MAX PLANCK INSTITUTE
FOR BIOGEOCHEMISTRY

Ecosystem Insights through Extreme Values: A Fresh Look at Meteorological Drivers

Jena, November 3, 2025

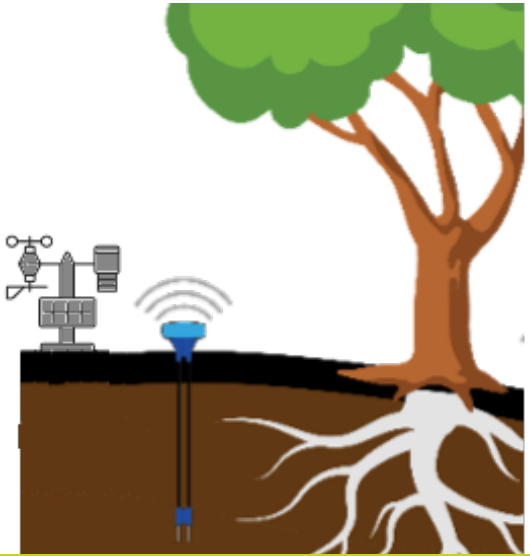
Christian Reimers, Claire Robin and Alexander Winkler

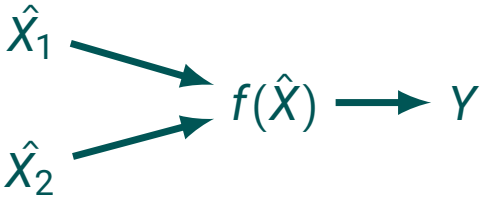
Research Gap

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

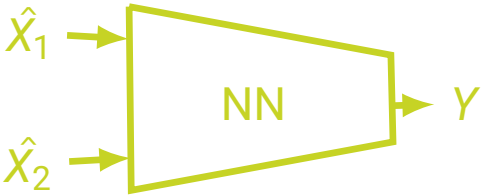
NEURAL INFORMATION PROCESSING SYSTEMS

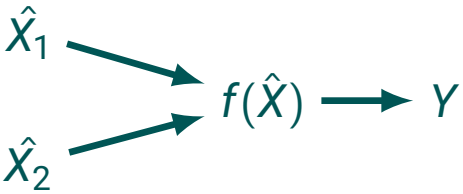
- The main purpose of data-driven models is predicting some response to specific inputs.
- For example, the amount of CO_2 uptake given meteorological forcings, or the amount of evaporation in a situation.
- However, in many disciplines researchers also build data- driven models in order to **identify the correct mechanism**.



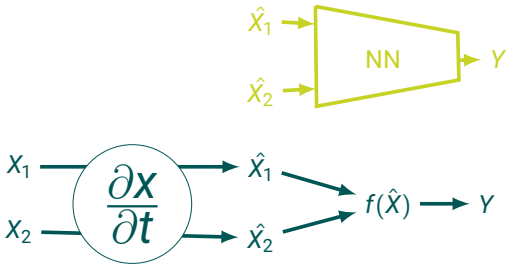


- When we build a data-driven model, we want to identify a mechanism $f(\hat{X})$ that determines some outcome Y from some drivers X_1 and X_2.

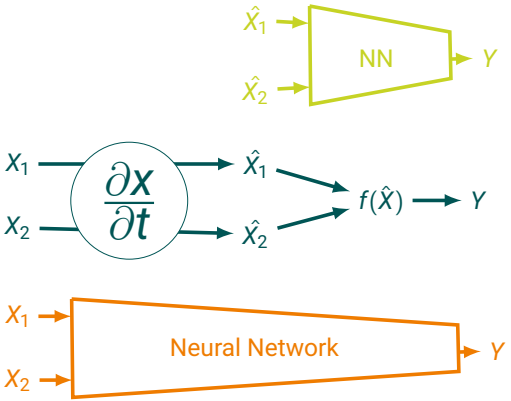




- We can understand this mechanism by training a neural network.



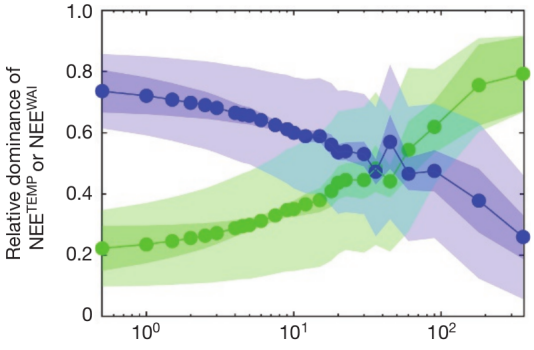
- However, in real applications, we do not measure the inputs directly when they enter the mechanism.
- Instead, the variables interact between the point we measure and the point of entering the mechanism.



- Therefore, when building a data-driven model, we model the combination of the dynamic system and the target mechanism.
- **Resulting in a biased prediction**

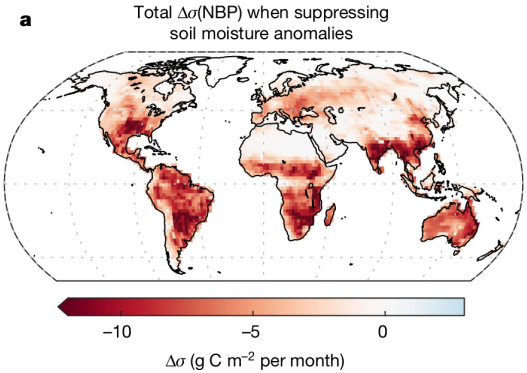
Where is this Relevant?

- (Jung et al., 2017) find that plant productivity is locally mostly influenced by water availability but globally by temperature.



Where is this Relevant?

- (Jung et al., 2017) find that plant productivity is locally mostly influenced by water availability but globally by temperature.
- (Seneviratne et al., 2013) find that global models lose more than 80% of the variability in plant productivity when supressing the variability in soil moisture.

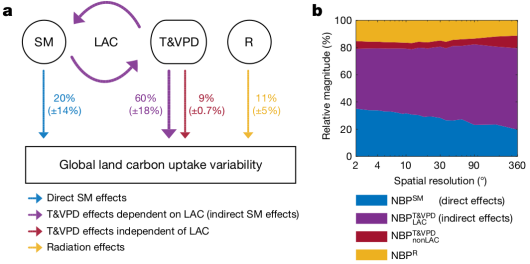


Where is this Relevant?

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

NEURAL INFORMATION PROCESSING SYSTEMS

- (Jung et al., 2017) find that plant productivity is locally mostly influenced by water availability but globally by temperature.
- (Seneviratne et al., 2013) find that global models lose more than 80% of the variability in plant productivity when supressing the variability in soil moisture.
- (Humphrey et al., 2021) claim that this difference is due to the land–atmosphere feedback.



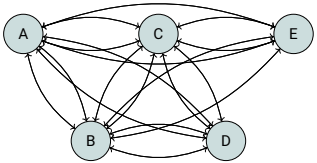
We Propose

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

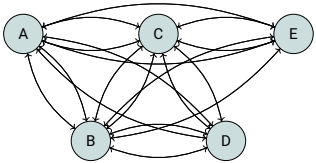
NEURAL INFORMATION PROCESSING SYSTEMS

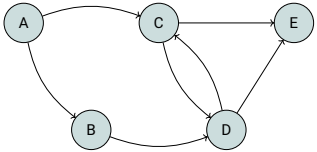
We propose to train only on extreme values.

Why?



Why?

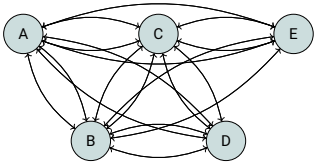


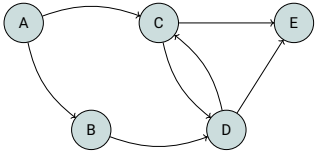


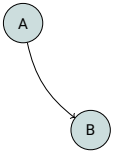
Why?

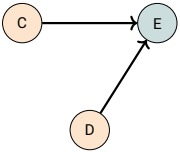
MAX PLANCK INSTITUTE
FOR BIOGEOCHEMISTRY

NEURAL INFORMATION
PROCESSING SYSTEMS

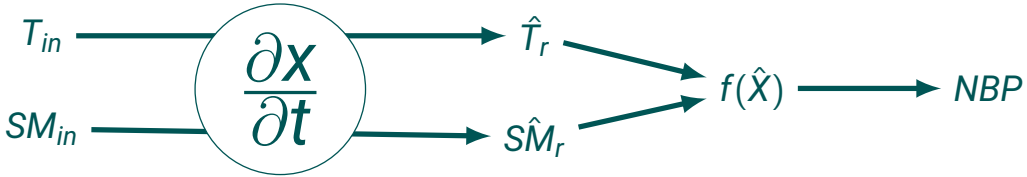




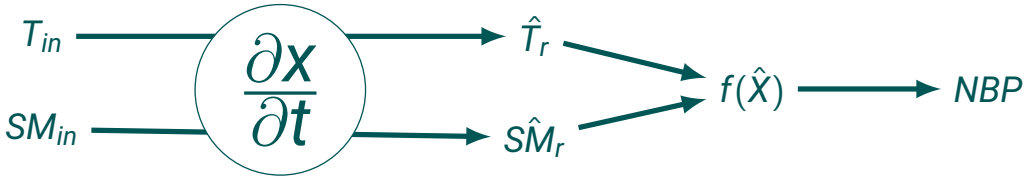




Does it work?

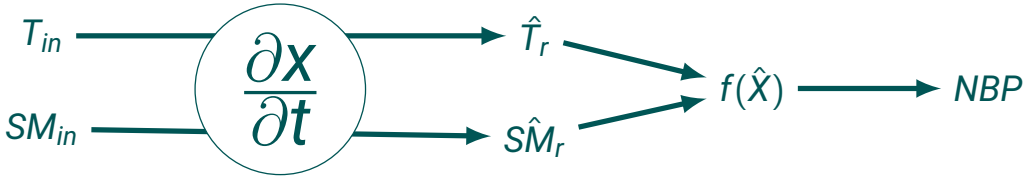


Does it work?



$$NBP = \alpha T_r + \beta SM_r + \gamma + \epsilon$$

Does it work?



$$T_r = T_{in} - \frac{C_T}{1 + e^{-(SM_{in}T_{in} - C_0)}}$$

$$SM_r = SM_{in} - \frac{C_{SM}}{1 + e^{-(SM_{in}T_{in} - C_0)}}$$

$$NBP = \alpha T_r + \beta SM_r + \gamma + \epsilon$$

Theoretical Result

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

NEURAL INFORMATION PROCESSING SYSTEMS

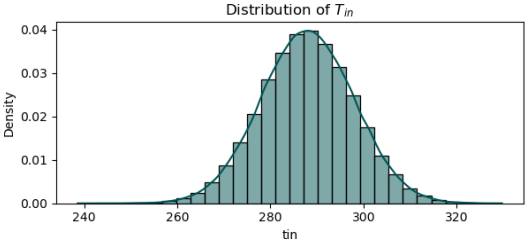
If we can make reasonable assumptions, a linear regression on T_{in} and SM_{in} does not return the correct values α and β but

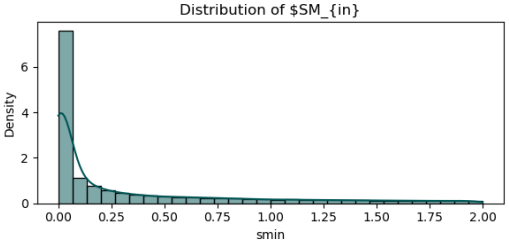
$$\alpha^* = \frac{\alpha}{1 - C_T SM_{in}} \quad \textit{and} \quad \beta^* = \frac{\beta}{1 - C_{SM} T_{in}}.$$

However, if we train only on extreme values, we find

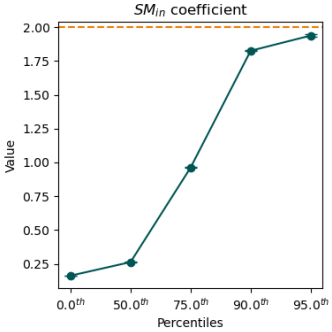
$$\alpha^* = \alpha \quad \textit{and} \quad \beta^* = \beta.$$

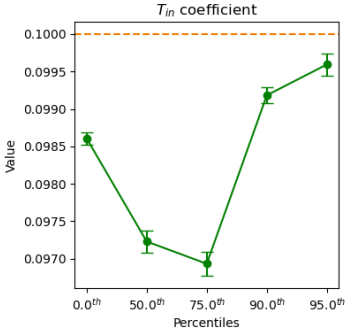
Empirical Results

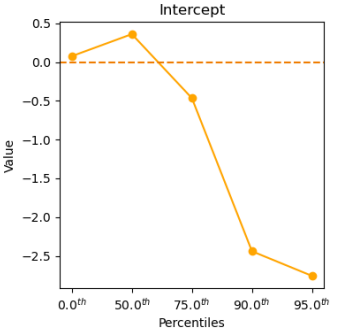




Empirical Results







Conclusion and Future Work

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

NEURAL INFORMATION PROCESSING SYSTEMS

- Training on extreme values recovers the true mechanism while training on all values does not.

Conclusion and Future Work

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

NEURAL INFORMATION PROCESSING SYSTEMS

- Training on extreme values recovers the true mechanism while training on all values does not.
- For the example we showed this theoretically as well as empirically.

Conclusion and Future Work

MAX PLANCK INSTITUTE FOR BIOGEOCHEMISTRY

NEURAL INFORMATION PROCESSING SYSTEMS

- Training on extreme values recovers the true mechanism while training on all values does not.
- For the example we showed this theoretically as well as empirically.
- The next step is to test this with more realistic models and, afterward, on real data.