Predicting the temperature

A study of time series analysis

By Victoria Lagerstedt & Kajsa Hansson Willis

Table of contents

01

Problem formulation and introduction

Goal: Predict the temperature in Tåstrup in Denmark

Part A

Predicting the temperature using only previous measurements of temperature

Part B

Predicting the temperature using previous measurements of temperature and other historical data as input

Part C

Predicting the temperature recursively based on the model in part B

02

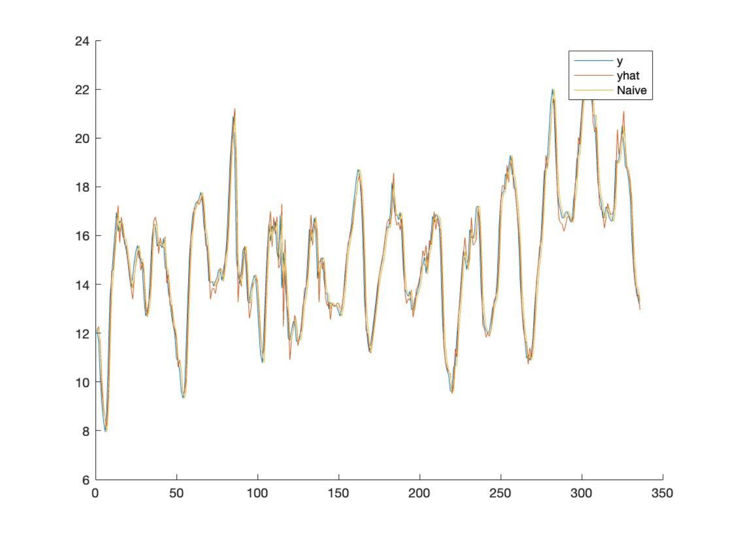
Part A

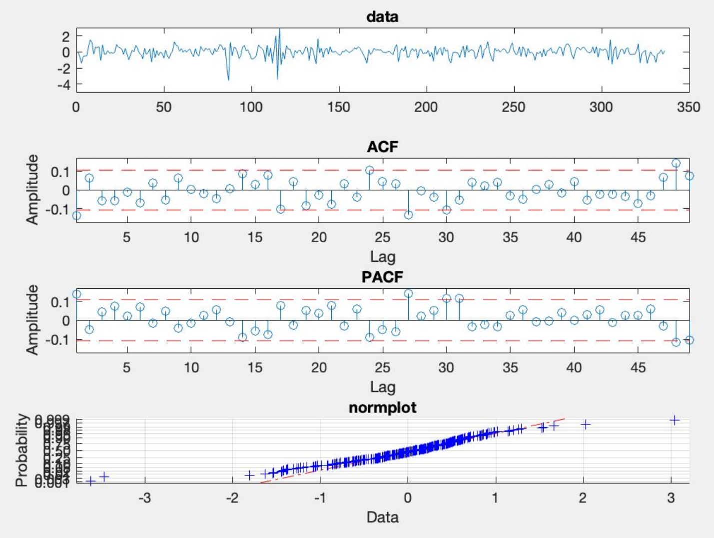
The first predictor

- Found evidence of trend and season but decided to try a simpler model first and see if that would work
- ARMA model was created by looking at the data ACF and PACF and trial and error
    - Parameters were added and insignificant ones were removed
- When the model that worked for the modeling data it was tested on the validation data

The validation data looked like…

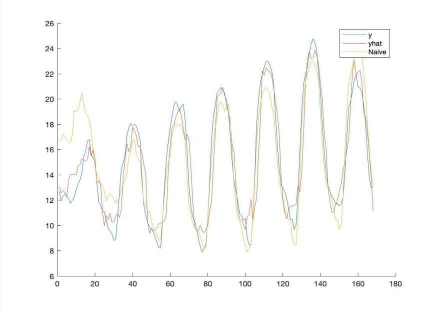
... This (for lag one)

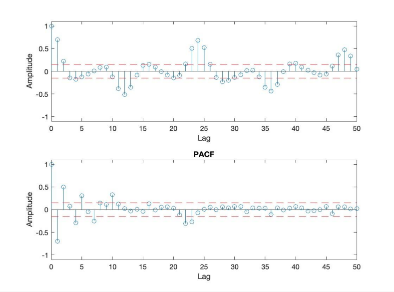




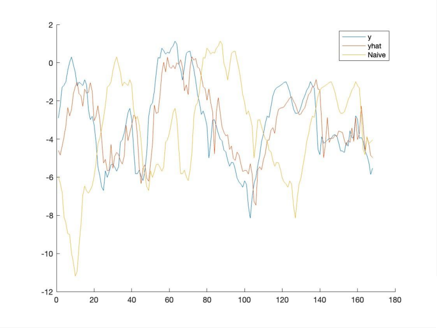
This was very promising so the test data was tested as well, this time lag 3 is showed

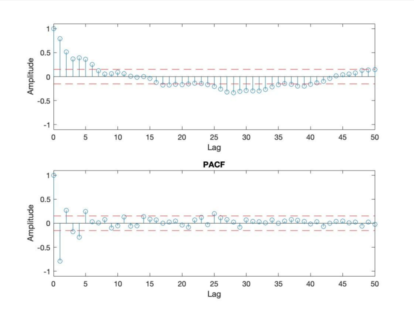
First test data





Second test data





The ARMA(5,8) model looked good!

A(z) = 1 - 2.953 (± 0.08302) z-1 + 2.631 (± 0.1603)z-2 - 1.041 (± 0.1601) z-4+
+ 0.3623 (± 0.08282) z-5

C(z)= 1 - 1.556 (± 0.08968) z-1 - 0.04917 (± 0.08146) z-2 + 0.8044 (± 0.1203) z-3
- 0.1043 (±0.08919) z-4- 0.01965 (±0.05811) z-7 - 0.0466 (±0.04175) z-8

where the process is described by $A(z)y_t = C(z)e_t$

03

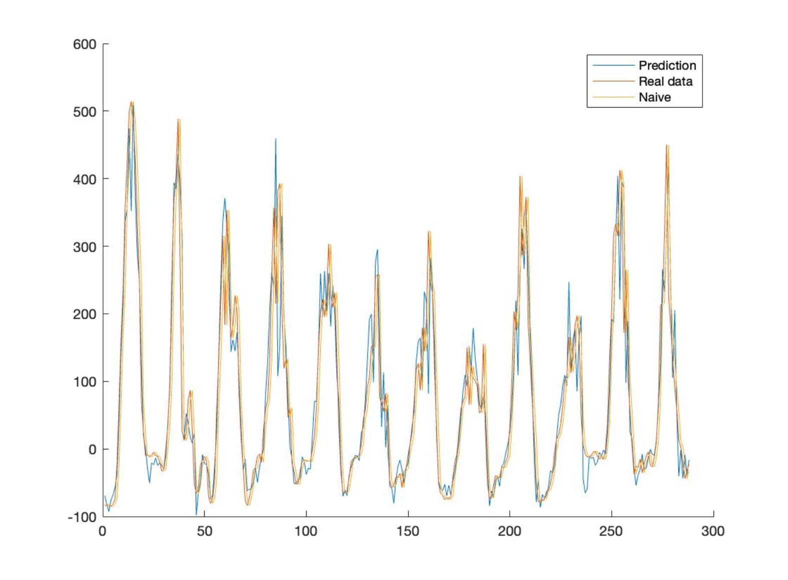
Part B

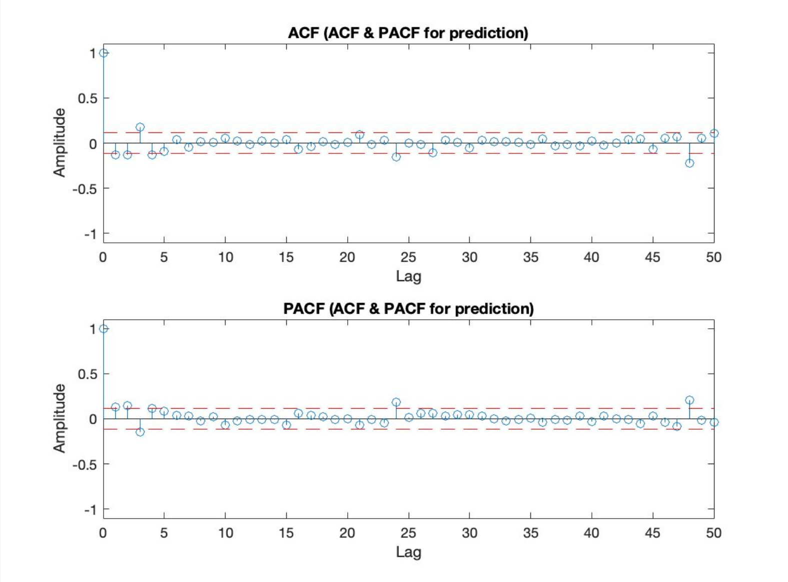
The second predictor

- The net radiation data was added as an input
- An input model was estimated first
  - This model was estimated on differentiated data
  - Estimated through estimateARMA and perfected until the parameters were significant and results good
  - Tested to ensure it worked well on different data sets
- The output model was then estimated
  - Box-Jenkins methodology was applied

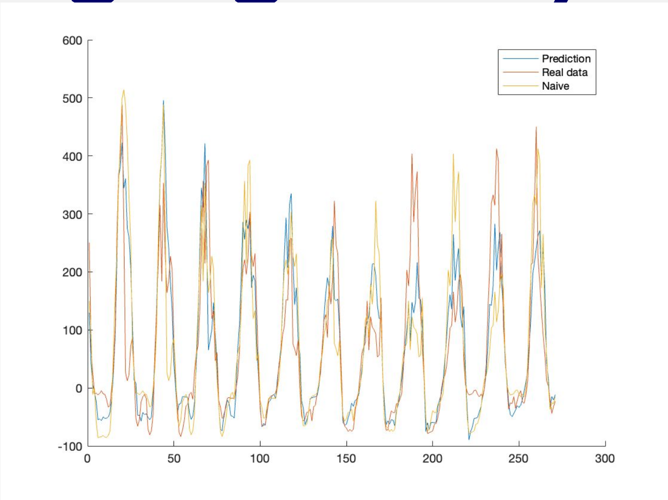
The input model looked like...

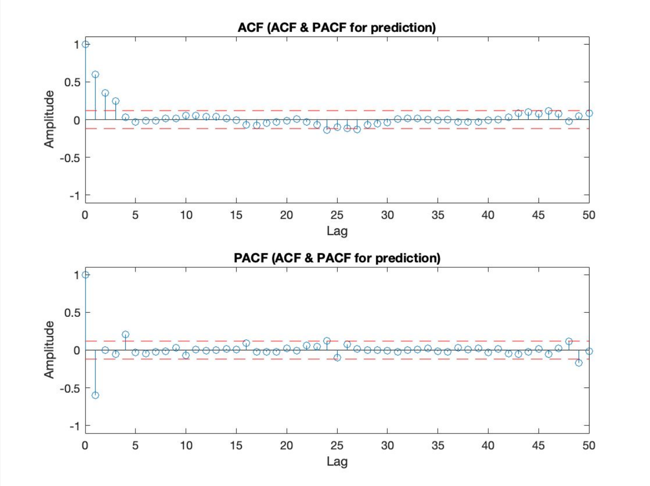
... This for the validation data (for lag one)



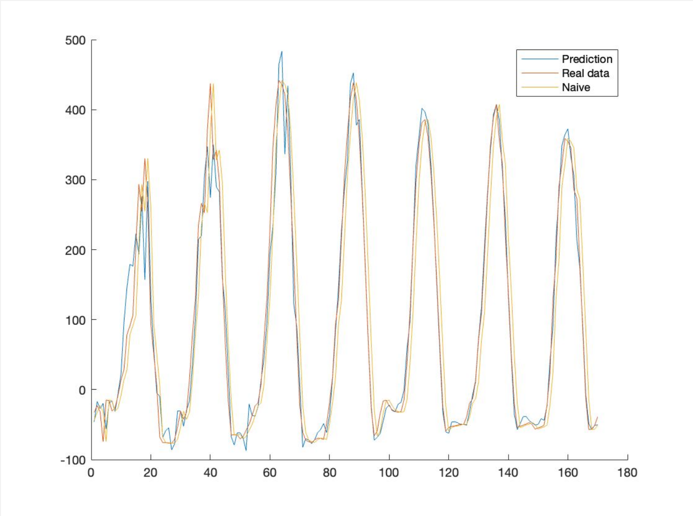


… This for the validation data (for lag eighteen)





This was very promising so the test data was tested as well, this time lag 1



The ARMA(24,9) model:

A(z) = 1 + 0.05284 (±0.03745) z^-1 - 0.1012 (±0.03679) z^-2 + 0.2068 (±0.03789) z^-3 + 0.5606 (±0.02286) z^-24

C(z) = 1 + 0.8191 (±0.04665) z^-1 + 0.5513 (±0.06221) z^-2 + 0.5943 (±0.06603) z^-3 + 0.4171 (±0.05822) z^-4 + 0.2914 (±0.0519) z^-5 + 0.2071 (±0.04621) z^-6 + 0.1534 (±0.04043) z^-7 + 0.1111 (±0.03673) z^-8 + 0.0381 (±0.02828) z^-9

where the process is described by A(z)yt = C(z)et

(This is excluding the differentiation - ARMA(48,9) with it)

For the output model:

The Box–Jenkins was applied and the (d,r,s) was estimated as (0,0,4) from the cross-correlation

The polynomials were then estimated to be:

Let’s test it!

B(z) = 0.00625 (+/- 0.0002288) + 0.00420 (+/- 0.0002401) z^-1 + 0.00182 (+/- 0.0002399) z^-2 + 0.00105 (+/- 0.0002276) z^-3

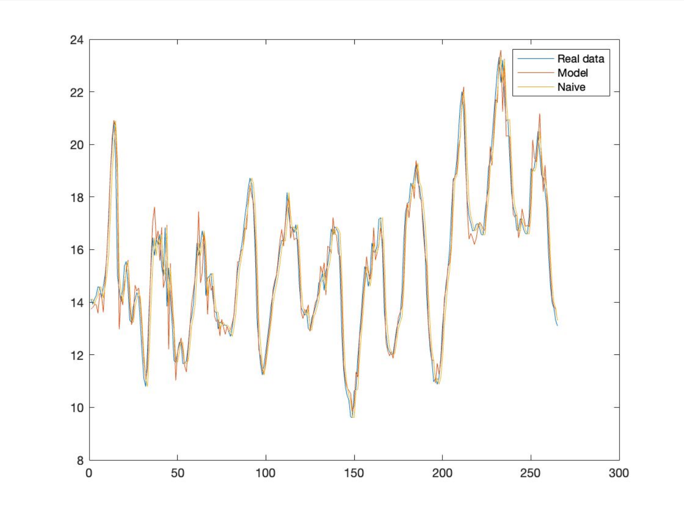
C1(z) = 1 - 1.698 (+/- 0.1172) z^-1 + 1.346 (+/- 0.1655) z^-2 - 1.333 (+/- 0.1631) z^-3 + 1.178 (+/- 0.1348) z^-4 - 0.1203 (+/- 0.09701) z^-5 - 0.3175 (+/- 0.08824) z^-6 + 0.1102 (+/- 0.08502) z^-7 - 0.1688 (+/- 0.08511) z^-8 + 0.07483 (+/- 0.08558) z^-9 - 0.1028 (+ /- 0.08502) z^-10 + 0.1213 (+/- 0.08649) z^-11 - 0.1267 (+/- 0.08604) z^-12 + 0.1316 (+/- 0.08718) z^-13 - 0.1172 (+/- 0.08628) z^-14 + 0.02185 (+/- 0.08637) z^-15 + 0.0567 (+/- 0.08572) z^-16 - 0.05738 (+/- 0.08567) z^-17 + 0.09134 (+/- 0.08512) z^-18 - 0.09669 (+/- 0.08569) z^-19 + 0.04958 (+/- 0.08089) z^-20 - 0.003307 (+/- 0.07201) z^-21 + 0.06523 (+/- 0.06191) z^-22 - 0.03656 (+/- 0.03613) z^-23

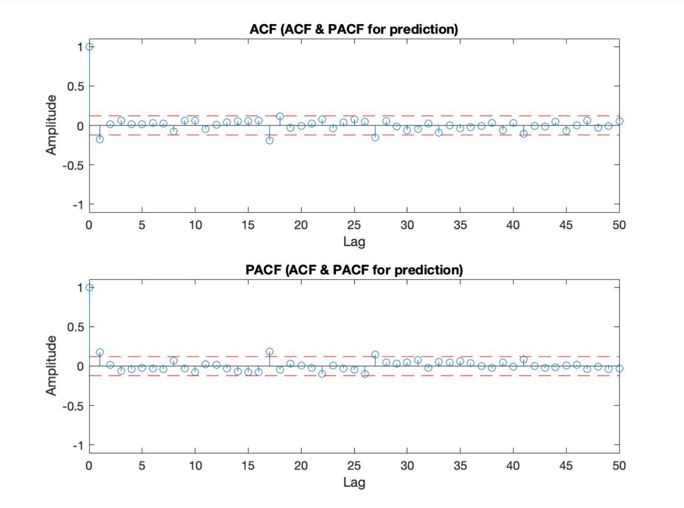
A1(z) = 1 - 3.056 (+/- 0.1136) z^-1 + 4.086 (+/- 0.2981) z^-2 - 3.931 (+/- 0.3872) z^-3 + 3.692 (+/- 0.3576) z^-4 - 2.554 (+/- 0.2491) z^-5 + 0.7638 (+/- 0.09512) z^-6

A2(z) = 1

Where yt=B(z)/A2(z)xt-d+C1(z)/A1(z)et

The validation data (lag one)

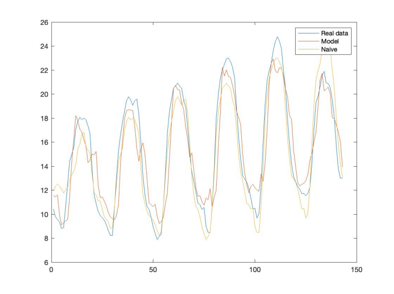


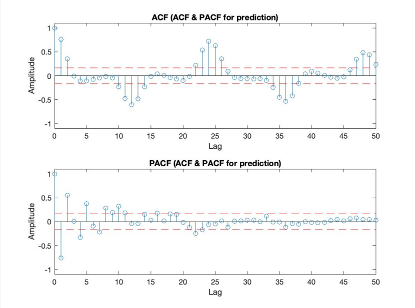


Looks great!

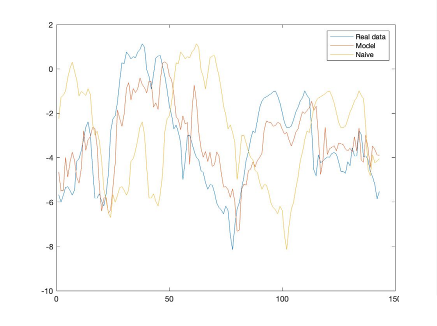
This was very promising so the test data was tested as well, this time lag 3 is showed

First test data

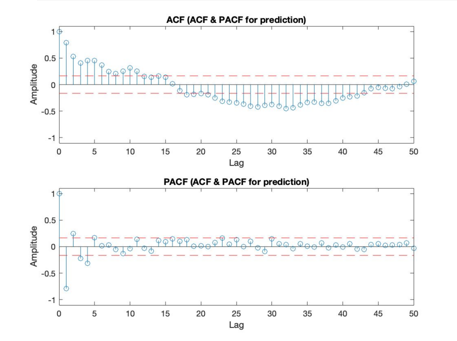




Second test data



This is also quite promising!



04

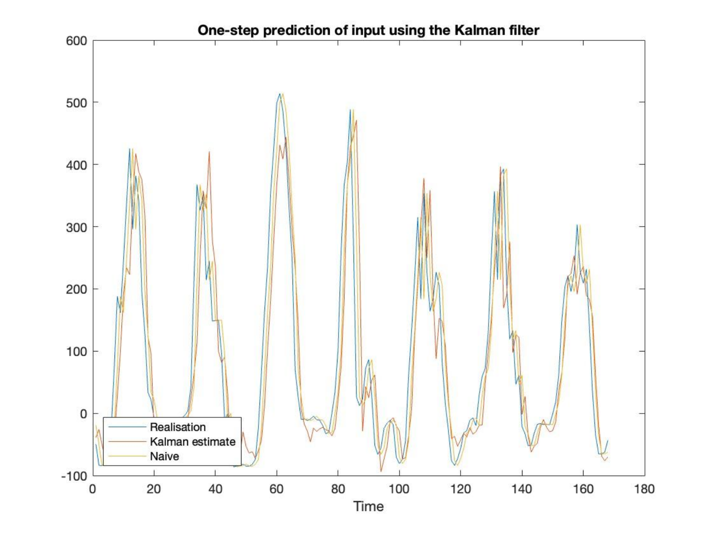
Part C

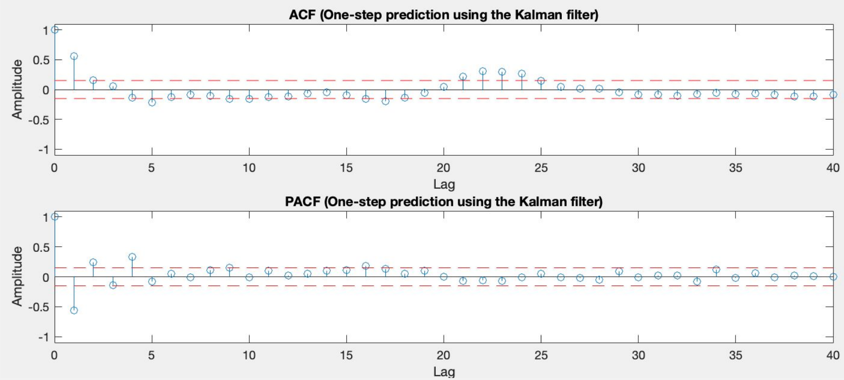
The third predictor

- Recursive predictor with input
- Kalman filter was used to estimate input model first
  - Based on model from part B
  - Ran through a Kalman filter to recursively improve the predictions
- Then Kalman filter estimated output model
  - Based on model from part B
  - Used input data from temperature and radiation that was run through the filter to create predictions

The input model looked like...

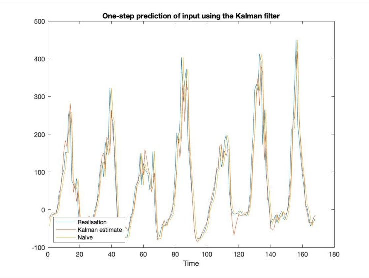
… This for the validation data (for lag one)

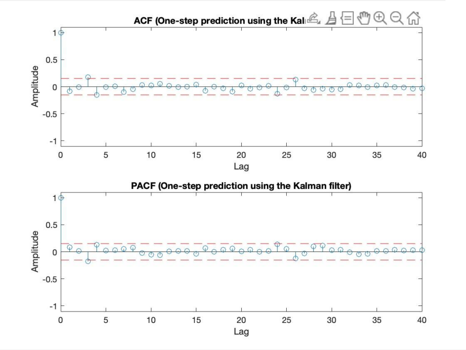




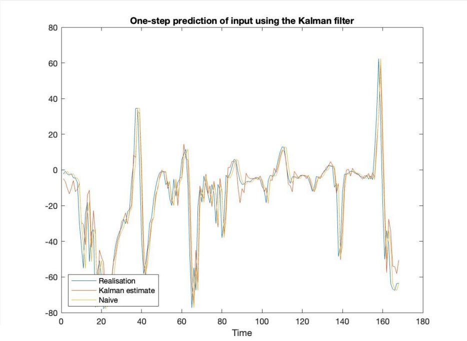
The test data was tested as well for the input (lag 1)

First test data

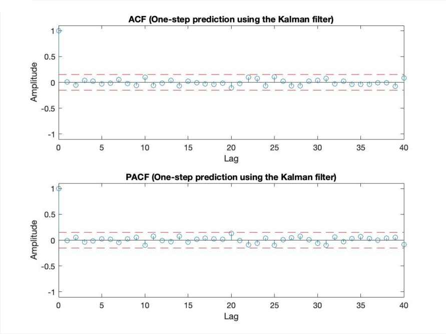




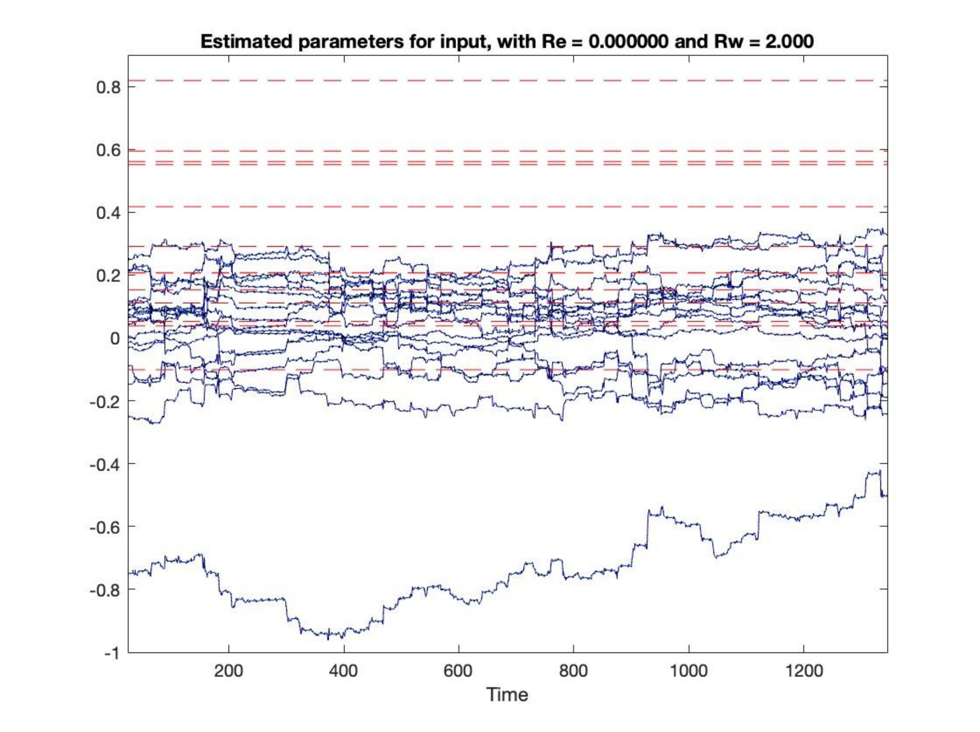
Second test data



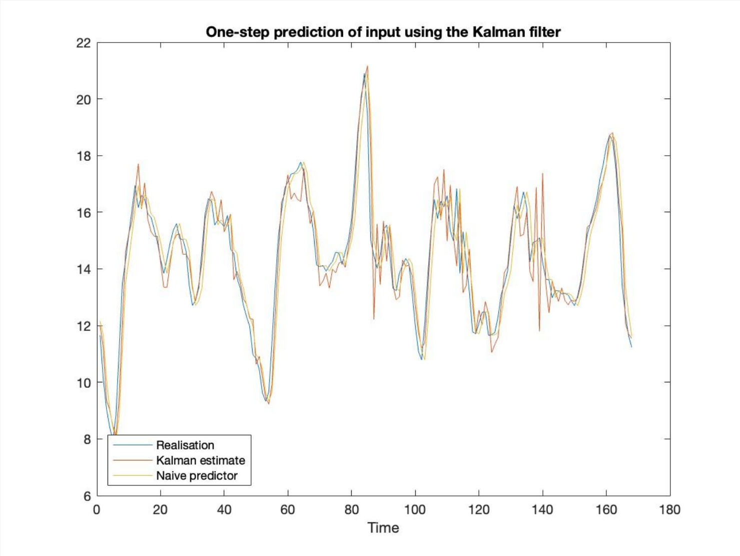
This is very promising and can possibly make up for deficiencies with the validation data!

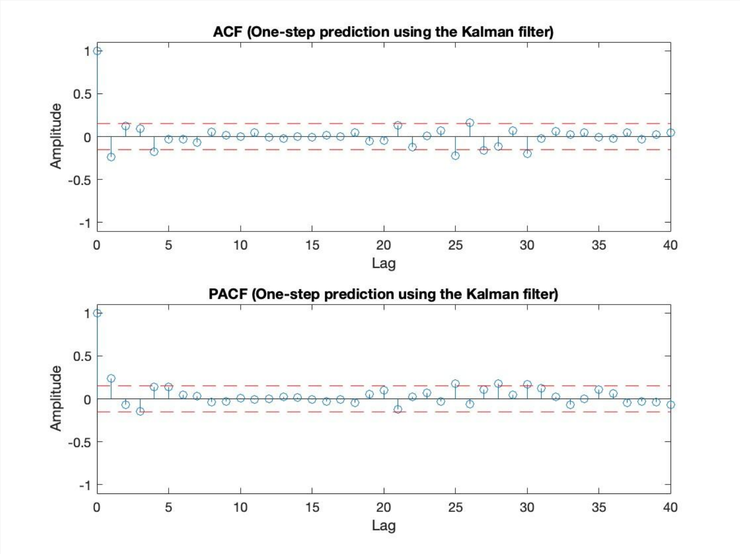


The input parameters:



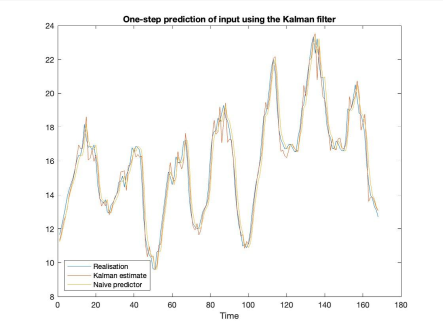
For the output model validation data (lag one):

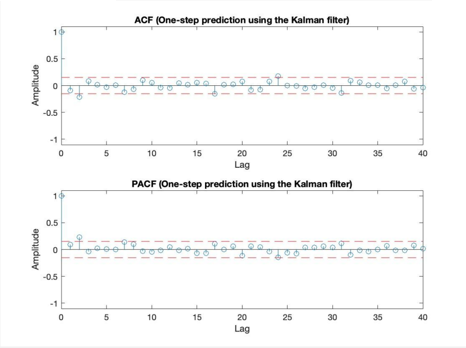




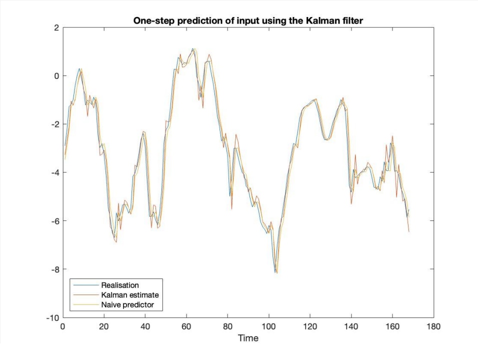
The test data was tested as well for the output (lag 1)

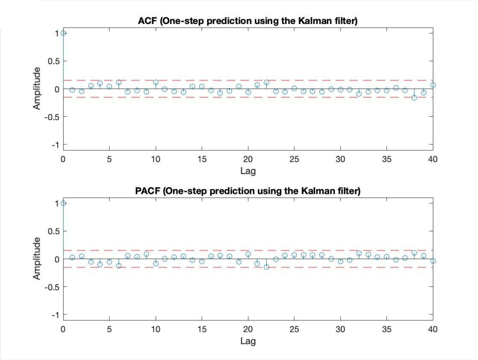
First test data





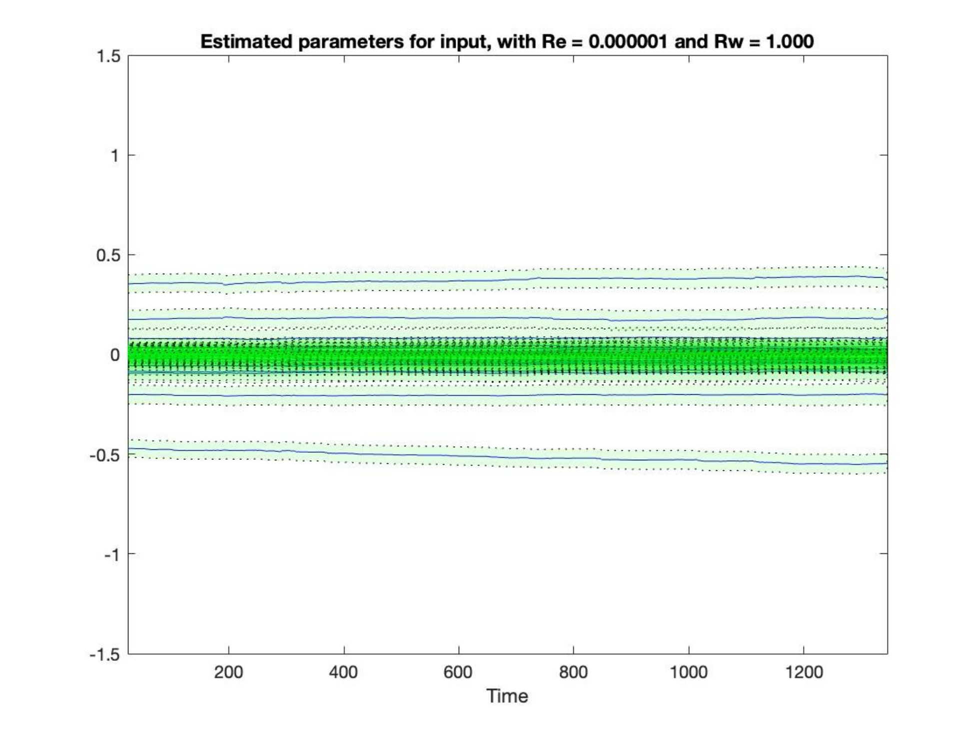
Second test data





This looks very promising!

The input parameters:



05

Analysis

Comparison

- A typically has the lowest normalized variance
  - It is possible that the input does not have a large impact after all
- There is no clear winner between B and C
  - Their specific traits complement each other and they work differently well on different data sets and lags
- The Kalman filter has the strongest advantage for long lags

Improvements

It is very time-consuming to create a good prediction
We spent a lot of time trying to find the best one, but it was impossible to test all kinds that we would have wanted to

The different test data sets worked differently well with different models
The test data sets we had were randomly chosen (or directly after the validation period) but how they look have a huge impact on how the model is judged
In many cases we tried many test data sets, but one always wants to test more

06

Conclusions

Conclusions

- All three models were reasonable and mostly better than the naive predictor
  -> Success!!

- The models worked differently well on different predictions
  -> Hard to pronounce a clear winner

- The input model did not have a strong impact on the results
  -> Maybe the net radiation does not affect the temperature very strongly

- It is very difficult to predict the future

Thank you!