Humboldt State University
Environmental Resources Engineering
ENGR 326 (41787-2) – Computational Methods for Engineers III
Beth Eschenbach, Ph. D., David Narum, Ph. D.

Trends in COVID-19 transmission using a modified susceptible, infected, removed model to recommend Humboldt State University campus reopening date.

Client: Tom Jackson, Jr., - Humboldt State University
Submitted by Caleb Wegener – Wegener Engineering
December 11, 2020

Table of Contents

Abstract

The coronavirus pandemic has closed college campuses across the USA. Humboldt State University (HSU) president Tom Jackson, Jr. hired Wegener Engineering to model coronavirus transmission using a modified SIR model to determine a safe date to open HSU campus. Using the model and vaccination estimates, a recovered and vaccinated population reaches 67 percent of the total population by the end of 2021 resulting in herd immunity.

Introduction

Early in 2020, the coronavirus arrived in the USA and forced many universities to close their campuses without warning, resulting in an abrupt switch to online learning (OL). Students and faculty have been working to adapt to OL ever since. The question on many people's mind is, when will it be possible to return to face-to-face instruction (FFI)?

Due to the highly contagious coronavirus and subsequent social distancing mandates in the United States, Humboldt State University (HSU) has been forced to close its campus to the majority of campus activities, including FFI, sporting events, seminars, and many student services. The first case of coronavirus in the United States was reported on January 21, 2020 (Matrajt and Leung 2020). Since then, nearly 13 million Americans have contracted the virus and more than 260,000 of those individuals have died (Johns Hopkins). Following recommendations by epidemiological experts, strict social distancing measures have been put in place across the country. Universities, like the California State University (CSU) system, switched to OL in March 2020. Most classes at HSU are being held online for Fall 2020 and are scheduled to remain so for the Spring 2020 semester.

Client

Tom Jackson, Jr., approached Wegener Engineering on August 31, 2020, asking for analytical information about coronavirus transmission so that decisions can be made at HSU concerning reopening the campus to the student body. HSU is a CSU campus in Arcata, California, a town of 18,431 individuals (Census.gov), including 6,983 students and 534 faculty at HSU (Humboldt.edu). After subsequent meetings, Wegener Engineering was hired to investigate a time when the HSU campus may safely reopen. To safely reopen, the risk of community transmission of the virus must not be high, and coronavirus case counts should be at a manageable level across the state.

Method

A Susceptible (S), Infected (I), and Removed (R) (SIR) model (Cooper et al. 2020) for epidemics can be used to predict future number of cases. The SIR model used to determine the future transmission of the coronavirus is a modified SIR model; a standard SIR model operates on several assumptions that are not accurate in the case of the coronavirus pandemic (discussed further in the Review of the Problem below).

Outcome

Preliminary results suggest that the pandemic is not going away quickly; about four percent of the American population have contracted the virus so far and are the only population with known immunity. It is clear there is the potential for more peaks because 96 percent of the population still susceptible to the virus. Reopening campus by fall semester 2021 is a conservative estimate and one that relies heavily on rollout of a vaccine; although, HSU is determined to reopen campus as soon as it is safe.

Objectives

The main objective of the project is to recommend a campus reopening date that does not risk the health of any students, staff, or faculty at HSU, nor the residents of Humboldt County. To achieve this primary objective, a mathematical model employing numerical methods and historical data will be used to determine trends in virus transmission and recommend a date when campus can be reopened. A computer program will be provided to Mr. Jackson allowing the model to be updated to improve modeling as more data becomes available

Report structure

Next, more information about the problem will be given and the importance of solving the problem is discussed. Afterwards, the design and solution steps will be presented. Finally, the results are presented and discussed, followed by the final recommendation.

Review of the problem

While students and faculty alike are anxious to know when campus may be safely reopened, it is not clear when it will be safe to return to face-to-face instruction (FFI). Because many HSU students are from outside the local area, all computer modeling for the project will include data from the entire state of California. Until recently, the coronavirus data for Humboldt County did not reflected national or state trends in coronavirus case numbers.

Importance of solving the problem

It is important to accurately determine a campus reopening dates to: 1) protect the health and safety of the campus and local community, 2) get students back in the classroom, allowing social interaction once more and giving them access to campus resources, and 3) relieve the strain that has been placed on faculty, whose workload has increased with OL. The most important issue is saving lives and reducing the load on the health care system. Next, it is important to ensure students are getting a good education and that faculty have their needs met while trying to educate.

Community health impact

A critical issue to consider is the impact of the virus on the health of students, staff, faculty, and residents of Humboldt County and beyond. Loss of life or extreme illness impacts families, but it also affects health care workers, first responders, and other public servants. The ability of hospitals to handle the influx of patients, is a major concern. HSU is not located near any large cities, and residents may not be able to drive to other hospitals if Humboldt County's hospitals become overcrowded.

Students, staff and faculty

Students, staff, and faculty alike face challenges due to OL. Students may struggle to remain motivated or engaged in their studies (Aguilera-Hermida 2020, Meeter et al. 2020) or to maintain a distraction-free setting for OL (Nambiar 2020, Meeter et al 2020). Faculty may struggle with increased workload and with finding ways to engage students to learn (and learn collaboratively) (Chang and Fang 2020, Paulsen and McCormick 2020). At the same time, some staff may be without work due to campus closure. And students and faculty may face technical issues with OL that are beyond their control, such as internet connectivity or power outages; many universities lack the OL materials and resources required for effective online courses (Nambiar 2020). Educational gaps are likely widening among students as some students struggle more with autonomous, OL (Chang and Fang 2020). Creating a suitable environment at home can present challenges for some students (Meeter et al 2020).

Transmission of the coronavirus

Coronavirus transmission pathways are not completely understood and are an area of active research. The coronavirus is highly contagious and even brief contact with an infected person can transmit the virus. There is evidence that the coronavirus is spread by airborne particles that may travel across large rooms before settling (Morawska and Milton 2020). Because the virus can spread so easily, social distancing is considered the primary means of containing the virus. Different states have had varying level of success with social distancing; when individuals begin congregating, surges in infections have followed (Matrajt and Leung 2020). Social distancing delays, but does not prevent, infections.

Design and solution steps

To scientifically solve the problem of determining a safe HSU campus reopening day, five main tasks must be completed: 1) define the problem, 2) collect and analyze the data, 3) identify and implement the numerical method, 4) test and verify the model, and 5) present the results. When the risk of transmission of the virus is low enough, the campus may reopen. Using an SIR model, virus transmission is modeled, and a campus reopening date is recommended to Tom Jackson, Jr.

Problem statement

The problem is to determine a date when the risk of coronavirus transmission is low enough to justify reopening the HSU campus to the student body. A math model is developed to model historical data from January until November 2020 so that transmission of the virus can be forecasted into the future. If information about the virus transmission is known, policy decisions can be made at HSU based on the model. Observations will be made about trends in the historical data to guide the final recommendation. Furthermore, a Fortran program will be provided to Mr. Jackson so that future data may be used to update the model as needed.

Model (Susceptible, Infected, Removed)

The standard SIR model is a simple model. There are three subpopulations in the SIR model: 1) the susceptible (S) population, 2) the infected (I) population, and 3) the removed (R) population. Everyone starts off in the S-population; if they become infected, they move to the I-population but do not stay there long. The individuals in the I-population either recover, and are assumed to be immune from becoming reinfected, or they die from the illness. The I-population corresponds to the number of active cases in the total population. The R-population is everyone who has recovered or died from the virus.

<u>Model assumptions.</u> The SIR model works on several assumptions that make predicting future virus behavior with the SIR model complicated. First, the model assumes homogenous mixing among the populations (Cooper et al. 2020). Homogenous mixing has not occurred since the social distancing and travel restrictions took effect in March 2020. Another assumption is that there will be only one peak in infections; already there have been two peaks and active cases are currently peaking for the third time. The last assumption is that the S-population includes the entire population and the virus only goes away once everyone has become infected; this is particularly troublesome because if all Americans contracted the coronavirus, there would be widespread deaths, the hospitals would be completely overcrowded, and

infrastructure would crumble. Because of social distancing, many individuals of the S-population are kept from catching contracting the virus.

Modifying the model. To allow the model to fit historical data, the S-population is treated as a variable (Cooper et al 2020). Using S as a variable, allows the model to slow the rate of infections, and is justified because of the social distancing mandates that keep individuals separated; the standard SIR model assumes homogenous mixing among all populations (Cooper et al. 2020). Using model parameters from the literature with the SIR model and no adjustments to the S-population results in a 70 percent infection rate for a single day about 120 days into the model (See Figure 1 below). This type of scenario would probably have occurred if no social distancing, stay-at-home orders, or travel restrictions were put into place. California has not had 27 million individuals infected with the virus on a single day like the model suggest. When individuals are physically separated, there are less susceptible individuals coming into contact with infected individuals which has the same effect as manually reducing S in the model.

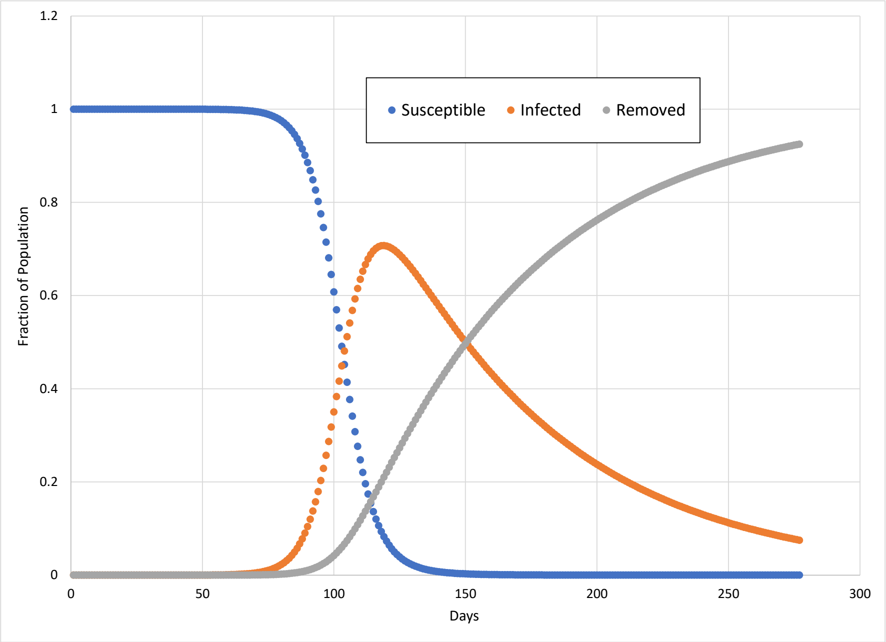


Figure 1. SIR-model run with parameters α = 0.178 and β = 0.015 over 277 days (Cooper et al) and no adjustments to S-population. Notice the R-population account for the majority of the total population by day 277 and around 120 days 0.7 of the population is infected at the same time but in reality, the R-population is 2.5 percent of the population at the time of this writing.

Model groups

There are three subpopulations of the SIR model. The S-population, the I-population, and the R-population. An overview of the three populations and model description is given below.

Susceptible (S). Initially, the S-population includes the entire California population. With the SIR model, all individuals are originally susceptible, and they remain in the S-population until they become infected. The three factors that affect the rate of change from the S-population to the I-population are: 1) model parameter α, 2) the number of I persons, and 3) the number of S persons. Cumulative cases (I + R) and cumulative removals (R) are modeled, and the value of the S is changed as historical data suggests. Reducing S reduces peak numbers and delays the peak (Hou et al. 2020).

Infected (I). The I-population is the number of active infections and is of interest in the analysis. The I-population changes on a daily basis as individuals are added from the S-population and as individuals leave to the R-population. Infected individuals are removed from the I-population by one of two methods, recovery or death. In the recorded data, results may be backlogged over the weekend or holidays; reported cases may include several days of testing making active infection data harder to model than the number of removals. A model assumption is that individuals who have recovered will not become reinfected and this is usually the case with coronavirus (Malkov 2020).

Removed (R). The R-population is cumulative number of recoveries and deaths as a result of the coronavirus. Because the number of recoveries and deaths is cumulative, it makes the data trend easier to model. Much has been said about heard immunity in the news since the beginning of the virus. A recovered population percentage of 67 percent has been estimated for immunity to the coronavirus (Frederiksen et al. 2020, Randolph and Barreiro 2020). The percentage of the recovered population should be higher still because recoveries and deaths are included in the R-population.

Vaccine modeling. There is news of several vaccines that may be approved by the Food and Drug Administration (CNN). If these vaccines are widely dispersed, experts feel that the risk of transmission of the virus will be diminished (Frederiksen et al. 2020). To model vaccinations using the SIR model, individuals are moved from the S-population to the R-population directly. It is assumed that the vaccine will only be given to individuals who have not previously been infected with the virus and that every individual vaccinated will develop immunity and never will be infected.

Defining a tolerable level of deaths and infections

Because most students, 89 percent of undergraduates, are from California and 25 percent of undergraduates are from northern California or local (Humboldt.edu), the data from California will be used. California Department of Public Health (CDPH) measures risk of community coronavirus transmission by two metrics: 1) rate per 100,000 (excluding prison cases,7-day average with 7-day lag), and 2) testing positivity rate. Based on the two metrics, risk of transmission is divided into four tiers. To reopen campus, the risk of community transmission should be in one of the lower two tiers as defined by the CDPH. Because looking at testing rate is outside of the scope of this project the maximum tolerable case rate per 100,000 is 3.9. Current national average is 49.6 cases per 100,000 (Barron 2020) and is at 6.1 cases per 100,000 persons in Humboldt County currently (Humboldt County Coronavirus Dashboard).

Herd immunity

The model should indicate reduced risk of transmission of the coronavirus before a safe reopening date can be determined. The large S-population makes the risk of transmission high until the S-population can be reduced. However, research shows that the herd immunity may be obtained when 67 percent of the population has an immunity (Frederiksen et al. 2020, Randolph and Barreiro 2020). Immunity may come from a previous infection or from a vaccination. It is determined that if herd immunity is reached, the risk of community transmission is diminished. When the number of individuals vaccinated, and the number of individuals recovered from the virus reaches 67 percent, herd immunity is assumed.

Constraints and scope of project

The scope of the project is defined as looking at historical data from the beginning of the pandemic until Nov 12, 2020 to determine an accurate model to estimate a campus reopening date for HSU. The budget for the project is $18,000.00 this is for 60 hours of engineer-billable hours at $300.00 per hour. The deliverables are a document outlining the project with a recommended date for campus reopening and a Fortran program for use by HSU to predict coronavirus transmission as new data becomes available.

Defining the system

Humboldt County is different from many other counties in California. Population per square mile in California is 239.1 persons but in Humboldt County there are 37.7 people per square mile (2010) (www.census.gov). Geographically HSU is isolated from the major population centers of the state. The student population of HSU is 6,983 (Humboldt.edu); total population of Arcata and Eureka is 45,141 individuals while the population of California is 39,512,223. Because a majority of HSU students live in central or southern California, examining Humboldt county data would not be sufficient. If the student body returned to campus from different areas of the state, or there was increased travel between Humboldt County and the different areas of the state, Humboldt County would experience increased exposure to individuals from other geographical areas. The data examined is for California and the state population is be used as the total population. Total days of data downloaded from the US Center for Disease Control (CDC) is 311 starting Jan 22, 2020 and running until Nov 27, 2020.

Data

To accurately model the coronavirus cases and deaths, the data used for modeling must reliable. There are many data sets available for coronavirus data, so it is easy to verify the numbers against multiple datasets (see Appendix II – Model and Data Verification). Because the virus appeared within the last year, much about the virus remains unknown. Coronavirus infections are not easily modeled, and the data is not expected to match the results from the SIR-model without modifying the model (Cooper et al. 2020).

Data needs. To compare the SIR-model to historical data, a daily sum of all infected and removed individuals is needed. The R-population includes individuals who have been infected by the coronavirus and died or recovered from the disease. However, the data for recoveries from coronavirus are not available from the CDC or the World Health Organization (WHO) for California. Humboldt County is reporting recovery data, so it is assumed that the ratio of recoveries to total cases is the same for California as it is for Humboldt County.

Data acquisition. Data is acquired from the CDC for 311 days starting January 22, 2020. Each day contains the cumulative total coronavirus cases, daily new case count, and cumulative total of deaths for every state. The cumulative deaths and case counts are verified with the data obtained from the World Health Organization (WHO), the California coronavirus dashboard, and John's Hopkins University. Data is downloaded in .csv directly from the CDC website.

Removed (R) population. The R-population includes all recoveries and deaths from the coronavirus. Data for coronavirus recoveries is available from the Humboldt County coronavirus dashboard but recovery data is not available for the USA or California. The ratio of Humboldt County recoveries to total cases is used to estimate the recoveries for California. Humboldt County began documenting recoveries as of April 20, 2020. The ratio of recoveries to total cases on April 20 is used to estimate the recoveries from

February 21 until April 20, 2020. There is likely some error in this assumption but over time the error should reduce to a negligible amount as cumulative case counts and total recoveries increase.

Vaccine data. The numbers of vaccines have been estimated although there are uncertainties associated with the estimates (Fox 2020). California is expected to get 375,000 doses of a coronavirus vaccine by the end of December 2020 (Ostrov and Ibarra 2020). Note: only a small percentage of the doses are heading to California (Fox 2020). The vaccines have not yet been approved, but vaccine suppliers have already produced doses of the virus and are ready to ship them once approved by the Federal Food and Drug Administration. The numbers of vaccines scheduled to be delivered in the USA are given (Fox 2020); to estimate the number of vaccines headed for California, the total number of vaccines are scaled according to California's population (see Table 2 below for vaccination numbers used in the model runs).

Model parameters

The two SIR-model parameters are: 1) the infection rate (α), and 2) the removal rate (β). The first model runs are completed with values found in the literature (Cooper et al. 2020). The model parameters may be changed as the data suggests. See Table 1 below for model parameter values. Because the data used is from of California, the population of California will be used for the total population (see Table 2 below for model run parameters). Because the model is calculating the fraction of the population that is in I or R, the total population is multiplied by the fraction from the model subpopulation resulting in the number of individuals. Because results are obtained in fractional value, it is possible to calculate the fraction of the HSU student body population in each subpopulation by multiplying the S, I, or R-fraction by the number of students.

Table 1. Parameter values for SIR-model. Values determined by trial and error and are in line with parameter values in the literature (Cooper et al. 2020, Toda 2020).

Symbol	Parameter	Value	Units
α	Infection Rate	0.23-0.35	n/a
β	Removal Rate	0.07	n/a

Table 2. Model run parameters for coronavirus modeling using a modified SIR model and Runga-Kutta Fehlberg (RKF) numerical method to solve the system of three ordinary differential equations (ODEs). Note: S, I, and R, are treated as fractions in the program (initial populations divided by the total population) and vaccinations are converted to fractions of the total population before being added to the R-population. Additional model run parameters listed in Table C2 in Appendix VI – Technical Methods and Application.

Symbol	Parameter	Value	Units
days	Number of days modeled	700	Days
poptotal	Total Population of California	39,512,223	Individuals
h	Step Size	1.0	Days
S(0)	Initial S-population	39,512,213	persons
I(0)	Initial I-population	10	persons
R(0)	Initial R-population	0	persons
Jan	Number of individuals vaccinated	375,000	persons
Feb	Number of individuals vaccinated	3,000,000	persons
Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec	Number of individuals vaccinated	1,800,000	persons

Numerical method

The numerical method is a way to solve a math model. The Runga-Kutta Fehlberg (RKF) numerical method is used to solve the equations of the SIR model. By using the numerical method, trends in coronavirus transmission can be examined and a reopening date recommended to Tom Jackson, Jr.

Fortran. Fortran is used to write the program solving the equations of the math model. Fortran is a programming language used for engineering applications. In Fortran, the interfacing of inputs and outputs is superior to Matlab. Fortran makes it easy to output data to .csv files for easy access in Excel and can accept input files as well. Fortran code can be compiled without any expensive software packages like Matlab. Fortran is also adaptable to many different types of models and situations.

Numerical method. The numerical method used to solve the equations of the SIR-model is the RKF method of integration (see Appendix II – Technical Methods and Application). By using the RKF method of integration in Fortran it is possible to evaluate the equations with a high degree of precision. The RKF intelligently slows itself down or speeds itself up depending the math being evaluated and does not overwork the computer with inefficiencies. There are infinite versions of the RKF but the one being used is more accurate that methods like Euler's that can be easily programmed in Excel.

The Code. The Fortran code for the RKF numerical method is already on file with Wegener Engineering from previous projects. This numerical method is a verified and accurate method for solving ordinary differential equation (ODEs). The SIR-model contains three ODEs and can be solved using the RKF. The only requirements for the RKF method to be completed is to code the equations and supply the RKF with model-run parameters. Model run parameters are found below in Table 2.

Using the program

The program is easy to use and gives instructions to the user at each step. See Appendix I – Using the Program for full instruction. The inputs and outputs of the program are discussed below.

Input. The program will accept user inputs for S-values and vaccine numbers at the end of December 2020 and the end of every month until September 2021. The number of days to simulated is 700 (model run parameters are listed in Table 2 above). The program prompts the user to name an output file which should be a .csv file to allow data manipulation in Excel. The program runs with the new values and outputs the results to the output file in comma-delimited form (see Appendix I – Using the Program).

Output. The program outputs the results in comma-delimited form to the output file of user choice. Output lists the day, the susceptible, infected, and removed populations across the top and then the values listed in rows with each day on its own row. The program will list the number of individuals in each population of the model (S, I, and R) for each day (see Appendix I – Using the Program).

Verifying the model

To assess the accuracy of the model, the SIR model results from an example problem (Callahan et al. 2008) are compared to values obtained in Scilab (See Appendix II – Program Verification). The model behaves like an SIR model where there is one peak of infections, removals start at zero and increase until the entire population is removed, and the S-population starts near one and decreases towards zero. Comparison of the model outputs to the historical data is done by visual inspection of R and I.

Model Fit. The S-population is generally kept low so that case numbers follow the data (see Table D1 in Appendix VIII – Additional Figures and Tables for all parameter updates). The numbers of individuals in each subpopulation can be compared to the data to find the percent relative error. The base case scenario with no S-value updates or changes to model parameters α or β is shown in Figure 1 above. Note: the standard model needs modifications to match the coronavirus data (see Figure 2 below).

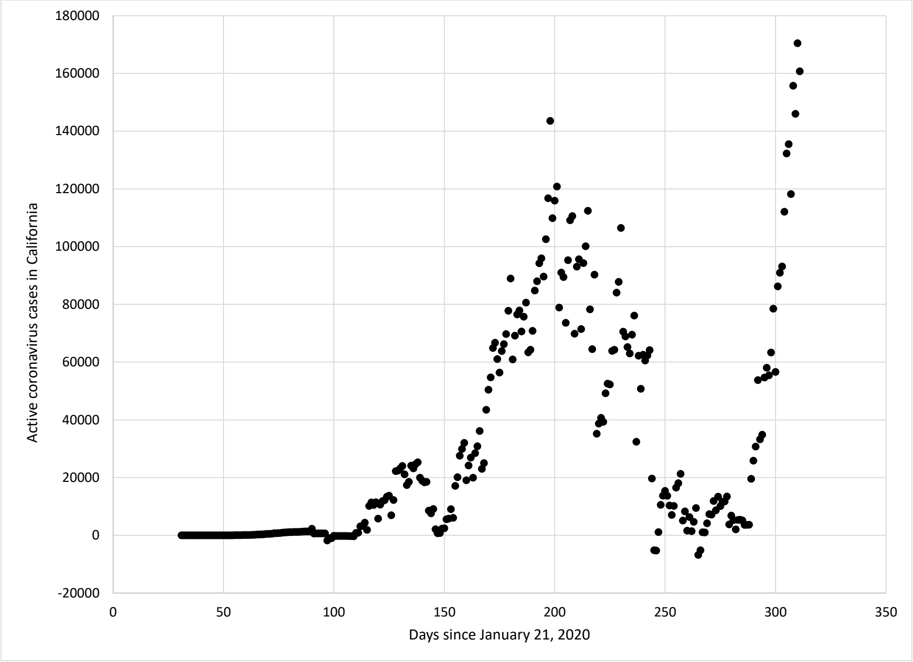


Figure 2. Active coronavirus cases in California estimated by removing the R-population and total deaths from the total case count (deaths and total case data from CDC). Some negative values are shown due to discrepancies in the R-population stemming from the assumption that recoveries in California matched the published recoveries in Humboldt County.

Sensitivity Analysis. A sensitivity analysis is performed on model parameters for all days after the recorded data ends on November 27, 2020. The sensitivity analysis is performed by evaluating the model with a 20 percent increase and decrease of parameters α and β. Sensitivity of the model to a 40 percent increase and decrease of S at every adjustment after November 27, 2020 is also examined. S-value updates begin at the end of January 2021 because the current peak in cases is expected to last through the beginning of the new year (Emmons 2020). The greatest uncertainty affecting the final recommendation is the number of individuals that may be vaccinated in the coming months. The number of vaccines administered is varied between 40 percent and 120 percent of the estimates.

Results and discussion

The math model is developed and used to make predictions about future virus transmission and recommend a campus reopening date to Tom Jackson, Jr. The model requires inputs from the user; the program is discussed further in Appendix I – Using the Program. Modeling the data shows that the virus is not going away soon without a vaccine. Luckily, there are two vaccines that may soon be available.

Section content. In this section, the model fit is discussed including results from the model runs and the modeling of vaccinations. Finally, results from the sensitivity analysis are presented. In the next section, the conclusion and final recommendation is given.

Model

Model fit. As expected, the coronavirus data is not modeled accurately by the standard SIR model. The SIR model results in one peak in infections and already there have been two peaks in California and a third peak is happening at the time of this writing (see Figure 3 below). The SIR model is modified and used to determine when herd immunity may be reached.

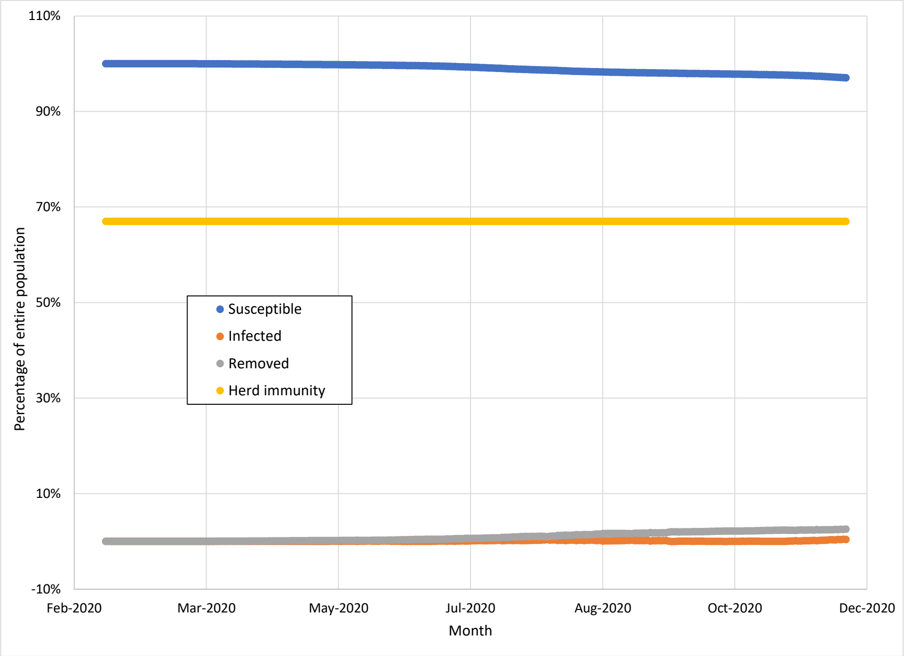


Figure 3. S, I, and R populations from February 22 to November 27, 2020 according to historical data (CDC). About 2.5 percent of the total population is in the R-population by November 27, meaning that they have recovered from the coronavirus or died; the rest of the population remains in S or I.

Model modifications. The SIR model is modified to allow S to be a variable that changes over time. These adjustments simulate the effects that social distancing and quarantines have had on virus transmission. Because of varying levels of success with maintaining social distancing across California, it is reasonable that S varies as individuals have more or less contact with others. Individuals who have tested positive for the coronavirus are told to stay at home for a period of time to allow the virus to run its course and not transmit the virus to others. The effect of wearing masks is another factor that would keep individuals in the S-population from contracting the virus from individuals in the I-population.

Delaying a peak. As discussed in the Design and Solution Steps section, homogenous mixing is not a valid assumption with the coronavirus and is a justification for modifying the SIR model. If the entire population is homogenously mixed and using model parameters from the literature, 70 percent of Californians would be infected around day 120 of the model (see Figure 1 above); a peak of this magnitude is not reflected in the data. With the current numbers of cases pushing hospitals to overcrowding, an additional influx of patients would be detrimental for the health care system. By keeping the S-population low in the model, the large peak in infections is delayed, but the potential for that peak is not diminished.

<u>Model parameters.</u> Model parameters α and β are determined through comparison of model outputs to historical data until November 27, 2020 and are in the range suggested by the literature (Cooper et al. 2020, Toda 2020). The model fits the data with updates to S, see Figure 4 below. If the stay-at-home orders continue in California long enough to get 67 percent of the population vaccinated, more peaks may be avoided. The first obstacle is flattening the curve of the current peak; flattening the curve may not be easy due to the approaching holidays and the social gatherings that traditionally go along with them.

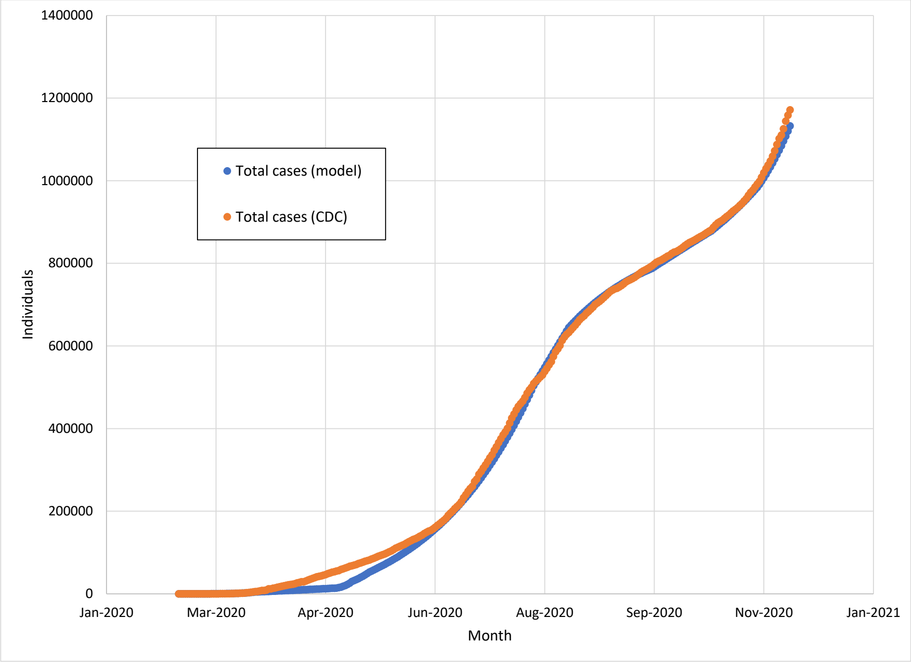


Figure 4. Total coronavirus cases in California according to the data and according to the model. Comparison of total coronavirus cases to historical data is considered the most accurate way to judge model fit because the total case count is coming directly from the CDC with no estimations such as when calculating the R-population.

Results

The model is verified and used to model coronavirus cases with updates to S. However, with forecasting coronavirus transmission, S is not known and assumed to be kept low. Through the enforcement of stay-at-home orders, quarantines, and wearing of face-coverings and personal protective equipment, infections are reduced and so S is reset at 0.2 every month of modeling in 2021 except for January 2021 because the current peak is likely to last into the new year (Emmons 2020).

Removed (R) population. Because data for the number of recoveries for California is not available, the recovery data from Humboldt County is used to estimate the recoveries for the state. Adding the estimated total recoveries to the total death count results in the R-population. As shown in Figure 2 above, the active cases are not accurate because the calculations result in negative values for several days. The overall trend follows the total daily case count, however, and the data is deemed useful.

Recovered population. The recovered population is used to monitor progress towards herd immunity. Because the deaths from the virus do not count towards herd immunity the deaths must be subtracted from the R-population. For modeling past November 27, 2020, the average recovery rate from the recorded data, 0.8883, is used to calculate the recoveries for coming months with model outputs.

Model verification. The modified SIR model is able to model the data well with adjustments to the S-population. In Figure 4 above, the total cases of the model are compared to the total cases from February 22 until November 27, 2020 and in Figure D3 in Appendix VIII – Additional Figures and Tables the I-population and the R-populations are compared to the data. The total estimated removals are compared to model removals along with the number of active infections from the model and from the data in Figure D2 in Appendix VIII – Additional Figures and Tables. Note: the removals are believed to have some error because of the calculations used to estimate California recoveries. The active case count is using the estimated removals and therefore the cumulative case count comparison, Figure 4 above, is likely a better indication if the model fits the data.

Model Fit. The SIR model fits the data with adjustments to the S-population as the data suggests. In Table D1 in Appendix VIII – Additional Tables and Figures, the S-values adjustments are given and in Figure D1, the S-values are shown throughout the modeling period of 700 days. In Table A2 in Appendix II – Program and Data Verification, the percent relative error for the cumulative case counts and removed populations on November 27, 2020 is given. See Figure 4 above for comparison of cumulative cases in California and total case count from the model.

Change in infection rate (α). Because of the rise in cumulative cases since September 2020, the infection rate, α, is increased to 0.3 after being constant for the first 218 days of the model (see Table D1 in Appendix VIII – Additional Figures and Tables and Figure 4 above). The reason for the change in the parameter value is due to trends in the data and may be caused by the changing of the seasons. As weather discourages individuals from spending as much time outdoors, there may be more opportunities for transmission between individuals.

Vaccine rollout. In Figure 5 below, the jumps in the R-population are modeling the administration of 375,000 vaccines at the end of December 2020, three million at the end of January 2021, and 1.8 million at the end of every month until the end of September 2021. If the vaccines are approved and at least 1.8 million vaccines are given each month, the herd immunity threshold is reached by December 10, 2021.

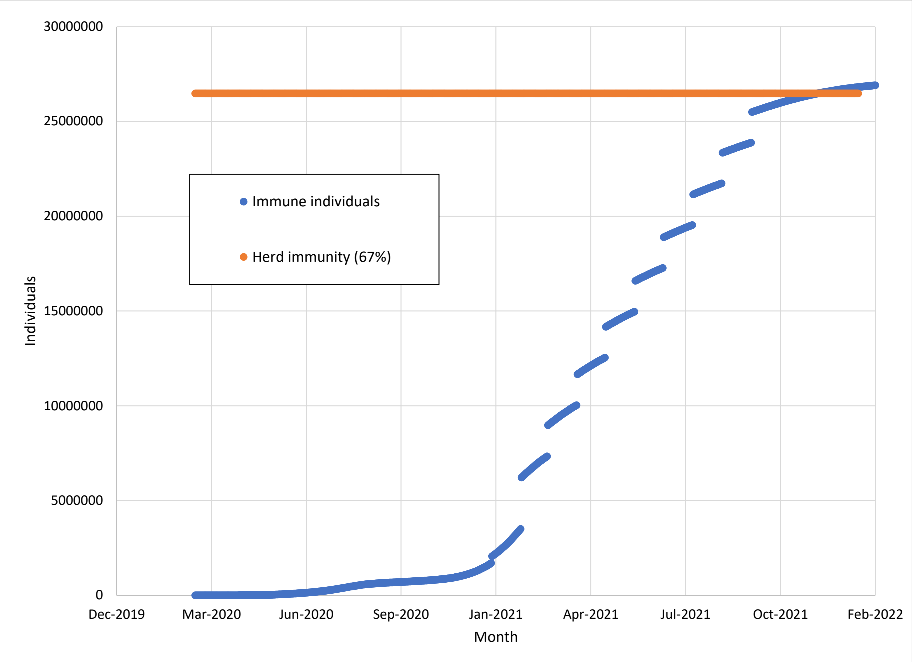


Figure 5. Recovered population is shown as recorded in the data from February 22, 2020 until November 27,2020. After November 27, the model is using S = 0.2 and adding vaccinated individuals to the recovered population. Herd immunity, 67 percent of the total population with immunity, is reached on December 10, 2021 after vaccination of 21,375,000 California residents.

Sensitivity analysis Sensitivity of the model is examined by varying model parameters α, β, S, and the vaccinations estimate. α and β are increased and decreased by 20 percent at five percent increments. The S-population is increased and decreased by 40 percent at increments of 10 percent. The resulting change in time before a campus reopening date shown below in Figure 6 and full results are given in Table B1 of Appendix III – Sensitivity Analysis. α and β are both increased and decreased by 20 percent but have inversed effects; a 20 percent decrease in α results in a 20 percent increase in campus reopening time and a 20 percent increase in α results in a 20 percent decrease in the time until the campus reopening date. A 20 percent decrease in β results in a 20 percent decrease in time and 20 percent increase, increases the time by 20 percent. Reducing vaccine numbers by 60 percent results in 50 percent more days to campus reopening and a 20 percent increase in vaccine estimates decreases the days by about 8 percent. Changing

S has a similar but weaker effect on time as changing α. See Table 2 above for the vaccination numbers and Appendix III – Sensitivity Analysis, for complete sensitivity analysis results.

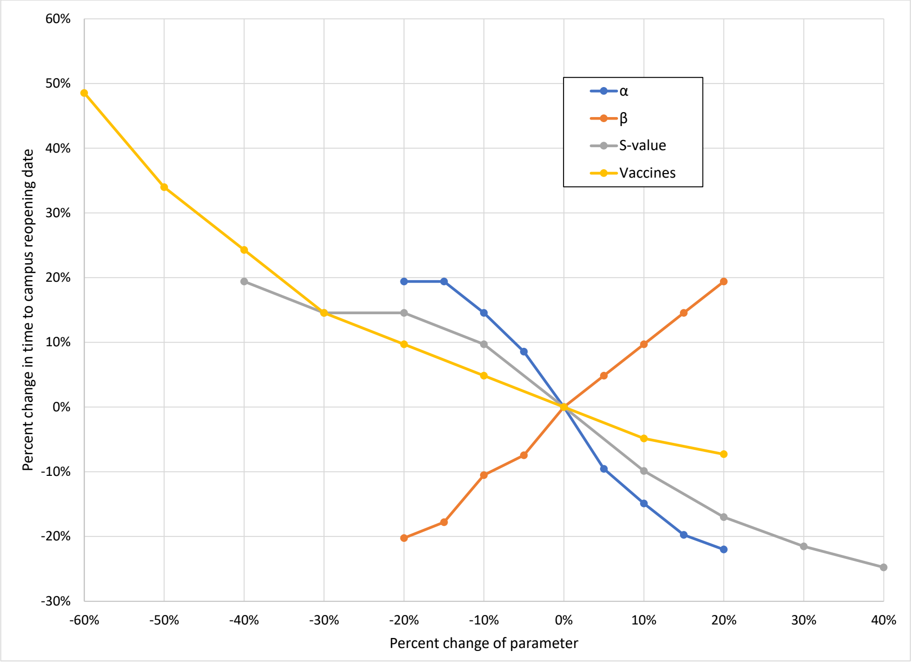


Figure 6. Sensitivity analysis on α and β, S, and the vaccine estimate. α and β are both increased and decreased by 20 percent having opposite effects; a 20 percent decrease in α results in a 20 percent increase in campus reopening time and a 20 percent increase in α results in a 20 percent decrease in the time until the campus reopening date. A 20 percent decrease in β results in a 20 percent decrease in time and 20 percent increase, increases the time by 20 percent. Reducing vaccine numbers by 60 percent results in about 50 percent longer time to campus reopening and a 20 percent increase in vaccine estimates lowers the time by about 8 percent. Changing S has a similar but weaker effect on time as changing α.

Considerations

Data. California cumulative positive cases and total deaths from five sources is compared to determine the percent error of each value. The data from CDC is assumed to be most accurate. The largest percent error is about five percent for cumulative case count and about two percent for total deaths (see Table A2 in Appendix II – Program and Data Verification).

Sensitivity analysis. Sensitivity of the model to its parameters is examined in the sensitivity analysis (see Figure 6 above and Table B1 in Appendix III – Sensitivity Analysis for full results). The model is moderately sensitive to model parameters α, β, and S. The model is sensitive to the number of vaccines particularly in the next few months. The final recommendation may be reduced by as much as 153 days or increase by as much as 120 days by changing one of the parameters. The recommendation may be reduced as much as 45 day and increased as much as 300 days with a change in the number of vaccines.

Vaccines. Vaccines may not be approved or may not reach Californians in the numbers estimated in this document. The program will accept different vaccination number for each month (see Appendix I – Using the Program). With the ability to model the recovered population with updated values, progress towards herd immunity may be monitored in the coming months. It is assumed that the vaccine is completely effective, which is not the case. Because a sensitivity analysis is performed on vaccine numbers, and the ability to model any number of vaccinations in the program, this discrepancy is overlooked.

Conclusion

Containment of the virus is contingent on vaccinations of Californians in large numbers. A recommended reopening date is recommended as December 10, 2021. The SIR model gives insights into virus behavior and potential for virus transmission. By updating the model in Fortran, the results can be refined as more historical data becomes available.

Uncertainties and sensitivity of the model

As discussed previously, the sensitivity of the model to parameters α and β and the S-population can change the recommended date for HSU campus reopening by as much as 300 days (see Appendix III – Sensitivity Analysis). If neither vaccine is approved or if the vaccine or vaccines approved do not get distributed at the estimated rate, the number of individuals moving from the S-population to the R-population without becoming infected will be reduced. If the R-population does not reach about 67 percent of the population, then there will be continued risk of community transmission. The number of vaccines scheduled for California by the end of December is a small percentage of the 20 million vaccines expected to be delivered across the nation. If California continues to get small percentages of the nation's vaccine supply, the recommendation will be influenced.

Current peak

It may seem like good news that herd immunity may be reached by December 2021, but in the model, the active coronavirus case count will be nearly five times greater than it was on November 27, 2020 at its next peak. The location of the peak according to the model is the end of January and beginning of February 2021. Considering the health care and essential workers, it is essential that the stay-at-home orders are met with compliance. If, however, the stay-at-home orders and other restrictions are successful, there is a chance that the R-population is not high enough for herd immunity by December 10, 2021.

Final recommendation

The final recommendation by Wegener Engineering to Tom Jackson, Jr., is that HSU can safely reopen its campus December 10, 2020. This recommendation is counting on the vaccination of 21,375,000 Californians before the end of November 2021 to bring the recovered population high enough that herd immunity can be assumed. There is nothing in the data or model outputs showing that the virus goes away without a major peak that would result in widespread cases and deaths or the vaccination of a large percentage of Californians. If at least 67 percent of Californians have recovered from the coronavirus or been vaccinated, then it is assumed that there will be herd immunity among the population and the virus will cease to be transmitted.

Final considerations

Considering the sensitivity analysis, if a vaccine is not delivered as scheduled or doses are not available as estimated for 2021, then the campus reopening date would need to be adjusted to a time coinciding with a 67 percent recovered population. It is recommended that Mr. Jackson continue to update the model by entering the number of individuals vaccinated each month into the program which will allow changes in the R-population to be analyzed with greater accuracy. The program will prompt the user for the S-value (set as s=0.2 from February until December 2021) and the number of vaccines administered at the end of each month to monitor progress towards herd immunity (see Appendix I – Using the Program).

References

Aguilera-Hermida, A. P. 2020. “College Students’ use and acceptance of emergency online learning due to COVID-19.” *International Journal of Educational Research Open.*

Barron, J. 2020. “Coronavirus Update.” *Wall Street Journal,* November 28, 2020.

Callahan, J., Hoffman, K., Cox, D., O’Shea, D., Pollatsek, H., Senechal, L. 2008. *Calculus in Context.* Five Colleges Inc

Census.gov. Humboldt County and California populations.

Chapra, S. C., and Canale, R. P. 2015. *Numerical Methods for Engineers,* 7th Ed., McGraw-Hill Education, New York, NY.

Cooper, I., Mondal, A., Antonopoulos, C. G. 2020. “A SIR model assumption for the spread of COVID-19 in different communities.” *Chaos, Solitons, and Fractals*, 139, 1-14.

Chang, C., Fang, M. 2020. “E-learning and Online Instructions of Higher Education during the 2019 Novel Coronavirus Diseases (COVID-19) Epidemic.” *Journal of Physics: Conference Series,* 1574, 1-5.

Emmons, N. 2020. “COVID Cases Surge in Worcester, Peak Expected in January.” *CBS: Boston.* December 10, 2020.

Fox, M. 2020. “When Can I get a Coronavirus Vaccine.” *CNN.* December 1, 2020.

Frederiksen, L. S. F., Zhang, Y., Foged, C., Thakur, A. 2020. “The Long Road Toward COVID-19 Herd Immunity: Vaccine Platform Technologies and Mass Immunization Strategies.” *Frontiers in Immunology,* 11, 1817.

Hethcote, H. W. (2000). “The Mathematics of Infection Diseases.” *SIAM Review,* 42(4), 599-653.

Hou, C., Chen, J., Zhou, Y., Hua, L., Yuan, J., He, S., Guo, Y., Zhang, S., Jia, Q., Zhao, C., Zhang, J., Xu, G., Jia, E. 2020. “The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis.” *Journal of Medical Virology,* 92, 841-848.

Malkov, E. 2020. “Simulation for coronavirus disease 2019 (COVID-19) scenarios with possibility of reinfection.” *Chaos Solitons Fractals,* 139.

Matrajt, L., Leung, T. 2020. “Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Infections.” *Emerging Infectious Diseases,* 26 (8), 1740-1748.

Meeter, M., Bele, T., Den Hartogh, C. F., Bakker, T., De Vries, R. E., Plak, S. 2020. ”College students‘ motivation and study results after COVID-19 stay-at-home orders.”

Morawska, L., Milton, D. K. 2020. “It is Time to Address Airborne Transmission of COVID-19.” *Clinical Infectious Diseases.*

Nambiar, D. 2020. “The impact of online learning during COVID-19: students’ and teachers’ perspective.” *International Journal of Indian Psychology,* 8 (2), 783-793.

Ostrov, B. F., Ibarra, A. B. 2020. “Vaccines getting closer: Governor outlines priorities for who’s up first.” *Cal Matters*. December 4, 2020.

Paulsen, J., McCormick, A. 2020. “Reassessing Disparities in Online Learner Student Engagement in Higher Education.” *Educational Researcher,* 49 (1), 20-29.

Randolph, H. E., Barreiro, L. B. 2020. “Herd Immunity: Understanding COVID-19.” *Immunity,* 52, 737-741.

“United States COVID-19 Cases and Deaths Data by State over Time.” CDC. Accessed November 27, 2020. www.cdc.org.

Appendices

There are eight appendices in total: 1) Using the program: how to use the program, and description of the inputs and outputs, 2) Program verification: where the accuracy of the model is discussed, 3) the design brief, 4) the technical memo, 5) the technical methods and application, 6) annotated bibliography where sources and corresponding descriptions can be found, 7) additional figures and tables, and 8) the Fortran code: including the RKF subroutine and SIR-model ODEs in an external function.

Appendix I – Using the Program

To use the program to model coronavirus cases moving forward, the user must launch the program and supply inputs to the program and then analyze the results. The program file, covid.f90, may be opened in the console in Linux or in SimplyFortran. The program will prompt the user to name an output file where the results will be stored (this file should have -.csv suffix to indicate the file should be opened with Excel). Next the program will prompt the user to input the number of vaccinated persons and the S-value at the end of December 2020 and every month until September 202. The program is set up to model 700 days including the 280 days from February 22 to November 27, 2020 where historical data is currently available.

To determine the S-value, the user should compare model outputs to historical data by trial and error. If the user wants to model the S-population of a homogenously mixed population, they should subtract the total number of positive cases from the California population. An S-value of 0.2 is used in the model runs to determine the campus reopening date due to lack of knowledge about future virus and sociological behavior.

Once the output file is compiled, the user can graph the data in Excel. To calculate the recovered population, the population that should reach 67 percent of the total population for herd immunity, the user subtracts total deaths from the R-population and adds any individuals who have been vaccinated. If the model predicts higher infections than is shown in the data, the S-value can be decreased to better follow the data and if the model is below the actual number of infections, S can be increased.

Appendix II – Program and Data Verification

To determine the accuracy of the model, the problem from Ch. 1 of *Calculus in Context* by (Callahan et al. 2020) is used and results compared to model results. Table A1 displays the results from textbook problem solved with the SIR model in Fortran and also solved in Scilab and presenting a percent relative error. The largest percent relative error found over the first three days of modeling is found to be 1.0 x 10-4 percent.

Table A1. Comparing Fortran SIR model outputs to results obtained in Scilab for the example problem from Callahan et al. 2008. The largest percent relative error in results is 1.0 x 10^{-4} percent for the removed population.

	Fortran SIR model output (persons)			Scilab verification			Error		
Days	**Susceptible**	**Infected**	**Removed**	**Susceptible**	**Infected**	**Removed**	**% Relative Error**		
0	45400	2100	2500	45400	2100	2500	0	0	0
1	44255.455	3062.163	2682.382	44255.459	3062.160	2682.381	-8.13E-06	8.82E-05	1.86E-05
2	42648.883	4404.610	2946.507	42648.892	4404.602	2946.506	-2.11E-05	1.75E-04	5.09E-05
3	40460.092	6217.079	3322.829	40460.111	6217.064	3322.825	-4.67E-05	2.49E-04	1.02E-04

Table A2. Data comparison for cumulative coronavirus cases and deaths for California from five different sources (CDC, John's Hopkins University, California Department of Public Health, Worldometers.com, and the New York Times). Assuming the CDC data is the most accurate, Worldometers and the New York Times has the largest percent error of about five percent for cumulative cases and around two percent for total deaths.

11-Dec-20

Data Source	**Total cases**	**Total deaths**	**% error (total cases)**	**% error (total deaths)**
CDC	1450235	20463	0.000%	0.000%
Johns Hopkins	1482551	20603	2.228%	0.684%
California Covid Dashboard	1485703	20622	2.446%	0.777%
Worldometers.com	1525790	20851	5.210%	1.896%
Google (New York Times)	1530000	20829	5.500%	1.789%

Table A2. Comparison of cumulative cases and R-population from the model to data from the CDC. A three percent relative error is found for cumulative case counts while a 0.1 percent error is found for the R-population

11/27/2020

	Cumulative cases	**R-population**	**% relative error (cum. Cases)**	**% relative error (R-population)**
model	1132704	1011953	3.297%	0.133%
CDC data	1171324	1010607	0.000%	0.000%

Appendix III – Sensitivity Analysis

A sensitivity analysis is performed on model parameters for all days after the recorded data ends on November 27, 2020. The sensitivity analysis is performed by evaluating the model with a 20 percent increase and decrease of parameters α and β. Sensitivity of the model to a 40 percent increase and decrease of S at every adjustment after November 27, 2020 is also examined. S-value updates begin at the end of January 2021 because the current peak in cases is expected to last through the beginning of the new year (Emmons 2020). The greatest uncertainty affecting the final recommendation is the number of individuals that may be vaccinated in the coming months. The number of vaccines administered is varied 40 percent and 120 percent of the estimates. The model is sensitive to model parameters α and β, S, and the number of infections. Decreasing α and S both increase the amount of time to reach heard immunity proportionately and increasing them has a similar but opposite effect (see Table B1 on the next page).

Table B1. Sensitivity analysis results from varying model parameters α and β, S, and number of vaccines administered. Largest percent change in time to reach herd immunity is when decreasing the number of vaccines by 60 percent. Reducing β by 20 percent reduces the time to reach herd immunity by 20 percent. Increasing α by 20 percent reduces the time by 22 percent and increasing S by 40 percent decreases the time by 25 percent.

Parameter changed	% change parameter	multiplication factor	parameter value	days until herd immunity	% change time
α	-20%	0.8	0.28	738	19.42%
	-15%	0.85	0.2975	738	19.42%
	-10%	0.9	0.315	708	14.56%
	-5%	0.95	0.3325	671	8.58%
	0%	1	0.35	618	0.00%
	5%	1.05	0.3675	559	-9.55%
	10%	1.1	0.385	526	-14.89%
	15%	1.15	0.4025	496	-19.74%
	20%	1.2	0.42	482	-22.01%
β	-20%	0.8	0.056	493	-20.23%
	-15%	0.85	0.0595	508	-17.80%
	-10%	0.9	0.063	553	-10.52%
	-5%	0.95	0.0665	572	-7.44%
	0%	1	0.07	618	0.00%
	5%	1.05	0.0735	648	4.85%
	10%	1.1	0.077	678	9.71%
	15%	1.15	0.0805	708	14.56%
	20%	1.2	0.084	738	19.42%
S-value	-40%	0.6	0.12	738	19.42%
	-30%	0.7	0.14	708	14.56%
	-20%	0.8	0.16	708	14.56%
	-10%	0.9	0.18	678	9.71%
	0%	1	0.2	618	0.00%
	10%	1.1	0.22	557	-9.87%
	20%	1.2	0.24	513	-16.99%
	30%	1.3	0.26	485	-21.52%
	40%	1.4	0.28	465	-24.76%
Vaccines	-80%	0.2		1000+	
	-70%	0.3		1000+	
	-60%	0.4		918	48.54%
	-50%	0.5		828	33.98%
	-40%	0.6		768	24.27%
	-30%	0.7		708	14.56%
	-20%	0.8		678	9.71%
	-10%	0.9		648	4.85%
	0%	1		618	0.00%
	10%	1.1		588	-4.85%
	20%	1.2		573	-7.28%

Appendix IV - Design Brief

Problem statement

Due to the coronavirus appearing in the United States in early 2020, this global pandemic has forced Humboldt State University to close its campus to the majority of campus activities including classes, sporting events, seminars, and many student services. Almost all classes are being held online; students and faculty alike are wondering when it will be safe to go back to campus. Tom Jackson, Jr., on behalf of HSU, has hired Wegener Engineering to examine the data surrounding the pandemic and to develop tools to determine when it will be safe to return to campus as a student body.

Design statement

The design will be a program that utilizes numerical methods to analyze the data and give administrators of HSU the ability to make decisions based on scientific and mathematical modeling. The safety and risks associated with reopening campus for classes and activities will be analyzed in light of current data from the virus; as the pandemic progresses, the program may be updated to reflect current trends. A report outlining recommendations will be given for a date of reopening along with best and worst-case scenarios. Global data will be analyzed and compared to local data and some assumptions will be made as to the acceptable rates of infections among the student body

Constraints

Global data sets will be analyzed from December 2019 until the end of September 2020. The report containing recommendation and instruction for model use will be issued by December 11, 2020. Assumptions will be made about the community based on information from Humboldt County and other areas of the USA. models will be tested against local and global data. The program will be easy for Mr. Jackson or members of his staff to use and results will be easy to interpret. Program will be accurate and versatile.

Science and engineering principles used

Scientific data modeling, numerical methods (1st Order Euler's Integration), solving ODE with numerical methods, and SIR-model for epidemics will all be approached throughout the project. The base SIR-model is shown in equations 1-3 (Chapra and Canale 2015):

$$\frac{dS(t)}{dt} = -aS(t)I(t) \tag{1}$$

$$\frac{dI(t)}{dt} = aS(t)I(t) - bI(t) \tag{2}$$

$$\frac{dR_m(t)}{dt} = bI(t). \tag{3}$$

where

S – susceptible individuals

I – infected individuals

R_m – removed individuals (immune or dead)

a – a real positive parameter of initial exponential growth

b – a real positive parameter of final exponential decay

t - time

The SIR model to be used for this project varies slightly in that the *S* is treated as a variable that changes in time (Cooper et al 2020).

Discussion

This problem is important to solve because Tom Jackson, Jr. needs accurate predictions about the effect of the pandemic as we move into 2021. Students and faculty alike are struggling to complete coursework and collaborate on projects which have scientific importance. The finances of the university are at stake; if more students stop taking classes the university will lose tuition funding and must reduce spending. If the student body returns too soon there is a risk of loss-of-life and overburdening of the health system here in Humboldt County.

Deliverables

A Fortran program will be delivered on December 11th, 2020 along with a report which outlines the findings and gives a recommendation for the date of campus reopening. The report will outline the program and how to use it to model new data as it becomes available.

Appendix V – Technical Memo

to: Tom Jackson, Jr.
from: Caleb Wegener (Wegener Engineering)
subject: Covid-19 Model for HSU campus reopening date
date: September 27, 2020
CC: Dr. Beth Eschenbach, Dr. David Narum

The team here at Wegener Engineering are working to develop a computer program to be used by HSU to make predictions about the COVID-19 pandemic. There are widely available graphics from the historical data of the virus, but HSU needs to be able to predict what the graphics and data will be in the future. The program developed by Wegener Engineering will give graphical and numerical predictions about the virus; the model will aid in making decisions about reopening campus to the student body and resuming classes.

The purpose of this memo is to discuss three topics: 1) the methods used to formulate the model for COVID-19 infections in the student body (and for Humboldt County), 2) the data that will be used to verify and update the model moving forward, and 3) anticipated results from the model and subsequent predictions of infection and death rates from COVID-19. The Susceptible, Infected, Removed (SIR) Model is the classic epidemic model (Hethcote 2000) and there is evidence which suggests a modified SIR-model is helpful in modeling COVID-19 (Cooper et al. 2020).

Methods (purpose and value): Three basic methods will be used to model COVID-19 data among the HSU student body and the population of Humboldt County: 1) develop a modified SIR-model which allows for updating the model as surges in COVID-19 cases occur, 2) compare results from the SIR-model to data in counties in California to validate and update the model parameters, and 3) use the verified model parameters in the modified SIR-model to make a prediction of a safe reopening date for HSU Campus. In the traditional SIR-model, the S-population decreases steadily towards zero, which is not valid for the COVID-19 pandemic. Therefore, some trial-and-error analysis must be performed on the modified SIR-model to determine the parameter values that best match the data for Humboldt County, so that the model can be used to make predictions about when the virus will begin to die out among the local population.

Data collection: The data used to verify the model is from the United States Center for Disease Control. The data is a daily log of new confirmed cases of COVID-19 and deaths resulting from the virus from January 2020 until the end of September 2020. Obtaining records of the HSU student body home county demographics may be useful in determining if Humboldt County data accurately reflects the student body population. We recognize obtaining this data may infringe on the privacy of the students or the university. Many students live outside of Humboldt County and may influence the parameter values used if the student body returns to Humboldt County for a campus reopening.

Anticipated results: Expected results are that the model parameters will closely correlate with the model parameters found in other studies (Cooper et al) for the United States. Results will likely point to a time in early 2021 when the decline in infection rates and deaths is large enough to justify reopening HSU's campus to the student body. One or more surge periods are likely to occur before December 11, and for that reason the model is designed to be updated if a surge occurs. Reopening campus will likely cause a surge to occur; the risks associated with reopening campus will be explored by considering different surge scenarios at different times (reopening dates).

I recommend providing Wegener Engineering with data of the student body's county residential demographics if possible. Demographics will allow model parameters to be estimated with higher accuracy.

Appendix VI – Technical Methods and Application

The methods outlined in this section are used to solve the system of ODEs which make up the SIR-model being used to model coronavirus infections and deaths for California to make predictions about the behavior of the virus moving forward. Accurately modeling the historical data provides assurance that the model is producing reliable results that can be used to determine a campus reopening date for HSU based on the mathematical model. Criteria for reopening the HSU campus is dependent on the definition of a tolerable level of disease transmission from California Dept of Public Health stating that if 3.9 cases per 100,000 is maintained, the risk of community transmission is “moderate” or “minimal”.

Approach to Solving the Problem.

SIR-model. The base SIR-model is given below. Equation 1 is the susceptible population and Equation 2 describes the movement of the population from susceptible population to the infected population and then on to the removed population. α is considered the infection rate and β is the removal rate (because the virus is deadly β cannot be considered a recovery rate). See Table C1 for model parameter values.

$$\frac{dS(t)}{dt} = -\alpha S(t)I(t) \tag{1}$$

$$\frac{dI(t)}{dt} = \alpha S(t)I(t) - \beta I(t) \tag{2}$$

$$\frac{dR_m(t)}{dt} = \beta I(t) \tag{3}$$

where

S – susceptible individuals
I – infected individuals
α – a real positive parameter of initial exponential growth
β – a real positive parameter of final exponential decay
t - time

Modified SIR-model. Modifying the standard SIR-model by keeping S as a variable where it can be updated to account for surge periods. Ssurge is set at 0.2 updated every 4 days during surge periods in the USA (Cooper et al. 2020).

For solving the ODEs of the SIR-model a numerical method in Fortran is used. Fortran is a programming language used by engineers and it has many strengths when computing mathematical systems. The RKF method of integration results in a high level of accuracy with minimal computing power.

Table C1. Parameter values for SIR-model use by Cooper for USA coronavirus cases.

Symbol	Parameter	Value	Units
α	Infection Rate (Cooper et al. 2020)	0.23-0.35	n/a
β	Removal Rate (Cooper et al. 2020)	0.07	n/a

Numerical method.

The numerical method used to evaluate the ODEs of the SIR-model for coronavirus infections in California is the Runga-Kutta Fehlberg (RKF) method of integration. This method is both accurate and efficient because it uses an estimate based on the 4th and 5th order estimates along with adaptive step sizes resulting in a global error of 4th order (Chapra and Canal 2015).

RKF estimates. The 4th and 5th order estimates are calculated using the Cash-Karp coefficients, which are the fractions in front of each slope estimate, and the six functional evaluations which are slope estimates for the functions. The 4th order RKF estimate is given in Equation 4 and the 5th order RKF estimate is given in Equation 5. y_i is the present value of the dependent variable and in the case of the first iteration is the initial condition required to solve the set of ODEs. The k_i values, Equations 6-11, are the slope estimates (functional evaluations) on which the method is based (Chapra and Canale 2015).

$$y_{i+1} = y_i + \left(\frac{37}{378}k_1 + \frac{250}{621}k_3 + \frac{125}{594}k_4 + \frac{512}{1771}k_4\right)h \tag{4}$$

$$y_{i+1} = y_i + \left(\frac{2825}{27648}k_1 + \frac{18575}{48384}k_3 + \frac{13525}{55296}k_4 + \frac{277}{14336}k_5 + \frac{1}{4}k_6\right)h \tag{5}$$

$$k_1 = f(x_i, y_i) \tag{6}$$

$$k_2 = f(x_i + \frac{1}{5}h, y_i + \frac{1}{5}k_1h) \tag{7}$$

$$k_3 = f(x_i + \frac{3}{10}h, y_i + \frac{3}{40}k_1h + \frac{9}{40}k_2h) \tag{8}$$

$$k_4 = f(x_i + \frac{3}{5}h, y_i + \frac{3}{10}k_1h - \frac{9}{10}k_2h + \frac{6}{5}k_3h) \tag{9}$$

$$k_5 = f(x_i + h, y_i - \tfrac{11}{54}k_1h + \tfrac{5}{2}k_2h - \tfrac{70}{27}k_3h + \tfrac{35}{27}k_4h) \tag{10}$$

$$\begin{gathered} k_6 = f(x_i + \frac{7}{8}h, y_i + \frac{1631}{55296}k_1h + \frac{175}{512}k_2h + \\ \tfrac{575}{13824}k_3h + \tfrac{44275}{110592}k_4h + \tfrac{253}{4096}k_5h) \end{gathered} \tag{11}$$

E is the relative error comparing 4th order estimate of the RKF method to the 5th order estimate and is given in Equation 12. *E* is used to check for convergence of the two estimates by comparing its value to eps1 and eps2, see Figure C1; the optimal range for E is between eps1 and eps2. If the value of E drops below eps1 the step size, h, is doubled and if the value of E exceeds the value of eps2 the step size, h, is reduced by half.

$$E = \frac{Y5 - Y4}{Y5} \tag{12}$$

Adaptive step size. Adaptive step sizes in RKF require calculating the local truncation error for each step to see if the step size should be adjusted or not, see Figure C1. If there are areas of the function where abrupt changes occur, the adaptive step size approach can more efficiently and accurately estimate these regions by taking larger steps when the data trends are smooth and taking smaller steps when slope is changing quickly.

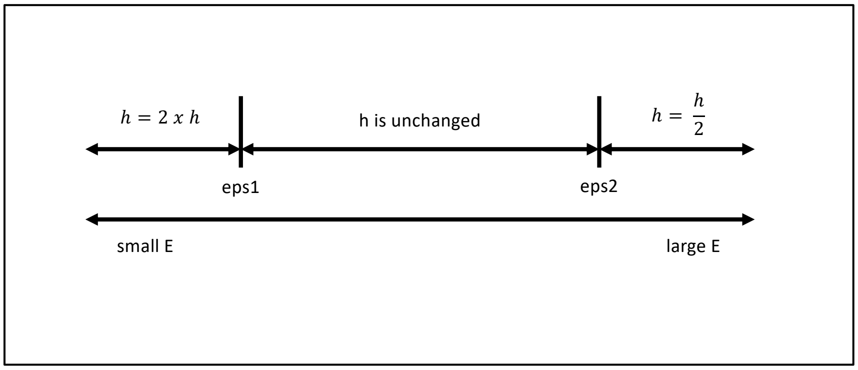


Figure C1. Adaptive step size graphic shows that when the relative error between 4th and 5th order estimates, E, is less than eps1 then the step size, h, is doubled; if the error is larger than eps2, the step size, h, is reduced by half. Adaptive step sizes allow for larger steps to be taken if the estimates are accurately modeling the ODEs but can reduce the step size selectively to deal with ODEs that have drastic changes in their slope.

Model runs

Purpose. The purpose of the model runs is to verify the model is matching the historical data well so that the time can be increase and coronavirus deaths and case counts may be forecasted. The S-variable will need to be changed and updated to account for surge period and model parameters α and β are adjusted as needed to fit the data (see Table C2). Initial step size is also adjusted to see if the model changes.

Table C2. Model run parameters for coronavirus modeling using a modified SIR model and Runga-Kuta Fehlberg (RKF) numerical method to solve the system of three ordinary differential equations (ODEs).

Symbol	Model Run Parameter	Value	Units
days	Number of days modeled	700	Days
h/hmax	Step Size/maximum step size	1	Day
hmin	Minimum step size (RKF)	0.00001	n/a
eps1	Small epsilon (lower bound)	0.0000001	n/a
eps2	Large epsilon (upper bound)	0.0001	n/a
tol	Convergence tolerance (RKF)	0.0000001	n/a

Finally, the model is run to perform the sensitivity analysis on model parameters α and β, the initial fraction of population infected, and the initial step size going into the RKF. α and β are increased and decreased by 20 percent and the initial fraction of population infected is increased and decreased by 50 percent and the initial step size is varied between 0.02 and 1.0. The resulting output is graphed and compared in the results section.

Program verification

The model is verified and follows standard trends for SIR-model. When compared to actual coronavirus cases and deaths, the values are within an acceptable range of error. The modifications to parameter of the model allow verification by visual inspection.

Appendix VII – Annotated Bibliography

Callahan, J., Hoffman, K., Cox, D., O'Shea, D., Pollatsek, H., Senechal, L. (2008). *Calculus in Context,* Five Colleges.

This textbook includes a section in chapter 1 about the SIR model for epidemics as an introduction to Calculus. The model is used for the measles (an illness that provides immunity for all recovered and is not deadly). The relationship between the rates of change of S, I, and R, are explored. The section talks about predicting change over time and two change parameters: 1) removal rate, and 2) transmission rate.

This text is useful for my project because the SIR model is broken down into simple language. Clear graphics illustrate the relationships between S, I, and R. The basic removal rate is obtained by dividing the number of infections by the time of illness duration. The transmission coefficient can be different for the same illness between different populations or with a quarantine enacted.

Chapra, S. C., and Canale, R. P. (2015). *Numerical Methods for Engineers,* 7th Ed., McGraw-Hill Education, New York, NY.

Cooper, I., Mondal, A., Antonopoulos, C. G. (2020). "A SIR model assumption for the spread of COVID-19 in different communities." *Chaos, Solitons, and Fractals*, 139, 1-14.

In this article, the authors demonstrate how a modified SIR model is used to fit coronavirus data. Predictions about future virus behavior are made. Assumptions of the SIR model are examined in light of the information available about COVID-19. The authors used removal rate, transmission coefficient, scaling factor, and the susceptible population as variables to fit the data by visual inspection. The S-variable is used to reset the susceptible population to a surge value during virus surge periods.

This article is applicable to my project because the authors are making future predictions about the virus. Parameter values for virus growth and decay as well as the scaling factor and a susceptible population at time = 1 are provided. The authors are scaling the values by the entire population so S + I + R = 1. Charts of deaths vs removals are useful for comparison. For USA S is varying between 0.06 and 0.20.

Hethcote, H. W. (2000). "The Mathematics of Infection Diseases." *SIAM Review,* 42(4), 599-653.

This article thoroughly discusses infectious diseases and their differences or similarities. A variety of mathematical models for infectious diseases are examined with their history and evolution. Epidemic models should be used for diseases that go away after one year. Endemic models are appropriate when: 1) disease last longer than one year, 2) the total population cannot be assumed to be steady, or 3) recovery does not guarantee immunity.

This article is useful for my project because the differences between models is discussed along with their formulation. Further discussion of the parameters used in the SIR model is found within. The basic reproduction ratio is introduced and discussed; reproduction ratio is equal to the contact number when the entire population is susceptible). The background on the different diseases may offer some context for the coronavirus in my document.

Jo, H., Son, H., Hwang, H. J., Jung, S. Y. (2020). "Analysis of COVID-19 spread in South Korea using the SIR model with time-dependent parameters and deep learning."

This article uses a SIR model for COVID-19 along with deep learning. Deep learning is an approximation method for non-linear functions. The authors use a forward-inverse problem approach to approximate S, I, R, beta, and gamma. Beta and gamma are the growth and decay coefficients of the SIR model.

This article is useful because beta and gamma values are determined and can be compared to the results of this project. The data is modeled with the deep learning technique; deep learning may prove to be the best method for determining the parameter values for SIR model. Figure X is a concise and clear graphic of the SIR model. Parameter values are a function of time instead of constant coefficients.

Wu, J. T., Leung, K., Leung, G. M. (2020). „Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study." *Lancet 2020*, 395, 689-697.

This article talks about the basic reproductive number, the R-value, for the coronavirus and provides doubling time for the disease. This article was published January 31, 2020 and thus there was not as much data available as there is at the time of writing this. In the article, data was examined from Dec 31, 2019 until Jan 28, 2020.

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., Guan, L., Wei, Y., Li, H., Wu, X., Xu, J., Tu, S., Zhang, Y., Chen, H., Cao, B. (2020). "Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study." *Annals of Oncology,* 31(7), 838-839.

This article discusses the medical outcome of 191 hospital patients in Wuhan, China diagnosed with COVID-19. Statistical analysis is performed to determine the mean time of hospital stay, onset to fever, onset to cough, onset to ICU admission, onset to death or discharge, and duration of viral shedding after onset for survivors and non-survivors. Vital statistics for the patients are determined. Health histories of patients and vital statistics are also documented.

This article is useful for the project because the mean time from onset to discharge or death is given. Recovery rate of the 191 patients may be used. Mean time of viral shedding may be used to determine the average time a coronavirus patient is contagious. Discussion points abound due to the comorbidities and demographics of the patients examined. Median duration of viral shedding 20.0 days.

Data retrieved from the USAfacts website will be used. Data is downloaded in .csv files. Every county of the USA is listed in columns and the rows list each day from January 22, 2020 until present. Cells are populated deaths from COVID-19 while a separate file contains confirmed cases from COVID-19.

Data from the WHO website contains nationwide deaths and confirmed cases data for the USA from January 3, 2020 until present in .csv file. Deaths and confirmed cases listed in rows and days listed in columns.

Census.gov

Population per square mile in California 239.1(2010) and for Humboldt County it is 37.7 people per square mile.

CollegeFactual.com

89 percent undergraduates are from California

Humboldt.edu

16.9 percent of students are local and another 11.1 percent are from Northern California

California Dept of Public Health (cdph.ca.gov)

Appendix VIII – Additional Figures and Tables

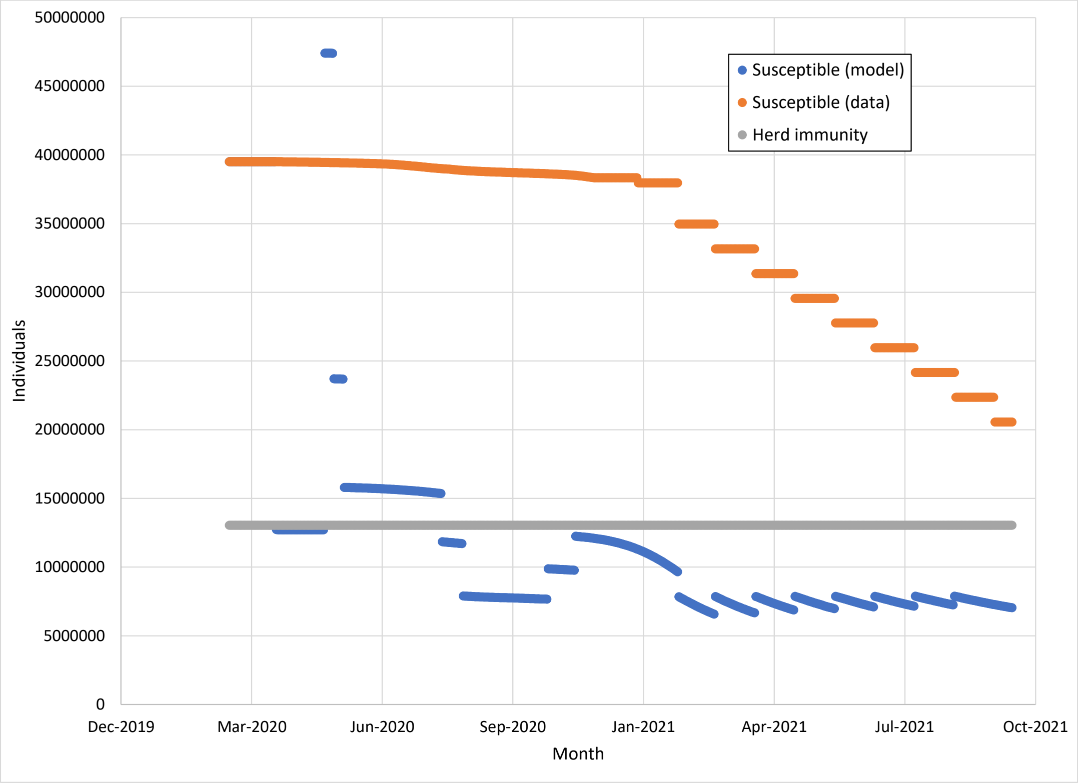


Figure D1. S population over time with reductions in the S population from simulation of vaccination on the S population but data ends on November 27,2020 so removals after that date are only from vaccines and so more removals are inevitable during that time period and may account for the gap between R population and herd immunity threshold. The S population from the model is kept low to delay a peak in infections and keep model R population in line with R estimated from the data.

Table D1. Model parameter updates and vaccination values from day 37 until day 587. Note S is kept constant but updating it every month with the vaccination update. The infection rate, α, is changed from 0.23 to 0.35 at day 218 to match historical data and is kept at 0.35 due to lack of scientific reason to change it for future model predictions.

Day	Date	Svalue	Infection Rate (α)	Immunizations	Duration (days)
37	3/29/2020	0.322	0.23	none available	37
74	5/5/2020	1.2	0.23	none available	37
81	5/12/2020	0.6	0.23	none available	7
89	5/20/2020	0.4	0.23	none available	8
164	8/3/2020	0.3	0.23	none available	75
180	8/19/2020	0.2	0.23	none available	16
218	9/26/2020	no change	0.35	none available	38
245	10/23/2020	0.25	0.35	none available	27
266	11/13/2020	0.31	0.35	none available	21
280	11/28/2020	no change	0.35	none available	14
314	12/31/2020	no change	0.35	375000	48
345	1/31/2020	0.2	0.35	3000000	31
373	2/28/2020	0.2	0.35	1800000	28
404	3/31/2020	0.2	0.35	1800000	31
434	4/30/2020	0.2	0.35	1800000	30
465	5/31/2020	0.2	0.35	1800000	31
495	6/30/2020	0.2	0.35	1800000	30
526	7/31/2020	0.2	0.35	1800000	31
557	8/31/2020	0.2	0.35	1800000	31
587	9/30/2020	0.2	0.35	1800000	30

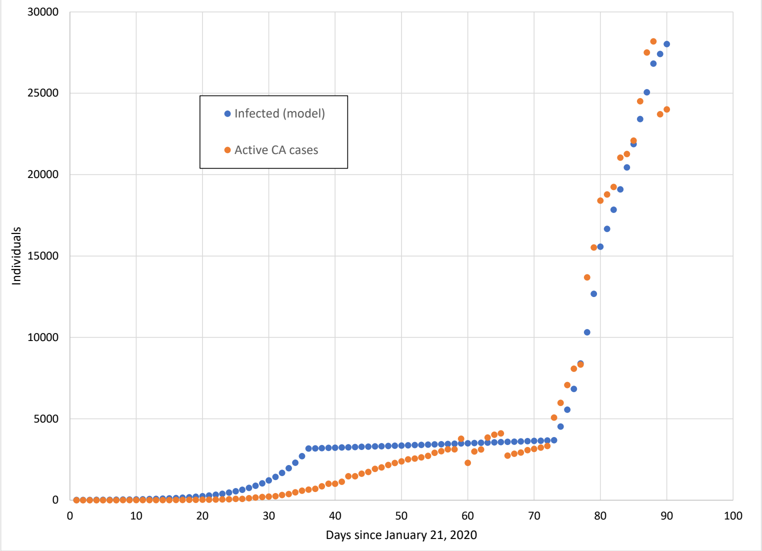


Figure D2. Active coronavirus cases and model infections are compared over a span of 90 days beginning January 21, 2020. The S-population is set at 1.0 for the first 36 days and is manually updated at day 37 accounting for the sudden change in slope of the I-population in the model; S was again altered at day 73 to follow the surge of cases recorded in the data (CDC).

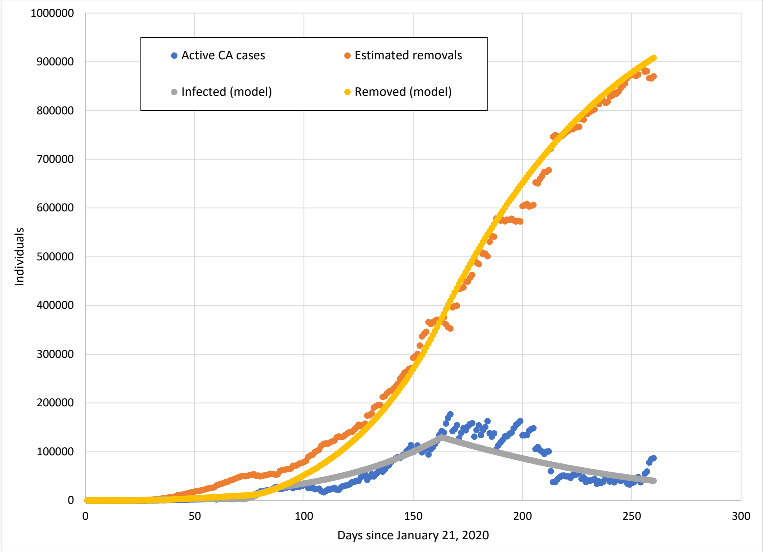


Figure D3. Modeling I and R-populations from the SIR model with historical data of removals and active cases (CDC). Removals are modeled with adjusting the S population as the data suggests; changes in S-population are indicative of varying levels of success with social distancing compliance.

Appendix IX – Logbook

This log is for the COVID-19 SIR MODEL project for Tom Jackson Jr president of HSU beginnings 09/01/2020 and ending on 12/11/2020.
Engineer: Caleb Wegener
Class: ENGR326
Valum Engineering
Humboldt State University

Data included spans from the beginning of 2020 until October giving adequate information for predictions ahead. Possibly up to 300 days after November first which would provide predictions until the beginning of fall semester 2021.

Mr Jackson is a man deeply committed to the university and the safety of all members of it's community. It is a pleasure that Mr. Jackson is choosing to contract Valum Engineering for this project and the importance of this project is second to none. The best data and most accurate approximations should be used to obtain the most accurate estimations for death and infection rates for HSU student population.

Assumptions for the project. HSU student body is relatively well represented by the demographics of California as a whole and therefore using state data in our model should prove accurate. Other scenarios might be tested such as if the student population is better represented by counties in Northern California, or Humboldt county because there is sufficient data on all regions of the state and country.

Data can be compared to findings in Cooper et al. for verification of program and model functionality.

TO DO

Begin writing code for system of equations including a,b,f

Get a reference list together. Length of average coronavirus infection?

Time (4 hours)

9/19/2020

-Design brief, constructing problem statement

4 hours

Time (1 hour)

09/24/2020

-data searching (obtained data from CDC all counties of USA for 2020 since until present) (deaths and infections)

Checklist before turning in any writing

-Did I cite all sources? All equations? Textbook?
-Do in-text citations match the references list
-Are references called "references"?
-All text matches? Style? Matches requirements?
-check all formatting
-check all topic sentences and supporting sentences.

Time(1.1 hours)

October 10/2020

Tasks: Literature review (cooper et al) and data analysis

In Cooper et al. the entire population is not considered. "Susceptible population is a variable that can be adjusted at various times to account for new infected individuals (Cooper et al)."

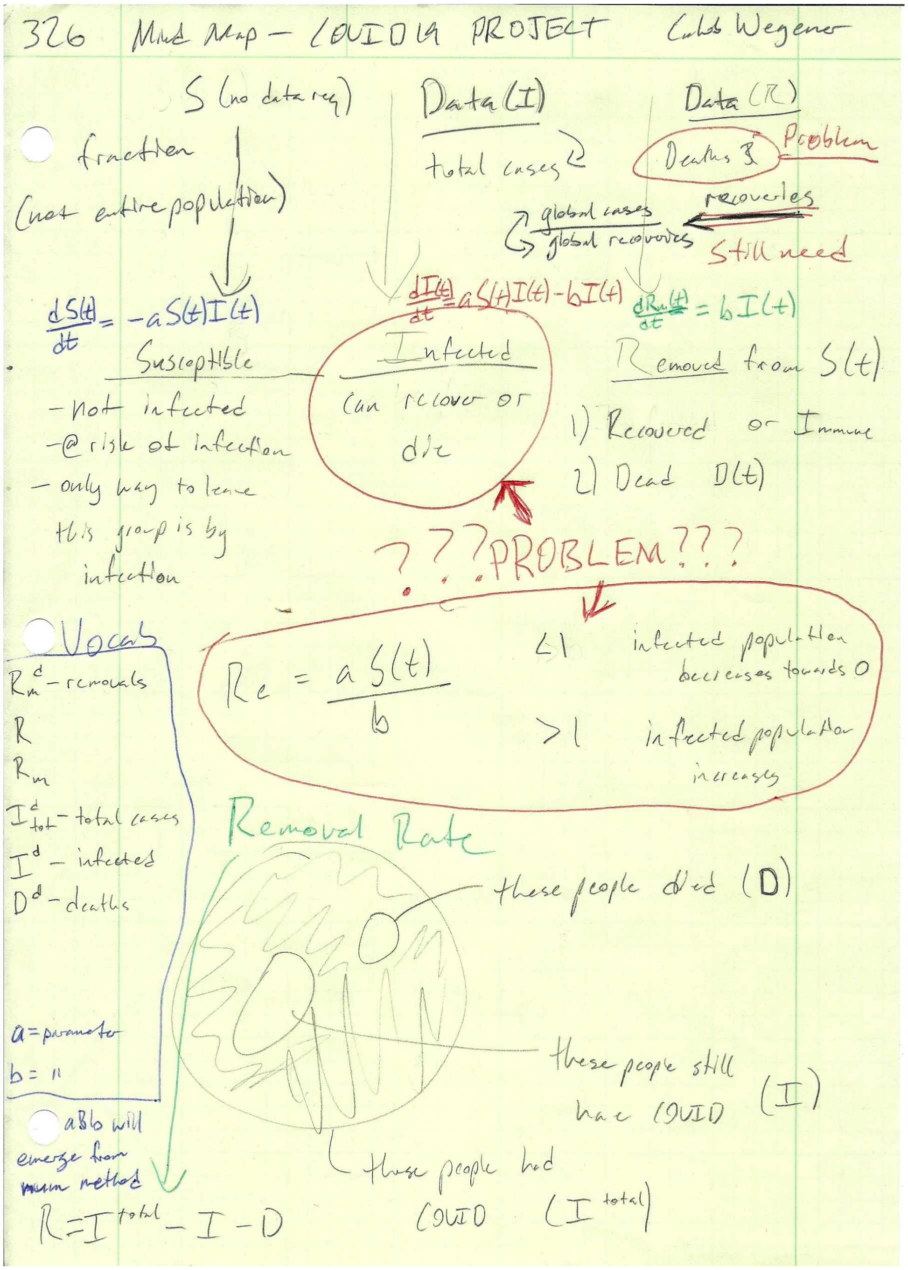
Zhou et al: recovery rates/death rates of hospitalizations due to covid in Wuhan china; time to discharge 22days

0.5 hours

10/18/2020

Tasks

Mind mapping. Reviewing data needs.



326 Mind Map — COVID19 PROJECT Caleb Wegener

S (no data req)

fraction

(not entire population)

$\frac{dS(t)}{dt} = -aS(t)I(t)$

Susceptible

- not infected
- @ risk of infection
- only way to leave this group is by infection

Data (I)

total cases

$\frac{dI(t)}{dt} = aS(t)I(t) - bI(t)$

Infected

Can recover or die

Data (R)

Deaths Problem

recoveries

global cases

global recoveries

Still need

$\frac{dR_m(t)}{dt} = bI(t)$

Removed from S(t)

1) Recovered or Immune

2) Dead D(t)

???PROBLEM???

$$R_e = \frac{aS(t)}{b}$$

<1 infected population decreases towards 0

>1 infected population increases

Vocab

R_m^d - removals

R

R_m

I_{tot}^d - total cases

I^d - infected

D^d - deaths

a = parameter

b = "

a&b will emerge from main method

Removal Rate

these people died (D)

these people still have COVID (I)

these people had COVID (I^{total})

$R = I^{total} - I - D$

Friday October 23, 2020

Time 2.5 hours

Tasks:

Setting up project document (title)

Creating Fortran Project

<u>Data</u>

Covid_confirmed_usafacts

-county by county for entire usa 1/22/2020-present(9/26)

-Cumulative confirmed cases

Covid_deaths_usafacts

-county by county for entire usa 1/22/2020-present(9/26)

-cumulative deaths

WHO-COVID

1/3/2020-present(10/23) UNITED STATES TOTALS
Cases and Deaths, cumulative and new
Row 66082 begins USA
USA has 20% of global cases WHO and 4% of population (wiki)
China has 0.2% of cases and 18% of population (wiki)

Saturday, October 24, 2020

Time 5 hours

Tasks:

-Writing the program
-Setting up data for trial run
-lit review
S + I + R = 1
(Biswas et al. 2020)
Biswas, K. Khaleque, A., Sen, P. (2020). “Covid-19 spread: Reproduction of data and prediction using a SIR model on Euclidean network.”
“Some health officials now consider anyone who was diagnosed with COVID-19 three or more weeks ago and has not died to be recovered from the disease. “ worldometer website
“whereas the median time to death was 18.5 days” (Zhou et al. 2020)
Time from illness onset to death or discharge, days (Zhou et al. 2020)
Domo data

Sunday, October 25, 2020

Time 4.5 hours

Tasks

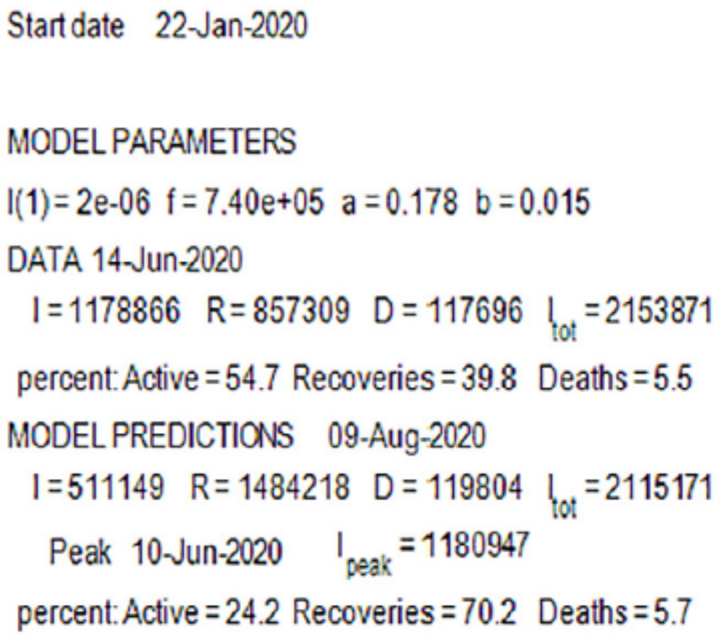
-Getting data in excel for humboldt county and the state of california.
-Looking for data for recoveries in time
-contacting worldometer via online web form and obtaining humboldt county contact for health and human services (will call monday)
-Setting up removal system by taking total cases minus total deaths minus cases more than 18 days old that did not result in death.
Worldometer says 465k recovered in CA and 424k still infected. I don’t think this is correct. I think excel data is looking pretty good. Consider revising for 14-day incubation period before recovery declaration.

Monday October 26, 2020

Time 1.5 hours

Tasks:

-literature review
-database searches
-waiting in Beths office hours for help
Data from Tableau app
git hub data
https://data.ca.gov/
Cdc
Who
Johns Hopkins University dashboard
Humboldt County Data
California only data
Cooper et al.



```
Start date    22-Jan-2020

MODEL PARAMETERS
I(1) = 2e-06  f = 7.40e+05  a = 0.178  b = 0.015
DATA 14-Jun-2020
  I = 1178866   R = 857309   D = 117696   I_tot = 2153871
 percent: Active = 54.7  Recoveries = 39.8   Deaths = 5.5
MODEL PREDICTIONS    09-Aug-2020
  I = 511149   R = 1484218   D = 119804   I_tot = 2115171
    Peak  10-Jun-2020      I_peak = 1180947
 percent: Active = 24.2  Recoveries = 70.2   Deaths = 5.7
```

Sunday November 8, 2020
Time 1.5 hours
Tasks:
-coding/debugging
-data review
Monday November 9, 2020
Time 0.5 hours
Tasks:
Office hours - discussing model and scope of work with model
To do:
Define r value
Use model with calculus problem and data set
Tuesday November 10, 2020
Time 0.5 hours
Tasks
Adding fileprompt module with fileopen subroutine
SIR model and graphing in excel with values for a and b from cooper.
Looking for data set in *Calculus in Context*
Backing up program
Wednesday November 11, 2020
Time 2.0 hours
Tasks:
Model runs: compare model to california data using a and b from Cooper et al.
To do:
Questions: does multiplying the deaths by a death rate produce more accurate numbers for infections?
Using those estimates, is the SIR more accurate?
Program SIR into excel
Calculate california infection rate
-make program user friendly with detailed instructions for inputs and outputs
Questions:

Can we make A and B a function of time? Allow them to change as data suggests
SIR is a cumulative model. How can I reset S to S(surge) without messing up my ratios?
https://med.stanford.edu/news/all-news/2020/09/few-americans-have-coronavirus-antibodies-study-finds.html
9% americans have antibodies or have been infected. Verify!

November 13, 2020

Time 10 hours

Tasks:

Writing - design steps and appendix #2 (METHODS)

To do:

Why is runga-kuta fehlberg great for solving ODE (besides adaptive step size)

Notes:

viral shedding time from onset 20.0 days (Zhou et al)

November 23, 2020

Time 4 hours

Tasks:
Recovery data in excel for humboldt county
-graphing and fitting
Humboldt county dashboard has cumulative cases, recoveries, and deaths
Data.ca.gov has cumulative cases and deaths and new cases and deaths for all counties from 3/18-11/22
Recovery ratio (recoveries/total cases) for humboldt county
Data for Humboldt county not in agreement between dashboard and CDC
Problems
Not sure how to scale the data to get it to fit into the graph (California data is ending with very high value for Susceptible population in recorded data)
Use ratio of recoveries from Humboldt dashboard data to estimate recoveries for california and use recoveries and deaths to determine removals
Active cases determined by subtracting total deaths and total recoveries from total case count

November 24, 2020

Time 5 hours

Tasks:

Use Fortan SIR model to fit data for california and humboldt county and maybe USA data as well.

Record any findings for scaling, start time, changes in parameter values etc.
Reviewing literature for information about fitting the model to the data.
Run model for different scenarios
I(0)=R(0) = Cooper value with a and b from cooper and comparing to california data (no scaling with entire population)
Try a = 0.0178
I(0)=R(0) = max infected/f and compare to USA data

Notes

S+I+R is not necessarily equal to 1 because of "recharging" the susceptible population
A = 0.178 B = 0.30 works good for first 52 days, after which the infected population skyrockets and california data goes down. (note until day 60 of model, the recoveries from data have not yet been accounted for so numbers of active infections before that are not accurate, so trying a smaller b value to see if I can match the model to the active cases on day 60 (3771)
It is possible that the model should be done in a piecewise fashion. Using different model parameters and starting with initial conditions as model run parameters. Since the circumstances around covid have been continually evolving, it makes sense that the model parameters will not be constant throughout.
Starting on February 20 and modeling 89 days of data with a = 0.23 and b = 0.07 and setting S at the following days to the following values. Comparing the total removals to the data the percent relative error

is less than 0.5%. Note the removal data is not available before 4/20/2020 because that is the day Humboldt county started updating their recoveries. Used total population to scale S, I, and R and 10 persons initially infected. S = 1 and R=0
(Day/S value)
37/0.322
74/1.2
81/0.6
91/0.2

November 25, 2020

Time 3.5 hours

Tasks:

-Document organization. Pulling together parts that are already written and arranging in final order. Starting at the top, assign one section per day to be written/revised/updated.

-Modeling starting at 92 days after February 20, 2020. (Determining a and b and any surge periods)

Notes:

COVID-19 and online teaching in higher education: A case study of Peking University Wei Bao

"In addition, based on an analysis of students' responses in social media, for such a large-scale online teaching, the challenges for students did not come from technical operational obstacles. Instead, they have difficulties due to the lack of a good learning attitude. Students often have problems such as lack of self-discipline, suitable learning materials, or good learning environments when they are self-isolated at home."
(https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbe2.191)

students' autonomous learning abilities and habits of learning behaviors ranked as having the biggest effect on effectiveness of online learning/comprehension. Increased workload on instructors.
To cite this article: Chiu-Lan Chang and Ming Fang 2020 J. Phys.: Conf. Ser. 1574 012166
(https://iopscience.iop.org/article/10.1088/1742-6596/1574/1/012166/pdf)

Online learning vs Face-to-face learning (case study)
Reassessing Disparities in Online Learner Student Engagement in Higher Education Justin Paulsen1 and Alexander C. McCormick1
http://baku8km.khazar.org/bitstream/20.500.12323/4406/1/reassessing%20disparities%20in%20online%20learner%20student.pdf

First Case of 2019 Novel Coronavirus in the United States (Holshue et al 2020)

https://www.nejm.org/doi/full/10.1056/NEJMoa2001191

To Do:

- Review sources for intro, start working on next section
- Modeling

November 26, 2020

Time 3 hours

Tasks:

-REVIEW OF THE PROBLEM and INTRO sections (writing)

-more research into cons of online learning (OL)

-more modeling (try to match total case counts and removals and intervals using less surge periods)

Notes:

- use acronyms for face-to-face instruction (FFI) and online learning (OL)
- Define "campus reopening" and "safe" in reference to campus reopening

New Sources

- Aguilera-Hermida 2020 College students' use and acceptance of emergency online learning due to COVID-19

- Nambiar - The impact of online learning during COVID-19: students' and teachers' perspective
-

November 27, 2020
Time 4 hours
Tasks:
-Modeling cases from day 90 using estimated removals and active cases to set infected and removed and susceptible population = 1 - I - R and starting with cooper values of a = 0.178 and b = 0.015
-Review of the problem and intro sections
-programming fortran to read in values for active cases and removals from 2/21-11/7
Notes:
Assumptions in data - removal ratio matches Humboldt county and days until 4/20 use the same ratio of removals and 4/20. Ratios are constantly changing after that date
To do:
-figure out why the first value of the csv input file is not getting into fortran
-get code to update a and b values on its own
November 28, 2020
2.5 hours (see excel)
Time log in excel (notes here in google docs
November 29, 2020
TO DO ASAP

- Define R and report values in results section (discuss in discussion)
- Get SIR charts in document ASAP (sir from historical data scaled- and unscaled)
- Get MODEL RUN chart in Document (
- talk about scaling factor)

Now Modeling 279 days since 2/21/2020 with a= 0.23 and b=0.07 with I(0)= 10 and R(0)=0
S(0) = 1 - I - R
S set at:
37, S=0.322
74, S=1.2
81, S=0.6
89, S=0.4
164, a = 0.3, S=0.2
November 30, 2020
Notes:

- Toda SIR finds a and b with historical data and least squares A = 0.2-0.4
- And b= 0.1
- Cooper a = 0.178 for USA and b= 0.015
- Add separate equations for the fractional form of the ODEs (toda)

Finish up modeling. (change a and b see if model can fit peaks or removals better with less changes to S)
Print charts and put into document and start talking about it
December 1, 2020
Intro notes

- **What is the general problem you are addressing, and why is the general sort of problem important or interesting to solve? The problem is the campus is closed and almost all classes are being held online. Students and faculty alike are anxious to return to the classroom.**
- **(learning gaps, less group/team work, less hands on learning, loss of resources for students and faculty)**
- **What are the objectives you intend to obtain? Obtain a model (for coronavirus) that can produce data identifying a possible campus reopening date**

- **What is the *specific* problem you mean to solve? Determine when COVID cases and deaths are low enough in California to safely reopen HSU campus by modelling the virus with SIR model.**
- **What is the method you will use to solve the problem? To solve the SIR model for epidemics, a numerical method will be applied and historical data compared the model output to determine model reliability.**
- **Why is it important for the client for you to solve the problem? Tom Jackson Jr needs reliable information to make decisions about reopening HSU campus. Lives may be lost by reopening the campus too soon; students and faculty alike will unnecessarily suffer from the campus being closed longer than is necessary.**
- **What are your expected final outcomes from your analysis? Expected final outcome is that model may not predict accurately into the future due to necessary updates to susceptible population. If model follows the data well, it likely cannot predict surge periods so some assumptions may be made based on surge periods from 2020 as model predicts behavior of the virus in 2021.**
- **If your project is successful, what will be the lasting effect of your efforts? The education and wellbeing of _______ thousand staff and faculty depends on choosing a safe reopening date for HSU campus and lives will not be lost unnecessarily within the staff and faculty but also within the community.**

What is the structure of the report? Following the introduction, more information about the problem will be given along with outlines from similar projects. The design and solution steps will be covered next and are followed by the results, discussion, and final recommendation.

Extra topics, odd pieces of writing from document:

Things we might learn from the experience are vast but are beyond the scope of this project. Here the focus is the struggles and challenges facing the HSU student body due to online classes and what is needed to return to face-to-face instruction.

Challenges facing HSU students and faculty

- Learning and teaching challenges: (Avoiding learning gaps (some students may excel online while others fall behind)), Added workload for professors (lack of online content/resources), environment (conditions at home, distractions, less interaction with peers and faculty)
- For students studying online, access to technology/connectivity may be a hinderance.

December 1, 2020

https://www.northcoastjournal.com/NewsBlog/archives/2020/03/17/hsu-campus-closed-to-general-public-instruction-going-virtual-until-end-of-year

HSU CLOSES on MARCH

https://onlinelibrary.wiley.com/doi/pdf/10.1002/jmv.25827 Hou et al 2020 (NEW SOURCE)

The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): A well-mixed SEIR model analysis

December 5, 2020

https://www.cnn.com/2020/11/30/health/covid-vaccine-questions-when/index.html

https://calmatters.org/health/coronavirus/2020/12/california-priorities-first-covid-vaccines/

FOR USA

Best case scenario for vaccine (both vaccines approved)

20 million vaccinated by the end of december
25 million by end of january
15 million by the end of feb
15 million by the end of march

Only one vaccine approved

10 million by end of december

12.5 million end of january
7.5 million end of Feb
7.5 million end of March
FOR CALIFORNIA (california population is 12% of USA population)
Best case scenario for vaccine (both vaccines approved)
327,000 vaccinated by the end of december (3.3% of decembers vaccines)
3 million by end of january
1.8 million by the end of feb
1.8 million by the end of march
Only one vaccine approved
327,000
1.5 million end of january
900,000 end of Feb
900,000 million end of March

39,512,223 (CA POPULATION)

330,668,187 (US POPULATION)

The total cases for California as of Nov 12, 2020 is 991,609 while 18,108 people have died from the virus.

December 8, 2020
Average recovered ratio for humboldt county is 0.88832
135,558 california population.
i

Appendix X – Fortran Code

```
module types
    Integer, parameter :: WP=selected_real_kind(15)
end module types
Module fileprompt   !File open utility
    contains
        subroutine fileopen(unitnumber,readfile)
            implicit none
            Integer, parameter :: wp=selected_real_kind(15)
            Integer, intent (in) :: unitnumber
            logical, intent (in) :: readfile
            !local variables
            Integer :: ioerror
            Character (3) :: fmt = "(A)"
            Character (60) :: filename
            do
                if (readfile) then
                    write (*,fmt,advance="no") "Input filename"
                else
                    write(*,fmt,advance="no") "Output filename"
                end if
                Read (*,fmt) filename
                if (readfile) then
                    open (unitnumber,file=filename,status="old",iostat = ioerror)
                else
                    open (unitnumber,file=filename,iostat=ioerror)
                end if
                if (ioerror==0) exit
                write (*,*) "problem opening file, try again"
            end do
        end subroutine fileopen
end module fileprompt
module mod2 !Home of the RUNGA-KUTA SUBROUTINE
    use types
    implicit none
contains
    subroutine RKF (tinit,tf,y,eps1,eps2,h,hmin,hmax,dy,exitflag,n,a,b)
        implicit none
        Integer, intent(in) :: n
        Real(wp), intent(in) :: tinit,tf,eps1,eps2,hmin,hmax,a,b
        Real(wp), intent(inout) :: h
        Integer, intent(out) :: exitflag
```

```
        Real(wp), dimension(:), intent(inout) :: y
        Real(wp), dimension(n) :: k1,k2,k3,k4,k5,k6,y4,ysave
        Real (wp) :: E, t, hsave, tol
        !Runga-Kuta Fehlberg Slope Estimate Fractions (Parameters)
        Real(wp), Parameter :: a2=(1._wp/5._wp)
        Real(wp), Parameter :: a3=(3._wp/10._wp)
        Real(wp), Parameter :: a4=(3._wp/5.)
        Real(wp), Parameter :: a5=1._wp
        Real(wp), Parameter :: a6=(7._wp/8._wp)
        Real(wp), Parameter :: b21=(1._wp/5._wp)
        Real(wp), Parameter :: b31=(3._wp/40._wp)
        Real(wp), Parameter :: b32=(9._wp/40._wp)
        Real(wp), Parameter :: b41=(3._wp/10._wp)
        Real(wp), Parameter :: b42=(9._wp/10._wp)
        Real(wp), Parameter :: b43=(6._wp/5._wp)
        Real(wp), Parameter :: b51=(11._wp/54._wp)
        Real(wp), Parameter :: b52=(5._wp/2._wp)
        Real(wp), Parameter :: b53=(70._wp/27._wp)
        Real(wp), Parameter :: b54=(35._wp/27._wp)
        Real(wp), Parameter :: b61=(1631._wp/55296._wp)
        Real(wp), Parameter :: b62=(175._wp/512._wp)
        Real(wp), Parameter :: b63=(575._wp/13824._wp)
        Real(wp), Parameter :: b64=(44275._wp/110592._wp)
        Real(wp), Parameter :: b65=(253._wp/4096._wp)
        Real(wp), Parameter :: c1=(37._wp/378._wp)
        Real(wp), Parameter :: c2=(250._wp/621._wp)
        Real(wp), Parameter :: c3=(125._wp/594._wp)
        Real(wp), Parameter :: c4=(512._wp/1771._wp)
        Real(wp), Parameter :: dc1=(2825._wp/27648._wp)
        Real(wp), Parameter :: dc2=(18575._wp/48384._wp)
        Real(wp), Parameter :: dc3=(13525._wp/55296._wp)
        Real(wp), Parameter :: dc4=(277._wp/14336._wp)
        Real(wp), Parameter :: dc5=(1._wp/4._wp)
interface
    function dy (t,y,n,a,b)
        use types
        implicit none
        Integer, intent(in) :: n
        Real (wp), intent(in) :: t,a,b
        Real(wp), dimension(:), intent(in) :: y
        Real(wp), dimension(n) :: dy
    end function dy
end interface
        tol = 0.0000001_wp
        !Error trapping
```

```fortran
        If (tinit>tf) then
            write (*,*) "The starting time is greater than ending time."
            stop
        end if
        if (eps1>eps2) then
            write (*,*) "The first stopping criteria is larger than the second stopping criteria."
            stop
        end if
        if (hmin>hmax) then
            write (*,*) "The minimum step size is larger than the maximum step size."
            stop
        end if
        if (h>(tf-tinit)) then
            h = tf - tinit
        end if
        hsave = h
        t = tinit
        do
            ysave = y(:n)
!            write (*,*) "ysave = ", ysave
            !Calculate Slope Estimates
            K1=dy(t,y(:n),n,a,b)
            K2=dy(t+a2*h,y(:n)+b21*k1*h,n,a,b)
            K3=dy(t+a3*h,y(:n)+b31*k1*h+b32*k2*h,n,a,b)
            K4=dy(t+a4*h,y(:n)+b41*k1*h-b42*k2*h+b43*k3*h,n,a,b)
            K5=dy(t+a5*h,y(:n)-b51*k1*h+b52*k2*h-b53*k3*h+b54*k4*h,n,a,b)
            K6=dy(t+a6*h,y(:n)+b61*k1*h+b62*k2*h+b63*k3*h+b64*k4*h+b65*k5*h,n,a,b)
     !Calculate 4th and 5th order estimates for t + h
            y4 = y(:n)+(c1*k1+c2*k3+c3*k4+c4*k6)*h
            y(:n) = y(:n) + (dc1*k1+dc2*k3+dc3*k4+dc4*k5+dc5*k6)*h
            !Check Convergence Criteria
            E = maxval (abs((y(:n)-y4)/y(:n)))
                if (E > eps2) then
                    if (abs(hmin-h) < tol) then
                        write (*,*) "hmin-h is less than tolerance"
                        exitflag = 1
                        return
                    end if
                    h = h/2
                    if (h < hmin) h = hmin
                    y (:n) = ysave
                else
                    t = t + h
                    if (E < eps1) then
                        h = h * 2._wp
```

```
                        if (h>hmax) then
                            h = hmax
                        end if
                    end if
                    if (abs(t-tf) < tol) then
                        exitflag=0 !all good in the hood
                        h = hsave !send back the original step size
                        return
                    end if
                    if (t+h > tf) then
                        hsave = h
                        h = tf - t
                    end if
                end if
        end do
    end subroutine RKF
end module mod2
Program covid19
!The objective of this program is to model the COVID-19 virus for the student population
!of Humboldt State University using a modified SIR model for  prediction infections from
!the virus using historical data and identified surge periods to better model data into
!the year 2021.
!
!Data used will be from the CDC and_________
!Input will be loaded in with numbers of infections, deaths, and susceptible will will be
!determined from the historical data modeling.
    use types
    use fileprompt
    use mod2
    implicit none
!Declare Local Variables
    Integer :: exitflag, i, n, days
    Real (wp), dimension(:), allocatable :: y, susceptible, infected, removed
    Real (wp) :: eps1, eps2, h, hmin, hmax, tinit, tf, initialinfected, initialremoved
    Real (wp) :: poptotal, a, b
    Real (wp) :: jan,feb,mar,apr,may,jun,jul,aug,sep,oct,nov,dec,jan1, sfuture, sens
    Real (wp) :: jans,febs,mars,aprs,mays,juns,juls,augs,seps,octs,novs,decs,jan1s
Interface
    Function SIR (t,y,n,a,b)
        use types
        Real(wp), Dimension (:), Intent(in) :: y
        Real(wp), Intent (in) :: t,a,b !Time (Days)
        Integer, Intent (in) :: n
        Real(wp), Dimension(n) :: SIR
```

```
    End Function SIR
End Interface
!Numerical Method Parameters (SIR model parameters a & b defined in external function)
    tinit = 0._wp
    tf = 1._wp
    eps1 = 0.0000001                !RKF lower bound for time step adjustments
    eps2 = 0.0001                   !RKF upper bound for time step adjustments
    h = 1._wp                       !Initial time step
    hmin = 0.00001_wp               !Min time step
    hmax = 1._wp                    !Max time step
    !sfuture = 0._wp                !s-value for all future updates
!MODEL RUN VALUES
    n = 3                           !Assign "n" number of ODEs
    days = 700                  !Assign number of days for model
    initialinfected = 10._wp
    initialremoved = 0._wp
    poptotal = 39512223._wp         !US CENSUS BUREAU - CALIFORNIA POPULATION (2019)
    a = 0.23_wp   !Declare parameter of initial exponential growth of Infected population (Cooper)
    b = 0.07_wp   !Declare parameter of exponential decay of Infected population           (Cooper)
!Allocate array for SIR-model with n = 3 ODEs
    Allocate (y(n))
    Allocate (susceptible(days))
    Allocate (infected(days))
    Allocate (removed(days))


    y = 0._wp
    susceptible = 0._wp
    infected = 0._wp
    removed = 0._wp
!Open I/O files
    Call fileopen (15,.false.)
!    Call fileopen (20,.true.)
!write(*,*) "y(2) = ", y(2)
    write (15,707)
    707 format ("Days,Susceptible(persons),Infected(persons),Removed(persons)")
    !Declare initial values of SIR model (S+I+R=1)
            y(2) = initialinfected/poptotal        ! Infected population (1 infected individual/california population)
            y(3) = initialremoved/poptotal         ! Removed population
            y(1) = 1-y(2)-y(3)         ! Susceptible population
!Prompt the user for vaccine numbers for January through September
        write(*,*) "Enter the number of vaccines administered by Jan 1, 2021"
        read (*,*) jan
        write(*,*) "Enter the number of vaccines administered by Feb 1, 2021"
        read (*,*) feb
```

```
write(*,*) "Enter the number of vaccines administered by Mar 1, 2021"
read (*,*) mar
write(*,*) "Enter the number of vaccines administered by Apr 1, 2021"
read (*,*) apr
write(*,*) "Enter the number of vaccines administered by May 1, 2021"
read (*,*) may
write(*,*) "Enter the number of vaccines administered by Jun 1, 2021"
read (*,*) jun
write(*,*) "Enter the number of vaccines administered by Jul 1, 2021"
read (*,*) jul
write(*,*) "Enter the number of vaccines administered by Aug 1, 2021"
read (*,*) aug
write(*,*) "Enter the number of vaccines administered by Sept 1, 2021"
read (*,*) sep
        write(*,*) "Enter the number of vaccines administered by OCT 1, 2021"
read (*,*) oct
        write(*,*) "Enter the number of vaccines administered by NOV 1, 2021"
read (*,*) nov
        write(*,*) "Enter the number of vaccines administered by DEC 1, 2021"
read (*,*) dec
        write(*,*) "Enter the number of vaccines administered by JAN 1, 2022"
read (*,*) jan1

write(*,*) "Enter the S-value for Jan 1, 2021"
read (*,*) jans
write(*,*) "Enter the S-value for Feb 1, 2021"
read (*,*) febs
write(*,*) "Enter the S-value for Mar 1, 2021"
read (*,*) mars
write(*,*) "Enter the S-value for Apr 1, 2021"
read (*,*) apr
write(*,*) "Enter the S-value for May 1, 2021"
read (*,*) mays
write(*,*) "Enter the S-value for Jun 1, 2021"
read (*,*) juns
write(*,*) "Enter the S-value for Jul 1, 2021"
read (*,*) juls
write(*,*) "Enter the S-value for Aug 1, 2021"
read (*,*) augs
write(*,*) "Enter the S-value for Sept 1, 2021"
read (*,*) seps
write(*,*) "Enter the S-value for OCT 1, 2021"
read (*,*) octs
write(*,*) "Enter the S-value for NOV 1, 2021"
read (*,*) novs
```

```
        write(*,*) "Enter the the S-value for DEC 1, 2021"
        read (*,*) decs
        write(*,*) "Enter the the S-value for JAN 1, 2022"
        read (*,*) jan1s
!For sensitivity analysis. For normal model should be set at 1.0
!sens = 1.0
!jan = 375000._wp * sens
!feb = 3000000._wp * sens
!mar = 1800000._wp * sens
!apr = 1800000._wp * sens
!may = 1800000._wp * sens
!jun = 1800000._wp * sens
!jul = 1800000._wp * sens
!aug = 1800000._wp * sens
!sep = 1800000._wp * sens
!oct = 1800000._wp * sens
sfuture = 0.2
        do i = 1, days
    !SET SUSCEPTIBLE POPULATION MANUALLY
                if (i == 37) then
                    y(1) = 0.322_wp
                end if
                if (i == 74) then
                    y(1) = 1.2_wp
                end if
                if (i == 81) then
                    y(1) = 0.6_wp
                end if
                if (i == 89) then
                    y(1) = .4_wp
                end if
                if (i == 164) then
                    y(1) = 0.3_wp
                end if
                if (i == 180) then
                    y(1) = .2_wp
                end if
               if (i == 218) then
                   a = 0.35_wp
                end if
                if (i == 245) then
                    y(1) = .25_wp
                end if
                if (i == 266) then
                    y(1) = .31_wp
```

```
                end if
!Sensitivity Analysis (RECORDED DATA STOPS HERE FOR PROJECT)
               if (i == 280) then
                   !a = 0.42_wp
                   !b = 0.084_wp
                end if
!Vaccinations end of december
                if (i == 314) then
                    y(3) = y(3) + jan/poptotal
                    y(1) = jans
                end if
!Vaccinations end of Jan
                if (i == 345) then
                    y(3) = y(3) + feb/poptotal
                    y(1) = febs
                end if
!Vaccinations end of Feb
                if (i == 373) then
                    y(3) = y(3) + mar/poptotal
                    y(1) = mars
                end if
!Vaccinations end of March
                if (i == 404) then
                    y(3) = y(3) + apr/poptotal
                    y(1) = aprs
                end if
                if (i == 434) then
                    y(3) = y(3) + may/poptotal
                    y(1) = mays
                end if
                if (i == 465) then
                    y(3) = y(3) + jun/poptotal
                    y(1) = juns
                end if
                if (i == 495) then
                    y(3) = y(3) + jul/poptotal
                     y(1) = juls
                end if
                if (i == 526) then
                    y(3) = y(3) + aug/poptotal
                    y(1) = augs
                end if
                if (i == 557) then
                    y(3) = y(3) + sep/poptotal
                    y(1) = seps
```

```
                end if
                if (i == 588) then
                    y(3) = y(3) + oct/poptotal
                    y(1) = octs
                end if
                if (i == 618) then
                    y(3) = y(3) + nov/poptotal
                    y(1) = novs
                end if
                if (i == 648) then
                    y(3) = y(3) + dec/poptotal
                    y(1) = decs
                end if
                if (i == 678) then
                    y(3) = y(3) + jan1/poptotal
                    y(1) = jan1s
                end if
    !Call RKF (tinit,tf,y,esp1,esp2,h,hmin,hmax,derivL,exitf,n)
                Call RKF (tinit,tf,y,eps1,eps2,h,hmin,hmax,SIR,exitflag,n,a,b)
                if ((0.88832_wp*(y(2)+y(3))*poptotal)>26473189.4_wp) then
                    write(*,*) "i = ", i
                end if
                susceptible(i) = y(1)
                infected(i) = y(2)
                removed(i) = y(3)
                !write(15,*) i, ",", y(1),",",y(2),",",y(3),",",y(1)+y(2)+y(3)

                tinit = tf
                tf = tf + 1._wp
        end do
    do i = 1, days
        !write(15,'(1000(f12.0))') yprint(:n,i)*totalpopulation
        !write(15,808) i, yprint(1,i)*poptotal,yprint(2,i)*poptotal,yprint(3,i)*poptotal
        write(15,808) i, susceptible(i)*poptotal, infected(i)*poptotal, removed(i)*poptotal
        808 Format (i6,",",f20.4,",",f20.4,","f20.4)
    end do
End Program covid19
Function SIR (t,y,n,a,b)    !Home of the SIR Equations
    use types
    implicit none
!Declare incoming variables
    Real(wp), Dimension (:), Intent(in) :: y
    Real(wp), Intent (in) :: t,a,b !Days
    Integer, Intent (in) :: n
!Declare Local Variables
```

```
    Real(wp), Dimension (n) :: SIR !n dimension array for storing and returning values to RKF
!    Real(wp) :: a = 0.23_wp   !Declare parameter of initial exponential growth of Infected population
!    Real(wp) :: b = 0.07_wp   !Declare parameter of exponential decay of Infected population
!Write (*,*) "y(1) = ", y(1)
!Write (*,*) "y(2) = ", y(2)
!Write (*,*) "y(3) = ", y(3)
!Calculate slopes
    SIR(1) = -a*y(1)*y(2) !change in  Susceptible population over change in time
    SIR(2) = a*y(1)*y(2)-b*y(2)  !change in Infected population over change in time
    SIR(3) = b*y(2)       !change in Removed population over change in time
!Write(*,*) " SIR = ", SIR
End Function SIR
```