TRANSNAV
http://www.transnav.eu
the International Journal on Marine Navigation and Safety of Sea Transportation
Volume 18
Number 4
December 2024

DOI: 10.12716/1001.18.04.13

Review of Ocean Wave's Uncertainties for Navigators

Z. Kalman
University of Zagreb, Zagreb, Croatia

ABSTRACT: The uncertainties of marine environments lastingly challenge navigation and safety of sea transportation. Therefore, the article tackles the extraction, assessment, and analysis as well as the perceptive presentations of probabilistic uncertainties of the random wind waves ocean-wide. The link of the probabilistic uncertainties and statistical variabilities is accomplished in the article by using the reported Global Wave Statistics of coherent and controlled wind wave data visually observed from ships in normal service. The probabilistic uncertainty is defined in the information theory most coherently with the human experience of randomness by the information entropy. The article reveals expressions, tables, graphs, and charts of information entropy which objectively express the uncertainties of observed wind wave directions, heights, and periods in all principal ocean areas. The combinations of areal entropy provide uncertainties of wider ocean zones, sectors, and shipping routes for the assessment of all-around exposures of ships and other objects in service at seas to random wind wave effects appropriately to sea-men's experience of randomness.

1 INTRODUCTION

The article considers the relevance of the information theory for the presentation of aleatory uncertainties of observable random wind wave properties. The study entrusts in the sequel on the information entropy measure for environmental probabilistic uncertainty assessment as well as on the Global Wave Statistics [8] (GWS) (Hogben, Dacunha, and Olliver, 1986) and online [3] BMT databases (2021). The article demonstrates in the sequel how to derive the information entropy from all tabulated annual and seasonal GWS ocean areas. The wave data sets contain observed frequencies of wave heights and periods for all principal wave directions. At the end, the results and examples are presented in the article by mathematical expressions, tables, graphs, and charts.

The article at the beginning refers to the information entropy concept that emerged earlier in the information theory [7,18,13] (Hartley, 1928; Wiener, 1948; Shannon & Weaver, 1949). The information entropy was generalized later in the probability theory as the probabilistic uncertainty measure [9,12,1,16] (Khinchin, 1957; Renyi, 1970; Aczel and Daroczy, 1975; Tsallis, 1988) and also recognized as information based on uncertainty [10] (Klir and Wierman, 1999). The usefulness of the information entropy concept in the probability theory for objective uncertainty assessment of various systems of random events in engineering and elsewhere was investigated earlier [20] Ziha (2000a). The information entropy may also appropriately express the redundancy and robustness of different operational and failure system states [20] Ziha (2000b). The entropy of marginal distributions was earlier considered for the uncertainty assessment of ocean wind waves [21] (Ziha, 2007). The probabilistic uncertainty can also be perceptively presented by average numbers of events or outcomes closer to the experience of gambling [23] (Ziha, 2013). More recently the information theory was applied to assessments of integral system safety in engineering [24] (Ziha, 2022).

Additional information to mere statistics of ocean wind waves provide a novel prospect for more objective assessments of the effects of uncertainties of maritime environments on ships and other ocean objects, structures and vessels.

2 SETS OF OBSERVATIONS OF RANDOM EVENTS

A discrete distribution $\boldsymbol{P}_N$ of N probabilities p_i, i=1,2, ...,N of outcomes of random events is presented as:

$$\mathbf{P}_N = (p_1, p_2, \cdots, p_N) \tag{1}$$

Distributions of N probabilities $\boldsymbol{P}_N$ (1) can also be interpreted as systems $\boldsymbol{S}_N$ of N random events E_i with probabilities p_i, i=1,2, ...,N.

The disjoint random events E_j configure a system $\boldsymbol{S}_N$ that can be written in a form of an N-element finite scheme (Khinchin, 1957) as follows:

$$\mathbf{S}_N = \begin{pmatrix} E_1 & E_2 & \cdots & E_j & \cdots & E_N \\ p(E_1) & p(E_2) & \cdots & p(E_j) & \cdots & p(E_N) \end{pmatrix} \tag{2}$$

The probability of N probabilities $\boldsymbol{P}_N$ (1) of a distribution or of a system of N events $\boldsymbol{S}_N$ (2) is then:

$$p_N(\mathbf{P}) \, or \, p_N(\mathbf{S}) = \sum_{i=1}^{N} p_i \le 1 \tag{3}$$

In (3), for complete distributions is $p_N(\mathbf{P})=1$ and for complete systems is $p_N(\mathbf{S})=1$.

3 VARIABILITY OF PROBABILITY DISTRIBUTIONS

Variability of a single random outcome of a probability distribution (1) can be experienced by an equivalent number ν_i of outcomes with hypothetically equal probabilities pi out of N possible in system $\boldsymbol{S}_N$ as defined:

$$\nu(p_i) = p_N(\mathbf{P}) / p_i = 1 / p_i \tag{4}$$

The following terms related to variability apply both for partial and complete distributions $\boldsymbol{P}_N$ (1) (3).

The mean value of a probability distribution $\boldsymbol{P}_N$ (1) is by definition:

$$\overline{p}_N = \overline{p}_N(\mathbf{P}) = (1/N) \cdot p_N(\mathbf{P}) = 1/N \tag{5}$$

The statistical variability of N probabilities p_i of a distribution $\boldsymbol{P}_N$ (1) is given by the average variance $V_N(\boldsymbol{P})$:

$$V_N(\mathbf{P}) = \sigma_N^2(\mathbf{P}) = (1/N) \cdot \sum_{i=1}^{N} (p_i - \overline{p}_N)^2 = (1/N) \cdot \sum_{i=1}^{N} p_i^{\,2} - \overline{p}_N^2 \tag{6}$$

The standard deviation of probabilities of distribution of probabilities using (6) is then:

$$\sigma_N(\mathbf{P}) = \sqrt{V_N(\mathbf{P})} \tag{7}$$

The reference value of variance can be calculated (Ziha, 2013) as the limiting value of (6) under the condition that one probability $p_j \to p_N(\mathbf{P})$ is dominant and all the others $p_{i \ne j} \to 0$ are vanishing:

$$V_{\text{Ref}}(\mathbf{P}_N) = p_N(\mathbf{P}) \cdot (N-1)/N^2 = (N-1)/N^2 \tag{8}$$

The coefficient of variation of N probabilities of a distribution PN (1) from (5 and 7) (Table 1) is:

$$CV_N(\mathbf{P}) = \sigma_N(\mathbf{P}) / \overline{p}_N = N \cdot \sigma_N(\mathbf{P}) \tag{9}$$

The reference value of the coefficient of variation follows from (9) as shown:

$$CV_{\text{Ref}}(\mathbf{P}) = \sqrt{p_N(\mathbf{P}) \cdot (N-1)} = \sqrt{N-1} \tag{10}$$

4 INFORMATION ENTROPY OF SYSTEMS OF EVENTS

The information entropy of a single event E_i in a system $\boldsymbol{S}_N$ (2) can be interpreted according to [18] Wiener (1948) either as a measure of the information yielded by the event or how unexpected the event was and can be defined based on the equivalent number of outcomes (4) $\nu(p_i)$ as follows:

$$\mu(E_i) = H(E_i) = \log_2 \nu(p_i) = \log_2 [1/p_i] = -\log_2 p_i \tag{11}$$

The information based on the uncertainty of a complete system $\boldsymbol{S}_N$ (2) of N events is the weighted sum of the unexpectednesses (11) expressed by the Shannon's entropy [13] (Shannon and Weaver, 1949) (Table 1):

$$H_N(\mathbf{S}) = \sum_{j=1}^{N} p_j \cdot \mu_j = -\sum_{j=1}^{N} p_j \log p_j \tag{12}$$

Shannon's entropy (12) has properties of uncertainty: continuity in its arguments, monotonic increase with a number of equiprobable outcomes and a composition rule [6] (Cover and Thomas, 2006). The information based on the uncertainty of an incomplete system SN (2) can be defined as the limiting case of the Renyi's entropy of order one (Renyi, 1970), using the probability of a system of event (3) and definition of the Shannon's entropy:

$$H_N^{R1}(\mathbf{S}) = -\frac{1}{p(\mathbf{S})}\sum_{j=1}^{N} p_j \log p_j \qquad (13)$$

Tsallis (1988) extended the information entropy formalism to the [12] Renyi's entropy (1970) (13) of order one.

Shannon's axiomatic derivation explains [13] (1949) why the entropy is an intuitive measure of uncertainty. The uniqueness theorem by [9] Khinchin (1957) states that the information entropy is the only function that measures the probabilistic uncertainty in agreement with experience.

Let us consider system S of L subsystems of L_i elements $\mathbf{S}_i = (s_{i1}\ s_{i2}\ \cdots\ s_{iL_i})$, i=1,2,…,$L_i$ (Ziha, 2000).

The probabilities of i^{th} subsystems are $p(\mathbf{S}_i) = s_{i1} + s_{i2} \cdots + s_{iL_i}$ and the conditional information entropy of the system S concerning for the subsystems S_i is as follows:

$$H(\mathbf{S}/\mathbf{S}_i) = -\sum_{k=1}^{L_i} \frac{s_k}{p(\mathbf{S}_i)} \cdot \log \frac{s_k}{p(\mathbf{S}_i)} \qquad (14)$$

The information entropy of the system of subsystems $\mathbf{S}' = (\mathbf{S}_1\ \mathbf{S}_2\ \cdots\ \mathbf{S}_L)$ can be presented as shown:

$$H_L(\mathbf{S}') = -\sum_{i=1}^{L} p(\mathbf{S}_i) \cdot \log p(\mathbf{S}_i) \qquad (15)$$

The theorem of mixture of distributions following [9] Khinchin (1957) and [12] Renyi (1970) provides the conditional information entropy of system S concerning for the system of subsystems S' using (12-15) (proof in APPENDIX B), as follows

$$H(\mathbf{S}/\mathbf{S}') = \frac{1}{p(\mathbf{S})} \cdot \sum_{j=1}^{N} p(\mathbf{S}_j) \cdot H(\mathbf{S}/\mathbf{S}_j) = \\ = \frac{1}{p(\mathbf{S})}\left[H_N(\mathbf{S}) - H_L(\mathbf{S}')\right] = H_N^{R1}(\mathbf{S}) - H_L^{R1}(\mathbf{S}') \qquad (16)$$

The additional knowledge of subdivision of system S to subsystems S' reduces unconditional information based on uncertainty (13) for an amount of (15). The incompleteness $p(\mathbf{S}) \le 1$ increases the system uncertainty (16). Conditional information (16) of system S may be viewed as the average entropy of subsystems S'.

The information entropy $H_N(S)$ is equal to zero when the state of the system S can be surely predicted, i.e., no uncertainty exists at all. This occurs when one of the probabilities of events p_i, i=1,2,...,N is equal to one, let us say p_k=1 and all other probabilities are equal to zero, p_i=0, $i \ne k$.

The information entropy is maximal for uniform distribution of event probabilities when all the events are equally probable with the probability equal to p_i= 1/N, for i=1, 2, ..., N, and it amounts to $H_N(S)_{max}$=log$N$ that is the [7] Hartley's entropy (1928). Hartley's entropy relates to Renyi's entropy of order 0.

The entropy increases as the number of events increases. The entropy does not depend on the sequence of events. The definition of the unit of information based on uncertainty measure according to [12] Renyi (1970) is not more and not less arbitrary than the choice of the unit of some physical quantity.

If the logarithm applied in (12-16) is of base two, the unit of information entropy is denoted as one "*bit*". One bit is the uncertainty of two equally probable events (a simple alternative) like flipping an ideal coin. If the natural logarithm is applied, the unit is one "*nit*". Relative measures for entropy $h_N(S)$=$H_N(S)$/log$N$ may be useful. Interpretations of statistical variability vs. probabilistic uncertainty are given in Table 1.

Table 1. Statistical variability vs. probabilistic uncertainty (information entropy)

Variability (Statistics)
Coefficient of variation of probabilities (9) $0 \le CV_N(\mathbf{P}) = \sqrt{N \cdot \sum_{i=1}^{i=N} p_i^2 - 1} \le \sqrt{N-1}$ Min: 0 – Fully invariable (*All N outcomes equally probable*) Max: $\sqrt{N-1}$ – Maximal variability (*One sure outcome all N-1 others impossible*) Unit: 1– (*One sure, one impossible N*=2) *1/p_i equivalent number of equiprobable random event*
Probabilistic uncertainty (Information theory)
The information entropy of system (12) $0 \le H_{N,b}(\mathbf{S}) = -\sum_{i=1}^{i=N} p_i \log_b p_i \le \log_b N$ Max: $\log_b N$ – Fully uncertain (*All N events equally probable*) Min: 0 – Maximal certainty (*One certain event, N-1 impossible*) Unit: 1 *bit* (*2 equally probable events*) *-log p_i unexpectedness of a random event*

4.1 *Average numbers of events*

It was mentioned earlier that the average number of equally probable events provides the same information as the considered system of events [1] (Aczel and Daroczy, 1975). The average number of equally probable events $F_N(S)$ follows from the condition of maximal information $H_N(\mathbf{S}) = \log_2 F_N(\mathbf{S})$ of a system of N events with average probability $1/F_N(\mathbf{S})$ as another perceptive uncertainty indicator defined as shown:

$$F_N(\mathbf{S}) = 2^{H_N(\mathbf{S})} \qquad (17)$$

Subsequently, uncertainties of random phenomena with arbitrary numbers of outcomes can be expressed by a number $F_N(S_N)$ (17) of equally probable events or outcomes. Recursively, the entropy of base B of the average number $F_N(S_N)$ (17) provides the entropy value of $\log_B[F_N(S_N)]=H_N(S_N)\log_B N$.

Relative measures for the average number of events $f_N(S)$=$F_N(\mathbf{S}_N)$/N may be useful. It is commonly perceptive that flipping an ideal coin is as uncertain as two events with the same probabilities equal to 1/2 of a system S_2=(1/2 12) with entropy $H_2(S_2)_{max}$=$\log_2 2$=1 *bit* and tossing a perfect die as six events with probabilities 1/6 of a system S_6=(1/6 1/6 1/6 1/6 1/6 1/6) with entropy $H_6(\mathbf{S}_6)_{max}$=$\log_2 6$=2.58 *bits*, which is equivalent to flipping of 2.58 coins. Since entropy (12-16) are, in general, real numbers, so are the average numbers of equally probable events $F_N(S)$ (17), providing continuous scales for interpretations of uncertainties (Ziha, 2013). More generally, the information entropy n using a logarithm of base b provides average numbers b^n of equally probable events (17) as gambling with n-sided ideal objects. Note how the numbers of equiprobable outcomes or events are generalized to not only integer values in a continuous scale.

5 INFORMATION AND UNCERTAINTY OF THE GWS

Visual observations of commercial ships have been archived since 1861. From 1961 the collection is systematized according to a resolution of the WMO. The compilation of these observations for each of the N_A=104 Marsden's squares (Appendix A, Fig. A1) is the Global Wave Statistics (GWS) [8] (Hogben, Dacunha and Olliver, 1986) that uses the past experiences to eliminate biases.

An introductory example presents the information based on uncertainty and statistical variability of observed calm p_{calm} and wavy p_{wavy} periods of sea states and unexpectedness μ (11) (Table 2).

The GWS integrated the wind/wave climate observations on the global level in scatter diagrams of joint distributions of wave heights against wind speed in terms of the Beaufort scale and separate sets of normalized wind frequencies. However, the GWS does not account for local climate conditions such as the size, the topography within/surrounding the region, the fetch, and ocean surface currents. Some of the local effects are directly or indirectly assessable from the attached GWS charts. Monthly frequency tables of wave heights and wind forces against the direction, together with information on ice conditions and the occurrences of tropical cyclones were used to decide upon seasonal subdivisions in the GWS. The advantages of GWS [4] (Choi and Hirayama, 2000) are the global approach, the duration of the collection period, and its suitability to maritime applications. The drawbacks are the lack of local wave/wind climate information particularly outside the oceanic areas, and the poor accuracy for periods, where heights are well estimated and enhanced by experienced observers. The accuracy of the Global Wave Statistics Data is checked with specific instruments [2] (Bitner-Gregersen and Cramer, 1994). The basic World Wide Waves Statistics (WWWS) consists of wave model time series for a great number of positions calibrated against Topex, Jason, or other satellite data such as the Atlas of the oceans: wind and wave climate from the GEOSAT satellite [19] (Young and Holland, 1996), The GWS often serves as a reference guide for wave data reported by other sources [11] (Nielsen and Ikonomakis, 2021). For long-term predictions of ship responses in ocean a practical method for comparison of GWS with other wave data was proposed [14] (Shinkai and Wan, 1996).

5.1 *Uncertainties of GWS areas*

The GWS wave properties data sets $A{:}s,d$ are available for each of 104 ocean area A and for N_s=4+1 annual and seasonal observations denoted s=(annual), ***M**arch-**M**ay*, ***J**une-**A**ugust*, ***S**eptember-**N**ovember*, ***D**ecember-**F**ebruary* as well as for N_d=8+1 wave directions denoted d=(*all*), *NW*, *N*, *NE*, *W*, *E*, *SW*, *S*, *SE*. Each data set $A{:}s,d,h,t$ present N_h=15 significant wave heights from 0 meters to >14 meters and N_t=11 zero-crossing wave periods from <4 seconds to >13 seconds, (for example Fig. 1 and Table 3 for A25:*s=annnual*,*d=all*).

The GWS ocean wave data sets in tabular form contain the jointly observed frequencies of wave heights h and periods t per 1000 of observations as $p_{A:s,d}(h,t)$. The unconditional information of a data set $A{:}s,d,h,t$ as complete systems (2) is defined by the Shannon's entropy of all observations (12) (Table 3):

$$H(A:s,d,h,t) = -\sum_{all\ h}\sum_{all\ t} p_{A:s,d}(h,t)\cdot\log p_{A:s,d}(h,t) \tag{18}$$

The unconditional information entropy of the marginal distributions [21] (Ziha, 2007) of all heights $h=h_w$ $p_{A:s,d}(h_w)=\sum_{all\,t} p_{A:s,d}(h_w,t)=1$ and for periods $t=t_z$ $p_{A:s,d}(t_z)=\sum_{all\,h} p_{A:s,d}(h,t_z)=1$ of the data set $A{:}s,d$ (Fig. 1, Table 3) is:

$$H(A:s,d,h) = -\sum_{all\,h_w} p_{A:s,d}(h_w)\cdot\log p_{A:s,d}(h_w) \tag{19}$$

$$H(A:s,d,t) = -\sum_{all\,t_z} p_{A:s,d}(t_z)\cdot\log p_{A:s,d}(t_z) \tag{20}$$

In addition to the unconditional entropy of the marginal distribution of heights (19) and periods (20), the conditional entropy with respect to a selected wave height h_w and period t_z is defined in (APPENDIX C).

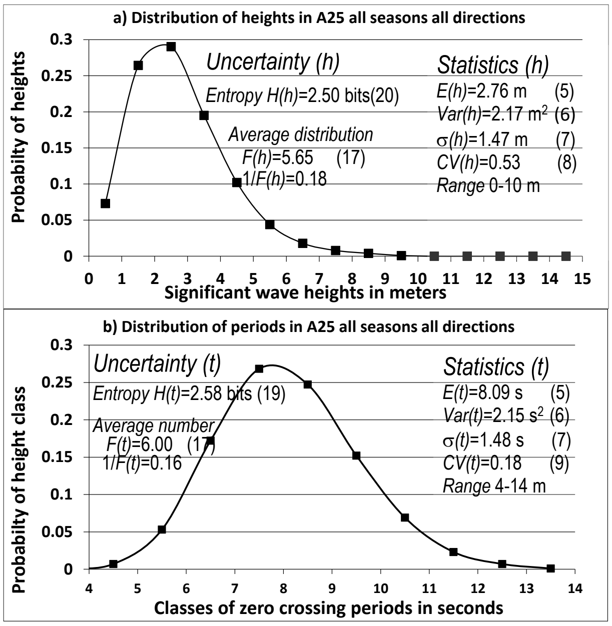


Figure 1. Distributions of wave heights (a) and periods (b) in GWS area A25

Table 2. Variability and uncertainty (information) of calm and wavy periods of sea states

p_{calm}	$\mu(p_{calm})$	p_{wavy}	$\mu(p_{wavy})$	Condition	CV(9)	H(12)	F(17)	Variability	Uncertainty
0	∞	1	0	Always wavy	1	0	1	Maximal	Certain
1/10	3.32	9/10	0.15	Prevailing wavy	0.82	0.47	1.38	High	Low
1/4	2	3/4	0.41	1/4 time calm 3/4 time wavy	0.5	0.81	1.75	Moderate	Moderate
1/2	1	1/2	1	1/2 time calm 1/2 time wavy	0	1	2	Invariable	Uncertain
4/5	0.3	1/5	2.32	3/4 time calm 1/t time wavy	0.6	0.72	1.65	Moderate	Moderate
19/20	0.07	1/20	4.32	Prevailing calm	0.90	0.29	1.22	High	Low
1	0	0		Always calm	1	0	1	Maximal	Certain

Table 3. GWS data set A25:*s=annual*, *d=all* of wave heights and zero-crossing periods (SUPPLEMENT 1)

$H(t)$(20)	2.583	$p_{A:s,d}(h_t)$	0.007	0.053	0.172	0.268	0.247	0.152	0.069	0.023	0.008	0.001	Σ=1
	>14					A25:*s=annual*, *d=all*			t_z				
	13-14					H(A25)=4.85 *bits* (18)							
	12-13					F(A25)=29 *ane* (17)							
	11-12					H(A25:h)=2.50 *bits*(19)							$p_{A:s,d}(h_w)$
	10-11					H(A25:t)=2.58 bits (20)							Σ=1
1	9-10					$p_{A:s,d}(h,t)$		0.001					0.001
4	8-9						0.001	0.001	0.001	0.001			0.004
5	7-8					0.001	0.002	0.002	0.002	0.001			0.008
6	6-7					0.002	0.004	0.005	0.004	0.002	0.001		0.018
7	5-6	h_w			0.002	0.006	0.012	0.012	0.008	0.003	0.001		0.044
9	4-5			0.001	0.006	0.019	0.029	0.025	0.014	0.005	0.002	0.001	0.103
8	3-4			0.002	0.018	0.048	0.059	0.041	0.019	0.006	0.002		0.195
8	2-3			0.008	0.046	0.089	0.082	0.044	0.016	0.004	0.001		0.290
8	1-2		0.002	0.022	0.073	0.088	0.053	0.020	0.005	0.001			0.264
6	0-1		0.005	0.020	0.027	0.015	0.005	0.001					0.073
N_t=10	h / t	<4	4-5	5-6	6-7	7-8	8-9	9-10	10-11	11-12	12-13	>13	2.497
N_{ht}=62	N_h=10		2	5	6	8	9	10	8	8	5	1	$H(h)$(19)

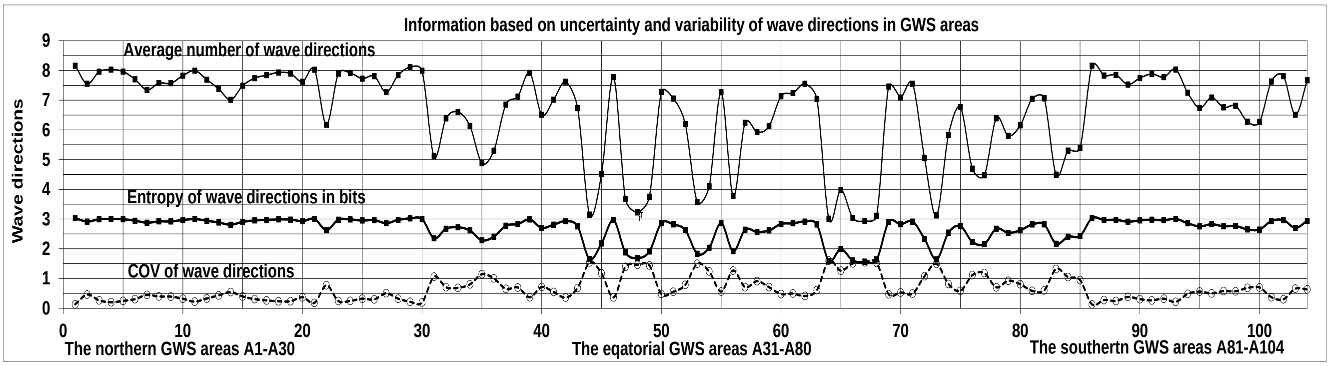


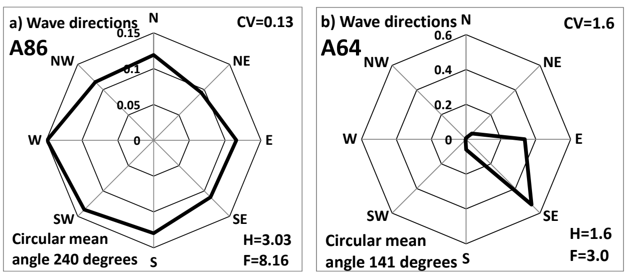
Figure 2. Information and variability of all wave directions on annual basis (21) (SUPPLEMENT 2)

5.2 *Uncertainties of wind wave directions of all GWS areas*

The entropy of a GWS area A with respect to the probabilities $p_{A,d}$ of eight wave directions can be assessed as an incomplete system by the unconditional [12] Renyi's entropy (13) where $p_A=\Sigma p_{A,d} \leq 1$ (Fig.2), as follows:

$$H^{R1}(A:s,d) = -\frac{1}{p_A} \cdot \sum_{all\,d} p_{A,d} \cdot \log p_{A,d} a \tag{21}$$

Circular statistics [5] (e.g. Fisher, 1993) provides the circular mean angle and coefficient of variation. Note how the high entropy H (12) and low coefficient of variation CV (9) indicate high uniformity of wave directions in GWS area A86 (Fig. 2a) compared to A64 (Fig. 2b).



Figures 2a and 2b. Statistics of wind wave directions in A86 and A64 GWS areas

5.3 *Uncertainties of combinations of GWS data sets*

The combinations of GWS data require the information of all GWS areas (18, 19, 20) (SUPPLEMENT 3).

A combination D consists of k GWS data sets $A_{i:s,d}$, i=1,2,…,k, of areas A, seasons s, and directions d.

The relative participations of all data sets $p_A=p_{A1:s,d}+p_{A2:s,d}+\ldots+p_{Ak:s,d}\leq 1$ provide the unconditional Renyi's entropy (13) of the area *A*, (e.g. Table 4, for A25:*annaul,all, NW, N, NE, W, E, SW, S, SE*), as follows:

$$H^{R1}(A) = -\frac{1}{p_A}\cdot\sum_{i=1}^{k} p_{A_i:s,d}\cdot\log p_{A_i:s,d} \qquad (22)$$

The conditional information entropy of *k* combined data sets of the combination *D* follows from the theorem of mixtures of distributions (16) with respect to aggregates *A* of all partaking areas A_i with participations $p_{A:s,d}$ that is the average information of jointly observed heights and periods (18) of all components, as shown:

$$H^{R1}(D/h,t) = \frac{1}{p_A}\cdot\sum_{i=1}^{k} p_{A_i:s,d}\cdot H(A_i:s,d,h,t) \qquad (23)$$

The average or conditional entropy of marginal distributions of heights $H(D/h)$ and of periods $H(D/t)$ of the combination D of the conditional entropy of heights $H(A:s,d,h)$ (19) and periods $H(A:s,d,t)$ (20) reported in GWS are as shown:

$$H^{R1}(D/h) = \frac{1}{p_A}\cdot\sum_{i=1}^{k} p_{A_i:s,d}\cdot H(A:s,d,h) \qquad (24)$$

$$H^{R1}(D/t) = \frac{1}{p_A}\cdot\sum_{i=1}^{k} p_{A_i:s,d}\cdot H(A:s,d,t) \qquad (25)$$

The overall or unconditional information of a combination *D* of all jointly observed probabilities of heights and periods $p_{Ai:s,d}(h,t)$ in proportion to the relative participations of components $p_{Ai:s,d}$, is equal to the sum of the unconditional (22) and the conditional information (23), as shown:

$$H^{R1}(Dht) = -\frac{1}{p_A}\sum_{i=1}^{k}\left[p_{Ai:s,d}\cdot p_{Ai:s,d}(h,t)\right]\cdot\log\left[p_{Ai:s,d}\cdot p_{Ai:s,d}(h,t)\right] = H^{R1}(D/h,t)+H^{R1}(A) \qquad (26)$$

The overall or unconditional entropy of marginal distributions of heights (24) and of periods (25) implies the unconditional information $H^{R1}(A)$ (22) according to the theorem of mixtures of distribution (16) as shown:

$$H^{R1}(Dh) = H^{R1}(D/h)+H^{R1}(A) \qquad (27)$$

$$H^{R1}(Dt) = H^{R1}(D/t)+H^{R1}(A) \qquad (28)$$

Succinctly, the combinations *D* of selected GWS data sets (26-28) comprise the average uncertainties of all componential data sets (23-25) augmented by the overall area *A* uncertainty *H*(*A*) (22).

Table 4. Information on combination of all wave directions for A25:s=annual,d= all (SUPPLEMENT 4)

d	N_{ht} GWS	$p_{A25:d}$ GWS	*H(A:h,t)bits GWS (18)*	*F(A:h,t)* 2^H	*Nh GWS*	*H(A:h) bits GWS (19)*	*F(A:h)* 2^H	N_t GWS	*H(A:t) bits GWS (20)*	*F(A:t)* 2^H
NW	66	0.1288	4.90	29.9	10	2.55	5.9	10	2.58	6.0
N	57	0.1770	4.67	25.5	9	2.35	5.1	10	2.53	5.8
NE	48	0.1887	4.39	20.9	8	2.19	4.6	8	2.38	5.2
W	66	0.1251	4.89	29.7	11	2.60	6.0	9	2.51	5.7
E	47	0.0914	4.32	19.9	8	2.16	4.5	8	2.33	5.0
SW	55	0.1046	4.53	23.0	10	2.36	5.1	8	2.35	5.1
S	50	0.1043	4.37	20.7	9	2.27	4.8	7	2.28	4.9
SE	44	0.0638	4.27	19.3	8	2.17	4.5	7	2.26	4.9
GWS all	433	p_A=0.9838	4.86	28.9	73	2.50	5.6	67	2.58	6.0
Conditional	54		(23) 4.57	23.7	9	(24) 2.34	5.1	8	(25) 2.42	5.36
Unconditional			(26) 7.52	183		(27) 5.29	39		(28) 5.37	41

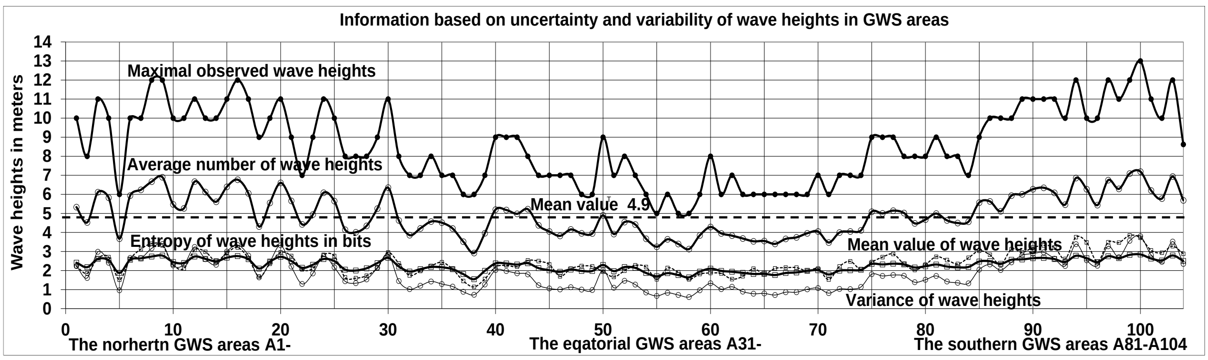


Figure 3. Information and variability of all wave periods on annual basis (19)

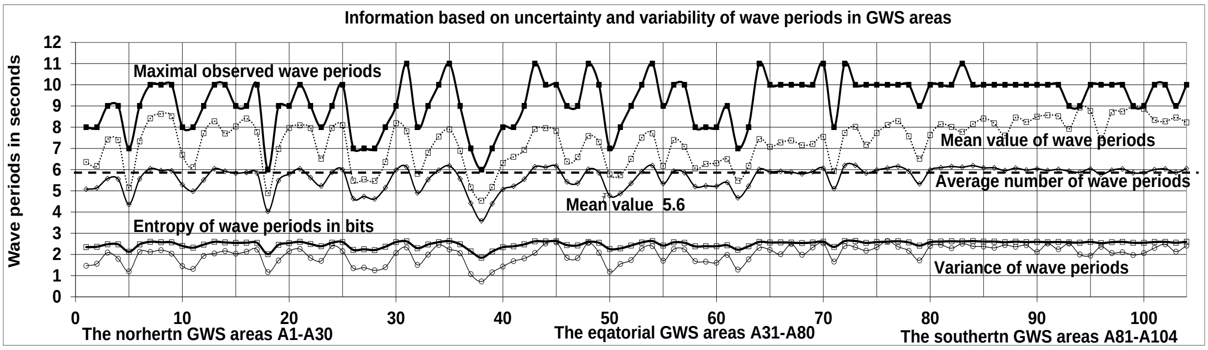


Figure 4. Information and variability of all wave periods on annual basis (20)

6 TRACING THE OCEAN-WIDE INFORMATION BASED ON UNCERTAINTIES

The article also presents the statistical means (5), variances (6) and maximal observed values as well as the entropy in bits and the average numbers (17) of wave heights (19) (Fig. 3) and periods (20) for all wave directions on annual basis (Fig. 4) (SUPPLEMENTS from A1-a-a to A104.a.a).

The study next presents the unconditional information (22), the observed maximal values (GWS), conditional and unconditional information of wave heights/periods (23, 26) as well as of heights (24, 27) and periods (25, 28) of the entire GWS and of zones of combinations *D* of selected GWS areas (Table 5).

The study also presents the unconditional information (22), the observed maximal values (GWS), conditional and unconditional information of wave heights/ (23, 26) as well as of heights (24, 27) and periods (25, 28) of the oceanic GWS zones of combinations *D* of selected GWS areas (Table 6).

The article also presents the charts of average numbers (17) based on information entropy of wave directions (21), jointly observed heights/periods (18), wave heights (19), and wave periods (20) on annual basis for wave directions in all 104 GWS areas (Appendix A, charts). (SUPPLEMENT 6)

The annual numbers of wind wave directions (21) are given in categories 3 to 8 (Appendix A, Fig. A2). The minimal number of wind-wave directions in the amount of 2.94 directions is encountered between the Brazilian and African coast in the Mid-Atlantic areas [A66-A68] where the E and SW wind-wave directions prevail with about 90%. The maximal numbers of wind-wave directions can even exceed the nominal amount of 8 due to reported incompleteness of observations (21) in areas where the wave directions are almost uniformly distributed, for example in the North Atlantic [A1, A4] and in the South Pacific areas [A86, A93].

The annual numbers of wave heights (19) are categorized from 2 to 7 (Appendix A, Fig. A3) in the range from 2.9 in the Persian Gulf [A38] to 7.2 in the south Indian Ocean [A99, A100] concerning the mean value of 4.9 in the whole GWS (Table 5, Fig. 3). The high uncertainty categories of 6-7 characterize the North Atlantic areas [A3, A8, A9, A15, A16, A24]. Lower and moderate wave height uncertainty categories are in gulfs and bays, for example, 2.9 in the Persian Gulf [A38], 3.1 in the Gulf of Guinea [A58], 3.5 in the Red Sea [A37], 3.9 in the Gulf of Mexico [A32] and in the Bay of Bengal [A51]. In the sea areas prevail the wave height category around 4, like 3.8 in the Philippine Sea [A52], 4 in the Arabian Sea [A39], 4-4.1 in the Mediterranean Sea [A26-A27], 4.2 in the Caribbean Sea [A47], 4.3 in the Sea of Japan [A18] and in the Yellow Sea [A28]. The wave height categories of 3.5-3.7 characterize the seas around the islands of Micronesia [A63] and Melanesia [A71].

The annual numbers of wave periods (20) are alienated in categories 4, 5, and 6 (Appendix A, Fig. A4) in the range from 3.6 in the Persian Gulf [A38] to maximally 6.2 in the mid-Pacific [A76-A77] concerning the 5.75 mean value of whole GWS (Table 5, Fig. 4). Wave period categories below 4 characterize only some gulfs and coastal seas, for example, 3.6 in the Persian Gulf [A38] and 4.0 in the Sea of Japan [A18].

Table 5. Information on all GWS areas and zones of selected ocean areas (SUPPLEMENT 5)

All GWS and ocean zones D	*% GWS*	*F(A) (22)*	*Max GWS*	*F(D/h,t) (23)*	*F(Dht) (26)*	*Max GWS*	*F(D/h) (24)*	*F(Dh) (27)*	*Max GWS*	*F(D/t) (25)*	*F(Dt) (28)*
All GWS areas [A1-A104]	100	91.4	34.2	24.2	2218	7.2	4.9	448	6.5	5.8	526
Northern areas [A1-A30]	21.7	25.1	34.0	27.0	677	6.9	5.6	141	6.1	5.6	140
Equatorial areas [A31-A80]	50.6	45.0	27.3	20.8	935	5.2	4.2	188	6.5	5.7	257
Southern areas [A81-A104]	27.7	22.1	34.2	29.7	656	7.2	6.0	131	6.2	7.0	132

Table 6. Information on zones of selected ocean areas

*Ocean zones D (83% of the GWS) *some coastal areas not included*	*% GWS*	*F(A) (22)*	*Max GWS*	*F(D/h,t) (23)*	*F(Dht) (26)*	*Max GWS*	*F(D/h) (24)*	*F(Dh) (27)*	*Max GWS*	*F(D/t) (25)*	*F(Dt) (28)*
North Atlantic Ocean (20 areas*)	15.5	19.0	34.2	24.8	473	6.9	5.0	96	6.5	5.7	109
South Atlantic Ocean (12 areas*)	12.8	8.8	33.1	25.6	289	6.75	5.0	57	2	6.0	67
North Pacific Ocean (17 areas*)	21.3	15.9	32.5	23.8	379	6.9	4.7	75	6.2	5.9	94
South Pacific Ocean (12 areas*)	13.2	10.7	32.8	25.8	276	6.8	5.0	53	6.2	6.0	65
Indian Ocean (16 areas*)	20.2	14.5	34.2	26.0	378	6.3	5.3	77	6.1	5.8	84

Table 7. Information of some northern hemisphere navigation routes (SUPPLEMENT 7)

R	Route description	GWS A:annual,all	H(A) (22); F(A)	F(D/ht)(23); F(h,t)(26)	F(D,h)(24); F(h)(27)	F(D,t)(25); F(t) (28)
A	English Channel-Gibraltar	A11 A16 A17	1.48; 2.80	27.1; 75.9	5.84; 16.3	5.4; 15.1
B	Gulf of Mexico-Gibraltar	A24 A25 A32 A33	1.86; 3.62	25.6; 92.7	5.22; 14.6	5.6; 20.5
C	Gibraltar - Port Said	A26 A27	0.88; 1.84	17.2; 31.7	4.1; 11.3	4.7; 8.6
D	Suez - Aden	A37	000; 0.00	14.1; 14.1	3.5; 3.5	4.4; 4.4
E	Aden - Arabian Gulf	A39 A50	0.97; 1.96	18.5; 36.4	4.5; 12.6	4.6; 9.0
F	Arabian Gulf-Colombo	A39 A50 A60	1.52; 2.87	18.4; 52.8	4.4; 12.3	4.7; 13.5
G	Colombo - Singapore	A61 A62	0.97; 1.96	17.9; 35.1	3.9; 7.6	5.1; 10.0
H	Singapore - Taiwan	A40 A62	0.97; 1.96	19.9; 39.0	4.6; 12.9	4.9; 9.6
I	Taiwan - Japan	A29 A41	0.97; 1.96	23.6; 46.2	5.2; 14.6	5.2; 10.1
J	North Sea	A11	000; 0.00	23.0; 23.0	5.3; 5.3	5.0; 5.0
K	North Atlantic Ocean	A15 A16	0.97; 1.96	32.6; 63.8	6.6; 13.0	5.8; 11.5
L	North Pacific Ocean	A21 A22 A29 A30	1.72; 3.30	28.8; 95.0	5.7; 16.1	5.9; 19.4

Table 8. Uncertainties of some northern hemisphere combined navigation routes (SUPPLEMENT 6)

Compound route	Routes R (Table 7)	H(A) (22); F(A)	F(D/ht)(23); F(h,t)(26)	F(D,h)(24); F(h)(27)	F(D,t)(25); F(t) (28)
R1-Japan - Arabian Gulf	F-G-H-I	2.00; 4.00	43; 171	12; 32	11; 43
R2-Germany-Arabian Gulf	J-A-C-D-E	2.09; 4.25	29; 121	8; 23	7; 32
R3-Germany-Gulf of Mexico	J-A-B	1.05; 2.07	78; 162	13; 28	17; 35
R4-Gibraltar - Suez – Aden	C-D	0.99; 1.99	23; 45	7; 14	7; 13
R5-Colombo – Japan	GF-H-I	1.55; 2.94	40; 117	11; 32	10; 30

6.1 *Information based on uncertainties of navigation routes*

The effects of environmental uncertainties on navigation routes, uncertainties in collision avoidance maneuvering (Taylor, 1990), weather uncertainties in ship route optimization are of lasting interest andimportant for maritime safety [17] (Vettor, Bergamini, and Guedes Soares, 2021).

Routs normally pass through more ocean areas that may be jointly viewed as zones Z or a combination of zones D (21-28) in proportion to the length of a route or time spent in specific areas. The article recalculates the annual uncertainties of all wave directions, heights, and periods at some northern hemisphere shipping routes (Table 7).

Overall uncertainties of northern hemisphere combined navigation routes R are presented in Table 8.

Note that the entropy approach (21-28) is applicable to uncertainty assessment of navigation routes for different seasons, wave directions, heights, and periods.

For those with gambling experience, the navigation at the Mediterranean (Gibraltar - Suez – Aden) (*R4*) is uncertain concerning jointly observed heights and periods in all directions on an annual basis (26) as flipping $\log_2 171=7.41$ coins or dicing with a $\log_6 171=2.88$ six-sided die or drawing a card from a stock of 171 cards.

The navigation through the North Sea and the Atlantic Ocean from Germany to the Gulf of Mexico (*R3*) is uncertain as flipping $\log_2 162=7.35$ coins or dicing with $\log_6 162=2.84$ six sided dice or drawing from a stock of 162 cards (Table 8).

7 CONCLUSION

The article demonstrated that the probabilistic uncertainties expressed in terms of the information theory perceptively present the random properties of ocean wind wave climate data compiled in the Global Wave Statistics. The information entropy objectively defines the unexpectedness and uncertainty of ocean wind waves regarding the environmental and maritime experiences of randomness. Numerical calculations provide summarizing uncertainty formulae, tables, graphs, and charts of wind wave directions, heights, and periods in all relevant oceanic areas, zones, and routes. The ocean-wide distributions of wind-wave information based on environmental uncertainties are useful for maritime safety management, navigation, shipping, and various maritime activities as well as on exposures of ocean structures and vessels in service.

NOMENCLATURE

A	GWS areas (max 104)
CV	coefficient of variation
D	combined GWS data sets
d	GWS directiones (max 8+1)
E	directional exposure
H	information entropy (bits)
h	GWS wave heights (max 15)
F	average numbers of events, directional wave effects
μ, ν	unexpectedness, equivalent number of outcomes
N_A, N_s, N_d	Number of areas, seasons and directions
N_h, N_t,	Number of wave heights and periods
p, ***P***	probability , probability distributions
R	ocean navigation routes, robustness
S	systems of events, directional exposability
s	GWS seasons (max 4+1)
t	GWS periods (max 11)
V	variance

ACKNOWLEDGEMENT

This is curiosity-driven research of plausible application of information theory to assessments of probabilistic uncertainties of ocean-wide wind waves and their impact on navigation, shipping as well as on ocean structures and vessels in service.

APPENDIX A

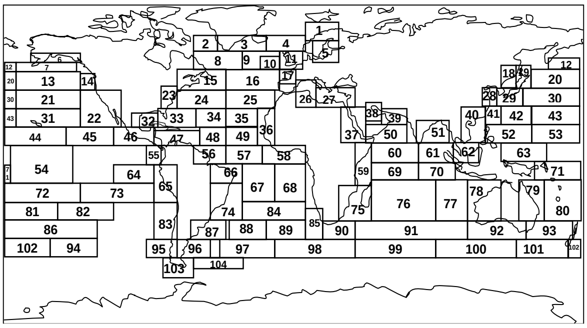


Figure A1. Marsden's squares in GWS Hogben, Dacunha and Olliver (1986)

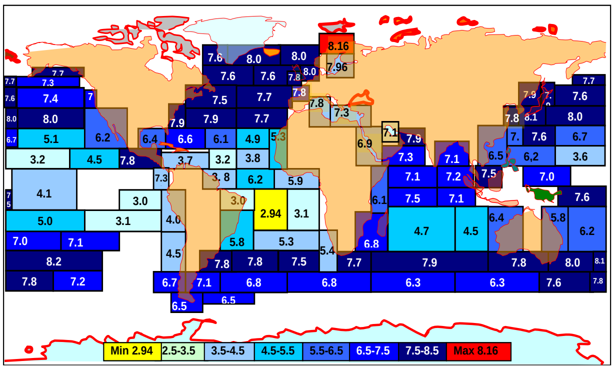


Figure A2. Average numbers of wave directions observed in GWS on annual basis

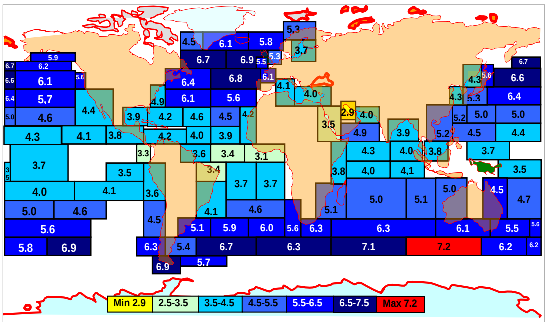


Figure A3. Average numbers of wave heights in GWS on annual basis for all wave directions

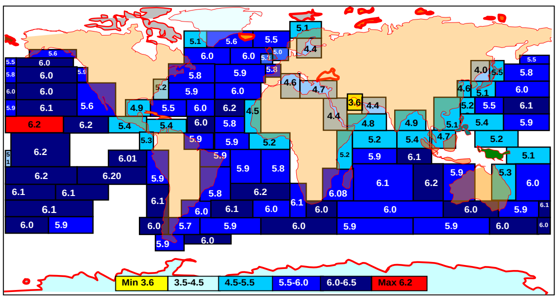


Figure A4. Average numbers of wave periods in GWS on annual basis for all wave directions

APPENDIX B

Proof of the relation between the conditional and unconditional entropies (16):

$$H(\mathbf{S}/\mathbf{S}') = -\frac{1}{p(\mathbf{S})}\cdot\sum_{j=1}^{N}\sum_{i=1}^{L} p(\mathbf{S}_i)\sum_{k=1}^{L_i}\frac{s_k}{p(\mathbf{S}_i)}\cdot\log\frac{s_k}{p(\mathbf{S}_i)} =$$

$$= \frac{1}{p(\mathbf{S})}\cdot\sum_{j=1}^{N}\sum_{i=1}^{L} - \left[\sum_{k=1}^{L_i} s_k\cdot\log s_k - \log p(\mathbf{S}_i)\cdot\sum_{k=1}^{L_i} s_k\right] = \tag{B1}$$

$$= -\frac{1}{p(\mathbf{S})}\cdot\left[\sum_{j=1}^{N} s_j\cdot\log s_j + \sum_{i=1}^{L} p(\mathbf{S}_i)\cdot\log p(\mathbf{S}_i)\right] =$$

$$= \frac{1}{p(\mathbf{S})}\left[H_N(\mathbf{S}) - H_L(\mathbf{S}')\right] = H_N^{R1}(\mathbf{S}) - H_L^{R1}(\mathbf{S}')$$

q.e.d.

APPENDIX C

Instead of the unconditional entropy of the marginal distribution of heights (19) and periods (20) the conditional entropy with respect to a selected wave height h_w and period t_z is:

$$H(A:s,d/h_w) = -\sum_{all\,t}\frac{p_{A:s,d}(h_w,t)}{p_{A:s,d}(h_w)}\log\frac{p_{A:s,d}(h_w,t)}{p_{A:s,d}(h_w)} \tag{C1}$$

$$H(A:s,d/t_z) = -\sum_{all\,t}\frac{p_{A:s,d}(h,t_z)}{p_{A:s,d}(t_z)}\log\frac{p_{A:s,d}(h,t_z)}{p_{A:s,d}(t_z)} \tag{C2}$$

Subsequently, the conditional entropy with respect to all wave heights h and for periods t is:

$$H(A:s,d/h) = \sum_{all\,h} p_{A,d}(h_w)\cdot H(A:s,d/h_w) \tag{C3}$$

$$H(A:s,d/t) = \sum_{all\,h} p_{A,d}(h_w)\cdot H(A:s,d/t_z) \tag{C4}$$

The relation among the unconditional entropy of joint distribution (18), unconditional entropy of of marginal distributions (19, 20) and the conditional entropy of heights(C3) and periods C4) holds:

$$H(A:s,d,h,t) = H(A:s,d,h) + H(A:s,d/h) =$$
$$= H(A:s,d,t) + H(A:s,d/t) \tag{C5}$$

The difference between the entropy of marginal distributions of heights (19) and periods (20) and the conditional entropy (C3) and (C4) are normally small.

REFERENCES

[1] Aczel, J. and Daroczy, Z. (1975). On Measures of Information and Their Characterization, Acad. Press, NY.

[2] Bitner-Gregersen, E.M. and Cramer, E.H. (1994). Accuracy of the Global Wave Statistics Data,

Proceedings of the Fourth International Offshore and Polar Engineering Conference, Osaka, Japan. http://publications.isope.org/proceedings/ISOPE/ISOPE%201994/Abstract%20Pages/I94v3p058.pdf.
[3] Global wave statistics BMT http://www.globalwavestatisticsonline.com/ (accessed March 2023).
[4] Choi Y. H. and Hirayama, T. (2000). Interrelationship Among the Long-Term Wave Data Bases. For the Unification of Wave Data Bases (in Japanese), Journal of the Society of Naval Architects of Japan, Vol.No. 188, pp 239-250.
[5] Fisher, N.I. (1993): Statistical Analysis of Circular Data, Cambridge University Press.
[6] Cover, A. and Thomas, J.A. (2006) Elements of Information Theory. Wiley-Interscience, Hoboken, 2 edition.
[7] Hartley, R. V. (1928). Transmission of Information, Bell Systems Tech. J, 7.
[8] Hogben, N., N. M .C. Dacunha and G. F. Olliver 1986. Global Wave Statistics, British Maritime Technology Ltd., Feltham.
[9] Khinchin, A. I. (1957). Mathematical Foundations of Information Theory, Dover Publications, NY.
[10] Klir, G. J. and Wierman, M. J. 1999. Uncertainty-Based Information, Springer-Verlag Berlin, Heidelberg GmbH.
[11] Nielsen U.D. and Ikonomakis, A. (2021). Wave conditions encountered by ships—A report from a larger shipping company based on ERA5, Ocean Engineering, 237, https://doi.org/10.1016/j.oceaneng.2021.109584.
[12] Renyi, A. (1970). Probability Theory, North-Holland, Amsterdam.
[13] Shannon, C. E. and Weaver, W. (1949). The Mathematical Theory of Communication, Urbana Univ. Illinois.
[14] Shinkai, A. and Wan, S., (1996). Statistical characteristics of the global wave statistics data and long-term predictions. United States: N. p., 1996. https://www.osti.gov/biblio/455449-statistical-characteristics-global-wave-statistics-data-long-term-predictions.
[15] Taylor, D. (1990). Uncertainty in Collision Avoidance Manoeuvring. Journal of Navigation, 43(2), 238-245. doi:10.1017/S0373463300009577
[16] Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics, J Stat Phys 52.
[17] Vettor, R., Bergamini, G. and Guedes Soares, C. (2021) A Comprehensive Approach to Account for Weather Uncertainties in Ship Route Optimization. J. Mar. Sci. Eng. 2021, 9, 1434. https://doi.org/10.3390/ jmse9121434.
[18] Wiener, N. (1948). Cybernetic, or Control and Communication, Bell System Tech. J, 27.
[19] Young, I. R. and Holland G. J. (1996). Atlas of the oceans: wind and wave climate, Pergamon.
[20] Ziha, K. (2000). Event Oriented System Analysis, Probabilistic Eng. Mech., 15 (3).
[21] Ziha, K. (2007). Entropy of marginal distributions, In: Luzar-Stiefler V, Hljuz Dobrić V, editors Proceedings of the 29th International Conference on Information Technology Interfaces; IEEE Catalog Number 07EX1589C; Cavtat, Croatia. Zagreb: SRCE University Computing Centre, University of Zagreb.
[22] Ziha, K. (2010) Effects of Loading Fields on Marine Objects, Journal of Ship Production, 26 (4). pp. 252-264.
[23] Ziha, K. (2013). Statistical variability vs. probabilistic uncertainty In: Luzar-Stiefler V, Hljuz Dobrić V, editors Proceedings of the 29th International Conference on Information Technology Interfaces; IEEE Catalog Number 07EX1589C; Cavtat, Croatia. Zagreb: SRCE University Computing Centre, University of Zagreb.
[24] Ziha, K. (2022). Uncertainty of Integral System Safety in Engineering, ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, 8(2), https://doi.org/10.1115/1.4051939.