DUKE

Deep-learning-based, hybrid, uncertainty-aware modeling of the coupled water and carbon cycle with Earth observation data

Basil Kraft

June 2023

MPI for Biogeochemistry in Jena
Department Biogeochemical Integration

Land surface processes: uncertainties and open questions

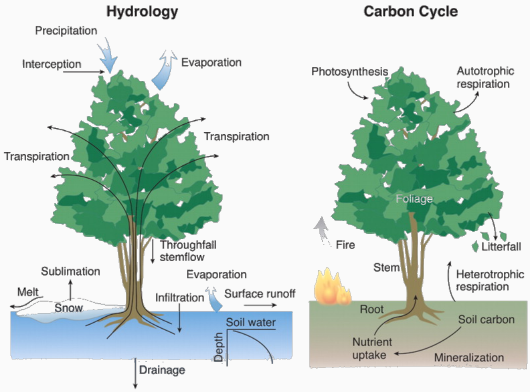


Figure 1: Some processes represented by modern land surface models, *Bonan (2008), Science.*

Objectives of land surface modeling

- Process understanding
- Running scenarios
- Simulate unobserved variables

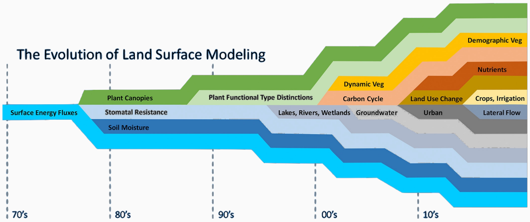
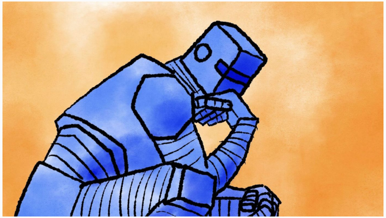


Figure 2: Land surface models grew in complexity, *Fisher & Koven (2020), JAMES.*

AI-Methods for EO: learning from data

Common AI applications in EO

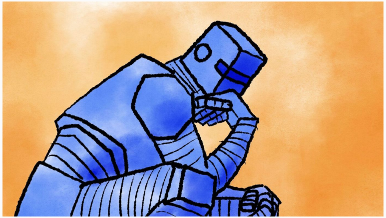
- Classification
- Object detection
- Time series prediction
- Outlier detection
- Gapfilling
- Downscaling
- ...



AI-Methods for EO: learning from data

Common AI applications in EO

- Classification
- Object detection
- Time series prediction
- Outlier detection
- Gapfilling
- Downscaling
- ...



(Some) limitations of AI

- Physical consistency (e.g., out-of-distribution prediction into future climate)
- Interpretability (e.g., understanding involved processes)

AI for land surface modeling?

Objectives of land surface modeling

- Process understanding →missing interpretability
- Running scenarios →missing physical consistency
- Simulate unobserved variables →missing interpretability

AI for land surface modeling?

Objectives of land surface modeling

- Process understanding →missing interpretability
- Running scenarios →missing physical consistency
- Simulate unobserved variables →missing interpretability

AI alone wont solve land surface modeling :(

Prior knowledge + big data + deep learning = ...

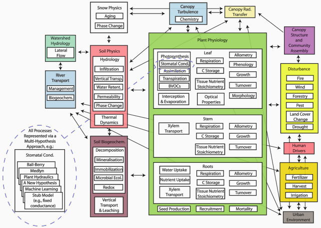


Figure 3: Prior knowledge: process parameterizations, causal pathways.

+

Figure 4: Big data from Earth observations, reanalysis, climate models.

+

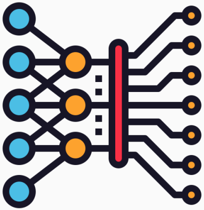


Figure 5: Deep learning: specialiced models, computation power.

Hybrid modeling

Hybrid modeling: Background and preliminary work

Hybrid modeling combines physically-based modeling and neural networks.

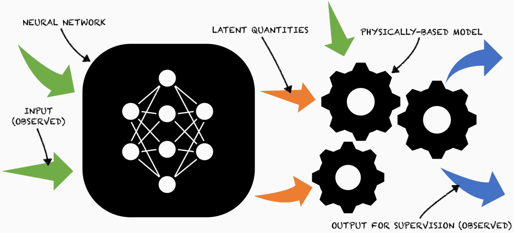


Figure 6: Hybrid models combine neural nets and physically-based modeling.

Literature

- *Reichstein et al. (2019), Combining System Modeling and Machine Learning into Hybrid Ecosystem Modeling, Nature*
- *Kraft et al. (2022), Towards hybrid modeling of the global hydrological cycle, HESS*

Hydrol. Earth Syst. Sci., 26, 1579–1614, 2022
https://doi.org/10.5194/hess-26-1579-2022

Hydrology and Earth System Sciences

Open Access

EGU

Towards hybrid modeling of the global hydrological cycle

Basil Kraft[1,2], Martin Jung[1], Marco Körner[2], Sujan Koirala[1], and Markus Reichstein[1]

[1]Department of Biogeochemical Integration, Max Planck Institute for Biogeochemistry, Germany
[2]Department of Aerospace and Geodesy, Technical University of Munich, Germany

Correspondence: Basil Kraft (bkraft@bgc-jena.mpg.de)

Received: 16 April 2021 – Discussion started: 25 May 2021
Revised: 25 January 2022 – Accepted: 31 January 2022 – Published: 23 March 2022

Abstract. State-of-the-art global hydrological models (GHMs) exhibit large uncertainties in hydrological simulations due to the complexity, diversity, and heterogeneity of the land surface and subsurface processes, as well as the scale dependency of these processes and associated parameters. Recent progress in machine learning, fueled by relevant Earth observation data streams, may help overcome

ory. The simulated contributions of groundwater, soil moisture, and snowpack variability to TWS variations are plausible and within the ranges of traditional GHMs. H2M identifies a somewhat stronger role of soil moisture for TWS variations in transitional and tropical regions compared to GHMs.
With the findings and analysis, we conclude that H2M provides a new data-driven perspective on modeling the global

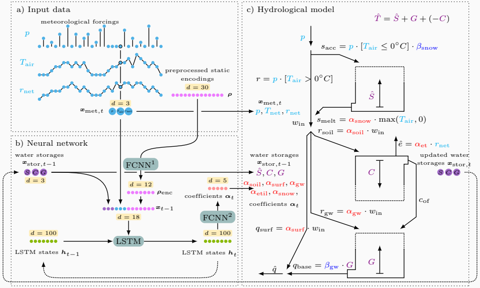


Figure 7: A dynamic neural network learns spatio-temporally varying parameters of a physical module in an end-to-end setup, *Kraft et al. (2022)*.

Hybrid modeling

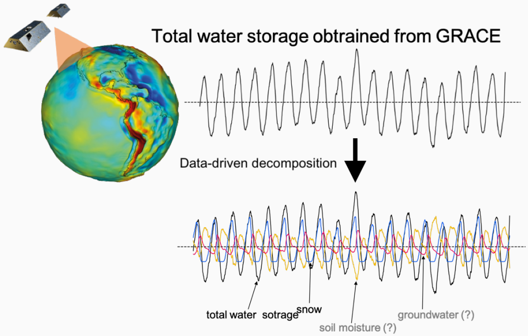


Figure 8: A data-driven estimate of water storage components regularized by physical constraints, *Kraft et al. (2022)*.

Main findings

- Hybrid modeling on large scale and complex systems works
- We got interesting insights into the hydrological cycle
- We identified a range of challenges, potential improvements → DUKE

The DUKE project

DUKE: team

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

(hybrid modeling, domain expertise)

TUM

(uncertainty quantification, ML expertise)

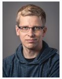
Zavud Baghirov (PhD student)

Basil Kraft (PI)

Martin Jung (CO-PI)

Markus Reichstein (CO-PI)

(PhD student)



Marco Körner (PI)

Bundesministerium für Wirtschaft und Klimaschutz

Funded by BMWK

DUKE

Objectives

→ Science Better represent and understand water × carbon cycle

→ Methods Advance hybrid modeling/uncertainty quantification

Block 1 (work in progress)

- Uncertainty quantification (TUM)
- Replicate results from previous work (MPI)
- Improve parameterizations of some processes (*e.g.*, soil moisture) (MPI)
- Include additional data constraints (MPI)

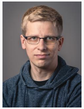
Block 2

- Couple water and carbon cycles
- Uncertainty-aware hybrid model

Uncertainty quantification

Uncertainty quantification

- Uncertainty quantification is useful for most applications of AI in EO
- Uncertainty in hybrid models could help assess robustness and identify equifinality, and ultimately develop better models.



→ Marco Körner

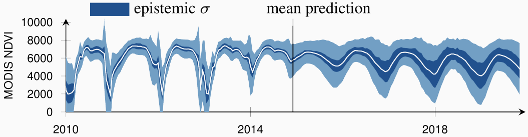
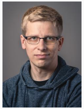


Figure 9: Episdemic and alleatoric uncertainty quantification for NDVI time-series prediction, *Rußwurm et al. (2020), IGARSS.*

Uncertainty quantification

- Uncertainty quantification is useful for most applications of AI in EO
- Uncertainty in hybrid models could help assess robustness and identify equifinality, and ultimately develop better models.



→ Marco Körner

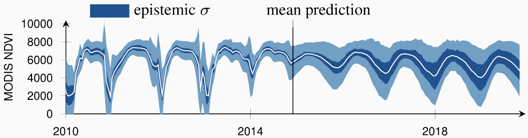


Figure 9: Episdemic and alleatoric uncertainty quantification for NDVI time-series prediction, *Rußwurm et al. (2020), IGARSS.*

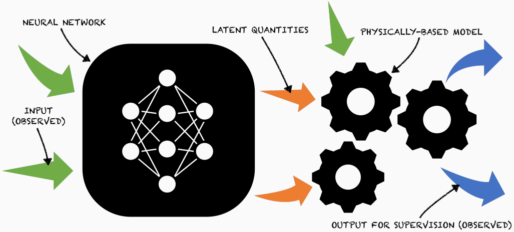


Figure 10: Can methods for quantifying uncertainty be applied to hybrid modeling?

Process understanding and products for downstream tasks

Estimating rootzone storage capacity

- A neural network f_{NN} learns a mapping from landscape features (soil, land cover, etc.) Z_s to maximum soil moisture capacity ($s_{\mathrm{max},s}$) at location s: $s_{\mathrm{max},s} = \mathrm{Softplus}(f_{\mathrm{NN}}(Z_s))$
- $s_{\mathrm{max},s}$ regulates soil moisture dynamics and is thereby indirectly constrained by observations of total water storage, evapotranspiration, and runoff.

Preliminary results!

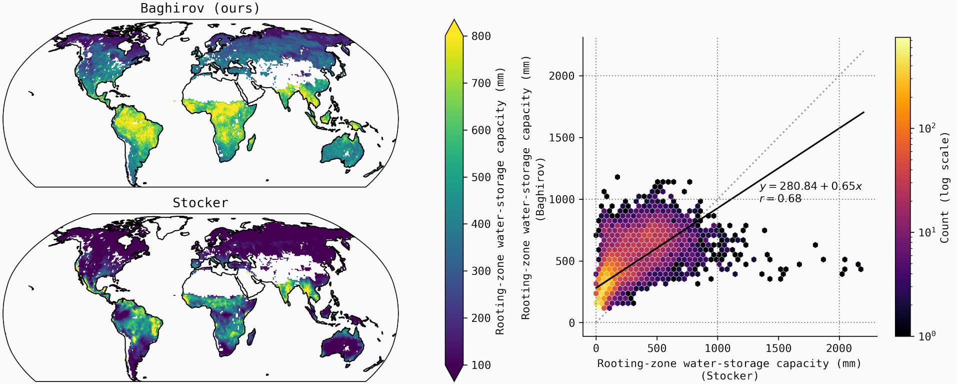


Figure 11: Comparison of hybrid (ours) and data-driven estimate of rooting-zone water-storage capacity from *Stocker et al. (2023), Nat. Geosci.*

Summary

- We are working on an improved hybrid model of the global hydrological cycle
- We plan to extend the model to cover the carbon cycle
- Uncertainty estimates could help to build model trust and identify weak points
- Hybrid modeling is complementary to existing approaches, it adds a datadriven yet physically constrained perspective