Evaluation of a model for mapping intense pluvial runoff hazard using proxy data of runoff-related impacts. Application to the railway context

Isabelle Braud[1], Lilly-Rose Lagadec[2,1,3], Loïc Moulin[3], Blandine Chazelle[4] , Pascal Breil[1],

[1]Irstea, UR RiverLy, Lyon- Villeurbanne Center, 5 Rue de la Doua, CS 20244, 69625 Villeurbanne cédex, France.
[2]SNCF Réseau (French Railway Company), Engineering and Projects South-West, PIEG General Studies, 54 bis rue Amédée Saint Germain, 33077 Bordeaux, France.
[3]SNCF Réseau (French Railway Company), Engineering & Projects Direction, Railways, Tracks & Environment Department, 6 avenue Francois Mitterrand, 93210 La-Plaine-Saint-Denis, France.
[4]SNCF Réseau (French Railway Company), Engineering and Projects South-East, General Studies, 31 Avenue Albert-et-Elisabeth, 63037 Clermont-Ferrand Cedex, France.

Correspondence to: Isabelle Braud (isabelle.braud@irstea.fr)

Abstract. The IRIP method, for "Indicator of intense pluvial runoff", in French, is a geomatics method which allows mapping the susceptibility of territory to surface runoff and that provides three maps of susceptibility to the generation, transfer and accumulation of runoff. It is based on the combination of binary maps that represent the impact of a given factor (favourable or not favourable) to runoff. These factors are summed up to provide susceptibility maps to runoff with levels ranging from 0 to 5. To be used for risk prevention, the quality and limitations of the produced maps must be assessed. However, runoff data are very scarce and not available everywhere in a territory. Proxy data of impacts related to runoff can provide information useful for the evaluation of the IRIP maps. However, both information cannot be compared directly and a specific methodology to compare susceptibility maps and proxy data must be proposed. This paper presents such a method that accounts for the hazard level, the vulnerability of the study area and possible mitigation actions taken to reduce the risk. The evaluation method is assessed using runoff-related impacts on railways as proxy data to compare with the maps. The evaluation is made on a 80 km railway line in Normandy (North of France), where a comprehensive database of runoff-related impacts on the railway has been gathered over the whole 20th century. The results show that the evaluation method is robust and relevant. The comparison with runoff-related impacts on the railway also shows that the susceptibility maps produced by the IRIP model provide relevant information related to runoff and that they can be used to design risk management strategies, as illustrated in the railway context.

1 Introduction

Runoff occurring outside the river network is a natural hazard that is often quite localized but has a high societal impact. In France, Moncoulon et al. (2014) mention that about half of insurance claims due to flooding occur outside areas mapped as at risk of river flooding in the framework of the European Union Flood Directive. Runoff consequences can be fatalities, damages to buildings or infrastructures, disruption of transport networks. In addition, surface runoff events are often associated with sediment transport and deposition, causing losses to agricultural land and increasing the damages to infrastructures. Linear transport networks, such as railways, are very sensitive to runoff hazards: they cross various small catchments and water and mud can damage the railway track and electric installations (Chazelle et al., 2014; Lagadec et al., 2018). Maurer et al. (2012) estimated that the median cost (direct and indirect) of an hydrometerological event on the European railway network was 2.69 M€.

As runoff can occur everywhere on a territory, there is a need to provide maps of susceptibility to surface runoff at the scale of a whole territory or an entire transport network. Physically based models may be deployed (e.g. Le Bissonnais et al., 2002; Schmocker-Fackel et al., 2007) but they require lots of data for their set up and calibration that may not be available

everywhere. An alternative solution, called IRIP for "Indicator of Intense Pluvial Runoff" was proposed by Dehotin and Breil (2011) for mapping the susceptibility to surface runoff. The model allows the creation of three maps representing three different phases of the surface runoff phenomenon: generation, transfer, and accumulation. It is based on a score method using a set of indicators derived from easily available information (digital terrain model, land use map and soil map). The

5   result is composed of three susceptibility maps with scores ranging from 0 (no susceptibility) to 5 (high susceptibility) for the generation, transfer and accumulation of runoff. The IRIP model and its evaluation is the focus of this paper.

To be used for hazard prevention and risk management, the validity of the produced maps must be assessed and the limits of the methodology clearly defined. However, runoff located outside the river network is a phenomenon that is difficult to observe as it can occur everywhere and over very short durations. There are therefore very few direct observations of runoff,

10   apart from artificial runoff simulation experiments or some rare research experiments (see Dehotin et al., 2015 for more details). On the other hand, indirect information of runoff-related impacts can be more easily available as runoff may have damaging consequences such as flooding of buildings or of transport networks (roads or railways), mud flows, erosion, landslides. Information about these impacts can be collected and reported based on various media: post-event surveys to collect the location of impacts on infrastructures or on transport networks (Versini et al., 2010b; Naulin et al., 2013;

15   Defrance et al., 2014; Lagadec et al., 2016b, 2018), insurance claims on buildings or infrastructures (Moncoulon et al., 2014; Le Bihan et al., 2017), analyses of the press and social media (Llasat et al., 2013; Saint Martin et al., 2018; Petrucci et al., 2019) or citizen science (Gourley et al., 2010; Le Coz et al., 2016). All these data are referred to as "proxy data" in this paper. Such data have been used for the evaluation of quantitative flash flood forecasting models (e.g. Gourley et al., 2010; Defrance et al., 2014; Javelle et al., 2014; Saint-Martin et al., 2016); road cutting warning models (Versini et al., 2010b;

20   Naulin et al., 2013) or flooding impact models on buildings (Le Bihan et al., 2017). The evaluation is based on criteria that are used for the evaluation of meteorological or hydrological forecasts (WWRP/WGNE Joint Working Group on Forecast Verification Research, 2015), i.e. the computation of probability of detection, false alarm ratio and success ratio. Such approach has been extended for the evaluation of non-quantitative prediction model such as the IRIP model by Lagadec et al. (2016b). It was further improved by Lagadec et al. (2018) using a comparison with expert judgement, taking into account the

25   vulnerability of the railway. However, these evaluations remained quite qualitative. It was necessary to generalize the evaluation methodology and to propose a more systematic and quantitative manner to deal with proxy data of localized runoff-related impacts. It was also necessary to use a large and comprehensive data set of runoff-related impacts to assess the relevance and limitations of the IRIP maps, before their use in risk management strategies.

Indeed, although proxy data provide useful information about the occurrence of runoff, these data cannot be compared

30   directly to susceptibility maps of runoff hazards because the information they carry is not the same. Impacts are related to the occurrence of a risk, thus taking into account the vulnerability of the stakes (for instance an infrastructure will be less vulnerable to runoff if there is a protection structure), whereas susceptibility maps are only describing hazards. Susceptibility maps to runoff are continuous in space whereas proxy data are generally point data (impacts on buildings or transport networks). Impacts are generally observed where there are stakes. So the information may not be comprehensive, in

35   particular when runoff occurred without stakes. This comprehensive information would be required to accurately estimate false alarm ratio (Calianno et al., 2013). In addition, proxy data are not always well geolocalized and the description of impacts is subjective and depends on the observer.

The paper focuses on one kind of proxy data that are localized runoff-related impacts, such as impact on transport networks. The first objective of the paper is to propose a methodology to use these proxy data for the evaluation of a non-quantitative

40   method of runoff hazard mapping, such as the IRIP method. This implies identifying which data and information processing are required to perform such an evaluation, and the criteria that can be used for a quantitative comparison. The second objective is to assess the feasibility and relevance of the proposed methodology using a well-documented case study in the railway context. The study takes advantage of the availability of databases of damages and incidents on the French railway

network. The case study is a particularly well-documented 80 km railway between Rouen and Le Havre in northern France, where the IRIP model was set up and where a comprehensive database of runoff-related impacts has been collected for about one century. This provided a comprehensive proxy data set that allowed the assessment of the robustness and applicability of the proposed evaluation methodology. The data set also allowed the assessment of the relevance of the IRIP model for runoff

5   hazard mapping on a wide area.

The paper is organized as follows. The "Materials and Methods" section presents the IRIP model, the proposed evaluation methodology, the case study and how the IRIP model and the evaluation methodologies were set up on the case study. Then the "Results" section presents the results of the IRIP model and of the evaluation on the 80km railway between Rouen and Le Havre. In the "Discussion" section, we discuss the hypotheses behind the evaluation method and the genericity of the

10   proposed methodology. The use of the IRIP model for risk assessment in the railway context is also discussed, before providing the main conclusions of this study.

2. Materials and methods

2.1 The IRIP mapping model

The IRIP model is briefly described here, but more details can be found in the literature (Dehotin and Breil, 2011; Lagadec
15   et al., 2018). The present description is mainly taken from Lagadec et al. (2018) that retained improvements proposed by Lagadec (2017) to the IRIP model. The IRIP model provides three maps representing three processes involved in storm runoff hazard: generation corresponding to areas with low infiltration capacity and where production of runoff at the soil surface is likely; transfer corresponding to areas where water can be transferred downwards, accelerate and produce erosion; accumulation corresponding to areas where water can slow down, concentrate and produce floods and deposits. The IRIP
20   model focuses on runoff occurring outside the river network and is therefore complementary to flooding risk mapping along river networks. The IRIP maps combine indicators derived from geographic information layers (Figure 1, Table 1). Each indicator is classified into two categories: favorable to runoff, where 1 is attributed to the pixel, or not favorable to runoff, where 0 is attributed. This yields 5 binary maps. The maps are then added to create a susceptibility map with 6 levels, from 0 (not susceptible) to 5 (very susceptible). The indicators used for each of the three susceptibility maps are presented in Figure
25   1. The generation map is created thanks to three indicators derived from a soil map, one indicator derived from a land use map, and one indicator derived from the topography. The latter is a combination of the slope and the topographic index (Beven and Kirkby, 1979): 1 if both are favorable, 0 if one is not favorable. The generation map is then considered as an input indicator for the two other maps of susceptibility to transfer and accumulation in order to account for the need of runoff generation, before its possible transfer and/or accumulation. Maps of susceptibility to transfer and accumulation of runoff are
30   created mainly by associating indicators based on topography. But the indicators have opposed conditions for being favorable to runoff. For example, the slope indicator is favorable for transfer in the case of steep slopes, and for accumulation in the case of low slopes. The break of slope indicator is favorable for transfer in the case of convex break of slopes and for accumulation in the case of concave break of slopes. Topographic indicators are computed for each pixel relatively to their upstream sub-catchment in order to follow the hydrological upstream to downstream water transfer. The
35   resolution of the susceptibility maps retains the resolution of the Digital Elevation Model (rasterized topography map) used as input data. To determine the thresholds separating the topographic indicator values into values favorable or not to runoff, a classification method (Rubin, 1967), dividing the observed values in two classes, is used. Thus, the threshold value, separating the two classes (favorable or not to runoff), depends on the study area, but their impact on the final maps is reduces as they only influence four indicators out of 15 in Figure 1. The IRIP model can therefore be applied to various
40   territories without a priori local knowledge.

2.2 The evaluation framework

The proposed evaluation framework is shown in Figure 2. It extends the work of Lagadec et al. (2016b, 2018) but remains based on the use of contingency tables and the computation of a detection rate and a false alarm ratio (see details below) to propose a quantitative comparison between the IRIP maps and the localized runoff-related impacts. The method takes into

5 account the following elements: the different nature of the impacts (localized) and the IRIP maps (continuous score maps), the vulnerability of the stakes for which runoff-related impacts are reported, and the existence of mitigation measures that may reduce the occurrence of risks. The four steps of the method are detailed below.

2.2.1 Step 1: Definition of the evaluation area

The IRIP maps can be computed over a whole territory. The evaluation area, i.e. the area where quantitative measures are

10 computed must be relevant with regards to the available runoff-related impact data. This is particularly important to get a reliable estimate of false alarm ratio. The evaluation area will therefore depend on the runoff-related impact database, as illustrated by the following examples. In case of impacts following a specific localized rainfall event, the evaluation area may be defined as the area experiencing rainfall larger than a specified intensity, as if there is no rain, there is no runoff. In the case of a transport network, the IRIP maps are established for all the catchments that are intercepted by the transport

15 network. If impacts are only recorded on the transport network, the evaluation area will be the transport track itself, with a buffer zone consistent with the resolution of the DTM used to compute the IRIP maps. This buffer accounts for inaccuracy in the DTM and in the geolocalization of the impact data. When a comprehensive database of runoff-related impacts is available over a territory, for a long historical period, it can be assumed that all the territory may have been affected by a rainfall event, and the whole catchment can be considered as the evaluation area for the application of the evaluation method.

2.2.2 Step 2: Characterization of the vulnerability and hazard in the evaluation zone

The IRIP model provides susceptibility maps with score values ranging from 0 (no susceptibility) to 5 (high susceptibility). To compare these scores with runoff-related impact data it is necessary to choose which levels of susceptibility computed by the IRIP model are considered at risk. As the IRIP model provides three maps, it also means choosing the maps that will be considered in the evaluation. Previous experience (Lagadec et al., 2016b; 2018) showed that, when compared to localized

25 impact data, susceptibility maps to transfer and accumulation were relevant, with the susceptibility map to transfer generally associated with erosion, and the susceptibility map to accumulation associated to sediment deposition and flooding. Therefore a composite of the two maps was built for defining the hazard.

In addition, the notion of risk depends on the exposure and vulnerability of the stakes that are considered. Several methods can be used to assess the vulnerability of stakes. Saint-Martin et al. (2016) used the AHP (Analytic Hierarchy Process, Saaty,

30 1990) method to rank several stakes. Another possibility is to use a vulnerability tree based on expert judgement. For buildings, a vulnerability value can be affected to each building and the hazard level computed in a buffer zone around the building. In the case of a transport network, it is necessary to divide the network into sections that are meaningful with regards to runoff risk, and to assign a vulnerability value to each section. The hazard level must also be computed in each section. As a buffer zone around an impact or a transport network section contains several IRIP pixels with different values

35 of susceptibility, a rule must be chosen to assign a susceptibility value to the buffer area or the section. For instance, the hazard level can be computed as the maximum value of the susceptibility inside the buffer area or the transport network section.

Once established, the vulnerability scale is converted to a limited number of vulnerability classes. The latter are then combined with the hazard levels to define which combinations are at risk. The notion of area at risk is defined according to

40 the principle that "the higher the vulnerability, the lower the hazard level triggering a risk". Figure 3 provides examples of

such choices, the one in Figure 3a corresponding to a case where the vulnerability of the stakes is not taken into account and where IRIP levels of hazard 4 and 5 are chosen at risk everywhere.

2.2.3 Step 3: Quantitative evaluation of the maps

5   The third step is the comparison between the areas identified at risk in the previous step and runoff-related impact data. For this purpose, a contingency table (Table 1) is built for a sample containing all the buffer areas or sections for which a risk level has been assigned in step 2. If an impact has been observed in an area considered at risk, the impact is counted as 'hit'. If no impact has been observed in an area not considered at risk, the impact is counted as 'correct negative'. If an impact has been observed and the area not declared at risk, the impact is counted as 'miss'. Finally, if no impact has been observed but the area declared at risk, the impact is counted as 'false alarm'. Based on the contingency table, three quantitative measures

10   of performance are computed (Table 2): the probability of detection (POD) which represents the fraction of impacts that have been correctly identified in an area at risk. The false alarm ratio (FAR) indicates the proportion of areas at risk with false alarms. If the method was perfect, the POD would be equal to one and the FAR to zero. The χ^2 test is used to define if the dependency between risk levels and the occurrence of impacts is significant. For that, the χ^2 is compared to that of the theoretical distribution with full independency of risk and impacts. For a contingency table with one degree of freedom (as in

15   our case), the probability to get a $\chi2$ larger than 10.83 is lower than 0.1%. This means that if the χ^2 is larger than 10.83, the relationship between risk levels derived from the IRIP maps and impacts is highly significant.

2.2.4 Step 4: Taking into account risk mitigation measures

This fourth step is necessary to properly take into account the fact that, if an area is at risk, the stakeholder may have taken mitigation measures that may explain the absence of observed impact. Such mitigation measures are therefore likely to

20   explain a certain amount of false alarms. For instance, mitigation structures can be protection to buildings, retention basins, hydraulic works crossing below transport infrastructures, etc… They can also be resilience actions like a reinforced supervision in case of high rainfall amount warning. If an area classified at risk has a specific supervision measure or mitigation structures have been built, it is moved from 'false alarm' to 'hit' as the implementation of mitigation measures means that the area was indeed at risk, but that no impact was recorded due to the efficiency of the mitigation measure. The

25   performance measures are then recomputed, based on this modified contingency table.

After these four steps, the final values of the quantitative performance measures are obtained, characterizing the performance of the IRIP mapping model.

2.3 Case study

2.3.1 Presentation of the study area

30   The case study is the 80 km railway line between Rouen and Le Havre (Figure 4). This railway has been operating since 1847. It is a strong stake for the region, as it connects Paris to Le Havre within about 2 hours, and connects Paris to the major fluvial and sea ports of Rouen and Le Havre. It is located in Pays de Caux, an area known for being affected by intense surface runoff (e.g. Cerdan et al., 2002; Martin et al., 2010). The land use is mainly agricultural land. Soils, composed of silts and clays are sensitive to slacking (formation of a crust lowering significantly the infiltration capacity, Cerdan et al., 2002).

35   The catchment intercepted by the railway has a total area of about 500 km^2. Only two streams cross the railway but the dry talweg network that can be activated during a rainfall event is very dense (Figure 4).

2.3.2 Application of the IRIP model

Three input maps were used to produce the IRIP maps. The retained GISs layers are easily available and allow testing the IRIP model with standard data. The topography was described using the IGN BD ALTI© Digital Terrain Model (DTM), with a 25 m resolution. Land use was described using a 1/2500 land use map of Haute-Normandie region from 2009.

5 Pedology was taken from the European Soil Database (ESDB) V2.0 with 500 m resolution. The IRIP model parameterization used in this study is presented in Table 3. Given the low local knowledge of the study area, the thresholds defining the classes favorable or not favorable to runoff were computed using the classification method proposed by default in the IRIP model, contrarily to the application by Lagadec et al. (2018) that used values derived from local expertise. Thresholds for topographic index and slope indicators were therefore defined using the classification method. The threshold

10 of the drained area indicator was fixed to 2.5 ha following sensitivity tests performed by Lagadec (2017) to identify, in this specific catchment, a minimum surface from which significant surface runoff can be generated.

2.3.3 The database of impacts on the railway

For its internal needs in terms of risks management, SNCF, the French railway company, has set up a quite systematic archiving system of all the incidents and disruptions of the train traffic, as well as of all the works carried out on the railway

15 tracks. This information is archived either in digital databases since the 1990s, either in paper format. Paper archives are located in several places in France according to the date the documents were produced. For the present study, the objective was to gather, on the Rouen- Le Havre railway, all the registered impacts related to runoff and the associated information, since the creation of the line until today.

The database was created in two steps: a data collection step and a data processing step. The collection of impact data was

20 carried out with the support of archival expertise. Four archive sites were visited, depending on the age of the documents. It was necessary to define a limited number of information that had to be collected and that were relevant with respect to runoff. They are provided in Table 4 that describes the two tables that were filled by the archivist when consulting the archives and that were relevant for the next phase of data processing. The first table describes the source documents of interest and the second table describes the runoff-related events. Archives are organized according to railway kilometric

25 points (KP) and earthworks. The description of the location of runoff-related impacts makes reference to the KP and earthworks, so this information was retained in the event description (Table 4). Indeed, earthworks are relevant elements to divide the railway tracks into meaningful sections, with regards to its hydraulic functioning. Earthworks are designed to insert the railway track within its environment, while respecting technical constraints such as a maximum allowed slope to ensure electric traction and braking in good conditions. Earthworks modify the natural surface topography and therefore

30 water flow paths. Four types of earthworks can be distinguished: embankment to cross talwegs or valleys; excavations to follow longitudinally valleys or to cross small ridges; mixed profile to cross hillslopes and quasi-flat profiles (see Figure 5). Note that a digital GIS layer describing the railway tracks and the location of the various earthworks and their characteristics was created for this study. The challenge in data collection has been to manage the diversity of formats (paper, digital) and to manage duplicates. The document collection work took four months. 506 documents were retained and inventoried, dating

35 from 1903 to 2017.

The data processing step consisted in retracing the history of each impacted area, the circumstances of the incidents, the work undertaken and this, up to the current situation. One difficulty was to manage the uncertainties, particularly for the location of impacts and to determine whether they were direct consequences of an intense runoff event. Finally, the database consists of 62 sections impacted at least once, ranging from point zones to a 1.3 km long section. All the sections impacted at

40 least once represent a cumulative length of 12 km, over the 80 km of the railway, or 15% of the length studied. A geographic information layer of georeferenced impacts with their date, type and uncertainty has finally been created, allowing its overlay with the IRIP maps.

There are two assumptions behind the use of this database for the evaluation of the IRIP maps. First: the duration of the period over which impacts are recorded (about one century) is long enough so that each section may have experienced a possible damaging event. As a consequence, the database can be assumed to be comprehensive. Second, it is assumed that land use types have not changed drastically in this area dominated by agricultural land.

2.4 Application of the evaluation methodology to the case study

Step 1: The database of runoff-related impacts covers more than one century. We can thus assume that all the catchments intercepted by the railway may have experienced a runoff event. We could therefore consider all the catchments intercepted by the railway as the evaluation area. However, recorded impacts are only located on the railway track. Thus, the evaluation area must be restricted to the railway track itself. However, to account for uncertainty on the location of impacts and of the DTM inaccuracy, a 25 m buffer area was considered, on both sides along the railway line. Other values of the buffer were tested in Lagadec (2017), but a 25 m buffer was considered as the most relevant value, given the 25 m resolution of the DTM.

Step 2: The railway track vulnerability was defined using a decision tree where scores were assigned to each branch of the tree (Figure 5). The decision tree was built from expert judgement and verified using the impact data (Lagadec, 2017). Four criteria were considered to build the decision tree for the 182 sections of earthworks that correspond to the railway profile composed of four types of earthworks: embankment, excavation, mixed profile, quasi-flat profile. Furthermore, this segmentation of the railway is consistent with impact recording that are assigned to an earthwork type (see section 2.3.3). The four criteria considered to compute the vulnerability scores are:

- (1) The exposure: unexposed areas are sections of long tunnels or large viaducts. The other sections are considered as exposed and get a vulnerability score of 1.
- (2) The type of profile and (3) its length: The types of profiles considered to be the most vulnerable are long excavations and short embankments. Long excavations are prone to flooding, they have to handle more water from the slopes they intersect and are more likely to experience malfunction of the drainage structures. Short embankments are suitable for loading by runoff. They can play a role of hydraulic barrier. The median earthwork length of the sample was used to separate short and long earthworks.
- (4) The presence of a singularity: The singularities are either level crossings, road bridges or tunnel inlets or outlets These are areas likely to experience arrival of water on the platform. These singularities generally constitute discontinuities in the topography of the work. If one of these singularities is present in the envelope of an earthwork, its vulnerability score is increased by 1.

For defining the hazard, two IRIP maps were considered: the susceptibility maps to transfer and to accumulation that were combined to a unique map. They can be related to erosion and flooding respectively (Lagadec et al., 2016b). A value of hazard level was assigned to each 182 section of the railway track where a vulnerability score was also assigned. This value corresponds to the maximum value of susceptibility to transfer or to accumulation in the 25 m buffer zone on both sides of the railway section. Vulnerability and hazard were combined following Figure 3c, where red boxes are considered at risk. Performance indicators were also computed for combination 3c (where vulnerability of the railway track is not taken into account and levels 4 and 5 are considered at risk) in order to illustrate the impact of taking vulnerability into account in the evaluation method.

Step 3: Performance criteria of the IRIP model were computed using the measures defined in section 2.2.3.

Step 4: Mitigation measures were considered in a second step to try to explain the false alarms. Structural and non-structural mitigation measures were considered and inventoried along the whole railway. Structural measures includes all the hydraulic structures (drainage structures along or below the railway track, retention pond, etc…) that were built to help water flow circulation. At SNCF, non-structural measures include surveillance patrols in case of bad weather. These patrols target as a

priority the sections registered in what is called the 'bad weather tours'. The latter are defined using local knowledge about the hazard exposure or about the specific infrastructure vulnerabilities. They provide increased and targeted monitoring in case of bad weather and early response if needed. If an hydraulic structure was present in one section, or if the section was registered in the 'bad weather tours', it was considered as evidence of hazard exposure of the section. This allowed the

5   transformation of the corresponding 'false alarm' values to 'hit'. Then, the performance indicators were recomputed, providing the final evaluation criteria of the IRIP model.

3. Results

3.1 The IRIP maps

Figure 6 presents the IRIP susceptibility maps for the generation, transfer and accumulation of runoff. The map of

10   susceptibility to runoff generation shows a high sensitivity to the genesis of runoff on agricultural plateaus, an even larger one in urbanized areas and a lower one on slopes that are more vegetated. The IRIP map of susceptibility to runoff transfer shows strong potential of erosion to produce mud and mass flows along the uphill slopes of the main talwegs. Other small talwegs with high susceptibility level to runoff transfer are also scattered throughout the area, showing areas potentially sensitive to erosion. The map of susceptibility to runoff accumulation shows all the preferential flow paths, from the

15   perennial streams to dry talwegs, the latter being located in headwater catchments with flatter areas.

3.2 Results of the evaluation method

Using, Figure 3c to combine hazard and vulnerability, Figure 7 provides illustrations of the application of the evaluation method and of the building of the contingency tables.

On the left figure (Area A), for the two 'H' (see caption in Fig.7) sections, the vulnerability scores are greater than or equal

20   to 2 and there are pixels with IRIP levels greater than or equal to 4, these sections have already been impacted at least once, they are therefore 'hits'. The 'CN' section has a vulnerability of 2, its maximum IRIP level is 3, there has been no impact, so the section is a 'correct negative'. The section 'FA' has a maximum vulnerability score of 4, so a level IRIP 3 is sufficient to consider this section at risk, nevertheless no impact was observed here. The section is then counted as 'false alarm'.

The right figure (Area B) illustrates how mitigation measures are taken into account in the analysis. The 'H' section has a

25   vulnerability score of 2 and a maximum IRIP level of 5, so it is considered at risk, yet no impact has been reported. This section should be assigned to the 'false alarm' class, but we note that it is equipped with a crossing work under the railway which could play a role of protection against the hazard. The location of this infrastructure shows that the hazard is indeed present at this location and that the IRIP map is correct, so the section is finally rated 'hit', once mitigation measures are taken into account. The 'M' section has a vulnerability of 1 which requires a IRIP level of 5 to be considered at risk. But, the

30   maximum IRIP level is 4, so the section is not considered at risk. However, as an impact has occurred, the section is considered as 'miss', although a mitigation structure is present.

The results of the evaluation along the whole railway lines are reported in Table 5. Three results are presented: first column - when not taking into account neither the vulnerability neither the mitigation measures (meaning that hazard and vulnerability are combined following figure 3a); second column - when only the vulnerability is taken into account following Figure 3c);

35   and third column - when the vulnerability and the mitigation measures are taken into account. The results show that the POD increases when more information (vulnerability, mitigation measures) is taken into account in computing it: from 86 %, to 93%, and to 96%. In the same way, the FAR decreases when taking into account more information: from 62%, to 58%, and to 28 %. The results in terms of χ^2 present the same trend, with the significance increasing with the number of information considered. It can be noticed that from the first to second column, 4 'missed' impacts change into 'hit'. These are impacts

40   that occurred on highly vulnerable railway sections which can experience damages with a very low hazard exposure level.

Likewise, 6 'false alarms' change into 'correct negatives', because the railway sections were not enough vulnerable. From the second column to the third, 'correct negatives' and 'missed' do not change. Only 'false alarms' that are related to mitigation measures change into 'hit'. It accounts for 40 railway sections. The values of the χ^2 shows that the relation between risk and impacts is significant in the first column and highly significant in columns 2 and 3. The results in the last

5   column present very encouraging values, highlighting the added value of the IRIP maps, and of the vulnerability and mitigation measures characterization, for the evaluation of the IRIP model.

4. Discussion

4.1 Relevance and limitations of the evaluation method

The results of the evaluation method, as applied to the IRIP maps, show the interest of considering both the vulnerability of

10   the railway and mitigation measures taken to lower risks. These two factors are essential to get an accurate comparison between runoff-related impacts data and the IRIP maps. When both are not considered, POD values are high, but the rate of false alarms is also very high. Note that the term false alarm, does not mean that the information is false but that it cannot be proven. Even if no past impact occurred, an impact could possibly happen in the future.

When both vulnerability and mitigation measures are taken into account, probability of detection remains high, but false

15   alarm ratio is much lower, reaching performance that are satisfactory. When evaluating the performance of a road cutting warning system, Versini et al. (2010a, b) and Naulin et al. (2013) also showed that it was essential to incorporate road vulnerability in their computation of POD and FAR to get meaningful results. However, the results of the performance criteria depend on the choice made to assign hazard to the section (here we chose to use the maximum hazard (as discussed in section 2.2.2) value within the section) and on the choice made to combine hazard and vulnerability (i.e Figure 3c).

20   Lagadec (2017) compared combinations shown in Figures 3b and 3c but found similar performance criteria. The difference between performance criteria when vulnerability is not taken into account (column 1 in Table 5) or is taken into account (columns 2 and 3 in Table 5) shows that considering the vulnerability of the railway has a much higher impact on the performance criteria values than the way hazard and vulnerability are combined.

Lagadec et al. (2018) tested another way to assign the hazard value to one section by declaring that a section was at risk if

25   the percentage of the section area with values larger than a susceptibility threshold (4 in Lagadec et al., 2018) was higher than a percentage threshold (10% in Lagadec et al., 2018). In this case, the user must choose two thresholds, the values of which will strongly impact the evaluation measures values. When assigning the maximum susceptibility value as hazared level to one section, as used in this study, these subjective choices are avoided. However, when runoff-related impact data are available, such adjustment of the method to assign IRIP hazard may be useful for operational purpose, in order to define

30   which areas are at risks in a given territory and to prioritize areas requiring protection measures.

There are however limitations of the vulnerability/hazard combination, as illustrated in Figure 7 (right, Zone B). The section 'M' is rated as very little vulnerable with a score of 1. So a level 5 is required in the IRIP maps to consider this zone at risk. This is not the case as the maximum level of the IRIP maps is 4. However, this area has already been impacted by intense runoff as impacts were recorded, and it is also equipped with a hydraulic structure crossing under the track. This section is

35   therefore at risk but classified as missed impact. We can see that the IRIP map shows a specific arrival of runoff, so the map looks correct. But the section vulnerability score is only one, leading to consider it not at risk. Therefore, the vulnerability classification is obviously deficient in this example and should be modified to better take into account the specificities of this type of configuration (quasi-flat profile). Further discussion with railway experts could make increasing the vulnerability score of quasi-flat profiles.

40   Another limitation of the study is related to the runoff-related impact database itself. As mentioned before, the location of impacts is sometimes not very accurate and may alter the computation of the performance measures. Furthermore, although

it covers more than one century of data, the database may not be comprehensive, which could affect the false alarm ratio if all the occurred impacts have not been recorded. Another question is: had each portion of the railway the chance to experience a runoff event? To answer the question, we can compute the probability of not having experienced an event of a given return period over one century. This probability is lower than 0.001% for a 10-year return period event, and lower than

5   1% and 13% for a 20-year and 50-year return period respectively. Therefore, it can be assumed that each section of the railway had a chance to experience a rare event at least once. This means that, if the database is comprehensive enough, the performance measures are not biased. Another point that must be considered is the assumption of a constant land use map for the IRIP map building. It is clear that land use has changed over a whole century, with the development of intensive agriculture and urbanization. Indeed large field crops have replaced the mosaic of small fields crops since the second world

10   war. The IRIP model considers that urban and crop lands are both favorable to intense runoff generation. The largest cities of Rouen and Le Havre being located at the start and end of the railway line, urban growth has no major effect in this context. Loss of grassland and forest is more sensitive. As the IRIP maps were established with the 2009 land use, change in land use over the last century would lead to a possible overestimation of false alarms, as current land use is more prone to runoff than it used to be in the past. Land use change could also explain the increasing number of impacts in the recent decade. But this

15   increase could also be explained by a more comprehensive recording of impact statements in SNCF practices during this period.

4.2 Impact of the uncertainty and resolution of the IRIP maps.

Regarding the influence of the resolution of the input maps on the final maps, several resolutions and qualities of DTMs were compared in Lagadec (2017), with five DTMs ranging from 250 m to 5 m on only 30 km of railway for which an

20   accurate Lidar DTM at 5 m resolution was available. It showed that there is a spatial persistence of information from higher resolutions to coarser resolutions. The analysis also showed that the data acquired by Lidar provide very relevant information that helps to understand the phenomenon of runoff. This information is relevant as it provides detailed explanatory elements about runoff pathways. Although it would be recommended to have similar resolutions for the three input maps, accuracy of the DTM is essential for the IRIP model application, as DTM is used to compute 3 factors maps in the transfer susceptibility

25   map and 4 factors in the accumulation susceptibility map. On the other hand, 3 factors over 5 use the soil map in the building of the susceptibility to runoff generation map. The quality of this data is therefore essential for the interpretation of the susceptibility to runoff generation map. Efforts to spend in collecting accurate input data depends on the use of the final maps. Input data resolution also depends on the size of the study area and must be chosen to facilitate map reading and to optimize computing resources. For a large study area, it is recommended to zoom in through successive applications of the

30   IRIP model: identify the most exposed areas with coarser resolutions first and then zoom in with higher resolutions.
When considering coarser resolutions, it becomes difficult to apply the evaluation method proposed in this paper (mainly 75 and 250 m resolution DTMs) as the evaluation zone must be enlarged to account for the larger pixel resolution. The size of the railway sections become small as compared to the pixel resolution, so it becomes more difficult to overlay point impacts and IRIP maps pixels. In the same way, high resolution maps imply to adjust some choices made for the evaluation process,

35   such as the size of the buffer area both sides of the railway. The way hazard level is assigned to a section should also be reconsidered as the chance to get one pixel with a high hazard level is larger if the resolution is higher and it may not be relevant anymore to mark the whole section at risk with only one pixel with a high hazard level. Computing a percentage of the section with high hazard level may be more relevant in this case. For these reasons, quantitative evaluation has not been tested yet with high resolution maps. Only qualitative analyses are provided in Lagadec (2017).

40

The data set used in this study is very comprehensive and includes the three pieces of information required for the application of the evaluation methodology described in section 2. Lagadec (2017) used the quantitative measures to compare

and improve the IRIP model by testing alternative indicators for the building of IRIP maps. This led to recommendations to improve the method as proposed by Lagadec et al. (2018) and used in the present study. One example is the 2.5 ha threshold value of the drained area chosen to separate the conditions favorable and not favorable to runoff transfer and accumulation (see Figure 1 and Table 3). To choose this value, Lagadec (2017) performed six simulations with a drained area threshold

5   ranging from 0.5 to 100 ha. A threshold value between 1 and 5 ha was a good compromise between performance in explaining impacts and the visual aspect of the maps. Thus, the 2.5 ha threshold was retained in the analysis.

Up to now, all the five factors involved in the IRIP maps are given the same weights. The evaluation methodology could be used to compare non-equal weights in the building of the maps. Methods such as the one proposed by Neuhäuser et al.

10   (2011) could be used for this purpose.

4.3 Genericity of the evaluation method

The evaluation method presented in this paper was applied using proxy data of runoff-related impacts on the railway. In this case, the evaluation area was defined as a buffer zone along the railway to account for inaccuracy in the impact location and DTM. Apart from the compilation of a database of impacts on the railway, the combination of hazard and vulnerability

15   required a complete analysis of the vulnerability of the railway and of its characteristics, and an inventory of all the hydraulic structures set up to limit impacts related to runoff. This was a huge effort, as all the corresponding information was not digitized yet, but such effort was very valuable and can be used for other studies. As more and more companies or administrations are setting up databases of the infrastructures they are controlling, such database are becoming more and more common. We have seen that information about vulnerability and mitigation measures was necessary to decrease the

20   false alarm ratio, but had little impact on the probability of detection. Therefore, without information about vulnerability, but compilation of impact data, it is possible to compute reliable estimates of the probability of detection (POD), but not of false alarm ratio (FAR).

In this paper, the evaluation methodology was applied to the railway context, with proxy data related to runoff-related impacts on the railway. This led to a very specific definition of the evaluation area, that was restricted to a buffer zone on

25   both sides of the railway. The approach can easily be extended to road networks or any other linear stakes.

Other applications of the evaluation method are in progress, that show that the method can be applied to other types of localized impact data. Two contexts can be distinguished:

- *The availability of a long term database of impacts over a territory*, like impacts on the protected forested domain managed by the French ONF agency (National Forest Organization) (see also Defrance et al., 2014). This database

30   contains information about damages and protection infrastructure against landslides, gullying and flooding in the Alpes and the Pyrenees. It has been launched in the 1980s but also contains information about historical events. Given the duration of data collection, assumption that the whole surveyed territory may have been impacted can be made and the evaluation area can be defined as the whole forested area covered by the survey. Using these data, the probability of detection can be computed with a good degree of confidence, provided impacts localization is

35   accurate enough, and especially if information about protection infrastructures can be incorporated in the analysis. It is much difficult to get information about vulnerability of the territory, and this information should be defined with local stakeholders.

- *The availability of impact data for a given hydrometeorological event*. In this case, information about rainfall is necessary to define the evaluation area, as no impact will be observed if no rainfall or only low intensity rainfall was recorded. The evaluation area can be defined using rainfall data and a given rainfall intensity threshold, like a

40   ten-year return period rainfall over a 1 h duration, that is used for natural disaster declaration in France. Given that the event are often very localized, the use of radar rainfall data with a short time step (e.g. 5 min time step) is

recommended, as shown by Marra et al. (2016) for landslides. If data collection about impacts is comprehensive and information about protection infrastructure is available, POD and FAR values can be computed with a good accuracy.

5   The evaluation methodology has been designed for localized impact data. However, the avenue of very high resolution remote sensing information, opens perspectives for the evaluation of a mapping method, like the IRIP model. Indeed, those techniques are now able to provide accurate information about erosion and gullying (e.g. Desprats et al., 2013; Eckert et al., 2017), that could be used as verification data for the IRIP maps. Such analysis is in progress using data from satellite Pleiades 1A and 1B which provide 0.7 m image resolution, for the 15-16 October 2018 in the Aude department in France, with 15 fatalities and damages of about 220 M€. Even if river flooding is responsible of a large part of damages, runoff

10   outside the river network has also been observed, that particularly affected agricultural land. A database of insurance claims related to damages in agricultural land is being collected. It will allow the assessment of the added value of high resolution satellite images in the evaluation of the IRIP maps as compared to the use of localized impacts. The evaluation method proposed in this paper may however need some adaptation to be used with remote sensing images.

4.4 Relevance of the IRIP maps for risk assessment and design of mitigation measures in the railway context

15   The evaluation method presented in this paper, and the results obtained with a comprehensive database of runoff-related impacts raise confidence on the relevance of the IRIP maps and of its potential for a use in risk management studies. This confirms the potential of the IRIP maps in various risk management tasks in the railway context that were highlighted in previous studies. Lagadec et al. (2016a) showed qualitatively the usefulness of IRIP maps combined with high resolution radar images of rainfall, for post event survey after an intense rainfall event that damaged the St-Germain-des-Fossés -

20   Nîmes line in the Gard in 2014. Lagadec et al. (2018) showed that IRIP maps were consistent with an expert hydraulic diagnostic of the Bréauté - Fécamp line in Normandy, in prioritizing railway track rehabilitation works. There are several advantages in using IRIP maps: time can be saved and accuracy on the results increased by using these maps as a new source of information to inform field expertise. The maps help to prepare the expertise, to better understand the context and once on site, they allow to focus on specific areas and to move to certain points in the catchment area (areas of runoff generation,

25   erosion, deposition or stagnation). For studies on larger linear areas, the method automatically identifies all exposed areas. By crossing with the infrastructure configuration, a pre-diagnosis of the areas at risk is obtained. IRIP is therefore a relevant tool to help identify runoff hazard.

    For the risk management and technical solution definition phases, runoff maps can also provide useful information. The maps represent the hydrological surface processes over the entire watershed around the railway. This can help in the

30   implementation of actions on hillslopes, in choosing the location of solutions, and in adapting them to hydrological processes according to the zones (erosion, deposits, stagnation). Working outside the railway right-of-way is still difficult today. On the one hand, from a legal point of view, it is necessary to obtain the agreement of the plot owners and to establish contracts for maintenance operations. In addition, such nonstandard technical solutions are often not referenced in quality control procedures, as the latter rather propose dimensioning (in flow and volume) of complete networks bringing water to an outlet.

35   On the other hand, the implementation of alternative techniques favors locally adapted solution, that can be validated through risk analysis. Nevertheless, acting on the slope is sometimes the only sustainable solution to manage runoff. Sediment inputs are difficult to manage by conventional hydraulic structures and the lack of space in the railway right-of-way makes it difficult to implement appropriate solutions. As the constraints of discharges into the environment are very restrictive, managing runoff may require the creation of retention basins, a solution that is often incompatible with the

40   available space. It would be interesting to reduce inflows, for example by creating retention pools in accumulation areas, rehabilitating or creating wetlands (Fressignac et al, 2016), by setting up fascines on the transfer axes to trap sediments and avoid soil losses, by developing grassy stripes on the deposit areas to allow fines and sludge to spread, by avoiding bare land

or by favoring vegetation to increase infiltration capacity on areas sensitive to runoff generation. These soft hydraulic techniques are to be used in addition to the traditional hydraulic techniques used to manage exceptional events. In the long term, the actions in the hillslope limit the degradation of railway infrastructure elements, increase safety and reduce economic losses. The interest is also ecological by creating wet or wooded areas with an improved social perception of

5   railways in the landscape. The runoff problems encountered at a point of the railway infrastructure generally also have an impact upstream of the infrastructure. Whether for urban areas (flood risk) or agricultural areas (erosion risk), runoff also needs to be controlled upstream. It would seem more relevant and technically more efficient to manage runoff in a distributed manner throughout the watershed. There are many obstacles to such control: complex legislation, difficulty of communication, differences in deadlines and budget according to the actors. Having a visual and educational tool,

10   illustrating the downstream impact of an upstream action, and the interconnection of issues on the territory, can promote such a synergy. IRIP map can be a tool to assist in such eco-design.

5. Conclusions

The paper presents an evaluation method, suitable for assessing the relevance of susceptibility maps to intense runoff, using proxy data of localized runoff-related impacts. The methodology was validated using a comprehensive database of runoff-

15   related impacts on an 80 km railway in northern France, covering more than one century of functioning, and applied to the maps produced using the IRIP "Indicator of Intense Pluvial Runoff" maps. The evaluation method requires not only knowledge of the hazard, but also knowledge of the vulnerability of the study area with regards to the considered hazard, and of mitigation measures taken to lower the risk. This information can be time consuming to collect, but is essential to get meaningful performance measures characterizing the accuracy of the map and to get a good appraisal of the risk. It would be

20   interesting to gather the same type of information in other climatic, pedologic and land use contexts. The methodology proposed in the paper is generic enough and can be extended to other sources of localized impact data and to other mapping method of susceptibility to runoff. In order to capitalize on runoff-related impact data acquisition, a perspective is to build a platform where stakeholders could provide their runoff-related impact data and benefit from an on-line QGIS plugin implementing the IRIP model. This could contribute to increase runoff knowledge and understanding and to improve runoff

25   risk management.

Acknowledgements

We thank Judicaël Dehotin, a key proponent of the development of the IRIP method. Corentin Descourts helped in mapping the Rouen – Le Havre railway profile and its vulnerability during its internship. Sixtine de Bejarry performed the data collection of impacts/damages on the Rouen-Le Havre railway. We thank ANRT (Agence Nationale de la Recherche et de la

30   Technologie) for its partial funding of this work. Funding from MTES (French Ministry of ecological and solidarity transition) also contributed to part of this study.

Data availability

The impact/damage data used in this study are confidential and property of SNCF Réseaux and cannot be made available publicly. The study was performed using the iRIP© software, property of SNCF that cannot be made available either. A new

35   version of the IRIP model is however under development as an open-source QGIS plugin.

Competing interest: The authors declare that they have no conflict of interest.

Authors contributions. Most of the work presented in this study was conducted by L.R. Lagadec during her PhD thesis under the supervision of I. Braud, L. Moulin, P. Breil and B. Chazelle. I. Braud wrote a first English version of the manuscript that was revised and complemented by the other authors.

References

5 Beven, K. J., and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology / Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Bulletin, 24, 43-69, 10.1080/02626667909491834, 1979.

Calianno, M., Ruin, I., and Gourley, J. J.: Supplementing flash flood reports with impact classifications, Journal of Hydrology, 477, 1-16, 10.1016/j.jhydrol.2012.09.036, 2013.

10 Cerdan, O., Le Bissonnais, Y., Souchère, V., King, C., Antoni, V., Surdyk, N., Dubus, I., Arrouays, D., Desprats, J.-F. : Guide méthodologique pour un zonage départemental de l'érosion des sols Rapport no. 3: Synthèse et recommandations générales (No. BRGM-RP-55104-FR). BRGM – INRA, 2006. In French, available at http://infoterre.brgm.fr/rapports/RP-55104-FR.pdf

Cerdan, O., Souchère, V., Lecomte, V., Couturier, A., and Le Bissonnais, Y.: Incorporating soil surface crusting processes in

15 an expert-based runoff model: Sealing and Transfer by Runoff and Erosion related to Agricultural Management, CATENA, 46, 189-205, https://doi.org/10.1016/S0341-8162(01)00166-7, 2002.

Chazelle, B., Lambert, L., and Capoccioni, C. P.: Railway vulnerability in case of extremes floods. Knowledge and risk management, Houille Blanche-Revue Internationale De L Eau, 48-54, 10.1051/1hb/2014016, 2014.

Defrance, D., Javelle, P., Organde, D., Ecrepont, S., Andréassian, V., and Arnaud, P.: Using damage reports to assess

20 different versions of a hydrological early warning system, Hydrol. Earth Syst. Sci. Discuss., 2014, 4365-4401, 10.5194/hessd-11-4365-2014, 2014.

Dehotin, J., and Breil, P., 2011: Technical report of the IRIP project: mapping the flooding by runoff (Technical report). IRSTEA Hydrology-Hydraulic Research Unit.

Dehotin, J., Breil, P., Braud, I., de Lavenne, A., Lagouy, M., and Sarrazin, B.: Detecting surface runoff location in a small

25 catchment using distributed and simple observation method, Journal of Hydrology, 525, 113-129, http://dx.doi.org/10.1016/j.jhydrol.2015.02.051, 2015.

Desprats, J.-F., Raclot, D., Rousseau, M., Cerdan, O., Garcin, M., Le Bissonnais, Y., Ben Slimane, A., Fouché, J., and Monfort-Climent, D.: Mapping linear erosion features using high and very high resolution satellite imagery, Land Degradation & Development, 24, 22-32, 10.1002/ldr.1094, 2013.

30 Eckert, S., Ghebremicael, S. T., Hurni, H., and Kohler, T.: Identification and classification of structural soil conservation measures based on very high resolution stereo satellite data, Journal of Environmental Management, 193, 592-606, 10.1016/j.jenvman.2017.02.061, 2017.

Fressignac, C., Breil, P., Matillon, Y., Nullans, A., Chazelle, B., Sarrazin, B. and Vallod, D. : Assurer la maitrise du ruissellement grâce aux zones humides au voisinage des infrastructures de transport dans une perspective de conservation de

35 la biodiversité, VertigO - la revue électronique en sciences de l'environnement [En ligne], Hors-série 24 | juin 2016, mis en ligne le 10 juin 2016, consulté le 27 mai 2019. URL : http://journals.openedition.org/vertigo/17406 ; DOI : 10.4000/vertigo.17406

Gourley, J. J., Erlingis, J. M., Smith, T. M., Ortega, K. L., and Hong, Y.: Remote collection and analysis of witness reports on flash floods, Journal of Hydrology, 394, 53-62, http://dx.doi.org/10.1016/j.jhydrol.2010.05.042, 2010.

Javelle, P., Demargne, J., Defrance, D., Pansu, J., and Arnaud, P.: Evaluating flash-flood warnings at ungauged locations using post-event surveys: a case study with the AIGA warning system, Hydrological Sciences Journal, 59, 1390-1402, 10.1080/02626667.2014.923970, 2014.

Lagadec L.-R. : Développement et évaluation de la méthode IRIP de cartographie du ruissellement intense pluvial.
5 Application au contexte ferroviaire. Doctoral School Earth, Universe, Environment. Communauté Université Grenoble Alpes, 336 pp, https://tel.archives-ouvertes.fr/tel-01778052, 2017.

Lagadec, L. R., Breil, P., Chazelle, B., Braud, I., and Moulin, L.: Use of post-event surveys of impacts on railways for the evaluation of the IRIP method for surface runoff mapping, in: 3rd European Conference on Flood Risk Management, edited by: Lang, M., Klijn, F., and Samuels, P., E3S Web of Conferences, 2016a.

10 Lagadec, L.-R., Patrice, P., Braud, I., Chazelle, B., Moulin, L., Dehotin, J., Hauchard, E., and Breil, P.: Description and evaluation of a surface runoff susceptibility mapping method, Journal of Hydrology, 541, Part A, 495-509, http://dx.doi.org/10.1016/j.jhydrol.2016.05.049, 2016b.

Lagadec, L.-R., Moulin, L., Braud, I., Chazelle, B., and Breil, P.: A surface runoff mapping method for optimizing risk assessment on railways, Safety Science, 110, 253-267, https://doi.org/10.1016/j.ssci.2018.05.014, 2018.

15 Le Bihan, G., Payrastre, O., Gaume, E., Moncoulon, D., and Pons, F.: The challenge of forecasting impacts of flash floods: test of a simplified hydraulic approach and validation based on insurance claim data, Hydrology and Earth System Sciences, 21, 5911-5928, 10.5194/hess-21-5911-2017, 2017.

Le Bissonnais, Y., Montier, C., Jamagne, M., Daroussin, J., and King, D.: Mapping erosion risk for cultivated soil in France, CATENA, 46, 207-220, https://doi.org/10.1016/S0341-8162(01)00167-9, 2002.

20 Le Coz, J., Patalano, A., Collins, D., Guillen, N. F., Garcia, C. M., Smart, G. M., Bind, J., Chiaverini, A., Le Boursicaud, R., Dramais, G., and Braud, I.: Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand, Journal of Hydrology, 541, 766-777, 10.1016/j.jhydrol.2016.07.036, 2016.

Llasat, M. C., Llasat-Botija, M., Petrucci, O., Pasqua, A. A., Rossello, J., Vinet, F., and Boissier, L.: Towards a database on societal impact of Mediterranean floods within the framework of the HYMEX project, Natural Hazards and Earth System
25 Sciences, 13, 1337-1350, 10.5194/nhess-13-1337-2013, 2013.

Marra, F., Nikolopoulos, E. I., Creutin, J. D., and Borga, M.: Space-time organization of debris flows-triggering rainfall and its effect on the identification of the rainfall threshold relationship, Journal of Hydrology, 541, 246-255, 10.1016/j.jhydrol.2015.10.010, 2016.

Martin, P., Joannon, A., and Piskiewicz, N.: Temporal variability of surface runoff due to cropping systems in cultivated
30 catchment areas: Use of the DIAR model for the assessment of environmental public policies in the Pays de Caux (France), Journal of Environmental Management, 91, 869-878, https://doi.org/10.1016/j.jenvman.2009.11.003, 2010.

Maurer, H., Rudzikaite, L., Kiel, J., Partzsch, I., Pelikan, V., Sedlacek, N., Mitsakis, E., Stamos, I., Papanikolaou, A., Celano, M., Mezzasalma, P., Pavan, V., Alberoni, A., Genovese, E., Przyluski, V., Hallegatte, S.: Weather Extremes: Assessment of Impacts on Transport Systems and Hazards for European Regions, doi:<hal-00803668> , 2012.

35 Moncoulon, D., Labat, D., Ardon, J., Leblois, E., Onfroy, T., Poulard, C., Aji, S., Remy, A., and Quantin, A.: Analysis of the French insurance market exposure to floods: a stochastic model combining river overflow and surface runoff, Natural Hazards and Earth System Sciences, 14, 2469-2485, 10.5194/nhess-14-2469-2014, 2014.

Naulin, J. P., Payrastre, O., and Gaume, E.: Spatially distributed flood forecasting in flash flood prone areas: Application to road network supervision in Southern France, Journal of Hydrology, 486, 88-99, 10.1016/j.jhydrol.2013.01.044, 2013.

40 Neuhäuser, B., Damm, B., and Terhorst, B.: GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model, Landslides, 9, 511-528, 10.1007/s10346-011-0305-5, 2012.

Petrucci, O., Papagiannaki, K., Aceto, L., Boissier, L., Kotroni, V., Grimalt, M., Llasat, M. C., Llasat-Botija, M., Rossello, J., Pasqua, A. A., and Vinet, F.: MEFF: The database of MEditerranean Flood Fatalities (1980 to 2015), Journal of Flood Risk Management, 12, 10.1111/jfr3.12461, 2019.

Rubin, J.: Optimal classification into groups: An approach for solving the taxonomy problem. Journal of Theoretical Biology 15, 103–144. doi:10.1016/0022-5193(67)90046-X, 1967.

Saaty, T.L.: How to make a decision: the analytic hierarchy process, European journal of operational research, 48(1),9-26, 1990.

Saint-Martin, C., Fouchier, C., Javelle, P., Douvinet, J., and Vinet, F.: Assessing the exposure to floods to estimate the risk of flood-related damage in French Mediterranean basins, in: 3rd European Conference on Flood Risk Management, edited by: Lang, M., Klijn, F., and Samuels, P., E3S Web of Conferences, 2016.

Saint-Martin, C., Javelle, P., and Vinet, F.: DamaGIS: a multisource geodatabase for collection of flood- related damage data, Earth System Science Data, 10, 1019-1029, 10.5194/essd-10-1019-2018, 2018.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrology and Earth System Sciences, 11, 891-906, 2007.

Versini, P. A., Gaume, E., and Andrieu, H.: Assessment of the susceptibility of roads to flooding based on geographical information – test in a flash flood prone area (the Gard region, France), Nat. Hazards Earth Syst. Sci., 10, 793-803, 10.5194/nhess-10-793-2010, 2010a.

Versini, P. A., Gaume, E., and Andrieu, H.: Application of a distributed hydrological model to the design of a road inundation warning system for flash flood prone areas, Nat. Hazards Earth Syst. Sci., 10, 805-817, 10.5194/nhess-10-805-2010, 2010b.

WWRP/WGNE Joint Working Group on Forecast Verification Research: Description of methods for forecast verification, 2015. [WWW Document]. URL http://www.cawcr.gov.au/projects/verification/ (accessed 2019-05-28).

Table 1: Contingency table

	Observed impact	No observed impact
Area declared at risk	"Hits"	"False alarms"
Area declared not at risk	"Misses"	"Correct negatives"

Table 2: Evaluation criteria used in the study

	Formula	Interpretation
Probability of detection (POD)	$\dfrac{(Hits)}{(Hits)+(Misses)}$	Varies from 0 to 1 Perfect score:1
False alarm ratio (FAR)	$\dfrac{(False\ alarms)}{(Hits)+(False\ alarms)}$	Varies from 0 to 1 Perfect score:0
χ^2 test	$\sum \dfrac{(Obtained-Theoretical)^2}{Theoretical}$	$P(\chi^2 \geq 10{,}83) = 0{,}001$ « Highly significant » $P(\chi^2 \geq 7.88) = 0{,}005$ « Very significant » $P(\chi^2 \geq 6.63) = 0{,}01$ « Significant »

Table 3: Parameterization of the IRIP model for the case study. The table provides values of the thresholds used for each indicator when condition is favorable (score = 1)

IRIP maps	Indicators	Thresholds used for favorable conditions (score =1)
Generation	Soil permeability	Saturated hydraulic conductivity (Ks) < 10^{-6} m s^{-1} + urban areas
	Soil thickness	Soil thickness < 50 cm + urban areas
	Soil slaking	Urban areas + slacking ≥ 3 Slacking computed according to Cerdan et al. (2006)
	Topography	Slope > Threshold_1 OR topographic index > Threshold_2 Threshold_1 and Threshold_2 determined using a classification algorithm (Rubin, 1967)
	Land use	Urban areas and agricultural lands
Transfer	Upstream generation susceptibility	Modal value of the upstream sub-catchment ≥ 3
	Slope	Slope > Threshold_1
	Break of slope	Convex break of slope ≥ 0,0018 (GRASS GIS r.param.scale function ; 3 pixels)
	Drained area	Drained area ≥ 2.5 ha (Lagadec, 2017)
	Soil erodibility	Urban areas + erodibility ≥ 3 Erodibility computed according to Cerdan et al. (2006)
Accumulation	Upstream generation susceptibility	Modal value of the upstream sub-catchment ≥ 3
	Slope	Slope ≤ Threshold_1
	Break of slope	Concave break of slope ≤ -0,0018 (GRASS GIS r.param.scale function ; 3 pixels)
	Topographic index	Topographic index > Threshold_2
	Drained area	Drained area > 2.5 ha (Lagadec, 2017)

Table 4: Information collected about the documents collected in the archives about runoff-related impacts (left) and about the events related to runoff (right).

Information about documents	Information about runoff-related events
Document number (unique identifier)	Event number (unique identifier)
Location of the document (where the document is archived)	Start Kilometric Point (KP)
Cote of the document	End Kilometric Point (KP)
Date of the document	Name of the earthwork
Typology of the document (correspondance, report, …)	Date
Link to the numerical copy	Type (incident, works, observation, …)
Remarks	Remarks

Table 5: Performance of the IRIP model for three configurations: without taking vulnerability into account according to Figure 3a when taking vulnerability into account according to Figure 3c, but not mitigation measures; when taking vulnerability and mitigation measures into account.

	Without taking vulnerability into account (cf Fig 3a)	When taking vulnerability into account but not mitigation measures	When taking vulnerability and mitigation measures into account
Number of 'Hit'	51	55	95
Number of 'False Alarm'	83	77	37
Number of 'Correct Negative'	40	46	46
Number of 'Miss'	8	4	4
Probability of Detection: POD (%)	86	93	96
False Alarm Ratio: FAR (%)	62	58	28
χ^2	7	19	60
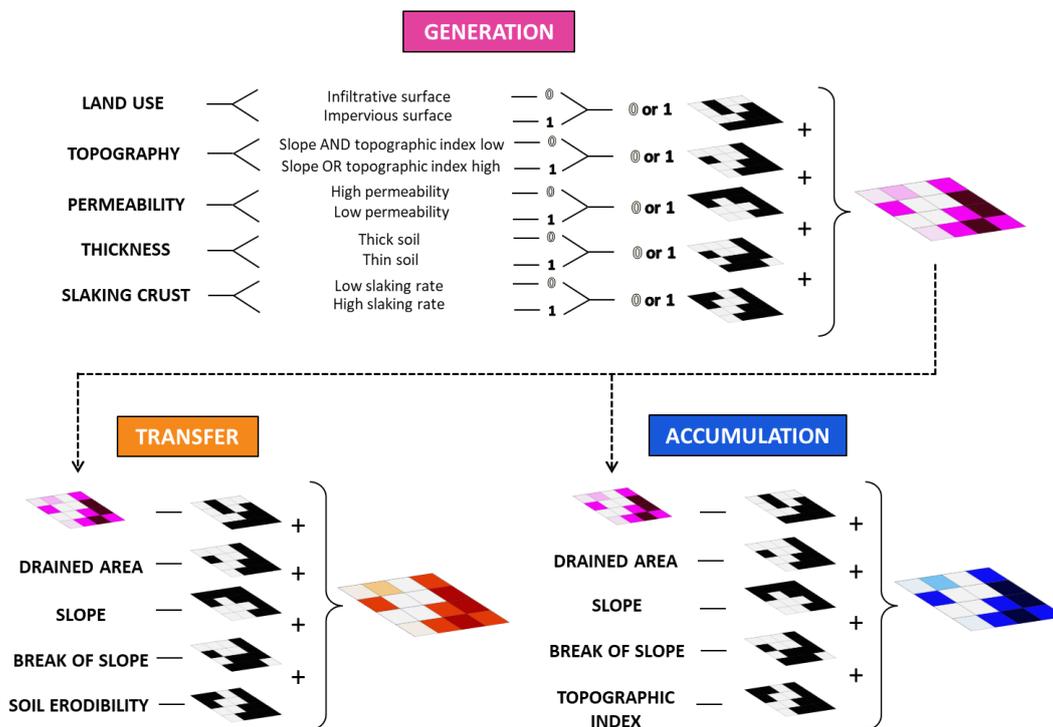
Figure 1: Scheme of the IRIP model presenting the various indicators computed to produce the susceptibility maps to runoff generation (top), transfer (bottom left) and accumulation (bottom right).
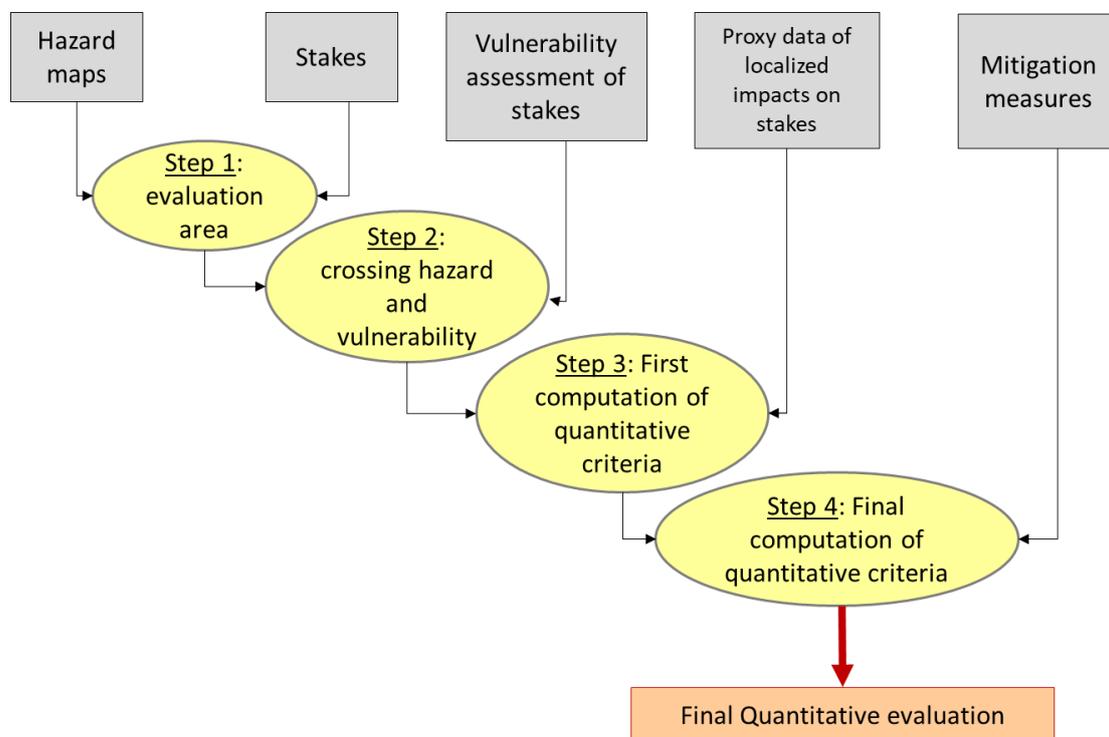
Figure 2: Scheme of the evaluation methodology to assess the relevance of susceptibility maps to runoff, using localized runoff-related impacts proxy data.

(a) (b) (c)
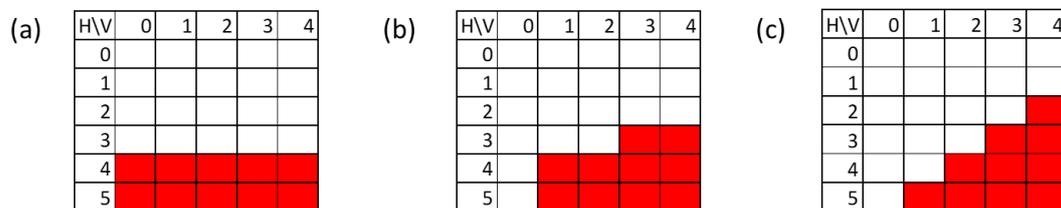
Figure 3: Examples of combination of hazard (H) (vertical) and vulnerability (V) (horizontal) to define the areas at risk with regards to runoff (red cells) when (a) vulnerability of stakes is not taken into account and IRIP hazard levels 4 an 5 are considered at risk, (b) and (c) vulnerability of stakes is taken into account in various manners based on the principle that "the higher the vulnerability, the lower the hazard level triggering risk".
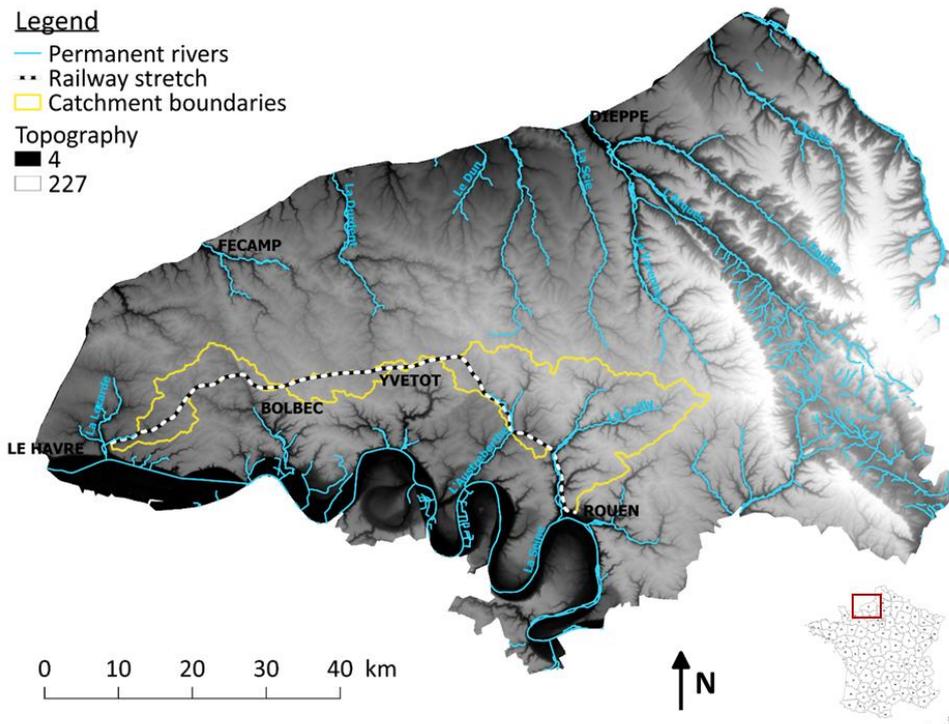
Figure 4: Map of the study area in Normandy (northern France). The yellow contour is the boundary of the catchments intercepted by the Rouen- Le Havre railway (line in black and white). Blue lines are the permanent river courses and the black lines correspond to a dense network of dry talwegs that can be activated during a rainfall event.
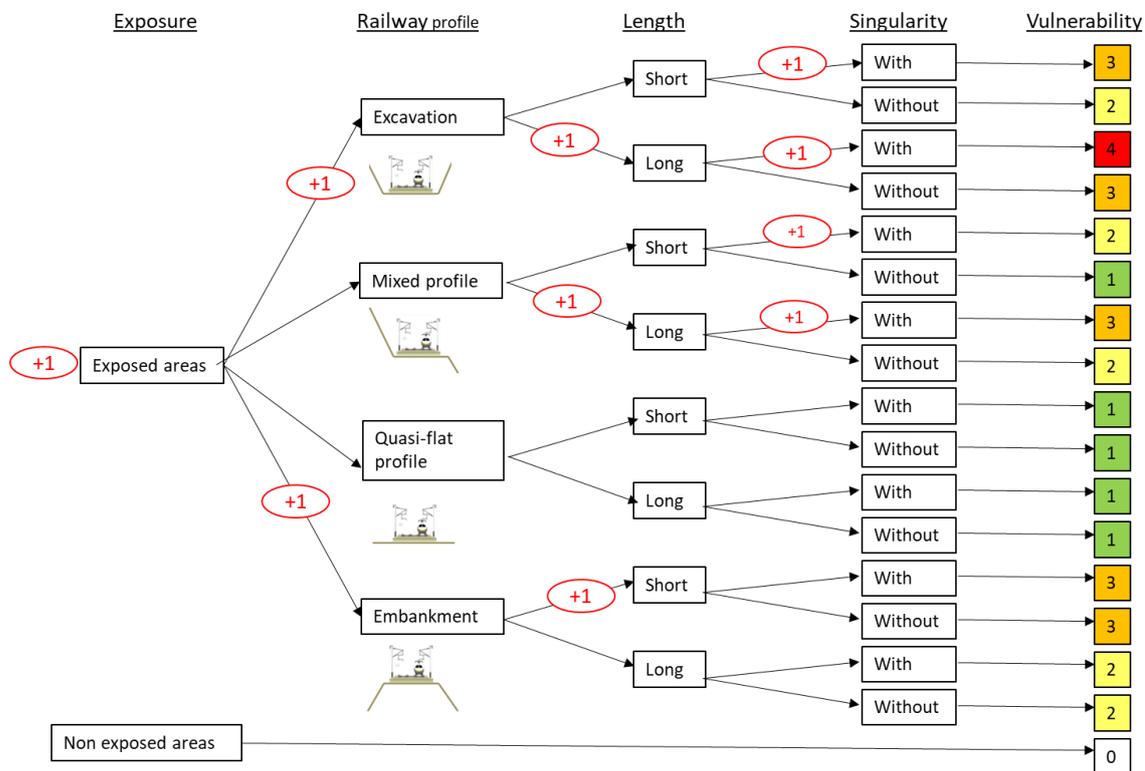
Figure 5: Vulnerability tree of the railway based on expert judgment
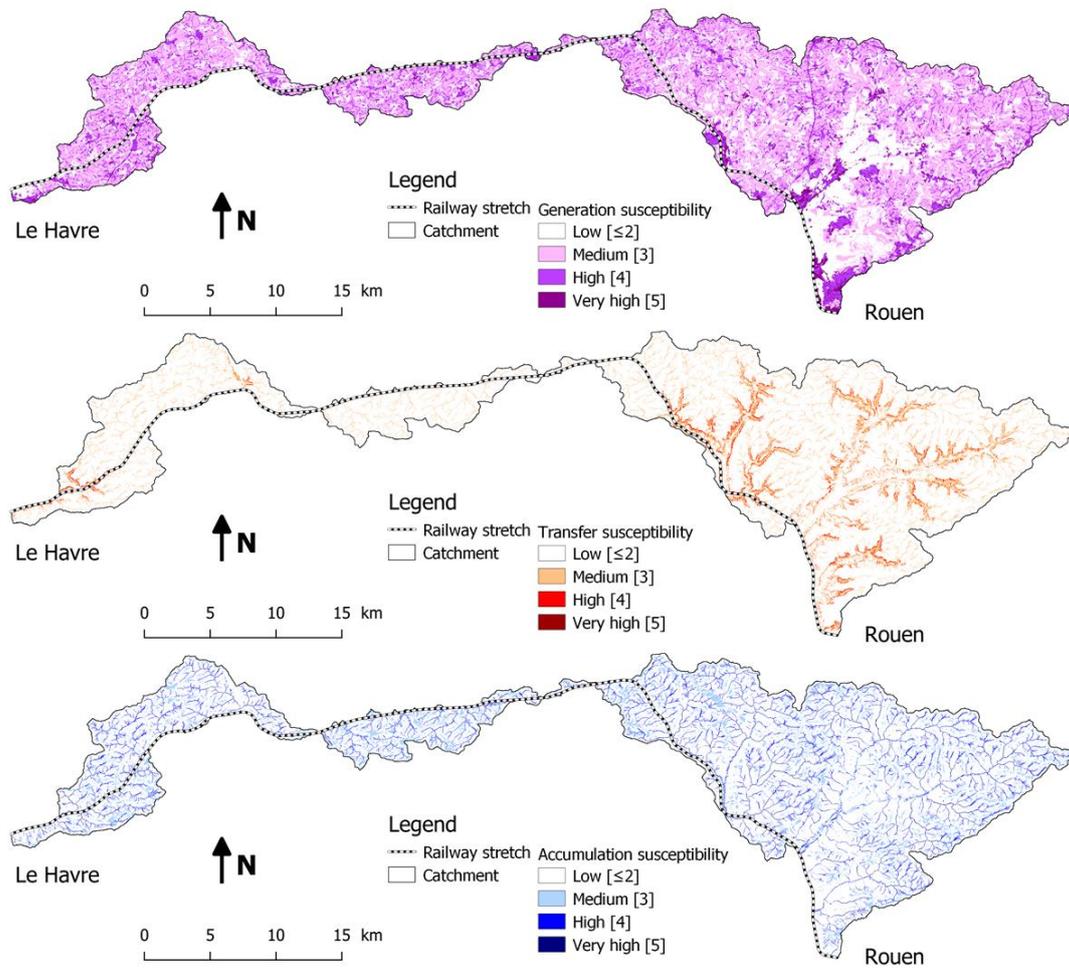
Figure 6: IRIP susceptibility maps to runoff generation (top), transfer (middle) and accumulation (bottom)
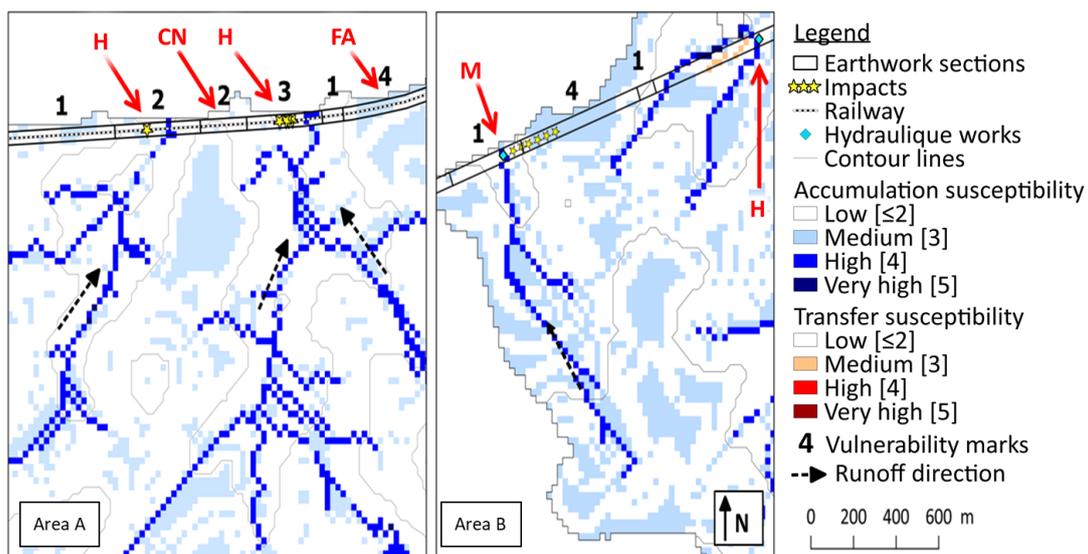
Figure 7: Illustration of the evaluation area and of the building of contingency tables on two sub-areas. The evaluation area is the buffer (black contours) along the railway line (black dotted line). The latter is divided into sections (earthwork sections) to which a vulnerability score is assigned (black numbers in the figure). Impacts are the yellow stars. Hydraulic infrastructures are figured with blue diamonds. IRIP susceptibility levels appear in red colors for transfer and blue colors for accumulation (only levels 3, 4 and 5 are drawn in the figure). Red arrows show the value assigned to the railway sections in the contingency table, where H means "Hit", CN means "Correct Negative", M means "Miss" and FA means "False Alarm".