WP5.1 Machine learning to advance climate model evaluation and process understanding

Authors: L. Bock [1], A. Lauer [1] and V. Eyring [1,2]

WP 5.1.1 Enhancing observational products for climate model evaluation with machine learning

An approach based on machine learning is developed and applied to derive cloud classes from high-resolution satellite data and coarse-resolution climate models. The aim is to allow for an improved evaluation of clouds in climate models by analysing cloud properties by cloud type. This enables to evaluate the different underlying processes driving formation and evolution of these cloud types in climate models. The method is then applied to ESACCI-Cloud data that are coarse-grained to the resolution of a typical climate model. As a proof of concept, the resulting timeseries of cloud class information from ESACCI-Cloud is then used for comparison with results from a simulation with the ICON-A model.

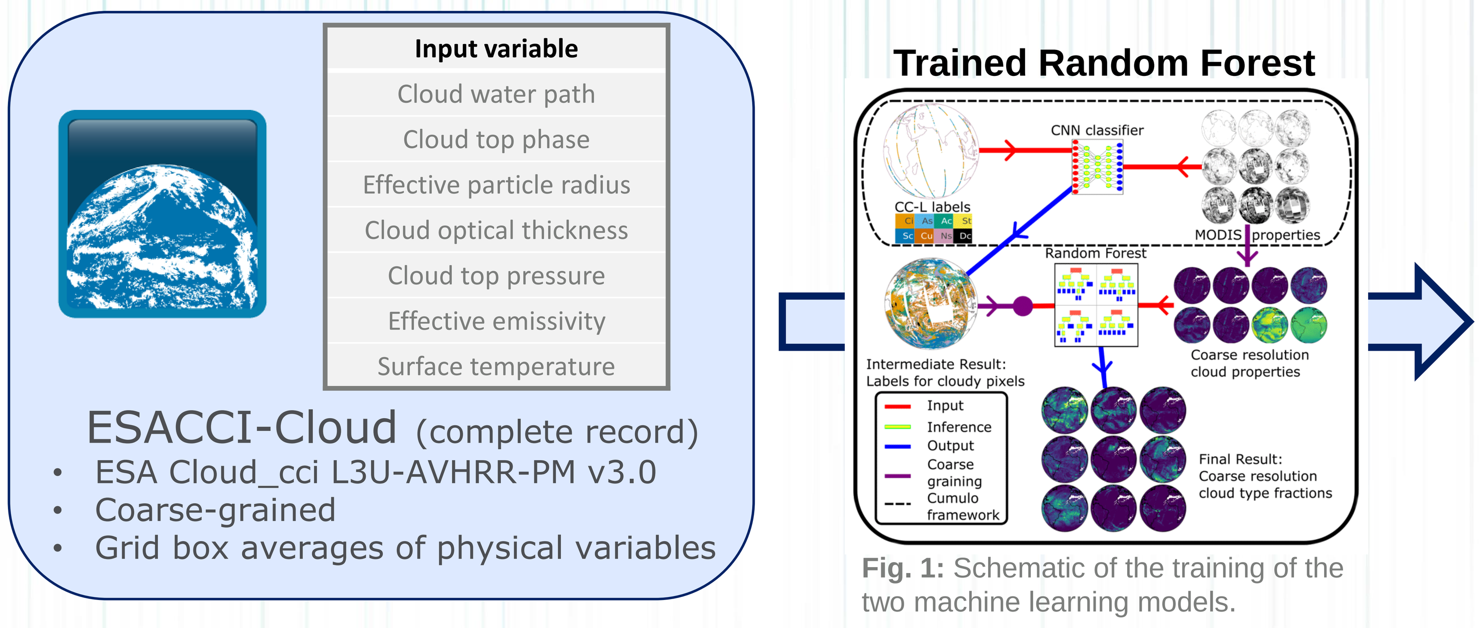


Fig. 1: Schematic of the training of the two machine learning models.

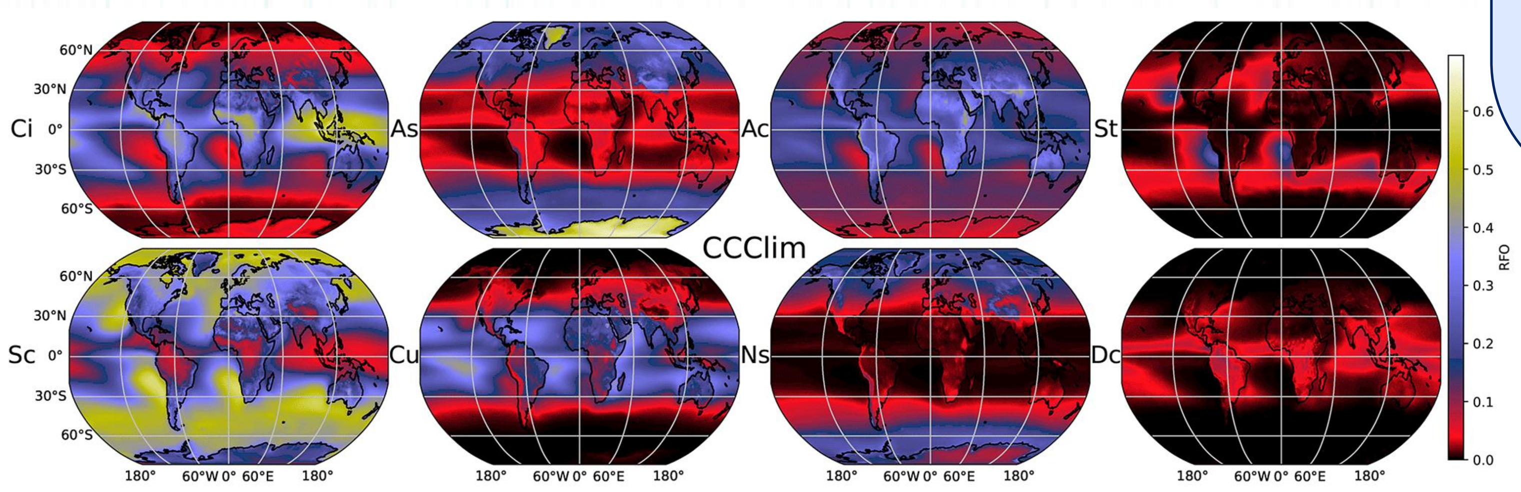


Fig. 2: Average geographical distribution of the relative frequencies of occurrence (RFOs) for all cloud types in CCClim.

Cloud Class Climatology dataset (CCClim, 1982-2016)

https://doi.org/10.5281/zenodo.8369201

- 8 WMO-like cloud types with a long coverage period (35 years) and high spatial resolution (1° x1°) as daily samples
- consistent seasonal variations, sensible regional distributions and little drift over the complete period
- all cloud types can be associated with relevant physical quantities

Kaps et al., 2024: Characterizing clouds with the CCClim dataset, a machine learning cloud class climatology, Earth Syst. Sci. Data

WP 5.1.2 Quantifying the causal effect of cloud controlling factors on marine stratocumulus clouds

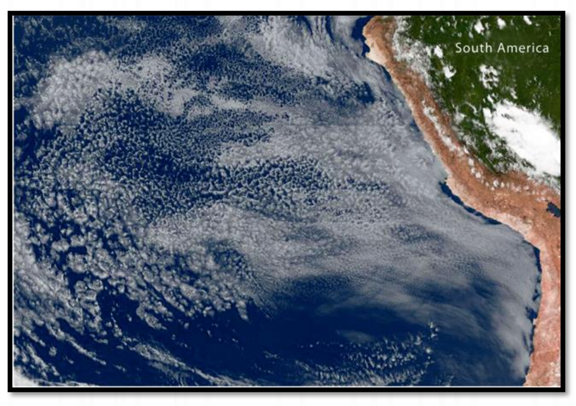


Fig. 3: Geostationary Operational Environmental Satellites (GOES) (Wood, 2012)

As a key component of the hydrological cycle and the Earth's radiation budget, clouds play an important role in both weather and climate. Our incomplete understanding of clouds and their role in cloud-climate feedbacks leads to large uncertainties in climate simulations. Using causal inference as an unsupervised machine learning method we aim to systematically analyse and quantify causal interdependencies and links between cloud properties and their controlling factors. This approach goes beyond correlation-based measures by systematically excluding common drivers and indirect links. By estimating the causal effect of each of the cloud controlling factors for different cloud regimes we expect to be able to better understand the dominant processes which determine the micro- and macro-physical properties of clouds.

Data

For this study we use 5 years (2003-2007) of daily data, which is averaged over 5°x5° regions (ocean grid cells only) and anomalized.

	Variable	Dataset
Cloud Properties	Total Cloud Fraction (clt), Total Cloud Water Path (clwvi), Cloud Optical Depth (cod), Cloud Effective Radius (reff), Cloud Top Height (ctp)	ESACCI-Cloud (v3.0, L3U, AVHRR-PM, NOAA-16, daily instantaneous data) (Stengel et al., 2020)
Cloud-controlling factors	Sea Surface Temperature (tos)	ESACCI-SST (v3.0, Level 4 Analysis Product, daily) (Good et al., 2024)
	Water Vapour Path (prw)	ESACCI-Watervapour (CM SAF/CCI TCWV-global (COMBI), v3.1, daily mean data) (Schröder et al., 2023)
	Vertical Velocity at 700hPa (wap700), Lower Tropospheric Stability (LTS), Sea Surface Pressure (psl), Sensible Heat Flux at Surface (hfss), 10m Horizontal Wind Speed (sfcWind)	ERA5 (daily average from hourly data) (C3S, 2017)

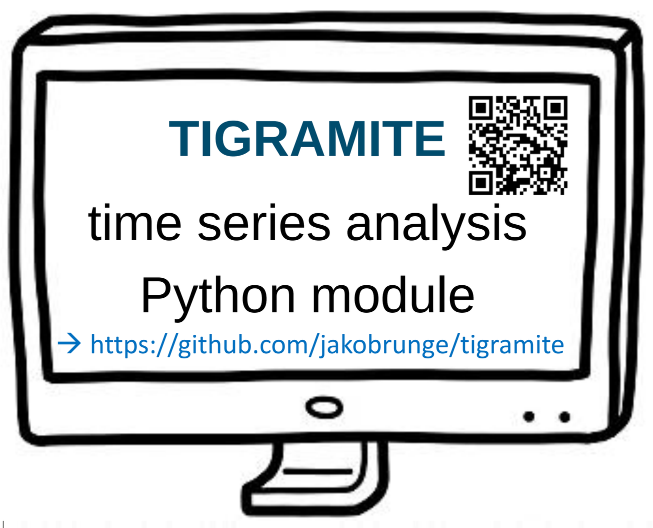
Method

PCMCI (*Runge et al., 2019*)

→ identifies causal relationships and quantifies their strengths from time series data

→ unsupervised machine learning

→ approach goes beyond correlation-based measures by systematically excluding common driver effects and indirect links

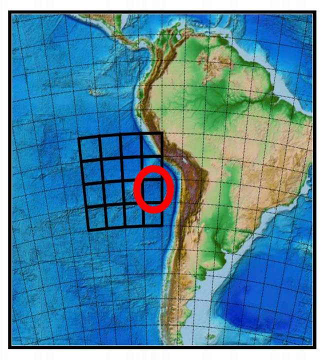


LPCMCI based on **Fast Causal Inference (FCI)** Algorithm: constraint-based causal discovery with conditional independence tests equal to Peter Clark (PC)-Algorithm but in the presence of unobserved variables (possibility of latent confounders)

CausalEffects class: allows to estimate (conditional) causal effects and mediation based on assuming a causal graph.

First results

First, we applied LPCMCI to one of the 16 regions separately. The results are shown here.



TIGRAMITE settings:
Method: LPCMCI
Conditional independence test: RobustParCorr
pc_alpha = 0.01
tau_max = 2
n_preliminary_iterations = 5

Causal network

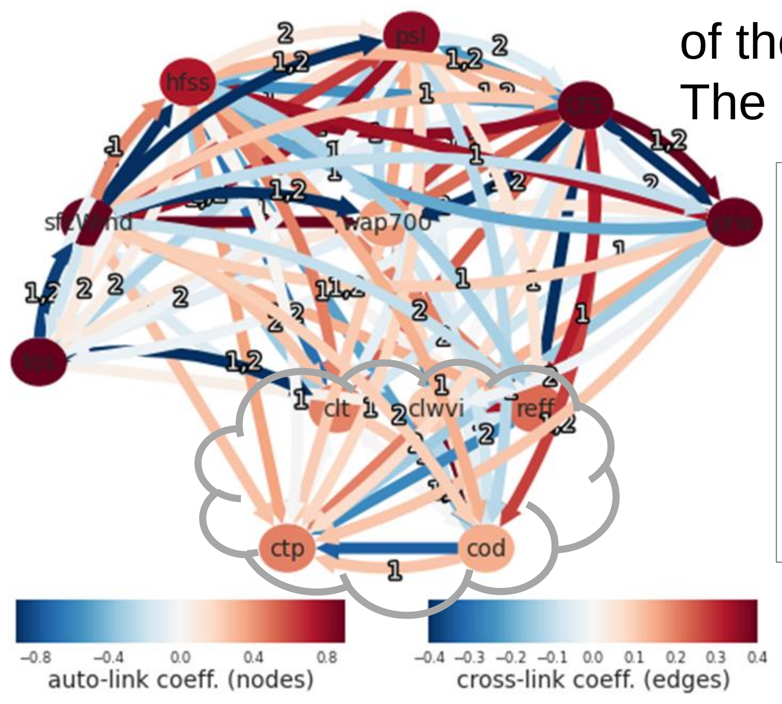


Fig. 4: Process graph of causal links. The link colour refers to the cross-link coeff. and the node colour denotes the auto-link coeff. at the lag with maximum absolute value.

Mediation analysis is a powerful tool to quantify not only direct causal links, but also indirect pathways as the total causal effect.

Total causal effect

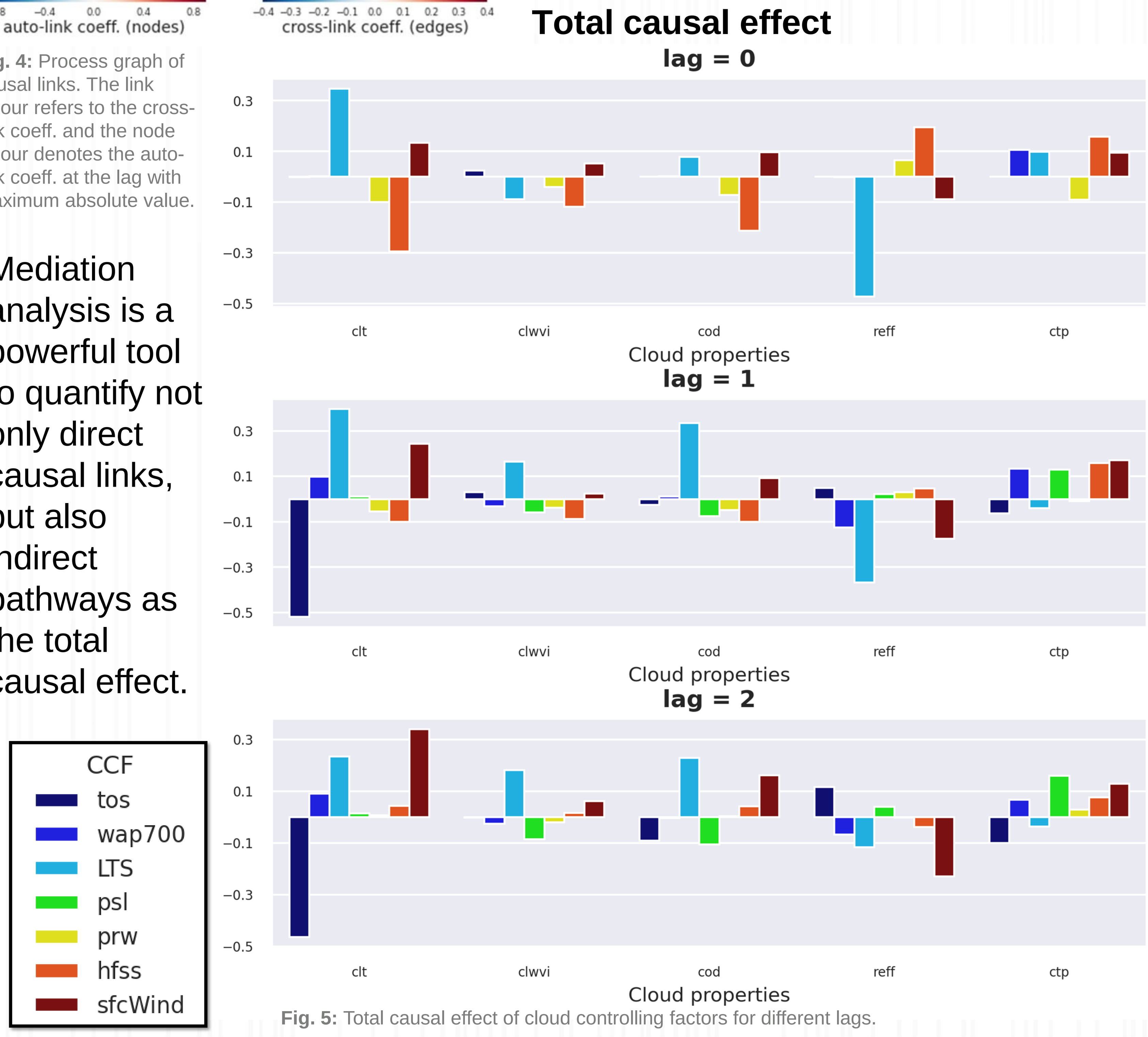


Fig. 5: Total causal effect of cloud controlling factors for different lags.

[1] Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany
[2] University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany

References
C3S (2017): ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate, https://cds.climate.copernicus.eu/cdsapp#!/home.
Good, S. A. and Embury, O.: ESA Sea Surface Temperature Climate Change Initiative (SST_cci): Level 4 Analysis product, version 3.0, https://doi.org/10.5285/4A9654136A7148E39B7FEB56F8BB02D2, [Titel anhand dieser DOI in Citavi-Projekt übernehmen] 2024
Kaps, A., Lauer, A., Kazeroni, R., Stengel, M., and Eyring, V.: Characterizing clouds with the CCClim dataset, a machine learning cloud class climatology, Earth Syst. Sci. Data, 16, 3001-3016, https://doi.org/10.5194/essd-16-3001-2024, 2024.
Runge, J. et al. (2019): Detecting and quantifying causal associations in large nonlinear time series datasets.Sci. Adv.5, eaau4996, https://doi.org/10.1126/sciadv.aau4996
Schröder, M. et al. (2023): A combined high resolution global TGWV product from microwave and near infrared imagers - COMBI, Satellite Application Facility on Climate Monitoring, DOI:10.5676/EUM_SAF_CM/COMBI/V001, https://doi.org/10.5676/EUM_SAF_CM/COMBI/V001.
Stengel, M. et al. (2020): Cloud_cci Advanced Very High Resolution Radiometer post meridiem (AVHRR-PM) dataset version 3: 35-year climatology of global cloud and radiation properties, Earth Syst Sci Data Earth Syst Sci Data, 12, 41–60.
Wood, R. (2012): Stratocumulus Clouds. Mon. Wea. Rev., 140, 2373–2423, https://doi.org/10.1175/MWR-D-11-00121.1.