Leveraging diversity in all dimensions to improve the prediction of multiscale meteorology and its impacts

Katherine (Kate) Evans

Oak Ridge National Laboratory

ORNL is managed by UT-Battelle LLC for the US Department of Energy

U.S. DEPARTMENT OF ENERGY

Thank you to my US Dept. of Energy and Air Force sponsors

Outline:

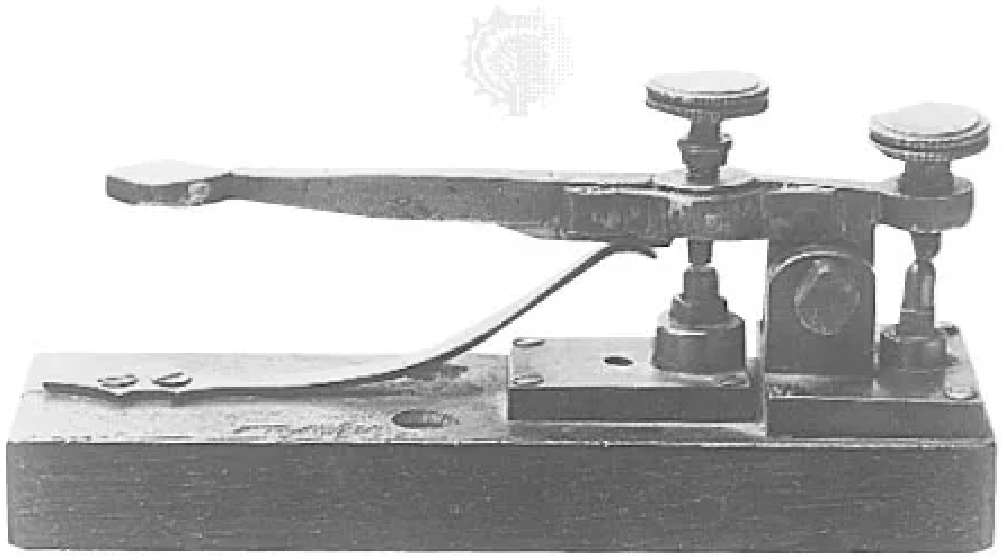
We are making good use of HPC for weather and climate prediction

A convergence of diverse science domains connected to weather using HPC will provide new knowledge and impact

We are already moving in this direction: We are making strides to combine scales, physics, and models

Climate/weather scientists have been early adopters of data collection for forecasting - since the inception of telegraphy

- Electrical telegraphy was used as early as the mid 1800's.
- Provided experimental data to confirm and refute theories of meteorology
- Could do very basic forecasts
- Data security was an issue even then, as accurate weather forecasts provided an advantage in the battlefield. Many used codebooks.



Some good discussion of the drama of these developments in Peter Moore. *The Weather Experiment:* The Pioneers Who Sought to See the Future. New York: Farrar, Straus & Giroux, 2015.

Climate/weather modelers have also been very early adopters of HPC

- First coupled ocean–atmosphere general circulation model, eventually used for global warming experiments in 1975
- Used 1200 CPU hours (50 days) on the UNIVAC 1108 @ GFDL

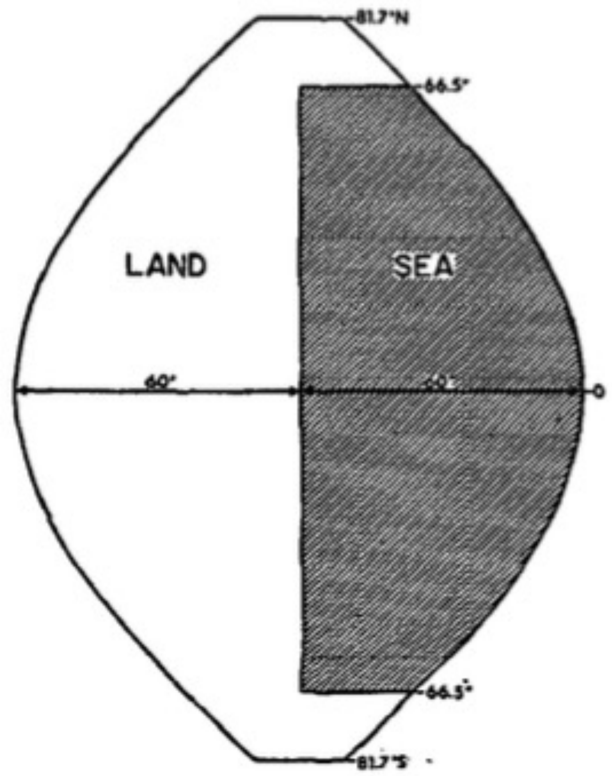


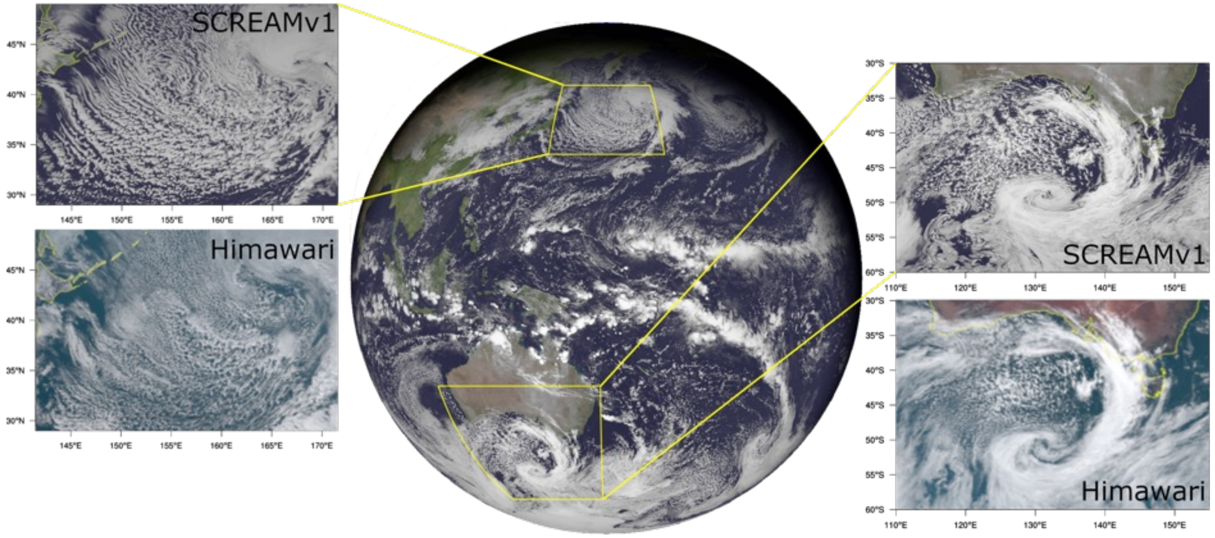
FIG. 1. Ocean-continent configuration of the model.

Manabe and Bryan, JAS, 1969

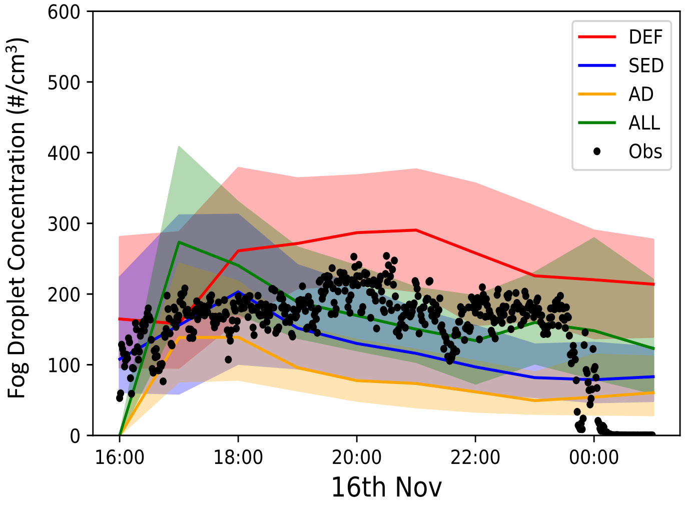
Image courtesy of Computer History Museum. http://www.computerhistory.org/collections/catalog/102621838

- 9 atmosphere levels and 5 ocean levels
- Continental boundary of atmosphere used over land as seen above
- Ocean has an ice pack!

Today: weather and climate modelers provide a host of accurate and reproducible science

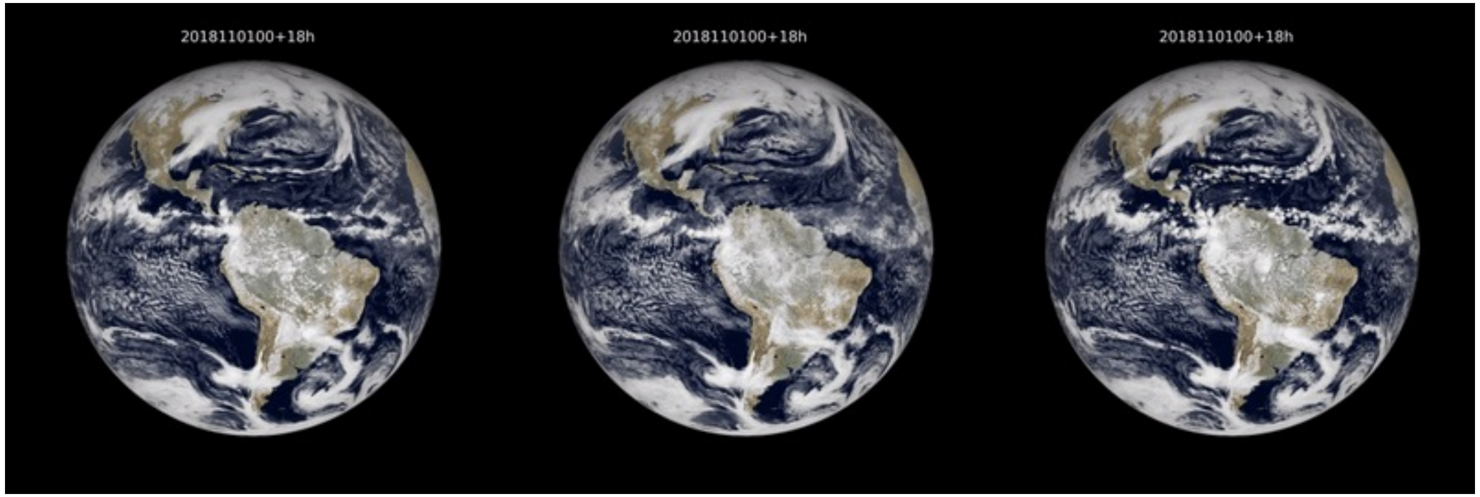


E3SM SCREAM climate 3.25km simulation on Frontier >1 SYPD (shown 2 days after start for comparison)



Above: Several simulations using the Met Office's UM with several different cloud droplet methods compared to observations, to determine the one that best represents fog

Right: ECMWF full season 1.4km (left) and 9km operational simulations run on Summit/OLCF. Animation courtesy of Philippe Lopez/ECMWF



Going forward, our stakeholders need answers about weather impacts

- Water cycle
- Oceans
- Sea ice and ice sheets
- Land surface
- Subsurface
- Plants/Ecosystem/Soils

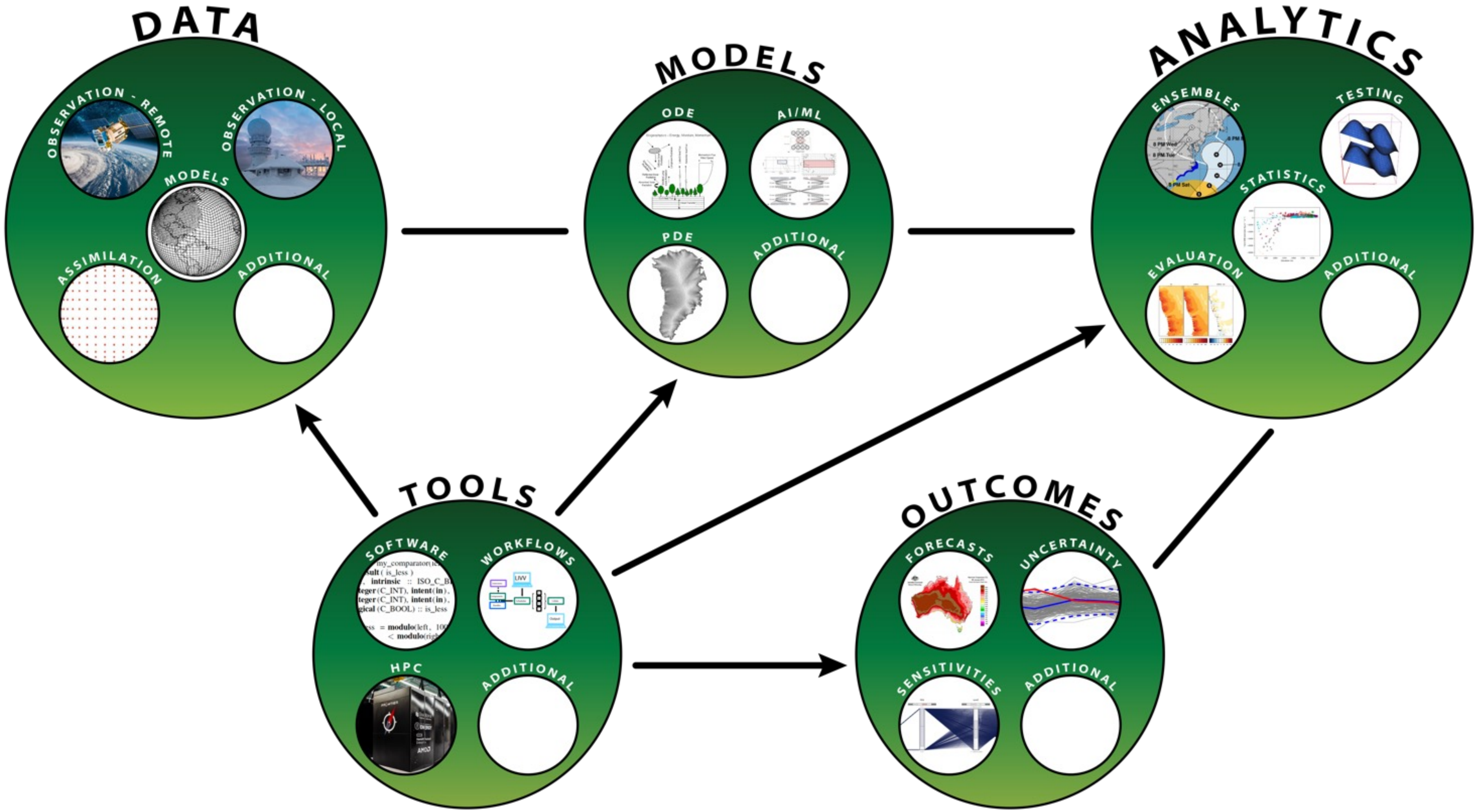
- Survival
  - Flooding
  - Heat waves
  - Wind events
- Disease propagation
- Habitat: sea level rise, storms, drought
- Air/water quality

- Electrical grid
- Buildings/Structures
- Urban infrastructure
- Transportation
- Agriculture

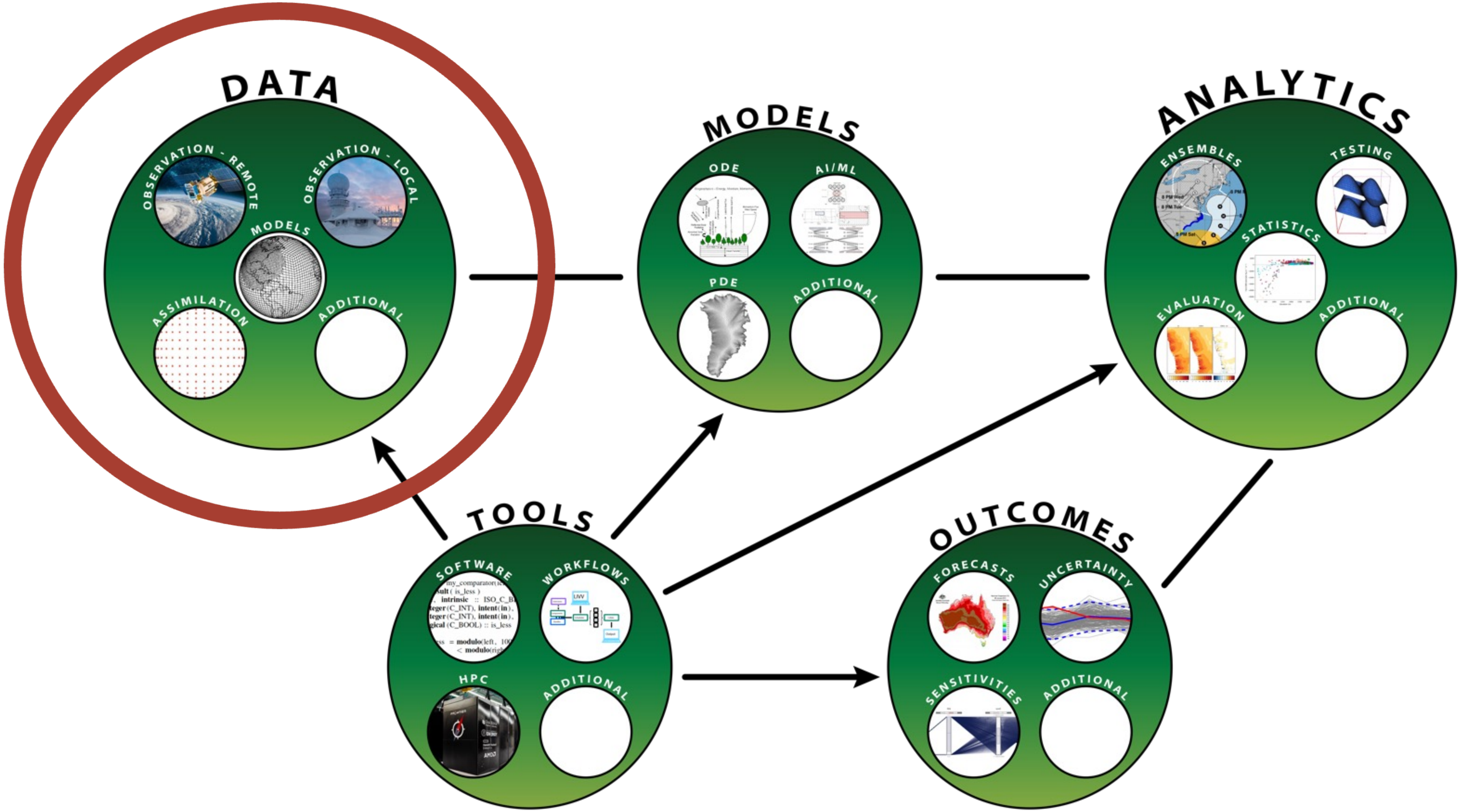
Vision: create an interacting, robust, flexible, modular computing environment of diverse components that allows us to ask and answer science questions

Will the melting Greenland ice sheet alter the MOC	We need climate models with good atmosphere and ocean coupling, high resolution and basin wide ocean data, GIS models with accurate streamflow, etc.
Will malaria spread to new areas of the globe?	We need high resolution climate models, downscaled, and/or regional models with precipitation/moisture and elevation data and create "ponding" and other mosquito habitat conditions metrics and couple to population and urban models
Will climate change induce changes in transportation?	We need climate models with urban scale inputs and resolution, heat index with humidity data, weather extremes, coupled to population data, energy costs, etc.

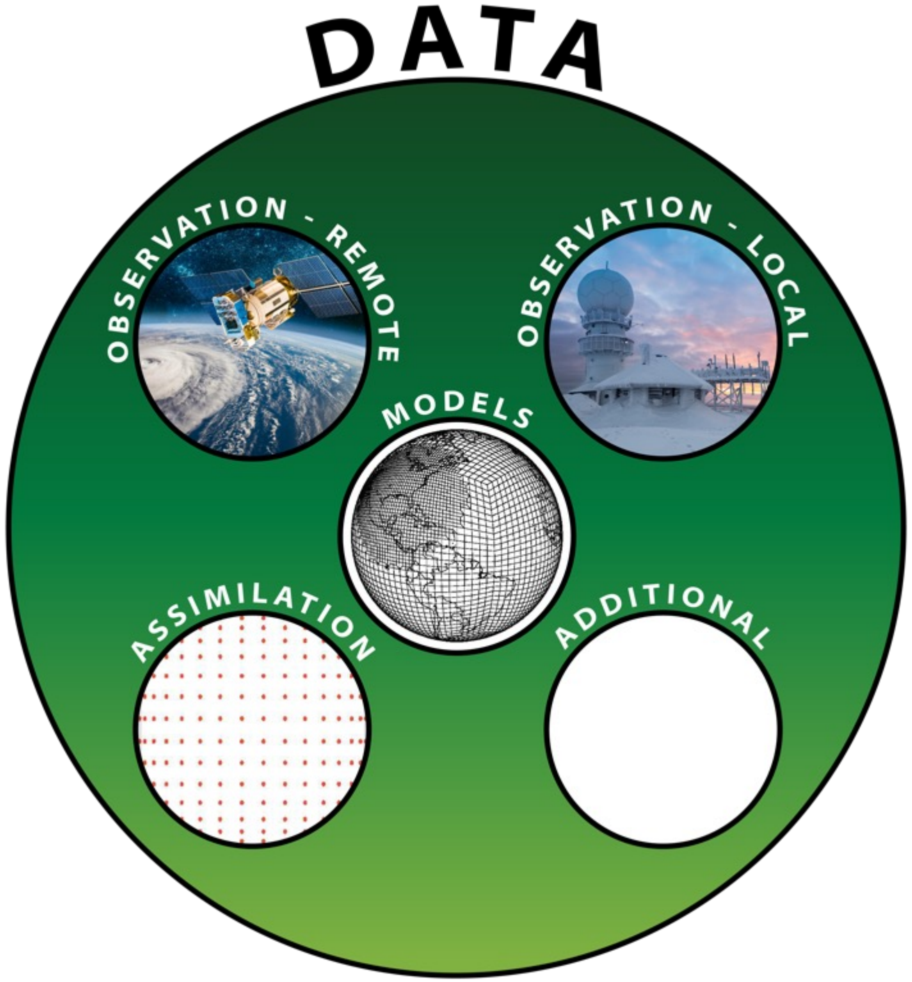
What could this environment this look like?



Diversity of the experimental setup



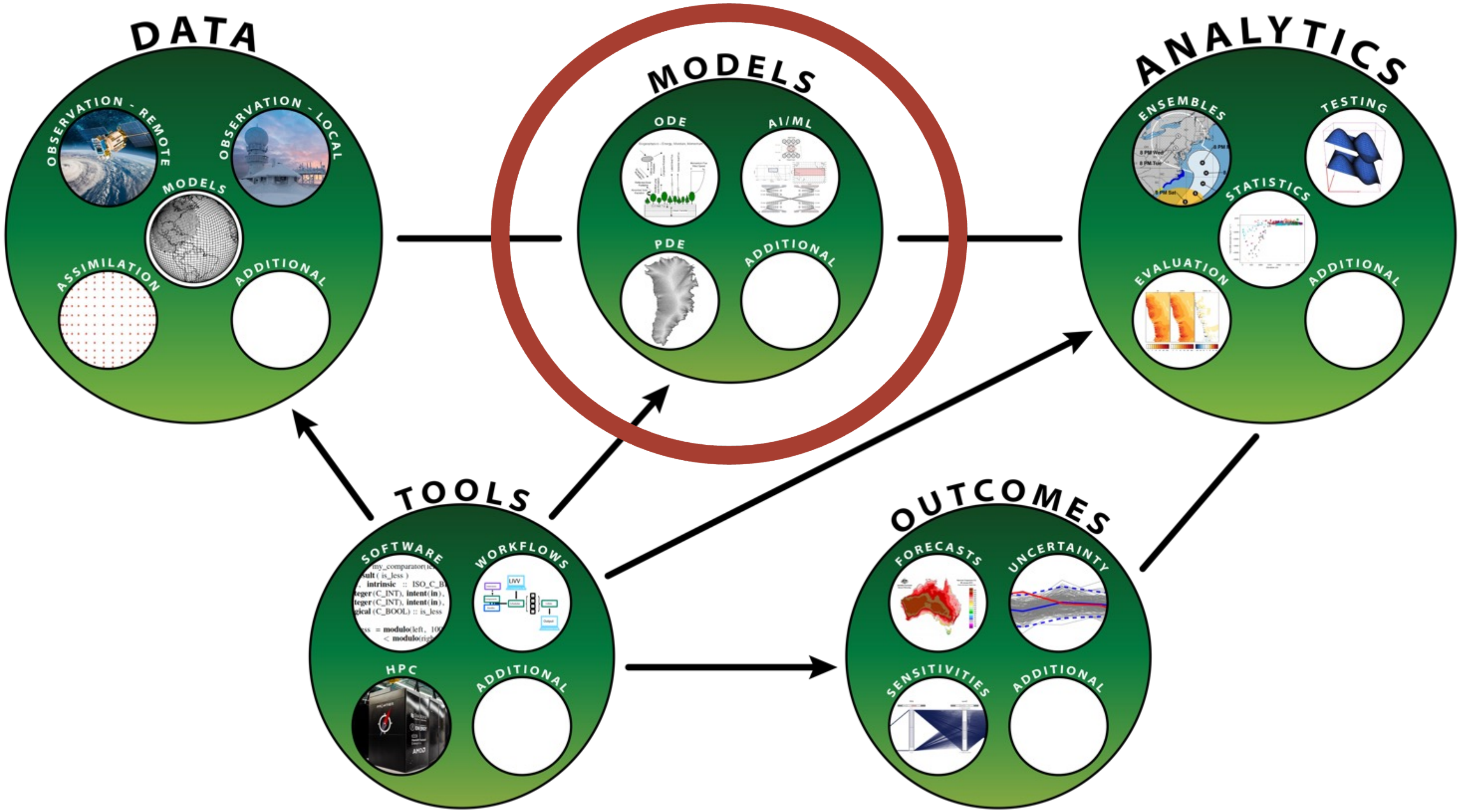
Diversity of the experimental setup



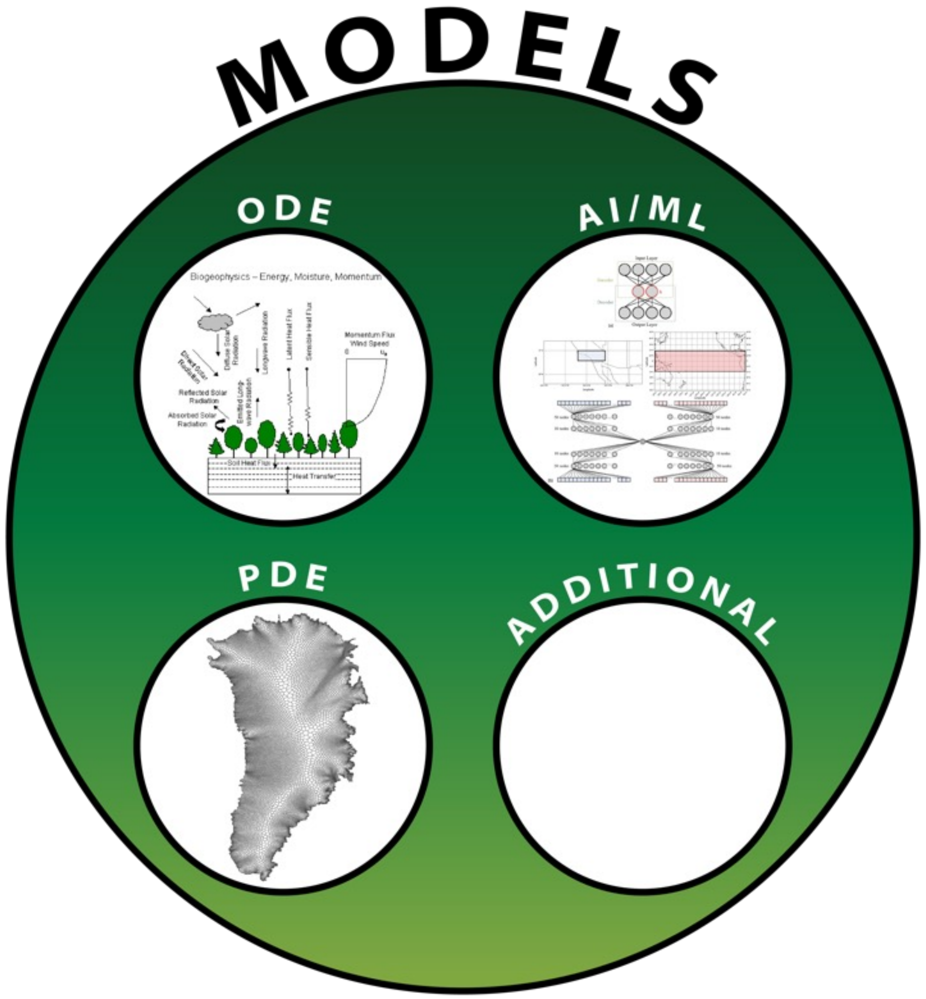
- Gather observational data from more and diverse sources
  - Personal weather stations
  - Social media
  - Generate missing data using ML methods
- More diverse and better quality model data (resolution, accuracy)
- Improved assimilated data
  - From more and unique, non-gaussian distributed sources
  - Better assimilation methods (using models) with existing data

OAK RIDGE National Laboratory

Diversity of models



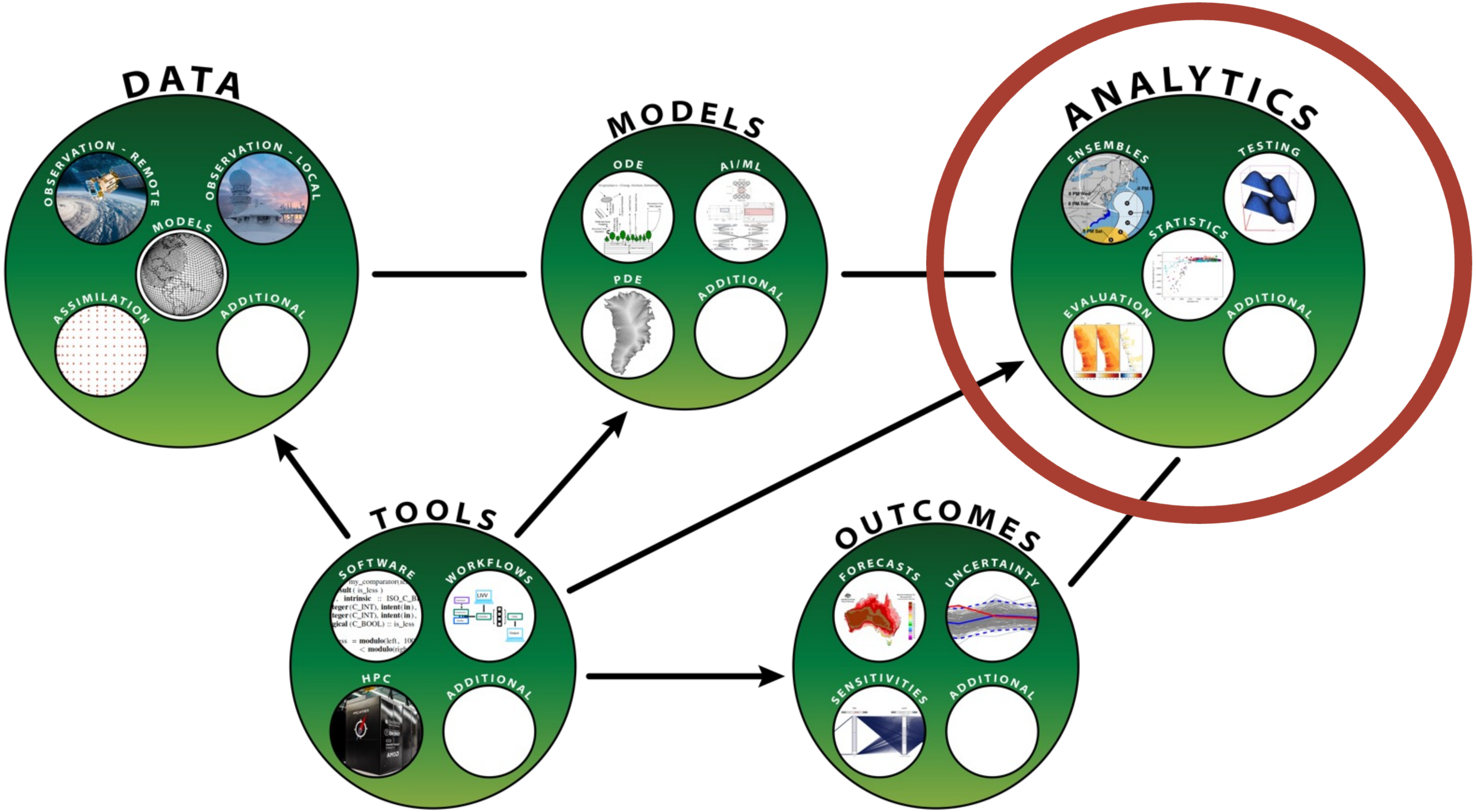
Diversity of models



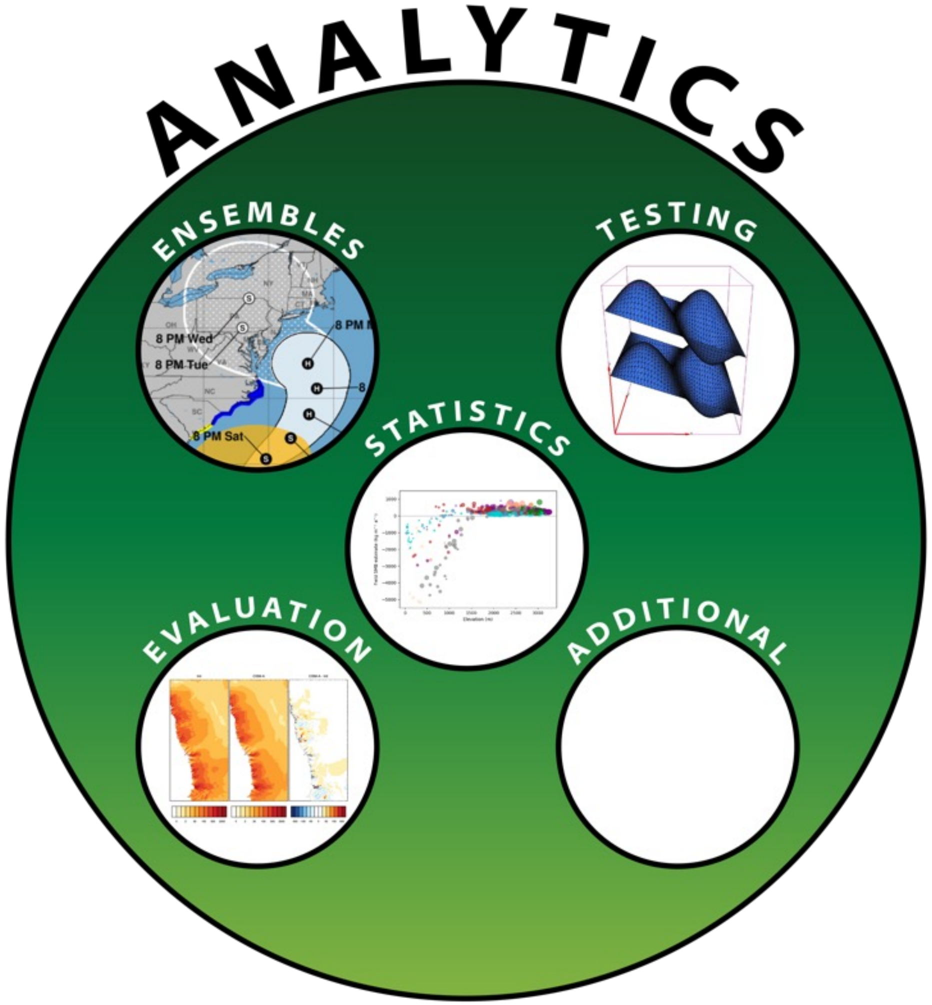
With data we can run models to understand our earth system

- PDE: Partial Differential Equations (time and space) of the atmosphere, ocean, ice sheets, subsurface flow, etc.
- ODE: Ordinary Differential Equations of the land surface, disease propagation, chemical reactions etc.
- AI/ML models cover a very wide space and are being explored for a number of applications, e.g. unsupervised learning to detect events and surrogates for subsurface flow
- Systems models, e.g. life cycle, land use land cover change (LULCC) models
- Optimization models, single or multilevel, linear or nonlinear, often used by our stakeholders

Diversity of analytics



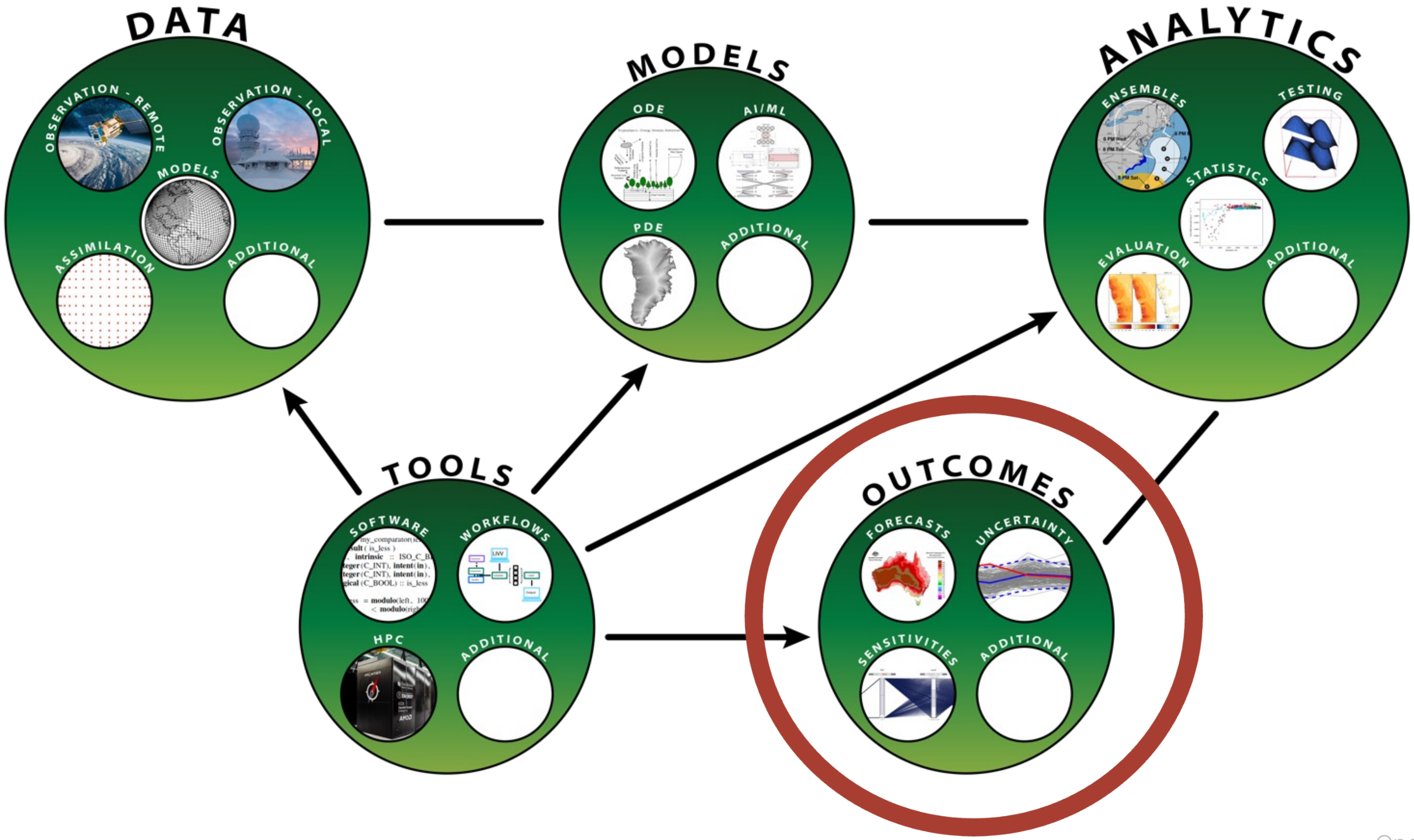
Diversity of analytics



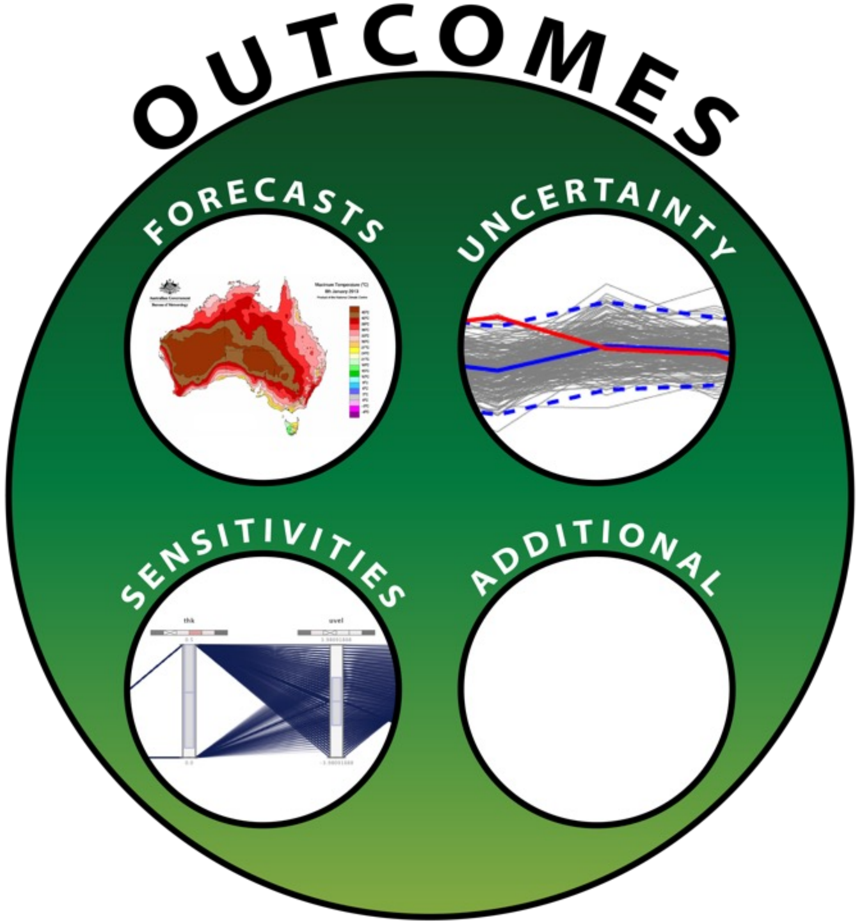
Analysis provides insight from models and data and leads to discovery

- Verification and validation of models and the computing environment to build confidence
- Community tests and evaluation of models and their couplings
- Ensembles for sensitivity
- ML/AI applied to model output (rather than serving as the model of the Earth system itself)
- Statistics: the collections of methods for all of the above

Diversity of outcomes

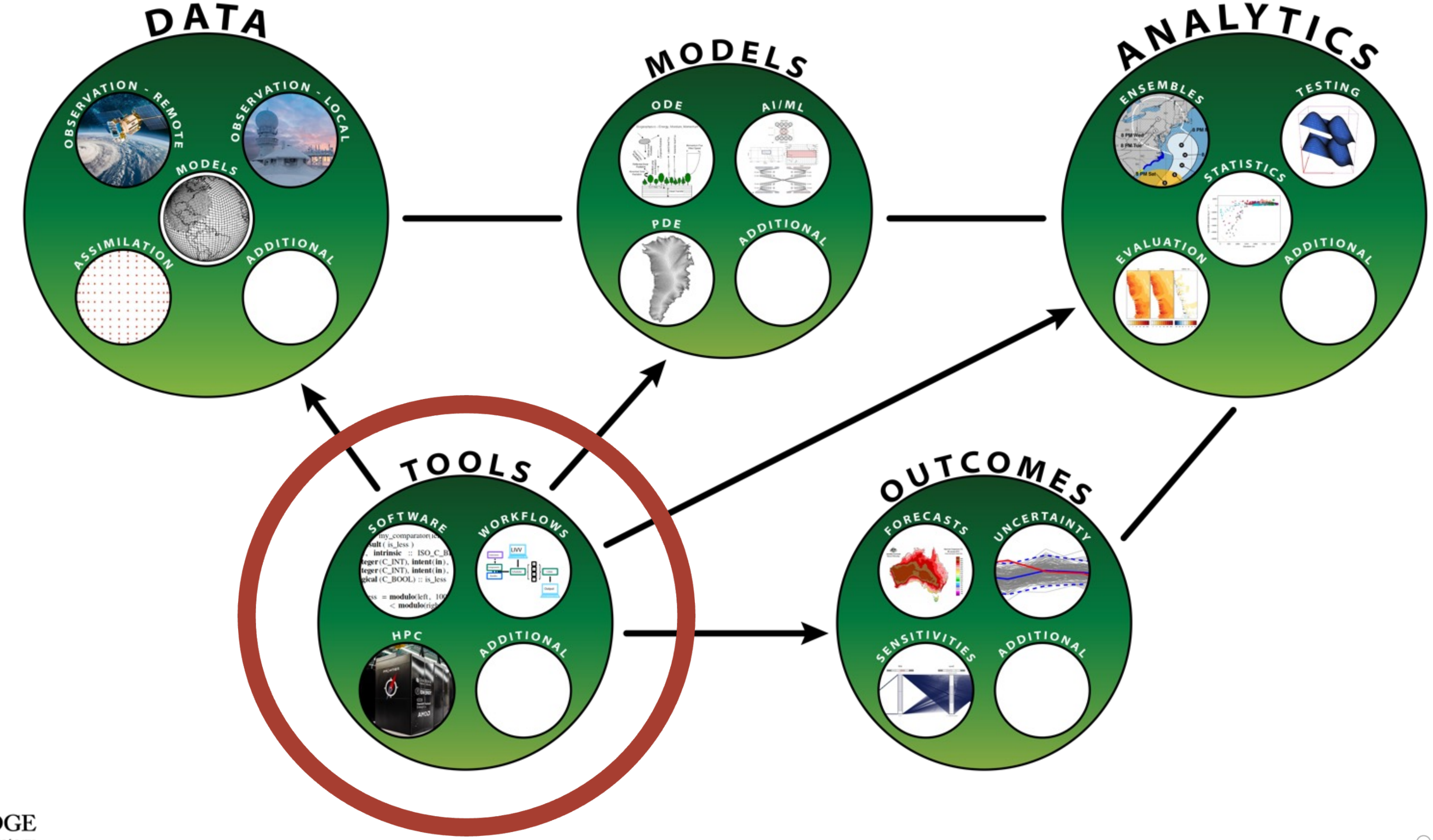


Diversity of outcomes

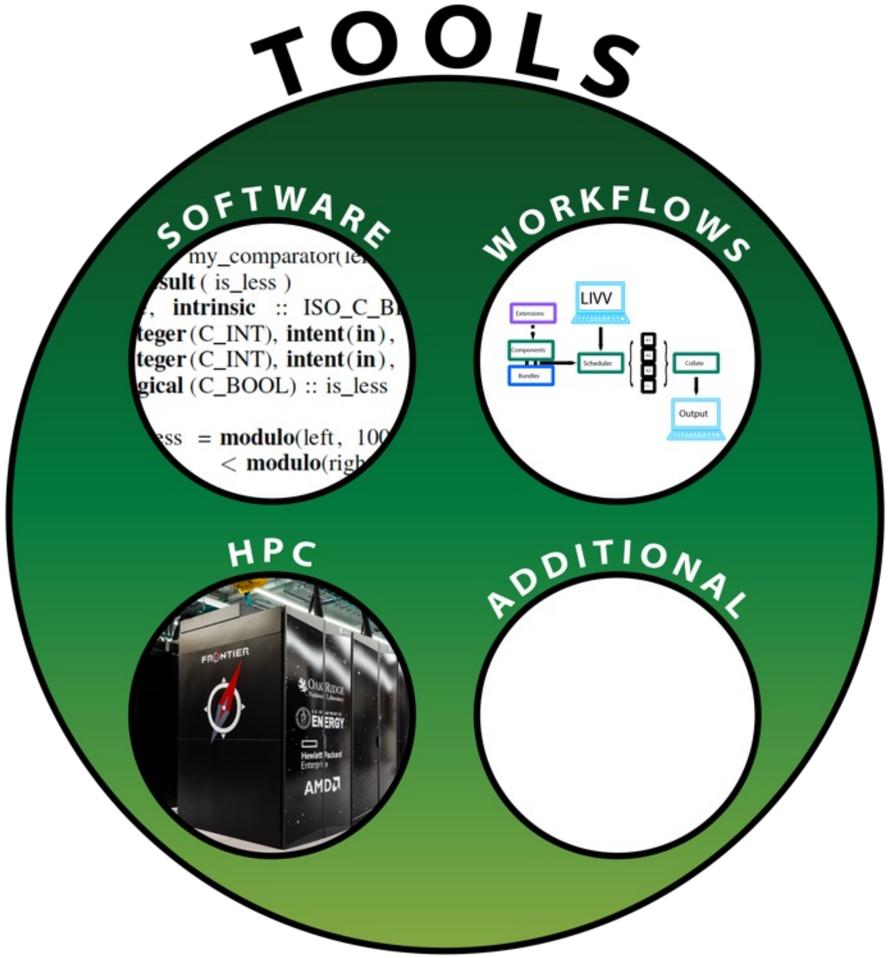


- Forecasts: from individual and combined models/analytics
- Projections: with input from our stakeholders
- Uncertainty: both within and across models, data, and analytics. Define the range of outcomes
- Sensitivities: how much does output vary? How do results change when we create a complex computing environment?
- Products: how do we package, distribute, and message the information from our computing environment?

Diversity of tools

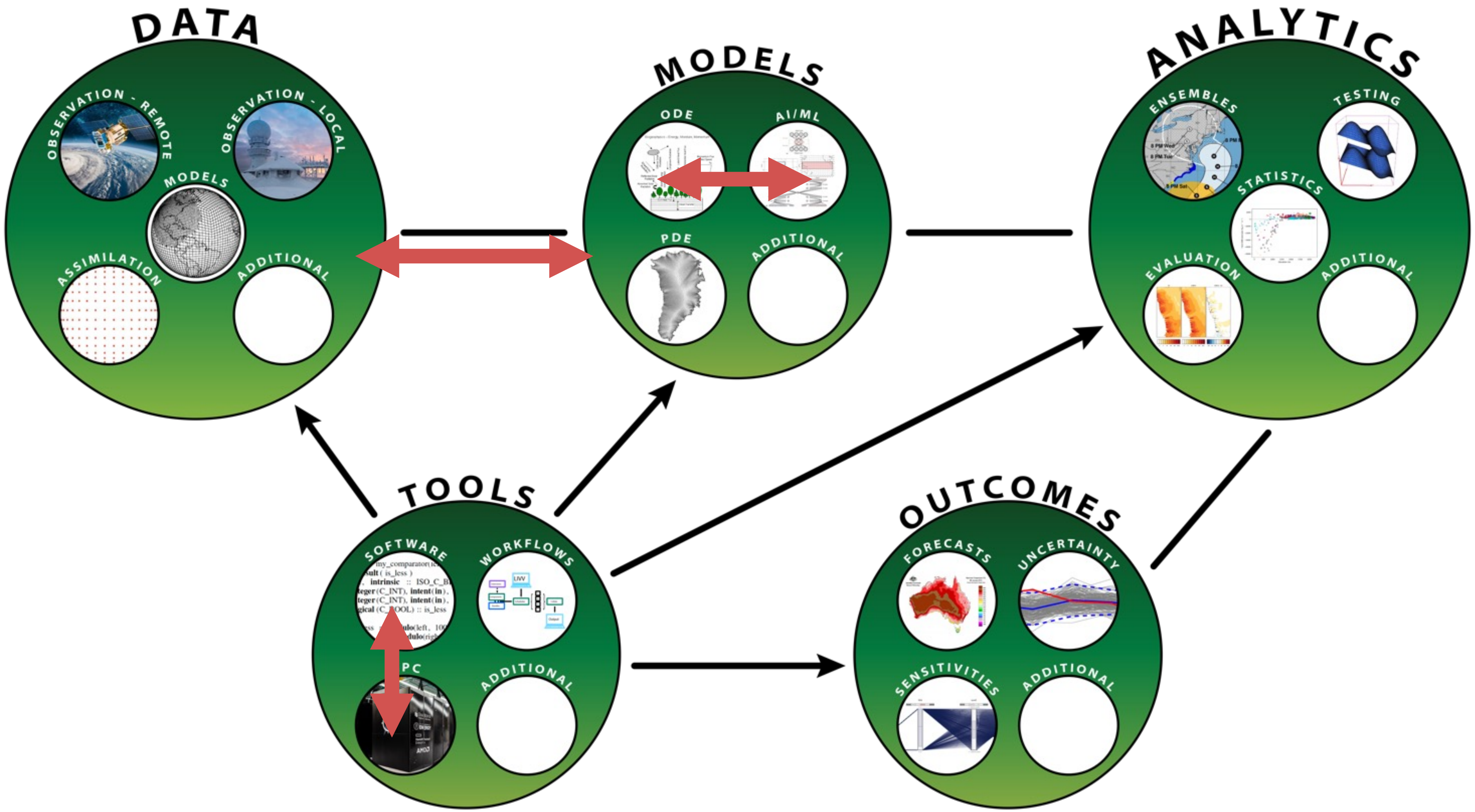


Diversity of tools



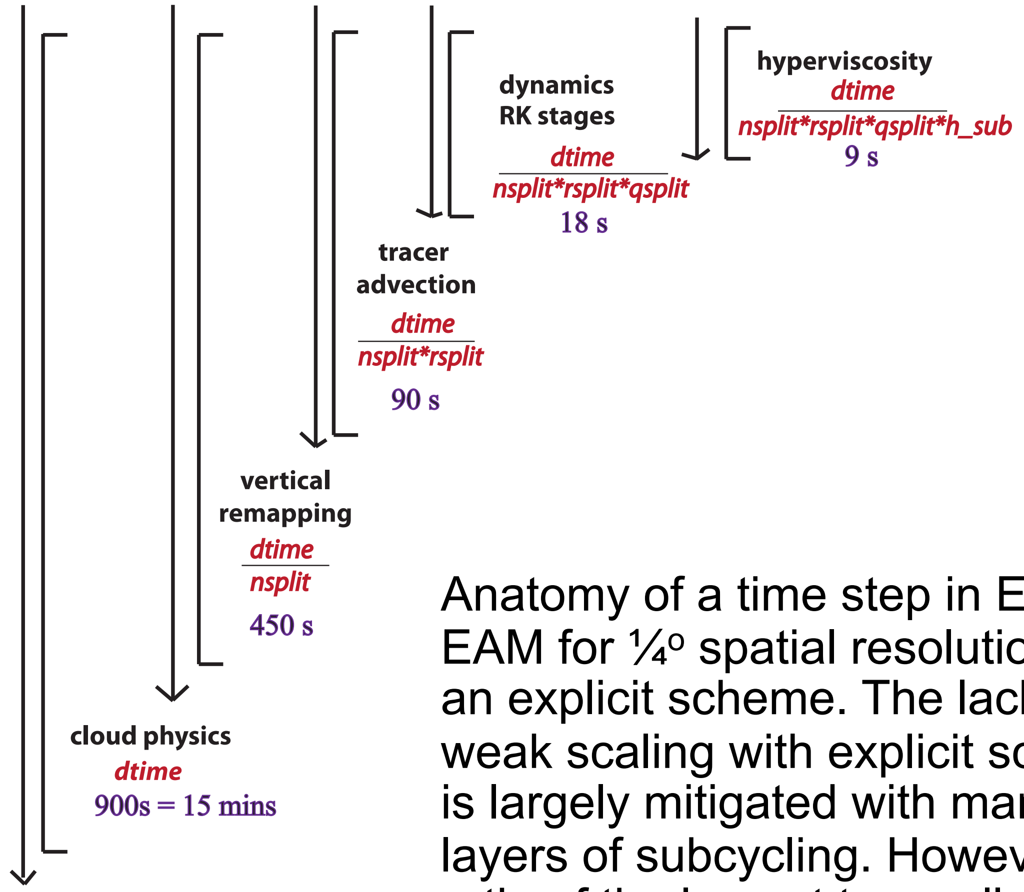
- Computing tools are needed for all of the pieces of our computing environment
- We all know we need good software. Does it have to be a certain language?
- How do we combine software from multiple communities?
- Future HPC systems are changing in a way that could benefit or stymie our efforts to build a computing environment
  - Connected/integrated computing
  - Quantum/neuromorphic
  - Cloud computing
  - Sensors to collect data
- Data transfer, access control, cyber infrastructure ….

Coupling: enabling a complex yet valuable computing environment

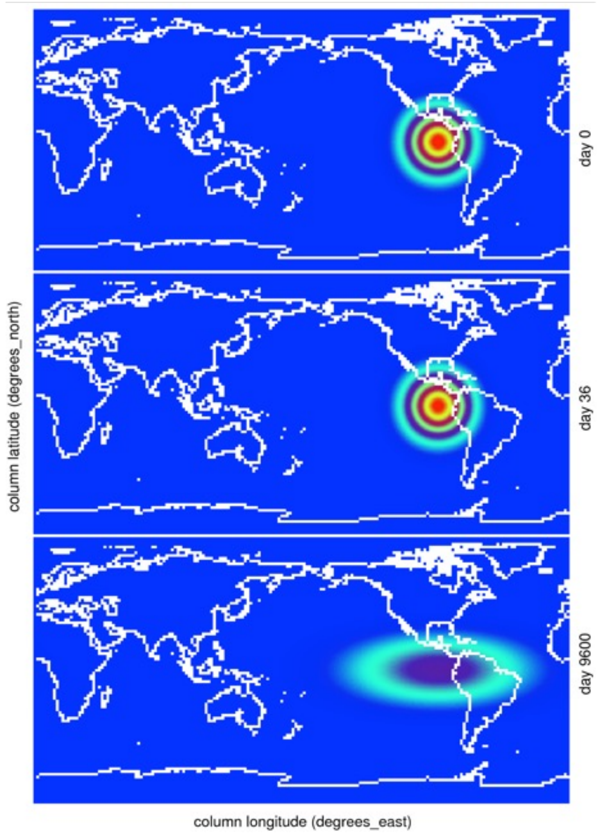


Where is our community now in realizing this vision?

Currently: The community has been working to address multiple space and time scales **within** PDE models



Anatomy of a time step in E3SM's EAM for ¼° spatial resolution using an explicit scheme. The lack of weak scaling with explicit schemes is largely mitigated with many layers of subcycling. However: the ratio of the largest to smallest time step size covers **2 orders of magnitude**



With overly diffusive time schemes, weather features dissolve over climate time scales.

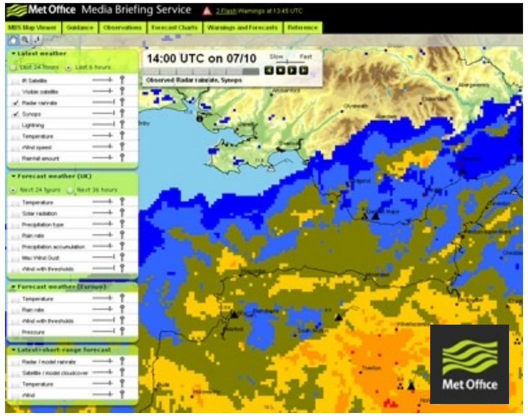
Planet: Coupling weather predictions with flood models

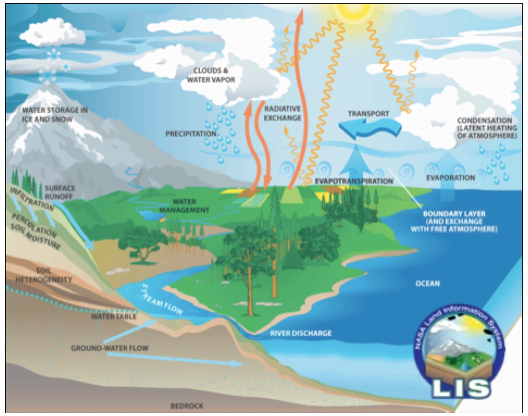
Rainfall
- **Met Office**
- Weather Forecasting Models

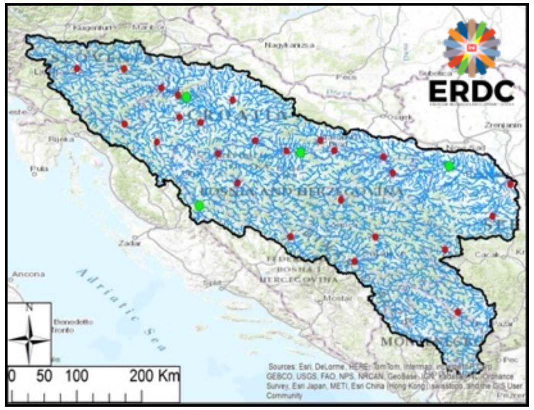
Runoff
- **NASA**
- Land Information System (LIS)

Streamflow
- **ERDC**
- Streamflow Prediction System (SPT)

Inundation
- **ORNL/TTU**
- TRITON inundation model







Opportunity: improve and automate the workflow. The model coupling crosses many institutional boundaries and is currently loose and manual

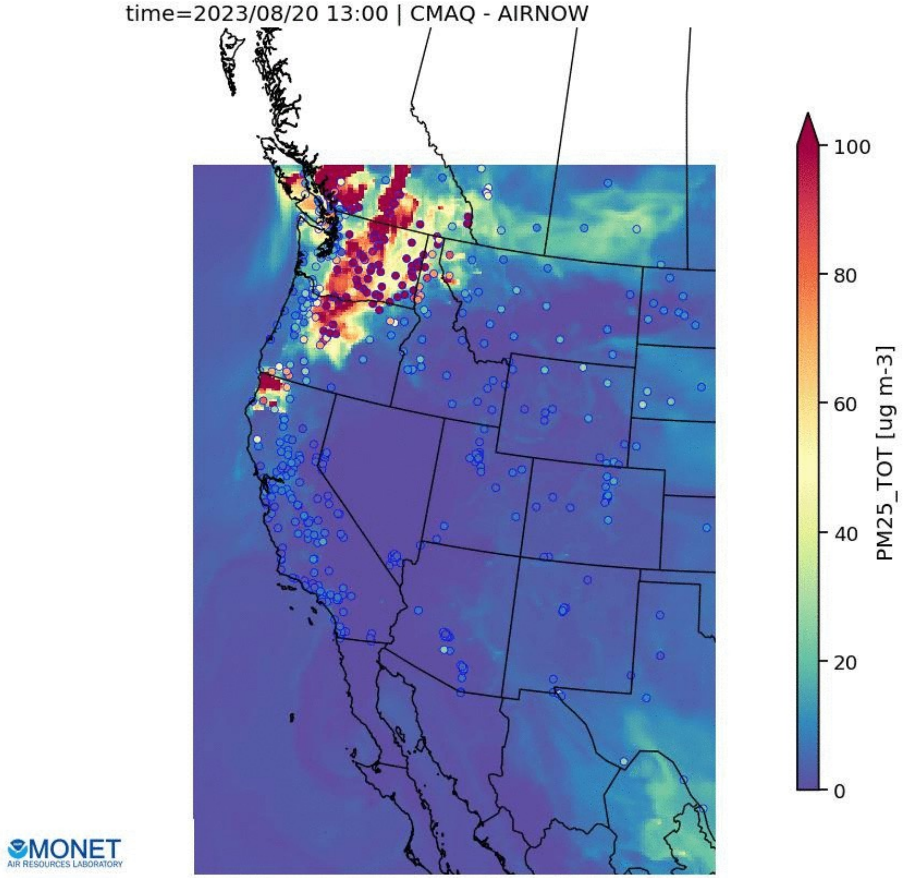
Impact: flood prediction at the neighborhood level ahead of large rain events

NGA provides the elevation data

This involves 5 independent organizations with a common goal, with U.S. Air Force as the sponsor

People: Coupling data with models: air quality

- NOAA: coupling weather models and air quality https://www.arl.noaa.gov/news-pubs/west-coast-wildfire/
- Uses Hysplit, a hybrid Lagrangian / Eulerian methodology to compute multiple interacting pollutants transported, dispersed, and deposited over local to global scales.



Opportunity: Incorporate uncertainty information and higher resolution

Impact: produce air quality information for multiple stakeholders in the presence of forest fires, dust, volcanic ash, locusts, and uses back trajectory to determine the source of release, as maps in apps, widgets, weathercasters, educators

Animation from wildfire smoke forecast for August 20-23, 2023. This forecast is also compared to the AirNow stations (shown as circles), to measure the performance of the model in real-world conditions.

AirNow is a partnership of the U.S. EPA, NOAA, NPS, NASA, CDC, tribal, state, and local air quality agencies

Infrastructure: coupling climate models to energy related applications

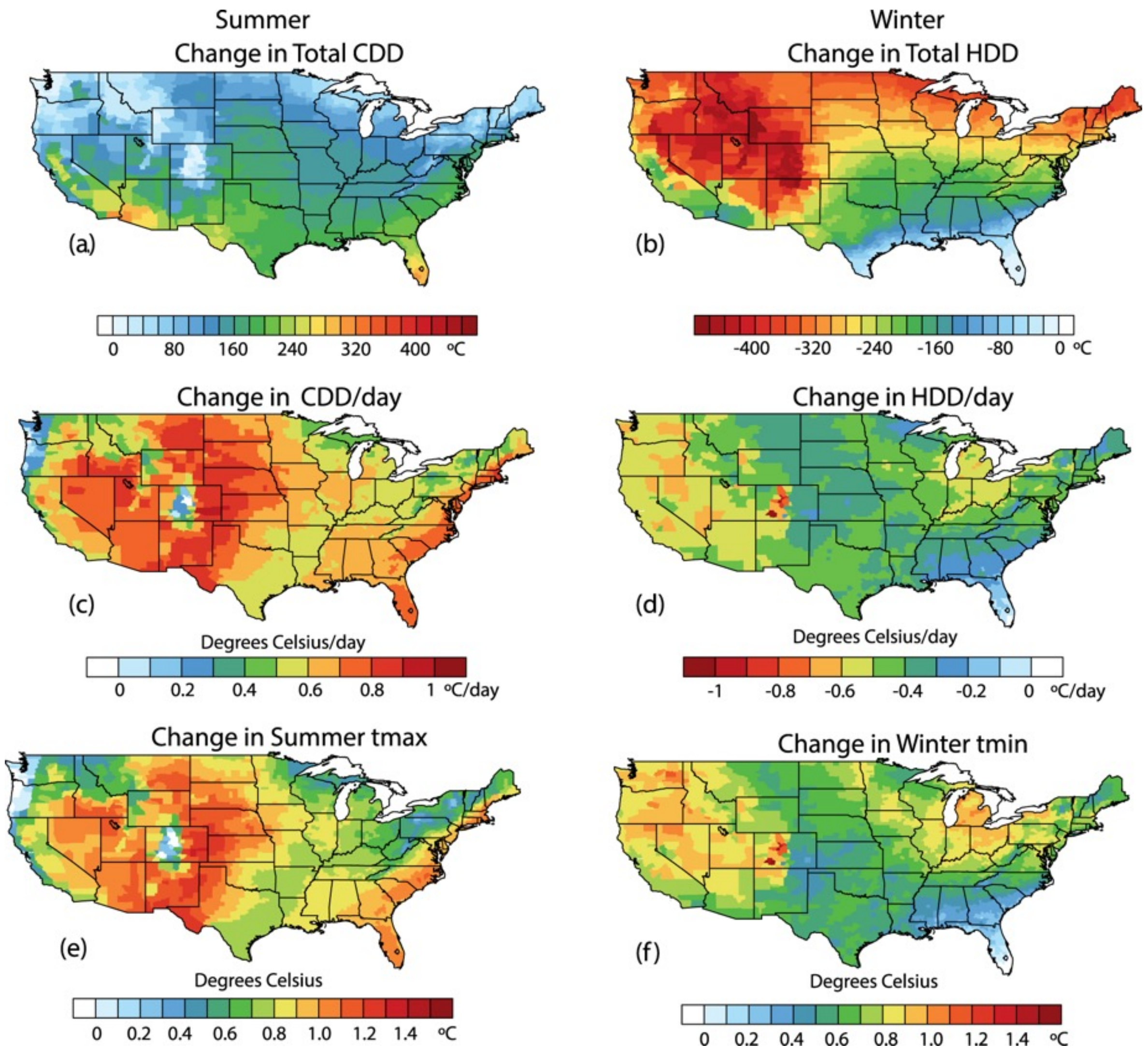
Changes in Temps by mid-century will shift US electricity and natural gas demand

Connected downscaled climate projections to models of US residential energy demand from heating degree days (HDD) and cooling degree days (CDD)

More intense and prolonged warm conditions will drive an increase in electricity demand while a shorter and milder cold season will reduce natural gas demand by the mid 21st century

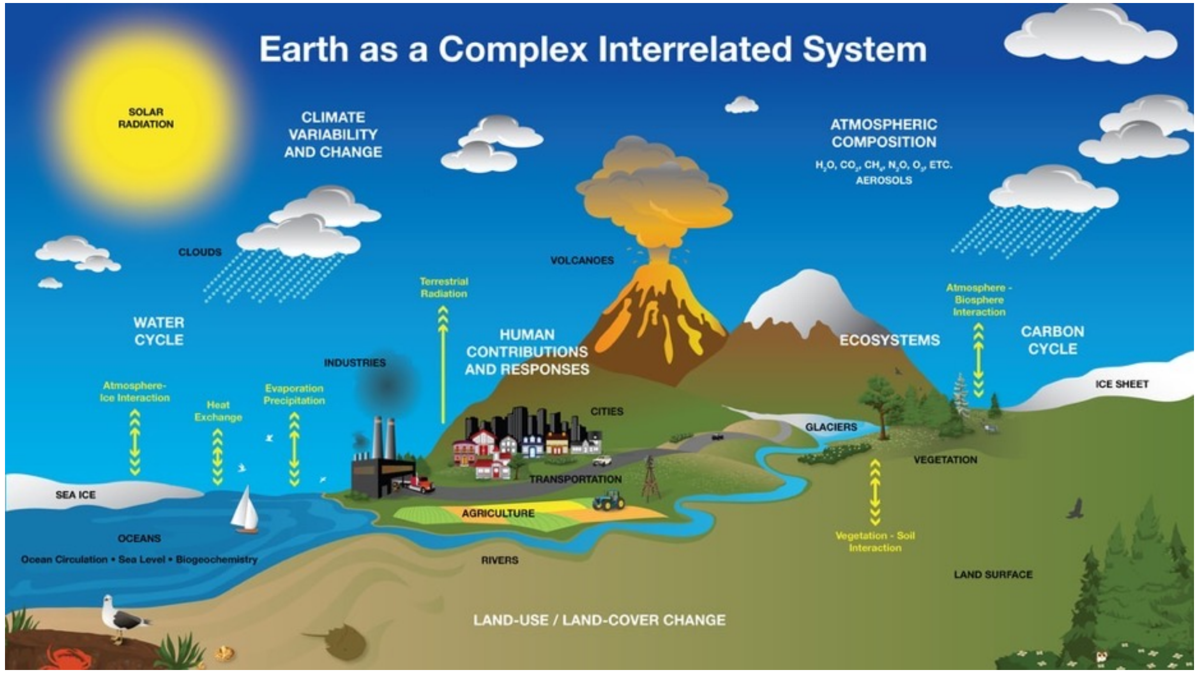
Opportunity: Incorporate alternative scenarios, rigorous coupling, access to energy use data

Impact: Comprehensive analyses of residential energy demand in response to climate change



Rastogi et al. (2019) doi.org/10.1088/1748-9326/ab22d2

Each of these are a great start toward the vision of a computing environment that allows us to answer our science questions. But there is a lot more work to do to combine disparate pieces:



- **Time and space scales** → atmosphere and oceanic flow
- **Algorithms** → atmosphere PDEs and disease model ODES
- **Data types** → Open source climate data and secure energy grid data
- **Science domains** → Chemistry and biology of agriculture

Coupling well across scales matters

- *CMIP6 models suffer from temporal inconsistencies at the ocean-atmosphere interface*
- *IRINA group tested an iterative coupling algorithm for atm and ocean based on the Schwarz iterative method*
- *Difference between the two numerical solutions is often larger than 100% of the solution, even with a small coupling period, thus suggesting that significant errors are potentially made with current coupling methods*
- *Need to balance cost and accuracy. They suggest just 2 iterations of Schwarz*

A Schwarz iterative method to evaluate OAC schemes: implementation and diagnostics in IPSL-CM6-SW-VLR Marti et. Al. GMD, 14, 2959–2975, 2021

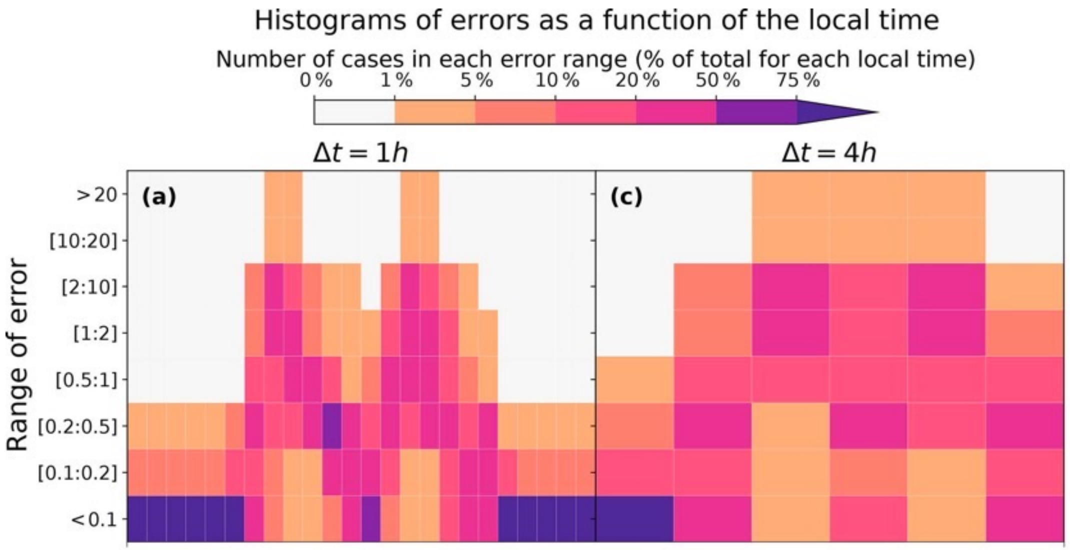


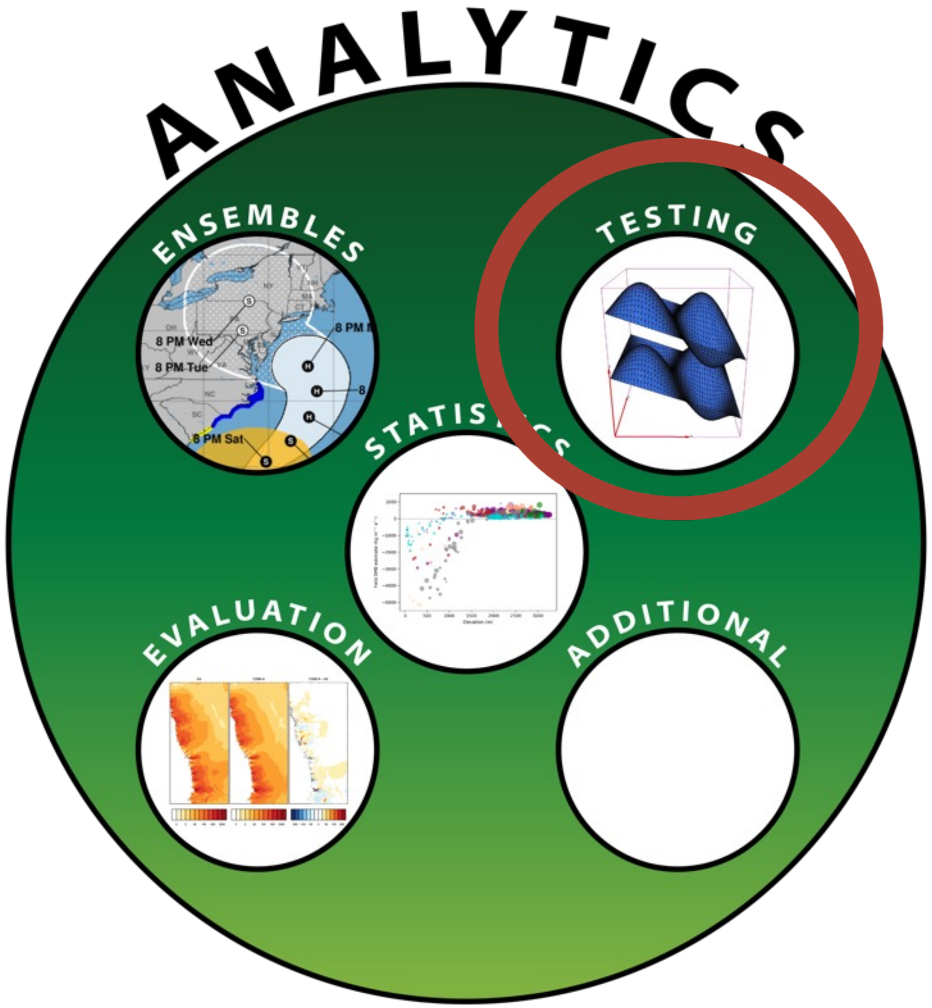
Figure 7 from Marti et al. paper: Histograms of errors as a function of the Roman local time and error classes for the parallel experiments. Top (a, c) panels show the percentage of cases in space-time in each range of error as a function of the Roman local time. The % are computed with respect to the total number of cases for each local time.

Coupling across many diverse models: What is the best way to exchange data?

- Example: Malaria and other Mosquito borne diseases.
- Review article (right shows the cover) outlines the challenges and opportunities to understand changes in disease transmissions due to climate change
- Studying requires understanding and incorporating models of socio-economics, shifting weather patterns, and background climate and coupling statistical and mechanistic models and coupling different time scales
- **Another example:** Due to increased population mobility and surface T increases, Chagas disease has been detected and is growing in USA, Canada, and many European and some African, Eastern Mediterranean and Western Pacific countries.

OAK RIDGE National Laboratory

Final thought about becoming diverse with models and data: Testing is crucial

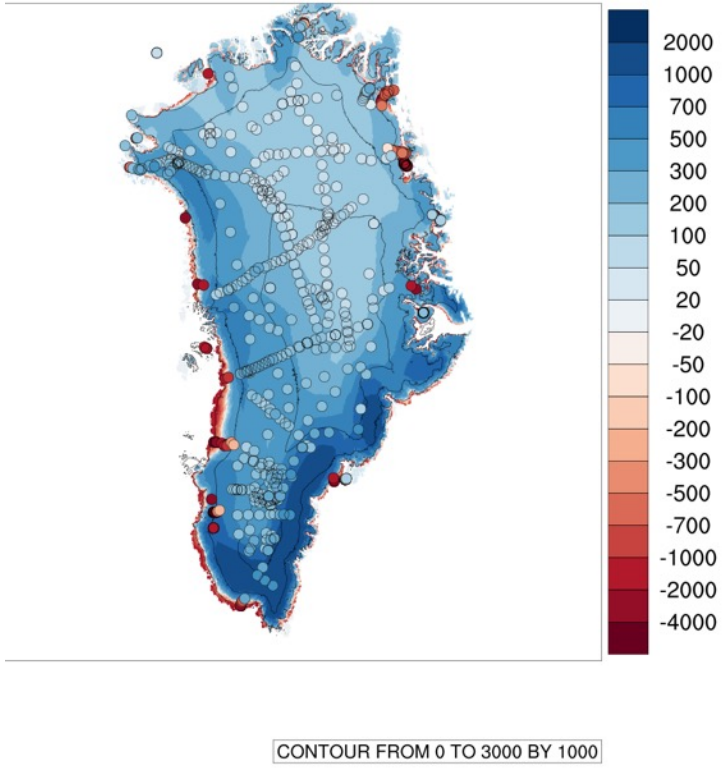


Oberkampf and collaborators provide a baseline philosophy for creating predictive models we trust in computational science:

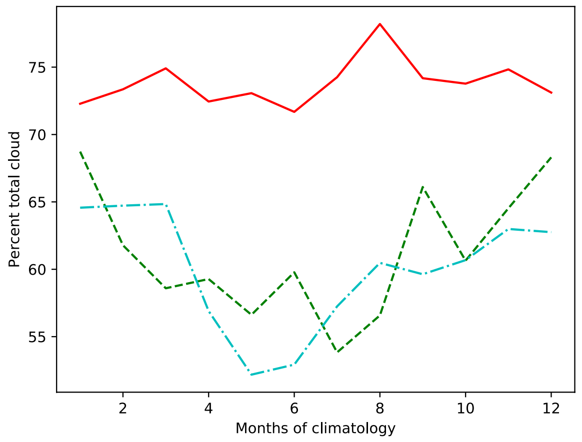
- What can go wrong?
- How likely is it to happen?
- What are the consequences?

In the context of weather and climate models, what do we have in terms of: Fidelity, Software quality, Numerical accuracy, Comparison to experimental data, Estimation of uncertainty, Understanding

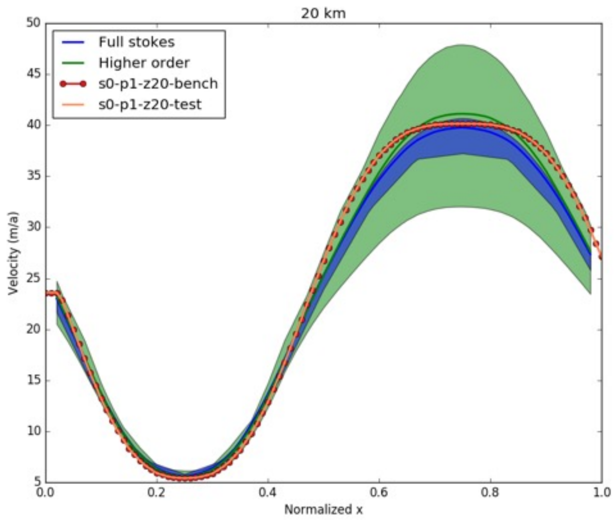
The community is making tools to help!
LIVkit: Comprehensive V&V of ice sheets and coupled models



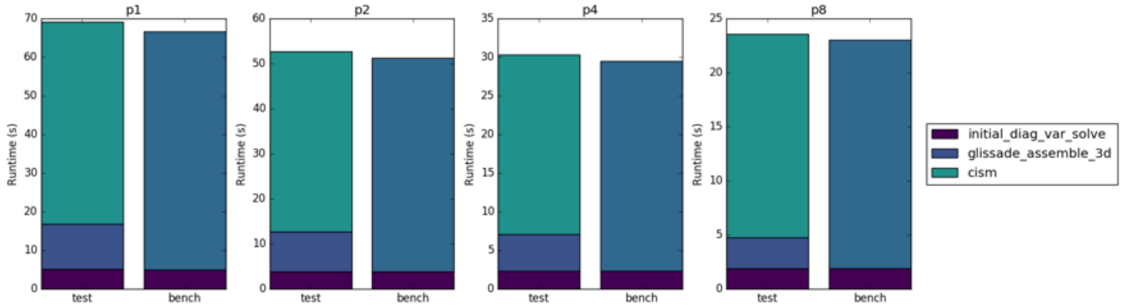
Surface mass balance simulations CISM compared to data from ice cores or snow pits.



E3SM atmosphere over CISM versus ISCCP and CLOUDSAT low cloud %



Verification of CISM with simplified analytic tests

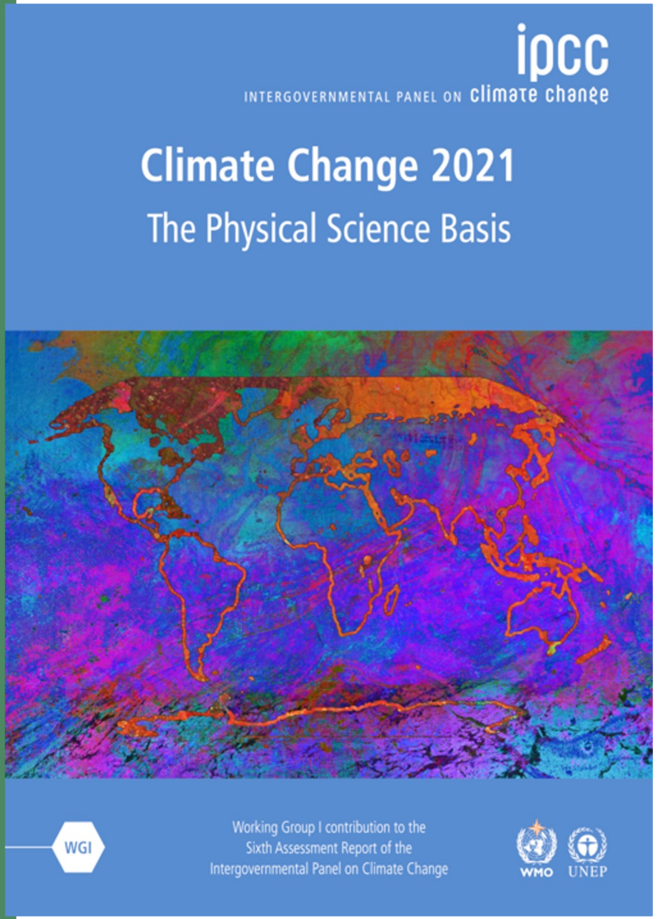


Left: Performance verification is also critical: A stacked bar chart showing the timing breakdown for CISM dome test compared to benchmark.

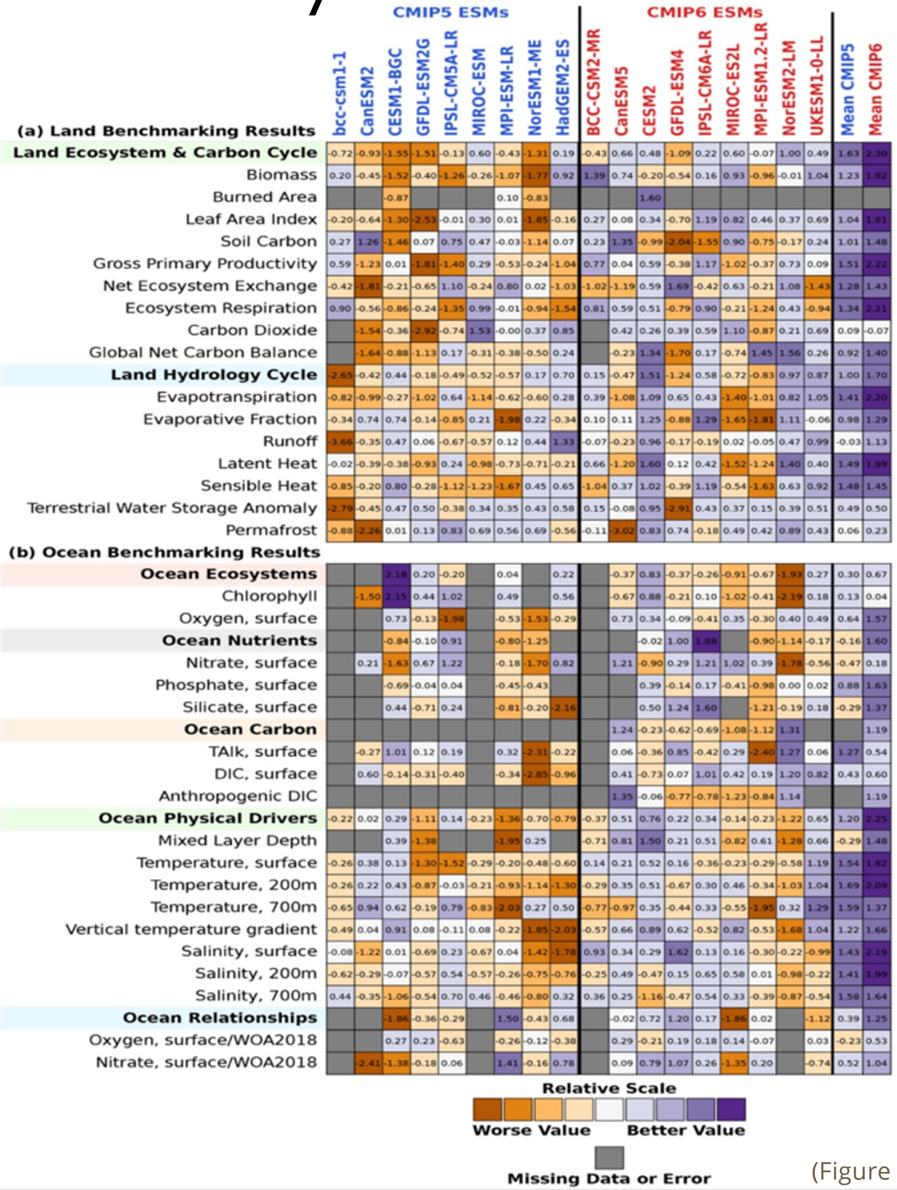
“git clone git@github.com:LIVVkit/LIVVkit.git”

ILAMB/IOMB Model Evaluation used by IPCC AR6

RUBISCO



- UN Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) from Working Group I released on August 9, 2021
- Figure 5.22 contains a detailed evaluation of land and ocean models, performed by DOE's RUBISCO project, that indicates CMIP6 models have improved over their CMIP5 progenitors, yielding more realistic land and ocean representations in climate change projections



(Figure 5.22)

Argonne NATIONAL LABORATORY
BROOKHAVEN NATIONAL LABORATORY
BERKELEY LAB
Los Alamos NATIONAL LABORATORY
NCAR NATIONAL CENTER FOR ATMOSPHERIC RESEARCH
OAK RIDGE National Laboratory
UCI
MICHIGAN

From diversity, to convergence to solutions.

Jackson Pollock
“Convergence”
1952

We need to embrace the diversity of tools, capabilities, and scientists, and immerse ourselves in this new paradigm of scientific discovery to make an impact.

PASC24 theme: Synthesizing Applications Through Learning and Computing

PASC 24

Zurich Switzerland | 3-5 June 2024

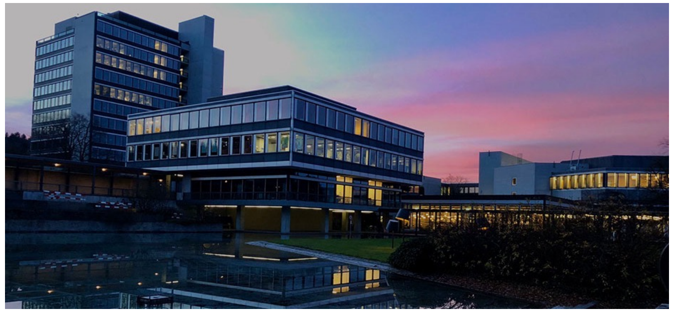
The program includes **keynote presentations**, **papers**, **minisymposia** and **poster sessions**, a **panel discussion**, and an **interdisciplinary dialogue**.

PASC24 is an international conference committed to fostering interdisciplinary exchange of knowledge in scientific computing and computational science, with a strong focus on methods, tools, algorithms, application challenges, and novel techniques in the usage of high-performance computing.

https://pasc24.pasc-conference.org/

Upcoming Submission Deadlines

- **21 October 2023**: Deadline for **Minisymposium Expressions of Interest**
- **01 December 2023**: Deadline for **full paper submissions**
- **22 December 2023**: Deadline for **poster submissions**



Association for Computing Machinery

sighpc

CSCS

OAK RIDGE National Laboratory