

MAPPING FLOOD SUSCEPTIBILITY USING RANDOM FOREST EXPLOITING SATELLITE OBSERVATIONS AND GEOMORPHIC FEATURES

Jorge Saavedra Navarro¹, Ruodan Zhuang¹, Cinzia Albertini, Salvatore Manfreda¹

¹ Department of Civil, Construction and Environmental Engineering (DICEA), University of Naples Federico II, Naples, Italy
Corresponding author: Jorge Saavedra Navarro (jorgeandres.saavedranavarro@unina.it)

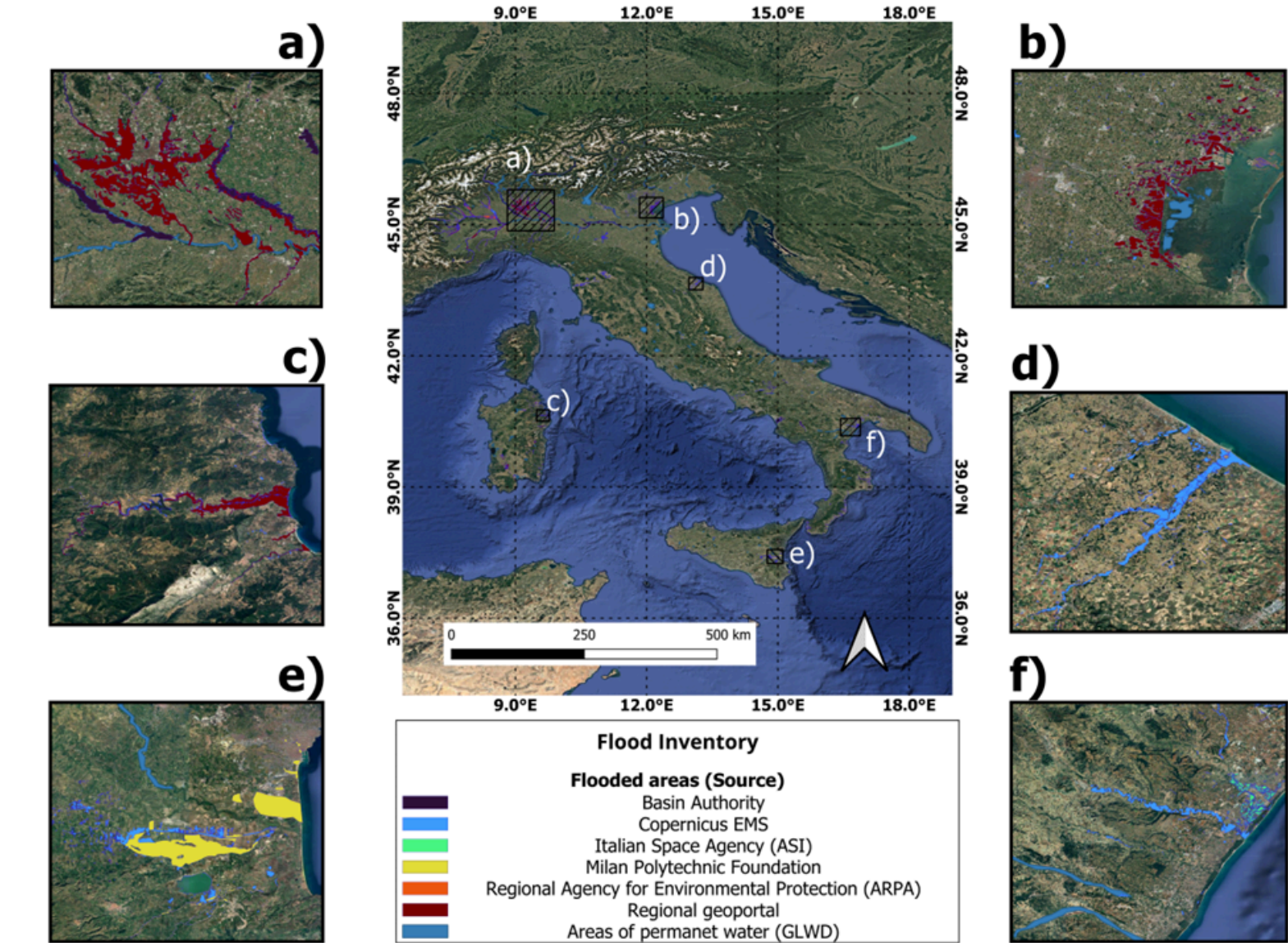
ABSTRACT

Flood events rank among the most destructive natural hazards, necessitating comprehensive risk management strategies to mitigate their impact on human health, the environment, cultural heritage, and economic activities.

This study investigates the potential of the **Random Forest (RF) model** to map flood-prone areas and assess the flood susceptibility of the Italian territory. The optimal set of predictor factors (SoF) was determined by evaluating a total of 25 variables spanning hydrologic, topographic, and categorical information. Particular attention was given to maximizing the information and avoiding redundancy and high correlations among factors through innovative methods such as the Average Merit of Information (**AMI**) and Variance Inflation Factor (**VIF**).

The model Calibration and validation were conducted using Copernicus Emergency Management Service maps, along with regional data from historical flood event records and permanent water areas. The results highlight that AMI effectively identifies the most relevant factors. Furthermore, the RF model trained with Set of Factors - 1 (SoF-1) demonstrated superior generalization capacity compared to other tested sets, as evidenced by a quantitative comparison with official flood hazard maps (AUCROC greater than 0.9).

The inclusion of **GFI** significantly improved prediction accuracy in unexplored areas, though challenges persist in extensively flat regions where geomorphic indicators are less distinct. Therefore, the integration of satellite-derived information and complementary datasets facilitates the accurate identification of flood-prone areas, streamlining computational processes and enabling preliminary analyses for decision-makers. These findings underscore the importance of leveraging advanced modelling techniques and continuously updated data to inform effective flood risk management policies and practices.



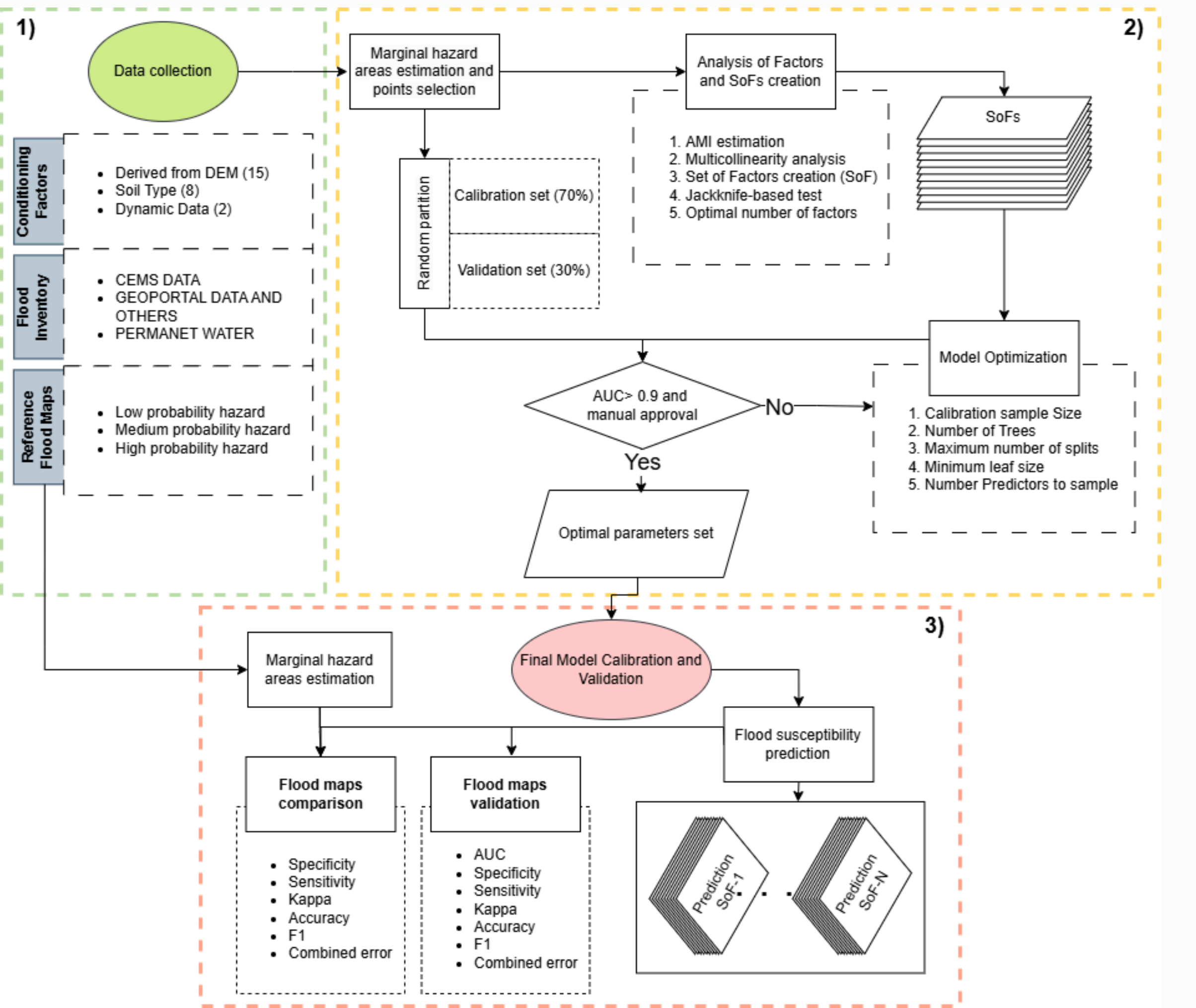
INTRODUCTION AND OBJECTIVES

In the literature, various factor combinations are possible, but these factors are often evaluated based on a single metric (model importance, IGR, VIF, etc). In this study, AMI is introduced as a metric to assess the information of continuous variables [-inf to inf] using a simple average of three different metrics (RF importance, Cohen's d and IGR). The categorical variables were analysed with the most important continuous factors using a jackknife-based test. Finally, the selected factors were used to create eleven sets of factors, which were evaluated with the internal set (validation) and an external set (comparison with official flood maps).

The flood conditioning factors used are elevation, tangential slope, aspect, tangential and profile curvature, TWI, SPI, Topographic Position Index (TPI), Surface Roughness Factor (SRF), geomorphology (terrain forms), geomorphic features including HAND, GFI, potential floodwater depth (hr), Distance to the nearest rivers (DNR), Contributing Area (CA), Flow Accumulation (FA), percentage of organic matter, sand, silt, and clay particles of the topsoil, soil texture, Hydrologic Soil Groups (HSG), a geolithological map, land cover (LC), hourly precipitation to calculate the Mean Maximum Daily Precipitation (MMDP), and the NDVI to estimate the mean minimum NDVI.

- Analyse the importance of the flood conditional factor and its influence,
- Map flood-prone areas based on different sources (Satellite observations and others)
- Explore the transferability skills of the RF model.

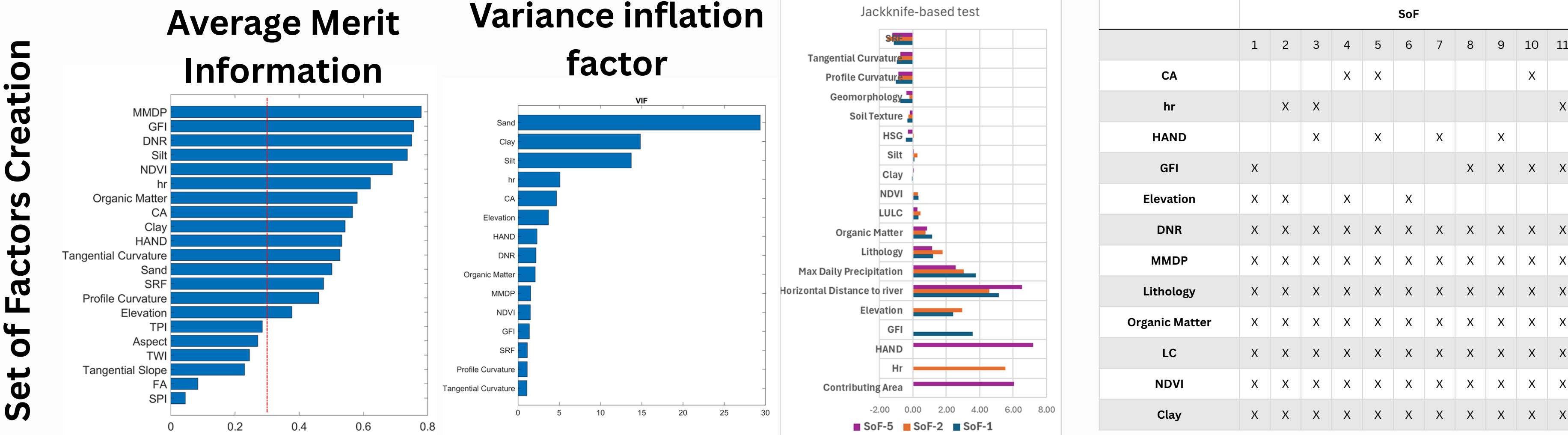
METHODOLOGY



Acknowledgments

This study was funded by the Partnership RETURN Extended Partnership and received funding from the European Union Next-Generation EU (National Recovery and Resilience Plan - NRRP, Mission 4, Component 2, Investment 1.3 - D.D. 1243 2/8/2022, PE0000005).

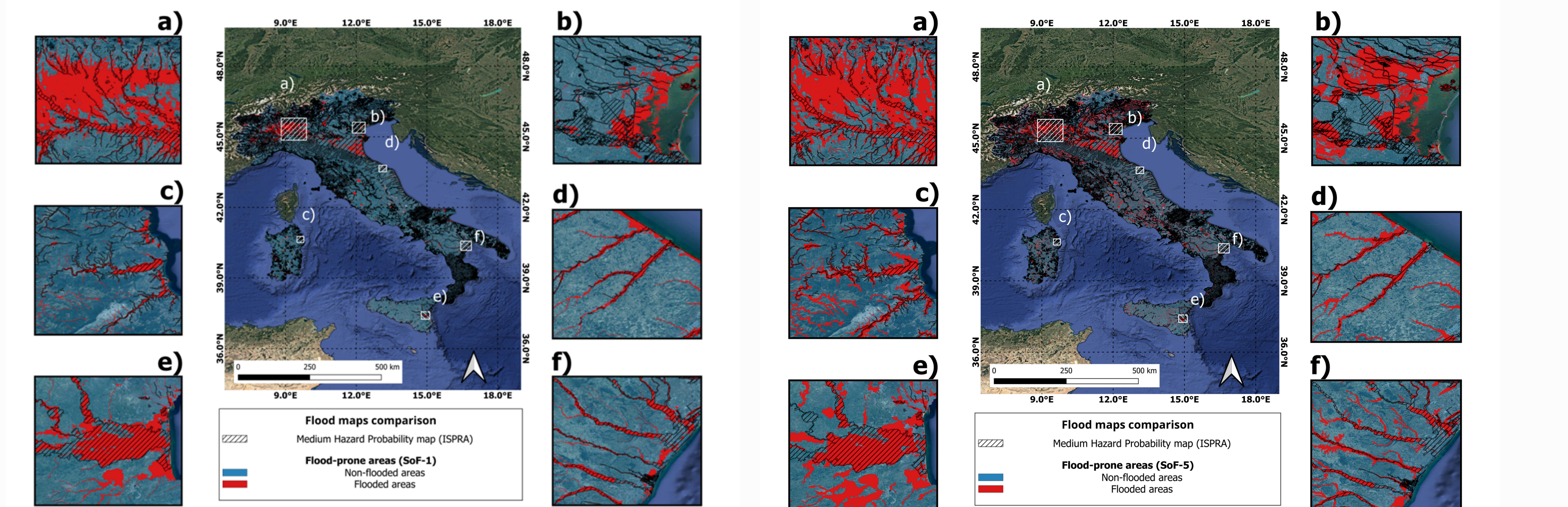
RESULTS



The selection of the most relevant factors was based on: AMI value greater than 0.3, ensuring relevance was detected by at least two of the three applied approaches, and VIF below 10 to discard multicollinearity. This process resulted in the exclusion of only the "Sand" factor. Subsequently, eleven different SoFs were created and evaluated using a Jackknife-based test to determine the influence of categorical factors. The analysis revealed that HSG, Curvature, Geomorphology, Soil Texture, and SRF had a negligible or negative impact on the results. Only the factors with a positive impact were used for the final prediction

SoF	AUC	Specificity	Sensitivity	Kappa	Accuracy	F1	Combined
1	0.94	0.86	0.88	0.74	0.87	0.87	0.26
2	0.95	0.88	0.88	0.76	0.88	0.88	0.24
3	0.96	0.9	0.89	0.79	0.89	0.89	0.21
4	0.95	0.88	0.88	0.76	0.88	0.88	0.24
5	0.96	0.9	0.89	0.79	0.89	0.89	0.21
6	0.93	0.85	0.87	0.71	0.86	0.86	0.29
7	0.94	0.87	0.86	0.73	0.87	0.86	0.27
8	0.93	0.85	0.87	0.72	0.86	0.86	0.28
9	0.95	0.88	0.87	0.75	0.88	0.88	0.25
10	0.96	0.9	0.88	0.78	0.89	0.89	0.22
11	0.96	0.9	0.88	0.78	0.89	0.89	0.22

SoF	Specificity			Sensitivity			Kappa			Accuracy			F1			Combined Error		
	LPH	MPH	HPH	LPH	MPH	HPH	LPH	MPH	HPH	LPH	MPH	HPH	LPH	MPH	HPH	LPH	MPH	HPH
1	0.63	0.52	0.36	0.8	0.86	0.92	0.34	0.34	0.3	0.77	0.8	0.82	0.86	0.88	0.89	0.57	0.62	0.73
2	0.51	0.4	0.26	0.79	0.84	0.9	0.26	0.24	0.19	0.73	0.75	0.77	0.82	0.85	0.86	0.7	0.76	0.83
3	0.52	0.43	0.29	0.82	0.87	0.93	0.33	0.32	0.26	0.74	0.76	0.77	0.82	0.84	0.86	0.66	0.7	0.78
4	0.5	0.4	0.26	0.79	0.84	0.9	0.26	0.24	0.19	0.73	0.75	0.77	0.82	0.84	0.86	0.71	0.76	0.84
5	0.52	0.43	0.29	0.82	0.87	0.93	0.34	0.32	0.26	0.74	0.76	0.76	0.82	0.84	0.85	0.66	0.7	0.79
6	0.58	0.47	0.31	0.78	0.84	0.9	0.27	0.26	0.23	0.75	0.79	0.81	0.84	0.87	0.89	0.63	0.69	0.78
7	0.58	0.47	0.31	0.81	0.87	0.92	0.36	0.34	0.27	0.76	0.78	0.79	0.84	0.86	0.87	0.61	0.66	0.77
8	0.62	0.51	0.35	0.8	0.86	0.92	0.34	0.34	0.3	0.77	0.8	0.82	0.85	0.88	0.89	0.58	0.63	0.73
9	0.57	0.47	0.31	0.82	0.87	0.92	0.36	0.35	0.28	0.76	0.78	0.79	0.84	0.86	0.87	0.61	0.66	0.76
10	0.54	0.43	0.3	0.82	0.87	0.92	0.34	0.32	0.26	0.75	0.76	0.77	0.83	0.85	0.86	0.65	0.7	0.78
11	0.53	0.42	0.29	0.82	0.87	0.92	0.34	0.31	0.25	0.74	0.75	0.76	0.83	0.84	0.85	0.66	0.71	0.79



Flood prone area Prediction using SoF-1 (The best performance refers to the Comparison)
Flood-prone area Prediction using SoF-5 (The best performance refers to the validation)

CONCLUSION

This study developed a dynamic flood susceptibility map for Italy using an optimized Random Forest (RF) model, fed by historical flood data (1920-2023) and satellite imagery. By incorporating importance analyses such as AMI, the study identified that a limited group of key geomorphic factors (like GFI) are the most critical predictors, surpassing the performance of models overloaded with less relevant variables.

Key strengths and findings include:

- A higher number of predictors does not always improve model performance, a feature selection based on coherence can help alancing the performance and efficiency. **A maximum of 10 factors** is recommended for accurate predictions.
- Satellite data effectively identified flooded areas, aligned well with hazard maps provided by ISPRA and proposed susceptibility levels in regions with limited or no prior information. However, data scarcity in certain regions constrained the model predictive capacity. Integrating multiple satellite data sources could help mitigate these limitations.
- The incorporation of a secondary verification step (comparison) utilising hazard maps facilitated the evaluation of the model generalisation capacity, thereby uncovering limitations that were not apparent during internal validation. This methodology improved the dependability of the outcomes and is advised as an auxiliary practice for future studies.
- Geomorphic features such as **GFI**, **DNR**, and **MMDP** consistently emerged as highly relevant factors for flood susceptibility. The analysis demonstrated that while certain factors are indispensable, those of low importance can be excluded to streamline models without sacrificing predictive accuracy.

The result is a cost-effective and robust methodology capable of accurately identifying flood-prone areas across large, diverse territories, offering a valuable tool for risk assessment that can be rapidly updated.

References

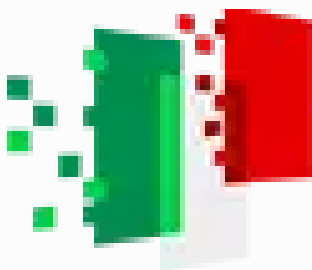
Tehrany et al.(2019). Identifying the essential flood conditioning factors for flood prone area mapping using machine learning techniques. CATENA.
Saavedra Navarro et al. (2025) Mapping Flood Susceptibility Using Random Forest Exploiting Satellite Observations and Geomorphic Features. ESS Open Archive.
Magnini et al.(2023). Geomorphic flood hazard mapping: From floodplain delineation to flood hazard characterization. Hydrological Sciences Journal.
Manfreda et al. (2015), Flood-Prone Areas Assessment Using Linear Binary Classifiers based on flood maps obtained from 1D and 2D hydraulic models. Natural Hazards.
Samela et al.(2017), Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. Advances in Water Resources.



Finanziato dall'Unione europea
NextGenerationEU



Ministero dell'Università e della Ricerca



Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

