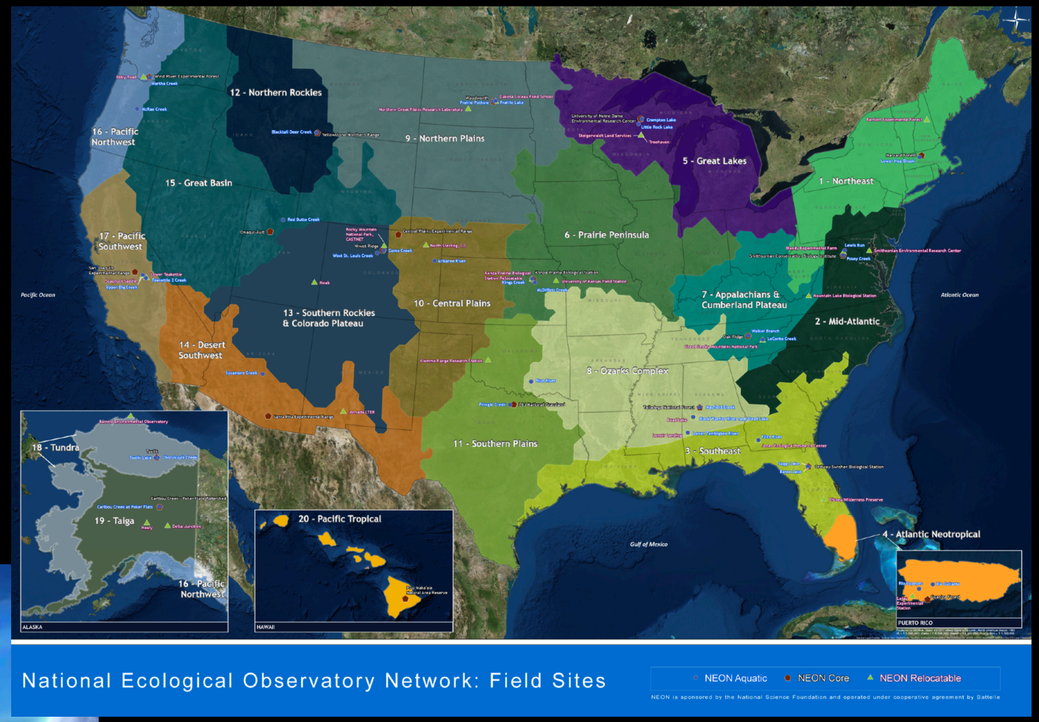
PROTOTYPING A DISTRIBUTED, ASYNCHRONOUS WORKFLOW FOR ITERATIVE NEAR-TERM ECOLOGICAL FORECASTING

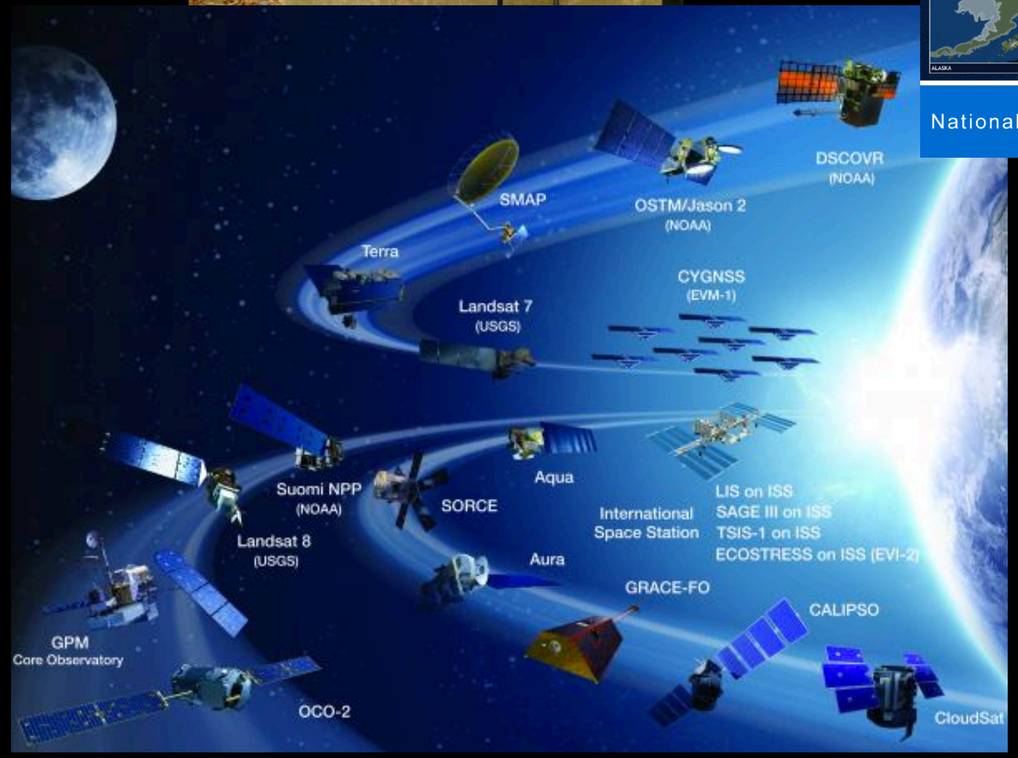
NSF 1458021, 1638577, 1926388
NASA 16-CMS16-0007, 20-CMS20-0030

Red Hat
Research Interest Group
April 4, 2023

- Michael Dietze (BU, Earth & Environment)
- Christopher Tate (Red Hat)
- Yannis Paschalidis (BU, Hariri Institute)
- BU SAIL team (Will Tomlinson, Greg Frasco, Jeff Simeon)

REAL TIME SCIENCE



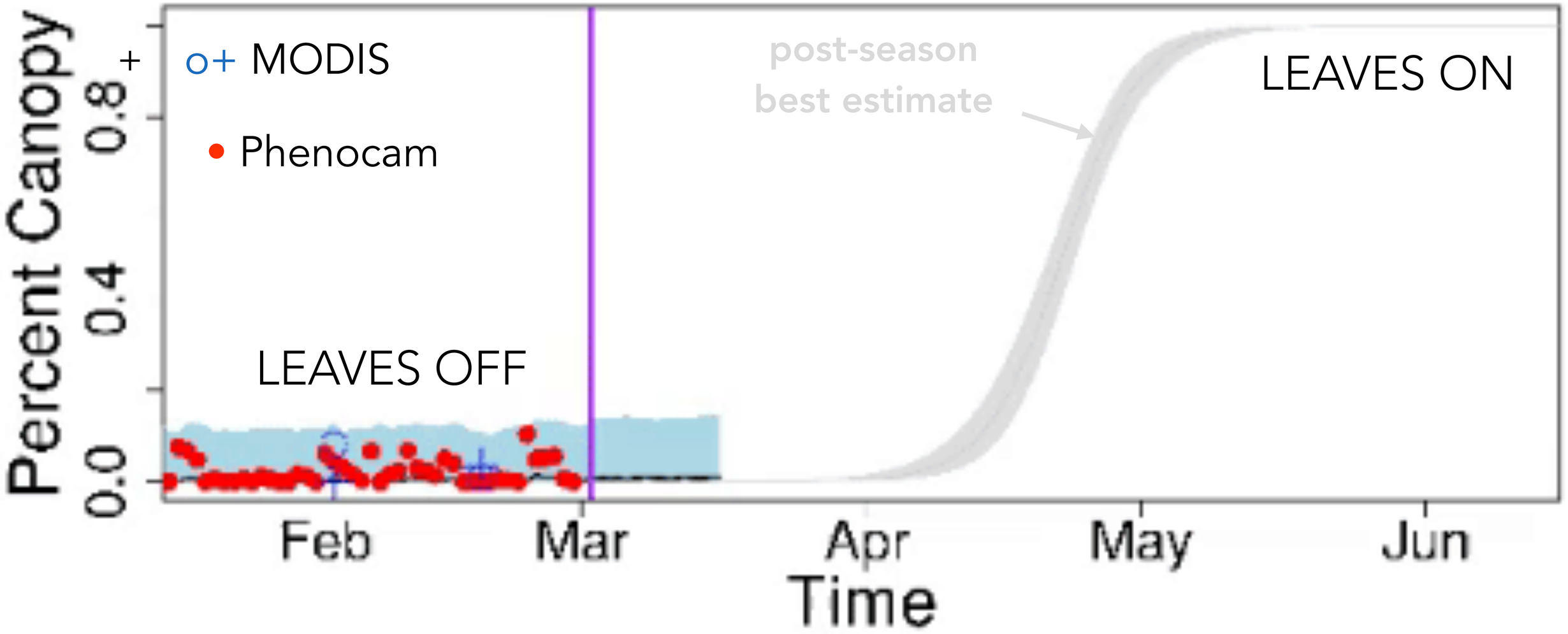


USA npn National Phenology Network
Taking the Pulse of Our Planet

eBird

iNaturalist
Explore Learn Record

GLOBAL SOUNDSCAPES


o+ MODIS
• Phenocam
post-season
best estimate
LEAVES ON
LEAVES OFF
Percent Canopy
0.0
0.4
0.8
Feb
Mar
Apr
May
Jun
Time

CLIMATE CHANGE

Stationarity Is Dead: Whither ~~Water~~ Environmental Management?

P. C. D. Milly,[1]* Julio Betancourt,[2] Malin Falkenmark,[3] Robert M. Hirsch,[4] Zbigniew W. Kundzewicz,[5] Dennis P. Lettenmaier,[6] Ronald J. Stouffer[7]

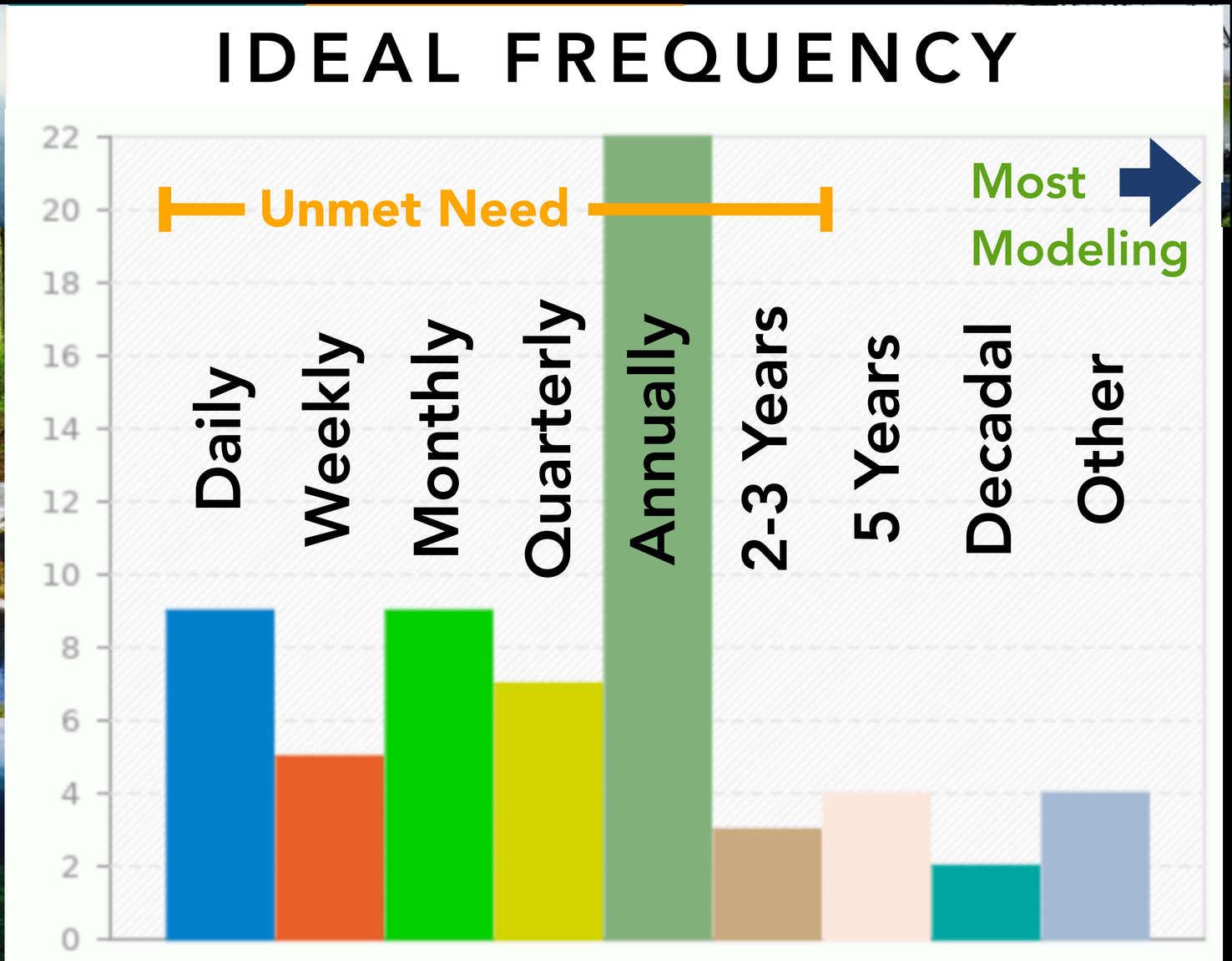
Climate change undermines a basic assumption that historically has facilitated management of water supplies, demands, and risks.

Science 2008

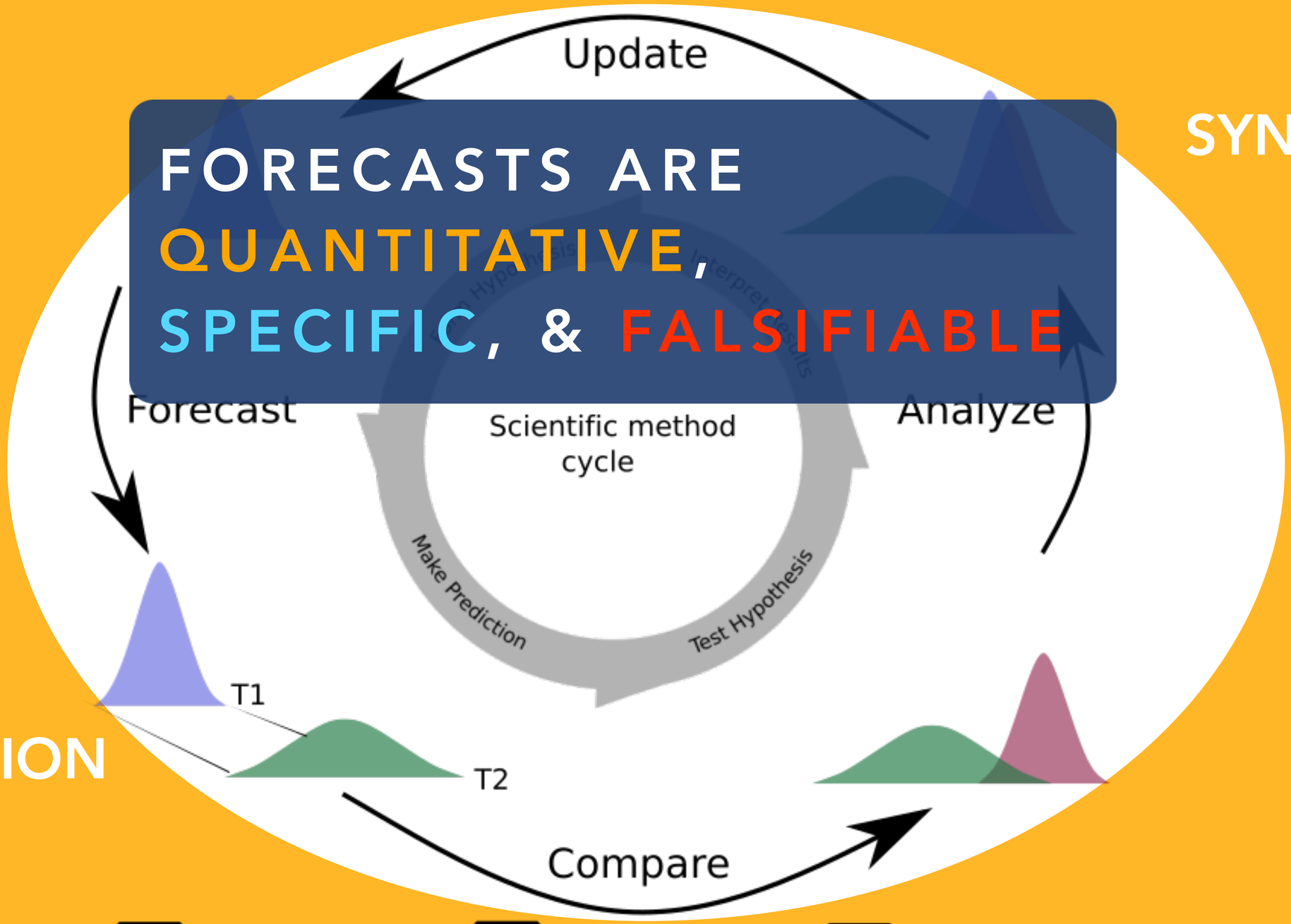
DECISIONS ARE ABOUT THE FUTURE

NASA Carbon Monitoring **Stakeholder** Survey

data courtesy Edil Sepulveda Carlo



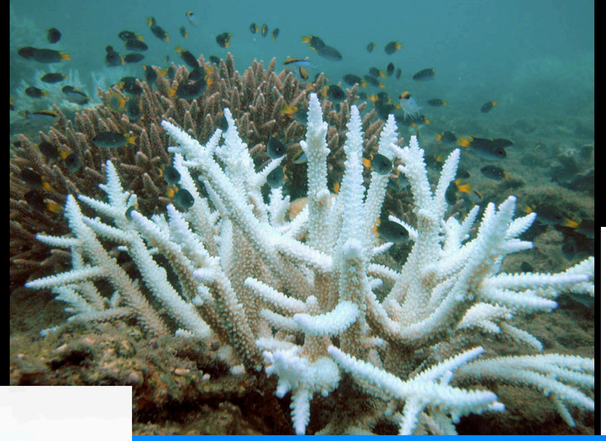
FORECASTS ARE
QUANTITATIVE,
SPECIFIC, & FALSIFIABLE

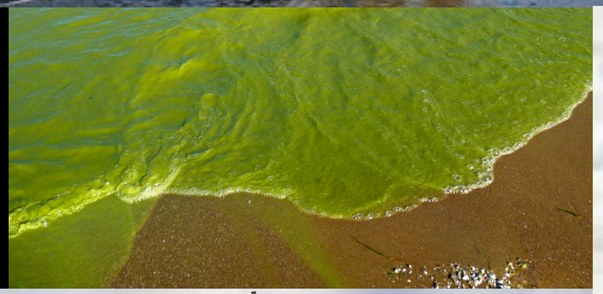
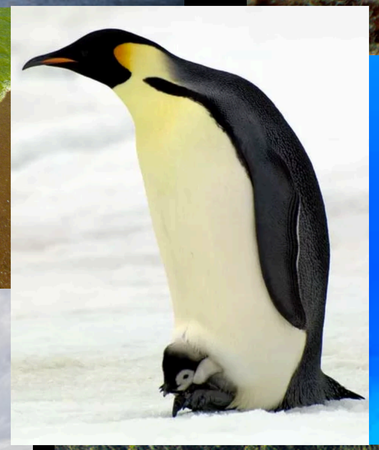


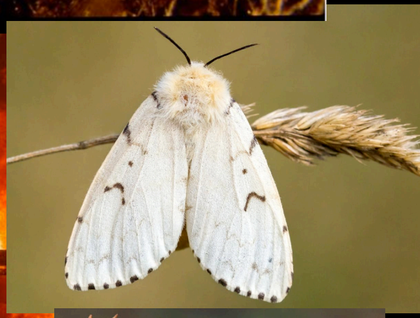
SYNTHESIS

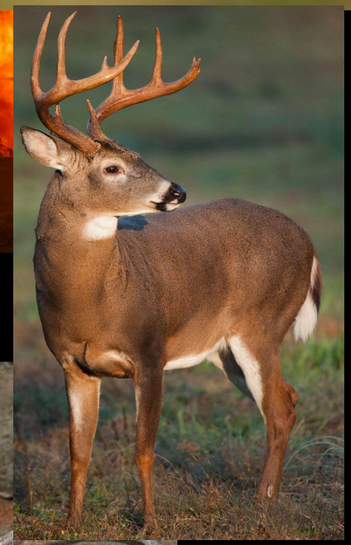
PREDICTION

Figure: T. McCabe
Dietze et al 2018 PNAS





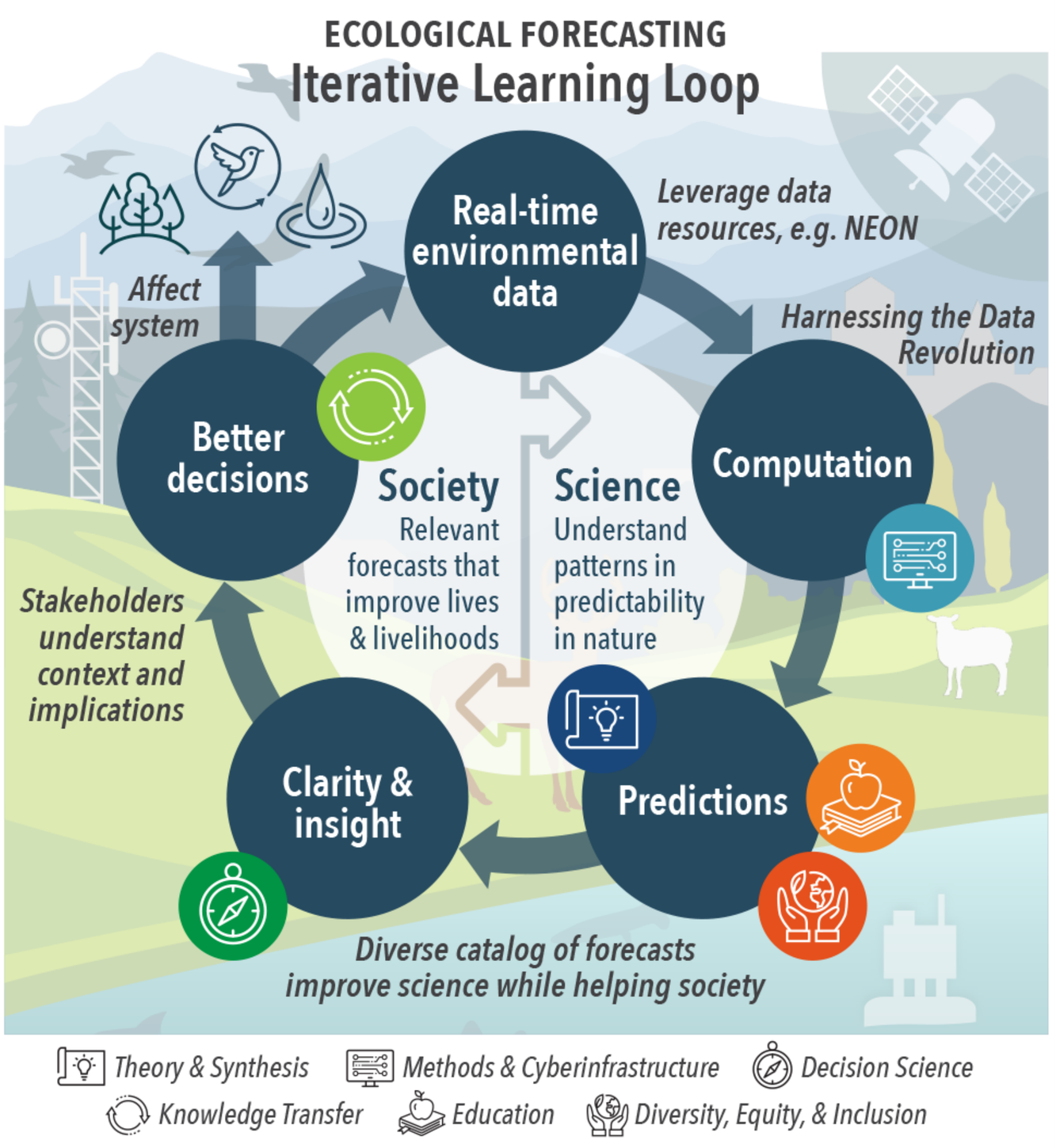


ecoforecast.org

@eco4cast

Ecological Forecasting Initiative

UNDERSTAND · MANAGE · CONSERVE


ECOLOGICAL FORECASTING
Iterative Learning Loop
Real-time environmental data
Leverage data resources, e.g. NEON
Harnessing the Data Revolution
Computation
Predictions
Clarity & insight
Better decisions
Affect system
Stakeholders understand context and implications
Society
Relevant forecasts that improve lives & livelihoods
Science
Understand patterns in predictability in nature
Diverse catalog of forecasts improve science while helping society
Theory & Synthesis
Methods & Cyberinfrastructure
Decision Science
Knowledge Transfer
Education
Diversity, Equity, & Inclusion

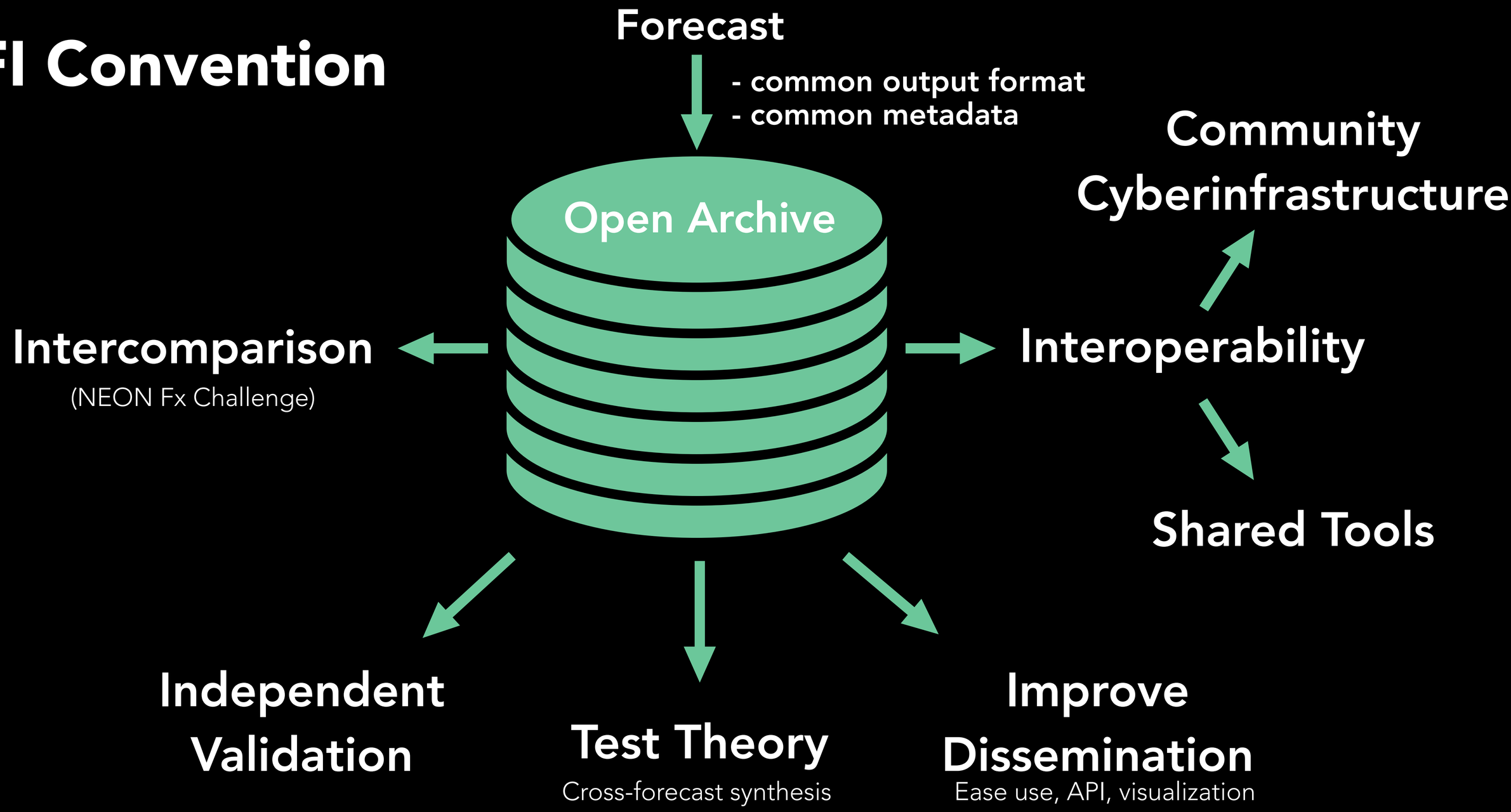
CYBERINFRASTRUCTURE BOTTLENECK

- Existing forecast workflows have been developed indepently
  - “Boutique” solutions
  - Redundancy
  - Barrier to entry (technical, financial)
  - Costly to sustain
- Need: Community Tools
  - Scalable, Reusable, Easier to learn, Cheaper to maintain

NEON FORECAST CHALLENGE

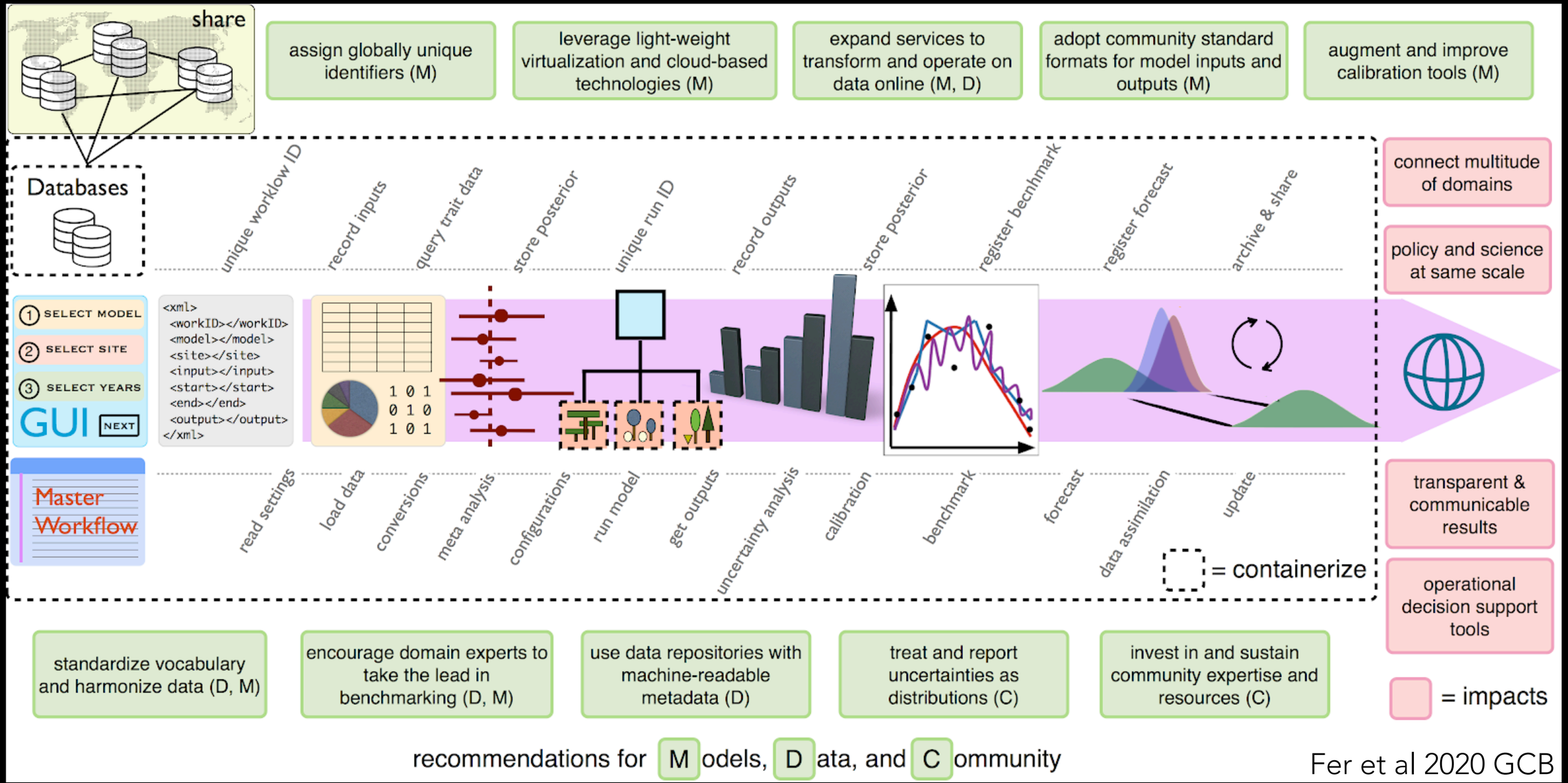
- Goal: Predict NEON site-level observation before they're collected
- Open to all participants and approaches (statistical, machine learning, mech., etc)
- Standard format, CI, "null" models: persistence, historical means
- Round 1 & 2: ~7000 forecasts from ~100 teams (11 courses)
- Round 3: 2023 https://ecoforecast.org/efi-rcn-forecast-challenges/
  - Tick population dynamics
  - Aquatic DO and water temperature
  - Phenocam vegetation phenology
  - Ground beetle biodiversity
  - Terrestrial carbon and water flux, soil moisture

EFI Convention



Dietze et al 2021 EcoEvoRxiv https://doi.org/10.32942/osf.io/9dgtq

BEYOND ECOSYSTEM MODELING:
A ROADMAP TO COMMUNITY CYBERINFRASTRUCTURE FOR ECOLOGICAL DATA-MODEL INTEGRATION



Fer et al 2020 GCB

RED HAT COLLABORATORY PROJECT OBJECTIVES

1. High-level system design & test deployment of a core CI stack

2. Minimal cloud-based test case using PEcAn workflow: 1 model forecasting 1 site with 1 input & 1 constraint

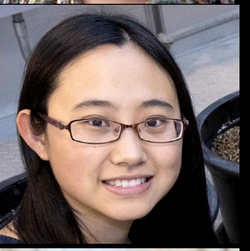
3. Scaling up: Asynchronous, distributed, event-driven scheduling across many sites, inputs, data constraints.

NASA CMS CARBON REANALYSIS

Hamze Dokoohaki
UIUC

Shawn Serbin
BNL

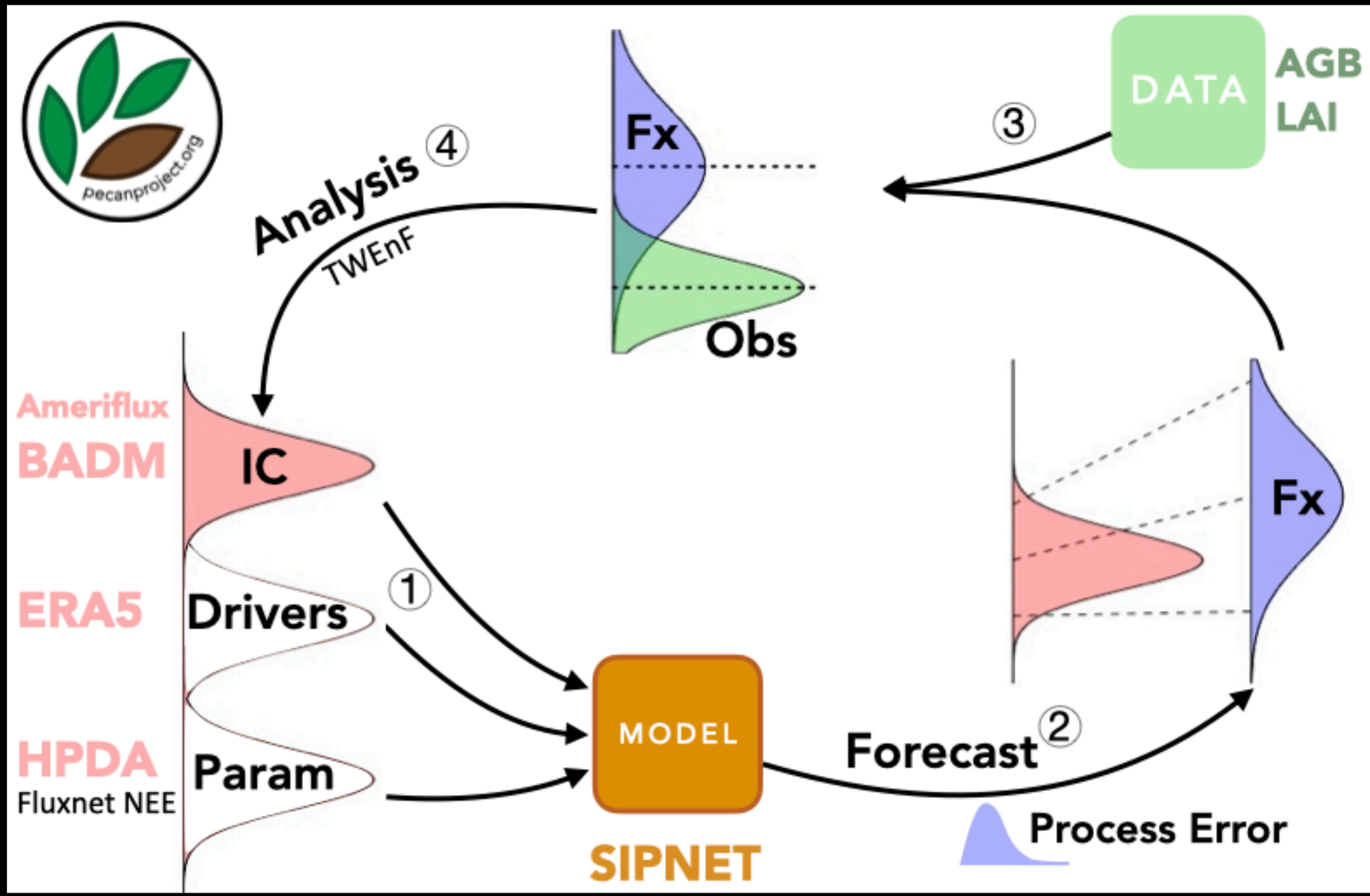
Bailey Morrison
UC Merced



Qianyu Li
BNL

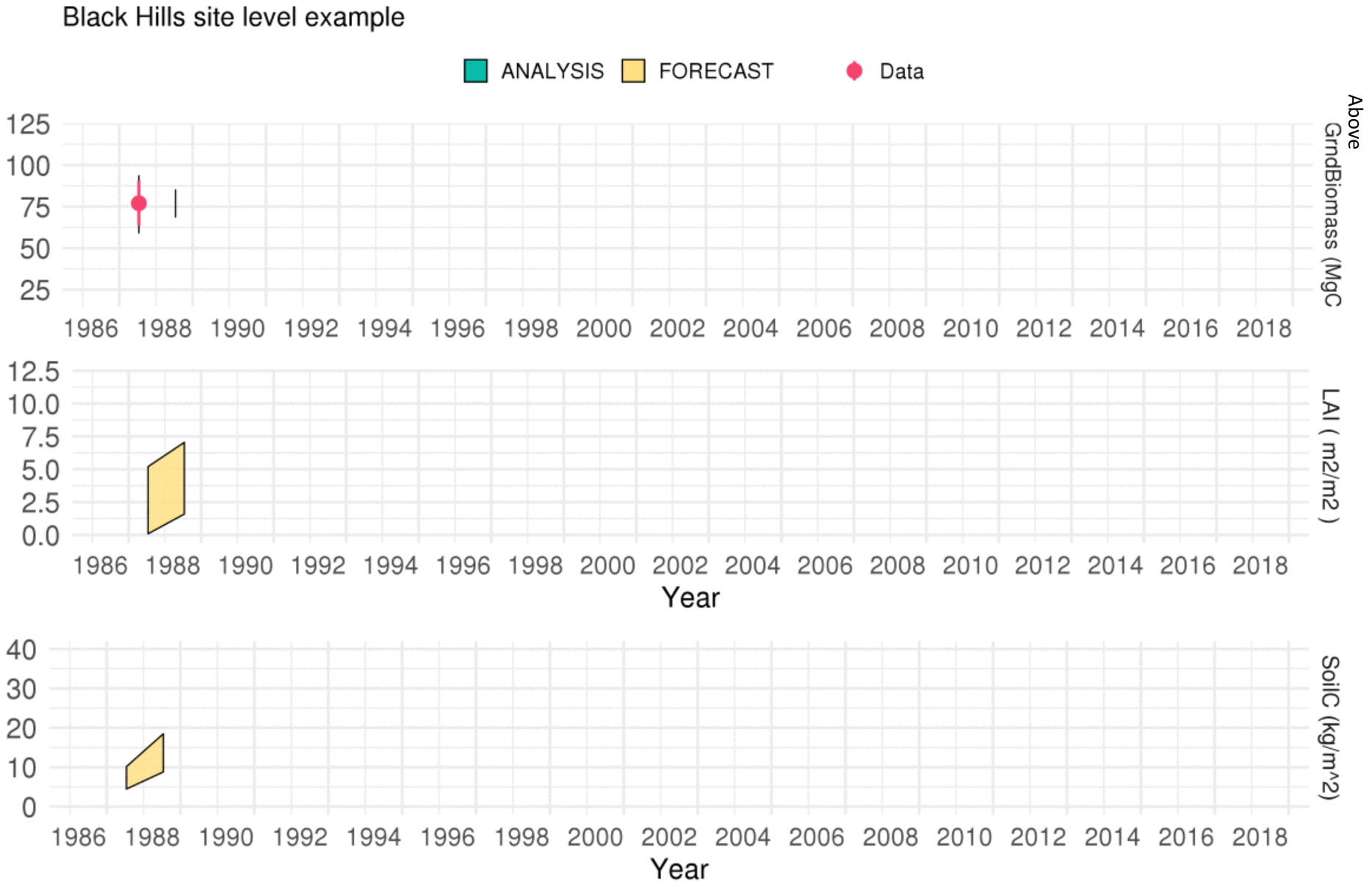
Arlyn Andrews
NOAA

Helgeson & Zhang (BU)



Carbon Pools

MODEL ENSEMBLE = COVARIANCE

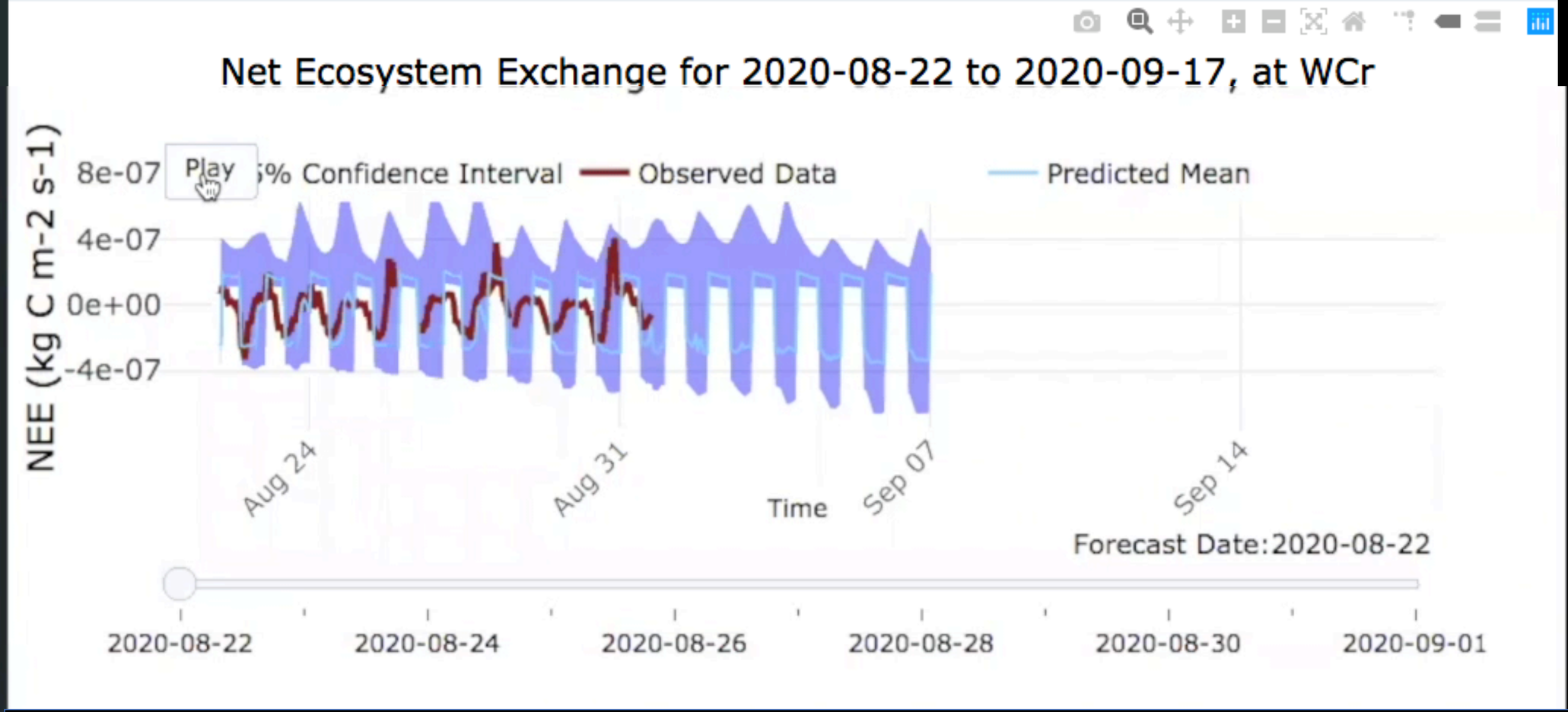


Flux Dashboard

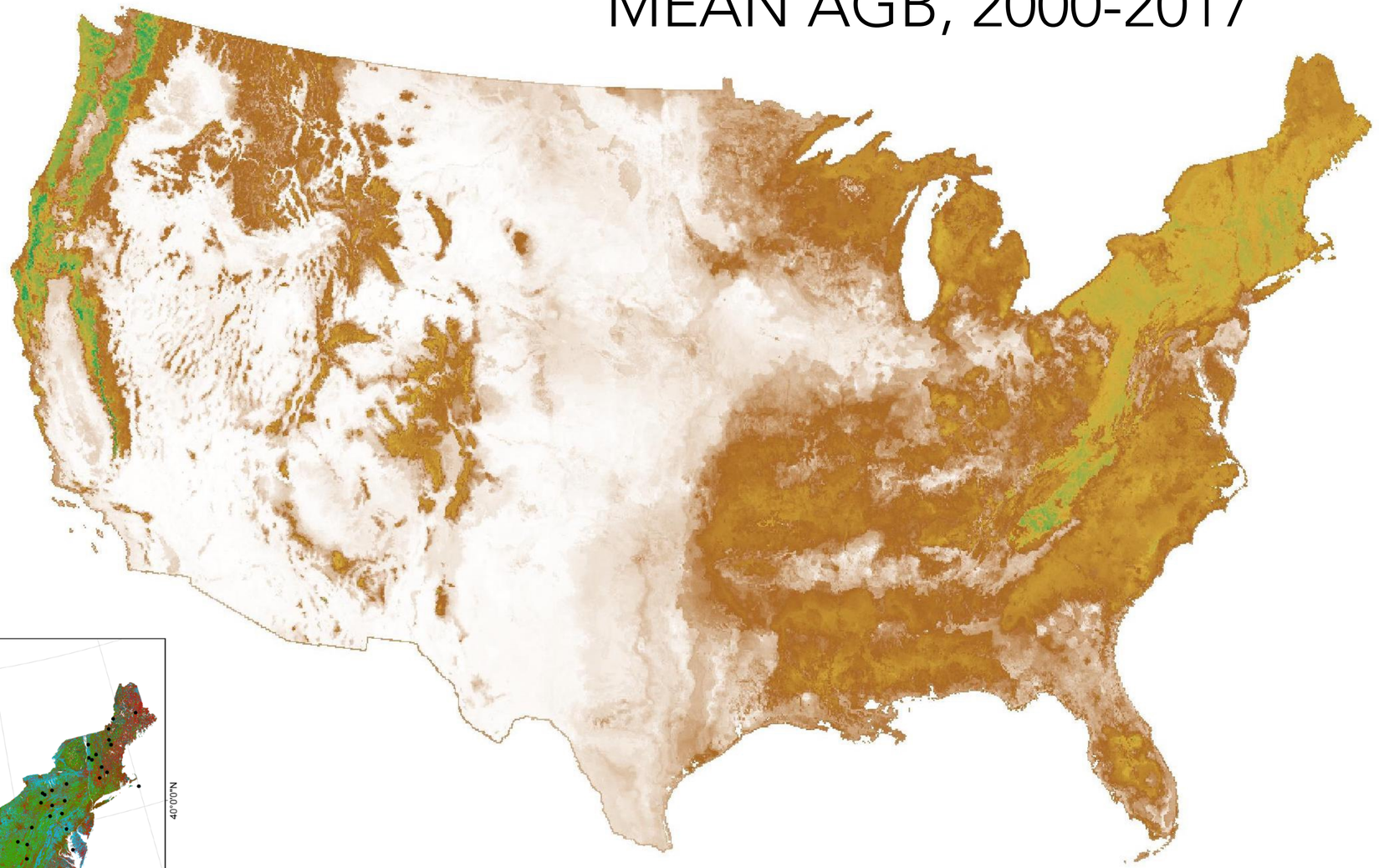
Select Site:

Willow Creek

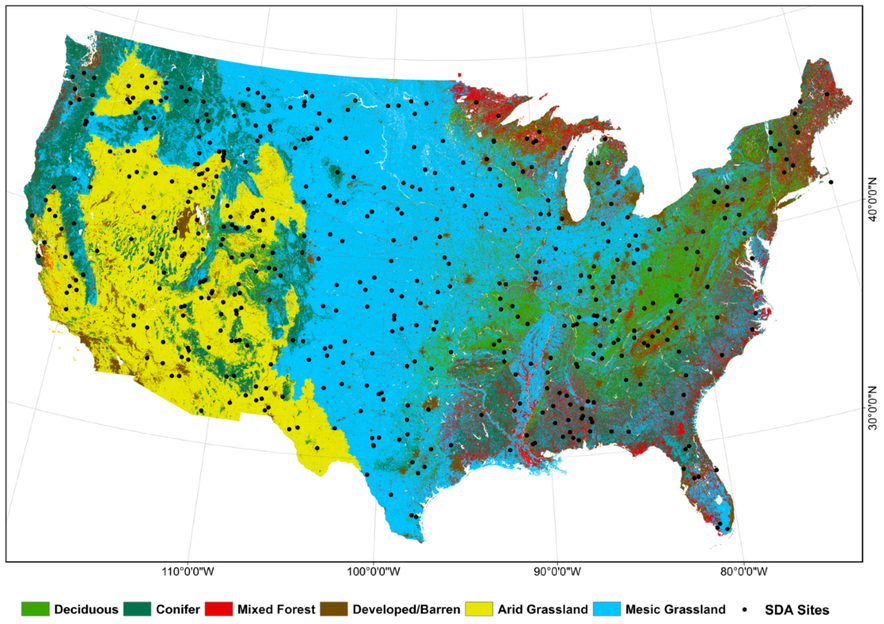
- Site Map
- Forecast
- Model Performance
- Model by Start Date
- NOAA GEFS



MEAN AGB, 2000-2017



CONUS PFT Classification

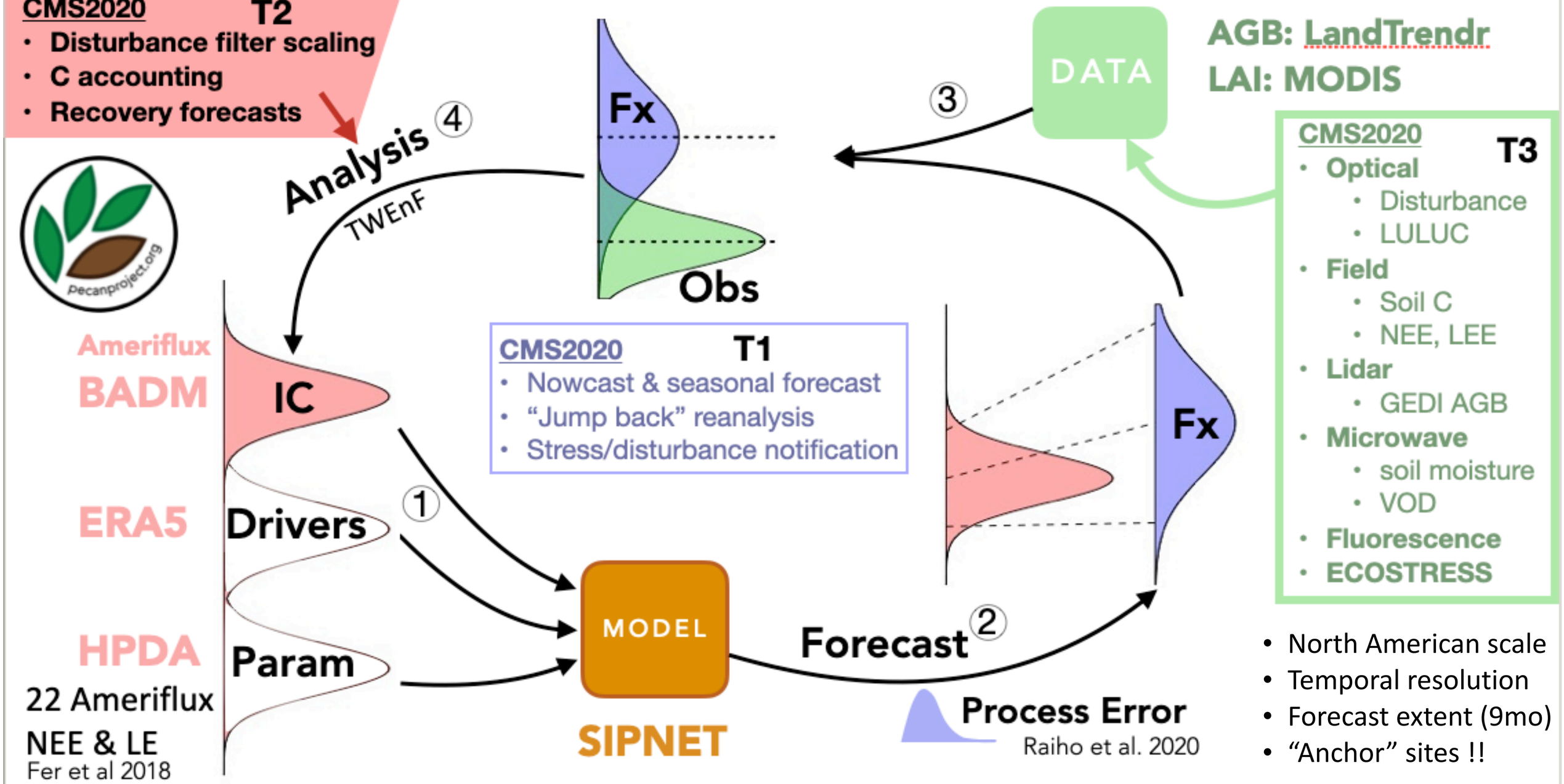


Dokoohaki et al 2022 GMD

National Aeronautics and Space Administration

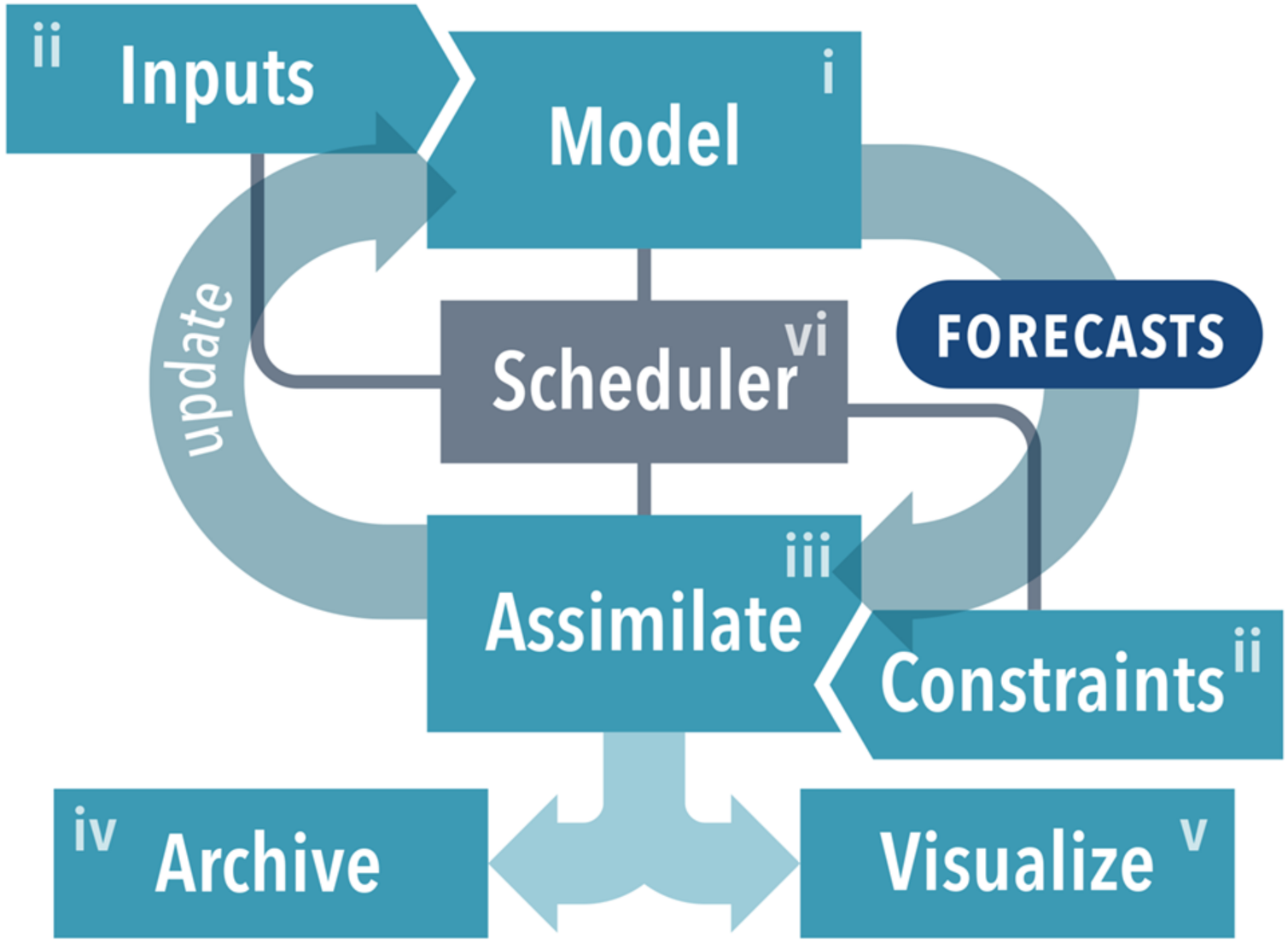
CMS CARBON MONITORING SYSTEM

CMS2020 **T2**
- Disturbance filter scaling
- C accounting
- Recovery forecasts



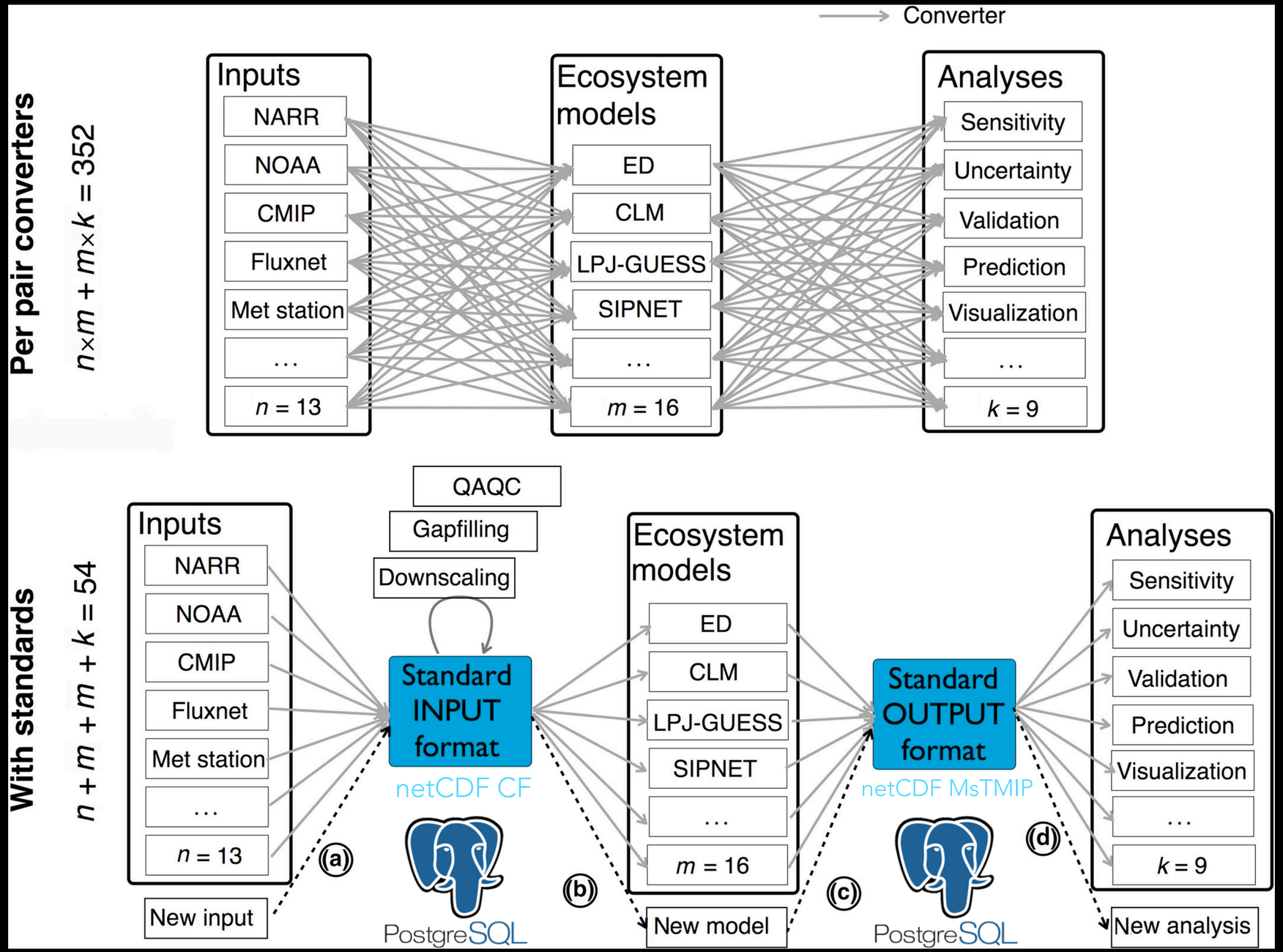
- North American scale
- Temporal resolution
- Forecast extent (9mo)
- "Anchor" sites !!

Generic Ecological Forecasting Workflow

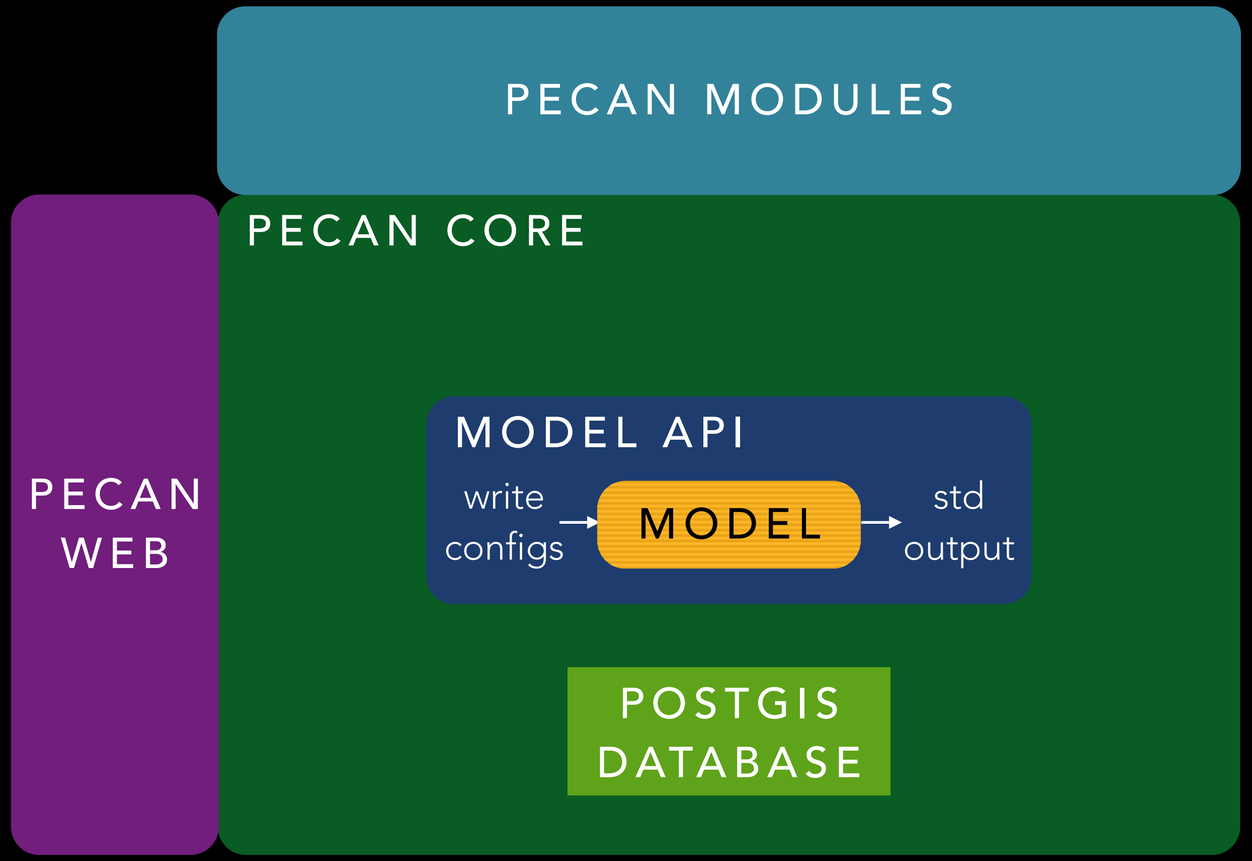


- Iterative
  - Prep inputs & data constraints
  - Run model ensemble
  - Run data assimilation
  - Distribution (vis, API, alerts)
  - Adaptive monitoring [beta]
  - Integration testing
- Occasional
  - Initialization
  - Model improvement
  - Calibration
  - Validation
  - Scenario-based projections
  - Uncertainty analysis
  - Reanalysis
  - Synthesis/Analysis

HOW DO YOU HANDLE INPUTS?



https://github.com/pecanproject/pecan/



Standardized inputs and outputs

Provenance: Transparent & Repeatable

Accessible interface

Reusable tools for ingest, execution, analysis, visualization