Spontaneous ELLIS Hackathon

Exploring

Identifiability – Explainability – Extrapolation – Causality

How much solar energy is converted into chemical energy?

(1) $F_{GPP} = \epsilon_{max} \times PAR$ → Photosynthetic Active Radiation

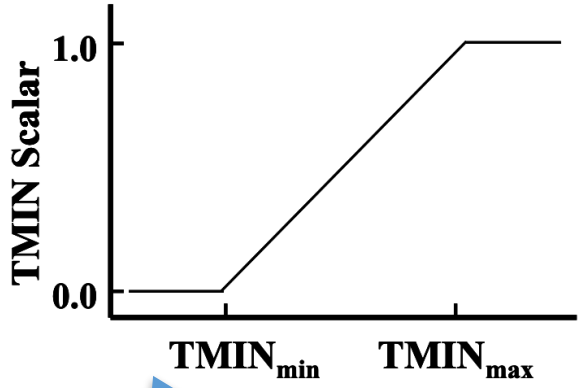
↘ Maximum Radiation-Use Efficiency

(2) $F_{GPP} = \epsilon_{max} \times PAR \times fPAR$ → Fraction of Absorbed Photosynthetic Active Radiation

Add temperature dependency

(3) $F_{GPP} = \epsilon_{max} \times PAR \times fPAR \times f_1(T_{min})$

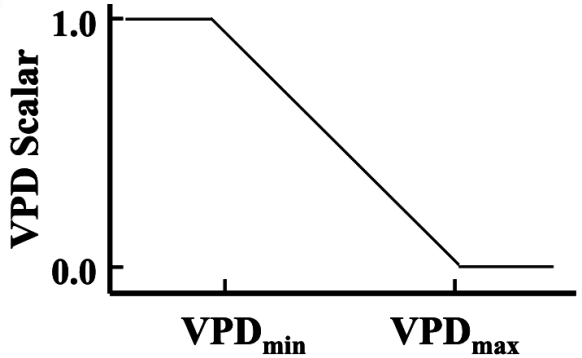
↙ Daily Minimum Temperature



Add water stress dependency

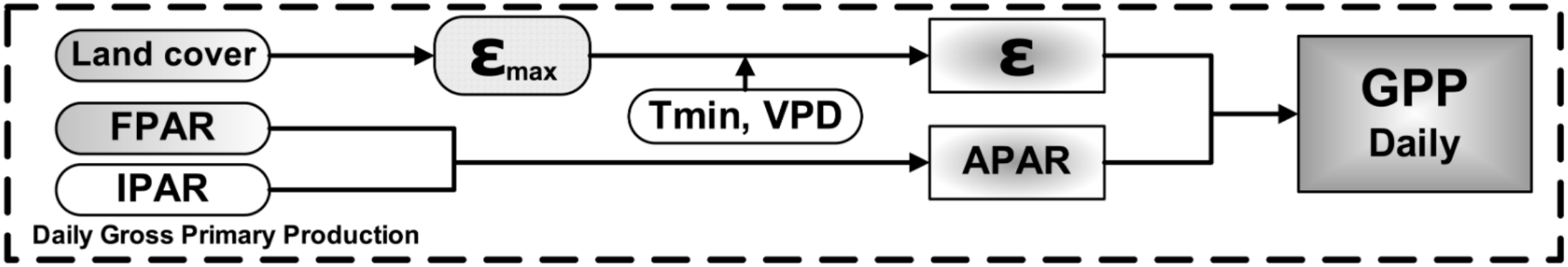
(4) $F_{GPP} = \epsilon_{max} \times PAR \times fPAR \times f_1(T_{min}) \times f_2(VPD)$

↙ Vapor Pressure Deficit



MODIS GPP Model (MOD17)

The MOD17 GPP Model at one glance



More details: https://www.ntsg.umt.edu/project/modis/user-guides/mod17c61usersguidev11mar112021.pdf

How?

Join the mattermost group

https://minervamessenger.mpdl.mpg.de/signup_user_complete/?id=5yc6iq67tigduxko8yd9snaine

pinplex/**Hackathon-Attribution**

This repository contains code (toy models) and data for the attribution and hybrid modeling hackathon.

1 Contributor 0 Issues 0 Stars 0 Forks

Get the GitHub repo

https://github.com/pinplex/Hackathon-Attribution

FOUR TASKS:

1. Hybrid Modeling
2. Extrapolation
3. Explainable AI
4. Causal Inference

Can we fit parts of the MODIS model using hybrid modeling?

$$\text{GPP} = \epsilon_{\text{max}} \times \underbrace{\text{Temperature Scalar}}_{f_{TS}(t_{min})} \times \underbrace{\text{Water Stress Scalar}}_{f_{WSS}(\text{vpd})} \times \underbrace{\text{Radiation Scalar}}_{f_{RS}(\text{rad},\text{FAPAR})}$$

Can we fit parts of the MODIS model using hybrid modeling?

$$\text{GPP} = \epsilon_{\text{max}} \times \underbrace{\text{Temperature Scalar}}_{f_{TS}(t_{min})} \times \underbrace{\text{Water Stress Scalar}}_{f_{WSS}(vpd)} \times \underbrace{\text{Radiation Scalar}}_{f_{RS}(rad,FAPAR)}$$

Pure ML Model

Model GPP as a function:

$$\text{GPP} = \text{NN}(t_{min}, vpd, rad, FAPAR)$$

Can we fit parts of the MODIS model using hybrid modeling?

$$\text{GPP} = \epsilon_{\text{max}} \times \underbrace{\text{Temperature Scalar}}_{f_{TS}(t_{min})} \times \underbrace{\text{Water Stress Scalar}}_{f_{WSS}(vpd)} \times \underbrace{\text{Radiation Scalar}}_{f_{RS}(rad,FAPAR)}$$

Pure ML Model

Model GPP as a function:

$$\text{GPP} = \text{NN}(t_{min}, vpd, rad, FAPAR)$$

Hybrid Model

Replace only f_{WSS}:

$$\text{Water Stress Scalar} = \text{NN}(vpd)$$

$$\text{GPP} = f_{TS}(t_{min})\text{NN}(vpd)f_{RS}(rad, FAPAR)$$

Can we fit parts of the MODIS model using hybrid modeling?

$$\text{GPP} = \epsilon_{\text{max}} \times \underbrace{\text{Temperature Scalar}}_{f_{TS}(t_{min})} \times \underbrace{\text{Water Stress Scalar}}_{f_{WSS}(vpd)} \times \underbrace{\text{Radiation Scalar}}_{f_{RS}(rad,FAPAR)}$$

Pure ML Model

Model GPP as a function:

$$\text{GPP} = \text{NN}(t_{min}, vpd, rad, FAPAR)$$

Hybrid Model

Replace only f_{WSS}:

$$\text{Water Stress Scalar} = \text{NN}(vpd)$$

$$\text{GPP} = f_{TS}(t_{min})\text{NN}(vpd)f_{RS}(rad, FAPAR)$$

- **Task 1:** Does the hybrid model output the correct GPP?

- **Task 1:** Does the hybrid model output the correct GPP?
- **Task 2:** Does the hybrid model identify the correct f_{WSS}?

- **Task 1:** Does the hybrid model output the correct GPP?
- **Task 2:** Does the hybrid model identify the correct f_{WSS}?
- **Bonus Task 1:** If we use a more complex function for f_{WSS}, can the hybrid model still identify it?

- **Task 1:** Does the hybrid model output the correct GPP?
- **Task 2:** Does the hybrid model identify the correct f_{WSS}?
- **Bonus Task 1:** If we use a more complex function for f_{WSS}, can the hybrid model still identify it?
- **Bonus Task 2:** Can we identify f_{WSS} and f_{RS} simultaneously?

Does the model generalize across different regimes?

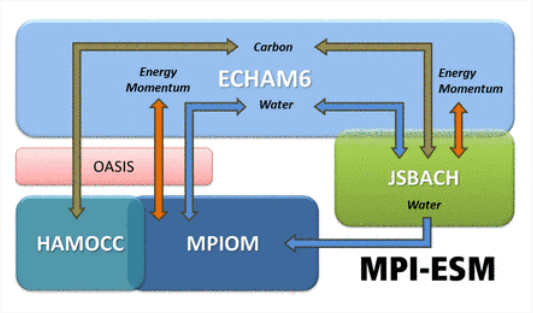
Does the model generalize across different regimes?

Today

Variables from ERA5 in Jena, Germany

Transient

Output of a transient run of an ESM at Jena, Germany



- **Task 1:** Does the model generalize to the different regimes if we only use one in training?

- **Task 1:** Does the model generalize to the different regimes if we only use one in training?
- **Task 2:** Does the hybrid model generalize better than the pure ML model?

Did the pure ML model identify the correct function f_{WSS}?

- **Task 1:** Use your favorite explanation method for interpreting neural networks to determine whether the pure ML model identified the correct function f_{WSS}.

Did the pure ML model identify the correct function f_{WSS}?

- **Task 1:** Use your favorite explanation method for interpreting neural networks to determine whether the pure ML model identified the correct function f_{WSS}.
- **Task 2:** Use your second favorite explanation method and compare the results.

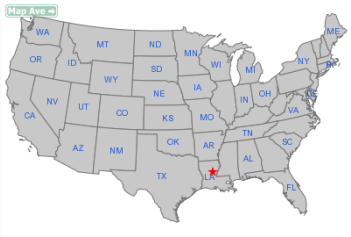
We have data (GPP, t_{min}, vpd, rad, FAPAR) for three locations:

Jena, Germany

We have data (GPP, t_{min}, vpd, rad, FAPAR) for three locations:

Jena, Germany

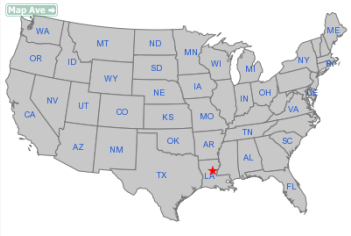
Jena, Louisiana



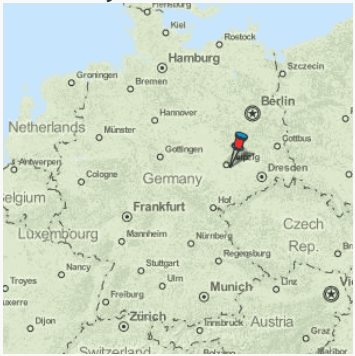
We have data (GPP, t_{min}, vpd, rad, FAPAR) for three locations:

Jena, Germany

Jena, Louisiana



The pixel next to Jena, Germany



- **Task 1:** The GPP in Jena, Germany causally only depends on the meteorological variables in Jena, Germany. Can you correctly identify this causal structure when we use the variables from Jena, Germany and Jena, USA?

- **Task 1:** The GPP in Jena, Germany causally only depends on the meteorological variables in Jena, Germany. Can you correctly identify this causal structure when we use the variables from Jena, Germany and Jena, USA?
- **Task 2:** The same as Task 1 but using Jena, Germany and its neighboring grid-point?

- **Task 1:** The GPP in Jena, Germany causally only depends on the meteorological variables in Jena, Germany. Can you correctly identify this causal structure when we use the variables from Jena, Germany and Jena, USA?
- **Task 2:** The same as Task 1 but using Jena, Germany and its neighboring grid-point?
- **Task 3:** Can we use XAI to identify the correct drivers?

- **Task 1:** The GPP in Jena, Germany causally only depends on the meteorological variables in Jena, Germany. Can you correctly identify this causal structure when we use the variables from Jena, Germany and Jena, USA?
- **Task 2:** The same as Task 1 but using Jena, Germany and its neighboring grid-point?
- **Task 3:** Can we use XAI to identify the correct drivers?
- **Bonus Task:** Do the results change if we change the sampling frequency of GPP?