DEPARTMENT OF MATHEMATICAL SCIENCES

RISK ESTIMATION FOR PERCEPTION FAILURES IN AUTOMATED DRIVING

Love Nilsson

Degree project for Master of Science with a major in Mathematical Statistics
2021, 120 HEC
Second Cycle

Abstract

The failure of sensors to perceive the environment correctly is one of the primary sources of risk that needs to be quantified in the development of active safety features for autonomous vehicles. By extracting training data from the CARLA simulator, an object detector was trained to simulate a perception system of an autonomous vehicle. Using the detection model and gathering data for incorrect detections, various extreme value models were created and compared to investigate if extreme value theory is a viable option for estimating the risk of sensor failures of the perception system. An analysis of the extreme value's dependency on the velocity of the vehicle is performed and a risk measure is presented.

Acknowledgements

I would like to thank my supervisors Holger Rootzén and Andrew Backhouse for all the helpful discussions we had. I would also like to thank everyone at Zenseact for being very supportive in the process. Finally, I would like to thank my family and friends for their unconditional help and support.

Contents

1 Introduction

Quantifying risk is a principal part of the creation and analysis of various systems in our society. In particular, the industry of autonomous vehicles has produced new areas of research where the modeling of risk is especially important.

Around 1.4 million people die in traffic incidents each year, [1] which places this cause of death in the top ten. However, from another perspective, human drivers are safe, since in Sweden a traffic-related death occurs only 3.8 times per billion vehicle-km [2]. For the autonomous vehicle industry, it is a requirement to ensure that their vehicles are at least as safe as an average human driver, which is a challenging task.

A key component in the safety system of autonomous vehicles is the perception system, which must fulfill very strict safety requirements. The errors of a perception system can mainly be divided into two groups. On the one hand, objects that should be detected may be missed by the system. This could, for example, include objects that are uncommon, such as certain specific agricultural vehicles. On the other hand, the system might incorrectly report an object that does not exist. Two unrelated light sources at night might incorrectly be interpreted as the headlights of a car, or a tree might be identified as a pedestrian. Incorrectly detected objects are a frequent occurrence, however, most incorrectly detected objects appear far away or for a very short period of time. An essential part of the design of a safe perception system is estimating the risk of incorrect detections that could lead to dangerous situations. Mathematically speaking this amounts to estimating the probability of random events that are very rare, and might not ever have been observed.

Extreme value theory is the branch of mathematical statistics that is concerned with estimating such rare random events. If the quantity of interest can be represented by a one-dimensional value, there are mainly two different ways of studying the tail of its probability distribution. One can model the maxima M_n of an i.i.d sample $(X_1, ..., X_n)$, or the values above certain thresholds, the so-called peaks over thresholds. In the limit $n \to \infty$, the normalized M_n follows a generalized extreme value distribution $G_\xi(x)$. This family of probability distributions has three members that each are represented by various values of the parameter ξ, which characterizes the tail of the distribution. The distribution of the values above a threshold is the generalized Pareto distribution, which is closely related to the generalized extreme value distribution.

For certain applications, the risk of interest depends on more than one random quantity and thus requires the estimation of multivariate extreme value distributions, for which a parametric representation is impossible in the general case. To combat this problem one can either resort to various parametric subclasses or create semi-parametric models using the theory of copulas.

Copulas are distribution functions that describe dependence structures and provide a link between multivariate distributions and their marginals. Due to Sklar's theorem, given continuous marginals there exists a unique copula C that allows for a representation of a multivariate probability distribution in terms of

the copula and the marginals. For a bivariate distribution F, with marginals F_1, F_2 this representation reads

$$F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

The copula of an extreme value distribution is completely characterized by its dependence structure, which can be represented in various ways, for example using Pickands dependence function. This allows for the possibility of estimating the multivariate extreme value distribution, from which risk estimates can be inferred.

In this thesis, the possibility of applying these methods to estimate the risk of sensor failures of the perception system is explored. Using sensor data, we investigate modeling this risk using various random descriptors, such as velocity, the cumulative distance traveled while detecting an object, and the remaining distance to the object.

To obtain sensor data, the driving simulator CARLA was used. CARLA has been developed to support the development, training, and validation of autonomous driving systems. CARLA consists of two parts, a graphical simulation and an API where the user controls the environment by writing scripts in the python programming language. The weather conditions and environment can freely be varied to simulate real-life events for research. It is also possible to place sensors on the vehicle, from which data can be gathered. Using such sensors in CARLA, a data set was collected and an object detector was trained on this data set, resulting in a detection system from which sensor data could be collected and analyzed.

The thesis begins with a description of how the sensor data was gathered Following that, the necessary mathematical theory is presented. This includes extreme value theory and copulas, but also dependence and risk measures that are to be used. The procedure of estimating the bivariate extreme value distributions is then described and the risk estimates are presented. In the section titled model comparison, an attempt is made to compare the various model's ability to accurately predict extreme values when only being exposed to a small amount of data. This procedure involves splitting the data set and using a small part of it to fit the various models. The accuracy is quantified by using a metric to compare how close the so-called quantile curves are. The thesis is concluded with an analysis of how the velocity of the vehicle influence the estimated extreme values and a risk measure based on the braking distance is presented.

2 Sensor data

The scope of this thesis is to examine if extreme value theory is a viable option for measuring the risk of sensor failures of the perception system. This raises the question of what data to use. In a real-life application, the data would be collected from real-life situations and it would be manually annotated, which is a task that requires resources. Since the purpose is to explore if extreme value theory can be applied in this context, it will suffice to use simulated data. This is also convenient as it is easier to obtain a large data set.

With this in mind, the CARLA simulator is a suitable option. CARLA is an open-source autonomous vehicle research simulator. It provides flexible data generation and the sensor data obtained using CARLA is fairly realistic. Using a simulated environment also gives access to ground truth information about objects in the simulation. This allows for benefits such as automatic annotation of the training data for an object detector or unique identifiers for distinguishing objects. In addition, there is less variation in the obtained data, which allows for the concepts of extreme value theory to be applied on a smaller scale to start with. For information on the CARLA simulator, see [3].

CARLA offers a wide variety of sensors, such as RGB, LIDAR, and semantic segmentation, from which data can be retrieved each frame. Using RGB sensors, training data was gathered by running several simulations and retrieving images at constant intervals. Automatic annotation of the images was done using bounding boxes, which are available for most objects in CARLA, and a semantic segmentation sensor to distinguish objects.

The gathered training data consisted of around 200 1024x1024 RGB images and the classes it was trained to detect were vehicles, motorcycles, bicycles, and traffic lights. The reason for the small size of the training data set is that in this particular case, the detection model should not work perfectly, and thus generate more data for sensor failures.

To achieve a somewhat realistic detection model, an object detector was trained using the Tensorflow object detection API developed by Google. The training process of the neural net starts from a checkpoint of a pre-trained neural net. The neural net used for the object detector was the Faster R-CNN Inception ResNet V2 1024x1024, which is a region-based convolutional neural network (R-CNN). It has a speed of 236 ms and a COCO mAP of 38.7.

Figure 1: CARLA simulator

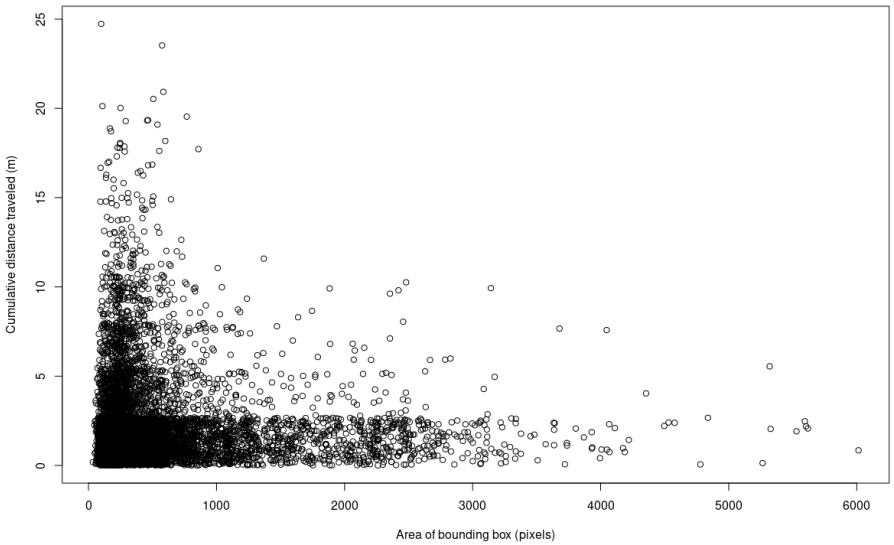
Using the trained object detector, new simulations were made and the object detector was applied to each obtained RGB image. Around 200 simulations were run, totaling 100 km of driving and an image data set of around 180 GB. Figure 1 depicts a frame of the CARLA simulator where the object detection model has been applied.

Each detection from the model is in its rawest form an object label and a bounding box. As mentioned before, the CARLA simulator provides unique identifiers for each object which allows for the identification of objects between different frames. Thus, incorrectly detected objects and objects that were missed by the detection system can be identified and data for such detections generated. Incorrect detections largely outweighed the number of missed detections and so the models presented in this thesis are for incorrect detections.

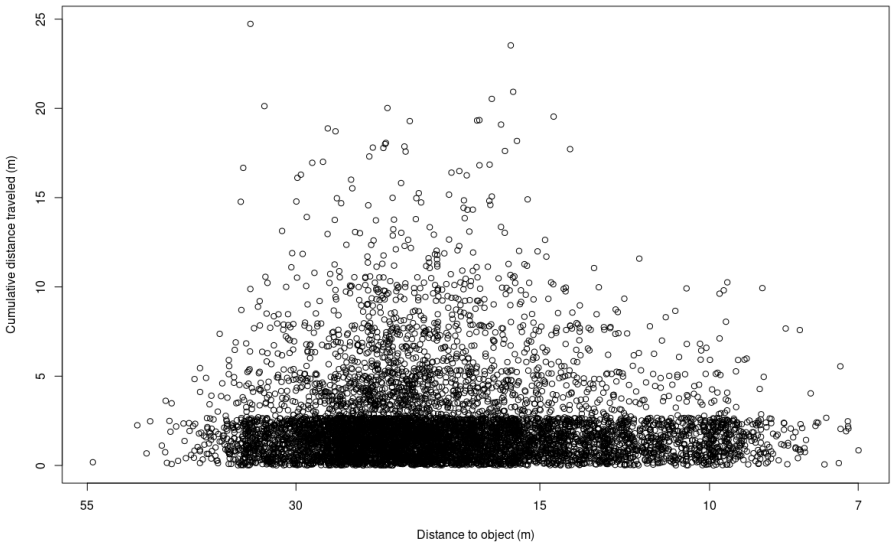
For each incorrect detection the cumulative distance traveled between the consecutive frames the object was detected, and the area of the bounding box is observed. The values are recorded directly when the object stops being detected. Mathematically it is an independent sample of the random vector consisting of two random variables; the area of the bounding box (measured in pixels), and the cumulative distance traveled while incorrectly detecting the object (measured in meters). Estimating the risk of sensor failures then corresponds to estimating the probability that these two random variables take on large values simultaneously.

The idea is that the area of the bounding box will serve as a measurement of how close the object is. A large area corresponds to a small distance to the object and vice versa. The reason for using the area of the bounding box as a distance measure is that is easy to calculate and does not require measuring the distance using LIDAR or any similar technology. The correspondence between the area of the bounding box and the distance to the object was obtained using a regression analysis for data of distances and areas. Thus, the analysis can be carried out using the area of the bounding box, and then any results in terms of bounding box areas can be translated into distances in meters.

Measurements of the area of the bounding box are dependent on what type of object is being detected. Therefore, mathematical models for the various objects need to be separated, but the mathematical treatment is the same. The object detector generated very few incorrect detections for vehicles, motorbikes, and bicycles. Most incorrect detections were labeled as traffic lights. Over the 200 simulations, the detection model produced 7654 incorrect detections traffic light detections. Because of this surplus of data, the models created were made using data for incorrect traffic light detections. The collected data is presented in figure 2.



(a)



(b)

Figure 2: (a) Scatterplot of bounding box area and distance traveled until the incorrect detection disappears. (b) Scatterplot of the distance to incorrect detection and the distance traveled until it disappears. The x-axis is log transformed and the scale is reversed due to the relationship between area and distance.

In figure 3 is a depiction of the detection model's incapacity to distinguish black pixels in the upper part of the image from what is actually a traffic light. This case is extreme in the sense that the area of the bounding box and the distance traveled until the detection disappears are both large simultaneously.

Figure 3: A traffic light incorrectly detected during several consecutive frames

Most incorrect detections are only observed in one frame, and not the next, resulting in the distance traveled until it disappears being 0 meters. In actuality, the distance could be any value between 0 and the distance the vehicle moved between the frames. The discretization created by measuring frame to frame is necessary to keep the amount of data to reasonable levels but is problematic when trying to fit distributions to the data. A solution to this problem was to give the values a uniformly distributed value within the discretization.

3 Theory

In this section the necessary mathematical theory is presented. Univariate extreme value theory is presented briefly and then the extension to the multivariate case is discussed. Most of the multivariate theory is described without using copulas. The theory of copulas and how the theory translates using them is introduced afterwards. The theory section is concluded with simulation methods for copulas as well as methods for estimating dependence structures and quantile curves. The main reference for the sections on extreme value theory is [4], and the main reference for the section on copulas is [7].

3.1 Extreme value theory

Extreme value theory is concerned with describing the probability distribution of 'extreme' observations. What is meant by extreme is ambiguous, but usually the distribution of $\max(X_1, X_2, ..., X_n)$ or the set of observations that are larger than some threshold is of interest. These notions of what is meant by extreme observations are cornerstones in the two leading methods, referred to as block maxima and peaks over threshold (PoT). Both methods will be reviewed briefly.

3.1.1 Block maxima

In the block maxima methods, the data is divided into blocks of certain length and for each block, the maxima is extracted. In the uni-variate case where $\{X_i\}_{i=1}^n$ is an i.i.d sample with common distribution function F, the distribution of $M_n = \max(X_1, X_2, ..., X_n)$ can be obtained as

$$P(M_n \leq x) = \prod_i^n P(X_i \leq x) = (F(x))^n$$

In practice, however, one might not have access to an analytical expression for F and thus one relies on asymptotical results in the limit $n \to \infty$ to estimate the distribution of M_n. Such a result is the Fisher–Tippett–Gnedenko theorem.

Theorem 1 (Fisher–Tippett–Gnedenko)**.** If there exists a sequences $\{a_n\}$ and $\{b_n\}$ with $a_n > 0$ and $b_n \in \mathbb{R}$ for all n such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G_\xi(x), \quad n \to \infty,$$

where G_ξ is a non-degenerate distribution, then $G_\xi(x)$ is necessarily one of the three possible extreme value distributions.

The possible limiting distributions are referred to as generalized extreme value distributions. They are parametrized by location, scale and shape parameters μ, σ, ξ, respectively. It is convenient to simply write x instead of $\frac{y-\mu}{\sigma}$. The normalizing constants $\{a_n\}$ and $\{b_n\}$ are needed to ensure that this convergence

in distribution is not to that of a degenerate distribution with all of its mass on the point $\inf\{x : F(x) = 1\}$.

It is a result that all $G_\xi, \xi \in \mathbb{R}$ can occur in this limit [4]. The parameter ξ is the extreme value index and is central to extreme value analysis because it is related to the tail of the distribution. There are three cases; $\xi > 0$ gives the Fréchet distribution

$$F(x) = \begin{cases} \exp[-(1+\xi x)^{\frac{1}{\xi}}] & 1+\xi x > 0 \\ 0 & 1+\xi x \leq 0 \end{cases}$$

$\xi < 0$ gives the Weibull distribution

$$F(x) = \begin{cases} \exp[-(1+\xi x)^{\frac{-1}{\xi}}] & 1+\xi x > 0 \\ 1 & 1+\xi x \leq 0 \end{cases}$$

and $\xi = 0$ gives the Gumbel distribution

$$F(x) = e^{-e^{-x}}$$

Apart from determining the possible limits G_ξ, the question that remains is under what conditions on the distribution of the X_i one precisely obtains a specific distribution G_ξ in the limit. The set of distributions F such that this occurs is denoted by $\mathcal{D}(G_\xi)$ and is referred to as the domain of attraction. One writes $F \in \mathcal{D}(G_\xi)$. Under the assumption that F is continuous, de Haan gave necessary and sufficient conditions for F to be in the domain of attraction of G_ξ for general ξ in 1970. In terms of F this condition is

$$\frac{1 - F(y + b(y)v)}{1 - F(y)} \to (1+\xi v)^{\frac{-1}{\xi}} \text{ as } y \to \inf\{x : F(x) = 1\}$$

where the function b is the so-called auxiliary function and $(1 + \xi v) > 0$. The full treatment of these conditions are given in [4].

3.1.2 Peaks over threshold

The peaks over threshold model, introduced by Pickands, instead makes use of all observations that exceed a certain threshold. This generally produces more data than the block maxima method. For a one dimensional random variable X, it is based on the fact that

$$P(X > t + x | X > t) = \frac{P(X > t + x)}{P(X > t)} \sim F(x; \mu, \sigma, \xi)$$

for $x > 0$, and sufficiently large threshold t. $F(x; \mu, \sigma, \xi)$ is the generalized pareto distribution, parametrized by the location, scale and shape parameters $\mu \in \mathbb{R}, \sigma > 0, \xi \in \mathbb{R}$, respectively. With $z = \frac{x-\mu}{\sigma}$ the distribution function is

$$F(x) = \begin{cases} 1 - (1 + \frac{\xi(x-\mu)}{\sigma})^{\frac{-1}{\xi}} & \xi \neq 0 \\ 1 - \exp(-\frac{(x-\mu)}{\sigma}) & \xi = 0 \end{cases}$$

where the support is $x \geq 0$ for $\xi \geq 0$ and $\mu \leq x \leq \mu - \frac{\sigma}{\xi}$ for $\xi < 0$.

3.2 Multivariate extreme value theory

In multivariate extreme value theory one faces the problem of not having a natural ordering. There is no longer an unambiguous way to describe what is meant by a large or small value. Furthermore one must now take the dependence structure of the random variables into account. In particular, it is crucial to capture the possible dependence in the extremes. Thus, the focus shifts to the marginal distributions and the structure of their dependence.

The extension of the block maxima and PoT methods to the multivariate case is rather natural. A useful and natural way to think about ordering is component-wise ordering, i.e for two vectors x and y one writes $x \leq y$ iff $x_i \leq y_i$ for all i and one can similarily define the component-wise maxima of x, y as the vector with components $\max(x_i, y_i)$. In this section inequalities for vectors are to be interpreted in this way. For a multivariate random sample $\{X_i\}_{i=1}^m$ where each $X_i \in \mathbb{R}^n$, one can then analogously to the univariate case define the sample maximum M_m as the vector with components $\max_{1 \leq i \leq m} X_{i,j}$, for $j = 1, ..., n$. An immediate observation is that this maxima need not be in the sample. However, this definition can still lead to useful results. For the PoT model, individual thresholds can be defined for each marginal and one can consider points that are above some or all of the thresholds.

Similarly to the univariate case one can now formulate the corresponding statement regarding the convergence of M_m to an extreme value distribution. For an i.i.d sample with distribution function F it is stated as finding a sequence of vectors $a_n > 0$ and b_n such that $F^m(a_m x + b_m)$ converges in distribution to a multivariate extreme value distribution $G(x)$ as $m \to \infty$. Again the set of all distribution functions for which the above holds make up the domain of attraction for G, $\mathcal{D}(G)$. It can be noted that the above convergence in distribution imply the convergence in distribtuion for each of the marginals, which means that each marginal distribution converges to an extreme value distribution as described in the univariate case.

A fundamental difference is that a general multivariate extreme value distribution $G(x)$ no longer can be indexed using a vector of finite dimension, which poses a problem in the characterization of $G(x)$. An informal reason for this is that the set of dependence structures is simply too large. There are several remedies to the problem of characterizing dependence structures, each resulting in a different way of expressing the dependence structure.

A popular approach which will be used here is to index $G(x)$ using certain convex functions.To motivate this indexation we will require the concept of max-stability. It is an important concept as it relates to the domain of attraction of an extreme value distribution.

Definition 1. A multivariate distribtuion $G(x)$ is called max-stable if for all positive integers k, there exists vectors $\alpha_k > 0$ and β_k such that

$$G^k(\alpha_k x + \beta_k) = G(x)$$

The above equality means that the distribution of the component-wise max-

ima is the same for all sample sizes.

The classes of max-stable and extreme value distributions are the same. Indeed, by combining the definition of max stability with the definition of the domain of attraction mentioned earlier, it is easy to see that a max-stable distribution is in its own domain of attraction and so is an extreme value distribution. In other words, all max-stable distributions are in $\mathcal{D}(G)$.

Now, to show the reverse inclusion, assume $F \in \mathcal{D}(G)$ and partition the sample size m into n blocks of size k, i.e write $m = nk$. Since $F \in \mathcal{D}(G)$, we then have

$$\lim_{n\to\infty} F^m(a_m x + b_m) = G(x)$$

on the other hand, for any positive integer k

$$\lim_{m\to\infty} F^{mk}(a_{mk}x + b_{mk}) = [\lim_{m\to\infty} F^m(a_{mk}x + b_{mk})]^k = G^k(x)$$

By the convergence of types theorem there then exists $\alpha_k > 0, \beta_k$ such that $\frac{a_{mk}}{a_m} \to \alpha_k$, $\frac{(b_{mk}-b_m)}{a_m} \to \beta_k$ and $G^k(\alpha_k x + \beta_k) = G(x)$ so that G is max-stable.

If a distribution is max-stable, then $G^{\frac{1}{k}}$ is a distribution function for all k, and the exponent measure μ, introduced by Balkema and Resnick in 1977, defined on on $[-\infty, \infty)$, allows for a representation of G as $G(x) = \exp[-\mu([-\infty, \infty) \setminus [-\infty, x])]$.

For max-stable (or extreme value) distributions the the stable tail dependence function can be defined.

Definition 2 (Huang 1992)**.** Let G be a max-stable distribution and let $G_j^{-1}(u) = \inf\{x : G_j(x) \geq u\}$ denote the generalized inverse of the marginal G_j. Then the stable tail dependence function, for $v \in \mathbb{R}^n$, is

$$\ell(v) = -\log G(G_1^{-1}(e^{-v_1}), G_2^{-1}(e^{-v_2}), ..., G_n^{-1}(e^{-v_n}))$$

Depending on the assumptions of the marginal distributions the stable tail dependence functions has various properties. Some of them include homogeneity, convexity and certain boundary conditions.

One can obtain the original max-stable distribution G through its marginal distributions and the stable tail dependence function, by writing

$$-\log G(x) = \ell(-\log G_1(x_1), ..., -\log G_n(x_n))$$

for $x \in \mathbb{R}^n$ one obtains

$$G(x) = \exp(-\ell(-\log G_1(x_1), ..., -\log G_n(x_n)) \tag{1}$$

showing how multivariate extreme value distributions can be indexed using the stable tail dependence function.

For the practical application considered in this thesis, a special case of the stable tail dependence function will be of greatest interest. It was introduced by Pickands and has been popularized due to it being a helpful tool in the estimation of bivariate extreme value distributions. As mentioned above, ℓ satisfies a homogeneity condition. It is thus possible to restrict ℓ to the unit simplex, i.e the set of points in $[0, 1]^n$ whith components summing to 1.

Definition 3 (Pickands 1981)**.** Let G be a max stable distribution of dimension $n+1$, with stable tail dependence function ℓ. Let $\triangle^n = \{(t_1, t_2, ..., t_{n+1}) \in [0,1]^{n+1} : \sum_{i=1}^{n+1} t_i = 1\} \subset \mathbb{R}^{n+1}$ be the unit simplex. Then Pickands dependence function, $A : \triangle^n \mapsto [\frac{1}{n}, 1]$ is defined as

$$A(z_1, ..., z_{n+1}) = \frac{\ell(x_1, x_2, ..., x_{n+1})}{\sum_{i=1}^{n+1} x_i}, \quad z_i = \frac{x_i}{\sum_{i=1}^{n+1} x_i}$$

Since A is defined through the stable tail dependence function, it inherits its properties. In particular, Pickands dependence function is necessarily convex and and satisfies $\max(z_1, ..., z_{n+1}) \leq A(z_1, ..., z_{n+1}) \leq 1$.

For $t \in \triangle^n$ it is possible to express one of the components of t through the other components, since they sum to 1. Thus, A can be written as a function of n variables. To see this, let G be a $n+1 = 2$ dimensional max-stable distribution. By equation (1) we have

$$G(x) = \exp(-\ell(-\log G_1(x_1), -\log G_2(x_2))$$

for $x = (x_1, x_2) \in \mathbb{R}^2$. Then, by definition of A,

$$\begin{aligned} G(x) &= \exp(-\ell(-\log G_1(x_1), -\log G_2(x_2)) \\ &= \exp\left[-\sum_{i=1}^{2} -\log G_i(x_i) A\left(\frac{-\log G_1(x_1)}{\sum_{i=1}^{2} -\log G_i(x_i)}, \frac{-\log G_2(x_2)}{\sum_{i=1}^{2} -\log G_i(x_i)} \right) \right] \\ &= \exp\left[\log(G_1(x_1)G_2(x_2)) A\left(\frac{\log G_1(x_1)}{\log(G_1(x_1)G_2(x_2))}, \frac{\log G_2(x_2)}{\log(G_1(x_1)G_2(x_2))} \right) \right] \end{aligned}$$

In this case the unit simplex is simply the line from $(1,0)$ to $(0,1)$ which can be parametrized as $(1-t, t), t \in [0,1]$. Then, if we let $t = \frac{\log G_2(x_2)}{\log G_1(x_1)G_2(x_2)}$ we can write A as a function of a single variable,

$$A(t) = \ell(1-t, t) = \ell(\frac{\log G_2(x_2)}{\log G_1(x_1)G_2(x_2)}, \frac{\log G_2(x_2)}{\log G_1(x_1)G_2(x_2)})$$

and get that

$$G(x) = \exp(\log(G_1(x_1)G_2(x_2))A(t)) = (G_1(x_1)G_2(x_2))^{A(t)} \tag{2}$$

Thus, one is able to reduce the inference of bivariate extreme value distributions to that of estimating a function of a single variable.

3.2.1 Parametric subclasses of bivariate extreme value distributions

Although, the class of multivariate extreme value distributions is not parametrizable, there are, however, parametric subclasses. Some examples of such distributions that are used in the modeling will be discussed here briefly.

For ease of notation, let $y_i = y_i(x_i) = (1 + \xi_i(x_i - \mu_i)/\sigma_i))^{\frac{-1}{\xi_i}}$, where $1 + \xi_i(x_i - \mu_i)/\sigma_i) > 0$. Introduced by Joe (1990), the asymmetric negative

logistic distribution is parametrized by the dependence parameter $r > 0$ and the asymmetry parameters $0 < t = (t_1, t_2) \leq 1$. The distribution function is

$$G(x_1, x_2) = \exp[-y_1 - y_2 + ((t_1 y_1)^{-r} + (t_2 y_2)^{-r})]^{\frac{-1}{r}}$$

Independence corresponds to either parameter approaching 0, complete independence is obtained for fixed $t_1 = t_2 = 1$, with $r \to \infty$.

A special case of the bivariate negative logistic distribution that will also be used is due to Galambos (1975). It is the special case when $t_1 = t_2 = 1$ and thus have distribution function

$$G(x_1, x_2) = \exp[-y_1 - y_2 + (y_1^{-r} + y_2^{-r})^{\frac{-1}{r}}],\ r > 0$$

with independence and complete dependence for the limits $r \to \infty$ and $r \to 0$, respectively.

In both cases, the marginal distributions are generalized extreme value distributions.

3.3 Copulas

In essence, copulas are distribution functions which contain all of the dependence structure between components of a random vector. There are more technical definitions available, see [7], but here the following simple definition will be used.

Definition 4 (Copula)**.** A copula C is any multivariate distribution function with uniform marginal distributions

Example 1. For any multivariate distribution function F with continuous marginal distributions one can obtain a copula by letting $U_i = F_i(X_i)$, where $F_i(X_i)$ is the i'th marginal distribution. Then, since each U_i is uniformly distributed, we have that

$$C(u_1, u_2, ..., u_n) = P(U_1 \leq u_1, U_2 \leq u_2, ..., U_n \leq u_n)$$

is a copula and F can be written as,

$$\begin{aligned} F(x_1, x_2, ..., x_n) &= P(X_1 \leq x_1, X_2 \leq x_2, ..., X_n \leq x_n) \\ &= P(U_1 \leq F_1(x_1), U_2 \leq F_1(x_2), ..., U_n \leq F_1(x_n)) \\ &= C(F_1(x_1), ..., F_n(x_n)) \end{aligned}$$

It is common to assume continuity of the marginals and write the copula C as a function of $u_i = F_i(x_i)$. From now on, this will be the case unless otherwise stated.

A copula that is very useful in practice is the empirical copula. If the marginal distributions are unknown, which usually is the case, the empirical copula provides a non-parametric estimation by using the empirical distribution functions $\hat{F}_i(x_i) = \hat{U}_i$.

Example 2 (Empirical copula)**.**

$$\hat{C}(u_1, u_2, ..., u_n) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbb{I}_{\{\hat{U}_1 \leq u_1, \hat{U}_2 \leq u_2, ..., \hat{U}_n \leq u_n\}}$$

As shown in in example 1, given a distribution F with marginals F_i, $i = 1, ..., n$, it is possible to represent it through the copula C and the marginals. That this copula is unique, given that the marginal distributions are continuous, is the cornerstone of the theory of copulas. It was first proved by Abel Sklar.

Theorem 2 (Sklar 1958)**.** Given a multivariate distribution F with continuous marginal distribution functions, F_i, $i = 1, ..., n$ there exists a unique copula C such that

$$F(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n))$$

The use of copulas is to describe the dependence structure of random variables. An extreme case is independence.

Example 3 (Independence copula)**.**

$$C(u_1, u_2, ..., u_n) = \prod_{i=1}^{n} u_i$$

The other extreme cases of dependence structures are perfect positive or negative correlation. In the two-dimensional case, consider positive dependence, for example $X_2 = 2X_1$. Then,

$$F_1(x) = P(X_1 \leq x) = P(2X_1 \leq 2x) = P(X_2 \leq 2x) = F_2(2x),$$

and we have

$$U_1 = F_1(X_1) = F_2(2X_1) = F_2(X_2) = U_2$$

so the copula becomes

$$\begin{aligned} C(u_1, u_2) &= P(U_1 \leq u_1, U_2 \leq u_2) \\ &= P(U_1 \leq u_1, U_1 \leq u_2) \\ &= P(U_1 \leq \min\{u_1, u_2\}) \\ &= \min\{u_1, u_2\}. \end{aligned}$$

The calculation for perfect negative dependence, for example $X_2 = -2X_1$ gives $U_1 = 1 - U_2$ and thus

$$\begin{aligned} C(u_1, u_2) &= P(U_1 \leq u_1, U_2 \leq u_2) \\ &= P(U_1 \leq u_1, 1 - U_1 \leq u_2) \\ &= P(1 - u_2 \leq U_1 \leq u_1) \\ &= \max\{u_1 + u_2 - 1, 0\}. \end{aligned}$$

These extreme cases provide upper and lower bounds for any copula, a result due to Fréchet and Hoeffding.

Theorem 3 (Fréchet-Hoeffding bounds)**.** For any n-dimensional copula C, the following inequalites hold:

$$\max\{u_1 + u_2 + ... + u_n + 1 - n, 0\} \leq C(u_1, u_2, ..., u_n) \leq \min\{u_1, u_2, ..., u_n\}$$

Now the theory of multivariate extreme value distributions will be connected to the theory copulas. In the previous section, multivariate extreme value distributions were described as the limit distribution of the sample maximum M_m of an i.i.d multivariate random sample $\{X_i\}_{i=1}^m$ where each $X_i \in \mathbb{R}^n$. The theory translates to copulas in the following way, where we restrict ourselves to the bivariate case for ease of notation.

We will begin by showing that the copula of the component-wise maxima of i.i.d pairs of random variables $(X_1, Y_1), ..., (X_m, Y_m)$ is $C_m(u_1, u_2) = C^m(u_1^{\frac{1}{m}}, u_2^{\frac{1}{m}})$.Too see this, Let their common distribution function be H, their copula be C. Let F and G be the marginals of X_i and Y_i, respectively. Finally, let m be a fixed positive integer.

Now, the distribution functions $F_{(m)}$ and $G_{(m)}$ of $X_{(m)} = \max_{i=1,...,m} X_i$ and $Y_{(m)} = \max_{i=1,...,m} Y_i$, respectively are given as $F_{(m)}(x) = [F(x)]^m$ and $G_{(m)}(y) = [F(y)]^m$ due to independence. Therefore, the joint distribution function $H_{(m)}$ of the component-wise maxima $(X_{(m)}, Y_{(m)})$ is

Now, we have

$$\begin{aligned} H_{(m)}(x, y) &= [H(x, y)]^m \\ &= [C(F(x), G(y))]^m \\ &= \left[C\left([F_{(m)}(x)]^{\frac{1}{m}}, [G_{(m)}(y)]^{\frac{1}{m}}\right)\right]^m \end{aligned}$$

As before, the question now is whether this converges in distribution as $m \to \infty$. If it does, then the copula is an extreme value copula and C_F is in the domain of attraction $\mathcal{D}(C)$.

Definition 5 (Extreme value copula)**.** A copula C_G is the copula of an extreme value distribution, or simply an extreme value copula if there exists a copula C_F such that

$$C_G(u_1, u_2) = \lim_{m \to \infty} C_F^m(u_1^{1/m}, u_2^{1/m})$$

The concept of max-stability is stated as

Definition 6. The copula C is max-stable if for all positive integers m, we have

$$C(u_1, u_2) = C^m(u_1^{1/m}, u_2^{1/m})$$

As before, the conditions that a copula is an extreme value copula and that it is max-stable are equivalent.

As in equation (2), a bivariate extreme value copula C can be represented using Pickand's dependence function, A, as

$$C(u_1, u_2) = \exp(\log(u_1 u_2) A(t) = (u_1 u_2)^{A(t)} \tag{3}$$

where $t = \frac{\log(u_2)}{\log(u_1 u_2)}$.

3.4 Simulating from a copula

When estimating an unknown distribution using a copula, generating random variates from this distribution might of interest. There are a number of different methods to simulate from the copula. A common method where the approach is to use conditional distributions, [7]. To simulate from the estimated distribution one first generates two uniformly distributed numbers u_1, r on $(0, 1)$. Since the copula has uniform marginals, we have

$$P(U_2 \leq u_2 \mid U_1 = u_1) = \frac{\partial C(u_1, u_2)}{\partial u_1}$$

Denoting this function of u_2 as $C_{u_1}(u_2)$ one then sets $v = C_{u_1}^{-1}(r)$. Finally mapping u, v to x, y using the inverses of the marginals produces random variates with the dependence structure of the copula.

The function $C_{u_1}^{-1}(u_2)$ exists and is non decreasing almost everywhere, [7]. If C is an extreme value copula given as in equation (3), the partial derivative has the closed form expression

$$\frac{\partial C(u_1, u_2)}{\partial u_1} = \frac{C(u_1, u_2)}{u_1}(A(t) - tA'(t))$$

where $t = \frac{\log(u_2)}{\log(u_1 u_2)}$.

3.5 Dependence measures

3.5.1 Tail dependence

In the study of extreme values and estimation of risk, an important concept is that of tail-dependence, [15], [11], it is a way of measuring dependence in the tail of a multivariate distribution. It is possible for random variables with seemingly no correlation to exhibit tail dependence, the common example being stock returns. When studying tail dependence, the so-called survival function, $S(x_1, ..., x_n) = P(X_1 > x_1, X_2 > x_2, ..., X_n > x_n)$ is a natural object to consider.

The copula of the survival function is called the survival copula, or sometimes the reflected copula. In the two dimensional case it is defined as $\bar{C}(u_1, u_2) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2)$. To motivate this definition, consider the joint survival function of (X_1, X_2) with copula C and marginals F_1, F_2. Let the univariate survival functions of F_1, F_2 be S_1, S_2. Then we have

$$\begin{aligned} S(x_1, x_2) &= 1 - P(X_1 \leq x_1 \text{ or } X_2 \leq x_2) \\ &= 1 - F_1(x_1) - F_2(x_2) + F(x_1, x_2) \\ &= S_1(x_1) + S_2(x_2) - 1 + C(F_1(x_1), F_2(x_2)) \\ &= S_1(x_1) + S_2(x_2) - 1 + C(1 - S_1(x_1), 1 - S_2(x_2)). \end{aligned}$$

and thus $S(x_1, x_2) = \bar{C}(S_1(x_1), S_2(x_2))$ The reason for this discussion is the central concept of tail dependence; the coefficients of upper and lower tail dependence.

Definition 7 (Coefficients of tail dependence)**.**

$$\lambda_\ell = \lim_{q\to 0^+} P(X_2 \le F_2^{-1}(q) | X_1 \le F_1^{-1}(q)),$$
$$\lambda_u = \lim_{q\to 1^-} P(X_2 > F_2^{-1}(q) | X_1 > F_1^{-1}(q)),$$

Both coefficients are in the interval $[0, 1]$. They can elegantly be described in terms of the copulas C and $\bar{C}$ as

$$\begin{aligned}\lambda_U &= \lim_{u\to 1^-} P(U_2 > u | U_1 > u) \\ &= \lim_{u\to 1^-} \frac{P(U_2 > u, U_1 > u)}{P(U_1 > u)} \\ &= \lim_{u\to 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \\ &= \lim_{u\to 1^-} \frac{\bar{C}(1 - u, 1 - u)}{1 - u}.\end{aligned}$$

and similarly

$$\lambda_L = \lim_{u\to 0^+} \frac{C(u, u)}{u}.$$

For the application considered here, the coefficient of upper tail dependence is mainly of interest. This coefficient can be represented using Pickands dependence function [16] as

$$\lambda_U = 2(1 - A(\frac{1}{2})) \tag{4}$$

If $\lambda_U > 0$, C has upper tail dependence, and if $\lambda_U = 0$, C has no upper tail dependence.

3.5.2 The function $\chi(u)$

The upper coefficient of tail dependence can also be obtained as the limit of another function, $\chi(u) = 2 - \frac{\log C(u,u)}{\log u}$, [11]. Using the above expression for the upper tail coefficient it is also possible to write

$$\begin{aligned}\chi(u) &= 2 - \frac{1 - C(u, u)}{1 - u} + o(1) \\ &= \frac{1 - 2u + C(u, u)}{1 - u} + o(1) \\ &= P(U_2 > u | U_1 > u) + o(1)\end{aligned}$$

Thus, the limit of $\chi(u)$ as $u \to 1$ is λ_U, but this function can also be thought of as a way of measuring dependence for various quantiles, where the sign of $\chi(u)$ is an indicator of negative or positive dependence. The Fréchet-Hoeffding bounds gives

$$\max(2u - 1, 0) \le C(u, u) \le u$$

for $0 < u < 1$ so therefore

$$2 - \frac{\log[\max(2u-1,0)]}{\log u} \leq \chi(u) \leq 1$$

where the lower bound is interpreted as $-\infty$ for $u \leq \frac{1}{2}$ and 0 for $u = 1$.

For the special case when C is a bivariate extreme value copula, given as in equation (3), one obtains $C(u,u) = u^{2A(\frac{1}{2})}$ and thus $\lim_{u\to 1^-} \chi(u) = \lambda_U = 2(1 - A(\frac{1}{2}))$ as before. Furthermore $\chi(u) = 2(1 - A(\frac{1}{2}))$ is constant for all u, a property that is true for any bivariate extreme value distribution, [11]. This gives a diagnostic tool for ascertaining membership of the bivariate extreme value class.

3.6 Inference on Pickands dependence function

Representing a bivariate extreme value copula through the use of Pickands dependence function allows for estimation of the copula by instead estimating $A(t)$. There exists many estimators in the litterature, [9], [8].

A classical estimator is the Pickands estimator, [8]. It is based on the fact that for (X_1, X_2) with bivariate copula C with marginals $U_i = F(X_i)$, the random variables $E_i = -\log(U_i)$ have a standard exponential distribution. One then defines

$$\xi(t) = \min(\frac{E_1}{1-t}, \frac{E_2}{t})$$

and shows that the survival function of $\xi(t)$, $P(\xi(t) > x)$, is equal to $e^{-xA(t)}$ and so is exponentially distributed with parameter $A(t)$. Thus, $\mathbb{E}(\xi(t)) = \frac{1}{A(t)}$ and the Pickands estimator is obtained through the method of moments simply as

$$\frac{1}{\hat{A}^P(t)} = \frac{1}{m}\sum_{i=1}^{m} \xi_i(t)$$

A problem is that in general, no estimator will fulfill the boundary conditions or be convex. An estimator that fulfills these boundary conditions by construction is due to Tajvidi and Hall. It is given by

$$\hat{A}^{HT}(t) = \left(\frac{1}{m}\sum_{i=1}^{m} \bar{\xi}(t)\right)^{-1}$$

where $\bar{\xi}(t) = \min(\frac{\bar{S}_i}{1-t}, \frac{\bar{T}_i}{t})$ and

$$\bar{S}_i = \frac{-n\log \hat{F}_1(X_{1,i})}{\sum_{k=1}^{m} -\log \hat{F}_1(X_{1,k})} \quad \bar{T}_i = \frac{-n\log \hat{F}_2(X_{2,i})}{\sum_{k=1}^{m} -\log \hat{F}_2(X_{2,k})}$$

To combat the non-convexity it is then convenient to consider the greatest convex minorant of $\hat{A}^{HT}(t)$. Combining these modifications provides legitimate estimates of Pickands dependence function that fulfill all conditions. Other popular estimators also exist, most notable is perhaps the Capéraà Fougères Genest (CFG) estimator [9].

3.7 Quantile curves

An interesting way to analyze and visualize bivariate extreme value distributions is through its quantile curves [4]. Let G be an estimated bivariate extreme value distribution. As was shown earlier, G can be expressed through its marginals and Pickands dependence function as

$$G(x_1, x_2) = \exp[\log(G_1(x_1)G_2(x_2))A(t)]$$

where $t = \frac{\log G_2(x_2)}{\log G_1(x_1)G_2(x_2)}$. The quantile curves of the distribution are

$$Q(p) = \{(x_1, x_2) : G(x_1, x_2) = p\}$$

for $0 < p < 1$. Now, using the above representation of G, it can be seen that $G(x_1, x_2) = p$ if and only if there exists $\alpha \in [0, 1]$ such that $G_1(x_1) = p^{\frac{1-\alpha}{A(\alpha)}}$ and $G_2(x_2) = p^{\frac{\alpha}{A(\alpha)}}$, so that the quantile curve can be written as

$$Q(p) = \left\{\left(G_1^{-1}(p^{\frac{1-\alpha}{A(\alpha)}}), G_2^{-1}(p^{\frac{\alpha}{A(\alpha)}})\right) : \alpha \in [0, 1]\right\}$$

for the generalized inverses G_1^{-1}, G_2^{-1}.

A special case is independence, where A is identically equal to 1. The endpoints of the quantile curve correspond to $w = 0$ and $w = 1$, for which the points on the curve are $(G_1^{-1}(p), G_2^{-1}(p)))$ and $(G_1^{-1}(1), G_2^{-1}(p))$. $G_i^{-1}(1)$ is approximated as a finite number in the discretization of the quantile curve. Thus, the probability of (X_1, X_2) being in the region defined by the quantile curve is approximately equal to the probability of being in the recangle defined by the vertical and horizontal lines given by the marginal quantiles, $x_1 = G_1^{-1}(p)$ and $x_2 = G_2^{-1}(p)$. This probability is in turn equal to $P(X_1 > x_1, X_2 > x_2) = (1 - p)^2$.

4 Estimation of bivariate extreme value distributions

In essence there are four different models that will be considered. The data used for modeling the tail will be chosen both by taking the component-wise maxima of blocks of the data, and by using threshold exceedances.

In each case, the extreme value distribution will be estimated using two various methods; a semi-parametric model and a parametric model. In the semi-parametric model the marginals are modeled as generalized extreme value or generalized pareto distributions and a copula is used to estimate the bivariate distribution. In the parametric models, the component-wise maxima and threshold exceedances are used to fit the parametric subclasses presented in section 3.2.1 of the bivariate extreme value distributions. The parametric models were fitted using the evd package [5] and the POT package in R [6].

For each model 10^4 random variates are generated and plotted to show the general characteristics of the distribution. The quantile curves for $p = 0.95, 0.99, 0.995$, which are generated from 10^6 simulated random variates, are also depicted. The quantile curves are used as the risk measure and furthermore are used to compare the estimated distributions in the section titled model comparison.

As shown in the section on quantile curves, for independent, or roughly independent data, the probability of an observation being in this region is approximately $(1-p)^2$ since the region determined by the quantile curves could be approximated by a rectangular region. This rectangular region is formed by the vertical and horizontal lines that intersect each endpoint of the quantile curve and will be referred to as the threshold excess regions for $p = 0.95, 0.99, 0.995$. The reason for introducing this region is because it is possible to represent it using only two numbers, one for each marginal. These numbers can be viewed as thresholds for which approximately for which the probability of an observation being larger in both components is $(1-p)^2$, i.e $0.05^2 = 2.5 \cdot 10^{-3}, 0.01^2 = 10^{-4}$ and $2.5 \cdot 10^{-5}$ of observations are larger in both components.

For each model the thresholds of the threshold excess regions are presented, with the area of the bounding box converted to distances in meters. Also presented are the empirical probabilities and the estimated number of detections per 10^5 km for each of these regions. These were estimated by simulation of 10^6 random variates.

4.1 Analysis of the dependence structure

When modeling extreme value it is important to be aware of this structure, in particular the dependence in the extremes. Therefore, before turning to estimating the extreme value distributions, a brief analysis of the dependence structure of the data will be made.

The data for the incorrect detections is presented in figure 4.

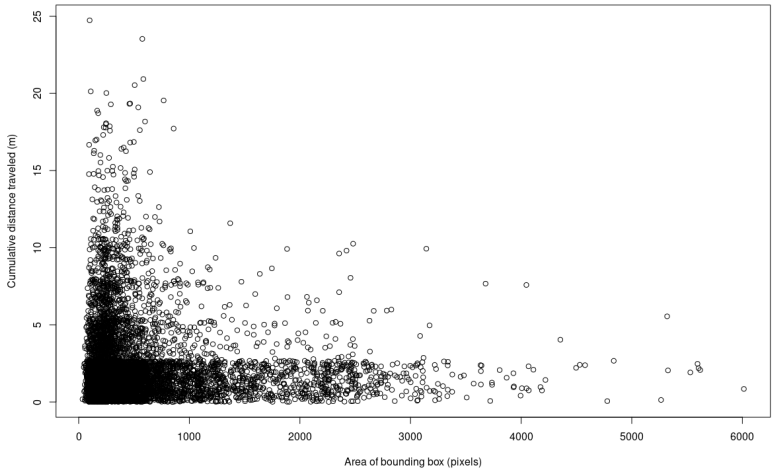


Figure 4: Scatterplot of bounding box area and distance traveled when the incorrect detection disappears.

By inspection, the data shows signs of independence and there is no evidence for extremal dependence in the upper tail. To further investigate the dependence structure one can plot the function $\chi(u)$ for various quantiles u. Figure 5 shows estimates and 95% point-wise confidence intervals for $\chi(u)$, which are given by bootstrap sampling the data as in [14].

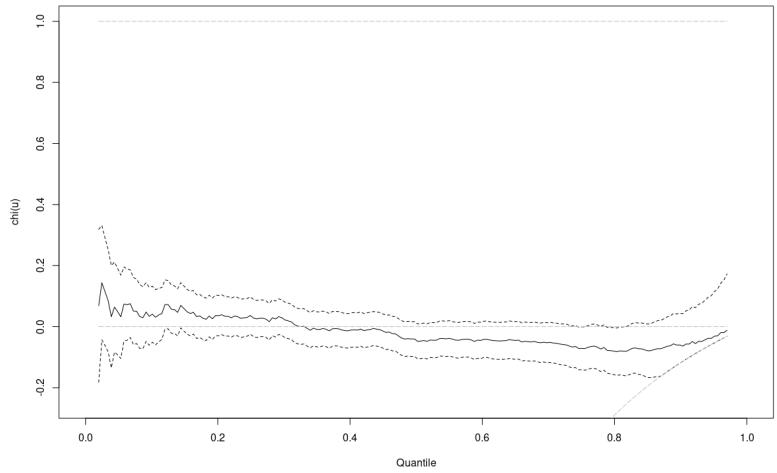


Figure 5: The estimated dependence measure $\chi(u)$ (solid line) and 95% confidence intervals (dashed line). The cases for complete dependence and independence are also shown as dashed lines at 1 and 0, respectively.

The 95% confidence intervals contain 0, which is the case for independence, for all quantiles. However, as the quantiles increases, the sign changes from positive to negative which indicates that there is some negative dependence for higher quantiles. The limit of $\chi(u)$ as $u \to 1$, the coefficient of upper tail dependence, is 0, which confirms that there is no upper tail dependence.

The constancy of $\chi(u)$ is used to check if the dependence structure of the data is in the domain of attraction of a bivariate extreme value copula. The function $\chi(u)$ does vary to some degree but the 95% confidence intervals does include 0 for all quantiles. It is however approximately constant for quantiles larger than 0.5, indicating that a threshold model taking the slight negative dependence into account is more suitable.

In summary, the dependence measure $\chi(u)$ suggests that a bivariate extreme value distribution fitted using component-wise maxima might not be the best model for the data, and that a model using threshold exceedances is a better choice.

4.2 Component-wise maxima

In figure 6 is the raw data for the incorrect detections (grey points) and the component-wise maxima (black points) for 50 blocks each of size 150. The blocks are chosen such that when fitting generalized extreme value distributions to the marginals, the Smirnov-Kolmogorov test statistic was jointly minimized over 1000 randomized trials.

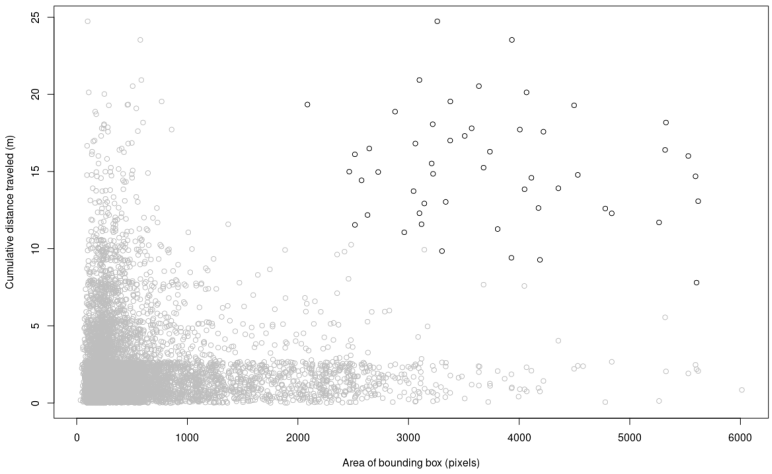


Figure 6: Data (grey points) and 50 component-wise maxima (black points) of 150 blocks

4.2.1 Parametric model

Generalized extreme value distributions were fitted to the marginals of the component-wise maxima by fitting the data to a parametric subclass of bivariate extreme value distributions. The model with the lowest AIC, which therefore was chosen, was the negative logistic model. The ML estimates of the parameters are given in table 1. The density and QQ-plots are presented in figure 7.

Parameter	Area of bounding box	Cumulative distance traveled
Location (μ)	3312 ± 125	13.85 ± 0.52
Scale (σ)	765.5 ± 90.5	3.33 ± 0.36
Shape (ξ)	-0.053 ± 0.127	-0.18 ± 0.09
Dependence parameter (α)	0.053 ± 0.013	

Table 1: ML estimates and standard errors for the fitted generalized extreme value distributions.

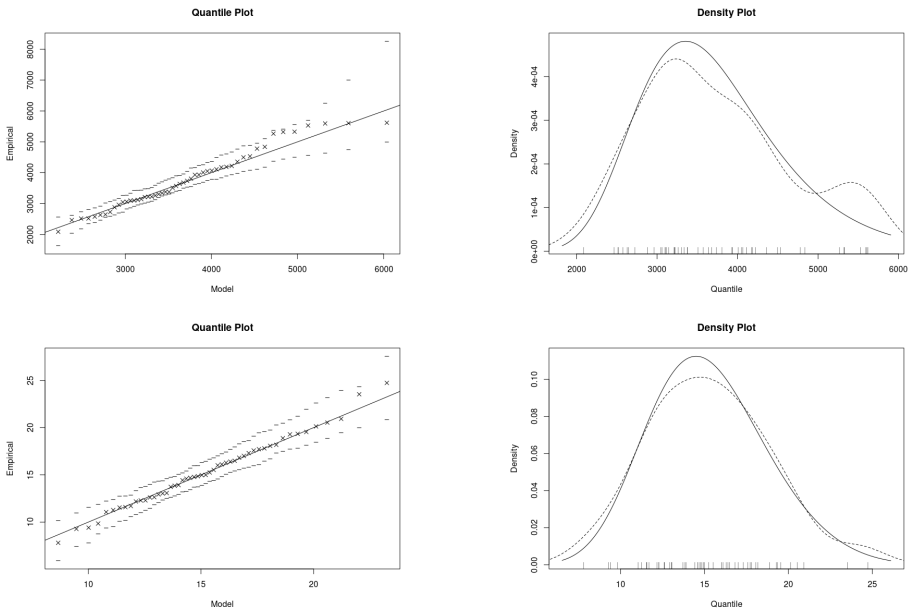


Figure 7: QQ and density plots of the GEV fits to the marginals

Both generalized extreme value distributions have a negative shape parameter, which means they are type III or Weibull distributions and so the distributions have upper bounds.

In figure 8, random variates from the distribution are shown with the estimated quantile curves.

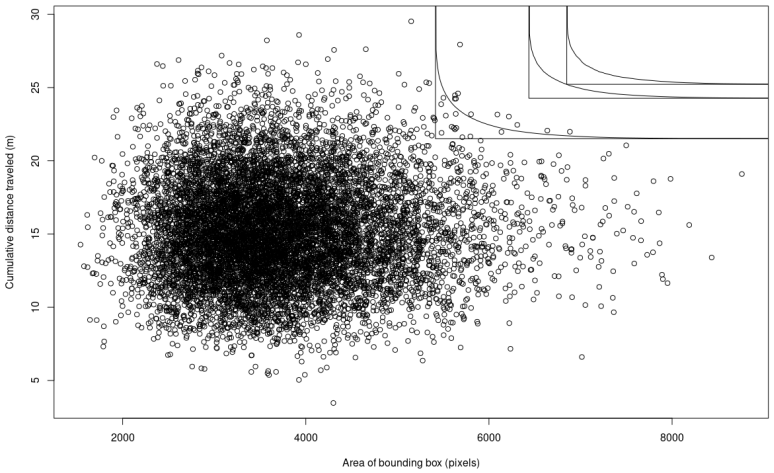


Figure 8: 10 000 random variates from the fitted negative logistic distribution together with quantile curves and threshold excess regions for $p = 0.95, 0.99, 0.995$.

The empirical probabilities for an observation to be in the threshold excess regions, denoted $\hat{p}_T$, and the estimated number of detections, $\hat{d}$, per 10^5 km for each region are presented in table 2.

p	Distance to object	Traveled distance	$\hat{p}_T$	$\hat{d}/10^5$ km
0.95	7.1	21.5	$2.5 \cdot 10^{-3}$	125
0.99	6.3	24.3	$8.2 \cdot 10^{-5}$	4.1
0.995	6	25.2	$3.1 \cdot 10^{-5}$	1.6

Table 2: Thresholds which determine the threshold excess regions corresponding to the quantile curves for $p = 0.95, 0.99, 0.995$ *and the empirical probability of an observation being in the region. Also presented is the estimated number of detections per* 10^5 *km for each region.*

4.2.2 Semi-parametric model

Using the estimated marginal distributions and Pickands dependence function, a extreme value distribution was estimated using a extreme value copula. Random variates were generated from this extreme value distribution, by simulating from the extreme value copula using the conditional distribution approach. The simulated values are presented in figure 9 with quantile curves and threshold excess regions corresponding to $p = 0.95, 0.99, 0.995$.

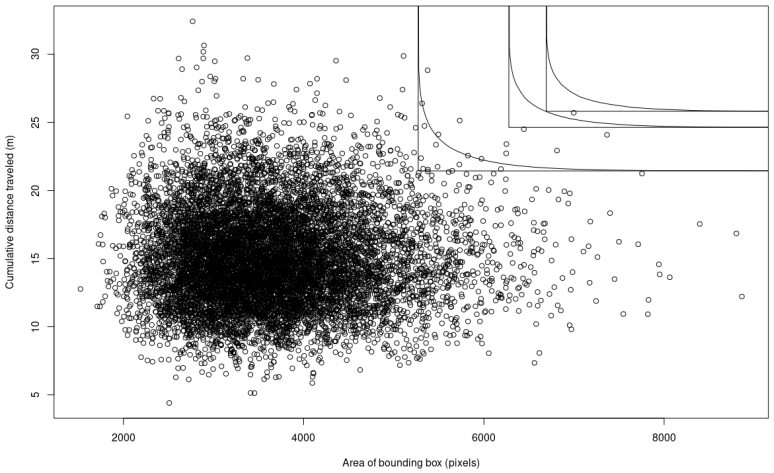


Figure 9: 10 000 random variates from the copula model together with quantile curves and threshold excess regions for $p = 0.95, 0.99, 0.995$.

The empirical probabilities, $\hat{p}_T$, for the threshold excess regions obtained from the quantile curves and the estimated number of detections per 10^5 km for each region are presented in table 3.

p	Distance to object	Traveled distance	$\hat{p}_T$	$\hat{d}/10^5$ km
0.95	7.2	21.4	$2.5 \cdot 10^{-3}$	125
0.99	6.4	24.6	$9.3 \cdot 10^{-5}$	4.7
0.995	6.1	25.8	$2.2 \cdot 10^{-5}$	1.1

Table 3: Thresholds which determine the threshold excess regions corresponding to the quantile curves for $p = 0.95, 0.99, 0.995$ *and the empirical probability of an observation being in the region. Also presented is the estimated number of detections per* 10^5 *km for each region.*

4.3 Peaks over threshold

In figure 10 is the data with the thresholds $(700, 3)$ indicated. Using QQ-plots as the diagnostic tool, the thresholds were chosen as the lowest thresholds (giving the largest amount of data), for which the generalized pareto distributions still provide a good fit.

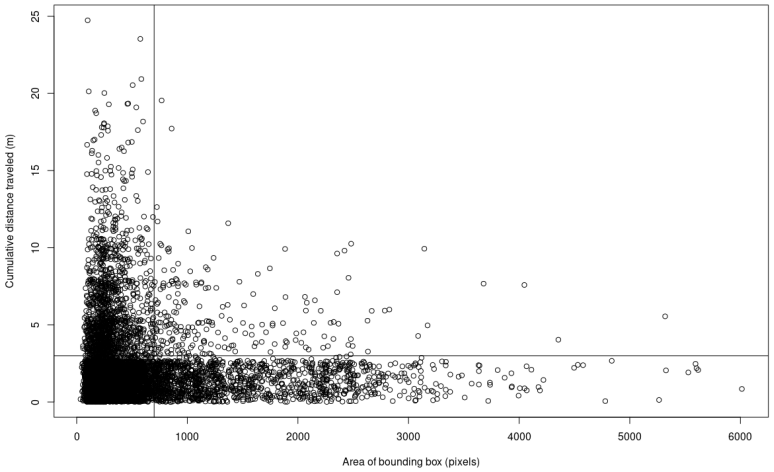


Figure 10: Data for incorrect detections together with the thresholds for the generalized pareto distributions

4.3.1 Parametric model

To the data above both thresholds, a bivariate extreme value distribution were fitted. Out of the six parametric distributions available in the POT package in R, the one with the lowest AIC, the negative logistic distribution, was chosen. The ML estimates of the parameters are presented in table 4, together with the standard errors.

Parameter	Area of bounding box	Cumulative distance
Location (μ)	700	3
Scale, (σ)	842 ± 32	3.25 ± 0.084
Shape, (ξ)	-0.025 ± 0.028	0^*
Dependence parameter (α)	$0.01569 \pm 2.072 \cdot 10^{-6}$	

*Table 4: ML estimates and standard errors for the fitted generalized pareto distributions. A * means the parameter was fixed*

The marginal corresponding to the distance traveled by the vehicle until the incorrect detection disappeared was fitted with the shape parameter fixed at

0, which corresponds to an exponential distribution. This produced the best fit and was furthermore motivated by a 95% confidence interval for the shape parameter, $(-0.075, 0.02)$, containing 0. Figure 11 contains the QQ and density plots of both fits.

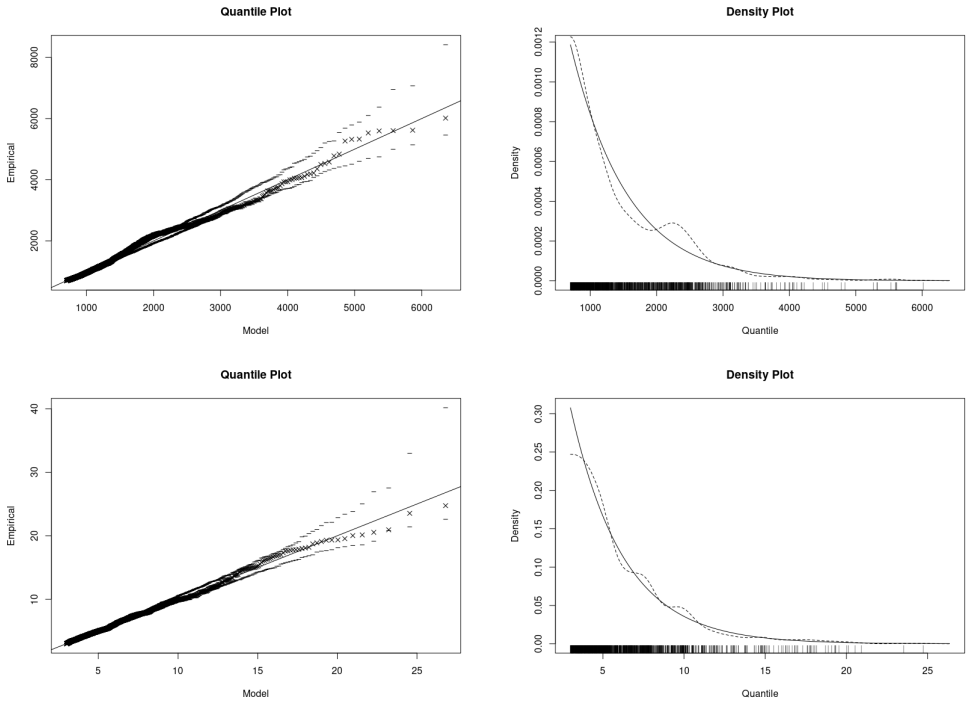


Figure 11: QQ and density plots of the GPD fits to the marginal distributions

In figure 12 are 10 000 random variates, simulated from the estimated model, together with estimated quantile curves corresponding to $p = 0.95, 0.99, 0.995$.

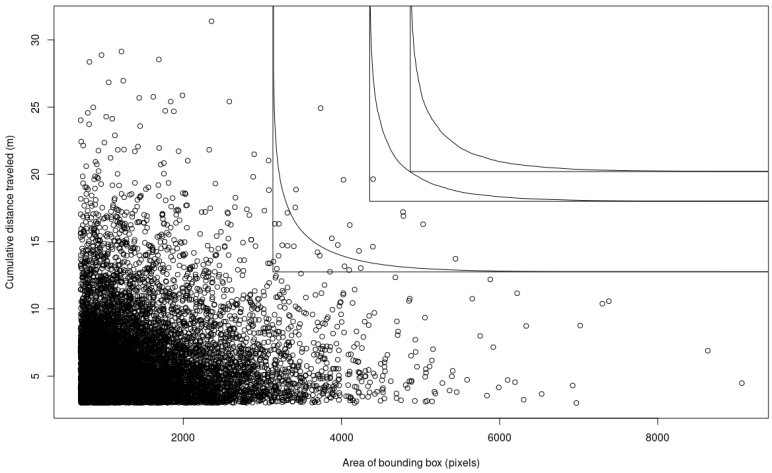


Figure 12: 10 000 random variates from the fitted negative logistic model together with quantile curves and threshold excess regions for $p = 0.95, 0.99, 0.995$

.

The empirical probabilities, $\hat{p}_T$, for the threshold excess regions obtained from the quantile curves and the estimated number of detections, $\hat{d}$, per 10^5 km for each region are presented in table 5.

p	Distance to object	Traveled distance	$\hat{p}_T$	$\hat{d}/10^5$ km
0.95	8.9	12.7	$2.5 \cdot 10^{-3}$	540
0.99	8	17.9	$9.7 \cdot 10^{-5}$	21
0.995	7.5	20.2	$2.5 \cdot 10^{-5}$	5.4

Table 5: Thresholds which determine the threshold excess regions corresponding to the quantile curves for $p = 0.95, 0.99, 0.995$ *and the empirical probability of an observation being in the region. Also presented is the estimated number of detections per* 10^5 *km for each region.*

4.3.2 Semi-parametric model

To create a semi parametric model using threshold exceedances, the marginals are modeled as generalized pareto distributions with the parameters that were estimated in the previous section. A bivariate distribution with these marginals is then obtained using a copula, as in [12]. A suitable copula to use because of its lower tail dependence is the Kimeldorf and Sampson copula,[15]

$$C(u, v) = [u^{-\delta} + v^{-\delta} - 1]^{\frac{-1}{\delta}}$$

where $\delta \geq 0$.

The dependence parameter δ can easily be estimated by using the connection to Kendall's tau, $\tau = \frac{\delta}{\delta+2}$ [15]. Kendall's tau for the data of the exceedances is $\tau = -0.044$ which gives $\delta = -0.084$. The generalization of the Kimeldorf and Sampson copula which handles negative correlation is the widely popular Clayton copula,

$$C(u, v) = \max(0, [u^{-\delta} + v^{-\delta} - 1])^{\frac{-1}{\delta}}$$

where $\delta \geq -1$

For negative δ, C is 0 if $v < (1 - u^{-\delta})^{\frac{-1}{\delta}}$, which makes such pairs (u, v) an area where the copula is not useful. If this area is too big, the copula can not be used. For $\delta = -0.18$, 10^6 random uniform numbers were generated, and the occurence of C vanishing was estimated to be 0.002%, which is considered to be negligible.

Assuming for notational simplicity that F_1 and F_2 are standard GPD marginals, the estimated joint distribution function is

$$F(x, y) = C(F_1(x), F_2(y)) = \max(0, [(1-(1-\xi_1 x)^{\frac{-1}{\xi_1}})^{-\delta} + (1-(1-\xi_2 y)^{\frac{-1}{\xi_2}})^{-\delta}])^{\frac{-1}{\delta}}$$

In figure 13 is 10 000 random variates from the estimated distribution, simulated from the copula using the conditional distribution approach.

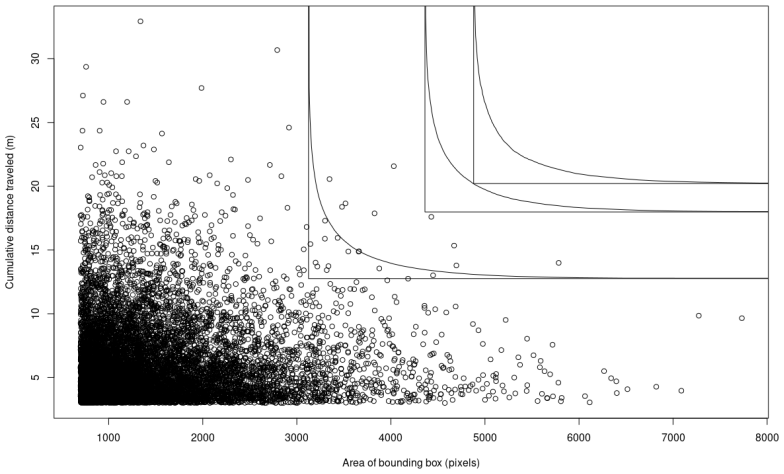


Figure 13: 10 000 random variates from the estimated bivariate distribution with GP margins, together with quantile curves and threshold excess regions for $p = 0.95, 0.99, 0.995$.

The empirical probabilities, $\hat{p}_T$, for the threshold excess regions obtained from the quantile curves and the estimated number of detections, $\hat{d}$, per 10^5 km for each region are presented in table 6.

p	Distance to object	Traveled distance	$\hat{p}_T$	$\hat{d}/10^5$ km
0.95	8.7	12.8	$2.3 \cdot 10^{-3}$	495
0.99	8	18	$9 \cdot 10^{-5}$	19.4
0.995	7.5	20.2	$2.5 \cdot 10^{-5}$	5.4

Table 6: Thresholds which determine the threshold excess regions corresponding to the quantile curves for $p = 0.95, 0.99, 0.995$ and the empirical probability of an observation being in the region. Also presented is the estimated number of detections per 10^5 km for each region.

5 Model comparison

In this section the various models will be compared in terms of their abilities to accurately estimate the risk of an observation being large in both components.

The data set is randomly split into a smaller and larger set in the ratio 1:9. The non-parametrically estimated quantile curves of the larger set is used as ground truth. Using the smaller set the various models were fitted. The quantile curves were then estimated and compared to the ground truth quantile curves. The number of blocks for the component-wise maxima were set to 50 as before and the thresholds for the threshold models were chosen using a function available in the evd package in R.

To measure how accurate the estimated quantile curves are, a similarity measure of these curves is needed. A suitable metric, which takes the location and ordering of the points along the curves into account is the Fréchet distance [13]. Let X be a metric space with metric d and γ_1, γ_2 be two curves in X. Furthermore let α and β be any reparametrizations of γ_1, γ_2. The Fréchet distance between two curves γ_1, γ_2 is defined as

$$d_F(\gamma_1, \gamma_2) = \inf_{\alpha,\beta} \max_{t\in[0,1]} d(\gamma_1(\alpha(t)), \gamma_2(\beta(t))$$

where d is the metric in X. Here, $X = \mathbb{R}^2$ and d is the euclidean metric.

A common way of explaining the definition of the Fréchet distance is that it corresponds to the length of the shortest leash needed for a man walking along one curve to be connected to his dog, walking on the other curve. They can both vary their speed, but not walk backwards.

The mean of the distance between the quantile curves for $p = 0.95, 0.99, 0.995$ quantile curves and the sum of these distances over 10^5 repetitions of the above procedure is presented in table 7.

p	Param. POT	Semi-param. POT	Param. Maxima	Semi-param. Maxima
0.95	9.31	9.47	18	17.45
0.99	9.21	9.39	17.83	17.2
0.995	9.19	9.36	17.66	17
Σ	27.71	28.22	53.49	51.65

Table 7: Mean Fréchet distance between the estimated quantile curves and the quantile curves of the data for $p = 0.95, 0.99, 0.995$. The last row contains the sum of the distances for each model.

The Fréchet-distance measures how far away the predicted quantile curve is from the quantile curves of the data. Since d_F is a metric, a perfect prediction of the quantile curves would correspond to $d_F = 0$. Thus, the results are to be interpreted as a lower mean distance being better. The sum of the distances over $p = 0.95, 0.99, 0.995$ is given as an over all score for the model.

6 The extreme value's dependency on velocity

So far the velocity of the vehicle has not been taken into account. The velocity is of course natural to include in any risk analysis regarding vehicle safety. As the other quantities, the velocity was measured directly when the incorrect detection disappeared.

It was not possible to fit any extreme value distribution to the velocity data, as the self driving autopilot of the CARLA simulator produced little variation in the velocity. The autopilot is also set to keep a velocity of around 70% of the current speed limit. This in combination with the environment being an inner city, the resulting data for the velocity is lower than it would be in a real life application.

A question is whether or not the bivariate extreme value distributions that were estimated earlier depends on the velocity of the vehicle. To answer this question the data was divided into groups with mean velocities v_1, v_2, v_3 equal to $10, 18, 40$ (km/h), respectively. Again, the best performing model, the negative logistic model was fitted using threshold exceedances. The thresholds for the threshold excess region obtained from the quantile curves are presented in table 8.

p	Group 1 ($v_1 = 10$)	Group 2 ($v_2 = 18$)	Group 3 ($v_3 = 40$)
0.95	8.3, 10.8	9, 12.8	9.4, 19.8
0.99	7.2, 13.9	8.3, 17.7	7.9, 24.7
0.995	6.7, 15	7.9, 19.7	7, 29.2

Table 8: Thresholds for the threshold excess region for $p = 0.95, 0.99, 0.995$. The numbers correspond to the threshold for the distance to the object in meters and the cumulative distance traveled until the object disappears.

7 A risk measure based on velocity

One might be interested in a risk measure that incorporate the velocity. This could be achieved by considering the braking distance, which is a quadratic function of the velocity. The idea is that a safety system of a self driving vehicle could for example brake abruptly if an object is incorrectly detected and the braking distance is under a certain threshold.

The formula is easily derived by equating the work $W = \mu mgd$ done by braking and the kinetic energy $E = \frac{mv^2}{2}$ of the vehicle, where μ is the coefficient of friction between the road surface and the tires, g is the gravity of the earth, and d is the distance travelled. Thus, the braking distance can be written as a function of the velocity:

$$b(v) = \frac{v^2}{2\mu g}$$

An assumption is needed regarding the coefficient of friction. Here, it is set as $\mu = 0.8$, which covers most normal situations [17]. A critical region can then be

defined by considering points for which $b(v) + t > d$ where d is the distance to the object and t is a tolerance.

As an example, consider the data for the third group in the previous section. In figure 13 is a plot of the area of the bounding box and the velocity, with the region for $t = 0$ displayed. This critical region contains five detections, which gives an empirical probability of $6.5 \cdot 10^{-4}$ for such an event.

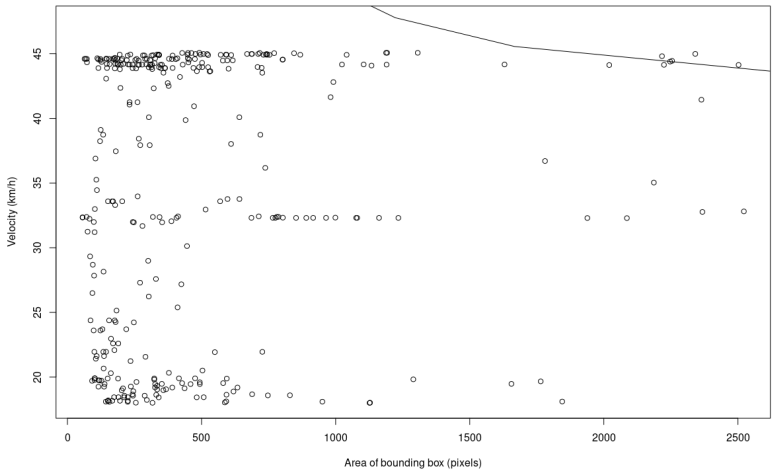


Figure 14: Scatterplot of bounding box area and velocity for incorrect detections, together with the critical region

8 Conclusions

Extreme value theory has the potential to be a valuable tool in estimating the risk of sensor failures, and some of the various techniques and methods that could be employed in such extreme value models have been presented here. The methods that mainly are used in extreme value theory, the block maxima method and peaks over threshold, were both investigated. For the data considered here, it is clear that a model using threshold exceedances is to be preferred. This is not only because of preserving more of the data, but mostly because the distribution of the data above both thresholds serves as a better approximation of the general distribution than the component-wise maxima.

This agrees with the results of the model comparison, where the models using threshold exceedances on average produced quantile curves with smaller Fréchet distance to the non-parametric quantile curves of the data set which was used as ground truth. It should be noted that since the data set itself was used for generating the ground truth quantile curves which are used to compare the models, the results of this section are not to be interpreted as the model's abilities to accurately predict extreme observations. Comparing the models using a larger data set, or even data for extreme observations is to be preferred, but such data is usually scarce or non-existent, a general problem in extreme value theory. However, the method of using the Fréchet distance to compare quantile curves (or other risk measures that can be realised as simple curves) could be used as a general strategy for comparing risk estimates of models to real, observed data in further research.

In general, the threshold excess regions are slightly larger for the threshold models, meaning that the rare and potentially dangerous incorrect detections of the sensors are estimated to be less severe. The estimated frequencies of observations large in both components are thus also higher for these models. The frequencies decrease rapidly as p increases since the probability of an observation being in the threshold excess region corresponding to a probability p is $(1-p)^2$.

This difference in the risk estimates stems from the common criticism of models based on component-wise maxima, which is that it produces synthetic observations that are not in the data set, something that is clear from figure 6. For independent data, these synthetic observations are different from the observed data and the relevance of such a model must be questioned. However, the component-wise maxima inherit the independence of the data, and thus produces a model with results not that different from the threshold models. This type of model may still have value in applications concerning safety, as it can be used to provide a small overestimation of the risk estimates. If such a model estimates risks that are within tolerable limits, then it can be seen as an indication that the process of generating a large data set to be used for validation is worthwhile, as the number of observed extreme observations is more likely to also be within the limits.

There is no real difference in the risk estimates for semi-parametric and parametric models. The reason for this is the dependence structure of the data. Although a semi-parametric model using threshold exceedances can incorporate

negative dependence structures of data more flexibly, for example by reflecting the copula, the dependence structure of the data is still too close to independence for it to make a real difference. It shows however the flexibility of modeling dependence using copulas.

For all models, the estimated regions include incorrectly detected objects that are within 10 meters before disappearing. A reason for the distance not being larger is that on average, the velocity determined by the autopilot in CARLA is lower than it would be in reality. The extreme value's dependency on the velocity was investigated and the analysis indicated that the risk estimates depend on the velocity, with the dependence differing for the traveled distance and the distance to the object. As the velocity increases, the vehicle travels a longer distance while incorrectly detecting an object and this influenced the risk estimates presented in this section. On the other hand, the distance to the object becomes larger as the velocity increases. This might seem counter-intuitive but could be explained by the fact that a higher velocity increases the chance of driving past the object between frames, resulting in the measurement being made in an earlier frame where the distance to the object was larger. In any case, the methodology of investigating factors that potentially influence extreme values is an important one and should be carried out extensively. Methods similar to the one used here, where data is divided into discrete groups and extreme value analysis is performed for each group, can serve as an instructive way of investigating such dependence in real-life applications.

To motivate how risk estimates from extreme value models such as those presented here can be used in validating the safety of a perception system, what is meant by sufficiently safe needs to be defined. This is an open question but as a minimum requirement, one can refer to the ethical guidelines from [21] stating that self-driving vehicles should be at least as safe as human drivers. However, such a requirement is not very precise and it is not exactly clear what types of accidents should be included.

To further motivate the complexity of proving that a perception system is sufficiently safe, the amount of time 100 autonomous vehicles would need to drive at an average speed of 25 miles per hour to have sufficient data was estimated to be 225 years [22]. In this estimate the authors used the number 1.09 deaths per 100 million miles driven and they note that this number decreases with time, thus making the requirement harder and harder to fulfill. An estimation of the amount of data collected from an 8-megapixel camera recording at 10 frames per second is $170 \cdot 10^6$ Terabytes, which is impossible to store. One should also note that the number 1.09 is estimated from all types of situations which shows that in individual cases there might be more aspects to consider and the requirements need to be more conservative.

A possible way to regard such safety requirements is thus to divide accidents into subsets with varying tolerances for each subset. For example, incorrect detections might cause false braking which could lead to a rear-end crash. Approximately 1.8 million rear-end crashes occur in the U.S annually [23]. With 5 trillion driven kilometers per year estimated by the federal highway administration, this gives a frequency of 2.78 rear-end crashes per million km. Out

of these, it is estimated that 0.2 percent lead to an injury on the maximum abbreviated injury scale (MAIS) corresponding to the worst type [24]. Thus, rear-end crashes resulting in such injuries occur less frequently than 5.6 per 10^9 km and so an autonomous vehicle would need to have a frequency lower than that to be considered safer than an average human driver.

Not all false braking will lead to rear-end crashes and potentially a severe injury. Accurate estimations of this probability are hard to make. For simplicity and sake of argument assume that this probability is 0.01. Then false braking should occur less frequently than 5.6 per 10^7 km. Now, as a simplified model assume that incorrectly detected objects lead to false braking if the vehicle drives a distance larger, and the object is closer than some pre-determined thresholds. Then, to fulfill the requirement of a false deceleration occurring less than 5.6 times per 10^7 km, a large amount of data needs to be collected. Before starting this process of data collection, risk estimates from extreme value models such as those presented here can be used as an indication of whether or not the perception system fulfills the requirement, which can save time and resources by not having to restart the process.

As a final note, this thesis also shows that what is being measured, and how it is measured is of great importance in creating extreme value models. The distance traveled while incorrectly detecting an object and the remaining distance to it when it disappears resulted in an approximately independent dependence structure, with no upper tail dependence. Thus, if this methodology is applied in a real-life setting, care should be taken in the analysis of the dependence structure, especially when modeling the dependence in the extremes.

9 Further research

Further research should be carried out using a non-simulated data. As was investigated in the later sections of the thesis, different notions of what is considered to be a dangerous situation, using more than two random variables, should be determined. Such models would result in more complicated dependence structures for which tools outside the scope of this thesis are needed. An example of such a tool is vine copulas, [18], which are described as a flexible way of handling higher-dimensional settings. The idea is based on using bivariate copulas as building blocks for higher-dimensional distributions. Thus, the copula methods presented here could be applied for pairs of random variables, which then in turn can be used to obtain a vine copula.

The risk measure considered here, the quantile curves of a bivariate extreme value distribution, is a natural generalization of the traditional univariate risk measures such as value at risk or expected shortfall. An interesting topic to explore is other kinds of risk measures, both in the bivariate and the multivariate case. This was briefly touched upon by presenting a risk measure based on velocity, which is only relevant for this particular application. It should be noted that this risk measure only takes the velocity of the vehicle and the distance to the object into account, neglecting other factors that decide the severity of

the situation. Other multivariate risk measures that might be suitable in this area are perhaps those based not only on the frequency of observations over certain thresholds but also on the mean of how large the threshold exceedances are when they do occur. This is indeed an active area of research in modern times, [19], [20]. Although often developed with financial models in mind, they should translate with ease to the current application.

Finally, the data that was obtained for missed detections using this detection model was insufficient to create any extreme value models. An interesting and important topic for further research is models for objects that should be detected but are missed by the detection model. As a first step, marginals similar to those considered here could be used together with the tools presented for estimating dependence structures.

References

[1] WHO, ed. 2018. *Global Status Report on Road Safety 2018* Geneva: World Health Organization (WHO). pp. xiv–xv, 1–13, 91ff (countries), 302–313 (table A2), 392–397 (table A11). ISBN 978-92-4-156568-4. Retrieved 5 May 2019. Tables A2 & A11, data from 2016

[2] OECD/ITF, ed. (2018). *Road Safety Annual Report 2018*. Paris: International Transport Forum (itf). p. 21.

[3] Alexey Dosovitskiy and German Ros and Felipe Codevilla and Antonio Lopez and Vladlen Koltun. *CARLA: An Open Urban Driving Simulator*. Proceedings of the 1st Annual Conference on Robot Learning, p.1-16, 2017

[4] Beirlant, J. ; Goegebeur, Y. ; Segers, J.J.J. ; Teugels, J. / Statistics of Extremes : Theory and Applications. Chichester : Wiley, 2004. 522 p.

[5] A. G. Stephenson. evd: Extreme Value Distributions. R News, 2(2):31-32, June 2002. URL: `https://CRAN.R-project.org/doc/Rnews/`

[6] M. Ribatet and C. Dutang, POT: Generalized Pareto Distribution and Peaks Over Threshold, 2019, R package version 1.1-7, `https://CRAN.R-project.org/package=POT`,

[7] Roger B. Nelsen. 2010. An Introduction to Copulas. Springer Publishing Company, Incorporated.

[8] G. Gudendorf, J. Segers, 2011 *Nonparametric estimation of multivariate extreme-value copulas* `https://arxiv.org/abs/1107.2410`

[9] Bücher, Axel and Dette, Holger and Volgushev, Stanislav, 2011,
New estimators of the Pickands dependence function and a test for extreme-value dependence
The Annals of Statistics, Institute of Mathematical Statistics URL: `http://dx.doi.org/10.1214/11-AOS890`

[10] Tawn, J.A., *Bivariate extreme value theory: models and estimation*, Biometrika 75, 397-415, 1988

[11] Coles, S., Heffernan, J. & Tawn, J. Dependence Measures for Extreme Value Analyses. Extremes 2, 339–365 (1999). `https://doi.org/10.1023/A:1009963131610`

[12] I. Ghosh, O. Banks, 2020, *A Study of Bivariate Generalized Pareto Distribution and its Dependence Structure Among Model Parameters*, Sankhya B. `https://doi.org/10.1007/s13571-020-00224-z`

[13] Erin Wolf Chambers, Éric Colin de Verdière, Jeff Erickson, Sylvain Lazard, Francis Lazarus, Shripad Thite, 2010, *Homotopic Fréchet distance between curves or, walking your dog in the woods in polynomial*

time, Computational Geometry, Volume 43, Issue 3, Pages 295-311, https://doi.org/10.1016/j.comgeo.2009.02.008

[14] Fermanian JD, Radulovic D and Wegkamp M, 2004 *Weak convergence of empirical copula processes*, Bernoulli.

[15] H. Joe, 2001, *Multivariate models and dependence concepts*, Boca Raton, FL, Chapman & Hall/CRC.

[16] Frahm, Gabriel, Junker, Markus and Schmidt, Rafael, 2005, *Estimating the tail-dependence coefficient: Properties and pitfalls*, Insurance: Mathematics and Economics, 37, issue 1, p. 80-100, https://EconPapers.repec.org/RePEc:eee:insuma:v:37:y:2005:i:1:p:80-100

[17] G. Taoka, 1989, *Brake Reaction Times of Unalerted Drivers*, ITE Journal, p. 19-21.

[18] Harry Joe, Haijun Li, Aristidis K. Nikoloulopoulos, 2010, *Tail dependence functions and vine copulas*, Journal of Multivariate Analysis, Volume 101, Issue 1, Pages 252-270, ISSN 0047-259X, https://doi.org/10.1016/j.jmva.2009.08.002

[19] Tomer Shushi, Jing Yao, *Multivariate risk measures based on conditional expectation and systemic risk for Exponential Dispersion Models*, 2020 Insurance: Mathematics and Economics, Volume 93, Pages 178-186, ISSN 0167-6687, https://doi.org/10.1016/j.insmatheco.2020.04.014

[20] Chen, Y., Hu, Y. *Multivariate coherent risk measures induced by multivariate convex risk measures*, 2020, Positivity 24, 711–727. https://doi.org/10.1007/s11117-019-00703-2

[21] Lütge, Christoph. (2017). The German Ethics Code for Automated and Connected Driving. Philosophy & Technology. 10.1007/s13347-017-0284-0.

[22] Kalra, Nidhi and Susan M. Paddock, Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?, Santa Monica, Calif.: RAND Corporation, RR-1478-RC, 2016. As of June 04, 2021:

https://www.rand.org/pubs/research_reports/RR1478.html

[23] Distner, Martin & Bengtsson, Mattias & Broberg, Thomas & Jakobsson, Lotta. (2021). *City safety - a system addressing rear-end collisions at low speeds*

[24] Eis, Volker & Sferco, Raimondo & Fay, Paul. (2005). *A deatailed analysis of the characteristics of european rear impacts*