Title: Deep learning techniques to disentangle water use efficiency, climate change, and carbon sequestration across ecosystem scales

1. Authors/Affiliations: Jennifer A. Holm, Research Scientist, LBNL
Trevor Keenan, Assistant Professor, University of California Berkeley [Not Confirmed]
Daniel Ricciuto, Research Staff, ORNL
Vincent Emanuele, Founder & CTO, Embedded Intelligence

2. Focal Areas: We plan to use machine learning (ML) to disentangle the impact of climate change on plant water use efficiency (WUE) across the leaf level, structured canopy, and community ecosystem scales. Due to the complex behavior of modeling plant WUE's connection to carbon storage (and hidden co-varying interactions) we suggest that physics-based deep learning must be utilized. The second topic we propose is, based on a better understanding of WUE response to climate change, there is a need to predict how WUE will limit or enhance the recycling of water from transpiration by vegetation back to the land surface worldwide, and thus impact water storage capacity (i.e., natural and reservoirs), using machine learning optimization techniques such as genetic algorithms.

3. Science Challenge: In a world with an increasingly changing climate, a critical process that links terrestrial carbon and water cycles is water use efficiency (WUE), defined as the ratio of carbon dioxide (CO_2) assimilation, or C gain, to water loss and is essential for understanding ecosystem response to climate change and droughts (Baldocchi 1994, Beer et al. 2009, Keenan et al. 2013). Multiple climate conditions impact plant water use, and with many complex interactions and dependencies across plant physiology processes, community competition, and environmental conditions. Thus the full, global WUE 'picture' is very difficult to model, even with state-of-the-art process-based models.

4. Introduction and Rationale: Droughts associated with heat waves are becoming more intense, longer, and more extreme (Dai, 2013; Trenberth et al., 2014; Williams et al., 2020). How vegetation responses to drought influences global water resources, productivity, and land-atmosphere coupling (Bart et al., 2016; Green et al., 2017; Liu et al., 2020). The most direct effect of drought on vegetation behavior is its influence on stomatal conductance, which in turn affects evapotranspiration (ET) and photosynthesis. In a world with increasingly changing climate, a critical process that links terrestrial carbon (C) and water cycles is WUE. Ecosystem WUE is predicted to decrease under climate warming (De Boeck et al. 2006, Bell et al., 2010), yet increase with rising CO_2 concentrations due to partial closure of stomata and maintaining constant CO_2 concentrations inside the leaf (Keenan et al. 2013) as seen over the past two decades.

The modeling of the ecosystem C cycle response to extreme droughts is particularly challenging (Zscheischler et. al., 2014b; Piao et. al., 2019). Drought- induced ecosystem changes can feed back to the climate system, amplifying or modulating drought impacts (Anderegg et. al., 2019; Zhou et al., 2019). Another point of concern is that 13 terrestrial biosphere models did not predict long-term trends in ecosystem scale WUE consistent with observations (Keenan et. al., 2013), and underestimated WUE by not matching the observed trend in increasing annual C uptake. Masri et. al., (2019) reported that in response to climate impacts, discrepancies existed between 10 ecosystem models (i.e., decreasing WUE) and up-scaled flux observations (i.e., constant WUE). Overall, many challenges and unknowns still exist between models and observations due to the complex behavior of WUE.

In general there are four major climate change factors that will affect plant water use, including: increasing CO_2 concentrations, increasing temperatures, more variable precipitation, and

variations in humidity; but with many complex interactions and dependencies across plant physiology processes, community competition, and environmental conditions. Rising global temperatures, which lead to increased surface evaporation will increase atmospheric water demand by plants and potential for limitations in soil water availability, because of the increased variation in precipitation during the growing season and even more so in soils with limited water holding capacity (Xiao et al. 2013, Collins et al. 2013). Increasing drought due to projected climate change (Fu and Feng, 2014) has been shown to substantially reduce global net primary productivity (Zhao and Running, 2009).

4.1 Major challenge 1 – Complex behavior of water use efficiency: We are only beginning to scratch the surface with how WUE responds to climate change at different scales (i.e., leaf, canopy, and ecosystem level), all which have been found to have varying results compared to each other (Niu et. al., 2011). At the leaf level, the instantaneous WUE is calculated as the net photosynthetic rate (A_n) divided by transpiration rate. Alternatively, the intrinsic WUE, which is calculated as A_n divided by stomatal conductance g_s. These variables are then regulated by the available energy controls on the leaf, vapor pressure deficit (VPD), and aerodynamic exchange. While at the canopy level, the processes involving energy exchange at the soil surface and the plant canopy and water loss is a combination of evaporation from the soil surface and transpiration from the plant canopy. An ecosystem-level estimation of WUE requires more assumptions such as needing to incorporate the ratio of atmospheric evaporative demand to ecosystem ET, gross ecosystem photosynthesis (GEP) typically estimated from flux measurements, and the canopy integrated concentration of intercellular CO_2 to ambient CO_2 (Keenan et. al., 2013).

An array of environmental conditions can impact WUE at one or more scales (i.e., leaf, canopy, and ecosystem level). Leaf level driving factors include interacting effects of temperature (e.g., heat stress) and solar irradiance, with the exchange of CO_2 and water. These abiotic factors then impact net photosynthetic rate, stomatal conductance, transpiration rate, and volatile emissions and defense mechanisms against cellular damage. While at the canopy and ecosystem level driving factors include the complex interactions between shifting canopy structure, and surface and soil energy exchange.

WUE also varies significantly depending on a plant's photosynthetic pathway, with plants being classified as using C_3 photosynthesis, C_4 photosynthesis, and Crassulacean acid metabolism (CAM). Compared to C_3 species, C_4 species have higher optimum temperatures for photosynthesis and higher WUE (Taylor et al., 2014; Ripley et al., 2007). Therefore, one can expect C_4 species to be more likely to survive and even replace C_3 species during drought. However, recent work has found evidence for new behaviors of shifting community dynamics of C_3 and C_4 grasses during droughts. Knapp et al. (2020) revealed that during the Great Dust Bowl, a mega-drought event, C_3 grasses replaced C_4 species across large regions of the US Great Plains, a *paradoxical response* that was also confirmed by manipulative extreme drought experiments. Investigating this opposite switch in WUE is paramount because most crops are C_4 species, and might be a contributor to future food scarcity.

This white paper argues that due to the complex nature of these 'ecosystem sustaining processes' and flux exchanges, the automation of ML through supervised learning of training datasets, learning hidden data features, and using optimizers is required to achieve the best insight.

4.2 Machine Learning Techniques: Under traditional methods to try to *fully* address complex ecological questions raised here, there is a high likelihood of high-dimensional data (from leaf level measurements to flux towers) and leads to overfitting where the model captures both real and random effects. Datasets with high-dimensions can cause an algorithm to model the random noise in the training data, rather than the intended outputs (i.e., overfitting) and lead to misinterpreted results. Therefore, we need to go beyond traditional statistical approaches and utilize machine learning. We propose a combination of physics informed deep learning (a new emerging technique), generative

adversarial networks (GANs), and genetic algorithms will be key to predictive modeling of complex interactions between rising temperatures, CO_2, and WUE at scales that are difficult to measure.

Physics-informed neural networks (NN) – that are trained to solve supervised learning tasks, but also while respecting any given laws of physics described by general nonlinear partial differential equations (Raissi et. al., 2019) are ideal for these research problems. The resulting NNs form a new class of data-efficient universal function approximators that naturally encode any underlying physical laws as prior information. This approach directly tackles nonlinear problems without the need for committing to any prior assumptions, linearization, or local time-stepping, by using automatic differentiation (Baydin et. al., 2017). A technique that uses multiple NNs, but differentiates them with respect to their input coordinates and model parameters to obtain physics informed NNs. Such NNs are constrained to respect any symmetry, invariance, or conservation principles originating from the physical laws that govern the observed data. This approach is designed to inform the loss function in ML, because the governing equations in our case will be based on the WUE and climate in question. The loss function is trained by minimizing the mean square error loss across multiple NNs; one that denotes the initial and boundary data, the other that specifies the physical function that we are trying to solve (typically using Burgers' equation).

4.3 Research Question 1: How, when, and where will WUE be limited or enhanced by increasing CO_2, temperature, and changing precipitation conditions with a focus on droughts, and as a result ET, carbon sequestration, and soil water retention?

4.4 Major challenge 2 – Complex behavior of water recycling for Earth sustainability: Forests receive a portion of their water supply through naturally occurring recycled rain, or acting as conduits of their own 'water pumps' (Figure 1 from Aragao 2012). This occurs when water is pumped by the trees from the soil and into the atmosphere, through ET dynamics. For example 25-56% of the rainfall in tropical forests can be recycled for reuse within the ecosystem (Aragao 2012, Eltahir and Bras 1996). However, deforestation has weakened this critical ecosystem service, through increased runoff and sensible heat, and reduced ET and atmospheric moisture (Spracklen et al. 2012; Shukla et. al. 1990). Maintaining water recycling in intact forests is critical, because they remove ~30% of anthropogenic CO_2 emissions through photosynthesis and return ~40% of incident precipitation back to the atmosphere via transpiration. The trade-off between photosynthesis and transpiration through WUE under droughts, has profound effects on climate Earth's water sustainability, leading to enhanced or weakened water recycling and drought severity. Changes to the rainfall-recycling ratio can be traced by evaporated water molecules in ESMs, which is promising for monitoring implications of land cover changes on water storage for human use (e. g., reservoirs, groundwater, snow water equivalent).

4.5 Research Question 2: (a) Given that we have a better understanding of WUE response to

climate change, how will WUE limit or enhance water recycling worldwide, and thus impacts to water storage (i.e., natural and reservoirs)? **(b)** What are the consequences of reduced local water recycling processes on vegetation hydraulic-stress mortality and water transport (i.e., irrigation, runoff, and atmospheric transport)? **(c)** Use genetic algorithms (see Section 5.1) to determine what is the threshold of deforestation in which terrestrial water recycling is reduced below a limit to sustain forest recovery to pre-disturbance conditions, in multiple regions of the Earth?
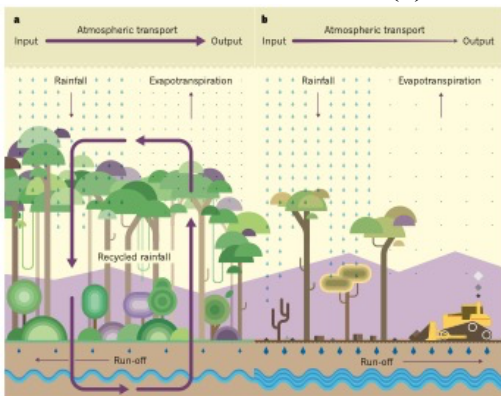
5. Methodology and Approach: We propose to integrate deep learning techniques with process-based plant physiology modeling and ESM climate capabilities. Included in this approach is large-scale global data synthesis to detect patterns, relationships, and dependencies between a myriad of variables influencing WUE and impacts to water recycling, water storage, carbon sequestration under changing climates.

5.1   Generative adversarial networks (GANs) & genetic algorithms: We propose using GANs based on their combination of generative ML which uses fully probabilistic models, or distributions, of observed data to "generate" as realistic solutions as possible to deceive the discriminator, and adversarial ML which applies a discriminator to try and "classify" fake solutions from the true solutions (Gui et. al., 2020). This method will assist in fine tuning the accuracy of global predictions at which increased CO_2 stops being a benefit to WUE, and increased temperature, heat stress, or other causal factors like changes to VPD and ET lead to detrimental WUE. In addition, the GAN structure of using a generator to try and capture the distribution of true examples, the discriminator locating true examples as accurately as possible, and using strong optimization of the training data (e.g., spatial maps of WUE, NEE, GPP from flux towers and satellite products like MODIS EVI, OCO-2 SIF, ECOSTRESS, SMAP soil moisture) based on maximum likelihood game, will be useful for climate-WUE-carbon feedbacks estimates. For example, when/how C sequestration becomes limited due to climate change, WUE, mortality feedbacks.

Genetic Algorithms (GA) are an optimization procedure inspired by genetics and evolution (Mitchell 1996, Coley 1999). Although GA optimizations cannot be guaranteed to find the global optimum, they are effective techniques for finding good solutions to high-dimensional problems. Ecosystem modelers have used GA techniques to optimize inputs for classification trees and NNs (D'Heyere et al., 2006), forest harvest scheduling (Cropper et. al., 2012), and landscape habitat management (Holzkamper et al., 2006). Our objective is to use inverse analysis using a GA in ELM-FATES (section 5.2) to optimize model parameters associated with forest water recycling benefits given degrees of deforestation, based on improved processes of WUE and transpiration, and observed data. Inverse modeling is an approach that emphasizes using data to determine the model rather than the forward problem of simulating the dynamics of a model from an initial value. Uncertainty can never be eliminated, but it is useful to find a set of parameter values that best describes observed data using optimization techniques.

5.2 Ecosystem Modeling: Vegetation demographic models have increased the representation of plant functional diversity and simulate transient changes in vegetation in response to ecosystem perturbations, via mechanistic demographic processes while being explicitly linked to a biogeochemical model (Fisher et al. 2015; Holm et al. 2020; Koven et al. 2020). However, with increased complexity comes increased uncertainty, and these models have poor performance because high dimensionality causes multi-variate interactions between WUE, plant physiology, and abiotic climate and environmental conditions described above that are difficult to parameterize. Instead, using ML techniques that extract important features and meaningful relationships by interactively transforming the model data could help to simulate dynamics of WUE driven shifts in community competition (e.g., C_3 natural vegetation vs. C_4 crops), that are known to occur in response to multi-year periods of extreme drought (Stocker et al., 2019). We will use the modeling platform ELM-FATES (E3SM Land Model (ELM) - Functionally Assembled Terrestrial Ecosystem Simulator (FATES)) developed by DOE. ELM-FATES now represents cohort-based plant trait variation, dynamic growth and survival rates, and plant-soil hydrodynamics, key features for projecting vegetation response in water stressed climates. Community dynamics and vertically stratified tree size- and age-class structure are emergent properties from competition for light, water and nutrients which dynamically and explicitly scale up from the tree, to stand, to ecosystem level, required for addressing the scale issue of WUE.

6. Suggested Partners/Experts: As this is a newly developing research focus area, I am still working on a final list of collaborating partners and experts, from university, laboratories, and industry partners. Potential partners who have expertise on the topics proposed here include: Dr. Trevor Keenan who is an expert on climate, CO_2, and WUE interactions, Dr. Daniel Ricciuto who specializes in the application of uncertainty quantification (UQ) techniques, and Dr. Vincent Emanuele is the Co-Founder and CTO of Embedded Intelligence (EI), a Machine Intelligence R&D company, with a PhD in Electrical and Computer Engineering from Georgia Tech.

7. References

Aragão, L. The rainforest's water pump. *Nature* 489, 217–218 (2012). https://doi.org/10.1038/nature11485

Bart, R. R., Tague, C. L., & Moritz, M. A. (2016). Effect of Tree-to-Shrub Type Conversion in Lower Montane Forests of the Sierra Nevada (USA) on Streamflow. *PLOS ONE*, *11*(8), e0161805. https://doi.org/10.1371/journal.pone.0161805

Baydin, A.G., Pearlmutter, B.A., Radul, A., & Siskind, J. (2017). Automatic differentiation in machine learning: a survey. *J. Mach. Learn. Res., 18*, 153:1-153:43.

Coley, D.A., 1999. An introduction to genetic algorithms for scientists and engineers. World Scientific, Singapore.

Cropper, W. P. Jr., J. Holm, C. Miller. 2012. An inverse analysis of a matrix population model using a genetic algorithm. Ecological Informatics. 7, 41-45.

Dai, A. Increasing drought under global warming in observations and models. *Nat Clim Chan*, 3, 52-58 (2013).

D'Heyere, T., Goethals, P.L.M., De Pauw, N., 2006. Genetic algorithms for optimisation of predictive ecosystems models based on decision trees and neural networks. Ecological Modelling 195, 20–29.

Eltahir, E. A. B., and Bras, R. L. (1996), Precipitation recycling, *Rev. Geophys.*, 34(3), 367– 378, doi:10.1029/96RG01927.

Fisher, R. A., Muszala, S., Verteinstein, M., Lawrence, P., Xu, C., McDowell, N. G., Knox, R. G., Koven, C., Holm, J., Rogers, B. M., Spessa, A., Lawrence, D., and Bonan, G.: Taking off the training wheels: the properties of a dynamic vegetation model without climate envelopes, CLM4.5(ED), Geosci. Model Dev., 8, 3593–3619, https://doi.org/10.5194/gmd-8-3593-2015, 2015.

Fisher, R.A., C.D. Koven, W.R.L. Anderegg. 2018. Vegetation demographics in Earth System Models: A review of progress and priorities. Global Change Biology.

Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., et al. (2017). Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geoscience*, *10*(6), 410–414.

Gui, J., Z. Sun, Y. Wen, D. Tao, and J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," arXiv preprint arXiv:2001.06937, 2020.

Holm, J. A., Knox, R. G., Zhu, Q., Fisher, R. A., Koven, C. D., Nogueira Lima, A. J., et al. (2020). The central Amazon biomass sink under current and future atmospheric CO_2: Predictions from big-leaf and demographic vegetation models. *Journal of Geophysical Research: Biogeosciences*, 125, e2019JG005500. https://doi.org/10.1029/2019JG005500

Holzkamper, A., Lausch, A., Seppelt, R., 2006. Optimizing landscape configuration to enhance habitat suitability for species with contrasting habitat requirements. Ecological Modelling 198, 277–292.

Keenan, T., Hollinger, D., Bohrer, G. *et al.* Increase in forest water-use efficiency as atmospheric carbon dioxide concentrations rise. *Nature* 499, 324–327 (2013). https://doi.org/10.1038/nature12291

Koven, C. D., Knox, R. G., Fisher, R. A., Chambers, J. Q., Christoffersen, B. O., Davies, S. J., Detto, M., Dietze, M. C., Faybishenko, B., Holm, J., et al.: Benchmarking and parameter sensitivity of physiological and vegetation dynamics using the Functionally Assembled Terrestrial Ecosystem Simulator (FATES) at Barro Colorado Island, Panama, Biogeosciences, 17, 3017–3044, https://doi.org/10.5194/bg-17-3017-2020, 2020.

Knapp, A. K., Chen, A., Griffin-Nolan, R. J., et al. Resolving the Dust Bowl paradox of grassland responses to extreme. *PNAS*, 2020.

Liu, Y., Kumar, M., Katul, G. G., Feng, X., & Konings, A. G. (2020). Plant hydraulics accentuates the effect of atmospheric moisture stress on transpiration. *Nature Climate Change*, https://doi.org/10.1038/s41558-020-0781-5

Mitchell, M., 1996. An introduction to genetic algorithms. MIT Press, Cambridge, Mass.

Raissi, M., Perdikaris, P., & Karniadakis, G. (2017). Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations. *ArXiv, abs/1711.10561*.

Ripley, B. S., Gilbert, M. E., Ibrahim, D. G. & Osborne, C. P. Drought constraints on C4 photosynthesis: Stomatal and metabolic limitations in C3 and C4 subspecies of Alloteropsis semialata. *J. Exp. Bot.* 58, 1351-1363 (2007).

Spracklen, D., Arnold, S. & Taylor, C. Observations of increased tropical rainfall preceded by air passage over forests. *Nature* 489, 282–285 (2012). https://doi.org/10.1038/nature11390

Stocker, B. D. et al. Drought impacts on terrestrial primary production underestimated by satellite monitoring. *Nat. Geosci.* 12, 264–270 (2019).

Taylor, S. H. et al. Physiological advantages of C4 grasses in the field: A comparative experiment demonstrating the importance of drought. *Glob. Chang. Biol.* 20, 1992-2003 (2014).

Trenberth, K. E., Dai, A., Schrier, G. van der, Jones, P. D., Barichivich, J., Briffa, K. R., & Sheffield, J. (2014). Global warming and changes in drought. *Nature Climate Change*, *4*(1), 17–22.

Williams, A. P., Cook, E. R., Smerdon, J. E., Cook, B. I., Abatzoglou, J. T., Bolles, K., et al. (2020). Large contribution from anthropogenic warming to an emerging North American megadrought. *Science*, *368*(6488), 314–318.