CHARACTERIZATION OF TRANSPORT EQUATIONS WITH FORENSIC APPLICATIONS (NUCLEAR AND SOCIAL)

by

Nickolas A. Duncan

A thesis submitted to the Faculty and the Board of Trustees of the Colorado School of Mines in partial fulfillment of the requirements for the degree of Doctor of Philosophy (Nuclear Science and Engineering).

Golden, Colorado
Date: July 17, 2020

Signed: _____

Nickolas A. Duncan

Signed: _____

Dr. Mark Deinert
Thesis Advisor

Golden, Colorado
Date: July 17, 2020

Signed: _____

Dr. Mark Jensen
Professor and Director
Nuclear Science and Engineering Program

ABSTRACT

Convection and diffusion processes are used to understand transport in a wide range of contexts including the spread of diseases, the adoption of ideas within populations and the classical applications to heat and mass transfer. While much attention is typically paid to formulating the appropriate equations to accurately capture the underlaying processes, the parameters that go into these mathematical models are equally important and receive far less attention. The SARS-CoV-2 emerged in late 2019 and caused a worldwide pandemic. Epidemiological models are playing a key role in guiding public health interventions. The SIR model (susceptible, infected, recovered) is used to predict the number of infections over time. Their ability to accurately predict the number of people who will become infected depends on input parameters that are poorly understood. Here the effects of uncertainty on predicted outcomes are explored. The diffusion of ideas on social media is also studied in this context. How ideas propagate can affect societal trends, norms, behaviors, influence markets and the outcomes of elections. The SIR model is again used, but here in combination with sentiment analysis to understand tweet behavior. Different sentiment messages spread at different rates through social media.

Parameter estimation in the classical domain is conducted here to understand subsurface transport models that are used for post detonation nuclear forensics. Subsurface gas transport depends on accurately estimating the depth of the underground explosion as well as the geology that surrounds the explosion. The site of the explosions are likely to be denied access sites and parameter estimations must be done remotely. The depth at which a test occurs is known to be a critical parameter, affecting not only the migration time for gases to reach the surface but also their subsequent isotopic ratios. Bayesian data synthesis can improve depth of burst estimates by considering local topology, geology, the presence of surface deformation, yield, and a safety factor (for US tests). Here a method is developed to characterize fracture width, spacing, tortuosity, permeability and porosity at a denied access site. Fractures are treated as fractals with their respective fractal dimensions determined using surface images. The input parameters were applied to a subsurface gas transport model for six underground nuclear explosions conducted by the Democratic People's Republic of Korea (DPRK).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
INTRODUCTION

Transport models are used to predict the movement for a variety of phenomena. The context for this study includes understanding how diseases move through a population, how ideas move through a population and how radioactive xenon gas moves through the geology following an underground nuclear explosion. Transport models are used by leaders to provide information for which they can make decisions for actions or responses to a situation.

The majority of the attention for transport models is placed on the development of the model itself. The developer needs to ensure the underlying physics and processes the model uses are correct. For subsurface gas transport, this includes understanding the convection-diffusion equations and knowing the numerical discretization scheme. The models need to be verified and validated using analytical solutions and compared to problems with known solutions.

However, little attention is placed on the input parameters for the transport models. Regardless of the complexity and accuracy of a transport model, if the input parameters are inaccurate and don't represent the system in which that it is attempting to apply them to, the corresponding output of the model becomes worthless because it will be inaccurate. Therefore, when utilizing a transport model, a deliberate effort should be made to understand the appropriate input parameters and the uncertainty for each.

Understanding the uncertainty in each input parameter allows you to better understand the results from the transport model. The range of the transport models results will be directly correlated to the uncertainty in the input parameters. This effort focuses on quantifying and estimating input parameters for three different topics. The topics include the spread of a disease and ideas through a population, and subsurface xenon transport. For subsurface xenon transport, the context is for a denied access site. A denied access site makes it difficult to estimate the input parameters and thus the expected uncertainty for each is expected to be large.

For the spread of a disease through a population, the three-compartment model known as the SIR (Susceptible, Infected, and Recovered) model is used. For the SIR model, the infection and recovery rates are required. We conduct a case study using the SAR-CoV-2 pandemic that started in late 2019 and early 2020. A Monte Carlo simulation is conducted using a range of distributions for the infection and recovery rates based upon medical studies conducted during early stages of the epidemic.

For the spread of ideas on social media, the SIR model is used to model the behavior of messages being adopted by users. The infection and recovery rates are estimated using actual data from twitter. The infection and recovery rates are adjusted until a best fit is found by minimizing the sum of the squared errors between the predicted and actual data from twitter.

The input parameters for subsurface xenon transport will be applied to a denied access site. The key input parameter for subsurface gas transport is the depth of burst for the underground nuclear explosion. The five geological parameters that surround the underground nuclear explosion are fracture spacing, fracture width, permeability, porosity and tortuosity. All these parameters are estimated using remote sensing data to include seismology data, imagery, and geological maps.

The objective is to quantify each input parameter into a distribution in order to conduct Monte Carlo simulations. The Monte Carlo simulations provide a method to understand the range of possible outcomes for an event given a range of input parameters. This provides an understanding of what leaders can do to influence a response.

CHAPTER TWO

PARAMETER UNCERTAINTY AND ITS EFFECT ON PREDICTING OUTCOMES IN THE SARS-COV-2 PANDEMIC

SARS-CoV-2 emerged in late 2019 as a zoonotic infection of humans, and proceeded to cause a worldwide pandemic of historic magnitude. Epidemiological models are playing a key role in guiding public health interventions. However, their ability to accurately predict the number of people who will become infected depends on input parameters that are poorly understood. Here we explore the effects of uncertainty on a standard epidemiological model using valid inputs from within accepted ranges. The resulting predictions for peak infection and hospitalization rates vary by more than an order of magnitude. While the range is large, low impact epidemics are a small fraction of the possible outcomes. Importantly, social distancing is seen to do more than flatten the epidemic curve, it can reduce the total number infected before herd immunity is achieved. This study shows that an understanding of the virus and disease are the main hinderances to accurate modeling.

2.1 Introduction to SARS-CoV-2 Pandemic

The 20th century saw the emergence of the Spanish Flu, MERS, HIV, Ebola, Swine Flu, avian flu, Lasa, Marburg fever and many more. In 2002 SARS emerged and the latest global threat is now from SARS-CoV-2 [1]. Epidemiological models are playing an important role in guiding the response to the global pandemic. However, important input parameters are poorly quantified with emergent pathogens like SARS-CoV-2, leading to a wide range of predicted outcomes.

The first mathematical model used to study disease, and the impact of intervention, is widely attributed to Daniel Bernoulli for his work on inoculation against smallpox [2]. Charles Picquet used data on cholera death rates in 48 districts of Paris to perform the first geospatial analysis of disease progression [3] and this was followed by John Snow's analysis of cholera in London [4]. Models for how infectious diseases move through populations began to appear in the early 1900s and were developed to understand the drivers for epidemics and the number of people who could be affected. The simplest of these are compartmental models which group a population into categories (e.g. susceptible, infected, recovered). The basic concept dates to at least 1908 and the work of Ronald Ross on malaria [5], and the common formulation to Kermack and McKendrick [6]. Here the susceptible and infected groups are coupled by an infection rate,

α, and the infected and recovered groups by a recovery rate, β [7].

Compartmental models can be fit to case incidence rates in the early part of epidemics to estimate infection and recovery rates and the reproduction number (R_0) for the disease. For a standard SIR model, $R_0 = \alpha/\beta$ [7]. However, models like these are considered a poor choice for estimating R_0 (and the underlying infection and recovery rates) during the initial phase of an epidemic because they assume homogeneous mixing and uniform susceptibility of the population. The first of these assumptions is particularly incorrect during the initial phase of an epidemic as the infection process here is typically stochastic. As a result, other methods such as those based on contact networks (e.g. [8]) or stochastic branching processes are often employed [9].

Whether people are uniformly susceptible to SARS-CoV-2 is poorly understood at this time. Work on SARS-CoV showed some antibody cross reactivity with common human coronaviruses [10], [11]. It is possible, and perhaps likely, that this would also be true for SARS-CoV-2 and could afford some protection or immunity. Other work has recently shown that there could be cross-reactive T-cell recognition between common human corona viruses and SARS-CoV-2 in up to 50% of people [12]. Studies that followed individuals who recovered from SARS-CoV showed measurable antibodies in half of recovered people at 4 years and over 90% at 2 years [13]. This suggests that some level of immunity should be conferred though it may fade with time.

The reproduction number is often estimated using the serial interval time, γ, and data on the exponential growth of an epidemic where the growth rate $g = (R_0-1)/\gamma = \alpha - \beta$ [9]. However, for SARS-CoV-2, estimates for the exponential growth from studies in the first four months of the pandemic vary considerably, as do those for the serial interval. Estimates for the recovery and incubation period are similarly broad. Table 2.1 (page 5) gives examples for epidemic parameters from studies published in the first several months of the SARS-CoV-2 pandemic.

Uncertainty also exists in how infection rates, and R_0, would change with population size or density. Past work has shown that the reproductive number had a weak correlation with population density for the Spanish Flu [22] and recent work has shown scaling of flu epidemic intensities with population size [23]. Work on social contacts, drivers for transmission of infectious diseases, shows that the number of contacts ~ N^a where N is the population size.

Table 2.1 Estimates for critical parameters for modeling the SARS_CoV-2 pandemic. Data on the growth rate, serial interval time, R_0 recovery and incubation period from peer reviewed studies published in the first 4 months of the pandemic. The 9.8-day recovery period listed is post-onset of symptoms, while the 7.5-day period is reported as the time to become non-infectious.

Parameter	**Units**	**Estimated**	**95% CI**	**Reference**
Growth rate (g)	1/day	0.29	0.21 – 0.37	[14]
		0.19	0.09 – 0.69	[15]
		0.10	0.05 – 0.16	[1]
Serial interval (γ)	day	7.5	5.3 - 19	[1]
		5.8	4.8–6.8	[16]
		4.6	3.5 - 5.9	[17]
		4.0	3.53–4.39	[18]
R_0	-	5.7	3.8 - 8.9	[14]
		2.6	2.1 - 5.1	[15]
		2.2	1.4 – 3.9	[1]
Recovery period	day	9.8	8.5 - 21.8	[19]
		7.5	5.0 -15.2	[15]
Incubation period	day	5.2	1.8–12.4	[20]
		5.1	4.5 – 5.8	[21]

The exponent ‘a’ has been found to range between 0.1 and 0.14 [24] which is similar to the scaling effects observed with flu for populations greater than 1 million [23]. Common human coronaviruses can also exhibit significant seasonality which could affect the course of the SARS-CoV-2 pandemic in the northern and southern hemispheres (e.g. [25]).

A critical issue with emergent diseases is the stress that they can place on medical personnel and hospitals. This has been seen in many regions where SARS-CoV-2 has high infection rates. Estimates from China indicated a median hospital stay for SARS-CoV-2 of 11 days (IQR 10-14) [26] with severe cases having a median of 13 days (IQR 11.5-17) and non-severe cases a median of 10 days (IQR 10-13). Estimates for the percentage of symptomatic

SARS-CoV-2 cases requiring hospitalization, with critical care and mortality rates are available for the US (e.g. [27]) and other countries (e.g. [1]), though these data are likely skewed as they are based only on people seeking care. Actual case fatality, hospitalization and critical care requirements appear to vary geographically as well [28] but this could also be influenced by differences in testing levels. Table 2.2 gives ranges for hospitalization, critical care and fatality rate by age of symptomatic infected people in the US for data available as of 16 March 2020 [27]. The proportion of hospitalized cases, critical conditions cases and fatality varies by age.

Table 2.2 Proportion of infections leading to hospitalization in the US [27] Data ranges are for hospitalizations in the US between February 12–March 16, 2020.

Age category	**Percentage range for symptomatic infected individuals**		
	Hospitalization	**Intensive Care Unit**	**Fatality**
0-19 years	1.6% - 2.5%	0	0
20-44 years	14.3% - 20.8%	2.0% - 4.2%	0.1% - 0.2%
45-54 years	21.2% - 28.3%	5.4% - 10.4%	0.5% - 0.8%
55-64 years	20.5% - 30.1%	4.7% - 11.2%	1.4% - 2.6%
65-74 years	28.6% - 43.5%	8.1% - 18.8%	2.7% - 4.9%
74-85 years	30.5% - 58.7%	10.5% - 31.0%	4.3% - 10.5%
85 years and above	31.3% - 70.3%	6.3% - 29.0%	10.4% - 27.3%

Because data on hospitalization and mortality are being reported on the basis of symptomatic cases, it is also important to understand what fraction SARS-CoV-2 infections are asymptomatic. This has been estimated at ~18% for passengers on the Diamond Princess cruise ship [29] and as high as 78% for unconfirmed reports coming out of China [30]. A preliminary study on the municipality of Gangelt, Germany (population ~12,529) looked at a random sample of ~ 1000 people [30]. Blood tests and PCR were used to determine the fraction of people with active infections and those who were SARS-CoV-2 IgG positive. The results showed 2% of the sample with an active infection, but 14% who were IgG positive, suggesting previous infection

[31]. A recent study of 215 obstetric patients in the US presenting for delivery showed 33 testing positive for SARS-CoV-2 but with 87.9% of them being asymptomatic [32].

Because clinical data show hospitalization rates to be age dependent, it is also important to understand the relative susceptibility of different age groups to SARS-CoV-2 and if transmission rates are age dependent. A recent study on potential hospitalization rates in the US [33] used a contact matrix [34] which shifts simulated infections to the younger part of the population. However, a population screening study in Iceland, published subsequently, showed the opposite result; younger people (and children under 10 in particular) were found less likely to be infected than were older segments of the population. Here two screens were performed, one targeted at people considered high risk (symptomatic persons, those having traveled to high risk areas, etc) and one targeted at the general population [35]. The targeted screen showed more uniform infection rates across age groups, but with children again showing lower infection rates. A study of 3,712 people in Germany showed children to have a lower rate of infection than adults but to carry similar viral loads when they do [36].

Here we show how uncertainty in infection and recovery rates, seasonality, hospitalization and case fatality rates, affect the predicted impact of the SARS-CoV-2 pandemic in the United States. Because of the wide uncertainty in all key input parameters, we use the simplest compartment model, SIR, to simulate the fraction of the US population infected as a function of time. This helps avoid assuming information, required for more complex models, which is often poorly established in the early phases of an emergent pathogen (e.g. [37]). We use a Monte Carlo simulation to show the spread in the predicted outcomes for the fraction of the population infected as a function of time, hospital beds needed, and mortality rates with and without seasonal effects.

2.2 Methods for Monte Carlo Simulation

Data for the epidemic growth rate, and time to become noninfectious, were taken from the studies listed in Table 2.1. Because of the symmetry of the confidence intervals, the 0.10/day and 0.29/day exponential growth rates were assumed to follow a normal distribution with the standard deviations chosen to best fit the confidence intervals. A third exponential growth rate (mean 0.16, CI: 0.11 - 0.21) was computed by fitting an exponential to the data from [1]. These were used in conjunction with the recovery period to determine the possible distributions for the infection rate, α. The 7.5 day recovery period was used here because it was given as specific to

the time to become noninfectious and because its distribution (normal) was specified [15]. The infection rate was assumed to be uniform across age groups in keeping with the general results of the targeted screen in Iceland and Germany [35], [36]. The infectious period was assumed to overlap with the incubation period to be consistent with observations of asymptomatic infection [16], [38]. It was assumed that reinfection is not possible during the one year simulated time frame [13]. Monte Carlo simulations for the SIR equations were used to randomly sample the infection and recovery rates under the assumption that SARS-CoV-2 is not affected by time of year, or has the same average seasonality as other human coronaviruses. The effect of seasonality was modeled using data on the fraction of positive tests for human coronaviruses as a function of the time of year with the infection rate scaled accordingly. The portion of a population that would be hospitalized at any one time was determined using the fraction of the infected population that would be symptomatic, and the fraction of these that would require hospitalization or ICU care as well as their residence time in each. For each of these a uniform distribution was assumed due to the absence of other information. A complete description of the implementation is given in Supplementary Information Note 1.

2.3 Results and discussion for SARS-CoV-2.

Figure 2.1 shows the results of 1,000 Monte Carlo simulations for the fraction of the US population that would be infected as a function of time (left) in the absence of any interventions. The horizontal axis is days after the appearance of the first case, which is thought to be January 20th for the US [39]. Depending on the combination of α and β, the peak fraction of the population infected can vary by more than a factor of 6. Figure 2.2 (right) shows the corresponding fraction of the population that would become infected before the epidemic burns itself out, which again varies considerably. The results show that in the absence of interventions a very high fraction of the population could be infected by SARS-CoV-2 across the full range of simulated epidemics.

When all the simulations in Fig. 2.1 (left) are combined with the distribution of symptomatic people and the age dependent distributions for hospitalization rates shown in Table 2, the result can be used to determine the fraction of combinations which give a specific hospitalization rate for the US population. This is shown in Fig. 2.2 (left). Here the output of these combinations was filtered to remove the simulated epidemics producing the 2.5% highest and lowest peak hospitalization rates. The left panel then shows the range for the fraction of a

population that would need to be hospitalized as a function of time along with the epidemic that gives 97.5th percentile, 50th, and 2.5th percentile mortality rates. All other possibilities between these limits sit within the shaded boundary. A heat map for their density is given in Supplementary Note 1.

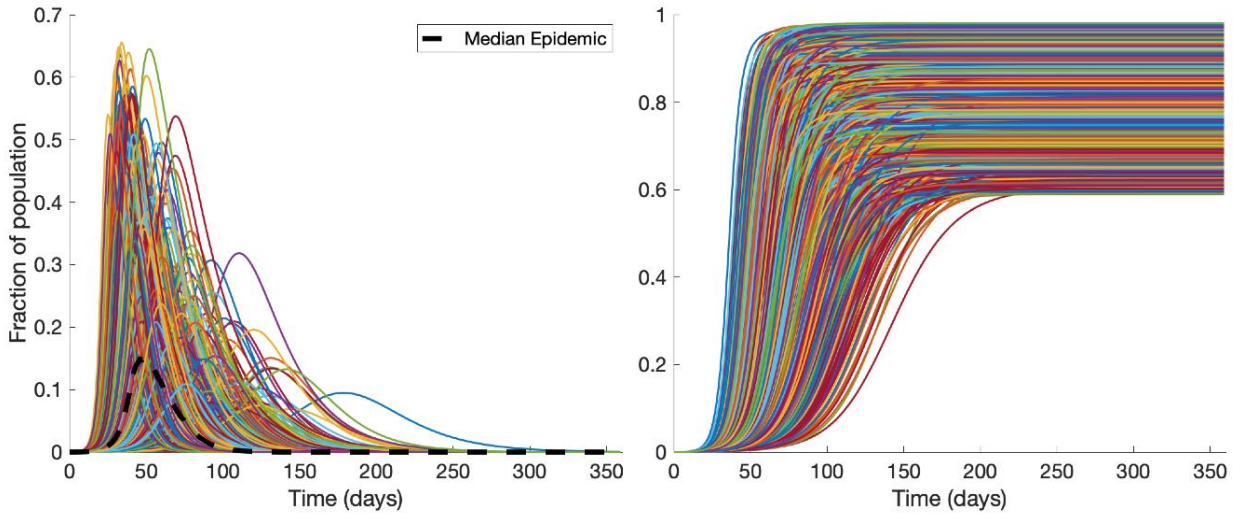

Figure 2.1 Monte Carlo simulation of the range of possible infection and recovery rates. (left) The results from 1,000 iterations randomly sampling distributions for the infection rate and recovery rates. Depending on the combination of α and β the peak can vary by more than a factor of 10. (right) The cumulative fraction of the population that would become infected. The results assume no social distancing.

The same method is applied to the mortality rate, with the output of the combination of Monte Carlo simulation with the symptomatic distribution and the age specific mortality distributions filtered to remove the 2.5% highest and lowest mortality rates. This is shown in Fig. 2.2 (right). The corresponding heat map is given in Supplementary Note 1. These results show that, in the absence of interventions, low impact epidemics are a small subset of the predicted outcomes.

Figure 2.2 (right) shows the corresponding range of results for predicted mortality by day. The uncertainty in both is large with the daily hospitalization and mortality rates falling in a range more than an order of magnitude wide. However, at the low end one would expect mortality of 450,000 people in the US, and at the upper end more than 4.4 million with a median of 2.0 million. The median is just below the total number of fatalities seen in the US from the

Spanish flu in the 1918-1919 flu if the US population were scaled to its current level. This range of potential fatalities supports the serious social distancing efforts put in place by many US states. It is important to note that there could be scaling effects that reduce infection rates for SARS-CoV-2 as urban population sizes decrease. This effect has been shown to occur with flu [23] and its potential influence on SARS-CoV-2 requires further investigation.

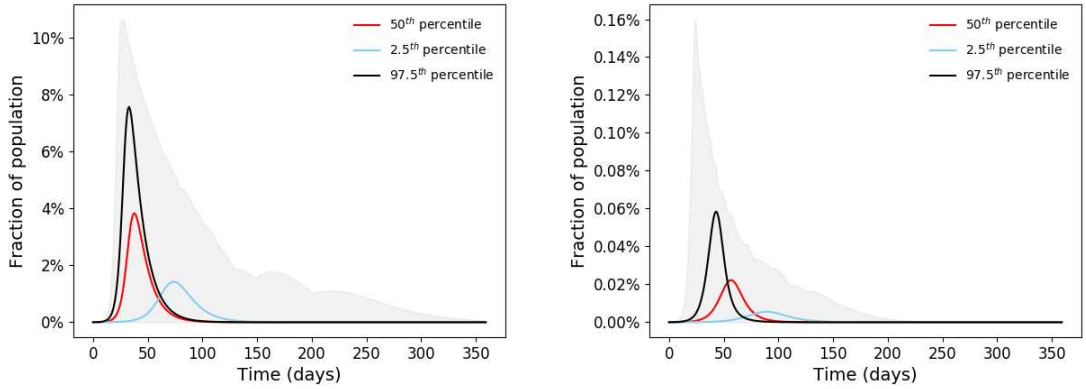

Figure 2.2 Monte Carlo simulation of the range of possible hospitalization and fatality rates**.** (left) The daily hospitalizations results from 1,000,000 iterations randomly sampling distributions for the infection rate, recovery rate, symptomatic rate and hospitalization rate. (right) The daily fatalities results from 10^3 simulations randomly sampling distributions for the infection rate, recovery rate which are inputs to 10^3 simulations sampling the symptomatic rate and fatality rate. The results assume no social distancing.

The effect that seasonality could have on the time course for SARS-CoV-2 is shown in Fig. 2.3. Here the simulated epidemic starts on January 20, the date of the first reported case in the US [38] Figure 2.3 (left) shows different predictions for how the epidemic could have evolved in the absence of intervention. Seasonality itself could have a strong suppressive effect and drive new infections close to zero during the summer months. Figure 2.3 (right) shows a subset of the simulation results to highlight the dynamics. Strong epidemics early in the year lead to dampened out ones later in the year. But weak early epidemics can result in strong spikes in the fall. Importantly though, in the absence of herd immunity the epidemic would rekindle in the fall, though it would be less intense due to immunity from those previously infected. The possibility of seasonal effects with SARS-Cov-2 is important to consider as the disappearance of new infections during the Northern hemisphere summer could give an unwarranted sense that threats from the virus have passed.

The results of the Monte Carlo simulations shown in Fig. 2.1 (left) can also be used to

predict the median epidemic profile. This can be used to understand effects of social distancing, and this is shown in Fig. 2.4. For a 40% reduction in contact frequency the peak for the fraction of the population infected would drop by a factor of four. This is roughly equivalent to reduction reported by Google for several different categories of 'mobility' after social distancing efforts were put in place by many states [39]. However, the relationship between reduction in contact frequency and 'mobility' has yet to be established, and would likely vary across populations. It is important to note that while reducing social contact rates has a clear effect on a simulated SARS-CoV-2 epidemic, the rate of infection would rise when social distancing measures are relaxed. This would be true until herd immunity is achieved, or a vaccine developed. Figure 2.4 (right) also shows the herd immunity level for the average pandemic. Social distancing can reduce the total number of infections to that required to achieve herd immunity. This assumes that immunity persists, which is still an open question for SARS-CoV-2.

The results in Figs. 2.1 – 2.4 use a subset of the data from Table 2.1 corresponding to an average infection rate generated using the distributions for three estimates of the exponential growth for the epidemic in Wuhan and the distribution for the estimated recovery period of 7.5 days (Supplemental Information Note 1). However, it could also be that the different exponential growth rates correspond to different potential outcomes for a SARS-CoV-2 epidemic. In this case it would be more appropriate to sample three separate distributions for the infection rate instead of an average one and the result is a slightly wider range of outcomes. The results for this are shown in Supplementary Note 2. The wide range of uncertainty in both cases points to the need for more work to better understand the parameters on which epidemic models rely. This is not an issue of model fidelity. Instead, fundamental properties of the virus, and the disease itself, need to be better understood to avoid a false sense of precision in predicted outcomes. This will require laboratory and clinical study that reduces parameter uncertainty and establishes their statistical distribution (e.g. not just mean values).

The assumption made here that all age groups are similarly susceptible to infection is generally consistent with the results from screening in Iceland and Germany [34], [36].

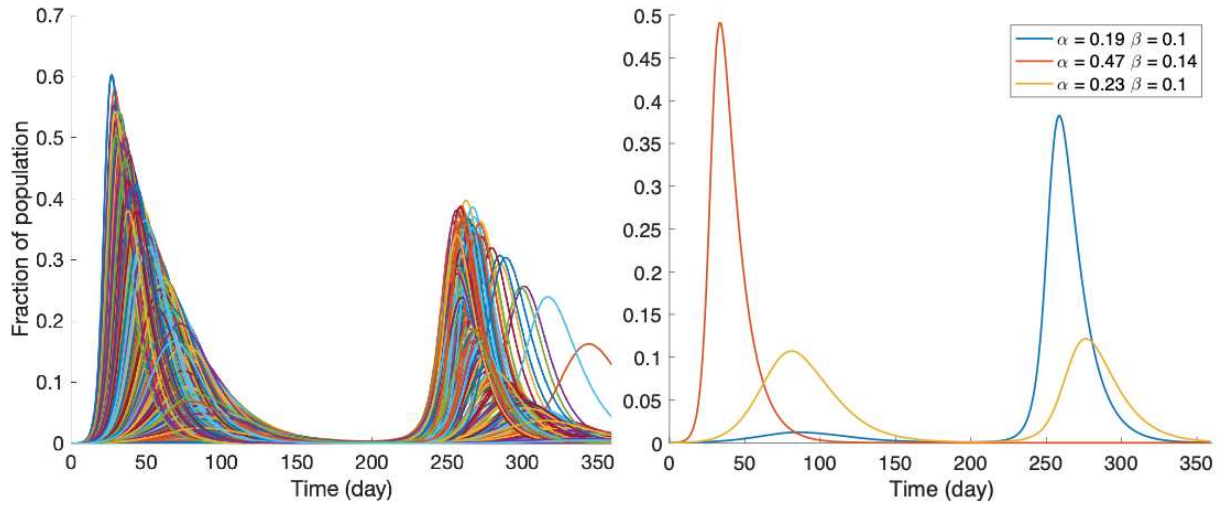

Figure 2.3 The possible effect of seasonality. If SARS-CoV-2 behaves like other human coronaviruses, infection rates would be expected to drop significantly during summer and early fall, but then increase again during late fall and winter. (left) This effect on the range of Monte Carlo simulations is shown. (right) Three scenarios with different infection (α) and recovery rates (β) are shown for clarity. Large infection rates early in the year have correspondingly smaller ones in the fall. By contrast, small epidemics in the early part of the year would rebound to much higher rates in the fall. While the results here assume no social distancing, they are consistent with what would happen if social distancing is ended during the summer when seasonal effects might naturally dampen infection rates.

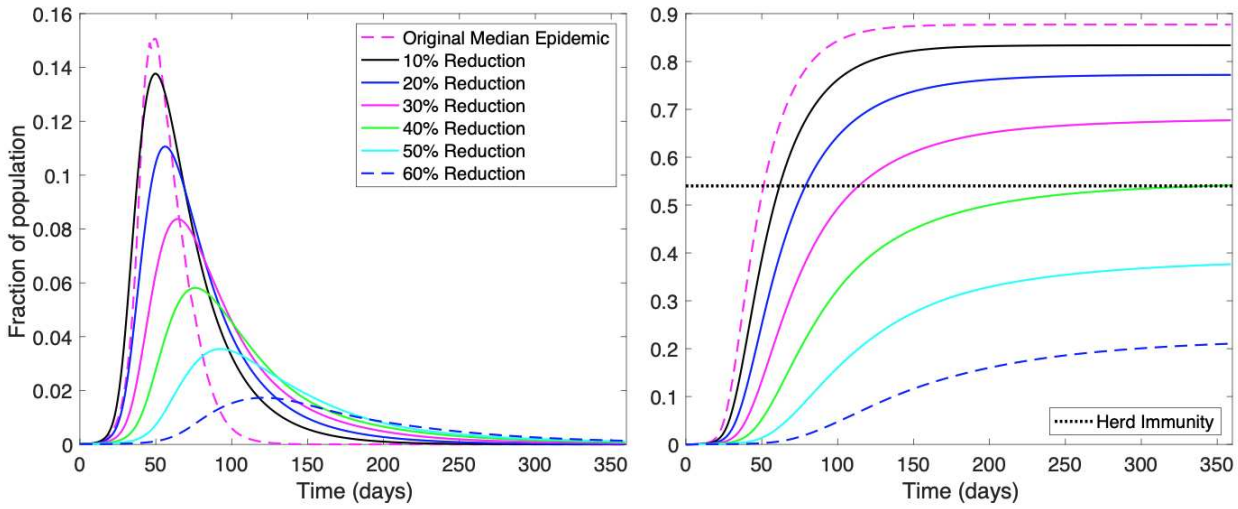

Figure 2.4 Effect of social distancing on the time course of the epidemic. Social distancing here is assumed to reduce the contact rate between people and have a corresponding effect on the infection rate. R_0 for the median epidemic is 2.2 which corresponds to a herd immunity level of 0.54 which is shown by the horizontal black dotted line (right).

However, other studies have employed age-dependent infection rates based on data for contact frequencies between age groups derived from surveys [32]. The result is an infection profile concentrated at the younger end of the population. However, data subsequently published has shown lower infection rates in children [34], [36]. While the approach is in principle more realistic, it is only more accurate if the age-dependent infection rates are correct.

The modeling approach used here makes no distinction between the ability of symptomatic and asymptomatic people to cause infection and uses an aggregate infection rate for the susceptible class as a whole. Other studies have used network population models, and estimates for the number of mild or asymptomatic cases, and concluded that people in this category have lower infection rates than do symptomatic people but constitute 46-62% of total transmissions [40]. Data published subsequently showed the that people become most infectious just before they become symptomatic and also that 46-55% of transmissions occurred from this state [16]. If the infection rates for symptomatic and asymptomatic classes Li et al (2020) [40] are combined in population weighted manner, the aggregate infection rate falls within the range used in the current study.

The effect of seasonality on the progression of the epidemic is important. Should SARS-CoV-2 behave as other human coronaviruses, the results in Fig. 2.3 show new infections would largely disappear during the summer months but reemerge in the fall. Reemergence would be inevitable until a vaccine becomes available or herd immunity is achieved. The simulations done here assume that people once infected and recovered are immune. However, the lifetime of antibody response has not yet been established with SARS-CoV-2. If it fades with time, then a cyclic pattern of reemergence could occur.

The effect of social distancing is clear in Fig 2.4. Importantly, this has potential to do more than just “flatten the curve” and reduce peak hospital demand. Without a vaccine SARS-CoV-2 will continue to be a problem unless herd immunity is achieved. In the absence of social distancing the epidemic can move so quickly through a population that many people get infected before the immunity people gain as they recover can reduce the spread. By reducing the contact frequency between people one can also reduce the degree to which the epidemic overshoots the population fraction needed to achieve herd immunity.

2.4 Conclusions.

Accurately modeling the progression of an epidemic requires understanding many properties of a virus that are poorly understood with emergent pathogens like SARS-CoV-2. The first four months of this pandemic have seen a flood of studies that are advancing the state of knowledge for this pathogen at a truly impressive rate. However, as Table 2.1 shows the range of values for the reproductive number, exponential growth rates, as well incubation and recovery periods is large. This leads to a correspondingly large range for infection and recovery rates and considerable uncertainty in predictions for the course of the SARS-CoV-2 epidemic from even the simplest epidemic model (SIR). As Robert May observed, more complex models are attractive, but only if their input parameters are understood [37]. Key questions remain about the relationship between age and susceptibility to this virus, seasonality, the effect of population scaling on infection rates, and the lifetime of antibody response. For epidemic models, these are all areas of significant uncertainty. Even so, the results presented here show that the implementation of social distancing measures was well founded.

Acknowledgements. Special thanks to Sara Sawyer for discussions about the virology of SARS-CoV-2, editorial comments and suggestions.

2.5 Description of Monte Carlo Simulations.

$$Total\ Population = N(t) = S(t) + I(t) + R(t) \quad (2.1)$$

$$\frac{dS(t)}{dt} = -\alpha \cdot S(t) \cdot I(t) + \mu\, S(t) \quad (2.2)$$

$$\frac{dI(t)}{dt} = \alpha \cdot S(t) \cdot I(t) - \beta \cdot I(t) \quad (2.3)$$

$$\frac{dR(t)}{dt} = \beta \cdot I(t) \quad (2.4)$$

Here *N(t)* is the population size , *S(t)* is the susceptible population, *I(t)* the infected population and *R(t*) the recovered population. The infection rate is α(1/day) and β (1/day) is the recovery rate. In this work, the size of the population, *N(t)* is a constant with no growth (i.e. μ=0).

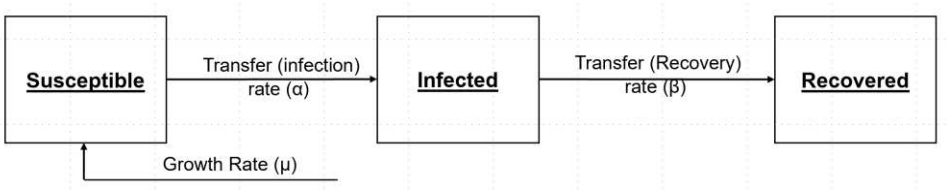

Figure 2.5 The susceptible, infected and recovered (SIR) model. The basic SIR model couples a susceptible class of people, with growth rate μ, to infected and recovered classes through transfer rates.

Distributions for α were generated using the data for the epidemic growth rate, and the time to become noninfectious. Three distributions were used for the former, two from Table 2.1 of the main paper (0.10/day and 0.29/day) and one obtained by fitting an exponential to the data from Li et al (2020) which gave a mean value of 0.16/day (SD: 0.03). All three were assumed to follow a normal. The 7.3 day recovery period was used here because it was given as specific to the time to become noninfectious and because its distribution (normal) was specified [15].

The Monte Carlo simulations were done assuming that the three distributions for α were independent samples from an underlying distribution, and that they were independent possible infection rates for SARS-CoV-2. In the former case the Monte Carlo simulations were done on an average distribution generated by averaging the means and summing the variances for the infection rate. In the latter case the three individual distributions for the infection rate were sampled.

One thousand Monte Carlo simulations were used to randomly sample the distributions for α and β to generate the time course in Figs. 2.2 - 2.5. The portion of a population that would be hospitalized at any one time was determined using:

$$\frac{dH(t)}{dt} = -f_s \cdot f_h \frac{dI(t)}{dt}\Delta t - r \cdot H(t) \qquad (2.5)$$

Here f_s {-} is the fraction of infected people who are symptomatic, f_h {-} is the fraction of these who require hospitalization, Δt is the time step and r {1/day} is the discharge rate for people from the hospital. The discharge period was taken to be the inverse of the hospitalization time and was assumed to follow a uniform distribution with a range of 10-14 days and Δt was set to one day (the resolution of the data). Uniform distributions were also used for f_s and f_h with ranges of 0.22-0.82 for the former. The ranges for f_h were age specific and taken from Table 2.1 and weighted by the average age demographic of the United States [41].

The 1000 Monte Carlo simulations used to generate Fig. 2.1 of the main paper, and the results in Fig. 2.1 (left) were used as inputs for determining hospitalization rates. Figure 2.6

shows the histogram of infection rates produced by this sampling. For each of the results show in Fig. 2.1 (left) of the main paper, 1000 Monte Carlo simulations were run where the distributions for f_s and f_h were sampled. The output was used to generate Figs. 2.2 and 2.3 of the main paper.

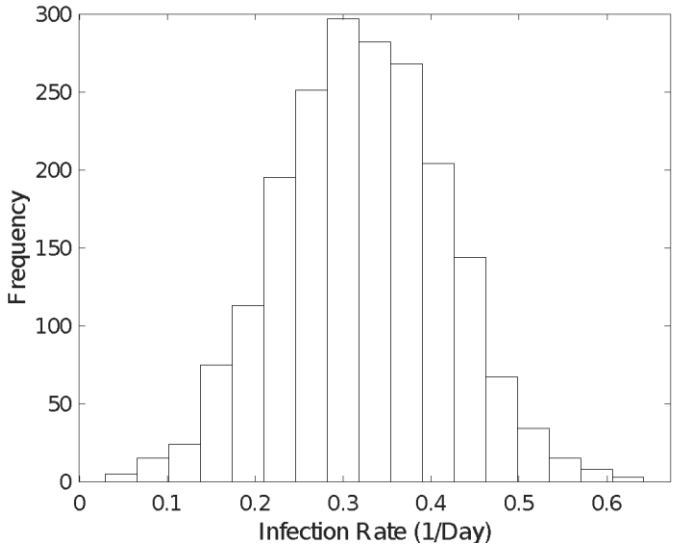

Figure 2.6 Histogram of infection rates produced by sampling the average infection rate distribution 1000 times.

The density of epidemic hospitalization and morbidity trajectories is given in Fig. 2.7. The Monte Carlo simulations are used to create epidemic trajectories, and each of these is used as an input to another Monte Carlo simulation that gives the range for hospitalization and morbidity. The boundary of these figures is the same as those in left and right panels of Fig. 2.2, but a heat map shows the regions where the hospitalization and morbidity trajectories have their greatest and lowest density.

Seasonality. The possible effect of seasonality was determined using clinic data on the fraction of people testing positive for human coronaviruses (CoV 229E, Cov NL63, Cov HKU1/OC43) in Switzerland by month over the course of a year [25]. The relative infectivity was determined by scaling the positive test fraction to the month of January, Fig. 2.8, as this was the month with the dominant emergence of the SARS-CoV-2 epidemic in Wuhan. The relative infectivity was used to scale the infection rate, α, by month in Eqs. (2.2 – 2.3).

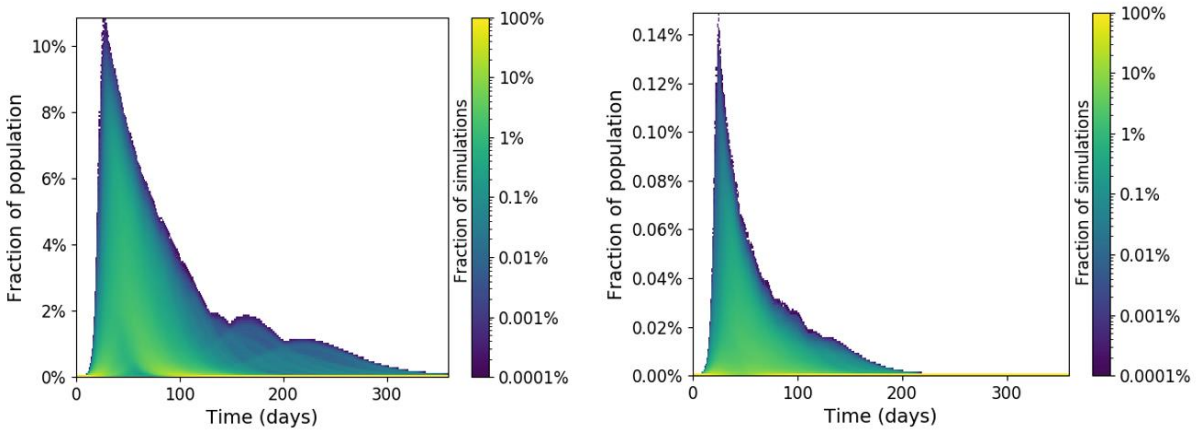

Figure 2.7 Density of hospitalization and morbidity trajectories. (left) Density of hospitalization trajectories as a function of location within the Monte Carlo output space. (right) Density of morbidity trajectories as a function of location within the Monte Carlo output space. Both figures show the regions of the graph where the majority of simulations overlap.

2.6 Effect of Independent Infection Rate Distribution.

Monte Carlo simulations were performed in which three individual distributions for the infection rate, α, added together (not averaged). The collective distribution was sampled 3000 times, Fig. 2.9. This produced the epidemic profile shown in Fig. 2.10 (left) and the cumulative fraction of the population infected, Fig. 2.10 (right).

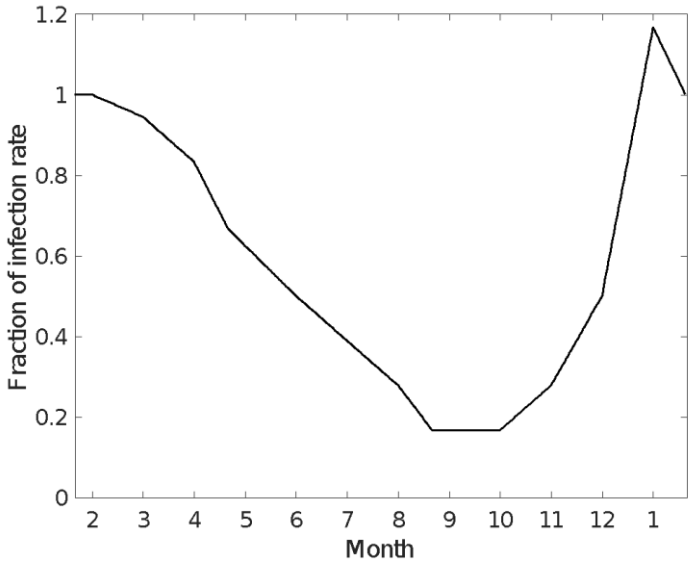

Figure 2.8 Relative infectivity used to model the possible effect of SARS-CoV-2. The relative infectivity was determined using data on the fraction of people testing positive for human coronaviruses CoV 229E, Cov NL63, Cov HKU1/OC43 by month in Switzerland.

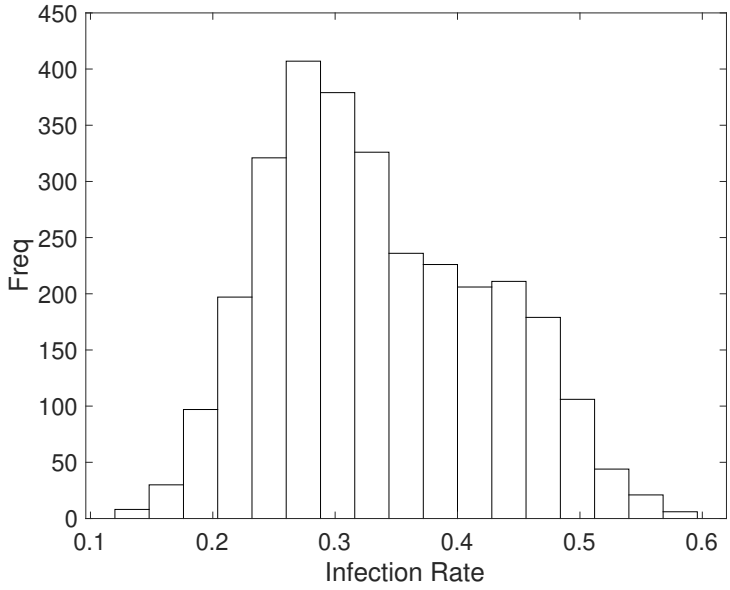

Figure 2.9 Histogram of infection rates produced by sampling the average infection rate distribution 3000 times. Three times the sample rate was used for these simulations as three separate distributions for α were added together to get the distribution for infection rate.

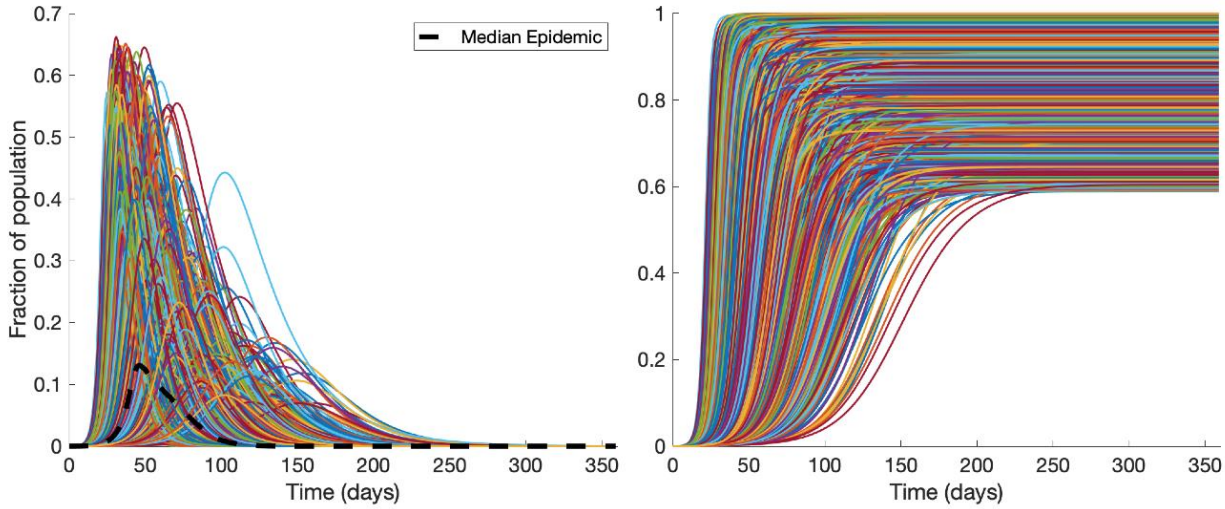

Figure 2.10 Results of sampling three independent distributions for the infection rate. The results are very similar, though slightly broader, than those in Fig. 1.1 of the main paper.

CHAPTER THREE

EFFECT OF SENTIMENT ON THE TRANSMISSION OF IDEAS THROUGH SOCIAL MEDIA

The diffusion of ideas on social media can affect societal trends, norms, behaviors, influence markets and the outcomes of elections. How and why some ideas spread farther and faster than others has been the subject of considerable research. Cascade size, width and depth of penetration into social networks are standard parameters used to quantify diffusion. However, a more refined measure of how susceptible populations are to particular ideas, and how quickly they lose interest in them, would be of considerable value. Here we measure the infection and recovery rates of twitter messages using an SIR model with tweets grouped into positive, neutral or negative sentiment categories. Bayesian data synthesis is used to determine the effect of sentiment when the size of follower networks is also taken into consideration. Our analysis of 4,252 initial messages, and 1.5 million retweets, shows that positive messages have lower recovery rate and a higher probability of being retweeted.

3. 1 Introduction to idea adoption and using social media.

A viral message on social media is one that spreads rapidly through a population by being frequently shared with others [42]. An important question is why some ideas 'go viral' while others do not and what modes can be used to understand how a population receives, values and spreads one idea over another. Quantifying the effect of a particular message is often done by measuring cascade size, width and depth of penetration into social networks along with corresponding statistics [43], [44]. While this type of analysis is informative, quantifying how susceptible populations are to particular ideas, and how quickly they lose interest in them, would be of considerable use. Here epidemiological models can be of great value and provide a way to measure the rate at which people are infected by, and recover from social media messages.

How ideas move through populations has been studied extensively [45]. Work in this area has its roots in the early 20th century [45], [46] though it's mentioned in this context comes much earlier [47]. Analyses of what drives idea propagation dates to at least 1950 and an empirical study of how rumors move through social structures [48]. A 1957 study analyzed the influence physicians have on the adoption of a new drug, measured the rate at which if occurred, and proposed a simple kinetic model to explain it [49]. The 1960's saw the development of

models for the rate at which innovation spreads between firms [50] and the development of a theory of adoption to predict product growth [51].

The concept of an idea or behavior spreading like a contagion has been in use since at least the mid 19th century [47]. The first epidemiological models to describe how diseases move through populations were developed in the early 1900's [5] and the widely used SIR model was formulated by Kermack and McKendrick in 1927 [6]. Its first application to the transmission of ideas was given by Goffman and Newell in 1964 [52] and the approach subsequently found use in the developing area of information science.

Early work on the adoption of ideas and behaviors was limited by access to sufficient data with which to understand the underlaying dynamics [53]. The growth of the internet, and social media in particular, changed this. A key advantage here was data volume and the ability to map the network through which an idea moved, its person to person transmission [43], [53]–[56] and who followed whom [57]. Models were subsequently developed for the spread of ideas through social media that treated all nodes within a network the same [58]–[60]. Other approaches identified nodes that had more influence than others [60], [61] an observation which was made in the mid 1960's in a study of curriculum adoption [62]. The diffusion of ideas through social media has also been modeled as a threshold process [63]–[65], a cascade effect [25], [28] and using epidemiological models [67].

Access to large amounts of data, and the ability to automate content analysis, has enabled studies of what drives the propagation of messages on social media. Work rooted in epistemology has looked at the effect that veracity has. Figgeri et al (2014) [68] studied rumors on Facebook which had been found by Snopes to be, true, false or mixed. Their analysis showed that true rumors elicited the largest cascades and showed the greatest virality. Vosoughi et al (2018) [69] studied messages on Twitter and found the opposite. Using six independent fact checking organizations they categorized tweet content as true, false, or mixed. False messages were seen to be more novel, spread faster, wider and deeper into a network and have a higher probability of being retweeted. However, whether the retweet agreed with the false content of the message was not measured [69].

Sentiment has also been looked at as a driver for virality [70]. Jenders et al (2013) [71] studied multiple aspects of messages to understand what increased the probability of retweet. They used SentiStrength [72] to categorize the emotional content of tweets and found that

negative messages are more likely to be retweeted than positive ones [71]. However, they also found that the number of followers that a person had showed greatest effect on retweet probability. Figgeri et al (2014) [68] showed that the number of followers associated with true and false messages differed and it is possible that this is true for different sentiments as well. If this is the case, the effect that negative sentiment was found to have could be driven by the size of the follower networks instead of the sentiment itself.

Here we look at how sentiment effects the differential diffusion of Twitter messages. We use an SIR model to measure the infection and recovery rates for messages, and the retweet probability when the effect that followers have is explicitly taken into consideration. We look at 4,252 initial messages and 1.5 million retweets and retweets across categories of politics, weather, drones, nuclear, sports, military, covid-19, academics and business.

3. 2 Methods to analyze social media behavior by sentiment.

Data for this study was collected through the Twitter API. Tweets were grouped into positive, negative and neutral sentiment categories. Messages were assessed for sentiment using SentiStrength [72] with its lexicon assigning an integer value between -5 and +5 to words. The sign was used to group tweets into positive or negative sentiment categories. A score of zero was taken to indicate neutral sentiment. Sturges method was used to bin the time dependent Twitter data. A time series of data started with the initial tweet. At the end of the first bin, all the retweets were added up within each allocated bin time interval. In this manner, binned distributions were created for messages by sentiment, as well as for the number of followers.

An SIR model was fit to time series Twitter data for messages corresponding to each bin in the follower distributions. This gave a measure of the infection and recovery rates for each message as a function of the number of followers for the user from which the messages were sent. Bayes theorem was used to calculate the probability that messages would exceed the median infection and recovery rates, reproductive number, and retweet probability as a function of positive, negative or neutral sentiment. A Kolmogorov-Smirnoff test was used to test whether distributions were statistically different. Supplemental Note 1 gives complete description of the methods.

3.3 Results and discussion for SIR and sentiment analysis.

The ratio of infection to recovery rates in an SIR model gives the basic reproduction number, R_0, which indicates how many secondary cases are generated by a single primary case.

In order for an 'epidemic' to form R_0 must be greater than one. Figure 3.1 shows the distributions for infection and recovery rates, as well as the corresponding R_0, for messages with greater than 300 retweets as a function of sentiment. Table 3.1 lists the median values of these as well as the R^2 for the fit of the SIR model to the data across the message categories.

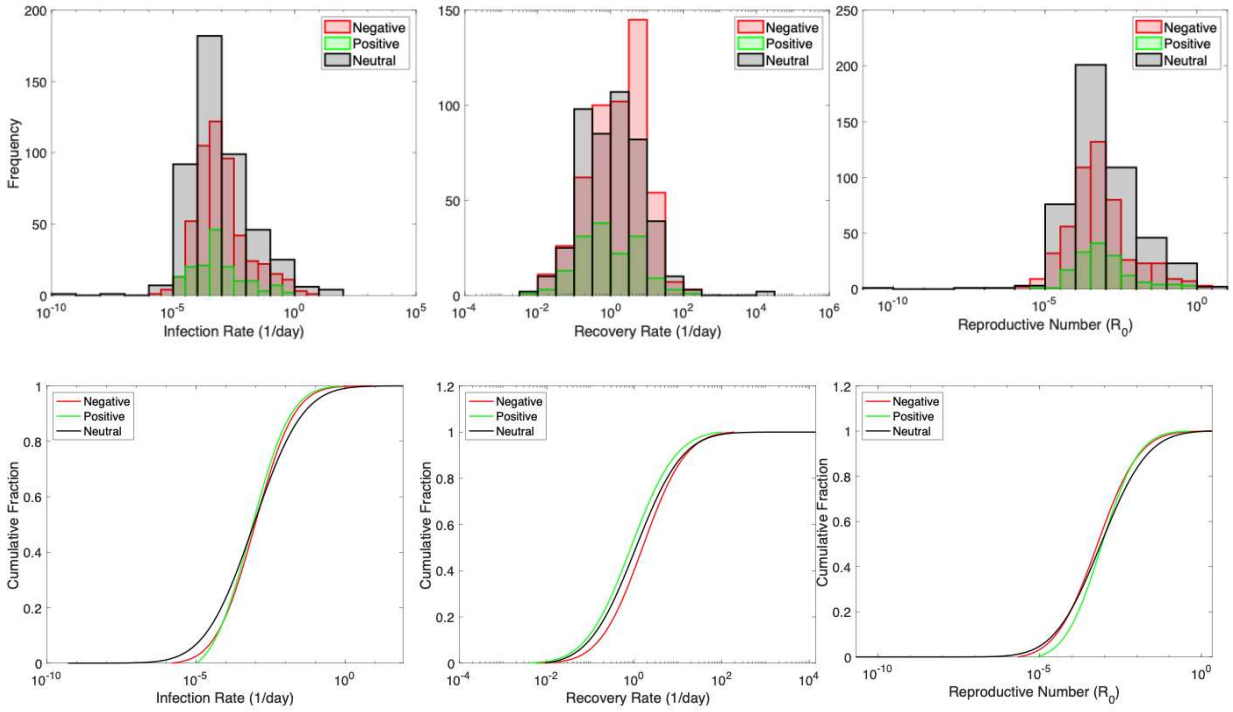

Figure 3.1 Distributions for infection rate the recovery rate, and reproductive number for messages with > 300 retweets. The top figures show the histograms for infection and recovery rates along with the reproductive number. The bottom figure shows the corresponding cumulative distributions. A KS2 test shows the distributions for the positive and negative sentiment infection rates to be essentially the same. However, the distributions are different for recovery rate with positive sentiment having a lower overall recovery rate.

A two-sample Kolmogorov-Smirnov (KS2) test was used to determine which distributions are statistically the same and which are different. Negative and positive messages show statistically different distributions for recovery rates, but similar distributions for infection rate. For the recovery rate, the p values were, Neg. – Pos. $p = 10^{-5}$, Neg. – Neut. $p = 10^{-5}$, and Pos. – Neut. P = 0.11. For the infection rate the p values were, Neg. – Pos. p = 0.20, Neg. – Neut. p = 0.03, and Pos. – Neut. P = 0.7. Supplemental Note 2 has the KS2 test for all distributions in figures 3.4 – 3.9 and a summary in Table 3.4.

Table 3.1 Lists the median SIR metrics by topic for messages with > 300 retweets.

	R^2	Infection Rate (1/days)	Recovery Rate (1/days)	Infection Period (days)	R_0	Total Followers
Nuclear	0.93	0.0013	3.88	0.26	$2.4 \cdot 10^{-4}$	2M
Drones	0.93	0.0015	2.64	0.39	$3.5 \cdot 10^{-4}$	1.5M
Weather	0.92	$2.4 \cdot 10^{-4}$	2.88	0.36	$2.9 \cdot 10^{-4}$	2.1M
Impeachment	0.97	$8.2 \cdot 10^{-4}$	5.1	0.19	$1.9 \cdot 10^{-4}$	4.9M
Corona Virus	0.96	0.001	3.57	0.28	$2.8 \cdot 10^{-4}$	2.7M
Business	0.91	0.18	3.4	0.29	$2.9 \cdot 10^{-3}$	1.3M
Politics	0.90	0.06	14.1	0.07	$6.1 \cdot 10^{-3}$	6.1M
Military	0.80	0.01	14.4	0.07	$1.5 \cdot 10^{-3}$	1.2M
Sports	0.89	0.01	5.7	0.18	$1.1 \cdot 10^{-3}$	9.9M

Estimates for infection period and basic reproduction number for Covid-19 pandemic are 5.0 -15.2 days and 2.1 - 5.1 respectively [15]. The range of R_0 values shows that messages investigated are not 'viral' in the conventional sense where one generation of infected people creates a larger second generation of infection, and so on. Instead, the very low R_0 values show that there is very little penetration into the twitter network which is consistent with recent findings [58], [69].

Overall, 39% of the initial tweets (not retweets) were negative, 46% neutral and 15% positive. If sentiment and size of follower networks had no effect, one would expect a similar distribution for retweets. However, this turns out not to be the case. The vast majority of messages investigated were found to have negative sentiment. This might suggest that negative messages have a higher probability of being retweeted. However, negative tweets were also found to occur in networks with a larger number of followers, Fig. 3.2, and this has been noted previously to correlate with retweet probability [68].

Bayesian analysis is helpful for determining effect that tweet sentiment, and the number of followers a sender has, on retweet probability, reproductive number and infection and recovery rates. The results are shown in Table 3.2. The analysis shows that positive messages have the highest probability of exceeding the median infection and recovery rates, and R0. The Bayesian analysis also shows that positive messages have a higher probability to have more followers and to be retweeted more.

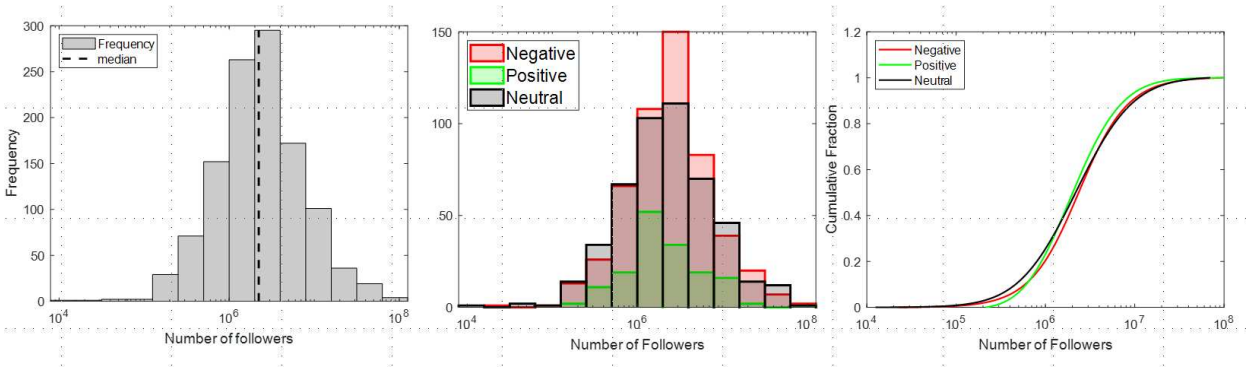

Figure 3.2 Follower distributions for Twitter messages by sentiment. (left) distribution of total number of followers for the messages studied. (middle) follower distributions by sentiment. (right) cumulative distributions. The distributions for the number of followers for users tweeting positive and negative messages showed a statistically significant difference with a KS test ($p = 0.008$). The difference between positive-neutral and neutral-negative messages was less significant ($p = 0.06$ and 0.12 respectively).

The primary difference between negative and positive messages is the longer infection period for the latter. Figure 3.3 shows the effec on the 'infected' class in the SIR model. The top row figures show the median infection rate and recovery rate by sentiment category and number of followers. Figure 3.3 (top left, middle and right) use the lowest, median and top bin from Fig. 3.2 (middle) for number of followers. The median infection and recovery rates for each sentiment category are then used to generate the respective figures for the time response of people 'infected' by a message. The results show that the period over which people are likely to retweet positive messages is longer than it is for neutral or negative ones. Figure 3.3 (bottom) shows the time response of people 'infected' by a message but using the average infection and recovery rates. Here too the positive messages are more persistent in categories of follower size, except the highest, where neutral messages are most persistent. Even here though, positive messages fade more slowly than negative ones.

The analysis here shows that sentiment influences how twitter messages diffuse and the SIR model provides a way to measure how infectious they are and how quickly people recover. We calculated the overall SIR metrics for all messages with > 300 messages. The low reproductive number indicates that twitter messages are not viral in the conventional sense, and that penetration into the network is low.

Overall, there are far more negative messages than positive ones in the tweet streams that were tracked. Negative messages are retweeted more often, but this appears to come from the larger networks in which they spread. When adjusted for network size, positive messages exceed

the mean infection rate, recovery period, and reproductive number. Despite fewer positive sentiment messages, positive messages were more likely to exceed the median retweet probability, and were more persistent across networks of all sizes.

Table 3.2 Probability of message retweet when sentiment valence is taken into account. The highlighted row within each SIR metric indicates the highest probability. Posterior probabilities for retweets were generated using Bayes theorem and data on tweet and retweet sentiment.

	sentiment	$\mathbf{P(Rec\ Rate > median|sentiment)}$
Recovery period	Positive	21%
	Negative	9%
	Neutral	12%
	sentiment	$\mathbf{P(Inf\ Rate > median|sentiment)}$
Infection Rate	Positive	42%
	Negative	19%
	Neutral	41%
	sentiment	$\mathbf{P(R0 > median|sentiment)}$
Reproductive Number	Positive	71%
	Negative	22%
	Neutral	49%
	sentiment	$\mathbf{P(followers > median|sentiment)}$
Followers	Positive	40%
	Negative	17%
	Neutral	15%
	sentiment	$\mathbf{P(Retweet > median|sentiment)}$
Retweet	Positive	32%
	Negative	18%
	Neutral	22%

This indicates that positive messages are retweeted with additional time and negative messages are not retweeted with additional time. With the longer infection period, positive messages have a higher probability of having a larger network and being retweeted more. For users that want to influence the largest number of people, the data suggests positive messages have a higher probability of having larger metrics despite the low number of overall messages.

The tweets gathered for this study do not represent a random sample from the Twittersphere. Instead they are drawn from a limited set of topics, Table 3.2, and were selected because they generated >300 retweets. This could skew the results and further investigation is required to understand if the results shown here persist across a broader tweet domain. The limited number of positive messages that were tracked, as compared to negative and neutral

ones, could also affect the results. Importantly, why positive messages would perform as they do needs to be understood.

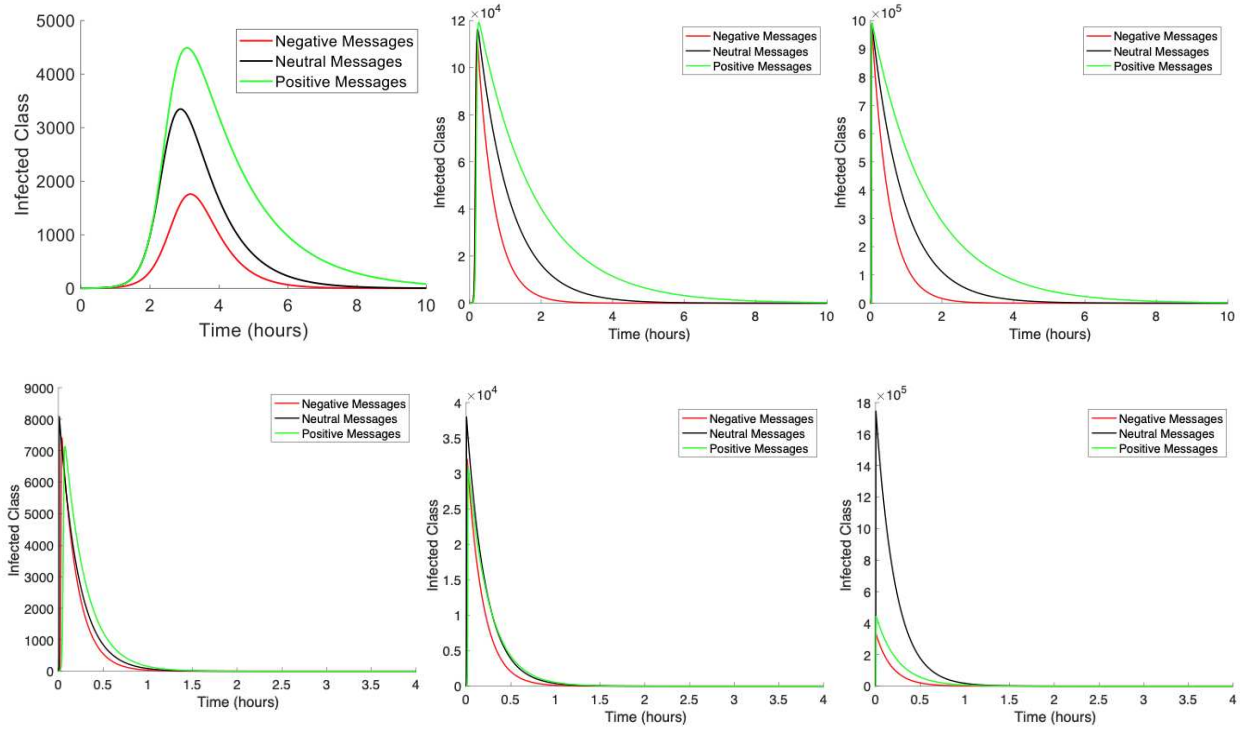

Figure 3.3 Effect of sentiment on rate of recovery. (top figures) The median infection and recovery rates for each sentiment category by follower bin, Fig. 3.2 (middle) are shown. The left figure corresponds to the number of followers from the low bin shown in Fig. 3.2 (left). The middle figure uses the median bin from Fig. 3.1 (left). The right figure uses the high bin from Fig. 3.2 (left). Negative messages show superior behavior on smaller networks and it decreases as the network size increases. (bottom figures) The average infection and recovery rates for each sentiment category by follower bin, Fig. 3.2 (middle) are shown. The left figure corresponds to the number of followers from the low bin shown in Fig. 3.2 (left). The middle figure uses the median bin from Fig. 3.2 (left). The right figure uses the high bin from Fig. 3.2 (left). Negative messages show superior behavior on smaller networks and it decreases as the network size increases.

Acknowledgements. Special thanks to Tracy Camp and Thyago Mota for discussions about mining social media data.

3.3 SIR model, sentiment analysis and Bayesian data synthesis.

Equations (3.1–3.5) describe the SIR model. Here N(t) is the total population, S(t), I(t) and R(t) are the susceptible, infected and recovered populations [6], [74]. The coefficients of α, β, and μ are the infectious, recovery and growth rates respectively. The SIR model is fit to

Twitter data by adjusting these coefficients to minimize the sum of the square errors, Eq. (3.5). A constrained nonlinear multivariable function in MATLAB (fmincon) is used to determine the SIR parameters of α, β, and μ to minimize equation 3.5 below [75]. The constraints were α and β > 0.

$$Total\ Population = N(t) = S(t) + I(t) + R(t) \quad (3.1)$$

$$\frac{dS(t)}{dt} = -\alpha \cdot S(t) \cdot I(t) + \mu \cdot S(t) \quad (3.2)$$

$$\frac{dI(t)}{dt} = \alpha \cdot S(t) \cdot I(t) - \beta \cdot I(t) \quad (3.3)$$

$$\frac{dR(t)}{dt} = \beta \cdot I(t) \quad (3.4)$$

$$SSE = \sum_{i=1}^{n} (I_{i,actual} - I_{i,estimate})^2 \quad (3.5)$$

Figure 3.4 shows how the SIR model can be applied to social media. For the current study, the social media platform is Twitter and it's assumed that all people recover. Out of the 4,252 messages collected, 1,252 of the messages were selected by a random sample of the data-feeds that had > 300 retweets to analyze with the SIR model. If an author tweets an idea or retweets an idea, that person is considered part of the infected class. The followers of those authors are considered part of the susceptible class. Finally, if you are not part of the Susceptible or Infected class, you are part of the recovered class by utilizing Eq. (3.1) over time [53], [56].

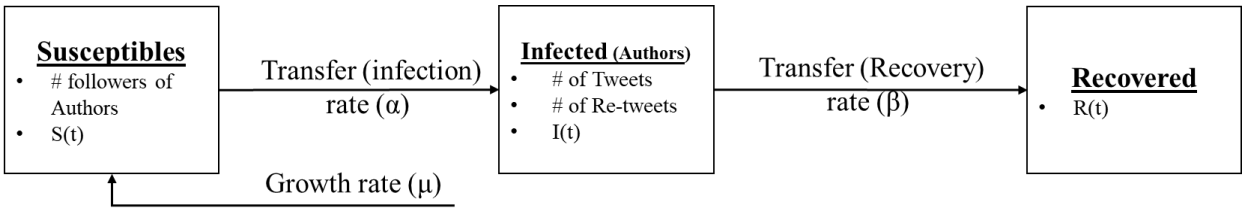

Figure 3.4 Application of the SIR model to Social Media (Twitter).

We use Bayes theorem to couple the effects of the SIR model metrics with the sentiment of the message [71], [76]:

$$P(Metric = med|sentiment) = \frac{P(sentiment|SIR\ Metric=med) \cdot P(SIR\ Metric=med)}{P(sentiment)} \quad (3.6)$$

Here, $P(SIR\ Metric = med|sentiment)$ is the posterior probability of the SIR metric under

consideration having a value within the median histogram bin.

- $P(sentiment)$ is the probability the message has the sentiment and is taken from table S1.
- $P(SIR\ Metric = med)$ is the prior distribution which is equal to the probability of the bin based upon the histogram of all the messages. For example, in figure 3.2, this would be taken from the left histogram based upon the probability of being in the median bin. For followers, this was equal to 26%.
- $P(valence|SIR\ Metric = med)$ is the likelihood probability. The likelihood probability is determined by the probability of the bin of the histogram by sentiment. For example, in Figure 3.4, this would be taken from the middle histogram based upon which sentiment you are analyzing for the bin that corresponds to the median value. For positive messages this was equal to 22%.

From our data set we calculated the posterior probability for all the SIR metrics given the sentiment to include infection rate, infectious period, reproductive number and the total number of followers. Additionally, we calculated the probability that valence influences the message following epidemic behavior using the calculated R^2 of the SIR model to the twitter data. Results of all posterior probabilities is found in Table 3.4.

Table 3.3 Lists the $P(sentiment)$ values to use with Bayes theorem calculations. These values were determined by dividing the number of messages of a given sentiment by the total number of messages.

	Negative	**Positive**	**Neutral**
$P(sentiment)$	45%	14%	41%

Table 3.4 lists the data collected from Twitter that we analyzed using the SIR model. Three data-feeds with 333 messages each were collected by the University of Oklahoma, The Center for Risk and Crisis Management (CRCM). These data-feeds are "Nuclear, "Drones," and "Weather" [77]. The other data was collected via the twitter application programming interface. 100 messages were collected based upon key words during the impeachment of President Donald Trump. 100 messages were from key words during the COVID-19 pandemic. 52 messages were collected from 5 different topics that include politics, business, military, sports and apologies [57]. For this analysis, the minimum threshold for a tweet to be collected was that it had to be retweeted at least 300 times. In all, 1,252 total messages were analyzed that were retweeted over

1.5 million times.

Table 3.4 Lists the twitter data collected for this study.

Group	Data Collection	Number of messages	Time Period
Nuclear	16 Keywords/phrases	333	1 Jan – 31 Dec 2019
Drones	5 Keywords/phrases	333	1 Jan – 31 Dec 2019
Weather	9 Keywords/phrases	333	1 Jan – 31 Dec 2019
Coronavirus	1 Keyword	100	1 – 5 Feb 2020
Impeachment	2 Keyword	100	1 – 5 Feb 2020
Business	Various	12	2017 - 2019
Politics	Various	19	2017 - 2019
Military	Various	13	2017 - 2019
Sports	Various	8	2017 - 2019

3.4 Analysis of SIR metrics, Sentiment Analysis, and KS2 Test

Figures 3.5 – 3.10 show the analysis for each of the SIR metrics with sentiment. For each figure, the left histogram shows the overall distribution, the center histogram shows the distribution separated by sentiment, and the right figure shows the two sample Kolmogorov-Smirnov (KS2) test. The right figure uses the cumulative distributions for each sentiment to conduct a KS2 test. The results are summarized in Table 2.5 that lists which distributions are statistically the same and which are different. Negative and positive messages are statistically different for all SIR metrics except the infection rate. Therefore, the sentiment doesn't influence the infection rate but does for all other metrics. The results of the KS2 test indicate that sentiment, specifically negative vs. positive influence the spread of messages. This analysis is used to determine the

probability that sentiment has on the behavior of an idea spreading through social media.

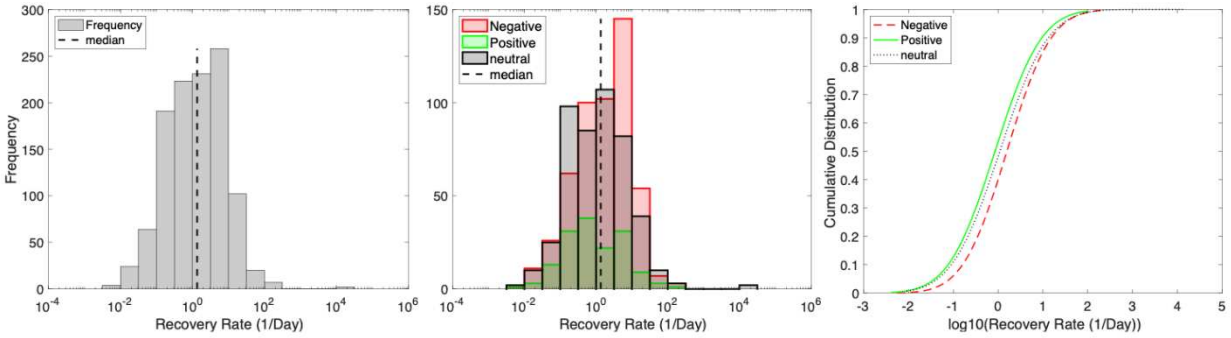

Figure 3.5 Distribution of recovery rate by sentiment. For each created histograms of the distribution of the values. The overall distribution is shown on left histogram. The middle histogram separates the messages by sentiment. The right figure shows the cumulative distribution of a best fit probability distribution for each sentiment histogram. Negative messages have the highest median recovery rate of 2.7 {1/day} compared to positive 0.6 {1/day} and neutral 1.1 {1/day}.

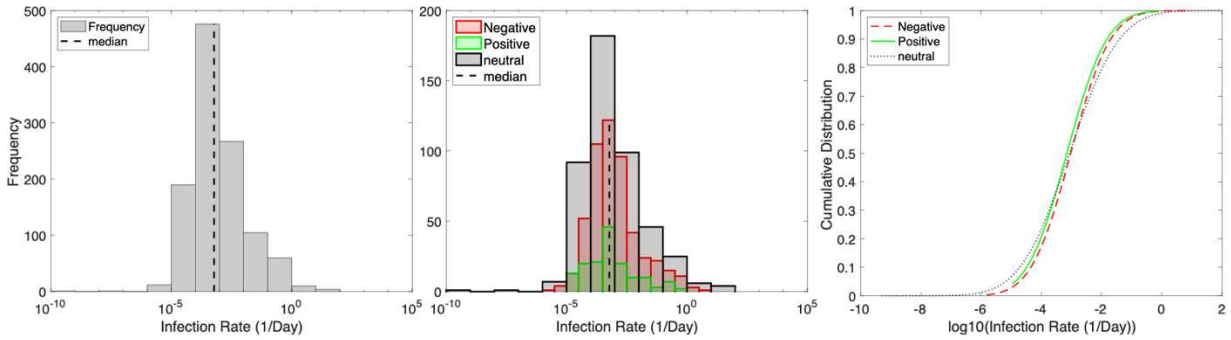

Figure 3.6 Analysis of the effect of sentiment on the spread of ideas on social media for the Infection Rate**.** For each created histograms of the distribution of the values. The overall distribution is shown on left histogram. The middle histogram separates the messages by sentiment. The right figure shows the cumulative distribution of a best fit probability distribution for each sentiment histogram. Table 3.5 lists the results of the KS2 test.

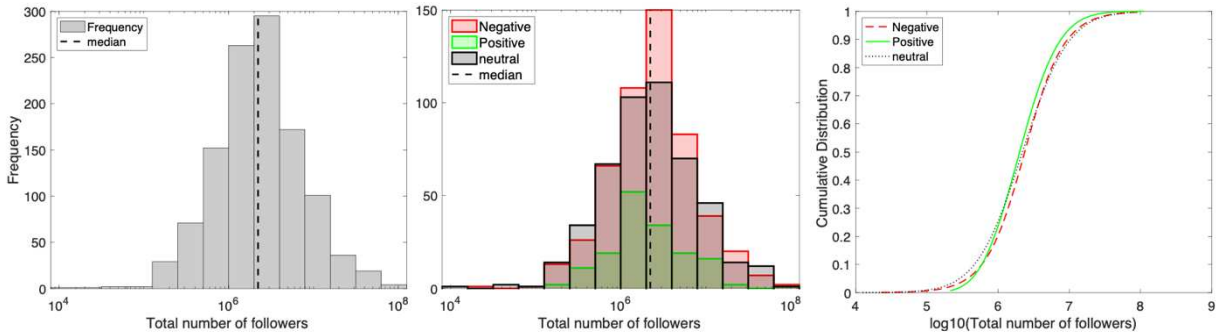

Figure 3.7 Histogram of total number of follower (left), histogram of total number of followers separated by sentiment (middle), and the cumulative KS2 test (right) for total number of followers separated by sentiment. The histograms and the KS2 tests facilitated the understanding of how the total number of followers are different for different sentiment messages. Table 3.5 lists the results of the KS2 test.

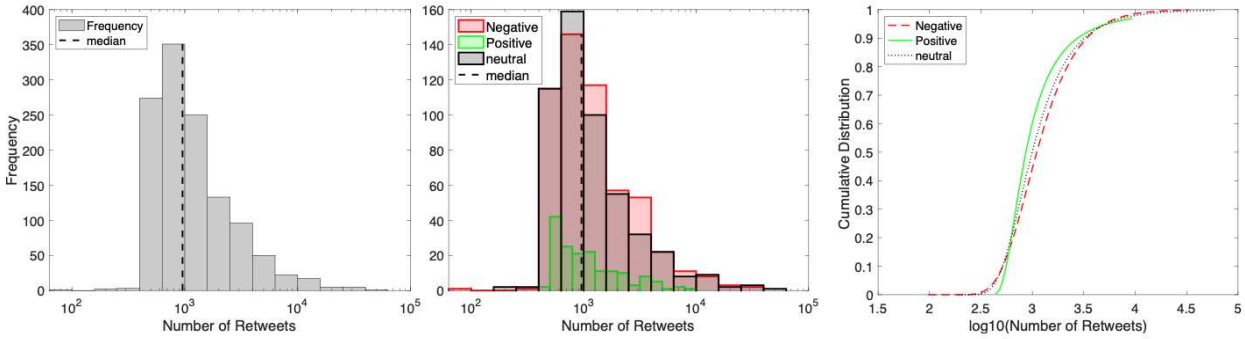

Figure 3.8 Histogram of number of retweets (left), histogram of number of retweets separated by sentiment (middle), and the cumulative KS2 test (right) for number of retweets separated by sentiment. The histograms and the KS2 tests facilitated the understanding of how the number of retweets are different for different sentiment messages. Table 3.5 lists the results of the KS2 test.

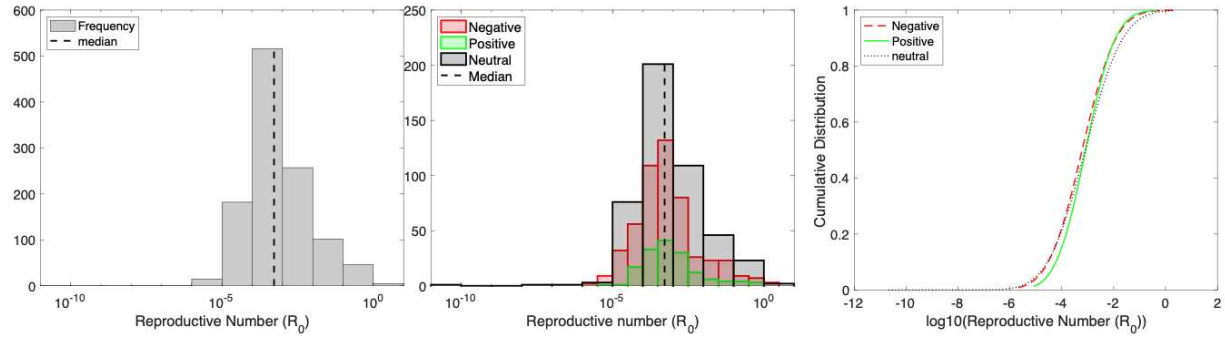

Figure 3.9 Histogram of the Reproductive number (left), histogram of Reproductive number separated by sentiment (middle), and the cumulative KS2 test (right) for Reproductive number separated by sentiment. The histograms and the KS2 tests facilitated the understanding of how the Reproductive number's are different for different sentiment messages. Table 3.5 lists the results of the KS2 test.

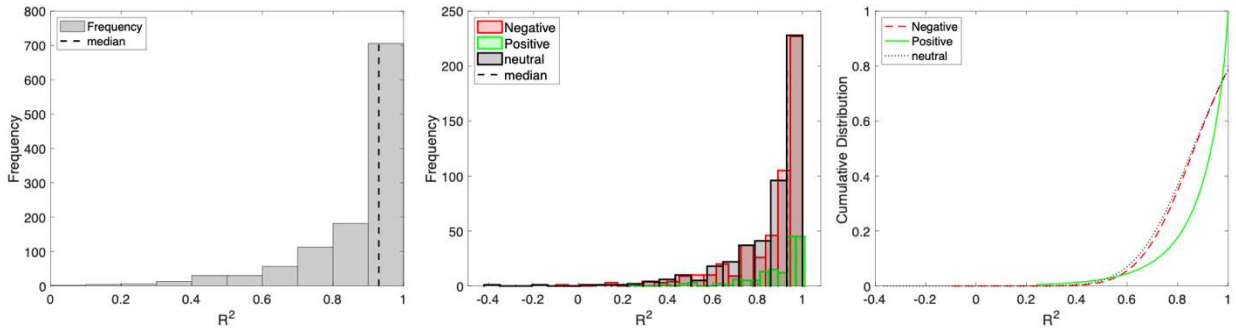

Figure 3.10 Histogram of the R^2(left), histogram of R^2 separated by sentiment (middle), and the cumulative KS2 test (right) for R^2 separated by sentiment. The histograms and the KS2 tests facilitated the understanding of how the $R^2$are different for different sentiment messages. Table 3.5 lists the results of the KS2 test.

Table 3.5 Lists the results of the KS2 test to determine which distributions are different based upon sentiment.

	Comparison	**KS2 Result**	**D statistic**	**p-value**
Recovery Rate	Positive - Negative	Different	0.21	$4 \cdot 10^{-5}$
	Negative – Neutral	Different	0.14	$9 \cdot 10^{-5}$
	Positive - Neutral	Different	0.11	$3 \cdot 10^{-4}$
Infection Rate	Positive - Negative	Same	0.10	0.20
	Negative – Neutral	Different	$9 \cdot 10^{-2}$	0.02
	Positive - Neutral	Same	0.07	0.43
Reproductive Number	Positive - Negative	Same	0.11	0.09
	Negative – Neutral	Same	0.08	0.10
	Positive - Neutral	Same	0.07	0.48
Followers	Positive - Negative	Different	0.15	0.01
	Negative – Neutral	Same	0.08	0.09
	Positive - Neutral	Same	0.11	0.09
R^2	Positive - Negative	Different	0.10	1E-11
	Negative – Neutral	Same	0.06	0.14
	Positive - Neutral	Different	0.78	1E-15
Retweet	Positive - Negative	Same	0.09	0.29
	Negative – Neutral	Same	0.06	0.25
	Positive - Neutral	Same	0.03	0.82

CHAPTER FOUR

BAYESIAN DATA SYNTHESIS IMPROVES DEPTH OF BURST ESTIMATES FOR NUCLEAR TESTS AT DENIED ACCESS SITES

Anthropogenic radioisotopes are the only definitive signature that a below ground nuclear test has been conducted. The depth at which a test occurs is known to be a critical parameter, affecting not only the migration time for gases to reach the surface but also their subsequent isotopic ratios. Suspected tests typically occur at sites to which the international community has no access, and their suspected location and depth must be computed. Current methods for doing this have a resolution of a few kilometers at best. Here we show that Bayesian data synthesis can improve depth of burst estimates by considering local topology, geology, the presence of surface deformation, yield, and a safety factor (for US tests). The method is tested against actual depth of burst data from 36 Soviet tests and 14 in the US. Bayesian data synthesis reduces the mean error to less than 53m.

4.1 Introduction to depth of burst of an underground nuclear explosion.

The last above ground nuclear test was carried out by China in 1980 [78]. In today's international climate, a country that conducts a nuclear test is likely to do so below ground and at a site to which the international community has no access. This would prevent direct measurements that could be used to both confirm that a nuclear test has taken place, and also to establish the characteristics of the explosion and the device that produced it. In cases such as these, remote sensing techniques come into play and are central to the monitoring done by countries and the Comprehensive Nuclear-Test-Ban Treaty Organization [79], [80].

Nuclear explosions produce shock waves, sound, direct radiation, electromagnetic pulses as well as fission products [81]. However, only the production of anthropogenic radioisotopes is considered to be a definitive signature of a nuclear test. Because of their ability to move through geological structures, noble gas radioisotopes play a key role [82], [83] in nuclear detection. Studies with nonradioactive tracers have shown that heavy gases released below ground will diffuse to cracks along which they then move to the surface. The rate of movement within cracks is itself strongly affected by both their width and variations in barometric conditions at ground level [84]. In this way, local geology and atmospheric conditions combine to affect the rate at which gases are transported to the surface. Because of decay, this can then affect the ratios of radioxenon isotopes that are used to distinguish between a nuclear test and other

anthropogenic sources such as nuclear reactors or reprocessing facilities [85], [86]. The depth of burst is a particularly important parameter here as it affects the time required for radioactive gases to reach the surface as well as the expected ratios of radioxenon isotopes.

The timeliness, accuracy, and level of effort required to estimate the depth of burst for a suspected nuclear test is important for international decision makers. The time it takes a radioactive gas to reach the surface will dictate how quickly a team from the Comprehensive Nuclear-Test-Ban Treaty Organization has before it must be on-site or taking air samples [80], [87]. If depth is underestimated or overestimated, the on-site inspection team will arrive at the wrong time and the window for a measurable signal may be missed. If the depth of burst estimate arrives too late for decision makers, it becomes useless.

Common techniques for estimating Depth of Burst can be inaccurate and labor intensive. Seismology is typically used to determine the region in which a suspected test has taken place but can also be used to gauge depth. Here, the travel times for seismic waves measured at multiple monitoring stations are used to converge on a location [88]. A simple analysis of differences between measured P and S waves can be used to classify a burst [89] as originating in a surface or a non-surface region. However, "surface origin" is typically defined as within ~10 kilometers of the actual surface [89]–[92]. Newer 'array processing' techniques can improve the depth estimate that can be had with seismology. Here primary and surface reflected P waves, as well as surface reflected S waves that convert to P waves, are used to determine the depth of burst. Differences in the arrival times of the latter two depend on the depth of burst and are used to improve estimates of it. Even so, predictions are only to within a few kilometers of actual depths [93]. Why is there a difference in the arrival times of these waves?

Site monitoring can also be used to estimate the depth of burst. This approach includes gathering imagery of the underground facility to look at construction activity, spoil, and movement [94], [95]. This process may not be possible if indicators are not present to start the site monitoring activities. Site monitoring has the possibility of missing key evidence if the area under surveillance is large. The possibility could lead to systematic error with a depth estimate.

Nuclear tests are often conducted in mountainous or hill regions. Here horizontal drilling is typically used to place a device. Horizontal tunnels allow for better measurement of explosion effects testing, specifically on missile silos and underground command centers, and also ease the process of emplacing devices [96]. In mountainous or hilly test regions the difference between

the surface and base elevations within a region can be used to estimate the possible range for a test's depth of burst. While this is a common method of analysis [97], the resulting range can still be large, particularly in mountainous regions.

A method for estimating the *Minimum Depth of Burst* at which a device would need to be buried for a test to remain contained was developed by the Joint Soviet-American Experiment on Verification program [98]. The intent here was to ensure that underground nuclear tests would be conducted deeply enough for them to comply with the Limited Test Ban Treaty. Here, the *Minimum Depth of Burst* was defined as:

$$\text{Minimum Depth of Burst} = C\, Y^{1/\alpha} \tag{4.6}$$

where C is a constant of proportionality, Y is the yield (kilotons) and α is a scaling parameter. The Soviets used the expected yield in Eq. (4.1) while the US used the 'maximum credible yield' to better ensure that a test was contained. The scaling parameter, α, was taken as 3 by the US and 3.4 by the Soviets [96], [99]. The yield itself can be estimated by measuring the amplitude and period of the seismic wave [98]–[100]. With different reported body wave magnitudes from separate seismic stations, a distribution of the estimated yield is created. The constant C was originally given as a range which depended only on whether or not surface deformation from the test was evident [101], [102]. Teller (1968) [103] refined the estimated ranges based on the type of geology in which the test took place. Later in its testing period the US set C = 122m [96].

At Soviet test sites, devices were typically buried no deeper than the *Minimum Depth of Burst* [99] possibly because of the added expense of going deeper [104]. However, for US tests Eq. (4.1) was used as a guide with other factors also taken into consideration [96]. In both cases, a version of Eq. (4.1) can be used to narrow the possible range for the actual Depth of Burst while still allowing for the possibility that a device could be placed much deeper. Here we show that Bayesian data synthesis can improve depth of burst estimates by combining seismic data with local geology, topology, imaging (to determine presence of surface deformation) and estimates for the *Minimum Depth of Burst*. The approach is tested against data provided by the Joint Soviet-American Experiment on Verification program, as well as tests conducted at the Nevada Test Site. With the continued proliferation of nuclear weapons, the necessity to revisit the characterization of underground nuclear explosions at a denied site is required to hold rogue nations accountable.

4.2 Methods for estimating the depth of burst.

Figure 4.1 summarizes both the data used to estimate the actual Depth of Burst for a below ground nuclear test as well as the order in which it is processed. Open data sources are used to generate distributions for yield estimates, and the latitude and longitude of a suspected test [105], [106]. From this, a distribution for terrain elevation is generated, with satellite imagery giving an indication as to whether surface deformation occurred in the test. An initial estimate for the Depth of Burst is then obtained using Eq. (4.1), or the US form of it. The location estimate is used in conjunction with global geology maps to estimate local geology. The Depth of Burst estimate is then refined using Bayesian data synthesis to fold in how C changes with rock type and the effect of local terrain elevation [76]:

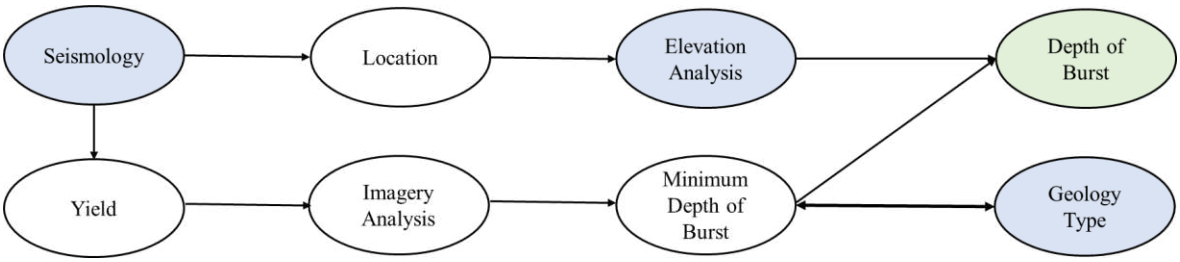

Figure 4.1 Process to estimate the depth of burst of an underground nuclear explosion. Depicts the process to use to estimate the depth of burst. The blue nodes indicate data required to make an estimate, white nodes indicate an intermediate estimate, and green nodes indicate the final estimate of depth of burst.

Bayes theorem provides a method for understanding the probability of an event given a known conditional probability of another event. For continuous probabilities, Bayes theorem can be written as [76].

$$P(\theta|data) = \frac{P(data|\theta)\cdot P(\theta)}{P(data)} \quad (4.2)$$

Here θ is the Depth of Burst, $P(\theta|data)$ the posterior distribution which gives the probability distribution for the Depth of Burst, $P(data|\theta)$ is the 'likelihood distribution,' $P(\theta)$ is the prior distribution (what the depth was initially thought to be before 'data' was considered), and $P(data)$ is the marginal probability.

Prior Distribution P(θ). Once an underground explosion happens, its shock wave will be detected at seismic stations that surround the explosion [107]. The arrival times of the shock waves can be used to triangulate a location area, and Bayesian Seismic Locator (Bayesloc) was

used to do this [106]. The inputs for Bayesloc are the seismic arrival times for the shock waves from seismic stations that surround the explosion, and seismic stations closest to the test site are preferred as are ones with a full 360-degree coverage [108]. For tests in the Joint Soviet-American Experiment on Verification program, seismic arrival times were taken from the international seismology center online at http:www.isc.ac.uk [105], [109].

Bayesloc gives a latitude and longitude distribution for a suspected test. The location area is defined as a rectangular box centered at the mean of the latitude and longitude distributions. The height and widths are taken to be 1.4 standard deviations away from the mean latitude and longitude. This was found to correspond to the 95% confidence interval for the actual test being within the location area, Fig. 4.2.

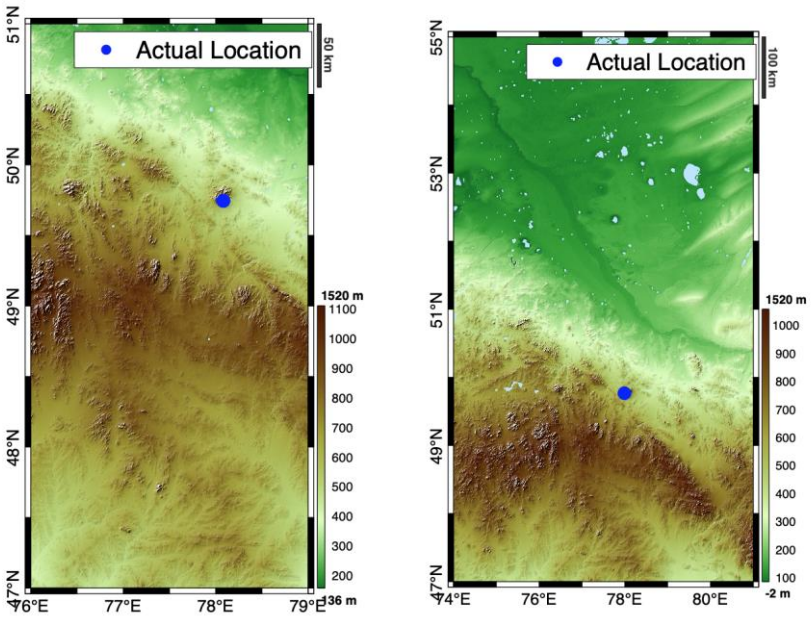

Figure 4.2 Estimated locations for three tests in the Joint Soviet-American Experiment on Verification program. The location area is determined by using the results of Bayesloc. The center is the mean of each distribution and the height and widths are 1.4 standard deviation from the mean, respectively, which gives the 95% confidence interval for the test having been within the specified region. The blue dot shows the actual ground zero location. [98], [99], [110]

Elevation data with 30m spatial resolution within suspected test regions was obtained from the United States Geological Survey, which has elevation data for the entire world available at https://dds.cr.usgs.gov/srtm/version1 [111], [112]. A test can then be taken as occurring anywhere between the surface and base elevation:

$$D(x,y) = Surface\ Elevation(x,y) - Minimum\ Elevation \quad (4.3)$$

where *D(x,y)* is a matrix of possible depths within the location area. Here, *D(x,y)* is taken over the latitude (x) and longitude (y) coordinates within the suspected test region at 30m spatial resolution and '*Minimum Elevation'* is the lowest Base Elevation within the location area, Fig. 4.3.

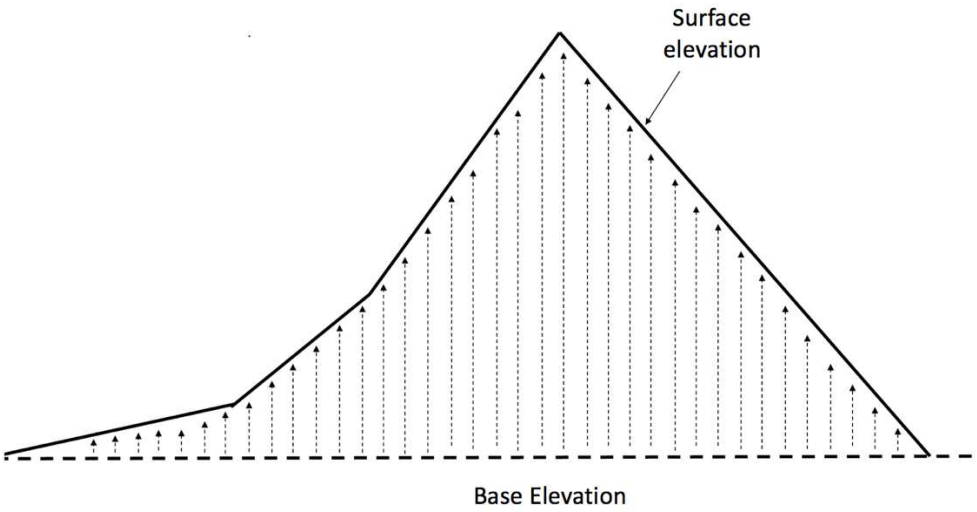

Figure 4.3 Terrain analysis to get a prior distribution for Depth of Burst. Subsurface tests are often done in regions where horizontal drilling would allow easy placement of a test device. A first estimate for the Depth of Burst, P(θ), is the distribution for the difference between the Surface Elevation and the Base Elevation within a suspected test region.

The results for *D(x,y)* are binned using Sturges' method to create a histogram [113], [114], and the maximum likelihood estimation is used to fit a distribution to the histogram. Matlab has 17 built in probability density functions that can be used to fit to data and the best distribution according to the Akaike Information Criterion [115] then becomes the prior distribution, *P(θ)*, for use with Bayes theorem, i.e. Eq. (4.2). This prior gives an overall assessment of the topology of the surrounding area in which the explosion took place and thus an idea of what the possible depths could be. See Supplemental Information, Fig. 4.5 for example histogram and distributions for the prior distribution.

Likelihood Distribution $P(data|\theta)$. The likelihood distribution is developed using Eq. (4.1) and data on estimated yields:

$$Y = 10^{\frac{(M_b - \gamma)}{\beta}} \tag{4.4}$$

Here M_b is the body wave magnitude, Y is yield and γ *and* β are seismic station specific constants. Following an underground explosion, many seismic stations calculate the body wave magnitude and publish them at the international seismic center online at http:www.isc.ac.uk

[105]. Body wave magnitudes are recorded for each seismic station that publishes a value and therefore you will have "n" number of body wave magnitudes labeled as $M_{b1}, M_{b2}, \ldots M_{bn}$. Through testing conducted in the United States, China and the former Soviet Union, the values of γ and β ranged in value from 3.91 – 4.35 and 0.75 and 1.0 respectively [98], [99], [102]. The number of seismic stations (n) ranges from 7 to 18.

The yield distribution is created by Monte Carlo sampling. For each Monte Carlo iteration, a random value for the body wave magnitude is selected from one of the "n" stations, along with a random value of γ and β assuming uniform distributions.[114]. Here, γ comes from $U(3.91, 4.35)$ and β from $U(0.75, 1.0)$. Sturges' method is used to bin the results from 1,000 iterations and create a histogram of the yield results [113]. Maximum likelihood estimation is again used along with the Akaike Information Criteria to choose the best of Matlab's 17 distributions. See Supplemental Information for histograms and distributions for the yield estimation.

From the yield distribution, a distribution for the *Minimum Depth of Burst* is determined using Monte Carlo sampling with Eq. (4.1). For the 14 US tests considered here, an 'announced yield' was reported which we take here as the maximum credible yield. When geology is ignored, two possible ranges of the constant *C* exist. The first range is $C \sim U(55, 80)$ and the second range is $C \sim U(80, 110)$. The first range is selected if the depth of the underground explosion was shallow and surface deformation exists. The second range is used if the depth is deep enough and no evidence of surface deformation exists. The surface deformation is analyzed through imagery analysis within the location area [96]. If the local geology is alluvium, tuff, dolomite, or granite, the range for the constant *C* can instead be drawn from Table 4.1. See Supplemental Information for the histogram for the minimum depth of burst distribution (Fig. 4.6). The minimum depth of burst is the likelihood distribution.

Data Synthesis. Equation (6) is used to create a distribution for the estimated *Depth of Burst*. Here *P(data)* takes the form:

$$P(data) = \sum P(data|\theta) \cdot P(\theta) \tag{4.5}$$

Combining Eqs. (3.2,3.5) gives:

$$P(\theta|data) = \frac{P(data|\theta) \cdot P(\theta)}{\sum P(data|\theta) \cdot P(\theta)} \tag{4.6}$$

Supplemental Information, shows two example results from an updated depth of burst estimate using Bayes theorem.

Table 4.1 Range values for the constant C in the Eq. (4.1).

Rock Type	Constant C for Surface Deformation (50 – 80)	Constant C for No Surface Deformation (80 – 110)
Alluvium	64 – 76	94 – 106
Tuff	74 – 77	104 – 107
Dolomite	51	81
Granite	57 – 60	87 - 90

Statistical Analysis for depth of burst estimates. Statistical analysis is used to test the accuracy and precision of the methodology. The following is a summary of the statistics used to compare to the actual depth of burst values reported through the Joint Soviet-American Experiment on Verification program. The above methodology results in 1 to 3 estimates of the depth of burst. The first estimate is using the elevation analysis (EA) only. The second method is updating the EA with estimating the yield. This method is referred to as EA + Yield. The final updated depth of burst estimate is with a known geology type. This estimate is referred to as EA + Yield + Geology.

The first statistic gives the distance of the mean of the posterior distribution to the actual value. Eq. (4.7) shows the formula to calculate the average difference:

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_i) \tag{4.7}$$

Where x_i is the mean of the posterior distribution of each test, $\bar{x}_i$ is the true depth of burst for each test, and n is the number of tests.

Additionally, a p-value is calculated to determine if a null hypothesis is valid. The null hypothesis is that the mean of the distribution is equal to 50 meters from the true value, which is approximately equal to the standard error of terrain analysis. Therefore, the alternative hypothesis is the mean of the distribution is less than 50 meters from the true value. To calculate the p-value the standard deviation and t-statistics must be calculated. Eqns. (4.8) and (4.9) are used to calculate the standard deviation and t-statistic, respectively:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}{n}} \tag{4.8}$$

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} \tag{4.9}$$

Where μ_0 is the null hypothesis. The p-value assumes a t – distribution of the differences of the mean of the distribution and the actual values. The p-value provides the probability the

difference is 50 meters. If the probability is lower than 0.05, then the alternative hypothesis is accepted and the mean is < 50 meters from the true value.

Other statistics to show the accuracy and precision are the standard error and probable error. Eqns. (4.10) and (4.11) show the formulae for standard error and probable error, respectively. The standard error and probable error are two more measures of how close the mean is to the actual depth of burst values:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x}_l)^2}{n(n-1)}} \tag{4.10}$$

$$r = 0.6745 \cdot \sigma_m \tag{4.11}$$

Validation. The former Soviet Union (USSR) detonated over 500 nuclear weapons underground during the Cold War [101]. Through the Joint Soviet-American Experiment on Verification program data on *Depth of Burst*, yield, location and a general geology description is given for 96 underground tests conducted between 1961 and 1972 [116]. However, no seismic data was available for tests before 1966. This excluded 57 tests and 2 additional tests were removed because of multiple reported depths of burst. For the United States test, 14 tests were used where seismic data for both location and body wave magnitude were included in the International Seismic Center database along with actual data of the depth of burst [105], [117].

4.3 Results and discussion.

Risk tolerance and technology capability are important considerations when estimating the Depth of Burst. The Soviets used the design yield in Eq. (4.1) and an $\alpha = 3.4$. However, the Soviet Union and the US differed in their approach to the depth at which they buried their nuclear devices. From 1963 – 1970, the United States contained its nuclear explosions and its subsequent radioactive releases within the Nevada Test Site. In 1970, the test codenamed "Baneberry" resulted in an immediate release and the radioactive material was tracked as far as the Canadian border [96]. This event prompted the United States to take extra caution in containing the underground nuclear explosions. The United States made safety changes to how it selected a Depth of Burst [96]. This included using the "maximum credible" in siting as well as setting C = 122m in Eq. (4.1) [96]. However, the US also used the minimum depth of burst as a guide and other factors were taken also into consideration [96]. The result was that different US weapons labs were noted to differ in how deep they placed their devices [104].

Tables 4.2 – 4.3 give a summary of predicted Depth of Burst for Soviet tests conducted at Semipalatinsk, Kazakstan and US tests conducted at the Nevada Test Site. The results for the Soviet tests used yield estimates from multiple seismic stations. There is a clear decrease in the mean difference between predicted and actual depths of burial as more information is brought to bear. Pure elevation analysis gives a mean difference of -219m whereas elevation analysis plus the effect of a minimum depth of burst, with geology considered, reduces that to -53m. The standard deviation also goes down considerably. A z-test shows that the distribution for latter predictions are the same as the distribution of actual Depth of Burst values (p=0.025).

Table 4.2 Comparison of predicted and actual depths of burst for Soviet tests. The first column (Elevation Analysis) includes statistics from the distribution created through Eq. (4.3). The second column (EA + Yield) includes statistics from the distribution created through Eq. (4.6) with unknown geology type. The last column (EA + Yield + Geo) includes statistics from the distribution with the constant C from Table 4.1.

Estimate Type	**Elevation Analysis (EA)**	**EA + Yield**	**EA + Yield + Geo**
Mean of the Differences	-219m	-56m	-53m
Mean of the Distribution Standard Deviation	219m	74m	72m
Average z-score	1.17	1.65	1.11
p - value	0.95	0.028	0.025
Average 95% CI	511m	246m	159m
Standard Error	52m	18m	17m
Probable Error	35m	12m	11m

The situation is different with the 14 tests summarized in Table 4.3. Here elevation analysis by itself gives a mean difference between predicated and actual depths of burst of 134m. However, factoring in the measurements of the yield from seismic observations, and geology, make the predictions worse. None of the three approaches shown has a distribution that is statistically similar to the actual values. The issue here is that the depth of burial for US tests was based on the maximum credible yield that a device could realistically be expected to have. This would be higher than the design and very likely higher than the measured yield as well.

The data in Table 4.3 use the maximum yield estimated from seismic observations, even so are well outside the actual depths of burst.

Table 4.3 Comparison of predicted and actual depths of burst for United States Tests. The first column (Elevation Analysis) includes statistics from the distribution created through Eq. (4.3). The second column (EA + Yield) includes statistics from the distribution created through Eq. (4.6) with unknown geology type. The last column (EA + Yield + Geo) includes statistics from the distribution with the constant C from Table 4.1. The yield used here was the maximum yield estimated from seismic observations.

Estimate Type	**Elevation Analysis (EA)**	**EA + Yield**	**EA + Yield + Geo**
Mean of the Differences	134m	-169m	-158m
Mean of the Distribution Standard Deviation	245m	117m	82m
Average z-score	1.9	2.1	3.7
p - value	0.92	0.93	0.93
Average 95% CI	909m	502m	607m
Standard Error	147m	93m	89m
Probable Error	98m	62m	59m

Table 4.2 and 4.3 lists the statistical analysis for estimating the results using the Bayesian analysis. The results from the Soviet tests are more accurate but the variance is larger for the Soviet Union tests. The better accuracy with the Soviet tests is based upon its compliance with the limited test ban treaty. The larger variances for the Soviet Union tests were likely due to the tests being conducted in a more remote location and the location areas were larger for these tests. The primary difference in accuracy is largely due to an assessment of risk levels for the country conducting the test. The United States had a lower risk tolerance for venting and safety was its primary concern [96]. The stated goal was to prevent venting at the surface of the test location. While the Soviet Union stated the objective was to prevent elevated radiation levels above background radiation from leaving its nation's borders which is simply staying compliant with the Limited Test Ban Treaty [98], [118].

Data on the 14 US tests used here also gave an 'announced yield' which in every case was much higher than the values predicted from seismic observations. Using this as an estimate of the maximum credible yield gives the results in Table 4.4. Here again the combination of

elevation analysis, minimum depth of burst and geology give the best result. The mean difference between predicted and actual Depth of Burst falls to -22m with the predicted and actual depth distributions being statistically the same (p=0.04). While a country conducting a clandestine nuclear test is unlikely to give an announced yield, CTBTO countries may well be able to do their own estimate for the maximum credible yield. This would then give two measures for estimated depth of burial. One based on yields measured using seismic observations, and one based on an estimated maximum credible yield.

Table 4.4 Effect of announced yield on predicted and actual depths of burst for United States Tests. The first column (Elevation Analysis) includes statistics from the distribution created through Eq. (4.3). The second column (EA + Yield) includes statistics from the distribution created through Eq. (4.6) with unknown geology type. The last column (EA + Yield + Geo) includes statistics from the distribution with the constant C from Table 4.1.

Estimate Type	**Announced yield with $122 \cdot Y^{1/3}$**	**EA + Yield**	**EA + Yield + Geo**
Mean of the Differences	84m	107m	-22m
Mean of the Distribution Standard Deviation	23m	25m	18m
Average z-score	0.5	4.4	5.1
p - value	0.8	0.88	0.04
Average 95% CI	25m	214m	189m
Standard Error	58m	60m	42m
Probable Error	39m	40m	28m

When 71 nations signed the limited test ban treaty in 1963, part of the objective was to prevent radioactive venting [118]. The nuclear weapon states (primarily the United States and the Soviet Union) conducted empirical tests and relationships to determine how deeply a device should be buried [96], [98], [99]. However, Depth of Burst turns out also to be important for treaty verification. It is the most important parameter in predicting when anthropogenic gases will arrive at the surface after a below ground nuclear test, and what their isotopic composition will be. The method described here for predicting how deep a device was buried for a test shows remarkable accuracy when the yield required in Eq. (4.1) can be estimated. For Cold War era

seismic technology, Bayesian synthesis of terrain, yield and geological data gave a mean error of < 50 meters and will aid the CTBTO in its monitoring efforts.

Acknowledgements. Special thanks to John McClory for help in finding data and discussions of its use, to Jenifer Shafer for suggestions on test location analysis and to Andrew Osborne for editorial comments and suggestions.

4.4 Note 1 for depth of burst

Figures 4.4 – 4.7 show results of the intermediate steps to estimate the depth of burst at a denied access site. The process utilizes Bayes theorem (Eqn. 4.7). The intermediate steps include estimating a prior distribution through terrain analysis, and a likelihood distribution using a form of Eq. (4.1).

Prior distribution: Figure 4.4 shows results of the prior distribution ($P(\theta)$) for two tests.

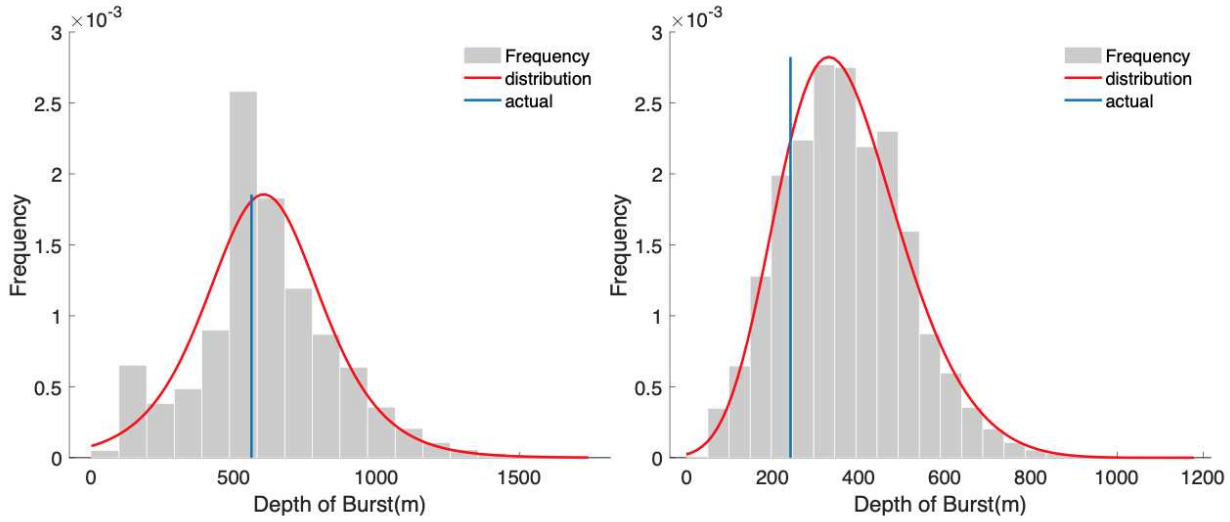

Figure 4.4 Example result for elevation analysis depth of burst estimate which is the prior distribution in Bayes theorem. The histogram is on the left is from a test conducted on March 26, 1970 at the Nevada Test Site. The histogram on the right is from a test conducted January 30, 1967 in Kazakhstan. The elevation analysis distributions are the result of fitting a distribution to the histogram for each elevation data point used in Eq. (4.2). The vertical blue line indicates the actual depth of burst provided by the Joint Soviet-American Experiment on Verification program [96 test].

Likelihood Distribution: The likelihood distribution is estimated using Eq. (4.1) [98], [99], [110]. We must first estimate the yield distribution through seismology and the body wave magnitude from various seismic stations that surround the underground nuclear explosion.

Figure 4.5 shows examples of yield estimations from nuclear tests conducted by the former Soviet Union using Eq. (4.5).

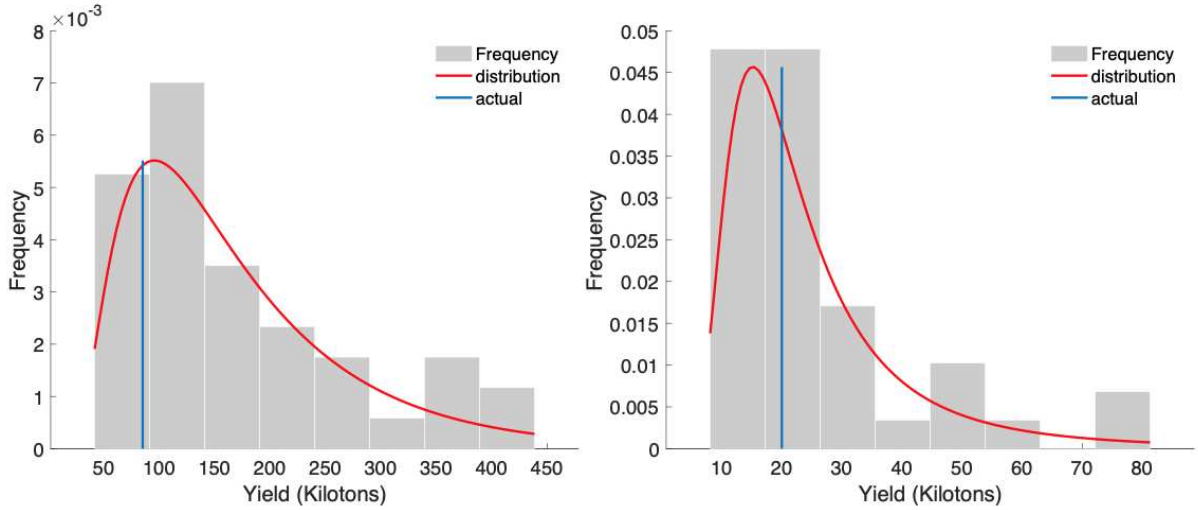

Figure 4.5 Example results for a histogram and yield distribution that is used in Eq. (4.4) to calculate the likelihood distribution. The gray boxes are the histogram created from the 1,000 iterations of calculating the yield according to Eqn. 4.5 through a Monte Carlo sampling technique.

Figure 4.6 depicts results of the likelihood distribution using the results of the yield (Fig. 4.5) and Eq. (4.1).

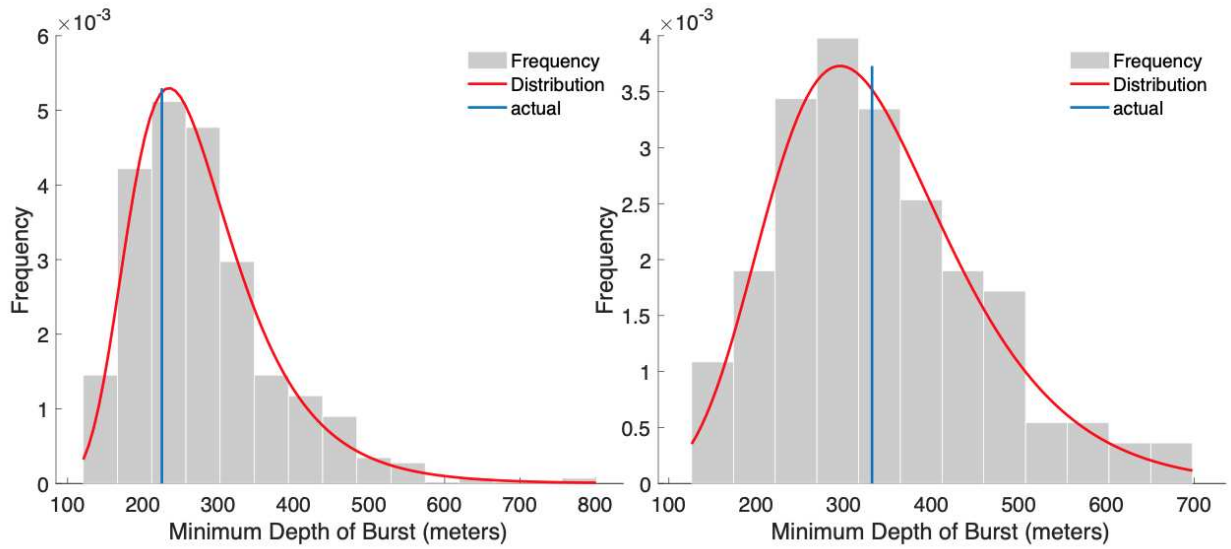

Figure 4.6 Example results for the likelihood distribution using Eqn. 4.1. Eqn. 4.1 shows a relationship between yield (figure 4.5) and depth of burst.

Figure 4.7 shows the results of the Bayesian analysis using the prior (Fig. 4.4) and

likelihood distributions (Fig. 4.6) with Eq. (4.7). The red distribution is the prior distribution, the blue dotted line is the likelihood distribution and the green distribution is the posterior distribution. The posterior distribution is what is used as the initial depth in a stochastic subsurface radioactive gas transport simulation.

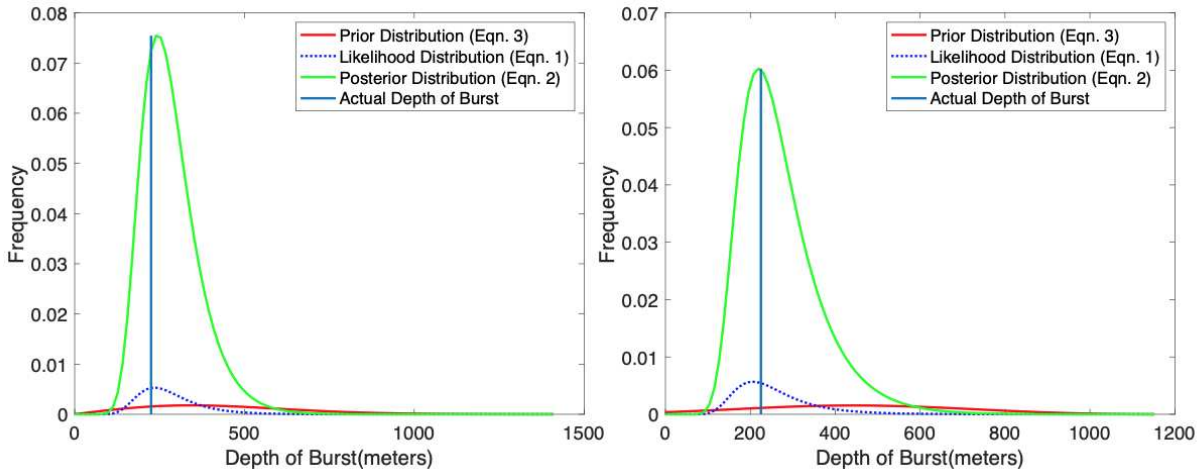

Figure 4.7 Example results for a depth of burst using Bayes theorem. The prior distribution is the distribution developed from figure S1 using elevation analysis. The likelihood distribution is developed through Eqn. (4.1) using the yield to depth of burst relationship. The vertical blue line indicates the actual depth of burst. The posterior distribution developed from Eq. (4.2) is the final depth of burst estimation.

CHAPTER FIVE

METHOD TO CHARACTERIZE THE GEOLOGY OF DENIED ACCESS SITES

Past work has shown that the depth at which a nuclear test occurs can affect not only the migration time for gases to the surface but also their isotopic ratios. Accurately modeling gas transport and isotopic fractionation requires knowledge of the geology. However, suspected tests are typically conducted at sites to which the international community has no access. Here we describe a method to characterize fracture width, spacing, tortuosity, permeability and porosity at a denied access site. Fractures are treated as fractals with their respective fractal dimensions determined using surface images. Tortuosity in the geological matrix (i.e. bulk media) is determined using rock class with the latter also being used to determine distributions for permeability and porosity. The approach is tested using areal images and rock class for the Nevada Test Site and validated by comparison of the results to field and core measurements. The method is then used to characterize the location of nuclear tests in the Democratic People's Republic of Korea.

5.1 Introduction to geology of a denied access site

Nobel gas radioisotopes are used by the international community to help determine if a clandestine nuclear test has taken place [119], [120]. The time it takes for these gases to arrive at the surface is affected by atmospheric conditions as well as the local geology. Variations in barometric pressure cause a pumping effect which can draw post detonation gases up and push them back down [84], [121]. Subsurface gases will migrate through the matrix and move by convection through fractures to the surface [121]. Modeling below ground transport requires estimates for the tortuosity, porosity and permeability of the media as well as the width, and spacing of fractures [86]. Unfortunately, clandestine nuclear tests invariably take place at sites to which the international community has no access which means that characterization must be done remotely [94].

Geological properties that affect subsurface gas transport are typically determined using direct measurements on cores from boreholes [122]. Even then, geology is rarely homogeneous, and many cores are needed to establish the statistical distribution for properties like porosity, permeability, fracture width and spacing. This can be costly and time consuming under normal circumstances, and impossible at a denied access site. A nation that has a nuclear weapons program and conducts a clandestine test is unlikely to allow foreign verification organizations,

like the Comprehensive Nuclear Test Ban Treaty Organization, access to characterize the site [123].

Current efforts to estimate geological parameters of denied access sites typically use subject matter expertise along with state sponsored reconnaissance platforms. With these resources, the process involves using satellite imagery and geomorphometric analysis methodologies. Additionally, geological maps of the area or adjacent area can be used to establish a geological history of the area [124], [125]. This method requires time, expertise and resources. The accuracy of the estimates depends on the assumptions made and judgements about the use of adjacent site comparisons as a proxy for unknown variables. A systematic method based on published data would provide points of references and potentially reduce error.

Multiple properties of porous systems have been shown to have fractal character including pore size distributions, fracture composition, and tortuosity [126]–[130]. In each case a scaling relationship applies [131]:

$$F = X^D \tag{5.1}$$

Here, F feature in question {length}, X is its frequency and D is the feature's fractal dimension. This same relationship also holds for fracture width where F is then the frequency of width X {millimeters}. Fracture composition, and tortuosity are particularly relevant to below ground gas transport as they affect convection to the surface and diffusivities [84], [86]. Methods based on lineation and pore size distributions can be used to estimate both [127], [129], [130].

Several global studies have been done that assigned a rock class to every location on Earth [132]. Durr et. al. (2005) [133] produced a global map at 0.5º x 0.5º resolution, Fig. 5.1. Hartmen et. al. (2012) [134] did a study of North America that sub-divided one of the rock classes into two. Both porosity and permeability can be estimated based on rock class. Gleeson et. al. (2011) [132] combined the 15 rock classes of Durr et. al. (2005) [133] into 6 and assigned a horizontal permeability distribution to each that is applicable to the near subsurface (< 100m). Kuang et al (2014) provided a method for extending these permeabilities to greater depth [135].

Conventional methods to characterize denied access sites are time consuming and relies on input that can be subjective. Here we show that the fractal analysis, and global maps for rock class can be used to characterize the geology of a denied access site. The methodology presented gives a systematic way to estimating geological properties that affect subsurface gas transport using vetted data that are within the public domain.

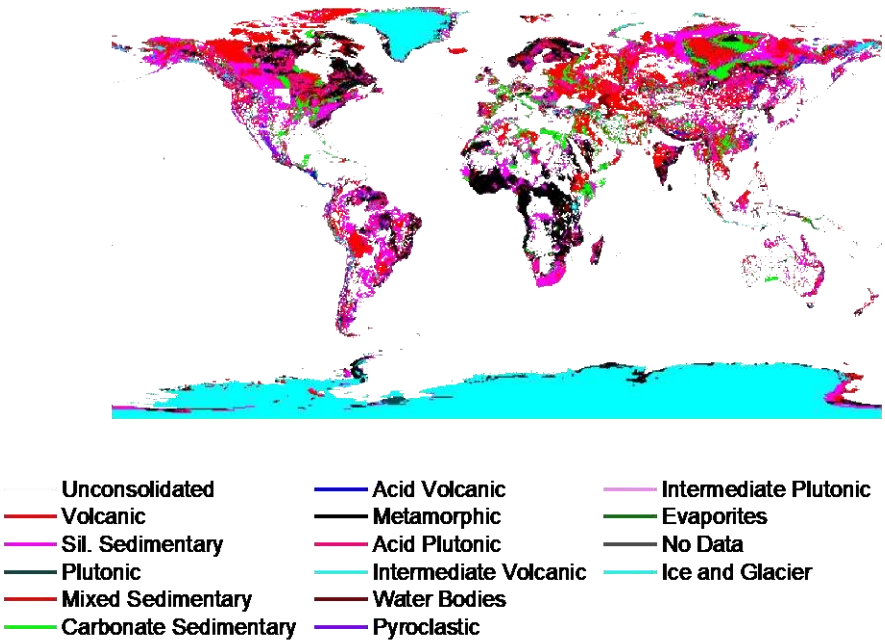

Figure 5.1 Depicts the rock classes for the planet earth [133]. The rock classes group different rock types into a general category to simplify and begin the process of predicting geological parameters such as permeability and porosity.

5.2 Method to estimate geological parameters.

The location of a suspected underground nuclear test can be identified using seismology [88], [105], [106]. Bayesian Seismic Locator (Bayesloc) is an example and uses the seismic arrival times of the shock waves to triangulate a location area [106]. The inputs for Bayesloc are the seismic arrival times which can be obtained from the international seismology center online at http:www.isc.ac.uk [105]. For best results, seismic stations closest to the test site are preferred as are ones with a full 360-degree coverage [105], [108]. Fracture spacing and tortuosity are determined from lineation's on images taken at different scales combined with regression analysis to measure the corresponding fractal dimensions [136], [137]. For fracture width, power law distributions are used that have been correlated with rock type to give the corresponding fractal dimension [130], [136], [138]–[142]. Estimates for permeability and porosity are made using worldwide geological maps and distributions based on rock class [133], [134], [143]. Tortuosity within matrix is estimated using published measurements pore space fractal dimensions for samples from different rock classes. The methodology for characterizing a denied access site is validated with a characterization of the Nevada Test Site and a comparison of the results with measurements made using samples from bore holes.

Fracture Spacing. The fracture spacing is measured using area imagery at 1:50,000, 1:25,000, and 1:15,000 scales, Fig. 5.2. This process was implemented using ArcGIS with world map imagery [144]. Overlays were created at each of the above imagery scales and then the lineation spacing analyzed on the overlays. Lineation's are any natural lines seen on imagery. For each lineation overlay a histogram is created for the spacing's between lineation's and then a distribution is then fit to the histogram[145]. Seventeen different probability distribution functions are fit to the histogram and the best fit according to the Akaike information criterion is selected [113], [115].

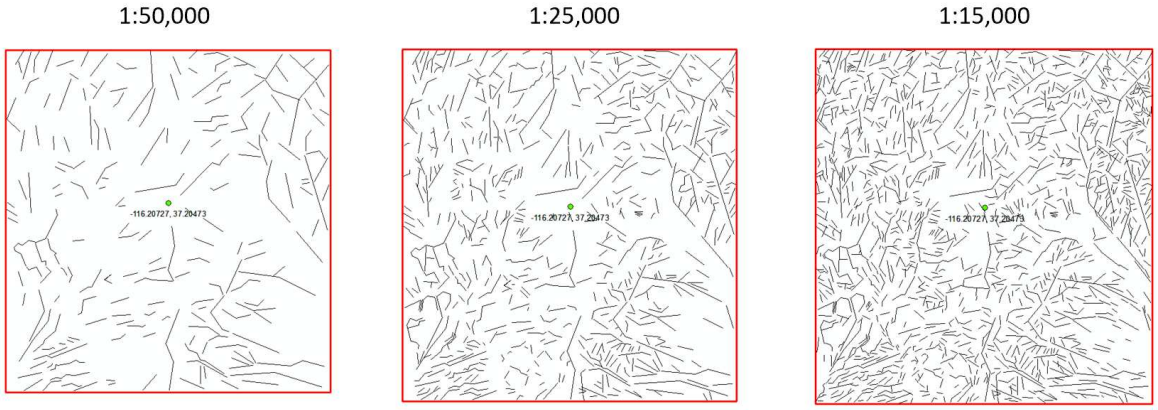

Figure 5.2 Lineation identification for the Nevada Test Site at three different scales. The lineation overlays were created in ArcGIS with world map imagery at different scales [144]. The scale of each imagery was from left to right 1:50k, 1:25k, and 1:15k. Each scale uses the imagery for the entire location area. As the imagery scale decreases additional lineation's are observed and this follows fractal behavior.

The mean and standard deviation from the fitted distribution is used to conduct a regression analysis and will be extrapolated down to the 1:1 scale to estimate a normal distribution for fracture spacing. A log-log plot of the natural log of the mean lineation spacing (y-axis) vs the natural log of the imagery scale (x-axis) is created. From this log-log plot, a linear best fit line is found. This linear best fit line is used to predict the value at the ground surface (x = 0 in the regression analysis) [129]. The fracture spacing fractal dimension is given by the exponential of the slope, D_{fs}. This process is repeated for the standard deviation. The fracture spacing fractal dimension is given by the exponential of the slope. See Supplementary Note 1 for additional detail.

Fracture Tortuosity. The fracture tortuosity is estimated by calculating the fractal dimension from the lineation overlays using the box counting method. The fracture tortuosity can then be estimated using [146]:

$$\tau = L^{D_{ft}} \tag{5.2}$$

where D_{ft} is the fracture tortuosity fractal dimension and L is the scale of the box dimension. Here, individual fractures in Fig. 5.2 are isolated and their fractal dimension determined to create a representative average for the location in question. See Supplementary Note 2 for additional details.

Fracture Width. Fracture width has been studied by rock type to identify the power law exponent that describes fracture width distributions in rock samples. Several studies identified a power law exponent by rock type [130], [138], [141], [142], [147]–[152]. Table 5.1 lists several of these studies that assign a power law exponent to a rock type. In each of these studies, several core samples were taken from a location area. These samples were taken to the lab and each rock fracture width was characterized. In each case, the associated fracture width distributions were best described by a power law distribution where the exponent is the fracture width fractal dimension, D_{fw}. It should be noted that the scale range for the power law may be limited based upon a limitation with the rock samples they used to develop the power law distribution [130]. Here the sample size may not represent the actual fracture widths of the geological rock type in terms of the applicable to the full fracture width distribution.

Each of the power law distributions by rock type are placed in the appropriate rock class (Table 5.1). For each rock class, the rock type exponents are averaged to use as the rock class fractal dimension (Table 5.1, column 4). For a denied access site, the weighted average of each rock class fractal dimension based upon percent area within the location area is used for a power law distribution for fracture width. Figure 5.3 shows the power law distributions for the rock classes given in Table 5.1.

Permeability. Table 5.2 gives the logarithm of the mean and standard deviation for normal distributions for permeability and the associated rock class to which distributions apply [132]. Table 5.2 lists the porosity values for each rock class [143]. A weighted average is used for the permeability of a denied site:

$$Denied\ site\ permeability = \sum_{i=1}^{5} \frac{a_i}{A} k_i \tag{5.3}$$

Where "A" is the total area, $a_i$is the area of the individual rock class. And k_i is the individual rock class permeability.

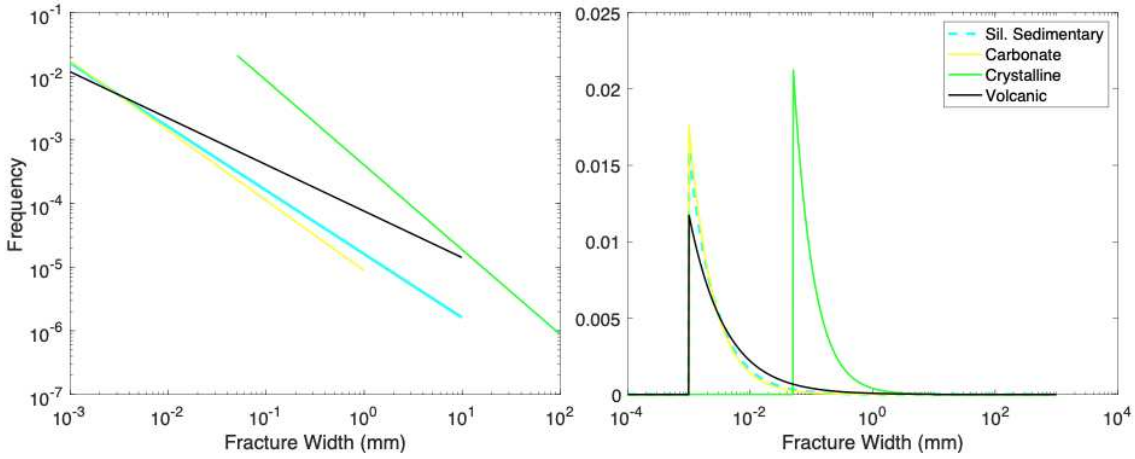

Figure 5.3 Fracture width distributions for different rock classes. The distributions are shown on a log-log axis (left) and a log-normal axis (right). The distributions correspond to the values associated with Table 12. The red line indicates the universal power law for Sandstone of 0.8 which is the recommended power law to use with no knowledge of the rock type [130], [138], [141], [142], [147], [148], [150]–[152].

Table 5.1 Fractal dimensions by rock class and associated rock types. The range of values indicated in the third column.

Rock Class	**Rock Types**	**Exponent D_{fw}**	**Average Exponent D_{fw}**	**Range**	**Reference**
Unconsolidated*	Alluvial, loess, dunes sediments	N/A	N/A	N/A	
Sil. Sedimentary	Sandstones, conglomerates	0.8 1.2	1.0	$1 \cdot 10^{-3}$ – 10 $1 \cdot 10^{-3}$– 0.2	[130] [152]
Carbonate	Limestone, dolomite	1.1 0.54 0.7 2.0	1.1	$1 \cdot 10^{-3}$– 1 0.075 – 0.75 $1 \cdot 10^{-3}$ – 10	[138][141][142]
Crystalline	Quartz, gneiss	1.33 1.23	1.28	$1 \cdot 10^{-3}$– 100 0.05 - 10	[150], [153] [140]
Volcanic	Tuff	0.73, 1.35	0.73	$1 \cdot 10^{-3}$ – 10	[139] [154]
Not Assigned	Water bodies, ice and glaciers, evaporites	-		-	

Table 5.2 Permeability and porosity values for the six rock classes**.** Each rock class has a mean permeability along with a standard deviation for a normal distribution, however the porosity values are deterministic values for each rock class [132].

Rock Class	**Rock Type**	$\log \mu[m^2]$	$\log \sigma[m^2]$	**Porosity**
Unconsolidated	Alluvial, loess, dunes sediments	-13.0	2	0.22
Sil. Sedimentary	Sandstones, conglomerates	-15.2	2.5	0.19
Carbonate	Limestone, dolomite	-11.8	1.5	0.06
Crystalline	Quartz, gneiss	-14.1	1.5	0.01
Volcanic	Tuff	-12.5	1.8	0.09
Not Assigned	Water bodies, ice and glaciers, evaporites	-	-	-

Table 5.2 lists values applicable to the near surface region (<100m). However, nuclear tests are typically conducted much deeper [116], [117]. Scaling for depth can be done using [135]:

$$\log k = \log k_r + (\log k_s - \log k_r)(1+z)^{-\alpha} \tag{5.4}$$

Here k { m^2} is the permability at depth z {km}, k_r {m} is the permeability at very large depth, $10^{-25.4}$ for Eq. (5.4), k_s {m} is the permeability at the surface, and α is a decay index and is equal to 0.25 [135].

Matrix tortuosity. The matrix tortuosity is estimated using the pore space fractal dimension [146]:

$$\tau = \theta^{D_{ps}-2} \tag{5.5}$$

Here θ is the porosity, and D_{ps} is the fractal dimension which can be estimated by rock class, see Table 5.5 for matrix fractal dimensions by rock class. The fractal dimension ranges in values from 1.01 – 2.

Porosity. The assigned porosity values, Table 5.2, by rock class are deterministic values. The porosity distribution is created by rock class weighted by the percent area of the rock class in the location area. A histogram is created for the percent area of each rock class within the location area. Replace the rock class with the porosity values assigned to each rock class. Once the histogram is built, fit a normal distribution over all the porosity values with a restriction of porosity ≥ 0.

Validation. These processes are used in two locations, the first being a site with known values and the second site is used in the area surrounding the underground test for the DPRK. The Nevada Test Site has had a geological characterization in which we can validate the process. The geological characterization collected borehole samples and measured the fracture spacing at outcrops in and around the Nevada Test Site to include Pahute Mesa, Ranier Mesa and Yucca Mountain. These measurements will be used to compare against the process estimated distributions for fracture spacing and width. This validation gives confidence in the estimations for the DPRK.

5.3 Results for the geological characterization.

Both the Nevada and DPRK test sites have played a critical role in the development of nuclear weapons for their respective countries. Unlike the US, the DPRK pursued these weapons after pulling out of the Nuclear Non-Proliferation Treaty. Monitoring by the Comprehensive Test Ban Treaty Organization has been essential to help understand the state of their program and verify the veracity of claims being made by the North Korean government. The detection of anthropogenic radioisotopes provides the only definitive signature that a nuclear weapons test has occurred and can also be used to inform aspects of the design [85]. Accurately predicting the time at which post detonation gases arrive at the surface, and their composition, is an essential part of these efforts. However, this requires detailed knowledge of the subsurface environment, which is missing for a denied access site like the one used by the DPRK.

The methodology described here provides a systematic way to develop the input parameters for subsurface gas transport models. Unlike North Korea, the US has made public considerable information about the geology of the Nevada Test Site, which makes it an ideal location with which to validate the approach. Geological characterization of the Nevada Test Site resulted actual core sample measurements. These measurements are described by either a distribution or a range of values for each parameter. The ranges and distributions are used to compare to the predicted distributions [155]. Matrix and fracture tortuosity are left out of the comparison as direct measurements of could not be found.

Fracture spacing. Figure 5.2 shows lineation overlays at three different scales for the location area surrounding the Rainer Mesa Mountain at the Nevada Test site. The overlays were developed using ArcGIS with world map imagery [144]. The imagery was imported at each described scale to identify each lineation. As shown, the number of lineation's increases as the

scale decreases. This results with the spacing between each lineation decreasing as the scale decreases. See Fig 5.14 – 5.15 for the histogram and fitting distributions for the lineation overlays. Table 4 lists the mean and standard deviations at each imagery scale for the Nevada Test Site.

Figure 5.4 (left) shows a plot of the natural log of the mean spacing distribution against the natural log of the scale map. Figure 5.4 depicts a regression analysis on the three points on a plot. The result of the regression analysis was an R^2 of 0.99 and an equation of the line was $y = 0.6314 \cdot x + 0.5262$. Where *x* is the natural log of the map scale and *y* is the natural log of the fracture spacing in meters. Therefore, the geology surrounding the Rainer Mesa suggests a fractal dimension of 1.8. Extending this regression line down to a 1:1 scale, the mean of the fracture spacing distribution is 1.7 m. Figure 5.5 (right) shows the same analysis on the standard deviation at each imagery scale. The result of the regression analysis was $R^2 = 0.98$ and the equation of the line is y = 0.0551*x* + 1.1377. Here, *x* is the natural log of the imagery scale in kilometers and y is the natural log of the standard deviation of the fracture spacing distribution in meters. Therefore, the standard deviation at the 1:1 scale is 3.1 $meters$.

Using the mean and the standard deviation with a normal distribution results in a fracture spacing distribution depicted in Fig. 5.4. The normal distribution is constrained to fracture spacing ≥ 0 meters. The blue dotted lines in Fig. 5.4 show the actual fracture spacing at the Nevada Test Site reported from a geological characterization from boreholes and outcrop measurements [156], [157]. The blue uniform distribution depicts the actual values based upon sample measurements [155]. The 95% confidence interval for the predicted distribution is 0.3 and 7.3 meters respectively.

Fracture Width. The rock class for the Nevada Test Site is volcanic tuff for which fractal dimensions of 0.73 and 1.35 have been measured [139] [154]. Figure 5.6 shows the corresponding distributions for both along with the range of fracture widths from field measurements. The limited range for the actual measurements shown likely comes from the limited number of samples from which the measurements were made. The power law distribution allows for extrapolation outside of the measured widths assuming that fracture widths exhibit a natural fractal behavior. This allows for a better match of the predicted and actual fracture width distributions.

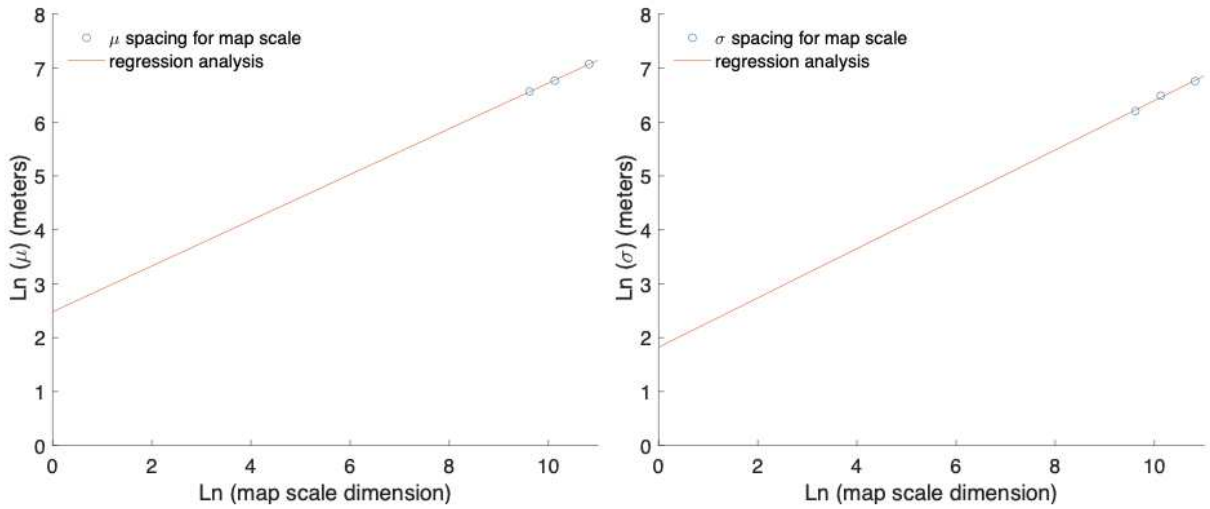

Figure 5.4 An extrapolation for the mean (left) and standard deviation (right) for fracture spacing at the Nevada Test Site. The regression line extends from the three scale points to the actual fracture spacing at the location site. The extrapolation creates a normal distribution with mean (μ) and standard deviation (σ) for the fracture spacing within the location area.

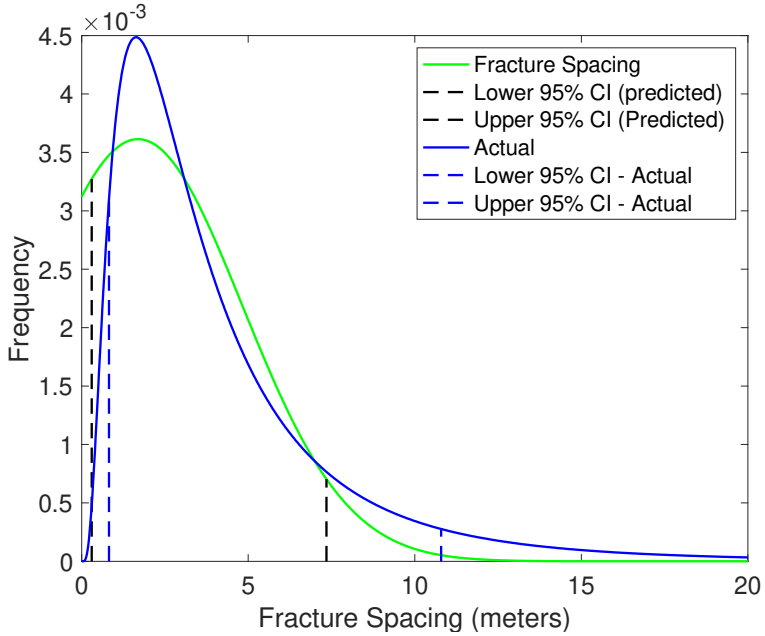

Figure 5.5 Fracture spacing distribution for the Nevada Test Site. Depicted in the green based upon a normal distribution with mean and standard deviation determined through the fractal analysis. The true values are from outcrop and borehole measurements throughout the Nevada Test Site [1] [2].

Table 5.3 Comparison of the predicted vs actual fracture spacing distributions for the Nevada Test Site. The mean, standard deviation and the 95% confidence intervals are listed based upon the distributions shown in Fig. 5.5.

	Predicted [m]	**Actual [m]**
Mean {m}	1.7	1.1
Standard Dev {m}	3.1	2.2
95% CI (low, high) {m}	(0.3, 7.3)	(0.8, 10.8)

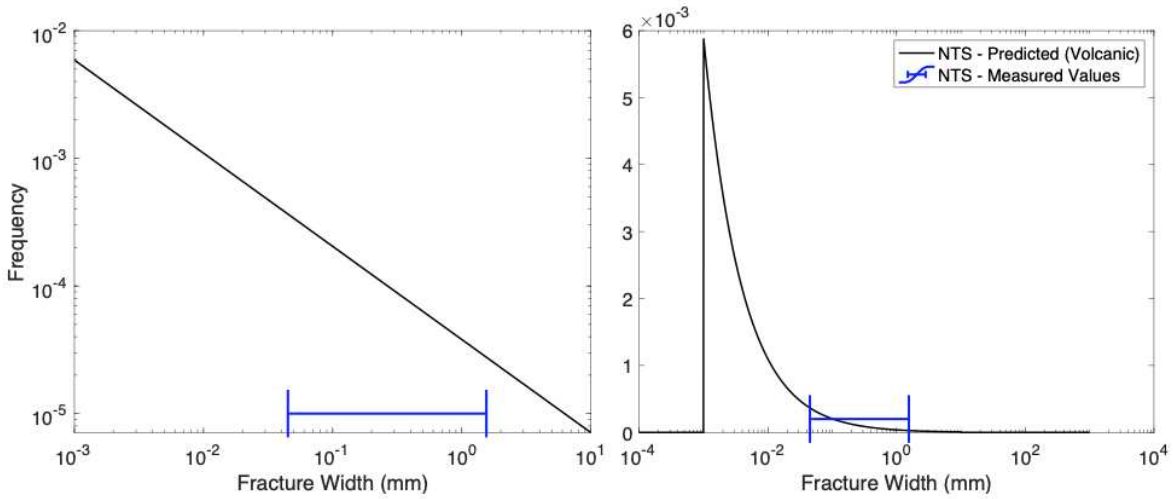

Figure 5.6 Fracture width distribution for the Nevada Test Site. The left plot shows the power law distribution on a log-log plot for the predicted fracture width along with the reported actual range. The right plot shows it on a semi-log plot with the reported range of measured values [139], [158].

Permeability. Figure 5.7 shows the predicted permeability distribution for the Nevada Test Site. This location area has three different rock classes and each individual permeability distribution is weighted by the amount of area (left) with their weighted sum shown in black. Figure 5.7 (right) shows the normalized final distribution along with the range of permeability values reported by McCord (2007) [155]. The distribution in black Fig. 5.7 (right) is the distribution to use in the subsurface gas transport model for a stochastic analysis.

Porosity. The predicted porosity distribution for the Nevada Test Site is depicted in Fig. 5.8 in red with 95% confidence interval. The blue distribution is the actual distribution with 95% confidence interval. The predicted distribution is skewed to the left with predicted lower porosity values.

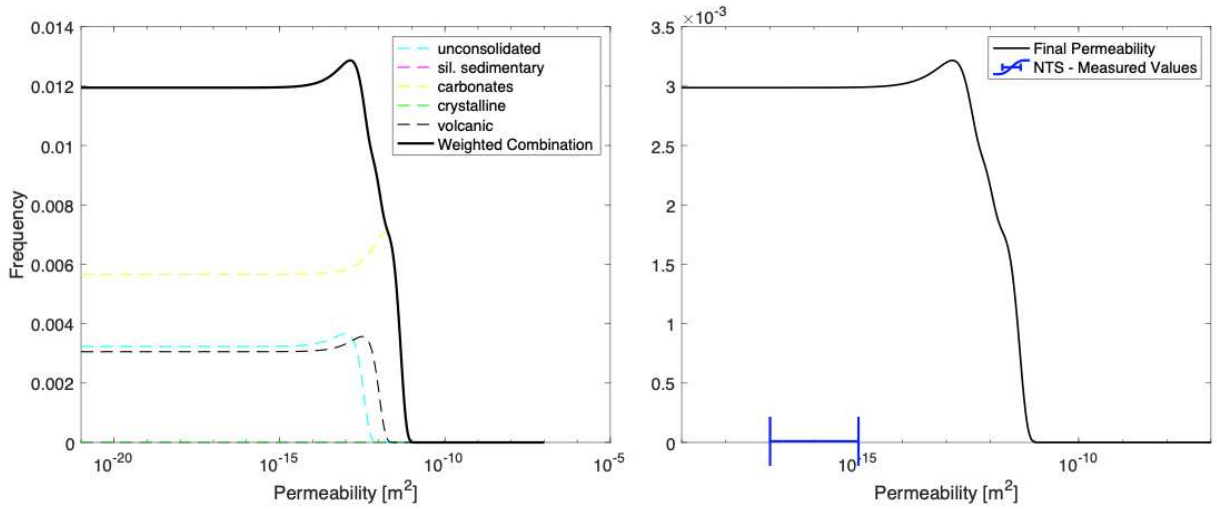

Figure 5.7 Permeability distributions for the Nevada Test Site. The left shows the individual permeabilities that are included in the location area weighted by the amount of area. The black line indicates the sum of the individual permeabilities. The right plot shows the final normalized distribution. The blue error bars display the range of measured values of the permeability from Nevada Test Site along with a reported median permeability [155].

5.4 Results for the area surrounding the DPRK underground nuclear test site.

Fracture Spacing. Figure 5.9 shows three lineation overlays for the location area that surround the Punggye-Ri test site in DPRK. The same scaling effects are observed. The analysis described in *supplemental note 1* for the histogram and distribution fitting of the lineation's overlays for the Nevada Test Site was conducted on the overlays for the DPRK shown in Fig. 5.9.

Figure 5.10 (left) depicts a regression analysis on the three distribution means from the lineation overlays on a plot. The result of the regression analysis was an $R^2$of 0.99 and an equation of the line was y=0.4244 • x + 2.4791. Where x is the natural log of the map scale and y is the natural log of the fracture spacing in meters. This results in a fractal dimension of 1.42 for the geology in the DPRK Punggye-Ri test site. Extending this regression line down to a 1:1 scale, the mean of the fracture spacing distribution is 11 meters.

To determine the standard deviation for a normal distribution, Figure 5.10 (right) depicts the same analysis on the standard deviation at each map scale. The results of the regression analysis had a $R^2 = 0.99$ and the equation of the line is $y = 0.4575 \cdot x + 1.822$. Therefore, the standard deviation at the 1:1 scale is 6.2 $meters$. Using the mean and the standard deviation with a normal distribution results in a fracture spacing depicted in Fig. 5.10.

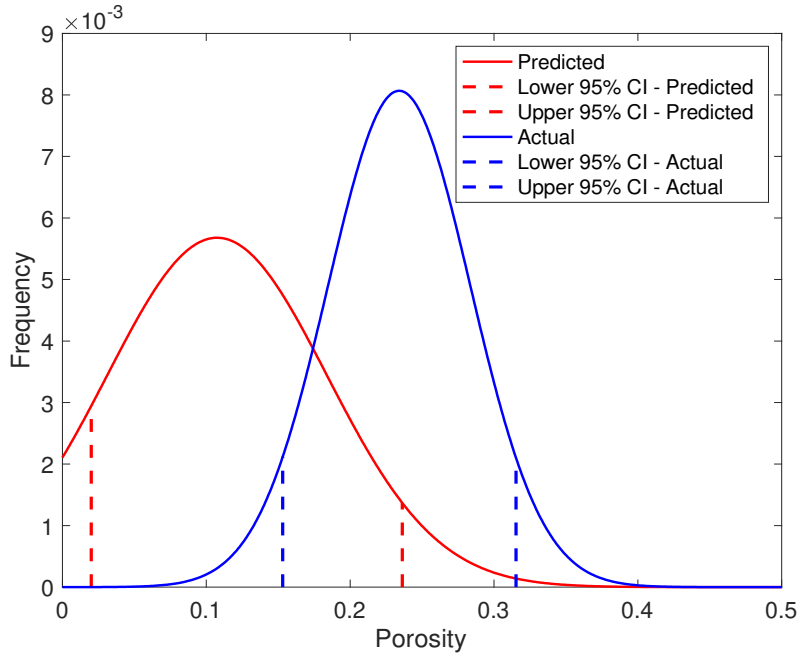

Figure 5.8 Porosity distribution for the area around Nevada Test Site. The distribution shows a normal distribution about which was a result of the analysis of the class within the location area surrounding the underground nuclear explosion.

Table 5.4 Comparison of the predicted vs actual porosity distributions for the Nevada Test Site. The mean, standard deviation and the 95% confidence intervals are listed based upon the distributions shown in Fig. 5.8.

	Predicted [m]	**Actual [m]**
Mean {m}	0.12	0.2
Standard Dev {m}	0.08	0.05
95% CI (low, high) {m}	(0.3, 7.3)	(0.15, 0.32)

Fracture width. The rock class for the DPRK test site consists of Silicate Sedimentary, Crystalline and unconsolidated geology. The fracture width for the DPRK test site is a power law distribution with a weighted average of the fractal dimension for the exponent. This results in a fractal dimension of 1.1.

Permeability. Figure 5.12 shows the permeability distributions for each rock class (left) and the final permeability distribution (right), based upon the location area analysis for a weapon test at the denied access site within the DPRK. As shown, this location area had three different rock classes. The weighted sum is depicted in black on the left plot. The right plot shows the

normalized final distribution to use in the subsurface gas transport model for a stochastic analysis.

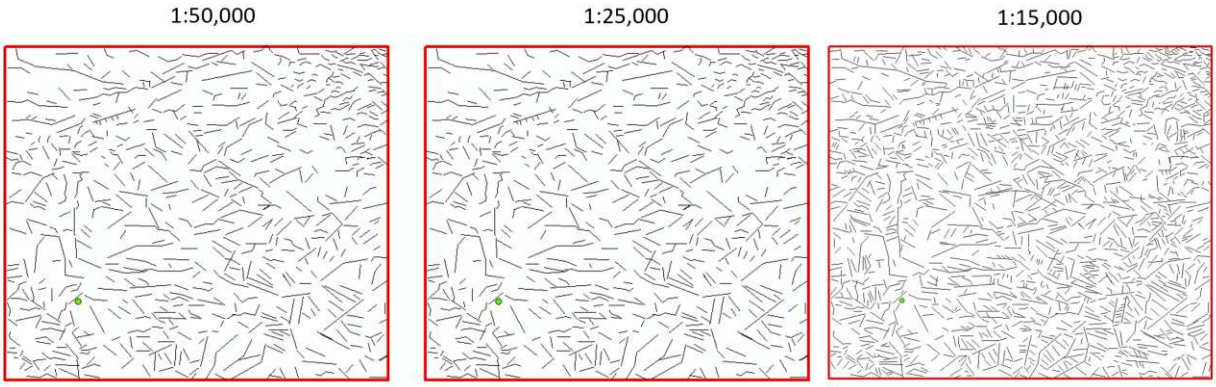

Figure 5.9 Lineation identification for three different scales for the test site in the DPRK. The lineation overlays were created in ArcGIS with world map imagery at different scales [144]. The scale of each imagery was from left to right 1:50k, 1:25k, and 1:15k.

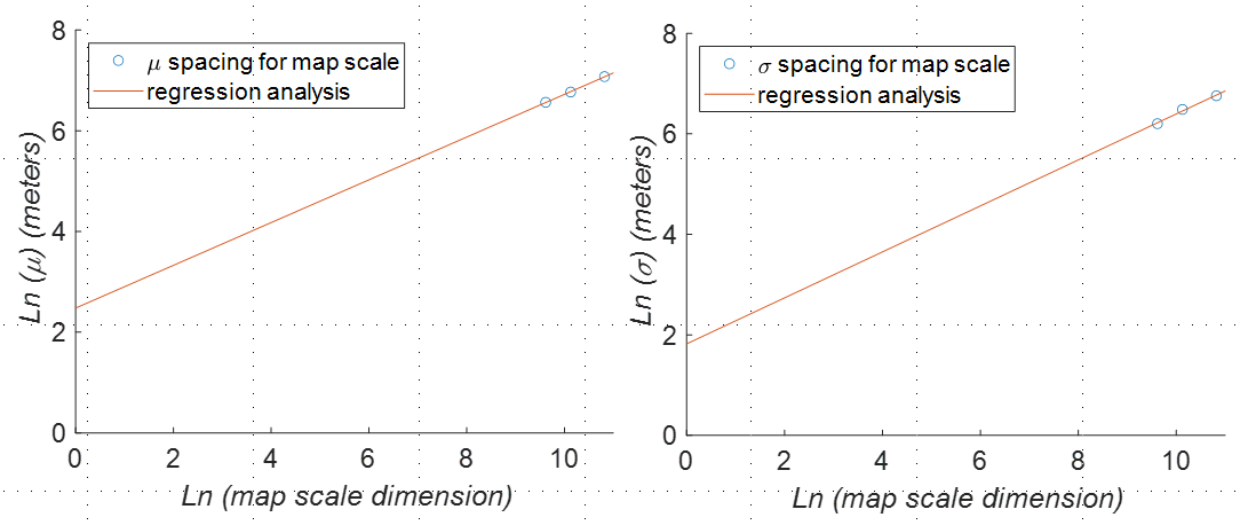

Figure 5.10 Extrapolating the fracture spacing distribution down to the 1:1 scale for the location area surrounding Punggye-Ri. Left, depicts the extrapolation of the mean fracture spacing (μ). Right, depicts the extrapolation of the standard deviation (σ). The method uses a power law regression method.

The median permeability for the DPRK test site at Punggye-Ri is $4.8 \cdot 10^{-18} m^2$. The permeability for the DPRK test site is lower than the predicted permeability at Semipalatinsk.

This provides additional evidence the geology surrounding the DPRK test site is hard solid rock that does not facilitate subsurface gas transport. The ideal location for a clandestine nuclear test site to prevent the international community from knowing the results of its tests.

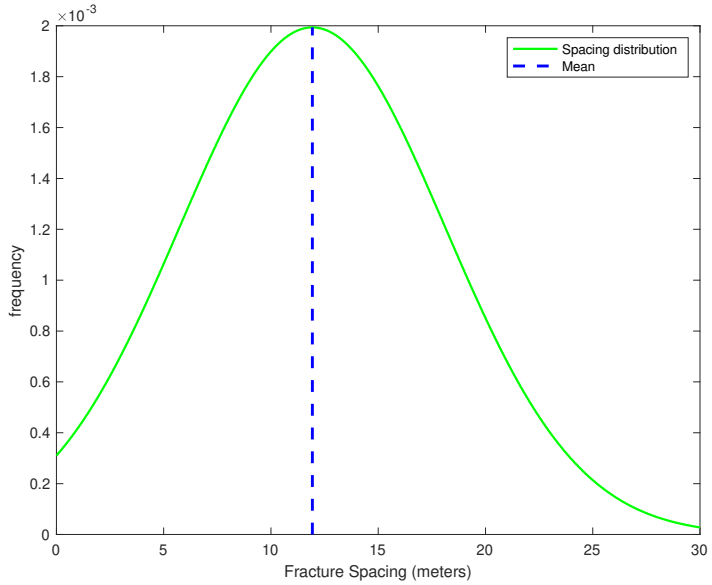

Figure 5.11 The final fracture spacing distribution for the location area surrounding the Punggye-Ri underground nuclear explosion complex in the DPRK. This distribution is based upon a normal distribution with mean and standard deviation extrapolated from the analysis depicted in Fig. 5.10.

Porosity. The porosity distribution for the DPRK test site at Punggye-Ri is shown in Fig. 5.13. The average porosity for the DPRK test site is 0.05. An average porosity value of 0.05 is an indication of hard solid rock. The DPRK rock type included Crystalline rock that has an assigned porosity of 0.01. This shifted the porosity distribution to the left to lower porosity values. This is support for hard solid rock that is an ideal location for an underground nuclear explosion.

Fracture tortuosity. The fracture spacing overlays for the DPRK (Fig. 5.9) were used to calculate the fracture tortuosity fractal dimension (D_{ft}). The analysis resulted in an average fractal dimension of 1.25 for the DPRK test site. Using Eq. 5.2, this results in a fracture tortuosity of 1.13 for the test site at DPRK.

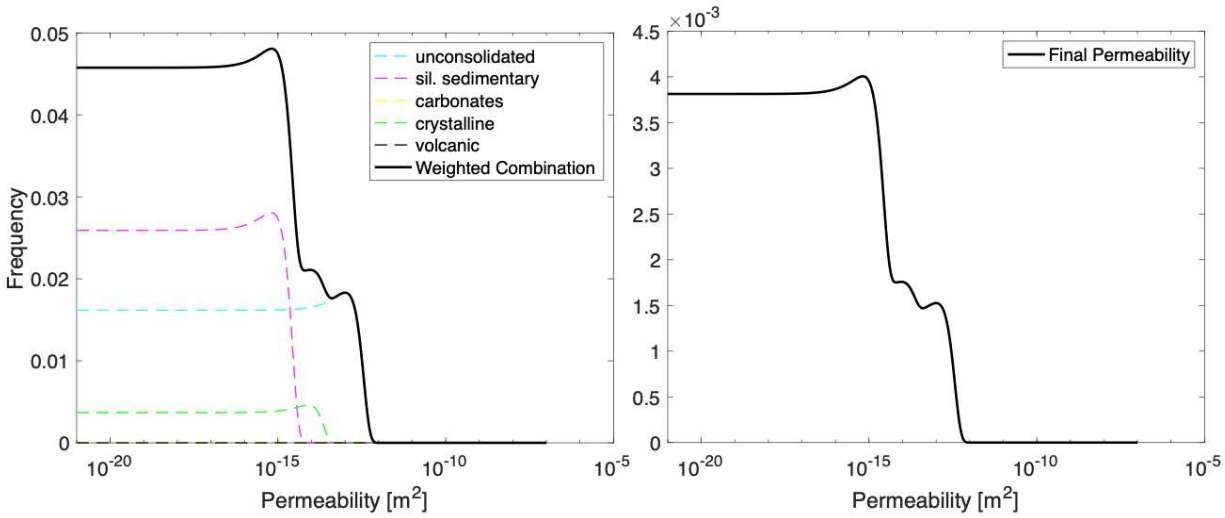

Figure 5.12 Permeability distributions for the test site in the DPRK. The left shows the individual permeabilities that are included in the location area weighted by the amount of area. The black line indicates the sum of the individual permeabilities. The right plot shows the final normalized distribution.

5.5 Conclusions for geology at a denied access site.

The methodology developed for characterizing a denied access site can be applied using open source data and avoids the need to rely on expert judgement. Characterization of the Nevada Test Site was used for validation. All the predicted property distributions overlapped with direct field measurements. The lineation fractal analysis for the Nevada Test Site produced a normal distribution that resulted with a mean fracture spacing within 60 centimeters of the actual mean. The standard deviation was 90 centimeters larger than the actual distribution. The fracture width for the Nevada Test Site also resulted in a larger distribution but recent characterization methodologies look to extrapolate measured fracture widths to other widths based upon limitations of the rock samples [130]. Utilizing the geological rock class maps with the associated permeability distributions resulted in a final permeability distribution for the Nevada Test Site that encompassed the entire range of the actual permeability measurements. The porosity distribution was shifted to smaller porosity values than the actual porosity distribution. The mean of the porosity was -0.08 from the actual and the standard deviation was 0.03 larger. Additionally, the tortuosity was estimated treating it as a natural fractal.

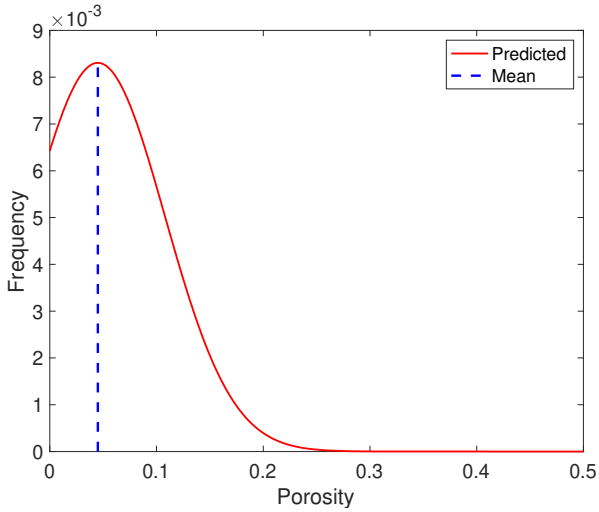

Figure 5.13 Porosity distributions for the area that surrounds the DPRK underground nuclear explosions. The plot analyzed the rock classes within the location area and the assigned porosity values to each rock class.

The methodology developed was also applied to the underground nuclear test site (Punggye-Ri) identified by Bayesloc in the DPRK and gives the first full estimate for geological conditions in the open literature. The results for the DPRK test site are consistent with hard solid rock. The geologic properties contain the explosion effects and slow gas transport to the surface through barometric pumping. The geology would be a prime environment to conduct an underground nuclear explosion in a clandestine environment to prevent the international community from gathering insight into your technology and capabilities.

5.6 Lineation identification discussion and intermediate steps to estimating the fracture spacing

Lineation identification: Lineation's are seen in imagery as lines of various lengths and distinctness crossing the landscape. The natural lines can be from a change in color, a straight segment of a tree line, or stream segment[137]. A detailed explanation to describe how to identify lineation's from air photo's or imagery is described by Rinker and Corl, and Ehlen [137], [159]. The primary concern in some areas is to ensure the lineation is a naturally occurring linear feature. The primary misidentification of a lineation is a manmade feature [137]. Such manmade features that can be confused for a naturally occurring lineation are roads, farm crop fields, and power lines. However, this makes the method a useful method because most underground nuclear test sites are in remote areas with very little manmade features. This

is evident at the Nevada Test Site and the DPRK test site. Other sites in which nations have conducted underground tests are also remote and lack manmade features.

Fracture spacing fractal dimension. For each lineation overlay, a histogram for the spacings between lineations is created. The spacing should be measured along vertical and horizontal segments [160]. In order to get a good representative vertical and horizontal spacing data, ten equally spaced vertical and horizontal lines were used to measure the lineation spacing in each overlay.

Figure 5.14 (left) shows a histogram of the lineation spacing distributions found in the 1:50,000 scale for the analysis done for the area surrounding Rainer Mesa at the Nevada Test Site. The distribution is used to determine the mean and standard deviation of the lineation spacing within the location area at a specific imagery scale. Figure 5.14 (right) depicts the fitted distributions for the three different scales for the surrounding area around Rainer Mesa of the Nevada Test site.

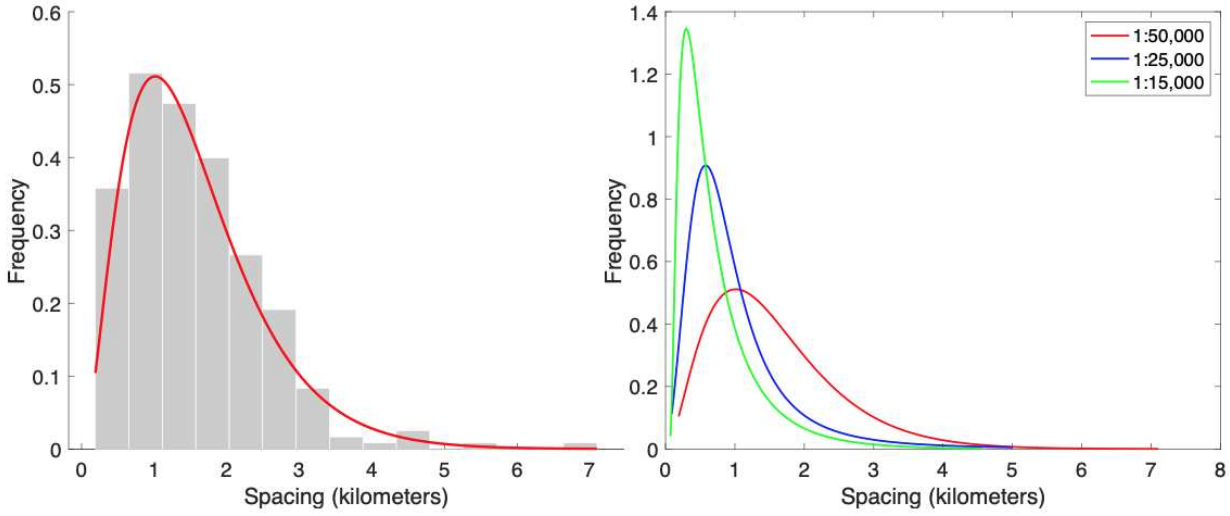

Figure 5.14 Creating distributions of the fracture spacing for the Nevada Test Site. A histogram of the lineation distribution at the 1:50,000 imagery scale for the Nevada Test Site (left). The histogram is fitted with a distribution to determine the mean and standard deviation lineation spacing (left, red). Spacing distributions for the Nevada Test Site for three different imagery scales (right).

The same analysis to determine a fracture spacing distribution was conducted for the location area surrounding the DPRK's underground nuclear explosions. The results of the lineation spacing distributions are depicted in Fig 5.15. The DPRK test site is a denied access site so no true values are known for fracture spacing.

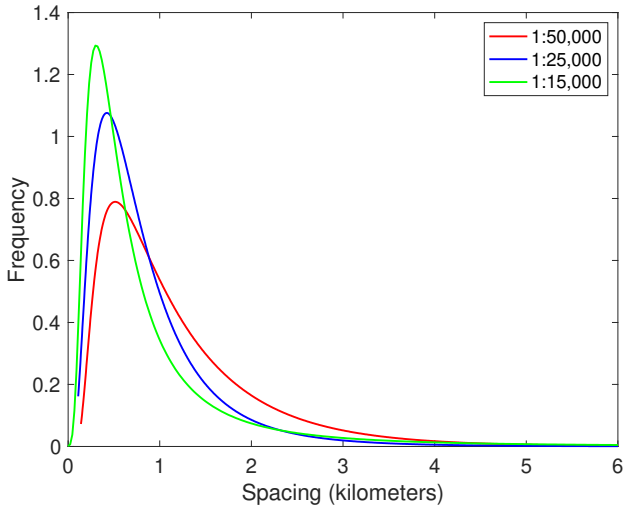

Figure 5.15 Fracture spacing distributions for Punggye-Ri. Three different imagery scales for the location area surrounding the Punggye-Ri underground nuclear explosion complex in the DPRK.

The fractal dimension is determined by taking the exponential of the slope of the best fit line (Fig. 5.4 for NTS and Fig. 5.10 for DPRK). The slope for the NTS analysis was 0.55. Therefore, the fractal dimension for fracture spacing at NTS is equal to 1.7.

5.7 Fracture tortuosity fractal dimension estimation by the box counting method

The tortuosity is estimated using Eq. 5.2. Equation 5.2 requires the fractal dimension of the fractures within the subsurface. This is estimated using the lineation overlays created for the fracture spacing (Fig. 5.2 for NTS and Fig. 5.9 for DPRK). With the overlays, a single lineation that represents a fracture is taken and analyzed using the box counting method. Figure 5.16 depicts the process of randomly selecting lineations for analysis with the box counting method. The box counting method utilizes a process of adjusting the box size and counting the number of boxes in which the fracture is in each box to do a regression analysis to determine the fractal dimension. The fracture tortuosity fractal dimension was 1.2 and 1.25 for NTS and DPRK respectively.

5.8 Permeability distributions by rock class and intermediate steps

Figure 5.17 shows a variation of the global map shown in Fig. 5.1, however it assigns a permeability group according to the lithologies of Table 5.1. The right side of Fig. 5.17 is a histogram for the prevalence of each permeability group according to the global map on the left side of Fig. 5.17. Figure 5.18 shows the actual permeability distributions associated with each rock class on a semi-log plot using the mean and standard deviations from Table 5.1.

Figure 5.16 Fracture tortuosity fractal dimension estimation. The lineation overlay for the DPRK for the 1:50,000 scale is shown on the left. Random lineation's are isolated and analyzed with the box counting method.

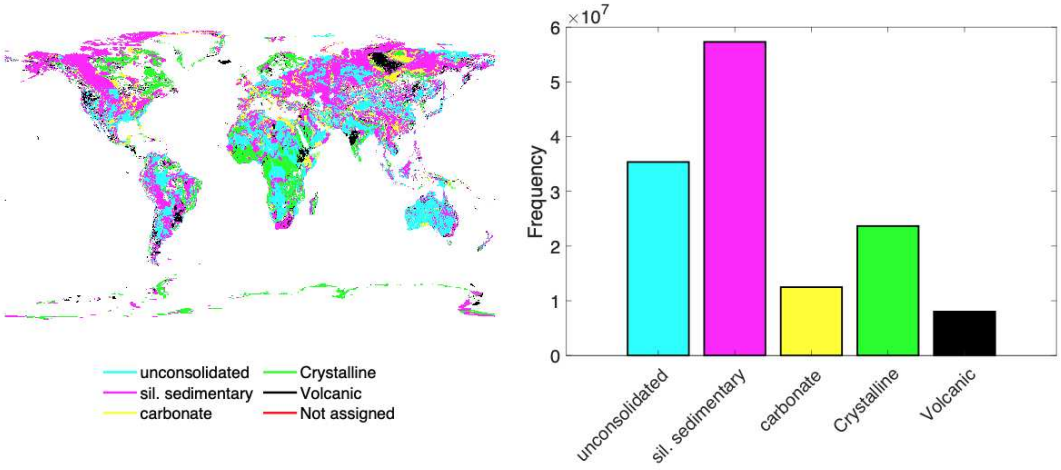

Figure 5.17 A variation of the rock classes from table 5.1 with rock classes grouped according to the permeability groups shown in Table 5.1. The left is a global map of the 5 rock classes and the right shows a histogram to indicate the prevalence of each rock class [132], [134]. The colors on the global map correlate with the histogram x-axis labels and colors.

Figure 5.19 shows a blowout of the location area and the geology types within the location area. The two areas are the USSR test site at Semipalatinsk, and the DPRK test site at Punggye-Ri.

5.8 Matrix tortuosity values by rock class.

Table 5.5 shows the fractal dimension from various studies for each rock class.

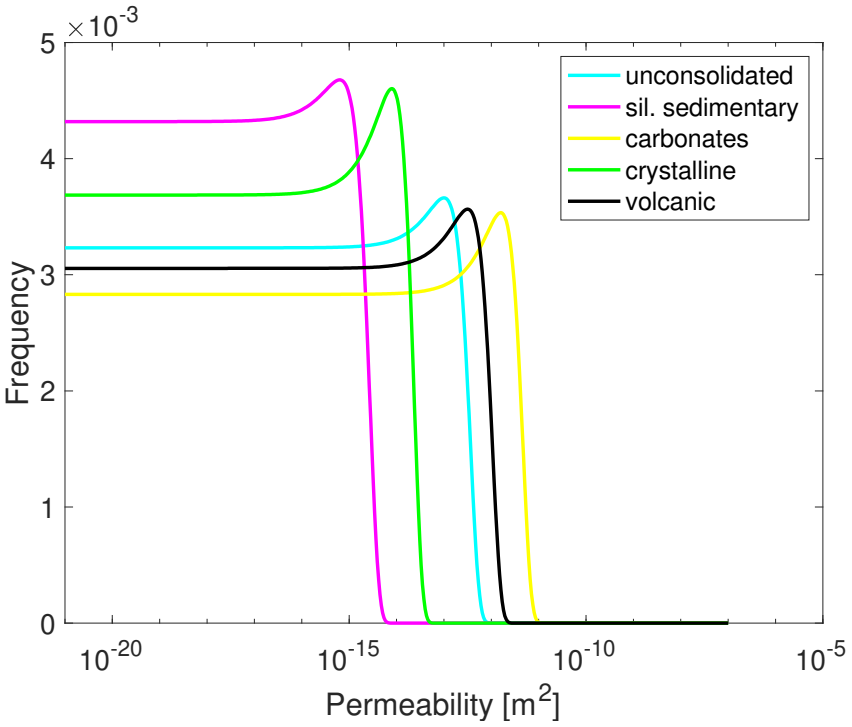

Figure 5.18 Permeability distributions according to table 5.1 [3]. The permeability distributions were developed by rock class.

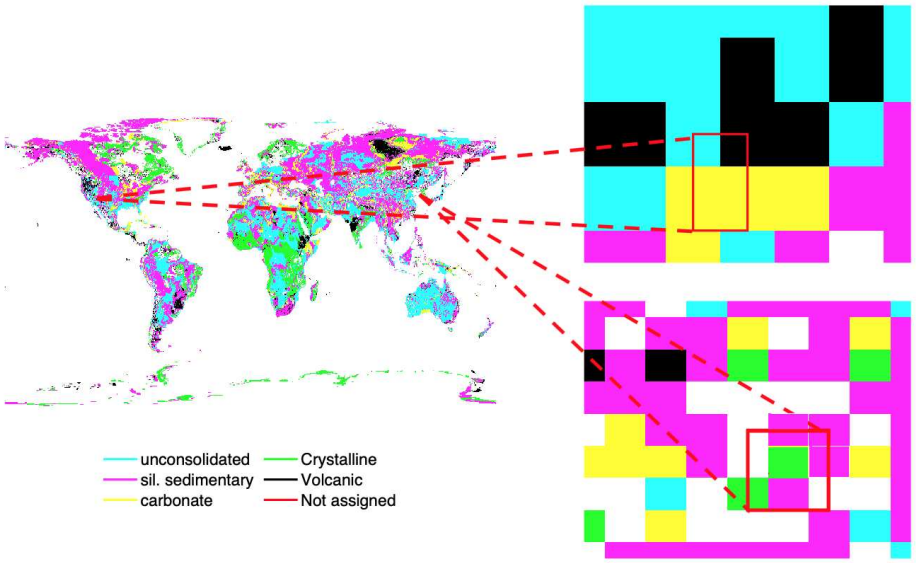

Figure 5.19 A permeability map for two location areas. The upper right pop-out location area is for the Nevada Test Site. The lower location area is the DPRK Punggye-Ri test site. The location areas are determined using seismic arrival data and using the software known as Bayesloc [88], [106].

Table 5.5 Fractal dimension by rock class for matrix tortuosity [140], [161]–[164]**.**

Rock Class	**Fractal Dimension**
Unconsolidated	1.01 – 1.2
Sil. Sedimentary	1.1 – 1.3
Carbonate	1.3 – 2
Crystalline	1.4 – 1.8
Volcanic	1.2 – 2
Not Assigned	-

CHAPTER SIX

SUBSURFACE XENON TRANSPORT OF THE NORTH KOREAN UNDERGROUND NUCLEAR TESTS

The Democratic People's Republic of Korea claimed six underground nuclear tests between 2006 and 2017. Following the tests in 2006, 2013 and 2016, elevated levels of radioactive xenon were detected by the Comprehensive Nuclear Test Ban Treaty Organization's International Monitoring System. Past work has used reverse atmospheric transport to identify possible sources for the radioxenon but the results have not been definitive. The ratios of Xe135/Xe133 and Xe133m/Xe131m in particular can be used to help establish if the radioxenon is from a weapon or another source. However, following the six claimed North Korean tests, none detected all four isotopes. Here we used a double porosity model and estimates for the subsurface geology and barometric history at the Punggye-ri nuclear site, to analyze when xenon venting is most likely to have occurred. The results show that emission of all four radioxenon isotopes is a near term phenomena and that Xe135 concentrations would have fallen below detectable limits at Punggye-ri within 20 days after the claimed tests. The results show that the use of radioxenon ratios to discriminate between sources is of limited use unless one gains access to the test site.

6.1 Introduction to North Korean underground nuclear explosions.

Between 2006 and 2017 the Democratic People's Republic Korea (DPRK) claims to have conducted six underground nuclear tests [165]. The International Monitoring System of the Comprehensive Nuclear Test Ban Treaty Organization (CTBTO) uses infrasound, seismic, and hydroacoustic signals to detect explosions. However, only the presence of anthropogenic radionuclides can be used to definitively determine whether a nuclear weapon was the cause. The ratios of Xe^{135}/Xe^{133} and Xe^{133m}/Xe^{131m} can be used to distinguish the signature of a weapon from that of another source [85]. However, all four isotopes have never been detected after the suspected DPRK tests [166]. Following the first, third and fourth tests the International Monitoring System detected elevated levels of Xe^{133} as well as elevated levels of Xe^{131m} for the first and third test [167]–[169]. Monitoring for radioxenon is an important component of the International Monitoring system, but it is unclear if it can provide the type information hoped for if a suspected test occurs at a site to which the international community has no access.

The detection of radioxenon following the first DPRK test occurred at the border of North and South Korea five days after the explosion [170], [171]. A mobile detection unit was deployed to the border to take samples [171] and Xe^{133}, Xe^{133m} and Xe^{131m} were detected. However, background observations showed that Xe^{131m} was frequently present in that area, and at comparable concentration levels, so it could not be attributed to a test five days before [171]. For the third and fourth underground explosions, the CTBTO's International Monitoring System station in Tasaki, Japan detected elevated levels of Xe^{133} several weeks following the underground explosion [166], [172]. For the third test, Xe^{131m} was also detected in Tasaki Japan as well as station in Ussuriysk, Russia also detected elevated levels of Xe^{133} and Xe^{131m} [166]. Figure 6.1 summarizes the six claimed tests, what was detected, and their estimated yields.

	Underground Explosion	Surface Venting	IMS Detection	
Test Number	Date of Explosion	Date Range of Surface Venting	Location and Date of Detection	Date of Detection
1	October 9, 2006	October 9 - 11, 2006	ROK - DPRK Border	October 11 - 14 2006
2	May 25, 2009			
3	February 12, 2013	Mar 21 - April 15 2013	Takasaki JP Ussuriysk, RS	April 7 - 9 2013 (JP) April 12 - 13 2013 (RS)
4	January 6, 2016	February 11 - 19, 2016	Takasaki JP	Feb 16 - 18 2016
5	September 9, 2016			
6	September 3, 2017			
Test Number	Explosion Characteristics	Estimated Depth of Burst	Sample description	Comments
1	1.8 KT	126 m	Five air samples collected during	Xe - 131m, Xe - 133m, and Xe 133, detected
2	3.7 KT	149 m		
3	8.8 KT	213 m	Three samples JP 2 samples in RS	Xe 131m and Xe 133 seven to eight weeks after explosion
4	8.5 KT	213 m	Elevated levels of in five samples	Xe 133 only,
5	13. 6 KT	192 m		
6	290 KT	481 m		

Figure 6.1 A summary of the DPRK underground nuclear tests. The first column lists the sequence of the test and then the associated date of the explosion. The surface venting is determined from reverse modeling from the IMS detection platform from atmospheric models. This effort will attempt to model the surface venting from the time of the explosion forward to corroborate the IMS detection possibility [166], [171]–[173].

Reverse atmospheric modeling was used after the 1st, 3rd and 4th tests claimed by the DPRK to determine if these noble gas radioisotopes came from the Punggye-ri nuclear test [166], [172]. The results are consistent with Punggye-ri as a source for the radioxenon, but they could not rule out the possibility of a different origin. Monitoring stations failed to pick up the Xe^{135}

and Xe^{133m} that would have allowed for source discrimination and it remains to be understood why. Independent estimates for when these radioxenon isotopes would have arrived at the surface are important here and also in understanding how detector systems could be better deployed in the future [174], [175]. Double porosity models can be used to do this if the geological conditions, depth of burst, yield and barometric conditions can be estimated [85].

Here we estimate the time to venting for the first, third and fourth tests claimed by the DPRK. We use the 2-D subsurface radioxenon transport and depletion model developed by Lowery (2013) [85]. This is a double porosity model with barometric pumping and is used to predict the location and isotopic concentration of each xenon isotope at a time period post detonation [86]. Through barometric pumping, the xenon gas will move vertically through fractures of the geology. At the same time, radioactive decay contributes to the production depletion of radioxenon isotopes.

6.2 Methods for analysis of the DPRK underground nuclear explosion.

The 2-D subsurface radioxenon transport model developed by Lowery (2013) requires seven inputs [85], Supplemental Note 1. The first six are depth of burst, yield, fracture width, fracture spacing, permeability, porosity. All input variables are assigned a distribution for each underground explosion conducted by the DPRK from previous studies according to Duncan et. al. [176], [177]. Monte Carlo sampling was used to generate possible outcomes for radio xenon transport to the surface for each of the claimed DPRK tests. The results were used to create a distribution for the fraction of simulations that predict venting of Xe^{135}, Xe^{133}, Xe^{135m} and Xe^{131m} at concentrations above the minimum detectable limit as a function of time. The seventh input parameter is the barometric pressure. Barometric pressure is taken from Changbai, China weather station [178]. The weather station reports the atmospheric pressure has an average every three hours [178]. Interpolation is used to determine the atmospheric pressure between reported pressure times. See *Supplemental Note 2* and figure S1 for a comparison of barometric pressure at three different weather stations that surround the DPRK underground test site.

For each claimed test at Punggye-ri Monte Carlo runs were done in which the distributions for depth of burst and geological parameters were sampled and combined with data on barometric history. The resulting radioxenon concentrations as a function of depth were computed for 2, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90 days post detonation. A threshold for detection was set at the minimum detectable limit, defined as $1\ mBq/m^3$, and based on

detection limits of the CTBTO radionuclide detectors [82]. The results were used to determine the fraction of models for which the radioxenon isotope concentrations were above the minimal detection limit. This was then used as a measure of the probability of exceeding this detection limit. See *Supplemental Note 2* for examples of histograms, probability distribution functions and cumulative distribution functions.

6.3 Results and discussion for Monte Carlo simulation.

Figure 6.2 shows the tests that the CTBTO believes xenon gas vented to the surface and was detected by International Monitoring System along with the respective isotopes found. These tests were test one, three and four conducted in 2006, 2013, and 2016. The seismic data for these tests indicates that all the tests were under 20 kilotons [179]. The sixth test predicted a much higher yield than the first five tests [179]. This would indicate an advancement in weapons technology. However, the CTBTO did not detect any xenon gas following this test [168]. However, for a yield greater than 100 kilotons, detecting the shock wave alone should provide enough evidence to prove that it was a nuclear explosion because chemical explosions have not been achieved at those yields.

Figure 6.2 shows the cumulative distribution function for Xe^{133} for the first DPRK test (left) along with the time dependent probability of exceeding the minimum detectable limit at the surface of the Punggye-ri site (right). Almost 50% of the Monte Carlo simulations predict that Xe^{133} will exceed minimum detectable limits during the first 70 days post-detonation. However, the number of simulations that show Xe^{133} exceeding minimum detectable limits drops to zero within 20 days. It is important to note that the small yield of the device in the first claimed test suggests that it was buried at shallow depth, which would increase the likelihood that this isotope would make it to the surface before decaying away. It is also important to note that simulation results are relative to concentrations at the surface in Fig. 6.2 (right). These would of course drop as radioxenon is transported through the atmosphere.

Isotopes of Xe^{135}, and Xe^{133m} reach a peak probability and then start to decrease in probability. This decrease is associated with the radioactive decay and the decay of the isotopes that results in the minimum detectable limit depth moving down towards the cavity until it has decayed away from detectable levels. The probabilities of isotopes of Xe^{131m} and Xe^{133} continue to increase for 50 days. The peak probability for all isotopes is 47% between day 40 and 70 for

Xe^{133}. With this analysis it is likely that xenon vented from the explosion site and vented into the atmosphere.

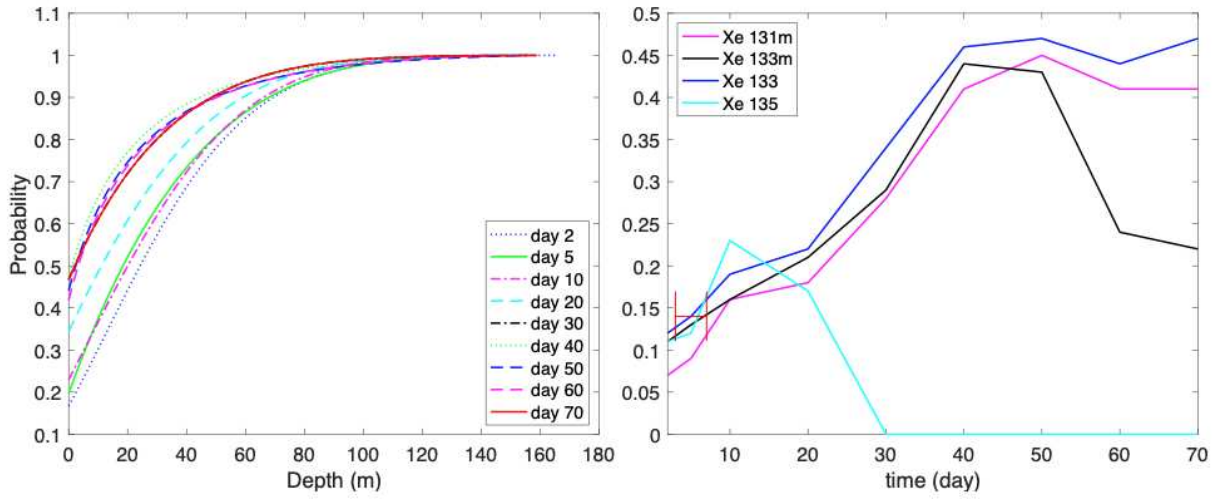

Figure 6.2 Cumulative Distribution of the depth of xenon-133 for the first DPRK underground nuclear explosion (left and the probability of venting vs time (right). The right figure shows $Xenon^{133}$ had the highest probability to vent to the surface. The right figure takes the fraction of the simulations that vented to the surface (depth =0 on the left figure) as a function of time.

However, results also indicate that at a lower probability other isotopes of xenon should have been detected by the same mobile detection platform. In particular, Xe^{135} is seen to disappear very quickly. For the third DPRK nuclear explosion the CTBTO detected elevated levels of Xe^{131m} and Xe^{133} at International Monitoring System stations in Takasaki, Japan and Ussuriysk, Russia [166]. The detection happened several weeks after the explosion. With reverse modelling, the expected venting of gas to the surface would have occurred at ~61 days post explosion [166]. Figure 6.3 (left) shows the fraction of the Monte Carlo simulations that predict the concentration of Xe^{133} exceeding the minimum detectable limit as a function of depth, and at the surface as a function of time (right). These results indicate that it is likely the xenon gas was contained from the initial explosion. Ten days following the explosion there is less than a 10% chance of all isotopes venting to the surface.

Xe^{131m} and Xe^{133} indicate 10% and 12% of simulations show venting at 60 days. Following the 60 days, the probability of Xe^{131m} and Xe^{133} venting increases. This fraction indicates a possibility the elevated levels of xenon at the International Monitoring System station originated from the DPRK nuclear explosion. The results show that 60 days following the

underground explosion is when the peak number of simulations predict venting to occur. At this point, there would be no detectable Xe^{135}. Figure 6.4 shows the results of the fourth DPRK test.

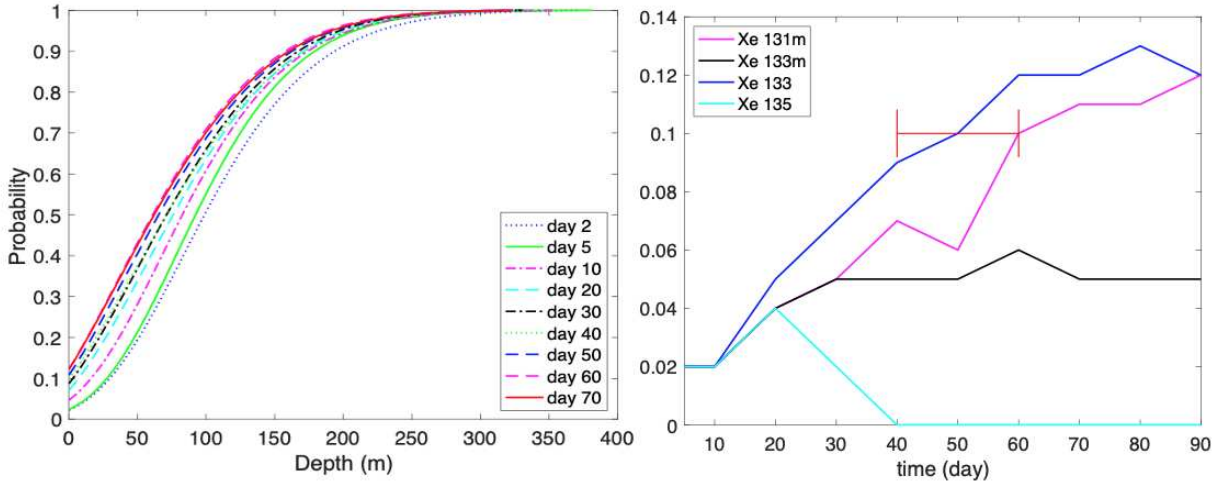

Figure 6.3 Cumulative Distribution of the depth of Xe^{133} (left) and as a function of time for four xenon isostopes (right). The right figure shows that Xe^{135} decays away prior starting after ~ 20 days due to the half -life . The half-life of Xe^{133} is 5.25 days, and the half-life of Xe^{135} is 9.1 hours.

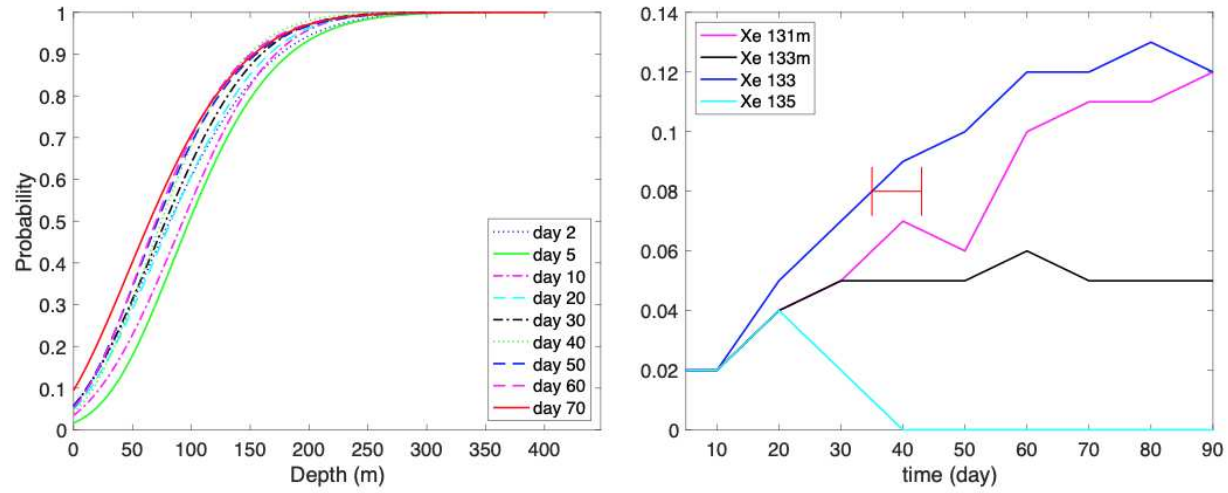

Figure 6.4 Cumulative Distribution for the fourth DPRK underground nuclear test. The cumulative distribution function of the depth of Xe^{133} (left) and as a function of time for four xenon isostopes (right). The right figure shows that Xe^{135} decays away prior starting after ~ 20 days due to the half -life . The half-life of Xe^{133} is 5.25 days, and the half-life of Xe^{135} is 9.1 hours.

For all three of the tests that are likely to have vented xenon gas, Xe^{133} has the highest probability of reaching the surface. This is due to the long half-life and the largest initial concentration. All isotopes will move vertically, however the speed of the vertical movement is

slowed compared to the decay of Xe^{135}. These two isotopes decay away before they reach the surface. In order to detect Xe^{135}, an onsite inspection team would have to take samples at a significant depth.

Figure 6.5 shows a multi isotopic ratio correlation plot for the expected signals at the surface for tests 1, 3 and 4 for the DPRK. The red is for the first test and only goes out to five days post detonation which is when the detection occurred on the border. All four of the xenon isotopes would still be present in detectable quantities. The blue is for the third test and goes until the detectable limit is reached for the shortest half-life would be detectable which is Xe^{133m}. There was only a five percent chance of detecting Xe^{135} prior to it decaying away. The green shows the expected signal for the fourth tests.

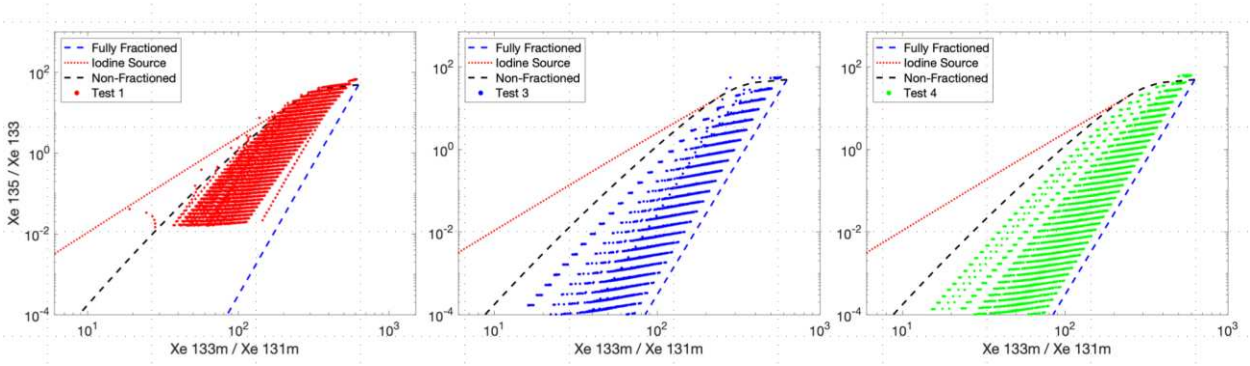

Figure 6.5 Multi isotopic ratio correlation plot of the expected signal from tests 1 (left), 3 (middle) and 4 (right). These are the tests that had elevated levels of xenon and the CTBTO detected elevated levels with the IMS.

The results assume that monitoring is done on-sight. However, when the gas comes to the surface it disperses within the atmosphere. Therefore, it is possible this diffusion and dispersion will decrease the chances of each isotope to be detected at an International Monitoring System station. It is likely each isotope vented at slightly different times and this would lead to a different atmospheric transport scenario.

The results presented here show that the claimed test conducted between 2006 and 2013 would have vented in time for the detection of radioxenon by International Monitoring Stations. However, even for the first test, the likelihood that all four isotopes needed to discriminate a civilian source from a weapon would have been small. The results are also relative to monitoring at site, which was not possible in North Korea. Atmospheric convection would reduce the concentrations making it very unlikely that all four isotopes would be detected offsite

after more than days. These results suggest that the ratios of Xe^{135}/Xe^{133} and Xe^{133m}/Xe^{131m} are of limited use unless the CTBTO gains access to the site. Baring this, mobile detectors that can be moved into proximity with plumbs emitting from the suspected tests site should be considered.

6.4 Conclusions from the Monte Carlo simulation.

The stochastic subsurface gas transport simulation corroborated two out of the three detections by the International Monitoring System for the underground nuclear explosions in the DPRK. The explosions in 2006 indicated a high probability of rapid venting to the surface within days of the explosion. The test conducted in 2013 resulted in a high probability the xenon gas was initially contained and weeks later vented from a process of barometric pumping. The nuclear explosion conducted in 2016 resulted in a low probability of venting for 90 days following the explosion. After this time, radioactive decay is likely to bring the concentrations below detectable levels.

The simulations utilized input parameter estimations that were estimated by remote sensing techniques. The subsurface gas transport model utilized a fast running, simple numerical model that models barometric pumping from the cavity to the surface based upon variations in atmospheric pressure. The atmospheric pressure was the only input parameter that used deterministic values. This was justified based upon looking at the variations in pressure at three stations that surround the location area. All three stations reported a similar variation in atmospheric pressure.

The depth of the minimum detectable level was found after each iteration for each batch run. These depths were fit to a probability distribution function. The cumulative distribution functions of the depth of the minimum detectable limit were utilized to determine the probabilities of the depth of the xenon gas. The cumulative distributions were used with the assumption that the concentration increases as you get closer to the cavity.

The signature of the xenon isotopes could not be corroborated utilizing the Multi-isotopic ratio correlation plots. The International Monitoring System detected only one isotope for two of the tests, and two isotopes for one tests. Plus, the International Monitoring System is located outside of the borders, so this would become a complicated issue of subsurface transport that effects the expected isotopic ratios, but it is unknown how atmospheric transport would effect the isotopic ratios.

Acknowledgements. Special thanks to John McClory for help in finding data and discussions of its use, to Jenifer Shafer for suggestions on test location analysis and to Andrew Osborne for editorial comments and suggestions.

6.5 Description of Underground Transport of Environmental Xenon

The underground transport of environmental xenon (UTEX) models the transport of 5 isotopes of radioxenon. These isotopes are Xe – 131m, Xe – 133m, Xe – 133, Xe – 135m and Xe – 135. The position of the concentration for the each isotope is given by [180]:

$$\varphi_m \frac{\partial C(x,y)}{\partial t} + \frac{\partial}{\partial x}\big(C(x,y)v(x,y)\big) = \frac{\partial}{\partial x}\Big(\varphi_m D \frac{\partial C(x,y)}{\partial x}\Big) - \varphi_m \lambda C(x,y) \quad (6.1)$$

$$\varphi_f \frac{\partial C(x,y)}{\partial t} + \frac{\partial}{\partial y}\big(C(x,y)u(x,y)\big) = \int_0^{\frac{\delta}{2}} \varphi_m \frac{\partial C(x,y)}{\partial t} dx + \frac{\partial}{\partial y}\Big(\varphi_m D \frac{\partial C(x,y)}{\partial y}\Big) - \varphi_f \lambda C(x,y) \quad (6.2)$$

Where φ_m and φ_f are the matrix and fracture porosities. *C(x,y)* is the conecntraion $\{Ci/m^3\}$. *v(x,y)* is the bulk flow velocity {m/s} through the matrix, D $\{m^2/s\}$ is the diffusion coefficient (isotope specific). λ is the decay constant {1/s}. *u(x,y)* is the bulk flow velocity {m/s} in a fracture centered at x, and δ is the fracture spacing {m}. Equation 1 describes the horizontal transport of gas in the matrix at a given height y. Equation S2 describes transport along a fracture where x is taken to be 0.

Equations 6.1 and 6.2 take into account advection, diffusion, and radioactive decay. Figure 6.6 shows a conceptual model for the double porosity fracture matrix system that Eqns 6.1 and 6.2 describe.

The radioisotopes are assumed to be contained within the contaminated matrix shown in Fig. 6.6. The initial concentration of each tracked isotope was uniformly distributed within the contaminated matrix which has width equal to the fracture spacing. This was used to assume a volumetric concentration with the third dimension being equal to the discretized horizontal grid.

The diffusion coefficients in Eqns. 6.1 and 6.2 are adjusted according to the tortuosity. Therefore, the diffusion coefficient is adjusted within the fracture and the matrix with appropriate tortuosity's for the fracture and the matrix. The diffusion coefficient (D) is multiplied by the ratio of porosity divided by tortuosity.

The variations in the atmospheric pressure are dependent upon the differential pressure caused by variations in the barometric pressure at the surface. The pressure response in the system is described by Eqns 6.3 and 6.4 [180].

$$\frac{\partial p(x;y)}{\partial t} = \frac{\partial}{\partial x}(\alpha_m \frac{\partial p(x;y)}{\partial x}) \quad (6.3)$$

$$\frac{\partial p(0,y)}{\partial t} = \frac{\varphi_m}{\varphi_f}\int_0^{\frac{\delta}{2}} \frac{\partial p(x,y)}{\partial t}\,dx + \frac{\partial}{\partial y}\left(\alpha_f \frac{\partial p(0,y)}{\partial y}\right) \tag{6.4}$$

A semicolon in the notation indicates the symbol to its right is held constant. Equations 6.5 and 6.6 show how the pneumatic diffusivities are calculated and are held constant for the whole system [180].

$$\alpha_f = \frac{\delta_f^2}{12}\frac{p_0}{\mu\varphi_f} \tag{6.5}$$

$$\alpha_m = \frac{k_m p_0}{\mu\varphi_f} \tag{6.6}$$

Where p_0 is the mean pressure of the system {Pa}, μ is the dynamic viscosity of air {Pa s}. and k_m is the permeability of the bulk matrix $\{m^2\}$.

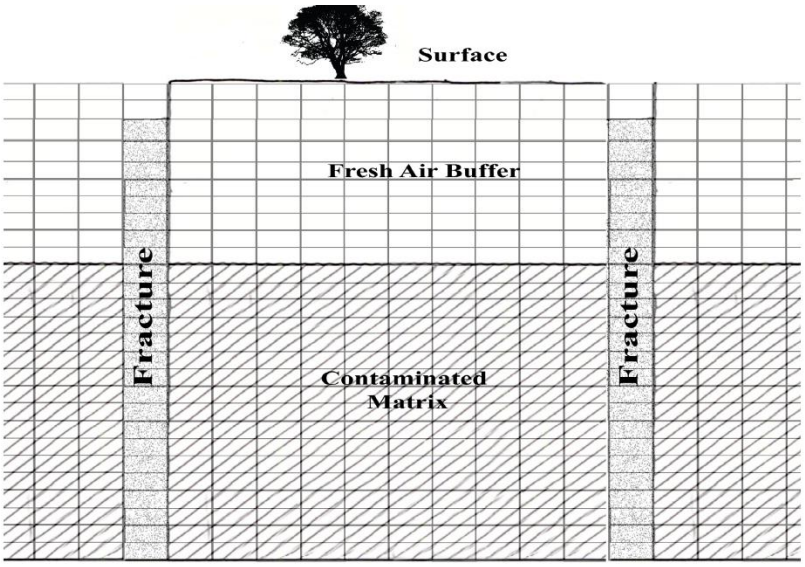

Figure 6.6 Conceptual model for the fracture matrix system. The fracture matrix system is divided into a grid system where x denotes horizontal rows and y denotes vertical columns.

Equations 6.1 – 6.6 are solved using backwards differencing. The solution to the equations for each time step use Gaussian elimination with periodic boundary conditions at the interior of the matrix and a closed bottom boundary. No interaction between vertical layers is assumed. Therefore, each horizontal layer x is solved separately, and again for the fracture [180].

6.6 Comparison of barometric pressure at surrounding sites.

Three stations were considered for atmospheric pressure for the time period post detonation [178]. Ideally, a weather station that is at the site of the explosion would be available

and could be used. However, the pressure change should be similar for weather stations that surround the explosion site. The weather stations that surround the explosion site are Kimchaek, DPRK, Changbai, China and Hamheung, DPRK.

These three stations are the closest weather stations that report atmospheric pressure. Kimchaek is located 40 miles south of the test location, Changbai 40 miles west of the test location, and Hamheung 180 miles southwest. Figure 6.7 shows the reported atmospheric pressures for 92 days following the second test for the three weather stations. From this plot, all three pressure profiles appear to follow a similar pattern. The pressure change is the primary mechanism that will move the gas vertically over time. The change in pressure over time for each weather station is similar. Therefore, the reported pressure from the weather station at Changbai, China is used as the atmospheric pressure input variable for the UTEX model.

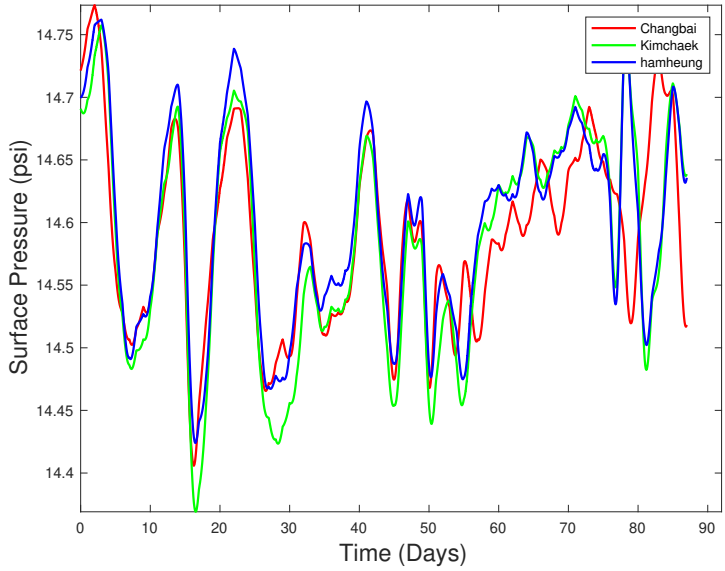

Figure 6.7 Surface pressure for weather stations that surround Punggye-Ri, DPRK. This example goes for 92 days following the second underground nuclear explosion [178].

6.7 Monte Carlo sampling Intermediate steps.

Figure 6.8 depicts a histogram and probability distribution function for third test after 60 days following the nuclear explosion. The histogram shows the depths of the minimum detection limit for the 1,000 iterations. The best fit for this example was the Nakagami distribution [115].

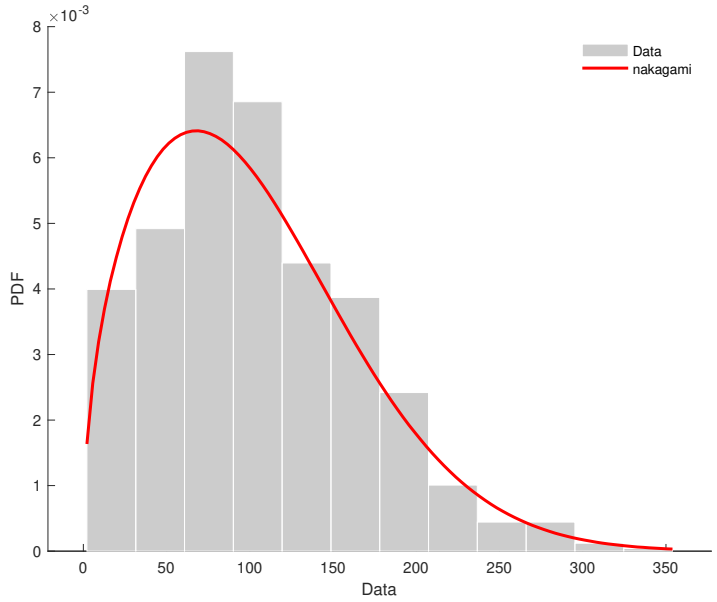

Figure 6.8 Depicts a histogram of the depth of the MDL from the third test after 60 days following the nuclear explosion. This histogram is fit with 17 different probability distribution function's and the best fit is selected. For this particular example, the Nakagami distribution fit the best.

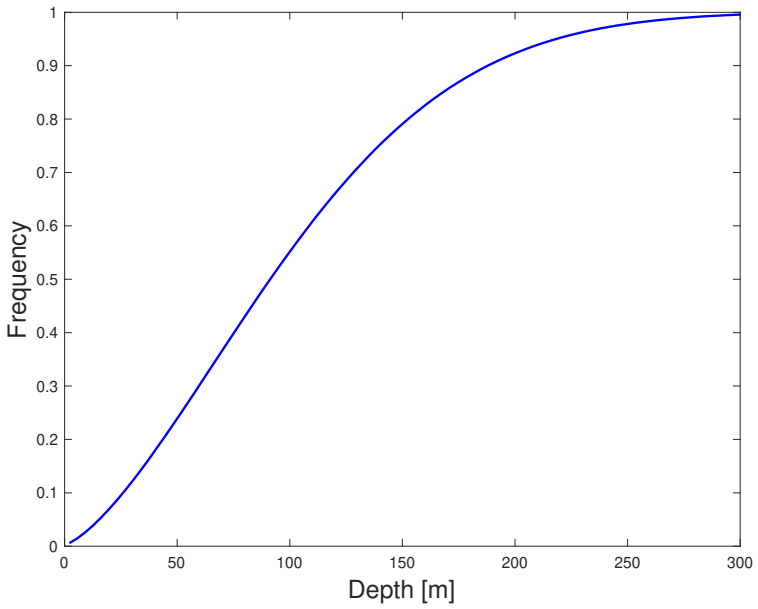

Figure 6.9 The cumulative distribution function developed from probability distribution function depicted in Fig. 6.7 for the third DPRK test after 60 days following the nuclear explosion.

Figure 6.9 depicts a cumulative distribution function created from the probability distribution function (Fig. 6.3). The cumulative distribution function allows us to assign a

probability to the depth of the xenon gas. The cumulative distribution function is useful with the assumption that the concentration increases the deeper you go.

REFERENCES

[1] Q. Li *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *N Engl J Med*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, doi: 10.1056/NEJMoa2001316.

[2] B. Daniel, "Réflexions sur les avantages de l'inoculation," *Mercure de France*, pp. 173–190, 1760.

[3] C. Picquet, *Rapport sur la marche et les effets du choléra-morbus dans Paris et les communes rurales du Département de la Seine*. Royal College of Physicians of London, 1834.

[4] J. Snow, "On the Mode of Communication of Cholera," *Edinburgh Medical Journal*, pp. 668–670, 1855.

[5] R. Ross, "Report on the prevention of Malaria in Mauritius," London, 1908.

[6] W. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. Lond. A*, vol. 115, no. 772, pp. 700–721, Aug. 1927, doi: 10.1098/rspa.1927.0118.

[7] James Holland Jones, "Notes on R0," May 01, 2007.

[8] M. E. J. Newman, "Spread of epidemic disease on networks," *Phys. Rev. E*, vol. 66, no. 1, p. 016128, Jul. 2002, doi: 10.1103/PhysRevE.66.016128.

[9] F. Brauer, "Mathematical epidemiology: Past, present, and future," *Infectious Disease Modelling*, vol. 2, no. 2, pp. 113–127, May 2017, doi: 10.1016/j.idm.2017.02.001.

[10] Z. F. Sun and X. J. Meng, "Antigenic Cross-Reactivity between the Nucleocapsid Protein of Severe Acute Respiratory Syndrome (SARS) Coronavirus and Polyclonal Antisera of Antigenic Group I Animal Coronaviruses: Implication for SARS Diagnosis," *Journal of Clinical Microbiology*, vol. 42, no. 5, pp. 2351–2352, May 2004, doi: 10.1128/JCM.42.5.2351-2352.2004.

[11] X. Che *et al.*, "Antigenic Cross-Reactivity between Severe Acute Respiratory Syndrome–Associated Coronavirus and Human Coronaviruses 229E and OC43," *J INFECT DIS*, vol. 191, no. 12, pp. 2033–2037, Jun. 2005, doi: 10.1086/430355.

[12] A. Grifoni, "Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals," p. 33.

[13] Q. Lin, L. Zhu, Z. Ni, H. Meng, and L. You, "Duration of serum neutralizing antibodies for SARS-CoV-2: Lessons from SARS-CoV infection," *Journal of Microbiology, Immunology and Infection*, p. S168411822030075X, Mar. 2020, doi: 10.1016/j.jmii.2020.03.015.

[14] S. Sanche, Y. T. Lin, C. Xu, E. Romero-Severson, N. Hengartner, and R. Ke, "High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2," *Emerg. Infect. Dis.*, vol. 26, no. 7, Jul. 2020, doi: 10.3201/eid2607.200282.

[15] A. Lai, A. Bergna, C. Acciarri, M. Galli, and G. Zehender, “Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2,” *J Med Virol*, vol. 92, no. 6, pp. 675–679, Jun. 2020, doi: 10.1002/jmv.25723.

[16] X. He *et al.*, “Temporal dynamics in viral shedding and transmissibility of COVID-19,” *Nat Med*, Apr. 2020, doi: 10.1038/s41591-020-0869-5.

[17] H. Nishiura *et al.*, “Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19),” Epidemiology, preprint, Feb. 2020. doi: 10.1101/2020.02.03.20020248.

[18] Z. Du, X. Xu, Y. Wu, L. Wang, B. J. Cowling, and L. A. Meyers, “Serial Interval of COVID-19 among Publicly Reported Confirmed Cases,” *Emerg. Infect. Dis.*, vol. 26, no. 6, Jun. 2020, doi: 10.3201/eid2606.200357.

[19] R. Wölfel *et al.*, “Virological assessment of hospitalized patients with COVID-2019,” *Nature*, Apr. 2020, doi: 10.1038/s41586-020-2196-x.

[20] J. Zhang *et al.*, “Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study,” *The Lancet Infectious Diseases*, p. S1473309920302309, Apr. 2020, doi: 10.1016/S1473-3099(20)30230-9.

[21] S. A. Lauer *et al.*, “The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application,” *Ann Intern Med*, Mar. 2020, doi: 10.7326/M20-0504.

[22] K. H. Grantz, M. S. Rane, H. Salje, G. E. Glass, S. E. Schachterle, and D. A. T. Cummings, “Disparities in influenza mortality and transmission related to sociodemographic factors within Chicago in the pandemic of 1918,” *Proc Natl Acad Sci USA*, vol. 113, no. 48, pp. 13839–13844, Nov. 2016, doi: 10.1073/pnas.1612838113.

[23] B. D. Dalziel *et al.*, “Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities,” *Science*, vol. 362, no. 6410, pp. 75–79, Oct. 2018, doi: 10.1126/science.aat6030.

[24] M. Schläpfer *et al.*, “The scaling of human interactions with city size,” *J. R. Soc. Interface*, vol. 11, no. 98, p. 20130789, Sep. 2014, doi: 10.1098/rsif.2013.0789.

[25] R. A. Neher, R. Dyrdak, V. Druelle, E. B. Hodcroft, and J. Albert, “Potential impact of seasonal forcing on a SARS-CoV-2 pandemic,” *Swiss Med Wkly*, Mar. 2020, doi: 10.4414/smw.2020.20224.

[26] W. Guan *et al.*, “Clinical Characteristics of Coronavirus Disease 2019 in China,” *N Engl J Med*, p. NEJMoa2002032, Feb. 2020, doi: 10.1056/NEJMoa2002032.

[27] CDC COVID-19 Response Team *et al.*, “Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) — United States, February 12–March 16, 2020,” *MMWR Morb. Mortal. Wkly. Rep.*, vol. 69, no. 12, pp. 343–346, Mar. 2020, doi: 10.15585/mmwr.mm6912e2.

[28] “Coronavirus Resource Center,” *Johns Hopkins University of Medicine*. https://coronavirus.jhu.edu/ (accessed Apr. 01, 2020).

[29] K. Mizumoto, K. Kagaya, A. Zarebski, and G. Chowell, “Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020,” *Eurosurveillance*, vol. 25, no. 10, Mar. 2020, doi: 10.2807/1560-7917.ES.2020.25.10.2000180.

[30] M. Day, “Covid-19: four fifths of cases are asymptomatic, China figures indicate,” *BMJ*, p. m1375, Apr. 2020, doi: 10.1136/bmj.m1375.

[31] H. Streeck, G. Hartmann, M. Exner, and M. Schmid, “Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-ClusterStudy (Gemeinde Gangelt),” Apr. 2020. [Online]. Available: Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-ClusterStudy (Gemeinde Gangelt).

[32] D. Sutton, K. Fuchs, M. D’Alton, and D. Goffman, “Universal Screening for SARS-CoV-2 in Women Admitted for Delivery,” *N Engl J Med*, p. NEJMc2009316, Apr. 2020, doi: 10.1056/NEJMc2009316.

[33] S. M. Moghadas *et al.*, “Projecting hospital utilization during the COVID-19 outbreaks in the United States,” *Proc Natl Acad Sci USA*, vol. 117, no. 16, pp. 9122–9126, Apr. 2020, doi: 10.1073/pnas.2004064117.

[34] K. Prem, A. R. Cook, and M. Jit, “Projecting social contact matrices in 152 countries using contact surveys and demographic data,” *PLoS Comput Biol*, vol. 13, no. 9, p. e1005697, Sep. 2017, doi: 10.1371/journal.pcbi.1005697.

[35] D. F. Gudbjartsson *et al.*, “Spread of SARS-CoV-2 in the Icelandic Population,” *N Engl J Med*, p. NEJMoa2006100, Apr. 2020, doi: 10.1056/NEJMoa2006100.

[36] T. C. Jones *et al.*, “An analysis of SARS-CoV-2 viral load by patient age,” p. 19.

[37] R. M. May, “Uses and Abuses of Mathematics in Biology,” *Science*, vol. 303, no. 5659, pp. 790–793, Feb. 2004, doi: 10.1126/science.1094442.

[38] Y. Zhu and Y. Q. Chen, “On a Statistical Transmission Model in Analysis of the Early Phase of COVID-19 Outbreak,” *Stat Biosci*, Apr. 2020, doi: 10.1007/s12561-020-09277-0.

[39] M. L. Holshue *et al.*, “First Case of 2019 Novel Coronavirus in the United States,” *N Engl J Med*, vol. 382, no. 10, pp. 929–936, Mar. 2020, doi: 10.1056/NEJMoa2001191.

[40] R. Li *et al.*, “Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2),” *Science*, p. eabb3221, Mar. 2020, doi: 10.1126/science.abb3221.

[41] U. C. Bureau, “American Community Survey Data,” *The United States Census Bureau*. https://www.census.gov/programs-surveys/acs/data.html (accessed May 01, 2020).

[42] “Urban Dictionary,” *Urban Dictionary*. Urban Dictionary, "Urban Dictionary," 1 July 2019. [Online]. Available: https://www.urbandictionary.com/define.php?term=going+viral. (accessed Jul. 01, 2019).

[43] R. L. F. Coelho, D. S. de Oliveira, and M. I. S. de Almeida, “Does social media matter for post typology? Impact of post content on Facebook and Instagram metrics,” *Online Information Review*, vol. 40, no. 4, pp. 458–471, Aug. 2016, doi: 10.1108/OIR-06-2015-0176.

[44] S. Belew and J. Hoctor, *The art of social selling: finding and engaging customers on Twitter, Facebook, Linkedin, and other social networks*. 2014.

[45] E. M. Rogers, “Diffusion of preventive innovations,” *Addictive Behaviors*, vol. 27, no. 6, pp. 989–993, Nov. 2002, doi: 10.1016/S0306-4603(02)00300-3.

[46] G. Tarde, *The laws of imitation*. Henry Hold and Company, 1903.

[47] C. Mackay, “Memoirs of Extraordinary Popular Delusions and the Madness of Crowds, by Charles Mackay.” https://www.gutenberg.org/files/24518/24518-h/24518-h.htm (accessed May 20, 2020).

[48] K. Back, L. Festinger, B. Hymovitch, H. Kelley, S. Schachter, and J. Thibaut, “A method of studying rumor transmission.,” in *Theory and experiment in social communication.*, Michigan: Research Center for Dynamics Institute for Social Research, 1950, pp. 118–123.

[49] J. Coleman, E. Katz, and H. Menzel, “The Diffusion of an Innovation Among Physicians,” *Sociometry*, vol. 20, no. 4, p. 253, Dec. 1957, doi: 10.2307/2785979.

[50] E. Mansfield, “Technical Change and the Rate of Imitation,” *Econometrica*, vol. 29, no. 4, p. 741, Oct. 1961, doi: 10.2307/1911817.

[51] F. M. Bass, “A New Product Growth for Model Consumer Durables,” *Management Science*, vol. 50, no. 12_supplement, pp. 1825–1832, Dec. 2004, doi: 10.1287/mnsc.1040.0264.

[52] W. Goffman and V. A. Newill, “Generalization of Epidemic Theory: An Application to the Transmission of Ideas,” *Nature*, vol. 204, no. 4955, pp. 225–228, Oct. 1964, doi: 10.1038/204225a0.

[53] J. Woo and H. Chen, “Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog,” *SpringerPlus*, vol. 5, no. 1, p. 66, Dec. 2016, doi: 10.1186/s40064-016-1675-x.

[54] E. Adar and L. A. Adamic, “Tracking Information Epidemics in Blogspace,” in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI’05)*, Compiegne, France, 2005, pp. 207–214, doi: 10.1109/WI.2005.151.

[55] J. Huang and Q. Su, “A rumor spreading model based on user browsing behavior analysis in microblog,” in *2013 10th International Conference on Service Systems and Service Management*, Hong Kong, China, Jul. 2013, pp. 170–173, doi: 10.1109/ICSSSM.2013.6602630.

[56] J. Woo, J. Son, and H. Chen, “An SIR model for violent topic diffusion in social media,” in *Proceedings of 2011 IEEE International Conference on Intelligence and Security Informatics*, Beijing, China, Jul. 2011, pp. 15–19, doi: 10.1109/ISI.2011.5984043.

[57] “Twitter,” *Twitter*. https://developer.twitter.com/en/docs/tweets/search/api-reference/premium-search.

[58] S. Goel, D. J. Watts, and D. G. Goldstein, “The structure of online diffusion networks,” in *Proceedings of the 13th ACM Conference on Electronic Commerce - EC ’12*, Valencia, Spain, 2012, p. 623, doi: 10.1145/2229012.2229058.

[59] S. Ma, L. Feng, and C.-H. Lai, “Mechanistic modelling of viral spreading on empirical social network and popularity prediction,” *Sci Rep*, vol. 8, no. 1, p. 13126, Dec. 2018, doi: 10.1038/s41598-018-31346-0.

[60] D. J. Watts, “A simple model of global cascades on random networks,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 9, pp. 5766–5771, Apr. 2002, doi: 10.1073/pnas.082090499.

[61] D. Kempe, J. Kleinberg, and E. Tardos, “[No title found],” *Theory of Comput.*, vol. 11, no. 1, pp. 105–147, 2015, doi: 10.4086/toc.2015.v011a004.

[62] R. O. Carlson, *ADOPTION OF EDUCATIONAL INNOVATIONS*. 1965.

[63] M. Granovetter, “Threshold Models of Collective Behavior,” *The University of Chicago*, 1978.

[64] J. Leskovec, “Diffusion and Cascading Behavior in Networks,” p. 14.

[65] C. Moore and M. E. J. Newman, “Epidemics and percolation in small-world networks,” *Phys. Rev. E*, vol. 61, no. 5, pp. 5678–5682, May 2000, doi: 10.1103/PhysRevE.61.5678.

[66] M. Gladwell, *The tipping point: how little things can make a big difference*, 1st Back Bay pbk. ed. Boston: Back Bay Books, 2002.

[67] S. Abdullah and X. Wu, “An Epidemic Model for News Spreading on Twitter,” in *2011 IEEE 23rd International Conference on Tools with Artificial Intelligence*, Boca Raton, FL, USA, Nov. 2011, pp. 163–169, doi: 10.1109/ICTAI.2011.33.

[68] A. Friggeri, L. A. Adamic, D. Eckles, and J. Cheng, “Rumor Cascades,” presented at the Weblogs and Social Media, 2014.

[69] S. Vosoughi, D. Roy, and S. Aral, “The spread of true and false news online,” *Science*, vol. 359, no. 6380, pp. 1146–1151, Mar. 2018, doi: 10.1126/science.aap9559.

[70] R. Pfitzner, A. Garas, and F. Schweitzer, “Emotional Divergence Influences Information Spreading in Twitter,” p. 4.

[71] M. Jenders, G. Kasneci, and F. Naumann, “Analyzing and predicting viral tweets,” in *Proceedings of the 22nd International Conference on World Wide Web - WWW ’13 Companion*, Rio de Janeiro, Brazil, 2013, pp. 657–664, doi: 10.1145/2487788.2488017.

[72] M. Thelwall, K. Buckley, and G. Paltoglou, “Sentiment strength detection for the social web,” *J. Am. Soc. Inf. Sci.*, vol. 63, no. 1, pp. 163–173, Jan. 2012, doi: 10.1002/asi.21662.

[73] F. checchi, “Principles of infectious disease transmission,” Mar. 30, 2009, [Online]. Available: https://www.who.int/diseasecontrol_emergencies/publications/idhe_2009_london_inf_dis_transmission.pdf.

[74] M. Martcheva, *An Introduction to Mathematical Epidemiology*, vol. 61. Boston, MA: Springer US, 2015.

[75] "Find minimum of constrained nonlinear multivariable function - MATLAB fmincon." https://www.mathworks.com/help/optim/ug/fmincon.html (accessed May 21, 2020).

[76] P. D. Hoff, *A First Course in Bayesian Statistical Methods*. New York, NY: Springer New York, 2009.

[77] H. Jenkins-Smith, "Twitter Data," 2019.

[78] "List of nuclear weaons tests," *Wikipedia Contributors*. https://en.wikipedia.org/wiki/List_of_nuclear_weapons_tests (accessed Jul. 09, 2019).

[79] "1994-1996: Reaching critical mass: CTBTO Preparatory Commission." https://www.ctbto.org/verification-regime/building-theinternational-monitoring-system/1994-1996-reaching-critical-mass/ (accessed Apr. 28, 2020).

[80] "The final verification measure: CTBTO Preparatory Commission." https://www.ctbto.org/verification-regime/on-site-inspection/the-final-verification-measure/ (accessed Apr. 28, 2020).

[81] S. Glasstone and P. J. Dolan, *The Effects of Nuclear Weapons*, Third Edition. United States Department of Defense and United States Department of Energy, 1977.

[82] T. W. Bowyer *et al.*, "Detection and analysis of xenon isotopes for the comprehensive nuclear-test-ban treaty international monitoring system," p. 13, 2002.

[83] M. B. Kalinowski *et al.*, "Discrimination of Nuclear Explosions against Civilian Sources Based on Atmospheric Xenon Isotopic Activity Ratios," *Pure Appl. Geophys.*, vol. 167, p. 23, 2010.

[84] C. R. Carrigan, R. A. Heinle, G. B. Hudson, J. J. Nitao, and J. J. Zucca, "Trace gas emissions on geological faults as indicators of underground nuclear testing," *Nature*, vol. 382, no. 6591, pp. 528–531, Aug. 1996, doi: 10.1038/382528a0.

[85] J. D. Lowrey, S. R. Biegalski, A. G. Osborne, and M. R. Deinert, "Subsurface mass transport affects the radioxenon signatures that are used to identify clandestine nuclear tests," *Geophysical Research Letters*, vol. 40, no. 1, pp. 111–115, 2013, doi: 10.1029/2012GL053885.

[86] J. D. Lowrey, S. R. Biegalski, and M. R. Deinert, "UTEX modeling of radioxenon isotopic fractionation resulting from subsurface transport," *J Radioanal Nucl Chem*, p. 6, 2013.

[87] A. Ringbom and H. Miley, "Radionuclide Monitoring." Comprehensive Nuclear Test Ban Treaty Organization, [Online]. Available: https://www.ctbto.org/fileadmin/user_upload/pdf/ISS_Publication/Radionuclide_23-28.pdf.

[88] S. C. Myers, G. Johannesson, and W. Hanley, "Incorporation of probabilistic seismic phase labels into a Bayesian multiple-event seismic locator," p. 12, 2009.

[89] J. R. Murphy, B. W. Barker, M. E. Marshall, and W. L. Rodi, "Improved Focal Depth Determination For Use in Seismic Monitoring of the Underground Nuclear Explosions," Science

Applications International Corporation, Hanscom AFB MA, AFRL-VS-HA-TR-2004-1031, Nov. 2003.

[90] S. G. Kim, “Depth estimate of the DPRK’s 2006-10-09, 2009-05-25 and 2013-02-12 underground nuclear tests using local and teleseismic arrays,” *Journal of Asian Earth Sciences*, p. 15, 2018.

[91] D. Bowers and N. D. Selby, “Forensic Seismology and the Comprehensive Nuclear-Test-Ban Treaty,” *Annu. Rev. Earth Planet. Sci.*, vol. 37, no. 1, pp. 209–236, May 2009, doi: 10.1146/annurev.earth.36.031207.124143.

[92] S. K. Arora, T. K. Basu, and C. A. Krishnan, “SOURCE DEPTH AS A USEFUL PARAMETER IN THE DISCRIMINATION OF EARTHQUAKES AND UNDERGROUND EXPLOSIONS,” *Tectonophysics*, vol. 91, pp. 29–52, 1983.

[93] W. N. Junek, J. Roman-Nieves, R. C. Kemerait, M. T. Woods, and J. P. Creasey, “Automated Source Depth Estimation Using Array Processing Techniques,” Air Force Technical Applications Center, Patrick AFB, FL.

[94] F. V. Pabian and J. Liu, “North Korea’s Punggye-ri Nuclear Test Site: No Activity Spotted,” *38 North*, Mar. 15, 2019.

[95] F. V. Pabian, J. Liu, and J. S. Bermudez, “More Potential Questions About the Punggye-ri Nuclear Test Site Destruction,” *38 North*, Jun. 08, 2018.

[96] “The Containment of Underground Nuclear Explosions,” U.S. Congress, Office of Technology Assessment U.S. Government Printing Office, Washington DC, OTA-ISC-414, Oct. 1989.

[97] “North Korea’s Punggye-ri Nuclear Test Site: Analysis Reveals Its Potential for Additional Testing with Significantly Higher Yields | 38 North: Informed Analysis of North Korea,” *38 North*, Mar. 10, 2017. https://www.38north.org/2017/03/punggye031017/ (accessed Jun. 01, 2020).

[98] V. Adushkin V. and W. Lieth, “The Containment of Soviet Underground Nuclear Explosions,” U.S. Geological Survey, Institute of Dynamics of Geospheres of the Russian Academy of Sciences, 01–312, Sep. 2001.

[99] V. Adushkin V. and A. Spivak, “Underground Explosions,” U.S. Department of State, Patrick AFB, FL, ADA627744, Sep. 2015.

[100] F. Ringdal, P. D. Marshall, and R. W. Alewine, “Seismic yield determination of Soviet underground nuclear explosions at the Shagan River test site,” *Geophysical Journal International*, vol. 109, no. 1, pp. 65–77, Apr. 1992, doi: 10.1111/j.1365-246X.1992.tb00079.x.

[101] R.-S. Jih, R. R. Shumway, D. W. Rivers, R. Wagner, and T. W. McElfresh, “MAGNITUDE:YIELD RELATIONSHIP AT VARIOUS NUCLEAR TEST SITES --- A MAXIMUM-LIKELIHOOD APPROACH USING HEAVILY CENSORED EXPLOSIVE YIELDS,” Teledyne Geotech and University of California, Alexandria, VA, GL-TR-90-0107, May 1990.

[102] E. Chael, “Local Magnitudes of Small Contained Explosions,” Sandia National Laboratories, Albuquerque,NM, SAND2009-7941, Jan. 2010.

[103] E. Teller, W. Talley, G. H. Higgins, and G. Johnson, *Constructive Uses of Nuclear Explosives*, First Edition. McGraw-Hill, 1968.

[104] “Need For Increased Management Attention To Certain laboratory Operating Practices In The Nuclear Weapons Testing Program,” United States General Accounting Office, Washington DC, B-165546, Sep. 1971. [Online]. Available: https://www.gao.gov/assets/210/201758.pdf.

[105] “International Seismological Centre.” Accessed: Apr. 27, 2020. [Online]. Available: http://www.isc.ac.uk/.

[106] S. C. Myers, G. Johannesson, and W. Hanley, “A Bayesian hierarchical method for multiple-event seismic location: Bayesian stochastic multiple-event location,” *Geophysical Journal International*, vol. 171, no. 3, pp. 1049–1063, Nov. 2007, doi: 10.1111/j.1365-246X.2007.03555.x.

[107] “Overview of the verification regime: CTBTO Preparatory Commission.” https://www.ctbto.org/verification-regime/background/overview-of-the-verification-regime/ (accessed Apr. 27, 2020).

[108] I. Bondár, S. C. Myers, E. R. Engdahl, and E. A. Bergman, “Epicentre accuracy based on seismic network criteria,” *Geophys J Int*, vol. 156, no. 3, pp. 483–496, Mar. 2004, doi: 10.1111/j.1365-246X.2004.02070.x.

[109] “Seismic Data of Nuclear Tests.” Accessed: Apr. 27, 2020. [Online]. Available: https://www.ldeo.columbia.edu/res/pi/Monitoring/Arch/BRV_arch_exp.html.

[110] G. H. Higgins, “Underground Nuclear Explosions,” Lawrence Radiation Laboratory, University of California, XA04N0741, 1970. [Online]. Available: https://www.osti.gov/servlets/purl/4106885.

[111] *SRTM/version1*. United States Geological Survey, 2003.

[112] F. Beauducel, *READHGT: Import/download NASA SRTM data files*. 2020.

[113] D. W. Scott, “Sturges’ rule,” *WIREs Computational Statistics*, vol. 1, no. 3, pp. 303–306, 2009, doi: 10.1002/wics.35.

[114] R. W. Shonkwiler and F. Mendivil, *Explorations in Monte Carlo methods*. Dordrecht ; New York: Springer, 2009.

[115] F. de Castro, *fitmethis*. 2020.

[116] V. S. Bocharov, S. A. Zelentsov, and V. N. Mikhailov, “Characteristics of 96 underground nuclear explosions at the Semipalatinsk proving ground,” *At Energy*, vol. 67, no. 3, pp. 691–695, Sep. 1989, doi: 10.1007/BF01123212.

[117] “Sandia National Laboratory Official List of UNEs in Nevada.” https://nuclearweaponarchive.org/Usa/Tests/Nevada.html (accessed Jun. 23, 2020).

[118] "Limited Test Ban Treaty (LTBT)," *U.S. Department of State*. //2009-2017.state.gov/t/avc/trty/199116.htm (accessed Jun. 01, 2020).

[119] "The Final Verification Measure," *Comprehensive Nuclear Test Ban Treaty Organization*. https://www.ctbto.org/verification-regime/on-site-inspection/the-final-verification-measure/ (accessed Jul. 10, 2019).

[120] "Overview of the Verification Regime," *Comprehensive Nuclear Test Ban Treaty Organization*. https://www.ctbto.org/verification-regime/background/overview-of-the-verification-regime (accessed Jul. 09, 2019).

[121] Y. Sun and C. R. Carrigan, "Modeling Noble Gas Transport and Detection for The Comprehensive Nuclear-Test-Ban Treaty," *Pure Appl. Geophys.*, vol. 171, no. 3–5, pp. 735–750, Mar. 2014, doi: 10.1007/s00024-012-0514-4.

[122] D. B. Wood, "Digitally Available Interval-Specific Rock Sample Data Compiled from Historical Records, Nevada National Security Site and Vicinity, Ny County, Nevada," Data Series DE-A152-07NA28100, 2007.

[123] "A Way Forward With North Korea: The Comprehensive Test Ban Treaty," *War on the Rocks*, Jun. 11, 2018. https://warontherocks.com/2018/06/a-way-forward-with-north-korea-the-comprehensive-nuclear-test-ban-treaty/ (accessed Jun. 19, 2020).

[124] D. Coblentz and F. Pabian, "Revised Geologic Site Characterization of the North Korean Test Site at Punggye-ri On-line Appendix," p. 13, 2015.

[125] D. Coblentz and F. Pabian, "Geologic Site Characterization of the North Korean Nuclear Test Site at Punggye-ri: A Reconnaissance Mapping Redux," Los Alamos, NM, Nov. 2013.

[126] B. Ghanbarian, A. G. Hunt, R. P. Ewing, and M. Sahimi, "Tortuosity in Porous Media: A Critical Review," *Soil Science Society of America Journal*, vol. 77, no. 5, pp. 1461–1477, Sep. 2013, doi: 10.2136/sssaj2012.0435.

[127] T.-E. Wang, C.-Y. Lin, C.-C. King, and W.-C. Lee, "Estimating Pathogen-specific Asymptomatic Ratios:," *Epidemiology*, vol. 21, no. 5, pp. 726–728, Sep. 2010, doi: 10.1097/EDE.0b013e3181e94274.

[128] M. R. Deinert, A. Dathe, J.-Y. Parlange, and K. B. Cady, "Capillary pressure in a porous medium with distinct pore surface and pore volume fractal dimensions," *Phys. Rev. E*, vol. 77, no. 2, p. 021203, Feb. 2008, doi: 10.1103/PhysRevE.77.021203.

[129] J. Ehlen, "Predicting fracture properties in weathered granite in denied areas," in *Reviews in Engineering Geology*, vol. 14, Geological Society of America, 2001, pp. 61–74.

[130] J. N. Hooker, S. E. Laubach, and R. Marrett, "A universal power-law scaling exponent for fracture apertures in sandstones," *Geological Society of America Bulletin*, vol. 126, no. 9–10, pp. 1340–1362, Sep. 2014, doi: 10.1130/B30945.1.

[131] B. B. Mandelbrot, *The fractal geometry of nature*. New York: W.H. Freeman, 1983.

[132] T. Gleeson *et al.*, “Mapping permeability over the surface of the Earth: MAPPING GLOBAL PERMEABILITY,” *Geophys. Res. Lett.*, vol. 38, no. 2, p. n/a-n/a, Jan. 2011, doi: 10.1029/2010GL045565.

[133] H. H. Dürr, M. Meybeck, and S. H. Dürr, “Lithologic composition of the Earth’s continental surfaces derived from a new digital map emphasizing riverine material transfer: LITHOLOGY COMPOSITION OF EARTH’S CONTINENTAL SURFACES,” *Global Biogeochem. Cycles*, vol. 19, no. 4, p. n/a-n/a, Dec. 2005, doi: 10.1029/2005GB002515.

[134] J. Hartmann and N. Moosdorf, “The new global lithological map database GLiM: A representation of rock properties at the Earth surface: TECHNICAL BRIEF,” *Geochem. Geophys. Geosyst.*, vol. 13, no. 12, Dec. 2012, doi: 10.1029/2012GC004370.

[135] X. Kuang and J. J. Jiao, “An integrated permeability-depth model for Earth’s crust: Permeability of Earth’s crust,” *Geophys. Res. Lett.*, vol. 41, no. 21, pp. 7539–7545, Nov. 2014, doi: 10.1002/2014GL061999.

[136] S. F. Jr. Clark, R. B. Moore, E. W. Ferguson, and M. Z. Picard, “Criteria and methods for fracture-trace analysis of the New Hampshire bedrock aquifer,” United States Geological Survey, OFR 96-479, 1996.

[137] J. Ehlen, “Comparison of Air Photo Lineations and Joint Patterns, Dartmoor, Southwest England:,” Defense Technical Information Center, Fort Belvoir, VA, May 1992. doi: 10.21236/ADA252017.

[138] K. Bisdom, G. Bertotti, and H. M. Nick, “The impact of different aperture distribution models and critical stress criteria on equivalent permeability in fractured rocks: THE IMPACT OF FRACTURE APERTURE ON FLOW,” *J. Geophys. Res. Solid Earth*, vol. 121, no. 5, pp. 4045–4063, May 2016, doi: 10.1002/2015JB012657.

[139] R. Marrett, O. Ortega, and C. Kelsey, “Extent of power-law scaling for natural fractures in rock,” *Geology*, vol. 27, no. 9, pp. 799–802.

[140] C. A. Barton and M. D. Zoback, “Self-similar distribution and properties of macroscopic fractures at depth in crystalline rock in the Cajon Pass Scientific Drill Hole,” *J. Geophys. Res.*, vol. 97, no. B4, p. 5181, 1992, doi: 10.1029/91JB01674.

[141] L. A. Gomez and S. E. Laubach, “Rapid digital quantification of microfracture populations,” *Journal of Structural Geology*, vol. 28, no. 3, pp. 408–420, Mar. 2006, doi: 10.1016/j.jsg.2005.12.006.

[142] T. S. Miranda *et al.*, “Quantifying aperture, spacing and fracture intensity in a carbonate reservoir analogue: Crato Formation, NE Brazil,” *Marine and Petroleum Geology*, vol. 97, pp. 556–567, Nov. 2018, doi: 10.1016/j.marpetgeo.2018.07.019.

[143] T. Gleeson, N. Moosdorf, J. Hartmann, and L. P. H. van Beek, “A glimpse beneath earth’s surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity,” *Geophys. Res. Lett.*, vol. 41, no. 11, pp. 3891–3898, Jun. 2014, doi: 10.1002/2014GL059856.

[144] “World Imagery.” https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9 (accessed Jun. 19, 2020).

[145] J. Ehlen, “A proposed method for characterizing fracture patterns in denied areas,” in *Reviews in Engineering Geology*, vol. 13, Geological Society of America, 1998, pp. 151–164.

[146] B. Wang, Y. Jin, Q. Chen, J. Zheng, Y. Zhu, and X. Zhang, “DERIVATION OF PERMEABILITY–PORE RELATIONSHIP FOR FRACTAL POROUS RESERVOIRS USING SERIES–PARALLEL FLOW RESISTANCE MODEL AND LATTICE BOLTZMANN METHOD,” *Fractals*, vol. 22, no. 03, p. 1440005, Sep. 2014, doi: 10.1142/S0218348X14400052.

[147] J. Cai, W. Wei, X. Hu, R. Liu, and J. Wang, “FRACTAL CHARACTERIZATION OF DYNAMIC FRACTURE NETWORK EXTENSION IN POROUS MEDIA,” *Fractals*, vol. 25, no. 02, p. 1750023, Apr. 2017, doi: 10.1142/S0218348X17500232.

[148] E. Bonnet *et al.*, “Scaling of fracture systems in geological media,” *Rev. Geophys.*, vol. 39, no. 3, pp. 347–383, Aug. 2001, doi: 10.1029/1999RG000074.

[149] J. N. Hooker, S. E. Laubach, and R. Marrett, “Fracture-aperture size—frequency, spatial distribution, and growth processes in strata-bounded and non-strata-bounded fractures, Cambrian Mesón Group, NW Argentina,” *Journal of Structural Geology*, vol. 54, pp. 54–71, Sep. 2013, doi: 10.1016/j.jsg.2013.06.011.

[150] M. B. Clark, S. L. Brantley, and D. M. Fisher, “Power-law vein-thickness distributions and positive feedback in vein growth,” *Geology*, vol. 23, no. 11, pp. 975–978, Nov. 1995, doi: 10.1130/0091-7613(1995)023<0975:PLVTDA>2.3.CO;2.

[151] M. R. Gross and T. Engelder, “Strain accommodated by brittle failure in adjacent units of the Monterey Formation, U.S.A.: scale effects and evidence for uniform displacement boundary conditions,” *Journal of Structural Geology*, vol. 17, no. 9, pp. 1303–1318, Sep. 1995, doi: 10.1016/0191-8141(95)00011-2.

[152] L. Gong *et al.*, “Characterization and Prediction of Complex Natural Fractures in the Tight Conglomerate Reservoirs: A Fractal Method,” *Energies*, vol. 11, no. 9, p. 2311, Sep. 2018, doi: 10.3390/en11092311.

[153] K. J. W. McCaffrey, R. E. Holdsworth, J. Pless, B. S. G. Franklin, and K. Hardman, “Basement reservoir plumbing: fracture aperture, length and topology analysis of the Lewisian Complex, NW Scotland,” *Journal of the Geological Society*, pp. jgs2019-143, Jun. 2020, doi: 10.1144/jgs2019-143.

[154] B. Cox and J. S. Y. Wang, “Single fracture aperture patters: Characterization by slit-island fractal analysis,” International High-level radioactive waste management conference, Las Vegas, NV, 1993.

[155] J. McCord, “Phase I Contaminant Transport Parameters for the Groundwater Flow and Contaminant Transport Model of Corrective Action Unit 97: Yucca Flat/Climax Mine, Nevada Test Site, Nye County Nevada,” U.S Department of Energy, Las Vegas, NV, S-N/99205--096, Sep. 2007.

[156] L. B. Prothro, “Analysis of Fractures in Cores from the Tuff Confining Unit beneath Yucca Flat, Nevada Test Site,” DOE/NV/25946--351, Mar. 2008.

[157] Donald Sweetkind and Ronald M. Drake II, “Characteristics of Fault Zones in Volcanic Rocks Near Yucca Flat, Nevada Test Site, Nevada,” OFR 2007-1293, 920108, Nov. 2007. doi: 10.2172/920108.

[158] S. L. Jr. Drellack, L. B. Prothro, and K. E. Roberson, “Analysis of fractures in volcanic cores from Pahute Mesa, Nevada Test Site,” Washington DC, DOE/NV/11718--160, Sep. 1997.

[159] J. N. Rinker and P. A. Cori, “Air Photo Analysis, Photo Interpretation Logic, and Feature Extraction,” p. 97.

[160] J. Ehlen, “Predicting fracture properties in weathered granite in denied areas,” in *Reviews in Engineering Geology*, vol. 14, Geological Society of America, 2001, pp. 61–74.

[161] D. A. Silva *et al.*, “Fractal dimension and Shannon’s entropy analyses of the architectural complexity caused by the inflammatory reactions induced by highly crystalline poly(vinyl alcohol) microspheres implanted in subcutaneous tissues of the Wistar rats,” *J. Biomed. Mater. Res.*, vol. 101A, no. 2, pp. 326–339, Feb. 2013, doi: 10.1002/jbm.a.34334.

[162] B. Mukherjee and P. N. S. Roy, “Fractal analysis of logs to characterize the hydrocarbon and non-hydrocarbon zones of Bhogpara oil field, Northeast India,” *Arab J Geosci*, vol. 10, no. 22, p. 497, Nov. 2017, doi: 10.1007/s12517-017-3282-8.

[163] A. Pola, G. Crosta, N. Fusi, V. Barberini, and G. Norini, “Influence of alteration on physical properties of volcanic rocks,” *Tectonophysics*, vol. 566–567, pp. 67–86, Sep. 2012, doi: 10.1016/j.tecto.2012.07.017.

[164] L. Luo, J. Jin, W. Wei, and J. Cai, “An analysis of fractal dimension and tortuosity based on 2D numerical reconstruction model of reservoir rocks,” *Interpretation*, vol. 7, no. 4, pp. SJ1–SJ6, Nov. 2019, doi: 10.1190/INT-2019-0033.1.

[165] “2017 Sept DPRK: CTBTO Preparatory Commission.” https://www.ctbto.org/the-treaty/developments-after-1996/2017-sept-dprk/ (accessed Jun. 23, 2020).

[166] P. De Meutter, J. Camps, A. Delcloo, and P. Termonia, “Source localisation and its uncertainty quantification after the third DPRK nuclear test,” *Sci Rep*, vol. 8, no. 1, p. 10155, Dec. 2018, doi: 10.1038/s41598-018-28403-z.

[167] “CTBTO detects radioactivity consistent with 12 February announced North Korean nuclear test: CTBTO Preparatory Commission.” https://www.ctbto.org/press-centre/press-releases/2013/ctbto-detects-radioactivity-consistent-with-12-february-announced-north-korean-nuclear-test/ (accessed Jun. 23, 2020).

[168] “Technical Findings: CTBTO Preparatory Commission.” https://www.ctbto.org/the-treaty/developments-after-1996/2016-dprk-announced-nuclear-test/technical-findings/ (accessed Jun. 23, 2020).

[169] “The CTBT verification regime put to the test - the event in the DPRK on 9 October 2006: CTBTO Preparatory Commission.” https://www.ctbto.org/press-centre/highlights/2007/the-ctbt-verification-regime-put-to-the-test-the-event-in-the-dprk-on-9-october-2006/ (accessed Jun. 23, 2020).

[170] A. Becker, G. Wotawa, A. Ringbom, and P. R. J. Saey, “Backtracking of Noble Gas Measurements Taken in the Aftermath of the Announced October 2006 Event in North Korea by Means of PTS Methods in Nuclear Source Estimation and Reconstruction,” *Pure Appl. Geophys.*, vol. 167, no. 4–5, pp. 581–599, May 2010, doi: 10.1007/s00024-009-0025-0.

[171] A. Ringbom *et al.*, “Measurements of radioxenon in ground level air in South Korea following the claimed nuclear test in North Korea on October 9, 2006,” *J Radioanal Nucl Chem*, vol. 282, no. 3, pp. 773–779, Dec. 2009, doi: 10.1007/s10967-009-0271-8.

[172] P. De Meutter, J. Camps, A. Delcloo, and P. Termonia, “Assessment of the announced North Korean nuclear test using long-range atmospheric transport and dispersion modelling,” *Sci Rep*, vol. 7, no. 1, p. 8762, Dec. 2017, doi: 10.1038/s41598-017-07113-y.

[173] J. Medalia, “North Korea’s 2009 Nuclear Test: Containment, Monitoring, Implications,” Congressional Research Service Report to Congress 7–5700, Nov. 2010.

[174] C. R. Carrigan, Y. Sun, and M. D. Simpson, “The characteristic release of noble gases from an underground nuclear explosion,” *Journal of Environmental Radioactivity*, vol. 196, pp. 91–97, Jan. 2019, doi: 10.1016/j.jenvrad.2018.10.015.

[175] C. R. Carrigan *et al.*, “Delayed signatures of underground nuclear explosions,” *Sci Rep*, vol. 6, no. 1, p. 23032, Sep. 2016, doi: 10.1038/srep23032.

[176] N. A. Duncan and M. R. Deinert, “Method to characterize the geology of denied access sites,” *In Process*, 2020.

[177] N. A. Duncan, M. F. Piazzoli, and M. R. Deinert, “Bayesian Data Sythesis Improves Depth of Burst Estimates for Nuclear Tests at Denied Access Sites,” *in process*, 2020.

[178] N. C. for E. Information (NCEI), “Integrated Surface Dataset (Global).” https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00532 (accessed May 31, 2020).

[179] “International Seismological Centre.” http://www.isc.ac.uk/ (accessed Apr. 27, 2020).

[180] J. D. Lowrey, “Subsurface radioactive gas transport and release studies using the UTEX model,” Aug. 2013, Accessed: Apr. 27, 2020. [Online]. Available: https://repositories.lib.utexas.edu/handle/2152/21571.

.