DOE AI4ESP: **Land Surface Modeling 2.0**

for agricultural climate change impact assessments

James A. Franke[1,*], Beth A. Drewniak[2,*], Alexandre Renchon[2], Zhenong Jin[3], Vipin Kumar[4], Kaiyu Guan[5], Bin Peng[5], Jules F. Cacho[2], Leroy J. Walston[2], and Elisabeth J. Moyer[1]

[1] *University of Chicago, Department of the Geophysical Sciences*
[2] *Argonne National Laboratory, Environmental Science Division*
[3] *University of Minnesota, Department of Bioproducts and Biosystems Engineering*
[4] *University of Minnesota, Department of Computer Sciences*
[5] *University of Illinois at Urbana-Champaign, Dept. of Natural Resources and Env. Sciences*
[*] *Correspondence: jfranke@uchicago.edu | bbye@anl.gov*

Focus Area

This white paper addresses DOE AI4ESP focus area #2 by providing a sketch blueprint for a next-generation, hybrid AI/process-based global Land Surface Modeling (LSM) framework to improve projections of climate change impacts on the land surface system including agriculture and the hydrological cycle.

Science Challenge

Climate change impacts on agriculture are highly uncertain: -50% to +150% global production changes for major grains under high-end climate change [1], for example. The core challenge in land surface system predictability is the large gap between the scale at which the relevant biological processes act and the scale at which the risks need to be assessed. Conventional empirical and process-based modeling approaches are insufficient. A new multi-scale modeling paradigm which employs AI/ML to learn from new streams of remote sensing data and targeted 'gene-to-global' simulations [2] is needed.

Rationale

Agriculture is a major component of the Earth System. Over 70% of total freshwater withdrawals [3] and half of the Earth's habitable land (not including glaciers or deserts [4]) is agricultural. Water availability issues driven by climate variability are therefore almost invariably agricultural water shortages. Agriculture accounts for an even higher fraction of consumptive water use. Total freshwater withdrawals for thermoelectric generation can exceed that of agriculture in some energy-intense economies, but the consumptive fraction is very low (~2.5% [5]) making it hydrologically unimportant. Agricultural impacts on the hydrological cycle and climate are therefore substantial; irrigation in South Asia is changing the characteristics of the monsoon [6] and evapotranspiration (ET) from maize in the US Midwest substantially reduces maximum dry bulb summertime temperature in the region [7].

Agricultural climate risks are typically assessed using one of two broad categories of model: statistical (empirical) and so-called process-based models, both with strengths and weaknesses. On one hand, doubt can be cast on the generalizability of using the relationships between historical agronomic data residuals and historical weather to predict the impacts of future mean changes in climate. Estimated relationships, while possibly very accurate in out-of-historical-sample validation, may be learning incorrect weight and cannot account for changes in management and the effects of elevated CO_2. On the other hand, the relevant plant growth processes act at the scale of

DOI: https://doi.org/10.2172/1769734
DOE BER, AI4ESP 2021 #AI4ESP1045 (https://www.ai4esp.org)

plant DNA, and the governing equations are largely unknown. Heterogeneity in technology (e.g. cultivar genetics) and management decisions (e.g. growing seasons) are impossible to explicitly code into process-based simulations at the relevant resolution so projections from global process-based models remain very uncertain. A gap remains between the modeling approaches with little cross-pollination between research domains. AI/ML can be employed to bridge the research gaps between empirical and process based models and improve land surface system predictability.

Narrative

We intend to develop a next-generation LSM framework analogous to *Earth System Modeling 2.0: ClimateMachine.jl* [8]. The key scientific problem in climate modeling is the computational cost of running global simulations at the horizontal and vertical resolution needed to accurately represent boundary layer eddies. The proposed solution is to break open the standard climate model so that it can learn from satellite data and outputs from targeted high resolution convection simulations. The DOE is currently funding efforts in this vein at the Computational Clouds and Climate Lab at UCI and through other AI/ML applications in *Superparameterization* or *Ultraparaterization* of clouds and boundary layer dynamics [9]. Similarly, next-generation land surface models should be able to learn from two critical streams of data: remote sensing and targeted, high-resolution farm system modeling from 'gene-to-global' scales [2].

The first major area of promise in land surface modeling involves using satellite images to better understand the relationship between environment (soil and weather) and vegetative production [10]. Earth observation data is growing at ∼100TB/day [11], with satellite images the largest source. Satellite images are primarily being used in the agricultural modeling domain to cost-effectively improve yield data where little or none exists (e.g. to get more regression or process tuning targets [12]) and map land use areas more accurately [13]. Biophysical estimations, such as leaf area index, gross primary productivity, and ET, have recently become available over global cropland with improved resolutions and accuracy [14, 15, 16]. Early efforts of integrating these satellite measurements into regional or global process-based land surface models are promising, but still amount to parameter calibration in a manual set-up in most cases at the moment [17, 18].

The second area of active development is applications of AI/ML in agricultural models. Deep learning is being used in place of conventional statistical frameworks in standard empirical agricultural models, where the regression target is the (government) reported yield at some administrative level. These supervised methods have shown to outperform conventional statistical models and process based models at predicting out of sample historical yields [19, 20]. The volume of incoming data makes conventional assimilation techniques inadequate, deep learning must be applied to automatically learn the relevant patterns and features. Finally, agricultural models need to account for changing genetics and management in order to accurately predict future threats from climate change. The next-generation model framework must be able to learn from targeted, or fine-scale simulations where observations are unavailable (or not possible) and to inform genetic development and management interventions.

We propose building a hybrid AI/process-based model framework from the ground up to synthesize these three research themes. A core component of the approach will be leveraging AI/ML to learn solutions to the PDEs representing photosynthesis and plant growth under heterogeneous conditions using the rapidly expanding wealth of satellite data and existing process knowledge. Misfits in conventional empirical models are likely due to coarse temporal aggregation of weather regressors and targets. Typically, growing season temperatures are aggregated up to the growing season, with a variety of engineered features intended to represent heat and water stress. A key innovation of our data-driven approach involves using AI/ML to predict plant growth and ET across the entire season –not just the final yield, increasing the number of targets by 2 orders of magnitude.

AI/ML can be employed to improve prediction power in land surface models in at least five concrete ways [11]. First, classification deep learning algorithms can be used to generate new high-resolution input datasets, especially of management conditions such as land-use type, sowing and harvest dates, irrigation application, tillage practices, and cropping systems. Second, AI/ML can be used to improve existing parameterizations within the model such as applying data-driven spatially explicit ET parameterizations. Applicable learning algorithms for this process will include OLS gradient descent, Bayesian inversion, and ensemble Kalman methods [8]. Third, clustering may be used to reduce the computational burden of high resolution simulations through spatial aggregations of grid cells by homogeneity criteria. Fourth, AI/ML can be used to study existing process-based model simulation outputs by analyzing the residuals on different outputs like ET. Finally, entire sub-modules in the process-based model (e.g. the plant phenological development) can be replaced with AI/ML wholesale.

Relevant processes that need additional research development include: diffuse radiation (to study the effect of potential geoengineering [21], pests and diseases [22], temporal soil moisture dynamics [23], heavy precipitation and water logging [24, 25], changing irrigation technologies, surface and groundwater irrigation availability, and non-annual managed plant types. Another key feature of next-generation LSM needs to be ESM coupling. Current representations of land surface processes in ESMs are crude, with most GCMs not explicitly representing managed vegetation [26]. Our new framework will tightly couple with the E3SM model to provide boundary conditions and study the critical feedbacks between the land surface and the climate.

Critical data challenges to improving Earth System predictability by this approach include limited *ground truth* data outside of North America and Europe, especially that of high-quality crop cuts and management choices. While satellite data can fill this gap to some extent, these data are limited by satellite return rates and cloud cover in the tropics. Additionally, the crucial impacts of CO_2 on plant growth are not easily derived from observational datasets and may seriously hamper any data-driven approaches. Expanded networks of field-level data collection under uniform standards[27], advanced synthesis of multiple satellite data products, and innovative use of Free-Air Carbon Enrichment (FACE) data [28] are requried to address these challenges.

Finally, we aim to reshape the Earth System Model (ESM) development paradigm in general. Most ESMs are written in legacy languages (e.g. FORTRAN) and custom HPC environments that are unavailable in practice to most researchers and cannot readily capture the power of new computational architectures (e.g. GPUs or TPUs). Next-generation models need to be built from the ground up to be able to capitalize on new high performance compute environments and to efficiently incorporate the AI/ML features outlined above. While climate model skill at representing certain phenomena has improved over the last decade [e.g. 29], overall sensitivity to CO_2 forcing has seen little consensus improvement between CMIP5 and CMIP6 [30]. ESMs become harder to improve through model intercomparison as they increase in complexity, because the root causes become harder to diagnose. To accelerate model development and improve Earth System predictability, the next generation should not consist of many models being compared to one another, but a single, modular model framework which is improved by everyone. A next-generation model should therefore be fully open source code and utilize a community-focused developmental strategy following the principles of *FLOSS* and *Open Science*.

Suggested Partners/Experts

1. Christoph Müller, PIK - process-based crop modeling
2. Senthold Asseng, Technical University of Munich - process-based crop modeling
3. Vikram Adve and Jingrui He, AIFARMS at UIUC - AI/ML and agriculture
4. Supratik Guha, UChicago, ANL, and AIFARMS
5. Lawrence Livermore - E3SM and specifically DOE land surface modeling efforts
6. AgMIP: the Agricultural Model Intercomparison and improvement Project led by Cynthia Rosenzweig
7. ISIMIP / PROCLIAS: International unified frameworks for process-based climate impact modeling led by Katja Frieler
8. Climate Modeling Alliance (Caltech): next-gen earth system model framework led by Tapio Schneider

References

[1] A. J. Challinor, J. Watson, D. B. Lobell, S. M. Howden, D. R. Smith, and N. Chhetri. A meta-analysis of crop yield under climate change and adaptation. *Nature Climate Change*, 4(4):287–291, April 2014.

[2] Bin Peng, Kaiyu Guan, Jinyun Tang, Elizabeth A. Ainsworth, Senthold Asseng, Carl J. Bernacchi, Mark Cooper, Evan H. Delucia, Joshua W. Elliott, Frank Ewert, Robert F. Grant, David I Gustafson, Graeme L. Hammer, Zhenong Jin, James W. Jones, Hyungsuk Kimm, David M. Lawrence, Yan Li, Danica L. Lombardozzi, Amy Marshall-Colon, Carlos D. Messina, Donald R. Ort, James C. Schnable, C. Eduardo Vallejos, Alex Wu, Xinyou Yin, and Wang Zhou. Towards a multiscale crop modelling framework for climate change adaptation assessment. *Nature Plants*, 6(4):338–348, April 2020.

[3] FAO. AQUASTAT: FAO's information system on water and agriculture, 1999.

[4] Erle C. Ellis, Kees Klein Goldewijk, Stefan Siebert, Deborah Lightman, and Navin Ramankutty. Anthropogenic transformation of the biomes, 1700 to 2000: Anthropogenic transformation of the biomes. *Global Ecology and Biogeography*, pages no–no, June 2010.

[5] P Torcellini, N Long, and R Judkoff. Consumptive Water Use for U.S. Power Production. Technical Report NREL/TP-550-33905, 15005918, National Renewable Energy Laboratory, December 2003.

[6] Benjamin I. Cook, Sonali Shukla McDermid, Michael J. Puma, A. Park Williams, Richard Seager, Maxwell Kelley, Larissa Nazarenko, and Igor Aleinov. Divergent Regional Climate Consequences of Maintaining Current Irrigation Rates in the 21st Century. *Journal of Geophysical Research: Atmospheres*, 125(14), July 2020.

[7] Ethan E. Butler, Nathaniel D. Mueller, and Peter Huybers. Peculiarly pleasant weather for US maize. *Proceedings of the National Academy of Sciences*, 115(47):11935–11940, November 2018.

[8] Tapio Schneider, Shiwei Lan, Andrew Stuart, and João Teixeira. Earth System Modeling 2.0: A Blueprint for Models That Learn From Observations and Targeted High-Resolution Simulations. *Geophysical Research Letters*, 44(24), December 2017.

[9] P. Gentine, M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis. Could Machine Learning Break the Convection Parameterization Deadlock? *Geophysical Research Letters*, 45(11):5742–5751, June 2018.

[10] Elinor Benami and et al. Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nature Reviews Earth and Enviroment*, page 20, 2021.

[11] Markus Reichstein, Gustau Camps-Valls, Bjorn Stevens, Martin Jung, Joachim Denzler, Nuno Carvalhais, and Prabhat. Deep learning and process understanding for data-driven Earth system science. *Nature*, 566(7743):195–204, February 2019.

[12] Zhenong Jin, George Azzari, Calum You, Stefania Di Tommaso, Stephen Aston, Marshall Burke, and David B. Lobell. Smallholder maize area and yield mapping at national scales with Google Earth Engine. *Remote Sensing of Environment*, 228:115–128, July 2019.

[13] Yaping Cai, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sensing of Environment*, 210:35–47, June 2018.

[14] Hyungsuk Kimm, Kaiyu Guan, Chongya Jiang, Bin Peng, Laura F. Gentry, Scott C. Wilkin, Sibo Wang, Yaping Cai, Carl J. Bernacchi, Jian Peng, and Yunan Luo. Deriving high-spatiotemporal-resolution leaf area index for agroecosystems in the U.S. Corn Belt using Planet Labs CubeSat and STAIR fusion data. *Remote Sensing of Environment*, 239:111615, March 2020.

[15] Chongya Jiang, Kaiyu Guan, Ming Pan, Youngryel Ryu, Bin Peng, and Sibo Wang. BESS-STAIR: a framework to estimate daily, 30 m, and all-weather crop evapotranspiration using multi-source satellite data for the US Corn Belt. *Hydrology and Earth System Sciences*, 24(3):1251–1273, March 2020.

[16] Chongya Jiang, Kaiyu Guan, Genghong Wu, Bin Peng, and Sheng Wang. A daily, 250 m, and real-time gross primary productivity product(2000–present) covering the Contiguous United States. preprint, Biosphere – Biogeosciences, May 2020.

[17] Jianxi Huang, Jose L. Gómez-Dans, Hai Huang, Hongyuan Ma, Qingling Wu, Philip E. Lewis, Shunlin Liang, Zhongxin Chen, Jing-Hao Xue, Yantong Wu, Feng Zhao, Jing Wang, and Xianhong Xie. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agricultural and Forest Meteorology*, 276-277:107609, October 2019.

[18] Xiuliang Jin, Lalit Kumar, Zhenhai Li, Haikuan Feng, Xingang Xu, Guijun Yang, and Jihua Wang. A review of data assimilation of remote sensing and crop models. *European Journal of Agronomy*, 92:141–152, January 2018.

[19] Andrew Crane-Droesch. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13(11):114003, October 2018.

[20] Saeed Khaki, Lizhi Wang, and Sotirios V. Archontoulis. A CNN-RNN Framework for Crop Yield Prediction. *Frontiers in Plant Science*, 10:1750, January 2020.

[21] Jonas Jägermeyr, Alan Robock, Joshua Elliott, Christoph Müller, Lili Xia, Nikolay Khabarov, Christian Folberth, Erwin Schmid, Wenfeng Liu, Florian Zabel, Sam S. Rabin, Michael J. Puma, Alison Heslin, James Franke, Ian Foster, Senthold Asseng, Charles G. Bardeen,

Owen B. Toon, and Cynthia Rosenzweig. A regional nuclear conflict would compromise global food security. *Proceedings of the National Academy of Sciences*, 117(13):7071–7081, March 2020.

[22] Curtis A. Deutsch, Joshua J. Tewksbury, Michelle Tigchelaar, David S. Battisti, Scott C. Merrill, Raymond B. Huey, and Rosamond L. Naylor. Increase in crop losses to insect pests in a warming climate. *Science*, 361(6405):916–919, August 2018.

[23] Ariel Ortiz-Bobea, Haoying Wang, Carlos M Carrillo, and Toby R Ault. Unpacking the climatic drivers of US agricultural yields. *Environmental Research Letters*, 14(6):064003, May 2019.

[24] Yan Li, Kaiyu Guan, Gary D. Schnitkey, Evan DeLucia, and Bin Peng. Excessive rainfall leads to maize yield loss of a comparable magnitude to extreme drought in the United States. *Global Change Biology*, 25(7):2325–2337, 2019.

[25] Corey Lesk, Ethan Coffel, and Radley Horton. Net benefits to US soy and maize yields from intensifying hourly rainfall. *Nature Climate Change*, August 2020.

[26] S. S. McDermid, L. O. Mearns, and A. C. Ruane. Representing agriculture in Earth System Models: Approaches and priorities for development: AGRICULTURE IN ESMS. *Journal of Advances in Modeling Earth Systems*, 9(5):2230–2265, September 2017.

[27] R.P. Rötter, M. Appiah, E. Fichtler, K.C. Kersebaum, M. Trnka, and M.P Hoffmann. Linking modelling and experimentation to better capture crop impacts of agroclimatic extremes—A review. *Field Crops Research*, 221:142–156, May 2018.

[28] Andrea Toreti, Delphine Deryng, Francesco N. Tubiello, Christoph Müller, Bruce A. Kimball, Gerald Moser, Kenneth Boote, Senthold Asseng, Thomas A. M. Pugh, Eline Vanuytrecht, Håkan Pleijel, Heidi Webber, Jean-Louis Durand, Frank Dentener, Andrej Ceglar, Xuhui Wang, Franz Badeck, Remi Lecerf, Gerard W. Wall, Maurits van den Berg, Petra Hoegy, Raul Lopez-Lozano, Matteo Zampieri, Stefano Galmarini, Garry J. O'Leary, Remy Manderscheid, Erik Mencos Contreras, and Cynthia Rosenzweig. Narrowing uncertainties in the effects of elevated CO2 on crops. *Nature Food*, 1(12):775–782, December 2020.

[29] T. J. Bracegirdle, C. R. Holmes, J. S. Hosking, G. J. Marshall, M. Osman, M. Patterson, and T. Rackow. Improvements in Circumpolar Southern Hemisphere Extratropical Atmospheric Circulation in CMIP6 Compared to CMIP5. *Earth and Space Science*, 7(6), June 2020.

[30] Mark D. Zelinka, Timothy A. Myers, Daniel T. McCoy, Stephen Po-Chedley, Peter M. Caldwell, Paulo Ceppi, Stephen A. Klein, and Karl E. Taylor. Causes of Higher Climate Sensitivity in CMIP6 Models. *Geophysical Research Letters*, 47(1), January 2020.