

LSTM-Regionalmodell für Luxemburg

Autorenschaft: Administration de la gestion de l'eau Luxembourg

Das Ingenieurbüro HYDRON hat im Auftrag der AGE Luxembourg ein sogenanntes LSTM- (Long Short Term Memory) Regionalmodell für Luxemburg aufgebaut. Nach einer ersten Machbarkeitsstudie für einen einzigen Pegel sind im Regionalmodell nun Simulationen und Vorhersagen des Abflusses für insgesamt 27 Pegel in Luxemburg möglich (vgl. Abbildung 1).



Abbildung 1: Karte mit den 27 Pegeln in Luxemburg, für die im LSTM-Regionalmodell Simulationen und Vorhersagen des Abflusses möglich sind.

Das Modell ist mit validierten Abflussdaten und den meteorologischen Parametern (Niederschlag, Lufttemperatur, Globalstrahlung, relative Luftfeuchte, Luftdruck und

Windgeschwindigkeit) im Zeitraum 2002-2022 aufgebaut worden. Der Trainingszeitraum umfasst 15 Jahre (2003-2017), der Testzeitraum fünf Jahre (2018-2022), also auch das extreme Hochwasser vom Juli 2021.

Bei der Modellerstellung sind aus Überlegungen zur späteren operationellen Nutzung des LSTM als Eingangsdaten die gleichen Eingangsgrößen wie für LARSIM gewählt worden. Dies ermöglicht Synergien im operationellen Betrieb, sollte das LSTM parallel zu LARSIM genutzt werden. Diese dynamischen, meteorologischen Eingangsdaten werden in Form von in Zeitreihen umgewandelten, räumlich interpolierten Einzugsgebietsmittelwerten pro Pegel und Parameter ins Modell eingelesen. Zusätzlich dazu werden der gemessene Abfluss bis zum VZP und verschiedene statische Gebietseigenschaften wie Einzugsgebietsgröße oder die mittlere Geländehöhe (zur Unterscheidung der Pegel in den unterschiedlichen Einzugsgebieten) eingelesen.

Beim Modelltraining handelt es sich noch um Simulationen mit gemessener Meteorologie. Mit dem Modell sollen nun ebenfalls stündliche Abflussvorhersagen mit einer Vorhersagetiefe von bis zu 45 Stunden erstellt werden, vorerst nur mit den deterministischen Kurzfrist Vorhersagen des DWD ICON-D2 Modells (48 Stunden Vorhersagetiefe alle drei Stunden aktualisiert), später sind auch Ensemble-Vorhersagen weiterer Produkte möglich.

Dafür wird erstmalig ein neuer Ansatz eines MF-LSTM (Multi Frequency-LSTM) verwendet. Diese spezielle Erweiterung des LSTM ermöglicht eine operationelle Abflussvorhersage mit Stundenwerten, als Input für die meisten Tage des zurückliegenden Jahres kann dabei allerdings, anders als bisher, auf Tageswerte zurückgegriffen werden, was den Rechenaufwand erheblich verringert.

Beim Modelltraining und der Modelloptimierung werden für einen Teil der Trainingsdaten Abflussvorhersagen erstellt. Die berechneten Differenzen zwischen Messungen und Vorhersagen werden an das LSTM übermittelt und die internen Parameter des LSTM (Gewichtungen) angepasst/optimiert. Ist dies oft genug durchgeführt worden, werden Vorhersagetests ohne Optimierung für den Testdatensatz durchgeführt und dieser Prozess so oft wiederholt, bis keine weitere Verbesserung zu erwarten ist.

Als weiterer Datensatz zur Verbesserung der Güte und Extrapolationsfähigkeit des Modells wurde ein Teil des CAMLES-DE-Datensatzes verwendet, allerdings in geringerer Gewichtung als die Daten der luxemburgischen Pegel.

Die Auswertung der Simulationen erfolgt anhand zusammengesetzter Ganglinien für die einzelnen Vorhersagetiefen. Bereits ohne die gemessenen Abflüsse zeigen sich sehr gute Ergebnisse mit einem NSE (Median) von 0,89, auch für Pegel mit denen nicht trainiert wurde. Unter Einbeziehung der gemessenen Abflüsse und des CAMLES-DE-Datensatzes steigt die Modellgüte weiter an auf einen NSE (Median) von 0,96 in allen Vorhersagetiefen. Ohne den CAMLES-DE-Datensatz verschlechtern sich vor allem Pegel mit nur wenigen verfügbaren Daten im Trainingszeitraum.

Es zeigen sich allerdings noch Unterschiede in der Simulationsgüte zwischen einzelnen Ereignissen, teilweise sogar bei der Verwendung gemessener Abflüsse. Die Extrapolationsfähigkeit funktioniert nur eingeschränkt, das extreme Hochwasser vom Juli 2021 wird nicht optimal simuliert, falls kein vergleichbares Ereignis in den Trainingsdaten vorliegt (2021 ist nur im Testzeitraum enthalten, nicht im Trainingszeitraum). Häufig ist dieses Ereignis allerdings das mit Abstand höchste an den Pegeln im Gesamtzeitraum. Die Verwendung des CAMELS-DE-Datensatzes kann die Extrapolationsfähigkeit teilweise verbessern.

In einem nächsten Schritt soll nun ein Konzept für die operationelle Umsetzung der LSTM-Vorhersagen erstellt werden. Diese sollen möglichst Synergien zum operationellen Betrieb von LARSIM und den bereits vorhandenen Eingangsdaten nutzen. Herausforderungen und Ansatzpunkte für mögliche Weiterentwicklungen bleiben allerdings unterschiedliche Datenverfügbarkeit und Datenlücken an einigen Pegeln in bestimmten Jahren und der Umgang mit fehlenden Daten im operationellen Betrieb sowie die Abbildung anthropogener Einflüsse, wie beispielsweise von Talsperren, und die Extrapolationsfähigkeit extremer Ereignisse.

Modèle régional LSTM pour le Luxembourg

Auteur: Administration de la gestion de l'eau Luxembourg

À la demande de l'AGE Luxembourg, le bureau d'études HYDRON a développé un modèle régional pour le Luxembourg, dit LSTM (Long Short Term Memory). Après une première étude de faisabilité pour une seule station, le modèle régional permet désormais de réaliser des simulations et des prévisions de débit pour 27 stations au total au Luxembourg (voir la figure 1).



Figure 1 : Carte des 27 stations au Luxembourg pour lesquelles des simulations et des prévisions de débit sont possibles dans le modèle régional LSTM.

Le modèle a été élaboré à partir de données de débit validées et de paramètres météorologiques (précipitations, température de l'air, rayonnement global, humidité

relative de l'air, pression atmosphérique et vitesse du vent) sur la période 2002-2022. La période d'entraînement s'étend sur 15 ans (2003-2017), tandis que la période de test dure cinq ans (2018-2022) et inclut donc la crue extrême de juillet 2021.

Lors de l'élaboration du modèle, les mêmes paramètres d'entrée que pour LARSIM ont été choisis comme données d'entrée pour des raisons liées à l'utilisation opérationnelle ultérieure du LSTM. Cela permet de créer des synergies dans le fonctionnement opérationnel si le LSTM est utilisé en parallèle avec LARSIM. Ces données d'entrée dynamiques et météorologiques sont importées dans le modèle sous la forme de valeurs moyennes de bassin versant converties en chroniques et interpolées par station et par paramètre. En outre, le débit mesuré jusqu'à l'instant de prévision est importé, ainsi que différentes caractéristiques statiques du bassin, telles que la taille du bassin versant ou la hauteur moyenne du terrain (pour distinguer les stations situées dans les différents bassins versants).

L'entraînement du modèle consiste à effectuer des simulations à l'aide de la météorologie mesurée. Ce modèle doit désormais permettre d'établir des prévisions horaires de débit avec un délai de prévision allant jusqu'à 45 heures, dans un premier temps uniquement avec les prévisions déterministes à court terme du modèle DWD ICON-D2 (délai de prévision de 48 heures actualisé toutes les trois heures). Ensuite, il sera également possible d'établir des prévisions d'ensemble d'autres produits.

Pour ce faire, une nouvelle approche de MF-LSTM (LSTM multi-fréquences) est utilisée pour la première fois. Cette extension spéciale du LSTM permet une prévision opérationnelle des débits avec des valeurs horaires. Contrairement à ce qui se faisait auparavant, il est désormais possible d'utiliser des valeurs journalières comme données d'entrée pour la plupart des journées de l'année écoulée, ce qui réduit considérablement l'effort de calcul.

Lors de l'entraînement et de l'optimisation du modèle, des prévisions de débit sont établies pour une partie des données d'entraînement. Les différences calculées entre les mesures et les prévisions sont transmises au LSTM et les paramètres internes du LSTM (pondérations) sont adaptés/optimisés. Une fois cette opération répétée suffisamment de fois, des tests de prédiction sans optimisation sont effectués pour le jeu de données de test. Ce processus est répété jusqu'à ce qu'aucune amélioration supplémentaire ne soit attendue.

Une partie des données CAMLES-DE a été utilisée comme jeu de données supplémentaire pour améliorer la qualité et la capacité d'extrapolation du modèle, mais avec une pondération moindre que les données des stations luxembourgeoises.

L'évaluation des simulations est réalisée à l'aide d'hydrogrammes composites pour les différents délais de prévision. Même sans les débits mesurés, les résultats sont très bons, avec un NSE (médiane) de 0,89, y compris pour les stations qui n'ont pas été entraînées. En intégrant les débits mesurés et le jeu de données CAMELS-DE, la qualité du modèle s'améliore encore pour atteindre un NSE (médiane) de 0,96 à tous les délais de prévision.

Sans le jeu de données CAMELS-DE, ce sont surtout les niveaux avec peu de données disponibles qui se détériorent pendant la période d'entraînement.

On constate toutefois encore des différences dans la qualité de la simulation entre les différents événements, parfois même en utilisant les débits mesurés. La capacité d'extrapolation ne fonctionne que de manière limitée : la crue extrême de juillet 2021 n'est pas simulée de manière optimale, s'il n'y a pas d'événement comparable dans les données d'entraînement (2021 n'est inclus que dans la période de test, pas dans la période d'entraînement). Cependant, cet événement est souvent de loin le plus important enregistré dans les stations sur l'ensemble de la période. L'utilisation du jeu de données CAMELS-DE permet d'améliorer en partie la capacité d'extrapolation.

La prochaine étape consistera à élaborer un concept pour la mise en œuvre opérationnelle des prévisions LSTM. Ces dernières doivent exploiter autant que possible les synergies avec le fonctionnement opérationnel de LARSIM et les données d'entrée déjà disponibles. Les principaux défis et les pistes d'amélioration possibles concernent toutefois la disponibilité variable des données, les lacunes pour certaines stations et certaines années, le traitement des données manquantes lors de l'exploitation opérationnelle, la prise en compte des impacts anthropiques, tels que les barrages, ainsi que la capacité d'extrapolation des événements extrêmes.