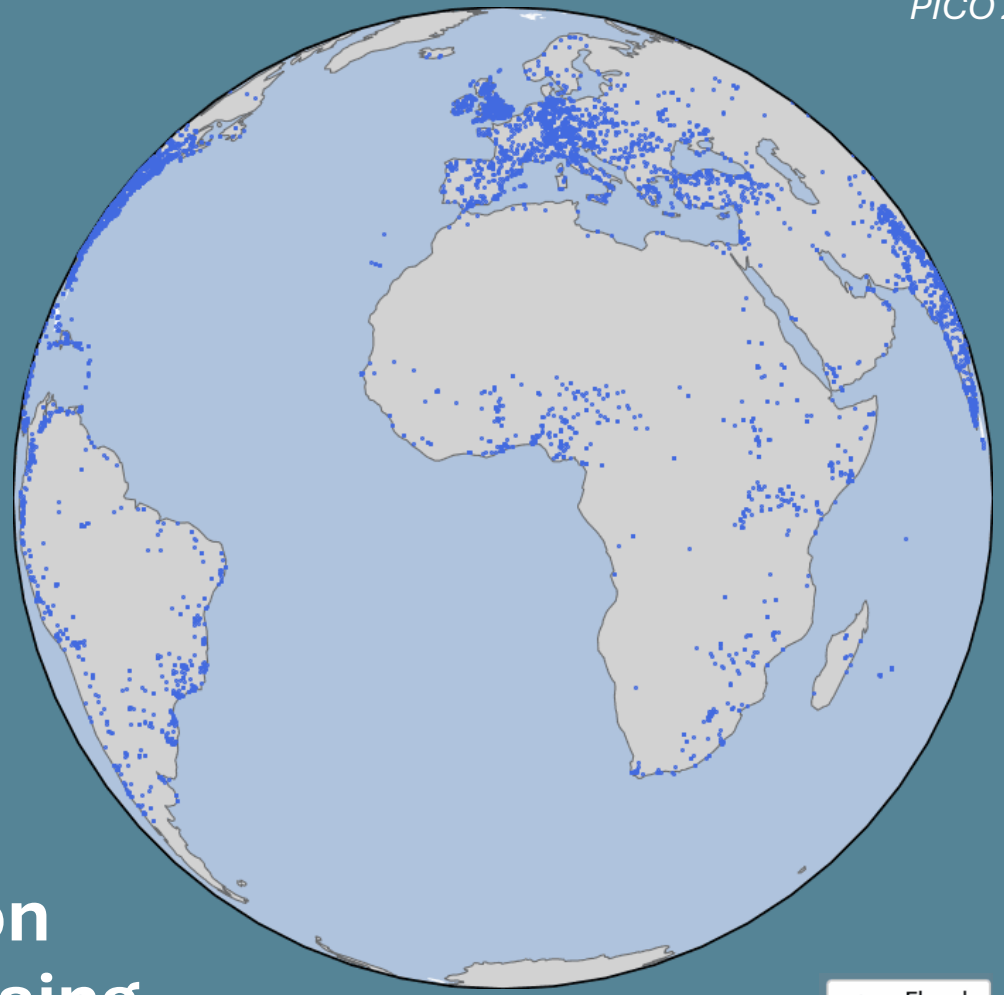
PICO 2.4



A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

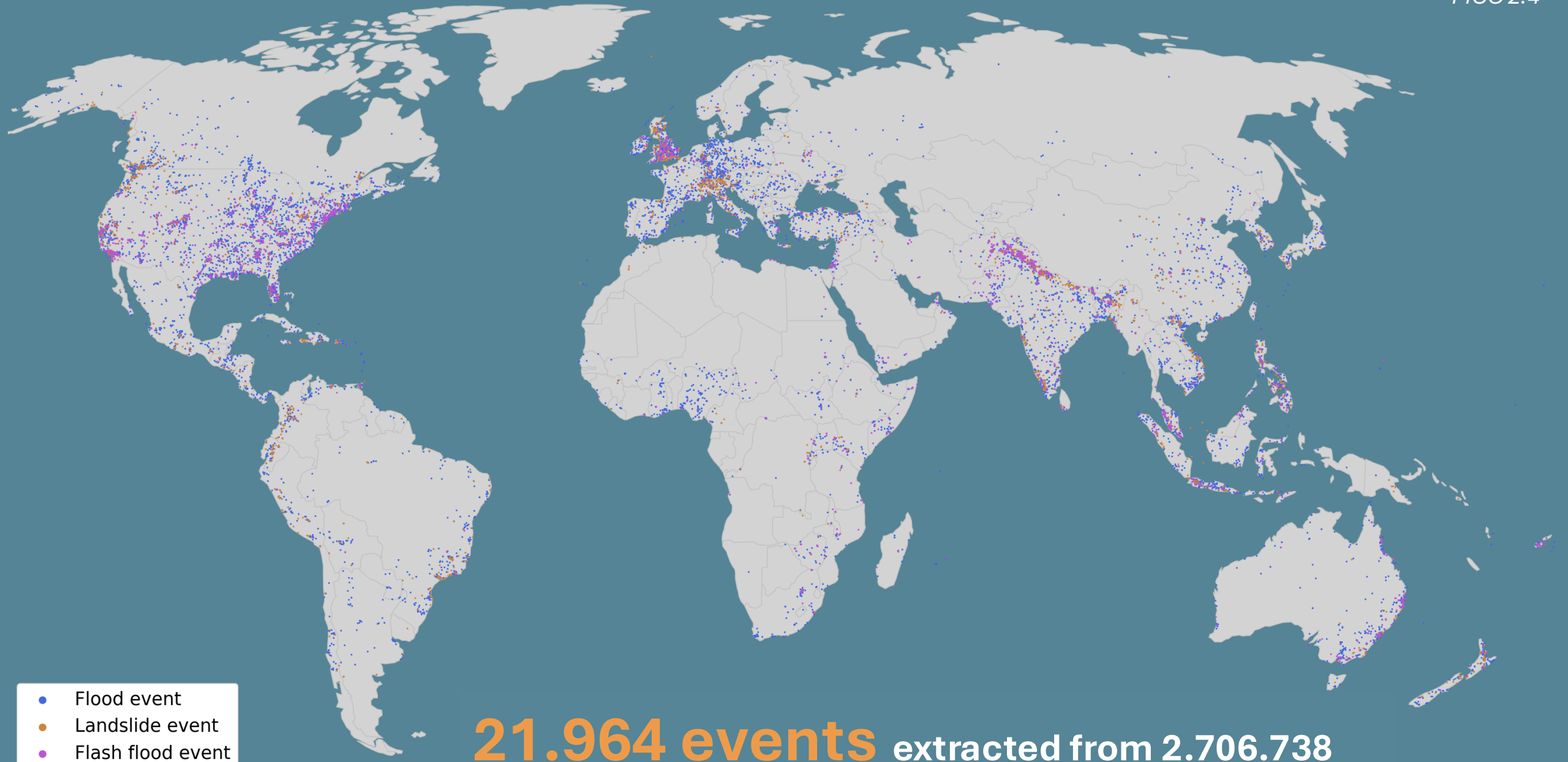
Bram Valkenborg [1, 2], *Olivier Dewitte* [2], *Benoît Smets* [1, 2]

[1] Cartography and GIS Research Group, Department of Geography, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium

[2] Natural Hazards and Cartography, Department of Earth Sciences, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium

AFRICA museum

VUB VRIJE UNIVERSITEIT BRUSSEL



21.964 events extracted from 2.706.738 articles of 15.059 sources

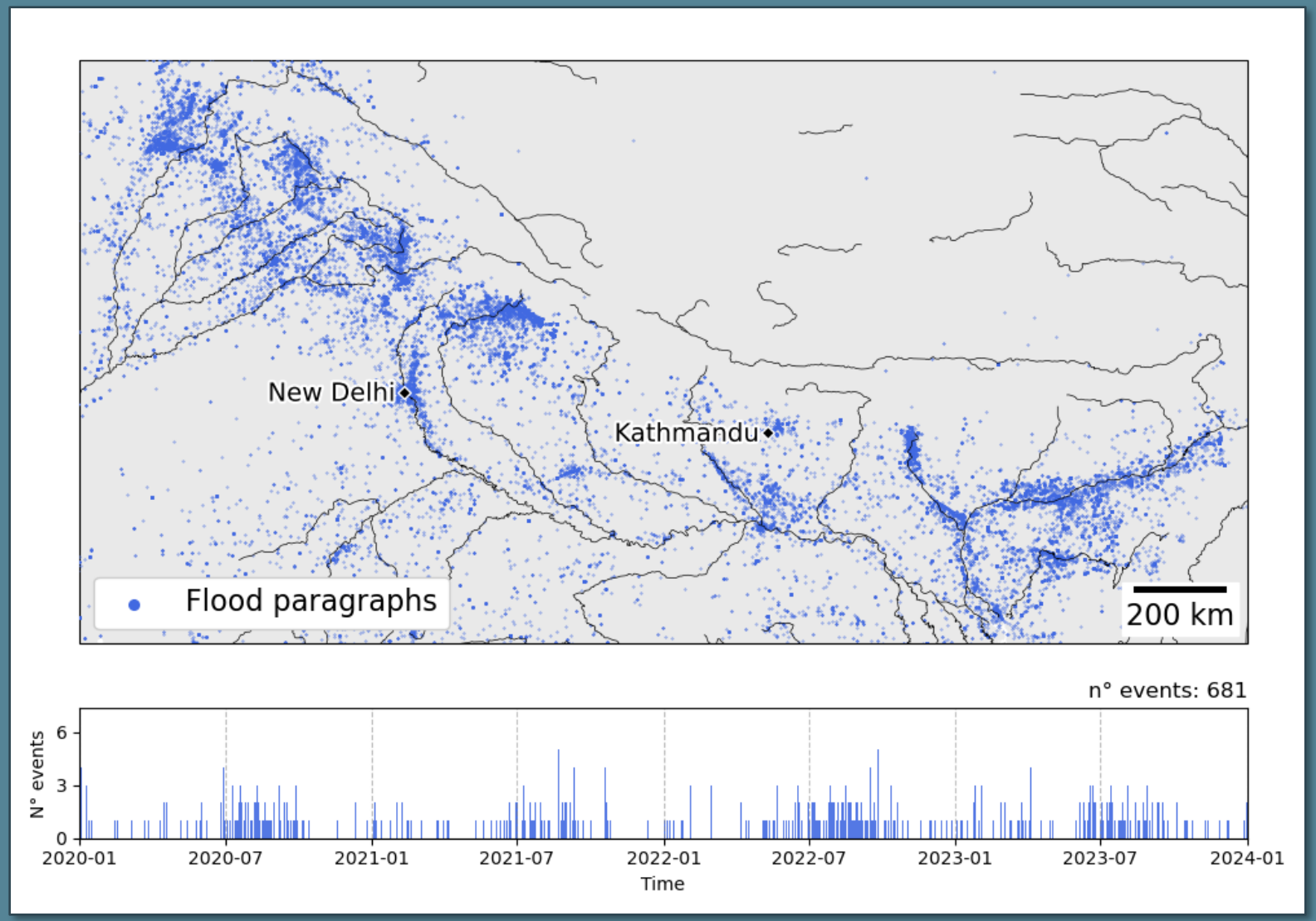
AFRICA museum
VUB VRIJE UNIVERSITEIT BRUSSEL

High level of detail

- Floods align with river basins
- Seasonal trend

Goal of the algorithm

- Fighting data scarcity
- Increase information on natural hazards
- Complementary approach to citizen science observer networks and remote sensing



A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

***Bram Valkenborg*[1,2], *Olivier Dewitte*[2], *Benoît Smets*[1,2]**

[1]Cartography and GIS Research Group, Department of Geography, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium
[2]Natural Hazards and Cartography, Department of Earth Sciences, Royal Museum for Central Africa, Leuvensesteenweg 13, 3080 Tervuren, Belgium

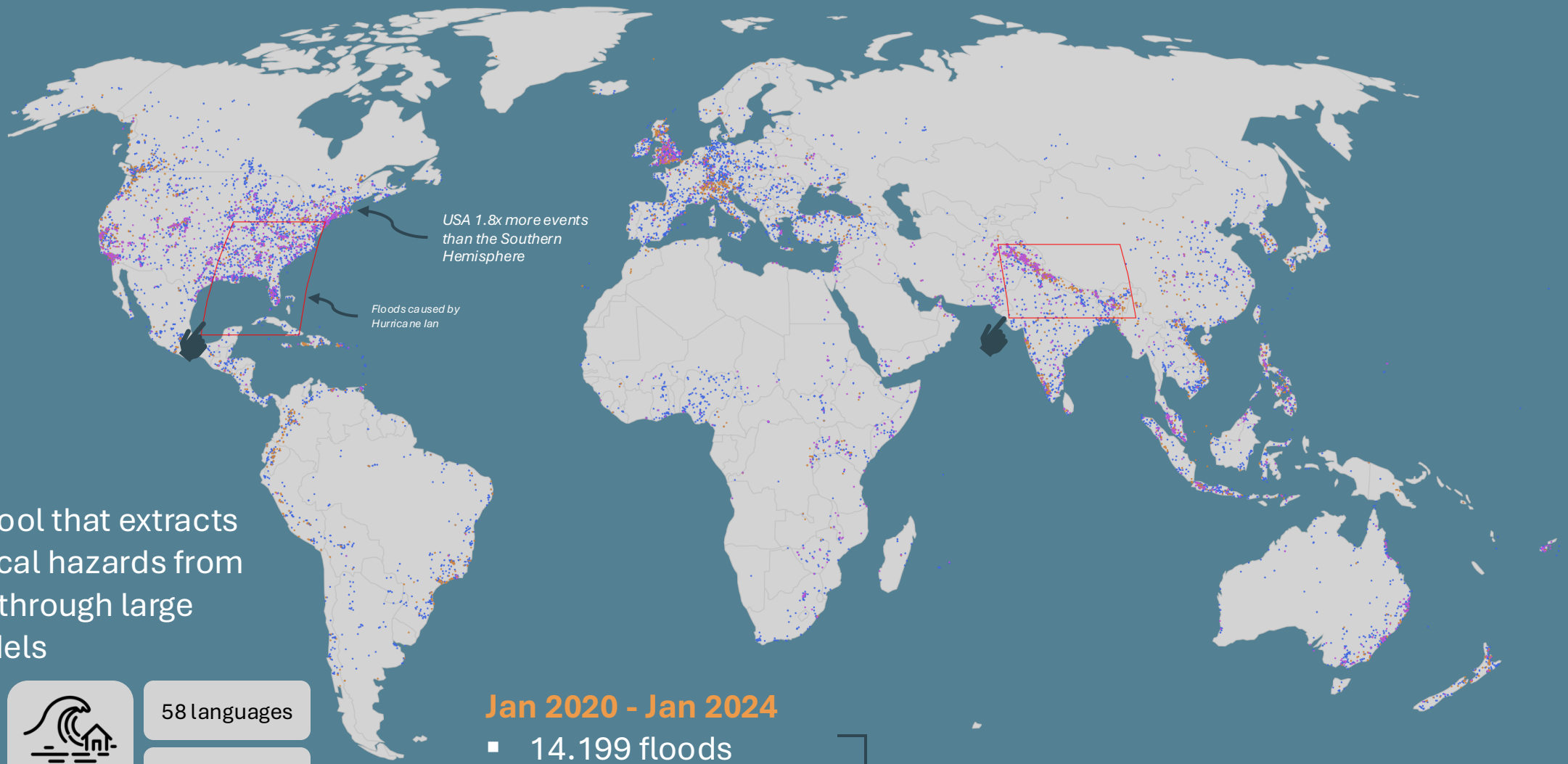
HazMiner

Context

Method

Validation

Conclusion



HazMiner: a tool that extracts geo-hydrological hazards from news articles through large language models

● FLOOD ● LANDSLIDE ● FLASH FLOOD

58 languages

Time series

Jan 2020 - Jan 2024

- 14.199 floods
- 4.054 landslides
- 3.711 flash floods

Extracted from 2.706.738 articles

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

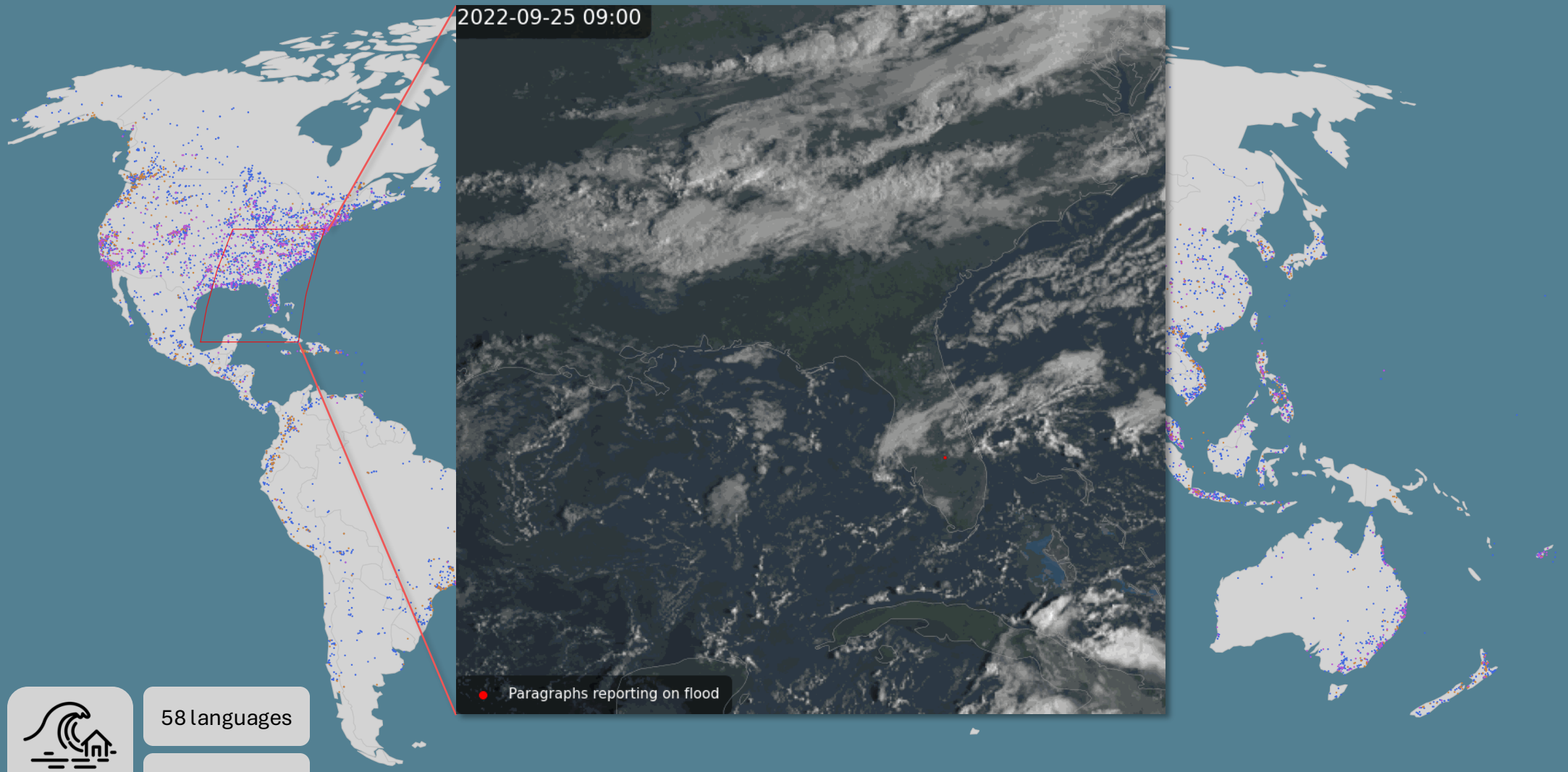
HazMiner

Context

Method

Validation

Conclusion



● FLOOD ● LANDSLIDE ● FLASH FLOOD

58 languages

Time series

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

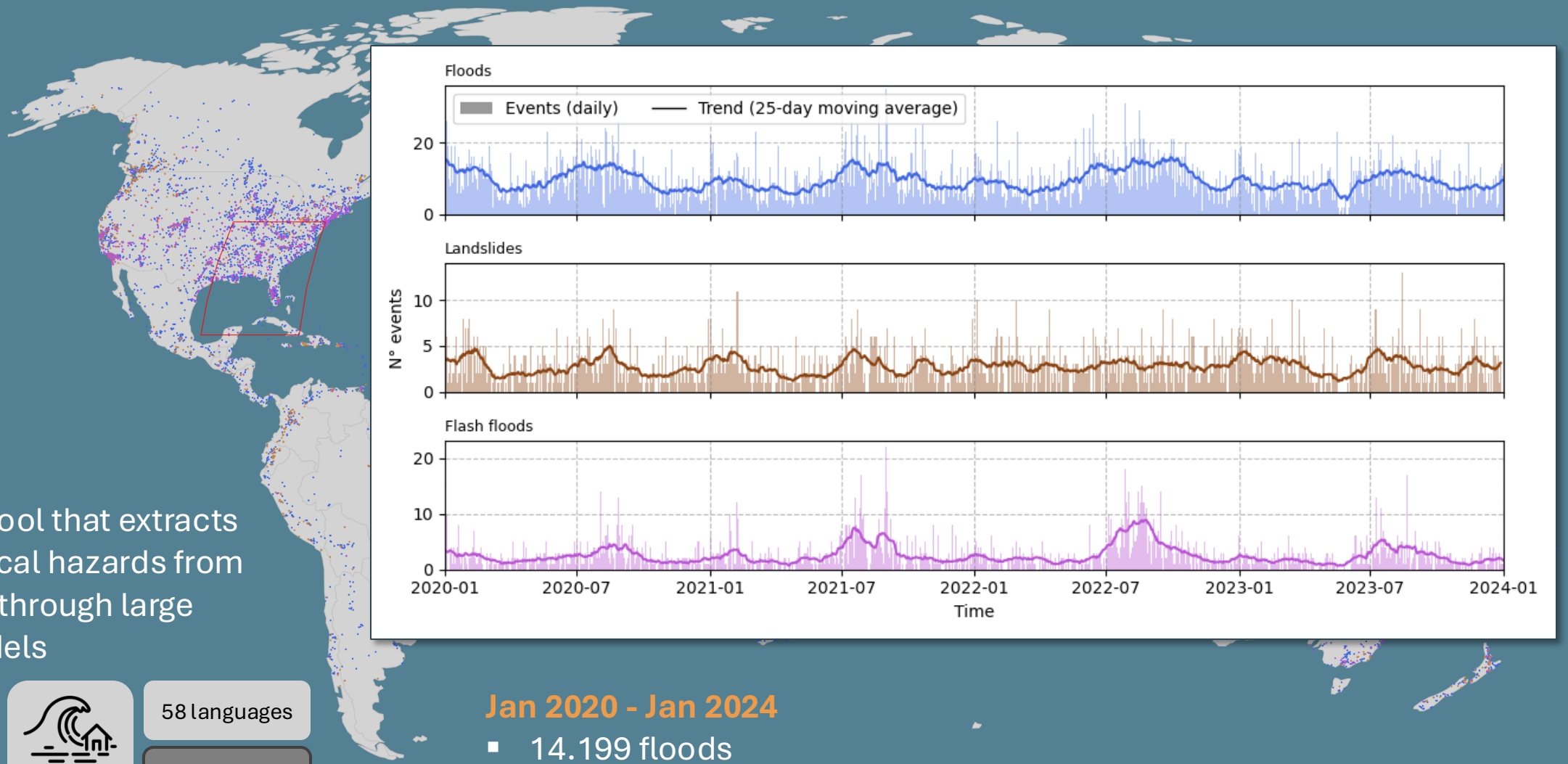
HazMiner

Context

Method

Validation

Conclusion



HazMiner: a tool that extracts geo-hydrological hazards from news articles through large language models

● FLOOD ● LANDSLIDE ● FLASH FLOOD

58 languages

Time series

Jan 2020 - Jan 2024

- 14.199 floods
- 4.054 landslides
- 3.711 flash floods

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

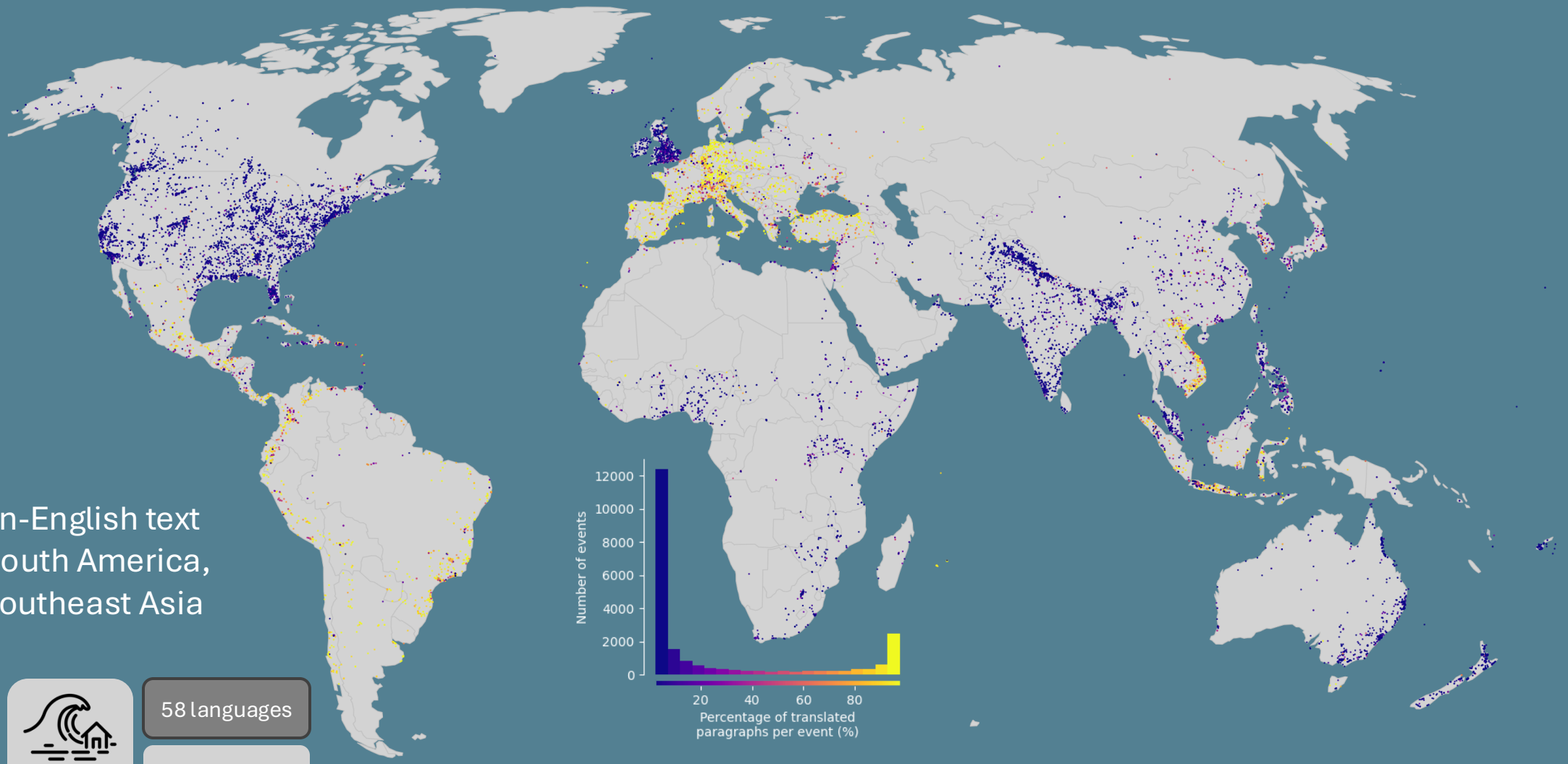
HazMiner

Context

Method

Validation

Conclusion



Language: non-English text important in South America, Europe, and Southeast Asia

● FLOOD ● LANDSLIDE ● FLASH FLOOD

58 languages

Time series

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

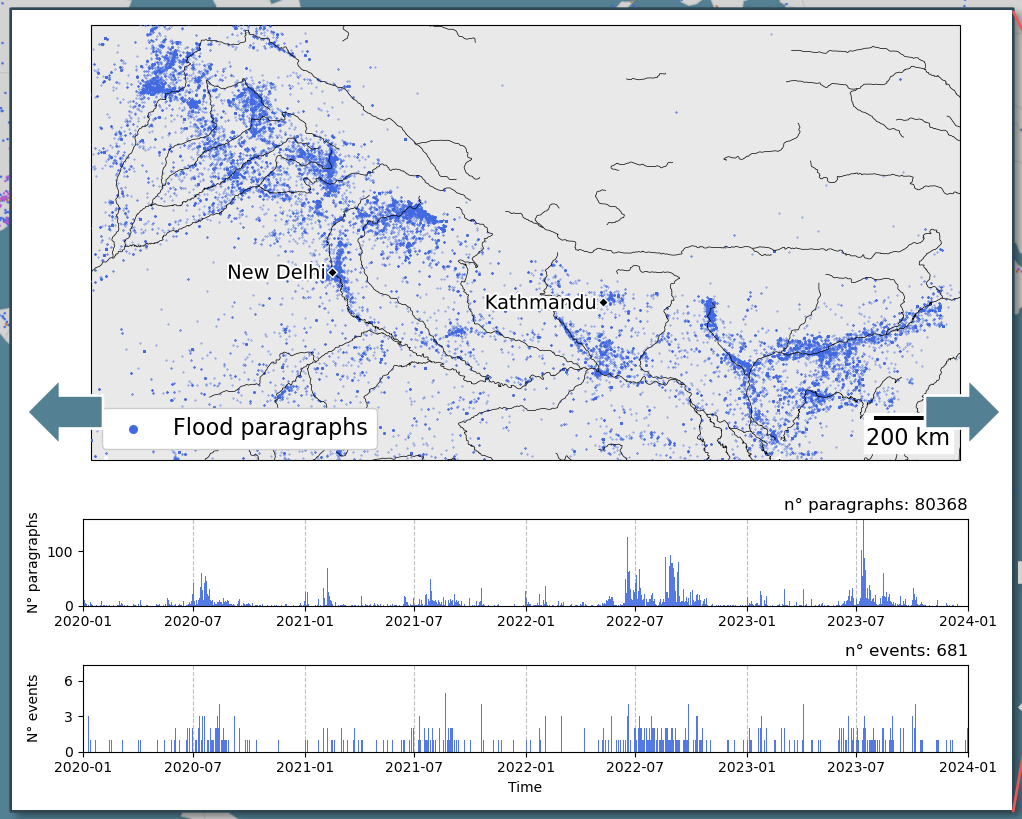
Context

Method

Validation

Conclusion

Paragraphs → quantity
Events → quality



Reduces media bias

● FLOOD ● LANDSLIDE ● FLASH FLOOD

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

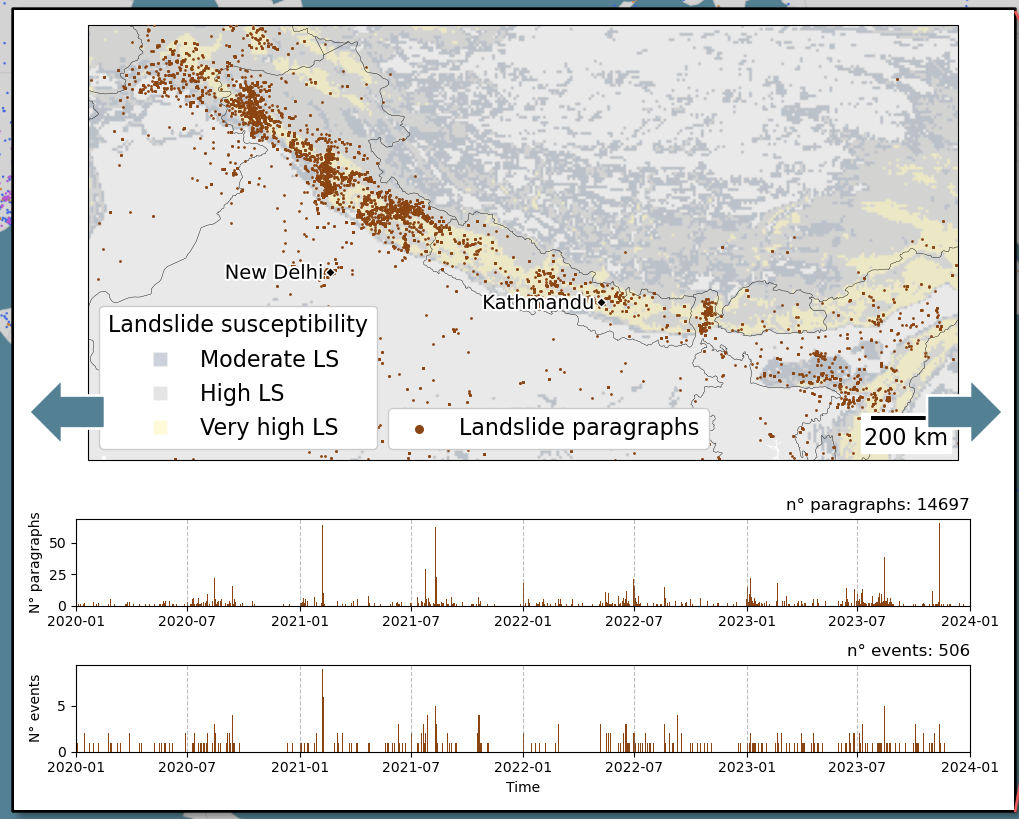
HazMiner

Context

Method

Validation

Conclusion



Paragraphs → quantity
Events → quality

Reduces media bias

● FLOOD ● LANDSLIDE ● FLASH FLOOD

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

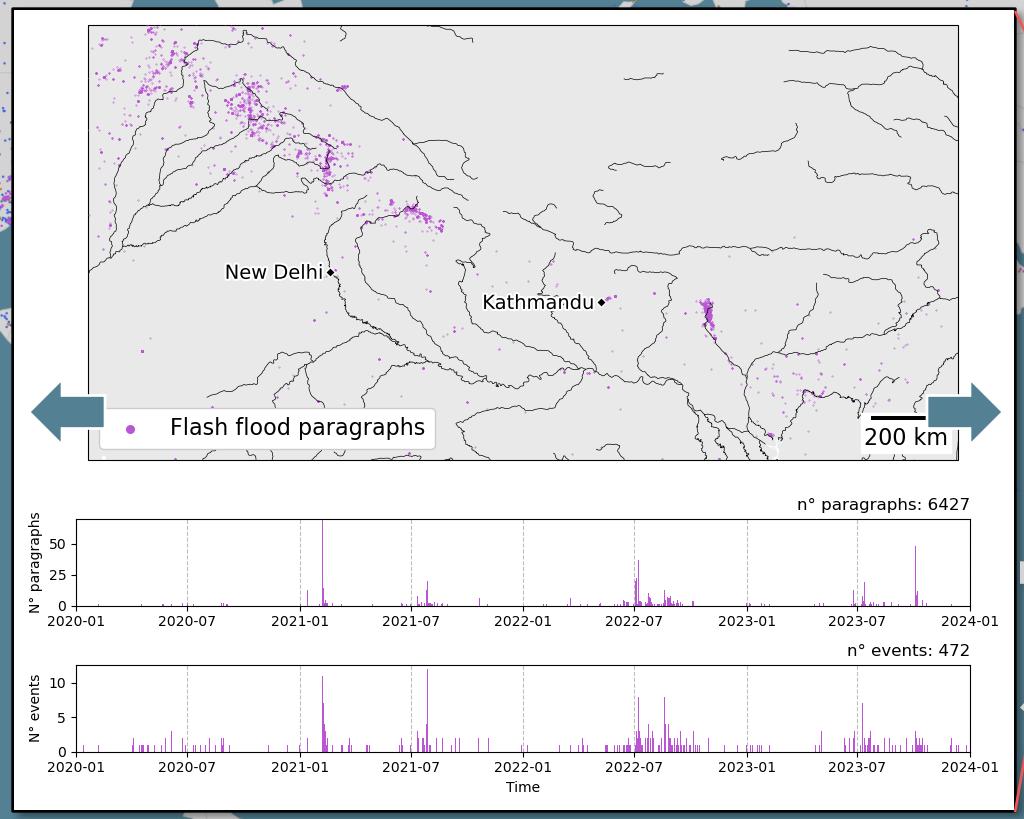
Context

Method

Validation

Conclusion

Paragraphs → quantity
Events → quality



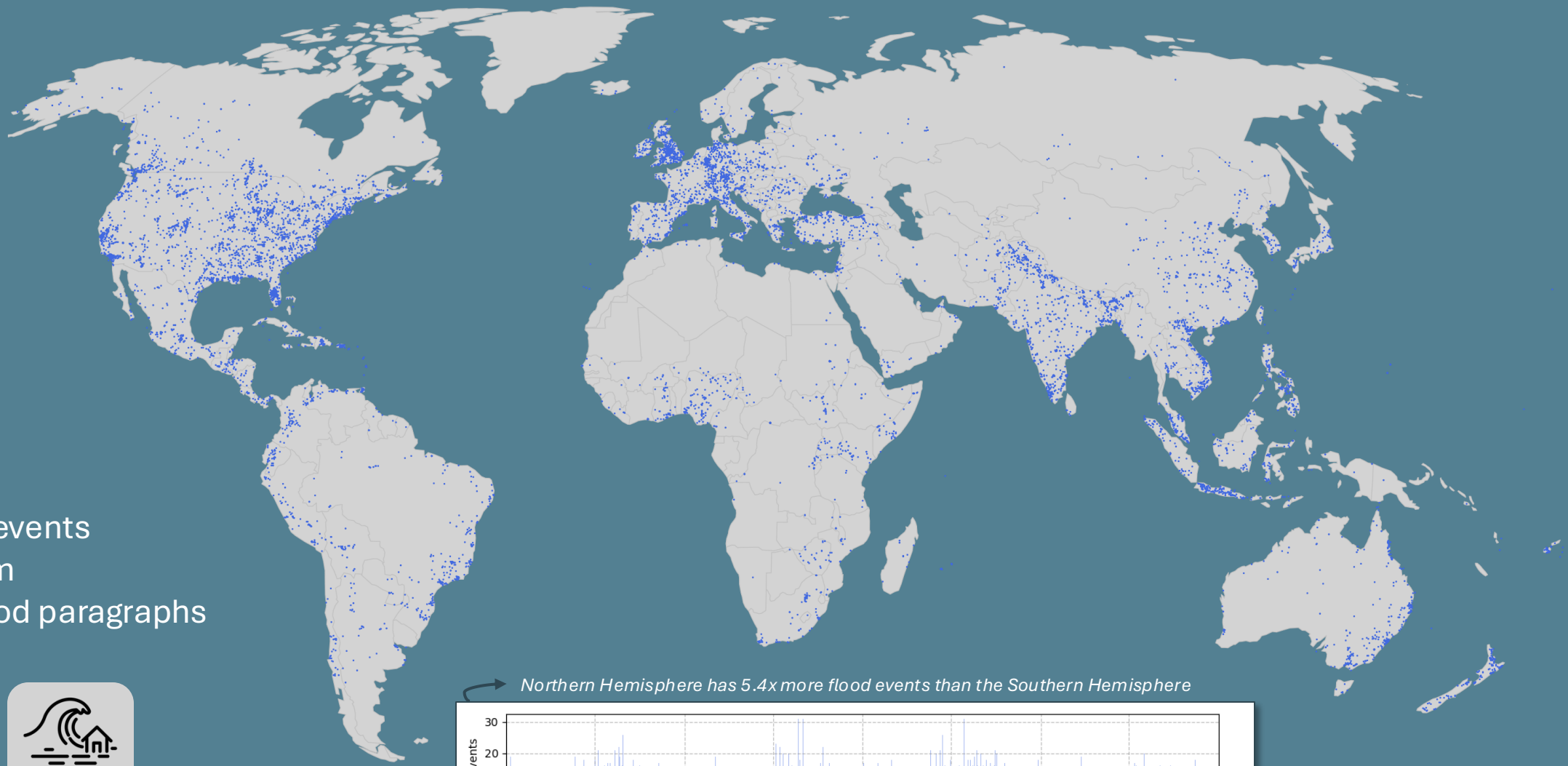
Reduces media bias

● FLOOD ● LANDSLIDE ● FLASH FLOOD

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

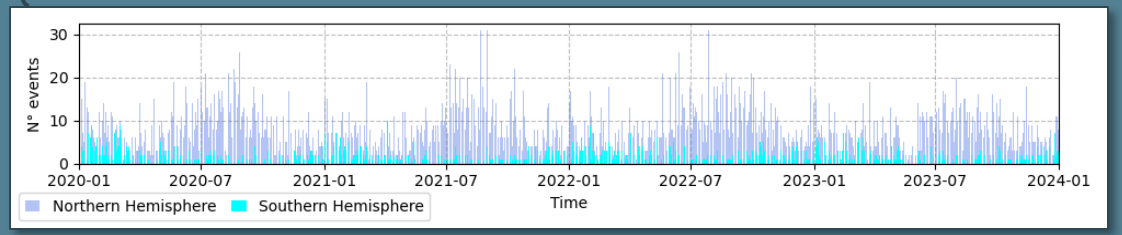
A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

- HazMiner
- Context
- Method
- Validation
- Conclusion



14.199 flood events
extracted from
2.935.881 flood paragraphs

Northern Hemisphere has 5.4x more flood events than the Southern Hemisphere



VUB VRIJE UNIVERSITEIT BRUSSEL
AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

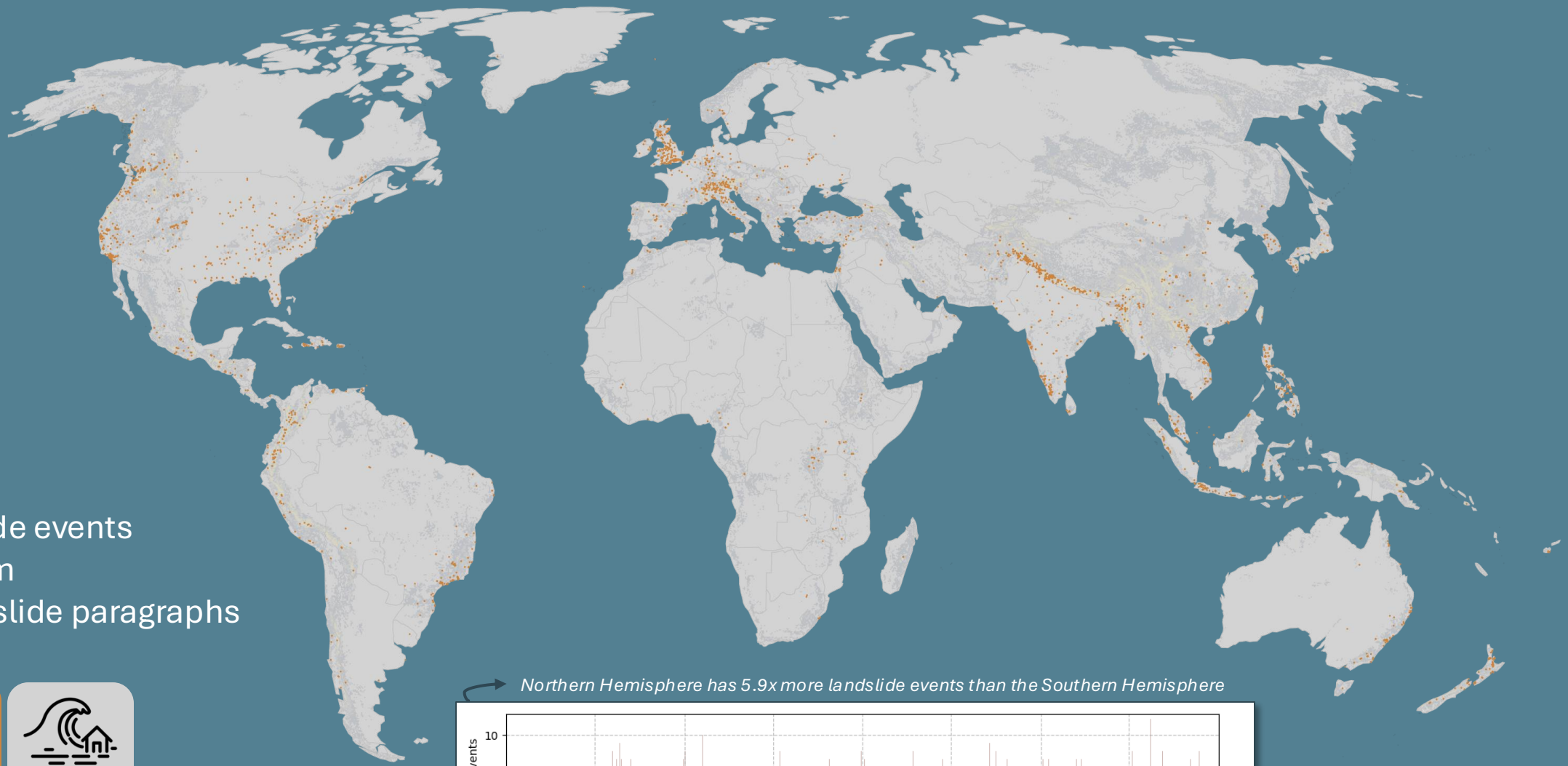
HazMiner

Context

Method

Validation

Conclusion

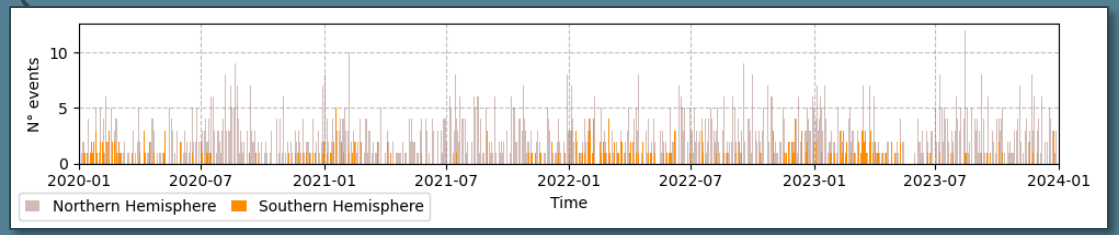


4.054 landslide events
extracted from
233.742 landslide paragraphs

● FLOOD ● LANDSLIDE ● FLASH FLOOD

Landslide susceptibility
- Moderate LS
- High LS
- Very high LS

Northern Hemisphere has 5.9x more landslide events than the Southern Hemisphere



VUB VRIJE UNIVERSITEIT BRUSSEL

AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

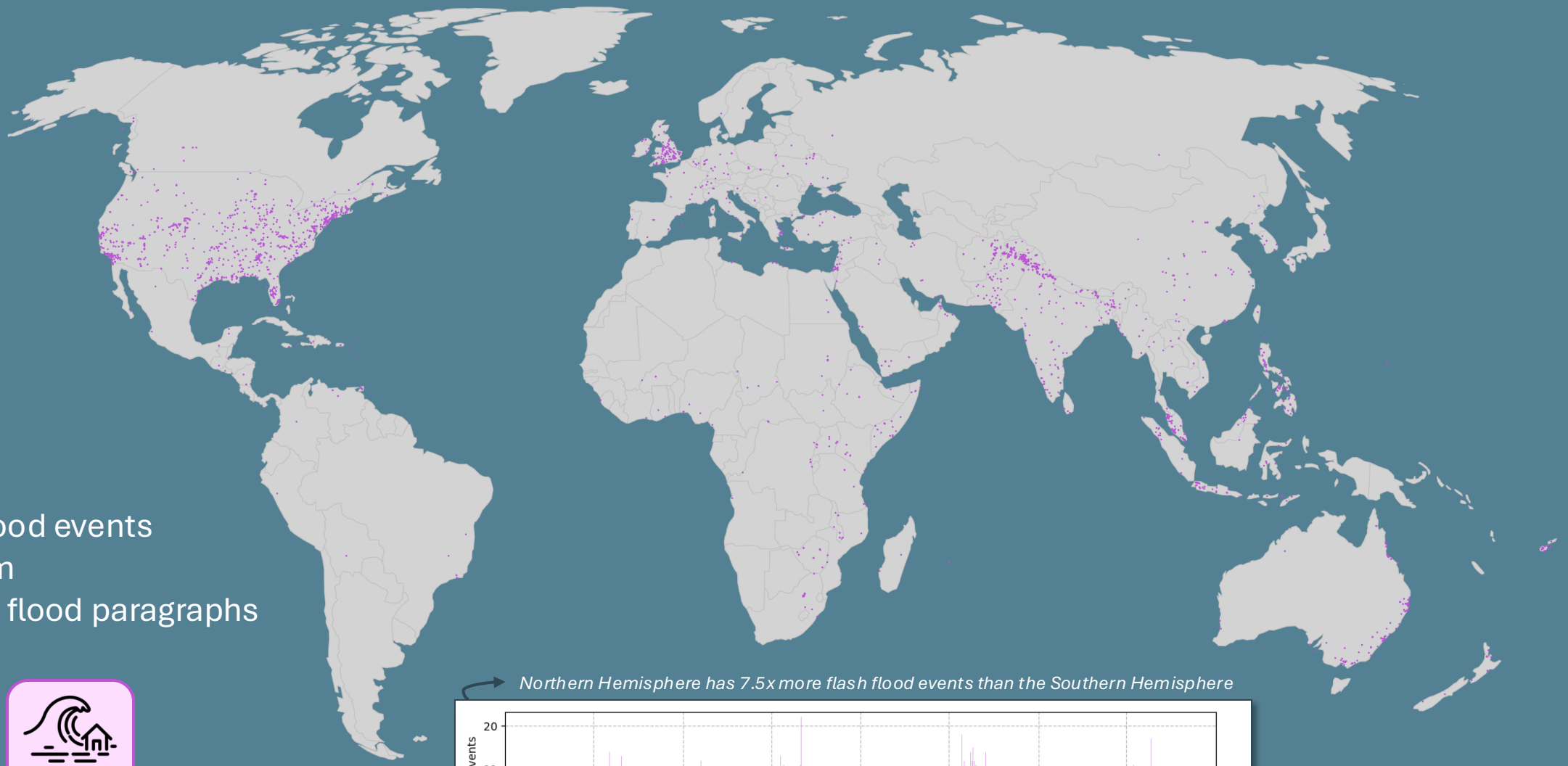
HazMiner

Context

Method

Validation

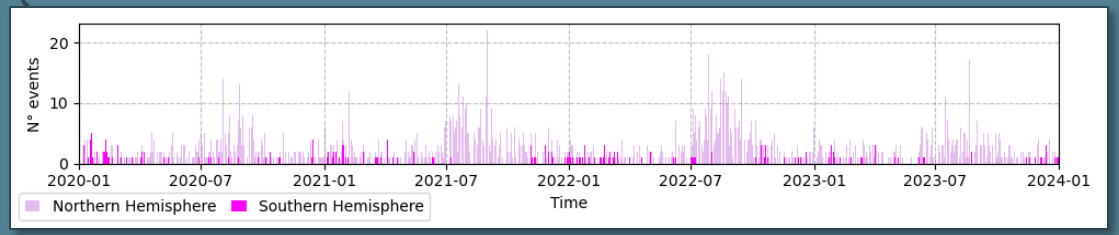
Conclusion



3.711 flash flood events
extracted from
117.205 flash flood paragraphs

● FLOOD ● LANDSLIDE ● FLASH FLOOD

Northern Hemisphere has 7.5x more flash flood events than the Southern Hemisphere



VUB VRIJE UNIVERSITEIT BRUSSEL

AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

Validation

Conclusion

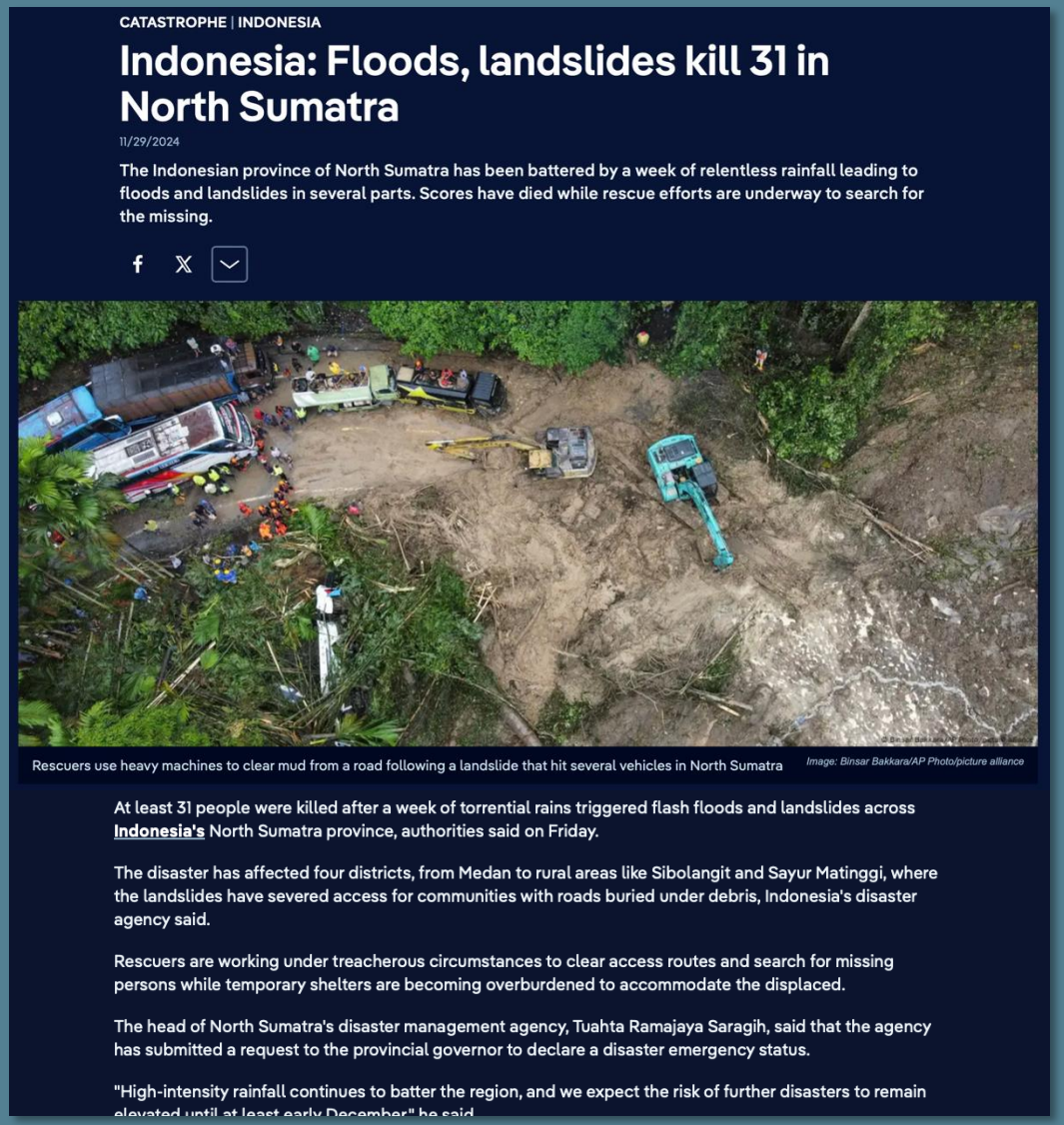
Data scarcity: incomplete documentation of geo-hydrological hazards at the global scale

Limitations of traditional methods: time-consuming, resource-intensive, and may fall short regarding data needs, especially at large scales

Need for innovative solutions:

- Large Language Models can be used to extract information from text
- Online news articles provide unstructured documentation of events

Research question: Can multilingual natural language processing of online news articles enhance documentation of geo-hydrological hazards?



CATASTROPHE | INDONESIA

Indonesia: Floods, landslides kill 31 in North Sumatra

11/29/2024

The Indonesian province of North Sumatra has been battered by a week of relentless rainfall leading to floods and landslides in several parts. Scores have died while rescue efforts are underway to search for the missing.

Rescuers use heavy machines to clear mud from a road following a landslide that hit several vehicles in North Sumatra

Image: Binsar Bakkara/AP Photo/picture alliance

At least 31 people were killed after a week of torrential rains triggered flash floods and landslides across Indonesia's North Sumatra province, authorities said on Friday.

The disaster has affected four districts, from Medan to rural areas like Sibolangit and Sayur Matinggi, where the landslides have severed access for communities with roads buried under debris, Indonesia's disaster agency said.

Rescuers are working under treacherous circumstances to clear access routes and search for missing persons while temporary shelters are becoming overburdened to accommodate the displaced.

The head of North Sumatra's disaster management agency, Tuahta Ramajaya Saragih, said that the agency has submitted a request to the provincial governor to declare a disaster emergency status.

"High-intensity rainfall continues to batter the region, and we expect the risk of further disasters to remain elevated until at least early December" he said.

VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

Validation

Conclusion

1. Text extraction

- Keyword-based from Global Database Events, Language and Tone (2020-2023)
- Translation from 58 languages into English

2. Filtering

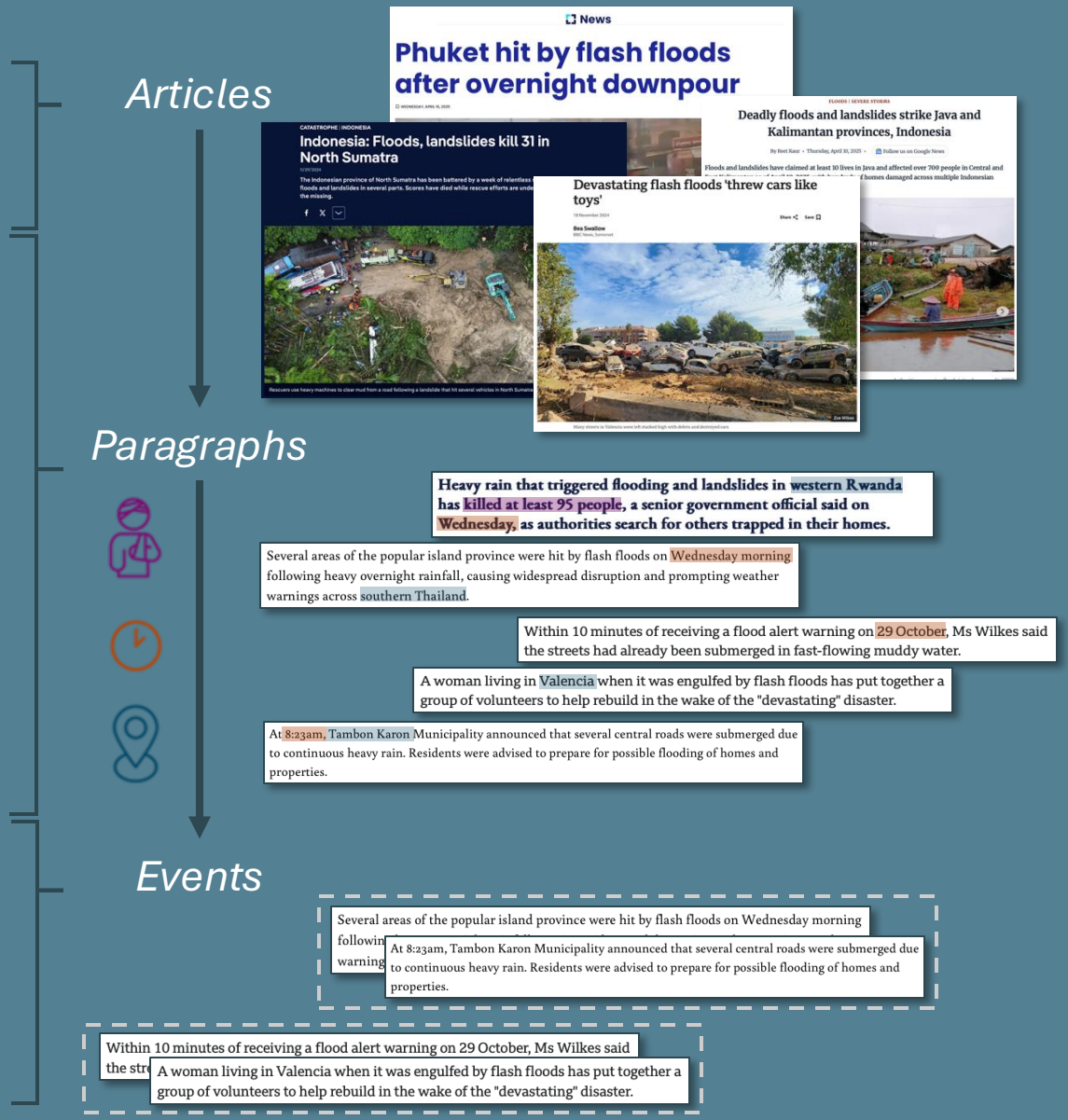
- Hazard-related text filter: binary classification
- Hazard (zero-shot) classification

3. Information extraction

- Impact: Q&A algorithm
- Timing: absolute and relative timing
- Geoparsing: on the original language

4. Event creation

- Events = clustered paragraphs
- H-DBSCAN algorithm
- Space and time



VUB VRIJE UNIVERSITEIT BRUSSEL AFRICA museum

A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

Validation

Conclusion

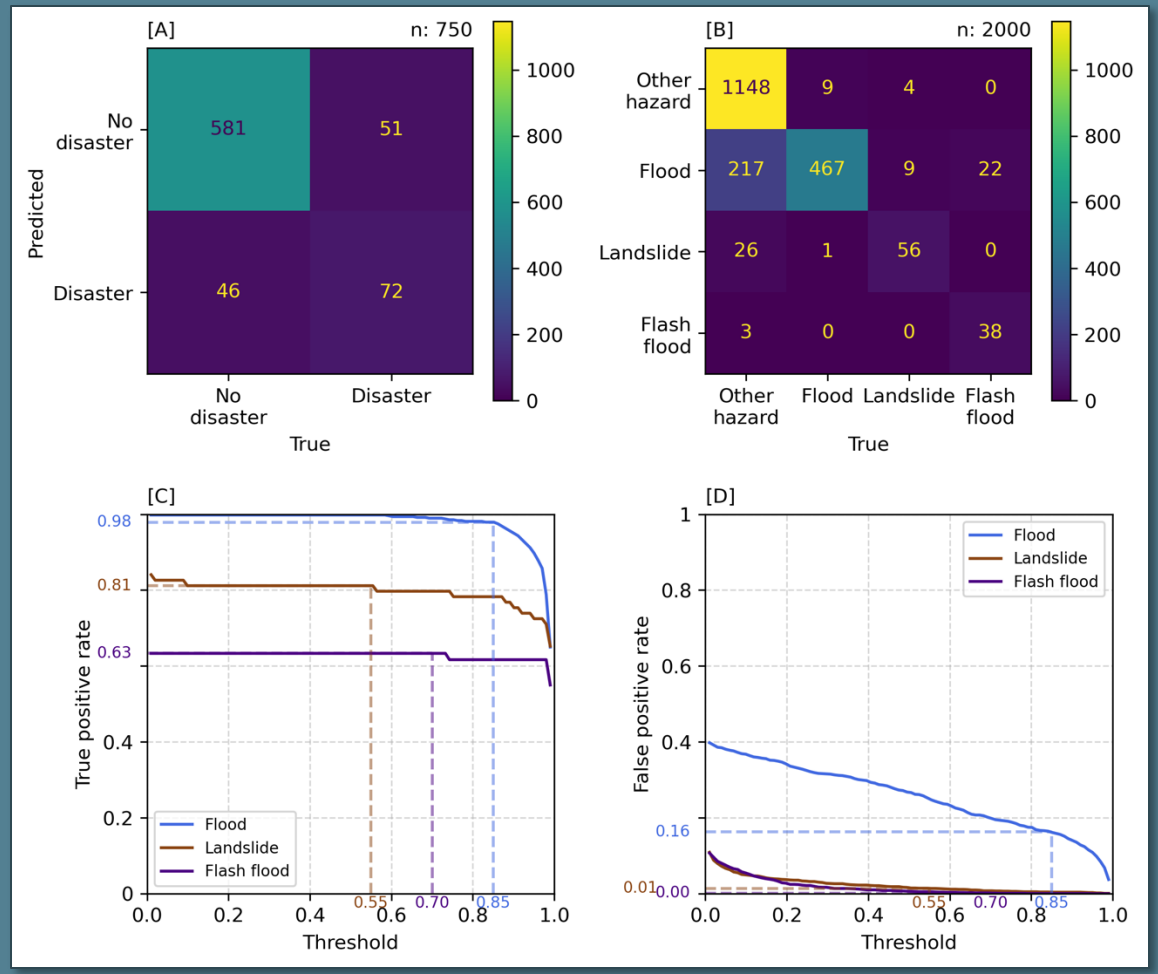
1. Filtering | 2. Geoparsing | 3. Impact

Manual validation

- Reading paragraphs
- Binary classification of hazard-related paragraphs
- Hazard classification

Manual calibration of LLMs

- Hazard dependent thresholds
- High true positive rate
- Low false positive rate



A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

Validation

Conclusion

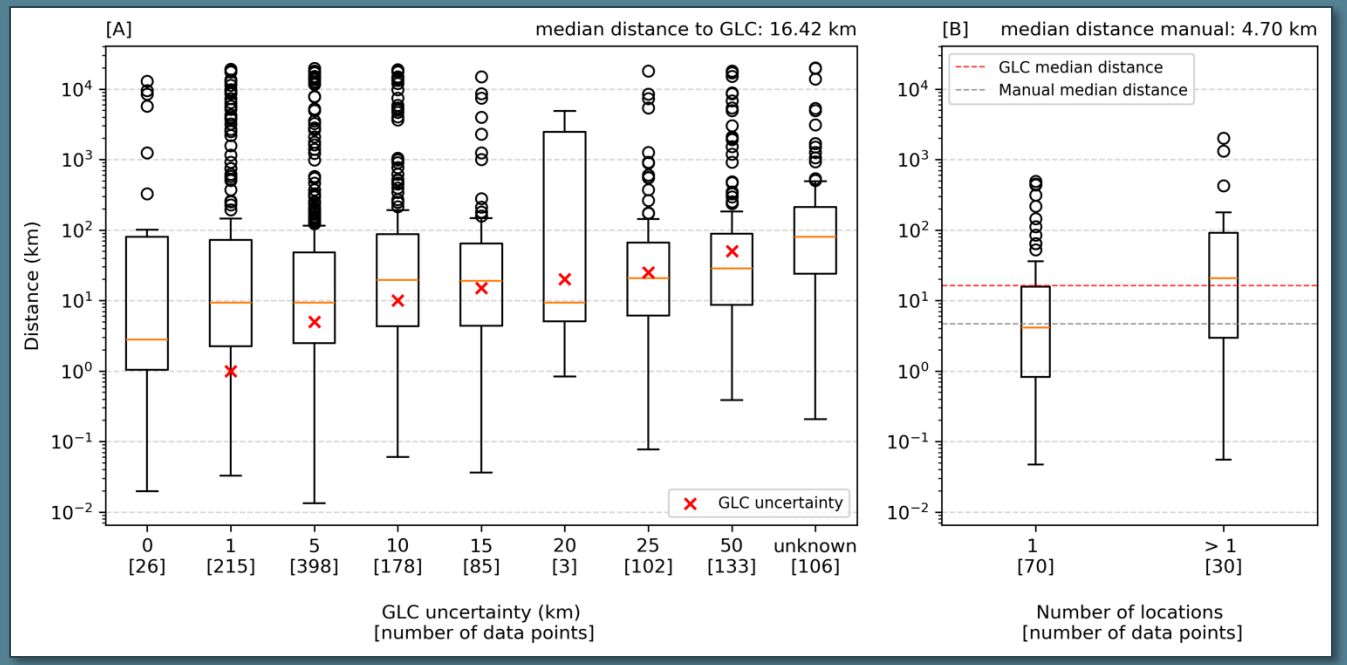
1. Filtering | 2. Geoparsing | 3. Impact

[A] Global Landslide Catalog (GLC) validation

- Geoparsing algorithm on GLC articles
- GLC: higher accuracy at low uncertainty
- Our data: higher accuracy at high uncertainty

[B] Manual validation

- Reading paragraphs
- Location of hazard events
- Google Maps
- Multiple locations: average



A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

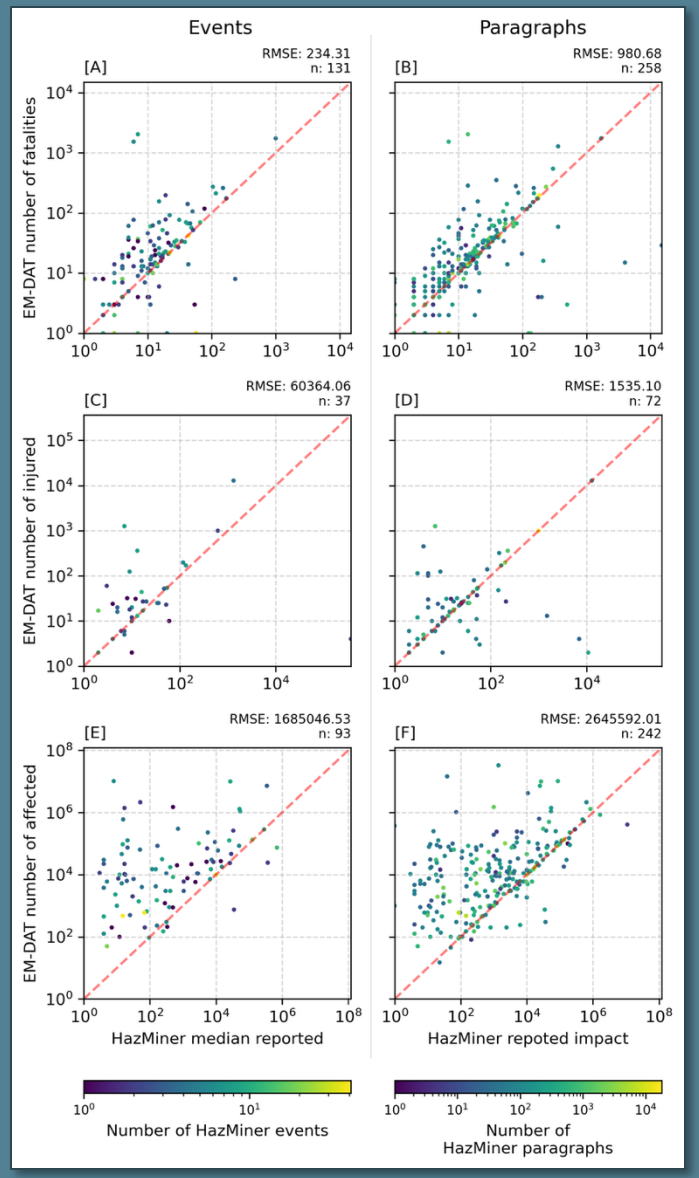
Validation

Conclusion

1. Filtering | 2. Geoparsing | 3. Impact

EM-DAT validation

- Fatalities, injuries, and affected
- Best match in time and space
  - Time: 5 days extra for HazMiner event, since there is a delay in impact reporting
- Impact reporting
  - EM-DAT: country level
  - HazMiner: event level
- Fatalities & Injuries: same trends
- Affected: underestimation by HazMiner



A multilingual tool for the documentation of impactful geo-hydrological hazards using online news articles: a worldwide application

HazMiner

Context

Method

Validation

Conclusion

Conclusion

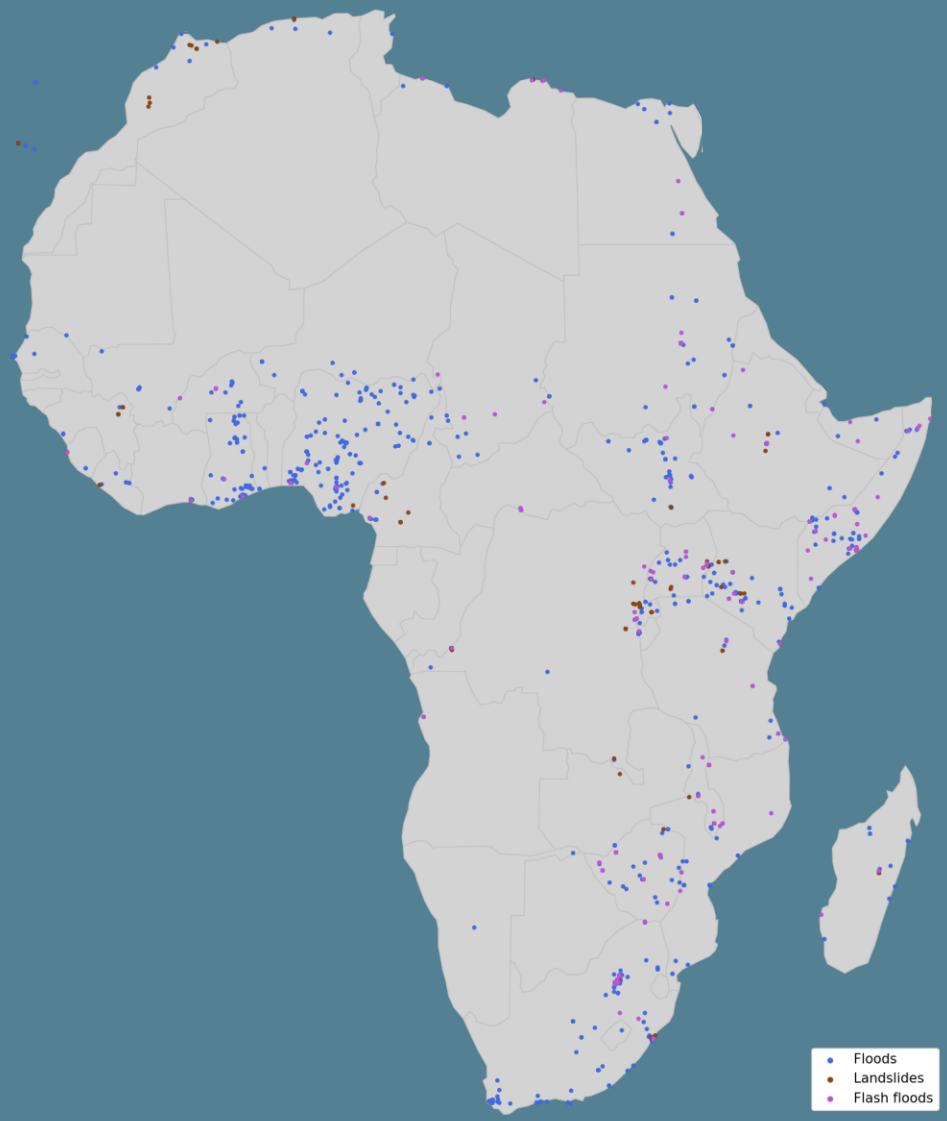
- Documentation of geo-hydrological hazards from news articles aligns with expected patterns and trends
- Analyzing at the paragraph level increases the level of detail extracted, but contains media bias
- Multilingual analysis increases the amount of information in non-English regions

General challenges

- Multilingual large language models
- Classification of hazards from text
- Improving impact detection

Next steps

- Publication of the method and dataset
- Python Package
- Is HazMiner able to fill the gap in data-scarce regions?



VUB VRIJE UNIVERSITEIT BRUSSEL
AFRICA museum