Sensitivity to the lowest data value

What's wrong? Let's carry out an accurate sensitivity study with respect to the threshold (Fig. 1 - top) using the classical IAHR Haltenbanken dataset: the quantile (and parameter) ML estimates are unstable between 2 consecutive data values (“virgates”). A slight translation of the sample may lead to a significant change in the estimated quantiles! The maximum likelihood of the GPD-2 model (Fig. 1 - bottom) is continuously increasing with the threshold between two consecutive data values and shows a positive jump when the threshold reaches a new sample value.

How do we explain that? Considering that:
- You still work with the same physical events (storms) when \( \text{u}_2 \) varies between two data so the final result shouldn’t change;
- Letting \( \text{u}_2 \) vary between two data sets the origin of the distribution while this role is usually devoted to a location parameter;
- The role of the threshold should be limited to the selection of extreme data to be fitted;

we claim the need for a location parameter \( \mu \), distinct from the threshold, to improve the stability of the results.

Comparison with a GPD-3 For comparative purpose, we fit a 3-parameter GPD (GPD-3) with the L-Moments estimator (Hosking, 1990). Quantile (Fig. 1 - top) as well as shape \( \kappa \) and scale \( \sigma \) parameter estimates are stable! But the 100-yr \( \mu \) is 1 higher…

Examination of the GPD-3 likelihood Test: random generation of an ordered sample of simulated data \( Y_{N} \) (N = 25), then computation of the likelihood for each triple \((\mu, \sigma, \kappa)\) within \(x=10^{-12}\ldots1\) with the condition \( \sigma > 3(\mu - Y_{N}) \). For each constant value of \( \mu \), there is a maximum likelihood in the \((\kappa, \sigma)\) plane (Fig. 2). But the global maximum for the full (3D) parameter space is reached at the open upper bound of the interval of \( \mu \) (i.e. \( Y_{N} \)), with non-null derivatives (Fig. 3), while \( \mu_0 \) doesn’t converge (Fig. 4). Still, the asymptotic properties of the MLE require that the maximum be reached on an interior point of an open set (Lehmann, 1983). These properties are not proven!

> The use of MLE is quite dubious.
> An accurate and robust estimation of \( \mu \) is necessary!

Conclusions

The statistical threshold \( \text{u}_2 \) selects the data to be fitted and defines the “extreme domain”, while the location parameter \( \mu \) sets the origin of the distribution. These two roles are distinct and \( \text{u}_2 \) and \( \mu \) should not be confused.

A comparison MLE/GPD-2 vs LMOM/GPD-3 shows that introducing \( \mu \) yields stable results between two consecutive storm peaks, consistent with the physics.

References:


Context:

Franck Mazas - ARTELIA Maritime
6, rue de Lorraine
38130 Echirolles – FRANCE
franck.mazas@arteliagroup.com