integrate models, users, and tools. So, how does the NextGen support collaboration between users and model developers?
- It is important to measure accuracy in NextGen models, but we also need to measure impacts on the real outcomes considered.
- Pre-competitive space vs competitive: what about vested interests? How to be careful about agenda setting and defining these mechanisms?
- Need to include more space for SMEs, as well as include the potential for businesses *that do not exist yet*. The NextGen is about building a domestic eco-system that keeps the feedback loop alive: the conception of private sector should include space for businesses that do not exist yet.

Breakouts for Creative Networking

In order to leverage the unique mix of individuals and common understanding developed over the course of the convening, a creative networking session was added to the agenda. During this session attendees informally shared how they could collaborate and take the next steps towards NextGen modelling in their own fields. Several participants shared their visions for how NextGen models could be applied to their work. One spoke on monitoring and data standards, particularly integrating ecological, pest, and diseases so that techniques and tools do not evolve in parallel with NextGen initiatives. Another shared her data collection experiences, and stressed the importance of data collection, harmonization and curation in an efficient fashion. The comments focused on enhancing the ability to model farming systems while recognizing there are boundaries on a farm, people, livestock, crops, climate, and land. The challenge is to get all of these aspects communicating with each other. In terms of steps, there needs to be a "bare bones" decision support system that is a pre-cursor to the NextGen which incorporates heterogeneity and "what-if" analysis. Furthermore, the information provided needs to be relevant at the farm level, which could mean using a crowdsourcing mechanism for local knowledge. There is no single application that is an answer, rather a panoply of applications existing in competition or complementing each other. Dave spoke on developing an application to support sustainable nutrition security. This application included metrics for characterizing sustainability and dietary quality outcomes of food systems. Previous applications did not incorporate many of the concepts presented, nor use open source data and methods, which are integral for a NextGen application.

Stan Wood's Concluding Comments

Stan Wood's final comments included a vision of what a NextGen modelling tool could look like, incorporating many of the themes discussed at the meeting. Stan then elaborated the Foundation's role in the next steps of NextGen modeling (advocacy, agenda setting), and where the initial focus communities should be. Stan pointed out the value in leveraging existing programs in order to take the next step to the NextGen of agriculture and food system modeling.

Stan suggested a user interface with graphical visualization components that provide the user with various capabilities such as a drag and drop tool to assemble model components and design a farming system model. The user would be able to define the study region, geographies and spatial and temporal scales. Data would be real time if applicable (weather, price, expectations on yield), and could incorporate elements of machine learning. Software would adhere to best practices and a code of standards for interoperability. In addition to further developing the science underpinning model components, the academic community would establish credibility through validation and model inter-comparison using standard data. The donor community will need to help finance developments, especially for work focusing on smallholder systems.

The Foundation's role should not be to pick "winners" but rather to facilitate communities, support standards and protocols, and accelerate model improvements and development of applications. One way to advance the work would be to take a Use Case in a priority region and use it to test and further develop models using the above approaches.

Supplemental Material Available in the Convening Dropbox

1. Background papers
2. Use Case Summaries (original 5 plus those developed at the convening)
3. PowerPoint presentations by Session[1]
    a. Use Cases
        i. Farm Extension Africa (Sander Janssen)
        ii. Early Warning Systems
        iii. Precision Agriculture (Bruno Basso)
        iv. Government (Joshua Elliott)
        v. Nutrition (Ian Foster)
        vi. Agribusiness, service providers, financial services (Richard Conant)
    b. Technical & Institutional Challenges
        i. Big Data Collection & Analytics (Bruno Basso)
        ii. Community Building (Charles Godfray)
        iii. Data Access & Interoperability (Medha Devare)
        iv. Model Assessment & Credibility (Joshua Elliott)
        v. Model Components & Connectivity (Richard Howitt)
        vi. Model Improvement & Tool Development (Amy Faye)
        vii. Private Public Partnerships (Pierre Sibiry Traore)
        viii. User Engagement & Experience (Jamie Kinney)

[1] Dropbox link available upon request.

Use Case Summaries

#1 [Author Team Use Case]

Title: Farm Extension in Africa

Farming System: small-holder

User: farm advisor

Decision: use new drought-tolerant varieties effectively

Capabilities: traditional knowledge and new information from research.

Limitations/Challenges: same models? Diverse conditions need to tailor advice to users. Pests & diseases.

Beneficiaries: farm households

Outcomes: higher, more stable yields. Improved nutrition.

#2 [Author Team Use Case]

Title: Developing & Evaluating Improved Crop and livestock models

Farming System: Small holder

User Ag research team/program

Decision: Investment/research priorities

Capabilities: existing crop, livestock, and systems models, optimization models

Limitations/Challenges: yield, limiting factors beyond water and nitrogen, livestock models: species composition, rangelands, feed estimates, tradeoffs, individual uncertainty and risk

Beneficiaries: research institution/farm population

Outcome: Improved technology

#3 [Author Team Use Case]

Title: Investment in Ag Development to support sustainable development

Farming System: small-holder

User: Analyst/Advisor

Decision: Investment priorities to achieve specific outcomes

Capabilities: crop, animal production, farming system behavior models, landscape decision models

Challenges: integrating multiple model components, data

Outcome: sustainable technology and adoption

Beneficiaries: NGO & Clients, donors

#4 [Author Team Use Case]

Title: Support for precision agriculture to include profit and sustainability

Farming System: commercial crops

User: management consultant

Decision: seeding, fertilization, pest management

Capabilities: mobile data, point based models

Limitations/Challenges: integrate data and model, data ownership/management, integrate with economic and environmental processes

Beneficiaries: Farm businesses, advisory services, the environment.

Outcomes: higher profit, less environmental impact

#5 [Author Team Use Case]

Title: Supplying food products that meet corporate sustainability goals

Farming system: Commercial crop

User: Corporate analyst

Decision: farm management and cropping best practices

Capabilities: real time weather and historical climate data, crop models, integrated platforms for farmers

Challenges:

Outcome: Sales, profit, sustainability objectives

Beneficiaries: Agri-business firms

#6

Title: Government sector evaluating investment alternatives/policies

Farming System: Mixed small-holder and commercial farms.

Information Users: Agencies, NGOs, Industry

Decision: Investment/policies to pursue (production/profit/health)

Beneficiaries: Value chain, national, farmers, society-environment

Information Requirements: High resolution data (soil, weather, management), resource availability and cost (cultivars, water, fertilizer, mechanization), current land use, local preferences, market demand, infrastructure (storage, roads, transport)

Challenges: cloud infrastructure, reading & feeding systems, web APIs, data collection and output interpretation, spatialize data

Outcomes: trade-offs, transparency

#7

Title: Sustainable Nutrition for the Planet under climate change and resource scarcity

Farming System: All, global-national scale nutrition security

User: National governments, international bodies

Decision: Climate change policy, food and ag policy, natural resource conservation, nutritional security, international agreements, trade policy

Capabilities: data availability mixed, model availability mixed, credibility a challenge, needs focus and innovation, model inter-comparison, model data, fusion, back casting, big data science

Challenges: Cross disciplinary, simplicity, need for public private partnerships.

Outcome: Our kids are not screwed.

Beneficiaries: Everyone

#8: Farm-level, Management Decision Support for small-holder farms

Farming System: mixed crop-livestock system

Decision Temporal Scale: pre-season – what crop/variety to grow

In-season – when to apply fertilizer/pesticide fungicide; livestock buy/sell; market conditions; grain buy/sell, Yield-risk: profit

Information Users: Patricia- young farmers, early adopters

Data Requirements: Good quality & short term season forecast; soils relevant (generalized for area); varieties/crops grown; livestock breeds, rangeland

Current Capabilities: models can currently do most of this, but require a lot of work to generate data

Challenges: data credibility/need for feedback locally valid; dissemination; pest markup; crowd source data

Outcomes: management options to buy/sell livestock

Beneficiaries: small-holder farmers; bi directional info flow; support mechanism (tech support/help) sms text

#9

Title: Technology Assessment for Dual Production systems for irrigated rice in Ghana

Farming System: small-scale out growers, subsistence & commerce.

Information Users: Agronomists with rice buyer, farmer (homesteaders)).

Decision: Management/Input, package rice variety. Land/labor allocation (rice, subsistence ag).

Information Requirement: Water req./availability. Seasonal climate forecast. Risks: water, pests, weeds, disease. Markers, Yield range, soil/geographic variability, labor availability.

Challenges: Seasonal (real time), monitoring to assess potential yield, disease/pest early warning (IPM), model integration, market projections, training in use/interpretation of new technologies, synthesis of model outputs, potential transformation of landscape or other users, transfer learning, adaptive management (information requirements), sustainability (pest/fertilizer application, soil management)

Outcomes: Reducing risk of undertaking rice farming, increasing discretionary income, reducing poverty

Beneficiaries: farmers and families, rice buyers, local communities

#10

Title: Modeling Supporting Early Warning Systems

Farming System: small-holder farmers in sub-Saharan Africa

Information Users: Extension Agents, NGOs

Decision: Livestock – when to sell? What to feed? How to plan for a drought? When to plant? What to plant? (Decisions throughout the whole value chain)

Information Requirements: Weather forecast, variety information, seed availability, soils, livestock management practices, stock density

Key Partners: Food and seed industry, input and output across the value chain

Scale: Farm and landscape

Challenges & Limitations: Vital science, early warning for US, Europe, and Australia. Requires High quality data, systems complex with heterogeneity. Weak institutions, lack of high quality data, lack of credibility and support systems (lots of risks). Understanding of uncertainty (risk?).

Beneficiaries: Smallholder farmers, consumers

Outcomes: improved farmer income. Increased food security in drought years.

#11

Title: Peri-urban & urban food Production

Farming System: Vegetable/fruit production with & around cities

Information Users: Urban land use planner

Decision: Investments in urban & peri-urban infrastructure to support local veggie production

Information Requirements: Resource needs to support local food production systems

Capabilities & limitations: Highly distributed base of "farmers" without an significant investment in data needs at the system level. Current capabilities and limitations are largely unknown.

Challenges: Outcomes: Greater availability of fresh fruits/veggies in cities

Beneficiaries: Consumers, new producers

#12

Title: Avoiding Jevon's paradox: can sustainable intensification really be sustainable?

Farming/System: Preserving eco-system services in East-Africa as intensification proceeds

Information Users: Government land use minister/stakeholder

Decision: Policy-setting & adaptation to evaluate/optimize trade-offs

Information Requirements: Recent real-time land, water, and resource use data in a map format.

Challenges/Limitations: General land-use models OK often difficult due to data limitations. Privacy Concerns with Big Data socio-economic, data currently very limited ability to model "governance" is weak

Beneficiaries: Public good maximized.

Outcomes: Improved capacity for governance, trade-offs. Income more equitably distributed; better management of natural resources.

#13

Title: Medium Term (3-5yr) decision support system

Farming System: Small holder (or independent larger); has several activities that would increase productivity; crop selection/rotation, dig a well multi-year crops (coffee, cacao), terracing, deep cultivation, soil amendments, equipment

Information User: extension or independent farmer

Decision: identify, prioritize activities over 3.5 years' time frame

Information Requirement: soil, weather, climate, market, cost/benefits of improvements

Capacities: use many current models, but over different scenarios over 3-5 years.

Challenges/Limitations: Data and trust in long term planning

Beneficiaries: Not identified

Outcomes: Longer term productivity, more effective, less “knee jerk” decisions. Grower makes decision for and over longer term period

#14

Title: My government is taking care of me

Farming System: Small-holder with rice/livestock mixed farming

Information Users: national & provincial governments

Decision: Subsidy (N fertilizer) and laying out options

Information Requirements: demographics; baseline, nutritional status; natural resources such as soil, weather etc, human habits (food preferences)

Current Capability/Limitations: Rice model (Eg FAO tool basket model); USDA/FAS data on Ag production; livestock models; human resources to communicate outcomes crop (livestock linkages)

Beneficiaries: general population, farmer groups

Outcomes: Information for policy makers to make better decisions on subsidies, environment and nutritional outcomes

#15

Title: Small-holder in Nepal woman farmer

Farming System:

Information Users: small-holder woman farmers

Decision: Crop choice/diversification to enable profit-making

Information Requirement: Use & market info, variety info/durability, transport costs, weather

Limitations: Advisor skills, availability/usability/reliability of local data (soils, weather, price)

Challenges/Limitations: info-interpretation/presentation (usability); info exchange (rather than “delivery’); technical accessibility (inequalities/digital)

Beneficiaries: farmer/families; community

Outcomes: increased profits, livelihoods; decrease water resources; increase decision-making/empowerment

#16

Title: Sustainable modeling of agriculture returns and tradeoffs (“SMART”)

Farming System: small-holder family in Sub-Saharan Africa

Information Users: Farms, Extension Agents, NGO, Policy Makers

Decision: Optimize income and land use

Information Requirements: Crop yields, soil types, water, weather, market access price data, farm size, management systems, labor costs

Current Capabilities & Limitations: Crop models, farming system models

Challenges: feedback loops, ground-truthing, ensuring local ownership, relevance of models

Beneficiaries: Farmers

Outcomes: Not identified

Next Generation Agricultural Systems Models

August 12-14, 2014
Bill & Melinda Gates Foundation
500 5th Avenue North, Seattle, WA
For any assistance, please contact Thomas Bogan (206) 770-2446

Context:

The purpose of the meeting is to gather a wide group of stakeholders to contribute to a scoping project called *Next Generation Agricultural Systems Models.*

We aim to generate interest and establish a community that can articulate a vision for what *Next Generation* tools can and must deliver and a roadmap of activities to build those tools and begin delivering the information products required. The scoping project began with the drafting of three background papers (provided to all participants of the workshop). We have invited people to the workshop who might not find each other in daily business, but whose skills could be applied to the challenges of planning and decision-making for global food security and poverty reduction for the world's smallholder farmers, the particular interest of the Bill & Melinda Gates Foundation agricultural development program. The three draft background papers on Next Generation models introduce use cases and ideas from an expert author team. Our goal is to have a broader conversation around use cases, technical challenges, data, and how we can build practitioner and user communities. A key outcome from the workshop will be an agenda for a body of work that will deliver new ways of analyzing agricultural systems and delivering improved information products for decision-makers (policymakers, investors, companies with products and supply chains involving agriculture, companies providing information services, etc).

Overview and Objectives:

The convening will take place over two and a half days. The ***first day*** will be focused around developing and discussing use cases that can respond to the information demands set forth by the global community addressing agriculture and food security. The ***second day*** will address technical challenges, identify potential partnerships, and think about how a community can be fostered to advance this work. The ***third (half) day*** will review the draft papers and gather the community's input and feedback to guide their completion.

The **expected outcomes** of this convening are to:

- Generate realistic use cases that meet real information demands and a roadmap of projects and partnerships that could implement them and address related technical challenges.
- Begin to develop a multi-sectoral and inter-disciplinary community dedicated to applying their diverse skills to delivering data, analysis, and information that serves global agriculture, food security, and poverty reduction decision-making.
- Solicit input and feedback into the papers which will begin to establish an agenda for this body of work.

Agenda

Day 1: August 12, 2014

8:30 – 9 **Registration**

9 – 9:30 **Welcome address & Setting the stage** Stanley Wood, John Antle
We have a diverse group of participants and are looking forward to some out-of-the-box thinking. We ask all discussants to participate genuinely as individuals and to remain in the "pre-competitive" sphere, and that participants keep focus on our main objectives.

9:30 – 10:45 **Opening Panel:**
"What do Next Generation Models need to deliver?"
Panelists: *Marc Sadler (World Bank), Hilary Parsons (Nestle), Rob Bertram (USAID), Lystra Antoine (DuPont/Pioneer), Wendy-Lin Bartels (University of Florida)*
Moderator: *Brian Keating*

10:45 – 11:15 **Break**

11:15 – 11:30 **Use cases intro** Sander Janssen
An introduction to the purpose of use cases, structure/format, and an example.

11:30 – 12:30 **Brainstorm: coming up with use cases**
Each table will be given a context to guide the brainstorm and develop a use case within that context. For example, farm advisory services in a developing world context. Please select a discussion leader at each table and a rapporteur who is responsible for capturing main ideas on the flipchart provided. If desired, you may move tables after the first 10-15 minutes of discussion. We ask that the discussion leaders and rapporteurs remain at their table for the duration of the exercise.

12:30 – 1:30 **Lunch**

1:30 – 2 **Use Cases Gallery Walk, Commenting & Voting**
All use cases, developed today and those from the circulated papers, are detailed on flip charts hung around the room. Please circulate and use the post-its to leave comments. In your participant packet, you will also find 2 stickers; please vote for your 2 favorite use cases by placing your stickers on the relevant flip charts. The use cases with the most votes will be discussed further in the breakout sessions that follow.

2:00 – 2:15 **Instructions**

2:15 – 3:45 **Use cases breakouts**
Based on the votes, the top 8 use cases will be discussed in breakouts. Participants are asked to select which breakouts to attend. Please select a discussion leader and rapporteur in each breakout group, which may or may not be the same people who served in those roles earlier, and who commit to remaining in that breakout for the duration fo the session. Others are free to circulate as desired. Please refer to the flipcharts and comments as a starting point for the discussion. Breakouts are then asked to address the following questions:

- *Who are the users? What are their decisions?*
- *Who will benefit, and how?*
- *What are some of the technical challenges involved?*

- *What input data are needed, and at what time and space resolution?*
- *How does the information need to be delivered?*
- *In what ways are current systems inadequate? Where are we currently failing to deliver for this use case?What are the gaps that need to be filled?*
- *What are the outstanding or new challenges that need to be addressed to implement this case (technical, parthersphips, institutional, etc)?*

3:45 – 4 **Break**

4 – 5 **Reporting back (with challenges for day 2)**
Rapporteurs from each use case provide a short report back (5-7 minutes each).

5 – 5:15 **Wrap-up** Peter Craufurd

6:30 – 9 **Dinner**
The Libarary Bistro (at the Alexis Hotel), 1007 1st Ave.
Transportation will be provided from in front of the Gates Foundation at 6:00pm

Day 2: August 13, 2014

8:30 – 9 **Opening session: Recap from Day 1** Peter Craufurd

9 – 10:15 **Responsive Panel:**
"What can the tech industry deliver to ag and food systems modeling today?
What needs to be developed for tomorrow? How do we better deliver information to users?"
Panelists: *Ashish Kapoor (Microsoft Research), Jason Cawley (Wolfram), Nancy Harvey (University of Chicago)*
Moderator: *Sander Janssen*

10:15 – 10:45 **Break**

10:45 – 12:15 **Technical and Institutional Challenges breakouts**
The themes for these breakout session will be selected from those identified during the use case breakouts on Day 1, and will be announced at the beginning of Day 2. This session will then provide an opportunity to work collaboratively towards a roadmap of potential partnerships and projects that could address them. Breakout groups should appoint a chair and rapporteur for the conversation who remain in the discussion for the duration of the session, others are free to circulate as desired.

12:15 – 1:30 **Lunch**

1:30 – 2:30 **Challenges report back with summary and recommendations**
Rapporteurs report back (5-7 minutes each).

2:30 – 3:30 **Panel: "Opportunities for Innovation in Open Environments"**
Panelists: *Jamie Kinney (Amazon), Liz Carolan (Open Data Institute), Dave Gustafson (ILSI Research Foundation), Caroline Figueres (IICD), Scott Malcolm (USDA)*
Moderator: *Tim Wheeler*

3:30 – 4 **Break**

4 – 5 **How do we foster this community?** Peter Craufurd

We are interested to see a diverse community and partnerships emerge, how can we foster this community going forward? A few "straw men" ideas will be described to get the discussion started, and then we will open to the group for thoughts, reactions, and ideas. Please consider the following questions:

- *Are there other communities and platforms already that are aligned?*
- *What new collaborations or partnerships are needed?*
- *What would individuals and organizations need/want to stay engaged?*

5 – 5:15 **Wrap-up** Peter Craufurd

5:15 – 5:20 **Housekeeping Announcements** Thomams Bogan

5:30 – 7 **Outdoor Reception**

Gates Foundation, Campus Heart (weather permitting)

Day 3: August 14, 2014

8:30 – 8:45 **Opening session: Recap from Day 2** Peter Craufurd

8:45 – 10 **World café on papers**

Breakout discussions on each of the papers, chaired by lead authors and joined by their collaborating authors. We ask all participants to cycle through multiple discussions in whichever order you choose. The goal of the session is for the authors to have a chance to solicit specific input and feedback on the papers from the group, in particular:

- *Do the papers articulate the vision?*
- *Are the challenges that have been highlighted by the group represented?*
- *What is missing?*
- *Should new use-cases be added to the papers; if so which one(s) and why?*
- *Link to use cases? Which elements should be elaborated more given the discussion on the previous days?*

10 – 10:30 **Break**

10:30 – 11:45 **Author Response** John Antle, Cheryl Porter, Jim Jones

Lead authors on each of the three papers will give a 15 minute summary of what they have heard and their plans to move forward with the papers.

11:45 – 12 **Closing words and thank you** Stanley Wood

12 – 1 **Lunch**

For anyone who does not have a flight to catch immediately, we will provide an informal lunch.

List of Attendees

Name	Organization
Phillip Alderman	International Maize and Wheat Improvement Center (CIMMYT)
Sandy Andelman	Conservation International
John Antle	Department of Applied Economics, Oregon State University
Lystra Antoine	DuPont
Ioannis Athanasiadis	Democritus University of Thrace
Elizabeth Bandason	Lilongwe University of Agriculture and Natural Resources
Wendy-Lin Bartels	University of Florida
Bruno Basso	Michigan State University
Robert Bertram	US Agency for International Development (USAID)
Thomas Bogan	Bill & Melinda Gates Foundation
John Bolte	Oregon State University
Kenneth Boote	University of Florida
Susan Capalbo	Oregon State University
Liz Carolan	The Open Data Institute
Jason Cawley	Wolfram Solutions
Richard Conant	Colorado State University
Peter Craufurd	International Maize and Wheat Improvement Center (CIMMYT)
Medha Devare	Consultative Group for International Agricultural Research (CGIAR) Consortium Office
Patrick Donahue	Mondelez International
Lance Donny	OnFarm
Kofikuma Dzotsi	University of Florida
Joshua Elliott	University of Chicago
Amy Faye	ISRA (Senegalese Institute of Agricultural Research)
Caroline Figuères	International Institute for Communication and Development (IICD)
Ian Foster	University of Chicago and Argonne National Laboratory
Giampero Genovese	European Commission
Charles Godfray	Oxford University
Dave Gustafson	ILSI Research Foundation
Daniel Halperin	University of Washington eScience Institute
Nancy Harvey	Institute for Entrepreneurial Studies at University of Chicago
Mario Herrero	Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia
Jerry Hjelle	Monsanto
Richard Howitt	University of California, Davis
Sander Janssen	Alterra, Wageningen University
Jim Jones	University of Florida
Ashish Kapoor	Microsoft Research
Brian Keating	Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia
Ermias Kebreab	University of California, Davis
Jamie Kinney	Amazon Web Services
Raymond Layton	DuPont Pioneer
Dilys MacCarthy	University of Ghana
Job Kihara Maguta	International Center for Tropical Agriculture (CIAT)
Scott Malcolm	Economic Research Service, US Department of Agriculture
Patricia Masikate	International Crop Research Institute for the Semi-Arid Tropics (ICRISAT)
Isabel Meirelles	Northeastern University
Andrew Moore	Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia

Siwa Msangi	Internationall Food Policy Research Institute (IFPRI)
Rafa Muñoz-Carpena	University of Florida
Carolyn Mutter	Agricultural Models Intercomparison and Improvement Project (AgMIP)
Vasey Mwaja	Bill & Melinda Gates Foundation
Tuu-Van Nguyen	Bill & Melinda Gates Foundation
Hilary Parsons	Nestlé S.A.
Cheryl Porter	University of Florida
Stephanie Price	Oregon State University
Wilhemina Quaye	Council for Scientific and Industrial Research, Ghana
Alan Rennison	Bill & Melinda Gates Foundation
Cynthia Rosenzweig	NASA GISS
Marc Sadler	The World Bank
Kate Schneider	Bill & Melinda Gates Foundation
Jetse Stoorvogel	Wageningen University
Pablo Tittonell	Wageningen University
Sibiry Traore	International Crop Research Institute for the Semi-Arid Tropics (ICRISAT)
Tim Wheeler	University of Reading
Christian Witt	Bill & Melinda Gates Foundation
Stanley Wood	Bill & Melinda Gates Foundation