

KI-HopE-De – LSTM-Operationalisierung

Autorenschaft: Landesamt für Umwelt Rheinland-Pfalz (LfU RLP)

Das vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) geförderte Forschungsprojekt KI-HopE-De beschäftigt sich mit der Entwicklung eines universellen überregionalen LSTM-Modells für die operationelle Hochwasservorhersage in Deutschland. Der Fokus liegt dabei auf der Verbesserung der Extrapolationsfähigkeit und Qualität für den Hochwasserfall sowie der Kurzfrist-Abflussvorhersage in kleinen, teilweise unbeobachteten Gebieten.

Im ersten Arbeitspaket des Projekts wird der Aufbau eines standardisierten deutschlandweiten Datensatz in stündlicher Auflösung als Erweiterung des bestehenden CAMELS-DE Datensatzes (Dolich et al. 2024) – CAMELS-DE-1h –realisiert. Die CAMELS-Datensätze (Catchment Attributes and MEteorology for Large-sample Studies) nach Addor et al. 2017 haben sich bereits im Bereich der hydrologischen Modellierung und der Verwendung maschineller Lernmethoden etabliert. Die Datensätze umfassen Einzugsgebietsmerkmale, hydrologische Pegelzeitreihen sowie räumlich aggregierte meteorologische Zeitreihen (Reanalyse-Daten). Ergänzend dazu werden auch die meteorologischen Vorhersagen des ICON-D2 Modells des Deutschen Wetterdienstes integriert.

Im Zentrum der Modellentwicklung steht die Multi-Frequency-LSTM- (MF-LSTM) Architektur (Acuña Espinoza et al. 2024), die auch im Regionalmodell für Luxemburg zum Einsatz kommt und in Beitrag 4 der vorliegenden Ausgabe des Au fil de l'eau näher beschrieben wird. Dieser Modelltyp ermöglicht ein effizientes Training eines deutschlandweiten LSTM-Modells, das derzeit 1.116 Pegeleinzugsgebiete umfasst. Ziel ist es, die MF-LSTM-Architektur mit einem Mixture-Density-Network (MDN) zu kombinieren, um als Modellausgabe eine Wahrscheinlichkeitsverteilung zu erzeugen und damit die natürliche Variabilität des hydrologischen Systems abzubilden. Ergänzt wird dieser Ansatz durch die Verwendung von Niederschlags-Ensembles als Modelleingang. Auf diese Weise können sowohl die Modellunsicherheit als auch die Vorhersageunsicherheit explizit berücksichtigt werden.

Parallel zur Modellentwicklungen erfolgen am LfU RLP zusätzliche Modelltests, ein regionales Fine-Tuning sowie die Implementierung des LSTM-Modells in das operationelle Hochwasservorhersagesystem. Ziel ist es, das deutschlandweite „Basis-LSTM“ auf die für Rheinland-Pfalz relevanten Einzugsgebiete und die operationellen Datenflüsse anzupassen. Das LSTM wird an die Ausdehnung des operationellen LARSIM-Modells angepasst und umfasst damit neben Rheinland-Pfalz auch das gesamte Mosel-Einzugsgebiet. Hierfür ist die Erstellung eines grenzübergreifenden Datensatzes (vgl. Abbildung 4) für das Modelltraining erforderlich – dieser wird analog zu CAMELS-DE-1h aufgebaut.

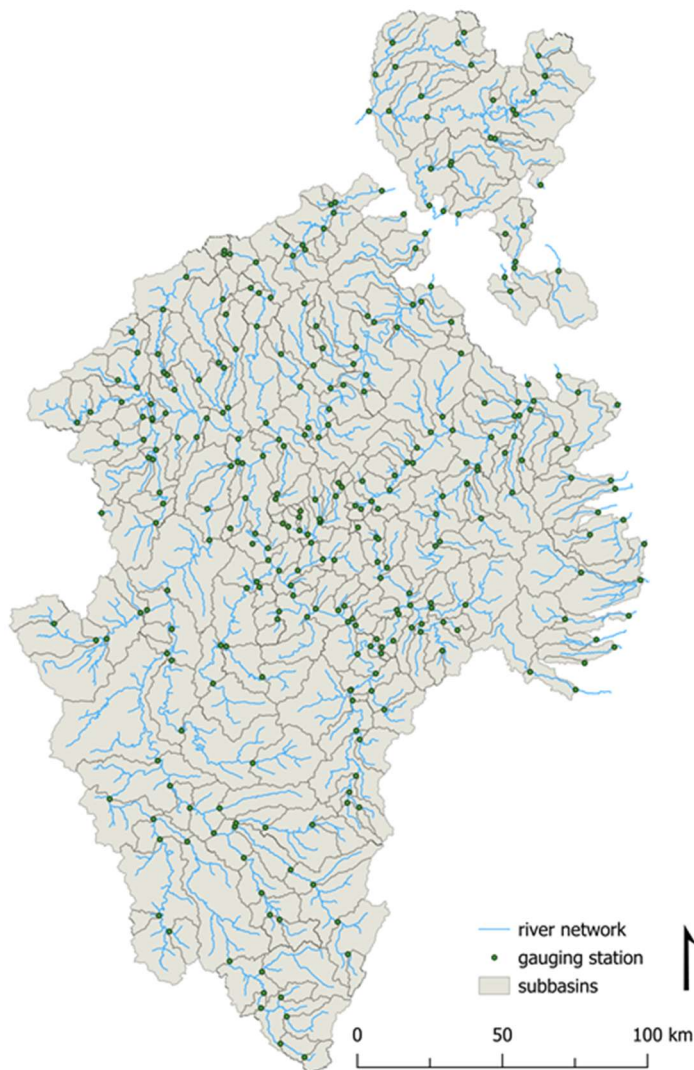


Abbildung 4: Pegeleinzugsgebiete des grenzübergreifenden hydrometeorologischen Datensatzes für das Rheinland-Pfalz- und Mosel-LSTM-Regionalmodell.

Für den operationellen Einsatz muss das Modell für den „Hindcast“ auf hydrologische und meteorologische Echtzeit-Daten zugreifen können, die am LfU bereits für LARSIM aufbereitet werden. Dabei handelt es sich insbesondere um interpolierte Stationsdaten (InterMet) sowie die Niederschlagsradarprodukte RADOLAN/ANTILOPE. Aus der langen LARSIM-Modellhistorie steht am LfU ein umfangreicher, langjähriger Datenbestand dieser Produkte zur Verfügung und kann, ergänzend zu den DWD-Reanalyse-Produkten RADKLIM und HOSTRADA, in das Modelltraining einfließen.

In ersten Tests wurden verschiedene MF-LSTM-Modelltypen und Experiment-Settings für 100 Einzugsgebiete in Rheinland-Pfalz im Vorhersagemodus trainiert und validiert. Dabei wurden unter anderem unterschiedliche zeitliche Auflösungen (stündlich, täglich, wöchentlich), die Verwendung einer autoregressiven Verarbeitung der Abflussdaten sowie die Ergänzung zusätzlicher dynamischer Attribute (z.B. InterMet-Variablen) untersucht. Der Schwerpunkt der Auswertung liegt auf der Abbildungsqualität während Hochwasserphasen. Zum aktuellen Zeitpunkt lässt sich jedoch keine generelle Überlegenheit eines bestimmten Modelltyps für alle betrachteten Einzugsgebiete und Ereignisse feststellen. So zeigen in einigen Fällen nicht-autoregressive Architekturen

bessere Ergebnisse, während in anderen Fällen autoregressive Ansätze überlegen sind. Der Einsatz zusätzlicher meteorologischer Eingangsdaten, insbesondere der interpolierten Stationsniederschläge, führt hingegen überwiegend zu einer Verbesserung der Vorhersagegüte.

In den nächsten Arbeitsschritten sollen die vielversprechendsten Modellkonfigurationen weiter verfeinert und auf das grenzübergreifende Untersuchungsgebiet angewendet werden. Zudem ist eine vertiefte Ereignisanalyse extremer Hochwasser vorgesehen, um die Robustheit der Modelle unter Extrembedingungen besser bewerten zu können. Perspektivisch wird das LSTM schrittweise in den operationellen Vorhersagebetrieb integriert und parallel zu LARSIM getestet.

Literatur

- Acuña Espinoza, E., Kratzert, F., Klotz, D., Gauch, M., Álvarez Chaves, M., Loritz, R., and Ehret, U. (2024). Technical note: An approach for handling multiple temporal frequencies with different input dimensions using a single LSTM cell, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2024-3355>
- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P. (2024). The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, <https://doi.org/10.5194/hess-21-5293-2017>
- Dolich, A., Espinoza, E. A., Ebeling, P., Guse, B., Götze, J., Hassler, S., Hauße, C., Kiesel, J., Heidebüchel, I., Mälicke, M., Müller-Thomy, H., Stölzle, M., Tarasova, L., & Loritz, R. (2024). CAMELS-DE: hydrometeorological time series and attributes for 1582 catchments in Germany (1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13837553>

KI-HopE-De - Operationalisation du LSTM

Auteur: Landesamt für Umwelt Rheinland-Pfalz (LfU RLP)

Das vom Bundesministerium für Forschung, Technologie und Raumfahrt (BMFTR) geförderte Forschungsprojekt KI-HopE-De beschäftigt sich mit der Entwicklung eines universellen überregionalen LSTM-Modells für die operationelle Hochwasservorhersage in Deutschland. Der Fokus liegt dabei auf der Verbesserung der Extrapolationsfähigkeit und Qualität für den Hochwasserfall sowie der Kurzfrist-Abflussvorhersage in kleinen, teilweise unbeobachteten Gebieten.

Im ersten Arbeitspaket des Projekts wird der Aufbau eines standardisierten deutschlandweiten Datensatz in stündlicher Auflösung als Erweiterung des bestehenden CAMELS-DE Datensatzes (Dolich et al. 2024) – CAMELS-DE-1h –realisiert. Die CAMELS-Datensätze (Catchment Attributes and MEteorology for Large-sample Studies) nach Addor et al. 2017 haben sich bereits im Bereich der hydrologischen Modellierung und der Verwendung maschineller Lernmethoden etabliert. Die Datensätze umfassen Einzugsgebietsmerkmale, hydrologische Pegelzeitreihen sowie räumlich aggregierte meteorologische Zeitreihen (Reanalyse-Daten). Ergänzend dazu werden auch die meteorologischen Vorhersagen des ICON-D2 Modells des Deutschen Wetterdienstes integriert.

Im Zentrum der Modellentwicklung steht die Multi-Frequency-LSTM- (MF-LSTM) Architektur (Acuña Espinoza et al. 2024), die auch im Regionalmodell für Luxemburg zum Einsatz kommt und in Beitrag 4 der vorliegenden Ausgabe des Au fil de l'eau näher beschrieben wird. Dieser Modelltyp ermöglicht ein effizientes Training eines deutschlandweiten LSTM-Modells, das derzeit 1.116 Pegeleinzugsgebiete umfasst. Ziel ist es, die MF-LSTM-Architektur mit einem Mixture-Density-Network (MDN) zu kombinieren, um als Modellausgabe eine Wahrscheinlichkeitsverteilung zu erzeugen und damit die natürliche Variabilität des hydrologischen Systems abzubilden. Ergänzt wird dieser Ansatz durch die Verwendung von Niederschlags-Ensembles als Modelleingang. Auf diese Weise können sowohl die Modellunsicherheit als auch die Vorhersageunsicherheit explizit berücksichtigt werden.

Parallel zur Modellentwicklungen erfolgen am LfU RLP zusätzliche Modelltests, ein regionales Fine-Tuning sowie die Implementierung des LSTM-Modells in das operationelle Hochwasservorhersagesystem. Ziel ist es, das deutschlandweite „Basis-LSTM“ auf die für Rheinland-Pfalz relevanten Einzugsgebiete und die operationellen Datenflüsse anzupassen. Das LSTM wird an die Ausdehnung des operationellen LARSIM-Modells angepasst und umfasst damit neben Rheinland-Pfalz auch das gesamte Mosel-Einzugsgebiet. Hierfür ist die Erstellung eines grenzübergreifenden Datensatzes (vgl. Abbildung 4) für das Modelltraining erforderlich – dieser wird analog zu CAMELS-DE-1h aufgebaut.

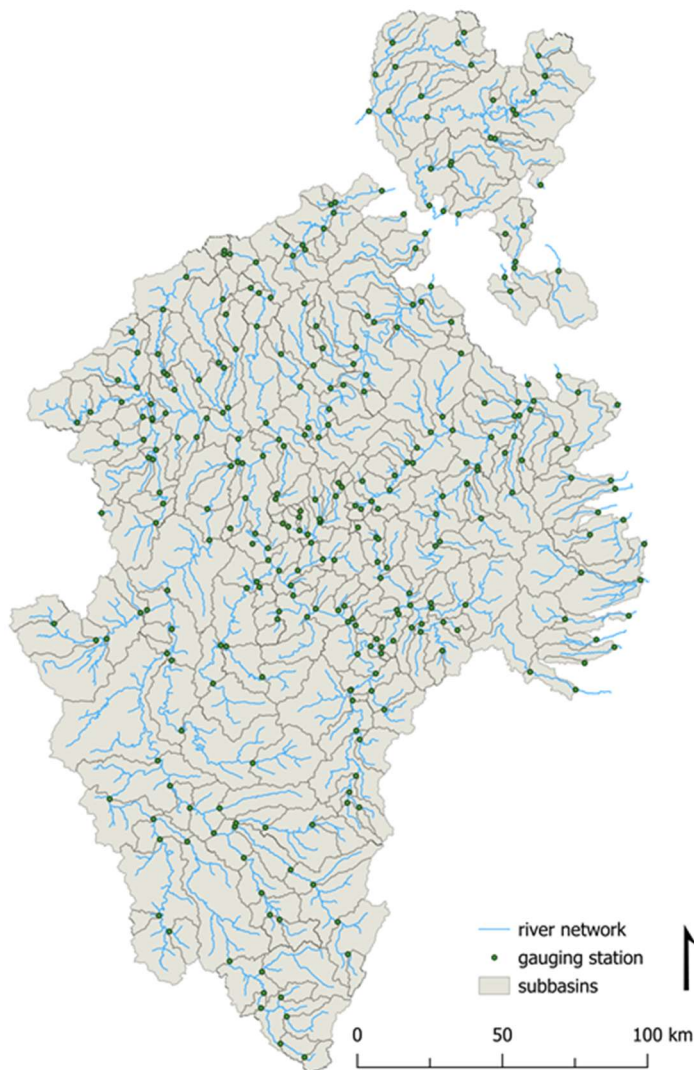


Figure 4 : Bassins versants du jeu de données hydrométéorologiques transfrontalier pour le modèle LSTM régional (Rhénanie-Palatinat et Moselle).

Pour une utilisation opérationnelle aux fins du « hindcast », le modèle doit pouvoir accéder à des données hydrologiques et météorologiques en temps réel, qui sont déjà traitées pour LARSIM au sein du LfU. Il s'agit notamment de données de station interpolées (InterMet) ainsi que des produits radars de précipitations RADOLAN/ANTILOPE. Grâce à la longue histoire des modèles LARSIM, le LfU dispose d'une vaste base de données de longue date de ces produits et peut, en complément des produits de réanalyse du DWD RADKLIM et HOSTRADA, les intégrer dans l'entraînement des modèles.

Lors des premiers tests, différents types de modèles MF-LSTM et paramètres expérimentaux ont été entraînés et validés en mode prévisionnel pour 100 bassins versants situés en Rhénanie-Palatinat. Différentes résolutions temporelles (horaire, quotidienne, hebdomadaire), l'utilisation d'un traitement autorégressif des données de débit et l'ajout d'attributs dynamiques supplémentaires (par ex. variables InterMet) ont notamment été étudiés. L'évaluation se concentre sur la qualité de l'imagerie pendant les phases de crue. À l'heure actuelle, il n'est toutefois pas possible de conclure à la supériorité générale d'un type de modèle particulier pour l'ensemble des bassins versants

et des événements considérés. Ainsi, dans certains cas, les architectures non autorégressives donnent de meilleurs résultats, alors que dans d'autres cas, les approches autorégressives sont supérieures. En revanche, l'utilisation de données d'entrée météorologiques supplémentaires, en particulier les précipitations interpolées des stations, améliore largement la qualité des prévisions.

Dans les prochaines étapes de travail, les configurations de modèles les plus prometteuses seront encore affinées et appliquées à la zone d'étude transfrontalière. En outre, une analyse approfondie des événements de crues extrêmes est prévue afin de mieux évaluer la robustesse des modèles dans des conditions extrêmes. En perspective, le LSTM sera progressivement intégré dans les opérations de prévision opérationnelles et testé parallèlement à LARSIM.

Littérature

- Acuña Espinoza, E., Kratzert, F., Klotz, D., Gauch, M., Álvarez Chaves, M., Loritz, R., and Ehret, U. (2024). Note technique : An approach for handling multiple temporal frequencies with different input dimensions using a single LSTM cell, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2024-3355>
- Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P. (2024). The CAMELS data set : catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci, 21, 5293-5313, <https://doi.org/10.5194/hess-21-5293-2017>
- Dolich, A., Espinoza, E. A., Ebeling, P., Guse, B., Götze, J., Hassler, S., Hauße, C., Kiesel, J., Heidebüchel, I., Mälicke, M., Müller-Thomy, H., Stölzle, M., Tarasova, L., & Loritz, R. (2024). CAMELS-FR : hydrometeorological time series and attributes for 1582 catchments in Germany (1.0.0) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.13837553>