Title: AI for advanced sensor data collection, automation, and processing for the methane cycle

Authors: Maruti K. Mudunuru (PNNL), Nikolla P. Qafoku (PNNL), Andre Coleman (PNNL), Satish Karra (PNNL), Mariefel Olarte (PNNL), Sarah Barrows (PNNL), Tamas Varga (PNNL), Odeta Qafoku (PNNL), Glenn Hammond (PNNL), Behzad Ghanbarian (KSU), and Mahantesh Halappanavar (PNNL).

Focal Area(s): Topic-4 – Automated or real-time data capture and processing or federated learning for improvements in measurement coverage.

Science or Technological Challenge: *How to use recent advances in AI to obtain automated measurements of methane flux from distributed sensor networks (particularly in soil and agricultural systems) and process the collected data efficiently?*

Rationale:

Research needs and challenges: One of the main contributors to the methane (CH_4) cycle is the CH_4 gas emitted by microbes in soils is significantly impacted by human activities[1]. Microorganisms from other sources, such as landfills, livestock, and the exploitation of fossil fuels, also emit CH_4. To better understand the methane flux under a wide range of environmental conditions and ecological stressors, various programs (e.g., FLUXNET-CH_4, COSORE)[2,3] are actively collecting data spatiotemporally that are commonly used in process models (e.g., PFLOTRAN, GCAM)[4,5] in a coupled modeling-experimental (ModEx) approach. However, there are some challenges associated with this traditional ModEx approach, some of which were recently disclosed within the AI4ESP workshop report[6] highlights. In the report, how to use recent advances in AI to overcome some of the traditional ModEx approach challenges was also highlighted. However, many data analysis challenges still need to be answered[6]. Within the context of the methane cycle, we believe there are knowledge gaps that AI would enable us to address by integrating modeling and analysis activities across the field- and lab-scale experiments, particularly related to soil and agricultural systems. Those gaps are:

- *Quality of the collected data from sensor networks –* This includes identifying methane flux signatures (e.g., microbial activity due to anthropogenic stressors, extreme events) from sensor data, filling in data gaps, and associated data worth analysis.
- *When, how, and where to collect data –* It is not feasible to measure fluxes all the time. Thus, we need a way to decide when and how to measure (and possibly even where) methane flux smartly and efficiently.
- *Dealing with big data –* When advanced sensors, such as multispectral cameras are used, the amount of the data collected is substantial. So, efficiently processing this data at the sensor edge is needed.

Our proposed approach to address the above challenges is to develop self-aware and intelligent sensor nodes. This self-awareness is achieved by advancing and tailoring our AI@SensorEdge workflow (e.g., edge-to-cloud intelligence)[7-9] as depicted in Figure 1.

Narrative:

Scientific and technical description: An AI@SensorEdge workflow provides a transformational way to integrate multi-modal data through sensor fusion (e.g., combining geophysical, geochemical, and hydrological sensor data sampled at different frequencies). Moreover, efficiently harnessing the connectivity of intelligent sensors through edge and fog computing will result in an advanced understanding of soil and agricultural systems under disturbances and extreme events in near real-time. Development in advanced flux data acquisition systems, sensor network design for soil and farming systems, hardware-related efforts (e.g., AI-enabled accelerators), lightweight AI

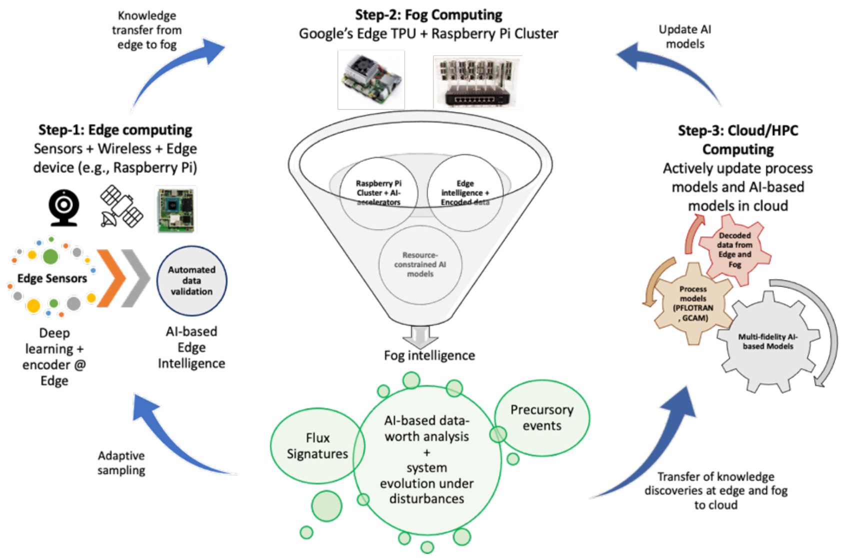

Figure 1. AI@SensorEdge workflow to extract actionable information and discover knowledge from methane flux sensor networks under disturbances.

models (e.g., energy-efficient), and cyber security for edge computing will advance the proposed science[10]. This workflow will recognize:

- *Data quality* – AI-based local data worth analysis will determine if sensor data or signals might contain useful information to detect flux signatures and underlying patterns. The discovered signatures will be provided to process models (e.g., PFLOTRAN) and then converted to actionable intelligence (e.g., system evolution) at the edge.
- *Data collection* – AI@SensorEdge can accelerate the collection of informative data by creating a digital twin for soil and agricultural systems (e.g., through IoT). We will optimize the location of sensors by exploring the system behavior in digital space.
- *Data volume* – Edge computing-based AI models (e.g., RNNPool[11], SmartTensors AI platform[12]) can be leveraged to compress data efficiently. This compressed data can be transferred to the cloud and HPC systems through 5G-enabled AI@SensorEdge programming models[13-17].

AI@SensorEdge workflow interfacing with FAIR data sources: The real-time flux measurements collected from sensor networks can be interfaced with existing resources and databases such as

- Soil chemistry from Web Soil Survey (NRCS) – To understand the impact of impact of soil chemistry to methane output.
  - https://www.nrcs.usda.gov/conservation-basics/natural-resource-concerns/soils/soil-geography
- Farming data (e.g., types of crops planted in the ground, livestock)
  - https://quickstats.nass.usda.gov/
- FLUXNET-CH_4 – Methane flux measurements
  - https://fluxnet.org/data/fluxnet-ch4-community-product/
- COSORE – Soil respiration and greenhouse gas flux data, https://github.com/bpbond/cosore
- Soil respiration database: https://github.com/bpbond/srdb
- Flux measurement sites for carbon, water, and/or energy
  - AmeriFlux Network – https://ameriflux.lbl.gov/, LeafWeb – https://www.leafweb.org/
  - SPRUCE experimental databases – https://mnspruce.ornl.gov/

Pre-trained AI models can be embedding on these distributed sensor networks through smart computing devices such as Raspberry Pi CM4+. These intelligent edge devices also provide a venue to interface with next generation WiFi and 5G network. The flux data acquired from these sensor networks and processed using AI algorithms can be made reusable and findings reproduceable through FAIR data sources such as ESS-DIVE and EMSL-GitHub.

References (Optional):

1. Nazaries, L., Murrell, J.C., Millard, P., Baggs, L. and Singh, B.K., 2013. Methane, microbes and models: fundamental understanding of the soil methane cycle for future predictions. Environmental microbiology, 15(9), pp.2395-2417.
2. FLUXNET-CH4: a global, multi-ecosystem dataset and analysis of methane seasonality from freshwater wetlands. https://essd.copernicus.org/articles/13/3607/2021/
3. Bond-Lamberty, B., Christianson, D.S., Malhotra, A., Pennington, S.C., Sihi, D., AghaKouchak, A., Anjileli, H., Altaf Arain, M., Armesto, J.J., Ashraf, S. and Ataka, M., 2020. COSORE: A community database for continuous soil respiration and other soil-atmosphere greenhouse gas flux data. Global change biology, 26(12), pp.7268-7283.
4. PFLOTRAN -- https://www.pflotran.org/
5. GCAM -- https://jgcri.github.io/gcam-doc/
6. Hickmon, Nicki L., Varadharajan, Charuleka, Hoffman, Forrest M., Collis, Scott, and Wainwright, Haruko M. Artificial Intelligence for Earth System Predictability (AI4ESP) Workshop Report. United States: N. p., 2022. Web. doi:10.2172/1888810.
7. C. J. Talsma, K. C. Solander, M. K. Mudunuru, B. Crawford, and M. R. Powell, Frost Prediction using Machine Learning and Deep Neural Network Models for Use on IoT Sensors, Frontiers in Artificial Intelligence, https://doi.org/10.3389/frai.2022.963781, 2023.
8. M. K. Mudunuru (2019). EDGEip - Intelligent Processing at the Edge to Enhance Efficiency, LA-UR-19-29757.
9. M. K. Mudunuru (2019). IoGES: Internet of Things for Geophysical and Environmental Sensing; LA-UR-19-30467; 2019. DOI: 10.13140/RG.2.2.32470.40006
10. Friha, O., Ferrag, M.A., Shu, L., Maglaras, L., Choo, K.K.R. and Nafaa, M., 2022. FELIDS: Federated learning-based intrusion detection system for agricultural Internet of Things. Journal of Parallel and Distributed Computing, 165, pp.17-31.
11. O. Saha, A. Kusupati, H. V. Simhadri, M. Varma, P. Jain (2020). RNNPool: Efficient Non-linear Pooling for RAM Constrained Inference. arXiv preprint arXiv:2002.11921.
12. SmartTensors AI Platform: https://github.com/SmartTensors/
13. P. Beckman, C. Catlett, M. Ahmed, M. Alawad, L. Bai, P. Balaprakash, K. Barker, P. Beckman, R. Berry, A. Bhuyan, and G. Brebner (2020). 5G Enabled Energy Innovation: Advanced Wireless Networks for Science, Workshop Report. USDOE Office of Science (SC)(United States).
14. Waggle: An open Platform for AI@Edge Computing and Intelligent Sensors https://wa8.gl/science/array-of-things/
15. AoT: Array of Things -- https://arrayofthings.github.io/
16. EdgeML: https://microsoft.github.io/EdgeML/
17. TensorFlow Lite: https://www.tensorflow.org/lite