Building AI-based Earth Systems Models and the TERRAHydro Terrestrial Digital Twin

*Craig Pelissier (PI)
Grey Nearing (Co-PI)
Carlos Cruz
Kia Saeedi
Brandon Smith
Vanessa Valenti

*presenter

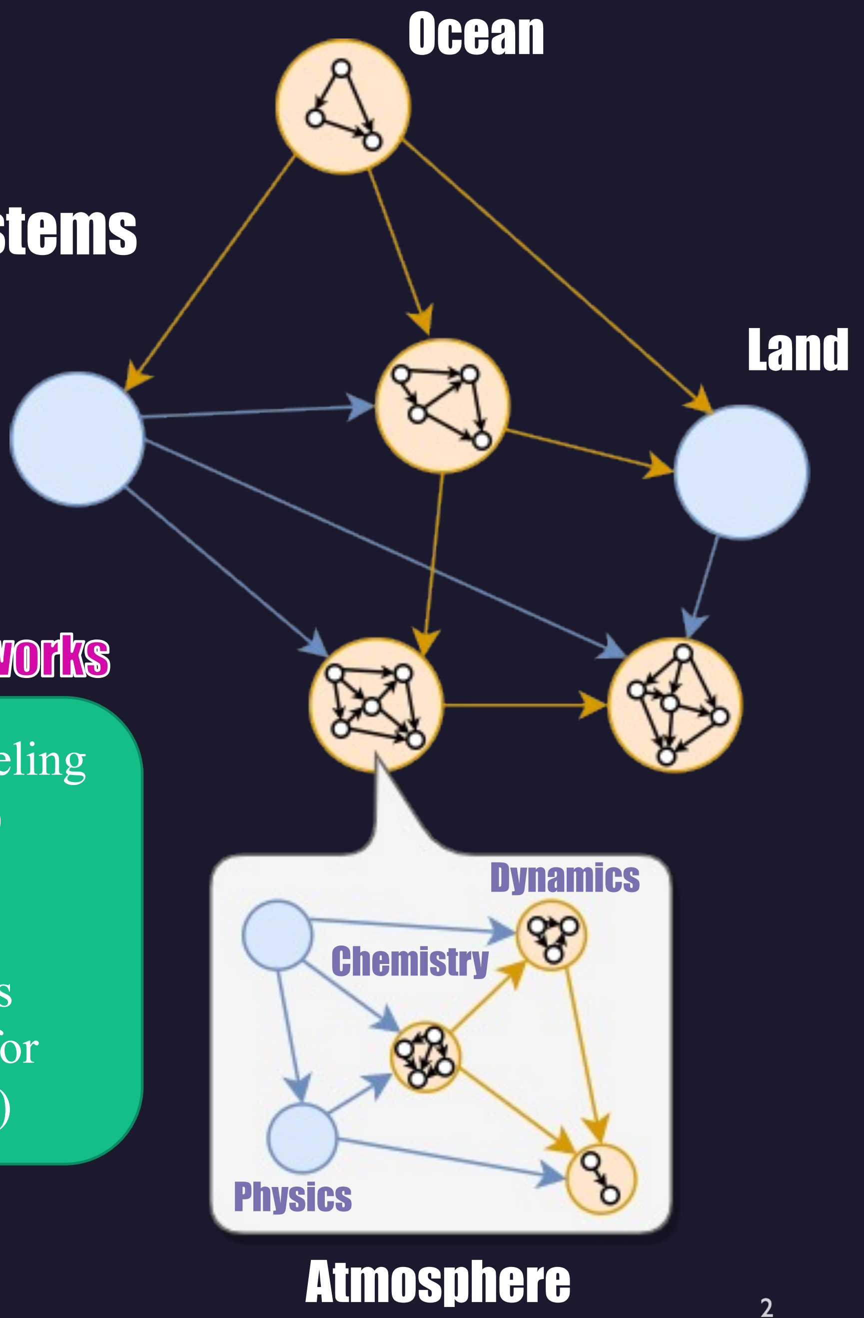
The Structure of Earth Systems Models

- Composed of a coupled set of sub-systems or models of physical processes.
- Organizable into a hierarchal graph with components developed and maintained by domain experts across multiple organizations (e.g., government, academia, industry).
- Frameworks for coupling, organization, and operation of large ESMs/climate models are essential to developing and maintaining large open-source open-science systems.
- Existing frameworks mostly written in/for Fortran and parallelized with MPI for scalability on large CPU-based clusters and not designed for integrating AI technogies.

Some existing frameworks

- Earth Systems Modeling Framework (GEOS)
- MAPL (GEOS)
- NASA's Land Information Systems (operational driver for land surface models)

A Diagram of a Hierarchal Systems Model



Adoption of AI-based Technology

- Data-Driven models have emerged across many Earth Science domains and have been shown to be a powerful alternative to traditional modeling.
- Very few operational AI-based Earth Systems models exists today (one counter example is Google's flood forecasting system).
- Integration into existing modeling frameworks is a significant challenge to the adoption AI-based technologies.

This is relevant today. AI technologies already exist that offer improved capabilities.

Some Barriers to Adoption

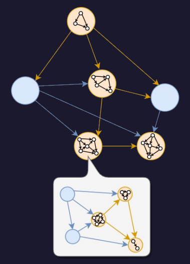
- Existing frameworks do not provide gradients to train fully integrated AI technologies.
- The AI community leverages powerful (Python) tensor-based software which is not easily integrated within existing Fortran frameworks.
- AI-models use differently structured data (sequence-to-one sequence-to-sequence, image stacks/series)
- Assimilation techniques in AI can use tensor gradients techniques.

A Coupled Reusable Earth System Tensor (CREST) Framework

An AI-First Python framework for building coupled open-science Earth Systems Models

Combines the idea of ESM hierarchal graphs with Tensor Network graphs.

Tensor Network Backend



Earth Systems Model Hierarchal Graph

CREST Hierarchal Tensor Graph/Model

- Provides a specification/map between a Tensor Network graph and ESM graph.
- Standardizes how data is passed between sub-systems.
- Internally handles different types of AI-models: Keras, Pytorch, Jax, Black-box models.
- Creates a layer between the tensor backend to enable flexible, extensible, agnostic infrastructure to be built.

CREST Operational Infrastructure

Infrastructure to build, maintain, and operate ESMs.

- **Data Management** for loading large geo-spatial-temporal data.
- **Data Asssimilation** for ingestion of near real-time data and adaptability.
- **Easy-to-use GUI-based driver** for operation and analysis.
- **Training, inference, and scenario analysis workflows** to perform operational simulations.

Traditional Modeling In CREST

- It is easier to express physical processes in Tensor Networks than integrate tensor networks in physical-based model.
- Any physical-based or process-based parameterized model can be written using tensor-based software.
- Some work on incorporating physical constraints within AI models.
- Some approaches to estimate gradients of black-box models which would allow wrapping of Fortran-based models into python for inclusion in CREST.
- Some efforts towards Fortran auto-diff.

Integrating traditional models in CREST is targeted for future development.

NEED GRADIENTS

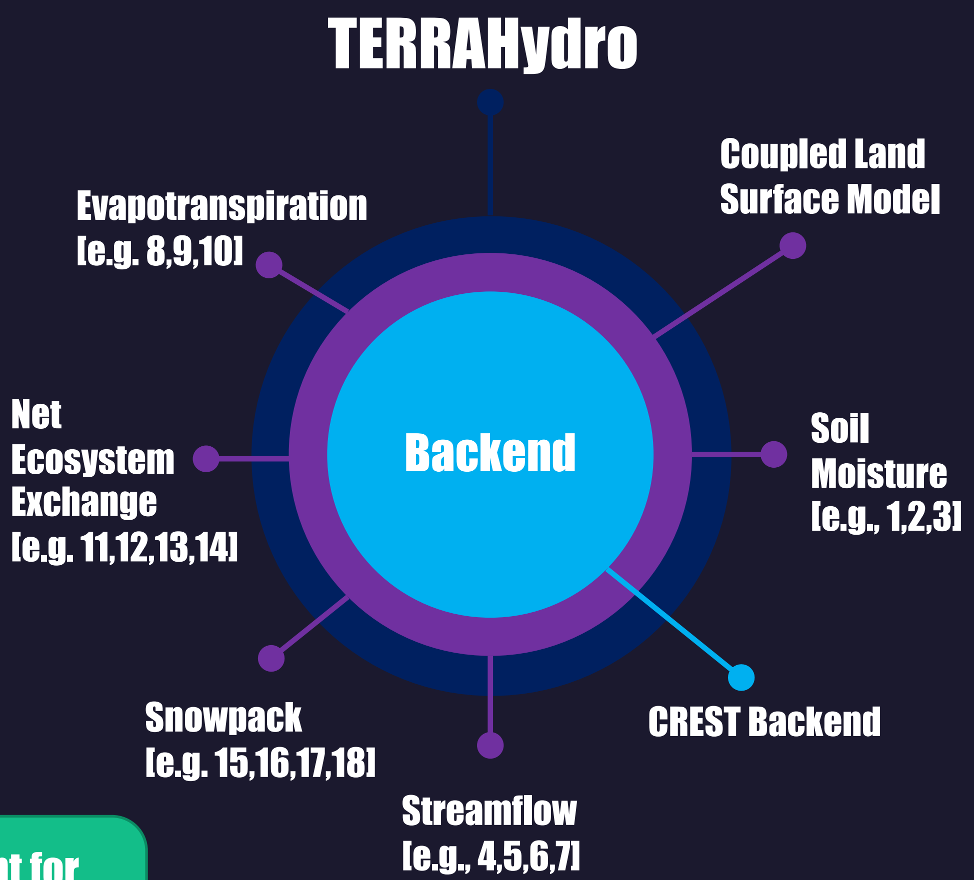
Potential Approaches

- **Rewrite (Fortran) models in a tensor language.**
- **Leverage black-box gradient estimates.**
- **Development of a reliable auto-differentiator for Fortran.**
- **Apply physical constraints on AI models.**

Emerging packages like JAX offer the use of numpy-like numerical packages that make this easier.

TERRAHydro

- Takes the next step in data-driven hydrology by coupling together 5 existing AI-based land surface components to assemble a land surface digital twin.
- Demonstrates an application of the CREST framework for assembling coupled AI-based Earth Systems Models --- guides the development.
- Enables new tensor-gradient-based data assimilation techniques in addition to traditional Bayesian approaches for near-real time ingestion of data.
- Computational efficiency and use of hardware accelerators enables rapid adaptability and scenario analysis beyond current capabilities.
- Potential for improved accuracy. References can be found on the last slide.



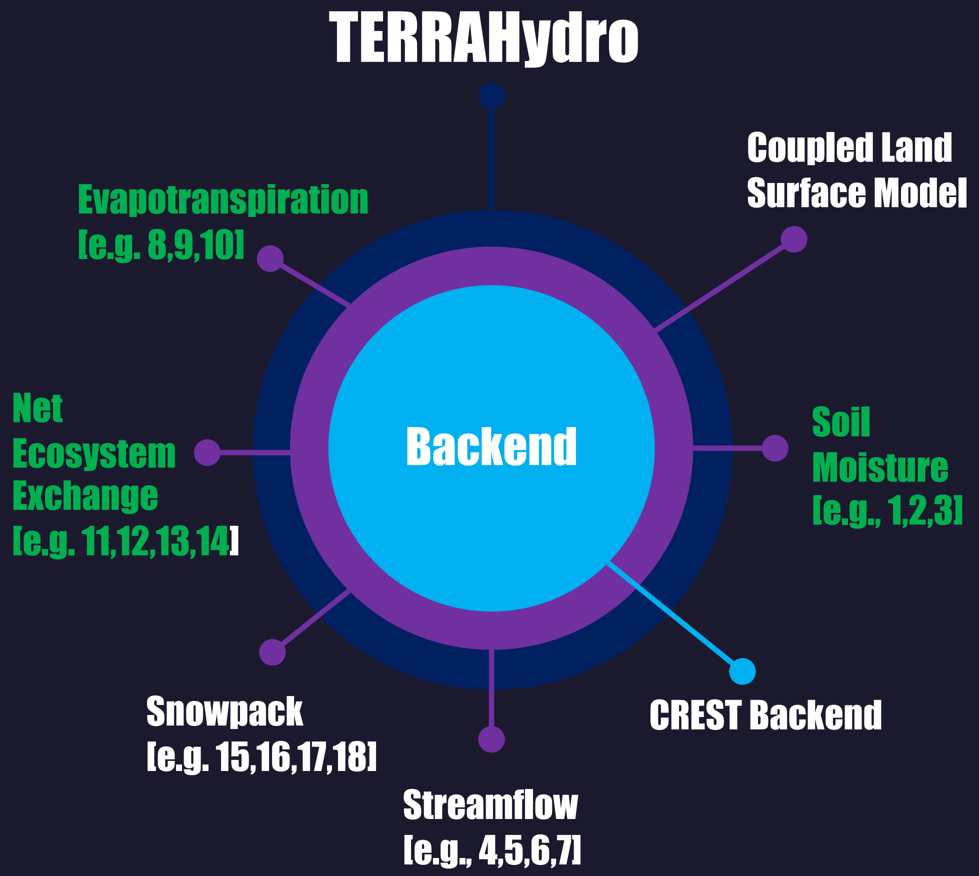
Important for Digital Twin capabilities

CREST Progress

Data Loading	Model Specification	Training and Validation	Inference
crest.Dataset crest.StructuredDataset	crest.HierarchalTensorGraph	crest.Batcher crest.Model	TBD
Efficiently loads large heterogeneous gridded geospatial-temporal data	Currently restricted to TensorFlow (Keras) AI-based models.	Builds Keras models and efficiently batches data contained in Dataset.	Inference can be performed but software design still being conceptualized

TERRAHydro Progress

- Soil Moisture, Evapotranspiration, Net Ecosystem Exchange (3 of 5) land surface components have been implemented and moderately validated.
- ESMWF (ERA5), Soil Moisture Active Passive (SMAP), and FluxNet data has been ingested for training.
- Large scale validation to reproduce paper results underway.
- A coupled Soil Moisture + Evapotranspiration model is underway with a target demonstration of coupled Soil Moisture + Evapotranspiration + Net Ecosystem Exchange model for an 18-month demo.



Timeline

September, 2022	1st year milestone	18 month demo	2nd year milestone	3rd year
Project kick-off.	End-to-end capabilities in CREST and a simplified coupled model with SM+ET.	TERRAHydro retrospective demonstration of a coupled land surface (SM+ET+NEE) model for a hydrological event.	TERRAHydro coupled land surface model and assimilation completed along with all relevant CREST features for the third-year demo.	Full scale retrospective demonstration including data assimilation and scenario analysis on relevant hydrological events.

Thank You

Craig Pelissier, craig.s.pelissier@nasa.gov

Lead, Advanced Software Technology Group (ASTG)

NASA Goddard/SSAI

ASTG website

References

[1] Kuai Fang, Chaopeng Shen, and Daniel Kifer. Evaluating aleatoric and epistemic uncer- tainties of time series deep learning models for soil moisture predictions. arXiv preprint arXiv:1906.04595, 2019.
[2] Kuai Fang, Chaopeng Shen, Daniel Kifer, and Xiao Yang. Prolongation of smap to spa- tiotemporally seamless coverage of continental us using a deep learning neural network. Geophysical Research Letters, 44(21):11–030, 2017.
[3] O Sungmin and Rene Orth. Global soil moisture data derived through machine learning trained with in-situ measurements. Scientific Data, 8(1):1–14, 2021.
[4] Frederik Kratzert, Daniel Klotz, Mathew Herrnegger, Alden K Sampson, Sepp Hochre- iter, and Grey S Nearing. Toward improved predictions in ungauged basins: Exploiting the power of machine learning. Water Resources Research, 55(12):11344–11354, 2019.
[5] Frederik Kratzert, Daniel Klotz, Claire Brenner, Karsten Schulz, and Mathew Herrneg- ger. Rainfall–runo↵ modelling using long short-term memory (lstm) networks. Hydrology and Earth System Sciences, 22(11):6005–6022, 2018.
[6] Frederik Kratzert, Daniel Klotz, Guy Shalev, Gu ̈nter Klambauer, Sepp Hochreiter, and Grey Nearing. Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. Hydrology and Earth System Sciences, 23(12):5089–5110, 2019.
[7] Dapeng Feng, Kuai Fang, and Chaopeng Shen. Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. Water Resources Research, 56(9):e2019WR026793, 2020.
[8] Grey S Nearing, Benjamin L Ruddell, Martyn P Clark, Bart Nijssen, and Christa Peters- Lidard. Benchmarking and process diagnostics of land models. Journal of Hydromete- orology, 19(11):1835–1852, 2018.
[9] Wen Li Zhao, Pierre Gentine, Markus Reichstein, Yao Zhang, Sha Zhou, Yeqiang Wen, Changjie Lin, Xi Li, and Guo Yu Qiu. Physics-constrained machine learning of evapo- transpiration. Geophysical Research Letters, 46(24):14496–14507, 2019.
[10] Andrew Bennett and Bart Nijssen. Deep learned process parameterizations provide better representations of turbulent heat fluxes in hydrologic models. Water Resources Research, 57(5):e2020WR029328, 2021.
[11] Phuong Nguyen and Milton Halem. Machine learning for inferring co2 fluxes: The new metaphysics of neural nets. UMBC Faculty Collection, 2019.
[12] O Reitz, A Graf, M Schmidt, G Ketzler, and M Leuchner. Upscaling net ecosystem exchange over heterogeneous landscapes with machine learning. Journal of Geophysical Research: Biogeosciences, 126(2):e2020JG005814, 2021.
[13] Jianzhao Liu, Yunjiang Zuo, Nannan Wang, Fenghui Yuan, Xinhao Zhu, Lihua Zhang, Jingwei Zhang, Ying Sun, Ziyu Guo, Yuedong Guo, et al. Comparative analysis of two machine learning algorithms in predicting site-level net ecosystem exchange in major biomes. Remote Sensing, 13(12):2242, 2021.
[14] Ni Huang, Li Wang, Yuelin Zhang, Shuai Gao, and Zheng Niu. Estimating the net ecosystem exchange at global fluxnet sites using a random forest model. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14:9826–9836, 2021.
[15] Ludovica De Gregorio, Mattia Callegari, Carlo Marin, Marc Zebisch, Lorenzo Bruzzone, Begu ̈m Demir, Ulrich Strasser, Thomas Marke, Daniel Gu ̈nther, Rudi Nadalet, et al. A novel data fusion technique for snow cover retrieval. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12(8):2862–2877, 2019.
[16] Changyu Liu, Xiaodong Huang, Xubing Li, and Tiangang Liang. Modis fractional snow cover mapping using machine learning technology in a mountainous area. Remote Sensing, 12(6):962, 2020.
[17] Jinliang Hou, Chunlin Huang, Ying Zhang, Jifu Guo, and Juan Gu. Gap-filling of modis fractional snow cover products via non-local spatio-temporal filtering based on machine learning techniques. Remote Sensing, 11(1):90, 2019.
[18] Soni Yatheendradas and Sujay Kumar. A novel machine learning-based gap-filling of fine-resolution remotely sensed snow cover fraction data by combining downscaling and regression. Journal of Hydrometeorology, 2021.