Project 3: Hydrological system in Switzerland

MATH-516 Applied Statistics

Linda Mhalla

2025-03-17

Introduction: Local causal discovery

For a target variable T, local causal discovery methods aim at learning its *Markov blanket*, i.e., the direct causes (parents), direct effects (children), and spouses (direct causes of the direct effects) of T, yielding its "neighbourhood" in the causal DAG

$\Rightarrow$ not interested in the causal DAG of the entire system of $(T, \mathbf{X})$ but only on the causal DAG around T

Popular constraint-based methods:

- PC algorithm can be used to extract the Markov blanket of any variable of interest
- HITON-PC (Aliferis et al. 2003): identifies the Markov blanket of a target variable using conditional independence tests

Data: Hydrologically simulated dataset

- Data simulated using the hydrological modelling system PREVAH (PREecipitation-Runoff-EVApotranspiration Hydrotope model); see this paper and this one

$\Rightarrow$ developed to improve the understanding of the spatio-temporal variability of hydrological processes in catchments with complex topography; see this paper

- Dataset consists of 307 catchments in Switzerland for which

- discharge [mm]
- precipitation [mm]
- snowmelt [mm]
- soil moisture [mm]
- temperature [°C]
- (actual) evapotranspiration [mm]
- radiation [W m^{-2}]

were simulated at a daily-resolution from 1981 to 2016

Data: Hydrologically simulated dataset

- Data were calibrated and validated for each catchment by running it with observed meteorological input data (precipitation and temperature) and assessing the simulated discharge values against observed values
- Catchments' flood events are mainly driven either by snowmelt (Alps) or rainfall (Jura, Plateau, and Southern Alps) or by their mixture (Pre-Alps); see this paper

⇒ system of the hydrological and climate variables is spatially dynamic

Source: Part of the data is downloaded from Envidat and the other part is courtesy of Massimiliano Zappa from the WSL

The goal

Aim of the study: We are interested in understanding the hydrological and climatological drivers of extreme discharges during **Spring (March-April-May)**

To make the analysis simpler, for a given catchment, consider a binary variable stating whether the discharge variable exceeds its 90% empirical quantile or not.

$\rightarrow$ stationarity is implictely assumed. Is it reasonable? Maybe we can do better...

The goal of this project is thus to investigate

- *which variables intervene on the propensity of flood events*
- *how far can causal discovery pay off in terms of prediction accuracy*

Tasks: For a given catchment

1. Determine the 90% quantile of the discharge levels (stationary or non-stationary?) and create the binary flood event variable
2. Consider all the daily variables (except discharge at time t) as well as their lagged versions by three days (you can investigate if this lag is informative about floods and consider alternatives if not)

Tasks: For a given catchment

3. Model the effect of these variables on the binary flood event variable (target variable T) by performing classification using logistic regression, where

   - 3.1. you consider all the variables defined in 2.,

and reduce the set features/covariates by

   - 3.2. performing **standard variable selection**, such as, e.g., by using a L_1 penalty (LASSO)
   - 3.3. using the **PC algorithm** for causal discovery and selecting only variables in the Markov blanket around T
   - 3.4 using **a local causal discovery method** (such as implemented in the `R` packages `bnlearn` and `MXM`) to get the Markov blanket around T and keeping only the corresponding variables

4. Compare the prediction performance on a test set (the last two years, say)

Tasks

Do this

a) for a low-elevation and a high-elevation catchments

b) adding a **brief description of the chosen local discovery method** to the report

Assigned catchments

Name of student	Low Catchments	High Catchments
Ahou Samuel	128	218
Boissier Charles Louis Pierre Bogdan	298	136
Bouhadra Kalil Brahim	70	242
Caldarone Alex John	294	81
Carron Léo Jérémy	215	7
Chu Tianle	91	126
Do Alexis	284	201
Ferrera Alessandro	306	146
Frasa Nina	161	274
Garcia Averell Regina	199	56
Gauché Maurice	230	253
Giuli Daniele	116	221
Hengl Stephan	9	28
Khella Georg	207	106
Kriem Zayed	141	232
Loukaidis Andronikos	205	197
Mailänder Lennart Wolfgang	264	85
Olaye Whaboaman Joel Thierry E	127	212
Pettersen Julie Sofie	4	139
Pfander Mila	94	155
Salahshour Arya	115	77
Seixas Aires Joao	150	278
Sigillo' Massara Vincenzo	80	302
Sikking Juliette Joséphine	237	21
Zakharov Daniil	46	273

Assigned catchments