A data-driven approach to flood hazard zonation: an application to Italy

Ivan Marchesini[(1)], Mauro Rossi[(1)], Paola Salvati[(1)], Marco Donnini[(1)], Simone Sterlacchini[(2)], Fausto Guzzetti[(1,3)]

(1) CNR IRPI, via della Madonna Alta 126, I-06128 Perugia, Italy

(2) CNR IGAG, piazza della Scienza 1, I-20126 Milano, Italy

(3) Dipartimento della Protezione Civile, via Vitorchiano 2, I-00189 Roma, Italy

e-mail: ivan.marchesini@irpi.cnr.it

Context

- Flood hazard and risk zoning are commonly implemented based on the results of hydro-dynamic flood routing and inundation modelling
- Hydro-dynamic flood routing and inundation models remain poorly suited to model very large areas at high resolution
- Reasons are not only (or no more) computational accuracy and efficiency but amount and quality of the information they need,
    - VHR DEM or LR hydrologically conditioned DEM, river cross-sections topography, hydro-meteorological data, boundary conditions, roughness values, ..

Context

- Alternative (or completion): data-driven models. Functional relationships between environmental variables and the presence or absence of inundations
- No standard data-driven approach exists to delineate flood-prone areas or to estimate water depth.
- At national scale, experiments exist using medium to low resolution data
  - Greece and and Canada (20 m × 20 m)
  - Slovakia (50 m × 50 m)
  - Portugal and the USA (90 m × 90 m)
  - China (10 km × 10 km)

Research questions

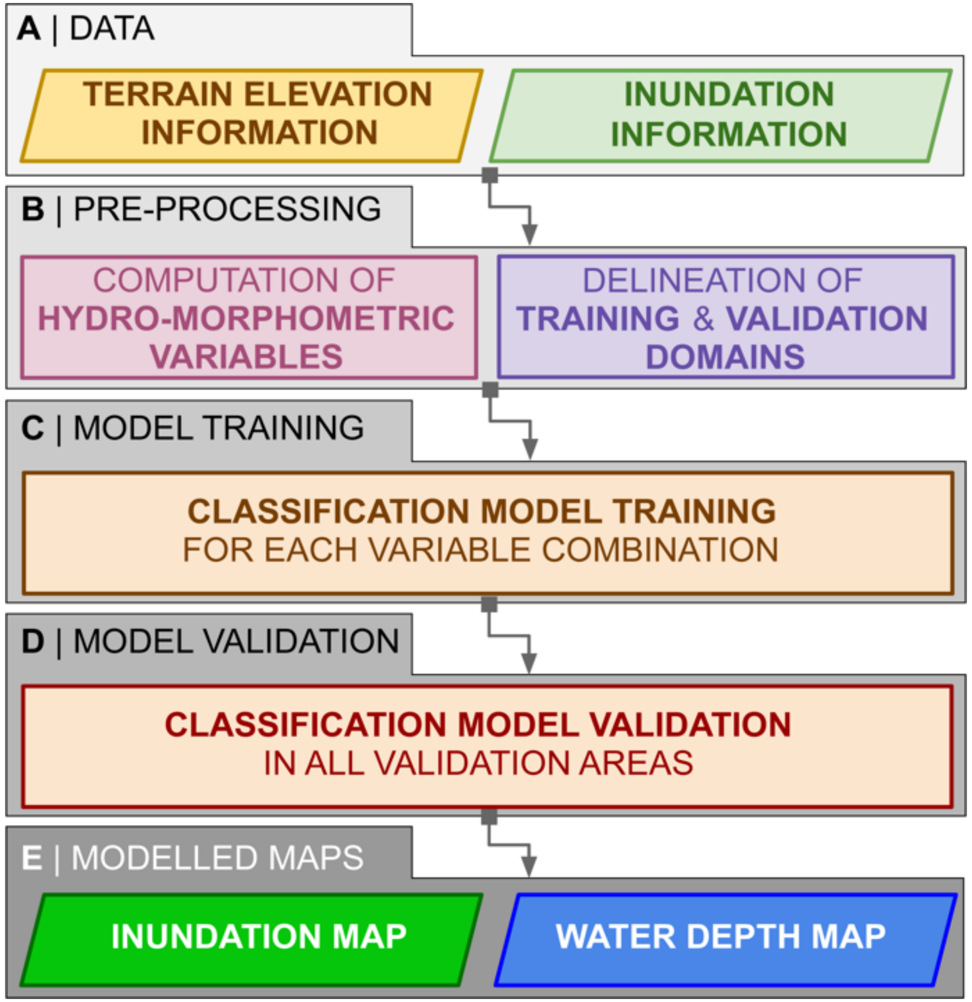
Some issues are still open about data-driven flood-hazard zonings:

- Is the performance homogeneous over large study areas (national scale)?
- What most influences the quality of the models?
- Do the models work adequately in predicting the location of sites with historical flood consequences?
- Can modelled flood-prone areas be used to estimate water depths, thus allowing for proper flood-hazard estimation?

We implemented the **procedure Flood-SHE** (Flood - Statistical Hazard Evaluation) **throughout Italy** to deal with these research questions

Marchesini, I., Salvati, P., Rossi, M., Donnini, M., Sterlacchini, S., & Guzzetti, F. (2021). Data-driven flood hazard zonation of Italy. *Journal of Environmental Management*, 294, 112986. https://doi.org/10.1016/j.jenvman.2021.112986

Design/methods: Flood-SHE procedure

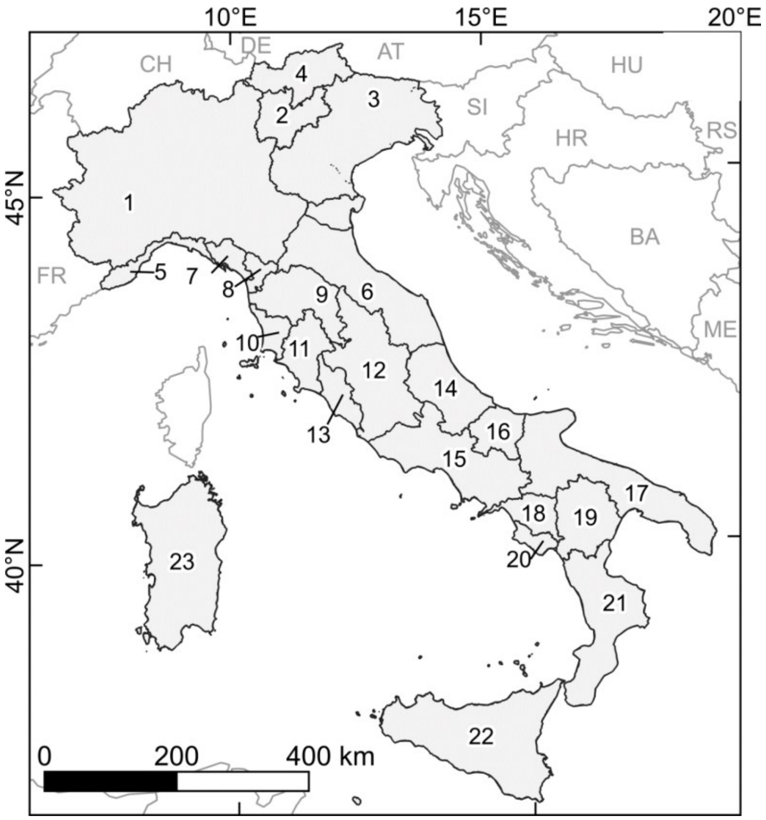


A. First step consists of the acquisition of the relevant data. **DEM** and inundation information ("**Flood Prone Areas**" - **FPAs**)
B. Second step includes (i) preparing **covariates** and (ii) delineating **training and validation domains**
C. The third step consists in **training models**. We used **logistic regression** and multiple combinations of covariates
D. The fourth step validates the trained models. **Validation** allows to select the **optimal combinations of covariates**
E. The last step prepares "**Potentially Inundated Areas**" (**PIAs**) maps, which are used to prepare **water-depth** maps

Design/methods: Data

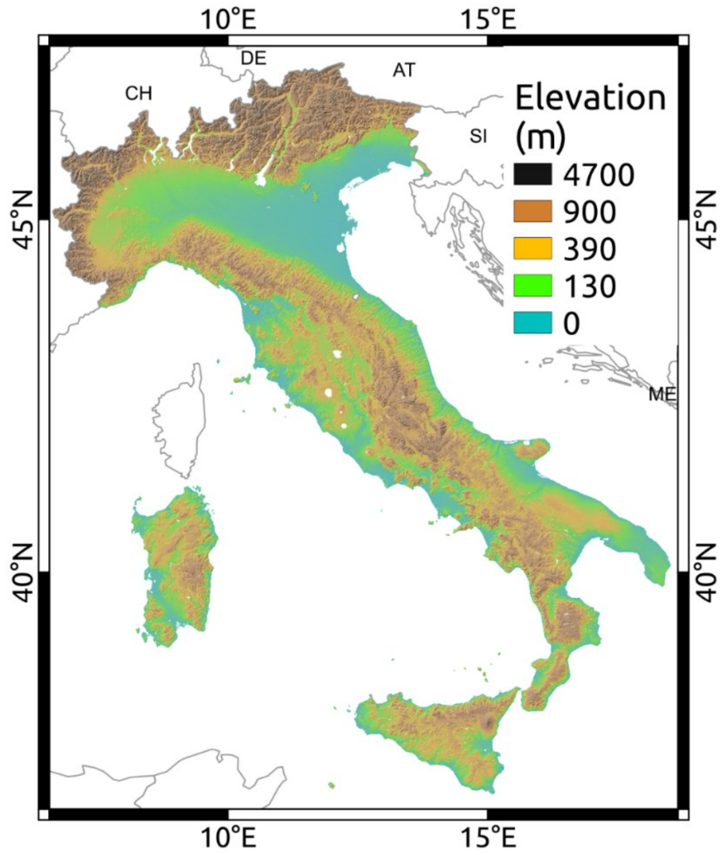
23 River Basin Authorities (**RBAs**)

Total study area: ~301,300 km^2



TinItaly **DEM** (Tarquini et al., 2012)

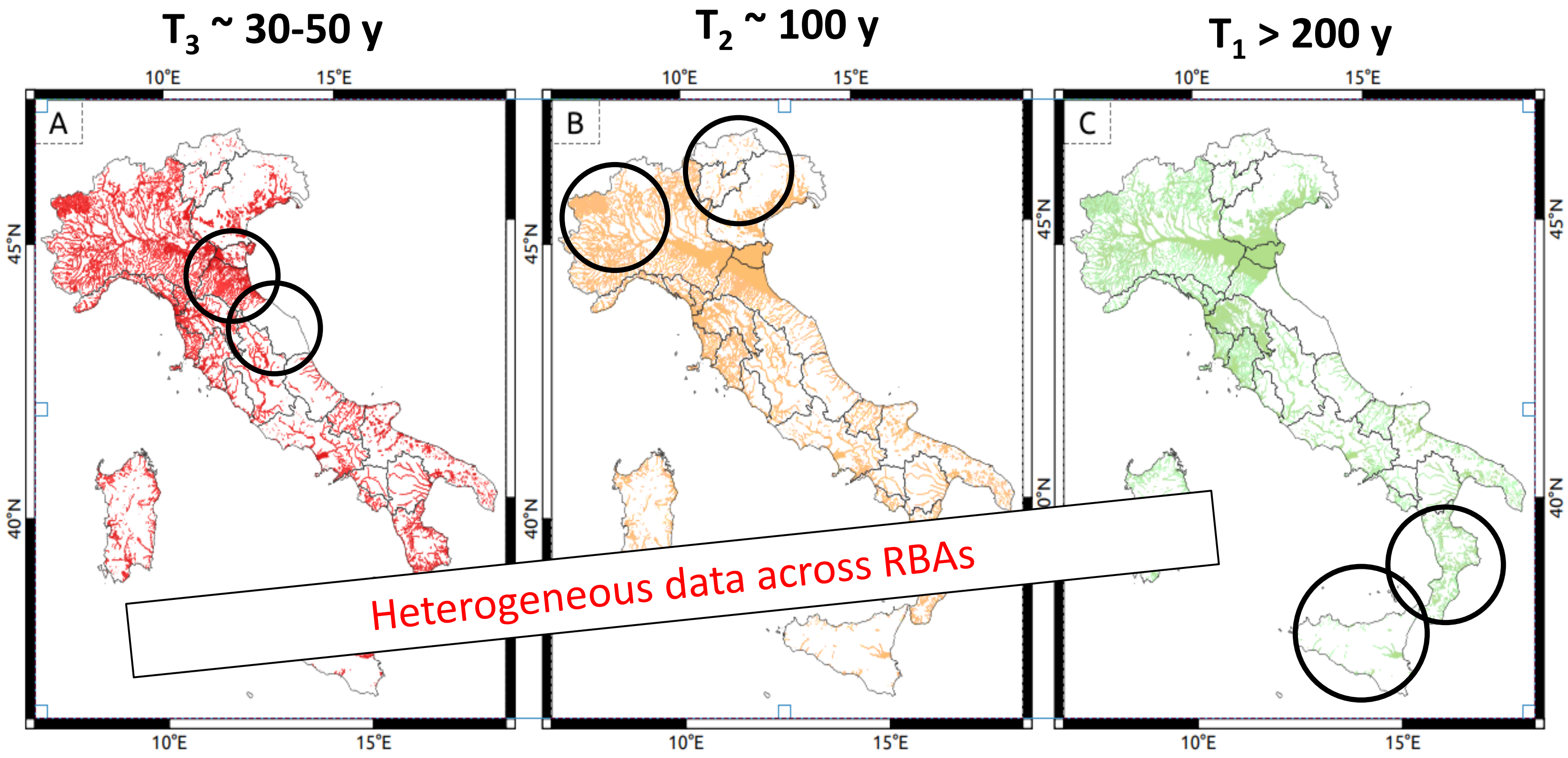
Resolution: 10x10 m



Design/methods: Data

Flood prone areas **(FPAs)**

Polygons delineated through hydro-dynamic flood routing and inundation modelling for different return periods (T [years]) and using different approaches in the different RBAs



ISPRA (2015): http://www.sinanet.isprambiente.it

Design/methods: Pre-processing (Conditioning Factors)

Seven (7) Hydro-Morphometric conditioning factors:

The **planimetric distance L**, and the **elevation difference H**, between each grid cell and the **nearest stream**, measured along the hydrologic flow direction (L and H were always included in the models)

the **planimetric distance D**, and the **elevation difference V**, between each grid cell and the **basin outlet** measured along the hydrologic flow direction

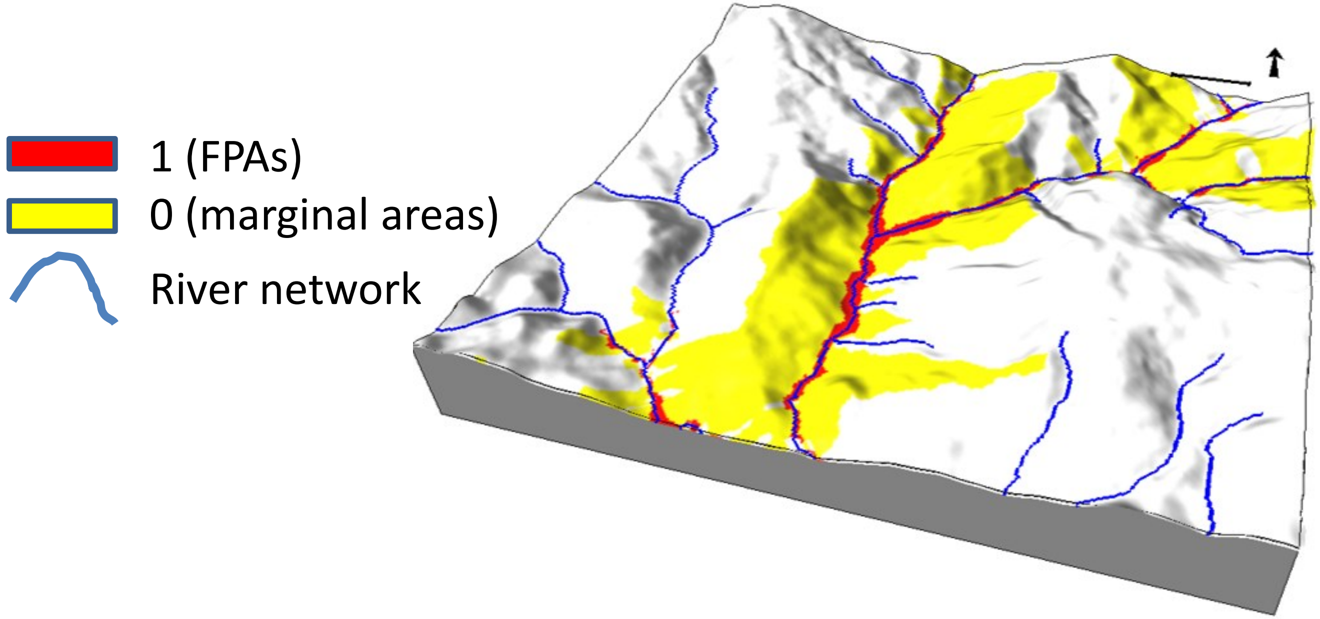
the **local terrain slope S**, calculated on a 3 × 3-cell kernel

the **order of the stream channel O** into which each grid cell drains, using the Shreve (1966) ordering scheme.

the **local terrain roughness G**, measured by the mean length of the unit normal vectors to the topographic surface

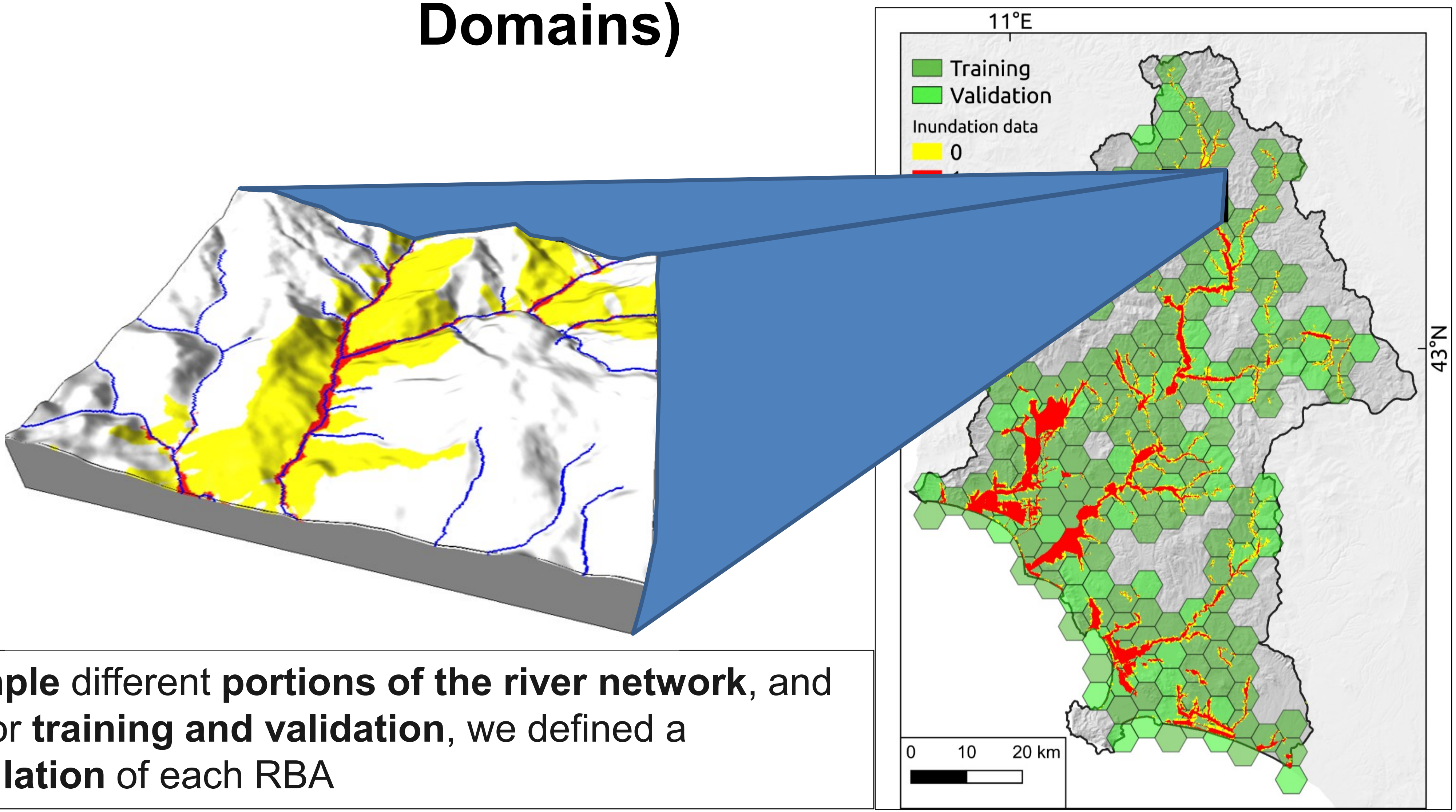
Design/methods: Pre-processing (Marginal Areas)

- **Cells** located **inside the FPAs** were **labelled as 1** (Flood Prone Area)



- Among the **remaining cells**:
  - those **hydrologically draining into the FPAs** were considered **“Marginal”** and **labeled to 0**
  - all the **other cells** were **not used** for training the models

Design/methods: Pre-processing (Training and Validation Domains)



To **randomly sample** different **portions of the river network**, and thus have areas for **training and validation**, we defined a **hexagonal tessellation** of each RBA

Design/methods: Models training

Training domain:

random **70% of the hexagons** in each RBA

- in each one of the **23 RBAs**
- for **32** different **covariates combinations**
- for **3** different **return periods**

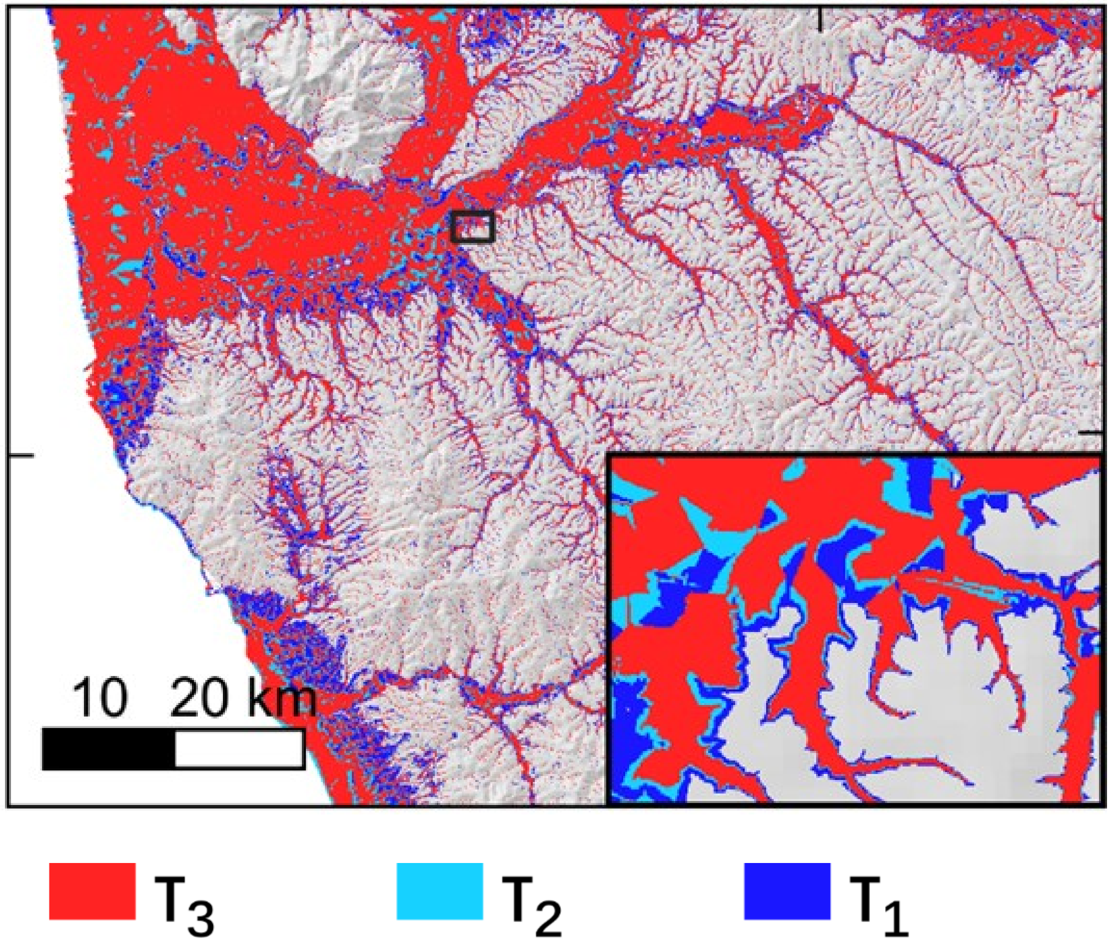
Overall we generated
2,208
logistic models and
probabilistic
flood susceptibility maps

- Susceptibility maps were transformed to **binary maps** (**0/1**, Marginal/Flood-Prone areas)
- Binary maps were **combined**, obtaining **736 PIAs** maps (**0/1/2/3**, Marginal/Flood-Prone areas for different return periods T1, T2, T3)

Design/methods: Models training

- **Example**

Portion and further enlargement of the **PIA** «Potentially Inundated Areas map» generated for **RBA #9** and **LHGO covariates combination**



T_1: return period > 200 years

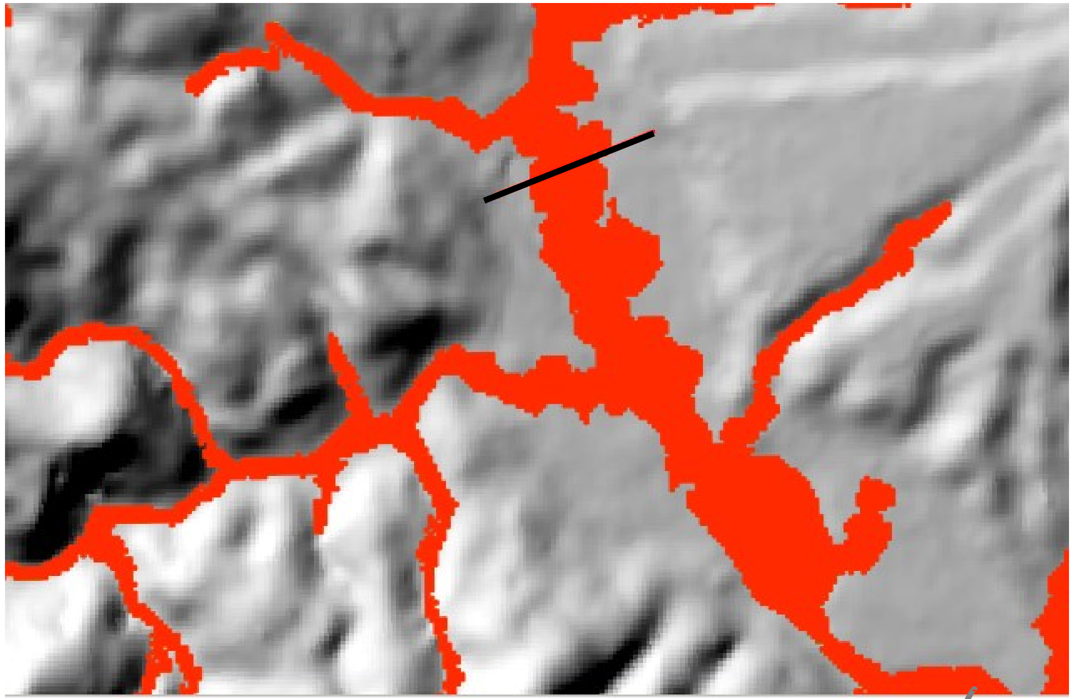
T_2: return period ~ 100 years

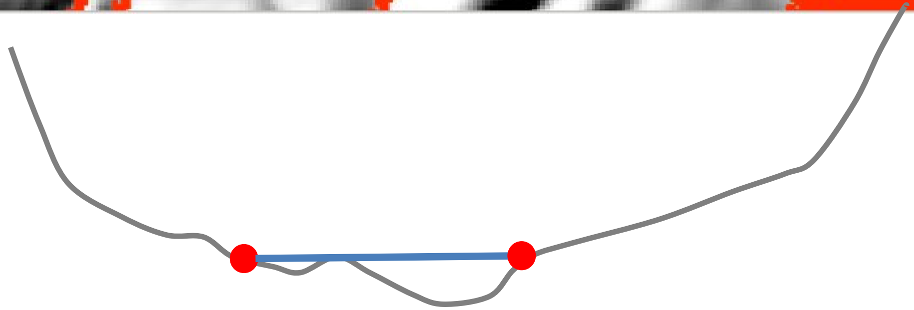
T_3: return period ~ 30-50 years

Design/methods: water depth

- **Water depth** was calculated for the **three return periods**.

- **Simple geometrical approach:**
  - Interpolation of the terrain elevation measured at the boundaries of the PIAs
  - Calculation of the difference with DEM





Design/methods: Models Validation

Validation domain:
random **30% of the hexagons** in each RBA

- For **each one of the 736 PIA** we generated:
  - **32** different random **validation sets**

Overall we generated
23,552
Validation sets

- **Inside each validation hexagon** we calculated the **True Rate (TR)** for a **four classes classification model**

$$TR = \left(\sum_{c=0}^{c=3} \frac{TP_c}{TP_c + FN_c} \right) \Big/ 4$$

Design/methods: Performance index

- We used the **Coefficient of Variation (CV)** to evaluate the performance on the total number of hexagons of each validation set

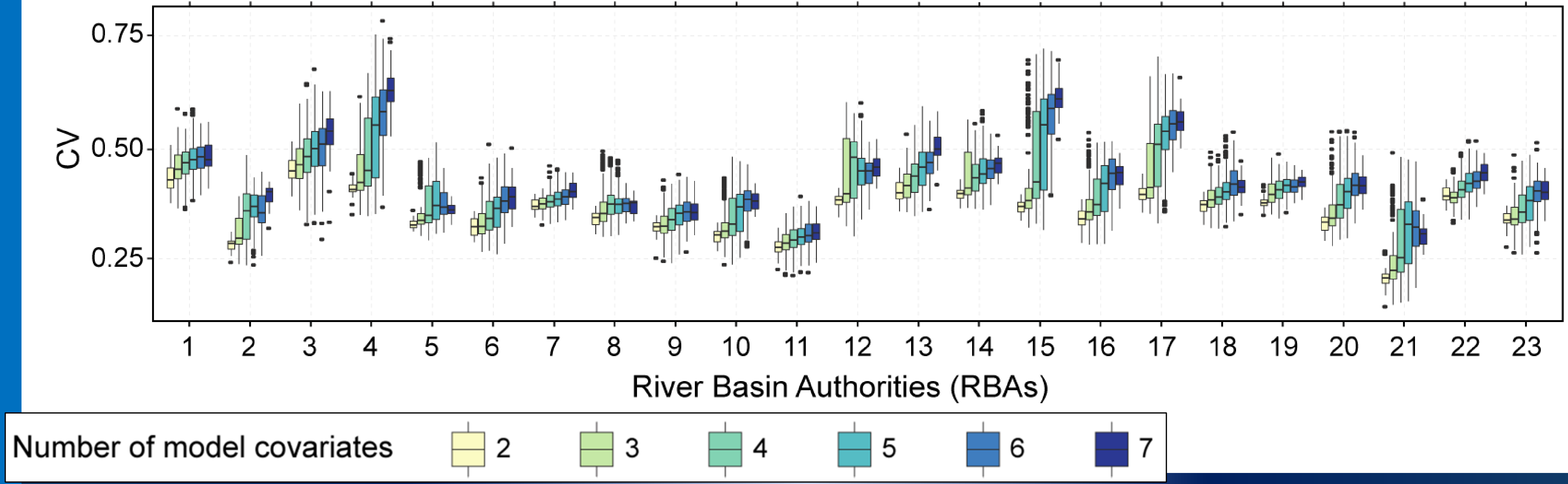
$$CV = TR_{\sigma}/TR_{\mu}$$

- where $\boldsymbol{TR_{\mu}}$ and $\boldsymbol{TR_{\sigma}}$ are the **mean** and the **standard deviation** of the **TR** computed in the **different hexagons** of the validation set.
- $\boldsymbol{TR_{\mu}}$ measures the **average performance** and $\boldsymbol{TR_{\sigma}}$ **the spatial homogeneity of the prediction** across all the hexagons of the validation set
- **Low CV values indicate better performance** and imply an appropriate balance between accuracy and prediction precision in the validation set

Results

- Research question: <u>Is the performance of the data-driven models homogeneous over large study areas (national scale)?</u>

- CV metric largely variates across the different RBAs



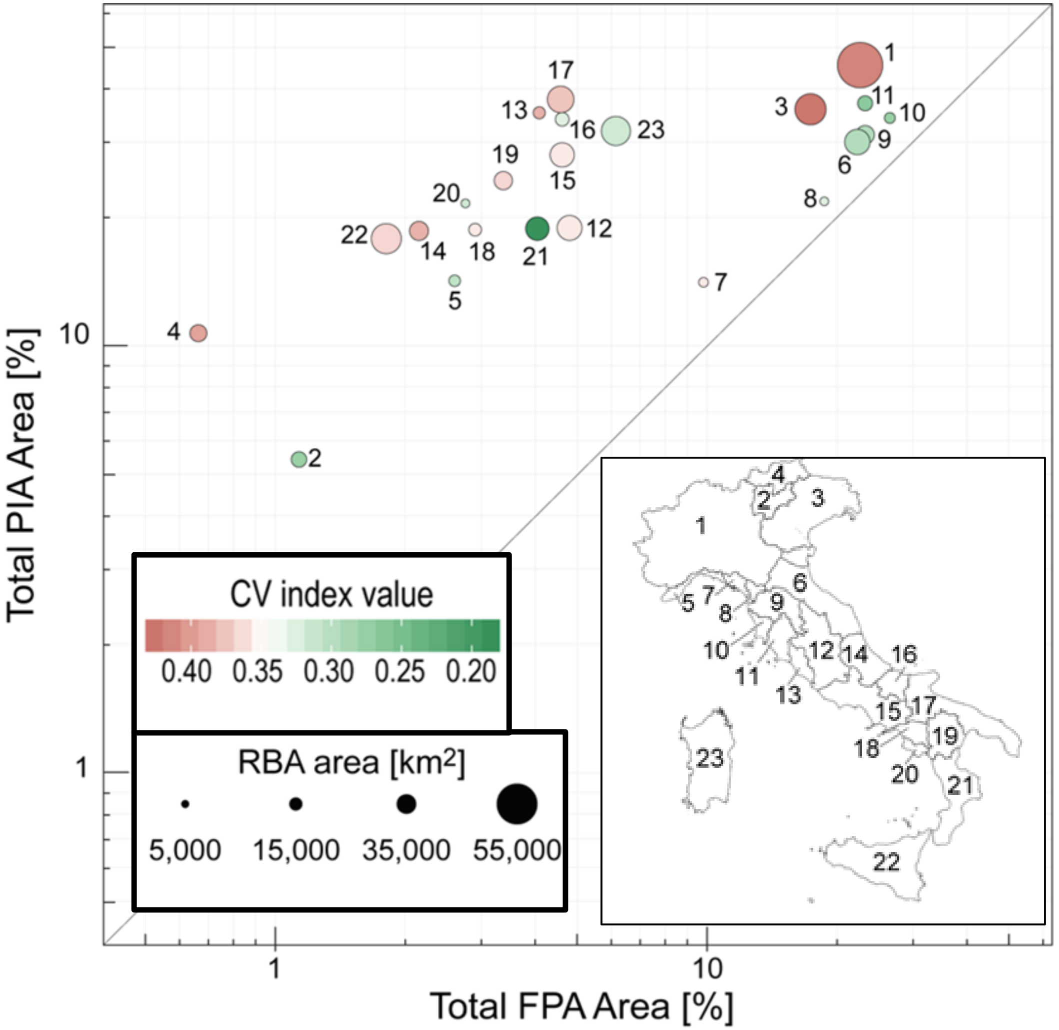
Results

- Research question: What most influences the quality of the models?

- Quality of the existing FPA
- Morphological homogeneity (small areas)

There is a considerable increment in the extent of the PIAs compared to the corresponding FPAs.

- Mean FPAs area is 9.13% of RBA area
- Mean PIAs area is 24.67% of RBA area



Results

- Research question: Do the models work adequately in predicting the location of sites with historical flood consequences?
- Reply: Yes.
  - In Italy, a catalogue of flood events with damage to the population exists (Salvati et al., 2018).
  - We confronted a portion of a catalogue (206 events from 1920 to 2019) with the extent of the FPAs and the PIAs.
  - We found that **65.5% were inside the FPAs**, and **94.2% were inside the PIAs**.

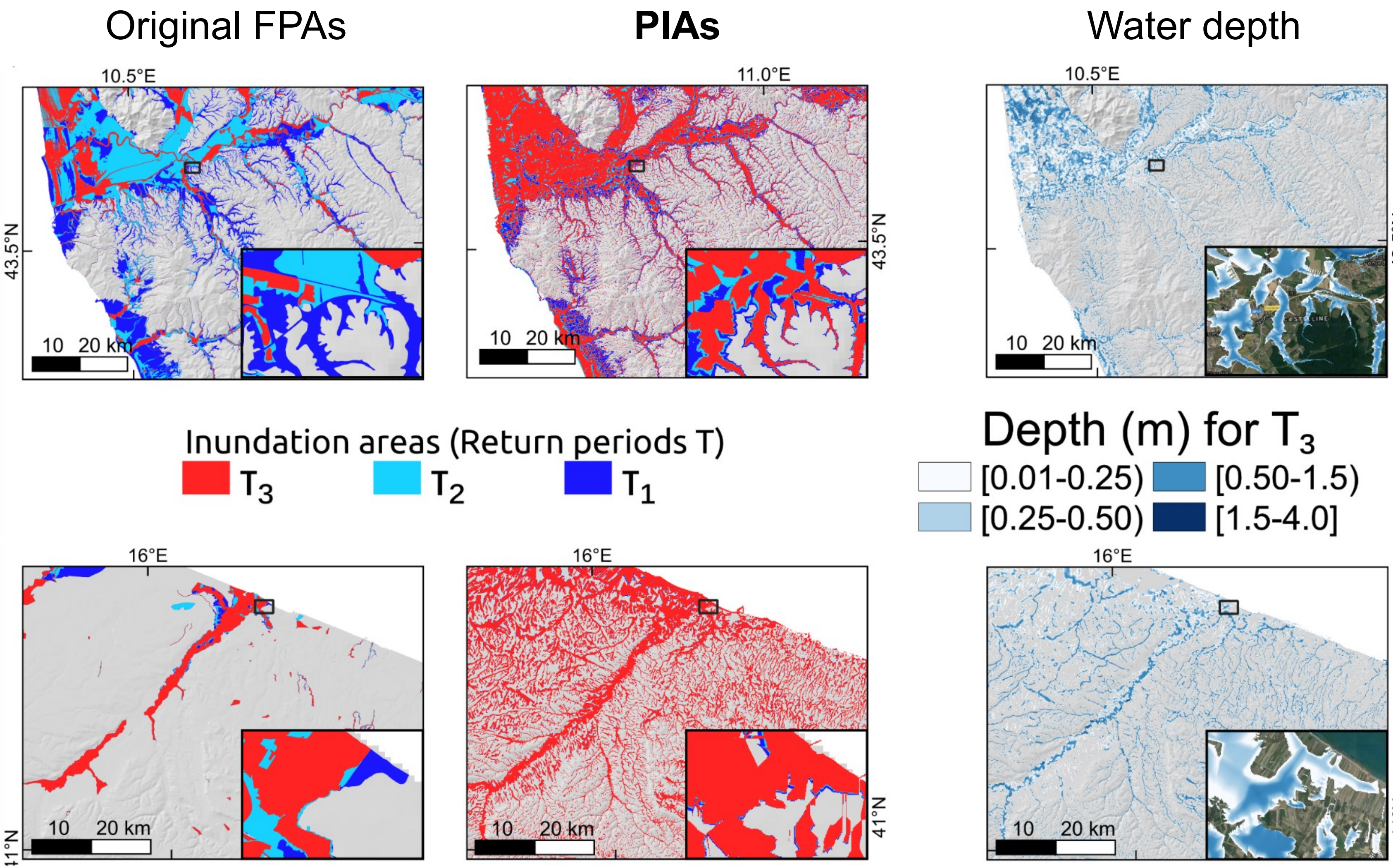
Results

- Research question: Can modelled flood-prone areas be used to estimate water depths, thus allowing for proper flood-hazard estimation?
- Reply: Yes
  - we compared average Flood-SHE water depths to similar maps prepared for Europe by Alfieri et al. (2015) at a 100 m × 100 m ground resolution and using a hydro-dynamic model

	TR_3	TR_2	TR_1
Alfieri (2015)	1.21 m	1.41 m	1.48 m
Flood-SHE	0.67 m	1.33 m	1.55 m

  - In the common areas, the two estimates are similar but we note that the PIAs maps cover a larger area

Results: Examples of the products



Conclusions

- Flood-SHE procedure is entirely based on **Open Source Software** and was used for the **data-driven based flood-prone area zoning of Italy,** at 10x10m spatial resolution
- In the 23 RBAs in Italy, our **data-driven approach has outlined larger** to much larger **areas as potentially subject of being inundated when compared to** the corresponding **flood zonings prepared by the hydro-dynamic flood routing** and inundation models
- We expect the **PIAs zonings prepared in this work** can be use **to complement** and **to expand the pre-existing FPA zonings** where they are not available or are poor

Thank You for your attention

ELSEVIER

Journal of Environmental Management

Volume 294, 15 September 2021, 112986

Data-driven flood hazard zonation of Italy

Ivan Marchesini [a], Paola Salvati [a], Mauro Rossi [a], Marco Donnini [a], Simone Sterlacchini [b], Fausto Guzzetti [a, c]

https://doi.org/10.1016/j.jenvman.2021.112986