

**Advisor(s):**

Prof. Dr.-Ing. Wolfgang Nowak  
apl. Prof. Sergey Oladyshkin  
Dr. rer. nat. Jochen Seidel

**Research group / department:**

Chair of Stochastic Simulation and Safety Research for Hydrosystems (LS<sup>3</sup>)  
Institute for Modelling Hydraulic and Environmental Systems (IWS)  
Stuttgart Centre for Simulation Technology (SC SimTech)

## **End-to-End Hydrological Modelling Using Machine Learning: Leveraging Transformers and Opportunistic Sensor Data for Hydrological Predictions**

**Keywords**

Hydrological modelling, Model development, Deep learning, Transformers, Data-driven modelling

**Introduction / Background**

The majority of hydrological models rely heavily on the principle of mass balance, often represented through Ordinary Differential Equations (ODEs). These models encapsulate the conservation of mass within hydrological systems, ensuring that the inflows, outflows, and storage changes are accurately accounted for. This fundamental principle, coupled with assumed relationships between various components of the hydrological cycle (such as precipitation, evapotranspiration, runoff, and infiltration), forms the core of traditional hydrological modeling approaches. Although powerful, these models often use simplified, linearized assumptions, limiting their capacity to capture the nonlinear complexities inherent in real-world hydrological processes.

Recently, there has also been the branch of neural hydrology, where hydrological models are directly learned from data via machine learning (e.g., LSTM neural networks, [1]). Initially, these models ignored all physical background knowledge and did not necessarily conserve mass, as they used black-box model structures far away from that of rainfall-runoff models. For these reasons, neural hydrology is often criticized by the conventional hydrological community. Alternatively, it is well known that Neural ODEs [2] are capable of representing dynamic systems that are coded in ODEs. They can encapsulate the complex temporal dependencies and dynamics inherent in hydrological systems, offering a promising direction for integrating machine learning with physical modelling. The potential of neural ODE models in hydrology has been discussed in [3].

Nevertheless, both conventional hydrological approaches and neural hydrology typically depend on interpolated input data, e.g., precipitation. Real-world data are often sparse and primarily obtained from meteorological stations and occasionally supplemented by opportunistic sensors. The integration of diverse opportunistic sensor data sources [4], such as private weather stations, further broadens the capability of hydrological models to reliably forecast extreme hydrological events. However, these data sets must initially undergo interpolation before integration into hydrological models.

Current research is developing combining Transformers [5] with Neural ODEs, resulting in an end-to-end modelling framework. Transformers are particularly effective at capturing complex long-range spatiotemporal dependencies within heterogeneous and opportunistic sensor networks. Therefore, such an approach may significantly improve rainfall and runoff predictions.

## References

- [1]. Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, 22(11), 6005-6022. (2018).
- [2]. Chen, R. T., Rubanova, Y., Bettencourt, J., & Duvenaud, D. K. Neural ordinary differential equations. *Advances in Neural Information Processing Systems*, 31. (2018).
- [3]. Höge, M., Scheidegger, A., Baity-Jesi, M., Albert, C., & Fenicia, F. Improving hydrologic models for predictions and process understanding using neural ODEs. *Hydrology and Earth System Sciences*, 26(19), 5085-5102. (2022).
- [4]. Bárdossy, A., Seidel, J., & El Hachem, A. (2021). The use of personal weather station observations to improve precipitation estimation and interpolation. *Hydrology & Earth System Sciences*, 25(2), 583–601
- [5]. Meo, C. et al. (2024). Extreme precipitation nowcasting using transformer-based generative models. *ICLR Workshop: Tackling Climate Change with Machine Learning*.

## Research goals

Our primary goal is to improve the accuracy and prediction reliability of hydrological models through an innovative end-to-end modelling framework. Specifically, we aim to develop and evaluate a novel hybrid model that combines Transformers with Neural ODEs, integrating both data-driven machine learning techniques and established physical hydrological principles. This novel hybrid approach should be then rigorously validated and compared against conventional hydrological modelling methods that explicitly use interpolation to pre-process precipitation data. In addition, the research should explore the integration of opportunistic sensor data to improve forecast performance, particularly during extreme events, addressing limitations inherent in traditional meteorological monitoring networks.

## Methods to be used

The research will focus on the integration of Transformer architectures [5] and Neural ODEs [2], creating a cohesive end-to-end hydrological model. Transformers will be utilized to pre-process and analyse the raw input data, capturing essential spatial and temporal relationships in rainfall data. Neural ODEs (or initially more traditional hydrological models like HBV) will then enforce physically consistent mass balances to translate the pre-processed rainfall into runoff reliably. These methodologies will be employed to model dynamic systems, where the dependencies between state variables and their temporal evolution will be learned. In the Neural ODEs, state variables will be defined similarly to those in existing conceptual hydrological rainfall-runoff models, such as HBV. However, the storage terms and fluxes, as functions of the current model states, will be learned using Neural ODEs, which inherently follows the principle of mass balance, but have more flexibility in matching the complex rainfall-runoff relation. To validate and test the proposed hybrid framework, selected case studies will be implemented and compared against suitable baseline models, such as standard HBV with traditional rainfall interpolation. The selection of methods and case studies will be tailored to identify the most effective combination for addressing the challenges posed by the proposed research. This

methodological flexibility is crucial for optimizing the model's performance and ensuring its applicability to real-world hydrological systems.

### **Prerequisites**

- MSc in hydrology, environmental sciences, hydrogeology, water management (or similar) or in data sciences, statistics, applied mathematics.
- Skills in programming (e.g. python, matlab, julia)
- Skills at scientific writing and presentation
- Ability to work independently and in a team
- Willingness to learn new concepts and methods
- Experience (e.g., coursework, thesis work) in hydrological modelling or in machine learning
- Willingness to contribute to the goals and culture of the research group

### **Research Environment**

This research will be embedded into the Chair of Stochastic Simulation and Safety Research for Hydrosystems (LS<sup>3</sup>) at the IWS, Faculty of Civil and Environmental Engineering. Depending on qualification of the candidate, a formal association of the project to the SC SimTech is possible.