**Infectious Disease Intervention Analysis; Interactive visualization of infectious disease models**

Ashlynn Daughton

University of Colorado, Boulder

Ashlynn.daughton@colorado.edu

**ABSTRACT**

Models are often used by public health practitioners to forecast the possible courses of infectious diseases. Compartmental models are one flavor that group individuals into compartments based on their disease status. They are computationally inexpensive, commonly used for first pass explorations of emerging outbreaks, and rely on only a handful of differential equations to run. However, there are no freely available interface to allow those in public health to explore those models without learning differential equations, or developing this programming background required to build the models. This work shows a web-based interface that allows a user to interact with a common compartmental model in a number of ways that facilitate data exploration, and decision making.

**1.0 Introduction & Related Work**

Despite substantial public health improvements in the last century, infectious diseases remain one of the leading causes of both morbidity and mortality [1-2]. During an infectious disease outbreak, standard protocol is to perform assessments, analyze surveillance data, identify resources and interact with subject matter experts to identify and implement appropriate control measures [2,3]. These approaches rely heavily on good surveillance systems, access to experts, and good intuition about which control measures to use. As such, they are largely subjective and time consuming, and the infrastructure required is often not present in high disease burden areas.

Modeling is an attractive supplemental method. Compartmental models are commonly used, historically as a first pass attempt to characterize outbreaks or infections quickly [4]. They divide individuals into categories based on their disease status. The most common variant is the SIR model, named after the categories used—“susceptible”, “infectious” and “recovered”. However, there is a systematic lack of collaboration between the researchers developing models and those making policy decisions during an outbreak that results in a disconnect between public health and policy [5,6].

Specific work on design for public health has been generally limited. Sedig et al. are a notable exception and focus specifically on visualizations for the public health analyst community [8]. This work focuses largely on what visualizations should enable, rather than what the specific visualizations ought to be. They find that the goals of data interaction and exploration are prominent, and emphasize that it is this interaction with the data that is necessary for analysts to create their mental models [8]. Other common cognitive models used for decision support also emphasize the importance of interacting data to create schemas [9]. When these schemas are externalized, they manifest as data visualizations (diagrams, charts etc) [10]. Sedig et al. hypothesize that by ”offloading” this information, the mental models required to make decisions are more easily refined, which may improve the ability to make decisions [8].

**2.1 Designing for public health**

The design of this application specifically targeted two goals. The first was to allow a user to interact with a simple model that might be useful for decision support. The second was to support exploration within that model, rather than projecting one outbreak path. The hope is that this exploration allows the users to (1) understand the effects of particular parameters and (2) gain an intuitive feel for the level of uncertainty present in the model.

**2.1.1 SIR Models**

The specific model used needed to be simple, and commonly used in the domain. As such, the simple SIR model was selected. This model is named from the compartments individuals are groups into (‘Susceptible’, ‘Infectious’ and ‘Recovered’) based on their disease status. It is governed by three differential equation (see equations 1-3), which use two parameters. β describes the force of the infection (similar to how quickly an outbreak spreads), and 𝛾 is related to the infectious period of the disease.

(1)  (2)  (3) 

**2.2 Supporting data exploration**

Previous work (e.g., [7,8]) focus on the importance of data exploration in decision making. Thus, designing ways to support this exploration within the application was key. In addition, however, it is difficult to describe uncertainty in these models in a way that doesn’t create information overload. By showing the influence of a range of parameters rather than a single outbreak trajectory, the visualizations support exploration and an intuitive feel for the effects of small perturbations (e.g., uncertainty) in the model (akin to [11]).

**3.0 The application**

Given the related work described above, the application was designed with a focus on data exploration, and manipulation. As described, the underlying model and metrics used to assess control measures had been developed previously. However, the interface designed was difficult to use, clunky, and prone to error. As such, several improvements were identified. They are as follows:

1. Improved user interface such that forms were small, terms used were generic, or defined, and errors filling out forms were handled appropriately.
2. Addition of visualizations that allow the user to interact with (1) the parameters specified by the user in the form, as well as (2) visualize how slight perturbations of the parameters affect the overall outbreak.

The final product is shown in Figure 1. The left side of the application shows the parameters required to run the model: the range of infectiousness period, range of reproductive number, control measure effectiveness and the control measure start. The right side then allows the user to interact with the model in three specific, but unique ways (see below).

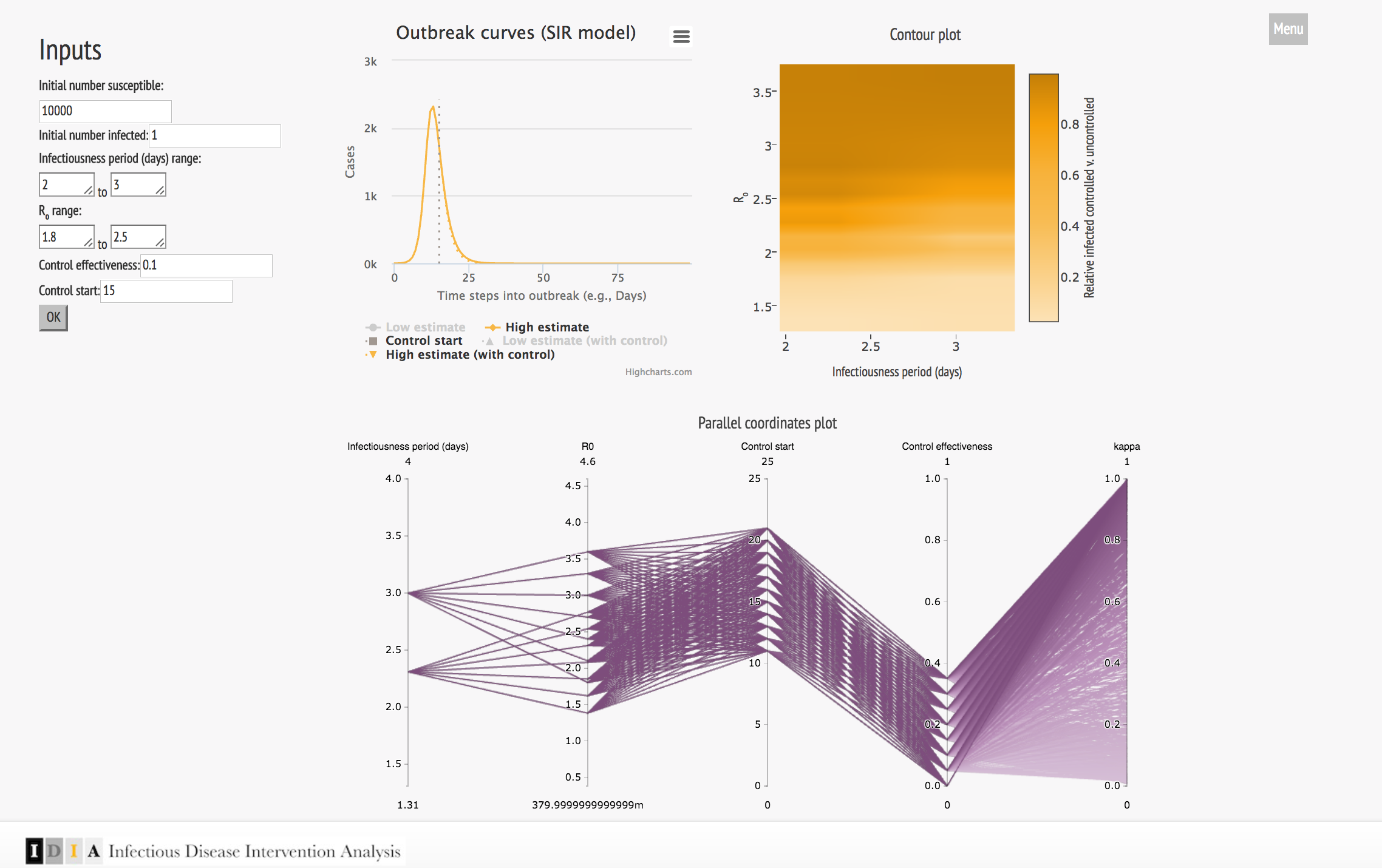


Figure 1: Screenshot of application. The left side shows the form for user’s to input the parameters required to run the model, while the right side shows application output. Three graphs are shown: A epidemic curve (familiar to many in public health) that compares controlled and uncontrolled outbreaks, a contour plot that shows how controlled and uncontrolled outbreaks vary across the user’s parameter ranges, and a parallel coordinates plot that allows a user to see how perturbations in the model parameters affect outcomes.

**3.1 Design considerations**

Application output was designed to be easy to digest, while also allowing the user to interact with the model in specific ways. Specific attention was paid to colors, which were chosen with the goal of maximizing interpretability, but also cohesion, such that the final product looks nice.

**3.1.1 Epidemic curve**

The first graph (top left) shows the epidemic curve. These curves are familiar to those in public health, and show the number of cases at each time point (days in this context). There are four epidemic curves shown: the largest possible outbreak with and without a control, and the smallest possible outbreak with and without a control. These curves are calculated based on the smallest and largest possible outbreaks given the values provided by the users. Then, if given, corresponding curves show the effect of the control measure. This graph was largely intact from the initial iteration of the application. However, I added a horizontal line that shows when the control is initiated, and modified the colors such that there is a cohesive color scheme.

**3.1.2 Contour plot**

Because the user can enter a range of parameters, and because those parameters interact in interesting ways in the model, it was also important to see how a given control measure might affect the epidemic curve across the range of parameters given. This contour plot shows the relative number of infected in a controlled outbreak compared to the number infected in an uncontrolled outbreak, across the entire set of parameters given by the user (i.e., this shows all the curves in between the two large epidemic curves in 3.1.1, and provides some analysis of the actual impact of the control measure). The plot was created using *plotly.js*, and includes a tooltip, panning, zooming, and exporting features.

**3.1.3 Parallel coordinates plot**

In order to explore the effective of perturbations on the parameters, the parallel coordinates plot is shown. Here, a range around the minimum and maximum parameter values is shown. This allows the user to investigate potential hypotheses (e.g., ‘If I implement the control measure 3 days sooner, how much more reduction in case count can I achieve?’). The plot was also created using *plotly.js*, and allows a user to select specific ranges of interest on each axis, as well as reorder axes as necessary.

**3.2 Application validation, A case study**

To explore the use of this tool, I conducted a small case study. For this example, I used a case study from the Centers for Disease Control and Prevention, that is commonly used for students learning about epidemiology.[[1]](#footnote-1) It describes an outbreak of measles in Texarkana, a city that lies on the Texas-Arkansas border.

Using this case study, I wanted to see if:

1. The data required to run the model would be available
2. The application would allow hypothesis generation about possible control measures
3. The application could help facilitate decision making

Results were as follows:

1. The case study indicates that there are roughly 50,000 in the population (the number initially susceptible). Over the course of one week, 295 cases of measles were diagnosed (number infected). The other parameters were not explicitly available in the case study. I had to use other work [FIXME-CITE], to identify the related parameters. Further, there is no mention of control measures in the case study.
2. I investigated the use of a control measure that was 10% effective, that occurred 7 days after the health department was notified of the 295 cases. The model indicated that if the higher range parameters were correct, the control measure would be completely ineffective. According to the contour plot, if the reproductive number was under 4, the above control measure would reduce the outbreak by at least 50%.
3. As might be expected, the parallel coordinates plot indicated that later control starts would require higher control effectiveness (and low reproductive numbers) to substantially reduce the outbreak.

**4.0 Discussion**

Overall, the case study does indicate that the application might provide a useful way of interacting with a compartmental model. However, there are several limitations:

1. During the case study, I realized the first graph is too small and is difficult to read.
2. It is unclear if a public health person will have all the data required to run the model, as not all values were reported in the case study. However, the values required for the model are commonly known by experts in the field. Thus, the extent to which this is problematic requires additional work with more field users.
3. If the parameter ranges are extremely broad, my python server crashes. Thus, there are clearly performance issues to be dealt with.
4. The parallel coordinates plot is interesting, but also quite difficult to interpret. Perhaps more disparate colors would be more useful here.

**5.0 References**

1. Lopez AD. *et al*. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *The Lancet* 367, 1747–1757 (2006).

2. Abdallah S. and Panjabi R. Control of communicable diseases. In *Public health guide in emergencies*, 284–369 (Johns Hopkins Bloomberg School of Public Health, 2008), 2 edn.

1. Frieden TR *et al*. A CDC framework for preventing infectious diseases - Sustaining the essentials and innovating for the future. Tech. Rep., Centers for Disease Control and Prevention (CDC) (2011).
2. Brauer F. *et al*. (eds.) *Mathematical Epidemiology*, vol. 1945 of *Lecture Notes in Mathematics* (Springer Berlin Heidelberg, 2008).
3. Wagner WE *et al*. Misunderstanding Models in Environmental and Public Health Regulation. SSRN Scholarly Paper ID 1711766, Social Science Research Network (2010).
4. Margevicius, KJ *et al*. The Biosurveillance Analytics Resource Directory (BARD): Facilitating the Use of Epidemiological Models for Infectious Disease Surveillance. *PLOS ONE* 11, e0146600 (2016). DOI 10.1371/journal.pone.0146600
5. Daughton AR *et al*. An approach to and web-based tool for infectious disease outbreak intervention analysis. Scientific Reports. 2017.
6. K. Sedig and others, “Beyond information access: Support for complex cognitive activities in public health informatics tools,” Online Journal of Public Health Informatics, vol. 4, Dec. 2012.
7. G. Klein, “A recognition-primed decision (RPD) model ofrapid decision making,” in Decision Making in Action: Models and Methods (G. Klein and others, eds.), pp. 138–147, Norwood: Ablex Publishing Corporation, 1993.
8. P. Pirolli and S. Card, “The sensemaking process and leverage points for analyst technology as identified through cognitive tasks analysis,” 2005.
9. Krause, J., Perer, A. & Ng, K. Interacting with Predictions: Visual Inspection of Black-box Machine Learning Models. in 5686–5697 (ACM Press, 2016). doi:10.1145/2858036.2858529

1. Available at: https://www.cdc.gov/eis/downloads/xtexark-711-903-student.pdf [↑](#footnote-ref-1)