

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

Morales Napoles, Oswaldo; Paprotny, Dominik; Worm, D; Abspoel-Bukman, L; Courage, W

DOI

[10.1061/AJRUA6.0000914](https://doi.org/10.1061/AJRUA6.0000914)

Publication date

2017

Document Version

Accepted author manuscript

Published in

ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering

Citation (APA)

Morales Napoles, O., Paprotny, D., Worm, D., Abspoel-Bukman, L., & Courage, W. (2017). Characterization of Precipitation through Copulas and Expert Judgement for Risk Assessment of Infrastructure. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3(4), Article 04017012. <https://doi.org/10.1061/AJRUA6.0000914>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

1 **Characterization of precipitation through copulas and expert judgement for risk**
2 **assessment of infrastructure**

3
4 Oswaldo Morales-Nápoles¹, Dominik Paprotny², Daniël Worm³, Linda Abspoel-Bukman⁴ and Wim
5 Courage⁵

6
7 ¹ Assistant Professor of Probabilistic Design, Department of Hydraulic Engineering, Faculty of Civil
8 Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
9 (corresponding author). E-mail: o.moralesnapoles@tudelft.nl

10 ² PhD student, Department of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft
11 University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands. E-mail: d.paprotny@tudelft.nl

12 ³ Researcher, TNO, Anna van Buerenplein 1, 2595 DA Den Haag, The Netherlands. E-mail:
13 daniel.worm@tno.nl

14 ⁴ Researcher, TNO, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands. E-mail:
15 linda.abspoel@tno.nl

16 ⁵ Researcher, TNO, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands. E-mail:
17 wim.courage@tno.nl

18
19 In this paper two methodologies are investigated that contribute to better assessment of risks related to
20 extreme rainfall events. Firstly, we use one-parameter bivariate copulas to analyze rain gauge data in the
21 Netherlands. Out of three models considered, the Gumbel copula, which indicates upper tail dependence,
22 represents the data most accurately for all 33 stations in the Netherlands. We notice seasonal variability, with
23 rank correlation reaching maximum in winter and minimum in summer as well as other temporal and spatial
24 patterns. Secondly, an expert judgment elicitation was undertaken. The experts' opinions were combined using
25 Cooke's classical method in order to obtain estimates of future changes in precipitation patterns. Experts
26 predicted mostly around 10% increase in rain amount, duration, intensity and the dependence between amount
27 and duration. The results were in line with official national climate change scenarios, based on numerical
28 modelling. Applicability of both methods was presented based on an example of an existing tunnel in the
29 Netherlands, contributing to better estimates of the tunnel's limit state function and therefore the probability of
30 failure.

31 **Introduction**

32 Extreme precipitation is a major source of threat to society and infrastructure. It causes, for example,
33 disastrous flash floods, which in Europe were responsible for up to 1000 casualties during a single event
34 (Barredo 2007). Excessive rainfall is also problematic in any areas covered by artificial surfaces, where water
35 does not infiltrate the soil, but instead is removed by the drainage system. If the capacity of those systems is
36 insufficient, water accumulates, resulting in not only direct damage, but also cessation of services. An example is
37 the inundation of roads or tunnels, which apart from damaging their surfaces brings the traffic to a halt. This is of
38 particularly high concern in the Netherlands, which is flat and relies on an extensive network of channels for
39 water management. Moreover, it has high population density resulting in increased fraction of land covered by
40 artificial surfaces—12% compared to an European Union average of 5%—and high volume of traffic on more
41 than 2600 km of freeways and numerous other roads (Eurostat 2016). Therefore, precipitation is one of the most
42 relevant variables investigated by the Royal Netherlands Meteorological Institute (KNMI) and other Dutch
43 research organizations in the context of future climate scenarios for the Netherlands.

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

50 A typical mathematical solution to this problem is to utilize depth-duration-frequency (DDF) or intensity-
51 duration-frequency curves (IDF). These describe either rainfall depth (i.e. total amount that has rained) or
52 intensity as a function of duration for given return periods or probabilities of exceedance. The curves are derived
53 by fitting a parametric probability distribution function to precipitation data for fixed durations, e.g. 1, 2, 4, 8,
54 12 and 24 hours. Then through a regression analysis a linear relation between the parameters of the underlying
55 distribution function and the duration (or a transformation of it) is found. This approach was used as early as the
56 1930s (Bernard 1932) and a large number of studies are available covering a vast selection of countries (e. g.
57 Alam and Elshorbagy 2015, Ben-Zvi 2009, Haddad and Rahman 2014, Kotowski and Kaźmierczak 2013,
58 Koutsoyiannis and Baloutsos 2000, Modesto Gonzalez Pereira 2014, Overeem et al. 2008). This standard
59 approach relies on a linear regression based on a limited number of observation, though improvements using
60 Bayesian statistics were also proposed (Van de Vyver 2015).

61 Another issue related to extreme rainfall is predicting the change in their frequency due to climate change.
62 This can be done with numerical modelling of climate similarly as for other meteorological variables. However,
63 the models are still not as reliable for precipitation as they are for temperature, for example (Jacob et al. 2007),
64 especially the extreme sort which is of interest for us (Lenderink 2010). Furthermore, there is large variation in
65 output between models and emission scenarios, leaving substantial uncertainty about how the distribution of
66 heavy precipitation will change in the future (Rojas et al. 2011, Jacob et al. 2014).

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

73 Both approaches are shown on examples from the Netherlands, with rainfall data used for analyzing the
74 joint distributions (Sections “Materials and methods” & “Dependence of rain amount and duration”), and an
75 expert judgment session with local experts assessing the influence of climate change on rainfall patterns in the
76 Netherlands (Section “Expert judgment on precipitation”). Finally, it is presented how this information can be
77 used in practice in infrastructure management (section “Impact of rain on infrastructure”). The case presented in
78 that section refers to a hypothetical tunnel. As stated before, our goal is to present a generic methodology
79 available to researchers for the characterization of precipitation and its use in risk assessment of infrastructure.

80 **Materials and methods**

81 **Precipitation data and climate scenarios**

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

87 For each station, rain data show the fraction of the hour during which rainfall was recorded and the amount
88 of rain. The fractions are in 0.1 increments, ranging from 0.1 to 1, while the amount is given in mm. The

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

102 In the experts' elicitation (see section "Structured expert judgment") KNMI's climate projections for the 21st
103 century, known as KNMI'14, were used additionally to rain gauge data. They include a large number of variables
104 and their expected changes between the reference period 1981–2010 and 30-year periods centered around 2030,
105 2050 and 2085. The projections were made in four scenarios, derived from the results of 250 global climate
106 simulation runs with EC-Earth model (<http://www.ec-earth.org/>). Detailed climate modelling for the Netherlands
107 was done with RACMO2, KNMI's in-house regional climate model (KNMI 2014). Four scenarios were
108 considered: GL, GH, WL and WH. The 'G' scenarios correspond to an increase in global temperature by 1°C
109 while the 'W' scenarios to an increase by 2°C. 'L' stands for small changes in air circulation patterns while 'H'
110 indicates large changes for the same variable. The resulting projections were compared with the experts'
111 opinions on trends in average and extreme rainfall, for example.

112 **Bivariate copulas**

113 As expressed in the introduction, it is intended to model the dependency between rainfall amount and
114 duration using copulas. A copula can be loosely defined as the joint distribution on the unit hypercube with
115 uniform (0,1) margins. For the most comprehensive description of copulas the reader is referred to Joe (2014).
116 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:

$$F_{X_i, X_j}(X_i, X_j) = C_\theta \left(F_{X_i}(X_i), F_{X_j}(X_j) \right) \quad (1)$$

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

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

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

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

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

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

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

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

129 The last one, the Clayton copula is parameterized by α :

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

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

$$\lambda_U = \lim_{u \rightarrow 1} P\left(X_i > F_{X_i}^{-1}(u) | X_j > F_{X_j}^{-1}(u)\right) = \lim_{u \rightarrow 1} P(U > u | V > u) \quad (7)$$

133 Roughly, a value of $\lambda_U > 0$ in equation 7 indicates that it is likely (more than normal) to observe high values
 134 of, say, X_D (rain duration) together with high values of X_A (rain amount). Lower tail dependence would be
 135 defined similarly, but for low values of the marginals. In the Gaussian copula, there is no tail dependence $\lambda_U =$
 136 0, while in Clayton lower tail dependence $\lambda_U = 2^{-\frac{1}{\alpha}}$ is represented and the Gumbel copula presents upper tail
 137 dependence $\lambda_U = 2 - 2^{\frac{1}{\delta}}$. These three types of copulas already cover a range of dependence structures that are

138 typically observed in the data.

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

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

143 where $B(\mathbf{u}) = \sum 1(U_i \leq \mathbf{u})$ is the empirical copula and $C_{\hat{\theta}_n}(\mathbf{u})$ is a parametric copula with parameter $\hat{\theta}_n$
144 estimated from the sample. It can be noted that this statistic is the sum of squared differences between the
145 empirical copula and the parametric estimate.

146 As another diagnostics tool, semi-correlations are analyzed, which is an approach suggested by Joe (2014).
147 The semi-correlations are the Pearson’s product moment correlation coefficients computed in the upper and
148 lower quadrants of the normal transforms of the original variables. If the correlations are positive, semi-
149 correlations in the upper right (NE) and lower left (SW) quadrants are computed using the following formulas:

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

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

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

154 Application of copulas to estimation of rainfall patterns is still fairly new (see e.g. De Michele and Salvadori
155 2003). Large part of the applications focused on correcting climate model data (Laux et al. 2011), filling gaps in
156 data series (Bárdossy and Pegram 2014), analyzing dependencies between stations (Schölzel and Friederichs
157 2008) with some studies exploring the relationship of rainfall amount and duration (Serinaldi 2009, Balistocchi
158 and Bacchi 2011, Cantet and Arnaud 2014). In this study however, as pointed out earlier, we provide an
159 approximation to the bivariate copula of amount and duration and corroborate it with data for the Netherlands.
160 We also show how expert opinions regarding future trends for precipitation may be obtained in a structured way.

161 **Structured expert judgment**

162 Structured expert judgment is a method of quantifying and generalizing opinions of experts. Here, the so-
163 called Classical, or Cooke’s, method (Cooke 1991) is applied, which aims to derive rational consensus from

164 experts' judgments. Roughly, experts are asked to provide their assessment over a continuous quantity.
165 Importantly, they do not give a single 'best estimate', but rather their uncertainty distribution over certain
166 quantities. The experts make their estimates in certain percentiles, most commonly the 5th, 50th and 95th
167 percentile. Plainly, but not entirely strictly, speaking, the expert giving the 5th percentile expresses that he would
168 be very surprised if the actual value of the variable in question was smaller than his 5th percentile estimate.
169 Conversely, the 95th percentile estimate will be the value for which the expert would be surprised if the variable
170 have exceeded it; the 50th percentile is thus the expert's 'best estimate'.

171 The experts are asked two types of questions which they answer in the way described above (see section 1
172 of Supplement 1 and Table S1). First group are the seed, or calibration, variables. These are quantities whose
173 value is known, or will be known within the time frame of the research, to the analysts but not to the experts at
174 the moment of the elicitation. The variables are used to assess the reliability of each expert. In Cooke's model,
175 two measures of performance are computed: the calibration and information scores (section 2 of Supplement 1).
176 Roughly, calibration measures the degree to which experts are statistically accurate, while information measures
177 the degree to which experts' uncertainty estimates are concentrated relative to a background measure. Good
178 expertise in the classical method refers to highly calibrated (typically calibration scores > 0.05) and highly
179 informative experts.

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

187 The elicitation of dependence estimates from experts is a subject of active research. It has been observed in
188 recent studies that though Cooke's method is an appropriate estimate to get empirical validation of experts
189 estimates over uncertainty distributions it fails to provide empirical evidence regarding expert's ability to
190 quantify dependence (Morales-Nápoles et al. 2014). For this reason the *dependence-calibration* score introduced
191 in (Morales Nápoles O and Worm 2013) was used (also discussed in section 2 of Supplement 1). The dependence
192 calibration score is a measure of distance between a certain dependence structure used for calibration purposes
193 and estimates of this dependence structure provided by experts. A value close to one would indicate expert's

194 ability to quantify dependence.

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

- 205 • *For each of the following two stations: **Rotterdam** and **De Bilt**. Consider the shower data as described*
206 *in [section “precipitation data and climate scenarios” of the paper] and think of the shower with the*
207 *maximum value of **rain amount** in the period of interest (**January 1st 1975 – December 31st 2013**).*
208 *What would this maximum value **of rain amount** be (in mm)?*

209 The experts gave here their estimates for the two stations providing the 5th, 50th and 95th percentile of their
210 uncertainty distribution. The exact same question was asked for the 2015–2053 time period as a variable of
211 interest. Having discussed the main methods and materials to be used in this research we now present the
212 principal findings of the study. Additional details on expert judgment methodology and a summary of all
213 questions used in the elicitation may be found in Supplement 1.

214 **Dependence of rain amount and duration**

215 **Results for De Bilt station**

216 De Bilt weather station is located at the KNMI's headquarters and has the longest data series of the entire
217 network. It was chosen to present here the copula approach to analyzing showers; spatial and temporal variation
218 in the Netherlands is shown in the next section. In Figure 3 the transformation from observations (left graph) to
219 pseudo-observations (right graph) is shown. The pseudo-observations are the samples transformed to the interval
220 (0,1) through the empirical margins.

221 For this station, the rank correlation coefficient for the whole data series is 0.66. Ties, i.e. samples realizing

222 the same pair of values, are visible in the lower tail of the distribution – that is, for small values of both rain
223 amount and duration. It also appears that more samples are concentrated in the upper right tail of the joint
224 distribution than elsewhere. That feature suggests upper tail dependence. However, the upper tail dependence
225 does not necessarily lead to the most intense showers. Even though the upper right corner of the distribution
226 contains the highest amounts of rain per shower, it also contains the longest shower durations. There are 13
227 showers with durations larger than half hour, which is about the 60th percentile of the distribution of X_D , with an
228 intensity larger than 20 mm per hour (marked red in Figure 3). These correspond to amounts of rain between
229 12.6 and 31 mm.

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

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

246 **Temporal and spatial variability in correlation**

247 Analysis of shower time series indicated noticeable temporal variability. In Figure 5, the annual variations
248 for De Bilt station are presented alongside seasonal and yearly correlations over two time periods, 1951–1980
249 and 1981–2013. From year to year, Spearman's rank correlation coefficient between rain amount and duration
250 changes in the range of 0.54–0.78. Over time, there is a noticeable decline in the correlation. In the first time
251 period it amounts to 0.69, while in the other it decreases to 0.64. The mean rank correlations were found

252 significantly different at the 5% confidence level with a Tukey-Cramer differences test (Duncan 1955,
253 Ramachandran and Tsokos 2009). Significant difference is also visible for two seasons: winter and summer; it is
254 also statistically significant at the 5% confidence level. In winter months (December, January, February) the
255 correlation is almost 0.05 higher than the average for the whole in both time periods, while in summer (June,
256 July, August) the link between the variables is weaker. In spring and fall the correlation is similar to the one
257 observed when analyzing the whole year. Similarly, empirical estimates of $P(U>u/V>u)$ for $u=0.95$ give the
258 highest value in winter (0.79) and lowest in summer (0.62). However, the probability of joint occurrence of
259 extreme rainfall amount and duration increased slightly between the two time periods, from 0.59 and 0.61, as
260 opposed to a decline in correlation. That points to possible changes in dependency structure over time.

261 The rainfall data also shows spatial variability, though due to relatively uniform terrain and small size of the
262 Netherlands they are not much pronounced. Fig. 6 and Table 2 presents the rank correlations of X_A and X_D for the
263 entire series available for each station for the whole year and by season. Most of the correlations are within the
264 0.6–0.7 range, with the lowest (0.61 for the whole year) recorded in Eindhoven (370) and the highest in
265 Soesterberg (265). The patterns of dependency between rain amount and duration show similarities to station De
266 Bilt. In all but four stations the correlation is the highest in winter, and in all except one it is the lowest in the
267 summer. Those outlying station had the highest correlation in spring, and lowest in fall. Empirical estimates of
268 $P(U>u/V>u)$ for $u=0.95$ are also relatively similar (0.55–0.64). Furthermore, of the one-parameter copulas
269 analyzed here, Gumbel's fits best to the data from all stations. Rainfall data from Soesterberg, located next to De
270 Bilt, show almost identical properties as the highlighted station. More generally, coastal stations have slightly
271 higher correlations, especially in spring and summer, while the lowest are observed in the south and central
272 regions of the country. In Table 2 a drop in correlation over time can be noticed for all seasons for most stations.
273 However, much less diversity between stations is observed in the most recent decades than in the preceding
274 period.

275 **Expert judgment on precipitation**

276 **Calibration, information and dependence calibration scores**

277 In this section the results of the expert judgment elicitation are analyzed. Firstly, results regarding the expert
278 calibration and information scores are presented. In Table 3 (Table S2 in the supplement) the results of the
279 analysis are presented based on all variables. Calibration and information scores are presented for the eight

280 experts as well as for three “decision-makers” (DM). The equal weight combination, which is a simple average
281 of experts’ answers for each question, already gives a better calibration (0.092) than any individual expert (the
282 highest individual calibration score is less than 0.001). On the other hand, the EWDM is less informative than
283 any of the experts (0.16 compared to 0.60 for the expert with the lowest information score), since simple
284 arithmetic averaging results in a very wide distribution. Weighting the expert opinions by their performance in
285 seed variables gives a combination that improves both calibration and information, even though the information
286 score is still lower than for the worst expert. Expert 7 has the best calibration score, but is the least informative,
287 which is a typical pattern in the classical method. Nevertheless, the expert was given the highest weight when
288 constructing the performance-based DM, with 66%. Expert 3, second most highly calibrated and in the middle in
289 information, had 28% weight, and expert 8 had 5%. The remaining experts received a weight of a fraction of a
290 percent. This happens, however, only without optimizing for the DM’s calibration score ($\alpha = 0$). When
291 optimizing for the DM’s calibration score ($\alpha = 0.0002236$, which is the calibration score of the second best
292 expert) only two experts are included in the decision-maker (last column of Table 3). Expert 7 received a 70%
293 weight and expert 3 a 30% weight. This didn’t make a noticeable difference to the calibration score, but the
294 information score increases again. The difference between the equal weight and performance-based (with DM
295 optimization) decision-maker can be also seen in the example in Fig. 7. Both decision-makers are close in their
296 means, but the latter has much lower uncertainty. Therefore, this performance-based combination (with DM
297 optimization) was used to analyze the experts’ judgments on the variables of interest together with the equal
298 weight combination to give a broader overview of experts’ thoughts on the subjects.

299 In Fig. 7 it can also be seen how large the differences between the experts can be. Three of the experts were
300 convinced that the maximum rain amount that occurred during a single shower in Rotterdam between 1975 and
301 2013 was less than 60 mm, while two experts were certain it is above this value. Expert 1 had the largest
302 uncertainty about the answer, with a very high mean of 200 mm. The experts who gave very small estimates
303 were informative, but missed the actual value in their uncertainty distributions. Still, the performance-based DM
304 gives a mean of 67.5 mm, which is not far from the observed value of 76.4 mm. The EWDM provided a fair
305 estimate 58.1 mm, but again with a very large uncertainty distribution (12–247 mm), as it is substantially
306 influenced by expert 1.

307 In general, the score of individual experts are low. This together with a robustness analysis (Table S3 in
308 Supplement 1) indicates that the exercise is not very robust to the choice of calibration variables. One reason for
309 this is the fact that many of the methods used in this exercise (dependence elicitation and copula modelling for

310 example) are of very recent use in the field of interest. A second possible explanation is the definition used in
311 our exercise for showers which as explained earlier overestimates the actual number of separate rain events and
312 is not of common use in this field.

313 Additionally to the standard calibration, we analyzed the questions regarding dependence with a different
314 approach. Extending the method discussed in Morales Nápoles, O. & Worm, D (2013), we use the Hellinger
315 distance to compare the Gumbel copula generated from the precipitation data and one constructed from an
316 expert's assessment of the tail dependence. We use the same measure under the Gaussian copula assumption in
317 order to combine correlation matrices of experts. The methodology of this calculation is included in Supplement
318 1, and the results are presented in Table 4. Gaussian ($1 - H$) and Hellinger ($1 - H_G$) distances are shown here as
319 a value in the interval $[0, 1]$, where 1 is a perfect match of the copulas. Because the experts did not directly
320 assess the rank correlation between rain amount and duration, this was inferred from two estimates they
321 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
322 et al 2008. In case of using the former estimate, Expert 3 achieved the highest result, while with the latter
323 judgment, it is Expert 5 with highest score. Combining the results using the EWDM does not give satisfactory
324 results, but the performance-based DM is much better. Using the estimate of $P(U > 0.95 / V < 0.50)$ gives better
325 results, too, than using rank correlations based on $P(U > 0.95 / V > 0.95)$.

326 In previous studies regarding elicitation of dependence Morales Nápoles et al (2014) it was noticed that
327 experts with highest calibration score are not necessarily the same experts with highest dependence calibration
328 score. In this exercise experts 7 and 3 receive the highest weight in the combination according to Cooke's
329 method. Expert 7 is amongst the experts with lowest performance in assessing dependence. Expert 3 in contrast
330 performs high in both elicitation of dependence and uncertainty. It is however observed that similarly to other
331 exercises, a combination of expert opinions based on the dependence-calibration score typically outperforms
332 individual expert opinions.

333 **Climate change predictions of experts**

334 As noted earlier, one of the main scopes of the EJ exercise was to obtain the experts' assessment of future
335 changes in precipitation patterns in the Netherlands. For that purpose, most questions asked both the historical
336 occurrence (1975–2013) of rain as seed variables, and future occurrence (2015–2053) of rain as variables of
337 interest. Having calculated the DM for each question, it is possible to know what is the experts' consensus on
338 changes in climate properties in the country and at particular locations. The projections, based on the 50th
339 percentiles of the DM's solutions to the questions, are presented in Table 5. For the vast majority of questions,

340 the experts predict an increase in extreme and average rainfall. There are, however, differences in assessment
341 between the two types of DMs. For most questions, the EWDM shows higher increase than the performance-
342 based DM, except for predictions of shower duration.

343 The first set of questions regarded maximum rain amount in Rotterdam and De Bilt. Experts' estimate of this
344 variable for 2015–2053 was typically 10 mm higher than for 1975–2013 in the 50th percentile and 20 mm higher
345 in the 95th percentile. Most experts did not modify their lower (5th percentile) estimate, or adjusted it only
346 slightly. Depending on the station and DM combination, the experts projected an 8–12% increase in rain amount.
347 In contrast, most experts did not expect an increase in maximum shower duration. Three experts expected some
348 increase, while the same number thought an opposite trend will happen. The performance-based DM with
349 optimization shows only a 3% increase between the time periods, while the EWDM indicated almost no change
350 in Rotterdam and even a decrease in De Bilt. It should be noted, however, that most experts significantly
351 underestimated the historical maximum shower duration, which was 30–33 hours. A large spread of estimates of
352 maximum rain intensity was also observed, though the DM's 50th percentile was close to the actual value. In
353 experts' consideration, a 10–17% increase in maximum intensity may occur. All but one expert forecasted this
354 increase. Meanwhile, four experts did not expect the average number of showers per year to increase. The same
355 number grossly over- and underestimated the observed values in this variable. No expert considered the average
356 number of showers to be different between the two analyzed locations, which is reasonable (in De Bilt there are
357 only 7% more showers per year than in Rotterdam).

358 Further questions analyzed the ability of experts to assess the joint distribution of rainfall amount and
359 duration. Firstly, they were asked to estimate in what percentage of events, that the rainfall duration exceeded the
360 95th percentile, the rainfall amount will also be above the 95th percentile. This is the same as estimates of
361 $P(U > u_{95} / V > u_{95})$ described in section "Dependence of rain amount and duration". The experts mostly
362 underestimated the upper tail dependence between the two variables as described by the probability of interest.
363 Only one of the experts considered any difference between Rotterdam and De Bilt. The performance-based DM's
364 consensus was 30% for the probability of interest, which is about half the actual value. This result supports the
365 conclusion from previous studies that experts performance in the assessments of dependence should be
366 calibrated differently than estimates of uncertainty. Moreover the combination of experts' dependence estimates
367 should also be performed with different procedures as those deign to assess uncertainty.

368 The two experts with highest overall performance did not forecast any change in the dependence, while the
369 EWDM indicated a 9–10% increase, or about 2.5 percentage points. Another question was identical, but for the

370 rainfall duration being below the 50th percentile. It is very rare for rainfall amount to be higher than the 95th
371 percentile, as it occurs only for about 0.2% of showers shorter than the median. This is nonetheless of large
372 interest, as those showers have very high intensity. All experts overestimated the occurrence of such events and
373 the average answer in the 50th percentile was about 1.5%. Only three experts thought that there will be an
374 increase in the percentage of such events in the future, and one expert's 50th percentile was lower for 2015–2030
375 than for 1975–2013, however his upper and lower estimate remained the same. Average relative increase for the
376 EWDM was 19%, but that corresponds to only 0.3 percentage points. For the performance-based DM, the
377 difference is smaller than 0.2 percentage points.

378 Some final questions were more general and related to the whole Netherlands or a set of 11 stations with
379 long records. Those questions could be compared with some official KNMI predictions based on climate models.
380 Firstly, the experts did well estimating the mean annual rainfall and predicted a 6–10% increase by 2015–2053.
381 In the KNMI'14 scenarios, the increase is lower, only 2.5–5.5%. Secondly, more variation in uncertainty
382 estimates was observed across experts for the maximum winter precipitation (mostly underestimated it by
383 experts). The comparison is presented in Figure 8. The observed value during 1975–2013 was 351 mm. Green
384 bars show KNMI's observations and predictions (2050s) for the minimum–mean–maximum precipitation. The
385 change in the winter maximum is predicted to be about 3.5–17%, depending on scenario. The largest increase is
386 forecasted in the WH scenario (2°C temperature rise, high change in circulation), in contrast to mean annual
387 precipitation, where the highest increase was found in WL scenario. The estimate of the DM of the maximum
388 precipitation (red bars) is closer to the observed mean, however the range of the EWDM is very uninformative
389 (100–530 mm). Large spread of individual experts' estimates can be seen (orange bars). The uncertainty
390 distributions of half of the experts are below the observed maximum. Interestingly, the performance-based DM's
391 prediction is a 9.9% increase by 2015–2053, which is almost identical to the average of the four KNMI scenarios
392 (9.4% by 2050s) presented by KNMI (2015). The EWDM indicates a smaller increase (8.7%), while all experts
393 indicated a rise in maximum winter precipitation by 10–50 mm.

394 The final question regarded maximum summer rainfall intensity. Again, there was a large spread in the
395 answers, but all experts predicted an increase in intensity in the future. The performance-based DM's consensus
396 in the 50th percentile for 2015–2053 is 9.3% larger than for 1975–2013 (for EWDM, 15.4%). According to the
397 KNMI, there is large uncertainty in the predictions of this parameter. The institute predicts that rain intensity
398 considered in an hourly resolution (as opposed to 6 minutes used here) will change at least by 5.5–11% (GL
399 scenario) up to 13–25% (WH scenario). The average of those predictions is 13.8%, so not far from the EWDM's

400 estimate.

401 **Impact of rain on infrastructure and adaptation measures**

402

403 In this section we present an example application of the methodology described in the previous section. It is
404 a model for flooding of a tunnel based on a Gumbel copula (see section “Dependence of rain amount and
405 duration”) quantified with measurements for Rotterdam (1975–2013). This example is based on the model
406 presented in Huibregtse et al. (2013, 2016). Although simplified data for an existing tunnel in the Netherlands is
407 used, the example should be considered as a hypothetical case. Also the use of Rotterdam weather data does not
408 imply that the tunnel is located anywhere near the meteorological station.

409 The tunnel has two entrances, presented in Figure 9, and a drainage system comprised of pumps and cellars.
410 During intensive rainfall, water may flow through the entrances at a rate higher than the pumping capacity. This
411 will not cause a flooding, since the excess water is stored in the cellars. Inundation will occur then, only if the
412 cellars’ capacity is exceeded during a long, intense shower. The tunnel has three cellars, each with three pumps
413 (all elements with different capacity). The water is pumped between cellars if any is full. In such a configuration,
414 the limit state (Z) function can be written as:

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

415 where V_{cell} is the capacity of the cellars (387 m³), A is the area from which the rainfall is collected (26265
416 m²), Q_{pump} is the capacity of the pumps (14 m³ per minute), X_I is the rainfall intensity and X_D is the rainfall
417 (shower) duration. The parameters used in the model are summarized in Table 6. We can now perform a Monte
418 Carlo simulation by randomly sampling the Gumbel copula. Out of 10,000,000 samples, the limit state was
419 reached only in 63 samples. Given the average annual number of showers in Rotterdam (623.8), that corresponds
420 to an annual probability of occurrence of 0.39%, or 1 in 254 years. In contrast, in station De Bilt, the return
421 period is even lower – 792 years. The annual probability of occurrence in Rotterdam (return period of 1 in 254
422 years) is similar to the desired standard for Dutch tunnels, which is a failure probability of 1 in 250 years
423 (Huibregtse et al. 2013). The low probability of failure and the difference between the stations can be explained
424 by the very extreme parameters of the shower. An average event in Rotterdam resulting in flooding of the tunnel
425 discharged 29 mm of rain in 18 minutes. Very high intensity is therefore needed, of 1–4 mm per minute, to
426 inundate the tunnel. None such event was observed in the empirical data, as shown in Fig. 3, but the copula
427 indicates that such a possibility exists.

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

437 Discussion

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

454 With respect to the case study, we have illustrated a single infrastructure element (a tunnel) subject to a single
455 hazard (extreme precipitation intensity). For decision makers, a complex system consisting of multiple elements
456 (roads, bridges, harbors, buildings, etc.) subject to different hazards (coastal flood, river flood, compound flood,

457 earthquakes, etc.) may be of interest. If this is the case the assessment of dependence would become even more
458 critical. Yet, the methods discussed here may still be used to assess experts' performance as dependence
459 assessors at this scale if required. It should be noted that in this case the dependence between different
460 infrastructure elements in the particular system under investigation, together with the hazard to which they are
461 exposed, need also be addressed.

462 **Conclusions**

463

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

476 In the second part of the study, an expert judgment elicitation was undertaken. The experts' opinions were
477 combined in a structured manner using Cooke's (1991) classical method in order to obtain estimates of future
478 changes in precipitation patterns. Experts predicted mostly around 10% increase in variables such as rain
479 amount, duration, intensity and the dependence between amount and duration.. For three variables, which could
480 be compared with KNMI'14 scenarios, based on numerical modelling, the experts' forecasts were similar. Their
481 expectation is a higher increase of annual precipitation than KNMI's models predict; for maximum winter
482 precipitation amount and maximum summer rain intensity, the difference depended on the method of combining
483 of experts' opinions. Experimental results of calibrating experts' opinions on dependency structure were also
484 presented. They were based on the difference between copula parameter estimates by the experts and taken from
485 actual data. Similar conclusions as observed in previous studies may be provided. The calibration and

486 combination of expert's dependence estimates must be done with measures different than those used in Cooke's
487 method. This is an active research area and the methods are still in their infancy. Much remains to be done in this
488 area.

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

494 **Acknowledgments**

495

496 This research was financed by TNO project "GrAphical MEthods for Systems' Risk and Reliability"
497 (GAMES2R) under research program "Enabling Technologies – Models". Authors are grateful to all eight
498 experts who participated in this study.

499 **Supplemental Data**

500 The supplement together with Tables S1–S3 are available online in the ASCE Library (ascelibrary.org).

501

502 **References**

- 503 • Alam, M. S., and Amin Elshorbagya, A. (2015). "Quantification of the climate change-induced
504 variations in Intensity–Duration–Frequency curves in the Canadian Prairies" *J. Hydrol.*, 527, 990–1005.
- 505 • Balistrocchi, M. and Bacchi, B. (2011). "Modelling the statistical dependence of rainfall event variables
506 through copula functions." *Hydrol. Earth Syst. Sci.*, 15, 1959–1977.
- 507 • Bamber, J. L., and Aspinall, W. P. (2013). "An expert judgement assessment of future sea level rise from
508 the ice sheets." *Nature Clim. Change*, 3, 424–427.
- 509 • Bárdossy, A. and Pegram, G. (2014). "Infilling missing precipitation records – A comparison of a new
510 copula-based method with other techniques." *J. Hydrol.*, 519, 1162–1170.
- 511 • Barredo, J. I. (2007). "Major flood disasters in Europe: 1950–2005." *Nat. Hazards*, 42(1), 125–148.
- 512 • Ben-Zvi, A. (2009). "Rainfall intensity-duration-frequency relationships derived from large partial

- 513 duration series.” *J. Hydrol.*, 367(1–2), 104–114.
- 514 • Bernard, M.M. (1932). “Formulas for rainfall intensities of long duration.” *Trans. Am. Soc. Civ. Eng.*,
- 515 96, 592–624.
- 516 • Cantet, P., Arnaud, P. (2014). “Extreme rainfall analysis by a stochastic model: impact of the copula
- 517 choice on the sub-daily rainfall generation.” *Stoch. Environ. Res. Risk Assess.*, 28(6), 1479–1492.
- 518 • Cooke, R. M. (1991). *Experts in uncertainty*. Oxford University Press.
- 519 • Cooke, R. M. (2013). “Expert judgement assessment: Quantifying uncertainty on thin ice.” *Nature*
- 520 *Clim. Change*, 3, 311–312.
- 521 • De Michele, C., and Salvadori, G. (2003). “A generalized Pareto intensity-duration model of storm
- 522 rainfall exploiting 2-copulas.” *J. Geophys. Res.*, 108(D2), 4067.
- 523 • Duncan, D. B. (1955). “Multiple range and multiple f tests.” *Biometrics*, 11(1), 1–42.
- 524 • Genest, C., Rémillard, B., Beaudoin, D. (2009). “Goodness-of-fit tests for copulas: A review and a
- 525 power study.” *Insur. Math. Econ.*, 44, 199–213.
- 526 • Haddad, K, and Rahman, A (2014). “Derivation of short duration design rainfalls using daily rainfall
- 527 statistics.” *Nat. Hazards*, 74, 1391–1401.
- 528 • Huibregtse, J. N., Morales Nápoles, O., de Wit, M. S. (2013). “Flooding of tunnels: quantifying climate
- 529 change effects on infrastructure.” *Proc., 11th International Conference on Structural Safety and*
- 530 *Reliability (ICOSSAR2013)*, New York, 1463–1470.
- 531 • Huibregtse, J. N., Morales Napoles, O., Hellebrandt, L., Paprotny, D., de Wit, M. S. (2016). “Climate
- 532 change in asset management of infrastructure: A risk-based methodology applied to disruption of traffic
- 533 on road networks due to the flooding of tunnels.” *Eur. J. Transport. Infrastruct. Res.*, 16(1), 98–113.
- 534 • Jacob, D., Bärring, L., Christensen, O., Christensen, J., de Castro, M., ..., and Hurk, B. (2007). “An
- 535 inter-comparison of regional climate models for Europe: Model performance in present-day climate.”
- 536 *Climatic Change*, 81, 31–52.
- 537 • Jacob, D., Petersen, J., Eggert, B., Alias, A., Christensen, O., ..., and Yiou, P. (2014). “EURO-
- 538 CORDEX: new high-resolution climate change projections for European impact research.” *Reg.*
- 539 *Environ. Change*, 14(2), 563–578.
- 540 • Joe, H. (2015). *Dependence Modeling with Copulas*, Chapman & Hall/CRC, London, U.K.
- 541 • KNMI (2014). “KNMI’14: Climate Change scenarios for the 21st Century – A Netherlands
- 542 perspective”. *Scientific Report WR2014-01*, KNMI, De Bilt, The Netherlands.

- 543 • KNMI (2016). “KNMI Data Centre” <<https://data.knmi.nl/>> (10 May 2016).
- 544 • Kotowski, A., and Kaźmierczak, B. (2013). “Probabilistic Models of Maximum Precipitation for
- 545 Designing Sewerage”. *J. Hydrometeorol.*, 14(6), 1958–1965.
- 546 • Koutsoyiannis, D., and Baloutsos, G., (2000). “Analysis of a long record of annual maximum rainfall in
- 547 Athens, Greece, and design rainfall inferences.” *Nat. Hazards*, 29, 29–48.
- 548 • Laux, P., Vogl, S., Qiu, W., Knoche, H. R., and Kunstmann, H. (2011). “Copula-based statistical
- 549 refinement of precipitation in RCM simulations over complex terrain.” *Hydrol. Earth Syst. Sci.*, 15,
- 550 2401–2419.
- 551 • Lenderink, G. (2010). “Exploring metrics of extreme daily precipitation in a large ensemble of regional
- 552 climate model simulations.” *Clim. Res.*, 44, 151–166.
- 553 • Modesto Gonzalez Pereira, M., Sanches Fernandes, L., Barros Macário, E., Gaspar, S., and Pinto, J.
- 554 (2014). "Climate Change Impacts in the Design of Drainage Systems: Case Study of Portugal." *J. Irrig.*
- 555 *Drain Eng.*, 141(2), 05014009.
- 556 • Morales Nápoles, O. and Worm, D. (2013). “Hypothesis testing of Multidimensional probability
- 557 distributions WP4 GAMES2R.” TNO Report 0100003764, TNO, The Netherlands.
- 558 • Morales Nápoles, O., Hanea, A. M, Worm, D. (2013). “Experimental results about the assessments of
- 559 conditional rank correlations by experts: Example with air pollution estimates.” *Safety, Reliability and*
- 560 *Risk Analysis: Beyond the Horizon*, CRC Press/Balkema, Amsterdam, The Netherlands, 1359–1366.
- 561 • Morales Nápoles, O., Kurowicka, D., and Roelen, A. (2008). “Eliciting conditional and unconditional
- 562 rank correlations from conditional probabilities.” *Reliab. Eng. Syst. Safe.*, 93, 699–710.
- 563 • Morales Nápoles, O., Worm, D., Abspoel-Bukman, L., Huibregtse, J. N., Courage, W. (2015).
- 564 “Characterization of present and future precipitation through bivariate copulas and its use in risk
- 565 assessment of infrastructure.” *Safety and Reliability of Complex Engineered Systems*, CRC
- 566 Press/Balkema, Leiden, The Netherlands, 4303–4311.
- 567 • Oppenheimer, M., Little, C. M., and Cooke, R. M. (2016). “Expert judgement and uncertainty
- 568 quantification for climate change.” *Nature Clim. Change*, 6, 445–451.
- 569 • Overeem, A., Buishand, A., and Holleman, I. (2008). “Rainfall depth-duration-frequency curves and
- 570 their uncertainties.” *J. Hydrol.*, 348(12), 124–134.
- 571 • Ramachandran, K. M. and Tsokos, C. P. (2009). *Mathematical Statistics with Applications*. Elsevier.
- 572 • Rojas, R., Feyen, L., Dosio, A., and Bavera, D. (2011). “Improving pan-European hydrological

573 simulation of extreme events through statistical bias correction of RCM-driven climate simulations.”
574 *Hydrol. Earth Syst. Sci.*, 15, 2599–2620.

575 • Schölzel, C. and Friederichs, P. (2008). “Multivariate non-normally distributed random variables in
576 climate research – introduction to the copula approach.” *Nonlin. Processes Geophys.*, 15, 761–772.

577 • Serinaldi, F. (2009). “A multisite daily rainfall generator driven by bivariate copula-based mixed
578 distributions.” *J. Geophys. Res.*, 114, D10103.

579 • Van de Vyver, H. (2015). “Bayesian estimation of rainfall intensity–duration–frequency relationships.”
580 *J. Hydrol.*, 529, 1451–1463.

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

584

Figure Captions List

585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603

- Fig. 1.** Location of KNMI weather stations used in this study.
- Fig. 2.** Aggregation of rainfall data into showers.
- Fig. 3.** Rain amount (X_A) and duration (X_D): original observations (left) and pseudo-observations (right). Showers lasting more than 0.5 h with intensity higher than 20 mm/h are marked red.
- Fig. 4.** Left: rain amount and duration transformed to standard normal, with semi-correlation for each quadrant. Right: empirical and parametric estimates of $P(U>u/V>u)$ for different probabilities u at De Bilt.
- 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.
- 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.
- 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.
- Fig. 9.** Schematized (not to scale) representation of the tunnel used as test case .

604 **Table 1.** List of KNMI weather stations used in this study. All stations have an end date of 31-12-2013.

No.	Name	Start date	No.	Name	Start date	No.	Name	Start date
210	Valkenburg	1-1-1973	273	Marknesse	1-1-1994	323	Wilhelminadorp	1-1-1994
235	De Kooy	1-1-1957	275	Deelen	1-1-1983	330	Hoek van Holland	1-1-1996
240	Schiphol	1-1-1974	277	Lauwersoog	1-1-1994	344	Rotterdam	1-1-1974
249	Berkhout	1-1-2000	278	Heino	1-1-1994	348	Cabauw	1-1-1987
251	Hoorn (Terschelling)	1-1-1995	279	Hoogeveen	1-1-1994	350	Gilze-Rijen	1-1-1977
257	Wijk aan Zee	1-1-2002	280	Eelde	1-1-1957	356	Herwijnen	1-1-1994
260	De Bilt	1-1-1951	283	Hupsel	1-1-1994	370	Eindhoven	1-1-1985
265	Soesterberg	1-1-1975	286	Nieuw Beerta	1-1-1994	375	Volkel	1-1-1975
267	Stavoren	1-1-1994	290	Twenthe	1-1-1975	377	Ell	1-1-2000
269	Lelystad	1-1-1994	310	Vlissingen	1-1-1957	380	Maastricht	1-1-1958
270	Leeuwarden	1-1-1975	319	Westdorpe	1-1-1994	391	Arcen	1-1-1994

605

606

607 **Table 2.** Temporal and seasonal variability of rank correlation between rain amount and duration at weather
608 stations in the Netherlands with more than 38 years of record, split into two equal periods. The data are averages
609 from annual correlations. * indicates that the difference in correlation is significant at $p=0.05$ between the two
610 time period.

No.	Name	Correlation between X_A and $X_D - 1^{st}$ period					Correlation between X_A and $X_D - 2^{nd}$ period				
		Year	DJF	MAM	JJA	SOM	Year	DJF	MAM	JJA	SOM
210	Valkenburg	0.711	0.768	0.694	0.695	0.712	0.639*	0.697*	0.654	0.600*	0.636*
235	De Kooy	0.692	0.697	0.730	0.715	0.675	0.641*	0.683	0.663*	0.633*	0.623*
240	Schiphol	0.687	0.744	0.717	0.652	0.666	0.625*	0.663*	0.665*	0.595*	0.619*
260	De Bilt	0.703	0.743	0.705	0.682	0.699	0.638*	0.688*	0.643*	0.602*	0.649*
265	Soesterberg	0.737	0.751	0.736	0.711	0.762	0.663*	0.718	0.648*	0.610*	0.675*
270	Leeuwarden	0.590	0.580	0.638	0.574	0.603	0.645*	0.682*	0.673	0.604	0.646*
280	Eelde	0.693	0.724	0.718	0.662	0.696	0.614*	0.642*	0.633*	0.596*	0.621*
290	Twenthe	0.582	0.626	0.599	0.541	0.600	0.635*	0.709*	0.626	0.580	0.659*
310	Vlissingen	0.678	0.714	0.674	0.672	0.679	0.640*	0.696	0.655	0.601*	0.631*
344	Rotterdam	0.603	0.649	0.608	0.588	0.603	0.652*	0.712*	0.666*	0.596	0.650*
375	Volkel	0.585	0.621	0.592	0.564	0.602	0.636*	0.701*	0.632	0.590	0.656*
380	Maastricht	0.675	0.705	0.684	0.655	0.695	0.616*	0.665*	0.634*	0.551*	0.656

611

612

613

614 **Table 3.** Calibration, information and weights for expert judgment elicitation of rainfall.

Expert	Calibration	Information		Weights for performance-based combination (%)	
		All variables	Calibration Variables	Without DM optimization	With DM optimization
Experts					
Expert 1	$1.95 \cdot 10^{-10}$	1.12	1.26	<0.01	-
Expert 2	$1.25 \cdot 10^{-7}$	1.21	1.44	0.02	-
Expert 3	$2.24 \cdot 10^{-4}$	0.96	1.08	28.49	30.16
Expert 4	$1.81 \cdot 10^{-8}$	0.93	1.01	<0.01	-
Expert 5	$7.58 \cdot 10^{-8}$	1.32	1.47	0.01	-
Expert 6	$5.73 \cdot 10^{-12}$	0.85	1.05	<0.01	-
Expert 7	$8.47 \cdot 10^{-4}$	0.60	0.66	65.98	69.84
Expert 8	$4.46 \cdot 10^{-5}$	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

615

616

617

618 **Table 4.** Results of calibration based on rank correlation, Gaussian ($1 - H_G$) and Hellinger ($1 - H$) distance.

Category	Rotterdam	De Bilt	Rotterdam	De Bilt
	$V > 0.95$	$V > 0.95$	$V < 0.5$	$V < 0.5$
$1 - H_G$				
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

619

620

621 **Table 5.** Projected change in precipitation according to expert judgments (performance-based DM with
 622 optimization), 50th percentile.

Variable	Projected change (2015–2023 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*	+0.2%	+2.2%
$P(U>0.95/V>0.95)$ in Rotterdam	0.0%	+9.1%
$P(U>0.95/V>0.95)$ in De Bilt	0.0%	+10.4%
$P(U>0.95/V<0.50)$ in Rotterdam/De Bilt*	+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%

624 * The same values of rainfall given by the experts for both locations.

625

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 <i>(Multiplying length of entrances by width of tunnel)</i>	26265	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