Revisiting the Sellmeijer method for backward erosion piping

Sina Ramezanifouladi & Jean Côté
Université Laval, Québec, Québec, Canada

GeoCalgary 2022 October 2-5
Reflection on Resources

ABSTRACT
Backward erosion piping, one of the initiating mechanisms of internal erosion and piping which classically occurs at the contact of dam foundation and upper cohesive layer, is responsible for lots of embankment dam incidents and failures. Sellmeijer was the first to develop a 2D theoretical criterion for the progression of the pipe considering its hydraulic properties. It was later improved by means of experimental data and statistical tools evaluating the influence of each incorporated parameters on the critical gradient and adding more soil properties into the criterion. However, further analysis on the experimental data reveals that the regression model from which the improved criterion was derived suffers from overfitting. In this article, considering multiple regression analysis concepts, the shortcoming with the improved criterion is highlighted. The data is re-analyzed, and more robust equation is presented.

RÉSUMÉ
Érosion régressive de conduit, l'un des mécanismes initiateurs de l'érosion interne de conduit qui se produit classiquement au contact de la fondation du barrage et de la couche cohésive supérieure, est responsable de nombreux incidents et ruptures de barrages en remblai. Sellmeijer a été le premier à développer un critère théorique 2D pour la progression de la conduite en tenant compte de ses propriétés hydrauliques. Il a ensuite été amélioré au moyen de données expérimentales et d'outils statistiques évaluant l'influence de chaque paramètre incorporé sur le gradient critique et ajoutant plus de propriétés du sol dans le critère. Cependant, une analyse plus approfondie des données expérimentales révèle que le modèle de régression à partir duquel le critère amélioré a été dérivé souffre de surajustement. Dans cet article, compte tenu des concepts d'analyse de régression multiple, la lacune du critère amélioré est mise en évidence. Les données sont réanalysées et une équation plus robuste est présentée.

1 INTRODUCTION

Internal erosion and piping are among the most important causes of earthfill dam incidents and failures. It occurs when soil particles within an embankment dam or its foundation are carried downstream by seepage flow leading to a continuous pipe from upstream to downstream side of an embankment (ICOLD 2015).

One of the initiating mechanisms of internal erosion is backward erosion piping (BEP) which generally occurs at the interface of either a dam body and its foundation or a cohesive layer overlaid on a non-cohesive layer in foundation. It is initiated at the downstream toe nearby an existing defect and will progress toward upstream under the roof of cohesive soil if the critical condition is met (Figure 1).

By taking into account the three main area of flow in the foundation medium, flow in the progressing pipe and equilibrium of grains at pipe bed, Sellmeijer (1988) proposed a theoretical criterion to determine the critical gradient across the structure required for progression of the pipe. The model was later improved based on a vast series of experimental and field tests exploiting multiple regression tools (Sellmeijer et al. 2011).

From statistical point of view, including variables with small weights in a regression model could lead to overfitting issues. This is the case observed in Sellmeijer's improved criterion which triggered the need to revisit this formula. The objective of this research is to propose an improved criterion. This will be done by multiple regression re-analysis of the data provided by Sellmeijer et al. (2011).

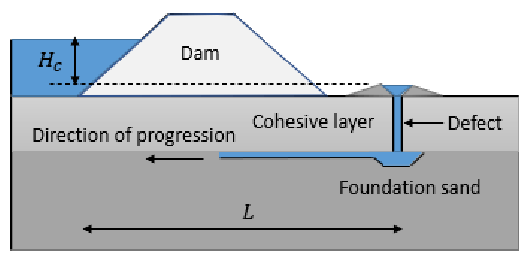

Figure 1. Schematic illustration of backward erosion piping

2 SELLMEIJER'S IMPROVED CRITERION

Sellmeijer's initial model was developed for a flow toward a pipe in an infinitly deep aquifer which later was modified to a finite layer. The model was implemented in MSEEP, a numerical piping program, so that backward erosion for more complex geometries could have been analyzed. By performing thousands of piping calculations in MSEEP and curve fitting of numerical outcomes, the design criterion of Equation 1 was proposed. In Eq. 1, H_c is critical head difference, F_R is resistance factor, F_S is scale factor, F_G is geometry factor, η is coefficient of White, θ is White's bedding angle, γ_p' is submerged unit weight of particles, γ_w is unit weight of water, d is particle diameter, K is intrinsic permeability, L is erosion length and D is the height of sand layer.

To assess the effect of sand characteristics on the critical gradient and to validate the formula of Sellmeijer

(Eq. 1), a series of small-, medium-, and full-scale tests of BEP were conducted as a part of a research program called Strength and Loading of Flood Defence Structures (SBW). Test procedure, material properties, observations and the outcomes were collected and analysed in Knoeff et al. (2009), Van Beek et al. (2011), and Sellmeijer et al. (2011).

$$\frac{H_c}{L} = F_R F_S F_G \begin{cases} F_R = \eta \frac{\gamma_P'}{\gamma_W} \tan\theta \\ F_S = \frac{d_{70}}{\sqrt[3]{KL}} \\ F_G = 0.91 (\frac{D}{L})^{\frac{0.28}{(\frac{D}{L})^{2.8} - 1} + 0.04} \end{cases} \quad [1]$$

Figure 2 shows the 10×30×50 cm^3 small scale test setup (erosion length of 34 cm) that was used in this project. For a compacted soil in the setup with specified soil properties, the overall head difference is increased gradually until critical gradient is obtained by which the progressing pipe reaches upstream without any stop.

Knoeff et al. (2009) demonstrated statistically the influence of different soil properties including relative density (RD), conductivity (k), soil grain size (d_{70}), uniformity (C_u), and soil angularity (KAS) on the critical hydraulic gradient ($\frac{H_c}{L}$) over different scales that the tests were performed. The final conclusion was based only on the results of statistical analysis for small-scale (47 tests) as not sufficient number of tests were carried out for medium-scale (7 tests).

Based on 38 sets of tests multivariate regression analysis with normalized variables was carried out. From the 47 tests, 5 tests in which another type of erosion (forward erosion) than the classical backward erosion was observed, as well as 2 more tests that were outliers and was thought might distort the direction of results, were put out of the analysis. The weights of each of 5 soil properties were calculated as were indicated in Equation 2. The properties subscripted with "m" in Eq.2 represents the mean value of the same properties which were summarized in Table 1.

The goodness of the regression model was demonstrated by the small distances of data to the diagonal line in the plot of predicted critical gradient versus corresponding experimentally measured critical gradients. Another sign of its goodness was the reduced amount of error for this plot compared to the same plot calculated by Sellmeijer's original criterion (Knoeff et al. 2009).

To assess the influence of each soil property on the measured critical gradient, Knoeff et al. (2009) tried to remove one variable each time and perform the regression analysis again in order to observe its effect on the weights of other variables. It was realized that the most influential variables were relative density, conductivity, and soil grain size. As a result, it was recommended not to include the influences of uniformity and angularity in the calculation rule. Nevertheless, Sellmeijer et al. (2011) incorporated the influences of all variables in the improved criterion that is shown in Equation 3.

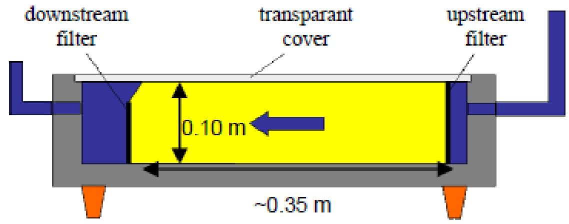

Figure 2. Schematic illustration of small-scale test setup (Van Beek et al. 2011)

$$\frac{H_c}{L} = e^{-0.713} \times (\frac{RD}{RD_m})^{0.354} \times (\frac{k}{k_m})^{-0.357} \times (\frac{d_{70}}{d_{70\,m}})^{0.398} \times (\frac{C_u}{C_{u\,m}})^{0.132} \times (\frac{KAS}{KAS_m})^{-0.021} \quad [2]$$

Table 1. Summary of the properties of the 38 tested soils (Sellmeijer et al. 2011)

Soil properties	Minimum	Maximum	Mean
RD (%)	50	100	72.47
$k \times 10^{-3}$ $(\frac{m}{s})$	0.027	0.372	0.102
d_{70} (μm)	150	430	207.92
C_u	1.3	2.6	1.81
KAS	35	70	49.76

$$\frac{H_c}{L} = F_R F_S F_G \begin{cases} F_R = \eta \frac{\gamma_P'}{\gamma_W} \tan\theta (\frac{RD}{RD_m})^{0.35} (\frac{C_u}{C_{u\,m}})^{0.13} (\frac{KAS}{KAS_m})^{-0.02} \\ F_S = \frac{d_{70}}{\sqrt[3]{KL}} (\frac{d_{70\,m}}{d_{70}})^{0.6} \\ F_G = 0.91 (\frac{D}{L})^{\frac{0.28}{(\frac{D}{L})^{2.8} - 1} + 0.04} \end{cases} \quad [3]$$

3 REGRESSION METHOD

Regression analysis is a statistical technique for modeling and investigating the relationship between two or more variables (Hines et al. 2003). Regression Equation 4 shows the relation between dependent variable, y, and independent variables, x_i and is called multiple linear regression model. A regression coefficient (β_i) represents the amount of change in dependent variable according to a unit change in the corresponding independent variable while keeping other independent variables constant. ε, is the error term which indicates the inefficiency of the model to fit the observation data. Regression equation can be expressed whether in a format as Equation 4, which proposes regression coefficients, or in a logarithmic domain which returns weights (Eq. 5). In order to obtain weights, multiple regression model must be generated based on the natural logarithm of each variable. Thereafter, the calculated unknown parameters are considered regression weights, which are usually determined by means of least squares method.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i + \varepsilon \quad [4]$$

$$y = \exp^{\beta_0} \times x_1^{\beta_1} \times x_2^{\beta_2} \times \ldots \times x_k^{\beta_k} \times \varepsilon \quad [5]$$

It is worth noting that adding more independent variables to the model could result in error reduction and obtaining better fit line, though, it could end up overfitting, and prediction problem for new observation data. Therefore, it is always intended to build a model with the fewest possible number of independent variables.

To assure that a regression model exists between a specified number of independent and dependent variables, the results must be statistically significant. The significancy of a model is investigated by "P-value" which is the measure of explained variation to unexplained variation by a regression. P-value of a significant regression must be greater than a determined confidence level, which is conventionally considered 95 % and it can be increased to 99.9 % (for more details, readers are referred to Hines et al. 2003).

Also, having two independent variables in a regression model that are highly correlated could result in poor determination of actual effect of each independent variables on the results which is called multicollinearity. Although it does not affect overall fitting process, it could make misleading prediction as the correct amount of contribution of each intercorrelated variables are not determined. In statistical analysis, a high "VIF" (variance inflation factor), a considerable difference between R^2_{adjusted} and $R^2_{\text{Prediction}}$, and high intercorrelation between independent variables could indicate existence of multicollinearity in the regression model. VIF demonstrates how much of the variance explained by an independent variable is captured by other existing independent variables in the regression model. R^2_{adjusted} is the adjusted version of R^2 which takes into account the number of independent variables incorporated in the model and $R^2_{\text{Prediction}}$ indicates how well the model can predict the new observation data.

Finally, the variance of errors of a model must be homogenously distributed. A non-homogenous model is not able to capture all the information in observation data by its independent variables which results in a non-random pattern in the residual errors of the model.

There is no unique statistical procedure to obtain the best regression model, and at some point, personal judgment has to interfere. An analyst can make benefit of model building procedure such as "Stepwise method" or "Best subsets method", nevertheless, comparing models' parameters including "P-value" of the model and coefficients, the amount of reduced error, number of independent variables, standard error, adjusted and prediction coefficient of determination and VIF must be taken into account.

4 RESULT

In this section first, the same regression model that was reported by Knoeff et al. (2009) and used to improve the Sellmeijer's original criterion is reproduced, and the shortcoming with their regression analysis is reviewed. Then, other alternative equations for the prediction of critical gradient are proposed based on the re-analysis of the SBW's small-scale test data.

4.1 Reproduction of regression model of Knoeff et al.

Soil properties and critical hydraulic gradient of BEP small-scale tests required for multiple regression analysis are collected from Sellmeijer et al. (2011). All five independent variables must be normalized dividing by their mean values summarized in Table 1. Although intrinsic permeability was used in Sellmeijer's criterion, for building the regression model, hydraulic conductivity can also be used in calculation, since each independent variable is divided by its mean value, therefore, there would be no difference between intrinsic permeability and hydraulic conductivity.

Table 2 and 3 show statistical parameters of the regression model and the explanatory powers or the weights of each five independent variables that were incorporated in the model as well as their P-values.

It must be noticed that Knoeff et al. (2009) and Sellmeijer et al. (2011) have obtained different weights than the ones collected in Table 3 which was also mentioned by Van der Zee (2011). Most important is that the P-values of the coefficients are considerably greater than the significance level of 5%. In other words, based on the 38 sets of observation data, incorporating all five independent variables in the model would result in overfitting, therefore, obtaining a better model must be investigated. Although Knoeff et al. (2009) recommended not to use angularity and uniformity in the formula, removing the two variables from formula would change the explanatory power of other variables. This in turn requires performing the regression again which would adjust the weights of other three variables. The values of VIF for conductivity and grain size are also high which indicate these two variables are highly correlated and having both in the model may result in misleading prediction.

4.2 Current research's proposed criteria

In the research for an improved regression model, the best subset selection approach was used. All possible combinations of independent variables are built and analysed. The best models are then picked out based on the comparison of the models' statistical parameters. Table 4 summarizes selected combinations of the five independent variables among several other possible combinations. Table 4, in fact shows only the two best ones for each possible number of variables. The cross sign under each variable indicates the one that is used in the model. The model highlighted in red are selected for further investigation and outcome comparison. In selection of the highlighted models, their statistical parameters were compared to all other available combinations whether put in Table 4 or not. Table 5 also summarizes the P-value of the model and coefficients as well as the VIF for coefficients. It is worth noticing that the other regression equations than models 3 and 6 were not selected despite their better statistical parameters, namely R^2 or $R^2_{adjusted}$ (according to Table 4) because P-values of their coefficients were greater than the 5% acceptable significance level (according to Table 5). As examples models 5 and 7 in Table 4 can be mentioned. In a few of them such as the last model which incorporates all five variables, high VIF was observed representing multicollinearity. Equations 6 and 7 indicate the two proposed regression models based on the results of SBW's small-scale tests.

$$\frac{H_c}{L} = e^{-0.952} \times (\frac{k}{k_m})^{-0.539} \times (\frac{d_{70}}{d_{70\,m}})^{0.694} \quad [6]$$

$$\frac{H_c}{L} = e^{-0.883} \times (\frac{RD}{RD_m})^{0.604} \times (\frac{k}{k_m})^{-0.148} \times (\frac{C_u}{C_{u\,m}})^{0.326} \quad [7]$$

For the models to be applicable for larger scales and taking particularly the effect of scale factor $(\frac{d_{70}}{\sqrt[3]{K\,L}})$ into account, the influence of each independent variable is empirically embedded in the Sellmeijer's original criterion. Equations 8 and 9 show the current research's modified criteria in accordance with the two achieved regression models.

Table 2. Reproduced regression model's information

P-value	R^2	$R^2_{adjusted}$	Standard error
1.22E-07	0.7	0.66	0.154

Table 3. Reproduced model variable's information

	$e^{\beta 0}$	$(\frac{RD}{RD_m})^{\beta 1}$	$(\frac{k}{k_m})^{\beta 2}$	$(\frac{d_{70}}{d_{70\,m}})^{\beta 3}$	$(\frac{C_u}{C_{u\,m}})^{\beta 4}$	$(\frac{KAS}{KAS_m})^{\beta 5}$
Eq. 3 (Sellmeijer):						
Weight	-0.079	0.35	-0.35	0.4	0.13	-0.02
Analysis of this study:						
Weight	-0.9	0.418	-0.306	0.292	0.168	-0.025
P-value	-	0.060	0.067	0.319	0.405	0.903
VIF	-	4.070	16.780	14.430	2.650	1.200

Table 4. Regression information of different combinations of independent variables by means of best subsets method

Model number	Number of variables	R^2	$R^2_{adjusted}$	$R^2_{Prediction}$	Standard error	RD	K	D70	Cu	KAS
1	1	46.6	45.1	41.3	0.19448	X				
2	1	38.0	36.3	30.4	0.20949		X			
3	2	65.9	63.9	59.4	0.15767		X	X		
4	2	63.1	61.0	57.4	0.16400	X	X			
5	3	69.5	66.8	62.5	0.15117	X	X	X		
6	3	69.2	66.5	62.1	0.15183	X	X		X	
7	4	70.3	66.7	61.8	0.15153	X	X	X	X	
8	4	69.6	65.9	61.1	0.15317	X	X	X		X
9	5	70.3	65.6	60.1	0.15384	X	X	X	X	X

Table 5. Other statistical parameters of different regression models of Table 4

Model number	Model's P-value	P-value of coefficients					VIF of coefficients				
		RD	K	D70	Cu	KAS	RD	K	D70	Cu	KAS
1	0.000	0.000					1.00				
2	0.000		0.000					1.00			
3	0.000		0.000	0.000				2.78	2.78		
4	0.000	0.000	0.000				1.14	1.14			
5	0.000	0.051	0.000	0.011			2.13	5.87	5.24		
6	0.000	0.000	0.001		0.013		1.14	1.21	1.07		
7	0.000	0.041	0.051	0.294	0.366		3.65	14.42	12.15	2.48	
8	0.000	0.070	0.000	0.013		0.732	2.24	6.45	5.89		1.13
9	0.000	0.060	0.067	0.319	0.405	0.903	4.07	16.78	14.43	2.65	1.20

$$\frac{H_c}{L} = F_R F_S F_G \begin{cases} F_R = \eta \frac{\gamma'_P}{\gamma_W} \tan\theta \\ F_S = \frac{d_{70}}{\sqrt[3]{K\,L}} (\frac{k}{k_m})^{-0.539} (\frac{d_{70}}{d_{70\,m}})^{0.694} \\ F_G = 0.91 (\frac{D}{L})^{\frac{0.28}{(\frac{D}{L})^{2.8}-1}+0.04} \end{cases} \quad [8]$$

$$\frac{H_c}{L} = F_R F_S F_G \begin{cases} F_R = \eta \frac{\gamma'_P}{\gamma_W} \tan\theta (\frac{RD}{RD_m})^{0.604} (\frac{C_u}{C_{u\,m}})^{0.326} \\ F_S = \frac{d_{70}}{\sqrt[3]{K\,L}} (\frac{k}{k_m})^{-0.148} \\ F_G = 0.91 (\frac{D}{L})^{\frac{0.28}{(\frac{D}{L})^{2.8}-1}+0.04} \end{cases} \quad [9]$$

From now on for the ease of calling throughout the paper, Equations 8 and 9 are named model 3 and 6, respectively. In model 3, both variables, k and d_{70}, already existed in the original criterion which are being affected by the obtained weights in this research, while in model 6, the influences of RD and C_u are added to the criterion for the first time.

The final model is built considering the scale factor $(\frac{d_{70}}{\sqrt[3]{K\,L}})$ itself as a variable along with the other five independent variables in the regression modelling procedure. In view of the fact that the scale factor plays the most important role in piping criterion for transition from small-scale to larger scales, incorporating it in the model might eliminate the need for its adjustment according to the original criterion. In this way, another regression model was built already incorporating the scale factor (Equation 10 and model 10). When the scale factor is considered as one of the possible independent variables among other five variables, surprisingly, combination of scale factor and conductivity indicates comparable statistical parameter to the two proposed combinations. Table 6 shows the regression information of model 10.

$$\frac{H_c}{L} = e^{-1.163} \times (\frac{k}{k_m})^{-0.307} \times (\frac{d_{70}}{\sqrt[3]{K\,L}})^{0.691} \quad [10]$$

Table 6. Regression information of model 10

P-value	R^2	$R^2_{adjusted}$	Standard error
7.37E-09	0.66	0.64	0.158

4.3 Precision of the models in prediction

To indicate precision of the proposed models in comparison with the existing criteria, the percentage of error in prediction for each model and small-scale test is calculated. In order to have a tangible insight over the prediction outcomes of the models, absolute values of errors in prediction are summed (SoE) according to Equation 11, and an average value of error for 38 tests is calculated and summarized in Table 7.

Table 7. Sum of error and its average for small-scale tests for the different models

Criteria	(SoE)	Average
Sellmeijer’s original criterion	5.004	0.132
Sellmeijer's improved criterion	4.801	0.126
Model 3 incorporating k, d_{70} (Eq. 8)	7.145	0.188
Model 6 incorporating RD, k, C_u (Eq. 9)	5.650	0.149
Model 10 incorporating k, Scale factor (Eq. 10)	2.057	0.054

$$SoE = \sum_{i=1}^{i=n} Abs[(\frac{H_c}{L})_{Measured\ i} - (\frac{H_c}{L})_{Prediction\ i}] \quad [11]$$

Compared to Sellmeijer’s original criterion and Sellmeijer’s improved criterion, model 10 reduces errors considerably and it predicts critical gradients more precise. Figures 3 and 4 demonstrate this fact clearly. While errors of the Sellmeijer’s improved criterion are scattered mostly between 0.1 to 0.2 (Figure 3c), errors of model 10 mostly distributed between -0.1 to 0.1 (Figure 4e), which is considerably lower. Regarding the other two proposed models, while multiple regression analysis of small-scale tests proposed two satisfactory equations (Eq. 6 and 7), their adjusted versions (models 3 and 6) poorly predict gradients for the small-scale (Figures 4a and 4c), though, overall prediction behavior of model 6 is comparable to the existing criteria of Sellmeijer.

The last step for validation of a given model is to investigate the homogeneity of errors. The distribution of the errors must not follow a pattern, otherwise it means predictive information of the model leaked over into the errors. This can be checked by plotting measured gradient versus predicted gradient which are shown in Figures 3 and 4. Data overlaying on the diagonal line of this plot represents normal distribution of errors. In Figure 3 (b and d), although no evident pattern of data scatter is observed for Sellmeijer’s original and improved criteria, the predicted gradients are greater than the measured ones in a non-conservative manner. The same observation is valid for model 3 and 6 in Figure 4 (b and d). Conversely, no obvious pattern or biasness of the errors is observed for model 10, and the data are more tightly distributed around the diagonal line in the proposed model (Figure 4f).

5 DISCUSSION

It was demonstrated that the Sellmeijer’s widely accepted improved criterion could not be perfectly reproduced by the provided SBW’s small-scale test data. As was explained by Van der Zee (2011) the main source of discrepancy observed between the weights provided by Knoeff et al. (2011) or Sellmeijer et al. (2011) and the analysis performed by Van der Zee or in the present research is that they have kept erosion length constant to 34 cm for all tests in their analysis while the length of erosion varied between 32.5 to 34.5 which led to obtain different critical gradients.

Moreover, by re-analyzing of the SBW’s data in this study, three new regression models were developed among which two equations (Eq. 6 and 7) ended up to the development of two new adjusted models in accordance with the Sellmeijer's original model: the first one being model 3, (Eq. 8) and the second one being model 6 (Eq. 9). The third regression equation (Eq. 10) simply incorporated the scale factor and is designated as model 10.

The two regression equations of Eq. 6 and 7 can be ranked at the same level of effectiveness as model 10 in statistical term of view because of almost identical P-value, R^2 and standard error. Since these models do not take into account the scale factor, gradient prediction of larger scale tests cannot be relied on them. On the other hand, their corresponding adjusted criteria (models 3 and 6) are not exactly fitted for the small-scale data. As a matter of fact, it is expected that implementing each of these four equations and models (Eq. 6 to 9) as a criterion for assessment of BEP may come with some shortages. On the contrary, the last proposed model (model 10) is more precise for small-scale, and having incorporated scale factor, makes it predict better for larger scale as well. However, this model was achieved only by analyzing the small-scale test data, hence, it is recommended that in the future research it is validated with larger scale test results and field data. It should be noted that to determine independent variables required for prediction of critical gradient for larger scale, each variable must be divided by its corresponding mean value for small-scale tests provided in Table 1. Furthermore, another achievement made with this model is the considerably fewer number of variables that it takes into account. Almost all the resistance factor, geometry factor, the effect of relative density, uniformity and angularity were excluded from the Sellmeijer's improved criterion. Although the new simpler model requires to be further validated, it is evident from the results shown in Figure 3 and 4 that it yields better predictions.

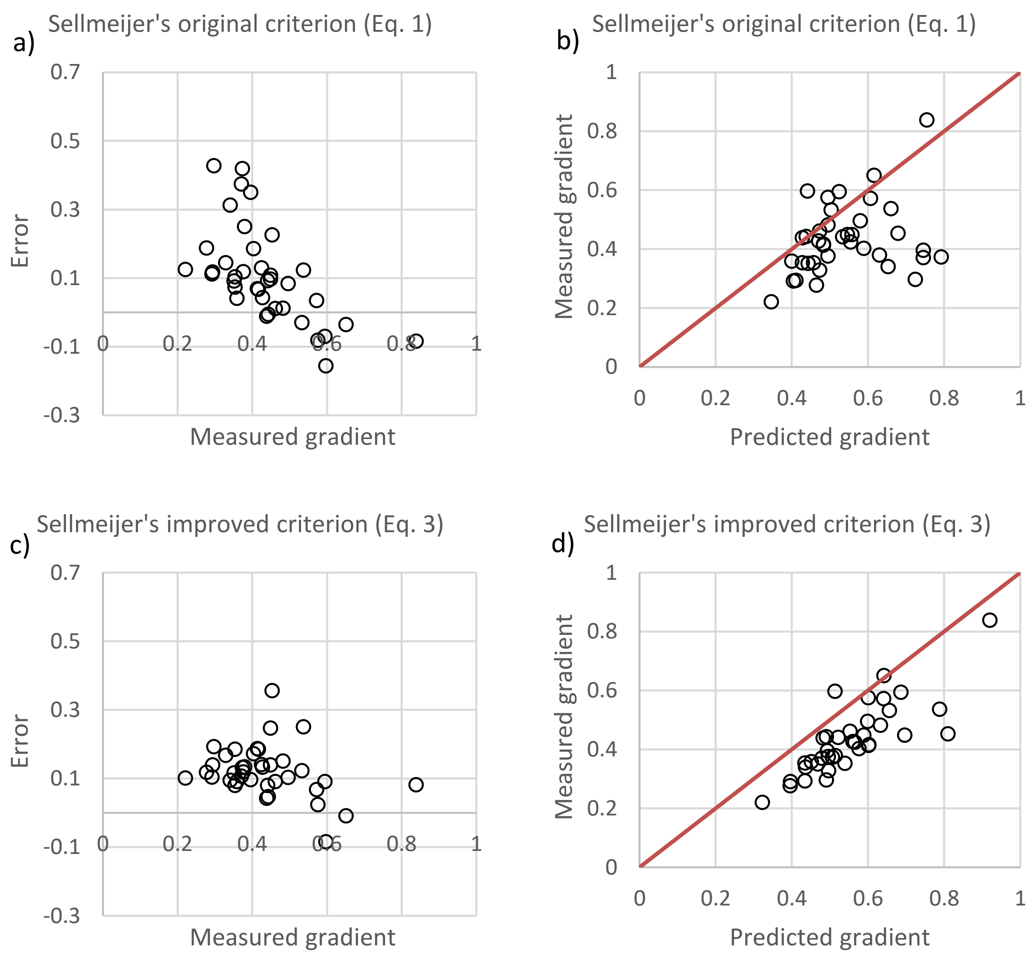

Figure 3. Plots of error versus measured gradient (a, c) and measured gradient versus predicted gradient (b, d) for the Sellmeijer's original and improved criteria

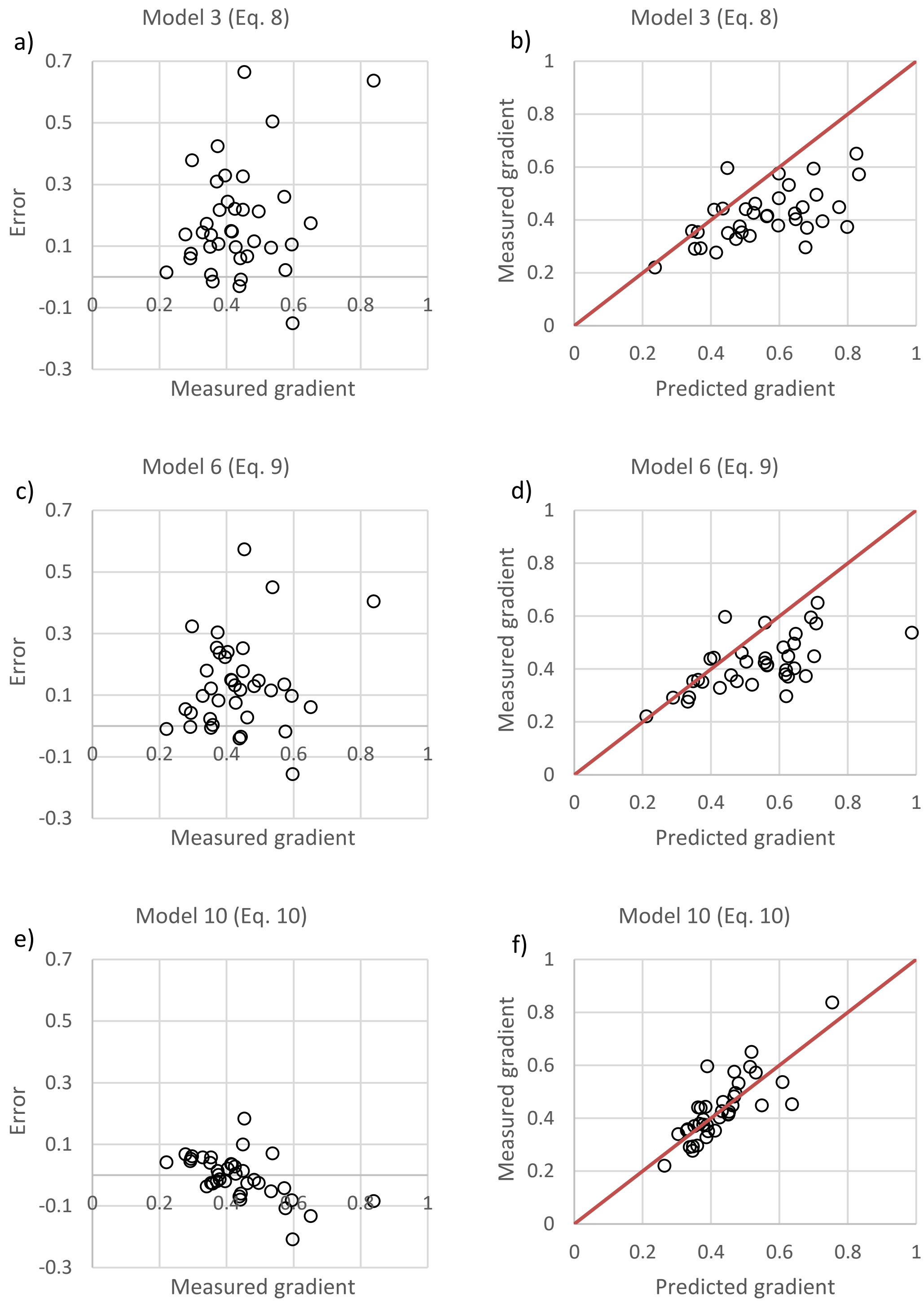

Figure 4. Plots of error versus measured gradient (a, c, e) and measured gradient versus predicted gradient (b, d, f) for the three proposed criteria

Sellmeijer's model relied on several parameters often difficult to assess and many times corelated. Also, Van Beek et al. (2019) demonstrated that the assumption of constant value of bedding angle for different particle diameter is incorrect. Therefore, reducing the number of variables ensures that only the most significant parameters are involved, reducing thus the negative effects of uncertainties related to the determination of several soil properties.

It is noted that both Sellmeijer's criteria and the proposed models in the current research are valid for prediction of critical gradient responsible for progression of backward erosion piping in a one-layer homogenous granular soil which may not be always the case in the field. Although several research has been carried out investigating the influence of heterogeneity on the critical gradient required for pipe progression, no such criterion has been developed so far. Further research is therefore required to better understand field conditions leading to piping.

6 CONCLUSION

The well-known Sellmeijer's criterion that was improved by means of experimental and field tests and statistical tools was revisited in this paper. The statistical re-analysis of SBW's small-scale data indicated that the weights obtained herein for independent variables were different than those obtained by Sellmeijer et al. (2011) because of using the variable erosion length of the tests in the regression analysis. In addition, it was demonstrated that Sellmeijer's improved criterion was suffering from overfitting indicated by the greater P-value of the coefficients than the acceptable significance level. Hence, in this study, systematic multiple regression analysis of the same data helped developing alternative more compacted models for critical gradient prediction. As a result, three new regression equations were built. Two of them took into account the effect of $\mathrm{k}, \mathrm{d}_{70}$ and $\mathrm{RD}, \mathrm{k}, \mathrm{C_u}$, and the third one incorporated scale factor.

The three equations were the best regression models that could be achieved by the SBW's small-scale test data according to the P-value of their model and coefficients, standard errors so on. However, the first two equations were adjusted empirically in accordance with the formulation of Sellmeijer's original criterion to ensure that they can predict the critical gradients of the larger scale tests. Comparison of sum of absolute error of the three proposed models (models 3, 6, and 10) with those of the Sellmeijer's criteria indicated that model 10, which incorporates the scale factor, gives the best prediction results for the small-scale tests. Although equations 6 and 7 were statistically better alternatives than the regression equation (Eq. 2) that Sellmeijer used to develop his model (Eq. 3), their prediction behavior with small-scale data did not product satisfactory results after implementation within models 3 and 6. Furthermore, the plot of measured gradients versus predicted gradients for the two Sellmeijer models and the three models proposed herein demonstrated the existence of bias in all models except for model 10. This is clearly shown in Figure 4f where the data are homogenously distributed around the diagonal line.

The objective of proposing a new simpler model has been reached. Nevertheless, it needs to be validated for larger scale and field data.

7 REFERENCES

Hines, W.W., Montgomery, D.C., Goldsman, D.M. and Borror, C.M. 2003. *Probability and Statistics in Engineering*, 4th edition, Wiley.

ICOLD 2015. Internal erosion of existing dams, levees, and dikes, and their foundations, *International Commission On Large Dams*, Paris, France, Paris, Bulletin 164, 1.

Knoeff, H., Sellmeijer, H., Lopez, J. and Luijendijk, S. 2009. SBW Piping - Hervalidatie piping, *technical report*, Deltares.

Sellmeijer, J.B. 1988. On the mechanism of piping under impervious structures, *Doctoral dissertation*, Technische Universiteit Delft, the Netherlands.

Sellmeijer, J.B., Lopéz de la Cruz, J., Van Beek, V.M. and Knoeff, J.G. 2011. Fine-tuning of the piping model through small-scale, medium-scale and IJkdijk experiments, *European Journal of Environmental and Civil Engineering* 15(8): 1139-1154.

Van Beek, V.M., Knoeff, H. and Sellmeijer, H. 2011. Observations on the process of backward erosion piping in small-, medium- and full-scale experiments, *European journal of Environment and Civil Engineering*, 15(8): 1115-1137.

Van Beek, V.M., Robbins, B.A., Hoffmans, G., Bezuijen, A. and Van Rijn, L.C. 2019. Use of incipient motion data for backward erosion piping models, *International Journal of Sediment Research*, 34(5): 401-408.

Van der Zee, R. 2011. Influence of sand characteristics on the piping process, *Master thesis*, Technische Universiteit Delft, the Netherlands