TU Delft | Deltas, Infrastructures & Mobility Initiative

SmartPhones 4 Water

Return Level Analysis of Hanumante River using Structured Expert Judgment: A reconstruction of historical water levels

W.S. Brouwer[1], A. van Hamel[1], M. van Haren[1], P.E. Kindermann[1], R.P. Verboeket[1]
Dr.ir. G.F. Nane[2] Dr. T. Bogaard[3] and Dr.ir. J.C. Davids[4]

[1]Delft University of Technology, Master programme Civil Engineering

[2]Delft Institute of Applied Mathematics (DIAM), TU Delft

[3]Delft University of Technology, Department of Watermanagement

[4]California State University, Department of Water Resources and Agricultural Engineering

Abstract. Like other cities in Kathmandu Valley, Bhaktapur faces rapid urbanisation and population growth. Unsafe, new settlements are partly located at the floodplains and the government lags behind in implementing proper land-use policy to control unrestrained settlement. The rivers are not only constrained by uncontrolled settlements, but also by insufficient width and freeboard of bridges, and waste blockages causes problems. Combined with more extreme rain events during the monsoon due to climate change, flooding has become a reoccurring problem in Bhaktapur. To gain better understanding of the river and the corresponding flood risk, historical data is essential. Unfortunately, historical databases of water levels are non-existent for this river. Only starting from monsoon 2019, water levels and discharge have been measured on a regular basis. To reconstruct the missing historical data for a return level analysis, this research introduces the Classical Model for Structured Expert Judgment (SEJ) in combination with citizen science (CS). The objective of this research was to use Structured Expert Judgment in a flood risk analysis for the city of Bhaktapur. As a result of using SEJ, we were able to obtain sufficient water level data and estimate the return levels of extreme water levels of Hanumante river by fitting a Generalized Extreme Value distribution (GEV). This eventually led to a reverse Weibull fit, which in this case does not seem accurate. This research discusses in detail the advantages and issues of using Structured Expert Judgement in situations like this and also discusses the reliability of the results.

Keywords: Flood Risk, Kathmandu Valley, Structured Expert Judgment, Hydraulic Engineering

I. Introduction

IN July 2018, precipitation stations in Bhaktapur recorded the highest amount of rainfall ever documented in the last decade [1]. This extreme rainfall event caused Hanumante river to flood the entire area, affecting local population, blocking the transport system and leaving the city in disarray. Flooding has become the major problem in Bhaktapur in the recent years and the two biggest flood events were recorded in 2015 and in 2018 [1]. Bhaktapur is the third biggest city in Kathmandu Valley (Valley) and it is located in the eastern part of the Valley. Alike other cities in the Valley, Bhaktapur faces rapid urbanisation and a population growth rate of 2.3 percent [2]. With this, land use patterns are changing steadily. According to ICIMOD's study on land-use changes, the built area has increased from 16.9% in 1990 to 43.5% in 2010, an increase of more than 250% over 20 years [3]. Unsafe, new settlements are partly located at the floodplains and the government lags behind in implementing proper land-use policy to control unrestrained settlement. The river is not only constrained by uncontrolled settlements, but also insufficient width and freeboard of bridges and waste blockages causes problems. Combined with more extreme rain events during the monsoon due to climate change, flooding has become a reoccurring problem in Bhaktapur.

A. *Limited Data*

In order to do research on the probability of floods, it is important to completely understand the system. Unfortunately, the Hanumante river system is a relatively ungauged river system and little data is available. Historical databases of water levels are non-existent. Except of four precipitation stations that have been installed in by the Nepali Department of Hydrology and Meteorology (DHM), there are no hydrological stations to measure water level and discharge within the watershed of Hanumante river. Only starting from monsoon 2019, water levels and discharge have been measured on a regular basis, when the collaborating non-governmental organisation Smartphones4Water-Nepal (S4W)[1] got involved. S4W combines citizens, mobile technology, and young researchers to collect water data (e.g. rainfall, groundwater levels, water quality, etc.) as alternative for the struggling traditional approach that requires permanent sensors [4]. However, to gain a better understanding of the river system and the corresponding flood risk, historical data is essential. This situation is a challenging starting point that calls for creativity and solution-oriented thinking. It was therefore decided to find an alternative solution in the form of using Structured Expert Judgment (SEJ), which is a standard approach. We use the Classical Model for SEJ, a method which objectively evaluates and aggregates experts assessments [5]. As discussed in The Wisdom of Crowds by James Surowiecki, there is a lot of potential in group intelligence [6]. These promising considerations mark the motivation of using SEJ during this research by addressing both specialists in the field of water and citizens as experts. The objective of this research was to investigate the possibilities of using SEJ in a flood risk analysis for the city of Bhaktapur in terms of extreme monthly water levels and their corresponding return periods.

II. Study Area

We conducted this study in the Bhaktapur Municipality area from August 2019 to October 2019. The district of Bhaktapur is located twelve kilometers east of Kathmandu and is still located in the Valley (see figure 1). The town is enclosed by the Hanumante river to the south and Khasankhusang river to the Northeast. The city spreads over an area of 6.88 square kilometers at an elevation of 1,401 meter above sea level [7].

The Hanumante river is one of the tributaries of the Bagmati River, the main river in the Valley, and has a catchment area of 143 km^2, according [7]. The major water sources for the Hanumante river are rainfall and natural springs. It is the main natural river in the district of Bhaktapur and is the most important water source of the city. Besides, it is of great ecological, cultural and religious importance. The Hanumante river has multiple tributaries with their own sub-basins [7]. In pre-monsoon months, the river can be almost dry in some areas, while it can transform into a broad and fast-flowing river during monsoon months, with water levels up to two to five meters. According to Rajaram Prajapati et al. (2018), the river shrunk from six to two meters since 1964. In the past, the lower parts of the area were only used for agriculture. Now, many people have built their houses in the low lands close to rivers and some are even located within the river's flood plains [1]. This resulted in a significant increase of the flood risk in recent years. Starting from June 2019, S4W decided to measure water levels in Hanumante river on a daily basis. The only daily water level data that exists covers the 2019 monsoon months: June, July and August. For Hanumante river, there are three sites that are daily measured, named HM01, HM04, and HM06. Additional to the daily water level measurements, also monthly discharge measurements by using a SonTek Flowtracker are available.

[1] https://www.smartphones4water.org/

III. Methodology

In order to carry out a flood risk analysis, we divided our research in two main parts. First, we used SEJ to obtain the yet unknown water level time series of monthly maxima. Secondly, the resulting time series were analysed to determine its Generalized Extreme Value (GEV) distribution in order to provide an estimation of the return levels of the Hanumante river.

A. *Structured Expert Judgment*

We have used the Classical Model (CM) for SEJ to create a time series of the monthly maximum water levels during the monsoon months for the period: *1990- 2019*. CM is a rigorous method for SEJ that evaluates expert opinion based on two objective measures, statistical accuracy and informativeness and uses these measures to aggregate assessments [8]. The method is widely accepted and has been used in fields like the nuclear sector, chemical, and gas industry [9]. Within the CM, experts are asked to assess their uncertainty for quantities of interest, denoted as target variables/questions. Moreover, the experts also provide uncertain assessments for quantities which are not known to them, but are known to the analysts; these are referred to as seed questions/variables. Instead of providing a probability distribution, the experts are asked for 3 quantiles of the distribution, that is, the 5th, 50th, and 95th quantile.

The target questions refer to the uncertain and unknown monthly maximum water levels for June, July and August for the timespan of interest. The seed variables were questions about water levels that have been measured by S4W. The seed variables were used to give the experts a weight based on their performance. Afterwards, these weights were used to weigh the experts assessments at the target questions. The experts have been informed which questions were the seed variables and which questions were the target questions.

The experts' assessments have been evaluated by two measures of performance: *statistical accuracy* (or calibration score) and *information score*. The calibration score is a p-value of a statistical hypothesis test that the realizations (the real values for the seed variables) correspond statistically with the expert's assessments [5]. Suppose, an expert is asked one hundred seed questions. Statistically, it is then expected

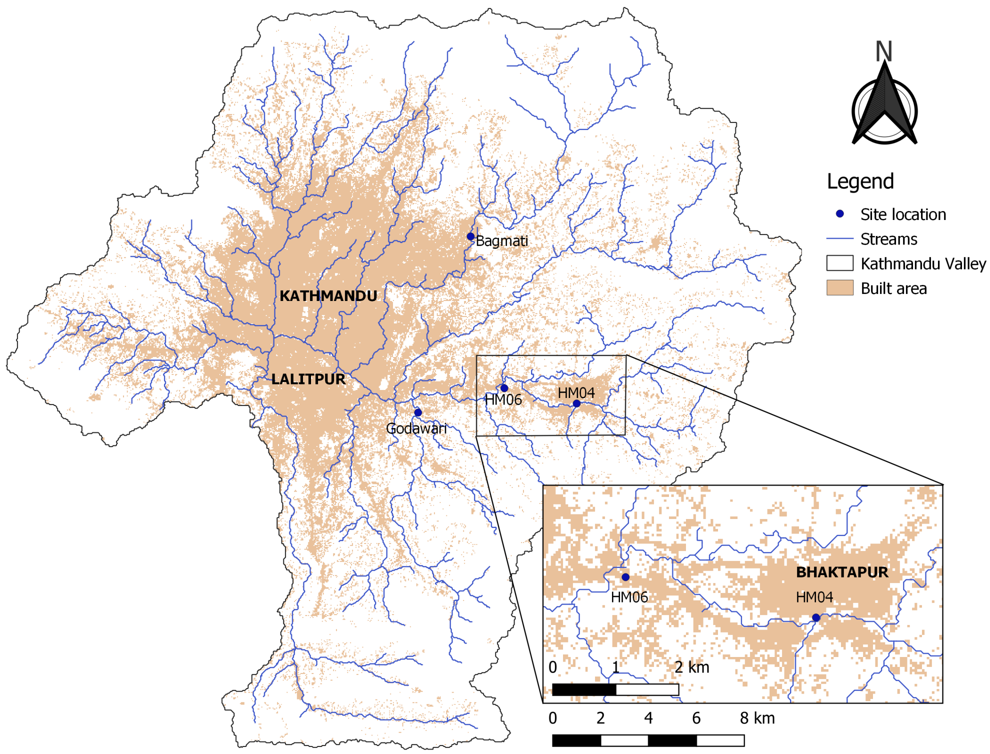

Fig. 1. Overview Kathmandu Valley and Bhaktapur. HM04 and HM06 have been used in seed variables

that five times the realization will fall below the 5% quantile assessments and five times the realization will fall above the 95% quantile assessments. For the other ninety questions the realizations should statistically fall in the second quantile (ranging from 5% to 50%) and the third quantile (ranging from 50% to 95%) evenly. The more the expert deviates from these expected frequencies, the lower his/her calibration score will be. Formally, for computing the calibration score, an expert is treated as a statistical hypotheses, where the calibration score is the result of a hypothesis test. The null hypothesis (H_0) can be defined as follows: "The inter quantile interval containing the realization for each variable is drawn independently from the probability vector p, where $p = (0.05, 0.45, 0.45, 0.05)$" [8] [9]. Informally, the null hypothesis can be formulated as: "The expert is well calibrated" [10].
The calibration score is determined considering all seed variables. If an expert is perfectly statistically accurate he or she will receive a calibration score of 1. The less statistically accurate the assessments of the expert are, the lower the calibration score. The calibration score ranges from 0 to 1 where the lowest calibration score is 0.

The information score is a measure for the concentration of experts' assessments with respect to a background measure, which can be a uniform or a logarithmic uniform distribution. We will use the uniform distribution as the background measure, meaning that every assessment is treated as equally possible. So, the relative information of experts' assessments with respect to the uniform background measure determines the information score [10]. The information score is determined per seed variable and target question. An overall information score is computed by averaging the information score over all seed variables.

The Decision Maker

The overall performance of every expert's assessments is characterised by a combined score which is the product of the information (I) and calibration (C) score. This combined score will provide a normalized weight (w) for each expert.
Finally, a weighted combination of experts' distributions leads to a so-called performance-based Decision Maker (DM). Experts' distributions are aggregated for the target questions, but can be also aggregated for the seed variables. The resulting DM's assessments can be assessed with respect to the calibration and information score, just as any expert's assessments. The two scores can be used to evaluate the performance of the DM. Other weights are possible, for example, one can consider equal weights, which lead to the so-called equal weight DM. The weights in the performance-based DM result from accounting for the calibration score as well as for the overall information score. Accounting for the information score for each seed variable leads to item weights, which lead to an item weight DM.

Experts can be removed from the pool, using the calibration score criterion and the performance of the resulting DM can be assessed. The DM which leads to the highest combined score is called the optimal DM.

A more elaborate description of the scores and method, including formulas, can be found in *Supplementary information for Return Level Analysis of Hanumante River using Structured Expert Judgment: reconstruction of historical water levels*.

Expert selection & Questionnaires

The Hanumante river is a small river that was not studied much by hydrologists. Only in recent years, the river has attracted attention because of problems with urban flooding, due to habitation of the flood plains. This lack of knowledge made it difficult to select experts. We therefore assumed that people living or working close to the river, could be just as well an expert as hydrologists and engineers. It was decided that the list of experts should consist of a variety of people: citizens that live close to the river, young S4W scientists who regularly measure the rivers in the Valley, and hydrologists and engineers working for several governmental institutions related to water. This variety in background of the experts made it challenging and very important to create a questionnaire that was easy to understand and feasible for every expert.

As mentioned before, the aim of the Structured Expert Judgment was to create a time series of extreme monthly water levels of Hanumante river, that could be used for a return level analysis. In order to draw reliable conclusions from a time series regarding return levels, it is important to have a time series that is sufficiently long. It was therefore decided that the time series should cover the monsoon months June, July and August for the period 1990 - 2019, which is equal to at least ninety target questions. This would mean that experts had to answer one hundred questions, including the seed variables. We concluded that this was too much when the expert is expected to give thoughtful responses for every variable, especially given the missing scientific background of some experts. Therefore, we have split the target questions over four groups and assigned experts to each group. In appendix B an overview of the different questions asked to the different groups is presented.

To validate whether the responses of one group would be significantly different compared to the three other groups, we included ten overlapping target questions that all groups had to answer. These overlapping questions were later treated as one group and the computed DM of this dataset was then compared to the DM's of the individual groups to conclude whether or not there were clear distinctions between the responding groups. Eventually, every questionnaire consisted of ten seed variables, 28 or 31 target variables (including overlapping questions).

Because of the unfamiliarity of Structured Expert Judgment in Nepal, the importance of explanation and communication of the method was regarded as a key factor for the executed interviews. Most of the experts' assessments were obtained by visiting the experts at their working place. An elaborate explanation of the method was given through translations by the people of S4W-Nepal, who were accompanying the visits. An example question was also provided to assure that the experts understand the procedure in order to answer the target and calibration questions. Some of the specialists gave their responses via an online survey. We informed these experts about the method beforehand. The aim was to have at least ten assessments by specialists in the field of water within this SEJ. The surveys and questionnaires, that were divided into four groups, were equally distributed among the specialists to obtain at least two specialists' assessments for every group.

Seed Variables & Target Questions

The lack of data made it difficult to select enough seed variables, given that only three monthly maxima were recorded yet at three locations of the Hanumante river; HM01, HM04, and HM06 respectively. Thus, focussing on one site in the SEJ could not provide sufficient seed variables. Moreover, not every site was considered recognizable enough for the experts. Luckily, S4W-Nepal provided us with a few more locations where the water level was well-recorded for at least a month. However, these recordings were for different rivers and different periods. We chose to focus on four different sites in the Valley that, together, would provide eight of the total of ten seed variables. The most recognisable site, HM04, would feature as main research point for which return levels would be calculated. HM06 and two sites, in the Bagmati river and Godawari river respectively, were added as the other three sites for the seed variables. These different sites all featured water level staff gauges that were attached to familiar bridges that could be used as local reference points for the questionnaires, see figure 1 for the final locations that were used. For the final two missing seed variables, we needed to consult the media. The Hanumante river floods of 2018 and 2015 were recorded graphically in newspapers and on the internet. Eventually, some imagery could be traced back to HM04 and HM06. The corresponding high water level as seen in the videos and pictures were then measured using Precise Levelling, taking the water level staff gauge at the nearby site as reference point. These efforts led to a total of ten seed variables in the end, which are considered sufficient to evaluate the performance of experts' assessments. We aimed to make the questions as clear and as short as possible and we provided the experts with enough background information to estimate the water levels. The background information given in the questionnaires consisted of (1) a map and description of the site location, (2) a picture of the bridge at the site location, (3) the average water level during the monsoon, and (4) the water level at which the bridge would be inundated. This information was obtained partially from the ODK data and partially from site visits and field measurements. Eventually the seed variables the target questions were constructed as: "What was the highest water level in [month][year]?". An example of the questionnaires can be found in the *Supplementary information for Return Level Analysis of Hanumante River using Structured Expert Judgment: reconstruction of historical water levels*.

Software

The software we used to perform SEJ was ANDURIL. ANDURIL is a MATLAB toolbox which has many functions which includes an extensive list of functions for the CM of

SEJ, including the performances of several DMs. [9].

B. Return levels

The SEJ, as explained above, resulted in a time series of maximum monthly water levels, created by the chosen DM for every target question. It was necessary to convert these monthly maxima to yearly maxima, by taking the maximum of the three monsoon maxima that were estimated by experts in a specific year. The final objective was to determine the return levels and corresponding return periods, in which the return periods should be expressed in years. The next step was to fit a GEV to the obtained estimates and to extrapolate these with respect to their probability of occurrence. The underlaying relation between GEV's, return water levels, and return period is shown in formula 1.

$$G(Z_p) = 1 - p \tag{1}$$

In which G is the chosen GEV, Z_p is the return level and the return period is defined as $1/p$, where p is the probability of occurrence of a water level Z_p.

In general, there are three types of possible distribution within GEV, for extreme (maxima) values: Gumbel (type I), Fréchet (type II), and reverse Weibull (type III). Each GEV type is characterised by a location parameter μ, a scale parameter σ and a shape parameter k. For $k = 0$ the GEV is the Gumbel distribution, for $k > 0$ it is the Fréchet distribution, and for $k < 0$ it is the Weibull distribution [11].

MATLAB features the built-in function 'GEVfit', which provides maximum likelihood estimates of the location and scale parameters. We used these parameters to plot the inverse of the GEV, using MATLAB's built-in function GEVinv, showing exactly which return water levels can be expected for which return periods.

The flood risk is defined here as '*the expected damage due to a certain water level multiplied by the probability of occurrence p of that water level*'. The lower the acceptable probability p, and thus the higher the return period $1/p$, the safer the area. Consequently, the flood defences surrounding that area can be designed based on the corresponding (acceptable) return water level. Therefore, when an area is statistically allowed to be flooded only once every hundred years, the flood defences should be at least as high as the water level that corresponds to this return period of a hundred years.

IV. Results

A. Questionnaires

We conducted a total of 62 questionnaires in September 2019 in the city of Bhaktapur during fieldwork excursions and through an online survey. The details of the respondents and their functions/affiliations can be found in table I. The majority of the respondents were citizen of Bhaktapur, meaning that they live and/or work close to the Hanumante river. Most of these people were shop owners and school teachers. From the table, it can also be observed that the 'specialists' were well divided over the different groups. We ourselves also took part in the survey.

TABLE I
OVERVIEW OF THE RESPONDENTS

		Questionnaire			
Function	**total**	**1**	**2**	**3**	**4**
Number of participants	**62**	16	16	16	14
Number of specialists	**13**	3	2	4	4
Number of citizens	**45**	12	13	11	9
Number of students	**4**	1	1	1	1
Average age	**35.8**	39.6	32.8	34.5	36.3
Male/Female/Unknown	**40/19/3**	8/8	10/5	11/4	11/2

B. Structured Expert Judgment

In this section the results of the SEJ are presented for every group, as well as for all 62 experts (which will be referred to as the 'combined group'). For every expert, the calibration score, information score, combined score and normalized weight, when the optimized DM's were computed, are presented in appendix C.

1) Group 1: The calibration scores for the experts within group 1 range from $6.13 \cdot 10^{-13}$ to 0.395 (see table VIII in appendix C). An optimized performance-based combination of weights leads to an α value of 0.395 in which only one expert received a non-zero weight. This expert turned out to be a local shop owner in Bhaktapur.

We also computed DM's based on global-, item-, and equal weights. The results of the different DM's are summarized in table II. We note first that all DMs except the Equal weight DM obtain a calibration score higher than the significance level of 0.05. Furthermore, the highest informative DM is the optimized DM, which is also the DM with the highest combined score. Concluding, the optimized DM was the best performing DM for group 1 and therefore we decided to continue with the results of the optimized DM for the target questions.

TABLE II
PERFORMANCE OF THE DIFFERENT DECISION MAKERS FOR GROUP 1

	Calibration. score	**Information score**	**Combined score**
Optimized DM	0.3946	0.8287	0.3270
Global weight DM	0.2894	0.2717	0.0786
Item weight DM	0.4735	0.4038	0.1912
Equal weight DM	0.0012	0.1405	0.0002

2) Group 2: Within group 2, the calibration scores range from $6.17 \cdot 10^{-9}$ to 0.061. From table IX in appendix C, it can be observed that the assessments of the experts of group 2 are less statistically accurate than the assessments of the experts in group 1. We also computed the optimized DM for group 2. With an optimized α-value of 0.047 five experts were granted a non-zero weighting.

The results of the different computed DM's are summarized in table III. It is remarkable that the calibration scores of all DM's are quite low, but still above the 0.05 siginificance level threshold, which is the result of the low calibration scores of the experts. Another remarkable observation is that the

DM based on item weights has a slightly higher combined score compared to the optimized DM. From this, we could conclude that for group 2 the DM based on item weights is preferred over the optimized DM. However, as presented from other studiets, a DM based on item weights starts working properly only when experts had a good training in probabilistic assessment [10]. A training most of our experts never had. Therefore, it was decided to continue working with the optimized DM during this research, also since the the difference in weight between the optimized DM and item weight DM was low.

TABLE III
PERFORMANCE OF THE DIFFERENT DECISION MAKERS FOR GROUP 2

	Calibration score	Information score	Combined score
Optimized DM	0.2894	0.3924	0.1136
Global weight DM	0.1242	0.4819	0.0599
Item weight DM	0.2441	0.4974	0.1214
Equal weight DM	0.0031	0.1475	0.0005

3) Group 3: In the third group, the calibration scores for the experts ranged from $6.13 \cdot 10^{-13}$ to 0.036 (see also table X in appendix C). Compared to the first and second group, the highest individual calibration score for group 3 was relatively low. When we computed the optimized DM an α-value of 0.0063 was found, resulting in three experts with non-zero weights.

The results of the different DM's are again summarized in table IV. It can be observed that the calibration score for the optimized DM is still relatively low (0.2441), but much higher than any of the expert's calibration score. However, compared to the calibration scores of the other DM's, the value was significantly better. Consequently, the optimized DM obtained the highest combined score.

TABLE IV
PERFORMANCE OF THE DIFFERENT DECISION MAKERS FOR GROUP 3

	Calibration score	Information score	Combined score
Optimized DM	0.2441	0.3739	0.0913
Global weight DM	0.0357	0.6502	0.0232
Item weight DM	0.0357	0.6502	0.0232
Equal weight DM	0.0012	0.1542	0.0002

4) Group 4: The calibration scores for the experts within the final group ranged from $1.29 \cdot 10^{-10}$ to 0.493 (see table XI in appendix C. When we computed the optimized DM, an optimized α-value of 0.493 was found. Considering this value, only one expert received a non-zero weight. Just as in group 1, this was a local shop owner from Bhaktapur.

The results of the other computed DM's are summarized in table V. Similar to group 1 and 3, the optimized DM was performing best when compared to the other DM's. Again notice the low calibration and information score for the Equal weight DM.

5) Combined group: For the data set consisting of all the 62 experts, we again computed the four different DM's. In table VI, the corresponding scores are presented. It can be observed that the optimized DM performed best with a relatively high calibration score of 0.6828. For the optimized DM, the α-value was equal to 0.3946. This resulted in the fact that only two experts received a non-zero weight. Not surprisingly, these were the same experts that were selected for group 1 and 4 for the optimized DM's. This can be explained by the fact that the value of α determines which expert will receive a non-zero weight, based on his/her calibration score. Since the calibration score only dependents on the answers in the seed variables, the calibration scores of the experts did not change. Once again, the results show the poor performance of combining experts'assessments using equal weights.

TABLE V
PERFORMANCE OF THE DIFFERENT DECISION MAKERS FOR GROUP 4

	Calibration score	Information score	Combined score
Optimized DM	0.4926	0.4848	0.2388
Global weight DM	0.4926	0.3015	0.1485
Item weight DM	0.4926	0.3109	0.1531
Equal weight DM	0.0237	0.1489	0.0035

TABLE VI
PERFORMANCE OF THE DIFFERENT DECISION MAKERS APPLIED ON THE DATA SET WITH ALL 62 EXPERTS

	Calibration score	Information score	Combined score
Optimized DM	0.6828	0.5644	0.3854
Global weight DM	0.2894	0.3542	0.1025
Item weight DM	0.4735	0.4671	0.2212
Equal weight DM	0.0012	0.2487	0.0003

Comparison of groups To see whether one group performed better with respect to the other groups, we compared the calibration scores, information scores and combined scores of the optimized DM's. The results are presented in figure 2. The figure shows that the optimized DM based on all experts performed best compared to the other optimized DM's, the calibration scores and combined scores. Therefore, we assumed that the final estimates of this DM for the target questions were the most reliable. A remarkable aspect was the low calibration scores and weights of group 2 and 3.

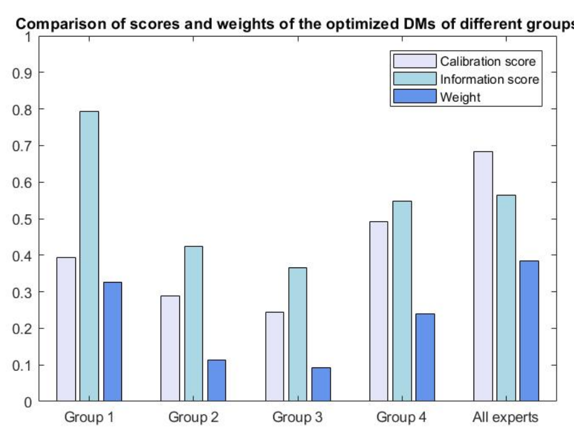

Fig. 2. A comparison of the optimized Decision Makers of the different groups

These were the result of the low calibration scores of the individual experts within those groups. We selected the best performing DM per group, resulting in the optimized DM for group 1, 3, and 4. Since the difference between item weight DM and optimezed DM for the combined score in group 2 is not that significant, we have used the optimal DM for consistency purposes.

Moreover, we checked whether there was no group significantly over- or underestimating. We did this by comparing the final assessments of the optimized DM's of the four groups with the final assessments for seed variables of the optimized DM of the combined group (based on all 62 experts). Generally, the result of an optimized DM always contains an assessment of the $lowerbound$, 5%, 50%, 95% and $upperbound$ quantiles of the water levels per seed variable and target question. The best estimate of a DM for item i can be assumed to the be the value of the $50^{th}\%$ quantile of that assessment. We were able to compare the estimates of the different groups on the seed variables as well as on the ten overlapping target questions. For the seed variables, we also compared the assessments of the DM's with the realizations.

The results of the comparisons are presented in figure 3 and figure 4. For the seed variables, it can be observed that no group is constantly over- or underestimating, compared to the realizations and the optimized DM's based on all experts (the combined group). Based on this analysis, we decided to combine the water levels estimated by the four groups for different years to one data set containing all water levels of the years 1990 till 2019.

C. Maximum water levels

When we combined the data of the different groups, this resulted in monthly maximum water levels for the monsoon months for the years 1990-2019. Out of all 90 monthes, there were nine months, separated over three years, that were estimated by all 62 experts. For those months we used the assessments of the optimized DM's based on the complete group of the 62 experts (since this DM has the highest calibration score). For the other months, the assessments that were obtained by the optimized DM's per group were used. As mentioned before, to determine return levels, yearly maximum water levels are needed. These were determined as the maximum of the three monsoon months of a year. The results are presented in figure 5. The final estimates of the water levels are given as the black line, being the 50% quantile estimates assessed by the DM's. To emphasize the large uncertainties of the final assessments of the DM's, the 90% confidence interval is also given by the upper and lower limit provided by the 95% respectively 5% quantile values. The maximum water level of 2019 was not obtained from SEJ since water level measurements of Smartphones4Water were available for the monsoon of 2019. According to expert data, the mean maximum water level for 1990-2019 is 2.57 meter. The highest maximum water level was 4.0 meters and occurred in 2017. The lowest maximum water level was 0.4 meter and occurred in 1992.

D. Return Level Analysis

With the time series now available, it is possible to fit a GEV distribution over the distribution that results from the yearly maxima. The resulting best-fitting GEV and its characterising parameters are presented in figure 6. The numerical value of the shape parameter, k, is negative (-0.455), which means that the corresponding GEV is a type III extreme value distribution, also referred to as reverse Weibull. The location and scale parameters are respectively: 2.36 and 0.81.

To obtain confidence intervals for the return levels, we also fitted a GEV distribution for the 5% and 95% quantile estimates of the water levels. The results of the fitted GEV's are presented in appendix D.

By taking the inverse of the fitted GEV, we were able to calculate the return periods corresponding to the extreme water levels, including the confidence intervals. In figure 7 the results are presented. The black line represents the final return levels, the 90% confidence interval is also shown. It can be observed that the confidence interval is very shaded, which results from the large uncertainty in the DMs of the SEJ. We found that a water level of 3.25 meter has a return period of five years. A water level of 3.51 meter has a return period of ten years and finally, a water level of 3.84 meter would statistically occur once every fifty years.

V. Conclusion

By using SEJ, we were able to reconstruct historical water levels for the Hanumante river (see figure 5). Still, the question remains whether these water levels are reliable. Since there is almost no historical water level data, the correctness of the reconstructed water level data is hard to validate. However, we know that there were some extreme rain events that occured in the years 1990, 2015, and 2018 [1], for which it is assumed that they led to extreme water levels. Looking at figure 5, the highest estimated maximum water level occurred in 2017 and was 4.0 meter, while the years for which higher water levels were expected did not show higher water levels at all. Although high, the water levels of 2015 and 2018 completely blend in with the other extremes and the 1990 extreme water level was even estimated to be the second lowest value (0.46 meters). Based on this, we could conclude that experts were not able to remember water levels of specific years (keeping in mind that the measurements for 1990 fully relied on the conservative estimates of two experts).

Another issue in the obtained results were the large differences in the final optimized DMs for the different groups for the overlapping questions. Although there were no groups consistently over- or underestimating, it can be observed that the differences in the final assessments were quite large.

On the other hand, there is evidence that the order of magnitude of the estimated water levels is correct. As mentioned, the water levels for monsoon of 2019 were measured. Up to the moment of writing, October 2019, the maximum water level that was measured in the Hanumante river by S4W during the 2019 monsoon was 2.4 meters. With respect to the water levels obtained by SEJ this value is comparable to the average of the

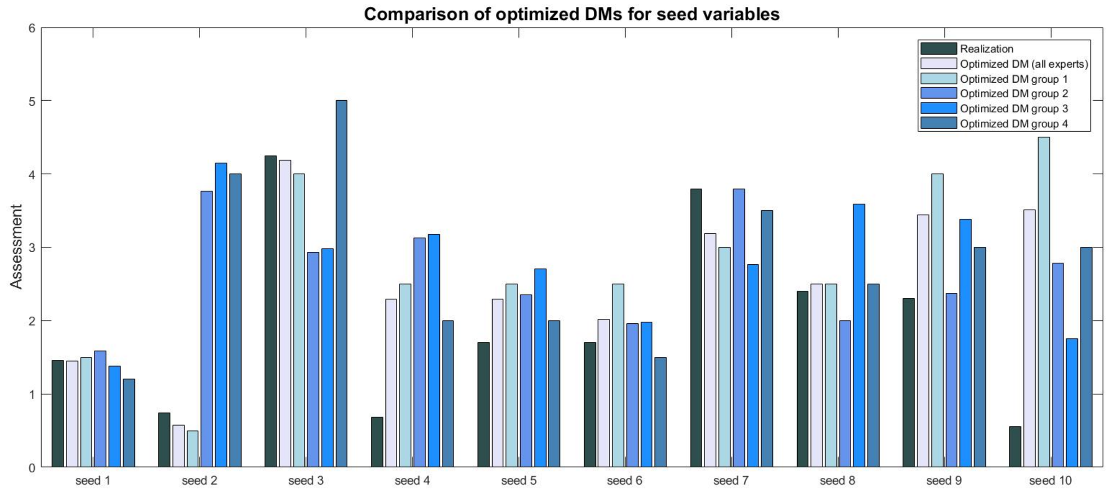

Fig. 3. A comparison between the optimized DM's on the seed variables

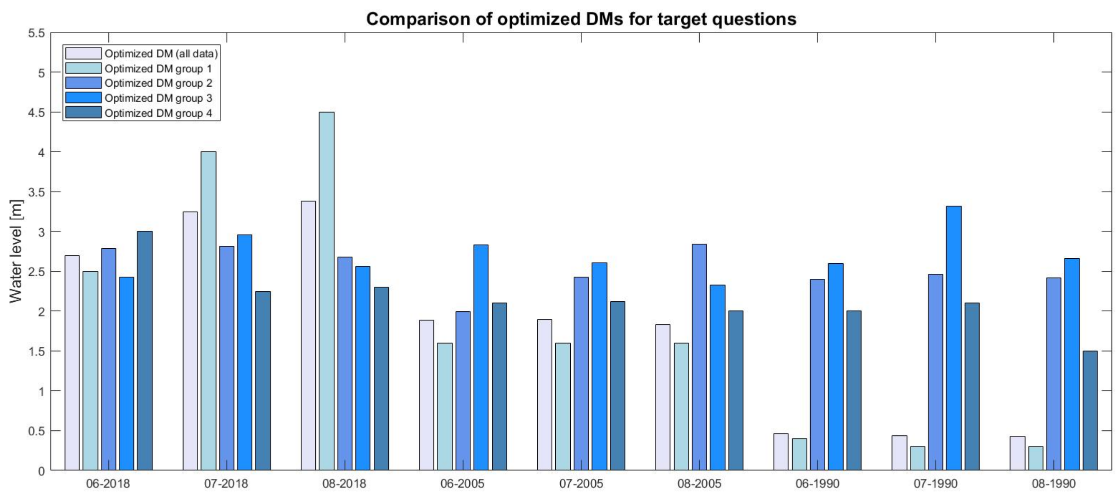

Fig. 4. A comparison between the optimized DM's on the target questions

water levels between 1990 and 2018 which equals 2.57 meters. This is also in line with our expectations, since, based on the precipitation data from 2019 this monsoon seems slightly less heavy than average. From this, we concluded that the order of magnitude of the yearly water levels obtained with SEJ can be regarded as fairly accurate.

There was one more remarkable conclusion that became clear from the results. Namely, the fact that specialists did not necessarily perform better than citizens. The two experts that the optimized DM of the combined group relied on, were both citizens of Bhaktapur. A possible explanation could be that specialists are more certain about their estimations, which results in smaller confidence intervals, which denotes overconfidence that might result in exclusion.

Concluding, it is possible to use estimations of both citizens and specialists to fill in historical data gaps of water levels in the Hanumante river by using SEJ. However, the question remains wether or not these results are accurate. While the exact extreme water levels of specific years could not be found by this method, the results might still be useful, regarding the global range of values for the water levels and their return periods. We therefore conclude that the method definitely has potential when some improvements would be made, as will be elaborated in further detail in the discussion. Since there were no other opportunities to provide ourselves with better data, we decided to continue working with the results of SEJ under the condition that the uncertainty should be emphasized. In this case, SEJ was the only possibility we had to obtain historical water level data.

Next, We were able to construct a time series of maximum water levels for the years 1990-2019. As already stated, in this case the data set was considered useful enough to be analysed with respect to return periods.

The time series of these yearly water levels, as shown in

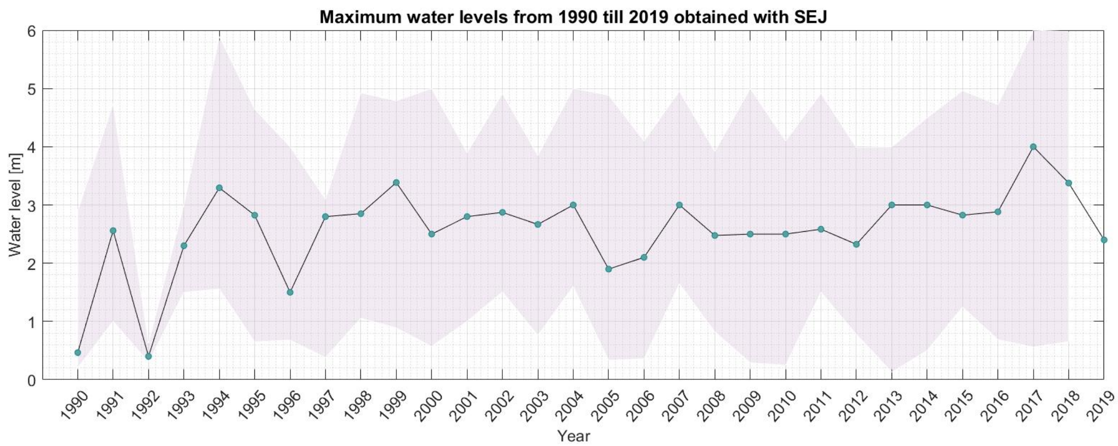

Fig. 5. The maximum water levels obtained with Structured Expert Judgment for 1990-2019 (black line), with 5% and 95% quantile estimates (shaded)

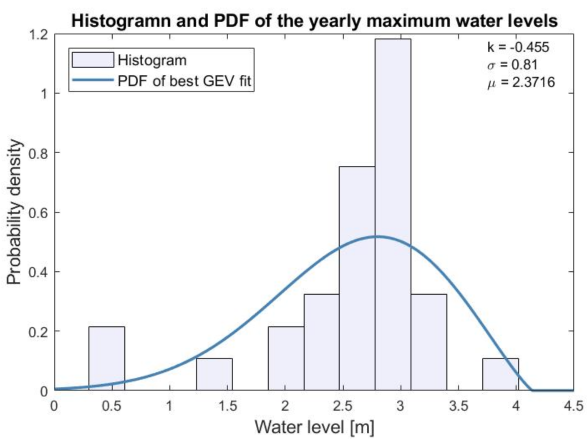

Fig. 6. The probability density function according to the GEV

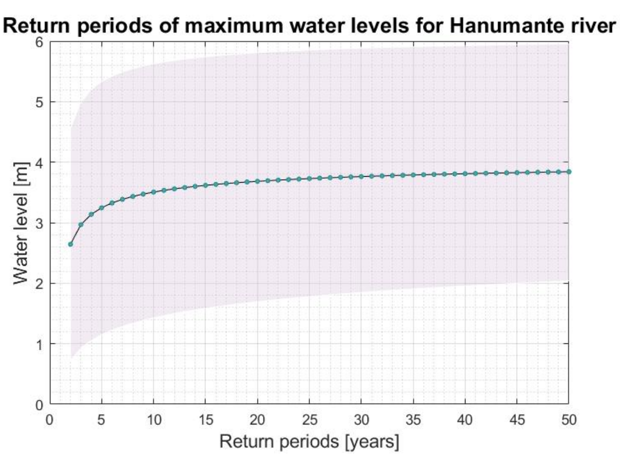

Fig. 7. Return periods for the water levels at the Hanumante river (black line), together with the 90% confidence intervqal resulted from DMs 5% and 95% quantiles (shaded)

figure 5, showed more downward peaks than upward, from which the most obvious peaks can be found in 1990, 1992, and in 2017 with water levels of 0.46 m, 0.4 m and 4.0 m respectively. As a result, the Probability Density Function (PDF) of the fitted GEV has a tail towards the lower values and aborts quite abruptly at the maximum of 4 meters, which is similar to the shape of a type II (reverse Weibull) extreme value distribution. However, it was expected that the best GEV was more likely to be an extreme value distribution type I (Gumbel), since this one is often used for extreme water levels [12]. With the reverse Weibull GEV, the return period graph shows a horizontal asymptote for the value of 4.0 m, suggesting that the Hanumante river would never exceed this water level and that all flood defences should be up to this level to guarantee full safety. The same observation could be made when basing the return periods on the 95% quantile values, this time with an asymptote for the value of 6.0 m. For the results obtained by the 5% quantile, it was more difficult to say whether it has an asymptote or not. Obviously, it is physically not possible that water levels have an upper bound (e.g. 4.0 or 6.0m). So, even though the order of magnitudes of the water levels were promising at first, the underlying distributions were not. The reason for this non-correct fitted GEV, is the histogram obtained from SEJ that was different than expected.

We should conclude that within the scope of this research, the water levels of Hanumante river, obtained with SEJ can be used for the determination of return levels, but with a very high uncertainty and under the condition that the return levels are only used to indicate the order of magnitude of occurring water levels.

VI. Discussion

One of the main issues we encountered when using SEJ was the fact that the optimized DM's were based on only one person for some of the groups. When these persons would not remember certain years with extreme rainfall, the final result would thus exclude an extreme water level. Moreover, other people could remember this extreme rainfall event but did not answer the seed variables correctly and thus obtain no calibration score. When there is such a big difference in individual scores, the risk exists that informative assessments get lost when optimized DM's are used.

Also, during the field trips we observed that experts took a long time to fill in the questionnaire. The average duration of a questionnaire turned out to be around thirty minutes and some of the experts could not bring up the time and effort to give thoughtful responses until the final question. Consequently, the answers for the final years, say 1995 till 1990, did not show a lot of variation. People started repeating their guesses for these consecutive years. Looking back at the field trips, it might have been more sensible to not overwhelm the experts with such a vast questionnaire, but instead discuss with them the implications of the research and the importance of their recollections of the extreme water levels. Then, if experts were still willing to help, the questions might be answered more thoroughly. Now, the essence of the method was lost in translation and it was clear that some experts struggled with the answers and their corresponding confidence intervals.

Another problem with translation was the calendar that was used. In our questionnaire we used the Gregorian calendar instead of the Nepali calendar. This caused confusion since the Nepali month Asar equals mid June to mid July, and also the counting of years is different. The Greogorian year 2019 equals the Nepali year 2076. Using the Nepali calendar from the beginning on, could have resulted in less confusion.

Furthermore, we assumed that there would be no significant changes in results when other rivers would be considered and therefore we used four different sites for the seed variables. However, it could be deducted that the experts could not so easily answer questions for other location than Hanumante river. Although we provided the experts with average water levels and some other references for every location, they still seemed rather doubtful and had great difficulty answering for the locations that were not in Bhaktapur. With only ten seed variables these kind of issues have a major effect on the scores of the experts.

Lastly, the reliability of the time series is a major point of discussion. By looking at the overlapping years, we could conclude that there are sometimes larges differences in the final assessment of the different groups. These large differences in the overlapping years, might indicate that the resulting time series would have looked much different if the other years were divided differently over the groups. This fact gives rise to doubt about the resulting time series, and its applicability for the return level analysis, as discussed in the Conclusions.

VII. Recommendations

After conducting this research we think that the method has potential, but that there are several opportunities to improve the application for a SEJ in a situation like this. We would therefore recommend to consider the following aspects for any future research including SEJ in combination with SC:

Firstly, the experts should be better prepared for the questionnaires. An elaborate (oral) explanation of the method is extremely important. Especially the importance of the confidence intervals should be well explained. Moreover, it is advisable to provide the experts with even more background information. It would be useful to give the value of one recent monthly maximum water level. The experts could use this value to refer months of the past to a month that they remember and to understand how much the maximum can deviate from the mean.

Secondly, it could be useful to ask the experts to start by estimating the water levels of months in which they remember that a flood or very high water level occurred. In this way, it is avoided that experts oversee to assess those years with relatively high values while they are working themselves trough the long list of years. Of course, it cannot be avoided that experts might just forgot flood events of the past.

Thirdly, it is important to choose the seed variables thoroughly. If possible, all seed variables should be related to the location of interest. If this is not possible, the locations should be close to the location of interest and the experts should be familiar with them. Besides, it is important that the behaviour of the variable of interest is similar at the other locations. So, the mean and maximum water levels should be comparable at the different locations.

Moreover, it would be very interesting to see whether people's estimations are close to reality or not. So, if possible, it would be great to have any possibility for validation.

Next to an improvement of the method, we would recommend further research concerning the flood risk of Hanumante river. Further research could include the following aspects:

Firstly, it would be very useful to find a way to validate the results of the SEJ or to reconstruct the historical data by another method. One possibility for both aspects would be precipitation data. It would be interesting to see if the peaks in the water level time series, resulting from the SEJ, match with peaks in precipitation data. We've had a look at using precipitation data to reconstruct the historical data, but we were not successful. This does not mean, however, that it is not possible. We still see potential in using precipitation data as validation.

Secondly, we would recommend to continue on evaluating the flood risk by Hanumante river. For example, further research could be done on the expected damages due to floods or about the possibilities of an early-warning-system.

Finally, probably the most important recommendation, it is extremely important to continue measuring the water levels and discharge of Hanumante river, if possible on a daily basis. That would make us less dependent on methods like SEJ.

VIII. Acknowledgment

We would like to thank the S4W-Nepal team for hosting us during our research period and providing us with all the data from their measurement campaigns. We had a great time and learned a lot about Nepali and Newari culture. Another acknowledgment should be made towards the Delft Deltas, Infrastructures and Mobility Initiative (DIMI), this project had not been possible without their financial support.

REFERENCES

[1] B. R. T. Rajaram Prajapati and R. Talchabhadel, "What flooded bhaktapur?" *My Republica*, 2018.
[2] C. B. of Statistics, "National population and housing census 2011. tech. rep." 2012.
[3] ICIMOD. (2010) Land cover distribution for bhaktapur.
[4] Smartphones4Water. (October, 2019) Projects s4w-nepal.
[5] R. M. Cooke, *Experts in Uncertainty, Opinion and Subjective Probability in Science*. Oxford University Press, 1991.
[6] J. Surowiecki, *The Wisdom of Crowds*. Anchor, 2004.
[7] R. Sada, "Hanumante river: Emerging uses, competition and implications," *Journal of Science and Engineering Nepal*, vol. 1, pp. 17–24, 2011.
[8] R. M. Cooke and L. L. Goossens, "Tu delft expert judgment data base," *Reliability Engineering & System Safety*, vol. 93, no. 5, pp. 657–674, 2008.
[9] G. Leontaris and O. Morales-Nápoles, "Andurila matlab toolbox for analysis and decisions with uncertainty: Learning from expert judgments," *SoftwareX*, vol. 7, pp. 313–317, 2018.
[10] M. E. Wittmann, R. M. Cooke, J. D. Rothlisberger, and D. M. Lodge, "Using structured expert judgment to assess invasive species prevention: Asian carp and the mississippi great lakes hydrologic connection - appendix b," *Environmental science & technology*, vol. 48, no. 4, pp. 2150–2156, 2014.
[11] S. N. Jonkman, R. D. J. M. Steenbergen, O. Morales Napoles, A. C. W. M. Vrouwenvelder, and J. K. Vrijling, *Probabilistic Design: Risk and Reliability Analysis in Civil Engineering*. Delft University of Technology, 2017.
[12] J. K. Vrijling and P. H. A. J. M. van Gelder, "Probabilistic design in hydraulic engineering," pp. II–8, 2002.

APPENDIX A
ANALYSIS OF EXISTING PRECIPITATION AND WATER LEVEL MEASUREMENTS

For this research, the initial objective was to find the correlation between the obtained water level measurements and corresponding precipitation data, also provided by S4W-Nepal. Starting from June 2019 daily water level measurements were conducted at the three different sites of the Hanumante river. Near the same sites, precipitation data was already being collected daily and we choose to analyse these two variables together, without regarding the effects of precipitation on upstream parts of the river and corresponding watershed. Next, we wanted to use these dependencies in order to predict past water levels by using the historical precipitation data (1971-2018) provided by DHM [7]. However, there were some major difficulties that arose using this kind of regression:

During analysis of the water levels in the Hanumante River and the nearby precipitation data, both from ODK, some notable oddities were observed when plotting both time series in one plot (see figure 8).

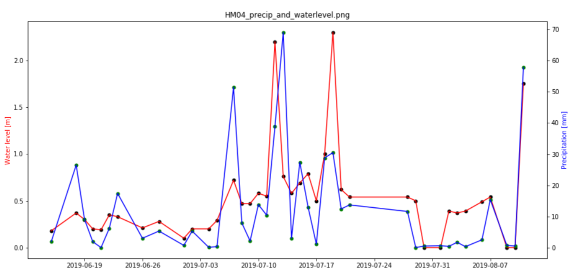

Fig. 8. Plot of Precipitation and Water level during the 2019 monsoon

Apart from the fact that some dates were missing, the main observation was that high peaks in precipitation followed one day after a high peak in water level. Judging by the nature of both phenomena this would be highly unlikely. The inverse would be expected. An explanation was found in the time at which the data was collected. Usually, the precipitation data was checked in the early morning and the water levels in the late afternoon. When, for instance the rainfall happened at lunchtime or during the afternoon, the precipitation would be recorded the next morning. However, the resulting high water would be recorded the same day of the rainfall event. To solve for these mistakes, the data could be manually altered by dragging the precipitation values to one day earlier. However, this led to the need of checking the vast collection of precipitation data and to secondary problems in the new data set (e.g. multiple measurements at one day or no measurements at all for a day). Therefore, it was chosen not to compute alterations in the data. Also, by basing thirty years of water levels on only three recent monsoon months, our research outcomes could be labeled as unreliable and would include additional research to compare the behaviour of precipitation data of both ODK and DHM and also account for non-stationary trends that should be happening over time. Still the scatter plots were computed and several regression lines were fitted over the plots. The root mean squared error was then used to judge whether or not the line was fitted successfully. It turned out that the absence of high water level recordings led to insufficient dependence in the high value regions especially. This resulted in great deviations and large errors for every regression line that was used. Eventually, we decided that a regression model without high water level data could not be used for this data set as the main interest for this research was in fact the high water levels. It was therefore concluded that the correlation between water levels and precipitation data was not the right method to generate historical water levels.

APPENDIX B
OVERVIEW OF THE QUESTIONS ANSWERED BY DIFFERENT GROUPS

In table VII the different seed variables and target questions are presented. It can also be seen which target questions are answered by which group. All seed variables are answered by all the groups. There were also 10 overlapping questions, namely the three monsoon months for the years 2018, 2015, 1990 as well as a prediction of the water level for the year 2025. The numbers in the table are presenting the specific question number. It can be seen that group 1 and 2 answered 41 questions where group 3 and 4 answered 38 questions.

TABLE VII
OVERVIEW OF THE SEED VARIABLES AND TARGET QUESTIONS ANSWERED BY DIFFERENT GROUPS

Question nr	Question	All	1	2	3	4
Seed variable 1	Bagmati July 2017	1	1	1	1	1
Seed variable 2	Godawari July 2017	2	2	2	2	2
Seed variable 3	HM06 July 2018	3	3	3	3	3
Seed variable 4	HM06 June 2019	4	4	4	4	4
Seed variable 5	HM06 July 2019	5	5	5	5	5
Seed variable 6	HM06 August 2019	6	6	6	6	6
Seed variable 7	HM04 August 2015	7	7	7	7	7
Seed variable 8	HM04 August 2019	8	8	8	8	8
Seed variable 9	HM04 July 2019	9	9	9	9	9
Seed variable 10	HM04 June 2019	10	10	10	10	10
Target question 1	August 2018	11	11	11	11	11
Target question 2	July 2018	12	12	12	12	12
Target question 3	June 2018	13	13	13	13	13
Target question 4	August 2017		14			
Target question 5	July 2017		15			
Target question 6	June 2017		16			
Target question 7	August 2016			14		
Target question 8	July 2016			15		
Target question 9	June 2016			16		
Target question 10	August 2015				14	
Target question 11	July 2015				15	
Target question 12	June 2015				16	
Target question 13	August 2014					14
Target question 14	July 2014					15
Target question 15	June 2014					16
Target question 16	August 2013		17			
Target question 17	July 2013		18			
Target question 18	June 2013		19			
Target question 19	August 2012			17		
Target question 20	July 2012			18		
Target question 21	June 2012			19		
Target question 22	August 2011				17	
Target question 23	July 2011				18	
Target question 24	June 2011				19	
Target question 25	August 2010					17
Target question 26	July 2010					18
Target question 27	June 2010					19
Target question 28	August 2009		20			
Target question 29	July 2009		21			
Target question 30	June 2009		22			
Target question 31	August 2008			20		
Target question 32	July 2008			21		
Target question 33	June 2008			22		
Target question 34	August 2007				20	
Target question 35	July 2007				21	
Target question 36	June 2007				22	
Target question 37	August 2006					20
Target question 38	July 2006					21
Target question 39	June 2006					22
Target question 40	August 2005	14	23	23	23	23
Target question 41	July 2005	15	24	24	24	24
Target question 42	June 2005	16	25	25	25	25
Target question 43	August 2004		26			
Target question 44	July 2004		27			
Target question 45	June 2004		28			
Target question 46	August 2003			26		
Target question 47	July 2003			27		
Target question 48	June 2003			28		
Target question 49	August 2002				26	
Target question 50	July 2002				27	
Target question 51	June 2002				28	
Target question 52	August 2001					26
Target question 53	July 2001					27
Target question 54	June 2001					28
Target question 55	August 2000		29			
Target question 56	July 2000		30			
Target question 57	June 2000		31			
Target question 58	August 1999			29		
Target question 59	July 1999			30		
Target question 60	June 1999			31		
Target question 61	August 1998				29	
Target question 62	July 1998				30	
Target question 63	June 1998				31	
Target question 64	August 1997					29
Target question 65	July 1997					30
Target question 66	June 1997					31
Target question 67	August 1996		32			
Target question 68	July 1996		33			
Target question 69	June 1996		34			
Target question 70	August 1995			32		
Target question 71	July 1995			33		
Target question 72	June 1995			34		
Target question 73	August 1994				32	
Target question 74	July 1994				33	
Target question 75	June 1994				34	
Target question 76	August 1993					32
Target question 77	July 1993					33
Target question 78	June 1993					34
Target question 79	August 1992		35			
Target question 80	July 1992		36			
Target question 81	June 1992		37			
Target question 82	August 1991			35		
Target question 83	July 1991			36		
Target question 84	June 1991			37		
Target question 85	August 1990	17	38	38	35	35
Target question 86	July 1990	18	39	39	36	36
Target question 87	June 1990	19	40	40	37	37
Target question 88	Year 2025	20	41	41	38	38

APPENDIX C
CALIBRATION AND INFORMATION SCORES PER EXPERT

Per group we presented a table containing all the experts within this group and their corresponding calibration and information scores as well as their normalized weights. The calibration score is a value for the likelihood that the realizations (the real values for the seed variables) correspond statistically with the expert's assessments [10]. The information score is a measure for the spreading of the distribution of an expert's response with respect to a background measure. Information can only be measured with respect to the background measure. The column with the normalized weights represents the weight an expert received when the optimized Decision Maker was computed.

TABLE VIII
CALIBRATION AND INFORMATION SCORES FOR THE EXPERTS OF GROUP 1

Expert	Function	Calibr. score	Inf. score	Norm. weight
Expert 1.1	Citizen of Bhaktapur	$6.085 * 10^{-2}$	2.118	0
Expert 1.2	Student	$1.371 * 10^{-8}$	2.300	0
Expert 1.3	Citizen of Bhaktapur	0.114	0.664	0
Expert 1.4	Citizen of Bhaktapur	0.314	0.789	0
Expert 1.5	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.633	0
Expert 1.6	Citizen of Bhaktapur	$1.497 * 10^{-11}$	1.517	0
Expert 1.7	Citizen of Bhaktapur	$6.131 * 10^{-13}$	1.398	0
Expert 1.8	Citizen of Bhaktapur	$6.174 * 10^{-9}$	1.154	0
Expert 1.9	Citizen of Bhaktapur	$2.638 * 10^{-4}$	0.774	0
Expert 1.10	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.277	0
Expert 1.11	Citizen of Bhaktapur	0.395	0.829	1
Expert 1.12	Citizen of Bhaktapur	$1.150 * 10^{-3}$	0.694	0
Expert 1.13	Citizen of Bhaktapur	$3.098 * 10^{-3}$	0.438	0
Expert 1.14	Citizen of Bhaktapur	$6.131 * 10^{-13}$	2.032	0
Expert 1.15	Student	0.228	0.356	0
Expert 1.16	Water specialist	$6.289 * 10^{-3}$	1.301	0

TABLE IX
CALIBRATION AND INFORMATION SCORES FOR THE EXPERTS OF GROUP 2

Expert	Function	Calibr. score	Inf. score	Norm. weight
Expert 2.1	Citizen of Bhaktapur	$7.994 * 10^{-4}$	1.466	0.012
Expert 2.2	Citizen of Bhaktapur	$6.085 * 10^{-2}$	0.915	0.563
Expert 2.3	Citizen of Bhaktapur	$4.488 * 10^{-7}$	0.925	0
Expert 2.4	Citizen of Bhaktapur	$7.284 * 10^{-4}$	0.799	0
Expert 2.5	Citizen of Bhaktapur	$6.174 * 10^{-9}$	0.708	0
Expert 2.6	Citizen of Bhaktapur	$6.174 * 10^{-9}$	1.190	0
Expert 2.7	Citizen of Bhaktapur	$4.704 * 10^{-2}$	0.602	0.286
Expert 2.8	Citizen of Bhaktapur	$4.937 * 10^{-7}$	1.537	0
Expert 2.9	Citizen of Bhaktapur	$1.543 * 10^{-7}$	0.769	0
Expert 2.10	Water specialist	$2.389 * 10^{-8}$	1.294	0
Expert 2.11	Citizen of Bhaktapur	$5.992 * 10^{-3}$	1.894	0.115
Expert 2.12	Citizen of Bhaktapur	$2.500 * 10^{-6}$	1.107	0
Expert 2.13	Citizen of Bhaktapur	$1.543 * 10^{-7}$	0.951	0
Expert 2.14	Citizen of Bhaktapur	$1.543 * 10^{-7}$	0.892	0
Expert 2.15	Student	$2.083 * 10^{-5}$	1.429	0
Expert 2.16	Water specialist	$1.311 * 10^{-3}$	1.855	0.025

TABLE X
CALIBRATION AND INFORMATION SCORES FOR THE EXPERTS OF GROUP 3

Expert	Function	Calibr. score	Inf. score	Norm. weight
Expert 3.1	Citizen of Bhaktapur	$9.855 * 10^{-7}$	1.109	0
Expert 3.2	Citizen of Bhaktapur	$6.131 * 10^{-13}$	1.859	0
Expert 3.3	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.195	0
Expert 3.4	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.336	0
Expert 3.5	Citizen of Bhaktapur	$3.574 * 10^{-2}$	0.650	0.573
Expert 3.6	Citizen of Bhaktapur	$2.042 * 10^{-3}$	0.434	0
Expert 3.7	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.643	0
Expert 3.8	Citizen of Bhaktapur	$8.214 * 10^{-3}$	0.845	0.171
Expert 3.9	Citizen of Bhaktapur	$1.579 * 10^{-5}$	0.995	0
Expert 3.10	Citizen of Bhaktapur	$1.543 * 10^{-7}$	0.904	0
Expert 3.11	Water specialist	$1.543 * 10^{-7}$	0.746	0
Expert 3.12	Citizen of Bhaktapur	$7.284 * 10^{-4}$	0.625	0
Expert 3.13	Water specialist	$6.131 * 10^{-13}$	2.423	0
Expert 3.14	Student	$3.500 * 10^{-8}$	2.038	0
Expert 3.15	Water specialist	$1.066 * 10^{-6}$	2.284	0
Expert 3.16	Water specialist	$6.289 * 10^{-3}$	1.645	0.255

TABLE XI
CALIBRATION AND INFORMATION SCORES FOR THE EXPERTS OF GROUP 4

Expert	Function	Calibr. score	Inf. score	Norm. weight
Expert 4.1	Citizen of Bhaktapur	$5.992 * 10^{-3}$	0.722	0
Expert 4.2	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.192	0
Expert 4.3	Citizen of Bhaktapur	0.493	0.485	1
Expert 4.4	Citizen of Bhaktapur	$5.544 * 10^{-2}$	0.228	0
Expert 4.5	Citizen of Bhaktapur	$4.488 * 10^{-7}$	1.172	0
Expert 4.6	Citizen of Bhaktapur	$1.293 * 10^{-10}$	1.684	0
Expert 4.7	Citizen of Bhaktapur	$1.628 * 10^{-4}$	0.811	0
Expert 4.8	Citizen of Bhaktapur	$3.321 * 10^{-2}$	0.483	0
Expert 4.9	Citizen of Bhaktapur	$1.893 * 10^{-6}$	2.119	0
Expert 4.10	Water specialist	$1.293 * 10^{-10}$	1.908	0
Expert 4.11	Student	$1.293 * 10^{-10}$	1.706	0
Expert 4.12	Citizen of Bhaktapur	$1.579 * 10^{-5}$	1.916	0
Expert 4.13	Water specialist	$4.937 * 10^{-7}$	2.733	0
Expert 4.14	Water specialist	$2.638 * 10^{-4}$	0.567	0

Appendix D
Best GEV for the 5% and 95% quantiles of the water levels

The results for the characterising parameters of the best GEV based on the 5% and 95% quantiles of the water levels are presented in figure 9 and figure 10 respectively. For the 5% quantile value of the water levels, the numerical value of the shape parameter, k, is approximately zero (-0.0052), which means that the corresponding GEV is a Gumbel extreme value distribution. For the 95% quantile value of the water levels, the numerical value of the shape parameter, k, is negative (-0.58), which means that the corresponding GEV is Weibyll extreme value distribution.

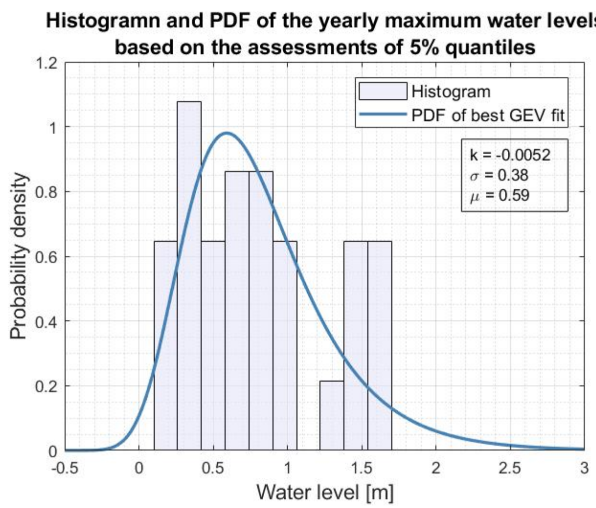

Fig. 9. The probability density function according to the GEV based on the 5% quantiles of the water levels

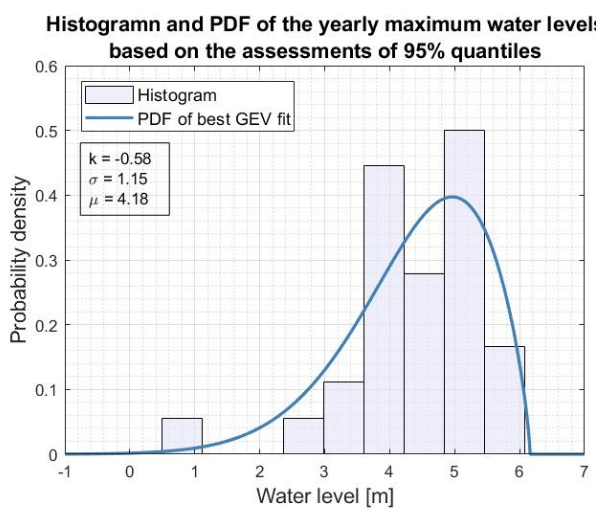

Fig. 10. The probability density function according to the GEV based on the 95% quantiles of the water levels