Rhine Level Data

Paul Jonas Jost, Annabell Gros

July 16, 2019

Table of contents

The Setting

The Rhine Level Data gives us the water level of the river Rhine in the following way:

- for 15 stations
- in 15 min-time intervals
- over the course of 32 years (1981-2013)

Goal: Prediction of the water level in Bonn 12h into the future.

The Preprocessing

Basic Idea: The water obviously travels, so water which is at a station X will be in Bonn after time $\frac{dist_{XB}}{v_{Water}}$. So we shifted our data in the following way:

- Shift Bonn by -12h
- create new columns with the data of stations X shifted by $S = \frac{dist_{XB}}{v_{Water}} - 12\text{h}$, ignoring all stations where $S < 0$ (we would use data from the future)
- fill up gaps of missing data of a maximal length of 24h
- drop all times where the data from Bonn are missing (i.e. for each remaining time we can compare our prediction with the true water level)

Selection of the Prediction Data

1. pick a time T to get the prediction at time $T+12$h
2. throw out all stations with missing data at time T
3. throw out all incomplete data sets of the remaining stations
4. split training and test data by the standard rate of 0.25

Selection of the Prediction Data II

1. Problem I: T is part of the training data for our model (Problem: Correlation with training data, unrealistic time travel scenario)
2. Problem II: train model on data up to time T (Problem: still correlation with training data)
3. Our approach: train model on data up to time T, consider predictions on a lengthy time period after time T

The Model

Simple neural network:

1. one dense layer, one flatten layer , one output layer
2. optimizer: Nadam, loss: mean squared error

Training data sets

We used four slightly different data sets to train the neural network:

- we used all stations and all shifted data
- a we dropped the stations too near or too far away
- b we used only the shifted data
- c we used no shifted data at all

Normal prediction

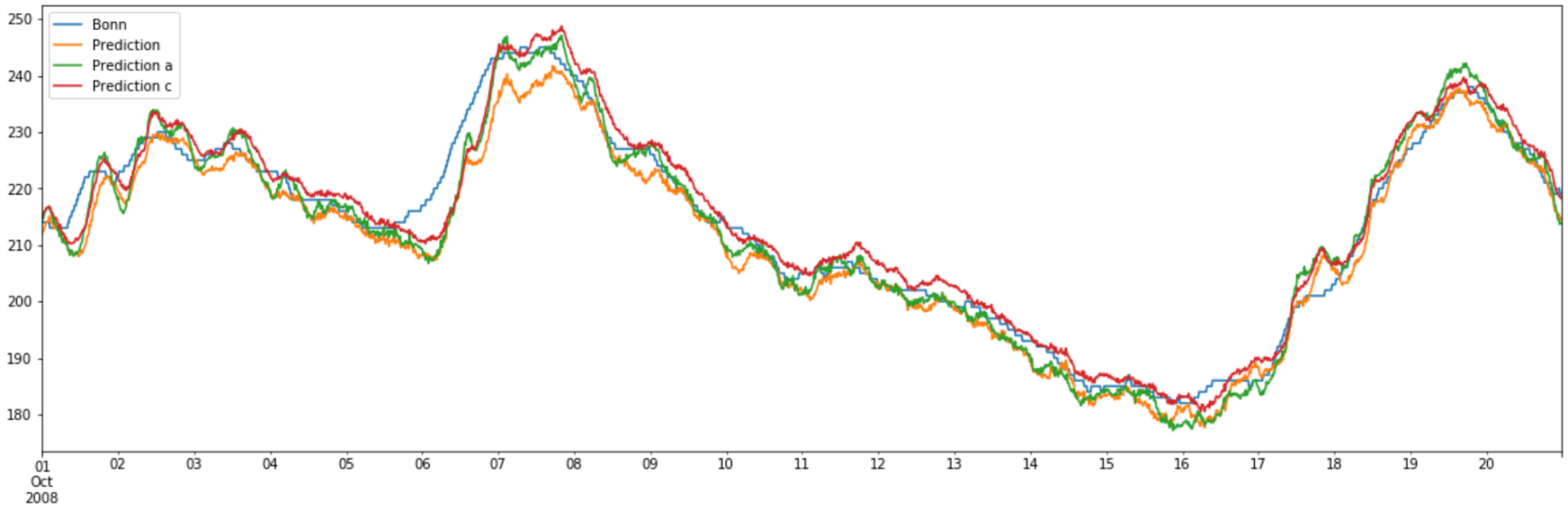

Figure: October 2008, model trained on data before 2000

Flood prediction

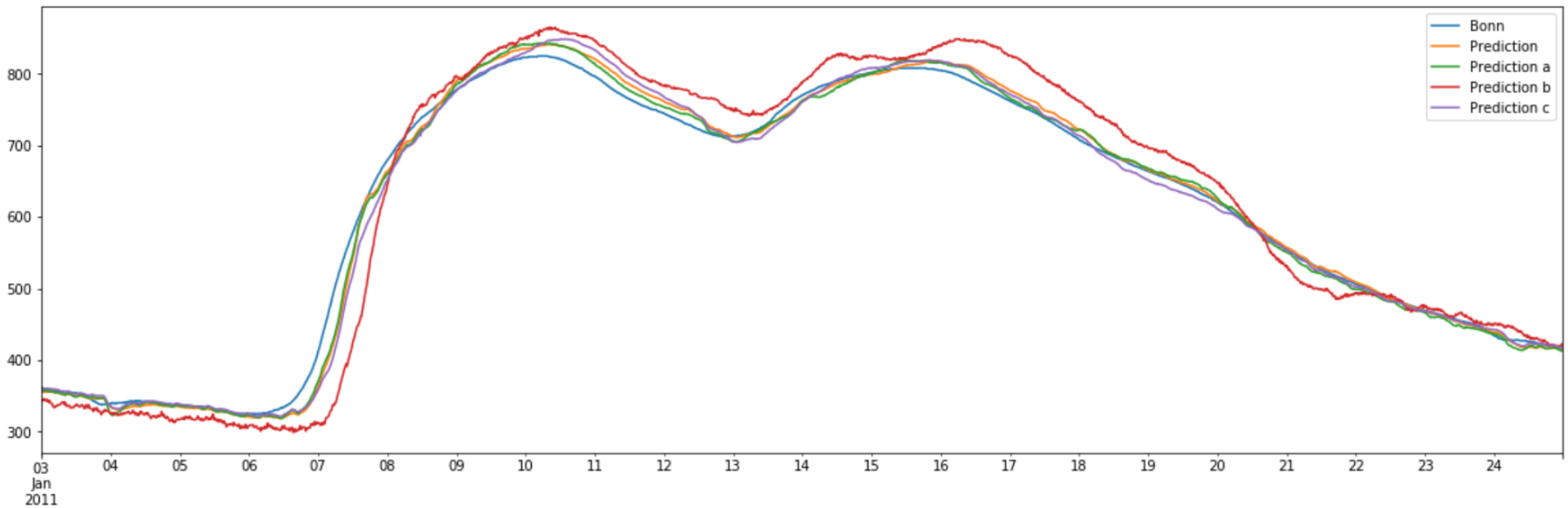

Figure: January 2011, model trained on data before 2000

Results

1. MSE of ca. 20 cm at the model fit (for most data sets), good predictions for normal data
2. time delay and too high values in flood prediction
3. worst predictions with data set b (only shifted data, MSE 200 cm)