Towards scalable design rainfall in Europe: building a continental dataset for local design applications

Pluies de projet évolutives en Europe : un ensemble de données pour les applications locales

Sigrid Schødt Hansen[a,b]*, Sara Maria Lerer[b], Roland Löwe[a], Hjalte Jomo Danielsen Sørup[c], Jonas Tranberg Hansen[b] and Peter Steen Mikkelsen[a]
[a]Department of Environmental and Resource Engineering, Technical University of Denmark
[b]Scalgo ApS
[c]Danish Meteorological Institute

RÉSUMÉ

La pluviométrie de projet est généralement estimée localement à partir d'observations pluviométriques de longue durée et à haute résolution, au moyen de courbes Intensité-Durée-Fréquence (IDF). Toutefois, les données nécessaires pour caractériser les extrêmes de courte durée ne sont pas disponibles de manière homogène en Europe, et les jeux de données nationaux existants sont souvent difficiles ou coûteux d'accès.

Pour combler cette lacune, nous constituons une vaste base de données transnationale d'observations pluviométriques sub-horaires. Elle permet d'analyser et de comparer les extrêmes pluviométriques entre régions et pays, et sert de données d'entraînement pour un cadre de modélisation spatiale fondé sur l'apprentissage automatique. Les premiers résultats confirment la pertinence de la distribution généralisée des valeurs extrêmes (GEV) pour modéliser les séries d'annuelles maximales.

À partir de ces éléments, nous développons un modèle basé sur des graphes pour prédire les courbes IDF à l'échelle européenne en utilisant des données climatologiques et géographiques accessibles au public. Le cadre obtenu fournira des données homogènes et à haute résolution pour soutenir une planification et une gestion de l'eau plus résilientes face au climat.

ABSTRACT

Design rainfall is typically estimated locally from long-term, high-resolution rain gauge observations using Intensity-Duration-Frequency (IDF) curves. However, the data required to capture short-duration extremes are not consistently available across Europe, and existing national datasets are often difficult or expensive to access. We address this gap by compiling a large transnational dataset of sub-hourly rainfall observations. The dataset enables comparative analyses of rainfall extremes across regions and countries and serves as training data for a spatial machine learning framework. Preliminary results confirm the suitability of the Generalised Extreme Value (GEV) distribution for modelling annual maximum series.

Building on these insights, we develop a graph-based model to predict IDF curves across Europe using publicly available climatological and geographical data. The resulting framework will provide consistent, high-resolution design rainfall data to support climate-resilient water management and planning.

KEYWORDS

climate adaptation, design rainfall, extreme precipitation, machine learning, urban hydrology

1 INTRODUCTION

Traditional regionalisation approaches for interpolating Intensity-Duration-Frequency (IDF) curves require dense networks of rain gauges with long records (Koutsoyiannis et al., 1998; Lanciotti et al., 2022), and, for urban applications, a temporal resolution that can capture the peak intensity of short-duration events. These data are not consistently available across national borders, meaning that the traditional regionalisation techniques cannot be applied at a continental scale.

To generate design rainfall consistently across Europe, we must therefore rely on methods that can generalize beyond the areas with gauge coverage, making use of the large-scale datasets that exist everywhere: reanalysis products and geographical predictors such as elevation. Reanalysis products integrate weather observations and climate models to create coherent datasets describing the recent past climate.

We aim to develop a scalable framework for estimating local design rainfall across Europe by:

- compiling a large transnational database of sub-hourly rainfall observations that can constitute training data,
- training a machine learning model using various climatological and geographical variables as input data to predict the probability of extreme rainfall events, represented by IDF curves, across multiple European countries, and
- extending the developed methodology to predict the impacts of climate change on the probability of extreme rainfall events, enhancing the models' utility in long-term water management planning.

This contribution presents the dataset foundation, initial analyses, and the ongoing modelling work.

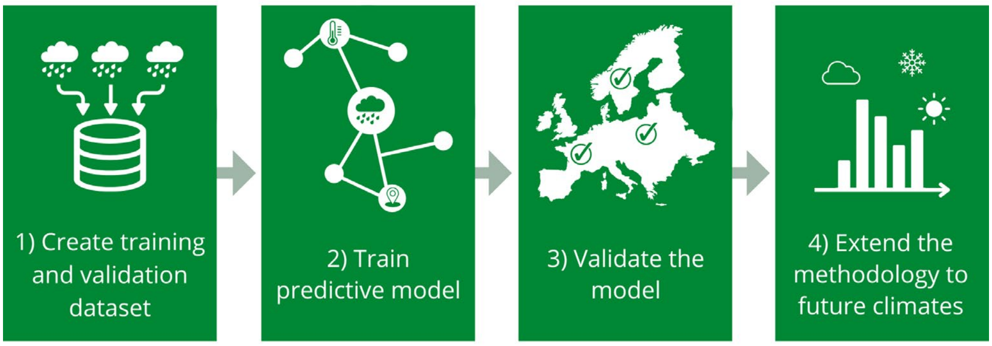


Figure 1. Project overview. Phases have been split into work packages.

2 DATASET: EUROPEAN SUB-HOURLY RAINFALL MEASUREMENTS

2.1 Resolution and coverage

Currently, we have collected rainfall observations from more than five thousand rain gauges across six different countries (Denmark, France, Germany, Norway, Poland and Sweden). The spatial distribution of these data is visualised in Figure 2. The dataset includes both weighting and tipping-bucket gauges with temporal resolutions from 1 to 15 minutes. Record availability varies across countries: some networks extend back to the late 1960s (Norway), while others were established more recently (Germany, Poland and France), and roughly three thousand of the gauges provide more than ten years of observations.

2.2 Quality control and harmonisation

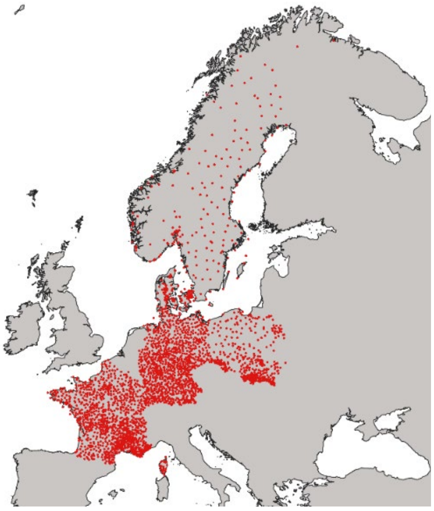
Preliminary analysis of the collected rainfall observations showed that, in addition to the quality control conducted by the national meteorological institutes, more quality control was needed for the data to be of comparable quality. We have chosen a pragmatic approach to quality control and removed:

- unrealistically high observations, i.e. 1-hour rainfall intensities higher than the world record of 400 mm/h (Blenkinsop et al., 2017),

- station years if more than 20% of the observations in the period from May to October were missing,

time series with record lengths of less than 10 years.

Figure 2. Spatial distribution of rain gauges (red dots) in the compiled dataset.



2.3 Insights from the compiled dataset

The rainfall data collected so far already reveal insights that have guided our methodological choices. The suitability of the Generalized Extreme Value (GEV) distribution as a parent distribution for annual maximum series was confirmed by L-moment ratio diagrams (Hansen et al., in review). The resulting return levels were generally comparable to national design rainfall values. The fitted parameters also showed a clear north-south gradient in the location parameter, consistent with findings in previous studies (e.g. Shehu et al. (2023), indicating that large-scale climatic gradients are reflected in the statistical properties of extremes.

3 SPATIAL MODELLING FRAMEWORK

We are developing a machine learning model for spatial prediction of IDF curves (cf. Figure 3). The model will combine the compiled dataset with geographical and climatological covariates such as elevation (EU-DEM), land cover (CORINE), and reanalysis-based climate statistics (CERRA). Because the gauges are irregularly spaced, a graph representation is used to capture spatial proximity and allow nearby stations to influence each other.

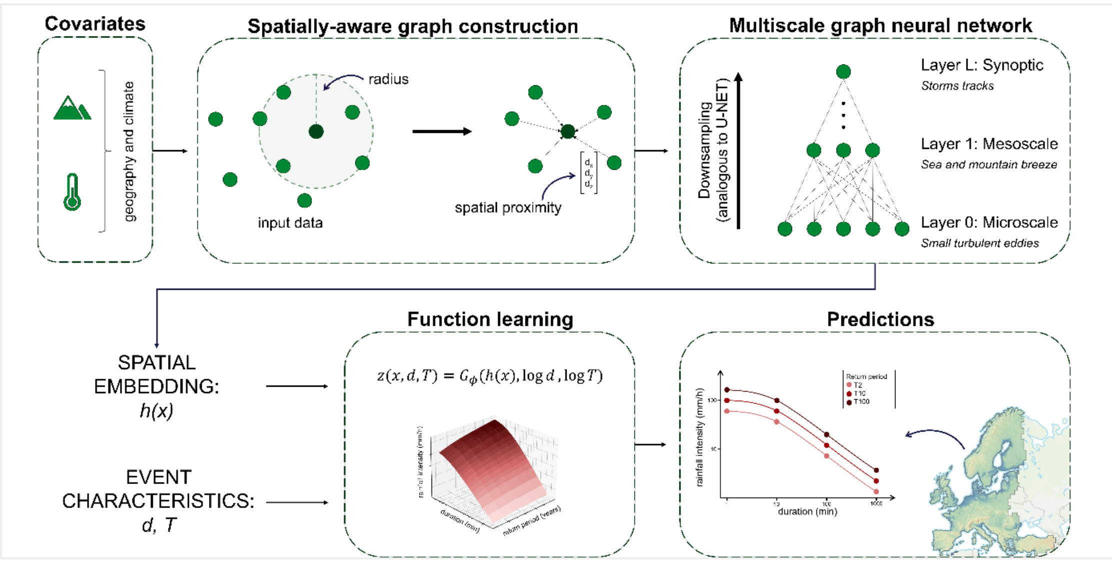


Figure 3. Conceptual modelling framework

To represent atmospheric processes acting across different spatial scales, the model will follow a hierarchical structure analogous to the U-Net architecture, where information is progressively downsampled and aggregated from local to regional contexts (Ronneberger et al., 2015) . Finally, the model is designed to learn a continuous IDF function rather than predicting a fixed set of return levels. This follows the function-learning paradigm of DeepONet (Lu et al., 2020), enabling the model to produce rainfall intensities for any combination of duration and return period at any location.

4 NEXT STEPS AND EXPECTED OUTCOMES

In the coming months, we will continue the modelling experiments, systematically evaluating how well the model generalises to unseen countries and how its return level estimates compare with national design guidelines. Additional rain gauges will be incorporated in regions where the model shows limited generalisability to strengthen the spatial representativeness of the training data.

Our aim is to deliver a spatially distributed dataset of IDF curve predictions for present climate conditions across Europe, as well as corresponding IDF projections under future climate scenarios. These products will provide consistent and locally relevant design rainfall information across borders. When integrated into digital planning tools such as Scalgo Live, design rainfall generation can occur seamlessly, making advanced analyses accessible to non-experts. The practical significance of this research lies in its potential to improve the accessibility of design rainfall to local planners, directly contributing to the climate resilience of infrastructure. And from a scientific perspective, this research advances our understanding of the spatial behaviour of sub-hourly extremes on a continental scale.

ACKNOWLEDGEMENTS

The project is funded by Innovation Fund Denmark and Scalgo. We thank the Danish Water Pollution Committee of the Danish Society of Engineers (SVK), Météo-France, the Deutscher Wetterdienst (DWD), the Norwegian Meteorological Institute, the Polish Institute of Meteorology and Water Management – National Research Institute (IMGW-PIB), and the Swedish Meteorological and Hydrological Institute (SMHI) for providing access to the sub-hourly rainfall observations used in this study.

LIST OF REFERENCES

Blenkinsop, S., Lewis, E., Chan, S.C., Fowler, H.J., 2017. Quality-control of an hourly rainfall dataset and climatology of extremes for the UK. International Journal of Climatology 37, 722–740. https://doi.org/10.1002/joc.4735

Hansen, S.S., Lerer Maria, S., Löwe, R., Sørup, H.J.D., Mikkelsen, P.S., n.d. Estimation of intensity-duration-frequency curves in Northern Europe using sub-hourly rainfall observations.

Koutsoyiannis, D., Kozonis, D., Manetas, A., 1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships. J Hydrol (Amst) 206, 118–135. https://doi.org/10.1016/S0022-1694(98)00097-3

Lanciotti, S., Ridolfi, E., Russo, F., Napolitano, F., 2022. Intensity–Duration–Frequency Curves in a Data-Rich Era: A Review. Water (Switzerland) 14, 3705. https://doi.org/10.3390/w14223705

Lu, L., Jin, P., Karniadakis, G.E., 2020. DeepONet: Learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. https://doi.org/10.1038/s42256-021-00302-5

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation.

Shehu, B., Willems, W., Stockel, H., Thiele, L.B., Haberlandt, U., 2023. Regionalisation of rainfall depth-duration-frequency curves with different data types in Germany. Hydrol Earth Syst Sci 27, 1109–1132. https://doi.org/10.5194/hess-27-1109-2023