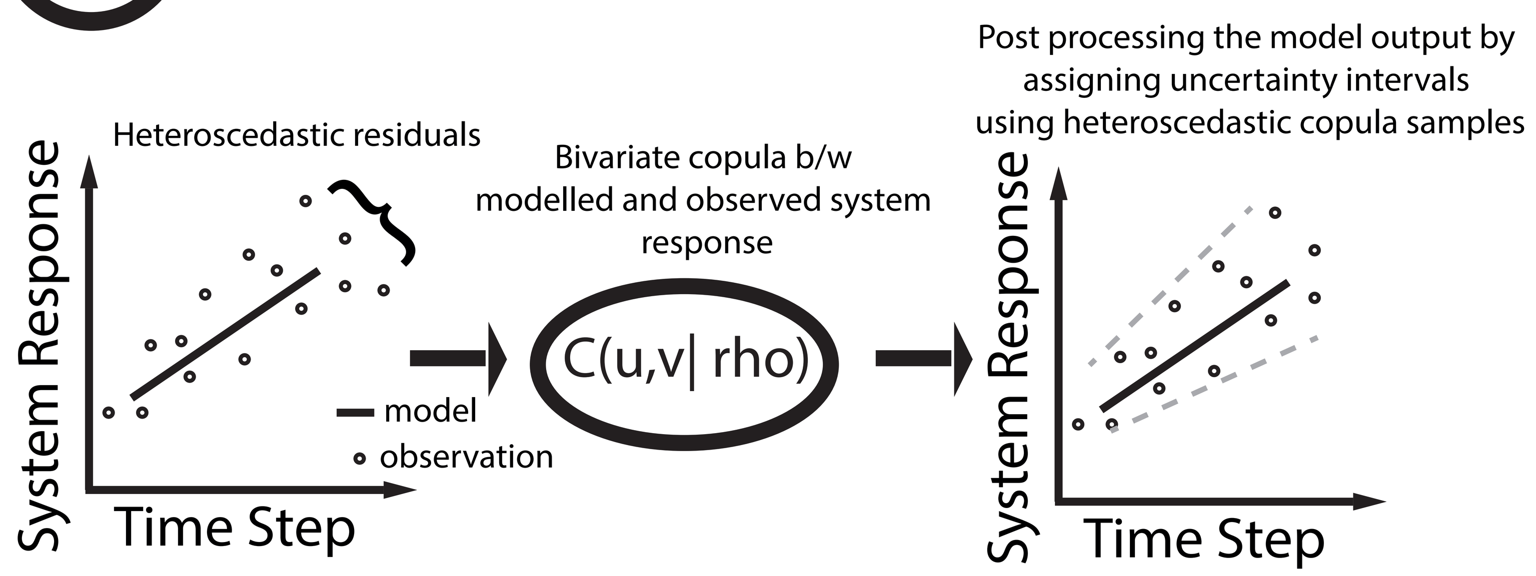
Gaussian copula as a post processor for environmental models

Omar Wani,
Gabriel Espadas, Francesca Cecinati, Jörg Rieckermann

1 Graphical abstract



2 Problem statement

a) Can residual errors be faithfully represented by a semi-parametric distribution?

b) Can heteroscedasticty be captured without using predefined tranformations?

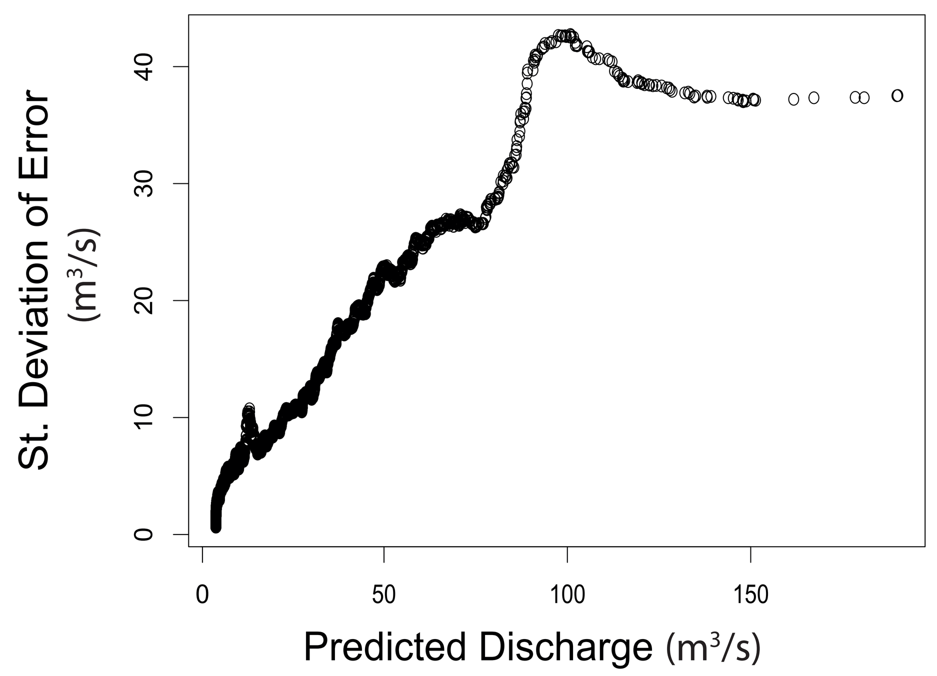


Fig 1. A representative example of heteroscedasticity in rainfall-runoff models

3 Bivariate Gaussian copula

An approximated probability density function of two correlated variables can be written using a bivariate Gaussian copula.

Parameter rho captures the linear correlation between the two variables. Along with the respective marginal cumulative distributions of the two variables, their joint distribution is uniquely defined. Here we formulate a Gaussian copula between modelled discharge (Q_m) and observed discharge (Q_o):

$$p(Q_m, Q_o|rho) = c(u, v|rho)p(Q_m)p(Q_o)$$

where c is the copula density, (u,v) are the probability integral transformed (Q_m,Q_o), using the emperical cumulative distribution function. After this description, p(Q_o|Q_m,rho), formally a likelihood function, is sampled to generate uncertainty intervals. We use an adaptive MCMC algorithm for sample generation.

4 River Brue as a case study

Area: 135 sq. km
Avg. rainfall: 867 mm/year
Model: HBV for rainfall-runoff
Calibration: Jun 94 - May 95
Validation: Jun 95 - May 96

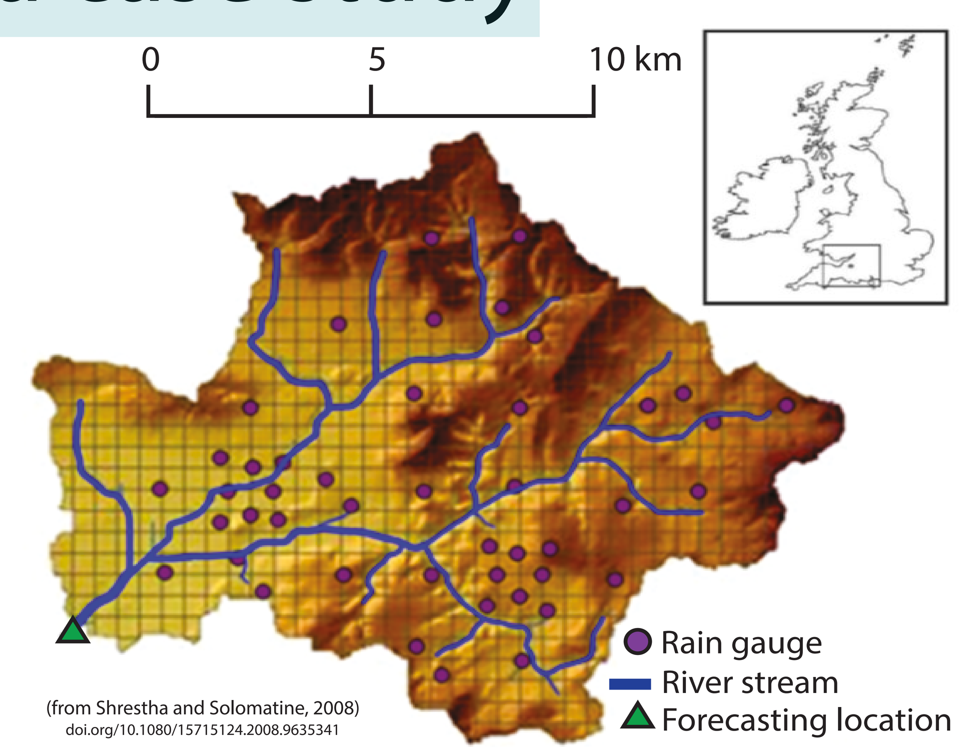
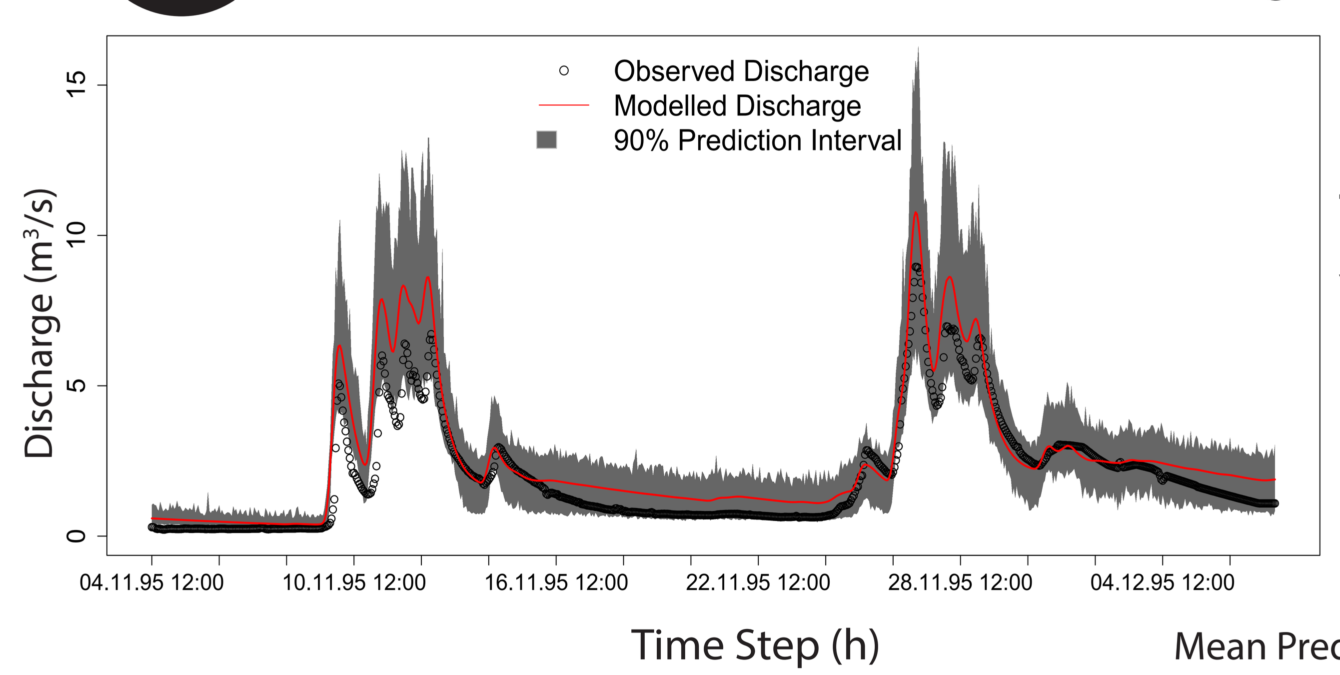


Fig 2. Brue catchment, UK.

5 Results



Mean Prediction Interval: 1.56 m^3/s
Interval Coverage: 74.85%

Fig 3. Hydrograph for two events in the validation timeseries.

The prediction interval for a year of model simulation shows the required property of heterscedastcity. The hydrograph for two events is plotted in figure 3 and the scatter plot for the entire year in figure 4 and 5.

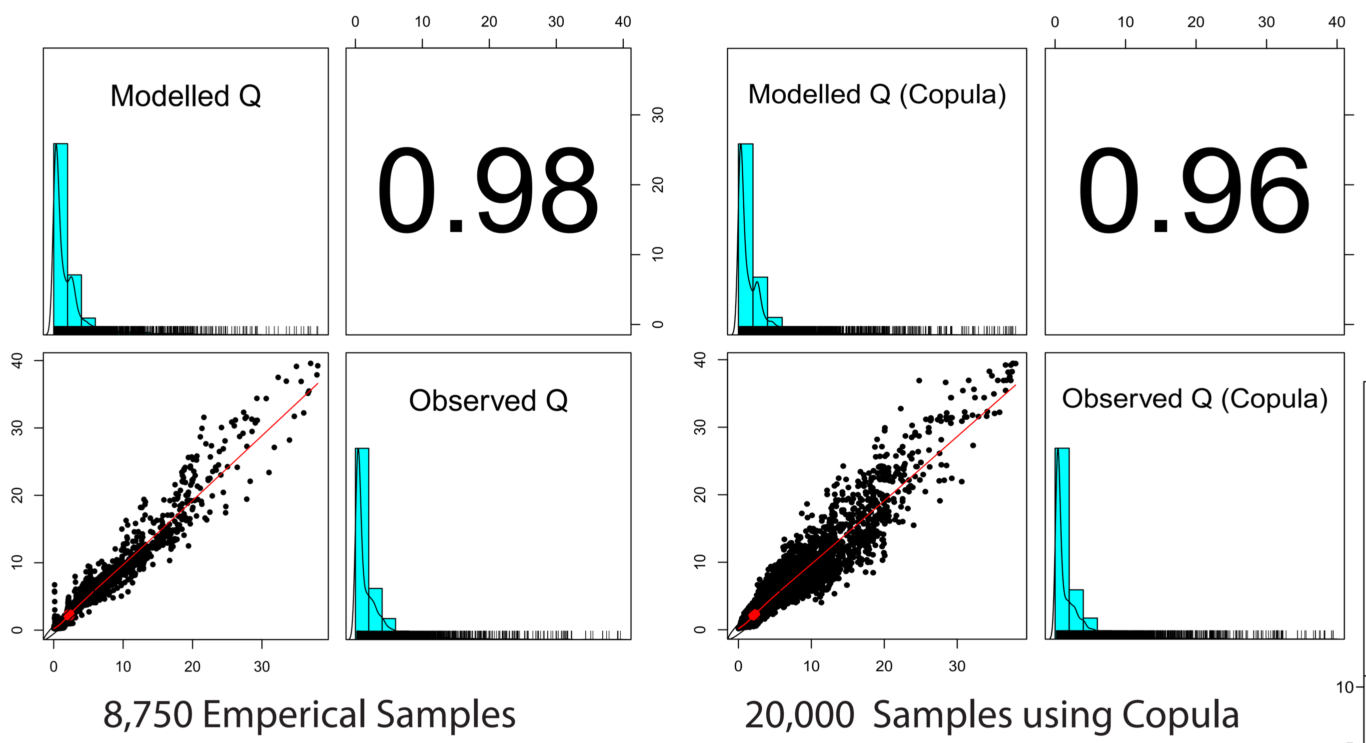


Fig 4. (left) Plot showing the scatter and correlation between observed and modelled discharge.

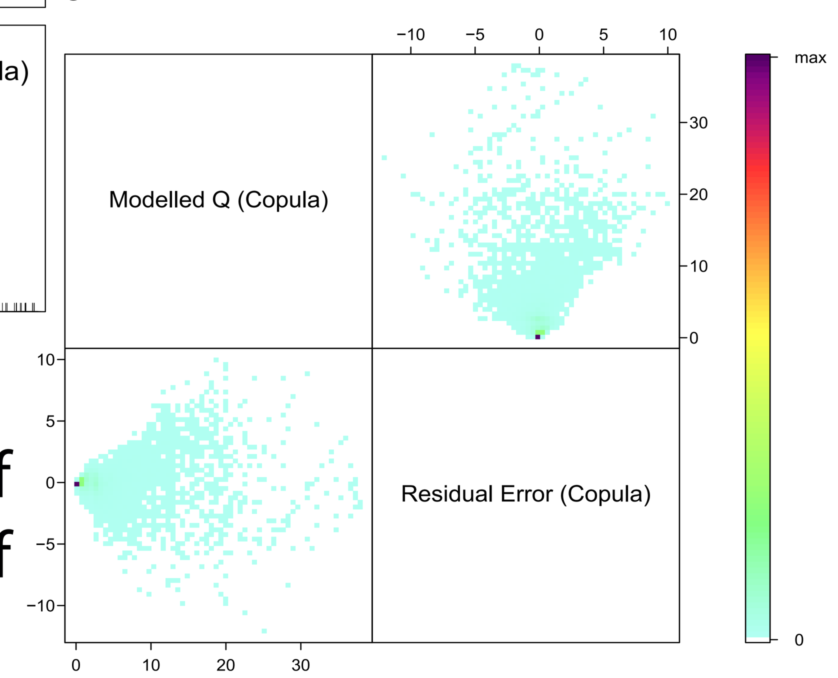


Fig 5. (right) Plot showing the dependence of residual error variance on the magnitude of modelled discharge.

6 Conclusions

a) Gaussian copula was able to generate heteroscedastic uncertainty intervals for the case study - without using transformations explicitly for this purpose.

b) Samples can be generated to assign uncertainty intervals corresponding to any probabilty (e.g. 0.75, 0.90, 0.99 etc), unlike many other post processor techniques like kNN resampling and quantile regression.

7 Ongoing research

a) Identifying appropriate copula families for heteroscedastcity.
b) Analyzing the limiting cases where such copula usage performs poorly.
c) Model parameter inference using copula as error descriptors.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

eawag aquatic research

University of BRISTOL

QUICS

This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no 607000.

Marie Skłodowska-Curie Actions Fellowship