A LATENT MODEL OF STREET-LEVEL URBAN FLOOD PREDICTION USING OPEN DATA

Yue Qin [1]*, Xiaowei Jia [2] and Rosta Farzan [1]

[1] School of Computing and Information, University of Pittsburgh, Pittsburgh, USA
[2] Department of Computer Science, Rutgers University, New Brunswick, USA
* yuq9@pitt.edu

Keywords: Urban flood, flood prediction, crowdsourcing, FloodNet sensor

Introduction

As urban floods become more frequent and damaging, understanding when and where floods will occur is crucial for hazard management and mitigation. Most flood models adopt observations from either physical sensors or social sensors (crowdsourced information) to predict sensor measurements or flood reports (e.g., Liu et al., 2024b, Mohamadiazar et al., 2024). These models can suffer from limited spatial transferability due to fixed sensor locations (Wang et al., 2025) or biased predictions due to heterogeneous crowdsourcing behaviours (Liu et al., 2024a). Moreover, the sparsity of publicly available data has greatly limited the capabilities of existing data-driven approaches in making street-level predictions (Yuan et al., 2022), which provides fine-grained flood information for decision-making in urban planning and transportation. A model leveraging open data collected from both physical and social sensors and predicting unbiased flood occurrences with a fine spatiotemporal resolution is still lacking in the current literature.

In this paper, we propose an urban flood prediction model that decomposes the task of flood prediction into flood generation and flood observation, which are connected through a latent state representing true flood occurrences. Our main contributions include: (1) we collected, processed, and publicly released two fine-grained datasets on street segment-level urban floodings using open data available for Pittsburgh and New York City (NYC), (2) we proposed a latent-state flood prediction model that allows inferring true flood occurrences at the street-segment level while uncovering reporting bias in crowdsourcing, (3) we conducted a comprehensive evaluation of our model using the NYC data and demonstrated the model's capabilities of capturing different types of floods, predicting sensor-detected and citizen-reported floods, and inferring road network status during extreme events.

Datasets

We have made public two datasets on urban floods at the street-segment level by integrating publicly available data, one for Pittsburgh (Qin, 2025b) and the other for NYC (Qin, 2025a). They represent data-sparse and data-intensive cities with different topological characteristics (hilly versus flat) and major flood types (pluvial and fluvial versus pluvial and coastal). For data-sparse cities, the main sources of data we identified include crowdsourced reports from official records (NOAA storm event database) and citizens (city 311). While these sources also cover data-intensive cities like NYC, such regions could involve additional data from physical flood sensors (FloodNet, Mydlarz et al., 2024) and multiple crowdsourcing platforms (MyCoast).

The data processing pipeline involves automatic extraction of geographical information from natural language descriptions using LLMs (GPT-4o-mini), mapping the extracted street information on the Open Street Map (OSM) network, and fusion of the mapped segments from heterogeneous data sources. The outcome datasets include 488 segment floodings from June 2015 to December 2024 in Pittsburgh, and 12,604 segment floodings from October 2020 to December 2024 in NYC. They allow aggregation at different spatiotemporal resolution, informing decision-making and flood modelling across multiple levels. We demonstrate the utility of the datasets below using the NYC data.

Method

Proposed model

We identified **flood generation** and **flood observation** as two distinct processes connected by a **latent**, binary state of true flood occurrences $Z_{e,d}$ at segment e on day d, as shown in Figure 1. The flood generation process adopts a multi-layer perceptron (MLP) with 2 hidden layers to predict the binary state $Z_{e,d}$ from common flood drivers (e.g., Mohamadiazar et al., 2024; Roy et al., 2025), including weather features (*one-day, 3-day, and 7-day total precipitation* at the Central Park station), tide features (*daily maximum water level* and *maximum surge* of The Battery and Kings Point gauges), spatial features (*grade/slope, elevation, distance to waterbody, distance to streamline, and height above the nearest drainage (HAND)*), and street network topologies (*16-dimension node2vec embeddings* (Grover and Leskovec, 2016) trained on the entire OSM street network (Boeing, 2025)).

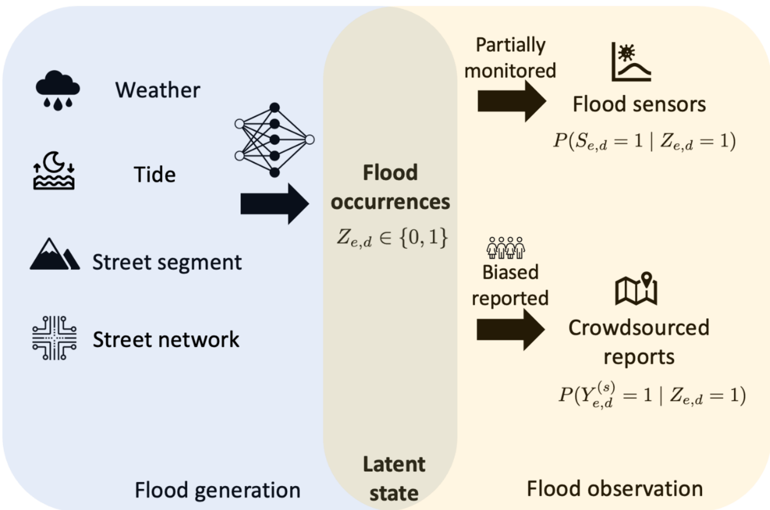

Figure 1: Our proposed latent flood prediction model.

For flood observation, we hypothesize that true flood occurrences can be indirectly observed by flood sensors and/or crowdsourced reports modelled by the marginalized probabilities conditioned on $Z_{e,\,d}$. The conditional probability of sensor-detected floods remains constant under the same sensor configurations, while that of receiving reports for each crowdsourcing platform s is modelled as a function of census-tract demographics D_{tract} (including population, area, and socioeconomics derived from Census data) based on findings from biased reporting (Agostini et al., 2025):

$$P(Y_{e,d}^{(s)} = 1 \mid Z_{e,d} = 1) = \sigma(\alpha_s + \beta_s^{\mathrm{T}} D_{tract})$$

The training loss combines marginalized probabilities of observing floods through sensors and/or crowdsourced reports. The model is called "report-only" when λ_{sensor} is 0 and "sensor-only" when λ_{report} is 0.

$$\mathcal{L} = \lambda_{report}\mathcal{L}_{report} + \lambda_{sensor}\mathcal{L}_{sensor}$$

Evaluation

We used AUC-ROC as our evaluation metric to examine prediction performance due to its suitability for class imbalance when false negatives are costly (McDermott et al., 2024). We constructed 22 sliding windows from October 2020 to December 2024, each containing 30 consecutive months with a 24:6 train-test split. The evaluation paradigm includes investigating the *entire latent structure, flood generation, flood observation*, and *latent state inference*. All models were trained on a single CPU, with a learning rate of 1e-3 and a batch size of 2,048.

Results

Latent model versus non-latent baseline

Model	Training loss	AUC (Sensor) ↑	AUC (Reports) ↑
Proposed latent model	Report-only	0.852 [0.820, 0.884]	**0.805** [0.787, 0.823]
	Report + sensor	0.800 [0.750, 0.850]	0.718 [0.680, 0.756]
Baseline non-latent model	Report-only	0.608 [0.527, 0.688]	0.764 [0.744, 0.784]
	Sensor-only	**0.931** [0.916, 0.946]	0.489 [0.432, 0.545]
	Report + sensor	0.927 [0.911, 0.944]	0.712 [0.668, 0.755]

Table 1: AUC results (mean and 95% CIs) pooled across 22 sliding windows. Sensor-only latent model is not comparable due to not adopting demographic features.

We first constructed a non-latent baseline model with the same set of features and MLP architecture as the proposed model and added separate output layers for sensors and reports. The results in Table 1 indicate that flood reports and sensors represent distinct flood observation mechanisms, as baseline models trained on either of them performs poorly in predicting the other. We also find a trade-off between prediction accuracy and interpretability (i.e., propensity of reporting and true flood occurrences) comparing the baseline and the proposed model jointly trained on sensors and reports, but the report-only latent model achieves both when compared to the report-only baseline. Finally, adding sensor supervision harms the performance in the latent model, which may due to reporting delays (Liu et al., 2024a) or non-binary latent states (e.g., different flood depth).

Feature ablation in flood generation

We find from Table 2 that precipitation features are more predictive of reports. By contrast, tide features are more predictive of sensor-detected floods, which is consistent with the purpose of the FloodNet project in monitoring coastal floods (Mydlarz et al., 2024). In general, dynamic features are more important in prediction performance than static, spatial features.

Features	AUC (Sensor) ↑	AUC (Reports) ↑
Full (precipitation + tide + segment + node2vec)	**0.852** [0.820, 0.884]	**0.805** [0.787, 0.823]
w/o node2vec	0.845 [0.808, 0.881]	0.797 [0.776, 0.818]
w/o segment	0.833 [0.802, 0.863]	0.789 [0.768, 0.810]
w/o precipitation	0.820 [0.786, 0.854]	0.731 [0.705, 0.757]
w/o tide	0.814 [0.774, 0.854]	0.796 [0.776, 0.815]

Table 2: AUC results (mean and 95% CIs) for the best-performing latent model (report-only) under different prior feature ablations.

Reporting bias in crowdsourced observations

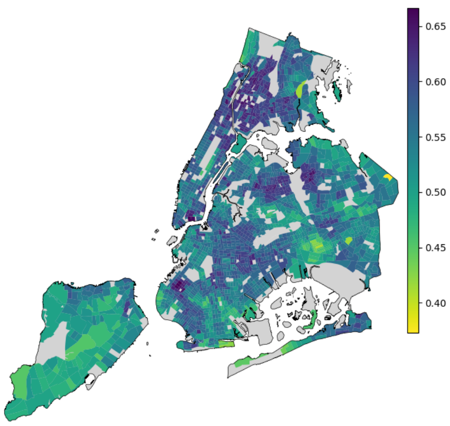

Figure 2. Reporting propensity of census tracts learned by the proposed model.

Census-tract demographics	Mean ± 95% CI
Log Population	0.050 ± 0.018
Median Age	−0.047 ± 0.014
Census Tract Area (m²)	−0.056 ± 0.020
Median Household Income	0.012 ± 0.015
Bachelor's Degree or Higher (%)	−0.024 ± 0.019
White Population (%)	−0.051 ± 0.017
Renter-Occupied Households (%)	0.044 ± 0.012

Table 3: Demographic coefficients of 311 reports pooled across the models.

Our findings from Figure 2 and Table 3 are mostly consistent with the literature on demographic bias in urban crowdsourcing (Agostini et al., 2024; Balachandar et al., 2026; Liu et al., 2024a). Specifically, areas with lower age, higher population, and larger tract area are associated with higher reporting propensity. However, different from the existing findings, our results indicate that the reporting propensity is negatively related to the percentage of white population, while being positively related to the percentage of renter-occupied households. It may be attributed to the different types of requests and modelling approaches between our study and existing literature.

Latent status of street network during storms

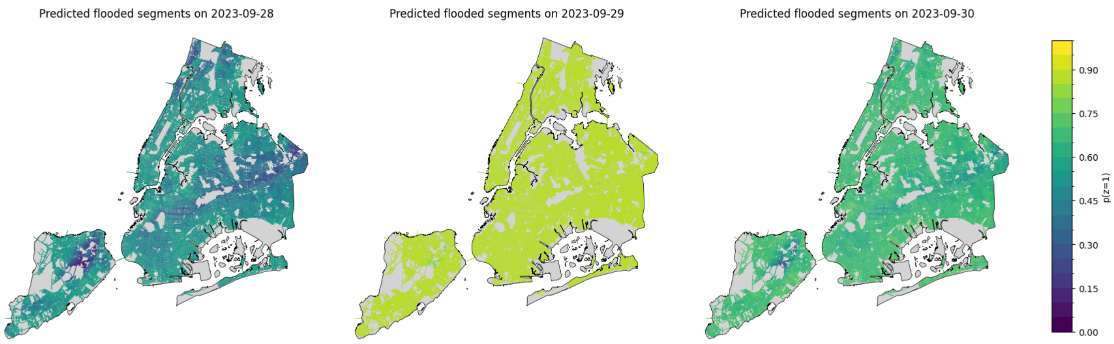

Figure 3. Latent true flood occurrences during Tropical Storm Ophelia inferred by the model.

Figure 3 shows the model inference on the flooded segments a day before, during, and following Tropical Storm Ophelia (September 2023). It demonstrates that the model can capture sudden onsets of storm-driven disruptions and distinguish among normal (non-flooded), peak flooding, and flood recession periods. The model also learns the spatial pattern of flood risks, including locations near the coastline and those at a higher elevation.

Conclusion

We proposed two fine-grained street-level flood datasets and a latent-state flood prediction model capable of inferring true flood occurrences at the street-segment level and interpreting demographic bias in crowdsourcing. Future work could investigate the spatial transferability of the model applied to other cities and compare the prediction performance when adopting more complex deep learning architectures in flood generation.

Acknowledgements:

The authors declare no funding sources for the research.

References

Agostini, G., Pierson, E., & Garg, N. (2024, March). A Bayesian spatial model to correct under-reporting in urban crowdsourcing. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 20, pp. 21888-21896).

Balachandar, S., Sadhuka, S., Berger, B., Pierson, E., & Garg, N. (2026, March). Urban incident prediction with graph neural networks: Integrating government ratings and crowdsourced reports. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 45, pp. 38194-38204).

Boeing, G. (2025). Modeling and analyzing urban networks and amenities with OSMnx. Geographical Analysis, 57(4), 567-577.

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864).

Liu, Z., Bhandaram, U., & Garg, N. (2024a). Quantifying spatial under-reporting disparities in resident crowdsourcing. Nature Computational Science, 4(1), 57-65.

Liu, Z., Felton, T., & Mostafavi, A. (2024b). Interpretable machine learning for predicting urban flash flood hotspots using intertwined land and built-environment features. Computers, Environment and Urban Systems, 110, 102096.

McDermott, M. B., Zhang, H., Hansen, L. H., Angelotti, G., & Gallifant, J. (2024). A closer look at AUROC and AUPRC under class imbalance. Advances in Neural Information Processing Systems, 37, 44102-44163.

Mohamadiazar, N., Ebrahimian, A., & Hosseiny, H. (2024). Integrating deep learning, satellite image processing, and spatial-temporal analysis for urban flood prediction. Journal of Hydrology, 639, 131508. https://doi.org/10.1016/j.jhydrol.2024.131508

Mydlarz, C., Sai Venkat Challagonda, P., Steers, B., Rucker, J., Brain, T., Branco, B., Burnett, H. E., Kaur, A., Fischman, R., Graziano, K., Krueger, K., Hénaff, E., Ignace, V., Jozwiak, E., Palchuri, J., Pierone, P., Rothman, P., Toledo-Crow, R., & Silverman, A. I. (2024). FloodNet: Low-Cost Ultrasonic Sensors for Real-Time Measurement of Hyperlocal, Street-Level Floods in New York City. Water Resources Research, 60(5), e2023WR036806. https://doi.org/10.1029/2023WR036806

Qin, Y. (2025a). NYC floods at street-segment level from open data (2020-2024) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.18062104

Qin, Y. (2025b). Pittsburgh (PA, USA) floods at street-segment level from open data (2015-2024) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.17562033

Roy, B., Goodall, J. L., McSpadden, D., Goldenberg, S., & Schram, M. (2025). Forecasting Multi-Step-Ahead Street-Scale Nuisance Flooding using a seq2seq LSTM Surrogate Model for Real-Time Application in a Coastal-Urban City. Journal of Hydrology, 656, 132697. https://doi.org/10.1016/j.jhydrol.2025.132697

Wang, M., Zhuang, J. A., Zhao, J., Sun, C., Li, J., Zhou, L., ... & Qi, J. (2025). A cross-city transferable convolutional neural network framework for assessing street-scale flood risks in urban networks. Journal of Environmental Management, 395, 127977.

Yuan, F., Xu, Y., Li, Q., & Mostafavi, A. (2022). Spatio-temporal graph convolutional networks for road network inundation status prediction during urban flooding. Computers, Environment and Urban Systems, 97, 101870.