Methodology for Identification of the Key Levee Parameters for Limit-State Analyses Based on Sequential Bifurcation

Rossi, Nicola; Bačić, Mario; Librić, Lovorka; Kovačević, Meho Saša

sustainability

MDPI

Article

Methodology for Identification of the Key Levee Parameters for Limit-State Analyses Based on Sequential Bifurcation

Nicola Rossi *, Mario Bačić, Lovorka Librić and Meho Saša Kovačević

Faculty of Civil Engineering, University of Zagreb, 10000 Zagreb, Croatia
* Correspondence: nrossi@grad.hr

Abstract: Levees are linear structures that are continuously reconstructed throughout the years and whose construction and behavior depends on local soil conditions, as well as requirements regarding impermeability and mechanical resistance. This results in various levee cross sections, even within the same levee. In situations of extreme water events, when timely actions are required, this variability poses a problem for decision-making based on observed behavior, which is highly dependent on the specific section parameters. Creating models for each problematic section becomes impractical, and because of that, in this study, 91 different cross sections from 16 levees are considered to identify the key levee parameters with the largest effects on three observed mechanisms: deformations, exit hydraulic gradients, and factors of safety. The implemented factor screening methodology is based on the sequential bifurcation method (SB) and numerical analyses. The SB method successively investigates groups of factors and uses their cumulative effects to identify the important groups and to discard the unimportant based on a previously selected parameter Δ, until the groups are reduced to single factors that may be deemed important. It is found that approximately 30% of all the factors used to describe the most complex sections are considered important by at least one of the investigated mechanisms.

Keywords: river levees; sequential bifurcation; factor screening

Citation: Rossi, N.; Bačić, M.; Librić, L.; Kovačević, M.S. Methodology for Identification of the Key Levee Parameters for Limit-State Analyses Based on Sequential Bifurcation. *Sustainability* **2023**, *15*, 4754. https://doi.org/10.3390/su15064754

Academic Editor: Jongmuk Won

Received: 24 January 2023
Revised: 1 March 2023
Accepted: 3 March 2023
Published: 7 March 2023

1. Introduction

There are many types of levees being constructed, considering their different cross sections and the materials used for each part of the levee. This also includes the subsoil on which they are constructed, as the behavior of the levee is also affected by it. Each of these components is susceptible to some factors that may or may not be controlled, such as locally available materials, construction space requirements, ULS and SLS requirements, local practices, etc. Combinations of those factors yield the many different cross sections that can be encountered. Such variations may even occur within a single levee, which is why levees are often divided into what we call "reaches" when conducting analyses. Reaches are segments of a levee that have similar characteristics such that they can be considered using one single cross section [1]. These cross sections are often defined by 5 to 10 parameters selected based on theoretical and empirical prior knowledge and experience, without looking too much into it. However, if we were to describe any section uniquely and use advanced constitutive laws such as the Hardening Soil model, then even a simplified geometry (including the subsoil) would require over 100 parameters, which can be geometrical, physical, mechanical, or hydraulic. Granted, most of those parameters do not have a noticeable effect on the desired quantity (such as deformations), but a relatively quick preliminary analysis might indicate some parameters that may be overlooked and, more importantly, detect the relative importance of each one so the analyst knows where to focus. These parameters and their order of importance may change with any specific case being analyzed and with the response parameter of interest. In contrast to determining the factors that affect the levee's response to high-water events, Mirosław-Świątek et al. [2]

showed a method for determining the effects of eight groups of factors, defined by factor types, on the technical condition and safety of levees after a period of exploitation, based on experts' opinions.

This paper presents a methodology for an efficient evaluation of the important parameters that affect the behavior of all the levees found within a specific area of interest so that further analyses may be conducted based on the obtained results of the presented methodology. The methodology relies on statistical techniques for the design and analysis of experiments. Generally, to design an experiment, one needs to know what is to be studied; that is, to have a clear statement of the problem and objectives, how will the data be collected, and how the data will be analyzed [3]. The type of experiment required for detection of the key parameters is called factor screening, and in this study we will specifically focus on the sequential bifurcation (SB) method. Designs of experiments can be generally divided into classical and sequential designs. Classical designs focus on creating one big experiment that answers some specific questions, i.e., the whole experiment setup must be known in advance before starting. On the other hand, sequential design starts with one smaller experiment, and the knowledge gained from its results is used to "adjust" the parameters of the later experiments. One such method is the SB method developed by Bettonvil [4] in his doctoral thesis in the field of economics. Throughout the years, SB has been improved within various fields of study, and other variants have emerged, such as the extension by Cheng [5] where the response is allowed to be stochastic and subject to significant error, and then the controlled SB (CSB) [6], the CSB-X [7], the fractional factorial CSB (FFCSB) [8,9], the use of second-order metamodels in SB as explained by Kleijnen [10], as well as the multi-response SB (MSB), which allows for a simulation to have multiple response types [11,12]. However, the authors did not come across the usage of the SB within the field of geotechnics. The SB method is a similar idea to the binary search in computer science, which divides a sorted array into two halves in each iteration and eliminates one (even though SB does not necessarily eliminate one in each step). In the SB, the elimination of one (or none, or both) of the halves is controlled by a control parameter, Δ, set in advance, which can be chosen arbitrarily by the analyst's judgement, or by techniques described in the mentioned papers. On top of that, SB also identifies the magnitudes of each effect. The goal of this study is to identify the most important parameters of the levees, within a large area of interest, that affect their behavior, but also contain enough information to describe the different cross sections we may encounter. These parameters will be later used to create generalized predictive models to predict the behavior of levees during high-water events based on monitoring data. The methodology is demonstrated on Croatian levees in areas prone to floods, sometimes with catastrophic outcomes [13,14], which are also being managed with very limited funds. As many of these levees were constructed when height requirements were substantially lower than today's standards, many are being reconstructed and many more are waiting their turn. This methodology is the first step in generalizing the levees' behavior in the analyzed area, which would help improve decision-making before, during, and after high-water events.

2. Methodology

The methodology presented in this study is based on a popular factor screening method—the sequential bifurcation (SB)—with slight changes to the procedures, coupled with statistical analyses and complex geotechnical numerical analyses. The results indicate a list of factors which are found to be most important in the behavior of levees, with regards to the observed mechanisms. The importances are given as percentages for a specific factor's effect on a specific mechanism. However, when conducting analyses it is always useful to consider the previous knowledge and local expertise regarding the subject, and thus this list is not exclusive.

2.1. The Sequential Bifurcation Procedure

The classic SB method relies on the assumption that the metamodel that simulates the system's behavior is linear, as shown in Equation (1), where y is the response, β represents the model parameters, x represents the independent variables that are then coded, and e is the error. Each of the selected variables (parameters, factors) is observed at two levels, which are encoded as 0 and 1 based on the effect that the value has on the response; i.e., if the value decreases the response, it is coded with 0, and if it increases the response, it is coded with 1. This leads to the second assumption of the SB method, which states that the signs of the model parameters are known. In other words, for each levee parameter, we must know its effect on the response. In accordance with this assumption, the function is assumed to be monotonically increasing. Otherwise, some effects may cancel each other out within the SB groups and may thus go unnoticed.

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + e \tag{1}$$

SB starts with two analyses, which consist of one where all the parameters are set to the "high" value coded with 1, and the second where all are set to the "low" value coded with 0. Continuing after the first two runs, the factors are then split into groups labelled w_j, where j indicates that the first j out of the k parameters is set to the "high" value, while the rest $(k - j)$ are set to their "low" values. Each group's effect is compared to the parameter Δ to check whether the effect is important or not. This continues until the groups are reduced to individual factors and all groups are exhausted. Thus, during the SB, the model parameters β are estimated for groups of factors and for single factors, as shown in Equation (2), where $j \neq j'$.

$$\widehat{\beta_{j'-j}} = \frac{(w_j - w_{j'-1})}{2} \qquad \widehat{\beta_j} = \frac{(w_j - w_{j-1})}{2} \tag{2}$$

If the previous assumptions are satisfied, the first two runs should yield the highest and lowest values of all subsequent runs. However, it is not uncommon to be uncertain about some factors' effects and encode them incorrectly. In such cases, Kleijnen [10] suggests excluding these parameters from the SB analysis and analyzing them individually (in groups of 1). Instead, Oh et al. [8] in their version of the SB, suggest conducting a preliminary fractional factorial design of resolution III to identify the sign of each effect, then perform the SB with all the correctly encoded factors first, and then with the others. Due to the efficiency of the SB method, in our case, the identification of the eventually incorrect signs is done with just the SB method itself. If some factors are incorrectly labelled, the results of the first two analyses will not be the minimum/maximum, but will be higher/lower than the same runs with correctly labelled parameters. The amount by which they would differ depends on the cumulative importance of the incorrectly assigned factors. Figure 1 shows the estimated response of the "all high" and "all low" runs for two situations with increasing number of incorrectly assigned factors, for a general case of SB. In one situation, the first incorrectly encoded factor is the most important one, with each subsequent incorrectly encoded factor being the second most important one, and so on until the least important one is reached (termed "decreasing" in the figure). The second situation is the opposite (termed "increasing" in the figure). The realistic situation where seemingly random errors in factor encoding exist is found between these boundary curves. Firstly, all factor combinations that yield results on the right side of the curve-pair intersections, where the "minimum" response is higher than the "maximum", are immediately obviously incorrect. For the left part, if we define the parameter Δ as a fraction of the difference between the first two SB runs, which means it is a function of the incorrectly assigned factors, then the Δ is lower when there are more incorrectly assigned factors, which makes it more likely for SB to identify those parameters that yield a response lower than the "all low" run or higher than the "all high" run. The high and low values of such parameters should be switched. This is all done to avoid effect cancellation, which is a main concern in

SB. However, as found by Dean and Lewis [15], cancellation occurs very rarely under effect sparsity, which states that a system is mostly controlled by main effects and two-factor interactions. This ties in with the third and final assumption of SB—heredity—which states that if a factor does not have an important main effect, then it also does not have important interactions. This means that all the important interactions are found only between the parameters that have important main effects, and no significant interaction exists between a detected and undetected factor.

From Equation (2), it can be seen that, to estimate a factor, say factor 6 out of 10, one would need to perform an analysis with the first 6 factors at their high values and the rest on low, then use the results from the analysis where the first 5 factors are at the high values and the rest on low, to subtract one from the other and get the effect of only factor 6 being at the high value. This works for systems that are truly linear. Wan et al. [6] used an approach where the groups consist of having only the parameters of interest at their high values, which directly estimates their effects. The latter is implemented in this study. It seems that the standard method of estimating a group's or single factor's effect is to have it on "high", while all the others are on "low". However, due to the complex stress–strain behavior of levees, the situations with all the factors at "low" and with all the factors at "high" produce completely different behavior mechanisms. Since we are interested in predicting what happens at the "high" values, we instead use the approach of estimating the factor importances by having those factors set to their "low" values, while all the others are on their "high" values. This way, we do not observe how much a group or single factor increases the response compared to the situation of all on "low", but how much that group or single factor decreases the response compared to the situation of all on "high". In our case, these two situations are not equivalent. However, certain responses are better observed in one way, and other responses in the other way, so in this study both are implemented, as described later in this section.

Figure 2 shows the structure of the described sequential bifurcation method with the required analyses and decisions to be made in each step.

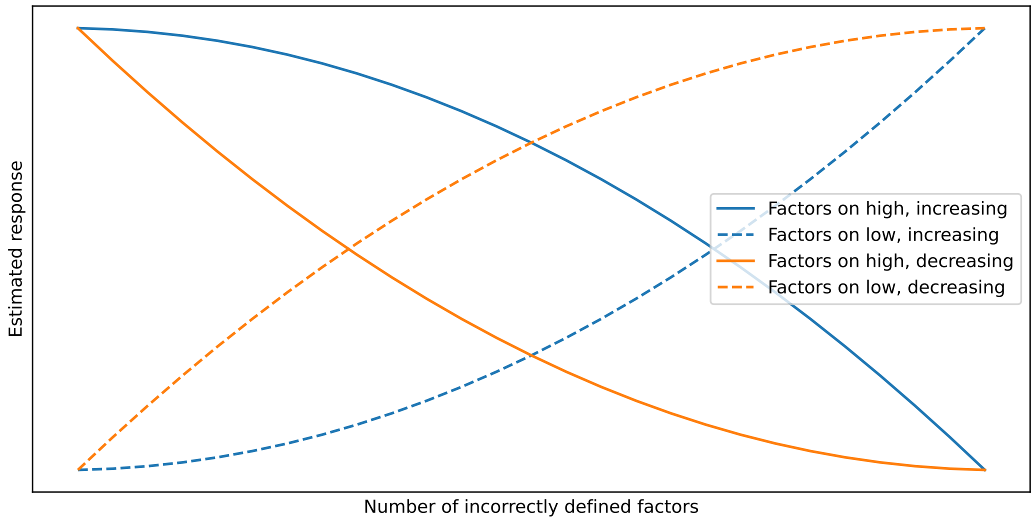

Figure 1. Bounds of the first two analyses in the SB method.

First two runs "All high" run (w_k, all factors coded with 1): gives max. deformations, max. gradients, min. factor of safety

"All low" run (w_0, all factors coded with 0): gives min. deformations, min. gradients, max. factor of safety

$\Delta = (w_k - w_0) \times 0.05$

Middle runs Take a list of factors from the queue and set all of these factors to either $+$ or $-$, depending on the selected approach, while all the rest are set to the opposite sign. After performing the analysis, the result is the group's effect, $w_{j'-j}$, where j' through j are the parameters set to the selected level, with all the others set to the opposite level.

Check if the specific list (group) of factors is important by comparing the results: $w_k - w_{j-j'} \geq \Delta$ or $w_{j'-j} - w_0 \geq \Delta$. If the group is important, divide the list of factors into two lists, and add them to a LIFO (Last-In-First-Out) queue, such that the first list to be analyzed always contains factors of assumed higher importances. Optionally, the effect of the other half of the original list can be checked by superposition and either discarded if it seems to be not important or left as it is if it seems important. Repeat until the queue contains lists of length 1.

End runs When the list contains only one factor, it is set to either $+$ or $-$, while the rest are set to the opposite sign. Its effect is estimated by: $w_k - w_i \geq \Delta$ or $w_i - w_0 \geq \Delta$, where w_i is the response for that single factor set at that specific level, with the rest set at the opposite level. Repeat until the queue is empty.

Figure 2. Structure of the described sequential bifurcation method.

2.2. Implementation

The first step in the implementation of this methodology is to define the area of interest based on the problem of interest, i.e., which structures to consider in this factor screening analysis. The covered area (or structures) may be defined in terms of geographic locations, a specific class of structures if some classification exists, same owners and managers, or any other specific criteria. In our case, the problem of interest is the identification of the key factors affecting the levees' behavior with regards to several limit states, and thus our structures of interest are river levees found across various locations. The next step is to define all the parameters that may be used to describe the more or less complex geometry of these levee sections—the geometric, hydraulic, physical, and mechanical parameters of each component, as well as the relevant hydrological and hydrogeological data. The parameters must be chosen based on the desired complexity of the models, and again the problem of interest that defines the parameters required to conduct the analyses. Then, the data should be gathered from all relevant sources: project designs, field and laboratory investigation works, technical drawings, jurisdictive institutions, etc. Often, the data are incomplete, i.e., not all parameters are available for all the sections. In such cases, they can either be estimated based on experience and knowledge of local regulations and/or practices, or just left blank, requiring work with less data.

When a database is formed from the gathered data, exploratory data analysis is conducted to facilitate further analyses. This includes factor normalization to overcome impossible or unrealistic combinations that result from a random selection of factors' values, finding factor correlations, and data transformation. The factors are normalized with regard to other factors of the same type (geometric, hydraulic, physical, mechanical). Most geometric parameters are normalized to the levee height or crown-width. Some physical and mechanical parameters that need to be in specific relation to other parameters are also defined this way, e.g., hydraulic conductivities and Young's moduli of the various components.

After having each parameter defined in the desired units, the range bounds of each factor for the analyses need to be defined. For some parameters which may not be available, such as flood durations and inundation time, the bounds can be estimated from experience. For the available parameters, the bounds are first taken as the extreme values of the gathered data for each parameter and are set as the two levels of each. However, it should be noted that if the levee's behavior is not linear, then the selected range may greatly affect the calculated model parameter β. Additionally, some factors may display lower or higher effects only due to their ranges being smaller or larger in comparison to other factors' ranges [16]. For example, if the levee height is varied within a small range, say ± 0.25 m from a mean value, while berm width is varied in a large range, say 1–5 m, then the berm width might show higher importance than the levee height, which would generally not be expected when searching for parameters that affect levee behavior. Thus, it is important to clearly define the purpose of conducting these analyses. One purpose may be to deepen the understanding of the behavior of levees in general, and the other to detect the parameters that have the most effect on the behavior of levees found within a specific area of interest. In the latter case, the ranges chosen based on the data gathered within that area intrinsically contain information about the importance of parameters in the area, i.e., parameters which we should focus on. Now, if we look again at the previous example with the latter purpose in mind, then the fact that the SB identified the height as unimportant and the berm width as important means that we should focus more on the berm width, because all the levees are practically the same height, so a mean value could be a good enough representation. However, even by using the normalized values, the ranges may be too large such that using the extreme values can still yield numerical model failures due to unrealistic parameter combinations. Additionally, the highly non-linear nature of the levees' behavior may cause the linear metamodel of SB to be unfit for this system if the behavior cannot be linearized over the selected ranges. Because of this, data transformations are used to create reduced ranges if needed, as follows. First, all of the heavily skewed data are transformed by logarithmic, square root, or Box–Cox transformations to make the data symmetrically distributed around the mean. Then a normal probability distribution is created for each parameter by setting $\pm 3\sigma$ of the transformed variables as the ranges bounds, and then selecting the new ranges as $\pm 1\sigma$ away from the mean, which is either the value between the range bounds or the 50th percentile. The average of the range bounds is used when the data are approximately uniformly distributed. The real parameters are then found by reversing the transformed values of $\pm 1\sigma$. With this method, the whole range of identified parameters is covered, which is important because there are no outliers in these data, only extreme but realistic values. For data that look normally or uniformly distributed, the transformation step is skipped.

Now, with the defined ranges, factor screening analyses using the described SB method are conducted. The factor Δ used to decide whether a factor or a group is important is usually a fraction of the value of the response type/performance-measures (cost of the supply chain, cycle-time of a system, deformations of a system, etc.). However, the decision can also be based on multiple responses. This study utilizes the described SB procedure, and multiple responses are observed simultaneously only when all the factors had equal effect signs for both responses (which does not require the usage of MSB), while for other response variables of interest, separate SB analyses are conducted. Through the responses of interest, all the relevant geotechnical analyses are considered—deformation analyses through the maximum model deformations (regardless of the location of its occurrence in the section; D), flow analyses through the exit hydraulic gradients (G), and safety analyses through the factor of safety (F). Since a goal of this study is to identify parameters that may be used to create predictive models that will rely on monitoring data, one more response variable is added. This response is the horizontal deformation at a specific point in the levee section (the landside part of the crown in this case; Ux), because it is a quantity that can be directly measured by conventional monitoring, which is one of the main aspects of levee safety control and management. The horizontal deformations of a specific point are observed simultaneously with the model deformations. Other similar responses can be added based on the available monitoring equipment, such as piezometer results, etc. Finally, the model deformation, the deformation at a specific point, and the gradient are observed by comparing the effect of each parameter at its low value with all the others at their high value, while the factor of safety is observed by comparing the parameter's high value with all the others at their low value. The reasoning behind the way the factor of safety is observed is that often the slip surface generated by the strength reduction method (SRM) coupled with the previous flow–deformation analyses with all the parameters set to their high values (lowest factor of safety), develops as a shallow slip surface that does not encompass the levee at all, but forms within the channel or river slopes. On the other hand, the slip surface generated with all the factors set to their low values (highest factor of safety) develops as a larger surface that encompasses the levee body. By turning a group or single factor to its high value while all the others are on their low values, it is more likely that the generated surface will encompass the levee.

It should be noted that the parameters identified as important will vary depending on the stress/strain component of interest and its location in the levee section.

3. Case Study

For this study, 16 levees were considered, and from each, 1–14 cross sections. The chosen levees were all found in Croatia and varied from segments of only a few hundred meters in length to over 10 km. Their positions are shown in Figure 3. In total, 91 different cross sections were taken for the statistical analysis. The criteria for choosing the cross sections were that each cross section must have some specific geometry compared to other cross sections of the same levee, and/or in the same way have specific subsoil conditions. The criterion for saying that something is "specific" compared to other sections was a subjective choice based on the analyst's experience. Some levees simply contain less variability in their geometry, subsoil stratigraphy, and material parameters due to the construction methods, locations, and lengths. Because of this, for some levees it was possible to identify more specific-sections than for others. To uniquely define a levee section consisting of the foundation soil, a relatively thin surface soil layer, an impermeable levee core, the levee's body, and two berms, 102 parameters have been identified, which may be divided into three groups, as shown in Tables 1–3, while the general geometry in Figure 4 also shows all of the geometric factors. Out of the 91 levee sections analyzed, not all sections contained all of the mentioned components, but instead contained different combinations of them. The numbers of levee sections containing each specific component along the levee body and foundation soil are shown in Figure 5. All of the geometric factors with the unit of length were normalized to either the crown height from the landside toe, or to the crown width, with three exceptions being: the crown width of the core, which was normalized to the levee width at the core height; the thickness of the top soil layer, which was not normalized; and the channel width, which was also not normalized. Some mechanical factors have also been normalized. These are Young's modulus of the core, normalized to the modulus of the levee body, and the vertical conductivities of the core and both berms, which were normalized to the vertical conductivity of the levee body. For each factor in the tables, the sign of their effect is mentioned, where the plus (+) or minus (−) signs indicate which of those operations (increase or reduction) on the factor's value increased the responses. The increase in a response is here defined as the change towards a more unfavorable value. The indicated signs are the ones used to obtain the highest and lowest responses for the appropriate runs and response types.

Table 1. Common physical, mechanical, and hydraulic parameters.

Physical & Mechanical		Hydraulic	
unit weight	− − − −		
porosity	+ + + +		
cohesion	− − − − [1]		
friction angle	+ + − − [2]		
dilatation	− − − − [3]	vertical perm.	+ + + + [5]
Poisson ratio	+ + + +	anisotropy	+ + + +
OCR	− − − −		
Young's modulus	− − − −		
unload./reload. modulus	− − − −		
power *m*	− − − − [4]		
	D Ux G F [6]		D Ux G F

[1] Except for the foundation soil, where the effect is + + − −; [2] Except for both berms, where the effect is − − − −; [3] Except for the foundation soil, where the effect is − − − −; [4] Except for the levee body and core, where the effect is + + − −; [5] Except for the land. berm (− − − −), top soil (+ + − +), found. soil (− − − +); [6] Response types: "D"—deformations, "Ux"—crown displacement, "G"—hydraulic gradient, "F"—factor of safety; The responses are increased by either increasing (+) or reducing (−) the factor's value.

Table 2. Specific geometric parameters.

Component	Geometric Parameters	D Ux G F
Levee body	waterside slope	− − − +
	landside slope	+ + + +
	crown width	− − − −
	crown height from waterside toe	+ + + +
	crown height from landside toe	+ + + +
Impermeable core	waterside slope	− − − +
	landside slope	− − − −
	crown width	− − − −
	crown height from landside toe	− − + +
	core position within the body [1]	+ + − +
Landside berm	height	− − − −
	width	− − − −
	slope angle	+ + + +
	berm angle	+ + + +
Waterside berm	height	+ + − −
	width	− − − +
	slope angle	+ + + +
	berm angle	+ + + +
Surface layer	thickness	+ + + −
Foundation soil		

[1] − moves the core towards the water side, and + towards the land side.

Table 3. Geometric and hydraulic parameters not associated with any specific material.

Geometric		Hydraulic	
river depth	+ + − −		
river bank slope	+ + + +	landside water level	+ + − −
river dist. from waterside toe	− − − −	inundation time	+ + + +
channel depth	+ + + +	duration of max. water level	+ + + +
channel bank slope	+ + + +	height of water event	
channel width	+ + + +		
channel dist. from landside toe	− − − −		
	D Ux G F		D Ux G F

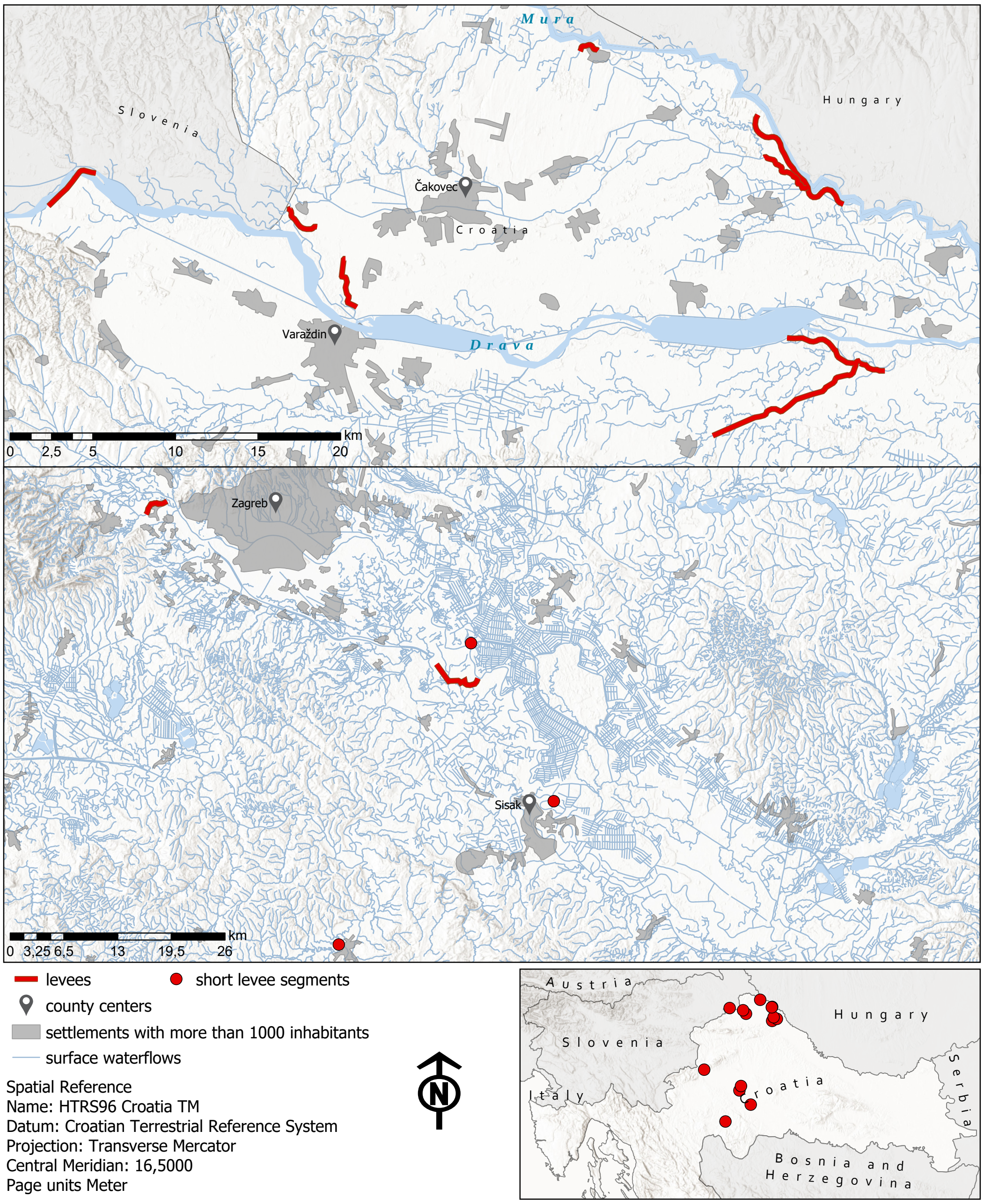

Figure 3. Map showing the locations of the chosen levees.

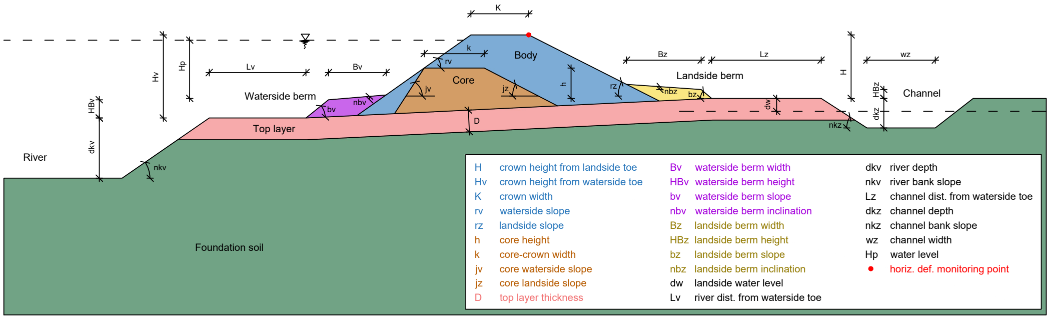

Figure 4. General model with identified geometric characteristics.

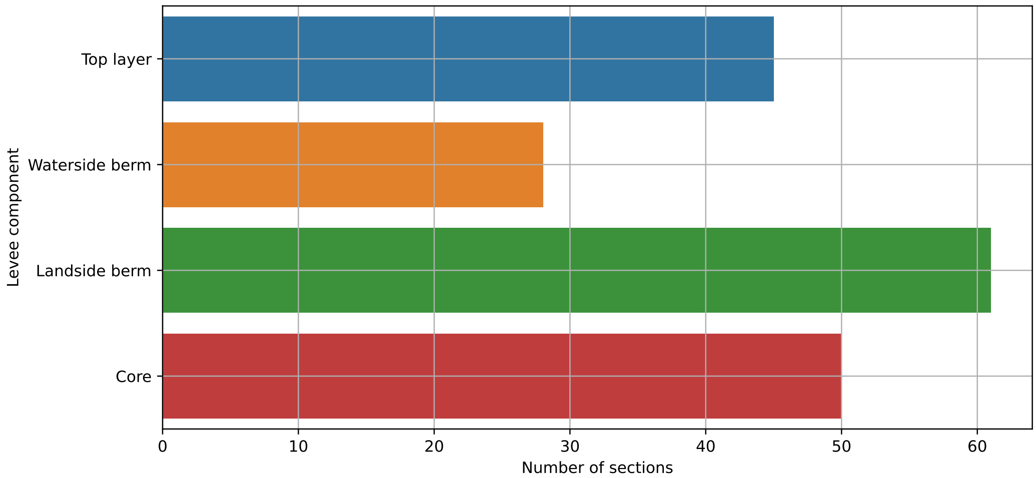

Figure 5. Number of each levee component found in the 91 analyzed cross sections.

The ranges for the factors were determined from statistical analysis. As previously described, SB takes only the limiting values of the ranges, i.e., the minimum and the maximum. A combination of only the observed extreme values often leads to unrealistic levee section and material parameters, and if numerical analyses are used to obtain results, this leads to numerical problems and thus results are impossible to acquire. Additionally, the relationships of deformations and other responses to all the input parameters are highly non-linear, similar to the function of safety factors shown in Rossi et al. [17]. Since the SB is based on a linear metamodel, this would cause the model to be unfit for the data and as a result give incorrect factor importances. To get around this problem, in this study, the ranges were reduced as previously described (the data in the ranges were already normalized prior to the reduction). The only exception to the described reduction method is the water level on the land side, where one value was taken at the surface level (as it is the most commonly used value in numerical analyses), and the other value as the median of all the other observed values. In this way, we obtained much more realistic levees, while also reducing the range covered by the function, which can thus be better linearized. However, care must be taken not to reduce the range too much in order for the results of the numerical analyses to be distinguishable enough from one another and not fall within errors contained in the results of each analysis.

The soil's behavior was modelled using the Hardening Soil model (HS), which requires both Young's modulus and the constrained modulus, with no fixed relationship between them. Due to some limitations imposed by the HS model, some combinations of these two moduli are not possible. In such cases, various modifications to the parameters are possible, where some compromises need to be made. Due to the loading conditions, where an existing levee is pushed by water loads, focus may be put more on shear-hardening than on compression-hardening, which is represented in the HS model by Young's modulus. Thus, the choice is made to set the constrained modulus equal to Young's modulus, which complies with the limitations. It should be noted that even though 102 parameters were identified, only 101 were varied, while one—the water level—was kept constant at the crown height because it is the main action on the levee and its variation should be included in later analyses anyway. This means that all of the factors' effects are, in this phase, determined only for the most critical water level. There may be some complex factors to analyze, one of them being the thickness of the top layer. The reason for its complexity is that an increase it may have a favourable effect when, e.g., the conductivity and Young's modulus of the top layer are larger than their foundation soil counterparts, but have an unfavourable effect in the opposite case. Such factors should either be isolated and analyzed separately, or left in the analyses and be kept in mind during the analysis of the results. The factors' effects should always be arranged such that the "all-low" run gives the lowest possible response and the "all-high" run the highest possible response.

To model the levee sections' response to the input variables, fully coupled flow-deformation analyses were conducted to analyze the stress–strain behavior, and safety analyses based on the SRM to find the factor of safety, both conducted with Plaxis 2D, 2019. Because many numerical analyses are required to make a factor screening experiment, Python programming was used to automatically generate the models, set calculation parameters, run analyses, gather results, and decide what to do in the next SB step. For the SB method, it is unknown in advance how many runs it will take to finish, which, among others, depends on the chosen Δ and factor arrangement—the factors should be sorted by assumed importance, with the most important being at the end of the list. In this study, the Δ was chosen as 5% of the difference between the maximum and minimum values of the response type of interest. Due to the SB's efficiency, this procedure was applied multiple times, each time observing a different response. The reason these had to be separate analyses is that a handful of factors have reversed effects for different response variables (model deformations, exit gradients, factors of safety). These SB analyses for identification of the most important factors were performed for the most complex and the most simple levee geometries, i.e., the cross section containing the impermeable core, both berms, and the top soil layer (Figure 4), and the cross section containing only the homogeneous levee body and foundation soil, without all the other components. In total, 6 SB analyses were performed, 3 for each cross section, and the resulting number of runs varied from as little as 20 up to 100.

4. Results and Discussion

With the six conducted SB runs, it was found that the parameters' individual effects varied up to over 90%. From the direct results of the analyses and the sum of all the single effects, is seems as though factor interactions reduce the individual effects when estimating the effect by how much the factors reduce the response, i.e., the sum of the effects of two factors is greater than the effect of those two factors acting together. If the model was indeed linear and interactions were unimportant, then superposition could be applied to get the lowest deformation from individual effects. However, this is not the case, and thus it is likely that some two-factor interactions are important to consider. Assuming the heredity assumption still holds, some of the identified factors must have an important interaction with another identified factor. The identified factors are shown in Table 4 (sorted by the effect on deformations). The table also shows for which response variable (mechanism) each factor is deemed important. As mentioned previously, the water level is not identified because it was kept constant at the crown height in these analyses. The Δ was set arbitrarily as 5% of the difference between the maximum and minimum runs' results.

By using the method of identifying incorrectly assigned factors described in Section 2, a few factors were identified for the different response variables, which were appropriately corrected for each analysis, which were then run again. Out of these factors, the most surprising were the friction angles of both the foundation soil and the top layer, which have reversed effects to the deformations, i.e., the higher friction angles yield higher deformations. After analyzing the specific factors, it is concluded that a possible reason for this behavior is the plastification of the foundation soils caused by the lower friction angles. Because the maximum deformations are in most cases found somewhere on the waterside part of the levee body, in the upwards and landside direction, the plastification of the foundation soils reduces the deformations at those places caused by the water pressure. Table 4 also shows the original ranges of the identified parameters found by analyzing the selected levee cross sections. These are the ranges which have been reduced by using the described methodology. Still, the results found in this study apply for the whole range, because the reduced ranges are all directly affected by the original ones. Levees whose geometric, physical, mechanical, and hydraulic parameters approximately vary from one bound to the other will have results similar to these, but if the values are only found within the ranges and the bounds are not actually part of the possible parameters, then the values will differ from the ones in this paper. However, the methodology can still be applied to find the parameters on which to focus for the levees found within an area of interest.

The identified parameters' importance, shown in percent effect, only holds for situations where all the other factors are at either end of their spectrum (most critical or least critical). This means that the percentages shown here are not fixed on a specific factor but can vary depending on the other factors.

It may look like there are some factors missing from the ones identified as important for the factor of safety, e.g., the strength properties of other materials. This is because, firstly, in the coupled flow–deformation analysis using the Hardening Soil model, plastic failure occurs on various occasions within the river and channel slopes. The safety analyses that come afterwards identify these positions as critical for stability failure, and a small failure surface develops at these positions. Those failure surfaces are not affected by the levee and core materials. Using such analyses instead of simple slope stability analyses where a failure region can be manually defined is limiting in this regard, but they provide the most realistic results, even though it may be harder to analyze the results and retrieve the wanted data. This, however, can be partly mitigated by creating other analyses where such section components are missing, so that failure surfaces cannot occur in these positions. Also, additional parameters can be added to the list of important parameters based on theoretical knowledge, experience, etc.

A total of 24 parameters were identified as important by the 5% effect criteria, where 13 (more than half) were geometric characteristics, 4 were mechanical, and 7 were hydraulic. It can be seen from Table 4 that individual factors have the most effect on the maximum exit gradient, where the single most important factor has an effect of over 95%, while the most important factor in the deformation response type has an effect of around 60%.

Additional analyses were performed for a geometry that includes only the levee body and a homogeneous foundation soil, and the results are shown in Table 5. It can be seen that the strength parameters are indeed important for the stability analyses, now that the slip surfaces cannot develop away from the levee. It can be noticed from the table that for this simple situation no singular factors are identified as important for the hydraulic gradients; however, it is clear that instead the interactions are high, as there is a large difference between the "all maximum" and "all minimum" runs.

Finally, Table 6 shows the subset of the input parameters that are deemed important for the analyzed levees, as well as others that fall within the selected ranges, after modifications are made to account for prior experience and theoretical knowledge. The table contains the factors identified in Tables 4 and 5, and also includes the minimum number of parameters needed to describe a levee component if they are not already identified by SB.

Table 4. Identified important factors with main effects.

Factors		Percent Effect [%]				Ranges	
	crown height from landside toe	52.9	59.5	96.7	24.0	0.5–7.3	[m]
LEVEE BODY:	vertical permeability	35.6	37.8	91.7	∼	1×10^{-8} to 5×10^{-3}	[m/s]
FOUND. SOIL:	Young's modulus	23.6	22.4	∼	∼	10–80	[MPa]
	water depth in the landside soil	21.3	12.0	79.4	12.0	0–5.5	[m]
CORE:	position inside the levee	20.6	24.3	∼	∼	−1 to 1	[%]
	crown height from waterside toe	18.8	20.8	∼	27.0	0.3–3.34	[%]
FOUND. SOIL:	unloading/reloading modulus	15.0	13.8	∼	∼	2–5	[MPa]
TOP LAYER:	vertical permeability	12.0	9.3	87.5	31.6	1×10^{-10} to 1×10^{-2}	[m/s]
LEVEE BODY:	waterside slope	9.7	5.8	∼	∼	0.18–0.55	[/]
CORE:	height	7.3	9.6	91.0	∼	0.3–0.9	[%]
LEVEE BODY:	Young's modulus	7.2	8.2	∼	∼	20–50	[MPa]
FOUND. SOIL:	friction angle	6.9	5.4	∼	19.8	21.9–42.4	[°]
FOUND. SOIL:	vertical permeability	5.5	7.9	94.3	∼	1×10^{-8} to 1×10^{-3}	[m/s]
	river depth	5.1	8.0	∼	∼	0.8–5.4	[m]
LEVEE BODY:	crown width	∼	6.0	91.7	∼	1.4–12.4	[m]
	inundation time	∼	∼	91.7	∼	1–6	[d]
CORE:	landside slope	∼	∼	80.5	∼	0.25–2	[/]
CORE:	vertical permeability	∼	∼	79.9	∼	1×10^{-10} to 5×10^{-5}	[m/s]
	channel width	∼	∼	32.9	∼	0.5–7	[m]
	duration of water event	∼	∼	19.4	∼	1–6	[d]
	river bank slope	∼	∼	∼	46.7	0.06–0.75	[/]
TOP LAYER:	thickness	∼	∼	∼	43.5	0.5–5.5	[m]
	channel depth	∼	∼	∼	35.2	0.8–2.4	[m]
	channel dist. from landside toe	∼	∼	∼	12.0	0–4.9	[%]
		D	Ux	G	F [1]		

[1] Response types: "D"—deformations; "Ux"—crown displacement; "G"—hydraulic gradient; "F"—factor of safety.

Table 5. Identified important factors with main effects for the simple geometry.

Factors		Percent Effect [%]				Ranges	
	crown height from landside toe	91.4	78.3	∼	43.0	0.5–7.3	[m]
LEVEE BODY:	vertical permeability	89.4	71.5	∼	40.2	1×10^{-8} to 5×10^{-3}	[m/s]
	crown height from waterside toe	83.2	37.4	∼	∼	0.3–3.34	[%]
	water depth in the landside soil	72.5	42.0	∼	52.5	0–5.5	[m]
FOUND. SOIL:	vertical permeability	70.0	42.0	∼	34.5	1×10^{-8} to 1×10^{-3}	[m/s]
LEVEE BODY:	landside slope	65.5	27.3	∼	∼	0.29–0.57	[/]
	inundation time	52.9	9.2	∼	∼	1–6	[d]
LEVEE BODY:	waterside slope	50.0	29.9	∼	∼	0.18–0.55	[/]
FOUND. SOIL:	anisotropy	34.7	∼	∼	∼	2–5	[/]
FOUND. SOIL:	friction angle	33.4	∼	∼	24.4	21.9–42.4	[°]
LEVEE BODY:	cohesion	27.5	27.0	∼	8.1	1–29.1	[kPa]
FOUND. SOIL:	unloading/reloading modulus	17.4	22.2	∼	∼	2–5	[MPa]
FOUND. SOIL:	Young's modulus	17.4	22.2	∼	∼	10–80	[MPa]
LEVEE BODY:	anisotropy	6.0	9.5	∼	∼	2–5	[/]
LEVEE BODY:	friction angle	5.2	∼	∼	∼	26–35	[°]
LEVEE BODY:	crown width	∼	16.3	91.7	∼	1.4–12.4	[m]
LEVEE BODY:	Young's modulus	∼	16.3	∼	∼	20–50	[MPa]
LEVEE BODY:	power m	∼	10.5	∼	∼	0.5–1	[/]
	duration of water event	∼	∼	∼	8.7	1–6	[d]
		D	Ux	G	F		

Table 6. Final selected important parameters.

Component	Parameter		
	Geometric	**Physical/Mechanical**	**Hydraulic**
LEVEE BODY	waterside slope crown width landside slope	Young's modulus cohesion friction angle power *m*	vertical permeability anisotropy
CORE	position inside the levee height landside slope crown width *		vertical permeability
TOP LAYER	thickness		vertical permeability
FOUND. SOIL		Young's modulus unload./reload. modulus friction angle	vertical permeability anisotropy
BERMS	landside berm width * landside berm height *		
GENERAL	crown height from landside toe crown height from waterside toe river depth river bank slope channel width channel depth channel dist. from landside toe water height *		water depth in the landside soil inundation time duration of water event

* Factors not identified by SB.

5. Conclusions

This paper presents a methodology for assessing the most important parameters affecting levee behavior within an area of interest, in terms of deformation, flow, and stability, but it is not limited to these mechanisms by any means. The methodology utilizes complex numerical analyses coupled with the sequential bifurcation method to identify these parameters. The first step requires the analyst to define the area of interest for which the analysis results will hold. Then a statistical analysis of all the desired levee parameters is done to identify the appropriate ranges, and then if needed, data transformation is conducted to modify the identified ranges. Analyses by SB are then conducted by using said numerical analyses and observing the desired response variable or variables. The results of analyses conducted by this methodology provide a list of parameters that should be focused on when analyzing the specific levees of interest and any levee whose parameters fall within the ranges specified by the analysts. These parameters (or factors) should not be confused with the parameters that in general are known—from theoretical knowledge, empirical data, and intuitively—to control any specific mechanism, but they are instead parameters which, for a large number of levees, are most important for the generalization of their behavior. They obviously do have some overlap, but failure to understand this point could lead to incorrect conclusions about the system. It should be noted that the methodology encourages adding parameters to the list that are not identified by the methodology but that the analyst deems important. The resulting list of important factors may be used for various purposes, some of which are: understanding the behavior of levees within the specific area of interest, creating models by using only the most important information, optimizing the field and laboratory investigations required for a specific project, and optimizing the control measures during and after construction. Results from this paper's case study are valid for any set of levee cross sections that are part of a larger set of levees, whose parameters can reach approximately the same values as the bounds specified in this paper. In any other case, the percentages would be misleading; instead, new analyses should be conducted following the same methodology. Further improvements can be made by defining the corresponding levee parameters as random variables with known distributions [18–22] and then conducting a variation of the SB method by Cheng [5], which allows for the output to be stochastic. This would surely affect the classification of the important parameters as their variations are taken into consideration, contrary to the analyses in this paper, where the inherent soil variation is considered negligible.

One of the biggest challenges regarding this methodology, besides the need for correct sign identification, is that some parameters change sign depending on other parameters' values. This means that it is impossible to assign a "correct" sign for these parameters' effects. If such parameters are known, they may be isolated and analyzed separately; otherwise, they are left in the analyses and should not have too much of an effect on the desired results.

Author Contributions: Conceptualization, N.R., M.S.K. and M.B.; methodology, N.R. and M.S.K.; software, N.R.; validation, N.R., M.B. and M.S.K.; formal analysis, N.R.; investigation, N.R., M.B., M.S.K. and L.L.; resources, M.S.K.; data curation, N.R.; writing—original draft preparation, N.R. and M.B.; writing—review and editing, L.L. and M.S.K.; visualization, N.R.; supervision, M.S.K. and M.B.; project administration, M.B.; funding acquisition, M.S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was co-funded by European Union Civil Protection Mechanism, under UCPM-2021-PP call, Grant Agreement Number 101048395, CROSScade project (Cross-border cascading risk management for critical infrastructure in Sava River basin).

Data Availability Statement: The input data used for numerical analyses in the study represented by input factors and their range bounds, as well as the output data represented by the maximum model deformations, horizontal displacements of the crown, maximum hydraulic gradients along the landside surface, and minimum factors of safety within the model, generated by coupled flow-deformation analyses and safety analyses, are openly available in Mendeley Data at DOI: 10.17632/n2zff5ts89.1, https://data.mendeley.com/datasets/n2zff5ts89/draft?a=c9a5974c-fcf5-4d76-84e1-45f9f42902b5 with CC BY 4.0.

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

Abbreviations

The following abbreviations are used in this manuscript:

ULS	Ultimate Limit State
SLS	Serviceability Limit State
SB	Sequential Bifurcation
SRM	Strength Reduction Method

References

1. Wolff, T.F. Reliability of Levee Systems. In *Reliability-Based Design in Geotechnical Engineering*; Phoon, K.K., Ed.; Taylor & Francis: London, UK, 2008; pp. 448–496.
2. Mirosław-Świątek, D.; Popielski, P.; Śliwiński, P.; Cwalina, T.; Skutnik, Z. Analysis of Factors Influencing Levee Safety Using the DEMATEL Method. *PLOS ONE* **2021**, *16*, e0255755. [CrossRef] [PubMed]
3. Montgomery, D.C. *Design and Analysis of Experiments*, 9th ed.; Wiley: Hoboken, NJ, USA, 2017.
4. Bettonvil, B. Detection of Important Factors by Sequential Bifurcation. Ph.D. Thesis, Tilburg University, Tilburg, The Netherlands, 1990.
5. Cheng, R. Searching For Important Factors: Sequential Bifurcation Under Uncertainty. In Proceedings of the Winter Simulation Conference Proceedings, Atlanta, GA, USA, 7–10 December 1997; pp. 275–280. [CrossRef]
6. Wan, H.; Ankenman, B.E.; Nelson, B.L. Controlled Sequential Bifurcation: A New Factor-Screening Method for Discrete-Event Simulation. *Oper. Res.* **2006**, *54*, 743–755. [CrossRef]
7. Wan, H.; Ankenman, B.E.; Nelson, B.L. Improving the Efficiency and Efficacy of Controlled Sequential Bifurcation for Simulation Factor Screening. *INFORMS J. Comput.* **2010**, *22*, 482–492. [CrossRef]

8. Oh, R.P.T.; Sanchez, S.M.; Lucas, T.W.; Wan, H.; Nissen, M.E. Efficient Experimental Design Tools for Exploring Large Simulation Models. *Comput. Math. Organ. Theory* **2009**, *15*, 237. [CrossRef]
9. Sanchez, S.M.; Wan, H.; Lucas, T.W. Two-Phase Screening Procedure for Simulation Experiments. *ACM Trans. Model. Comput. Simul.* **2009**, *19*, 7:1–7:24. [CrossRef]
10. Kleijnen, J.P. *Design and Analysis of Simulation Experiments*, 2nd ed.; International Series in Operations Research & Management Science; Springer International Publishing: Cham, Switzerland, 2015; Volume 230.
11. Shi, W.; Kleijnen, J.P.; Liu, Z. Factor Screening for Simulation with Multiple Responses: Sequential Bifurcation. *Factor Screen. Simul. Mult. Responses* **2012**, *2012*, 37. [CrossRef]
12. Shi, W.; Shang, J.; Liu, Z.; Zuo, X. Optimal Design of the Auto Parts Supply Chain for JIT Operations: Sequential Bifurcation Factor Screening and Multi-Response Surface Methodology. *Eur. J. Oper. Res.* **2014**, *236*, 664–676. [CrossRef]
13. Tadić, L.; Bonacci, O.; Dadić, T. Analysis of the Drava and Danube Rivers Floods in Osijek (Croatia) and Possibility of Their Coincidence. *Environ. Earth Sci.* **2016**, *75*, 1238. [CrossRef]
14. Bačić, M.; Lovorka, L.; Kovačević, M.S. Application of Geosynthetics as Hydraulic Barriers in Flood Protection Embankments. In Proceedings of the 1st International Conference CoMS_2017, Zadar, Croatia, 19–21 April 2017; p. 677.
15. Dean, A.M.; Lewis, S.M. Comparison of Group Screening Strategies for Factorial Experiments. *Comput. Stat. Data Anal.* **2002**, *39*, 287–297. [CrossRef]
16. Kleijnen, J.P.C.; Bettonvil, B.; Persson, F. Screening for the Important Factors in Large Discrete-Event Simulation Models: Sequential Bifurcation and Its Applications. In *Screening: Methods for Experimentation in Industry, Drug Discovery, and Genetics*; Dean, A., Lewis, S., Eds.; Springer: New York, NY, USA, 2006; pp. 287–307. [CrossRef]
17. Rossi, N.; Bačić, M.; Kovačević, M.S.; Librić, L. Fragility Curves for Slope Stability of Geogrid Reinforced River Levees. *Water* **2021**, *13*, 2615. [CrossRef]
18. Phoon, K.K.; Kulhawy, F.H. Characterization of Geotechnical Variability. *Can. Geotech. J.* **1999**, *36*, 612–624. [CrossRef]
19. Phoon, K.K.; Kulhawy, F.H. Evaluation of Geotechnical Property Variability. *Can. Geotech. J.* **1999**, *36*, 625–639. [CrossRef]
20. Wolebo, A.P. Advanced Probabilistic Slope Stability Analysis on Rissa Slope. Master's Thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2016.
21. Rossi, N.; Bačić, M.; Kovačević, M.S.; Librić, L. Development of Fragility Curves for Piping and Slope Stability of River Levees. *Water* **2021**, *13*, 738. [CrossRef]
22. Reale, C.; Kovačević, M.S.; Bačić, M.; Gavin, K. Cone Penetration Testing 2022. In *Assessment of the Spatial Variability of a Croatian Flood Embankment Using the Cone Penetration Test*; CRC Press: London, UK, 2022; pp. 1053–1057. [CrossRef]