Design Project - SIE 2026

Comparison of Flow Prediction Models for the Arve River in Geneva
Evaluation of Operational Flood Forecasting Systems

Students:
Raphaëlle Simonet
Inès Laury

External partners:
OCEau Geneva
EPFL Supervisor:
Sara Bonetti

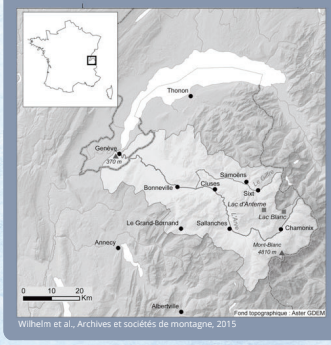
Context & Objectives

Flood forecasting on the Arve River is essential for flood risk management in Geneva (close bridges to traffic in case of flood). Several operational forecasting systems are currently used by the cantonal authorities, but their performances had not previously been evaluated within a unified framework.

Objectives

- Inventory and compare the operational forecasting systems
- Evaluate deterministic and probabilistic forecast performance
- Assess flood detection capability
- Provide operational recommendations for OCEau / CELSUCR

Study area



Wilhelm et al., Archives et sociétés de montagne, 2015

Arve River Basin: Alpine catchment
Strongly influenced by snow and glaciers
Basin area ≈ 2,000 km²
Elevation range: 370 m to more than 4,800 m
Confluence with the Rhône River in Geneva

Methodology

- Python-based evaluation workflow
- Comparison with observed discharge at Genève Bout du Monde station for the 2020-2024 period
- Flood threshold: Q > 400 m³/s
- Evaluation up to 48 h lead times

Deterministic Evaluation

- Continuous metrics: NSE, KGE, RMSE, MAPE
- Flood detection metrics: peak discharge error, timing error, relative volume error
- Event detection metrics: POD (Probability of Detection), FAR (False Alarm Ratio), CSI (Critical Success Index), Confusion matrices
- Visual Analysis: Time series, Flood-event hydrographs, Scatterplots

Probabilistic Evaluation (FOEN)

- Probabilistic Metrics : CRPS / CRPSS, Brier Score,
- Ensemble Diagnostics : Reliability diagrams, Rank histograms, Spread-skill analysis.

Models Inventory

Framework	Type	Forecast	Horizon	Main characteristics
FOEN	Fully distributed physical model	Probabilistic ensemble	Up to 10 days	Explicit uncertainty representation using NWP forcings, uses a WaSiM model
Hydrique Physique	Semi-distributed physical model	Deterministic	Up to 14 days	Uses Routing System, a semi-distributed conceptual model with process-based components
Hydrique ML	Multi Layer Perceptron	Deterministic	+24 h	Data-driven approach using historical discharge, meteorological inputs and simplified Hydrique Phy. state variables
SIG-CNR	Hydrological + statistical models	Deterministic	Up to 5 days	Expert-based workflow combining GRP rainfall-runoff reservoir and ARX statistical models

Results

SIG-CNR

Best deterministic reference system
✓ Highest overall deterministic skill
✓ Strong flood detection capability
⚠ More false alarms during flood situations

Hydrique physique

Most balanced operational model
✓ Stable performance across lead times
✓ Good compromise between POD and FAR
✓ Reliable during flood warning situations

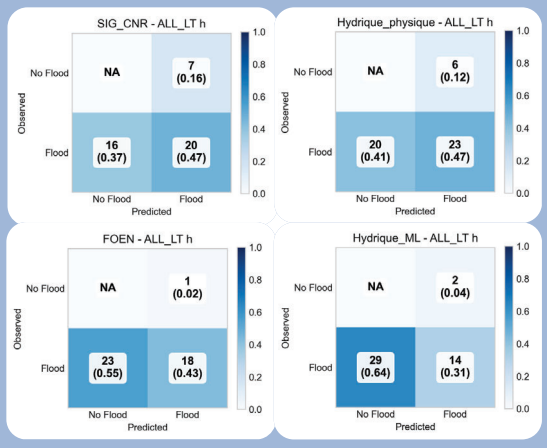
FOEN

Probabilistic decision-support tool
✓ Low false alarm rate
✓ High continuous performances with newer forcing configuration
⚠ Ensemble underdispersive

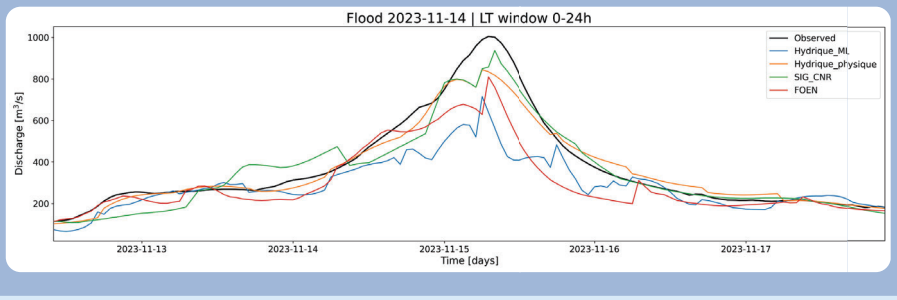
Hydrique ML

Complementary short lead-time model
✓ Good short-term behaviour
⚠ Strong underestimation of flood peaks
⚠ Lowest flood detection capability

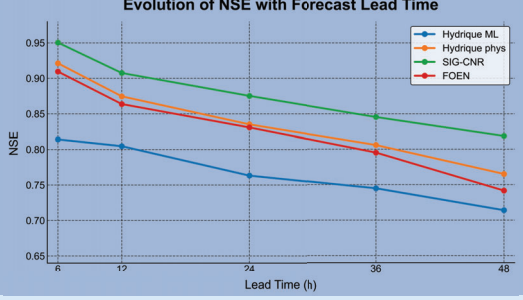
Event-based confusion matrices comparing observed to predicted floods for all lead times analyzed



November 15, 2023 flood: highest discharge at Genève Bout-du-Monde since records began in 1904 (~200-year return period)



Evolution of NSE with Forecast Lead Time



Operational recommendations

Situation	Recommended System	Main Advantage
Normal conditions	SIG-CNR	Best overall deterministic performance
Flood warning	Hydrique Phy. + SIG-CNR	Best flood detection capability
Long lead times / uncertainty	FOEN Ensemble	Probabilistic uncertainty information

Limitations and future perspectives

⚠ Analysis based on historical data, interpret with caution for unseen conditions
⚠ Different forcings, structures, and temporal samples limit inter-model comparison
⚠ Lead times rounded to nearest hour

→ Extended evaluation period to stabilise event-based metrics and test alternative flood thresholds
→ Ensemble dispersion correction (spread inflation, lagged ensemble or EMOS)
→ Ongoing model improvements (Hydrique ML retraining on ICON data underway)