



Model error correlation in a regression model of the median annual maximum flood

T. R. Kjeldsen and D. A. Jones

Centre for Ecology & Hydrology, Maclean Building, Crowmarsh Gifford, Wallingford, OX10 8BB, UK (trkj@ceh.ac.uk / daj@ceh.ac.uk)

Use of regression models linking flood statistics to a set of catchment descriptors is a long-established practice in applied hydrology. On recognising that the total regression error is a combination of both sampling and model error, a generalised least square (GLS) technique for estimating the regression model has previously been used based on the added assumption that the model error components (in contrast to the sampling error components) are uncorrelated. In a study involving the median annual maximum flood from 602 rural catchments it is shown that i) model error correlation is present when pairs of catchments are geographically close and ii) the degree of model error correlation is strongly dependent on the particular combination of catchment descriptors used as covariates in the regression model. For the first part of the study, an exploratory data analysis was conducted based on a set of re-weighted regression residuals whose theoretical covariance structure should be essentially similar to the model error covariance. By grouping pairs of catchments in bins defined by geographical distance between catchment centroids, it was found that the correlation could be modelled in relation to geographical distance using a generic function based on a weighted sum of two simple exponential functions. Having identified a generic description of model error correlation, estimation of the overall model, including both this correlation and the regression components, was accomplished using the maximum-likelihood method. A sensitivity analysis showed that while some of the model error correlation could be removed by adding more catchment descriptors to the model, some degree of correlation remained despite having identified a reasonable set of descriptors.