Human-Machine Collaboration for World Modeling

James Allen[1,2] and Choh Man Teng[1]

[1]Institute for Human and Machine Cognition, Pensacola, FL 32502
[2]Univeristy of Rochester, Rochester, NY 14627

Abstract

Modeling ongoing events and processes in the world requires integrating research and systems developed in diverse fields. These fields use different concepts and vocabularies and each system uses its own idiosyncratic parameters and data formats. Attempts to construct large-scale simulations by hand have proven to be prohibitively time consuming and labor intensive, signaling a great need for automation. The inherent vagueness and uncertainty in the problems, however, make fully-automated model building and execution highly implausible. We argue that effective world modeling will result only from human-machine collaborative systems. We explore different levels of human-machine interaction, focusing on truly collaborative systems that we believe are necessary for world modeling. We also describe our preliminary efforts towards building such systems.

1 Introduction

The current state of the art in World Modeling involves an extremely labor intensive process, requiring person-years of effort by highly trained modelers. The bulk of this effort does not involve developing new modeling software (referred to as Quantitative Reasoning Engine (QRE)), but rather involves determining how a given scenario can be modeled using a combination of existing QREs and available data. Even once a network of models has been configured, the mechanics of running multiple models over a set of scenarios can still be very time consuming, taking up for example over 50% of the manual effort in the case of the Australian National Outlook (Hattfield-Dodds et al., 2015).

Current practice requires human analysts to do the bulk of the work, including identifying the capabilities and requirements of each QRE, researching historical data, harmonizing the data for input and connecting the output of some engines to other engines, possibly via some data transformation, and configuring the network of modeling engines to perform the analyses. It is not feasible to construct and run large scale models quickly enough to support the evaluation of options for time critical situations. For instance, when riots spread across Africa during the global food crisis of 2007/8, real-time analyses of the problem were not feasible and could only be performed retrospectively (Headey, 2011).

We believe the solution does not lie in building ever more complex systems that attempt world modeling autonomously. Realistic predictions and effective interventions require insights on modeling decisions, approximations and tradeoffs. It stretches the imagination to believe machines could identify relevant elements, approximate missing data and model functionalities, and build an effective computational model, all without human mediation. Furthermore, even if such a system existed, it is unlikely humans would believe the results and act on the system's recommendations unless the system can present intelligible descriptions of the analysis, answer clarification questions about its model choices and processes, and potentially iterate on the analysis based on initial results and feedback.

Significance We describe a model of collaborative human-machine interaction that can tackle problems too complex for either humans or machines to solve alone. We demonstrate this with CWMS, a prototype system that is able to integrate a handful of existing QREs into a single framework, and use collaborative interactions to build fairly complex models, iteratively explore situations and evaluate intervention strategies quickly and effectively.

Table 1. Interaction Styles with Increasingly Mixed Initiative between the User and the System.

Style of Initiative	Types of Interactions	Example	Possible Task Complexity
System Initiative (Single Initiative)	System prompts responses from user	Automated telephone interfaces (e.g., press 1 for sales, 2 for customer service,…)	Simple, predetermined sequences
User Initiative (Single Initiative)	User selects commands for system to perform	GUI-based interfaces (e.g., point and click) on laptops	No task models: user identifies the task and instructs system on individual steps
User Initiative with Clarification Sub-dialogues	User initiates system actions, which may result in questions clarifying information needed to perform task	Siri is asked about nearby restaurants, and responds by asking for the type of restaurant desired before answering	Simple tasks encoded by a set of information requirements that system must satisfy before performing task
Delegated Mixed-Initiative Task Execution	A task model directs who takes initiative for each subtask	Mars Rover autonomously performs navigation tasks while human decides high-level trajectory	Complex hierarchical task libraries, predefined for each task
Agent-based Collaborative Model Building and Execution	Human and system together define plans/models that then can be executed, modified and reevaluated	Research systems only, such as TRIPS, Bob and CWMS	System behavior defined by agents that engage in problem solving operations. Expandable.

2 Mixed Initiative Interaction and Collaboration

By collaboration we mean a form of mixed-initiative interaction where the machine and human contribute equally, analyzing a problem, developing solutions and experimenting with different conditions. There is a continuum of human-machine interaction styles (Table 1), ranging from single initiative systems to complex agent-based collaborative problem solving systems. Most interactive systems in use today are single initiative, with either the system or the user directing the interaction. For example, when you telephone a company you typically have to answer a series of automated questions before you can get anything done. This is a *system initiative* model. On the other hand, GUI-based interfaces on laptops and phones are essentially based on *user initiative* models. The machine presents a range of menus and buttons, and the user selects the next operation. In single initiative models, the system does not need to model and track the task the user is performing.

Conversational agents such as Siri and Alexa contain rudimentary models of tasks the user might perform (e.g., find a restaurant, buy a product) and use these models to clarify the information needed (e.g., what type of food, what price range). Once all the necessary information is collected, the task is executed by the machine. In such systems the task model is a set of information requirements (often called a frame) the machine needs to know before it can carry out an action.

Systems that move beyond frame-based models exploit AI planning techniques, where tasks are represented as sets of hierarchical and/or partially ordered actions, such that performing the actions in a manner consistent with the ordering constraints would accomplish the task. With such models the machine can reason about goals and ways to achieve them, and ultimately plan its own actions. This enables delegated mixed-initiative task execution, where a human sets the higher level goals and the machine works out the details. A good example is the MAPGEN system for the Mars Rover (Ai-Chang et al., 2004). In early rover missions, the spacecraft was controlled from Earth. It would receive simple instructions (e.g., move a meter forward, send back sensor readings) and then wait for the next instruction. Even without the 10 to 15 minute transmission delay, this was a very inefficient process. By instilling the machine with an ability to reason, execute actions and monitor results, communication needs are greatly reduced as the human only needs to specify the higher level goals. In these models, the system is acting as an *agent* rather than a passive computational device. In other words, the system is goal driven and reasons about its own actions.

When agent-based systems can reason not only about the task, but also about the interaction itself, we have *collaborative systems.* Such systems do not work from a predetermined set of plans, but rather create plans and models on the fly in a mixed initiative fashion. Examples include TRIPS (Ferguson & Allen, 1998), in which the system interacted with a human to jointly manage a transportation system; more recently Bob that interacts with a biologist to construct and execute pathway models (Gyori et al., 2017); and CWMS (described below). Such systems greatly enhance human problem solving capabilities by automating much of the details of constructing and analyzing models, including finding, accessing and approximating data, filling in model details not explicitly mentioned, coordinating model execution/simulation and generating explanations and displays summarizing the results and examining the assumptions.

This is not just a matter of providing a human with better tools, but rather designing systems that take initiative and actively collaborate. The human analyst brings in intuitions that help develop strategy, bridge modeling and data gaps, and suggest alternative approaches. The human also has a more global and general view of problems and can identify innovative strategies. The intelligent system, on the other hand, provides capabilities to perform quantitative simulations of specific phenomena and automatically construct complex model networks composing of individual modeling engines. It can also support extraction of relevant data from formal (i.e., databases) and informal (i.e., reports and papers) sources.

Modes of Communication As problems get more complex, point-and-click interactions become mired in a Byzantine maze of options requiring highly trained experts to operate. Our models support conversational natural language interaction, providing an intuitive interface such that the user can focus on solving the task instead of hunting down the next operation in the GUI. For example, suppose a user would like a map of Darfur in Sudan showing the accumulated rainfall in 2018. In a GUI-based system, the user might need to find and navigate to a page for map displays, look up the geographic coordinates (or standard code) for the area and enter these into different fields, select the parameter to display, and perhaps also the data source and calculation methods, and the format of the map. In contrast, with a language-based interface, this same task can be accomplished using simple natural expressions, for instance "*Show me a heat-map display of accumulated rainfall in Darfur in 2018*" or "*Can you find out how much rain there was in Darfar last year?*". Language allows users to state their goals in their own words, and have the system manage the details. Where applicable the system can provide defaults and ask followup questions ("*Darfar, Sudan or Darfar, Minnesota?*"), conducted also in natural language, thus alleviating the need for the user to master the intricacies of the technical terminology needed to use the system.

We believe the most effective systems are multi-modal, combining traditional graphical interfaces with natural language conversation. Language provides a natural and intuitive means to define goals, discuss the problem solving approach and refocus analyses based on initial results, whereas GUIs help ground the interaction and are ideal for summarizing and presenting information.

3 Information and Ontologies

Some information needed for a modeling problem exists in structured databases, but much more has to be obtained from more informal sources such as research papers and reports. Up-to-date knowledge about current situations (e.g., troop movement, Ebola outbreaks in new areas) is often only available from news and social media sources. A human analyst would need to identify the relevant databases and ways to gain access. For informal sources they would need to read the documents and extract the relevant data. This is typically such a laborious process that a comprehensive investigation would be a graduate student's thesis work, not achievable with the few hours available to the analyst.

Automated systems can offer a centralized server or data catalogue for existing data resources. However, each dataset is organized in its own idiosyncratic way and uses its own data vocabulary. For example, wheat is coded as WH in DSSAT (Jones et al., 2003) and "wheat" in APSIM (Holzworth et al., 2014). In the ICASA data dictionary (White et al., 2013), there are two possibilities: WHB (bread wheat) and WHD (durum wheat). As another example, "rainfall" could correspond to RAIN, RAIY, RTOT, PRCM and PREC, all just in DSSAT. Developing an automated query processing and data retrieval service for the centralized data server is challenging: The mappings between variables used in different resources may not

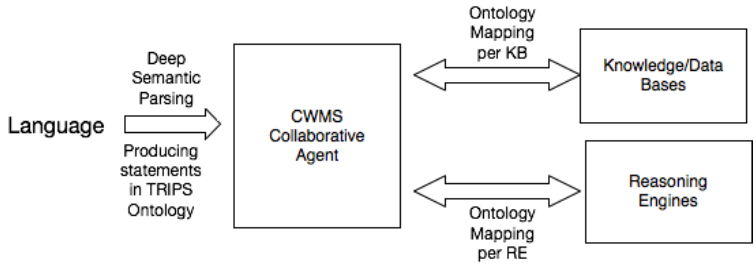

Figure 1: System Architecture Using Ontology Mappings

be one-to-one (e.g., wheat), and each variable could denote the same concept but with different ranges and units of measurement (e.g., rainfall) and may include a superset or subset of the target concept (e.g., RAIN includes both rain and snow).

We need to be able to characterize the content of available databases with a common general ontology that is intuitive and accessible for humans, such that the analyst can easily query the service for relevant resources and obtain access with data transformations appropriate for the intended models. Furthermore, unless this common ontology can be accessed via natural language queries, the user is back to browsing vocabulary lists and guessing what might be relevant. We approach this problem in two thrusts. First, the TRIPS deep broad-coverage semantic parser is able to map English sentences to a semantic representation in which the predicates are drawn from a comprehensive language-based upper ontology (Allen & Teng, 2017). Second, we have developed tools to map between our general TRIPS ontology and third-party ontologies. Figure 1 shows the abstract organization. Language is mapped to statements in the TRIPS ontology within which all collaborative problem solving operates. Data queries and service requests to knowledge bases and QREs similarly are represented in terms of the TRIPS ontology and then mapped to the specialized ontologies native to each resource. The responses are converted back to the TRIPS ontology.

The ontology mappings can be constructed automatically by reading definitions in the documentation (Allen et al., 2013; Allen & Teng, 2013). For example, HWAH in DSSAT is defined as "yield at harvest". Parsing this definition using the TRIPS parser gives a semantic representation in terms of the TRIPS ontology concept CAUSE-PRODUCE-REPRODUCE, situated in the COMMERCIAL-ACTIVITY of harvest. This enables us to link HWAH into our general ontology. Parsing "How much did the harvest produce?" the system would be able to understand this is a question relevant to HWAH in DSSAT. Using this ontology mapping framework, the analyst does not need to know the exact names used to index the items in a specialized data resource or QRE procedure. Rather they can provide a natural description of the concept and the system can identify and retrieve the appropriate data based on the mappings.

4 Building Models and Intervention Planning

The TRIPS parser is domain-independent, and needs only minor parameter adjustment to be deployed effectively for understanding text from different domains. It encompasses a range of conversational styles and vocabularies (e.g., collaborating with biologists, texting with teens) and can read documents of different styles and complexities (e.g., short stories, scientific papers). A version has been used for extracting and building causal models in biology (i.e., molecular pathways). Coupled into a collaborative problem solving system called **Bob** (Allen et al., 2015; Gyori et al., 2017), preliminary evaluations showed that biologists were able to build models to elucidate biology problems of significance much more efficiently and, most important, with a much higher success rate.

We have developed a prototype system for collaborative world modeling called **CWMS** (**C**ollaborative **W**orld **M**odeling **S**ystem). CWMS supports collaborative planning and simulations, integrating a number of existing QREs. These include (1) **DSSAT** (Jones et al, 2003): A point-based crop modelling system with detailed crop models for over 40 crops; (2) **pSIMS** (Elliott et al., 2014): A framework for climate impact simulations and global grid-based crop modelling using DSSAT and others; (3) **ABN** (Marchand et al., 2016): An agent-based out-of-equilibrium food shock model that computes changes in import, export, consumption and reserves due to an injected crop production shock; (4) **TWIST** (Schewe et al., 2017): A short term equilibrium model for computing market prices based on reserves, supply, demand, production and consumption; and (5) **IMPACT** (Robinson et al., 2015): A partial equilibrium multi-market economic model for analyzing longer term scenarios (30 years) involving agriculture, trade and food security.

Figure 2 shows an actual dialogue with CWMS. The user starts by suggesting a goal (1), which is accepted. The goal is refined by the system (2) and an initial course of action suggested (4). To support this interaction, the system reasons that malnourishment is an indicator of food insecurity and it is an indicator that can be computed by one of the available QREs. The system then constructs a plan to compute expected childhood malnutrition rates, based on baseline information on expected availability of food, etc. The answer is given in (6). The user then expresses a desire (7) to elaborate the scenario based on predictions that the coming year will be an El Niño year. The system recognizes the intention to run a new analysis with the new assumption.

To estimate the amount of change, CWMS needs to construct a plan to estimate the change in crop yields due to El Niño. A key problem in prediction is that future data (e.g., weather) is of course unknown. Thus CWMS constructs a plan to estimate the effect of an El Niño year by interpolating from simulations of the past 30 years, weighting the result of each year based on whether there was an El Niño event that year and how distant the year was. The change in crop yield is then fed to an economic model to compute the effect on food availability, which is then fed to a food security model to compute malnourishment rates. The system presents this plan to the user for inspection and asks the user whether to execute it (8). After confirmation (9), the system runs the plan and presents intermediate results shown in Figure 3 (10 and 11), and the final result that the number of malnourished children would increase by 5% (12).

The user follows up on the analysis with an open-ended question about what could be done to mitigate the effect of an El Niño event. The system does not have a specific answer. However, it does know a strategy to help the user explore intervention options by examining the known causal connections, and so it generates a causal graph and generates utterance (14). Using this information the user asks about the effect of

1 U: *Can you analyze food insecurity in Sudan next year.*
2 S: *Should we look at child malnourishment rates?*
3 U: *OK.*
4 S: *Should I compute a baseline estimate based on available data?*
5 U: *yes*
System identifies a reasoning engine (IMPACT) that can estimate childhood malnourishment, and uses existing knowledge about Sudan (population, food availability, etc.) to estimate the number of malnourished children.
6 S: *OK. The percentage of malnourished children is about 39%.*
7 U: *What if 2019 is an El Nino year?*
System builds a plan to estimate the difference in crop yields in Sudan in El Nino years by interpolating simulation results from 30 years of prior weather data.
8 S: *I can estimate this effect using the following plan. Is it OK?*
9 U: *Yes*
10 S: *OK. I'm running the plan …*
11 S: *OK. This graph shows the expected difference in production.*
12 S: *OK. The percentage of malnourished children increases by 5%.*
13 U: *What can we do to mitigate the increase?*
System creates and displays a causal graph based on the knowledge it has acquired from the declarations of the reasoning engines.
14 S: *Here is what I know about causal influences of malnourishment.*
15 U: *What would happen if we change the planting date?*
16 S: *Shall I run simulations varying the planting date?*
17 U: *Yes*
System creates a plan to run a series of simulations that vary the planting date parameter, again estimating yields for an El Nino year by interpolating over results obtained from simulations of the past 30 years. The plan then assembles the results and presents them to the user in a chart.
18 S: *The best scenario involves planting crops one week earlier than usual.*

Figure 2: A dialogue between a user (U) and the CWMS system (S).

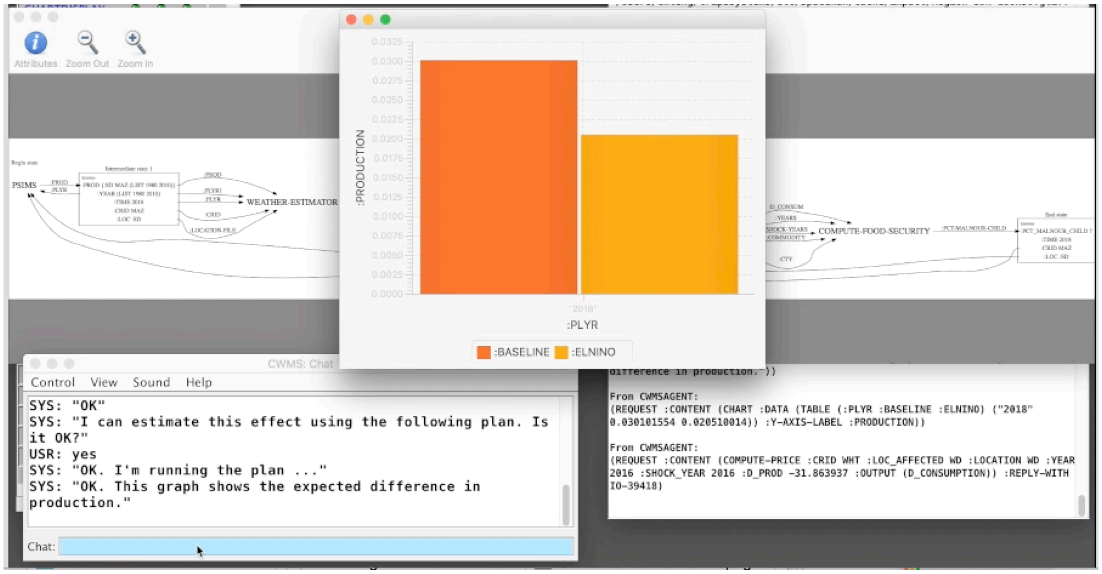

Figure 3: System presents a plan (in background) and then presents results of simulation of the model (i.e., the difference in expected crop production in an El Nino year)

changing the planting date (15). The system knows a problem solving strategy for investigating the effects of changing variables, and so asks whether it should construct a simulation experiment that estimates the crop yields for a range of different planting dates (16).

Once the user concurs (17), the system builds a plan for the experiment. It chooses a range of planting dates one week apart, starting one month before the typical planting date and ending one month after. For each of these dates, it estimates the expected yield using an interpolation plan similar to the one presented in (8). Once completed, it shows a plot of the results to the user and identifies the best option as planting one week earlier (18). While this is as far as we have space to discuss, the user could easily continue on, for instance asking for elaboration (e.g., OK, and how would that affect the malnourishment rates?), or exploring other options (e.g., What if we could increase the amount of fertilizer that was available?), or pursuing some new strategy for dealing with the problem, such as shipping more food aid to the region.

5 Conclusion

We have examined a range of possible forms of human-machine interaction in the context of world modeling problems and shown examples of how by harnessing the unique strengths of humans and intelligent agent-based machines, this approach has great promise for constructing systems far more complex and encompassing than previously possible. Furthermore, we are able to construct them in a timely manner that could allow their use for analyzing and planning responses to ongoing situations. The technical challenges to fully achieve this vision are substantial, but we have made a strong foundation and believe significant progress can be made in the next few years with a focused effort. More details on this approach can be found in Allen, Teng & Galescu (to appear).

6 Resources

Additional detail on the underlying technology can be found in Galescu et al. (2018). Much of the system is available online. Versions of the parser customized for specific domains can be tried at http://trips.ihmc.us/parser/. The code for the generic agent architecture including the parser (called cogent) is at https://github.com/wdebeaum/cogent.

Acknowledgements

This work has been supported in part by DARPA CwC and World Modelers programs (ARO contracts W911NF-15-1-0542, W911NF-17-1-0047 and W911NF-18-1-0464).

References

M. Ai-Chang, J. Bresina, L. Charest, A. Chase, J. C.-J. Hsu, A. Jonsson, B. Kanefsky, P. Morris, K. Rajan, J. Yglesias, B. G. Chafin, W C. Dias, P. F. Maldague, (2004) MAPGEN: Mixed-initiative planning and scheduling for the Mars Exploration Rover mission, IEEE Intell. Syst., vol. 19, no. 1, pp. 8-12.

Allen, J., de Beaumont, W., Galescu, L., Orfan, J., Swift, M. and Teng, C.M. (2013). Automatically deriving event ontologies for a commonsense knowledge base. In P*roceedings of the 10th International Conference for Computational Semantics (IWCS-2013)*, Potsdam, Germany.

Allen, J., de Beaumont, W., Galescu, L. and Teng, C.M. (2015). Complex Event Extraction using DRUM. In *Proceedings of the Workshop on Biomedical Natural Language Processing (BioNLP-2015),* pp.1-11.

Allen, J. and Teng, C. M. (2013). Becoming different: A language-driven formalism for commonsense knowledge. *International Symposium on Logical Formalizations of Commonsense Reasoning*.

Allen, J. F. and Teng, C. M. (2017). Broad coverage, domain-generic deep semantic parsing. In Proceedings of the *AAAI Spring Symposium on Computational Construction Grammar and Natural Language Understanding*.

Allen, J.F., Teng, C.M., and Gaelscu, L. (to appear) Dialogue as Collaborative Problem Solving: A Case Study, in *Advances in Cognitive Systems, 7.*

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

Ferguson, G. and J. Allen (1998). TRIPS: An Integrated Intelligent Problem-Solving Assistant. *National Conference on Artificial Intelligence (AAAI)*, Madison, WI, MIT Press.

Galescu, L., Teng, C. M., Allen, J. & Perera, I. (2018). COGENT: A generic dialogue system shell based on a collaborative problem solving model. *Proc. 19th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (pp. 400–409). Melbourne, Australia: ACL.

Gyori, B.M.; Bachman, J.A.; Subramanian, K.; Muhlich, J.L.; Galescu, L.; Sorger, P.K. (2017). From word models to executable models of signaling networks using automated assembly. *Molecular Systems Biology,* 13(11):954.

Hatfield-Dodds, S., Adams, P. D., Brinsmead, T. S., Bryan, B. A., Chiew, F. H. S., Finnigan, J. J., Graham, P. W., Grundy, M. J., Harwood, T., McCallum, R., McKellar, L. E., Newth, D., Nolan, M., Schandl, H., & Wonhas, A. (2015). In *Australian National Outlook 2015: Economic activity, resource use, environmental performance and living standards* (1970–2050). CSIRO, Canberra.

Headey, D. (2011) Rethinking the global food crisis: The role of trade shocks. *Food Policy*, 36(2):136-146.

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. and Ritchie, J.T. (2003). The DSSAT cropping system model. *European Journal of Agronomy*, 18(3), pp.235-265.

Holzworth, Dean P., Neil I. Huth, Peter G. deVoil, Eric J. Zurcher, Neville I. Herrmann, Greg McLean, Karine Chenu, et al. (2014) APSIM—Evolution towards a New Generation of Agricultural Systems Simulation. *Environmental Modelling & Software* 62 (December 2014): 327–350.

Marchand, P., Carr, J. A., Dell'Angelo, J., Fader, M., Gephart, J.A., Kummu, M., Magliocca, N.R., Porkka, M., Puma, M.J., Ratajczak, Z. and Rulli, M.C. (2016). Reserves and trade jointly determine exposure to food supply shocks. *Environmental Research Letters*, 11(9), p. 095009.

Robinson, S., Mason-D'Croz, D., Sulser, T., Islam, S., Robertson, R., Zhu, T., Gueneau, A., Pitois, G. and Rosegrant, M.W. (2015). The international model for policy analysis of agricultural commodities and trade (IMPACT): model description, version 3. *IFPRI Discussion Paper 1483*. Washington, DC: IFPRI.

Schewe, J., Otto, C. and Frieler, K. (2017). The role of storage dynamics in annual wheat prices. *Environmental Research Letters*, 12(5), p.054005.

White, Jeffrey & Hunt, L.A. & J. Boote, Kenneth & Jones, James & Koo, Jawoo & Kim, Soonho & Porter, Cheryl & W. Wilkens, Paul & Hoogenboom, Gerrit. (2013). Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and Electronics in Agriculture*. 96. 1-12.