Universität Stuttgart

WAREM
Water Resources Engineering and Management

akkreditiert durch ZevA
im Auftrag des Akkreditierungsrats (KMK / HRK)

Universität Stuttgart WAREM, Pfaffenwaldring 7, 70569 Stuttgart

Pfaffenwaldring 7
70569 Stuttgart
Telefon: (0711) 685 - 66615 / 66616
Telefax: (0711) 685 – 66600
warem@iws.uni-stuttgart.de
http://www.warem.uni-stuttgart.de/

Anne Weiss M.A., M.Sc.
(Durchwahl: - 66616)

Master Thesis Abstract

of Stella Nadine Steidl

Comparison of Multiple Linear Regression and Artificial Neural Networks for Infilling Missing Precipitation Records

Despite today's near-infinite data storage capacity and a plethora of information only a click away, collecting and pre-processing hydrological data can still remain critical and tedious tasks, due to, especially present in the developing world, sparse rain gauge networks and incomplete time series data. Rainfall data, as a crucial input to design, measures and models of various fields such as hydrology, urban hydrology or the management of water resources, governs important decisions derived from the analysis of a model's outcome. Therefore, missing data, ranging from a single day to several months can have a deleterious impact and gaps need to be filled prior to any analysis undertaken.

This study investigates and compares two different methods for infilling missing daily precipitation records in the area along the Neckar River in Baden-Württemberg, Germany. For a rather traditional statistical approach a Multiple Linear Regression (MLR) was chosen and its performance compared with that of a recent computational model using an Artificial Neural Network (ANN) from the field of machine learning.

The results indicated, that both models provided a good estimation of the missing precipitation records, when used to fill in daily precipitation as well as cumulative precipitation sums of two, three and five days. Given different input variants, for some instances, both models derived slightly negative precipitation estimates for periods where there has been no precipitation observed at neighbouring stations, that were used to fill in the missing data. Furthermore, the choice of neighbouring stations considered to fill in data at a given point, was found to have a significant influence on the accuracy and reliability of the two models. Overall, the ANN performed well on the given infilling tasks once a suitable architecture was found for the model. However, due to the resource-intensive nature of neural networks and their accuracy being highly dependent on the set-up of the network's architecture, further research is required for this technique to gain acceptance among hydrologists.