Article

Tailings Dams Failures: Updated Statistical Model for Discharge Volume and Runout

Paulina Concha Larrauri [1,*] and Upmanu Lall [1,2]

[1] Columbia Water Center, Columbia University, 842 S.W. Mudd, 500 West 120th Street, New York, NY 10027, USA; ula2@columbia.edu

[2] Department of Earth and Environmental Engineering, Columbia University, 500 W. 120th St., 918 Mudd, New York, NY 10027, USA

[*] Correspondence: pc2521@columbia.edu; Tel.: +1-212-854-1695

Abstract: This paper presents a statistical model to estimate the volume of released tailings (V_F) and the maximum distance travelled by the tailings (D_{max}) in the event of a tailings dam failure, based on physical parameters of the dams. The dataset of historical tailings dam failures is updated from the one used by Rico et al., (Floods from tailings dam failures, Journal of Hazardous Materials, 154 (1) (2008) 79–87) for their regression model. It includes events out of the range of the dams contained in the previous dataset. A new linear regression model for the calculation of D_{max}, which considers the potential energy associated with the released volume is proposed. A reduction in the uncertainty in the estimation of D_{max} when large tailings dam failures are evaluated, is demonstrated. Since site conditions vary significantly it is important to directly consider the uncertainty associated with such predictions, rather than directly using these types of regression equations. Here, we formally quantify the uncertainty distribution for the conditional estimation of V_F and D_{max}, given tailings dam attributes, and advocate its use to better represent the tailings dam failure data and to characterize the risk associated with a potential failure.

Keywords: tailings storage facilities; dam break; risk analysis; mining waste

1. Introduction

The failure of tailings storage facilities can have disastrous consequences for nearby communities, the environment, and for the mining companies, who may consequently face high financial and reputational costs. Tailings are waste resulting from mining operations and are commonly deposited as slurry behind earthen or masonry dams. We refer to this form of tailing storage facility as TSF. In 2015, the breach of the Fundão TSF at Samarco mine in Minas Gerais (jointly owned by BHP Billiton Brasil and Vale S.A.) resulted in 19 fatalities, and was declared the worst environmental disaster in Brazil's history. The company entered an agreement with the Federal Government of Brazil and other public authorities to remediate and compensate for the impacts over a 15 years period. Jointly, BHP and Vale recognized a US$ 2.4 billion provision for potential obligations under the agreement [1,2]. Twenty-one company executives were charged with qualified murder, and up until September 2017 the mine had not resumed operations. The ecosystems impacts caused by a TSF failure can last for many years depending on the nature of the tailings. Samarco is in the process of restoring 5000 streams, 16,000 hectares of Permanent Conservation Areas along the Doce River basin, and 1200 hectares in the riverbanks [3]. It is estimated that the livelihoods of more than 1 million people were affected because of the failure [4]. Improvements in the design, monitoring, management, and risk analysis of TSFs are needed to prevent future failures and to estimate the consequences of a breach.

The design of tailings dams has changed significantly from the 1930s to the present [5–7]. Construction of the early TSFs was done by trial and error [8]. During the 1960's and 1970's geomechanical engineering started to be used to assess the behavior of the tailings and the stability of the impoundments [8]. Currently, various studies are required to approve a TSF design and increasingly the plans for remediation and closure of the impoundments have to be included in the feasibility phase. Breach assessments are now part of the requirements in the permitting process of a new TSF or an expansion in many countries. Different parameters need to be estimated while conducting these assessments [9]. These include the volume of tailings (V_F) that could potentially be released, and the distance to which the material may travel in a downstream channel, called the run-out distance (D_{max}). Empirical regression equations for this purpose were developed by Rico et al. [10] using historical TSF failure data, and are commonly used to characterize such failures (similar empirical relationships have been developed for dams holding water [11,12], but the lack of tailings and differences in design and construction make them inapplicable to tailings dams). However, at site conditions in the mines can vary substantially and there is considerable residual uncertainty associated with the conditional mean value estimated by these equations. In this paper, we rigorously update these regression equations using an updated data set, and characterize the uncertainty associated with the prediction. Using the uncertainty distribution for the conditional estimation of V_F and D_{max} using TSF parameters provides a better way to interpret the TSF failure data and to characterize the risk associated with a potential failure.

The calculation of V_F is of particular importance for inundation analyses. Typically, TSFs are not totally emptied in case of failure (as opposed to water dams), and only a portion of the tailings are released [10]. In TSFs containing a large amount of water (supernatant pond), the breach would usually result in an initial flood wave followed by mobilized/liquefied tailings [9]. Therefore, the methods developed to estimate the released volume of water or the inundation extent from a regular dam (such as the water dam break-flood analyses methods in [13, 14]), do not apply to tailings dams. Empirical equations based on past failures, dam height, and the impounded volume of tailings, are commonly used to get a first estimate of the volume of tailings that could be released and the run-out distance. In Rico et al. [10] V_F is calculated using the total impounded volume (V_T) in Mm³ as in Equation (1)

$$V_F = 0.354 \times V_T{}^{1.01} \quad R^2 = 0.86 \tag{1}$$

and the outflow run-out distance travelled by the tailings in km (D_{max}) is obtained using V_F and the dam's height in meters at the time of failure (H) as in Equation (2)

$$D_{max} = 1.61 \times (H\,V_F)^{0.66} \quad R^2 = 0.57 \tag{2}$$

Many investigators directly use such regression equations in a deterministic way to specify exposure. However, at site conditions vary significantly, and there is considerable uncertainty that needs to be quantified. This uncertainty increases as we consider TSF volumes that are near or beyond the range of the data included in the regression equation. Equation (1) predicts that approximately a third of the tailings in the impoundment (including water) will be the outflow volume. This approach may result in unrealistic estimates when liquefaction is a known risk as it does not take into account the tailings mass rheology (viscosity and yield stress) [9]. As Rico et al. [10] point out, some parameters contributing to the uncertainty in the predictions include sediment load, fluid behavior (depending on the type of failure), topography, the presence of obstacles stopping the flow, and the proportion of water stored in the tailings dam (linked to meteorological events or not). Therefore, it is important to account for the uncertainty in these estimates to derive a probabilistic measure of risk that also accounts for how well the regression fits in a certain range of values of the predictors.

Additional information about TSF failures since Rico et al. [10] published the above-mentioned equations is available. In this paper we update the original data used by Rico from 22 complete cases (including height, storage volume in m³, released volume in m³, and distance traveled) to 29 complete cases with data compiled in Chambers and Bowker [15]. We compare the results of the

original linear regressions done by Rico et al. [10] with the results using the updated dataset. A new model for the calculation of D_{max} is proposed introducing the predictor (H_f), which is defined as:

$$H_f = H \left(\frac{V_F}{V_T} \right) V_F \tag{3}$$

This variable was introduced to consider that the potential energy associated with the release volume, may be better related to the fractional volume released as opposed to the total volume of the TSF.

For each of the models we consider, and for the final model we recommend, we consider the uncertainty analysis of prediction. We compare the predicted intervals and observed values of V_F and D_{max} of three TSF failures across the models that were evaluated to see how well the prediction intervals fit the observed data. The indicated probability of exceedance of the observation as per each model was also assessed.

2. Materials and Methods

The height at time of failure (H), TSF capacity (V_T), released volume (V_F), and the distance traveled by the tailings (D_{max}) are inputs in Rico's equations (Equation (1) and Equation (2)). The data used in this analysis is a combination of the cases used by Rico and others compiled by Chambers and Bowker [15], including failures post 2008. Seven of the 29 incidents used by Rico et al. [10] do not have complete information or the information for volume is included in million tons, which cannot be used in the analysis without density data. These cases are in red letters in Table 1. It is important to note that the original data did not include releases as large as the ones experienced in Samarco and Mt Polley (cases 15 and 19 in Table 1), and this becomes relevant for future estimations of the potential risk of larger TSFs. The data on reported failures have variations in different sources; some of these variations are included as footnotes in Table 1.

Table 1. Data in Rico et al. [10] and Chambers and Bowker (CB) [15]. Entries in red are incomplete.

No	Mine	Year	H (m)	V_T (×10^6 m³)	D_{max} (km)	V_F (x 10^6 m³)	Failure Type [a]	Source
1	(unidentified), Southwestern USA	1973	43	0.5	25	0.17	SI	Rico
2	Aitik mine, Sweden (Boliden Ltd.)	2000	15	15	5.2	1.8	ER	CB
3	**Arcturus (Zimbawe)**	**1978**	**25**	**1.7–2 Mt**	**0.3**	**0.0211** [b]	**OT**	**Rico**
4	Bafokeng, South Africa	1974	20	13 [c]	45	3	SE	Rico
5	Balka Chuficheva, Russia	1981	25	27	1.3	3.5	SI	CB
6	Bellavista, Chile	1965	20	0.45	0.8	0.07	EQ	Rico
7	Bonsal, North Carolina, USA	1985	6	0.038	0.8	0.011	OT	CB
8	Cerro Negro No. (3 of 5)	1965	20	0.5	5	0.085	EQ	Rico
9	Cerro Negro No. (4 of 5)	1985	40	2	8	0.5	EQ	Rico
10	Churchrock, New Mexico, United Nuclear	1979	11	0.37	110 [d]	0.37	FN	Rico/CB
11	Cities Service, Fort Meade, Florida	1971	15	12.34	120	9	SE	Rico

12	Deneen Mica Yancey County, North Carolina, USA	1974	18	0.3	0.03	0.038	SI	CB
13	El Cobre New Dam	1965	19	0.35	12	0.35	EQ	CB
14	El Cobre Old Dam	1965	35	4.25	12	1.9	EQ	Rico
15	Fundão-Santarem, Minas Gerais, Brazil (Samarco)	2015	90	55	637	32 [e]	ST	CB
16	**Galena Mine (1974)**	**1974**	**9**	**NA**	**0.61**	**0.0038**	**OT**	**Rico**
17	**Gypsum Tailings Dam (Texas, USA)**	**1966**	**11**	**7 [f]**	**0.3**	**0.085**	**SE**	**CB**
18	Hokkaido, Japan	1968	12	0.3	0.15	0.09	EQ	Rico
19	Imperial Metals, Mt Polley, British Columbia, Canada	2014	40	74	7	23.6	FN	CB
20	**Itabirito (Brazil)**	**1986**	**30**	**NA**	**12**	**0.1**	**ST**	**Rico**
21	**La Patagua New Dam (Chile)**	**1965**	**15**	**NA**	**5**	**0.035**	**EQ**	**Rico**
22	Los Frailes, near Seville, Spain (Boliden Ltd.)	1998	27	15	41	6.8 [g]	FN	CB
23	Los Maquis No. 3	1965	15	0.043	5	0.021	EQ	Rico
24	Merriespruit, South Africa (Harmony)-No. 4A Tailings Complex	1994	31	7.04	4 [h]	0.6 [h]	OT	CB
25	Mochikoshi No. 1, Japan (1 of 2)	1978	28	0.48	8	0.08	EQ	Rico
26	**Mochikoshi No. 2 (Japan)**	**1978**	**19**	**NA**	**0.15**	**0.003**	**EQ**	**Rico**
27	Olinghouse, Nevada, USA	1985	5	0.12	1.5	0.025	SE	Rico
28	Omai Mine, No. 1, 2, Guyana (Cambior)	1995	44	5.25	80	4.2	ER	Rico
29	Prestavel Mine-Stava, North Italy, 2, 3 (Prealpi Mineraria)	1985	29.5	0.3	8 [i]	0.2	SI	Rico
30	Sgurigrad, Bulgaria	1996	45	1.52	6	0.22	SI	Rico
31	Stancil, Maryland, USA	1989	9	0.074	0.1	0.038	SI	Rico
32	Taoshi, Linfen City, Shanxi province, China (Tahsan Mining Co.)	2008	50.7	0.29	2.5	0.19	U	CB
33	**Tapo Canyon (USA)**	**1994**	**24**	**NA**	**0.18**	**NA**	**EQ**	**Rico**
34	Tyrone, New Mexico (Phelps Dodge)	1980	66	2.5	8	2	SI	Rico
35	Veta de Agua (Chile)	1985	24	0.7	5	0.28	EQ	Rico

[a] SI = Slope instability, EQ = Earthquake, OT = Overtopping, ER = Erosion, FN = Foundation, SE = Seepage, U = Undefined. [b] CB report 0.039×10^6 m^3 V_F. [c] In Rico 13×10^6 m^3 as [16] and [17]; in CB $17 \times$

10^6 m^3. [d] In CB 110 km as in [18] and [17], in Rico 96.5–112.6 km. [e] 43×10^6 m^3 in [19]. [f] Rico in tones. [g] In Rico 4.6×10^6 m^3; in [20] 10^6 m^3, in [21] 5.5×10^6 m^3 of tailings and 1.9×10^6 m^3 of acid water; in [17] 5.5 Mm^3. [h] In Rico 2.5 Mt V_F and 2 km in D_{max} as in [22]; in [5] 0.6×10^6 m^3. [i] In Rico 4.2 km D_{max}; in [13] 8 km.

Figure 1 shows the relationships between V_F and V_T, and D_{max} with $H \times V_F$ (called dam factor in [10]) using the updated dataset (plots in log scale), and D_{max} with H_f, (Equation (3)). V_F and V_T show a linear relationship in the log form, while for D_{max}, there is greater dispersion with the dam factor and H_f.
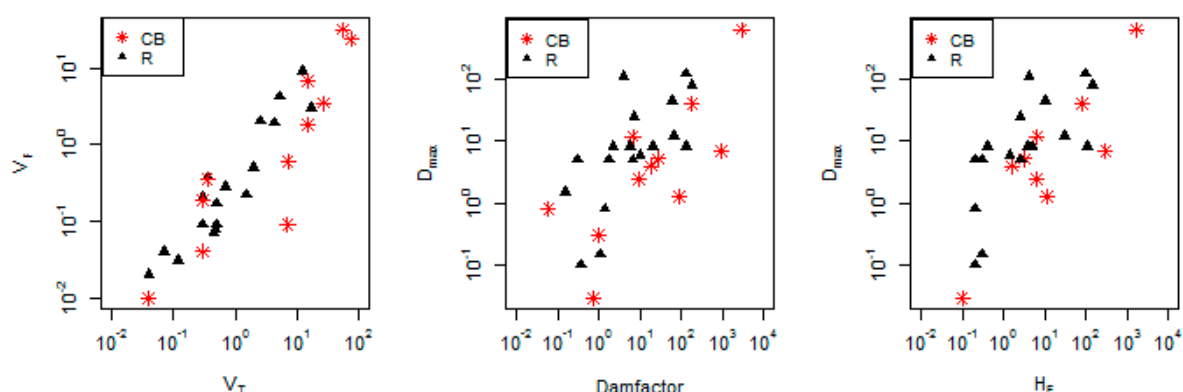

Figure 1 Left: Relationship between V_F and V_T in $\times 10^6$ m^3. Center: D_{max} in km in relation to the dam factor ($H \times V_F$). Right: D_{max} in km in relation to H_f ($H \times (V_F/V_T) \times V_F$). All plots are in the log-log scale. Note that the D_{max} vs H_f plot is much tighter than the D_{max} vs the dam factor plot. CB are added points from Chambers and Bowker [15] and R are from Rico [10].

V_F was estimated in the same way as in Rico et al. [10] using the predictor V_T with a log-log (power) transformation and the updated data. For the estimation of D_{max} three models were considered:

1. A model titled D_{max}.1 which is similar in functional form to the one used by Rico et al. [10] (D_{max}.1 in Table 2).
2. A generalized linear model (glm) with the Gaussian family using a log link function (D_{max}.2 in Table 2).
3. A model D_{max}.3 which uses the new predictor H_f.

Table 2. Fitted models.

Name	Model
V_F.1	$\log(V_F) \sim \log(V_T)$
D_{max}.1	$\log(D_{max}) \sim \log(H \times V_F)$
D_{max}.2	$D_{max} \sim \log(H \times V_F)$, glm, Gaussian, log link
D_{max}.3	$\log(D_{max}) \sim \log(H_f)$

The observed value of V_F was used to fit the D_{max} regressions.

The goodness of fit of each model was analyzed with residual plots, outlier identification, analysis of influential observations using Cook's distance, and computing the root mean square error(RMSE) using a 5-fold cross validation (CV). The prediction intervals and probability of occurrence of V_F and D_{max} in three historical failures was compared across models using the original and the updated datasets.

3. Results and Discussion

3.1. Released Volume of Tailings

The presence of the new data points in the updated dataset has an effect in the uncertainty of V_F.1 as seen in the R^2, standard error and 5-fold cross validation results in Table 3. In Samarco, 58% of the tailings contained at the Fundão dam were released (P.15 in Table 1), whereas in the incident of the Gypsum tailings dam in Texas only 1.2% of the contained tailings were released (P.17 in Table 1). The Gypsum tailings dam incident (P.17) was not included in the updated regression (Equation (4)) because it was a minor release, different from the characteristics of the rest of the dataset (identified as an outlier with high influence, Table 3), and had a strong effect in the normality of the residuals.

Table 3. Results for V_F.1 with the original Rico dataset (O) and updated (U) datasets (including P.17).

Data	R^2	p-Value	Standard Error	5-fold CV, 100 reps	Outliers	Leverage	Cook's Distance
O	0.87	$1.209e^{-9}$	0.288	3.3		P.11, P.12	P.9
U	0.815	$1.285e^{-11}$	0.402	11.5	P.17	P.19, P.15	P.17

The final regression equation with the updated dataset excluding P.17 is,

$$\log(V_F) = -0.477 + 0.954 \log(V_T) \tag{4}$$

R^2 = 0.887; standard error: 0.315

Using Rico's notation, this may be expressed as $V_F = 0.332 \times V_T^{0.95}$. However, since the model is fit in terms of $\log(V_F)$ it is important to consider the uncertainty of prediction in terms of $\log(V_F)$, and then transform the prediction intervals to real space to determine the proper uncertainty intervals for V_F. Tests for the residuals from the fit provided by Equation (4) indicate that a Gaussian distribution cannot be rejected for the residuals at the 5% level (Shapiro-Wilk test p-value = 0.1161). The prediction intervals are then computed at the desired significance level for $\log(V_F)$ and then transformed to real space for V_F.

Table 4 shows the prediction of V_F and 90th prediction interval for Samarco, Mt. Polley and the Gypsum TSF using Equation (4). These cases were selected based on the influence they have in the regression of V_F with the updated data (Table 3). The prediction intervals are wider and the predicted mean is larger for Samarco and Mt. Polley when the original dataset is used. In the case of the Gypsum tailings dam, the observed value is not within the prediction interval in neither regression because the volume released was so small and the probability of exceeding it was very high (the same was observed even when that data point was included in the regression).

As mentioned in the introduction, the volume of released tailings will vary greatly depending on the type of failure and the composition of the tailings but for a first estimate of potential damage, the scenarios within the prediction interval are useful to assess a range of potential consequences. The upper prediction interval for V_F should be constrained by the total volume of tailings available since in all the cases presented in Table 4, the 95th percentile is more than what is physically possible (more than the total impounded are released). In this case, finding the probability associated with totally emptying the dam would be a better approach for risk estimation. We developed an online pp that has the capability of computing the probability of exceeding a value of V_F specified by the user (available at https://columbiawater.shinyapps.io/ShinyappRicoRedo/). In this manner, the uncertainty around V_F can be considered when estimating D_{max}. The app also provides Q5, Q50 and Q75 of V_F. From Table 1, P.10, P.13, P.28, and P.29 were nearly or totally emptied.

Table 4. Predicted and observed V_F values in $\times 10^6$ m^3 for select cases (using all the points in the original Rico dataset (O) and updated (U) data sets except P.17 in U).

Data	Points	Median Prediction *	Q5	Q95	Observed	Probability $V_F >$ Observed	Probability $V_F > V_T$
O	Samarco (P.15)	19.8	5.2	75.7	32	0.27	0.1
	Mt Polley (P.19)	26.8	6.8	104.8	23.6	0.56	0.1
	Gympsum (P.17)	2.42	0.71	8.2	0.09	0.99	0.07
U	Samarco (P.15)	15.2	4.1	57.3	32	0.17	0.05
	Mt Polley (P.19)	20.3	5.3	76.9	23.6	0.42	0.05
	Gypsum (P.17)	2.13	0.6	7.6	0.09	0.99	0.08

* This corresponds to the mean prediction in the log-space, which is not the mean when it is back transformed to the real space. For further explanation review [24].

3.2. Run-Out Distance (D_{max})

The analysis of D_{max}.1 (the original model by Rico et al. [10]) using the updated and original datasets, shows that the uncertainty increases when the new data points are introduced; R^2 is reduced and the cross validated error increases (Table 5). The Samarco failure is an influential observation in all the D_{max} regressions except for D_{max}.3 (Table 5). The distance traveled by the tailings reached 637 km, although more than 90% of the tailings stayed within 120 Km of the dam, the rest were transported in the Doce river all the way to the Atlantic Ocean [15]. The Bonsal TSF (P.7) is also identified as an influential observation (Table 5), this incident had a large D_{max} (close to 1 km) compared to the released volume of tailings (the released volume of tailings was approximately 0.01 m^3). The failure mode reported at P.7 was overtopping so it is likely that the tailings dam had a large proportion of water at the time of failure. P.12 also appears as an influential observation; in that case the distance traveled by the tailings was only 30 meters, which is small considering the dam height (18 m) and the released volume of tailings (0.038 Mm^3).

Based on the results from Table 5 (R^2, and 5-fold CV) and the analysis of the residual plots, the best model found was D_{max}.3, which uses the new predictor H_f and has the form:

$$\log(D_{max}) = 0.484 + 0.545 \log(H_f) \tag{5}$$

Residual standard error = 0.658. Using Rico's notation, this may be expressed as $D_{max} = 3.04 \times H_f^{0.545}$

Table 5. Results of the fitted models for D_{max}.

Model	Data *	R^2	p-Value	5-Fold CV, 100 reps	Significant Outliers	Leverage	Cook's Distance
D_{max}.1	U	0.44	5.335e^{-5}	249.5	P.12	P.7, P.15	P.19, P.12
D_{max}.1	O	0.55	5.39e^{-6}	54.5		P.12	P.12, P.28
D_{max}.2	U	NA	NA	284.9	P.11, P.15, P.19	P.15, P.7	P.15
D_{max}.3	U	0.53	4.415e^{-6}	230.3	P.12	P.7, P.15	P.7, P.12, P.19

* U = updated dataset, O = original dataset.

Table 6 includes the results of the prediction and prediction intervals using the original model (D_{max}.1) with the original and the updated datasets, and includes the results of D_{max}.3 for three of the influential observations. The uncertainty distributions are obtained as before by considering the residuals associated with Equation (5), testing for Gaussian structure (Shapiro-Wilk test *p*-value = 0.388), and then computing the prediction intervals at the desired significance levels, and transforming them to real space. Figure 2 has examples of the prediction intervals obtained from D_{max}.1 O and D_{max}.3 U compared to the observations. From the results in Table 6 and Figure 2 is evident that the additional data used to fit the regressions of D_{max} dramatically reduces the uncertainty bounds in the prediction of large events such as Samarco and Mt. Polley, although in smaller D_{max} events such as Bonsal, Los Frailes or Omai, the uncertainty is similar or it increases. The probability of exceeding the observed run-out distance shown in Table 6 was calculated transforming the observation to the log space, and evaluating its location in the distribution associated to the prediction interval (*t* distribution). This is exemplified in Figure 3.

Table 6. Predicted values (in km) using all the data points for training the models (using the observed V_F).

Model	Points	Median Prediction *	Q5	Q95	Observed	Probability D_{max} > Observed
D_{max}.1	Samarco (P.15)	294	19	4595	637	0.3
O	Mt Polley (P.19)	141	10	2012	7	0.96
	Bonsal (P.7)	0.25	0.02	3.25	8	0.014
D_{max}.1	Samarco (P.15)	141	6	3130	637	0.21
U	Mt Polley (P.19)	74	3.6	1525	7	0.9
	Bonsal (P.7)	0.28	0.01	6	0.8	0.23
D_{max}.3	Samarco (P.15)	174	10	2933	637	0.22
U	Mt Polley (P.19)	68	4	1054	7	0.92
	Bonsal (P.7)	0.3	0.02	5	0.8	0.28

* This corresponds to the mean prediction in the log-space, which is not the mean when it is back transformed to the real space. For further explanation review [24].
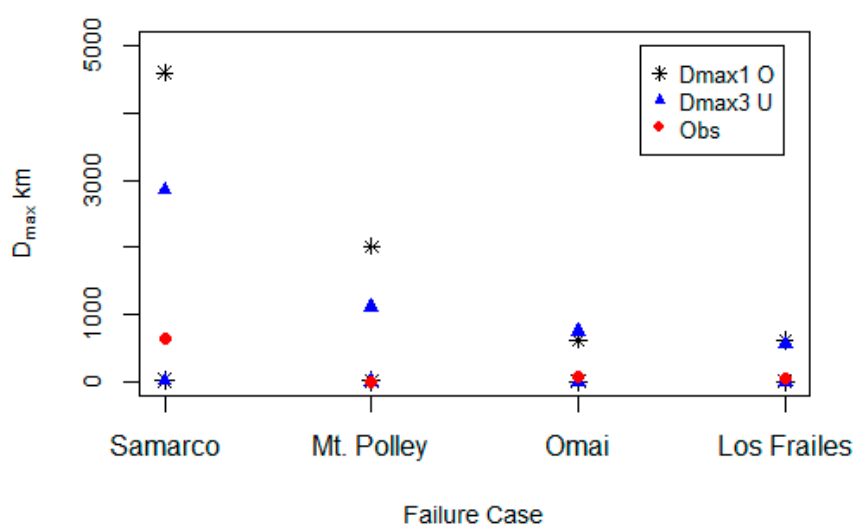

Figure 2. Examples of prediction intervals from D_{max}. 1 O (Rico's original equation), D_{max}.3 U and the observations for D_{max} of past failures (obs).

The importance of considering the uncertainty distribution around the regression of D_{max}, rather than using the conditional mean directly is illustrated by the examples in Table 6. For the Samarco incident, the mean value of the predicted D_{max} is 174 km using D_{max}.3, while the predicted 5th (95th)

percentile is 10 (2933) km. The D_{max} reported from the actual failure was 637 km downstream, which based on the uncertainty distribution associated with the regression equation, has a probability of exceedance of approximately 22% using model D_{max}.3. In this case the tailings were deposited directly in the Doce River [4], transporting the tailings all the way to the Atlantic Ocean, whereas for other TSF failures an immediate river receptor may not be there, limiting the travel distance. Consequently, if just the conditional mean of the regression equations such as those developed by Rico et al. [10] is used, then one would be rather poorly informed as to the range of potential consequences of a failure. For a probabilistic risk evaluation then, for Samarco, the concern would have been the greater than 600 km impact with a 22% chance rather than the modest 174 km indicated by the regression. It is important to highlight that the observed V_F was used to fit all the D_{max} regressions but in reality this value might not be known prior to a failure. Therefore, the uncertainty of the estimation of V_F from Equation (4) has to also be considered in the predictions of D_{max}, which will increase the uncertainty. This is also an issue with Rico's approach reported in Equations (1) and (2), and it is acknowledged in their paper. In the online app we developed, this is addressed and D_{max} can be calculated taking into account the uncertainty around V_F in a two-step model.
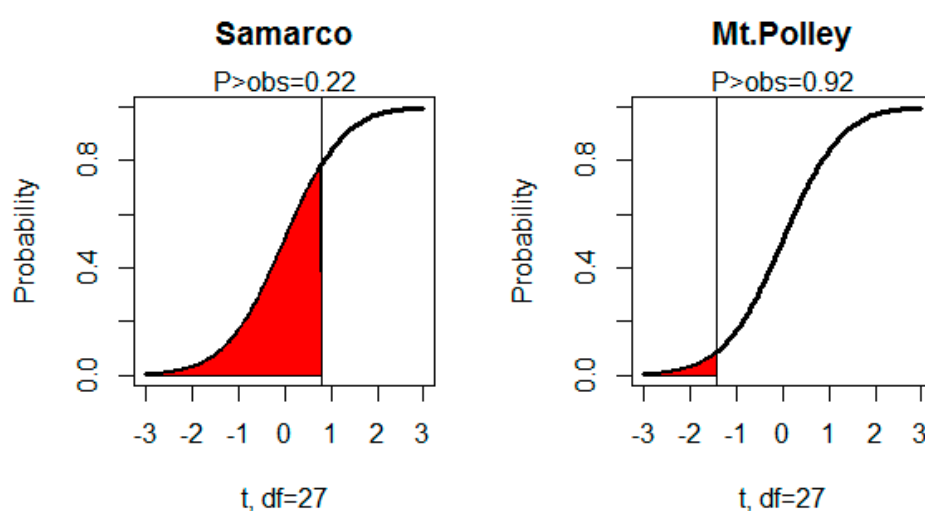

Figure 2. *t* distributions showing the probability of exceedance of the Samarco and Mt. Polley observed D_{max} and the distributions obtained with model D_{max}.3 around the predicted mean value of D_{max} in the log space.

4. Conclusions

The empirical equations developed by Rico et al. [10] to estimate the volume of tailings released in a tailings dam failure and the run-out distance of the tailings were reviewed. An updated dataset provided information on dam failures that happened after the Rico et al. [10] paper was published and includes cases of dams with larger storage capacity and height than the points in the original dataset. The introduction of the new data points in the regression reduces the uncertainty of the prediction of large failure incidents such as the one occurred in Samarco.

An improved model to estimate the run-out distance is proposed. The model uses the predictor H_f that considers the potential energy associated with the released volume as opposed to the whole tailings impoundment volume. The model proposed has a better linear fit than the original model when using the updated dataset. The updated model to calculate V_F is presented in Equation (4), and the new model to estimate D_{max} in Equation (5). We recommend using the app we provide which contains the equations (available at https://columbiawater.shinyapps.io/ShinyappRicoRedo/). Since we recommend using the uncertainty distribution for each "prediction" it is easiest for the user to use our web app. As data on other failures becomes available, it can be brought into the app and the model can then be automatically updated.

This paper emphasizes that these are empirical regression equations with significant uncertainty about the mean. Some investigators directly use such regression equations in a deterministic way to specify exposure. However, at site conditions vary significantly (rheology, water content, failure type, etc.), and even with the log-log regressions presented here, there is considerable uncertainty that needs to be quantified. It is important to account for the uncertainty in these estimates to derive a probabilistic measure of risk that also accounts for how well the regression fits in a certain range of values of the predictors.

Author Contributions: Paulina Concha gathered and analyzed the data. The conception, design, and writing of the paper were done between Paulina Concha and Upmanu Lall.

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. BHP Billiton Results for The Year Ended 20 June 2016. Available online: http://www.bhp.com/-/media/bhp/documents/investors/news/2016/160816_bhpbillitonresultsyearended30june2016.pdf?la=en (accessed on 1 July 2017).
2. Form 20-F, Annual Report Pursuant to Section 12 or 15(d) of the Securities Exchange Act of 1934. Available online: http://www.vale.com/EN/investors/information-market/annual-reports/20f/20FDocs/Vale_20-F_FY2016_-_i.pdf (accessed on 1 July 2017).
3. Renova Foundation Update. Available online: http://www.bhp.com/-/media/documents/media/reports-and-presentations/2017/170607_renovafoundationupdate.pdf (accessed on 1 June 2017).
4. Fernandes, G.W.; Goulart, F.F.; Ranieri, B.D.; Coelho, M.S., Dalesf, K., Boesche, N.; Bustamante, M.; Carvalho, F.A.; Carvalho, D.C.; Dirzo, R.; et al. Deep into the mud: Ecological and socio-economic impacts of the dam breach in Mariana, Brazil. *Nat. Conserv.* **2016**, *14*, 35–45.
5. Design of Tailings Dams and Impoundments. Available online: http://www.infomine.com/library/publications/docs/Davies2002b.pdf (accessed on 1 April 2017).
6. Caldwell, J.A.; Van Zyl, D. Thirty years of tailings history from tailings & mine waste. In Proceedings of the 15th International Conference on Tailings and Mine Waste, Vancouver, BC, Canada, November 2011.
7. Morgenstern, N.R. Improving the safety of mine waste impoundments. In *Tailings and Mine Waste 2010*; CRC Press: Vail, CO, USA, 2010; pp. 3–10.
8. Strachan, C.; Caldwell, J. New directions in tailings management. In *Tailings and Mine Waste 2010*; CRC Press: Vail, CO, USA, 2010; pp. 41–48.
9. Martin, V.; Akkerman, A. Challenges with conducting tailings dam breach studies. In Proceedings of the 85th Annual Meeting of International Commission on Large Dams, Prague, Czech Republic, July 2017.
10. Rico, M.; Benito, G.; Diez-Herrero, A. Floods from tailings dam failures. *J. Hazard. Mater.* **2008**, *154*, 79–87.
11. Pierce, M.W.; Thornton, C.I.; Abt, S.R. Predicting peak outflow from breached embankment dams. *J. Hydraul. Eng*. **2010**, *15*, 338–349.
12. MacDonald, T.; Langridge-Monopolis, J. Breaching characteristics of dam failures. *J. Hydraul. Eng*. **1984**, *110*, 567–586.
13. Álvarez, M.; Puertas, J.; Peña, E.; Bermúdez, M. Two-dimensional dam-break flood analysis in data-scarce regions: The case study of Chipembe dam, Mozambique. *Water* **2017**, *9*, 432.
14. Cannata, M.; Marzocchi, R. Two-dimensional dam break flooding simulation: A GIS-embedded approach. *Nat. Hazards* **2012**, *61*, 1143–1159.
15. TSF Failures 1915–2017 as of 16 Aug 2017. Available online: http://www.csp2.org/tsf-failures-1915-2017 (accessed on October 2017).
16. Blight, G.E.; Robinson, M.G.; Diering, J.A.C. The flow of slurry from a breached tailings dam. *J. South. Afr. Inst. Min. Metall.* **1981**, *1*, 1–10.

17. Tailings Dams Risk of Dangerous Occurrences. Lessons Learnt from Practical Experiences. Available online: http://www.unep.fr/shared/publications/pdf/2891-TailingsDams.pdf (accessed on 11 January 2018).

18. Chronologies of Major Tailings Dams Failures. Available online: http://www.patagoniaalliance.org/swp-content/uploads/2014/09/Chronology-of-major-tailings-dam-failures.pdf (accessed on 4 September 2014).

19. Fonseca do Carmo, F.; Kamino, L.H.Y.; Tobias, R.; de Campos, I.C.; Fonseca do Carmo, F.; Silvino, G.; de Castro, K.J.S.X.; Mateus Leite, M.; Rodrigues, N.U.A.; et al. Fundão tailings dam failures: The environment tragedy of the largest technological disaster of Brazilian mining in global context. *Perspect. Ecol. Conserv.* **2017**, *15*, 145-151.

20. Los Frailes Tailings Incident, Mining & Environment Research Network. Available online: http://www.pebblescience.org/pdfs/LosFrailes.pdf (accessed on 25 April 1998).

21. La Catástrofe de Aznalcóllar, X Aniversario: ¿Una Lección Aprendida? Available online: http://assets.wwf.es/downloads/_informe_2008.pdf (accessed on 1 April 2008).

22. Van Niekerk, H.J.; Viljoen, M.J. Causes and consequences of the Merriespruit and other tailings-dam failures. *L. Degrad. Dev.* **2005**, *16*, 201–212.

23. Feng, C.; Wang, H.; Lu, N. Log-transformation and its implications for data analysis. *Shanghai Arch. Psychiatry* **2014**, *26*, 105–109.