



## **Data-mining for multi-variate flood damage modelling with limited data**

Dennis Wagenaar and Laurens Bouwer

Deltares, Delft, Netherlands (laurens.bouwer@deltares.nl)

Flood damage assessment is usually done with damage curves only dependent on the water depth. Recent studies have shown that data-mining techniques applied to a multi-dimensional dataset can produce significantly better flood damage estimates. However, creating and applying a multi-variate flood damage model requires an extensive dataset, which is rarely available and this can limit the application of these new techniques. In this paper we enrich a dataset of residential building and content damages from the Meuse flood of 1993 in the Netherlands, to make it suitable for multi-variate flood damage assessment. Results from 2D flood simulations are used to add information on flow velocity, flood duration and the return period to the dataset, and cadastre data is used to add information on building characteristics. Next, several statistical approaches are used to create multi-variate flood damage models, including regression trees, bagging regression trees, random forest, and a Bayesian network. Validation on data points from a test set shows that the enriched dataset in combination with the data-mining techniques delivers a significant improvement over a simple model only based on the water depth. We find that with our dataset, the trees based methods perform better than the Bayesian Network, which is in contrast to other studies.