Project Title: A machine-learning data assimilation method to improve lake methane prediction

Authors: Zeli Tan (PNNL), Johnny Li (University of Idaho), Hoang Viet Tran (PNNL), Dalei Hao (PNNL)

Focal Areas: The proposed research will demonstrate the potential to assimilate satellite and unmanned aerial vehicle (UAV) data into a lake methane model to improve the accurary of lake methane prediction in the temperate climate regions of the Northern Hemisphere.

Science or Technological Challenge: Methane emissions from lakes are one of the largest natural methane sources, comprising as much as one third of global methane emissions based on a recent estimate [1]. Lakes are prominent landscape elements in the temperate regions of the Northern Hemisphere. Due to climate warming, methane emissions from northern lakes are expected to rise rapidly in this century [2]. However, our capability to model methane emissions from northern lakes is still very limited. First, many related lake physical processes, such as water mixing and ice phenology, have not been well constrained in the current large-scale lake models [3]. Second, lake primary production not only provides labile organic carbon for methanogens to produce methane in anoxic conditions but also fuel methanogenesis in oxygen-rich conditions. It is found to be a critical factor for methane emissions from northern lakes but hasn't been well represented in the current large-scale lake models [4]. Third, satellite-based methane data do not have sufficient signal-to-noise ratios and spatiotemporal resolutions to detect methane plumes from lake surface [5].

Rationale: Although data assimilation is widely used to improve the performance of numerical models, there are only limited applications for methane models. It is mainly because traditional data assimilation (DA) methods, such as ensemble Kalman filter (EnKF), have high computational and implementation cost. Also, traditional DA methods are not efficient to assimilate different types of data. To bridge the gaps, we propose to develop a Long Short-Term Memory (LSTM) neural-network-based DA method that harnesses quality, multi-source satellite data of ice cover, lake surface water temperature, Secchi depth and chlorophyll and high-quality UAV data of methane fluxes to: (1) optimize the lake model's parameters and (2) improve estimations of lake methane emissions. Compared to traditional DA methods, the proposed machine-learning based method will be computationally efficient, code-change free, and bias-proofed against ill-selected likelihood functions [6]. Once validated, the method can be extended to the whole temperate regions and even the high-latitude regions of the Northern Hemisphere to constrain lake methane emissions from these regions. The effort will strongly benefit to the accomplishement of the Global Methane Pledge.

Narrative: We will develop a LSTM neural-network-based DA method to assimilate satellite and UAV data to improve the estimation of lake methane emissions in temperate regions of the Northern Hemisphere.

Description of the machine learning based DA method: To overcome the limitations of traditional DA methods, we will develop a LSTM based surrogate model to efficiently assimilate different types of lake data. Specifically, the neural density estimator will adjust the prior distribution of model parameters using satellite observations and UAV measurements to provide desiable sets of model parameters (Figure 1).

We will construct the DA framework in three steps:

1. First, on the study lake, we will run a small number of lake model ensembles and then train a LSTM surrogate model using climate inputs, simulated lake dynamics and model parameters. This step makes the surrogate LSTM model to learn the ice phenology, mixing, primary production and methane flux physics of the process-based lake methane model.
2. Next, we will use a neural density estimator to assess the uncertainty of the LSTM ensembles based on satellite and UAV observations and produce the posterior distribution of the parameters. The neural density estimator is a special approach to the 'inverse problem'. Simulators use parameters (θ) to simulate data (Y). Inference goes the other direction, using observed data (Y_{True}) to get back the parameters (θ). Here, we will utilize the neural density estimator with satellite and UAV observations to infer better values for lake parameters with physical meaning.
3. Finally, lake parameters based on the posterior distributions will be fed to the LSTM surrogate model to produce estimations of lake thermal and methane dynamics that are close to observations. The estimated lake thermal and methane dynamics will be validated against in-situ observations.

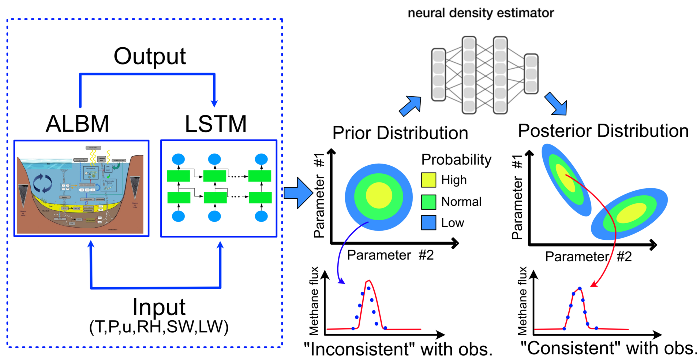

Fig. 1. Workflow of training a ML-based surrogate model of a 1-D lake methane model and using the surrogate model and a neural density estimator to assimilate satellite and UAV data to constrain simulated lake methane emissions. Climate inputs for Arctic Lake Biogeochemistry Model (ALBM) and LSTM are air temperature (T), precipitation (P), wind speed (u), relative humidity (RH), shortwave radiation (SW) and longwave radiation (LW).

<u>Description of the lake model</u>: The Arctic Lake Biogeochemistry Model (ALBM) model is a 1-D process-based lake methane model that was developed by Dr. Zeli Tan [7]. In this research, we will use the ALBM model to produce prior lake thermal and methane estimates for training the LSTM surrogate model. Here, we briefly describe the model processes and structure that are related to lake stratification simulations. ALBM is an integral energy lake model based on the Hostetler diffusivity parameterization, with depth-resolved 1-D water and sediment columns. Both the water and sediment columns have variable layer thickness, with thinner layers at surface to represent more intense thermal dynamics. In the model, lake methane emissions are governed by methane production, oxidation and transport (via both diffusion and ebullition). The ALBM model has demonstrated good performance in simulating methane emissions of specific northern lakes [7, 8].

<u>Description of the satellite and UAV observations</u>: We will use the satellite and UAV data of different lake thermal and biogeochemical variables together for data assimilation. All satellite data operations will be executed using Google Earth Engine. **(1)** For lake surface temperature, we will use the Advanced Very High Resolution Radiometer (AVHRR) sensor based GloboLakes product for large lakes and the Landsat-8 thermal data for small lakes. **(2)** For ice cover, we will use the Advanced Microwave Scanning Radiometer (AMSR) sensor based daily ice phenology data for large lakes and the Landsat-based ice phenology data for small lakes. **(3)** For Secchi depth, we will extract the values for the studied lakes from multiple satellite sensors, including Moderate Resolution Imaging Spectroradiometer (MODIS), Landsat-8 Operational Land Imager (OLI), Sentinel-2 MultiSpectral Imager (MSI), and MEdium Resolution Imaging Spectrometer (MERIS), by adopting a well-established quasi-analytical algorithm. **(4)** For chlorophyll, we will use the data from the Landsat-8 OLI and the Ocean and Land Color Instrument (OLCI) aboard the Sentinel 3 satellite. **(5)** For methane fluxes, we will use drones carrying multimodal instruments to measure lake water surface temperature and map methane plumes close to sources and use coincident wind measurements to derive flux accurately and reliably. The approach will be developed at the University of Idaho.

References

[1] Rosentreter, J. A. et al. Half of global methane emissions come from highly variable aquatic ecosystem sources. Nat. Geosci. 14, 225-230 (2021).

[2] Tan, Z., & Zhuang, Q. Arctic lakes are continuous methane sources to the atmosphere under warming conditions. Environmental Research Letters, 10, 054016 (2015).

[3] Guseva, S. et al. Multimodel simulation of vertical gas transfer in a temperate lake. Hydrol. Earth Syst. Sc. 24, 697-715 (2020).

[4] West, W. E., Creamer, K. P., & Johns, S. T. Productivity and morphometry regulate lake contributions to atmospheric methane. Limnology and Oceanography, 61, S51–S61 (2016).

[5] Tan, Z. et al. Inverse modeling of pan-Arctic methane emissions at high spatial resolution: what can we learn from assimilating satellite retrievals and using different process-based wetland and lake biogeochemical models?. Atmospheric Chemistry and Physics, 16, 12649-12666 (2016).

[6] Tsai, W. P., et al. From calibration to parameter learning: Harnessing the scaling effects of big data in geoscientific modeling. Nat. Commun. 12, 5988 (2021).

[7] Tan, Z., Zhuang, Q., & Walter Anthony, K. Modeling methane emissions from arctic lakes: Model development and site-level study. Journal of Advances in Modeling Earth Systems, 7, 459-483 (2015).

[8] Guo, M. et al. Rising methane emissions from boreal lakes due to increasing ice-free days. Environmental Research Letters, 15, 064008 (2020).