

Extracting historical flood locations from news media data by the named entity recognition (NER) model to assess urban flood susceptibility

Shengnan Fu, Heng Lyu, Ze Wang, Xin Hao, Chi Zhang

School of Hydraulic Engineering, Dalian University of Technology, Dalian, 116024, China



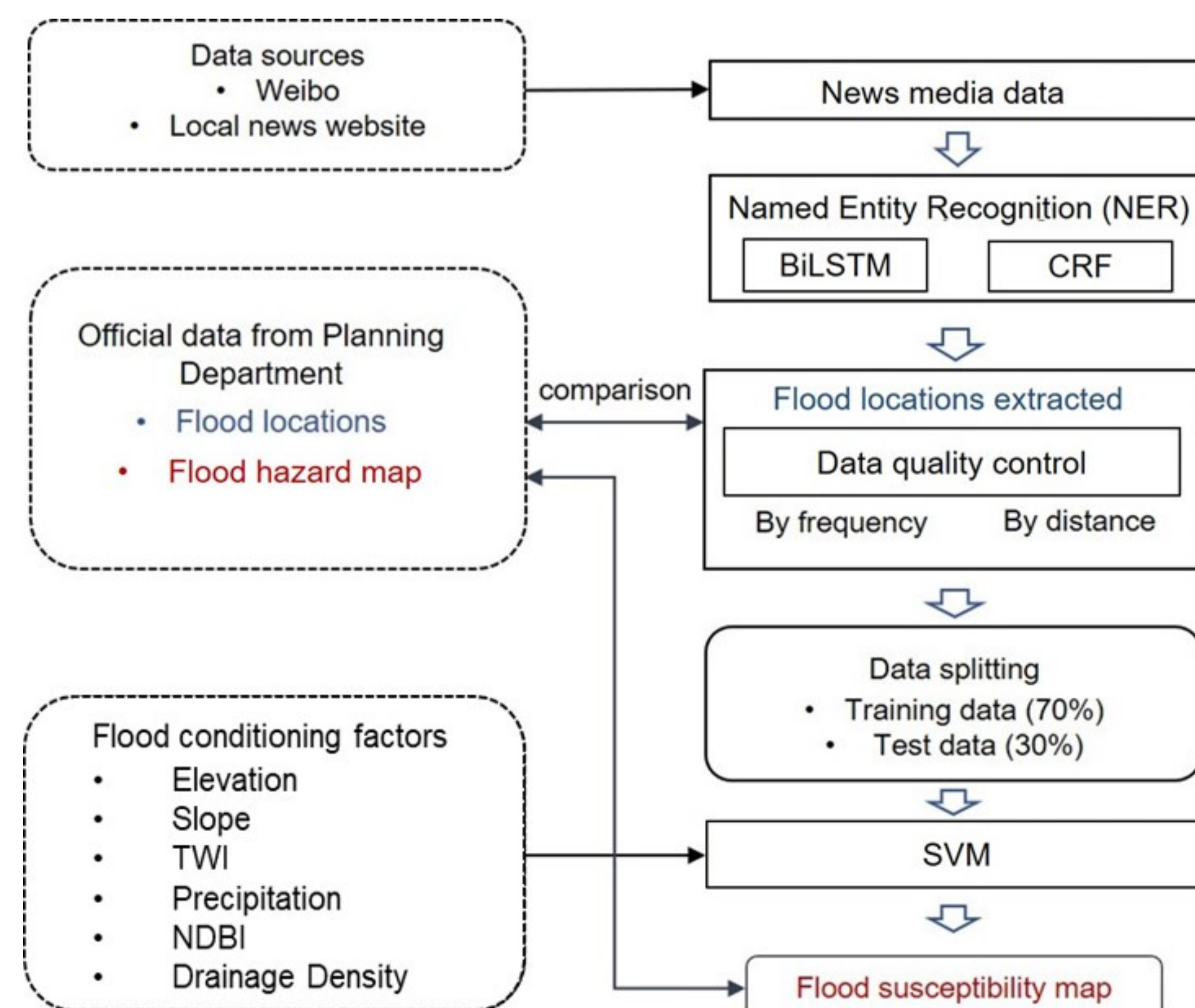
Introduction

- Flood susceptibility assessment for identifying flood-prone areas plays a significant role in flood hazard mitigation.
- Machine learning is an optional assessment method because of its high objectivity and computational efficiency, but how to get enough and accurate information of historical flood locations to train the machine learning models has been a key problem.
- In recent years, news media data from both news websites and social media accounts has emerged as a promising source for natural science studies (Avellaneda et al., 2020).
- However, the application of news media data in urban flood susceptibility assessment is still inadequate. This study proposed an approach to fill this gap.



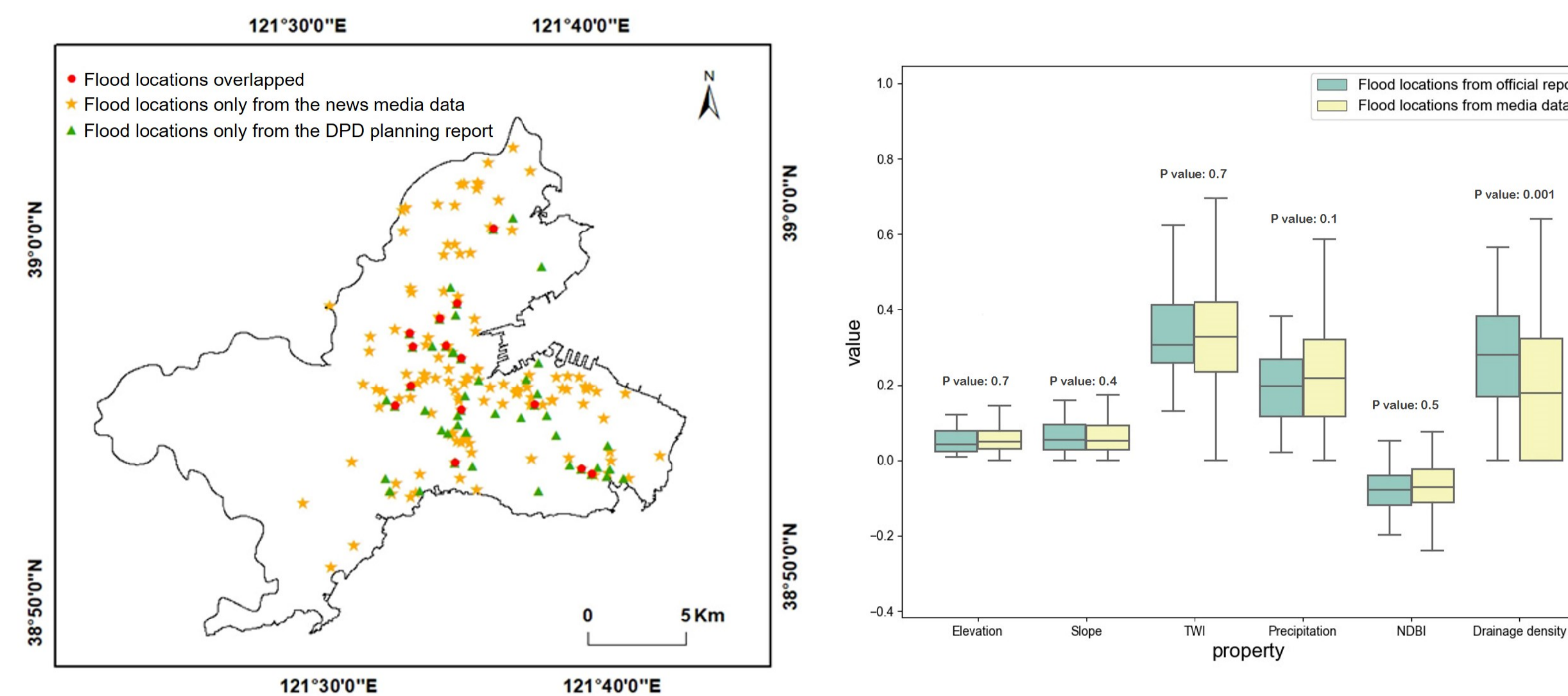
Methods

- Flood locations were recognized by a NER (BiLSTM-CRF) model.
- A frequency-based or distance-based data quality control method was used to improve the representativeness of the extracted flooded locations.
- The flood susceptibility map was generated with the help of an SVM model.



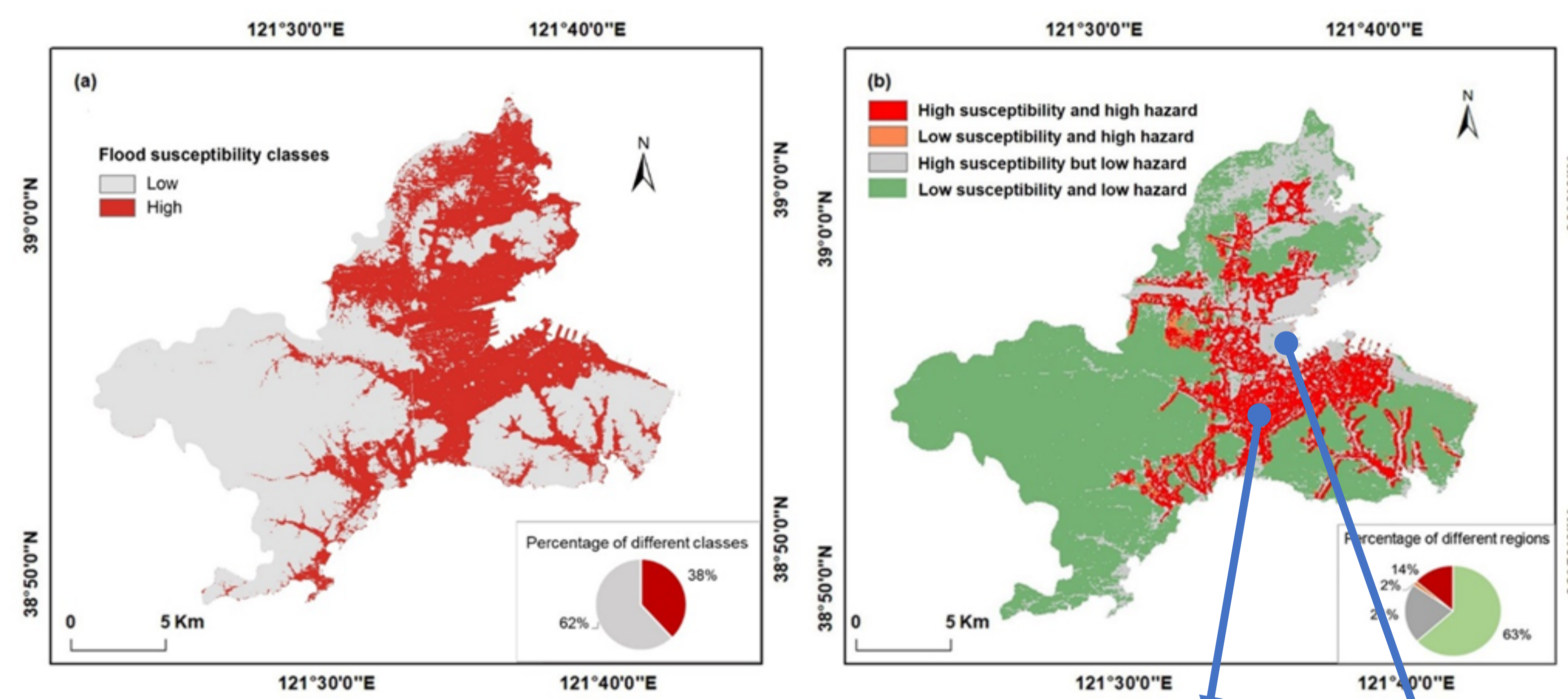
Results

◆ Flood locations from the news media data and the validation



The distributions of most flood conditioning factors at flood locations extracted from news media data are not significantly different from those of the official report ($P > 0.05$).

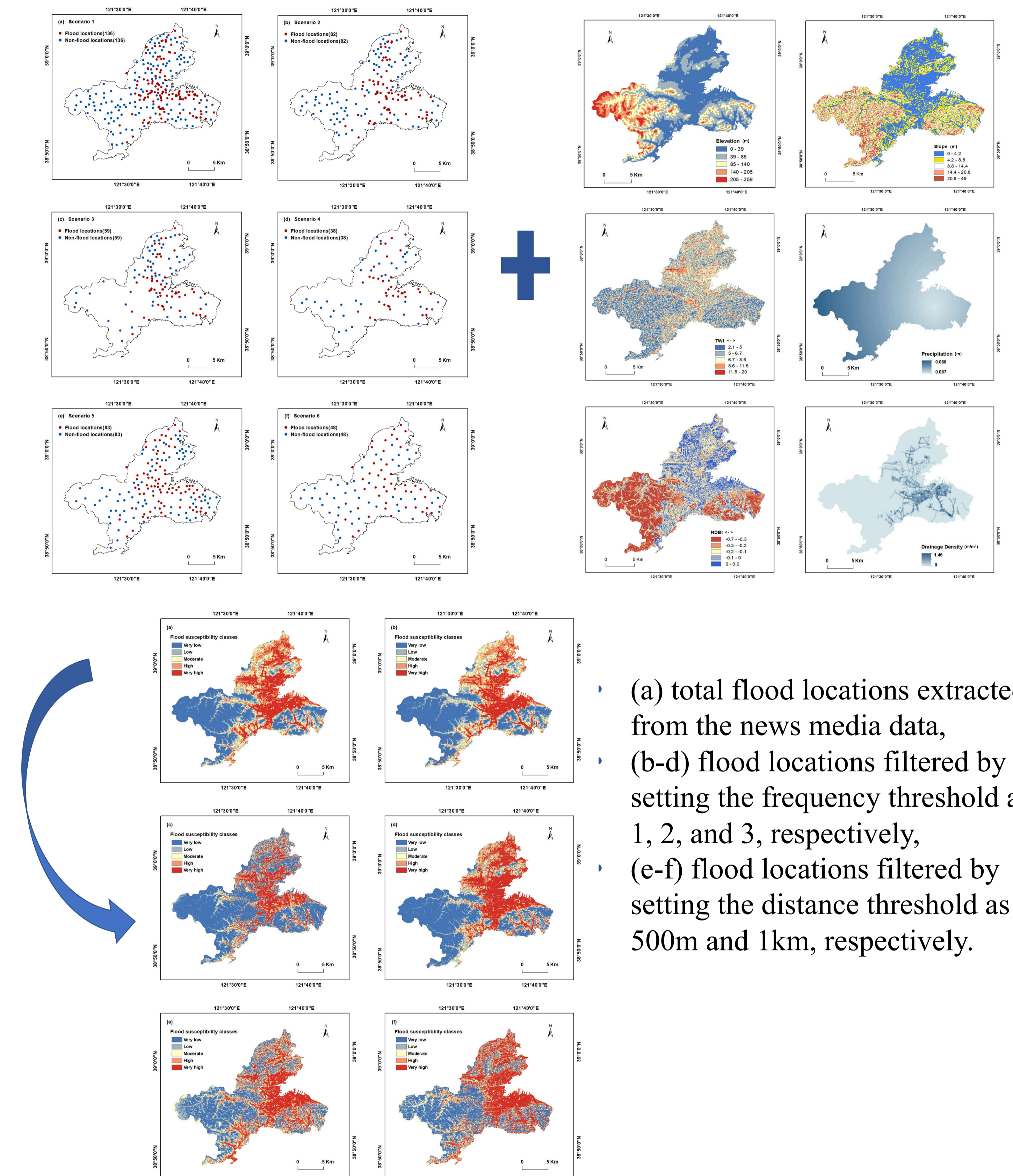
◆ Flood susceptibility map and the validation



- A high recall index of 90% indicates that most high hazard areas in the planning report have been identified as high susceptibility.
- The mismatched high susceptibility areas were mostly distributed in the built-up areas with low elevation and low drainage capacity.

✓ *This study verified news media data to be capable of supplementing historical flood inventory and mapping urban flood susceptibility.*

◆ Flood susceptibility maps obtained using different quality control methods



- (a) total flood locations extracted from the news media data,
- (b-d) flood locations filtered by setting the frequency threshold as 1, 2, and 3, respectively,
- (e-f) flood locations filtered by setting the distance threshold as 500m and 1km, respectively.

- ✓ *The frequency-based method can improve the precision of the high flood susceptibility results by up to 5% to some extent.*
- ✓ *The distance-based method could not lead to a more accurate flood susceptibility, which indicates that over-uniformly dispersed samples cannot reflect the heterogeneity of the urban area.*

References

Avellaneda, P.M., Ficklin, D.L., Lowry, C.S., Knouft, J.H., Hall, D.M., 2020. *Improving Hydrological Models With the Assimilation of Crowdsourced Data*. Water Resources Research 56.