A severe local flood and social events show a similar impact on human mobility

Simone Loreti^{1,2,*}, Margreth Keiler^{3,4}, and Andreas Paul Zischg^{1,2}

¹University of Bern, Institute of Geography, Bern, 3012, Switzerland

²University of Bern, Oeschger Centre for Climate Change Research, Mobiliar Lab for Natural Risks, Bern, 3012, Switzerland

³University of Innsbruck, Department of Geography, Innsbruck, 6020, Austria

⁴Austrian Academy of Sciences, Institute of Interdisciplinary Mountain Research, Innsbruck, 6020, Austria

*To whom correspondence should be addressed. E-mail: simone.loreti@unibe.ch

Date: Wednesday 27th March, 2024

ABSTRACT

While a social event, such as a concert or a food festival, is a common experience to people, a natural disaster is experienced by a fewer individuals. The ordinary and common ground experience of social events could be therefore used to better understand the complex impacts of uncommon, but devastating natural events on society, such as floods. Based on this idea, we present a comparison - in terms of human mobility -, between an extreme local flood that occurred in 2017 in Switzerland, and social events which took place in the same region, in the weeks before and after the inundation. Using mobile phone location data, we show that the severe local flood and social events have a similar impact on human mobility, both at the national scale and at a local scale. At the national level, we found a small difference between the distributions of visitors and their travelled distances among the several weeks in which the events took place. At the local level, instead, we detected the anomalies (in time series) in the number of people travelling each road and railway, and we found that the distributions of anomalies, and of their clusters, are comparable between the flood and the social events. Hence, our findings suggest that the knowledge on ubiquitous social events can be employed to characterise the impacts of rare natural disasters on human mobility. The proposed methods at the local level can thus be used to analyse the disturbances in complex spatial networks and, in general, as complementary approaches for the analyses of complex systems.

Introduction

Being an innate need of human beings (1), mobility represents one of the essential components of human societies and their economic activities (2). Understanding the complex underlying mechanisms of human mobility is crucial for a wide range of applications, such as urban planning, traffic prediction, diseases spread, disaster evacuation and response (3-7). The growth and proliferation of technological innovations, as Internet, mobile phones (8–10), Wi-Fi, Global Positioning System (GPS), which led to an unprecedent availability of geo-referenced data on human mobility, enhanced our observations, assessments and modelling of human mobility dynamics (5, 11-13). In particular in the last decade, human mobility was primarily investigated through mobile phone data, i.e. the Call Detail Records (CDR) and Global Positioning System (GPS), the latter including geo-located social media (5, 7, 13-17). The need to evaluate policies and interventions during the COVID-19 period, and likely the recent concerns about a possible increase of climatic hazards (related to a worldwide rapid urbanization and climate change), spurred further the usage of mobile phone data for human mobility studies (13, 18). Forecasts about climatic hazards (weather and climate extreme events) refer to a possible increase in both their probability of occurrence and impacts (19-22), and can be justified by the sharp increase in the number of natural loss events - mostly climate-related disasters -, that occurred between the period from 1980 to 1999, and the following twenty years, i.e., from 2000 to 2019 (23, 24). A similar trend continued in 2020 and 2021, where climate and weather-related loss events and disasters represented the most frequent hazards (not considering the COVID-19 pandemics (25)), outstripping geological and technological disasters (26). More recently, in 2022, floods and storms had the highest number of occurrences, causing the largest economic losses, and droughts affected the largest number of people (27). Despite the abundance of studies that leveraged mobile phone data to analyse mobility patterns during disaster events, we still lack: (i) a universal framework for modeling the impacts of natural hazards related disasters, crossing geographical regions and different natural hazards related disaster events (28, 29), (ii) a full understanding of the complexity, dynamics, and interdependencies among social, economic and technical systems, which could be explored by integrating advanced techniques (e.g. AI) (6) and different data sources, such as CDR, social media and satellite imagery, (iii) an effective and simplified transfer of mobility insights to policy-makers and stakeholders (29), (iv) real-time tools to predict

mobility patterns during real-world disasters, (v) a comprehensive analysis of long-term recovery and resilience patterns, (vi) a transparent procedure for data collection and elaboration (13), and (vii) a clear comprehension of the fluctuations of human activities (30). Instead of comparing different disaster events among each other (28, 29), in this work, we compare one natural loss event, i.e. a flood, with social events, both occurred in Switzerland, between June and July 2017. Indeed, we argue that possible similarities or differences among scheduled artificial events and unscheduled natural loss event would help us in better understanding the transition between the laws of ordinary and unperturbed mobility (5, 31) and the laws of perturbed mobility (e.g. that one proposed by Li et al. (2022) (28)). By using mobile phone data provided by a major telecommunications company of Switzerland (i.e. Swisscom, with a market share of around 60%, at the end of 2017 (32)), we show that a severe local flood, causing more than 90 million Swiss Francs of property damage (33) (i.e. ~ 93 million USD, in July 2017), has an impact on human mobility similar to that one of major social events, as food festivals and concerts, both at a national and at a local scale.

Results

Similarities between social events and a local flood at a national scale

We consider the entire Swiss territory to describe the long-range impacts of medium-to-large social events (such as music and food festivals) and the impacts of a severe local flood, with a return period of 100 years, on human mobility (see Fig. 1(a)). The investigated social events took place between the 17th of June and the 30th of July 2017, i.e. in a time window of 44 days. The flood, instead, occurred on Saturday the 8^{th} of July 2017, i.e. exactly in the middle of the temporal window under investigation (see Supplementary Text and Supplementary Table 1 for details on the flood and social events, respectively). Both types of events occurred within a geographical region that measures around 20 km times 15 km, and that crosses three Swiss cantons, named, Solothurn, Aargau, and Lucerne. Throughout the entire manuscript, we will call that region as the local area of study, to distinguish it from the national area of study, the latter corresponding to the entire Switzerland. To capture potential differences of mobility among the diverse types of events, we could simply focus, for each event, on the exact days and hours in which a specific event occurred, and then measure the corresponding impacts on mobility. However, the examined events occurred both during weekends (most of them) and weekdays, i.e. during days that are usually characterised by different mobility patterns. Indeed, for example, people go to work and to schools during weekdays, while they are generally more involved in leisure activities during the weekends. In addition, the studied events lasted for different lengths of time, making the assessment of their impacts on mobility more challenging. One possibility to overcome potential biases could be to analyse the six weeks, composing the 44-day period, each week lasting from Monday to Sunday, both days included. This approach would require an observation period that is multiple of 7 (days), such as a 42-day period. We therefore neglect the first two days, a Saturday and a Sunday, of the examined temporal window of 44 days. For each of the six weeks, we then build two relative frequency distributions, displayed through histograms (12 histograms in total). One related to the number of visits (see Fig. 1(b), and Methods about the datasets details), and a second one related to the distances travelled by those individuals who visited the local area of study, during the 42 days (see Fig. 1(c), and Methods about the datasets details). The paths in Fig. 1(a) show that, during the entire period of 42 days, people arrived to the local area of study from all around Switzerland. Then, we compare the 6 distributions of visits – among themselves – pairwise (see Fig. 1(b)) and, lastly, we compare the 6 distributions of distances travelled, still pairwise (see Fig. 1(c)). Such comparisons represent, at a national scale, our idea of assessment of the impacts of both social and flood events, on human mobility. Indeed, in one of the examined weeks, we would anticipate a considerable perturbation of mobility, either in the number of visits or in the distances travelled by visitors, if the event responsible for such a perturbation were impactful enough. For example, we might expect a strong perturbation of mobility due to a severe flood, as that one occurring at the end of the 3^{rd} week. To measure such differences among weeks, both for the distributions related to the visits and the distributions related to the distances, we use a set of common statistical measures of distributional dissimilarity. The values returned by both the Chi-square test and the one-sample Kolmogorov-Smirnov test show that the null hypothesis is rejected at the 5% significance level, for each pair of estimated probability distributions P_i and P_j at different weeks, i.e. at i-week and j-week. This would indicate that the 6 distributions containing the counts of visits - to the local area of study -, would be different. The same consideration would hold for the 6 distributions representing the distances travelled by visitors to the local area of study. However, by using less strict measures of dissimilarity among distributions, the differences of mobility would appear less significant among the different weeks, implying a similar impact of social and flood events on human mobility. Indeed, the values returned by the Jensen-Shannon divergence, the Wasserstein distance, W_1 , the Hellinger distance, the total variation distance, the 1-Szekely's distance correlation and the 1-Cosine similarity (please see Methods for details), are similar for each pair of distributions, both for the number of visits (see Fig. 1(d)) and the travelled distances (see Fig. 1(e)). In addition, both graphics in Fig. 1(d-e) show very small values for all the above-mentioned metrics. The values related to the comparison among distributions at different weeks, and indicating the number of visits (see Fig. 1(d)), vary from 0 and 0.3, while the values related to the distributions representing distances covered by visitors at different weeks, vary from 0 and 0.1. Values equal to 0 would indicate identical distributions, while values equal to 1 would represent completely dissimilar distributions. The graphical representation of the above-mentioned metrics showed in Fig. 1(d-e), would correspond

to a matrix of vectors, $D = (d_{ijk})$, with i, j = 1, 2, ..., 6, indicating the estimated probability distributions P_i and P_j related to the *i*-week and *j*-week, respectively, and k = 1, 2, ..., 6 indicating the k^{th} -function of distance, or divergence, among the pair of distributions, that we can write as $f_k(P_i, P_j)$. Intuitively, the generic matrix element denotes how different P_i and P_j are, and it could be written as $d_{ijk} = f_k(P_i, P_j)$. In addition, the matrix elements are symmetric, i.e. $d_{ijk} = d_{jik}$ for each k^{th} -function, non-negative, i.e. $d_{ijk} \ge 0$, and they vanish on the diagonal, i.e. $d_{ii} = 0$, making the matrix D a dissimilarity matrix (of vectors) (34-36).

Similarities between social events and a local flood at a local scale

We now focus on the number of people who travelled each road or railway lines every hour, during the entire period of 44 days, and within the local area of study (please see Methods for details about the datasets). Those intra-town or intra-city travels typically follow the daily circadian rhythm (37), leading to regular variation of the number of travellers on each road and railway track. When a disruptive event - like a social or natural (loss) event - occurs, those variations might become irregular for a while, returning to the unperturbed state afterwards. If we consider the temporal evolution of the number of travellers as a "time series", and the irregular variations of the number of travellers as "anomalies", then, a disruptive event would cause anomalies in a time series. We therefore use the anomalies discovered in travelers-related time series to measure the impact of social and flood events on human mobility, at a local scale. Clearly, an anomaly can represent either an excess of travelers (e.g. overflow of traffic) or a dearth of people (e.g. empty roads due to road closures or disruptions), but in this manuscript we do not deal with the two types of anomalies separately, i.e. we do not specify if an anomaly is related to a surplus or to a lack of travelers. Indeed, we are interested in the overall patterns of anomalies, without a distinction between the different types of anomalies. Among the most relevant events that we found within the local area of study and throughout the 44 days, seven of them are distinct social events, named Aargau Cantonal Shooting Festival (ACSF), Bio Marché, New Orleans Meets in Zofingen (NOMZ), Tattoo concert & Children Festival, Ski Festival, Beach Festival and Summer Party, and only one of the largest events is a distinct natural event, i.e. the flood of the 8th of June. With "distinct events" we mean events that occured singularly, without any temporal overlap with other significant events. A few of the most significant events were instead "multiple events", i.e. different major events that occurred at the same time, or with some temporal overlap (details about all the social events are contained in the Supplementary Table 1). Among the distinct social events, ACSF is the only one that occured in 7 different municipalities. This means that the corresponding attendees did spread out over 7 municipalities, instead of gathering in one municipality as in the other distinct social events. Therefore, due to the large dispersion of its attendees, we do not consider ACSF in the comparison among major events, throughout this work. For both distinct and multiple events, we detect anomalies with the Discord Aware Matrix Profile (DAMP) algorithm, a numerical approach based on the well-recognised concepts of "discord" – that defines the notion of anomaly -, and "matrix profile" - a structure for data storage. DAMP requires some training data and the period of our times series, called subsequence length, which is calculated automatically by the algorithm (please see Methods for details). We set a training period equal to 6 days, i.e. from Saturday 17^{th} of June to Thursday 22^{nd} of June 2017, both included. This means that the anomalies start to be detected by DAMP just after the midnight between the 22^{nd} of July and the 23^{rd} of July 2017. In addition, for each road or railway track, DAMP seeks for the largest anomalies, up to 10, if present.

We observe that the largest amount of anomalies is detected both during the course of all the major recorded events, and on Sundays in which no significant event was reported (please see Fig. 2(a)). Since the majority of the most significant recorded events took place over the weekends (among them, only NOMZ and the Tattoo Concert that took place on weekdays), we can therefore say that the anomalies are concentrated around weekends, rather than on weekdays. Please note that this statement would be still valid even including a build-up of anomalies occurring on Tuesday 25^{th} of July, which origin is unknown to us (and that we did not mention so far). Interestingly, we note that all Sundays are characterized by a peak of anomalies in the morning, at 08 a.m. At least partially, the peaks on Sunday 25^{th} of June and on Sunday 2^{nd} of July are due to, respectively, Bio Marché in Zofingen, and the School & City Festival in Olten (which started in the morning, around 07 a.m.). However, no relevant documented events took place on the other Sundays, early in the morning. We therefore associate those peaks of anomalies, that regularly happen at 08 a.m. every Sunday, to recurrent leisure activities, typical of weekends (38–40). Hence, if every Sunday features regular leisure activities, regardless the occurrence of any event, when an event then takes place on Sunday, it would contribute to increase the peak of anomalies. This would be quite evident for the peak of anomalies observed on Sunday 25th of June at 08 in the morning, during Bio Marché. This observation would suggest that a relationship exists between the number of anomalies and the number of attendees, but it would be difficult to find an accurate one, since we just have rough estimates of the number of participants to the events (and not even for all the most significant events). However, if we consider the largest local maxima in Fig. 2(a) and the corresponding approximate numbers of attendees in Supplementary Table 1, we see that the largest local maxima in the distribution of anomalies are directly proportional to the estimated numbers of attendees. Still looking at the local maxima we can notice that the height of the flood peak (8^{th} of July) is very similar to the peaks' height of both the Tattoo & Children Festival on Friday 7^{th} of July (with ~ 1000 attendees) and of the Ski Festival

on Saturday 15^{th} of July (with ~ 1500 attendees). Also, the height of the flood peak (8^{th} of July) is larger than the peaks' height of both NOMZ (3^{rd} of July) and Tattoo & Children Festival on Thursday 6^{th} of July, and smaller than all the most prominent peaks' height of Bio Marché (23^{rd} -to- 25^{th} of June, with ~ 12000 attendees per day), Beach Festival (22^{th} of July, with ~ 5000 attendees), and of the multiple events that occurred the 1^{st} of July (with ~ 6500 attendees). Therefore, such a visual comparison among the local maxima in Fig. 2(a) allow us to identify a similarity between the impact of social events, with an estimated number of attendees between ~ 1000 and ~ 1500 individuals, and the impact of a local flood, on human mobility.

Although the sole number of anomalies might give already an indication on the similarities between the social events and the flood, it does not convey any spatial information, except the name and border of the municipality where an event took place. As a first spatial information about the roads and railways' anomalies, we observe - for example from the Supplementary Figures 1-4 – that the anomalies are scattered all over the local area of study, with varying density, and showing different patterns every hour. Given such variations of density, we hypothesize that a sufficiently significant event would be responsible for the occurrence of anomalies mainly in the surroundings of the same event. In other words, we expect that anomalies group around the location where an event takes place (for example, please see in Supplementary Figure 2, in the time range between 6 p.m. of Thursday 6th of July and 1 a.m. of Friday 7th of July, and in Supplementary Figure 3, between 7 and 8 p.m. of Saturday 22^{nd}) Since in data mining and machine learning the groups of data-points in (high-dimensional) data spaces are called "clusters" – which are detected with techniques of data clustering – (41), we define the collections of nodes that represent the groups of roads and railway tracks with anomalies as "clusters". At the same time, roads and railways tracks represent the edges of the corresponding transport networks, and the word "community" could be used here to indicate a group of edges that are "densely connected" within the community and "sparsely connected" to the other communities (42). This would be a common terminology in network science, where a group of nodes is called community if those nodes are more densely connected within the community than with the rest of the network (42). But then, can we call those groups of roads or railway tracks with anomalies as communities, if densely connected? Still from Supplementary Figures 1-4, we notice that some edges with anomalies are relatively far from the rest of the edges with anomalies, and they could be then considered as "outliers". Indeed, those isolated anomalies could be caused by a single road disruption, or by a single technical issue on the railway line, and not related to any major event. Also, we can notice that relatively close edges with anomalies, might not have common nodes, but the corresponding anomalies could have been originated by the same event. The latter observation would mean that those edges with anomalies might belong to the same community (i.e. resulting from the same event), but being internally disconnected within it, i.e. a part of the community could be reached only through a path outside the community. Although there is no universally accepted definition of community in network science (43, 44), the connectedness among nodes is considered as a basic requirement to define a community (43), and internally disconnected communities are generally deemed to be bad partitions of a network (45, 46). Therefore, we could use the noun community if we extend its definition (44) to those ones that might be internally disconnected, or if we just consider definitions of community based on node similarity (43). In this way, the terms "cluster" and "community" might be used interchangeably.

By following our observations on the detected anomalies positions, we use RNN-DBSCAN (47) – a data clustering algorithm - as a density-based community detection technique (48, 49). Indeed, RNN-DBSCAN can detect disconnected communities of arbitrary shapes and various densities, and filter noise, as those isolated edges with anomalies that would unlikely result from major events (50, 51). Once the clusters are detected, we can assess our hypothesis (that anomalies occur around the events) for every event, by measuring the distances between the clusters centers and the event center. A cluster center might be calculated by simply averaging the coordinates of the cluster's data points. However, the resulting center would be influenced by the number of data points, and would be driven towards denser areas of data points. Thus, we draw a tight 2-D boundary (52) that surrounds the cluster's data points, and we calculate the centroid, or center of mass, of the resulting polygon (53). A centroid is indeed less influenced by different densities of the cluster's data points (53). For each event, the resulting distribution of the distances among the clusters centers and the event center is shown in Fig. 2(b), while the distribution of the clusters areas (i.e. the polygon areas) is illustrated in Fig. 2(c). Please note that some of the events in Fig. 2(b-c) took place over two or three days, and, therefore, the number of distributions is larger than the number of distinct events. Alongside the distributions we show the corresponding boxplots, which provide descriptive statistics and facilitate a visual comparison among (the distributions of) different events. Probably, the first thing that strikes the reader about the boxplots in Fig. 2(b) (related to the distributions of the distances) is that 7 out of 11 boxplots are remarkably similar to each other, i.e. those ones related to Bio Marché (all days), Tattoo concert & Children Festival (on Friday the 7th of July), the flood, and to the Ski Festival (both days). We used the expression "remarkably similar" since the boxplots look very similar to each other, despite the great difference in the estimated number of attendees who participated in those events. Indeed, ~ 12000 people were estimated for each day of Bio Marché, ~ 1000 people for the Tattoo concert & Children Festival (on Friday the 7th of July) and ~ 2000 people for the

Ski Festival. But, how could events of different magnitude (i.e. expressed through the number of attendees) lead to similar spatial dispersions of anomalies (i.e. similar boxplots)? We can exclude any reason related to the process of clustering the edges with anomalies through RNN-DBSCAN, since the boxplots in Fig. 2(b) are very similar to the boxplots in Supplementary Figure 5 (related to the distributions of distances among the edges with anomalies and the event centers, i.e. before any usage of RNN-DBSCAN). We then suppose that the spatial dispersion of anomalies might depend on the network topology, since it is the same for all the events, rather than on the number of attendees. In particular, all the events with similar boxplots, took place in the same town (Zofingen), except the Ski Festival which was held in a town around 5 km far, in a beeline (Rothrist).

Still comparing the boxplots in Fig. 2(b) we notice that the boxes representing the interquartile ranges of NOMZ and the beach festival, are slightly lower than the previous ones. However, we might still include those two events in the previous list of similar events (i.e. Bio Marché, Tattoo concert & Children Festival on Friday the 7th of July, flood, and Ski Festival). Indeed, if we use the "average median" calculated over the initially mentioned 7 boxplots, i.e. $\langle Q_2 \rangle \approx 5.38$ km, as one of the reference measures to uniquely characterize all the similar events, the medians related to NOMZ and the beach festival, i.e. $Q_2 \approx 4.05$, can be still considered reasonably close to the average median. The boxplots related to the Tattoo concert (on Thursday the 6th of July) and to the Summer Party, instead, look quite dissimilar to the previously mentioned boxplots, and their medians, which are $Q_2 \approx 2.30$ km and $Q_2 \approx 8.86$ km, respectively, are farther from the average median, than the NOMZ and beach festival's medians. About the dissimilarity of the boxplot related to the Summer Party, we notice that a large number of anomalies detected during that event are quite far from the site where that event took place. The large distances between the edges (or clusters) of anomalies and the event location is then reflected in the values of the boxplot which are larger than all the other boxplots values. The dissimilarity of the Summer Party's boxplot might be due to the event size (~ 500 attendees) which was not sufficiently large for producing a considerable number of anomalies in the surroundings of the event, and at the same time, another event might have occurred, about which, however, we did not find any record. Overall, the boxplot associated to the Tattoo concert (on Thursday the 6^{th} of July) is the "lowest" one in Fig. 2(b), i.e. it has the smallest boxplot's summary statistics (except the minimum) than the other boxplots. This case would support our hypothesis about the occurrence of anomalies mainly in the surroundings of an event. However, why aren't the other boxplots closer to distance zero as the boxplot related to the Tattoo concert? Or, other way around, if we take the 7 similar boxplots as reference, why does the Tattoo concert's boxplot is considerably lower than them? We know that the number of trips (i.e. trip frequency) is usually higher during weekdays than during weekends (54–57) (with some exception (58)), and that weekday trips are generally shorter (56, 57, 59, 60) and more regular (61-66) than weekend trips. Despite there is some discordant results on trips regularities (60, 67), we think that the relative stability and the short distances of weekday trips limit the occurrence of irregularities, i.e. the anomalies. Therefore, given a less occurrence of anomalies during the weekdays, when a significant weekdays event takes place, most of the anomalies would be originated by that event. This means that, generally, a weekdays event would be more distinguishable - i.e. showing a clear(er) grouping of anomalies around the event - than other events occurring during weekends. This is the case of the Tattoo concert (on Thursday the 6^{th} of July), which is the only major event that took place on a weekday not belonging to "long weekends" (i.e. Fridays and Mondays). Here, we exclude Fridays and Mondays, since the Monday travel behaviour might be similar to the weekend one (63), and the patterns of activity-travel behaviour on Fridays could be dissimilar to the behaviour of other weekdays (65, 68, 69). Now, if an event that takes place on weekdays shows clearer patterns of anomalies around and close to the event due to the stability of weekdays trips, once an event occurs during the weekend or in the run-up to the weekend (or either side of the weekend), the clusters of anomalies originated by the major event might be partially overlapping or mixed with the clusters of anomalies caused by irregular and long-distance trips, not necessarily related to the event. Therefore, the spatial distribution of the (clusters of) anomalies would include the irregular travel behaviour, occurring potentially anywhere. Consequently, once we show the spatial distribution of anomalies related to a major weekend event, this would be larger than the distribution of anomalies related to a weekday major event. However, despite a possible mix of anomalies resulting from major events and anomalies given by unpredictable weekend trips, in Supplementary Table 2 we notice that almost all the spatial distributions of anomalies' distances (in Fig. 2(b)) show a general tendency to skew towards the event centers (i.e. positive skewness). This is another commonality between the flood and social events. Indeed, the Bowley's coefficient of skewness (BS), and the Fisher-Pearson coefficient of skewness (FPS) in Supplementary Table 2 are positive for all the events, with the only exceptions of the Ski Festival on Saturday the 15^{th} of July, where BS and FPS have discordant sign, and the Summer Party, where both BS and FPS are negative. Perhaps, such negative signs for both the BS and FPS might be interpreted as a lack of connection between the detected anomalies and the Summer Party. The reason why we use both BS and FPS is that, often, (strongly) skewed distributions have skewed boxes. Indeed, the Bowley's coefficient assesses the skewness of the central box, while the Fisher-Pearson coefficient quantifies the skewness of the entire distribution. However, the skewness should generally refer to the entire distribution, which would bring our attention primarily to the Fisher-Pearson coefficient (even though the Bowley's skewness is preferable in presence of outliers (70)). As possible qualitative descriptions, some authors (71, 72) suggest that distributions need to have a $|FPS| \ge 1$ to be considered

significantly skewed. Therefore, we can say that only the Tattoo concert (on Thursday the 6^{th} of July) and the Beach Festival are fairly skewed (with FPS ≈ 0.97 and FPS ≈ 1.25 respectively), while the flood and the Ski Festival on Saturday 15^{th} of July are instead moderately skewed (with FPS ≈ 0.54 and FPS ≈ 0.56 , respectively). All the other distributions show a light positive skewness, with $0.05 \leq \text{FPS} \leq 0.39$ (excluding the Summer Party). However, values of FPS less than 0.5 are often associated to approximately symmetric distributions.

While the distributions of the distances exhibit a light-to-moderate departure from symmetry (or they are approximately symmetric), all the distributions of the clusters areas – including the flood-related one – show a very high positive skewness, with a large concentration of values below the median Fig. 2(c) (please see Supplementary Table 2 for the exact values of BS and FPS). This means that (at least) half of the detected clusters are small or very small, in comparison to the entire range of values included in the boxplots. If we look at the medians of all the events, we can see that the approximate average median and the largest one would be, respectively, ≈ 0.21 km² and ≈ 0.33 km², if we do not include the Beach Festival (otherwise they would be, respectively, 0.23 km^2 and 0.46 km^2). Indeed, although the difference between the second quartile (i.e. median) and the first quartile in the Beach Festival is similar to other events, the first quartile of the Beach festival is the largest one among all the events ($Q_1 \approx 0.18 \text{ km}^2$), and it is around double of the second largest first quartile, which belongs to the Bio Marché (on Friday 23^{th} of June). Therefore, the clusters of anomalies in the lowest 25% of the Beach festival's distribution, reach at least twice the size of clusters in the lowest 25% of other events' distributions. In other words, the lowest 25% of the Beach festival's distribution shows a larger range of cluster sizes than in the lowest 25% of other events' distributions, where the clusters have instead a similar size among each other. Another feature that stands out in Fig. 2(c) – and that demans some attention – is the large quantity of boxplots outliers. Indeed, although some outliers are associated to proper and correct assessments of the polygon sizes, other ones might be related to overestimates of the polygon sizes, and should therefore be interpreted with caution. The latter case occurs when long network's edges, i.e. railways or motorways, have anomalies, and they are included into polygonal shapes, i.e. the clusters of anomalies (from which we calculate the clusters areas). Basically, the long edges stretch the polygons towards themselves, leading to an increase of the polygon area.

Although the boxplots in Fig. 2(b) and Fig. 2(c) provide a synthetic summary of the distributions of the clusters distances and areas, they do not directly represent variance or other measures of data variability. It is indeed a common practice to report mean and standard deviation for symmetric or approximate symmetric distributions, as in Fig. 2(b), and median and interquartile range for skewed distributions, as in Fig. 2(c) (73, 74). In addition, so far, our evidence of similarity among the boxplots in Fig. 2(b) was mainly based on a visual inspection, which could be less robust than a numeric comparison. We therefore calculate and compare a set of measures of dispersion, both based on the mean, i.e. standard deviation (STD) and mean absolute deviation (MeanAD), and based on the median, i.e. interquartile range (IQR) and Rousseeuw-Croux scale estimators (S_n and Q_n). If we focus on Fig. 2(d), we notice that our previous observation related to the similarity among 7 out of 11 boxplots – i.e. those ones associated to Bio Marché (all days), Tattoo concert & Children Festival (on Friday the 7^{th} of July), the flood, and to the Ski Festival (both days) -, is still fairly valid regardless the measure of dispersion used, and despite the slightly larger values of the Ski Festival (both days). Also, we mentioned that the boxplots related to NOMZ and the Beach Festival might still be considered similar to the previous 7 boxplots, in terms of dispersion. This is quite true for NOMZ, if we consider all the measures of dispersion, except the interquartile range. About the Beach Festival, the answer depends on the measure of dispersion. If we do not consider both the interquartile range and the standard deviation, the Beach Festival shows a spatial dispersion similar to the previously mentioned 7 events. In addition, the standard deviation is here large due to presence of outliers on the right tail (please see Fig. 2(b)). About the remaining two events, i.e. the Tattoo concert (on Thursday the 6th of July) and the Summer Party, they still look dissimilar to the previous events for any measure of dispersion (and not just for the interquartile range), showing, respectively, significantly lower and rather higher values of dispersion. Therefore, if we use any of the proposed measures of dispersion, except the interquartile range (that yields the largest excursion of values among all the measures of dispersion), we can still fairly assert that 9 out of 11 events are approximately similar to each other, in terms of spatial variability of anomalies. But then, should we avoid any usage of the interquartile range in Fig. 2(b)? Although there is nothing stopping us from employing the interquartile range, this measure of dispersion is not generally applied to approximate symmetric or light-to-moderate skewed distributions as in Fig. 2(b), since it does not describe the entire distribution, as the standard deviation or the mean absolute deviation would do, but only the central half of the distribution. The usage of the interquartile range is instead advised in cases of skewed distributions, as in Fig. 2(c). In particular, with the presence of outliers, which significantly affect the mean and all the mean-based measures of dispersion (in this case, the standard deviation and the mean absolute deviation). Indeed, it is the standard deviation that yields the largest (and the broadest range of) value(s) in Fig. 2(e), followed by the mean absolute deviation. On the contrary, the interquartile range, and in particular the Rousseeuw-Croux scale estimators, exhibit more robustness against the outliers, and thus less variations among the events. The fact that limited variations occur across the different events, might indicate a certain degree of similarity among

the events (as between the flood and the social events), in terms of variability of the clusters areas (in particular the values provided by the Rousseeuw-Croux scale estimators). Interestingly enough, the Rousseeuw-Croux scale estimators yield values of dispersion which are very similar to the standard deviation and the mean absolute deviation ones in Fig. 2(d), and, as we have just mentioned, they provide some of the most robust values of dispersion in Fig. 2(e). This means that Rousseeuw-Croux scale estimators might be used as common tool for both types of dispersion, i.e. the dispersion of distances and the dispersion of the clusters areas.

Empirical relationships among the clusters of anomalies

The distributions in Fig. 2(b) and Fig. 2(c) provide us a coarse-grained description of the events, since they include, respectively, all the distances and all the areas measured during the entire events, i.e. from the starting hour to the ending hour of the event. With this approach we therefore loose the details related to every single hour (in any event), that might be useful to identify possible relationships among distances, areas, and number of clusters. Indeed, for (every event and) every hour, we know the number of clusters detected by RNN-DBSCAN, as well as their distances from the event center, and their areas. We can then extend our local scale-related analysis by using these hourly-based data as fine-grained information. If we consider for instance the distances between the event center and the farthest and closest clusters, i.e. r_{max} and r_{min} , respectively, we observe that r_{max} and $\Delta r = r_{max} - r_{min}$ are connected by a a linear relationship (please see Fig. 3(a-b)):

$$r_{\max} = a_1 \Delta r + b_1 \tag{1}$$

where a_1 and b_1 are listed in Supplementary Table 3, for each event. Although Equation (1) does not show a direct relationship between r_{max} and r_{\min} (since we prefer to use Δr instead of r_{\min} , for reasons of clarity), we can always transform it into $r_{\max} = \left(\frac{a_1}{a_1-1}\right)r_{\min} - \left(\frac{b_1}{a_1-1}\right)$, but bearing in mind that it would be undefined for $a_1 = 1$ (i.e. we can use it only when $a_1 \neq 1$). By consulting Supplementary Table 3, we notice that both slopes corresponding to NOMZ and the Beach Festival are $a_1 \approx 1$. This means that – for both social gatherings and their entire duration – the closest cluster has a constant distance from the event center, $r_{\min} = b_1$ (see the red line in Fig. 3(a)). All the other slopes, instead, are smaller than the NOMZ and the Beach Festival's ones, being $0.45 \leq a_1 \leq 1$ (here, we exclude that one related to the Summer Party, due to the low coefficient of determination in Equation (1), i.e. $R^2 \approx 0.37$). In this case, if we read from left to right the Fig. 3(a-b), not just the farthest cluster goes farther, but the closest cluster gets closer to the event center as well (see the blue line in Fig. 3(a)). In other words, the space occupied by all the clusters, and identified by Δr , increases radially towards both directions, i.e. inwards and outwards. And, this behaviour is common to the vast majority of the events here presented, flood included. Obviously, the upper bound for any cluster's distance is given by the perimeter of the local area of study, while the lower bound is zero, and it is reached when the cluster center overlaps the event center (see the black line in Fig. 3(a-b), i.e. any point on the $r_{\max} = \Delta r$ line, also called the $r_{\min} = 0$ line).

About the number of clusters, n_c , we observe that it decreases with the increase of the mean area of the clusters (averaged every hour), $\langle A_c \rangle$:

$$n_c \sim \frac{1}{\langle A_c \rangle}.\tag{2}$$

which might be somehow obvious. However, the way the number of clusters decreases might not be so trivial. We see that some events exhibit a power law, $n_c = a_2 \langle A_c \rangle^{b_2}$, where the exponent ranges around $b_2 \approx -0.3$ (more precisely between $-0.45 \leq b_2 \leq -0.26$) for those events that have a relatively high coefficient of determination (i.e. Bio Marché, on Saturday 24^{th} of June, with $R^2 \approx 0.6$, Flood with $R^2 \approx 0.65$, and Ski Festival, both days, with $R^2 \approx 0.94$ and $R^2 \approx 0.68$, respectively). Still focusing on the areas of the clusters, we also observe a mild direct proportionality among the distance of the closest cluster, $d_{c(r_{min})}$, and its area, $A_{c(r_{min})}$:

$$d_{c(r_{min})} \sim \log(A_{c(r_{min})}) \tag{3}$$

In particular for Bio Marché (both Friday 23^{rd} and Saturday 24^{th} of June), the Tattoo & Children Festival (on Thursday 6^{th} of July) and the Summer Party (on Saturday 29^{th} of July), we can write a logarithmic relationship, i.e. $d_{c(r_{min})} = a_3 log(A_{c(r_{min})}) + b_3$, where the coefficients are shown in the Supplementary Table 3.

Anomalies in the framework of multi-layer transportation networks

A multi-layer network (75, 76) is a framework consisting of coupled networks that are connected (to each other) through inter-links, and it allows to approach the complex reality of interconnections and interdependencies among the diverse networks, typical of real-world systems. Although the literature on coupled multi-layer networks is extensive (77-85), there is a scarce

evidence about the usage of multi-layer approaches for investigating the impact of floods on transportation networks (86, 87). However, this framework was employed to model multimodal transportation (88–91), where each mode of transportation (e.g. bus, tram, metro) was represented by a distinct network (layer). A typical analysis in multi-layer networks is related to their robustness, i.e. one investigates if the network functionality is preserved during random failures or targeted attacks (e.g. starting from nodes with highest degree or betweenness), up to the shattering of the coupled networks into small clusters. Once fragmented, a multi-layer network might still maintain a certain degree of functionality within the "mutually-connected components", i.e. when there is an intra-layer path between two nodes in all of the intra-layer networks (76, 82). Here, we do not perform any analysis of robustness (i.e. percolation) among the coupled networks, but we just focus on the final stage of the process of networks fragmentation, i.e. on all the possible smallest mutually connected components, where the change of transport mode and the return to the initial starting place are guaranteed. As shown in Fig. 4(a), in our case we have three layers, i.e. rails, urban roads, and motorways, and the mutually connected components are represented by the differently coloured "cycles". In this framework, interlinks would be "functionally interdependent" (92), and would start both at train stations (for the railway network) and at the motorways entrances or exits (for the motorways layer). What is surprising here is the presence of anomalies (detected with DAMP) in the smallest mutually connected components, and for all the events (see Fig. 4(b)). Indeed, the mutually connected components should be the smallest elements in networks where the functionality should be preserved. Although anomalies could represent both a congestion or a smaller amount of people, in case of congestion, the functionality would be compromised. Also, it is interesting to notice the similarity among the flood and other social events, in terms of mean number of anomalies, for both rail-urban roads cycles and motorways-urban roads cycles. Interestingly, all the events except the Tattoo & Children Festival (on Thursday 6th of July), have averagely more anomalies in the rails-urban roads cycles than in motorways-urban roads cycles. This might be interpreted as a larger impact of the events on the railway network, than on the motorways layer. This aspect would also make the flooding event look similar to the social events. The Tattoo & Children Festival (on Thursday 6^{th} of July) is not just the only event that impacts slightly more the motorway network than the railway one, but it also exhibits the lowest amount of anomalies in both rails-urban roads cycles and motorways-urban roads cycles among all the events in this study. This feature means that the event has a fairly small effect on both motorways and railways. In this case, the multi-layer approach for assessing the robustness of coupled networks might not be necessary (since the amount of anomalies in both types of cycles is quite small), while the traditional measures of robustness for single layers (monoplex) might be more appropriate, or at least sufficient.

Conclusions

In this work, we presented a comparison between a severe flood event, i.e. a flood that occurred in 2017 in Switzerland, and major social events, which took place in the weeks before and after the flood, but in the same geographical area, that we called as the "local area of study". We showed that, both at a national scale and at a local scale, a local but highly damaging flood and social events have a similar impact on human mobility. In particular, at the national scale, we compared the weekly-based distributions of visits and the weekly-based distributions of distances travelled by visitors, who reached the local area of study, and we found minimal differences among them (see Fig. 1(b-e)). At the local scale, we instead focused on the irregularities – i.e. the anomalies – of the number of people who travelled each road or railway line, within the local area of study, as well as on the spatial clusters of those irregularities. Surprisingly, we observed that the largest number of anomalies that occurred during the flood (i.e. the local maxima in the distributions of anomalies) is comparable to those ones of major social events (see Fig. 2(a)). In fact, one might expect that a flood would easily surpass other social events, in terms of number of anomalies (related to mobility). About the clusters of anomalies, that we detected every hour and for each event, we calculated both the distances between the clusters centers and the event center, and the areas of those clusters. We then compared the flood with the social events through the resulting distributions of clusters-event distances (see Fig. 2(b)) and the distributions of clusters areas (see Fig. 2(c)), finding compelling similarities between the corresponding summary statistics, i.e. boxplots and measures of skewness and dispersion (see Fig. 2(b-e)). For example, we found that the flood-related distribution of distances (see Fig. 2(b)) is positively skewed as most of the social events-related ones (which exhibit a light-to-moderate skewness), meaning that the anomalies tend to group around the event center. Still at a local scale, we found that all the events, except the Summer Party, follow a linear empirical relationship between the distances of the farthest and the closest clusters of anomalies (see Equation (1)), allowing us to distinguish two types of spatial distributions of clusters. One, where the distance between the closest cluster and the event center is constant during the event (at NOMZ and at the Beach Festival), and a second type, where the distance of the closest cluster varies during the event. We observed the second behaviour for most of the events, meaning that the flood is similar, in this aspect, to the majority of the social events. A last comparison that we performed among the flood and social events was related to the framework of multi-layers, in network science. Indeed, we found that, during the flood, the average number of anomalies in the smallest mutually connected components of the multi-layer networks (i.e. that we called "cycles"), is comparable to the average number of anomalies occurring during the social events (see Fig. 4). We also found that the average number of anomalies is larger in the rails-urban roads cycles than in motorways-urban roads cycles, in

all the events except the Tattoo & Children Festival (on Thursday 6^{th} of July), denoting another commonality between the flood and the social events. In addition, regardless the comparison among different events, the presence of anomalies in the smallest mutually connected components of a multi-layer network is a surprising fact by itself, since those components should preserve the functionality of the coupled network.

All our observations were supported by a robust methodology. At the national scale we used six measures of (dis)similarity among distributions (please see Fig. 1), while at the local scale we used two state-of-the-art algorithms of data mining and machine learning (i.e. DAMP and RNN-DBSCAN) which require, overall, only one parameter. Indeed, to detect the anomalies in the number of people who travelled each road or railway lines (please see Figs. 2–4), DAMP uses a set of training data, which size needs to be selected by the user before running the algorithm (and only once). RNN-DBSCAN, instead, employs a heuristics approach to automatically select the single required parameter, and to ultimately find the clusters of anomalies (please see Figs. 2 and 3). Despite the solid methodology, this work has limitations, starting from the mobile phone data. Indeed, our mobile phone data represent only around 60% of the entire set of Swiss mobile phone data, and this is due to the market share of Swisscom (at the end of 2017, Swisscom held 60% of the entire Swiss mobile network (32)). Also, to preserve user anonymity, the mobile phone datasets provided by Swisscom were hourly-based and did not include the number of people travelling a certain road or a railway line, if less than 20 travellers were observed. Arguably, a hourly time granularity might be considered a little coarse for social or natural events. Another limitation of this study is related to the lack of documentation (mainly on the Web) about the minor events that took place in the local area of study (e.g. a road closure due to maintenance works). Although we were able to identify the origin of almost all the anomalies (see Fig. 2(a)), a few anomalies remain undocumented, and are likely related to minor events. Also, the presence of undocumented minor events during the course of a major event, could be the cause (at least partial) of the dissimilarity of the Summer Party's boxplot in Fig. 2(b). Indeed, the presence of (clusters of) anomalies, relatively far from a major event, and likely related to unknown (since undocumented) minor events, would affect both the distribution of distances between the clusters center and the event center, and the distribution of the clusters areas, as well as the corresponding summary statistics (Fig. 2(b-e)). And this could still happen despite RNN-DBSCAN is a noise-resistant algorithm. One way to avoid such a situation would be to prevent it with an adequate documentation about the minor events. In that case, the documents would help to identify the minor events, which will be then classified, together with the Summer Party, as "multiple events", and therefore not treated with our proposed methodologies (that are designed for "distinct events"). However, future efforts may develop further our methodologies for "multiple events", to better understand the physics of anomalies during both natural and human-made events, and strengthen our empirical relationships. For example, the physics of anomalies could be further investigated for "multiple" social events, being scheduled and frequent, and then applied to the assessment of rare flood events. In conclusion, our work represents a bridge between the studies on human mobility during natural disasters (28, 93) and human mobility during social events (94, 95), and can therefore have an impact on both (flood) risk management and city and event management, supporting urban planners, social scientists, city authorities, and traffic engineers. We hope that our findings could also inspire the development of more sophisticated data mining and machine learning algorithms for time series anomaly detection and clustering, as well as stimulate the advancement of the physics of anomalies - related to human mobility -, likely in the context of network science.

Methods

Nomenclature

Please refer to Supplementary Table 4 for a complete list of symbols used in this work, and their description.

Datasets

To perform our analysis at the national scale, we employ both the estimated (by Swisscom) number of users who, coming from the entire Switzerland were detected in the local area of study during the period of 44 days, and the postal code of their municipality of residence. In particular, for each of the 44 days, our dataset is composed of (i) the postal code of the municipality of residence, (ii) the name of the municipality of residence, (iii) the hour of the day during which travellers were detected, and (iv) the estimated count of detected people. We therefore derive the frequency of visits by using the estimated number of travellers, and, both time and day in which they were detected in the local area of study. About the way to derive the frequency of distances travelled by the users, we just employ the postal code of their municipality of residence. We then calculate those distances by selecting the shortest path, on the Swiss road network, between the centroid of every municipality of residence of the users, and the centroid of the local area of study. At the local scale, we use both the estimated number of people who travelled each road or railway of the transport networks (which includes ~ 5000 edges, and ~ 9000 nodes), and the features of the network's edges. In particular, our dataset is composed of (i) the edge's start node, (ii) the edge's end node, (iii) the edge path, i.e. the nodes constituting the polyline edge, (iv) the edge type, i.e. if urban road, motorway or railway track, (v) the hour of the day during which travellers were detected, and (vi) the estimated count of detected, and (vi) the estimated count of detected people.

Measures of dissimilarity among distributions

Given a measurable space $(\mathcal{X}, \mathcal{A})$, i.e. a pair consisting of a set \mathcal{X} (called sample space) and a σ -algebra \mathcal{A} , we consider two probability measures, P_1 and P_2 , on $(\mathcal{X}, \mathcal{A})$, and their respective densities, ρ_1 and ρ_2 .

Jensen–Shannon divergence. Lin (96) defines the Jensen–Shannon divergence through the Shannon entropy function, $f_{SE}(P_i) = -\sum_{x \in X} P_i(x) \log P_i(x)$, and the weights of P_1 and P_2 , i.e. ω_1 and ω_2 (96, 97):

$$f_{JS}(P_1, P_2) \stackrel{\text{def}}{=} f_{SE}(\omega_1 P_1 + \omega_2 P_2) - \omega_1 f_{SE}(P_1) - \omega_2 f_{SE}(P_2)$$

We then rewrite the Jensen–Shannon divergence in terms of the Kullback–Leibler divergence, $f_{KL}(P_1 \parallel P_2)$, using uniform weights, $\omega_1 = \omega_2 = \frac{1}{2}$:

$$\begin{split} \stackrel{\text{def}}{=} & -\sum_{x \in X} \left(\frac{P_1(x)}{2} + \frac{P_2(x)}{2} \right) \log \left(\frac{P_1(x)}{2} + \frac{P_2(x)}{2} \right) \\ & + \frac{1}{2} \sum_{x \in X} P_1(x) \log P_1(x) + \frac{1}{2} \sum_{x \in X} P_2(x) \log P_2(x) \\ & = & -\sum_{x \in X} \left(\frac{P_1(x)}{2} \right) \log \left(\frac{P_1(x)}{2} + \frac{P_2(x)}{2} \right) - \sum_{x \in X} \left(\frac{P_2(x)}{2} \right) \log \left(\frac{P_1(x)}{2} + \frac{P_2(x)}{2} \right) \\ & + \frac{1}{2} \sum_{x \in X} P_1(x) \log P_1(x) + \frac{1}{2} \sum_{x \in X} P_2(x) \log P_2(x) \end{split}$$

We use $H(x) = \frac{P_1(x) + P_2(x)}{2}$,

$$= -\frac{1}{2} \sum_{x \in X} P_1(x) \log H(x) - \frac{1}{2} \sum_{x \in X} P_2(x) \log H(x) + \frac{1}{2} \sum_{x \in X} P_1(x) \log P_1(x) + \frac{1}{2} \sum_{x \in X} P_2(x) \log P_2(x)$$

$$= \frac{1}{2} \left(-\sum_{x \in X} P_1(x) \log H(x) - \sum_{x \in X} P_2(x) \log H(x) + \sum_{x \in X} P_1(x) \log P_1(x) + \sum_{x \in X} P_2(x) \log P_2(x) \right)$$

$$= \frac{1}{2} \left(\sum_{x \in X} P_1(x) \log \frac{P_1(x)}{H(x)} + \sum_{x \in X} P_2(x) \log \frac{P_2(x)}{H(x)} \right)$$

$$= \frac{1}{2} \left(\sum_{x \in X} P_1(x) \log \frac{P_1(x)}{H(x)} + \sum_{x \in X} P_2(x) \log \frac{P_2(x)}{H(x)} \right)$$

$$(4)$$

In this study, we use Equation (4) to calculate the Jensen–Shannon divergence.

Wasserstein distance. We consider the Monge–Kantorovich minimization problem (98), also called the Kantorovich's optimal transportation problem (99):

minimize
$$\int_{\mathcal{X}\times\mathcal{Y}} \cos(x,y) d\pi(x,y)$$
 for $\pi \in \Pi(\mu,\nu)$

That would correspond to find the optimal transport cost between two probability measures μ and ν (98, 99) which is defined as:

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{Y}}\operatorname{cost}(x,y)d\pi(x,y)$$

If both our two probability measures, μ and ν , and the cost function are on \mathbb{R} , and the cost function is a convex non-negative symmetric function that takes the form $\cot(x, y) = \cot(x - y)$, then, the optimal transportation cost can be written as follows (details in (99), from both Theorem 2.18 and Remarks 2.19):

$$\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{Y}}\operatorname{cost}(x,y)d\pi(x,y) = \int_0^1 c(F^{-1}(t) - G^{-1}(t))\,dt$$

where F^{-1} and G^{-1} are the quantile functions of the probability measures μ and ν , respectively. F and G are instead the two cumulative distribution functions (CDFs), associated with μ and ν , respectively. If the cost function takes the form $\cos(x, y) = |x - y|$, by applying the Fubini's theorem, we would then get the following optimal transportation cost:

$$\int_0^1 \left| F^{-1}(t) - G^{-1}(t) \right| dt = \int_{\mathbb{R}} \left| F(x) - G(x) \right| dx$$

We now consider a Polish metric space, \mathcal{X} , equipped with a metric d(x, y), and we define the cost function as the p-th power of d(x, y), i.e. $cost(x, y) = d(x, y)^p$, where $p \in [1, \infty)$. Then, we can establish a distance between μ and ν , that we call the Wasserstein distance of order p between μ and ν (98, 99):

$$W_p(\mu,\nu) = \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathcal{X}} d(x,y)^p d\pi(x,y)\right)^{\frac{1}{p}}$$

The Wasserstein distance represents the metric side of optimal transportation. Therefore, when two probability measures, μ and ν , are on the real line, i.e. \mathbb{R} , then the Wasserstein distance has a closed form, and reads as $W_p(\mu,\nu) = \left(\int_0^1 |F^{-1}(t) - G^{-1}(t)|^p dt\right)^{\frac{1}{p}}$. When both μ and ν are on \mathbb{R} , and the exponent is p = 1, the Wasserstein distance takes the W_1 form, also called the Kantorovich–Rubinstein distance:

$$W_1(\mu,\nu) = \int_{\mathbb{R}} |F(x) - G(x)| \, dx \tag{5}$$

In this study, we use Equation (5) to calculate the Wasserstein distance W_1 .

Total Variation distance and Hellinger distance. The total variation distance is defined as (100):

$$f_{TV}(P_1, P_2) \stackrel{\text{def}}{=} \sup_{A \in \mathcal{A}} |P_1(A) - P_2(A)|$$

If \mathcal{X} is finite, then, any σ -algebra \mathcal{A} is finite (since the power set is finite). When \mathcal{A} is finite, the $\max_{A \in \mathcal{A}}$ exists, and it is equivalent to $\sup_{A \in \mathcal{A}}$. Then, if \mathcal{A} is the power set, $\max_{A \in \mathcal{A}}$ is equivalent to $\max_{A \subseteq \mathcal{X}}$ leading to $f_{TV}(P_1, P_2) = \max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)|$. As an alternative characterization, the total variation distance can be expressed as half of the L^1 -norm (101). This can be shown by using subsets of \mathcal{X} , i.e. $B = \{x : P_1(x) > P_2(x)\}$, and $B^c = \{x : P_1(x) < P_2(x)\}$ (102):

$$f_{TV}(P_1, P_2) = \max_{A \subseteq \mathcal{X}} |P_1(A) - P_2(A)|$$

= $P_1(B) - P_2(B)$
= $\frac{1}{2} (P_1(B) - P_2(B) + P_1(B^c) - P_2(B^c))$
= $\frac{1}{2} \left(\sum_{x \in B} \rho_1(x) - \rho_2(x) + \sum_{x \in B^c} \rho_1(x) - \rho_2(x) \right)$
= $\frac{1}{2} \sum_{x \in \mathcal{X}} |\rho_1(x) - \rho_2(x)|$
 $\stackrel{\text{def}}{=} \frac{1}{2} ||P_1 - P_2||_1$ (6)

The Hellinger distance between two probability measures is defined as the L^2 -norm of the difference between the square roots of the densities, $\sqrt{\rho_1}$ and $\sqrt{\rho_2}$ (100, 103):

$$f_{H}(P_{1}, P_{2}) \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}} \left(\sum_{i=1}^{\infty} \left(\sqrt{\rho_{1}(i)} - \sqrt{\rho_{2}(i)} \right)^{2} \right)^{\frac{1}{2}}$$
$$\stackrel{\text{def}}{=} \frac{1}{\sqrt{2}} \|P_{1} - P_{2}\|_{2} \tag{7}$$

In this study, we use Equation (6) and Equation (7) to calculate, respectively, the total variation distance and the Hellinger distance.

Complementary empirical distance correlation. Proposed by Székely et al. in 2007 (104), the distance correlation f_{DC} measures the degree of dependence between two multivariate random variables, or random vectors, i.e. X in \mathbb{R}^u and Y in \mathbb{R}^v , with u and v positive integers. If we consider $X \sim P_1$ and $Y \sim P_2$, we can write the empirical distance correlation among two probability measures as:

$$f_{DC}(X,Y) = \begin{cases} \sqrt{\frac{\mathcal{V}^2(X,Y)}{\sqrt{\mathcal{V}^2(X)\mathcal{V}^2(Y)}}}, & \mathcal{V}^2(X)\mathcal{V}^2(Y) > 0\\ 0, & \mathcal{V}^2(X)\mathcal{V}^2(Y) = 0 \end{cases}$$
(8)

where

$$\mathcal{V}^{2}(X,Y) = \frac{1}{v^{2}} \sum_{i,j=1}^{v} A_{ij}B_{ij} \qquad A_{ij} = a_{ij} + \frac{1}{v} \sum_{i=1}^{v} a_{ij} + \frac{1}{v} \sum_{j=1}^{v} a_{ij} + \frac{1}{v^{2}} \sum_{i,j=1}^{v} a_{ij}$$

$$\mathcal{V}^{2}(X) = \mathcal{V}^{2}(X,X) = \frac{1}{v^{2}} \sum_{i,j=1}^{v} A_{ij} \qquad B_{ij} = b_{ij} + \frac{1}{v} \sum_{i=1}^{v} b_{ij} + \frac{1}{v} \sum_{j=1}^{v} b_{ij} + \frac{1}{v^{2}} \sum_{i,j=1}^{v} b_{ij}$$

$$\mathcal{V}^{2}(Y) = \mathcal{V}^{2}(Y,Y) = \frac{1}{v^{2}} \sum_{i,j=1}^{v} B_{ij} \qquad a_{ij} = |X_{i} - X_{j}|_{u} \qquad b_{ij} = |Y_{i} - Y_{j}|_{v}$$

and n is the sample size. In Equation (8), $f_{DC}(X,Y) = 0$, when $\mathcal{V}^2(X)\mathcal{V}^2(Y) = 0$, means that the distance correlation is zero if X and Y are independent. In this work, we use Equation (8) to calculate the complementary empirical distance correlation, defined as $f_{DC}^{\complement}(X,Y) \stackrel{\text{def}}{=} 1 - f_{DC}(X,Y)$.

Cosine distance Among the measures of distributional (dis)similarity (105), we also consider the cosine similarity, with $X \sim P_1$ and $Y \sim P_2$:

$$f_{CS}(X,Y) = \frac{\sum_{i=1}^{v} (X_i, Y_i)}{\sqrt{\sum_{i=1}^{v} X_i^2 \sum_{i=1}^{v} Y_i^2}} \\ \stackrel{\text{def}}{=} \frac{X \cdot Y}{\|X\|_2 \|Y\|_2}$$
(9)

In this study, we use Equation (9) to compute the cosine distance $f_{CD}(X,Y) = 1 - f_{CS}(X,Y)$.

Discord Aware Matrix Profile (DAMP) algorithm

The DAMP algorithm (106, 107) is based on two state-of-the-art techniques for time series anomaly detection (TSAD), i.e. the "time series discord" (108–111), proposed by Keogh et al. (112) in 2005, and the "matrix profile" (113–116), introduced by Yeh et al. (117) in 2016. Intuitively, a time series discord defines the anomaly in a time series, while the matrix profile is a data structure used for time series analysis. Despite there being no single method that is superior in every case of anomaly detection (118–120), the matrix profile-based approaches – of which DAMP is one –, performed well in different benchmarks (118, 121). On a large benchamark containing 13766 univariate time series (i.e. TSB-UAD), Paparrizos et al. (118) compared 12 methods for anomaly detection, which belonged to different method families, i.e. deep learning (CNN, AutoEncoder, LSTM), outlier detection (Isolation Forest, LOF), classic machine learning (OCSVM, PCA, HBOS) and data mining (matrix profile, NormA). They found that, on average, the best methods for anomaly detection were the matrix profile-based algorithms and NormA (122). Another comparison between anomaly detection algorithms was made by He et al. (121), on a collection of 250 univariate time series, called the KDD CUP 2021 (KDD21) dataset (123). They used algorithms based on deep learning (LSTM, USAD, TranAD), seasonal-trend decomposition (NSigma, OnlineSTL, OneShotSTL) and matrix profile (NormA, STOMPI, SAND, DAMP), and they showed the superiority of DAMP, in terms of accuracy.

Basic concepts and definitions in DAMP. We consider a time series T, defined as a sequence of temporal data, i.e. $T = (t_k)_{k=1}^{n}$, and a time subsequence $T_{i,m} = (t_k)_{k=i}^{i+m-1}$, where $k, n, i, m \in \mathbb{N}$. In particular, n and m are the length of the time subsequence, respectively. A time subsequence $T_{i,m}$ is therefore characterised by two values: one is the starting position i on the temporal line, while the second is its length m. We then consider the Euclidean distance among two subsequences, $T_{i,m} = (t_k)_{k=i}^{i+m-1}$ and $T_{j,m} = (t_k)_{k=j}^{j+m-1}$, that is defined through the L^2 -norm, i.e. $||T_{i,m} - T_{j,m}||_2 \stackrel{\text{def}}{=} \sqrt{\sum_{l=0}^{m-1} (t_{i+l} - t_{j+l})^2}$. Here, we used $t_i = (t_k)_{k=i}$ and $t_j = (t_k)_{k=j}$ to indicate the first elements of $T_{i,m}$ and $T_{j,m}$, respectively. More precisely, for this algorithm, we will use the z-normalised Euclidean distance, $\sqrt{\sum_{l=0}^{m-1} (\tilde{t}_{i+l} - \tilde{t}_{j+l})^2}$, where $t_{i+l} \to \tilde{t}_{i+l} = \left(\frac{t_{i+l} - \mu_i}{\sigma_i}\right)$ and $t_{j+l} \to \tilde{t}_{j+l} = \left(\frac{t_{j+l} - \mu_j}{\sigma_j}\right)$, and, for ease of notation, we will denote it with the same L^2 -norm notation of the non-normalized Euclidean distance, i.e. $||T_{i,m} - T_{j,m}||_2$. Distances are calculated only for non-overlapping subsequences $T_{i,m}$ and $T_{j,m}$, also called as "non-self match" subsequences, i.e. when $|i-j| \ge m$. We can name $|i-j| \ge m$ as the "non-overlapping condition" as well. If we focus on one specific $T_{i,m}$, i.e. we fix the index i to a certain number, and we only vary the index j such that $|i-j| \ge m$, we can calculate all the distances among our chosen subsequence $T_{i,m}$ and all the other subsequences $T_{j,m}$ forming the same time series. We define the set of distances among our chosen subsequence $T_{i,m}$ and all the other subsequences $T_{j,m}$, of the same time series, as "distance profile" δ_i :

$$\delta_i = \left\{ \|T_{i,m} - T_{j,m}\|_2 : 1 \le j \le (n - m + 1) \land |i - j| \ge m \right\}$$
(10)

Given the "distance profile" δ_i of a reference sebsequence $T_{i,m}$, we can then find the "nearest neighbour" of $T_{i,m}$, i.e. that $T_{j,m}$ subsequence that has the minimum Euclidean distance from $T_{i,m}$. In other words, the "nearest neighbour" is defined by $\min(\delta_i)$. Obviously, we can find the nearest neighbour for each reference subsequence $T_{i,m}$, and the collection of all the minimum distances, $\min(\delta_i)$, is called as the "matrix profile", M (117):

$$M = \{\min(\delta_1), \min(\delta_2), \dots, \min(\delta_{n-m+1})\}$$
(11)

The subsequence corresponding to the maximum value of the matrix profile, i.e. the subsequence with the largest Euclidean distance to its non-overlapping nearest neighbour, is called as "time series discord". Indeed, a discord is a subsequence which is a maximally different from all the other subsequences, and can therefore be thought as an anomalous subsequence. Despite the discovery of time series discords is a well recognised and popular anomaly detection approach (108–111), it fails when unusual subsequences (discords) occur more than once in the same time series. This is a widely known issue, called as the "twin freak problem" (124, 125). However, the current algorithm is able to mitigate this problem, by computing the "left-discords" (106, 107). Essentially, those discords are calculated as previously described, but by adding the $1 \le j \le (i - m)$ condition, i.e. by considering only subsequences $T_{j,m}$ on the left of the reference subsequence $T_{i,m}$. With this additional condition, the profile distance would become the "left distance profile":

$$\delta_i^* = \left\{ \|T_{i,m} - T_{j,m}\|_2 : 1 \le j \le (i-m) \le (n-2m+1) \land |i-j| \ge m \right\}$$
(12)

and, consequently, the matrix profile would become the "left matrix profile":

$$M^* = \{\min(\delta_1^*), \min(\delta_2^*), \dots, \min(\delta_{n-m+1}^*)\}$$
(13)

Therefore, the DAMP algorithm mainly computes Equation (12) and Equation (13) to find the time series left-discords. Then, those discords can be sorted to identify the "top left-discords", i.e. the largest anomalies.

Parameters in DAMP. We use DAMP_topK (126), a variant of DAMP that computes the top left-discords, and that splits the data into two intervals, i.e. the "training data" and the "test data" (106, 107). With such a division of data, we compare the current subsequence $T_{i,m}$ with those subsequences $T_{j,m}$ that are located within the training interval. We set the split point I between the training data and the test data – called "CurrentIndex" –, equal to $I = 6 \times 24$, i.e. 6 days. During those 6 days, ranging from Saturday 17^{th} of June to Thursday 22^{nd} of June, we did not find significant events that would affect the anomaly detection analysis. For details on the events that occurred during the entire examined temporal window of 44 days, within the local area of study, please see Supplementary Table 1. Once the CurrentIndex is given by the user, DAMP_topK calculates automatically the subsequence length. In our case, it would result in m = 24, corresponding to one period, i.e. 1 day, of our time series. With these numbers, we would get a ratio between the CurrentIndex and the subsequence length equal to 6, which is larger than the recommended ratio, $\frac{I}{m} > 4$ (127). We then set the number of top left-discords we want to extract with DAMP_topK equal to 10. This means that for every road and railway track, DAMP would identify up to the 10 largest left-discords (anomalies), if they exist.

Reverse Nearest Neighbor Density-Based Spatial Clustering of Applications with Noise (RNN-DBSCAN) algorithm

DBSCAN is one of the most popular algorithm for clustering detection, it is resistant to noise and can handle clusters of different shapes and sizes. However, it does not work well with varying densities and it is sensitive to its two parameters. RNN-DBSCAN (47) is an extension of DBSCAN, based on the construction of a directed network over the dataset of points. RNN-DBSCAN reduces the amount of input parameters from two to one, and has the ability to handle regions that can vary greatly in density, showing a great accuracy in cluster detection (128).

Basic concepts and definitions in DBSCAN. Let Υ be a set of points in a 2-dimensional space \mathbb{R}^2 , where the generic point *i* and the generic point *j* have, respectively, coordinates (x_i, y_i) and (x_j, y_j) . Please note that, here, we use *i*, $j \in \mathbb{N}$ to label our points, while in DAMP, we use *i* and *j* as 1-dimensional coordinates over the time line, i.e. to indicate the position of subsequences $T_{i,m}$ and $T_{j,m}$. Also, we use the same indices, *i* and *j*, for two different contexts, to facilitate those readers who have a background in the respective scientific areas, and are therefore used to see *i* and *j* as common notations. After this brief digression, we continue with DBSCAN and we define the ϵ -neighborhood of point *i* as a subset of points, whose Euclidean distance from point *i* is less than or equal to ϵ , i.e., $\epsilon N_i = \{j : d_{ij} \leq \epsilon\}$. Intuitively, the ϵ -neighborhood would allow us to distinguish points inside a cluster, that we call "core points", from points on the border of a cluster, that we name "border points", and isolated points, simply identified as "noise". Indeed, the ϵ -neighborhood of a border point would contain less points than the ϵ -neighborhood of a core point. But what is the exact number of points, within a ϵ -neighborhood, to

distinguish core points from border points (or noise points)? We can unambiguously identify core points from border points by introducing a minimum number of points, γ , that lie within a ϵ -neighborhood, i.e. by assigning a minimum local density. Therefore, a core point needs to contain at least γ points within a ϵ -neighborhood, i.e. $|\epsilon N_i| \geq \gamma$, while a border point contains less than γ points within a ϵ -neighborhood, i.e. $|\epsilon N_i| < \gamma$. Any point *j* that belongs to the neighborhood of a core point *i*, i.e. $\{j : j \in \epsilon N_i \land |\epsilon N_i| \geq \gamma\}$, is called "directly density-reachable" from *i*. This definition can be then extended to pairs of points *i* and *j* that are not directly reachable. Indeed a point *j* is "density-reachable" from a point *i*, if we can form a chain of "directly density-reachable" points, starting at point *i* and ending at point *j*. If both points *i* and *j* are core points, the property of density-reachability is symmetric, i.e. holds for both directions, either starting at point *i* and moving towards point *j*, or starting at point *j* and moving towards point *i*. However, if *i* and *j* are both border points. Nevertheless, two border points can be "density-connected" if they are both density-reachable from a common point between them. The notion of density-reachability allows us to define a cluster as a set of points that are density-reachable from a core point (and, consequently, density-connected among each other). Indeed, when DBSCAN starts to retrieve all points that are density-reachable from an arbitrary point, if it is a orce point, DBSCAN detects the entire cluster, to which that core point from it, and goes to visit the next point in the dataset.

Basic concepts and definitions in RNN-DBSCAN. RNN-DBSCAN is a graph-based interpretation of DBSCAN, where the concept of ϵ -neighborhood (with a fixed radius) of a point *i*, ϵN_i , is replaced by the graph's notion of reverse k-nearest neighborhood of node i, $RkNN_i$. The concept of reverse k-nearest neighborhood, $RkNN_i$, is related to the construction of a k-nearest neighbor graph (k-NNG) over our set of points Υ . Indeed, for any set of points Υ , we can build up a directed graph k-NNG with nodes equal to Υ , and directed edges that connect every node to each of its k-nearest neighbors. Once the k-NNG graph is built, we can introduce the k-nearest neighborhood of an arbitrary node i, i.e. $kNN_i = \left\{j: d_{ij'}, \forall j \in \Upsilon \setminus (\{i\} \cup \{j'\}), \forall j' \in \Upsilon \setminus (\{i\} \cup \{j\}) \land |\{j\}| = k\right\} \text{ and, likewise, } kNN_j, \text{ i.e. the } k-\text{nearest} \in \mathbb{C}$ neighborhood of node j, that we can use to define the "outgoing degree" of node i, $k_i^{\text{out}} = |\{(i,j): j \in kNN_i\}|$ and the "ingoing degree" of the same node $i, k_i^{\text{in}} = |\{(j,i): i \in kNN_j\}|$. Here, (i,j) denotes the edge from node i to node j, while (j,i) symbolizes the edge from node j to node i. The outgoing degree k_i^{out} , abbreviated as "out-degree", represents the number of edges that point from node i to other nodes, while the ingoing degree k_i^{in} , abbreviated as "in-degree", represents the number of edges that point to node i. Therefore, the out-degree k_i^{out} corresponds to the size of the k-nearest neighborhood of an arbitrary node i, i.e. $k_i^{\text{out}} = |kNN_i|$, while the in-degree k_i^{in} corresponds to the size of the reverse k-nearest neighborhood of node *i*, i.e. $k_i^{\text{in}} = |RkNN_i|$. A meticulous reader would confute both $k_i^{\text{out}} = |kNN_i|$ and $k_i^{\text{in}} = |RkNN_i|$, since kNN_i and $RkNN_i$ would be sets of nodes and not a sets of edges, while both k_i^{out} and k_i^{in} would refer to edges. However, every node j would be part of an edge (i, j) or (j, i), by construction (i.e. definition) of the k-NNG graph, and we could then use a set of edges (i, j) - having the same starting node i, and $j \in kNN_i$ -, to indicate kNN_i , and a set of edges (j, i) - having the same ending node i, and $i \in kNN_j$ –, to indicate $RkNN_i$. For example, in this perspective, $j \in RkNN_i$ means that node j is the starting node of an edge (j,i) that points towards node i, such that $i \in kNN_i$. In other words, $j \in RkNN_i$ means $j \in \{(j,i) : i \in kNN_j\}$, or, more precisely, $j \in \{j : j \in (j,i) \land i \in kNN_j\}$. Due to the definition of k-NNG graph, all nodes in k-NNG have the same out-degrees, $k_i^{\text{out}} = |kNN_i| = k$, but they have different in-degrees, k_i^{in} . We can now replace the DBSCAN core point condition, $|\epsilon N_i| \ge \gamma$, with the RNN-DBSCAN core point condition, i.e. $k_i^{\text{in}} = |RkNN_i| \ge k$. Similarly to what we did in DBSCAN with its core point condition, here, we can use the RNN-DBSCAN core point condition to re-define the core and border points (and noise points), as well as the notions of "directly density-reachable", "density-reachable", "density-connected" and "cluster". For example, the set of points j that are "directly density-reachable" from i would read as $\{j: j \in RkNN_i \land k_i^{\text{in}} = |RkNN_i| \ge k\}.$

Parameters in RNN-DBSCAN. We use a Matlab implementation of RNN-DBSCAN (129), which requires the sole input parameter k, i.e. the number of nearest neighbours in the k-NNG graph, and we use a heuristics approach proposed by Bryant et al. (47), to automatically select an appropriate value for k. For each $k = 1, 2, \dots, 100$, we run RNN-DBSCAN and detect the number of resulting clusters, c_k . We therefore obtain a sequence of 100 elements, that we denote with $C = (c_k)_{k=1}^{100}$. Since we commonly get the same number of clusters for different values of k, i.e. $c_k = c_{k'}$ for $k \neq k'$, the sequence C will contain duplicate elements. The duplicate elements can be then gathered to form subsequences of C, where each subsequence contains the same element, repeated a number of times. We then use the image of C and the preimage of C to build up those subsequences. The image of C is a set, i.e. an unordered collection of distinct objects, $img(C) = \{c_k\}_{k=1}^{100}$, and it is usually denoted using curly braces. Through the image of C we therefore get the unique elements of C, i.e. unique values of the number of clusters. Indeed, for example, if we had $img(C) = \{3, 1, 3, 7, 7, 3, 1, 1, 3, 2\}$, this would be equivalent to $\{1, 2, 3, 7\}$. Then, for any element of the image we can introduce the preimage of that element, $C^{-1}(\{c_k\})$, i.e. the set of k indices that indicate the position of that element's duplicates within C. Hence, the preimage of C allows us to see if those unique elements

are repeated, while the cardinality of the preimage of C, denoted by $|C^{-1}(\{c_k\})|$, tells us about the number of times those unique elements are repeated. Basically, the cardinality of the preimage of C is the frequency of occurrence of the number of clusters. Obviously, the union of preimages of singletons $\{c_k\}$, with respect to C, will give us the original sequence, i.e. $C = \bigcup_{c_k \in img(C)} C|_{C^{-1}(\{c_k\})}$, where $C|_{C^{-1}(\{c_k\})}$ is the restriction of the function C (a sequence is a function) to the preimage of $\{c_k\}$, with respect to C. Once we know the frequency of occurrence for each number of clusters, we essentially have a distribution of the number of clusters calculated over the range $1 \le k \le 100$. The heuristics approach proposed by Bryant et al. (47) consists of considering the leftmost local maxima in the distribution, to find first the best value of k, and then get the best clustering in a given dataset. Indeed, when they tested RNN-DBSCAN over a range $1 \le k \le 100$, on a set of artificial datasets with given ground truth, they noticed that the Adjusted Rand Index (ARI) performance was maximum at the leftmost local maxima of the distribution. Also, they observed a positive correlation between the maximum ARI and the ARI performance at minimum k. This means that, among all the k input parameters which contributed to the construction of the leftmost local maxima of the distribution, the minimum k could be safely used as input parameter for generating the correct number of clusters in a dataset of points. Please note that here, and only for this distribution of the number of clusters, we use the same convention employed by Bryant et al. (47) about the direction of the x-axis, where positive numbers increase towards left. Obviously, if we had used the classical convention for the x-axis, where positive numbers increase towards right, we would have focused on the rightmost local maxima of the distribution. Our tests on the proposed heuristic approach, and on the same artificial datasets used by Bryant et al. (47), are illustrated in Supplementary Figure 6. There, we show pretty much the same results to those ones of Bryant et al. (47).

Acknowledgements

We thank Eamonn Keogh, Yue Lu, and Trevor Vannoy for useful discussions, and Swisscom for providing the data.

Author contributions statement

S.L. designed research, analyzed data, performed research, created the collaboration with Swisscom, and wrote the paper. M.K. and A.Z. reviewed the paper, M.K. initiated the project, A.Z. performed geodata analysis and flood exposure analysis.

Additional information

Competing interests: The authors declare no competing interest.

Costs for data: Mobile phone data were purchased by Swisscom in November 2019, for a total cost of 32400 CHF (i.e. \sim 36000 USD, in October 2023). The costs included 1800 CHF for "Scope & Alignment" (1-man day of work), 9000 CHF for "Data mining & preparation" (5-man days of work), and 21600 CHF for "Data processing & analysis" (12-man days of work).

Data availability: The Swisscom mobile phone data will be available on Zenodo.



Figure 1. (a) Graphical representation of the impacts of social and flood events at a national scale. The paths show the distances travelled by individuals who visited the local area of study, during 6 weeks. (b) Relative frequency histograms representing the empirical distributions of visits to the local area of study, during 6 weeks. (c) Relative frequency histograms representing the empirical distributions of distances travelled by people who visited the local area of study, during 6 weeks. (d) Comparison of the 6 histograms that represent the number of visits in each week. (e) Comparison of the 6 histograms that represent the distances covered by visitors, in each week.



Figure 2. (a) Temporal evolution of anomalies over 44 days. (b) Distribution of distances among clusters of anomalies and an event center, per each type of event. (c) Distribution of areas of clusters of anomalies, per each type of event. (d) Statistical dispersion of distances between the clusters of anomalies and an event center, per event type. (e) Statistical dispersion of cluster's areas, per event type. 18/27



Figure 3. (a) Graphical explanation of the relationship between the largest and the smallest distances among clusters of anomalies and an event center. (b) Largest distance between a cluster of anomalies and an event center, against the difference between the largest and the smallest distances among clusters of anomalies and an event center. (c) Number of clusters of anomalies versus the mean value of the areas of the clusters of anomalies. (d) Distance of the closest cluster of anomalies to an event center against the area of the same cluster of anomalies.



Figure 4. (a) Multi-layer network representing the local area of study. The network includes three layers: the urban roads layer, the motorways layer and the rails layer. The coloured edges represent the unique cycles of the multi-layer network. Please notice that a number of edges are multicoloured due to the overlap of some cycles. (b) Min, mean and max number of anomalies detected within the cycles of the multi-layer network, for each social and flood event (the colors in (b) refer to Fig. 2(a)).

References

- Colonna, P., Berloco, N. & Circella, G. The interaction between land use and transport planning: A methodological issue. *Procedia - Soc. Behav. Sci.* 53, 84–95, DOI: https://doi.org/10.1016/j.sbspro.2012.09.862 (2012).
- 2. Rodrigue, J.-P., Comtois, C. & Slack, B. The geography of transport systems (Routledge, 2016).
- González, M. C., Hidalgo, C. A. & Barabási, A.-L. Understanding individual human mobility patterns. *Nature* 453, 779, DOI: 10.1038/nature06958 (2008).
- **4.** Wang, Q. & Taylor, J. E. Quantifying human mobility perturbation and resilience in hurricane sandy. *PLOS ONE* **9**, e112608, DOI: 10.1371/journal.pone.0112608 (2014).
- 5. Barbosa, H. *et al.* Human mobility: Models and applications. *Phys. Reports* 734, 1–74, DOI: 10.1016/j.physrep.2018.01. 001 (2018).
- 6. Wang, S. *et al.* A bibliometric analysis and network visualisation of human mobility studies from 1990 to 2020: Emerging trends and future research directions. *Sustainability* **13**, DOI: 10.3390/su13105372 (2021).
- Wang, R., Zhang, X. & Li, N. Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities* 130, 103939, DOI: https://doi.org/10.1016/j.cities.2022.103939 (2022).
- 8. Barth, A.-K. & Heimeshoff, U. Does the growth of mobile markets cause the demise of fixed networks? Evidence from the European Union. *Telecommun. Policy* 38, 945–960, DOI: https://doi.org/10.1016/j.telpol.2014.03.003 (2014).
- 9. Gupta, G. K. Ubiquitous mobile phones are becoming indispensable. *ACM Inroads* 2, 32–33, DOI: 10.1145/1963533.19 63545 (2011).
- **10.** Smith, C. & Collins, D. *Wireless Networks: Design and Integration for LTE, EVDO, HSPA, and WiMAX* (McGraw-Hill International Enterprises, LLC., 2014).
- Lepri, B., Staiano, J., Sangokoya, D., Letouzé, E. & Oliver, N. The Tyranny of Data? The Bright and Dark Sides of Data-Driven Decision-Making for Social Good, 3–24 (Springer International Publishing, Cham, 2017).
- 12. Meekan, M. G. *et al.* The ecology of human mobility. *Trends Ecol. & Evol.* 32, 198–210, DOI: https://doi.org/10.1016/j. tree.2016.12.006 (2017).
- Yabe, T., Jones, N. K. W., Rao, P. S. C., Gonzalez, M. C. & Ukkusuri, S. V. Mobile phone location data for disasters: A review from natural hazards and epidemics. *Comput. Environ. Urban Syst.* 94, 101777, DOI: https://doi.org/10.1016/j.compenvurbsys.2022.101777 (2022).
- Wang, Z., He, S. Y. & Leung, Y. Applying mobile phone data to travel behaviour research: A literature review. *Travel. Behav. Soc.* 11, 141–155, DOI: https://doi.org/10.1016/j.tbs.2017.02.005 (2018).
- **15.** Muniz-Rodriguez, K. *et al.* Social media use in emergency response to natural disasters: A systematic review with a public health perspective. *Disaster Medicine Public Heal. Prep.* **14**, 139–149, DOI: 10.1017/dmp.2020.3 (2020).
- 16. Haraguchi, M. *et al.* Human mobility data and analysis for urban resilience: A systematic review. *Environ. Plan. B:* Urban Anal. City Sci. 49, 1507–1535, DOI: 10.1177/23998083221075634 (2022).
- Pappalardo, L., Manley, E., Sekara, V. & Alessandretti, L. Future directions in human mobility science. *Nat. Comput. Sci.* 3, 588–600, DOI: 10.1038/s43588-023-00469-4 (2023).
- Oliver, N. *et al.* Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Sci. Adv.* 6, eabc0764, DOI: 10.1126/sciadv.abc0764 (2020). https://www.science.org/doi/pdf/10.1126/sciadv.abc0764.
- Diffenbaugh, N. S. *et al.* Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci.* 114, 4881–4886, DOI: 10.1073/pnas.1618082114 (2017). https://www.pnas.org/doi/pdf/10.1073/pnas.16 18082114.
- **20.** on Climate Change (IPCC), I. P. *Weather and Climate Extreme Events in a Changing Climate*, 1513–1766 (Cambridge University Press, 2023).

- Lange, S. *et al.* Projecting exposure to extreme climate impact events across six event categories and three spatial scales. *Earth's Futur.* 8, e2020EF001616, DOI: https://doi.org/10.1029/2020EF001616 (2020). E2020EF001616 10.1029/2020EF001616.
- 22. Kreibich, H. *et al.* The challenge of unprecedented floods and droughts in risk management. *Nature* **608**, 80–86, DOI: 10.1038/s41586-022-04917-5 (2022).
- 23. CRED & UNDRR. The human cost of disasters: an overview of the last 20 years (2000-2019). Report (2020).
- 24. (WMO), W. M. O. Wmo atlas of mortality and economic losses from weather, climate and water extremes (1970–2019). Report (2021).
- 25. CRED & UNDRR. 2020: The non-covid year in disasters. Report (2021).
- 26. IFRC. World disasters report 2022. Report (2023).
- 27. CRED. 2022 disasters in numbers. Report (2023).
- Li, W., Wang, Q., Liu, Y., Small, M. L. & Gao, J. A spatiotemporal decay model of human mobility when facing large-scale crises. *Proc. Natl. Acad. Sci.* 119, e2203042119, DOI: doi:10.1073/pnas.2203042119 (2022).
- Yabe, T., Rao, P. S. C., Ukkusuri, S. V. & Cutter, S. L. Toward data-driven, dynamical complex systems approaches to disaster resilience. *Proc. Natl. Acad. Sci.* 119, e2111997119, DOI: doi:10.1073/pnas.2111997119 (2022).
- Farahmand, H., Wang, W., Mostafavi, A. & Maron, M. Anomalous human activity fluctuations from digital trace data signal flood inundation status. *Environ. Plan. B: Urban Anal. City Sci.* 49, 1893–1911, DOI: 10.1177/23998083211069990 (2022). https://doi.org/10.1177/23998083211069990.
- **31.** Schläpfer, M. *et al.* The universal visitation law of human mobility. *Nature* **593**, 522–527, DOI: 10.1038/s41586-021-034 80-9 (2021).
- **32.** Swiss Federal Office of Communications OFCOM. Market shares on mobile networks. https://www.bakom.admin.ch/bak om/en/homepage/telecommunication/facts-and-figures/statistical-observatory/structure-of-the-market-and-employm ent/market-shares-on-mobile-networks.html (2023).
- **33.** Swiss Federal Office for the Environment FOEN. Gap in flood protection closed: New map on surface runoff. https://www.bafu.admin.ch/bafu/en/home/topics/natural-hazards/news-releases.msg-id-71426.html (2018).
- 34. Arabie, P., Hubert, L. & De Soete, G. Clustering And Classification (World Scientific Publishing Company, 1996).
- 35. Carlsson, G. & Vejdemo-Johansson, M. Topological Data Analysis with Applications (Cambridge University Press, 2021).
- **36.** Mathar, R., Alirezaei, G., Balda, E. & Behboodi, A. *Fundamentals of Data Analytics: With a View to Machine Learning* (Springer International Publishing, 2020).
- 37. Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z. & González, M. C. Unravelling daily human mobility motifs. J. *The Royal Soc. Interface* 10, 20130246, DOI: 10.1098/rsif.2013.0246 (2013). https://royalsocietypublishing.org/doi/pdf/1 0.1098/rsif.2013.0246.
- Maeder, S., Stauffacher, M. & Knaus, F. Zooming in and out on everyday mobility practices in a rural, mountainous area of switzerland. *J. Transp. Geogr.* 112, 103680, DOI: https://doi.org/10.1016/j.jtrangeo.2023.103680 (2023).
- **39.** Mohammadi, N. & Taylor, J. E. Recurrent mobility: Urban conduits for diffusion of energy efficiency. *Sci. Reports* **9**, 20247, DOI: 10.1038/s41598-019-56372-4 (2019).
- **40.** Pappalardo, L. *et al.* Returners and explorers dichotomy in human mobility. *Nat. Commun.* **6**, 8166, DOI: 10.1038/ncom ms9166 (2015).
- **41.** Aggarwal, C. & Reddy, C. *Data Clustering: Algorithms and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series (Taylor & Francis, 2013).
- 42. Newman, M. Networks (OUP Oxford, 2018), 2 edn.

- **43.** Fortunato, S. Community detection in graphs. *Phys. Reports* **486**, 75–174, DOI: https://doi.org/10.1016/j.physrep.2009.1 1.002 (2010).
- 44. Fortunato, S. & Newman, M. E. J. 20 years of network community detection. *Nat. Phys.* 18, 848–850, DOI: 10.1038/s4 1567-022-01716-7 (2022).
- Traag, V. A., Waltman, L. & van Eck, N. J. From louvain to leiden: guaranteeing well-connected communities. *Sci. Reports* 9, 5233, DOI: 10.1038/s41598-019-41695-z (2019).
- **46.** Wolf, F. A. *et al.* Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol.* **20**, 59, DOI: 10.1186/s13059-019-1663-x (2019).
- Bryant, A. & Cios, K. Rnn-dbscan: A density-based clustering algorithm using reverse nearest neighbor density estimates. *IEEE Transactions on Knowl. Data Eng.* 30, 1109–1121, DOI: 10.1109/TKDE.2017.2787640 (2018).
- Bedi, P. & Sharma, C. Community detection in social networks. WIREs Data Min. Knowl. Discov. 6, 115–135, DOI: https://doi.org/10.1002/widm.1178 (2016). https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1178.
- **49.** Rostami, M., Oussalah, M., Berahmand, K. & Farrahi, V. Community detection algorithms in healthcare applications: A systematic review. *IEEE Access* **11**, 30247–30272, DOI: 10.1109/ACCESS.2023.3260652 (2023).
- **50.** Wang, B. *et al.* Network enhancement as a general method to denoise weighted biological networks. *Nat. Commun.* **9**, 3108, DOI: 10.1038/s41467-018-05469-x (2018).
- Jia, C., Li, Y., Carson, M. B., Wang, X. & Yu, J. Node attribute-enhanced community detection in complex networks. *Sci. Reports* 7, 2626, DOI: 10.1038/s41598-017-02751-8 (2017).
- 52. Mathworks. Boundary of a set of points in 2-d or 3-d (matlab® r2014b) (2014).
- Lance A. Waller, C. A. G. Spatial Data, chap. 3, 38–67 (John Wiley & Sons, Ltd, 2004). https://onlinelibrary.wiley.com/ doi/pdf/10.1002/0471662682.ch3.
- 54. Agarwal, A. A comparison of weekend and weekday travel behavior characteristics in urban areas. Master's thesis, University of South Florida, https://digitalcommons.usf.edu/etd/936/ (2004).
- **55.** Charlton, P. H., Samuel G.and Baas. Road user interactions: Patterns of road use and perception of driving risk. In *IPENZ Group Technical Conference* (2002). New Zealand.
- Lockwood, A. M., Srinivasan, S. & Bhat, C. R. Exploratory analysis of weekend activity patterns in the san francisco bay area, california. *Transp. Res. Rec.* 1926, 70–78, DOI: 10.1177/0361198105192600109 (2005).
- **57.** O'Fallon, C. & Sullivan, C. Understanding and managing weekend traffic congestion. In *26th Australasian Transport Research Forum Conference* (2003). Wellington, New Zealand.
- **58.** Raux, C., Ma, T.-Y. & Cornelis, E. Variability in daily activity-travel patterns: the case of a one-week travel diary. *Eur. Transp. Res. Rev.* **8**, 26, DOI: 10.1007/s12544-016-0213-9 (2016).
- **59.** Kagho, G. O., Balać, M. & Axhausen, K. W. Understanding of group travel and weekend travel behavior for on-demand mobility services. *Arbeitsberichte Verkehrs-und Raumplanung* **1771**, DOI: 10.3929/ethz-b-000562006 (2022).
- Yuan, Y., Raubal, M. & Liu, Y. Correlating mobile phone usage and travel behavior a case study of harbin, china. *Comput. Environ. Urban Syst.* 36, 118–130, DOI: https://doi.org/10.1016/j.compenvurbsys.2011.07.003 (2012). Special Issue: Geoinformatics 2010.
- 61. Schlich, R. Measurement issues in identifying variability in travel behaviour. In 1st Swiss Transport Research Conference STRC 2001 (2001).
- **62.** Schlich, R. & Axhausen, K. W. Habitual travel behaviour: Evidence from a six-week travel diary. *Transportation* **30**, 13–36, DOI: 10.1023/A:1021230507071 (2003).
- Buliung, R. N., Roorda, M. J. & Remmel, T. K. Exploring spatial variety in patterns of activity-travel behaviour: initial results from the toronto travel-activity panel survey (ttaps). *Transportation* 35, 697–722, DOI: 10.1007/s11116-008-917 8-4 (2008).

- 64. Susilo, Y. O. & Kitamura, R. Analysis of day-to-day variability in an individual's action space: Exploration of 6-week mobidrive travel diary data. *Transp. Res. Rec.* 1902, 124–133, DOI: 10.1177/0361198105190200115 (2005). https://doi.org/10.1177/0361198105190200115.
- **65.** McInerney, J., Stein, S., Rogers, A. & Jennings, N. R. Exploring periods of low predictability in daily life mobility. In *Mobile Data Challenge by Nokia (18/06/12 18/06/12)* (2012).
- **66.** Van Acker, Veronique and van Wee, Bert and Witlox, Frank. When transport geography meets social psychology : toward a new conceptual model of travel behaviour. *Transport Reviews* **30**, 219–240 (2010).
- Song, C., Qu, Z., Blumm, N. & Barabási, A.-L. Limits of predictability in human mobility. *Science* 327, 1018–1021, DOI: 10.1126/science.1177170 (2010). https://www.science.org/doi/pdf/10.1126/science.1177170.
- Jiang, S., Ferreira, J. & González, M. C. Clustering daily patterns of human activities in the city. *Data Min. Knowl. Discov.* 25, 478–510, DOI: 10.1007/s10618-012-0264-z (2012).
- **69.** Kumar, A. & Levinson, D. M. Temporal variations on the allocation of time. *Transp. Res. Rec.* 118–127 (1996). https://hdl.handle.net/11299/179849.
- **70.** Kim, T.-H. & White, H. On more robust estimation of skewness and kurtosis. *Finance Res. Lett.* **1**, 56–73, DOI: https://doi.org/10.1016/S1544-6123(03)00003-5 (2004).
- 71. Bulmer, M. Principles of Statistics. Dover Books on Mathematics Series (Dover Publications, 1979).
- 72. Hair Jr, J. F., Hult, G. T. M., Ringle, C. M. & Sarstedt, M. A primer on partial least squares structural equation modeling (*PLS-SEM*) (Sage publications, 2017), 2 edn.
- 73. Cooksey, R. W. Descriptive Statistics for Summarising Data, 61–139 (Springer Singapore, Singapore, 2020).
- 74. Lock, R. H., Lock, P. F., Lock Morgan, K., Lock, E. F. & Lock, D. F. *Statistics: Unlocking the power of data* (John Wiley & Sons, 2021), third edn.
- 75. Boccaletti, S. *et al.* The structure and dynamics of multilayer networks. *Phys. Reports* 544, 1–122, DOI: https://doi.org/10.1016/j.physrep.2014.07.001 (2014).
- 76. Kivelä, M. et al. Multilayer networks. J. Complex Networks 2, 203–271, DOI: 10.1093/comnet/cnu016 (2014).
- Parshani, R., Buldyrev, S. V. & Havlin, S. Interdependent networks: reducing the coupling strength leads to a change from a first to second order percolation transition. *physical review letters* 105, DOI: 10.1103/PHYSREVLETT.105.048701 (2010).
- 78. Shekhtman, L. M., Danziger, M. M., Vaknin, D. & Havlin, S. Robustness of spatial networks and networks of networks. *Comptes Rendus Physique* 19, 233–243, DOI: https://doi.org/10.1016/j.crhy.2018.09.005 (2018).
- Rinaldi, S., Peerenboom, J. & Kelly, T. Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Control. Syst. Mag.* 21, 11–25, DOI: 10.1109/37.969131 (2001).
- Havlin, S., Stanley, H. E., Bashan, A., Gao, J. & Kenett, D. Y. Percolation of interdependent network of networks. *Chaos, Solitons & Fractals* 72, 4–19, DOI: https://doi.org/10.1016/j.chaos.2014.09.006 (2015). Multiplex Networks: Structure, Dynamics and Applications.
- 81. Radicchi, F. Percolation in real interdependent networks. Nat. Phys. 11, 597-602, DOI: 10.1038/nphys3374 (2015).
- Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* 464, 1025, DOI: 10.1038/nature08932https://www.nature.com/articles/nature08932#supplementary-inf ormation (2010).
- **83.** Gao, J., Buldyrev, S. V., Stanley, H. E. & Havlin, S. Networks formed from interdependent networks. *nature physics* **8**, 40–48, DOI: 10.1038/NPHYS2180 (2012).
- 84. Gao, J., Buldyrev, S. V., Stanley, H. E., Xu, X. & Havlin, S. Percolation of a general network of networks. *physical review e* 88, DOI: 10.1103/PHYSREVE.88.062816 (2013).

- Li, J. *et al.* Recent progress on cascading failures and recovery in interdependent networks. *Int. J. Disaster Risk Reduct.* 40, 101266, DOI: https://doi.org/10.1016/j.ijdrr.2019.101266 (2019).
- **86.** Danziger, M. M. & Barabási, A.-L. Recovery coupling in multilayer networks. *Nat. Commun.* **13**, 955, DOI: 10.1038/s4 1467-022-28379-5 (2022).
- 87. Yadav, N., Chatterjee, S. & Ganguly, A. R. Resilience of urban transport network-of-networks under intense flood hazards exacerbated by targeted attacks. *Sci. Reports (Nature Publ. Group)* 10, DOI: http://dx.doi.org/10.1038/s41598-020-66049 -y (2020).
- Alessandretti, L., Karsai, M. & Gauvin, L. User-based representation of time-resolved multimodal public transportation networks. *Royal Soc. Open Sci.* 3, 160156, DOI: 10.1098/rsos.160156 (2016). https://royalsocietypublishing.org/doi/pdf /10.1098/rsos.160156.
- Aleta, A., Meloni, S. & Moreno, Y. A multilayer perspective for the analysis of urban transportation systems. *Sci. Reports* 7, 44359, DOI: 10.1038/srep44359 (2017).
- **90.** Bassolas, A., Gallotti, R., Lamanna, F., Lenormand, M. & Ramasco, J. J. Scaling in the recovery of urban transportation systems from massive events. *Sci. Reports* **10**, 2746, DOI: 10.1038/s41598-020-59576-1 (2020).
- 91. Gallotti, R., Bertagnolli, G. & De Domenico, M. Unraveling the hidden organisation of urban systems and their mobility flows. *EPJ Data Sci.* 10, 3, DOI: 10.1140/epjds/s13688-020-00258-3 (2021).
- 92. Ferrari, C. & Santagata, M. Vulnerability and robustness of interdependent transport networks in north-western italy. *Eur. Transp. Res. Rev.* 15, 6, DOI: 10.1186/s12544-023-00580-7 (2023).
- **93.** Wang, Q. & Taylor, J. E. Patterns and limitations of urban human mobility resilience under the influence of multiple types of natural disaster. *PLOS ONE* **11**, 1–14, DOI: 10.1371/journal.pone.0147299 (2016).
- 94. Calabrese, F., Pereira, F. C., Di Lorenzo, G., Liu, L. & Ratti, C. The geography of taste: Analyzing cell-phone mobility and social events. In Floréen, P., Krüger, A. & Spasojevic, M. (eds.) *Pervasive Computing*, 22–37 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2010).
- **95.** Pintér, G., Nádai, L. & Felde, I. Analysis of mobility patterns during a large social event. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY), 000339–000344, DOI: 10.1109/SISY.2018.8524674 (2018).
- **96.** Lin, J. Divergence measures based on the shannon entropy. *IEEE Transactions on Inf. Theory* **37**, 145–151, DOI: 10.1109/18.61115 (1991).
- 97. Menéndez, M., Pardo, J., Pardo, L. & Pardo, M. The Jensen-Shannon divergence. J. Frankl. Inst. 334, 307–318, DOI: https://doi.org/10.1016/S0016-0032(96)00063-4 (1997).
- 98. Villani, C. et al. Optimal transport: old and new, vol. 338 (Springer, 2009).
- 99. Villani, C. Topics in optimal transportation, vol. 58 (American Mathematical Soc., 2021).
- 100. Tsybakov, A. Introduction to Nonparametric Estimation. Springer Series in Statistics (Springer New York, 2008).
- 101. Levin, D. & Peres, Y. Markov Chains and Mixing Times. MBK (American Mathematical Society, 2017).
- 102. Daskalakis, C. Lecture notes in "6.896: Probability and Computation, Lecture 3". https://people.csail.mit.edu/costis/6896 sp11/lec3s.pdf (2011). Last access: 19/09/2023.
- **103.** Pollard, D. A User's Guide to Measure Theoretic Probability. Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, 2002).
- 104. Székely, G. J., Rizzo, M. L. & Bakirov, N. K. Measuring and testing dependence by correlation of distances. *The Annals Stat.* 35, 2769–2794 (2007).
- 105. Cha, S.-H. Comprehensive survey on distance/similarity measures between probability density functions. City 1, 1 (2007).

- 106. Lu, Y., Wu, R., Mueen, A., Zuluaga, M. A. & Keogh, E. Matrix profile xxiv: Scaling time series anomaly detection to trillions of datapoints and ultra-fast arriving data streams. In *Proceedings of the 28th ACM SIGKDD Conference* on Knowledge Discovery and Data Mining, KDD '22, 1173–1182, DOI: 10.1145/3534678.3539271 (Association for Computing Machinery, New York, NY, USA, 2022).
- 107. Lu, Y., Wu, R., Mueen, A., Zuluaga, M. A. & Keogh, E. Damp: accurate time series anomaly detection on trillions of datapoints and ultra-fast arriving data streams. *Data Min. Knowl. Discov.* 37, 627–669, DOI: 10.1007/s10618-022-00911-7 (2023).
- **108.** Hehir, C. & Smeaton, A. F. Calculating the matrix profile from noisy data. *PLOS ONE* **18**, 1–15, DOI: 10.1371/journal. pone.0286763 (2023).
- 109. Nilsson, F., Bouguelia, M.-R. & Rögnvaldsson, T. Practical joint human-machine exploration of industrial time series using the matrix profile. *Data Min. Knowl. Discov.* 37, 1–38, DOI: 10.1007/s10618-022-00871-y (2023).
- **110.** Zymbler, M., Polyakov, A. & Kipnis, M. Time series discord discovery on intel many-core systems. In Sokolinsky, L. & Zymbler, M. (eds.) *Parallel Computational Technologies*, 168–182 (Springer International Publishing, Cham, 2019).
- 111. Zymbler, M. & Kraeva, Y. High-performance time series anomaly discovery on graphics processors. *Mathematics* 11, DOI: 10.3390/math11143193 (2023).
- 112. Keogh, E., Lin, J. & Fu, A. Hot sax: efficiently finding the most unusual time series subsequence. In *Fifth IEEE International Conference on Data Mining (ICDM'05)*, 8 pp.–, DOI: 10.1109/ICDM.2005.79 (2005).
- **113.** De Paepe, D. *et al.* A generalized matrix profile framework with support for contextual series analysis. *Eng. Appl. Artif. Intell.* **90**, 103487, DOI: https://doi.org/10.1016/j.engappai.2020.103487 (2020).
- 114. De Paepe, D., Avendano, D. N. & Van Hoecke, S. Implications of z-normalization in the matrix profile. In De Marsico, M., Sanniti di Baja, G. & Fred, A. (eds.) *Pattern Recognition Applications and Methods*, 95–118 (Springer International Publishing, Cham, 2020).
- **115.** Tata, S. & Cagliero, L. A scalable approach to detect contrasting consecutive patterns in time series data. In 2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC), 388–394, DOI: 10.1109/COMPSAC54236 .2022.00061 (2022).
- 116. Quislant, R., Fernandez, I., Gutierrez, E. & Plata, O. Time series analysis acceleration with advanced vectorization extensions. *The J. Supercomput.* **79**, 10178–10207, DOI: 10.1007/s11227-023-05060-2 (2023).
- 117. Yeh, C.-C. M. *et al.* Matrix profile i: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In 2016 IEEE 16th International Conference on Data Mining (ICDM), 1317–1322, DOI: 10.1109/ICDM.2016.0179 (2016).
- 118. Paparrizos, J. *et al.* TSB-UAD: An end-to-end benchmark suite for univariate time-series anomaly detection. *Proc. VLDB Endow.* 15, 1697–1711, DOI: 10.14778/3529337.3529354 (2022).
- Schmidl, S., Wenig, P. & Papenbrock, T. Anomaly detection in time series: A comprehensive evaluation. *Proc. VLDB Endow.* 15, 1779–1797, DOI: 10.14778/3538598.3538602 (2022).
- 120. Chandola, V., Cheboli, D. & Kumar, V. Detecting anomalies in a time series database. Tech. Rep. TR 09-004, University of Minnesota, Department of Computer Science and Engineering, 4-192 EECS Building, 200 Union Street SE Minneapolis, MN 55455-0159 USA (2009).
- 121. He, X., Li, Y., Tan, J., Wu, B. & Li, F. Oneshotstl: One-shot seasonal-trend decomposition for online time series anomaly detection and forecasting. *Proc. VLDB Endow.* 16, 1399–1412, DOI: 10.14778/3583140.3583155 (2023).
- 122. Boniol, P. *et al.* Unsupervised and scalable subsequence anomaly detection in large data series. *The VLDB J.* 30, 909–931, DOI: 10.1007/s00778-021-00655-8 (2021).
- 123. Keogh, E., Dutta Roy, T., Naik, U. & Agrawal, A. Multi-dataset time-series anomaly detection competition, SIGKDD 2021. https://compete.hexagon-ml.com/practice/competition/39/ (2021).

- 124. Wei, L., Keogh, E., Xi, X. & Yoder, M. Efficiently finding unusual shapes in large image databases. *Data Min. Knowl. Discov.* 17, 343–376, DOI: 10.1007/s10618-008-0094-1 (2008).
- **125.** He, Y., Chu, X. & Wang, Y. Neighbor profile: Bagging nearest neighbors for unsupervised time series mining. In 2020 *IEEE 36th International Conference on Data Engineering (ICDE)*, 373–384, DOI: 10.1109/ICDE48307.2020.00039 (2020).
- 126. Lu, Y., Wu, R., Mueen, A., Zuluaga, M. A. & Keogh, E. DAMP_topK.m. https://sites.google.com/view/discord-aware-m atrix-profile/documentation (2023). Last access: 09/10/2023.
- 127. Loreti, S., Lu, Y. & Keogh, E. Exchange of emails between Simone Loreti, Ms. Yue Lu, and Prof. Eamonn Keogh from the Department of Computer Science and Engineering at the University of California, Riverside. (2022). Ms. Lu wrote: "Please note that CurrentIndex/SubsequenceLength is better to be greater than four. We recommend that you allow DAMP to see at least four cycles.". Email received: 22/06/2022.
- 128. Liu, Y., Liu, D., Yu, F. & Ma, Z. A double-density clustering method based on "nearest to first in" strategy. Symmetry 12, DOI: 10.3390/sym12050747 (2020).
- **129.** Vannoy, T. C. Matlab RNN-DBSCAN. https://doi.org/10.5281/zenodo.10279094, DOI: doi.org/10.5281/zenodo.10279094 (2023).

Supplementary Information for "A severe local flood and social events show a similar impact on human mobility"

Simone Loreti^{1,2,*}, Margreth Keiler^{3,4}, and Andreas Paul Zischg^{1,2}

¹University of Bern, Institute of Geography, Bern, 3012, Switzerland

²University of Bern, Oeschger Centre for Climate Change Research, Mobiliar Lab for Natural Risks, Bern, 3012, Switzerland

³University of Innsbruck, Department of Geography, Innsbruck, 6020, Austria

⁴Austrian Academy of Sciences, Institute of Interdisciplinary Mountain Research, Innsbruck, 6020, Austria

*To whom correspondence should be addressed. E-mail: simone.loreti@unibe.ch

Date: Wednesday 27th March, 2024

Supplementary Text

A German-to-English translation of notes about the flood, from the "WSL Swiss Flood and Landslide Damage Database" (1, 2)

"Several thunderstorms passed over the region in quick succession. This led to large amounts of rain. The Cantonal building insurance company of Aargau (AGV) received around 2500 damage reports for the entire canton of Aargau (floods including backwater and slides). In total, buildings suffered damage of around CHF 55 million.

Zofingen. A massive thunderstorm cell with hail and a lot of rain (around 3h, 85 - 90 mm over the catchment areas of the Zofingen city streams) unloaded over the city. Hail and leaves clogged the drains. There was flooding from surface water and overloaded sewer lines with backflows in the buildings. The city area was flooded over a wide area within 30 minutes. Landslides occurred (recorded separately). The areas of Riedtal and Mühlethal as well as areas along the Wigger river were particularly badly affected. Two people who slipped on the mud had to be taken to hospital. Because nine transformer stations were under water, the power failed in large parts of the city. The emergency call center and the fire department received about 550 damage reports. Around 250 - 450 rescue workers were deployed. The fire department pumped water out from more than 200 cellars. The storm flooded hundreds of businesses, basements, underground garages and underpasses. The Zofingen fire department performed 1800 person—hours, the civil defense organization of the region another 4700 hours. The AGV assumed 977 damages in Zofingen (incl. backwater). Based on this information, we calculated 42 million CHF (including furniture). Note: Taking this information into account, an initial estimate after the event of over CHF 100 million in damages seems too high.

Mühlehtal. Among the worst hit were the textile finishing company Bethge AG in Mühlethal, which was devastated by water and mud masses, and a private property in the same location, which was hit by a slides (recorded separately). Parts of the hospital in Zofingen were up to 40 centimeters under water. In the meantime, the elevators failed. The water ran from the sewage system into the hospital's archives -85000 patient files were destroyed. A data recovery team froze the files and then placed them in a vacuum tank. The work took several months. In the military accommodation Rosengarten there was a major fire department operation. The water damaged all the equipment and technology. Several walls and floor slabs were destroyed by the lateral pressure and buoyancy of the water. The accommodation had to be completely renovated; it was not expected to be rentable again until the beginning of 2018.

Industrial area (west of the railway tracks). Water penetrated into the industrial facility of Swissprinters: there was considerable damage to the printing devices, the building, and the logistics hall. The dirty water left a picture of destruction. Up to 2000 tons of paper were rendered unusable. The loss amounted of several millions Swiss Francs. At the pharmaceutical company Siegfried AG, the basements of production facilities, warehouses and administration buildings were under water. There was major damage. Production could be resumed in the night of the 9^{th} of June. There was only minor damage at Muller Martini and the Delta company. Cellars were also flooded on Henzmannstrasse. A traffic circle on Strengelbacherstrasse was under water.

Old town. At the Blumenheim retirement center, the basements of all buildings were affected (in part, 1 m high water and mud). Heating, elevator, washing machines, cooling systems and power distribution were damaged. The home's manager estimated the damage at over 1 million CHF. The rooms in the basement of the Lloyd Café were unusable for months. Huge damage also occurred in the bar "Key69". The entire wooden furnishings had to be replaced. In addition to a flooded basement, various stores had to deal with soaked and broken floors on the first floor. In the town hall, all carpets had to be replaced. Bar staff and guests were locked in for about 2h. Food and beverage storage, storage and tank rooms were flooded. The water in the basement was so high that the ladies' room was no longer visible. At the Xpert AG insurance company's store on Rathausgasse, the water penetrated through the toilets into the basement. The store was flooded. At the Leutwyler bakery, all the equipment was broken due to the flooding of the basement. The building of the regional administration was under water. The water penetrated into the parking facility of the main station and flooded two basements with over 100 parked vehicles (they suffered total damage). 7 vehicles of the regional police Zofingen were under water in the 2^{nd} basement of the parking station and had to be replaced. On the 21^{st} of July the parking garage could be reopened. The power supply and electrical controls, the ventilation, the sprinkler system as well as the fire doors and the ticket machines had to be repaired beforehand. From the Blumenheim retirement center, the water flowed onto General Guisan road and on to the Luzernerstrasse traffic circle. Pomernweg was under water. The Strengelbacher road underpass under the railway tracks was flooded several meters deep. A motorist who got stuck in the flooded Henzmann road underpass was able to rescue himself from the car. One vehicle got stuck in the Altachen road underpass. Due to the storm, the Bahnhofplatz and two railway tracks in Zofingen were also under water, which had an impact beyond the region: the line between Basel and Ticino was interrupted between Olten and Lucerne. The flood washed out and damaged several roads. As a result of the heavy hail, leaves, branches and ferns were cut off in the Zofingen forest, which was washed away by the rainwater and filled road ditches and blocked the passages. The water then sought a new path, sweeping away mud and gravel. Half of the road in the Brunngraben area was washed away. At the Stöckliacher junction, the tar pavement was damaged. The damage to roads was estimated at 0.2 million CHF."

Supplementary Tables and Figures

Day	Μ	Ionth	Hours	Event	Des	cription			Municipality
				30th Aargau Cantonal	A ta and of a day: end and and mar was Her	riget shooting co 300 m) and pis round 6500 sho s, which was sp s (i.e. 3 Friday nday). The firs Monday. Each ksmen took par held in 7 shooti e following the	ompetition that ir stol (25 and 50 m poters attended the read out over thr s, 3 Saturdays and t long weekend, it in the competiti ng ranges, in 7 di exact positions of	cluded both rifle (50) disciplines. A total le competition of ten ee consecutive week- nd 3 Sundays, plus 1 ncluded both Friday approximately 2300 on. The competition fferent municipalities.	
Friday	16 Ju	ine	08 - 19	Shooting Festival (ACSF) (3, 4)	-	Municipality Zofingen	Latitude (N) 47°16′59.09″	Longitude (E) 07°57′30.56″	Multiple
						Kölliken	$47^{\circ}19'15.96''$	08°01′12.40″	
						Aarburg	47°19′30.11″	07°54′33.44″	
						Brittnau	47°14′57.44″	07°57′08.93″	
						Reitnau	47°15′15.52″	08°01′53.98″	
						Staffelbach	47°16′39.54″	08°02′27.28″	
						Muhen	47°20′41.96″	08°02′48.12″	
					5	05 Attendees			
	08 - 19 ACSF (3, 4) On Saturdays, the shooting r only in Aarburg and in Zofin 902 Attendees		shooting match to 1 in Zofingen.	ook place until 18:00	Multiple				
Saturday 17	I7 Ju	ine …	9 - 12	Edi Schneider Cup (5, 6)	Hoc (IH0 ∎ 15	ckey tournament C) of Rothrist. 0 Attendees	t organised by the	e Inline Hockey Club	Rothrist
Sunday	18 Ju	ine	08 – 17	ACSF (3, 4)	As all Sundays, in Aarburg, the shooting was from 09:00 to 16:00. Only on Sunday 18^{th} , the afternoon session of shooting did not take place in Zofingen. = 444 Attendees			Multiple	
Thursday	22 Ju	ine		Senior citizens excursion (7)	3 0	00 Attendees			Oftringen
			08 - 19	ACSF (3, 4)		803 Attendees			Multiple
Friday	23 Ju	ine	14 — 21	Bio Marché (4, 8)	The betv The 23 ^{re} 24:0	largest Swiss of veen 35000 and fair was held in ^d , gastronomy 00.	organic trade fair 40000 visitors ov 1 the old town of events and conc	with approximately ver a 3-day weekend. Zofingen. On Friday erts took place until 12000 A.	Zofingen
			08 – 19	ACSF (3, 4)	On only	Saturdays, the s y in Aarburg and 969 Attendees	shooting match to l in Zofingen.	ook place until 18:00	Multiple
Saturday	24 Ju	ine	10 - 21	Bio Marché (4, 8)	On plac	Saturday 24 th , ce until 24:00.	gastronomy ever	tts and concerts took	Zofingen
Sunday	25 Ju	ine	08 – 17	ACSF (3, 4)	On 5	Sundays, the sho 00 only in Aarbu 593 Attendees	ooting match tool 1rg.	c place from 09:00 to	Multiple
Sunday	<i>20</i> 30		10 - 18	Bio Marché (4, 8)				12000 A.	Zofingen
				Day of People (5)	Reli	igious event of t	he Roman Catho	lic Parish of St. Paul.	Rothrist

Day	y Month Hours		Event	Description	Municipality	
		08 - 19	ACSF (3, 4)	948 Attendees	Multiple	
		20 - 02	School Festival & City Festival (9)	Two Festivals combined together, taking place over a $3-day$ weekend. The event was held among two fair- grounds ("Kirchgasse" and "Schützenmatte") in Olten, ap- proximately where the Dünnern river meets the Aare river. On Friday 30^{th} , City Lounge since 20:00. It was not possi- ble to estimate the number of attendees for the entire event (10).	Olten	
Friday	30 June	17 – 23	Youth Festival (11, 12)	Festival for young people, on Friday evening and Satur- day. The schedule for Friday evening included disco music (17:30 - 19 and 20:30 - 23), a concert $(19 - 20)$, and amusement park-related activities (from 17). \blacksquare 600 Attendees	Safenwil & Walterswil	
		18-23	Children's Festival (7)		Oftringen	
		$\frac{16^{30}}{20^{30}}$	Music Chat Reiden Mitte (13)	One evening of short concerts performed by four different music groups of the Reiden's Music Society (the Brassinis, the Junior Brass Band, the Brass Band, and the Senior Music).	Reiden	
		13 ³⁰ - 24	Children's Festival (5, 14)	Festival for the pupils of the Rothrist School, over a 3–day weekend. On Friday 30 th , a school Open Day was organised for public. An amusement park was set up for the evening. 1150 Attendees	Rothrist	
		08 – 19	ACSF (3, 4)	On Saturdays, the shooting match took place until 18:00 only in Aarburg and in Zofingen. 971 Attendees	Multiple	
		09 ³⁰ - 02	School Festival & City Festival (9)	Kids (09:30 $-$ 12:30) and adults (16 $-$ 18:30) races with office chairs, plus concerts and disco music (16 $-$ 24). Live musical performances (live-acts), In Kirchgasse, street food, bar service (09 $-$ 02), and live-acts, i.e. live musical performances (20 $-$ 02).	Olten	
	1 July	$10 - 22^{30}$	Youth Festival (11, 12)	Several concerts, exhibitions, and amusement park-related activities took place over the whole day. Fireworks were set off at 22:30. Although a road section was closed to traffic, and a detour was arranged by the fire department, no traffic congestion was reported (12). 2000 Attendees	Safenwil & Walterswil	
		10 - 24	Children's Festival (7)	1500 Attendees (btw. 1000 and 2000)	Oftringen	
		10 – 24	Children's Festival (5, 14)	On Saturday 1^{st} , the major activities included a children's parade (10 – 11:30), the speech of the President of the Canton of Aargau - Benjamin Giezendanner - (11:45), several concerts (13:30 – 17:30) and a recreational sport event (12 – 16). An amusement park was set up for the entire afternoon and evening (12 – 24).	Rothrist	
		17 - 02	Midsummer Night Festival (15)	Concerts by the Music Society of Brittnau.	Brittnau	
		$16^{30} - 22$	Village Festival (16)	Concerts by the Brass Band Uffikon-Buchs and by the Bärgblueme Dagmersellen yodelling club.	Dagmersellen	

Continued on next page

Day	Mont	h Hours	Event	Description	Municipality
		08 – 17	ACSF (3, 4)	On Sunday 2 nd , the shooting range of Reitnau was closed. In Aarburg, the shooting match lasted from 09:00 to 16:00, as all Sundays. 358 Attendees	Multiple
Sunday 2	2 July	07 – 21	School Festival & City Festival (9)	Drums concert in the city center (starting at 07:00), parade (starting at 09:45), Olten Music School's and Adrian Stern's concerts ($12:30 - 18:00$), kids games ($13 - 17$), and electronic music (up to 21).	Olten
		$09^{30} - 19$	Children's Festival (14)	Ecumenical worship (09:30 $-$ 10:30) and amusement park (11 $-$ 19).	Rothrist
Monday 3	3 July	17 – 00:30	New Orleans Meets in Zofingen (NOMZ) (17)	An evening-long jazz & blues festival $(17 - 00:30)$. Eight bands performed over three stages in the old town of Zofin- gen (in Thutplatz, in Kirchplatz and in Chorplatz). Food and beverage service provided by around 20 stands.	Zofingen
Thursday 6	5 Iuly	18 - 03	Tattoo Concert & Children's Festival (18)	Music parade of marching drummers in the old town of Zofingen (19:30 $-$ 20), followed by "Tattoo" concerts (typically called "Zapfenstreich" in the German-speaking countries) still in the old town of Zofingen (20 $-$ 02). Music stopped at 02 in the Sundays's morning. The overall event started with some opening activities, promoted by the Reformed Church's community (18 $-$ 19:30).	Zofingen
Inursday 6 Ji		17 – 20	Fashion Show by SMG Olten (19)	Two fashion shows, where the students of SMG Olten presented their own work to the public. Each show lasted around one hour. The shows took place at the main entrance staircases of the BBZ building, where approximately 150 seats were set, as well as an area were around 100 guests could stand. • 250 Attendees	Olten
Friday 7	7 July	07 – 23	Tattoo Concert & Children's Festival (18)	The morning and the first afternoon were marked by military-like parades of the Zofingen's Cadets Corp in the old town of Zofingen $(07 - 10, 11 - 11:45, and 13:45 - 14:45)$, and by a solemn celebration in the Reformed "Town" Church of Zofingen $(07 - 11:30)$. The afternoon was instead characterized by a battle reenactment on a parade ground, called "Heiternplatz", southeast of the Zofingen's old town (14:45 - 16), and by a number of concerts $(16 - 19:30)$. The Festival ended with a torch and lantern procession $(22 - 23)$.	Zofingen
		18 - 23	Dinner (20)	Fish dinner of the Music Society of Staffelbach	Staffelbach
Friday 1	14 July	19 - 03	Ski Festival (21, 22)	Food and drink, and music festival. 2000 Attendees	Rothrist
Saturday 1	15 July	18 - 03	Ski Festival (21, 22)	Food and drink, and music festival.	Rothrist
		$20^{30} - 22$	Variété Pavé (23)	Street theatre in the old town of Olten (Kirchgasse).	Olten
Friday 2	21 July	18 - 02	Beach Festival (24, 25)	Musical entertainment, artisan food products and amuse- ment park. 2000 Attendees	Aarburg

Continued on next page

Day	Month	Hours	Event	Description	Municipality
		$20^{30} - 22$	Variété Pavé (23)	Street theatre in the old town of Olten (Kirchgasse).	Olten
Saturday 22	July	18 - 02	Beach Festival (24, 25)	Concerts and artisan food products. On Saturday, the high- light of the Festival was a choreographed musical firework display, where the background music was synchronized with the fireworks. Shortly after darkening (sunset around 21:15), fireworks were set off from the banks of the Aare river, and they lasted around half an hour. 5000 Attendees	Aarburg
Sunday 23	July	$20^{30} - 22$	Variété Pavé (23)	Street theatre in the old town of Olten (Kirchgasse).	Olten
Saturday 29	July	17 - 04	Summer Party (26, 27)	Music performances and street food. A fireworks display in the evening. 500 Attendees	Reiden

Supplementary Table 2. Descriptive statistics about (**a**) the distributions of distances among the clusters centers and event centers, and (**b**) the distributions of clusters areas, for each major event. With Q_1 , Q_2 and Q_3 we indicate, respectively, the lowest quartile (i.e. the 25th percentile), the median (i.e. the 50th percentile), and the upper quartile (i.e. the 75th percentile) of the distribution. Those three quartiles are necessary to calculate the Bowley's coefficient of skewness (BS), i.e. $BS = \frac{(Q_3+Q_1-2\times Q_2)}{(Q_3-Q_1)}$. The Fisher-Pearson coefficient of skewness (FPS) needs instead the mean of the distribution, i.e. $FPS = \frac{\frac{1}{n_s}\sum_{i=1}^{n_s}(x_i-\langle x \rangle)^3}{(\sqrt{\frac{1}{n_s}\sum_{i=1}^{n_s}(x_i-\langle x \rangle))^3}}$, where x_i represents either each distance between a cluster and the event, or each cluster area,

 $\langle x \rangle$ indicates the mean of the distribution, and n_s is the sample size (i.e. $i = 1, 2, ..., n_s$). In FPS, we can recognise the standard deviation of the distribution, i.e. $\sigma = \sqrt{\frac{1}{n_s} \sum_{i=1}^{n_s} (x_i - \langle x \rangle)}$.

(a). Summary statistics about the distributions of distances among clusters centers and events centers.

Event $(\times 10^3 m \rightarrow)$	Min	Q1	Q2	Q3	Max	Mean	Bowley Skewness	FP Skewness
Bio Marché Fri	0.86	4.18	5.62	7.85	12.66	5.92	0.21	0.16
Bio Marché Sat	0.89	3.61	5.43	7.86	11.8	5.65	0.14	0.05
Bio Marché Sun	0.39	3.58	4.92	7.49	10.78	5.39	0.32	0.17
NOMZ	0.85	1.57	4.05	7.15	8.05	4.28	0.11	0.07
Tattoo Concert Thu	0.8	1.46	2.3	3.82	7.12	2.68	0.29	0.97
Tattoo & Child. Fest	0.95	3.54	4.96	7.82	12.29	5.5	0.34	0.39
Flood	0.32	3.55	5.52	7.6	11.69	5.75	0.03	0.54
Ski Fest. Fri	0.73	4.01	6.02	8.77	13.48	6.36	0.16	0.18
Ski Fest. Sat	0.76	3.02	5.2	7.32	12.55	5.34	-0.01	0.56
Beach Festival	0.27	2.72	4.05	6.18	11.12	5.03	0.23	1.25
Summer Party	1.64	5.62	8.86	11.93	16.33	8.72	-0.03	-0.1

(b). Summary statistics about the distributions of clusters areas.

Event $(\times 10^5 m^2 \rightarrow)$	Min	Q1	Q2	Q3	Max	Mean	Bowley Skewness	FP Skewness
Bio Marché Fri	0	0.89	2.86	10.9	25.23	11.19	0.61	4.68
Bio Marché Sat	0	0.7	2.44	6.3	13.32	11.01	0.38	10.58
Bio Marché Sun	0.01	0.81	3.22	10.3	24.12	11.7	0.49	8.69
NOMZ	0	0.14	0.34	2.52	4.58	7.23	0.83	4.22
Tattoo Concert Thu	0	0.41	1.56	5.44	10	9.92	0.54	3.74
Tattoo & Child. Fest	0.01	0.64	3.28	10.9	23.13	14.31	0.48	7.07
Flood	0	0.67	2.14	8.28	19.22	8.93	0.61	3.66
Ski Fest. Fri	0	0.28	1.38	13.08	29.83	21.5	0.83	6.12
Ski Fest. Sat	0	0.12	1.29	7.93	18.59	11.91	0.7	4.14
Beach Festival	0.02	1.85	4.6	15.03	33.36	16.32	0.58	3.88
Summer Party	0	0.59	2.38	6.27	14.22	8.23	0.37	4.42

Supplementary Table 3. Coefficients of the empirical relationships among distances (between the clusters centers and the event center), clusters areas, and number of clusters, for each event. The goodness of fit of the three relationships is assessed through the coefficient of determination, R^2 .

(a). Coefficients of $r_{\text{max}} = a_1 \Delta r + b_1$, where r_{max} is the distance of the farthest cluster (from the event center), and $\Delta r = r_{\text{max}} - r_{\text{min}}$ is the difference between the farthest and the clostest cluster distances.

Event	a ₁	b ₁	R ²
Bio Marché Fri	0.84	3020.22	0.87
Bio Marché Sat	0.79	3163.22	0.91
Bio Marché Sun	0.61	4525.32	0.91
NOMZ	0.99	1641.08	0.71
Tattoo Concert Thu	0.86	1596.33	0.93
Tattoo & Child. Fest	0.85	2802.61	0.92
Flood	0.81	3573.71	0.87
Ski Fest. Fri	0.45	7512.03	0.79
Ski Fest. Sat	0.69	4611.82	0.8
Beach Festival	1	1138.07	0.95
Summer Party	0.33	10554.52	0.37

(b). Coefficients of $n_c = a_2 \langle A_c \rangle^{b_2}$, where n_c is the number of clusters and $\langle A_c \rangle$ is the average size of the clusters areas.

Event	a ₂	b ₂	R ²
Bio Marché Fri	1078.56	-0.32	0.28
Bio Marché Sat	968.44	-0.32	0.6
Bio Marché Sun	16.7	-0.06	0.01
NOMZ	1.59	0.05	0.03
Tattoo Concert Thu	6.07	-0.03	0.03
Tattoo & Child. Fest	31.04	-0.09	0.03
Flood	5417.26	-0.45	0.65
Ski Fest. Fri	478.91	-0.31	0.94
Ski Fest. Sat	311.49	-0.26	0.68
Beach Festival	5.86	0.02	0.01
Summer Party	725459.48	-0.82	0.41

(C). Coefficients of $d_{c(r_{min})} = a_3 log(A_{c(r_{min})}) + b_3$, where $d_{c(r_{min})}$ is the distance closest cluster, and $A_{c(r_{min})}$ is the area closest cluster.

Event	a3	b3	R ²
Bio Marché Fri	204.84	-1265.51	0.65
Bio Marché Sat	268.55	-1955.41	0.4
Bio Marché Sun	82.13	743.56	0.04
NOMZ	-86.25	2149.05	0.28
Tattoo Concert Thu	58.14	391.11	0.7
Tattoo & Child. Fest	102.77	198.7	0.2
Flood	-111.98	3438.23	0.03
Ski Fest. Fri	533.33	-4724.38	0.37
Ski Fest. Sat	33.38	1280.12	0.02
Beach Festival	63.06	196.62	0.01
Summer Party	526.67	-4154.87	0.92

Supplementary Table 4	. List of symbols in	the order in which t	they appear.
-----------------------	----------------------	----------------------	--------------

Symbol	Description
i, j, k	Indices
P_i and P_j	Estimated probability distributions
$D = (d_{ijk})$	Dissimilarity matrix (of vectors), with element d_{ijk}
$f_k(P_i, P_j)$	k^{th} -function of distance or divergence
Q	Quartiles of a distribution
$\langle Q_2 \rangle$	Average median (boxplots)
$r_{\rm max}$ and $r_{\rm min}$	Maximum and minimum distances of clusters from the event center
Δr	$r_{\max} - r_{\min}$
n_c	Number of clusters
A_c	Area of clusters
$\langle A_c \rangle$	Mean area of the clusters
$d_{c(r_{min})}$	Distance of the closest cluster
$A_{c(r_{min})}$	Area of the closest cluster
$a_1, a_2, a_3, b_1, b_2, b_3$	Coefficients of the empirical relationships
${\mathcal X}$ and ${\mathcal Y}$	Sample spaces (Polish metric spaces in the Wasserstein distance)
\mathcal{A}	σ -algebra
$(\mathcal{X},\mathcal{A})$	Measurable space
ρ_1 and ρ_2	Densities of P_1 and P_2
ω_1 and ω_2	Weights of P_1 and P_2
H(x)	Mixture distribution of P_1 and P_2 .
cost(x,y)	Cost function
μ and ν	Probability measures
π	Joint probability measure
$\Pi(\mu, u)$	Set of all joint probability measures on $\mathcal{X}\times\mathcal{Y}$ whose marginals are μ and ν
F and G	Cumulative distribution functions (CDFs) associated with μ and ν
F^{-1} and G^{-1}	Quantile functions of μ and ν
(\mathcal{X},d)	Polish metric space equipped with a metric $d(x,y)$
<i>p</i>	p-th power of the metric $d(x,y)$
$f_{SE}(P_i)$	Shannon entropy function
$f_{JS}(P_1, P_2)$	Jensen–Shannon divergence
$f_{KL}(P_1 \parallel P_2)$	Kullback–Leibler divergence
$f_{TV}(P_1, P_2)$	Total variation distance
$f_H(P_1, P_2)$	Hellinger distance
$W_p(\mu, u)$	Wasserstein distance of order p between μ and ν
$W_1(\mu, u)$	W_1 form of the Wasserstein distance (Kantorovich–Rubinstein distance)
X and Y	Random variables
$f_{DC}(X,Y)$	Empirical distance correlation
$f_{DC}^{\complement}(X,Y)$	Complementary empirical distance correlation
$\mathcal{V}(X), \mathcal{V}(Y), \mathcal{V}(X, Y)$	Coefficients to calculate the empirical distance correlation
$a_{ij}, b_{ij}, A_{ij}, B_{ij}$	Coefficients to calculate the empirical distance correlation

Continued on next page

Symbol	Description
u and v	Dimensions of spaces \mathbb{R}^u and \mathbb{R}^v (in the empirical distance correlation)
$f_{CS}(X,Y)$	Cosine similarity
$f_{CD}(X,Y)$	Cosine distance
n	Length of the time series
m	Length of the time subsequence
$T = (t_k)_{k=1}^n$	Time series, with element t_k
$T_{i,m}$	Time subsequence of length m , starting at i
δ_i	Distance profile
М	Matrix profile
δ_i^*	Left distance profile
M^*	Left matrix profile
Ι	CurrentIndex (split point between the training data and the test data)
Υ	Set of points in \mathbb{R}^2
ϵ	Radius defining the neighborhood of point <i>i</i>
ϵN_i	ϵ -neighborhood of point i
γ	Minimum number of points that lie within a ϵ -neighborhood
k-NNG	k-nearest neighbor graph
kNNi	k-nearest neighborhood of an arbitrary node i
$RkNN_i$	Reverse k -nearest neighborhood of node i
$k_i^{ ext{out}}$	"Outgoing degree" of node i
$k_i^{ m in}$	"Ingoing degree" of node i
(i,j)	Edge from node <i>i</i> to node <i>j</i>
c_k	Number of clusters for different values of k (from RNN-DBSCAN)
$C = (c_k)_{k=1}^{100}$	Sequence of (100 dfferent) number of clusters, when k varies
$\operatorname{img}(C) = \{c_k\}_{k=1}^{100}$	Image of C
$C^{-1}(\{c_k\})$	Preimage of element c_k

Supplementary Table 4. List of symbols in the order in which they appear.



Supplementary Figure 1. Spatio-temporal dynamics of anomalies (pink edges) during NOMZ in the municipality of Zofingen (green patch).





Supplementary Figure 2. Spatio-temporal dynamics of anomalies (pink edges) during the Tattoo Concert & Children's Festival on Thursday the 7^{th} of July that occurred in the municipality of Zofingen (green patch).





Supplementary Figure 3. Spatio-temporal dynamics of anomalies (pink edges) during the Beach Festival that occurred in the municipality of Aarburg (green patch).





Supplementary Figure 4. Spatio-temporal dynamics of anomalies (pink edges) during the flooding event that occurred in the municipality of Zofingen (green patch) and its surroundings. Roads with flooded buildings are highlighted in blue.



Supplementary Figure 5. Distribution of distances among the edges with anomalies and the municipality center where an event took place, for every major documented event.



Supplementary Figure 6. Test of RNN-DBSCAN clustering over the range $1 \le k \le 100$, on a number of artificial datasets (28): (a) aggregation, (b) r15, (c) flame, (d) spiral, (e) d31, and (f) jain. On the left side of the figure, the number of clusters is shown against its frequency of occurrence (bars). This is depicted together with the maximum Adjusted Rand Index (ARI) performance (solid line) and the ARI performance at the minimum k (dash-dotted line). On the right side of the figure, a graphical representation of the resulting clusters is depicted, where the number of resulting clusters corresponds to the local maxima on the furthest left of the adjacent histogram (this local maxima is called as the "first frequency spike" by Bryant et al. (29)). Since different k values contributed to the construction of the bar corresponding to the leftmost local maxima, and since there is a positive correlation between the maximum Adjusted Rand Index (ARI) performance and the ARI performance at the minimum k (29), we use the minimum k that produced the leftmost local maxima, as input parameter for producing the clusters on the right side of the figure. This test shows similar results to those ones presented by Bryant et al. (29) (please see Figure 4 of their paper). 19/22

References

- 1. Swiss Federal Institute for Forest, S. & Landscape Research WSL. WSL Swiss Flood and Landslide Damage Database. Exchange of emails between Mobiliar Lab for Natural Hazards and Ms. Norina Andres, Researcher at WSL. www.wsl.ch/usdb (2019). Email received: 28/03/2019.
- Andres, N. & Badoux, A. The Swiss flood and landslide damage database: Normalisation and trends. *J. Flood Risk Manag.* 12, e12510, DOI: https://doi.org/10.1111/jf r3.12510 (2019). https://onlinelibrary.wiley.com/doi/pdf/ 10.1111/jfr3.12510.
- **3.** District Shooting Association of Zofingen. 30th Aargau Cantonal Shooting Festival (ACSF). https://www.aarg auerzeitung.ch/aargau/kanton-aargau/6000-schutzen-haben-den-kranz-im-visier-ld.1435201, https://www.shoot.ch/resultat/resksf.aspx?resid=AGKSF2017, and https://bsv-zofingen.ch/ (2017).
- **4.** Municipality of Zofingen. Exchange of emails between Simone Loreti and Ms. Elvira Zueger, Marketing & Events Specialist at the City Office of Zofingen. https://www.zofingen.ch/verwaltung/stadtmarketingstadtbuero.html/40/entityId/20/entityType/egov_unit (2022). Email received: 23/06/2022.
- 5. Municipality of Rothrist. Exchange of emails between Simone Loreti and Ms. Gabriela Meyer, Deputy Head of Resident Services of Rothrist. https://www.rothrist.ch/ve rwaltung-politik/verwaltung/einwohnerdienste/ (2023). Ms. Meyer mentioned the Children's Festival: "At the Children's Festival, there is no exact information about the number of participants from each individual day. The school has only found one report from that time [stating the following]: On the afternoon of games, all the teachers offered the approximately 1150 pupils an attractive range of games at over 40 [places]. On Saturday [the 1^{st} of July 2017] morning, the children's parade with over 30 numbers [?] was held in good weather. Afterwards, more than 3000 visitors took part in the festive speech [of] the President of the Grand Council, Benjamin Giezendanner, on the Dörfliareal. Various associations supported the school with food stands and activities for the pupils, on the festival grounds.". About the Edi Schneider Cup (i.e. an inline hockey tournament) that occurred on the 17^{th} of June 2017, Ms. Meyer stated that "there were a total of approximately 100 - 150 people (100 people on site at the same time, 150 [people in] total over the event), including approximately 35 players.". Email received: 07/03/2023.
- 6. Inline Hockey Club IHC of Rothrist. Edi Schneider Cup in Rothrist. https://www.facebook.com/photo/?fbid =1447543055306625&set=pcb.1447544905306440, https://www.facebook.com/permalink.php?story_fbid= pfbid0YhszMExiDTuhab5VeWi6Mqzx6arHyGSiaPgJ

CrPST5CencFDMdqzB61ZPxRY9DWNl&id=1017696 16550649, and https://1drv.ms/f/s!Au_0rvYOre9BgbRG RWhMCXPUc547ng (2023).

- 7. Municipality of Oftringen. News from websites and exchange of emails between Simone Loreti and Mr. Christoph Kuster, Communal Secretary of Oftringen. ht tps://www.mgok.ch/wordpress/Veranstaltung/kinderfest-oftringen-2/, https://www.oftringen.ch/gemeindekanzlei, and https://www.schule-elternhaus.ch/sektionen/kanton-aargau/veranstaltungen-1/ (2022). Email received: 30/06/2022.
- 8. Presseportal & Bio Marché AG. News from the news aktuell press agency (through presse portal) and exchange of emails between Simone Loreti and Ms. Bianca Braun, Advertising, Communication and Online Media Specialist of Bio Marché. https://www.presseportal.ch/de/pm/100010359/100801665, and https://www.biomarche.ch/ (2017).
- 9. Municipality of Olten. School Festival & City Festival. https://www.olten.ch/newsarchiv/363578, https://www.fa cebook.com/events/1531642226911002/?ref=newsfeed, https://www.facebook.com/buerostuhlrennenWM/phot os/a.134634097001839/281726865625894/, https://issu u.com/stadtanzeigerolten/docs/stadtanzeiger_olten_au sgabe_24__14., https://www.oltnertagblatt.ch/solothur n/olten/musik-ist-trumpf-am-stadt-und-schulfest-vielesternchen-proben-fur-einen-stern-ld.1436381 and https: //www.stadtanzeiger-olten.ch/fileadmin/user_upload /sao/pdf/Agenda/20170629_WOZ_STOANZ_014.pdf (2017).
- 10. Municipality of Olten. Exchange of emails between Simone Loreti and Ms. Miriam Siegrist, Student Administration Handler at the Municipality of Olten (2023). About a potential estimate of the number of attendees, Ms. Miriam Siegrist wrote: "Unfortunately, it is not possible for us to provide you with these numbers. The school festival took place in different places and there was also a parade on Sunday. However, we do not have any figures for the spectators." Email received: 22/02/2023.
- Schelshorn Sonja. Youth Festival. https://www.facebo ok.com/events/safenwil-switzerland/jugendfest-2017kreisschule-safenwil-walterswil/1343137335766384/ (2017).
- 12. Schelshorn Sonja. Exchange of emails between Simone Loreti and Ms. Sonja Schelshorn, Head of School Administration and Services at the District School of Safenwil-Walterswil. https://www.schule-sawa.ch/ (2023). Ms. Schelshorn provided an estimate of the number of attendees: "First day: approximately 600 [people]. Second day: circa 1800-2000 [people]". Then, she mentioned the road traffic circulation during the event: "There were no traffic jams. On Saturday, a section of the main road was

closed during the parade, as well as another small section. A traffic diversion was put in place by the fire department". Also, Ms. Schelshorn reported information about the main transportation means used by the attendees for reaching the Festival: "Since the majority of residents were in Safenwil and Waltersil, many came on foot. We also had enough parking spaces [and] the train [station] is right near the festival site (4 minutes on foot). However, there are no surveys.". Email received: 24/02/2023.

- Brass Band MG Reiden. Music Chat. http://www.mg reiden.ch/anlaesse/musikplausch-reiden-mitte-126, https://www.facebook.com/events/302212530231184/ ?active_tab=discussion, and https://www.gewerbevereinreiden.ch/wp-content/uploads/2017/06/2017-3.pdf (2017).
- **14.** School of Rothrist. Children's Festival of Rothrist. https: //www.schule-rothrist.ch/kinderfest (2017).
- Music Society of Brittnau. Midsummer Night Festival of Brittnau. https://www.facebook.com/events/3101964027 63453/?ref=newsfeed and https://www.facebook.com/1 38639826237283/photos/a.351764074924856/110150 6193283970/?type=3 (2017).
- 16. Brass Band Uffikon Buchs. Village Festival of Dagmersellen. https://www.facebook.com/721745341 246865/photos/gm.638836723166777/1784756958279 026/?type=3 and https://www.facebook.com/events/d41d 8cd9/dorffest/439238746459910/ (2017).
- 17. Municipality of Zofingen. New Orleans meets in Zofingen. http://zofingen.kcdb.net/uploads/dokumente/050717_NOMZ.pdf, https://www.mobilestage.ch/artikel/new-orleans-meets-zofingen/, https://www.facebook.com/events/682160515306176?active_tab=about, and https://issuu.com/ztonline/docs/mag_0217_web (2017).
- 18. Municipality of Zofingen. Exchange of emails between Simone Loreti and Ms. Elvira Zueger, Marketing & Events Specialist at the City Office of Zofingen. https: //www.pulverturm-zofingen.ch/assets/downloads/pdfund-powerpoint/Programm_Kinderfest_2017_def.pdf, https://www.aargauerzeitung.ch/aargau/kantonaargau/erbitterter-kampf-um-den-gestohlenenrathausler-ld.1293115, and https://issuu.com/ztonli ne/docs/mag_0217_web/22 (2022). Email received: 23/06/2022.
- 19. School of Fashion & Design (SGM) of Olten. Fashion Show SMG Olten. http://www.smg-olten.ch/wp-content/ uploads/2017/05/01_Var_A3_Plakat_Modeschau_SMG _2017_LR.pdf, https://www.facebook.com/BBZOlten/ posts/modeschau-smg-olten-eine-reise-um-die-weltdiehauptproben-sind-erfolgreich-abges/13987048535487 04/, and https://bbzolten.so.ch/aktuell/aktuell/news/einereise-um-die-welt/ (2017).

- **20.** Music Society of Staffelbach. Fish Dinner. https://staffe lbach.ch/scms/upload/Dokumente/2017/Veranstaltungs kalender_2017.pdf and https://issuu.com/ztonline/docs/ mag_0217_web (2017).
- 21. Ski Club of Rothrist. Ski Festival of Rothrist. https: //www.facebook.com/Skifest/posts/pfbid02CUxnzbr9 w7uUfb6h2wbcFrGYadFuVSghQP8dp5eJxZ8CX4D5 L1fmdiAJxL1tTGCTI (2023).
- 22. Ski Club of Rothrist. Exchange of emails between Simone Loreti and Mr. Fabian Senn, President of the Ski Festival, and of the Rothrist Ski Club. https://www.skiclu b-rothrist.ch/ (2023). Email received: 02/03/2023.
- **23.** Variété Pavé Theater Company. Variété Pavé. https: //issuu.com/stadtanzeigerolten/docs/stadtanzeiger_olte n_ausgabe_29, https://www.variete-pave.ch/ (2017).
- 24. Nautical Club of Aarburg. Beach Festival. https://ww w.nca-aarburg.ch/, https://www.facebook.com/eve nts/1327873550615382/?active_tab=about, https: //www.facebook.com/strandfestaarburg/posts/pfbid 032GAo6rkBMbs1My5KoSqNekjsT2rsZdEENWKQtf hjUBFZAJa3jdkEHKeLGEhZzp4TI (2017).
- 25. Nautical Club of Aarburg. Exchange of emails between Simone Loreti and Ms. Natascha Frutig, Actuary of the Nautical Club of Aarburg. https://www.nca-aarburg.ch/ (2023). Ms. Frutig estimated the number of people who attended the Beach Festival of Aarburg in 2017: "On Friday we were able to welcome around 2000 people to our party, [while] on Saturday evening there were an estimated [number of] 5000 people.". Email received: 18/03/2023.
- 26. Municipality of Reiden. Summer Party. https://www.will isauerbote.ch/wiggertal/reiden-langnau-richenthal/201 7-07-30/eine-sommerparty-von-reidern-fuer-reider and https://www.facebook.com/events/reiden-mitte/sommer party-/296622310796245/?paipv=0&eav=AfbxHqSrl5 IIiMI0Ds7Afp8YFEgOtGUTaLyRW7VuqZF6_qHyzf U0HQrhCu_heq9DdOw&_rdr (2017).
- 27. Municipality of Reiden. Exchange of emails between Simone Loreti and Mr. Daniel Loosli, Communal Secretary of the Municipality of Reiden. https://www.reiden.ch/ (2023). Mr. Loosli briefly described the event as follows: "The event took place in the village center [of Reiden, called] "Reiden Mitte". The forecourt of the triple gym "Johanniter", and the same sanitary facilities, served as infrastructure. It is estimated that around 400 600 "locals" attended this event. The festival lasted from about 17:00 to 04:00 [of the following morning]". Email received: 08/03/2023.
- **28.** Fränti, P. & Sieranoja, S. K-means properties on six clustering benchmark datasets (2018).

Bryant, A. & Cios, K. Rnn-dbscan: A density-based clustering algorithm using reverse nearest neighbor density estimates. *IEEE Transactions on Knowl. Data Eng.* 30, 1109–1121, DOI: 10.1109/TKDE.2017.2787640 (2018).