Characterization of Precipitation through Copulas and Expert Judgement for Risk Assessment of Infrastructure

Oswaldo Morales-Nápoles[1]; Dominik Paprotny[2]; Daniël Worm[3]; Linda Abspoel-Bukman[4]; and Wim Courage[5]

Abstract: In this paper two methodologies are investigated that contribute to better assessment of risks related to extreme rainfall events. Firstly, one-parameter bivariate copulas are used to analyze rain gauge data in the Netherlands. Out of three models considered, the Gumbel copula, which indicates upper tail dependence, represents the data most accurately for all 33 stations in the Netherlands. Seasonal variability is noticeable, with rank correlation reaching maximum in winter and minimum in summer as well as other temporal and spatial patterns. Secondly, an expert judgment elicitation was undertaken. The experts' opinions were combined using Cooke's classical method in order to obtain estimates of future changes in precipitation patterns. Experts predicted mostly an approximate 10% increase in rain amount, duration, intensity and the dependence between amount and duration. The results were in line with official national climate change scenarios, based on numerical modelling. Applicability of both methods was presented based on an example of an existing tunnel in the Netherlands, contributing to better estimates of the tunnel's limit state function and therefore the probability of failure. **DOI: 10.1061/AJRUA6.0000914.**

Introduction

Extreme precipitation is a major source of threat to society and infrastructure. For example, it causes disastrous flash floods, which in Europe were responsible for up to 1,000 casualties during a single event (Barredo 2007). Excessive rainfall is also problematic in any areas covered by artificial surfaces, where water does not infiltrate the soil, but instead is removed by the drainage system. If the capacity of those systems is insufficient, water accumulates, resulting in not only direct damage, but also cessation of services. An example is the inundation of roads or tunnels, which apart from damaging their surfaces brings the traffic to a halt. This is of particularly high concern in the Netherlands, which is flat and relies on an extensive network of channels for water management. Moreover, it has high population density resulting in increased fraction of land covered by artificial surfaces—12% compared to an European Union average of 5%—and high volume of traffic on more than 2,600 km of freeways and numerous other roads (Eurostat 2016). Therefore, precipitation is one of the most relevant variables investigated by the Royal Netherlands Meteorological Institute (KNMI) and other Dutch research organizations in the context of future climate scenarios for the Netherlands.

The analysis of precipitation in context of risk to infrastructure involves not only deriving the probable amount of rainfall, but also its duration and intensity. High intensities exceeding drainage capacity are as problematic as a long, but not intense shower generating higher volume of water than a structure's storage capacity. A joint distribution of the amount and duration of rainfall therefore calculates the probability of an event higher than the structure's resilience. Conversely, it can be used to define minimum design standards in a given location for certain types of infrastructure.

A typical mathematical solution to this problem is to use depth-duration-frequency (DDF) or intensity-duration-frequency curves (IDF). These describe either rainfall depth (i.e., total amount that has rained) or intensity as a function of duration for given return periods or probabilities of exceedance. The curves are derived by fitting a parametric probability distribution function is to precipitation data for fixed durations, e.g., 1, 2, 4, 8, 12, and 24 h. Then through a regression analysis a linear relation between the parameters of the underlying distribution function and the duration (or a transformation of it) is found. This approach was used as early as the 1930s (Bernard 1932) and a large number of studies are available covering a vast selection of countries (e.g., Alam and Elshorbagy 2015; Ben-Zvi 2009; Haddad and Rahman 2014; Kotowski and Kaźmierczak 2013; Koutsoyiannis and Baloutsos 2000; Modesto Gonzalez Pereira et al. 2015; Overeem et al. 2008). This standard approach relies on a linear regression based on a limited number of observations, though improvements using Bayesian statistics were also proposed (Van de Vyver 2015).

Another issue related to extreme rainfall is predicting the change in their frequency due to climate change. This can be done with numerical modeling of climate similarly as for other meteorological variables. However, the models are still not as reliable for precipitation as they are for temperature (e.g., Jacob et al. 2007), especially the extreme sort which is of interest for the authors (Lenderink 2010). Furthermore, there is large variation in output between models and emission scenarios, leaving substantial uncertainty about

[1]Assistant Professor of Probabilistic Design, Dept. of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft Univ. of Technology, Stevinweg 1, 2628 CN, Delft, Netherlands (corresponding author). ORCID: https://orcid.org/0000-0002-6764-4674. E-mail: o.moralesnapoles@tudelft.nl

[2]Ph.D. Student, Dept. of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft Univ. of Technology, Stevinweg 1, 2628 CN, Delft, Netherlands. E-mail: d.paprotny@tudelft.nl

[3]Researcher, TNO, Stieltjesweg 1, 2628 CK, Delft, Netherlands. E-mail: daniel.worm@tno.nl

[4]Researcher, TNO, Stieltjesweg 1, 2628 CK, Delft, Netherlands. E-mail: linda.abspoel@tno.nl

[5]Researcher, TNO, Stieltjesweg 1, 2628 CK, Delft, Netherlands. E-mail: wim.courage@tno.nl

Note. This manuscript was submitted on July 28, 2016; approved on March 6, 2017; published online on June 8, 2017. Discussion period open until November 8, 2017; separate discussions must be submitted for individual papers. This paper is part of the ***ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering***, © ASCE, ISSN 2376-7642.

how the distribution of heavy precipitation will change in the future (Rojas et al. 2011; Jacob et al. 2014).

In this paper it is proposed to improve the calculation of extreme rainfall probability of occurrence by two methods, namely copulas and structured expert judgment. First, this paper proposes an alternative method to IDF curves by describing the joint distribution of rain amount and rain duration per shower through bivariate copulas. Second, a method to assess and combine expert opinions regarding dependence is presented. The method is generic and used in this paper to give an idea how experts perceive future trends in extreme rainfall. The study is an extension of the work presented in Morales Nápoles et al. (2015).

Both approaches are shown on examples from the Netherlands, with rainfall data used for analyzing the joint distributions (sections “Materials and Methods” and “Dependence of Rain Amount and Duration”), and an expect judgment session with local experts assessing the influence of climate change on rainfall patterns in the Netherlands (section “Expert Judgment on Precipitation”). Finally, the paper presents how this information can be used in practice in infrastructure management (section “Impact of Rain on Infrastructure”). The case presented in that section refers to a hypothetical tunnel. As stated before, the goal is to present a generic methodology available to researchers for the characterization of precipitation and its use in risk assessment of infrastructure.

Materials and Methods

Precipitation Data and Climate Scenarios

The data of interest comes from publicly available measurements of precipitation taken at weather stations operated by the KNMI (2016). A total of 33 stations were used, which are listed in Table 1 and presented on the map in Fig. 1. For all stations the latest data are from 2013, but the length of the series varies from 12 (Wijk aan Zee) to 63 years (De Bilt), with an average of 30 years of data. For more details about the Dutch rain gauge network and methods for measuring rainfall the reader is referred to Overeem et al. (2008).

For each station, rain data show the fraction of the hour during which rainfall was recorded and the amount of rain. The fractions are in 0.1 increments, ranging from 0.1 to 1, while the amount is in given in mm. The information in which it has been raining for specific 6-min interval was not available. Therefore, certain assumptions had to be made in order to distinguish individual showers from the series. The procedure to aggregate hourly data into showers is presented in Fig. 2. All consecutive periods where the fraction was equal to 1 (i.e., it rained the full hour) were considered one shower. An hour with a fraction of 0.1 to 0.9 was joined to an adjacent hour with a fraction equal to 1. However, if during two subsequent hours it rained only partially, they were considered separate showers. For each shower, three quantities were derived: duration in hours (X_D), amount in mm (X_A), and intensity in mm/hour (X_I), which is the quotient of X_A and X_D. This definition of a shower likely overestimates the actual number of separate rain events, as in a warm or occluded front where rainfall is often patchy, raining only during a fraction of an hour, but for many hours. In the case of those low-intensity events, rainfall could also be erroneously not recorded. In this study only rain gauge measurements were used instead of modeled, radar, or satellite data, which could help in identifying rain events (showers) better in terms of a meteorological system. However, those drawbacks should not affect the high-intensity events which are of interest in risk management.

Table 1. List of KNMI Weather Stations Used in This Study

Number	Name	Start date
210	Valkenburg	1-1-1973
235	De Kooy	1-1-1957
240	Schiphol	1-1-1974
249	Berkhout	1-1-2000
251	Hoorn (Terschelling)	1-1-1995
257	Wijk aan Zee	1-1-2002
260	De Bilt	1-1-1951
265	Soesterberg	1-1-1975
267	Stavoren	1-1-1994
269	Lelystad	1-1-1994
270	Leeuwarden	1-1-1975
273	Marknesse	1-1-1994
275	Deelen	1-1-1983
277	Lauwersoog	1-1-1994
278	Heino	1-1-1994
279	Hoogeveen	1-1-1994
280	Eelde	1-1-1957
283	Hupsel	1-1-1994
286	Nieuw Beerta	1-1-1994
290	Twenthe	1-1-1975
310	Vlissingen	1-1-1957
319	Westdorpe	1-1-1994
323	Wilhelminadorp	1-1-1994
330	Hoek van Holland	1-1-1996
344	Rotterdam	1-1-1974
348	Cabauw	1-1-1987
350	Gilze-Rijen	1-1-1977
356	Herwijnen	1-1-1994
370	Eindhoven	1-1-1985
375	Volkel	1-1-1975
377	Ell	1-1-2000
380	Maastricht	1-1-1958
391	Arcen	1-1-1994

Note: All stations have an end date of 12-31-2013.

In the experts’ elicitation (see section “Structured Expert Judgment”) KNMI’s climate projections for the 21st century, known as KNMI’14, were used additionally to rain gauge data. They include a large number of variables and their expected changes between the reference period 1981 to 2010 and 30-year periods centered around 2030, 2050, and 2085. The projections were made by KNMI in four scenarios, derived from the results of 250 global climate simulation runs with EC-Earth model. Detailed climate modeling for the Netherlands was then done by KNMI with *RACMO2*, their in-house regional climate model (KNMI 2014). Four scenarios were considered: GL, GH, WL, and WH. The G scenarios correspond to an increase in global temperature by 1°C, while the W scenarios refer to an increase of 2°C. The L stands for small changes in air circulation patterns, while H indicates large changes for the same variable. The resulting projections were compared with the experts’ opinions on trends in average and extreme rainfall.

Bivariate Copulas

As expressed in the section “Introduction,” it is intended to model the dependency between rainfall amount and duration using copulas. A copula can be loosely defined as the joint distribution on the unit hypercube with uniform (0, 1) margins. For the most comprehensive description of copulas the reader is referred to Joe (2015). The bivariate copula of two continuous random variables X_i and X_j, for $i \neq j$, with joint distribution F_{X_i,X_j} is

Fig. 1. Location of KNMI weather stations used in this study

KNMI rainfall data

Hour	Fraction of the hour with rainfall	Amount of rainfall per hour [mm]
1	0	0
2	1	1
3	1	2
4	0.8	0.5
5	0	0
6	0.5	1
7	0.2	1

Aggregated shower data

Shower no.	Duration [h]	Amount [mm]	Intensity [mm/h]
1	2.8	3.5	1.25
2	0.5	1	2
3	0.2	1	5

Fig. 2. Aggregation of rainfall data into showers

$$F_{X_i,X_j}(X_i,X_j) = C_\theta[F_{X_i}(X_i),F_{X_j}(X_j)] \tag{1}$$

where the copula function C is indexed by a scalar or vector of parameters θ. Spearman's rank correlation coefficient $r(X_i,X_j)$, a familiar measures of dependence, may be expressed in terms of θ, provided that it is scalar as

$$r(X_i,X_j) = 12\int_{[0,1]^2} uv dC_\theta(u,v) - 3 \tag{2}$$

where u and v are marginal uniform variates on the interval [0, 1]. The rank correlation is the usual Pearson's product moment correlation ρ computed with the ranks of X_i and X_j, Pearson's product moment correlation coefficient is

$$\rho(X_i,X_j) = \frac{E(X_i,X_j) - E(X_i)E(X_j)}{\sqrt{\mathrm{var}(X_i)\mathrm{var}(X_j)}} \tag{3}$$

One parameter bivariate copulas have the convenient property of being parameterized by a single correlation value. However, different asymmetries in the joint distribution may be present in different types of copulas. In this paper, three of the most frequently used copula families are considered: Gaussian, Gumbel, and Clayton. First of those, the Gaussian copula, has the following cumulative distribution function:

$$C_\rho(u,v) = \Phi_\rho[\Phi^{-1}(u),\Phi^{-1}(v)], \quad (u,v)\in[0,1]^2 \tag{4}$$

where Φ = bivariate Gaussian cumulative distribution; and ρ = product moment correlation coefficient of the normal variates. Second, the Gumbel copula, which is parameterized by δ, is defined as

$$C_\delta(u,v) = \exp(-\{[-\log(u)]^\delta + [-\log(v)]^\delta\}^{1/\delta}), \quad \delta \geq 1 \tag{5}$$

The Clayton copula is parameterized by α

$$C_\alpha(u,v) = (u^{-\alpha} + v^{-\alpha} - 1)^{-\alpha}, \quad \alpha \in [-1,\infty) \tag{6}$$

The different copulas are used in extreme value to investigate certain dependence patterns related to the quantiles of the variables of interest. One such pattern is known as tail dependence. The upper tail dependence coefficient λ_U for two random variables X_i and X_j is defined as

$$\lambda_U = \lim_{u\to 1} P[X_i > F_{X_i}^{-1}(u)|X_j > F_{X_j}^{-1}(u)] = \lim_{u\to 1} P(U > u|V > u) \tag{7}$$

Roughly, a value of $\lambda_U > 0$ in Eq. (7) indicates that it is likely (more than normal) to observe high values of X_D (rain duration) together with high values of X_A (rain amount). Lower tail dependence would be defined similarly, but for low values of the marginals. In the Gaussian copula, there is no tail dependence $\lambda_U = 0$, while in Clayton lower tail dependence $\lambda_U = 2^{-1/\alpha}$ is represented, and the Gumbel copula presents upper tail dependence $\lambda_U = 2 - 2^{1/\delta}$. These three types of copulas already cover a range of dependence structures that are typically observed in the data.

The existence of a tail dependence can be corroborated using two methods. The statistic used in one of the blanket tests described by Genest et al. (2009) will be computed as a goodness-of-fit measure. It is based on the Cramèr–von Mises M statistic. It can be computed for a sample of length n using the following formula:

$$M_n(\boldsymbol{u},) = \sum_{|\boldsymbol{u}|}\{C_{\hat{\theta}_n}(\boldsymbol{u}) - B(\boldsymbol{u})\}^2, \quad \boldsymbol{u}\in[0,1]^2 \tag{8}$$

where $B(\boldsymbol{u}) = \sum 1(U_i \leq \boldsymbol{u})$ = empirical copula; and $C_{\hat{\theta}_n}(\boldsymbol{u})$ = parametric copula with parameter $\hat{\theta}_n$ estimated from the sample. This statistic is the sum of squared differences between the empirical copula and the parametric estimate.

As another diagnostics tool, semicorrelations are analyzed, which is an approach suggested by Joe (2015). The semicorrelations are the Pearson's product moment correlation coefficients computed in the upper and lower quadrants of the normal transforms of the original variables. If the correlations are positive, semicorrelations in the upper right (NE) and lower left (SW) quadrants are computed using the following formulas:

$$\rho_{ne} = \rho(Z_i,Z_j|Z_i > 0, Z_j > 0) \tag{9}$$

$$\rho_{sw} = \rho(Z_i,Z_j|Z_i \leq 0, Z_j < 0) \tag{10}$$

where (Z_i,Z_j) = standard normal transforms of (X_i,X_j). Semicorrelations in the upper left (NW) and lower right (SE) quadrants are denoted ρ_{nw} and ρ_{se} and defined similarly to Eqs. (9) and (10) if the correlation is negative. In general, larger absolute values of the semicorrelations in a particular quadrant compared to the correlation for the entire sample and the opposite quadrant indicate tail dependence.

Application of copulas to estimation of rainfall patterns is still fairly new (see e.g., De Michele and Salvadori 2003). A large part of the applications focused on correcting climate model data (Laux et al. 2011), filling gaps in data series (Bárdossy and Pegram 2014), and analyzing dependencies between stations (Schölzel and Friederichs 2008), with some studies exploring the relationship of rainfall amount and duration (Serinaldi 2009; Balistrocchi and Bacchi 2011; Cantet and Arnaud 2014). However, in this study, as pointed out earlier, an approximation to the bivariate copula of amount and duration is provided and corroborated with data for the Netherlands. It is also shown how expert opinions regarding future trends for precipitation may be obtained in a structured way.

Structured Expert Judgment

Structured expert judgment is a method of quantifying and generalizing opinions of experts. In this paper, the so-called classical, or Cooke's, method (Cooke 1991) is applied, which aims to derive rational consensus from experts' judgments. Roughly, experts are asked to provide their assessment over a continuous quantity. Importantly, they do not give a single best estimate, but rather their uncertainty distribution over certain quantities. The experts make their estimates in certain percentiles, most commonly the 5th, 50th, and 95th percentile. The expert giving the 5th percentile expresses that they would be very surprised if the actual value of the variable in question was smaller than their 5th percentile estimate. Conversely, the 95th percentile estimate will be the value for which the expert would be surprised if the variable has exceeded it; the 50th percentile is thus the expert's best estimate.

The experts are asked two types of questions that they answer in the way described previously (Appendix S1 and Table S1). The first group are the seed, or calibration, variables. These are quantities whose value is known, or will be known within the time frame of the research, to the analysts but not to the experts at the moment of the elicitation. The variables are used to assess the reliability of each expert opinion. In Cooke's model, two measures of performance are computed: the calibration and information scores (Appendix S2). Roughly, calibration measures the degree to which experts are statistically accurate, while information measures the degree to which experts' uncertainty estimates are concentrated relative to a background measure. Good expertise in the classical method refers to highly calibrated (typically calibration scores >0.05) and highly informative experts.

The combination of experts' assessments is known as the decision maker (DM). This is a weighted average of individual estimates. The experts could be weighted equally (EWDM) or the weights can be determined based on the performance of experts in the seed variables, as measured by information and calibration. The weights are then used to calculate an uncertainty distribution for the second group of questions, known as the variables of interest. These are unknown quantities that the analyst wants to derive based on experts' responses. Structured expert judgment was used in a variety of fields, most recently also in context of climate change, particularly sea level rise (Cooke 2013; Bamber and Aspinall 2013; Oppenheimer et al. 2016).

The elicitation of dependence estimates from experts is a subject of active research. It has been observed in recent studies that though Cooke's method is an appropriate estimate to get empirical validation of experts estimates over uncertainty distributions it fails to provide empirical evidence regarding expert's ability to quantify dependence (Morales-Nápoles et al. 2013). For this reason the *dependence-calibration* score introduced in Morales Nápoles and Worm (2013) was used (also discussed in Appendix S2). The dependence calibration score is a measure of distance between a certain dependence structure used for calibration purposes and estimates of this dependence structure provided by experts. A value close to 1 would indicate expert's ability to quantify dependence.

The elicitation that is the topic of this paper was carried out on October 16 and December 8, 2014 at Netherlands Organization for Applied Scientific Research (TNO) in Delft, the Netherlands. A total of eight experts participated, with an expertise ranging from a Ph.D. student to full professors in the fields of hydrology, climatology, and meteorology. They represented Wageningen University, Delft University of Technology, the national weather service KNMI, a private weather forecast company MeteoGroup, and HKV Lijn in Water, a consultancy firm based in the Netherlands. In this study a total of 15 calibration variables and 19 variables of interest were elicited. Some questions referred to historical climate (1975–2013) and were used for calibration, while the others inquired about experts' opinions on future climate (2015–2053). Most questions concerned rainfall occurring at De Bilt and Rotterdam weather stations, with a few referring to a larger group of stations. An example of a seed question is:

- For each of the following two stations: Rotterdam and De Bilt. Consider the shower data as described in [section "Precipitation Data and Climate Scenarios" of this paper] and think of the shower with the maximum value of rain amount in the period of interest (January 1, 1975–December 31, 2013). What would this maximum value *of rain amount* be (in mm)?

The experts gave their estimates for the two stations providing the 5th, 50th, and 95th percentile of their uncertainty distribution. The exact same question was asked for the 2015–2053 time period as a variable of interest. After the primary methods and materials to be used in this study have been discussed, the principal findings of the study are presented. Additional details on expert judgment methodology and a summary of all questions used in the elicitation may be found in Appendix S1.

Dependence of Rain Amount and Duration

Results for De Bilt Station

De Bilt weather station is located at the KNMI's headquarters and has the longest data series of the entire network. It was chosen for this study to present the copula approach to analyzing showers; spatial and temporal variation in the Netherlands is shown in the next section. In Fig. 3 the transformation from observations [Fig. 3(a)] to pseudo-observations [Fig. 3(b)] is shown. The pseudo-observations are the samples transformed to the interval (0, 1) through the empirical margins.

For this station, the rank correlation coefficient for the whole data series is 0.66. Ties, i.e., samples realizing the same pair of values, are visible in the lower tail of the distribution—that is, for small values of both rain amount and duration. It also appears that more samples are concentrated in the upper right tail of the joint distribution than elsewhere. That feature suggests upper tail dependence. However, the upper tail dependence does not necessarily lead to the most intense showers. Even though the upper right corner of

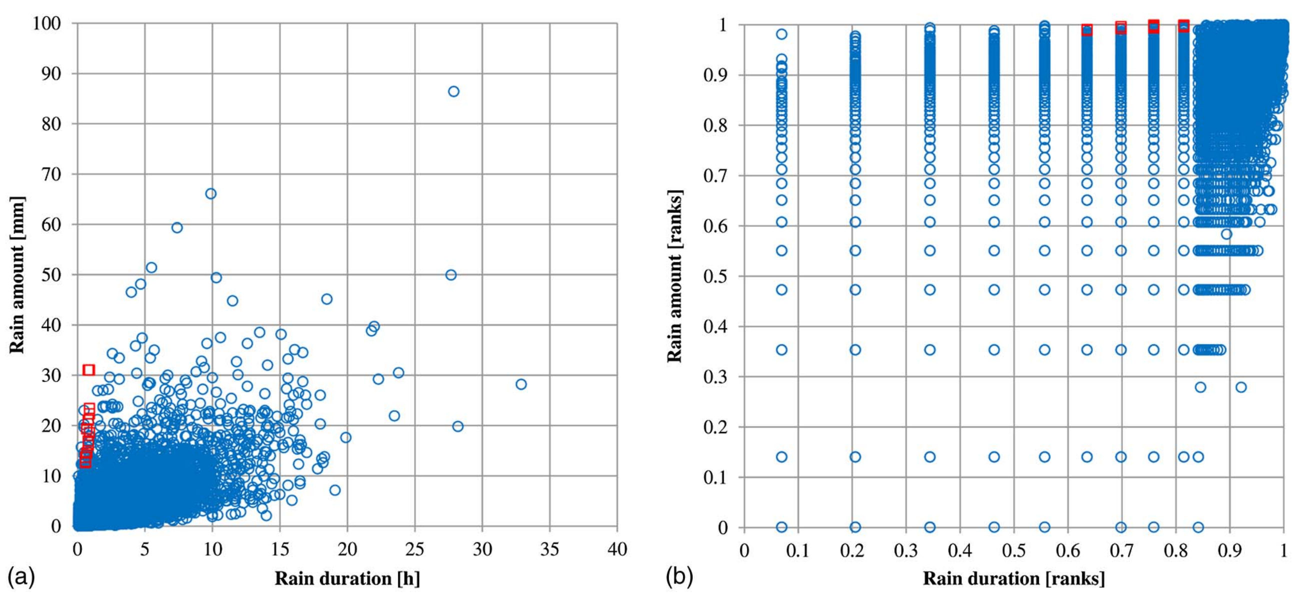

Fig. 3. Rain amount (X_A) and duration (X_D): (a) original observations; (b) pseudo-observations; showers lasting more than 0.5 h with intensity higher than 20 mm/h are marked with a square

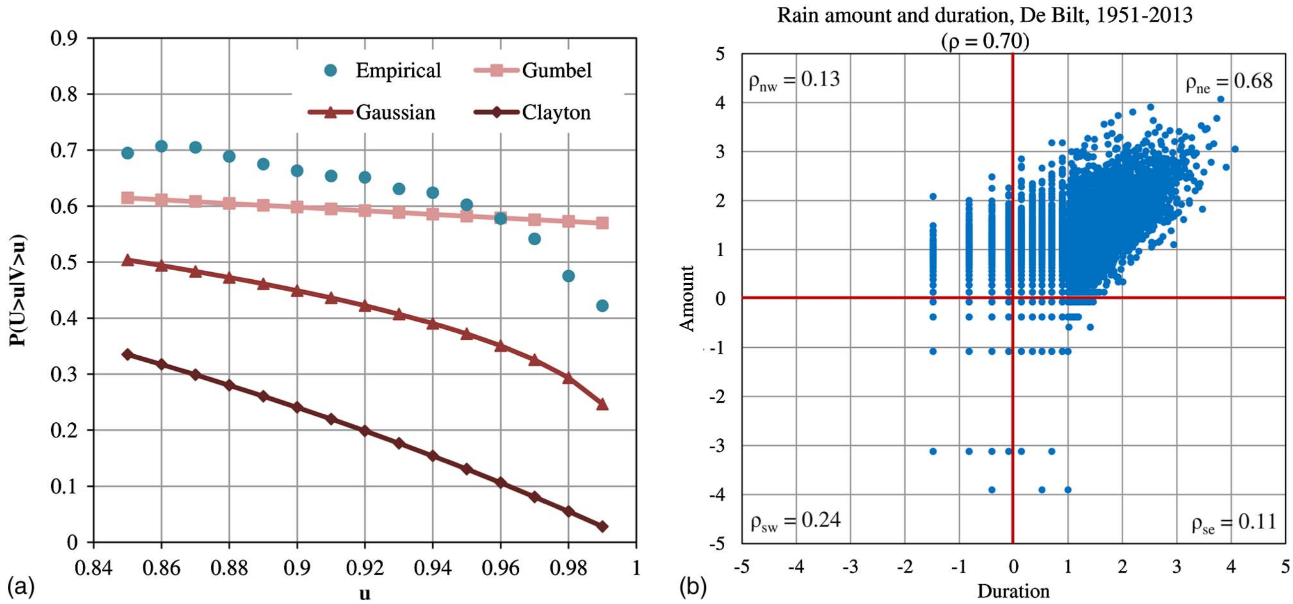

Fig. 4. (a) Rain amount and duration transformed to standard normal, with semicorrelation for each quadrant; (b) empirical and parametric estimates of $P(U > u|V > u)$ for different probabilities u at De Bilt

the distribution contains the highest amounts of rain per shower, it also contains the longest shower durations. There are 13 showers with durations larger than one-half hour, which is approximately the 60th percentile of the distribution of X_D, with an intensity larger than 20 mm per hour (marked with a square in Fig. 3). These correspond to amounts of rain between 12.6 and 31 mm.

The shower data from De Bilt were fitted to three types of copulas and evaluated. The value of the test statistic M for Clayton, Gumbel, and Gaussian copulas was 5.46, 4.33, and 4.42, respectively; the p-values amounted to 0.05, 0.097, and 0.082. Note that the Clayton copula is not a preferable model for this particular station. However, the difference between Gaussian and Gumbel copulas is small, therefore both could potentially be a fair model for De Bilt's shower data. More insights could be obtained by analyzing semicorrelations. In case of De Bilt, the overall Pearson's product moment correlation for data transformed to standard normal is approximately 0.70, while the upper right quadrant amounts to 0.68, and much less (0.11–0.24) in the other quadrants (Fig. 4). Thus, the semicorrelations indicate a preference for a model with upper tail dependence, such as the Gumbel copula.

The findings are further supported by comparing empirical and parametric estimates of conditional probabilities in the upper joint tail, as shown in Fig. 4(a). In this graph, empirical and parametric estimates of $P(U > u|V > u)$, i.e., the probability that rain amount will be larger than its uth percentile, given that the rain duration is observed above its uth percentile. It is shown that the empirical probability is closest to the parametric estimate from the Gumbel copula. For the 96th percentile ($u = 0.96$), the parametric (Gumbel) and empirical probabilities are the same. The conclusion is that out of the one-parameter copulas investigated in this study the Gumbel copula provides the best fit for data from De Bilt over the period 1951–2013.

Temporal and Spatial Variability in Correlation

Analysis of shower time series indicated noticeable temporal variability. In Fig. 5, the annual variations for De Bilt station are

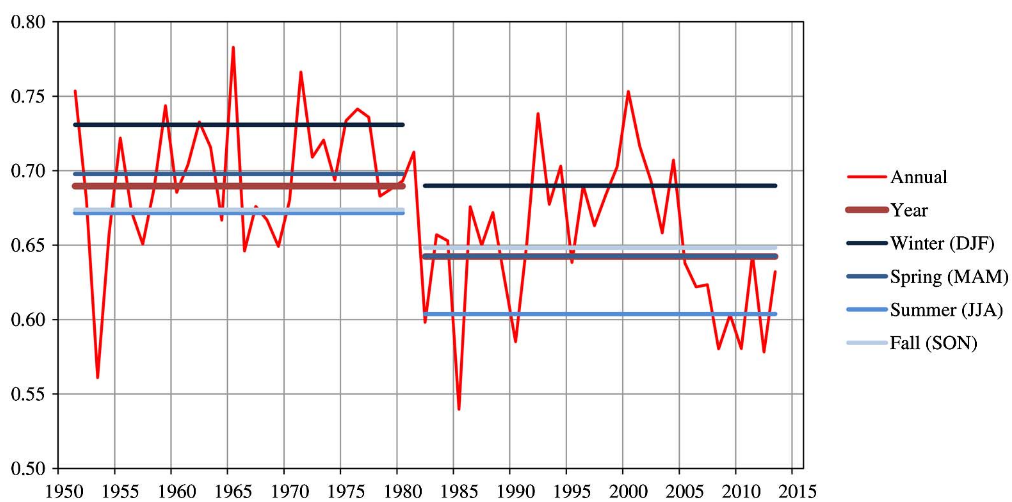

Fig. 5. Rank correlation between rain amount and duration for De Bilt by year and season

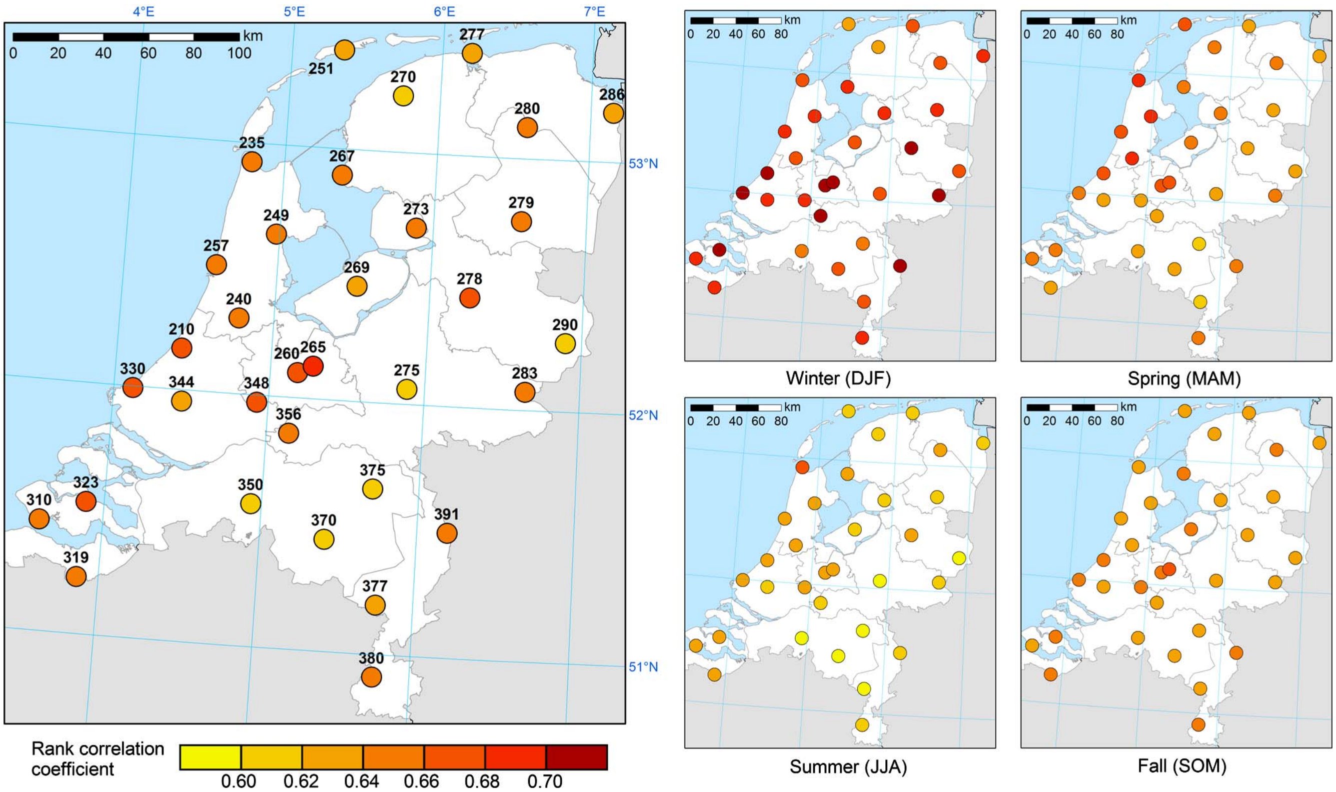

Fig. 6. Spatial and seasonal variability of rank correlation between rain amount and duration in the Netherlands

presented alongside seasonal and yearly correlations over two time periods, 1951–1980 and 1981–2013. From year to year, Spearman's rank correlation coefficient between rain amount and duration changes in the range of 0.54–0.78. Over time, there is a noticeable decline in the correlation. In the first time period it amounts to 0.69, while in the other it decreases to 0.64. The mean rank correlations were found significantly different at the 5% confidence level with a Tukey-Cramer differences test (Duncan 1955; Ramachandran and Tsokos 2009). Significant difference is also visible for two seasons: winter and summer; it is also statistically significant at the 5% confidence level. In winter months (December, January, February) the correlation is almost 0.05 higher than the average for the whole in both time periods, while in summer (June, July, August) the link between the variables is weaker. In spring and fall the correlation is similar to the one observed when analyzing the whole year. Similarly, empirical estimates of $P(U > u|V > u)$ for $u = 0.95$ give the highest value in winter (0.79) and lowest in summer (0.62). However, the probability of extreme rainfall amount given duration, increased slightly between the two time periods, from 0.59 and 0.61, as opposed to a decline in correlation. That points to possible changes in dependency structure over time.

The rainfall data also shows spatial variability, though due to relatively uniform terrain and small size of the Netherlands they are not much pronounced. Fig. 6 and Table 2 present the rank correlations of X_A and X_D for the entire series available for each station for the whole year and by season. Most of the correlations are

Table 2. Temporal and Seasonal Variability of Rank Correlation between Rain Amount and Duration at Weather Stations in the Netherlands with More Than 38 Years of Record, Split into Two Equal Periods

		Correlation between X_A and X_D—first period					Correlation between X_A and X_D—second period				
Number	Name	Year	DJF	MAM	JJA	SOM	Year	DJF	MAM	JJA	SOM
210	Valkenburg	0.711	0.768	0.694	0.695	0.712	0.639[a]	0.697[a]	0.654	0.600[a]	0.636[a]
235	De Kooy	0.692	0.697	0.730	0.715	0.675	0.641[a]	0.683	0.663[a]	0.633[a]	0.623[a]
240	Schiphol	0.687	0.744	0.717	0.652	0.666	0.625[a]	0.663[a]	0.665[a]	0.595[a]	0.619[a]
260	De Bilt	0.703	0.743	0.705	0.682	0.699	0.638[a]	0.688[a]	0.643[a]	0.602[a]	0.649[a]
265	Soesterberg	0.737	0.751	0.736	0.711	0.762	0.663[a]	0.718	0.648[a]	0.610[a]	0.675[a]
270	Leeuwarden	0.590	0.580	0.638	0.574	0.603	0.645[a]	0.682[a]	0.673	0.604	0.646[a]
280	Eelde	0.693	0.724	0.718	0.662	0.696	0.614[a]	0.642[a]	0.633[a]	0.596[a]	0.621[a]
290	Twenthe	0.582	0.626	0.599	0.541	0.600	0.635[a]	0.709[a]	0.626	0.580	0.659[a]
310	Vlissingen	0.678	0.714	0.674	0.672	0.679	0.640[a]	0.696	0.655	0.601[a]	0.631[a]
344	Rotterdam	0.603	0.649	0.608	0.588	0.603	0.652[a]	0.712[a]	0.666[a]	0.596	0.650[a]
375	Volkel	0.585	0.621	0.592	0.564	0.602	0.636[a]	0.701[a]	0.632	0.590	0.656[a]
380	Maastricht	0.675	0.705	0.684	0.655	0.695	0.616[a]	0.665[a]	0.634[a]	0.551[a]	0.656

Note: The data are averages from annual correlations.

[a]Indicates that the difference in correlation is significant at $p = 0.05$ between the two time periods.

within the 0.6–0.7 range, with the lowest (0.61 for the whole year) recorded in Eindhoven (370) and the highest in Soesterberg (265). The patterns of dependency between rain amount and duration show similarities to station De Bilt. In all but four stations the correlation is the highest in winter, and in all except one it is the lowest in the summer. Those outlying stations had the highest correlation in spring, and lowest in fall. Empirical estimates of $P(U > u|V > u)$ for $u = 0.95$ are also relatively similar (0.55–0.64). Furthermore, of the one-parameter copulas analyzed in this sudy, Gumbel's fits best to the data from all stations. Rainfall data from Soesterberg, located next to De Bilt, show almost identical properties as the highlighted station. More generally, coastal stations have slightly higher correlations, especially in spring and summer, while the lowest are observed in the south and central regions of the country. In Table 2 a drop in correlation over time can be noticed for all seasons for most stations. However, much less diversity between stations is observed in the most recent decades than in the preceding period.

Expert Judgment on Precipitation

Calibration, Information, and Dependence Calibration Scores

In this section the results of the expert judgment elicitation are analyzed. Firstly, results regarding the expert calibration and information scores are presented. In Table 3 (Table S2 in the Appendix S2) the results of the analysis are presented based on all variables. Calibration and information scores are presented for the eight experts as well as for three decision makers. The equal weight combination, which is a simple average of experts' answers for each question, already gives a better calibration (0.092) than any individual expert (the highest individual calibration score is less than 0.001). Conversely, the EWDM is less informative than any of the experts (0.16 compared with 0.60 for the expert with the lowest information score) because simple arithmetic averaging results in a very wide distribution. Weighting the expert opinions by their performance in seed variables gives a combination that improves both calibration and information, even though the information score is still lower than for the expert with lowest information score. Expert 7 has the best calibration score, but is the least informative, which is a typical pattern in the classical method. Nevertheless, the expert was given the highest weight when constructing the performance-based DM, with 66%. Expert 3, second most highly calibrated and in the middle in information, had 28% weight, and Expert 8 had 5%. The remaining experts received a weight of a fraction of a percent. However, this happens only without optimizing for the DM's calibration score ($\alpha = 0$). When optimizing for the DM's calibration score ($a = 0.0002236$, which is the calibration score of the second-best expert) only two experts are included in the decision maker (last column of Table 3). Expert 7 received a 70% weight and Expert 3 a 30% weight. This didn't make a noticeable difference to the calibration score, but the information score increases again. The difference between the equal weight and performance-based (with DM optimization) decision maker is shown in the example in Fig. 7. Both decision makers are close in their means, but the latter has much lower uncertainty. Therefore, this performance-based combination (with DM optimization) was used to analyze the experts' judgments on the variables of interest together with the equal weight combination to give a broader overview of experts' thoughts on the subjects.

Also in Fig. 7 it is shown how large the differences between the experts can be. Three of the experts were convinced that the maximum rain amount that occurred during a single shower in Rotterdam between 1975 and 2013 was less than 60 mm, while two experts were certain it was above this value. Expert 1 had the largest uncertainty about the answer, with a very high mean of 200 mm. The experts who gave very small estimates were informative, but missed the actual value in their uncertainty distributions. Still, the performance-based DM gives a mean of 67.5 mm, which is not far from the observed value of 76.4 mm. The EWDM provided a fair estimate of 58.1 mm, but again with a very large uncertainty distribution (12–247 mm), as it is substantially influenced by Expert 1.

In general, the score of individual experts are low. This together with a robustness analysis (Table S3 in Appendix S2) indicates that the exercise is not very robust to the choice of calibration variables. One reason for this is the fact that many of the methods used in this exercise (e.g., dependence elicitation and copula modeling) are of very recent use in the field of interest. A second possible explanation is the definition used in the exercise for showers that is explained earlier overestimates the actual number of separate rain events and is not of common use in this field.

Table 3. Calibration, Information, and Weights for Expert Judgment Elicitation of Rainfall

		Information		Weights for performance-based combination (%)	
Expert	Calibration	All variables	Calibration variables	Without DM optimization	With DM optimization
		Experts			
Expert 1	1.95×10^{-10}	1.12	1.26	<0.01	—
Expert 2	1.25×10^{-7}	1.21	1.44	0.02	—
Expert 3	2.24×10^{-4}	0.96	1.08	28.49	30.16
Expert 4	1.81×10^{-8}	0.93	1.01	<0.01	—
Expert 5	7.58×10^{-8}	1.32	1.47	0.01	—
Expert 6	5.73×10^{-12}	0.85	1.05	<0.01	—
Expert 7	8.47×10^{-4}	0.60	0.66	65.98	69.84
Expert 8	4.46×10^{-5}	1.02	1.04	5.49	—
		Decision maker (DM)			
Equal weight combination (EWDM)	0.092	0.16	0.17	X	X
Performance-based combination (without DM optimization)	0.127	0.29	0.31	X	X
Performance-based combination (with DM optimization)	0.127	0.38	0.42	X	X

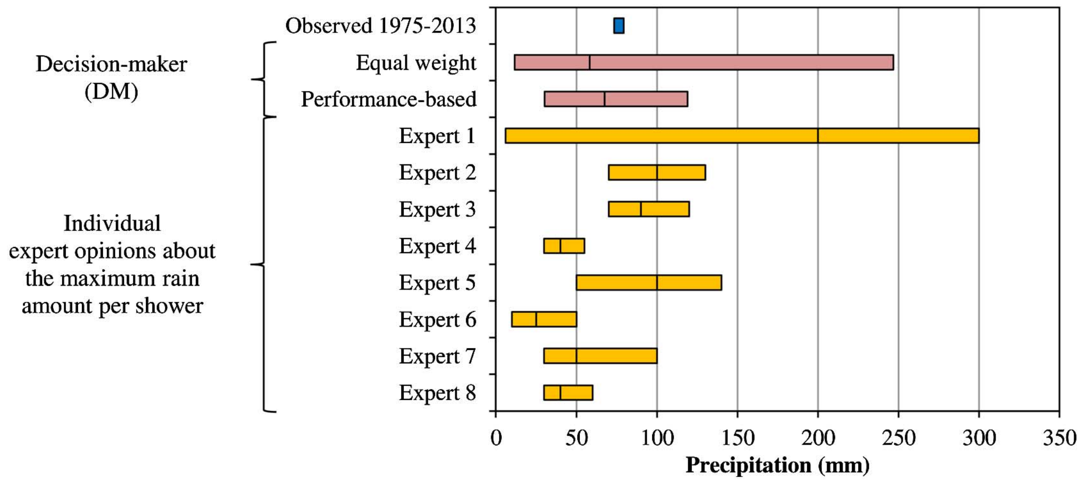

Fig. 7. Comparison of expert opinions, the consensus (decision maker), and observations for maximum rain amount for a shower in Rotterdam, 1975–2013; bars indicate the 5th, 50th, and 95th percentile

Additionally to the standard calibration, the questions regarding dependence with a different approach were analyzed. Extending the method discussed in Morales Nápoles and Worm (2013), the Hellinger distance is used to compare the Gumbel copula generated from the precipitation data and one constructed from an expert's assessment of the tail dependence. The same measure under the Gaussian copula assumption is used in order to combine correlation matrices of experts. The methodology of this calculation is included in Appendix S2, and the results are presented in Table 4. Gaussian $(1 - H)$ and Hellinger $(1 - H_G)$ distances are shown here as a value in the interval [0, 1], where 1 is a perfect match of the copulas. Because the experts did not directly assess the rank correlation between rain amount and duration, this was inferred from two estimates they provided, $P(U > 0.95|V > 0.95)$ and $P(U > 0.95|V < 0.50)$. The way to obtain these estimates is discussed in Morales Nápoles et al. (2008). In case of using the former estimate, Expert 3 achieved the highest result, while with the latter judgment Expert 5 had the highest score. Combining the results using the EWDM does not give satisfactory results, but the performance-based DM is much better. Also, using the estimate of $P(U > 0.95|V < 0.50)$ gives better results than using rank correlations based on $P(U > 0.95|V > 0.95)$.

Table 4. Results of Calibration Based on Rank Correlation, Gaussian $(1 - H_G)$ and Hellinger $(1 - H)$ Distance

Category	Rotterdam $(V > 0.95)$	De Bilt $(V > 0.95)$	Rotterdam $(V < 0.5)$	De Bilt $(V < 0.5)$
	$1 - H_G$			
Expert 1	0.809	0.812	0.894	0.897
Expert 2	0.889	0.892	0.766	0.769
Expert 3	0.960	0.963	0.853	0.856
Expert 4	0.746	0.769	0.960	0.963
Expert 5	0.832	0.812	0.979	0.982
Expert 6	0.733	0.736	0.730	0.733
Expert 7	0.787	0.790	0.730	0.733
Expert 8	0.809	0.812	0.894	0.897
	$1 - H$			
Expert 1	0.822	0.825	0.900	0.903
Expert 2	0.895	0.899	0.784	0.787
Expert 3	0.962	0.965	0.862	0.865
Expert 4	0.767	0.787	0.962	0.965
Expert 5	0.843	0.825	0.980	0.983
Expert 6	0.756	0.759	0.753	0.756
Expert 7	0.802	0.805	0.753	0.756
Expert 8	0.822	0.825	0.900	0.903
	Calibration score			
Equal-weight DM	0.814	0.817	0.837	0.841
Performance-based DM	0.960	0.963	0.979	0.982
	Rank correlation (solution)			
Equal-weight DM	0.264	0.264	0.326	0.326
Performance-based DM	0.578	0.578	0.608	0.608
Realization	0.622	0.617	0.622	0.617

In previous studies regarding elicitation of dependence Morales Nápoles et al. (2013) it was noticed that experts with highest calibration score are not necessarily the same experts with highest dependence calibration score. In this exercise Experts 7 and 3 receive the highest weight in the combination according to Cooke's method. Expert 7 is amongst the experts with lowest performance in assessing dependence. In contrast, Expert 3 performs high in both elicitation of dependence and uncertainty. However, it is observed that similarly to other exercises, a combination of expert opinions based on the dependence-calibration score typically outperforms individual expert opinions.

Climate Change Predictions of Experts

As noted earlier, one of the primary scopes of the expert judgment exercise was to obtain the experts' assessment of future changes in precipitation patterns in the Netherlands. For that purpose, most questions asked both the historical occurrence (1975–2013) of rain as seed variables, and future occurrence (2015–2053) of rain as variables of interest. Having calculated the DM for each question, it is possible to know what the experts' consensus is on changes in climate properties in the country and at particular locations. The projections, based on the 50th percentiles of the DM's solutions to the questions, are presented in Table 5. For the vast majority of questions, the experts predict an increase in extreme and average rainfall. However, there are differences in assessment between the two types of DMs. For most questions, the EWDM shows higher increase than the performance-based DM, except for predictions of shower duration.

The first set of questions regarded maximum rain amount in Rotterdam and De Bilt. Experts' estimate of this variable for the 2015–2053 time period was typically 10 mm higher than for the

Table 5. Projected Change in Precipitation according to Expert Judgments (Performance-Based DM with Optimization), 50th Percentile

Variable	Projected change (2015–2053 relative to 1975–2013) Performance-based DM (%)	Equal weight DM (%)
Maximum rain amount in a shower in Rotterdam	7.6	9.7
Maximum rain amount in a shower in De Bilt	9.0	12.1
Maximum shower duration in Rotterdam	3.1	0.2
Maximum shower duration in De Bilt	2.9	−1.3
Maximum rain intensity in Rotterdam	10.0	17.0
Maximum rain intensity in De Bilt	10.0	14.0
Average number of showers per year in Rotterdam/De Bilt[a]	0.2	2.2
$P(U > 0.95 \mid V > 0.95)$ in Rotterdam	0.0	9.1
$P(U > 0.95 \mid V > 0.95)$ in De Bilt	0.0	10.4
$P(U > 0.95 \mid V < 0.50)$ in Rotterdam/De Bilt[a]	4.1	18.9
Average yearly rain amount in the Netherlands	6.0	9.9
Maximum winter rain amount in the Netherlands	9.9	8.7
Maximum summer rain intensity in the Netherlands	9.3	15.4

[a]The same values of rainfall given by the experts for both locations.

1975–2013 time period in the 50th percentile, and 20 mm higher in the 95th percentile. Most experts did not modify their lower (5th percentile) estimate, or adjusted it only slightly. Depending on the station and DM combination, the experts projected an 8–2% increase in rain amount. In contrast, most experts did not expect an increase in maximum shower duration. Three experts expected some increase, while the same number thought an opposite trend will happen. The performance-based DM with optimization shows only a 3% increase between the time periods, while the EWDM indicated almost no change in Rotterdam and even a decrease in De Bilt. However, most experts significantly underestimated the historical maximum shower duration, which was 30–33 h. A large spread of estimates of maximum rain intensity was also observed, although the DM's 50th percentile was close to the actual value. In experts' consideration, a 10–17% increase in maximum intensity may occur. All but one expert forecasted this increase. Meanwhile, four experts did not expect the average number of showers per year to increase. The same number grossly overestimated and underestimated the observed values in this variable. No expert considered the average number of showers to be different between the two analyzed locations, which is reasonable (in De Bilt there are only 7% more showers per year than in Rotterdam).

Further questions analyzed the ability of experts to assess the joint distribution of rainfall amount and duration. First, they were asked to estimate in what percentage of events, that the rainfall duration exceeded the 95th percentile, the rainfall amount will also be above the 95th percentile. This is the same as estimates of $P(U > u_{95} \mid V > u_{95})$ described in section "Dependence of Rain Amount and Duration." The experts mostly underestimated the upper tail dependence between the two variables as described by the probability of interest. Only one of the experts considered any difference between Rotterdam and De Bilt. The performance-based DM's consensus was 30% for the probability of interest, which is about half the actual value. This result supports the conclusion from previous studies that experts' performance in the assessments of dependence should be calibrated differently than estimates of uncertainty. Moreover the combination of experts' dependence estimates should also be performed with different procedures as those deign to assess uncertainty.

The two experts with highest overall performance did not forecast any change in the dependence, while the EWDM indicated a 9–10% increase, or approximately 2.5 percentage points. Another question was identical, but for the rainfall duration being below the 50th percentile. It is very rare for rainfall amount to be higher than the 95th percentile, as it occurs only for approximately 0.2% of showers shorter than the median. This is nonetheless of large interest, as those showers have very high intensity. All experts overestimated the occurrence of such events and the average answer in the 50th percentile was approximately 1.5%. Only three experts thought that there will be an increase in the percentage of such events in the future, and one expert's 50th percentile was lower for the time period 2015–2030 than for 1975–2013, but his upper and lower estimate remained the same. Average relative increase for the EWDM was 19%, but that corresponds to only 0.3 percentage points. For the performance-based DM, the difference is smaller than 0.2 percentage points.

Some final questions were more general and related to the whole Netherlands or a set of 11 stations with long records. Those questions could be compared with some official KNMI predictions based on climate models. First, the experts did well estimating the mean annual rainfall and predicted a 6–10% increase by 2015–2053. In the KNMI'14 scenarios, the increase is lower, only 2.5–5.5%. Second, more variation in uncertainty estimates was observed across experts for the maximum winter precipitation (mostly underestimated it by experts). The comparison is presented in Fig. 8. The observed value during 1975–2013 was 351 mm. Bars show KNMI's observations and predictions (2050s) for the minimum–mean–maximum precipitation. The change in the winter maximum is predicted to be approximately 3.5–17%, depending on scenario. The largest increase is forecasted in the WH scenario (2°C temperature rise, high change in circulation), in contrast to mean annual precipitation, where the highest increase was found in WL scenario. The estimate of the DM of the maximum precipitation is closer to the observed mean; however the range of the EWDM is very uninformative (100–530 mm). Large spread of individual experts' estimates are shown. The uncertainty distributions of one-half of the experts are below the observed maximum. Interestingly, the performance-based DM's prediction is a 9.9% increase by 2015–2053, which is almost identical to the average of the four KNMI scenarios (9.4% by 2050s) presented by KNMI (2014). The EWDM indicates a smaller increase (8.7%), while all experts indicated a rise in maximum winter precipitation by 10–50 mm.

The final question regarded maximum summer rainfall intensity. Again, there was a large spread in the answers, but all experts predicted an increase in intensity in the future. The performance-based DM's consensus in the 50th percentile for 2015–2053 is 9.3% larger than for 1975–2013 (for EWDM, 15.4%). According to the KNMI, there is large uncertainty in the predictions of this

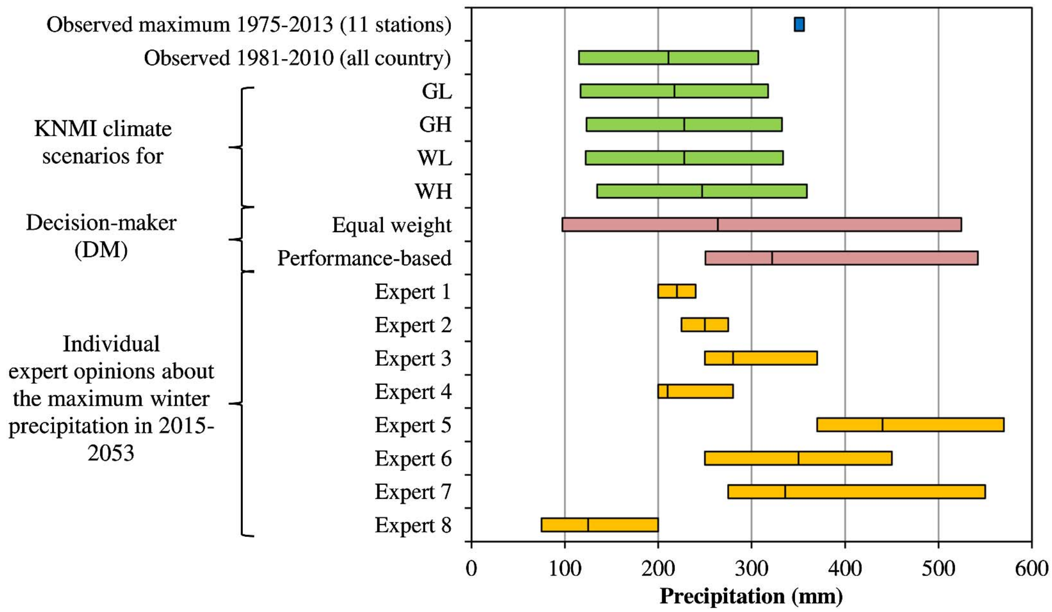

Fig. 8. Comparison of expert opinions, the consensus (decision maker), KNMI climate scenarios, and observations for winter precipitation amount in the Netherlands; bars indicate the 5th, 50th, and 95th percentile

parameter. The institute predicts that rain intensity considered in an hourly resolution (as opposed to 6 min used in this paper) will change at least by 5.5–11% (GL scenario) up to 13–25% (WH scenario). The average of those predictions is 13.8%, so not far from the EWDM's estimate.

Impact of Rain on Infrastructure and Adaptation Measures

In this section an example application of the methodology described in the previous section is presented. It is a model for flooding of a tunnel based on a Gumbel copula (see section "Dependence of Rain Amount and Duration") quantified with measurements for Rotterdam (1975–2013). This example is based on the model presented in Huibregtse et al. (2013, 2016). Although simplified data for an existing tunnel in the Netherlands is used, the example should be considered as a hypothetical case. Also the use of Rotterdam weather data does not imply that the tunnel is located anywhere near the meteorological station.

The tunnel has two entrances, presented in Fig. 9, and a drainage system comprised of pumps and cellars. During intensive rainfall, water may flow through the entrances at a rate higher than the pumping capacity. This will not cause a flooding because the excess water is stored in the cellars. Inundation will occur then, only if the cellars' capacity is exceeded during a long, intense shower.

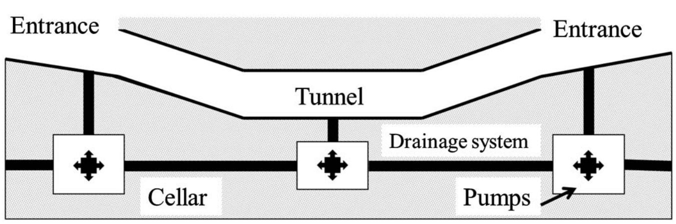

Fig. 9. Schematized (not to scale) representation of the tunnel used as test case

The tunnel has three cellars, each with three pumps (all elements with different capacity). The water is pumped between cellars if any is full. In such a configuration, the limit state (Z) function can be written as

$$Z = V_{\text{cell}} - (A \times X_I - Q_{\text{pump}}) \times X_D \tag{11}$$

where V_{cell} = capacity of the cellars (387 m^3); A = area from which the rainfall is collected (26,265 m^2); Q_{pump} = capacity of the pumps (14 m^3/min); X_I = rainfall intensity; and X_D = rainfall (shower) duration. The parameters used in the model are summarized in Table 6. A Monte Carlo simulation can now be performed by randomly sampling the Gumbel copula. Out of 10,000,000 samples, the limit state was reached only in 63 samples. Given the average annual number of showers in Rotterdam (623.8), this corresponds to an annual probability of occurrence of 0.39%, or 1 in 254 years. In contrast, in station De Bilt the return period is even lower (792 years). The annual probability of occurrence in Rotterdam (return period of 1 in 254 years) is similar to the desired standard for Dutch tunnels, which is a failure probability of 1 in 250 years (Huibregtse et al. 2013). The low probability of failure and the difference between the stations can be explained by the very extreme

Table 6. Characteristics Tunnel (Used as Test Case)

Characteristic	Value	Unit
Length entrance 1	320	m
Length entrance 2	530	m
Length of the closed part of the tunnel	547.5	m
Width tunnel	30.9	m
Area where water is collected (multiplying length of entrances by width of tunnel)	26,265	m^2
Volume of middle pump cellar	71	m^3
Volume main pump cellar 1	158	m^3
Volume main pump cellar 2	158	m^3
Number of pumps per cellar	3	—
Capacity per pump of middle cellar	1.332	m^3/min
Capacity per pump of main cellars	1.67	m^3/min

parameters of the shower. An average event in Rotterdam resulting in flooding of the tunnel discharged 29 mm of rain in 18 min. Therefore very high intensity is needed (1–4 mm/min) to inundate the tunnel. None such event was observed in the empirical data, as shown in Fig. 3, but the copula indicates that such a possibility exists.

Using the results of the expert judgment in order to investigate future trends in flooding, the marginal distribution of rain amount and correlation with rain duration can be modified. The combined expert opinion indicates an increase of annual rainfall by 6% in the Netherlands and increase in number of showers by 0.2% in Rotterdam (see Table 5 for performance-based DM), therefore an average shower will have 5.8% more rain amount. The increase in correlation $P(U > 0.95|V < 0.50)$, which is a good proxy for overall correlation as shown in Table 5, is expected to amount to 4.1% in Rotterdam. In this configuration the probability of failure increased only slightly, to 0.0043%, or 1 in 235 years. The difference is small because the increase in rain amount was offset by the change in the dependence structure. Higher correlation translates to a lower probability of an event occurring in the upper left quadrant, where most of the events causing failure in this example are positioned.

Discussion

In this paper the characterization of rain intensity through copulas was proposed. When modeling through copulas the marginal distributions and the dependence may be assessed separately. The assessment of marginal distributions by experts also has been extensively discussed in the past (Cooke 1991). The research related to the quantification of dependence models through expert judgments is only in its infancy (Werner et al. 2017). In this paper, a real-life exercise was provided for the first time where expert opinions regarding dependence are quantified and combined in a structured protocol using the dependence calibration measure. In the context of climate change the copula models can be investigated not only for future predictions in the presence of data, but also when expert judgments are the only option available, as it has been illustrated in the previous section and summarized in Table 5. The parameters used to investigate future predictions, including changes in marginal distributions or dependence, may be selected on the basis of empirical evidence, also when expert opinions serve as input for the models under investigation. Oppenheimer et al. (2016) noted that in predicting the future one should not be too surprised when it arrives. They also argue that the degree of surprise can be measured by comparing new observations with the probability assigned to them by the quantification of uncertainty. However, the probability assigned to them depends also on the dependence pattern of uncertain quantities. One method was proposed to test empirically the performance of experts as assessors of dependence rather than assessors of uncertainty, but other methods may also be possible.

With respect to the case study, a single infrastructure element (a tunnel) subject to a single hazard (extreme precipitation intensity) has been illustrated. For decision makers, a complex system consisting of multiple elements (roads, bridges, harbors, buildings, etc.) subject to different hazards (coastal flood, river flood, compound flood, earthquakes, etc.) may be of interest. If this is the case the assessment of dependence would become even more critical. Yet, the methods discussed in this paper may still be used to assess experts' performance as dependence assessors at this scale if required. Note that in this case the dependence between different infrastructure elements in the particular system under investigation, together with the hazard to which they are exposed, need also be addressed.

Conclusions

In this paper two methodologies were explored that contribute to better assessment of risks related to extreme rainfall events. The first part was based on fitting one-parameter bivariate copulas for precipitation measurements from rain gauges in the Netherlands. It has been shown that from the three models considered, the Gumbel copula, which indicates upper tail dependence, represents the data most accurately. Rank correlation coefficient in the interval (0.6, 0.7) was observed across all 33 measurement station in the Netherlands. Upper tail dependence was also identified in all stations. Seasonal variability is noticeable, with the highest rank correlation and upper tail dependence in the winter, and the lowest in the summer. For station De Bilt, the value of the mean yearly rank correlation coefficient for the periods 1951–1980 and 1981–2013 was found to be significantly different, though the absolute value could be within sampling fluctuation. This decrease was observed for all seasons, albeit upper tail dependence actually increased in the same time period. Yet, the analysis presented in this paper does not exhaust the wealth of copula types. For future research it is recommended to investigate also multiple-parameter copula families.

In the second part of the study, an expert judgment elicitation was undertaken. The experts' opinions were combined in a structured manner using Cooke's (1991) classical method in order to obtain estimates of future changes in precipitation patterns. Experts predicted mostly an approximate 10% increase in variables such as rain amount, duration, intensity, and the dependence between amount and duration. The experts' forecasts were similar for three variables, which could be compared with KNMI'14 scenarios based on numerical modeling. Their expectation results in a higher increase of annual precipitation than KNMI's models predict; for maximum winter precipitation amount and maximum summer rain intensity, the difference depended on the method of combining of experts' opinions. Experimental results of calibrating experts' opinions on dependency structure were also presented. They were based on the difference between copula parameter estimates by the experts and taken from actual data. Similar conclusions as observed in previous studies may be provided. The calibration and combination of expert's dependence estimates must be done with measures different than those used in Cooke's method. This is an active research area and the methods are still in their infancy. Much remains to be done in this area.

Applicability of both methods was presented based on an example of an existing tunnel in the Netherlands. A bivariate distribution of rain amount and duration is an efficient solution of the tunnel's limit state function, showing that the probability of failure in the historical climate is low (less than 1 in 250 years). Within the parameters analyzed in the expert judgment exercise, the effect of climate change was limited, though more research is needed with the use of multivariate copulas.

Acknowledgments

This research was financed by the TNO project "Graphical Methods for Systems' Risk and Reliability" (GAMES2R) under research program "Enabling Technologies—Models." The authors are grateful to all eight experts who participated in this study.

Supplemental Data

Appendixes S1 and S2, including Tables S1–S3, are available online in the ASCE Library (www.ascelibrary.org).

References

Alam, M. S., and Amin Elshorbagya, A. (2015). "Quantification of the climate change-induced variations in intensity-duration-frequency curves in the Canadian prairies." *J. Hydrol.*, 527, 990–1005.

Balistrocchi, M., and Bacchi, B. (2011). "Modelling the statistical dependence of rainfall event variables through copula functions." *Hydrol. Earth Syst. Sci.*, 15(6), 1959–1977.

Bamber, J. L., and Aspinall, W. P. (2013). "An expert judgement assessment of future sea level rise from the ice sheets." *Nat. Clim. Change*, 3(4), 424–427.

Bárdossy, A., and Pegram, G. (2014). "Infilling missing precipitation records—A comparison of a new copula-based method with other techniques." *J. Hydrol.*, 519, 1162–1170.

Barredo, J. I. (2007). "Major flood disasters in Europe: 1950–2005." *Nat. Hazards*, 42(1), 125–148.

Ben-Zvi, A. (2009). "Rainfall intensity-duration-frequency relationships derived from large partial duration series." *J. Hydrol.*, 367(1–2), 104–114.

Bernard, M. M. (1932). "Formulas for rainfall intensities of long duration." *Trans. Am. Soc. Civil Eng.*, 96(1), 592–624.

Cantet, P., and Arnaud, P. (2014). "Extreme rainfall analysis by a stochastic model: Impact of the copula choice on the sub-daily rainfall generation." *Stochastic Environ. Res. Risk Assess.*, 28(6), 1479–1492.

Cooke, R. M. (1991). *Experts in uncertainty*, Oxford University Press, New York.

Cooke, R. M. (2013). "Expert judgement assessment: Quantifying uncertainty on thin ice." *Nat. Clim. Change*, 3(4), 311–312.

De Michele, C., and Salvadori, G. (2003). "A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas." *J. Geophys. Res.*, 108(D2), 4067.

Duncan, D. B. (1955). "Multiple range and multiple f tests." *Biometrics*, 11(1), 1–42.

Eurostat. (2016). "Database." ⟨http://ec.europa.eu/eurostat/data/database/⟩ (Jul. 11, 2016).

Genest, C., Rémillard, B., and Beaudoin, D. (2009). "Goodness-of-fit tests for copulas: A review and a power study." *Insurance Math. Econ.*, 44(2), 199–213.

Haddad, K., and Rahman, A. (2014). "Derivation of short duration design rainfalls using daily rainfall statistics." *Nat. Hazards*, 74(3), 1391–1401.

Huibregtse, J. N., Morales Napoles, O., Hellebrandt, L., Paprotny, D., and de Wit, M. S. (2016). "Climate change in asset management of infrastructure: A risk-based methodology applied to disruption of traffic on road networks due to the flooding of tunnels." *Eur. J. Trans. Infrastruct. Res.*, 16(1), 98–113.

Huibregtse, J. N., Napoles, O. M., and Dewit, M. S. (2013). "Flooding of tunnels: Quantifying climate change effects on infrastructure." *Safety, reliability, risk and life-cycle performance of structures and infrastructures*, *Proc., 11th Int. Conf. on Structural Safety and Reliability (ICOSSAR2013)*, Taylor & Francis Group, London, 1463–1470.

Jacob, D., et al. (2007). "An inter-comparison of regional climate models for Europe: Model performance in present-day climate." *Clim. Change*, 81(Suppl 1), 31–52.

Jacob, D., et al. (2014). "EURO-CORDEX: New high-resolution climate change projections for European impact research." *Reg. Environ. Change*, 14(2), 563–578.

Joe, H. (2015). *Dependence modeling with copulas*, Chapman and Hall/CRC, London.

KNMI. (2014). "KNMI'14: Climate change scenarios for the 21st century—A Netherlands perspective." *Scientific Rep. WR2014-01*, De Bilt, Netherlands.

KNMI. (2016). "KNMI data centre." ⟨https://data.knmi.nl/⟩ (May 10, 2016).

Kotowski, A., and Kaźmierczak, B. (2013). "Probabilistic models of maximum precipitation for designing sewerage." *J. Hydrometeorol.*, 14(6), 1958–1965.

Koutsoyiannis, D., and Baloutsos, G. (2000). "Analysis of a long record of annual maximum rainfall in Athens, Greece, and design rainfall inferences." *Nat. Hazards*, 22(1), 29–48.

Laux, P., Vogl, S., Qiu, W., Knoche, H. R., and Kunstmann, H. (2011). "Copula-based statistical refinement of precipitation in RCM simulations over complex terrain." *Hydrol. Earth Syst. Sci.*, 15(7), 2401–2419.

Lenderink, G. (2010). "Exploring metrics of extreme daily precipitation in a large ensemble of regional climate model simulations." *Clim. Res.*, 44(2–3), 151–166.

Modesto Gonzalez Pereira, M., Sanches Fernandes, L., Barros Macário, E., Gaspar, S., and Pinto, J. (2015). "Climate change impacts in the design of drainage systems: Case study of Portugal." *J. Irrig. Drain. Eng.*, 10.1061/(ASCE)IR.1943-4774.0000788, 05014009.

Morales Nápoles, O., Hanea, A. M., and Worm, D. (2013). "Experimental results about the assessments of conditional rank correlations by experts: Example with air pollution estimates." *Safety, reliability and risk analysis: Beyond the horizon*, CRC Press/A.A. Balkema, Amsterdam, Netherlands, 1359–1366.

Morales Nápoles, O., Kurowicka, D., and Roelen, A. (2008). "Eliciting conditional and unconditional rank correlations from conditional probabilities." *Reliab. Eng. Syst. Saf.*, 93(5), 699–710.

Morales Nápoles, O., and Worm, D. (2013). "Hypothesis testing of multidimensional probability distributions WP4 GAMES2R." *TNO Rep. 0100003764*, TNO, Delft, Netherlands.

Morales Nápoles, O., Worm, D., Abspoel-Bukman, L., Huibregtse, J. N., and Courage, W. (2015). "Characterization of present and future precipitation through bivariate copulas and its use in risk assessment of infrastructure." *Safety and reliability of complex engineered systems*, CRC Press/A.A. Balkema, Leiden, Netherlands, 4303–4311.

Oppenheimer, M., Little, C. M., and Cooke, R. M. (2016). "Expert judgement and uncertainty quantification for climate change." *Nat. Clim. Change*, 6(5), 445–451.

Overeem, A., Buishand, A., and Holleman, I. (2008). "Rainfall depth-duration-frequency curves and their uncertainties." *J. Hydrol.*, 348(1–2), 124–134.

Ramachandran, K. M., and Tsokos, C. P. (2009). *Mathematical statistics with applications*, Elsevier, Burlington, MA.

Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011). "Improving pan-European hydrological simulation of extreme events through statistical bias correction of RCM-driven climate simulations." *Hydrol. Earth Syst. Sci.*, 15(8), 2599–2620.

Schölzel, C., and Friederichs, P. (2008). "Multivariate non-normally distributed random variables in climate research—Introduction to the copula approach." *Nonlin. Processes Geophys.*, 15(5), 761–772.

Serinaldi, F. (2009). "A multisite daily rainfall generator driven by bivariate copula-based mixed distributions." *J. Geophys. Res.*, 114(D10), D10103.

Van de Vyver, H. (2015). "Bayesian estimation of rainfall intensity-duration-frequency relationships." *J. Hydrol.*, 529, 1451–1463.

Werner, C., Bedford, T., Cooke, R. M., Hanea, A. M., and Morales-Nápoles, O. (2017). "Expert judgement for dependence in probabilistic modelling: A systematic literature review and future research directions." *Eur. J. Oper. Res.*, 258(3), 801–819.