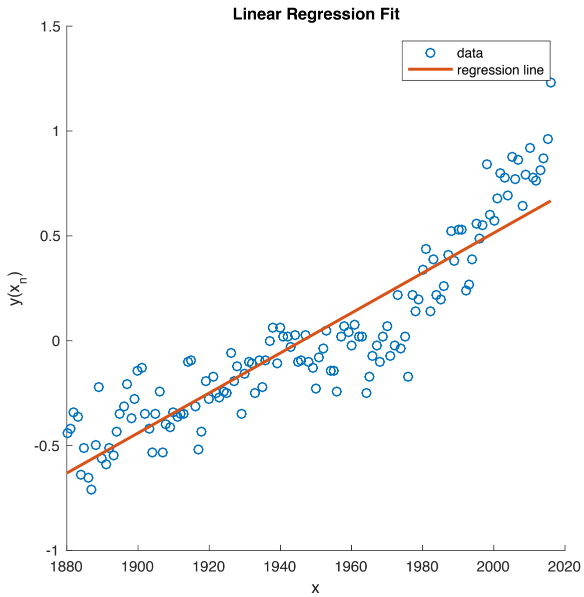
Lab #6

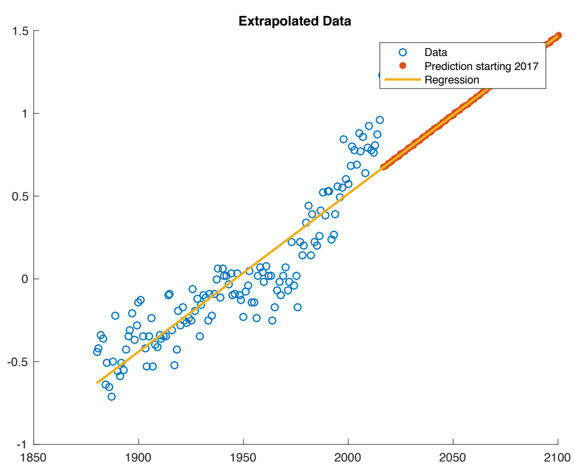
Experiment

1. The fitted linear regression is obtained by $y(x) = \mathbf{w}^\top \phi(x)$. The resulting regression line, together with a plot of the error obtained from $t_n = y(x_n) + \epsilon_n$, are the following:



The linear regression model y = wx + b is not an appropriate model for the data, which we can conclude from the error plot. The error should be zero-mean Gaussian, which means that error points should be symmetric with respect to zero, and that there should have no obvious pattern among the error points as x increases. But here the error points are not symmetrically scattered around zero, and there is a clear pattern starting from x = 1940. Therefore, the error is not Gaussian with zero mean, which makes the model inappropriate.

2. The extrapolated data is



The obtained values from 2017 are not accurate predictions in terms of the general trend. From year 2000 the actual data points are all above the regression line, and according to the trend the actual values from 2017 should be higher than the predictions. Also in #1 we noticed that the linear regression model I used to extrapolate data is not an appropriate model.

3. The kernel function is coded as:

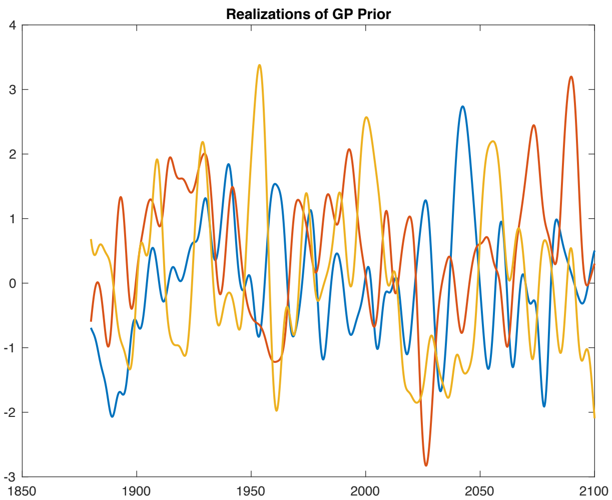
```
function [kernel_val] = compKernel(x,xp,tao_sq)
```

```
    temp = -(x-xp).^2/(2*tao_sq);
    kernel_val = exp(temp);

end
```

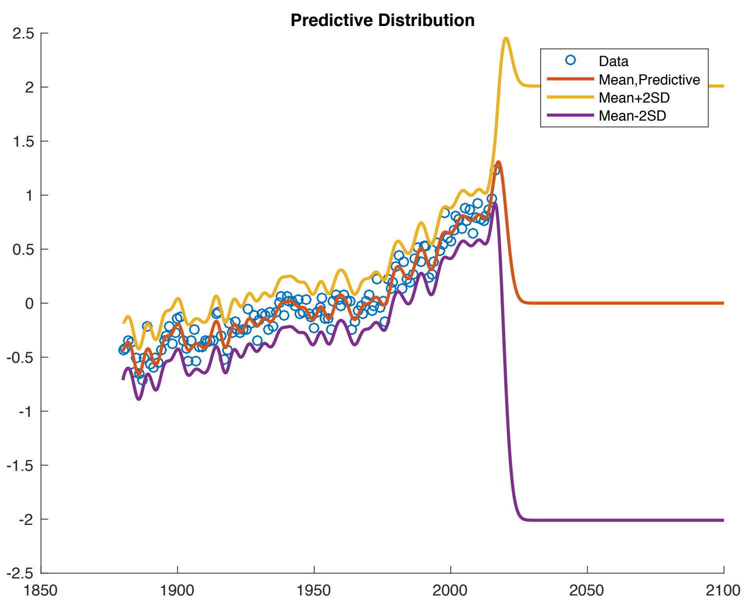
4.Sampling ten realizations of the corresponding GP prior using

```
%compute K
for i = 1:1000
    for j = 1:1000
        K_r(i,j) = compKernel(sample_(i),sample_x(j),10);
    end
end

sample_realization = mvnrnd(zeros(1000,1),K_r,3);
```

The resulting three realizations:



5. Here is the plot of the data and the mean of the predictive distribution:

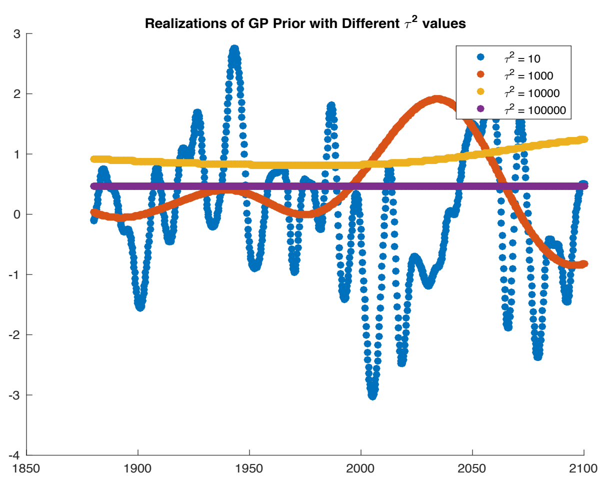


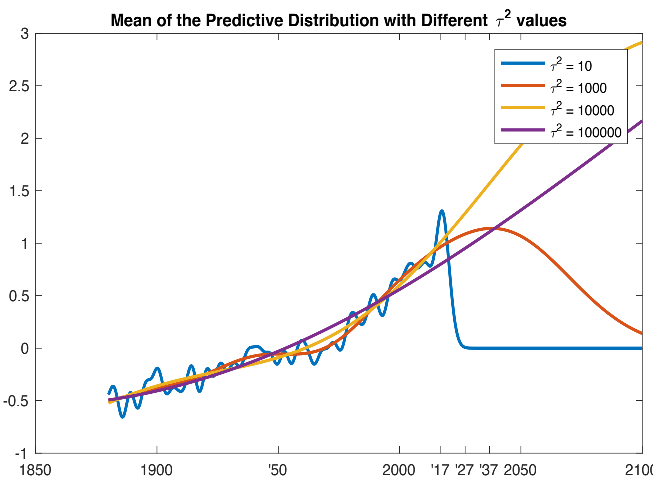
The obtained distribution provides a good estimate for data with x from 1880 to 2014, because the mean of the predictive distribution using 1000 data has a pattern that matches the pattern of the original (training) data. Also, in this time period, the lines corresponding to mean of the predictive distribution and the mean plus/minus two standard deviations are "parallel" to each other, because the majority of our original data fall in this time period, resulting in small variance of each data point. After year 2016, since where we do not have any information for t values, the mean of the predictive distribution, $m(\mathbf{x}_{N+1}) = \mathbf{k}^\top \mathbf{C}_N^{-1} \mathbf{t}$, goes to zero. The variances, $\sigma^2(\mathbf{x}_{N+1}) = c - \mathbf{k}^\top \mathbf{C}_N^{-1} \mathbf{k}$, increase and approach to c as $\mathbf{k}^\top \mathbf{C}_N^{-1} \mathbf{k}$ approaches to 0.

Although the obtained distribution provides good estimates for data corresponding to the time period in the past, it is not very effective when predicting values for yeaers to come.

6. Role of parameter τ^2: our kernel is $k_{\tau^2}(x, x') = \exp\left(-\frac{(x-x')^2}{2\tau^2}\right)$. Given the distance between x and x' unchanged, as τ^2 increases, the value of the kernel increases; as τ^2 decreases, the value of the kernel decreases. When τ^2 approaches to infinity, the kernel goes to 1; when τ^2 approaches to 0, the kernel is close to 0.

To see the role of τ^2, we plot realizations of GP prior and means of predictive distribution, using different values of τ^2.





From the left graph, I observe that when τ^2=10, the amplitude of fluctuation between points is the largest; when $\tau^2 = 1000$ the fluctuation is greatly reduced. And when τ^2 gets sufficiently large (100000), the data points in the realizations have a nearly "linear" pattern with very little fluctuation.

From the right graph, when $\tau^2 = 10$, the fluctuation in the mean of the predictive distribution is significant, and after 2017, the mean drops rapidly to 0. As τ^2 increases, the mean curve becomes smoother and smoother. The mean corresponding to $\tau^2 = 1000$ drops after x = 2037, but the means corresponding to τ^2=10000 and higher produce smooth curves that keeps increasing before x = 2100.

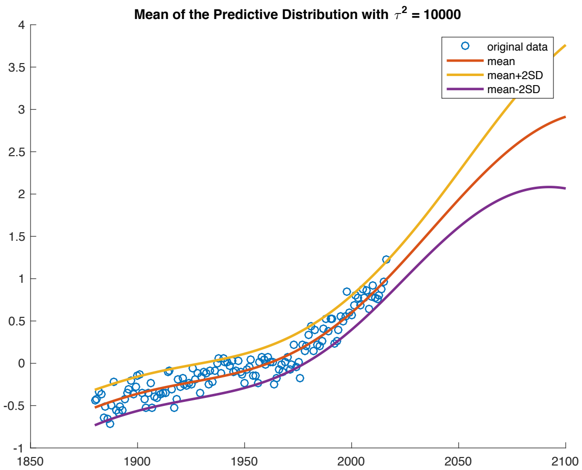
The graphs might indicate that when $\tau^2 = 10$, the model is good for predicting the temperature in year 2017 but not several years later. When $\tau^2 = 1000$, the model is good for predicting the temperature after year 2017 and before 2037 (the prediction for year 2017 is lower than what the trend tells it should be). When $\tau^2 = 10000$, the model is good for predicting the temperature after year 2037 before 2100. $\tau^2 = 100000$ should be used for prediction on temperature after year 2100.

Therefore, different values parameter τ^2 can be used to predict values corresponding to different time periods. If we want to predict the near future, a relatively small τ^2 value will give us more accurate predictions, but if we want to predict values in the long run, we should use a large τ^2 .

7. If $|x - x'|$ exceeds τ by a large amount, the value of our kernel will be very small and even close to 0, which makes the point of we want to predict insufficiently correlated with the original data points, based on

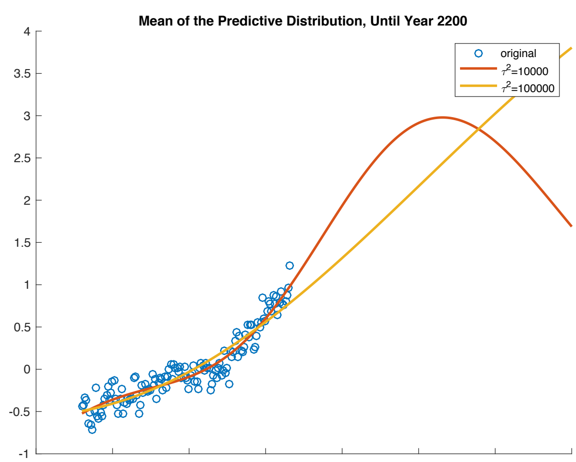
prior distribution. Thus, there will be no sufficient information can be used to make predictions, and extrapolate data.

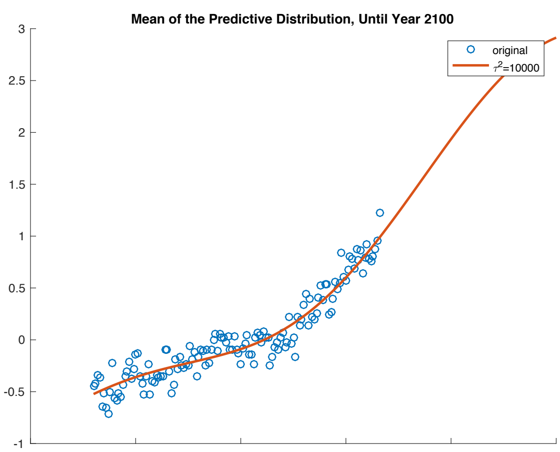
After trying different values of τ^2, I find that $\tau^2 = 10000$ will give a prediction for year 2100 that makes sense: it conforms the general trend from the original temperature data.



Uncertainty: Compared with the graph in #5, where τ^2=10, here the differences between curves mean plus two standard deviations and minus two standard deviations are much smaller after year 2016. This means that the uncertainty between predictions from year 2017 to year 2050 is still small, indicating that τ^2=10000 is an appropriate parameter to predict temperatures before 2050.

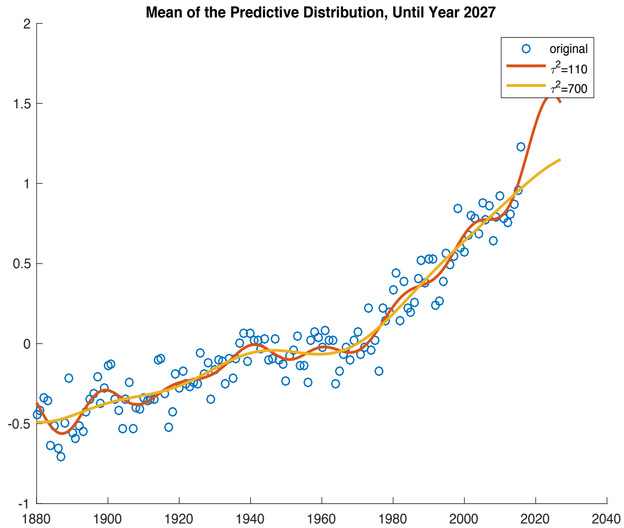
8. For extrapolation until year 2200, we should use $\tau^2 = 100000$. For extrapolation until year 2100, we should use $\tau^2 = 10000$. These choices of τ^2 make predictions of temperatures of year 2200 and 2100 conform the general trend of the original data, yet avoid the problem of overfitting caused by τ^2 values that are too small.





If we want to capture more localized changes in data, we should adopt smaller values of τ^2. For example, if we want to predict temperatures in the future ten years, we can pick small τ^2 as long as τ^2 is greater than 100. However, even if τ^2 is greater than 100, it cannot be too small because this will cause overfitting. From the graph below, when $\tau^2 = 110$, we see an overfitting. $\tau^2 =$

700 seems more reasonable.



Matlab code:

```
clear all
close all
load temperature_data

%%
%1. linear fit
phi = [ones(137,1),x];
w = inv((transpose(phi)*phi))*transpose(phi)*t;
er = t - phi*w;
figure
scatter(x,t)
hold on
plot(x,phi*w,'linewidth',2)
legend('data','regression line')
title('Linear Regression Fit')
hold off

figure
scatter(x,er)
hold on
refline(0,0)
title('Error')
hold off

%%
%2. extrapolate
x_ext = [x.',[2017:2100]].';
x_extra = [2017:2100].';
t_ext = w(1) + w(2)*x_ext;
t_extra = w(1) + w(2)*x_extra;
figure
scatter(x,t)
hold on
scatter(x_extra,t_extra,'filled')
hold on
plot(x_ext,t_ext,'Linewidth',2)
legend('Data','Prediction starting 2017','Regression')
title('Extrapolated Data')
hold off
```

```
%3. kernel function

%%
%4.realizations of Gaussian prior
sample_x = linspace(1880,2100,1000);
sample_x = sample_x.';

%compute K
for i = 1:1000
    for j = 1:1000
        K_r(i,j) = compKernel(sample_x(i),sample_x(j),10);
    end
end

sample_realization = mvnrnd(zeros(1000,1),K_r,3);

figure
plot(sample_x,sample_realization,'linewidth',1.5)

title('Realizations of GP Prior')



%%
%5.Mean and variance
%C = K + ?^2 * I_N

%compute K
for i = 1:137
    for j = 1:137
        K(i,j) = compKernel(x(i),x(j),10);
    end
end

C = K + 0.01 * eye(137);

for i = 1:1000
    c(i) = compKernel(sample_x(i),sample_x(i),10)+0.01;
end
c = c.';

for i = 1:1000
    for j = 1:137
        k(j,i) = compKernel(x(j),sample_x(i),10);
    end
end

for i = 1:1000
    mean(i) = k(:,i).'*inv(C)*t;
    sig_sq(i) = c(i)-k(:,i).'*inv(C)*k(:,i);
end

sig_sq = sig_sq.';

for i = 1:1000
    mean_plus(i) = mean(i)+2*sqrt(sig_sq(i));
    mean_minus(i) = mean(i)-2*sqrt(sig_sq(i));
end


figure
```

```matlab
scatter(x,t)
hold on
plot(sample_x,mean,'linewidth',2)
hold on
plot(sample_x,mean_plus,'linewidth',2)
hold on
plot(sample_x,mean_minus,'linewidth',2)
title('Predictive Distribution')
legend('Data','Mean,Predictive','Mean+2SD','Mean-2SD')
hold off

%%
%6. realizations
% tao = 10
for i = 1:1000
    for j = 1:1000
        K_1(i,j) = compKernel(sample_x(i),sample_x(j),10);
    end
end

r_1 = mvnrnd(zeros(1000,1),K_1,1);

% tao = 1000
for i = 1:1000
    for j = 1:1000
        K_2(i,j) = compKernel(sample_x(i),sample_x(j),1000);
    end
end

r_2 = mvnrnd(zeros(1000,1),K_2,1);


% tao = 10000
for i = 1:1000
    for j = 1:1000
        K_3(i,j) = compKernel(sample_x(i),sample_x(j),10000);
    end
end

r_3 = mvnrnd(zeros(1000,1),K_3,1);


% tao = 100000
for i = 1:1000
    for j = 1:1000
        K_4(i,j) = compKernel(sample_x(i),sample_x(j),100000);
    end
end

r_4 = mvnrnd(zeros(1000,1),K_4,1);



%%
%6. change values of tao
% tao = 10
for i = 1:137
    for j = 1:137
        K_10(i,j) = compKernel(x(i),x(j),10);
    end
end
```

```
sample_10 = mvnrnd(zeros(137,1),K_10,1);

% tao = 1000
for i = 1:137
    for j = 1:137
        K_1000(i,j) = compKernel(x(i),x(j),1000);
    end
end

sample_1000 = mvnrnd(zeros(137,1),K_1000,1);


% tao = 10000
for i = 1:137
    for j = 1:137
        K_10000(i,j) = compKernel(x(i),x(j),10000);
    end
end

sample_10000 = mvnrnd(zeros(137,1),K_10000,1);


% tao = 100000
for i = 1:137
    for j = 1:137
        K_100000(i,j) = compKernel(x(i),x(j),100000);
    end
end

sample_100000 = mvnrnd(zeros(137,1),K_100000,1);

[mymean1,myvar1] = mean_var(K_10, sample_x,10, x, t);
[mymean2,myvar2] = mean_var(K_1000, sample_x,1000, x, t);
[mymean3,myvar3] = mean_var(K_10000, sample_x,10000, x, t);
[mymean4,myvar4] = mean_var(K_100000, sample_x,100000, x, t);




%%
figure
scatter(sample_x,r_1,'filled')
hold on
scatter(sample_x,r_2,'filled')
hold on
scatter(sample_x,r_3,'filled')
hold on
scatter(sample_x,r_4,'filled')
legend('\tau^2 = 10','\tau^2 = 1000','\tau^2 = 10000','\tau^2 = 100000')
title('Realizations of GP Prior with Different \tau^2 values')
hold off

figure
plot(sample_x, mymean1,'linewidth',2)
hold on
plot(sample_x,mymean2,'linewidth',2)
hold on
plot(sample_x,mymean3,'linewidth',2)
hold on
plot(sample_x,mymean4,'linewidth',2)
legend('\tau^2 = 10','\tau^2 = 1000','\tau^2 = 10000','\tau^2 = 100000')
title('Mean of the Predictive Distribution with Different \tau^2 values')
hold off
```

```
%%
%7.
mymean3 = mymean3.';
mean_p_sd = mymean3+2*sqrt(myvar3);
mean_m_sd = mymean3-2*sqrt(myvar3);

figure
scatter(x,t)
hold on
plot(sample_x,mymean3,'linewidth',2)
plot(sample_x,mean_p_sd,'linewidth',2)
plot(sample_x,mean_m_sd,'linewidth',2)
legend('original data','mean','mean+2SD','mean-2SD')
title('Mean of the Predictive Distribution with \tau^2 = 10000')
hold off

%%
%8.
sample_new = linspace(1880,2200,1000);
sample_new = sample_new.';
% tao = 10000
for i = 1:137
    for j = 1:137
        Kn_10000(i,j) = compKernel(x(i),x(j),10000);
    end
end

sample_n10000 = mvnrnd(zeros(137,1),Kn_10000,1);


% tao = 100000
for i = 1:137
    for j = 1:137
        Kn_100000(i,j) = compKernel(x(i),x(j),100000);
    end
end

sample_n100000 = mvnrnd(zeros(137,1),Kn_100000,1);

[mymean3_n,mysample3_n] = mean_var(Kn_10000, sample_new,10000, x, t);
[mymean4_n,mysample4_n] = mean_var(Kn_100000, sample_new,100000, x, t);
%%
figure
scatter(x,t)
hold on
plot(sample_new,mymean3_n,'linewidth',2)
hold on
plot(sample_new,mymean4_n,'linewidth',2)
legend('original','\tau^2=10000','\tau^2=100000')
title('Mean of the Predictive Distribution, Until Year 2200')
hold off
%%
figure
scatter(x,t)
hold on
plot(sample_x,mymean3,'linewidth',2)
legend('original','\tau^2=10000')
title('Mean of the Predictive Distribution, Until Year 2100')
hold off

%%
sample_local = linspace(1880,2027,1000);
```

```
sample_local = sample_local.';
% tao = 110
for i = 1:137
    for j = 1:137
        K_30(i,j) = compKernel(x(i),x(j),110);
    end
end

sample_30 = mvnrnd(zeros(137,1),K_30,1);



% tao = 500
for i = 1:137
    for j = 1:137
        K_100(i,j) = compKernel(x(i),x(j),700);
    end
end

sample_100 = mvnrnd(zeros(137,1),K_100,1);

[mymean30,mysample30] = mean_var(K_30, sample_local,110, x, t);
[mymean100,mysample100] = mean_var(K_100, sample_local,700, x, t);
%%
figure
scatter(x,t)
hold on
plot(sample_local,mymean30,'linewidth',2)
hold on
plot(sample_local,mymean100,'linewidth',2)
legend('original','\tau^2=110','\tau^2=700')
title('Mean of the Predictive Distribution, Until Year 2027')
hold off
```