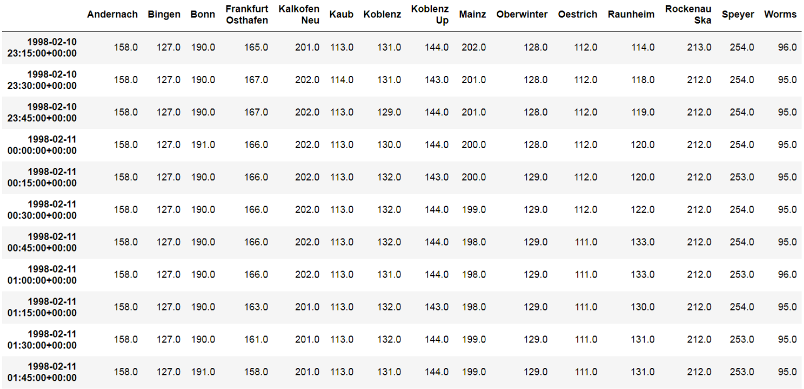
Rhine Level Prediction

Vaggelis Lamprou, Rico Krause

18.07.2019

The Data Set



	Andernach	Bingen	Bonn	Frankfurt Osthafen	Kalkofen Neu	Kaub	Koblenz	Koblenz Up	Mainz	Oberwinter	Oestrich	Raunheim	Rockenau Ska	Speyer	Worms
1998-02-10 23:15:00+00:00	158.0	127.0	190.0	165.0	201.0	113.0	131.0	144.0	202.0	128.0	112.0	114.0	213.0	254.0	96.0
1998-02-10 23:30:00+00:00	158.0	127.0	190.0	167.0	202.0	114.0	131.0	143.0	201.0	128.0	112.0	118.0	212.0	254.0	95.0
1998-02-10 23:45:00+00:00	158.0	127.0	190.0	167.0	202.0	113.0	129.0	144.0	201.0	128.0	112.0	119.0	212.0	254.0	95.0
1998-02-11 00:00:00+00:00	158.0	127.0	191.0	166.0	202.0	113.0	130.0	144.0	200.0	128.0	112.0	120.0	212.0	254.0	95.0
1998-02-11 00:15:00+00:00	158.0	127.0	190.0	166.0	202.0	113.0	132.0	143.0	200.0	129.0	112.0	120.0	212.0	253.0	95.0
1998-02-11 00:30:00+00:00	158.0	127.0	190.0	166.0	202.0	113.0	132.0	144.0	199.0	129.0	112.0	122.0	212.0	254.0	95.0
1998-02-11 00:45:00+00:00	158.0	127.0	190.0	166.0	202.0	113.0	132.0	144.0	198.0	129.0	111.0	133.0	212.0	254.0	95.0
1998-02-11 01:00:00+00:00	158.0	127.0	190.0	166.0	202.0	113.0	131.0	144.0	198.0	129.0	111.0	133.0	212.0	253.0	96.0
1998-02-11 01:15:00+00:00	158.0	127.0	190.0	163.0	201.0	113.0	132.0	143.0	198.0	129.0	111.0	130.0	212.0	254.0	95.0
1998-02-11 01:30:00+00:00	158.0	127.0	190.0	161.0	201.0	113.0	132.0	144.0	199.0	129.0	111.0	131.0	212.0	253.0	95.0
1998-02-11 01:45:00+00:00	158.0	127.0	191.0	158.0	201.0	113.0	131.0	144.0	199.0	129.0	111.0	131.0	212.0	253.0	95.0

- Task: Predict the Rhine Level 12 hours in the future.

We consider two different approaches:

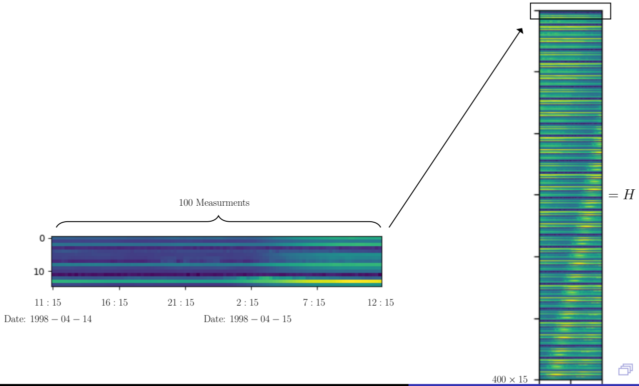
1. Exploiting the fact that the data is generated by an underlying unknown Dynamic System.
2. Extrapolating from the pure time series data.

Delay embedding

- A way to linearize a dynamic system is to consider its Delay embedding.

Definition

The m dimensional-delay embedding of a function $f : \mathbb{R} \to \mathbb{R}^n$ is the function $g : \mathbb{R} \to \mathbb{R}^{n \times m}$
$x \mapsto (f(x), f(x+h), f(x+2h), \dots, f(x+mh))$



Delay embedding

We test the linearity of the system H as follows.

- Consider its Singular value decomposition (svd) $H = USV^*$. V^* contains its Principal components (Dynamic Modes).
- Find the dominating singular values s_i.
- Test if the corresponding modes form a Fourier basis.

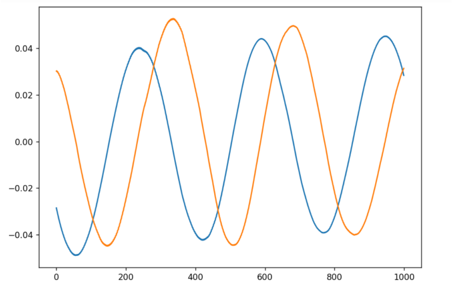


Figure 1: First two dominant modes of H

Dynamic mode decomposition

- Since H is now linear we can solve for a continuous well defined solution Φ over whole $\mathbb{R}$.
- We use the Dynamic mode decomposition (DMD) algorithm to solve it efficiently.
- DMD restricts the calculation to the few dominating modes of the system, thereby drastically reducing the dimensionality of the problem.
- Now evaluate Φ at the desired time to make prediction for the future.

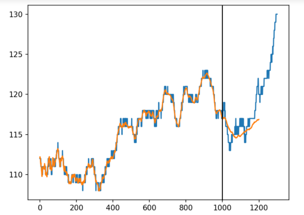


Figure 2: 20 Modes, Error = 0.38246460680858263

Dynamic mode decomposition

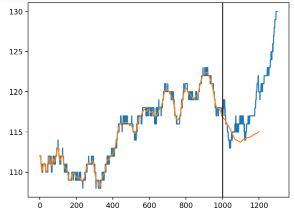


Figure 3: 200 Modes, Error = 1.0426185310642495

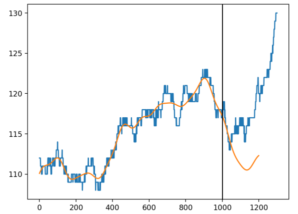


Figure 4: 800 Modes, Error = 1.2696253178717711

Predicting by sequential data

- For predicting sequential data Recurrent Neural Networks are widely used.

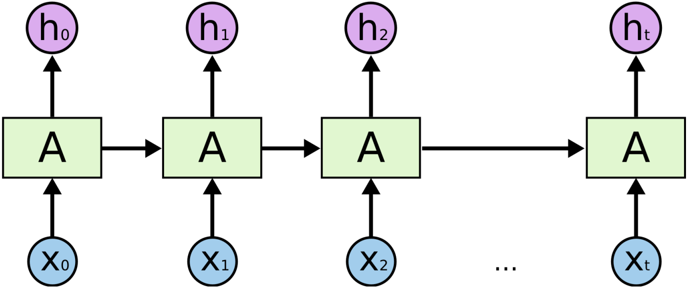


Figure 5: RNN

LSTM

- Regular RNN can lack an understanding of long time dependency.
- Long short-term memory Networks (LSTM's) solve this problem by introducing "Gates" that remember the useful information and separate it from the rest.

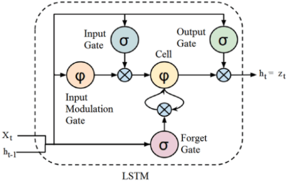


Figure 6: One LSTM Cell

Our Model

We try to predict the Future Rhine Level 12 hours ahead

- We used one looped LSTM Layer with 400 units
- One fully connected linear activated layer to gather the data into one dimension.
- As Training Data we use a stack of randomly starting sequences of length 1000 from the level data of Bonn. And as target date the Rhine Level 12 hour past that sequence.
- The model was fit using the ADAM optimizer with the mean squared error as the loss function.

Testing

Lets see how it preforms on 10 randomly starting time sequences of length 1000

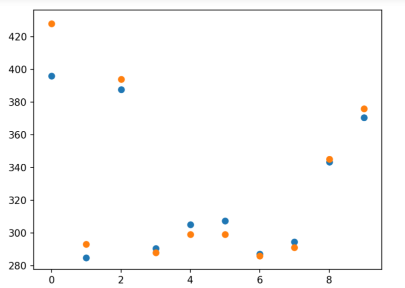


Figure 7: Yellow=real data, Blue=prediction

Mean Distance: 7.47