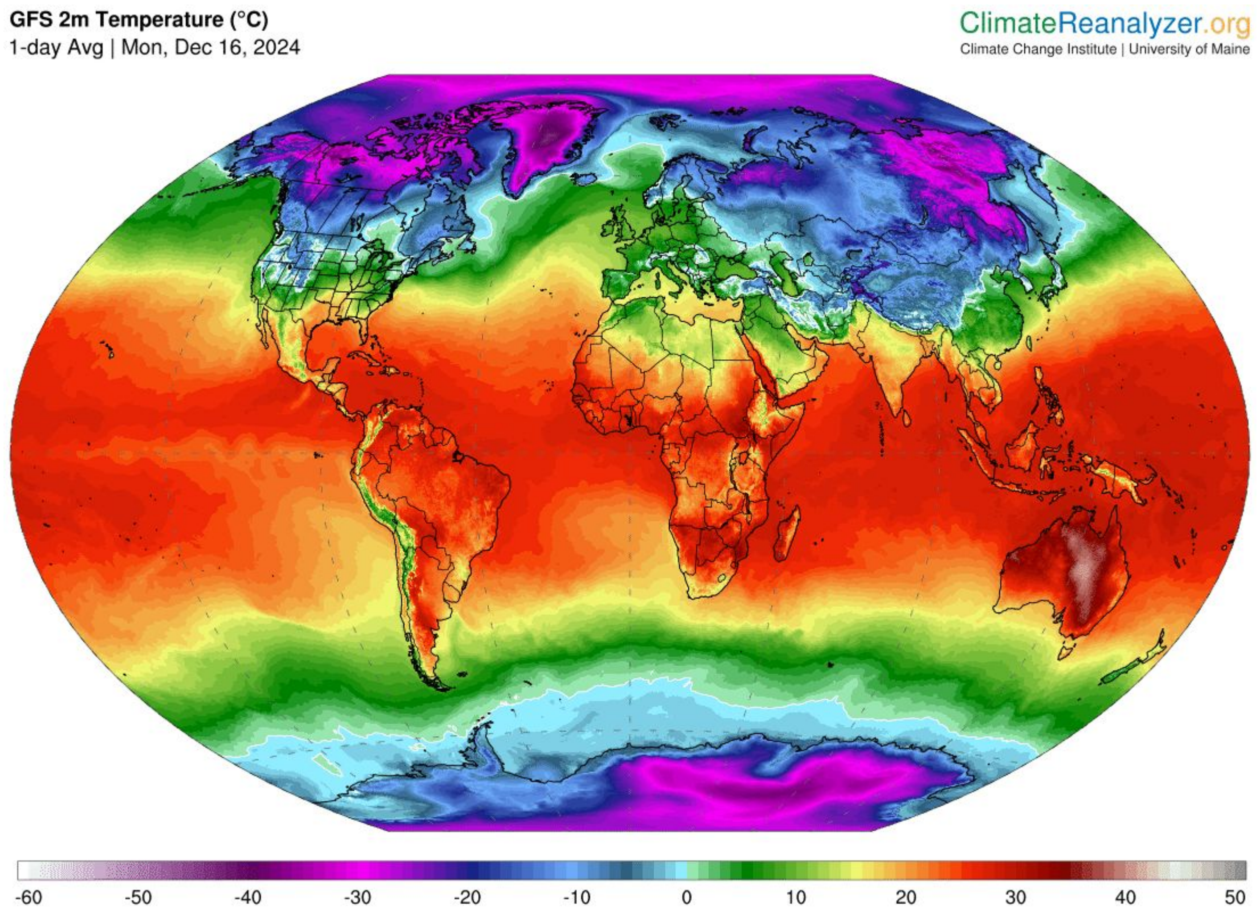
Weather prediction



By George Petropoulos, Thijs Spoor & Skipper van den Brekel

Universiteit van Amsterdam

Dataset

- Variables

- Only one location

Research question 1

To what extent does the Vector Autoregression improve the Linear Regression Model? Will capturing interdependencies between naturally related variables (e.g. temperature and humidity) improve modeling with VAR?

H0: Vector Autoregression does not improve the Linear Regression Model in weather prediction.

Research question 2

To what extent are periodic variables that follow seasonal patterns more predictable than variables that don’t exhibit these tendencies?

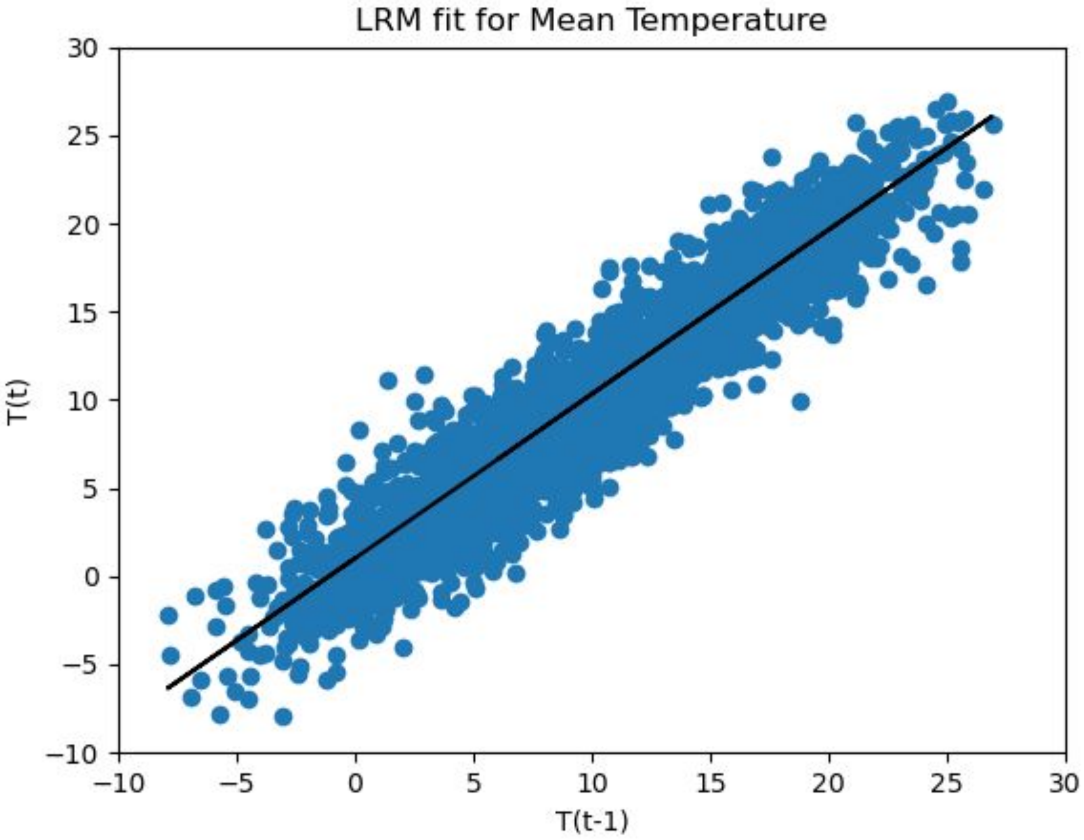
H0: Variables exhibiting seasonal periodicity do not have statistically lower forecasting errors than variables without such patterns.

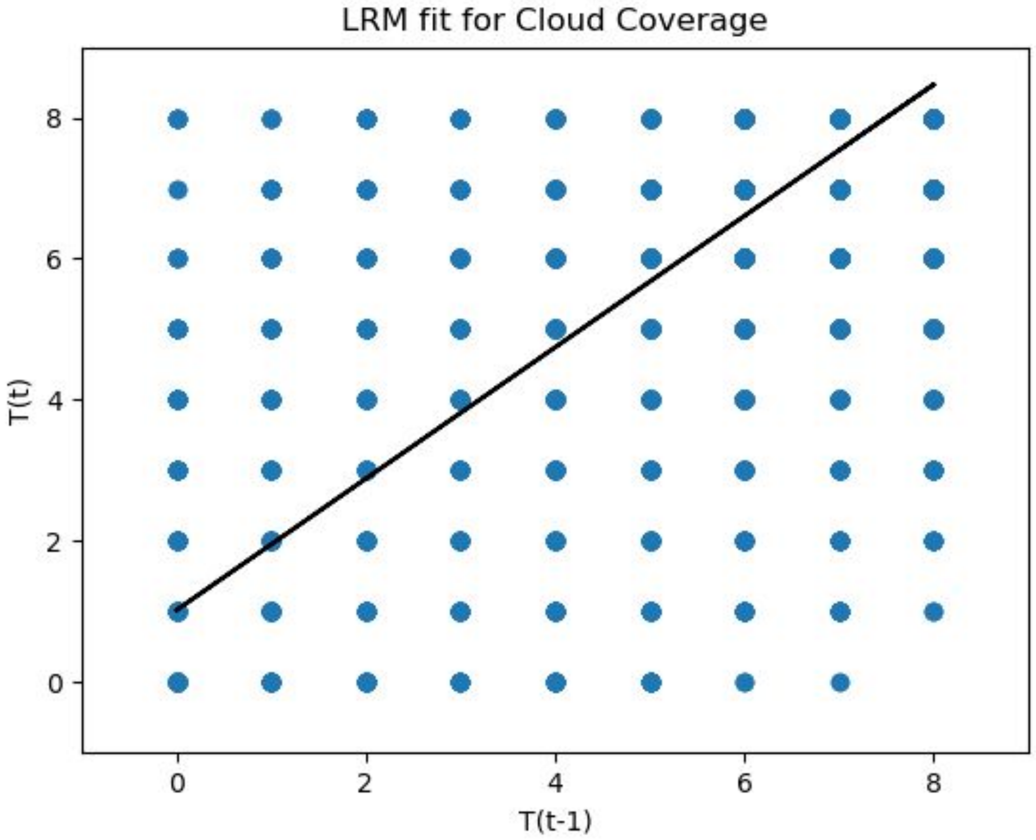
Research question 3

What is the effect of sample size on predicting power with a Vector Autoregression in weather prediction?

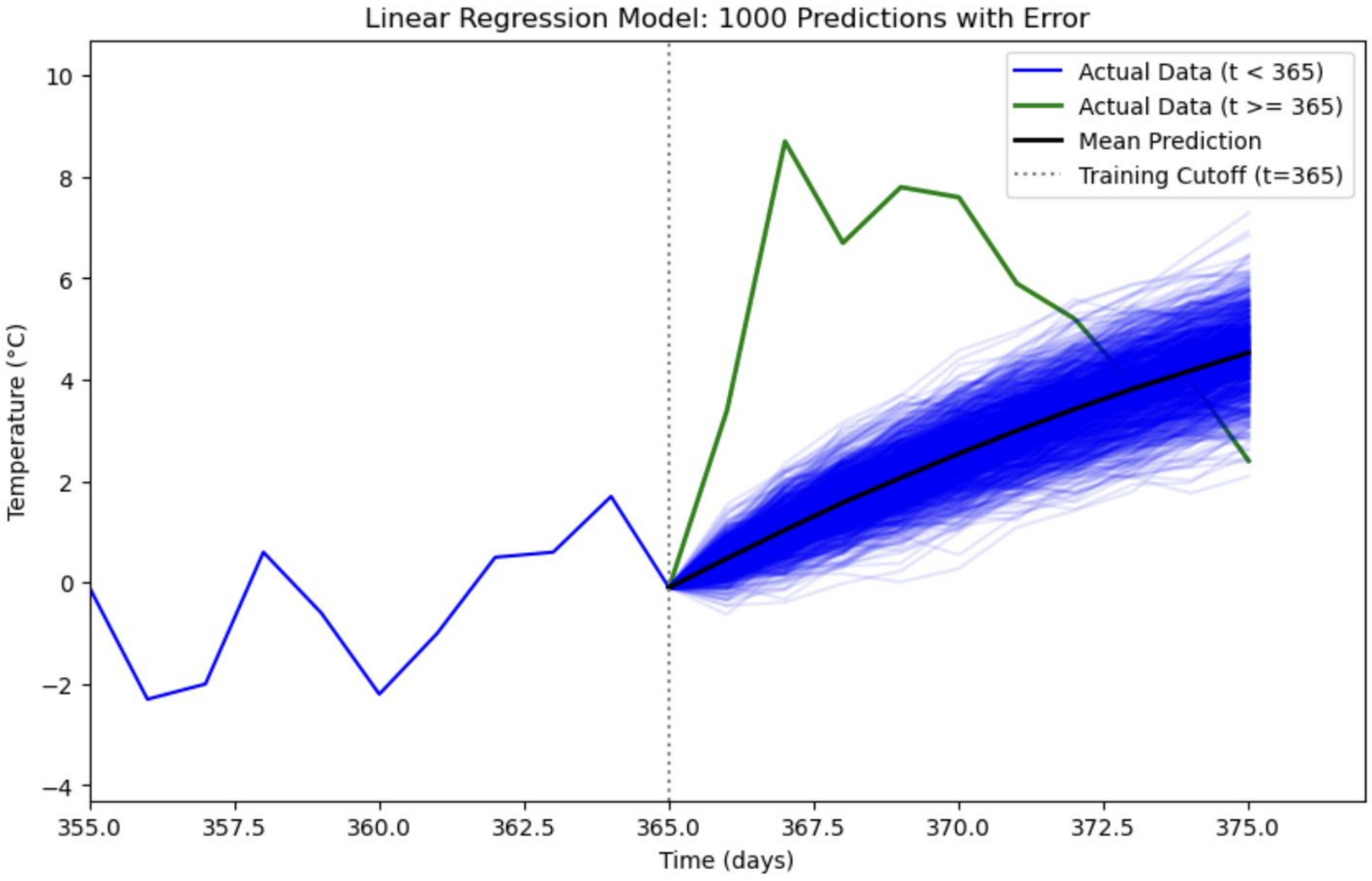
H0: Increasing the sample size does not have an effect on the predicting power of a Vector Autoregression.

LRM fit


LRM fit for Mean Temperature
T(t)
T(t-1)
30
25
20
15
10
5
0
−5
−10
−10
−5
0
5
10
15
20
25
30


LRM fit for Cloud Coverage
T(t)
T(t-1)
8
6
4
2
0
0
2
4
6
8

LRM prediction



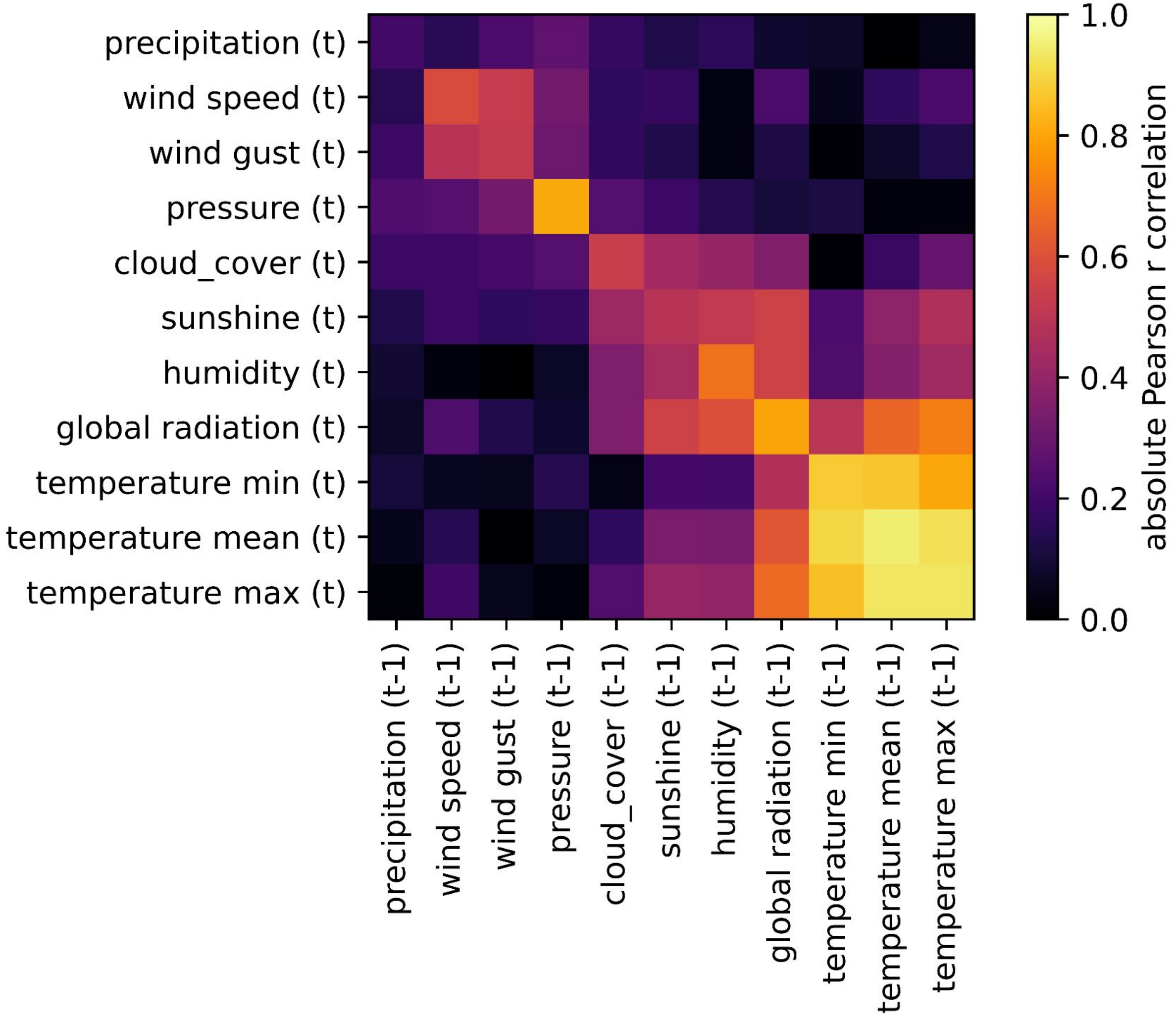
Universiteit van Amsterdam

VAR explanation

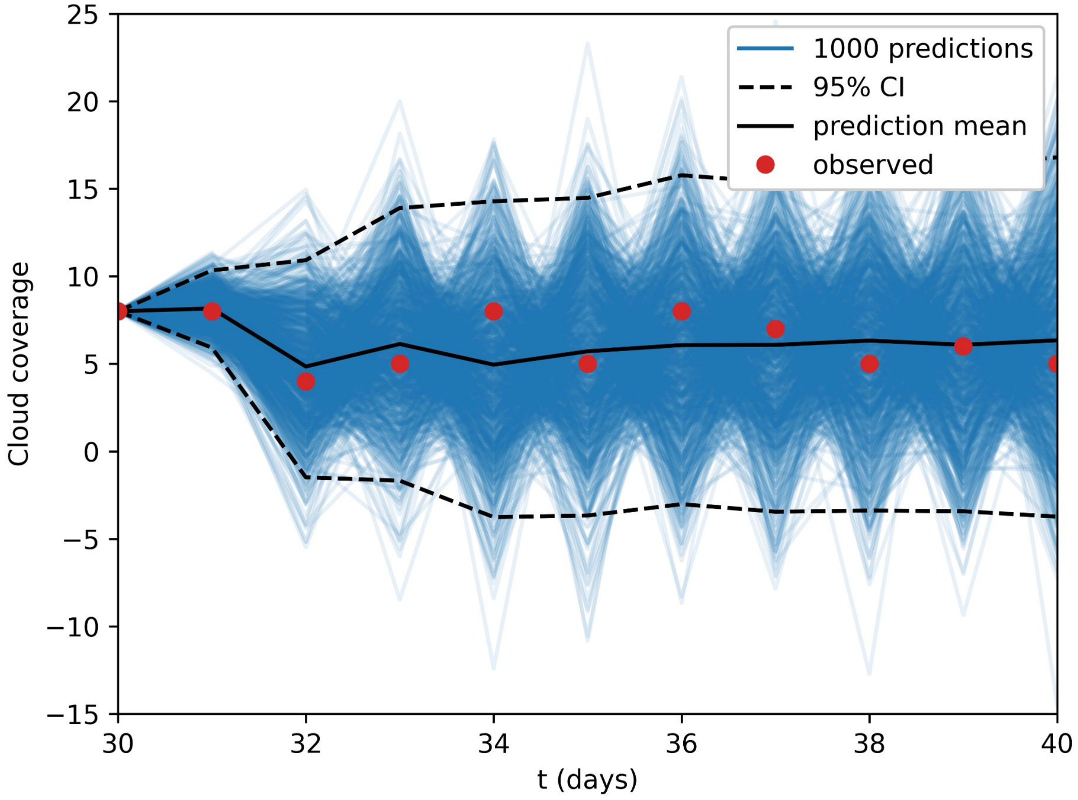
$$\vec{y}(t) = \vec{c} + M\vec{y}(t-1) + \vec{e}$$

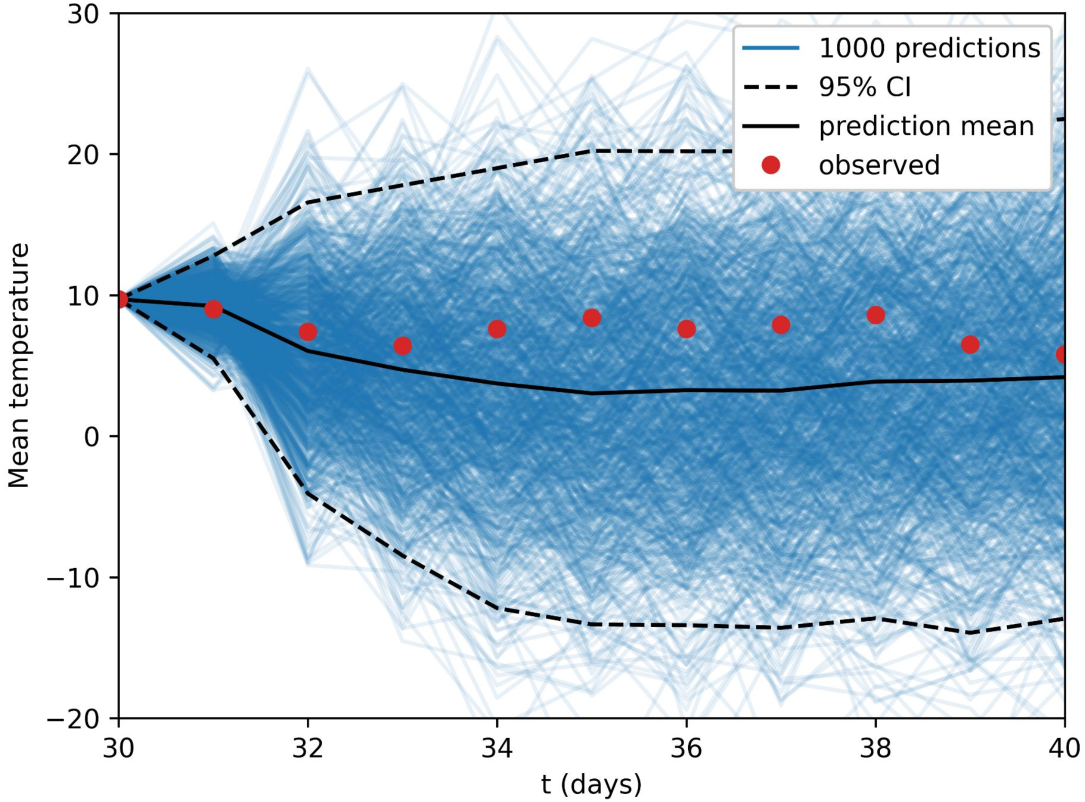
$$Y_t = C + M \cdot Y_{t-1}$$

$$\text{residuals} = Y_t - (C + M \cdot Y_{t-1})$$



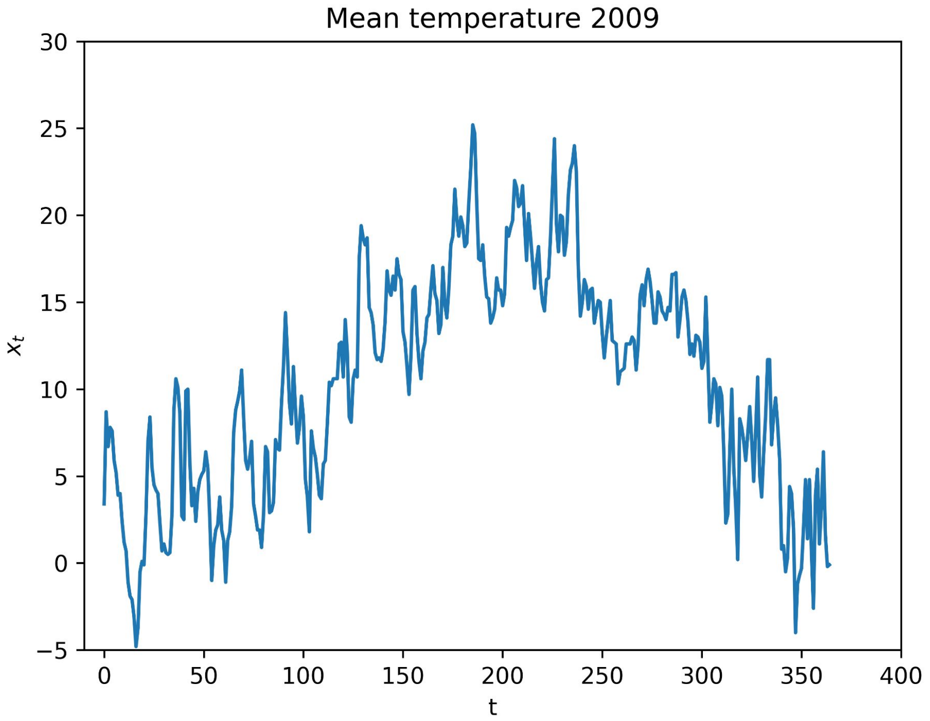
VAR predictions (alternative)

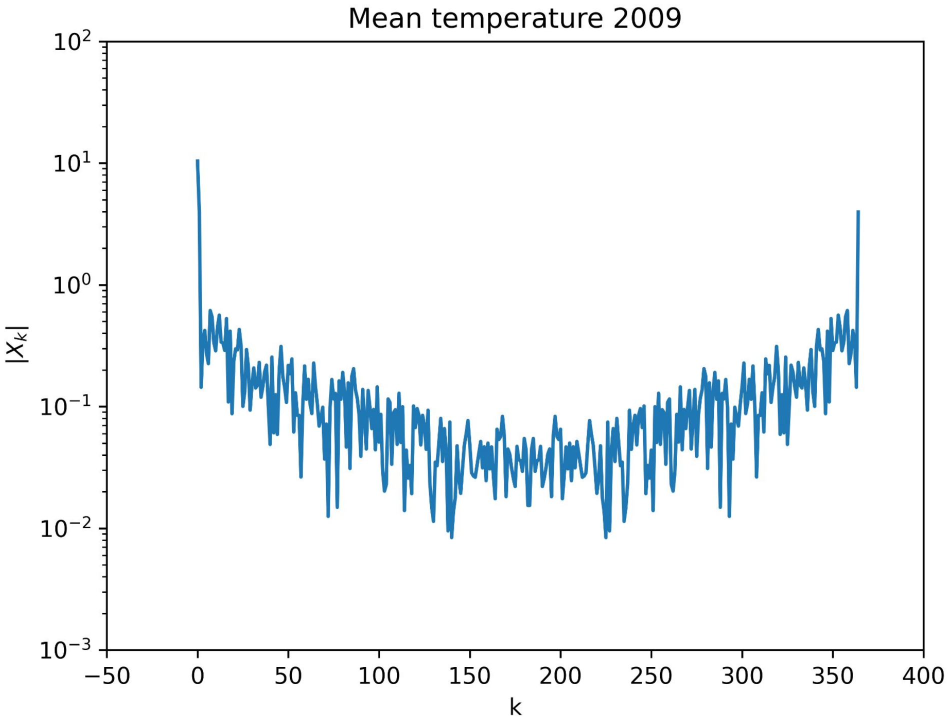




Universiteit van Amsterdam

Discrete fourier transform: $X_k = \sum_{t=0}^{N-1} x_n \cdot e^{\frac{-i2\pi kt}{N}}$

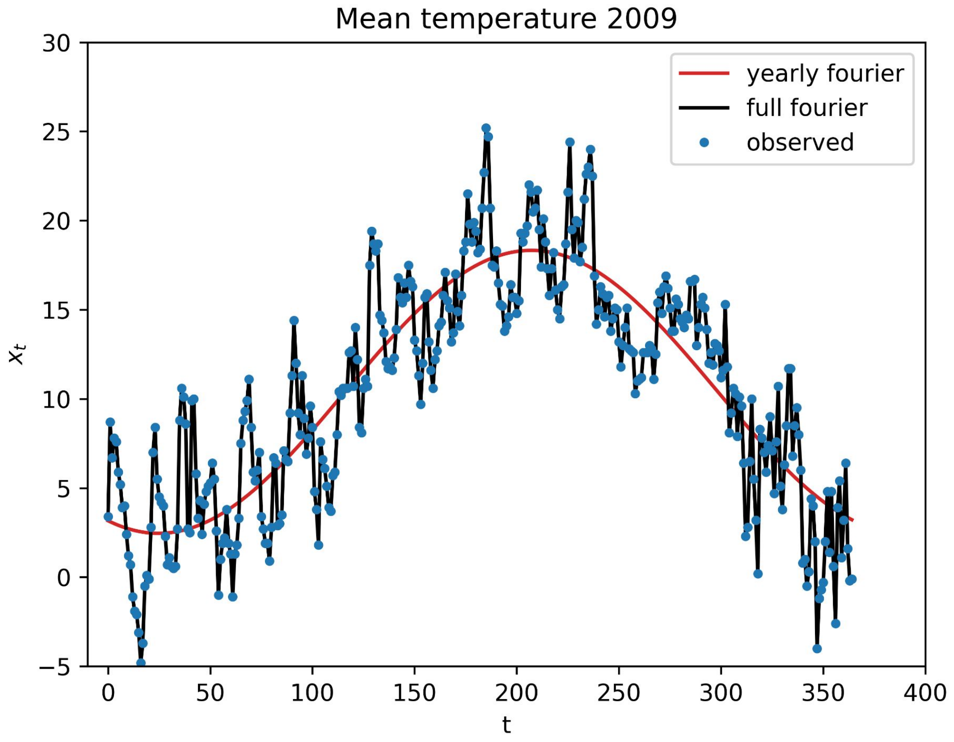


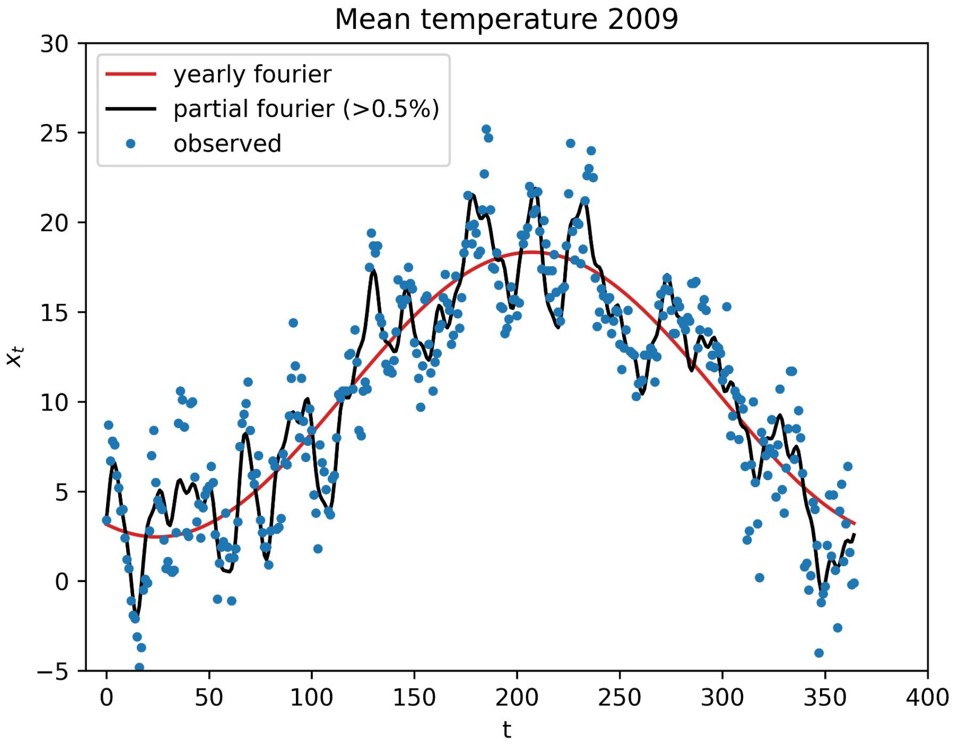


Reconstruction:

$$x_t = \sum_{k=0}^{N-1} A_k \cos\left(\frac{-i2\pi kt}{N} + \delta_k\right)$$

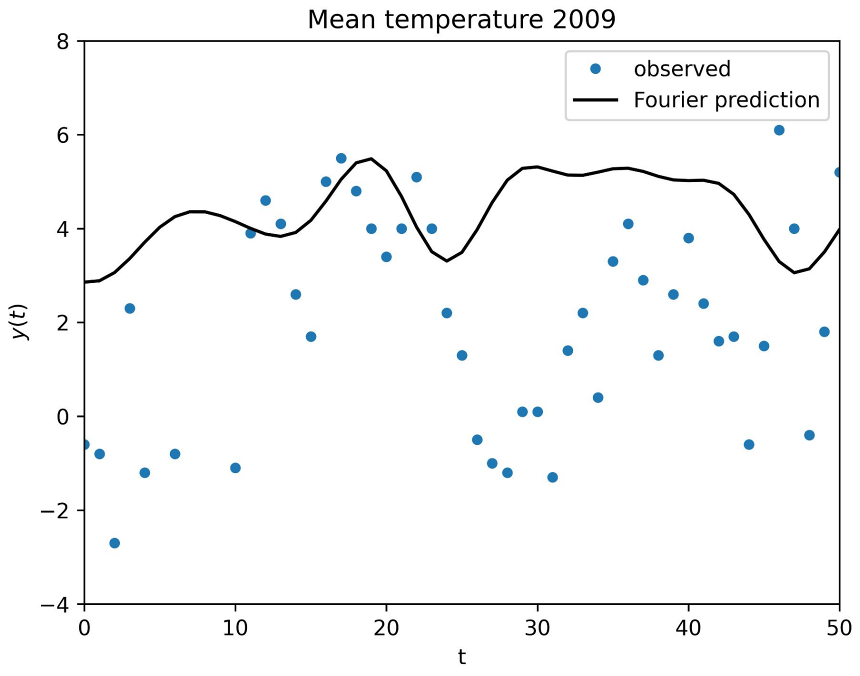
$$A_k = |X_k| \qquad\qquad \delta_k = \arg(X_k)$$

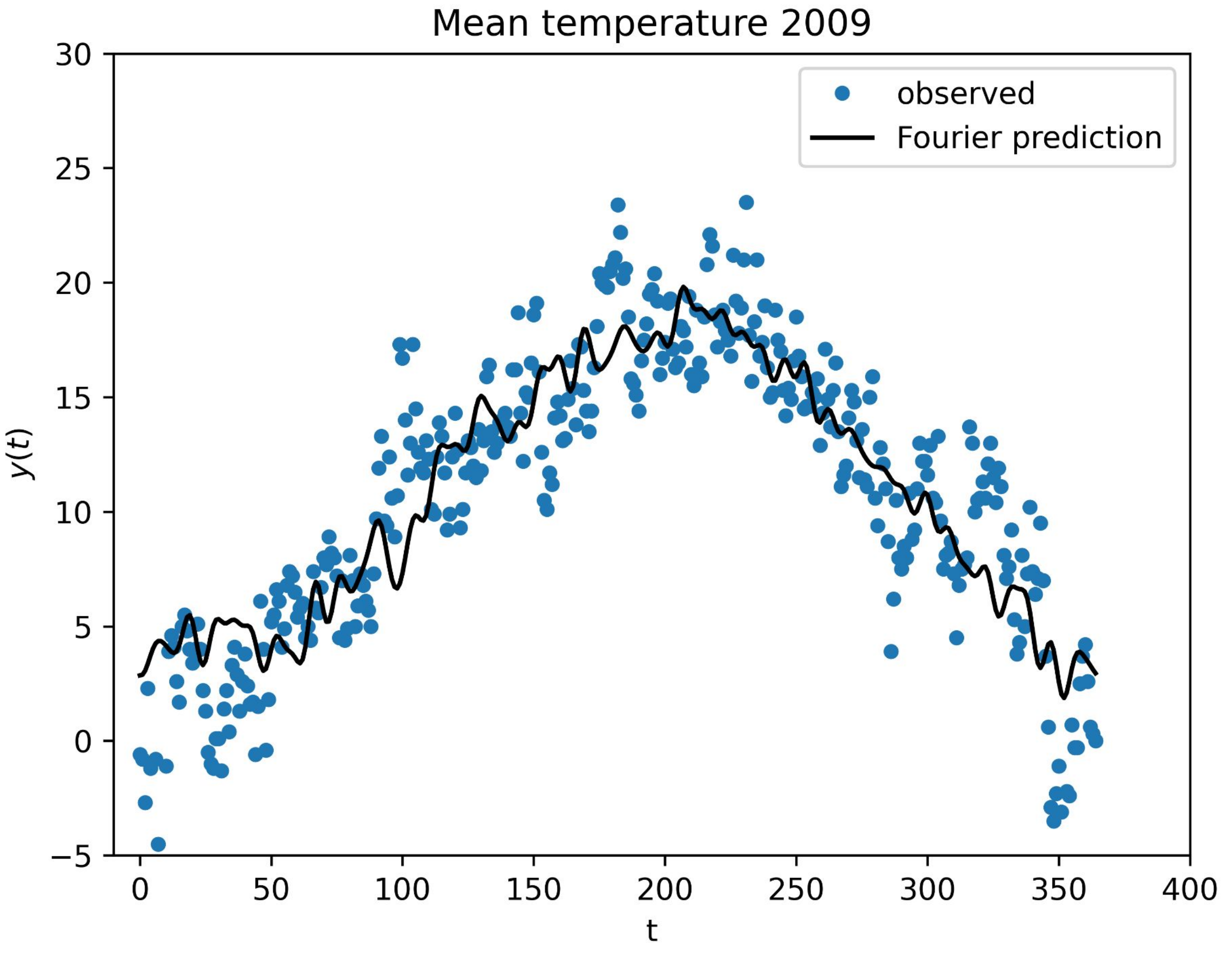




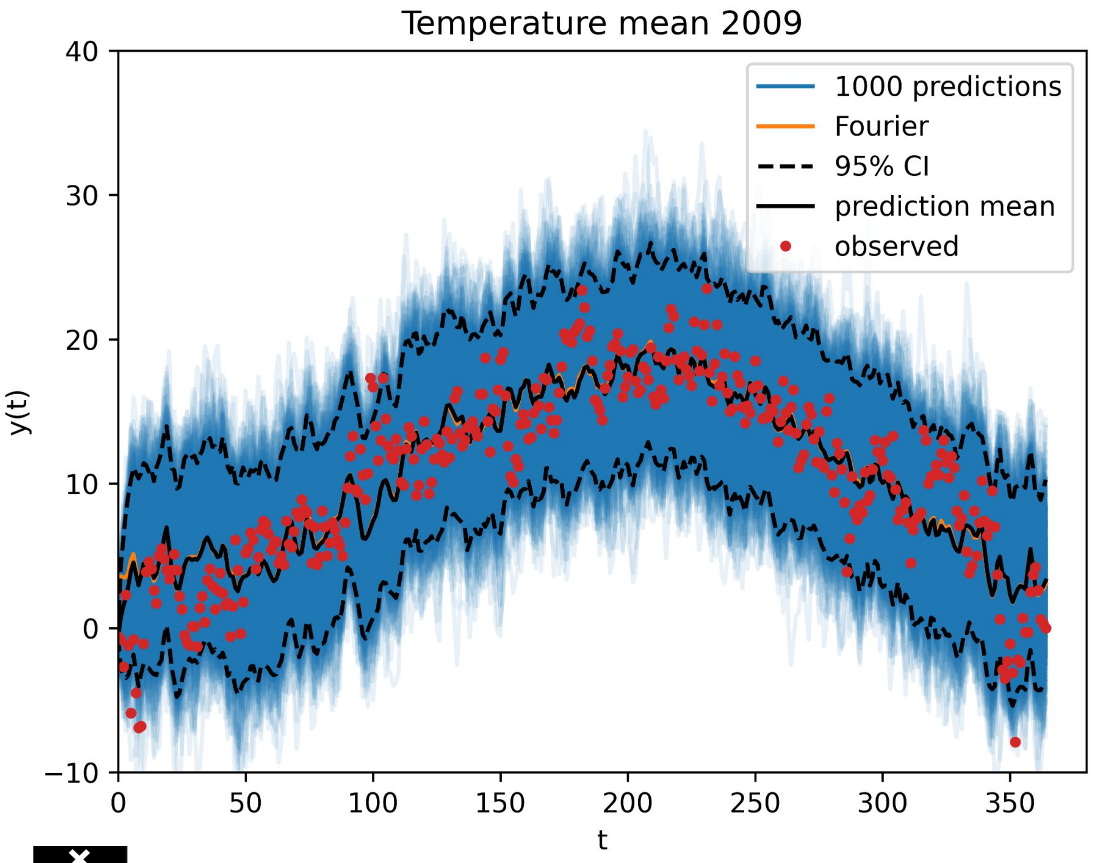
Fourier prediction

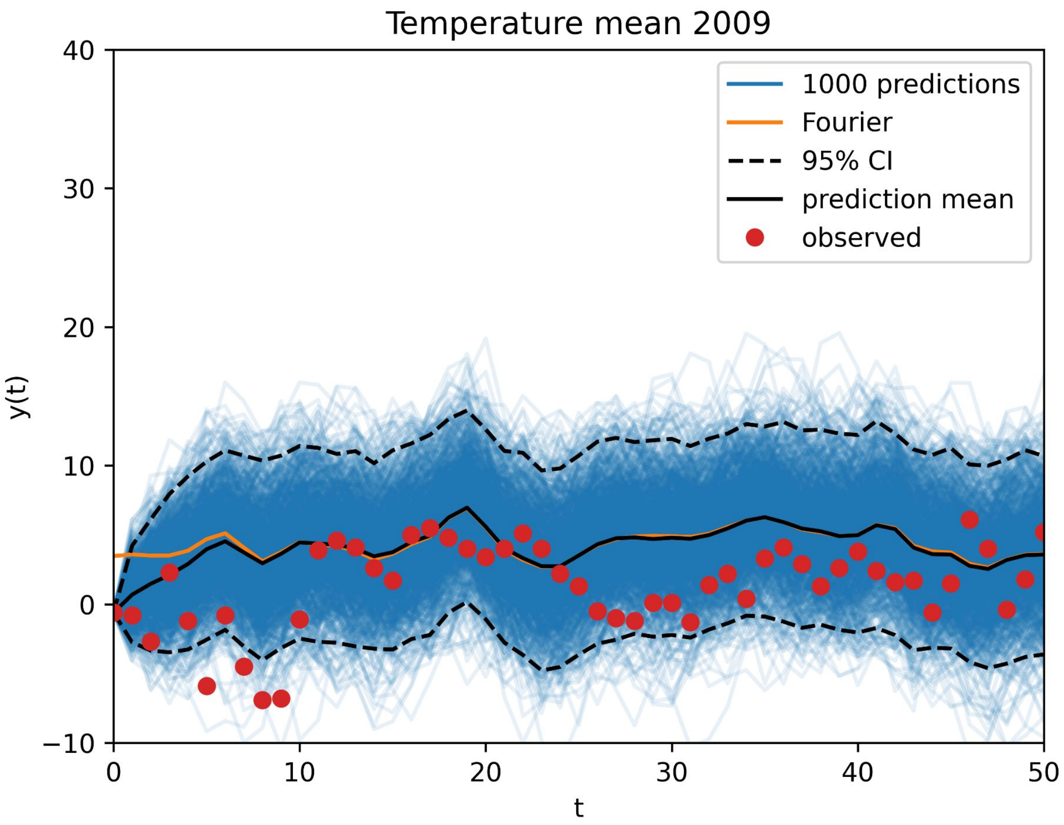
$$y(t) = F(t)$$



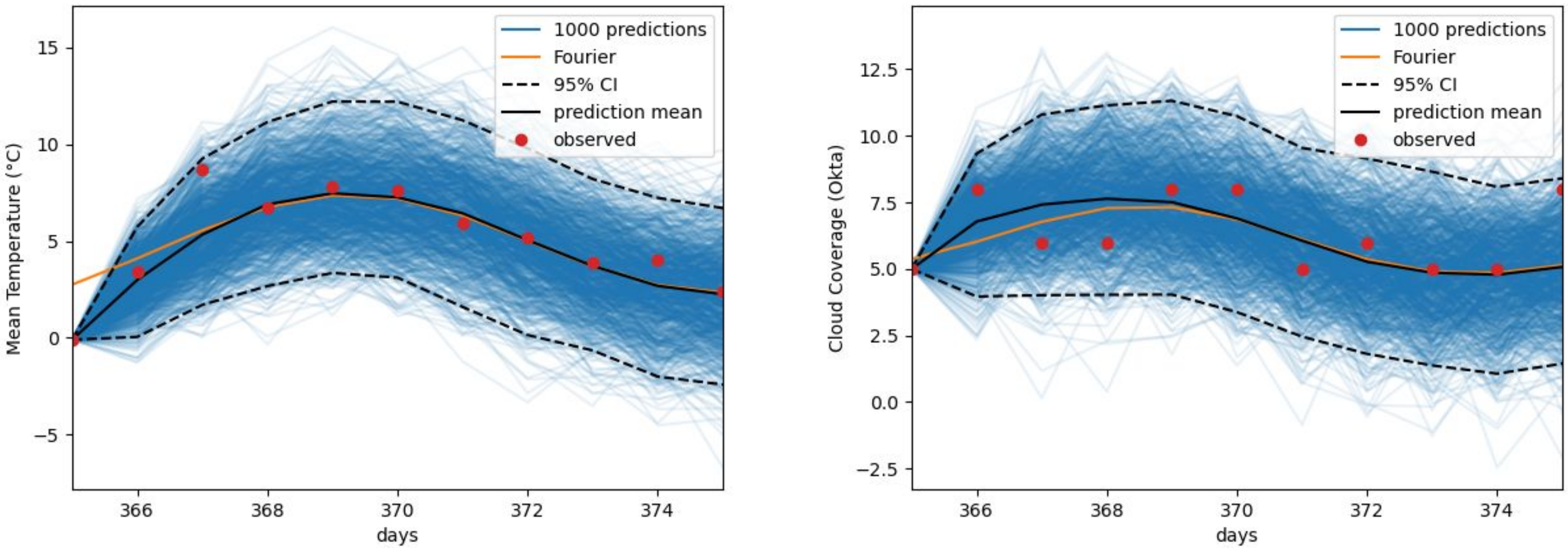


Hybrid prediction: $\vec{y}(t) = \vec{F}(t) + M \cdot (\vec{y}(t-1) - \vec{F}(t)) + \vec{e}(t)$





Hybrid (lucky) prediction



Question 1 Results:

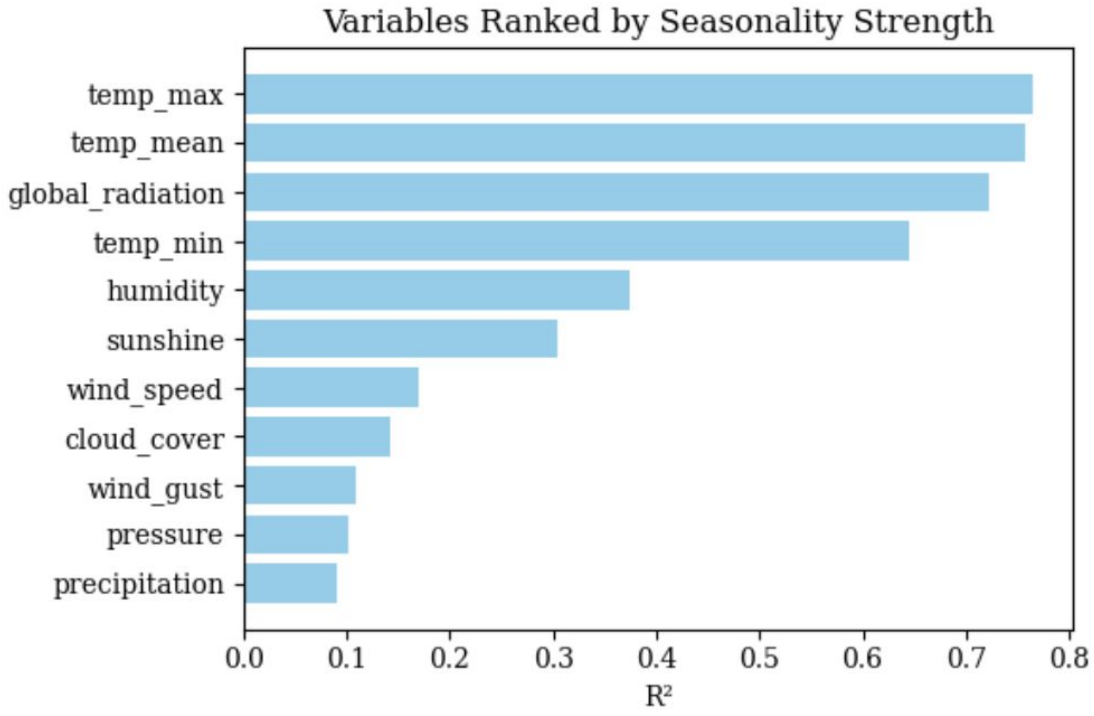
H0 : Vector Autoregression does not improve the Linear Regression Model in weather prediction. (Reject)

Diebold-Mariano test

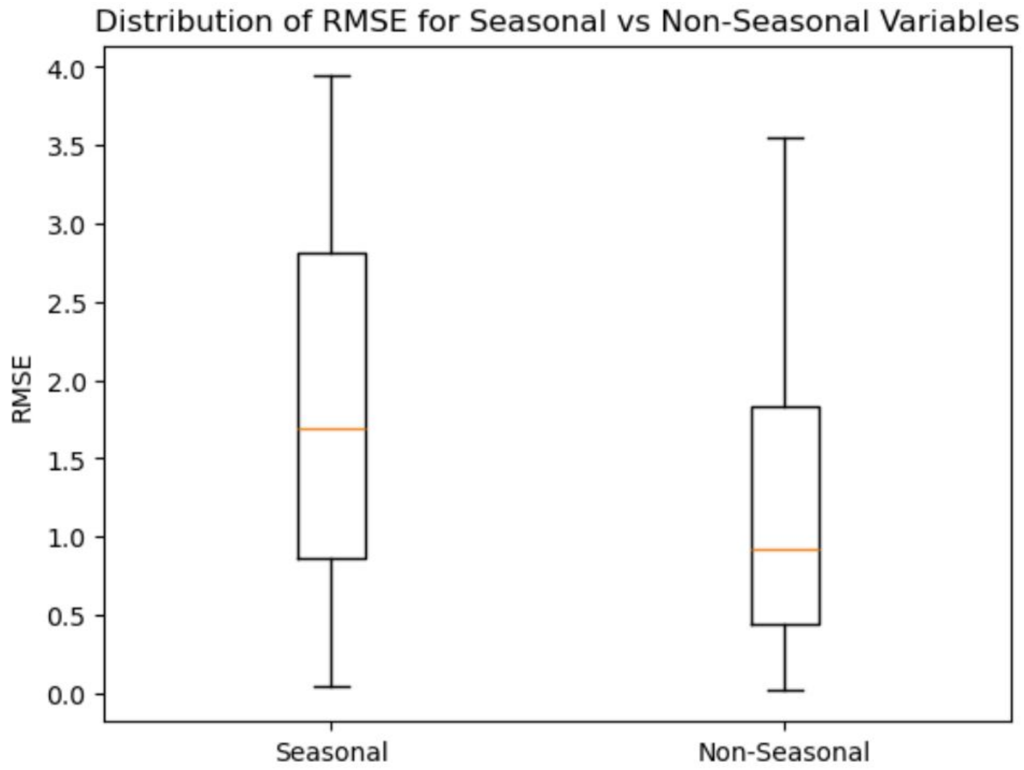
Variable	DM_stat	p_value	R^2 > 0.85	highly correlated r > 0.5
cloud_cover	0.843585	0.398902		x
wind_speed	1.733281	0.083046		x
wind_gust	0.881572	0.378008		
humidity	3.114168	0.001845		x
pressure	-2.239301	0.025136	x	
global_radiation	4.665723	0.000003		x
precipitation	1.354706	0.175511		
sunshine	2.621137	0.008764		
temp_mean	-2.450796	0.014254	x	
temp_min	-1.376901	0.168543	x	x
temp_max	-1.293419	0.195866	x	x

Question 2 Results

H0: Seasonality does not make a variable more predictable in the VAR framework (in terms of lower RMSE). (Accept)



Two Sample t-test



UNIVERSITEIT VAN AMSTERDAM

Question 3 Results

H0: Increasing the sample size does not have an effect on the predicting power of a Vector Autoregression. (Reject)

Sample Size	Aggregated MSE [1]
50	0.900 0.895
100	0.865 0.863
150	0.841 0.842
200	0.815 0.815
250	0.800 0.797
300	0.781 0.780
350	0.763 0.763
400	0.750 0.750

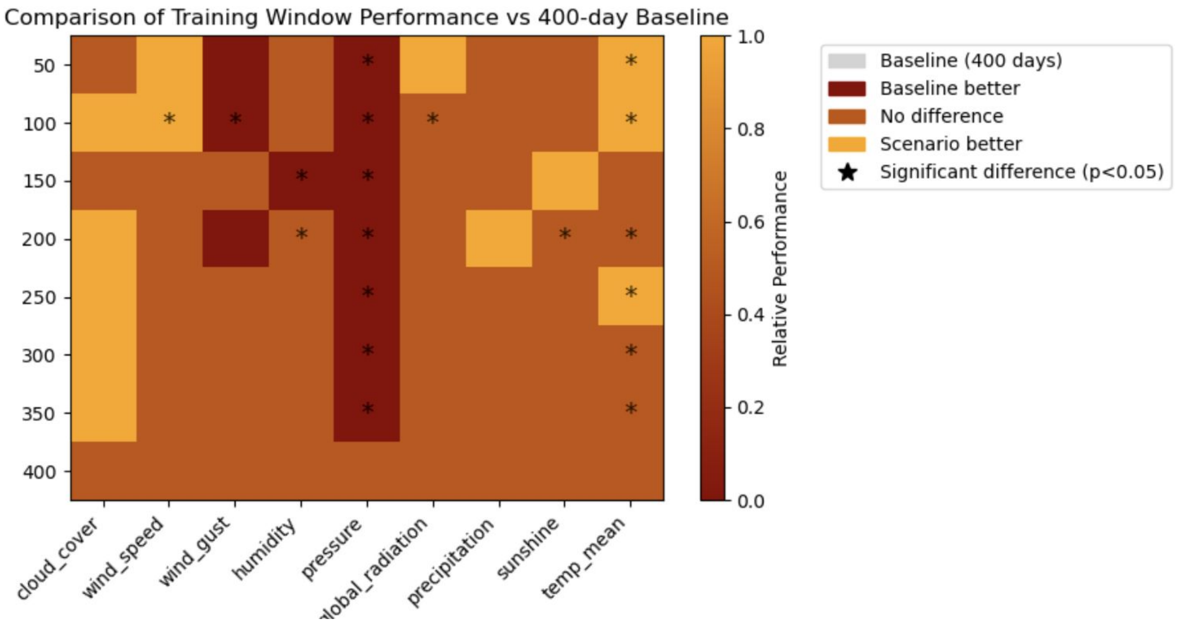
Pearson Correlation: -0.992

p-value: 0.000

Significance: Significant ($p<0.05$)

Conclusion: Reject H_0

Diebold-Mariano test



1. Please disregard the results of the Aggregate MSE. They are not statistically correct and should not have been included in the presentation.

Conclusion

- Weather Prediction is Hard!
- Var models do not provide consistent efficiency over standalone linear models, if that variable follows a strong linear trend.

Questions