39th IAHR WORLD CONGRESS GRANADA, SPAIN 2022 From Snow to Sea 19-24 June 2022

Proceedings of the 39th IAHR World Congress
19–24 June 2022, Granada, Spain
doi://10.3850/IAHR-39WC2521716X20221036

A Machine Learning-Based Surrogate Model for the Identification of Risk Zones due to Off-stream Reservoir Failure

Nathalia Silva Cancino[(1)], Fernando Salazar[(2)], Marcos Sanz-Ramos[(3)] and Ernest Bladé i Castellet [(4)]

[(1,2)] International Center for Numerical Methods in Engineering, Barcelona, Spain,
e-mail:nsilva@cimne.upc.edu[1], fsalazar@cimne.upc.edu[2]

[(3,4)] Institut Flumen, Universitat Politecnica de Catalunya- International Center for Numerical Methods in Engineering, Barcelona, Spain,
e-mail: marcos.sanz-ramos@upc.edu[3], ernest.blade@upc.edu[4]

Abstract

With the modification of the Regulations of the Hydraulic Public Domain of Spain in 2008, approximately 70.000 owners of off-stream reservoirs are obligated to present a classification assessment on the potential risk due to failure, which requires complex procedures. This work proposes a simplified methodology based on Machine Learning, which allows identifying risk zones at any point at the affected area based on the physical characteristics of the reservoir and the surrounding terrain. Random Forest algorithm is applied to two datasets generated with synthetic cases designed and modelled in Iber. Two methods were tested for balancing the datasets: synthetic minority over-sampling and random under-sampling. Results show high accuracy on both models, although the Random Forest model adjusted with random under-sampling presented better results for the estimation of risk zones. In conclusion, this work found that the simplified method based on Machine Learning can be a useful tool to owners and government administrations, having an equally reliable estimation than current methods and reducing the computational time and resources.

Keywords: Machine Learning; Iber; Off-stream reservoirs; Dam breach; Floods

1. INTRODUCTION

Off-stream reservoirs are a key factor for the regulation and supply of water, especially for irrigation. In contrast to dams, they are not located in water bodies and, generally, are not affected by surface runoff. Although this implies higher hydrological security, they are frequently located in high elevations near to urban areas and infrastructure that might be affected in case of failure. There are around 70.000 off-stream reservoirs in Spain (de Cea, 2021) that before 2008 were not obligated to present a classification assessment on the potential risk due to failure.

With the modification of the Regulation of the Hydraulic Public Domain with the Royal Decree 9/2008 (Ministry of the presidency, 2008), owners of off-stream reservoirs higher than 5 m or with volume storage above than 100.000 m^3 have to follow the same procedure as large dams to estimate its classification on the potential risk due to failure.

The Spanish Technical Guide for dam classification on the potential risk (Ministry for the ecological transition, 2021), hereinafter Technical Guide, lists the recommended methods to analyze the breach of the dam and propagation of the flood as a consequence of the failure.

One of the procedures described in the Technical Guide is the so-called 'Complete method'. It encourages the use of complete hydraulic models since they are the only ones capable of accurately representing the formation of the breach, the definition of the hydrograph after the break, the propagation of the flood wave and, therefore, the determination of the flood area, the dangerousness of the flow and the associated risks (Sanz-Ramos et al., 2019)

Hence, the necessary work to make a complete and accurate classification assessment on the potential risk due to failure includes relevant engineering knowledge, efforts and resources, often unavailable for many owners. Also, the competent administration that revises and approves the classification assessments and emergency plans requires a high demand of resources (time and people).

In the last years, Machine Learning (ML) techniques have been applied for dam breaching and safety analysis, processes that need a lot of human and time resources (Hariri-Ardebili & Pourkamali-Anaraki, 2018, 2018; Hooshyaripor et al., 2014, 2015; Salazar et al., 2015, 2021). These ML approaches can predict the output of a system by learning patterns from a training set. The main drawback is that large amount of data is needed, which is not always available. Other techniques have been applied for simplifying these analyses, such as GIS Automatization tools (Albano et al., 2019; Cannata & Marzocchi, 2012).

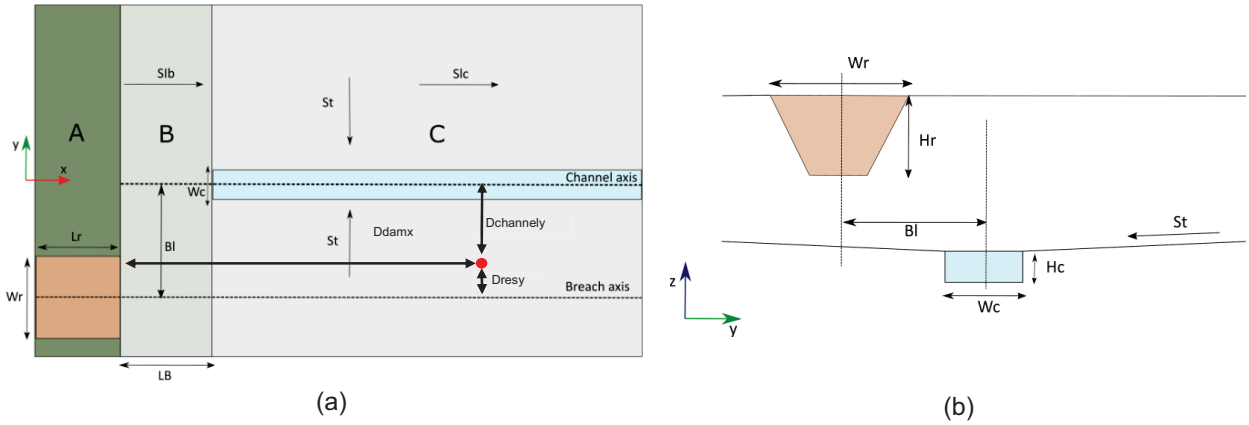
Hence, to facilitate the assessment of potential risk for off-stream reservoirs, this work proposes a simplified methodology based on ML. The required large dataset was created with the results of synthetic cases modelled using an automatized tool in Iber (Bladé Castellet et al., 2014; Bladé Castellet, Cea & Corestein, 2014). This methodology allows estimating the risk level at any point at the surrounding terrain of an off-stream reservoir based on the physical characteristics of the structure and terrain without the need to build full 2D hydraulic models.

2. DATA GENERATION

The geometry of each synthetic was designed based on ten parameters (Table 1). **Error! Reference source not found.**(a) shows the plan view and sections (A, B and C) of the synthetic geometry, while Figure 1(b) depicts frontal view. The dimensions of Table 1 are plotted in both figures. The model domain size was 4000 m of length and 1500 m of width. In section A (**Error! Reference source not found.**(a)), the off-stream reservoir is located and has the highest elevation. The location of the off-stream reservoir can vary along the y-axis. Section B (**Error! Reference source not found.**(a)) represents a transition zone between the off-stream reservoir and section C with a longitudinal slope between 0 and 1% and no transversal slope. In section C (**Error! Reference source not found.**(a)), in some cases include a main channel in the origin of the y-axis (y=0) with a rectangular cross section, and others only feature transversal slope without channel (W_c=0).

Table 1 Geometry parameters for synthetic cases and value ranges

Parameter	Symbol	Unit	Min.	Max.
Breach location concerning the axis of the main channel	B_l	m	-750	750
Off-stream reservoir width	W_r	m	50	250
Off-stream reservoir length	L_r	m	60	550
Off-stream reservoir height	H_r	m	5	15
Width preferred channel	W_c	m	0	10
Transversal slope	S_t	%	0	2
Longitudinal slope section C	Sl_c	%	0.1	10
Longitudinal slope section B	Sl_b	%	0	1
Depth preferred channel	H_c	m	0	2
Length section B	L_B	m	10	500



(a) (b)

Figure 1 Geometry of synthetic cases. a) Floor plan view, flow direction to the right on the x-axis. b) Frontal view.

Latin Hypercube Sampling (LHS) was used to generate the combination of the parameters. LHS is a statistical method for sampling, which ensures a full representation of the distribution of each input parameter using all areas of the sample space (Stein, 1987). Two sampling runs were performed using LHS. The first 50 samples were generated without taking into account channel width and depth (H_c=0 and W_c=0) and then 150 samples with all 10 parameters (Table 1). All variables were assumed to have a uniform distribution over the given ranges. With the combination of parameters, a text file with the geometry was generated using points with XYZ coordinates.

A second text file contains the breach parameters. It is composed of the start ($L_r/2$, $W_r/2$, H_r) and end ($3L_r/2$, $W_r/2$, H_r) coordinates of the breach, top elevation, bottom elevation and volume. The default maximum top elevation is 505 m for all cases. The bottom elevation is calculated as the top elevation minus H_r. The volume storage in the reservoir is estimated as a trapezoidal prism with H_r, W_r and L_r.

For the extraction of results, a third text file was created with the x and y coordinates of 200 points located along the terrain, where depths and velocities are extracted. As the points are given as a pair of independent coordinates, a parametrization based on their location was employed. Three variables were added ((a)). Dresx is the distance on the x-axis between the point and the off-stream reservoir, Dresy is the distance between the point and the breach axis on the y-axis, and Dchannely is the distance between the point and the channel axis on the y-axis.

3. IBER

Iber is a 2D numerical tool for modelling hydrodynamic and sediment transport (Bladé Castellet et al., 2014; Bladé Castellet, Cea & Corestein, 2014), which currently includes additional calculation modules for simulating hydrological processes (Cea & Bladé Castellet, 2015; Sanz-Ramos et al., 2021), pollutant propagation (Cea et al., 2016), large-wood transport (Ruiz-Villanueva et al., 2014) and physical habitat suitability assessment (Sanz-Ramos et al., 2019). Iber includes the Breach tool to simulate a gap in the dyke by entering parameters. The breach dimensions are calculated with the equations proposed in the Technical Guide.

A tool in Iber was developed to automatize the process of generation, calculation and extraction of results for massive cases. This tool reads the three text files for each synthetic case. This tool reads the three text files for each synthetic case. It generates the polygons, surfaces and mesh, runs the model and exports the results in CSV format. The tool also assigns the land use to the whole model. Six-terrain roughness are considered: 0.020 s/ $m^{1/3}$, 0.025 s/$m^{1/3}$, 0.032 s/$m^{1/3}$, 0.05 s/ $m^{1/3}$, 0.08 s/$m^{1/3}$ and 0.12 s/ $m^{1/3}$. Therefore, the tool calculates each synthetic case with six different Manning's coefficients, resulting in 1200 models (200x6).

The tool saved the depth and velocity every 300 seconds for the evaluation points, and the maximum value of both was extracted. The total calculation time was set at 16000 seconds. It was estimated based on the maximum terrain length at the x-axis (4000 m) and an assumed average velocity of 0.25 m/s. This maximum simulation time ensures that the model calculates the total breach rupture and the flood wave spreads over the entire area.

4. DEFINITION OF RISK OF DAMAGE

The Royal Decree 9/2008 in its article 9.2 states that a risk to human life is considered to exist when the depth of water is 1 m or more, the velocity is 1 m/s or more, or the product of depth and velocity is 0.5 m/s or more (Ministry of the presidency, 2008). Therefore, two classes were defined: class 0 when neither of the three criteria is true and class 1 when one or more criteria are true.

5. MACHINE LEARNING

The goal of building algorithms using Machine Learning (ML) is to produce a system that can reproduce results with data never seen before, based on patterns previously identified with a massive database. To develop a ML supervised algorithm, two databases are required: data for training and testing. ML algorithms with supervision are divided according to the type of desired output: regression or classification (Bonaccorso, 2017).

The problem at hand corresponds to classification, since the expected result is a label for each point indicating whether there is riks to human life. Among the different types of classification algorithms, for this work we considered Random Forest.

5.1 Random Forest

The Random Forest (RF) algorithm is a combination of different decision trees, each of which is trained on a bootstrapped sample taken from the original training sata set (Breiman, 2001). The bootstrapping method involves resampling the data with replacement, eliminating some of the samples and duplicating others. The RF method introduces an additional dimension of randomness by taking a random subsample of the input variables for computing every split point for each tree. The final prediction of the algorithm is calculated based on the average of the results of each of the decision trees. For the implementation of RF, the scikit-learn package developed for Python was used, which includes the Random Forests Classifier option (Pedregosa et al., 2011).

The training set for fitting the algorithm includes the results from 150 geometries (900 models) and the test set includes the results of 50 geometries (300 models). The models selected for each set were chosen randomly. Additionally, the results from 20 synthetic models not included in the training or testing set (each one with six different land uses), were used to create a validation dataset.

The distribution of classes of the training set presents 68% for class 0 (no risk) and 31% for class 1 (risk). Imbalance datasets are reported as a major problem to the development of good ML classifiers algorithms (Batista et al., 2004). Hence, the training data was balanced with the synthetic minority over-sampling technique (SMOTE) (Chawla et al., 2002) and the random under-sampling method, obtaining two modified training sets. Hereafter, the RF model fitted with the dataset balance with SMOTE is called RF-OS model and the model fitted with the dataset balanced with random under-sampling is called RF-US model.

6. RESULTS OF MACHINE LEARNING ALGORITHM

6.1 Calibration

The observations not taken into account in a bootstrap sample are called out-of-bag data (OOB). The OOB predictions are calculated only using the trees for which the observations are OOB (Zhang & Ma, 2012). The OOB error is calculated as the average of the OOB error rate for each tree. This result is nearly identical to the one estimated with cross-validation with N-folds (Hastie et al., 2017; Salazar et al., 2021a).

A calibration process was performed based on the OOB error rate to find the best combination of parameters for the RF models. With different possible combinations of the parameters, the OOB error rate was estimated and the combination with the lowest error rate was selected to fit the final RF model. All possible combinations of the number of trees: *n_estimators* (600,800,1000 and 1200) and the minimum number of samples required to split an internal node: *min_sample_split* (2,3 and 5), were considered to fit the RF models.

Figure 2 shows the OOB error rate for all combinations of parameters tested for the RF-OS model and RF-US. The variation between combinations of the parameters was low. However, the combination corresponding to the lowest value was selected for both cases. The best combination for RF-OS: n_estimators=1200 and min_sample_split=5 and for RF-US: n_estimators=1000 and min_sample_split=5.

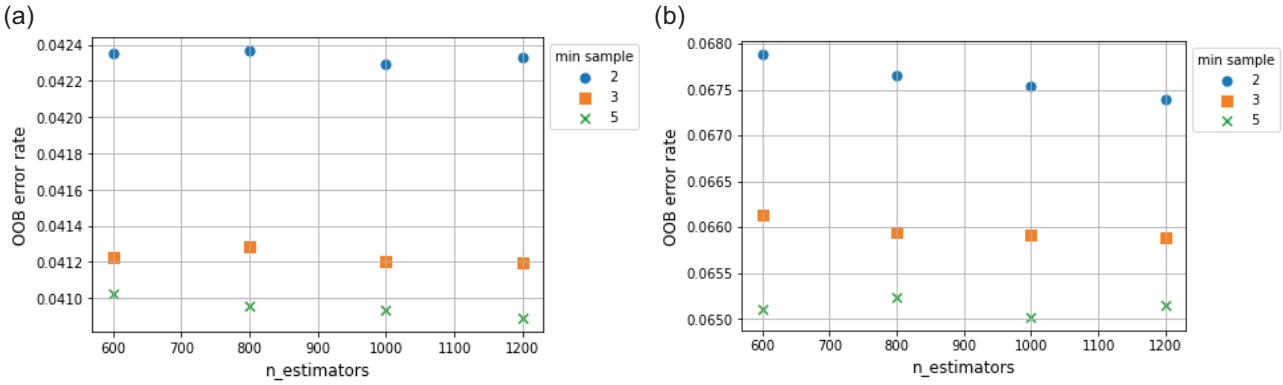


Figure 2 Results calibration process. The OOB error rate is plotted as a function of the number of trees (n_estimators) and min_sample_split parameters. (a) Results RF-OS. (b) Results RF-US.

6.2 Measures of Accuracy

Henceforth, the correct estimation of class 1 is considered as true positives (TP), while the correct prediction of class 0 corresponds to true negatives (TN). False positives (FP) are cases where the model predicts there is a risk of damage in an area where there is no risk, and false negatives (FN) are those when the model predicts no risk in areas where there is a risk. Eq. [1] to Eq. [4] present the calculated measures of accuracy.

$$Precision = \frac{TP}{TP + FP} \quad [1]$$

$$Recall = \frac{TP}{TP + FN} \quad [2]$$

$$Acuraccy = \frac{TP + TN}{TP + FP + TN + FN} \quad [3]$$

$$F\ score\ (F\beta) = (1+\beta^2)\frac{precision * recall}{\beta^2 * precision + recall} \quad [4]$$

Where,

$$\beta = 2$$

Table *2* and **Table *3*** present the confusion matrix generated with the test set for each RF model. The accuracy of the RF-OS model was 0.525 and 0.867 for the RF-US model. The RF-OS model showed a higher recall score and a very low precision score, this is due to an overestimation of class 1 areas (more TP values). The RF-US model presented a better capacity to identify class 0 and class 1, having a better precision and F2 score than RF-OS model.

Table 2 Confusion matrix for RF-OS model with the test set.

		Predicted class				
		0	**1**	**Precision**	**Recall**	**F2**
Observed class	**0**	12934	28051			
	1	124	18590	**0.399**	**0.993**	**0.760**

Table 3 Confusion matrix for RF-US model with the test set.

		Predicted class				
		0	**1**	**Precision**	**Recall**	**F2**
Observed class	**0**	37443	3542			
	1	4333	14381	**0.802**	**0.768**	**0.775**

Figure 3 and Figure 4 present maps of classification of the risk estimated with both models for two synthetic cases not included in the training set. The maps predicted with the RF-US model includes the majority of the points of class 1 estimated with Iber, although some class 0 points were predicted incorrectly. While the RF-OS model tended to estimate class 1 incorrectly in large areas. The reason might be that the balanced dataset with SMOTE has approximately 30.000 more samples corresponding to class 1 than the dataset balanced with random under-sampling. Thus the RF-US model is selected to carry out the study, guaranteeing the best performance to distinguish between classes.

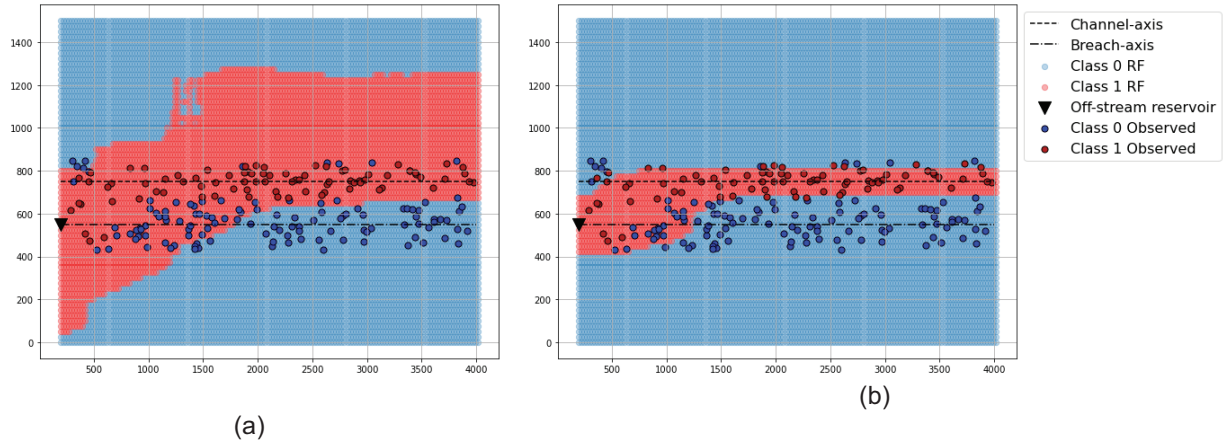
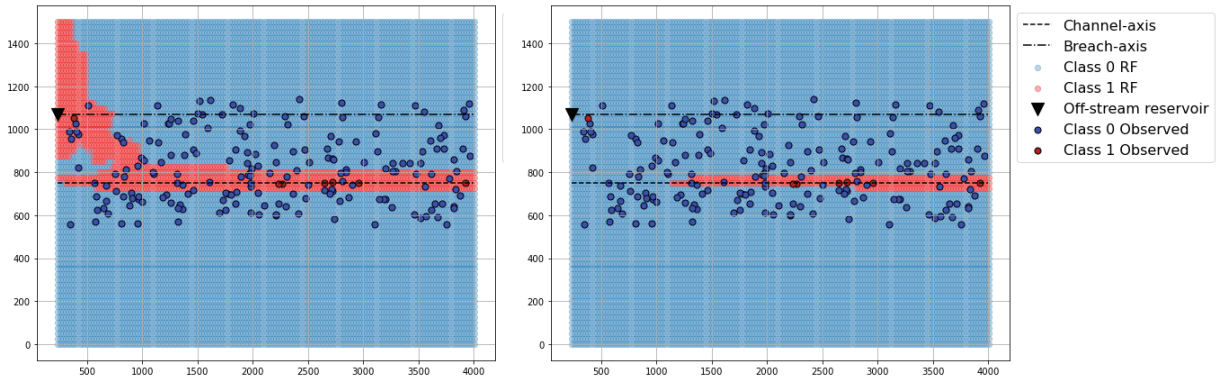


Figure 3 Map of classification risk t due to off-stream reservoir failure estimated with RF models and results from Iber for a synthetic case (ID:014 Manning coefficient:0.025 s/m$^{1/3}$). Red areas represent class 1 and blue areas class 0. (a) Results with RF-OS model. (b) Results with RF-US model.



(a) (b)

Figure 4 Map of classification risk t due to off-stream reservoir failure estimated with RF models and results from Iber for a synthetic case (ID:002 Manning coefficient:0.12 s/m1/3). Red areas represent class 1 and blue areas class 0. (a) Results with RF-OS model. (b) Results with RF-US model.

6.3 Validation

The performance of the chosen RF model (RF-US) was analyzed with the validation set. Table 4 presents the confusion matrix for the RF-US model with the prediction of the validation set. The results were similar to those estimated with the test set for the estimation of class 0. However, the prediction of class 1 showed a slight decrease in its performance (reduction of recall measure from 0.768 to 0.711).

Table 4 Confusion matrix for RF-US model with the validation set.

		Predicted class				
		0	**1**	**Precision**	**Recall**	**F2**
Observed class	**0**	11165	819			
	1	2399	5915	**0.878**	**0.711**	**0.740**

The validation dataset was divided based on the coefficient of Manning to analyze the influence of land use on the results. Figure 5 shows the results of precision, recall and F2 scores for the six Manning coefficients. The F2 score and recall tended to decrease while the Manning coefficient increases, meaning that more observed samples of class 1 are being predicted as class 0 for bigger coefficients. For the smaller coefficients (0.02 s/m$^{1/3}$, 0.025 s/m$^{1/3}$ and 0.032 s/m$^{1/3}$), the F2 score remains above 0.70. This implies that the model overestimates class 0 areas when the roughness coefficient is larger.

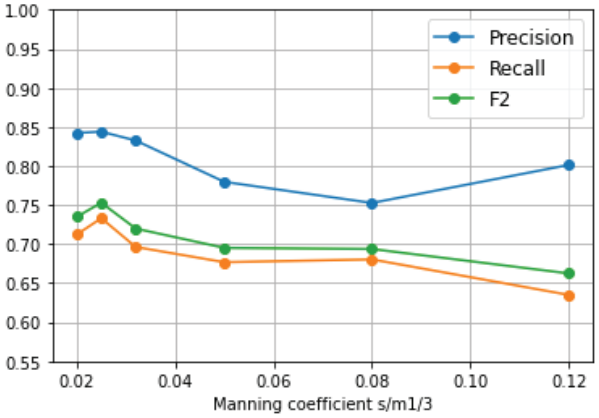


Figure 5 Accuracy measures as a function of the Manning coefficient.

6.4 Class Probability

RF model employees class probabilities to estimate the correct predicted class in each tree and overall prediction (Boström, 2007). The classifier generates a positive or negative prediction based on a probability threshold (Fernández et al., 2018). These probabilities can be extracted to obtain more detailed information. High probabilities can be expected to be more reliable than classes with similar probabilities.

Figure 6 presents probability class maps divided into the areas corresponding to class 0 and class 1 for two synthetic cases of the validation dataset. Areas with brighter colours represent higher probabilities and darker areas are corresponding to probabilities near 0.5. Most of the darker areas are in the limit between class 0 and class 1 areas, while areas far from the channel and breach axis, and near to the off-stream reservoir have higher probabilities.

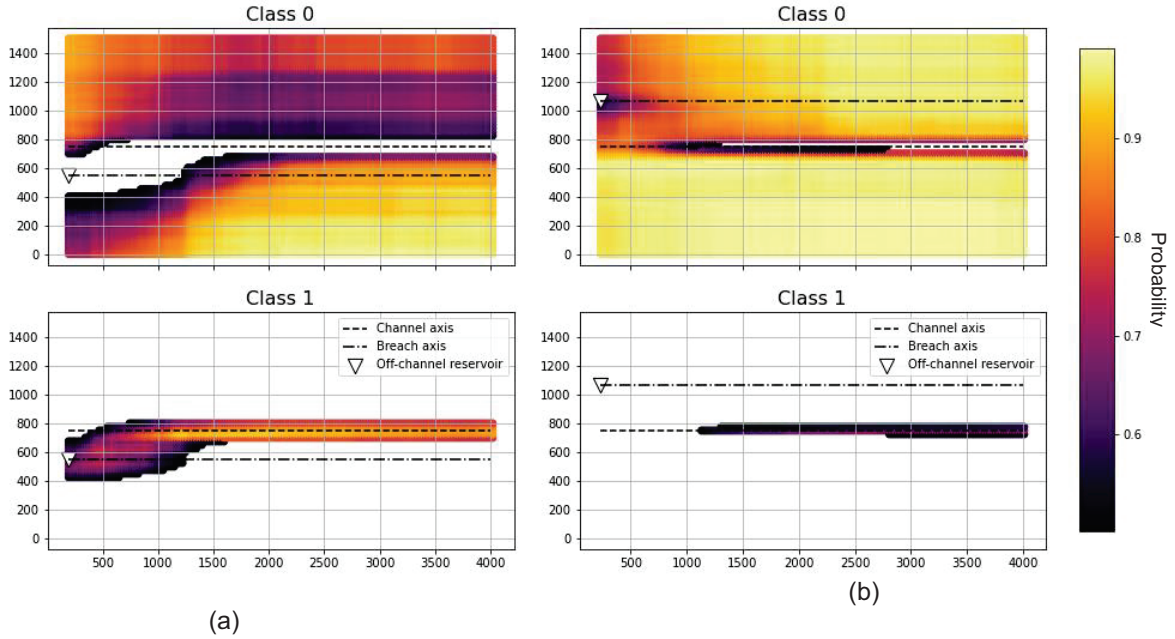


Figure 6 Probability class-map for synthetic cases estimated with the RF-US, divided into the areas corresponding to class 0 and class 1. a) ID: 014 Manning coefficient: 0.025 s/m$^{1/3}$, b) ID: 002 Manning coefficient:0.12 s/m$^{1/3}$

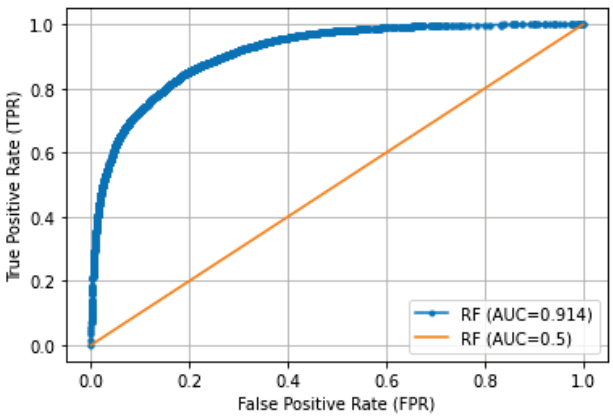


Figure 7 shows the ROC curve for the RF model. The Receiver Operating Characteristic (ROC) curve is a useful tool to visualize the performance of binary classifier algorithms, not dependent on a specific threshold. The ROC curve is generated by plotting the true positive rate (TPR) or recall, against the false positive rate (FPR). There is a point in the ROC curve for each possible threshold and the curve is generated by interpolation between these points (Fernández et al., 2018). The area under the ROC curve (AUC) is an additional measure for evaluating the predictive ability among classifier algorithms (Huang & Ling, 2005). Values of AUC nearest to 1 represent a good classifier and equal to 0.5 the model is not able to distinguish between classes. The AUC of the RF model is 0.914.

The optimal decision threshold or operational point is selected as the correspondent threshold of the nearest point to the left corner of the ROC curve. The optimal decision threshold for the RF-US model is 0.42,

similar to the default threshold. This operating point assumes that the importance of estimating correctly any of the classes is the same.

Decreasing the operating point ensures the estimation of more TP values but increase the number of FP. Different decision thresholds were used to estimate class 1 and the measures of accuracy were calculated. Figure 8 shows how the precision and recall vary as a function of the selected threshold. F2 score increases when the threshold is smaller while the precision decreases. With a decision threshold smaller than 0.5 is possible to minimize the FN and increase the TP samples, however, more FP values are estimated.

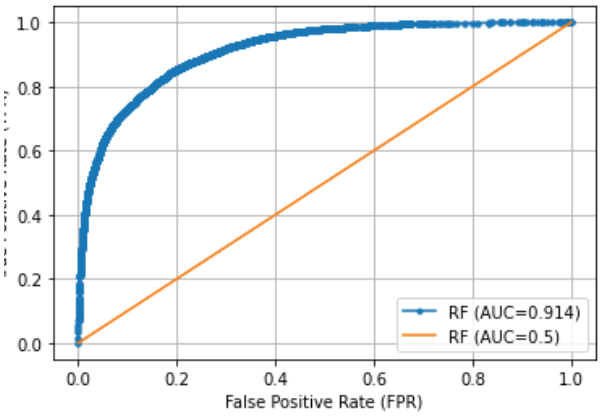


Figure 7 ROC curve and AUC for the RF-US model

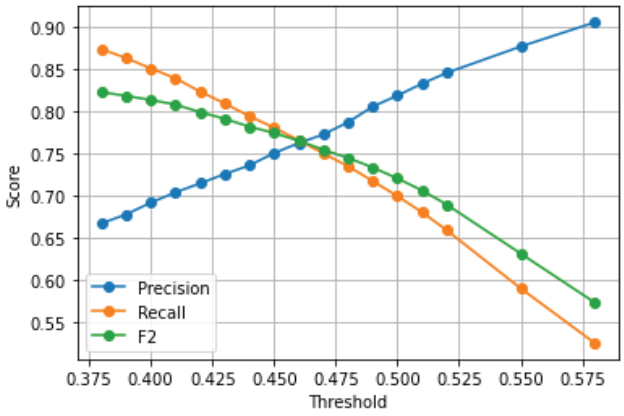


Figure 8 Accuracy measures as a function of the threshold

7. CONCLUSIONS

In this paper, we present a simplified methodology to give a first estimation of the zones with risk to human life in case of a failure of an off-stream reservoir, based on the Spanish regulations (Ministry of the presidency, 2008). The methodology was developed with an automatized tool in Iber, the design of 1200 synthetic off-stream reservoirs failure cases and the calibration and validation of RF models. The physical characteristics of the off-stream reservoir and the terrain are the only requirements to use the simplified method (ML model). The no requirement of 2D hydraulic modelling is one of the main advantages of this tool.

Two RF models were calibrated, fitted and compared. Each RF model was adjusted with different balanced datasets: balanced with SMOTE and random under-sampling. Although the accuracy and OOB error rate of both RF models are similar, the use of different balanced datasets affects the model's capacity to predict the risk areas. The model RF-US was selected as the better model, showing a superior capacity to estimate correctly class 1 areas. The model was validated with a validation dataset, analyzing the influences of the Manning coefficient and the class probability.

The Manning coefficient (roughness) influences the performance of the model. The RF model presents higher F2 scores for the smallest Manning coefficients. This implies the model has a high capacity to estimate correctly risk areas for scenarios with a more likely risk to damage (small roughness coefficient) due to the failure of the off-stream reservoir.

The analysis of the ROC curve and variation of the decision threshold to estimate class 1 areas showed that lowering the decision threshold can enhance the F2 score for class 1, however, this leads to an overestimation of class 1 areas. The selection of the decision thresholds has to be based on the nature of the case. While failures likely to compromise important infrastructure or with larger roughness can be estimated with a smaller decision threshold, in other cases the risk zones can be estimated using the default decision threshold or operating point.

The validity of the method is currently being verified for its application to real off-line reservoirs.

8. ACKNOWLEDGEMENTS

The authors acknowledge financial support from the Spanish Ministery of Science and Innovation through the the Project ACROPOLIS (RTC2019-007343-5).

9. REFERENCES

Albano, R., Mancusi, L., Adamowski, J., Cantisani, A., & Sole, A. (2019). A GIS Tool for Mapping Dam-Break Flood Hazards in Italy. *ISPRS International Journal of Geo-Information*, *8*(6), 250. https://doi.org/10.3390/ijgi8060250

Batista, G. E. A. P. A., Prati, R. C., & Monard, M. C. (2004). A study of the behaviour of several methods for balancing Machine Learning training data. *ACM SIGKDD Explorations Newsletter*, *6*, 21–29. https://doi.org/10.1145/1007730.1007735

Bladé Castellet, E., Cea, L., & Corestein, G. (2014). Numerical modelling of river inundations. *Ingeniería Del Agua*, *18*(1), 71–82. https://doi.org/10.4995/ia.2014.3144

Bladé Castellet, E., Cea, L., Corestein, G., Escolano, E., Puertas, J., Vázquez-Cendón, E., Dolz, J., & Coll, A. (2014). Iber: Herramienta de simulación numérica del flujo en ríos [Iber: Tool for numerical simulation of river flow]. *Revista Internacional de Métodos Numéricos para Cálculo y Diseño en Ingeniería*, *30*(1), 1–10. https://doi.org/10.1016/j.rimni.2012.07.004

Bonaccorso, G. (2017). *Machine Learning Algorithms: A reference guide to popular algorithms for data science and machine learning*. Packt Publishing.

Boström, H. (2007). *Estimating Class Probabilities in Random Forest*. 211–217. https://doi.org/10.1109/ICMLA.2007.64

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cannata, M., & Marzocchi, R. (2012). Two-dimensional dam break flooding simulation: A GIS-embedded approach. *Natural Hazards*, *61*(3), 1143–1159. https://doi.org/10.1007/s11069-011-9974-6

Cea, L., Bermudez, M., Puertas, J., Blade, E., Corestein, G., Escolano, E., Conde, A., Bockelmann-Evans, B., & Ahmadian, R. (2016). IberWQ: new simulation tool for 2D water quality modelling in rivers and shallow estuaries. *Journal of Hydroinformatics*, *18*(5), 816–830. https://doi.org/10.2166/hydro.2016.235

Cea, L., & Bladé, E. (2015). A simple and efficient unstructured finite volume scheme for solving the shallow water equations in overland flow applications. *Water Resources Research*, *51*(7), 5464–5486. https://doi.org/10.1002/2014WR016547

de Cea, C. (2021, May 19). *Evolución de la normativa en seguridad de balsas de riego [Developments in irrigation pond safety regulations]*. Jornada sobre la gestión de la seguridad en balsas de riego online.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced datasets*. Springer.

Hariri-Ardebili, M. A., & Pourkamali-Anaraki, F. (2018). Simplified reliability analysis of multi hazard risk in gravity dams via machine learning techniques. *Archives of Civil and Mechanical Engineering*, *18*(2), 592–610. https://doi.org/10.1016/j.acme.2017.09.003

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction* (Second edition). Springer.

Hooshyaripor, F., Tahershamsi, A., & Behzadian, K. (2015). Estimation of Peak Outflow in Dam Failure Using Neural Network Approach under Uncertainty Analysis. *Water Resources*, *42*, 723–736. https://doi.org/10.1134/S0097807815050085

Hooshyaripor, F., Tahershamsi, A., & Golian, S. (2014). Application of copula method and neural networks for predicting peak outflow from breached embankments. *Journal of Hydro-Environment Research*, *8*(3), 292–303. https://doi.org/10.1016/j.jher.2013.11.004

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transaction on Knowledge and Data Engineering*, *17*(3). https://doi.org/10.1109/TKDE.2005.50

Mckay, M. D., Beckman, R. J., & Conover, W. J. (2000). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics*, *42:1*, 55–61. https://doi.org/10.1080/00401706.2000.10485979

Ministry for the ecological transition, and the demographic challenge. (2021). *Guía Técnica para la clasificación de presas [Technical Guide Dam clasification]*.

Ministry of the presidency. (2008). *Real Decreto 9/2008, de 11 de enero, por el que se modifica el Reglamento del Dominio Público Hidráulico, aprobado por el Real Decreto 849/1986, de 11 de abril [Royal Decree 9/2008, of 11 January, amending the Regulations on the Public Hydraulic Domain, approved by Royal Decree 849/1986, of 11 April.]*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Ruiz-Villanueva, V., Bladé, E., Sánchez-Juny, M., Marti-Cardona, B., Díez-Herrero, A., & Bodoque, J. M. (2014). Two-dimensional numerical modeling of wood transport. *Journal of Hydroinformatics*, *16*(5), 1077–1096. https://doi.org/10.2166/hydro.2014.026

Salazar, F., Conde, A., Irazábal, J., & Vicente, D. J. (2021). Anomaly Detection in Dam Behaviour with Machine Learning Classification Models. *Water*, *13*(17), 2387. https://doi.org/10.3390/w13172387

Salazar, F., Toledo, M. A., Oñate, E., & Moran, R. (2015). An empirical comparison of machine learning techniques for dam behaviour modelling. *Structural Safety, 56*. https://doi.org/10.1016/j.strusafe.2015.05.001

Sanz-Ramos, M., Bladé Castellet, E., Palau Ibars, A., Vericat Querol, D., & Ramos-Fuertes, A. (2019). IberHABITAT: Evaluación de la Idoneidad del Hábitat Físico y del Hábitat Potencial Útil para peces. Aplicación en el río Eume [IberHABITAT: assessment of Physical Habitat Suitability and Weighted Usable Area for fishes. Application in the Eume River]. *Ribagua*, *6*(2), 158–167. https://doi.org/10.1080/23863781.2019.1664273

Sanz-Ramos, M., Bladé, E., González-Escalona, F., Olivares, G., & Aragón-Hernández, J. L. (2021). Interpreting the Manning Roughness Coefficient in Overland Flow Simulations with Coupled Hydrological-Hydraulic Distributed Models. *Water*, *13*(23), 3433. https://doi.org/10.3390/w13233433

Sanz-Ramos, M., Olivares Cerpa, G., & Bladé i Castellet, E. (2019). Metodología para el análisis de rotura de presas con aterramiento mediante simulación con fondo móvil [Methodology for dam failure analysis with grounding using moving bottom simulation]. *Ribagua*, *6*(2), 138–147. https://doi.org/10.1080/23863781.2019.1705198

Stein, M. (1987). Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics*, *29:2*, 143–151. https://doi.org/10.1080/00401706.1987.10488205

Zhang, C., & Ma, Y. (2012). *Ensemble Machine Learning: Methods and applications*. Springer.

