Prepared for:

Rijkswaterstaat

Dienst Getijdewateren

Definition of joint extreme condition statistics of hydraulic parameters off the coast of The Netherlands

Final Report

September 1991

H 957

delft hydraulics

Definition of joint extreme condition statistics of hydraulic parameters off the coast of The Netherlands

C.F. de Valk

delft hydraulics

CONTENTS

APPENDIX

EXECUTIVE SUMMARY

This report presents an attempt to define joint statistics of waves and water levels at the -20 m depth contour offshore the coast of the Netherlands. It covers
- listing and definition of relevant hydraulic parameters
- combinations of hydraulic parameters for which statistics will be produced
- the type(s) of statistics of these parameter combinations, the exact definitions of these statistics
- parameterization of these statistics, needed in order to estimate them from available field data and numerical hindcast data

For statistics at the -20 m contour, it is proposed to focus on the parameters high tide level, significant wave height, mean wave period, mean wave direction, wind speed, and wind direction. In addition, spectral width and directional spreading of the waves, as well as current can be considered. Joint statistics should pertain to the combination of the six most important parameters. Estimation of these from data is not feasible in practice, so some simplifying assumptions are required.

Since for assessment of the safety of the coast storm events with very low frequencies of occurrence are of interest, there is no difference between mean number of occurrences per year and probability of occurrence in an arbitrary year. Statistics of failure or damage of a water retaining structure can be derived from a single type of joint statistic of hydraulic parameters in cases of practical interest. This statistic is independent of failure mechanism, and can be estimated at the -20 m contour and then translated to the coast. Some ideas are presented about how to take limitations on the number and quality of available data into account into final estimates.

It is recommended to use methods for joint statistics of hydraulic parameters that are consistent with the most recent approach to statistics of high tide levels, based on a peak-over-threshold approach to data selection and the use of a Generalized Pareto distribution to specify exceedance frequencies.
Two different approaches are discussed:
- the current approach known as the method of Bruinsma/Van Aalst, which is based on simplification of the physics of generation of surges and waves
- direct estimation of the required statistics from data using the asymptotic shape for the probability that several hydraulic parameters simultaneously exceed high levels during a storm

It is recommended to test these approaches separately to identify strengths and weaknesses of each method, and then to compare them. Tests can be based on hindcast (NESS) data and on measured data. Based on this experience, improvements can be implemented and the final choice of method can be made.

1. Introduction

This report presents the results of a study aiming at the definition of joint statistics of waves and water levels at the -20 m depth contour offshore the coast of the Netherlands. The final goal is to come to a single consistent set of statistics to be used for evaluation of the safety of the coast of the Netherlands. Numerical models will be employed to translate these statistics at the -20 m contour to statistics at locations near the coast. At each location, the failure mechanisms pertaining to the protecting structure determine which statistic(s) are finally derived from the basic statistics.

The starting point for the development of a method for assessment of statistics of hydraulic boundary conditions at the -20 m contour is the approach developed at Rijkswaterstaat [*Vrijling and Bruinsma*, 1980; *Bruinsma*, 1982; *Van Aalst*, 1983]. This approach is considered valuable because it recognizes wind as the common source of surges and waves in the North Sea. Since that time, however, new sets of data became available, both measurements and numerical hindcast data. Moreover, developments in statistics of extreme events have taken place, and there appears a need for more rigorous definitions of parameters and statistics. This report covers

- listing and definition of relevant hydraulic parameters
- combinations of hydraulic parameters to be considered
- the type(s) of statistics of these parameter combinations, the exact definitions of these statistics
- methods for parameterization of these statistics, needed in order to estimate them from available field data and numerical hindcast data

Some attention is paid to estimation of statistics from data and to the accuracies of estimates. Also certain aspects of selection and validation of measured data and hindcast data are discussed that are related to estimation. In the conclusions, the various choices are summarized, and an outline is given of the proposed approach. New results presented in this report are preliminary in the sense that thorough checking of correctness and feasibility remain to be done.

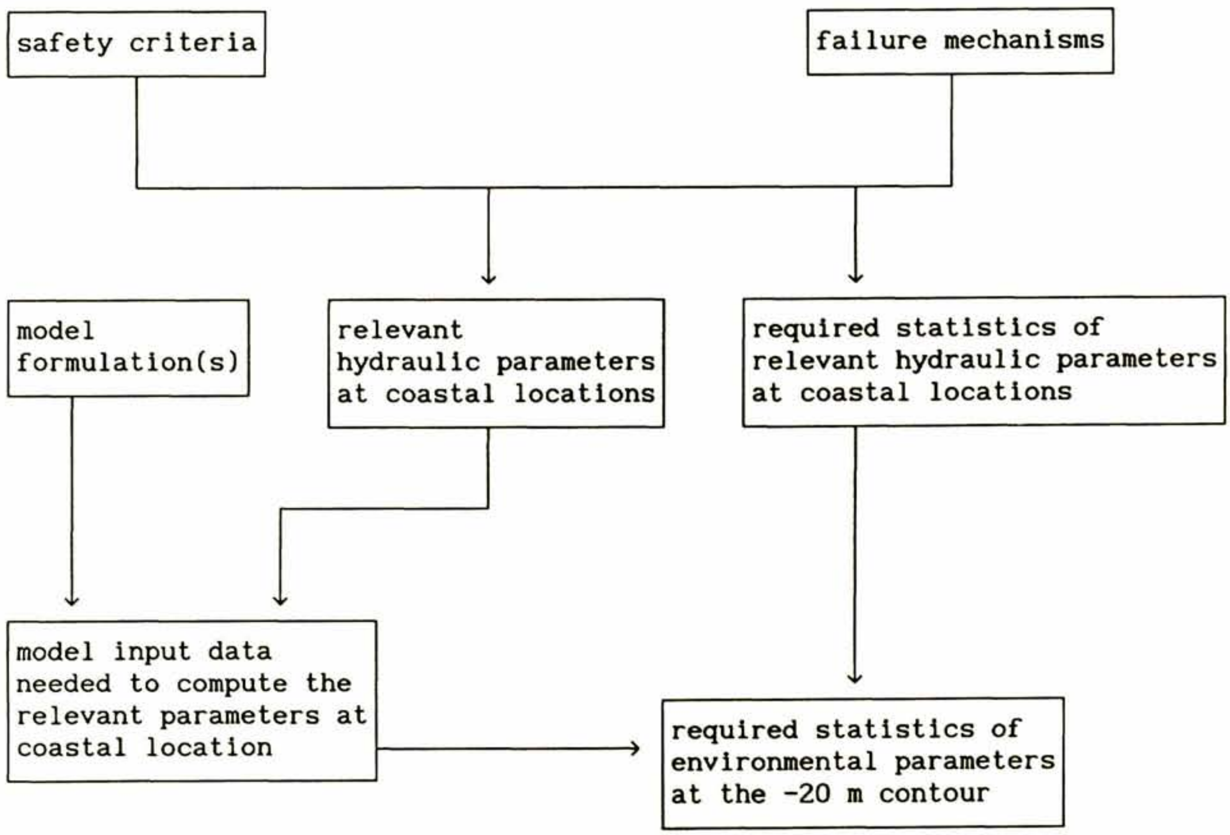


figure 1: logical model for determining the requirements on statistics at -20 m depth.

An alternative approach would be to interpolate available historical (measured or hindcast) data over the -20 m contour, and to generate historical data at all coastal locations using numerical models. Then at each separate location, the statistics required for the relevant failure mechanisms are estimated from these data. The advantage of this approach is that statistics do not need to be generic (applicable to all failure mechanisms). Several objections to this approach have already been indicated, which have lead to the decision to produce a single set of statistics for relevant environmental parameters at the -20 m contour, applicable to all failure mechanisms.
In the following chapters, some requirements will be discussed in more detail than is possible in this introductory chapter.

2. Requirements

The safety of the coast is threatened by extreme storm events generating high water levels and high waves, the combination of which can cause failure of dunes, sea walls, dikes and storm surge barriers.
In principle, there are no limits to the magnitudes of wave heights or water levels that may occur. As a consequence, safety criteria are based on an acceptable probability or frequency of failure of the protecting structure, rather than on the requirement of absolute safety. Therefore, statistics of the hydraulic conditions controlling failure of the structure must be assessed.
These hydraulic conditions involve water level and wave parameters such as wave height and wave period. Since the combined effect of waves and water level on a structure is generally complex, joint statistics of wave and water level parameters are required. The exact specification of the required statistics at coastal locations depends on the relevant failure mechanisms. Therefore, the set of statistics computed at the coastal locations should be general enough to cover all possible requirements.
In general, failure of a structure is not instantaneous. To assess the effects of hydraulic conditions on a structure, nowadays dynamical models are used. This implies that not only the magnitudes of certain hydraulic parameters, but also the persistence (duration) of extreme conditions has to be modeled statistically in order to assess the expected frequency of failure of the structure.
Waves undergo transformations when propagating from relatively deep water to the coast. Variation in coastline and depth close to the shore cause considerable variation in wave conditions nearshore. In order to obtain statistics of waves and water levels that are relatively invariant in space and time[1], they are preferably specified at relatively deep water, but not too far from the coast. Traditionally, the -20 m contour is chosen as a compromise. Another reason for specifying the statistics at relatively deep water is that it simplifies the description of wave spectra.

Shallow water wave models are used to translate environmental parameters at the -20 m contour to wave parameters near the coast. Therefore, the set of statistics to be specified at -20 m depth is determined by the required statistics near the coast, as well as by the model input data required to compute all relevant hydraulic parameters near the coast.

The dependence of the required statistics at the -20 m contour on safety criteria and failure mechanisms is summarized in the following diagram.

[1] i.e. invariant to changes in morphology of the sea bottom

3. Definition of parameters and statistics

3.1 Selection and definition of parameters

List of parameters

For statistics at the -20 m contour, it is proposed to restrict the set of physical parameters to:

1. high tide level
2. significant wave height
3. mean wave period
4. mean wave direction
5. wind speed
6. wind direction

Of these, extrapolation of statistics beyond the range of available data is necessary only for high tide level, significant wave height, mean wave period and wind speed; directions are limited to the interval $[-\pi,\pi]$. In addition, statistical information about

7. spectral width parameter
8. directional spreading of the waves
9. current speed
10. current direction

may be required. However, these parameters are not considered critical. The purpose of statistics of these parameters is merely to get an impression about the range of values occurring, in order to define proper boundary conditions for the shallow water wave models that are used to translate wave statistics from the -20 m contour to the shore.

Motivation

⟶1: For the *water levels*, a number of options are available besides high tide levels. The surface elevation measured at a fixed point in the plane is the basis of all definitions related to water level.
The first step is in general to exclude sea surface waves with periods of less than say 30 s, so when we talk about water level, a filtered surface elevation signal is meant in which all periods below 30 s are suppressed.
This signal is still not very useful for computation of statistics since it oscillates with the astronomical tide; since the astronomical components are predictable, there is a lot of redundant information in these data. So the first step is to remove the astronomical oscillation from the data. The most straightforward approach is to simply subtract the astronomical signal (as determined by tidal analysis) from the water level signal. However we will

still find features related to the astronomical tide, due to interactions between astronomical tide and variations caused by wind and by atmospherical pressure gradients. During a storm, it is in fact impossible to make a unambiguous distinction between astronomical tide and "meteorological tide" *at every instant*. However by focusing on *high tide level* (maximum of the water level between to tidal minima) only, the oscillation due to astronomical tide is removed. Moreover, it is much easier to make a meaningful distinction between astronomical and meteorological components: the meteorological component is simply the difference between measured high tide level and astronomical high tide level, and we do not have to bother about interactions (as far as the definition is concerned). The meteorological component thus defined can be the result of different physical mechanisms: we can distinguish set-up (slope of the surface of the North Sea built up by wind stress or atmospheric pressure gradients), and oscillations on different time scales, associated with the set-up, wave-wave interactions, etc. This implies that all variations on time-scales slower than those normally considered for sea surface waves contribute to the meteorological component. Currently, the water level signal is filtered by computing 10 minute averages. This means that all oscillations with periods suppressed by this filter are excluded from the analysis, and statistics of this fast component (short waves, long waves) have to be made separately. In this report, we will focus on high tide levels and short sea surface waves (with periods below say 30 s). Longer waves can be important too in practice (seiches). For statistics of long waves, see for example [*Vogel and De Valk*, 1991; *De Valk*, 1991]. Statistical relationships between long waves and high tide level can be treated in a similar way as discussed for short waves in this report.
The joint statistics to be produced will be of high tide level and wave parameters rather than of set-up and wave parameters (or astronomical tide and wave parameters). The motivation is that high tide level is the relevant parameter for the design of constructions. Statistical relationships between waves and set-up or astronomical tide may or may not be used at intermediate stages of the estimation procedure, but are not of practical relevance in the end.

⟶2&3: Most important wave parameters near the construction are a wave height parameter, and a wave period parameter, together also determining (by means of the dispersion relationship) a measure of wave steepness. Shallow water wave models describe the wave field in terms of wave spectra (variance spectra or action spectra). Therefore, the relevant parameters at the -20 m contour are spectral parameters. In wave models like HISWA [*Holthuijsen et al.* 1989], a parameterization of the wave spectrum is employed, which means that a few parameters are sufficient to fix the boundary conditions at the -20 m contour. This approach is probably also valid for models which do not parameterize the spectrum. Analyses of wave measurements and numerical wave hindcasts at -20 m depth produce spectral parameters too, so no problems are expected. For design purposes, relevant parameters are not always spectral parameters. For example, the maximum wave height or the wave height exceeded by the highest two percent of the waves can be more relevant. In general, such parameters can be computed once the shape of the spectrum is known, which usually depends on depth and possibly also on current.

⟶4: Information about wave direction is important mainly because wave directionality affects wave propagation in shallow water (in particular in the case of a complex bathymetry such as in estuaries) but also because this parameter is relevant for certain failure mechanism(s).

⟶5&6: Wind speed and direction determine wave growth on shallow water. Moreover, wind speed statistics are used in the method for estimation of joint statistics of waves and water levels employed until now by Rijkswaterstaat [e.g. *Bruinsma*, 1985].

An impression of the sensitivities of wave parameters nearshore computed by HISWA to wind, wave parameters and set-up at a deep-water boundary is given in [*Booij and Holthuijsen*, 1991].

⟶7&8: parameters related to spectral shape at deep water may be relevant for computation of wave propagation in shallow water. Near a construction, spectral width may be needed to compute wave height parameters other than H_{mo} (see definitions below) such as the maximum wave height over a time interval.

⟶9&10: current speed and direction are important for wave propagation in tidal inlets.

Definitions of parameters

The following notation will be used:
The symbols for parameters are placed between brackets (•), the units of parameters between square brackets [•].
The symbol $\triangleq$ means "is defined as".
For spectral parameters: in general, the notation is consistent with the IAHR list of sea state parameters [*IAHR*, 1986]. Among others,

f = frequency [Hz]
θ = direction of wave propagation [degrees]
$S(f,\theta)$ = directional spectrum (spectral density)

To simplify the notation, the inner product of two functions a and b on the spectral domain

$$\langle a,b\rangle \triangleq \iint a(f,\theta)\, b(f,\theta)\, df\, d\theta$$

will be employed frequently.

The definitions are:

1. High tide level (h) [m]
   maximum of the water level between two minima of the dominant (~12.5 hour) cycle of the astronomical tide.

related parameters:
1a. water level [m]:
   10-minute average of sea surface elevation relative to NAP

1b. Astronomical tide (a) [m]
sum of astronomical frequency components in water level signal (result of tidal analysis on long data records). Defined at every instant of time.

1c. Set-up (s) [m]:
difference between high tide level and the maximum of the astronomical tide between two minima of the dominant cycle (in Dutch: *schuine opzet*).

1d. Persistence of high tide p_h [tidal periods]:
length in no. of tidal periods over which the high tide level remains above the level h.

2. Significant wave height (H_{m0}) [m]

$$H_{m0} \triangleq 4(m_0)^{1/2}$$

m_0 [m^2] is the 20 minute average of the variance of the sea surface elevation (usually only for frequencies above a threshold, e.g. $(30s)^{-1}$). In general, m_0 is stored only every 3 hours (for example), but this is not part of the definition of the parameter, it pertain only to availability of the data.
The averaging interval is fixed here at 20 min. It can also be different in practice, usually within the range [10 min, 30 min].
m_0 is related to the spectrum by the definition of the i-th spectral moment

$$m_i \triangleq \langle f^i, S \rangle$$

related parameter:

2a. Persistance of significant wave height (p_H) [hours]:
length in hours of excursion of H_{m0} above a certain level.

3. Mean wave period ($T_{0,2}$) [s]

$$T_{0,2} \triangleq (m_0/m_2)^{1/2} \qquad \text{(see also 2.)}$$

alternative definitions of wave period which are also based on the spectrum are:

3a. $T_{0,1} \triangleq m_0/m_1$

3b. $T_{-1,0} \triangleq m_{-1}/m_0$

4. Mean wave direction (θ_0) [degrees]

$$\theta_0 = \mathrm{atan}_2(si, co)$$

with $si \triangleq \langle \sin\theta, S\rangle / \langle 1, S\rangle$ and $co \triangleq \langle \cos\theta, S\rangle / \langle 1, S\rangle$

(atan_2 is the mapping of sine and cosine to angle).

5. Wind speed (u_{10}) [ms^{-1}]
wind speed at 10 meters above the sea surface, averaged over 10 minutes.

6. Wind direction (θ_{10}) [degrees]
10 min. average direction of wind at 10 meters above sea surface.
Zero for wind vector pointing to the South, increasing with vector pointing toward S → W → N → E.

7. Spectral width parameter ν_2 [•]

$$\nu_2 \triangleq (m_0 m_2 m_1^{-2} - 1)^{1/2}$$

determines amplitude of complex envelope = one half of individual wave height in case of narrow band spectrum.
possibly useful for determination of other wave height parameters besides H_{m0}.

8. Directional spreading of the waves σ_0 [degrees]
or "circular standard deviation":

$$\sigma_0^2 \triangleq \frac{\langle 2(1-\cos(\theta-\theta_0)),\ S\rangle}{\langle 1,\ S\rangle}$$

In fact: $\sigma_0^2 = 2(1 - (a_1^2+b_1^2)^{1/2})$

with $a_1 \triangleq \langle \cos\theta,\ S\rangle / \langle 1,\ S\rangle$ and $b_1 \triangleq \langle \sin\theta,\ S\rangle / \langle 1,\ S\rangle$

9. Current speed (v) [ms^{-1}]
averaged over same interval as wave spectrum

10. Current direction (θ_v) [degrees]
averaged over same interval as wave spectrum

3.2 Listing of parameter combinations to be considered

In principle, a sufficient set of statistics would consist of a joint statistic of at least the first 6 parameters listed in 3.1. In practice it is impossible to assess such kind of statistic with reasonable accuracy. Therefore it is proposed to start with the following combinations of one or two parameters:

a. Univariate statistics of h, H_{m0}, $T_{0,2}$, θ_0, u_{10}, θ_{10}

and joint statistics of the pairs

b. high tide level h and significant wave height H_{m0}

c. significant wave height H_{m0} and mean wave direction θ_0

d. significant wave height H_{m0} and mean wave period $T_{0,2}$

e. wind speed u_{10} and wind direction θ_{10}

f. significant wave height H_{m0} and wind speed u_{10}

g. significant wave height H_{m0} and wind direction α_{10}

h. mean wave direction θ_0 and wind direction α_{10}

h	H_{m0}	$T_{0,2}$	θ_0	u_{10}	θ_{10}	
	×					h
		×	×	×	×	H_{m0}
						$T_{0,2}$
					×	θ_0
					×	u_{10}

The rationale behind this choice is that it is considered difficult enough to estimate joint statistics of two parameters; in the case of three parameters, it is expected that results will already be poor. When during execution of the study, it appears that it is possible to make meaningful estimates of statistics of more than two parameters, an attempt can be made, in the first place of

i. high tide level h, significant wave height H_{m0} and mean wave direction θ_0

and of

j. high tide level h, significant wave height H_{m0} and mean wave period $T_{0,2}$

To complete the joint statistics of all six parameters, simplifying assumptions like conditional independence can be tested, and applied if found to be appropriate. Physical arguments will be important in this process. Such assumptions are for example assumptions of complete dependence or independence. Also the sensitivities of the model that is used to translate parameters at the -20 m contour to parameters nearshore are important in determining which approximations are to be made.

Finally persistence statistics are important. In principle, we will focus on

k. persistence of high tide level

The reason for this choice is that a high water level is a necessary condition for other (wave) parameters to affect failure of a construction. Of course it is of interest to compare it with persistence of significant wave height. More detailed information on the behavior of water level and wave parameters as a function of time can be valuable for dynamical simulation of the behavior of a dune or dike during a storm. The results of such simulations can then be reduced to a more simple description of the behavior in terms of basic hydraulic parameters as mentioned in 3.1.

Of course the validity of the proposed restrictions on parameter combinations remains to be checked.

3.3 Summary of chapter 3

For statistics at the -20 m contour, it is proposed to focus on the following environmental parameters:

1. high tide level h
2. significant wave height H_{m0}
3. mean wave period $T_{0,2}$
4. mean wave direction θ_0
5. wind speed u_{10}
6. wind direction θ_{10}

In addition, spectral width and directional spreading of the waves, as well as current can be considered.

Joint statistics should pertain to the combination of all parameters 1-6. Estimation of these from data is not feasible in practice, so some simplifications are required. It is recommended to start with the following combinations of two parameters

(h,H_{m0}), $(H_{m0},T_{0,2})$, (H_{m0},θ_0), (H_{m0},u_{10}), (H_{m0},θ_{10}), (θ_0,θ_{10}), (u_{10},θ_{10})

and then to see if statistics of three parameters can be estimated for

(h,H_{m0},θ_0) and $(h,H_{m0},T_{0,2})$

The required joint statistics should be based on the results for combinations of not more than three parameters, using assumptions that have been thoroughly tested on available data.

4. Choice of statistics to be estimated

4.1 Probabilities and expected frequencies of occurrence of extreme conditions

Statistics of the frequencies or probabilities of occurrence of extreme (rare) events can take different forms.

The first starts with a subdivision of the time axis into equidistant intervals, say with a length of one year, and considers the probability of occurrence of an event during such an interval.
Usually it is assumed that these probabilities for different intervals of equal length are equal. This is valid if the process is seasonally stationary (as is the case for all phenomena related to the weather, when climate change is neglected), and the length of the interval is an integer number of years. Classical extreme value statistics deals with this kind of statistic.

Another kind of statistic is the expected number of occurrences per unit of time, for example the expected number of exceedances of a high level α per unit time. Again, this statistic can be converted to a probability if we consider only exceedances of a certain threshold ω, namely the probability of an exceedance of α during an excursion above the threshold ω.
The advantage of a threshold is that the statistic is not so much affected by small fluctuations, so it ensures that (generally) the maxima during different excursions above ω are independent, so only independent events are counted. The significance for statistics of hydraulic parameters is that a single storm does not show up several times in the statistics.
It may be even better to use two thresholds instead of one: a high threshold to mark the beginning of a storm event and a lower threshold to mark the end. This to make absolutely sure that small fluctuations do not affect the selection of storm events.
This idea of frequencies of occurrence and thresholds can be generalized straightforwardly to more than one parameter, e.g. high tide level and wave height. We will discuss this later.

In general, these two statistics are different. From both, a socalled *return period* can be computed, but these are different too:

For the first case, the probability of occurrence of an event in a fixed time-interval of length t, this is

$$r_t = P[k^{0,t} \geq 1]^{-1} t \qquad (1)$$

with t the length of the interval considered (say one year for example), and $k^{0,t}$ the number of occurrences of the extreme condition in the interval of length t. Apparently, this return period depends on t.

The return period in terms of expected number of occurrences per unit of time is simply μ^{-1}, with μ the expected number of exceedances of u per unit of time. Observe that

$$r_t \geq \mu^{-1} \tag{2}$$

When considering exceedance frequencies using a threshold w, the return period ρ_w is

$$\rho_w = \mu_w^{-1} \; P[k_w \geq 1]^{-1} \tag{3}$$

with k_w the number of exceedances of α during an arbitrary excursion above the threshold w, and μ_w the expected number of crossings of the threshold per unit of time.

Traditionally, the choice of statistic has been considered as a problem of choice of method, rather than as a choice of principle. In principle, the statistic should be used that is most relevant for the application. For example, if an oil production platform is designed which should be operational for a period of 30 years at most, than the user of the platform wants to know the probability that the construction will be damaged during its intended lifetime of 30 years. So the required statistic is r_t with t equal to 30 years.

If it is not clear a priori that a statistic should apply to a fixed finite time, then it does not make sense to chose an arbitrary value of t (such as one year) and compute a statistic. Rather, the expected number of upcrossings should be computed, since this is the statistic that is invariant to the choice of time-unit (the expected no. per 10 years is 10 times the expected no. per year). However, quite generally, provided μt is *small enough*, there is practically no difference between exceedance frequency and probability of exceedance within a time-interval of fixed length t (see for example [*Cramér and Leadbetter*, 1967], p.54). This applies to the problem of assessment of the safety of the coast. For example, suppose that upcrossings of a high level u are a realization of a Poisson process (in appendix b, weaker assumptions are given that are more appropriate for hydraulic parameters, but the result is the same). Let the time-interval t be fixed at 1 (one year), and let the frequency of failure be only 10^{-4} (yr^{-1}), then

$$P[k^1 = i] = (\mu)^i \exp(-\mu)/i! \simeq \mu^i/i! \qquad \forall i \tag{5}$$

which means that

$$\mu \simeq P[k^1 = 1] \simeq P[k^1 \geq 1] \tag{6}$$

so in this case, there is practically no difference between mean number of occurrences per year and the probability of at least one occurrence in a year. Frequencies of failure in the order of 10^{-4} (yr^{-1}) are typical values used in the design of coastal protection measures, so in the present context, there is essentially no difference between the two types of statistics.

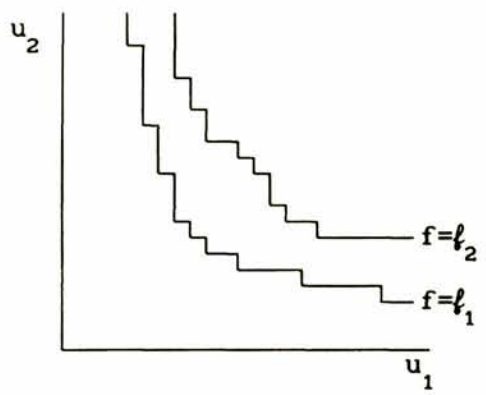
In appendix b, the mathematical relationship between these two different concepts of extreme condition statistics is discussed in some more detail, and the implications for parameterization of these statistics are reviewed.

4.2 How to define statistics of simultaneous occurrences of high tide levels and high waves

Suppose that the probability of failure of a construction during a storm is only determined by the most unfavorable condition occurring during that storm (so the duration of a storm is not relevant). Let's say that the total load on a construction is determined by two hydraulic parameters u_1 and u_2, such as significant wave height and high tide level. The total load at an instant t can be written as

$$f(u_1(t),u_2(t))$$

In this case, contours of equal total load in the parameter space can be determined, for example



The probability of failure PF in a time interval of fixed length, say [0,1] (for example one year) is the probability that the maximum of total load over [0,1],

$$m\ell = \max_{t\in[0,1]} f(u_1(t),u_2(t)) \tag{7a}$$

exceeds the strength of the construction, so PF is given by the convolution

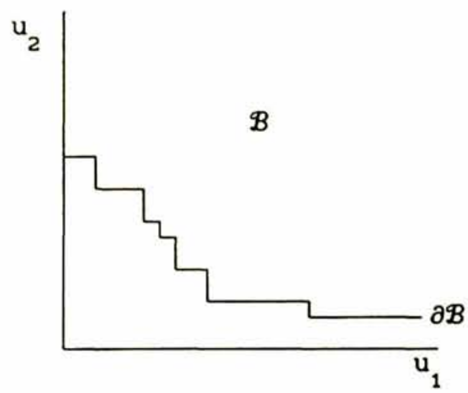
$$PF = \int_{s\in\mathbb{R}} P[m\ell > s]\ dH(s) \tag{7b}$$

with H the distribution function of the strength: strength is assumed statistically independent of load because it is the result of uncertainty about the properties of the construction.
By (7), the required statistic of the hydraulic parameters is:

$$P[m\ell > s] = P[\max_{t\in[0,1]} f(u_1(t),u_2(t)) > s] \quad (8)$$

This statistic depends on the failure mechanism. However, we need generic statistics, independent of failure mechanism, because these statistics will be determined at the -20 m contour, and then translated to the particular structure. Moreover, we want to avoid having to start the estimation of statistics from data all over again with every update of the model of the failure mechanisms. So what is needed is a statistic which is independent of failure mechanism and provides enough information to compute (8) for every relevant f.

In general, (8) above is computed from the probability that the load exceeds s during a single storm. Let a storm be defined as a time interval $I_{\mathcal{B}}$ during which (u_1,u_2) are in some region $\mathcal{B}$ in the plane which is far away from zero. In the case of wave height and high tide level, $\mathcal{B}$ contains only those wave height/high tide level combinations that occur during severe storms, so its boundary $\partial\mathcal{B}$ serves as a threshold. For example,



$\mathcal{B}$ should have such a shape that for all relevant failure mechanisms, all points in the plane corresponding to loads that may cause damage or failure are included in $\mathcal{B}$. In other words

$$\{ (a_1,a_2): f(a_1,a_2) > s \} \subset \mathcal{B} \quad (9)$$

for those values s of practical interest.
Now if the probability that the load f exceeds a value s during an arbitrary storm,

$$P[f(u_1(t),u_2(t)) > s \text{ for some } t \text{ in } I_{\mathcal{B}}] \quad (10)$$

is known, then the expected number of storms per year during which

$f(u_1(t),u_2(t))$ exceeds s can be computed by multiplying (10) with the expected number of storms per year. Then also (8) can be computed (see appendix b). In the present context, (8) is practically identical to (10) multiplied by the expected number of storms per year (see section 4.1 and also assumption (i) in appendix b). Observe that deriving statistics like (8) from (10) is similar to the approach called the POT (peak over threshold) method. In this case, the boundary of the set $\mathcal{B}$ acts as the threshold.

So for a particular type of construction, (10) is needed in order to compute statistics like (8). Under rather weak assumptions explained in appendix a, (10) can be computed for all relevant failure mechanisms from the statistic

$$\mathbf{F}_c(a_1,a_2) \triangleq \mathbf{P}[\ u_1(t)>a_1 \cap u_2(t)>a_2 \ \text{ for some t in } I_{\mathcal{B}}\ ] \tag{11a}$$

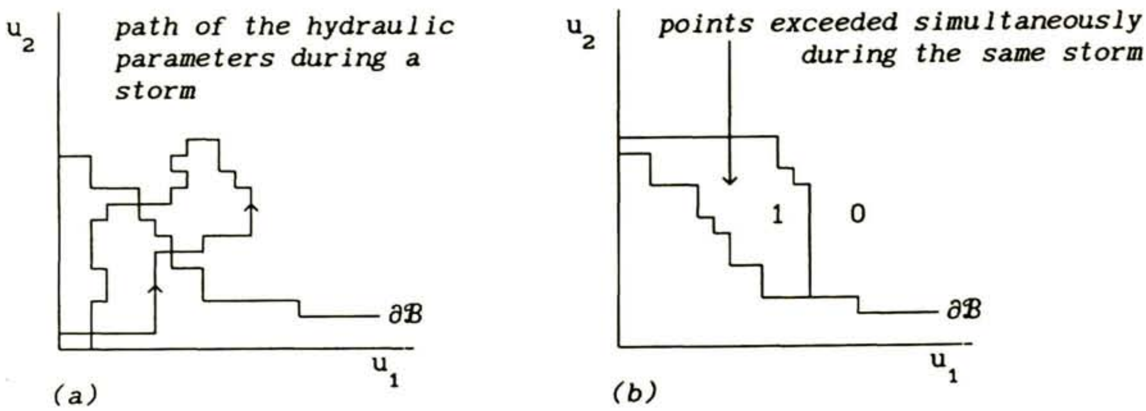
given as a function of a_1 and a_2, provided s is large enough for (9) to hold (see appendix a). The formula to compute (10) from (11a) is

$$\mathbf{P}[\ f(u_1(t),u_2(t)) > s \text{ for some t in } I_{\mathcal{B}}\ ] = \int_{\mathbb{R}^2} u[\ f(a_1,a_2)-s\ ]\ \ d\mathbf{F}_c(a_1,a_2) \tag{11b}$$

with u the unit step function defined as follows: $u(x)= 1$ if $x>0$ and $u(x)= 0$ otherwise. It can expected (11) is valid for all f of practical interest (see appendix a).

The function $\mathbf{F}_c$ is independent of failure mechanism. It can be determined at the -20 m contour; the probabilities of failure of a construction near the coast can be computed if we assume that the transformation from the -20 m contour to the coast is instantaneous (in view of 3.1, at least within a single interval between two tidal minima). Note that for waves, this assumption is implicit when the model HISWA [*Holthuijsen et al.*,1989] is used to translate wave parameters from deep to shallow water.

The meaning of F_c is explained graphically in the following diagram.



It shows a curve of the parameters u_1 and u_2 during a storm (see (a)) and the points that are simultaneously exceeded during that storm (see (b)). Defining a function which is unity at these points and zero everywhere else, then F_c is simply the expectation of this function.

Note that F_c need not to correspond to a proper probability measure on the plane, that is, its density may take negative values. However, this density is expected to be nonnegative in practice. In particular if u_1 and u_2 always reach their maxima at the same time, then

$$F_c(a_1,a_2) = P[\max_{t\in\mathcal{B}} u_1(t)>a_1 \cap \max_{t\in\mathcal{B}} u_2(t)>a_2] \tag{12}$$

the "joint survivor function" of the maxima of u_1 and u_2 during a storm. This makes the work easier in practice since we only have to deal with one observation per storm. (12) is however not essential. Another simple situation is that the hydraulic parameters are completely dependent. In that case, F_c in (11b) is nonzero only along a single curve in the plane. This simplifies the translation from one location to another considerably. For certain parameters like wave height and wave period this may turn out to be the case.

Generalization to the case of more than two hydraulic parameters is straightforward.

If one of the hydraulic parameters is a direction (mean wave direction, wind direction, current direction), then the approach as sketched above is not valid. However it can be modified to deal with directional data; statistics like the probability that during a (well-defined) storm H_{m0} exceeds a certain value while the mean wave direction θ_0 is in a particular interval can again be derived from a single statistic independent of failure mechanism. It will take us a little too far to discuss the technical aspects here.

Another more simple approach is to divide the circle into sectors of equal length and to consider for each sector the maximum of H_{m0} over the time in which θ_0 has been in that sector during a storm. In this way, the problem has been reduced to determining a univariate distribution for each directional sector. This type of statistic is not completely generic however, unless θ_0 remains in just one directional sector during a storm.

If failure depends on the time that a dune or dike has been under the attack of waves and surge, the approach sketched above will not be valid. There are two possible approaches in this case.
The first one is to specify completely the statistics of the time series of all relevant hydraulic parameters during a storm. Then a dynamical model computing the response of the structure can be run with all possible time series of these hydraulic parameters as input data to compute the statistics of failure or damage. This is only feasible if rather strict assumptions are imposed on the variation of the hydraulic parameters during a storm in order to simplify it.
The other approach is use the model first to find out exactly which characteristics of the time-series are essential for determining failure or damage. This results in a simplified description of the model's response in terms of a few parameters, which may include persistence. This description can be combined with the statistics of these parameters to obtain a probability of failure or damage.
The advantage of the second approach is that the results do not depend so much on assumptions on the variation of the hydraulic parameters during a storm. Moreover, the effort can be spent to obtain reliable statistics of the essential parameters rather than to figure out how to simplify a set-up curve, for example.
However, more detailed insight into the temporal variation of hydraulic parameters during a storm can be valuable in order to provide representative inputs for model simulations.

In this section, it has been shown that a rigorous definition of joint extreme condition statistics is possible in practically all circumstances. Elsewhere, parameterization (section 5.3) and estimation (appendix d) of these statistics is discussed; it appears that some modification of statistical methods is necessary. If the situation is relatively simple, that is, when the different hydraulic parameters reach their maxima during a storm at the same time, statistics like (12) are suitable and standard statistical methods can be used. The recommended strategy is therefore to see first if statistics like (12) are applicable. If not, then this section provides guidelines to obtain results without having to resort to ad hoc methods.

4.3 Dealing with the limited accuracies of estimates

Normal procedure in assessment of statistics of hydraulic parameters is to give an estimate of a probability (e.g. (9)) or expected frequency of occurrence, together with an indication of the accuracy of the estimate, usually a confidence interval. The engineer is then left with the problem what to do with this confidence interval, that corresponds to a statement like: "there is a probability of 0.05 that the level exceeded by the annual maximum wave height with a probability of 10^{-4} is higher than 6 m".
In this sentence, the word probability refers in the first instance to uncertainty due to the limited number of data with limited accuracies, and in the second instance to the unpredictable character of the weather in an arbitrary year in the future. Yet, for the engineer, the sources of these uncertainties do not matter; all that matters is the final uncertainty about whether or not the construction will fail, based on the information that is available.

This final estimate of a probability, including the uncertainty due to limitations on number and accuracies of data, is the conditional probability *relative to* the available data (or: conditioned on the available data). It is explained in detail in appendix e.
Normally, to estimate the probability that say wave height will exceed a level α in an arbitrary year, the data, for example of annual maximum wave height, are fitted by a distribution function of a certain shape, by estimating the parameters of this distribution by the maximum likelyhood method. The conditional probability relative to the data, however, is a weighted average of the probabilities obtained with different values of the parameters, reflecting the uncertainty in these parameters due to limitations on number and quality of the data. This implies that if only few inaccurate data are available, the probability of exceedance of a level tends to be high. By increasing the number and accuracy of the data, this probability may become lower (unless the values of the added data indicate otherwise).

This type of estimate is also quite robust to errors in assumptions on the general shape of the distribution: if the data are not fitted well by any distribution within the class considered, e.g. the Generalized Pareto distributions, then relatively high probabilities of exceedance result. This is a very important property of an estimator: it corrects errors in assumptions to a certain extend. In estimation of statistics of extreme conditions, the objective is to extrapolate a distribution beyond the range of available data. This implies that the selected class of distributions chosen should in the first place be suitable for the extrapolation. However another class of distributions may give a better fit to the data, for example because the data set is not representative, but may be unsuitable on theoretical grounds. If just the quality of fit is considered, then the second class of distributions is preferred. However it gives relatively low probabilities of exceedance because apparently the uncertainty associated to data volume and quality is small, whereas the first class, due to poor fit, leads to relatively high estimates. A cautious engineer will regard the low probabilities of exceedance with some suspicion.

Other properties of conditional estimates relative to the data are that data from different sources can be weighted according to their error statistics, and that prior knowledge can be incorporated straightforwardly.
An error analysis of the available data is required to assess the statistics of these data. Water level measurements can probably be regarded as error free: they are very accurate compared to other data. This may also be assumed for the astronomical tide; the accuracy of the tidal analyses needs to be checked. So errors in measurements are mainly limited to wave and wind parameters. The error statistics of wave parameters depend in general on spectral shape, and can be approximated. Other errors such as for example interpolation errors may be considered too. Hindcast errors should be determined by comparison with measurements from sensors with known error characteristics, taking the measurement errors into account. So there is a strong link between data validation and the estimation of statistics.

It is recommended to address the issues discussed in this section only after a satisfactory basic method for estimation of joint extreme condition statistics has been developed. In the mean time, the maximum likelyhood estimator (or equivalent methods) can be used.

4.4 Summary of chapter 4

The statistics that are ultimately required are statistics of a function of one or several hydraulic parameters near a water retaining structure. This function is the total load (in case failure is not dependent on the duration of a storm), or some equivalent which is also a function of persistence (see section 4.2). It is determined by the failure mechanisms.
In section 4.1, we found that these statistics can be the probability that the total load exceeds a level in an arbitrary year, or the expected number of exceedances of this level per year. In the present context, there is no difference between these two statistics.
In section 4.2, a method is presented to compute the probability that the total load exceeds a level s in an arbitrary year from a statistic which is *independent of failure mechanism*, given in equation (11a). Therefore (11a) is the statistic to be estimated at the -20 m contour and then translated to the coast.
Usually extreme condition statistics are made by fitting some curve to the available data. To give an impression of the accuracy of this estimate, a confidence interval is estimated also, which depends on the number and quality of the data. An alternative approach is described which incorporates the uncertainty due to limitations on sample size and accuracies into the final estimate of a probability. It is proposed to pursue this subject after a satisfactory basic method for estimation of joint extreme condition statistics has been implemented.

5. Parameterizations of joint statistics of waves and high tide levels

5.1 Statistics of high tide levels

In this section, a brief summary is given of the methods used to determine the statistics of high tide levels. This is important because the joint statistics of waves and high tide level will have to be consistent with the accepted approach to statistics of high tide levels.

Until recently, water level statistics were determined according to the Report of the Delta Committee [*Delta Committee*; 1960]. The method is as follows. The parameter of interest is high tide level (see section 3.2). No decomposition into set-up and astronomical tide is made. The statistic considered is the expected frequency of exceedance (in no. of exceedances per year) as a function of level relative to NAP. The shape of the exceedance frequency curve was assumed exponential (the logarithm of exceedance frequency is a linear function of level). This choice was made after comparison with some alternatives. Data are selected high tide levels at Hoek van Holland observed over the years 1888 to 1956. From the high tide levels, the maximum per storm was selected. Only storms during the months november, december and january were used, and of these, only the potentially dangerous storms were retained, as determined by the path followed by the pressure low. The purpose of the selection was to obtain a data set which can be expected to be statistically homogeneous, in the sense that the data are taken from the same distribution.

A recently developed approach differs in several aspects [*De Haan*, 1990; *Van der Made et al.*,1989; *Dillingh*, 1991]. The method of selection of 'potentially dangerous' storms is now rejected. Still only data collected in a particular season are used, now october 1 to march 15, in order to obtain a homogeneous data set. Only data of storms with set-up levels exceeding a threshold of 0.3 m above NAP are retained. Also, only data corresponding to different storm events are selected using a time window.
Statistics of set-up and astronomical high tide are estimated separately and then combined to statistics of high tide level instead of estimating statistics of high tide level directly from high tide level data. Astronomical tide and set-up are processes with very different statistical properties and can be regarded as statistically independent, so statistics of high tide level are easily derived from the statistics of astronomical tide and set-up.
This method is expected to yield more accurate statistics of high tide level than can be obtained from the high tide level data directly.
In the most recent method [*Dillingh*, 1991], the conditional probability of the maximum tide level during a storm given that it exceeds some level (the average 5 times per year exceeded level) is estimated according to the following four methods (see [*De Haan*, 1991] for an exposition of background and terminology):

[1] High tide level: moment estimator for high quantiles [*De Haan*, 1991]

(γ free, with γ a real number determining the curvature of the logarithm of exceedance frequency as a function of level, see also appendix c)
Confidence bands analytical (asymptotic expression)

[2] High tide level: moment estimator for high quantiles [*De Haan*, 1991] (γ fixed)
Confidence bands analytical (asymptotic expression)

[3] Set-up: Generalized Pareto distribution, estimated by Maximum Likelyhood.
Astronomical tide: distribution estimated by simulation.
High tide level: by convolution of distributions of set-up and astronomical tide.
Confidence bands by Monte Carlo simulation

[4] High tide level: Generalized Pareto distribution, estimated by Maximum Likelyhood directly from tide level data.
Confidence bands analytical (asymptotic)

Also, the distribution of annual maxima of high tide level has been estimated using the Generalized Extreme Value distribution.
All these methods yield about the same exceedance frequencies, but [3] gives a relatively narrow confidence region (probably due to the fact that the distribution of astronomical tide is assumed completely known, which is not unreasonable). The method [2] yields an even narrower confidence band, but this is not reliable because the uncertainty in γ has been ignored. This means that [1] and [3] are probably most accurate, with [3] probably giving the narrowest, yet reliable, confidence bands.

5.2 Parameterizations based on simplified physics (Bruinsma/Van Aalst)

Around 1980, Bruinsma and Van Aalst, working for Rijkswaterstaat, developed a method for assessment of joint statistics of significant wave height and high tide level. This method is based on the idea that surges and waves have a common source, the wind fields on the Atlantic Ocean and North Sea [*Bruinsma*, 1982; *Van Aalst*, 1983]

Definition of statistics

Result is a conditional distribution of significant wave height H_s, relative to the maximum high tide level during a storm. The exact definition of H_s is not specified. In particular, it is not clear weather it the H_s at the time of occurrence of the maximum high tide level during a storm, or maybe the maximum H_s during the excursion above a high level during a storm. Such ambiguities are also found for other parameters used in the computation (see below).

Computation of statistics

Following [*Bruinsma*, 1985], the first step is to determine the conditional distribution function of local wind speed relative to high tide level. This distribution function is determined by making use of estimates of the range of magnitudes c of the ratio of wind set-up s and the square of the local wind speed,

$$s = cu_{10}^2 \qquad (16)$$

and the observation that relatively high values of high tide level are found during spring tide.
Significant wave height is also assumed to depend on local wind speed, besides wind duration (or fetch), depth, and wave energy propagating from directions other than the wind direction (swell), by

$$H_s = \left(|g_{br}(d,F,u_{10})|^2 + |H_s^{sw}|^2 \right)^{1/2} \qquad (17)$$

with

$g_{br}(d,F,u_{10})$	the Brettschneider formula for significant wave height of the wind-sea
d	depth (actual, including set-up and astronomical tide)
F	fetch
H_s^{sw}	swell wave height

The fetch is assumed to be related to wind speed and wind duration τ by

$$F(\tau, u_{10}) = \tilde{F}|u_{10}|^2/g \qquad \tilde{F} = 60\tilde{\tau}^{10/7} \qquad \tilde{\tau} = g\tau/u_{10} \tag{18}$$

and τ is assumed independent of wind speed and set-up (or rather in physical terms: set-up is assumed not to depend on wind duration), and is assumed to have a lognormal distribution.

Now from the given distribution of wind duration, the distribution of swell wave height, and the conditional distribution of local wind speed relative to high tide level, the conditional distribution of H_s relative to high tide level h can be computed by straightforward integration.

The conditional distribution of wind speed relative to high tide level $F[u_{10}|h]$ is assumed Weibull, and defined for $u_{10} \geq u_{min}$ by

$$f(u_{10}|h) \triangleq \partial F[u_{10}|h]/\partial u_{10} =$$

$$(a-1)\,\frac{(u_{10}-u_{min})^{a-1}}{(u_{top}-u_{min})^{a}}\,\exp\left[\frac{(1-a)}{a}\,\frac{(u_{10}-u_{min})^{a}}{(u_{top}-u_{min})^{a}}\right] \tag{19}$$

with a = 2.6, and u_{min} and u_{top} the minimal wind speed and the windspeed corresponding to the mode of $f(u_{10}|h)$, respectively. They are computed as

$$u_{top} = [\ (h-a_{gem})/c_{gem}\ ]^{1/2} \tag{20a}$$

$$u_{min} = [\ (h-a_{spr})/c_{max}\ ]^{1/2} \tag{20b}$$

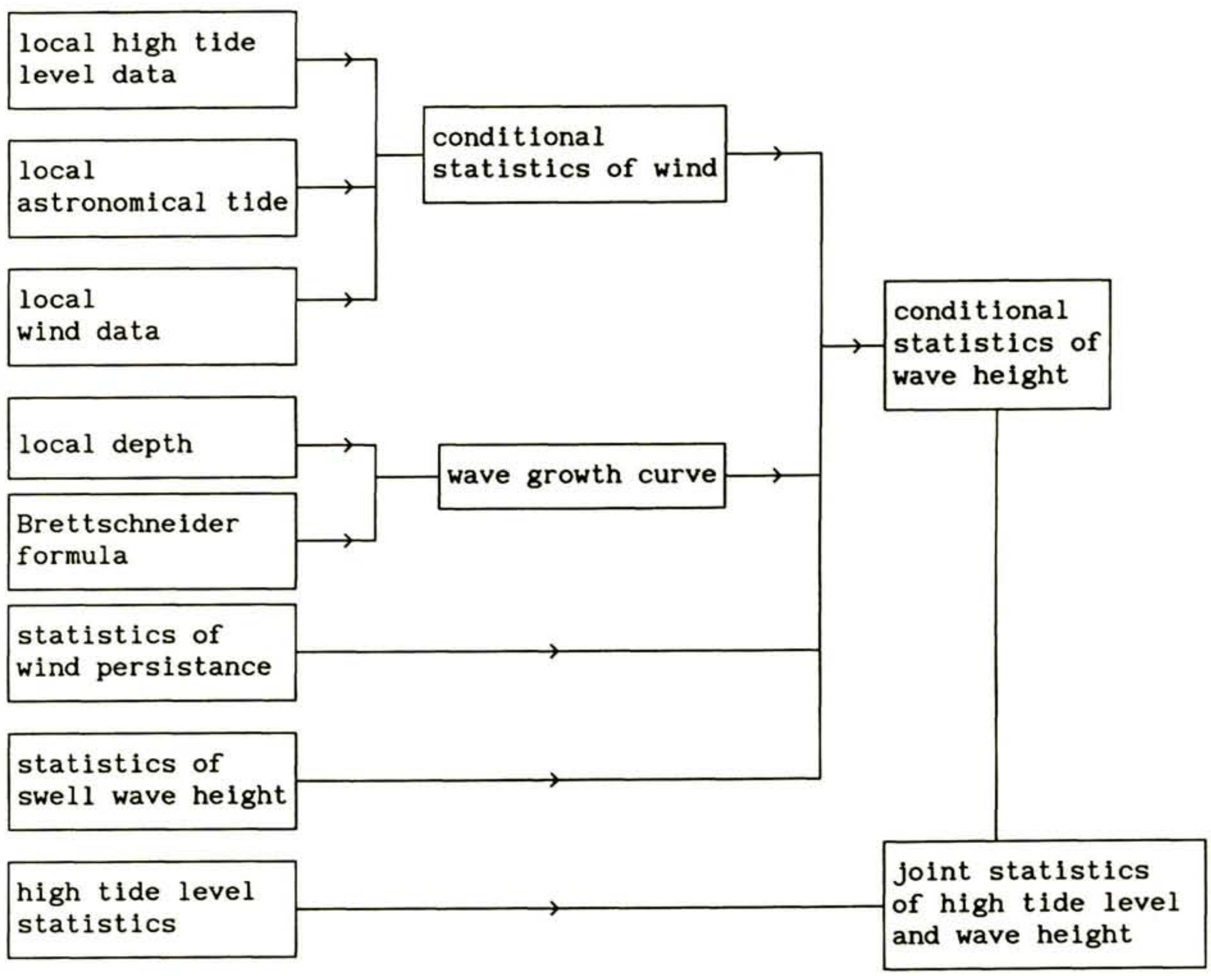
with

a_{spr} = maximal value of astronomical high tide

a_{gem} = mode of distribution of astronomical high tide

c^{gem} and c^{max} are location-dependent parameters: $c^{gem} = 0.5\ c^{max}$, and c^{max} is determined as the maximal value of $s|u_{10}|^{-2}$, with s the local set-up level.

summary of dependences:



Motivation for this approach: surges and waves have a common source, the wind fields on the Atlantic Ocean and North Sea. Therefore, to determine the conditional distribution function of H_s relative to the maximum high tide level during a storm, the fact can be used that a certain wind speed is required to generate a set-up (or that the wind speed cannot have exceeded a certain value otherwise the observed set-up level should have been higher). This has led to the use of a conditional distribution of wind speed relative to the high tide level, and a model to compute significant wave height from wind speed. This approach is extended somewhat, resulting in the formulas given above. Some more comments about the approach follow here.

1°

All variables are conditioned on the maximum high tide level during a storm, h. This parameter is the sum of astronomical tide and set-up. At the -20 m contour, the relationship between waves and astronomical tide is due to the effect of depth on wave dissipation and to wave/current interaction. The relationship between waves and set-up is mainly due to their common source, wind fields, and probably less to dissipation and wave/current interaction. It may be worthwhile to try to model these relationships separately, using the distributions of set-up and of astronomical tide instead of one distribution of high tide level.

2^o

It is reasonable to assume that the maximum local wind speed (or the during 9 hours continuously exceeded wind speed) has both an upper and a lower bound for a given set-up s or high tide level h. In the Weibull distribution (??), only a lower bound for the wind speed u_{min} is assumed, but no upper bound. This may be too conservative if there is indeed an upper bound. To some extend, including wind direction explicitly may reduce the uncertainty somewhat (that is, if it has not already been assumed implicitly that the wind should be North-West).

3^o

Both set-up and waves are related to the *local* wind speed, which is of course an approximation. To account for the case that the local wind speed is lower than the effective wind speed along the fetch (in the direction of the peak of the wave spectrum), socalled 'swell' has been introduced. This is probably not swell in the sense of wave energy travelling from a distant storm, since in the extreme storm events anticipated, the wave spectrum will be a typical wind sea spectrum. So it might be understood to take deviations of the local wind speed from the effective wind speed along the fetch into account. Possibly the swell term has been included after having observed that the wave height exceeded the Brettschneider curves in certain cases.
However, the local wind speed as derived from the set-up (i.e. by the conditional distribution of local wind speed given the set-up) is in fact also an effective wind speed, although 'effective' may have a different meaning for set-up than for waves. Moreover, a deviation of 'effective' wind speed from the local wind speed can be incorporated in the conditional distribution of the local wind speed as well, instead of adding a separate 'swell' term.

4^o

Brettschneider's curve is not valid for varying wind fields. For the North-Western storms causing high set-up levels at the coast of the Netherlands, this may not be a big problem.
Also, it is questionable that shallow water wave growth can be modeled universally by including scaled depth

$$\tilde{d} \triangleq dg|u_{10}|^{-2} \tag{21}$$

as an independent variable in addition to scaled fetch (or time), if depth varies along the fetch.

5^o

Wind duration and set-up level have been assumed independent but there should be a relationship, just as for wind duration and wave height.

6^o

The method of Bruinsma and Van Aalst is based on scaling rules, both for the waves and for set-up (or tide level). The idea behind scaling is that (statistical) relationships should in the first place be determined among the nondimensional variables (in which case there is a rather clear relationship expected). Then the statistics of the independent variables used to scale the other variables are taken into account.
Scaling using local wind speed has indeed been proven very useful when the wind is known, as in experimental studies of wave growth. However for joint statistics of wave height and tide level, we still need to plug in the statistics of the wind at the end, which is uncertain just as the statistics of wave height itself. So it is questionable whether scaling with the local wind still offers a substantial gain. In fact, it may be even more effective to replace the local wind-speed by the square root of set-up multiplied by a random number, and then calibrate all distributions required directly from data of wave height, set-up and astronomical high tide level.

As already indicated, there are a number of variations possible on the approach of Bruinsma and Van Aalst. However, this does not imply that significant improvements are still possible. It is not expected that the method of Bruinsma and Van Aalst underestimates the wave height for a given high-tide level since the approximations were chosen rather carefully in order to obtain conservative estimates. However, with the new data sets available now, at least some validation is possible, and more insight may be obtained as to what choices to make.

5.3 Simultaneous exceedance statistics: asymptotic shapes and estimation

Consider the case that statistics of failure are made for a particular construction, assuming a certain failure mechanism in which total load f is a function of the hydraulic parameters u_1 and u_2, as in section 4.2. Then the statistics of interest may be

$$P[\max_{t\in[0,\tau]} f(u_1(t),u_2(t)) > s] \tag{22}$$

for example for τ equal to one year. In this case, we have to deal with only a single parameter f as a function of time rather than with several parameters, due to the fact that the failure mechanism is specified. The data can be converted to data of total load, and usual methods for estimation of statistics of the form (22) can be applied. For example, we can focus on the probability that the load exceeds the value s during a storm, i.e.

$$P[\max_{t\in I_{\omega}} f(u_1(t),u_2(t)) > s] \tag{23}$$

with I_{ω} an arbitrary interval over which the load f exceeds the threshold ω serving to distinguish storms. Extreme value theory suggests to parameterize (23) by the generalized Pareto (GP) distribution. Then its parameters (or quantiles) can be estimated from the data and from the resulting distribution (23), (22) can be computed. Assumptions underlying this approach (called the peak-over-threshold (POT) method) are given in appendices b and the end of appendix c. See in particular *De Haan*(1990) for a recent application to high tide levels employing a new estimation technique.

In case the statistics to be produced should be applicable to all possible failure mechanisms, the situation is not that simple, as was already observed in section 4.2. In that case, the type of statistics to be produced is of the form (11a) in section 4.2:

$$F_c(a_1,a_2) = P[u_1(t)>a_1 \cap u_2(t)>a_2 \text{ for some t in } I_{\mathcal{B}}] \tag{24}$$

with $\mathcal{B}$ a region of the plane far away from zero, used to select and to distinguish storm events; the boundary of $\mathcal{B}$ serves the same purpose as the threshold ω in (23) above. In 4.2, it has already been explained how to apply (24).

It is essentially a multivariate statistic, as opposed to (23). In this section, the parameterization of (24) will be discussed. Parameterization means that the shape of (24) will be inferred from available knowledge or assumptions that are considered reasonable. The assumptions on u_1 and u_2 are given in detail in appendix b (equations (b3) and (b4)) and in appendix c (see (c2)). They can be described in words as follows.

(1) u_1 and u_2 are seasonal processes. This means that (taking the length of the seasonal cycle equal to 1, so one year) every statistic of the processes is the same when applied to the processes shifted over one or more years.

(2) Consider a region in the plane bounded by thresholds ω_1 and ω_2 as

$$\{ (\mathrm{a}_1,\mathrm{a}_2):\ \mathrm{a}_1>\omega_1 \cap \mathrm{a}_2>\omega_2 \} \tag{25a}$$

Again, a (randomly chosen) time-interval

$$I_{(\omega_1,\omega_2)} \tag{25b}$$

in which $(\mathrm{u}_1(t),\mathrm{u}_2(t))$ is continuously in (25a) can be called a storm. So the region (25a) has essentially the same function as the set $\mathcal{B}$ in (24). The assumption is that the probability of more than one storm (25b) in a fixed number of years will decrease more rapidly than the mean number of storms per year when ω_1 and/or ω_2 are increased (this assumption ensures that a condition similar to (6) in section 4.1 holds).

(3) Moreover, the probability that during a randomly picked storm (25b) in a fixed number of years certain even higher levels are exceeded simultaneously is practically independent of the number of storms over these years, for sufficiently high ω_1 and/or ω_2. This implies a lack of dependence between actual storm frequency and intensity during a randomly chosen storm in a period of an integer number of years.

(4) The probability that during a randomly chosen storm (25b) even higher levels are exceeded simultaneously by u_1 and u_2 converges in some sense, explained in detail in appendix c (see (c2)), to some function $\mathcal{F}_c$.

Under these assumptions, we have the following result (see appendix c, treating the case of an arbitrary number of variables):

The variables u_1 and u_2 can each be scaled as

$$u_i(t) = \gamma_i^{-1}\ln(\gamma_i \mathrm{u}_i(t)+1) \tag{27a}$$

with some fixed real numbers γ_1 and γ_2 , such that for each pair (a_1,a_2)

$P[\ u_1(t) > a_1 + \varepsilon\vartheta \cap u_2(t) > a_2 + (1-\varepsilon)\vartheta$ for some t in $\mathcal{I}_{(\varepsilon\vartheta,(1-\varepsilon))}]$

$$\longrightarrow \mathfrak{F}_c(a_1,a_2) \qquad \text{with } \vartheta \longrightarrow \infty \tag{27b}$$

with

$$\varepsilon = a_1/(a_1+a_2) \tag{27c}$$

and $\mathcal{I}_{(..)}$ defined similarly as $I_{(..)}$ in (25b) but now for the scaled processes u_1 and u_2.

The function $\mathfrak{F}_c$ in (27b) is of the form

$$\mathfrak{F}_c(a_1,a_2) = \exp(-(a_1+a_2)\phi_\varepsilon) \tag{27d}$$

with ϕ some nonnegative function on [0,1].
The result is stated more precisely and proved in appendix c. See also the following diagram.

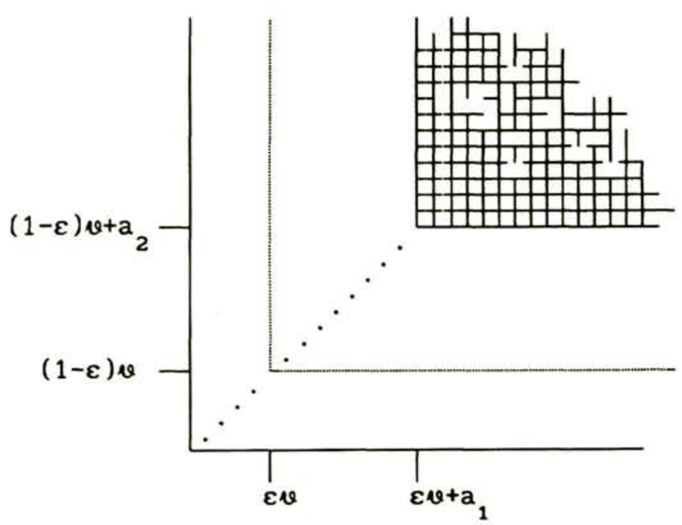


figure: illustration of (29).

What this result means is that after applying the appropriate scaling (27a) to the variables, the probability of reaching the checkered region in the figure will eventually decrease exponentially when this region is moved away along a straight line through the origin. The exponent depends on the direction of this line, and is given by the function ϕ.

In appendix c, it is shown that the number γ_1 in the scaling (27a) is a property of the process u_1 and is completely independent of the other process This can be seen by taking ε either equal to zero or equal to unity: the distributions of u_1 and u_2 converge to exponential distributions, implying that the asymptotic distributions of the original variables u_1 and u_2 are generalized Pareto distributions (see appnedix c).
The constants γ_1 in (27a) are most likely nonpositive for all hydraulic parameters. For high tide levels or set-up levels, the value 0 seems the most likely value both from data; a small negative value corresponding to a very high absolute maximum on high tide level or set-up might be acceptable too. In case γ_1 equals zero, the transformation (28a) reduces to the identity. For waves, a negative value of γ_1 is likely to come out of the data analysis as found in studies as for example [*Muir and Al-Shaarawi*, 1986]. There is a discussion whether nonzero values of γ_1 are acceptable, as negative values correspond to the existance of a maximum on the possble values of the hydraulic parameter.

To show how (27) can be used, an approximate expression for the statistic (24) in terms of the *scaled* parameters u_1 and u_2 and for a region $\mathcal{B}'$ of rather general shape is

$$P[\ u_1(t)>a_1 \cap u_2(t)>a_2 \text{ for some t in } \mathcal{I}_{\mathcal{B}'}\ ] \approx$$

$$r_{\varepsilon,\mathcal{B}'} \exp(-(a_1+a_2)\phi_\varepsilon) \qquad (28a)$$

with ε as in (27c),

$$r_{\varepsilon,\mathcal{B}'} \stackrel{\Delta}{=} \exp(\vartheta\phi_\varepsilon)\ P[\ u_1(t)>\varepsilon\vartheta \cap u_2(t)>(1-\varepsilon)\vartheta \text{ for some t in } I_{\mathcal{B}'}\ ] \qquad (28b)$$

and ϑ is chosen such that $(\varepsilon\vartheta,(1-\varepsilon)\vartheta)$ is on the boundary $\partial\mathcal{B}'$ of $\mathcal{B}'$.

This is illustrated in the following figure:

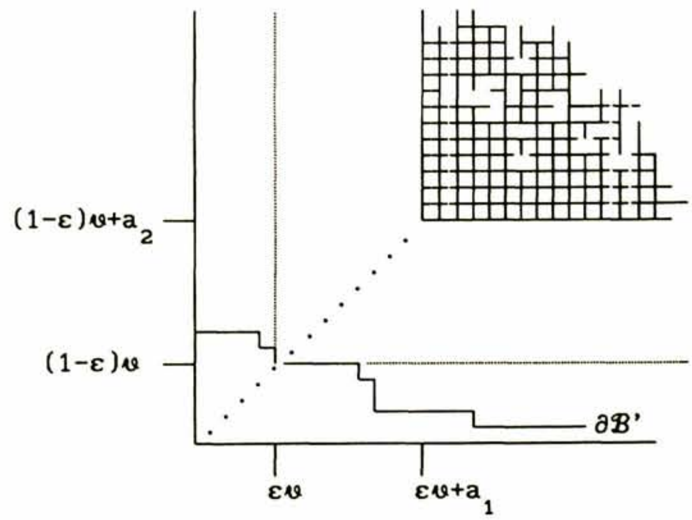


figure: illustration of (28).

The second factor on the right-hand side of (28b) is the probability of entering the region above the dotted line during a storm. The first factor is the inverse of what the second factor should be if the model (27) would hold in the entire plane and not just far away from zero. This second factor is not much smaller than unity, so it can be estimated from the data easily. The extrapolation is determined mainly by the exponent ϕ_ε, so this is the really critical parameter.
(28b) is easily converted to same statistic for the original variables u_1 and u_2 by transforming a_1,a_2 and the region $\mathcal{B}'$.

A procedure to estimate the statistics required at a coastal location from data at the -20 m contour from data only might look like this:

1° Data selection
Wave/high tide level data at the -20 m contour. Selected data are time-series of all relevant parameters during the "storm", defined as the interval in which the parameters are in a certain set. This approach is basically the POT method.
Simple approaches are to use a threshold for set-up to select and distinguish storm events, or to use a combination of high tide level and wave height. Both approaches suppress the effect of astronomical tide on data selection. At this stage, it should be decided whether the data can be reduced to a single point for each storm or not (see also 4.2).

2° Marginal distributions (including the scaling (27a))
First the marginals of (27) are estimated for the original variables. These are generalized Pareto distributions, which is consistent with the current method for statistics of high tide levels (see section 5.1). This determines also the scaling (27a) for each variable. Of course the agreement of the results for high tide levels or set-up with the statistics of high tide levels or set-up on a larger data set as described in [*De Haan*, 1990] and [*Dillingh*

et al, 1991] should be checked.

3° Estimation of parameters in (27)
This can be done in two ways: either all parameters are estimated from the original data, or the scaling (27a) is applied first (based on 2°) and then ϕ is estimated.
A rather straightforward method to estimate parameters is the maximum likelyhood method as for example in [*Tawn*, 1988] (see appendix d), which can be used if the data can be reduced to a single point for each storm. If not, then the modification indicated in equation (d8) of appendix d can be used. Experiments with different thresholds for data selection (from the data set already selected in step 1°) are carried out to check the sensitivity and stability.
If everything works fine, Bayesian estimates as described in 4.3 and appendix e may be computed or approximated.

4° Computation of the statistic (24) at the -20 m contour. This amounts to a computation similar to (28). It is rather straightforward.

5° Presentation and archiving of the joint statistics at the -20 m contour.

6° Translation of the statistics to coastal locations.
This will be reasonably straightforward if the transformation to the coast can simply be viewed as an instantaneous transformation of parameters at the -20 m contour to parameters at the coastal location (note that this is the case for wave parameters when the model HISWA [*Holthuijsen et al.*,1989] is used). This transformation should somehow be simplified, based on numerical model studies. The accuracy of this simplification may have consequences for the final statistics.
To translate joint statistics of wave height and high tide level (e.g. $\mathbf{F}_c$ in (24)), first the curve of highest density of $\mathbf{F}_c$ may be translated, for a start. Then it is early enough to decide whether other points need to be translated too. This is similar to translating curves of H_s versus high tide level of the form given in [*Technische Adviescommissie voor de Waterkeringen*, 1984]. Again results should be presented graphically and numerically.

7° Computation of statistics for specific types of constructions. See section 4.2.

8° Validation. In particular: comparison of the results with those of the method of Bruinsma/Van Aalst (section 5.2).

A final remark concerns the relationship with existing theory of multivariate extremes. It should be noted that the parameterization (27) cannot be compared to the parameterization of a joint distribution as described in [*De Haan*, 1990], basically because it applies to a different kind of joint statistic: the statistic considered in [*De Haan*, 1990; the right-hand side of his equation (24)] would in the present context be something like

$$P[\ u_1(t)>a_1 \cup u_2(t)>a_2 \ \text{ for some } t\in I_{\mathcal{B}}\ ] \tag{29}$$

(note the symbol $\cup$ denoting "and/or"), as opposed to (24) (with $\cap$ denoting "and"). The parameterizations of (29) and (24), although following by similar arguments, are different: the tail behavior of (29) is largely determined by its marginals, whereas for (24), there is more freedom which should be modeled. The differences are explained in detail in appendix f. Moreover, the approach differs also from e.g. [*Leadbetter et al.*,1983]: there, continuous-time processes are transformed to discrete-time sequences by taking maxima over intervals. This is not possible when considering simultaneous exceedances as in this section. Moreover, the assumptions invoked in [*Leadbetter et al.*, 1983] are less direct than the assumptions given in this report, but seem not very relevant to those interested in applications, since they can hardly be verified in practice. It seems easier to test assumptions that are stated directly in terms of the point process of occurrence of some extreme event, as in this report.

5.4 Summary of chapter 5

Statistics of high tide levels are currently based on data selection using a threshold on set-up levels and the use of the Generalized Pareto Distribution or the (related) moment estimator for set-up or directly for high tide levels. It is recommended to follow for joint wave/water level statistics an approach which is consistent with the methods for high tide level.

Two approaches to parameterization of the joint statistics of significant wave height and high tide level were discussed.

The first one (section 5.2) describes H_s as a function of local wind speed and fetch, and h as a function of local wind speed. Statistics of wind speed and fetch then determine the joint statistics of wave height and high tide level, so the joint statistics of wave height and high tide level can be computed from the conditional statistics of wind speed and fetch length for given high tide level. New sets of measurements and hindcast data are available now which may be used to validate (or maybe improve) the assumptions underlying the method. Some suggestions for possible improvement are given. An important point is to assess the sensitivity of the statistics to assumptions, such as the models of wave growth and set-up generation, and the distributions of wind speed and fetch.

A quite different approach to parameterization of this far region of the parameter space is presented in section 5.3. It is directly related to the approach to statistics of environmental conditions at the -20 m contour given in 4.2. The essence of the method is to estimate the required statistics directly from data, using an asymptotic shape derived for the probabilities of simultaneous exceedance of high levels during a storm. This is essentially the same approach as led in the univariate case to the POT method for extreme value statistics, using the Generalized Pareto distribution or related approaches such as [*de Haan*, 1990], as also used for high tide levels.

These two approaches seem completely different, which is only an advantage at this stage. There may also be ways to combine them and make use of the valuable aspects of both methods, but this should not be tried before the results both methods have been thoroughly checked and compared. For example, information about physical bounds of the set of possible wave height/high tide level combinations can be used in the statistical approach. On the other hand, more rigorous definitions and statistical methods may be used in the context of the method of Bruinsma and Van Aalst. However, combining different approaches is still a matter of speculation. The first things to do after data collection and validation is to tests both approaches on data.

6. Conclusions and recommendations

1. For statistics at the -20 m contour, it is proposed to focus on the parameters

1. high tide level h
2. significant wave height H_{m0}
3. mean wave period $T_{0,2}$
4. mean wave direction θ_0
5. wind speed u_{10}
6. wind direction θ_{10}

In addition, spectral width and directional spreading of the waves and current can be considered.
In principle, joint statistics of all parameters 1 to 6 above should be produced. Some simplifications are required: first bivariate statistics of the combinations

(h,H_{m0}), $(H_{m0},T_{0,2})$, (H_{m0},θ_0), (H_{m0},u_{10}), (H_{m0},θ_{10}), (θ_0,θ_{10}), (u_{10},θ_{10})

will be estimated. Then an attempt will be done to estimate the combinations

(h,H_{m0},θ_0) and $(h,H_{m0},T_{0,2})$

Assumptions need to be formulated and tested to derive additional statistics.

2. Since for assessment of the safety of the coast storm-related events with very low frequencies of occurrence are of interest, it can be assumed that there is no difference between mean number of occurrences per year and probability of occurrence in an arbitrary year.

3. Statistics required for evaluation of the risk of damage to or failure of a water retaining structure can in principle be derived from a single type of joint statistic of hydraulic parameters in cases of practical interest. This statistic, for the case of two hydraulic parameters given by equation (11a), is independent of failure mechanism. The idea is to estimate this statistic at the -20 m contour and to translate it to the coast.

4. In section 4.3, the advantages are explained of statistics that not only reflect the uncertainty due to the natural variability of weather phenomena but also reflect the uncertainty due to limitations on the number and quality of available data. It is proposed to pursue this subject after a satisfactory basic method for estimation of joint extreme condition statistics has been found.

5. Methods for data selection, parameterization and estimation of joint statistics of waves and high tide levels should be consistent with the most recent approach to estimation of statistics of high tide levels adopted by Rijkswaterstaat (see section 5.1). It is based on a peak-over-threshold approach to data selection and on parameterization of the exceedance frequencies of set-up or high tide level using the Generalized Pareto distribution.

6. It is recommended to explore two different approaches to the assessment of joint statistics of waves and water levels: these are

- The current approach known as the method of Bruinsma/Van Aalst, which is based on simplification of the physics of generation of surges and waves
- Direct estimation of the required statistics from data using the asymptotic shape for the probability that several hydraulic parameters simultaneously exceed high levels during a storm.

First, these two approaches can be tested separately to identify strengths and weaknesses of each method, and then the methods can be compared. Tests can be based on measured data and on hindcast (NESS) data. Based on this experience, improvements can be implemented and the final choice of method can be made.

REFERENCES

Booij, N. and L.H. Holthuijsen(1991): *Het golfveld langs de Nederlandse kust bij een zeespiegelrijzing*. Report no. 902, Digital Hydraulics Holland B.V., Zwijndrecht.

Bruinsma, J.(1982): *Golfhoogte-waterstand relatie t.p.v. de NAP -20 m lijn nabij Hoek van Holland t.b.v. de leidraad duinafslag*. Notitie WWKZ-82G.259, Rijkswaterstaat, Den Haag.

Bruinsma, J.(1985): *De waterstand-golfhoogterelatie t.b.v. duinafslagberekeningen*. In: Cursus Duinafslag, Postacademisch onderwijs in de technische wetenschappen, Delft.

Cramér, H. and M.R. Leadbetter(1967): *Stationary and related stochastic processes - Sample function properties and their applications*. John Wiley and Sons Inc., New York.

De Haan, L.(1990): *Fighting the arch-enemy with mathematics*. Statistica Neerlandica 44, nr. 2, pp.45-68.

Delta Committee(1960): *Report of the Delta Committee*. Staatsuitgeverij, Den Haag.

De Valk, C.F.(1991): *Long waves on the North Sea - measurement and data processing* (draft). DELFT HYDRAULICS, Delft.

Dillingh, D., Samenwerkingsverband Basispeilen 1991): *Voorstel voor de globale indeling van het eindrapport*. Rijkswaterstaat, Den Haag.

Holthuijsen, L.H., N. Booij and T.H.C. Herbers(1989): A prediction model for stationary short-crested waves in shallow water with ambient currents. *Coastal Engng*. 13, p. 23-45.

IAHR(1986): *List of sea state parameters*. International Association for Hydraulic Research, Brussels.

Leadbetter, M.R., G. Lindgren and H. Rootzén(1983): *Extremes and related properties of random sequences and processes*. Springer-Verlag, New York.

Muir, L.R. and Al-Shaarawi, A.H.(1986): On the calculation of extreme wave heights: a review. *Ocean Engng*, Vol. 13, pp. 93-118.

Pickands, J.(1981): Multivariate extreme value distributions. *Bull. Int. Statist. Inst.* 859-878.

Samenwerkingsverband Basispeilen [R. Helmers, A.L.M. Dekkers, L. De Haan, H. Daan, J.W. van der Made, J. van Malde, J.G. de Ronde] (1986): *Onderzoek Basispeilen-tussentijds verslag*. Nota GWIO-86.008, Rijkswaterstaat, Den Haag.

Tawn, J.A. (1988): *Bivariate extreme value theory: models and estimation.* Biometrica 75, pp. 397-415.

Technische Adviescommissie voor de Waterkeringen (1984): *Leiddraad voor de beoordeling van de veiligheid van duinen als waterkering.*

Van Aalst, W. (1983): *Golfhoogte-waterstand relatie t.p.v. de NAP -20 m lijn langs de Nederlandse kust.* Notitie WWKZ-82G.259, Rijkswaterstaat, Den Haag.

Van der Made, J.W., R. Helmers, A.L.M. Dekkers, L. De Haan, H. Daan, J. van Malde, J.G. de Ronde, D. Dillingh (1989): *Frequentielijnen kust - tussentijds verslag 2.* Nota GWIO 89.001, Rijkswaterstaat, Den Haag.

Vogel, J.A. and C.F. de Valk (1991): *Statistiek van seiches; rapportage van de derde fase van het seichesonderzoek t.b.v. een stormvloedkering in de Nieuwe Waterweg.* Nota GWAO-88.038, Rijkswaterstaat, Den Haag.

Vrijling, J.K. and J. Bruinsma, (1980): Hydraulic boundary conditions. *Symposium on hydraulic aspects of coastal structures*, Rotterdam.

SYMBOLS

physical parameters		first occurrence at page:
f	frequency	6
θ	direction of wave propagation [degrees]	6
S	directional spectrum (spectral density)	6
h	high tide level	6
a	astronomical tide	7
s	set-up	7
p_h	persistance of high tide level h	7
H_{m0}	significant wave height	7
m_0	variance of sea surface elevation	7
m_i	i-th moment of nondirectional sea surface spectrum	7
$T_{0,2}$	mean wave period defined in terms of zeroth and second spectral moments	7
$T_{i,j}$	idem, in terms of i-th and j-th moments	7
θ_0	mean wave direction	7
u_{10}	wind speed	8
θ_{10}	wind direction	8
ν_2	spectral width parameter	8
σ_0	directional spreading	8
v	current velocity magnitude	8
θ_v	current direction	8
t	time in years	12
d	actual depth	25
F	fetch	25
τ	wind duration	25
H_s	significant wave height	25
H_s^{sw}	swell component of H_s	25
g	gravitation constant	26

mathematical symbols

$\triangleq$	is defined as	6
$<a,b>$	inner product of functions a and b	6
$P[A]$	probability of the event A	12
$\{...\}$	set, decribed between the braces	16
$\subset$	is a subset of ...	16
$\in$	is an element of ...	15
$\cap$	intersection of sets, or "and" (depending on context)	17
$\cup$	union of sets, or "and/or" (depending on context)	36
γ, γ_i	parameter of the generalized Pareto distribution (see appendix c)	24

other symbols

r_t	return period based on probability of occurrence in a time-interval of fixed length t	12
$k^{0,t}$	number of occurrences of an event in the time-interval [0,t]	12
μ	expected number of occurrences of an event per year	13
ρ	return period derived from μ	13
u_i	an environmental parameter (identified by the label i) at a particular location	15
f	total load	15
$m\ell$	maximum of total load over an arbitrary year	15
PF	probability of failure in an arbitrary year	15
H	distribution function of strength	15
$\mathcal{B}$	general region in parameter space with boundary $\partial\mathcal{B}$ that serves as a threshold to distinguish storm events	16
$\partial\mathcal{B}$	the boundary of $\mathcal{B}$	16
$I_{\mathcal{B}}$	randomly selected time-interval during which the hydraulic parameters are in the region $\mathcal{B}$ ("storm")	16
I_{ω}	randomly selected time-interval during which the hydraulic parameter exceeds the threshold ω	30
F_c	probability that several hydraulic parameters simultaneously exceed certain levels during a storm	17
u_i	a scaled hydraulic parameter (see (27a))	31

APPENDIX

a: proof of equation (11) in section 4.2

Definition: let u be the unit step function defined here as follows: $u(a)=0$ for $a\le 0$ and $u(a)=1$ for $a>0$.
Consider the set of "jointly exceeded" points in the plane

$$\{ (a_1,a_2):\ u_1(t)>a_1 \cap u_2(t)>a_2 \text{ for some } t\in I_{\mathcal{B}} \} \tag{a1}$$

It can be written as

$$\{ (a_1,a_2):\ g(a_1,a_2)>0 \} \tag{a2a}$$

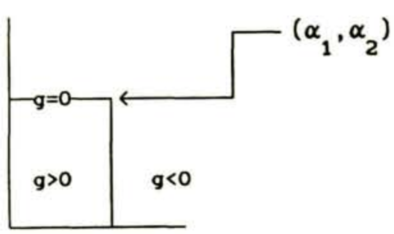
with

$$g(a_1,a_2) \triangleq \max_{t\in\mathcal{B}} \min[u_1(t)-a_1,u_2(t)-a_2] \tag{a2b}$$

Now let the integral

$$\int_{\mathbb{R}^2} u[f(a_1,a_2) - s]\ du[g(a_1,a_2)] \tag{a3}$$

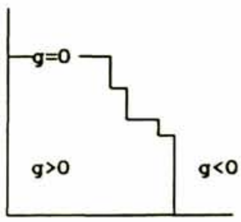
have the following meaning: if the curve g=0 is of the form



(which is the case when $u_1(t)$ and $u_2(t)$ reach their maxima at the same time on (a2b)), then the integral (a3) is clearly defined as

$$\int_{\mathbb{R}^2} u[f(a_1,a_2) - s]\ d\lambda(a_1,a_2) \tag{a4}$$

with λ a postive measure on the plane, which in this case is simply a delta function at the point (α_1,α_2). A more general curve $g=0$ is for example



In this case, the measure λ is the difference $\lambda^+ - \lambda^-$ of two positive measures λ^+ and λ^-: λ^+ consists of delta peaks at the corners of the form ┐ in the curve g= 0, and λ^- of delta peaks at the corners of the form └. For more general curves g=0 with g of the form (a2b), (a3) is defined by approximating the curve by one consisting of straight line segments parallel to the axes. Now the integral (a3) equals the number of upcrossings through zero of g along the curve f=s: assume that this number of upcrossings equals one. Then for the straight-line approximation of the curve g=0, the measure λ^+ of the set

$$\{ (a_1,a_2): f(a_1,a_2) > s \} \qquad \text{(a5)}$$

is always equal to unity plus the measure λ^- of the same set, as is found by by counting the corners of the curve g=0 in this set. See e.g. the following example in which there is a single upcrossing through zero of g along the curve f=s:

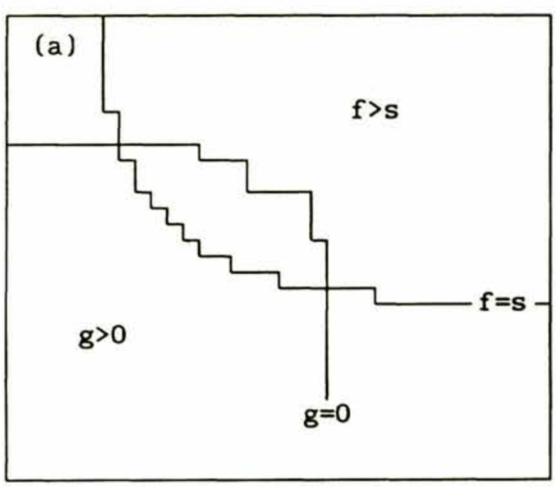


figure: (a) example of curves f=s and g=0

Therefore, the measure λ of the set (a5) always equals unity if there is just one upcrossing through zero of g along the curve f=s (and equal to zero if there is no upcrossing, of course). Therefore, with the meaning of (a3) as explained,

$$\mathcal{u}[\max_{t\in I_{\mathcal{B}}} f(u_1(t),u_2(t))> s] = \int_{\mathbb{R}^2} \mathcal{u}[f(a_1,a_2) - s]\, d\mathcal{u}[g(a_1,a_2)] \qquad (a6)$$

Assuming that for every f in a particular class there is with probability one not more than such upcrossing, then by taking expectations on both sides of (a3):

$$P[\max_{t\in I_{\mathcal{B}}} f(u_1(t),u_2(t))> s] =$$

$$\int_{\mathbb{R}^2} \mathcal{u}[f(a_1,a_2) - s]\, dP[u_1(t)>a_1 \cap u_2(t)>a_2 \text{ for some } t\in I_{\mathcal{B}}] \qquad (a7)$$

for all such f.

b: Probabilities of simultaneous exceedance during a fixed time interval, and during a storm

Consider a hydraulic parameter u at a fixed location as a stochastic process, its stochastic character reflecting the uncertainty due to the uncertain weather conditions affecting it. Also, u may be a scalar function of several hydraulic parameters.
We will give some assumptions on u that are just sufficient to derive the asymptotic shapes of statistics needed in the present context. This means in particular, that we will only need the probability of an extreme event in a rather short, fixed, time interval, such that the probability of occurrence of the event in this time-interval is very small. In section 4.1, it was mentioned that we are interested in for example the high tide levels that correspond to probabilities of exceedance in a year of as small as 10^{-4}. Closely related results covering also the case that these probabilities are larger can be derived under an additional assumption. This is discussed in appendix g.

Assume that u is a seasonal stochastic process, which means that the statistics of u are the same as of $L^k u$ if k is an integer, with L the foreward shift defined by

$$L^q u(t) = u(t-q) \tag{b1}$$

(time is measured in years).

Consider the point process of upcrossings of a threshold ω by u. The expected number of upcrossings in a year of the threshold ω by u is

$$\mu_{\omega}$$

Observe that since the process u is seasonal, μ_{ω} can be used to compute the mean number of upcrossings in a time-interval only for time-intervals that are an integer number of years, say n years.

μ_{ω} increases with increasing ω. Conversely, we can write the threshold as a function of the expected number of upcrossings, as

$$\omega(\mu)$$

The advantage of this is that it is more flexible: u may for example have an upper bound, and then we cannot give results for $\omega \longrightarrow \infty$, but we can still give results for $\mu \longrightarrow 0$. Let

$$k_{\omega}^{0,n}$$

be the number of upcrossings of ω by u in the time-interval [0,n].
Observe that, by the definition of μ,

$$(\mu n)^{-1} \sum_{m=1}^{\infty} m \, P[k^{0,n}_{\omega(\mu)} = m] = 1 \qquad \text{(b2)}$$

Now the assumptions on u are the following.

(i) $$\lim_{\mu\to 0} (\mu n)^{-1} P[k^{0,n}_{\omega(\mu)} > 1] = 0 \qquad \text{for every fixed } n>0 \qquad \text{(b3)}$$

This means that the probability of more than one upcrossing of $\omega(\mu)$ decreases faster than the probability of one upcrossing of $\omega(\mu)$ with μ going to zero. Sufficient conditions for (i) are given in (??).

(ii) for an arbitrary function $z \geq 0$,

$$P[\ u(t) > z(\mu)+\omega(\mu) \text{ for some } t\in[0,n] \ |\ k^{0,n}_{\omega(\mu)} = 1] \longrightarrow$$

$$P[\ u(t) > z(\mu)+\omega(\mu) \text{ for some } t\in I_{\omega(\mu)}] \qquad \text{with } \mu \to 0 \qquad \text{(b4)}$$

were I_{ω} is an arbitrary (randomly chosen) interval in [0,n] (an integer number of years) in which $u(t)>\omega$ for all t in I_{ω}.
What this assumption means is that say for a particular year, the probability that u(t) reaches the level $z+\omega$ during a randomly picked storm (i.e. an interval I_{ω}) in that year is independent of the number of storms in that year (in the limit, if the threshold ω becomes high enough).

The function z can be an arbitrary constant; the reason for including the case that it depends on μ is to account for those cases in which u has an upper limit. The important thing is that z is positive, so $z+\omega$ is higher than ω, and therefore reaching it is an event that occurs less frequently than reaching ω.

assertion: under the assumptions (i) and (ii):

$$(\mu n)^{-1} P[\ u(t) > a\pi(\omega(\mu)) + \omega(\mu) \text{ for some } t\in[0,n]] \longrightarrow$$

$$P[\ u(t) > a\pi(\omega(\mu)) + \omega(\mu) \text{ for some } t\in I_{\omega}\ ] \qquad \text{(b6)}$$

with $\mu \longrightarrow 0$

proof:

$$(\mu n)^{-1} P[\ u(t) > a+\omega(\mu) \text{ for some } t\in[0,n]] =$$

$$(\mu n)^{-1} \sum_{m=1}^{\infty} P[\ u(t) > a+\omega(\mu) \text{ for some } t\in[0,n] \mid k^{0,n}_{\omega(\mu)} = m]\ P[\ k^{0,n}_{\omega(\mu)} = m] \longrightarrow$$

$$P[\ u(t) > a+\omega(\mu) \text{ for some } t\in[0,n] \mid k^{0,n}_{\omega} = 1] \qquad \text{with } \mu\to 0 \qquad \text{(b7)}$$

by (b3), using (b2). Then apply (b4).

It is straightforward to generalize the assumptions (i) and (ii) above to more than one variable. In this case, define

$\omega = (\omega_1,\dots,\omega_k)$

and define I_ω as a (random) time-interval in which

$u_i(t) > \omega_i$ for $i= 1,\dots,k$ for all t in I_ω

The function $\mu \longrightarrow \omega(\mu)$ is now an arbitrary curve in $\mathbb{R}^k$,

$\mu \longrightarrow (\omega_1(\mu),\dots,\omega_k(\mu))$

such that at each point, the expected number of entrances in the region

$$\{\ (a_1,a_2):\ a_i > \omega_i(\mu)\ \forall i= 1,\dots,k\ \} \qquad \text{(b8)}$$

equals μ. Then by an obvious modification of the assumptions (i) and (ii), for any functions $z_1>0,\dots,z_k>0$:

$$(\mu n)^{-1} P[\ u_i(t) > z_i(\mu)+\omega_i(\mu)\ \forall i=1,\dots,k \quad \text{for some } t\in[0,n]\ ] \longrightarrow$$

$$P[\ u_i(t) > z_i(\mu)+\omega_i(\mu)\ \forall i=1,\dots,k \quad \text{for some } t\in I_\omega\ ]$$

$$\text{with } \mu\to 0 \qquad \text{(b9)}$$

c: <u>Parameterization of probabilities of simultaneous exceedance during a storm</u>

Assume that $\mathbf{u} = (u_1,..,u_k)$ satisfies the assumptions as in appendix b. Let

$w \triangleq (w_1,..,w_k)$

be a vector of thresholds, and let I_w be a (random) time-interval in which

$$u_i(t) > w_i \quad \text{for } i = 1,..,k \quad \text{for all } t \in I_w \tag{c1}$$

Now let $\mu \longrightarrow w(\mu)$ be a curve in $\mathbb{R}^k$ such that the expected number of entrances of the region

$$\{ \mathbf{a} \in \mathbb{R}^k:\ a_i > w_i(\mu) \quad \forall i = 1,..,k \} \tag{c2}$$

by $\mathbf{u}(t)$ equals μ, and assume that for every point $\mathbf{a}$, there is a curve $\mu \longrightarrow w(\mu)$ such that

$$P[\ u_i(t) > a_i \pi_i(w_i(\mu)) + w_i(\mu) \ \forall i \in \{1,..,k\} \text{ for some } t \in I_{w(\mu)}\ ]$$

$$\longrightarrow \mathcal{F}_c(\mathbf{a})$$

$$\text{with } \mu \longrightarrow 0 \tag{c3}$$

for some function $\mathcal{F}_c$, with $\pi_1,..,\pi_k$ nonnegative functions. The curves $\mu \longrightarrow w(\mu)$ are yet unspecified. Then

[a]

$$\mathcal{F}_c(\mathbf{a}) = \mathfrak{F}_c(\alpha) \tag{c4a}$$

with $\alpha = (\alpha_1,..,\alpha_k)$ defined by

$$\alpha_i \triangleq \gamma_i^{-1} \ln(\gamma_i a_i + 1) \tag{c4b}$$

and $\mathfrak{F}_c$ defined by

$$\mathcal{F}_c(a) \triangleq \exp(-\phi_\varepsilon \sum_{i=1}^{k} a_i) \qquad \text{(c4c)}$$

with $\varepsilon \triangleq (\varepsilon_1,..,\varepsilon_m)$,

$$\varepsilon_i \triangleq [\sum_{j=1}^{k} a_j]^{-1} a_i \qquad \text{(c4d)}$$

and ϕ some nonnegative function on the unit (k-1)-simplex (the set of points ε satisfying $\varepsilon_i \geq 0$ for i=1,..,k-1 and $\varepsilon_1 + .. + \varepsilon_{k-1} \leq 1$).

[b] The convergence in (c3) holds for example for curves of the form

$$\omega_i(\mu) = \gamma_i^{-1}(\mu^{-\gamma_i \varepsilon_i} - 1) \qquad \text{(c5a)}$$

and for functions π_i of the form

$$\pi_i(x) = \gamma_i x + 1 \qquad \text{(c5b)}$$

with $\gamma_1,..,\gamma_k$ some real constants.

[c] Moreover, by scaling the original processes $u_1,.., u_k$ by means of

$$\mathcal{u}_i(t) \triangleq \gamma_i^{-1} \ln(\gamma_i u_i(t)+1) \qquad \text{(c6a)}$$

also

$$P[\ \mathcal{u}_i(t) > a_i + \varepsilon_i \vartheta \quad \forall i \in \{1,..,k\} \text{ for some } t \in \mathcal{I}_{\varepsilon\vartheta}\ ] \longrightarrow \mathcal{F}_c(\mathbf{a}) \qquad \text{(c6b)}$$

$$\text{with } \vartheta \longrightarrow \infty$$

with $\mathcal{I}_\omega$ defined as I_ω but now for the scaled processes $\mathcal{u}_1,..,\mathcal{u}_k$ instead of $u_1,..,u_k$, and ε defined in (c4d).

proof:
Combining (c3) with (b9) of appendix b gives

$$(\mu n)^{-1} P[\ u_i(t) > a_i \pi_i(\omega_i(\mu)) + \omega_i(\mu)\ \forall i \in \{1,..,k\} \text{ for some } t \in [0,n]\ ]$$

$$\longrightarrow \mathcal{F}_c(\mathbf{a}) \qquad \text{with } \mu \longrightarrow 0 \qquad \text{(c7)}$$

As a consequence

$$\mu^{-1}\mathcal{F}_c(a_1\pi_1(\omega_1(\mu)) + \omega_1(\mu), \ldots) = \mathcal{F}_c(a_1, \ldots) \qquad \text{(c8)}$$

in which $\mathcal{F}_c(a_1,\ldots)$ is shorthand for $\mathcal{F}_c(\mathbf{a})$.

Repeating (c8) m times gives

$$\mu^{-m}\mathcal{F}_c(a_1[\pi_1(\omega_1(\mu))]^m + \omega_1(\mu)\sum_{j=0}^{m-1}[\pi_1(\omega_1(\mu))]^j, \ldots) = \mathcal{F}_c(a_1, \ldots) \qquad \text{(c9a)}$$

Evaluating (c8) at μ^m gives

$$\mu^{-m}\mathcal{F}_c(a_1\pi_1(\omega_1(\mu^m)) + \omega_1(\mu^m), \ldots) = \mathcal{F}_c(a_1, \ldots) \qquad \text{(c9b)}$$

To obtain the most general representation of $\mathcal{F}_c$, the arguments of $\mathcal{F}_c$ on the left-hand sides of (c9) are taken to be equal. Equivalently, for each point **a**, we consider only one curve that satisfies (c3), which implies that it should satisfy:

$$[\pi_1(\omega_1(\mu))]^m = \pi_1(\omega_1(\mu^m)) \qquad \text{(c10a)}$$

$$\omega_1(\mu)\sum_{j=0}^{m-1}[\pi_1(\omega_1(\mu))]^j = \omega_1(\mu^m) \qquad \text{(c10b)}$$

From (c10a),

$$\pi_1(\omega_1(\mu)) = \mu^{-\beta_1} \qquad \text{(c11)}$$

for some real number γ_1. Then the solution of (c10b) is

$$\omega_1(\mu) = d_1\beta_1^{-1}(\mu^{-\beta_1} - 1) \qquad \text{(c12a)}$$

and with (c11)

$$\pi_1(x) = d_1^{-1}\beta_1 x + 1 \qquad \text{for all real x} \qquad \text{(c12b)}$$

Since π_1 is a fixed function independent of the choice of curve $\mu \longrightarrow \omega(\mu)$, the number $d_1^{-1}\beta_1$ must be a constant, so defining

$$\gamma_1 \triangleq d_1^{-1}\beta_1 \qquad \text{(c13)}$$

we obtain

$$\pi_1(x) = \gamma_1 x + 1 \qquad \text{for all real x} \qquad \text{(c14a)}$$

and

$$\omega_i(\mu) = \gamma_i^{-1}(\mu^{-\gamma_i d_i} -1) \tag{c14b}$$

with $d_1,\ldots,d_m$ numbers that can be chosen to obtain different curves $\mu \longrightarrow \omega(\mu)$.

With these particular curves and functions π_i, we obtain from (c8)

$$\mu^{-1}\mathfrak{F}_c(a_i\mu^{-\gamma_i d_i} + \gamma_i^{-1}(\mu^{-\gamma_i d_i} -1),\ldots) = \mathfrak{F}_c(a_i,\ldots) \tag{c15}$$

so

$$\mu^{-1}\mathfrak{F}_c(\gamma_i^{-1}((\gamma_i a_i+1)\mu^{-\gamma_i d_i} -1),\ldots) = \mathfrak{F}_c(a_i,\ldots) \tag{c16}$$

Now define a function $\mathfrak{F}_c$ by

$$\mathfrak{F}_c(\ln x_i,\ldots) \triangleq \mathfrak{F}_c(\gamma_i^{-1}((x_i)^{\gamma_i} -1),\ldots) \qquad \forall \mathbf{x}\in\mathbb{R}^k;\ x_i>0 \text{ for } i=1,..,k \tag{c17}$$

then

$$\mu^{-1}\mathfrak{F}_c(\gamma_i^{-1}\ln(\gamma_i a_i+1) -d_i\ln\mu,\ldots) = \mathfrak{F}_c(\gamma_i^{-1}\ln(\gamma_i a_i+1),\ldots) \tag{c18}$$

or (noting that $\gamma_1,\ldots,\gamma_k$ are given constants) with

$$\alpha_i = \gamma_i^{-1}\ln(\gamma_i a_i+1) \tag{c19}$$

$$\mu^{-1}\mathfrak{F}_c(\alpha_i -d_i\ln\mu,\ldots) = \mathfrak{F}_c(\alpha_i,\ldots) \tag{c20}$$

There is (in this particular case) no reason for fixing $d_1,..,d_k$ at particular values, since they determine paths in the plane, and for μ small enough, the argument $\alpha_i-d_i\ln\mu$ of $\mathfrak{F}_c$ becomes dominated by $-d\ln\mu$. So the only way to obtain a solution is to take $(\alpha_1,..,\alpha_k)$ in the direction of $(d_1,..,d_k)$. In particular, setting $\alpha_1,..,\alpha_k$ to zero:

$$\mu^{-1}\mathfrak{F}_c(-d_i\ln\mu,\ldots) = \mathfrak{F}_c(0,\ldots) = 1 \tag{c21}$$

Define

$$\varepsilon_i = [\sum_{j=1}^{k} d_j]^{-1}d_i \tag{c22}$$

and let

$$\omega = -\ln \mu \; [\sum_{j=1}^{k} d_j] \qquad \text{(c23)}$$

then (c21) becomes

$$\mathcal{F}_c(\varepsilon_1\omega, \ldots) = \exp(-\omega\phi_{\varepsilon}) \qquad \text{(c24)}$$

with $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_{k-1})$ a point on the unit simplex (see the remark following (c4d)), and with ϕ some nonnegative function defined on the unit simplex. Combining (c24) with (c17) gives (c4). This proves [a]. [b] is proven by (c14b), and noting that

$$\mu^{-d_1} = [z(\mu)]^{\varepsilon_1} \qquad \text{with } z(\mu) = \mu^{-(d_1 + \ldots + d_k)} \qquad \text{(c25)}$$

Moreover, the magnitude of $(d_1 + \ldots + d_k)$ has no effect on the curves' image, so it can be replaced by an arbitrary constant, for example 1.
To prove [c], note that from (c3) and (c14)

$$P[\; u_i(t) > \gamma_i^{-1}((\gamma_i a_i + 1)\mu^{-\gamma_i d_i} - 1) \;\; \forall i \in \{1, \ldots, k\} \text{ for some } t \in I_{\omega(\mu)} \;]$$

$$\longrightarrow \mathcal{F}_c(a) \qquad \text{with } \mu \longrightarrow 0$$

The result (c6b) follows by applying (c6a) and noting that by (c22), $-d_1 \ln \mu = \varepsilon_1 \omega$ with $\omega \longrightarrow \infty$ as $\mu \longrightarrow 0$.

Univariate case: the generalized Pareto distribution.

Let I_ω be a (random) time-interval over which $u(t) > \omega$. Assume that for some function $\mathcal{F}_c$:

$$P[\; u(t) > a\pi(\omega(\mu)) + \omega(\mu) \text{ for some } t \in I_\omega \;] \longrightarrow \mathcal{F}_c(a) \qquad \text{(c26)}$$

with $\mu \longrightarrow 0$ and π some function greater than zero. Then

$$\mathcal{F}_c(a) = (1+a\gamma)^{-\phi/\gamma} \tag{c27}$$

for some real numbers $\phi>0$ and γ.

The reason for assuming convergence as in (c3) is that only under this condition, the probability of $u_1(t)>a_1 \cap u_2(t)>a_2$ in a fixed number of years converges in a manner similar to (c3) (either keeping the time-interval fixed, or letting it increase too). This convergence of extremes is the kind of regular behavior that is known to occur for many random processes satisfying assumptions as stated in the beginning. In particular when the continuous-time processes are replaced by a sequence of independent identically distributed random vectors, (c3) is a quotient of the form

$$\frac{F_c(a_1\pi_1(\omega_1)+\omega_1,\ldots)}{F_c(\omega_1,\ldots)} \tag{c28}$$

(with F_c the joint survivor function of u_1 and u_2) which converges quite generally. In fact, if a condition like (c3) does not hold, extrapolation of statistics from limited number of data to events outside the range of observations is not possible, and any attempt to do so would be useless.

In general, $\mathcal{F}_c$ (and so $\mathcal{F}_c$) does not need to correspond to a probability measure, which means that its density need not to be positive (remember this is the same with $\mathbf{F}_c$ in section 4.2). Therefore, apart from nonnegativity, there are no rules that the function ϕ should satisfy. However, as with $\mathbf{F}_c$, $\mathcal{F}_c$ does have a nonnegative density if the maxima of u_1 and u_2 during a storm always coincide.

d: Maximum likelyhood estimation of $\mathcal{F}_c$

The maximum likelyhood estimator is important as a practical tool, but also because the likelyhood function is a component in the type of estimator discussed in section 4.3. First we will discuss the likelyhood function and maximum likelyhood estimation of $\mathfrak{F}_c$ (when the data have exponential marginals), and then we will discuss maximum likelyhood estimation of $\mathcal{F}_c$ (arbitrary marginals).

Following [*Tawn*, 1988], who refers to [*Pickands*, 1981] (he discusses a class of models which in fact contain the limiting shape for $\mathfrak{F}_c$), and using the notation of appendix c: (see (c6))

$$P[\ u_i(t) > \varepsilon_i(\vartheta+a) \quad \text{for } i=1,..,k \text{ for some } t\in\mathcal{T}_{\varepsilon\vartheta}\ ] =$$

$$P[\ \varepsilon_i^{-1}u_i(t) > (\vartheta+a) \text{ for } i=1,..,k \text{ for some } t\in\mathcal{T}_{\varepsilon\vartheta}\ ] =$$

$$P[\ z_\varepsilon - \vartheta > a\ ] \tag{d1a}$$

with

$$z_\varepsilon \triangleq \sup_{t\in\mathcal{T}_{\varepsilon\vartheta}} \ \min_{i=1,..,k} \ \varepsilon_i^{-1}u_i(t) \tag{d1b}$$

so from appendix c:

$$\lim_{\vartheta\to\infty} P[\ z_\varepsilon - \vartheta > a\ ] = \mathfrak{F}_c(\varepsilon a) = \exp(-\phi_\varepsilon a) \tag{d2}$$

Therefore, the log likelyhood function for a single value ε, without posing any restrictions on the shape of ϕ, is, with m the number of samples:

$$\sum_{j=1}^{m} \ell\ell_{\phi_\varepsilon}(z_\varepsilon^j) \tag{d3a}$$

with

$$\ell\ell_{\phi_\varepsilon}(a) = \ln \phi_\varepsilon - \phi_\varepsilon(a) \tag{d3b}$$

and z_ε^j given by (d1b), with j referring to the j-th sample. Therefore, the maximum likelyhood estimator for ϕ_ε is

$$\hat{\phi}_{\varepsilon}^{ML} = m[\sum_{j=1}^{m} z_{\varepsilon}^{j} - \omega]^{-1} \tag{d4}$$

Without any restrictions on ϕ, this estimator may be fine, even if $\mathcal{F}_c$ is not a proper survivor function. The data need not be points, but may be curves. However, it can be expected that even if $\mathcal{F}_c$ is not a proper survivor function, ϕ will be smooth in some sense, so a sufficiently flexible parametric estimator for ϕ can be more efficient.

A possible restriction on ϕ is that $\mathcal{F}_c$ should be a proper survivor function (which means that the corresponding density is nonnegative). We will discuss here only the bivariate case. Then $\varepsilon = \varepsilon_1$, and ϕ must satisify:

$$\omega(\phi-\phi'\varepsilon)(\phi+\phi'(1-\varepsilon)) + \phi''\varepsilon(1-\varepsilon) \geq 0 \qquad \forall\omega\geq 0 \tag{d5}$$

so taking ω equal to zero:

$$\phi'' \geq 0 \tag{d6}$$

If (d6) holds, (d5) is also satisfied for all $\omega>0$, so ϕ must satistfy (d6), i.e.ϕ must be convex. [*Tawn*,1988] gives some examples of convex parameterizations of ϕ for the case that $\phi_0 = \phi_1 = 1$.
If the samples are points (as is most likely the case), $\mathcal{F}_c$ must be a proper survivor function so ϕ must be convex. To obtain in addition smoothness of the estimate of ϕ, a nonsymmetric parametric model is to be preferred [see *Tawn*,1988]. The maximum likelyhood estimator for the parameters θ of a parameterized function ϕ is

$$\hat{\theta}^{ML} = \underset{\theta}{\arg\max} \sum_{j=1}^{n} \ell\ell_{\theta}(u_1^j, u_2^j) \tag{d7}$$

with $\ell\ell_{\theta}$ the logarithm of the likelyhood function of a single sample.
If the samples are not points but curves, it is not straightforeward to define a meaningful estimator for a parameterized ϕ. The only sensible approach seems

$$\hat{\theta} = \underset{\theta}{\arg\max} \sum_{i=1}^{n} \min_{t\in\mathcal{I}_{\mathcal{B}}^{j}} \ell\ell_{\theta}(u_1(t), u_2(t)) \tag{d8}$$

This estimator chooses on each sample curve the point with the smallest likelyhood. There seems no reason why this estimator wouldn't be consistent just like the maximum likelyhood estimator in the case that the samples are points.

Now the estimation of $\mathcal{F}_c$ (for data with arbitrary marginals) will be discussed. This $\mathcal{F}_c$ is of the form (again in the bivarate case)

$$\mathcal{F}_c(a_1,a_2) = \exp(\phi_\varepsilon(\alpha_1+\alpha_2)) \qquad \varepsilon = \alpha_1/(\alpha_1+\alpha_2)$$

$$\alpha_1 = \gamma_1^{-1}\ln(\gamma_1 a_1+1) \qquad \text{(d9)}$$

For given parameterizations of these marginals, it is a simple matter to write down the likelyhood function of the data based on (d9), and maximize it to the parameters γ_1 and γ_2 of the marginals and the parameters of ϕ.
An alternative method [see *Tawn*, 1988] is to estimate $\gamma_1, \phi_0, \gamma_2$ and ϕ_1 first, then use these estimates to transform the data to obtain data with unit exponential marginals, and then estimate ϕ under the restriction that ϕ_0 and ϕ_1 are already fixed. In general, this method will produce different estimates than direct maximization of the likelyhood function of the original data. However, this method is not necessarily worse.

When applying maximum likelyhood estimators in the present context, it is not a matter of simply computing the estimate for a particular data set, but to examine the behavior of the estimates as a function of the thresholds on the data. This asymptotic behavior is what is needed to extrapolate the statistics beyond the range of observations.
This is recognized by [*De Haan*, 1990], applying new moment estimators to estimate γ_1. The alternative is to apply the maximum likelyhood estimator over a range of thresholds. Possible multivariate analogues of quantile estimators as in [*De Haan*, 1990] do not seem useful in the present context, since $\mathcal{F}_c$ is needed in parametric form, in order to be translated from one location to another and to be applied to estimate statistics of failure as in section 4.2.

e: Conditional probabilities as estimators

This appendix presents a more precise treatment of the approach introduced in section 4.3 to estimate probabilities that cover sources of uncertainty. It complements the text of section 4.3.

A final estimate of a probability, including the uncertainty due to limitations on number and accuracies of data, is the conditional probability *relative to* the available data. For example, let the distribution function of the annual maximum wave height be (with time in years)

$$P[\max_{t\in[0,1]} H_{m0}(t) \leq \alpha] = F_\theta(\alpha) \qquad \text{(e1)}$$

for some vector of parameters θ, so by letting θ run over $\mathbb{R}^m$, we obtain all possible distribution functions of the annual maximum wave height. Now θ must be estimated from the data, say $\mathbf{y} = (y_1,\dots,y_n)$, assumed independent annual maxima of n years (this is just an example).
The classical maximum likelyhood estimator for θ is

$$\hat{\theta}^{ml} = \arg\max_{\theta\in\mathbb{R}^m} \prod_{i=1}^{n} f_\theta(y_i) \qquad \text{(e2)}$$

with f_θ the probability density function corresponding to F_θ.
This estimate is converted to a value $\hat{\alpha}$ corresponding to a fixed probability of $F_{\hat{\theta}^{ml}}(\hat{\alpha})$. An (exact or approximate) confidence band about $\hat{\theta}^{ml}$ is then computed, and this is converted to a confidence band around $\hat{\alpha}$.

The estimate including all uncertainties is the conditional probability

$$P[\max_{t\in[0,1]} H_{m0}(t) \leq \alpha \,|\mathbf{y}] = \int F_\theta(\alpha)\; g[\theta|\mathbf{y}]\; d\theta \qquad \text{(e3a)}$$

with $g(\theta|\mathbf{y})$ the conditional probability density of θ relative to the data $\mathbf{y}$. If we have no information about θ besides the available data $\mathbf{y}$, it is

$$g[\theta|\mathbf{y}] = \left[\int h[\mathbf{y}|\theta]\; d\theta\right]^{-1} h[\mathbf{y}|\theta] \qquad \text{(e3b)}$$

with $h[\mathbf{y}|\theta]$ the conditional probability density of the data $\mathbf{y}$, relative to the parameter vector θ.

Clearly if few data are available or the data are inaccurate, $g[\mathbf{y}|\theta]$ will be dispersed over a large area so the effect will be that probabilities ofexceedance will turn out high. On the other hand, by increasing the number and accuracy of the data, $g[\mathbf{y}|\theta]$ will become more and more concentrated near the true value of θ so the estimate will converge to the true value of θ, and the only uncertainty left is the uncertainty due to the meteorology.

All this is based on the assumption that there is a θ such that the data are drawn from a population with a distribution function of the form F_θ for some θ. In reality, this is just a model for the data, a simplification. Often more than one choice is possible for the family of distributions (although extreme value theory can help in selecting an approppriate one). The estimate (e3) is also robust in the sense that if the data are not fitted well by any distribution of the form F_θ for some θ, relatively high probabilities of exceedance result because $g[\mathbf{y}|\theta]$ can never be concentrated near any value of θ.

f: Joint distribution of maxima during a fixed time-interval, and during a storm

The results in this appendix complement those of appendix c. Joint distributions of maxima are the kind of statistics usually considered in the literature on multivariate extreme value statistics. In this appendix, their parameterization is derived in a way that is very similar to the derivation of the parameterization of probabilities of simultaneous exceedance in appendix c. This serves to show more clearly the relationship between these two multivariate statistics.

It is straightforeward to generalize the assumptions in appendix b to the following case: define

$$\boldsymbol{\omega} = (\omega_1,\dots,\omega_k)$$

and define $J_{\boldsymbol{\omega}}$ as a (random) time-interval over which

$$\max_{i\in\{1,\dots,k\}} u_i(t)-\omega_i > 0 \quad \text{for all } t\in J_{\boldsymbol{\omega}} \tag{f1}$$

Consider curves

$$\mu \longrightarrow (\omega_1(\mu),\dots,\omega_k(\mu))$$

such that at each point, the expected number of entrances in the region

$$\{\ \mathbf{a}\in\mathbb{R}^k:\ a_i > \omega_i(\mu) \text{ for } \textit{some } i = 1,\dots,k\ \} \tag{f2}$$

equals μ. Then by an obvious modification of the argument in appendix b, for any functions $z_1>0,\dots,z_k>0$ and keeping n fixed:

$$(\mu n)^{-1}P[\ u_i(t)\le z_i(\mu)+\omega_i(\mu)\ \forall i\in\{1,\dots,k\} \text{ for all } t\in[0,n]\ ] \longrightarrow$$

$$P[\ u_i(t)\le z_i(\mu)+\omega_i(\mu)\ \forall i\in\{1,\dots,k\} \text{ for all } t\in I_{\boldsymbol{\omega}}\ ]$$

$$\text{with } \mu\to 0 \tag{f3}$$

Now the complement of appendix c for the joint distribution of maxima follows. Assume that for each **a**, there is at least one curve $\mu \longrightarrow \boldsymbol{\omega}(\mu)$ as above such that

$$P[\; u_i(t) \le a_i \pi_i(\omega_i(\mu)) + \omega_i(\mu) \;\; \forall i \in \{1,..,k\} \text{ for all } t \in J_{\omega(\mu)} \;]$$

$$\longrightarrow \mathfrak{G}(\mathbf{a}) \qquad\qquad \text{with } \mu \longrightarrow 0 \qquad\qquad \text{(f4)}$$

for some function $\mathfrak{G}$, with $\pi_1,..,\pi_k$ nonnegative functions. The curves $\mu \longrightarrow \omega(\mu)$ are not yet specified. Then

$$\mathfrak{G}(\mathbf{a}) = \hat{\mathfrak{G}}(\alpha) \qquad\qquad \text{(f5a)}$$

with $\alpha = (\alpha_1,..,\alpha_k)$ defined by

$$\alpha_i = \gamma_i^{-1}\ln(\gamma_i a_i + 1) \qquad\qquad \text{(f5b)}$$

with $\gamma_1,..,\gamma_k$ fixed constants, and $\hat{\mathfrak{G}}$ is of the form

$$1-\hat{\mathfrak{G}}(\alpha) = \int_{\mathbb{R}^{k-1}} \exp(-\min[d_1 a_1 - x_1,..,d_{k-1}a_{k-1} - x_{k-1}, d_k a_k + x_1 + .. + x_{k-1}])\, d\lambda(\mathbf{x})$$

(f5c)

with $d_1,..,d_k$ fixed constants, and λ a positive Borel measure on $\mathbb{R}^{k-1}$.

[b] The convergence in (f4) holds for example for curves of the form

$$\omega_i(\mu) = \gamma_i^{-1}(\mu^{-d_i\gamma_i} - 1) \qquad\qquad \text{(f6a)}$$

and for functions π_i of the form

$$\pi_i(x) = \gamma_i x + 1 \qquad\qquad \text{(f6b)}$$

[c] Moreover, by scaling the original processes $u_1,.., u_k$ by means of

$$\mathit{u}_i(t) \triangleq \gamma_i^{-1}\ln(\gamma_i u_i(t)+1) \qquad\qquad \text{(f7a)}$$

also

$$P[\, u_i(t) \le a_i + \vartheta \quad \forall i \in \{1,..,k\} \text{ and for all } t \in \mathcal{J}_{1\vartheta} \,]$$

$$\longrightarrow \mathcal{G}(a) \qquad \text{with } \vartheta \to \infty \tag{f7b}$$

with $\mathcal{J}_{\vartheta}$ defined as J_{ϑ} but now for the scaled processes $u_1,..,u_k$ instead of $u_1,..,u_k$, and with $1^{\vartheta} = (1, \ldots, 1)$.

Proof:
the proof follows most of appendix c, with $\mathcal{F}_c$ replaced by $1-\mathcal{G}$ and $\tilde{\mathcal{F}}_c$ replaced by $1-\mathcal{G}$. This means that (c20) is replaced by

$$\mu^{-1}[1 - \mathcal{G}(a_1 - d_1 \ln \mu, \ldots)] = 1 - \mathcal{G}(a_1, \ldots) \tag{f8}$$

In this case however, there is a clear reason to select particular values of $d_1,..,d_k$: setting $a_j = \infty$ for all j unequal to i in (f8), we obtain for the i-th marginal distribution $\mathcal{G}_i$ of $\mathcal{G}$:

$$\mu^{-1}[1 - \mathcal{G}_i(a_i - d_i \ln \mu)] = 1 - \mathcal{G}_i(a_i) \tag{f9}$$

which has the solution

$$1 - \mathcal{G}_i(a) = \exp(-d_i a) \tag{f10}$$

This means that the constants $d_1,..,d_k$ are fixed by the marginal distributions of $\mathcal{G}$. This is the essential difference with appendix c. Now we need to solve (f8) with fixed $d_1,..,d_k$. Following [*De Haan*, 1990], there is a positive measure ν on $\mathbb{R}^k$ such that

$$1-\mathcal{G}(d_1^{-1} a_1, \ldots) = \nu\{\, u \in \mathbb{R}^k: u_i > a_i \text{ for } \textit{some } i \in \{1,..,k\} \,\} \tag{f11}$$

Let

$$\mathbf{1} \triangleq (1, \ldots, 1) \tag{f12a}$$

and

$$\mathbf{1}_1 \triangleq (k,-1,\ldots,-1), \quad \mathbf{1}_2 \triangleq (-1,k,\ldots,-1), \ldots, \quad \mathbf{1}_k \triangleq (-1,\ldots,-1,k). \tag{f12b}$$

Observe that $\mathbf{1}_1,..,\mathbf{1}_{k-1}$ span the subspace orthogonal to $\mathbf{d}$, and that

$$\mathbf{1}_k = -\mathbf{1}_1 - \ldots - \mathbf{1}_{k-1} \tag{f13}$$

Then by (f11): (following [*De Haan*, 1990])

$$1-\mathcal{G}(d_1^{-1}a_1,\ldots) = \nu\{\ \mathbf{u}\in\mathbb{R}^k \mid \mathbf{1}^T\mathbf{u} > \min_{j=1,\ldots,k}[ka_j - \mathbf{1}_j^T\mathbf{u}]\ \} =$$

$$\nu\{\ \mathbf{u}\in\mathbb{R}^k \mid \mathbf{1}^T\mathbf{u} > \min[\min_{j=1,\ldots,k-1} ka_j - \mathbf{1}_j^T\mathbf{u},\ ka_k + \mathbf{1}_1^T\mathbf{u} + \ldots + \mathbf{1}_k^T\mathbf{u}\ \} \tag{f14}$$

Inserting (f14) in (f8) and generalizing, for every continuous real-valued function z on $\mathbb{R}^{k-1}$:

$$\nu\{\ \mathbf{u}\in\mathbb{R}^k \mid \mathbf{1}^T\mathbf{u} > z(\mathbf{1}_1^T\mathbf{u},\ldots,\mathbf{1}_{k-1}^T\mathbf{u}) + p\ \} =$$

$$\exp(-p)\ \nu\{\ \mathbf{u}\in\mathbb{R}^k \mid \mathbf{1}^T\mathbf{u} > z(\mathbf{1}_1^T\mathbf{u},\ldots,\mathbf{1}_{k-1}^T\mathbf{u})\ \} \tag{f15}$$

so

$$\nu\{\ \mathbf{u}\in\mathbb{R}^k \mid k^{-1}\mathbf{1}^T\mathbf{u} > w,\ k^{-1}(\mathbf{1}_1^T\mathbf{u},\ldots,\mathbf{1}_{k-1}^T\mathbf{u})\in\mathcal{B}\ \} = \lambda(\mathcal{B})\ \exp(-w) \tag{f16}$$

for every Borel set $\mathcal{B}$, with λ some borel-measure on $\mathbb{R}^{k-1}$, and

$$1-\mathcal{G}(d_1^{-1}a_1,\ldots) =$$

$$\nu\{\ \mathbf{u}\in\mathbb{R}^k \mid k^{-1}\mathbf{1}_0^T\mathbf{u} > \min[\min_{j=1,\ldots,k-1} a_j - k^{-1}\mathbf{1}_j^T\mathbf{u},\ a_k + k^{-1}\mathbf{1}_1^T\mathbf{u} + \ldots + k^{-1}\mathbf{1}_k^T\mathbf{u}\ \} =$$

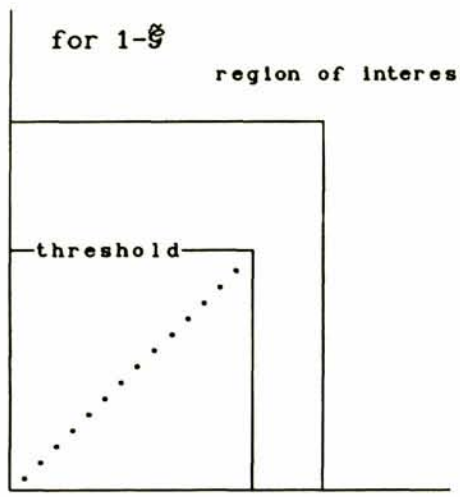
$$\int_{\mathbb{R}^{k-1}} \exp(-\min[a_1-x_1,\ldots,a_{k-1}-x_{k-1},a_k+x_1+\ldots+x_{k-1}])\ d\lambda(\mathbf{x}) \tag{f17}$$

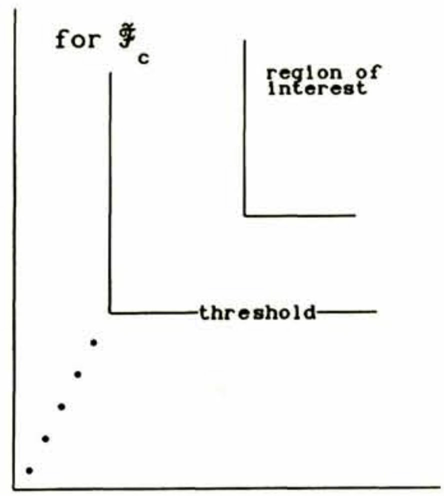
and (f5) follows. The other results follow in the same way as in appendix c. In the two-dimensional case:

$$1-\mathcal{G}(a_1,a_2) = \int_{\mathbb{R}} \exp(-\min[d_1a_1-x,d_2a_2+x])\ d\lambda(x) \tag{f18}$$

Note that there is a slight error in [*De Haan*, 1990, the equation following his equation (22)].

It is interesting to compare $1-\mathcal{G}$ with $\mathcal{F}_c$ of appendix c.
Since the scaling of the marginals is identical, we discuss only $1-\mathfrak{G}$ and $\mathfrak{F}$.
The thresholds defining $1-\mathfrak{G}$ and $\mathfrak{F}_c$ are of the form





This means that $1-\mathcal{G}$ is 'tied' to its marginals, but $\mathcal{F}_c$ is not. Information about correllation between variables is lost in computing $1-\mathcal{G}$, except if the correllation is practically complete. The different character of the two statistics is best explained with an example from the discrete-time case, assuming sequences of independent random vectors in $\mathbb{R}^2$ with independent components which have the unit exponential distribution. In that case,

$$\mathcal{F}_c(a_1,a_2) = \exp(-a_1-a_2) \tag{f19a}$$

and

$$1-\mathcal{G}(a_1,a_2) = (\exp(-a_1) + \exp(-a_2))/2 \tag{f19b}$$

An application for $1-\mathcal{G}$ in coastal engineering is for example to compute the probability that the high tide level somewhere along the coast exceeds a critical level in an arbitrary year, with "somewhere" meaning at one coastal location in a given set of locations. This statistic is naturally generalized to the distribution of the maximum of a variable over a certain time-interval and spatial region.
Applications of $\mathcal{F}_c$ are different. $\mathcal{F}_c$ applies to events like the entrance of regions in the parameter space of rather general shape, as discussed in section 4.2, such as a region that corresponds to failure of a construction.

$\mathcal{G}$ is always a distribution function corresponding to a probability measure on the plane. However $\mathcal{F}_c$ need not to be a proper survivor function.

Appendices c and f can be extended to obtain "classical" extreme value statistics, which give the probability of (joint) exceedance in a time-interval of n years as in (b9) but with the number of years n increasing instead of keeping it fixed, in such a way that μn remains bounded. This requires one more assumption about asymtotic independence of nonoccurrence of an 'extreme' event in disjoint time-intervals of an integer number of years. It is not relevant for the present application (see also section 4.1).

main office
Rotterdamseweg 185
p.o. box 177
2600 MH Delft
The Netherlands
telephone (31) 15 - 56 93 53
telefax (31) 15 - 61 96 74
telex 38176 hydel-nl

location ' De Voorst'
Voorsterweg 28, Marknesse
p.o. box 152
8300 AD Emmeloord
The Netherlands
telephone (31) 5274 - 29 22
telefax (31) 5274 - 35 73
telex 42290 hylvo-nl