Fall 2021

Critical behavior in evolutionary and population dynamics

Stephen Walter Ordway

Department: Physics

CRITICAL BEHAVIOR IN EVOLUTIONARY AND POPULATION DYNAMICS

by

STEPHEN WALTER ORDWAY

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

And

UNIVERSITY OF MISSOURI – ST. LOUIS

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

PHYSICS

2021

Approved by:

Sonya Bahar, Advisor
Alexey Yamilov, Co-Advisor
Ricardo Flores
Julia Medvedeva
Wendy Olivas

ABSTRACT

This study is an exploration of phase transition behavior in evolutionary and population dynamics, and techniques for predicting population changes, across the disciplines of physics, biology, and computer science. Under the looming threat of climate change, it is imperative to understand the dynamics of populations under environmental stress and to identify early warning signals of population decline. These issues are explored here in (1) a computational model of evolutionary dynamics, (2) an experimental system of decaying populations under environmental stress, and (3) a machine learning approach to predict population changes based on environmental factors. Through the lens of critical phase transition behavior, the non-equilibrium continuous transition of a neutral agent-based model is shown to exhibit power-law-like behavior for two control parameters in the critical regime. The model does not fall within the directed percolation universality class, despite exhibiting some characteristics of directed percolation. The results also compare a system exhibiting quenched randomness with one that does not. Experimentally, the impact of two stressors, temperature and NaCl stress, are examined in *S. cerevisiae*. Increased levels of NaCl in growth media result in a smooth transition from a survivable to an uninhabitable environment, whereas increased temperature stress results in a transition with signs of critical behavior. Lastly, population data from the Living Planet Index and weather data from NOAA are used to predict population changes based on weather attributes using classification and regression machine learning models. Results indicate that a machine learning approach is viable, but additional data and environmental factors are needed to improve the predictive models.

ACKNOWLEDGMENTS

First, I would like to thank Dr. Sonya Bahar for the support and patience through my years in graduate school and for the opportunity to investigate this work. Her guidance both personally and academically has helped me become a better researcher and a better person. I would not be where I am today without her support.

Thanks to Adam Scott for the time and effort educating me on the computational model and explaining the basis of which a great deal of this work has been built. Thank you to my lab partners, Tera Glaze, Scott Lewis and Shane Meyer for the discussion and perspective as I learned and implemented computational methods of analysis and for keeping me on task. Special thanks also to David Friend for all of your work on the yeast experiment and data presentation.

Thank you to Dr. Wendy Olivas and her lab for introducing me to the world of microbiology and all of the assistance as I learned my way around a wet lab. I would also like to thank my other committee members, Dr. Ricardo Flores, Dr. Julia Medvedeva, and Dr. Alexey Yamilov for their guidance and discussion.

To my wife, Dawn, thank you for the years of love and support in and outside of the lab, from introducing me to the computational model, to long nights of discussing ideas and all of the support during writing, to growing together overs these past years. Words cannot express my appreciation. To my son, Blake, thank you for reminding me not to be so serious all the time, and for pulling me away to do something fun with you during times of stress. To my parents, Fred and Nancy, and my sister and brother-in-law, Ellie and Josh, thank you for all of the love, encouragement and always taking the time to talk through topics as I worked through them.

Lastly, thank you to my funding sources for this work, the James S. McDonnell Foundation and the UM system IDIC grant.

TABLE OF CONTENTS

Page

LIST OF ILLUSTRATIONS

LIST OF TABLES

NOMENCLATURE

Symbol	Description
μ	Mutation Rate
δ	Death Rate
δ_{max}	Death Max Rate
c_b	Offspring Organism Coordinate
c_r	Parent Organism Coordinate
$P_s(t)$	Probabilty of Survival
$N(t)$	Number of Active Sites
$R^2(t)$	Mean Square Spreading
δ_e	Survival Probability Exponent
θ	Slip Exponent
$\tilde{z}$	Spreading Exponent
M	Mean Cluster Mass
T	Mean Survival Time
V	Mean Spatial Volume
S	Mean Size
Δ	Off-critical Measure
γ	Critical Relation – Mean Cluster Mass
τ	Critical Relation – Mean Survival Time
ν	Critical Relation – Mean Spatial Volume
σ	Critical Relation – Mean Size

$\upsilon_{\parallel}$	Critical Exponent – Correlation Time
$\upsilon_{\perp}$	Critical Exponent – Correlation Length
β	Critical Exponent – Population Density
β'	Critical Exponent – Survival Probability
z	Dynamical Exponent
n	Number of Doublings

1. INTRODUCTION

Climate-driven population declines have become increasingly prevalent in the age of the Anthropocene. Because of global-scale environmental changes, and the evidence that the Anthropocene marks the beginning of the sixth mass extinction event, understanding the impact on populations is of increasing importance (Kolbert, 2014). Investigation of population-level responses to stress casts a wide net, stretching across scientific disciplines where each field offers a different perspective by utilizing different analytical methods.

Computational modeling provides a platform to combine different fields of study. Even in the realm of computational modeling there are several approaches that can be taken. The physics-based approach is to model a phenomenon that closely resembles reality, which can be in the form of Newtonian equations for topics like trajectory analysis, statistical mechanics to model the behavior of particles and changes of states in matter, or simulations using rule-based, stochastic Markov processes to study systems-level behavior in time. Analyzing population behaviors from a physics-based perspective has proven fruitful in the past. For example, modeling population clustering on a neutral landscape has shown that clustering can occur in independently diffusing individuals, by reproduction and death processes (Houchmandzadeh, 2002, 2008; Houchmandzadeh & Vallade, 2003; Meyer *et al*., 1996; Young *et al*., 2001). In several cases, models have demonstrated evolution and speciation on a neutral landscape, supporting neutral theory (Alonso *et al*., 2006; de Aguiar *et al*., 2009; Derrida & Peliti, 1991; Lawson & Jensen, 2007). Models of both neutral theory and natural selection have demonstrated critical

phase transition behavior (King *et al*., 2017; Scott *et al*., 2013; Weinberger, 1987; Zhang *et al*., 1990). These models exemplify the how computational models can increase the understanding of population dynamics.

Laboratory environments allow for the isolation of a variable to test its effect independently on system-level behavior. Investigation of different environmental stressors in yeast has shown that the population response is dependent on a specific stressor, such as heat stress, dilution rate (as a proxy for death rate), nutrient concentration, or osmotic stress (NaCl concentration). (Dai *et al*., 2012, 2015; Mensonides *et al*., 2002). Slow recovery and early warning signals of an at-risk population under an environmental stress have been demonstrated from changes in population resilience and stability, identifying that the response can vary based on the type of stress on the population (Dai *et al.*, 2013, 2015). Observations like slow recovery and changes in resilience and stability near population collapse reveal critical phase transition behavior, critical slowing down, and fluctuations of an active state near the transition (Scheffer et al., 2009, 2012). The connection across scientific fields provides greater understanding of population collapse, and investigation from both biological- and physics-based approaches are complimentary to developing early warning signals and population recovery efforts.

Beyond physics-based modeling and laboratory work, there has been increasing demand and interest in building more "data-driven" models utilizing methods from the field of data science. The key to building a data-driven model is to take advantage of our data-rich world and use known data to build a model that attempts to predict future outcomes. The field of data science has established standard artificial

intelligence/machine learning techniques to accomplish inference (prediction) using real data. This approach has many applications, from identifying critical behavior of disease spread based on vaccination behavior to classifying individuals into pre-defined populations by detecting genetic differences between populations (Bridges *et al*., 2011; Pananos *et al*., 2017). The application of machine learning may provide valuable insight into the examination of populations in the wild. With so many environmental factors that can affect a population, traditional analytics have limited capability, but incorporating these environmental impacts as inputs or features in a machine learning model may function as an effective method for analysis.

1.1. SECTION TWO SUMMARY

This section examines absorbing state phase transition behavior of a neutral evolution model to determine whether the model belongs in the directed percolation universality class. Investigation of the transition includes the comparison of two control parameters, with and without quenched disorder, through the measurement of absorbing phase critical relations.

1.2. SECTION THREE SUMMARY

The work presented in Section 3 investigates the effect of multiple environmental stressors on yeast and bacterial populations. Yeast cultures were subjected to salt and temperature stress. The population declines, or death curves, were examined for characteristics of critical transition behavior. The temperature stress conditions were found to result in behavior similar to a critical phase transition, while the salt stress

conditions showed a more gradual decline. Similar results were observed by collaborators applying different antibiotics to bacterial populations.

1.3. SECTION FOUR SUMMARY

The final section explores the use of machine learning to predict population change in North American mammalian, amphibian, and reptilian species based on a number of weather attributes. Two classification models were trained to predict population increase or decline, and one regression model was trained to predict population change. Comparisons are made between a dataset with fewer features and a higher number of training examples and a dataset with more features but fewer training points.

2. POPULATION DYNAMICS IN A COMPUTATIONAL MODEL

Computational modeling allows investigation of complex real-world problems by providing better control than an experimental approach as well as data collection at a lower cost. The model discussed in this section began with the investigation of optimized speciation based on mutation size in an agent-based model of evolutionary dynamics on a two-dimensional phenotype space (Dees & Bahar, 2010). Dees developed a clustering algorithm analogous to speciation to investigate the effects of various fitness landscapes using an assortative (nearest neighbor) mating scheme. Fitness is defined here as the number offspring each organism reproduces based on different computational landscape definitions (varying fitness based on location, shifting fitness landscapes, and landscapes modulated by feedback). This model was further enhanced by Scott *et al*. (2013), who investigated and compared phase transition behavior among assortative mating, asexual fission, and random mating reproduction schemes on a neutral fitness landscape (fitness is equal for all organisms regardless of position in the phenotypic space). Scott *et al*. observed phase transition behavior in models using asexual fission and assortative mating as the maximum mutation size was varied, but not with randomized mating. This result, which derives from the local interaction (mating) rule, is consistent with the transition from extinction to survival belonging to the directed percolation universality class. Critical scaling behavior was further investigated through the lens of cluster-level dynamics, and power-law-like relationships through dynamically changing clusters in time were found (Scott, 2014; Scott *et al.*, 2013). This supported the concept that there could be multiple levels of organization in the model, which parallels biological theories

of multi-level evolution (Okasha, 2009). King (2015) expanded on this by investigating the effects of different control parameters and demonstrated the existence of multiple types of phase transitions occurring in one model. These included a continuum percolation transition, which is only spatially dependent, and a time-dependent directed percolation transition. King extended the investigation into multi-level evolution through analyzing phylogenetic tree structures of populations at the individual and species (cluster) level of biological organization, and also investigated the effects of mass extinction events on phylogenies as the model approached criticality (King, 2015; King *et al*., 2017).

This section will extend the previously described work by examining the model behavior in the absorbing phase of the critical regime for the critical phase transition previously classified as directed percolation (Scott, 2014). By analyzing the transition from the absorbing phase, the aim is to accomplish two things: first, to determine the accuracy of the previous assumption that the model demonstrates time-reversal symmetry, and second, to determine how the critical transition behavior is dependent on control parameter, in this case, the mutation parameter μ and death parameters δ, δ_{max}. By comparing the effects of δ, a constant death percentage, and δ_{max}, a death percentage that changes at each generation, the effects of temporal disorder are examined within the system.

2.1. PHASE TRANSITIONS AND UNIVERSALITY CLASSES

A phase transition occurs when a system changes from one state to another, driven by a feature of the system. An observable characteristic (or dependent variable) of

the model system is referred to as the order parameter, and the driving feature (independent variable) of the system is referred to as the control parameter. A system's phase transition is discontinuous if the transition demonstrates a discontinuity, or abrupt change, in the order parameter behavior and is continuous when the order parameter demonstrates no discontinuities as the system changes from one phase to another. In addition to categorizing a phase transition based on continuity, a transition can also be described as a non-equilibrium transition if there exists an "absorbing phase," and an equilibrium transition if no such absorbing phase exists. An absorbing phase is defined by the system's inability to return to the active fluctuating state, regardless of control parameter value. The transition that occurs in this model is a continuous, non-equilibrium phase transition.

Continuous phase transitions can be characterized by power-law scaling sufficiently close to the transition, in the critical regime. Power-law scaling of an order parameter is used to calculate critical exponents, which describe the system's universal behavior. During the 1950s and 1960s, it was recognized that quantities like a parameter's critical point depend heavily on the interaction details in a system, while critical exponents instead depend on a small number of general, or universal, system features like dimension or symmetries (Hinrichsen, 2000b). This led to the universality hypothesis, first clearly formulated by Kadanoff in 1971, which "reduces the large variety of critical phenomena to a small number of equivalence classes, so-called universality classes, which depend only on a few fundamental parameters" (Hinrichsen, 2000b). Every system in a universality class has the same critical exponents and exhibits the same scaling behavior near the critical point.

In its origins, universality classification was developed to explain universal behaviors in equilibrium transitions, but over the last several decades, these analytical methods have been applied to non-equilibrium transitions (Hinrichsen, 2000a). Directed percolation (DP) universality is a robust classification with respect to dynamic rules, making it the most understood class for non-equilibrium systems (Hinrichsen, 2000a). A model will be classified as directed percolation if it does not violate any parts of the DP-conjecture (Grassberger, 1982; Henkel *et al*., 2008; Janssen, 1981). The DP-conjecture, developed by Grassberger and Janssen, states that a system is belonging to directed percolation if:

1. the model displays a continuous phase transition from a *fluctuating* active phase into a *unique* absorbing state,
2. the transition is characterized by a non-negative one-component order parameter,
3. the dynamic rules are short-ranged, and
4. the system has no special attributes such as unconventional symmetries, conservation laws, or quenched randomness.

A number of theoretical models have fit into directed percolation universality, including Monte Carlo simulations in 2 + 1 dimensions, phase transitions in various lattice structure models, Brownian particles competing for resources, as well as modified versions of classical models like an Ising model with damage-spreading behavior and a Potts model with heat-bath dynamics (Grassberger, 1989; López *et al*., 2007; Marro & Dickman, 1999; Ódor, 2004). The first experimental realization of directed percolation was observed in turbulent liquid crystals (Takeuchi *et al*., 2009).

2.2. MODEL

The model is built on a 45 x 45 continuous, phenotypic landscape with time defined by generational steps forward. Active sites are analogous to organisms on the landscape, and each organism reproduces, creating the next generation, before the parent organism is removed and a series of death processes take place. The landscape is one of neutral fitness; each organism will produce two offspring and all removal processes are uniform across the landscape. For all analysis, simulations were initialized with an initial population placed on the landscape with a uniform random distribution. All simulations proceed through reproduction and death processes (described below) at each generation. Each simulation continues until the population goes extinct (zero), or the simulation reaches 2000 generations. Simulations were performed in MATLAB (The MathWorks) on PCs using a Windows 10 operating system.

2.2.1. Reproduction. All simulations follow asexual reproduction methods, as illustrated schematically in Figure 2.1; each organism creates two offspring without a mate. The offspring are placed within a 2μ x 2μ square centered around the parent. The coordinates of each offspring (c_{bx}, c_{by}) are determined by

$$c_{bx} = c_{rx} - \mu + 2r\mu \quad (1)$$

$$c_{by} = c_{ry} - \mu + 2r\mu \quad (2)$$

where (c_{rx}, c_{ry}) are the parent's coordinates, μ is the mutation parameter, and r is a random number drawn from a uniform distribution from the interval [0,1]. Once the offspring are created, the parents are removed, and the population goes through a series of death processes.

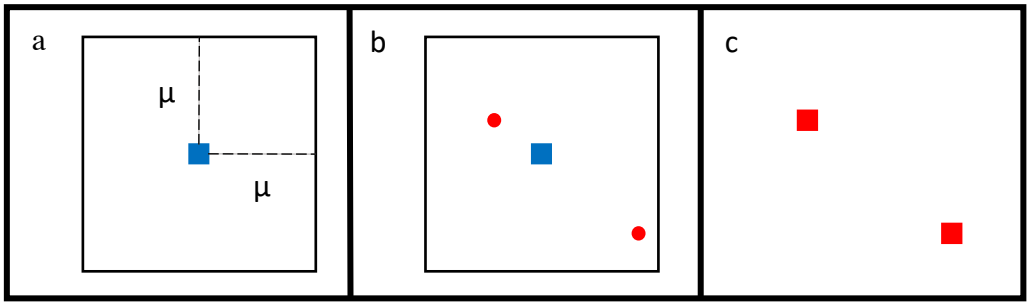

Figure 2.1 Asexual Reproduction. (a) The region of birth around the parent (blue square), a 2μ x 2μ square centered around the parent. (b) The offspring (red circles) are placed within the birth region. (c) The parent is removed, leaving the offspring to a series of death processes.

2.2.2. Death Processes. After reproduction occurs in each generation, a series of death processes follows. A competition limit is imposed for any organisms within the competition radius of κ=0.25 units, removing one of the organisms within the radius with no preference, simulating competition for resources in organisms that are too similar. All organisms that mutate to values outside of the landscape are also removed. Lastly, the death parameter removes some portion of the population. Two death parameters are compared in this work, δ, which removes a set percentage of the population at every generation, and δ_{max}, which removes a up to a certain percentage chosen from a uniform random distribution from 0 to δ_{max} for each generation. In both cases, an organism's probability of removal is uniform across the landscape because the model uses a neutral fitness landscape (meaning that each organism is subjected to same rule, even in the case of δ_{max}, where the percentage of removal may change from generation to generation).

2.3. METHODS

The first step is to identify the critical point for a control parameter. The critical point is the exact point at which the system transitions from an active or surviving phase to an absorbing or extinction phase. Simulations with initial populations of 300 were run at varying δ and δ_{max} in increments of 0.01, holding μ constant at 0.3, 0.35, and 0.4 to identify the approximate transition point for each value of μ. The transition point is identifiable as the region in which the population average at 2000 generations approaches zero. Once the approximate transition point was identified, 5000 simulations with an initial population of 1 were run in the parameter range near the transition in increments of 0.001 to gain more precision on the approach to criticality.

The scaling behavior of three observables allows for identification of a critical transition; the probability of survival $P_s(t)$, the number of active sites $N(t)$, and the mean square spreading (radius of gyration) from the origin $R^2(t)$ (Hinrichsen, 2000b, 2006). In each case, asymptotic power law behavior emerges at the critical point. The power law scaling for each observable is as follows:

$$P_s(t) \sim t^{-\delta_e} \tag{3}$$

$$N(t) \sim t^{\theta} \tag{4}$$

$$R^2(t) \sim t^{\tilde{z}} \tag{5}$$

where δ_e is the survival probability exponent, θ is the slip exponent, and $\tilde{z}$ is the spreading exponent (Henkel *et al*., 2008; Hinrichsen, 2000b, 2006). The traditional notation for the survivability exponent is δ, but in an effort to avoid confusion with the

death parameters δ and δ_{max}, the survivability exponent is defined as δ_e in this dissertation.

In practice, $P_s(t)$ is the percentage of survival at a given time t, calculated over all simulations, $N(t)$ is measured directly, and $R^2(t)$ is calculated by first calculating the mean spreading for each simulation and then averaging over all simulations. The mean spreading for a simulation, $R_k^2(t)$ is:

$$R_k^2(t) = \langle [x_i(t) - x_k(0)]^2 + [y_i(t) - y_k(0)]^2 \rangle \quad (6)$$

where $R_k^2(t)$ is the mean spreading for simulation k, $(x_i(t), y_i(t))$ are the coordinates of the ith organism at time t, and $(x_k(0), y_k(0))$ are the coordinates of the initial organism at $t = 0$. An average is then taken over all organisms active at time t, resulting in $R_k^2(t)$. Averaging over $R_k^2(t)$ for all simulations surviving at t gives $R^2(t)$.

The critical value of control parameters δ_c and $\delta_{max,c}$ are determined for $\mu = 0.3$, 0.35, and 0.4 by calculating best linear fit of $P_s(t)$, $N(t)$, and $R^2(t)$ on a double logarithmic scale. The critical parameter values exhibit the lowest chi-square linear fit (Garcia, 2000). In case of disagreement in critical value across the three observables, $N(t)$ is the most sensitive of the quantities and therefore determines the critical value (Hinrichsen, 2006). These critical value combinations of μ and δ or δ_{max} are used as critical points for both control parameters; when μ is acting as the control parameter, δ and δ_{max} are held constant at the corresponding δ_c and $\delta_{max,c}$, and vice versa. All comparisons of transition behavior based on control parameter are considered at parameter pairs of (μ_c, δ_c) or (μ_c, $\delta_{max,c}$).

After identifying the critical point, the scaling behavior of four additional quantities are used to characterize the system. Each of these quantities characterizes the scaling behavior of a system beginning with one active site in the absorbing phase. They are mean cluster mass *M*, mean survival time *T*, mean spatial volume *V*, and mean size *S*. These quantities scale as follows:

$$M \sim |\Delta|^{-\gamma} \tag{7}$$

$$T \sim |\Delta|^{-\tau} \tag{8}$$

$$V \sim |\Delta|^{-\upsilon} \tag{9}$$

$$S \sim |\Delta|^{-\sigma} \tag{10}$$

where Δ represents the off-critical measure, or the difference from the control parameter value and the critical point, and γ, τ, ν, and σ are critical relations. These scaling relations are used to calculate the four critical exponents, which identify the universality class of the system:

$$\gamma = \nu_{\parallel}(1 + \theta) = \nu_{\parallel} + d\nu_{\perp} - \beta - \beta' \tag{11}$$

$$\tau = \nu_{\parallel}(1 + \delta_e) = \nu_{\parallel} - \beta' \tag{12}$$

$$\upsilon = \nu_{\parallel}\left(\frac{d}{z} - \delta_e\right) = d\nu_{\perp} - \beta' \tag{13}$$

$$\sigma = \nu_{\parallel}\left(\frac{d}{z} + 1 - \delta_e\right) = \nu_{\parallel} + d\nu_{\perp} - \beta' \tag{14}$$

where, $\nu_{\parallel}$, $\nu_{\perp}$, β, and β' are the critical exponents used to determine universality class, d is the system's spatial dimension (in this case $d = 2$), and z is the is the dynamical exponent defined as $\tilde{z} = 2/z$.[1]

Because the field of critical phase transitions is almost entirely theoretical, the established method of determination for M, T, S, and V is to calculate them with the use of the system observables $P_s(t)$ and the pair-connectedness function (Henkel et al., 2008). This method of determination is well-defined compared to measuring M, T, S, and V directly from model data, and therefore will be the method used.

2.3.1. Mean Cluster Mass. The mean cluster mass, M, is calculated as the total integral of the pair connectedness function

$$M = \int d^d r \int_0^\infty dt\, c(r, t, \Delta) \tag{15}$$

where $c(r, t, \Delta)$ is the pair connectedness function, defined as "the probability that a cluster generated at site r_1 at time t_1 in an otherwise empty system activates r_2 at time t_2" (Henkel *et al.,* 2008). Applying this definition to our model, r_1 and t_1 correspond to the initial active site in an otherwise empty landscape and, at a given time t_2, any active site will have a probability of 1 and any non-active or empty sites will have a probability of 0.

---

[1] Notational inconsistencies are not uncommon with enough sources, sometimes even from the same author. The spreading exponent and dynamical exponent notation is no exception. Since the original work by Grassberger and de la Torre in 1978, a variety of combinations like Z and z, $\tilde{z}$ and z, and my personal favorite z and z, have been used to represent the spreading exponent and dynamical exponent respectively. For this work, I will use the notation defined in Henkel *et. al* (2008) where the spreading exponent will be represented as $\tilde{z}$ and the dynamical exponent as z.

By integrating over the landscape, the results become equal to the number of active sites at a given time *t*. Therefore, *M* can be rewritten as:

$$M = \int_0^\infty dt\, N(t) \tag{16}$$

which can be rewritten in discrete form for use with the generational data as:

$$M = \sum_t N(t) \tag{17}$$

where *N(t)* is the cluster mass (also called the average cluster mass)

$$N(t) = \langle \sum_i s_i(t) \rangle \tag{18}$$

where $s_i(t) = 1$ if a site is active and $s_i(t) = 0$ if a site is inactive.

2.3.2. Mean Survival Time, Mean Spatial Volume, and Mean Cluster Size.

The measurables of mean survival time *T*, mean spatial volume *V*, and mean cluster size *S* are all calculated by integrating over the probability of survival $P_s(t)$, with *S* and *V* having an additional multiplier before integration. The original equations as well as discrete form are included below:

$$T = \int dt\, P_s(t) \quad \Longrightarrow \quad T = \sum_t P_s(t) \tag{19}$$

$$V = \int dt\, P_s(t) t^{\frac{d}{z}-1} \quad \Longrightarrow \quad V = \sum_t P_s(t)\, t^{\frac{d}{z}-1} \tag{20}$$

$$S = \int dt\, P_s(t) t^{d/z} \quad \Longrightarrow \quad S = \sum_t P_s(t)\, t^{d/z} \tag{21}$$

where d is the spatial dimension and z is the dynamical exponent described above (Henkel *et al*., 2008).

2.4. RESULTS

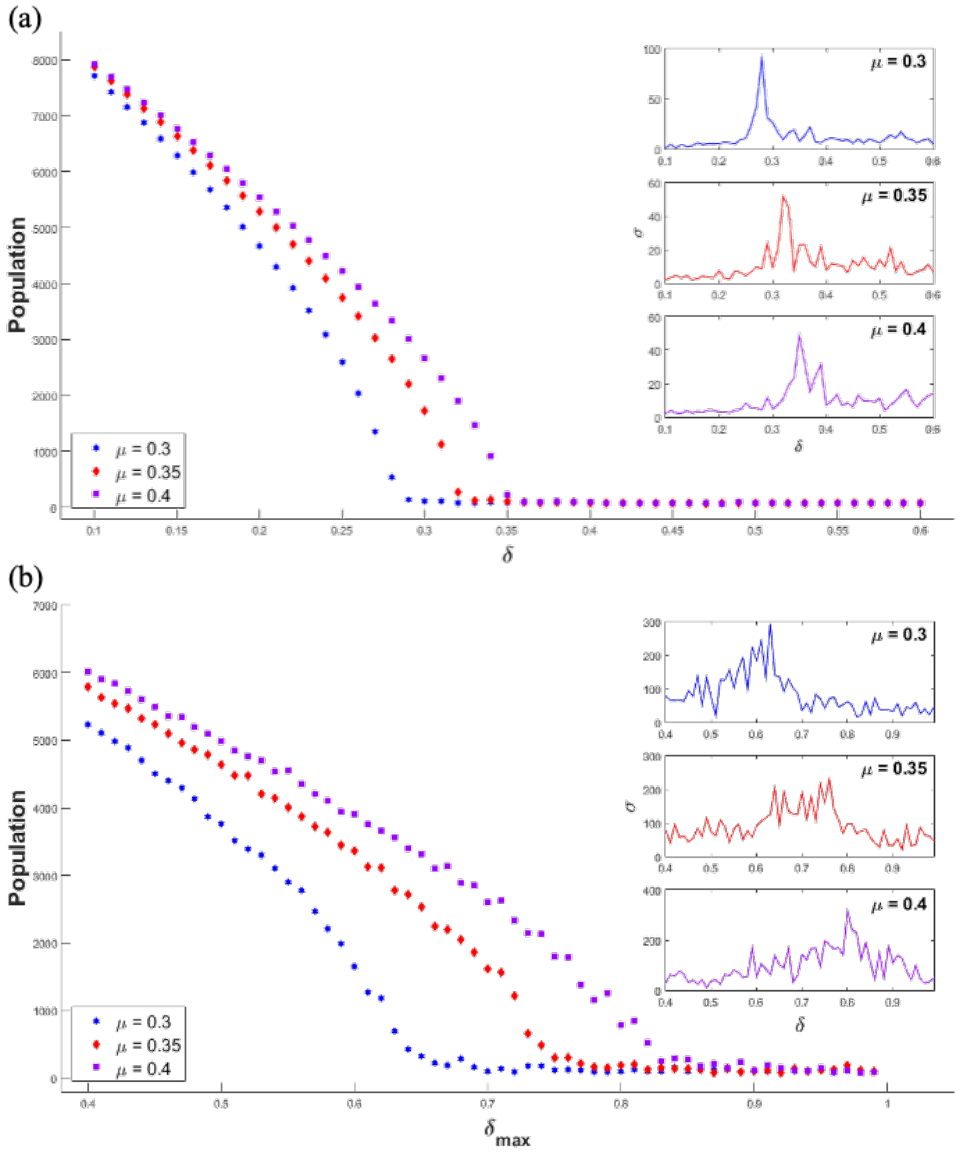

Figure 2.2 Phase Transition Behavior. Population sizes averaged over five simulations for δ (*a*) and δ_{max} (*b*) with insets of standard deviation are shown. Blue stars represent $\mu = 0.3$, red diamonds represent $\mu = 0.35$, and purple squares represent $\mu = 0.4$. The color of each standard deviation plot corresponds to the related dataset.

Figure 2.2 shows the continuous, nonequilibrium phase transitions for the population as a function of the control parameters $\delta(\alpha)$ and δ_{max} (b). Each subplot includes population decay curves for $\mu = 0.3$, 0.35 and 0.4 with corresponding standard deviations as insets. Characteristic behavior of a critical transition is shown for δ in Figure 2.2*a*, where each value of μ decays smoothly from the active phase to the absorbing phase with a spike in standard deviation at the transition point. In the case of δ_{max}, Figure 2.2*b* shows a trend of population decay from an active to an absorbing phase, but with greater fluctuations in the plotted averages and standard deviations. Even far from the transition point, the average population has noticeably higher fluctuations for δ_{max} compared to δ. The standard deviations show greater fluctuations about the transition point, but the insets in Figure 2.2*b* show these high fluctuations span a larger set of δ_{max} values, compared to the identifiable peaks seen in the insets of Figure 2.2*a*.

Following the determination of the approximate transition point for $\mu = 0.3$, 0.35, and 0.4, the survival probability, number of active clusters, and gyration radius were measured. An examination of each value through time demonstrated a shift from survival to population collapse, with the critical point demonstrating power-law-like behavior in survival probability and number of active sites. The critical point for each value of μ was determined by the death parameter value which minimized the chi-square linear fit test of the double logarithmic plot of each measurable vs. time (Garcia, 2000). Once the critical values δ_c and $\delta_{max,c}$ were determined for each μ, simulations were performed holding the death parameter constant at δ_c and $\delta_{max,c,}$ and varying μ as the control parameter.

Figure 2.3 shows the survival probability for control parameters δ and μ in the active and absorbing phases directly around the phase transition. Figure 2.3*a-c* corresponds to the control parameter δ for fixed values of μ = 0.3, 0.35 and 0.4 respectively, and Figure 2.3*d-f* corresponds to the control parameter μ for fixed values of δ = 0.281, 0.32, and 0.349 respectively. Each plot shows the average survival probability over 5000 simulations at the labeled order parameter value for t = [0, 2000]. The slope of the fit lines represents the critical relation exponent δ_e at the critical value of the control parameter. Each plot demonstrates the continued survival of population in the active phase, the critical transition exhibiting power-law-like behavior, and the decay of survival probability in the absorbing phase.

Similarly, Figure 2.4 represents the number of active sites for control parameters δ and μ in the active and absorbing phases in the immediate neighborhood the phase transition. Figure 2.4*a-c* corresponds to the control parameter δ for fixed values of μ = 0.3, 0.35 and 0.4 respectively, and Figure 2.4*d-f* correspond to the control parameter μ for fixed values of δ = 0.281, 0.32, and 0.349 respectively, with a time series from t = [0, 2000] and averaged over 5000 simulations. The slope of the fit lines represent the critical slip relation exponent θ. Both Figures 2.3 and 2.4 demonstrate that increasing values of μ_c and δ_c result in an increase in the domain of the critical regime; each subplot shows similar range in off-critical measure, but moving left to right, the plotted observables extend over a narrower range of the y-axis, suggesting that less of the critical regime is plotted.

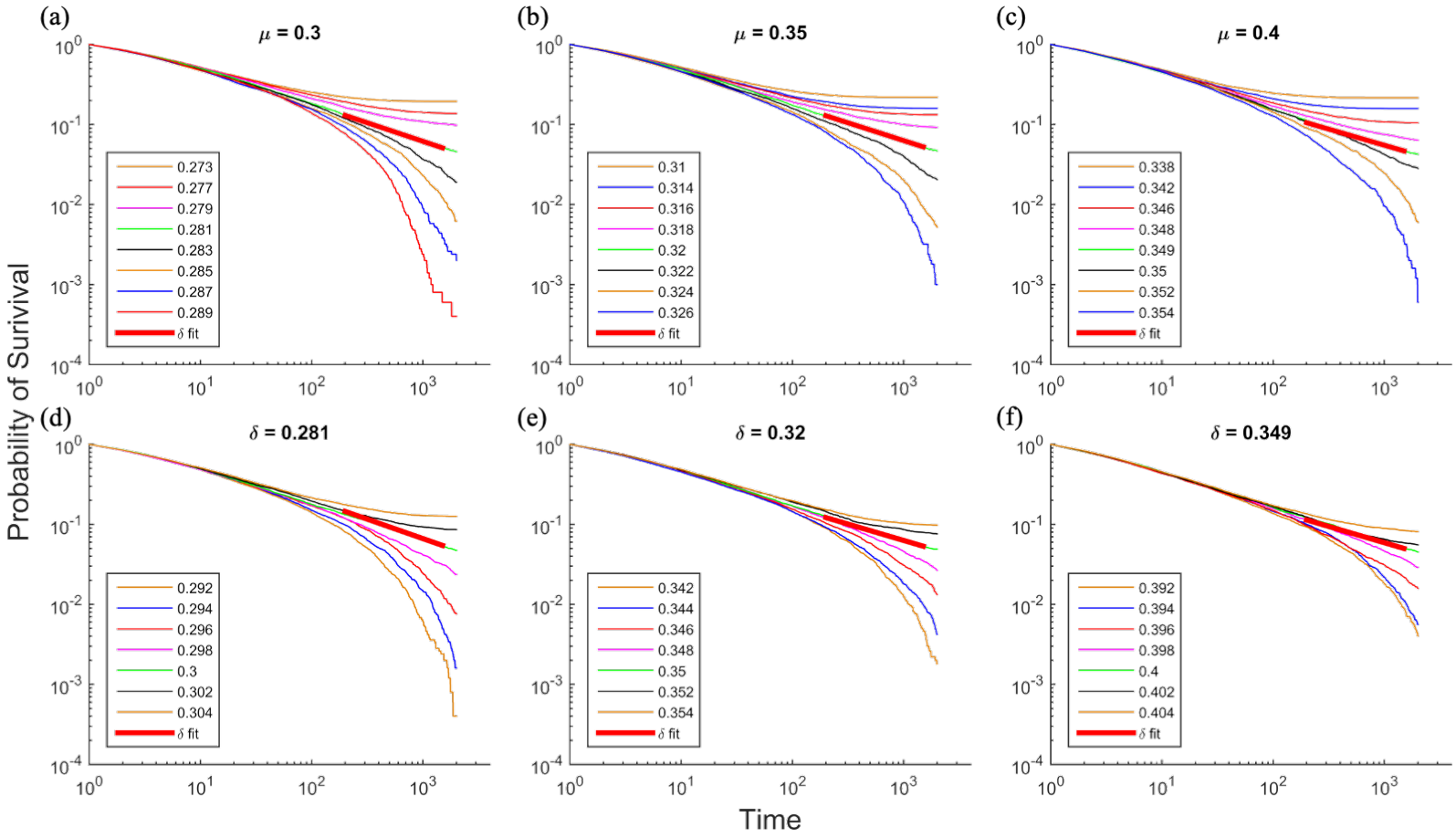

Figure 2.3 Survival Probability for μ and δ. Average probability of survival for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively). Legends show control parameter value. Thick red plots demonstrate the fit line, with the slope defining the survival probability exponent δ_e.

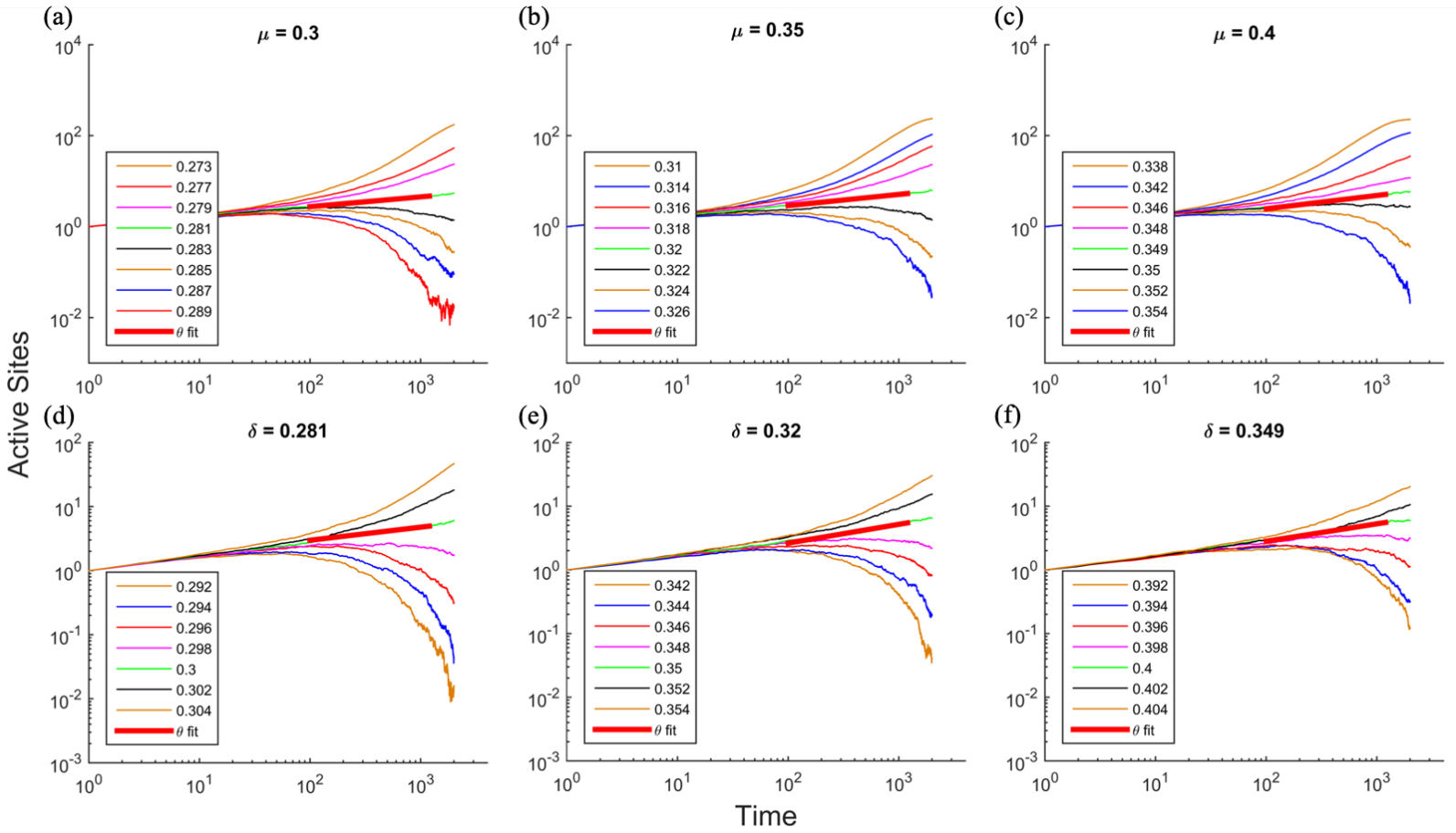

Figure 2.4 Number of Active Sites for μ and δ. Average number of active sites for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively) are shown. Legends show control parameter value. Thick red plots demonstrate the fit line, with the slope defining the critical slip exponent θ.

Figures 2.5 and 2.6 present the results for critical relations δ_e and θ for the control parameters μ and δ_{max}, with all system parameters the same as described for Figures 2.3 and 2.4 where *a-c* corresponds to the control parameter $\delta_{\max}$ for fixed values of μ = 0.3, 0.35 and 0.4 respectively, and *d-f* correspond to the control parameter μ for fixed values of δ = 0.636, 0.739, and 0.818 respectively, with a time series from t = [0, 2000] and averaged over 5000 simulations for both figures. The results are similar to those shown in Figures 2.3 and 2.4, with two exceptions. First, while the fluctuations of number of active sites, *N(t)* for both δ and δ_{max} are greater in the absorbing phase, the comparison of Figures 2.4 and 2.6 shows the results of δ_{max} demonstrate higher levels of fluctuations for each control parameter value in the active and absorbing phases. The results correlate to the increased disorder associated with δ_{max}. Figures 2.3 and 2.5, do not show the same fluctuation increase for δ_{max}, supporting the notion from Henkel *et al* (2008) that *N(t)* is a more sensitive measurement than $P_s(t)$. Secondly, the behavior around the transition spans further across the phase space for δ_{max}, noted by the larger intervals in control parameter values presented. The difference of highest and lowest δ values plotted in Figures 2.3 and 2.4 is 0.016 for each value of μ, and a difference in μ values of 0.012 for each value of δ. In Figures 2.5 and 2.6, the comparable difference of highest and lowest δ_{max} values is 0.028 for μ=0.3, 0.032 for μ=0.35 and 0.4, and a difference in μ values of 0.024 for δ_{max}=0.636, and 0.028 for δ_{max}=0.739 and 0.8181. Though the plotted ranges are larger for the plots of δ_{max} than δ, comparisons of Figures 2.3 and 2.5 and Figures 2.4 and 2.6 do not suggest that the increased ranges of δ_{max} include larger sections of the critical regime.

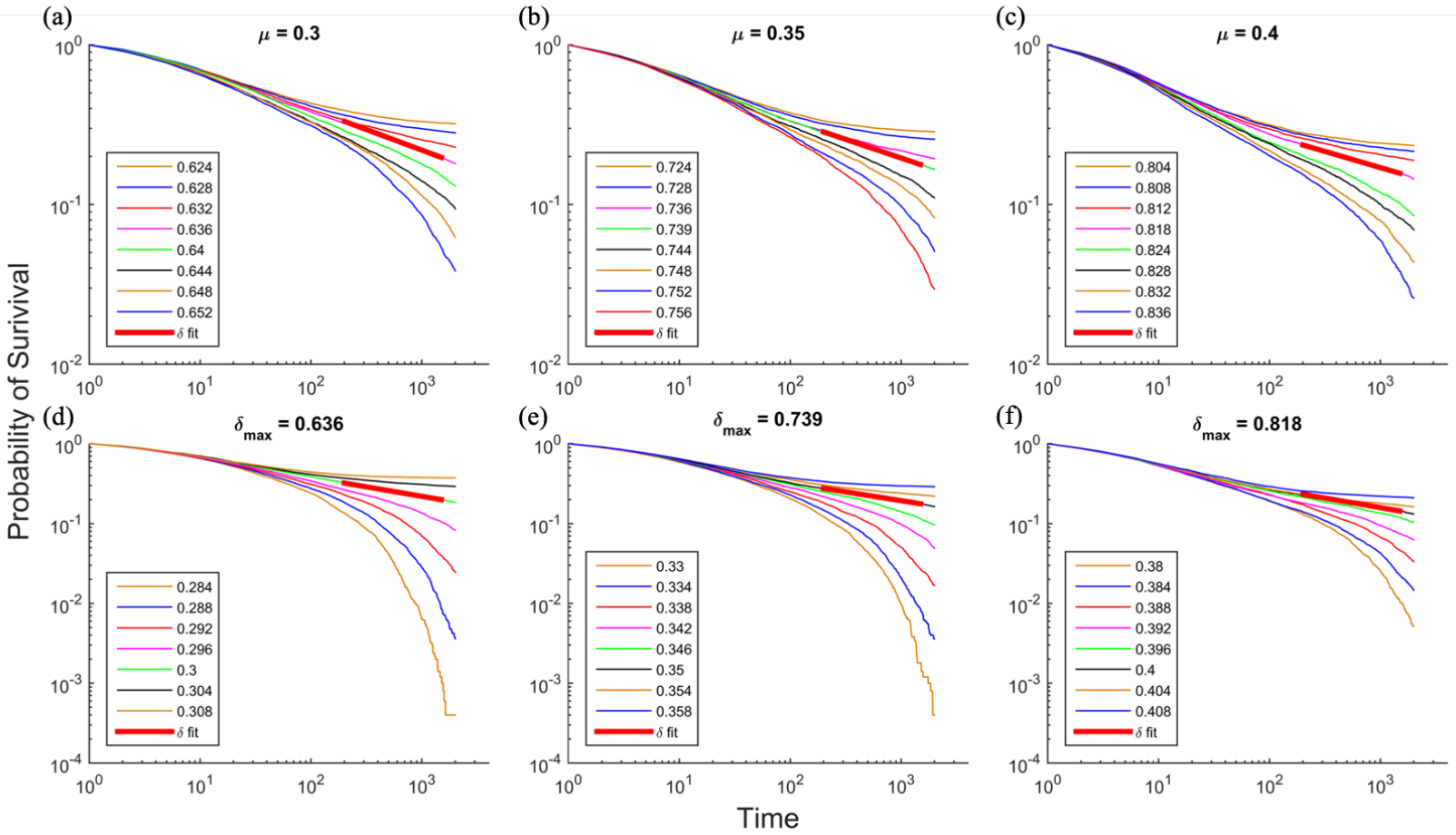

Figure 2.5 Survival Probability for μ and δ_{max}. Average probability of survival for control parameter δ_{max} with values of μ = 0.3, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of δ_{max} = 0.636, 0.739, and 0.818 (*d*, *e*, and *f* respectively). Legends show control parameter value. Thick red plots demonstrate the fit line, with the slope defining the survival probability exponent δ_e.

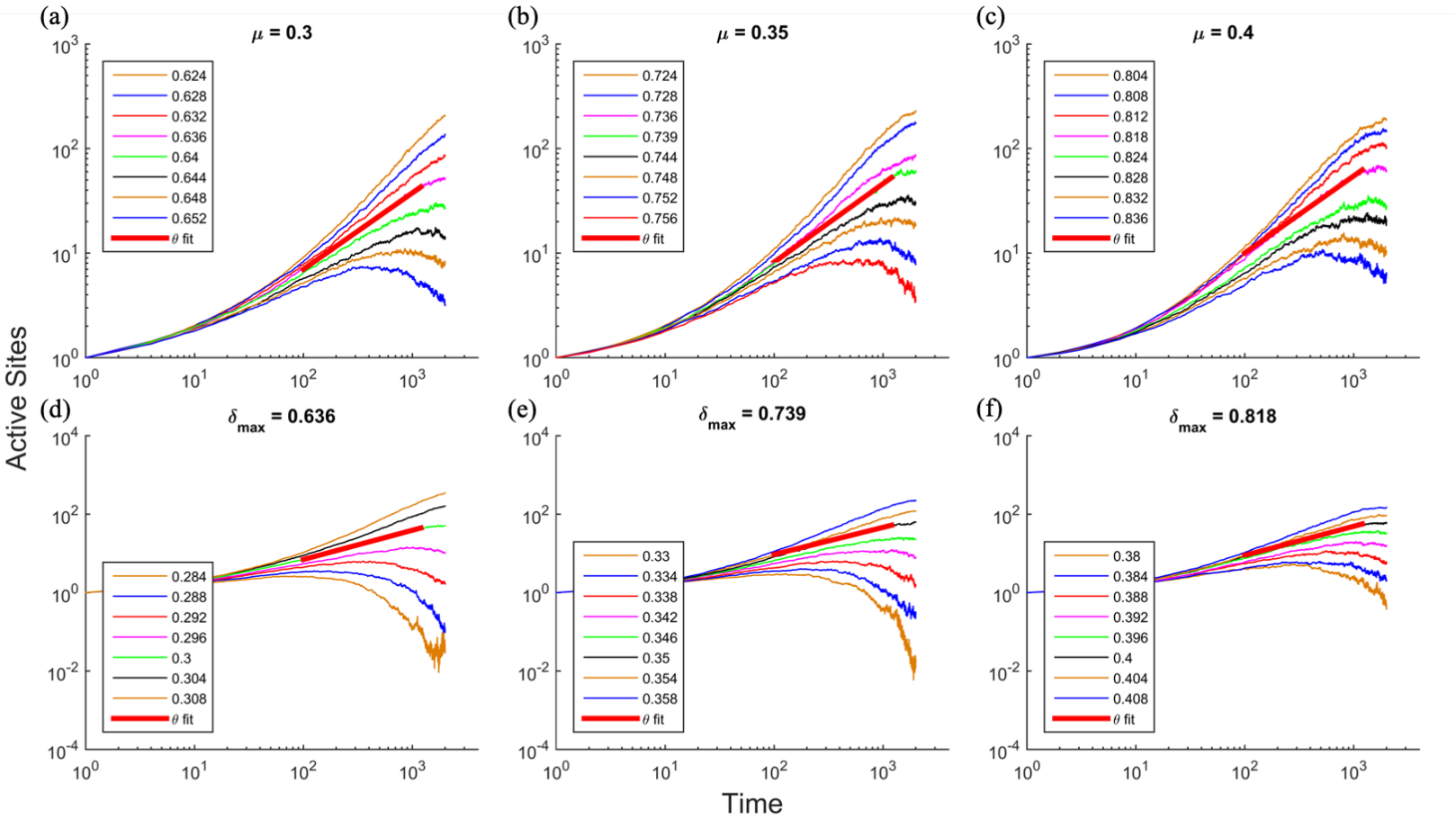

Figure 2.6 Number of Active Sites for μ and δ_{max}. Average number of active sites for control parameter δ_{max} with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta_{max} = 0.636$, 0.739, and 0.818 (*d*, *e*, and *f* respectively). Legends show control parameter value. Thick red plots demonstrate the fit line, with the slope defining the critical slip exponent θ.

The mean radius of gyration, $R^2(t)$, is presented for control parameters δ in Figure 2.7*a-c* and μ in Figure 2.7*d-f*. Figure 2.8*a-c* similarly demonstrates radius of gyration for δ_{max} and Figure 2.8*d-f* for μ. In both scenarios, all parameters near the critical point demonstrate nearly identical trends, but with slightly larger deviations near the end of the time series for results holding the death parameter constant and μ acting as the control parameter, an expected result as the increased mutation rate dictates spatial growth. Power law behavior is apparent in this system, regardless of death parameter value.

Table 2.1 presents the calculated values for the critical exponents δ_e, θ, and $\tilde{z}$ followed by accepted values for each exponent for 2 + 1 dimensional directed percolation (Grassberger & Yi-Cheng Zhang, 1996). When death parameter δ is used, measured exponents δ_e and θ are very similar to accepted values for directed percolation. As μ_c and δ_c increase, however, the gap between the accepted and measured values of δ_e and θ grows. Measurements of the spreading exponent, $\tilde{z}$, show a greater difference from the accepted value than that of δ_e and θ, with the gap between the accepted and measured values increasing as μ_c and δ_c increase, similar to the results of δ_e and θ. Results corresponding to δ_{max} however, do not match with accepted values of δ_e and θ for directed percolation, and demonstrate no consistent trend across the phase space. Interestingly, values of $\tilde{z}$ align more closely with accepted values for δ_{max} than for δ. Calculations for δ_e, θ, and $\tilde{z}$ are time dependent, but not dependent on the off-critical measure, therefore the similar scaling behavior of each exponent for each critical point pair is expected.

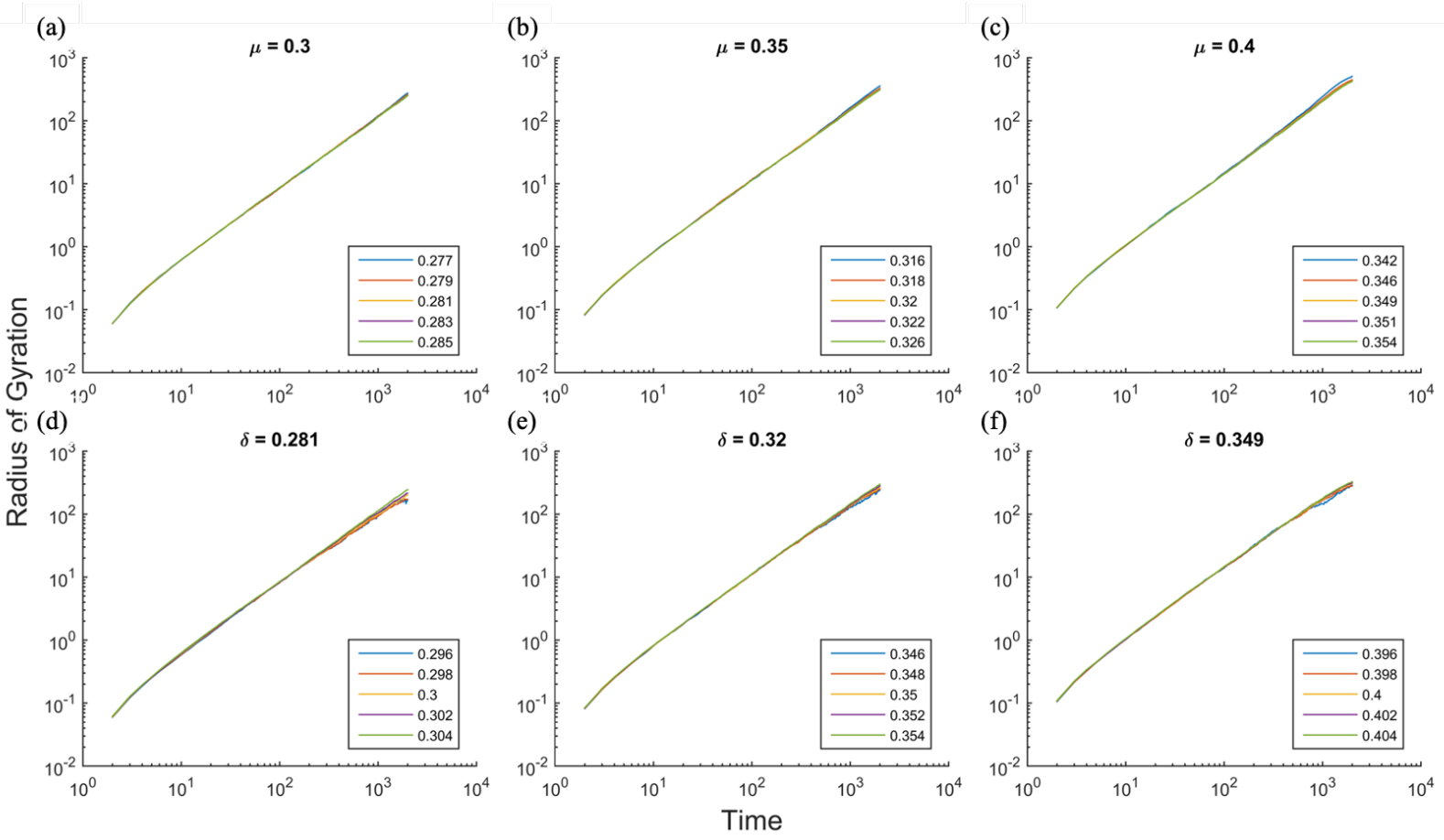

Figure 2.7 Radius of Gyration for μ and δ. Radius of Gyration for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively) are shown. Legends show control parameter value.

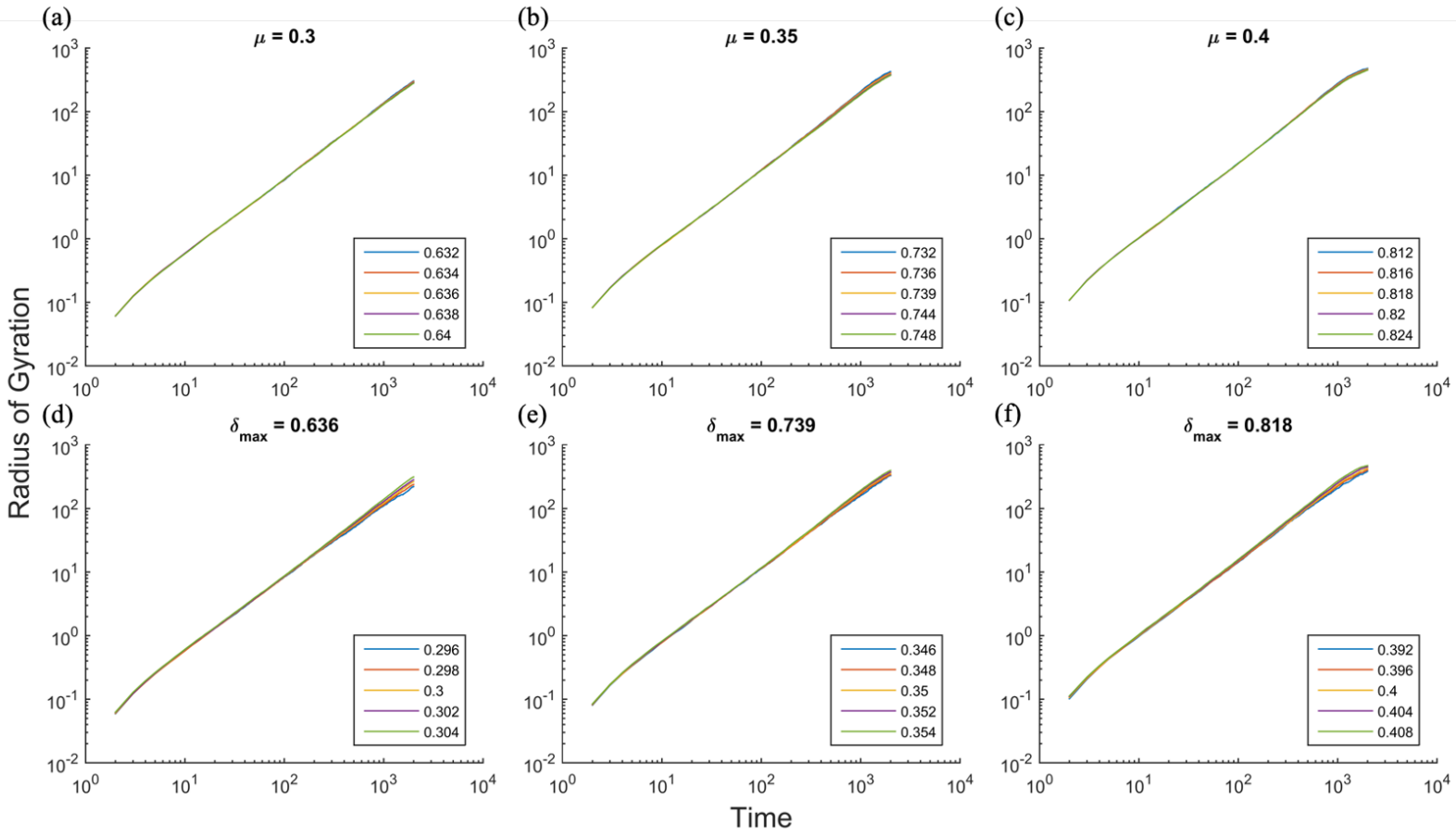

Figure 2.8 Radius of Gyration for μ and δ_{max}. Radius of Gyration for control parameter δ_{max} with values of μ = 0.3, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of δ_{max} = 0.636, 0.739, and 0.818 (*d*, *e*, and *f* respectively). Legends show control parameter value.

Table 2.1 Survival Probability, Slip, and Spreading Exponents

Critical Parameter Values	Control Parameter	δ_e	θ	$\tilde{z}$
$(\mu_c, \delta_c) = (0.3, 0.281)$	δ	0.4637(71)	0.2167(71)	1.079(14)
	μ	0.4861(58)	0.2043(86)	1.077(15)
$(\mu_c, \delta_c) = (0.35, 0.320)$	δ	0.4444(39)	0.2462(61)	1.048(18)
	μ	0.4090(43)	0.293(17)	1.059(18)
$(\mu_c, \delta_c) = (0.4, 0.239)$	δ	0.4026(63)	0.2925(82)	1.032(21)
	μ	0.4068(96)	0.2734(68)	1.053(19)
$(\mu_c, \delta_{max,c}) = (0.3, 0.636)$	δ_{max}	0.2582(40)	0.729(10)	1.1668(66)
	μ	0.2379(27)	0.7452(90)	1.1741(46)
$(\mu_c, \delta_{max,c}) = (0.35, 0.739)$	δ_{max}	0.2339(16)	0.751(13)	1.177(14)
	μ	0.2210(38)	0.690(18)	1.178(16)
$(\mu_c, \delta_{max,c}) = (0.4, 0.818)$	δ_{max}	0.2027(40)	0.736(16)	1.153(33)
	μ	0.2398(39)	0.741(12)	1.158(25)
Accepted Values (J. Wang *et al*., 2013)		0.451(3)	0.229(3)	1.1322(1)*

* $\tilde{z}$ calculated via $z = 2/\tilde{z}$; accepted value z = 1.7665(2)

The mean cluster mass is shown for μ and δ in Figure 2.9, and for μ and δ_{max} in Figure 2.10. In each subplot, the data shown with blue squares represents the calculated mean cluster mass, where the number of active sites is summed for each simulation, and then averaged over 5000 simulations beginning from a single organism at each off critical measure. As discussed above, the domain of the critical regime increases with increasing values of μ_c (and corresponding $\delta_c/\delta_{max,c}$). The ranges of off-critical measures, listed in Table 2.2, were determined by the range in system observables seen as the span along the y-axis in Figure 2.3 through Figure 2.6, with the intention to include the same span of the critical regime around each critical point pair. The fit line demonstrates the linear best fit of the double-logarithmic plot, the slope of which is the calculated exponent γ.

Comparison of the two figures demonstrates a more accurate fit of power-law behavior when the system is dictated by δ, as seen by the near linear data trends in Figure 2.9, as compared to the concave shape of the data trends seen in Figure 2.10 for δ_{max}.

Table 2.2 Ranges of Off-Critical Measures for *M*, *T*, *S*, and *V*

Constant parameter	Range	Increment
$\mu = 0.3$	$\delta = [0.282, 0.292]$	0.001
$\delta = 0.281$	$\mu = [0.289, 0.299]$	0.001
$\mu = 0.35$	$\delta = [0.321, 0.331]$	0.001
$\delta = 0.32$	$\mu = [0.339, 0.349]$	0.001
$\mu = 0.4$	$\delta = [0.35, 0.36]$	0.001
$\delta = 0.349$	$\mu = [0.389, 0.399]$	0.001
$\mu = 0.3$	$\delta_{max} = [0.638, 0.660]$	0.002
$\delta_{max} = 0.636$	$\mu = [0.278, 0.298]$	0.002
$\mu = 0.35$	$\delta_{max} = [0.74, 0.762]$	0.002
$\delta_{max} = 0.739$	$\mu = [0.328, 0.348]$	0.002
$\mu = 0.4$	$\delta_{max} = [0.820, 0.864]$	0.004
$\delta_{max} = 0.818$	$\mu = [0.378, 0.398]$	0.002

Figure 2.11 through Figure 2.16 shows data trends and fit lines, with the slope of the fit lines corresponding to the critical relation exponents listed in Table 2.3, for mean survival time (Figure 2.11 corresponding to δ, and Figure 2.12 for δ_{max}), mean cluster size, (Figure 2.13 for δ, and Figure 2.14 for δ_{max}) and mean spatial volume (Figure 2.15 for δ, and Figure 2.16 for δ_{max}). In all cases, 5000 simulations were performed for each value of the off-critical measure (as previously described for the calculation of mean cluster mass). Each data point in Figures 2.11 and 2.12 represents the summation of $P_s(t)$ for each simulation averaged over all simulations for each off-critical measure to

determine the mean survival time, T. Mean cluster size, S, shown in Figures 2.13 and 2.14 are calculated using Equation 21, and averaged over all simulations for each off critical measure. The data shown in Figures 2.15 and 2.16 represents the mean spatial volume calculated with Equation 20 and averaged over all simulations for each off-critical measure. When examining T, S, and V, the trend seen in M (Figures 2.9 and 2.10) with respect to the change in linearity of the data reemerges; the log-log plots for δ are more linear than those for δ_{max}.

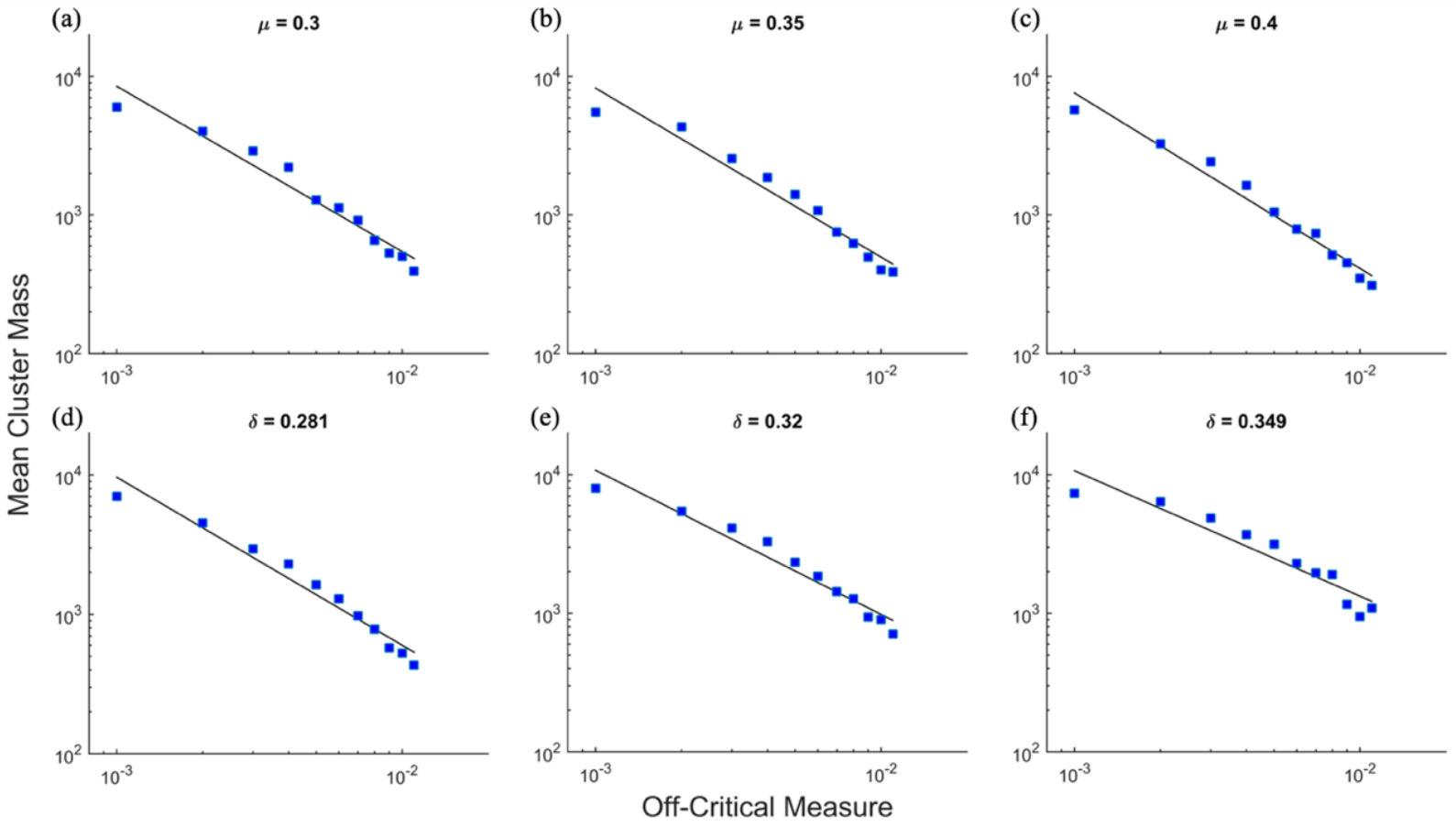

Figure 2.9 Mean Cluster Mass for μ and δ. Mean Cluster Mass for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_c = 0.281$, 0.32, and 0.349 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent γ are shown in black.

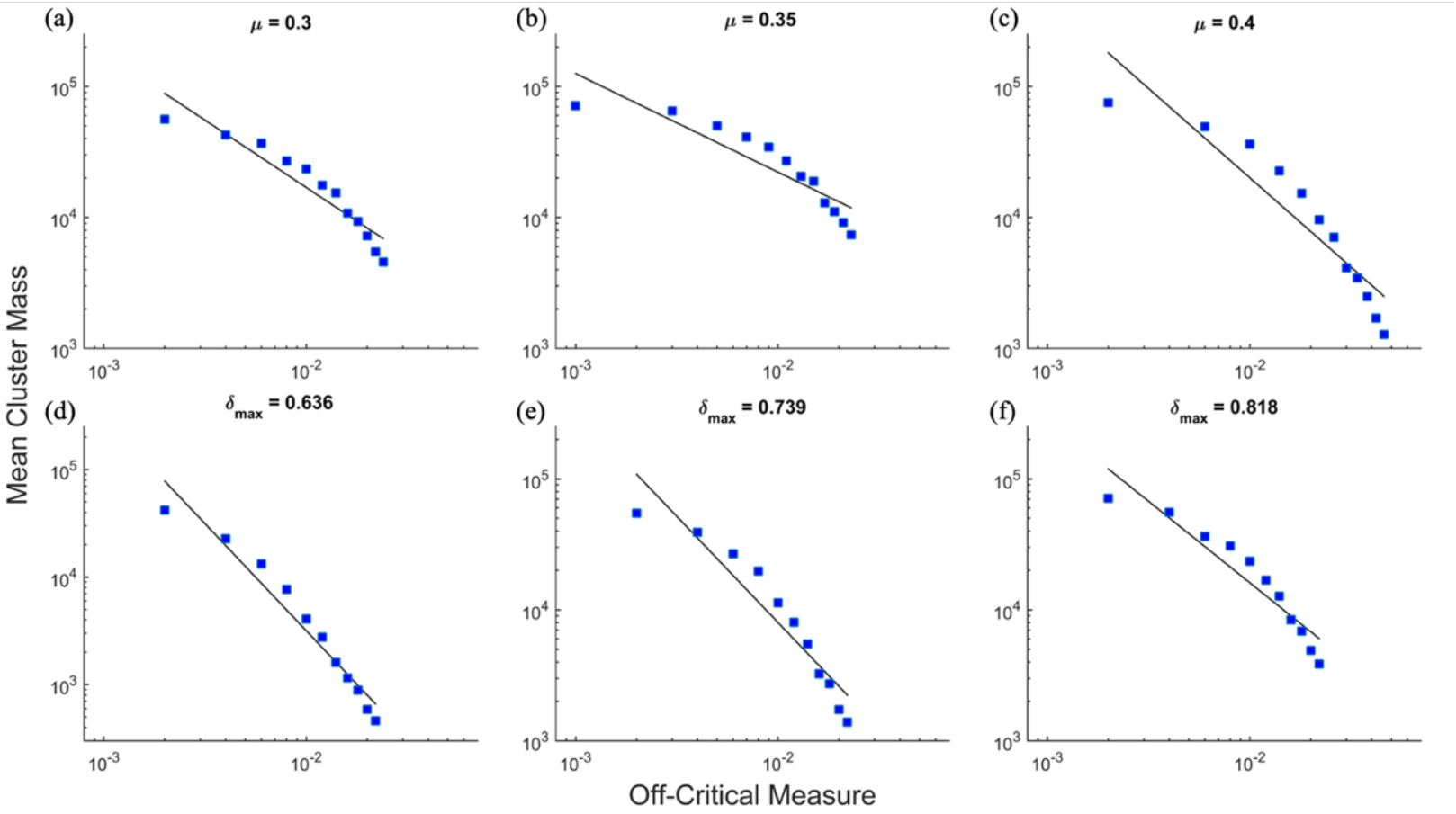

Figure 2.10 Mean Cluster Mass for μ and δ_{max}. Mean Cluster Mass for control parameter δ_{max} with values of μ = 0.3, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of δ_{max} = 0.636, 0.739, and 0.818 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_{max,c}$ = 0.636, 0.739, and 0.818 (*a*, *b*, and *c* respectively) and μ_c = 0.3, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent γ are shown in black.

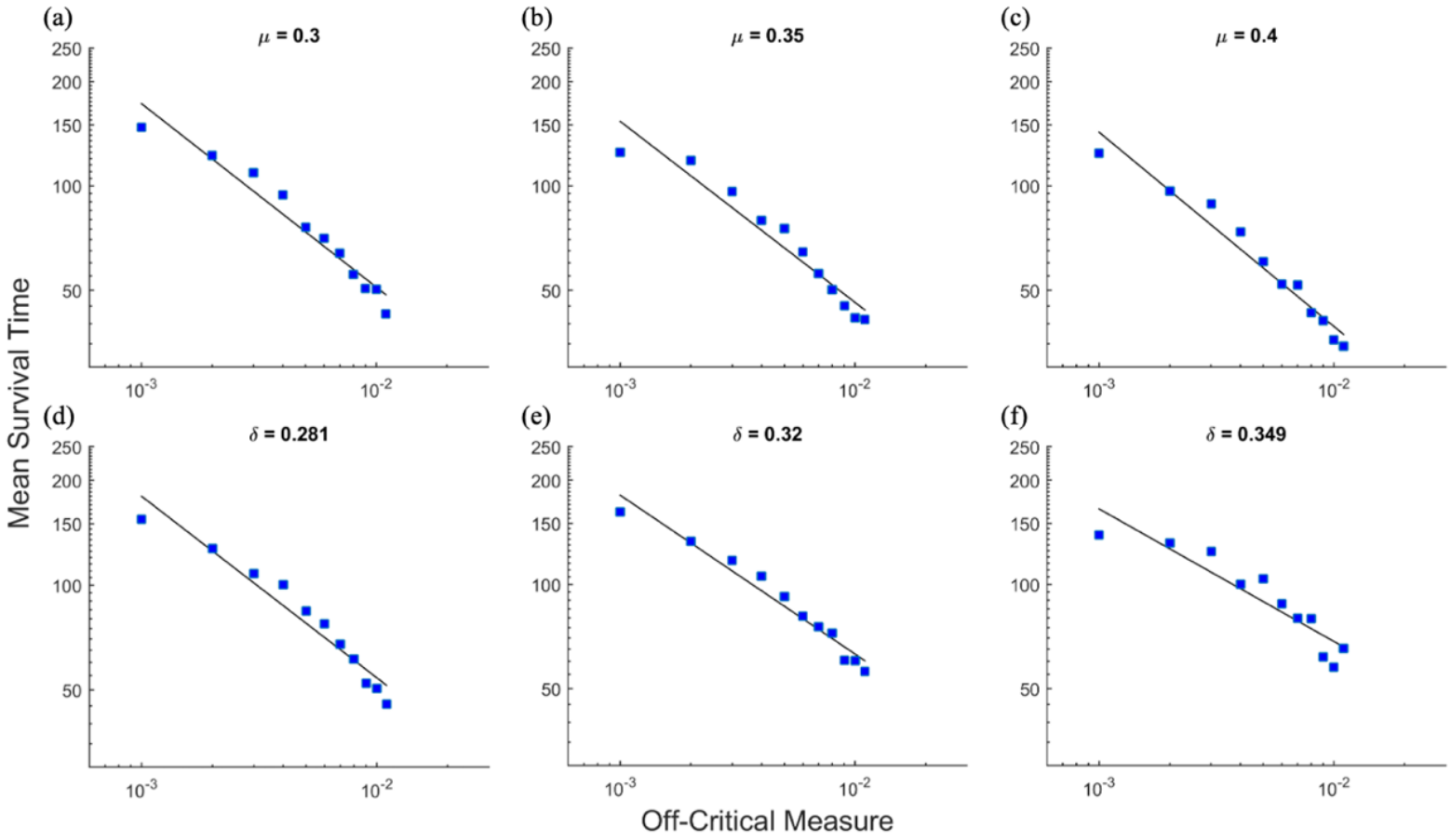

Figure 2.11 Mean Survival Time for μ and δ. Mean Survival Time for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively) are shown. The off-critical measure is the parameter distance from $\delta_c = 0.281$, 0.32, and 0.349 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent τ are shown in black.

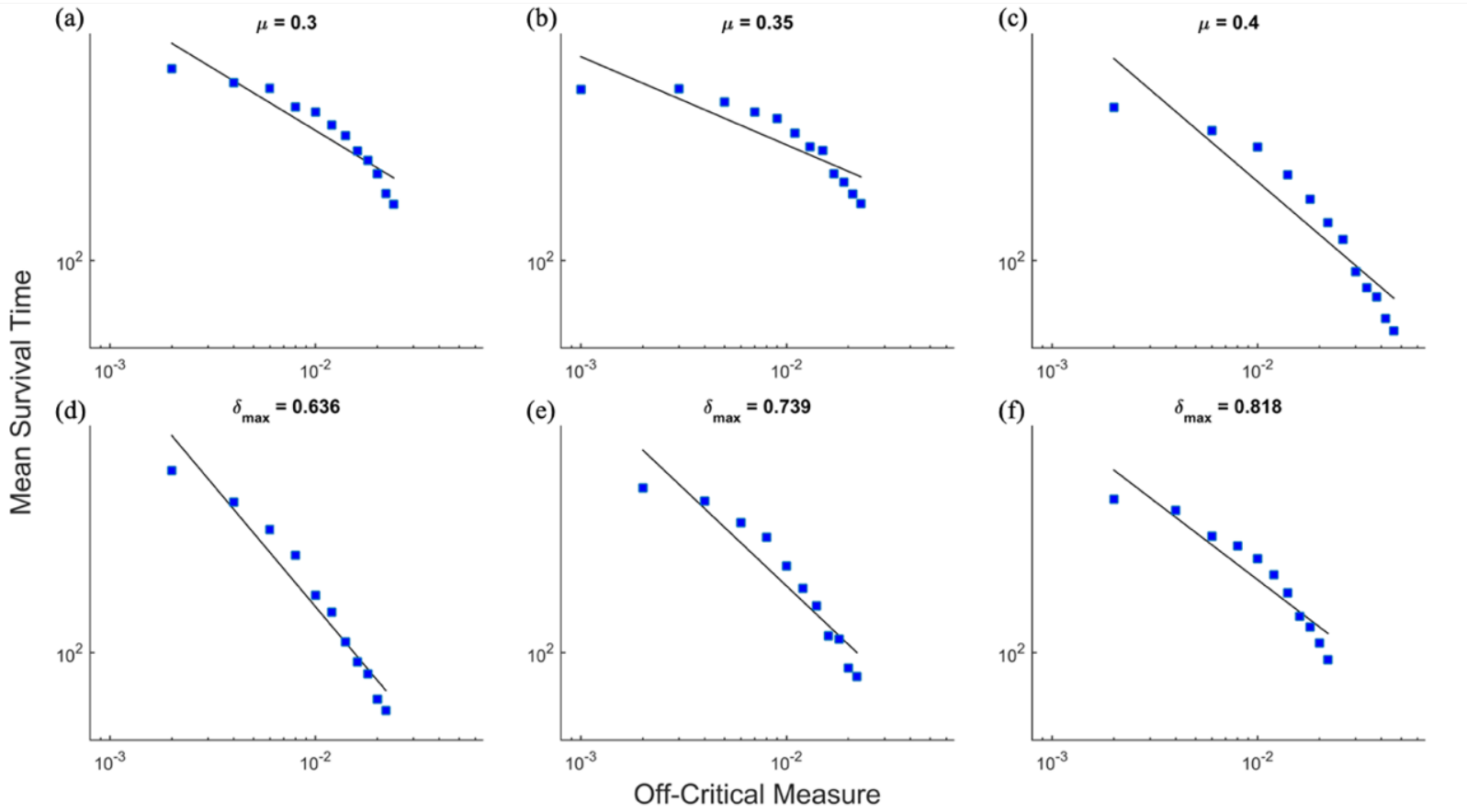

Figure 2.12 Mean Survival Time for μ and δ_{max}. Mean Survival Time for control parameter δ_{max} with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta_{max} = 0.636$, 0.739, and 0.818 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_{max,c} = 0.636$, 0.739, and 0.818 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent τ are shown in black.

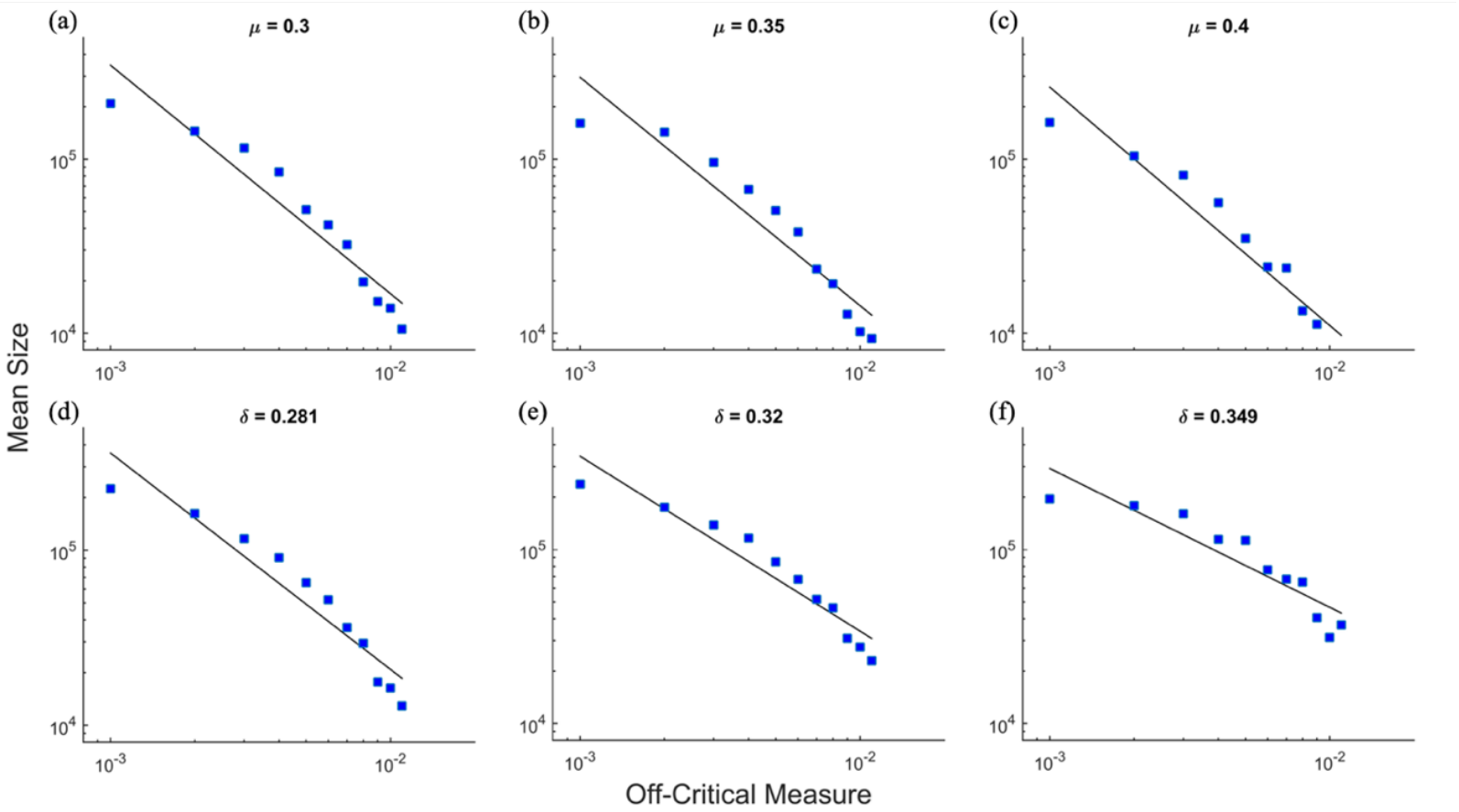

Figure 2.13 Mean Cluster Size for μ and δ. Mean Cluster Size for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_c = 0.281$, 0.32, and 0.349 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent σ are shown in black.

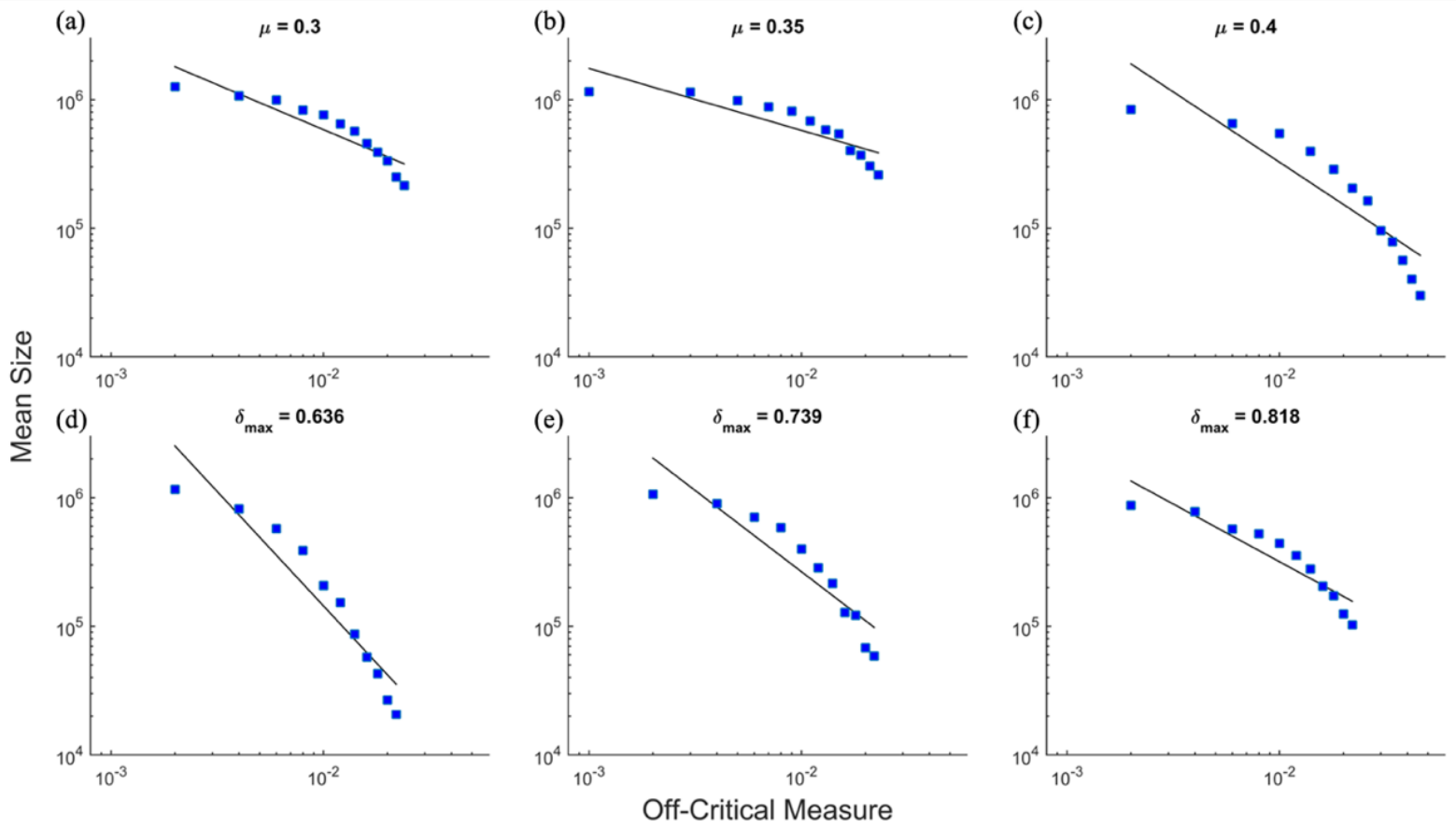

Figure 2.14 Mean Cluster Size for μ and δ_{max}. Mean Cluster Size for control parameter δ_{max} with values of μ = 0.3, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of δ_{max} = 0.636, 0.739, and 0.818 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_{max,c}$ = 0.636, 0.739, and 0.818 (*a*, *b*, and *c* respectively) and μ_c = 0.3, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent σ are shown in black.

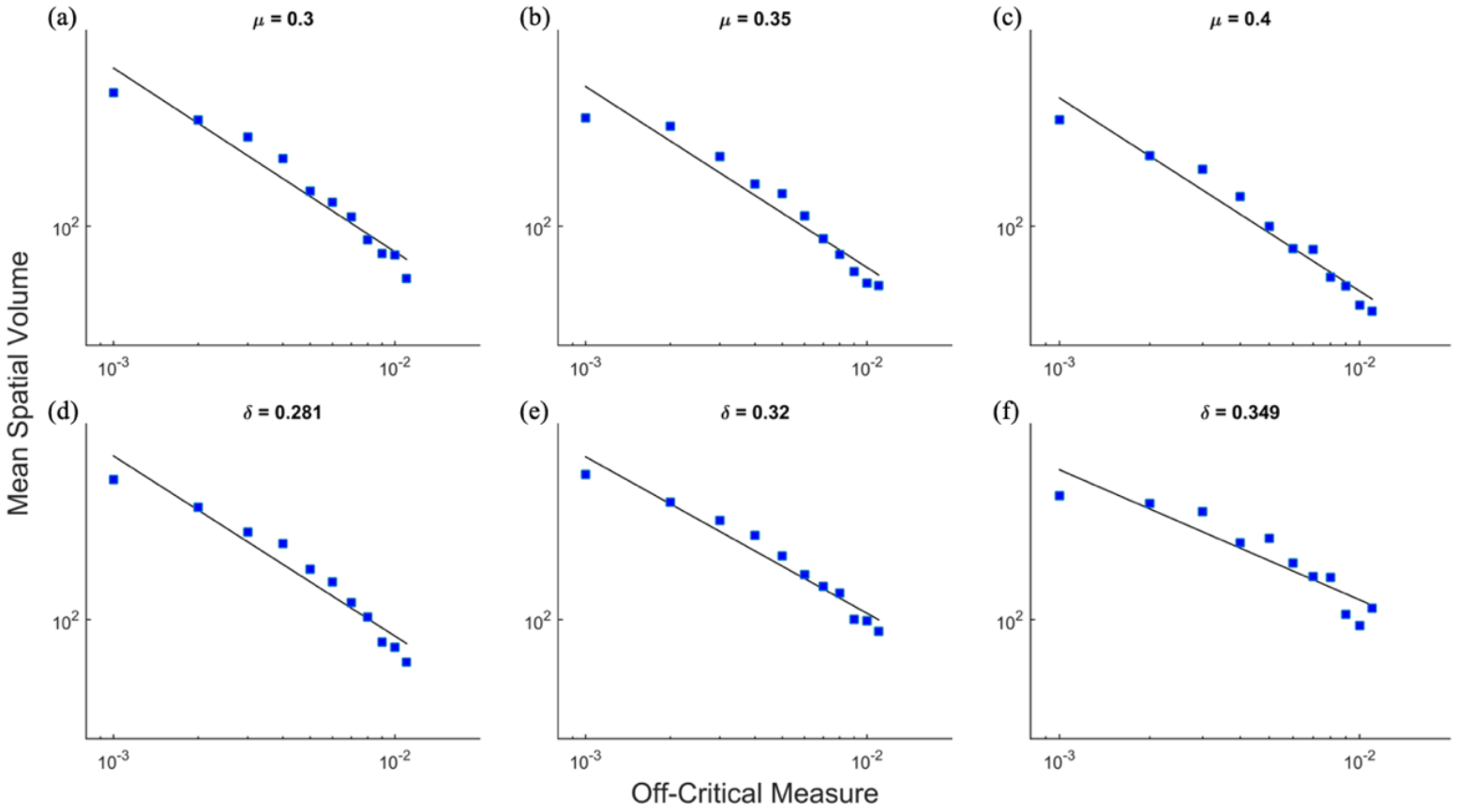

Figure 2.15 Mean Spatial Volume for μ and δ. Mean Spatial Volume for control parameter δ with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta = 0.281$, 0.32, and 0.349 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_c = 0.281$, 0.32, and 0.349 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent ν are shown in black.

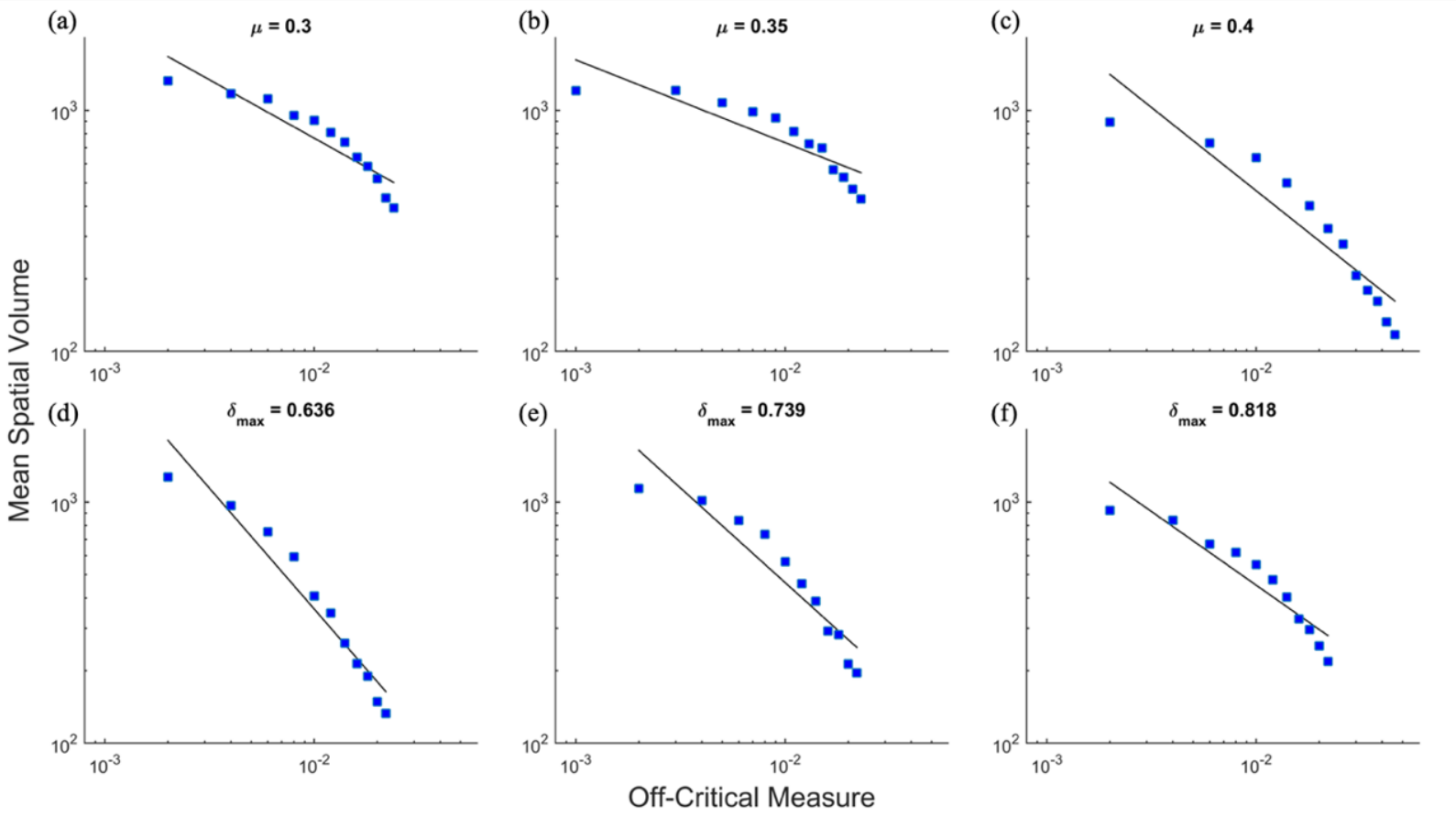

Figure 2.16 Mean Spatial Volume for μ and δ_{max}. Mean Spatial Volume for control parameter δ_{max} with values of $\mu = 0.3$, 0.35, and 0.4 (*a*, *b*, and *c* respectively) and for control parameter μ with values of $\delta_{max} = 0.636$, 0.739, and 0.818 (*d*, *e*, and *f* respectively). The off-critical measure is the parameter distance from $\delta_{max,c} = 0.636$, 0.739, and 0.818 (*a*, *b*, and *c* respectively) and $\mu_c = 0.3$, 0.35 and 0.4 (*d*, *e*, and *f* respectively). Fit lines with the slope defining the critical relation exponent ν are shown in black.

Table 2.3 Critical Relation Exponents

Critical Parameter Values	Control Parameter	γ	τ	σ	ν
(μ_c, δ_c) = (0.3, 0.281)	δ	1.193(81)	0.529(38)	1.313(124)	0.612(45)
	μ	1.205(75)	0.521(39)	1.236(121)	0.601(47)
(μ_c, δ_c) = (0.35, 0.320)	δ	1.219(84)	0.522(43)	1.312(138)	0.604(52)
	μ	1.040(73)	0.458(29)	1.004(96)	0.522(35)
(μ_c, δ_c) = (0.4, 0.239)	δ	1.266(81)	0.560(36)	1.369(118)	0.646(42)
	μ	0.905(93)	0.383(47)	0.798(106)	0.434(54)
$(\mu_c, \delta_{max,c})$ = (0.3, 0.281)	δ_{max}	1.027(115)	0.428(54)	0.699(98)	0.483(62)
	μ	1.989(139)	0.841(63)	1.778(186)	1.001(80)
$(\mu_c, \delta_{max,c})$ = (0.35, 0.739)	δ_{max}	0.753(137)	0.303(64)	0.481(106)	0.343(72)
	μ	1.619(165)	0.668(71)	1.263(166)	0.785(87)
$(\mu_c, \delta_{max,c})$ = (0.4, 0.818)	δ_{max}	1.363(208)	0.605(91)	1.093(206)	0.691(108)
	μ	1.243(132)	0.539(58)	0.897(117)	0.611(69)
Calculated Accepted Values* (J. Wang et al., 2013)		1.585	0.707	2.165	0.878

* Values calculated using Eq. 11-14 with accepted values $\beta = \beta' = 0.580(4)$, $\upsilon_{\parallel} = 1.287(2)$, and $\upsilon_{\perp} = 0.729(1)$.

Results for γ, τ, σ, and ν are presented for all critical point pairs with both mutation and death parameters acting as the control parameter in Table 2.3. Results for system parameters such that μ is held constant and δ acts as control parameter exhibit a reasonable consistency across all four critical relation exponents, whereas when μ acts as the control parameter and $\delta = \delta_c$, all four measured exponents decrease as μ increases. Interestingly, this decrease seems nearly linear, though with only three points, this may not be consistent across the whole parameter space. Results with a constant μ, with δ_{max} acting as control parameter, show a decrease in all four exponents from $\mu = 0.3$ to $\mu =$

0.35, but an increase in all cases with the highest measured values of all four exponents when $\mu = 0.4$. Similar to the results with a fixed δ, when δ_{max} was constant and μ was the control parameter, the values of all four exponents decreased as μ increased. None of the measured values align with the accepted values for directed percolation for any parameter at any of the measured transition points.

2.5. DISCUSSION

Scott (2013) approached the classification of the system into a universality class of directed percolation (DP) by holding δ_{max} at 70% and varying μ as the control parameter to examine the critical phase transition. The mutability μ was varied to find the transition from survival to extinction of the population. This approach utilized the measurement of the critical relation α, corresponding to the population density scaling, $\rho(t)$, and the direct measurement of critical exponents $\nu_{//}$ and $\nu_{\perp}$. These measurements were calculated using simulations of various landscape size, beginning with a completely full landscape, and observing the population decline to a steady state in the survival phase, or the population's transition to extinction in the absorbing phase. Observations and calculations to determine the system's universality classification were conducted in the survival phase of the critical regime (Scott, 2014; Scott et al., 2013).

One major assumption made by Scott was that the model demonstrated time-reversal symmetry, a known feature of the DP universality class. Time-reversal symmetry, also known as rapidity symmetry, holds when "the density $\rho(t)$ starting from a fully occupied lattice and the survival probability $P(t)$ for clusters grown from a seed are exactly equal for all t" (Hinrichsen, 2006). In other words, the dynamical behavior of the

density decay from a completely full (completely active) landscape scales exactly as the probability that the system will survive through time from a single active site. When time-reversal symmetry is present in a system, the scaling properties of density decay, corresponding to exponent β, and of survival probability, corresponding to β', are equal, making $\beta = \beta'$. Because the β, $\nu_{//}$, and $\nu_{\perp}$ matched the accepted values of DP, the assumption that time-reversal symmetry exists in the model, and therefore that $\beta = \beta'$, was reasonable.

The results presented above demonstrate that the system controlled by death parameter δ exhibits possible time-reversal symmetry for calculations in time at the transition point. The critical exponents, δ_e and θ, match closely with accepted values of directed percolation. Measurements of $\tilde{z}$ provide results similar to the accepted value of directed percolation but are slightly below the accepted value. The results for critical exponents δ_e and θ, when the system is controlled by δ_{max}, however, do not resemble accepted values for directed percolation. These initial results at the transition point suggest the possibility of time-reversal symmetry with the use of δ, but not with δ_{max} because the critical exponents in the former case more closely fit with the expected DP values.

The most straightforward explanation for this difference correlates with a violation, in the δ_{max} case, of the DP conjecture requirement that the system cannot have any special attributes, such as quenched randomness. Simulations with death parameter δ_{max} demonstrate quenched disorder, since δ_{max} acts as a fluctuating variable, but the fluctuations do not vary over time. The use of δ, for comparison, still provides a level of

fluctuation as the number of organisms removed varies from one generation to the next, but because this fluctuation is directly connected to population size, with the same percentage of removal at every generation, it is not considered a quenched disorder. This violation of the DP conjecture for δ_{max} becomes the most direct explanation for the mismatch of results with accepted directed percolation values. However, the results from Table 2.1 showing the critical exponent values of δ_e and θ, and the small deviation from accepted values for $\tilde{z}$ suggest that when the system is controlled by δ instead of δ_{max}, it could indeed undergo a directed percolation transition.

Previous work measuring scaling behavior in the absorbing phase only calculated M, T, S, and V using their defining functions, which resulted in a lack of clarity regarding the physical definition of each measurable. For this reason, measurement of each value through direct observation was impossible without establishing specific definitions. As an example, the accepted definitions for M and S only include each measurable's name, "mean cluster mass" and "mean cluster size", accompanied by their defining equations (Equations 15 and 21 respectively) (Henkel *et al*., 2008; Hinrichsen, 2000b, 2006). One additional definition for mean cluster size that appears on occasion in the literature states that S is the average size of all clusters present, with the size being defined as the sum of all active sites in a cluster and a cluster defined as sites connected in time (Henkel *et al*., 2008; Nachtrab, 2011).

In the absorbing phase, all clusters are generated from a single seed, which by definition means that each simulation is defined as one cluster (all active sites in time connect to the single seed organism), and therefore the mean cluster size, S, is the total number of active sites in each simulation. The mathematical definitions for M, mean

cluster mass, simplify to the total number of active sites in a cluster (in this case a simulation) as shown in section 2.3.1 Equations 15-17. This may not be the case in all situations, but when examining absorbing phase, system-level properties with an initial population of one, application of the definitions from Henkel *et al.* suggest that *M* and *S* scale identically, as they are both defined by the number of organisms in a cluster. However, this cannot be true, or the scaling relations $\gamma = \nu_{\parallel} + d\nu_{\perp} - \beta - \beta'$ and $\sigma = \nu_{\parallel} + d\nu_{\perp} - \beta'$ (from Equations 11 and 14) would also be equal. Additionally, there appears to be no agreed-upon definition for mean spatial volume, *V,* other than Equation 20. Mean survival time, *T,* is the only well-defined measurable value, though this may be because survival time is a common term.

The results of the global dynamic scaling in the absorbing phase side of the critical regime do not indicate that this system belongs to the DP universality class for either death parameter. Though δ demonstrates behavior close to directed percolation in time at the critical point, the scaling behavior of the observables in the absorbing phase of the critical regime do not support directed percolation classification. For both δ and δ_{max}, findings suggest that neither transition can be classified as directed percolation. The results of γ, τ, σ, and ν in Table 2.2 suggest that the dynamics in the absorbing phase of the critical regime may be dictated by power law behavior, but not within the directed percolation universality. Results for δ demonstrate linearity on the double-logarithmic scale (see Figures 2.9, 2.11, 2.13, and 2.15), whereas the results of δ_{max} show a more significant curvature in the plotted data (see Figures 2.10, 2.12, 2.14 and 2.16). Two explanations may support the reasoning for this: first, the range of the off-critical measures is larger for δ_{max}, and therefore may reach too far into the absorbing phase, and

include data outside of critical behavior about the transition, and second, that the quenched randomness, though it had little effect in the active phase in Scott's work, has a greater effect on system behavior in the absorbing phase of the critical regime, and the system does not exhibit power law-behavior in the case of δ_{max} (Scott, 2014; Scott *et al*., 2013). The range of the off-critical measure was determined based on the system behavior in the neighborhood of the transition, as seen in Figures 2.5 and 2.6, in an effort to encompass similar ranges based on dynamical behavior. Quenched disorders have been shown to modify the universality class of a critical point, and to spread the transition over a larger interval of the control parameter, further supporting the larger range of off-critical measures in the case of δ_{max} and the change in scaling behavior between δ and δ_{max} (Grinstein & Luther, 1976; Hooyberghs *et al*., 2003; Vojta, 2003).

The results shown above demonstrate that the behavior about a critical point can be dependent on which parameter is acting as the system control parameter. Comparison of μ acting as control parameter with the results of δ or δ_{max} as control parameter for γ, τ, σ, and ν show only one case where the exponents linked to each control parameter were approximately equal. The results about the critical point (μ_c, δ_c) = (0.3, 0.281) are remarkably close for γ, τ, σ, and ν, but the investigation of the transition point at (0.35, 0.32) shows notable differences for μ compared to δ as seen in Table 2.3: all four exponents are lower for control parameter μ than δ. The results corresponding to (0.4, 0.349) show a similar pattern of lower exponents for μ than δ, with the difference between measured exponent for μ and δ being nearly twice the comparable difference from (0.35, 0.32) for all four exponents. In all three cases, the results of all four

exponents show little variation when δ is the control parameter, but as μ increased, the exponents collectively decrease, which suggests that the critical behavior driven by μ is dependent on the value of δ, but behavior driven by δ demonstrates no such dependence on μ. Comparing the results of μ_c with corresponding $\delta_{max,c}$ demonstrates no notable similarities. The results for δ_{max} do not show consistency for any of the four exponents, in stark contrast to the results for δ. The only discernable pattern for δ_{max} is that all four exponents decrease from transition point (μ_c, $\delta_{max,c}$) = (0.3, 0.636) to (0.35, 0.739), but then all increase at (0.4, 0.818), where the exponents all have their highest values. The results for δ_{max} with μ as control parameter parallel the results for δ with μ as control parameter with a consistent decrease in all four exponents as μ increases, though the measured exponents for μ in each system were not equal. The results as μ increases may be demonstrating a violation of the DP conjecture that the dynamical rules are short-ranged; it is possible that the increased spatial range of new offspring on the phenotypic space defined by larger μ values outreach the "short-range" dynamics required for the directed percolation universality class.

2.6. CONCLUSION AND FUTURE WORK

Dynamics of the absorbing phase in the critical regime were investigated with aim to determine whether the presented model belongs to the directed percolation universality class in the presence and absence of quenched disorder, and to gain a better understanding of the differences in critical behavior based on control parameter. It was demonstrated that the model does not exhibit time-reversal symmetry (an assumption used in previous work to classify this system), and it further does not present evidence of

directed percolation in the absorbing phase of the critical regime, with quenched disorder (δ_{max}) and without quenched disorder (δ). The comparison of dynamics about the transition point of two control parameters, mutation size and the death parameter, demonstrated that transition behavior is dictated by the selection of control parameter, and in some cases (μ as the control parameter), the global dynamics of the system are also determined by the value of another parameter (δ or δ_{max}).

Examination of absorbing phase critical behavior in phase transitions is not commonplace, a possible reason for the lack of definitions in literature. Complex models can be limited by computational resources, and the ability to gain multiple critical relation values from a shared data set in the absorbing phase of a system is more efficient than the traditional approach requiring individual data sets for each critical exponent. Efforts to establish standardized definitions of mean cluster mass, mean cluster size, and spatial volume outside of their mathematical definitions may help strengthen the ties between model-based results and experimental results; values with well-defined definitions in the physical world will be more directly applicable to experimental data.

The work presented shows that directed percolation is not the correct universality classification for this model, but in the case of the model without a quenched randomness, power-law-like behavior still occurs. As multiple universality classes exist, the power-law-like behavior suggests that the model may fall into a different universality, though further investigation is required. Additionally, previous work has shown the model exhibiting continuum percolation which is spatially dependent, but only in cases where the landscape fitness is greater than 2 (King, 2015; King et al., 2017). The variation of critical relation exponents for different values of μ demonstrates a change in

scaling behavior that are spatially dependent. Examination of the spatial dynamics of model behavior with various mutation values may provide insight beneficial in determining population survivability.

3. MICROBIOLOGICAL POPULATIONS UNDER STRESS

As a result of the increasing environmental stressors of climate change, and the ecological effects of these stressors on ecosystems, identification of early warning signs of population collapse has become a priority (Carpenter, 2013; Dai *et al*., 2012, 2013, 2015; Dakos *et al*., 2019; Scheffer *et al*., 2009, 2012). Critical slowing down, a dynamical phenomenon associated with scale-free, power-law dynamics, has been suggested to accompany population decline in some studies, while others have observed population decline without any such critical signatures (Dai *et al*., 2015; Dakos et al., 2019; Rozek *et al*., 2017; Scheffer *et al*., 2009, 2012). Populations that experience critical slowing down recover more slowly from perturbations in the neighborhood of the system's tipping point between survival and population collapse (Dai *et al*., 2015).

Dai and colleagues have studied the population dynamics of yeast cultures to investigate the stability and resilience of populations subjected to environmental stressors (Carpenter, 2013; Dai *et al*., 2012, 2013, 2015). Examining populations from a nonlinear dynamics approach, they mapped the dynamics of stable and unstable populations under the influence of several environmental stressors including dilution factor (a proxy for death rate), nutrient concentration (a proxy for carrying capacity, using sucrose), and NaCl concentration (Dai *et al*., 2012, 2013, 2015). These studies resulted in a characterization of a population's stability and resilience, *i.e*., the ability to recover from a large environmental perturbation (Dai *et al*., 2015).

Section 2 outlined a computational approach to understanding population dynamics near "criticality" and examined system-level behavior of phase transitions. This

section investigates whether similar phase transition behavior exists in real biological populations due to an environmental stressor. The experimental approach used by Dai and colleagues has been adapted to investigate the decline of the budding yeast *Saccharomyces cerevisiae* and bacteria *Escherichia coli* populations under stress (Ordway *et al.*, 2020). The investigation focuses on the dynamics of yeast population decay in the presence of varying temperature and salt (NaCl) concentration, and the decay of *E. coli* populations in response to an assortment of antibiotics. All work with *E. coli* presented below was done by R. Fredrik Inglis and Holly Huelskamp, who collected the *E. coli* data presented and contributed significantly to the understanding declining population dynamics in relation to antibiotic mechanisms. Their work is presented here for completeness, with their permission. All the work discussed in this section appeared in Ordway *et al*. (2020), which can be found in APPENDIX B.

3.1. MATERIALS AND METHODS

3.1.1. *Saccharomyces cerevisiae* Experiments. All yeast experiments were performed using the yWO3 strain of *S. cerevisiae* (Ulbricht & Olivas, 2008). The selection of this strain was based on two factors: it is a well-documented wild-type laboratory strain, and it is neither thermophilic nor thermotolerant. Yeast cultures were placed under two environmental stressors, temperature and elevated NaCl concentration, to investigate each stressor's effect on population dynamics of the wild-type strain. For both stressors, *S. cerevisiae* was initially grown in standard medium of 10 g l^{-1} yeast extract, 20 g l^{-1} peptone and 2% dextrose (YEPD) at an optimum temperature of 30°C (Caspeta & Nielsen, 2015). Initially, these 50 ml cultures in liquid media were inoculated

from to an optical density (OD) of 0.0001, allowing multiple doublings in log phase over 24 h, and resulting in a large, but not saturated, concentration of cells, around an OD of 2.5. These starter cultures were inoculated into fresh media for both temperature stress and salt stress experimental cultures, as described below. All OD measurements were measured with a Turner visible spectrophotometer at 600 nm.

3.1.1.1. Temperature stress. Each initial culture was inoculated into 50 ml of YEPD into 250 ml flasks, resulting in the desired OD of 0.05. The samples were placed in orbital shaking water baths previously heated to the temperature of interest for a given set of measurements. Sample ODs were measured every 24 h over the course of eight days. At the time of each measurement, if a sample had increased by more than 0.01 OD, it was diluted back to an OD of 0.05 in a new 250 ml flask with YEPD, resulting in a final volume of 50 ml, and placed back into the water bath until the next measurement. If the sample had increased by less than 0.01 OD, the sample was placed back in the water bath until the next measurement. Population doubling rates were used as the metric of growth rates, where the number of doublings, n, is calculated using

$$n = \frac{\left(\log\left({}^{OD_{final}}/_{OD_{initial}}\right)\right)}{\log(2)} \tag{22}$$

where $OD_{initial}$ is the OD at the beginning of each 24 h cycle, and OD_{final} is the measured OD at the each of each 24 h growth period.

Optimal growth of *S. cerevisiae* is known to occur at 30°C, and exhibits a sharp decline in growth rate as temperatures exceed approximately 40°C (Caspeta & Nielsen, 2015; Mensonides *et al*., 2002). To investigate this transition, sample cultures were measured from 38°C to 44°C at 0.5°C intervals with sample size $N \geq 3$, with an

additional set of N = 3 measurements at 37°C. Measurements at 37°C were included to demonstrate behavior of the system at a steady-state growth rate in the survival regime.

3.1.1.2. NaCl stress. The effects of salt concentrations on the growth rate of *S. cerevisiae* were investigated by growing cultures in media with elevated NaCl concentrations ranging from concentrations which are documented to allow cells to grow normally, to concentrations known to cause considerable stress and even death of the yeast cells (Capusoni *et al.*, 2019; Melamed *et al.*, 2008; Murguía *et al.*, 1996; Szopinska *et al.*, 2011; Taymaz-Nikerel *et al*., 2016). NaCl was added to otherwise normally prepared YEPD in order to achieve desired concentrations. Culture growth was measured for NaCl concentrations from 66.5 to 104.5 g l^{-1} at approximately 5 g l^{-1} intervals with a sample size $N \geq 3$. Similar to the measurements for 37°C in the temperature trials, a set of N = 3 measurements at a concentration of 57 g l^{-1} was included, demonstrating steady-state growth further into the survival regime, and thus providing a reference to confirm emerging trends.

The initial culture in YEPD was inoculated into 50 ml of YEPD + NaCl (at desired concentration) in 125 ml flasks, resulting in an initial OD of 0.05. Samples were placed in orbital shakers at 30°C and measurements were taken at 24 h intervals for eight days. The procedure for measurements and dilutions was the same as for the temperature experiment; if growth over an OD of 0.01 occurred, the sample was diluted to an OD of 0.05, resulting in a volume of 50 ml YEPD + NaCl in a new 125 ml flask, and returned to the orbital shaker without any other action if no growth was measured. The number of doublings were calculated in the same way as the temperature experiment, using Equation 22.

Table 3.1 Mechanisms of Action of Antibiotics Used.

Antibiotic	Class	Mechanism	Mode of Action
Ampicillin	β-lactam	Inhibits cell wall synthesis	Bactericidal
Carbenicillin	β-lactam	Inhibits cell wall synthesis	Bactericidal
Chloramphenicol	Amphenicol	Protein synthesis (50S inhibitor)	Bacteriostatic
Ciprofloxacin	Fluoroquinolone	Inhibits DNA gyrase	Bactericidal
Gentamycin	Aminoglycoside	Protein synthesis (30S inhibitor)	Bactericidal
Kanamycin	Aminoglycoside	Protein synthesis (30S inhibitor)	Bactericidal
Rifampicin	Rifamycin	DNA-directed RNA polymerase	Bactericidal
Spectinomycin	Aminoglycoside	Protein synthesis (30S inhibitor)	Bacteriostatic
Streptomycin	Aminoglycoside	Protein synthesis (30S inhibitor)	Bactericidal
Tetracycline	Tetracycline	Protein synthesis (30S inhibitor)	Bacteriostatic

3.1.2. *Escherichia coli* Experiments. *Escherichia coli* (MG1655) (*E. coli*) were initially grown in M9 minimal glucose media, shaking at 37°C overnight. The initial culture was used to inoculate 96-well plates containing a dilution series of 10 different antibiotics, at a starting OD (600 nm) of 0.005 (i.e., 2 µl of culture in 198 µl media). The panel of 10 antibiotics, found in Table 3.1, spanned different antibiotic classes and included a variety of mechanisms of action. Each antibiotic was added at a beginning concentration of 100 µg ml^{-1} and serially diluted by ¾ for a total of 40 concentrations, from 100 to 0.001 µg ml^{-1}. For each antibiotic, a total of six replicates were measured. All 96-well plates were incubated at 37°C, shaking, for 24 h. All OD measurements were performed using a BioTek Cytation 3 multimode plate reader at a wavelength of 600 nm. Doubling times were calculated using Equation 22.

3.2. RESULTS

Figure 3.1 demonstrates the predicted behavior that environmental stressors result in a decrease in yeast growth rate, but a comparison of the data in Figure 3.1*a*,*b* reveal a significant difference in behavior as the stressors increase. The data shown in Figure 3.1*a* indicates that, as the salt concentration increases, the yeast grow more slowly, but with gradual change until the salt concentration is high enough to prevent growth altogether. In contrast, the data shown in Figure 3.1*b* presents a sharp drop in doubling rate as temperature increased; the cultures change from a quick recovery time in the 37-39°C range, to a recovery after a long lag time in the 39.5-40.5°C range, and no recovery for temperatures above 40.5°C. Data points taken at 37°C have been included to demonstrate steady-state growth well into the survival regime. Data points taken at 30°C (not shown) confirmed that the growth rate at this standard temperature for yeast growth was similar to the growth observed at 37°C.

Investigation of the population dynamics in the neighborhood of the tipping point requires an understanding of the steady state behavior in a system. In this case, the steady state is a steady growth rate, calculated by averaging the number of doublings over all remaining days once the system has reached an approximately constant growth rate. As can be seen in Figure 3.1, the populations reached an approximately constant growth rate by day 6 post-inoculation. The steady growth rate was thus calculated as the average number of doublings over the last three days for each experimental run. Steady growth rates are shown in Figure 3.2*a* as a function of NaCl concentration and Figure 3.2*b* as a function of temperature, with a grand average over all experiments at each condition, and error bars representing the standard deviation across all experiments at each condition.

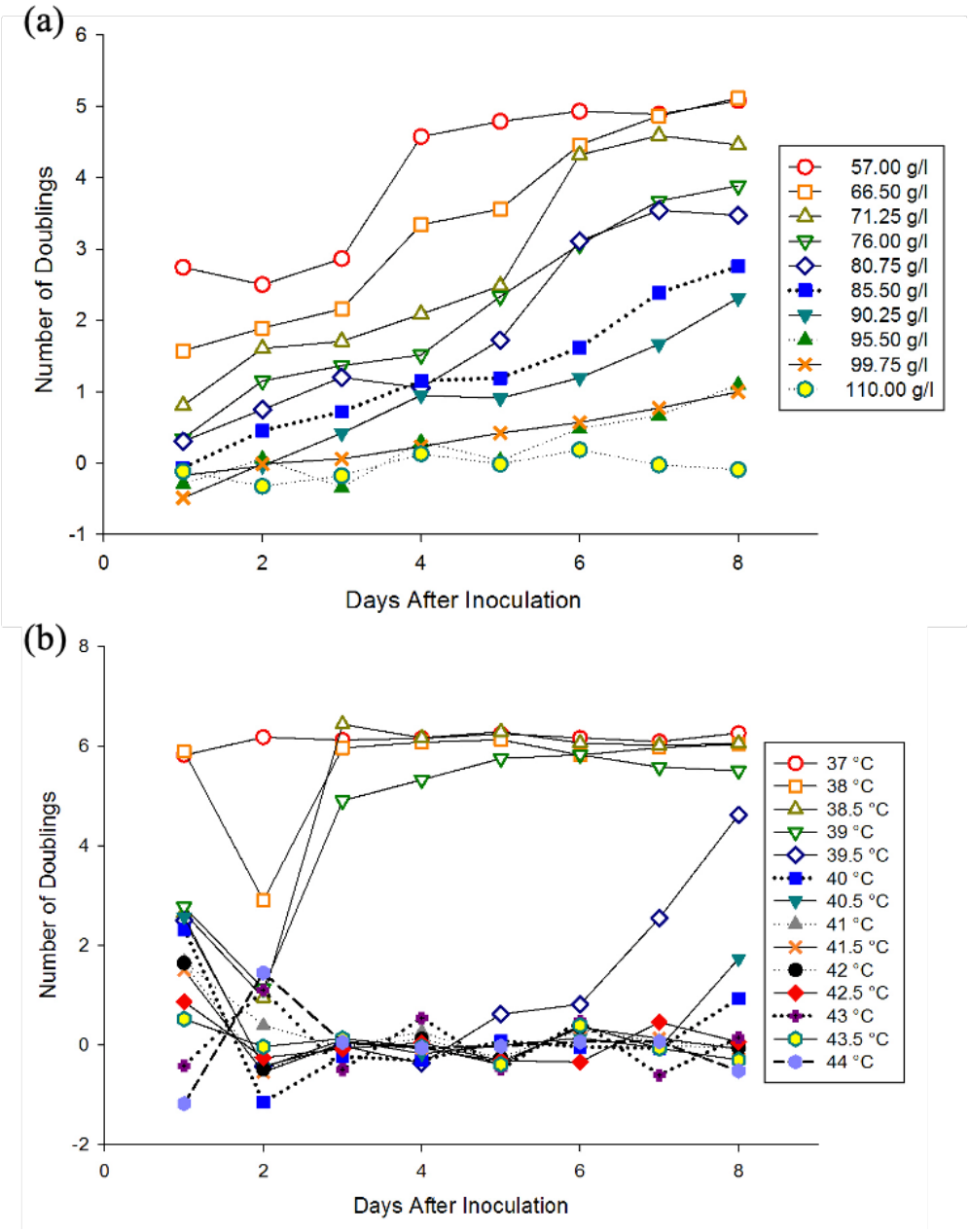

Figure 3.1 Daily Number of Doublings of *S. cerevisiae*. Number of population doublings for *S. cerevisiae* shown as a function of days after initial inoculation for values of NaCl concentration (a) and temperature (b). N ≥ 3 for each data point; error bars not displayed to avoid crowding of the figure.

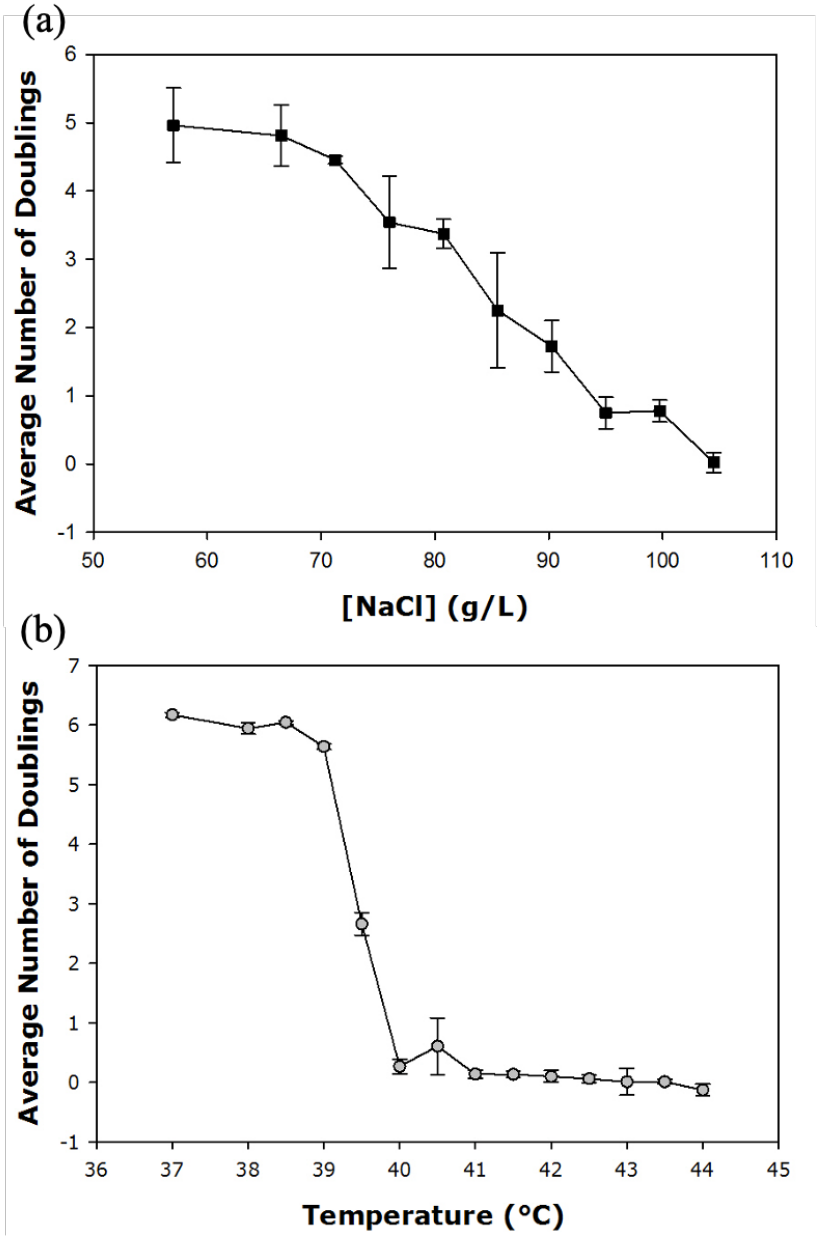

Figure 3.2 Steady Growth Rates of *S. cerevisiae*. Steady growth rates of the *S. cerevisiae* population for NaCl stress (a) and temperature stress (b) are shown. The steady growth rates equal the number of doublings averaged over days 6-8 post-inoculation, then over all replicate experiments. Error bars show standard deviation between replicates.

Figure 3.2 demonstrates, for both stressors, that the yeast population ultimately experienced an environment too harsh for survival with increased system stress. In the

case of increasing salt concentration, shown in Figure 3.2*a*, the system underwent a smooth, gradual decline with minimal fluctuations observed in the transition region, and with no discernable tipping point. In stark contrast, the system's response to temperature stress, shown in Figure 3.2*b*, showed an abrupt drop-in growth rate, with large fluctuations near the transition (as compared to the system's behavior away from the tipping point), which indicates a tipping point around 39.5°C.

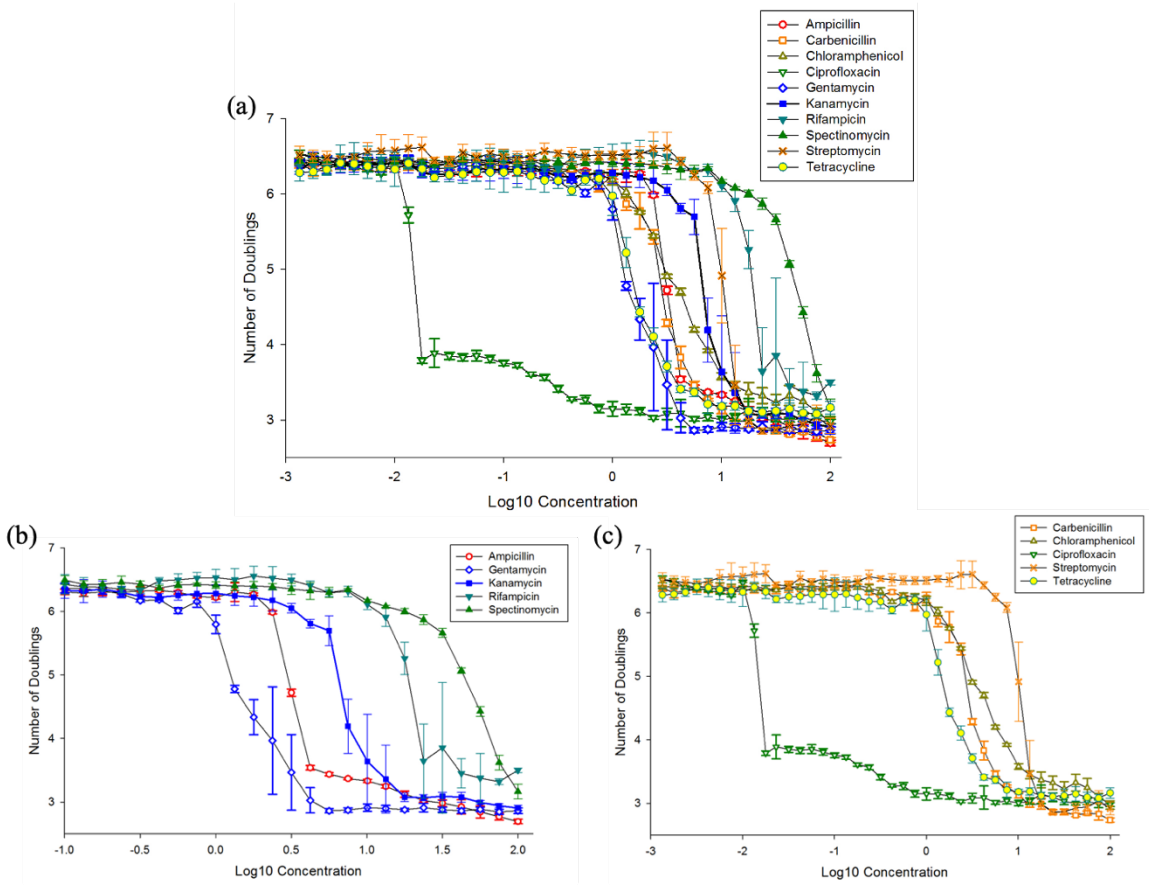

Figure 3.3 Number of Doublings of *E. coli*. Number of doublings of *E. coli* (MG1655 WT) populations as a function of $\log_{10}$(antibiotic concentration) for 10 antibiotics shown in (a). Error bars show standard deviation over six replicates for each antibiotic concentration. To show greater detail, (b) and (c) each show five of the studied antibiotics over a smaller antibiotic concentration range. Data collected by H. Huelskamp in the laboratory of R.F. Inglis.

Figure 3.3 shows the decay curves of *E. coli* populations in the presence of various antibiotics, where Figure 3.3*a* shows all of the antibiotics tested, and Figure 3.3*b*,*c* show the population responses to five each of the ten antibiotics studied to display the details of the responses more clearly. As expected, for all antibiotics tested, the antibiotic stress caused a decrease in bacterial growth. There is, however, a variety of patterns of decreased growth in response to antibiotic stress both within and across antibiotic classes and mechanisms of action. These patterns range from a very sharp drop in bacterial growth in the presence of ciprofloxacin (open green triangles in Figure 3.3*a*,*b*) to more gradual declines in the presence of spectinomycin (filled green triangles in Figure 3.3*a*,*c*). The broad range of response patterns seen across the selection of antibiotics is qualitatively similar to the observed behaviors in the yeast experiments.

3.3. DISCUSSION

When examining the responses to both stressors of *S. cerevisiae*, and the responses to the various antibiotics of *E. coli*, it is apparent that the nature of the transition is dependent on the specific stressor. A major difference in the presented *S. cerevisiae* and the studies of Dai *et al*. is the use of glucose as a nutrient source rather than sucrose (Dai *et al*., 2012, 2013, 2015). *S. cerevisiae* can metabolize glucose directly, whereas the disaccharide sucrose must be hydrolyzed outside the cytoplasm before it can be metabolized. During the process of hydrolysis, products of sucrose diffuse away before they can be taken back up into the cytoplasm, becoming public goods (Gore *et al*., 2009). Due to this mechanism, yeast cells fed with sucrose exhibit a classical Allee effect, with maximal growth at an intermediate population density (Dai *et al.*, 2012). The

population density is prevented from becoming a major influence in the population dynamics of interest by instead using glucose as the food source, enabling temperature and salt concentration to each serve as an isolated stress variable in the study presented here.

As expected, both stressors applied to *S. cerevisiae* cultures have a negative effect on yeast doubling rate for a sufficiently high level of stress, but the dynamics of the response to the two stressors differ greatly. As seen in Figure 3.2*a* and *b*, the response to NaCl stress is gradual, whereas temperature stress induces a sharp drop indicative of a critical phase transition with a well-defined tipping point. Accompanying this sharp change in population growth, the fluctuations of the system are much higher in the vicinity of the tipping point, compared to the response further into either phase of the system. Comparatively, the fluctuations of the system exposed to NaCl stress are uncorrelated with the vicinity of the tipping point. In the *E. coli* experiments (Figure 3.3), the responses to several antibiotics, including ciprofloxacin, streptomycin, and kanamycin, exhibit critical phase-transition-like behavior, demonstrating a well-defined transition point with high fluctuation about the tipping point. Other antibiotics, including chloramphenicol and spectinomycin, produced a gradual response with no well-defined tipping point. For other antibiotics, the responses were less clearly defined.

The observation of critical behavior requires an examination of the system after it has stabilized. As seen in Figure 3.1, for both stressors, the yeast had fluctuating growth behavior during the first few days after exposure to stress, and for most control parameters, the behavior stabilized after a few days. Figure 3.2 shows the averaged "long-term" stable behavior. The results presented are consistent with the findings of

Mensonides *et al*. (2002), who demonstrated the initial slowing of *S. cerevisiae* growth due to temperature for budding yeast, along with changes in metabolism, during the population's first 6 h after the introduction of various temperatures (Mensonides *et al*., 2002). In their work, they "observed a surprisingly 'thin line' for cells between growing, surviving and dying, with regard to growth temperature" (Mensonides *et al*., 2002). The sharp change indicating a tipping point between survival and death observed by Mensonides and colleagues occurred at a temperature between 42 and 43°C, slightly higher than the observed tipping point in this study of 39.5°C. The difference of observed tipping point temperatures is likely due to comparing short- versus long-term behavior; Figure 3.1 shows that all cultures below 42°C had a measurable amount of growth during the first 24 h period, but the populations were unable to recover and had no further growth. Mensonides *et al*. suggested that it cannot be concluded if the response observed was the result of the absolute temperature introduced or the difference between the initial and new temperature (Mensonides *et al*., 2002). The results seen in Figure 3.1 suggest the absolute temperature has an impact for long-term behavior. Notably, the initial temperature in their study was 28°C, compared to the 30°C used for this work.

Similar to the results shown here, Dai *et al*. observed that *S. cerevisiae* populations lose stability at different rates and exhibit different levels of resilience, based on different environmental stressors (Dai *et al*., 2015). For example, in the presence of sucrose, *S. cerevisiae* populations lost stability more rapidly when stressed by increasing dilution factor, a proxy for death rate in populations grown in sucrose, than in response to simple nutrient depletion (Dai *et al*., 2015).

To further investigate the reasons for the difference in response to the salt and temperature stressors, it is necessary to investigate the impact of the stressors on the yeast cells. Mensonides *et al*. proposed that the narrow temperature range over which the population decline is observed "may be explained by assuming the rapid denaturation of one or a very limited number of enzymes which are essential to growth" (Mensonides *et al*., 2002). This understanding is consistent with current understandings of heat shock response and protein denaturation in *S. cerevisiae* (Leuenberger *et al*., 2017; Verghese *et al*., 2012). In contrast, *S. cerevisiae* have a variety of known responses to modulate the response to salt stress. Change in salt concentration can have a number of effects on yeast, including osmotic shock, direct toxicity from Na^+ ions, and changes in membrane potential (Gore *et al*., 2009; Hohmann, 2002; Murguía *et al*., 1996). *S. cerevisiae* has been observed to respond to these effects with a range of responses, such as activation of the high-osmolarity glycerol (HOG) pathway, and membrane depolarization accompanied by decreased permeability (Capusoni *et al*., 2019; Verghese *et al*., 2012). Szopunska *et al*. proposed that protein internalization occurs rapidly after hyper-osmotic or ionic shock, enabling a cell to remain viable until a slower transcriptional response can be activated after observing that mild salt stress resulted in an increase in abundance of a dozen plasma membrane proteins (Melamed *et al*., 2008; Szopinska *et al*., 2011; Taymaz-Nikerel *et al*., 2016). These studies suggest that *S. cerevisiae* has better protective responses to salt stress than heat stress, which is consistent with the observations of temperature and salt stress presented here.

The explanation for the range responses of *E. coli* to various antibiotics is less clear. One might expect similar response to antibiotics with the same mode of action,

such as spectinomycin and kanamycin, which are aminoglycosides, but the decay curves present characteristically different responses. One might also expect that bacteriostatic antibiotics that only inhibit growth to have more gradual decay curves, as opposed to bactericidal antibiotics that kill *E. coli*. Surprisingly, there was a large variation in decay curves in both bacteriostatic (e.g., chloramphenicol and tetracycline) and bactericidal (e.g., rifampicin and ampicillin) antibiotics. More comprehensive screening would be required to illuminate the exact underlying mechanisms of each antibiotic that drives these differences in bacterial killing. The observed differences in decay curves might have implications for the evolution of antibiotic resistance. It is possible that mutations could impact the shape of the decay curve; this could be investigated using experimental evolution techniques.

The transition behavior seen in the temperature stress experiments and several types of antibiotics including ciprofloxacin, streptomycin and kanamycin demonstrate characteristics indicative of critical phase transition behavior, while the gradual transition in response to salt stress and other antibiotics including spectinomycin and chloramphenicol present none of these characteristics. The different responses to stress demonstrate that monitoring the dynamics of a population's decay may not only provide a warning sign of incipient collapse but may also be used to identify the stressor or combination of stressors causing the destabilization of the system. These differing responses also demonstrate that simple observation of a population in decline may not be sufficient to predict the course of the population's progress toward collapse without reference to the environmental stressors present. With the combined information of environmental stressors and information about individual stressor effects on a population,

it may be possible to identify the driving stressor(s) of a declining population and determine whether the decline is gradual or exhibiting critical behavior. Understanding the type of stress and corresponding transition behavior could have a major impact in designing appropriate intervention protocols.

In addition to greater understanding of populations approaching the tipping point, the type of stress response may also play a role in the dynamics of population recovery. Population recovery has been observed by King (2015) in a computational evolutionary model for a system near the tipping point in a critical phase transition and in studies of population resilience by Dai *et al*. in the presence of different stressors (Dai *et al*., 2012, 2013, 2015; King, 2015; King *et al*., 2017). A key behavior in systems with critical transition behavior is critical slowing down. If intervention is taken to save a population from collapse, any systems that exhibit critical slowing down will have significantly longer recovery times when compared to non-critical transition systems. For this reason, identification of transition type can have a major impact on resource allocation to rescue the highest number of populations approaching collapse and reduce the number of prematurely abandoned recovery efforts. Lastly, due to the cooperative effects, population size itself may modulate the dynamics of collapse and recovery and should be considered when designing any intervention practice.

3.4. CONCLUSION AND FUTURE WORK

Population declines exhibiting characteristics of a critical phase transition have been observed in both yeast (*S. cerevisiae*) and bacteria (*E. coli*). Yeast populations exhibit critical transition behavior when cultures are subjected to temperature stress, but

not when subjected to high salt concentration. Instead, yeast populations respond to increasing salt concentration with a gradual decline in reproduction, without any sign of criticality. Similarly, bacterial cultures respond to some antibiotics with population declines characteristic of a critical phase transition, while others respond with a gradual decline. Taken together, the yeast and bacterial results indicate that critical phase transition-like population dynamics may occur in response to a broad range of stressors in different organisms. Additionally, the disparity between the behavior indicative of a critical transition and more gradual population decline in both eukaryotic and prokaryotic cells for different stressors raises the possibility that divergent population dynamics may occur for a range of organisms and stressors.

These results indicate that critical phase transition behavior occurs in the presence of some, but not all stressors. The underlying causes of the different responses is an avenue of further investigation, including investigation of metabolic flux and imaging analysis of cells in the presence of different stressors over time. In the case of temperature and salt stress on yeast populations, experimental evolution studies, as well as the extensive literature on transcriptional response to stress in *S. cerevisiae* could point the way to identification of particular genetic and/or transcriptional determinants of population collapse dynamics (Nguyen Ba *et al*., 2019; Ratcliff *et al*., 2012, 2013, 2015; Taymaz-Nikerel *et al*., 2016). The repressors Nrg1 and Nrg2 have already been implicated in the adaptation to salt stress in *S. cerevisiae* (Vyas *et al*., 2005). Two decades ago, several hundred yeast genes were shown to respond to environmental stress by undergoing significant changes in expression (Causton *et al*., 2001; Gasch *et al*., 2000). The genetic pathways that mediate resilience of a population are likely to be

complicated, multilayered, and historically contingent. Berry *et al*. (2011) found that the changes in gene expression that confer tolerance to H_2O_2 stress in yeast differ based on the prior exposure the cells received (salt, heat shock or dithiothreitol (DTT)) (Berry *et al*., 2011). These complex interactions make investigating such problems daunting, but they also offer the possibility that mapping gene activation can reveal a cell's environmental history, similar to the way a human patient's complement of antibodies can show their history of disease exposure or vaccination record. Such studies could not only contribute to the development of population recovery protocols but may also facilitate the genetic engineering of species to be more resilient in the face of environmental stress.

4. POPULATIONS AND SUPERVISED MACHINE LEARNING

Section 2 investigated the response of populations based on system parameters in a computational model, and Section 3 presented a controlled laboratory experiment to identify system-level behaviors in real populations. Application of the methods presented in these sections has significant limitations in uncontrolled, real-world environments, such as time-consuming data collection, noisy data, control of environmental stressors, and cost. Therefore, an approach that can incorporate multiple environmental variables when analyzing uncontrolled impacts on a wild population is needed. With data availability in today's world, using analytic techniques that rely on large collections of data, even though this data may appear unconnected or not systematically collected in a controlled way, can potentially provide a new way of understanding system-level behavior. For this section, investigation will shift to using common data science techniques to compile disparate sources of data through standard machine learning practices in an effort to predict population decline.

The consolidation of a large number of individual studies like those found in The Living Planet Index (LPI) provides enough data that, after all necessary processing, a sufficiently large data set remains for analysis (*Living Planet Index*, n.d.). The use of this database has provided the necessary data to identify a trend of rapid decline in local vertebrate species abundance (Collen *et al*., 2008). Further investigation into these population trends show that both rare and common populations face multiple threats owed to human impact (Daskalova *et al*., 2020). The combination of information from LPI and a database on mammal traits, PanTHERIA, provides enough data to predict how

populations decline based on intrinsic traits and extrinsic factors and to demonstrate the changing rates of biodiversity loss (Collen *et al*., 2011).

With the development of this available data, analytical approaches have evolved with the use of machine learning to extract more information from these datasets. Machine learning is an inclusive term for computer programs that learn without explicit instruction, where "a computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E" (Mitchell, 1997). In the scientific community, machine learning has become a valuable tool, especially when investigating large biological data sets. In the late 1990s, machine learning played a role in modelling forest development based on the population dynamics of red deer (*Cervus elaphus* L.) and meteorological information (Stankovski *et al*., 1998). This work was extended to investigate the habitat suitability for red deer utilizing classification and decision trees (Debeljak *et al*., 2001). Recently, a number of machine learning algorithms were used in an effort to accurately predict winter wheat yield from multi-source data including historical yield records, climate data, satellite images, and soil maps, with results indicating that the machine learning methods all outperformed more traditional statistical methods of prediction (Y. Wang *et al*., 2020).

The field of modern machine learning can be broken down into four categories based on desired outcome and supplied data. *Supervised learning* utilizes input data to accurately classify or predict the provided output data (also known as labels). In contrast, *unsupervised learning* is not provided with the output data (or labels) and, instead of mapping to a known outcome, uses relations and connections in the input data to find

clusters of similar data and associations within the data (Çelik, 2018). *Semi-supervised learning* is used when a dataset includes data with and without outputs or labels, making both supervised and unsupervised learning inadequate. Lastly, *reinforcement learning* uses a "reward system" to determine the shortest and most correct way to reach the desired outcome (Çelik, 2018). The remainder of the discussion will be focused on supervised learning.

Supervised learning can be further separated into two categories: *regression models* are used to estimate output values based on the given parameters, while *classification models* distribute data into the categories defined in the dataset (Çelik, 2018). The work presented in this section explores the use of supervised machine learning models to determine the predictability of population dynamics from accessible local weather attributes. A regression model and two classification models were tested on the dataset comprised of population data and weather parameter data with the aim to predict future population changes based on the local weather.

4.1. METHODS

All population data was sourced from LPI, currently the largest open source repository of population studies (*Living Planet Index*, n.d.). Before data collection began, a few limitations were placed on this proof of concept. Proper analysis of migrating species would require general understanding of migration schedule and tracking the locations of the migrating populations to examine the impact of weather patterns at multiple locations. Since this is an investigation of population dynamics based on local weather behavior, any species that exhibit migration patterns significantly complicate the

analytic process. To address this concern, no avian species were included in the presented work. Secondly, all classes of fish and plants were discarded to focus the results to primarily terrestrial animals. Lastly, insects were removed from the scope of the work. In order to ensure that all weather data collection was performed uniformly as an effort to reduce inconsistency, limitations were placed geographically: all population studies used were limited to within the United States so all weather data was gathered from one source, the National Oceanic and Atmospheric Administration (NOAA) (NOAA, n.d.). Once all the limitations on the scope of work were implemented, the resulting data included all available amphibian, mammalian, and reptilian populations observed in the United States. Any study that did not exhibit population tracking for at least 5 years was also removed from the data set.

For each study in the database, the latitude, longitude, and year span were compared with the list of all weather stations to determine the closest weather station with an available global summary of the year (GSOY) for all years in the study. All GSOY datasets were retrieved for each year, and the location of each study from each weather station was gathered using NOAA's application programming interface (API). The GSOY has the possibility to include a plethora of weather attributes, though a number of attributes like those related to soil temperature are rarely recorded. A list of the attributes with enough data for analysis can be found in APPENDIX A. In the case that no weather station is within a window of ± 5° in latitude and longitude, the study was removed.

Upon initial data examination, it was determined there was a need for additional data cleaning and normalization before analytics could take place. The dataset included a

range of different units in population measurements across the included data. Some studies present data with units of counted individuals, while others used methods like the percentage of smell traps visited or approximate count from regular flyovers. In addition to variation from counting methods, some studies examined very large populations with measured units on the order of 10^5, while others measured units on the order of 10^1. This disparity between measurement style and number of units can be seen in Figure 4.1*a*. For this reason, each study was individually scaled (normalized) so all populations fall between the interval [0,1]. Scaling the population maintains the population dynamics of interest but allows for a common analytical application across all of the studies. The rescaled populations are visualized in Figure 4.1*b*. For each year in which the data of the previous year is available, the difference in scaled population value was calculated. This population difference will act as the dependent variable, or label.

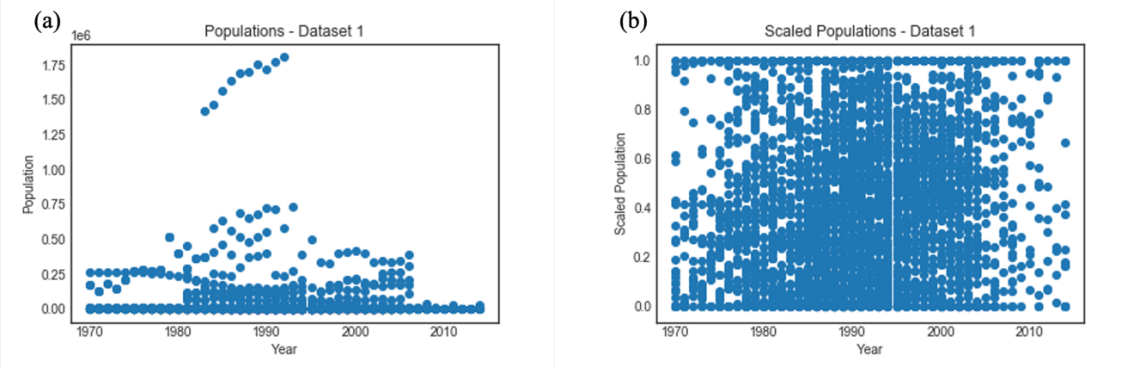

Figure 4.1 Scaled and Unscaled Populations. Populations from all studies are plotted as the observed value by year (a) and Populations scaled to [0,1] for each study (b).

Lastly, while machine learning is capable of handling messy or noisy data much more efficiently than traditional analytical methods, it still requires complete datasets.

For each year and location, it is common for multiple attributes of the GSOY to be missing. All attributes that include values for less than 80% of the dataset were removed (Dataset 1). In cases where feature values were available for 80% or more or the data set, the remaining missing information was approximated by averaging the attribute over the other provided years from that specific weather station. In all cases where no other data is available to average from the weather station, no reasonable estimations can be made to approximate the missing data. This results in an incomplete data point for all data collected at that location, and these data points must be removed. Because each study only covers one location, the removal of the incomplete data, which is location dependent, results in the removal of entire studies. The final size of Dataset 1 includes 2929 data points, or training examples, with five attributes from 258 studies. These remaining weather attributes act as the independent variables or features of the model. An additional version of the dataset was made following the same process but lowering the threshold of missing weather data to 70%, which includes a total of 16 features, but at the cost of the total number of training examples (Dataset 2). By reducing the threshold to 70%, the number of incomplete data points after missing attributes are approximated increases, resulting in a greater number of training examples removed. The final size of Dataset 2 includes 1833 training examples with 16 attributes from 200 studies.

Once the dataset contained only data with all complete or interpolated data for each weather attribute and the scaled population difference, it was properly prepared for a regression model. For the use of a classification model, the data were placed into two classes, population increase and decrease, with a population difference of 0 included in population decrease. For any choice of model, the data must be separated into a training

set, which will be given to train the chosen model, and a test set, which is reserved and will be used to test the performance of the model on "new" data. The test set consists of the final year from each study in the final data set. The remaining data act as the training set. In order to reduce chance of bias, only the training set is used for any data exploration, including examination of the features and labels for correlation. Highly correlated features can cause issues with a model's effectiveness and can often result in less accurate models when all correlated features are included, compared to the inclusion of just one (Bishop, 2006).

4.1.1. Classification Models. Two classification models were utilized, a decision tree classifier (DTC) and a random forest classifier (RFC). A decision tree classifies data by creating splits at nodes in order to separate the data into increasingly homogeneous groups. For example, a node may split the data based on the value of yearly precipitation, separating all training examples with a precipitation measurement above some value into the population increase group, and all training examples with a precipitation measurement lower than that value classified as population decrease. Each node can make classification predictions based on one or multiple features (weather attributes). Traditionally, decision trees select the feature or features to split the data based on the entropy of a given collection of training examples, *S*, relative to the Boolean classification of an attribute *A* (node). The entropy is calculated as:

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i \tag{23}$$

where p_i is the proportion of S belonging to class i, and c is the total number of classes (Mitchell, 1997). The selection of the splitting attribute is based on the maximum information gain where information gain is defined as:

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{24}$$

where $Values(A)$ is the set of possible values for the attribute A, and S_v is the subset of S for which A has the value v. Models can alternatively determine the splitting attribute with by minimizing the Gini impurity, with Gini is defined as

$$Gini = \sum_{i=1}^{c} p_i^2 \tag{25}$$

where p_i and c have the same definition and the Gini impurity is defined as

$$Gini\ Impurity = 1 - Gini. \tag{26}$$

As shown in Figure 4.2, the first node is referred to as the root node, with each branch corresponding to one of the possible values of the node attribute. The branching continues splitting the data at each node in an effort to separate all data into the correct class (Mitchell, 1997). The DTC works by passing the data from the root to the leaf (final grouping) and assigns labels according to the majority of samples in the leaf. An RFC is similar to a decision tree classifier, but it creates a number of decision trees which are aggregated to form a final output classification or label. The results of the trees are aggregated with a weighted system for each tree that generates labels, generally resulting in a better fit compared to a decision tree classifier (Ali *et al*., 2012). The scoring metric for classification models in this work is accuracy, or the percentage of correctly classified samples.

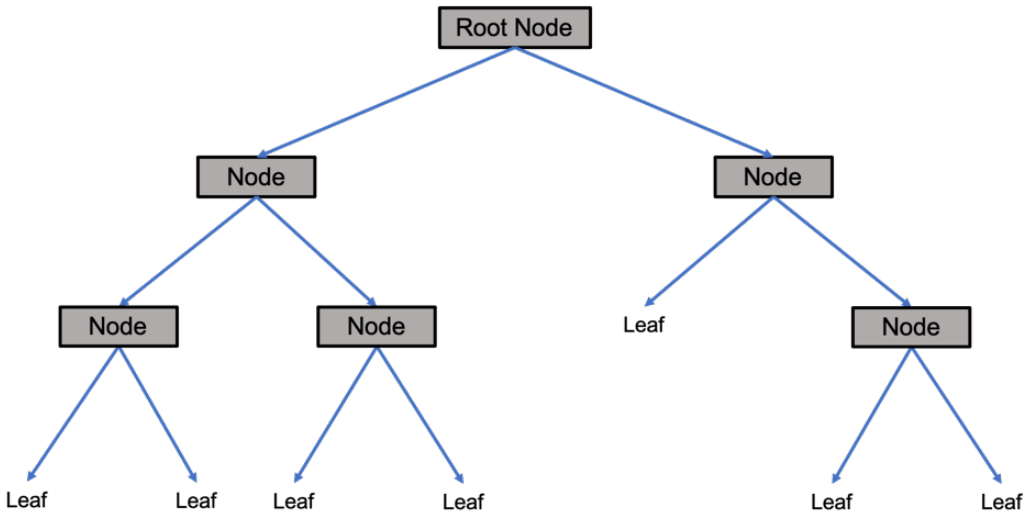

Figure 4.2 Generalized Decision Tree. Each node of the tree acts as a test of an attribute, separating the data based on the test. The initial split occurs at the root node, and the data in each branch continues to the next decision node. The depth of the tree refers to the number of nodes a datapoint can pass through.

4.1.2. Regression Models. In addition to the classification models, the use of a linear regression (LR) model was investigated. Linear regression modelling works to create a mapping from the list of weather attributes (features) to population differences (labels) taking the general form:

$$h(x^{(i)}) = \sum_{j=0}^{n} \theta_j x_j^{(i)} \tag{27}$$

where $x_j^{(i)}$ is the jth feature of the ith training example, $h(x^{(i)})$ is the hypothesis of the ith training example, θ_j is the jth fitting parameter and $x_0^{(i)} = 1$ for all i. The model is optimized by minimizing the cost function:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}\left(h\left(x^{(i)}\right) - y^{(i)}\right)^2 \tag{28}$$

where m is the number of training examples, and $y^{(i)}$ is the i[th] target variable for each iteration. During each iteration θ_j is updated as:

$$\theta_j \coloneqq \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \tag{29}$$

where $\coloneqq$ represents assignment, or the replacement of the old value with the new, and α is the learning rate (Ng, 2020). The scoring metric for regression models in this work is the coefficient of determination R2 (pronounced R squared). It should be noted that R2 differs from traditional definitions for coefficient of determination, r^2, in that negative values are possible; R2 scores are 0.0 when a model's predictability is comparable to a model that always predicts mean output value.

4.1.3. Hyperparameter Tuning. Most machine learning models accept a number of parameters that can affect the performance of the model, known as hyperparameters. For example, in the case of a model using decision trees, the maximum number of layers of nodes, known as the maximum depth, can be set to limit the model. Each model has a unique set of hyperparameters that can be initialized in an effort to improve model performance or can be left as default values. Hyperparameter tuning is the process of identifying specific parameter values in an effort to optimize model performance. A common method of determining optimal hyperparameters, which is used here, utilizes a grid search that cycles through specified hyperparameter ranges, and determines the set of hyperparameters values resulting in a model fit with highest scoring metrics. Hyperparameter tuning can however prevent a model from *generalizing*. A model fails to generalize when the optimization of the model to the training set results in a model that accurately predicts the training data but fits the test data poorly. The result is an

overfitting model that will give inaccurate predictions when presented with new data (Roelofs, 2019).

For the purposes of this study, the hyperparameters modified for DTC include maximum depth, minimum sample split (which defines the minimum number of samples required for a node to split the samples), minimum sample leaf (which defines the minimum number of samples in a leaf for it to be retained), and the maximum number of features to be included in one node. The hyperparameters that are modified for RFC include the list for DTC with the addition of the number of decision trees, known as estimator trees, and the criterion, which can either be ‘entropy’ or ‘gini.’

4.2. RESULTS

Figure 4.3 demonstrates possible correlations as a scatter plot matrix of the features (weather attributes and previous population) and labels (population difference) of Dataset 1. The diagonal holds plots of the kernel density estimation (KDE), a nonparametric estimation for the probability density function. The scatter plots corresponding to both population difference and previous population with any weather feature demonstrate a wide spread in the scatter plots, suggesting that no single weather attribute is correlated to the population difference and therefore cannot individually predict the population difference. The plots of any two weather features demonstrate a different, more focused spread, indicating higher correlation between the weather attributes, especially in the case of the number of days with >= 1.00 inch (25.4 mm) of precipitation in the year, DP1X, and the total annual precipitation, PRCP. Figure 4.4 shows the scatter plot matrix for Dataset 2, similarly demonstrating a lack of correlation

between the population difference and any weather features. With a larger number of features, Figure 4.4 includes a wider range of correlation between weather attributes when compared to Figure 4.3, including both positive and negative correlations.

The correlation values for each combination of features and the label are printed on the correlation heat maps for Dataset 1 (Figure 4.5) and Dataset 2 (Figure 4.6). For both datasets, the population difference label and previous population feature show some negative correlation with one another, but both population difference and previous population have near zero correlation with any given weather attribute. The weather attributes Dataset 1 all demonstrate some correlation to one another, whereas the inclusion of more features in Dataset 2 provides a much greater range in correlation values.

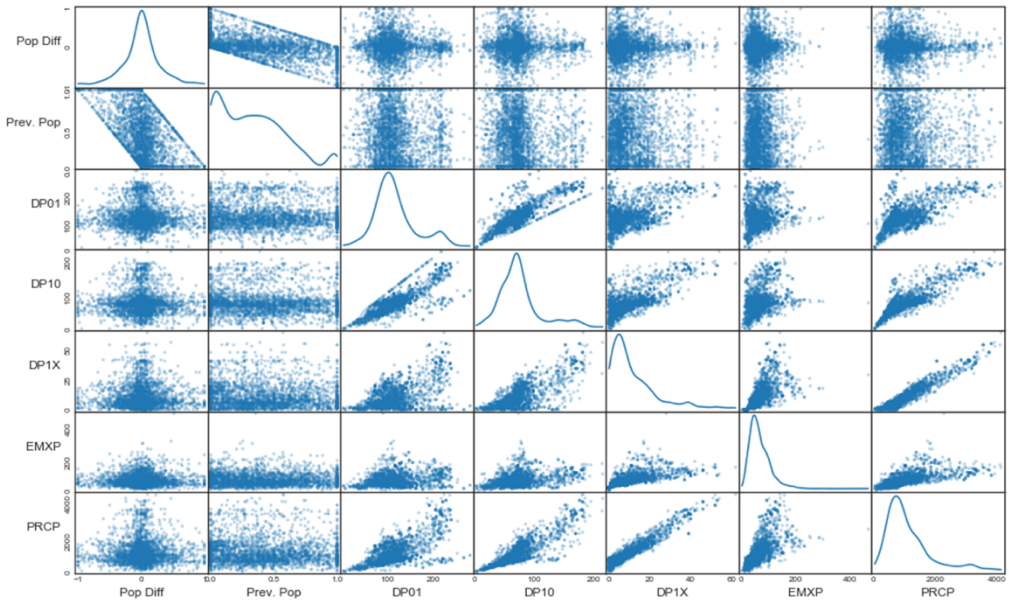

Figure 4.3 Feature Scatter Matrix – Dataset 1. A scatter plot matrix of the available features, previous population (labeled "Prev. Pop") and the weather attributes DP01, PD10, DP1X, EMXP, and PRCP for Dataset 1 and the label, population difference (labeled "Pop Diff"), are shown. Diagonal plots show the KDE for each variable.

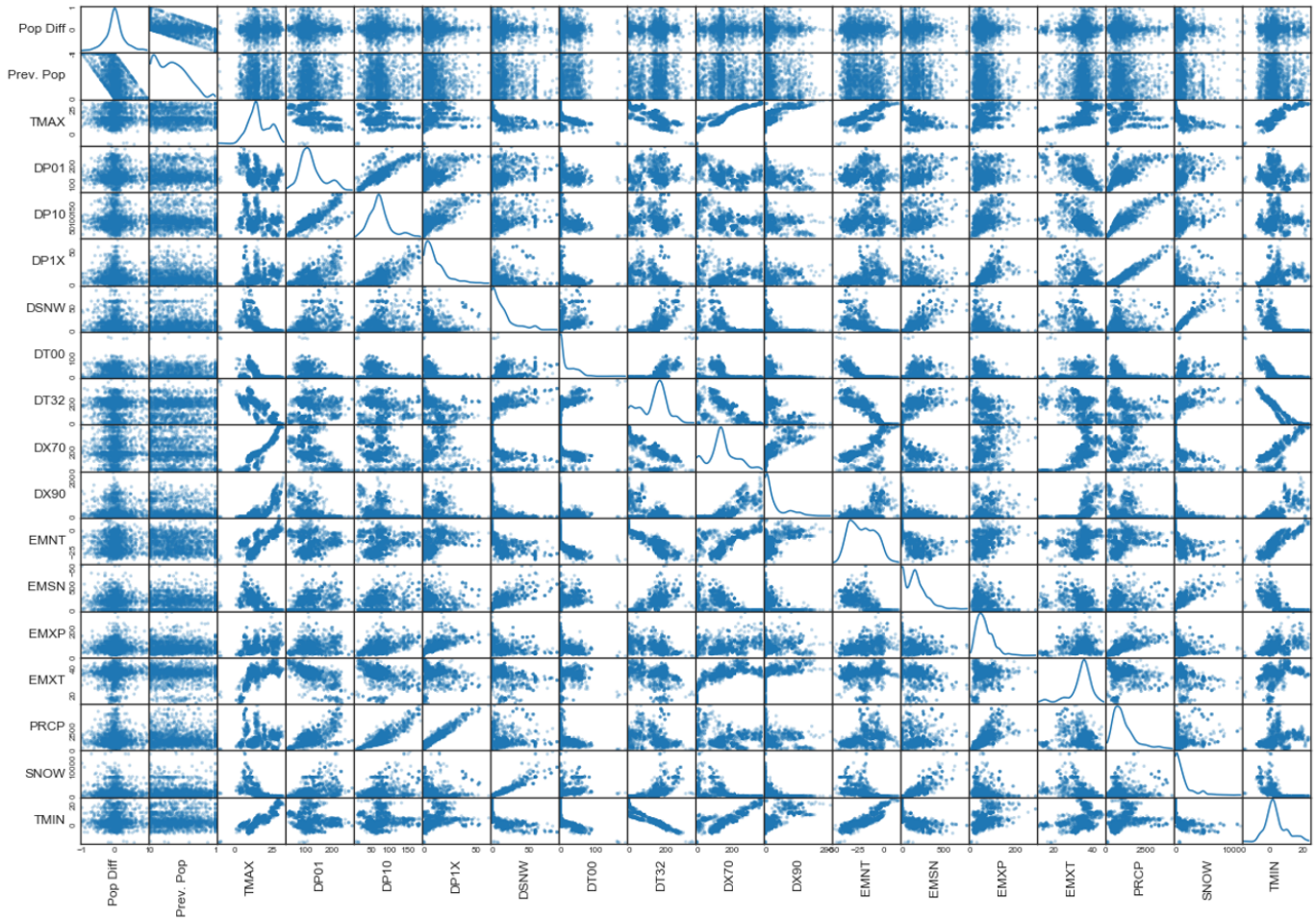

Figure 4.4 Feature Scatter Matrix – Dataset 2. A scatter plot matrix of the available features, previous population (labeled "Prev. Pop") and all weather attributes, for Dataset 2 and the label, population difference (labeled "Pop Diff"), are shown. Diagonal plots show the KDE for each variable.

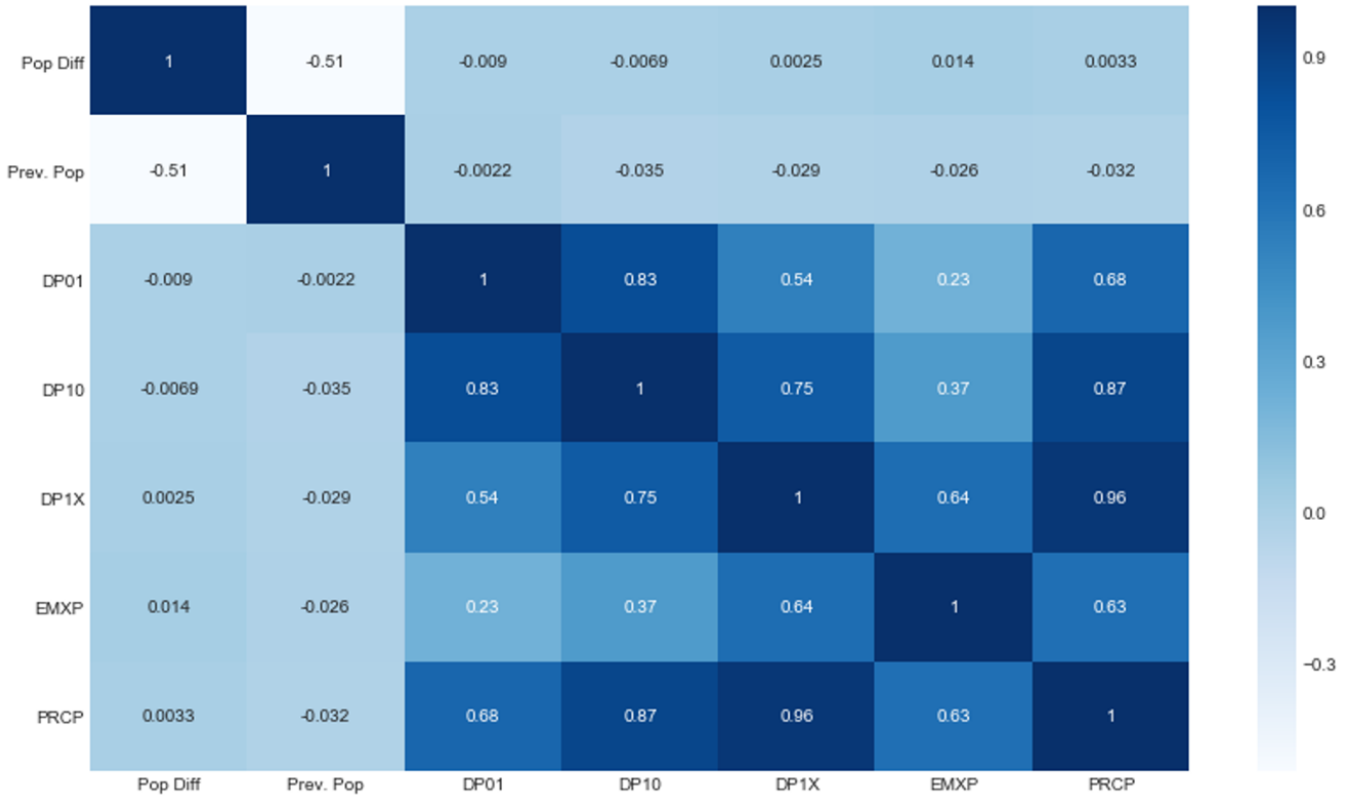

Figure 4.5 Feature Correlation – Dataset 1. The correlations between the population difference (labeled "Pop Diff"), previous population (labeled "Prev. Pop"), and the weather attributes of Dataset 1 are shown as numerical value in each grid location and visually as a heatmap using the color scale on the right.

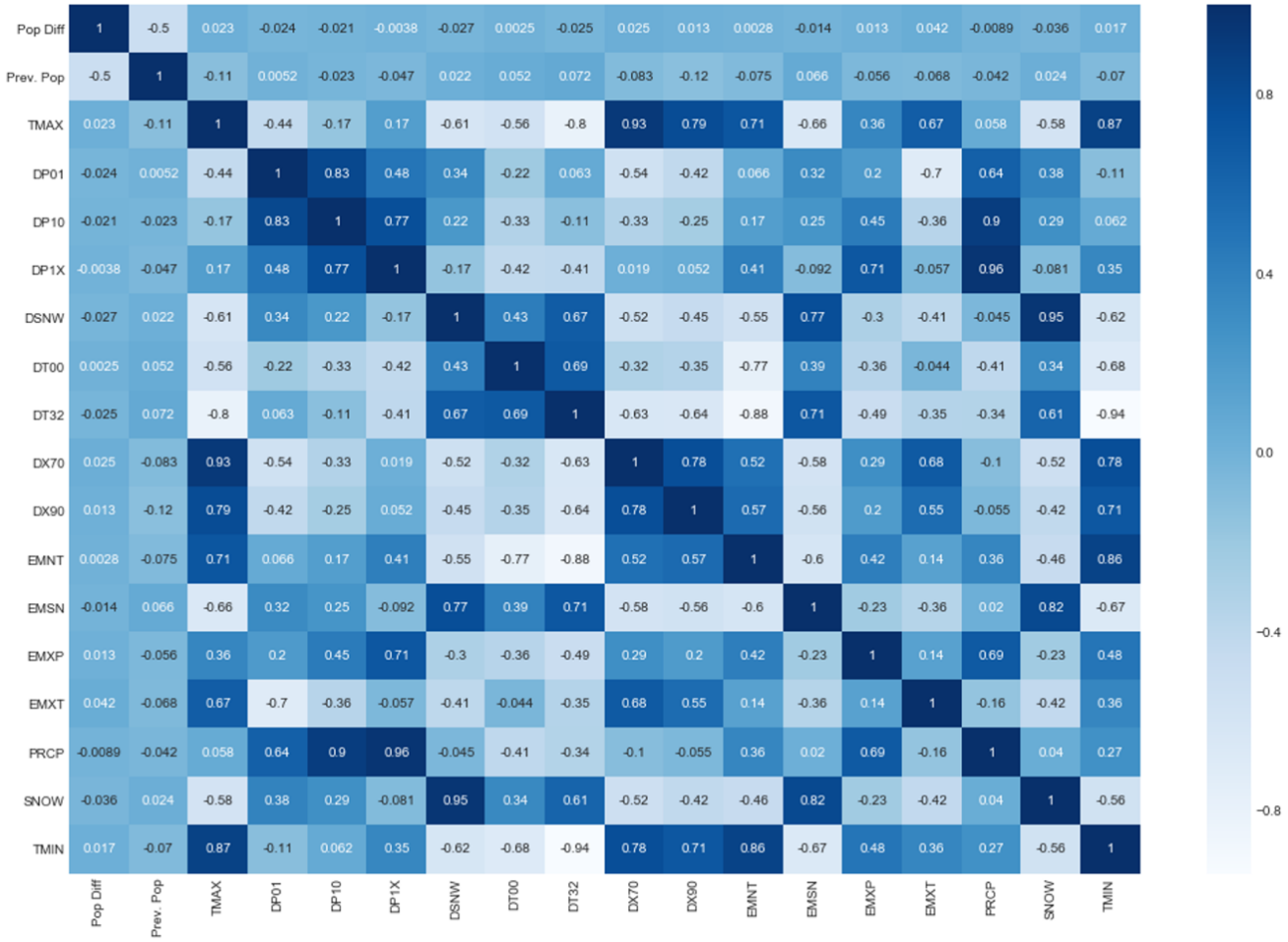

Figure 4.6 Feature Correlation – Dataset 2. The correlations between the population difference (labeled "Pop Diff"), previous population (labeled "Prev. Pop"), and the weather attributes of Dataset 2 are shown as numerical value in each grid location and visually as a heatmap using the color scale on the right.

Depending on the model type selected, correlated features can negatively affect the model fitting process, though some models have little response to the use or removal of correlated features. The comparison of Figure 4.7*a* showing the learning curves of DTC utilizing all features of Dataset 1 and Figure 4.7*b* which only includes weather attributes with a correlation below 0.9: previous population, the number of days with >= 0.01 inch (0.254 mm) of precipitation in the year, DP01, the highest daily total of precipitation in the year, EXMP, and the total annual precipitation, PRCP. Correlations are determined via the correlation plot and in the case that two features have a correlation value greater than 0.9, one of the correlated features is removed. This process is repeated until only correlations less than 0.9 remain. Learning curves demonstrate the model's accuracy of predicting population increase or decrease based on the number of training examples included in model training and the accuracy of the model in correctly predicting population change in the reserved validation set. Figure 4.7*a*,*b* includes subplots of the learning curves for DTC with defined maximum depths of 2, 3, 4, and 7. A well-fitting model will exhibit convergence of the training and validation scores, whereas a model that demonstrates a higher training score, but lower validation scores, is overfit. This occurs when the model is optimized to the data in the training set to achieve a higher score, but the optimization fails to improve the validation score as it is biased toward the data it is trained on. When all features are included during model training (Figure 4.7*a*), the model nearly converges for the maximum depth of 2 and 4 and shows convergence for a maximum depth of 3. When highly correlated (<0.9) features are removed (Figure 4.7*b*), the model convergence occurs for a maximum depth of 2, converges with less training examples for a maximum depth of 3, and nearly converges

for a maximum depth of 4. In both cases of feature use, a maximum depth of 7 demonstrates overfitting, with the suggestion of possible convergence with enough data points, but the trends do not demonstrate increasing the accuracy of population change predictions for the validation set compared to models trained with less depth. While the differences are minor and the model scores similarly with and without the correlated features, the training and validation scores converge with fewer training examples when the features are removed. Therefore, the reduced set of features for both Datasets 1 and 2 are used for all further investigation.

Hyperparameter tuning can impact the performance of a model, as seen in the learning curves for default hyperparameters, specifying maximum depth to 3, and tuned hyperparameters for DTC (Figure 4.8*a-c*) and RFC (Figure 4.8*d-f*). Hyperparameter tuning used a grid search over a range of values for multiple tuning parameters for each classifier, selecting the highest performing hyperparameter values. In both DTC and RFC, default hyperparameter settings fail to generalize, due to the lack of a maximum depth. Once the maximum depth is set to 3, the training score drops, but the models have a much better fitting model with signs of generalization, and when the models are set with the best performing hyperparameters, the model requires the least number of training values to converge.

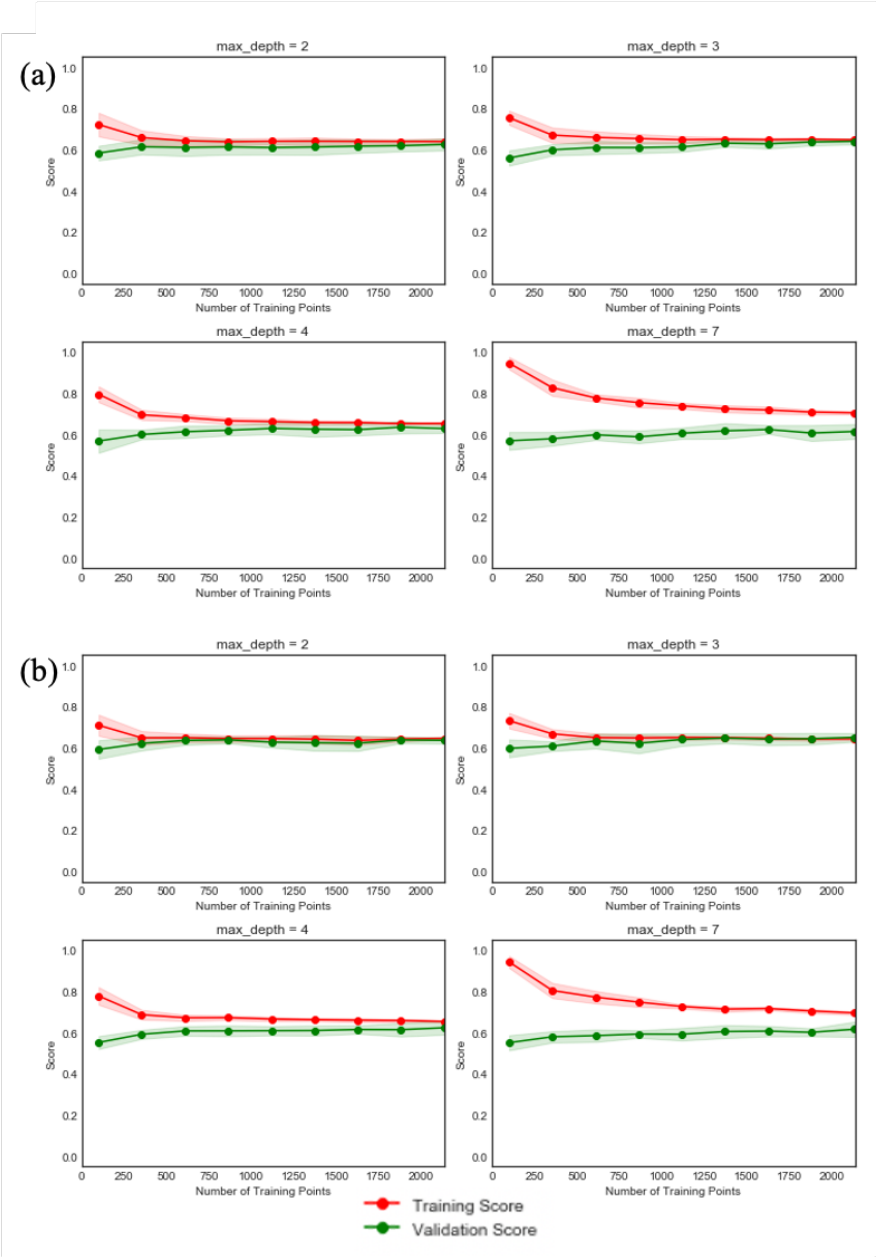

Figure 4.7 Decision Tree Learning Curves – Dataset 1. Learning curves of the Decision Tree Classifier for Dataset 1 including all features (a) and with the removal of highly correlated features DP1X, and DP10 (b) are shown. Each learning curve is created by randomly selecting 10% of the dataset to be reserved as a validation set a total of 10 times. Each training and validation set is fit to a model with solid red lines showing the mean score of the 10 training set iterations, and the green lines showing the mean score of the validation set. Standard deviations are shown as an opaque fill for the training (red) and validation (green) scores. Each subplot includes learning curves with the decision trees restricted to a maximum depth of 2, 3, 4, and 7. Accuracy is the designated scoring method.

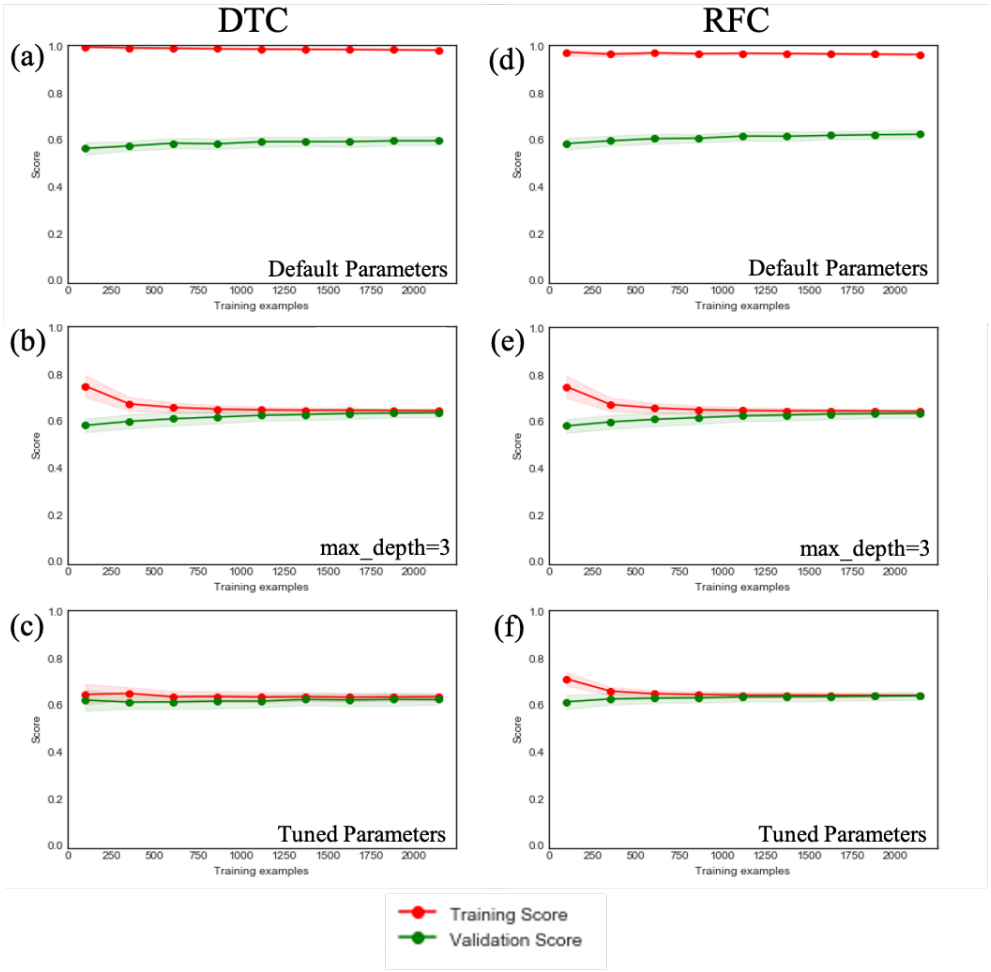

Figure 4.8 Learning Curves and Hyperparameter tuning – Dataset 1. Learning curves of DTC with default hyperparameters (a), default hyperparameters except for max depth (b), and multiple tuned hyperparameters (c) are shown and learning curves of RFC with default hyperparameters (d), default hyperparameters except for maximum depth (e), and tuned hyperparameters (f) are shown. In all cases, mean training scores are shown in red with standard deviation in shaded red, and the mean validation score is shown in green with standard deviations in shaded green.

Hyperparameter tuning has similar effects of model fitting for Dataset 2, as seen in Figure 4.9. Default hyperparameters for DTC (*a*) and RFC (*d*) fail to generalize, but the training and validation generalize for both models once a maximum depth is defined, as seen for DTC (*b*) and RFC (*e*). Surprisingly, optimizing DTC hyperparameters results

in a decrease in model generalization when compared to a maximum depth of 3 and otherwise default hyperparameters. The results of both models suggest that the model would likely converge with a greater number of training examples.

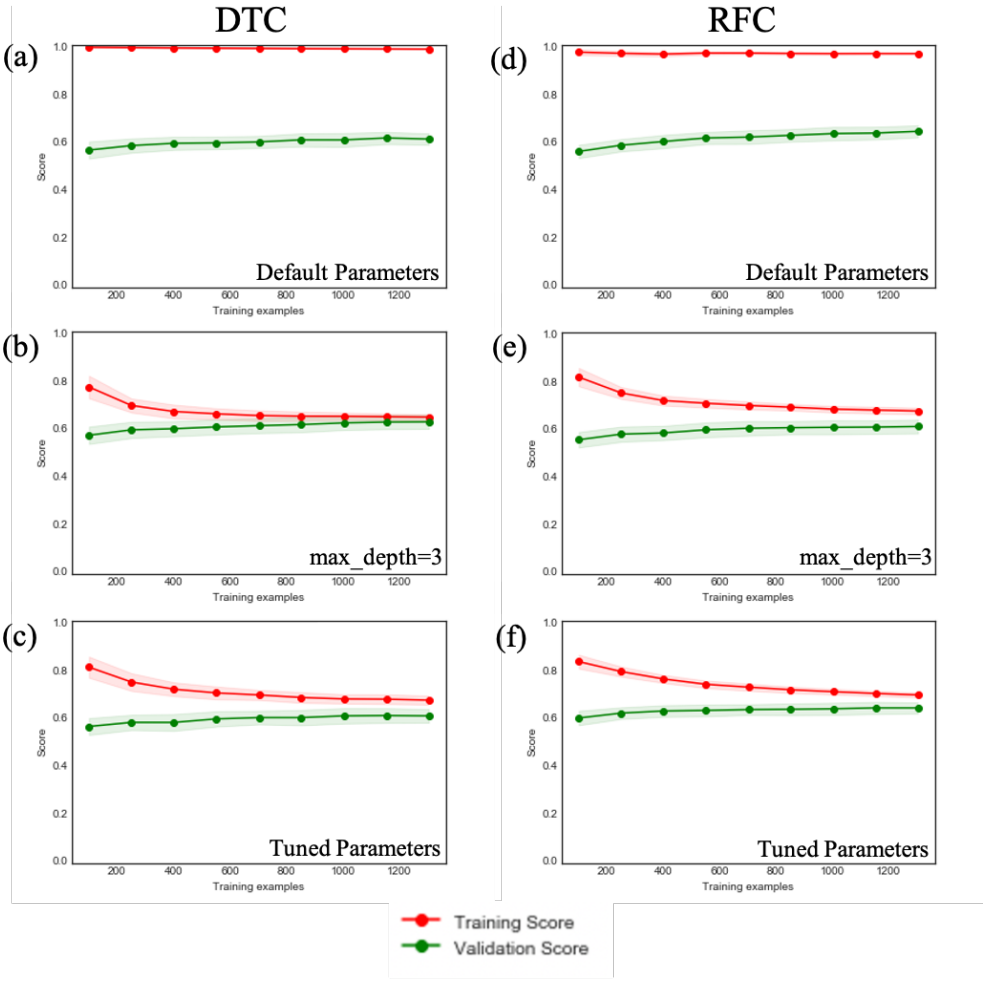

Figure 4.9 Learning Curves and Hyperparameter tuning – Dataset 2. Learning curves of DTC with default hyperparameters (a), default hyperparameters except for maximum depth (b), and multiple tuned hyperparameters (c) are shown and learning curves of RFC with default hyperparameters (d), default hyperparameters except for maximum depth (e), and tuned hyperparameters (f) are shown. In all cases, mean training scores are shown in red with standard deviation in shaded red, and the mean validation score is shown in green with standard deviations in shaded green.

Receiver operating characteristic (ROC) curves for Dataset 1 DTC (*a*) and RFC (*b*) and Dataset 2 DTC (*c*) and RFC (*d*) are shown in Figure 4.10. A ROC curve plots the true positive rate (TPR) corresponding to the percentage of training examples correctly classified as population increase against false positive rate (FPR) corresponding to the percentage of training examples incorrectly classified as population increase while classification threshold varies (Krzanowski & Hand, 2009). Models predict a probability of class membership which is then mapped to the predicted label of population increase or decrease, and the classification threshold defines the probability that determines classification. A high performing model would consist of almost entirely true positive cases, resulting in a curve near the y-axis until the TPR reaches 1, and a random classifier model is expected to fall on the diagonal line shown in red. For all threshold values, both models perform better than a random classifier, with the exception of RFC with a maximum depth of 3 fitting Dataset 2, which dips below the diagonal for high TPR and FPR conditions and the untuned RFC for low TPR and FPR. For Dataset 1, DTC appears to have the highest TPR when only the maximum depth is not default, with a more desirable curve than the tuned DTC. The RFC curves for Dataset 1 show little performance difference for all three hyperparameter sets tested. Hyperparameter tuning results in greater improvement of fit for Dataset 2 for both DTC and RFC. The ROC curve for Tuned DTC (Figure 4.10*c*) shows a steeper increase in TPR then begins to level out as FPR increases, whereas Figure 4.10*d* shows the Tuned RFC beginning at a higher TPR for FPR = 0.0 and increases more linearly until it reaches a TPR of 1.

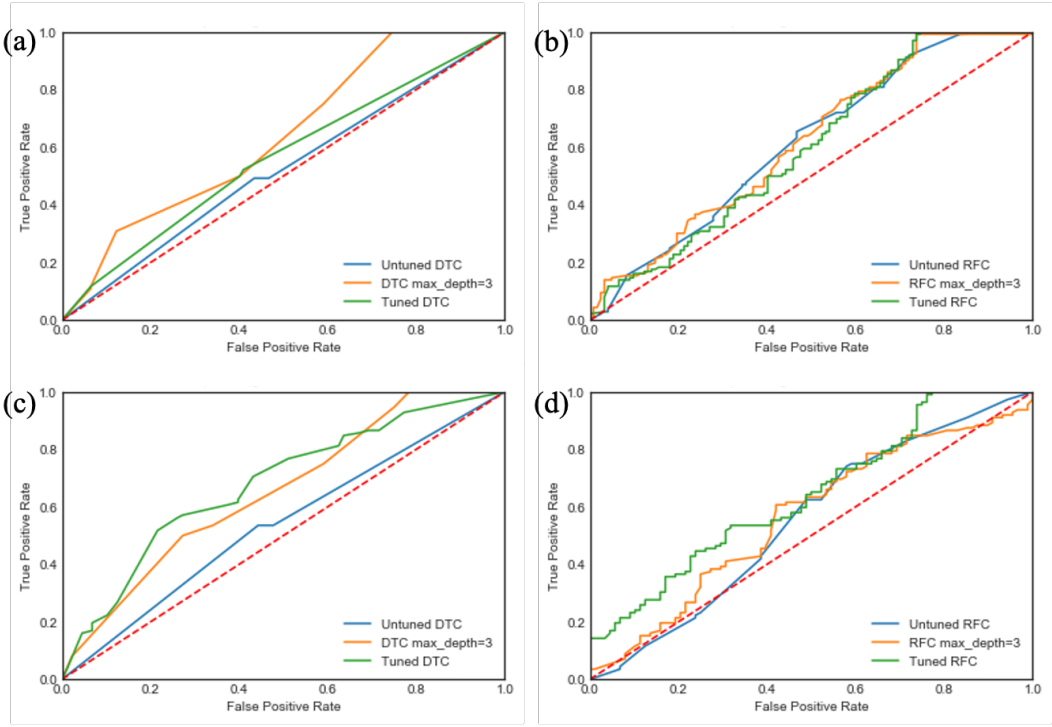

Figure 4.10 ROC Curves. The ROC curves for Dataset 1 DTC (a) and RFC (b) and Dataset 2 DTC (c) and RFC (d) are shown. Each plot contains curves for default tuning parameters (Untuned), specified max depth of 3 (max_depth=3) and for "best" tuning parameters (Tuned) for the respective classifier.

Moving from classification to regression, the learning curves in Figure 4.11 show the training and validation scores of a linear regression model. Figure 4.11*a* shows the learning curve of all features of Dataset 1 included, whereas *b* does not include the highly correlated features. Similarly, the linear regression learning curves for Dataset 2 including all features and removing highly correlated features are shown in Figure 4.10*c* and *d* respectively. For both datasets, removing correlated features increases the fit of the model resulting in closer training and validation scores with less training examples. In the case of Dataset 1, both curves show convergence, whereas the fit of Dataset 2 approaches convergence, but does not converge with the number of available training examples. Both

results for Dataset 2 suggest that additional training examples could improve the model fit, with a greater likelihood if correlated features are removed. Both instances demonstrate convergence with a score around R2 = 0.2, but the removal of the highly correlated features lowers the number of training examples required for the scores of the training and validation set to be equal.

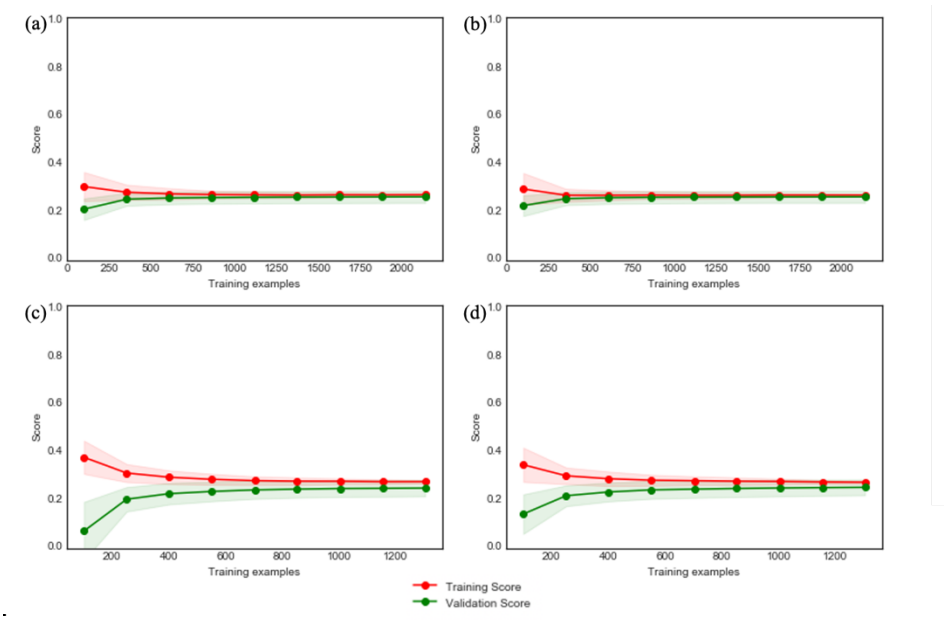

Figure 4.11 Linear Regression Learning Curves. Learning curves of the Linear Regression Model for Dataset 1 including all features (a) and with the removal of correlated features DP1X, and DP10 (b) are shown. Linear regression model results of Dataset 2 including all features and the removal of highly correlated features DP1X, DP10, DSNW, and DX70 are shown in (c) and (d) respectively. The training scores are plotted in red and validation scores in green. The scoring method is R2 for both plots.

Once models have been selected and fit to the datasets, the scores shown in Table 4.1 are calculated by comparing the model with the test data that was removed before model training began. This score represents the overall model performances for new data that was not used for training, simulating the performance of the model on new

population and weather data. Note that the scoring for the classification models is still based on accuracy of classification, and the scoring for regression uses R2 for the scoring metric. Overall, the scores for DTC and RFC are similar, with slight improvement from tuning the model hyperparameters. DTC fitting Dataset 2 with optimized hyperparameters has the best score overall of 0.6550. LR scored slightly higher with the inclusion of more features in Dataset 2 when compared to Dataset 1, though the improvement is minimal.

Table 4.1 Test Data Scores

Dataset	Model	Tuning parameters	Scores
Dataset 1	DTC	Default	0.5271
		Max depth=3	0.5465
		Optimized Parameters	0.5543
	RFC	Default	0.5581
		Max depth=3	0.5543
		Optimized Parameters	0.5349
Dataset 2	DTC	Default	0.5350
		Max depth=3	0.5900
		Optimized Parameters	0.6550
	RFC	Default	0.5100
		Max depth=3	0.5250
		Optimized Parameters	0.5700
Dataset 1	LR	Default	0.1979
Dataset 2	LR	Default	0.2153

4.3. DISCUSSION

The purpose of this proof of concept was to determine whether the use of a machine learning model would add benefit to predicting population changes as a function

of measured weather attributes. Each model was able to show improvement over random choice with the data available, confirming that population changes are at least partially described by weather parameters. General machine learning techniques improved the model's ability to predict population change, beginning with the removal of highly correlated weather attributes. This comes at a tradeoff of further reducing the available weather attributes, and with the limited number of attributes measured and available, only attributes with correlation values above 0.9 were removed. The comparisons of Figure 4.7*a* and *b* demonstrate that, in the case of a decision tree, the removal of the highly correlated features reduces the number of training examples required for the training and validation scores to converge. The removal of these correlated features did not, however, improve the model's ability to accurately predict population change. The decision tree was chosen for this initial effort as a decision tree model is one of the least sensitive to correlated features. Nodes are selected based on entropy, and therefore the model will select a feature with highest entropy; if correlated features perform equally well, only one will be used and the other will be ignored (Mitchell, 1997). Linear regression, however, is sensitive to correlations in the features since the hypothesis function is a summation with a term for each feature. Introducing highly correlated data is similar to introducing one feature multiple times. The impact of correlated features for LR are shown in Figure 4.11, and the number of training examples required for model generalization is less when the correlated features are removed. Removal of these features did not result in an improvement in the model's ability to predict population change.

The subplots in Figure 4.7 also demonstrate the need to establish a maximum depth to the DTC, and correspondingly the RFC, to prevent model overfitting. The subplots of training curves for maximum depths of 2, 3, and 4 demonstrate a well-fitting model, which predicts with the same accuracy from the training and the validation data, whereas the training curves for a maximum depth of 7 show the model is overfit. The additional decision levels result in a higher score for the training data, but this improvement does not translate to the model accurately predicting, and therefore does not signify increased model performance. Comparisons of learning curves of the DTC and RFC with default hyperparameters provide examples when no there is no specified maximum depth in Figure 4.8*a*,*d* and Figure 4.9*a*,*d*. In both datasets, the models were significantly overfit with default hyperparameters.

With limited available data, the "curse of dimensionality," that as the number of features increases, the required amount of training examples also increases, becomes apparent (Bellman & Dreyfus, 1962). Dataset 1 demonstrates model performance when higher priority is given to available training examples, and Dataset 2 demonstrates model performance with higher priority on available weather attributes. The learning curves for Dataset 2 in Figure 4.9 show that Dataset 2 can provide a well-fitting model with both DTC and RFC, but additional training examples would further improve model fitment. By comparison, Figure 4.8 demonstrates that for the available weather attributes, the number of training examples in Dataset 1 is sufficient for a well-fitting model. Comparisons of the results for all three models in Table 4.1 show that for the amount of data included in the study, both datasets have a limitation. The accuracy of the models on the test data suggests that an increase in the number of weather attributes will have a

greater benefit than an increase of training examples for model training. The ROC curves in Figure 4.10 further supports this, in that for most threshold values, models fit to Dataset 2 have a higher TPR and lower FPR than the same model's fit to Dataset 1. The ROC curves also show that parameter tuning made an impact on model performance, resulting in more desirable curves when compared to the untuned models, with a more noticeable improvement for Dataset 2.

Overall, in the case of classification, the previous population and lesser correlated weather attributes were able to most accurately predict a population increase or decrease with a DTC and hyperparameter tuning on Dataset 2. The scores in Table 4.1 show the accuracy of this model was highest with 0.655 or 65.5% of the population change predictions of the test set were correct, though all of the models showed some improvement in prediction accuracy over random choice. Both classifiers showed higher scores in the case of Dataset 2, indicating that the inclusion of additional weather attributes can improve the model's prediction accuracy.

Along with a greater number of weather attributes, Dataset 2 provides information about more environmental factors, giving possible explanation for improved model scores. Unfortunately, of the data at the time and location these studies were completed, the only weather attributes providing measured values for 80% of the datapoints were all associated with precipitation, whereas lowering the threshold to 70% included information about precipitation, temperature, and snowfall. The inclusion of data describing the temperature and snowfall provides a more complete picture of the local environment, giving further explanation for the improved model scores.

Thus far, discussion has focused primarily on classification models that predict if a population will increase or decrease. Predictions of the population change is a greater challenge, and one that cannot be handled with a classification model, as a classification model cannot predict a value it was not trained on. Regression models provide the ability to generalize to outputs that the model has not seen before, like a population change larger than the maximum in the training examples. However, predictions of a positive or negative population change is a simpler task, offering explanation to why results of LR were poor in comparison to the classification models. That said, the use of LR was still more advantageous than a model that consistently predicts the mean value, as shown by R2 scores above 0 in Table 4.1. In both Dataset 1 and Dataset 2, the LR model was able to generalize, suggesting that for the number of features available, the number of training examples was sufficient, but the features themselves do not linearly describe the population changes, and therefore LR is not an optimal approach.

Overall, the use of classification models to answer the simple question of "will a population increase or decrease" provides a benefit over a random selection model (a model that places the training examples into one of the classifications at random), but it is apparent that the included weather features, in combination with the previous year's population, does not fully describe the population change, suggesting that further optimization of the model requires additional feature inputs. The same holds true for the regression model tested.

4.4. CONCLUSION AND FUTURE WORK

The work presented in this section demonstrates that predictions of population change from yearly weather summaries of a location are possible but have significant room for improvement. The primary limitation in improving the performance of each model stems from the limited available weather attribute data. In the case that data collection improves the amount of available population and weather data, model performances would likely increase.

With the limited dataset, populations of all animals were consolidated, but this presents the assumption that all of the included species will be similarly impacted by the available weather attributes. An increase in available population data can offer the ability to divide the dataset by order, genus or even species, likely improving any model's performance. An increase in available population data may also allow for the examination of only declining populations in an effort to determine early warning signs of a struggling population. The identification of key weather attributes to predict population decline and even population collapse could allow for more efficient intervention efforts of populations approaching collapse.

Additionally, improvements in weather data collection would offer additional features and an improved picture of the local environment. As shown above, the inclusion of more features can improve model performance, but may also offer the opportunity to lower the correlation threshold with enough remaining features to create a model. The data that has been used here ranges from 1970 to 2016, suggesting that population data from more recent collection efforts may be accompanied by improved weather data,

resulting in more fruitful efforts in the future. The inclusion of other features such as terrain data, natural disaster information, and even pollution from human interaction in the area may offer additional improvements to model performance.

5. CONCLUSION

With a significant decline in global population abundance since 1970 (68% decrease), it is clear that environmental changes experienced in today's world can have a significant impact on population survival (WWF, 2020). The task of predicting and intervening in population decline is generally understood as a biological problem. However, an interdisciplinary approach incorporating physics-based and machine learning techniques can provide valuable insight into population dynamics. The results from both the computational model and the laboratory experiments presented in this dissertation demonstrate that the response of a population under stress correlates with system parameters or environmental stressors driving population decline. The computational model demonstrated differences in population decline for the different control parameters, while also highlighting the effects of additional noise in the form of quenched randomness. The results show that the transition from survival to extinction does not fall into the directed percolation universality class; the system's scaling behavior near the critical point was shown to depend on the particular control parameter causing the population decline. A difference in scaling behavior near the critical point was previously identified to correlate with differences in population decline and how a population may recover from environmental perturbations (King, 2015). This physics-based approach allowed for the use of computational models to gain greater understanding of global dynamics in a population efficiently.

Similarly, both the *S. cerevisiae* and *E. coli* experiments demonstrated that the mechanism driving population decline dictated the system's behavior during the

transition. The yeast population responded differently to salt stress and temperature stress, exhibiting phase transition-like behavior in the latter case. These different responses may be explained by the physiological response of the yeast to the stressor the population experiences (Capusoni et al., 2019; Gore et al., 2009; Hohmann, 2002; Leuenberger et al., 2017; Murguía et al., 1996; Szopinska et al., 2011; Verghese et al., 2012). As seen in the decay curves, the *E. coli* population response to antibiotics (shown in Figure 3.3) demonstrated that even with similar modes of action, the populations had different transition behavior. This suggests that similarly-acting antibiotics do not have similar effects on these bacteria. The microbiological approach shown here can be expanded upon in further investigation to explain why there are different responses to a particular population's stressor. This will help to increase the understanding of an individual stress on a specific population, and it could further provide an understanding as to why some species are able to adapt to some stressors more effectively than others.

Neither the physics-based nor the experimental approach directly affected the ability to identify a declining population, but each provided a quantitative characterization of individual species behavior based on environmental stressors or system parameters, and how these populations may decline or recover. Section 4 is a step towards accurate identification of struggling populations. When combined with an understanding of the transition behavior based on an identified stressor, the possibility of identifying declining populations and curating efficient and effective efforts to intervene becomes a more manageable struggle. Implementing machine learning techniques using a handful of weather attributes alone demonstrated some capability to predict population changes. Use of narrower species criteria will almost certainly increase model

performance, as it would increase specificity in the population response to stressors. Additionally, including intrinsic traits and external factors that have been identified as predictive of population decline may also provide increased prediction performance (Collen et al., 2011).

Each of the three disciplinary approaches provides an understanding of populations under stress, but the combination of tools and techniques across disciplines presented here provide a more complete picture in understanding population declines and potential response to recovery efforts. Combining the information accessible through each approach presents the possibility of more accurate computational models, a better understanding of species response, and presents a chance for optimization of predictive modeling, all of which encourage the creation of effective and efficient methods of population recovery intervention.

APPENDIX A.

NOAA WEATHER FEATURES

The following is a list of the weather features that have been included in the analysis of section 4 with definitions as stated by NOAA. A full list of all available weather attributes from NOAA can be found at https://www.ncei.noaa.gov/pub/data/metadata/documents/GSOYReadme.txt.

TMAX – Average Annual Maximum Temperature. Average of the mean monthly maximum temperatures given in Celsius or Fahrenheit. Missing if one or more months are missing or flagged.

DP01 – Number of days with >= 0.01 inch (25.4 mm) of precipitation in the year.

DP10 – Number of days with >= 0.1 inch (25.4 mm) of precipitation in the year.

DP1X – Number of days with >= 1.00 inch (25.4 mm) of precipitation in the year.

DSNW – Number of days with snowfall >= 1 inch (25 mm).

DT00 – Number of days with maximum temperature <= 0 °F/-17.8 °C.

DT32 – Number of days with minimum temperature <= 32 °F /0 °C.

DX70 – Number of days with maximum temperature >= 70 °F /21.1 °C.

DX90 – Number of days with maximum temperature >= 90 °F /32.2 °C.

EMNT – Extreme minimum temperature for the year. Lowest daily minimum temperature for the year. Given in Celsius or Fahrenheit.

EMSN – Highest daily snowfall in the year, given in inches or millimeter.

EMXP – Highest daily total of precipitation in the year, given in inches or millimeters.

EMXT – Extreme maximum temperature for the year. Highest daily maximum temperature for the year. Given in Celsius or Fahrenheit.

PRCP – Total Annual Precipitation. Given in inches or millimeters.

SNOW – Total Annual Snowfall. Given in inches or millimeters.

TMIN – Average Annual Minimum Temperature. Average of the mean monthly minimum temperatures given in Celsius or Fahrenheit. Missing if one or more months are missing or flagged.

APPENDIX B.

ORDWAY *ET AL.* (2020)

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Ordway SW, King DM, Friend D, Noto C, Phu S, Huelskamp H, Inglis RF, Olivas W, Bahar S. 2020 Phase transition behaviour in yeast and bacterial populations under stress. *R. Soc. Open Sci.* **7**: 192211.
http://dx.doi.org/10.1098/rsos.192211

Received: 9 January 2020
Accepted: 5 May 2020

Subject Category:
Physics and biophysics

Subject Areas:
biophysics

Keywords:
phase transitions, population collapse, *Saccharomyces cerevisiae*, *Escherichia coli*, temperature stress, salt stress

Author for correspondence:
Sonya Bahar
e-mail: bahars@umsl.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5063422.

THE ROYAL SOCIETY PUBLISHING

Phase transition behaviour in yeast and bacterial populations under stress

Stephen W. Ordway[1], Dawn M. King[1], David Friend[1,2], Christine Noto[2], Snowlee Phu[2], Holly Huelskamp[2], R. Fredrik Inglis[2], Wendy Olivas[2] and Sonya Bahar[1]

[1]Department of Physics and Astronomy, and [2]Department of Biology, University of Missouri – St. Louis, Saint Louis, MO, USA

DMK, 0000-0002-3665-3746; SB, 0000-0002-9064-9377

Non-equilibrium phase transitions from survival to extinction have recently been observed in computational models of evolutionary dynamics. Dynamical signatures predictive of population collapse have been observed in yeast populations under stress. We experimentally investigate the population response of the budding yeast *Saccharomyces cerevisiae* to biological stressors (temperature and salt concentration) in order to investigate the system's behaviour in the vicinity of population collapse. While both conditions lead to population decline, the dynamical characteristics of the population response differ significantly depending on the stressor. Under temperature stress, the population undergoes a sharp change with significant fluctuations within a critical temperature range, indicative of a continuous absorbing phase transition. In the case of salt stress, the response is more gradual. A similar range of response is observed with the application of various antibiotics to *Escherichia coli*, with a variety of patterns of decreased growth in response to antibiotic stress both within and across antibiotic classes and mechanisms of action. These findings have implications for the identification of critical tipping points for populations under environmental stress.

1. Introduction

As our planet edges ever closer to the 'Sixth Extinction' [1–4], environmental stressors are increasing, and the ecological effects of these stressors on ecosystems are becoming more apparent. As a result, identification of early warning signs of population collapse has become a priority [5–11]. Population decline has been suggested to be accompanied by critical slowing down [5–7,10], a dynamical phenomenon associated with scale-free, power-law dynamics, though other studies have observed population decline

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science.

in the absence of such critical signatures [12]. Importantly, due to critical slowing down, populations recover more slowly from perturbations in the neighbourhood of tipping points between survival and collapse [10]. The identification of indicators of incipient population decline is of increasing urgency in the light of collapse of pollinator [13–15] and avian [16] communities, among others. Recent studies have identified evidence for worldwide decline in insect populations [17–20], which could have catastrophic cascading effects on the global ecosystem.

Physics-based models of *phase transitions* provide useful models for extinction processes. In a phase transition, a system undergoes a change in the value of an *order parameter* characterizing the system's state, as a *control parameter* is varied. At a critical value of the control parameter, the system passes a tipping point and undergoes a change of state, such as from survival to extinction. The system's characteristic fluctuations in the neighbourhood of the transition can serve as a warning sign for the incipient state change. In *critical* (also called *continuous* or *second-order*) transitions, these fluctuations are characterized by a large standard deviation in the order parameter in the vicinity of the transition. The dynamics of the fluctuations can be quantified by *critical exponents*, which define the *universality class* of the phase transition, and hence the system's behaviour as it passes through the transition. Widely different systems can undergo phase transitions in the same universality class. In *non-equilibrium* phase transitions, a system undergoes a transition to an 'absorbing state', from which it cannot recover [21–22].

Computational evolutionary models of transitions from survival to extinction can be characterized as non-equilibrium phase transitions. A recent agent-based computational evolutionary model has been shown to exhibit a non-equilibrium phase transition with behaviour similar, but not identical, to the directed percolation universality class [23–27]. This behaviour has been observed during simulated transitions from survival to extinction as maximum mutation size [23–27] and death rate [26,27] are varied as control parameters.

Dai and colleagues have performed population dynamics experiments to investigate the stability and resilience of yeast cultures subjected to environmental stressors [8–11]. Studying their system through the lens of nonlinear dynamics, they mapped the dynamics of stable and unstable populations in response to environmental stressors such as dilution factor (a proxy for death rate) [8–10], nutrient (sucrose) concentration (a proxy for carrying capacity) [10] and osmotic stress (NaCl concentration) [8,10]. These studies resulted in a means of characterizing a population's resilience, i.e. its ability to recover from a large environmental perturbation [10].

In the present work, we adapt the experimental approach used by Dai and colleagues in order to investigate the decline of the budding yeast *Saccharomyces cerevisiae* and *Escherichia coli* populations under stress. We investigate the dynamics of yeast population decay in the presence of two different stressors, temperature and salt (NaCl) concentration, and the decay of *E. coli* populations in response to a range of antibiotics.

2. Material and methods

2.1. *Saccharomyces cerevisiae* experiments

Experiments were performed using the yWO3 [28] strain of *S. cerevisiae*. This strain was selected because it is a well-described wild-type laboratory strain that is neither thermophilic nor thermotolerant. To investigate the dynamics of yeast population growth under stress, two environmental stressors were used, temperature and elevated NaCl concentration. In both cases, *S. cerevisiae* was initially grown at 30°C, which is the optimum temperature for *S. cerevisiae* growth [29], in standard medium (YEPD) containing 10 g l^{-1} yeast extract, 20 g l^{-1} peptone and 2% dextrose. These initial 50 ml cultures in liquid media were inoculated from a plate with cells to an optical density (OD) of 0.0001 to allow multiple doublings in log phase over 24 h, thereby creating a large concentration of cells without saturating the culture. From the starting OD 0.0001 concentration, the cells typically grew to an OD of 2.5 in 24 h. These initial cultures were then used as starter cultures to inoculate fresh media for the temperature stress and salt stress experimental cultures. All OD measurements were taken at 600 nm and measured using a Turner visible spectrophotometer.

2.1.1. Temperature stress

Wild-type *S. cerevisiae* is known to grow optimally at 30°C [29] and to exhibit a sharp decline in growth rate at temperatures exceeding approximately 40°C [30]. To study the response to temperature stress, the

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

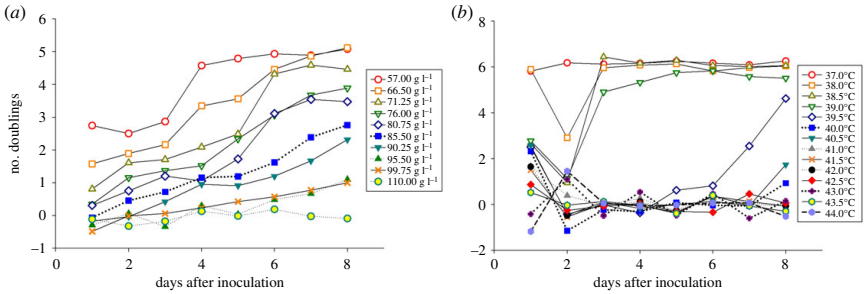

Figure 1. Number of *S. cerevisiae* population doublings is shown as a function of days after initial inoculation for various values of NaCl concentration (*a*) and temperature (*b*). $N \geq 3$ for each data point; error bars are not shown to avoid crowding the figure (see electronic supplementary material, table).

initial culture, prepared as described above, was inoculated into 50 ml of YEPD in 250 ml flasks to a resulting OD of 0.05. The samples were then placed in orbital shaking water baths that had already reached the particular temperature of interest for a given set of measurements. Sample ODs were then measured every 24 h. If the sample had grown, it was diluted back to an OD of 0.05 in a volume of 50 ml YEPD in a new 250 ml flask. If the sample had not grown significantly (less than 0.01 OD growth), or if the OD of the sample was lower than 0.05, the sample was placed back into the water bath for another 24 h. Measurements were taken over eight 24 h periods. Growth was measured by calculating the number of times the population doubled between measurement cycles, or number of doublings, using

$$n = \frac{(\log(\mathrm{OD}_{\mathrm{final}}/\mathrm{OD}_{\mathrm{initial}}))}{\log(2)}, \tag{2.1}$$

where $\mathrm{OD}_{\mathrm{initial}}$ is the OD at the beginning of each 24 h growth period, and $\mathrm{OD}_{\mathrm{final}}$ is the measured OD after each 24 h growth period.

For the measurements of growth under temperature stress, sample cultures were measured from 38°C to 44°C at 0.5°C intervals with a sample size $N \geq 3$, with an additional set of $N = 3$ measurements at 37°C. Since 37°C was well into the survival regime, as can be seen from the steady-state growth curve at this temperature in figure 1, and far from any population tipping points, measurements under this condition served as a reference point to confirm any emerging trends. Data points taken at 30°C (not shown) confirmed that the growth rate at this standard temperature for yeast growth was similar to the growth observed at 37°C.

2.1.2. NaCl stress

Saccharomyces cerevisiae growth was investigated over a range of salt concentrations, from concentrations under which cells are documented to grow normally to concentrations which are known to cause considerable stress, including cell death [31–35]. For the measurement of yeast growth under high salt stress, NaCl was added to freshly prepared YEPD in order to achieve the desired NaCl concentration. Culture growth was measured for NaCl concentrations from 66.5 to 104.5 g l^{-1} at approximately 5 g l^{-1} intervals with a sample size $N \geq 3$. An additional set of $N = 3$ measurements at 57 g l^{-1}, which is well into the survival regime, as can be seen from the steady-state growth curve at this salt concentration in figure 1, served as a reference point to confirm any emerging trends.

An initial culture in YEPD was inoculated into 50 ml of YEPD + NaCl in 125 ml flasks to a resulting OD of 0.05. The samples were then placed in an orbital shaking water bath at 30°C. As with the temperature studies described above, sample ODs were measured after 24 h at 600 nm using a Turner visible spectrophotometer. If growth had occurred, the sample was diluted back to an OD of 0.05 in a volume of 50 ml YEPD + NaCl in a new 125 ml flask. If the sample did not have significant growth (less than 0.01 OD), or if the OD was less than 0.05, the sample was returned to the incubator until the next measurement period. Measurements were taken over eight 24 h periods, and doublings were calculated using equation (2.1). All data, for both temperature and salt stress, are available in [36].

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

Table 1. Mechanisms of action of antibiotics used.

antibiotic	class	mechanism	mode of action
ampicillin	β-lactam	inhibits cell wall synthesis	bactericidal
carbenicillin	β-lactam	inhibits cell wall synthesis	bactericidal
chloramphenicol	amphenicol	protein synthesis (50S inhibitor)	bacteriostatic
ciprofloxacin	fluoroquinolone	inhibits DNA gyrase	bactericidal
gentamycin	aminoglycoside	protein synthesis (30S inhibitor)	bactericidal
kanamycin	aminoglycoside	protein synthesis (30S inhibitor)	bactericidal
rifampicin	rifamycin	DNA-directed RNA polymerase	bactericidal
spectinomycin	aminoglycoside	protein synthesis (30S inhibitor)	bacteriostatic
streptomycin	aminoglycoside	protein synthesis (30S inhibitor)	bactericidal
tetracycline	tetracycline	protein synthesis (30S inhibitor)	bacteriostatic

2.2. *Escherichia coli* experiments

Escherichia coli (MG1655) were grown in M9 minimal glucose media overnight, shaking at 37°C. This bacterial culture was used to inoculate 96-well plates containing a dilution series of 10 different antibiotics, at a starting OD (600 nm) of 0.005 (i.e. 2 µl of culture in 198 µl media). We selected a panel of 10 antibiotics across different antibiotic classes with a variety of mechanisms of action (table 1). Each antibiotic was added at a starting concentration of 100 µg ml^{-1} and subsequently serially diluted by ¾ to give 40 different concentrations, ranging from 100 to 0.001 µg ml^{-1}. There were six replicates for each antibiotic. The 96-well plates were incubated, shaking at 37°C, for 24 h. The OD (600 nm) for each well was measured using a Cytation 3 multimode plate reader (BioTek). Doubling times for each well were calculated as a function of initial and final OD using equation (2.1) as described above. All data are available in [36].

3. Results

As shown in figure 1, increases in both environmental stressors caused a decrease in yeast growth rate, as would be expected. However, the time course of the stress response differed significantly between the stressors. The data shown in figure 1*a* indicates that increasing salt concentrations caused the yeast to grow more slowly, but with a gradual shift until growth ceased altogether. By contrast, as shown in figure 1*b*, as the temperature was increased, a sharp drop in doubling rate was observed. Here, the cultures changed from a quick recovery time in the 37–39°C range, to a recovery after a long lag time (39.5–40.5°C range) and finally to being unable to recover at all for temperatures above 40.5°C. Standard deviations are provided in electronic supplementary material, Data tables 1 and 2, but not in figure 1, to avoid crowding the figures.

In order to investigate the population dynamics in the neighbourhood of the 'tipping point' into decline, average growth over the last 3 days was calculated as a function of NaCl concentration (figure 2*a*) and temperature (figure 2*b*). As can be seen in figure 1, in most cases, the populations had reached an approximately constant growth rate by post-inoculation days 6–8. Averaging the growth rates over these three days thus serves as a proxy for a 'steady-state' growth rate. In figure 2, these averages are shown, with a grand average taken over all experiments at each condition. This gives a measure of the system's overall response to each stressor.

It can be seen from figure 2 that, for both stressors, the yeast population ultimately experienced an environment too harsh for survival. However, in the case of increasing salt concentrations (figure 2*a*), the system underwent a smooth, gradual decline. Minimal fluctuations were observed in the transition region, and there was no identifiable tipping point. In stark contrast, the response to temperature stress (figure 2*b*) showed an abrupt drop in growth rate, with comparatively large fluctuations in the transition region, indicating a tipping point around 39.5°C.

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

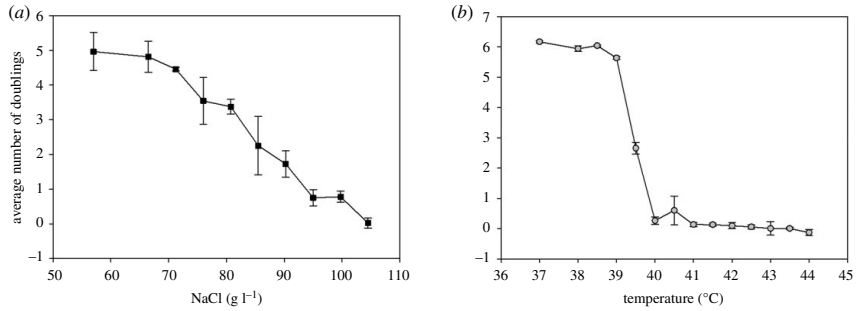

Figure 2. Steady states of the *S. cerevisiae* population for NaCl stress (*a*) and temperature stress (*b*) are shown. For each condition, the populations are first averaged over days 6–8 post-inoculation, then over all runs at that condition and finally averaged over all experiments at that condition. Error bars show standard deviation.

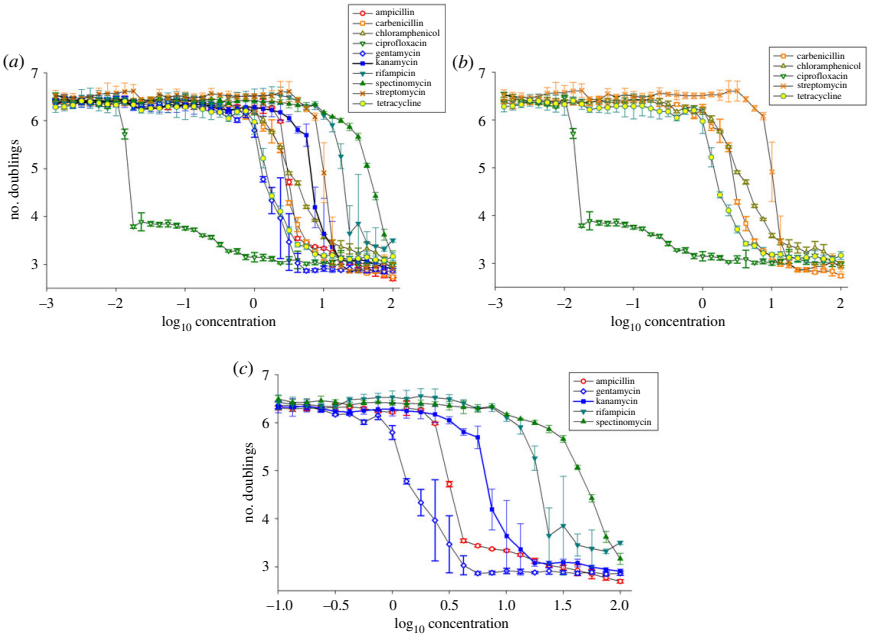

Figure 3. (*a*) Number of doublings of *E. coli* (MG1655 WT) populations as a function of $\log_{10}$(antibiotic concentration) for ten antibiotics. Error bars show standard deviation over six replicates for each antibiotic concentration. (*b*) Number of doublings for five of the studied antibiotics over a smaller concentration range, in order to show detail. (*c*) Number of doublings for the other five studied antibiotics over a smaller concentration range, in order to show detail.

In figure 3, we show the decay curves of *E. coli* populations in the presence of various antibiotics. All the antibiotics tested are shown in figure 3*a*. In figure 3*b* and *c*, the responses to five each of the ten antibiotics studied are shown (on a shorter horizontal scale, in the case of 3c), in order to display the details of the responses more clearly. The presence of antibiotic stress caused a decrease in bacterial growth across all the antibiotics tested, as expected. However, we observed a variety of patterns of decreased growth in response to antibiotic stress both within and across antibiotic classes and mechanisms of action. This ranged from a very sharp drop in bacterial growth in the presence of ciprofloxacin (open green triangles, figure 3*a* and *b*) to more gradual declines in the presence of spectinomycin (filled green triangles, figure 3*a* and *c*). This broad range of response patterns is qualitatively similar to those observed in the yeast experiments.

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

4. Discussion

We have examined the response of *S. cerevisiae* to two different environmental stressors, temperature and NaCl concentration, as well as the response of *E. coli* to various antibiotics. In contrast with the studies of Dai *et al.* [8–10], our yeast experiments used glucose as a nutrient source rather than sucrose. Glucose can be directly metabolized by *S. cerevisiae*, in contrast with the disaccharide sucrose, which must be hydrolysed outside the cytoplasm. However, most of the hydrolysis products of sucrose diffuse away before they can be taken back up into the cytoplasm, becoming public goods [37]. As a result, yeast cells fed with sucrose exhibit a classical Allee effect, with maximal growth at an intermediate population density [8]. Use of glucose rather than sucrose prevents population density from having a significant effect on survival, enabling temperature and salt concentration to serve as isolated stress variables in the studies described here.

We find that, while both high temperature and high NaCl concentration have a negative effect on the yeast doubling rate, the dynamics of the response to the two stressors is significantly different. In particular, the response to NaCl is gradual (figure 2*a*), while temperature stress induces a sharp drop reminiscent of a critical phase transition with a well-defined 'tipping point' (figure 2*b*). Accompanying this change in population growth, the fluctuations of the system are much higher in the vicinity of the tipping point as the temperature increases. In the *E. coli* experiments (figure 3), the responses to some of the antibiotics, such as ciprofloxacin, streptomycin and kanamycin, exhibited critical phase-transition-like behaviour, while others, such as chloramphenicol and spectinomycin, produced a gradual response without a well-defined tipping point. For other antibiotics, the response was less clearly defined.

In order to observe any critical behaviour, it is necessary to examine the system after it has stabilized. As figure 1 shows, the yeast had fluctuating growth behaviour in the first few days after exposure to both stressors. For most control parameter values, the behaviour stabilized after a few days. It is this stable 'long-term' behaviour that is averaged and shown in figure 2. Our results are consistent with those of Mensonides *et al.*, who demonstrated the initial slowing of *S. cerevisiae* population growth due to temperature for budding yeast, accompanied by changes in metabolism, by observing the populations for 6 h after introduction to various temperatures [30]. They 'observed a surprisingly "thin line" for cells between growing, surviving and dying, with regard to growth temperature' [30]. The sharp change between survival and death observed by Mensonides and colleagues occurred for a slightly higher temperature (between 42 and 43°C) than that observed in the present paper (39.5°C). This difference can be explained as the result of short- versus long-term behaviour; in all our cultures below 42°C, there was some measurable amount of growth in the first 24 h period, but the populations were unable to maintain consistent growth above 39.5°C, as can be seen in figure 1*a*. Mensonides *et al.* [30] suggested that 'it cannot be concluded if [the] effect on growth and viability is caused by the absolute temperature, or by the difference between initial and new temperature'. Notably, the initial temperature in their study was 28°C, while it was 30°C in our experiments. Dai *et al.* [10] also observed that *S. cerevisiae* populations lose stability at different rates and exhibit different levels of resilience, for different environmental stressors [10]. Specifically, for *S. cerevisiae* grown in the presence of sucrose, populations lost stability more rapidly (and early warning signals based on stability loss performed better) when stressed by increasing dilution factor, which serves as a proxy for death rate in *S. cerevisiae* populations grown on sucrose, than in response to simple nutrient depletion [10].

Why is the response to temperature stress much more abrupt than the salt stress response? Mensonides *et al.* proposed that the 'narrow temperature range' over which population decline is observed 'may be explained by assuming the rapid denaturation of one or a very limited number of enzymes which are essential to growth' [10]. This is consistent with the current understanding of heat shock response [38] and more specifically of protein denaturation [39] in *S. cerevisiae*. By contrast, these cells have a variety of responses that can modulate the response to salt stress. Changes in salt concentration can have multiple effects on yeast, including osmotic shock [40], direct toxicity from Na^+ ions [31] and changes in membrane potential [37]. *Saccharomyces cerevisiae* can respond to these diverse insults with a range of responses, such as activation of the high-osmolarity glycerol (HOG) pathway [39] and membrane depolarization accompanied by decreased permeability [32]. After observing an increase in abundance of a dozen plasma membrane proteins in response to mild salt stress, Szopinska *et al.* [33] proposed that protein internalization occurs rapidly after hyper-osmotic or ionic shock, enabling a cell to remain viable until a slower transcriptional response [34,35] can be activated. These various studies suggest a model in which *S. cerevisiae* is better able to modulate its response to salt stress than to heat stress, consistent with the more gradual response to salt stress we observe.

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

It is less clear, however, why *E. coli* responds so differently to antibiotics. Many of the antibiotics tested have the same mode of action but display completely different decay curves (see, for example, spectinomycin and kanamycin, which are both aminoglycosides). *A priori* one might expect that bacteriostatic antibiotics that only inhibit growth, as opposed to bactericidal antibiotics that kill *E. coli*, might have more gradual decay curves. However, there was a large variation in decay curves both within bacteriostatic (e.g. chloramphenicol and tetracycline) and bacteriocidal (e.g. rifampicin and ampicillin) antibiotics. It may be that the inherent pharmacokinetics of each antibiotic drives these differences in bacterial killing, and more comprehensive screening would be required to elucidate the exact underlying mechanisms. The differences observed in these bacterial decay curves might have important implications for the evolution of antibiotic resistance. It is possible that mutations could shift the shape of the decay curve; this could be tested in experimental evolution studies.

The system behaviour seen in the temperature stress experiments and some types of antibiotics (e.g. ciprofloxacin, streptomycin, kanamycin) match exceptionally well with critical phase transition behaviour, while the smooth transition in response to NaCl stress and other antibiotics (e.g. spectinomycin and chloramphenicol) show none of the signifying characteristics of a critical phase transition. These findings demonstrate that monitoring the dynamics of population decay can not only provide a warning sign of incipient collapse but could also be used in order to identify the type of stressor that is causing the destabilization. The difference in these two behaviours demonstrates that, without reference to the environmental stressors, simple observation of a population in decline may be insufficient to predict the course of the population's progress toward collapse. With detailed information about individual stressor effects, however, it may be possible to identify the driving stressor(s) of a population decline and determine whether it is approaching a critical collapse, or a gradual decline based on the transition behaviour. Understanding the type of stress response could be critical in designing appropriate intervention protocols.

The type of stress response may also play a role in the dynamics of population recovery. This has been observed by King *et al*. in a computational evolutionary model when a system is near the tipping point in a critical phase transition [26,27] and in the studies of population resilience in the presence of different stressors by Dai *et al*. [8–10]. As mentioned above, a key behaviour in systems with critical transition behaviour is critical slowing down. If intervention is taken to save a population from collapse, systems exhibiting critical slowing down will have significantly longer recovery times compared to systems undergoing gradual transitions. These systems will require more stringent monitoring to ensure that recovery continues through the long recovery period; it may also take far longer for significant signs of recovery to appear. Identification of transition type can be crucial in resource allocation to rescue the highest number of populations approaching collapse and prevent premature abandonment of a recovering population. Lastly, it is important to note that, due to cooperative effects, population size itself may modulate the dynamics of collapse and recovery [41] and should be taken into account when designing such interventions.

5. Conclusion

We observe population declines exhibiting characteristics of a critical phase transition in both yeast (*S. cerevisiae*) and bacteria (*E. coli*). In the case of yeast, phase transition behaviour is only observed when cultures are subjected to temperature stress, but not in the case of high salt concentration. In the latter case, the yeast populations exhibit a gradual decline without any sign of criticality. Similarly, in bacteria, some antibiotics produce population declines characteristic of a critical phase transition, while others produce more gradual declines. Taken together, the yeast and bacterial results suggest that phase-transition-like population dynamics may occur in response to a broad range of stressors in different organisms. Further, the disparity between phase-transition-like behaviour and more gradual population declines in both eukaryotic and prokaryotic cells for different stressors raises the possibility that divergent population dynamics may also occur for a range of organisms and stressors.

With increasing peril of current climate change and the approach of 'the Sixth Extinction', it is crucial to examine population dynamics for early indicators of collapse. Investigating population collapse through the lens of critical phase transitions can identify factors that drive population dynamics in uncontrolled environments and can facilitate the development of protocols for population rescue. This approach may also be used to pinpoint what stressor or combination of stressors is responsible for pushing a population to collapse, as well as to design intervention and rescue protocols.

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

Our results indicate that critical phase transition behaviour occurs in the presence of some stressors but not others. How and why this occurs is an avenue for future investigations, which could include, among other avenues, investigation of metabolic flux and imaging analysis of cells in the presence of different stressors over time. In the case of temperature and salt stress, recent experimental evolution studies in yeast [42–45], as well as the extensive literature on transcriptional response to stress in *S. cerevisiae* [35], could point the way toward identifying particular genetic and/or transcriptional determinants of population collapse dynamics. The repressors Nrg1 and Nrg2 have already been implicated in *S. cerevisiae* adaptation to salt stress [46]. Nearly two decades ago, a group of several hundred yeast genes were shown to undergo significant changes in expression in response to environmental stress [47,48]. The genetic pathways that mediate resilience are likely to be complicated, multilayered and historically contingent. Berry *et al*. [49] found that the changes in gene expression that confer tolerance to H_2O_2 stress in yeast differ depending on the prior mild insult the cells received (salt, heat shock or dithiothreitol (DTT)). While such complex interactions make investigating such problems daunting, they also offer the tantalizing possibility that mapping gene activation can reveal a cell's environmental history, much as a human patient's complement of antibodies can show their history of disease exposure. Such studies could ultimately contribute not only to the development of population rescue protocols but may also facilitate the genetic engineering of species more resilient in the face of environmental stress.

Data accessibility. Data are available at the Dryad Digital Repository, https://doi.org/10.5061/dryad.5hqbzkh2k [36].

Authors' contributions. S.W.O. carried out yeast experiments and data analysis, and participated in drafting the manuscript; D.M.K. designed the yeast study, performed preliminary yeast experiments and participated in drafting the manuscript; D.F., C.N. and S.P. performed yeast experiments and data analysis; H.H. performed *E. coli* experiments and data analysis; R.F.I. designed the *E. coli* experiments and participated in drafting the manuscript; W.O. designed the yeast study and participated in drafting the manuscript; S.B. designed the study, performed data analysis and participated in drafting the manuscript. All authors gave approval for publication.

Competing interests. The authors have no competing interests to declare.

Funding. S.W.O., D.M.K., S.B. and W.O. were supported by a University of Missouri Interdisciplinary Intercampus Research Grant.

References

1. Kolbert E. 2014 *The sixth extinction: an unnatural history*. New York, NY: Henry Holt and Company.
2. Barnosky AD *et al.* 2011 Has the Earth's sixth mass extinction already arrived? *Nature* **471**, 51–57. (doi:10.1038/nature09678)
3. Ceballos G, Ehrlich PR, Dirzo R. 2017 Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proc. Natl Acad. Sci. USA* **114**, E6089–E6096. (doi:10.1073/pnas.1704949114)
4. Ceballos G, Ehrlich PR, Barnosky AD, García A, Pringle RM, Palmer TM. 2017 Accelerated modern human-induced species losses: entering the sixth mass extinction. *Sci. Adv.* **1**, e1400253. (doi:10.1126/sciadv.1400253)
5. Scheffer M *et al.* 2012 Anticipating critical transitions. *Science* **338**, 344–348. (doi:10.1126/science.1225244)
6. Scheffer M *et al.* 2009 Early-warning signals for critical transitions. *Nature* **461**, 53–59. (doi:10.1038/nature08227)
7. Dakos V, Matthews B, Hendry AP, Levine J, Loeuille N, Norberg J, Nosil P, Scheffer M, De Meester L. 2019 Ecosystem tipping points in an evolving world. *Nat. Ecol. Evol.* **3**, 355–362. (doi:10.1038/s41559-019-0797-2)
8. Dai L, Vorselen D, Korolev KS, Gore J. 2012 Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science* **336**, 1175–1177. (doi:10.1126/science.1219805)
9. Dai L, Korolev KS, Gore J. 2013 Slower recovery in space before collapse of connected populations. *Nature* **496**, 355–358. (doi:10.1038/nature12071)
10. Dai L, Korolev KS, Gore J. 2015 Relation between stability and resilience determines the performance of early warning signals under different environmental drivers. *Proc. Natl Acad. Sci. USA* **112**, 10 056–10 061. (doi:10.1073/pnas.1418415112)
11. Carpenter SR. 2013 Spatial signatures of resilience. *Nature* **496**, 308–309. (doi:10.1038/nature12092)
12. Rozek JC, Camp RJ, Reed JM. 2017 No evidence of critical slowing down in two endangered Hawaiian honeycreepers. *PLoS ONE* **12**, e0187518. (doi:10.1371/journal.pone.0187518)
13. Lever JJ, van Nes EH, Scheffer M, Bascompte J. 2014 The sudden collapse of pollinator communities. *Ecol. Lett.* **17**, 350–359. (doi:10.1111/ele.12236)
14. Dainat B, Evans JD, Chen YP, Gauthier L, Neumann P. 2012 Predictive markers of honey bee colony collapse. *PLoS ONE* **7**, e32151. (doi:10.1371/journal.pone.0032151)
15. Steinhauer N, Kulhanek K, Antúnez K, Human H, Chantawannakul P, Chauzat MP, van Engelsdorp D. 2018 Drivers of colony losses. *Curr. Opin. Insect. Sci.* **26**, 142–148. (doi:10.1016/j.cois.2018.02.004)
16. Paxton EH, Camp RJ, Gorresen PM, Crampton LH, Leonard Jr DL, VanderWerf EA. 2016 Collapsing avian community on a Hawaiian island. *Sci. Adv.* **2**, e1600029. (doi:10.1126/sciadv.1600029)
17. Hallmann CA *et al.* 2017 More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLoS ONE* **12**, e0185809. (doi:10.1371/journal.pone.0185809)
18. Lister BC, Garcia A. 2018 Climate-driven declines in arthropod abundance restructure a rainforest food web. *Proc. Natl Acad. Sci. USA* **115**, E10397–E10406. (doi:10.1073/pnas.1722477115)
19. Simmons BI *et al.* 2019 Worldwide insect declines: an important message, but interpret with caution. *Ecol. Evol.* **9**, 3678–3680. (doi:10.1002/ece3.5153)
20. Sánchez-Bayo F, Wyckhuys KAG. 2019 Worldwide decline of the entomofauna: a review of its drivers. *Biol. Cons.* **232**, 8–27. (doi:10.1016/j.biocon.2019.01.020)
21. Hinrichsen H. 2000 Nonequilibrium critical phenomena and phase transitions into absorbing states. *Adv. Phys.* **49**, 815–958. (doi:10.1080/00018730050198152)

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

22. Henkel M, Hinrichsen H, Lübeck L. 2008 *Non-equilibrium phase transitions: volume 1: absorbing phase transitions*. Amsterdam, The Netherlands: Springer Science and Business Media B.V.
23. Dees ND, Bahar S. 2010 Mutation size optimizes speciation in an evolutionary model. *PLoS ONE* **5**, e11952. (doi:10.1371/journal.pone.0011952)
24. Scott AD, King DM, Marić N, Bahar S. 2013 Clustering and phase transitions on a neutral landscape. *Europhys. Lett.* **102**, 68003. (doi:10.1209/0295-5075/102/68003)
25. Scott AD. 2014 Speciation dynamics of an agent-based evolution model in phenotype space. PhD dissertation, University of Missouri at St. Louis/Missouri University of Science & Technology.
26. King DM. 2015 Evolutionary dynamics of speciation and extinction. PhD dissertation, University of Missouri at St. Louis/Missouri University of Science & Technology.
27. King DM, Scott AD, Bahar S. 2017 Multiple phase transitions in an agent-based evolutionary model with neutral fitness. *R. Soc. Open Sci.* **4**, 170005. (doi:10.1098/rsos.170005)
28. Ulbricht RJ, Olivas WM. 2008 Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA* **14**, 246–262. (doi:10.1261/rna.847408)
29. Caspeta L, Kielsen J. 2015 Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *mBio* **6**, e00431. (doi:10.1128/mBio.00431-15)
30. Mensonides FIC, Schuurmans JM, Teixeira de Mattos MJ, Hellingwerf KJ, Brul S. 2002 The metabolic response of *Saccharomyces cerevisiae* to continuous heat stress. *Mol. Biol. Rep.* **29**, 103–106. (doi:10.1023/A:1020392805411)
31. Murguía JR, Bellés JM, Serrano R. 1996 The yeast HAL2 nucleotidase is an *in vivo* target of salt toxicity. *J. Biol. Chem.* **271**, 29 029–29 033. (doi:10.1074/jbc.271.46.29029)
32. Capusoni C, Arioli S, Donzella S, Guidi B, Serra I, Compagno C. 2019 Hyper-osmotic stress elicits membrane depolarization and decreased permeability in halotolerant marine *Debaryomyces hansenii* strains and in *Saccharomyces cerevisiae*. *Front. Microbiol.* **10**, 64. (doi:10.3389/fmicb.2019.00064)
33. Szopinska A, Degand H, Hochstenbach JF, Nader J, Morsomme P. 2011 Rapid response of the yeast plasma membrane proteome to salt stress. *Mol. Cell. Proteomics* **10**, M111.009589. (doi:10.1074/mcp.M111.009589)
34. Melamed D, Pnueli L, Arava Y. 2008 Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *RNA* **14**, 1337–1351. (doi:10.1261/rna.864908)
35. Taymaz-Nikerel H, Cankorur-Cetinkaya A, Kirdar B. 2016 Genome-wide transcriptional response of *Saccharomyces cerevisiae* to stress-induced perturbations. *Front. Bioeng. Biotechnol.* **4**, 17.
36. Ordway SW, King DM, Friend D, Noto C, Phu S, Huelskamp H, Inglis RF, Olivas W, Bahar S. 2020 Data from: Phase transitions in yeast and bacterial populations. Dryad Digital Repository. (doi:10.5061/dryad.5hqbzkh2k)
37. Gore J, Youk H, van Oudenaarden A. 2009 Snowdrift game dynamics and facultative cheating in yeast. *Nature* **459**, 253–256. (doi:10.1038/nature07921)
38. Leuenberger P, Ganscha S, Kahraman A, Cappelletti V, Boersema PJ, von Mering C, Claassen M, Picotti P. 2017 Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825. (doi:10.1126/science.aai7825)
39. Verghese J, Abrams J, Wang Y, Morano KA. 2012 Biology of the heat shock response and protein chaperones: budding yeast (*Saccharomyces cerevisiae*) as a model system. *Microbiol. Mol. Biol. Rev.* **76**, 115–158. (doi:10.1128/MMBR.05018-11)
40. Hohmann S. 2002 Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol. Mol. Biol. Rev.* **66**, 300–372. (doi:10.1128/MMBR.66.2.300-372.2002)
41. Laman Trip DS, Youk H. 2019 Cells reshape habitability of temperature by secreting antioxidants to help each other replicate and avoid population extinction. 2019 *BioRxiv* https://doi.org/10.1101/726463.
42. Ratcliff WC, Denison RF, Borrello M, Travisano M. 2012 Experimental evolution of multicellularity. *Proc. Natl Acad. Sci. USA* **109**, 1595–1600. (doi:10.1073/pnas.1115323109)
43. Ratcliff WC, Pentz JT, Travisano M. 2013 Tempo and mode of multicellular adaptation in experimentally evolved *Saccharomyces cerevisiae*. *Evolution* **67**, 1573–1581. (doi:10.1111/evo.12101)
44. Ratcliff WC, Fankhauser JD, Rogers DW, Greig D, Travisano M. 2015 Origins of multicellular evolvability in snowflake yeast. *Nat. Commun.* **6**, 6102. (doi:10.1038/ncomms7102)
45. Nguyen Ba AN, Cvijović I, Rojas Echenique JI, Lawrence KR, Rego-Costa A, Liu X, Levy SF, Desai MM. 2019 High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature* **575**, 494–499. (doi:10.1038/s41586-019-1749-3)
46. Vyas VK, Berkey CD, Miyao T, Carlson M. 2005 Repressors Nrg1 and Nrg2 regulate a set of stress-responsive genes in *Saccharomyces cerevisiae*. *Eukaryot. Cell* **4**, 1882–1891. (doi:10.1128/EC.4.11.1882-1891.2005)
47. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D, Brown PO. 2000 Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257. (doi:10.1091/mbc.11.12.4241)
48. Causton HC *et al.* 2001 Remodeling of yeast genome expression in response to environmental changes. *Mol. Biol. Cell* **12**, 323–337. (doi:10.1091/mbc.12.2.323)
49. Berry DB, Guan Q, Hose J, Haroon S, Gebbia M, Heisler LE, Nislow C, Giaever G, Gasch AP. 2011 Multiple means to the same end: the genetic basis of acquired stress resistance in yeast. *PLoS Genet.* **7**, e1002353. (doi:10.1371/journal.pgen.1002353)

Figure B.1 Ordway *et al.* as Published in Royal Society Open Science (cont.)

BIBLIOGRAPHY

Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues(IJCSI)*, *9*.

Alonso, D., Etienne, R., & McKane, A. (2006). The merits of neutral theory. *Trends in Ecology & Evolution*, *21*, 451–457. https://doi.org/10.1016/j.tree.2006.03.019

Bellman, R. E., & Dreyfus, S. E. (1962). *Applied dynamic programming*. Princeton University Press; JSTOR. http://www.jstor.org/stable/j.ctt183pq99

Berry, D. B., Guan, Q., Hose, J., Haroon, S., Gebbia, M., Heisler, L. E., Nislow, C., Giaever, G., & Gasch, A. P. (2011). Multiple means to the same end: The genetic basis of acquired stress resistance in yeast. *PLOS Genetics*, *7*(11), e1002353. https://doi.org/10.1371/journal.pgen.1002353

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer Science+Business Media.

Bridges, M., Heron, E. A., O'Dushlaine, C., Segurado, R., The International Schizophrenia Consortium (ISC), Morris, D., Corvin, A., Gill, M., & Pinto, C. (2011). Genetic classification of populations using supervised learning. *PLOS ONE*, *6*(5), e14802. https://doi.org/10.1371/journal.pone.0014802

Capusoni, C., Arioli, S., Donzella, S., Guidi, B., Serra, I., & Compagno, C. (2019). Hyper-osmotic stress elicits membrane depolarization and decreased permeability in halotolerant marine debaryomyces hansenii strains and in Saccharomyces cerevisiae. *Frontiers in Microbiology*, *10*, 64. https://doi.org/10.3389/fmicb.2019.00064

Carpenter, S. R. (2013). Spatial signatures of resilience. *Nature*, *496*(7445), 308–309. https://doi.org/10.1038/nature12092

Caspeta, L., & Nielsen, J. (2015). Thermotolerant yeast strains adapted by laboratory evolution show trade-off at ancestral temperatures and preadaptation to other stresses. *MBio*, *6*(4), e00431-15. https://doi.org/10.1128/mBio.00431-15

Causton, H. C., Ren, B., Koh, S. S., Harbison, C. T., Kanin, E., Jennings, E. G., Lee, T. I., True, H. L., Lander, E. S., & Young, R. A. (2001). Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell*, *12*(2), 323–337. https://doi.org/10.1091/mbc.12.2.323

Çelik, Ö. (2018). *A research on machine learning methods and its applications*. https://doi.org/10.31681/jetol.457046

Collen, B., Loh, J., Whitmee, S., McRae, L., Amin, R., & Baillie, J. (2008). Monitoring change in vertebrate abundance: The Living Planet Index. *Conservation Biology : The Journal of the Society for Conservation Biology*, *23*, 317–327. https://doi.org/10.1111/j.1523-1739.2008.01117.x

Collen, B., McRae, L., Deinet, S., De Palma, A., Carranza, T., Cooper, N., Loh, J., & Baillie, J. E. M. (2011). Predicting how populations decline to extinction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1577), 2577–2586. https://doi.org/10.1098/rstb.2011.0015

Dai, L., Korolev, K. S., & Gore, J. (2013). Slower recovery in space before collapse of connected populations. *Nature*, *496*(7445), 355–358. https://doi.org/10.1038/nature12071

Dai, L., Korolev, K. S., & Gore, J. (2015). Relation between stability and resilience determines the performance of early warning signals under different environmental drivers. *Proceedings of the National Academy of Sciences*, *112*(32), 10056–10061. https://doi.org/10.1073/pnas.1418415112

Dai, L., Vorselen, D., Korolev, K. S., & Gore, J. (2012). Generic indicators for loss of resilience before a tipping point leading to population collapse. *Science (New York, N.Y.)*, *336*(6085), 1175–1177. https://doi.org/10.1126/science.1219805

Dakos, V., Matthews, B., Hendry, A. P., Levine, J., Loeuille, N., Norberg, J., Nosil, P., Scheffer, M., & De Meester, L. (2019). Ecosystem tipping points in an evolving world. *Nature Ecology & Evolution*, *3*(3), 355–362. https://doi.org/10.1038/s41559-019-0797-2

Daskalova, G. N., Myers-Smith, I. H., & Godlee, J. L. (2020). Rare and common vertebrates span a wide spectrum of population trends. *Nature Communications*, *11*(1), 4394. https://doi.org/10.1038/s41467-020-17779-0

de Aguiar, M. A. M., Baranger, M., Baptestini, E. M., Kaufman, L., & Bar-Yam, Y. (2009). Global patterns of speciation and diversity. *Nature*, *460*(7253), 384–387. https://doi.org/10.1038/nature08168

Debeljak, M., Džeroski, S., Jerina, K., Kobler, A., & Adamič, M. (2001). Habitat suitability modelling for red deer (Cervus elaphus L.) in South-central Slovenia with classification trees. *Ecological Modelling*, *138*, 321–330. https://doi.org/10.1016/S0304-3800(00)00411-7

Dees, N. D., & Bahar, S. (2010). Mutation size optimizes speciation in an evolutionary model. *PLOS ONE*, *5*(8), e11952. https://doi.org/10.1371/journal.pone.0011952

Derrida, B., & Peliti, L. (1991). Evolution in a flat fitness landscape. *Bulletin of Mathematical Biology*, *53*, 355–382. https://doi.org/10.1007/BF02460723

Garcia, A. (2000). *Numberical methods for physics* (2nd ed). Prentic Hall Inc.,.

Gasch, A. P., Spellman, P. T., Kao, C. M., Carmel-Harel, O., Eisen, M. B., Storz, G., Botstein, D., & Brown, P. O. (2000). Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell*, *11*(12), 4241–4257. https://doi.org/10.1091/mbc.11.12.4241

Gore, J., Youk, H., & van Oudenaarden, A. (2009). Snowdrift game dynamics and facultative cheating in yeast. *Nature*, *459*(7244), 253–256. https://doi.org/10.1038/nature07921

Grassberger, P. (1982). On phase transitions in Schlögl's second model. *Zeitschrift Für Physik B Condensed Matter*, *47*(4), 365–374. https://doi.org/10.1007/BF01313803

Grassberger, P. (1989). Directed percolation in 2+1 dimensions. *Journal of Physics A: Mathematical and General*, *22*(17), 3673–3679. https://doi.org/10.1088/0305-4470/22/17/032

Grassberger, P. & Yi-Cheng Zhang. (1996). "Self-organized" formulation of standard percolation phenomena. *Dynamics of Complex Systems*, *224*(1), 169–179. https://doi.org/10.1016/0378-4371(95)00321-5

Grinstein, G., & Luther, A. (1976). Application of the renormalization group to phase transitions in disordered systems. *Physical Review B*, *13*(3), 1329–1343. https://doi.org/10.1103/PhysRevB.13.1329

Henkel, M., Hinrichsen, H., & Lübeck, S. (2008). *Non-equilibrium phase transitions: Volume 1: absorbing phase transitions*. Springer Netherlands. https://www.springer.com/us/book/9781402087646

Hinrichsen, H. (2000a). On possible experimental realizations of directed percolation. *Braz. J. Phys*, *30*(1), 69–82. https://doi.org/10.1590/S0103-97332000000100007

Hinrichsen, H. (2000b). Nonequilibrium critical phenomena and phase transitions into absorbing states. *Advances in Physics*, *49*(7), 815–958. https://doi.org/10.1080/00018730050198152

Hinrichsen, H. (2006). *Non-equilibrium phase transitions* [Lecture notes].

Hohmann, S. (2002). Osmotic stress signaling and osmoadaptation in yeasts. *Microbiology and Molecular Biology Reviews*, *66*(2), 300. https://doi.org/10.1128/MMBR.66.2.300-372.2002

Hooyberghs, J., Igloi, F., & Vanderzande, C. (2003). Strong disorder fixed point in absorbing-state phase transitions. *Physical Review Letters*, *90*, 100601. https://doi.org/10.1103/PhysRevLett.90.100601

Houchmandzadeh, B. (2002). Clustering of diffusing organisms. *Phys. Rev. E*, *66*(5), 052902. https://doi.org/10.1103/PhysRevE.66.052902

Houchmandzadeh, B. (2008). Neutral clustering in a simple experimental ecological community. *Physical Review Letters*, *101*, 078103. https://doi.org/10.1103/PhysRevLett.101.078103

Houchmandzadeh, B., & Vallade, M. (2003). Clustering in neutral ecology. *Physical Review E*, *68*(6), 061912. https://doi.org/10.1103/PhysRevE.68.061912

Janssen, H. K. (1981). On the nonequilibrium phase transition in reaction-diffusion systems with an absorbing stationary state. *Zeitschrift Für Physik B Condensed Matter*, *42*(2), 151–154. https://doi.org/10.1007/BF01319549

King, D. M. (2015). *Evolutionary dynamics of speciation and extinction* [Doctoral Dissertation]. University of Missouri - Saint Louis, Missouri University of Science and Technology.

King, D. M., Scott, A. D., & Bahar, S. (2017). Multiple phase transitions in an agent-based evolutionary model with neutral fitness. *Royal Society Open Science*, *4*(4), 170005. https://doi.org/10.1098/rsos.170005

Kolbert, E. (2014). *The sixth extinction: An unnatural history* (1st ed.). Henry Holt and Company.

Krzanowski, W., & Hand, D. (2009). *ROC curves for continuous data*. CRC Press.

Lawson, D., & Jensen, H. (2007). Neutral evolution in a biological population as diffusion in phenotype space: Reproduction with local mutation but without selection. *Physical Review Letters*, *98*, 098102. https://doi.org/10.1103/PhysRevLett.98.098102

Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P. J., von Mering, C., Claassen, M., & Picotti, P. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*, *355*(6327), eaai7825. https://doi.org/10.1126/science.aai7825

Living Planet Index. (n.d.). Retrieved April 5, 2021, from http://livingplanetindex.org/search

López, C., Ramos, F., & Hernández-García, E. (2007). An absorbing phase transition from a structured active particle phase. *Journal of Physics: Condensed Matter*, *19*(6), 065133. https://doi.org/10.1088/0953-8984/19/6/065133

Marro, J., & Dickman, R. (1999). *Nonequilibrium phase transitions in lattice models*. Cambridge University Press; Cambridge Core. https://doi.org/10.1017/CBO9780511524288

Melamed, D., Pnueli, L., & Arava, Y. (2008). Yeast translational response to high salinity: Global analysis reveals regulation at multiple levels. *RNA*, *14*(7), 1337–1351. https://doi.org/10.1261/rna.864908

Mensonides, F. I. C., Schuurmans, J. M., de Mattos, M. J. T., Hellingwerf, K. J., & Brul, S. (2002). The metabolic response of Saccharomyces cerevisiae to continuous heat stress. *Molecular Biology Reports*, *29*(1), 103–106. https://doi.org/10.1023/A:1020392805411

Meyer, M., Havlin, S., & Bunde, A. (1996). Clustering of independently diffusing individuals by birth and death processes. *Phys. Rev. E*, *54*(5), 5567–5570. https://doi.org/10.1103/PhysRevE.54.5567

Mitchell, T. M. (1997). *Machine learning* (1st ed.). McGraw-Hill, Inc.

Murguía, J. R., Bellés, J. M., & Serrano, R. (1996). The yeast HAL2 nucleotidase Is an in vivo target of salt toxicity. *Journal of Biological Chemistry*, *271*(46), 29029–29033. https://doi.org/10.1074/jbc.271.46.29029

Nachtrab, S. (2011). *Percolation and elasticity of networks: From cellular structures to fibre bundles* [Doctoral Dissertation]. University of Erlangen-Nuremberg.

Ng, A. (2020). *CS229 lecture notes: Supervised learning*. http://cs229.stanford.edu/notes2020fall/notes2020fall/cs229-notes1.pdf

Nguyen Ba, A. N., Cvijović, I., Rojas Echenique, J. I., Lawrence, K. R., Rego-Costa, A., Liu, X., Levy, S. F., & Desai, M. M. (2019). High-resolution lineage tracking reveals travelling wave of adaptation in laboratory yeast. *Nature*, *575*(7783), 494–499. https://doi.org/10.1038/s41586-019-1749-3

NOAA. (n.d.). *National Oceanic and Atmospheric Administration: Climate Data Online*. Retrieved May 5, 2021, from https://www.ncdc.noaa.gov/cdo-web/api/v2/data

Ódor, G. (2004). Universality classes in nonequilibrium lattice systems. *Reviews of Modern Physics*, *76*(3), 663–724. https://doi.org/10.1103/RevModPhys.76.663

Okasha, S. (2009). *Evolution and the levels of selection*. Oxford University Press.

Ordway, S. W., King, D. M., Friend, D., Noto, C., Phu, S., Huelskamp, H., Inglis, R. F., Olivas, W., & Bahar, S. (2020). Phase transition behaviour in yeast and bacterial populations under stress. *Royal Society Open Science*, *7*(7), 192211–192211. PubMed. https://doi.org/10.1098/rsos.192211

Pananos, A. D., Bury, T. M., Wang, C., Schonfeld, J., Mohanty, S. P., Nyhan, B., Salathé, M., & Bauch, C. T. (2017). Critical dynamics in population vaccinating behavior. *Proceedings of the National Academy of Sciences*, *114*(52), 13762. https://doi.org/10.1073/pnas.1704093114

Ratcliff, W. C., Denison, R. F., Borrello, M., & Travisano, M. (2012). Experimental evolution of multicellularity. *Proceedings of the National Academy of Sciences*, *109*(5), 1595. https://doi.org/10.1073/pnas.1115323109

Ratcliff, W. C., Fankhauser, J. D., Rogers, D. W., Greig, D., & Travisano, M. (2015). Origins of multicellular evolvability in snowflake yeast. *Nature Communications*, *6*(1), 6102. https://doi.org/10.1038/ncomms7102

Ratcliff, W. C., Pentz, J. T., & Travisano, M. (2013). Tempo And mode of multicellular adaptation in experimentally evolved Saccharomyces cerevisiae. *Evolution*, *67*(6), 1573–1581. https://doi.org/10.1111/evo.12101

Roelofs, R. (2019). *Measuring generalization and overfitting in machine learning* [Doctoral Dissertation]. University of California, Berkeley.

Rozek, J. C., Camp, R. J., & Reed, J. M. (2017). No evidence of critical slowing down in two endangered Hawaiian honeycreepers. *PLOS ONE*, *12*(11), e0187518. https://doi.org/10.1371/journal.pone.0187518

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53–59. https://doi.org/10.1038/nature08227

Scheffer, M., Carpenter, S. R., Lenton, T. M., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I. A., Levin, S. A., van Nes, E. H., Pascual, M., & Vandermeer, J. (2012). Anticipating critical transitions. *Science*, *338*(6105), 344. https://doi.org/10.1126/science.1225244

Scott, A. D. (2014). *Speciation dynamics of and agent-based evolution model in phenotype space* [Doctoral Dissertation]. University of Missouri - Saint Louis, Missouri University of Science and Technology.

Scott, A. D., King, D. M., Marić, N., & Bahar, S. (2013). Clustering and phase transitions on a neutral landscape. *Europhysics Letters*, *102*(6), 68003. https://doi.org/10.1209/0295-5075/102/68003

Stankovski, V., Debeljak, M., Bratko, I., & Adami\vc, M. (1998). Modelling the population dynamics of red deer (Cervus elaphus L.) with regard to forest development. *Ecological Modelling*, *108*, 145–153.

Szopinska, A., Degand, H., Hochstenbach, J.-F., Nader, J., & Morsomme, P. (2011). Rapid response of the yeast plasma membrane proteome to salt stress. *Molecular & Cellular Proteomics*, *10*(11), M111.009589. https://doi.org/10.1074/mcp.M111.009589

Takeuchi, K. A., Kuroda, M., Chaté, H., & Sano, M. (2009). Experimental realization of directed percolation criticality in turbulent liquid crystals. *Physical Review E*, *80*(5), 051116. https://doi.org/10.1103/PhysRevE.80.051116

Taymaz-Nikerel, H., Cankorur-Cetinkaya, A., & Kirdar, B. (2016). Genome-wide transcriptional response of Saccharomyces cerevisiae to stress-induced perturbations. *Frontiers in Bioengineering and Biotechnology*, *4*, 17. https://doi.org/10.3389/fbioe.2016.00017

Ulbricht, R. J., & Olivas, W. M. (2008). Puf1p acts in combination with other yeast Puf proteins to control mRNA stability. *RNA*, *14*(2), 246–262. https://doi.org/10.1261/rna.847408

Verghese, J., Abrams, J., Wang, Y., & Morano, K. A. (2012). Biology of the heat shock response and protein chaperones: Budding yeast Saccharomyces cerevisiae as a model system. *Microbiology and Molecular Biology Reviews*, *76*(2), 115. https://doi.org/10.1128/MMBR.05018-11

Vojta, T. (2003). Disorder-induced rounding of certain quantum phase transitions. *Physical Review Letters*, *90*(10), 107202. https://doi.org/10.1103/PhysRevLett.90.107202

Vyas, V. K., Berkey, C. D., Miyao, T., & Carlson, M. (2005). Repressors Nrg1 and Nrg2 regulate a set of stress-responsive genes in Saccharomyces cerevisiae. *Eukaryotic Cell*, *4*(11), 1882. https://doi.org/10.1128/EC.4.11.1882-1891.2005

Wang, J., Zhou, Z., Liu, Q., Garoni, T. M., & Deng, Y. (2013). High-precision Monte Carlo study of directed percolation in (d+1) dimensions. *Physical Review E*, *88*(4), 042102. https://doi.org/10.1103/PhysRevE.88.042102

Wang, Y., Zhang, Z., Feng, L., Du, Q., & Runge, T. (2020). Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous united states. *Remote Sensing*, *12*(8). https://doi.org/10.3390/rs12081232

Weinberger, E. (1987). A model of natural selection that exhibits a dynamic phase transition. *Journal of Statistical Physics*, *49*(5), 1011–1028. https://doi.org/10.1007/BF01017557

WWF. (2020). *Living planet report 2020*.

Young, W. R., Roberts, A. J., & Stuhne, G. (2001). Reproductive pair correlations and the clustering of organisms. *Nature*, *412*(6844), 328–331. https://doi.org/10.1038/35085561

Zhang, Y.-C., Serva, M., & Polikarpov, M. (1990). Diffusion reproduction processes. *Journal of Statistical Physics*, *58*(5), 849–861. https://doi.org/10.1007/BF01026554

VITA

Stephen Walter Ordway was born in Saint Louis, MO to his loving parents, Fred Ordway and Nancy Ellis-Ordway. He attended high school in Jefferson City, MO where he received his high school diploma in 2009. Stephen attended the University of Missouri – St. Louis for his undergraduate degree, which included enrollment in the Pierre Laclede Honors College and a year abroad at the University of Birmingham, UK for the academic year of 2011-2012. He received his Bachelor of Sciences degree in Physics with a minor in Mathematics in 2013. He remained at the University of Missouri – St. Louis for graduate studies. Stephen earned the degree of Master of Science in Physics in August 2015. Stephen married Dawn King, Doctor of Physics, in the summer of 2016, and took on a parental role for Blake Warkenton. He received his Doctorate in Philosophy degree in Physics from the cooperative doctoral program for physics at the University of Missouri – St. Louis and the Missouri University of Science and Technology in December 2021.