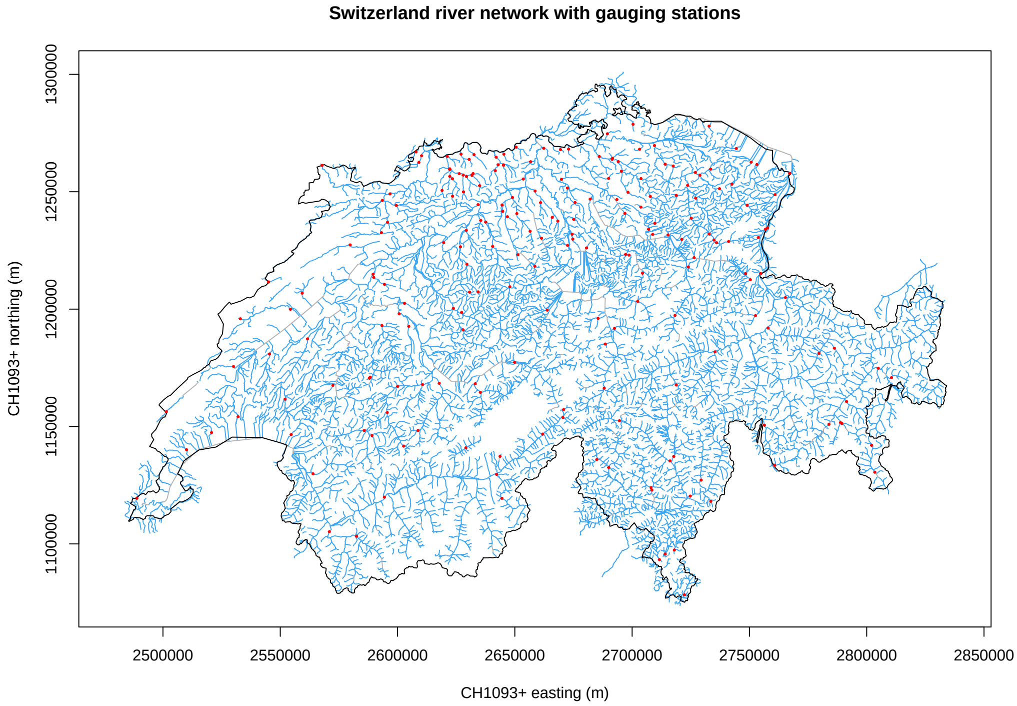
Gaussian Processes with Valid Covariance Defined on River Networks with Application to Discharge Prediction

William Aeberhard and Fernando Pérez-Cruz

January 2022



Project description

The prediction of river discharge (volume of water per time unit, e.g., in m^3/s) at an arbitrary location along a river stream is a challenging task. Traditional hydrological models, generally based on a water balance equation, focus on a single catchment (drainage basin) corresponding to a single location on a river stream, typically where a river gauging station is located. Predictions of water discharge at that particular location can be quite accurate with such models, but generalization to other (arbitrary, possibly ungauged) locations along the river network is often unsatisfactory.

A more empirical approach is to consider multiple river gauging stations, each recording discharge at regular time intervals, and use the directed graph nature of the river network to perform prediction. Taking into account this routing along the river network can be done with the definition of Gaussian processes (GPs) constrained to a network. In traditional geostatistical applications, GPs are defined on $\mathbb{R}^2$ and used to model spatial dependence. The guiding principle here is that phenomena occurring close to each other tend to correlate more positively than phenomena occurring far away. Thus the (stationary, isotropic) covariance of such GPs is typically a function of the Euclidean distance between any two locations. But such covariances do not extend directly to constrained domains like a river network [1]. In particular, the flow of water along the network is directed, distances computed along the river network are more meaningful than Euclidean distances, and flow-disconnected locations may be considered independent. Hence the need to specify non-trivial covariances in space-time that remain valid (positive semi-definite) when constrained to a river network. In addition, combinations of space-time covariances, as in e.g. Asadi, Davison, and Engelke [2], are worth exploring.

At the Swiss Data Science Center, the MACH-Flow project is a collaboration with scientists from the Institute for Atmospheric and Climate Sciences, ETH Zürich, with partners from the Swiss Federal Institute for Forest, Snow and Landscape Research (Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft, WSL). The project deals specifically with discharge prediction of river streams in Switzerland. The available data include 150+ river gauging stations spread over the country with daily average recordings dating back to 1990.

Additional information

- **Difficulty of the project:** moderate to challenging, depending on the student's background and interests
- **What will you learn?**
  - Good understanding of how GPs are used in geostatistics and more broadly in spatio-temporal modeling;
  - Usable skills in terms of packaging R code, writing algorithms, and visualization of predicted spatial fields and maps;
  - Experience in terms of hydrological modeling, complex spatio-temporal dependence specification, and model comparison.
- **Requirements:** Strong programming skills in R; good background in statistics; background in time series and geostatistical models useful but not required.
- **Supervisors:** Dr. William H. Aeberhard (william.aeberhard@sdsc.ethz.ch) and Prof. Fernando Pérez-Cruz (fernando.perezcruz@sdsc.ethz.ch).

Bibliography

[1] J. M. Ver Hoef, E. Peterson, and D. Theobald, "Spatial statistical models that use flow and stream distance," *Environmental and Ecological statistics*, vol. 13, no. 4, pp. 449–464, 2006.

[2] P. Asadi, A. C. Davison, and S. Engelke, "Extremes on river networks," *The Annals of Applied Statistics*, vol. 9, no. 4, pp. 2023–2050, 2015.