I still have some doubts about the robustness of the results, especially when applied to the Northern part of the Adriatic. ERA-40 is known to be quite deficient in enclosed areas near mountains. It is not obvious to me that a simple calibration could remedy these shortcomings. Moreover, I find it suspicious that there is an apparent change in some of the derived trends at about 1989, a time when satellite data became widely used in ERA40. Comments on the appropriateness of ERA40 for this analysis will be most welcome.

I found the first part of the paper (introduction, data set-up and calibration) very hard to read. It leaves the reader quite confused. On the other hand, the statistical analysis is
clearer.

Specific comments:

P2006, line 23: did you mean to say: The study of extreme events has been increas-
ingly motivated by environmental sciences due to the numerous applications of these studies, ranging . . .

P2007, line 1: rewrite to read “The aim of the scientific community at preventing ex-
treme events that harm people and infrastructures has been in trying to better under-
stand such extreme events and in the process, reduce the uncertainty related to their assessment.”

P2007, line 14: could be or not a risk → could be a risk or not

P2007, line 15: didn’t . . . yet → has not yet resulted in establishing standard criteria

P2008, line 12: runs → has been running

P2008, line 13: is an elaboration . . . mid 2002 → is a re-run from mid 1957 to 2002 of a fixed version of ECMWF atmospheric analysis system, including WAM, at a resolution of about 125 km. It uses all available observations to constraint the analysis.

P2008, line 15: which version of WAM was used, what is the spectral resolution, the first frequency? Was the shallow water physics option used? At what resolution were the winds obtained from ECMWF?

P2008, line 21: underestimate → underestimation

P2009, line 1: Were the satellite data corrected?

P2009, line 2-5: Not clear. Did you use Cavaleri and Sclavo results and then calibrate ERA-40 UNILE?

P2009, line 12: has length 42 yr going → spans a 42 year period from

P2009, line 24: I don’t think the underestimation is due to shallow water but rather to a
lack of shallow water details (0.5 degree is still very coarse) and also to the coarseness of the wind forcing.

P2010, line 4: fitting to the → fitting the

P2010, line 12: using more → using more data

P2010, line 18: provenance → origin

P2010, line 25: Actually . . . sample → Actually, even after the POT and rmax methods are applied to the same iid sample,

P2015, line 23: this finds agreement → this agrees

P2020, line 9: The obtained . . . employment of → The results also suggest a careful use of

P2020, line 26: Wind Model Data → Wind and Wave Model data

Table 1 would be better if it was made to correspond to figure 1, in which the points were to be labelled, with the corresponding values in the table.