City University of New York (CUNY)

CUNY Academic Works

International Conference on Hydroinformatics

2014

Data-Driven Surrogate Models For Flood Defence Failure Estimation : “Jarillon” Of Cali, Colombia Case Study

Juan Aguilar

Jord Warmink

Ralph Schielen

Suzanne Hulscher

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/cc_conf_hic/291

Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

DATA-DRIVEN SURROGATE MODELS FOR FLOOD DEFENCE FAILURE PROBABILITY ESTIMATION: CASE STUDY "JARILLÓN DE CALI, COLOMBIA"

J.P. AGUILAR LÓPEZ (1), J.J. WARMINK (1), R.M.J. SCHIELEN (1,2), S.J.M.H. HULSCHER(1)
(1): CTW-WEM, University of Twente, Drienerlolaan 5, 7522 NB Enschede,Netherlands
(2): Ministry of Transport, Public Works and Water Management (Rijkswaterstaat)6800 ED Arnhem, the Netherlands.

ABSTRACT
Reliability studies for safety assessment of flood defences now days demand a large amount of stochastic calculations. Therefore mathematical simplifications of the models are used to describe the failure state of the flood defence structures. The present study implemented emulation techniques of different flood defence failure mechanisms, in order to assess the impact in the failure probability by the change in operation of an upstream reservoir. It was found that for the assumed conditions, piping is the most probable failure to occur. However is the less sensitive towards an eventual change in the flow regime conditions. The calculation times where significantly reduced, and the influence in the failure probability distributions was assessed proving data driven models to be a powerful tool for flood defence safety assessment.

INTRODUCTION

Failure mechanisms are one of the main concerns for flood defence designers and managers as they have to be assessed to ensure the stability and functionality of the structure. These deterioration processes are evaluated by mathematical expressions that describe the state of the structure based on system state variables such as water loads, geometrical characteristics, and characteristics of the construction material. These expressions are also known as limit state equations (LSE) which are used to determine the failed or safe condition in a probabilistic way, for each failure mode or failure mechanism. Limit state equations have the general form Z = R-S, where (R) denotes the term of resistance to deterioration and (S) refers to the deterioration driving forces. Each failure mechanism has its own LSE which can result in a negative (unsafe) or positive (safe) value, after evaluating function Z. When a probabilistic approach is adopted, the terms (R) and (S) will be represented by probability distributions. In order to generate these distributions, stochastic procedures such as Monte Carlo can be implemented by the use of numerical models that estimate either the load and/or resistance terms. However, if these numerical models are too complex, the computational burden becomes a great challenge.

Different studies have shown that emulation methods are a feasible solution for the reliability analysis of flood defence structures (Kingston [5]). Therefore, the present study focuses on data driven surrogate model implementation as a tool to reduce the computational burden for the safety assessment of a riverine flood defence.

The main motivation for this study was a necessity to assess the impact of when changing the flow regime in the upstream part of a flood defence system. In order to achieve this goal, different emulation techniques are implemented for both load and resistance terms.

FAILURE MECHANISM DESCRIPTION

Overflow consists in the inflow of water to the protected area due to an extreme water level event that exceeds the height of the flood defence. Commonly this failure mechanism is analyzed in along with the “overtopping” failure. The last one consists in estimating the *water wave heights* originated during an extreme wind/water event that also will eventually erode the hinter part of structure as well.

Piping also known as backwater erosion, consists on the soil internal erosive deterioration of the foundation of the embankment. The erosion is derived from the water movement from the river side towards the inland side of the embankment. The occurrence of this kind of failure doesn’t occur instantaneously but the occurrence of sand boils in the inland side are assumed as a possible failure indication. In order for the erosion process to develop, a previous failure mechanism called “uplift” must occur as well. It consist in the lifting and breakage of the impervious layer above the foundation of the dike due to a high hydrostatic pressure originated by a high water level on the river side of the structure.

Macro stability failure mechanism consists in the displacement of a soil mass which derives the eventual collapse of the structure. This kind of failure occurs whenever the driven forces with respect a point of rotation are higher than the resistance forces with respect to the same point. For the case of earth embankments used for flood protection, the share stresses inside the body may change with the variation of the phreatic table level and the increase of the hydrostatic load.

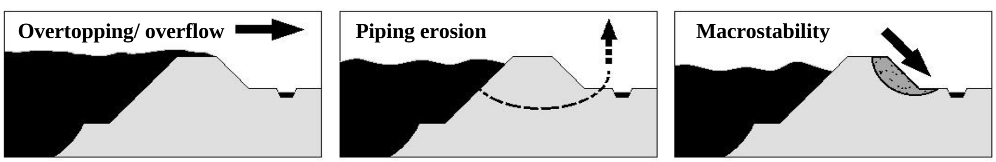

Figure 1. Most common failure mechanisms for embankment flood defences.

DATA-DRIVEN SURROGATE MODELS

Emulation modelling of water systems is a very broad practice as it can vary from a simple linear regression to more complex algorithms such as artificial neural networks Solomatine [8] , M5P model trees Bhattacharya [2], and time varying models Wolfs [9]. Most these studies showed that is possible to emulate the water levels (load term (S)) of a hydrodynamic system with sufficient accuracy while reducing the calculation time and the input requirements. However there is no evidence to the authors that such techniques have been used for the estimation of the water load probability distribution estimation of a river flood defence. For safety assessment studies of this kind of structures is more common to generate extreme value distribution such as Weibull, Gumbel or Log-Pearson III from measured data and then, generate random samples as input for the stochastic procedures. Yet the influence of an upstream operational change in the water level probabilistic distribution in front of the flood defence can also be an interesting measure for flood risk management.

The probability distribution estimation for the resistance term of the limit state equation in most of the different failure mechanisms, is commonly related to the geotechnical and geometrical behavior of the structure during a flooding event. Therefore porous media flow theory has been derived and tested for cases of riverine flood embankments such as piping backward erosion Sellmeijer [7], over topping breach Yu, *et al.* [10] and slope stability Zhiguo, *et al.* [11] for example. Still, most of these models are reduced to mathematical expressions that define the state of the system in order to simplify the safety assessment. In most cases, this simplification might also imply a greater uncertainty than by implementing a numerical model. Still, the accuracy of modelling methods such as the Finite elements where exploited by surrogate modelling in flood defence reliability studies by Rajabalinejad, *et al.* [6]. It showed to be a successful approach to reduce computational burden for reliability estimations. In the present study, the surrogate model approach for resistance terms will focus not only in the reducing calculation time but also in the input simplification. This means that surrogate modelling also allows to convert desired variables into probabilistic distributions that commercial packages might not allow to model as stochastic variables.

CASE STUDY: "JARILLON DEL RIO CAUCA-COLOMBIA"

The Cauca River is the second largest stream in Colombia. It has its origin in the high plateau of Sotara near the main city of Popayán between the central and western Andes chains, in the region known as the "Macizo Colombiano". Its 1350 kilometres extension, drain from South to North until it joins the "Magdalena" river. The city of Santiago de Cali, one of the largest cities in Colombia, was developed along the western side of this river.

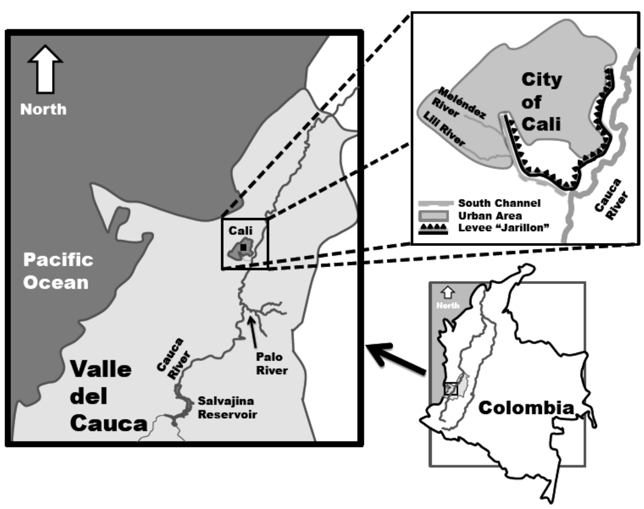

Figure 2. Schematic map of the Cauca river system over the "Valle del Cauca" province.

Since the early 60's, a large earth embankment or also named as "Jarillon", has been built and reinforced in order to protect new urban and agricultural developments that emerged around the city. According to HaskoningDHV [4], almost 20% of the population who has settled inside the river floodplain and are in potential risk of flooding. The "Jarillon" has a 18 kilometer longitudinal dimension and has a 100 year return period average height. 140 Kilometers upstream from the levee, the "Salvajina" reservoir (Figure 2) regulates the water produced by the upper catchment for Hydropower and flood management purposes.

Load term (S) surrogate model

A numerical model was built using the MIKE11 hydrodynamic package. 414 bathymetric cross sections where used to represent the 220 Kilometers that cross the "Valle del Cauca" province (Figure 2). This model include 48 tributaries along the main stream plus an upstream boundary condition which represents the outflow discharges coming from the "Salvajina" reservoir. The previously described model was reduced to 3 main inflow time series able to predict the flowing discharges in 9 different locations along the main stream. M5P decision tree models where trained and validated with the lagged time series for each of the interest locations(Aguilar Lopez, *et al.* [1]). This study proved prediction of discharges with the emulated model was sufficiently good while reducing the computational burden by almost 2 hours. The reduced input consisted in the outflows of "Salvajina" dam, the natural discharge from "Palo" river, and the waters coming by the "South Channel" from the pluvial system of the city of Cali (Figure 2). For the present study, stochastic random sampling of these three input where generated using daily time series for the period 1985-2010. The produced discharge values were routed through the emulated model. The discharge values produced for the location in front of the levee was transformed in water levels using the rating curve from the gauging station located in the same location. These values where compared with the actual measured values in the gauging station located in front of the flood defence (Figure 3).

Figure 3. Probability water load distributions (Observed and Emulated) in the location of existent gauging station "Juanchito", 140 kilometers downstream from the dam.

Resistance term (R)

The estimated water loads are the main driver of the 3 most important previously mentioned failure mechanisms. In some cases in the conceptual form of water heights like for the overflowing failure mechanism, or in other cases in the form of hydrostatic water pressures (Piping and Macro stability).

Overflow resistance term

For the present study, only overflowing mechanism was analyzed, as the probability of a extreme wind event for wave generation is very low. Therefore, the resistance term is assumed as the constant which is equivalent to the height of the "Jarillon" (6 meters). In reality, the uncertainty induced by settlement and compaction of the structure suggest that this term may also be represented as a probability distribution and therefore more complex models could be implemented if needed.

Piping backward erosion

Several models are commonly used for assessing the piping failure mechanisms such as Bligh, Lane and most recent Sellmeijer. The last one, is used for the present case study. According to Sellmeijer [7], the model accounts for the groundwater flow through the subsoil, pipe flow through the erosion channel and a limited particle equilibrium at the bottom of the channel. However for safety assessment, a limit state equation (Eq. (1)) was derived which describes the resistance as a proportion of the seepage length underneath the dike multiplied by 3 different factors. In this case there was no need of model emulation, but the resistance still was represented as probability distribution estimated by stochastic random sampling of the variables from Table 1 .

$$F_G = 0.91\left(\frac{D}{L}\right)^{\left(\frac{0.28}{\left(\frac{D}{L}\right)^{2.8}-1}+0.04\right)} \qquad F_R = \eta \frac{\gamma'_{sand}}{\gamma_w} \tan(\theta) \qquad F_S = \frac{d_{70m}}{\sqrt[3]{\left(\frac{\nu K}{g}\right)L}}\left(\frac{d_{70}}{d_{70m}}\right)^{0.4} \tag{1}$$

$$\mathrm{Hc} = m_p(F_G)\,(F_R)\,(F_S)L$$

Table 1. List of variables for estimating piping based in the Sellmeijer limit state equation

η	[-]	: Sand drag force factor (White's coefficient)
γ'_{sand}	[N/m^3]	: Unitary weight of sand particles
γ_w	[N/m^3]	: Unitary weight of water
θ	[deg.]	: Bedding angle of sand grains
d_{70}	[m.]	: 70 percent quintile value grain size distribution of sand layer
d_{70m}	[m.]	: Calibration reference value (2.08 x 10^{-4} m)
ν	[m^2/s]	: Kinematic viscosity of water at 20 °C
K	[m/s]	: Hydraulic permeability of sand
g	[m/s^2]	: Gravitational acceleration
D	[m.]	: Average thickness of sand layer
m_p	[-]	: Modelling uncertainty factor
FR	[-]	: Resistance factor
FS	[-]	: Scale factor
FG	[-]	: Geometric factor
L	[m.]	: Seepage length from entrance point to sand boil water exit

Macro stability

It was already mentioned that the macro stability of the flood defence can be compromised if the equilibrium of forces in one of the slopes is disturbed or nor correctly balanced. Not only the natural slope of the terrain increases the driving torsional moment but the hydrostatic load and the inner pore pressure fluctuation affect it as well. The resistance term for this failure mechanism is represented as the opposite torsional moment described inside a slip failure surface in the soil. For the present study, the Bishop stability method was chosen as a tool for the safety assessment of this mechanism. The "Jarillon" was modelled using the DGeoStability Software developed by Deltares. The software is capable of calculating the safety factor for different complex geometries for different phreatic conditions. It even allows the user to make probabilistic assessment of failure conditions by allowing to represent an model the materials as probabilistic distributions. Nevertheless it only allows a single hydrostatic condition per simulation when doing a probabilistic assessment. Therefore, an emulation technique was implemented in order to be able to represent the river and phreatic levels as a probabilistic distributions. The emulation consisted in constructing a base model configuration input file of the "Jarillon" flood defence with the mean assumed stochastic parameters. These parameters included , geometric characteristics, soil characteristics and hydrostatic loads.

Next, after assuming all distributions as uniform, 10.000 input files where created doing random sampling via a Matlab® text file generation routine. Afterwards, the 10,000 input files produced before, were used to calculate the safety factor using DGeoStability software.

This factor represents the proportion between the resistant forces and the driving forces. Nota that even though the water load is assumed as a load input to the model, the geometric and geotechnical parameters work as load term as well, as they affect the driving forces implicitly.

The 10,000 generated samples are used to train a neural network. The results of the model emulation are shown in Figure 4. A good correlation was achieved, and sufficient combinations related to different safety factors are obtained. One of the biggest challenges in reliability studies is to be able to generate random samples that represent low probability of occurrence events. For our case, a sufficient exploration of the failure region (S.F < 1) is represented by the model which indicates its suitable for probabilistic failure estimation.

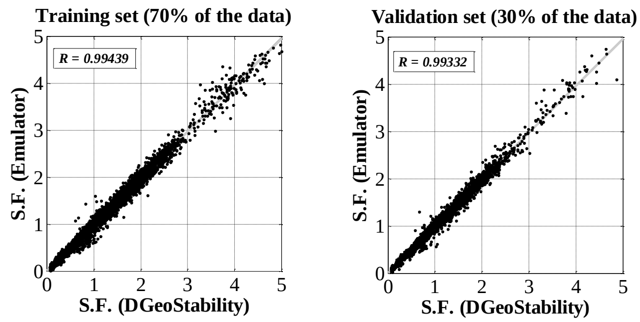

Figure 4. Artificial neural network emulator training, validation and test results for Safety factor prediction of "Jarillon" flood defence embankment.

During the training data calculation for the macro stability emulator it was observed that the calculation time for 500 samples was near 6 minutes. This mean that even with a simple method such as Bishop, a crude monte carlo simulation (at least 1,000,000 of samples required to ensure a low estimation error) will take around 200 hours. The present artificial neural network emulator takes 4.3 seconds to calculate 1,000,000 samples.

RESULTS

The first part of the study consisted in estimating the impact in the resultant probability distribution of the load term (S) after modifying the reservoir outflow discharges. This was achieved by affecting each discharge released from the dam by 25%. Afterwards, the obtained values were routed while generating random samples for the other two inflow tributaries of the emulated model (Palo and South channel).

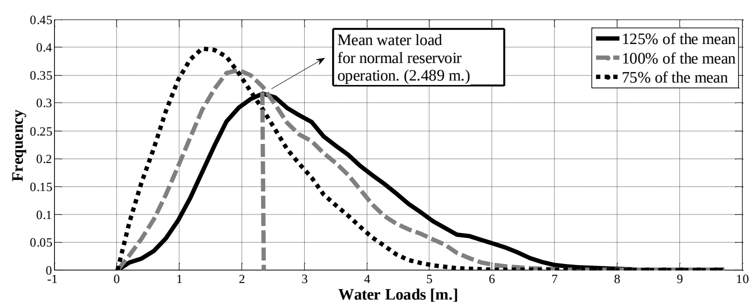

Figure 5. Experimental probability distributions of water loads in front of the flood defence location produced by the emulated model based in reservoir discharges.

It can also be observed that because of the increase of high level discharges for the 125% change in the mean operation of the reservoir, the shape of the pdf is less smooth. This can be attributed to the behavior of the emulator which implicitly considers the hysteresis of the rating curve when unsteady flow is recreated in the original Mike11 model.

The second part of the experiment consisted in estimating the resistant resultant probability distributions for each of the three failure mechanisms (Z) obtained for the different distributions produced in the first part of the experiment (Figure 5). For the case of overflowing, the probability can be estimated directly from the empirical load probability density function (Figure 5) as the resistance term is assumed as a constant value. The embankment is assumed to have 6 meters of average height from the bed bottom of the river to the crest.

The results show that the change in the operation of the reservoir has influences the mean value and shape for macro stability and piping. This results should be interpreted with care as when analyzing safety factors , many different combination of stochastic load and resistance values can result in similar safety factors. However if the failure function behaves monotonically, it can be assumed that a safety factor can be associated with a probability of occurrence (Ching [3]).

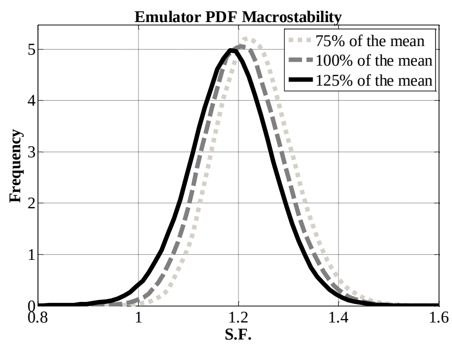

Figure 6. Experimental probability density for the estimated safety factor of Macro stability failure mechanism.

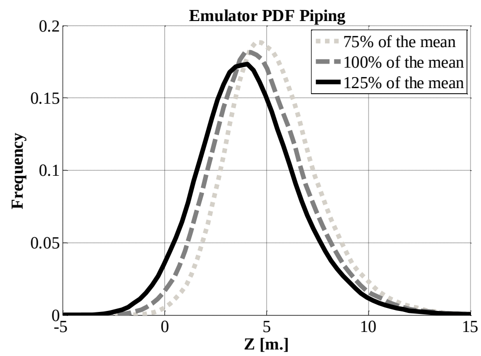

Figure 7. Experimental probability density function (Z) for piping failure mechanism.

The present structure is estimated to have an equivalent height to Tr =10 years + 2.50 Freeboard (HaskoningDHV [4]). For the present study, the freeboard is not taken into account for the failure estimation presented in Table 2. However in order to check that the present study goes along with the order of magnitude of the study previously cited, it is estimated that with a freeboard of 2.5 meters, the return period of the flood defence using the emulator is 5 years.

Table 2. Failure return periods for the eventual change in the reservoir operation

	Overflow (Tr)	**Piping (Tr)**	**Macrostability (Tr)**
75% of the mean	10,000	385	1257
100% of the mean	227	80	270
125% of the mean	31	29	57

The difference can be explained by two main reasons. The first one is that the levels generated from the emulator, don’t have the same probabilistic distribution as the one used in the HaskoningDHV study (Log-Pearson III). The second is that the levels used in the present study are generated stochastically whereas the levels used for fitting the distribution by HaskoningDHV are the real ones observed in the gauging station. Therefore different sources of uncertainty can be identified.

CONCLUSIONS

For the present study, the application of emulation techniques where implemented in both load and resistance terms of the different limit state equations. The calculation burden time was reduced significantly, in the case of Macro stability while considering an additional variable such as the pore pressure inside the body of the embankment. It also proved to be a useful tool for assessing the impact from the upstream reservoir release modification. Nevertheless, such methodology should be implemented with care as the produced emulators are only as good as the original model. If the training data generation is not sufficiently representative, the produced emulators will not represent correctly and eventual combination outside the training feasible space. The present case study was done by using average recommended values for the random sampling of the parameters. A more detailed experiment design is recommended for the original model data generation in order to ensure the correct representation of extreme events. Therefore the obtained numerical results and the resultant failure probabilities shouldn't be assumed valid for the "Jarillon" actual embankment. In terms of the study motivation (impact of dam operation change), it was shown that piping is always the failure mechanism most probable to happen for this case study configuration. However is the less sensitive of the failure mechanisms when changing the reservoir operation. The emulation techniques are a powerful method for linear large flood defences where several stochastic calculations are required per representative cross section before estimating the total probability of failure.

ACKNOWLEDGEMENTS

This work is part of the research program of the Technology Foundation STW, financially supported by the Netherlands Organization for Scientific Research (NWO). The authors would like to thank STW foundation, The MFFD research group, CVC environmental authorities in Colombia and The Cinara research institute in Colombia.

REFERENCES

[1] Aguilar Lopez, J. P., Van Andel, S. J., Werner, M., and Solomatine, D. P. (2014). "Hydrodynamic and water quality surrogate modelling for reservoir operation " *11th IWA/IAHR International Conference on Hydroinformatics* New York City, USA.

[2] Bhattacharya, B. S., D. P. (2005). "Neural networks and M5 model trees in modelling water level–discharge relationship." *Neurocomputing*, 63, 381-396.

[3] Ching, J. (2009). "Equivalence between reliability and factor of safety." *Probabilistic Engineering Mechanics*, 24(2), 159-171.

[4] HaskoningDHV, R. (2013). "Dique de Aguablanca a lo largo del río Cauca, Cali colombia."Colombia.

[5] Kingston, G. B. R., Mohammadreza; Gouldby, Ben P. ;Van Gelder, Pieter H. A. J. M. (2011). "Computational intelligence methods for the efficient reliability analysis of complex flood defence structures." *Structural Safety*, 33(1), 64-73.

[6] Rajabalinejad, M., Mahdi, T., and Gelder, P. (2010). "Stochastic methods for safety assessment of the flood defense system in the Scheldt Estuary of the Netherlands." *Natural Hazards*, 55(1), 123-144.

[7] Sellmeijer, H. d. l. C., Juliana López; van Beek, Vera M. Knoeff, Han; (2011). "Fine-tuning of the backward erosion piping model through small-scale, medium-scale and IJkdijk experiments." *European Journal of Environmental and Civil Engineering*, 15(8), 1139-1154.

[8] Solomatine, D. P. A. T. L. A. "Neural Network Approximation of a Hydrodynamic Model in Optimizing Reservoir Operation." *Proc., International Conference on Hydroinformatics*, Taylor & Francis, 500.

[9] Wolfs, V. W., Patrick; (2014). "Development of discharge-stage curves affected by hysteresis using time varying models, model trees and neural networks." *Environmental Modelling & Software*, 55(0), 107-119.

[10] Yu, M.-h., Wei, H.-y., Liang, Y.-j., and Zhao, Y. (2013). "Investigation of non-cohesive levee breach by overtopping flow." *Journal of Hydrodynamics, Ser. B*, 25(4), 572-579.

[11] Zhiguo, N., ri, Y., and Jun, L. (2012). "The Dike Stability Analysis based on Strength Reduction Method." *Procedia Engineering*, 28(0), 560-563.