Math 540: Project 2

Tilekbek Zhoroev

1. Repeat the analysis of Example 11.22 for the steady state heat model using the copper data in Table 1 and the values $k = 4.01$ for the thermal conductivity and $u_{amb} = 22.28$. Do your residuals appear to be iid? Use a Q-Q plot, as discussed in Definition 4.19 to establish whether the residuals are normally distributed.

x	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66
Temp	66.04	60.04	54.81	50.42	46.74	43.66	40.76	38.49	36.42	34.77	33.18	32.36	31.56	30.91	30.56

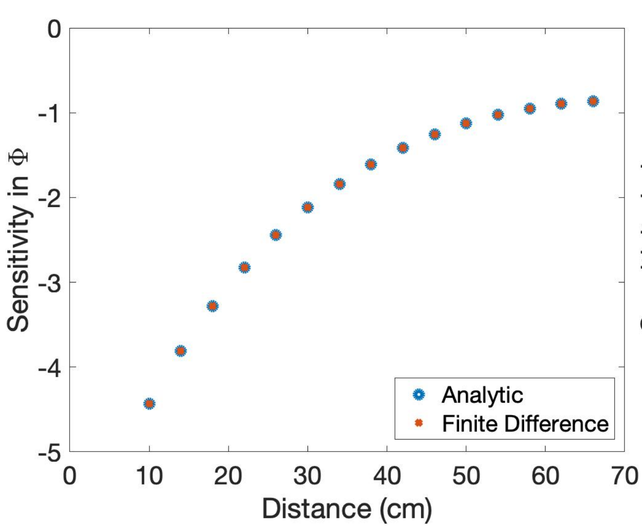
Table 1: Steady-state temperatures measured at locations x for a copper rod.

Solution.

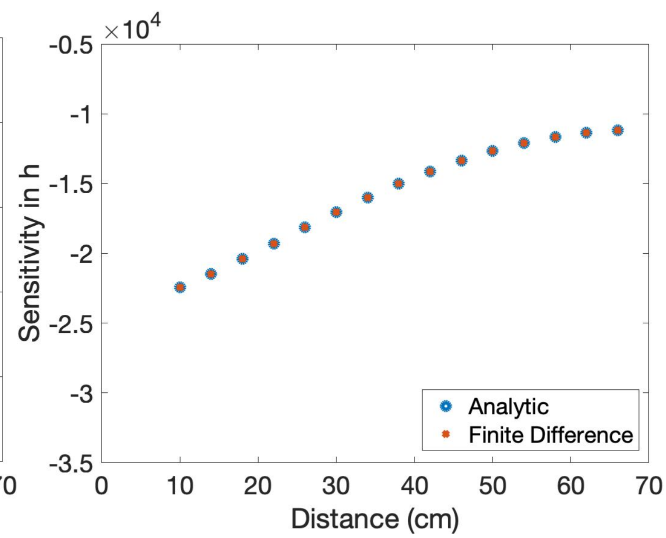
We have given the the temperature of the copper rote on equidistant points in Table 1. We also given the thermal conductivity model for the copper $k = 4.01$ and the ambient room temperature $u_{amb} = 22.38$. Using the analytic solution of the heat equation and given data we constructed least squares problem. Then, using **fminsearch.m** we obtained parameter estimations $\Phi = -9.9265$ and $h = 0.0014$.

Subsequently, we plot the graph of the our model using estimated parameters and data, to observe the fitting of model to the data. The Figure 2 (a) shows that we have good parameter estimated, since the fitting is good. However, we observe from Figure 2 (b) that residual is identical but not independent. Hence, we would like to analyse it using Q-Q plot of the residuals and it's presented in Figure 2 (c). We observe that the in Q-Q plot residuals are not follows straight line, so residuals are not from normal distribution.

We have given $n = 16$ points and in the heat model we have $p = 2$ parameters, so we accomplish the error variance estimate,



(a)



(b)

Figure 1: Sensitivity equation (a)$\dfrac{\partial u}{\partial \Phi}$; (b)$\dfrac{\partial u}{\partial h}$;

$$\sigma^2 = \frac{1}{n-p}\boldsymbol{R}^\top \boldsymbol{R} = 0.0529$$

where $\boldsymbol{R}$ is the vector of the residuals. Then using this along with sensitivity equations we obtain the covariance matrix,

$$\mathbf{V} = \sigma^2[\mathbf{X}^\top \mathbf{X}]^{-1} = \begin{bmatrix} 0.0090 & -1.2117 \times 10^{-6} \\ -1.2117 \times 10^{-6} & 1.7726 \times 10^{-10} \end{bmatrix}$$

where $\mathbf{X}$ is the sensitivity matrix. The graph of the sensitivities are presented in Figure 1. Then by utilizing sensitivity matrix we obtain the estimated standard deviations of the error and parameters,

$$\sigma = 0.2301, \sigma_\Phi = 0.0947, \sigma_h = 1.3314 \times 10^{-5}.$$

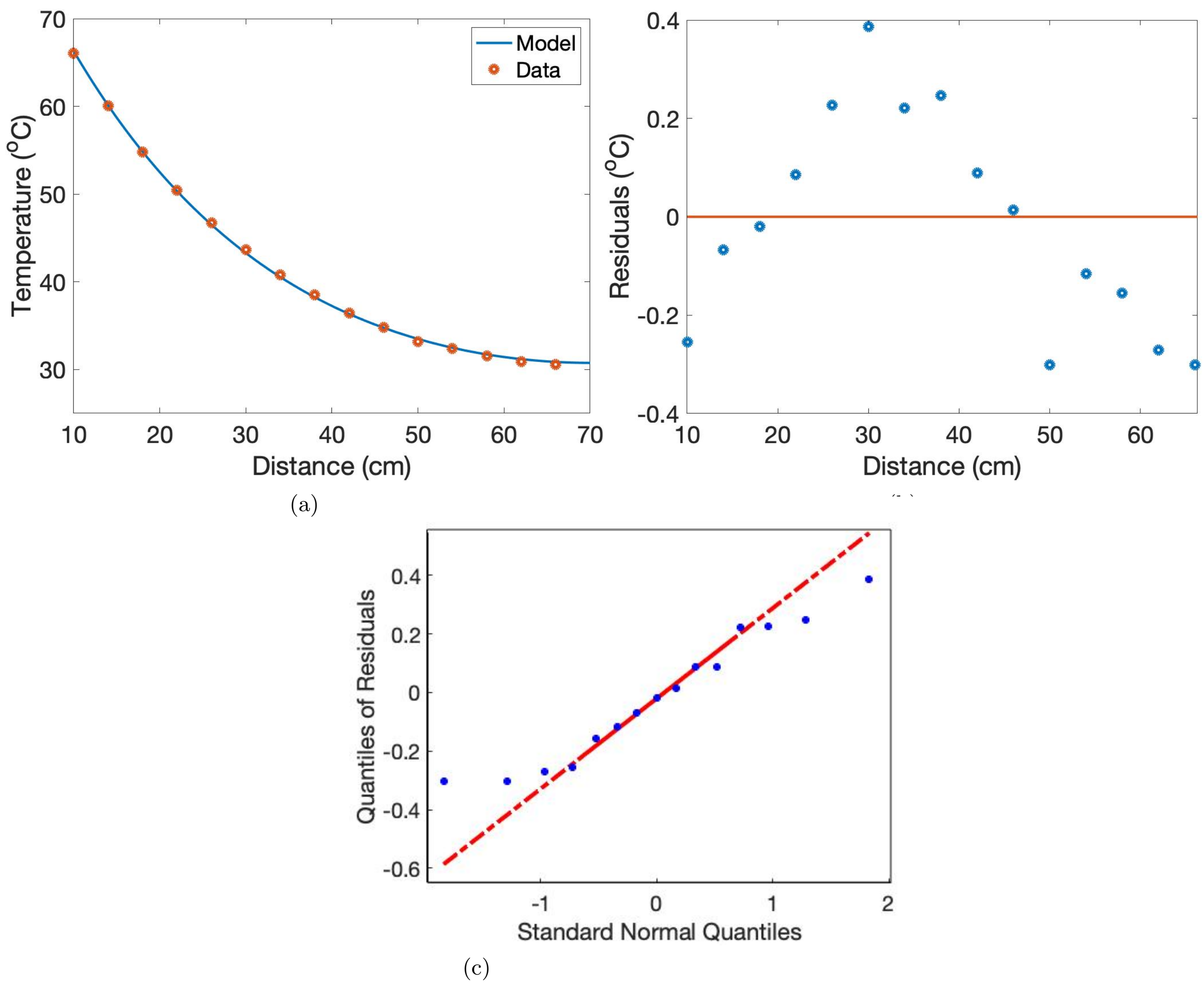


Figure 2: (a) The graph of the data and model for the uninsulated copper rod; (b) The residual plot; (c) Q-Q plot of the residuals.

Then using online t-table we obtain the 95% confidence intervals of each parameters

$$\Phi \in [-10.1310, -9.7219]$$
$$h \in [0.0014, 0.0015]$$

2. Consider the Helmholtz energy

$$\psi(P, \boldsymbol{\theta}) = \alpha_1 P^2 + \alpha_{11} P^4 + \alpha_{111} P^6,$$

where P is the polarization on the interval $[0, 0.8]$ and $\boldsymbol{\theta} = [\alpha_1, \alpha_{11}, \alpha_{111}]^\top$ are parameters. We will employ the nominal values $\alpha_1 = -389.4$,$\alpha_{11} = 761.3$ and $\alpha_{111} = 61.5$.

(a) For $n = 81$, 161 and 801 equally spaced polarization values $P_i = (i-1)\Delta P$, $\Delta P = \frac{0.8}{n-1}$ $i = 1, ..., n$, compute the model response $\psi(P_i, \boldsymbol{\theta})$ and observations

$$Y_i = \psi(P_i, \boldsymbol{\theta}) + \varepsilon_i,$$

where $\varepsilon_i \sim^{iid} \mathcal{N}(0, \sigma^2)$ with $\sigma = 2.2$. In each case, compute the OLS estimate for the observation variance σ^2 and compare it to the true value. What do you observe?

(b) For $n = 161$, use the normal equations to approximate the parameters and compare to the nominal values. Compute the covariance matrix estimate V and discuss the correlation of the parameters. Discuss why global sensitivity methods, with the assumption of mutually independent parameters, may give misleading results. Plot the residuals and 2σ intervals and discuss whether the intervals appear to be correct. Finally, plot the model and observations as a function of the polarization values.

Solution.

(a) For each case we computed the model response as described in the statement of the problem. Then using **normrnd.m** with $\mu = 0$ and $\sigma = 2.2$ to generate the noise term. Then by adding to the model response we obtained the synthetic observations $Y_i, i = 1, 2, ..., n$. Next, to compute the the estimated observation variance $\hat{\sigma}$ we used the residuals with $p = 3$,

$$\hat{\sigma}^2 = \frac{1}{n-p} \boldsymbol{R}^\top \boldsymbol{R}.$$

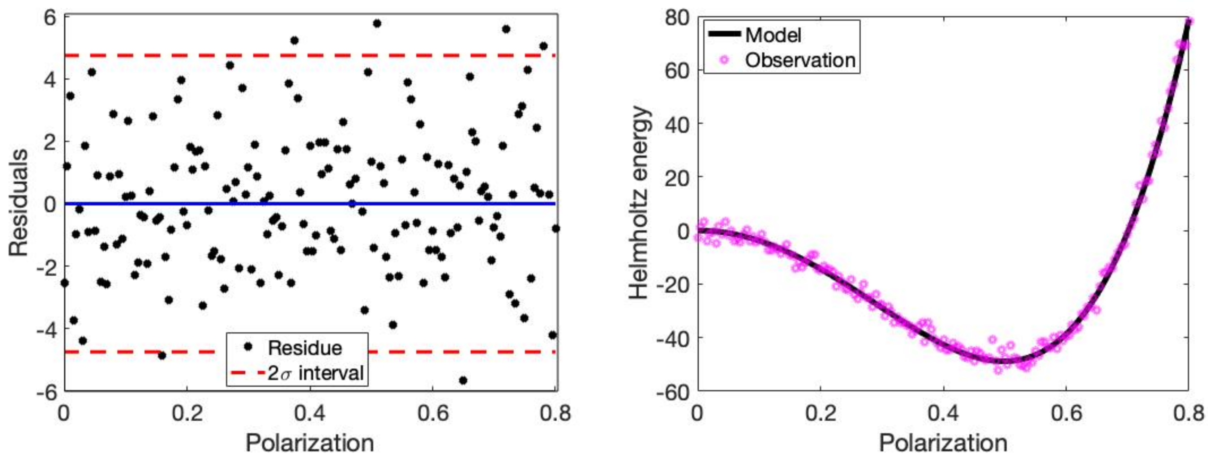
The estimated sigma's are

$$\hat{\sigma}_{n=81} = 2.3841,\ \hat{\sigma}_{n=161} = 2.2594,\ \hat{\sigma}_{n=801} = 2.1791.$$

From these results we observe that as data points are increase the estimated variance is converges to the true variance.

(b) Using the given observation formula

$$\begin{aligned} Y_i &= \psi(P_i, \boldsymbol{\theta}) + \varepsilon_i \\ &= [P_i^2, P_i^4, P_i^6] \begin{bmatrix} \alpha_1 \\ \alpha_{11} \\ \alpha_{111} \end{bmatrix} + \varepsilon_i \end{aligned}$$



(a) (b)

Figure 3: (a) Residual plot; (b) Graph of the model and observations

Thus, we obtain the design matrix,

$$\mathbf{X} = \begin{bmatrix} P_1^2 & P_1^4 & P_1^6 \\ P_2^2 & P_2^4 & P_2^6 \\ \vdots & \vdots & \vdots \\ P_n^2 & P_n^4 & P_n^6 \end{bmatrix}.$$

To find estimated parameters we used normal equations,

$$(\mathbf{X}^\top \mathbf{X})\hat{\boldsymbol{\theta}} = \mathbf{X}^\top \mathbf{Y} \implies \hat{\boldsymbol{\theta}} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{Y}$$

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_{11} \\ \hat{\alpha}_{111} \end{bmatrix} = \begin{bmatrix} -390.0709 \\ 764.5939 \\ 57.2555 \end{bmatrix}.$$

The covariance estimate is,

$$\mathbf{V} = \begin{bmatrix} 22.5543 & -110.0835 & 123.4682 \\ -110.0835 & 585.0489 & -690.2614 \\ 123.4682 & -690.2614 & 842.2265 \end{bmatrix}.$$

The correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with standard deviations σ_X and σ_Y and covariance $\text{cov}(X,Y)$ is defined as

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}.$$

By covariance matrix we obtained the correlation coefficients,

$$\rho_{\alpha_1,\alpha_{11}} = \frac{\text{cov}(\alpha_1,\alpha_{11})}{\sigma_{\alpha_1}\sigma_{\alpha_{11}}} = \frac{-110.0835}{\sqrt{22.543} \times \sqrt{585.0489}} = -0.9585$$

$$\rho_{\alpha_1,\alpha_{111}} = \frac{\text{cov}(\alpha_1,\alpha_{111})}{\sigma_{\alpha_1}\sigma_{\alpha_{111}}} = \frac{123.4681}{\sqrt{22.543} \times \sqrt{842.2265}} = 0.8960$$

$$\rho_{\alpha_{11},\alpha_{111}} = \frac{\text{cov}(\alpha_{11},\alpha_{111})}{\sigma_{\alpha_{11}}\sigma_{\alpha_{111}}} = \frac{690.2614}{\sqrt{585.0489} \times \sqrt{842.2265}} = 0.9833$$

From here we observe that all parameters are mutually correlated with strong correlation. Thus, we cannot assume this parameters mutually independent to employ global sensitivity analysis. In Figure 3 (a), we presented the residuals and it appear identically and independent distributed. Moreover, the most of the residuals are between $2\sigma-$interval as expected. Finally in Figure 3 (b) the plot of the model and observations displayed.

3. Consider the SIR model

$$\begin{aligned} \frac{dS}{dt} &= \delta N - \delta S - \gamma IS, \qquad & S(0) = 900 \\ \frac{dI}{dt} &= \gamma IS - (r+\delta)I, \qquad & I(0) = 100 \\ \frac{dR}{dt} &= rI - \delta R, \qquad & R(0) = 0 \end{aligned}$$

where γ, r and δ are each in the interval [0, 1]. The parameters $\boldsymbol{\theta} = [\gamma, r, \delta]$

(a) Use the data in the file **SIR.txt** to estimate the parameters $\boldsymbol{\theta}$ using the routine **fminsearch.m**. The first column contains times t_j and the second is corresponding values $I(t_j)$. You can approximate the solution to the ODE system using **ode45.m**, as you did in Project 1.

(b) Construct the local sensitivity matrix using the complex-step derivative approximation as illustrated in the MATLAB code for Example 8.10. Estimate the variance σ^2 of the observations, compute the covariance matrix **V** , and discuss the parameter correlation. Is **V** full rank and are your results to be expected.

(c) Using the parameter estimates and variances, use the command **normpdf.m** to plot the distributions for each of the three parameters. Additionally, you should plot your residuals, your model fit to the data $I(t_j)$ and the trajectories for $S(t)$, $I(t)$ and $R(t)$.

(d) Consider the influenza data in Table 2. This data was collected during a flu outbreak at a British boarding school. The total population of boys is $N = 763$ and you can take $S(0) = 760$, $I(0) = 3$ and $R(0) = 0$. Due to the short duration of the outbreak, it is appropriate to specify $\delta = 0$ and consider the parameters $\boldsymbol{\theta} = [\gamma, r]$. Repeat the analysis of (a) and (b). You should also compare the model fit to the data and plot the residuals.

Day	0	1	2	3	4	5	6	7	8	9	10	11	12	13
Confined to bed	3	8	26	76	225	298	258	233	189	128	68	29	14	4

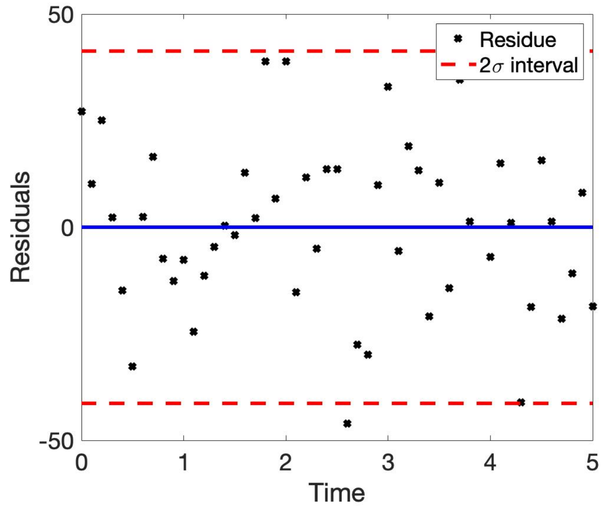
Table 2: Influenza data

Solution

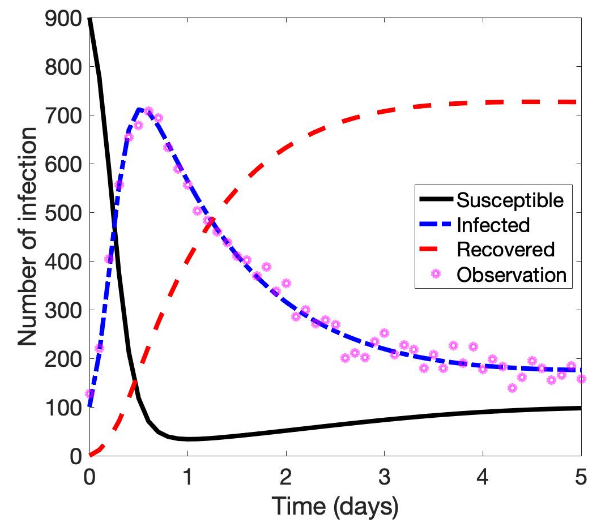
(a) Using given data and solution of the ode system we constructed the least squares problem. Then using **fminsearch.m** we have obtained the approximate parameters,

$$\gamma = 0.0100, \;\; \delta = 0.1953, \;\; r = 0.7970.$$

Following this we get the plot of the residuals and fitted model with data. From residual plot, Figure 4 (a) we observe that the errors are identically and independent distributed. Figure 4 (b) shows that we have good fitting model to the given observational data.



(a)



(b)

Figure 4: (a) Residual plot; (b) The graph of the SIR model and observational data.

(b) Using estimated parameters and complex-step approximation

$$f'(x) \approx \frac{Im(f(x+ih))}{h}$$

with $h = 10^{-16}$ we obtained the sensitivity equations. Then using these we obtained sensitivity matrix,$\boldsymbol{\chi}$. Moreover using data and obtained model solution we get the vector of residuals, $\boldsymbol{R}$. Using the given $n = 51$ data points and $p = 3$, we estimated the variance,

$$\sigma^2 = \frac{1}{n-p}\boldsymbol{R}^\top \boldsymbol{R} = 426.7780$$

and covariance matrix,

$$\mathbf{V} = \sigma^2(\boldsymbol{\chi}^\top\boldsymbol{\chi})^{-1} = \begin{bmatrix} 1.7357\times 10^{-8} & 1.8575\times 10^{-7} & 1.4540\times 10^{-7} \\ 1.8575\times 10^{-7} & 2.6153\times 10^{-5} & 3.1790\times 10^{-5} \\ 1.4540\times 10^{-7} & 6.1790\times 10^{-5} & 8.4797\times 10^{-5} \end{bmatrix}.$$

From covariance matrix and correlation formula we obtained,

$$\rho_{\gamma,\delta} = 0.2757,, \;\rho_{\gamma,r} = 0.1198, \;\rho_{\delta,r} = 0.6751.$$

The results are implies that δ and r are correlated, but others are weakly correlated. Moreover, the rank of the covariance matrix is 3, and this is expected result from Project 1.

(c) Using estimated covariance matrix we obtained estimated standard deviations of the parameters,

$$\sigma_\gamma = 1.3174\times 10^{-4}, \;\sigma_\delta = 0.0051, \;\sigma_r = 0.0092,$$

and 95% confidence intervals of each parameters,

$$\gamma \in [0.0097, 0.0103]$$
$$\delta \in [0.1850, 0.2056]$$
$$r \in [0.7785, 0.8155]$$

Using these results and estimated parameter values with use of **normpdf.m** we obtained the the sample points from normal distributions for each parameter. The plot of the distributions are given in Figure 5.

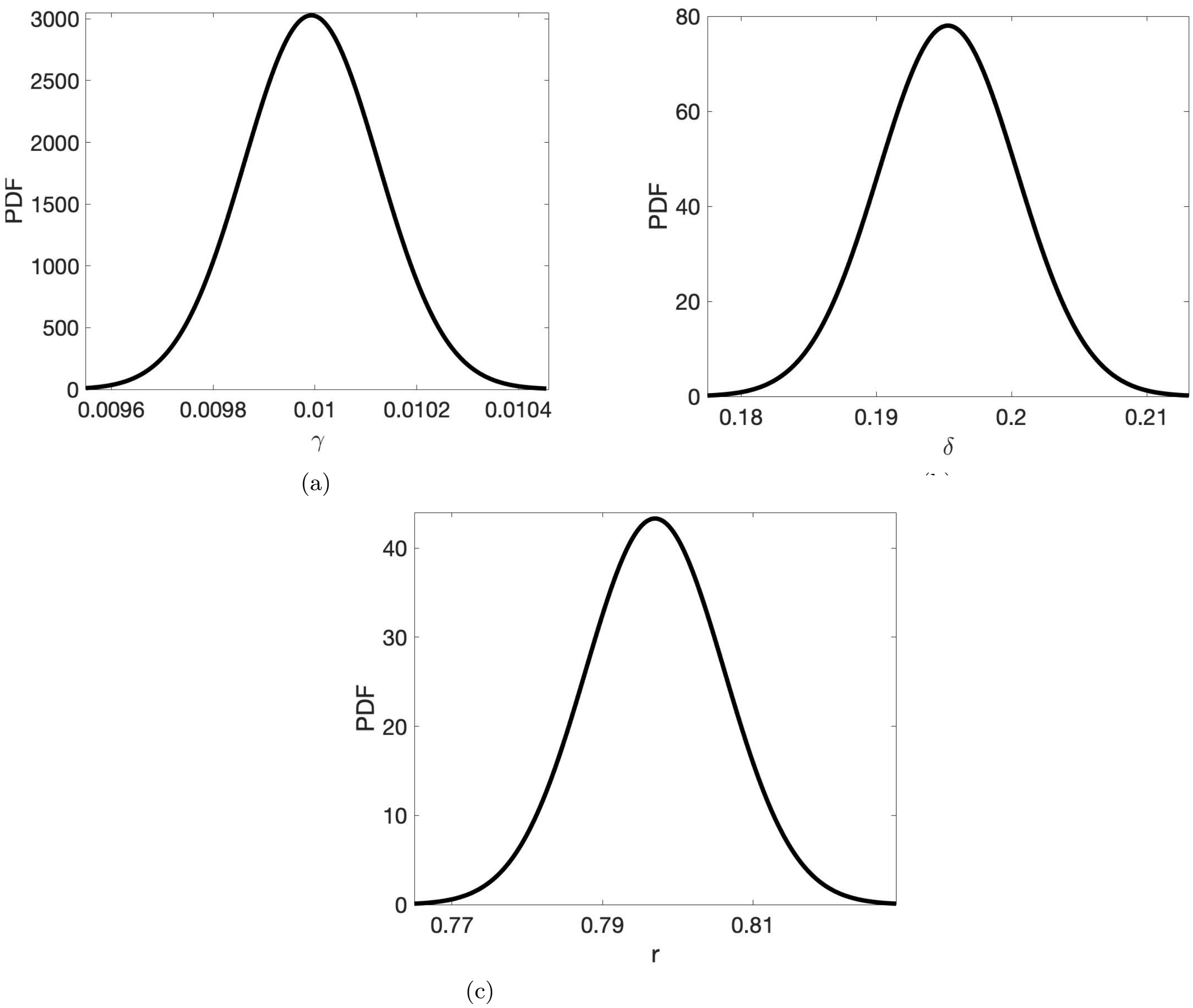


Figure 5: Graph of the distribution of (a) γ; (b) δ; (c) r.

(d) As previously using **fminsearch.m** we obtained the optimal parameter value,

$$\gamma = 0.0022, \; r = 0.4469.$$

Then, by utilizing these parameter values we obtained the sensitivity equations and sensitivity matrix, $\boldsymbol{\chi}$. In this case we have given $n = 14$ points and $p = 2$ parameters.

Subsequently, the variance estimate formula yields that,

$$\sigma^2 = 325.4893.$$

The covariance estimate is become,

$$\mathbf{V} = \begin{bmatrix} 3.3117 \times 10^{-10} & 1.3030 \times 10^{-7} \\ 1.3030 \times 10^{-7} & 1.5355 \times 10^{-4} \end{bmatrix}$$

and the correlation coefficient of two parameters is

$$\rho_{\gamma,r} = 0.5778.$$

Following it, we obtained the estimated standard deviation of the error and parameters,

$$\sigma = 18.0413, \sigma_\gamma = 1.8198 \times 10^{-5}, \sigma_r = 0.0124.$$

Then using online t-values we obtained the 95% confidence intervals of each parameters,

$$\gamma \in [0.00217, 0.00226], \quad r \in [0.4215, 0.4755].$$

Finally, the graph of the SIR model with observational data and residuals are given in Figure6.

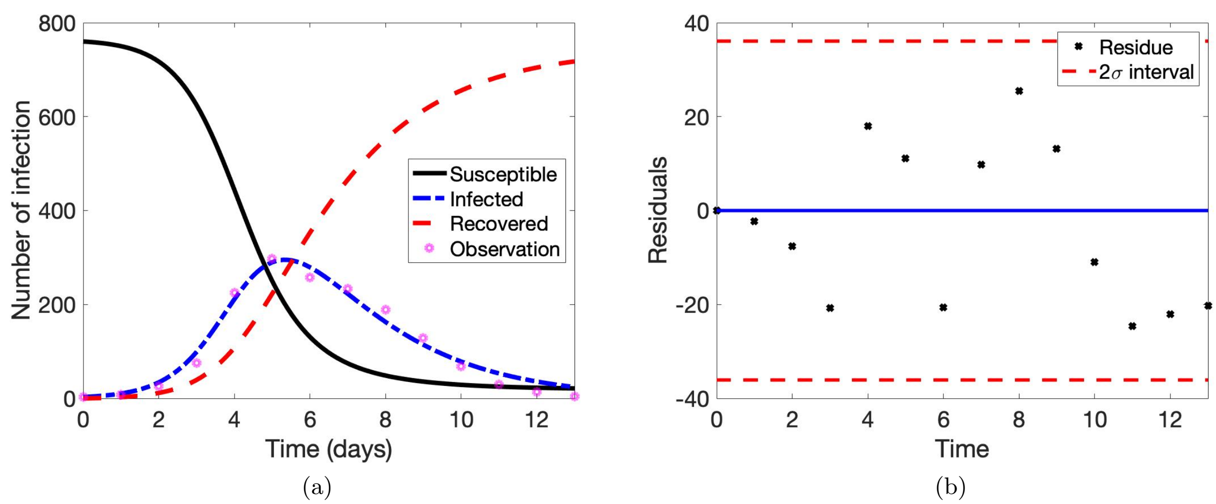


Figure 6: (b) The graph of the SIR model and observational data. (a) Residual plot;