



28 Models Later: Model Competition and the Zombie Apocalypse

Ian McGahan¹  · James Powell¹ · Elizabeth Spencer²

Received: 20 April 2020 / Accepted: 9 December 2020 / Published online: 16 January 2021

© The Author(s), under exclusive licence to Society for Mathematical Biology 2021

Abstract

Between Fall 2011 and Fall 2012 students at Utah State University played several rounds of Humans versus Zombies (HvZ), a role-playing variant of tag popular on college campuses. The goal of the game is for the zombies to tag humans, converting them into more zombies. Based on portrayals of ‘zombieism’ in popular culture, one might treat HvZ as a disease system. However, a traditional SIR model with mass-action dynamics does a poor job of modeling HvZ, leading to the natural question: What mechanisms drive the dynamics of the HvZ system? We use model competition, with Bayesian Information Criterion as arbiter, to answer this question. First, we develop a suite of models with a variety of transmission mechanisms and fit to data from fall 2011. We use model competition to determine which model(s) have the most support from the data, thereby offering insight into driving mechanisms for HvZ. Bootstrapping is used to both assess the significance of individual mechanisms and to determine confidence in the performance of our models. Finally, we test predictions of the best models with data from fall 2012. Results indicate that through both years of the game humans tend to cluster defensively, zombies tend to hunt in groups, some zombies are more proficient hunters, and some humans leave the game.

Keywords Model competition · Mathematical ecology · Infectious disease modeling · Predator prey modeling

National Science Foundation through the Division of Mathematical Sciences, Grant DMS-1614526.

✉ Ian McGahan
ian.mcgahan@aggiemail.usu.edu

¹ Department of Mathematics and Statistics, Utah State University, Logan, USA

² Department of Statistics and Applied Probability, University of California Santa Barbara, Santa Barbara, USA

1 Introduction

Humans versus Zombies (HvZ) began as a moderated game of tag at Goucher College in 2005 (Lewis and Powell 2016) and divides players into two groups, humans and zombies. Zombies attack humans and humans fight off zombies with Nerf guns and sock bombs, which are balled up socks that humans throw at zombies. Zombies hit with a Nerf gun or sock bomb are ‘stunned’ and cannot continue playing for 15 min. Successful attacks result in conversion from humans to zombies. A zombie who does not tag a human in a specified time period starves and is eliminated from play. Since its rise in popularity, the game has become more organized, with added rules for player safety (chiefly the removal of Nerf guns), an organizing body to ensure fairness, and other changes to keep the game dynamic. Throughout these changes, the underlying game play of zombies tagging humans to make more zombies and starving if they fail to tag a human remained. In this paper, we seek to offer insight into the driving mechanisms for the dynamics of humans and zombies in HvZ.

Mechanistic ordinary differential equation (ODE) models have long been used to gain understanding of drivers of population dynamics. ODEs have shown particular usefulness in both disease and predator prey systems. The portrayal of zombies in popular zombie films like *28 Days Later* suggests a disease model, like the Susceptible-Infected-Recovered (SIR) model of Kermack and McKendrick (1927), may be well-suited for modeling HvZ. Additionally, the population of players is readily split into susceptible (human) and infected (zombie) compartments with players following a disease-like progression between the two compartments. Compartmental ODE models have been used to model diseases like dengue fever (Stolerman et al. 2015), cholera (Tuite 2011; Tien and Earn 2010), ebola (Diaz 2017), Gambiense sleeping sickness (Rock 2018), SARS (Meyers 2005), STDs (Eames and Keeling 2002), H1N1 (Laguzet and Turinici 2015) and, most recently, the ongoing COVID-19 pandemic (Bertozzi 2020). This compartmental framework has also been used to simulate zombie outbreaks (Munz 2005). However, these disease models are essentially variants of Kermack and McKendrick’s compartmental SIR model as the underlying mechanisms for transitioning from one compartment to another generally depend on random, well-mixed contact between compartments.

Recently, the assumptions about how to model transmission have come into question (Hopkins 2020; Begon 2002; McCallum et al. 2001D), with theoretical results outnumbering empirical datasets. When susceptible population density is constant, model selection is known (Begon 2002; McCallum et al. 2001D). Unfortunately, it is often unclear as to whether assumptions about population size and density are appropriate. This is particularly true for spatially variable populations, e.g., wildlife. Generally speaking, modeling responses tend to be adding more compartments to the model, with mass action mechanisms within compartments, resulting in models with many dependent variables and only very simple interactions. We propose that models with fewer dependent variables, but slightly more complex transmission terms, can be more tractable.

HvZ offers an excellent venue to clearly illustrate techniques of model selection in the presence of uncertain prevalence and mechanism data since the game, and thus transmission pathway is well-understood, and high-quality data are available.

The nature of the game ensures that human population density is not constant, either spatially or in time. Mass action models perform poorly with regard to HvZ data, making a clear case for the need to consider alternative models for transmission and allow the data to select among possibilities. For this particular system, since zombies in HvZ are actively spreading the disease by hunting for humans, we consider a variety of predator–prey mechanisms to model emergent population dynamics.

Inspired by the seminal works of Lotka (1925), many mechanisms have been proposed to describe a variety of predator–prey interactions. The Lotka–Volterra model uses a mass action predation mechanism, like the transmission mechanism in the Kermack–McKendrick SIR model. Holling (1959) derived functional responses reflecting time lost handling prey. Beddington (1975) and DeAngelis et al. (1975) extended this work to incorporate time lost due to predator interference. May (1978) further introduced a mechanism to account for spatial distribution of populations without the complexity of partial differential equations. Kennedy and Dwyer (2018), building on Dwyer (2000), add a more complex mechanism to the SIR framework by considering variability in susceptibility as a driving mechanism for baculovirus in gypsy moth.

With a wealth of mechanisms to examine, the question becomes which, if any, of these mechanisms pertain to HvZ? Model competition, outlined by Hilborn and Mangel (1997), is a method for selecting the best model for a system, given data. Information theoretic criteria assess model performance against data, allowing comparison of multiple models. Kay et al. (2015) use model competition to assess driving mechanisms for observed predator–prey cycles between black-tailed jackrabbits (*Lepus californicus*) and coyote (*Canis latrans*) in Curlew Valley, Utah. Tien and Earn (2010) identified relevant transmission pathways for Cholera using model competition.

We use model competition, with mechanisms drawn from disease and predator–prey literature, to ascertain driving mechanisms for HvZ. For a round of the game played at Utah State University in the fall of 2011, we develop a variety of models with both predator–prey and disease mechanisms. Using model competition, with Bayesian Information Criterion (BIC) as our measure for goodness of fit, we determine which models, and therefore mechanisms, are best supported by the data. With bootstrapping, we produce distributions for both BIC values and parameter estimates to gain confidence in our results. Finally, we test predictions of the best models with data from a second round of HvZ played at Utah State University in fall 2012.

2 Humans Versus Zombies, Rules and Data

In early renditions of HvZ at Utah State University (USU), players would sign up in advance and be assigned roles as humans or zombies. Once the game started, no new players could join. A zombie needed at least one successful attack per day or it died of starvation and left play. If a group of zombies attacked a human, only one zombie earned credit for the attack. The goal for a human was to avoid zombie attacks by hiding, fleeing, or stunning zombies with Nerf guns or sock bombs to escape. Tagged humans became zombies, with successful attacks logged on a website. In later renditions of the game, humans were given missions to add storyline to the game and

force more encounters, while Nerf guns were banned for player safety and to avoid confusion with real-world firearms.

We have time-series data for both human and zombie populations collected over 80 consecutive hours of game play from a round of HvZ at USU in the fall of 2011 and a second time series from fall of 2012. In 2011, there were no missions and players were allowed the use of Nerf guns. The 2011 data, plotted in Fig. 1(top), reflect several aspects of the rules above. While the data is 80 consecutive hours of game play, these 80 h do not map directly to clock time. There were only 14 h Monday through Friday of game time and 10 h on Saturday. Hours between 9 pm and 7 am were off limits for sleep and study. Zombies could not attack in buildings or off campus. Time for zombie starvation was 24 h in game time, not real time. Data were collected through a self-reporting system in which zombies logged their attacks online. The zombie population was calculated by adding the number of successful attacks to the zombie population at the previous sampling time and subtracting the number of zombies without a successful attack in the previous 24-h period. The human population was calculated by subtracting the number of attacks from the human population at the previous sampling time. There was no mechanism for human players to register that they left the game, which they may have done because of lack of interest or the need to study. This may have artificially inflated the human population in our data. Additionally, there are some gaps in the data (approximately hours 5, 60 and 70) where the server went down and data were not recorded.

In fall 2012, missions were introduced. Missions were organized events in the game that gathered all players, humans and zombies, for a mini-game (e.g., capture the flag or defending a checkpoint). This forced a substantially higher number of contacts over a short period of time. In Fig. 1(bottom), the missions can be seen in the three jumps in the data that occur near hours 10, 25, and 39. While the duration of the game was shorter than the 2011 round, the sampling is much more frequent. Additionally, for the first several hours of the game, new players were allowed to join, causing the human population to grow before hour 10. Finally, game organizers removed humans who did not participate in the second mission (at $t = 24.5$) from both the game and the data. In Fig. 1(bottom), the effect of this removal is seen in the slide of the human population between $t = 28$ and $t = 30$ that is unmatched by growth in the zombie population over the same period. This removal provides better data as the artificial inflation of the human population possible in 2011 is less relevant in 2012.

3 Models for HvZ Population Dynamics

We consider a selection of nonlinear transmission and mortality model components to determine important mechanisms in the HvZ system, reflecting possible density-dependent mechanisms. It is possible that some players may choose to quit playing for reasons external to the game (e.g., homework). If this happened for a zombie, the loss would appear as starvation because a zombie who quits cannot register attacks and will subsequently be removed. However, human quitting is a hidden process as there is no in-game mechanism that removes inactive players from the game. (This was remedied midway through the 2012 round of HvZ, when game managers began

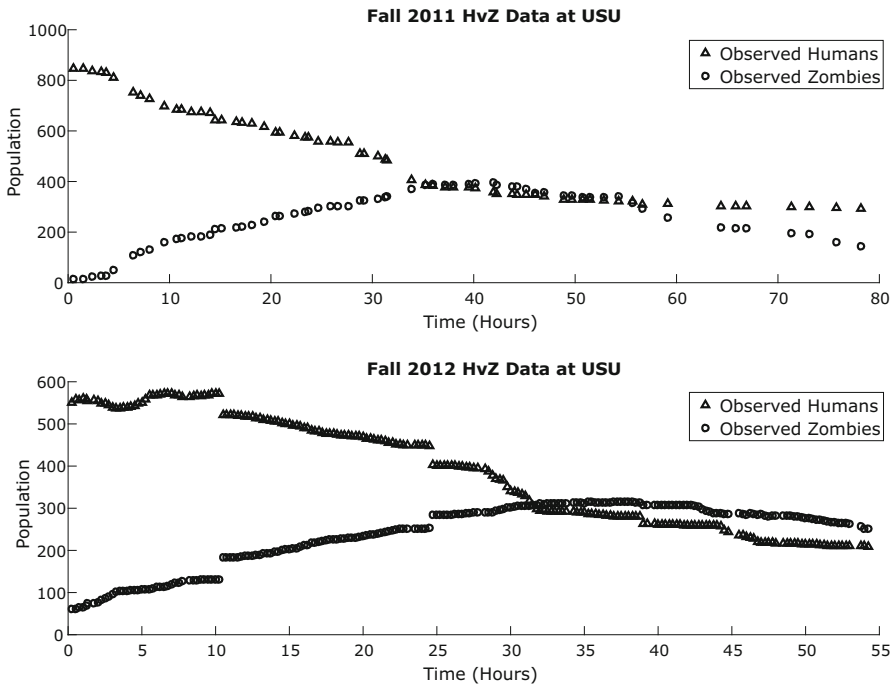


Fig. 1 Time series data from HvZ at Utah State University with triangles for human populations and circles for zombie populations. Top: Data from fall 2011. This round lasted for 80 in game hours with data recorded at intervals between 15 min and an hour and a half. There are some larger gaps in the data (approximately hours 5, 60 and 70) where the server went down, and data were not recorded. Bottom: Data from fall 2012. This round lasted for 55 in game hours with data recorded every 15 min. Missions occurred at $t = 10.25$, $t = 24.5$, 38.75 . Each mission gathered all players for a mini-game, drastically increasing the number of attacks for a short period of time, resulting in large population decreases for the humans and large increases for the zombies. Players were allowed to join as humans before the first mission, so between $t = 0$ and $t = 10.25$ the human population grew

removing inactive human players from the data set.) Thus, our data may report more humans than are actually playing the game. In order to deal with this, we add a compartment for those who have quit, resulting in the tracking of three populations: the number of humans reported in the data (H_{obs} , some of whom may have quit), the actual number of humans still playing the game (H which, as in many real-world diseases, is not observed), and the zombie population Z . The data should follow H_{obs} and Z , but H is the population that impacts the dynamics.

A model framework reflecting these mechanisms is

$$\begin{aligned}
 \dot{H}_{obs} &= F(H) - A(H, Z) \\
 \dot{H} &= F(H) - A(H, Z) - Q(H) \\
 \dot{Z} &= A(H, Z) - S(H, Z),
 \end{aligned}
 \tag{1}$$

where $F(H)$ is a human population growth term reflecting players joining the game late, $A(H, Z)$ reflects the attack mechanism, $S(H, Z)$ accounts for the starvation mechanism, and $Q(H)$ models the quitting mechanism. If there is little or no quitting, $Q(H) = 0$ and $H = H_{\text{obs}}$. Quitting is only a function of humans as zombies who quit will starve anyway. We consider seven possibilities for A , two for S , and one quitting mechanism in addition to no quitting ($Q = 0$). The twenty-eight models, consisting of all combinations of these mechanisms, compete to determine which mechanisms drive HvZ dynamics.

3.1 Zombie Attack Mechanisms

Zombie attacks dictate the transition from humans to zombies. The driving assumption for the SIR model is equal probability for any zombie and any human to have an encounter at any given time. However, this assumption may prove inaccurate in HvZ. Following, we consider several other attack mechanisms with the mass action attack rate of the SIR model as a tacit null hypothesis.

Random Encounters: Mass Action Simple assumptions about attacks are that they occur randomly, zombies are always hunting, and every zombie has the same probability of a successful attack. These assumptions were used by Lotka (1925) to model predator–prey attack dynamics. Analogous assumptions about disease transmission were used by Kermack and McKendrick (1927) to produce an identical model component for disease spread in their SIR model. An equation reflecting this mechanism is

$$A(H, Z) = aHZ, \quad (2)$$

where a is the attack rate between humans and zombies.

Zombie Time Budget Model Limiting Predation Mass action assumes zombies are always hunting, which may not be the case. Zombies spend some of their time hunting, but also spend time interacting with both humans and other zombies, potentially resulting in saturable consumption and competitive time lost. A model for saturable consumption, initially derived by Holling (1959), assumes each attack requires time for the predation interaction, termed handling time. In HvZ, this handling time can be interpreted as recovery time for a zombie after successfully chasing down and tagging a human, time needed to record the successful attack, and time for socializing. Beddington (1975) and DeAngelis et al. (1975) both derived a very similar model to Holling’s for interaction time between predators. In their derivation, this interaction time is ascribed to competition between predators. In HvZ, many zombies work together attacking the same human but only one may register the kill. We interpret this interaction time as time lost in a group hunt not registering in a kill. Following Beddington’s derivation (provided in “Appendix A”), a model for saturable consumption and competitive time lost is

$$A(H, Z) = \frac{aHZ}{1 + at_h H + a\beta Z}, \quad (3)$$

where a is the encounter rate, t_h is handling time, and β is proportional to time lost per zombie in group hunting.

If $\beta = 0$, i.e., zombies are not hunting in groups, then (3) reduces to Holling's type II model,

$$A(H, Z) = \frac{aHZ}{1 + at_h H}. \quad (4)$$

If $t_h = 0$, i.e., there is no time lost handling a human, then (3) reduces to

$$A(H, Z) = \frac{aHZ}{1 + a\beta Z}. \quad (5)$$

Spatial Clustering of Susceptibles: Negative Binomial Both mass action and the time budget models assume well-mixed systems, i.e., spatial homogeneity of both populations. Observation of game play shows that humans tend to gather in groups, violating the homogeneity assumption. May (1978) used a negative binomial to model spatial clustering of a host in a host-parasitoid system. The model describes implicit spatial dynamics without the use of partial differential equations. May suggested this, phenomenologically, in a discrete time model, but the same assumptions have been adapted to continuous time models (Kong et al. 2016; McCallum et al. 2001D) (derivation provided in "Appendix C.1"). In terms of HvZ, we have

$$A(H, Z) = k_A \ln \left(1 + \frac{aZ}{k_A} \right) H, \quad (6)$$

where k_A is a measure of the degree to which humans cluster. Large k_A corresponds to less grouping, limiting on spatial homogeneity. Small k_A correspond to highly clumped groups, i.e., few, small pockets of high density for humans. As with previous models, a is the encounter rate.

Variability in Susceptibility: The Theta-Logistic Model Another assumption of mass action is that all humans have an equal likelihood of being attacked by a zombie. Since humans can run away from zombies, and humans have different athletic abilities, some humans may be more likely to fall victim to zombie attack. This variability in susceptibility to zombie attack could alter attack dynamics. Kennedy and Dwyer (2018) showed variability in the susceptibility of healthy individuals plays a role in disease propagation in host-pathogen interactions. Assuming some distribution of susceptibility for humans, with V as the square of the coefficient of variation, Dwyer (2000) derive an equation for expected attack rates,

$$A(H, Z) = a \left(\frac{H}{H_0} \right)^V HZ, \quad (7)$$

where a is the attack rate and H_0 is the initial human population. A derivation is provided in “Appendix B”. If $V = 0$, then there is no variation in human susceptibility, reducing to mass action (2). On the other end of complexity, using (7) and the time budget of 3, we can incorporate the mechanisms variable susceptibility, time lost from group hunting, and handling time in the equation

$$A(H, Z) = \frac{a \left(\frac{H}{H_0}\right)^V H Z}{1 + at_h H + a\beta Z}. \tag{8}$$

3.2 Mechanisms for Zombie Starvation

In the HvZ game, zombies die of starvation when they fail to attack a human over the course of a day. Thus, zombie starvation is directly a result of the duration between successful attacks. In the SIR model, infected individuals are assumed to recover (or die) in the same average amount of time. Translated to the mechanics of HvZ, this assumes all zombies have the same likelihood of having no attacks in 24 h. However, if zombies have varying abilities to successfully attack humans, then starvation may not be equally likely. We consider two starvation mechanisms, proportional death and uneven attack distribution, with the proportional death mechanism of the SIR model as the null hypothesis.

Starvation Equally Likely for All Zombies: Proportional Starvation If starvation is equally likely for all zombies at any time, then the per capita rate of starvation is constant. Thus,

$$S(H, Z) = mZ, \tag{9}$$

where m is the per capita rate of starvation. This is also the recovery (or removal) mechanism of infected individuals for the SIR model.

Uneven Attack Distribution: Negative Binomial Model If hunting ability varies among zombies, then attacks are not evenly distributed among zombies and starvation is not equally likely. We modify May’s use of the negative binomial to model the impact of an uneven distribution of attacks on starvation with

$$S(H, Z) = m \frac{Z}{\left(1 + \frac{r_S H}{k_S}\right)^{k_S}}, \tag{10}$$

where r_S is related to the ratio between the mean number of attacks per zombie and the maximum number of attacks per zombie and k_S measures ‘clumpiness’ of attacks. A low k_S indicates some zombies get a disproportionately high number of attacks, i.e., a small number of zombies is responsible for a majority of the attacks on humans. Conversely, a large k_S means attacks are more evenly distributed. Although this starvation mechanism is based on a negative binomial distribution of events, like the spatial clustering attack mechanism (Sect. 3.1), the functional form is different. An explanation is provided with a derivation of this starvation mechanism in “Appendix C.2”.

3.3 Humans Joining or Quitting

In the fall 2011 round of HvZ, no players were allowed to join after the game started. Thus, we set $F(H) = 0$ for all of the 2011 models. In the fall 2012 round, new players were allowed to join until the first mission at $t = 10.25$, with new players added as humans. To reflect this, we use

$$F(H) = f. \tag{11}$$

The growth rate for humans, f , is set to the slope of the secant line for the total population (sum of humans and zombies) over the interval when humans were allowed to join the game, between $t = 0$ and $t = 10.25$. After the first mission, we set $F(H) = 0$.

To model human leaving the game, we consider a proportional quitting mechanism,

$$Q(H) = qH, \tag{12}$$

where q is the rate humans quit the game. The null hypothesis, that human quitting does not impact game dynamics, tacitly sets $q = 0$.

4 Parameter Estimation and Model Comparison

4.1 Maximum Likelihood Estimation for Parameters

Once we have a collection of possible models, we want to identify which models represent the data best from an information theoretic perspective. Model fit is based on Maximum Likelihood Estimation (MLE), popularized by Fisher (1920, 1922). For a given vector of parameters, θ , let $H_{\text{obs}}(t, \theta)$, $H(t, \theta)$, and $Z(t, \theta)$ describe the solution to (1). For a set of data with N observations, the relationship between the solution and the n th observation ($1 \leq n \leq N$), (H_n, Z_n) , is defined as follows. Let H_n be a normally distributed random variable with mean $H_{\text{obs}}(t_n, \theta)$ and variance σ^2 . Similarly, let Z_n be normally distributed with mean $Z(t_n, \theta)$ and variance σ^2 . Since the total initial population, P_0 , is fixed and $H = P_0 - Z$, the variance for both H and Z are the same. Thus, the probability of observing a particular set of data can be expressed via the density functions

$$f(H_n|\theta, t_n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(H_n - H_{\text{obs}}(t_n, \theta))^2}{2\sigma^2}} \tag{13}$$

and

$$g(Z_n|\theta, t_n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Z_n - Z(t_n, \theta))^2}{2\sigma^2}}. \tag{14}$$

The likelihood, \mathcal{L} , over the set of N data points is therefore

$$\mathcal{L}[\theta|\text{Data}] = \prod_{n=1}^N \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Z_n - Z(t_n, \theta))^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(H_n - H_{\text{obs}}(t_n, \theta))^2}{2\sigma^2}} \right). \quad (15)$$

The negative log likelihood (NLL) of θ given the data is

$$\text{NLL}[\theta|\text{Data}] = N \ln(2\pi\sigma^2) + \sum_{n=1}^N \frac{(Z_n - Z(t_n, \theta))^2 + (H_n - H_{\text{obs}}(t_n, \theta))^2}{2\sigma^2}, \quad (16)$$

which is minimized where parameters are most likely given the data (Lewis and Powell 2017).

4.2 Bayesian Information Criterion for Model Competition

Since an arbitrarily complex model can be tuned to fit complicated data, goodness of fit alone is a poor metric for model comparison. Bayesian Information Criterion (BIC), derived by Schwarz (1978), gives a way for determining the significance of one model over another given a particular data set by balancing goodness of fit with model complexity. BIC is given by

$$\text{BIC} = 2\text{NLL}(\hat{\theta}) + (K \ln(N)), \quad (17)$$

where $\hat{\theta}$ is the set of best fit parameters for a model, K is the number of parameters, and N is the number of data points. BIC captures behavior of the NLL while penalizing model complexity. Since a lower NLL implies a higher likelihood and a simpler model is preferred, a lower BIC denotes a better model. Further, we can compare two models by their relative difference of BIC. Given models A and B , with corresponding BIC_A and BIC_B , if $\text{BIC}_A < \text{BIC}_B$ then A is $e^{\frac{1}{2}\Delta\text{BIC}}$ times more probable than B (Neath and Cavanaugh 2012). The standard threshold of significance is $\Delta\text{BIC} = 5$ as this says model A is about $e^{\frac{1}{2}(5)} > 10$ times more likely than model B .

4.3 Bootstrapping and Uncertainty Quantification

Bootstrapping is a method for producing multiple sets of data by resampling available data with replacement, generating distributions for parameter values and BIC and quantifying uncertainty in results (Dixon 2006). Since our HvZ data are time series, we modify the method outlined by Dixon (2006) to preserve the ordering of samples as follows. We randomly draw from our data with replacement to produce a data set with the same number of data points as the fall 2011 data set and track the frequency of each data point in the resample. Models are then fit to the full data set with the error for a given data point weighted higher, the more frequently that point appears in the resample.

Table 1 BIC values for fittings of all combinations of mechanisms

$A(H, Z)$	$S(H, Z) = mZ$		$S(H, Z) = m \frac{Z}{\left(1 + \frac{r_S H}{k_S}\right)^{k_S}}$	
	$Q = 0$	$Q = qH$	$Q = 0$	$Q = qH$
aHZ	-163	-414	-163	-395
$\frac{aHZ}{1+at_h H}$	-159	-410	-146	-409
$\frac{aHZ}{1+at_z Z}$	-377	-464	-529	-577
$\frac{aHZ}{1+at_h H+at_z Z}$	-373	-460	-525	-573
$\frac{a}{H_0} \left(\frac{H}{H_0}\right)^V HZ$	-439	-437	-484	-316
$\frac{a}{H_0} \left(\frac{H}{H_0}\right)^V \frac{HZ}{1+at_h H+at_z Z}$	-437	-469	-561	-568
$k_A \ln \left(1 + \frac{aZ}{k_A}\right) H$	-392	-470	-525	-586

The bold value in the upper left, BIC of -163, is the BIC for Kermack and McKendrick’s SIR model. This is the second highest BIC of all the models tried, evidence that the SIR model does a poor job of explaining the data. The four boxed values (in the rightmost column) are the four lowest BIC values, corresponding to the best models by BIC competition (SC with a BIC of -586, GH with a BIC of -577, GHHT with a BIC of -573, and GHHTVS with a BIC of -568)

We bootstrap 10,000 data sets from the fall 2011 data set and fit all models to each. Each of the proposed mechanisms has a controlling parameter, like t_h for handling time that for some value causes the mechanism to vanish. The likelihood that a mechanism is unimportant is proportional to the mass of the distribution for the controlling parameter clustered at the vanishing value. From the BIC distributions, we can determine the probability that any model is better than any other. For models A and B , we find respective BIC values BIC_A and BIC_B for the original data set. The likelihood that model B is better than model A is the proportion of the model B ’s distribution of BIC that falls below BIC_A .

5 Results

We tested 28 models consisting of every combination of attack and starvation mechanism, both with and without quitting. BIC values for each model are presented in Table 1. The bold value in Table 1 is the BIC for the null hypothesis in our model competition, namely Kermack and McKendrick’s SIR model. The SIR model, with mechanisms of mass action attack rate, proportional starvation, and no quitting, is written

$$\begin{aligned} \dot{H} &= -aHZ, \\ \dot{Z} &= aHZ - mZ. \end{aligned} \tag{18}$$

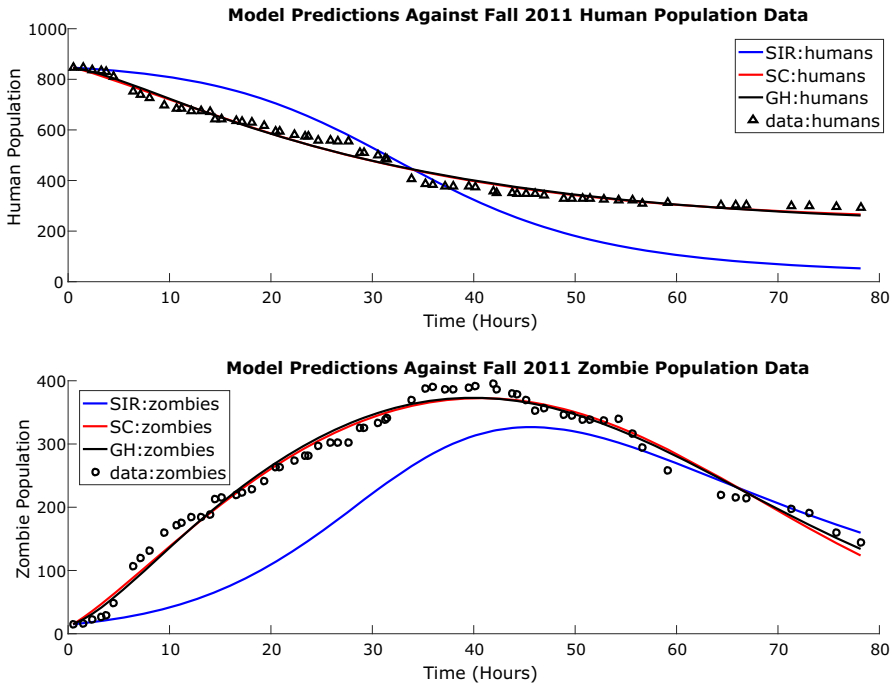


Fig. 2 Plotted are time series data from a fall 2011 round of HvZ at USU with the best fits of the SIR model (18) and the two best performing models: the group hunting model, GH, and SC, the susceptible clustering model (19). Top: Human population data (triangles) and model predictions. The SIR model (dark grey) does very poorly only matching the data at the start ($t = 0$) and the center ($t = 30$ to $t = 40$). SC (light grey) and GH (black) do significantly better at matching the data with SC performing slightly better. Bottom: Zombie population data (circles) and model predictions. The SIR model (dark grey) recovers some of the data at the end ($t = 57$ to $t = 80$) but does a poor job of capturing the dynamics overall. The SC (light grey) and GH (black) models match the data well with SC again slightly outperforming GH

This model performed poorly in the model competition, with the second highest BIC. The difference in BIC between the SIR model and the best performing model is 423, well above the threshold of significance in BIC difference of 5 (Table 1). Deviation from 2011 data is large (see Fig. 2), and the best fit solution fails to replicate dynamic characteristics of the outbreak.

All four of the best performing models share the same starvation mechanism and quitting mechanism but differ in attack mechanism. We refer to each model by its attack mechanism. The best performing model by BIC, with a BIC of -586 (Table 1), has spatial clustering of humans (6) in the attack mechanism, uneven attack distribution (10) for starvation, and quitting (12). The resulting susceptible clustering model (SC) is written

$$\begin{aligned}
 \dot{H}_{\text{obs}} &= -k_A \ln \left(1 + \frac{aZ}{k_A} \right) H, \\
 \dot{H} &= -k_A \ln \left(1 + \frac{aZ}{k_A} \right) H - qH, \\
 \dot{Z} &= k_A \ln \left(1 + \frac{aZ}{k_A} \right) H - m \frac{Z}{\left(1 + \frac{r_S H}{k_S} \right)^{k_S}}.
 \end{aligned}
 \tag{19}$$

The best fit of model SC is plotted in Fig. 2, for comparison with SIR and 2011 data.

The second best model by BIC, with a BIC of -577 , has group hunting (5) for the attack mechanism, uneven attack distribution (10) as the starvation mechanism, and quitting (12) and is given in (20), where we set $t_h = 0$ and $V = 0$. We plot the group hunting model (GH) in Fig. 2. The third best model, with a BIC of 573 , has attack mechanisms of group hunting and handling time (3), uneven attack distribution (10) for starvation, and quitting (12). The group hunting and handling time model (GHHT) is given in (20) where we set $V = 0$. The difference between BICs of GH and GHHT is 4 (Table 1), indicating that there is not a substantial difference in modeling power between GH and GHHT. The fourth best model, with a BIC of -568 , is

$$\begin{aligned}
 \dot{H}_{\text{obs}} &= \frac{-a \left(\frac{H}{H_0} \right)^V H Z}{1 + at_h H + at_c Z} \\
 \dot{H} &= \frac{-a \left(\frac{H}{H_0} \right)^V H Z}{1 + at_h H + at_c Z} - qH \\
 \dot{Z} &= \frac{a \left(\frac{H}{H_0} \right)^V H Z}{1 + at_h H + at_c Z} - m \frac{Z}{\left(1 + \frac{r_S H}{k_S} \right)^{k_S}}.
 \end{aligned}
 \tag{20}$$

This model has attack mechanisms of group hunting, handling time, and variation in susceptibility, and starvation and quitting mechanisms identical to those of SC, GH, and GHHT. The difference in BIC between the group hunting, handling time, variation in susceptibility model (GHHTVS) and GHHT is 5 (Table 1), just at the level of significance. The only difference between models GHHT and GHHTVS is the addition of variation in susceptibility as an attack mechanism (governed by V), indicating that there may not be a substantial difference in descriptive power between any of GH, GHHT, and GHHTVS.

The distribution of parameters from bootstrapping offers additional evidence that model fit is not improved by the addition of either the handling time or variation in susceptibility mechanisms. From the distribution of t_h in GHHT (Fig. 3), we get a 90% confidence interval for handling time of (1.3, 2.8) s. While this is realistic for HvZ, the loss of a few seconds from hunting time will not have a significant impact on zombie hunting time, which is at the scale of hours. The square of the coefficient of variation, V , for the variable susceptibility mechanism in GHHTVS fits to 1.48×10^{-5} . The

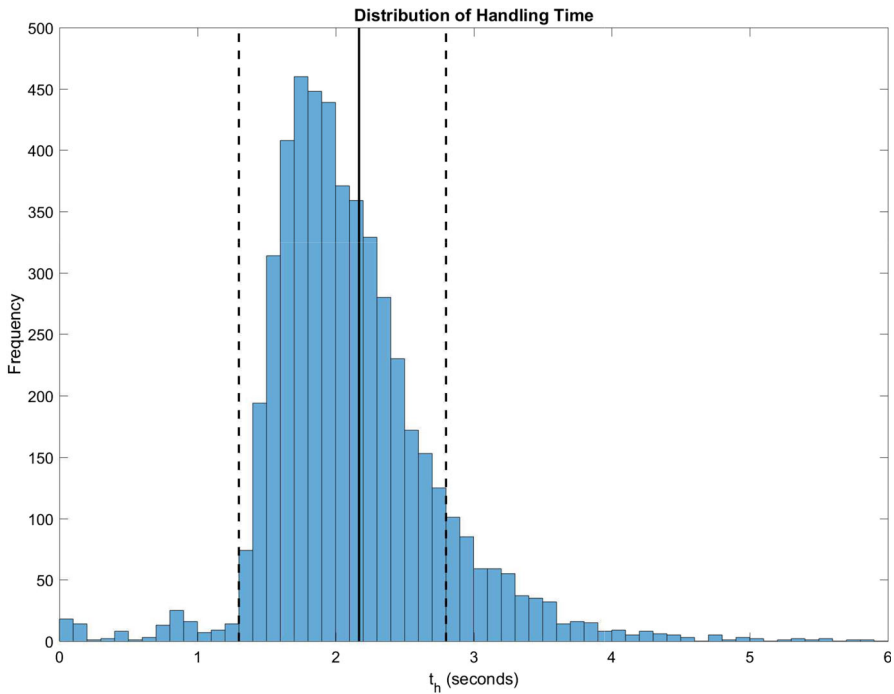


Fig. 3 Shown is the distribution for handling time, t_h , fitted to 10,000 bootstrapped data sets for the model with group hunting and handling time as attack mechanisms (GHHT). Dashed lines indicate the bounds on the 90% credible interval, (1.3, 2.8) s. The solid black line indicates the fitted value of $t_h = 2.17$ s for the fall 2011 data set. At the time resolution of the data (in hours) a loss of a couple seconds per zombie attack will not generate a significant impact on zombie hunting time

distribution of bootstrapped values for V in GHHTVS is displayed in Fig. 4. After removing outliers, we find a 90% confidence interval of $(1.1 \times 10^{-5}, 7.8 \times 10^{-5})$. Such a small coefficient of variation indicates that the level of variation in susceptibility is negligible.

All four of SC (19), GH, GHHT, and GHHTVS (20) share the same starvation mechanism of uneven attack distribution among zombies and the quitting mechanism. The starvation mechanism common to all four of the best performing models is uneven attack distribution (10). The measure of how clustered attacks are, k_S , is fit between 2.31 and 2.63 for all of the best four models indicating that there are zombies with a disproportionately high number of attacks. All four of our best models also have a common range of values for quitting rate, q , with between .007 and .008 of the human population quitting per hour. We do not consider any of the other models as the difference in BIC between any other model and model GHHTVS is at least seven, which is well over the threshold we consider for BIC competition. Full fitted parameterizations and BICs for SIR, NB, GH, GHHT, and GHHTVS models are provided in Table 2.

Figure 5 shows the distributions for the BIC of the SIR, GH, and SC models fitted to 10,000 bootstrapped data sets. The BICs for each of these models, given in Table 1, are

Table 2 BIC and fitted parameter values for the SIR model (BIC of -163), the model with uneven attack distribution (SC, BIC of -586), the model with group hunting (GH, BIC of -577), model with group hunting and handling time (GHHT, BIC of -573), and the model with group hunting, handling time, and variable susceptibility (GHHTVS, BIC of -568)

Model	BIC	Parameters									
		a ($\times 10^{-3}$)	t_h	β	V ($\times 10^{-5}$)	k_A ($\times 10^{-3}$)	m	r_S ($\times 10^{-2}$)	k_S	q ($\times 10^{-3}$)	
SIR	-163	0.181	-	-	-	-	0.0421	-	-	-	
SC	-586	1.56	-	-	-	6.09	0.121	1.19	2.55	8.00	
GH	-577	0.746	-	34.7	-	-	0.143	1.26	2.63	7.30	
GHHT	-573	0.746	2.17	34.7	-	-	0.143	1.26	2.62	7.30	
GHHTVS	-568	0.746	1.73	34.7	1.48	-	0.143	1.26	2.62	7.30	

In both of the models with a handling time mechanism, GHHT and GHHTVS, $t_h < 3$ s. The square of the coefficient of variation, V , in GHHTVS fits to 1.48×10^{-5} . This is small enough to be negligible. Given the nearly identical parameter values for common parameters between GH, GHHT, and GHHTVS, it seems likely that the difference in BIC is primarily due to the extra parameters in models GHHT and GHHTVS which do not contribute to the overall dynamics

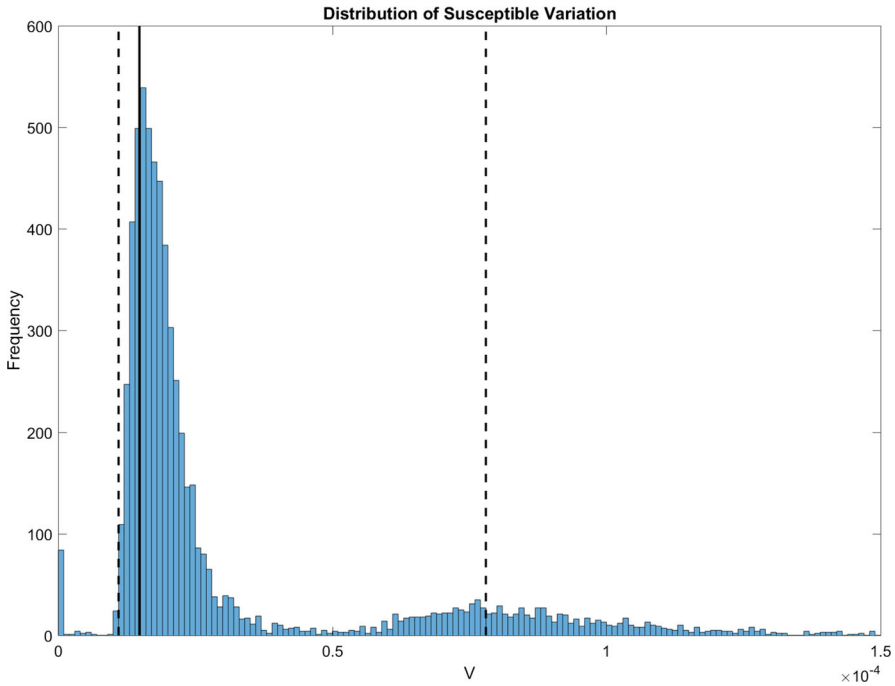


Fig. 4 Shown is the distribution for V , the square of the coefficient of variation for susceptibility fitted to 10,000 bootstrapped data sets for the model with group hunting, handling time, and variation in susceptibility as attack mechanisms (GHHTVS). Dashed lines indicate the bounds of the 90% credible interval, (1.1×10^{-5} , 7.8×10^{-5}). The solid black line indicates the fitted value of $V = 1.48 \times 10^{-5}$ for the fall 2011 data. Such a small coefficient of variation indicates that the level of variation in susceptibility is negligible

−163, −577, and −586, respectively. Discounting outliers that are the result of convergence in the MLE fitting to unreasonable parameter regimes, the means for the BIC distributions of models SIR, GH, and SC were −165, −564, and −572, respectively. This further supports that both the SC and GH models drastically outperform the SIR model, with the SC model slightly outperforming the GH model. The proportion of the BIC distribution for the SIR model that is below the BIC of either the GH or SC models is 0, meaning the SIR model cannot outperform either model for the 2011 data with a p value of $p < 10^{-4}$. Additionally, the proportion of the distribution for the GH model that is less than the BIC for the SC model is also 0, indicating that GH cannot outperform SC on the 2011 data with a p value of $p < 10^{-4}$. These p values were calculated directly from the histogram, p the fraction of bootstrapped trials of competing models that give an improvement over the nominal fit of a given model.

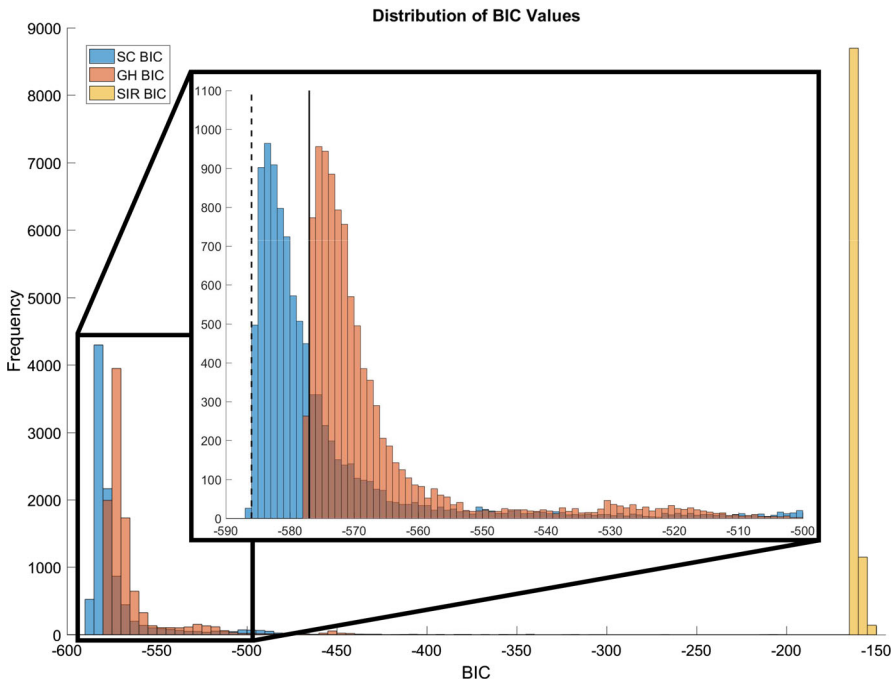


Fig. 5 Shown are the distributions for the BIC of the SIR model (18), the model with group hunting (GH), and the SC model with uneven attack distribution (19). The BICs for the fall 2011 data set of each model are -163 , -577 , and -586 , respectively. No bootstrapped BIC for the SIR model is less either of the BICs for SC or GH indicating that the SIR model cannot outperform either model for the 2011 data with a p value of $p < 10^{-4}$. The distributions for models GH and SC overlap and are shown again in the inset histogram to provide better resolution. The dashed black lines in the inset indicate the BIC for each model from the fall 2011 data set. No bootstrapped BIC for GH is less -586 , the BIC for SC, indicating that GH cannot outperform SC for the 2011 data with a p value of $p < 10^{-4}$

6 Model Predictive Power

With the results from fall 2011, we can make predictions about future rounds of the game. In order to compare our predictions with an independent data set from fall 2012 we must take into account changes in rules between 2011 and 2012. In 2012, new players were allowed to join until the first mission ($t = 10.25$). All late joining players were assigned the role of human. For this time interval, we add a human population growth term, as outlined in Sect. 3.3. Second, missions were added to HvZ at USU in 2012. There were three missions, at $t = 10.25$, $t = 24.5$, and 38.75 h (reflected by the jumps in Fig. 6). Since these events are a drastic change in game dynamics, we do not include the intervals in our models. Rather, we use our models to predict population for each of the four intervals between missions. To get the initial population for humans (zombies) at the start of each interval, we take the size in the jump in human population from before to after the mission and subtract (add) it from the final predicted human (zombie) population before the mission that started the new interval. Finally, the removal of humans who had stopped playing dampens the impact

Table 3 BIC and refitted quitting parameter for the four best models from 2011

	Models			
	SC	GH	GHHT	GHHTVS
BIC	-1486	-1250	-1445	-1291
$q(\times 10^{-3})$	2.60	8.10	0.189	7.90

Only the quitting parameter q was fit to fall 2012 data for the four best models. All unfit parameters were fixed as the best fit parameters from the fall 2011 data. For all but model GH, the quitting parameter was fit to a lower value in 2012 than in 2011. BIC indicates that model with susceptible clustering (SC) does the best at predicting 2012 data. The group hunting and handling time model (GHHT) and group hunting, handling time, and variable susceptibility model (GHHTVS) are second and third best with the group hunting model (GH) rounding out our strongest four

of quitting since the mechanism was only added to account for humans who had quit and were still in the data. Thus, we first fit a new quitting term to the fall 2012 data for each of our best models from 2011 before using these models to make predictions. We account for the jumps in the data resulting from the missions with the same method described above, fitting q to all four intervals between missions and tracking the jumps with the difference in the data from the start to the end of the mission. The quitting parameter is the only parameter fit to the fall 2012 data. All other model parameters are taken from the fall 2011 fits (Table 2). The newly fitted quitting parameters are given in Table 3.

The model with the strongest predictive power against the 2012 data is SC (19), with spatial clustering of humans (6) in the attack mechanism, uneven attack distribution (10) for starvation, and quitting (12). The quitting rate is greatly reduced (by a factor 4) between fall of 2011 and fall of 2012 for this model. Both model predictions match the human population data before the first mission and after around $t = 30$ but struggle in the interval from $t = 10.25$ to $t = 30$ (see Fig. 6). This may be the result of insufficient mechanisms to account for humans leaving the game. In the first interval the combination of humans joining and humans quitting causes some problems with identifiability. After around $t = 27$, game organizers started removing inactive players, reducing the impact of humans quitting but remaining in the data set. These two factors likely reduced the quitting term in the fitting, which may be closer to the 2011 value for the interval between $t = 10.25$ to $t = 30$. A better understanding of the impact quitting on the dynamics of HvZ is a subject for another analysis.

After SC, in order of strongest predictive power were GHHT, GHHTVS, and finally, GH. This seems to indicate that a handling time of a few seconds had a noticeable impact on model performance in 2012. This is likely due to the increased frequency in data collection from 2011 to 2012. With data collected on the order of minutes instead of hours a handling time of a few seconds may be reflected in the data. However, the difference in BIC between GHHT and GHHTVS indicates that variation in susceptibility is still not creating a significant impact on dynamics.

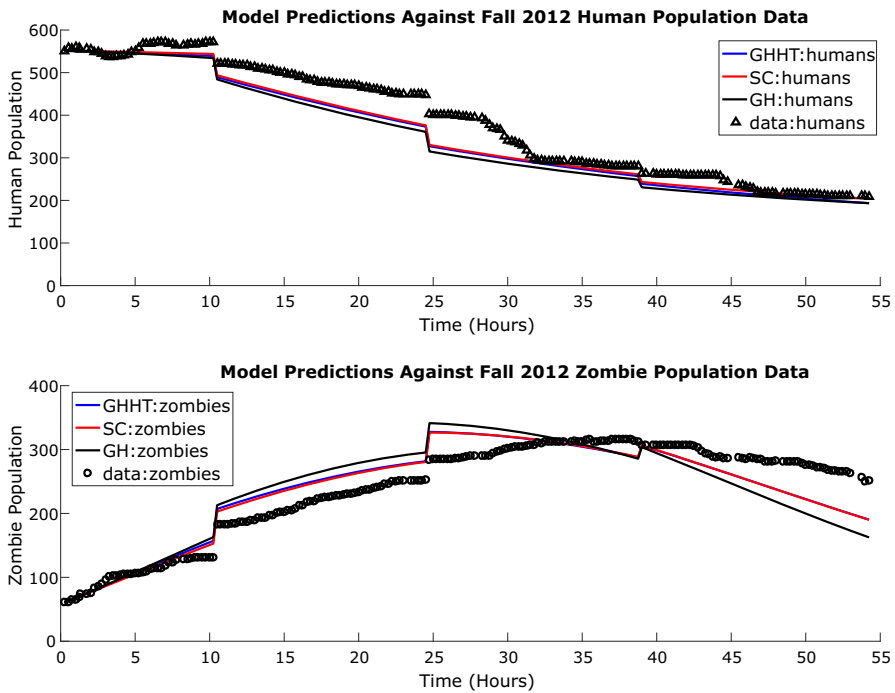


Fig. 6 Plotted are time series data from a fall 2012 round of HvZ at USU with model predictions from the best performing models in 2011. All parameters are from fall 2011 fitting (Table 2), with the exception of the quitting parameter. The missions added in fall 2012 (reflected in the jumps in the data) are accounted for by fitting around them since there none of the considered mechanisms can account for these jumps. Top: Human population data (triangles) and model predictions. The best model has attack mechanisms of spatial clustering of humans (SC, BIC of -1486) and is plotted in light grey. All three models most closely predict the data before the first mission when the growth term can offset quitting and after organizers started to account for quitting. Bottom: Zombie population data (circles) and model predictions. Of the three models considered, SC (light grey) has predictions that best align with the data

7 Discussion and Conclusions

A suite of models with a variety of contact and starvation mechanisms competed to describe the dynamics of time-series data from Humans versus Zombies. The de facto null model, Kermack and McKendrick's SIR, was among the most poorly performing models, supporting the hypothesis that more complex nonlinear dynamics drive HvZ. The best model by competition included mechanisms for spatial clustering of humans, heterogeneity in the distribution of attacks per zombie, and human quitting (model SC, Eq. 19). The next three best models (models GH, GHHT, and GHHTVS, Eq. 20) had attack mechanisms of group hunting, group hunting and handling time, and group hunting, handling time and variation in susceptibility. All three shared starvation and quitting mechanisms with the best model. The mechanisms for handling time and variation in susceptibility did not significantly improve performance for these models in comparison with the fall 2011 data. Of the 28 models we considered, those with a handling time mechanism consistently performed worse than those without for the

2011 data. However, in 2012, models with handling time generally did better than corresponding models without handling time (e.g., GHHT performed better than GH). This may be due to the improved time resolution for HvZ data in 2012. All four of the best models from 2011 struggled with predicting outcomes in 2012. We hypothesize that the removal of inactive human players from the data by game organizers after the second mission is likely a cause of this. Further analysis is required to better understand the full impacts of quitting and handling time.

The mechanisms of the best models are observable in behavior of HvZ players. Humans commonly traveled in groups for better protection, aligning with the spatial clustering mechanism of the strongest model. Group hunting occurred in multiple ways. First, when a zombie encountered a human and started chasing, other zombies in the area often noticed and joined the chase. Additionally, zombies sometimes would organize traps with a few chasing humans into a larger, hidden group of zombies. When groups of zombies chased a human the variability in speed between players gave faster zombies an advantage in tagging human players. This advantage resulted in some zombies recording significantly more attacks than others, matching the starvation mechanism common to all four of our best models. The lack of handling time can also be seen. Recording an attack only took a few seconds, and zombie players were generally eager to continue hunting and chasing humans.

In comparison to the severity of real diseases, application of model competition techniques to a game may seem trivial. While the popularity of zombies has been leveraged to provide nice educational examples with which to introduce students to epidemiological modeling (Lewis and Powell 2016; Munz 2005), this paper is not advancing modeling or understanding of detailed disease transmission. A large body of modern mathematical biology research has focused on important specifics of in-host immune responses, the influences of environmental heterogeneity, and varied effects of population structure (in age, susceptibility, infectivity, duration, lethality, etc.) on disease dynamics (Garlick 2014; Dwyer 2000; Kennedy and Dwyer 2018; Volz EM et al. 2011). Since HvZ is a game and all of these details are specified by clear rules, there is little to be learned from this paper in that regard. That said, this work highlights the need for alternative models of disease transmission, particularly when the assumptions of mass action transmission are not met. Using incorrect transmission mechanisms when modeling disease spread can have severe consequences. For example, a comparison between the predictions of the SIR model and the Fall 2011 data plotted in Fig. 2 shows the SIR model to significantly underpredict both the size and growth rate of the infected population. In the context of a real disease, such an underprediction can result in drastically underestimating its severity, leading to mismanagement and much higher societal impact (e.g., death toll).

Heterogeneity of connectivity, particularly of spatially dispersed populations, is another source of population structuring that impacts disease progress. To model spatial contact processes explicitly, researchers have used complex ODEs in patches (Tuite 2011; Tien and Earn 2010; Diaz 2017; Rock 2018; Laguzet and Turinici 2015; Meyers 2005; Stolerman et al. 2015; Kermack and McKendrick 1927; Miller 2009), partial differential equations (Turchin 1998; Pech and McIlroy 1990; Hosseini et al. 2006; O'Regan 2015; Hooten et al. 2013; Powell and Bentz 2014; Garlick 2011, 2014; Hefley Tj et al. 2017), integrodifference equations (Medlock and Kot 2003; Mollison 1972;

Kot and Schaffer 1986; Liu and Kot 2019; Wang et al. 2002), and stochastic individual based models (Chen 2018; Federico 2013; Robins et al. 2015; Gross and Miller 2001). Additionally, social networks and graph-theoretic measures have been used to characterize disease propagation indirectly (Colizza 2006; Brauer et al. 2008; Maheswaran et al. 2009; Luke and Harris 2007). To place our research in context, almost all of these approaches deal with transmission as a mass action effect on a sufficiently local scale. The blurring of boundaries between predator–prey and infected-susceptible dynamics in HvZ provides a natural avenue to tackle some of the modeling issues surrounding spatial contact processes with models of intermediate complexity. While it is difficult to deduce from first principles how the various spatial contact processes in play would result in a descriptive model, the use of model competition allowed us to clearly distinguish among potential contact mechanisms. Approaches like this could be of particular value in systems where spatial scaling is predetermined and cannot be selected to match mass action assumptions. For example, in the case of wildlife disease, the spatial scaling is often determined by the scale of landscape variation and choosing a scaling that allows for homogeneity in the population between susceptible and infected individuals is not possible.

Acknowledgements We would like to thank Michael Cortez, Luis Gordillo, and Guen Grosklos for their many helpful comments. Additionally, we thank two anonymous reviewers and the associate editor, Mark Lewis, for helpful suggestions. This work was supported by the National Science Foundation through the Division of Mathematical Sciences, Grant DMS-1614526.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

A Time Budget Model for Predation

Mass action assumes zombies are always hunting but this may not be the case. Zombies spend some amount of their time hunting, but also waste time through interactions with both humans and other zombies. There are two mechanisms here. The first, initially derived by Holling (1959), is handling time and assumes that each attack includes some interaction time between the human and zombie involved. The second, derived by both Beddington (1975) and DeAngelis et al. (1975), is predator interference and assumes that when two zombies encounter each other the interaction consumes time which detracts from hunting time. We interpret this interaction as time that multiple zombies hunt the same human. Following their derivations, we incorporate these two mechanisms by supposing, in a period of time T , each zombie spends its time divided between time spent searching for prey, T_s , wasted time interacting with humans, T_h , and wasted time interacting with zombies, T_c . The resulting time budget is then

$$T = T_s + T_h + T_c. \quad (21)$$

We assume zombies, on average, encounter humans at a rate a and encounter other zombies in a hunt at a rate a_z . Assuming the average time spent, an interaction between a human and zombie is t_h we can express the total time a zombie spends interacting with, or handling, humans as

$$T_h = at_h HT_s. \quad (22)$$

Similarly, assuming the average time, a zombie wastes in a hunt with other zombies that does not result in a kill is t_c we can express the time wasted in group hunts per zombie as

$$T_c = a_z t_c Z T_s. \quad (23)$$

Substitution of these terms into (21) gives

$$T = T_s (1 + at_h H + a_z t_c Z) \quad (24)$$

rewritten in terms of searching time as

$$T_s = \frac{T}{1 + at_h H + a_z t_c Z}. \quad (25)$$

Terming N the total number of attacks in a window of time T , we write

$$N = a T_s H Z. \quad (26)$$

After substituting the expression for searching time, we have the number of attacks in time T is

$$N = \frac{a T H Z}{1 + at_h H + a_z t_c Z}. \quad (27)$$

Rewriting a_z in terms of a , the human-zombie encounter rate, and defining $\beta = t_c(a_z/a)$, the resulting attack rate, $A(H, Z) = N/T$, is

$$A(H, Z) = \frac{N}{T} = \frac{a H Z}{1 + at_h H + a \beta Z}. \quad (28)$$

B Variability in Susceptibility: The Theta-Logistic Model

Variability in the susceptibility of healthy individuals has been shown to play a role in disease propagation in host–pathogen interactions. Variation in susceptibility could play a role in many other disease systems as well (explore super spreader question). To explore this, we begin with the derivation given by Dwyer (2000). First, we modify

a standard SIR model so that $S = S(t, v)$ is a function of both time, t , and level of susceptibility, v , giving

$$\begin{aligned} \frac{dS}{dt} &= -vSI \\ \frac{dI}{dt} &= I \int_0^\infty vS \, dxv - \mu I. \end{aligned} \tag{29}$$

Then, we can define

$$S_j = \int_0^\infty v^j S(t, v) \, dxv \tag{30}$$

to be the j th moment of the distribution of susceptibility. Taking a time derivative of (30) gives

$$\begin{aligned} \frac{dS_j}{dt} &= \int_0^\infty v^j \frac{d}{dt} S(t, v) \, dxv \\ &= \int_0^\infty v^j (-vSI) \, dxv \\ &= -I \int_0^\infty v^{j+1} S(t, v) \, dxv \\ &= -IS_{j+1}. \end{aligned} \tag{31}$$

Here, we define $m_j = \frac{S_j}{S_0}$, so that $m_1 = m$, the mean of the distribution of susceptibility. Taking a time derivative, we have

$$\frac{dm_j}{dt} = \frac{d}{dt} \frac{S_j}{S_0} \tag{32}$$

$$= \frac{S_0 \frac{dS_j}{dt} - S_j \frac{dS_0}{dt}}{S_0^2} \tag{33}$$

$$= -Im_{j+1} - (-Im_jm) \tag{34}$$

$$= -I(m_{j+1} - m_jm). \tag{35}$$

From (32), we see that the rate of change for any moment depends on the next higher moment, and to get around this we assume that the coefficient of variation remains constant. Thus, we can write the square of the coefficient of variation, V , as

$$V = \frac{m_2 - m^2}{m^2}. \tag{36}$$

In turn, this gives

$$\frac{dm}{dt} = -I(m_2 - m^2) \tag{37}$$

$$= -IVm^2. \quad (38)$$

Next, we define $\hat{S}(t)$ to be the total population density at time t . We can then write $\hat{S}(t) = S_0(t)$, and it follows that

$$\frac{d\hat{S}}{dt} = -m\hat{S}I \quad (39)$$

Here, it is convenient to find an explicit expression for m , so we start with

$$\frac{dm}{dt} / \frac{d\hat{S}}{dt} = V \frac{m}{\hat{S}} \quad (40)$$

and integrating both sides in t gives

$$m(t) = \bar{m} \left(\frac{\hat{S}(t)}{\hat{S}(0)} \right)^V, \quad (41)$$

where $\bar{m} = m(0)$, the mean of the initial distribution of susceptibility. The resulting dynamics can then be written

$$\frac{d\hat{S}}{dt} = -\bar{m} \left(\frac{\hat{S}(t)}{\hat{S}(0)} \right)^V \hat{S}I \quad (42)$$

$$\frac{dI}{dt} = \bar{m} \left(\frac{\hat{S}(t)}{\hat{S}(0)} \right)^V \hat{S}I - \mu I. \quad (43)$$

C Clustering in Attack or Starvation: Negative Binomial

C.1 Spatial Clustering of Susceptibles

The assumption that populations are well-mixed fails if individuals come into contact with only a small, clustered, subpopulation. To account for this, May (1978) replaced the Poisson distribution of the Nicholson–Bailey model with a negative binomial distribution. The resulting difference equation, in HvZ terms, relating host and zombie populations is

$$H_{t+1} = FH_t \left(1 + \frac{aZ_t}{k_A} \right)^{-k_A}, \quad (44)$$

where H_t (Z_t) is the human (zombie) population at time t , a is the encounter rate, k_A is a clumping parameter, and F is the mean number of surviving progeny of a healthy individual. Kong et al. (2016) extend this assumption to a continuous time model

by observing that $\left(1 + \frac{aZ_t}{k_A}\right)^{-k_A}$ is the probability that a human escapes all zombies between times t and $t + 1$. The risk of infection between successive time steps is then $1 - \left(1 + \frac{aZ_t}{k_A}\right)^{-k_A}$, or a rate of infection of $k_A \ln\left(1 + \frac{aZ}{k_A}\right)$ since $\text{rate} = 1 - \exp(\text{rate})$ (Vynnycky and White 2010). The result is an attack term of

$$A(H, Z) = k_A \ln\left(1 + \frac{aZ}{k_A}\right) H. \tag{45}$$

C.2 Uneven Attack Distribution

When zombies have significantly different rates of encounter with humans a mean starvation rate poorly represents the true starvation rate for a given zombie. Assume the number of mean number of encounters per unit time the i th zombie has is Gamma distributed with shape parameter k_S and mean $r_S H_t$ (r_S being the ratio between the mean number of attacks and the maximum number of attacks per zombie and H_t the number of humans at time t). Adapting the derivation of Kong et al. (2016), we see the conditional probability that the number of successful hunts (X_i) per unit time the i th zombie has, conditioned on the mean number of encounters the i th zombie has (θ_i) being θ , is expressed

$$P(X_i = x | \theta_i = \theta) = \frac{e^{-\theta} \theta^x}{x!}, \quad x = 0, 1, 2, \dots \tag{46}$$

and the marginal distribution of X_i is the PDF of a negative binomial distribution with mean $r_S H_t$. Then, the probability of starvation in a time window of Δt for a given zombie is the zero term of the negative binomial multiplied by the duration, namely

$$P(\text{starvation}) = \Delta t \left(1 + \frac{r_S H_t}{k_S}\right)^{-k_S}. \tag{47}$$

A difference equation for the zombie population that accounts only for starvation can then be written

$$Z_{t+\Delta t} = Z_t \left(1 - \Delta t \left(1 + \frac{r_S H_t}{k_S}\right)^{-k_S}\right). \tag{48}$$

In its limit, a rearrangement of the difference equation produces the starvation term

$$\dot{Z} = -Z \left(1 + \frac{r_S H}{k_S}\right)^{-k_S} \tag{49}$$

but since starvation is the product of consecutive unsuccessful hunts over a 24 h period this needs to be scaled, resulting in the uneven attack distribution starvation model

$$S(H, Z) = \frac{mZ}{\left(1 + \frac{r_S H}{k_S}\right)^{k_S}}. \tag{50}$$

It is worth noting that although both this starvation mechanism and the spatial clustering of susceptibles derived in ‘‘Appendix C.1’’ are based on a negative binomial distribution, they appear to have different limiting behavior. This is the result of tracking different terms from the negative binomial (i.e., the zero term or everything but the zero term). However, for small r_S/k_S , the logarithmic and negative binomial have essentially the same behavior.

D BIC

Now, $\mathcal{L}[\theta|\text{Data}]$ is the likelihood of a specific parameter set given particular data. Bayesian information criterion, formalized by Schwarz (1978), is a platform for model comparison that translates this probability of a parameter set into a probability for the model. Lewis and Powell (2017) provide a relatively elementary demonstration that the probability of a model M given data, $P(M|\text{Data})$, can be found by looking at the likelihood over the whole parameter space Θ . From Bayes’ theorem, we can say

$$P(M|\text{Data}) = \frac{1}{P[\text{Data}]} \int_{\theta \in \Theta} \mathcal{L}[\theta|\text{Data}] P(\theta) d\theta \tag{51}$$

$$\propto \int_{\theta \in \Theta} \mathcal{L}[\theta|\text{Data}] P(\theta) d\theta \tag{52}$$

$$\propto \int_{\theta \in \Theta} \exp\left[-N \ln(2\pi\sigma^2) - \sum_{n=1}^N \frac{(Z_n - Z(t_n, \theta))^2 + (H_n - H_{\text{obs}}(t_n, \theta))^2}{2\sigma^2}\right] P(\theta) d\theta \tag{53}$$

$$\propto \int_{\theta \in \Theta} \exp[-\text{NLL}(\theta)] P(\theta) d\theta. \tag{54}$$

Since we use the same data to fit each model, $P[\text{Data}]$ in (51) is common to each model and thus can be ignored for the purpose of model competition. This is important because it is impossible for us to determine the probability of the data occurring.

After fitting our models to the data, we need a way to measure which model fits best while still maintaining model simplicity, which we measure by the number of parameters. We have the probability of the model, related to the NLL, but need a fair way to penalize a model for complexity. We assume that NLL is at least twice continuously differentiable in a neighborhood of $\hat{\theta}$ and that all parameters are independent. We can express NLL by its Taylor series expansion

$$\begin{aligned} \text{NLL}(\boldsymbol{\theta}) &= \text{NLL}(\hat{\boldsymbol{\theta}}) + \sum_{n=1}^N \left[\sum_{i=1}^K \frac{\partial}{\partial \theta_i} \text{NLL}(\hat{\boldsymbol{\theta}})(\theta_i - \hat{\theta}_i) \right] \\ &+ \sum_{n=1}^N \left[\sum_{i=1}^K \frac{1}{2} \frac{\partial^2}{\partial \theta_i^2} \text{NLL}(\hat{\boldsymbol{\theta}})(\theta_i - \hat{\theta}_i)^2 \right] + \dots \end{aligned} \tag{55}$$

$$= \text{NLL}(\hat{\boldsymbol{\theta}}) + \sum_{n=1}^N \left[\sum_{i=1}^K \frac{1}{2} \frac{\partial^2}{\partial \theta_i^2} \text{NLL}(\hat{\boldsymbol{\theta}})(\theta_i - \hat{\theta}_i)^2 \right] + \dots \tag{56}$$

where K is the number of parameters. Independent parameters, in this context, means that all mixed partials are zero. Further, we can rescale parameters through non-dimensionalization to ensure that $\frac{\partial^2}{\partial \theta_i^2} \text{NLL}(\hat{\boldsymbol{\theta}}) = 1$, giving

$$\text{NLL}(\boldsymbol{\theta}) = \text{NLL}(\hat{\boldsymbol{\theta}}) + \frac{N}{2} \left[\sum_{i=1}^K (\theta_i - \hat{\theta}_i)^2 \right] + \dots \tag{57}$$

Substitution of (57) into (54) provides

$$P(M|\text{Data}) \propto \int_{\boldsymbol{\theta} \in \Theta} \exp[-\text{NLL}(\boldsymbol{\theta})] P(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{58}$$

$$\propto \int_{\boldsymbol{\theta} \in \Theta} \exp \left[-\text{NLL}(\hat{\boldsymbol{\theta}}) - \frac{N}{2} \left[\sum_{i=1}^K (\theta_i - \hat{\theta}_i)^2 \right] + \dots \right] P(\boldsymbol{\theta}) d\boldsymbol{\theta} \tag{59}$$

The dominant contribution to this will come from dominating exponential terms so as long as $P(\boldsymbol{\theta})$ is not exponential we can use $\int_{-\infty}^{\infty} e^{-ax^2} = \sqrt{\frac{\pi}{a}}$ to write

$$\begin{aligned} P(M|\text{Data}) &\sim e^{-\text{NLL}(\hat{\boldsymbol{\theta}})} \left(\sqrt{\frac{2\pi}{N}} \right)^K \\ &\sim \exp \left[-\text{NLL}(\hat{\boldsymbol{\theta}}) - \frac{1}{2} (K \ln(N) - K \ln(2\pi)) \right] \end{aligned} \tag{60}$$

It is true in our work (and is traditionally assumed) that $N \gg 2\pi$ which allows for the neglecting of the constant $K \ln(2\pi)$. The BIC is then defined as $\frac{1}{2} \text{BIC} = -\text{NLL}(\hat{\boldsymbol{\theta}}) - \frac{1}{2} (K \ln(N))$ or, as it is presented by Schwarz (1978)

$$\text{BIC} = 2\text{NLL}(\hat{\boldsymbol{\theta}}) + (K \ln(N)). \tag{61}$$

E Code for Selecting and Parameterizing a Model for Humans Versus Zombies

```

% % % % % Code used for parameterization of models
% % % % % with data from HvZ.
% % % % % This code is broken into four functions

% % % % % 1) HvZparameterization
% % % % % Reads in data, establishes bounds on parameter
% % % % % space, calls later functions

% % % % % 2) HZ__parametrization
% % % % % Establishes parameters to be fit (model
% % % % % dependent), calls the
% % % % % optimization function to determine optimal
% % % % % parameters, calls the
% % % % % parameterized ODE for population predictions,
% % % % % if desired
% % % % % 3) error
% % % % % The error function in the Maximum Likelihood
% % % % % Function. A
% % % % % different error function could be substituted
% % % % % here.

% % % % % 4) hzode
% % % % % The desired model to be parameterized. The
% % % % % example here is the
% % % % % best performing model (SC) with the Attack
% % % % % Mechanism of
% % % % % Suceptible Clustering, the Starvation
% % % % % Mechanism of Uneven Attack
% % % % % Distribution, and the Quitting Mechanism

function [] = HvZparameterization

load('HvZ2011.mat','HF2011','ZF2011','tF2011')
% % read in data from file

H = HF2011; % % Human population data from Fall 2011
Z = ZF2011'; % % Zombie population data from Fall
2011
t = tF2011; % % time vector data from Fall 2011

% % % % % normalize human and zombie populations by total
population at the
% % % % % start of the the game
P0 = H(1) + Z(1);
h = H./P0; % % normalized human population
z = Z./P0; % % normalized zombie population

% % % % % bounds on parameter space. Not required but
% % % % % makes paramterization
% % % % % more efficient. If fitted parameters are near
% % % % % bounds these need

```

```

% % % % % to be checked
lbb = 0;
ubb = 100;

% % % % Call parameterization function
[parameters,BIC] = HZ_parameterization(t,h,z,P0,lbb,ubb);

% % % % % Outputs
disp('Parameters')
disp(parameters)
disp('BIC')
disp(BIC)

end

function [parameters,BIC] = HZ_parameterization(t,h,z,
    P0,lbb,ubb)

IC=[h(1), h(1), z(1)]; % % Establish initial conditions for
    the ODE
parms=[.1, .1, .1, .1, .1, .1]; % % Initial guess for
    parameters
lb = lbb*ones(1,length(parms)); % % lower bound vector
ub = ubb*ones(1,length(parms)); % % upper bound vector

% % % % fmincon is the optimization function used to
% % % % % produce MLE estimates
% % % % in our MLE procedure. The error function called is
% % % % based on the NLL
parms = fmincon(@(x) error(t,h,z,...
    [x(1) x(2) x(3) x(4) x(5) x(6)],IC),parms,[],[],[],[],
    lb,ub);

% % % % To calculate BIC an estimate of variance (stderror)
% % % % % is needed
% % % % as described in the BIC Estimation Section.
[~,hz]=ode45(@(t,y)hzode(t,y,parms),t,IC); % %
N = length(h); % % number of data points
v = length(parms); % % number of fit parameters
errs2 = (hz(:,1)-h).^2+(hz(:,3)-z).^2;
stderror = 1/(2*N)*sum(errs2);%Variance of the data from
    the model estimate
NLL=N*log(2*pi*stderror)+1/(2*stderror)*sum(errs2);
BIC=2*NLL+v*log(N); % % BIC for parameterized model
    given the data

parameters=parms; % % parameter estimates for normalized
    data
parameters([1 4]) = parms([1 4])/P0; % % some
    parameters depend on

```

```

% % population
% % size,
% % so these
% % are
% % converted
% % values

end

function err=error(t,h,z,parms,IC)

% % % % For a given parameter set, the ODE is solved
% % % % numerically using
% % % % ode45
[~,hz]=ode45(@(t,y)hzode(t,y,parms),t,IC);

% % % % the error between the model estimates and the
% % % % data is calculated,
% % % % and subsequently minimized by fmincon
err=sum((hz(:,1)-h).^2)+sum((hz(:,3)-z).^2);

end

function rhs=hzode(~,x,parms)

% % % % Set of parameters for a given model. This is an
% % % % example and needs
% % % % to be modified depending on the model being fit.
a=parms(1);
ka=parms(2);
m=parms(3);
rd=parms(4);
kd=parms(5);
q=parms(6);

% % % % The model solved by ode45. The variable x =
% % % % [Hobs; H; Z]
% % % % consists of Humans present in the data (Hobs),
% % % % Humans present in
% % % % the game (H) to account for those that quit the
% % % % game,
% % % % and Zombies (Z)

rhs=zeros(3,1);
rhs(1)=-ka*log(1+(a*x(3)/ka))*x(2);
rhs(2)=-ka*log(1+(a*x(3)/ka))*x(2)-q*x(2);
rhs(3)=ka*log(1+(a*x(3)/ka))*x(2)-m*x(3)/((1+(rd*
x(2)/kd))^kd);

end

```

References

- Beddington JR (1975) Mutual interference between parasites or predators and its effect on searching efficiency. *J Anim Ecol* 44:331–340. <https://doi.org/10.2307/3866>
- Begon M et al (2002) A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiol Infect* 129:147–153. <https://doi.org/10.1017/S0950268802007148>
- Bertozzi AL et al (2020) The challenges of modeling and forecasting the spread of COVID-19. *PNAS* 117(29):16732–16738. <https://doi.org/10.1073/pnas.2006520117>
- Brauer F, van den Driessche P, Wu J (2008) *Mathematical epidemiology*. Springer, Berlin
- Chen S et al (2018) Pathogen transfer through environment-host contact: an agent-based queuing theoretic framework. *Math Med Biol* 35(3):409–425. <https://doi.org/10.1093/imammb/dqx014>
- Colizza V et al (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *PNAS* 103(7):2015–2020. <https://doi.org/10.1073/pnas.0510525103>
- DeAngelis DL, Goldstein RA, O'Neill RV (1975) A model for trophic interaction. *Ecology* 56(4):881–892. <https://doi.org/10.2307/1936298>
- Diaz P et al (2017) A modified SEIR model for the spread of Ebola in Western Africa and metrics for resource allocation. *Appl Math Comput* 324:141–155. <https://doi.org/10.1016/j.amc.2017.11.039>
- Dixon P (2006) Bootstrap resampling. In: Piegorisch WW, El-Shaarawi AH, Host G (eds) *Encyclopedia of environmental statistics*. Wiley, Hoboken. <https://doi.org/10.1002/9780470057339>
- Dwyer G et al (2000) Pathogen-driven outbreaks in forest defoliators re-visited: building models from experimental data. *Am Nat* 156(2):105–120. <https://doi.org/10.1086/303379>
- Eames K, Keeling M (2002) Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *PNAS* 99(20):13330–13335. <https://doi.org/10.1073/pnas.202244299>
- Federico P et al (2013) Optimal control in individual-based models: implications from aggregated methods. *Am Nat* 181(1):64–77. <https://doi.org/10.1086/668594>
- Fisher RA (1920) A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Mon Not R Astron Soc* 80(8):758–770. <https://doi.org/10.1093/mnras/80.8.758>
- Fisher RA (1922) On the mathematical foundations of theoretical statistics. *Philos Trans R Soc A* 222:309–368. <https://doi.org/10.1098/rsta.1922.0009>
- Garlick MJ et al (2011) Homogenization of large-scale movement models in ecology. *Bull Math Biol* 73:2088–2108. <https://doi.org/10.1007/s11538-010-9612-6>
- Garlick MJ et al (2014) Homogenization, sex, and differential motility predict spread of chronic wasting disease in mule deer in southern Utah. *J Math Biol* 69(2):369–399. <https://doi.org/10.1007/s00285-013-0709-z>
- Gross JE, Miller MW (2001) Chronic wasting disease in mule deer: disease dynamics and control. *J Wildl Manag* 65(2):205–215. <https://doi.org/10.2307/3802899>
- Hefley TJ et al (2017) When mechanism matters: forecasting the spread of disease using ecological diffusion. *Ecol Lett* 20(5):640–650. <https://doi.org/10.1111/ele.12763>
- Hilborn R, Mangel M (1997) *The ecological detective*. Princeton University Press, Princeton
- Holling CS (1959) Some characteristics of simple types of predation and parasitism. *Can Entomol* 91(7):385–398. <https://doi.org/10.4039/Ent91385-7>
- Hooten MB, Garlick MJ, Powell JA (2013) Computationally efficient statistical differential equation modeling using homogenization. *J Agric Biol Environ Stat* 18:405–428. <https://doi.org/10.1007/s13253-013-0147-9>
- Hopkins SR et al (2020) Systematic review of modelling assumptions and empirical evidence: does parasite transmission increase nonlinearly with host density? *Methods Ecol Evol* 11:476–486. <https://doi.org/10.1111/2041-210X.13361>
- Hosseini PR, Dhondt AA, Dobson AP (2006) Spatial spread of an emerging infectious disease: conjunctivitis in house finches. *Ecology* 87(12):3037–3046. [https://doi.org/10.1890/0012-9658\(2006\)87\[3037:SSOAEI\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[3037:SSOAEI]2.0.CO;2)
- Kay S, Powell JA, Knowlton F (2015) Canid social structure and density dependence improve predator-prey models of *Canis latrans* and *Lepus californicus* in Curley Valley, UT. *Open J Ecol* 5(4):120–135. <https://doi.org/10.4236/oje.2015.54011>
- Kennedy D, Dwyer G (2018) Effects of multiple sources of genetic drift on pathogen variation within hosts. *PLoS Biol* 16(3):e2004444. <https://doi.org/10.1371/journal.pbio.2004444>

- Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc A* 115(772):700–721. <https://doi.org/10.1098/rspa.1927.0118>
- Kong L et al (2016) Modeling heterogeneity in direct infectious disease transmission in a compartmental model. *Int J Environ Res Public Health* 13:253. <https://doi.org/10.3390/ijerph13030253>
- Kot M, Schaffer WM (1986) Discrete-time growth-dispersal models. *Math Biosci* 80(1):109–136. [https://doi.org/10.1016/0025-5564\(86\)90069-6](https://doi.org/10.1016/0025-5564(86)90069-6)
- Laguzet L, Turinici G (2015) Individual vaccination as Nash equilibrium in a SIR model with application to the 2009–2010 In uenza A (H1N1) epidemic in France. *Bull Math Biol* 77:1955–1984. <https://doi.org/10.1007/s11538-015-0111-7>
- Lewis M, Powell JA (2016) Modeling zombie outbreaks: a problem-based approach to improving mathematics one brain at a time. *PRIMUS* 26(7):705–726. <https://doi.org/10.1080/10511970.2016.1162236>
- Lewis M, Powell JA (2017) Yeast for mathematicians: a ferment of discovery and model competition to describe data. *Bull Math Biol* 79(2):356–382. <https://doi.org/10.1007/s11538-016-0236-3>
- Liu BR, Kot M (2019) Accelerating invasions and the asymptotic of a fat-tailed dispersal. *J Theor Biol* 471:22–41. <https://doi.org/10.1016/j.jtbi.2019.03.016>
- Lotka AJ (1925) *Elements of physical biology*. Williams and Wilkins, Baltimore
- Luke DA, Harris JK (2007) Network analysis in public health: history, methods, and applications. *Annu Rev Public Health* 28:69–93. <https://doi.org/10.1146/annurev.publhealth.28.02140.144132>
- Maheswaran R et al (2009) A graph-theory method for pattern identification in geographical epidemiology: a preliminary application to deprivation and mortality. *Int J Health Geogr* 8:28. <https://doi.org/10.1186/1476-072X-8-28>
- May RM (1978) Host-parasitoid systems in patchy environments: a phenomenological model. *J Anim Ecol* 47(3):833–844. <https://doi.org/10.2307/3674>
- McCallum H, Barlow N, Hone J (2001) How should pathogen transmission be modelled. *Trends Ecol Evol* 16:6. [https://doi.org/10.1016/S0169-5347\(01\)02144-9](https://doi.org/10.1016/S0169-5347(01)02144-9)
- Medlock J, Kot M (2003) Spreading disease: integro-differential equations old and new. *Math Biosci* 184(2):201–222. [https://doi.org/10.1016/S0025-5564\(03\)00041-5](https://doi.org/10.1016/S0025-5564(03)00041-5)
- Meyers LA et al (2005) Network theory and SARS: predicting outbreak diversity. *J Theor Biol* 232(1):71–81. <https://doi.org/10.1016/j.jtbi.2004.07.026>
- Miller JC (2009) Spread of infectious disease through clustered populations. *J R Soc Interface* 6(41):1121–1134. <https://doi.org/10.1098/rsif.2008.0524>
- Mollison D (1972) The rate of spatial propagation of simple epidemics. In: *Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 3: probability theory*. University of California Press, Berkeley, pp 579–614. <https://projecteuclid.org/euclid.bsm/1200514358>
- Munz O et al (2005) When zombies attack!: mathematical modeling of an outbreak of zombie infection. In: Tchuente JM, Chiyaka C (eds) *Infectious disease modelling research progress*. Nova Science Publishers Inc, Berkeley, pp 133–150
- Neath AA, Cavanaugh JE (2012) The Bayesian information criterion: background, derivation, and applications. *WIREs Comput Stat* 4(2):199–203. <https://doi.org/10.1002/wics.199>
- O'Regan SM et al (2015) Multi-scale model of epidemic fade-out: will local extirpation events inhibit the spread of white-nose syndrome? *Ecol Appl* 25(3):621–633. <https://doi.org/10.1890/14-0417.1>
- Pech RP, McIlroy JC (1990) A model of the velocity of advance of foot and mouth disease in feral pigs. *J Appl Ecol* 27(2):635–650. <https://doi.org/10.2307/2404308>
- Powell JA, Bentz BJ (2014) Phenology and density-dependent dispersal predict patterns of mountain pine beetle (*dendroctonus ponderosae*) impact. *Ecol Model* 273:173–185. <https://doi.org/10.1016/j.ecolmodel.2013.10.034>
- Robins J et al (2015) Agent-based model for Johne's disease dynamics in a dairy herd. *Vet Res* 46:68. <https://doi.org/10.1186/s13567-015-0195-y>
- Rock KS et al (2018) Assessing strategies against gambiense sleeping sickness through mathematical modeling. *Clin Infect Dis* 66(suppl 4):S286–S292. <https://doi.org/10.1093/cid/ciy018>
- Schwarz G (1978) Estimating the dimension of a model. *Ann Stat* 6(2):461–464. <https://doi.org/10.1214/aos/1176344136>
- Stolerman L, Coombs D, Boatto S (2015) SIR-network model and its applications to dengue fever. *SIAM J Appl Math* 75(6):2581–2609. <https://doi.org/10.1137/140996148>
- Tien J, Earn D (2010) Multiple transmission pathways and disease dynamics in a waterborne pathogen model. *Bull Math Biol* 72:1506–1533. <https://doi.org/10.1007/s11538-010-9507-6>

- Tuite AR et al (2011) Cholera Epidemic in Haiti, 2010: using a transmission model to explain spatial spread of disease and identify optimal control interventions. *Ann Intern Med* 154(9):593–601. <https://doi.org/10.7326/0003-4819-154-9-201105030-0033400334>
- Turchin P (1998) Quantitative analysis of movement. Sinaur Associates, Inc., Publishers, Sunderland
- Volz EM et al (2011) Effects of heterogeneous and clustered contact patterns on infectious disease dynamics. *PLoS Comput Biol* 7:6. <https://doi.org/10.1371/journal.pcbi.1002042>
- Vynnycky E, White R (2010) An introduction to infectious disease modelling. Oxford University Press, Oxford
- Wang M-H, Kot M, Neubert MG (2002) Integrodifference equations, Allee effects, and invasions. *J Math Biol* 44:150–168. <https://doi.org/10.1007/s002850100116>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.