Accounting for correlation in uncertainty propagation, a copula approach for water quality modelling.

Moreno-Ródenas, A.M[1, *], Langeveld, J.[1, 3] and Clemens, F[1, 4]

[1]Technische Universiteit Delft, Water Management Department, The Netherlands
(*Email: *a.m.morenorodenas@tudelft.nl)*
[3]Partners4urbanwater, Nijmegen, The Netherlands
[4]Deltares, Delft, The Netherlands

Keywords
Combined sewer overflow; Copulas; Integrated Catchment Modelling; Multivariate Non-Gaussian distributions; Uncertainty propagation

Modelling pollution loads is critical for assessing ecological impacts of sewer discharges. Sewer water quality simulations are subjected to significant uncertainties. Measured pollutant distributions are often non-Gaussian and present strong correlation. Those are characteristics seldom represented in uncertainty propagation (UP) sampling. This work investigates the effect of accounting for parameter correlation during UP in an integrated catchment model for dissolved oxygen simulation. This is done by comparing the use of a copula distribution (which respects marginal distributions and correlation of parameters) with random sampling of pollutant concentrations in sewer discharges.

1 INTRODUCTION

Pollutants concentration at combined sewer overflows (CSO) exhibit dynamics that are difficult to simulate. It is acknowledged that deterministic results of sewer water quality models are generally poor (Willems, 2006). For this reason, "mean value" pollution vectors are often used as input when modelling sewer overflow impacts in scarcely monitored environments. This simplification can render uncertain model outputs. Although, its effect can still be of the same order of magnitude using uncalibrated sewer water quality models. Assessing the uncertainty propagation (UP) of those assumptions is necessary when checking the reliability of the simulator.

Forward UP schemes rely on sampling parameters from probability ranges in order to evaluate its effect on the model output. Parameter probability distributions can be inferred, elicited from expert beliefs (O'Hagan, 1998) or when possible, derived from measured data.

In this study, a water quality model uses mean concentration values as a model input. CSO concentrations are seldom normally distributed, and present strong correlation. Sampling a non-Gaussian multivariate distribution is not a straightforward process. This paper proposes the use of a copula distribution as a means to generate pollutant concentration samples, which reproduce observed distributions and correlation. We check the effect of accounting for or neglecting parameter correlation in an UP scheme for an integrated catchment model (ICM) predicting dissolved oxygen (DO).

2 METHODS & MATERIALS

Model structure

This research is conducted in a conceptual ICM at the river Dommel, The Netherlands. This is a small river which receives the discharge of ~200 CSOs and of a WWTP of 750,000 p.e. The conceptual model is composed by ~110 km of river, a detailed WWTP and 28 urban catchments. In order to further separate its effect, the effluent of the WWTP and flow from the CSOs have been substituted by monitoring data.

Describing pollutants.

Distributions for each pollutant concentration were obtained from a monitoring campaign at several CSO structures (Moens et al, 2009). The parametric probability density functions (pdf) fitted to the measured data can be appreciated in Fig. 1.

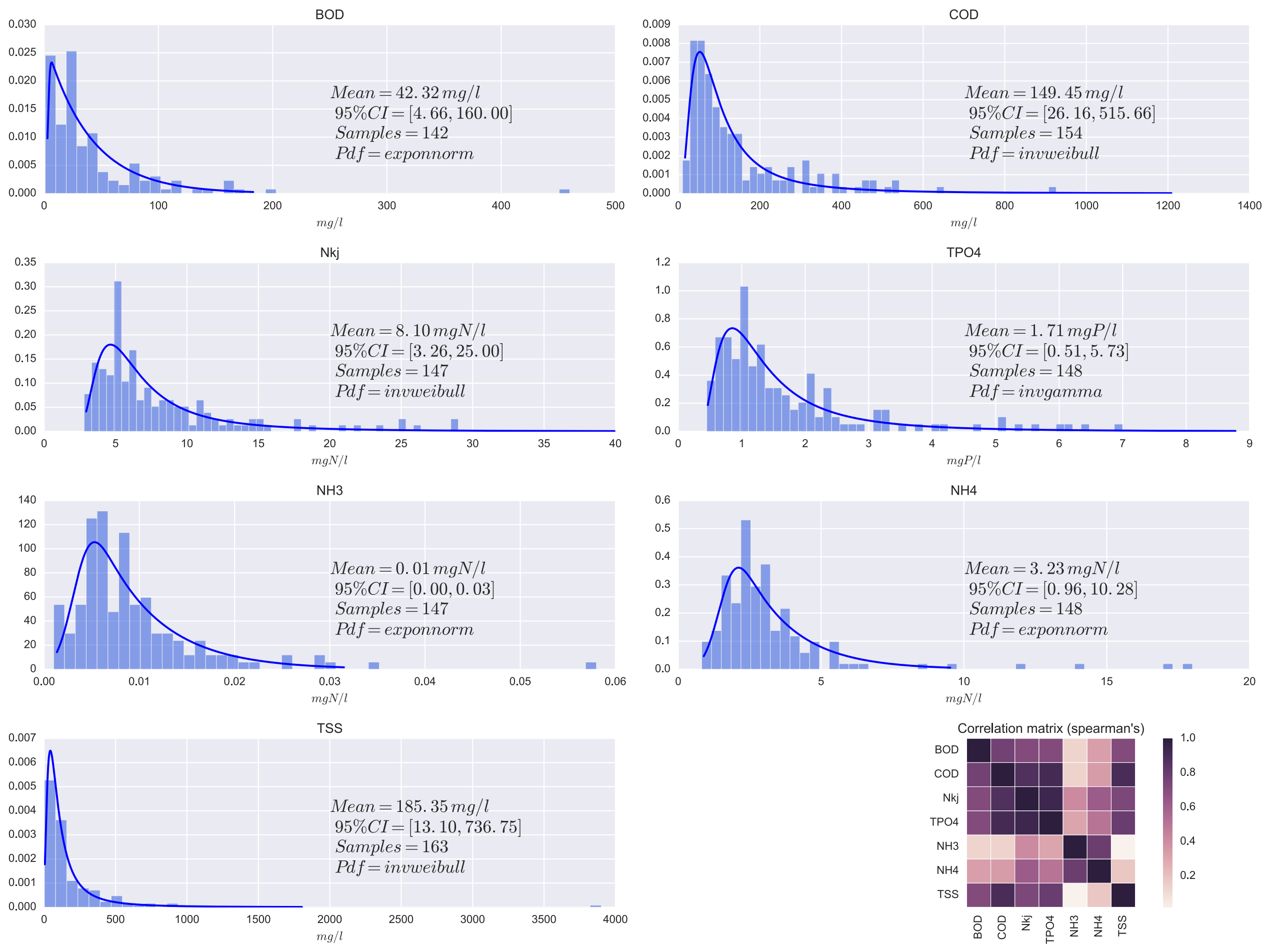

Fig 1 Histograms and correlation matrix of measured data, mean and 95% confidence intervals. Blue solid line; fitted pdf.

The copula approach

A detailed explanation on generating correlated multivariate samples can be found at Schoelzel (2008). In this study an elliptical Gaussian copula (GC) was used as a way to generate a joint distribution, which reproduces pollutant structural correlation. In order to construct the GC, let's $\boldsymbol{X}$ be a random vector, which components $X_1, \ldots X_n$ represent the set of pollutant concentrations with known marginal cumulative density functions (cdf) $F_{X1}, \ldots F_{Xn}$. Applying $F_{xi}(X_i) = U_i \sim U(0,1)$ which can be transformed by $F^{-1}_{N(0,1)}(U_i) = Z_i$. Thus the vector $\boldsymbol{Z}$ forms a Normal multivariate distribution $\sim N_k(0, \Sigma)$. Which density describes implicitly a GC with a correlation structure given by the covariance matrix Σ. In order to generate samples from the joint pdf, it is possible to do the inverse process, starting from a random realization of the copula density, and back transforming to the desired marginals.

A comparison of the rank (spearman's) correlation matrix for the measured data, along with 500 samples from the proposed GC and random sampling of the marginal pdf's is shown in Fig 2.

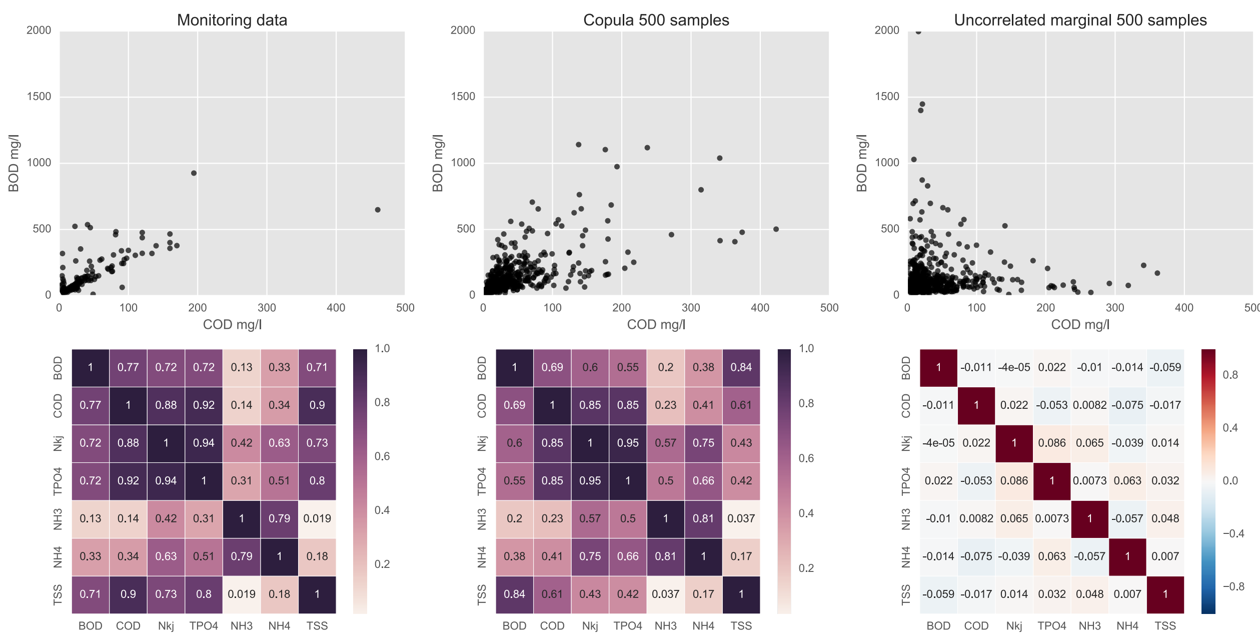

Fig 2 Comparison of measured data, GC and random sampling from marginals. Above, scatter plot for BOD-COD. Below, correlation matrix.

2.3 Uncertainty propagation

The model was evaluated using 500 samples from the GC distribution and from the uncorrelated marginals. The variables used were concentrations of COD, BOD and NH4. Predicted DO concentrations in a river section were compared with monitoring data for a period of 8 months (01-04-2012 / 31-09-2012).

3 RESULTS

Fig 3 depicts the comparison between the observed and simulated DO. In Fig 4 a comparison of the minimum DO depletion level for 8 selected events can be found. It can be noted that random sampling renders narrower bands compared with the correlated sampling.

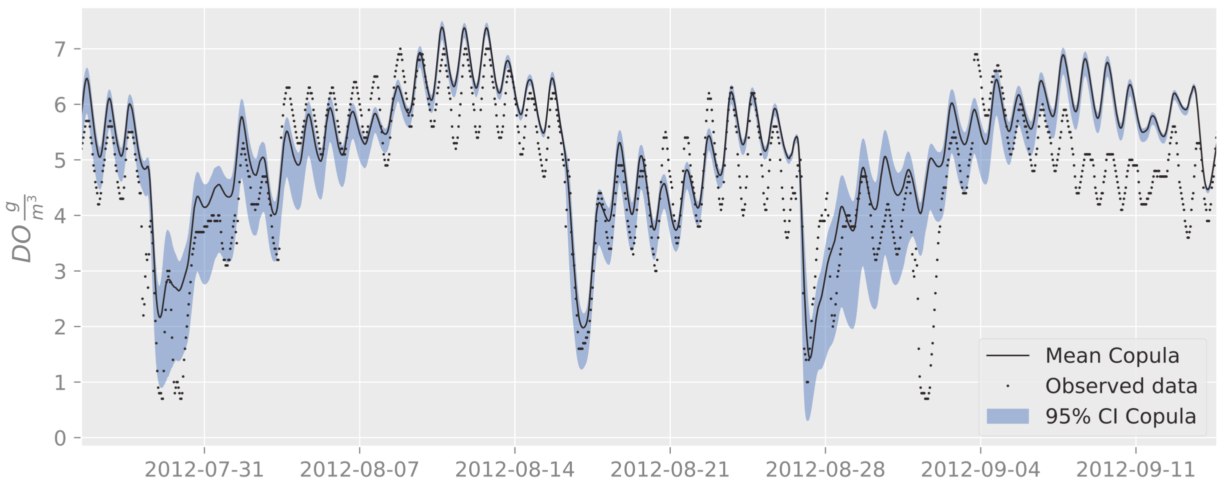

Fig 3 DO at the river (~17 km downstream of the WWTP). Observations vs simulated (GC).

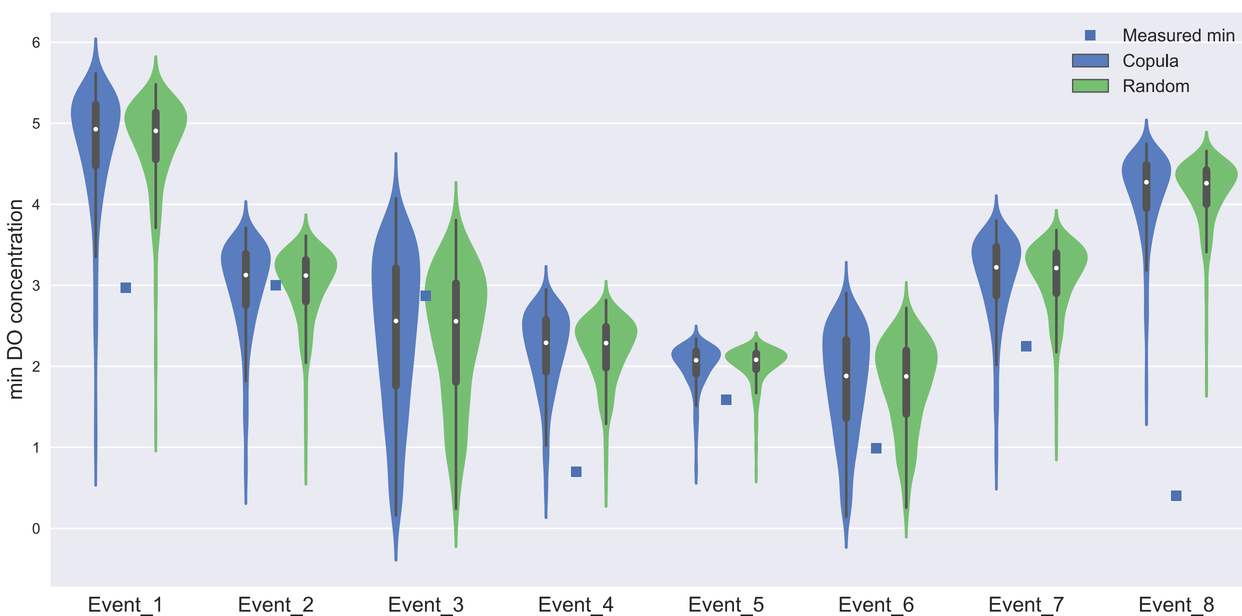

Fig 4 Minimum DO level for every event in which measurements are below 3 mg/l.

4 CONCLUSIONS

Uncertainties in CSO pollution load can affect significantly DO depletion patterns. The pollution load affects especially the intensity and the recovery of the DO-drop, this is relevant when characterizing river impacts through intensity-duration-frequency curves.

Not accounting for structural pollutant correlation in the UP scheme produced an underestimation of DO uncertainty bands. However, it has to be noted that the use of correlation didn't lead to a systematic improvement of observed values coverage. Nevertheless, when knowledge on the correlation structure is available it is recommended to include it in the UP process.

5 ACKNOWLEDGEMENTS
This research was done as part of the Marie Curie ITN – Quantifying Uncertainty in Integrated Catchment Studies (QUICS) project. This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 607000.

6 REFERENCES

Moens, M., Vroege, M. and Grendelman, R. 2009. Monitoring program for the River Dommel through Eindhoven, final report phase 1. Report 110502/ ZF9/2D3/300031. the Netherlands: Arcadis Nederland bv.

O'Hagan, A. 1998. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *47*(1), 21-35.

Schoelzel, C., Friederichs, P. 2008. Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlin. Processes Geophys.* 15 (5), pp.761-772.

Willems, P. 2008. Quantification and relative comparison of different types of uncertainties in sewer water quality modeling. *Water Res, 42*(13), 3539-3551. doi: 10.1016/j.watres.2008.05.006