Exploiting Artificial Intelligence for Advanced Earth and Environmental System Science

Forrest M. Hoffman¹,², Jitendra Kumar¹,², Zachary L. Langford²,¹, V. Shashank Konduri³,¹, Nathan Collier¹, and William W. Hargrove⁴

¹Oak Ridge National Laboratory, Oak Ridge, TN, USA
²University of Tennessee, Knoxville, TN, USA
³Northeastern University, Boston, MA, USA
⁴USDA Forest Service, Asheville, NC, USA
Introduction

- Observations of the Earth system are increasing in spatial resolution and temporal frequency, and will grow exponentially over the next 5–10 years (Langford et al. 2019).
- With Exascale computing, simulation output is growing even faster, outpacing our ability to evaluate and benchmark model results.
- Explosive data growth and the promise of discovery through data-driven modeling necessitate new methods for feature extraction, change detection, data assimilation, simulation, and analysis.
50 Phenoregions for year 2012 (Random Colors)

250m MODIS NDVI
Clustered from 2000 to present

50 Phenoregion Prototypes (Random Colors)

EarthInsights

Hargrove et al., in prep.
50 Phenoregions Persistence and 50 Phenoregions Max Mode (Similarity Colors)

Principal Components Analysis

PC1 ~ Evergreen
PC2 ~ Deciduous
PC3 ~ Dry Deciduous

Hargrove et al., in prep.
GSMNP: Spatial distribution of the 30 vegetation clusters across the national park

Extracted canopy height and structure from airborne LiDAR

EarthInsights

Kumar et al., in prep.
GSMNP: 30 representative vertical structures (cluster centroids) identified

- Tall forests with low understory vegetation
- Forests with slightly lower mean mean height with dense understory vegetation
- Low height grasslands and heath balds that are small in area but distinct landscape type
Regions that exhibit similar fire seasonality globally

From MODIS “Hotspots” from 2002–2018

Norman et al., in prep.
Arctic Vegetation Mapping from Multi-Sensor Fusion

Using Hyperion Multispectral and IfSAR-derived Digital Elevation Model Trained with Alaska Existing Vegetation Ecoregions (AKEVT)

EarthInsights

Langford et al., Remote Sensing, 2019
Watershed-Scale Plant Communities Determined from DNN and AVIRIS-NG

Legend
- Alder-Willow Shrub
- Birch-Ericaceous-Lichen Shrub Tundra
- Dryas-Lichen Dwarf Shrub Tundra
- Ericaceous Dwarf Shrub Tundra
- Mesic Graminoid-Herb Meadow
- Mixed Shrub-Sedge Tussock Tundra
- Sedge-Willow-Dryas Tundra
- Tussock-Lichen Tundra
- Wet Meadow Tundra
- Wet Sedge Bog-Meadow
- Willow Shrub
- Willow-Birch Shrub

EarthInsights

Konduri et al., in prep.
A novel multiscale strategy fusing process-resolving simulations and machine learning

• Tightly integrated role for machine learning
  • Synthesize spatially distributed model inputs from diverse data streams; use inverse modeling; apply surrogate models across scales; co-design distributed sensor networks through feedback from modeling

Machine learning approaches for integrating between and across scales.

• Process-explicit integrated surface/subsurface flow and reactive transport codes
  • Represent biogeochemical processes and their hydrologic controls at their native scales; leverage exascale computing; exploit high throughput model-data integration
Earth and Environmental Sciences

Forrest M. Hoffman (ORNL), Rao Kotamarthi (ANL), Haruko Wainwright (LBNL), and the EES Writing Team
Project environmental risk and develop resiliency in a changing environment

- Increasing frequency of weather extremes and changing environment pose risks to energy infrastructure and the built environment

- Sparse observations and inadequate model fidelity limit the ability to identify vulnerability, mitigate risks, and respond to disasters
Grand Challenge #1

• New tools are needed to accelerate projection of weather extremes and impacts on energy infrastructure

• Building resiliency to address evolving risks will benefit from integration of smart sensing systems, built-for-purpose models, ensemble forecasts to quantify uncertainty, and dynamic decision support systems for critical infrastructure
Grand Challenge #2

Characterize and modify subsurface conditions for responsible energy production, CO$_2$ storage, and contaminant remediation

- National energy security and transition to renewable energy resources relies on utilization of subsurface reservoirs for energy production, carbon storage, and spent nuclear fuel storage
- Subsurface data are uncertain, disparate, diverse, sparse, and affected by scaling issues
- Subsurface process models are incomplete, uncertain, and frequently unreliable for prediction
Grand Challenge #2

• We need to substantially increase hydrocarbon extraction efficiency, discover and exploit hidden geothermal resources, reduce induced seismicity and other impacts, improve geologic CO₂ storage, and predict long-term fate and transport of contaminants.

• Mitigating risks requires improved subsurface characterization and assimilation of real-time data streams into predictive models of geological and ecological processes.
Grand Challenge #3

Develop a predictive understanding of the Earth system under a changing environment

- To advance the nation’s energy and infrastructure security, a foundational scientific understanding of complex and dynamic hydrological, biological, and geochemical processes and their interactions is required (across atmosphere, ocean, land, ice)
- Knowledge must be incorporated into Earth system models to project future climate conditions for various scenarios of population, socioeconomics, and energy production and use
Grand Challenge #3

• Accurate predictions are needed to quantify changes in atmospheric and ocean circulation and weather extremes, to close the carbon cycle, and to understand responses and feedbacks of human, terrestrial, and marine ecosystems to environmental change.

• Advances in genomics and bioscience data need to be leveraged to provide detailed understanding of plant–microbial interactions and their adaptations and feedbacks to the changing environment.
Grand Challenge #4

Ensure global water security under a changing environment

• Water resources are critical for energy production, human health, food security, and economic prosperity

• Water availability and water quality are impacted by environmental change, weather extremes, and disturbances such as wildfire and land use change
Grand Challenge #4

• Methods are needed to integrate disparate and diverse multi-scale data with models of watersheds, rivers, and water utility infrastructure

• Predictions of water quality and quantity require data-driven models and smart sensing systems

• Water resource management must account for changes in weather extremes, population, and economic growth
Accelerating Development

The near-term (5–10 years) priorities are to:

• Develop hybrid process-based/AI modeling frameworks for Exascale systems
• Develop strategies for mapping hybrid components on GPU/CPU based on computational density and communications patterns
• Develop physics / chemistry / biology-constrained ML
• Develop explainable AI and ML methods for hypothesis generation and testing
Expected Outcomes

• Model testbeds and surrogate models are expected to yield insights into process understanding across all Grand Challenges
• Data-driven and physics-constrained hybrid models are expected to stimulate new discovery and bridge space and time scales
• Integrated models of Earth system processes and energy/built infrastructure will enhance national energy and water security through simulation
• AI methods will enable effective use of large data streams for energy production, predictive process understanding, and environmental resiliency