Statistical tests for the distribution of surface wind and current speeds across the globe

Salvatore Campisi-Pinto[a], Kaushal Gianchandani[b], Yosef Ashkenazy[a,*]

[a]*Department of Solar Energy and Environmental Physics, BIDR, Ben-Gurion University, Midreshet Ben-Gurion, Israel*
[b]*The Fredy and Nadine Herrmann Institute of Earth Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel*

Abstract

The distribution of surface winds and currents is important from climatic and energy production aspects. It is commonly assumed that the distribution of surface winds and currents speed is Weibull, yet, previous studies indicated that is assumption is not always valid. An inaccurate probability distribution function (PDF) of wind (current) statistic can lead to erroneous power estimation; thus, it is necessary to examine the accuracy of the PDFs employed. We propose statistical tests to check the validity of an assumed distribution of wind and current speeds. The main statistical test can be applied to any distribution and is based on surrogate data where the different moments of the data are compared with the moments of the surrogate data. We applied this and other tests to global surface wind and current speeds and found that the generalized gamma distribution fits the data distributions better than the Weibull distribution. The percentage of locations that fall within the confidence level of the assumed distribution varies with the moment. The third moment is used to estimate the potential power of winds and currents—we find that 89% (95%) of the wind (current) grid points fall within the 95% confidence interval of the generalized gamma distribution.

Keywords: surface winds, surface currents, speed statistics, Weibull distribution, generalized gamma distribution

*corresponding author
Email address: ashkena@bgu.ac.il (Yosef Ashkenazy)

1. Introduction

Surface winds and surface ocean currents play a crucial role in regulating the weather and climate systems. Driven by the energy of the Sun, winds are responsible for movement of air across the globe. Winds and currents span a wide range of temporal and spatial scales. By forcing the ocean surface, winds generate surface currents; currents transport ocean water (and hence heat and salt) and, in this way, affect regional and global climatic conditions and circulation. Winds are a major source of ocean kinetic energy—about half of the deep ocean energy ($\sim$1 TW) is attributed to winds, and the other half, approximately, is attributed to tides [1, 2].

The increasing interest in alternative forms of energy ("green" energy), as a step toward low carbon emissions, has led to a significant increase in the use of wind turbines, to convert the kinetic energy (power) of winds to electric energy (power). However, surface ocean currents have received much less attention as a potential source of energy [3, 4, 5, 6, 7]. Harnessing the kinetic energy of surface ocean currents may be a viable complement to wind energy because surface currents are less erratic and persist for a longer duration of time [8, 9].

Accurate information regarding the distributions of winds and currents can be utilized as a reference for improved ocean and climatic modeling. Accurate estimation of the probability density functions (PDFs) of surface wind and current speeds can be used to reliably estimate their potential power production. Moreover, precise PDFs are required to provide the recurrence time of extreme wind and current events, which are essential from an engineering perspective. Significant progress has been made in finding the PDFs of surface wind and current speeds [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24]; however, most of the studies on surface winds and ocean currents accept a simplifying hypothesis that the PDF under consideration follows the Weibull distribution [10, 11, 12, 13, 14, 15, 25, 16, 20, 26, 23] and the Weibull distribution can be used to characterize wind and current speed statics accurately. A few studies questioned the use of the Weibull distribution as the optimal PDF of surface winds and currents [27, 28, 29, 30, 31] and other distributions have been proposed to characterize the speed data.

For example, the following studies reported different distributions that should be used to fit wind speed data: (i) [32] used a mixture of two Weibull distributions (with two parameters for each distribution and one proportionality parameter) to study the wind statistics over the Eastern Mediterranean.

(ii) [33] studied the wind statistics of 178 off-shore stations (mainly over North America) using the Weibull, Kappa, Wakeby and other distributions, and suggested using different PDFs to describe different aspects of the wind statistics. (iii) [34] studied the wind speed distribution in the area of Palermo using the Weibull, Rayleigh, Lognormal, Gamma, Inverse Gaussian, Pearson type V, and Burr distributions. (iv) [22] studied the ERA-40 wind speed reanalysis data over Europe and found that the generalized gamma (GG) distribution better fits the data. (v) [35] studied wind speed statistics in the inner Mongolia region using the two-parameter Weibull, Logistic, and Lognormal distributions. (vi) [29] used a two-component mixture of Weibull distribution to fit bimodal distributed wind speed. (vii) [36] studied the performance of four different distributions (two- and three-parameter Weibull, Gamma, and Log-normal) to fit wind speed data from Dolný Hričov airport in Slovakia and found that the three-parameter Weibull distribution have the best fit to the data. (viii) [37] used 13 different distributions to study the statistics of hourly wind speed data from 9 stations in the United Arab Emirates and found that the (4-parameter) Kappa and the (3-parameter) Generalized Gamma distributions provide the best fit to the data; mixture of two Weibull distributions (with overall 5 parameters) yielded an even better fit.

The above studies concentrated on specific regions and focused on the statistics of wind speed data. A global analysis of winds above ocean areas was performed, e.g., in [38, 17], which suggested that the Weibull distribution is a good approximation for the PDF of the wind speed. [39, 17] also suggested a stochastic boundary layer model to explain the observed PDF of wind speed. The same author also compared the Weibull statistics (parameters and various moments) using various global and local data sources [18], such as wind estimations that are based on daily SeaWinds scatterometer and the NCEP-NCAR and ECMWF reanalysis.

In contrast to wind speed, the statistics of surface ocean currents have received much less attention. The parameters of the Weibull distribution over the global ocean were estimated based on geostrophic altimetry-based velocities [20, 40]. In addition, [19] discussed the Weibull parameters of the upper equatorial Pacific current speed estimated using six stations' hourly ADCP data. [41] analyzed ocean current statistics from the Gulf Stream (North Carolina shore) and found that the Weibull distribution properly fits the current speed PDF. The parameters of the Weibull distribution of high resolution surface current speeds were also estimated from radar (CODAR) data of the Gulf of Eilat, Israel [42] and of the Nan-Wan Bay, Taiwan [43]. Other studies [44, 45] investigated surface current velocity components that were based on altimetry data and found that the distribution varies from Gaussian when focusing on small ocean areas to exponential when dealing with extensive ocean areas—they proposed a model to explain their findings. The exponential distribution of the velocity components were also reported in [46], based on oceanic floats and numerical models [46, 47]. We note, however, that the relation between the distribution of the velocity components and the distribution of the current speed, which is the focus of this work, is not trivial, except when considering the idealized identical Gaussian distribution of the velocity components, which will result in the Rayleigh distribution (Weibull distribution with the shape parameter, $k = 2$).

The brief summary above indicates that the statistical analysis of surface winds has received much more attention than that of the surface ocean current speed, and here, we aim to extend the analysis of the latter. In addition, many distributions have been suggested to describe the observed PDF of the wind speed. This situation calls for a standard test. Following the above, the aim or this study is to present a procedure to quantify the level of agreement between an assumed PDF and the actual PDF of both wind and current speed data. The proposed procedure is not specific to either the Weibull or the GG PDF and depends on the moments of interest. We implemented this method on surface winds and currents around the globe using the Weibull and the GG PDFs. We found that the GG distribution more accurately fits the actual distribution of wind and current speed. In addition to the moment-dependent test, we studied other statistical tests.

The paper is organized as follows. Sec. 2 briefly elaborates the data analyzed for this study and in Sec. 3, we present the methodology of the present study. The results are then shown in Sec. 4. Sec. 5 discusses the estimation of the global distribution of the potential power of winds and currents when using the Weibull distribution in comparison to the GG distribution. The study is concluded and discussed in Sec. 6.

2. Data

We analyzed the ERA-Interim (a global atmospheric reanalysis) 6-hourly surface (10 m height) wind speed of the European Centre for Medium-Range Weather Forecasts (ECMWF) [48] from 1979 to 2016. The dataset spans the entire globe through a geographical grid of size 480×240 (spatial resolution of $3/4° \times 3/4°$).

The surface currents were acquired using satellite altimetry and made available by the Copernicus—Marine Environment Monitoring Service (CMEMS), http://marine.copernicus.eu and based on Topex/Poseidon between 1993-01-01 and 2002-04-23, Jason-1 between 2002-04-24 and 2008-10-18, and OSTM/Jason-2 since 2008-10-19; see [49, 50]. The spatial resolution of the altimetry data is much finer than that of the winds (grid size: 1440×720, spatial resolution of $1/4° \times 1/4°$); still, the temporal resolution is one day. The data spans 24 years, from 1993 to 2016. Both the datasets are freely available online and were download from the respective websites of ECMWF and CMEMS.

3. Methodology

The Weibull PDF is a two-parameter distribution,

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \tag{1}$$

where $x \geq 0$, and λ and k are the scale and shape parameters, respectively. The Weibull distribution reduces to the Rayleigh distribution when $k = 2$ and to the exponential distribution for $k = 1$. The GG PDF is a generalization of the Weibull PDF and has three parameters, λ, k, and ε

$$f(x; \lambda, k, \varepsilon) = \frac{1}{\Gamma(\varepsilon)} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{\varepsilon k-1} e^{-(x/\lambda)^k}, \tag{2}$$

where also here $x > 0$ and $\Gamma(\varepsilon)$ is the gamma function. The GG distribution reduces to the Weibull distribution for $\varepsilon = 1$ and to the gamma distribution for $k = 1$.

Figure 1(a),(b) depicts the Weibull PDFs for $\lambda = 1$ (scale parameter) and for different values of the shape parameter, k. The PDF decays faster for a larger k and, in this way, controls the "shape" of the PDF; the parameter λ only shifts the distribution along the x axis without altering the shape of the distribution. In Figure 1(c),(d), we present the GG PDF for $\lambda = 1$ and for $k = 1, 2$ and $\varepsilon = 1, 2, 3$. Figure 1(c) shows that in some cases ($k = 1$), the ε parameter also controls the shape of the distribution. Since the GG PDF has three parameters, it can potentially improve the fit to the PDF to the data. In Figures 1(e),(f), we present examples of the PDFs of two geographical locations surface wind speeds. In these examples, both the Weibull and the GG PDFs were fitted to the data using the maximum likelihood criteria. As expected, the GG distribution fits the data better than the Weibull distribution. Furthermore, the value of ε estimated for the GG fit was different than 1. If the time-series had been truly Weibull-distributed, the value of ε would have been about 1. In other cases (such as the case of Figure 1(f)), neither the Weibull nor the GG PDF properly fit the PDF of the actual wind speed data. The method we propose below aims to identify the locations at which either the Weibull or the GG distribution is suitable to fit the distribution of the data.

We used a methodological protocol based on the method of the moments used in conjunction with the method of the maximum likelihood estimation (MLE) [51, 52] to test the validity of the PDF (here Weibull and GG) hypothesis for a sample of measurements. The methods, as presented below, were applied to every single time series of the dataset at hand; i.e., the time series of every grid point were analyzed separately.

The method of the moments [52], first introduced by Chebyshev in the 19th century, is a method of estimating population parameters. Assuming a particular distribution, such as Weibull or GG, for a given sample of measurements, the method estimates the sample distribution parameters by solving a system of equations that relates the sample parameters to be estimated with the population moments. This method is used in Appendix B to find the Weibull and GG PDF parameters. In contrast, the MLE estimates the parameter values that maximize the likelihood function, given the observations—this method finds the best fit (and hence the optimal PDF parameters) to a given observed distribution. The MLE method is used throughout this paper.

The test we propose below is valid for any distribution; as an example, we consider the standard distribution for surface wind and current speed, the Weibull distribution. The analysis unfolds into the following steps:

(i) we start by assuming that the series at hand (x) is indeed Weibull-distributed (WBL);

(ii) we estimate the distribution parameters λ and k of x based on the MLE method;

(iii) by using these estimated parameters, λ and k, we generate a large number ($N = 300$) of surrogate Weibull-distributed series S_i for $i =$

$1, \ldots, N$ where the length of each surrogate series is equal to the length of the original series x;

(iv) we estimate the first $m_{\max}$ moments (i.e., $m = 1, \ldots, m_{\max}$) of each surrogate series S_i where the m^{th} moment is $\mu_m^{S_i} = \langle S_i^m \rangle$ (where $\langle \cdot \rangle$ represents the expected value);

(v) we calculate the first $m_{\max}$ moments of the original series x, μ_m^x;

(vi) in parallel, we estimate the 95% confidence intervals (CIs) of each moment CI_m using the 0.025 and the 0.975 quantiles of the distribution of $\mu_m^{S_i}$;

(vii) we benchmark μ_m^x against the corresponding CI_m of the surrogate data for all the moments. In other words, for each moment, we test whether μ_m^x falls within the boundary values (quantiles) defined by CI_m.

If the value of μ_m^x falls within the CI of the m^{th} surrogate moment, CI_m, the result of the benchmarking is positive, and the null hypothesis is not rejected; otherwise, the null hypothesis is rejected, and the conclusion is that the PDF of the data is not the assumed one. A positive result indicates that the hypothesized distribution, for example the Weibull, is a good approximation of the PDF of the data, for the specific moment at hand. It is worth emphasizing that the method is "moment-dependent" such that the same sample can score a positive result for a given moment and a negative result for a different one. We analyzed several moments for theoretical purposes, while for most practical applications (for instance, wind speed electric power generation), only moments up to three or four are of interest; the Skewness and Kurtosis are related to the first three and four moments respectively and were analyzed in previous studies [like, 19, 17]. Below, we show the implementation of the proposed test when assuming the Weibull and the GG distributions.

In addition to the general test proposed above, we propose two other tests that are specific to the Weibull and the GG distributions, and these are discussed in detail in Appendix B and Appendix C; we implement these tests on surface wind and current speed data. Essentially, in the first method, we estimate the parameters of either the Weibull or the GG distribution using the MLE, then generate surrogate series based on these parameters, then use the ratio between the different moments to estimate the parameters of the assumed distribution of both the original data and the surrogate data, and then check whether the moment-based parameters fall within the CI of the surrogate data moment-based parameters—see Appendix B. In the second method, we use the fact that the GG distribution reduces to the Weibull distribution when $\varepsilon = 1$. We estimate the Weibull parameters using the MLE, then use these parameters to generate surrogate series, and then estimate the GG parameters of these surrogate series. The ε parameter of the GG distribution should be scattered around 1; by comparing the ε parameter of the data to the CI of the ε of the surrogate data, one can conclude whether the data is Weibull-distributed or not (see Appendix C). We also applied the standard χ^2-test and the Kolmogorov-Smirnov test—see Sec. 6.

4. Results

We first show and discuss the estimated Weibull parameters for the surface wind speed and surface current speed. Figure 3 shows the MLE estimated scale and shape parameters, λ and k, over the entire globe. There is a clear difference in the λ of the wind speed over land and over the ocean where λ is much smaller over land due to the weaker winds there. This is since λ is closely related to the mean speed as the mean wind speed is $\langle s\rangle = \lambda\Gamma(1 + 1/k)$, and since $\Gamma(1 + 1/k) \sim 0.9$ for the relevant range of $k = 1 - 5$, λ is proportional to the mean speed; i.e., $\langle s\rangle \approx 0.9\lambda$. Thus, the scale parameter λ is large in regions of enhanced winds, such as storm tracks and over the Antarctic Ocean. Generally speaking, the shape parameter k of the wind speed Weibull distribution is smaller over land although there are some exceptions like Antarctica. We note that the winds over the tropical ocean are characterized by a large k.

Similarly, with reference to the ocean surface currents, the scale parameter λ also reflects the mean current distribution where, for example, the Gulf Stream, the Kuroshio Current, the Equatorial Current, and the Agulhas Current are clearly visible. Unlike the scale parameter λ, the shape parameter k is almost uniformly distributed over the ocean; no trivial geographical pattern can be extrapolated from the distribution of k. The distributions of the scale and shape parameters, λ and k, for the surface wind and current speed are presented in Fig. A.11 where it is clear that the range of k for the currents is smaller in comparison to the k parameter of the winds. This smaller k for the surface currents may be partially attributed to the fact that the surface of the ocean is forced by the wind stress whose value is, at least, the square of the surface wind speed. The zonal mean of the λ and k parameters of the winds and currents are presented in Fig. A.12 where the λ of the winds peak at the mid-latitudes of the southern ocean and the λ of the currents peak at the equator. The shape parameter k of the surface currents is almost uniformly distributed over almost all latitudes, in contrast to the large k for the surface winds for latitudes $\sim$40°S and at the tropical regions. The results described above are similar to the results discussed in [17] and in [40].

The GG distribution is a generalization of the Weibull distribution, and below, we show and discuss the MLE-fitted GG distribution parameters, λ, k and ε. Figure 4 depicts the estimated parameters of the surface wind speed data. In general, the λ and k parameters of the Weibull distribution (Fig. 3a,b) are comparable to the corresponding GG λ and k parameters presented in Fig. 4a,b; however, the GG parameters are typically larger and span a larger range than the Weibull-estimated parameters. This can be more easily seen in Fig. A.11a,b where the distribution of both λ and k is broader for the GG parameters. The zonal mean of the estimated parameters shown in Fig. A.12a,b indicates that while the pattern of the Weibull parameters is similar to the pattern of the GG parameters, the GG parameters span a larger range. For example, the value of λ is larger around 50°S-60°S, where the GG one is larger than the Weibull one. A similar situation is observed for the k parameter (shown in Fig. A.12b) where the GG k is much larger than the Weibull one for the tropics and around 50°S-60°S and is smaller than the Weibull one for the high latitudes. The GG ε parameter of the surface winds is shown in Fig. 4c, and it seems to be larger over land, in contrast to the k parameter. The relation between the k and ε parameters of the GG distribution is plotted in Fig. 4d, and it is clear that the two are not totally independent. The dependence between the two can be approximated by a power law relation, i.e., $\varepsilon \propto k^{-4/3}$, indicating a large ε for a small k and vice versa. We have no explanation for this apparent relation. Despite the above, one should remember that the approximate power law relation is not strict and that there is variability around this relation, making the GG distribution a better approximation for the PDF of the observed surface winds and surface currents; see below. We note that we could not identify a similar relation for other parameter combinations.

We repeated the estimation of the GG distribution parameters for the surface ocean current speed (Fig. 5). As for the surface wind speed field, also here the λ and k parameters of the Weibull are similar to the corresponding GG parameters, although the latter span a wider range of parameters, especially for the k parameters (Fig. A.11d,e). In comparison to the GG parameters of the surface wind speed, those of the surface currents are restricted to a narrower range of parameters, as we observed for the Weibull parameters of the winds and currents. The zonal mean of the surface current speed GG parameters is very similar to the Weibull ones. Large ε and small k are observed at the high latitudes, but these values could be due to the partial data coverage, both in space and time, at these latitudes. The relation of the ε parameter versus the k parameter is presented in Fig. 5d where the power law relation between the two ($\varepsilon \propto k^{-4/3}$) seems to hold here as well; however, the variability around this relation is not small, enabling a better fit of the GG distribution to the observed distribution of the surface current speed.

In Sec. 3 and in Fig. 2, we described a general method to verify whether a hypothesized PDF properly fits the PDF of data under investigation (in our case, wind and current speed). This method depends on the moment and on the prescribed CI. In Fig. 6, we present a map showing whether the third moment of the data falls within or outside the 95% CI of the third moment of the surrogate data. We use the third moment as it is often used to calculate the potential wind power. Fig. 6a,b depicts the results for the surface wind speed when assuming that the underlying PDF is Weibull (Fig. 6a) and GG (Fig. 6b). It is apparent that the null hypothesis of the Weibull distribution is not rejected over the ocean, while over extensive land areas (e.g., North and South America and Asia), the null hypothesis is rejected such that one cannot conclude that the underlying distribution is indeed Weibull. The Weibull null hypothesis is not rejected for 78% of the global area. When assuming that the GG PDF is the underlying distribution, the situation improves, and the null hypothesis is rejected only for 11% of the global area (Fig. 6b). Thus, as expected, the GG PDF better fits the distribution of the surface wind speed, especially over land. As for the surface current speed (Fig. 6c,d), here the situation is better, for both the Weibull and the GG distributions, where 80% (Weibull) and 95% (GG) of the analyzed area falls within the CI of the assumed distribution. Based the above, one can conclude that when focusing on the third moment (using the 95% CI), both the Weibull and the GG distributions are adequate distributions for both the surface wind and the current speed; the GG distribution performs better than the Weibull distribution by more than 10%, and thus is a better choice for the distribution of the data.

The ratio (or percentage) of the analyzed global area that falls within the 95% CI of the assumed distribution (in our case, either Weibull or GG) depends on the moment; here, we use the standard 95% CI, but obviously, the ratio will increase for larger CI and decrease for smaller CI. Fig. 7 shows this ratio as a function of the moment, for the Weibull and GG distributions of surface winds and surface currents. In general, there is a decreasing tendency of the ratio as the moment increases. In addition, there are more grid points that fall within the GG distribution CI (except $m = 1$ for GG winds) than within the Weibull ones, and the ratio for the surface current speed is larger than the surface wind speed. The above situation may vary for moments larger than $m = 7$. The ratio of the area that is within the CI drops to low values for large moments.

In this section, we considered the surrogate data test described in Sec. 3, which is applicable to general distribution and which tests each moment separately. In Appendix B and Appendix C, we present results that are specific to the Weibull and the GG distributions, where we use a set of moments to test the null hypothesis of underlying Weibull or GG distributions. These results indicate that a much smaller analyzed global area can be associated with the Weibull or the GG distribution. In addition, the χ^2-test and the Kolmogorov-Smirnov test yielded a limited area that falls within the CI; see Sec. 6 and Figs. 9, 10.

5. Winds and Oceans — Power Reservoirs

Apart from being pivotal to the dynamics of the ocean and the atmosphere, winds and currents are of economic importance. In particular, there is an increasing trend toward the use of green energy [53], to decrease greenhouse gas emissions (particularly carbon dioxide) into the atmosphere. Worldwide, wind turbines generate several hundred gigawatts of electrical power with China's contribution being the highest, about 30%; see https://www.worldenergy.org/data/resources/.

Winds, however, are not a stable source of electrical power due to their high spatial and temporal variability [54]. Energy can be harvested from the ocean through, for example, ocean waves, ocean currents [3, 4, 5, 6], ocean temperature [55], and tides. Marine energy devices, such as ocean current turbines, tidal turbines, ocean thermal energy converters, wave energy converters, and in-stream turbines, hold a huge potential for the generation of green energy.

Accurate knowledge of the distribution of both winds and currents is vital for cost-effective harnessing of the power available through these sources. The power per unit area generated from flowing fluid is [26, 56, 57, 58]:

$$P = \frac{1}{2}\rho\langle U^3\rangle \tag{3}$$

where ρ is the density of the fluid, and $\langle U^3\rangle$ is the third moment of the speed of the fluid under consideration.

To compare the performances of the Weibull and GG distributions in estimating the power, the percentage error in the power per unit area was calculated. More precisely, we computed the difference between the estimated power and the observed power (using either the Weibull or the GG estimated distributions) relative to the observed power, $\epsilon = \frac{P_{\text{Weibull or GG}} - P_{\text{observed}}}{P_{\text{observed}}}$. As is evident from Fig. 8, the GG distribution resulted in a more accurate estimate of the power per unit area to the actual value for both winds and currents. Fig. 8c,f, clearly shows that both the Weibull and GG distributions usually underestimate the power per unit area that can be generated by winds and currents. In addition, the distribution of the GG relative error is centered around the zero value, while the Weibull one is much wider, indicating smaller error when using the GG distribution. A comparison between Fig. 6 and Fig. 8a,b,d,e indicates, as expected, that the relative error is (relatively) large (indicated by the green-yellow colors in Fig. 8a,b,d,e) mostly over the regions that fall outside the CI (shown by the green color in Fig. 6), supporting the moment-based test we proposed above. We note, however, that in any case, the relative error is not large and typically is much smaller than 4%.

The use of the Weibull distribution as an approximation for the observed wind and current speed distributions may result in an inaccurate estimation of the power available for extraction for a particular location. The GG distribution instead provides a better estimate regarding the potential wind/current power. Other distributions that were not examined here may provide an even better estimation of the potential power.

6. Summary and conclusions

It is commonly assumed that surface winds and surface sea currents can be accurately modeled by Weibull probability density function over any given geographic location. In this study, we propose a method to test the validity of this assumption; in addition, an alternative distribution (namely the

Generalized Gamma) was tested. Specifically, we analyzed global 10 m surface wind speed ERA-Interim reanalysis data (6 hour interval from 1979 to 2016) and surface, altimetry based daily currents speed dataset (from 1993 to 2016).

At each grid point the tests were implemented as follows: (i) the parameters of the assumed distributions (Weibull and GG) were fitted to the available time series by the MLE method; (ii) the estimated parameters were used to generate a large number of surrogate (synthetic) data; (iii) the moments of the surrogate data were benchmarked against the moment of the original data; if the estimated moment of the original data falls within the confidence interval of the corresponding moment of the surrogate data then, for that moment, the distribution was regarded as truly Weibull (or GG depending on the initial hypothesis) such that the series passed the test (for that moment).

Overall, results showed that the GG distribution was likely to provide a better fit than the Weibull distribution for both winds and currents on a larger portion of geographical locations. In particular, with reference to the third moment of the data (which is used to calculate the potential power of winds and currents) results indicate that the portion of wind speed series passing the tests were respectively 78% when using a Weibull initial hypothesis and 89% when using GG hypothesis; on the other hand, the portion of sea current grid points passing the test were respectively 80% when using the Weibull hypothesis and 95% when using the GG hypothesis. It is worth reminding that the Weibull is a particular case of the GG distribution, when ε is about 1, therefore under appropriate conditions, both Weibull and GG distribution can fit accurately the same data.

In addition to the statistical test discussed above, which is valid for any given PDF, we applied another test that is specific to the Weibull and the GG distributions; see Appendix B and Appendix C. This approach (as described in Appendix B) resulted in a smaller percentage of geographical locations that fell within the CI of the surrogate data. Using this approach, we showed in Appendix C that only a small fraction of the available series ($\sim$10 %) was truly Weibull. Thus, for a large number of geographical locations, we cannot conclude that the Weibull or the GG were the best assumptions for wind and current speed; conversely, distributions other than Weibull and GG may provide a better fit to the particular data at hand.

We also performed standard statistical tests including the χ^2-test [59, 60] and the Kolmogorov-Smirnov test [61] as applied on a restricted dataset referring to Denmark—see, e.g., [13]. These approaches are completely different from the above mentioned tests. In particular, with reference to the χ^2-test, one basically sums the differences between the observed and the expected frequencies over the observed ranges of measured speeds. Therefore, even a small difference on the density estimated at the tail of the distribution can result in a large overall difference between the empirical and the theoretical distributions. According to the χ^2-test, results indicate that only a very small fraction of the global area falls within the CI interval of the theoretical PDF, indicating that only in a small portion of surface winds and currents are accurately approximated by Weibull (or GG) distributions (where the GG performs better than the Weibull). Results suggest that the tails of the observed distributions had a large impact on the test statistics; in practice, both the Weibull and the GG distributions were less accurate hypothesis for the highest regimens of wind and current speed.

In the Kolmogorov-Smirnov test, one basically computes the maximal difference between the observed and expected cumulative distributions where a larger difference indicates larger dissimilarity between the two distributions. Large differences are expected close to the center of the PDFs (where the PDFs are maximal), such that the Kolmogorov-Smirnov test is more sensitive to the central part of the distributions. The results of this test are presented in Fig. 10. In this case the percentage of the area falling within the 95% CI interval are much higher than what we obtained for the χ^2-test. In particular, the GG assumption yielded an area that is twice as large as the area obtained when using the Weibull assumption. In addition, when comparing the test statistics of surface wind speed against surface current speed, the current speed were likely to behave much better, in such a way that the Weibull and the GG assumptions were more accurate for currents than winds.

We translated these tests and assumptions into some of their practical consequences by calculating the potential power generated by surface winds and currents when assuming that the underlying distributions are either Weibull or GG. We estimated the error associated with calculations based on the third moment of the assumed distribution (either Weibull of GG) versus the expected power calculated from the original data. Results indicate that the magnitude of the errors associated with GG distributions are smaller than the errors associated with Weibull assumptions. Moreover, it is worth mentioning that in the context of this study, we focused on the analysis of low moments tested within the standard 95% CI interval. When considering higher moments and different CI intervals, results can change drastically.

Tuning the sensitivity of the statistical tests should be tailored to the specific application at hand.

In summary, we presented a general procedure to quantify the level of agreement between an assumed PDF and the actual PDF of the wind and current speed data. This procedure is based on comparison between the moments of the original and those of random time series which has the distribution of the assumed distribution. Other statistical tests were also presented and discussed. We found that the GG distribution more accurately fits the actual distribution of wind and current speed around the globe. We obtain better power estimation when using the GG distribution.

In this paper we used wind and current reanalysis time series as an approximation of in-situ measures. A potential limitation of this approach is inherent to the very nature of the data that we used for the statistical tests. However it is worth noticing that in-situ measurements are not evenly distributed around the globe, often, highly accurate observations are concentrated in some countries, while in other regions, in-situ measurements are very sparse, inaccurate or missing all together. In addition, observed measures over different location may not have the same temporal resolution, may not overlap over the the same time period, limiting or completely impairing the feasibility of a global analysis. Taken all this into consideration, reanalysis appeared to be an optimal choice for a global analysis. Yet, reanalysis data not always accurately estimate real winds. In addition, the surface current speeds we analyzed are based on remotely sensed daily altimetry data which are based on the assumption of geostrophy, which is not always accurate. Thus, we plan to analyze in-situ measurements of both surface winds and surface currents from different location around the globe and to compare these to the results reported here. Moreover, here we focused on *surface* winds and currents and in the future we plan to analyze the statistical properties of winds and currents of other vertical levels, both in the ocean and in the atmosphere; see, e.g., [23]. This can be performed on reanalysis data as well as on measured data. The vertical component of the wind and current vectors is related to the horizontal components via the continuity equation and we are planning to study the relation between these two. It will be also interesting if and how the parameters of the distributions vary with time; this can be accomplished but studying the CMIP5 models in recent history and under future different climate change scenarios.

In conclusion, one can ask: are surface wind and current speeds Weibull or GG distributed, if at all? The answer to this question is complex as it depends on the method of analysis and on the moment (or set of moments) of interest (where different application may focus on different moments). When focusing on low moments (smaller or equal to 3), we concluded that the GG distribution was likely to be a more accurate approximation of the distribution of the original wind and current speed.

Acknowledgments

We thank Golan Bel and Amos Zemel for helpful discussions.

References

[1] W. Munk, C. Wunsch, Abyssal recipes ii: energetics of tidal and wind mixing, Deep Sea Res. 45 (1998) 1977–2010.

[2] C. Wunsch, R. Ferrari, Vertical mixing, energy and thegeneral circulation of the oceans, Ann. Rev. Fluid Mech. 36 (2004) 281–314.

[3] P. Lissaman, The coriolis program., Oceanus 22 (4) (1979) 23–28.

[4] H. P. Hanson, S. H. Skemp, G. M. Alsenas, C. E. Coley, Power from the Florida Current: A new perspective on an old vision, Bulletin of the American Meteorological Society 91 (7) (2010) 861–868.

[5] A. E. Duerr, M. R. Dhanak, An assessment of the hydrokinetic energy resource of the Florida Current, IEEE Journal of Oceanic Engineering 37 (2) (2012) 281–293.

[6] X. Yang, K. A. Haas, H. M. Fritz, Evaluating the potential for energy extraction from turbines in the gulf stream system, Renewable Energy 72 (2014) 12–21.

[7] F. Chen, Kuroshio power plant development plan, Renewable and Sustainable Energy Reviews 14 (9) (2010) 2655–2668.

[8] H. L. Bryden, L. M. Beal, L. M. Duncan, Structure and transport of the Agulhas Current and its temporal variability, Journal of Oceanography 61 (3) (2005) 479–492.

[9] J. R. Lutjeharms, The agulhas current, Vol. 2, Springer, 2006.

[10] C. Justus, W. Hargraves, A. Mikhail, D. Graber, Methods for estimating wind speed frequency distributions, Journal of applied meteorology 17 (3) (1978) 350–353.

[11] E. S. Takle, J. Brown, Note on the use of weibull statistics to characterize wind-speed data, Journal of applied meteorology 17 (4) (1978) 556–559.

[12] G. Bowden, P. Barker, V. Shestopal, J. Twidell, The weibull distribution function and wind power statistics, Wind Engineering (1983) 85–98.

[13] K. Conradsen, L. Nielsen, L. Prahm, Review of weibull statistics for estimation of wind speed distributions, Journal of climate and Applied Meteorology 23 (8) (1984) 1173–1183.

[14] I. Troen, E. L. Petersen, European wind atlas, Risø National Laboratory, 1989.

[15] I. Y. Lun, J. C. Lam, A study of weibull parameters using long-term wind observations, Renewable Energy 20 (2) (2000) 145–153.

[16] K. Ulgen, A. Hepbasli, Determination of weibull parameters for wind energy analysis of izmir, turkey, International Journal of Energy Research 26 (6) (2002) 495–506.

[17] A. H. Monahan, The probability distribution of sea surface wind speeds. Part I: Theory and SeaWinds observations, J. Climate 19 (2006) 497–520.

[18] A. H. Monahan, The probability distribution of sea surface wind speeds. part ii: Dataset intercomparison and seasonal variability, Journal of Climate 19 (4) (2006) 521–534.

[19] P. C. Chu, Probability distribution function of the upper equatorial pacific current speeds, Geophys. Res. Lett. 35 (2008) L12606.

[20] P. C. Chu, Weibull distribution for the global surface current speeds obtained from satellite altimetry, in: Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International, Vol. 3, IEEE, 2008, pp. III–59.

[21] J. A. Carta, P. Ramirez, S. Velazquez, A review of wind speed probability distributions used in wind energy analysis: Case studies in the Canary Islands, Renewable and sustainable energy reviews 13 (5) (2009) 933–955.

[22] P. Kiss, I. M. Jánosi, Comprehensive empirical analysis of ERA-40 surface wind speed distribution over Europe, Energy Conversion and Management 49 (8) (2008) 2142–2151.

[23] M. Kelly, I. Troen, H. E. Jørgensen, Weibull-k revisited: ”tall” profiles and height variation of wind statistics, Boundary-layer meteorology 152 (1) (2014) 107–124.

[24] P. Drobinski, C. Coulais, B. Jourdier, Surface wind-speed statistics modelling: alternatives to the weibull distribution and performance evaluation, Boundary-Layer Meteorology 157 (1) (2015) 97–123.

[25] J. Seguro, T. Lambert, Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis, Journal of Wind Engineering and Industrial Aerodynamics 85 (1) (2000) 75–84.

[26] J. F. Manwell, J. G. McGowan, A. L. Rogers, Wind energy explained: theory, design and application, John Wiley & Sons, 2010.

[27] S. E. Tuller, A. C. Brett, The characteristics of wind velocity that favor the fitting of a Weibull distribution in wind speed analysis, Journal of Climate and Applied Meteorology 23 (1) (1984) 124–134.

[28] E. Bauer, Characteristic frequency distributions of remotely sensed in situ and modelled wind speeds, International Journal of Climatology 16 (10) (1996) 1087–1102.

[29] J. Carta, P. Ramirez, Analysis of two-component mixture weibull statistics for estimation of wind speed distributions, Renewable Energy 32 (3) (2007) 518–531.

[30] M. L. Morrissey, J. S. Greene, Tractable analytic expressions for the wind speed probability density functions using expansions of orthogonal polynomials, Journal of Applied Meteorology and Climatology 51 (7) (2012) 1310–1320.

[31] G. Bel, Y. Ashkenazy, The relationship between the statistics of open ocean currents and the temporal correlations of the wind stress, New Journal of Physics 15 (5) (2013) 053024.

[32] S. A. Akdağ, H. S. Bagiorgas, G. Mihalakakou, Use of two-component Weibull mixtures in the analysis of wind speed in the Eastern Mediterranean, Applied Energy 87 (8) (2010) 2566–2573.

[33] E. C. Morgan, M. Lackner, R. M. Vogel, L. G. Baise, Probability distributions for offshore wind speeds, Energy Conversion and Management 52 (1) (2011) 15–26.

[34] V. L. Brano, A. Orioli, G. Ciulla, S. Culotta, Quality of wind speed fitting distributions for the urban area of Palermo, Italy, Renewable Energy 36 (3) (2011) 1026–1039.

[35] J. Wu, J. Wang, D. Chi, Wind energy potential assessment for the site of Inner Mongolia in China, Renewable and Sustainable Energy Reviews 21 (2013) 215–228.

[36] I. Pobočíková, Z. Sedliačková, M. Michalková, Application of four probability distributions for wind speed modeling, Procedia engineering 192 (2017) 713–718.

[37] T. B. Ouarda, C. Charron, J.-Y. Shin, P. R. Marpu, A. H. Al-Mandoos, M. H. Al-Tamimi, H. Ghedira, T. Al Hosary, Probability distributions of wind speed in the UAE, Energy conversion and management 93 (2015) 414–434.

[38] E. G. Pavia, J. J. O'Brien, Weibull statistics of wind speed over the ocean, Journal of climate and applied meteorology 25 (10) (1986) 1324–1332.

[39] A. H. Monahan, A simple model for the skewness of global sea surface winds, Journal of the atmospheric sciences 61 (16) (2004) 2037–2049.

[40] P. C. Chu, Statistical characteristics of the global surface current speeds obtained from satellite altimetry and scatterometer data, IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing 2 (1) (2009) 27–32.

[41] A. Kabir, I. Lemongo-Tchamba, A. Fernandez, An assessment of available ocean current hydrokinetic energy near the north carolina shore, Renewable energy 80 (2015) 301–307.

[42] Y. Ashkenazy, H. Gildor, On the probability and spatial distribution of ocean surface currents, J. Phys. Oceanogr. 41 (2011) 2295–2306.

[43] Y. Ashkenazy, E. Fredj, H. Gildor, G.-C. Gong, H.-J. Lee, Current temporal asymmetry and the role of tides: Nan-Wan Bay vs. the Gulf of Elat, Ocean Science 12 (3) (2016) 733–742.

[44] S. T. Gille, S. G. L. Smith, Probability density functions of large-scale turbulence in the ocean, Phys. Rev. Lett. 81 (23) (1998) 5249–5252.

[45] S. T. Gille, S. G. L. Smith, Velocity probability density functions from altimetry, J. Phys. Oceanogr. 30 (1) (2000) 125–136.

[46] A. Bracco, J. Lacasce, C. Pasquero, A. Provenzale, The velocity distribution of barotropic turbulence, Physics of Fluids 12 (10) (2000) 2478–2488.

[47] A. Bracco, E. P. Chassignet, Z. D. Garraffo, A. Provenzale, Lagrangian velocity distributions in a high-resolution numerical simulation of the North Atlantic, J. Atmospheric and Oceanic Technology 20 (8) (2003) 1212–1220.

[48] D. P. Dee, S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al., The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society 137 (656) (2011) 553–597.

[49] P. Le Traon, F. Nadal, N. Ducet, An improved mapping method of multisatellite altimeter data, Journal of atmospheric and oceanic technology 15 (2) (1998) 522–534.

[50] N. Ducet, P.-Y. Le Traon, G. Reverdin, Global high-resolution mapping of ocean circulation from TOPEX/Poseidon and ERS-1 and-2, Journal of Geophysical Research: Oceans 105 (C8) (2000) 19477–19498.

[51] K. Pearson, Method of moments and method of maximum likelihood, Biometrika 28 (1/2) (1936) 34–59.

[52] V. K. Rohatgi, A. M. E. Saleh, An introduction to probability and statistics, John Wiley & Sons, 2015.

[53] O. Edenhofer, M. Kalkuhl, When do increasing carbon taxes accelerate global warming? a note on the green paradox, Energy Policy 39 (4) (2011) 2208–2212.

[54] I. Karagali, M. Badger, A. N. Hahmann, A. Peña, C. B. Hasager, A. M. Sempreviva, Spatial and temporal variability of winds in the Northern European Seas, Renewable Energy 57 (2013) 200–210.

[55] M. S. Mitchell, J. D. Spitler, Open-loop direct surface water cooling and surface water heat pump systemsA review, HVAC&R Research 19 (2) (2013) 125–140.

[56] J. Twidell, T. Weir, Renewable Energy resources, Routledge, 2015.

[57] A. S. Bahaj, Generating electricity from the oceans, Renewable and Sustainable Energy Reviews 15 (7) (2011) 3399–3416.

[58] R.-S. Tseng, Y.-C. Chang, P. C. Chu, Use of global satellite altimeter and drifter data for ocean current resource characterization, in: Marine Renewable Energy, Springer, 2017, pp. 159–177.

[59] K. Pearson, X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 50 (302) (1900) 157–175.

[60] W. G. Cochran, The $\chi 2$ test of goodness of fit, The Annals of Mathematical Statistics (1952) 315–345.

[61] A. Mood, F. Graybill, D. Boes, Introduction to the theory of statistics. 3rd mcgraw-hill (1974).

Appendix A. The parameter distribution of the Weibull and GG distributions

Fig. A.11 shows the distribution of the Weibull PDF parameters λ, k, for the surface winds and current speed as discussed in the main text and shown in Fig. 3. Similarly, Fig. A.11 shows the distribution of the GG PDF parameters, λ, k, and ε. As discussed in the main text, the scale parameter λ reflects the mean speed; this is roughly consistent with the range and center of the distributions shown in Fig. A.11a,d, which are typically 10 m s^{-1} and 10 cm s^{-1} for surface wind and surface current speeds, respectively. In all panels (except Fig. A.11d), the GG estimated parameters span a larger range than the Weibull ones. In addition, the k, and ε parameters span a smaller range for the surface currents. We note that very small and very large k and/or ε probably indicate that other distributions, rather than Weibull or GG, may better approximate the data distribution. When the GG parameter $\varepsilon \approx 1$, the GG PDF reduces to the Weibull PDF, and it is evident from Fig. A.11c,f that only a small portion of the distribution of ε is approximately 1, such that for the majority of the global area, the distribution is not Weibull. We elaborate more on this point below (Fig. B.14).

The zonal mean of the different parameters of the Weibull and GG distributions of the surface winds and currents are presented in Fig. A.12. We discuss these results in Sec. 4 of the main text. Also here, the λ parameter reflects the mean wind/current speed and is large at latitudes of large speeds (e.g., for currents at the equator and around $54°S$ for southern ocean winds). It is apparent that there is no clear relation between the λ of the winds and the λ of the currents, suggesting that the wind stress forces the ocean in a non-trivial way and that other sources of energy affect the ocean surface geostrophic currents.

Appendix B. Weibull and GG distribution-specific surrogate data test

In the main text (Sec. 3), we described a surrogate data test that can be applied to general distributions, for each moment and independently from other moments. Below, we suggest a test that is specific to the Weibull and the GG distributions; similar tests can be developed for other distributions too. We start by describing the method for the Weibull distribution with its parameters λ and k; a similar procedure, with the proper adjustments, is then repeated for the GG distribution. Assuming that the time series at hand, x, is Weibull-distributed, we apply the following steps for every geographic grid point:

(i) Estimate the Weibull distribution parameters, λ and k, of the original time series using the MLE method.

(ii) Generate many surrogate Weibull-distributed time series, y, using the λ and k of step (i).

(iii) Use the method of moments (MOM) to approximate the λ and k of the original data x and of the surrogate data y. In the case of a Weibull process, the m^{th} moment is:

$$\langle x^m \rangle = \mu_m = \lambda^m \Gamma\left(1 + \frac{m}{k}\right) \tag{B.1}$$

where $\langle\cdot\rangle$ represents the expected value, Γ is the gamma function, and λ and k are the parameters to be estimated. Based on the data (or surrogate data), we find the ratio, $r_{i,j}$ as follows

$$r_{i,j} = \frac{\mu_i^{j/i}}{\mu_j} = \frac{\Gamma\left(1 + \frac{i}{k}\right)}{\Gamma\left(1 + \frac{j}{k}\right)} \tag{B.2}$$

where i and j are the indexes of two different moments μ_i, μ_j that are calculated from the data (or surrogate data). By taking the ratio, we eliminate λ such that only the k parameter has be found by solving the transcendental equation (B.2); the λ parameter is then found by Eq. (B.1) using the first moment, for example.

(iv) Calculate the 95% CI (as the range of values between the 0.025 and 0.975 quantiles) of the k parameter of the surrogate data estimated in step (iii).

(v) Verify whether the MOM-based k parameters of the original data (from step (iii)) fall within the CI of the surrogate data (step (iv)); if positive, the null hypothesis is not rejected and the original data can be regarded as Weibull-distributed; otherwise, the Weibull hypothesis is rejected.

We now repeat the method described above for the GG distribution. Basically, the only difference is in step (iii) above, but, for the sake of completeness, we present the entire procedure from the beginning to end. Starting from the assumption that the time series at hand, x, is GG-distributed, we proceed as follows:

(i) Estimate the GG distribution parameters λ, k, ε of the original data using the MLE method.

(ii) Generate a large number of GG-distributed surrogate series, y, using the parameters of step (i).

(iii) Use the method of moments (MOM) to approximate the GG parameters (λ, k, ε) of the original data x and of the surrogate data y. The m^{th} moment of the GG PDF is:

$$\langle x^m \rangle = \mu_m = \frac{\lambda^m}{\Gamma(\varepsilon)}\Gamma\left(\varepsilon + \frac{m}{k}\right) \tag{B.3}$$

where $\langle\cdot\rangle$ represents the expected value, Γ is the gamma function, and λ, k, and ε are the GG parameters to be estimated. Based on the data (or surrogate data), we find the ratio, $r_{i,j}$ as follows

$$r_{i,j} = \frac{\mu_i^{j/i}}{\mu_j} = \frac{\Gamma(\varepsilon)^{1-j/i}[\Gamma\left(\varepsilon + \frac{i}{k}\right)]^{j/i}}{\Gamma\left(\varepsilon + \frac{j}{k}\right)}. \tag{B.4}$$

Then, we find the k and ε GG parameters by minimizing the following cost function:

$$\begin{aligned} f(r_{i_1,j_1}, r_{i_2,j_2}) = &\left[r_{i_1,j_1} - \frac{\Gamma(\varepsilon)^{1-j_1/i_1}[\Gamma\left(\varepsilon + \frac{i_1}{k}\right)]^{j_1/i_1}}{\Gamma\left(\varepsilon + \frac{j_1}{k}\right)}\right]^2 + \\ &\left[r_{i_2,j_2} - \frac{\Gamma(\varepsilon)^{1-j_2/i_2}[\Gamma\left(\varepsilon + \frac{i_2}{k}\right)]^{j_2/i_2}}{\Gamma\left(\varepsilon + \frac{j_2}{k}\right)}\right]^2 \end{aligned} \tag{B.5}$$

where (j_1, i_1) and (j_2, i_2) indicate two different sets of moments. The λ parameter is then found using Eq. (B.3), using the first moment, for example.

(iv) Calculate the 95% CI (using the 0.025 and 0.975 quantiles) of the k and ε of the surrogate data that were estimated using the MOM (detailed in step (iii)).

(v) Verify whether the original datas MOM-estimated parameters fall within the CI of the surrogate data (step (iv)); if positive, the null hypothesis is not rejected, and the data can be regarded as GG-distributed, while otherwise, the null hypothesis is rejected, and the data cannot be regarded as being GG-distributed.

We use the same datasets analyzed in the main text (Sec. 3), namely the ERA-Interim surface winds and geostrophic surface currents that are derived from altimetry measurements. The above tests were applied to every grid point separately. The results of the above Weibull MOM method are depicted in Fig. B.13a,b. The analysis is based on the first and second moments. The results indicate that the surface wind speed over the tropical ocean, Antarctica and Greenland are not Weibull-distributed as the k parameter of the assumed Weibull distribution falls outside the CI interval of the surrogate data. More generally, 60% of the global area of the k parameter of the Weibull distribution falls within the CI interval of the k parameter of the surrogate data. As for the surface currents, the k parameter of 78% of the analyzed area falls within the CI of the surrogate data. The results presented in B.13a,b are based on the first and second moments—other set of moments yielded different results, and the percentage of area that falls within the CI of the surrogate data decreases as the chosen moments increase; see Eq. (B.2).

Fig. B.13c,d,e,f depicts the results of the GG parameters. The analysis is based on moments $m = 1, 2, 3, 4$. Surprisingly, the more general GG distribution yielded a much larger area that falls outside the CI interval of the surrogate data; only for ∼28% of the analyzed areas did the k and ε GG parameters fall within the CI of the k and ε of the surrogate data. This is also valid for the k and ε GG parameters of the surface currents presented in Fig. B.13d,f where the area within the CI is ∼50%. These percentages, both for wind and currents, are much smaller than the percentages we obtained for the Weibull distribution (60% and 78% for the k parameters of the assumed Weibull distribution, Fig. B.13a,b) despite the fact that the GG is a more general distribution (compared to the Weibull distribution) that should result in a larger area that falls within the CI interval of the surrogate data. Most probably, these smaller percentages for the GG distribution are related to the fact that we used four moments ($m = 1, 2, 3, 4$) for the GG analysis and only two ($m = 1, 2$) for the Weibull distribution; generally speaking, higher moments yield a smaller area that falls within the CI of the surrogate data. This is consistent with Fig. 7.

In Fig. B.13c,d, we present the results of the k parameter of the GG distribution, while in Fig. B.13e,f, we present the results of the ε parameter of the GG distribution. As expected, the results of the two parameters are very similar, as the method solved the two parameters simultaneously. Thus, it is sufficient to concentrate on one of these parameters to make conclusions regarding the assumed probability.

Appendix C. A method of verifying whether a distribution is indeed Weibull

The GG distribution is a generalization of the Weibull distribution, such that when the ε parameter of the GG distribution is equal to 1, $\varepsilon = 1$, the GG distribution reduces to the Weibull distribution; see Eqs. (1), (2). We use this fact to verify whether an assumed Weibull distribution is indeed Weibull. Assuming that the time series at hand, x, is Weibull-distributed, we apply the following steps:

(i) Estimate the Weibull distribution parameters, λ and k, of the original data using the MLE method.

(ii) Generate (many) Weibull artificial time series with the same λ and k and the same length as the original time series.

(iii) Using the MLE, estimate the GG parameters, λ, k, and the ε of the time series from the previous step. The ε parameter should be scattered around 1, $\varepsilon \approx 1$.

(iv) Calculate the 95% CI interval of the ε parameter from step (iii).

(v) Estimate the GG distribution parameters of the original data and check whether the ε parameter of the original data is indeed close to 1 and falls within the CI interval of (iv). If positive, the data can be regarded as Weibull-distributed, while if negative, they are not.

The results of the method described above are presented in Fig. B.14. With regards to the surface wind speed, it is apparent that only 8% of the global area falls within the CI interval of $\varepsilon \approx 1$, indicating that only 8% of the globe can be considered as Weibull-distributed. The percentage is even lower for the ocean surface currents where only 7% of the analyzed area falls within the CI of $\varepsilon \approx 1$.

We note that in the above test, we assumed that the distribution is either Weibull or GG. It is possible that none of these distributions satisfactorily account for the distribution of the original data. This may be the reason for the low percentage we obtained in this test.

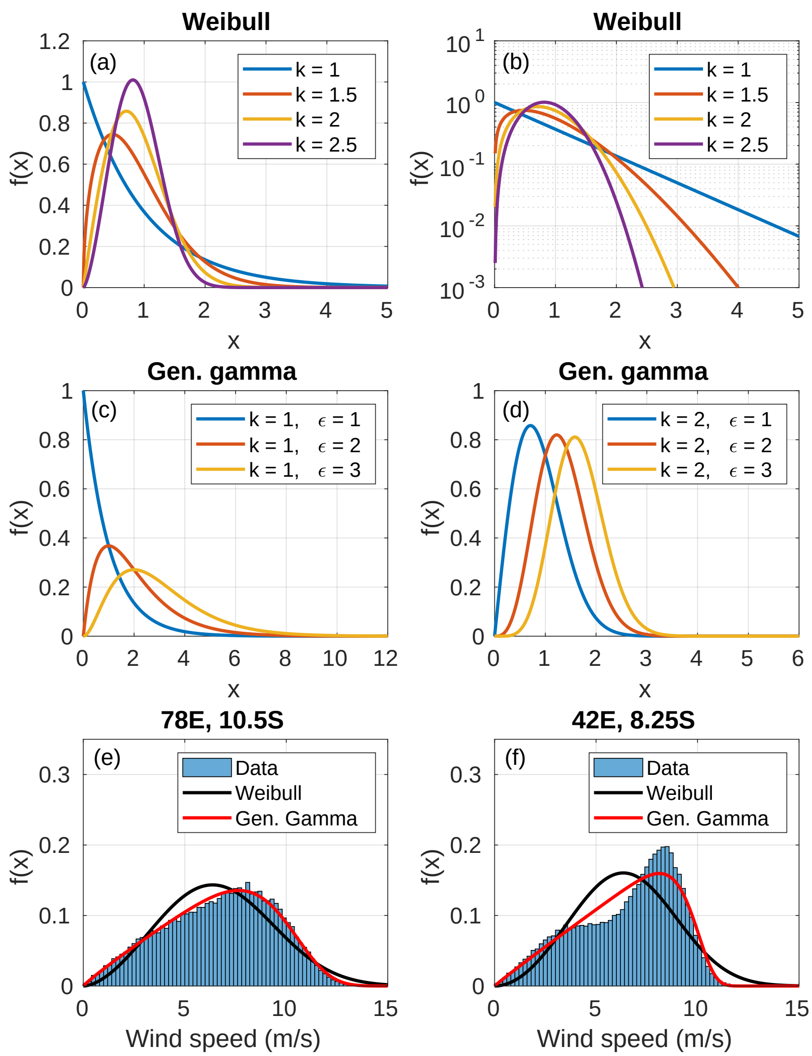

Figure 1: A few illustrative examples of the probability density function (PDF) of the Weibull distribution when the scale and shape parameters are $\lambda = 1$ and $k = 1, 1.5, 2, 2.5$, in (a) regular and (b) semi-log plots. (c) Examples of the PDFs of the GG distribution for $\lambda = 1$, $k = 1$ and $\varepsilon = 1, 2, 3$. (d) Same as (c) for $\lambda = 1$, $k = 2$ and $\varepsilon = 1, 2, 3$. Two particular instances of sample distributions of surface wind speeds (sampled from 1979 - 2016 at a frequency of 6 hours), as well as the corresponding Weibull and GG approximations at (e) 78°E, 10.5°S [the Weibull parameters are $\lambda = 7.5$m s^{-1}, $k = 2.7$, and the GG parameters are $\lambda = 10.4$m s^{-1}, $k = 6.4$, $\varepsilon = 0.3$] and (f) 42°W, 8.25°S [the Weibull parameters are $\lambda = 7.3$ m s^{-1}, $k = 3$ and the GG parameters are $\lambda = 10$ m s^{-1}, $k = 12.2$, $\varepsilon = 0.2$].

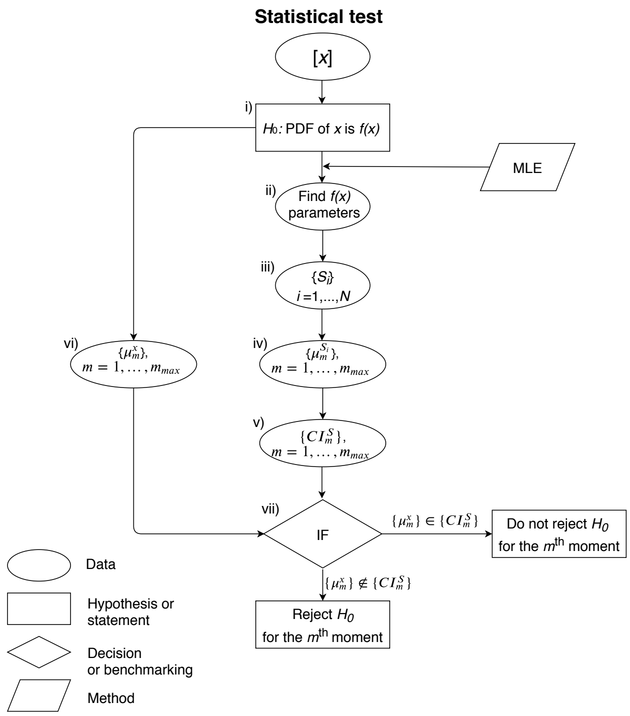

Figure 2: A flow chart showing the various steps of the analysis to test whether a specific assumed distribution $f(x)$ (either Weibull or GG) fits a given time series $[x]$ (in our case surface wind and current speed time series). The chart can be used to (i) test either a Weibull or a GG hypothesis. In step (ii) we apply a method of estimating the parameters of the hypothesis $f(x)$ by maximizing a likelihood function (MLE method). Therefore, using the assumed approximate distribution, (iii) we generate a large number $(i \sim 300)$ of surrogate (synthetic) time series $\{S_i\}$ where each series has the same length of the measured data. Thereafter (iv) we calculate $m = 1, \ldots, m_{max}$ moments μ_m of each individual surrogate S_i. On the basis of this set of surrogate moments, we estimate in (v) the 0.95 confidence interval (CI) of each moment. In step (vii) we check whether the value of the moment of original data x (calculated in vi) falls within the corresponding confidence interval, CI; the initial null hypothesis H_0 is not rejected if the moment of the original data falls within the CI of the surrogate data while otherwise the null hypothesis is rejected. The method was applied to all series at hand (surface wind and current speed) to test both Weibull and GG distributions but can be generalized to any given distribution that support the MLE method used as the initial estimator of distribution parameters.

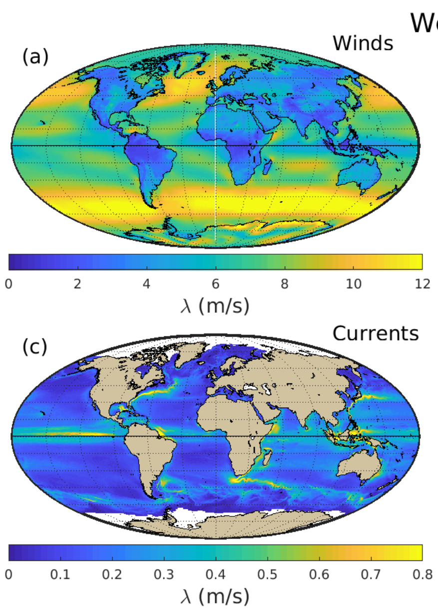

Figure 3: Maps of the Weibull distribution parameters, λ (left panels, in $\mathrm{m\,s^{-1}}$) and k (right panels) of surface wind speed (upper panels) and surface current speed (lower panels). The parameters were estimated based on the MLE method. The brown color in the lower panels indicates the land regions, while the white color indicates no available data.

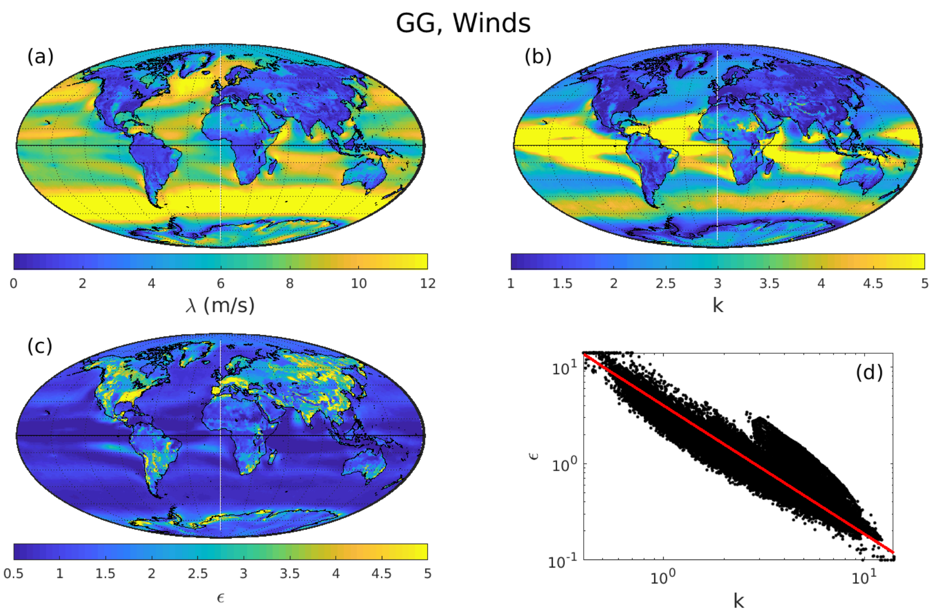

Figure 4: Maps showing the surface wind speed GG parameters (estimated using the MLE) (a) λ (in $\mathrm{m\,s^{-1}}$), (b) k, and (c) ε. (d) The ε GG parameter versus the k GG parameter showing that the two are not fully independent—the red line indicates the relation $\varepsilon = 4k^{-4/3}$.

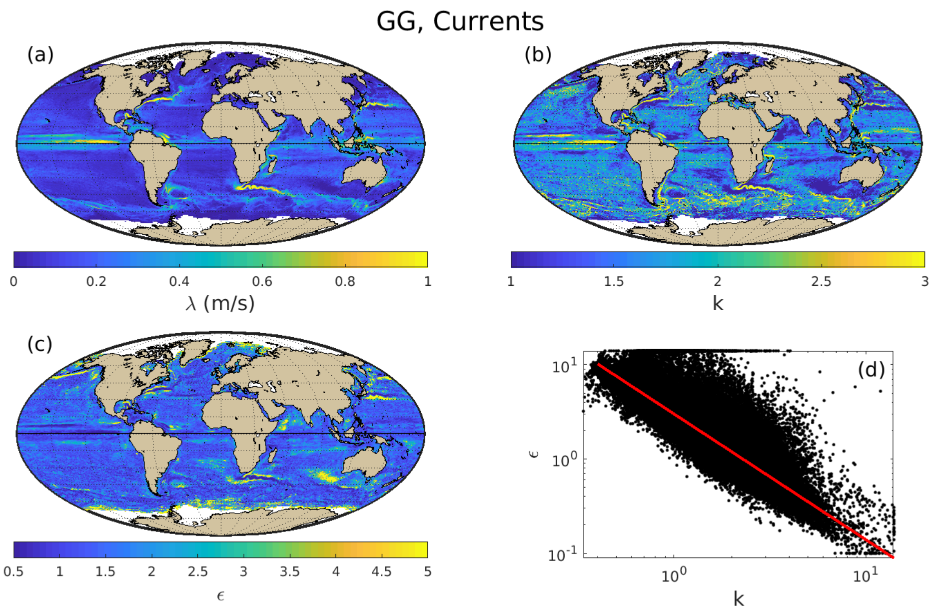

Figure 5: Same as Fig. 4 for the surface ocean current speed. The red line in (d) indicates the relation $\varepsilon = 3k^{-4/3}$.

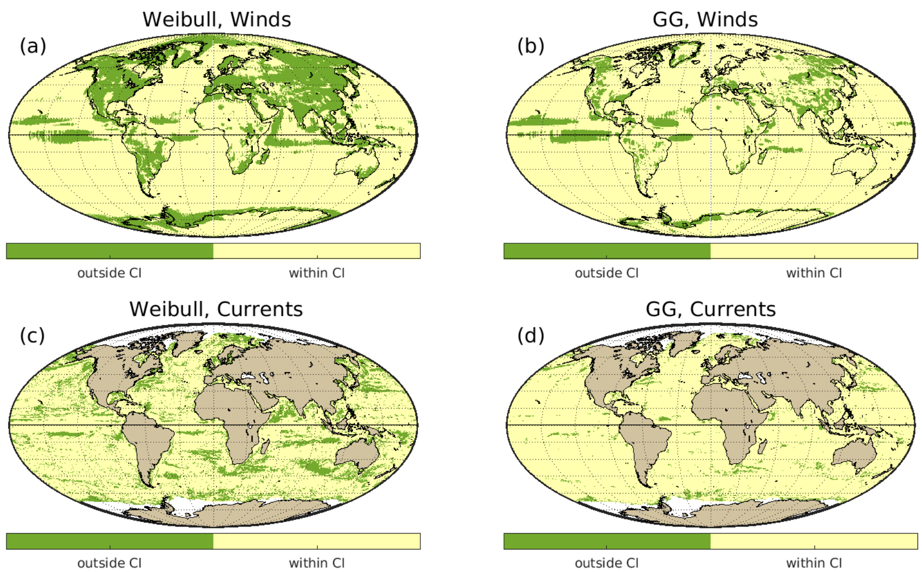

Figure 6: Maps showing the areas where the surface wind speed [(a),(b)] and the surface current speed [(c),(d)] are Weibull-distributed [(a),(c)] or GG-distributed [(b),(d)]: positive (within the 95% CI, yellow), negative (outside the 95% CI, green). The brown color indicates land areas, while the white color indicates no available data. Results are based on the surrogate data method (using the third moment) described in Sec. 3 and in Fig. 2. The percentage of the analyzed area that falls within the 95% CI is (a) 78%, (b) 89%, (c) 80%, and (d) 95%.

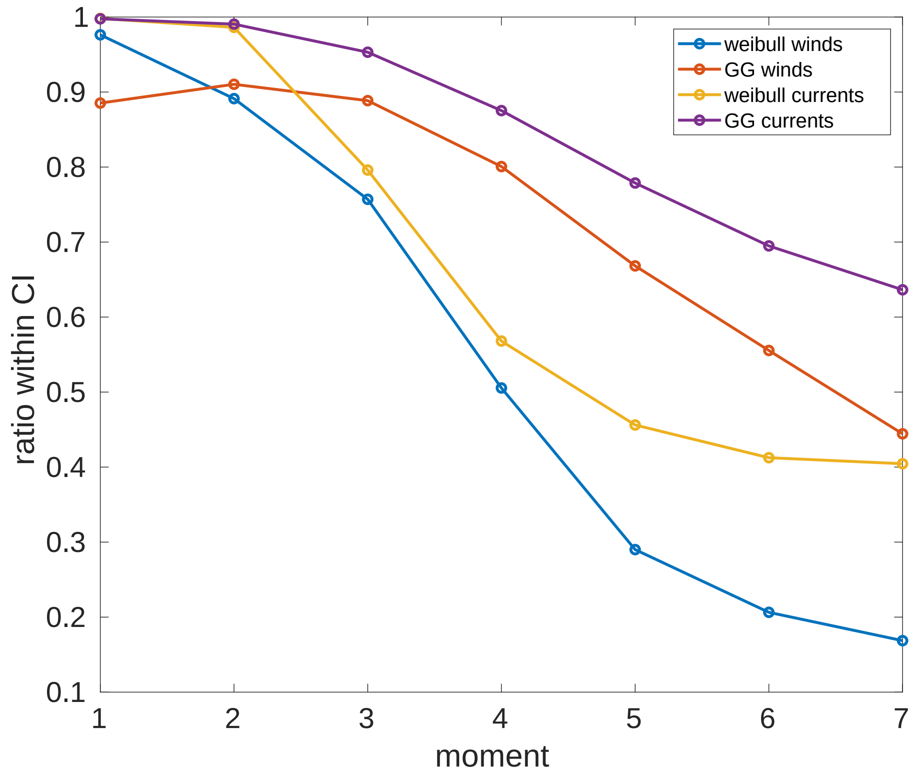

Figure 7: The proportion of the analyzed area that falls within the CI of the assumed distribution for the surface wind and current speed as a function of moment.

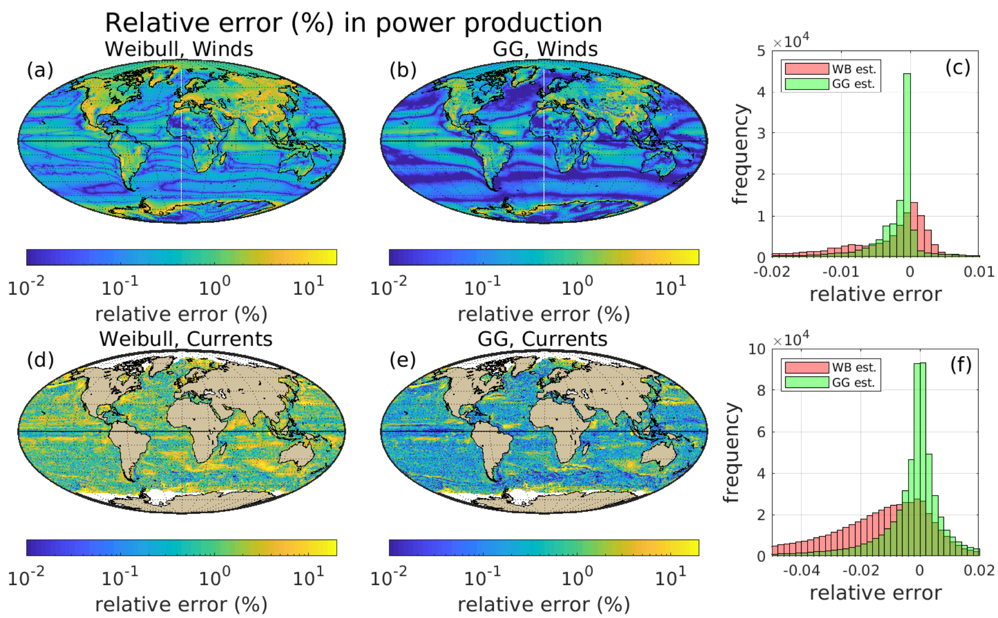

Figure 8: Maps showing the absolute value of the percentage error in the potential wind power per unit area as estimated using the Weibull [(a),(d)] and GG [(b),(d)] distributions for surface winds [(a),(b)] and surface currents [(d),(e)]. Frequency histograms showing the relative errors of the assumed Weibull (WB) and GG distributions for the surface winds (c) and surface currents (f). The red histograms indicate the errors obtained by estimation carried out using the Weibull hypothesis, while the green histograms indicate the error when assuming the GG hypothesis; the overlapping histogram region is indicated by the dark green color.

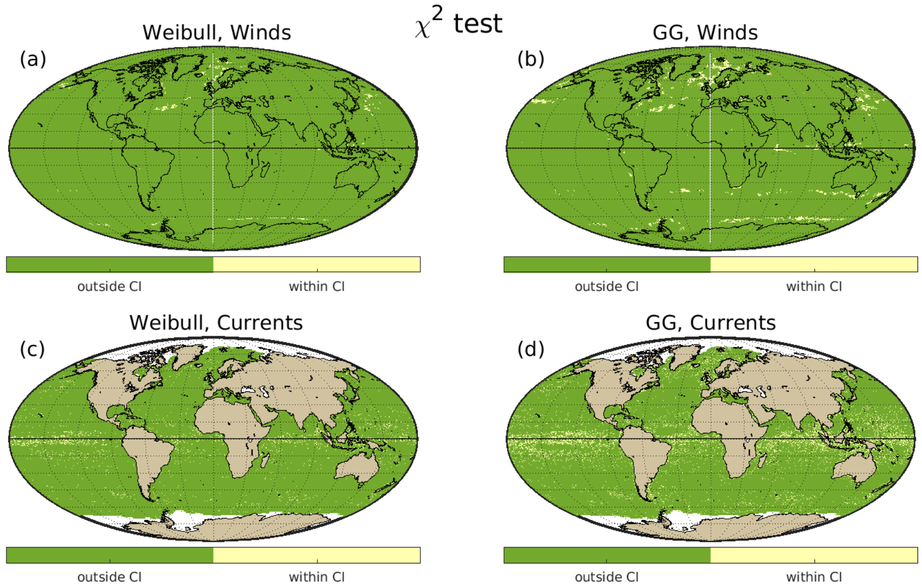

Figure 9: Results of the χ^2-test. Maps showing the areas where the surface wind speed [(a),(b)] and the surface current speed [(c),(d)] are Weibull-distributed [(a),(c)] or GG-distributed [(b),(d)]: positive (within the 95% CI, yellow), negative (outside the 95% CI, green). The brown areas refer to land areas, while the white color indicates no available data. The percentages of geographic grid points falling within the 95% CI are (a) 0.4%, (b) 1.6%, (c) 2.9%, and (d) 8.3%.

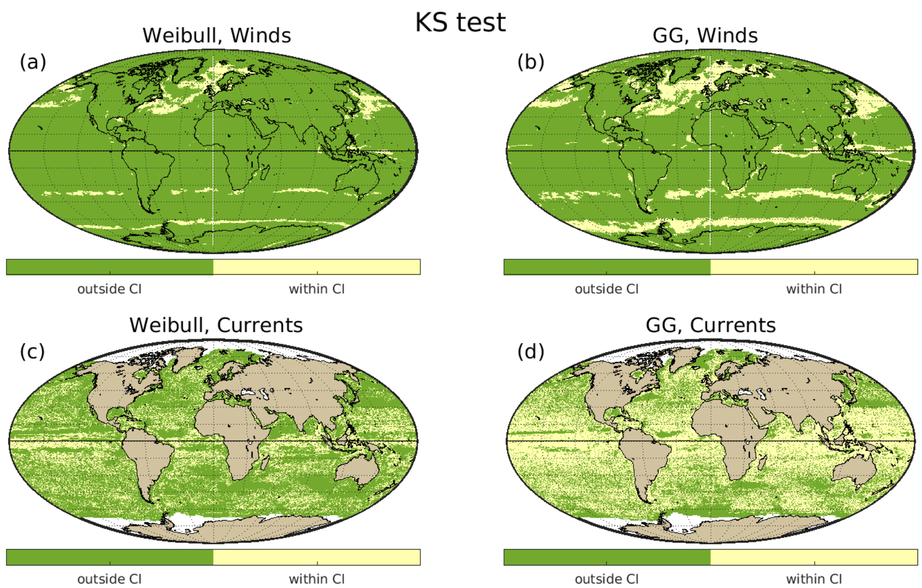

Figure 10: Results of the Kolmogorov-Smirnov test. Maps showing the areas where the surface wind speed [(a),(b)] and the surface current speed [(c),(d)] are Weibull-distributed [(a),(c)] or GG-distributed [(b),(d)]: positive (within the 95% CI, yellow), negative (outside the 95% CI, green). The brown color refers to land areas, while the white color indicates no available data. The percentages of geographic grid points falling within the 95% CI are (a) 4.7%, (b) 11.5%, (c) 29.8%, and (d) 60.1%.

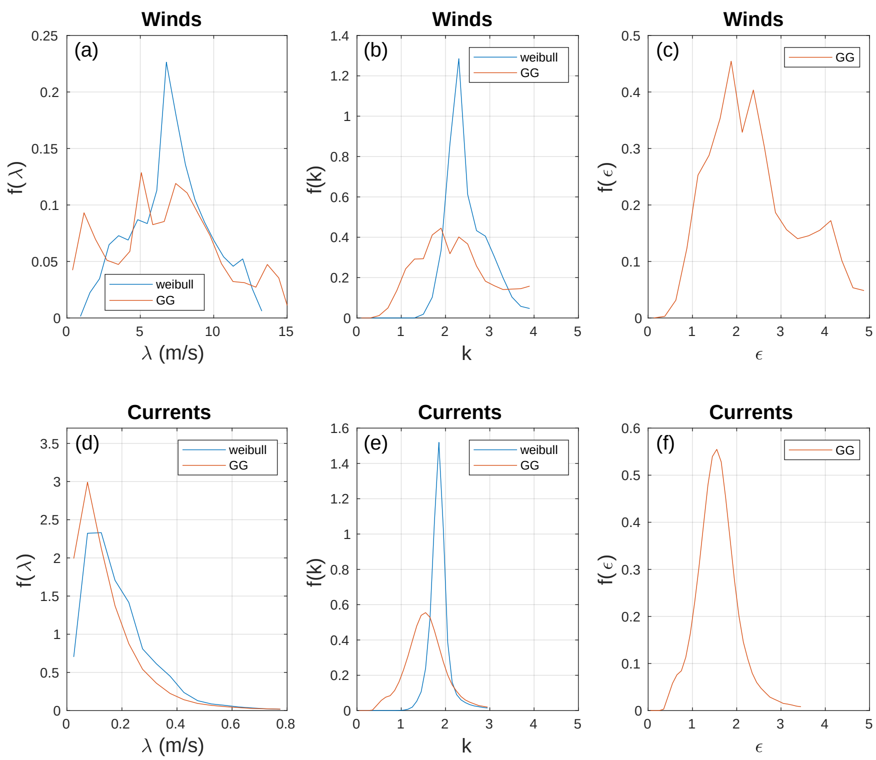

Figure A.11: The distribution of the parameters of the Weibull (blue) and GG (red) distributions estimated using the MLE method, for surface wind speed (upper panels) and surface current speed (lower panels). The λ parameter is shown in panels (a) and (d), the k parameter is shown in panels (b) and (e), and the ε parameter is shown in panels (c) and (f).

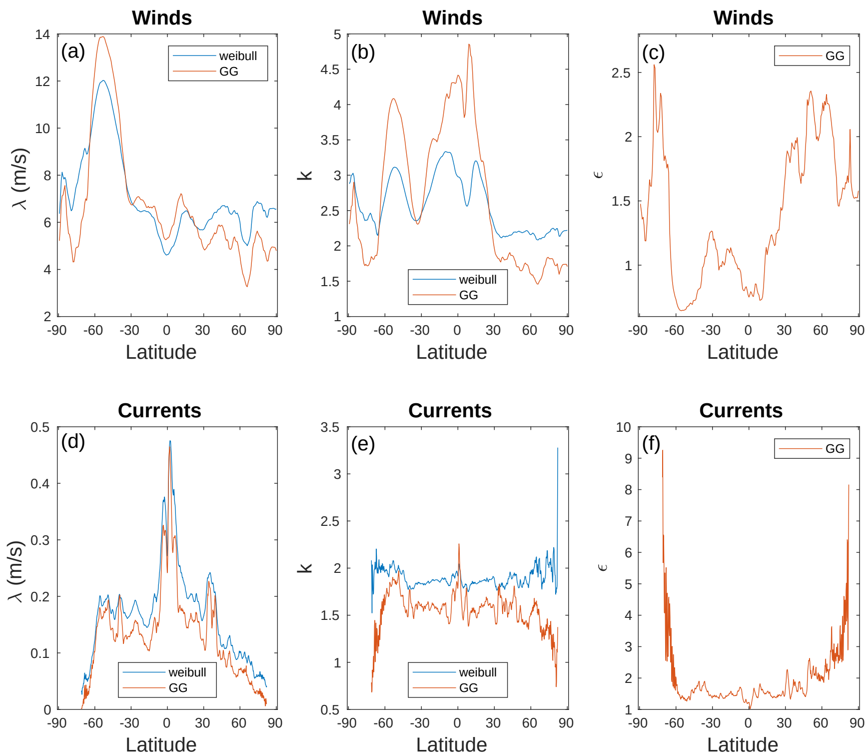

Figure A.12: Zonal mean of the MLE-fitted Weibull distribution parameters (blue) and GG distribution parameters (red) for surface wind speed (upper panels) and surface current speed (lower panels). (a),(d) λ parameter, (b),(e) k parameter, and (c),(f) ε parameter.

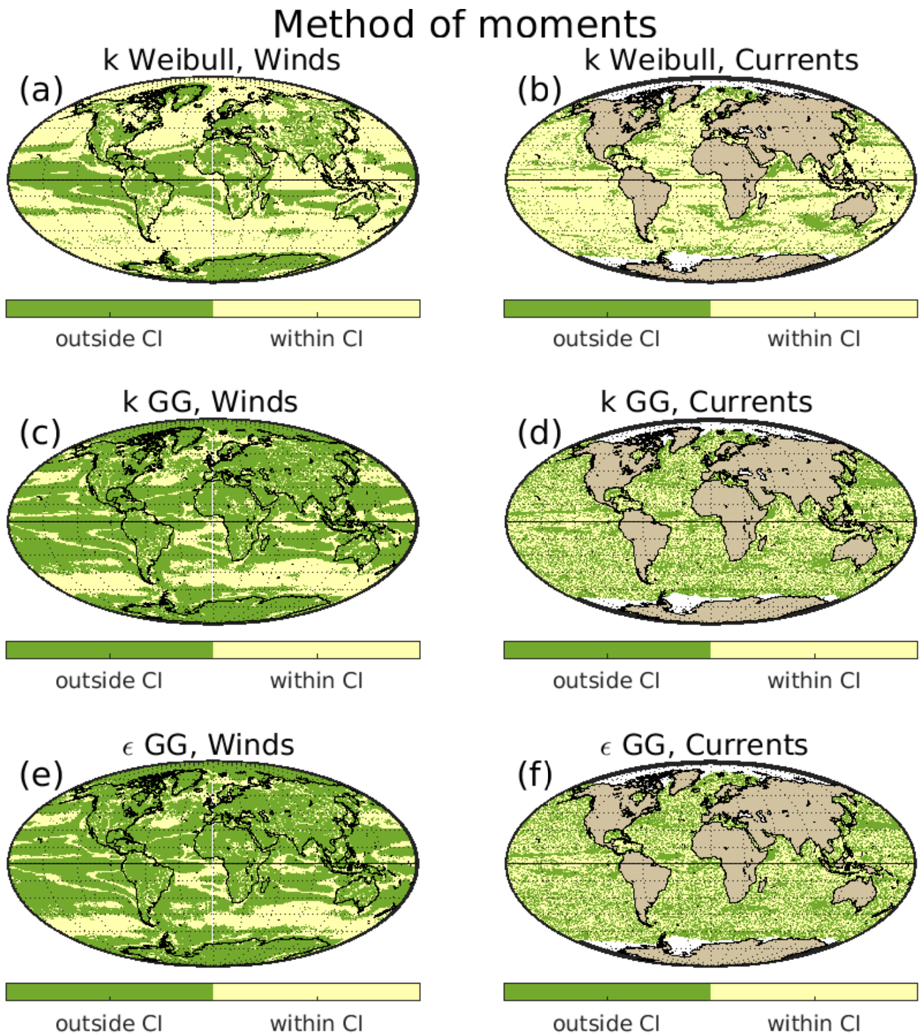

Figure B.13: Maps of the areas where the shape parameter, k, of the data falls within the CI of the k of the surrogate data. The maps are based on the method of moment (Appendix B) assuming a Weibull distribution of (a) surface wind speed and (b) surface ocean currents. (c),(d) same as (a),(b) for the k parameter of the GG distribution and (e),(f) are the same as (c),(d) for the ε GG parameter. The brown color in panels b,d,f indicates the land areas, while the white color indicates no available data. The percentage of the analyzed area that falls within the CI is: (a) 60%, (b) 78%, (c) 29%, (d) 50%, (e) 27%, (f) 49%.

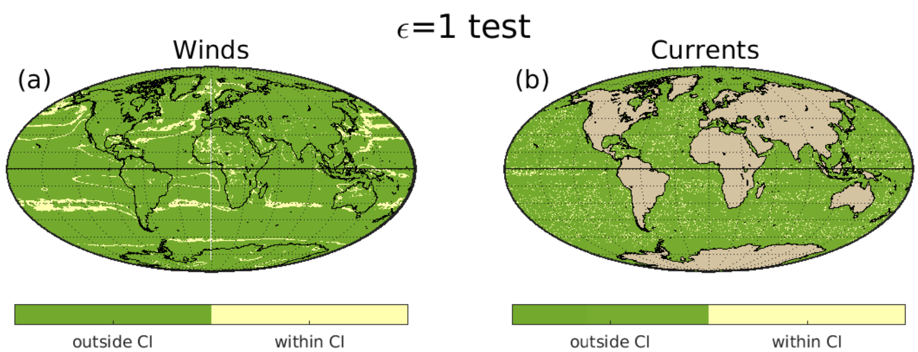

Figure B.14: Maps of the areas where the estimated ε of the GG distribution is within (or outside) the CI of the corresponding ε of the surrogate data with $\varepsilon \approx 1$. Maps for the (a) surface wind speed and (b) surface current speed. Areas with $\varepsilon \approx 1$ indicate that the underlying distribution is likely to be Weibull. The brown color in panel b indicates land areas, while the white color indicates no available data. The results are based on whether ε lies within the 95% CI of 1 as determined from 300 surrogate time series. The length of the surrogate is the same as the original data. The ratio (in %) of data falling within the CI is (a) 8%, (b) 7%.