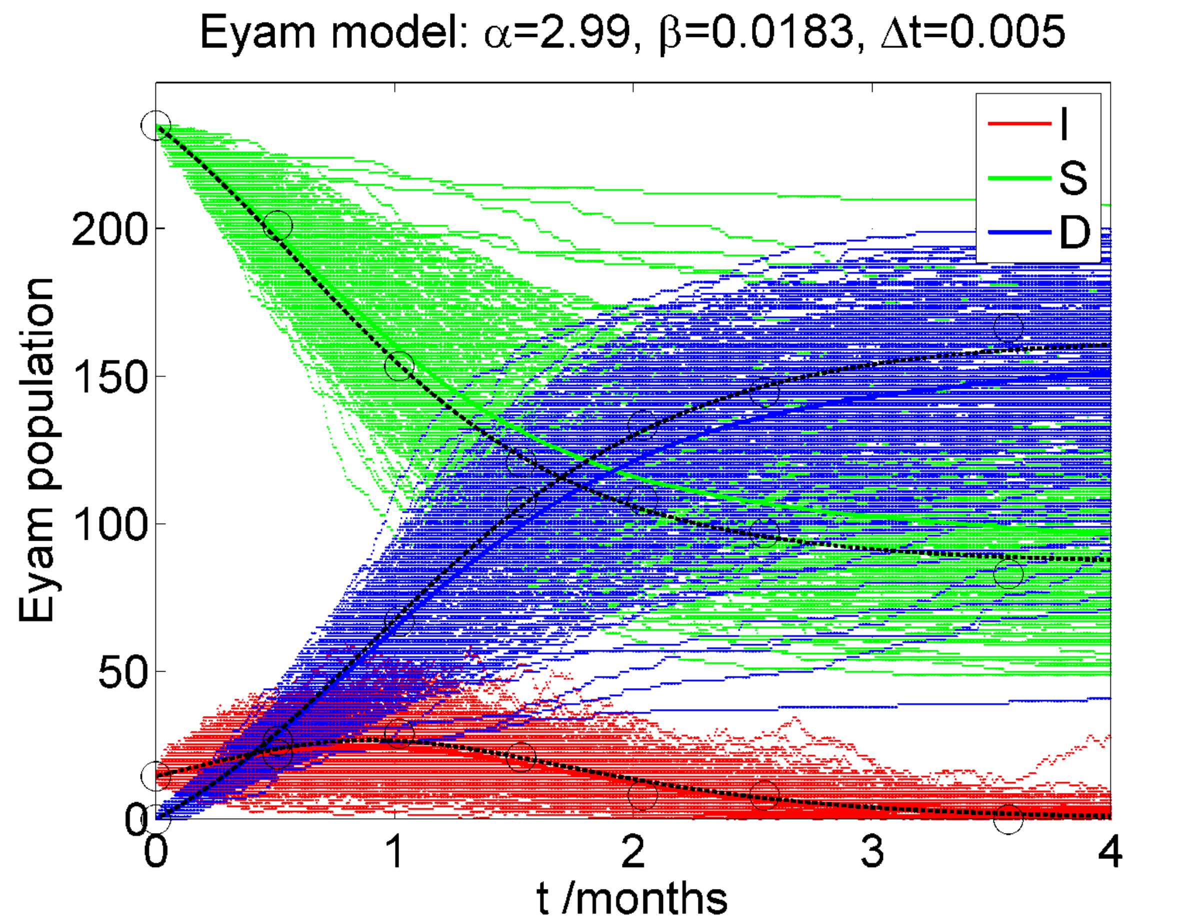
BPhO Computational Challenge

Seminar 11: The Epidemiology of Eyam

Dr Andrew French.
December 2021.

BPhO British Physics Olympiad

The Epidemiology of Eyam

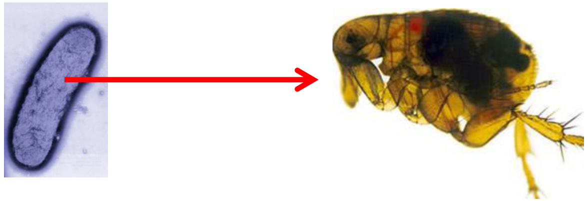


"How best to motivate students to expand their mathematical toolbox, and perhaps more importantly, gain experience of applying these ideas in the construction of quantitative models? A narrow focus on memorizing a long list of abstract procedures sufficient to pass an examination is a poor mechanism for producing the original thinkers of the future. It is also particularly harsh on those who have to struggle more than their peers to embed syllabus content in their minds. In this paper we celebrate the pedagogical power of context and storytelling, with the introduction of calculus ideas in an epidemiological scenario as an example."

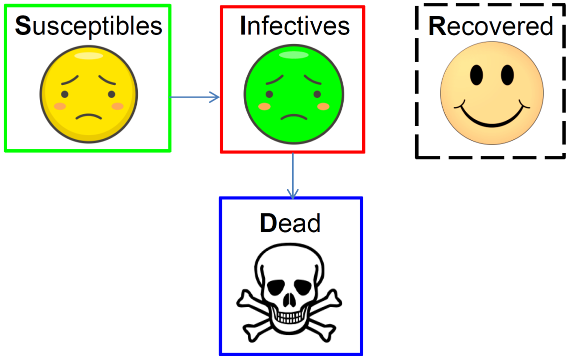
"The Pedagogical Power of Context: Iterative Calculus Methods and the Epidemiology of Eyam."
(French et al 2018 *J. Phys.Educ*.)

The Epidemiology of Eyam and the pedagogical power of context

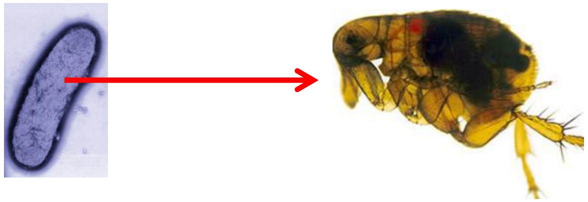
1. Context: The 1665 Plague of Eyam
2. The Eyam Equations
3. Iterative solution via the Euler numeric method
4. A semi-analytic solution, and Ebola
5. A stochastic model
6. COVID-19



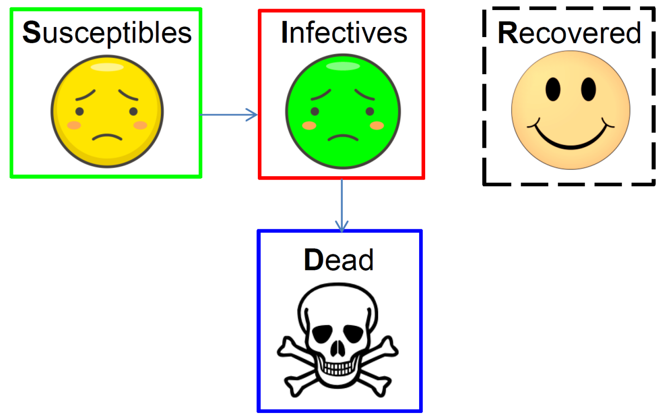
Context: The 1665 Plague of Eyam

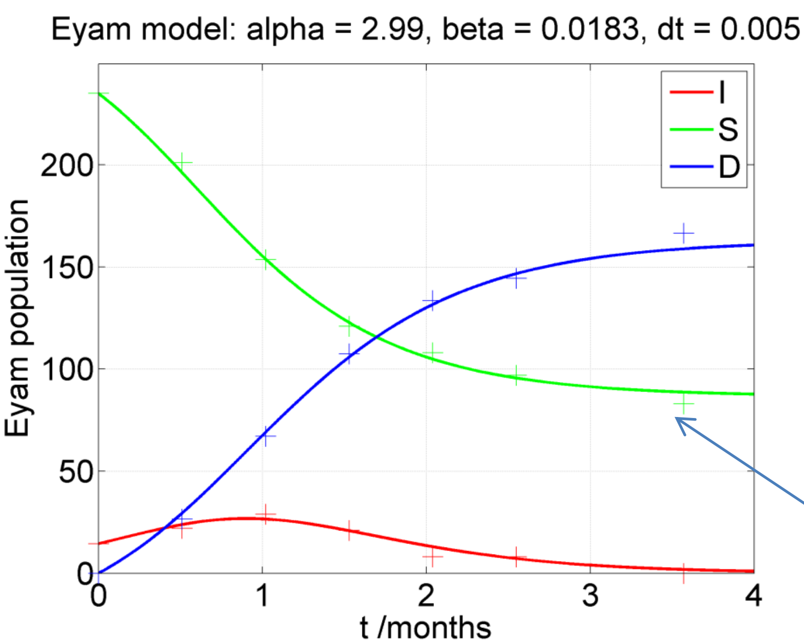


1665. A bale of damp cloth is delivered to the Derbyshire village of **Eyam**... George Viccars, the tailor's assistant, dries the cloth and releases fleas infected with *Yersinia Pestis* bacteria – **Plague**



Rector **William Mompesson** *quarantines* Eyam and records **Infected**, **Susceptible** and **Dead** populations *as time progresses*



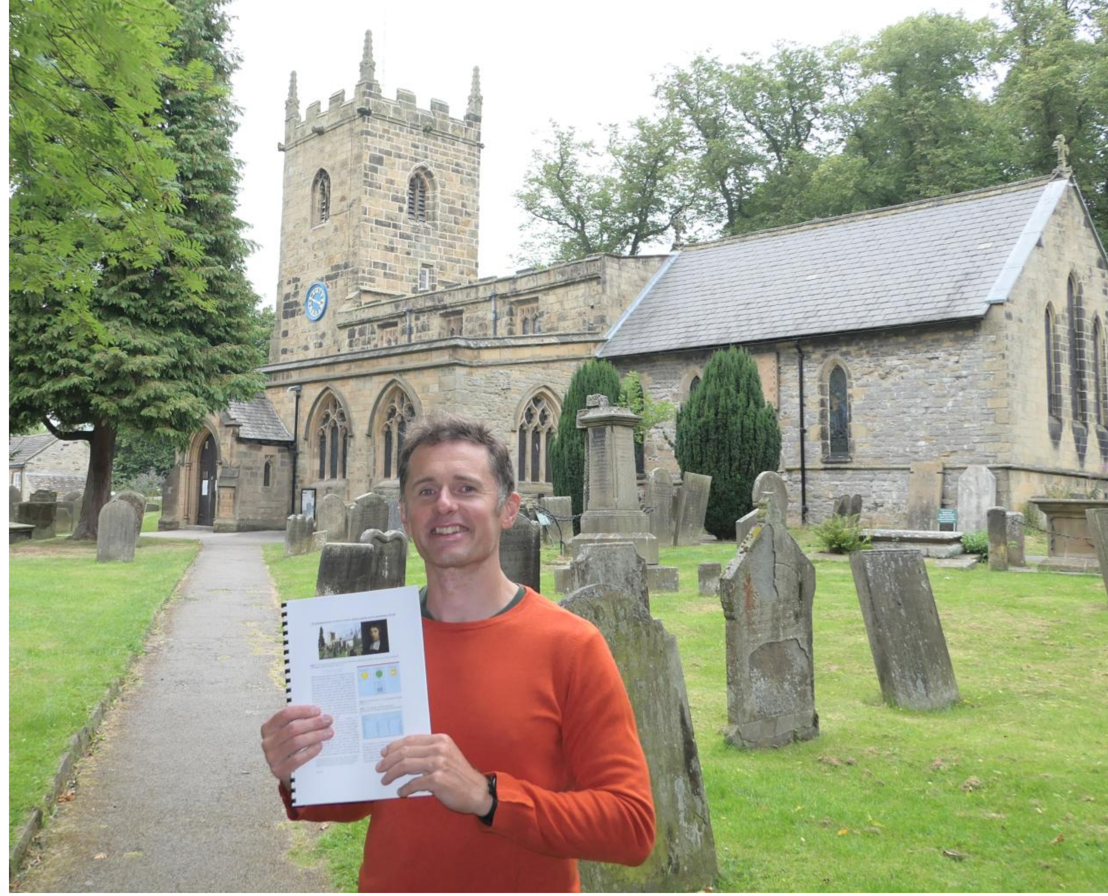


Can we develop a mathematical model to predict I,S,D vs time? What does this tell us about *Epidemiology* in general?

e.g Flu, Ebola

Calculus methods, differential equations numerical methods, line of best fit, iteration, loops ...

Yersinia Pestis



St Lawrence's Churchyard in Eyam
https://en.wikipedia.org/wiki/Eyam

A flea containing a blood meal infected with *Yersinia Pestis*
https://en.wikipedia.org/wiki/Yersinia_pestis

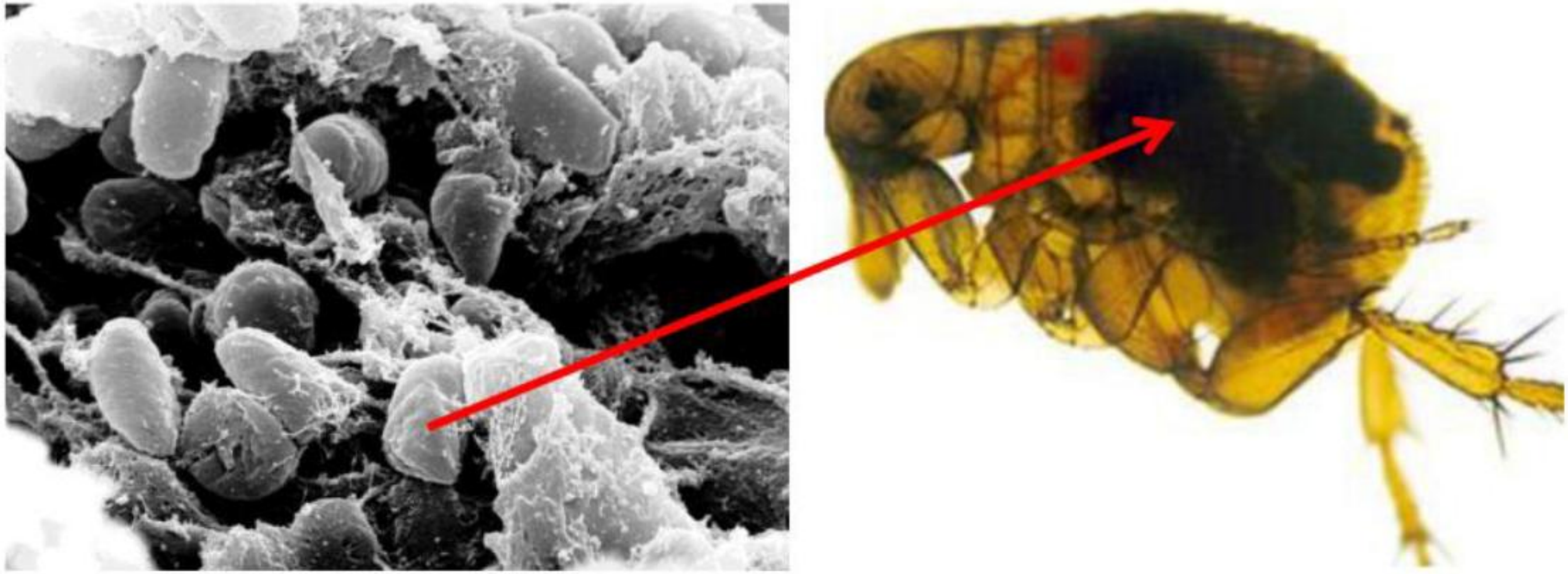


Figure 3. A flea containing a blood meal infected with the *Yersinia Pestis* bacterium (displayed at high magnification!) [13]

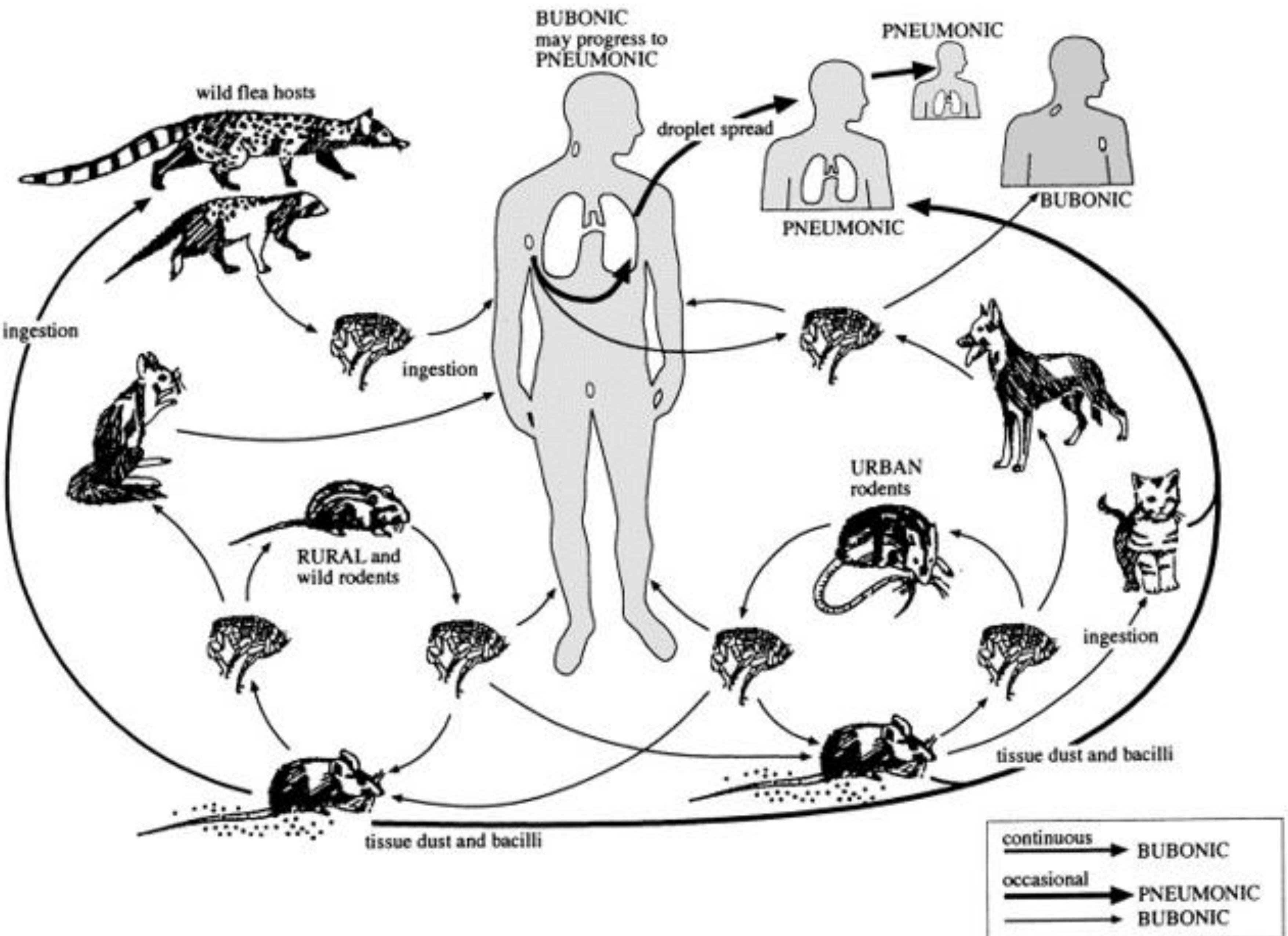
Yersinia Pestis

La Peste Bubonique à Hong-Kong
(Ann. Inst. Pasteur. 8: 662-667, 1894).

Alexandre Yersin 1863 - 1943

Unblocked, uninfected (panel A) and blocked, infected with an Hms+ *Y. pestis* strain, (panel B) *X. cheopis* fleas immediately after an uninfected blood meal. Bright red (fresh blood) throughout the digestive tract is indicative of unblocked fleas, while a dark-colored midgut due to digestion products from previous blood meals is diagnostic of proventricular blockage. Fresh blood in the esophagus of the blocked flea (panel B) shows that it recently attempted to feed.

Plague Pathways



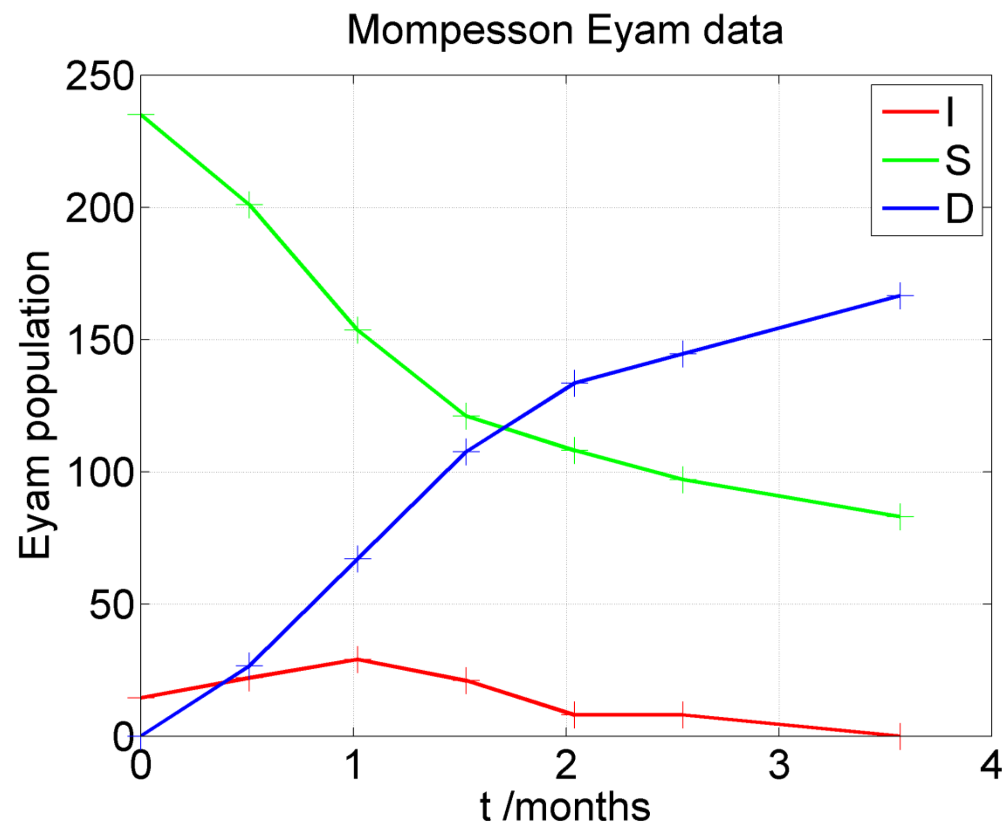
Buboes. A swollen inflamed lymph node in the armpit or groin.


Thankfully these are not real ...

Real buboes ...

Date	**Time /months**	S	I	D	$\ln(S_0/S)$	I_0+S_0-I-S
July 3-4 1666	0.00	235	14.5	0	0.00	0.00
July 19 1666	0.51	201	22	26.5	0.16	26.50
Aug 3-4 1666	1.02	153.5	29	67	0.43	67.00
Aug 19 1666	1.53	121	21	107.5	0.66	107.50
Sept 3-4 1666	2.04	108	8	133.5	0.78	133.50
Sept 19 1666	2.55	97	8	144.5	0.88	144.50
Oct 20 1666	3.57	83	0	166.5	1.04	166.50

Rev. William Mompesson
1639-1709



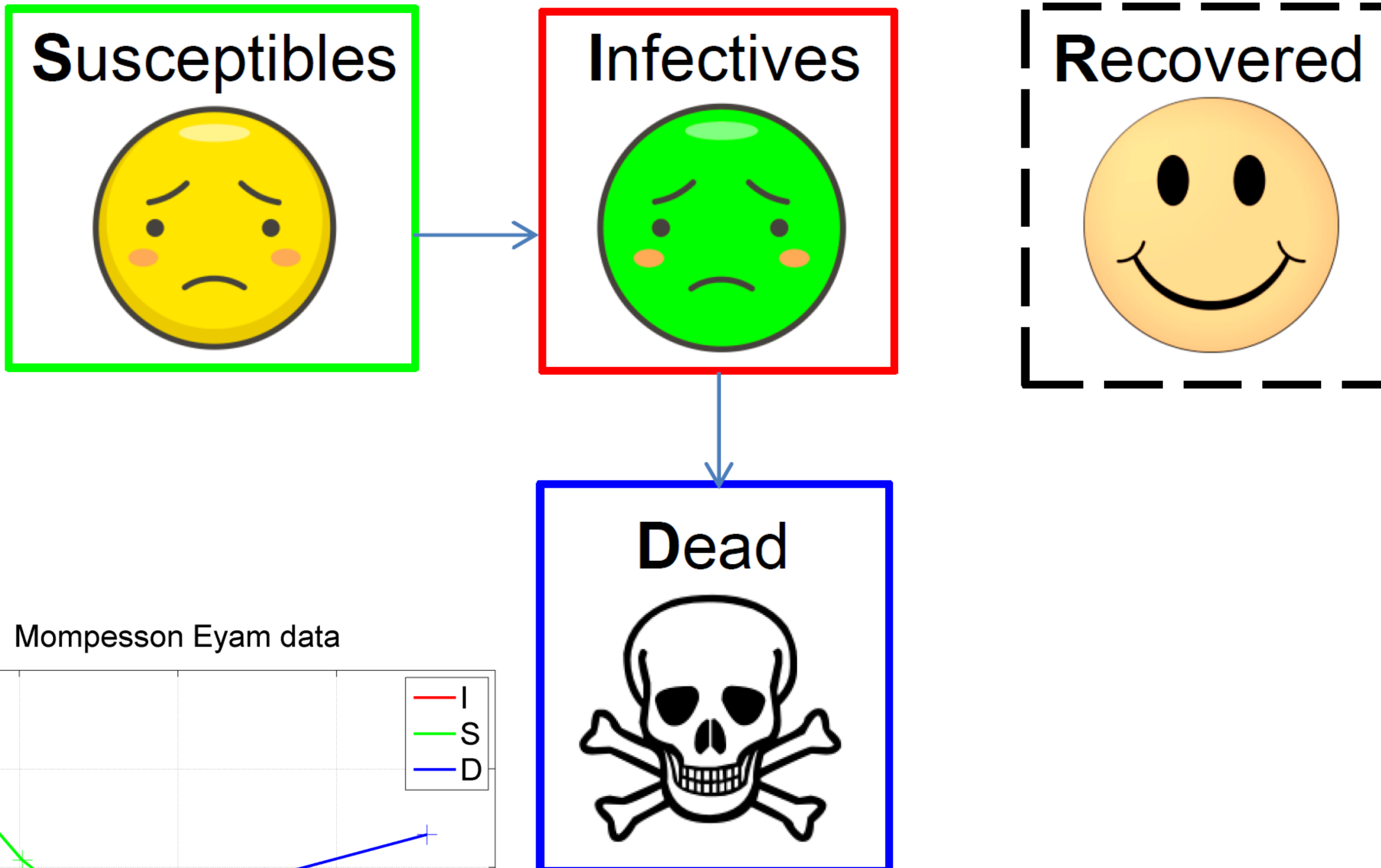
$$\boxed{\frac{d(R+D)}{dt} = \alpha I}$$

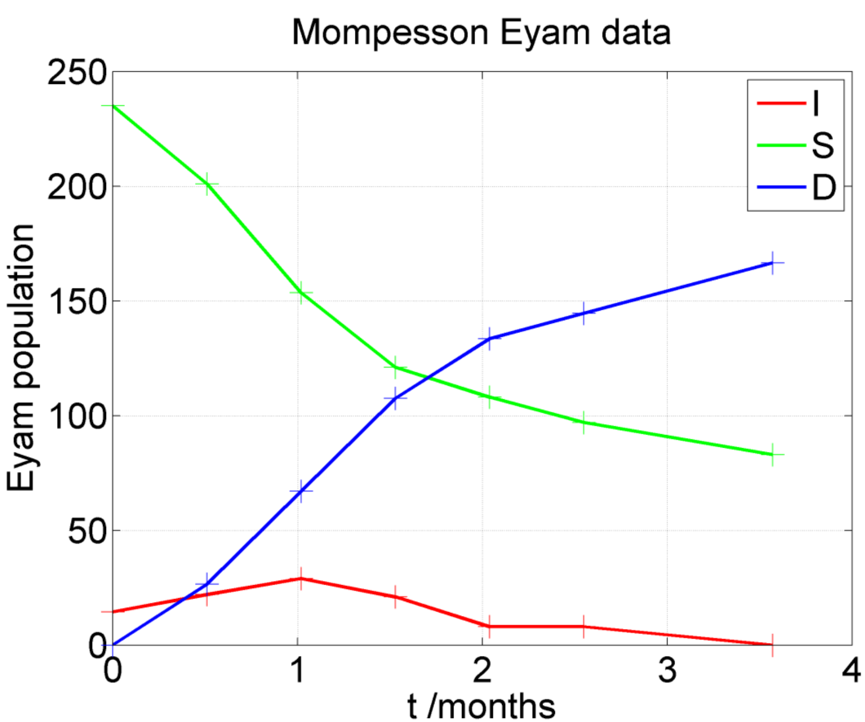
$$\boxed{\frac{dS}{dt} = -\beta SI}$$

The Eyam Equations

$$\boxed{S + I + D + R = \text{constant}}$$

$$\boxed{\frac{dI}{dt} = \beta SI - \alpha I}$$





What is the simplest model that describes the key features of Mompesson's data?

A population 'flow' from **S**usceptibles, to **I**nfectives, to **D**ead.

$$\boxed{\frac{dD}{dt} = \alpha I}$$

$$\boxed{\frac{dS}{dt} = -\beta SI}$$

D Dead

S Susceptible

I Infective

Assume $R = 0$. Or to generalize let D be $k(R + D)$

$$I + S + D = I_0 + S_0$$ Fixed population constraint

$$\therefore \frac{dI}{dt} + \frac{dS}{dt} + \frac{dD}{dt} = 0 \qquad \therefore \frac{dI}{dt} = -\frac{dS}{dt} - \frac{dD}{dt}$$

$$\therefore \boxed{\frac{dI}{dt} = \beta SI - \alpha I}$$

Eyam *Epidemiological* model

The Eyam equation for Infectives I is:

$$\frac{dI}{dt} = (\beta S - \alpha) I$$

$$\boxed{\frac{dI}{dt} = (\beta S - \alpha) I}$$

It is immediately apparent that $\frac{dI}{dt} = 0$ if $I = 0$ or $S = \frac{\alpha}{\beta}$. By performing a further time derivative, one can see that I is *maximized* when $S = \frac{\alpha}{\beta}$. This is the Susceptible population at the peak of the infection.

$$\frac{d^2 I}{dt^2} = (\beta S - \alpha)\frac{dI}{dt} + I\beta \frac{dS}{dt}$$

$$\therefore \frac{d^2 I}{dt^2} = (\beta S - \alpha)^2 I - I^2 \beta^2 S$$

$$\therefore \left.\frac{d^2 I}{dt^2}\right|_{S=\frac{\alpha}{\beta}} = \left(\beta \tfrac{\alpha}{\beta} - \alpha\right)^2 I - I^2 \beta^2 \tfrac{\alpha}{\beta} = -I^2 \beta\alpha$$

$$\therefore \left.\frac{d^2 I}{dt^2}\right|_{S=\frac{\alpha}{\beta}} < 0$$

$\rho = \frac{\alpha}{\beta}$ Susceptible **threshold**

$S > \rho$ Epidemic grows

$S < \rho$ Epidemic shrinks

Iterative solution via the Euler numeric method

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta SI - \alpha I$$

$$\frac{dD}{dt} = \alpha I$$

Leonhard Euler
1707-1783

Euler numerical *iterative* solution scheme

See later on how we worked out α

$$\alpha = 2.894, \quad \beta = \frac{\alpha}{163.3}$$

$$t_0 = 0, \; S_0 = 235, \; I_0 = 14.5, \; D_0 = 0$$

$$t_{n+1} = t_n + \Delta t$$

$$S_{n+1} = S_n - \beta S_n I_n \Delta t$$

$$I_{n+1} = I_n + \left(\beta S_n I_n - \alpha I_n\right)\Delta t$$

$$D_{n+1} = D_n + \alpha I_n \Delta t$$

```matlab
%Euler method solver for differential equations which
%describe model of Eyam epidemic.
function [t,I,S,D] = eyam_model( dt, I0, S0, alpha, beta, tmax )

%Initialize output vectors for t,I,S,D
t = 0 : dt : tmax;
N = length(t);
S = S0*ones(1,N);
I = I0*ones(1,N);
D = zeros(1,N);

%Loop through vectors to compute t, I, S, D.
%using the Euler first order differential equation method
for n=2:N
    t(n) = t(n-1) + dt;
    I(n) = I(n-1) + dt*( beta*S(n-1)*I(n-1) - alpha*I(n-1) );
    S(n) = S(n-1) - dt*beta*S(n-1)*I(n-1);
    D(n) = D(n-1) + dt*alpha*I(n-1);
end
```

$$\alpha = 2.894, \quad \beta = \frac{\alpha}{163.3}$$

$$t_0 = 0, \; S_0 = 235, \; I_0 = 14.5, \; D_0 = 0$$

$$t_{n+1} = t_n + \Delta t$$

$$S_{n+1} = S_n - \beta S_n I_n \Delta t$$

$$I_{n+1} = I_n + \left(\beta S_n I_n - \alpha I_n\right)\Delta t$$

$$D_{n+1} = D_n + \alpha I_n \Delta t$$

MATLAB code for Euler Eyam model

Eyam Equations

$$\boxed{\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dD}{dt} &= \alpha I \\ \frac{dI}{dt} &= \beta SI - \alpha I\end{aligned}}$$

$$\frac{dS}{dt} = -\beta SI, \quad \frac{dI}{dt} = \beta SI - \alpha I$$

$$\therefore \frac{dI}{dS} = -\frac{\beta SI - \alpha I}{\beta SI} = \tfrac{\alpha}{\beta}\frac{1}{S} - 1$$

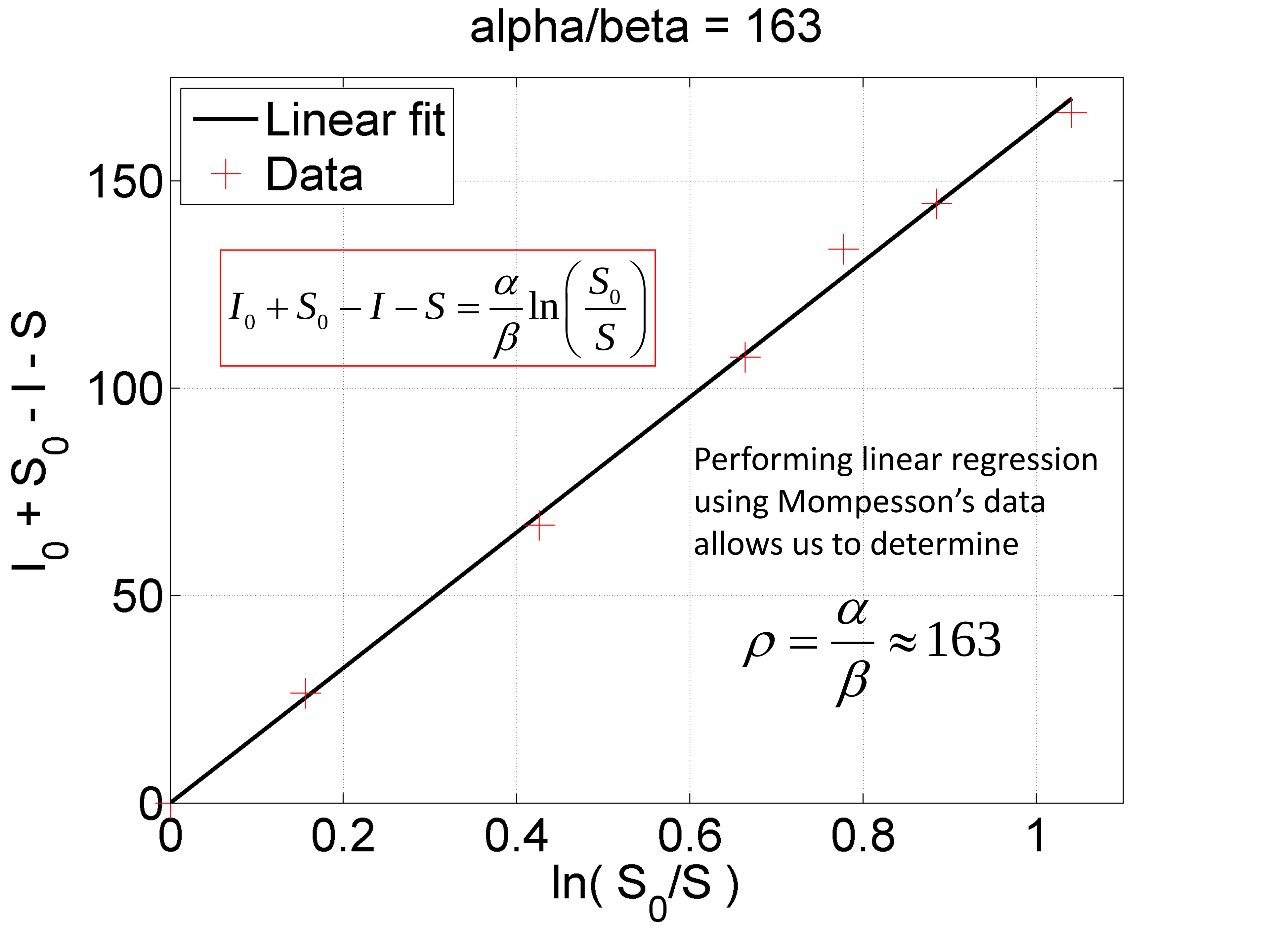
$$\therefore I - I_0 = \int_{S_0}^{S} \left(\tfrac{\alpha}{\beta}\frac{1}{S} - 1 \right) dS = \left[\tfrac{\alpha}{\beta} \ln S - S \right]_{S_0}^{S}$$

$$I = I_0 + \tfrac{\alpha}{\beta} \ln \frac{S}{S_0} - S + S_0$$

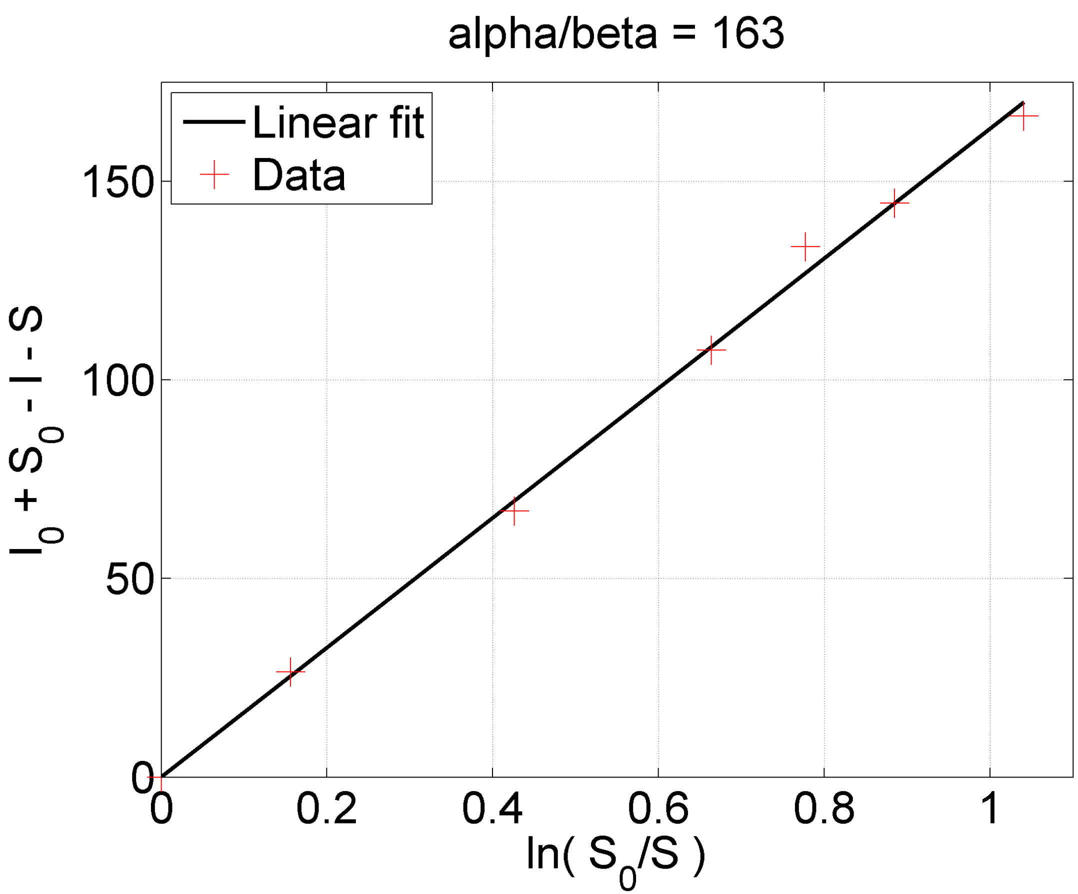
$$\boxed{\begin{aligned} &\tfrac{\alpha}{\beta} \underbrace{\ln \frac{S_0}{S}}_{x} = \underbrace{I_0 + S_0 - I - S}_{y} \\ &\therefore y = \tfrac{\alpha}{\beta} x \end{aligned}}$$

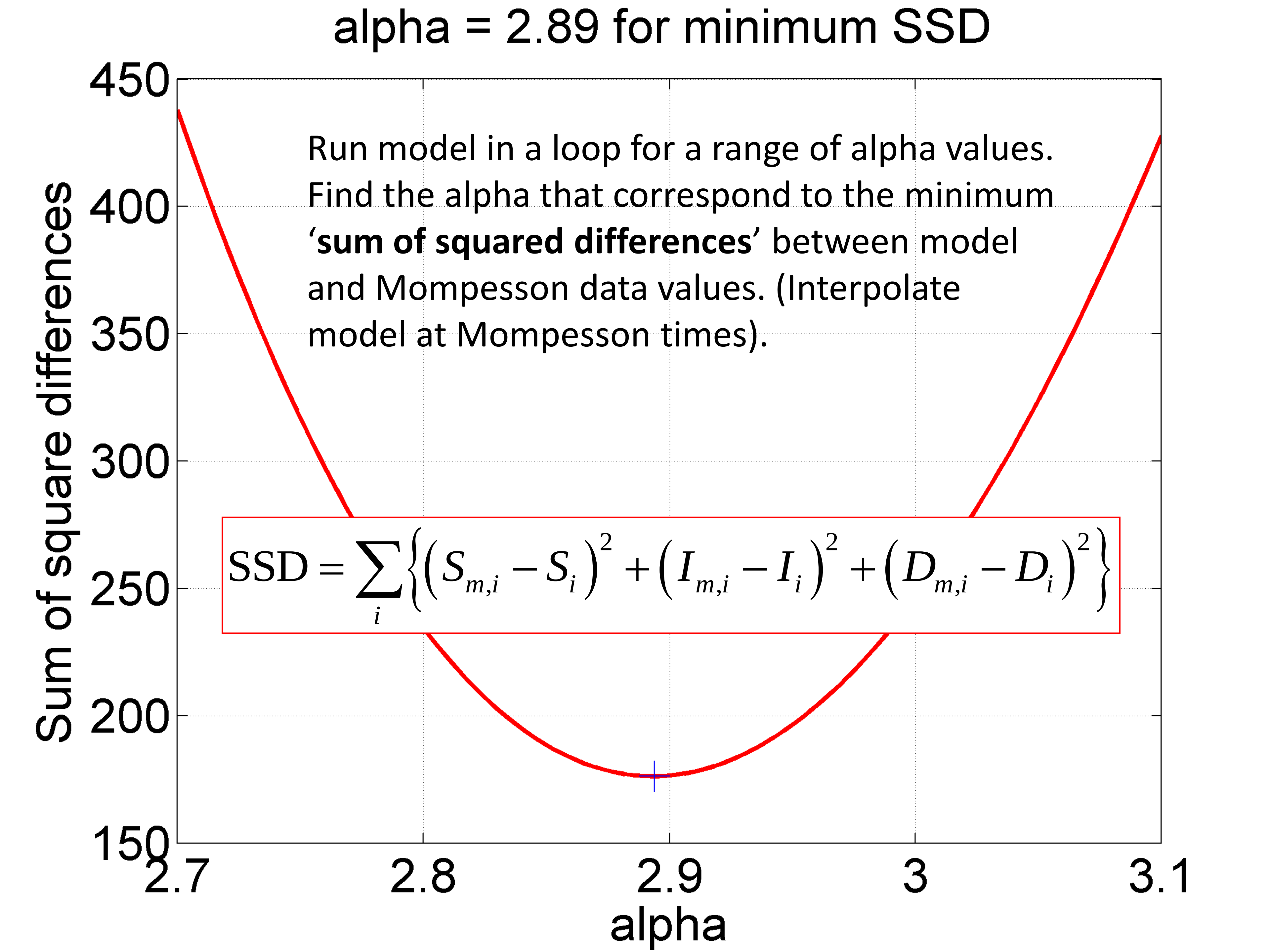
Note we can integrate to find *I(S)* analytically

.... But not *I(t)*, *S(t)*, *D(t)*

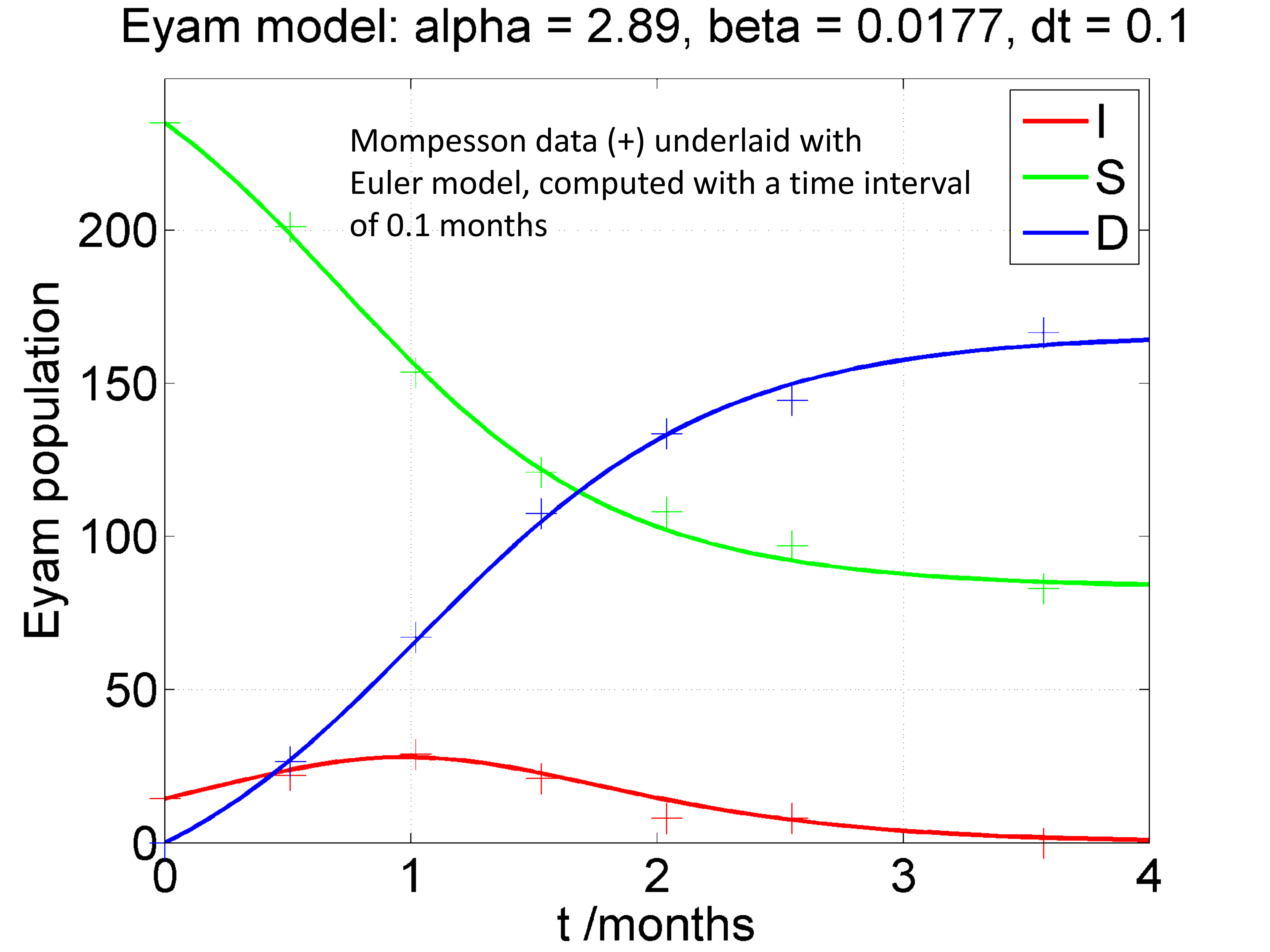

alpha/beta = 163
Linear fit
Data
$I_0 + S_0 - I - S = \frac{\alpha}{\beta}\ln\left(\frac{S_0}{S}\right)$
Performing linear regression using Mompesson's data allows us to determine
$\rho = \frac{\alpha}{\beta} \approx 163$
$I_0 + S_0 - I - S$
$\ln(S_0/S)$
0
50
100
150
0.2
0.4
0.6
0.8
1

```
%Line of best fit function yfit = m*x, with product moment correlation
%coefficient r
function [yfit,xfit,r,m] = bestfit(x,y)

%Find any x or y values that are NaN or Inf
ignore = isnan(abs(x)) | isnan(abs(y)) | isinf(abs(x)) | isinf(abs(y));
x(ignore) = [];
y(ignore) = [];

%Compute line of best fit
xybar = mean(x.*y);
xxbar = mean(x.^2 );
yybar = mean(y.^2 );
m = xybar/xxbar;
r = xybar/( xxbar*yybar );
yfit = m*x;
xfit = x;
```





alpha = 2.89 for minimum SSD
Sum of square differences
alpha
450
400
350
300
250
200
150
2.7
2.8
2.9
3
3.1
Run model in a loop for a range of alpha values. Find the alpha that correspond to the minimum '**sum of squared differences**' between model and Mompesson data values. (Interpolate model at Mompesson times).
$$\mathrm{SSD} = \sum_i \left\{ \left(S_{m,i} - S_i\right)^2 + \left(I_{m,i} - I_i\right)^2 + \left(D_{m,i} - D_i\right)^2 \right\}$$


Eyam model: alpha = 2.89, beta = 0.0177, dt = 0.1
Mompesson data (+) underlaid with
Euler model, computed with a time interval
of 0.1 months
I
S
D
Eyam population
t /months
0
50
100
150
200
0
1
2
3
4

Note $1/\alpha$ is a measure of a **time constant** for the Eyam plague.

In *days* it is:

$$\boxed{\tau = \frac{1}{\alpha} = \frac{365}{12} \times \frac{1}{2.894} = 10.5}$$

This could be used as a crude measure of ‘fatality time’ – i.e. an approximate number of days from infection till death.

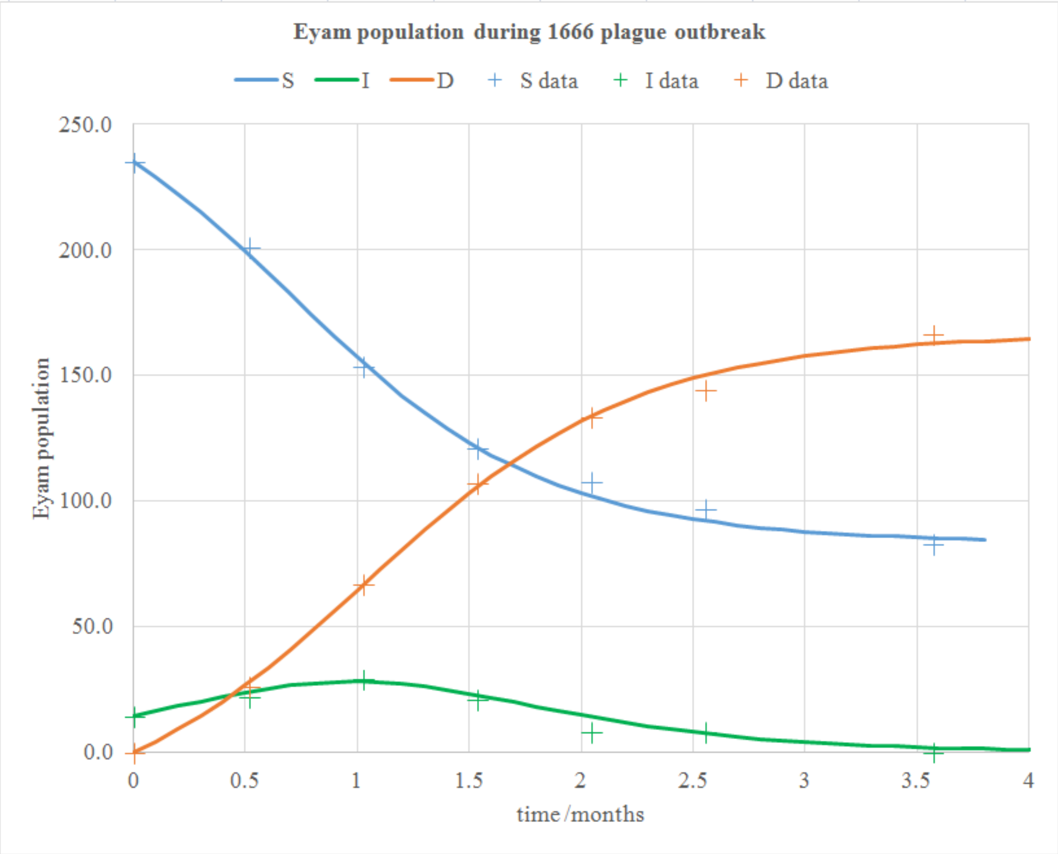
We performed the Eyam analysis in **Python**, then in **MATLAB**.
You can also construct an Euler model via a spreadsheet (**Excel)**.

Black Death Epidemiological model using the Eyam data

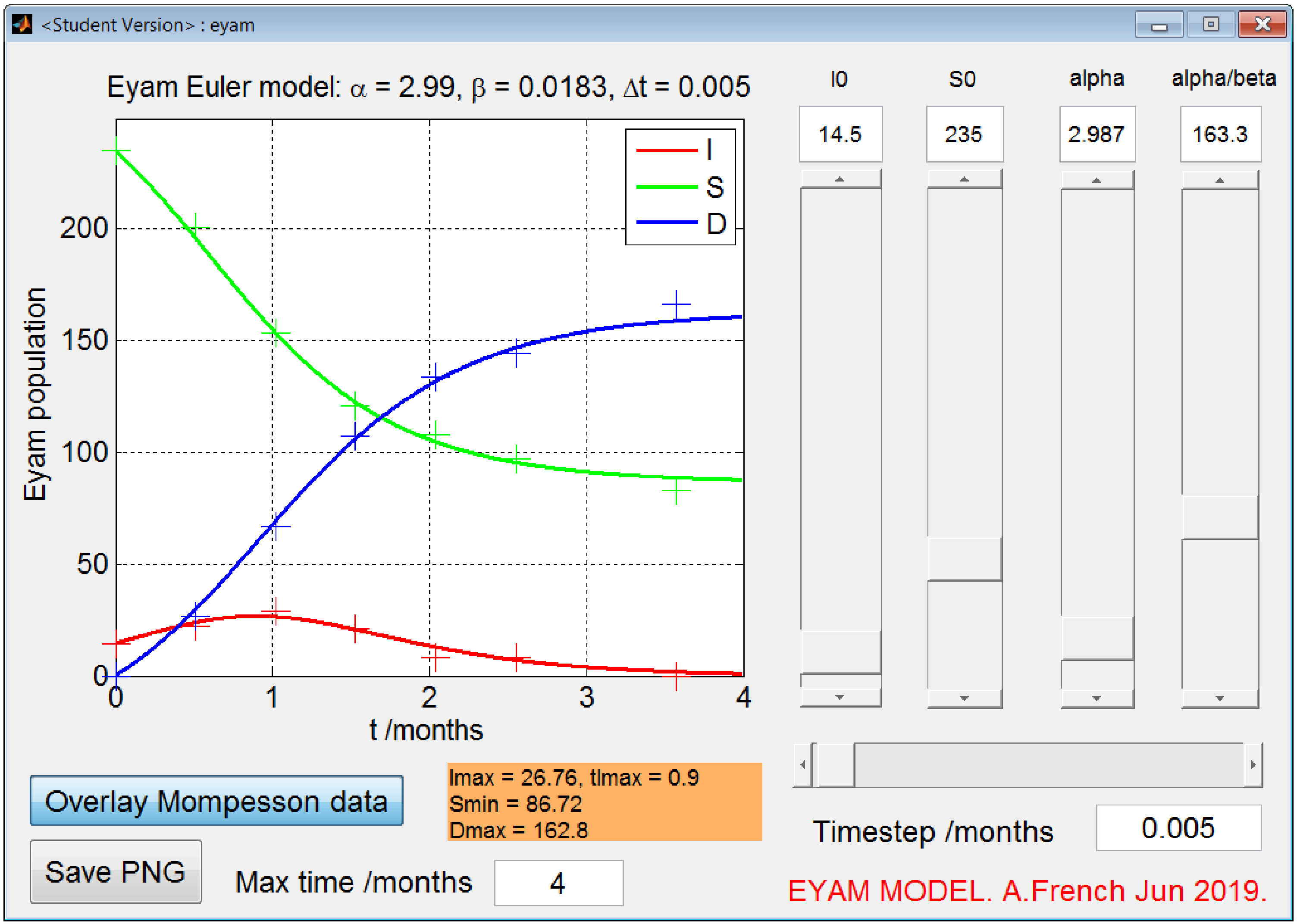
Andy French & John Cullerne. 24th February 2018.

Initial population N0	249.5
Initial number of succeptables S0	**235**
Initial number of infectives I0	**14.5**
Transmission rate constant beta	0.017759
Death rate constant alpha	**2.9**
timestep dt /months	**0.1**

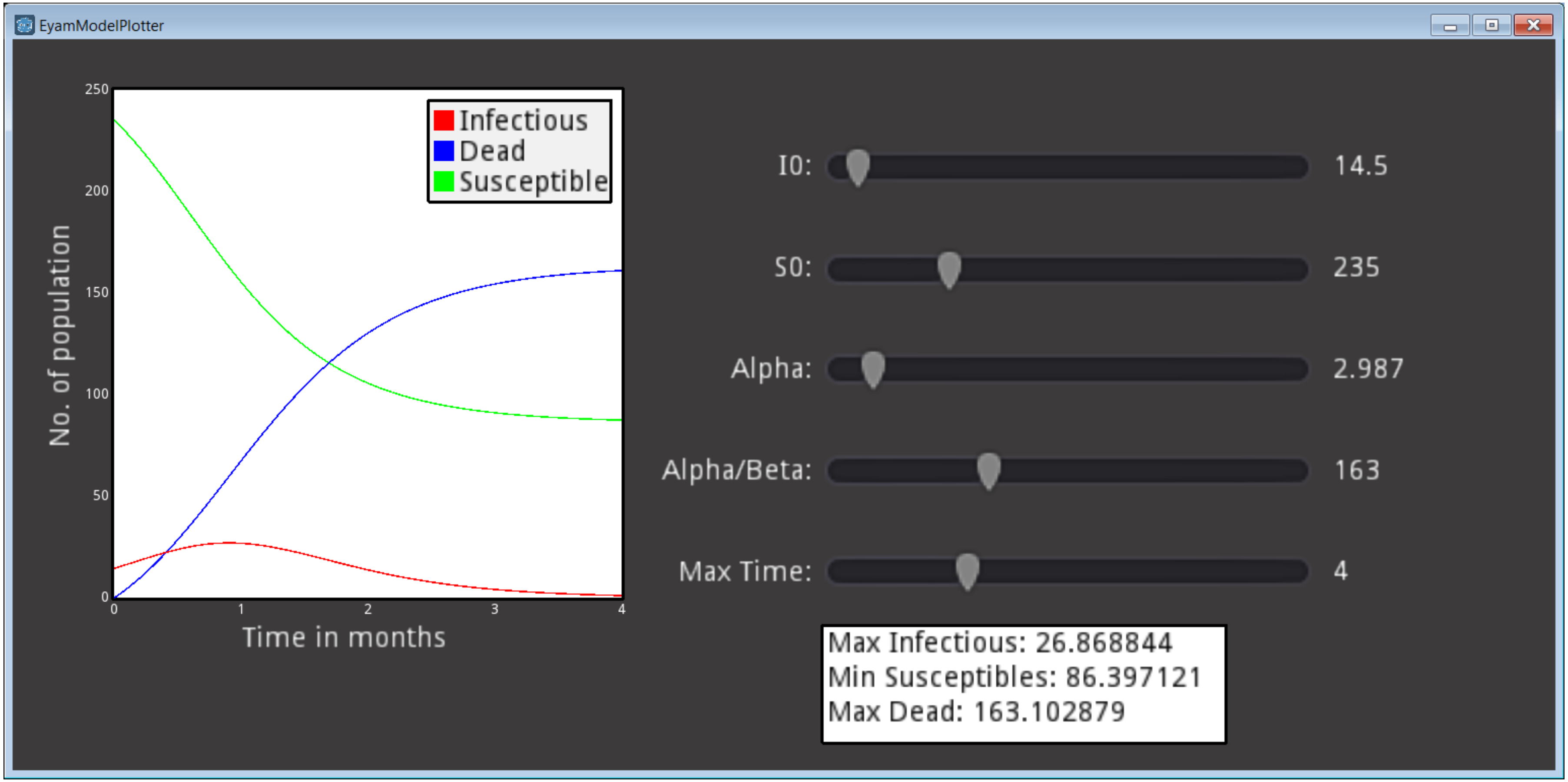
t /months	S	I	D	N	N+D = N0
0	235.0	14.5	0.0	249.5	249.5
0.1	228.9	16.3	4.2	245.3	249.5
0.2	222.3	18.3	8.9	240.6	249.5
0.3	215.1	20.2	14.2	235.3	249.5
0.4	207.4	22.0	20.1	229.4	249.5
0.5	199.3	23.7	26.5	223.0	249.5
0.6	190.9	25.3	33.4	216.1	249.5
0.7	182.3	26.5	40.7	208.8	249.5
0.8	173.7	27.4	48.4	201.1	249.5
0.9	165.3	27.9	56.3	193.2	249.5
1	157.1	28.0	64.4	185.1	249.5
1.1	149.3	27.7	72.5	177.0	249.5
1.2	141.9	27.0	80.6	168.9	249.5
1.3	135.1	26.0	88.4	161.1	249.5
1.4	128.9	24.7	95.9	153.6	249.5
1.5	123.3	23.2	103.1	146.4	249.5
1.6	118.2	21.5	109.8	139.7	249.5



Euler Eyam solver implemented in MATLAB with a Graphical User Interface (GUI).
Change the inputs via the sliders or edit boxes, and the curves are computed automatically.



Implementation of an Eyam model GUI by Barton Peveril student Alfie Baxter using the *Game Engine* development environment.



A semi-analytic solution, and Ebola

$$z_+ = -\ln(1-\eta) - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$z_- = -\ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$x_{\max} = -\frac{\ln(1-\eta)}{\eta} - 1 - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$\rho = \frac{I_{\max}}{x_{\max}}$$

$$\tau(z) = \int_0^z \frac{dz'}{x_{\max} + 1 - e^{-z'} - z'}$$

$$x = x_{\max} + 1 - e^{-z} - z$$

$$y = e^{-z}$$

$$t = \frac{\tau}{\alpha} + t_{\max},\ \ I = \rho x\ \ S = \rho y,\ \ D = \rho(z - z_-)$$

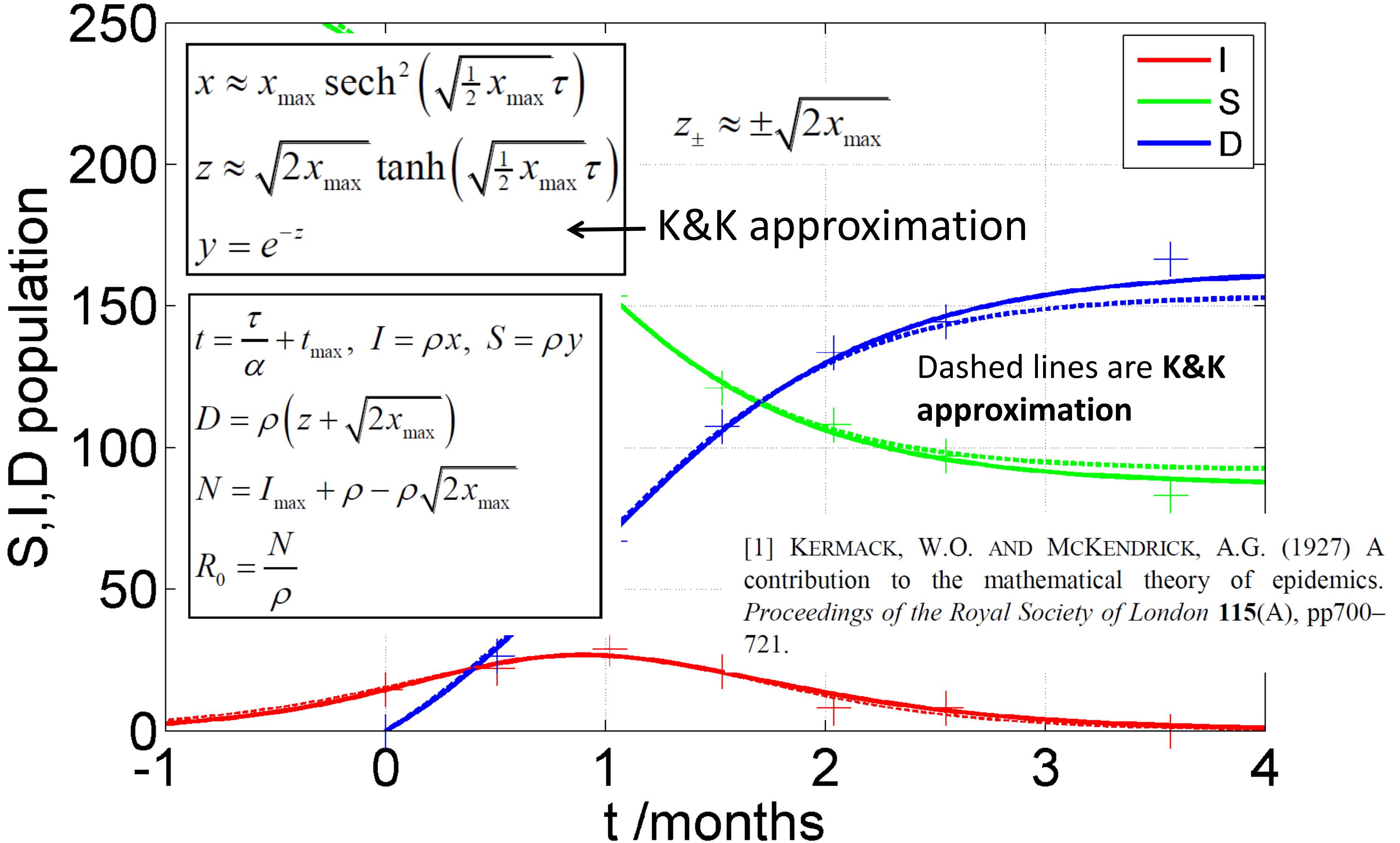
$$N = I_{\max} + \rho - \rho z_-$$

$$R_0 = \frac{N}{\rho}$$

Eyam model fit

N=275, I_{max}=26.76, R_0=1.68, t_{max}=0.9

α=2.987, ρ=163.3, η=0.6851, S_0=235, I_0=14.48



$$x \approx x_{\max} \operatorname{sech}^2\left(\sqrt{\tfrac{1}{2} x_{\max}}\, \tau\right)$$

$$z \approx \sqrt{2x_{\max}} \tanh\left(\sqrt{\tfrac{1}{2} x_{\max}}\, \tau\right)$$

$$y = e^{-z}$$

$$z_{\pm} \approx \pm\sqrt{2x_{\max}}$$

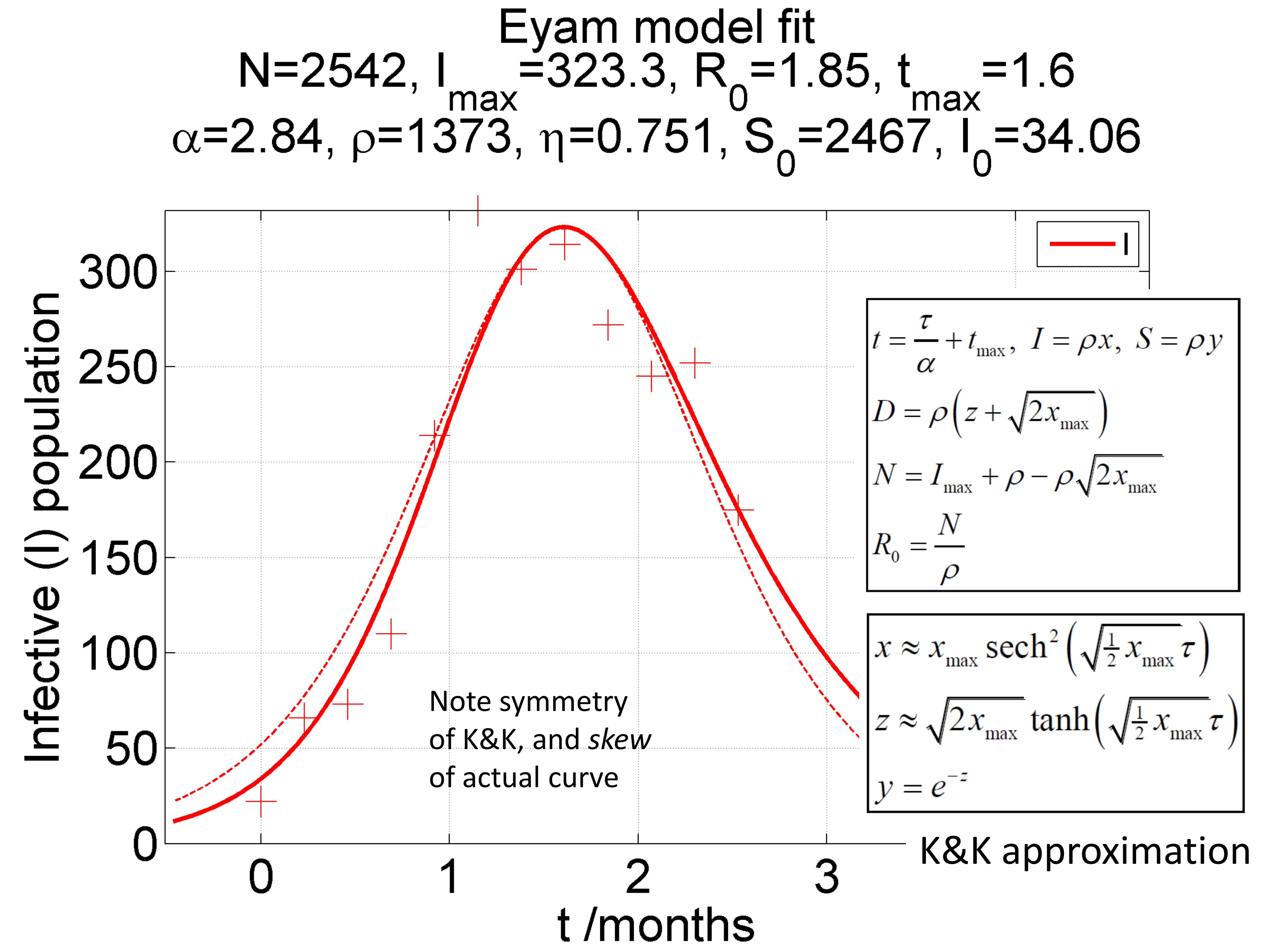
$$t = \frac{\tau}{\alpha} + t_{\max}, \ \ I = \rho x, \ \ S = \rho y$$

$$D = \rho\left(z + \sqrt{2x_{\max}}\right)$$

$$N = I_{\max} + \rho - \rho\sqrt{2x_{\max}}$$

$$R_0 = \frac{N}{\rho}$$

[1] Kermack, W.O. and McKendrick, A.G. (1927) A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London* **115**(A), pp700–721.


Eyam model fit
N=2542, I_{max}=323.3, R_0=1.85, t_{max}=1.6
α=2.84, ρ=1373, η=0.751, S_0=2467, I_0=34.06
Infective (I) population
300
250
200
150
100
50
0
0
1
2
3
t /months
I
$t = \frac{\tau}{\alpha} + t_{max}, \; I = \rho x, \; S = \rho y$
$D = \rho\left(z + \sqrt{2x_{max}}\right)$
$N = I_{max} + \rho - \rho\sqrt{2x_{max}}$
$R_0 = \frac{N}{\rho}$
$x \approx x_{max}\,\mathrm{sech}^2\left(\sqrt{\tfrac{1}{2}x_{max}}\,\tau\right)$
$z \approx \sqrt{2x_{max}}\,\tanh\left(\sqrt{\tfrac{1}{2}x_{max}}\,\tau\right)$
$y = e^{-z}$
K&K approximation
Note symmetry of K&K, and *skew* of actual curve

$$z_{+} = -\ln(1-\eta) - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$z_{-} = -\ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$x_{\max} = -\frac{\ln(1-\eta)}{\eta} - 1 - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$\rho = \frac{I_{\max}}{x_{\max}}$$

$$\tau(z) = \int_0^z \frac{dz'}{x_{\max} + 1 - e^{-z'} - z'}$$

$$x = x_{\max} + 1 - e^{-z} - z$$

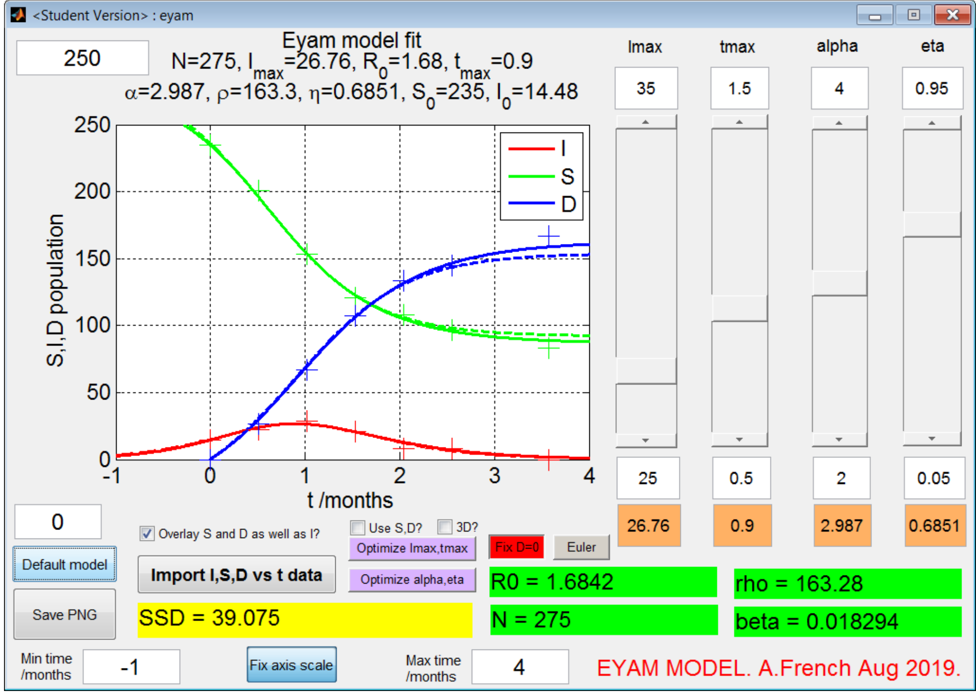
$$y = e^{-z}$$

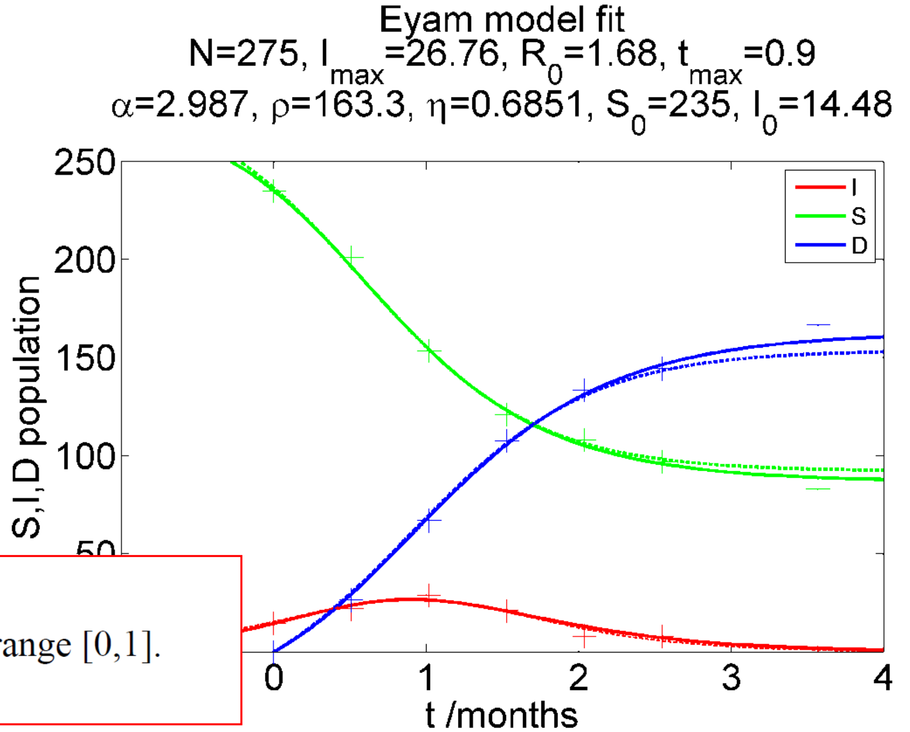
$$t = \frac{\tau}{\alpha} + t_{\max}, \quad I = \rho x \quad S = \rho y, \quad D = \rho(z - z_{-})$$

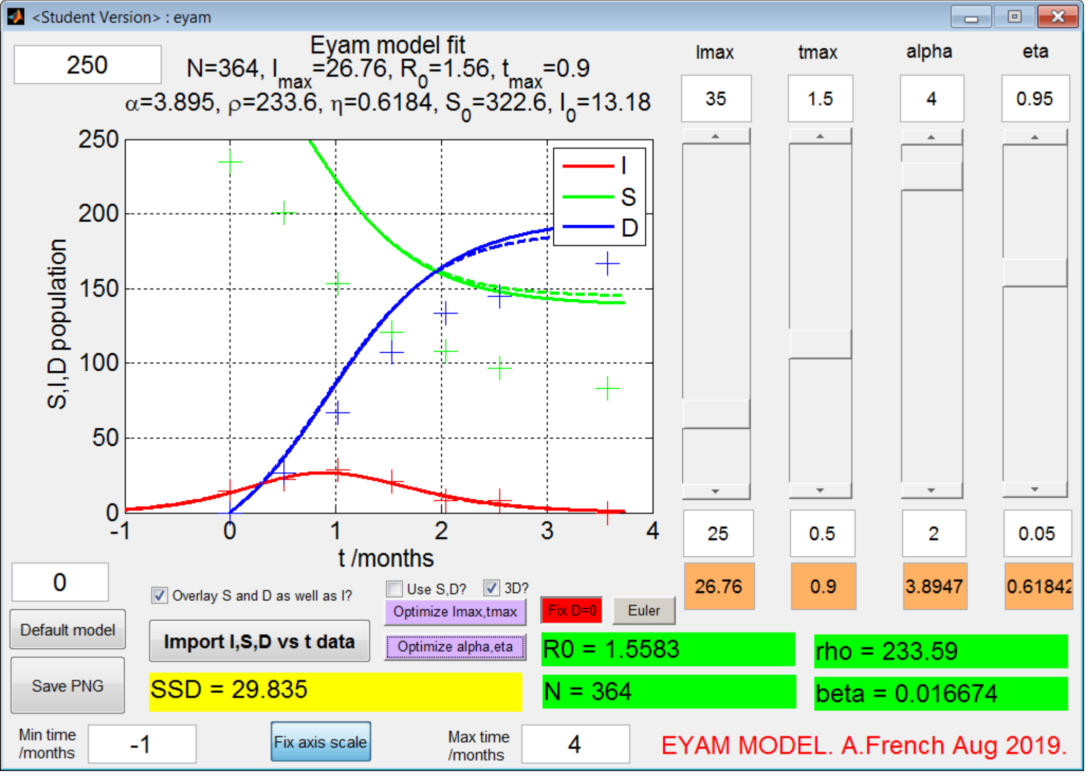
$$N = I_{\max} + \rho - \rho z_{-}$$

$$R_0 = \frac{N}{\rho}$$

Define $\boxed{\eta = \frac{z_{+} - z_{-}}{N/\rho}}$, which *must* be in the range [0,1].

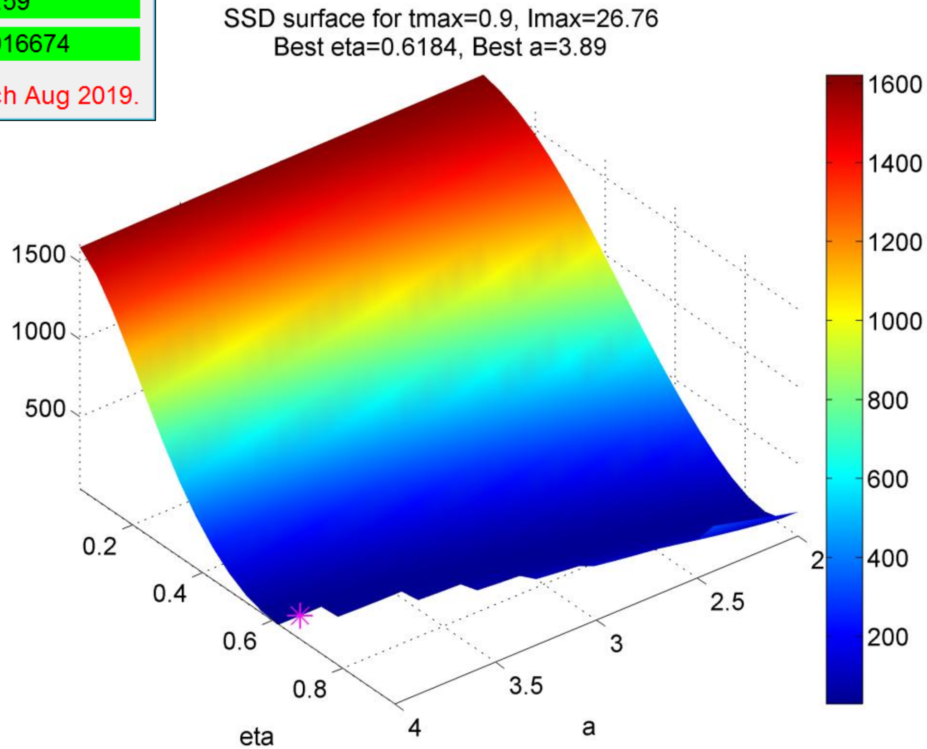






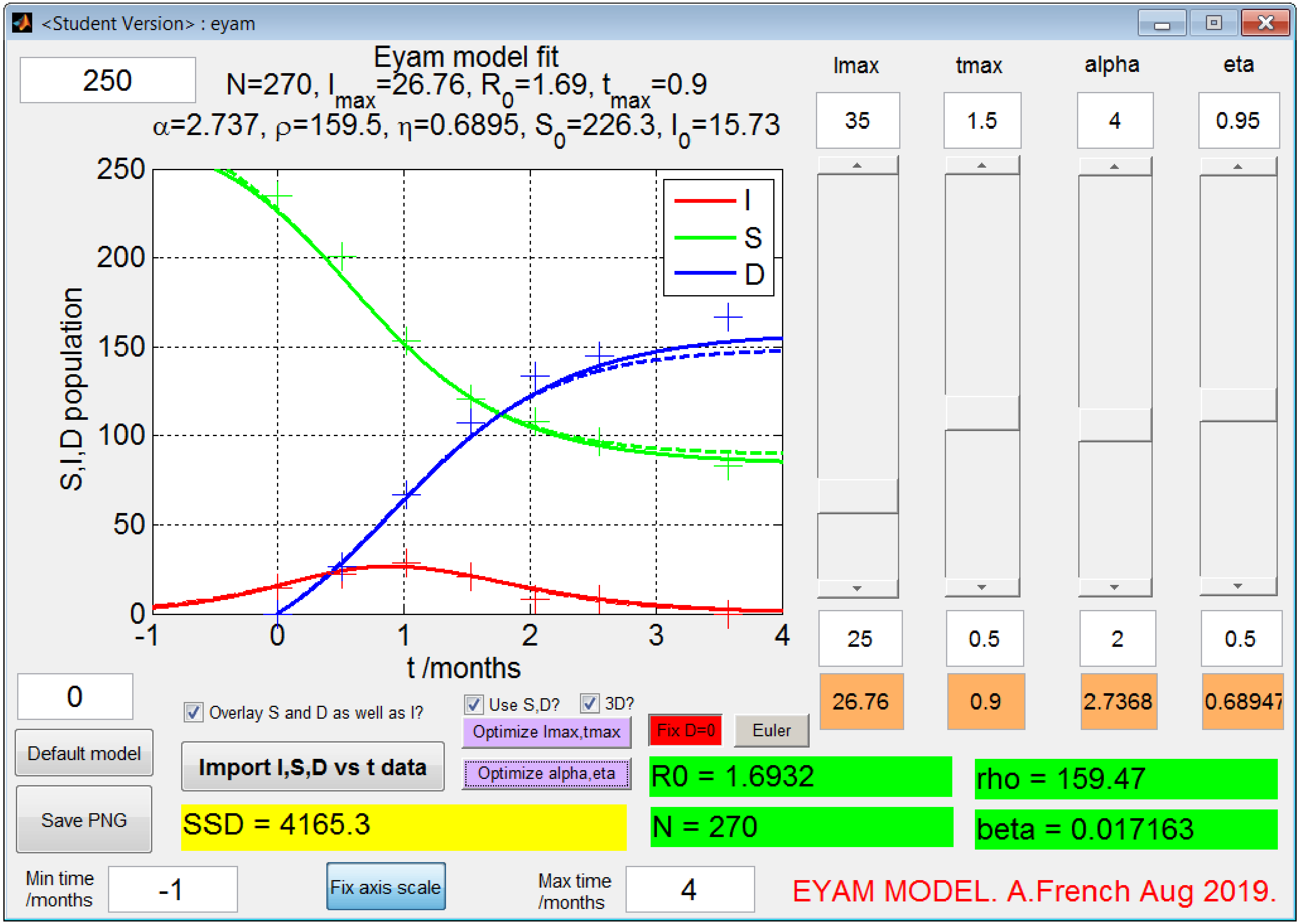
Run model over a range of alpha and eta parameters and determine sum of squared differences between model and data.

The surface minimum is the optimum parameter pair.

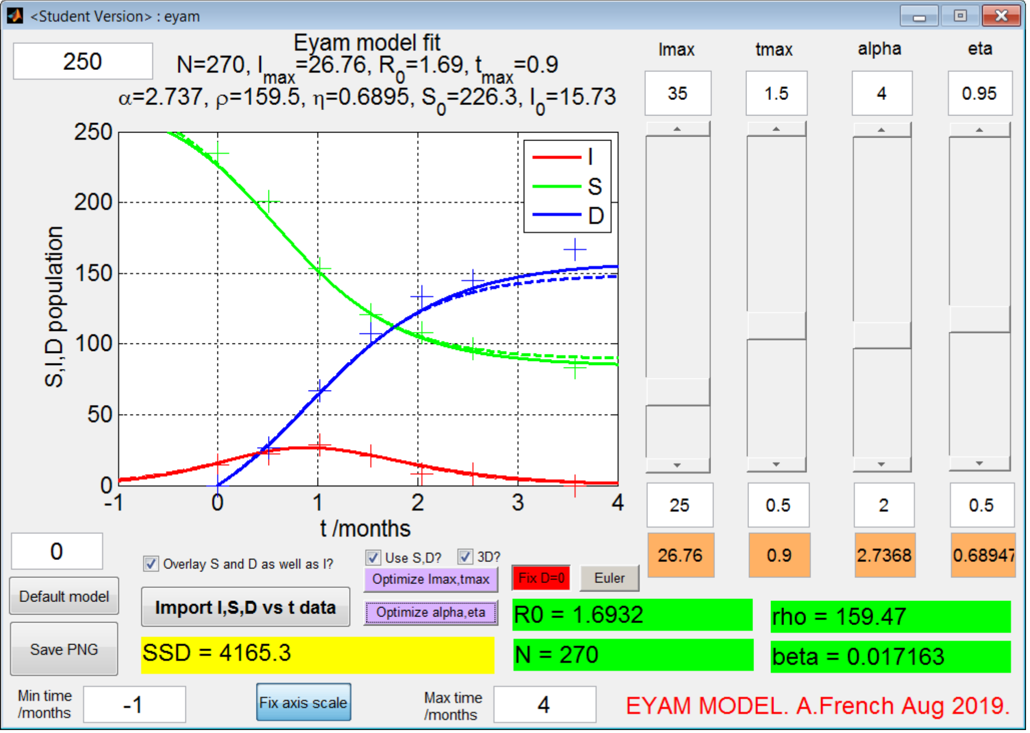


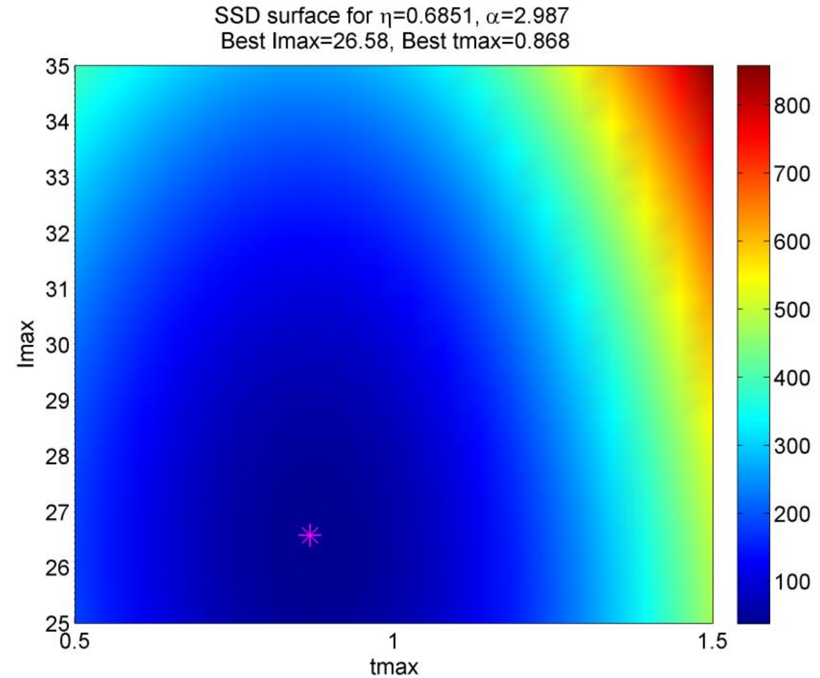
In this case the SSD is only computed using *I(t).*

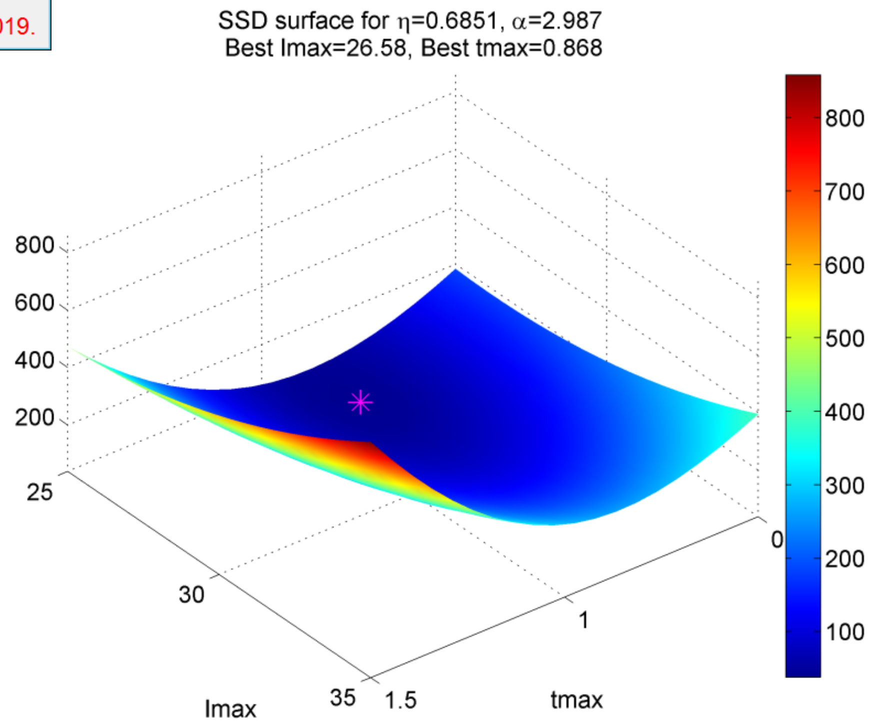
This is *not* a good fit to the Mompesson *S, D* data.



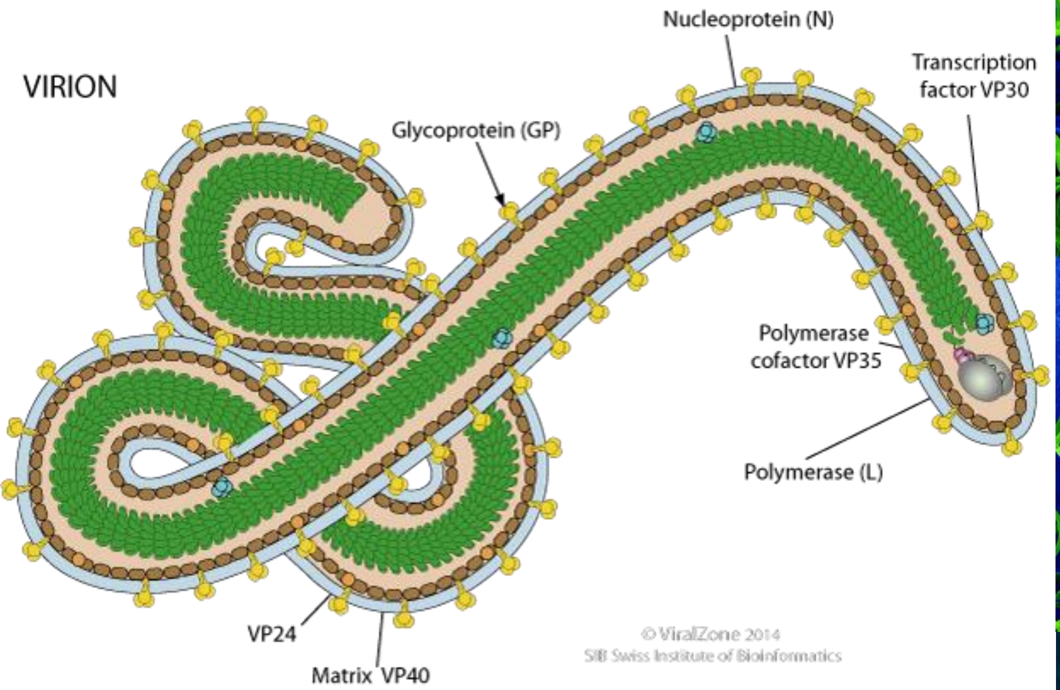
A much better fit if you use *I,S,D* data as well in the SSD computation and use t_{max} and I_{max} not eta and alpha as the SSD surface variables

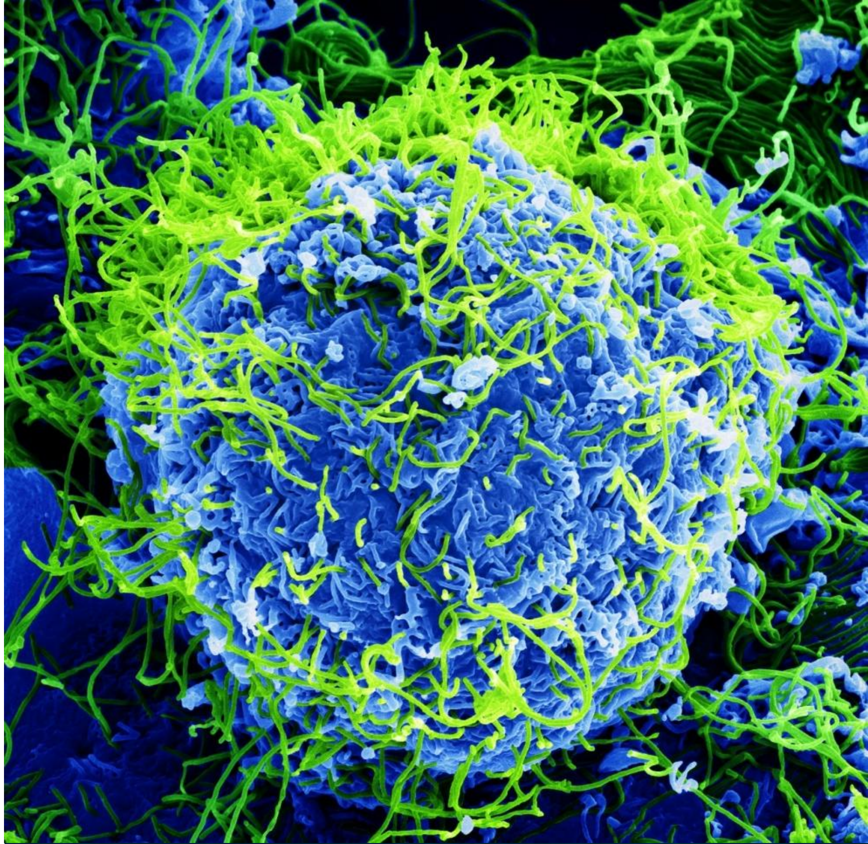






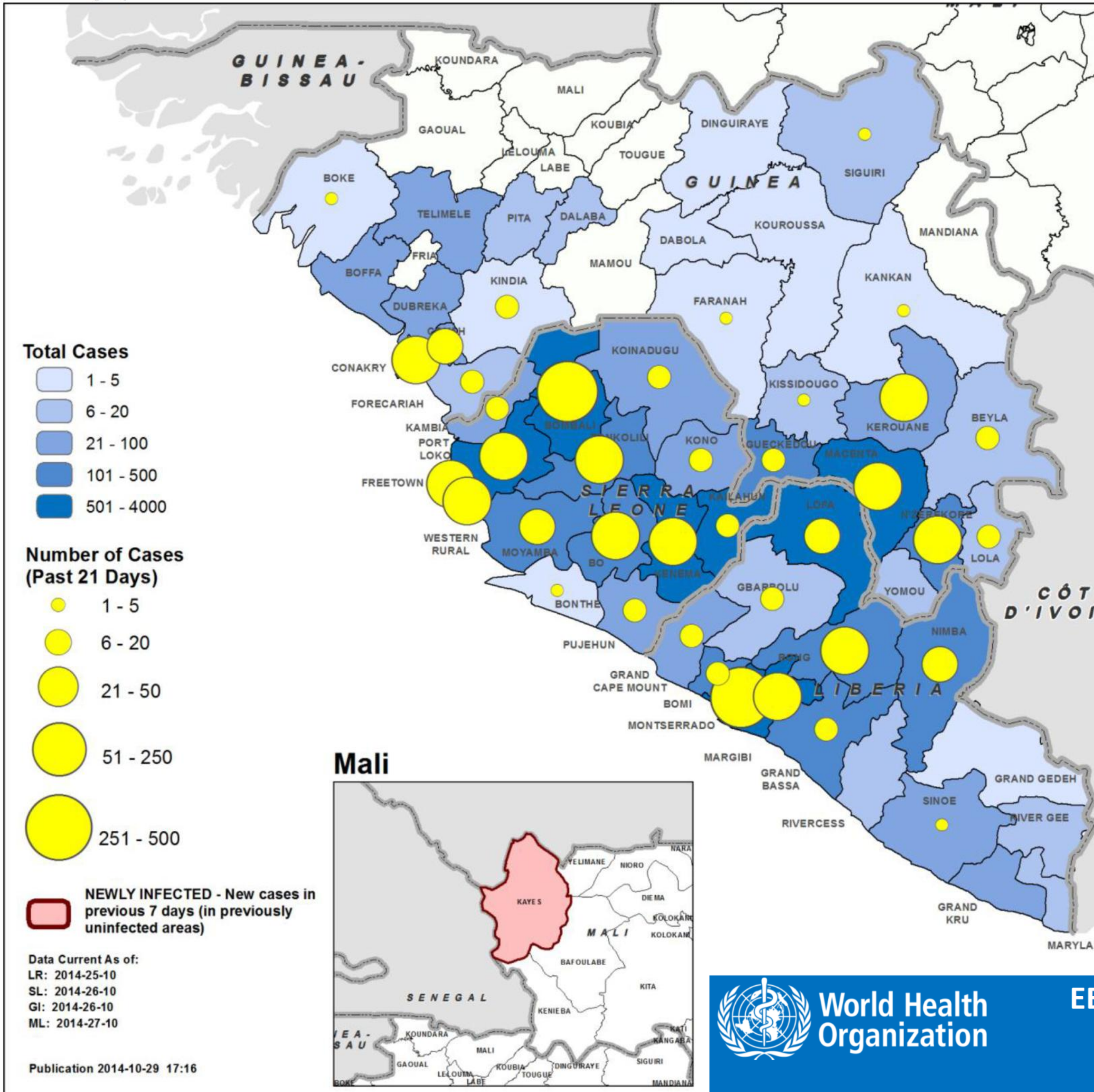
In this case a more clearly defined minimum of the SSD surface





Ebola virus

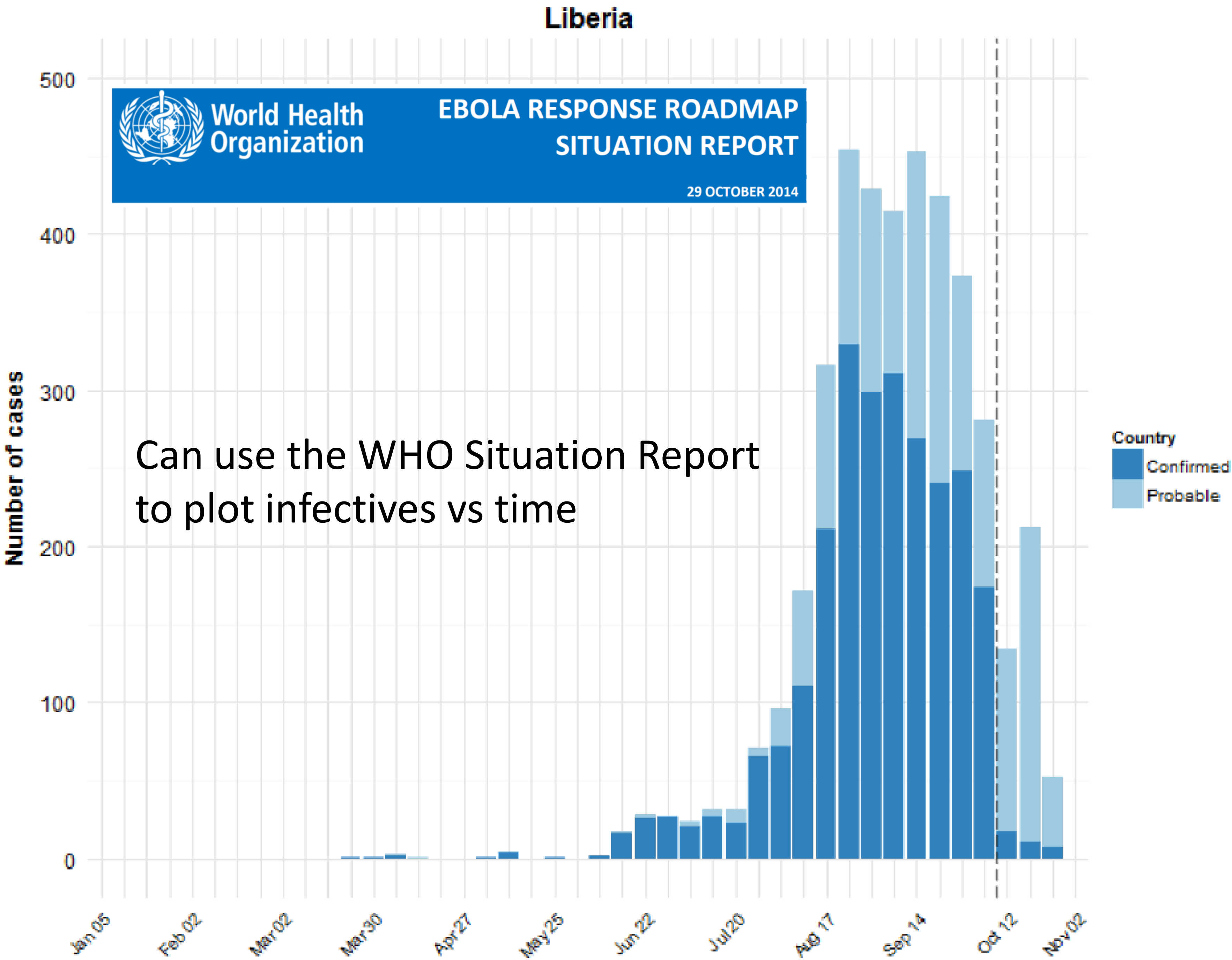
https://en.wikipedia.org/wiki/Zaire_ebolavirus



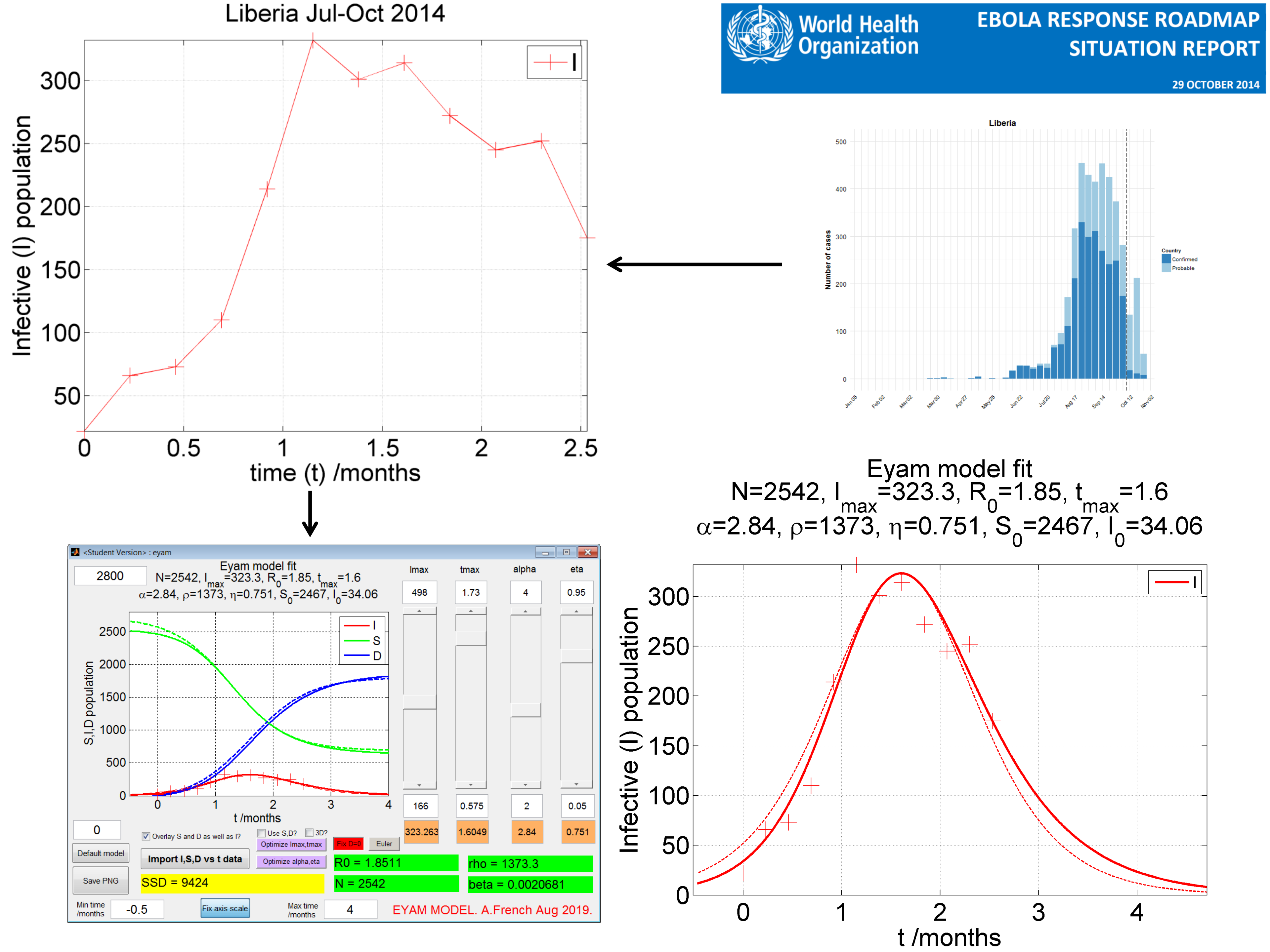
World Health Organization

EBOLA RESPONSE ROADMAP SITUATION REPORT

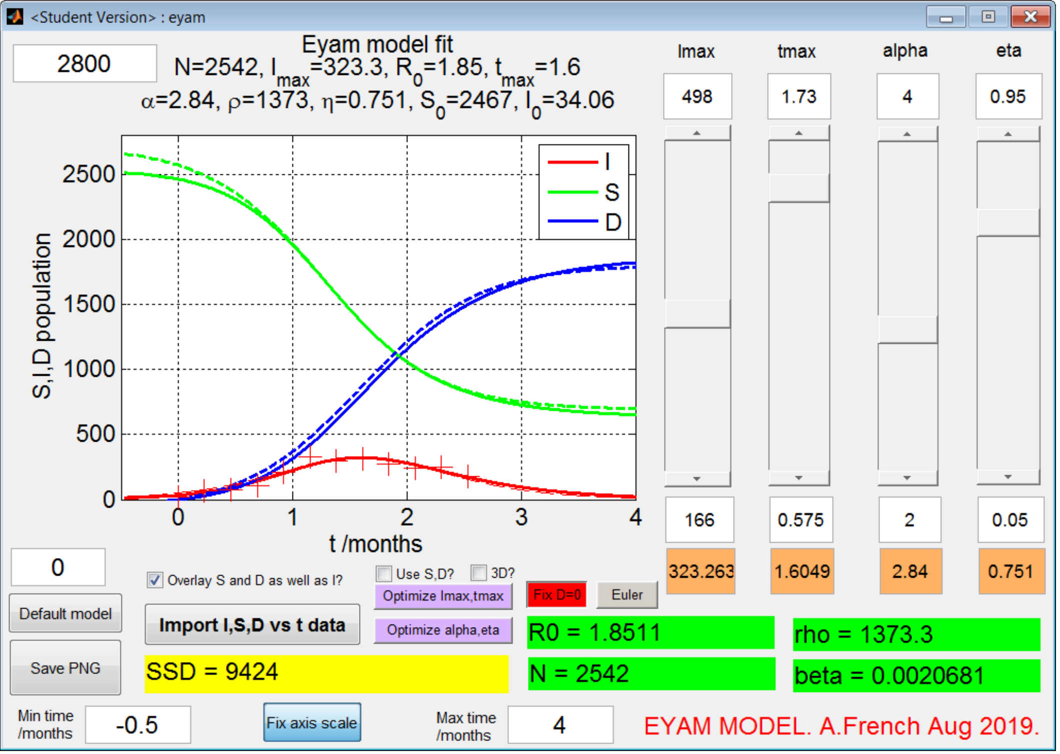
29 OCTOBER 2014

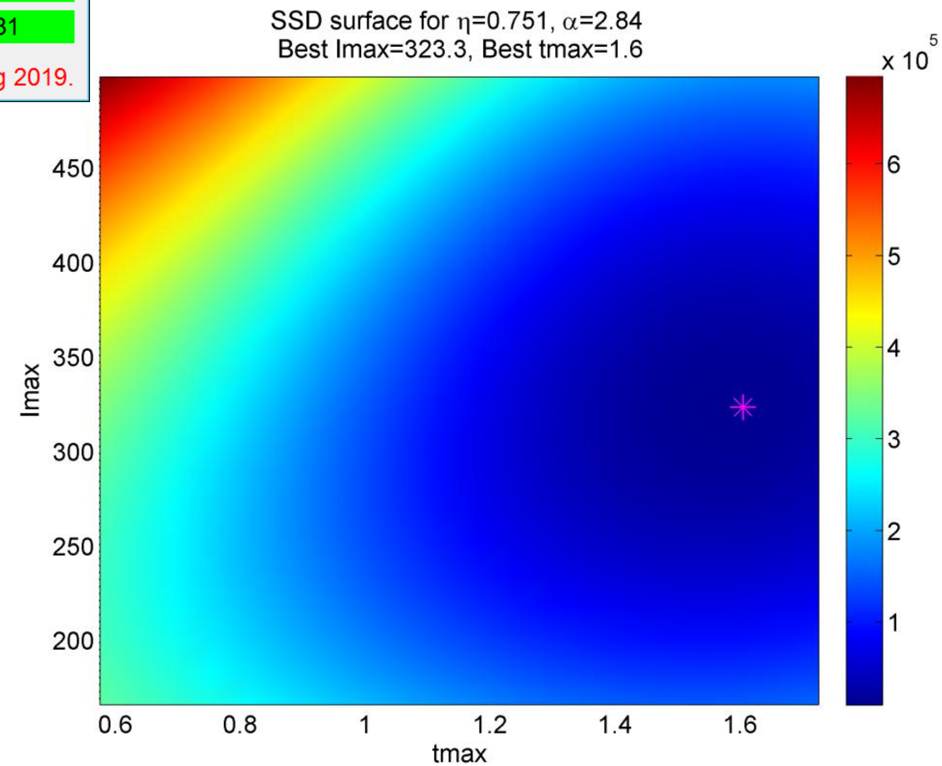


Can use the WHO Situation Report to plot infectives vs time

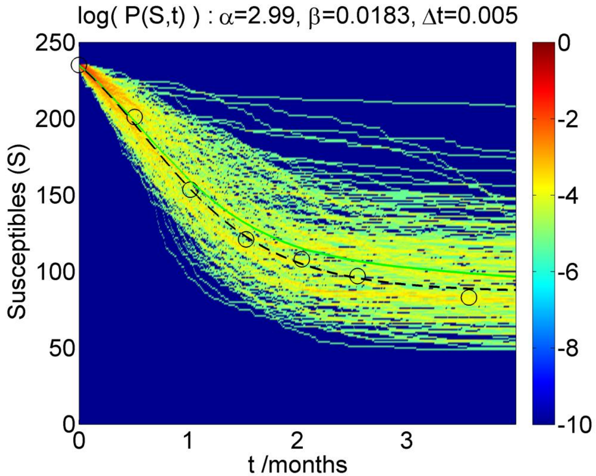


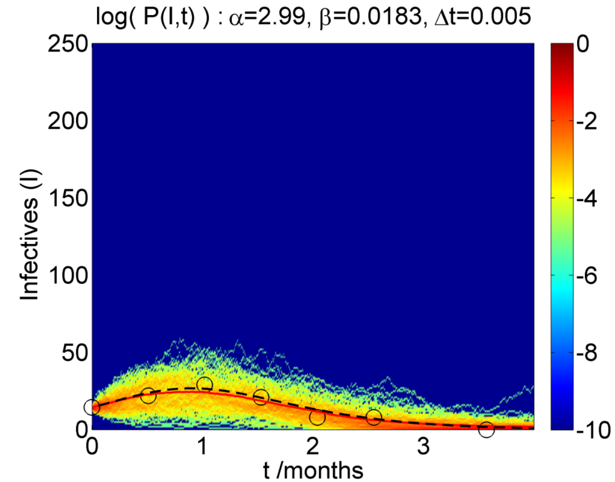
Liberia Jul-Oct 2014
Infective (I) population
time (t) /months
World Health Organization
EBOLA RESPONSE ROADMAP
SITUATION REPORT
29 OCTOBER 2014
Liberia
Number of cases
Country
Confirmed
Probable
Eyam model fit
N=2542, Imax=323.3, R0=1.85, tmax=1.6
α=2.84, ρ=1373, η=0.751, S0=2467, I0=34.06
t /months
<Student Version> : eyam
S,I,D population
Imax
tmax
alpha
eta
Overlay S and D as well as I?
Use S,D?
3D?
Optimize Imax,tmax
Optimize alpha,eta
Fix D=0
Euler
Default model
Import I,S,D vs t data
Save PNG
SSD = 9424
R0 = 1.8511
rho = 1373.3
N = 2542
beta = 0.0020681
Min time /months
Max time /months
Fix axis scale
EYAM MODEL. A.French Aug 2019.



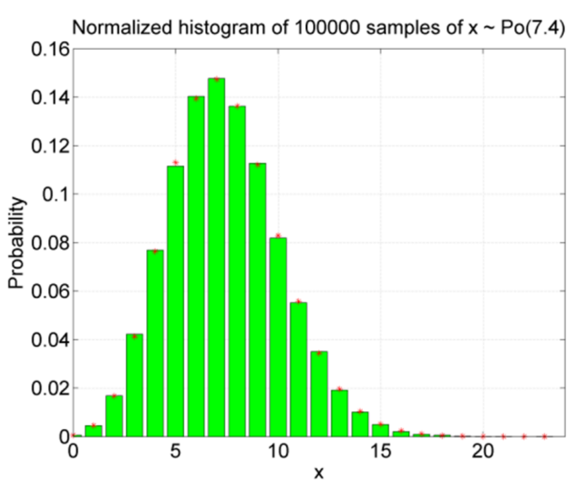


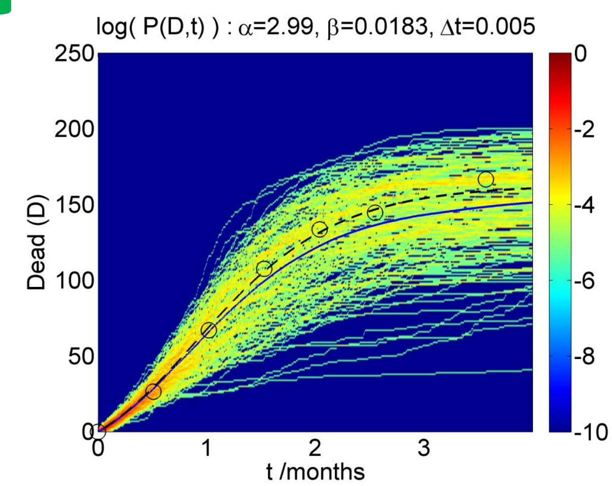
Optimizing parameters by finding the SSD surface minimum





A stochastic model





Stochastic Eyam model

Obviously the changes to S, I, D are **discrete**, *not* continuous values. Also, one expects the spread of infection to be a **random** process. Returning to Brauer's model, we can use the **expected** values of S,I and D changes within time interval Δt to be the mean (and variance) of a **Poisson distribution**. If we can sample this distribution, then *between each time step* we should have a *representative discrete change* of S,I,D that incorporates both the model and the idea of randomness.

$$\boxed{\begin{aligned}
&\Delta S = -x, \quad \Delta I_1 = x \\
&x \sim \mathrm{Po}\left(\beta S I \Delta t\right) \\
&\Delta D = y \\
&y \sim \mathrm{Po}\left(\alpha I \Delta t\right) \\
&\Delta I_2 = -y \qquad \therefore \Delta I = \Delta I_1 + \Delta I_2
\end{aligned}}$$

Poisson distribution

The random variable x is the number occurrences (e.g. goals, telephone calls ....) in a set interval of time, given a mean rate of occurrence λ.

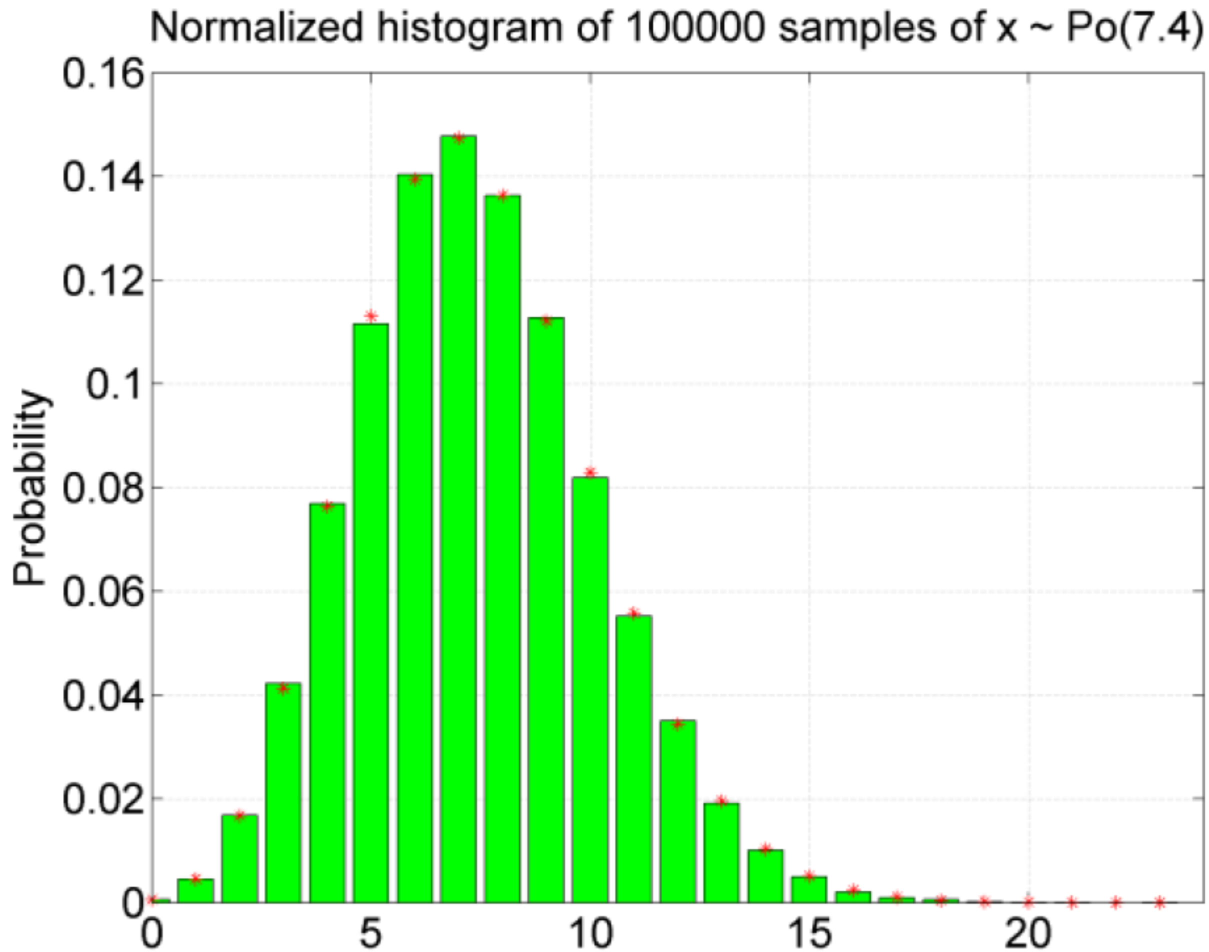
$$x \sim \mathrm{Po}(\lambda)$$

$$p(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

$$M_x(t) = e^{\lambda(e^t - 1)}$$

$$\mu = \lambda$$

$$\sigma^2 = \lambda$$

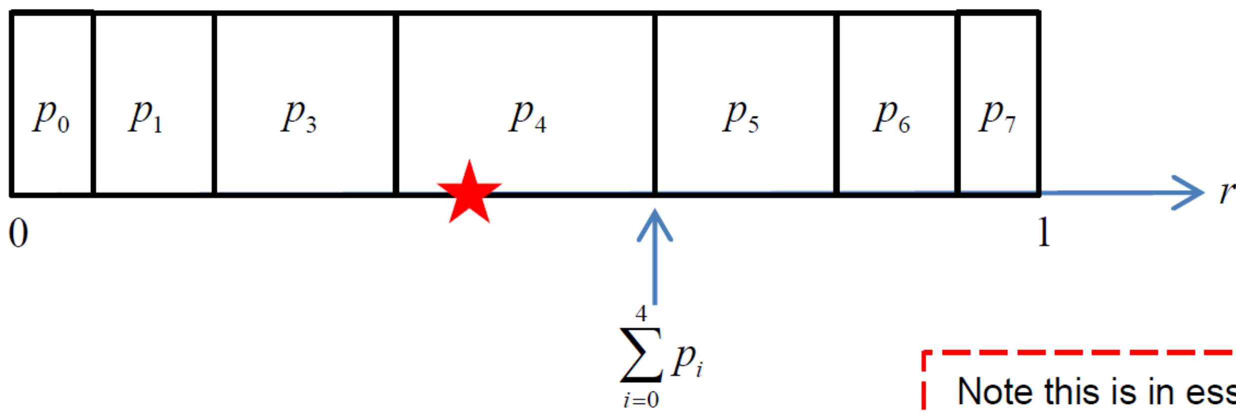



```
%Poisson distribution probabilities ( x is an array of integers >=0 )
function P = poisson(x,lambda)
P = (lambda.^x).*exp(-lambda)./factorial(x);
```

Generating random integers from discrete probability distributions

The fact that the sum of the probabilities in a discrete distribution must sum to unity can be used to generate random integers, assuming it is possible to generate a random number within the range [0,1].

Use the probabilities to form the edges of a series of 'boxes' which span the interval [0,1]. For every random fraction $\sim \mathrm{U}(0,1)$, determine the box number which encloses the fraction. This box number is the random variable.



Example: consider a discrete distribution of eight possible probabilities, for the random integers 0...7 as show above. The widths of the boxes correspond to the probabilities.

A random number $r \sim \mathrm{U}(0,1)$ is chosen. ★
This happens to be in the range:

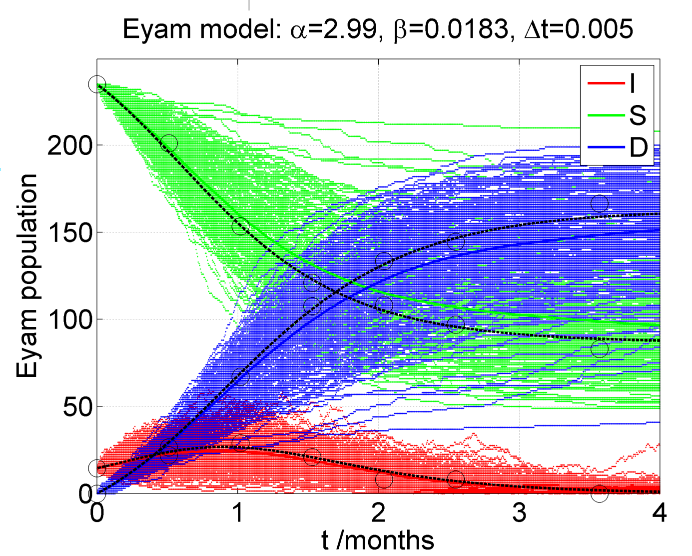
$$\sum_{i=0}^{3} p_i \leq r < \sum_{i=0}^{4} p_i$$

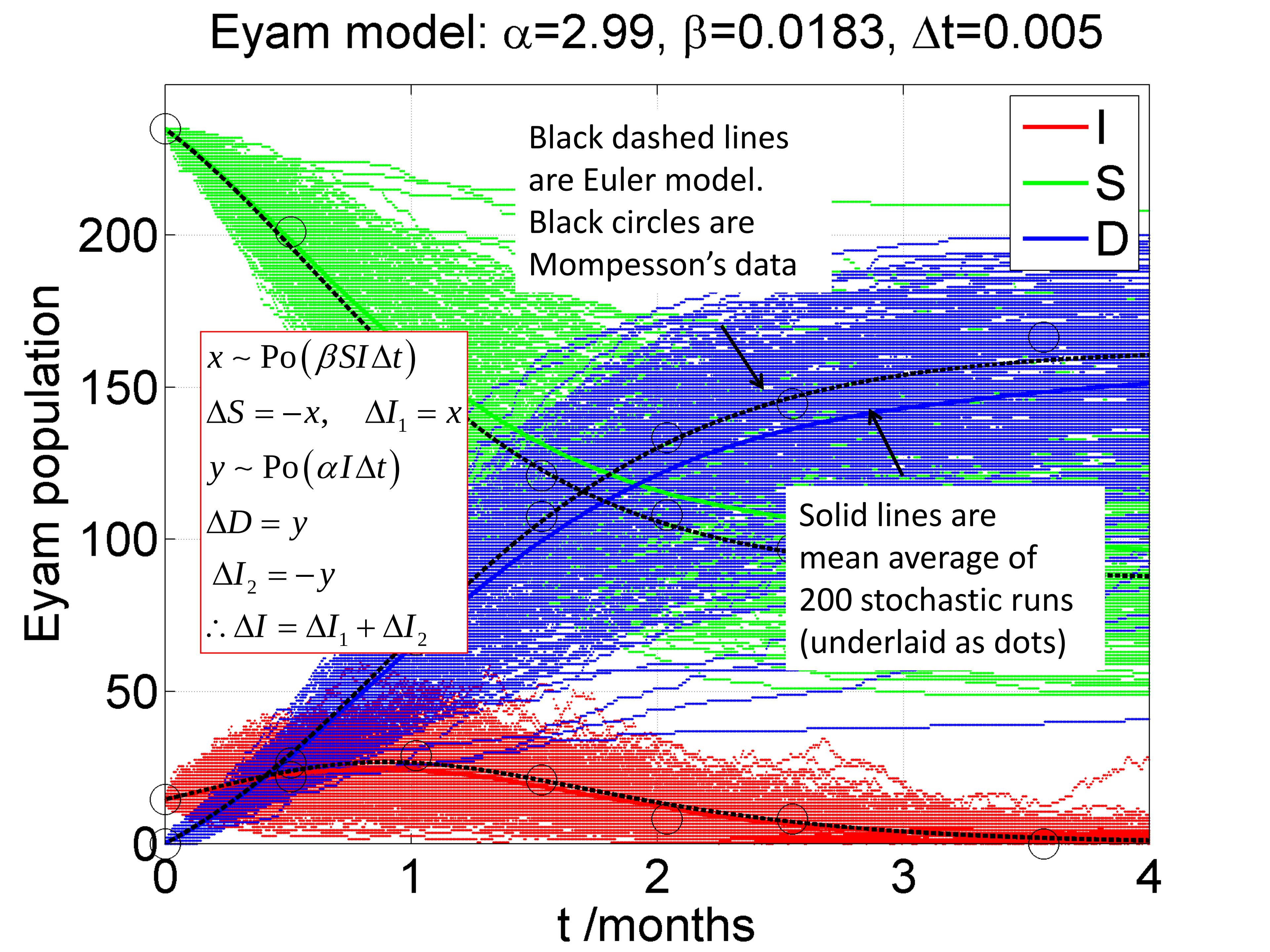
Cumulative distribution function

so in this case the random number $x = 4$ is selected.

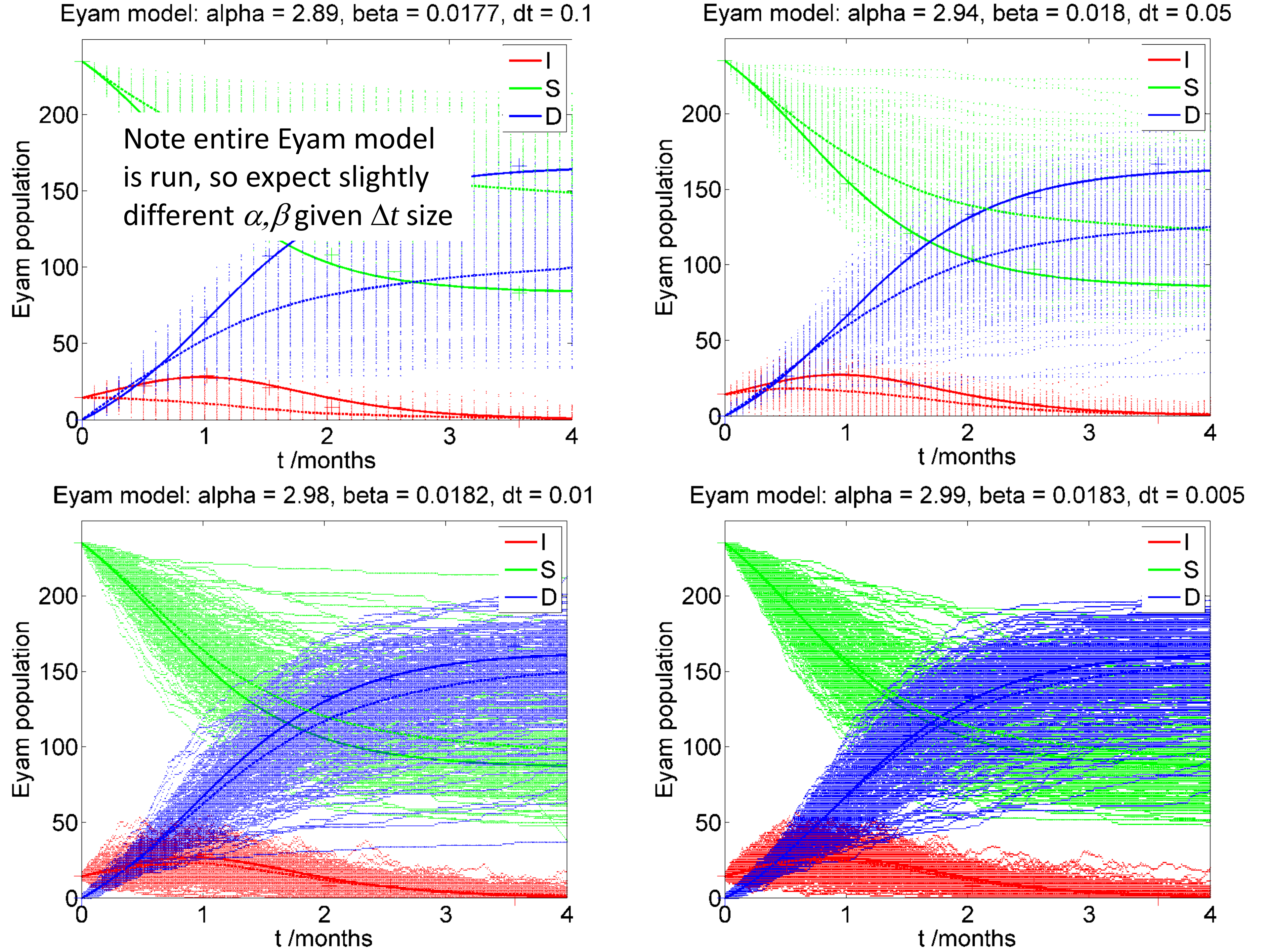
Note this is in essence the same process as random number generation from continuous distributions. i.e. the output of the **inverse cumulative distribution function** with input being a random number from $\mathrm{U}(0,1)$

```
%Stochastic model of Eyam SID model
function [t,I,S,D] = eyam_stochastic_model( dt, I0, S0, alpha, beta, tmax )

%Initialize output vectors for t,I,S,D
t = 0 : dt : tmax;
N = length(t);
S = S0*ones(1,N);
I = I0*ones(1,N);
D = zeros(1,N);

%Loop through vectors to compute t, I, S, D.
%using a Poisson probabilistic rule for S,I,D changes during timestep dt
for n=2:N
    t(n) = t(n-1) + dt;

    %Poisson probabalistic rule for transition from S to I
    lambda = dt*beta*S(n-1)*I(n-1);
    dS = -poisson_samples( lambda,1 );
    dI = -dS;

    %Update I and S
    I(n) = I(n-1) + dI;
    S(n) = S(n-1) + dS;

    %Probabilistic rule for transition from I to D populations
    lambda = dt*alpha*I(n);
    dD = poisson_samples( lambda, 1 );
    dI = -dD;

    %Update I and D ( note I(n) is to be modified )
    I(n) = I(n) + dI;
    D(n) = D(n-1) + dD;
end
```

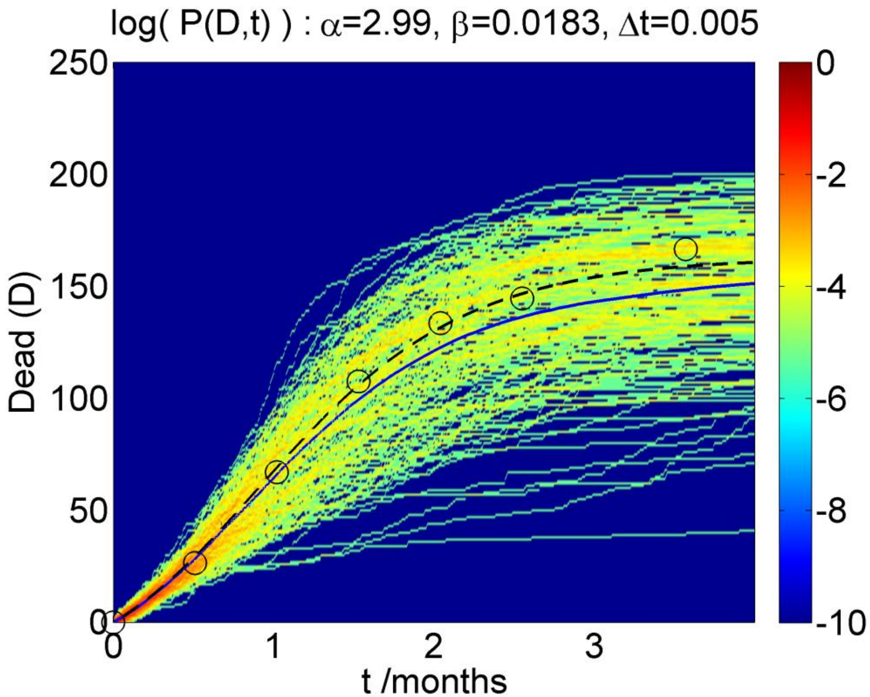


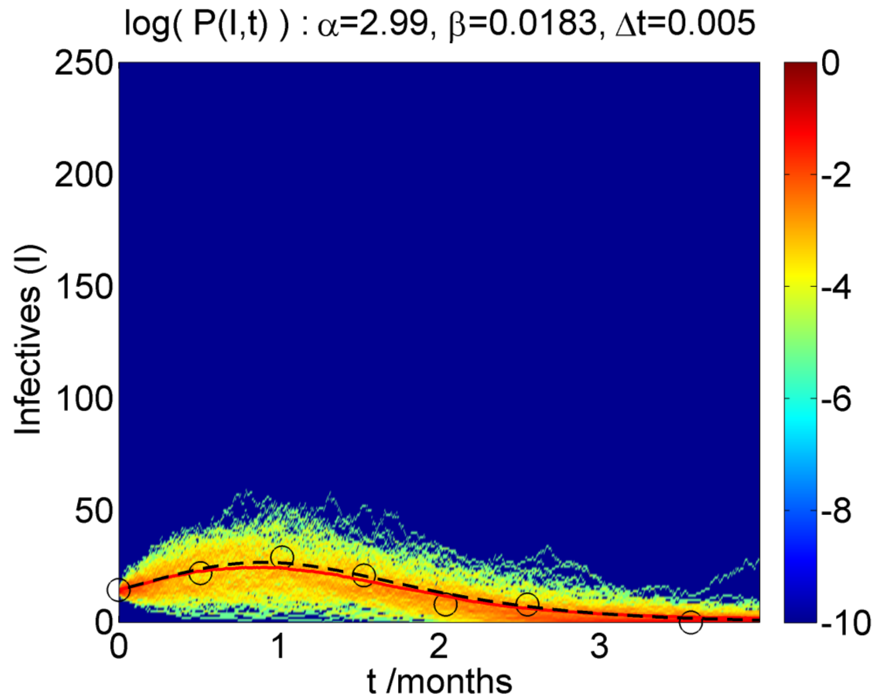


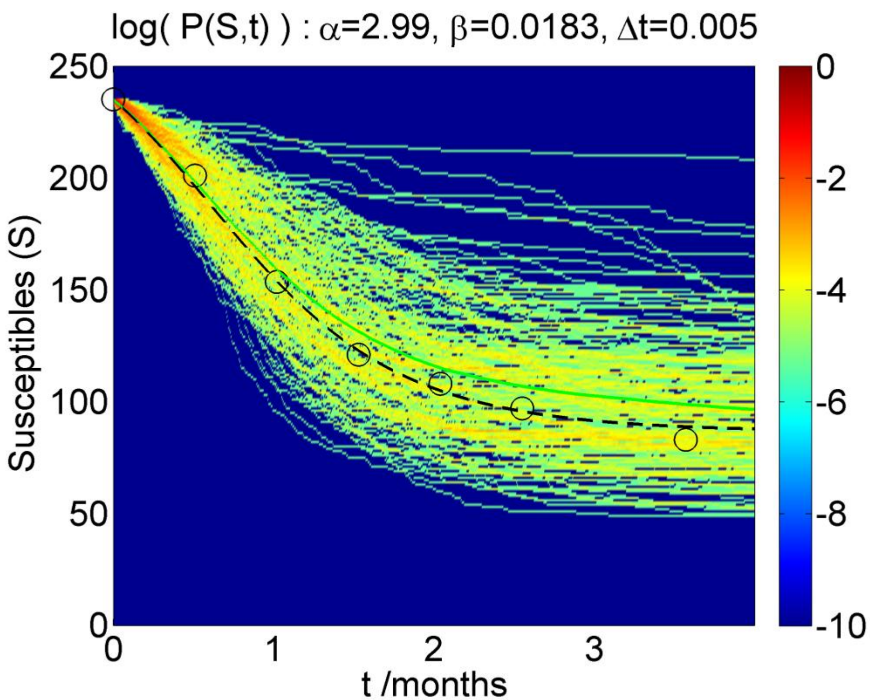
Eyam model: α=2.99, β=0.0183, Δt=0.005
Black dashed lines are Euler model.
Black circles are Mompesson's data
Solid lines are mean average of 200 stochastic runs (underlaid as dots)
$x \sim \mathrm{Po}(\beta S I \Delta t)$
$\Delta S = -x, \quad \Delta I_1 = x$
$y \sim \mathrm{Po}(\alpha I \Delta t)$
$\Delta D = y$
$\Delta I_2 = -y$
$\therefore \Delta I = \Delta I_1 + \Delta I_2$
I
S
D
Eyam population
t /months
0
50
100
150
200
0
1
2
3
4



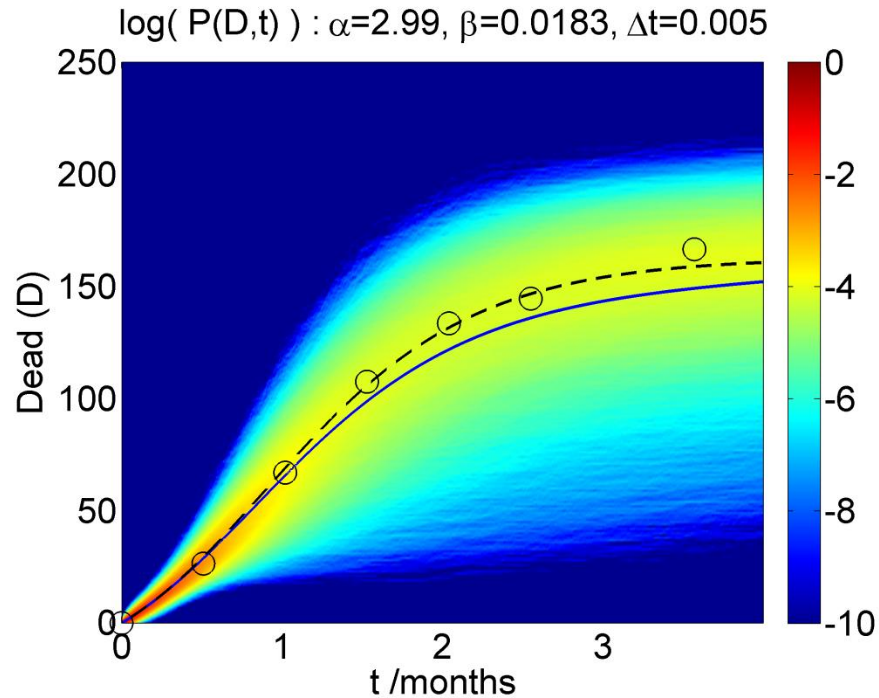
Eyam model: alpha = 2.89, beta = 0.0177, dt = 0.1
Note entire Eyam model is run, so expect slightly different α,β given Δt size
Eyam model: alpha = 2.94, beta = 0.018, dt = 0.05
Eyam model: alpha = 2.98, beta = 0.0182, dt = 0.01
Eyam model: alpha = 2.99, beta = 0.0183, dt = 0.005
I
S
D
Eyam population
t /months
0
1
2
3
4
0
50
100
150
200

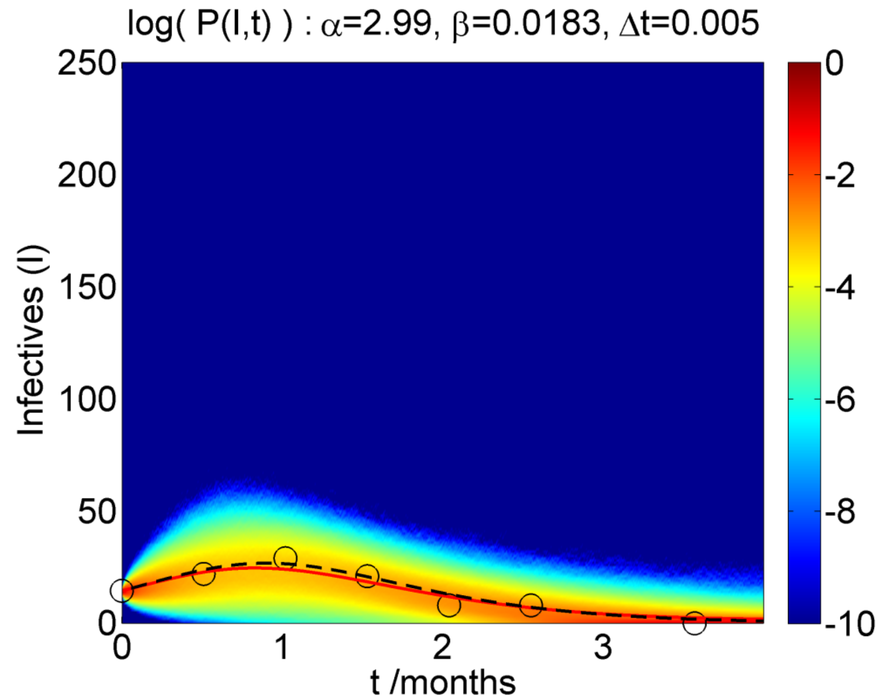


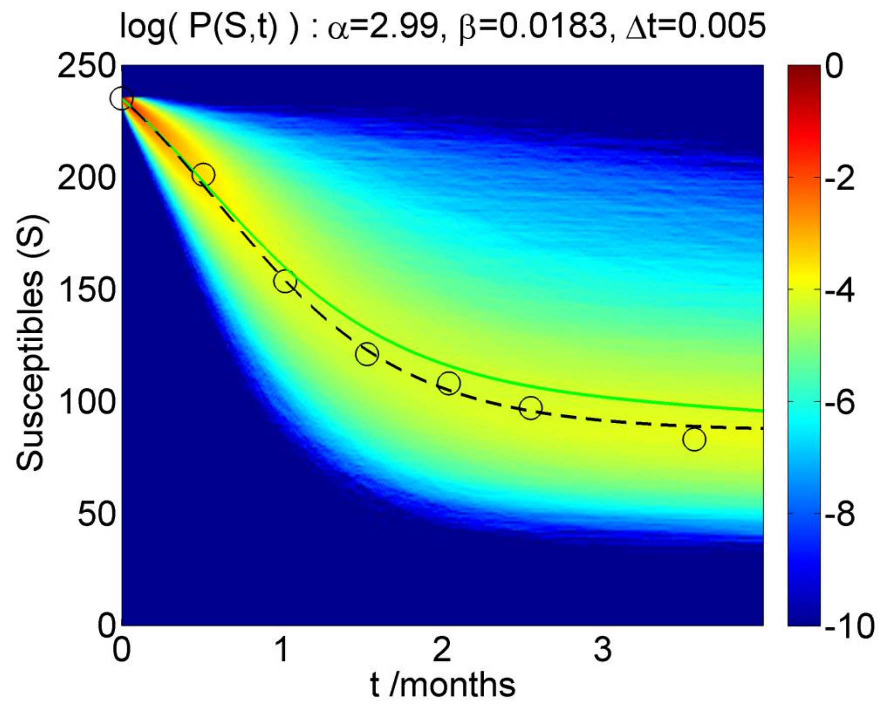




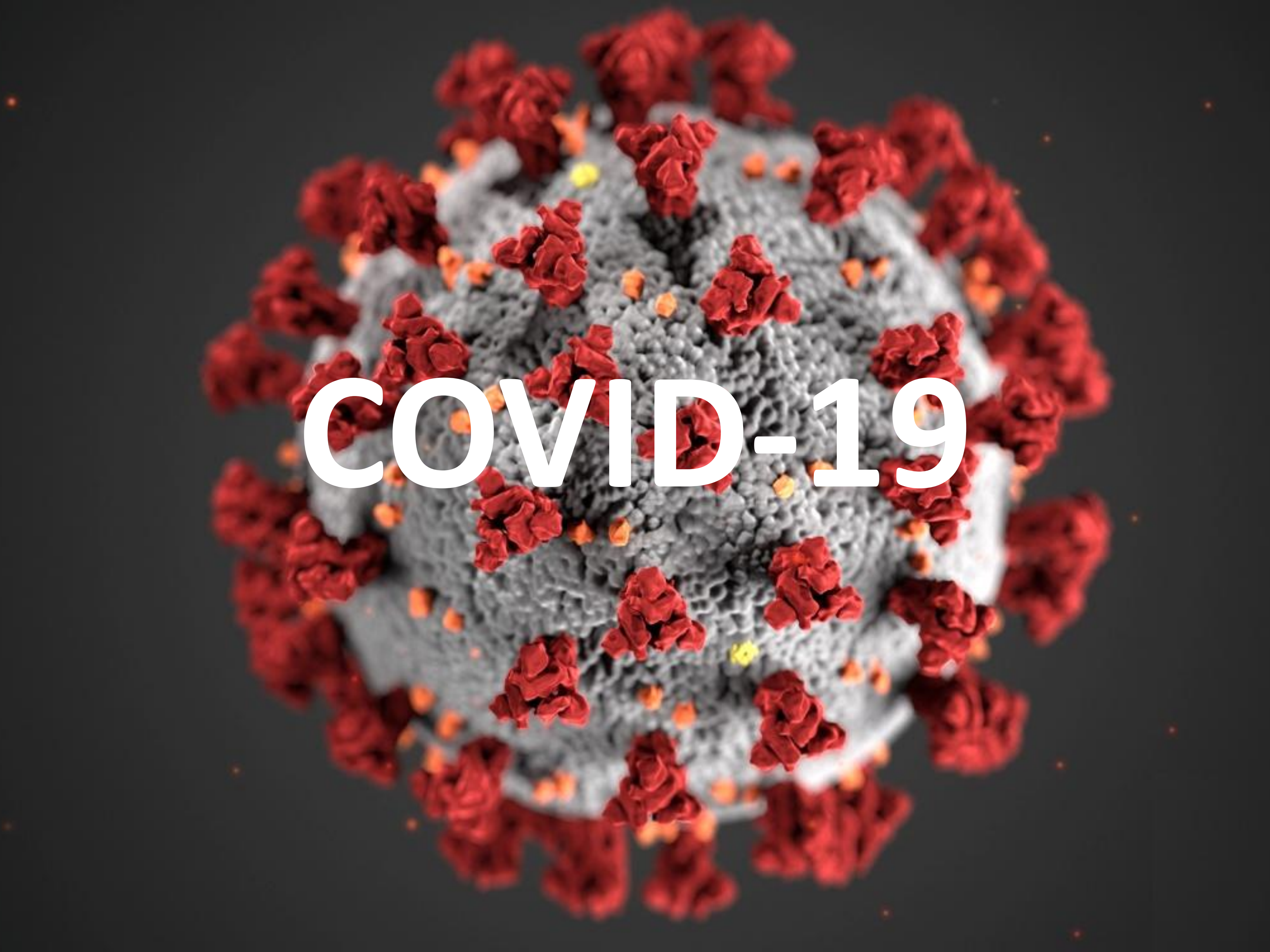
Probability map, computed from 200 iterations. Black circles are Mompesson data and black dashed lines correspond to the Euler model.







Probability map, computed from 50,000 iterations. Black circles are Mompesson data and black dashed lines correspond to the Euler model.


COVID-19

Eyam model fit for
Wuhan (China) COVID-19 outbreak
Jan 22 – Mar 16 2020
From [Oxford World in data]

EYAM EQUATIONS
Susceptible, **I**nfective
Removed (either **R**ecovered or **D**ead)

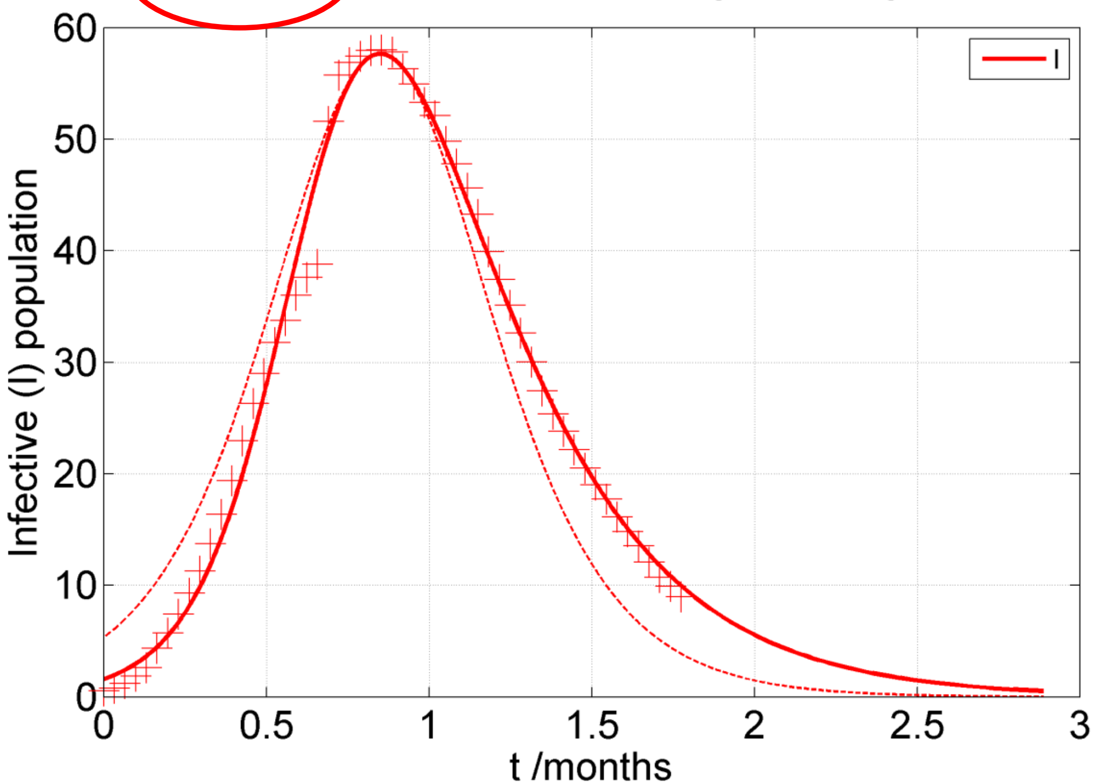
Assumes S to I to $R + D$ flow (one way) and a **fixed** total at-risk population N

$$\frac{dS}{dt} = -\beta S I$$

$$\frac{dI}{dt} = \beta\left(S - \tfrac{\alpha}{\beta}\right) I$$

$$\frac{d(R+D)}{dt} = \alpha I$$

Eyam model fit
N=192, I_{max}=57.65, R_0=3, t_{max}=0.849
α=3.263, ρ=63.92, η=0.9405, S_0=189.4, I_0=1.588



$$z_+ = -\ln(1-\eta) - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$z_- = -\ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$x_{\max} = -\frac{\ln(1-\eta)}{\eta} - 1 - \ln\left(-\frac{\ln(1-\eta)}{\eta}\right)$$

$$\rho = \frac{I_{\max}}{x_{\max}}$$

$$\tau(z) = \int_0^z \frac{dz'}{x_{\max} + 1 - e^{-z'} - z'}$$

$$x = x_{\max} + 1 - e^{-z} - z$$

$$y = e^{-z}$$

$$t = \frac{\tau}{\alpha} + t_{\max},\ \ I = \rho x\ \ S = \rho y,\ \ R + D = \rho(z - z_-)$$

$$N = I_{\max} + \rho - \rho z_-$$

$$R_0 = \frac{N}{\rho}$$

Model predicts a **Basic Reproduction number** R_0 of **3.00**

SEMI ANALYTIC EYAM MODEL. JPC/AF 2019

$$\frac{dS}{dt} = -\beta SI$$

$$\frac{dI}{dt} = \beta\left(S - \tfrac{\alpha}{\beta}\right) I$$

$$\frac{d(R+D)}{dt} = \alpha I$$

"**Eyam equations**" an S,I,R,D model of *population flows* to model an epidemic.

Time constant (from Wuhan data):

Assume this is a function of basic human biology and therefore an approximate *constant*, rather than something that might vary due to the proximity and social mixing of human populations.

i.e. *not* like β

$$T = 1/\alpha$$

$$T = \tfrac{1}{3.63} \text{ months}$$

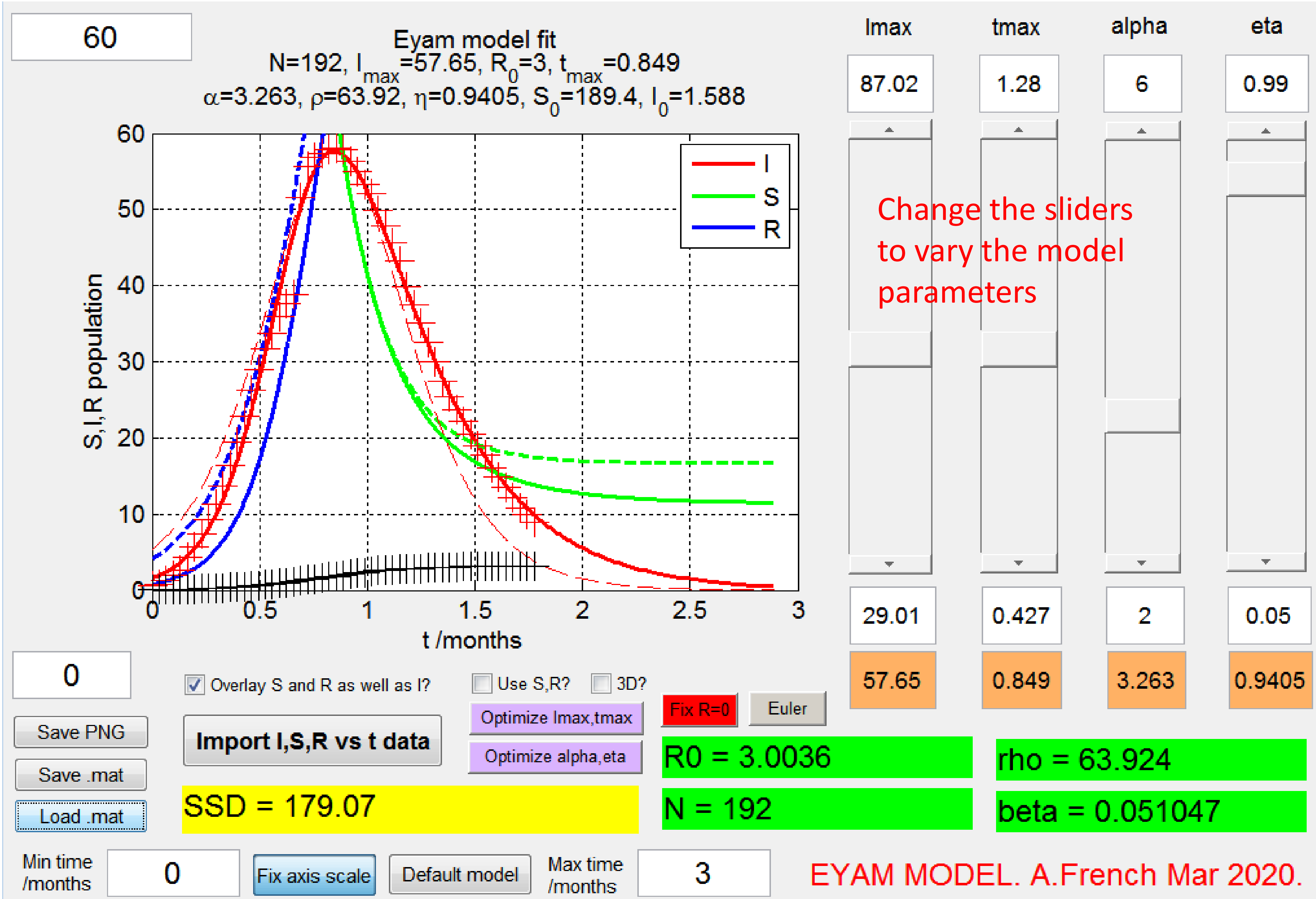
$$T = \frac{365/12}{3.263} \text{ days}$$

$$T = 9.32 \, \text{days}$$

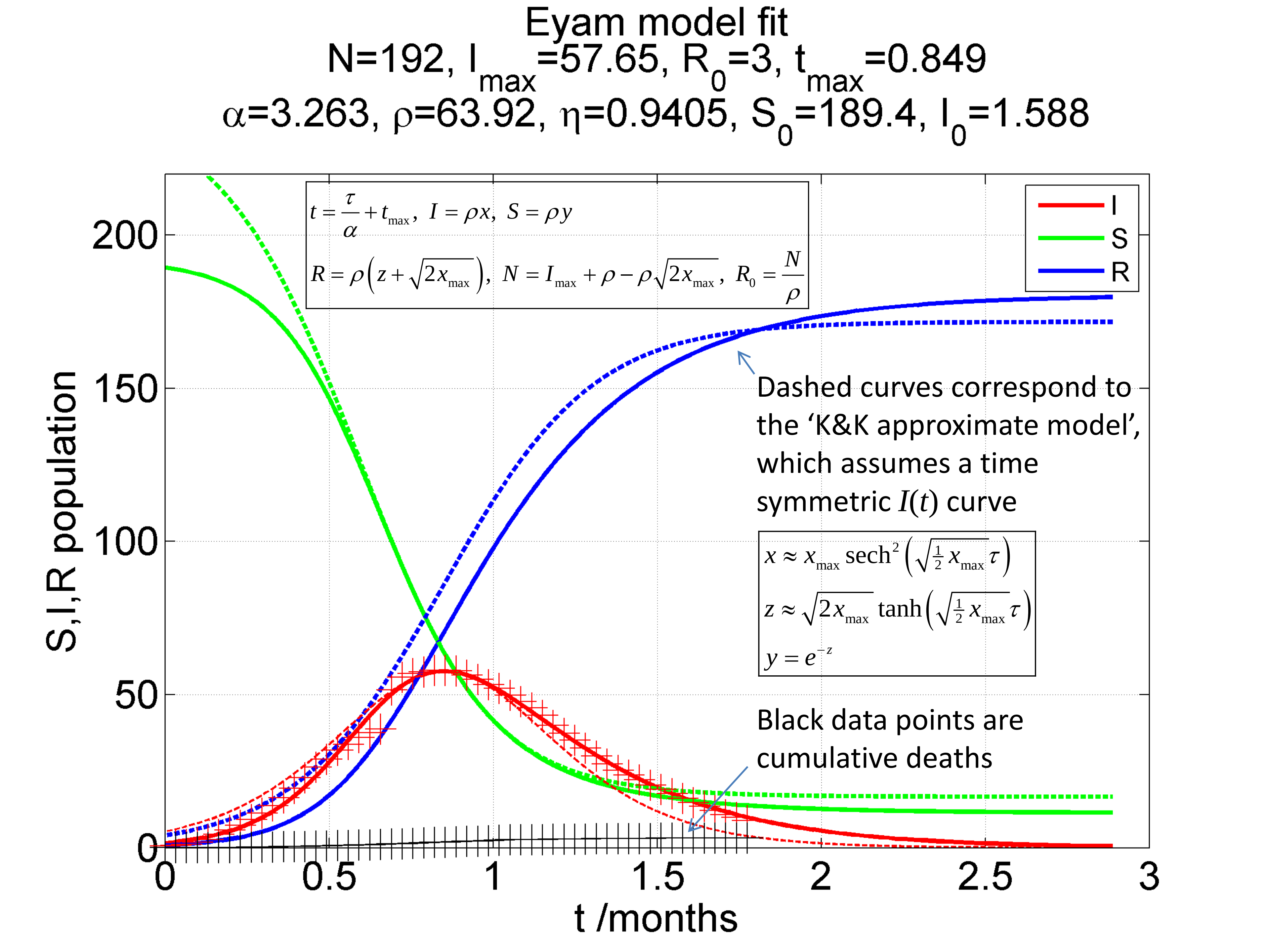
i.e. a measure of the characteristic time from infection till recovery (or death).

Assume **R**ecovered population can no longer spread COVID-19, and also have immunity so *cannot* become Susceptibles again.

This might not be true!



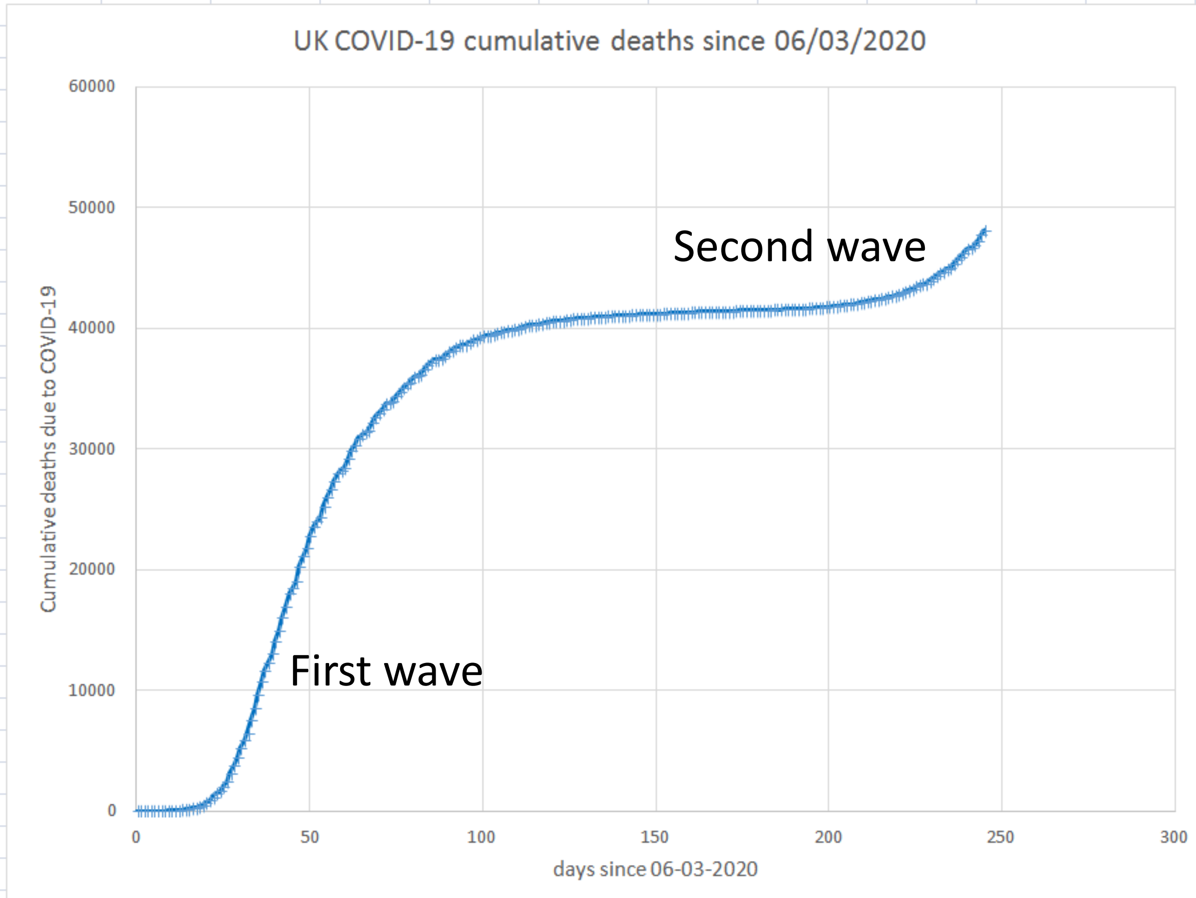
The EYAM MODEL predicts the S and R curves from the Infectives vs time data



Eyam model fit
N=192, I_{max}=57.65, R_0=3, t_{max}=0.849
α=3.263, ρ=63.92, η=0.9405, S_0=189.4, I_0=1.588
$t = \frac{\tau}{\alpha} + t_{max}$, $I = \rho x$, $S = \rho y$
$R = \rho\left(z + \sqrt{2x_{max}}\right)$, $N = I_{max} + \rho - \rho\sqrt{2x_{max}}$, $R_0 = \frac{N}{\rho}$
I
S
R
Dashed curves correspond to the 'K&K approximate model', which assumes a time symmetric $I(t)$ curve
$x \approx x_{max} \operatorname{sech}^2\left(\sqrt{\tfrac{1}{2}x_{max}}\,\tau\right)$
$z \approx \sqrt{2x_{max}} \tanh\left(\sqrt{\tfrac{1}{2}x_{max}}\,\tau\right)$
$y = e^{-z}$
Black data points are cumulative deaths
S,I,R population
t /months
0
50
100
150
200
0.5
1
1.5
2
2.5
3

UK COVID-19 curve of cumulative deaths (from Oxford *World in data*)

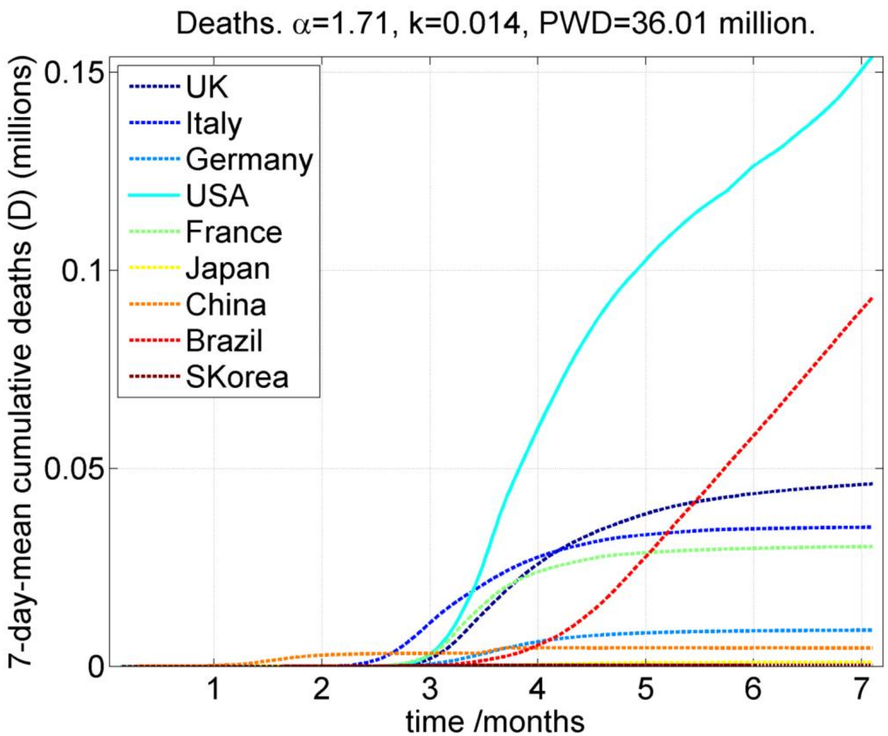
	A	B	C	D
1	date	days	total_deaths	
2	2020-03-06	0	0	
3	2020-03-07	1	1	
4	2020-03-08	2	2	
5	2020-03-09	3	2	
6	2020-03-10	4	3	
7	2020-03-11	5	7	
8	2020-03-12	6	7	
9	2020-03-13	7	9	
10	2020-03-14	8	10	
11	2020-03-15	9	29	
12	2020-03-16	10	43	
13	2020-03-17	11	65	
14	2020-03-18	12	82	
15	2020-03-19	13	116	
16	2020-03-20	14	162	
17	2020-03-21	15	194	
18	2020-03-22	16	252	
19	2020-03-23	17	288	
20	2020-03-24	18	364	LOCKDOWN BEGINS
21	2020-03-25	19	512	
22	2020-03-26	20	703	
23	2020-03-27	21	884	
24	2020-03-28	22	1172	
25	2020-03-29	23	1464	
26	2020-03-30	24	1676	
27	2020-03-31	25	2050	
28	2020-04-01	26	2453	
29	2020-04-02	27	3125	
30	2020-04-03	28	3782	
31	2020-04-04	29	4518	
32	2020-04-05	30	5274	

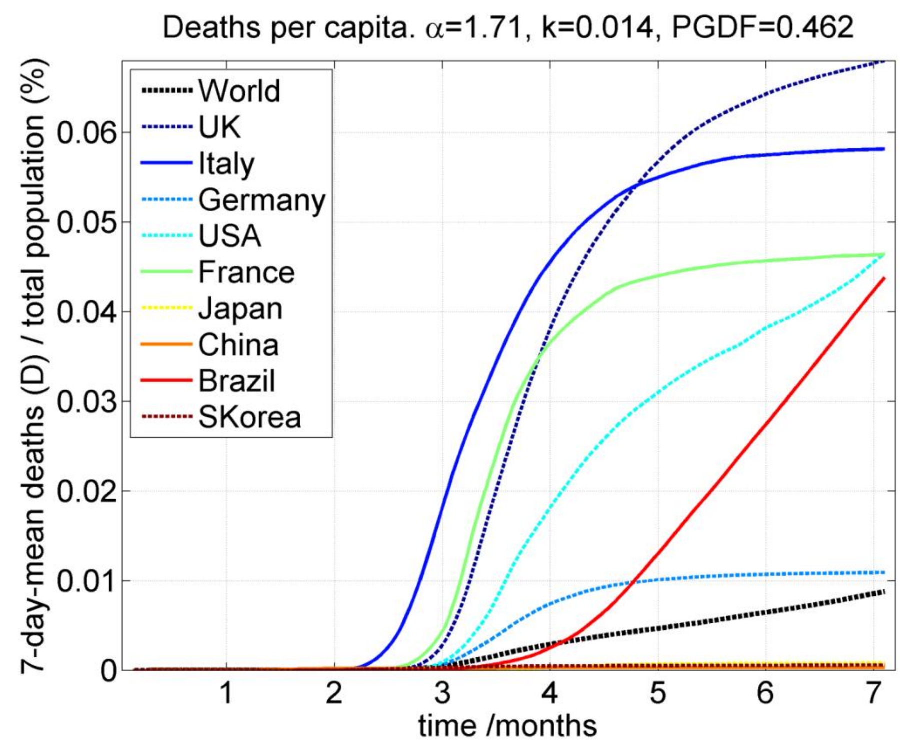


https://github.com/owid/covid-19-data/tree/master/public/data

Hasell, J., Mathieu, E., Beltekian, D. *et al*. A cross-country database of COVID-19 testing. *Sci Data* **7**, 345 (2020).

Compare 'first wave' COVID-19 deaths





Scale by population

To make sense of the COVID-19 epidemic, and for the epidemiology to match the narrative of "infection peaks" and "flattening the curve" (e.g. via a *lockdown* and increased social distancing and other interventions), **we ought to present the time variation of Infective population vs time.** *Other graphs are potentially confusing*. The graph of positive tests vs time (per day) is particularly problematic – since a rise might simply result from greater testing capacity rather than a rise in infective population.*

However, since testing is *not comprehensive*, i.e. the entire population is not tested regularly, which was certainly true at the start of the "first wave", **we can only <u>estimate</u> I vs t.**

The **Eyam equations** give us a means of achieving this, but only if we know the **time constant** T and hence α, and also the **mortality fraction** k. I shall assume both are biological in nature and *therefore constant*. Note the constancy of k **is probably a poor assumption**, since this will certainly vary among the population. Death from COVID-19 for a young healthy person is very likely to be much less probable than for someone elderly and frail, with possible multiple pre-existing health conditions. However, taking a crude average, let us assume k = 0.01. This is an educated guess, but informed by anecdotal evidence from NHS colleagues. Note the t vs I curve will look the *same* though, (just scaled slightly differently) as long as k is deemed to be a constant with time.

*The only other graph I think is useful to present is **new hospital admissions per day**, or perhaps even better, *fraction of maximum intensive care capacity per day*. This would give a sobering sense of the true human impact of COVID at the sharp end of things.
Note to **compare different countries**, one should plot **Infective population divided by total population**, i.e. 'per capita.'

$$\frac{d(R+D)}{dt} = \alpha I$$

Third “Eyam equation”

$$D = k(R+D) \quad \therefore R+D = \frac{D}{k}$$

$$\boxed{\begin{aligned} &\therefore D(1-k) = kR \\ &\therefore R = \frac{1-k}{k} D \end{aligned}}$$

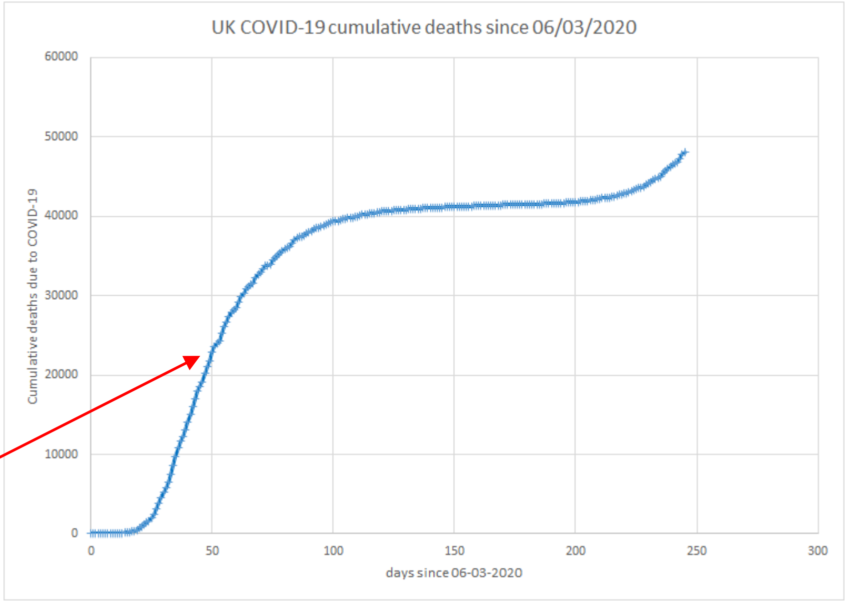
Recovered population, assuming a fixed mortality fraction.

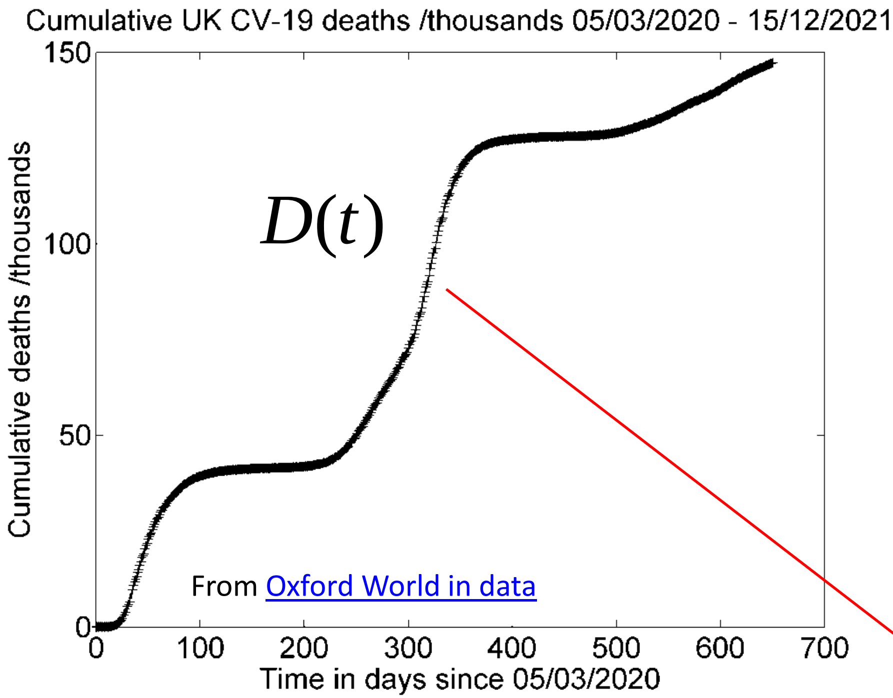
$$\frac{dD}{dt} = k\alpha I \quad \therefore \boxed{I = \frac{1}{k\alpha}\frac{dD}{dt}}$$

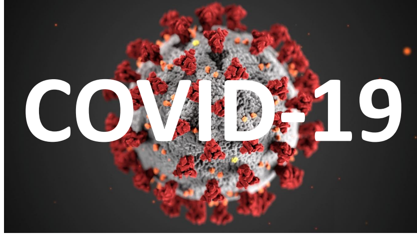
$$I_n = \frac{1}{k\alpha}\frac{dD}{dt} \approx \frac{1}{k\alpha}\frac{D_{n+1} - D_{n-1}}{t_{n+1} - t_n}$$

If we assume the cumulative deaths due to COVID-19 are accurate, then **numerically differentiating** this curve, and dividing by $k\alpha$, should yield an *estimate* for the Infective I population.

The cumulative deaths vs time is probably the most accurate statistic in the *World in Data* resource, since one assumes all UK deaths must have a death certificate and therefore a recorded cause of death (which if due to COVID-19, is represented in the data set).







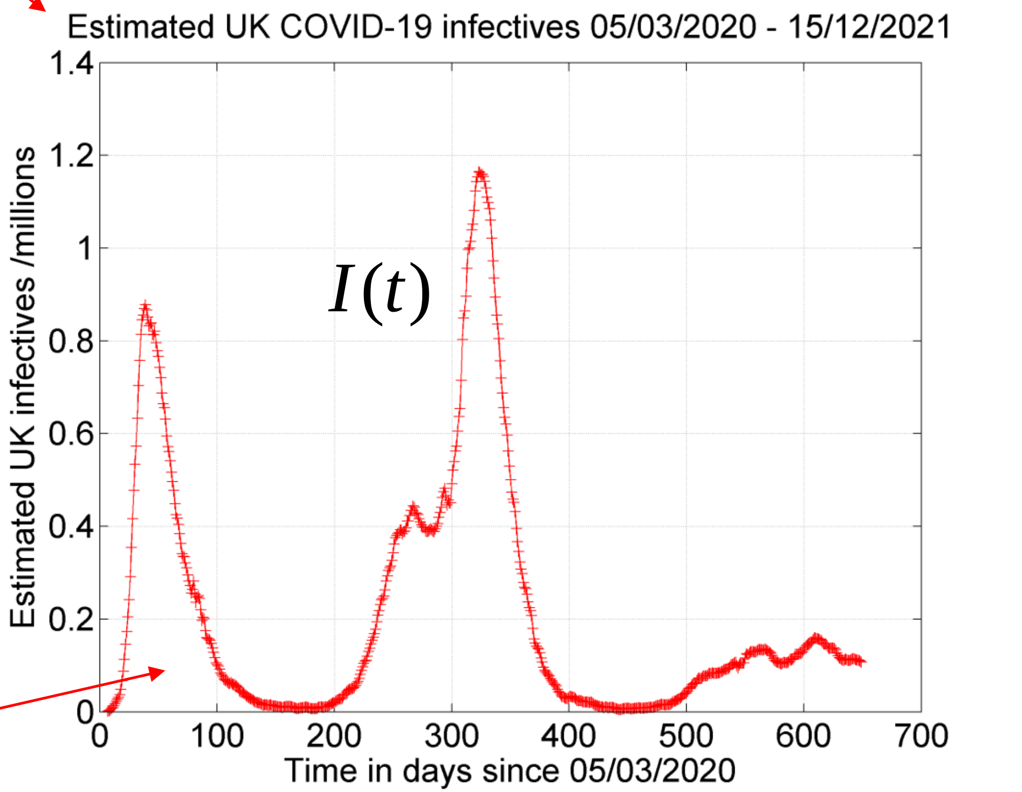
One can *estimate* the number of CV-19 **infectives** from the cumulative deaths:

$$I_n = \frac{1}{k\alpha}\frac{dD}{dt} \approx \frac{1}{k\alpha}\frac{D_{n+1} - D_{n-1}}{t_{n+1} - t_{n-1}}$$

Find the gradient and scale by:

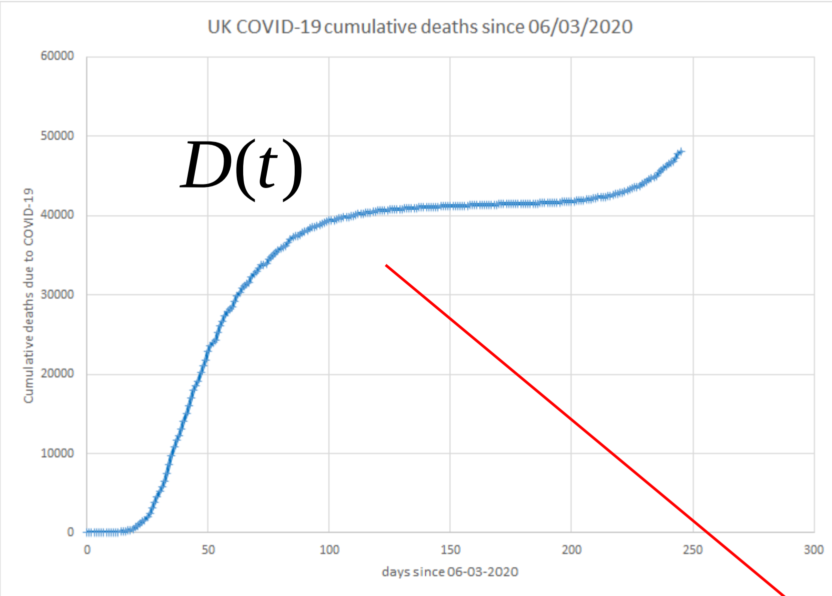
$$k\alpha = 0.01 \times \frac{1}{9.32} \text{ days}^{-1}$$

Note *mortality fraction k* and *disease time constant* α may vary considerably within a population and indeed post-vaccination – so treat with caution!



Note: as per the 'daily death rate' graphs in *World in Data*, we also apply a **seven-day moving average** to smooth the numerical derivative.

FIRST CV-19 WAVE IN UK

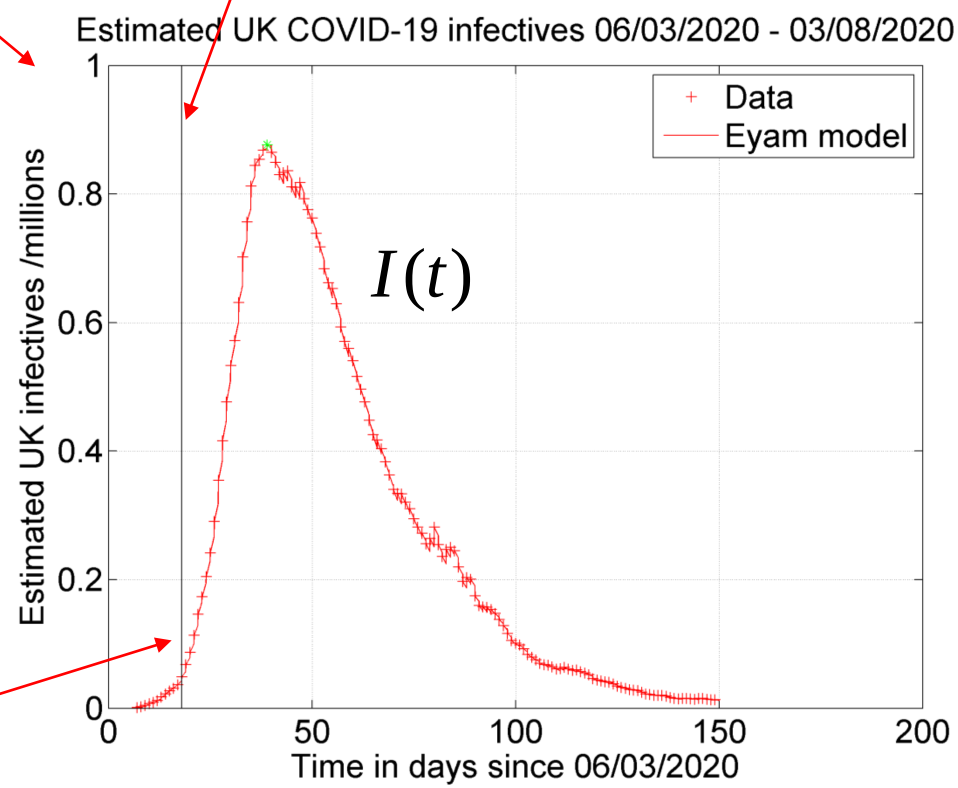


$$I_n = \frac{1}{k\alpha}\frac{dD}{dt} \approx \frac{1}{k\alpha}\frac{D_{n+1} - D_{n-1}}{t_{n+1} - t_{n-1}}$$

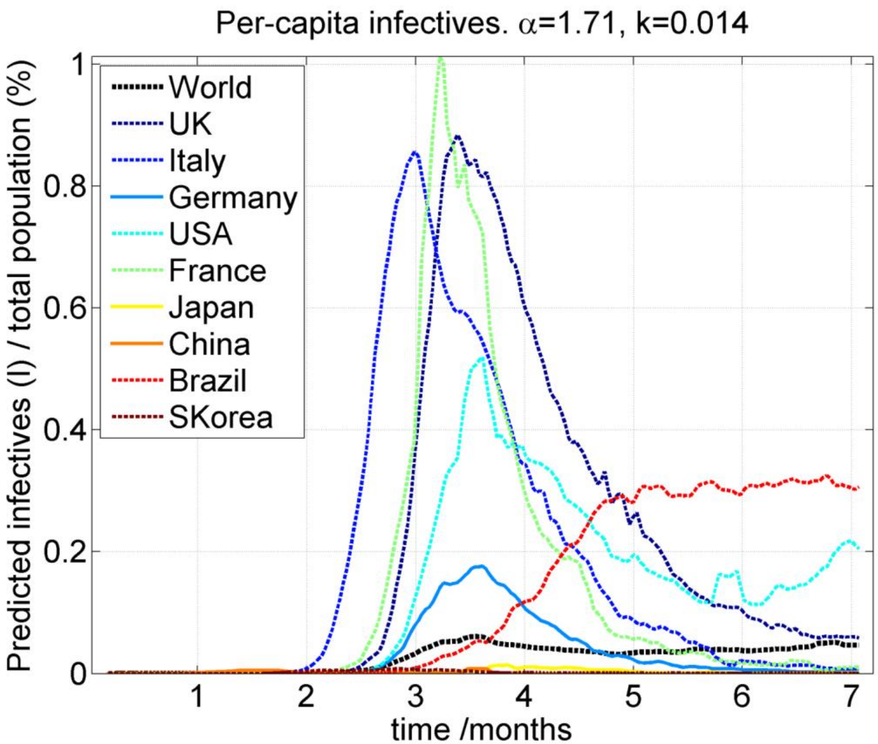
The vertical black line represents when the UK went into lockdown.

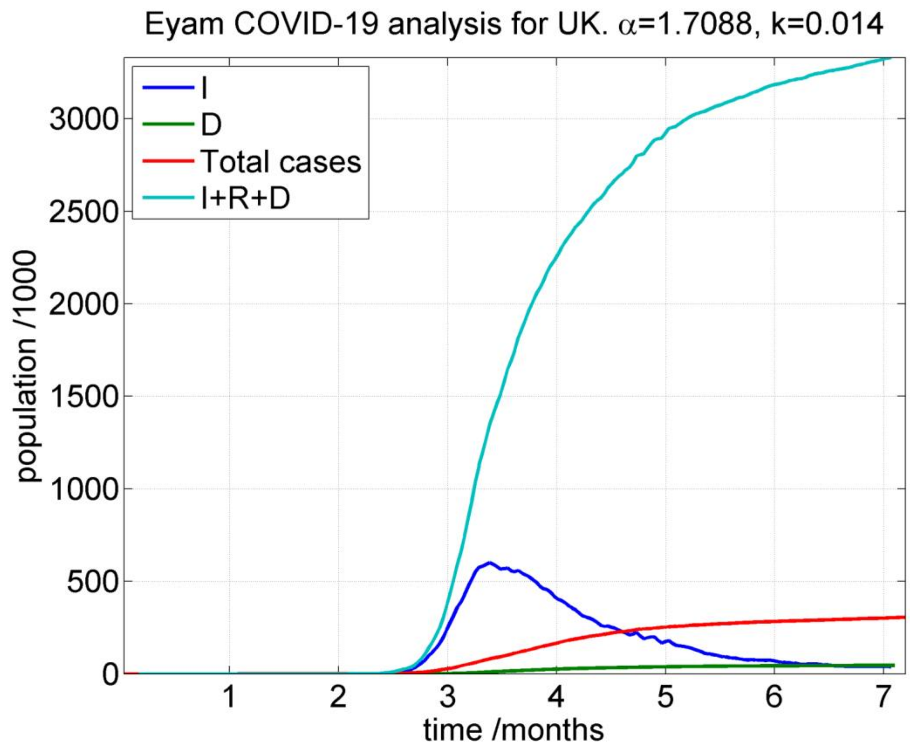
Find the gradient and scale by:

$$k\alpha = 0.01 \times \frac{1}{9.32}\ \text{days}^{-1}$$



Note: as per the 'daily death rate' graphs in *World in Data*, I also apply a **seven-day moving average** to smooth the numerical derivative.


Per-capita infectives. α=1.71, k=0.014
Predicted infectives (I) / total population (%)
time /months
World
UK
Italy
Germany
USA
France
Japan
China
Brazil
SKorea
0
0.2
0.4
0.6
0.8
1
1
2
3
4
5
6
7


Eyam COVID-19 analysis for UK. α=1.7088, k=0.014
population /1000
time /months
I
D
Total cases
I+R+D
0
500
1000
1500
2000
2500
3000
1
2
3
4
5
6
7

MATLAB UK COVID-19 data processing pipeline

```
%Get cumulative dead (D) and recovered (R) (in millions) from data.
%Time t in days.
[t,D,R] = get_D_R_from_data(k);

%Estimate Infective population vs time
[I,tmax,Imax] = estimate_I( t, D, k, a, W );

%Determine the beta value at the infection peak using a K&K model fit
bpeak = beta_from_peak( t,I,tmax,Imax,a,k );

%Use Eyam equations to step forward, and backward from peak to calculate S
%and beta (b) vs time
[b,S] = eyam_from_peak( t,I,tmax,Imax,bpeak,a );

%Determine 'at risk' population N
N = I + S + R + D;

%Define effective susceptible threshold at epidemic peak = alpha/beta
rho = a*(b.^-1);

%Apply Eyam numerical model for whole outbreak, using variable beta (b).
[tm,Im,Sm] = eyam_variable_b( I,S,a,b );

%Estimate basic reproduction number R0 vs time
R0 = estimate_R0(t,I);

%Calculate effective 'double-time' t_din days for infective population.
td = calculate_doubling_time( t,I );
```

This is what we have discussed so far…

To estimate β we can attempt a **K&K curve** fit near the peak. The K&K approximation works well near the Infective peak, but is less appropriate beyond it. (See "*Extending The Epidemiology of Eyam*")

$$\boxed{I \approx \frac{I_{\max}}{\cosh^2\left(\sqrt{\tfrac{1}{2} I_{\max} \dfrac{\beta}{\alpha}}\,(t-t_{\max})\,\alpha\right)}}$$

$$\cosh\left(\sqrt{\tfrac{1}{2} I_{\max} \frac{\beta}{\alpha}}\,(t-t_{\max})\,\alpha\right) = \sqrt{\frac{I_{\max}}{I}}$$

$$\sqrt{\tfrac{1}{2} I_{\max} \frac{\beta}{\alpha}}\,(t-t_{\max})\,\alpha = \cosh^{-1}\left(\sqrt{\frac{I_{\max}}{I}}\right)$$

$$\underbrace{\tfrac{1}{2} I_{\max}\,(t-t_{\max})^2\,\alpha}_{X}\,\beta = \underbrace{\left(\cosh^{-1}\left(\sqrt{\frac{I_{\max}}{I}}\right)\right)^2}_{Y}$$

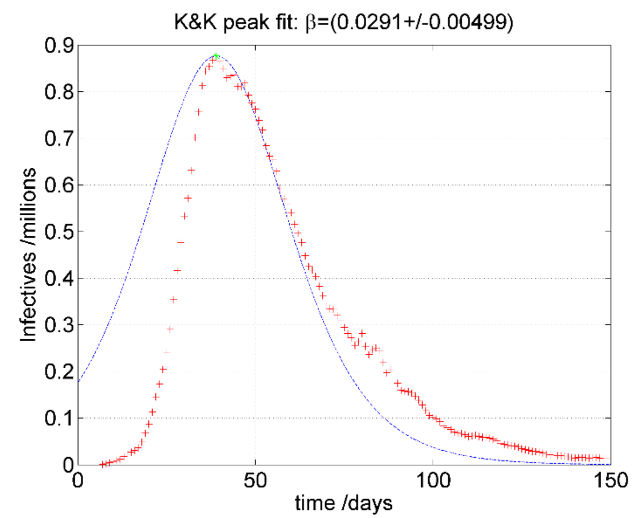
$$\beta X = Y$$

i.e. perform a **line of best fit** to find β

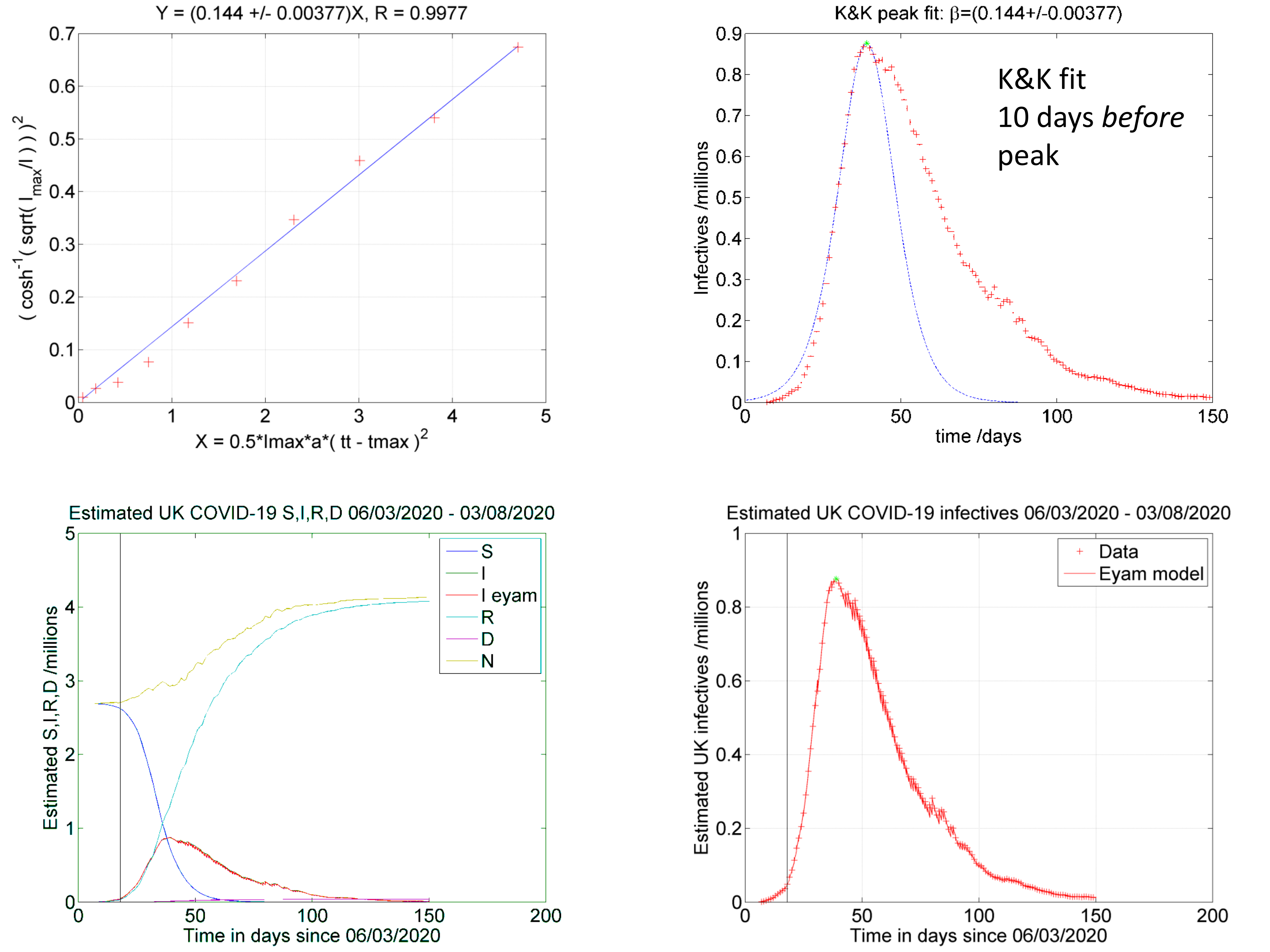
Plan:

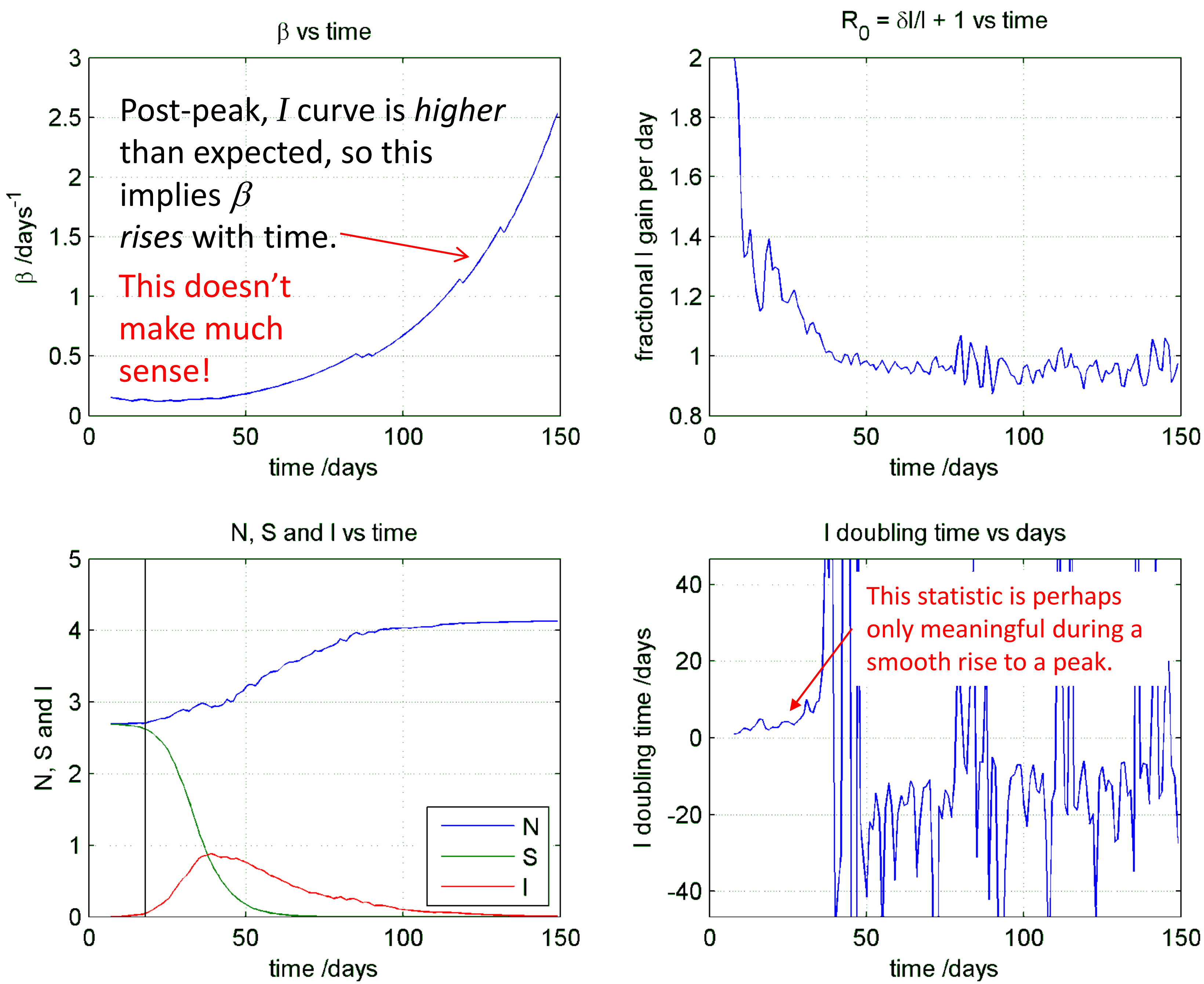
Do K&K ten days either side of the peak* to find beta, and **hence find S at peak = alpha/beta.**

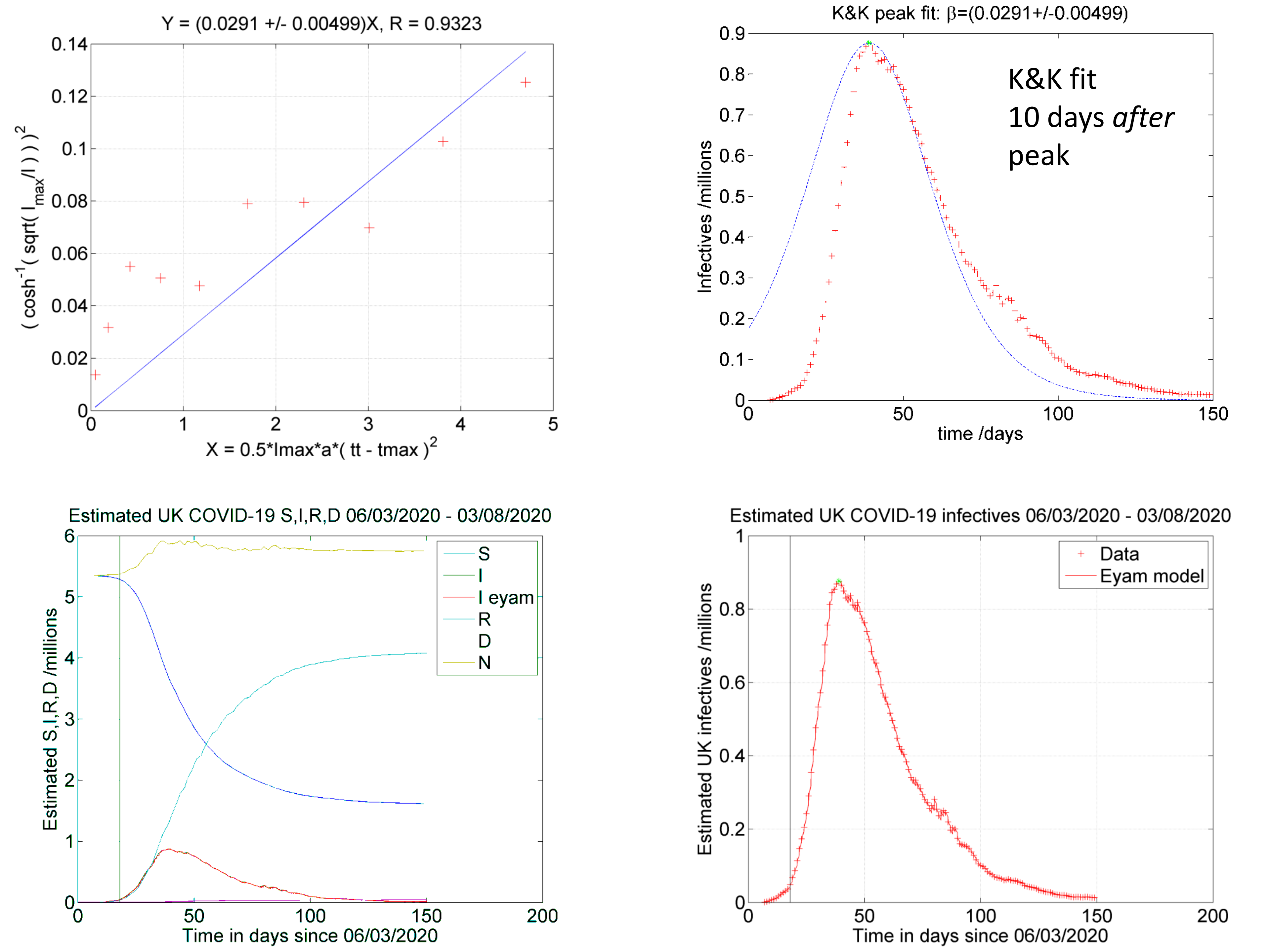
Then use Eyam equations forwards and backwards in time, and find beta from data.



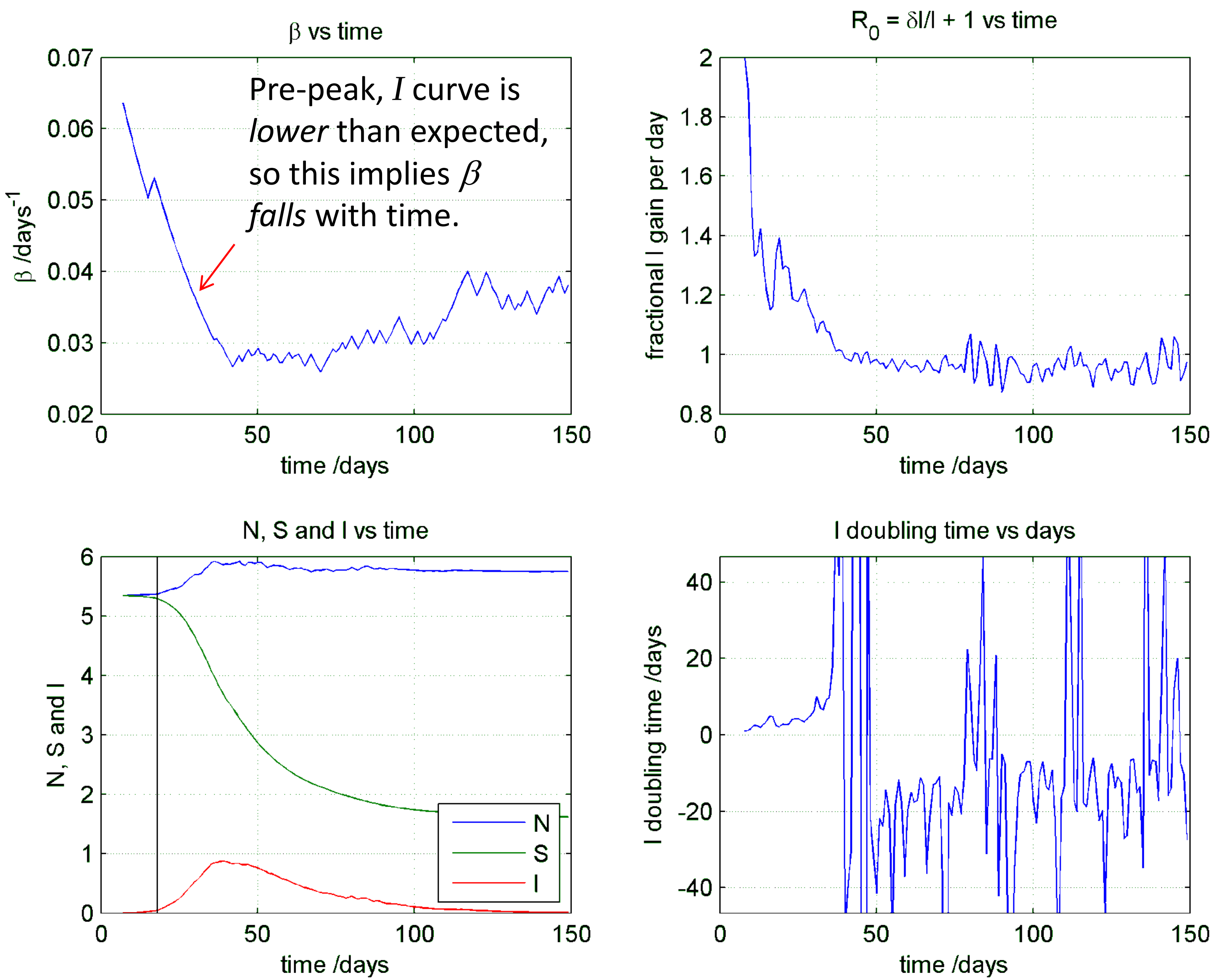
*We'll do 10 days before, and then 10 days after *separately*.


Y = (0.144 +/- 0.00377)X, R = 0.9977
(cosh^-1(sqrt(I_max/I)))^2
X = 0.5*Imax*a*(tt - tmax)^2
K&K peak fit: β=(0.144+/-0.00377)
K&K fit
10 days *before*
peak
Infectives /millions
time /days
Estimated UK COVID-19 S,I,R,D 06/03/2020 - 03/08/2020
Estimated S,I,,R,D /millions
Time in days since 06/03/2020
S
I
I eyam
R
D
N
Estimated UK COVID-19 infectives 06/03/2020 - 03/08/2020
Estimated UK infectives /millions
Time in days since 06/03/2020
Data
Eyam model


β vs time
Post-peak, I curve is *higher* than expected, so this implies β *rises* with time.
This doesn't make much sense!
β /days^{-1}
time /days
$R_0 = \delta I/I + 1$ vs time
fractional I gain per day
time /days
N, S and I vs time
N, S and I
N
S
I
time /days
I doubling time vs days
This statistic is perhaps only meaningful during a smooth rise to a peak.
I doubling time /days
time /days

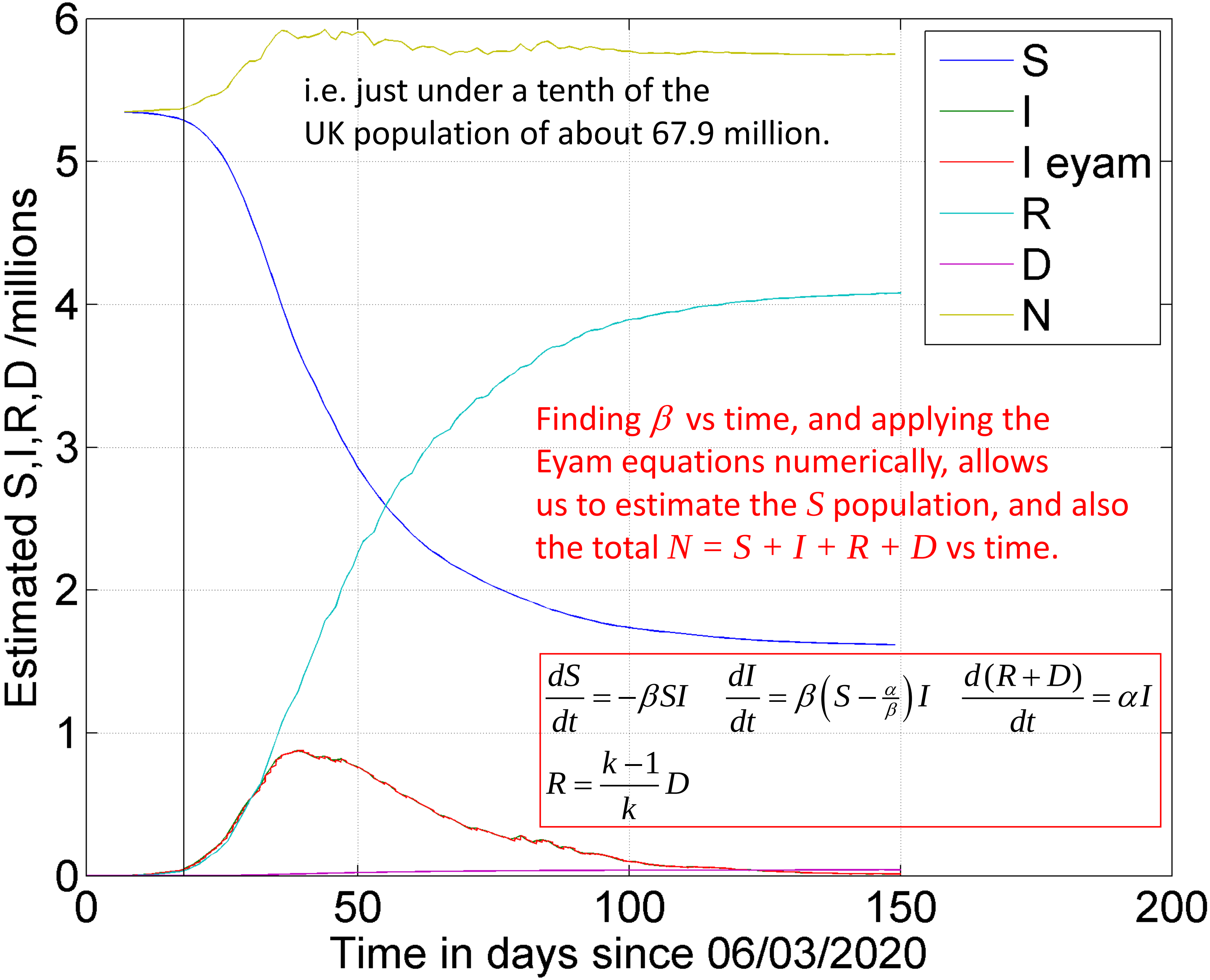


Y = (0.0291 +/- 0.00499)X, R = 0.9323
(cosh^-1(sqrt(I_max/I)))^2
X = 0.5*Imax*a*(tt - tmax)^2
K&K peak fit: β=(0.0291+/-0.00499)
K&K fit
10 days *after*
peak
Infectives /millions
time /days
Estimated UK COVID-19 S,I,R,D 06/03/2020 - 03/08/2020
S
I
I eyam
R
D
N
Estimated S,I,,R,D /millions
Time in days since 06/03/2020
Estimated UK COVID-19 infectives 06/03/2020 - 03/08/2020
Data
Eyam model
Estimated UK infectives /millions
Time in days since 06/03/2020



β vs time
0.07
0.06
0.05
0.04
0.03
0.02
β /days^-1
Pre-peak, I curve is lower than expected, so this implies β falls with time.
0
50
100
150
time /days
R0 = δI/I + 1 vs time
2
1.8
1.6
1.4
1.2
1
0.8
fractional I gain per day
time /days
N, S and I vs time
6
5
4
3
2
1
0
N, S and I
N
S
I
time /days
I doubling time vs days
40
20
0
-20
-40
I doubling time /days
time /days

Estimated UK COVID-19 S,I,R,D 06/03/2020 - 03/08/2020



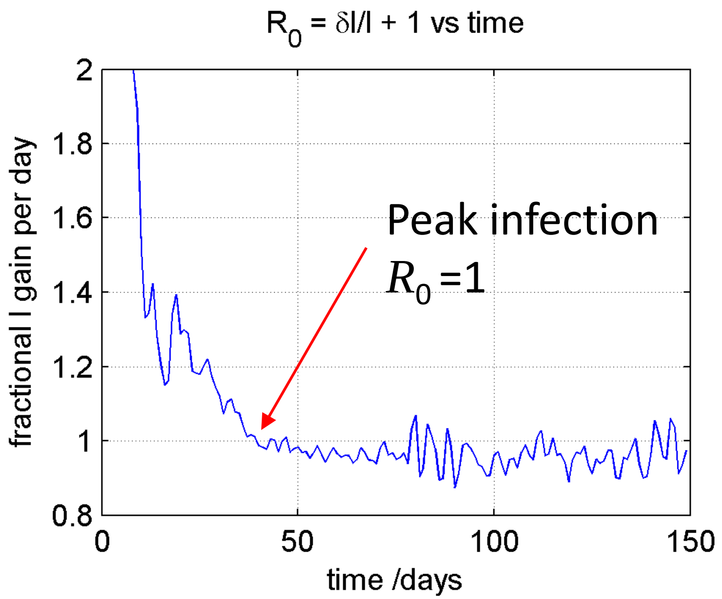
The **Basic Reproduction number** R_0 has been extensively quoted by UK Government during the COVID-19 pandemic – but what does it mean?

Well according to Pandit, it can have several meanings!

Mathematical notes on (various) meanings of Basic Reproduction number R_0
Pandit, J.J., "Managing the R0 of COVID-19: mathematics fights back." *Anaesthesia* 2020. doi:10.1111/anae.15151

During the pandemic, R_0>1 seemed to imply "the infectives rising" and R_0<1 implying "infectives falling". Although *annoyingly rarely defined rigorously*, this seems to imply the following definition: **the fractional change in infective population (per day), plus 1.**

$$R_0 = \frac{\delta I}{I} + 1$$

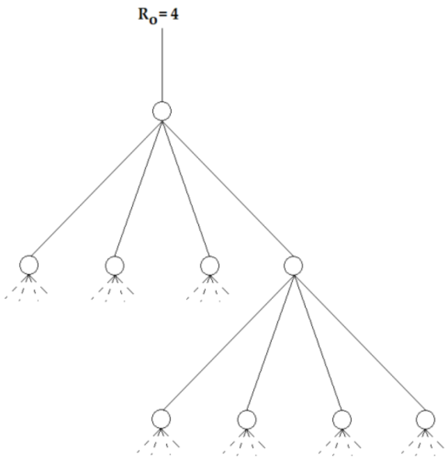


Alternative interpretations are on the following pages....

Basic reproduction number $R_0 = N/\rho$

[From analysis of Liberia Ebola outbreak]
https://iopscience.iop.org/article/10.1088/1361-6552/ab4a59

R_0 can be thought of as the number of susceptibles converted to infectives, for every one infective, per unit of time $\frac{1}{\alpha}$. You can see this from the 'Eulerization' of the Eyam equation $dS/dt = -\beta SI$:



$$\Delta S \approx -\beta SI\Delta t$$

$$\therefore \Delta S \approx -\beta S \times 1 \times \frac{1}{\alpha} = -\frac{S}{\rho} \approx -\frac{N}{\rho} = -R_0$$

So for $R_0 = 1.85$, this means Ebola will cause slightly less than two susceptibles to becomes infected for every infective, per unit time $\frac{1}{\alpha}$, which for our Ebola analysis is $\frac{1}{2.84} = 0.35$ months or ≈ 10.7 days.

R_0 is also directly related to a very important quantity, the minimum fraction $F_{\min}$ of the population to be immunized in order for 'herd immunity' (essentially a lack of susceptibles to catalyse an epidemic) to prevent the liklihood of a major epidemic.

$$\boxed{F_{\min} = P(\text{epidemic spreads}) = 1 - \frac{1}{R_0}}$$

Herd immunity as partial resistance, reflected in reductions in frequency of disease due to reductions in numbers of source cases and of susceptibles.

$$\boxed{F_{\min} = P(\text{epidemic spreads}) = 1 - \frac{1}{R_0}}$$

Ebola: $$F_{\min} = 1 - \frac{1}{1.85} = 45.9\%$$

Plague: $$F_{\min} = 1 - \frac{1}{1.68} = 40.5\%$$

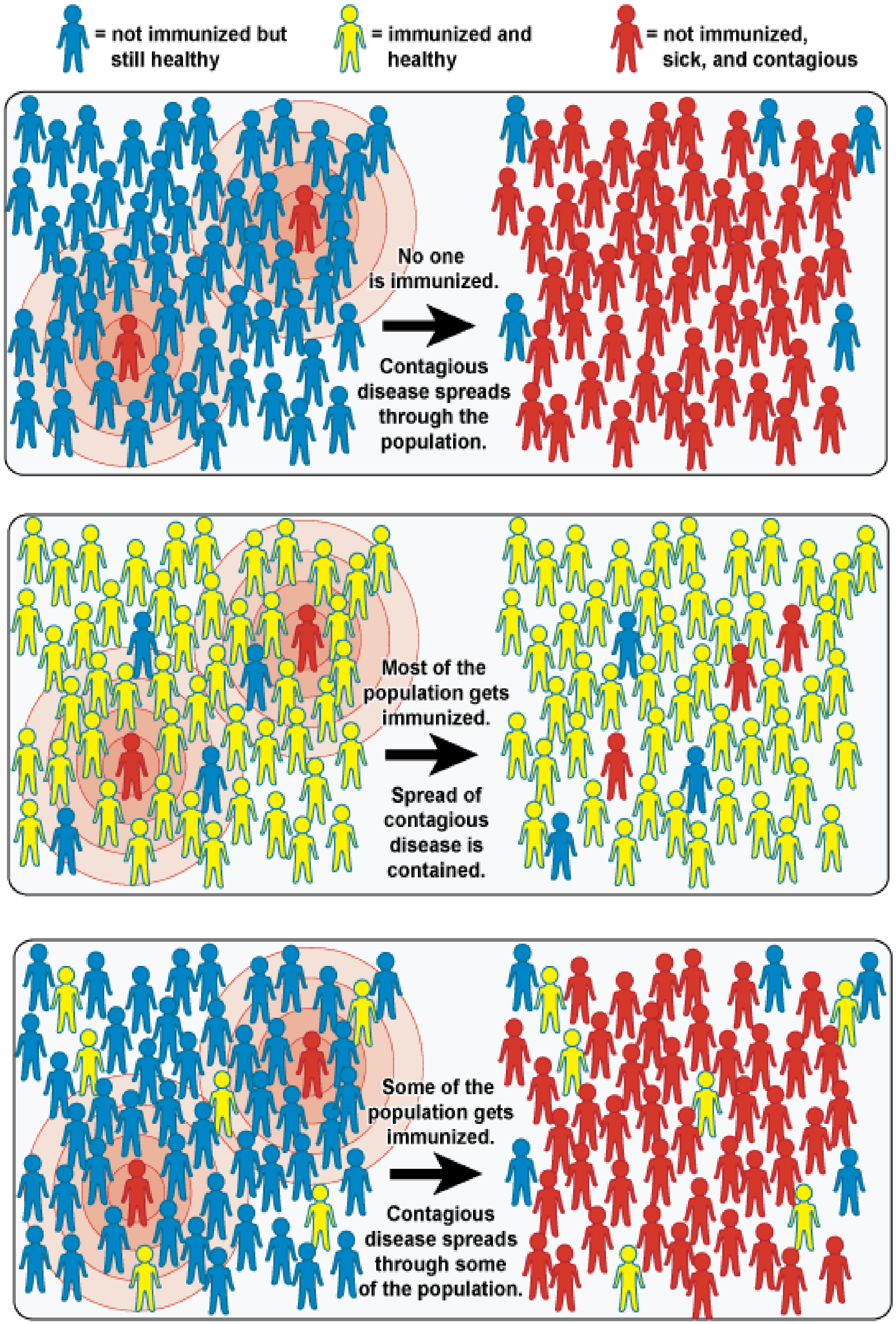
Measles: $$F_{\min} = 1 - \frac{1}{18} = 94.4\%$$

COVID-19?: $$F_{\min} = 1 - \frac{1}{3.00} = 66.7\%$$

Using Wuhan curve fit

vaccines.gov

A federal government Website managed by the U.S. Department of Health and Human Services



BPhO Computational Challenge

- Suggested homework
- Q&A

BPhO British Physics Olympiad