Environmental Modelling & Software 62 (2014) 495–508

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

Harmonization and translation of crop modeling data to ensure interoperability☆

Cheryl H. Porter [a,*], Chris Villalobos [a], Dean Holzworth [b], Roger Nelson [c], Jeffrey W. White [d], Ioannis N. Athanasiadis [e], Sander Janssen [f], Dominique Ripoche [g], Julien Cufi [g], Dirk Raes [h], Meng Zhang [a], Rob Knapen [f], Ritvik Sahajpal [i], Kenneth Boote [a], James W. Jones [a]

[a] Agricultural & Biological Engineering, University of Florida, PO Box 110570, Gainesville, FL 32611, USA
[b] CSIRO Ecosystem Sciences, Sustainable Agriculture Flagship, PO Box 102, Toowoomba, 4350 Qld, Australia
[c] Washington State University, Biological Systems Engr., PO Box 646120, Pullman, WA 99164-6120, USA
[d] USDA ARS, ALARC, 21881 North Cardon Lane, Maricopa, AZ 85138, USA
[e] Democritus University of Thrace, Greece, Elec & Comp Eng, Building A, GR 671 00 Xanthi, Greece
[f] Alterra, Postbus 47, 6700AA Wageningen, Netherlands
[g] INRA, US1116 AGROCLIM, Centre de Recherche PACA, 228, Route de l'Aérodrome, CS 40 509, Domaine Saint Paul, Site Agroparc, 84914 Avignon Cedex 9, France
[h] KU Leuven University, Faculty of Bioscience Engineering, Department of Earth and Environmental Sciences, Division of Soil and Water Management, Celestijnenlaan 200E, Post Box 02411, B-3001 Leuven, Belgium
[i] Department of Geographical Sciences, University of Maryland, College Park, MD 20740, USA

ARTICLE INFO

Article history:
Received 21 October 2013
Received in revised form
30 July 2014
Accepted 2 September 2014
Available online 30 October 2014

Keywords:
Data structure
Crop model
Collaborative software design
Database
Schema
Harmonization

ABSTRACT

The Agricultural Model Intercomparison and Improvement Project (AgMIP) seeks to improve the capability of ecophysiological and economic models to describe the potential impacts of climate change on agricultural systems. AgMIP protocols emphasize the use of multiple models; consequently, data harmonization is essential. This interoperability was achieved by establishing a data exchange mechanism with variables defined in accordance with international standards; implementing a flexibly structured data schema to store experimental data; and designing a method to fill gaps in model-required input data. Researchers and modelers are able to use these tools to run an ensemble of models on a single, harmonized dataset. This allows them to compare models directly, leading ultimately to model improvements. An important outcome is the development of a platform that facilitates researcher collaboration from many organizations, across many countries. This would have been very difficult to achieve without the AgMIP data interoperability standards described in this paper.

© 2014 Elsevier Ltd. All rights reserved.

Software and/or data availability

The data translation tools described herein are provided to AgMIP researchers at the AgMIP toolshed (http://tools.agmip.org/). These tools are in a continuous state of modification as translators for new models are added, new functions are included in the DOME, and known bugs are fixed. All source code for AgMIP projects, including applications described herein, is maintained under source control on GitHub. These repositories can be forked or downloaded from https://github.com/agmip and are further documented on the AgMIP research site at http://research.agmip.org/display/dev/Projects. The GitHub.com repositories relevant to this paper are:

☆ Thematic Issue on Agricultural Systems Modeling & Software.

* Corresponding author. Tel.: +1 352 392 1864x243; fax: +1 352 392 4092.
E-mail address: cporter@ufl.edu (C.H. Porter).

http://dx.doi.org/10.1016/j.envsoft.2014.09.004
1364-8152/© 2014 Elsevier Ltd. All rights reserved.

Acronyms

ACE	AgMIP Crop Experiment data schema
ACMO	AgMIP Crop Model Output data schema
AgMIP	Agricultural Model Intercomparison and Improvement Project
APSIM	Agricultural Production Systems sIMulator (crop model)
DOME	Data Overlay for Multi-model Export
DSSAT	Decision Support System for Agrotechnology Transfer (crop model)
EPIC	Environmental Policy Integrated Climate (crop model)
ICASA-MVL	International Consortium for Agricultural Systems Applications-Master Variable List
JSON	JavaScript Object Notation
RIA	AgMIP Regional Integrated Assessment
RRT	AgMIP Regional Research Team
RZWQM2	Root Zone Water Quality Model 2 (crop model)
SALUS	System Approach to Land Use Sustainability (crop model)
STICS	Simulateur mulTIdiscplinaire pour les Cultures Standard (crop model)
WOFOST	WOrld FOod STudies (crop model)

- Applications:
  - quadui
  - acmoui
- Libraries:
  - ACE data translators (input and output):
    - translator-apsim
    - translator-agmip
    - translator-wofost
    - translator-stics
    - translator-dist
    - translator-cropgrow
    - translator-dssat
    - translator-generic-csv
  - ACMO data translators:
    - acmo-apsim
    - acmo-dssat
    - acmo-cropgrow
  - agmip-common-functions
  - agmip-core
  - ace-core
  - ace-lookup
  - ace-translator-parent
  - agmip-parent
- Others:
  - json-translation-samples

AgMIP software tools above are made available under the BSD 3-Clause license.

1. Introduction

The Agricultural Model Intercomparison and Improvement Project (AgMIP, www.agmip.org, Rosenzweig et al., 2013a) seeks to improve the capability of biophysical and economic models to characterize the risks of hunger and food insecurity due to the increasing pressures of population, food price volatility, water scarcity, land degradation, competition for arable land, weather extremes and climate change. AgMIP promotes a consistent, long-term research approach to agricultural model testing, improvement and application across modeling disciplines, regions and scales. The project encourages use of ensemble modeling approaches with climate, biophysical and agricultural economics models.

AgMIP protocols emphasize the use of multiple models because ensembles allow better characterization of the uncertainty associated with model outputs and because ensemble means of crop responses are more accurate than outputs from single models (Asseng et al., 2013; Bassu et al., 2014); consequently, data harmonization is essential to facilitate interpretation, storage, access, interoperability and publication of data products. Similar problems of interoperability across models, scales and data sets have been researched in the past using different approaches and solutions in different domains. There has been extensive development of modeling frameworks, which allow plug-and-play use of different models (e.g., OMS, David et al., 2013; KEPLER; TIME, Argent, 2004), or in a slightly different approach the development of shared standards implemented as in OpenMI (Gregersen et al., 2007; Knapen et al., 2013). Other efforts have focused on the developments of semantic techniques based on ontologies (Janssen et al., 2011; Villa et al., 2009). Recently, attention has gone to re-developing models as components that can be plugged into existing modeling frameworks, see Donatelli et al. (2014) for an example of such a solution. Modeling frameworks and standards such as the OpenMI have advantages in aligning multiple models in the same code base and allowing re-use of the models as components in different configurations and applications. Although successful in their specific modeling domains these modeling frameworks have seen little adoption by the original model developers or others to put models to new uses. This is due to the fact that some of those frameworks are too narrow as they are tailored to specific domain needs, while others, which are of generic nature, never invested enough in community building. Knapen et al. (2013) discuss reasons for slow or no adoption of OpenMI, which is currently proposed as an OGC (Open Geospatial Consortium; OGC, 2014) standard. One of the breakthroughs of the AgMIP work is that it focused early in community building, brought together almost all major crop model developer groups, and developed tools that have been used for numerous AgMIP studies. Results presented here is an outcome of a community process.

Interoperability and exchange of data among multiple models potentially increase research efficiency, allow models to use a greater variety of datasets, and facilitate comparability and ensembles. This paper describes the AgMIP approach to achieving data interoperability across crop models, which consists of first, establishing an efficient standardized data exchange mechanism with specifications defined in accordance with international standards; second, implementing a flexibly structured data schema to store experimental datasets; and third, providing consistent procedures for filling gaps in model-required inputs.

The following section describes the technical architecture, standards and components as part of the AgMIP data solution, including a description of components for user interaction. This is followed by a description of the development process, which leads to the architecture, the development of the standards and the different components. The usefulness of these software and modeling products is demonstrated in a section on applications, and finally followed by Discussion and Conclusions.

The work described herein has introduced several innovations to the crop modeling community by implementing state of the art technology. The design of the system, which was truly demand-driven and community-developed, has been adopted and tested extensively by a large research community. The system works with off-the-shelf models and with tools that are designed and maintained by individual crop model development teams. The AgMIP tools are based on recent technologies, including JSON data structures, a NoSQL database and a REST API to allow web-based access to the data. The implementation of a standardized data format with associated data translation libraries has not previously been successfully implemented in a large, truly global, agricultural modeling community.

2. Crop model data harmonization

This section presents a system of specifications, tools and methods of application for achieving crop data harmonization across models in order to facilitate simulation interoperability and long-term archiving. The system involves four parts: (1) developing harmonized formats for the data types associated with AgMIP crop modeling activities; (2) writing a library of translation tools to allow input of raw data, generation of model-ready input files and harmonizing model outputs; (3) providing a means of storing and accessing these harmonized data; and (4) providing applications and services to allow use of the data and the translation libraries.

2.1. AgMIP crop modeling data

Fig. 1 illustrates the challenges facing AgMIP researchers who wish to use datasets from diverse sources with ensembles of crop models and to use the simulated model outputs for comparison, model improvement, and assessments. These challenges include harmonizing the diverse data, translating data to the specific formats required for each model, translating model outputs to a harmonized format, and supplying model assumptions in a uniform way.

Diversity in specifications exists for datasets used to calibrate and corroborate crop models due to the influences of the original purpose of the experiment or survey and the measurement equipment and methodologies used. Well-designed field experiments may collect detailed records of soil properties, weather

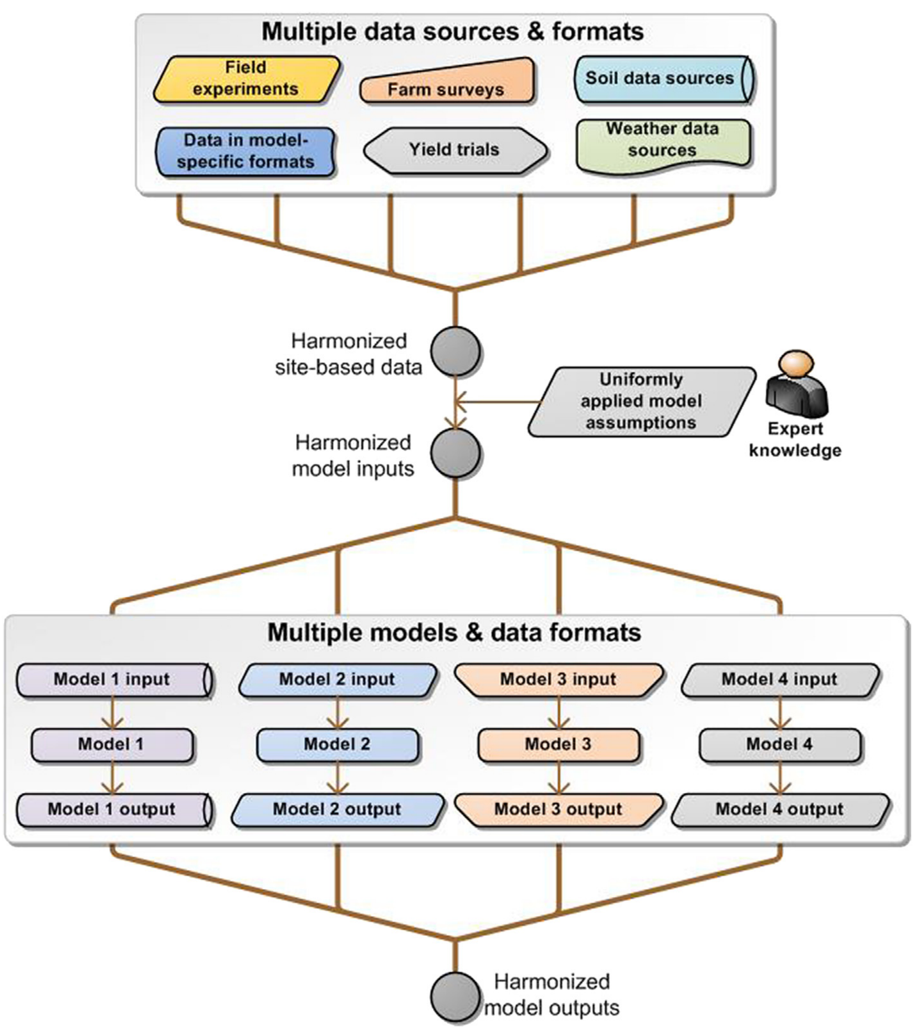

Fig. 1. Data flow diagram for AgMIP crop modeling activities. Source data, model inputs and model outputs have diverse and inconsistent formats, hampering ensemble modeling approaches.

conditions, and management practices such as tillage, fertilizer and water application dates, rates, methods and materials. Conversely, farm and household surveys may only query general management practices and estimations of seasonal total yields with a dearth of specifics and details. More poignantly with respect to crop model biophysical parameters, field studies often record crop phenology, canopy and biomass development and response to management practices, whereas such information is seldom solicited in surveys. These data sources also vary in storage and schema implementation formats, including databases, spreadsheets, tabular text files, structured XML and other specialized formats.

Most crop models are based on similar concepts (Boote et al., 2013), although they currently lack the ability to share inputs and parameters. While the various models may implement different algorithms for crop, soil and atmospheric processes, the driving data are generally similar. They typically include daily weather records, soil physical and chemical properties, and information related to management practices such as planting, fertilization, irrigation and harvest. These inputs can have quite disparate specifications, as models adopt different representations of the same phenomena. This could be a matter of design choice, preference, availability or even understanding of the underlying phenomena. As an example, soil data specifications may differ in the soil classification used (sometimes incompatible with each other) and in how vertical differences in a given soil profile are specified (e.g., number and depth of soil layers).

Likewise, although the core outputs from crop models are often similar, the formats are diverse and inconsistent. Model output formats generally parallel the input format structure. For example, models running within the Decision Support System for Agrotechnology Transfer (DSSAT; Jones et al., 2003; Hoogenboom et al., 2010) rely heavily on column formatted text for both input and output files, while the Agricultural Production Simulator (APSIM, Holzworth et al., 2014; Keating et al., 2003) relies on XML files, with a syntactic heterogeneity on the variable names and definitions.

In the past, there has been little coordination among crop modeling groups and hence, there is minimal standardization among formats associated with modeling datasets. The International Consortium for Agricultural Systems Applications (ICASA, White et al., 2013; Hunt et al., 2001) developed comprehensive standards for documenting field experiments. Although the ICASA definitions and formats were developed with input from multiple modeling groups, the standards have seen limited adoption, at least in the naming and formatting of input variables. A similar situation holds for the SEAMLESS ontologies (Athanasiadis et al., 2009; Janssen et al., 2011), which were developed in the SEAMLESS project (Van Ittersum et al., 2008). They have been developed to facilitate the agricultural model integration as scientific workflows. SEAMLESS ontologies define knowledge structures for crops, agricultural feasibility filters, agricultural management, and economic valuation of crop products, and agricultural and environmental policy, which are in principle the main types of data exchanged by the models. Issues related to translating data structures among model programming languages have been tackled by employing annotations in the ontology. However modular and rich these ontologies are, there was limited adoption in follow-up activities. Even though it was realized early in the process that the SEAMLESS ontologies offered added value for groups outside the project, there was never a coordinated effort to reach the community and involve these other groups. In contrast, in AgMIP, modelers from many modeling groups were involved from the beginning of the project.

Compounding the problem of model formatting inconsistencies is the issue of providing a consistent set of assumptions for information that is required by models but is not available in the datasets. Low information data sources in particular, but also highly detailed records, require modelers to make assumptions about the cropping system modeled in order to provide minimum input requirements to the models. For ensemble modeling, it is necessary to harmonize not only the recorded information but also the assumptions used to infer any missing model-specific input parameters. Examples of these types of model assumptions include initial soil water and nitrogen content, planting densities, and fertilizer type and application method.

Thus, three types of harmonized data have been identified for ensuring interoperability in AgMIP crop modeling activities: (1) site-based agricultural records collected from various sources; (2) uniform assumptions to supply minimum required inputs to crop models; and (3) crop model simulation outputs.

2.2. AgMIP Crop Experiment (ACE) harmonized format

The AgMIP Crop Experiment (ACE) harmonized data format was designed to provide a means of storing the highly variable types of data associated with AgMIP crop modeling exercises in a flexible and efficient schema. A data dictionary was needed to provide a common vocabulary, independent of data source and level of detail provided. The standards developed by the International Benchmark Sites Network for Agrotechnology Transfer project (IBSNAT, ICRISAT, 1984; Uehara and Tsuji, 1998) and subsequently revised by the International Consortium for Agricultural Systems Applications (ICASA) were designed to provide a comprehensive means of describing field experiments (White et al., 2013). The ICASA standards emphasize common vocabularies, clear relations among variables, and the ability to implement the specifications in various formats including text files, relational databases, spreadsheets, and XML. The foundation of the ICASA standards is the Master Variables List (ICASA-MVL, http://research.agmip.org/display/it/Data+Interoperability), which is organized in a hierarchical arrangement with major separations among descriptions of management practices, soil characteristics, weather, and measurements of crop and soil responses. The ICASA dictionary can be used to provide detailed descriptions of management practices and traits of soils and plants for crop experiments and other site-based agricultural data. Therefore, it was selected as the foundation of the ACE data definitions, but extended and modified for use within AgMIP. Changes included defining model variables that were not previously described, clarifying definitions that could have distinct interpretations in different models and accommodating metadata specifically required for AgMIP simulations.

AgMIP site-based agricultural datasets have a wide range of quality and quantity of records, which do not easily fit in a rigid schema. Therefore a flexible, non-relational architecture was selected. Data are managed using type-agnostic JSON (JavaScript Object Notation, www.json.org) key-value structures. A key-value structure is a data model that stores all data in `<attribute name, value>` tuples. This accommodates the storage of information for both simple observations and complex structured objects in an efficient, consistent and open-ended format. The key (attribute name) in each key-value pair corresponds to an ICASA-MVL variable; the value conforms to the variable definition, including units. All keys and values are stored as text strings so that formatting and significant figures remain as originally recorded. The ACE core library enforces data specifications, referential integrity, and formatting.

The core structure of ACE divides input elements to be provided to simulation models into three compartments that reflect the common drivers in most models, i.e. (a) experimental management, (b) soil characteristics, and (c) weather-related information. Each compartment is assigned an identification key, or hash code

(Secure Hash Standard, 2012), generated from the contents. Weather and soil structures are associated with experimental site data via these identification keys. This allows imposing one-to-many relations for efficient storage.

Additional structure is maintained by using the hierarchy implemented in the ICASA-MVL categories of Dataset, Subset, Group and Sub-group. These groupings allow initial conditions, management event details, soil layer parameters, and daily weather records to be recognized as consolidated structures. For example, a fertilizer event structure could consist of fertilization date, material, application method, nitrogen, phosphorus and potassium amounts, and other particulars. ACE only contains data that were actually recorded. Thus structures of the same kind may vary depending on the data source or purpose. This flexible structure allows missing information simply to be omitted without the need for placeholder values when there are no associated field measurements. In contrast, a table-style structure, such as found in a relational database, could result in large areas of the database containing null values.

Fig. 2 presents a sample JSON fragment showing a) an expanded experiment node and b) an expanded soils node. This dataset includes four soil nodes ([0] through [3]), five experiment nodes and four weather nodes. Within the expanded experiment node [0] in Fig. 2a, are key-value pairs including 'fl_lat: "28.38"', which indicates that the ICASA variable 'fl_lat' (field latitude) has a value of 28.38°. The units for each variable are defined by the ICASA standards. Also within experiment node [0], there are three nested structures containing grouped key-value pairs for observed data, initial conditions and management data. The nested structure, 'initial_conditions' is expanded to show the internal key-value pairs, such as 'icrn : "0.80"', indicating that the N content of the initial above-ground residue (ICASA variable 'icrn')

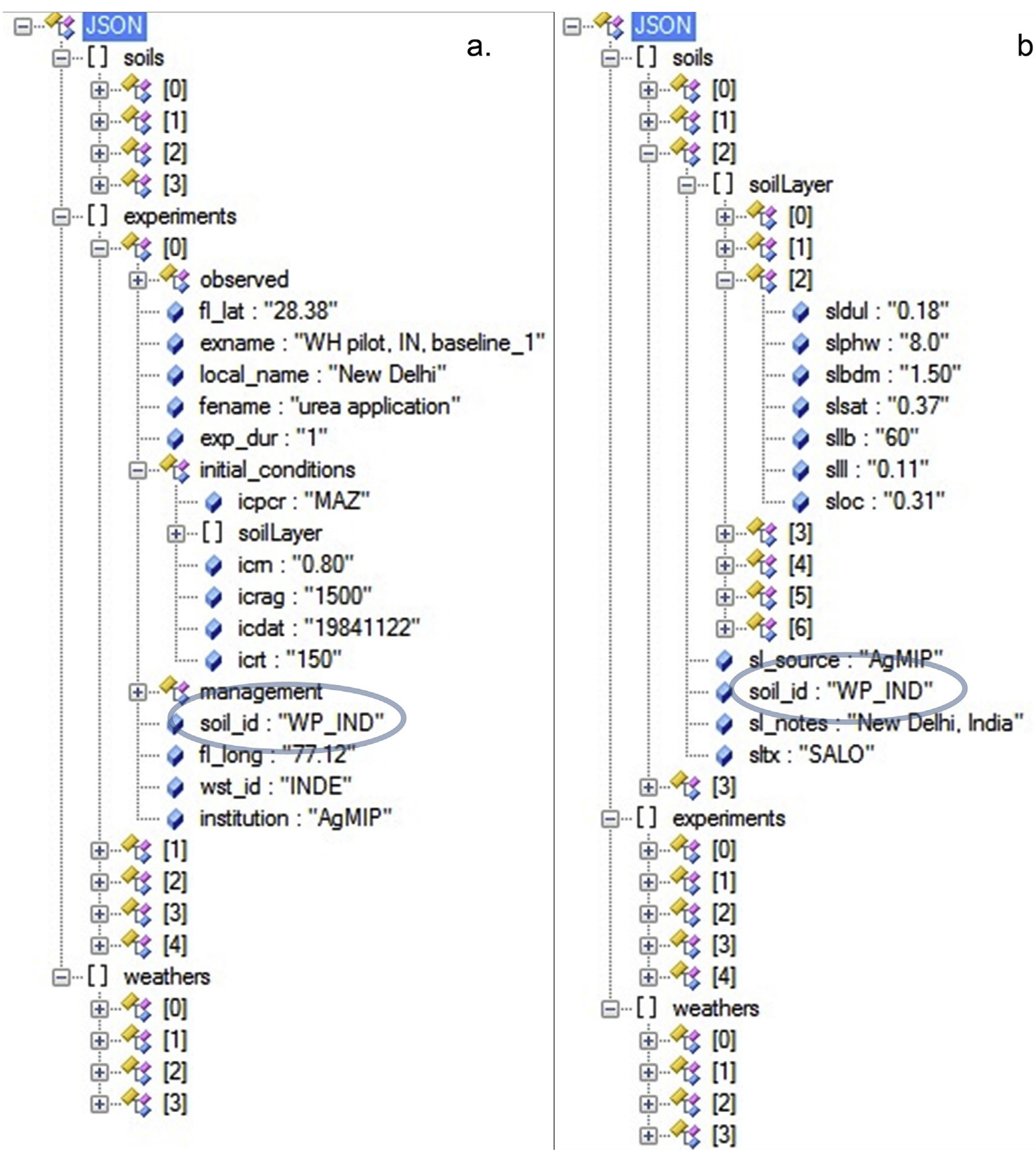

Fig. 2. Fragment of ACE JSON structure showing (a) expanded experiment data and (b) expanded soil data. The circled soil identifiers show how a specific soil profile is linked from the description of crop management to the section describing individual profiles.

was recorded as 0.80% by mass. The 'soil_id' for the expanded experiment in Fig. 2a corresponds to that of the expanded soil layer in Fig 2b, linking these two structures. In Fig. 2b, soil node [2] is expanded to show soil profile data, such as 'sltx', or soil texture, with a text value of "SALO", which is identified by the ICASA codes as a sandy loam. Soil layer [2] is expanded to show data for that layer, including soil bulk density (ICASA variable 'slbdm') of 1.50 g/cm^3.

Because of the diverse quality and content of datasets converted to the ACE format, the required content for datasets is minimal. The only mandatory data are field latitude and longitude (to indicate location) and crop species. All other data are considered to be optional, although a dataset with only these minimum data would have minimal value.

2.3. Data Overlay for Multi-model Export (DOME)

The Data Overlay for Multi-model Export, or DOME, provides a mechanism for representing model-specific assumptions for required inputs and has several potential uses for model intercomparisons. Foremost, this feature allows researchers to provide supplemental information in order to utilize incomplete datasets. For example, consider a farm survey dataset. From a crop modeling perspective, it is considered incomplete as is it lacks detailed information. However with additional interpretation, the data may be useful as the basis for a modeling exercise, particularly if many farms in a region were surveyed. Missing parameters required by models can be supplied through one or more DOMEs which allow researchers to make assumptions uniformly based on the best agronomic knowledge of cultural practices in a region. This is implemented as an overlay, which allows supplemented information to remain detached from field-measured observations and ensures that all models are provided with a consistent set of tenable assumptions. A DOME which is specifically used to provide a complete dataset for crop model simulations is referred to as a "field overlay".

In addition to filling in missing information, DOMEs can be used to impose hypothetical management regimens, in order to simulate adaptation or other "what-if" scenarios. These might be simulations of field conditions over multiple years with existing or imposed management and/or climate scenarios for seasonal strategy analysis. A "seasonal strategy" DOME encodes associated metadata describing the climate scenario, imposed management or adaptation strategies, socioeconomic strata and other keys which describe the hypothetical system.

"Rotational strategy" DOMEs are used to simulate long-term, continuous crop rotations. Like seasonal strategy analyses, the climate scenario and management regimen can represent historical, baseline or hypothetical future conditions. Unlike seasonal strategy analyses, the initial conditions of soil water, soil organic carbon, soil nitrate-ammonium, and prior crop residue are set only once at the beginning of the continuous simulation period, rather than at the beginning of each planting season.

DOMEs operate through the specification of functions, which compute model inputs based on other available data. Examples of DOME functions include an exponential decay function for distribution of soil layer-specific data such as organic carbon and root distribution; automatically computed planting dates based on rainfall records; and distribution of fertilizer applications when only the total fertilizer amount was reported. All DOME functions are fully documented on the AgMIP research site (http://research.agmip.org/display/it/Data+Interoperability). DOME data are stored as JSON structures consisting of metadata, function calls and associated function arguments.

2.4. AgMIP Crop Model Output (ACMO) harmonized format

Simulated outputs from the crop models are harmonized and archived in the ACMO (AgMIP Crop Model Output) format to be used for further analysis, aggregation and input to economic models. As with ACE, the ACMO schema conforms to the ICASA data dictionary. Unlike ACE, the ACMO data are uniform, with each dataset containing exactly the same elements. Crop and economic modelers collaborated in the selection of the ACMO variables to ensure that these data are sufficient for use in the economic models, yet readily available from the outputs of most crop models. The full set of ACMO variables is listed and defined on the AgMIP research site (http://research.agmip.org/display/it/Data+Interoperability). Because of the consistent content of metadata and simulated outputs, regardless of model or simulation scenario, ACMO data are stored in a table format with each row in the table representing a simulation for a single site and single season.

The ACMO data consist of a total of 52 variables; but only 10 of these are simulated outputs from the models. The remaining variables represent database hash codes (e.g., for weather and soil data), simulation metadata (e.g., climate id, population stratum represented by the simulation), field metadata (e.g., latitude, longitude, research institution), field data or model input parameters (e.g., cultivar, planting date, observed data). These data and metadata fields are stored with the model outputs to provide the necessary data provenance requirements for the workflow.

2.5. Other aspects

2.5.1. Data provenance

Reproducibility and integrity of archived data is extremely important for AgMIP researchers. The data at the end of the analysis pipeline must be linked to the sources as well as to all intermediate modifications. The metadata for the ACE, DOME and ACMO enforce these provenance requirements.

Metadata for each ACE dataset include information regarding the source of field, weather and soils records. Weather metadata also include flags which denote modifications to weather observations or inclusion of data from other sources, such as solar radiation from satellite observations (Chandler et al., 2013), where data were not recorded at the weather station.

Simulated model outputs in the ACMO file are connected via metadata to the raw data collected by researchers; the transformations and additions supplied by DOMEs; and the model name and version used for each simulation. The metadata fully identify the modeled scenario and provide the identification of simulated model outputs with the ACE datasets and DOMEs used to generate the model inputs.

Archived ACE, DOME and ACMO datasets are tagged with unique hash codes which are generated from the data contents. This allows verification of the data integrity at every step in the processing chain. Hash codes are used to link the ACMO simulation outputs to the ACE and DOME data used to drive the simulation. These hash codes are generated upon writing the ACE and DOME files, prior to executing DOME transformations or translations to prevent impacting runtime performance. When an existing archived ACE dataset is modified in a way that would affect simulated outputs for one or more models, a new unique hash code is generated and the new dataset is stored, linked to the original. In this way, a history of modifications is maintained. Modified datasets are available to users, but a warning message is issued that the data may have been superseded.

2.5.2. Collaborative design process

AgMIP has demanded a high level of cooperation at all levels among researchers around the globe. This is particularly evident in the collaborative design of the data schemas used for ACE, DOME and ACMO. The initial ACE data structure was developed at the University of Florida and described by Villalobos (2012). Five software development sprints were held in 2012 and 2013 to develop the translation tools specific to each participating crop simulation model. During each sprint, crop model developers or experienced model users were paired with software developers in a one-week intensive workshop to rapidly develop the model-specific translation utilities. At the early workshops, in-depth discussions took place among the modelers to improve and solidify the form and function of the ACE data schema so that it could represent a wide range of cropping systems, field experiments and models. This has resulted in a robust, flexible, extensible and efficient design.

The ICASA-MVL has evolved and will continue to evolve as more models and modeling capability are included in AgMIP activities. Changes to the ICASA-MVL were often the outcome of the development sprints, where collaborative work revealed errors, omissions and deficiencies, thus allowing improvements to the list. AgMIP scientists are currently discussing similar data dictionaries and harmonization methods for livestock, pest, and disease models to allow inclusion of these as components in future modeling frameworks.

The list of variables included in the crop simulation model outputs, or ACMO variables, were developed collaboratively with AgMIP crop and economic modelers in an iterative design process. This was an important first step for the interdisciplinary team to begin harmonizing scientific vocabularies so that economists and crop modelers could understand the data flows intersecting these two very different domains.

AgMIP data formats and metadata standards are also being coordinated with other groups that deal with crop modeling and site-based agricultural information. AgMIP members worked collaboratively with representatives of the Climate Change, Agriculture and Food Security (CCAFS) research program of the Consultative Group on International Agricultural Research (CGIAR) partnership to ensure that data and metadata are compatible and can be shared between the online databases for the two organizations. Some AgMIP data are currently shared on the CCAFS online site-based database (http://agtrials.org/).

The MACSUR project (www.macsur.eu/) was developed to contribute to the AgMIP initiative with a European focus, but also to contribute to global AgMIP activities. Researchers from AgMIP and MACSUR are collaborating on the development of databases, translation tools and user interfaces as an efficient use of the limited resources available to both groups.

As part of the collaborative efforts for AgMIP software development, all AgMIP software products are developed in an open source environment, hosted on GitHub (www.github.com/agmip). The AgMIP development protocols ensure that these tools can be implemented in multiple computing environments and on multiple platforms for modeling, data preparation and analysis.

3. Implementation and demonstration

3.1. Data translation libraries

Three types of translation tools were developed to support AgMIP crop modeling activities: 1) ACE input translators, which allow data to be imported from various formats into the ACE harmonized format, 2) ACE output translators, which take combined ACE and DOME data and generate crop model-ready files for crop simulation models and 3) ACMO translators, which convert crop model outputs into harmonized format. In most cases, these programs were developed as a library of translation tools which can be implemented in multiple ways including desktop applications, Web services and parallel processing platforms for large-scale modeling applications. Table 1 lists the status of ACE input, ACE output and ACMO translators that are completed or in development.

Model developers who want to participate in AgMIP multi-model activities generally attend an AgMIP development sprint and begin with development of an ACE output translator, which allows their model access to AgMIP datasets. Development of an ACMO translator is the next step, which allows outputs from a particular model to be used in analyses, visualizations and comparisons with other models. ACE input translators are an optional development path for modeling teams who wish to make their site-based crop data available to the wider modeling community.

The utility and accuracy of the DSSAT translators was tested by taking a DSSAT-format experiment, converting it to ACE format using the DSSAT ACE input translator, then converting back to DSSAT format using the DSSAT ACE output translator. The resulting input and output files were compared. The simulation output files were equivalent for the pre- and post-translation versions of the input data, with the only differences being the date-time stamp on the output files. Pre- and post-conversion input files were actually more dissimilar than the output files, but they contained equivalent information. For example, the translated DSSAT input files produced more complete labeling of missing data for some data fields which were not actually used by model; and the soil file contained

Table 1
Status of translator development for 13 participating models. These translators are in various stages of development including “Operational”, “Nearing completion”, “Under development”. Use of a “–” indicates that development of the translator has not begun.

Crop model	ACE input translator status	ACE output translator status	ACMO status	Reference for crop model
APSIM	Under development	Operational	Operational	Holzworth et al., 2014; Keating et al., 2003
AquaCrop	–	Nearing completion	Under development	Steduto et al., 2009; Raes et al., 2009; Hsiao et al., 2009
CropGrow-NAU	Operational	Operational	Operational	Yan et al., 2004
CropSyst	Under development	Operational	Operational	Stöckle et al., 2003
DSSAT	Operational	Operational	Operational	Hoogenboom et al., 2010; Jones et al., 2003
EPIC	–	Under development	–	Izaurralde et al., 2006
InfoCrop	–	Under development	–	Aggarwal et al., 2006
ORYZA2000	–	Under development	–	Bouman and Van Laar, 2006; Bouman et al., 2001
RZWQM2	–	Under development	–	Ahuja et al., 2000
SALUS	–	Operational	–	Basso and Ritchie, 2012; Basso et al., 2010
SarraH	–	Under development	–	Sultan et al., 2013; Oettli et al., 2011
STICS	Under development	Operational	Operational	Brisson et al., 2009; Brisson et al., 2003
WOFOST	–	Operational	Operational	Van Diepen et al., 1989

only the single soil profile needed by the simulation, rather than the complete soil database. Other AgMIP models lack input translators, so similar tests could not be performed.

3.1.1. ACE input translators

Input translators facilitate importing raw data from a variety of formats including spreadsheet templates with ICASA variable headers and definitions, model input files, and AgMIP weather data.

The AgMIP Climate Team developed a standard format for disseminating daily weather records, as described in the AgMIP Protocols (Rosenzweig et al., 2011). Translators to convert these weather data to the ACE harmonized format were the first AgMIP input translators to be developed. Additional ACE input translators have since been developed to convert daily weather data from NetCDF format and from several model-specific formats.

Several spreadsheet templates are provided to AgMIP users to accommodate differing levels of detail and widely varying management options, including templates for farm survey data, rice paddy management systems, fertilizer trials, breeder trials and other options. The primary formatting requirement is that the data elements entered into a spreadsheet must fall under a variable name column which uses the ICASA standard naming approach. All values entered into the template must be in the format and units specified in the ICASA-MVL.

Model-specific input translators import model data directly from the input files for some of the models actively used in AgMIP research. As of this publication, only the DSSAT and CROPGROW-NAU ACE input translators were operational. Additional translators for APSIM, CROPSYST and STICS are under development. These model-specific input translators will allow many existing datasets associated with each modeling platform to be harmonized and made available to AgMIP researchers, in the formats specific to each model. The time requirement for development of model-specific ACE input translators is large. Development of these translators has not been a priority in AgMIP because of the urgent need and limited resources to provide ACE output translators for use by crop modelers. Future AgMIP development sprints, if funded, would allow input translators to be developed for additional models, thus facilitating rapid addition to the AgMIP Crop Site database of calibrated field experiments from the databases of modeling groups around the world.

The main incentive to the modeling teams for developing ACE input translators is that the database of crop experimental data, which has been calibrated and used extensively for that particular model, can be made available to the wider AgMIP community for model intercomparison and improvement activities. These activities have already resulted in the sharing of ideas between modeling teams which are leading towards model improvements among participating groups. In addition, the ability to share good quality data, including calibrations for cultivars, will likely result in wider acceptance and use of the source model.

3.1.2. ACE output translators

Translators from the ACE format to model-ready formats were developed for 13 models at the time of this publication (see Table 1). Five of the translators (APSIM, CropGrow-NAU, DSSAT, STICS, WOFOST) are linked into a desktop translation utility and are currently available for use by AgMIP researchers. Additionally, the CropGrow-NAU, CropSyst, and EPIC models have translators linked to the model user interface thus allowing these models to directly use the ACE data format. Other ACE output translators are in various stages of development and will be available for future AgMIP modeling activities.

The amount of time required to develop model-specific ACE output translator varies with the complexity of the model, the programming abilities of the translator development team, and the software approach used for development of the translator. In the case of the APSIM ACE output translator, a working translator was developed by a two-person APSIM team during the course of a 5-day development sprint using a template approach for a single crop. Additional crop and management capabilities were added to the translator in subsequent months as the need arose from the AgMIP crop modeling teams.

The primary incentive for developing an ACE output translator is facilitate use of a given model in AgMIP ensemble modeling activities, particularly the Regional Integrated Assessments, where the number of simulations is too large to allow manual translation of input data (see Section 4.1).

3.1.3. ACMO translators

ACMO translators are also currently available to AgMIP researchers for the APSIM, CropGrow-NAU, DSSAT, and WOFOST models. Other participating models, listed in Table 1, are also developing or have developed translators which are integrated into their model user interfaces. The ACMO file is partially generated when ACE and DOME data are translated to model-ready formats. The translation application (described in section 3.3) compiles all metadata which describe the simulation, including links to the ACE and DOME data used in the simulation; selected model inputs; and the crop model for which the translation was done. After crop model simulations are complete, the model-specific ACMO translators combine these pre-compiled metadata with the simulated outputs into the ACMO harmonized file, in comma delimited format. The number of simulated outputs stored in the ACMO format was intentionally limited to a carefully selected subset (10 variables total) which represent biomass, phenology, and environmental conditions and which are readily available outputs from most models. All units are converted as necessary by the model-specific ACMO translators to ensure conformance with the ICASA MVL specifications. In addition, the specific model version used for the simulations is recorded in the ACMO to facilitate reproducibility of results.

Model-specific ACMO translators can be programmed in a few hours, typically, due to the relatively simple data and formatting requirements and the fact that the majority of data are pre-compiled by the ACE-output translation application.

3.2. AgMIP data storage solution

The database for ACE, DOME and ACMO data was designed using a hybrid system of PostgreSQL and Riak (http://basho.com/riak/). PostgreSQL is an enterprise-class, open source SQL database. Riak is an open source, distributed database designed for scalable, fault-tolerant operation and is within a class of NoSQL or non-relational databases. The AgMIP implementation consists of multiple data nodes at facilities involved with AgMIP activities. The first data node is in testing phase at the University of Florida, with future nodes planned for University of Passo Fundo, Brazil; the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India; and other locations.

Fig. 3 illustrates the data services platform for AgMIP. The distributed data nodes comprise the bottom layer of the three-tier system. A separate metadata layer stores and indexes a searchable subset of the ACE data enabling rapid searches. The hierarchical structure allows for expedited import and retrieval of experiment parameters.

The top tier of the data services platform involves the user access controls, user interfaces and public websites. A single sign-on (SSO) system links user authentication for all AgMIP sites. A REST API is used to implement a standard interface for tools to query,

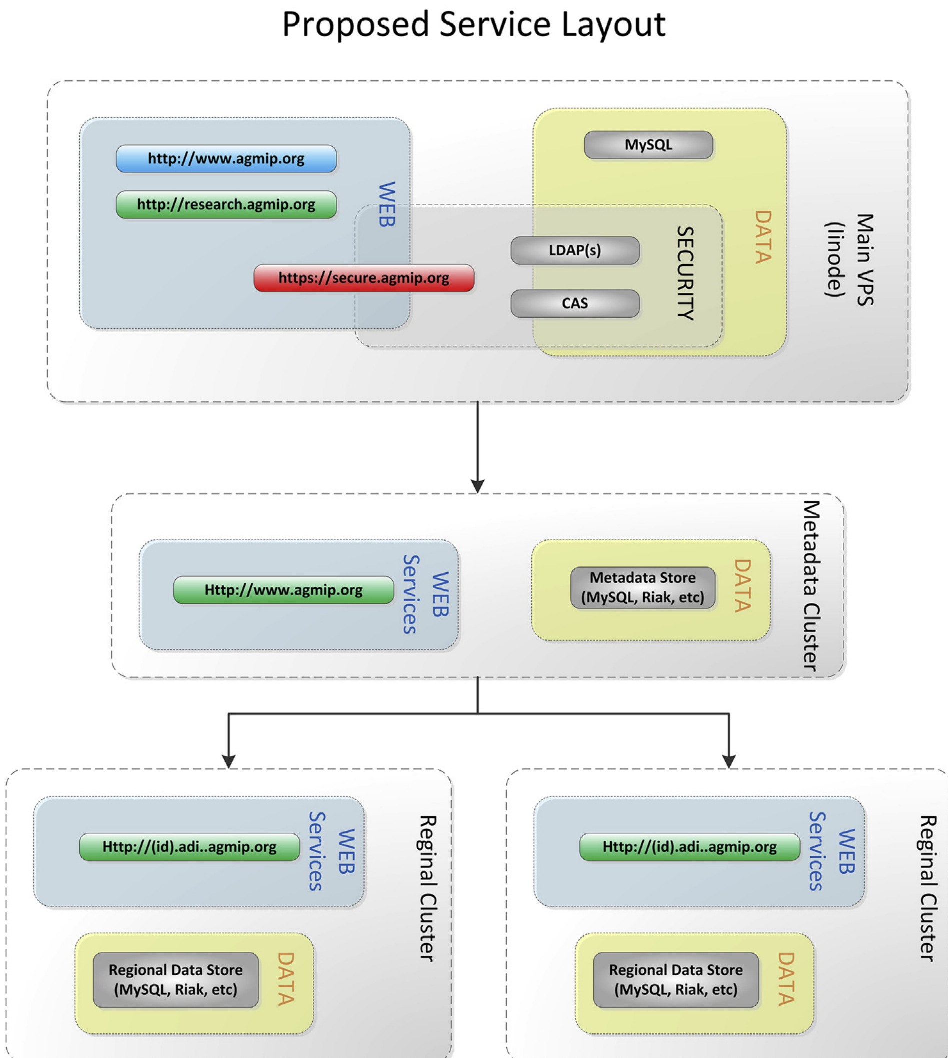

Fig. 3. Proposed AgMIP data services. Three tiers of data include the main virtual private server, the metadata cluster and multiple data nodes.

download and upload datasets to the database. In addition to the data.agmip.org site, this API can be used be external entities (such as data sharing partners) to discover AgMIP data or enable tool developers to interact with the database. Searching, download and viewing of datasets are not restricted, but only registered AgMIP users may upload to the database.

AgMIP researchers are active in many countries and regions around the world. The data which are served by AgMIP must be made readily available in a format that can be searched and accessed efficiently by researchers who may lack access to high speed Internet connections. Many ensemble modeling approaches use 30 years or more of daily weather data associated with each climate scenario, climate model, downscaling method, and time slice. Thus, weather datasets required for Regional Integrated Assessments (RIAs) can be large. All data are uploaded and downloaded from the server use a compressed format to minimize transfer latency.

3.3. Software applications for AgMIP harmonized data

The AgMIP Regional Research Teams (RRTs, see http://www.agmip.org/regional-integrated-assessments-handbook) in South Asia and Sub-Saharan Africa began work on their RIAs simultaneously with the AgMIP Information Technologies (IT) team's development of the data translation tools intended to enhance the researchers productivity. For this reason, it was necessary to quickly implement rudimentary tools to allow raw data to be converted to ACE format and thence to formats needed by the multiple models used in RIAs. At the same time, the AgMIP IT team began development of a more robust web-based user interface to allow users to upload, search and download harmonized data from an online database. The following examples illustrate the flexibility of the designed protocol which allows implementation using different tool sets.

3.3.1. Desktop utilities

All desktop data translation applications described herein are available for download from the AgMIP toolshed at http://tools.agmip.org/. AgMIP RRTs currently use these tools in their RIAs.

The QuadUI desktop utility implements ACE input translators, DOME functions and ACE output translators in an interactive java-based application. Users prepare raw data, translate to harmonized format, optionally apply DOMEs, and translate the combined ACE-DOME data stream to model-ready formats for multiple models. Translated model-ready files are written to a directory selected by the user, with separate sub-directories for each model. Metadata files are also created at the time of ACE output translation for later use in generation of the ACMO files. QuadUI generates the compressed JSON files for ACE and DOME data (ACE-binary format), which can be later uploaded to the AgMIP online database.

The ACMOUI application is a separate desktop utility for harmonizing crop model outputs from multiple models using the model-specific ACMO translators. A harmonized ACMO file, containing both metadata and simulated model outputs, is generated with a unique name, based on simulation metadata.

Use of these desktop applications is discussed in Section 4.1 on the AgMIP Regional Integrated Assessments use case.

3.3.2. Model-integrated translators

The CROPGROW-NAU model, developed recently at Nanjing Agricultural University, reads the AgMIP ACE format directly, with development of I/O routines based entirely on the documentation on the AgMIP research website (research.agmip.org). This illustrates that ACE input and ACE output translators are not always necessary but could be incorporated within the model implementation.

The developers of the CropSyst and EPIC models developed translators that are integrated into their respective model user interfaces. This allowed these developers the flexibility to develop translation utilities in a language that was not necessarily compatible with the AgMIP Java applications, in these cases, C++ for CropSyst and Python for EPIC. The Python based translator for EPIC is estimated to require another man-month of development and testing. The C++ translator for CropSyst is operational, although further testing is required. An ACE input translator is also under development for CropSyst.

3.3.3. AgMIP Data Interchange

The AgMIP Data Interchange is a web-interface to access online AgMIP databases. As of this publication the interface was linked only to site-based crop modeling data (ACE, DOME and ACMO) but eventually it will link to other AgMIP databases including quality controlled climate data for historical and future scenarios, regional and global economic data, and GIS data to drive large-scale spatial modeling. The search interface for site-based data allows a user to search for and select datasets based on specified criteria including location and crop. The user may elect to view the metadata for the selected datasets; to download data in ACE-binary format (compressed JSON) or to download the metadata in human-readable format. Registered AgMIP users may sign on to upload their data in ACE-binary format, after agreeing to the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).

Access to some AgMIP data may be restricted either permanently or temporarily due to intellectual property rights, governmental restrictions or a researcher wishing to delay until publication of their study. In this case, only the metadata are stored and made available for viewing and searching. If a user wishes to access restricted data, they will be directed to contact the data owner.

3.3.4. Future software enhancements

The first generation desktop utilities (QuadUI and ACMOUI), currently used in AgMIP RIAs, were developed quickly in response to an immediate need by RRTs. It is recognized that the iterative process for generating and manipulating ACE and DOME data to produce successful and accurate simulations with multiple models can be a cumbersome process. The ICASA-MVL is extensive and the number of variables can be overwhelming for a user who is not familiar with the terms and structure of this data dictionary. The format of the DOME is essentially a Domain Specific Language (DSL), which can be difficult to learn for many researchers. For these reasons and others, we recognize that the next generation user interfaces must make data entry, manipulation and translation much easier for the user.

The AgMIP IT team has begun design of an enhanced desktop utility for data management that will combine the capabilities of QuadUI, ACMOUI and the AgMIP Data Interchange. Researchers will use this multi-function application for direct input of ACE and DOME data and to perform the iterative process of refining the inputs to generate useful and accurate simulations. Models which can be run with a command line will be accessed directly from the interface, with ACMO outputs generated automatically. Direct linkage to the server will allow a user to search the database, download datasets, manipulate the data, and upload results to the server, all within a single desktop application.

Under a National Science Foundation project called Framework to Advance Climate, Economic and Impact Investigations with Information Technology (FACE-IT, www.faceit-portal.org), researchers are developing a web-based, workflow platform to support AgMIP modeling activities. This platform will allow a researcher to create complex workflows within a visual interface, connect the workflow to diverse data sources and share the workflow, data and outputs with other researchers conducting similar types of research. The platform uses the Galaxy workflow engine, used extensively in data intensive biology applications, but under modification for use in Earth Science domains. For this project, the AgMIP Regional Integrated Assessment process (section 4.1) represents a use case and the data translation utilities developed for AgMIP have been implemented as workflow apps within the FACE-IT toolshed. It is envisioned that this platform will facilitate ensemble modeling efforts for researchers worldwide as it becomes more fully operational.

4. AgMIP use cases for data interoperability

4.1. AgMIP Regional Integrated Assessments (RIA)

The AgMIP RRTs in South Asia and Sub-Saharan Africa use AgMIP protocols to conduct Regional Integrated Assessments to quantify the effects of climate change on food security in their regions. The teams use multiple climate, crop and economic models to answer three key questions:

1. What is the sensitivity of current agricultural production systems to climate change?
2. What is the impact of climate change on future agricultural production systems?
3. What are the benefits of climate change adaptations?

The AgMIP RIA process requires collaboration among a multidisciplinary team to provide consistent and cohesive inputs at each phase of the process for climate, crop and economic analyses. These processes are described in detail in the AgMIP Regional Integrated Assessment Handbook (Rosenzweig et al., 2013b). This use case focuses only on the crop modeling process, which uses outputs of climate models as input and which generates inputs for the economic modeling phase of the assessments. A simplified data flow diagram for the crop modeling process is shown in Fig. 4, which shows the use of the three types of data translators and the two desktop utilities, QuadUI and ACMOUI.

Many sets of crop modeling simulations are required in order to evaluate current climate and technology conditions, future climate conditions with current technology trends and future climate conditions with adaptation. Each of these systems is simulated for multiple climate models and climate scenarios, multiple crop models, and multiple site-years. This results in a large number of

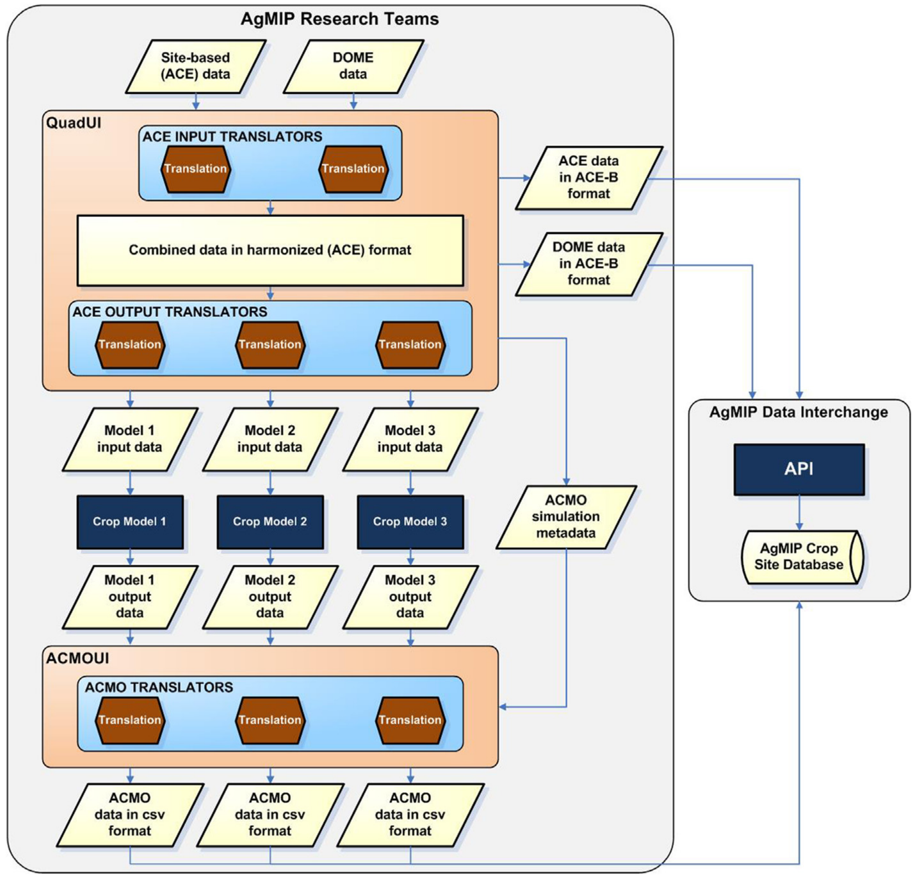

Fig. 4. Crop modeling data flow diagram for AgMIP Regional Integrated Assessments showing ACE input, ACE output and ACMO translators as implemented in desktop utilities QuadUI and ACMOUI.

simulations which are to be evaluated and compared and then used as input to regional economic models. A single modeler may generate several hundred thousand model simulations. Over 3 million model simulations were done by the AgMIP RRTs in Sub-Saharan Africa and South Asia. The processes described below in detail are for analysis of the historical conditions under which the survey data were collected. The data translation process for the analyses of seasonal strategies and future climate and adaptation scenarios is similar, but includes the application of additional DOMEs. The interested reader should refer to the AgMIP Regional Integrated Assessment Handbook (Rosenzweig et al., 2013b) for details of these analyses.

The use of translation tools in the RIAs is part of an iterative process. The site-based data are collected, converted to ACE format with units as specified in the ICASA-MVL, and input to a survey data template, typically as a spreadsheet. The templates are modified by the crop modelers to include the appropriate ICASA variables associated with the available survey data for each regional analysis.

A field overlay DOME is created to supply data required by the crop models, but not supplied in the farm or household survey data. DOME templates are available for users to modify with their site- and region-specific crop model inputs. Multiple DOMEs may be necessary to apply selective parameters to different soil types or socio-economic strata or to impose spatial variability among sites in an assessment.

ACE and DOME data are converted from spreadsheet format to comma-delimited format, then converted to a compressed ZIP format, either manually or using the AgMIP Data Assistant (ADA) desktop utility, also available on the AgMIP toolshed site.

Generation of ACE and DOME data which produce good, error-free simulations for multiple models and multiple sites is an iterative, often frustrating process. QuadUI is used to read the survey and DOME data from the zipped archives, combine them into a single data stream, and translate into user-selected model formats. Syntax errors, misspelling of key variable names or DOME names, or omission of critical information can all cause the translation to fail. When translations occur without error, the models may still fail if required model inputs are not provided. Or the models may run without error but produce results which do not adequately describe actual regional yields or other observed data. At each stage, the user must troubleshoot the problems using the QuadUI log, model outputs and error messages and by revising the survey and DOME sheets as needed.

Once the simulations for historical conditions have been successfully generated, the user runs ACMOUI to generate the ACMO harmonized data, which can then be analyzed and used as input to the regional economic models.

Analyses of multi-season systems for current and future climate conditions and adaptation scenarios follow the same process, and use the same field overlay DOMEs. Seasonal strategy DOMEs are applied to impose automatic planting date rules, future management regimens, and to simulate 30 weather years with current or future climatology for each site.

A final step is for the user to upload the working datasets using the AgMIP Data Interchange web interface. ACE and DOME data are uploaded in compressed ACE format; ACMO files are uploaded in comma-delimited format.

Results of the AgMIP Regional Integrated Assessments are being prepared for publication in the Handbook of Climate and Agroecosystems (Hillel and Rosenzweig, in preparation).

4.2. Model intercomparisons

Crop-specific model intercomparisons form one type of collaborative activity undertaken by AgMIP researchers. These model comparisons allow researchers to quantify the uncertainty in the crop modeling process and also to highlight areas where models can be improved to better predict the responses to climate variables. The AgMIP wheat model intercomparison project (Asseng et al., 2013) began in 2011 and included 27 wheat models, each with different input requirements and formats. Both the wheat and maize (Bassu et al., 2014) model intercomparisons showed that the mean of the ensemble predicted yields better than any single model over a range of environments, even for a small number of models. Additionally, an ensemble approach allows researchers the ability to quantify the uncertainty associated with the crop models, which can be high in comparison with the uncertainty associated with climate models and downscaling methods. AgMIP rice and sugarcane crop intercomparison teams are progressing similarly. Development of the translation tools had not yet begun at the time that the wheat model intercomparison was initiated and so the tools were not available for automated preparation of inputs for the multiple models. As a result, each modeling group converted project data manually to the various model-specific formats. In some cases, the groups may have applied inconsistent assumptions in order to supply inputs required by their model, but not explicitly provided to the modelers.

A second phase of the wheat model intercomparison, using observed elevated temperature datasets, is underway currently and is using AgMIP harmonized data entry templates to supply data to the participating modeling groups. The model groups who have working translators are able to quickly generate model input files. AgMIP IT intends to continue development of data translation tools and adding more model translators so that future model intercomparison and improvement activities may have all the benefits of the harmonized data structure and the multi-model overlays for consistent application of model assumptions and simulation boundary conditions.

4.3. Large-scale gridded modeling efforts

The pSIMS (parallel System for Integrating Impact Models and Sectors) framework, developed at the University of Chicago (Elliott et al., 2014), uses a massively parallel computation system to run the DSSAT crop simulation model over a large scale for gridded (raster) inputs. Recent integration of the AgMIP data translation tools has enabled this system to run multiple models on high performance computing platforms using identical inputs. The system allows data in NetCDF format to be converted to ACE format and thence to various model input formats. The pSIMS framework with multi-model capability was tested using DSSAT and APSIM for a regional study in East Africa as part of a USAID project and for a global analysis as part of the AgMIP Global Gridded Crop Model Intercomparison project (http://www.agmip.org/ag-grid/ggcmi/ and Rosenzweig et al., 2014).

A demonstration of the multi-model capability of pSIMS for multi-scale assessments of crop growth and climate impacts is presented by Elliot et al. (2014). In that paper, the authors describe four pSIMS campaigns that were conducted for maize in Africa from 1980 to 2010. The four campaigns involve two models (pDSSAT and pAPSIM), each simulating a) the full continent at 0.5° spatial resolution (10,301 unique grid cells) and b) the Southern/Eastern African countries of Zimbabwe, Malawi, Zambia, Tanzania, Mozambique, Ethiopia, Burundi, Rwanda, Uganda, Kenya, and Somalia at 0.25° spatial resolution (7778 unique grid cells). These campaigns are small compared to a standard global climate impact run (56,537 grid cells run over 150 years), but are useful to convey the versatility of the framework and the multi-model capability using AgMIP data translation utilities.

5. Discussion and conclusions

The AgMIP harmonized data solution was developed in response to a need by the AgMIP community of researchers for interoperability of data for multiple crop models for assessment of climate change and food security issues. Ensemble modeling in crop model impact assessments has two major benefits: (1) the mean from multiple models is a better predictor of yield than any single model, when applied over a range of environments and (2) the uncertainty from crop models can be assessed. To date, crop model uncertainties are generally greater than those associated with downscaled general circulation models (Asseng et al., 2013; Bassu et al., 2014).

The ACE and ACMO data formats coupled with the concept of the DOME and a library of translators provide a simple, flexible, and extensible means of handling data from diverse sources and for making them available to multiple crop models. The design of the data structures and modifications to the underlying ICASA-MVL were collaborative efforts between the crop modeling groups that participated in AgMIP Development Sprints in 2012 and 2013 and the economists who use outputs from the crop models. The ACE key-value schema provides an efficient and flexible means of defining and archiving the inconsistent site-specific data from diverse sources. The DOME provides a means for users to supply information such as assumed management details, initial conditions, simulation time span, or hypothetical management regimens consistently to multiple models. Harmonization of simulated model outputs enables consistent analysis of the results from multiple models and makes the data more readily available for input to economic models.

Our main scientific innovation involves the collaboration among diverse research groups to develop a community-driven standard for interoperability of data for use in multiple cropping system models. The use of these standards and software applications, as discussed in Section 4, is becoming widespread through AgMIP modeling activities. As of publication of this manuscript, approximately 50 AgMIP crop modelers in South Asia and Sub-Saharan Africa are finalizing results of Regional Integrated Assessments using the procedures outlined in Section 4.1 to analyze the impacts of climate change on food production systems in their regions. These analyses include over 3 million multi-model simulations using datasets prepared for DSSAT, APSIM and STICS models using the standards and applications described in this paper. These research teams are currently interpreting the results of their analyses, which will appear in Hillel and Rosenzweig (in preparation). It is beyond the scope of this paper to present these data and results, but our experiences convince us that model dataset processing for multiple models on this scale would have been impractical without the data interoperability standards and applications developed for AgMIP.

Another primary innovation associated with the design and implementation of the AgMIP data interoperability standards and tools, is the in the ability to conduct ensemble crop model simulations with confidence that the input data supplied to each model conform to a standard and represent equivalent information regardless of model data formatting differences. Current implementation of the tools by the AgMIP regional research teams use only two or three models to analyze hundreds of sites using low-quality farm survey data. Uniform application of assumed data via DOMEs is critical to the use of multiple models in these studies. As more model translators are developed by the crop modeling teams, analyses using larger model ensembles will be possible. This may be especially beneficial for the crop-specific model intercomparisons which emphasize high quality data from a relatively few sites for use by a larger number of models.

Crucial to the design of the AgMIP data interoperability tools is the expertise of the modeler in the use of their crop model. The tools and data provide only the input files for running the model, after which the modeler must step in, and check for any missing values and out-of-bounds parameters, then correct these through iterative passes with DOME inputs. The modeler remains responsible for calibrating the model, validating the outputs, and generating appropriate scenarios for different climate, technology or variety specifications. This focus on the practices of the modeler has an important implication in that the tools must be flexible enough to handle both the data used in the experiment and the model running the experimental data. This flexibility comes at a cost because for each combination of an experimental dataset and a modeling tool, an intervention of the modeler is required to fill parameters and to validate the model. This requires consideration of whether a given experiment can be run with a given model, and which experimental data can be adequately included. For example, it does not make sense to run an experimental dataset with detailed observations of nutrient dynamics through a crop model that lacks modules to compute nutrient dynamics.

The current implementation of AgMIP data interoperability tools uses spreadsheet templates and desktop utilities to allow researchers to harmonize and translate site-based data. These tools filled an immediate need for the researchers but remain difficult to use. Next generation tools will include more interactive user interfaces and web-based approaches. The existing data schemas and translators can be implemented in these next generation utilities without redesign.

The AgMIP translation tools depend on a large enough collection of datasets archived in a harmonized format to be of interest to modelers, and to stimulate new groups to join the efforts of the AgMIP community. Many of these site-based datasets have traditionally been poorly curated, sitting at experimentalists' desks in paper, personal hard drives and reports. Lately, there has been more attention to curating the data for long term preservation. AgMIP is well positioned to facilitate the sharing and harmonization of crop modeling data within its large global community of modelers. As a second step, links could be established with experimentalists gathering new data sets potentially useful for crop modeling, to share the data harmonization formats to allow ease of exchange and application. For example, developments of the crop ontology (http://www.cropontology.org/) developed by scientists involved in plant breeding could be linked to the ICASA-MVL.

The AgMIP data management system was designed to address the needs of a large, distributed, multi-model research community by adopting a rich, extensible vocabulary to describe site-based agricultural data from diverse sources; flexible data structures which allow both detailed and lesser quality records to be stored in an efficient database; and the ability to provide a consistent set of assumptions and parameters which can be extrapolated to multiple models. For the first time ever, researchers and modelers are able to use these tools to run an ensemble of models on multiple, harmonized datasets. This allows them to compare models, leading ultimately to model improvements. Perhaps the most important outcome is that the AgMIP project has provided a platform that facilitates researcher collaboration from many organizations, across many countries. The level of cooperation between these groups is unprecedented and has already resulted in data interoperability tools that will benefit the large crop modeling community of researchers by making data from a wide variety of sources available to any of the participating model users. This would have been very difficult to achieve without the AgMIP data standards described in this paper. As AgMIP grows to include more regions, researchers and modeling groups, the harmonized data format and translation tools will continue to be valuable resources.

Acknowledgments

The authors acknowledge the efforts of numerous model developers who participated in code sprints and otherwise contributed to the developing, testing and implementation of ACE, DOME and ACMO data schemas. In addition to the authors listed, significant contributions to the underlying concepts in the DOME were guided and supported by John Hargreaves, CSIRO. Steve Welch made possible the AgMIP Austin Hackathon, which was sponsored by iPlant Collaborative. We also want to recognize the larger AgMIP community; including the members of the AgMIP Regional Research Teams in South Asia and Sub-Saharan Africa, whose suggestions guided the development of the software products. Author contributions: C.P. and J.W. helped develop and implement the ICASA standards; C.V. designed the AgMIP Crop Site Database architecture, the ACE data structure, the DOME syntax, and most of the user interfaces and applications; M.Z. wrote much of QuadUI, ACMOUI and DOME functions. The table below lists contributions for development of the various model-specific data translation libraries. Much of the support for development of the ACE data format, the data translation tools and the desktop and web-based user interfaces was provided under a grant by the UK Department for International Development.

Crop model	Translator developers
APSIM	D. Holzworth, I. Athanasiadis
AquaCrop	D. Raes, R. Knapen
CropGrow-NAU	Z. Yan, L. Liu, X. Qiu
CropSyst	R. Nelson
DSSAT	C. Porter, M. Zhang
EPIC	R. Sahajpal
InfoCrop	N. Kalra, S. Pamuru
ORYZA2000	E. Manalo, B. Sailaja
RZWQM2	P. Bartling
SALUS	B. Baer
SarraH	H. Songoti, M. Zhang
STICS	D. Ripoche, J. Cufi
WOFOST	J. teRoller, D. van Kraalingen, S. Janssen

References

Aggarwal, P.K., Kalra, N., Chander, S., Pathak, H., 2006. InfoCrop: a dynamic simulation model for the assessment of crop yields, losses due to pests, and environmental impact of agro-ecosystems in tropical environments. I. Model description. Agric. Syst. 89 (1), 1–25.

Ahuja, L.R., Rojas, K.W., Hanson, J.D., Shaffer, M.J., Ma, L. (Eds.), 2000. The Root Zone Water Quality Model. Water Resources Publications LLC, Highlands Ranch, CO, 372pp.

Argent, R.M., 2004. An overview of model integration for environmental applications – components, frameworks and semantics. Environ. Model. Softw. 19 (3), 219–234.

Asseng, S., Ewert, F., Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Rötter, R.P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P.K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A.J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L.A., Ingwersen, J., Izaurralde, R.C., Kersebaum, K.C., Müller, C., Kumar, S.N., Nendel, C., O'Leary, G.,

Olesen, J.E., Osborne, T.M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M.A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J.W., Williams, J.R., Wolf, J., 2013. Uncertainties in simulating wheat yields under climate change. Nat. Clim. Change 3, 827–832.
Athanasiadis, I.N., Rizzoli, A.E., Janssen, S., Andersen, E., Villa, F., 2009. Ontology for seamless integration of agricultural data and models. In: Metadata and Semantic Research. Springer, Berlin Heidelberg, pp. 282–293.
Basso, B., Cammarano, D., Troccoli, A., Chen, D., Ritchie, J.T., 2010. Long-term wheat response to nitrogen in a rainfed mediterranean environment: field data and simulation analysis. Eur. J. Agron. 33, 132–138.
Basso, B., Ritchie, J.T., 2012. Assessing the impact of management strategies on water use efficiency using soil-plant-atmosphere models. Vadose Zone J. 11 (3) http://dx.doi.org/10.2136/vzj2011.0173.
Bassu, S., Brisson, N., Durand, J.L., Boote, K.J., Lizaso, J.I., Jones, J.W., Rosenzweig, C., Ruane, A.C., Adam, M., Baron, C., Basso, B., Biernath, C., Boogaard, H., Conijn, S., Corbeels, M., Deryng, D., De Sanctis, G., Gayler, S., Grassini, P., Hatfield, J., Hoek, S., Izaurralde, C., Jongschaap, R., Kemanian, A.R., Kersebaum, K.C., Kim, S.H., Kumar, N.S., Makowski, D., Müller, C., Nendel, C., Priesack, E., Pravia, M.V., Sau, F., Shcherbak, I., Tao, F., Teixeira, E., Timlin, D., Waha, K., 2014. How do various maize crop models vary in their responses to climate change factors? Glob. Change Biol. 20, 2301–2320. http://dx.doi.org/10.1111/gcb.12520.
Boote, K.J., Jones, J.W., White, J.W., Asseng, S., Lizaso, J.I., 2013. Putting mechanisms into crop production models. Plant Cell Environ. 36, 1658–1672.
Bouman, B.A.M., Kropff, M.J., Tuong, T.P., Wopereis, M.C.S., Ten Berge, H.F.M., Van Laar, H.H., 2001. ORYZA2000: Modeling Lowland Rice. International Rice Research Institute, Los Baños, Philippines, and Wageningen University and Research Centre, Wageningen, Netherlands, 235 pp.
Bouman, B.A.M., Van Laar, H.H., 2006. Description and evaluation of the rice growth model ORYZA2000 under nitrogen-limited conditions. Agric. Syst. 87, 249–273.
Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussiere, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillere, J.P., Henault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309–332.
Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2009. Conceptual Basis, Formalisations and Parameterization of the STICS Crop Model. Editions QUAE.
Chandler, W.S., Stackhouse Jr., P.W., Hoell, J.M., Westberg, D., Zhang, T., 2013. NASA prediction of worldwide Energy resource high resolution meteorology data for sustainable building design. In: Proceedings of the Solar 2013 Conference (American Solar Energy Society), April 16–20, Baltimore, Maryland. http://power.larc.nasa.gov/solar/publications/solar2013_0319.pdf.
David, O., Ascough II, J.C., Lloyd, W., Green, T.R., Rojas, K.W., Leavesley, G.H., Ahuja, L.R., 2013. A software engineering perspective on environmental modeling framework design: the Object Modeling System. Environ. Model. Softw. 39, 201–213.
Elliott, J., Glotter, M., Best, N., Chryssanthacopoulos, J., Kelly, D., Wilde, M., Foster, I., 2014. The Parallel System for Integrating Impact Models and Sectors (pSIMS). Environ. Model. Softw. Agric. Syst. Model. Softw. 62, 509–516.
Gregersen, J.B., Gijsbers, P.J.A., Westen, S.J.P., 2007. OpenMI: open modelling interface. J. Hydroinf. 9 (3), 175–191.
Hillel, D., Rosenzweig, C. (Eds.), 2014. Handbook of Climate and Agroecosystems: the Agricultural Model Intercomparison and Improvement Project (AgMIP). IPC Series on Climate Change Impacts, Adaptation, and Mitigation, vol. 3. Joint Publication with American Society of Agronomy and Imperial College Press-World Scientific (in preparation).
Holzworth, D., Huth, N., deVoil, P., Zurcher, E., Herrmann, N., McLean, G., Chenu, K., van Oosterom, E., Snow, V., Murphy, C., Moore, A., Brown, H., Whish, J., Verrall, S., Fainges, J., Bell, L., Peake, A., Poulton, P., Hochman, Z., Thorburn, P., Gaydon, D., Dalgliesh, N., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F., Wang, E., Hammer, G., Robertson, M., Dimes, J., Carberry, P., Hargreaves, J., MacLeod, N.C.M., Harsdorf, J., Wedgewood, S., Keating, B., 2014. APSIM – evolution towards a new generation of agricultural systems simulation. Environ. Model. Softw. Agric. Syst. Model. Softw. 62, 327–350. http://dx.doi.org/10.1016/j.envsoft.2014.07.009.
Hoogenboom, G., Jones, J.W., Wilkens, P.W., Porter, C.H., Boote, K.J., Hunt, L.A., Singh, U., Lisazo, J.I., White, J.W., Uryasev, O., Royce, F.S., Ogoshi, R., Gijsman, A.J., Tsuji, G.Y., 2010. Decision Support System for Agrotechnology Transfer Version 4.5 [CD-ROM]. University of Hawaii, Honolulu, Hawaii.
Hsiao, T.C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D., Fereres, E., 2009. AquaCrop-The FAO crop model to simulate yield response to water: III. Parameterization and testing for maize. Agron. J. 101 (3), 448–459.
Hunt, L.A., White, J.W., Hoogenboom, G., 2001. Agronomic data: advances in documentation and protocols for exchange and use. Agric. Syst. 70, 477–492.
ICRISAT (International Crops Research Institute for the Semi-Arid Tropics), 1984. In: Proceedings of the International Symposium on Minimum Datasets for Agrotechnology Transfer. 21–26 March 1983. ICRISAT Center, India, Patancheru, A.P. 502 324, India: ICRISAT.
Izaurralde, R.C., Williams, J.R., McGill, W.B., Rosenberg, N.J., Jakas, M.C., 2006. Simulating soil C dynamics with EPIC: model description and testing against long-term data. Ecol. Model. 192 (3), 362–384.
Janssen, S., Athanasiadis, I.N., Bezlepkina, I., Knapen, R., Li, H., Domínguez, I.P., Rizzoli, A.E., van Ittersum, M.K., 2011. Linking models for assessing agricultural land use change. Comput. Electron. Agric. 76, 148–160.
Jones, J.W., Hoogenboom, G., Porter, C.H., Boote, K.J., Batchelor, W.D., Hunt, L.A., Wilkens, P.W., Singh, U., Gijsman, A.J., Ritchie, J.T., 2003. The DSSAT cropping system model. Eur. J. Agron. 18 (3–4), 235–265.
Keating, B.A., Carberry, P.S., Hammer, G.L., Probert, M.E., Robertson, M.J., Holzworth, D., Huth, N.I., Hargreaves, J.N.G., Meinke, H., Hochman, Z., McLean, G., Verburg, K., Snow, V., Dimes, J.P., Silburne, M., Wang, E., Brown, S., Bristow, K.L., Asseng, S., Chapman, S., McCown, R.L., Freebairne, D.M., Smith, C.J., 2003. An overview of APSIM, a model designed for farming systems simulation. Eur. J. Agron. 18, 267–288.
Knapen, R., Janssen, S., Roosenschoon, O., Verweij, P., de Winter, W., Uiterwijk, M., Wien, J., 2013. Evaluating OpenMI as a model integration platform across disciplines. Environ. Model. Softw. ISSN: 1364-8152 39, 274–282. http://dx.doi.org/10.1016/j.envsoft.2012.06.011.
Oettli, P., Sultan, B., Baron, C., Vrac, M., 2011. Are regional climate models relevant for crop yield prediction in West Africa. Environ. Res. Lett. 6 pp014008. http://stacks.iop.org/1748-9326/6/014008.
OGC, 2014. Open Geospatial Consortium (accessed 31.3.14.). http://www.opengeospatial.org/.
Raes, D., Steduto, P., Hsiao, T.C., Fereres, E., 2009. AquaCrop-the FAO crop model to simulate yield response to water: II. Main algorithms and software description. Agron. J. 101 (3), 438–447.
Rosenzweig, C., Elliott, J., Deryng, D., Ruane, A.C., Arneth, A., Boote, K.J., Folberth, C., Glotter, M., Müller, C., Neumann, K., Piontek, F., Pugh, T., Schmid, E., Stehfest, E., Jones, J.W., 2014. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. PNAS 111 (9), 3268–3273.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Antle, J.M., Nelson, G.C., Porter, C.H., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorria, G.A., Winter, J.M., 2013a. The Agricultural Model Intercomparison and Improvement Project (AgMIP): protocols and pilot studies. Agric. For. Meteorol. 170, 166–182.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Antle, J.M., Ruane, A.C., Boote, K.J., Thorburn, P.J., Valdivia, R., Porter, C.H., Janssen, S., Mutter, C.Z., 2013b. AgMIP Guide for Regional Integrated Assessments: Handbook of Methods and Procedures Version 5.0. http://www.agmip.org/wp-content/uploads/2013/10/AgMIP-Regional-Integrated-Assessment-Handbook-v5.pdf.
Rosenzweig, C., Jones, J.W., Hatfield, J.L., Ruane, A.C., Boote, K.J., Thorburn, P.J., Antle, J.M., Nelson, G.C., Porter, C.H., Janssen, S., 2011. AgMIP Protocols. http://research.agmip.org/display/research/AgMIP+Protocols.
Secure Hash Standards (SHS). 2012. National Institute of Standards and Technology, FIPS 180–2.
Steduto, P., Hsiao, T.C., Raes, D., Fereres, E., 2009. AquaCrop-the FAO crop model to simulate yield response to water: I. Concepts and underlying principles. Agron. J. 101 (3), 426–437.
Stöckle, C.O., Donatelli, M., Nelson, R., 2003. CropSyst, a cropping systems simulation model. Eur. J. Agron. 18, 289–307.
Sultan, B., Roudier, P., Quirion, P., Alhassane, A., Muller, B., Dingkuhn, M., Ciais, P., Guimberteau, M., Traore, S., Baron, C., 2013. Assessing climate change impacts on sorghum and millet yields in the Sudanian and Sahelian savannas of West Africa. Environ. Res. Lett. 8 http://dx.doi.org/10.1088/1748-9326/8/1/014040, 014040. http://iopscience.iop.org/1748-9326/8/1/014040/article.
Uehara, G., Tsuji, G.Y., 1998. Overview of IBSNAT, pp. 1–7. In: Tsuji, G.Y., Hoogenboom, G., Thornton, P.K. (Eds.), Understanding Options for Agricultural Production. Kluwer Academic Publishers, Dordrecht, The Netherlands.
Van Diepen, C.A., Wolf, J., van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5 (1), 16–24. http://dx.doi.org/10.1111/j.1475-2743.1989.tb00755.x.
Van Ittersum, M.K., Ewert, F., Heckelei, T., Wery, J., Olsson, J.A., Andersen, E., Bezlepkina, I., Brouwer, F., Donatelli, M., Flichman, G., Olsson, L., Rizzoli, A.E., van der Wal, T., Wien, J.E., Wolf, J., 2008. Integrated assessment of agricultural systems–a component-based framework for the European Union (SEAMLESS). Agric. Syst. 96.1–3, 150–165.
Villa, F., Athanasiadis, I.N., Rizzoli, A.E., 2009. Modelling with knowledge: a review of emerging semantic approaches to environmental modelling. Environ. Model. Softw. 24, 577–587.
Villalobos, C., 2012. Using JSON & Riak for AgMIP. AgMIP White Paper. University of Florida. http://research.agmip.org/pages/viewpage.action?pageId=1212592.
White, J.W., Hunt, L.A., Boote, K.J., Jones, J.W., Koo, J., Kim, S., Porter, C.H., Wilkens, P.W., Hoogenboom, G., 2013. Integrated description of agricultural field experiments and production: the ICASA version 2.0 data standards. Comput. Electron. Agric. 96, 1–12.
Yan, Z., Weixiang, S., Weixing, C., Yongchao, T., 2004. Dynamic knowledge model and decision support system for rapeseed cultivation management. Trans. Chin. Soc. Agric. Eng. 6, 141–144.

Further reading

Donatelli, M., Confalonieri, R., Bregaglio, S., Cerrani, I., Fanchini, D., Fumagalli, D., Acutis, M., Rizzoli, A.E., 2014. The BioMA Platform for Biophysical Modelling in Agriculture: Enhancing Model Reuse via a Component-centered Architecture (in this issue).