Mapping public investments targeting hydrologic risk mitigation in Italy*

This version: April 11, 2025

Antonella Fabrizio[1], Michele Penza[1], Sedric Zucchiatti[1]

[1]Prometeia S.p.A., Piazza Trento e Trieste, 3, Bologna, 40137, Italy.

Contributing authors: antonella.fabrizio@prometeia.com; michele.penza@prometeia.com; sedric.zucchiatti@prometeia.com;

Abstract

The increasing frequency and severity of extreme weather events and floods highlight the urgent need to monitor hydrologic risk mitigation investments and understand their potential effectiveness. However, such efforts are limited by the absence of integrated databases combining diverse and essential data sources. Focusing on the Italian context, this paper presents a newly constructed dataset that merges information on national hydrography, public investments in flood risk mitigation, and historical flood events. Organized as a shapefile at the municipality–river pair level, the dataset provides detailed records of both investment allocations and flood occurrences. We describe the dataset's construction and offer preliminary descriptive insights, while outlining possible extensions, its value for future analyses, and relevance to policy design in the face of increasing hydrologic risks.

Keywords: Flood, Mitigation, Climate adaptation, Hydrography, Public Investments, Italy

1 Introduction

Climate change is expected to significantly increase the frequency and severity of climate-related extreme events, including floods (Calvin et al., 2023). These events are likely to impose increasing costs on the economy, individuals, and companies, affecting both the productive capacity and the overall economic stability (Fatica et al., 2024; Clò et al., 2024; Benincasa et al., 2024).

The rising frequency of hydrologic risks underscores the need for effective mitigation and adaptation measures aimed at minimizing the adverse consequences of climate change. Among these measures, state-level interventions, such as investments in infrastructure designed to reduce

*Funded by the European Union – Next Generation EU - PNRR - MISSION 4 COMPONENT 2 in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP J53C22003140001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the affiliated institutions.

hydrologic risks, can play a crucial role in enhancing economic resilience and safeguarding productive assets.

Italy represents a particularly relevant case for studying the impacts of hydrologic risk mitigation investments. The country is highly exposed to the effects of climate change, with its diverse topography and extensive river network, making it especially prone to flood-related hazards (Trigila et al., 2021). Moreover, the Italian productive system is primarily composed of micro, small, and medium enterprises, which tend to be more vulnerable to extreme weather events and often lack adequate insurance coverage against such risks (ECB and EIOPA, 2024). These factors make it crucial to assess the effectiveness of public investments in flood risk mitigation in Italy, as such initiatives can provide valuable protection for businesses and local economies.

Evaluating the effectiveness of public investments in mitigating hydrologic risk requires information on public expenditures, flood risk levels, and recorded flood events.

Focusing on public investments, a key initiative is the *Repertorio Nazionale degli interventi per la Difesa del Suolo* (the National Repository of Soil Defense Interventions, hereinafter ReNDiS) (ISPRA, n.d.). This is a database compiled by the *Istituto Superiore per la Protezione e la Ricerca Ambientale* (Italian Institute for Environmental Protection and Research, hereinafter ISPRA) that tracks interventions aimed at mitigating hydrologic risk across Italy. While ReNDiS is well aligned with the scope of this study, it has limitations, as it mainly includes projects funded by the *Ministero dell'Ambiente e della Sicurezza Energetica* (Ministry of Environment and Energy Security). On the contrary, OpenCUP (DiPE and MEF, 2025), an open-access portal provided by the *Dipartimento per la programmazione e il coordinamento della politica economica* (Department for the Planning and Coordination of Economic Policy, hereinafter DiPE) of the *Presidenza del Consiglio dei Ministri* (Presidency of the Council of Ministers), accounts for public interventions financed through local, national, European, and even private funds. Nevertheless, it includes investments in a wide variety of topics, hence the necessity of a specific filtering before using it for the purposes of this study.

In terms of hydrologic risk mapping, a valuable source of information is the *Mosaicatura ISPRA delle aree a pericolosità idraulica* (ISPRA mosaic of hydraulic risk areas) (ISPRA, 2020b), mapped by the *Autorità di Bacino* (River Basin Authorities). Last updated in 2020, the ISPRA Mosaic identifies high, medium, and low flood risk zones in Italy. The classification is based on three scenario analyses that simulate extreme events of increasing severity, identifying areas that could be inundated under each scenario.

As regards the exposure to hydrologic risk, it is essential to consider the entire network of watercourses flowing within the Italian territory. To this end, multiple data sources exist. The dataset *Reticolo Idrografico Nazionale* (National Hydrographic Network) (ISPRA, 2020a), compiled by ISPRA, provides a representation of Italian hydrography limited to natural watercourses, available in shapefile format at a 1:250,000 scale. In contrast, two additional sources also include artificial canals. Firstly, Humanitarian OpenStreetMap Team (2025), within the Humanitarian Data Exchange project, offers the monthly updated Italy Waterways map. The second source is instead a map developed within the framework of the *Strati Prioritari di Interesse Nazionale*[1] (Priority layers of National Interest, hereinafter DBPrior10k), as detailed in Section 2.1. The choice of a hydrographic network which includes artificial canals, as well as natural rivers, is crucial for the purposes of this study, since both types of watercourse are affected by flood events.

Finally, regarding flood occurrences in Italy, there are various potential sources. ISPRA's annual reports on climate events (Berti and Lucarini, 2024) - which we detail in Section 2.3 - provide information on actual flood events, reporting the affected river basin, but not the

[1] For more information: https://github.com/ondata/DBPrior10K.

impacted municipality. Differently, the dataset by Gatto et al. (2023) collects all hydrogeological disasters (both landslides and floods) that occurred in Italy between 2013 and 2022, which caused such severe impacts as to require the declaration of national-level emergencies. Although the dataset provides detailed data at the municipal level, it is not possible to distinguish between landslides and floods, nor there is any information about the rivers affected by such events. In addition, Copernicus Emergency Management Service (n.d.) offers mapping services in cases of natural hazards. The platform documents the extent of flooded areas for selected major flood events worldwide. However, a key limitation is that it does not provide a comprehensive inventory of all flood events, instead focusing only on the most significant ones. Despite this, it can be valuable for case studies, as it offers precise spatial data on affected areas. Similarly, the institute *Popolazione a Rischio da Frana e da Inondazione in Italia* (Population at risk of landslides and floods in Italy, hereinafter Polaris) (CNR, 2024), part of the Consiglio Nazionale delle Ricerche (National Council of Research, CNR), publishes detailed descriptions of selected flood events, providing information on the most affected municipalities and rivers, although without offering precise data on the extent of flooded areas.

Despite the availability of valuable data sources, analyzing the distribution and effectiveness of flood risk mitigation and adaptation measures in Italy remains challenging. This is primarily due to the absence of a comprehensive database that integrates all the necessary information for such assessments.

To address this limitation, our work makes a key contribution by developing a novel dataset that integrates multiple information sources essential for evaluating the effectiveness of public investments in mitigating flood risks. Specifically, we combine data on the Italian hydrography, records of public investments targeted at mitigating hydrologic risks, and historical data on flood events across Italy. Among the available possibilities in terms of hydrography, we opted for DBPrior10k, because it reports both natural and artificial watercourses. As for public interventions, we selected OpenCUP: despite encompassing a much broader scope than the one required for this study, it offers the flexibility to be filtered according to our specific needs. Finally, as regards flood events, we think that ISPRA's annual reports on climate events are the most suitable option for our case: while municipal-level information is not readily available, we can reconstruct it using hydrography data.

The resulting dataset provides a unique spatial representation of flood mitigation investments at the river-municipality level together with their evolution over time.

The construction of this dataset represents an initial step in a broader project aimed at informing policy discussions on the benefits and cost-effectiveness of public investments in hydrologic risk management, as well as identifying potential investment gaps that require further attention.

We intend to use this dataset to conduct additional analyses in future work. One possible extension involves integrating data on extreme rainfall events in Italy to assess whether public investments in hydrologic risk mitigation effectively reduce the likelihood of flooding. Another promising direction would be to combine this foundational dataset with firm-level data to evaluate whether, in the event of a flood, firms located in areas with greater public investment experience less severe consequences.

The remainder of the paper is structured as follows. Section 2 describes the data sources, while Section 3 details the methodology used to construct the dataset. Section 4 presents the final dataset, while Section 5 presents the next steps and possible use cases of the dataset.

2 Data

We compile a novel dataset by integrating multiple sources, including Italian hydrography data, records of investments in flood risk mitigation, and historical flood events data in Italy. In the following subsections, we introduce each data source individually.

2.1 Hydrography

We primarily collect data on the Italian hydrography network from DBPrior10k[2]. This project provides comprehensive coverage of road and rail networks, river maps, and administrative boundaries in shapefile format at a 1:10,000 scale across the Italian territory.

Focusing on watercourse data, DBPrior10K provides 1,410,450 observations (river geometries) spanning both the primary and secondary river network as well as artificial canals. Avoiding a focus on only the primary river network is essential as often flood events are associated to secondary networks and artificial canals.

The DBPrior10K river network provides information about river geometries, names, and relevant attributes such as whether the river is natural or artificial and the type of water element (e.g., river, stream, canal).

Several river geometries lack a recorded name, posing a challenge for our analysis. Without a river name, it becomes impossible to associate data on public investments made and flood events occurred to the river geometry. Furthermore, unnamed river geometries are largely concentrated in specific regions — namely, Piedmont, Aosta Valley, and Liguria — likely due to the dataset primarily leveraging data from sub-national entities such as the River Basin Authorities. This uneven distribution may introduce territorial biases in the analysis.

To address these issues, we supplement the DBPrior10K hydrography data with additional sources: the historical DBPrior10K for Piedmont (Regione Piemonte, 2009), the regional hydrography for Liguria (Regione Liguria, 2025), and the ISPRA National Hydrographic Network dataset for Aosta Valley (ISPRA, 2020a).

2.2 Public Investments

Data on public investments aimed at mitigating hydrologic risks is sourced from OpenCUP (DiPE and MEF, 2025), an open-access portal provided by the Department for the Planning and Coordination of Economic Policy of the Presidency of the Council of Ministers. The platform aims to enhance transparency and accessibility of information on public investments and their implementation status.

The platform provides the data divided into four datasets: *Progetti* (Projects), *Localizzazione* (Localization), *Soggetti* (Subjects), and *Fonti di Copertura* (Funding Sources). The main communication key between these datasets is the *Codice Unico di Progetto* (Unique Project Code Identifier, or CUP), a 15-character alphanumeric string[3]. This identification code is mandatory under the *Sistema di Monitoraggio degli Investimenti Pubblici* (Public Investments Monitoring System) and was introduced by Law No. 3/2003.

The Projects dataset contains detailed information on each investment, including the associated CUP, the responsible public body, the year of the investment decision, the project's current status, total cost and funding, a detailed intervention description, and various investment categorizations. The Localization dataset reports one or more observations for each CUP, with

[2]For more information: https://github.com/ondata/DBPrior10K.

[3]Only the dataset Subjects does not report information on the CUP. Hence, it can be matched to the Projects dataset by using the *partita IVA* (VAT code) of the public body in charge of the project.

information on the region, province, and municipality in which the investment was made, and their relative codes. The Subjects dataset provides additional details regarding the responsible public bodies, while the Funding Sources dataset reports whether the project used EU, national, regional, provincial, municipal, private, or other funding.

Using the CUP as communication key, we integrate the Projects dataset with the Localization dataset to obtain information on the geographical location of investments.

For this research, the reference time window spans from 2000 to 2024. As of the time of writing this report, the total number of projects since 2000 amounts to 10,341,570, with a total allocation of more than 3,099 billion euros. The OpenCUP portal categorizes projects into several classifications: Public Works, Purchase of Goods, Grants, Services, Training Courses, Business Incentives, Financial Instruments, and Research Projects. Our analysis focused exclusively on Public Works, reducing the total number of projects to 1,599,288 and the total allocated funds to approximately 1,595 billion euros.

2.3 Flood events

We gather information on past flood events in Italy from ISPRA's annual reports on climate events (Berti and Lucarini, 2024). These reports primarily rely on technical assessments from ISPRA and *Agenzie Regionali per la Protezione Ambientale* (Regional Environmental Protection Agencies). Specifically, we extract data on flood events that occurred between 2003 and 2023.

For the period under consideration, the dataset organizes information on flood events into three distinct tables. The first table outlines the key rainfall characteristics of each flood event (such as duration, total cumulated precipitation, and maximum daily precipitation), along with details on the affected areas. While it lists the impacted provinces, municipality-level data is generally unavailable. However, the table often includes information on the affected river basins, which is crucial to our analysis, as it enables us to accurately link flood events to individual observations in our municipality-river database. The second table focuses on the socio-economic impact of flood events, detailing the number of injuries or fatalities, the financial resources required for recovery, and the related legislative measures. The third table documents soil effects and provides damage estimates relative to GDP.

3 Methodology

We now illustrate the methodology employed to integrate the three main data sources presented in Section 2 into a single dataset.

3.1 Hydrography cleaning

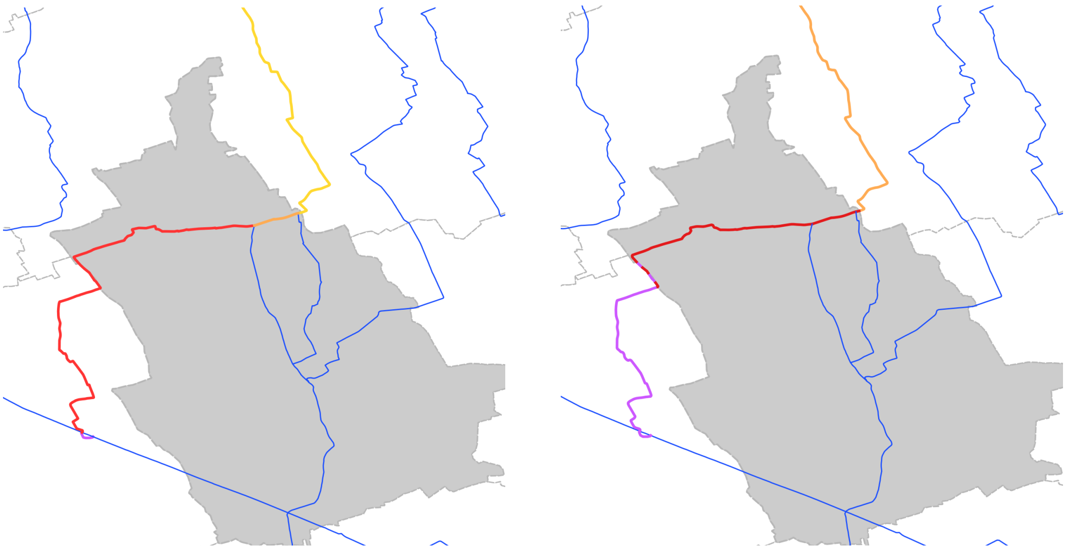
As already mentioned in Section 2.1, we employ the DBPrior10K hydrography integrated with regional hydrographies for Piedmont, Liguria, and Aosta Valley to overcome issues with missing river names in these regions.

By integrating these datasets, we obtain a rich and detailed hydrography which however requires an extensive cleaning process to ensure consistency, usability, and avoid duplicated geometries. From the integrated sample of 1,721,778 watercourse observations, we first remove all geometries located in Piedmont, Liguria, and Aosta Valley associated to the primary DBPrior10K dataset to avoid duplicates with the entries provided by the supplementary regional datasets. Further, we remove entries with missing or undefined river names. For the remaining river geometries, we clean their names by standardizing their format and eliminating general terms that could obscure river identification or hinder the association to the investment data. The cleaned hydrography contains 660,491 geometries.

3.2 Construction of the Municipality-River dataset

Even after cleaning, the hydrography shapefile remains unsuitable for integration with public investment and flood event datasets. This is because each river geometry is defined based on the characteristics of the river network, and rivers are often segmented into distinct geometries at intersection points with other rivers.

To address this issue, we restructure the cleaned hydrography using the shapefile of Italian municipalities as of 1st January 2024, provided by ISTAT (the Italian National Institute of Statistics). The restructuring process involves two steps: first, splitting river geometries that span multiple municipalities at municipal borders, and second, consolidating river geometries with the same name within each municipality into a single entity. In Figure 1 we contrast the cleaned hydrography with the resulting municipality-river shapefile.



(a) Hydrography (b) Municipality-river

Fig. 1: Comparison between hydrography and municipality-river shapefiles. In both panels, hydrography is depicted in blue, while the river of interest is shown in multiple colors (yellow, orange, red, and violet), with each color representing distinct river geometries. Municipality boundaries are indicated by dashed gray lines, with the municipality of interest shaded in gray. In the hydrography shapefile (panel (a)), the river is segmented at intersections with other rivers, resulting in distinct geometries. In contrast, in the municipality-river shapefile (panel (b)), segmentation occurs where the river crosses municipality boundaries, merging all sections of the same river within a municipality into a single geometry. Notably, a single geometry in the municipality-river dataset may consist of non-contiguous river segments, as seen with the red and violet geometries.

This process generates the foundational dataset for our project: a municipality-river pair-level hydrographic network. This dataset comprises 110,015 unique municipality-river combinations across 7,458 (out of 7,899) municipalities, representing distinct river segments, each identified by a unique name within the corresponding municipality. It is important to note that for municipalities where no river could be associated, data on public investment in hydrologic risk mitigation and flood events will remain unmatched all throughout.

To integrate public investments and flood events data while maintaining their timeline, we convert the municipality-river dataset into a panel format. This process generates multiple observations for each municipality-river pair, with each entry corresponding to a specific year from 2000 to 2024.

3.3 Public Investments cleaning

While the OpenCUP database, described in Section 2.2, serves as a comprehensive repository of project investments, its scope extends beyond hydrologic risk mitigation.

To refine the investment dataset in line with our research objectives, we explore the variables *Area Intervento* (Intervention Area), *Settore Intervento* (Intervention Sector), *Sottosettore Intervento* (Intervention Sub-sector), and *Categoria Intervento* (Investment Category). This allows us to further filter the database and narrow the selection to investments specifically targeting hydrologic risk mitigation.

Within the Intervention Area, we filter for Environment and Energy, excluding other out-of-scope categories such as Productive Areas, Real Estate, Research, ICT, Training, Other Services for Public Administration and the Community, and Transport.

Once the Intervention Area is defined, we work on the Intervention Sector and Sub-sector. For the sector, we select Environmental Infrastructure and Water Resources, while discarding the Energy Sector Infrastructure. Within this scope, we then explore the sub-sector categories: Soil Protection, Environmental Protection, Enhancement, and Enjoyment, Reorganization and Rehabilitation of Urban and Industrial Sites, Water Resources and Wastewater, and Waste Disposal. We opt for the Soil Protection sub-sector as it is the one that suits our goal best.

Finally, we further filter the investment database according to the Investment Category. Among Settlements, Other Soil Defense Infrastructures/Structures, Site Remediation, Watercourses, Forests, Water Regulation, Beaches and Seismic Risk Structures/Infrastructures, we focus only on Watercourses and Water Regulation. Thus, filtering for the Environment and Energy area, Environmental Infrastructure and Water Resources sector, Soil Protection sub-sector and Watercourses and Water Regulation categories results in a total of 59,657 investment projects, corresponding to approximately 37.45 billions of euro allocated.

The selection and filtering process of public investments aimed at mitigating hydrologic risks is illustrated in Figure 2.

After filtering the database, we harmonize the administrative data integrated from the Localization dataset. Notably, the information regarding the region, province, and especially the municipality of most investments is particularly noisy due to changes in administrative boundaries. For instance, some provinces and several municipalities underwent unification, suppression, renaming, or changes in area codes. To address this, we utilize the ISTAT shapefile of Italian municipalities as of January 1, 2024, mapping unified, suppressed, or renamed municipalities to their current equivalents. This process ensures that all investments are associated with presently existing municipalities, guaranteeing compatibility and consistency with the municipality-river dataset.

Furthermore, information on the relevant administrative unit is missing for some investments. Specifically, 1,819 projects have an undefined municipality; 114 projects have an undefined municipality and province; while 9 projects have an undefined municipality, province, and region.

To address the lack of precise administrative information for these investments, we expand the dataset as follows. We assign all municipalities within the province to the 1,819 projects that

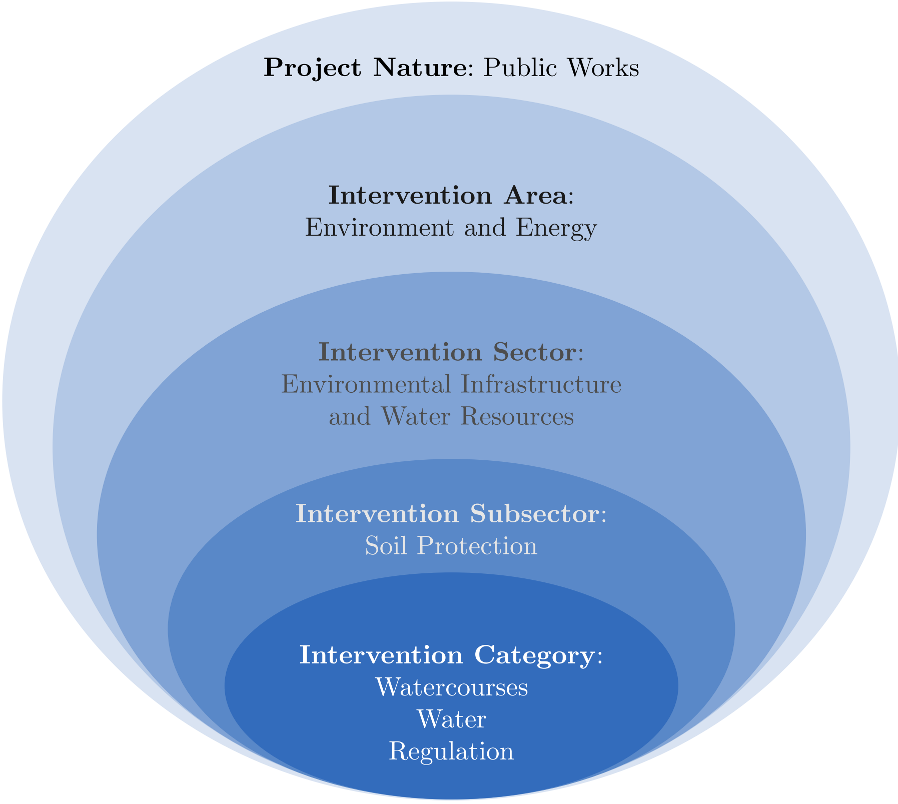


Fig. 2: Visual representation of the filtering process of the OpenCUP database.

provide data up to provincial level. Then we associate all municipalities across all provinces within the region to the 114 projects that provide data up to regional level. Finally, we assign all Italian municipalities to the 9 projects lacking any geographical information. This ensures that each investment project is associated with at least one existing and correctly addressed municipality, preventing any loss of information.

While this expansion procedure may generate some inaccurate observations — since it is unlikely that all municipalities were affected by the investment — this does not pose a significant limitation to the analysis. When merging all data sources to construct the final database, as described in Section 3.5, each investment project is linked to municipality-river combinations only if the investment description explicitly mentions the river under consideration and that river flows through the municipality in which the investment was made, according to the municipality-river dataset. This ensures that inaccurate observations are effectively eliminated.

The resulting dataset on public investment addressed at mitigating hydrologic risks comprises 59,657 investment projects for a total funding of 37.45 billions of euros. The dataset is composed of 392,867 observations resulting from the unique investment-municipality combinations retrieved.

3.4 Extraction of river names from investments

We cannot directly integrate the public investment dataset into the municipality-river dataset, as it lacks a field specifying the targeted river. To achieve this, we extract river names from the intervention descriptions, enabling us to associate investments with the corresponding water bodies.

We start by cleaning and standardizing the intervention descriptions to address issues caused by abbreviations and special characters. Then, we create lists of search expressions to identify targeted rivers within these descriptions. The lists include two types of expressions: (1) general patterns such as the regular expression "FIUME [A-Z]*", which captures all instances where the word *fiume* (river) is followed by any word, and (2) specific river names from the municipality-river dataset, searching within the basin of the municipality where the investment occurs. By applying the search expressions to the intervention descriptions, we identify and extract all instances of these patterns within each description. The extracted instances are stored as lists within each row, meaning that each investment-municipality combination is associated with a list - potentially empty if no river name is found in the description - of identified river names.

Since a single CUP can be linked to multiple rivers, achieving a CUP-municipality-river level of detail is necessary for the public investment dataset. We expand the dataset by breaking down river lists into individual entries, creating one observation for each unique combination of investment, municipality, and river name. We then standardize river names to match those in the hydrography shapefile, ensuring consistency across datasets. The extraction procedure enables us to associate at least one river with 44,049 out of 59,657 CUP (approximately 74%). As a result, the dataset grows from 392,867 CUP-municipality pairs to 473,832 CUP-municipality-river combinations.

3.5 Matching public investments to municipality-river pairs

After expanding the investments dataset, we perform a statistical matching between public investment records and the municipality-river shapefile. We employ a statistical matching rather than an exact matching to identify rivers that are present in the shapefile but may have been misreported in the investment description due to typos or alternative naming conventions. Specifically, for each CUP-municipality-river entry in the public investments dataset, we identify the most similar municipality-river combination within the same municipality, based on river name similarity, measured using the Jaro-Winkler string distance. We consider a match valid only if the similarity score is at least 95%.

Through this process, we successfully link at least one municipality-river pair from the municipality-river shapefile to at least one CUP-municipality-river entry in the public investments dataset for a total of 34,590 CUP records — approximately 58% of all CUP and 79% of those for which we extracted at least one river from the project descriptions. These matched investment projects total approximately 22.52 billion euro, accounting for around 60% of the reference investment dataset.

We may be unable to assign an investment to a municipality-river pair for three main reasons. First, if the investment takes place in a municipality not covered by our municipality-river dataset — meaning our hydrography shows no river geometries there — we cannot establish a match. This applies to 727 projects, totaling approximately 0.6 billion euros. Second, if we cannot extract a single river name from the investment description, assignment becomes impossible. This affects 15,124 CUPs, amounting to about 9 billion euros. Finally, even when we identify a river in a covered municipality, we may still fail to match it if no sufficiently similar river name exists in our dataset for the specific municipalities where the investment occurs. This issue affects 9,173 projects, totaling around 4.7 billion euros.

In Table 1 we report the ten largest unmatched investment projects and the reason why the investment has not been matched. As the Reason column shows, we have been unable to extract a river name from most of these projects. However, this seems mainly caused by the absence of a river name to extract from the Description rather than the inability to find available river names.

Notably, among the largest investments by financed amount, none remain unmatched due to the municipality being absent from the municipality-river dataset.

CUP	Funding	Description	Reason
C92B15000000003	235	Olbia and its waters: works for mitigating hydraulic risk and restoring the city's relationship with its rivers. Hydrographic network of the city of Olbia, urban area: works for mitigating hydraulic risk, construction of embankments, diversion channels, and crossings.	No river found
B71H04000280002	195	Repair works for flood damage in october on the Aia and Palaselva streams in the municipality of Otricoli.	Low score
F13C22002190002	191	Dismantling of the motor vessel La Franca Real and restoration of the hydraulic docks near the Due Giugno bridge in the municipality of Fiumicino, via Ponte Due Giugno: dismantling of the motor vessel and restoration of hydraulic docks.	No river found
D93B04000000002	184	Località Macchia, municipal road Coscia Ponte: road infrastructure maintenance.	No river found
D93B04000010002	166	Località cervara, Coscia Ponte municipal road: road improvement.	No river found
H81H02000040002	155	Hydraulic project: routine maintenance works on hydraulic structures, annual installment.	No river found
H88G00000000002	145	Technical office via Statale: completion of the gutter channel in Località Bellavista for the disposal of rainwater.	Low score
D85D04000030002	135	Località Annicchiali: hydrogeological works.	No river found
F77B15000220003	125	Works to mitigate hydraulica risk in the municipal territory of Olbia: General study drawn up following the flood event of november involving a variation of the PAI and defining the framework of works to mitigate hydraulica risk in the municipal territory of Olbia.	No river found
I24B04000120002	124	Hydraulic system interventions on the rio San Giuseppe: basin intervention.	Low score

Table 1: Top 10 unmatched investment projects: project identifier (CUP), funding in million euro, description, and reason for the missing match. In the Description, actual river names in red.

We associate each investment project with either none, one, or multiple municipality-river pairs. If no match is found, the investment is not assigned to any municipality-river pair[4]. If a single match exists, we allocate the entire funding to that municipality-river pair. When multiple matches occur, we distribute the total funding proportionally based on the length of each river segment.

We then merge this panel dataset with the public investment data at the CUP-municipality-river level, using municipality, river, and year as linking keys. For each combination, we aggregate investments — where present — calculating both annual and cumulative totals for each municipality-river pair across Italy from 2000 to 2024.

3.6 Flood data cleaning

As described in Section 2.3, the ISPRA annual reports on climate events are reported into three distinct tables. We merge these tables to create a comprehensive database containing information on flood events. We concatenate the tables based solely on the event period, using the reported starting and ending event day as aggregation variable.

One potential limitation of using the exact starting and ending period to identify distinct flood events is that events affecting multiple regions within the same perturbation, but occurring on slightly different days, are treated as distinct events. This could potentially bias analyses based on event counts. For example, in the database, the identification code EV_01_2003 uniquely

[4]Alternatively, the funding could be distributed proportionally based on river length across all municipality-river pairs within the municipalities affected by the investment project.

identifies the first flood event of 2003, which affected the Abruzzo and Molise regions from 23rd to 26th January. Similarly, EV_02_2003 uniquely identifies the second flood event reported in 2003, which affected the Campania and Apulia regions from 24th to 26th January. This example illustrates that two separate records are created, despite the fact that, most probably, the same perturbation impacted the four regions in southern Italy within a similar time window.

After merging the three tables to compile a comprehensive dataset, we proceed with a thorough data-cleaning process. Each row in the database represents a single combination of event and affected region, with affected provinces and river basins reported as lists within the same row[5]. Since this structure makes it challenging to merge the data with the river-municipality dataset, we expand the dataset to create one observation for each unique province-river combination within a flood event.

Firstly, we expand the province variable by splitting province lists into their individual elements, assigning a separate row to each unique province. We then harmonize administrative data by assigning to the provinces their currently corresponding name and code based on the ISTAT administrative units as of January 1, 2024.

Secondly, we expand the river basins variable by splitting river lists into their individual elements, assigning a separate row to each unique combination of province and river name. We then standardize the river names to match those used in the hydrography shapefile, ensuring consistency across datasets.

The unique identifier for the cleaned flood events dataset is the combination of the flood event identification code, the province code, and the river name. This structure produces a total of 6,247 triplets and 585 unique river names reportedly involved in overflow events.

We then perform consistency checks between the municipality-river dataset and the flood events dataset. This process allows us to identify and exclude river names that, despite the name cleaning phase, still do not match any rivers in the municipality-river dataset. Additionally, we filter out instances where the reported rivers matched the names in the municipality-river dataset but were incorrectly assigned to the wrong province. This discrepancy likely stems from a mismatch between the reported provinces and the associated rivers in the flood records.

Once the provinces and river names are cleaned and unequivocally associated, we expand the flood events dataset to retrieve information at the municipality level. To achieve this, we utilize the municipality-river dataset, assigning to each combination of province and river the municipalities through which the respective watercourse flows. This process results in the final flood events dataset providing 17,555 triplets of flood event code, municipality and river. The unique river names reportedly involved in overflow events are 484, and the municipalities affected are 3,830 in 105 (out of 107) provinces.

3.7 Flood data integration

Once the flood events dataset is cleaned, we associate each municipality-river pair from the municipality-river dataset with information on whether it was impacted by flooding in a given year, along with the number of flood events that occurred in that year. We do this by merging the municipality-river with aggregated investments and the flood events dataset at municipality-river-year level. Note that, for flood events where specific municipal-level information is unavailable, we assume that all municipality-river combinations along the course of the flooded rivers within the affected province were impacted by the event. However, this assumption may sometimes be incorrect, potentially leading to an overestimation of the flooded area for a given event.

[5] There are a few cases of event-region duplicates, but this is just due to the structure of the data table after the merge. With the dataset expansions for province and rivers later on, these duplicates will appear as any other observation.

By integrating both public investments and flood events records into the municipality-river dataset, we obtain our final dataset.

4 Resulting dataset

The final dataset is a municipality-river-year panel, comprising 2,750,375 observations: 110,015 unique municipality-river pairs over a 25-year period (2000–2024).

Primarily, the dataset provides detailed information on each municipality-river pair, including the river name, municipality, and corresponding province and region. It also includes the geometry of the municipality-river pair and its total length.

On the other hand, regarding public investments in hydrologic risk mitigation, the dataset records yearly investment amounts. Additionally, it includes total investments allocated to each municipality-river pair over the entire observation period, along with moving averages of investments.

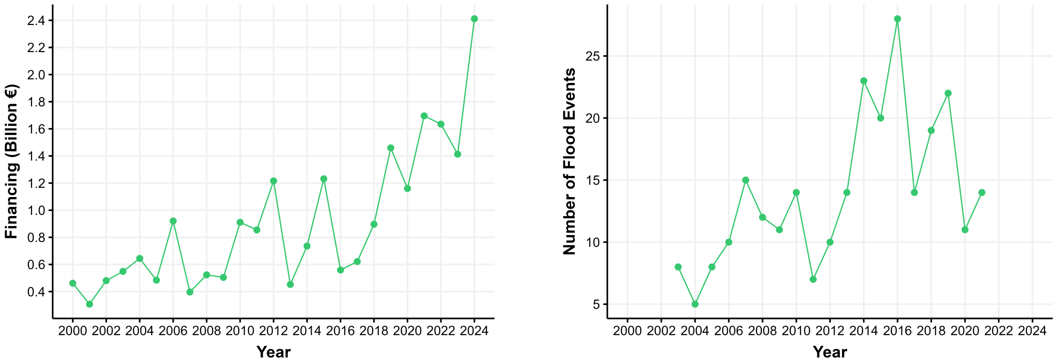
For flood events, the dataset indicates whether a municipality-river pair was affected by a flood in a given year and, if so, reports the total number of flood events that occurred that year.

We provide a comprehensive list of variables in Table 2.

Name	**Description**	**Format**
municipality_river_id	Unique municipality-river identifier	String
year	Year to which financing and flood events data refers to (2000-2024)	Numeric
river_name	Cleaned river name	String
municipality_code	ISTAT municipality code as of 1st January 2024	Numeric
municipality	ISTAT municipality name as of 1st January 2024	String
province_code	ISTAT province code as of 1st January 2024	Numeric
province_acronym	ISTAT province 2-letters acronym as of 1st January 2024	String
province	ISTAT province name as of 1st January 2024	String
region_code	ISTAT region code as of 1st January 2024	Numeric
region	ISTAT region name as of 1st January 2024	String
fin	Total yearly financing associated to the municipality-river	Numeric
fin_municipality_river	Total financing to the municipality-river, 2000-2024	Numeric
fin_ma3	Average financing to the municipality-river over the last three years	Numeric
fin_ma5	Average financing to the municipality-river over the last five years	Numeric
flood_any	Indicator of whether the municipality-river experienced a flood over the year	Numeric
flood_number	If flood_any = 1, number of flood events	Numeric
length	Total length in meters of the municipality-river geometry	Numeric
geometry	Municipality-river geometry	Geometry

Table 2: List and description of the baseline variables included in the final dataset.

In Figure 3, we present the evolution of public investments in mitigating hydrologic risk, alongside the occurrence of flood events. Panel (a) illustrates a significant rise in annual public funding, increasing from approximately 0.4 billion euros in 2000 to over 2.4 billion euros in 2024. The data also reveal considerable year-to-year variability. Panel (b) shifts the focus to the annual frequency of flood events in Italy, which shows an overall upward trend. Notably, there was a sharp increase between 2011 and 2014: while the average number of floods per year was around 10 before 2011, it nearly doubled from 2014 onward, though recent years suggest a possible reversal of this trend. Interestingly, the sharp rise in funding for hydrologic risk mitigation since 2014 may indicate a response to increased flood damage or a heightened awareness of hydrologic risks. At the same time, the reduction of flood events from 2016 onward may reflect the positive impact of increased public spending in hydrologic risk mitigation. The two possible interpretations highlight the presence of a reverse causality problem. On one hand, a higher frequency of floods might have triggered more investments. On the other hand, more public spending to mitigate hydrologic risk may have reduced the occurrence of floods, which is also one of the research questions in this study. An econometric analysis is essential to investigate the true direction of the effect.



(a) Financing

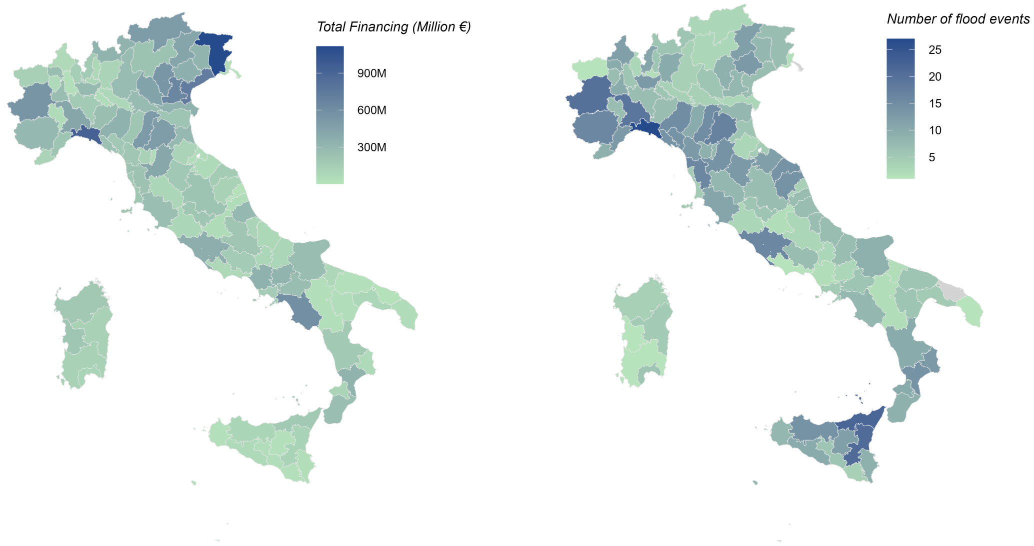
(b) Flood events

Fig. 3: Evolution of total investment funding for hydrologic risk mitigation (panel (a)) and the number of flood events (panel (b)) in Italy.

In Figure 4, we examine the provincial (NUTS 2 level) distribution of public financing and flood events over the analyzed period. Panel (a) reveals that investment levels are highest in northeastern Italy, totaling over 8.15 billion euros, followed by the northwest (5.43 billion), the south (4.57 billion), and the center (4.38 billion). The five regions receiving the highest funding are Veneto (3.49 billion euros), Emilia-Romagna (2.45 billion), Lombardy (2.08 billion), Tuscany (1.93 billion), and Piedmont (1.75 billion). However, the distribution of investments does not closely align with the geographic pattern of flood events shown in panel (b). Flood occurrences are more concentrated in the northwest, as well as in Emilia-Romagna, Tuscany, and Sicily. Interestingly, almost all provinces have been hit by a flood event over the period considered.

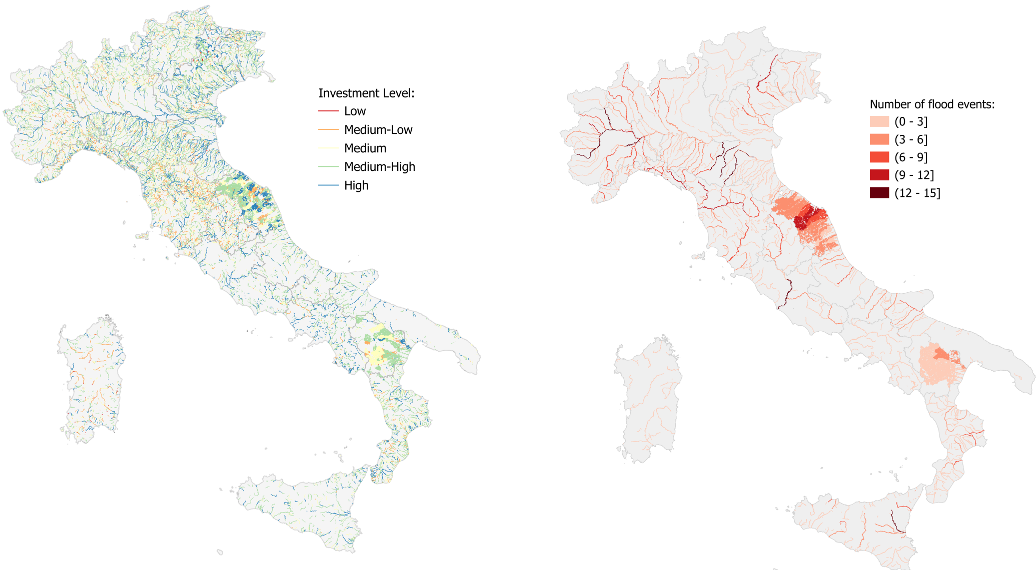
Figure 5 presents the same data at the municipality-river level. Panel (a) illustrates the territorial distribution of municipality-river pairs that received non-null funding during the study period, with investment levels categorized into quintiles. A total of 17,173 municipality-river pairs (approximately 15.6%) received funding. From 2000 to 2024, the average cumulative investment among these pairs was 1,311,479 euro, while the median was significantly lower at 155,959 euro, highlighting a highly skewed distribution. The highest recorded cumulative investment was 374,025,840 euro, allocated to *Torrente Bisagno* in Genoa.

Panel (b) maps municipality-river pairs affected by at least one flood event, with darker colors representing a higher total number of floods. Out of 110,015 municipality-river pairs, 4,978 (approximately 4.5%) have experienced flooding. However, nearly 51.4% of municipalities in the dataset have been affected by at least one flood event. This contrast highlights that while relatively few municipality-river pairs have flooded, a significant portion of Italy's territory has been exposed to flooding, underscoring the country's high hydrologic risk.



(a) Financing (b) Flood events

Fig. 4: Total investment funding for hydrologic risk mitigation (panel (a)) and number of flood events (panel (b)) by province.



(a) Financing (b) Flood events

Fig. 5: Municipality-river pairs with non-null financing for hydrologic risk mitigation, grouped by investment quintile (panel a) and municipality-river pairs affected by at least one flood event, categorized by the number of flood events experienced (panel b).

Displaying data at the municipality-river combination level enables a more precise and detailed analysis, but it also shows certain limitations of the dataset. This is particularly evident in the regions of Marche (central Italy) and Basilicata (southern Italy), where some areas show a high density of both investment and flood events. These areas appear as the most colored regions in panel (a) and even more so in panel (b) of Figure 5.

This primarily occurs for a specific reason. Since hydrography data is provided by the River Basin Authorities, for the regions Marche and Basilicata information on rivers was likely recorded at the basin-wide level, meaning that river geometries reported the name of the river basin rather than that of the specific river. As a result, after cleaning river names and merging geometries within the same municipality under the same name, it was no longer possible to distinguish between main rivers and their tributaries, but only the entire basin.

For instance, rivers such as *Chienti*, *Esino*, *Foglia*, *Metauro*, *Misa*, *Musone*, *Potenza*, and *Tronto* in the Marche region, as well as *Agri*, *Basento*, *Bradano*, and *Cavone* in the Basilicata region, are often recorded with their name followed by a sequential number. After cleaning these names, most geometries within the same basin are merged under a single name, effectively consolidating them into one river in the municipality-river dataset. This constitutes a non-negligible caveat that may distort the analysis by misallocating investments and identifying areas as flooded that, in principle, should not be affected.

5 Conclusions and next steps

The rise in extreme weather events driven by climate change, combined with Italy's high exposure to hydrologic hazards, underscores the need to monitor both the distribution and effectiveness of public investments in hydrologic risk mitigation.

To enable such an analysis, we have created a novel dataset by consolidating and integrating multiple data sources, including the Italian hydrography shapefile, the public investment dataset for hydrologic risk mitigation, and the ISPRA flood event dataset.

Leveraging this dataset, we would like to investigate the role of public investments in reducing hydrologic risks, focusing on two key aspects. First, analyze whether such investments contribute to prevent flood events when extreme weather conditions occur. Second, assess whether - in cases where floods do take place - these investments help in mitigating the financial losses incurred by affected firms. Addressing these questions would provide valuable insights into the protective role of public infrastructure in reducing disaster-related economic disruptions.

The novel dataset we have constructed has the potential to contribute significantly to the existing literature in multiple ways. First, it provides a detailed mapping of public investments dedicated to mitigating hydrologic risks, offering a fine-grained geographic distribution of investment allocations at the river-municipality level. This enables an in-depth analysis of the spatial dispersion of investments and their temporal evolution from 2000 to 2024. Moreover, by retrieving river-level information, the dataset enhances the precision of identifying flood-affected areas, even in cases of incomplete data. This, in turn, allows for a localized mapping of flood-related damages incurred by firms, and a comprehensive assessment of both the actual and potential positive spillover effects generated by public investments in hydrologic risk mitigation.

We plan to integrate additional data sources in the dataset. First of all, we have begun collecting data on monthly 1-day and 5-day maxima of cumulated precipitation from Copernicus (Copernicus Climate Change Service, 2024), aiming to identify extreme rainfall events. Additionally, we plan to incorporate multiple control variables related to physical features, such as land use and land cover, soil moisture, and soil elevation.

The next phase of the research involves collecting firm-level balance sheet data to link firms' performance to public investments and flood events, based on their proximity to rivers. To achieve this, we plan on locating firms and their production units. Once this is done, we will consider a firm affected by a flood event if it is sufficiently close to inundated rivers, with proximity determined by a buffer applied along the watercourses. We acknowledge that the size of the chosen buffer may significantly influence the econometric results and, thus, plan on performing sensitivity analyses changing the buffer size.

The final dataset can be further enhanced by incorporating details on the nature of the investment — distinguishing between prevention and restoration measures — and including additional sections of the original OpenCUP database to identify firms that applied for incentives and assess whether they were affected by flooding.

Moreover, collecting precise geospatial data from Copernicus on past flood events, when available, would add significant value by enabling use case applications by leveraging information such as flood extent and depth, the evolution of flooded areas over time, and the impact on buildings in specific areas of interest.

Developing detailed use cases would allow us to refine our assumptions and methodologies, enhancing the accuracy of firm identification within flooded areas by leveraging precise flood extent maps. This approach would also help verify whether the reported rivers correspond to the actual basin responsible for triggering the flood event and assess the impact of extreme rainfall. By applying this method, we could analyze investments at a granular level and gain deeper insights into the functionality of installed infrastructure — such as the activation of overflow pipes during extreme rainfall events — to determine their effectiveness in mitigating floods. This underscores the broader importance of targeted use cases in achieving geospatial precision for evaluating flood prevention and response strategies.

References

E. Benincasa, F. Betz, and L. Gattini. How do firms cope with losses from extreme weather events? *Journal of Corporate Finance*, 84:102508, Feb. 2024. ISSN 09291199. doi: 10.1016/j.jcorpfin.2023.102508. URL https://linkinghub.elsevier.com/retrieve/pii/S0929119923001578.

D. Berti and M. Lucarini. Eventi alluvionali [data set], 2024. URL https://indicatoriambientali.isprambiente.it/it/pericolosita-da-alluvione/eventi-alluvionali.

K. Calvin, D. Dasgupta, G. Krinner, A. Mukherji, P. W. Thorne, C. Trisos, J. Romero, P. Aldunce, K. Barret, G. Blanco, et al. *IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland.*, pages 1–34, 2023.

S. Clò, F. David, and S. Segoni. The impact of hydrogeological events on firms: Evidence from Italy. *Journal of Environmental Economics and Management*, 124:102942, Mar. 2024. ISSN 00950696. doi: 10.1016/j.jeem.2024.102942. URL https://linkinghub.elsevier.com/retrieve/pii/S0095069624000160.

CNR. Polaris - popolazione a rischio da frana e da inondazione in italia [data set], 2024. URL https://polaris.irpi.cnr.it/event/.

Copernicus Climate Change Service. Gridded dataset underpinning the copernicus interactive climate atlas. copernicus climate change service (c3s) climate data store (cds) [data set], 2024. URL https://cds.climate.copernicus.eu/datasets/multi-origin-c3s-atlas?tab=overview.

Copernicus Emergency Management Service. [data set], n.d. URL https://mapping.emergency.copernicus.eu/activations/.

DiPE and MEF. Opencup [data set], 2025. URL https://www.opencup.gov.it/portale/web/opencup/accesso-agli-open-data.

ECB and EIOPA. Towards a european system for natural catastrophe risk management. 2024. URL https://www.eiopa.europa.eu/document/download/d8c87070-f602-4bf7-b8d8-726ec0b5c173_en?filename=eiopa-ecb-climate-insurance-protection-gap.pdf.

S. Fatica, G. Katay, and M. Rancan. Floods and firms: vulnerabilities and resilience to natural disasters in Europe. *SSRN Electronic Journal*, 2024. ISSN 1556-5068. doi: 10.2139/ssrn.4796097. URL https://www.ssrn.com/abstract=4796097.

A. Gatto, S. Clò, F. Martellozzo, and S. Segoni. Tracking a Decade of Hydrogeological Emergencies in Italian Municipalities. *Data*, 8(10):151, Oct. 2023. ISSN 2306-5729. doi: 10.3390/data8100151. URL https://www.mdpi.com/2306-5729/8/10/151.

Humanitarian OpenStreetMap Team. Italy waterways [data set], 2025. URL https://data.humdata.org/dataset/hotosm_ita_waterways.

ISPRA. Reticolo idrografico nazionale [data set], 2020a. URL https://geodati.gov.it/resource/id/ispra_rm:01Idro250N_DT.

ISPRA. Mosaicatura ispra (2020) delle aree a pericolosità idraulica – scenari d. lgs. 49/2010 [data set], 2020b. URL https://idrogeo.isprambiente.it/app/page/open-data.

ISPRA. Repertorio nazionale degli interventi per la difesa del suolo (rendis) [data set], n.d. URL http://www.rendis.isprambiente.it/rendisweb/vistepub.jsp.

Regione Liguria. Reticolo idrografico e bacini idrografici sc. 1:10000 dgr n.1280/2023 [data set], 2025. URL https://geoportal.regione.liguria.it/catalogo/mappe.html?typeEvent=detailFromHome&idmap=2542.

Regione Piemonte. Reticolo idrografico regionale - db prior 10k - nodi idrici (storico) [data set], 2009. URL https://geodati.gov.it/resource/id/r_piemon:104e43cd-5c50-4345-b508-5ba83a92ba0b.

A. Trigila, L. B. Iadanza, C and, M. Bussettini, and A. Barbano. Dissesto idrogeologico in italia: pericolosità e indicatori di rischio - edizione 2021. Technical report, ISPRA, 2021.