



THE IMPACT OF LAND COVER DATA ON RAINFALL-RUNOFF PREDICTION USING AN ENTITY-AWARE-LSTM

EGU21 - Session HS3.4: Deep learning in hydrological science

Sebastian Drost ¹, Fabian Netzel ², Andreas Wytzisk-Ahrens ¹, and Christoph Mudersbach ²

¹ Department of Geodesy, Bochum University of Applied Sciences, Bochum, Germany (sebastian.drost@hs-bochum.de)

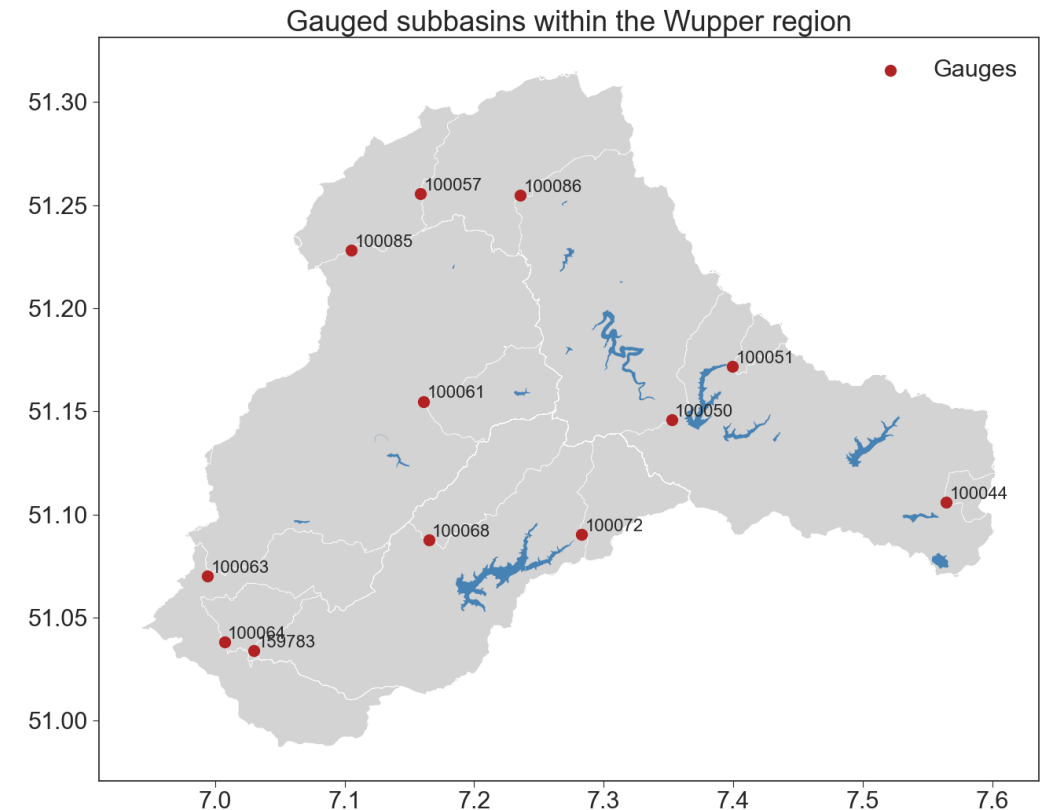
² Department of Civil and Environmental Engineering, Hydraulic Engineering and Hydromechanics Section (LWH), Bochum University of Applied Sciences, Bochum, Germany (fabian.netzel@hs-bochum.de)

Research Topics

- Studying the effect of land cover changes data on rainfall-runoff prediction using an Entity-Aware-LSTM (EA-LSTM)
 - Suitability of EA-LSTM for assessing the effect of land-cover changes in modeling Rainfall-Runoff models
- Overcoming the regional modelling problem
 - Optimizing the applicability of models trained outside of the selected area of study for ungauged basins using land-cover data
 - Addressing catchment regions with a small number of gauged basis
- Comparison of EA-LSTM (land-cover data included) and basic LSTM (land-cover data excluded)

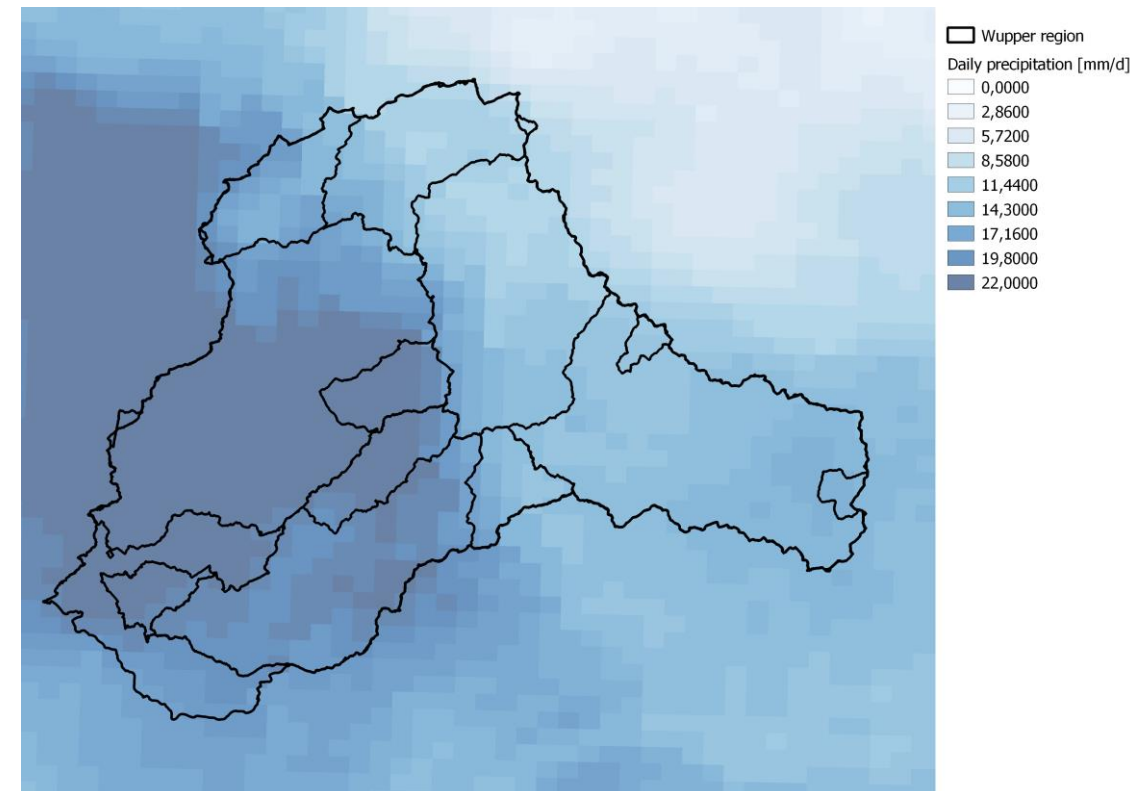
Area of Study

- Wupper region in Germany, part of Rhine-catchment
- Catchment size: 814 km², 21 subcatchments > 10 km²
- Total length of streams: 2.300 km, length of river Wupper: 115 km
- Mean annual precipitation: 1.200 mm
- 14 dams for process and drinking water
- Altitude: 36 m NHN to 483 m NHN
- Land use:
 - Urban area: 27 %, Grassland: 29 %, Forest: 32 %, Fields and crops: 9 %, other: 3 %



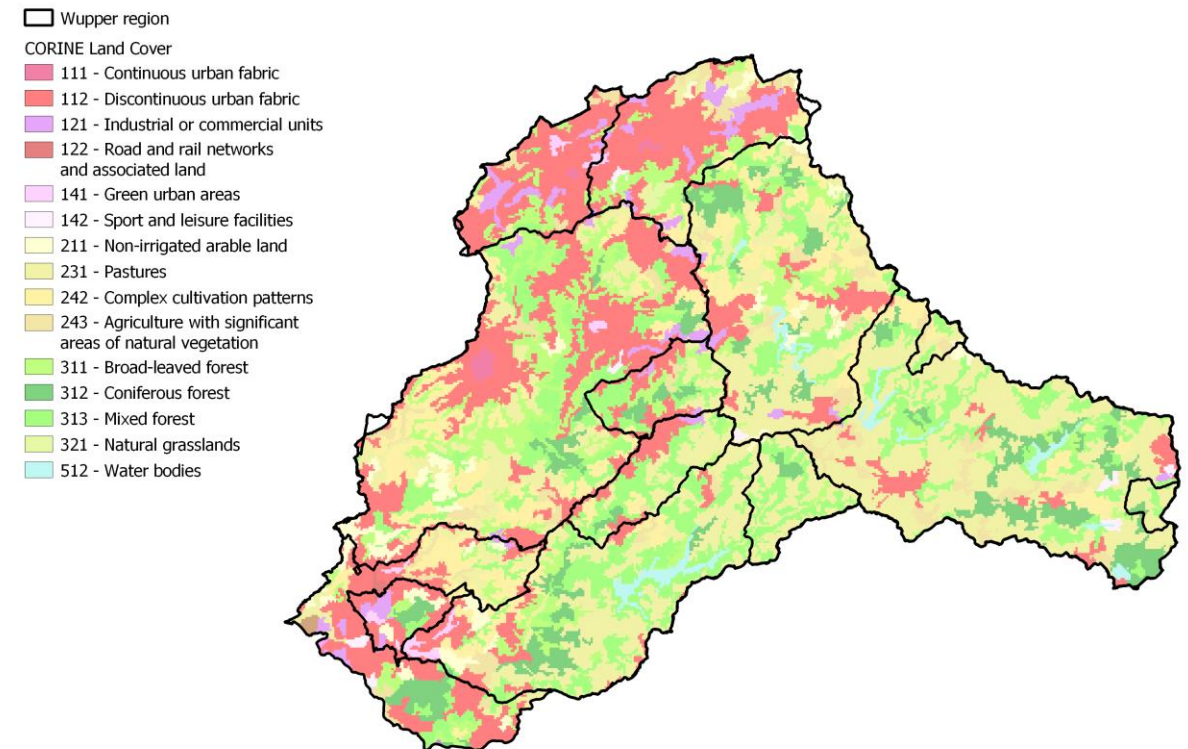
Datasets

- Catchment specific precipitation timeseries derived from REGNIE
 - REGNIE: Regionalized precipitation amount (**REG**ionalisierte **NIE**derschlagshöhe) (Rauthe et al., 2013)
 - Gridded daily precipitation since 1931 with a spatial resolution of 1 km x 1km
 - Method: Regionalized interpolation of stationary observations
- Preprocessing
 - Calculation of the average precipitation for all raster cells within a certain catchment area of the Wupper region



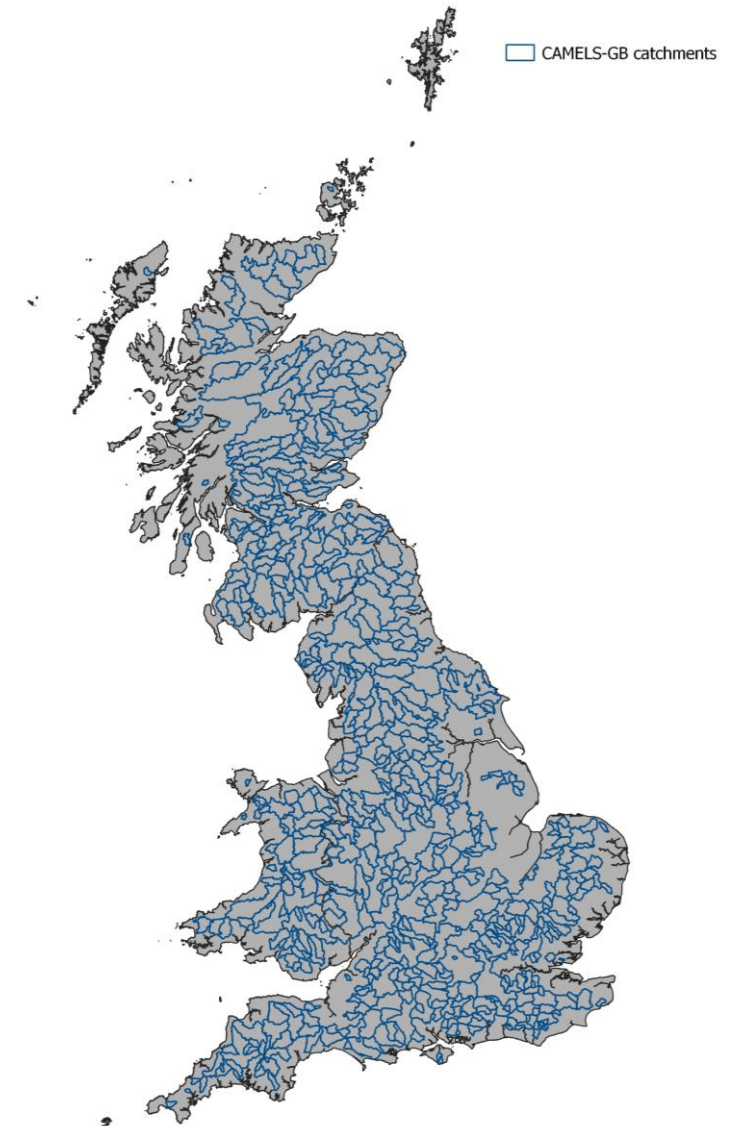
Datasets

- Static catchment attributes derived from CORINE Land Cover
 - Europe-wide timeseries of land cover in 44 classes as well as land cover changes for different time periods (Büttner, 2014)
- Preprocessing
 - Determination of the percentage proportion of each land cover class within single catchments
 - Relative amount of 14 land cover classes from year 2000 for all catchments of the Wupper region and the CAMEL-GB dataset



Datasets

- CAMELS-GB: large-sample catchment hydrology dataset for Great (Coxon et al., 2020)
 - Daily hydrological and meteorological timeseries data from 1970-2015
 - E.g., discharge, precipitation, temperature
 - Attributes of 671 catchments
 - E.g., area, mean elevation, land cover and soil attributes
 - Used as large sample dataset to learn regional discharge similarities from land cover data
- Preprocessing
 - Determination of the percentage proportion of land cover classes from CORINE Land Cover 2000

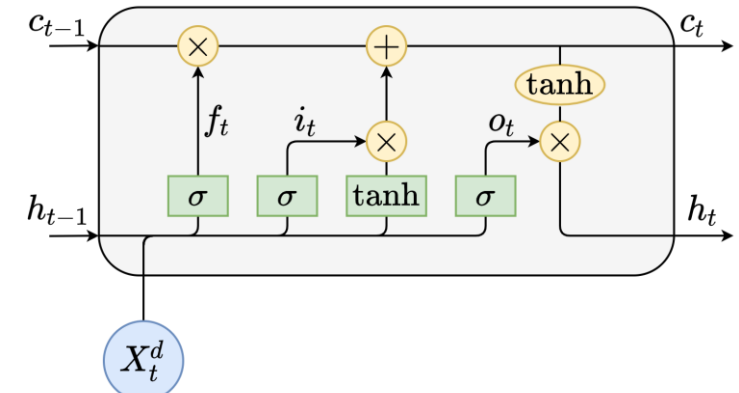


Methods

Models

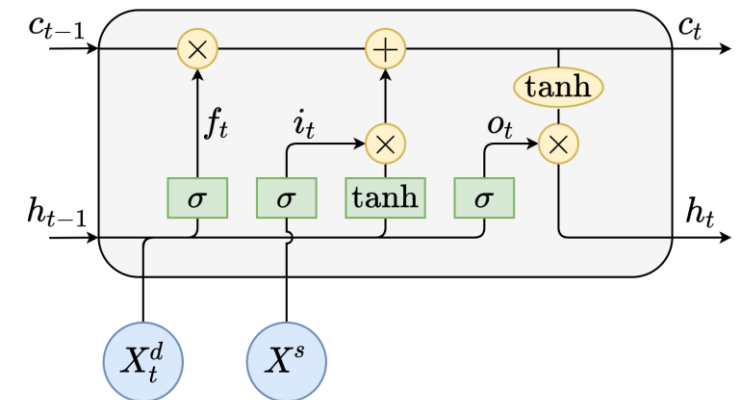
- **LSTM** (Long short-term memory network)
 - Dynamic inputs (X_t^d): Daily precipitation [mm/d] timeseries of the last 365 days
 - Output: Discharge [mm/d] of the current day
- **Entity-Aware-LSTM** (EA-LSTM) as proposed by Kratzert et al. (2019)
 - Dynamic inputs (X_t^d): Daily precipitation [mm/d] timeseries of the last 365 days
 - Static inputs (X_t^a): percentage proportion of each land cover class within a subbasin
 - Output: Discharge [mm/d] of the current day

Classic LSTM



Conceptual visualization of the classic LSTM adapted from Olah (2015) and Kratzert et al. (2019)

Entity Aware LSTM



Conceptual visualization of an Entity-Aware-LSTM adapted from Olah (2015) and Kratzert et al. (2019)

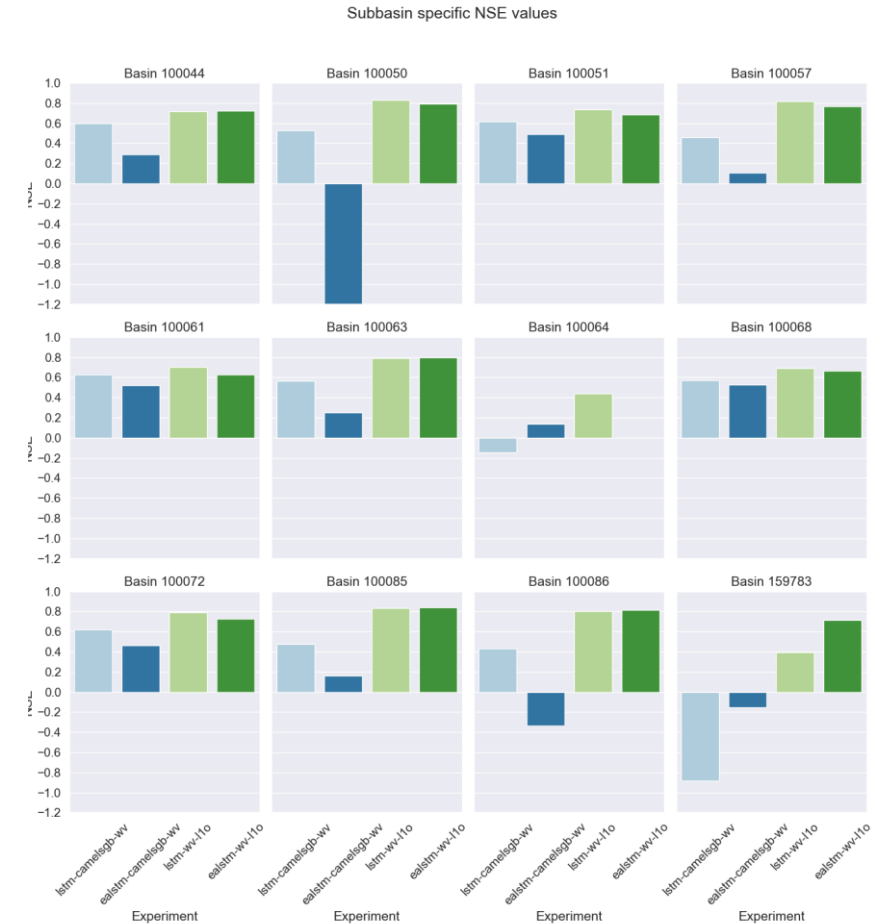
Methods

Experimental Setup

- We set up 4 different experiments to analyze if regional discharge behavior can be learned from land cover data used as static catchment attributes
 - 1./2. Training on all 671 CAMELS-GB catchments and testing on 12 Wupper region catchments using classic LSTM/EA-LSTM
 - 3./4. Training and testing on Wupper region catchments applying leave-one-out strategy using classic LSTM/EA-LSTM
- Training period for CAMELS-GB catchments: from 1 October 1970 to 31 December 2010
- Training and testing period for Wupper region catchments: from 1 January 1990 or 1 December 2010 to 31 December 2020
- All models are evaluated using the Nash–Sutcliffe efficiency measure

Results

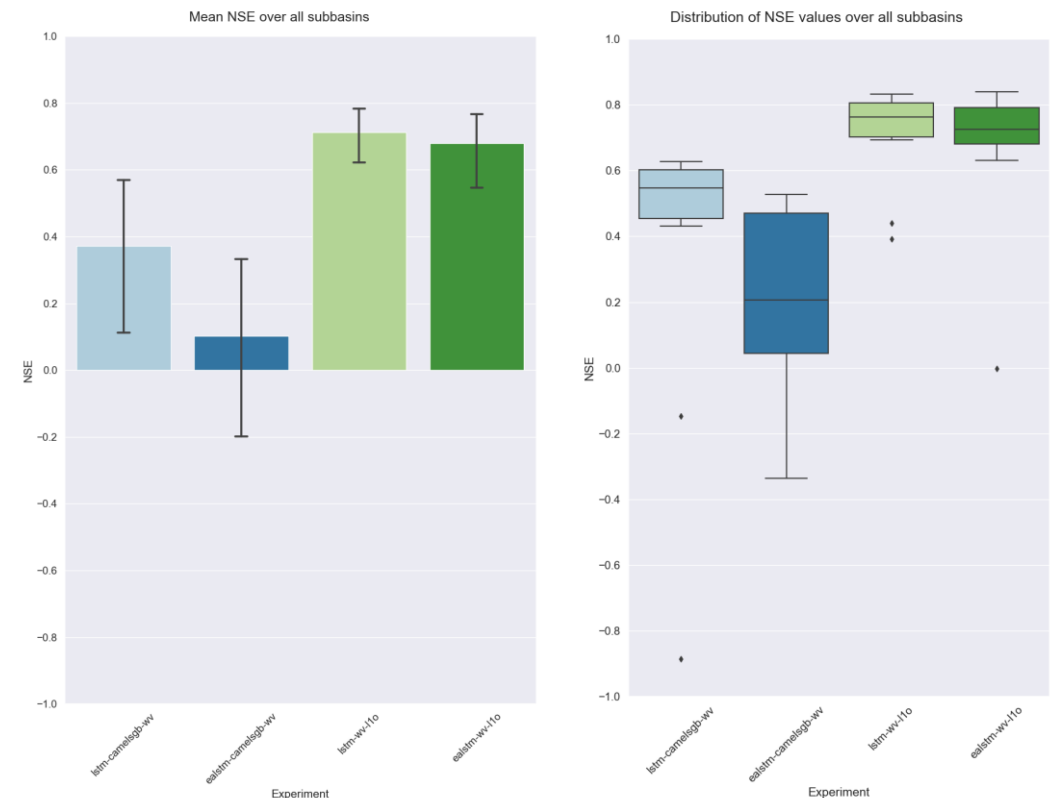
Specific NSE values for each subbasin



Results

NSE summary statistics for all subbasins

- Models trained on basins of Wupper region (leave-one-out experiments) outperforms CAMELS-GB models
- EA-LSTM and LSTM show no significant differences in prediction performance
 - No prediction improvements using static land cover data
- EA-LSTM trained on CAMELS-GB performs worst and shows a high variance
 - Regional similarities could not be learned from land cover data using catchments outside the area of study



Results

- Catchments with discharge regulated by dams show a higher variation in yearly NSE metrics
 - E.g., basins 100050, 100064, 159783
- The variation shows a dependency on the distance of gauges to dams in a certain catchment area



Conclusion

- Regional discharge behavior could not be learned from catchments outside (CAMELS-GB in this study) the area of study
 - EA-LSTM models seems to learn inappropriate inter-regional relationships between discharge behavior and land cover
- Land cover data as static catchment attributes did not improve discharge predictions
 - Wupper region dataset seems to be too small to learn intra-regional discharge behavior
- Yearly NSE is influenced by the location of gauges in connection with dams (the farther away, the better the NSE)

Outlook

- To check if EA-LSTM can learn intra-regional discharge behavior, training and testing should be done solely with the large CAMELS-GB datasets using land cover data as static catchment
- The correlation of varying NSE and dam regulated catchments should be analyzed using time information about regulations of flow rates
- To improve regional modeling, we aim to consider the spatial distribution of dynamic and static catchment attributes
 - In contrast to lumped models, input features are varying in both time and space (in accordance to distributed models)
 - E.g., grid-based datasets (such as REGNIE and CORINE land cover within this study) could be directly used as inputs instead of using aggregated forcings

Data and code Availability

- **Paper code**

<https://github.com/SebaDro/egu21-rrp-landcover-lstm>

- Provides code and datasets to reproduce results and graphics related to the model evaluations via Jupyter Notebooks

- **Preprocessing:**

<https://github.com/SebaDro/rainfall-runoff-preprocessing>

- Hosts R-based preprocessing routines that were used for calculating daily precipitation means from REGNIE as well as for calculating CORINE land cover proportions for Wupper region and CAMELS-GB catchments

- **DWD REGNIE:**

https://opendata.dwd.de/climate_environment/CDC/grids_germany/daily/regnie/

- **CORINE Land Cover 2000:**

<https://land.copernicus.eu/pan-european/corine-land-cover/clc-2000>

- **CAMELS-GB**

<https://catalogue.ceh.ac.uk/documents/8344e4f3-d2ea-44f5-8afa-86d2987543a9>

References

Büttner, G. (2014). CORINE Land Cover and Land Cover Change Products. In: Manakos & M. Braun (Hrsg.), *Land Use and Land Cover Mapping in Europe* (Bd. 18, S. 55–74). Springer Netherlands. https://doi.org/10.1007/978-94-007-7969-3_5

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., & Woods, R. (2020). CAMELS-GB: Hydrometeorological time series and landscape attributes for 671 catchments in Great Britain. *Earth System Science Data*, 12(4), 2459–2483. <https://doi.org/10.5194/essd-12-2459-2020>

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., & Nearing, G. (2019). Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets. *Hydrology and Earth System Sciences*, 23(12), 5089–5110. <https://doi.org/10.5194/hess-23-5089-2019>

Olah, C. (2015). *Understanding LSTM Networks*. <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A & Gratzki, A (2013): A Central European precipitation climatology – Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS), *Meteorologische Zeitschrift*, Vol 22, No 3, 235–256. <https://doi.org/10.1127/0941-2948/2013/0436>



ANY QUESTIONS OR SUGGESTIONS?

Sebastian Drost

sebastian.drost@hs-bochum.de