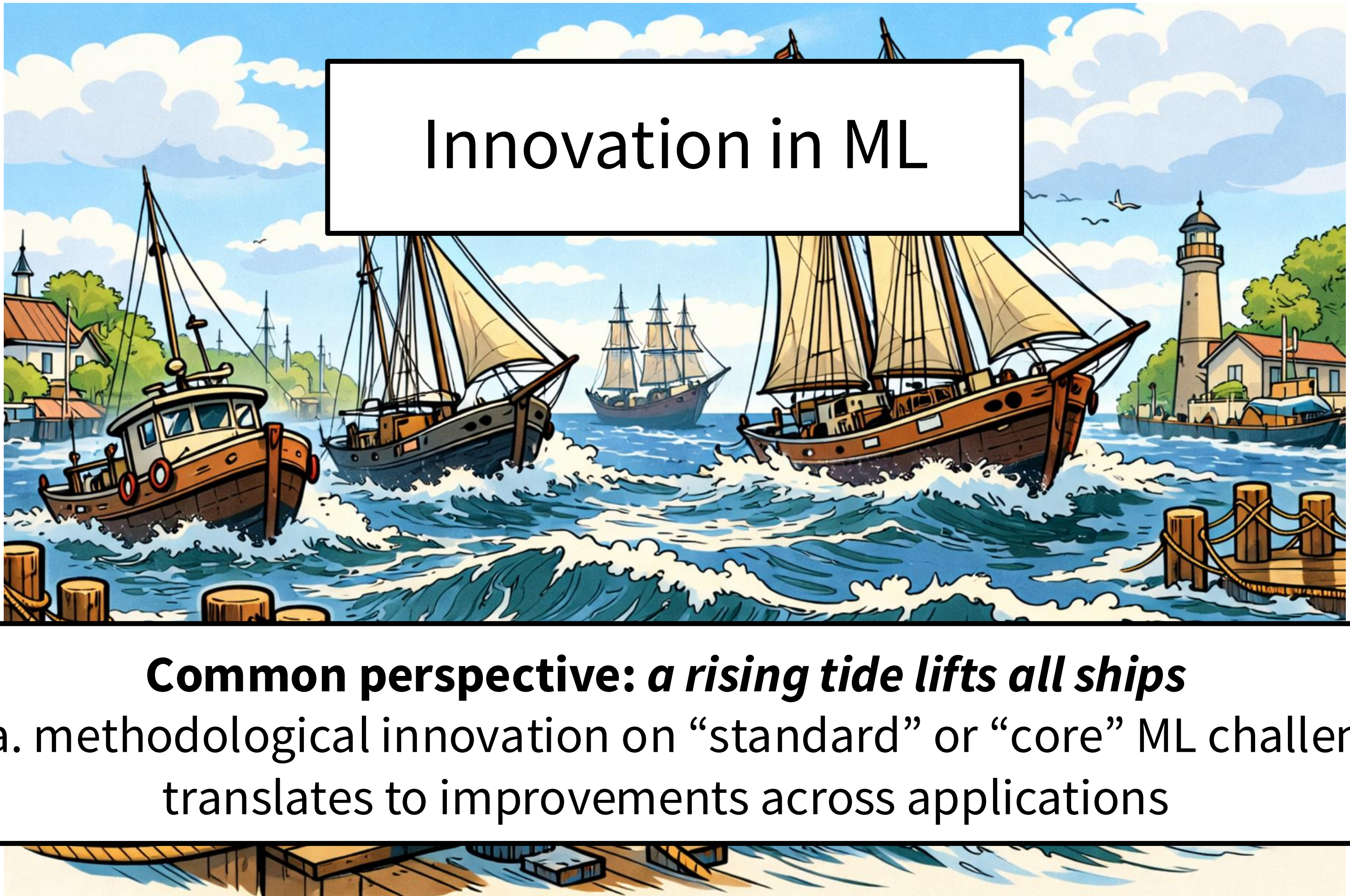
Application Driven Innovation in Machine Learning

6.S891/12.S992/6.S893: AI for Climate Action

Spring 2026

Speaker: Sara Beery

Innovation in ML



Common perspective: *a rising tide lifts all ships*
a.k.a. methodological innovation on “standard” or “core” ML challenges translates to improvements across applications

When has the tide lifted ships?

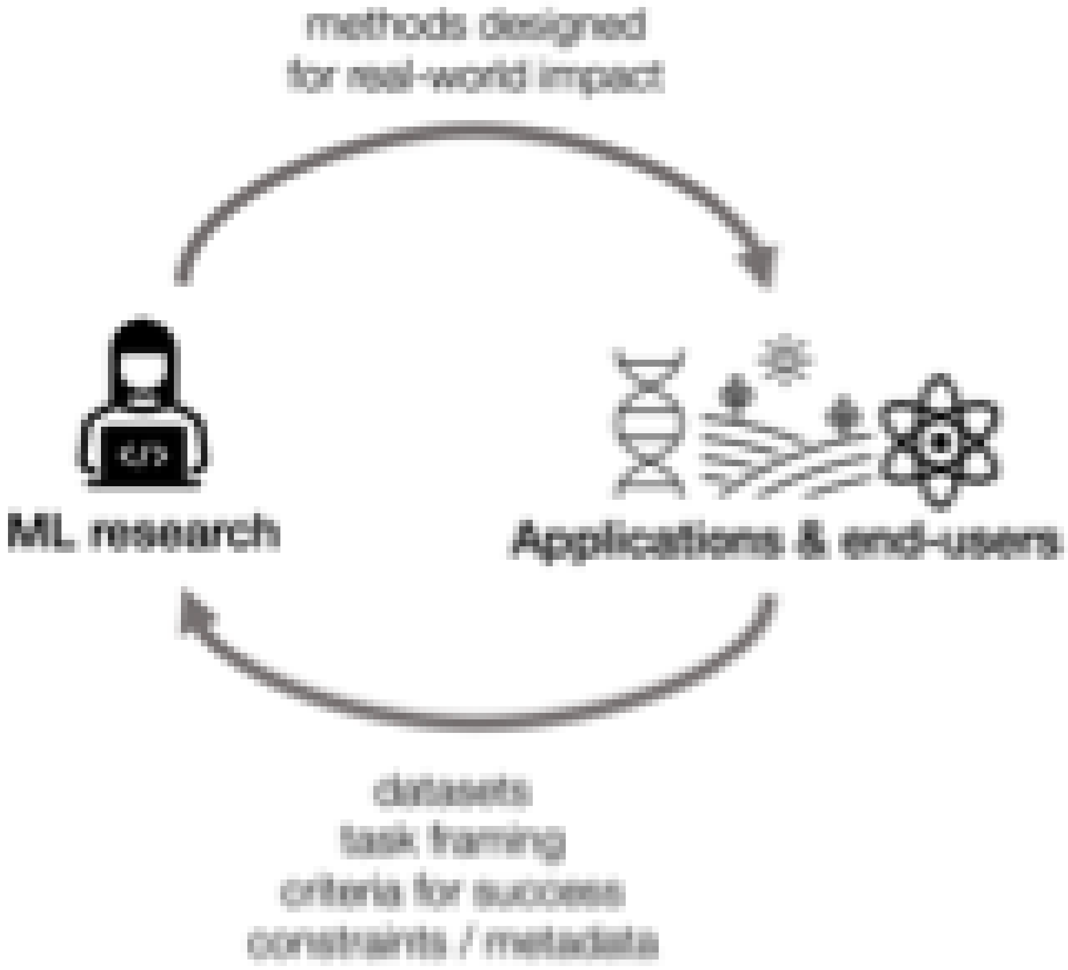
- Advances in one application that transfer broadly
  - Ex: Advances in general-purpose AI architectures
    - ResNets were developed for natural images, they work well for audio, microscopy, satellite data, etc
    - Transformers were developed for NLP, they are now widely used for everything from video to molecular graphs
- Advances in AI infrastructure - pytorch, GPUs, model zoos, training datasets
- Industry investment in scale - foundation models like ChatGPT, SAM

Arguably, *all* of these innovations were inspired by an application

Applications define needs, constraints, and metrics of success

Importantly, *new and diverse* applications often help outline failures of current AI systems

How have applications shaped ML Methods?


methods designed
for real-world impact
ML research
Applications & end-users
datasets
task framing
criteria for success
constraints / metadata

Case study: Data scarcity

- Applications include: biodiversity, healthcare, manufacturing anomaly detection
- Resulting innovations
  - Transfer learning
  - Self-supervised learning
  - Few-shot & zero-shot learning
  - Active learning
  - Training on synthetic data

Case study: Changing distributions

- Applications include: biodiversity, healthcare,
- Resulting innovations
    - Domain adaptation
    - Active calibration, active inference
    - Distribution matching w/ optimal transport
    - Distributionally robust optimization

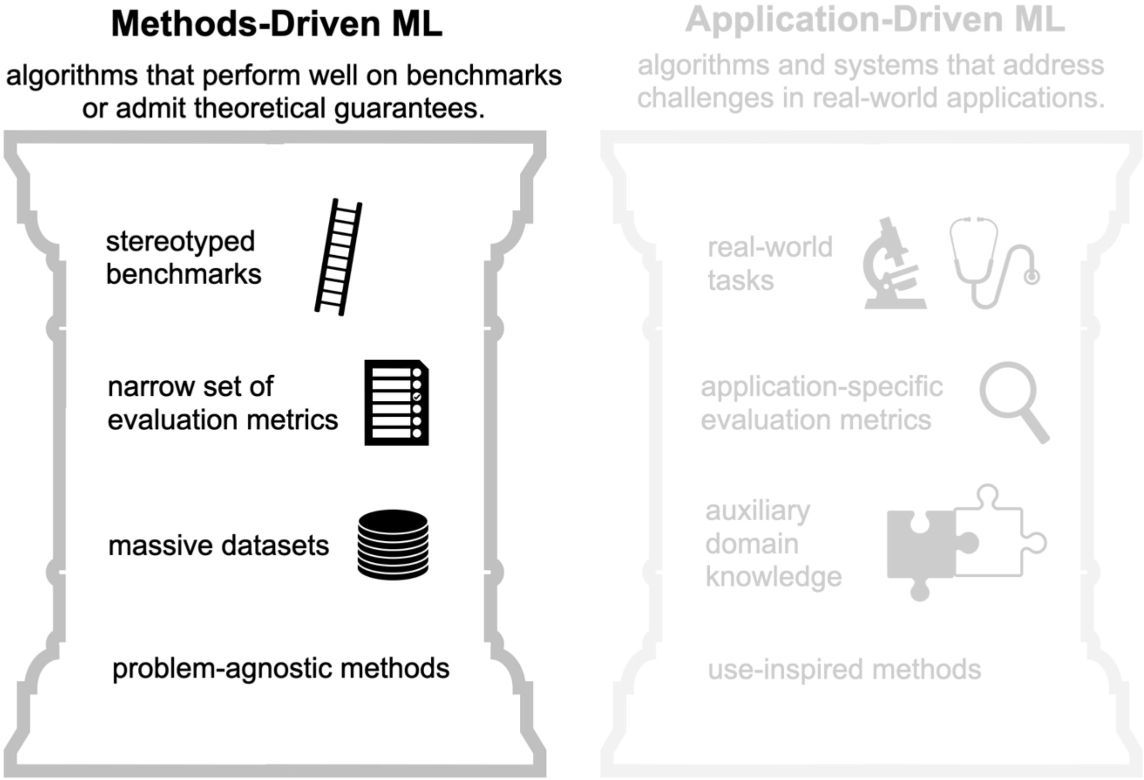
Case study: Real-time constraints

- Applications include: Autonomous driving, fraud detection, recommender systems
- Resulting innovations
  - Online learning
  - Model compression & distillation
  - Approximate inference

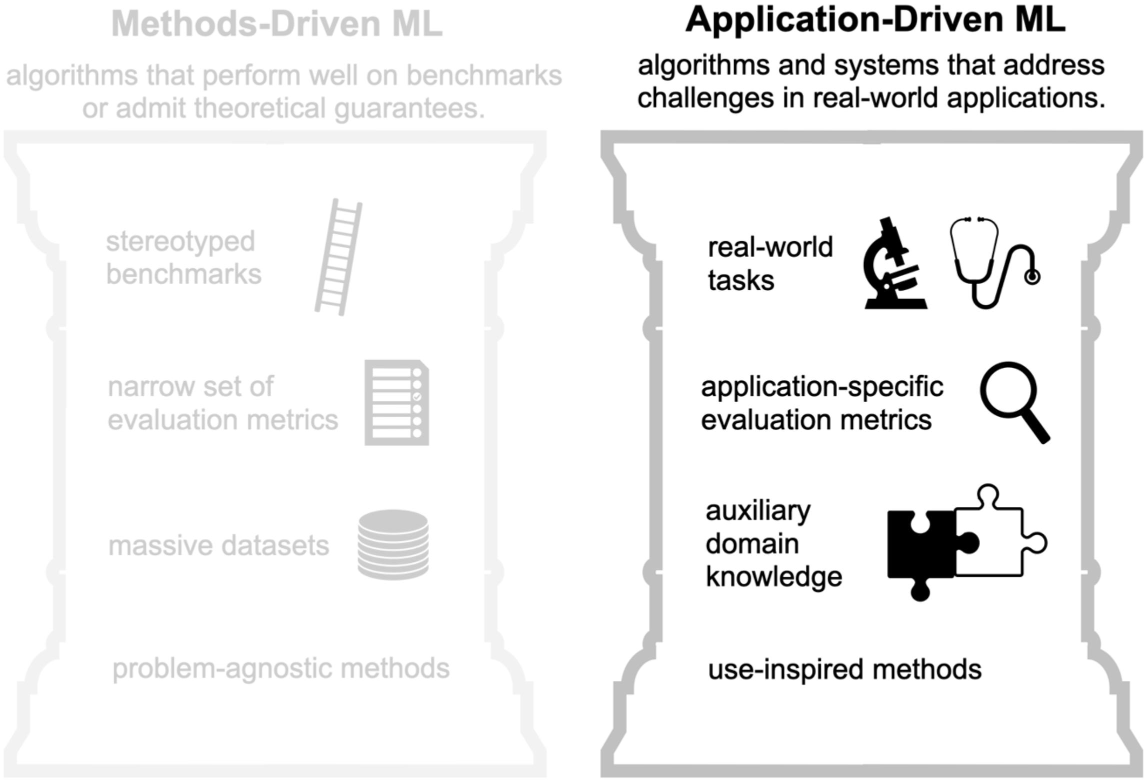
Case study: High-stakes decisions

- Applications include: healthcare, finance, justice systems, resource allocation after natural disasters
- Resulting innovations
  - Interpretability methods
  - Uncertainty estimation
  - Causal ML
  - Selective prediction
  - Alignment

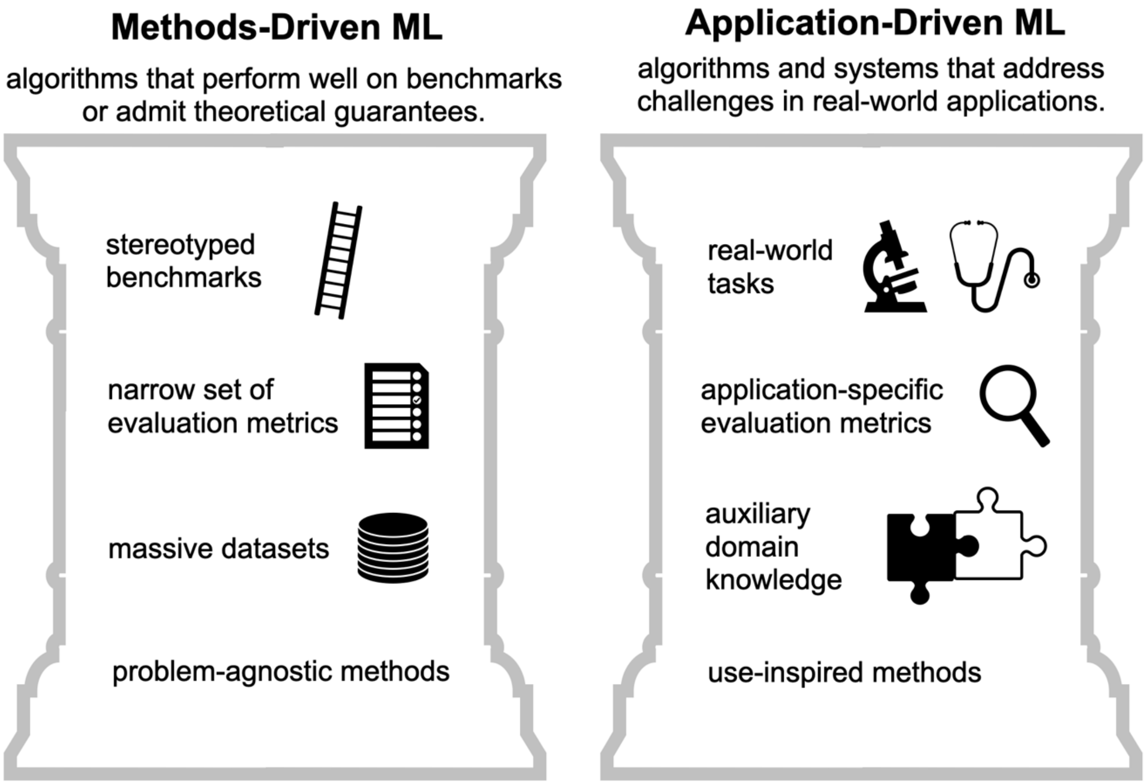
So, what defines a “core” ML challenge?



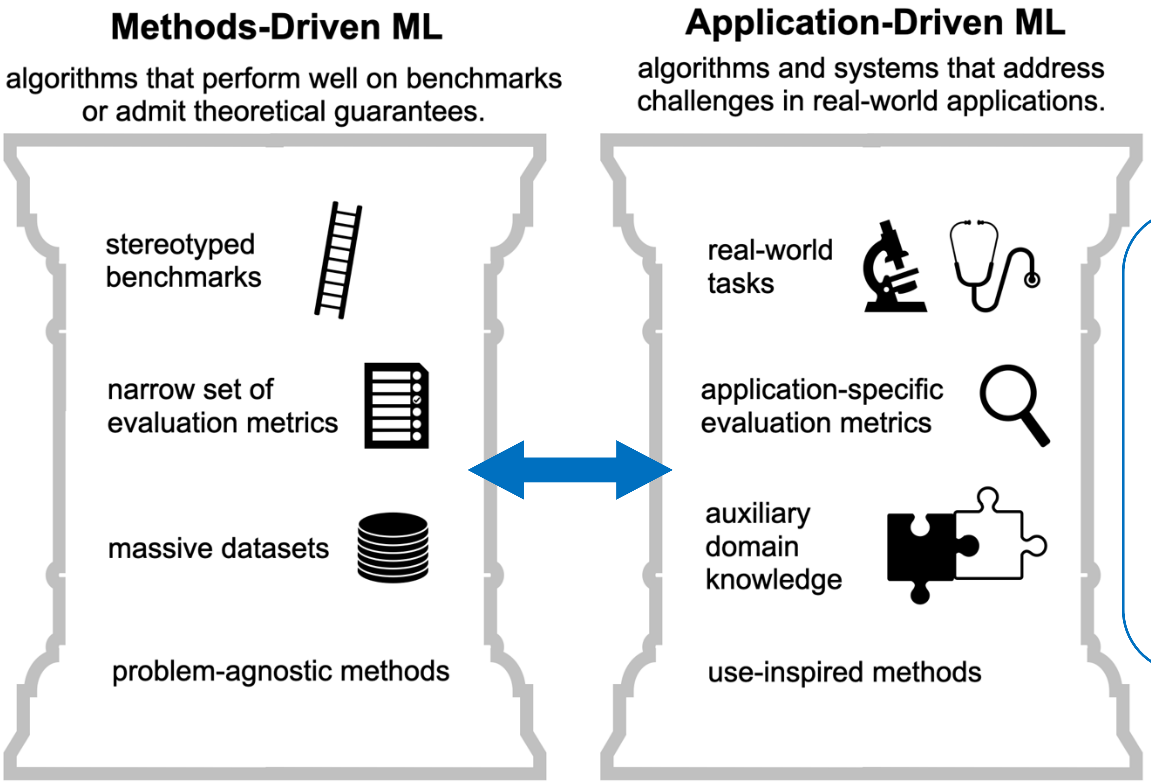
Rolnick, et al. “Application-driven Innovation in Machine Learning”,
International Conference on Machine Learning (ICML) 2024.



Rolnick, et al. "Application-driven Innovation in Machine Learning", International Conference on Machine Learning (ICML) 2024.



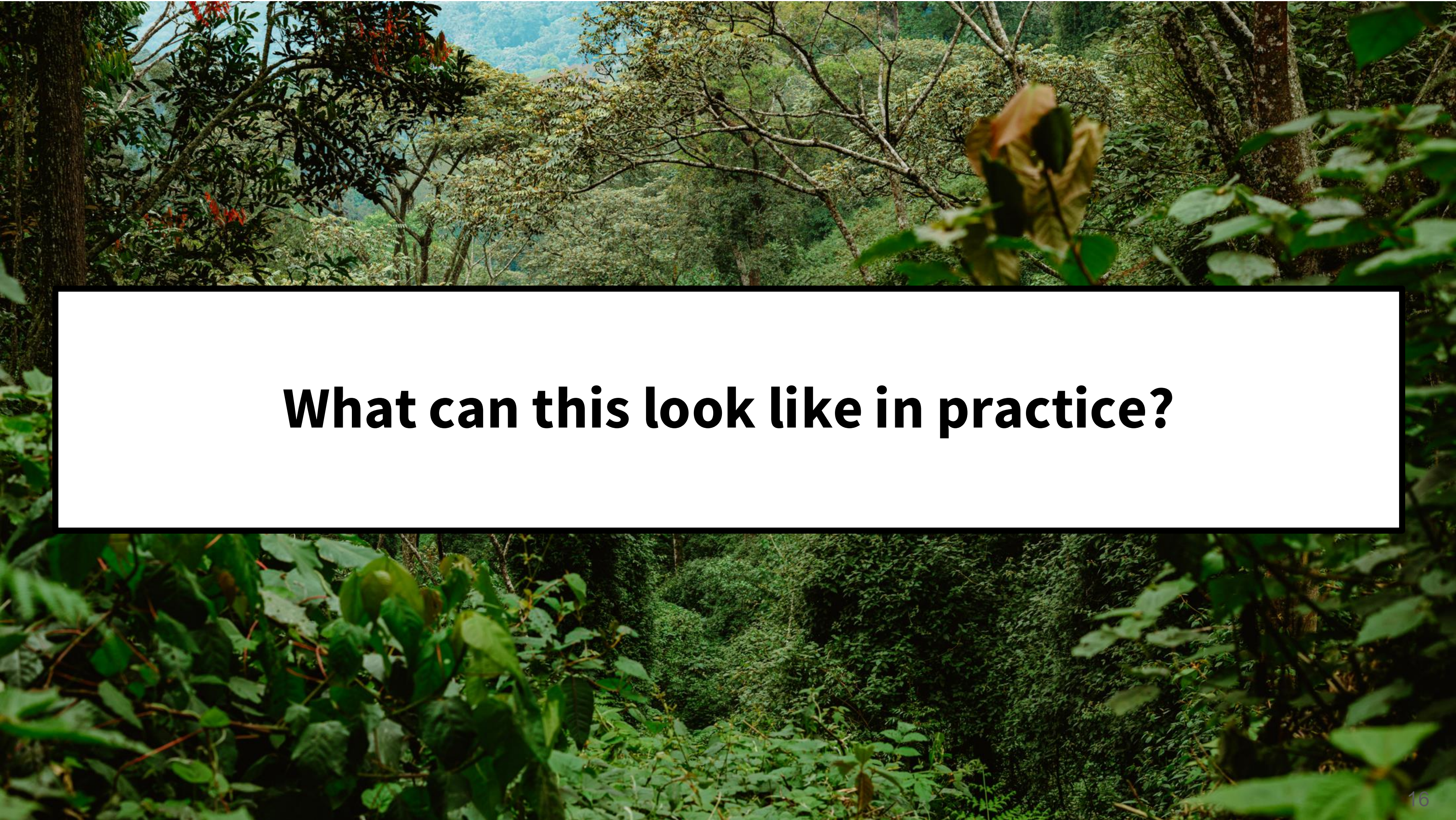
Rolnick, et al. “Application-driven Innovation in Machine Learning”, International Conference on Machine Learning (ICML) 2024.



Challenges for ML

OOD generalization
Interpretability
Lightweight models
Physical constraints
Limited labels
Multi-modal data
...

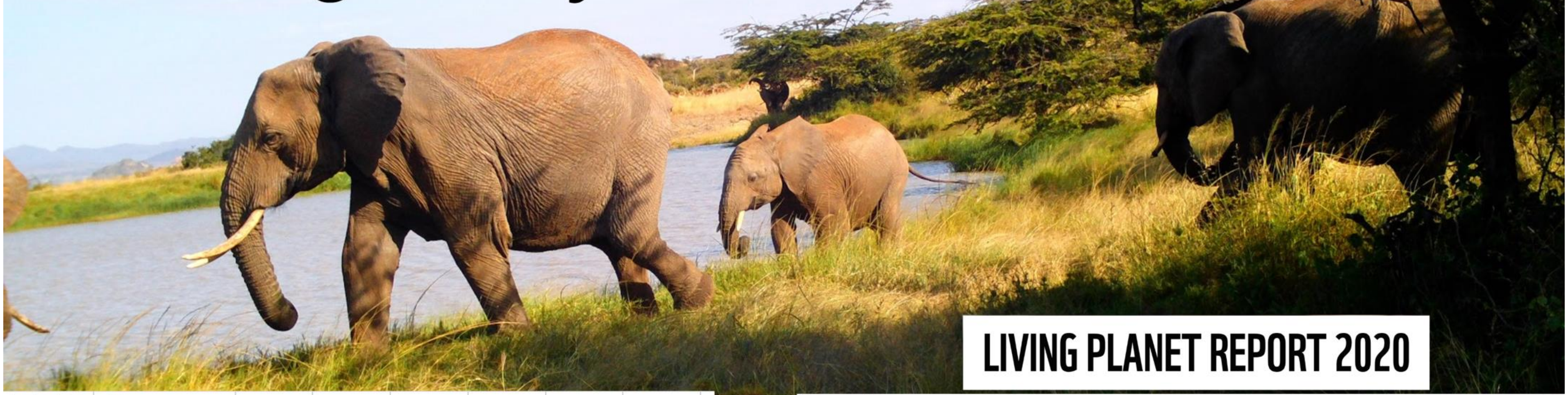
Rolnick, et al. “Application-driven Innovation in Machine Learning”, International Conference on Machine Learning (ICML) 2024.



What can this look like in practice?

Monitoring wildlife

Biodiversity is in decline globally



BBC | Sign in | Home | News | Sport | Reel | Worklife | Travel

NEWS

Home | US Election | Coronavirus | Video | World | US & Canada | UK | Business | Tech | Science | Stories

Science

Wildlife in 'catastrophic decline' due to human destruction, scientists warn

16:3

WWF | OUR WORK | PEOPLE | PLACES | WILDLIFE | About | How to help | DONATE + | ADO

PRESS RELEASES

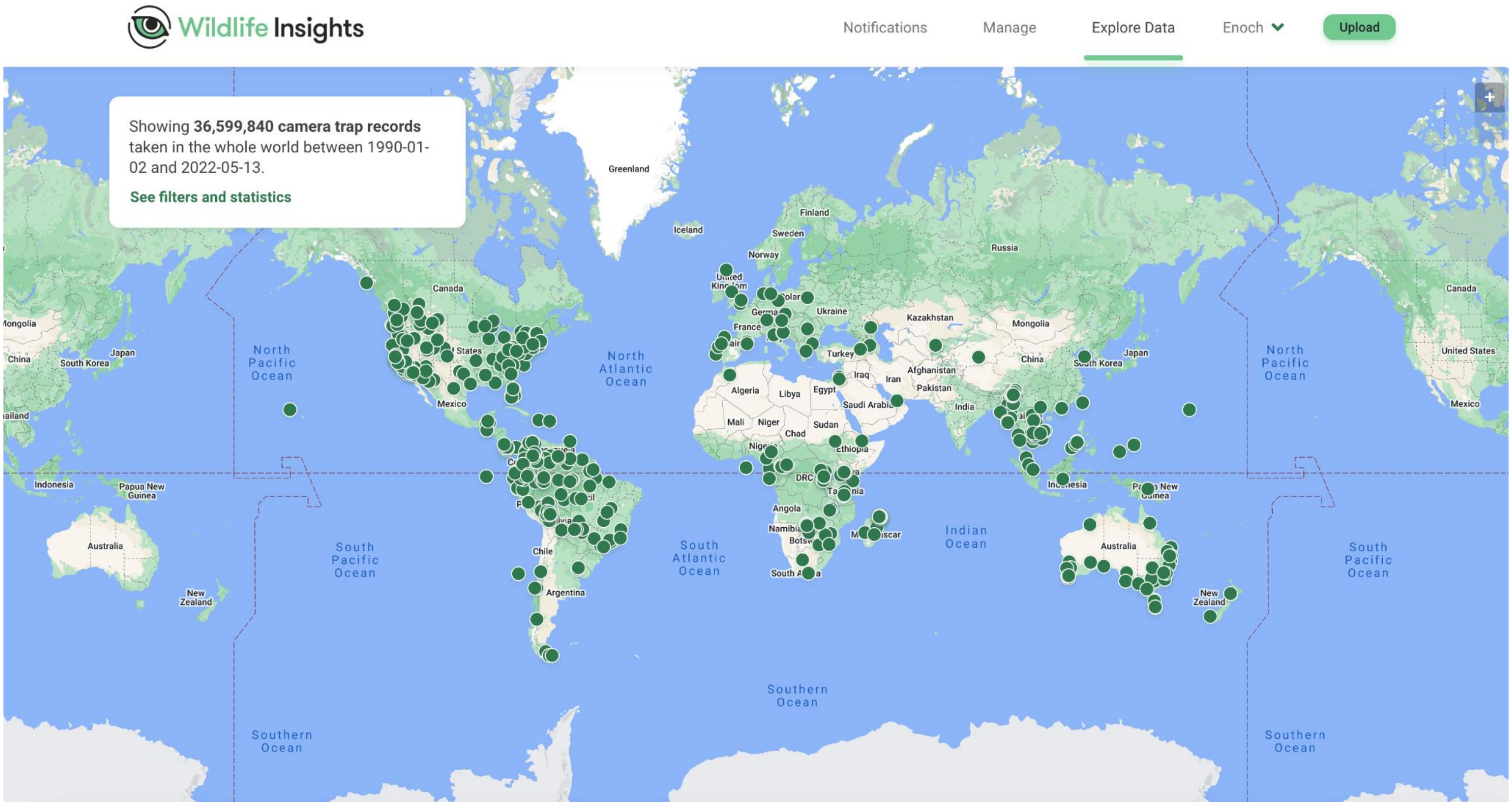
68% Average Decline in Species Population Sizes Since 1970, Says New WWF Report

Declines in monitored populations of mammals, fish, birds, reptiles, and amphibians present a dire warning for the health of people and the planet

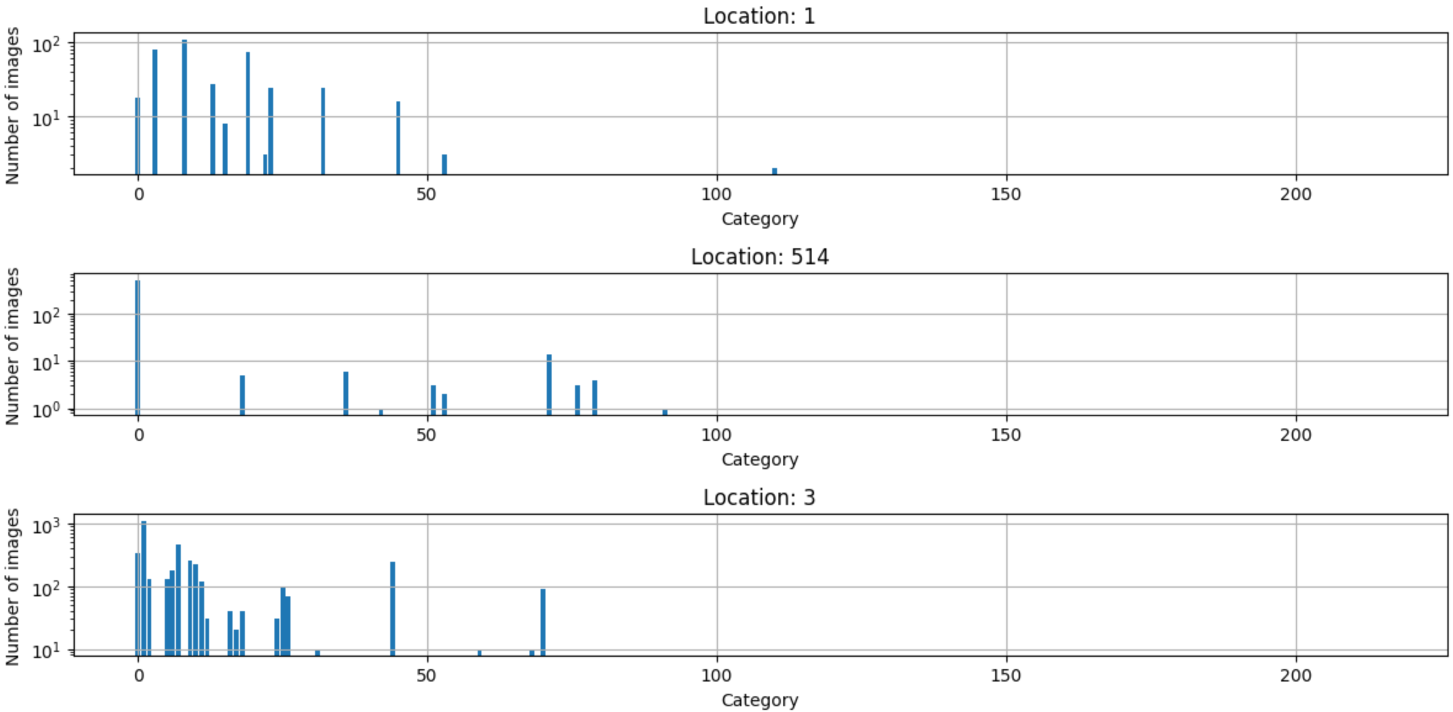
Biodiversity data collection is increasing in quantity and diversity

Perspectives in machine learning for wildlife conservation, Beery*, et al., Nature Comms 2022

AI can help ecologists process huge volumes of data

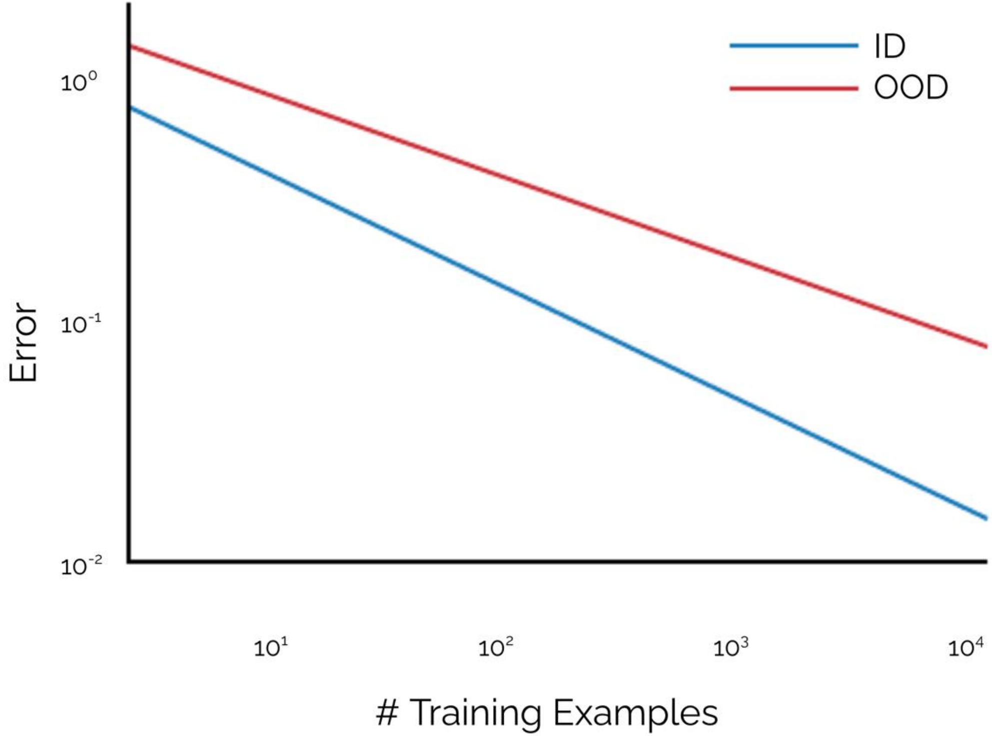


Distribution shift makes this hard



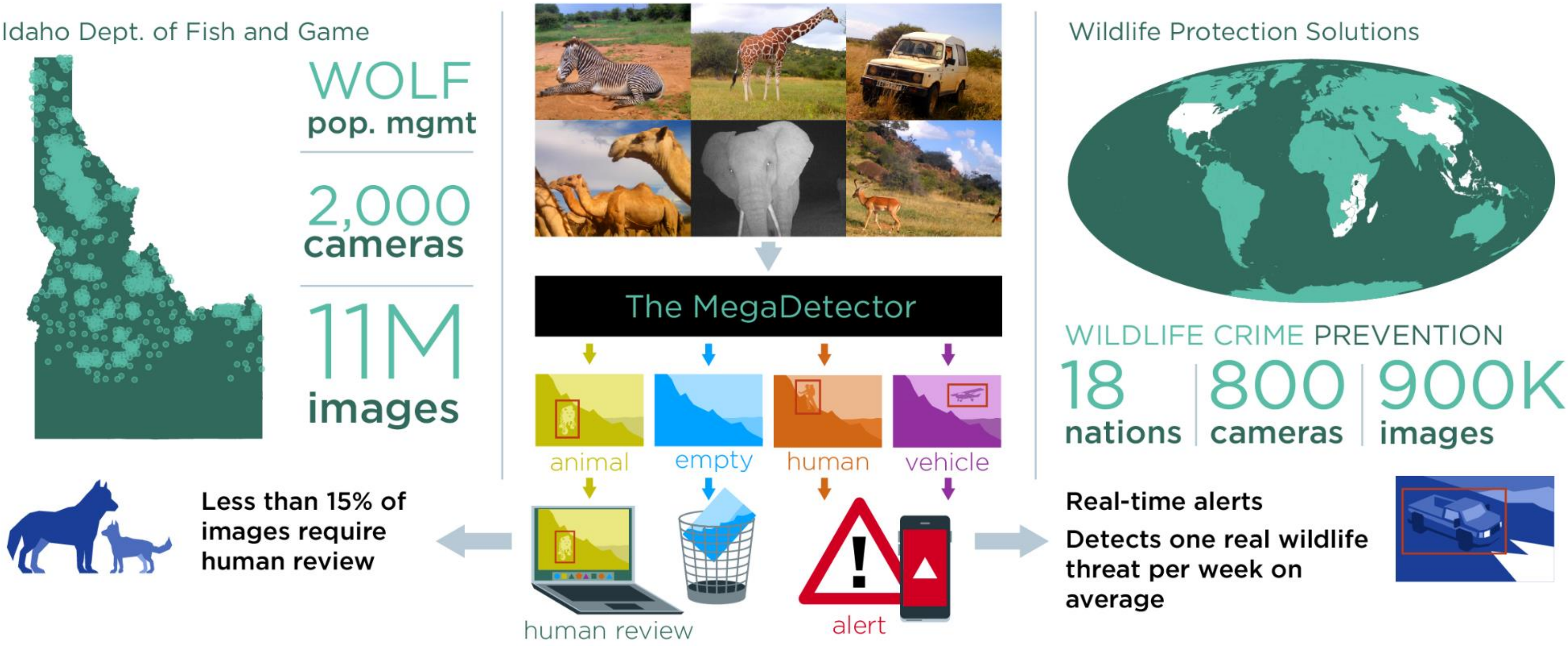
Recognition in Terra Incognita, Beery, et al. ECCV 2018

Models don’t generalize



Recognition in Terra Incognita, Beery et al., ECCV 2018

Focus on developing methods that generalize leads to impact

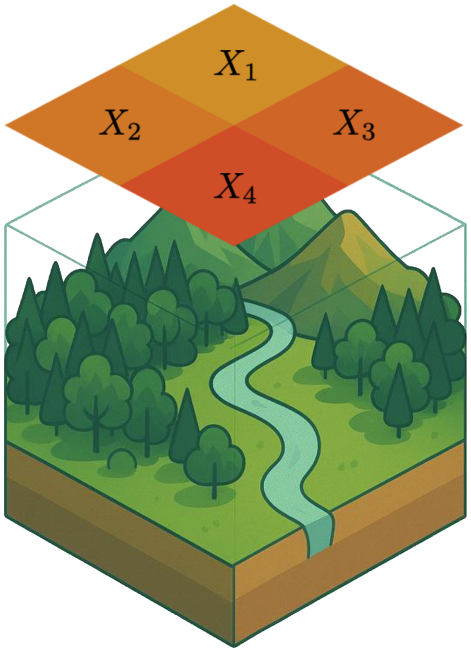


Efficient pipeline for camera trap image review, Beery, et al. KDD 2019

Moving beyond image processing – what do we need?

Occupancy Models

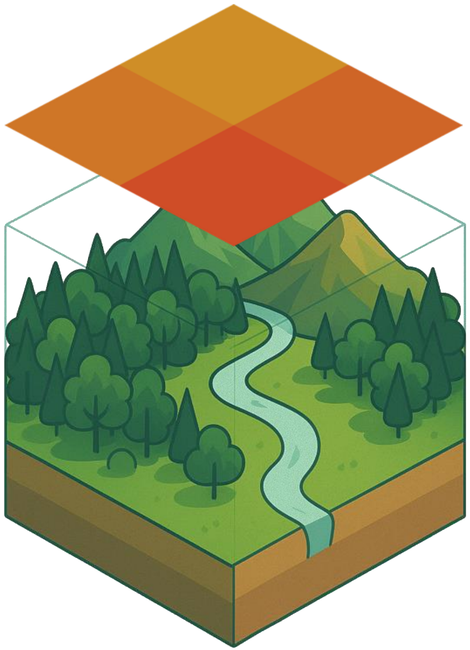
E.g. average temperature



$$\psi(X_i) \in [0, 1]$$

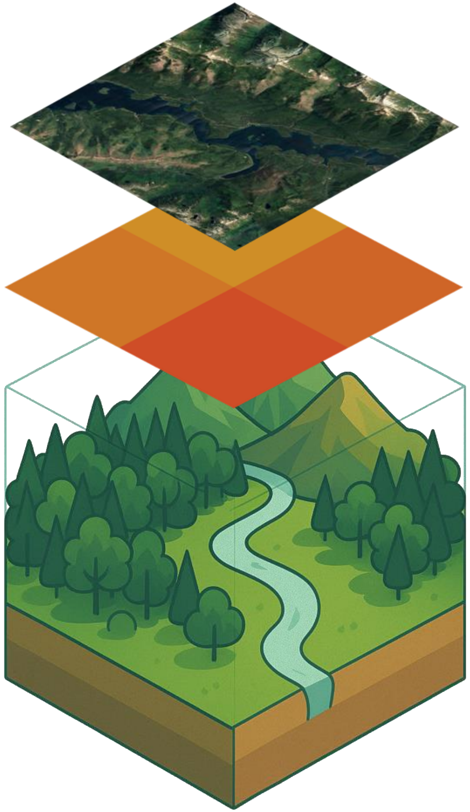
Occupancy probability

Limitations of environmental variables



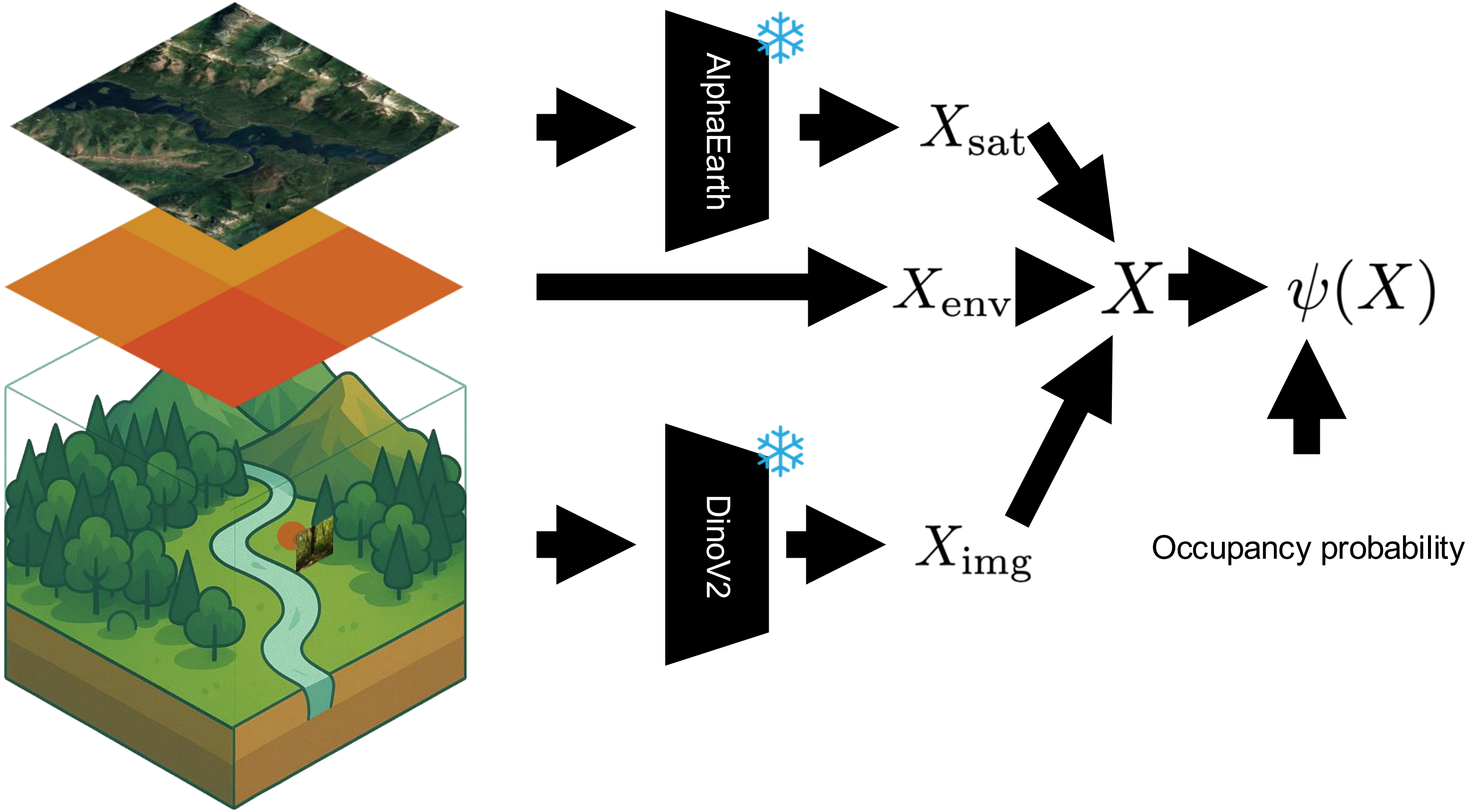
- Interpolated from far-away measurements → Low resolution
- Frequently fail to capture micro-habitat conditions
  - Micro-climates
  - Below canopies

High-resolution satellite imagery



- Shown to be helpful for species distribution modeling (e.g. SatBird, Teng et al. 2023)
- Can we do even better?

Multi-modal habitat descriptions



Occupancy Model Formulation

$$\psi(X) = \sigma(\beta^\top X)$$

"Habitat suitability" / occupancy probability

$$z \sim \text{Bernoulli}(\psi(X))$$

Discrete occupancy state

$$z \in \{0, 1\}$$

$$p_{\text{det}}$$

Probability of detection

$$y \sim \text{Bernoulli}(z \cdot p)$$

To make a detection, the site needs to be occupied, and the animal needs to be detected

$$z = 1$$

Fitting

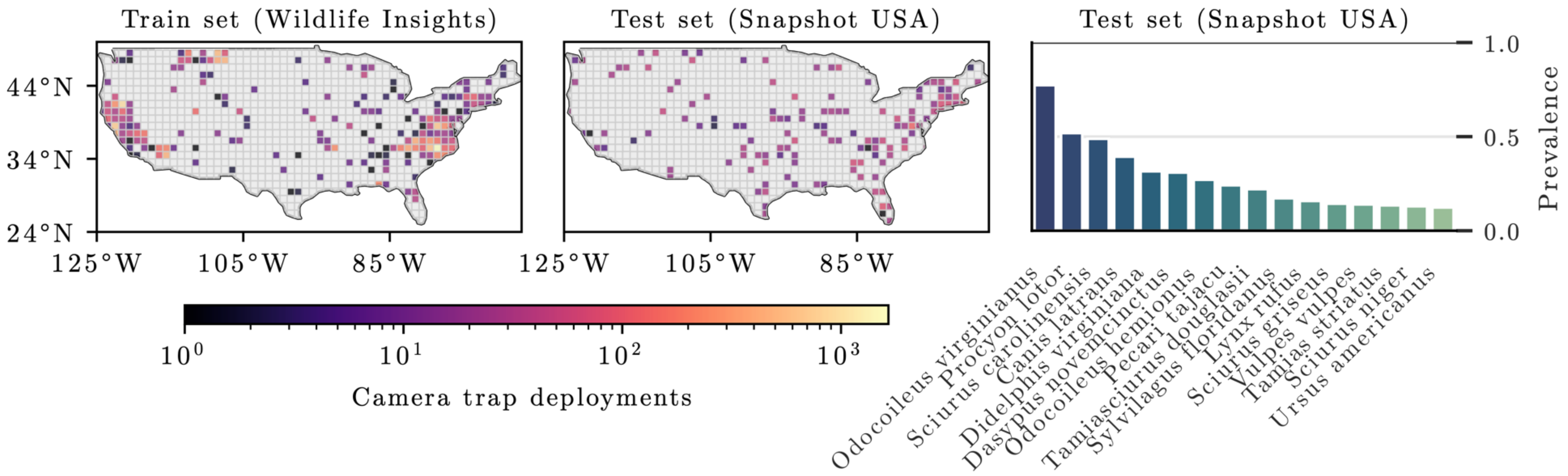
- Fit using Biolith (github.com/timmh/biolith)
- Fully Bayesian
- Using MCMC

Evaluation — Data

Snapshot USA

(Rooney et al. 2025)

Wildlife Insights



Evaluation — Metric

- Don’t know the “true” occupancy, only have observations
- Instead, see how well our model predicts observations on held-out sites

$$\text{LPPD} = \sum_{i=1}^{n} \log\left(\frac{1}{Q}\sum_{q=1}^{Q} p(y_i|\theta^{(q)})\right)$$

Sum over sites

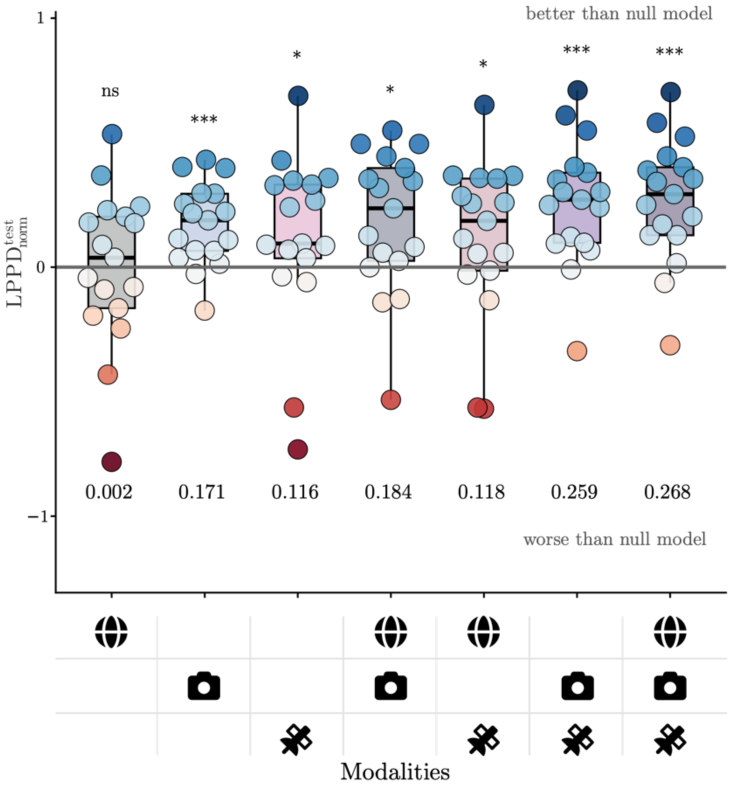
Sum over posterior samples

Likelihood of observation

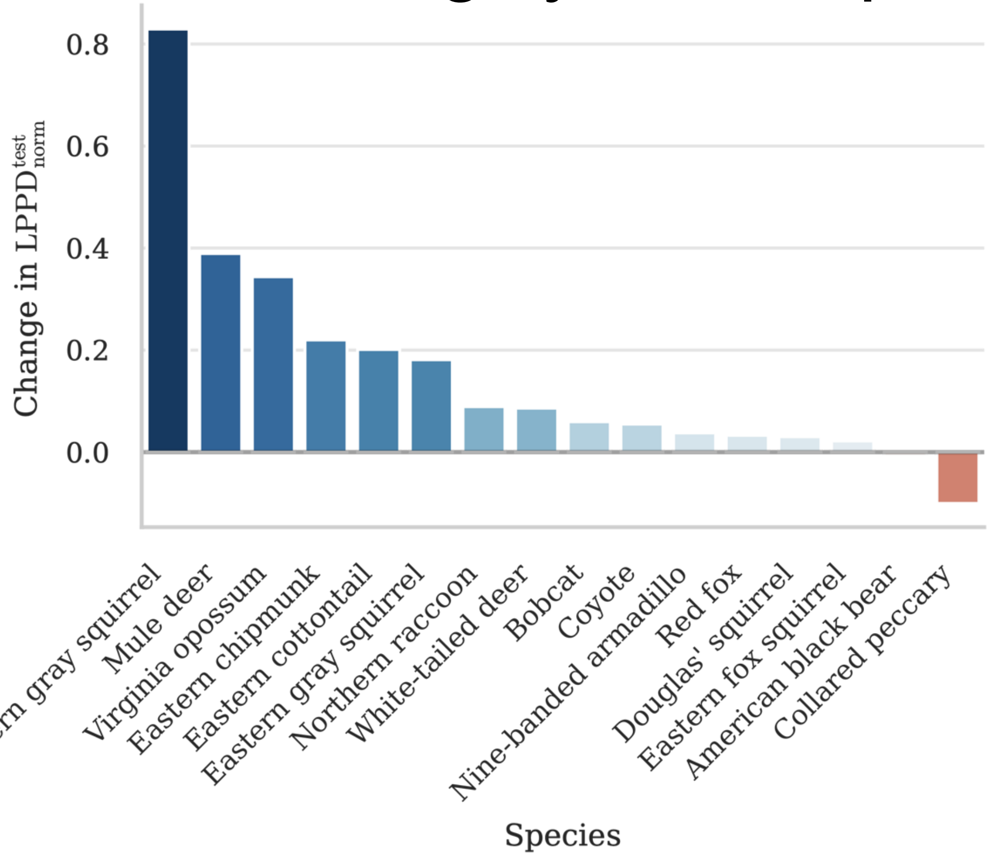
Model parameters

- Absolute value of LPPD is difficult to interpret
- Normalize to a 0-1 scale defined by null (intercept-only) and oracle (trained on test data) models

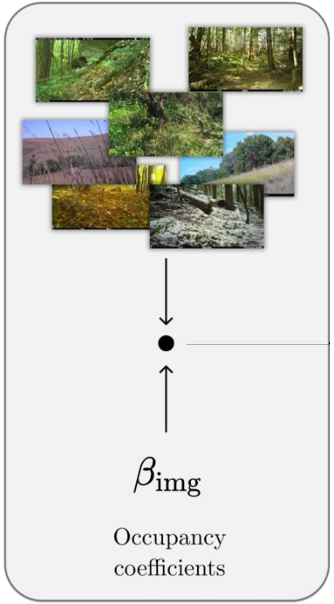
How predictive is each modality?



Value of ground-level imagery across species



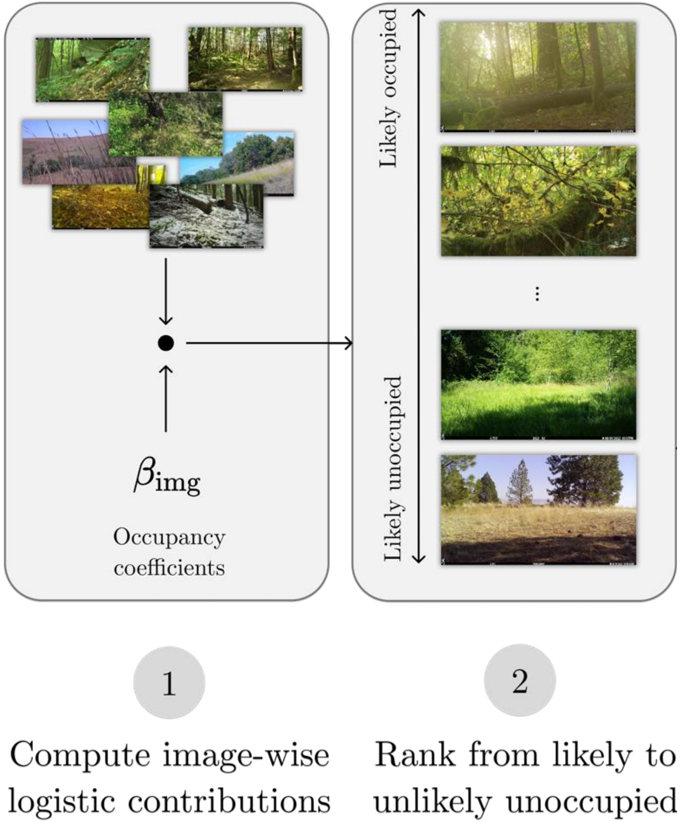
Explaining and Simplifying Deep Multi-modal models



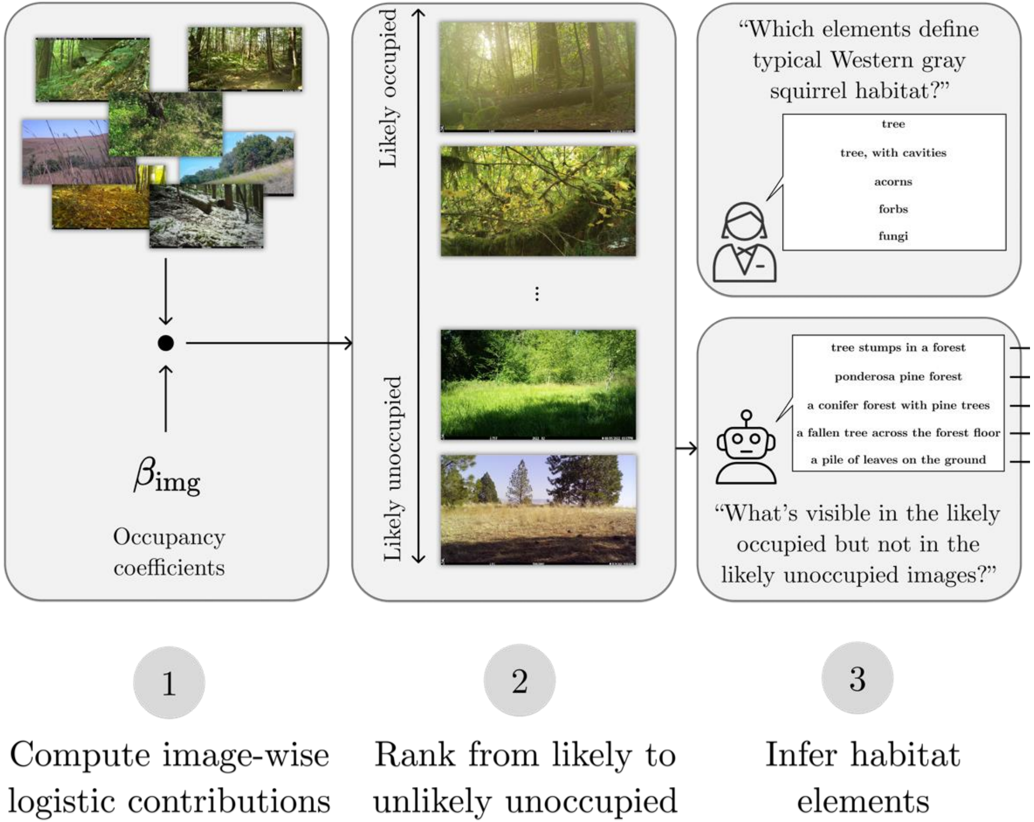
1

Compute image-wise logistic contributions

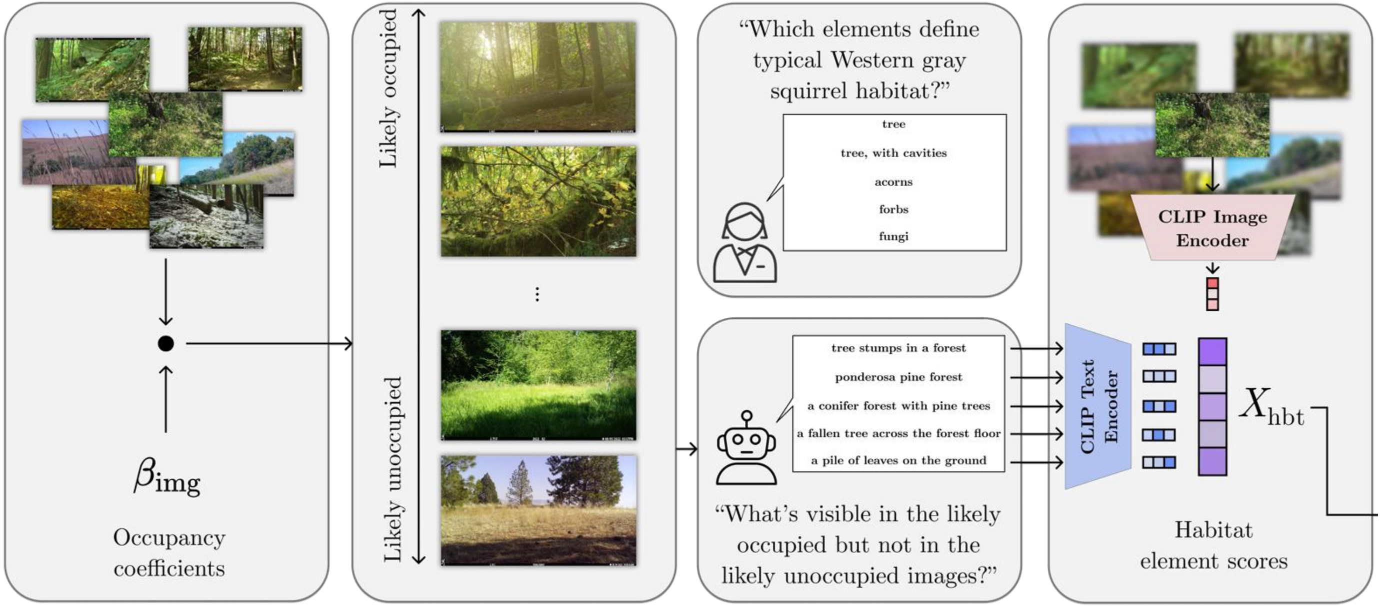
Explaining and Simplifying Deep Multi-modal models



Explaining and Simplifying Deep Multi-modal models

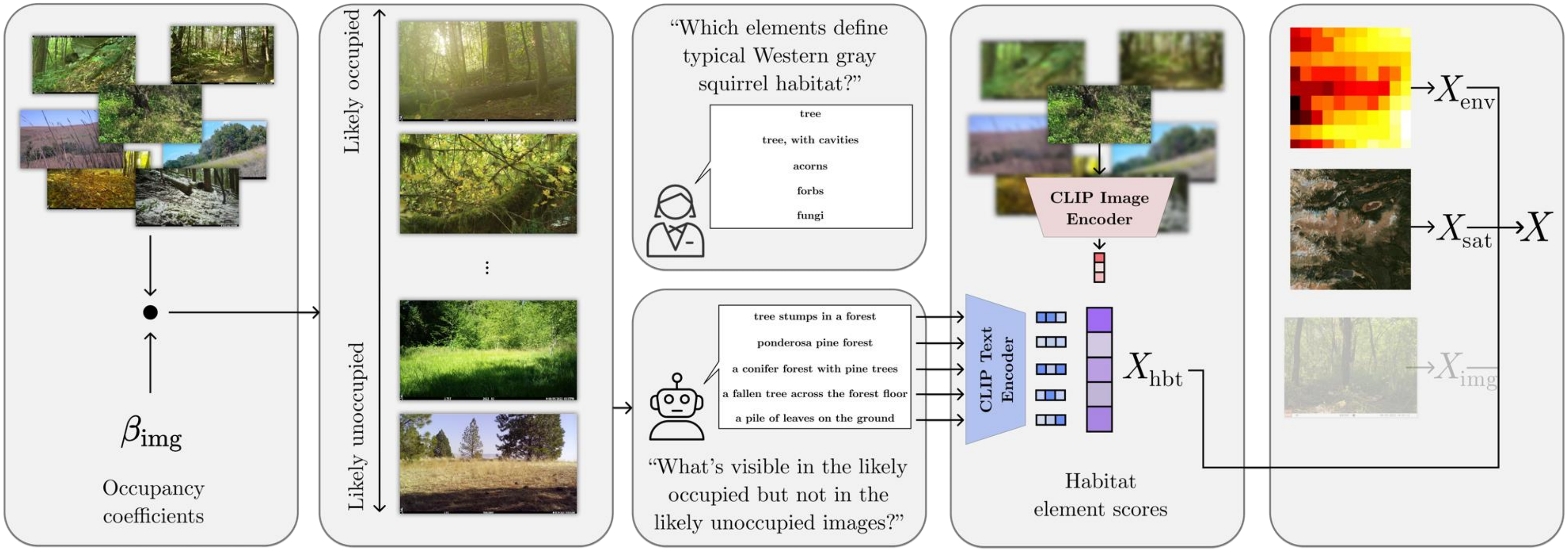


Explaining and Simplifying Deep Multi-modal models



1. Compute image-wise logistic contributions
2. Rank from likely to unlikely unoccupied
3. Infer habitat elements
4. Compute habitat element scores

Explaining and Simplifying Deep Multi-modal models



1 Compute image-wise logistic contributions

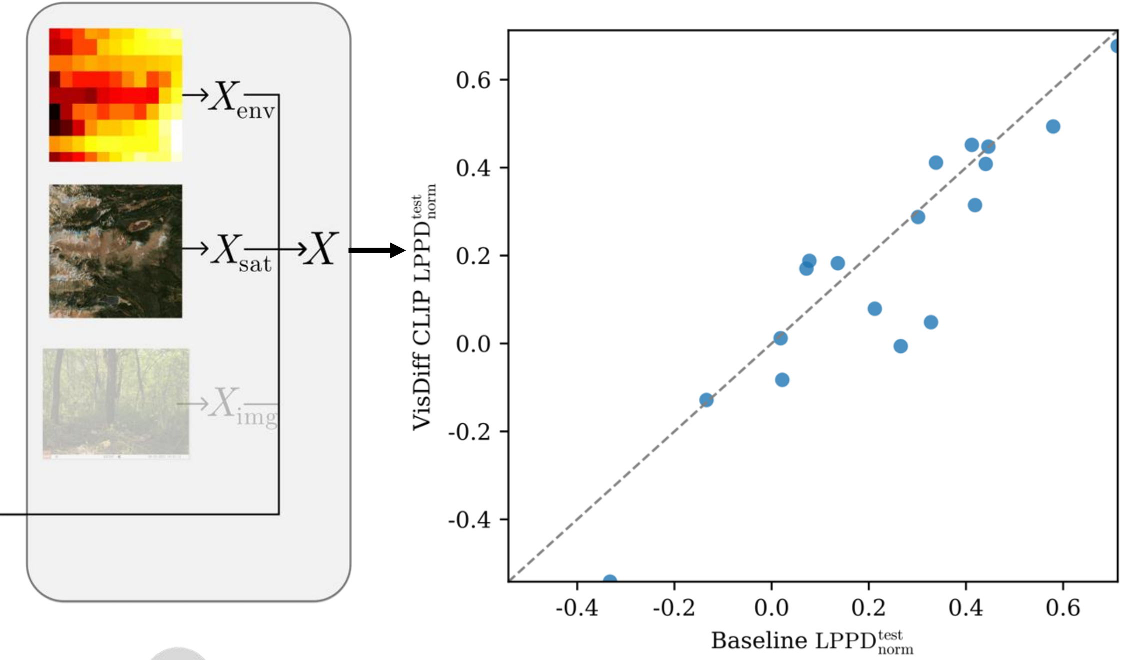
2 Rank from likely to unlikely unoccupied

3 Infer habitat elements

4 Compute habitat element scores

5 Re-fit models with habitat element scores

Explaining and Simplifying Deep Multi-modal models



5

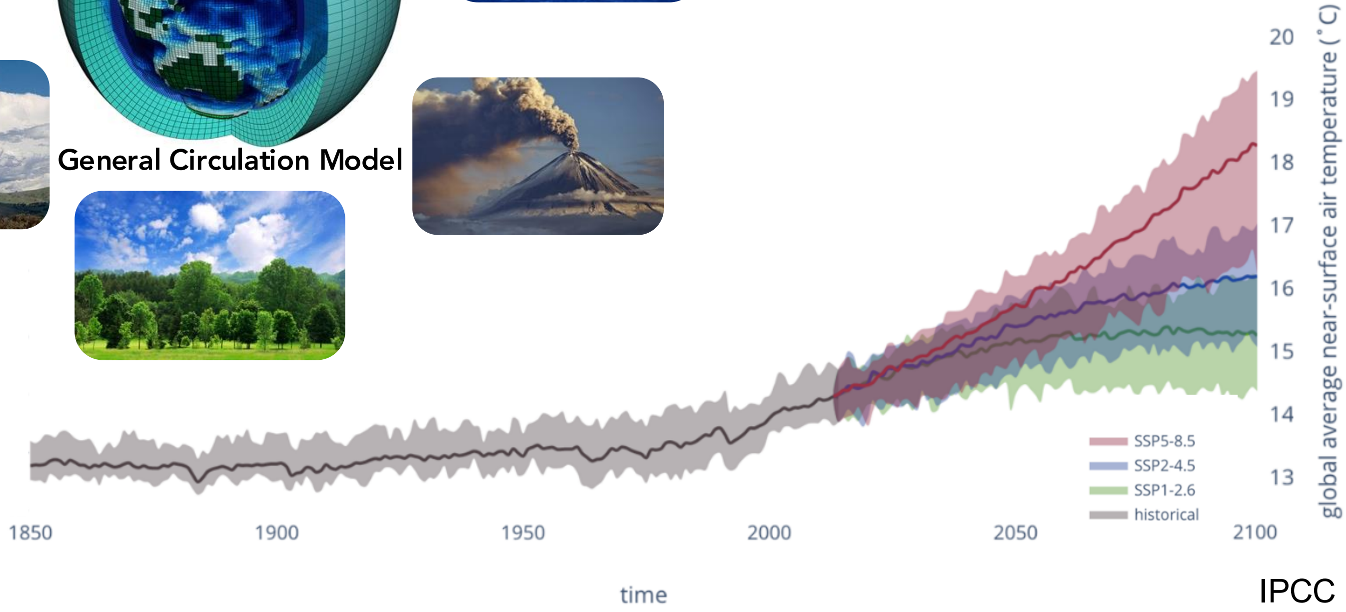
Re-fit models with habitat element scores

Application-driven innovation in ML

- Real-world tasks
- Application-specific evaluation metrics
- Auxiliary domain knowledge
- Use-inspired & problem-informed methods

Modeling the ocean (example from Abigail Bodner)

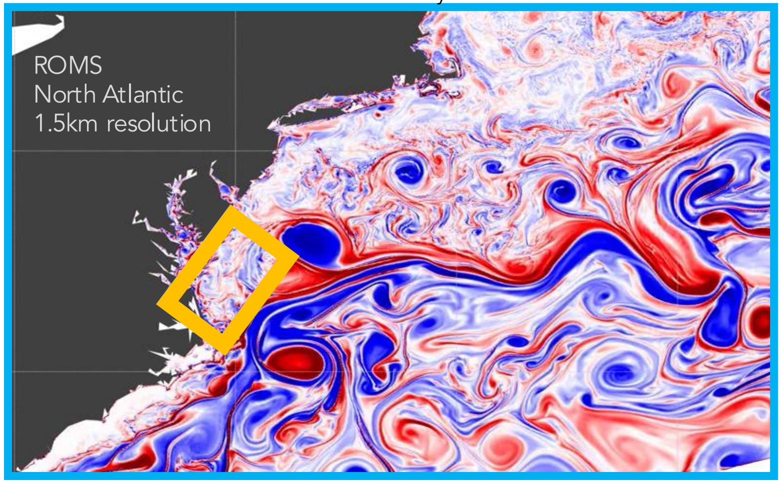
Global Climate Change

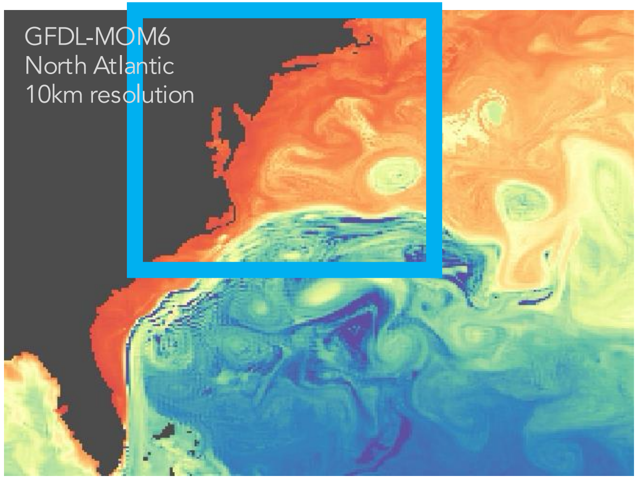


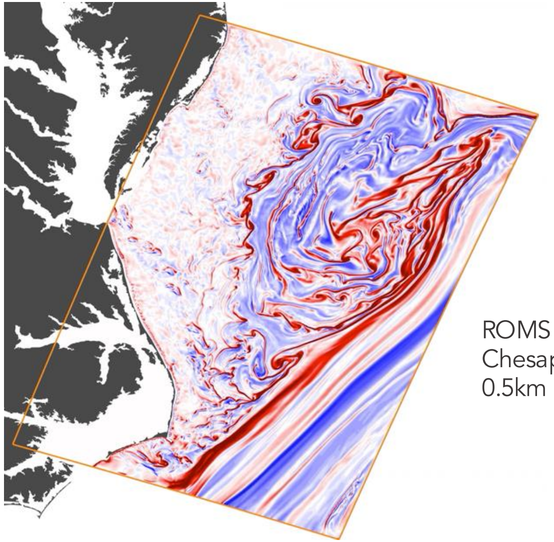
IPCC

Sea surface Temperature

Vorticity





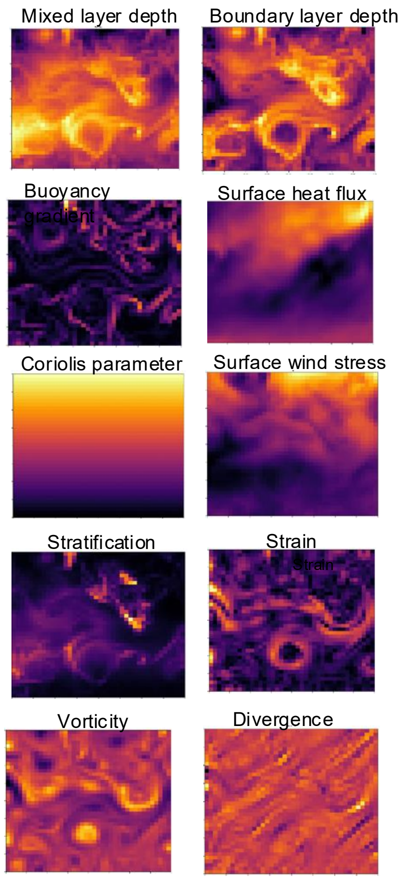


ROMS
Chesapeake Bay
0.5km resolution

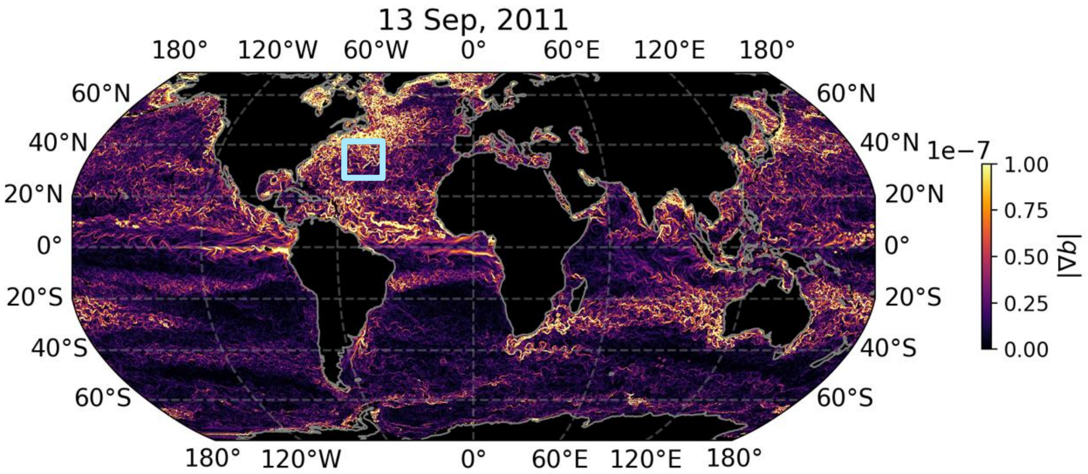
Courtesy of Julius Busecke

Courtesy of Jacob Wenegra

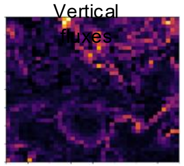
Data-driven submesoscale parameterization



MITgcm-llc4320 (horizontal resolution $1/48^o$ ~2km)



Given a set of relevant variables: predict vertical fluxes directly computed from data

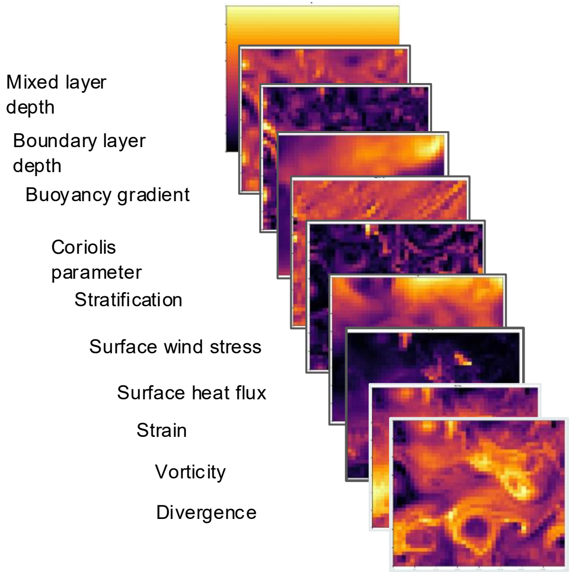


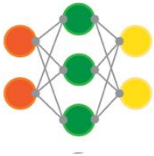
Bodner, Balwada, Zanna (2025)

Data-driven submesoscale parameterization

MITgcm-llc4320 (horizontal resolution $1/48^o$ ~2km)

Inputs
Variables resolved by
General Circulation Models

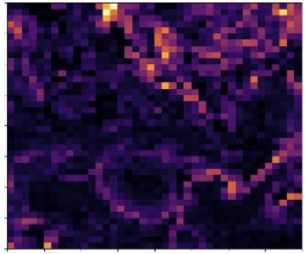




Fully Convolutional
Neural Network

Output
Submesoscale vertical
buoyancy fluxes

$\overline{w'b'}$

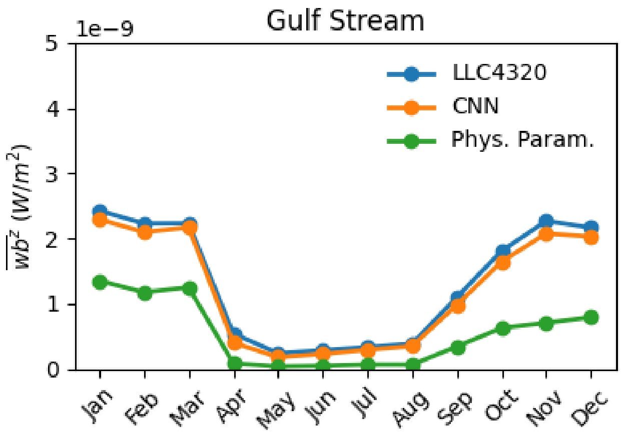


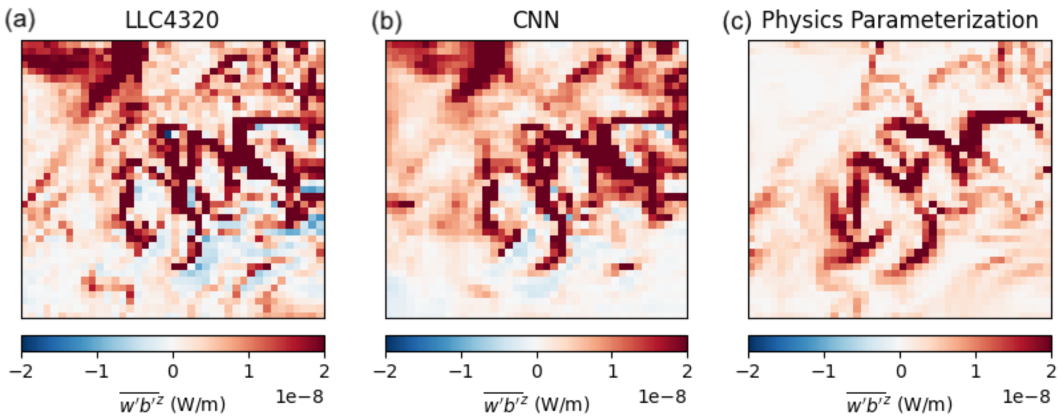
Given a set of relevant variables: predict vertical fluxes directly computed from data

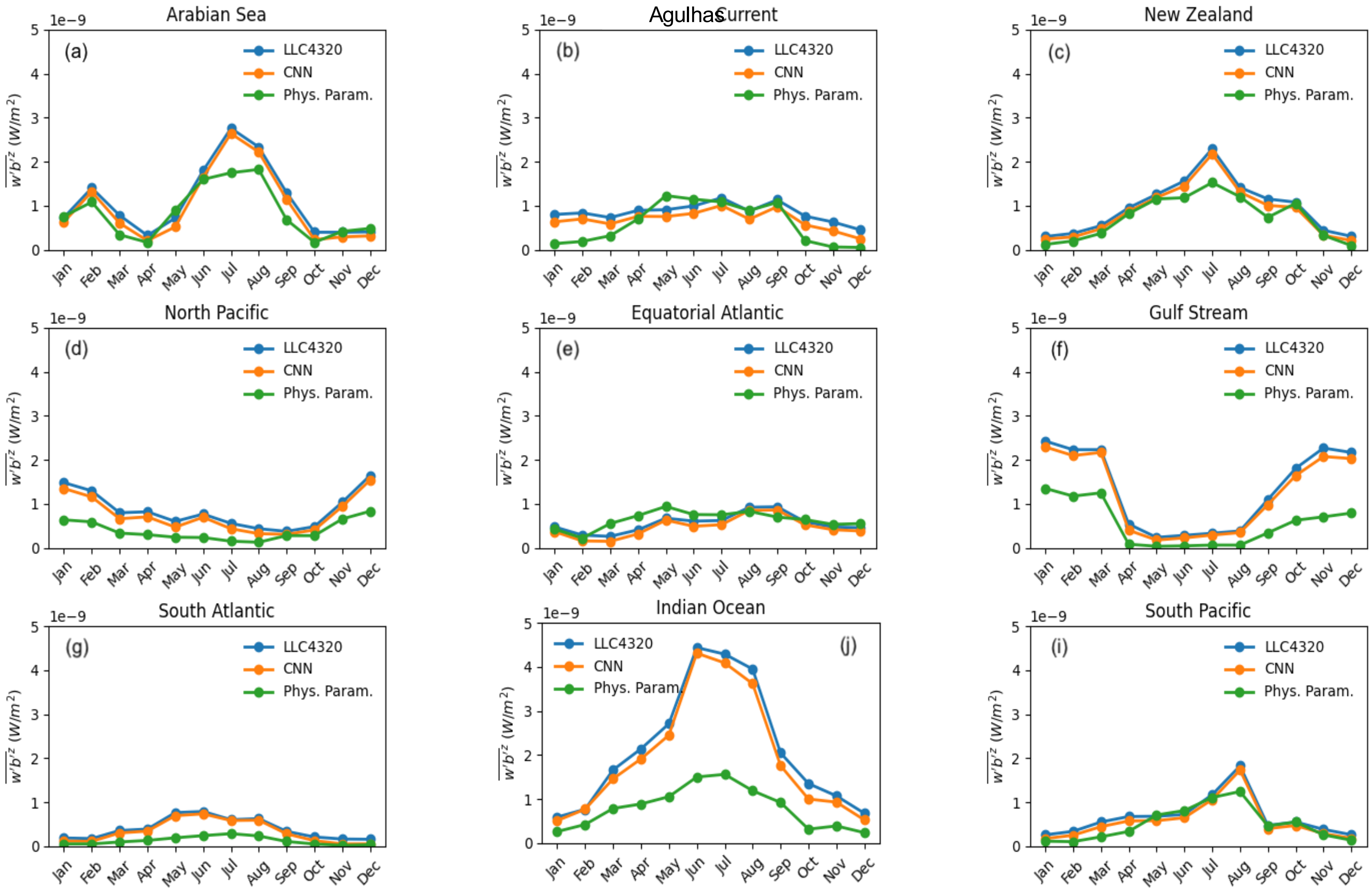
Bodner, Balwada, Zanna (2025)

Prediction on unseen data

- The CNN captures the overall structure, including negative fluxes
- the CNN predictions outperforms the physics-based parameterization, particularly during months of strong submesoscale fluxes.







Arabian Sea
(a)
Agulhas Current
(b)
New Zealand
(c)
North Pacific
(d)
Equatorial Atlantic
(e)
Gulf Stream
(f)
South Atlantic
(g)
Indian Ocean
(j)
South Pacific
(i)
LLC4320
CNN
Phys. Param.
1e−9
$\overline{w'b'^z}$ (W/m^2)
Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

CNN submesoscale parameterization

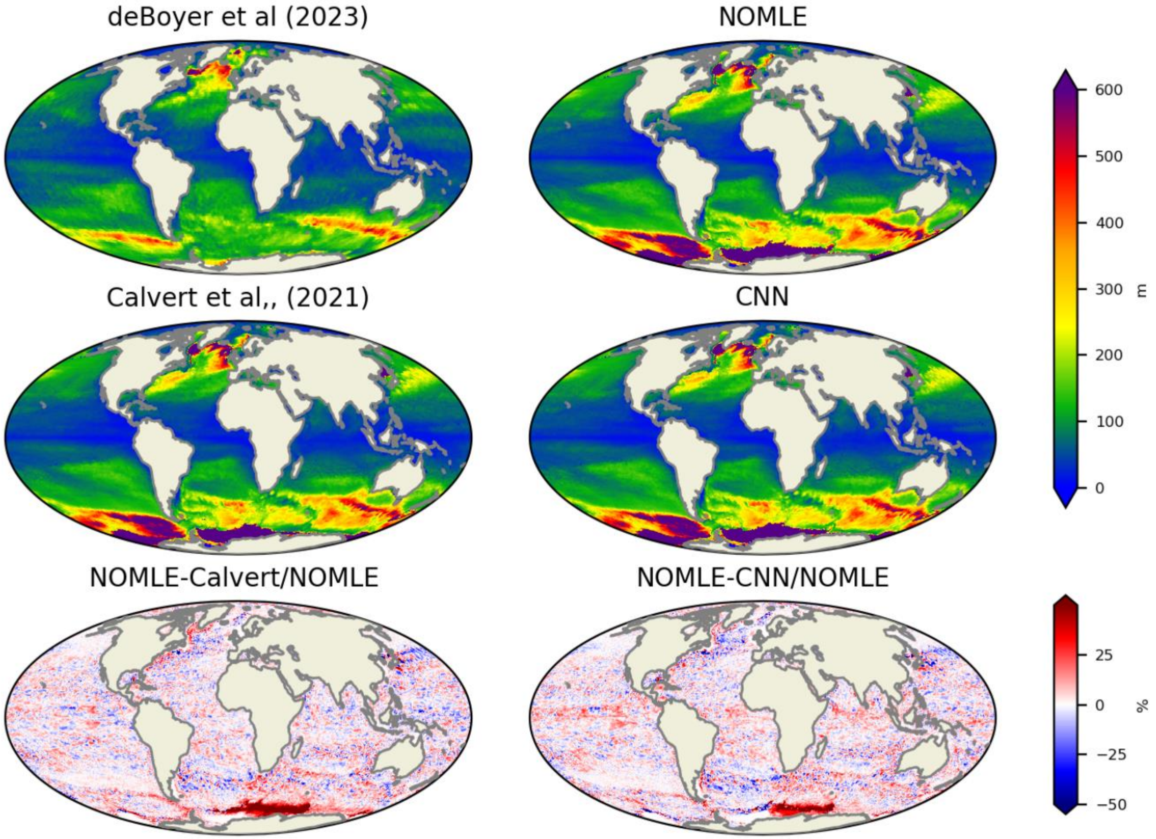
Bodner, Balwada, Zanna (2025)

Contreras et al. (in prep.)

- CNN implemented in NEMO
- Streamfunction inverted from predicted fluxes

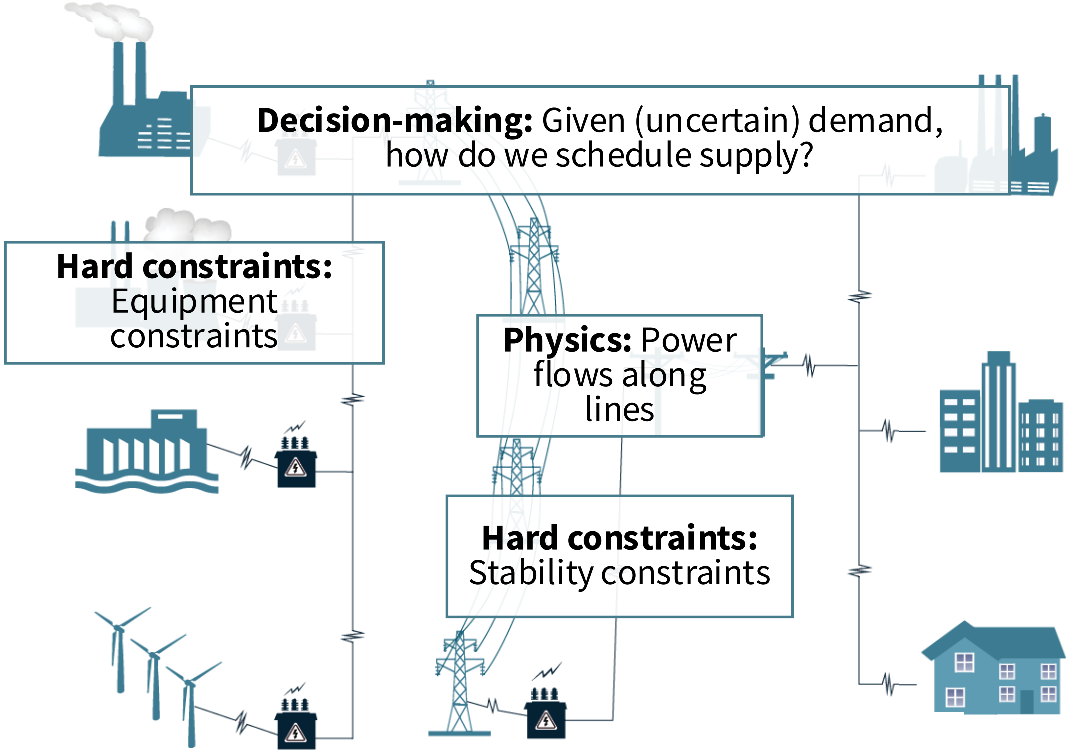
$$\Psi = \frac{\boxed{\overline{w'b'}^{z}}}{\overline{|\nabla b|}^{z}}$$

- Online performance compared with Calvert MLE parameterization reveals not much change



Power grids (example from Priya Donti)

ML with engineering constraints (power grids)



Trad. optimization & control
- Satisfies (many) constraints
- Struggles with speed / scale

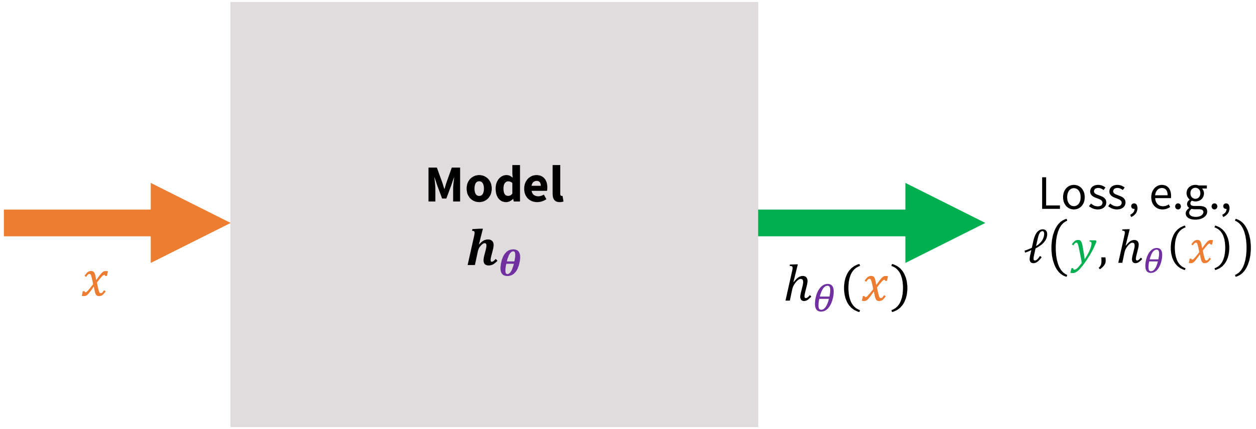
Machine learning (ML)
- Fast and scalable
- Struggles with constraints

Donti, Priya L. *Bridging Deep Learning and Electric Power Systems*. Diss. Carnegie Mellon University, 2022.

Figure adapted from: US Congressional Budget Office

Optimization-in-the-loop ML

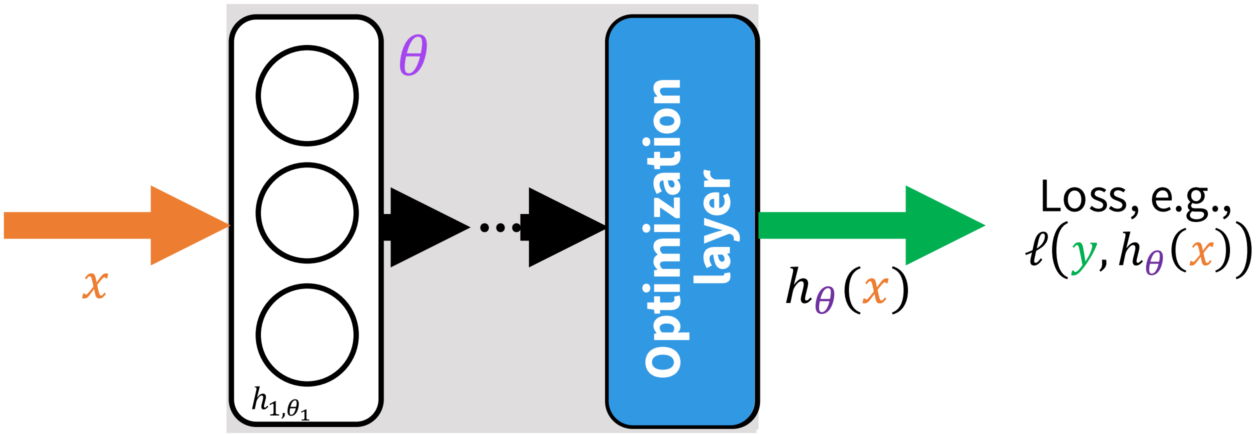
Framework for developing ML methods that incorporate knowledge of physics, hard constraints, or downstream decision-making procedures, via optimization problems



Donti, Priya L. *Bridging Deep Learning and Electric Power Systems*. Diss. Carnegie Mellon University, 2022.

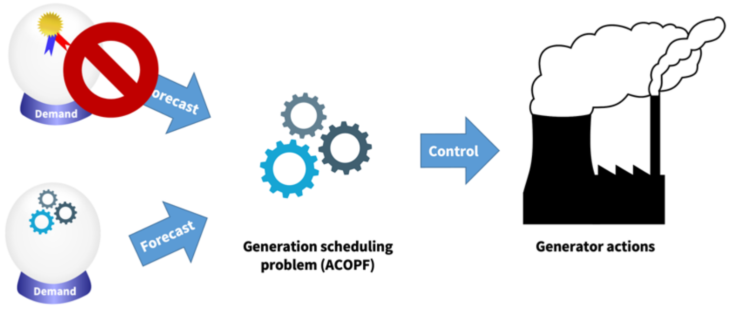
Optimization-in-the-loop ML

Framework for developing ML methods that incorporate knowledge of physics, hard constraints, or downstream decision-making procedures, via optimization problems

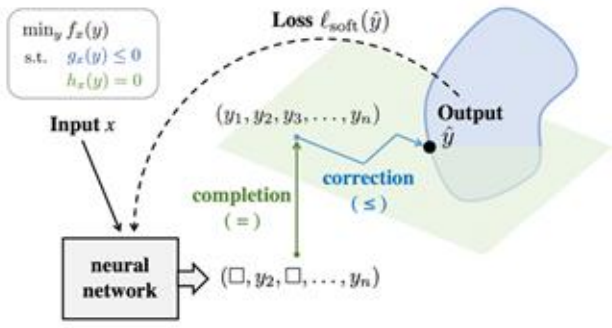


Donti, Priya L. *Bridging Deep Learning and Electric Power Systems*. Diss. Carnegie Mellon University, 2022.

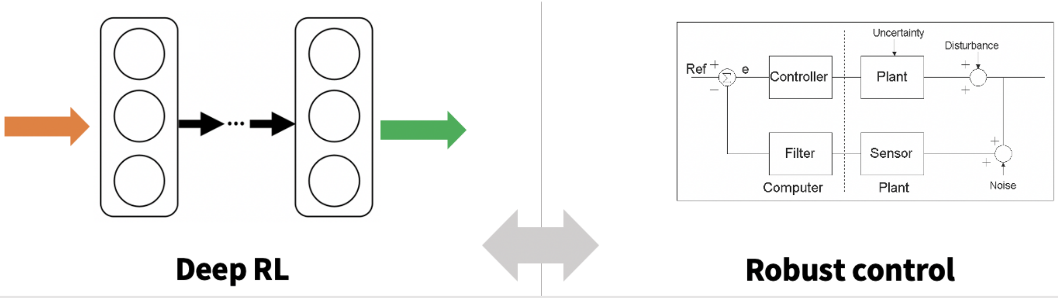
Optimization-in-the-loop ML for power systems



Decision-cognizant forecasting of supply & demand



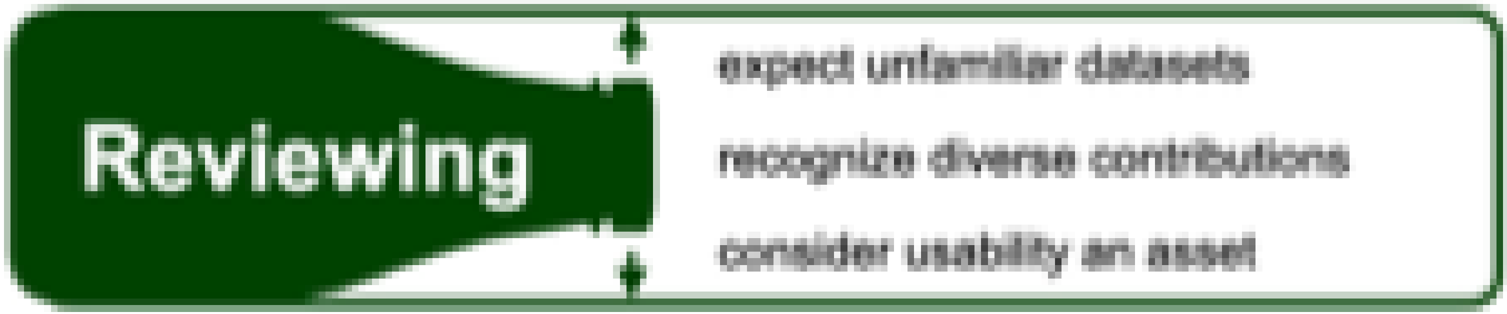
Fast, feasible approximations to power systems optimization (ACOPF, SCOPF)



Provably robust control via deep reinforcement learning (power, buildings)

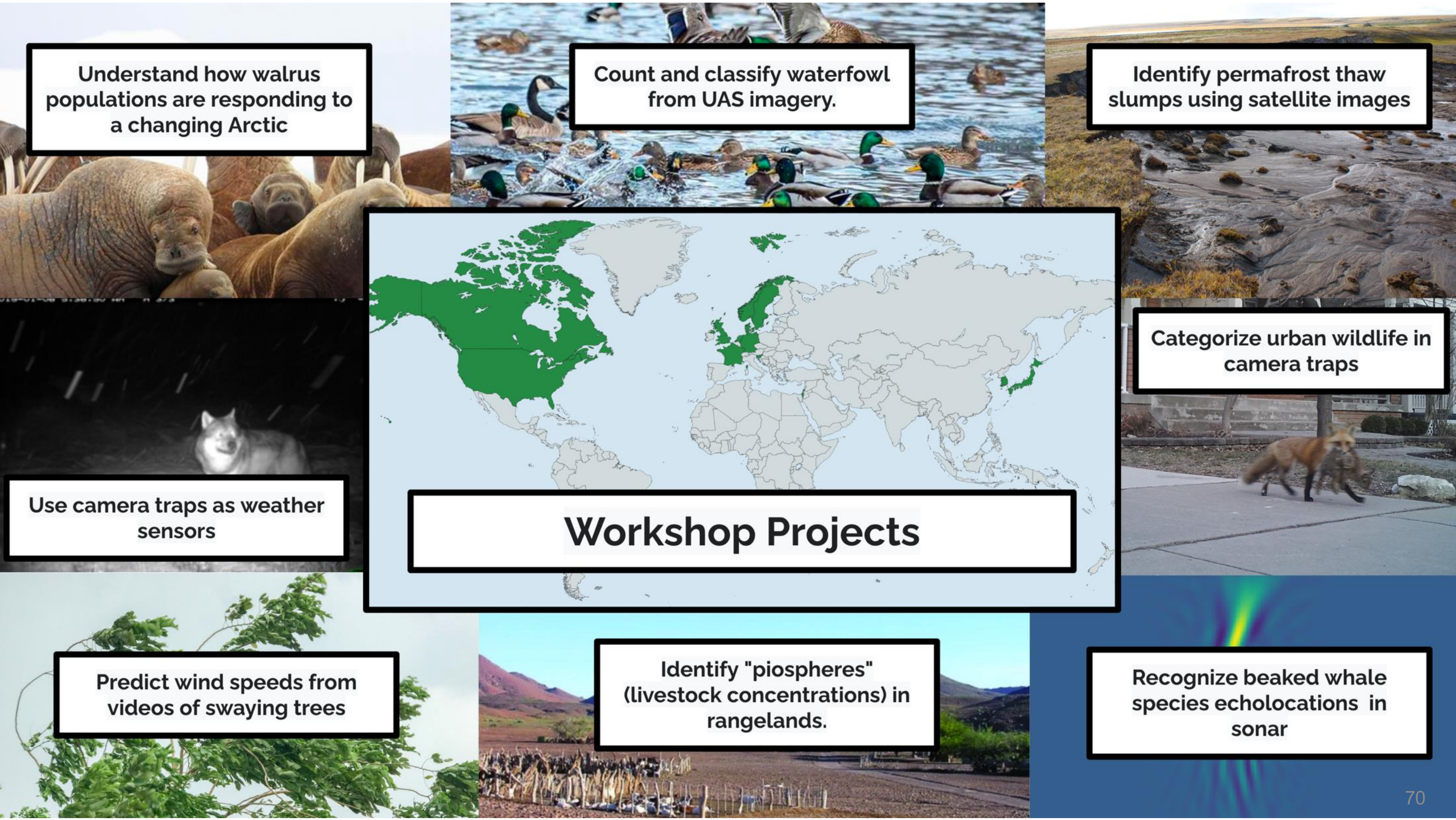
Donti, Priya L. *Bridging Deep Learning and Electric Power Systems.* Diss. Carnegie Mellon University, 2022.

How do we support this kind of research?


Reviewing
expect unfamiliar datasets
recognize diverse contributions
consider usability an asset

Expanding interdisciplinary capacity

http://cv4ecology.caltech.edu/


Understand how walrus populations are responding to a changing Arctic
Count and classify waterfowl from UAS imagery.
Identify permafrost thaw slumps using satellite images
Categorize urban wildlife in camera traps
Use camera traps as weather sensors
Workshop Projects
Predict wind speeds from videos of swaying trees
Identify "piospheres" (livestock concentrations) in rangelands.
Recognize beaked whale species echolocations in sonar

Climate applications *do* and *will* drive innovation in ML

- Diverse tasks
- Significant constraints
- Complex evaluations