Ebola Virus Disease in Sierra Leone:
Efficiency of control measures and Transmission after death.

The Model

N = Population size.
S = Number of susceptible people in the population.
E = Number of people exposed to the disease, but are not infectious.
I = Number of infected people, capable of transmitting the disease.
D_I = Number of dead people who are still infectious.
T_X = Incubation time = Time from exposure to the appearance of symptoms.
T_I = Infection time = Time from appearance of symptoms to time of death or recovery.
T_D = Time during which a dead person remains infectious.
T_c = Time when control measures are implemented.
f = Mean fatality rate.
k = Efficiency of control measures.
β_I = Transmission rate from infectious people.
β_D = Transmission rate from dead people who are still infectious.
$\rho_I = \beta_I/T_I$ = Partial reproduction number due to infectious people.
$\rho_D = \beta_D/T_D$ = Partial reproduction number due to dead people who are still infectious.
$\rho = \rho_I + f\rho_D$ = Total reproduction number.

With the simple approximation of uniform mixing, the probability of coming in contact with an infectious person is I/N. Similarly, the probability of coming in contact with a dead person who is still infectious is D_I/N. Let β_I and β_D be the transmission rate of the disease due to contact with sick people and dead people who are infectious. E represents the number of people exposed to the disease and who do not show symptoms, or transmit the disease. People who are exposed will become sick after an incubation time T_X, and will thereafter become infectious. Then, the number of people exposed to the disease per unit time is:

$$\frac{dE}{dt} = \beta_I S\,(I/N) + \beta_D S\,(D_I/N) - \frac{E}{T_X}. \tag{1}$$

Let us make the approximation that the population N is much larger than the number of

people exposed or infected. Thus $N \gg I, E$. Therefore, we have $S \approx N$ and $dS/dt \approx 0$. We can rewrite Eq. 1 as:

$$\frac{dE}{dt} \approx \beta_I I + \beta_D D_I - \frac{E}{T_X}. \tag{2}$$

Infected people are able to transmit the disease over a period T_I, after which they either die, or recover from illness (and are neither infectious, nor susceptible). The number of infected people per unit time is given by:

$$\frac{dI}{dt} = \frac{E}{T_X} - \frac{I}{T_i}. \tag{3}$$

People who die from the disease can still transmit the disease for a period T_D. The number of dead people who remain infectious is given by:

$$\frac{dD_I}{dt} = \frac{fI}{T_X} - \frac{D_i}{T_D}, \tag{4}$$

where f is the mean fatality rate, and T_D is the time period during which a dead person can transmit the disease.

The parameter that determines whether an epidemic dies out, or increases exponentially is the *reproduction number* defined as:

$$\begin{aligned} \rho &= \rho_I + f\rho_D \\ &= \beta_I T_i + f\beta_D T_D. \end{aligned} \tag{5}$$

The set of equations that need to be solved are:

$$\frac{d\vec{V}}{dt} = R\,\vec{V}, \tag{6}$$

where the vector $\vec{V}^T = [E, I, D_I]$, and the matrix A is given by

$$A = \begin{bmatrix} -\frac{1}{T_X} & \frac{\rho_I}{T_i} & \frac{f\rho_D}{T_D} \\ \frac{1}{T_X} & -\frac{1}{T_i} & 0 \\ 0 & \frac{f}{T_X} & -\frac{1}{T_D} \end{bmatrix}$$

. The solution to Eq. 6 is an exponential. There exists a growing solution if one of the eigenvalues is positive, i.e. the epidemic is exponentially increasing provided the reproduction number $\rho > 1$.

The reproduction number determines the number of people infected by a sick person. For the epidemic to subside, it is necessary to have $\rho < 1$. This can be achieved through control

measures, e.g. isolation, quarantine, wearing of masks, etc. Let T_c be the time when control measures become effective. We assume that the time variation of the reproduction rate can be expressed in the simple form:

$$\begin{aligned} \rho &= \rho(t=0) \quad \text{for time t} < \mathrm{T_c} \\ &= \rho(t=0)\, e^{-k(t-T_c)} \quad \text{for t} \geq \mathrm{T_c} \end{aligned}$$
$$. \tag{7}$$

The parameter k represents the efficiency of control measures, and quantifies the time taken for the epidemic to subside once control measures are implemented. The cumulative number of cases and deaths are given by:

$$\begin{aligned} C(t) &= \int_0^t dt' \, \frac{E(t')}{T_X} \\ D(t) &= \int_0^t dt' \, \frac{fI(t')}{T_I} \end{aligned} \tag{8}$$

Method

We use publicly available data from the World Health Organization:

`http://www.afro.who.int/en/clusters-a-programmes/dpc/epidemic-a-pandemic-alert-and-response/sitreps.html`

to obtain tables of the cumulative number of cases and deaths over time, for Sierra Leone. We solve Eq. 6 numerically and compute the disease parameters using a Markov Chain Montecarlo analysis of the data using Python's PyMC package. For the sake of simplicity, we do not vary all parameters: We fix the incubation time T_X, infection time T_I, and mean fatality rate f from the analysis done by the WHO Response Team ("Ebola Virus Disease in West Africa - The First 9 Months of the Epidemic and Forward Projections", *New England Journal of Medicine*, Oct 2014): $T_X = 8.6$ days, $T_I = 9$ days, $f = 0.316$. We vary 5 parameters: Efficiency of control measures k, time when control measures become effective T_c, infection period after death T_D, and the two partial reproduction numbers ρ_I and ρ_D.

Explanation of files, and how to run the code

The folder data/ contains data from the publicly available WHO website, for Sierra Leone, Guinea, and Liberia. The .dat files contain 3 columns: Day, Cumulative cases,

Cumulative deaths. The cvs file contains data for all countries.

File **ebola.py**
You must edit this file before running the code!
"fileName" is the input (ascii) data file name.
"outFolderName" is the output folder where results will be stored (remember to create this folder before running the code!)
"output" is the name of the output file.
"numSamples" is the number of MCMC samples
"numBurn" is the number of samples to ignore.
"trnLength": Set to -1 to use all data points, else it sets the training size.

When starting a run, set numSamples to e.g., 50,000 and numBurn to e.g. 25,000. This will output a file and a covariance matrix. This chain has not converged and must be discarded.

For the next run, set numSamples to e.g. 100,000 and numBurn to e.g. 50,000. The output will be a mostly converged chain, and can be used for analysis.

For subsequent runs, set numSamples to e.g. 50,000 and numBurn = 0.

File **ebolaModel.py**
This files solves the disease model.
File **ebolaMC.py**
Runs the MCMC.
File **ebolaResults.py**
This program is used to analyze the chains. When you have a sufficient number of files in your output folder (each file having a number of chains), you may want to analyze the MCMC data and produce confidence contours:

python ebolaResults.py <data folder> <output file name> <burn>
e.g. python ebolaResults.py chainsSL SL.png 1

Results for Sierra Leone

Median efficiency of control measures $k = (7.87 \pm 0.276) \times 10^{-3}$ per day.

Characteristic time $\tau = 1/k = 127$ days.

Median time when control measures became effective $T_C = 142.34 \pm 1.68$ days.

Median value of partial reproduction number $\rho_I = 1.176 \pm 0.0616$

Median value of partial reproduction number $\rho_D = 0.085 \pm 0.12$

Total reproduction number $\rho = 1.2$

Median value of infection time after death $T_D = 1.44 \pm 1.35$

Reproduction number in April 2015 $= \rho(t = 0) \exp -k(t - T_C) = 0.33$

Conclusion:

1. Control measures are effective. The reproduction number at the present time < 1, and the epidemic will subside in the coming months.

2. There is no statistically significant evidence for disease transmission after death.

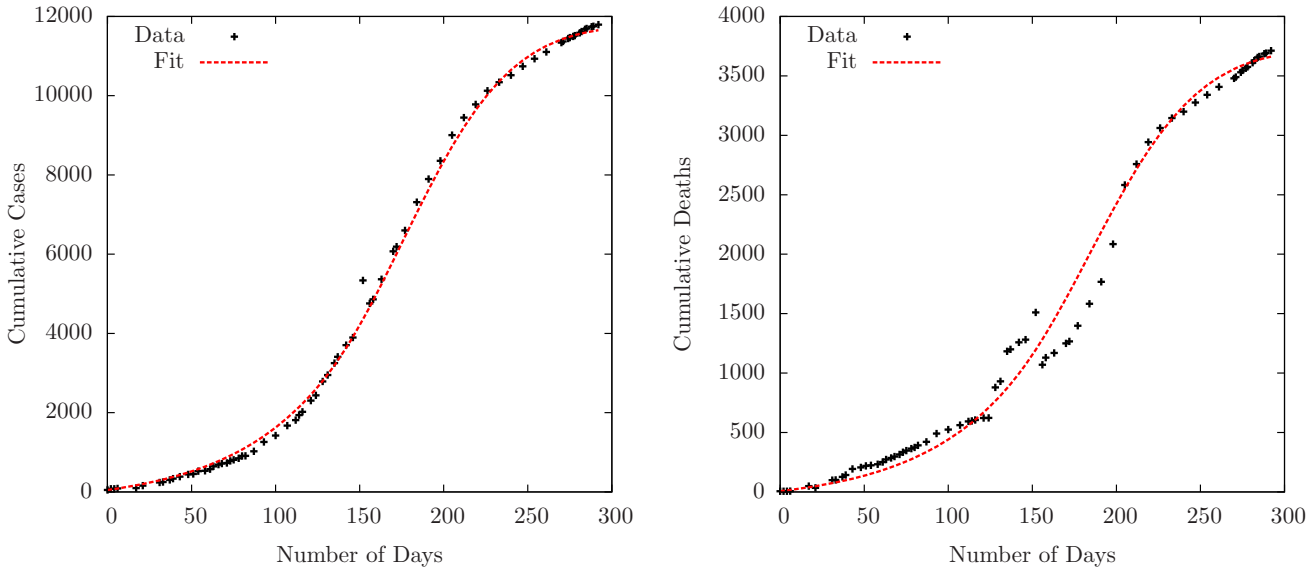

FIG. 1: Cumulative cases and deaths - Data and Fit.

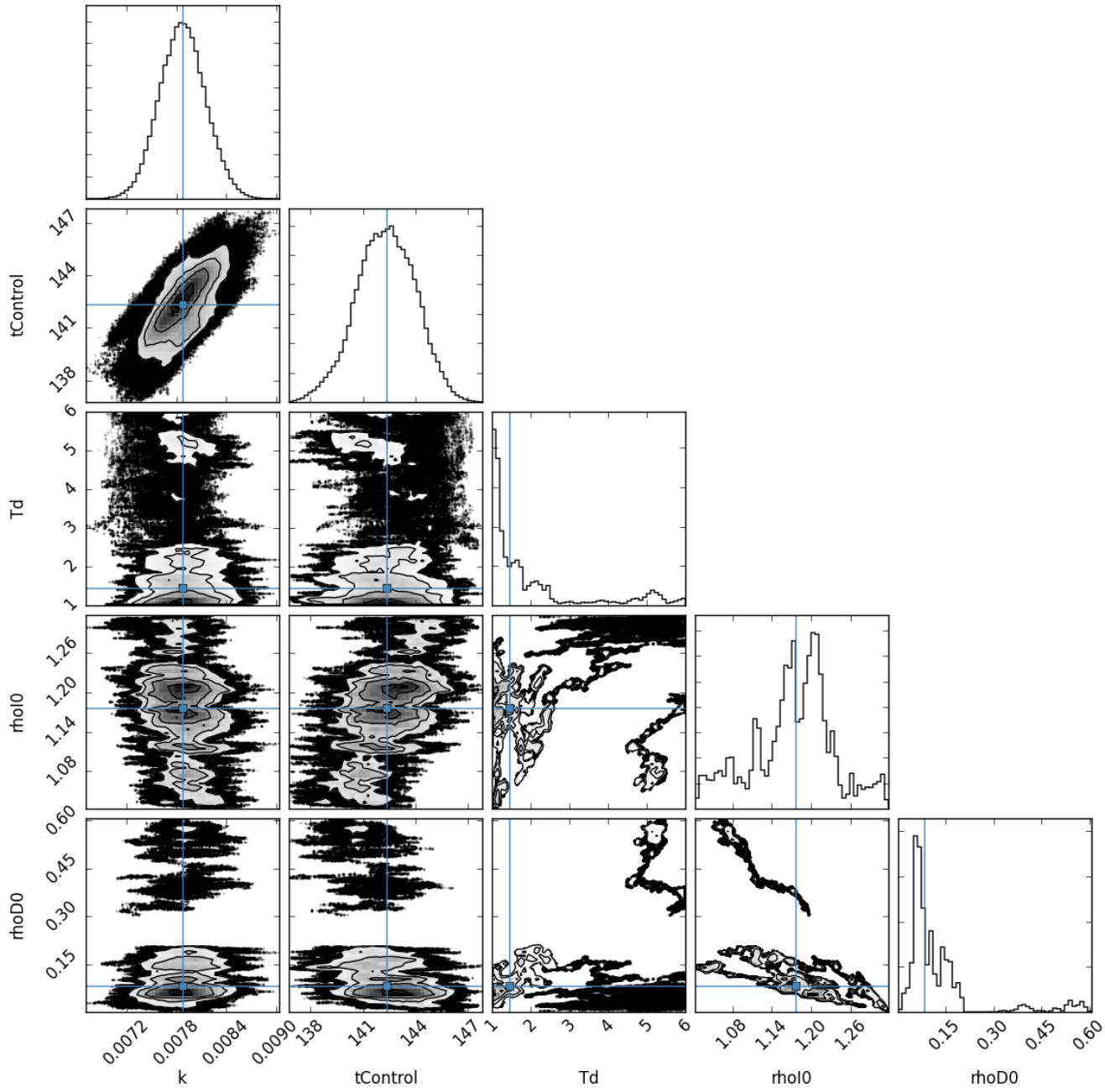

FIG. 2: Joint posteriors.