KELLY CAYLOR
BREN SCHOOL OF ENV. SCIENCE & MANAGEMENT
DEPARTMENT OF GEOGRAPHY
UNIVERSITY OF CALIFORNIA, SANTA BARBARA

LYNDON ESTES
GRADUATE SCHOOL OF GEOGRAPHY
CLARK UNIVERSITY

A MILE WIDE AND A PIXEL DEEP: INTEGRATING MACHINE LEARNING, COMPUTER VISION, AND SATELLITE IMAGERY FOR COUPLED-NATURAL HUMAN SYSTEM MODELING

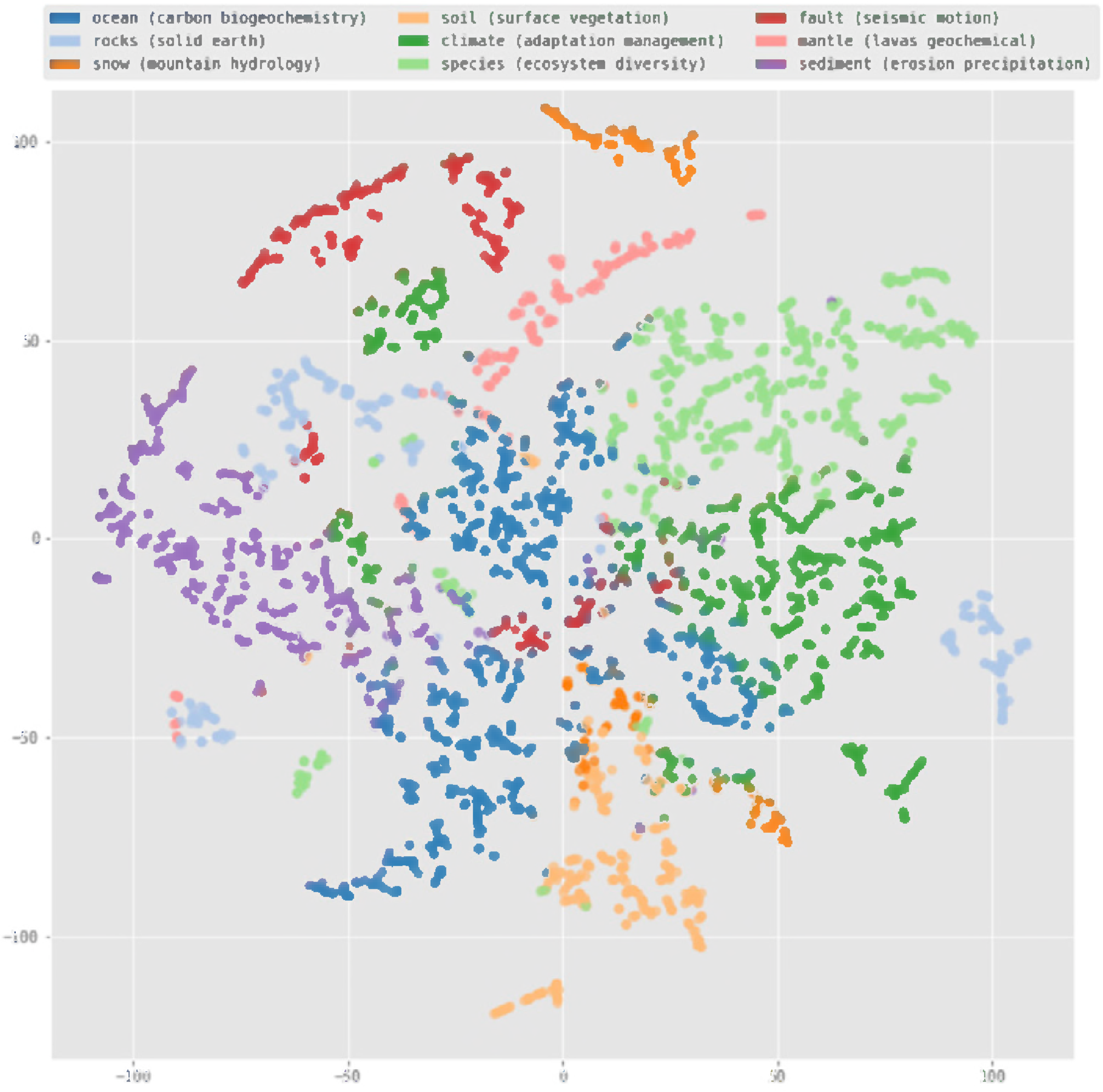
21 April 1889, Arkansas Gazette
'Nye Waxes Eloquent'

"The Platt River… has a <u>very large circulation, but very little influence</u>. It covers a good deal of ground, but is not deep. In some places it is a mile wide and three-quarters of an inch deep."

Edgar Wilson "Bill" Nye (1850 –1896)

Earth Research Institute at UCSB

Topics derived from Non-negative matrix factorization (NMF) of 3,770 research manuscript abstracts and titles. Plotted using stochastic neighbor embedding (t-SNE).



Patterns

CellPress OPEN ACCESS

Article

Mapping research topics at multiple levels of detail

Sara Lafia,[1,3,*] Werner Kuhn,[2] Kelly Caylor,[2] and Libby Hemphill[1]
[1]ICPSR, University of Michigan, Ann Arbor, MI 48109, USA
[2]Department of Geography, University of California, Santa Barbara, CA 93106, USA
[3]Lead contact
*Correspondence: slafia@umich.edu

https://doi.org/10.1016/j.patter.2021.100210

THE BIGGER PICTURE Research institutes and organizations are interested in communicating the impact of their work and its value to a broader audience. However, quantifying impact and providing high-level views of interdisciplinary research trends are challenging. To address this, we leverage distant reading methods from the digital humanities to model the topics of a large body of interdisciplinary research products and visualize them in maps. We analyze 3,770 academic publications and grants affiliated with an interdisciplinary earth science research institute over a 10-year period and model its research topics. We then map the topics at two distinct levels of detail and evaluate the interpretation of the maps through a survey of leading researchers. We show that the topic maps reveal insights including the emergence of interdisciplinary collaboration areas and evolving areas of expertise over time.

1 2 3 4 5 **Proof-of-Concept:** Data science output has been formulated, implemented, and tested for one domain/problem

Interactive Data Visualization: https://bit.ly/eri_research

Lyndon Estes,
Associate Professor
Clark University

Su Ye

Lei Song

Sitian Xiong

Boka Luo

Qi Zhang

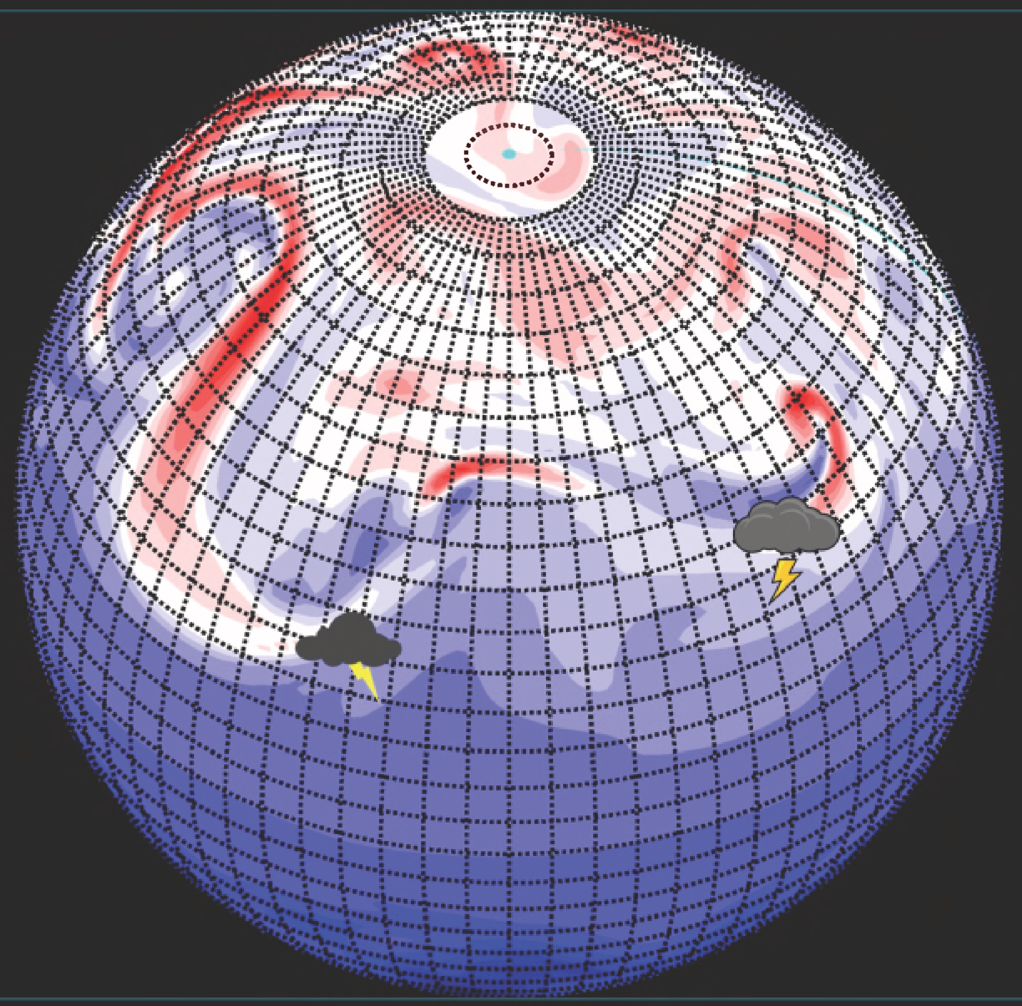
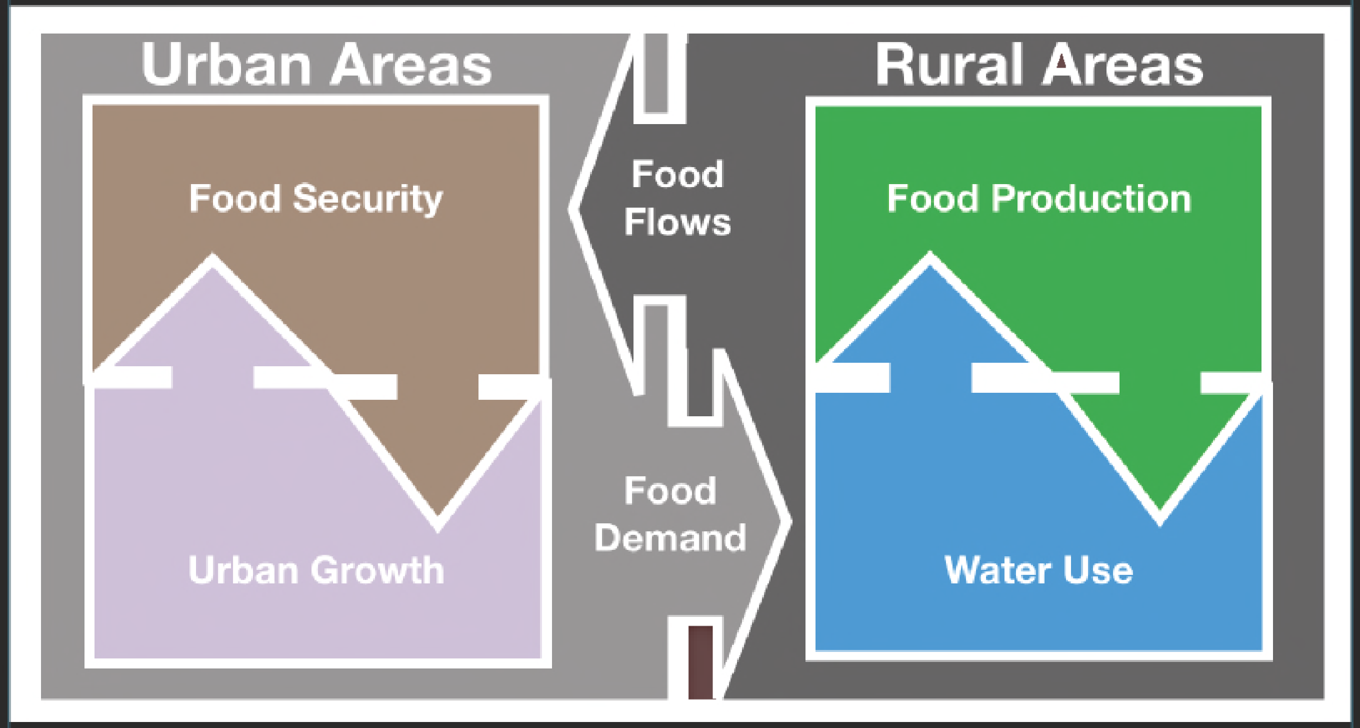
Ryan Avery
Development Seed

Stephanie Debats
Google

Fredrick Mwawaza, Angeline Amukoa, Jackson Makuru, Julius Mucha, Joseph Wanyoike, Francis Muinde, Purent Oduor, Charles Juma, Thuo Wanjiku, Adelide Mugami, Isaac Mutisya Muasa, Jeff Ochieng, Ben Mbatia

ERI CLARK UNIVERSITY NSF NATIONAL GEOGRAPHIC NASA planet. OMIDYAR NETWORK aws MERIDIA

EARTH SYSTEMS & COUPLED NATURAL-HUMAN SYSTEMS

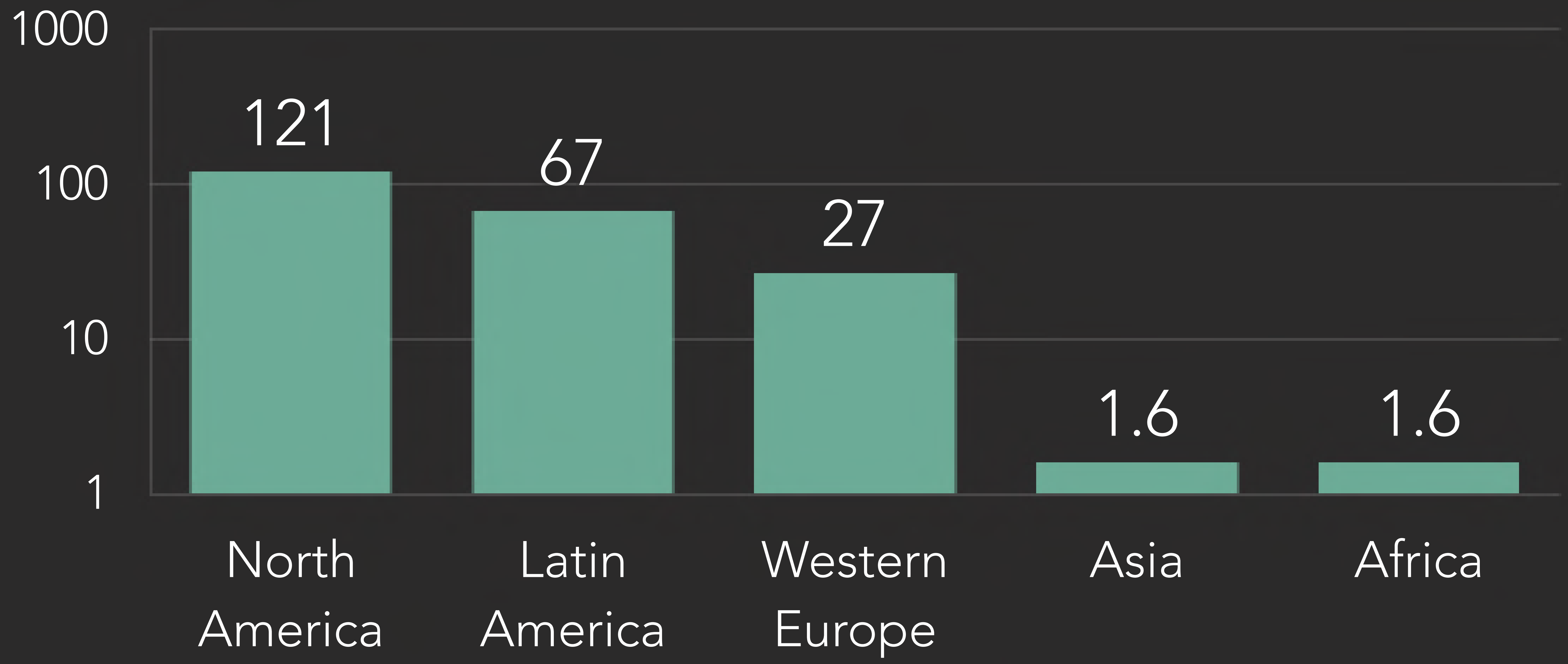
DOMAIN		MODELS	DATA	ML/AI PATTERN APPLICATIONS	ML/AI PROCESS APPLICATIONS
EARTH SYSTEMS		Numerical Physics-Based	Big > Small	Pattern Emulation	Process Representation
COUPLED NATURAL-HUMAN SYSTEMS		Conceptual Empirically-Derived	Small > Big	Pattern Detection	Process Estimation

Smallholder Agriculture as a Coupled Natural-Human System

2.5 billion smallholders managing
500 million small farms worldwide

Family farms constitute over 98%
of all farms

Average farm size (ha) of selected world regions

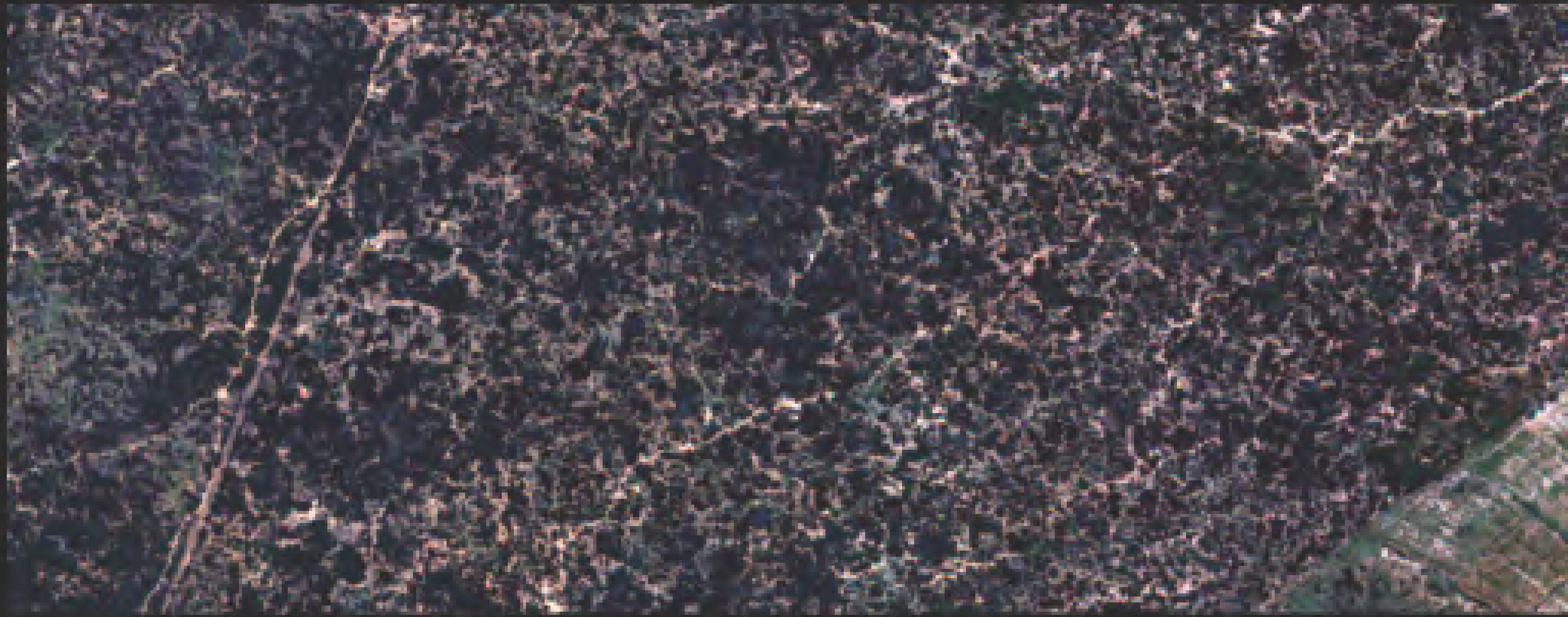


Food demand to triple by 2050

At least **140 m ha new cropland**

1/3 of world's potential cropland

Searchinger, Estes et al (2015)

WHERE IS SMALLHOLDER AGRICULTURE **OCCURRING**?


HOW IS SMALLHOLDER AGRICULTURE **CHANGING**?

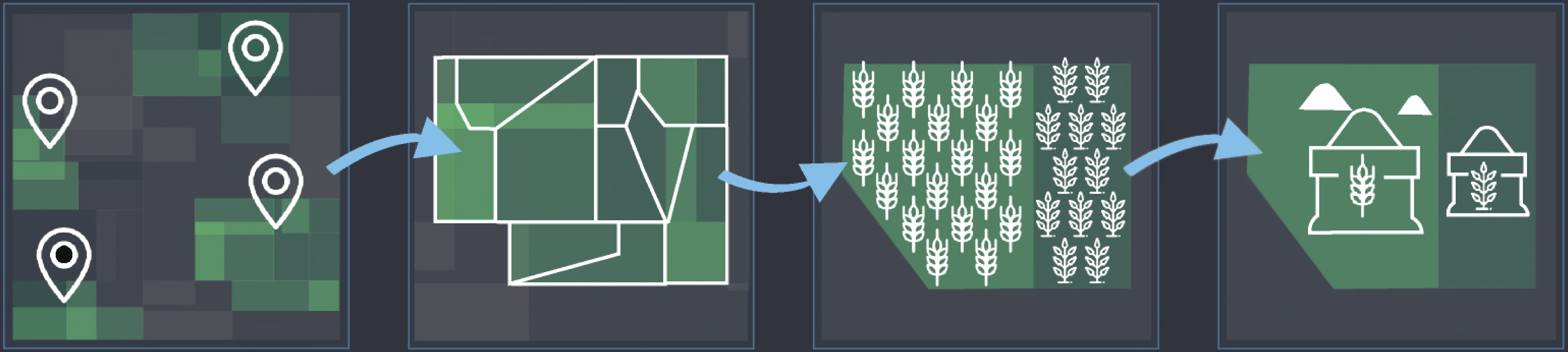
Understanding agricultural pattern and process depends on an inter-dependent series of datasets

Where is Agriculture Occurring?

How are Fields Being Managed?

Which Crops are Being Grown?

What Yields are Being Obtained?



Agricultural clearing in Madagascar

Moderate (100-250m) and High Resolution (10-60m) Satellite Data are routinely used to captured large-scale landcover patterns and transitions

Google Earth Timelapse (Google, Landsat, Copernicus)

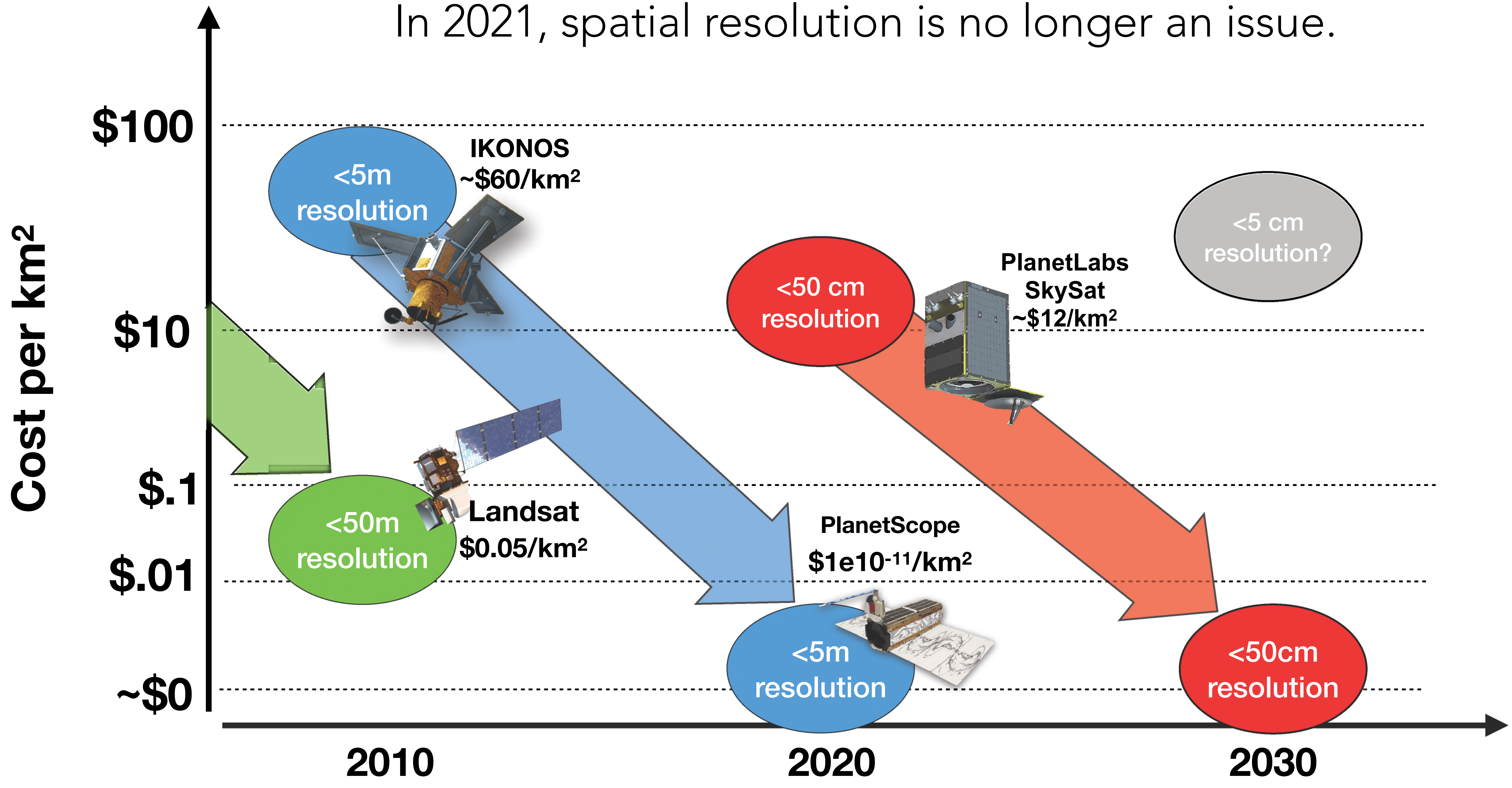
Barriers to the use of ML/AI in Smallholder Agriculture

01

SPATIAL/TEMPORAL IMAGERY

EVERY 10 YEARS, THE SPATIAL RESOLUTION OF ESSENTIALLY FREE IMAGERY INCREASES BY 10X

In 2021, spatial resolution is no longer an issue.

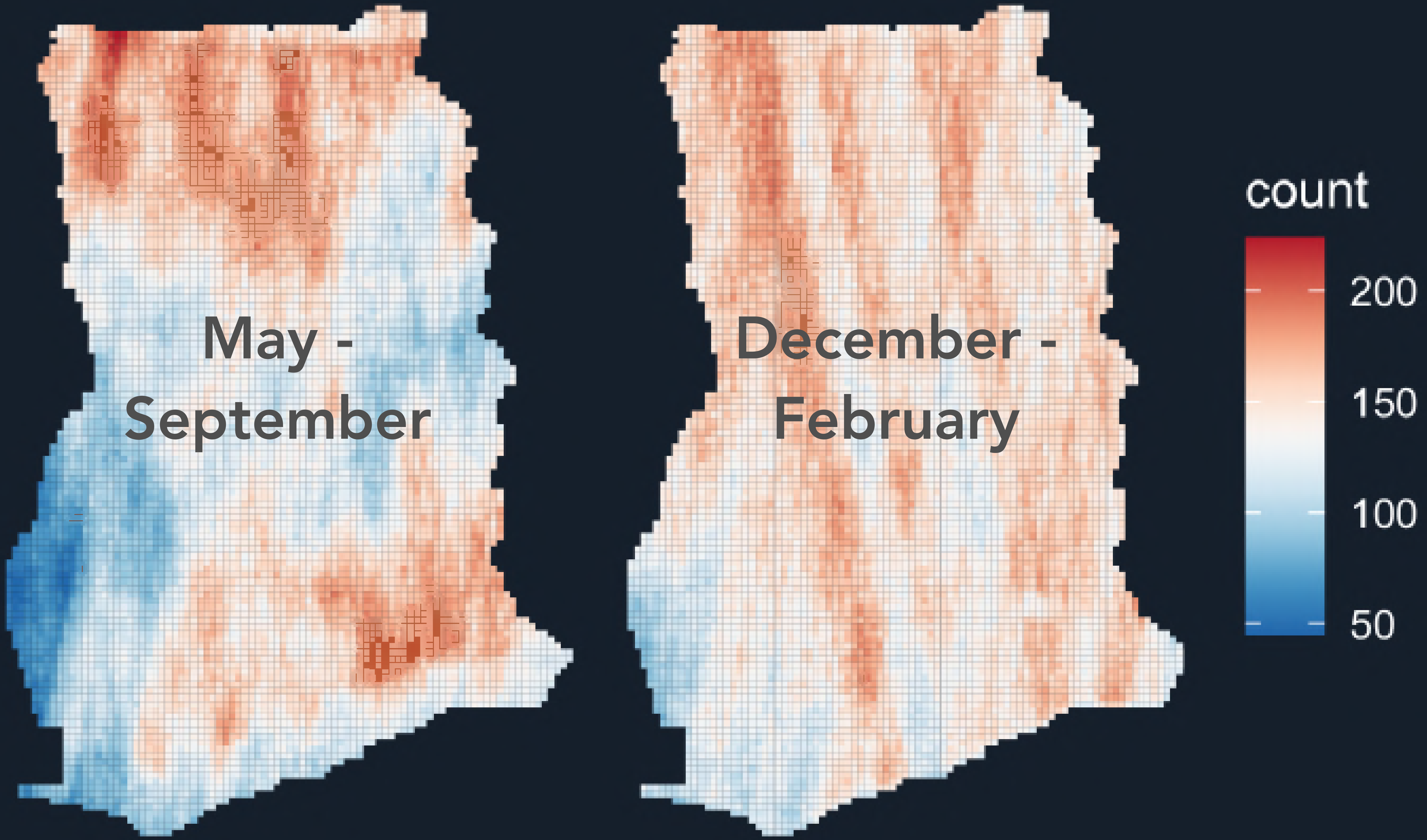


01

SPATIAL/TEMPORAL IMAGERY

Characterizing temporal dynamics remains challenging with optical sensors

PlanetScope Seasonal Availability of Imagery Ghana, May 2018 - February 2019



02 DEVELOPING TRAINING DATA

TRADITIONAL TRAINING DATA AREN'T NECESSARILY SUITABLE FOR REMOTE SENSING

ImageNet Sample Images

02 DEVELOPING TRAINING DATA

Most labels are image classifications or object instances; very few segmentation labels for Environmental Systems

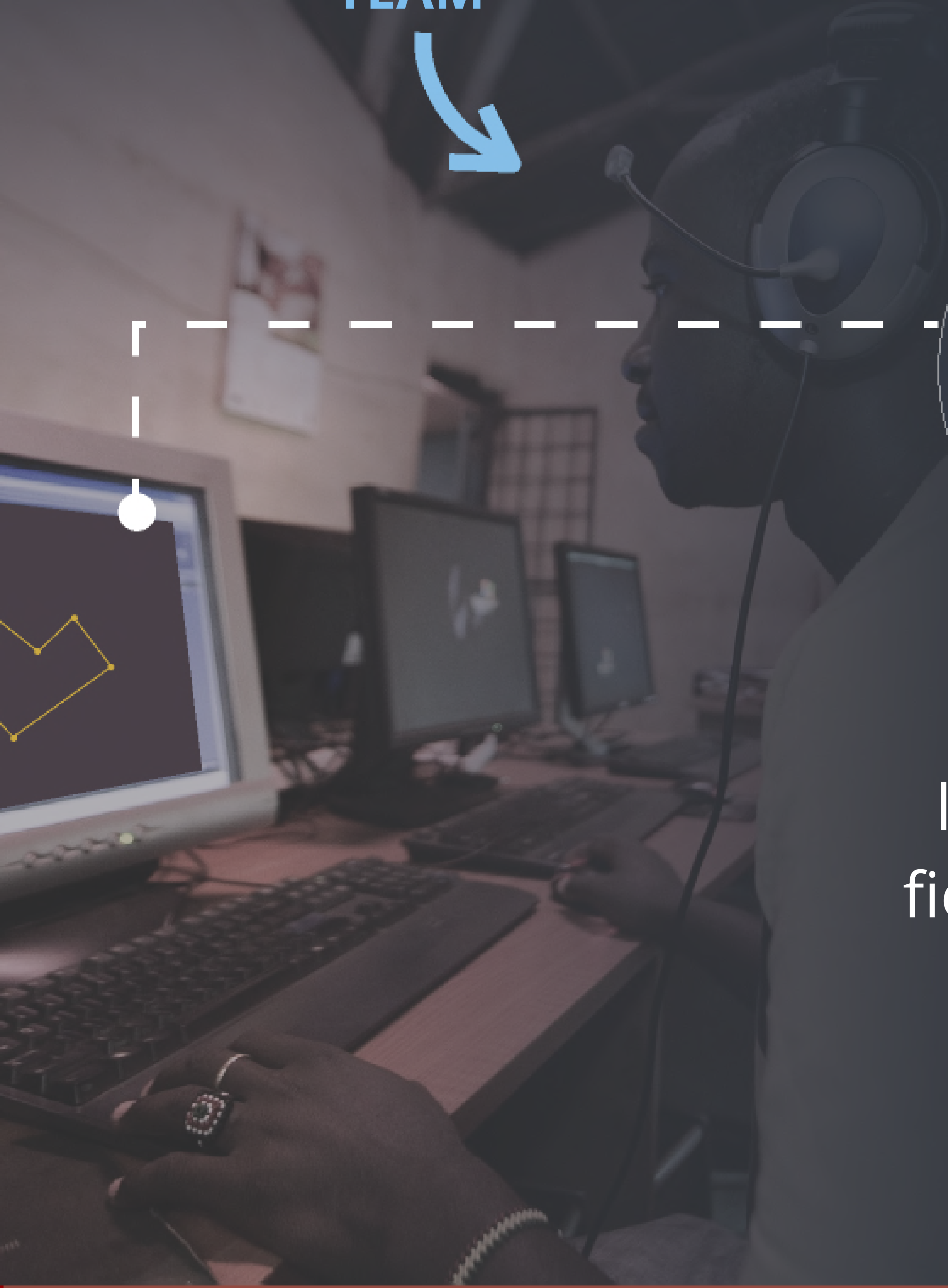
ISPRS Semantic Segmentation Dataset

Potsdam , Germany

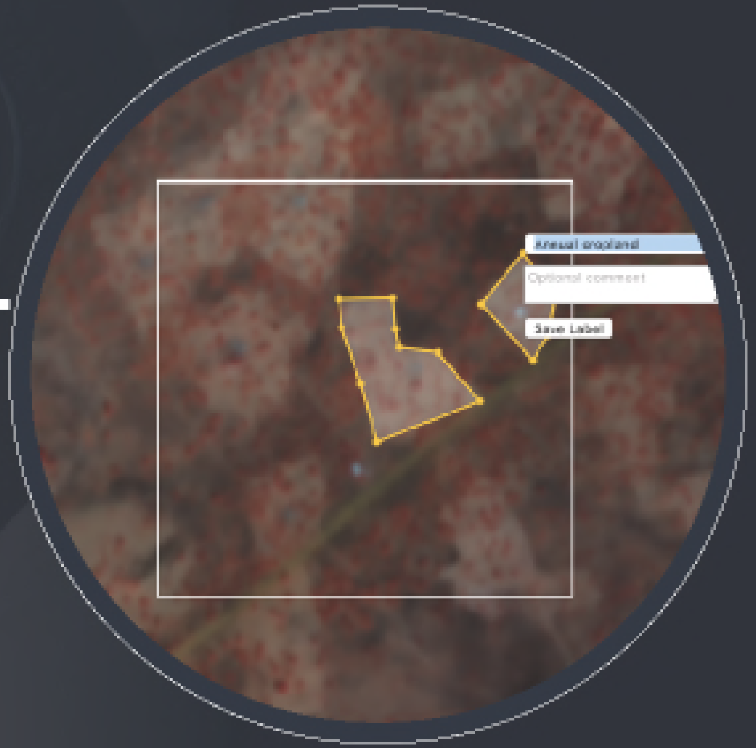
02 DEVELOPING TRAINING DATA

DEVELOPING TRAINING DATA FOR SMALLHOLDER AGRICULTURE

LABELLING TEAM



Cropland Delineation



Trained map labellers trace field boundaries in imagery

Consensus Labelling

Several people's labels are combined into a consensus label to reduce individual mapping error

More Accurate Training Data

02 DEVELOPING TRAINING DATA

INSTANCE LABELLING IS HARD

02 DEVELOPING TRAINING DATA

INSTANCE LABELLING IS HARD

Elmes, A.; Alemohammad, H.; Avery, R.; Caylor, K.; Eastman, J.R.; Fishgold, L.; Friedl, M.A.; Jain, M.; Kohli, D.; Laso Bayas, J.C.; Lunga, D.; McCarty, J.L.; Pontius, R.G.; Reinmann, A.B.; Rogan, J.; Song, L.; Stoynova, H.; Ye, S.; Yi, Z.-F.; Estes, L. Accounting for Training Data Error in Machine Learning Applied to Earth Observations. *Remote Sens.* **2020**, *12*, 1034. https://doi.org/10.3390/rs12061034


02
DEVELOPING TRAINING DATA
LABELS AREN'T STATIONARY
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
-16.867407°, 26.901499°


02
DEVELOPING TRAINING DATA
LABELS AREN'T STATIONARY
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
-16.867407°, 26.901499°

02 DEVELOPING TRAINING DATA

Labels aren't Stationary

02 DEVELOPING TRAINING DATA

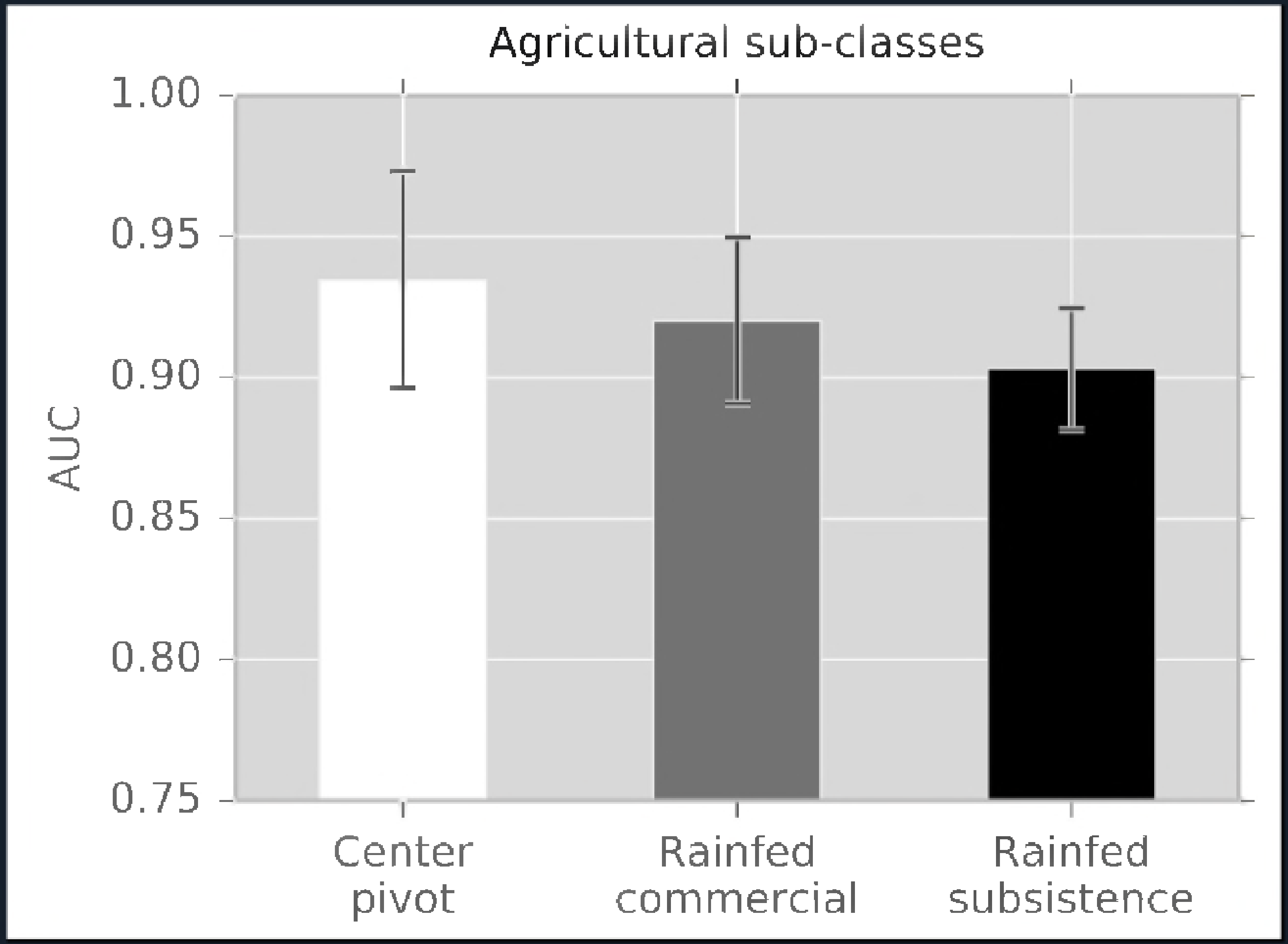
LABELS AREN'T STATIONARY

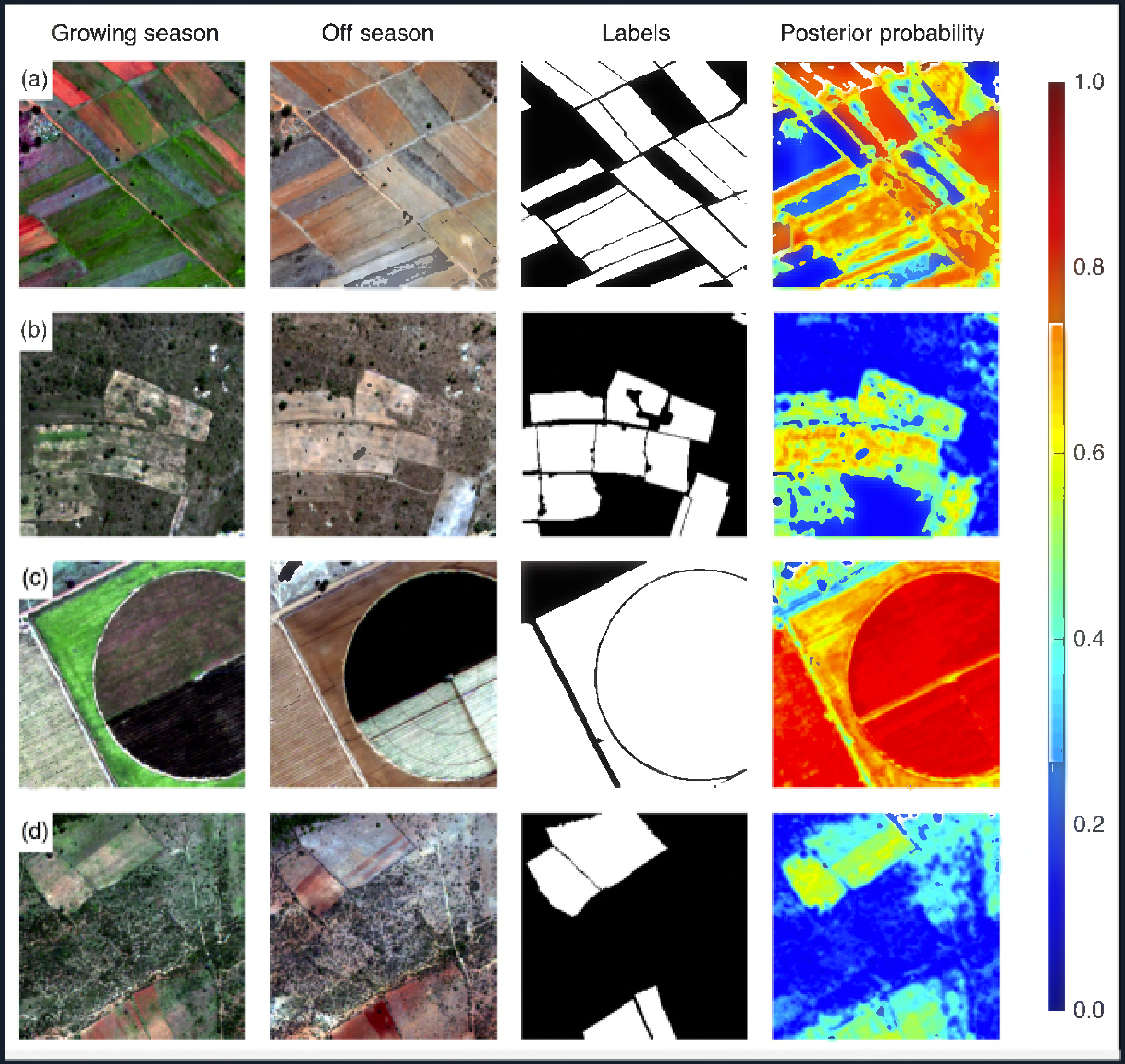
03

EVOLVING ML/AI ALGORITHMS

MACHINE LEARNING CAN CHARACTERIZE SMALL-SCALE PATTERNS

High-accuracy, 1-m random forest classifier across agricultural types

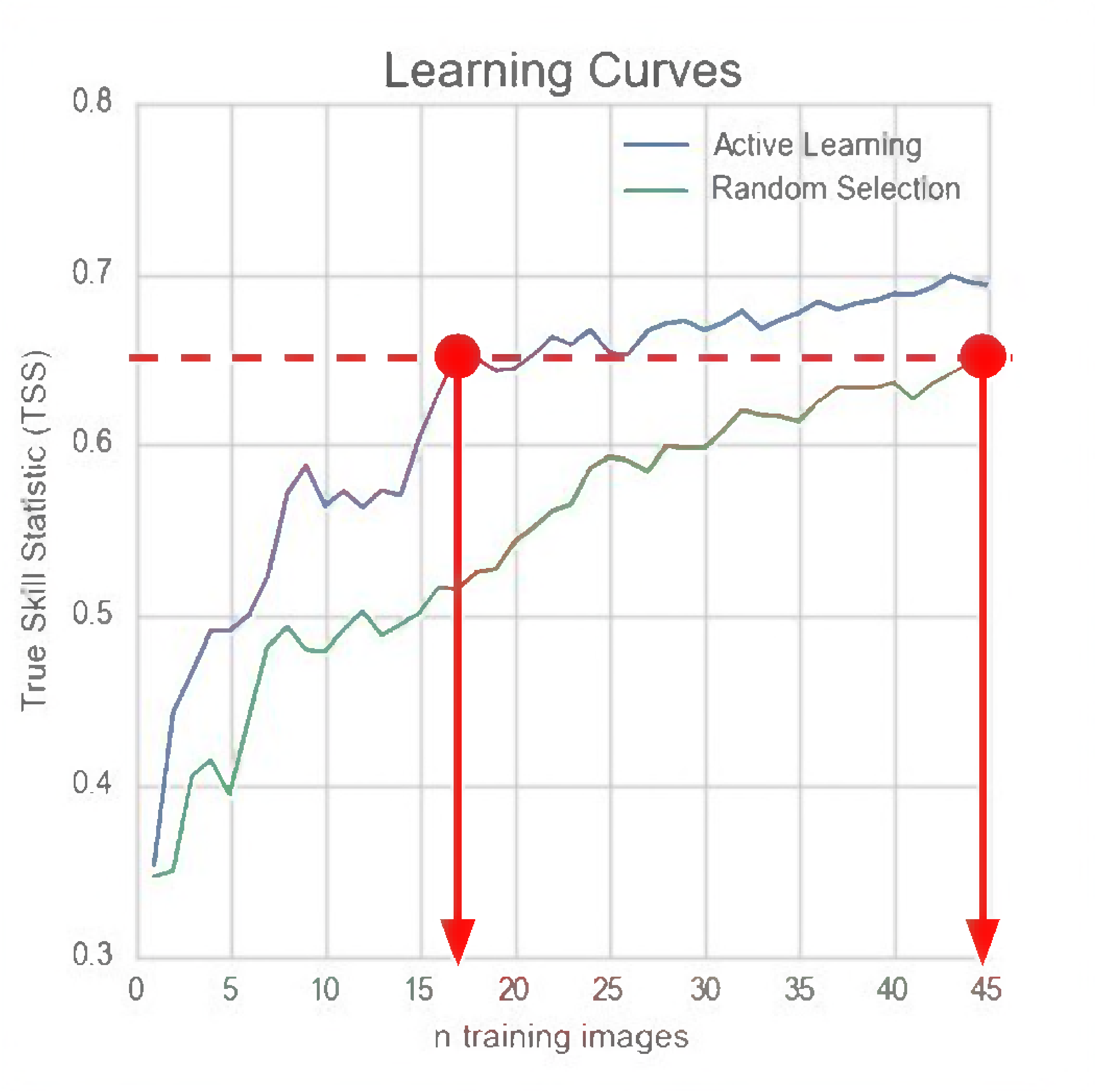




*Debats, S., +Luo, D., *Estes, L.D., Fuchs, T, **Caylor, K.K.** (2016) “A generalized computer vision approach to mapping agricultural fields in heterogeneous landscapes”. Remote Sensing of Environment, doi:10.1016/j.rse.2016.03.010.

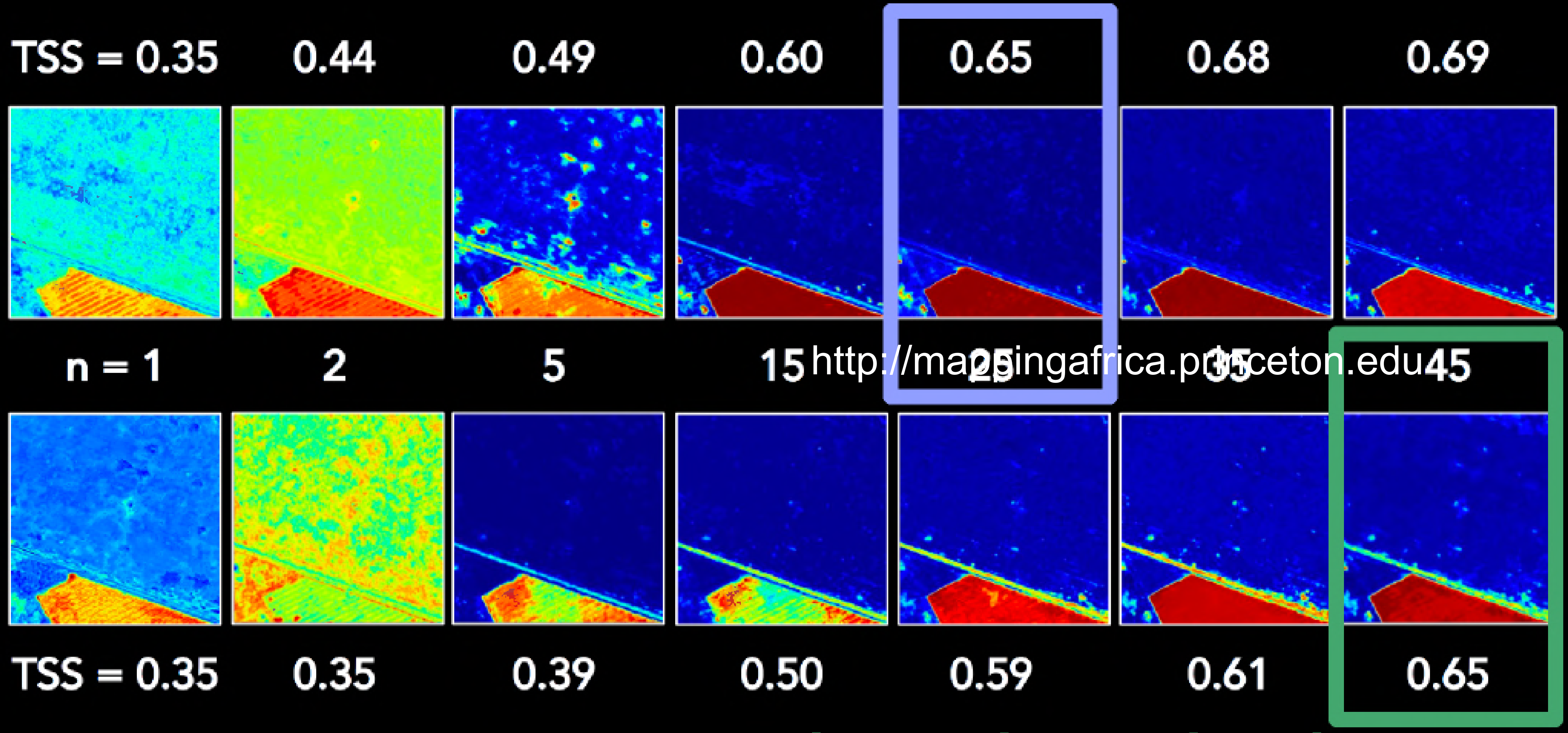
03 EVOLVING ML/AI ALGORITHMS

ACTIVE LEARNING ACCELERATES MODEL DEVELOPMENT, ACCURACY CEILING IS LOW



Estes, L.D., Yee, S., Song, L., Avery, R.B., McRitchie, D., Eastman, R., Debats, S.R., Caylor, K.K. "Improving maps of smallholder-dominated croplands through tight integration of human and machine intelligence." AGU Fall Meeting Abstracts, vol. 2019, pp. IN42A-04. 2019.

Training Images Selected via Active Learning

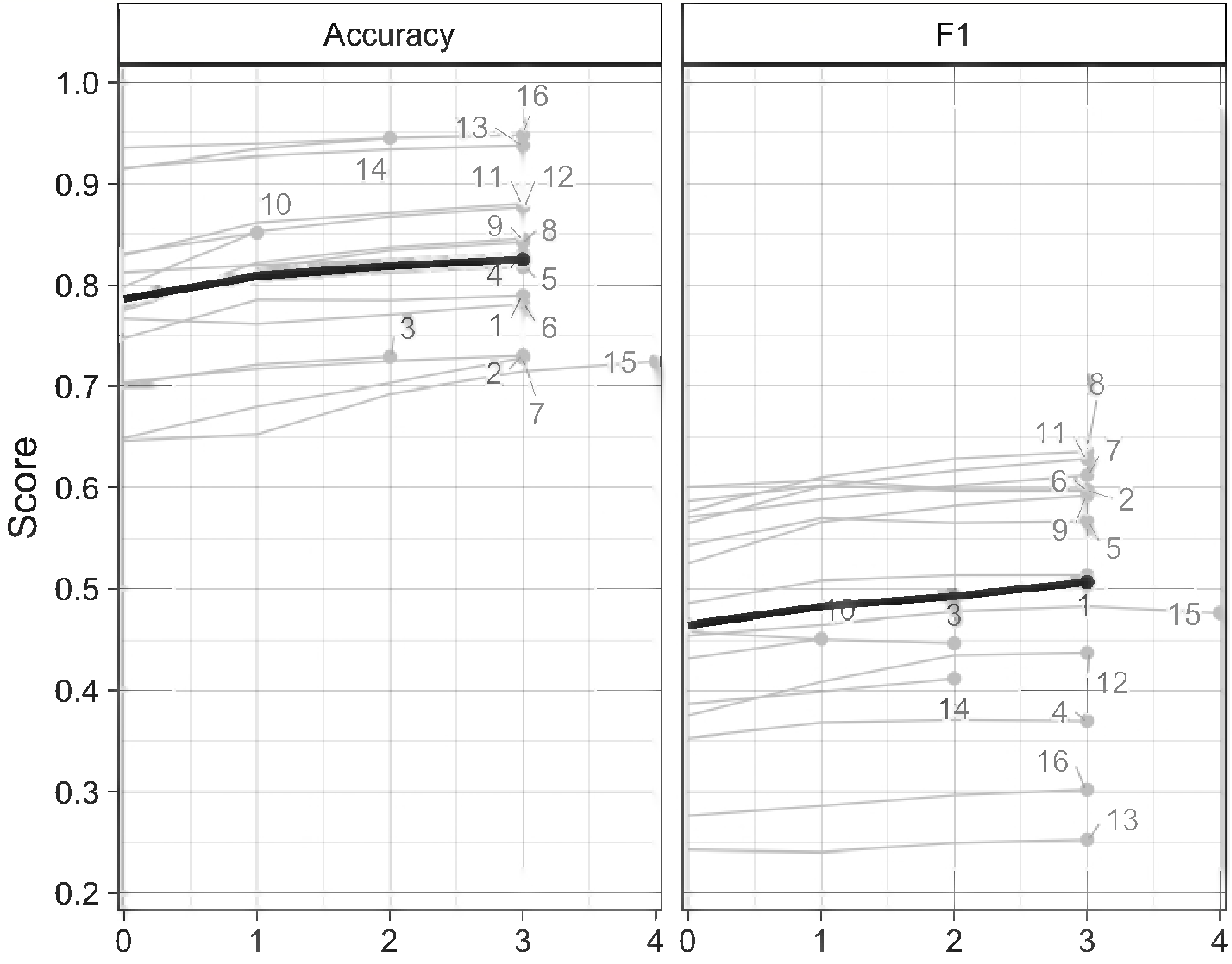


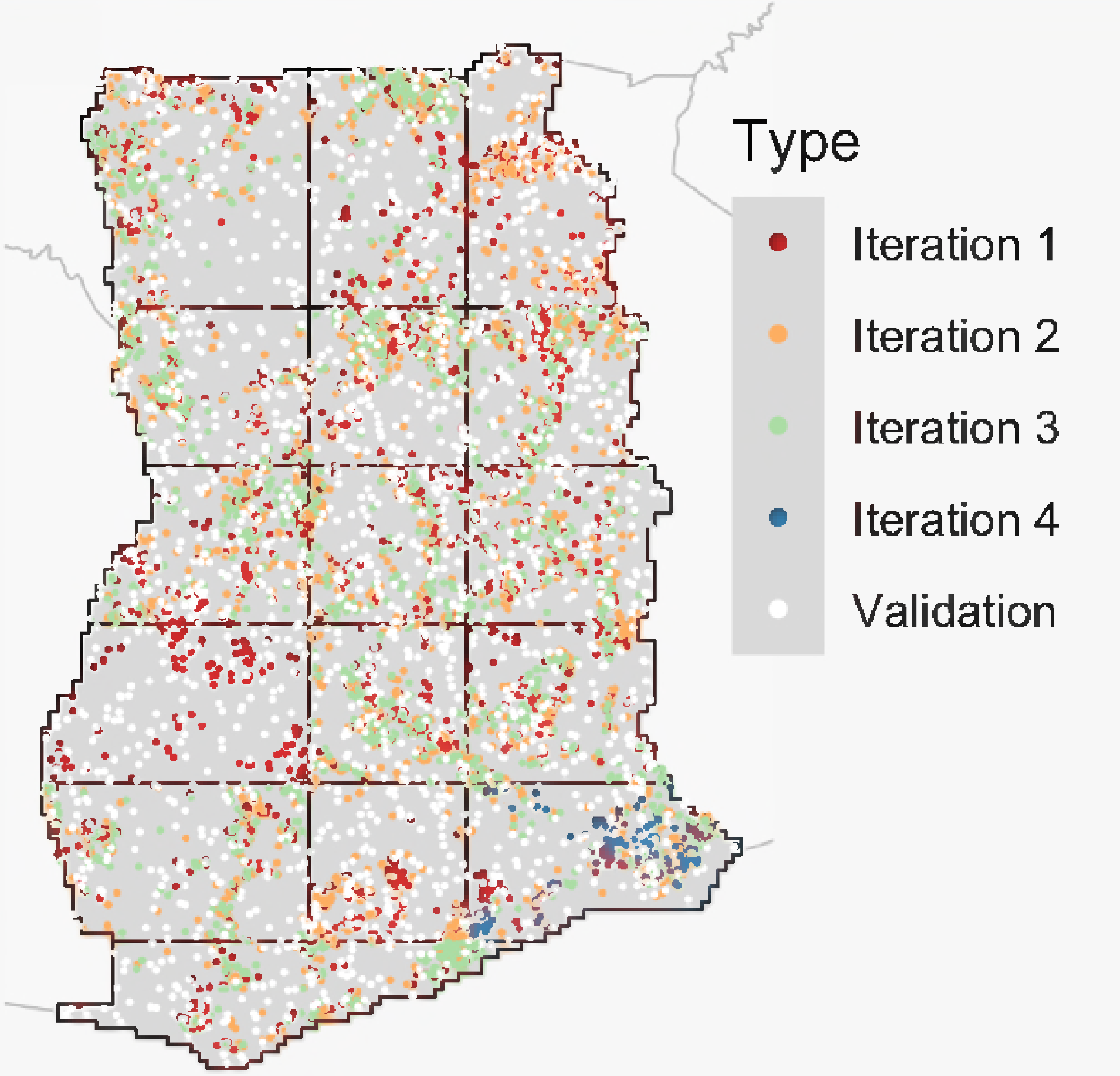
Training Images Selected Randomly

03

EVOLVING ML/AI ALGORITHMS

GENERALIZATION OF ML MODELS TO LARGE-SCALE APPLICATIONS IS CHALLENGING





03 EVOLVING ML/AI ALGORITHMS

Deep Learning Models resolve patterns more distinctly and accurately

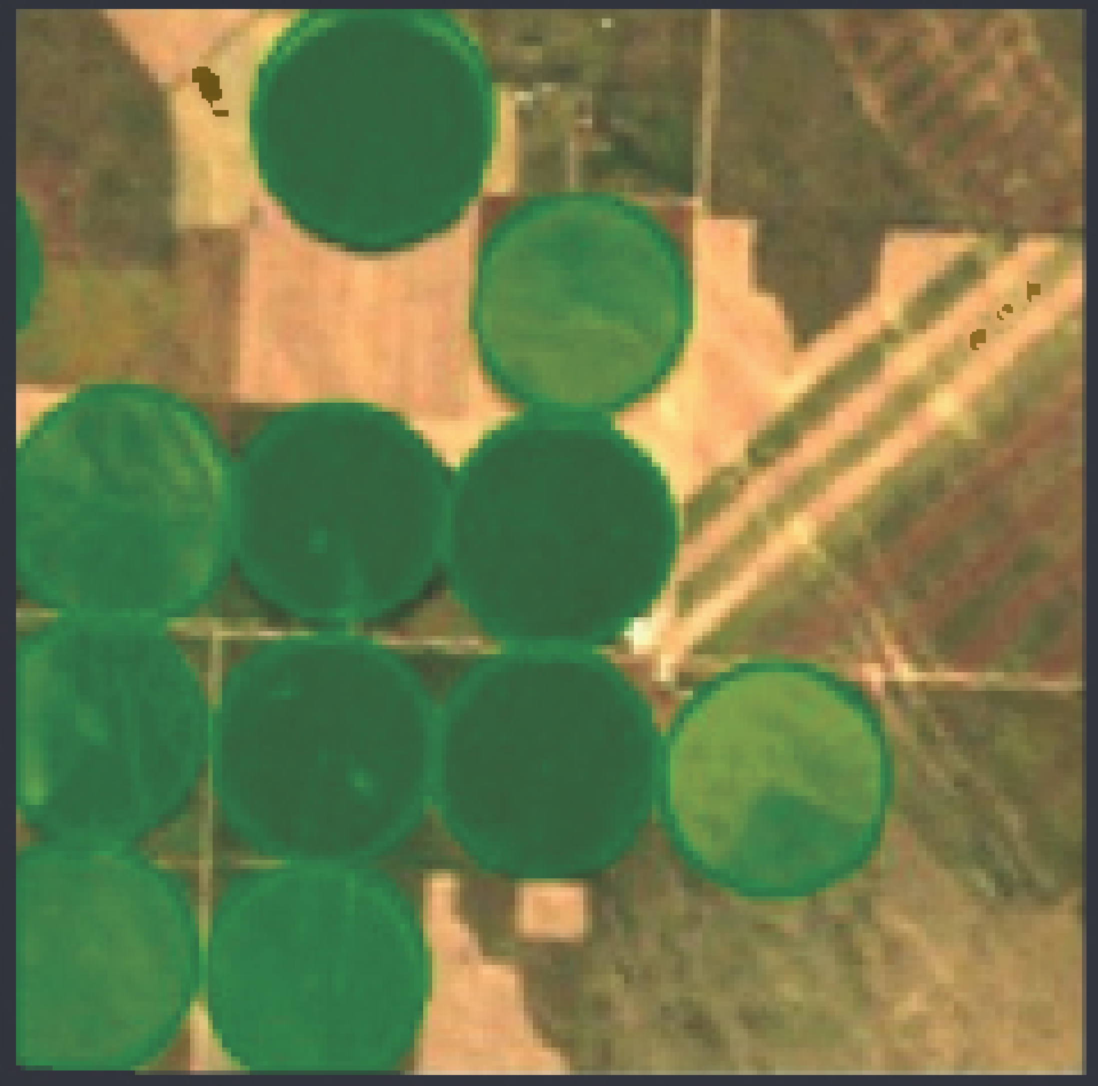
PlanetScope composite | RandomForest | U-Net (semantic segmentation)

Model	TSS	Accuracy	Precision	Recall	FPR	TPR	AUC
Random Forest	0.67	0.79	0.37	0.89	0.22	0.89	0.90
U-Net (Balanced Dice + Cross Entropy)	0.82	0.91	0.61	0.91	0.08	0.91	0.97

03 EVOLVING ML/AI ALGORITHMS

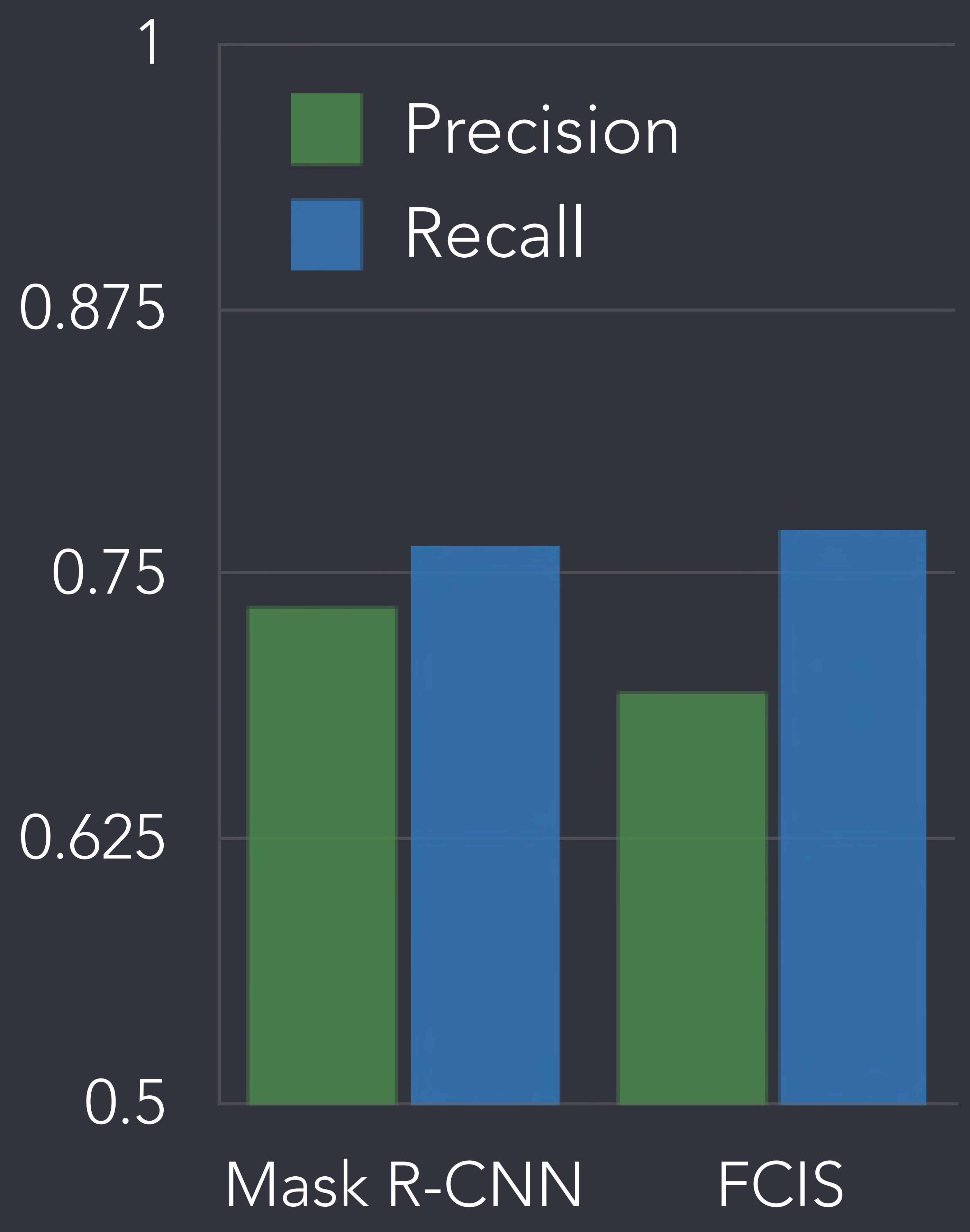
BENCHMARK TESTS DO NOT NECESSARILY INDICATE REMOTE SENSING SUITABILITY

2005 Labels
(Nebraska, USA)



Mask R-CNN
Likelihood Score

(Instance Segmentation)



Algorithm choices ↔ Tool choices ↔ Implementation choices

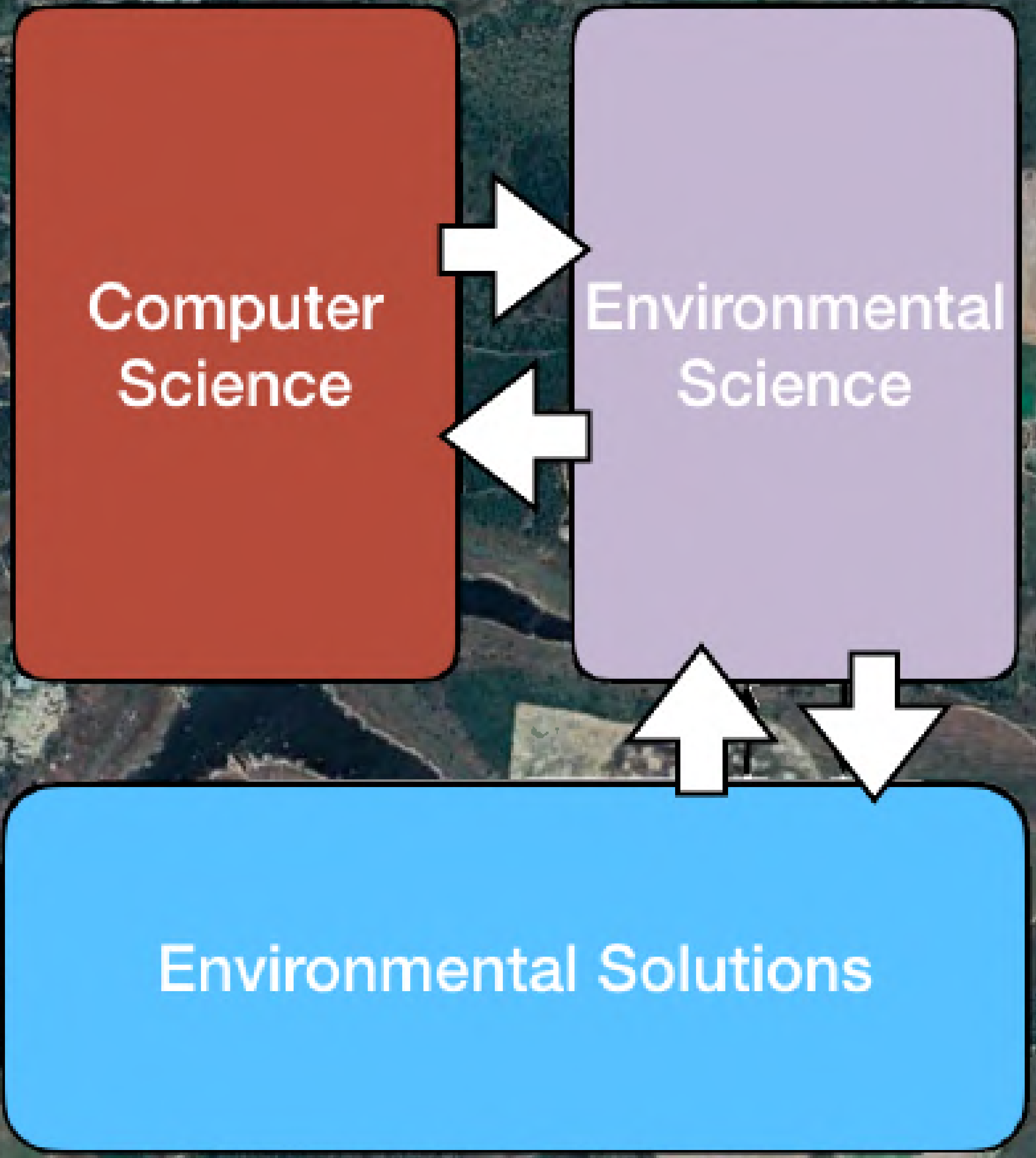
Emerging opportunities in coupled natural-human systems

Coming soon: Label-free training of panoptic segmentation models built on cloudy/incomplete data


Computer Science
Environmental Science
Environmental Solutions
September, 2004
-15.170849°, 28.430749°

How to Accelerate our Human Training Frameworks?

Closed,
Disciplinary
System



April, 2020
-15.170849°, 28.430749°


How to Accelerate our Human Training Frameworks?
Open,
Dissipative
System
Environmental Solutions
Environmental
Data
Science
April, 2020
-15.170849°, 28.430749°


Kelly Caylor
@kcaylor
caylor.eri.ucsb.edu
UCSB
Thank You!
http://bit.ly/Caylor-HelmholtzAI