ARTIFICIAL INTELLIGENCE FOR THE METHANE CYCLE

Pamela Weisenhorn (ANL), James Ang (PNNL), Jaydeep Bardhan (PNNL), Max Grover (ANL), Forrest Hoffman (ORNL), Daniel Ricciuto (ORNL), Charuleka Vardharajan (LBNL)

Olga Tweedy and Paul Bayer

BERAC, April 12, 2024

AI for Earth System Predictability

AI4ESP

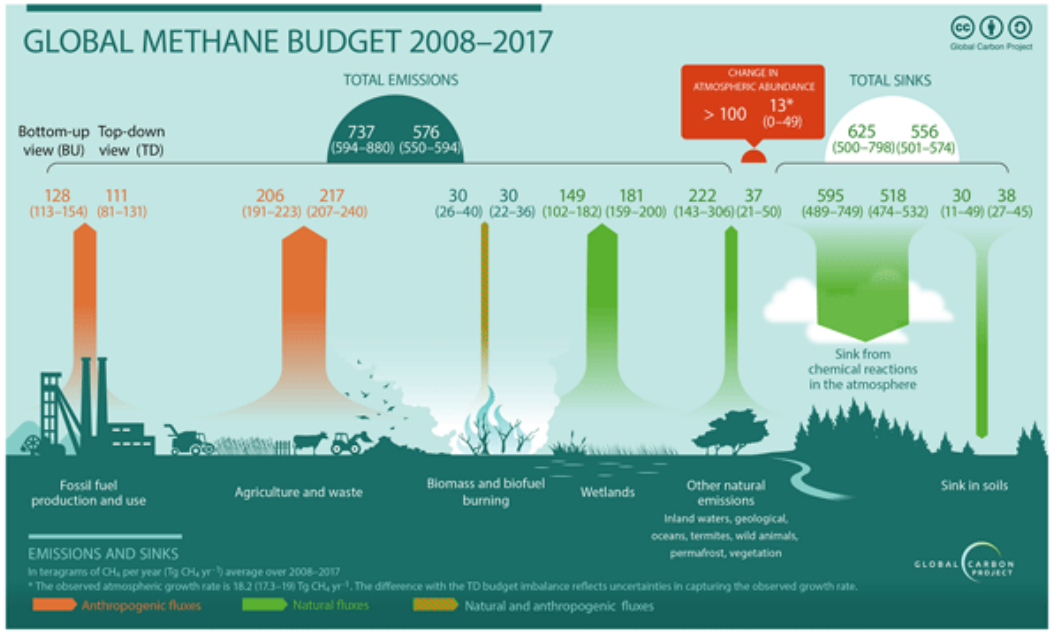
- October-December 2021
- Report released Sept 2022
- Over 740 participants from 178 institutions

- Huge workshop with huge findings

"need to incorporate AI into models, analytics, and data generation as a means to accelerate advancement, create new scientific opportunities, and revolutionize new approaches to predictive capabilities and capacity"

Methane as a Case Study

- Methane has a GWP-100 of 27-30
- Despite intense study, high model uncertainty persists
- Challenges spanning biological and environmental fields



Generalizable

- Understanding the physiology, activity, and impact of plants and microbes at ecosystem scales and larger.
- Connecting measurements and insights between laboratory and field.
- Capturing hot spots and hot moments.
- Integrating field measurements across scales.
- Establishing benchmarks for model development and intercomparisons.

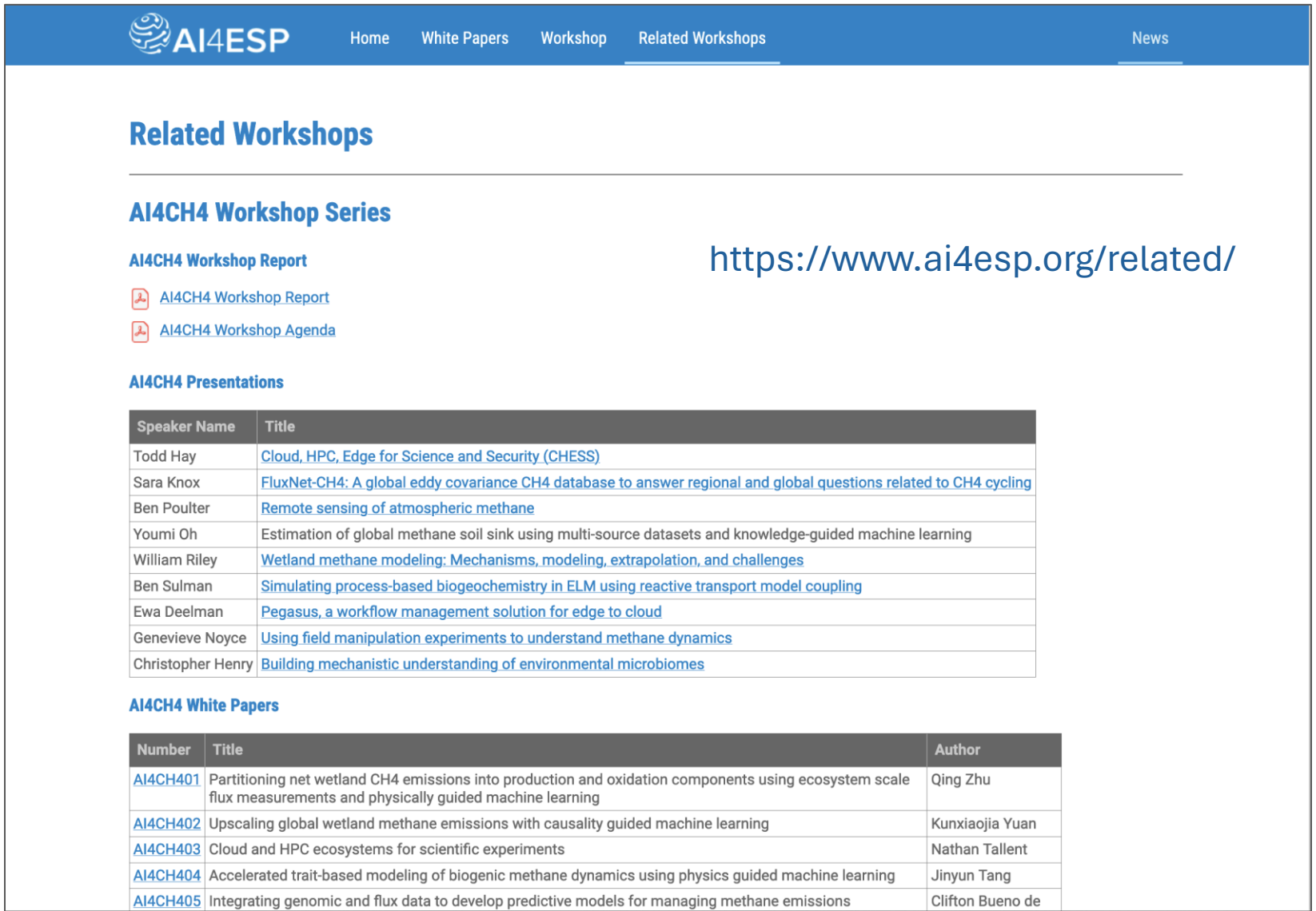
Workshop Goals

Workshop goals included identifying challenges and opportunities in data and modeling for the methane cycle and charting potential paths toward incorporating AI into future BER- supported research implementing the ModEx framework.

Workshop Structure

- Call for White Papers
  - 22 white papers received
- Workshop was held virtually over 4 days, 5-hour sessions
- Total of 102 contributors as white paper authors or participants at the workshop representing 30 institutions, including 7 national labs, NOAA, and NASA

Workshop Content Available



AI4ESP Home White Papers Workshop Related Workshops News

Related Workshops

AI4CH4 Workshop Series

https://www.ai4esp.org/related/

AI4CH4 Workshop Report

- AI4CH4 Workshop Report
- AI4CH4 Workshop Agenda

AI4CH4 Presentations

Speaker Name	Title
Todd Hay	Cloud, HPC, Edge for Science and Security (CHESS)
Sara Knox	FluxNet-CH4: A global eddy covariance CH4 database to answer regional and global questions related to CH4 cycling
Ben Poulter	Remote sensing of atmospheric methane
Youmi Oh	Estimation of global methane soil sink using multi-source datasets and knowledge-guided machine learning
William Riley	Wetland methane modeling: Mechanisms, modeling, extrapolation, and challenges
Ben Sulman	Simulating process-based biogeochemistry in ELM using reactive transport model coupling
Ewa Deelman	Pegasus, a workflow management solution for edge to cloud
Genevieve Noyce	Using field manipulation experiments to understand methane dynamics
Christopher Henry	Building mechanistic understanding of environmental microbiomes

AI4CH4 White Papers

Number	Title	Author
AI4CH401	Partitioning net wetland CH4 emissions into production and oxidation components using ecosystem scale flux measurements and physically guided machine learning	Qing Zhu
AI4CH402	Upscaling global wetland methane emissions with causality guided machine learning	Kunxiaojia Yuan
AI4CH403	Cloud and HPC ecosystems for scientific experiments	Nathan Tallent
AI4CH404	Accelerated trait-based modeling of biogenic methane dynamics using physics guided machine learning	Jinyun Tang
AI4CH405	Integrating genomic and flux data to develop predictive models for managing methane emissions	Clifton Bueno de

Focal Topics

Improving predictions from fundamental microbiology

Environmental controls and empirical relationships

Targeting field measurements and observations

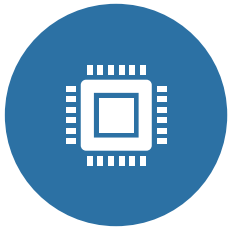
Data-model integration

Multiscale modeling

Discussion Questions

KNOWLEDGE GAPS AND SCIENTIFIC QUESTIONS

CHARACTERISTICS AND CHALLENGES OF SPECIFIC DATA AND MODELS



RELATED ALGORITHMS, INFRASTRUCTURE, AND THEIR POTENTIAL TO ADDRESS GAPS

NEW OBSERVATIONS, MEASUREMENTS, AND EXPERIMENTAL APPROACH

DATA PRODUCTS (QC, UQ, HARMONIZATION, BENCHMARKS)

Enhance observation and experimentation

- Need to develop autonomous sensor systems that can change the frequency or location of sampling based on real-time conditions
- Development and application of surrogate modeling approaches to improve model efficiency and enable data assimilation
- Autonomous and self-driving laboratories have the potential to quickly and robustly perform evaluation of AI model output and modify training data

Supplement contextual data

- Often needed data exist, but are difficult to efficiently use because of issues with access or metadata
- Metadata can be supplemented from existing data sources if those sources were more readily accessible
- Metadata could be generated through the development of models that take advantage of existing data sources

Expand findability and usability of data

- Need for cross-domain exchange of data and access to interpretable model outputs and ability to run models

- Expand the findability and usability of data with quantified uncertainties and complete metadata, including derived quantities and quality control methods

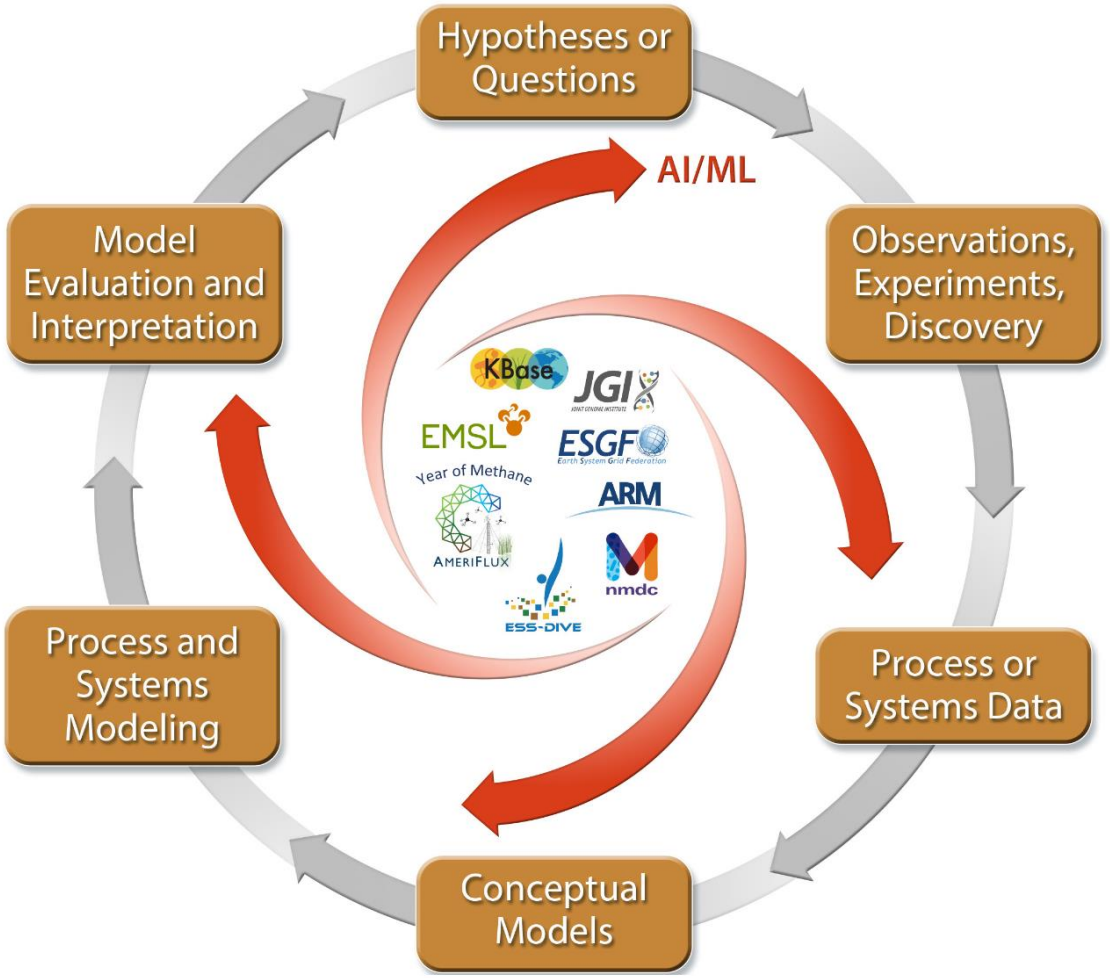
Optimize sampling strategies

- Greater efficiency of surrogate models enables global sensitivity analysis and finer pinpointing of model-critical data

- Improved ability to integrate knowledge across scales
  - Connect ground and radar-based observations
  - Integrate microbes into Earth scale models

Develop and support scientific workflows

- Development and support of scientific workflows that can be used across the computing continuum (HPC to cloud to edge) and across distributed systems involves many challenges: Computational, Communication, Ownership, and Provenance
- Implementation of AI approaches to automate feedback to instruments will create new edge computing capabilities
- ML approaches can combine with domain expertise to address challenges and build an interpretable model, but this requires growing a workforce able to work across these boundaries



Adapting the ModEx Framework to AI Models