Water Resources Management

Supplementary information to

"How do decision-tree-based machine learning techniques compare to hybrid approaches for predicting fluvial dike breach discharge?"

Vincent Schmitz[1*], Renaud Vandeghen[2], Sébastien Erpicum[1], Michel Pirotton[1], Pierre Archambeau[1], and Benjamin Dewals[1]

[1]Hydraulics in Environmental and Civil Engineering, Urban and Environmental Engineering, University of Liege, Belgium.

[2]Department of Electrical Engineering and Computer Science, University of Liège, Belgium

[*]Corresponding author: Vincent Schmitz (v.schmitz@uliege.be)

Text S1 – Experimental data

The experimental setup consists of a horizontal trapezoidal straight main channel (10 × 1 m) (Figure S1). A 3-m-long trapezoidal dike built with uniform sand of median diameter d_{50} = 1 mm was erected along the right side of the main channel. A horizontal floodplain was present beside the dike (4.3 × 2.5 m). Its bottom and the main channel bottom were at the same level. The main channel and floodplain were covered with an impermeable whitewash coating to ensure roughness continuity between the flume, floodplain, and sand dike (Rifai et al., 2017). A drainage system was installed at the dike bottom to control the seepage through the dike body. A perforated plate, followed by a reservoir, was placed at the downstream extremity of the main channel. The perforated plate was adjusted such that, at the beginning of each test, the water level corresponded to the dike crest elevation. To trigger breaching, a 2-cm-deep and 10-cm-wide initial notch was created in the dike crest, 0.8 m from its upstream end.

Three ultrasonic sensors measured the water level in the main channel, one in the outflow tank, and one in the drainage tank (accuracy of ±1 mm). A spatially averaged water level was computed based on the measurements from the three sensors located in the main channel. This value was considered as the main channel water level for the rest of the analysis. The inflow discharge Q_{in} was measured using an electromagnetic flowmeter (accuracy of ± 0.75 l/s). The outflow discharge was derived from the discharge passing through a V-notch weir, i.e., by means of its rating curve, and mass balance in the outflow tank. The drainage discharge Q_d was estimated from the evolution of the water level measured in the drainage basin. The breach discharge Q_b was determined from mass balance in the main channel.

The 3D breach geometry was recorded for 13 tests using the laser profilometry technique presented by Rifai et al. (2020), with a resolution of 1cm x 1cm. Alternatively, the breach widening in the other tests was monitored using a simplified version of this technique: the laser sheet was not rotated, which enabled to reliably monitor the breach widening at the crest level as the laser sheet was aligned with the dike crest, but not to produce 3D reconstructions of the breach evolving geometry.

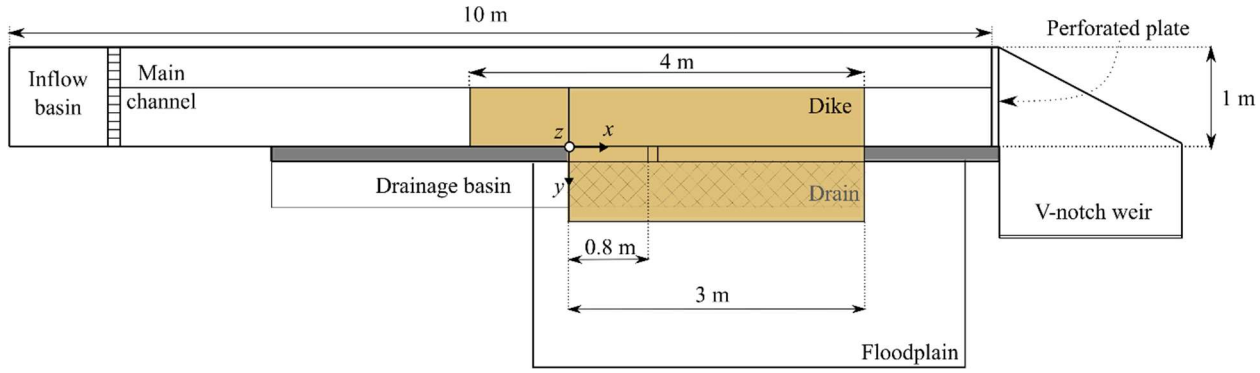

Figure S1 Laboratory setup used to collect experimental data. Not at scale.

Table S1 Experimental tests features. The target Froude number, F_{target}, is computed using the target inflow discharge, Q_{in}, and the water level equal to the dike crest. S_u and S_d stand for the dike slope on the main channel and floodplain side, respectively, while L_k is the dike crest width. The three tests corresponding to underlined values of the inlet discharge Q_{in} were repeated twice. The test corresponding to the bold underlined Q_{in} value was repeated three times. The stars depict tests for which the full 3D breach geometry was recorded.

S_u (-)	S_d (-)	L_k (m)	Q_{in} (l/s)			F_{target}		
2.0	2.0	0.15	<u>25</u>	<u>40</u>	55	<u>0.083</u>	<u>0.133</u>	0.183
1.5	1.5	0.15	25	40	55	0.071	0.114	0.157
1.5	2.0	0.15	25	40	55	0.071	0.114	0.157
2.0	1.5	0.15	25	40	<u>55</u>	0.083	0.133	<u>0.183</u>
2.0	2.5	0.15	25	40	55	0.083	0.133	0.183
2.0	3.0	0.15	25	40	55	0.083	0.133	0.183
2.0	3.0	0.30	25	-	55	0.083	-	0.183
2.0	2.0	0.00	25	40	55	0.083	<u>0.133</u>	0.183
2.0	2.0	0.30	25	40	55	0.083	0.133	0.183
2.0	2.0	0.60	-	-	55	-	-	0.183
2.0*	2.0*	0.1*	20	21	28	0.066	0.07	0.093
			30	31	**<u>40</u>**	0.100	0.103	**<u>0.133</u>**
			41	47	50	0.136	0.156	0.166
			51	55		0.169	0.182	

Table S2 Values of the hyperparameters used in the machine learning models. Values labelled with a star (*) are recommended in the *scikit-learn* documentation.

	Maximum tree depth	Split criterion	Minimum samples for split	Minimum samples in a leaf	Number of estimators
Decision trees	15	Mean square error*	2*	1*	-
Random forests					1000
Extremely randomized trees					

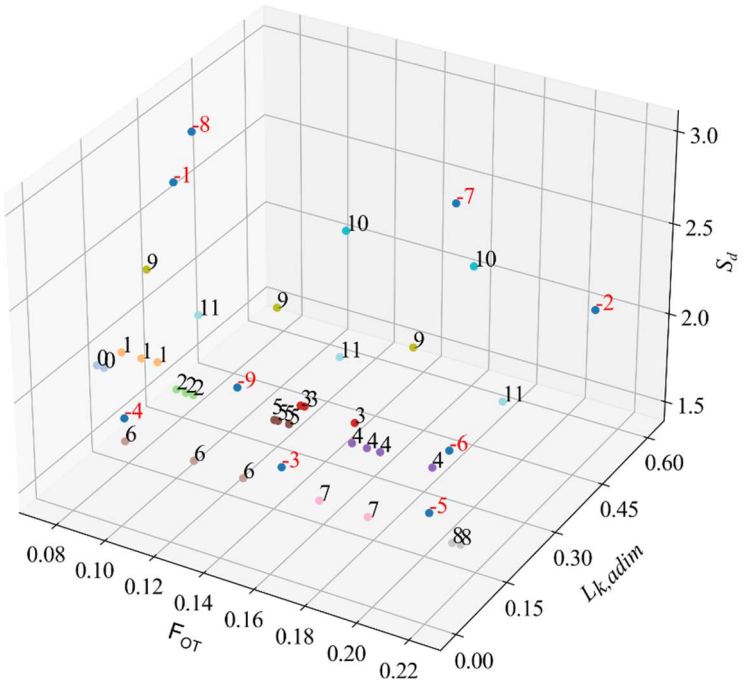

Figure S2 Clustering of the laboratory experiments. Each cluster is depicted in a different color. Tests contained in each cluster are listed in Table S3 in Supplement.

Table S3 Definition of the clusters. S_u = dike slope on the main channel side; S_d = dike slope on the floodplain side; Q_{in} = target inflow discharge; F_{OT} = main channel Froude number at overtopping initiation.

Cluster ID	S_u (-)	S_d (-)	L_k (m)	Q_{in} (l/s)	F_{OT} (-)
-9	1.5	2	0.15	40	0.125
-8	2	3	0.3	25	0.086
-7	2	3	0.3	55	0.191
-6	1.5	2	0.15	55	0.208
-5	2	2	0	55	0.221
-4	2	2	0	25	0.103
-3	2	2	0	40	0.165
-2	2	2	0.6	55	0.208
-1	2	3	0.15	25	0.100
0	2	2	0.1	20	0.075
				21	0.078
1	1.5	2	0.15	25	0.078
	2				0.086
	2				0.092
2	2	2	0.1	28	0.108
				30	0.112
				31	0.115
3	2	2	0.15	40	0.150
				40	0.152
				55	0.173
4	2	2	0.1	47	0.178
				50	0.184
				51	0.189
				55	0.209
5	2	2	0.1	40	0.147
				40	0.147
				40	0.149
				41	0.153
6	1.5	1.5	0.15	25	0.078
	2				0.106
	1.5			40	0.127
7	2	1.5	0.15	40	0.157
	1.5			55	0.177
8	2	1.5	0.15	55	0.210
					0.213
9	2	2.5	0.15	25	0.089
				40	0.142
				55	0.195
10	2	3	0.15	40	0.168
				55	0.216
11	2	2	0.3	25	0.087
				40	0.146
				55	0.209

Text S2 – New analytical model for the breach discharge

Instead of predicting the breach discharge by machine learning directly from the dike and flow parameters, an analytical expression can be derived based on sound assumptions and physics reasoning. Here, we consider shallow water flows and hydrostatic pressure distribution everywhere, and a subcritical flow in the main channel. If the backwater effect in the hinterland is negligible, a critical section must occur somewhere in the breach, so that

$$\mathrm{F}_b = \frac{U_c}{\sqrt{g\frac{A_c}{L_c}}} = 1 \tag{1}$$

where F_b and U_c stand for the Froude number and the mean flow velocity across the critical section in the breach, and A_c and L_c, the area and free surface width of this section.

The breach discharge Q_b can then be derived as

$$Q_b = U_c A_C = \sqrt{g\frac{A_c^3}{L_c}} \tag{2}$$

To simplify A_c and L_c expressions while being consistent with experimental observations (Rifai et al., 2018), the breach cross section is approximated by a trapezoid with flat bottom and side slopes equal to the dike material repose angle.

Neglecting head losses between the main channel and the breach critical cross section, the energy conservation equation writes

$$(z_r - z_b) + \frac{U_r^2}{2g} = h_c + \frac{U_c^2}{2g} \tag{3}$$

with z_b the breach invert level, U_r the mean flow velocity in the main channel upstream from the breach, h_c the critical water depth above the breach bottom, and g the gravity acceleration (Figure S3a). Injecting Eq. (2) into Eq. (3) yields

$$(z_r - z_b) + \frac{U_r^2}{2g} = h_c + \frac{A_c}{2L_c} \tag{4}$$

Using a trapezoidal breach shape allows for explicitly linking A_c and L_c with the critical water depth h_c, leading to one equation with one single unknown. Nonetheless, laboratory experiments (Rifai et al., 2017), field tests (Kakinuma, Tobita, Yokoyama, & Takeda, 2013) and detailed 2D computations (Charrier, 2015) show that the flow velocity is highly non-uniform through the breach section, especially when the dike breach is large. Kakinuma & Shimizu (2014) observed that the flow concentrates close to the breach downstream extremity whilst recirculations take place at the upstream extremity. This suggests that the whole wetted breach area should not be considered when computing the critical breach section as only a reduced section of the breach conveys most of the flow. Similarly to Schmitz et al. (2023b), we introduce a parameter α that represents the fraction of the breach width that conveys most of the breach discharge and that should be considered as the breach critical section top width :

$$L_c = \alpha L_b = \alpha\left(b_{bot} + 2mh_c\right) \tag{5}$$

with L_b the total breach width at free surface level, b_{bot} the breach bottom width and m the breach side slopes (Figure S3b).

Note that shortly after the breaching onset, the breach is narrow, and the entire breach section is actively used by the flow. Stilmant et al. (2013) showed that the critical section at the breach could be curved, leading to a critical section width that is larger than the one of the breach straight cross-section. In this case, α might be larger than unity.

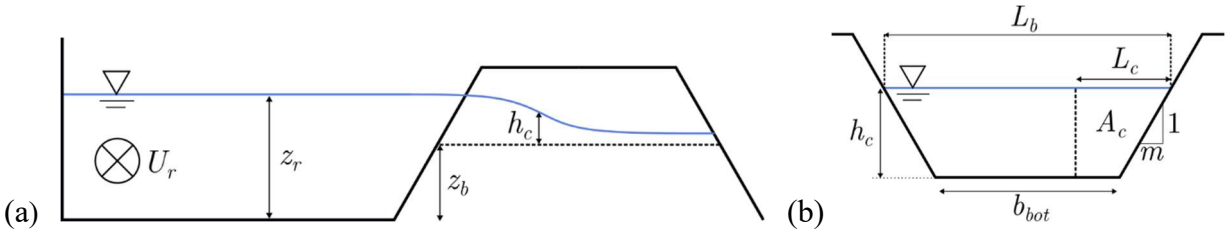

Figure S3 (a) Channel cross-section and (b) breach cross-section.

The critical height h_c is then derived from Eq. (4) as a function of α. For detailed computations, see Text S3. From then, A_c and L_c are determined and injected in Eq. (2) to obtain the breach discharge.

Similarly, α can be obtained if the breach discharge and geometry are known. The breach bottom width, b_{bot}, and bottom elevation, z_b, can be derived from the experimental breach top width, i.e., the only breach feature recorded in all experiments. This is done by assuming a uniform erosion over the entire breach surface and constant side slopes equal to the repose angle of the dike wet material. That way, an "experimental" set of α values can be generated from the experimental data in hand (Figure S4). Machine learning models presented in Section 2.1 can then be applied to predict α, following the same procedure as the one presented for the breach discharge (Section 2.2). In this work, we limited the maximum value of α to 2. Without any limit, inconsistent extreme values are encountered at the beginning of the breaching event. During this period, the breach discharge is very small, and the relative error induced by experimental uncertainties becomes substantial (Schmitz et al., 2023a), which largely impacts the value of α computed through the analytical model.

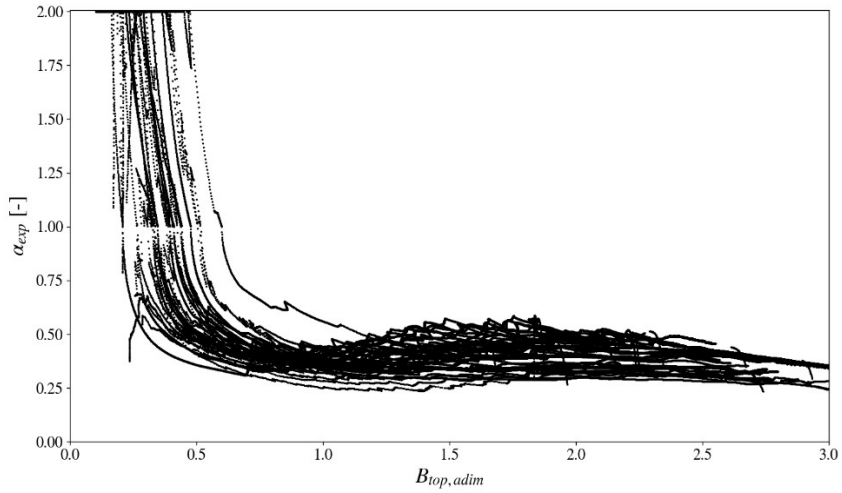

Figure S4 α values derived from experimental data as a function of the non-dimensional breach top width.

Text S3 – Detailed computation of the critical height

In the new analytical model, the energy conservation between the main channel and the breach cross-section was simplified as

$$H_r = (z_r - z_b) + \frac{U_r^2}{2g} = h_c + \frac{A_c}{2L_c},$$

with $L_c = \alpha L_b = \alpha(b_{bot} + 2mh_c)$ and $m > 0$. The effective critical section, A_c, should then be expressed as a function of the critical height, h_c, i.e. the only unknown. Three cases are encountered, which depend on the value of h_c. For this reason, an iterative solving scheme should be adopted to solve the equation.

$L_c < m\,h_c$

In this case, A_c corresponds to a triangle and writes

$$A_c = \frac{L_c^2}{2m} = \frac{\alpha^2}{2m}\left(b_{bot} + 2mh_c\right)^2,$$

which leads to

$$h_c = \frac{-\frac{\alpha}{4m}b_{bot} + H_r}{1 + \frac{\alpha}{2}}.$$

$m\,h_c \le L_c \le b_{bot} + m\,h_c$

In this case, A_c corresponds to a trapezoid and writes

$$A_c = \left(L_c - mh_c\right)h_c + m\frac{h_c^2}{2} = \left[\alpha b_{bot} + mh_c\left(2\alpha - 0.5\right)\right]h_c.$$

The resulting expression of the critical height is

$$h_c = \frac{4\alpha}{12\alpha - 1}\left[H_r - 0.75\frac{b_{bot}}{m} + \sqrt{\left(0.75\frac{b_{bot}}{m}\right)^2 + \frac{H_r\left(6\alpha - 1\right)}{4\alpha}\frac{b_{bot}}{m} + H_r^2}\right].$$

$b_{bot} + m\,h_c < L_c \le b_{bot} + 2\,m\,h_c$

In this configuration, A_c corresponds to a pentagon and writes

$$A_c = \left(L_c + b_{bot}\right)\frac{h_c}{2} - \frac{\left(L_b - L_c\right)^2}{2m} = b_{bot}h_c + mh_c^2 - \frac{\left(1-\alpha\right)^2}{2m}\left(b_{bot} + 2mh_c\right)^2$$

It comes

$$h_c = \frac{\left(2\alpha^2 - 6\alpha + 1\right) b_{bot} + 4\alpha m H_r + \sqrt{\Delta}}{-m\left(4\alpha^2 - 16\alpha + 2\right)},$$

$$\text{with } \Delta = \left[\left(2\alpha^2 - 6\alpha + 1\right) b_{bot} + 4\alpha m H_r\right]^2 + 4\left[m b_{bot} - m\left(2\alpha^2 - 8\alpha + 1\right)\right]\left[\frac{(1-\alpha)^2}{2m} b_{bot}^2 + 2\alpha H_r b_{bot}\right].$$

$L_c > b_{bot} + 2 m h_c$

This last case is encountered when the critical section, A_c, is larger than the entire breach cross-section, A_b. It yields $A_c = \alpha A_b$, and

$$h_c = \frac{2}{5}\left[H_r - \frac{3}{4}\frac{b_{bot}}{m} + \sqrt{\left(\frac{3}{4}\frac{b_{bot}}{m}\right)^2 + \frac{3}{4}\frac{b_{bot}}{m} H_r + H_r^2}\right].$$

Table S4 Mean absolute relative error (MARE) [%] related to each ML model with Q_b or α as the target variable, and when computing $Q_b(\alpha)$ using the predicted values of α. The values presented here are averaged MARE obtained when all clusters are successively considered as test sets.

ML models	Q_b	α	$Q_b(\alpha)$
MLR	13.4	34.1	38.3
Decision tree	7.6	7.9	10
Random forest	6.7	7.3	9.3
Extremely randomized trees	5.8	7.1	8.8

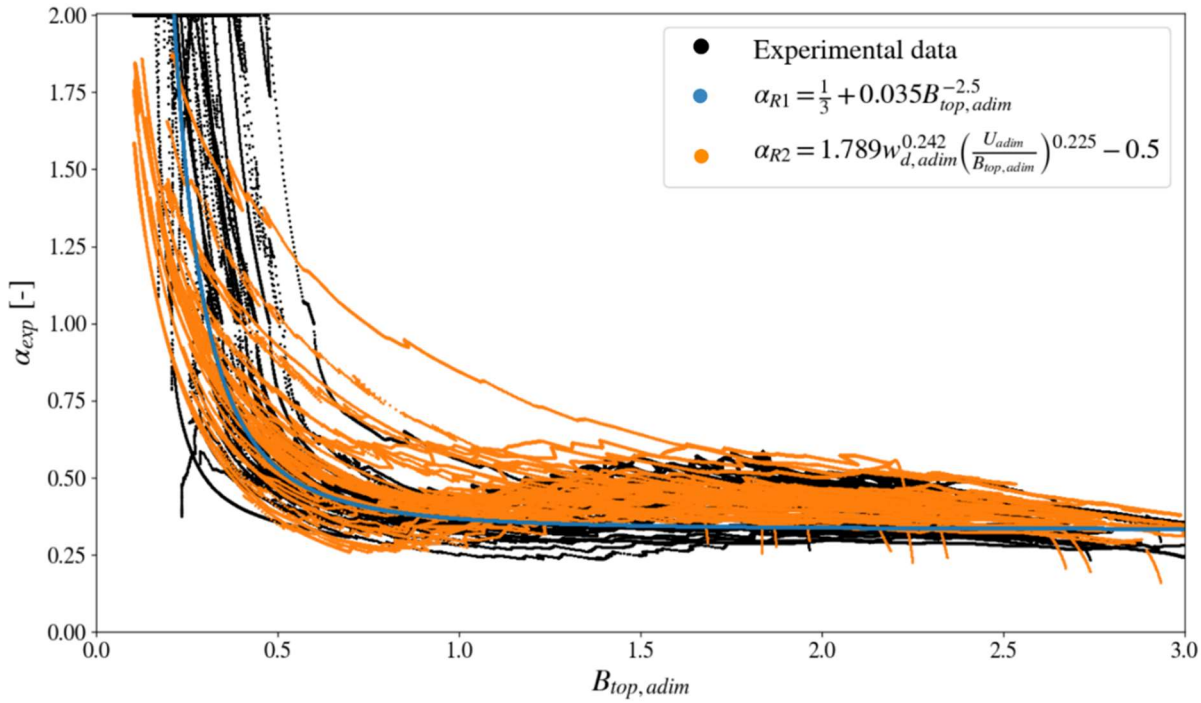

Figure S5 Fitting of the α regression formulas.

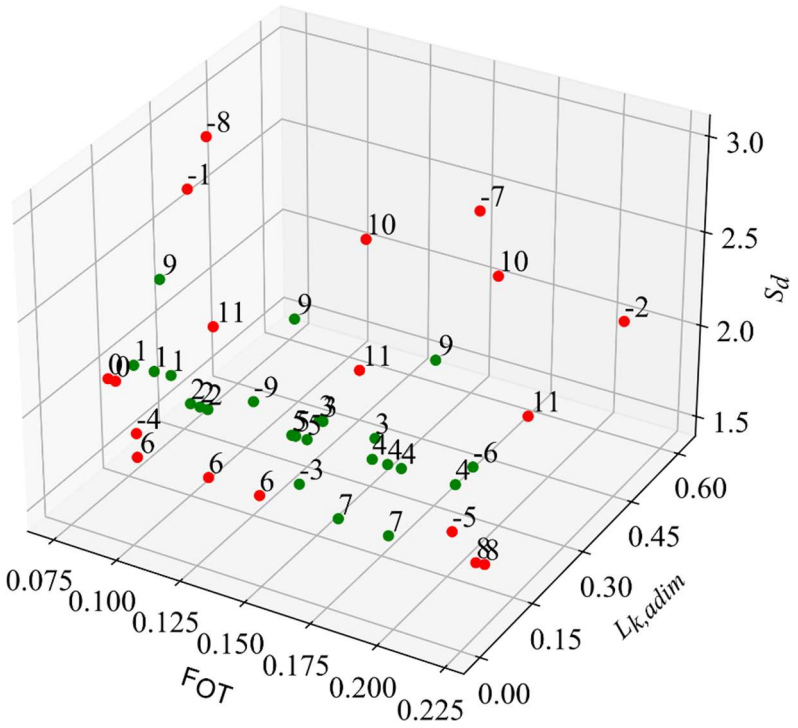

Figure S6 Clusters in extrapolation (red dots), and in interpolation (green dots) when used as test sets. Numbers refer to the clusters ID of each laboratory experiment.

Table S5 Averaged mean absolute relative error (MARE) [%] for each model and target variable when evaluated in interpolation and extrapolation.

	Q_b		$Q_b(\alpha)$		Q_b^+		$Q_b^+(\alpha)$	
	Interp.	Extrap.	Interp.	Extrap.	Interp.	Extrap.	Interp.	Extrap.
Extremely randomized trees	4.5	7	5.8	11.4	-	-	-	-
Singh et al. (1994) $C_d = 0.33 - 0.18\ \mathrm{F} + 0.49\frac{z_b}{h}$	36.7	38.1	-	-	5.5	9.9	-	-
Jalili & Borghei (1996) $C_d = 0.71 - 0.41\ \mathrm{F} - 0.22\frac{z_b}{h}$	75.9	68.5	-	-	9.9	15.4	-	-
α regression 1 (Eq. 1.4)	-	-	16.5	20.9	-	-	12.2	13.9
α regression 2 (Eq. 1.5)	-	-	10.2	24.2	-	-	6.9	10.5

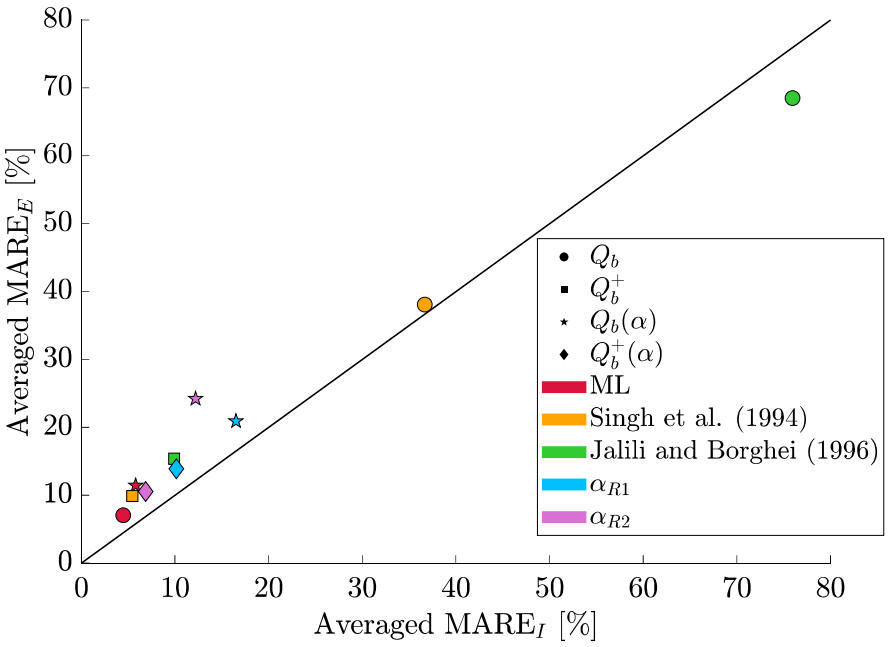

Figure S7 Averaged MARE obtained in interpolation (MARE$_I$) and extrapolation (MARE$_E$).

References associated to the "Supplementary information"

Charrier, G. (2015). *Etude expérimentale des ruptures de digues fluviales par surverse (Doctoral dissertation, Aix-Marseille)*. Aix-Marseille.

Jalili, M. R., & Borghei, S. M. (1996). Discussion: Discharge Coefficient of Rectangular Side Weirs. *Journal of Irrigation and Drainage Engineering*, *122*(2), 132–132. doi:10.1061/(asce)0733-9437(1996)122:2(132)

Kakinuma, T., & Shimizu, Y. (2014). Large-scale experiment and numerical modeling of a riverine levee breach. *Journal of Hydraulic Engineering*, *140*(9).

Kakinuma, T., Tobita, D., Yokoyama, H., & Takeda, A. (2013). Levee breach observation at Chiyoda experimental flume. *12th International Symposium on River Sedimentation (ISRS), IRTCES, Kyoto, Japan*.

Rifai, I., El Kadi Abderrezzak, K., Erpicum, S., Archambeau, P., Violeau, D., Pirotton, M., & Dewals, B. (2018). Floodplain Backwater Effect on Overtopping Induced Fluvial Dike Failure. *Water Resources Research*, *54*(11), 9060–9073.

Rifai, I., Erpicum, S., Archambeau, P., Violeau, D., Pirotton, M., El Kadi Abderrezzak, K., & Dewals, B. (2017). Overtopping induced failure of noncohesive, homogeneous fluvial dikes. *Water Resources Research*, *53*(4), 3373–3386.

Rifai, I., Schmitz, V., Erpicum, S., Archambeau, P., Violeau, D., Pirotton, M., Dewals, B., et al. (2020). Continuous Monitoring of Fluvial Dike Breaching by a Laser Profilometry Technique. *Water Resources Research*, *56*(10), e2019WR026941. doi:https://doi.org/10.1029/2019WR026941

Schmitz, V., Arnst, M., Abderrezzak, K. El kadi, Pirotton, M., Erpicum, S., Archambeau, P., & Dewals, B. (2023a). Global sensitivity analysis of a dam breaching model: To which extent is parameter sensitivity case-dependent? *Water Resources Research*, e2022WR033894.

Schmitz, V., Rifai, I., Kheloui, L., Erpicum, S., Archambeau, P., Violeau, D., Pirotton, M., et al. (2023b). Main channel width effects on overtopping-induced non-cohesive fluvial dike breaching. *Journal of Hydraulic Research*, *61*(5), 601–610.

Singh, R., Manivannan, D., & Satyanarayana, T. (1994). Discharge Coefficient of Rectangular Side Weirs. *Journal of Irrigation and Drainage Engineering*, *120*(4), 814–819. doi:10.1061/(asce)0733-9437(1994)120:4(814)

Stilmant, F., Pirotton, M., Archambeau, P., Roger, S., Erpicum, S., & Dewals, B. (2013). Dike-break induced flows: a simplified model. *Environmental fluid mechanics*, *13*, 89–100.