Table 6. Performance criteria assessing the predictive power of the presence of hydraulic works or bad-weather tours in identifying sections with proven risk.

	(1) Impacts explained by hydraulic works	(2) Impacts explained by bad-weather tours	(3) Impacts explained by hydraulic works or bad-weather tours
Number of hits	22	29	40
Number of false alarms	33	21	51
Number of correct negatives	90	102	72
Number of misses	37	30	19
Probability of detection: POD (%)	37	49	68
False-alarm ratio: FAR (%)	60	42	56
χ^2	2 (not significant)	20.6	11.1

account not only the hazard parameter but also the vulnerability and the criticality of the stake. A bad-weather tour is preferably designed on a section that is critical regarding the train traffic management or on sections with known structural weaknesses. Bad-weather tours are not precise; they often involve long linear areas of the railway, and all sections of the tours may not be relevant to runoff risk. This is why we used them in step 4 to refine the computation of performance criteria, in particular as explanatory factors of false alarms and not as elements proving the existence of a risk. The results of the evaluation criteria obtained after step 4.2 assume that the existence of mitigation measures means proven risk, which is not the case. Therefore, they provide the most pessimistic POD values for the evaluation of the IRIP maps. The results presented in Table 6 also highlight that the bad-weather tour is more reliable proxy data for runoff-related risk than hydraulic works. But this may be due to the number of railway sections concerned in one bad-weather tour, whereas local mitigation measures only affect one section.

One of the reasons for this low predictive power of hydraulic works could be the following. Blockage of culverts or drainage pipes is a common problem in the railway context. In addition to blockage related to a particular intense event, progressive filling of the infrastructure by diffuse sediment transport is also a difficulty, since there are a large number of small hydraulic works that are difficult to maintain. Unfortunately, the information on blockage of hydraulic works is rarely documented in the reports about the impacts found in the archives. On the other hand, other hydraulic works are well dimensioned and very efficient. Thus, hydraulic works can sometimes increase the vulnerability and sometimes decrease it. Therefore, it was not possible to consider this information to be a reliable source of information for proven risk in the evaluation methodology or in the vulnerability tree. On the other hand, the IRIP map of susceptibility to transfer, by highlighting areas prone to sediment transport, can allow management and warning to be concentrated on these areas.

Another limitation of the evaluation presented in the paper 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 1 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. Moreover, the evaluation was conducted assuming (see Sect. 2.3.3) that each section of the railway had the opportunity to be affected by runoff, i.e. that each section of the railway had the opportunity to be affected by an intense rainfall event. If it were not the case, the IRIP model could indicate a risk in a section that would not have been impacted in the absence of any intense rainfall event at that location. To assess the validity of this working hypothesis, in which "each section had the opportunity to be affected by an intense runoff event", we can calculate the probability of not having experienced a rainfall event of a given return period during 1 century. This probability is less than 0.001 % for a 10-year return period ($(1/10)^{100}$), less than 1 % ($(1/20)^{100}$) for a 20-year return period, and 13 % ($(1/50)^{100}$) for a 50-year return period. Therefore, it can be assumed that each section of the railway had the opportunity to experience a rare event at least once during the data collection period. This shows that, if the database is long enough and of course comprehensive (i.e. all the occurred runoff-related impacts were properly reported), the working hypothesis can be accepted, and, therefore, performance measures can be considered to be not biased. In the present case, the comprehensiveness of the database is exceptional but far from perfect. However, it was the best that could be collected, and the duration of data collection (more than 1 century) ensures that the chosen case study was relevant for assessing the accuracy of the proposed evaluation methodology.

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 with hedgerows (the so-called bocage) since the second world war. The IRIP model considers that urban and crop lands are both favourable to intense runoff generation.