Signal Processing **Michaelmas 2014**

Time Series Forecasting

Student: Ankush Gupta

CO_2 Level Forecasting

I followed two approaches for forecasting this time series — (1) decomposing the series into long-term and short-term trends (2) using gaussian process regression.

Series Decomposition

Clearly, there are (at least) two trends in the given time series — (1) $f_L(t)$ Long term increase (2) $f_S(t)$ Short term almost sinusoidal trend. Hence, my analysis for this time series was aimed at finding these two components such that $f_{CO_2}(t) \approx f_L(t) + f_S(t)$

Fitting Long Term Trend

Since, many natural phenomena follow the exponential trend because of their dynamics being $\frac{dx}{dt} \propto \alpha x$, I first tried fitting an exponential curve $f_L(t) = \beta e^{\alpha t}$, by taking logs and using ordinary least squares. The fit didn't look right [1], so I tried decomposing it in terms of the polynomial basis $\{1, x, x^2, x^3, ...\}$ using least-squares.

To get rid of the sinusoidal trend, I used a low-pass filter to uncover the long-term trend [2] (more specifically, I used a Butterworth filter with the cut-off frequency determined heuristically and then applied the filter using `filtfilt` to get rid of phase-shift effects). Then I found the coefficients by regressing the curve on the polynomial basis. Using cross-validation, it was found that degree 2 basis basis $\{1, x, x^2\}$ gave the best fit and *generalisation* into the future.

In conclusion, the long term trend was determined to be $\boxed{f_L(t) = a_0 + a_1 x + a_2 x^2}$

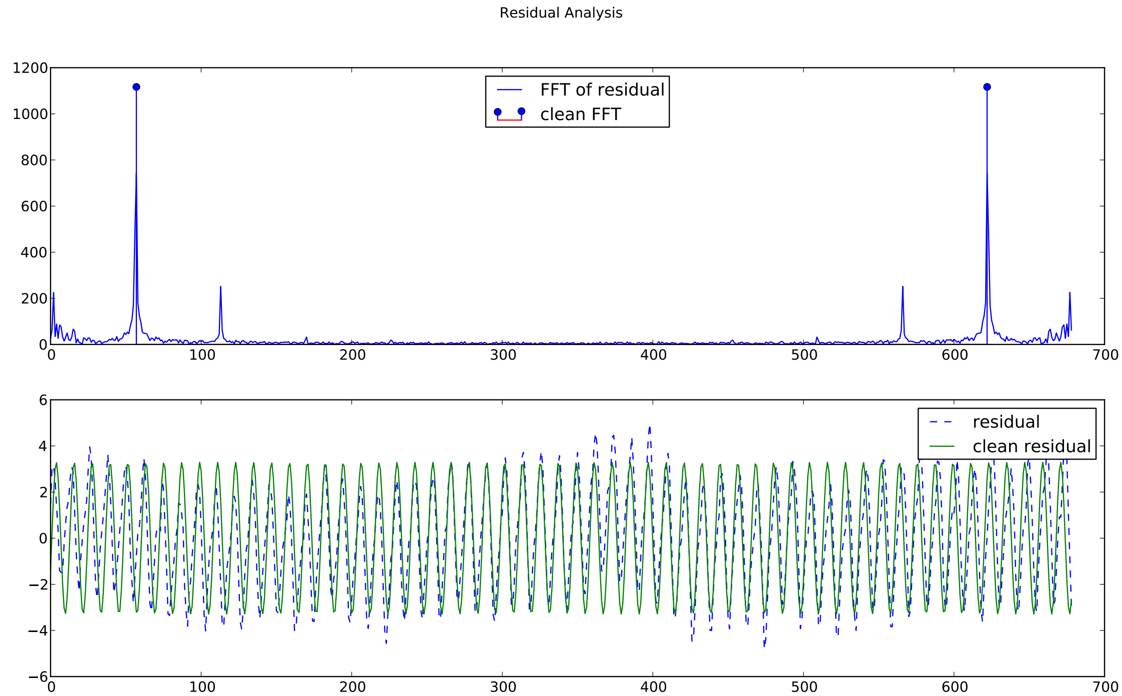
Fitting the Short Term (almost) Sinusoidal Trend

The short-term trend was modelled as a pure sinusoidal i.e., $f_S(t) = f_{CO_2} - f_L(t)$ was modelled as a sine wave of **one** frequency. This decision of using just one frequency was based on analysing the spectrum of the residual $f_{CO_2} - f_L(t)$ (Figure 1.1a). As the maximum peak is relatively bigger than the rest of coefficients, only one frequency was used to model the residual.

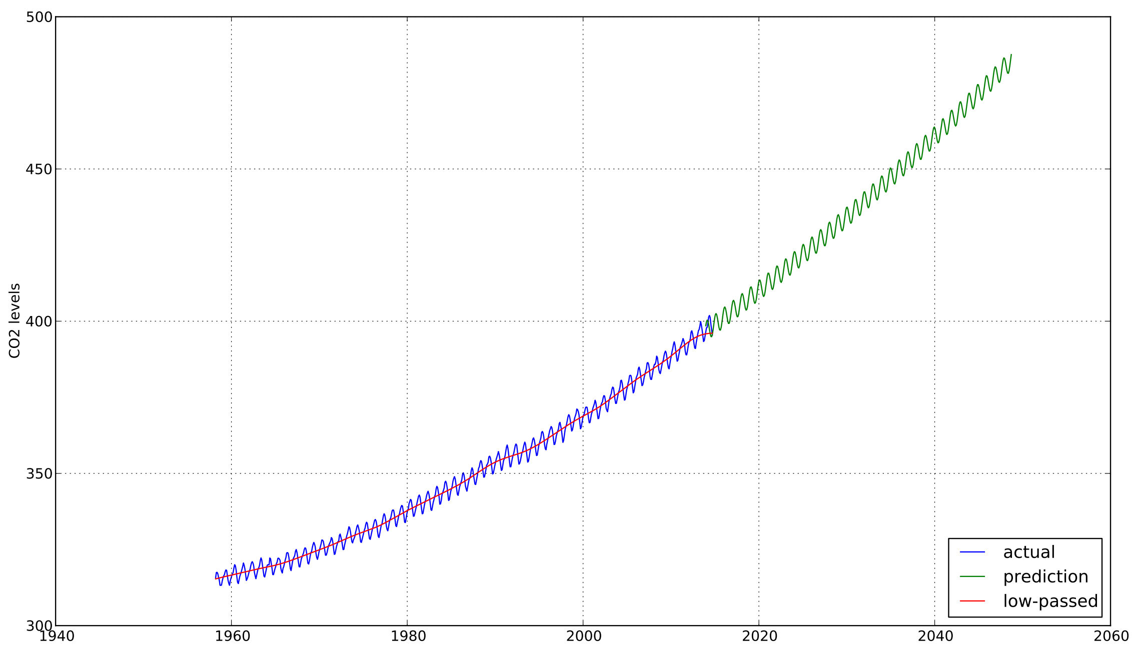
Figure 1.1b shows the final result — predictions up to the year 2050. Also shown is the low-passed version of the original signal used for fitting the long-term trend.

[1] which was quite surprising for me.

[2] Note: I could have used the raw-curve (without any pre-processing/ filtering) to find coefficients of the polynomial basis elements.



(a) Spectrum of the residual (short-term trend) = $f_{CO_2}(t) - f_L(t)$.



(b) Prediction of CO_2 time series data.

Gaussian Process Regression for CO_2

The code written for the *Data Estimation and Inference* course was used here. The polynomial fit for the long-term trend was used as the mean-function. For the covariance, I analysed two Kernels — (1) Square-Exponential $= \sigma^2 exp\left(-(x_1 - x_2)^2/2a\right) + \sigma_y^2 I$ and, (2) Periodic $= \sigma^2 exp\left(-l^2 \sin^2(2\pi/b|x_1 - x_2|)\right)$.

Figure 1.2 shows the results (please see the figure caption for discussion).

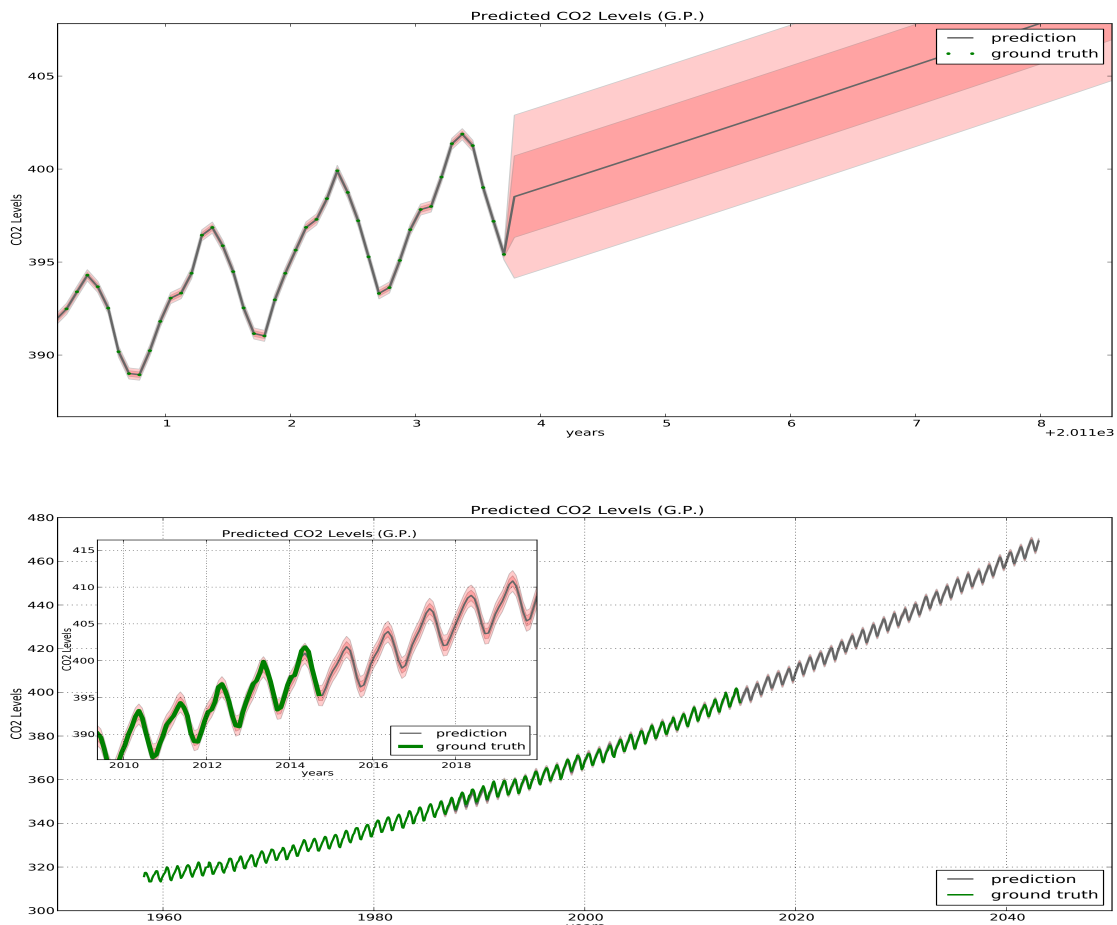


Figure 1.2: [Top] GPR using the square-exponential kernel. Note the high-uncertainty in prediction. The future predictions basically just track the mean function. [Bottom] GPR prediction using periodic kernel. Since this periodic kernel captures the periodicity in the data, the future predictions have low uncertainty (inset). There is some finite variance at the points where observations are available because an observation uncertainty term was added to the kernel to make it numerically robust. The two red intervals indicate regions of one and two standard deviations.

In this lab we investigate using Gaussian Processes for regression on weather data prediction. Following are the main features of this implementation.

- The code was implemented in Python.
- Two mean functions were analyzed : (1) constant mean = mean of the data (2) cubic spline fit to the data.[3]
- Two covariance functions were analyzed :
  1. Squared-Exponential $= \sigma^2 exp\left(-\frac{(x_1 - x_2)^T M^{-1}(x_1 - x_2}{2}\right) + \sigma_y^2 I$
  2. Periodic $= \sigma^2 exp\left(-l^2 \sin^2(2\pi/b|x_1 - x_2|)\right)$
- The hyperparameters were tuned by maximizing the marginal likelihood. Analytical gradients were used with the conjugate gradient optimization method.
- Sequential predictions were also carried out for Tide Height and Temperature time series.

Please note all figures indicate intervals of 1 and 2 standard deviations.

RMS Error with Different Means/ Covariances

The table below indicates the root-mean-squared error of the **tide-height** predcitions with respect to the ground-truth data. The covariance hyper-parameters used for these experiments were optimised through marginal likelihood as detailed in the next section.

Mean \Covariance	Squared-Exponent	Periodic
Mean(data)	**0.158**	0.175
Spline	0.240	0.290

Table 1.1: RMS Error for tide-height predictions.

[3]Note that these are strictly not bayesian methods as we are using the data itself to specify the prior.