Improved Understanding of Coupled Water and Carbon Cycle Processes through Machine Learning Approaches

Authors

Debjani Sihi (Emory University, debjani.sihi@emory.edu), Kanad Basu (University of Texas at Dallas, Kanad.Basu@UTDallas.edu*), Debjani Singh (Oak Ridge National Laboratory, debd@ornl.gov)*

Focal Area(s)

This white paper addresses how explainable machine learning (ML) algorithms can improve insights gained from complex data (Focal Area 3). We will also address how ML approaches related to sensor compression and low energy AI hardware can be used for efficient data acquisition (Focal Area 1).

Science Challenge

We will focus on the coupled water cycle and carbon cycle processes in terrestrial ecosystems and terrestrial-aquatic interfaces. Thus, our approaches will be centered around several data-model integration challenges indicated in EESD strategic plans for two science focus areas: Terrestrial Ecosystem Science and Hydrobiogeochemistry. We can use AI to correct systematic model errors due to biases in either data or model structure/processes. Error patterns in model predictions usually vary by region, model, season etc [4]. But there are systematic patterns in them. *e.g.*, soil moisture tends to be overestimated in the arid western continental United States underestimated in wetter eastern USA [11]; some land surface models tend to underestimate moisture in wet seasons and overestimate in dry seasons. Moreover, the consequences of extreme events (*e.g.*, drought, extreme flooding, storm surges associated with tropical storms, hurricanes) on carbon cycle processes are not well represented in ecosystem and Earth system models [5]. Redox-sensitive processes (*e.g.*, rapid oxidation/reduction of iron and other redox-sensitive elements in soil microsites subjected to fluctuated hydrology) in terrestrial-aquatic interfaces further challenge model predictions of hot-spots (and hot-moments) due to poor understanding of underlying mechanisms. Thus, AI/ML approaches can be used to learn patterns in the data and model errors and use them to build model equations and correct process-based model errors.

Rationale

Use AI to correct systematic model errors due to biases in the data: Complex, non-linear processes are difficult to unravel and represent in process-based Earth System Models like E3SM. AI/ML can help understand the predictor-response relationship across broad spatial and temporal scales for coupled water and carbon cycle processes [3-6, 9-10]. However, existing AI/ML applications in Earth and Environmental Sciences rely on a *"black box"* approach, which hinder our capability to glean insight from complex ecological data. To that end, explainable ML algorithms will help evaluate model intercomparison projects (*e.g.*, ILAMB) and model benchmarking. We also expect that our explainable ML approach can be implemented into the MODEX (Model-Experimental Coupling) framework for improved prediction of extreme events (*e.g.*, droughts, extreme precipitation, and storm surge from hurricanes). Since we are focusing on coupled water and carbon cycle processes, this work also contributes to the recently launched theme year, Year of Water, by the AmeriFlux community.

Scaling and synthesis of multiple data types across multiple projects: ML can be used to scale plot-scale measurements to regional- and global-scale such that spatial heterogeneity can be captured [11]. To that end, synthesis of different data types (csv, netCDF, HDF5) is needed for improved Earth system predictability. For instance, synthesis of knowledge through the application of explainable ML approaches across DOE-funded flagship projects (Next Generation Ecosystem Experiments, AmeriFlux, SPRUCE, and COMPASS) and NASA-funded projects (e.g., ABoVE, AirMOSS, SMAP, GRACE) are crucial for better understanding of water and carbon cycle processes in critical ecosystems.

DOI: https://doi.org/10.2172/1769721
DOE BER, AI4ESP 2021 #AI4ESP1122 (https://www.ai4esp.org)

Improved Understanding of Coupled Water and Carbon Cycle Processes through Machine Learning Approaches

Use AI for efficient collection and storage of high-frequency sensor data: The growing volume of sensor data to study water cycle (*e.g.*, soil moisture, PPT, VPD, RH, ET) and carbon cycle (*e.g.*, eddy-covariance data on NEE, and partitioned flux components like GPP, Ecosystem respiration, automated chamber data for soil respiration) processes, challenge our capability to utilize and integrate big data in existing ESMs. Multiple sources of sensor data are available for the USA and worldwide. Often these sensor data are collected at high temporal frequency (half-hourly) and at multiple spatial resolutions (plot-scale measurements for soil moisture sensors installed at different soil depths in eddy covariance tower sites from AmeriFlux/FluxNet network vs landscape-scale measurements taken from air-borne platforms like drones, flights for LiDAR data, and satellite). High-frequency sensor data also comes with a high cost in terms of storage space and resource/energy need. HardCompress or low energy AI hardware (a novel hardware-based low power compression solution) and energy savings by different sensor compression technologies can be useful in addressing these challenges related to automated collections of sensor data ranging from field-based soil moisture sensors to satellite-based images.

Narrative

Explainable AI for improved Earth system predictability: While there is prior work in explainable machine learning in computer vision and related domains, utilizing explainable machine learning for ecosystem and Earth system modeling is not common. Figure 1 shows an overview of our proposed method that consists of four major steps: Model Training, Perturbation, Non-linear regression, and Interpretation [7, 8]. The last step is the key to perform *outcome interpretation using explainable ML approaches, which sets our algorithm apart from the traditional "black-box" approach*. The top features, ranked by the size of coefficients will provide users the crucial timing information of coupling and decoupling of water and carbon cycles during extreme events. Moreover, we can check the clock cycle distribution of these top features; it will provide us with extra information about the coupled water-carbon cycle processes. For example, if we observe adjacent clusters of top features, then the time slot within which they reside shall provide coarse time information about when coupling and decoupling of water and carbon cycles happened. If the features related to water and carbon cycle processes reside within the same timeframe, water and carbon cycles will be considered as coupled; otherwise, the two processes will be considered decoupled. Another useful scenario is that clock cycle numbers are periodically separated. In this case, the coupling of water-carbon cycles is likely to repeat their sneak operation periodically. The explanation given by our model would be able to illustrate which trace values are most likely relevant to the critical behavior, thereby aiding coupled water-carbon fluxes.

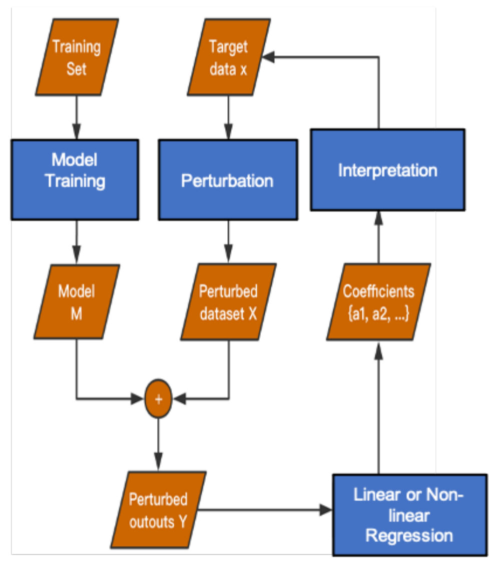

Figure 1: Proposed explainable ML-process for improved understanding of coupled water and carbon fluxes

Improved Understanding of Coupled Water and Carbon Cycle Processes through Machine Learning Approaches

Sensor Compression: To handle efficient data collection, we propose a novel sensor compression framework, a model- and platform-agnostic solution that involves performing subsampling and supersampling of sensor images using two sampling algorithms, namely Nearest Neighbor (NN) interpolation and Bilinear (BL) interpolation [1] (Fig. 2). Subsampling involves reducing the image dimension by either deleting entire pixel rows and columns (NN) or replacing a 2x2 pixel neighborhood with the average value (BL). Our solution works because of two main reasons. First, the images being captured by the modern sensors are much higher resolution than required for computer vision applications. Second, due to the intrinsic error resilience characteristic of DNNs, the compression of sensor images results in negligible performance degradation. The advantages of our framework are multifold – substantial energy savings in the sensor subsystem, with minimal degradation in model performance and additional energy savings in the communication subsystem and memory subsystems by virtue of the reduced data storage requirement. This sensor compression will be very helpful to acquisition and synthesis of high-resolution images collected by air-borne platforms and satellites in the coming few decades as they reflect ecosystem states of water and carbon on a high spatio-temporal frequency.

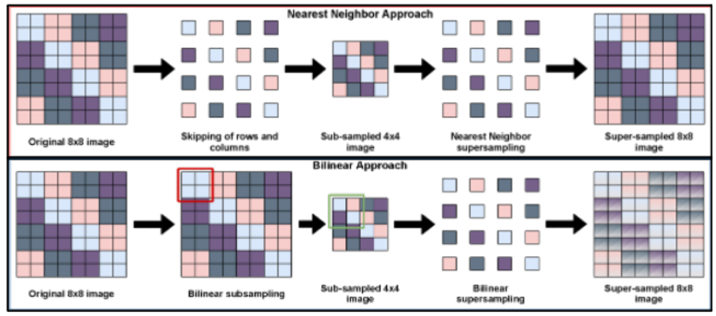

Figure 2: Proposed Sensor Compression Methodology

Low Energy AI hardware: Traditional CPU- and GPU-based deep learning implementations incur high overhead in terms of latency, computation, and power. This is especially true for edge-based sensor devices used in Earth system science research. Towards this end, we propose a novel compression strategy pertaining to commercial DNN accelerators, called HardCompress [2] (Fig. 3). The three-step approach involves hardware-based post-quantization trimming of weights, followed by dictionary-based compression of the weights and subsequent decompression by a low-power hardware engine during inference in the accelerator. Owing to their large size, the trained parameters of DNNs are stored in the larger off-chip memory (Dynamic RAM or DRAM) instead of the smaller on-chip memory (Static RAM or SRAM). Fetching these parameters from DRAM is 128x more energy intensive. The proposed technique will address this problem by facilitating the storage of DNN models in SRAM. Our initial results on popular image-based DNNs demonstrate that HardCompress can reduce energy up to 99%. Hence, we are confident that our HardCompress approach will be revolutionary for acquisition of any kind of data of file formats (*e.g.*, csv, txt, netCDF, and HDF5) as big handling data is one of the biggest challenges in the era of Earth and environmental system sciences.

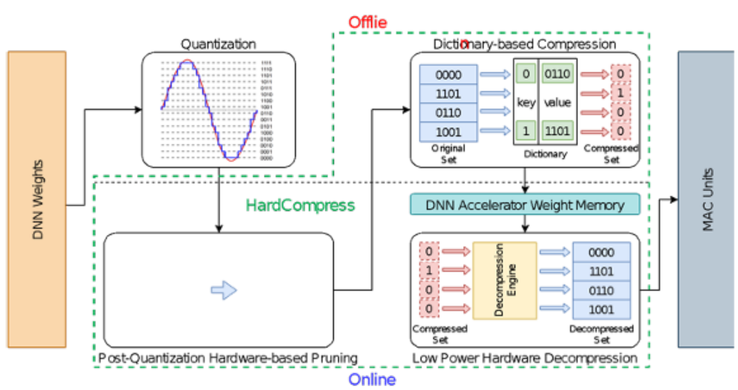

Figure 3: Proposed HardCompress Methodology

Improved Understanding of Coupled Water and Carbon Cycle Processes through Machine Learning Approaches

References

[1] Arunachalam, A., Kundu, S. and Basu K., 2021. SeNNse: Energy Efficient Sensor Approximations for DNN Inference at the Edge. In 2021 IEEE International Conference on Artificial Intelligence Circuits and Systems.

[2] Arunachalam, A., Kundu, S., Raha, A., Banerjee, S, Natarajan. S. Basu K., 2021. HardCompress: A Novel Hardware-based Low Power Compression Scheme for DNN Accelerators. In 2021 IEEE International Symposium on Quality Electronic Design.

[3] Clark, M.P., Fan, Y., Lawrence, D.M., Adam, J.C., Bolster, D., Gochis, D.J., Hooper, R.P., Kumar, M., Leung, L.R., Mackay, D.S. and Maxwell, R.M., 2015. Improving the representation of hydrologic processes in Earth System Models. Water Resources Research, 51, 5929-5956. https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR017096

[4] Fatichi, S., Vivoni, E.R., Ogden, F.L., Ivanov, V.Y., Mirus, B., Gochis, D., Downer, C.W., Camporese, M., Davison, J.H., Ebel, B. and Jones, N., 2016. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology, 537, 45-60.

[5] Gentine, P., Green, J.K., Guérin, M., Humphrey, V., Seneviratne, S.I., Zhang, Y. and Zhou, S., 2019. Coupling between the terrestrial carbon and water cycles—a review. Environmental Research Letters, 14, 083003.

[6] Hardin, J., Burrows, S., Kotomarthi, R., Comstock, J., Shrivastava, M., Riemer, N., Lier-Walqui, M. v., Theisen, A. and Kumar, J., 2020. A Summary of the Machine Learning and Statistical Methods for Observations and Modeling Constraint Workshop. Available at https://drive.google.com/file/d/17CK2eWmjlUxQd2M7L6atnKgxht4jYAgq/view

[7] Kuruvila, A. P., Meng, X., Kundu, S. and Basu K., 2021. Explainable Machine Learning for Intrusion Detection via Hardware Performance Counters. In 2021 IEEE Design Automation Conference.

[8] Pan, Z., Sheldon, J. and Mishra, P., 2020. Hardware-Assisted Malware Detection using Explainable Machine Learning. In 2020 IEEE International Conference on Computer Design.

[9] Shen, C., 2018. A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54, 8558-8593. https://agupubs.onlinelibrary.wiley.com/doi/pdfdirect/10.1029/2018WR022643

[10] Stevens, R., Taylor, V., Nichols, J., Maccabe, A. B., Yelick, K. and Brown, D., 2019. AI for Science, Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science. Available at https://www.anl.gov/ai-for-science-report

[11] Yuan, S. and Quiring, S.M., 2017. Evaluation of soil moisture in CMIP5 simulations over the contiguous United States using in situ and satellite observations. Hydrology and Earth System Sciences, 21, 2203-2218.