Microsoft Research
Summit 2022

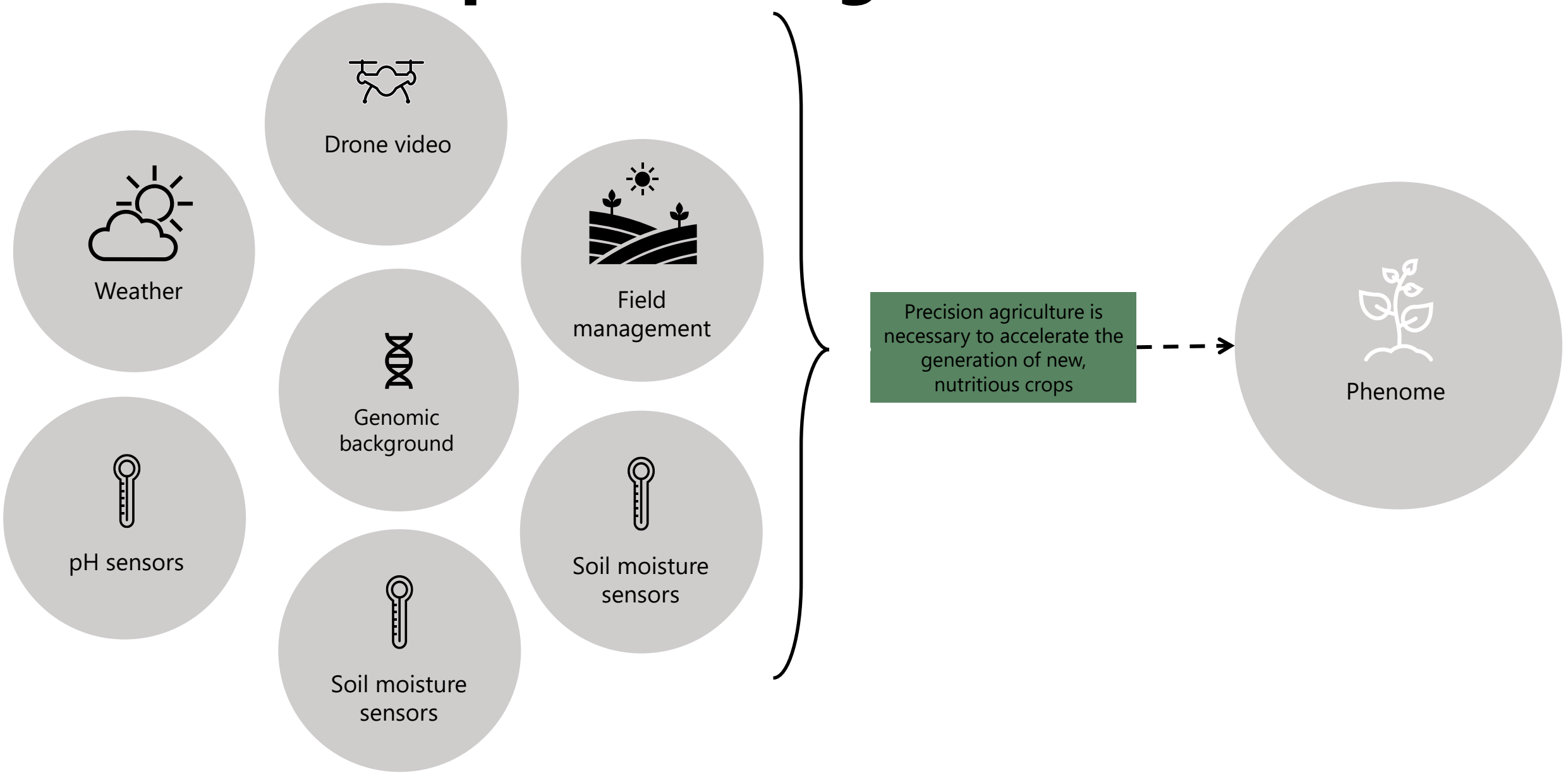
Leveraging AI to improve crop production

Swati Sharma (swatisharma@microsoft.com)
Microsoft Research

Microsoft Research
Summit 2022

DeepG2P: Multi-modality fusion for yield estimation

Role of precision agriculture



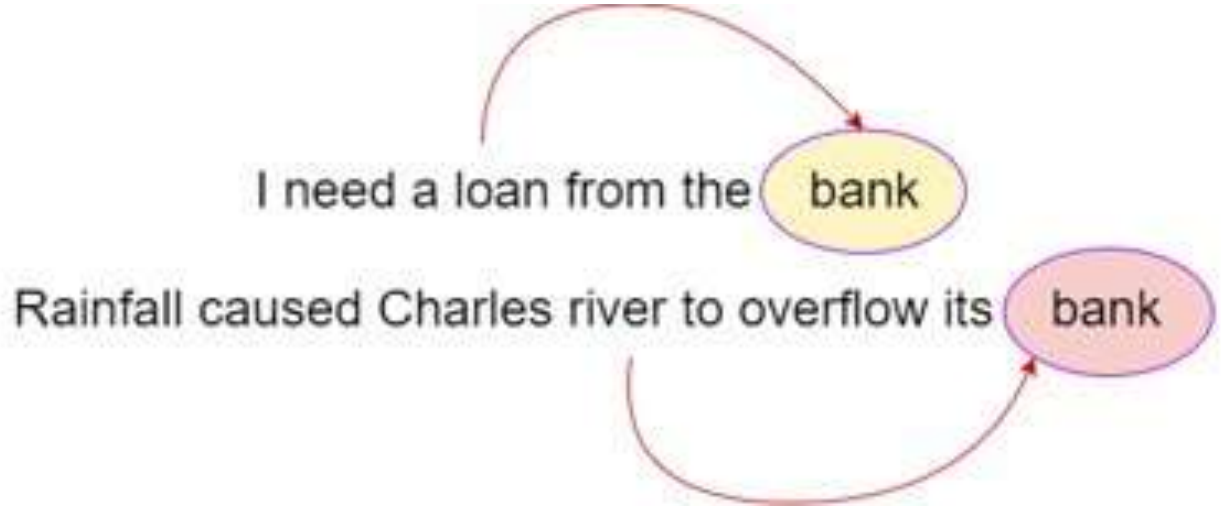
DeepG2P: A deep learning approach for yield prediction

G (Genome) + E (Environment) + M (Field Management) = P (Phenotype)

Representing genomic data

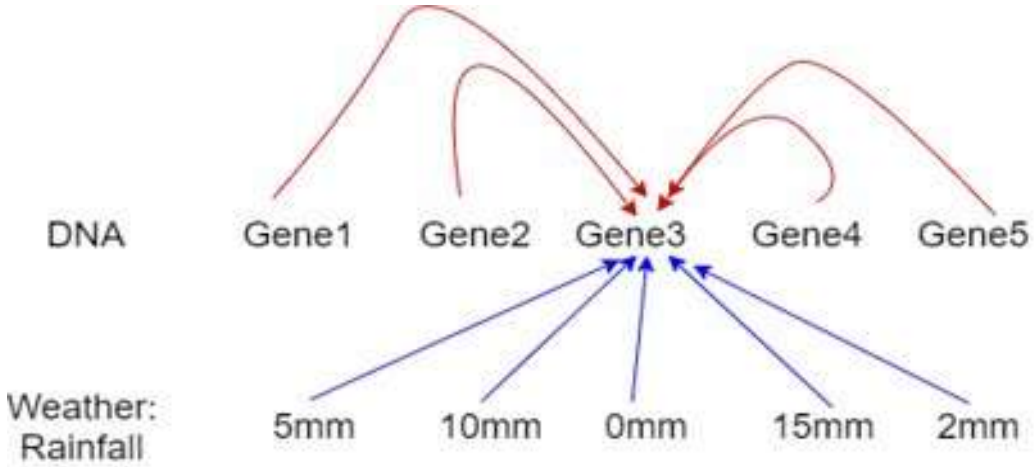
DNA is a natural language!

Genome is a sentence with genes/DNA sequences as words

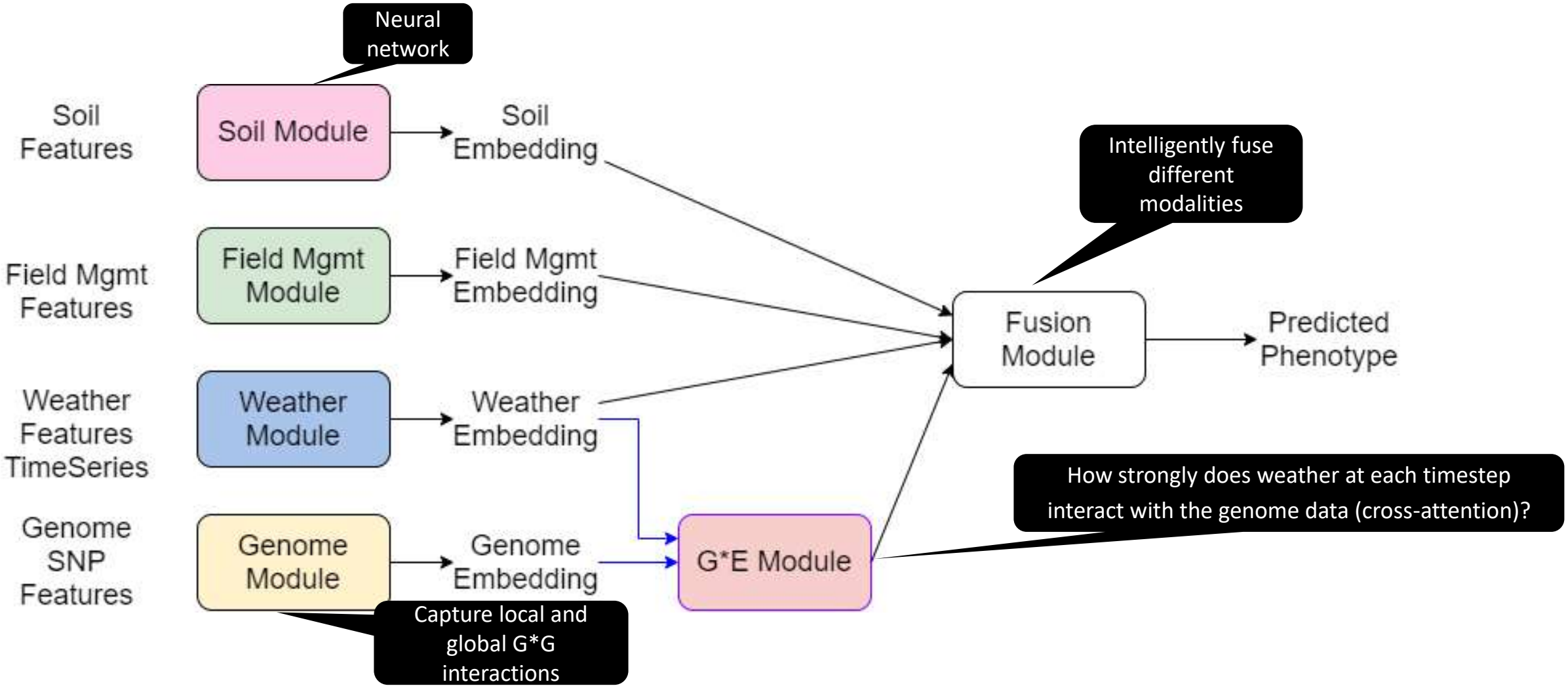


DNA as a natural language

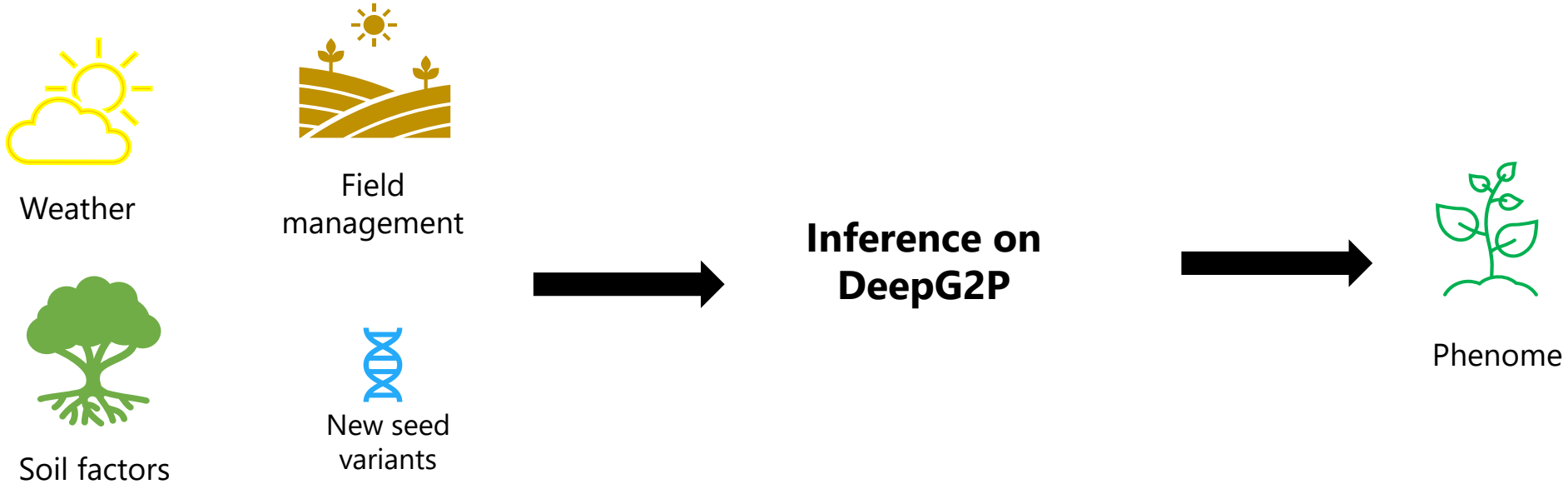
- **Vocabulary:** Genes/DNA sequences
- **Context:**
  - Neighborhood of a sequence
  - Weather, soil, field management factors add additional context to the DNA sequence /sentence.
- **Meaning:** Phenotype is the 'meaning' of the sentence



DeepG2P Framework



Applications



- **Automate discovery of new crops through precision agriculture**
  - ***Which seed variety is drought resistant?***
  - ***Which seed variety is resistant to a disease(s)?***

- **Determine the best management techniques to apply**
  - ***After planting, decide whether to irrigate or not, based on weather forecast and soil condition?***
  - ***Compare fertilizers for a seed variety and soil composition***

Microsoft

Microsoft Research
Summit 2022

Causal Modeling of Soil Processes for Improved Generalization

How important is soil carbon?

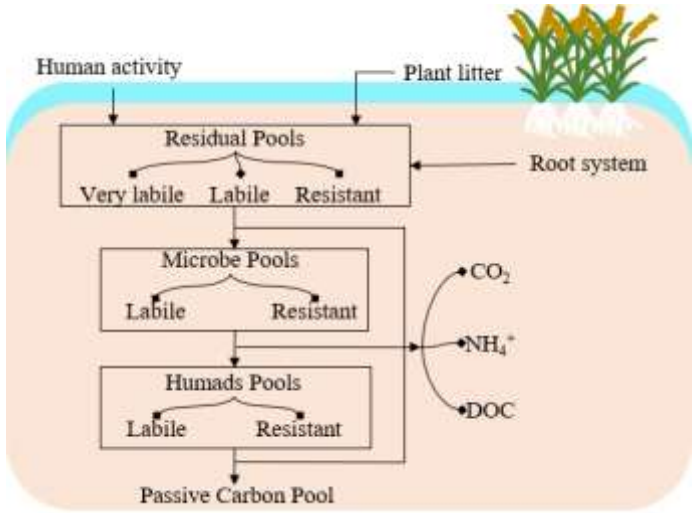
Soil health

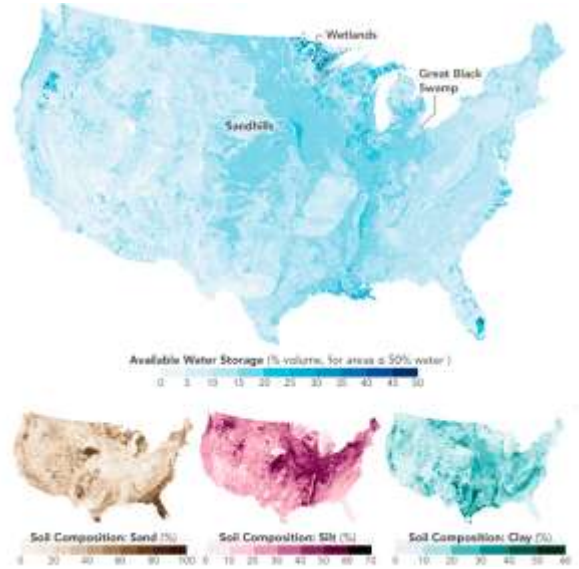
Co-benefits

Improve agricultural productivity

Enhance climate resilience

Challenges



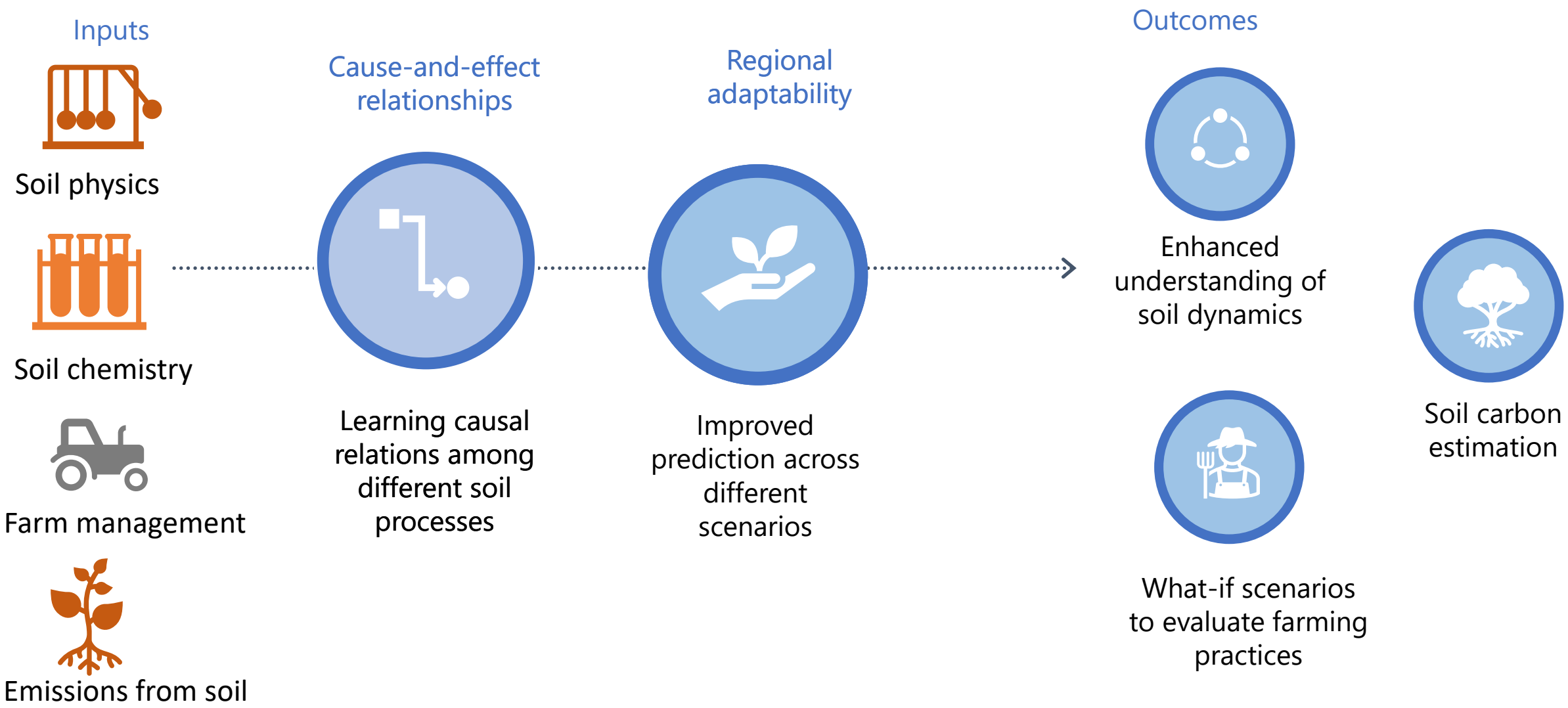


Complex “living” medium

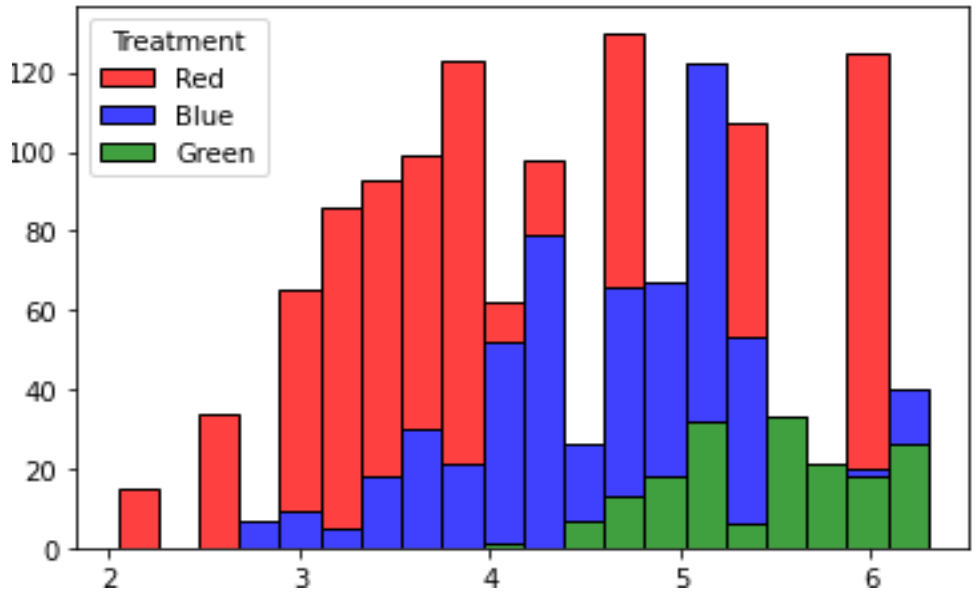
No universal model

Different environments

Goal


Inputs
Soil physics
Soil chemistry
Farm management
Emissions from soil
Cause-and-effect relationships
Learning causal relations among different soil processes
Regional adaptability
Improved prediction across different scenarios
Outcomes
Enhanced understanding of soil dynamics
What-if scenarios to evaluate farming practices
Soil carbon estimation

How does causal ML approaches help soil carbon modeling ?



Soil organic carbon distribution

Data from: North Wyke farm platform, Rothamsted Research, UK

ML approaches do not work well on OOD test data

Farms either received the red, blue or green treatment

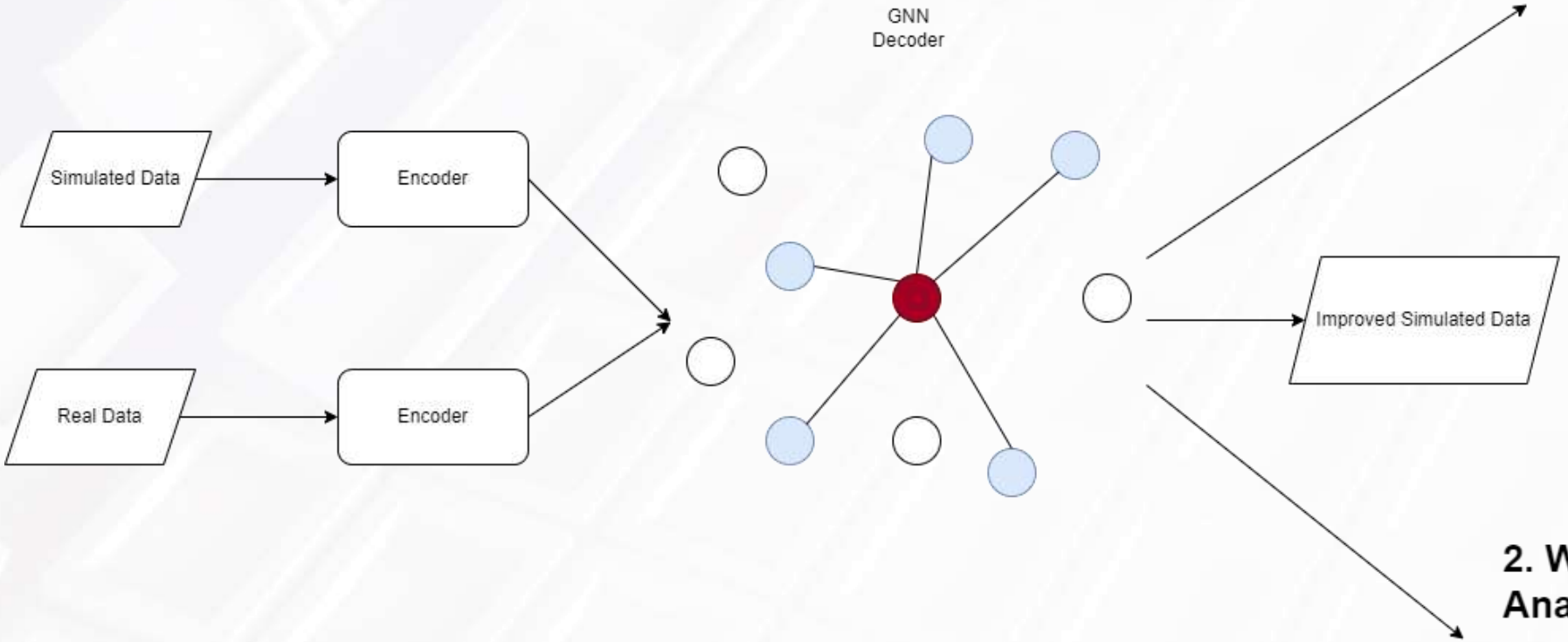
Red: High sugar grasses

Blue: High sugar grass-white clover mix

Green: Permanent Pasture

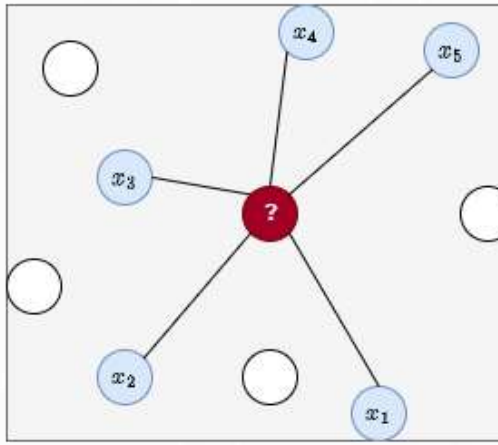
Causal ML approaches generalize better than ML approaches

Knowledge Guided Representation Learning



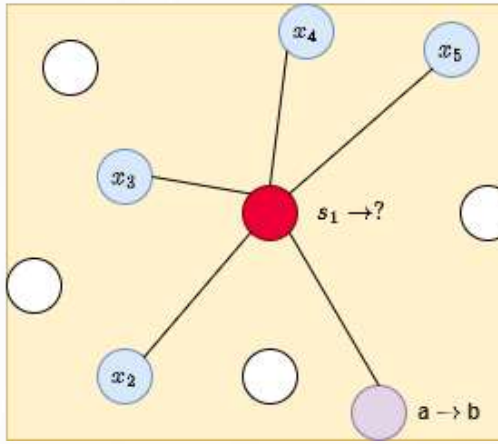
1. Soil Carbon Prediction

Causal Graph Neural Network



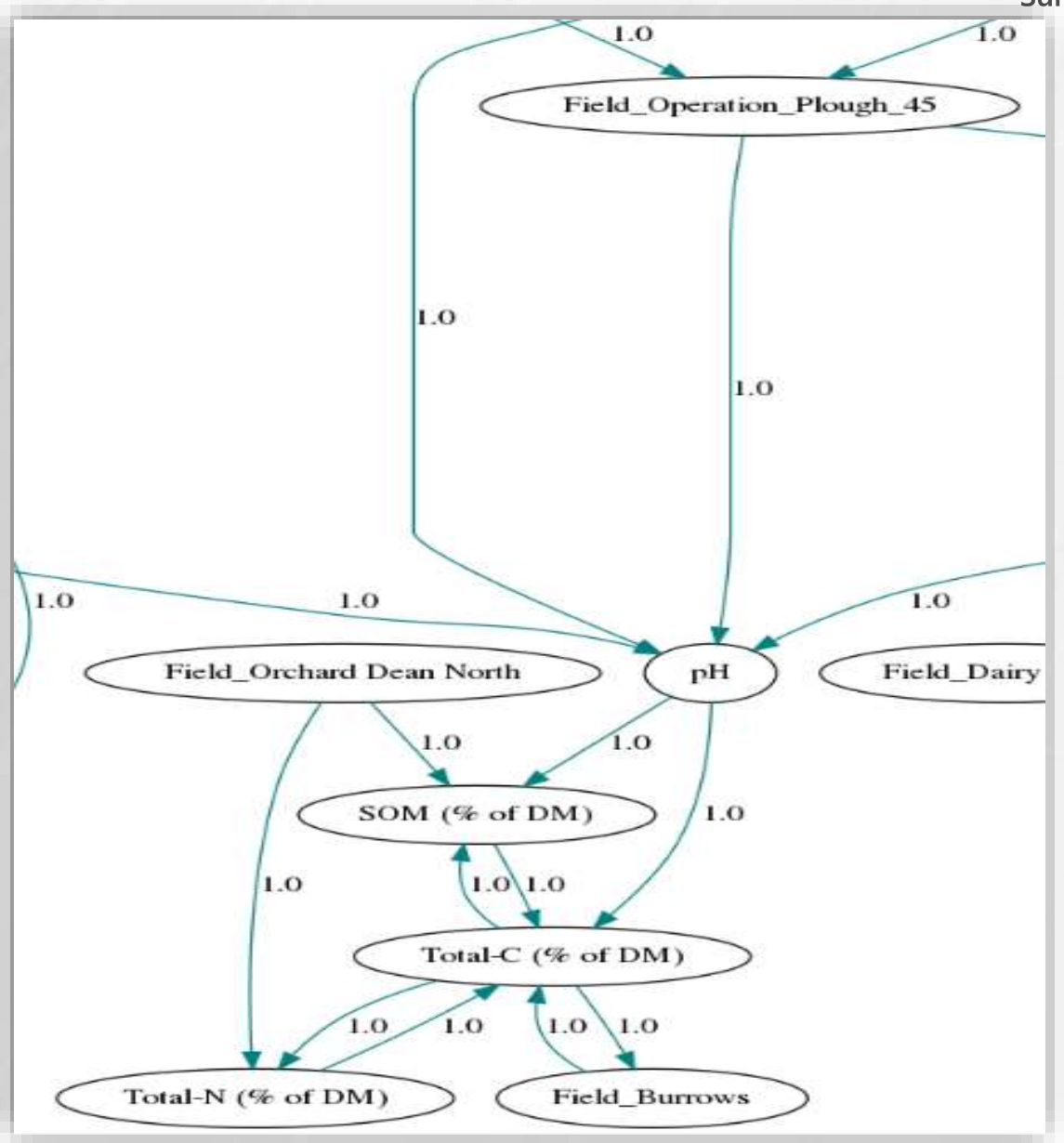
2. What-if Analysis

SCM based Causal Inference



Causal graphs discover some well-known links

- Relationship between carbon and nitrogen in soil
- Ploughing/Tilling changes pH of the soil that affects soil organic carbon and total carbon, nitrogen
- Iron, Zinc, Manganese affects total carbon



Applications

Provide farmers or food companies insight about their farms given the local conditions

I am growing wheat on clay soil. Do I need a nitrogen-based fertilizer?

Does shallow tilling lower a farm's soil organic carbon?

Given a farm's soil conditions, weather and management practices, what is the soil organic carbon content?

Does adding a cover crop affect the nitrogen and carbon in the soil?

Microsoft Research
Summit 2022

Thank you

Stay in touch:

@Username

www.Website.com