Data, Dynamics and COVID-19 summary

Justin Bennett

August 16, 2020

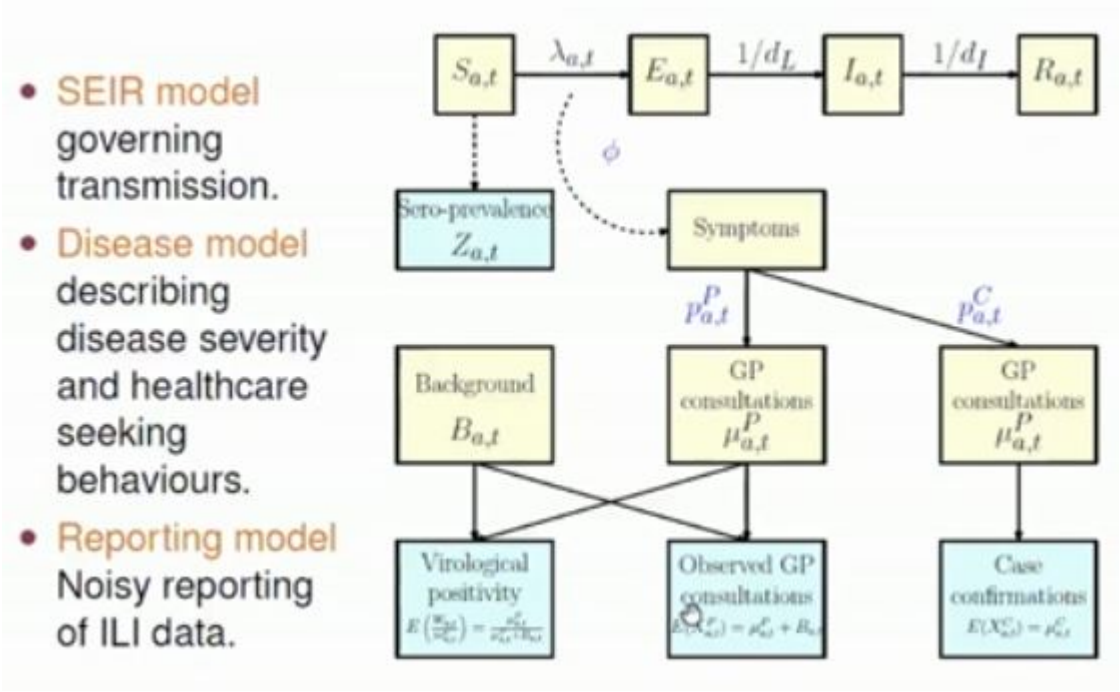
Contents

1 Week 1 Data and COVID-19 Introduction

The first week offered a general introduction to the topic of the summer school: the dynamics of epidemics, data, and COVID-19. We received an overview of the relevant problems, heard from experts in the field, and got a taste of the available data and models. There was not much mathematics and not much depth as of yet, but the week involved hearty preparation for the real work that is yet to come. Participants also got to know each other, learned how to navigate the virtual conference center, and explored options for future projects. A particular article of interest that we read was "The Mathematics of Infectious Diseases" by Herbert W. Hethcote, which published in SIAM Review in 2000. However, primary tasks for the week included participation in seminars by relevent experts, and work together to begin formulating research questions.

1.1 Nowcasting and forecasting COVID-19 in England

Biostatistician Daniela De Angelis (University of Cambridge) presented at the Isaac Newton Institute at Cambridge on May 22, 2020. De Angelis spoke about the development timeline and mentioned details of the simulator that is used to "nowcast" and forecast the COVID-19 pandemic in the UK. The simulator manifested from a code that was developed for the H1N1 virus pandemic in 2009; the government supported further development and the code was "ready" to go when COVID-19 hit. The code employs an age-structured Susceptible-Exposed-Infected-Recovered (SEIR) model, and parameters are updated as data become available.



1.2 Research question formulation

Experts, whose specialties ranged from mathematical epidemiology to biomathematics and statistics, included Cordelia McGehee (University of Minnesota), Jack O'Brien (Bowdoin College), Pauline van den Driessche (University of Victoria), and Jianhong Wu (York University). Andrew Roberts and Nick Ma (both of Cerner Corporation) served as practitioners. What are the important questions and considerations? The experts offered many suggestions. Here are a few (in no particular order):

• Be aware of the different ways clinical trial data are reported

• How can we capture resilience—of an infrastructure, for example—in a mathematical model? What type of data do we need to assess resilience?

• Can we model the influence of human activities (e.g., land use changes) on the spread of zoonotic diseases?

• There seems to be a one-week cycle in daily infection rates; is there a possible correlation with weekend sociability (stochastic forcing)?

• Develop granular (meta)population models for specific communities: long-term care facilities, homeless people, prison population, etc. • Introduce spatial variation, group differences, and delays in ODE models

• Design and evaluate strategies for de-escalation: social distancing, re-opening the economy, stratified lockdown, school reopening

• Devise data collection schemes for global monitoring of infectious diseases (citizen science?) and use livestock for the pilot project

• Study the effect of multiple simultaneous epidemics via co- infection models

• Investigate scenarios pertaining to the timing of recovery transitions for long-term care facilities separately from the general population

• Different diseases may have shared symptoms. How do we distinguish them?

• Parameter estimation, role of parameters and their uncertainties in compartmental models

• What have we learned? Epidemics will continue to occur, so we must develop strategies for avoiding future outbreaks.

2 Week 2 - Models and Mathematics

During week two, participants learned about models as a tool for understanding the evolution of an epidemic: how it begins with one infected individual among a population of susceptible individuals, spreads through contact between infected and susceptible individuals, reaches a maximum, and ultimately descends until the probability of infection becomes too small and the epidemic is no longer sustained. The magic number is R_0 (R naught) — the cumulative number of infected individuals generated through the course of the epidemic by a single infected person.

2.1 Compartmental models

The Susceptible-Infectious-Recovered (SIR) model is the fundamental model of epidemiology. The population of interest is subdivided into three groups of individuals: Susceptible, Infected, and Recovered. Each group is homogeneous—no spatial or other dependencies—and occupies a compartment. One can then think of an epidemic as flowing through the compartments, from S to I to R. Depending on the problem of interest and the desired level of granularity, researchers can generalize or modify the SIR model in many ways; well-known variants include SEIR (accounting for Exposed individuals), SIS (accounting for reinfection), and MSEIR (accounting for newborn individuals who inherit temporary immunity from their Mothers). These compartmental models are all deterministic and mathematically described by ordinary differential equations. They are also known as conceptual models — caricatures of the real world, in the sense that they are simplified to be amenable to mathematical analysis while retaining an epidemic's characteristic features. One characteristic of an epidemic is that the number of infected individuals initially increases, then decreases after reaching a maximum.

2.2 Stochastic models

Linda Allen (Texas Tech University) showed participants how to formulate stochastic models and numerically simulate sample paths via Markov Chain Monte Carlo (MCMC) techniques. Stochastic models work particularly well for smaller populations and provide quantitative information about uncertainties. Below are two examples, DTMC (discrete time markov chain) and CTMC (continuous time Markov Chain)

```
clear all
%close all
set(0,'DefaultAxesFontSize', 18)
set(gca,'fontsize',18);

init=1; %I(0)=init;
%Parameters
g=0.25;
beta=2*g; %Change constant 1.5, 2, 4
R0=beta/g;
N=100;
```

```
sim=3;
time=80;
ddt=.01;  %Timestep
count=0; % Checks if any Prob>1
  for k=1:sim
    clear  t s i r tot totev
    cas=init;
    t(1)=0;
    i(1)=init;
    s(1)=N-init;
    r(1)=0;
    tot=N;
    j=1;
    while i(j)>0;
        u=rand; %uniform random number
        ev1=beta*i(j)*s(j)/tot*ddt;
        ev2=ev1+g*i(j)*ddt;% should be <1
        if ev2>1
            count=count+1;%should always be zero
        end
        if (u<=ev1) % transmission
           i(j+1)=i(j)+1;
           s(j+1)=s(j)-1;
           r(j+1)=r(j);
           cas=cas+1; %count total number of cases
        elseif u>ev1 && u<=ev2 %recovery
            i(j+1)=i(j)-1;
            r(j+1)=r(j)+1;
            s(j+1)=s(j);
        else             %%no change
            s(j+1)=s(j);
            i(j+1)=i(j);
            r(j+1)=r(j);
        end
        j=j+1;
        tot=s(j)+i(j)+r(j);
        t(j)=t(j-1)+ddt;
    end
    casetot(k)=cas; %total cases
    Tend(k)=t(j); %time epidemic ends
    figure(1)
   if k==1
     stairs(t,i,'r-','Linewidth',2)
     xlabel('Day')
     ylabel('Infectious Individuals')
   elseif k==2
     stairs(t,i,'b-','Linewidth',2);
   elseif k==3
    stairs(t,i,'g-','Linewidth',2);
   end
    hold on

end
x(1)=N-init;
```

```
y(1)=init;
z(1)=0;
tot=N;
dt=.005;
sumcases=init;
for k=1:time/dt % Euler's method
    x(k+1)=x(k)+dt*(-beta*x(k)*y(k)/tot);
    y(k+1)=y(k)+dt*(beta*x(k)*y(k)/tot-g*y(k));
    z(k+1)=z(k)+dt*(g*y(k));
    sumcases=sumcases+dt*(beta*x(k)*y(k)/tot); %%total number of cases
    tot=x(k+1)+y(k+1)+z(k+1);
end
plot([0:dt:time],y,'k-','LineWidth',2);
hold off
 str = sprintf( 'DTMC SIR, R_0 = %.2f',R0);
 sgtitle(str);
disp((sprintf('--------------------------------------------------------------------------------
disp((sprintf('N= %5.0f, gamma= %5.2f, beta= %5.2f, R0= %2.2f, I(0)= %2.0f', N, g, beta, R0
disp((sprintf('Case1=%3.0f, Case2=%3.0f, Case3=%3.0f, ODECase=%3.2f', casetot(1), casetot(2
disp((sprintf('Tend1=%5.2f, Tend2=%5.2f, Tend3=%5.2f', Tend(1), Tend(2), Tend(3))))
disp((sprintf('Prob>1 = %5.0f', count)));
disp((sprintf('--------------------------------------------------------------------------------
```

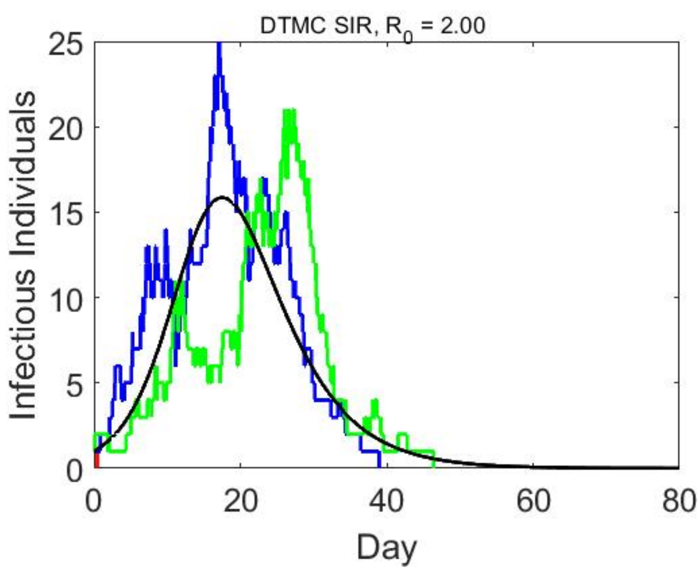


Figure 1: SIR model using discrete time markov chain

```
clear all
%close all
set(0,'DefaultAxesFontSize', 18)
set(gca,'fontsize',18);

init=1; %I(0)=init;
%Parameters
g=0.25;
beta=2*g; %CHANGE constant 1.5, 2, 4
R0=beta/g;
N=100;
sim=3;
time=80;
  for k=1:sim
    clear  t s i r tot totev
    cas=init;
    t(1)=0;
    i(1)=init;
    s(1)=N-init;
    r(1)=0;
    tot=N;
    j=1;
    while i(j)>0;
        u1=rand;
        u2=rand;
        totev=(beta/tot)*i(j)*s(j)+g*i(j);
        t(j+1)=-log(u1)/(totev)+t(j); %Interevent time
        ev1=beta*i(j)*s(j)/tot/totev;
        ev2=ev1+g*i(j)/totev; %%should equal 1
        if (u2<=ev1) %Transmission
           i(j+1)=i(j)+1;
           s(j+1)=s(j)-1;
           r(j+1)=r(j);
           cas=cas+1; %count total number of cases
        elseif u2>ev1 && u2<=ev2 %recovery
            i(j+1)=i(j)-1;
            r(j+1)=r(j)+1;
            s(j+1)=s(j);
        end
        j=j+1;
        tot=s(j)+i(j)+r(j);
    end
    casetot(k)=cas; %total number of cases
    Tend(k)=t(j); %time epidemic ends
    figure(2)
   if k==1
     stairs(t,i,'r-','Linewidth',2)
     xlabel('Day')
     ylabel('Infectious Individuals')
   elseif k==2
     stairs(t,i,'b-','Linewidth',2);
   elseif k==3
    stairs(t,i,'g-','Linewidth',2);
   end
```

```
    hold on
end
x(1)=N-init;
y(1)=init;
z(1)=0;
tot=N;
dt=.005;
sumcases=init;
for k=1:time/dt % Euler's method
    x(k+1)=x(k)+dt*(-beta*x(k)*y(k)/tot);
    y(k+1)=y(k)+dt*(beta*x(k)*y(k)/tot-g*y(k));
    z(k+1)=z(k)+dt*(g*y(k));
    sumcases=sumcases+dt*(beta*x(k)*y(k)/tot); %%total number of cases
    tot=x(k+1)+y(k+1)+z(k+1);
end
plot([0:dt:time],y,'k-','LineWidth',2);
hold off
 str = sprintf( 'CTMC SIR, R_0 = %.2f',R0);
 sgtitle(str);

disp((sprintf('----------------------------------------------------------------------------------------------------
disp((sprintf('N= %5.0f, gamma= %5.2f, beta= %5.2f, R0= %2.2f, I(0)= %2.0f', N, g, beta, R0
disp((sprintf('Case1=%3.0f, Case2=%3.0f, Case3=%3.0f, ODECase=%3.2f', casetot(1), casetot(2
disp((sprintf('Tend1=%5.2f, Tend2=%5.2f, Tend3=%5.2f', Tend(1), Tend(2), Tend(3))))
disp((sprintf('----------------------------------------------------------------------------------------------------
```

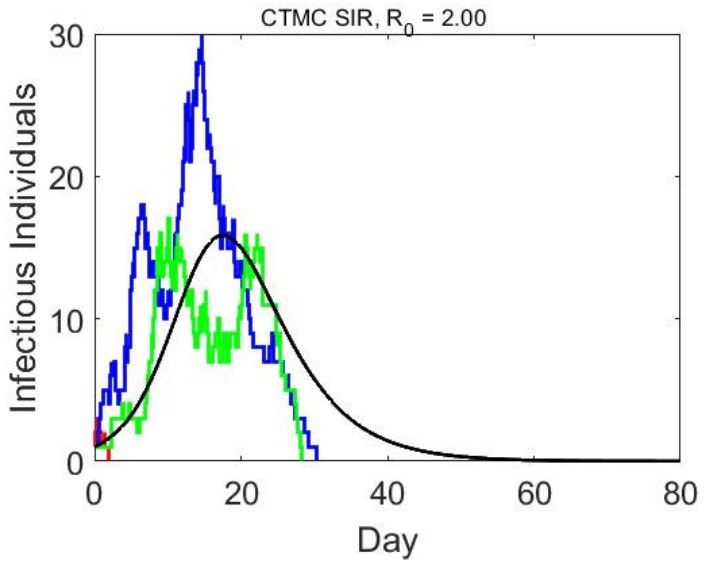
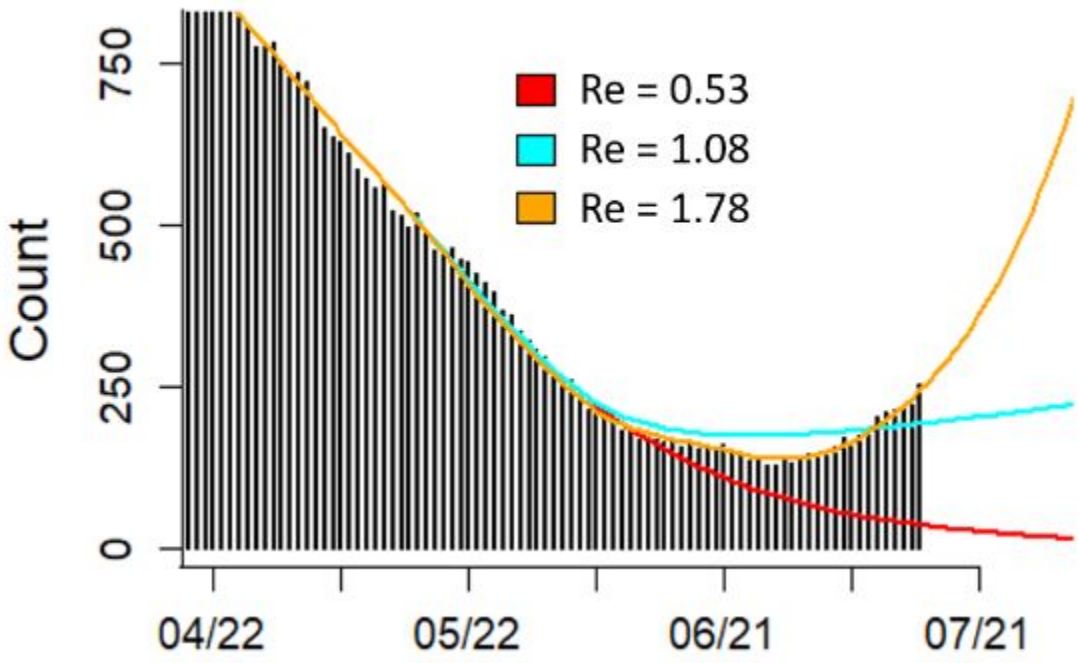


Figure 2: SIR model using continuous time markov chain

2.3 Mathematical Modeling of COVID-19 in Colorado

David Bortz's (University of Colorado Boulder) presented a talk on "Mathematical Modeling of COVID-19 in Colorado," in preparation for group discussions and a QA session with David on Thursday morning. Bortz is a member of the Colorado Covid-19 Task Force, so he spoke from experience about how state officials (including the governor) use science to make informed decisions about social distancing and mask policies.



Model documentation and updated parameter estimates are hosted on ColoradoCoronavirus-Model.com under the "Documentation" tab.
Code for the model (in R) is available on GitHub: https://github.com/agb85/covid-19

2.4 Introduction to heterogeneous frameworks

Whether deterministic or stochastic, SIR and similar models assume population homogeneity. The question of how to account for heterogeneity then arises. Nancy Rodriguez (University of Colorado Boulder) explained four "frameworks" that scientists popularly use to incorporate spatial interactions:

- Agent-based models / interacting-particle systems[3]
- Metapopulation models / patch models[1]
- Reaction-advection-diffusion equations / partial differential equations[2]
- Integro-differential equations[4]

Selected references that were studied during the workshop have been added to the end of this report

3 Week 3 Data Assimilation

Chris Jones (North Carolina - Chapel Hill) started the week with an introduction to data assimilation (DA) that emphasized three perspectives: DA as an application of optimization (variational methods), DA as an interpolation scheme (Kalman filters), and DA as an application of probability theory (Bayesian inference). Some offered applications and examples were improved numerical weather prediction, explanations of the cause of the fish kill in Kinneret Lake in Israel, and use of satellite data to visualize the ozone hole around the South Pole.

3.1 Ensemble Kalman Filter

To effectively practice Data assimilation, we coded DA on a susceptible- infectious-recovered (SIR) model using MATLAB. For this particular example, I coded Steve Schecter's (North Carolina State) SIX model:

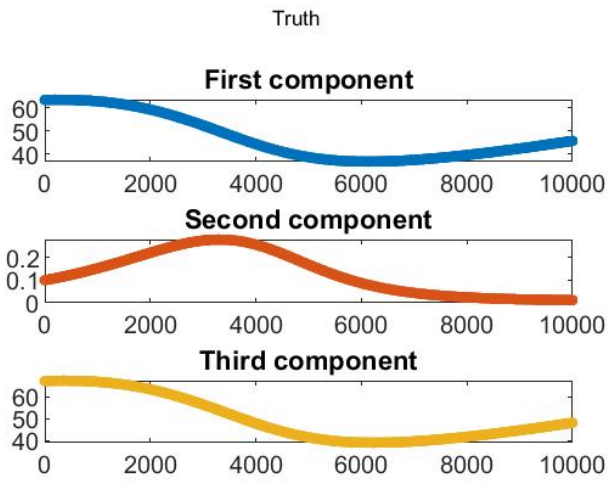
```
%% Ensemble Kalman Filter (EnKF) Example
% This function uses the EnKF to make predictions of an n-dimensional
% dynamical system, defined by F. It uses a possibly nonlinear observation
% operator H and linearizes where appropriate.

%% Set parameters
dim      = 4;              % dimension of physical model problem
obsdim      = 4;              % dimension of observation space
NEns = 5;
numsteps = 1e5;            % number of timesteps
t0       = 0;
tf       = 1e4             % final time
dt = tf/numsteps;          % step size

gamma    = 5e-1*eye(obsdim);   % variance of obs noise
sigma    = 2e-1*eye(dim); % variance of model noise
C0       = 5e-1*eye(dim); % prior initial condition variance
```

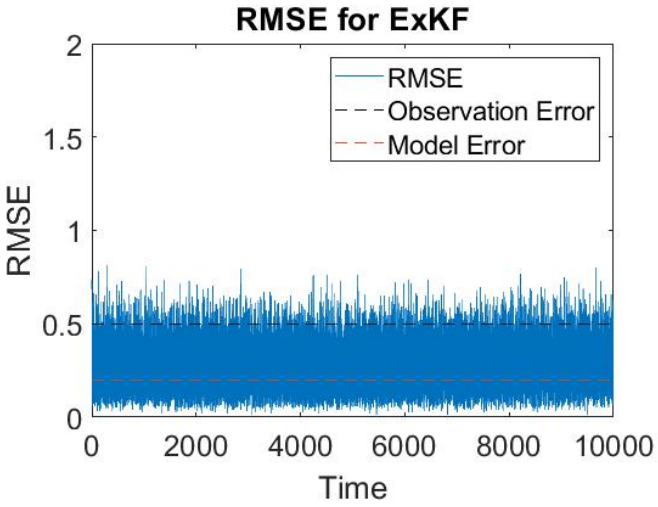
```
18
19 sd       = 1; rng(sd);         % set seed for replicating results
20 color = get(gca,'colororder'); % for plotting
21
22 tpn = 1e3; %For printing during DA
23 dtp = tpn;
24
25 %% Initialize
26 %IC=[1 1 1];
27 %Ffun = @lor63
28
29 %W. Liu, M. Li model
30 %IC=[35 0.09555 67];
31 %Ffun = @sir
32
33 %S. Schechter model
34 IC=[.97,.03,.98,.02];
35 Ffun = @six
36
37 % spin up to generate initial condition
38 [times,a]=ode45(Ffun,[0,100],IC);
39 IC=a(end,:);
40
41 %Display the "Truth"
42 runod45(Ffun,IC,tf);
43
44 v = IC + randn(1,dim); % initial truth
45 m = IC + randn(1,dim)*sqrt(C0); % initial mean/estimate
46 y = H(0,v) + (gamma*randn(obsdim,1));
47
48 %For SIR type models only: Cutoff for negative values
49 y = cutoff(y);
50
51 %Initialize the ensemble members
52 Ens = zeros(NEns,dim);
53 for j=1:NEns
54     Ens(j,:) = IC + randn(1,dim)*sqrt(C0); % + Noise from N(0,C0)
55     Ens(j,:) = cutoff(Ens(j,:));
56 end
57
58 eta = 2e-1;               % stabilization coefficient 0 < eta << 1
59 c = C0/eta;          % covariance
60
61 NormC = norm(c)
62
63 %% Assimilation loop
64 t = t0;
65 for j = 1:numsteps
66
67     % push truth forward in time
68     v=formod(Ffun,t,dt,v);
69     %For SIR type models only: Cutoff for negative values
70     v = cutoff(v);
71
```

```
    % get observation
    y=H(t,v) + (gamma*randn(obsdim,1));
    %For SIR type models only: Cutoff for negative values
    y = cutoff(y);

    %Extended KF
    %[m,c] = ExKF(t,dt,Ffun,m,y,c,sigma,gamma);
    %Ensemble KF
    [m,c,Ens] = EnKF(t,dt,Ffun,m,Ens,y,c,sigma,gamma);


    NormC=max(NormC,norm(c));

    rmse(j) = sqrt(sum((m-v).^2)/dim);
    t=t+dt;

    if t > tpn
       fprintf('t = %d\n',t)
       tpn = tpn+dtp;
    end

end
%% End: assimilation loop

Tfinal = t
NormC

%Display the RMSE vs. time
figure(2)
plot(linspace(t0,tf,numsteps),rmse)
hold on
plot([t0 tf], [norm(gamma) norm(gamma)],'k--')
plot([t0 tf],[norm(sigma) norm(sigma)],'--','Color',color(2,:))
axis([t0 tf 0 2])
legend('RMSE','Observation Error','Model Error')
xlabel('Time')
ylabel('RMSE')
title('RMSE for ExKF')
```

```
%% SIx model from S. Schechter's paper %%
% Inputs:
% T = time, unused but required for Matlab's ode45
% a = state variable at time T
%
% Outputs:
% aprime = derivative given by SIX model
%%
parameters
epsilon = 2.E-3;
ba = 0.1;
bn = 0.5;
gamma = 1/6;
k = 0.3;
```

```
mn = 5;
ma = 2;

%aprime= [-(bn*x(3)+ba*(1-x(3)))*x(1)*x(2); ...
%          (bn*x(3)+ba*(1-x(3)))*x(1)*x(2) - gamma*x(2); ...
%         x(3)*(1-x(3))*(ba-bn)*x(2) + x(3)*(1-x(3))*(k-(mn-ma)*x(2))/epsilon; ...
%         -(x(3)*(1-x(3))*(ba-bn)*x(2) + x(3)*(1-x(3))*(k-(mn-ma)*x(2))/epsilon)];
%In t: (4.1)-(4.3)
%aprime= [-(bn*x(3)+ba*x(4))*x(1)*x(2); ...
%          (bn*x(3)+ba*x(4))*x(1)*x(2) - gamma*x(2); ...
%         x(3)*x(4)*(ba-bn)*x(2) + x(3)*x(4)*(k-(mn-ma)*x(2))/epsilon; ...
%         -(x(3)*x(4)*(ba-bn)*x(2) + x(3)*x(4)*(k-(mn-ma)*x(2))/epsilon)];
%In tau: (4.4)-(4.6)
aprime= [epsilon*(-(bn*x(3)+ba*x(4))*x(1)*x(2)); ...
         epsilon*((bn*x(3)+ba*x(4))*x(1)*x(2) - gamma*x(2)); ...
         epsilon*(x(3)*x(4)*(ba-bn)*x(2) + x(3)*x(4)*(k-(mn-ma)*x(2))/epsilon); ...
         epsilon*(-(x(3)*x(4)*(ba-bn)*x(2) + x(3)*x(4)*(k-(mn-ma)*x(2))/epsilon))];
end
```





3.2 An international assessment of the COVID-19 pandemic using ensemble data assimilation

Dr. Evensen authored a paper which provides an overview of the Ensemble Kalman filter (EnKF) technique for parameter estimation, with applications to the SEIR model for COVID-19. The SEIR model is simple to understand and formulate, and tends to successfully model epidemics. The code $EnKF_seir$ (available on GitHub) was developed as a tool for decision-making in Norway, especially with regard to school reopening plans. It is an age-structured model with 11 age groups and uses data from hospitals and care homes. Evensen's paper includes results for many countries and states. In his talk, Evensen summarized the technique and illustrated the results for Norway, the Netherlands, and four states in the US.

3.3 Data Fusion and Forecasting

Sara Del Valle (Los Alamos National Laboratory) offered a planetary talk at the virtual 2020 SIAM Conference on the Life Sciences. Del Valle spoke about "Real-time Data Fusion to Guide Disease Forecasting Models" and offered an overview of Los Alamos' work for real-time forecasting of epidemics. The work was inspired by a challenge issued by the Centers for Disease Control and Prevention (CDC) to forecast the onset, peak timing, and peak intensity of flu epidemics, based on influenza-like illness data that the CDC and other data streams make available on a weekly basis. The basic model is an SIR model with various statistical analysis techniques. Del Valle also shared results of dengue fever in Brazil and COVID-19 in the U.S.

4 Week 4-6 Group Projects

The final three weeks were spent more on specific group research and project work, and less on limited lectures and seminars. The learning process, which had occupied most of Weeks 1-3, was continued on a smaller scale with two special lectures. However, Steffen Eikenberry (Arizona State) gave a talk on "To Mask or Not to Mask," and Linda Allen (Texas Tech) returned with a tutorial talk on "Branching Processes." Linda's talk was a follow-up on her talk during Week 2. This time, she focused on estimating the probability of an outbreak of an epidemic using branching processes.

4.1 To Mask or Not to Mask

Steffen's talk addressed the effectiveness of facemasks to curtail the spread of a pandemic. His talk was based on a recent paper published online in Infectious Disease Modeling (April 21, 2020). Apparently, the topic was already discussed in an early paper by A.J. Jessup, published in Scientific American Supplement No. 143 (1878) but remains controversial and is sometimes painted as a new symbol of tyranny. The talk included details about the Manchurian pneumonic plague (1910-'11), the Spanish influenza pandemic (1918-'19), the SARS epidemic (2002-'04), and the current Covid-19 pandemic. Steffen emphasized that the effectiveness of wearing masks can be studied effectively with an extended version of the SEIR model. The bottom line was clear: universal application of mask use is critical to curtail the spread of a pandemic.

4.2 Spatial Modeling of COVID-19

Henri Berestycki (EHESS, Paris) gave a lecture on "Spatial Modeling of Covid-19", describing mathematical models with nonlocal infections, leading up to the Pandemic Threshold Theorem. Other topics covered the pandemic spread in a periodic medium, traveling waves, SIR with diffusion, and a recent case study by Gatto et al. (PNAS, 2020) in Italy of the effect of expressways on the spread of disease. For the latter study, the authors used an augmented SIR model (SIRT) for the spread of an epidemic in a half-space, including Traveling by infected individuals on the boundary.

4.3 Two-population model

My group was interested in exploring the application of ensemble data assimilation to real data and an epidemiological model - for this purpose, we sought to extend Geir Evensen's "International Assessment..." paper. The summary is separate from this report.

References

[1] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani, *Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model*, Journal of Computational Science, 1 (2010), p. 132–145.

[2] H. Berestycki, J.-M. Roquejoffre, and L. Rossi, *Propagation of epidemics along lines with fast diffusion*, arXiv: Analysis of PDEs, (2020).

[3] M. E. Halloran, N. M. Ferguson, S. Eubank, I. M. Longini, D. A. T. Cummings, B. Lewis, S. Xu, C. Fraser, A. Vullikanti, T. C. Germann, and et al., *Modeling targeted layered containment of an influenza pandemic in the united states*, Proceedings of the National Academy of Sciences, 105 (2008), p. 4639–4644.

[4] J. Medlock and M. Kot, *Spreading disease: integro-differential equations old and new*, Mathematical Biosciences, 184 (2003), p. 201–222.