02417: Time Series Analysis
Rune Grønborg Junker
Spring 2022

DTU Compute
Technical University of Denmark

Assignment 1: Forecasting temperature anomalies

Global warming is a great concern around the globe. Scientists at the Climatic Research Unit (CRU) at University of East Anglia have made several data sets with temperatures on the globe covering the past +150 years. The temperatures are expressed as anomalies from 1961-90 and in this assignment the focus will be on the changes in the average annual anomalies for the Northern hemisphere. The temperatures are estimated based on a number of measurement stations and do include measurement errors.
The data is provided in `A1_annual.txt` and includes three columns:

year Year for the observations

sh Temperature anomality for the Southern hemisphere (Not used)

nh Temperature anomality for the Northern hemisphere

You should not use the observations from 2014 through 2018 (Last five observations) for estimations - only for comparisons.

Question 1.1: *Plot the temperature anomalies for the Northern hemispheres as a function of time. Do indicate which data is used for training and testing. Comment on the evolution of the temperature over time.*

Figure 1 shows the data of the measured annual temperatures. It seems as though three distinct periods can be observed, as indicated by the vertical dashed lines. Until around 1920 the temperature seems steady, then between 1920 and 1970 an excursion is observed where the temperature increased steadily and then returned to a level close to before the excursion. Finally, for the last period, including the present, a clear linear trend of increasing temperatures is observed. This could adequately be called global warming and might be a consequence of increased CO2-levels. The test data seems to mostly follow the same pattern as the preceding training data (perhaps a little above), and so we expect decent performance when predicting it.

Question 1.2: *Use a global linear trend model to predict the temperature anomalies on the Northern hemisphere. Plot the data and the corresponding one step predictions for all observations in the training data. Plot the one step prediction errors. Make a plot and a table with the predictions for the five years that were left out - do include a 95% prediction interval. Compare with the test data. Comment on the results.*

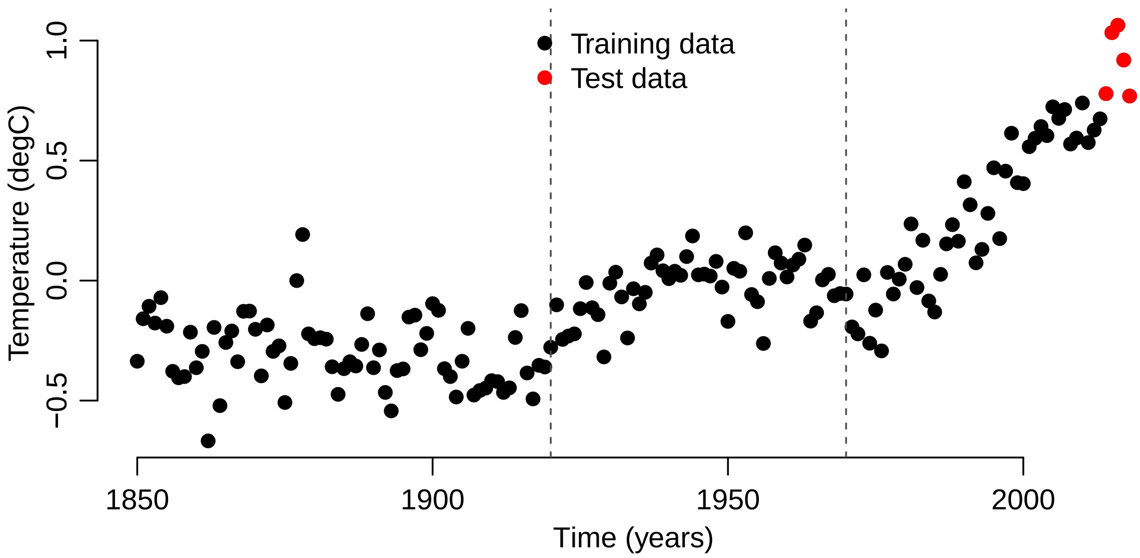

Figure 1: Annual mean temperatures measured as difference to reference year 1961-1990.

Although we just observed that the data does not all follow a linear trend, our teacher has told us to make a global linear trend model, and so this is what we do. The model is thus on the form

$$Y_t = \alpha + \beta t + \varepsilon_t,$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ i.i.d. Notice that the parameters are re-estimated whenever a new observation is available using ordinary least squares, but that it is assumed that there are true constant parameters and that hopefully as the amount of available data increases, so does the proximity to these. The model is then used to produce 1-step predictions which can be seen alongside the data in Figure 2. There are no predictions for years 1850 and 1851 as two observations are needed to make a prediction - this holds throughout the document.

We observe a number of things on this figure:

1. Until around year 1990 the model seems to fit the data quite well, in the sense that most observations are within the prediction intervals.
2. As will be further commented on below, the excursion in temperature between 1920 and 1970 is not captured by the model. Having this many observations above the model in a row makes it obvious that this could have been captured by a more appropriate model.
3. The nature of the global model means that in the beginning only a few observations are used to fit the parameters, while in the end way more observations have been used. For this reason the model is very non-smooth in the beginning but quick to adapt, while in the end it is very smooth but also only adapts very slowly to the data.
4. For the reason stated in 3) the model does not capture the last part of the training data well, and does even worse for the test data where all observations are above the prediction intervals.

Figure 3 shows the residuals of the global model. We would like these to show no patterns at all, but we see a clear pattern especially starting from year 1920. This pattern is

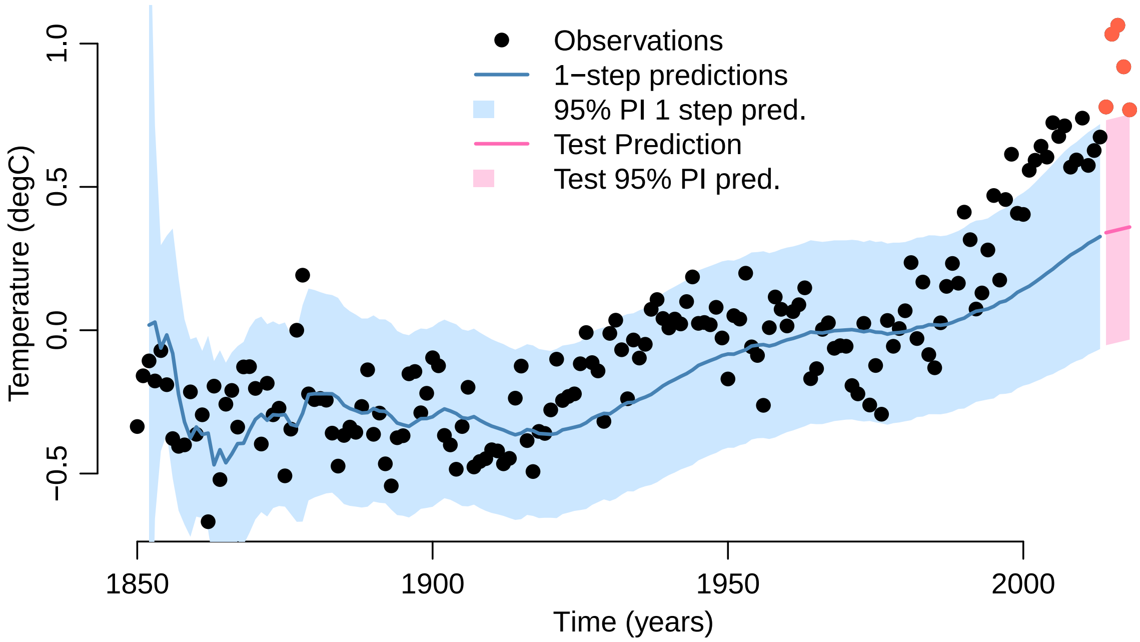

Figure 2: 1-step predictions from global model.

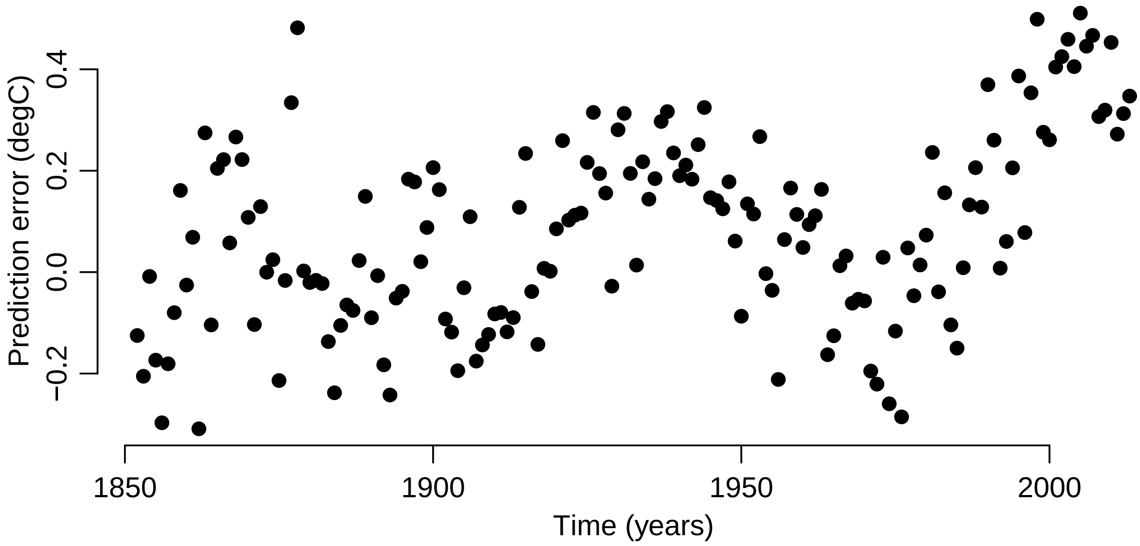

Figure 3: 1-step residuals from global model.

obvious enough that it should be able to be incorporated in the model. Therefore, we are not satisfied with the model.

For good measure we include a table of the predictions of the test data as well, in case people need the specific numbers for something. For example as an input to some other algorithm. This can be seen in Table 1. Notice that the standard deviation of the errors have been estimated using OLS as $\sigma = 0.1964$, while the standard deviation in the table is slightly larger since it also includes the uncertainty coming from the parameter estimates.

Table 1: Prediction of the Test data using the global model.

	Prediction	Std. Dev.	2.5 %	97.5 %	Observed
2014	0.3402	0.1988	-0.0523	0.7327	0.7790
2015	0.3451	0.1988	-0.0475	0.7377	1.0330
2016	0.3501	0.1988	-0.0426	0.7427	1.0640
2017	0.3550	0.1989	-0.0378	0.7478	0.9190
2018	0.3599	0.1989	-0.0329	0.7528	0.7690

Question 1.3: *Use a local linear trend model to predict the temperature anomalies on the Northern hemisphere using $\lambda = 0.8$. Plot the data and the corresponding one step predictions for all observations in the training data. Plot the one step prediction errors. Make a plot and a table with the predictions for the five years that were left out - again including a 95% prediction interval. Compare with the test data. Comment on the results.*

We saw that the global trend model was not able to adapt to the data in a convincing matter. To remedy this we try a local trend model instead with a forgetting factor of $\lambda = 0.8$. Notice that this is a rather small value leading to a total memory of only around $\frac{1}{1-\lambda} = 5$, meaning that it only uses around 5 observations to fit the trend. Figure 4 shows the 1-step predictions along with predictions of the test data using a linear local trend model. Compared to the global model, we see, not surprisingly, that the trend is much less smooth, since the effective amount of used observation is much lower. However, the predictions are tracking the observations much better with no obvious bias. The prediction seems wider although this might partly be due to the much more varying tendency of the predictions, making the interval simply look wider than it actually is.

Figure 5 shows the 1-step prediction errors which seem a lot more independent than before. It might be that the variance is larger in the beginning of the period. For this reason it might be preferred to use the local estimator of the variance, although using the global estimator is also reasonable. Here both estimates are provided: $\sigma_{local} = 0.0978$ and $\sigma_{global} = 0.1163$ (Using observations 3 to 164). Notice that actually these are quite a bit smaller than for the global model, already indicating that the local model is better. These lead to the following predictions and 95% prediction intervals given as quantiles. Notice how the local estimator provides larger prediction intervals than the global estimator of σ, even though the variance is estimated to be smaller using the local method. This is due to the fewer degrees of freedom in the student's t-distribution when using the local estimator of the variance.

The results are also presented using the local estimator of σ in table 2 and the global estimator in table 3.

The local trend model captures the test data better, since it includes the positive trend in the end of the training data. However, like the global model, it underestimates the test data. At least the prediction interval based on the local estimator of σ includes all 5

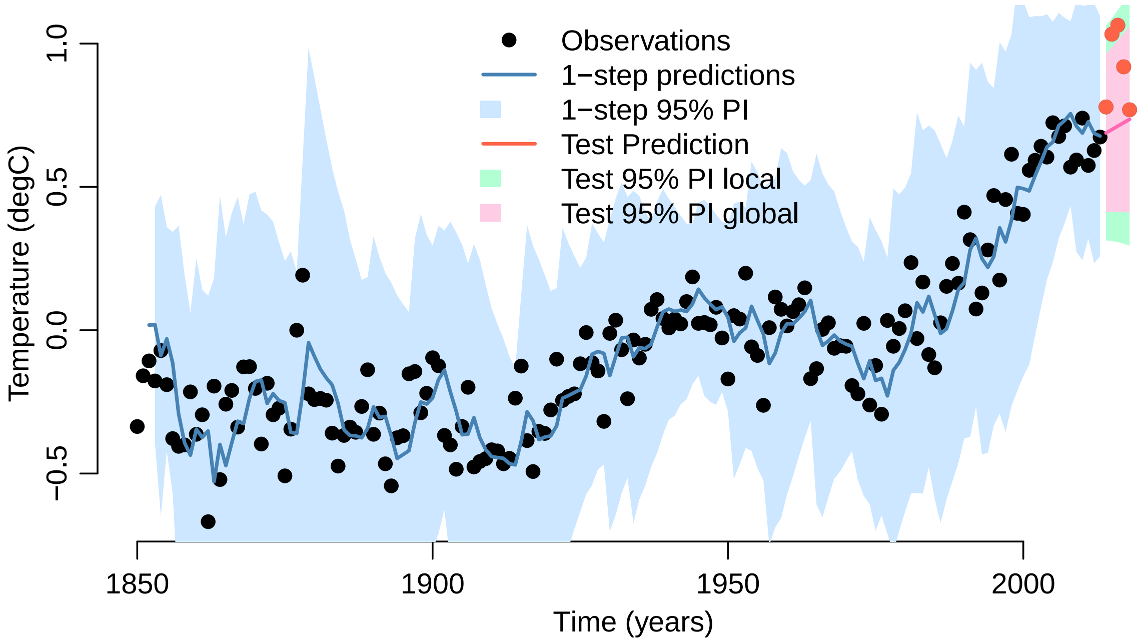

Figure 4: 1-step predictions of local trend model.

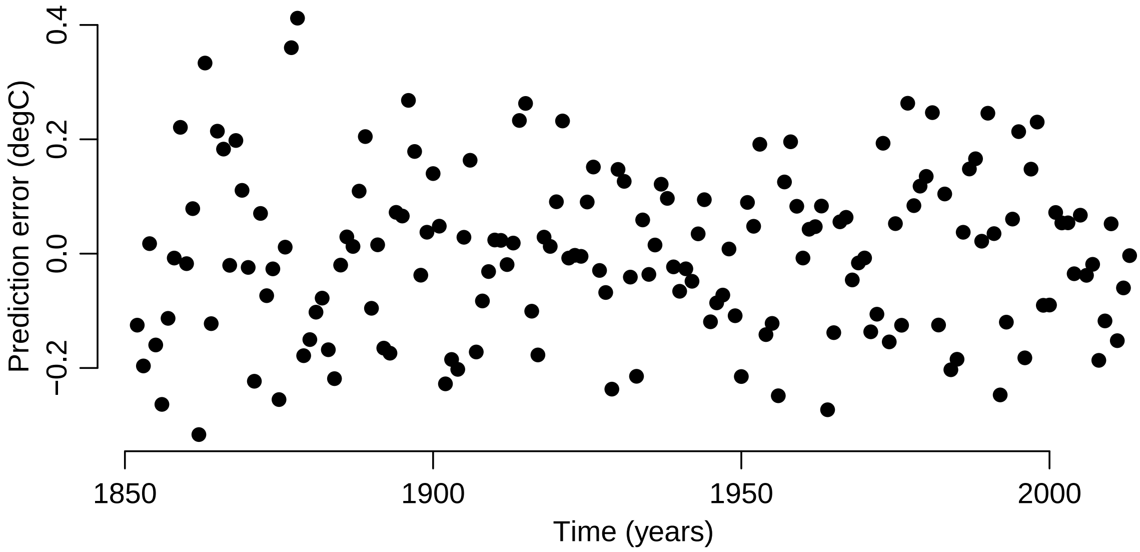

Figure 5: 1-step residuals of local trend model.

Table 2: Prediction of the Test data using the local model with a local estimate of σ.

	Prediction	Std. Dev.	2.5 %	97.5 %	Observed
2014	0.6882	0.1178	0.3133	1.0630	0.7790
2015	0.7001	0.1222	0.3113	1.0889	1.0330
2016	0.7121	0.1272	0.3074	1.1167	1.0640
2017	0.7241	0.1327	0.3018	1.1463	0.9190
2018	0.7360	0.1387	0.2947	1.1773	0.7690

Table 3: Prediction of the Test data using the local model but a global estimate of σ.

	Prediction	Std. Dev.	2.5 %	97.5 %	Observed
2014	0.6882	0.1401	0.4115	0.9648	0.7790
2015	0.7001	0.1453	0.4131	0.9871	1.0330
2016	0.7121	0.1513	0.4134	1.0108	1.0640
2017	0.7241	0.1578	0.4124	1.0357	0.9190
2018	0.7360	0.1650	0.4103	1.0618	0.7690

observations, while the prediction interval based on the global estimator includes only 3. This is acceptable although we can hardly call the performance great.

Question 1.4: *Find an optimal value of the forgetting factor for use in the local trend model suggested in the previous question. (Optimize 1-step prediction errors. Do disregard the first 5 1-step prediction errors as burn in period.)*

Present the performance with the optimal value as above.

The sum of squared one step prediction errors as a function of λ can be seen in Figure 6. The red point indicates the optimal value found through an optimiser, and equals $\lambda = 0.8444$. This is quite close the 0.8 that we used previously, so we should expect very similar performance using this value.

None the less, we re-run the local trend model using this value and obtain the one-step predictions shown in Figure 7. Visually it is hard to see any difference between this and the previous local trend model. However, for the test data we do see slightly better performance since the last 4 observations in the training data does not impact the estimated trend as much, and thus it is slightly larger for this model, resulting in a better capture of the test data. The residuals of this model look the same as for the previous one and there is nothing to gain from plotting them again.

The results are also presented using the local estimator of σ in table 4 and the global estimator in table 5.

Table 4: Prediction of the Test data using the local model with a local estimate of σ and $\lambda = 0.8444$.

	Prediction	Std. Dev.	2.5 %	97.5 %	Observed
2014	0.7103	0.1132	0.3499	1.0706	0.7790
2015	0.7261	0.1158	0.3576	1.0946	1.0330
2016	0.7420	0.1187	0.3643	1.1197	1.0640
2017	0.7579	0.1218	0.3701	1.1457	0.9190
2018	0.7738	0.1253	0.3751	1.1725	0.7690

Question 1.5: *Comment on the results. Which model do you prefer? Do you trust the forecasts? Do you have ideas for improving the forecast method?*

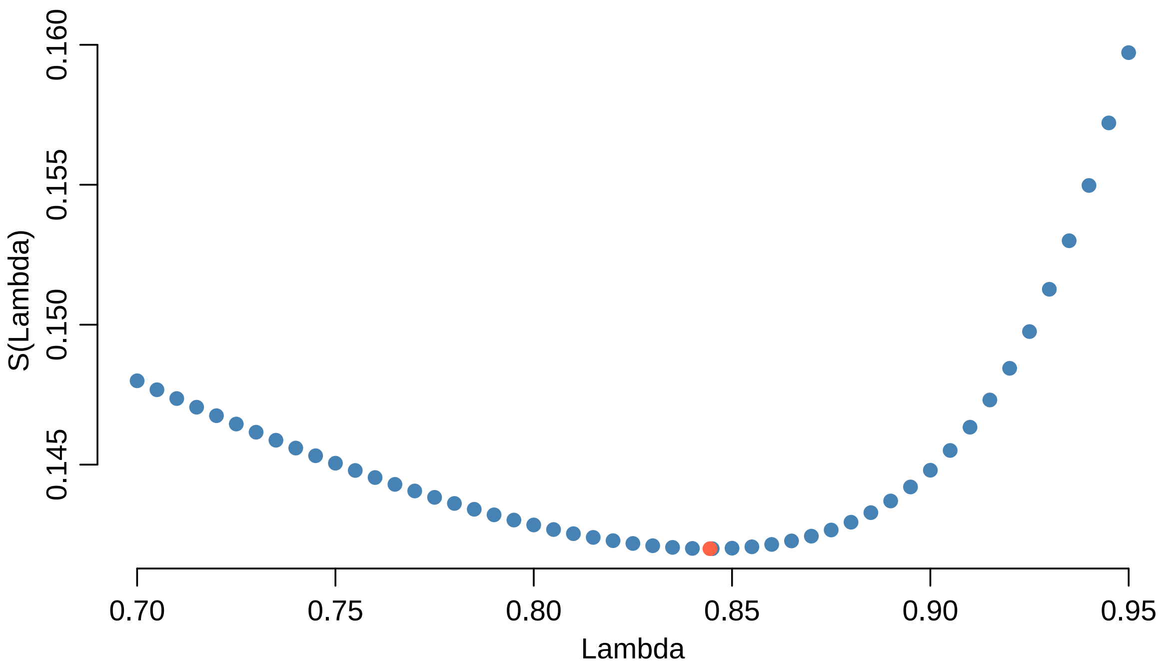

Figure 6: Sum of squared 1-step prediction errors as a function of λ.

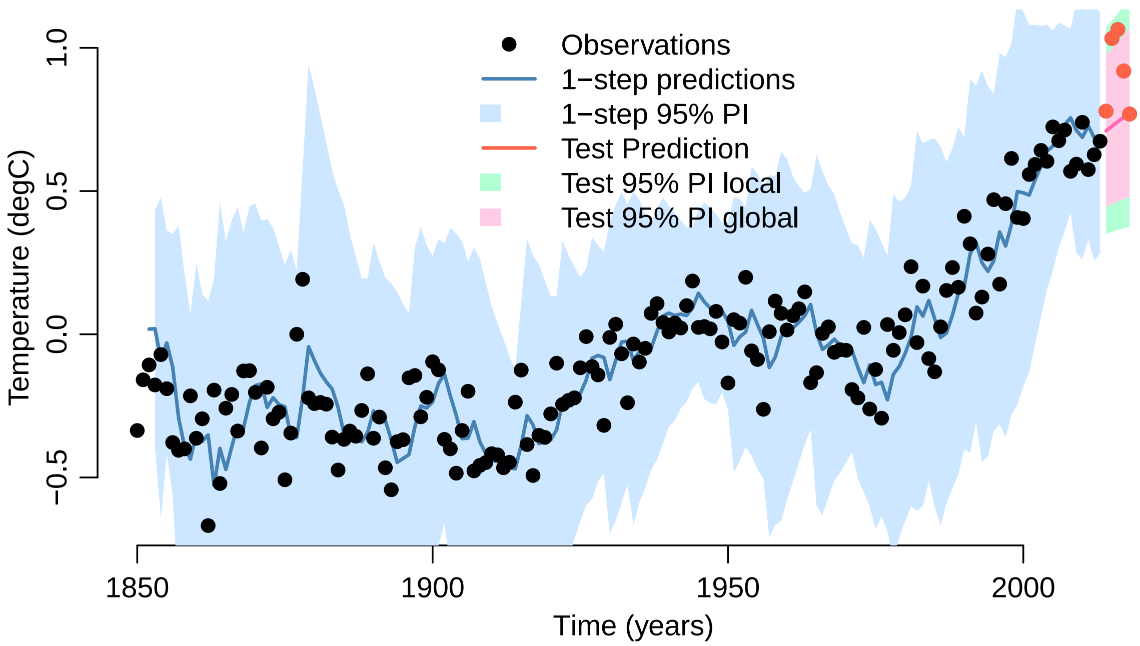

Figure 7: One-step predictions using $\lambda = 0.8444$.

Table 5: Prediction of the Test data using the local model but a global estimate of σ and $\lambda = 0.8444$.

	Prediction	Std. Dev.	2.5 %	97.5 %	Observed
2014	0.7103	0.1347	0.4443	0.9762	0.7790
2015	0.7261	0.1377	0.4541	0.9982	1.0330
2016	0.7420	0.1412	0.4632	1.0208	1.0640
2017	0.7579	0.1449	0.4716	1.0441	0.9190
2018	0.7738	0.1490	0.4795	1.0681	0.7690

Prefered models

Since the data does not seem to follow a single overall trend, the global model performed poorly, and I would never use it to forecast anything. The local models do a much better job, and for one-step predictions the one with optimised λ is to be preferred since it minimises the sum of square errors while not seeming to overfit the data or experience any other problems. For short term predictions this model is able to provide decent predictions and I would be content with providing these to a customer.

For prediction horizons greater than one time step λ should be re-estimated for each horizon of interest. This leads to a local linear trend model for tailored for each horizon. However, I would not expect the linear trend to increase for long, and thus for longer term predictions it seems more appropriate to use models based on physics to describe the temperature anomalies.

Improving the model

In this assignment a linear trend model is used. One should consider if other models could perform better. E.g. For the present data a global quadratic trend model might do a good job but it would probably (hopefully!) overestimate the future increase in temperature.

There may also be other indicators that can be used to predict the temperature anomalies, e.g. the concentration of green house gasses in the atmosphere. Therefore, incorporating some regressors might improve performance.

As already mentioned if multi-step predictions are required then the forgetting factor should be re-optimised. (*It is also fine to mention this here.*)