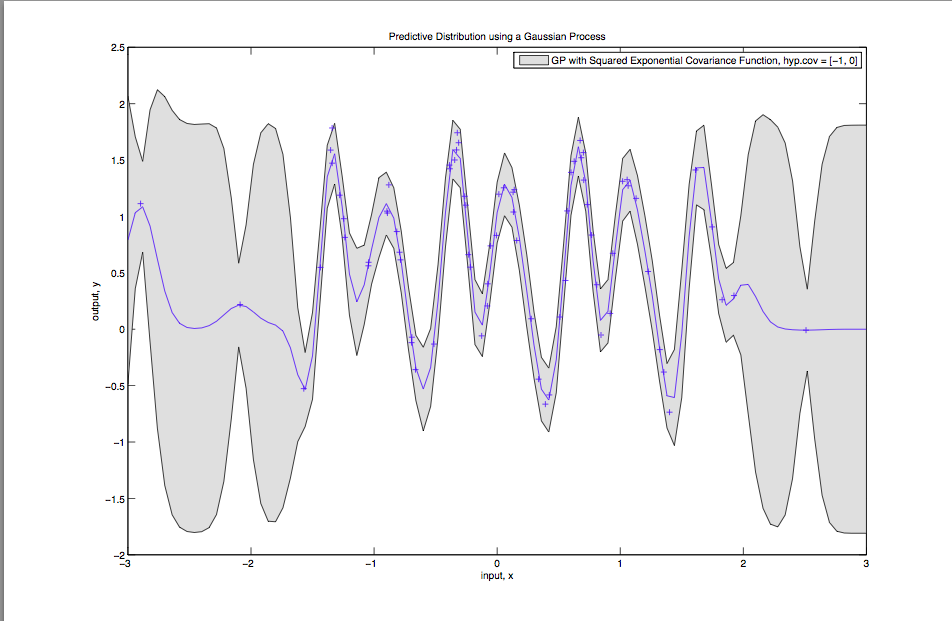
**Question A**

The negative log marginal likelihood is 11.8990  
For the squared exponential covariance function, in GPML toolbox, the hyperparameters [-1 0] are encoded in logarithms.  
  
Initial hyperparameters:

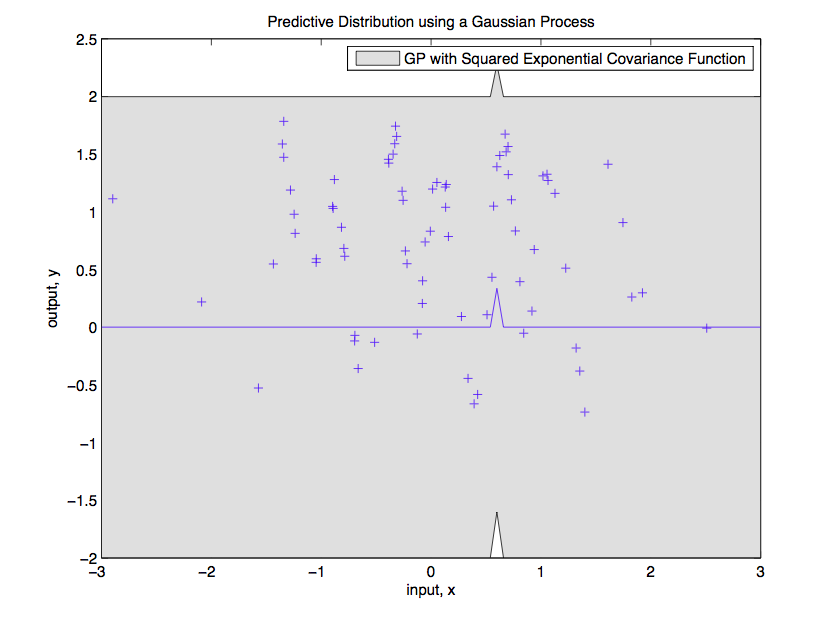
Length\_scale = 0.3679  
Signal Variance = 1

Optimized hyperparameters:  
Length\_scale = 0.1282  
Signal Variance = 0.8046  
  
The plot below is for the optimized mean and variance at the sample test locations, with the plot for the mean function plus/minus two standard deviations corresponding to the 95% confidence interval.

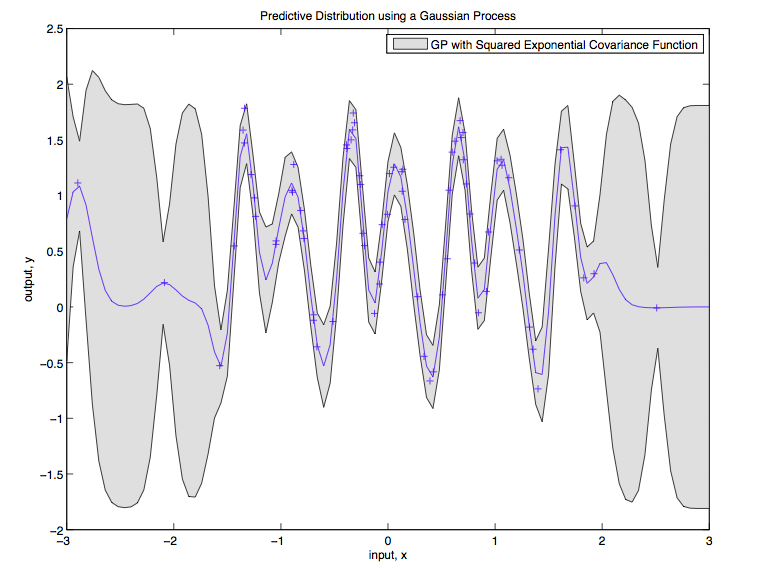
The error bars (confidence interval) shows that in regions where there are less test points, the GP function is more uncertain about the possible functions, hence more uncertainty in predictions. In regions of more points, the predictive distribution has reasonable confidence, with a certain level of uncertainty. The model is quite different from the data generating process in regions of few test points, whereas there is a much better fit in regions of more test data.   
  
The initialized hyperparameters hyp.cov=[-1 0] for the squared exponential covariance function is too high to fit the data; therefore, the optimized hyperparamter values are reduced to make covSEiso suitable.   
  
  
  
*Figure 1*

**Question B**

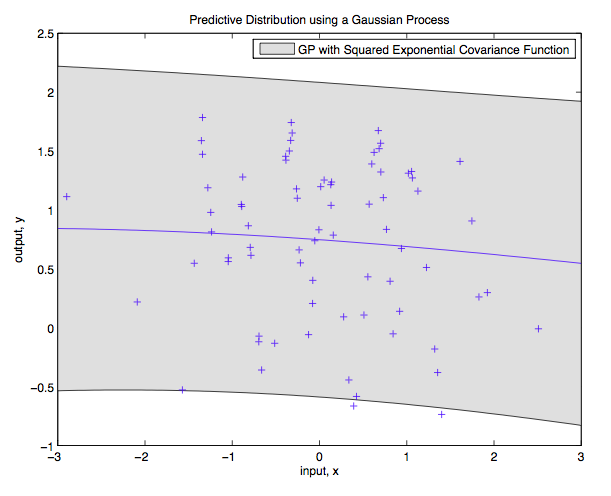
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| L init | L opt | Variance init | Variance opt | NLML |
| 0.3679 | 0.1282 | 1 | 0.8046 | 11.899 |
| 4.54e-5 | 4.54e-5 | 1 | 0.4991 | 106.3493 |
| 0.1353 | 0.1282 | 1 | 0.8046 | 11.899 |
| 1 | 8.3488 | 1 | 0.4884 | 78.2203 |
| 1.6487 | 7.1644 | 54.5982 | 0.4562 | 78.2217 |
| 20.0855 | 8.0421 | 1 | 0.4843 | 78.2207 |

The table above shows different local optimum values for the hyperparameters dependent on the initialization.   
Figures 2 and 4 shows that the model is a poor fit (model quite differnet from the generating process) indicated by the shaded regions of the graph (high uncertainty in predictions). Row 2 and 4 corresponding to these graphs shows that if the length scale initialization is very small or very high, then the negative marginal likelihood is very high (indicating a very low confidence), with the model fitting the data very poorly. Figure 3 shows that when length scale and variances are initialized to reasonable values, the NLML is not too high or low. As shown in table above and the figures below, notice how the optimized hyperparameter values are significantly different and dependent on initializations of parameters.   
  


*Figure 2: Corresponding to Row 2 of table above (L\_opt = 4.54e-5, Var\_opt=0.4991) and NLML = 106.35*



*Figure 3: Corresponding to Row 1 of table above (L\_opt = 0.1282, Var\_opt=0.8046) and NLML = 11.899*



*Figure 4: Corresponding to Row 4 of table above (L\_opt = 8.3488, Var\_opt=0.4884) and NLML = 78.2203*

It can be concluded that initalizations of hyp.cov=[-1 0] or hyp.cov=[-2 0] are reasonable hyperparamter initial values, giving a much better fit to the data, compared to initial values being too high or too low.

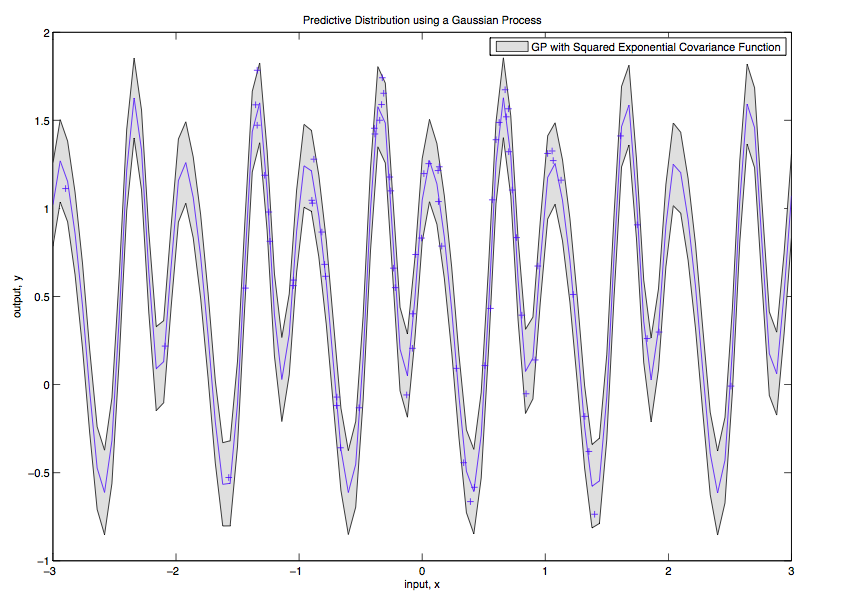
**Question C**

Figure 5 below is for the following hyperparameter values

Length\_scale init = 1.2840

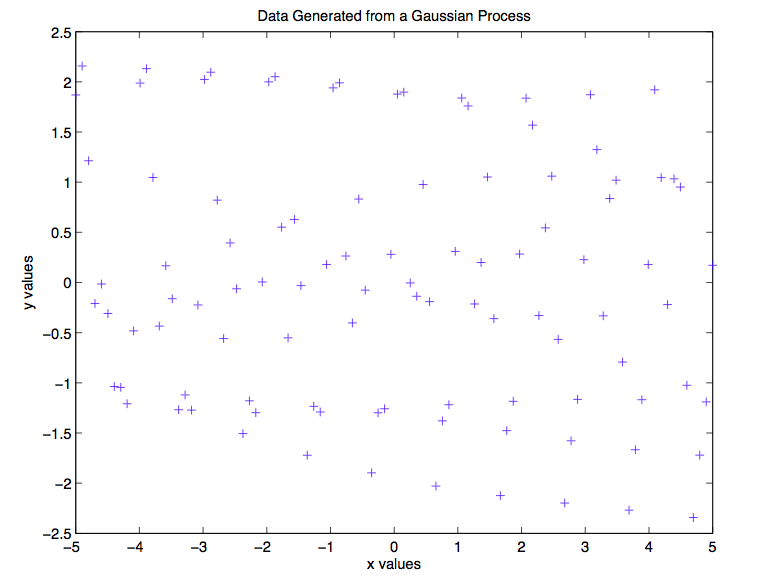
Length\_scale opt = 0.9397

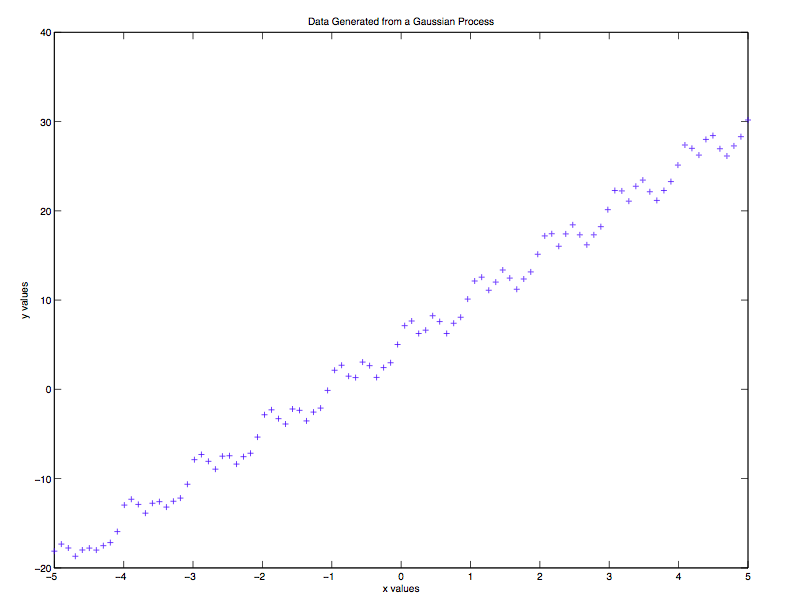
Variance\_init = 7.3891  
Variance\_opt = 0.9675  
NLML = -35.1724

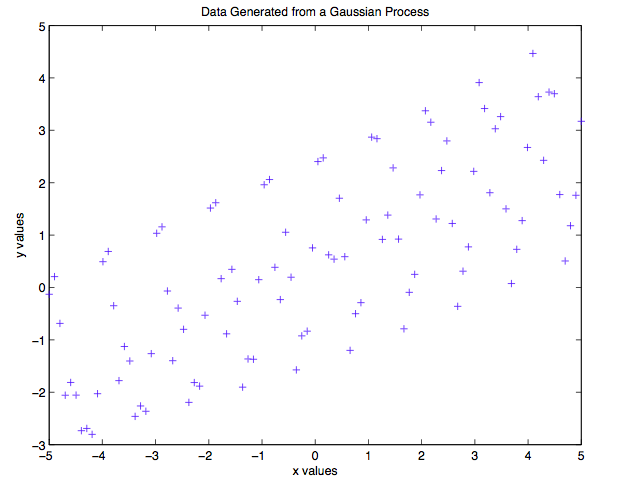


The 95% confidence interval (error bars) are much lower for the periodic covariance function, compared to the squared exponential.   
The periodic covariance function fits the model very well, with high confidence, suggesting that the data generating mechanism is periodic. However, this is not supported by the NLML, having a very small value, which shows that too high and very confident fit to the data is not supported by the marginal log likelihood.  
  
Compared to the fit in Question A, the model fits with too high confidence with very low uncertainty even in regions of less test points. Therefore, it is overfitting the data. Comparing the NLML values with Question A, the NLML value suggests that the data generating mechanism is in fact not suppose to be periodic.

**Question D**

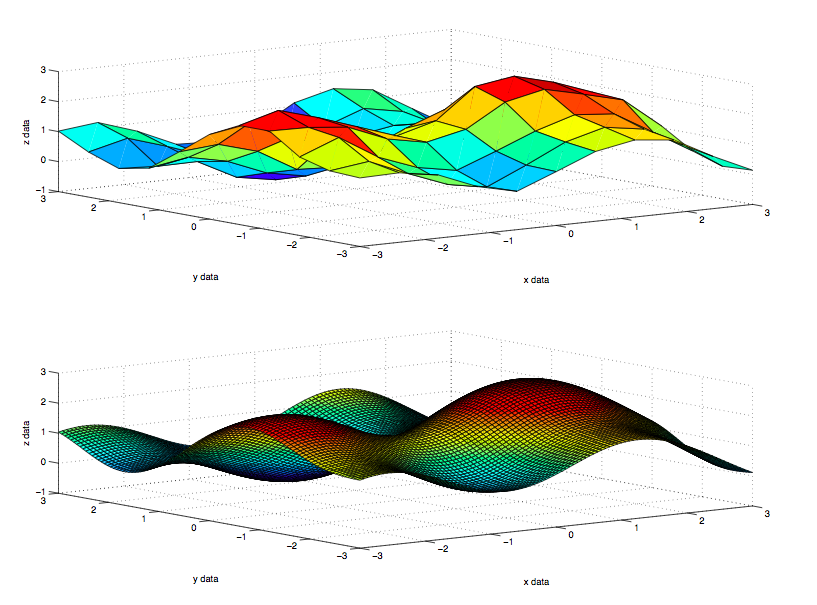
We need to add the small diagonal matrix because when doing the negative log marginal likelihood, or generate data points from a GP, we need to enforce constraints that the covariance matrix must be positive definite.  
  
The plot below shows a sample function when no mean is added to the y generated poins  
  


The graphs below are for plots when we add a mean function.  
  
On adding a very high mean to the generated data points  


On adding low mean values to the generated points  
  


**Explain the behavior of the above functions**

**Question E**

****

The graph shows a GP fit of the cw1e.mat data using the squared exponential covariance function with Automatic Relevance Determination.  
  
The plot shows that   
Comment on the values of the hyperparameters here

Comment on the fit of the data

Question: How much noise is there in the data?

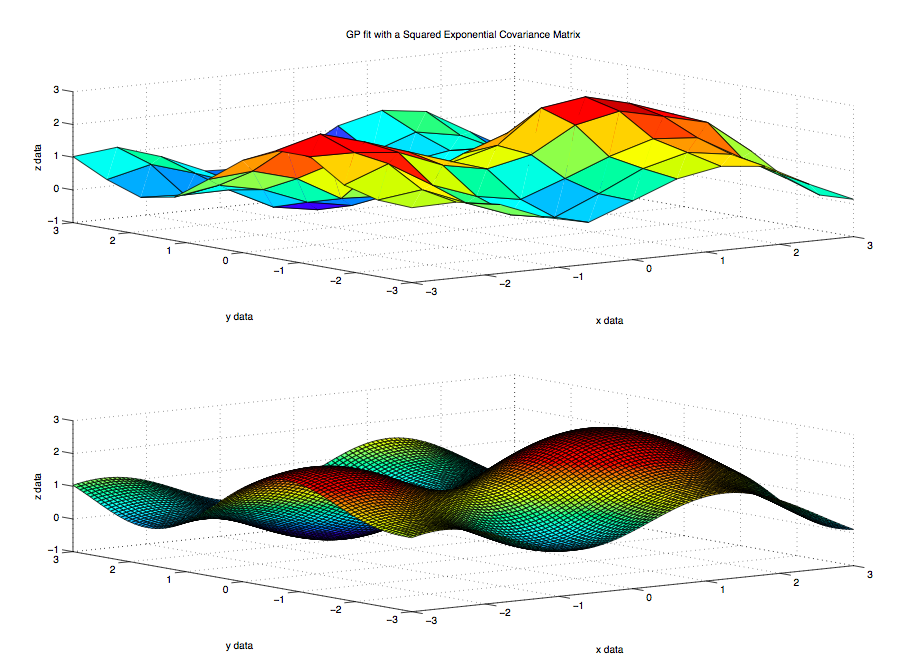
**Question F**

With covSEard, the marginal likelihood was -19.2187

With covSEiso, NLML is -18.07

Based on the marginal likelihood of the two model fits, the covSEiso GP fit behaves similarly as the one above.

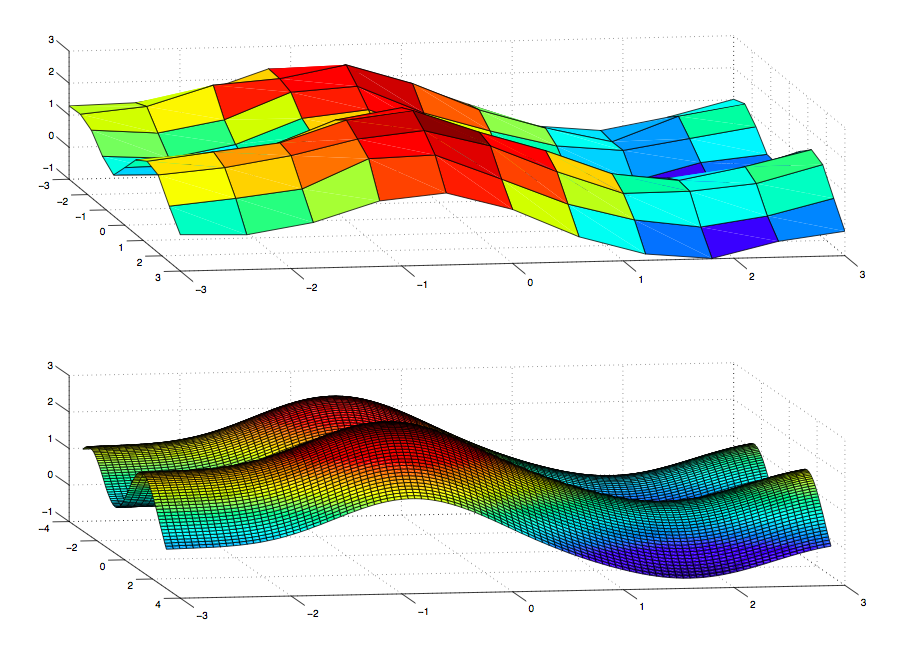
The relative probability is : -19.2187 – (-18.07) = -1.1496



**Question G**

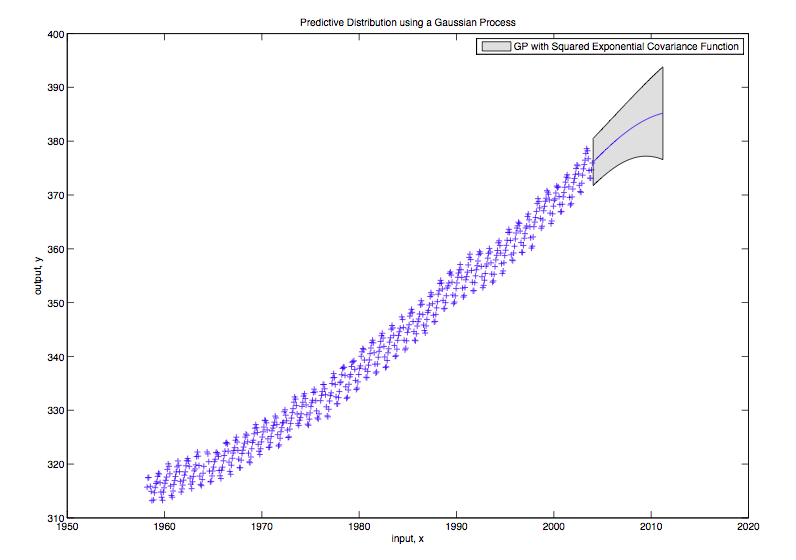
With the sum of covSEard covariance matrices, the model fit is extremely over confident, which is not suggested by the marginal likelihood.  
  
Compared to Q.e and Q.f, the NLML now is -66.33, which suggests that the model fit is not favored by the marginal likelihood, as the predictive distribution has very low uncertain (over confident).

Therefore, the model here is very different, even though from the graphs, significant differences cannot be ascertained.

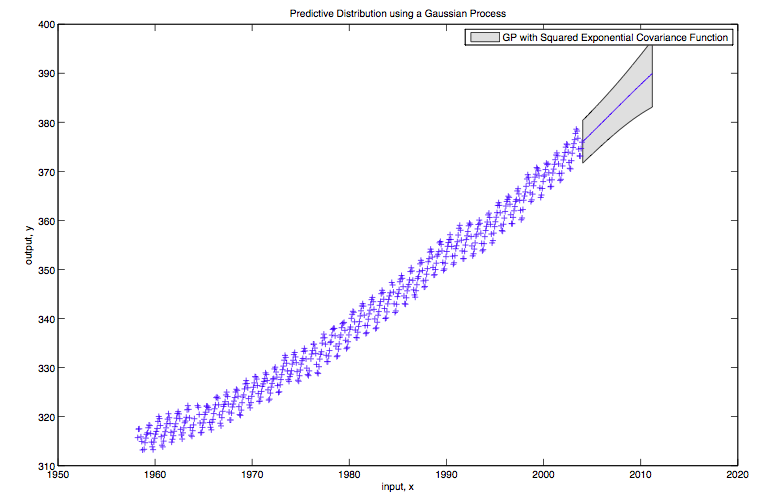


**Question H**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| L\_init | L\_opt | Var\_init | Var\_opt | Mean\_init | Mean\_opt | NLML |
| 1.6487 | 12.73 | 2.72 | 15.20 | [0.2 2] | [0.174 1.99] | 1992 |
| 1.6487 | 0.5099 | 2.72 | 10.512 | [5 10] | [0.1671 9.997] | 963.38 |
| 12.72 | 18.71 | 15.18 | 22.93 | [0.2 2] | [0.1761 1.9989] | 1198 |



*Figure corresponding to Row 1*



*Figure corresponding to Row 3*

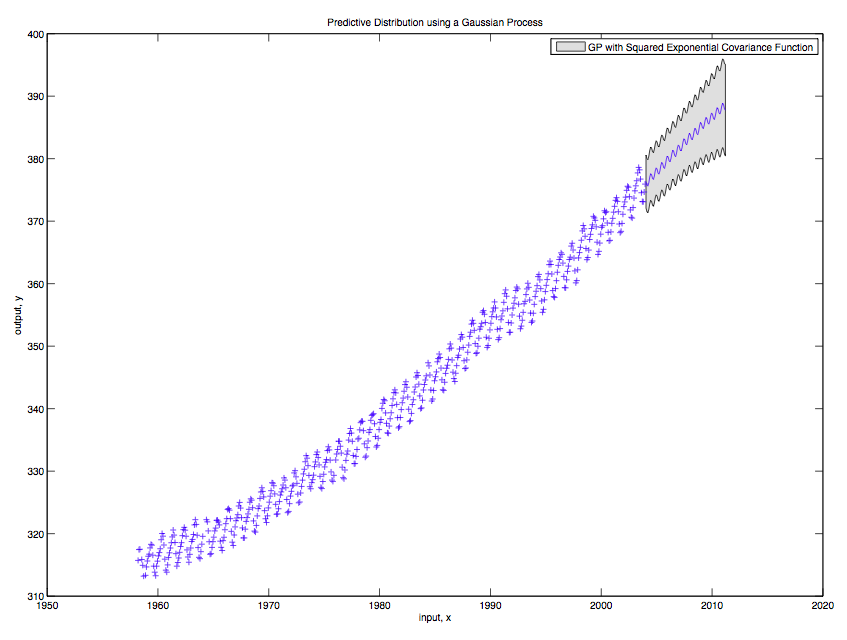
**Question I**

The following is a plot of a sum of a squared exponential and a periodic covariance matrix, with NLML=527.6

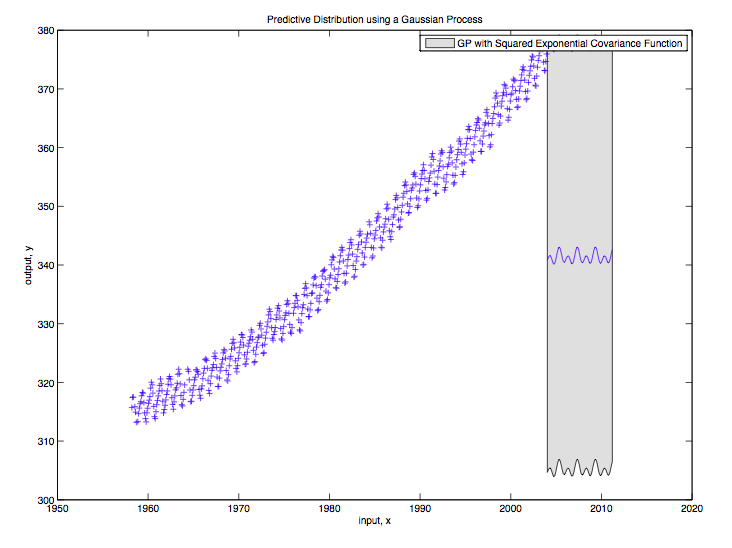
The figure below shows that the sum of these two additive covariance matrices is not suitable to model the predictive distribution.



With a different initialization of the hyperparameters, we obtain the following plot with a NLML=1207



The figure below shows that two periodic covariance functions are not ideal to model the predictive distribution either.



Finally, we added a Linear Covariance function with a squared exponential, which models the predictive distribution of the data better

