**Quantifying drivers of uncertainty in land model predictions at global scales using machine learning**

**Daniel Ricciuto1, Khachik Sargsyan2, Dan Lu1, Jiafu Mao1, Anping Chen3**

1 Oak Ridge National Laboratory

2 Sandia National Laboratories

3 Colorado State University

Uncertainty quantification (UQ) algorithms link observations and models together to produce credible simulations of Earth system behavior with uncertainty estimates. Dimension-reduction techniques in combination with machine learning can be used to build fast-to-evaluate surrogate models of spatiotemporally varying model output fields using a relatively small number of ESM ensemble members and higher accuracy than traditional methods. Here we focus on uncertainty quantification of the Energy Exascale Earth System land model (ELM) parameters related to ecosystem processes considering uncertainty in key model parameters related to fluxes of carbon and energy, phenology and drought response. Parameter uncertainty ranges are determined from trait databases and the literature across a range of 13 naturally occurring plant functional types. A 275-member ensemble of global ELM simulations is performed at 2x2 degree spatial resolution. We then create a temporally and spatially resolved surrogate model of gross primary productivity using the dimension-reduction techniques described above. Global sensitivity analysis performed using the surrogate model indicates different parameters drive model prediction uncertainty depending on time of year and environmental conditions. In warmer and drier climates, parameters controlling stomatal conductance and rooting depth distribution are strong drivers of productivity, while in colder climates phenology and temperature sensitivity parameters are more important. We investigate how these parameter sensitivities change under extreme conditions. Finally, we perform a calibration on the surrogate model using Bayesian methods to demonstrate how ELM parameters and predictions may be improved using gridded observation datasets.