02417: Time Series Analysis
Rune Grønborg Junker
Spring 2023

DTU Compute
Technical University of Denmark

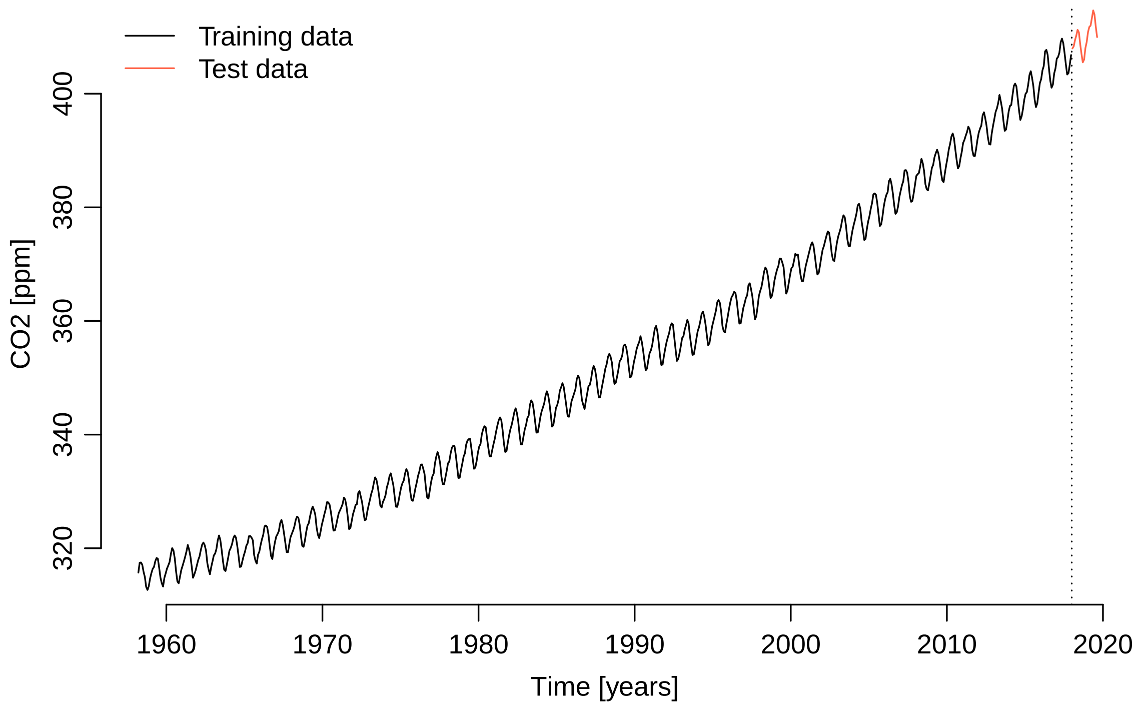
Assignment 1: Atmospheric CO2

In this assignment you will estimate different linear models for the development of atmospheric CO_2 over the past 60 years. Monthly observations from Mauna Loa, Hawaii, are provided by NOAA and the goal is to predict the concentration a couple of years ahead.
The data is provided in `A1_co2.txt` and includes four columns:

- year: Year measurement was made
- month: Month measurement was made
- time: Decimal year and month
- co2: Atmospheric CO2 [ppm]

The data is provided by NOAA/ESRL: `www.esrl.noaa.gov/gmd/ccgg/trends/`
You should not use the observations for years 2018 and 2019 (Last 20 observations) for estimations/training - only for comparisons/testing.
In this assignment you are expected to code your own implementations of the relevant algorithms. Do include the code representing your algorithms in your main report.

Question 1.1: *Read the data and plot the CO_2 as a function of time. Do indicate which data is used for training and testing. Comment on the evolution of the values over time. It is OK if the plot is combined with results from the following questions.*



A global upward trend is seen and there is also a clear annual cycle. There is an indication that the trend is increasing with time, although this is not very pronounced. To capture this with a linear model it is clear that we will need atleast one term for the trend (assuming that it is linear) and a term for the yearly variation.

Question 1.2: *First OLS and WLS models are to be tested. The models should contain an intercept, a slope and an annual harmonic part. The anticipated model is of the form:*

$$Y_t = \alpha + \beta_t t + \beta_s \sin\left(\frac{2\pi}{p}t\right) + \beta_c \cos\left(\frac{2\pi}{p}t\right) + \varepsilon_t$$

1. *Estimate the parameters in the above linear regression model (OLS).*
2. *Present the estimated parameters including a measure of uncertainty.*
3. *Plot the fitted values together with the data.*
4. *Now, we assume that the correlation structure of the residiuals is an exponential decaying function of the time distance between two observations, i.e.*

$$Cor(\varepsilon_{t_i}, \varepsilon_{t_j}) = \rho^{|t_i - t_j|} \tag{1}$$

   for some $0 < \rho < 1$*, which results in the following correlation matrix:*

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots \\ \rho & 1 & \rho & \cdots \\ \rho^2 & \rho & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{2}$$

5. *Estimate* ρ *using five iterations of the relaxation algorithm.*
6. *Present the estimated parameters based on this WLS model. Provide a measure of uncertainty for each of the estimates.*
7. *Again plot the fitted values.*
8. *Compare the two estimates and comment on the results.*

We fit a model of the proposed form, and since we observed a yearly variation, p is chosen to equal one year. Using yearly time steps is preferred, since they are easier to interpret, but using months is only a minor disadvantage. Table 1 shows the estimated parameters along with standard deviations of the estimates when using yearly time steps while Table 2 shows the same but for monthly time steps, where the first month of the data set is defined as month number 1. Notice that α depends on what time is chosen as the origin and a deviation in this paramter alone does not matter for the evaluation. We see that the estimate of β_t is approximately 12 times larger for the monthly time steps as expected, and, likewise, β_s is almost identical in both cases. For β_c we see that the sign has been changed. Notice that this means that the amplitude is the same, but a different phase was chosen by the model. The observed differences are caused by us forcing an arbitrary choice of phase for the harmonic functions.

The fitted values from the OLS model are shown in Figure 1, which we will get back to once we have dealt with the WLS estimates as well.

For the WLS model we assume the correlation structure proposed by the assignment. We estimate ρ using the relaxation algorithm, again assuming both monthly and yearly time steps. For the yearly time steps we get the following estimate of ρ for the first 5 iterations of the relaxation algorithm:

	Estimate	Std.
α	-2709.6696	14.9649
β_t	1.5405	0.0075
β_s	2.6138	0.1837
β_c	-1.0502	0.1840
σ^2	12.1343	

Table 1: OLS Estimates when using yearly time steps.

	Estimate	Std.
α	306.8000	0.2603
β_t	0.1284	0.0006
β_s	2.5909	0.1837
β_c	1.1058	0.1840
σ^2	12.1340	

Table 2: OLS Estimates when using monthly time steps.

```
## [1] 0.8050072
## [1] 0.8065016
## [1] 0.8065192
## [1] 0.8065194
## [1] 0.8065194
```

For the monthly time steps the first five iterations give:

```
## [1] 0.9820921
## [1] 0.9822411
## [1] 0.9822427
## [1] 0.9822428
## [1] 0.9822428
```

In both cases we see that the algorithm converges almost instantanously. We would expect that $\rho_{\text{month}} = \rho_{\text{year}}^{1/12}$, where ρ_{year} is row estimated using a yearly time step and vice-versa for ρ_{month}. Indeed, we get that $\rho_{year}^{1/12} = 0.9822406$.

Using these estimated values of ρ we do WLS-estimates of the parameters. The results can be seen in Table 3 for the year-based estimates and Table 4 for the month-based estimates.

	Estimate	Std.
α	-2743.9792	128.1808
β_t	1.5583	0.0645
β_s	2.6477	0.0676
β_c	-1.0188	0.0675
σ^2	12.2299	

Table 3: WLS Estimates when using yearly time steps.

Comparing the WLS estimates to the OLS estimates we see that they are quite similar. However, the estimated uncertainty of the estimates is very different. For example, when

	Estimate	Std.
α	307.3410	2.3189
β_t	0.1298	0.0054
β_s	2.5926	0.0676
β_c	1.1527	0.0674
σ^2	12.2274	

Table 4: WLS estimates for monthly time-steps.

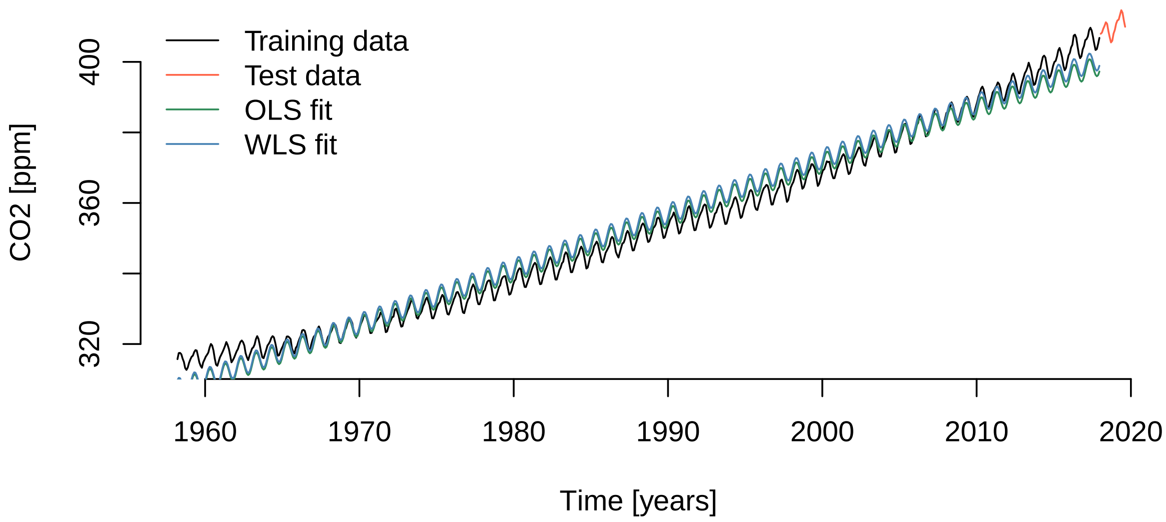


Figure 1: Fitted values using both OLS and WLS estimated parameters.

considering the year-based estimates, the estimated standard deviation of the estimate of β_s is 0.0075 in the OLS case while it is 0.0645 in the WLS case. This difference of almost a factor 10 is caused by the WLS recognising that it is basing the estimates on observations with correlated noise. This large correlation could give rise to coincidental trends, which of course impacts our parameter estimates. On the other hand the WLS estimator seems more confident about the estimates of the harmonic terms.

The OLS and WLS fit are plotted together in Figure 1. Both models fail to catch the increasing trend. However, they both capture the seasonality quite well. Normally, one should plot the residuals to check model assumptions. In this case it is clear from the plot above that the residuals cannot be assumed to be independent as the black line stays on the same side of the green and blue lines most of the time. This also means that these models can not be trusted for predictions. If we wanted to make a serious model for this data we would have to include at least one more term to capture the change in the trend. The two models are so similar that it does not make sense to talk about their differences.

Question 1.3: *Use a local linear trend model with* $\lambda = 0.9$.

1. *Provide* L *and* $f(0)$ *for the trend model corresponding to the linear model in the previous question.*

2. *Filter the data with the chosen model.*
3. *Plot the one step prediction errors and the resulting estimates of σ_t for each observation (Do skip the first 10 observations as transient).*
4. *Plot the data and the corresponding one step predictions for all observations in the training data. Include a 95% prediction interval.*
5. *Zoom in on the time since 2010 and include predictions for the test data. Again including a 95% prediction interval.*
6. *Make a table presenting the predictions 1, 2, 6, 12 and 20 months ahead.*
7. *Compare with the test data and comment on the results*
8. *Plot the data and the estimated mean for each time step (Typically the first element in θ_t for all t).*

No matter whether time is assumed to be in years or months, the matrix L and fector $f(0)$ are the same, namely:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0.866 & 0.5 \\ 0 & 0 & -0.5 & 0.866 \end{pmatrix}$$

$$f(0) = [1;0;0;1]^T$$

We use the local linear trend model with $\lambda = 0.9$, to obtain the one-step predictions shown on top of the data in Figure 2 and in a zoomed version on Figure 3. It is evident that the local linear trend model captures the data much better than the previously explored models, even for this non-tuned value of λ. Notice that the estimated prediction intervals shown on the plots have been calculated using a global estimate of the variance. The choice of using a global estimate was made based on the fact that it seemed like the variance is constant over time, and that the data experiences no rapid changes in behaviour. Figure 4 shows a plot of the estimated standard deviation over time using the global estimator. Since it almost converges (except for maybe a slight upwards trend) this is clearly a better choice than a local estimator. Also, note how most observations are captured within the prediction intervals as expected. The sharp eye will also notice how the prediction interval grows for the longer horizons of the test data.

We now take a look at the residuals which are plotted in Figure 5. Although it is not that clear from such a zoomed out plot, for the experienced data analyst it is evident that they exibit a lot of auto-correlation. This is seen by consequent points typically having the same sign. This shows that although we get a decent fit with the model, there is short term variation not described by it.

	Prediction	σ	2.5 % PI	97.5 % PI	Observed
Y2018M1	408.21	0.76	406.72	409.71	407.96
Y2018M2	410.08	0.80	408.51	411.64	408.32
Y2018M6	410.83	0.82	409.22	412.44	410.79
Y2018M12	409.00	0.83	407.37	410.62	409.07
Y2019M8	410.64	0.94	408.81	412.48	409.95

Table 5: Predictions from Local Linear Trend model using $\lambda = 0.9$.

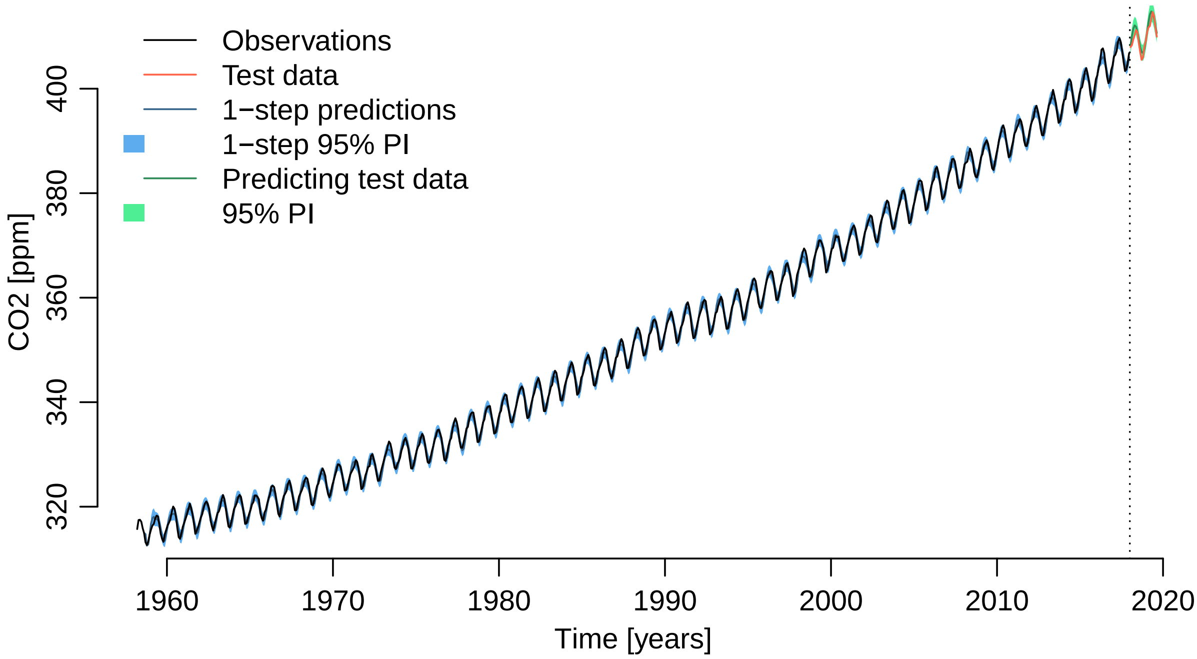


Figure 2: Predictions from local linear trend model using $\lambda = 0.9$.

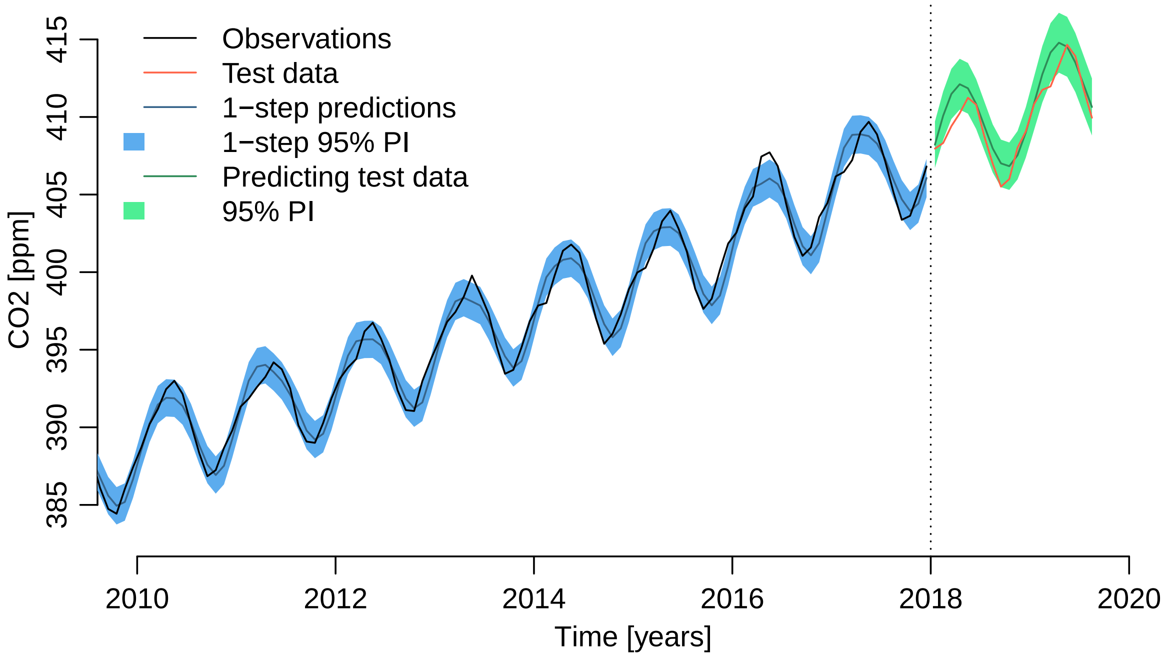


Figure 3: Zoomed plot of local linear trend predictions with $\lambda = 0.9$.

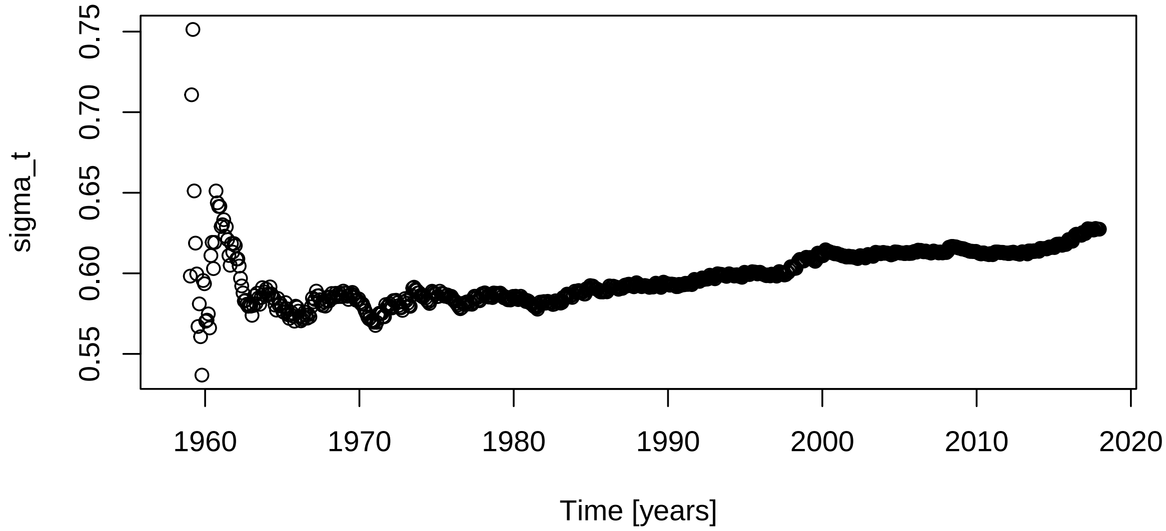


Figure 4: Estimated standard deviation over time.

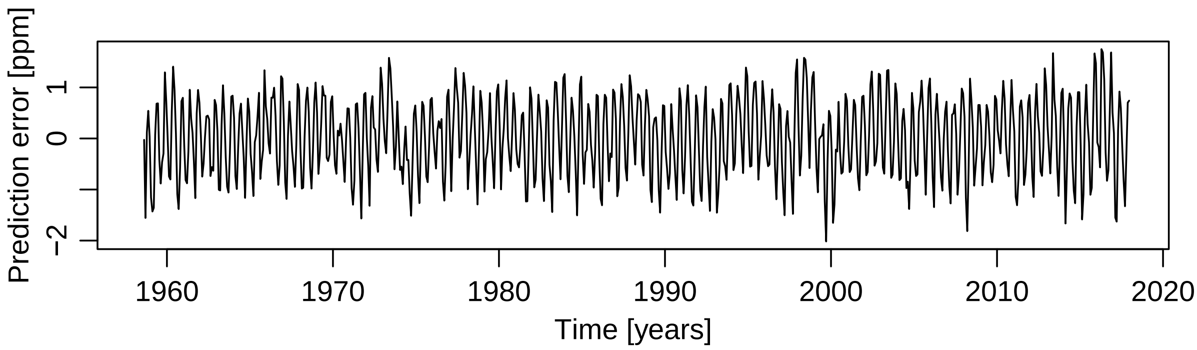


Figure 5: 1-step prediction errors of local trend model.

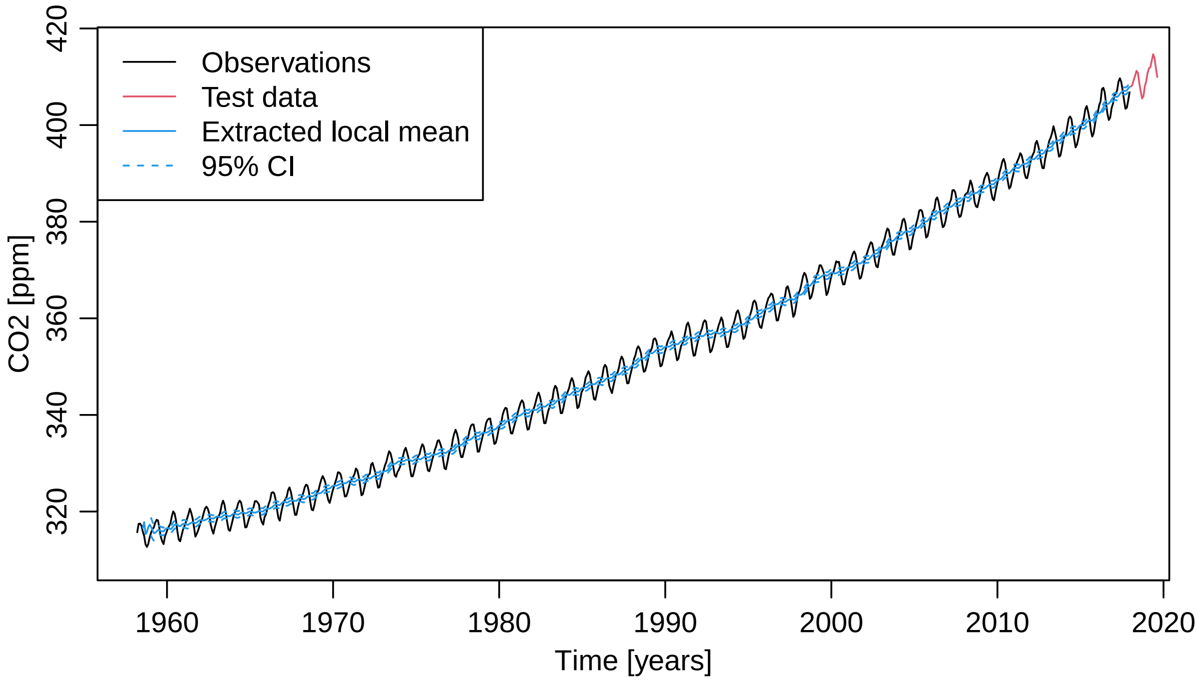


Figure 6: Estimated mean value over time.

Table 5 shows the prediction for the specified horizons. It is seen that the observation 2 months ahead is just outside the 95% PI - the remaining observations are well within the prediction intervals.

Finally, Figure 6 shows the estimated mean value as a function of time. This series is one way to make seasonally adjusted data, where this estimated mean is subtracted from the data points to obtain what is essentially trend-less data. It is worth noting that there is a rather small uncertainty for this estimate. The estimated uncertainty is calculated using the inverse of F_N at all time points and the associated estimate of σ. (With this many observations) T quickly converges to 10 and four parameters are estimated making a t-distribution with df=6 a conservative estimate. Since this is only a small amount of parameters of freedom, assuming a normal distribution is quite wrong.

Question 1.4: *In the previous question you used a given λ. In this question you should find the optimal λ (minimizing squared one step prediction errors) with a burnin period of 100 months.*

1. *Make a plot with the optimal λ showing the data and predictions from 2010 and onwards (As in the previous question).*
2. *Make a table predicting the test data as in the previous question.*
3. *Comment on the results.*

Figure 7 shows the sum of squared 1-step errors for different values of lambda with a burn-in period of 100 months. The optimal value is found to equal 0.934. As is typical we see that having smaller values of lambda only reduces performance slightly, but larger values ruin the performance rapidly. Looking at the sum of squared errors for $\lambda = 0.9$, we see that it is only marginally worse than the optimal value of λ, so we do not expect that our predictions will be much different, let alone much better than what we saw previously.

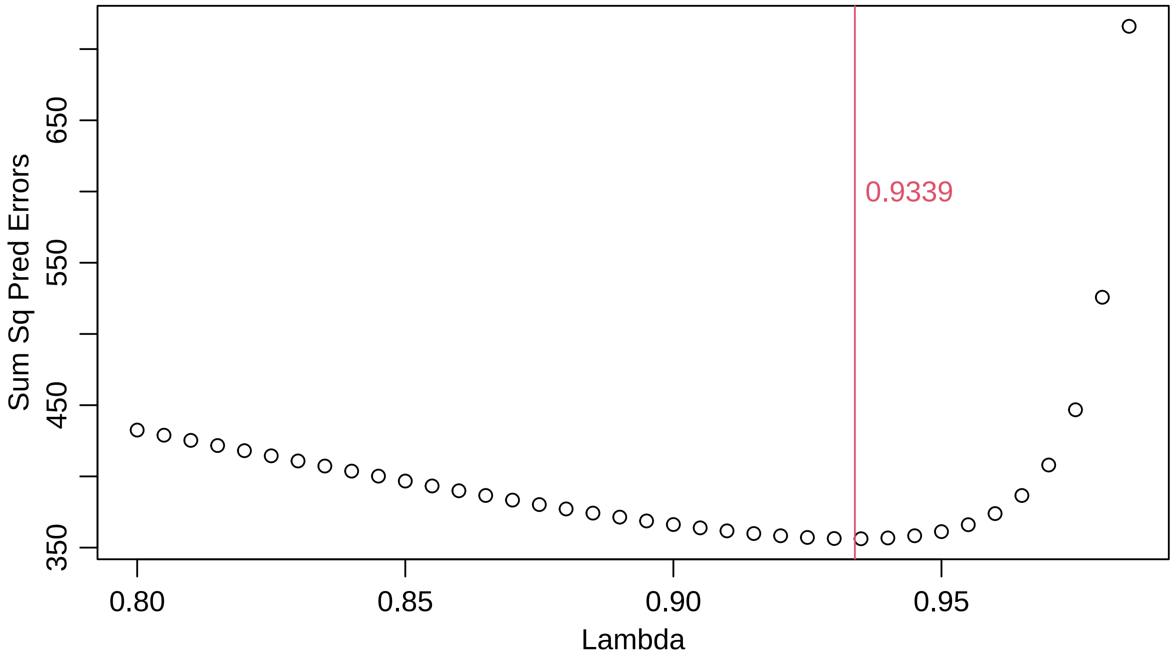
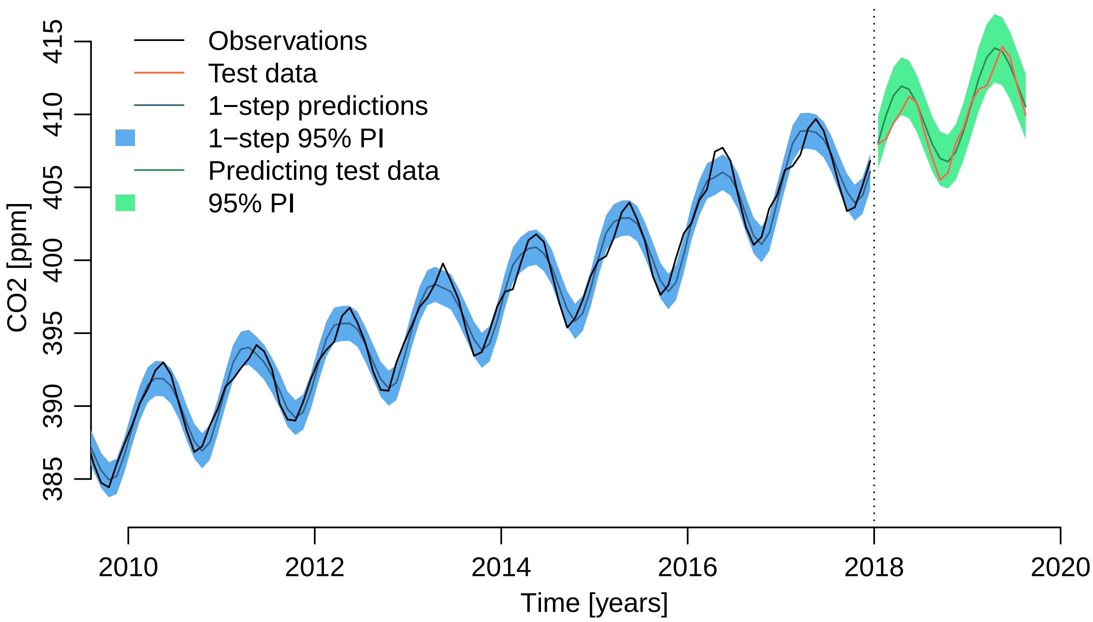


Figure 7: Sum of squared 1-step prediction errors for different values of λ.



	Prediction	σ	2.5 % PI	97.5 % PI	Observed
Y2018M1	408.07	0.92	406.26	409.87	407.96
Y2018M2	409.90	0.96	408.01	411.79	408.32
Y2018M6	410.74	0.99	408.79	412.69	410.79
Y2018M12	408.83	1.00	406.87	410.80	409.07
Y2019M8	410.54	1.13	408.32	412.76	409.95

Table 6: Predictions of local linear trend model using optimal value of λ

Table 6 shows the predictions from the local linear trend model using the optimal value of λ. The main difference between using $\lambda = 0.9$ and the optimal value is that the estimated σ is a little larger which is clearly seen from the table where all five observations are within the prediction intervals.

Question 1.5: Over all *Which model do you prefer and why?*

Suggest extensions to the model and how you find they can improve the model.

The assumptions for the OLS and WLS models are not fulfilled so they should not be used as is. Even with better predictors the assumption of independent residuals for the OLS model will most likely be violated.

For short term predictions the local trend model with the optimized λ should be prefered. And "short term" means not too much further ahead than the effective memory which here is about 15 months.

Suggested extensions:
The WLS model has a problem capturing the change in the trend. A second order polynomial in time does improve the fit a lot.

When having a closer look at the prediction errors then oscillations are seen with a period of six months. Therefore, it is worth testing higher harmonics in the Fourier series, e.g. p=6 months. This applies both to the WLS and the local trend model.

Other climate variables could be included in the WLS model - but not in the local trend model - however, we'll return to this when we look at Adaptive Recursive Least Squares.