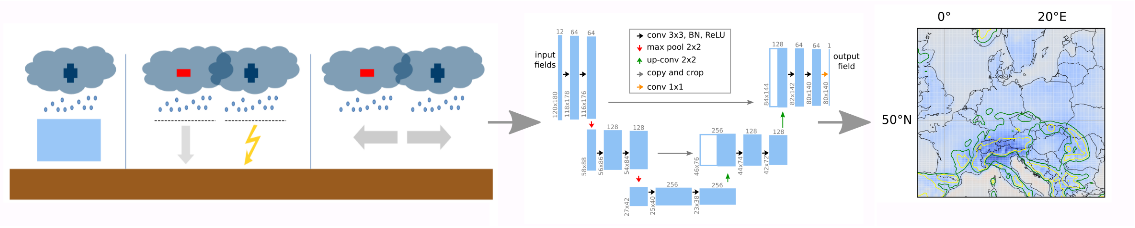


Causal DL models for studying relations in the Earth system

Impact of soil moisture changes on precipitation

ECMWF ML Workshop 2022 | Tobias Tesch, Stefan Kollet, Jochen Garcke | IBG-3, Forschungszentrum Jülich

JÜLICH Forschungszentrum

Motivation

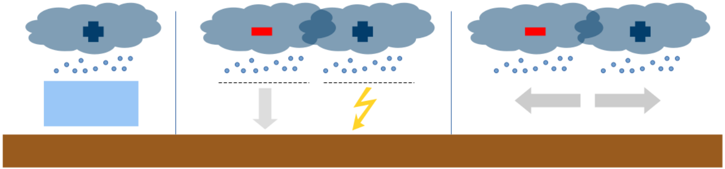
Many processes and relations between variables in the Earth system are still poorly understood.

Motivation

Many processes and relations between variables in the Earth system are still poorly understood.

Example: Impact of increased soil moisture on precipitation.

Different effects of soil moisture increases and their impact on precipitation (+/-).



(a) Increased atmospheric water content (+).

(b) Decreased atmospheric boundary layer (ABL) height (-), increased moist static energy in the ABL (+).

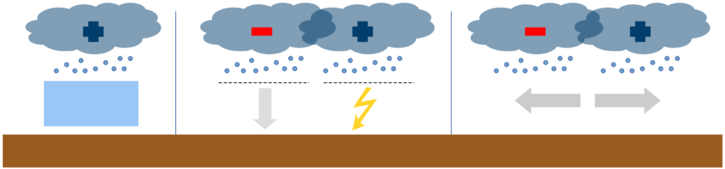
(c) Effects of horizontal soil moisture inhomogeneities on mesoscale circulation (+/-).

Motivation

Many processes and relations between variables in the Earth system are still poorly understood.

Example: Impact of increased soil moisture on precipitation.

Different effects of soil moisture increases and their impact on precipitation (+/-).



(a) Increased atmospheric water content (+).

(b) Decreased atmospheric boundary layer (ABL) height (-), increased moist static energy in the ABL (+).

(c) Effects of horizontal soil moisture inhomogeneities on mesoscale circulation (+/-).

⇒ Better understanding might improve precipitation prediction with numerical models.

Motivation

Limitations of current approaches for studying relations in the Earth system:

Motivation

Limitations of current approaches for studying relations in the Earth system:

- Approaches based on numerical simulations:
  - high computational costs,
  - rely on correct representation of considered relations in the numerical model.

Motivation

Limitations of current approaches for studying relations in the Earth system:

- Approaches based on numerical simulations:
  - high computational costs,
  - rely on correct representation of considered relations in the numerical model.
- Current statistical approaches:
  - strong assumptions, e.g. linearity and locality of relations,
  - ignore discrepancy between correlation and causation.

Motivation

Limitations of current approaches for studying relations in the Earth system:

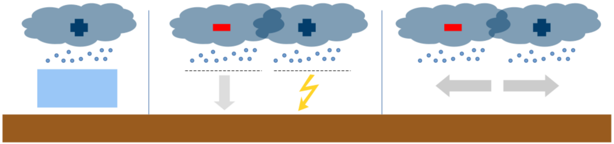
- Approaches based on numerical simulations:
  - high computational costs,
  - rely on correct representation of considered relations in the numerical model.
- Current statistical approaches:
  - strong assumptions, e.g. linearity and locality of relations,
  - ignore discrepancy between correlation and causation.

Combining DL and causality research, we can overcome these limitations!

Method

1 - Given a complex relation between two variables, ...

Different effects of soil moisture increases and their impact on precipitation (+/-).

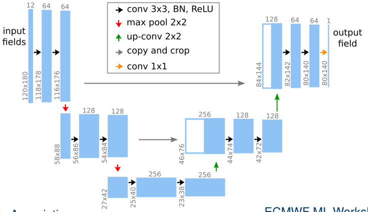


(a) Increased atmospheric water content (+).

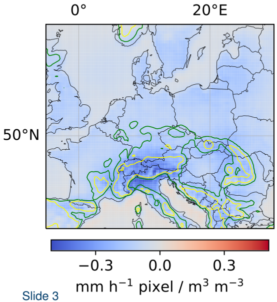
(b) Decreased atmospheric boundary layer (ABL) height (-), increased moist static energy in the ABL (+).

(c) Effects of horizontal soil moisture inhomogeneities on mesoscale circulation (+/-).

2 - ... we train a ***causal*** deep learning model to predict one variable given the other, ...



3 - ... and perform a sensitivity analysis of the trained model.



Causal impact

What is meant by the *causal impact* of some variable $X \in \mathbb{R}^d$ on another variable $Y \in \mathbb{R}^n$?

Causal impact

What is meant by the *causal impact* of some variable $X \in \mathbb{R}^d$ on another variable $Y \in \mathbb{R}^n$?

- It is the (expected) response of Y to intervening into the considered system and changing the value of X.

Causal impact

What is meant by the *causal impact* of some variable $X \in \mathbb{R}^d$ on another variable $Y \in \mathbb{R}^n$?

- It is the (expected) response of Y to intervening into the considered system and changing the value of X.
- We need to determine the expected value of Y given that one intervened into the system and set X to some value x, referred to by $\mathbb{E}[Y|do(X = x)]$.

Causal impact

What is meant by the *causal impact* of some variable $X \in \mathbb{R}^d$ on another variable $Y \in \mathbb{R}^n$?

- It is the (expected) response of Y to intervening into the considered system and changing the value of X.
- We need to determine the expected value of Y given that one intervened into the system and set X to some value x, referred to by $\mathbb{E}[Y|do(X = x)]$.
- In general, it holds $\mathbb{E}[Y|do(X = x)] \neq \mathbb{E}[Y|X = x]$.

Causal impact

What is meant by the *causal impact* of some variable $X \in \mathbb{R}^d$ on another variable $Y \in \mathbb{R}^n$?

- It is the (expected) response of Y to intervening into the considered system and changing the value of X.
- We need to determine the expected value of Y given that one intervened into the system and set X to some value x, referred to by $\mathbb{E}[Y|do(X = x)]$.
- In general, it holds $\boxed{\mathbb{E}[Y|do(X = x)] \neq \mathbb{E}[Y|X = x].}$

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t+3$ h].

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t + 3$ h].

Increased soil moisture[t]

Forschungszentrum

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t+3$ h].

Increased soil moisture[t]

$\Rightarrow$ likely increased recent precipitation

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t+3$ h].

Increased soil moisture[t]
$\Rightarrow$ likely increased recent precipitation
$\Rightarrow$ likely increased precipitation[$t+3$ h] (precipitation persistence).

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t+3$ h].

Increased soil moisture[t]
$\Rightarrow$ likely increased recent precipitation
$\Rightarrow$ likely increased precipitation[$t+3$ h] (precipitation persistence).

- The right expectation contains this confounding effect.

Causal impact

In general, it holds $\boxed{\mathbb{E}[Y|do(X=x)] \neq \mathbb{E}[Y|X=x].}$

Example: X - soil moisture[t], Y - precipitation[$t+3$ h].

Increased soil moisture[t]
$\Rightarrow$ likely increased recent precipitation
$\Rightarrow$ likely increased precipitation[$t+3$ h] (precipitation persistence).

- The right expectation contains this confounding effect.
- The left expectation does not, because the do-operator represents an intervention into the system that breaks the link between soil moisture and recent precipitation.

Causal deep learning model

Given a target variable $Y \in \mathbb{R}^n$, input variables $X \in \mathbb{R}^d$ and $\{C_i\}_{i=1}^k \in \mathbb{R}^{d_i}$, and a suitable loss function, a DL model approximates the map

$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y|X = x, \{C_i = c_i\}_{i=1}^k].$$

Causal deep learning model

Given a target variable $Y \in \mathbb{R}^n$, input variables $X \in \mathbb{R}^d$ and $\{C_i\}_{i=1}^k \in \mathbb{R}^{d_i}$, and a suitable loss function, a DL model approximates the map

$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y|X = x, \{C_i = c_i\}_{i=1}^k].$$

We call a DL model *causal*, if it approximates the map

$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y| \boxed{\text{do(X=x)}}, \{C_i = c_i\}_{i=1}^k].$$

Causal deep learning model

Given a target variable $Y \in \mathbb{R}^n$, input variables $X \in \mathbb{R}^d$ and $\{C_i\}_{i=1}^k \in \mathbb{R}^{d_i}$, and a suitable loss function, a DL model approximates the map

$$(x, \{c_i\}_{i=1}^k) \rightarrow \mathbb{E}[Y|X = x, \{C_i = c_i\}_{i=1}^k].$$

We call a DL model *causal*, if it approximates the map

$$(x, \{c_i\}_{i=1}^k) \rightarrow \mathbb{E}[Y| \boxed{\text{do(X=x)}}, \{C_i = c_i\}_{i=1}^k].$$

We can obtain a causal DL model by choosing suitable additional input variables C_i, because then it holds

$$\mathbb{E}[Y|do(X = x), \{C_i = c_i\}_{i=1}^k] = \mathbb{E}[Y|X = x, \{C_i = c_i\}_{i=1}^k].$$

Causal deep learning model

Theorem

Pearl[1]: "For multivalued variables X and Y, finding a sufficient set S of multivalued variables $C_i \in \mathbb{R}^{d_i}, i = 1, \ldots, k$, permits us to write

$$\mathbb{E}[Y|do(X = x), \{C_i = c_i\}_{i=1}^{k}] = \mathbb{E}[Y|X = x, \{C_i = c_i\}_{i=1}^{k}]. \text{"} \tag{1}$$

Sufficient set:

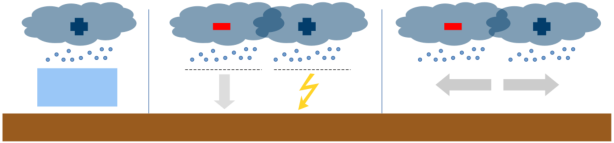
- "no element of S is a descendant of X",
- "the elements of S block all back-door paths from X to Y, namely all paths that end with an arrow pointing to X".

[1] Pearl, J. "Causal inference in statistics: An overview." Statist. Surv. 3, 96, 2009.

Method

1 - Given a complex relation between two variables, ...

Different effects of soil moisture increases and their impact on precipitation (+/-).

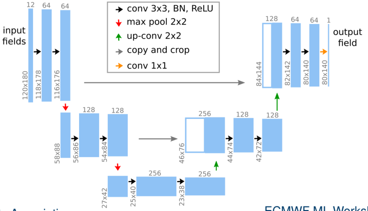


(a) Increased atmospheric water content (+).

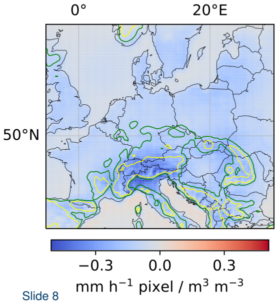
(b) Decreased atmospheric boundary layer (ABL) height (-), increased moist static energy in the ABL (+).

(c) Effects of horizontal soil moisture inhomogeneities on mesoscale circulation (+/-).

2 - ... we train a ***causal*** deep learning model to predict one variable given the other, ...



3 - ... and perform a sensitivity analysis of the trained model.



Sensitivity analysis of the trained model

- From step 2, we have a causal DL model, i.e. a DL model that approximates the map
$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y|do(X = x), \{C_i = c_i\}_{i=1}^k].$$

Sensitivity analysis of the trained model

- From step 2, we have a causal DL model, i.e. a DL model that approximates the map
$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y|do(X = x), \{C_i = c_i\}_{i=1}^k].$$
- To determine the causal impact of $X \in \mathbb{R}^d$ on $Y \in \mathbb{R}^n$, we consider the partial derivatives
$$s_{i_1 i_2} = \frac{\partial \mathbb{E}[Y_{i_1}|do(X = x), \{C_i = c_i\}_{i=1}^k]}{\partial X_{i_2}}, \text{ for } i_1 \in \{1, \ldots, n\}, i_2 \in \{1, \ldots, d\}.$$

Sensitivity analysis of the trained model

- From step 2, we have a causal DL model, i.e. a DL model that approximates the map

$$(x, \{c_i\}_{i=1}^k) \to \mathbb{E}[Y|do(X = x), \{C_i = c_i\}_{i=1}^k].$$

- To determine the causal impact of $X \in \mathbb{R}^d$ on $Y \in \mathbb{R}^n$, we consider the partial derivatives

$$s_{i_1 i_2} = \frac{\partial \mathbb{E}[Y_{i_1}|do(X = x), \{C_i = c_i\}_{i=1}^k]}{\partial X_{i_2}}, \text{ for } i_1 \in \{1, \dots, n\}, i_2 \in \{1, \dots, d\}.$$

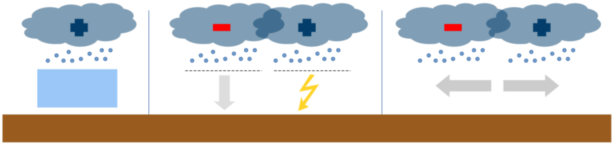
- To answer how Y_{i_1} changes on average if we intervened into the system and changed X_{i_2}, we consider the expected value of $s_{i_1 i_2}$ w.r.t. the joint distribution of X and $\{C_i\}_{i=1}^k$

$$\overline{s_{i_1 i_2}} = \mathbb{E}_{x, \{c_i\}_{i=1}^k}[s_{i_1 i_2}] = \mathbb{E}_{x, \{c_i\}_{i=1}^k}\left[\frac{\partial \mathbb{E}[Y_{i_1}|do(X = x), \{C_i = c_i\}_{i=1}^k]}{\partial X_{i_2}}\right].$$

Method

1 - Given a complex relation between two variables, …

Different effects of soil moisture increases and their impact on precipitation (+/-).

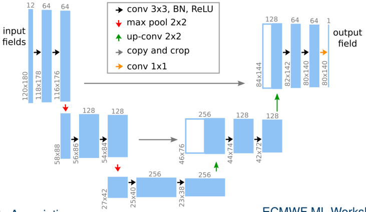


(a) Increased atmospheric water content (+).

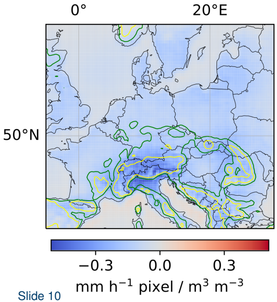
(b) Decreased atmospheric boundary layer (ABL) height (-), increased moist static energy in the ABL (+).

(c) Effects of horizontal soil moisture inhomogeneities on mesoscale circulation (+/-).

2 - … we train a ***causal*** deep learning model to predict one variable given the other, …



3 - … and perform a sensitivity analysis of the trained model.



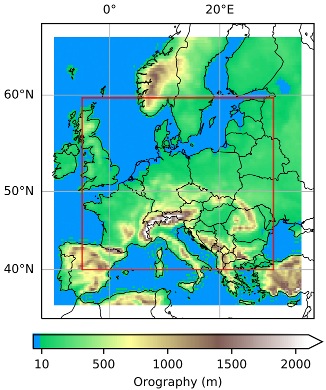
JÜLICH Forschungszentrum

Train a causal DL model to predict one variable given the other

Given soil moisture[t] and further variables[t] , which approximate a sufficient set, at the 120x180 pixels in the input region,

predict precipitation[t+3 h] at the 80x140 pixels in the target region (red box).

Data: ERA5. Summer months (JJA). Target variable between 11 am and 11 pm UTC.



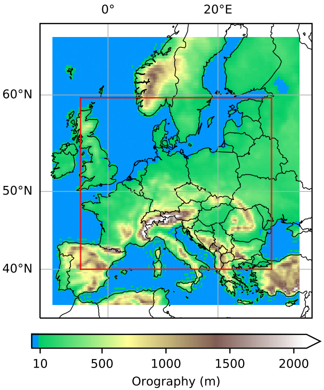
JÜLICH Forschungszentrum

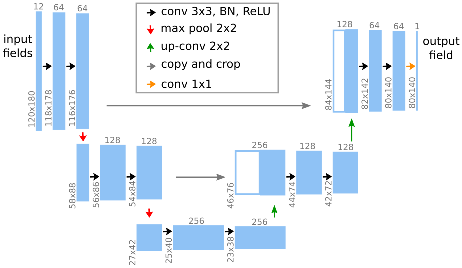
Train a causal DL model to predict one variable given the other

Given soil moisture[t] and further variables[t] , which approximate a sufficient set, at the 120x180 pixels in the input region,

predict precipitation[t+3 h] at the 80x140 pixels in the target region (red box).

Data: ERA5. Summer months (JJA). Target variable between 11 am and 11 pm UTC.



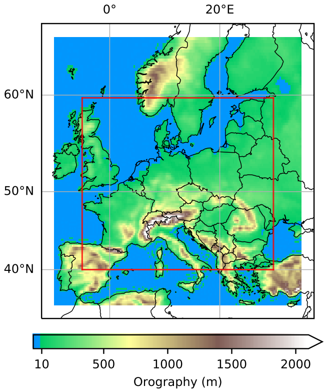


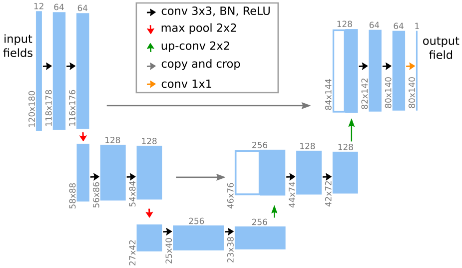
Train a causal DL model to predict one variable given the other

Given soil moisture[t] and further variables[t], which approximate a sufficient set, at the 120x180 pixels in the input region,

predict precipitation[t+3 h] at the 80x140 pixels in the target region (red box).

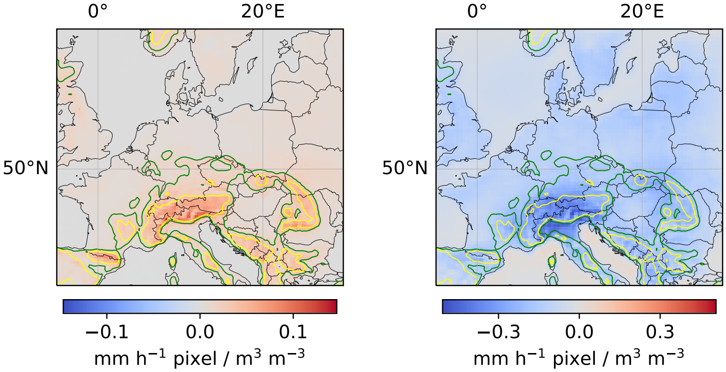
Data: ERA5. Summer months (JJA). Target variable between 11 am and 11 pm UTC.





Sensitivity analysis of the trained model

Impact of an increase in local soil moisture on local (left) and regional (right) precipitation.



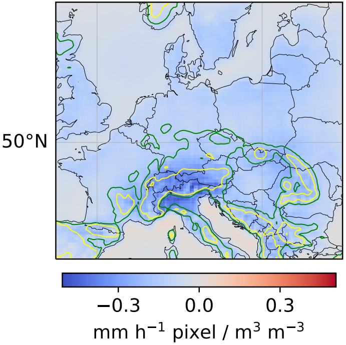
For each pixel ij: $\overline{\frac{\partial P_{ij}}{\partial x_{SM,ij}}(x)}$

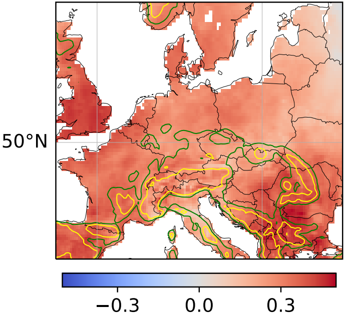
For each pixel ij: $\overline{\frac{\partial \sum_{nk} P_{nk}}{\partial x_{SM,ij}}(x)}$

JÜLICH Forschungszentrum

Comparison to linear correlation

Left: impact of an increase in local soil moisture on regional precipitation (our method).
Right: linear correlation between local soil moisture and regional precipitation.

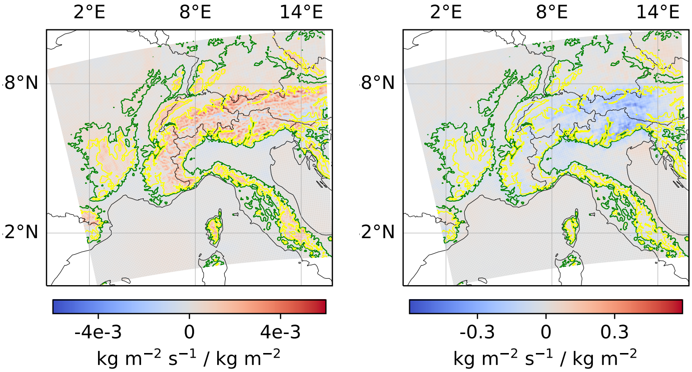




For each pixel *ij*: $\overline{\frac{\partial \sum_{nk} P_{nk}}{\partial x_{SM,ij}}(x)}$ For each pixel *ij*: $corr_t(SM_{ij}, \sum_{nk} P_{nk})$

Data from convection-permitting simulations

Impact of an increase in local soil moisture on local (left) and regional (right) precipitation.



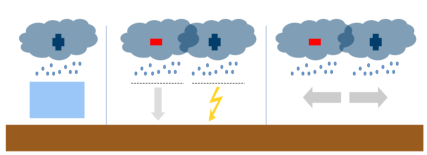
For each pixel ij: $\overline{\dfrac{\partial P_{ij}}{\partial x_{SM,ij}}(x)}$ For each pixel ij: $\overline{\dfrac{\partial \sum_{nk} P_{nk}}{\partial x_{SM,ij}}(x)}$

Data courtesy: CORDEX Flagship Pilot Study on convection-permitting regional climate modelling, WRF simulation results from Aristotle University of Thessaloniki-Dept Meteorology & Climatology, see also Ban, N. *et al.* The first multi-model ensemble of regional climate simulations at kilometer-scale resolution, part I: evaluation of precipitation. *Clim Dyn* **57**, 275–302 (2021).

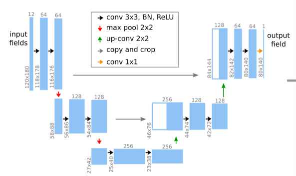
Summary & take home message

Causal DL for studying relations in the Earth system: impact of soil moisture changes on precipitation.

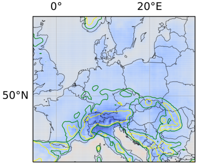
1 - Given a complex relation between two variables, …



2 - … we train a ***causal*** DL model to predict one variable given the other, …



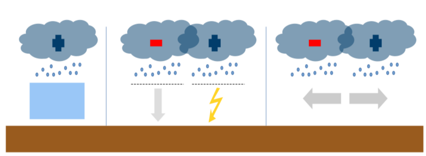
3 - … and perform a sensitivity analysis of the trained model.



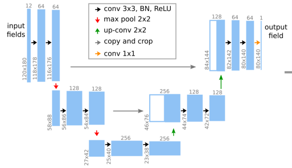
Summary & take home message

Causal DL for studying relations in the Earth system: impact of soil moisture changes on precipitation.

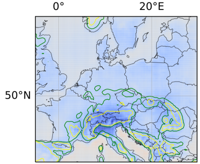
1 - Given a complex relation between two variables, ...



2 - ... we train a ***causal*** DL model to predict one variable given the other, ...



3 - ... and perform a sensitivity analysis of the trained model.

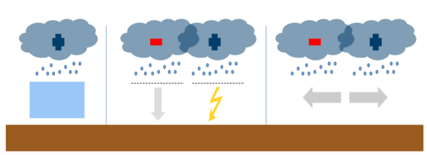


⇒ **Take home message:** Deep learning is a powerful tool for gaining scientific insights into the Earth system.

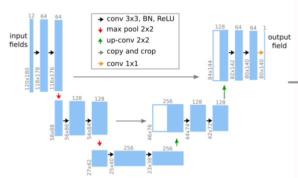
Summary & take home message

Causal DL for studying relations in the Earth system: impact of soil moisture changes on precipitation.

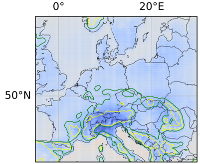
1 - Given a complex relation between two variables, ...



2 - ... we train a ***causal*** DL model to predict one variable given the other, ...



3 - ... and perform a sensitivity analysis of the trained model.



⇒ **Take home message:** Deep learning is a powerful tool for gaining scientific insights into the Earth system.

Get in touch: t.tesch@fz-juelich.de