



Statistical flood frequency analysis – an evaluation of the subjectivity in model selection

Kolbjorn Engeland (1), Thordis Thoraninsdottir (2), and Ingelin Steinsland (3)

(1) The Norwegian Water Resources and Energy Directorate, Oslo, Norway (koe@nve.no), (2) Norwegian Computing Center, Oslo, Norway, (3) Norwegian University of Science and Technology, Trondheim, Norway

The operational aim of flood frequency analysis is to estimate design floods that are used for areal planning and the design of important infrastructure. In most operational guidelines, annual maximum data are fitted to one or several distributions. Typical distribution that are used include the Generalized extreme value-, Gumbel-, Log_pearson-III-, and Generalized logistic distributions. Typical estimation methods include Bayesian, ordinary moments, l-moment and maximum likelihood approaches.

In an operational setting, the assessment of how well the model fit to data, based on visualizing the fitted distribution together with the empirical distribution of the data. If no specific requirements are given, this visualization is used to select the distributions providing the best fit to data.

Several elements of a return level plot are influenced by uncertainties. Firstly, the fitted distribution is uncertain since a limited dataset is used in the estimation. Secondly the observed flood data are uncertain since they are derived from a rating curve. Thirdly, the plotting position used for estimating the empirical distribution is based on order statistics and is also uncertain due to the limited sample size.

The aim of this presentation is to investigate the subjectivity in the choice of extreme value distribution based on visual inspections of return level plots. In particular, we aimed to assess if visualisation of uncertainties in plotting positions would affect the subjective choices. Based on a survey with 40 participants, we wanted to answer the following questions:

- 1: Does the plotting-position influence model selection?
- 2: Does the information about uncertainty in plotting position change the strategy for fitting distributions.
- 3: Which numerical criterion reflects best the subjective selection of distribution.?
- 4: Can we identify the true underlying distribution based on a sample of 100 or less members?
- 5: Can we identify different strategies people use for fitting distributions?

The results show that plotting position and information about uncertainty influence the model selection. We found also that combining the choices from all the 40 participants, more frequently resulted in identification of the true distribution. We could also identify different strategies used for fitting distributions.