

Weibull Statistics in Ocean Analysis and Prediction

Peter C. Chu

Naval Ocean Analysis and Prediction (NOAP) Laboratory
Naval Postgraduate School
Monterey, California 93943, USA
pcchu@nps.edu

Abstract—This Ocean processes are typically sparsely observed, and their dynamics are high dimensional and nonlinear, so that the statistics are non-Gaussian. Successful and efficient analysis and prediction require these characteristics to be directly addressed in data assimilation, data fusion, and uncertainty quantification since methods using Gaussianity assumptions perform poorly in many applications. In this study, the Weibull distribution is proved more realistic in ocean analysis and prediction such as analysis of near ocean surface current and significant wave height data as well as numerical modeling in wave prediction. Use of the Weibull distribution will improve ocean analysis and prediction.

Keywords—Weibull Statistics; Ocean Currents; Satellite Altimeter; OSCAR Data; Ocean Analysis and Prediction

I. INTRODUCTION

The Gaussian distribution is generally assumed in ocean analysis and prediction as an easy way of modeling the probability density function (PDF) of experimental and modeling data. The one-dimensional Gaussian distribution is determined by just two parameters: the mean and the standard deviation of the data. The high-order moments are given (skewness: 0; kurtosis: 3). When we refer to ocean data (either observed or modeled) by its mean and variance, we are in fact compressing the original data. We delete all additional information possibly hidden in the data set. But we gain simplicity. Since the skewness of ocean observed and modeled data is usually not zero (either positive or negative), it is naturally to choose the Weibull distribution as the first step in extending from the Gaussian into non-Gaussian statistics. This is because the Weibull distribution can represent variety of distributional shapes from negatively to positively skewed. The stochastic theory for ocean surface current dynamics leads to the Weibull distribution.

II. STOCHASTIC DIFFERENTIAL EQUATIONS FOR SURFACE CURRENTS

Let (x, y) be the horizontal coordinates and z be the vertical coordinate. Vertically averaged horizontal velocity components (u, v) from the surface to a constant scale depth (h) of surface mixed layer are given by^[1]

$$\frac{\partial u}{\partial t} = \frac{1}{h} \Lambda_u - \frac{K}{h^2} u, \quad (1)$$

$$\frac{\partial v}{\partial t} = \frac{1}{h} \Lambda_v - \frac{K}{h^2} v, \quad (2)$$

where

$$\Lambda_u \equiv fV_E + \frac{\tau_x}{\rho}, \quad \Lambda_v \equiv -fU_E + \frac{\tau_y}{\rho}, \quad (3)$$

represent the residual between the Ekman transport and surface wind stress. Here, K is the eddy viscosity; f is the Coriolis parameter; (τ_x, τ_y) are the surface wind stress components; and (U_E, V_E) are Ekman transports computed by

$$(U_E, V_E) = \int_{-h}^0 (\tilde{u} - u_g, \tilde{v} - v_g) dz, \quad (4)$$

where $(\tilde{u}, \tilde{v})$ are the vertically varying horizontal velocity components; and (u_g, v_g) are the geostrophic velocity components. With absence of horizontal pressure gradient, e.g., $u_g = v_g = 0$, Equations (1) and (2) reduce to the commonly used wind-forced slab model^[2]. For the sake of convenience, we assume that the residual between the Ekman transports (U_E, V_E) and surface wind stress does not depend on the horizontal current vector (u, v) . Away from the equator, this approximation is similar to a small Rossby number approximation^[3]. If the forcing (Λ_u, Λ_v) is fluctuating around some mean value,

$$\Lambda_u(t) = \langle \Lambda_u \rangle + \dot{W}_1(t)h\Sigma, \quad \Lambda_v(t) = \langle \Lambda_v \rangle + \dot{W}_2(t)h\Sigma, \quad (5)$$

where the angle brackets represent ensemble mean and the fluctuations are taken to be isotropic and white in time:

$$\langle \dot{W}_i(t_1) \dot{W}_j(t_2) \rangle = \delta_{ij} \delta(t_1 - t_2), \quad (6)$$

with a strength that is represented by Σ . Note that the Ekman transport is determined by the surface wind stress for time-independent case, and therefore the ensemble mean values of (Λ_u, Λ_v) are zero,

$$\langle \Lambda_u \rangle = 0, \quad \langle \Lambda_v \rangle = 0. \quad (7)$$

Substitution of (5)-(7) into (1) and (2) gives

$$\frac{\partial u}{\partial t} = -\frac{K}{h^2} u + \dot{W}_1(t)\Sigma, \quad (8)$$

$$\frac{\partial v}{\partial t} = -\frac{K}{h^2} v + \dot{W}_2(t)\Sigma, \quad (9)$$

which is a set of stochastic differential equations for the surface current vector. The joint PDF of (u, v) satisfies the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = \left(\frac{\Sigma^2}{2} \right) \left(\frac{\partial^2 p}{\partial u^2} + \frac{\partial^2 p}{\partial v^2} \right) + \frac{\partial}{\partial u} \left[\left(\frac{K}{h^2} u \right) p \right] + \frac{\partial}{\partial v} \left[\left(\frac{K}{h^2} v \right) p \right], \quad (10)$$

which is a linear second-order partial differential equation with the depth scale (h) taken as a constant. Transforming from the orthogonal coordinates (u, v) to the polar coordinates (w, φ) respectively the current speed and direction,

$$u = w \cos \varphi, \quad v = w \sin \varphi. \quad (11)$$

The joint PDF of (u, v) is transformed into the joint PDF of (w, φ),

$$p(u, v) du dv = p(u, v) w dw d\varphi = \tilde{p}(w, \varphi) dw d\varphi. \quad (12)$$

Integration of (12) over the angle φ from 0 to 2π yields the marginal PDF for the current speed alone,

$$p(w) = \int_0^{2\pi} \tilde{p}(w, \varphi) d\varphi. \quad (13)$$

For a constant eddy viscosity (K) at $z = -h$, the steady state solution of equation (10) is given by

$$p(u, v) = A \exp \left[-\frac{K}{\Sigma^2 h^2} (u^2 + v^2) \right], \quad (14)$$

where A is a normalization constant. Substitution of (14) into (12) and use of (13) yield the Rayleigh distribution

$$p(w) = \frac{2w}{a^2} \exp \left[-\left(\frac{w}{a} \right)^2 \right], \quad a \equiv \frac{\Sigma h}{\sqrt{K}}, \quad (15)$$

with the scale parameter a . The basic postulation of constant K may not be met always at the upper ocean. Hence we require a model that can meet the twin objectives of (a) accommodating Rayleigh distribution whenever the basic hypothesis (constant K) that justifies it is satisfied and (b) fitting data under more general conditions. This requirement is supposed to be satisfied by the Weibull probability density function,

$$p(w) = \frac{b}{a} \left(\frac{w}{a} \right)^{b-1} \exp \left[-\left(\frac{w}{a} \right)^b \right], \quad (16)$$

where the parameters a and b denote the scale and shape of the distribution. This distribution has been recently used in investigating the ocean model predictability^{[4] [5]}.

III. PARAMETERS OF THE WEIBULL DISTRIBUTION

The four parameters (mean, standard deviation, skewness, and kurtosis) of the Weibull distribution are calculated by^[6],

$$\text{mean}(w) = a \Gamma \left(1 + \frac{1}{b} \right), \quad (17)$$

$$\text{std}(w) = a \left[\Gamma \left(1 + \frac{2}{b} \right) - \Gamma^2 \left(1 + \frac{1}{b} \right) \right]^{1/2}, \quad (18)$$

where Γ is the gamma function. The parameters a and b can be inverted^[7] from (17) and (18),

$$b \approx \left[\frac{\text{mean}(w)}{\text{std}(w)} \right]^{1.086}, \quad a = \frac{\text{mean}(w)}{\Gamma(1 + 1/b)}. \quad (19)$$

The skewness and kurtosis are computed by

$$\text{skew}(w) = \frac{\Gamma \left(1 + \frac{3}{b} \right) - 3 \Gamma \left(1 + \frac{1}{b} \right) \Gamma \left(1 + \frac{2}{b} \right) + 2 \Gamma^3 \left(1 + \frac{1}{b} \right)}{\left[\Gamma \left(1 + \frac{2}{b} \right) - \Gamma^2 \left(1 + \frac{1}{b} \right) \right]^{3/2}} \quad (20)$$

$$\text{kurt}(w) = \frac{\Gamma \left(1 + \frac{4}{b} \right) - 4 \Gamma \left(1 + \frac{1}{b} \right) \Gamma \left(1 + \frac{3}{b} \right)}{\left[\Gamma \left(1 + \frac{2}{b} \right) - \Gamma^2 \left(1 + \frac{1}{b} \right) \right]^2} + \frac{6 \Gamma^2 \left(1 + \frac{1}{b} \right) \Gamma \left(1 + \frac{2}{b} \right) - 3 \Gamma^4 \left(1 + \frac{1}{b} \right)}{\left[\Gamma \left(1 + \frac{2}{b} \right) - \Gamma^2 \left(1 + \frac{1}{b} \right) \right]^2} - 3 \quad (21)$$

which depend on the parameter b only [see (20) and (21)] for the Weibull distribution. The relationship between the kurtosis and skewness can be determined from (20) and (21). Thus, the two parameters are easily determined from modeled and/or observed data. Fig. 1 shows the dependence of the shape of the PDF on the shape parameter (b) from positively ($b < 3.5$) to negatively ($b > 3.5$) skewed.

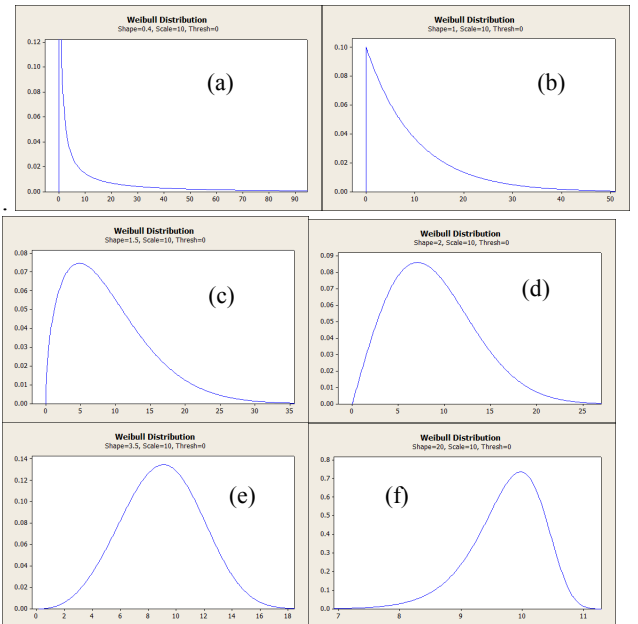


Figure 1. Dependence of the Weibull distribution (with $a = 10$) on the shape parameter: (a) $b = 0.4$, (b) $b = 1$, (c) $b = 1.5$, (d) $b = 2$, (e) $b = 3.5$, and (f) $b = 20$.

IV. THE OSCAR DATA

The near real-time surface current fields (OSCAR) are available for the world oceans from 60°N to 60°S on 1°×1° grid to a broad-based user community via a web-based interactive data selection interface on a time base with exactly 72 steps per year (about 5 day interval) starting from October 1992. The velocity is automatically computed from gridded fields of surface topography and wind derived on the base of the Ekman dynamics from satellite altimeter (JASON-1, GFO, ENVISAT) and scatterometer (QSCAT) vector wind data. See website: <http://www.oscar.noaa.gov/> for detailed information.

V. STATISTICAL CHARACTERISTICS

The data depicted in Section 4 are used to investigate the statistical features of the global surface current speeds (w). The four parameters (mean, standard deviation, skewness, and kurtosis) can also be calculated from the observational data (w)

$$\begin{aligned} \text{mean}(w) &= \frac{1}{N} \sum_{i=1}^N w_i, \quad \text{std}(w) = \sqrt{\text{mean}[w - \text{mean}(w)]^2}, \\ \text{skew}(w) &= \frac{\text{mean}\{[w - \text{mean}(w)]^3\}}{\text{std}^3(w)}, \\ \text{kurt}(w) &= \frac{\text{mean}\{[w - \text{mean}(w)]^4\}}{\text{std}^4(w)} - 3, \end{aligned} \quad (22)$$

for the each grid point. The mean, standard deviation, skewness, and kurtosis fields of w estimated from the OSCAR data are displayed in Fig. 2. Large values of $\text{mean}(w)$ occur in the western boundaries such as in the Gulf Stream, Kuroshio, and Somali Current, Malvinas Current and; secondary maxima in the equatorial zones especially in the western and central equatorial Pacific. Minima of $\text{mean}(w)$ occur in the subtropical horse latitudes. The standard deviation of w is also large near the western boundaries and in the equatorial zones. In general, w is positively skewed in the most part of the global oceans and negatively skewed in the equatorial zones and Southern Ocean. The kurtosis field is much noisier than those of $\text{mean}(w)$, $\text{std}(w)$, or $\text{skew}(w)$ ^[8].

VI. WEIBULL PARAMETERS

The Weibull parameters (a , b) were calculated from $\text{mean}(w)$ and $\text{std}(w)$. The distribution of the parameter a over the global oceans (Fig. 3a) is quite close to the distribution of $\text{mean}(w)$, i.e., with large values in western boundaries and equatorial zone. The distribution of the Weibull parameter b is shown in Fig. 3b. Thus, a four-parameter dataset has been established each location. The scatter diagrams were drawn for global oceans during all or different time periods.

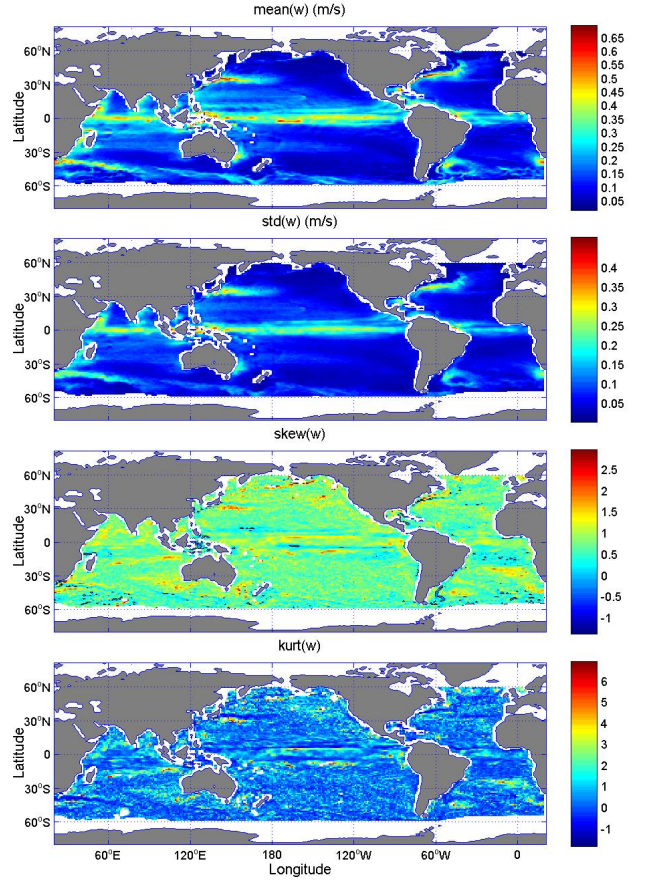


Figure 2. First four parameters of the surface current speeds calculated from the OSCAR data (1992-2007) (from Chu 2009).

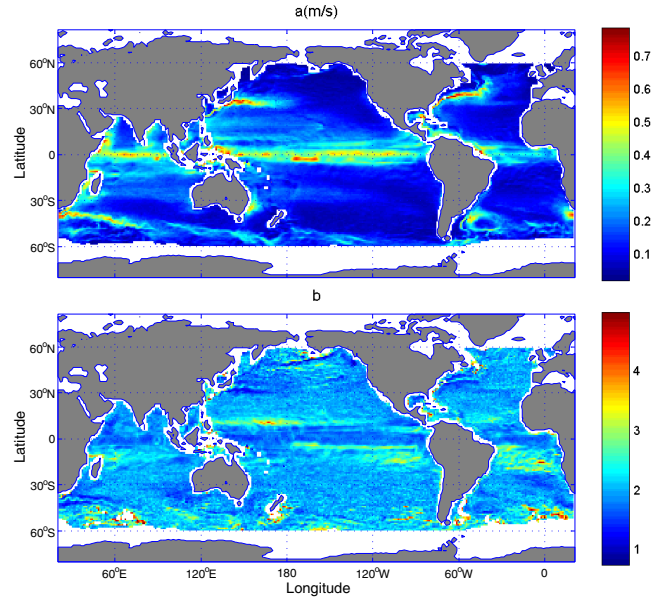


Figure 3. Same as in Fig. 2, but for Weibull parameters (a , b) (from Chu 2009).

We may use the relationships between $\text{skew}(w)$ and $\text{mean}(w)/\text{std}(w)$ (representing the parameter b) and between the $\text{kurt}(w)$ and the $\text{skew}(w)$ to identify the fitness of the Weibull distribution for observational w -PDFs. The solid curve on these figures shows the relationship for a Weibull variable. Fig. 4 shows the kernel density estimates of joint PDFs of $\text{mean}(w)/\text{std}(w)$ and $\text{skew}(w)$ for (left) January and (right) July OSCAR data from 1992 to 2008. The contour intervals are logarithmically spaced. The thick black line is the theoretical curve for a Weibull variable. (bottom) Same as in top, but for the joint PDF of $\text{skew}(w)$ and $\text{kurt}(w)$. For the observational surface w -PDF, the $\text{skew}(w)$ is evidently a concave function of the ratio $\text{mean}(w)/\text{std}(w)$ (the same as the Weibull distribution), such that the theoretical function is positive for small values of this ratio and negative for large values. However, for the core of the kernel with the joint probability higher than 0.32, $\text{mean}(w)/\text{std}(w)$ is always less than 2.2 and $\text{skew}(w)$ is always positive (Fig. 4, upper panes). Similarly, the relationship between $\text{skew}(w)$ and $\text{kurt}(w)$ in the observations is similar to that for a Weibull variable (lower panels, Fig. 4) with smaller kurtosis. The agreement between the moment relationships in the OSCAR data and those for a Weibull variable reinforces the conclusion that these data are Weibull to a good approximation.

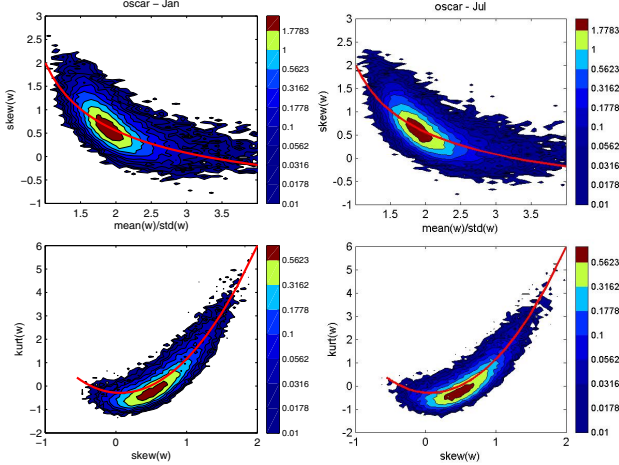


Figure 4. (top) Kernel density estimates of joint PDFs of $\text{mean}(w)/\text{std}(w)$ and $\text{skew}(w)$ for (left) January and (right) July OSCAR data from 1992 to 2008. The contour intervals are logarithmically spaced. The thick black line is the theoretical curve for a Weibull variable. (bottom) Same as in top, but for the joint PDF of $\text{skew}(w)$ and $\text{kurt}(w)$ (from Chu 2009).

VII. CONCLUSIONS

This study has investigated the probability distribution function of the surface current speeds (w), using long-term (1992-2008) 5 day “Ocean Surface Current Analyses – Real Time (OSCAR)” (OSCAR) data; and theoretically, using a

stochastic model derived using upper boundary layer physics. The following results were obtained.

(1) Probability distribution function of the surface current speeds (w) approximately satisfies the two-parameter Weibull distribution. In the upper ocean with a constant eddy viscosity K , the probability distribution function satisfies a linear second-order partial differential equation (i.e., the Fokker-Planck equation) with an analytical solution – the Rayleigh distribution (special case of the 2 parameter Weibull distribution).

(2) Four moments of w (mean, standard deviation, skewness, kurtosis) have been characterized. It was found that the relationships between $\text{mean}(w)/\text{std}(w)$ and $\text{skew}(w)$ and between $\text{skew}(w)$ and $\text{kurt}(w)$ from the data are in fairly well agreement with the theoretical Weibull distribution for the upper (0 - 50 m) tropical Pacific for the whole period. The OSCAR data also show that the ratio $\text{mean}(w)/\text{std}(w)$ is generally less than 2.2 and the skewness is generally positive for the whole global oceans.

(3) The Weibull distribution provides a good approximation to the PDF of w , which presents the possibility of improving ocean analysis and prediction.

ACKNOWLEDGMENT (Heading 5)

This research was supported by the Naval Oceanographic Office, and Naval Postgraduate School. The author thanks Mr. Chenwu Fan for computational assistance.

REFERENCES

- [1] P.C. Chu, “Probability distribution function of the upper equatorial Pacific current speeds,” *Geophysical Research Letters*, 35, doi:10.1029/2008GL033669, 2008.
- [2] R.T. Pollard, and R. C. Millard, “Comparison between observed and simulated wind-generated inertial oscillations,” *Deep Sea Research*, 17, 795-812, 1970.
- [3] A.E. Gill, “*Atmosphere-Ocean Dynamics*,” Academic Press, San Diego, 662 pp, 1982.
- [4] L.M. Ivanov,, and P.C. Chu, “On stochastic stability of regional ocean models to finite-amplitude perturbations of initial conditions,” *Dynamics of Atmosphere and Oceans*, 43, 199-225, doi:10.1016/j.dynatmoce.2007.03.001, 2007a.
- [5] L.M. Ivanov, and P.C. Chu, “On stochastic stability of regional ocean models with uncertainty in wind forcing,” *Nonlinear Processes of Geophysics*, 14, 655-670, 2007b.
- [6] N. Johnson, S. Kotz, and N. Balakrishnan, “*Continuous Univariate Distributions*,” Vol. 1. Wiley, 756 pp, 1994.
- [7] A.H. Monahan, “The probability distribution of sea surface wind speeds. Part-I: theory and SeaWinds observations,” *Journal of Climate*, 19, 497-519, 2006.
- [8] Chu, P. C., “Statistical Characteristics of the Global Surface Current Speeds Obtained from Satellite Altimeter and Scatterometer Data,” *IEEE Journal of Selected Topics in Earth Observations and Remote Sensing*, 2 (1), 27-32, 2009.