Unlocking the Future of Soil Health:

Blending Data-Driven and Process-Based Models to Predict N_2O Emissions Across Sub-Saharan African Soils

M.A.Tufail[1], P. Agredazywczuk[1], T. Ouma[1,2,3], M. Barthel[3], A. Otinga[4], R. Njoroge[3], K. Butterbach-Bahl[5], R. Daelman[6,7], J. E. Hickman[8,9], W. Ibrahim[10], M. Laub[3], S. Leitner[10], A. Shumba[11,12,13], K. L. Tully[14], S. Wachiye[15,16], J. Zheng[17], G. Obozinski[2], J. Six[3], E. Harris[1]

[1] Climate and Environmental Physics, University of Bern, Sidlerstrasse 5, 3012 Bern, Switzerland. [2] Swiss Data Science Center (EPFL and ETH Zurich), 8092 Zurich, Switzerland. [3] Sustainable Agroecosystems Group, Department of Environmental Systems Science, ETH Zurich, Switzerland. [4] Department of Soil Science, University of Eldoret, Eldoret, Kenya. [5] IMK-IFU, Karlsruhe Institute of Technology, Germany. [6] ISOFYS, Department of Green Chemistry and Technology, Ghent University, Belgium. [7] Q-Forest lab, Department of Environment, Ghent University, Belgium. [8] Center for Climate Systems Research, Columbia University, New York, NY. [9] NASA Goddard Institute for Space Studies, New York, NY. [10] Mazingira Centre for Environmental Research and Education, International Livestock Research Institute (ILRI), Naivasha Rd, PO 30709, 00100 Nairobi, Kenya. [11] Fertilizer, Farm Feeds and Remedies Institute, Department of Research and Specialist Services, Ministry of Lands, Agriculture, Fisheries, Water and Rural Development, Harare, Zimbabwe. [12] CIRAD, UPR AIDA, Harare, Zimbabwe. [13] Department of Plant Production Sciences and Technologies, University of Zimbabwe, Harare, Zimbabwe. [14] University of Maryland: College Park, MD, US. [15] Department of Agro-Forestry, Environmental Studies, and Integrated Natural Resources, University of Kabianga, Kenya. [16] Department of Agricultural Sciences, University of Helsinki, Helsinki, Finland. [17] Crop, Livestock and Environment Division, Japan International Research Center for Agricultural Sciences, Tsukuba, Japan.

Introduction

Inorganic fertilisation is expected to increase in sub-Saharan Africa (SSA) to boost agricultural productivity, potentially elevating nitrous oxide (N_2O) emissions, a potent greenhouse gas. Given the high variability and spatial heterogeneity of N_2O emissions, including the formation of emission "hotspots," identifying the key environmental and management drivers is critical for accurate prediction. However, SSA remains an understudied region with limited data availability. To address this gap, there is a pressing need to **calibrate and optimize data-driven and process-based models under SSA-specific conditions**. This will enhance understanding of N_2O drivers and aid region-specific, climate-smart strategies.

Aims:

1. Use of data driven and process-based models to assess drivers of N_2O emissions across environments
2. To quantify annual N_2O emissions and corresponding emission factors, and to project future emission scenarios across varying environments
3. Test, calibrate, and validate the N_2O emissions from each process-based model on Sub-Saharan Arican soils

Methodology

Literature Search

Keywords: N_2O, Nitrous Oxide, Sub-Saharan Africa, Fertiliser
Environments: Forest, Grassland, Cropland

Data Collection

Datasets: 11 continuous or semi-continuous N_2O sets with 64 different sites or different fertiliser treatment within a site

Parameter Interpolation

Parameters: A total of 17, including temperature, rain, radiation, fertilisation, days since fertilisation etc.
Data Handling: Interpolated missing parameters using global modelled datasets

Data Driven Machine Learning based Modeling

Models: Random Forest (RF), XGboost, Artificial neural Network (ANN)
Data Split: Temporal block-based split (80/20 ratio)
Feature Selection: 5-fold cross validation with recursive feature elimination (RF & XGB)

Process Based Modeling

Models: DNDC, DayCent, APSIM, QUINCY, STICS, DAISY, cnmodel

Results: Data-Driven Modelling – (Agredazywczuk et al., 2025 in preparation)

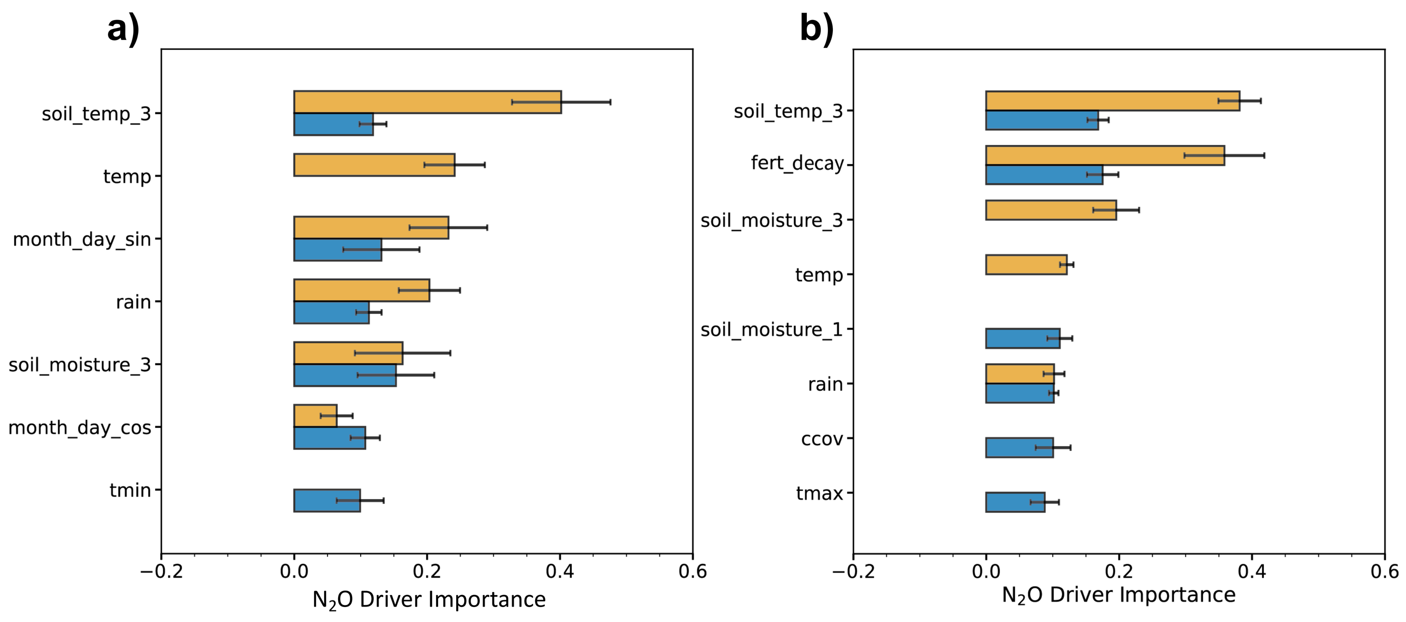


Figure 1: Comparison of important drivers of N_2O production from RF and XGB feature selection processes. **a)** Forest, and **b)** Cropland

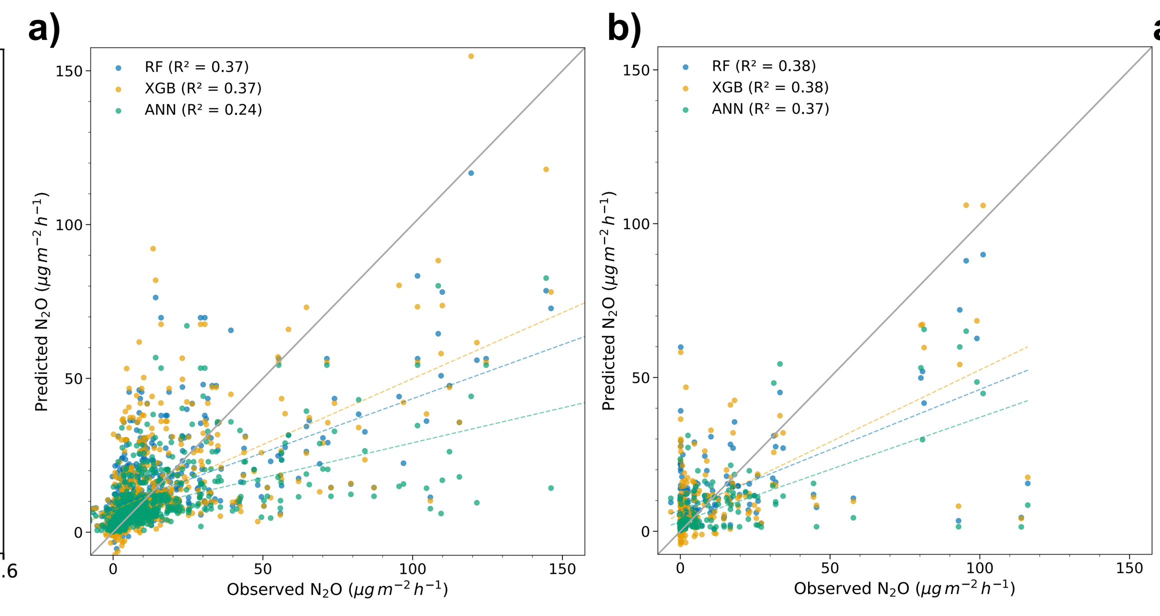


Figure 2: Comparison of observed and predicted daily instantaneous N_2O fluxes on the test set: **a)** Forest, and **b)** Cropland

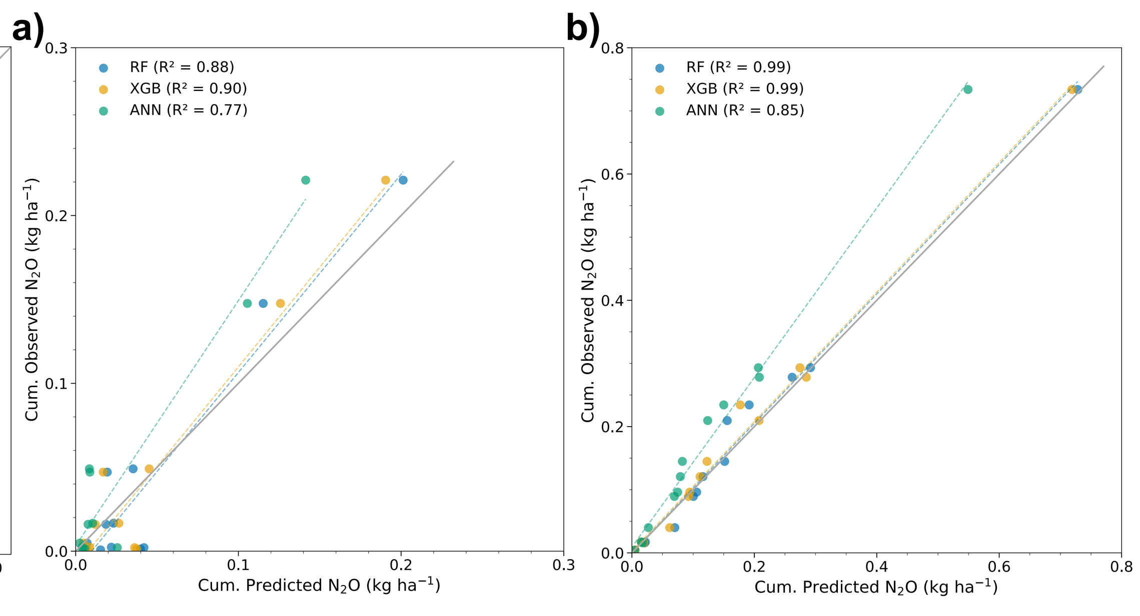


Figure 3: Comparison of cumulative observed and predicted N_2O fluxes on the test set: **a)** Forest, and **b)** Cropland

- RF showed the best R^2 performance across land uses
- Models consistently underestimate emission peaks, but cumulative emissions are captured very well.

- **Temperature and moisture variables:** most important N_2O drivers across all land uses
- **Cropland:** fertiliser exponential decay important driver for both models

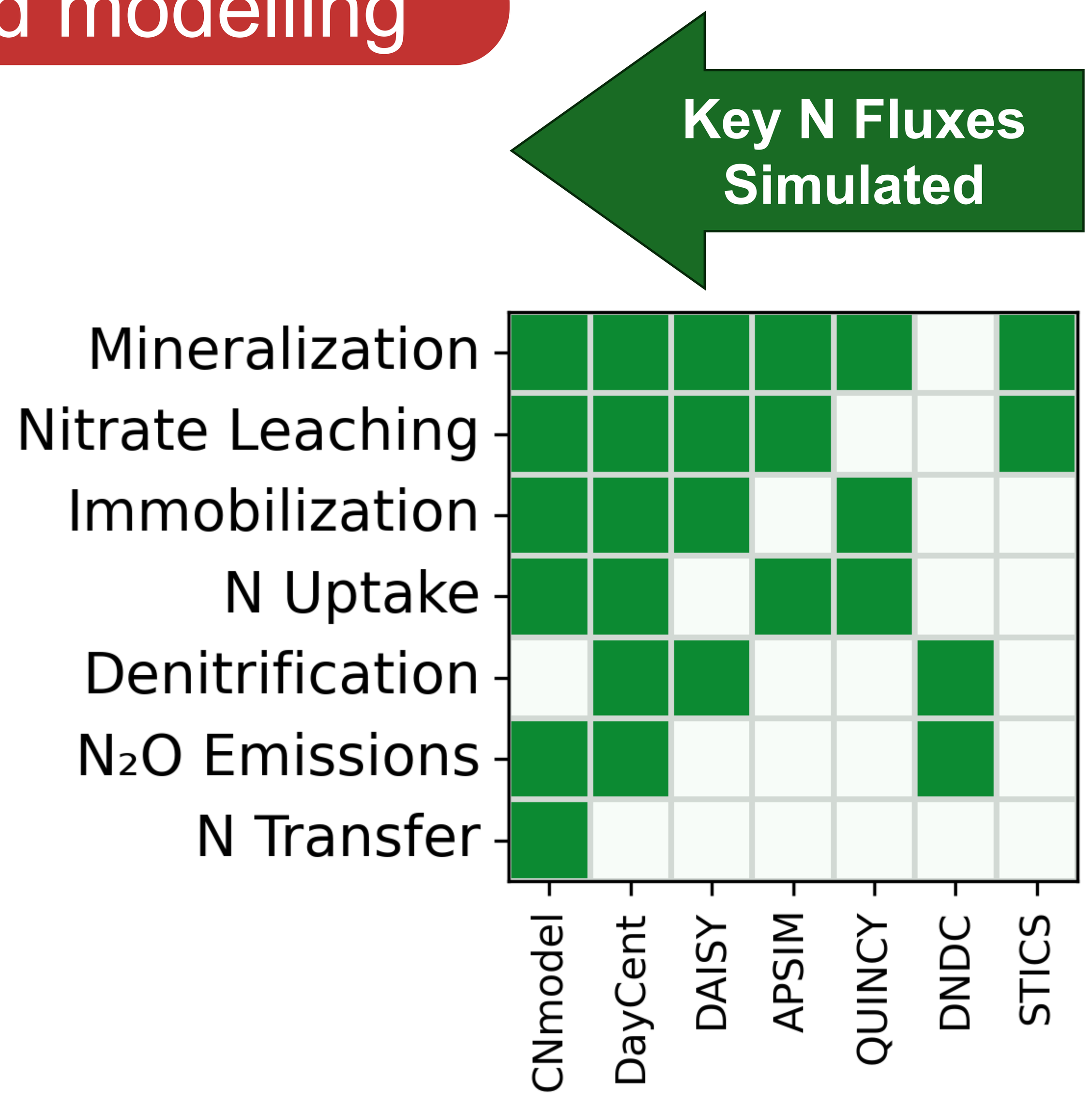
Outlook: Process-based modelling

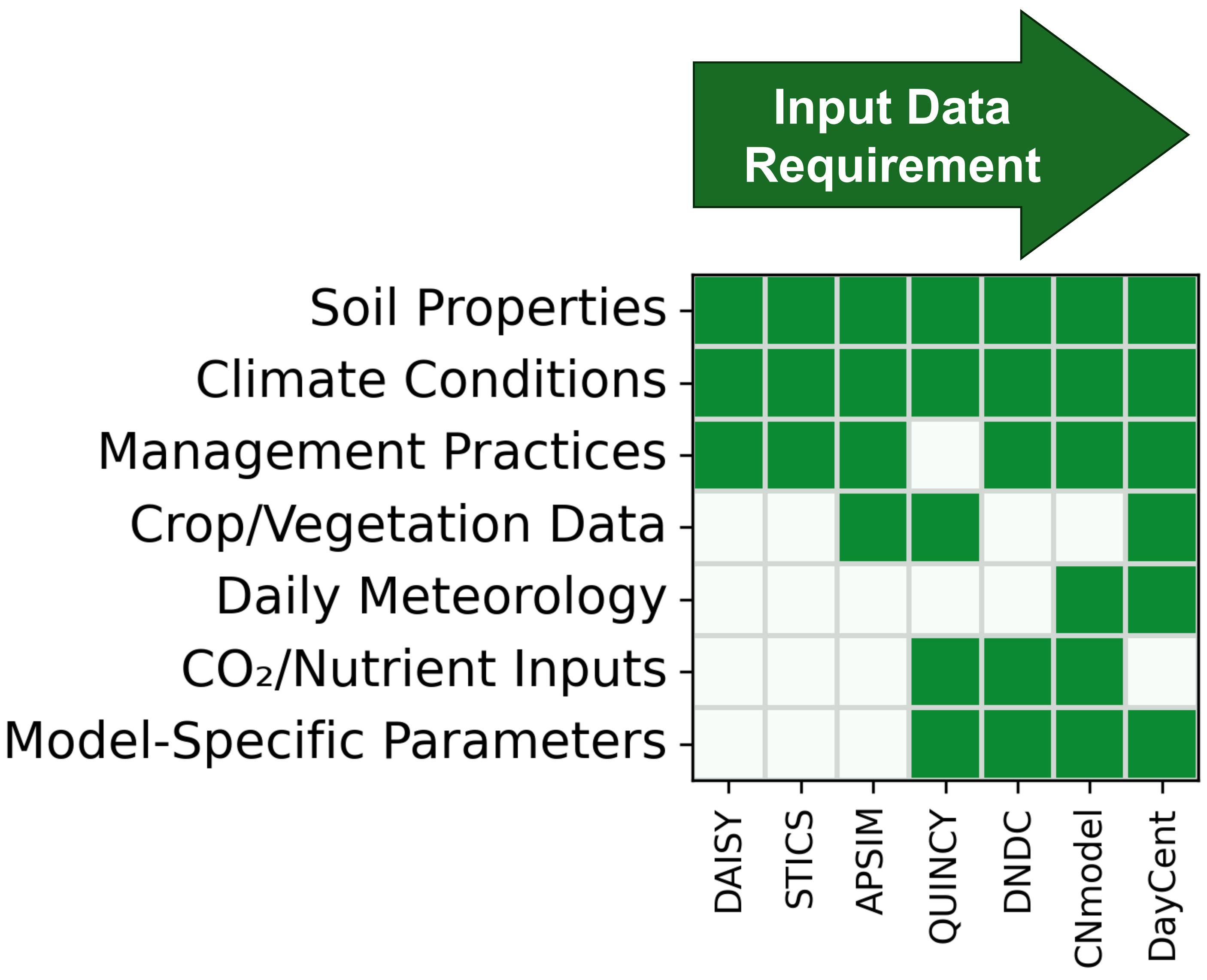
For data-scarce regions:
- Use DNDC or APSIM (simpler inputs, broad applicability).

For hotspot analysis:
- Prioritize DayCent (denitrification) or STICS (leaching).

For seasonal dynamics:
- CNmodel or DAISY (daily meteorology required).

CNmodel uniquely requires CO_2/N-deposition data and daily meteorology, making it the most data-intensive.
DNDC and QUINCY need nutrient inputs (e.g., fertilizer, manure) to simulate N_2O hotspots.
APSIM and STICS prioritize soil properties and management practices for crop-specific simulations.





Conclusion

- **Data-driven models** show strong predictive potential; however, their application is constrained by limited data availability in many regions.
- **Process-based models** are often calibrated for temperate environments and may not be well-suited for data-scarce, tropical regions.
- We will consider to **evaluate the suitability and performance** of various process-based models for simulating nitrogen fluxes in **sub-Saharan Africa**, with a focus on their adaptability to local conditions.

Swiss National Science Foundation
Grant Number: 200021 207348

muhammad.tufail@unibe.ch