Ocean Sci. Discuss., 11, C1398–C1405, 2015
www.ocean-sci-discuss.net/11/C1398/2015/

Ocean Science
Discussions
Open Access

Interactive comment on "A Monte Carlo simulation of multivariate general Pareto distribution and its application" *by* L. Yao et al.

L. Yao et al.

yaoluo@scsio.ac.cn

Received and published: 10 March 2015

Reply to Anonymous Referee #2:

General comments: The manuscript presents a novel method, called MGPD, to the world of Ocean Engineering. It presents some of the theory of MGPD, and gives an example how to model the extremes of wind and waves, taking their correlation into account. I find the theoretical background hard to understand, as will be the case for most of the readers of Ocean Science Discussions as well (I guess). On the other side, the theoretical background is not comprehensive enough to be self-explaining. The same holds for the wind/wave example, which does not convince me that the MGPD works as expected. My conclusion is that the manuscript is hard to understand, and will not invite readers to use the MGPD approach in Ocean Science.

Thank you very much for your suggestions and helpful comments to our work. Here we copy your original comments, which are followed by our replies. The new paper has made large changes, and many statements have been modified.

Specific comments: p 2734, line 13: "rather small" should be avoided: either quantify the error, or remove the word "rather". Accepted, Page 1

p 2735, line 13: Change into "Rootzen and Tajvidi (2006) suggest, based on the research by Tajvidi (1996), that... Accepted, line 35 p 2.

p 2736, line 16: either use the abbreviation MGPD, or write all words fully. Now, it is a mixed of full words and abbreviations. Accepted,

p 2736, line 19: "more stable calculation results". One of the disadvantages of (M)POT over (M)GEV/Gumbel is that one has to choose a threshold, which may influence the fit considerably. It is thus not trivial that (M)POT results in more stable results than (M)GEV/Gumbel Besides, Luo et al. (2012) selected 20 or 30-year 6 samples from about 60-year wave raw data arbitrarily. The 6 samples were analyzed by POT and AMS respectively, and the result shows that the return levels with the POT method are closer to the return levels from 60-year wave raw data and have smaller fluctuations than AMS. Because the POT method can get as much extreme information as possible from raw data, its result may be more stable.

p 2739, line 14: "..were not proposed..." Sentence seems rather cripple to me. Accepted,

p 2741, line 8: The selection procedure is not clear to me. If a 5-day interval is chosen, the interval between two extremes cannot be (less than) 2 day, so I don't understand line 8. Additionally, is the extreme wind chosen, and the corresponding wave height at the same moment added, or is the extreme wind in a 5-day interval connected to the extreme wave height in the same 5-day interval? Please clarify. 5-day is time window, the maximums from the adjacent time windows may be found in 2 days. The corresponding wave height is at the same moment. Means the concurrent wind and wave.

p 2741, line 19. Add "probability plot of THE marginal distributionS." p 2741, Eq. 13. Why is the Pearson Type III chosen to model the return levels, and not the GEV distribution (which is mentioned several times to be the natural distribution for extremes).? Because the GEV distribution can't fit the annual maximum series of the raw data well, so the Pearson Type III is chosen to model the return levels. The deeper causes of the phenomena have not found. Some articles also show that the GEV distribution can't fit the annual maximum series sometimes.

p 2742, line 1. The motivation for the choice of the bivariate logistic generalized Pareto distribution is missing. The motivation will be showed, because "The dependence models between extreme variables have been suggested: Logistic, Bilogistic and Dirichlet. However, it appears that the choice of dependence model is not usually critical to the accuracy of the final model (Morton and Bowers, 1996)."

p 2742, line 16: "The simulation results are in agreement with the actual situation". I disagree with the conclusion, as Figure 4 shows that the observations show much more 40m/s events than the stochastic simulation indicates. Additionally, these 40m/s events correspond with much lower wave heights than the simulation indicates. The simulation method is done base on the distribution function, and the probability of the simulation results is in basic agreement with the actual situation. Table 2&3 illustrate some too. Due to restrictions with observation technologies at the time, the wind speed and the wave height were kept only the integer and one decimal place respectively. So the 40m/s events can't be fitted well.

p 2743, line 11. Please add the return levels to the table as well (p 2747, table 2) Accepted

p 2744, line 10. I disagree with the conclusion that, by Monte Carlo simulation, a better estimate can be made than based on the traditional extreme-value methods, as no ex-

tra information is added. Accepted. These conclusions have been changed.P13∼P14

p 2747, Table 2&3. I think that addition of V=30 and V=40 (instead of V=1 and V=2) makes the table more interesting. I suspect that especially at these high levels, the difference between the analytic solution and the simulation increases. If not, it would be worthwile to intercompare the direct observation with the simulation as well. V=1 means the V=VP%=1? If so that, V is so small that the return period is meaningless in the actual projects. If V >45 or more, the table 2 shows the absolute error may become larger, but the relative error is small still.

p 2750, Figure 1. Add a note why the empirical distributions are discrete. I assume that the observations are stored with a limited amount of digits. Due to restrictions with observation technologies at the time, the wind speed and the wave height were kept only the integer and one decimal place respectively. This will influence the level of precision of extreme value, so in fig. 2 (a), the all the wind values in the range 0.4 and 0.6 (standard unit) show the same probability

p 2750, Figure 2. Please add at the upper X-axis the return level values. Accepted

p 2750, Figure 2. It seems to me that the wind distribution shows a 'kink' around 20m/s, which might be attributed to a different phenomenon that causes the wind: normal depressions for lower winds, typhoons for higher winds. If so, this violates the assumption of EVT that all extremes originates from a single parent distribution. In figure 2, the data is the annual maximum series. The lower winds and higher winds is caused by lower-winds years and high-wings years. The all big typhoons maybe far away from our observation point, so the annual maximum is about 20m/s. So the parent distributions may be the same still.

Please also note the supplement to this comment:
http://www.ocean-sci-discuss.net/11/C1398/2015/osd-11-C1398-2015-supplement.pdf

Interactive comment on Ocean Sci. Discuss., 11, 2733, 2014.

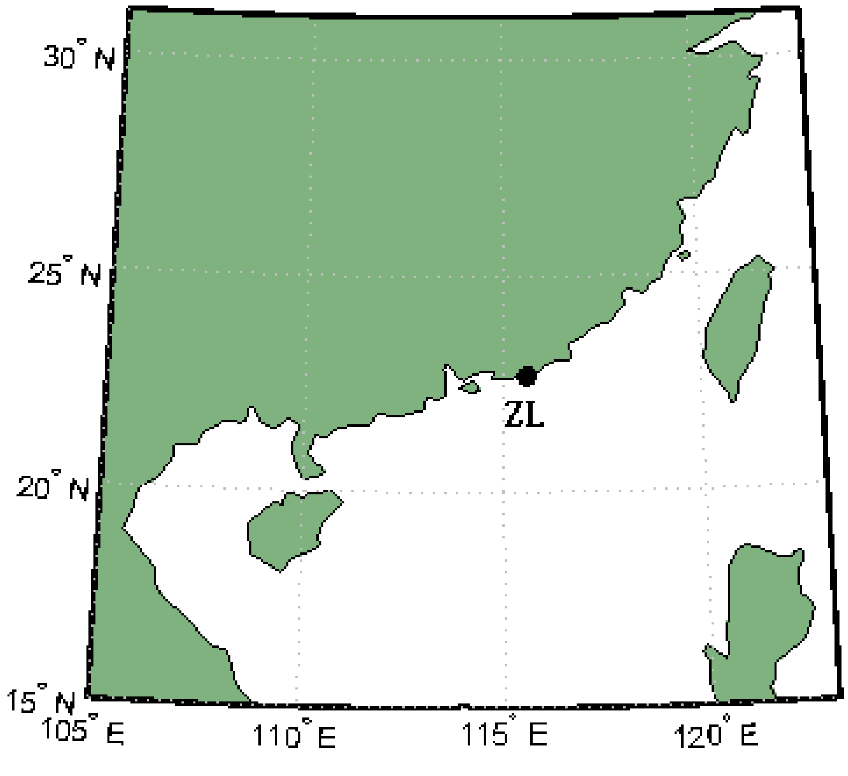

Fig. 1.

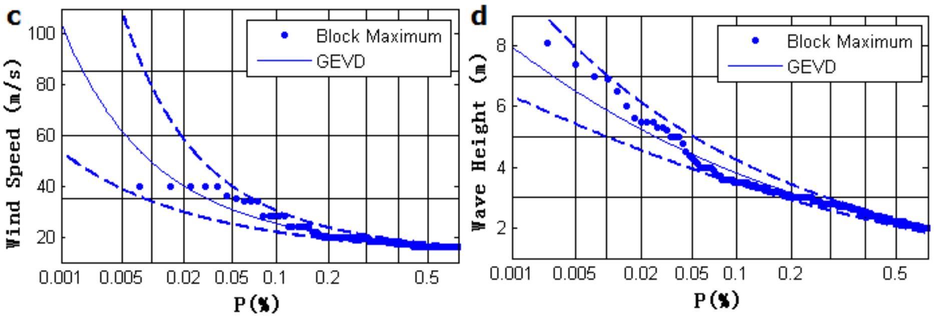
c
d
Block Maximum
GEVD
Wind Speed (m/s)
Wave Height (m)
P(%)

Fig. 2.

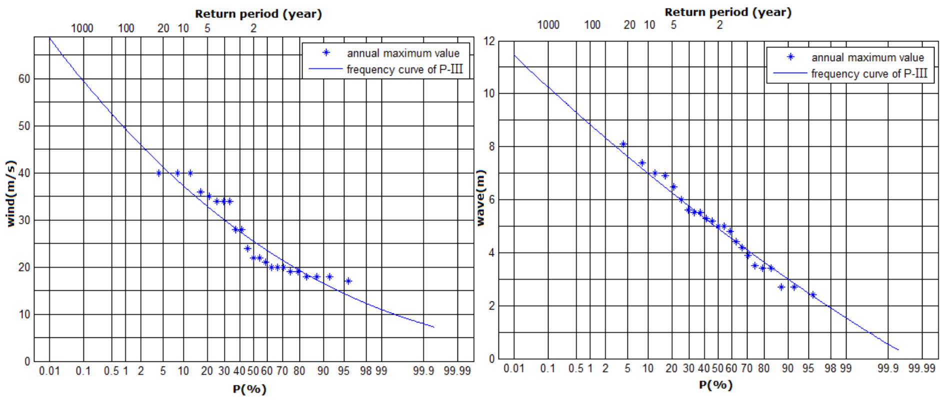
Return period (year)
annual maximum value
frequency curve of P-III
wind(m/s)
wave(m)
P(%)

Fig. 3.