Analysis of the Infection Time from a Potential H7N9 Influenza Pandemic Outbreak

Walter Silva-Sotillo[1,2], MSc., Mingyang Li[2], PhD., Tapas K. Das[2], PhD.
Pontificia Universidad Catolica del Peru, Peru, [2]University of South Florida, USA
silvasotillo@mail.usf.edu, mingyangli@usf.edu, das@usf.edu

Abstract - Avian influenza viruses have been affecting human populations for a long time since the outbreak in the year 1580 as the first recorded in history. Since then, other mutations and reassortments of the influenza viruses (e.g., H1N1, H3N2) have emerged causing pandemics. Recent emergence of H7N9 influenza virus in China resulted in 1307 laboratory-confirmed cases of human infections causing 489 deaths (37.4%fatality rate). Researchers have developed early esti-mates of some of the epidemiological parameters to characterize H7N9 virus in China. In this research we examine the distribution that characterizes the time to infection from a potential H7N9 influenza pandemic outbreak using results from an agent-based (AB) simulation model. The AB model replicates the dynamics of contacts between susceptibles an infected individuals. We considered some of the common con-tinuous probability distributions and conclude, based on the negative log-likelihood, that the lognormal dis-tribution provides a good fit to characterize the time to be infected.

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2017.1.1.288
ISBN: 978-0-9993443-0-9
ISSN: 2414-6390

Analysis of the Gnfection Rime from a Potential H7N9 Gnfluenza Nandemic Mutbreak

Walter Silva-Sotillo[1,2], MSc., Mingyang Li[2], PhD., Tapas K. Das[2], PhD.

[1]Pontificia Universidad Catolica del Peru, Peru, [2]University of South Florida, USA

silvasotillo@mail.usf.edu, mingyangli@usf.edu, das@usf.edu

Abstract

Avian influenza viruses have been affecting human populations for a long time since the outbreak in the year 1580 as the first recorded in history. Since then, other mutations and reassortments of the influenza viruses (e.g., H1N1, H3N2) have emerged causing pandemics. Recent emergence of H7N9 influenza virus in China resulted in 1307 laboratory-confirmed cases of human infections causing 489 deaths (37.4%fatality rate). Researchers have developed early esti-mates of some of the epidemiological parameters to characterize H7N9 virus in China. In this research we examine the distribution that characterizes the time to infection from a potential H7N9 influenza pandemic outbreak using results from an agent-based (AB) simulation model. The AB model replicates the dynamics of contacts between susceptibles an infected individuals. We considered some of the common continuous probability distributions and conclude, based on the negative log-likelihood, that the lognormal distribution provides a good fit to characterize the time to be infected.

1 Introduction

AH7N9 is a subtype of influenza A virus that is found commonly in birds and poultry. Since March of 2013, A(H7N9) has been noted to infect humans in several regions of China, especially those who are in close contact with poultry either at farms or at markets dealing with poultry. So far, there has been five waves of infections since March 2013. A total of 1307 laboratory confirmed cases of A(H7N9) infections have been recorded in several regions of China causing 489 deaths [1]. The outbreaks have occurred in relatively densely populated regions of China that have over 54% of the country population. The average age of those infected in the first three waves have been reported to be 61, 57, and 56, respectively. However, the relative high age of those infected has not been attributed as an epidemiological characteristics of the A(H7N9) virus. Instead, it is conjectured to be a function of the higher level of exposure to poultry for elderly men in particular. Though most of the reported infections are known to be isolated cases of animal to human transmissions, researchers have noted exceptions where they believe human-to-human transmission may have occurred. However, it is concluded that there is still a lack of sustained evidence of human-to-human transmission [2].

It is feared though that A(H7N9) influenza virus could gain the ability to mutate or reassort to become human-to-human transmittable and cause a pandemic. Similar situation happened with H5N1 during the years 2003-2009 where scientists believed that H5N1 was highly likely to become human-to-human transmittable and cause a pandemic. Though an H5N1 pandemic did not occur as yet, this virus is still in circulation and, as reported by WHO, has caused 145 infections and 42 deaths in three countries in 2015 [3].

An important observation made so far about A(H7N9) is that, though it is highly pathogenic both in humans and birds, infected poultry remain asymptomatic and do not die. This makes it difficult

Digital Object Identifier (DOI): http://dx.doi.org/10.18687/LACCEI2017.1.1.288
ISBN: 978-0-9993443-0-9
ISSN: 2414-6390

to identify the spread of A(H7N9) among poultry. In recent years, the Chinese government has applied containment measures to limit the spread of the virus including culling birds and closing live poultry-markets and trading areas.

The objective of the research underlying the content of this paper was to estimate the time to be infected from a potential A(H7N9) influenza pandemic outbreak. We used data from recent reports and an updated version of our previously developed AB simulation model [4, 5, 6] to simulate an outbreak and estimate which distribution provides better understanding of the time to be infected. The simulation model replicates the dynamics of pandemic outbreak in a selected area incorporating the demographic information (households, schools, workplaces, and communities), human behavior (including contacts, compliance to quarantine and other public health measures, and travel), epidemiological parameters of the virus (e.g., force of infection, incubation and latent periods, basic reproduction number (R_0), and fatality rate), and non-pharmaceutical intervention strategies (NPI) for containment and mitigation. The AB model considers detailed information about the households and their member compositions (age, sex, work, parental status), distance between individuals and their daily movements, contact processes, infection process, and disease natural history. We assumed that during the outbreaks, an *ad-hoc* or an *optimal* NPI strategy (which was found as optimal for a generic influenza virus) was in place with a goal to contain the spread. The NPI strategy comprised measures like isolation, quarantine, school and workplace closures.

The remaining part of the paper is organized as follows. In Section 2, we describe the methods used in this research. Section 3 presents the results. Section 4 outlines the conclusions and Section 5 discusses important points from this research and propose future work.

2 Methods

In this section we will describe the disease natural history, the AB simulation model, the interventions and the statistical modeling of the infection time.

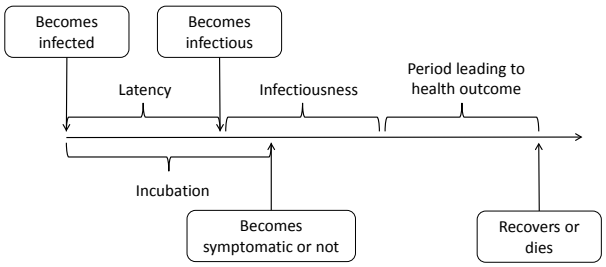

Figure 1: Typical influenza disease natural history showing the progression of the disease from the moment of exposure until health outcome.

Once a person becomes infected, s/he starts a period of latency. During this period, infected individuals cannot infect other persons. After this, the person becomes infectious and can spread virus to susceptibles individuals in a symptomatic or asymptomatic way which is defined by the incubation period. Finally, after the infectiousness period is over the person either recovers and becomes immune or dies. Figure 1 describes the influenza disease natural history. That schema is used in the AB model to simulate the contact and infection process for each individual.

2.1 Agent-based simulation model

The agent-based (AB) simulation model provides output data to be used to analyze the behavior of the time to be infected. Figure 2 depicts the process used in the AB model. Individuals and households are generated and associated. Then schools, workplaces, community locations and schedules are also assigned to the persons. After generating all population, some infected individuals are released in the model to trigger a pandemic. The model keeps

daily trace of susceptible and infected individuals and checks for contacts in households, workplaces/schools and community places.

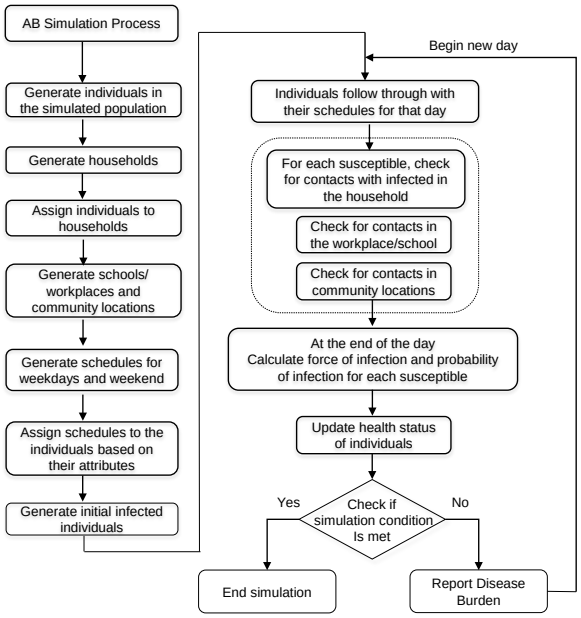

Figure 2: Schematic of the AB model.

By the end of the day, the model calculates the force of infection for each susceptible individual, calculates the probability of infection and update the health status for each susceptibles person. After having no more infected and reaching the simulation condition the model stops and generates output files containing relevant daily statistics (e.g.,number of infected persons, number of deaths, number of persons that visit the doctor, number of persons recovered, etc.).

2.2 Non-pharmaceutical interventions

We assume that during an outbreak, a NPI strategy will be in place with a goal to contain the spread (that is, to keep the reproduction number $R_0 < 1$ and infection attack rate $IAR < 0.1$). A NPI strategy comprises measures like social distancing, isolation, quarantine, school and workplace closure, and travel restrictions. Consequently, there can be numerous possible strategies based on the chosen parameter values for the measures. We implement two NPIs strategies, one *ad-hoc* and the other that was recommended in [5]. We refer to the strategies as NPI(1) and NPI(2). Table 1 shows the 16 factors involved in the NPI and their corresponding values used in the AB model.

Global thresold is associated with the number of cases needed to declare an outbreak of influenza. *Deployment Delay* is the time needed to fully deploy NPIs after the onset of an outbreak. *Case isolation* is related to isolate an infected individual at home. *Household quarantine* measures the restriction to leave the house to the household members of an infected person. *School closure* defines the number of students infected in a class to close a class, the number of classes closed to close a school and the duration of such closure. *Workplace closure* requires the number of cases to close a department in a workplace, the percentage of departments closed to close the workplace and the duration of the workplace closure.

Factor	Intervention	NPI(1)	NPI(2)
1	Global Threshold	10	10
2	Deployment delay	3 days	7 days
3	Case isolation threshold	1 day	1 day
4	Case isolation duration	7 days	10 days
5	Case isolation compliance for workers	75%	75%
6	Case isolation compliance for non-workers	84%	57%
7	Household quarantine threshold	1 day	1 day
8	Household quarantine duration	7 days	7 days
9	Household quarantine compliance workers	75%	53%
10	Household quarantine compliance non-workers	84%	84%
11	Cases to close a class in a school	4	1
12	Classes to close a school	6	3
13	School clsure duration	10 days	21 days
14	# cases to close a department in a workplace	6	3
15	% of departments to close a workplace	60%	30%
16	Workplace closure duration	10 days	7 days

Table 1: Parameters for two NPI strategies

2.3 Statistical modeling of infection time

Due to the right-skewness of infection time data, in this section, we investigated different survival distributions, such as exponential, weibull and log-normal, to characterize the time to infection data. The last two distributions are more flexible in representing the various heavy-tailed infection time data. All results are implemented in the statistical computing environment of R software [7, 8].

The maximum likelihood estimation (MLE) is employed to estimate the distribution parameters by maximizing the data likelihood (LIK) function described below:

$$LIK = [F(t_1)^{r_1}]\{\prod_{i=2}^{m}[F(t_i) - F(t_{i-1})]^{r_i}\} \quad (1)$$

where $F(t_i)$ is the cumulative number of infected persons up to time i, r_i is the number of infections in interval i, n is the total number of infected persons and m is the day when there is no more infected individuals. In our case, m can vary depending on the scenario evaluated and our analysis of infections considers only days with infected persons.

2.3.1 Exponential distribution

The exponential distribution is the simplest distribution in the analysis of reliability/survival data and has a constant hazard. Totlife is defined as the time elapsed before getting infected.

$$Totlife = \sum((start + end)/2 * infected) \quad (2)$$

The exponential hazard rate (i.e., the instantaneous probability of being infected) and the corresponding mean time to be infected (MTTI) are presented in equations 2 and 3.

$$\lambda = \sum(infected/Totlife) \quad (3)$$

$$MTTI = 1/\lambda \quad (4)$$

2.3.2 Weibull distribution

Weibull distribution is a generalization of the exponential distribution and is often used in biomedical applications. It can be used to model infection data with a decreasing, a constant or an increasing hazard rate.

Based on the MLE, distribution parameters of weibull, namely the rate parameter α and the shape parameter β, can be simultaneously estimated. The correponding MTTI is given by,

$$MTTI = \alpha\Gamma(1 + 1/\beta) \quad (5)$$

2.3.3 Lognormal distribution

The lognormal distribution offers flexible shapes for the probability density functions (pdfs) and hazard rate functions. This distribution is preferred when there is a hypothetical multiplicative and progressive increments that could trigger a critical event, e.g., the ocurrence of infection. In the context of pandemic outbreaks, such progressive increments can be interpreted as the increasing number of infections during the first days of pandemics. This means also an increasing number of contacts that a susceptible individual faces which feeds the amount of virus ingested and hence, it affects the probability of getting infected.

If $T \sim lognormal(T_{50}, \sigma)$, then $lnT \sim N(lnT_{50}, \sigma)$, where T_{50} is the median infection time for the population of lognormal infection time observations and σ is the standard deviation of the logarithmic transformation of infection times (i.e., shape parameter). Equivalent, if $X \sim N(\mu, \sigma)$, then $e^x \sim lognormal(e^\mu, \sigma)$. Both parameters can be estimated as follows.

$$T_{50} = e^\mu \quad (6)$$

$$\sigma = ln(e^{\sigma\phi^{-1}F(t)})/(\phi^{-1}F(t)) \quad (7)$$

where $F(t)$ is the cumulative distribution function (cdf) of the lognormal distribution. The MTTF is estimated from:

$$MTTI = T_{50} e^{\sigma^2/\mu} \tag{8}$$

2.4 Problem description and data

The data comes from the AB model that use epidemiological information from H7N9 outbreak in China and combined that with U.S. demographical parameters to replicate a potential H7N9 outbreak in the U.S. One of the outputs of the AB simulation model is the daily number of infected persons which is used in this study. This value is considered up to the day where no further infections are detected.

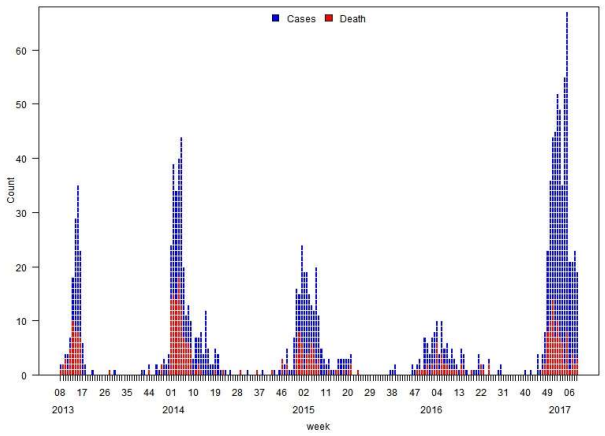

Figure 3: Laboratory-confirmed cases and deaths of human infection with A(H7N9) reported by WHO.

The data used in this research was simulated for a population of 1.27 million of persons (Hillsborough county area in Tampa Bay) considering a moderate force of infection. The number of incidences of H7N9 in China has been reducing since 2013 but researchers think that a potential pandemic can occur at any time and more dangerous is that the virus mutate or reassort to a human-to-human transmittable. Figure 3 shows the number of cases since initial outbreak.

We used readout data consisting of *interval*, *start*, *end* and *infected*. This data is obtained from the AB model output report and is used in the three statistical distributions proposed.

Finally, this study considers 6 scenarios associated with the variation on the NPI used (as described in section 2.2) and the potential force of spread which is related with the IAR, beeing 33%, 50% and 65% the values more considered in the literature.

3 Results

In the present project we considered three main distributions as potential candidates to describe the time to infection from an influenza pandemic outbreak: exponential, weibull and log-normal. After running the AB model, we obtained as output the daily number of infected cases. This information is used as readout data to proceed estimating the main parameters, the LIK and the MTTI for each case.

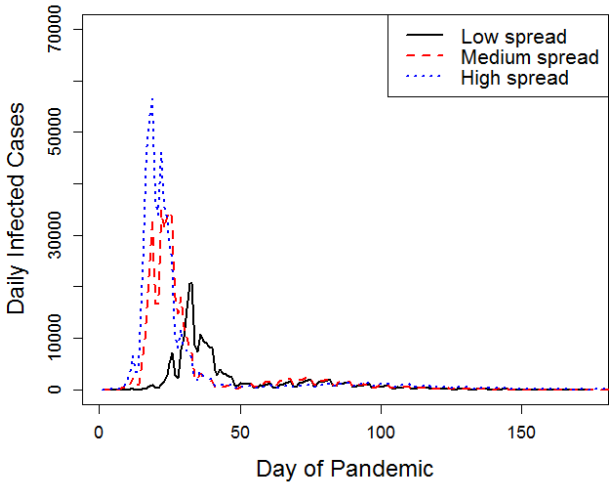

Figure 4: Simulated daily infected cases for NPI(1).

Simulated infected cases are shown in figure 4 and

figure 5 for NPI(1) and NPI(2) respectively and for the three levels of spread. From figure 3, for a given year, each one of the waves displayed seems not to differ significantly from the shape obtained as result from our AB simulation model depicted in figures 4 and 5. It is important to notice that in figure 5, we observe the presence of waves in all spread cases. Our procedure takes in consideration all time frame, hence all waves in the estimations. Parameters estimated for the exponential (λ), weibull (α, β) and lognormal (T_{50}, σ) distributions are respectively displayed in Table 2 for NPI(1) and in Table 3 for NPI(2).

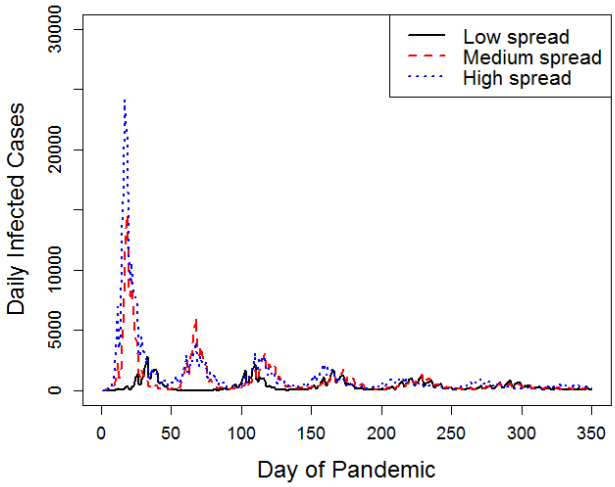

Figure 5: Simulated daily infected cases for NPI(2).

Spread	Distribution	Parameters	MTTI	LIK
Low	Exponential	0.00069	1440.05	-1439.40
	Weibull	(707.99,0.51)	1366.07	-1343.747
	Lognormal	(5.51,2.21)	96.40	-1342.92
Medium	Exponential	0.00038	2642.62	-1607.19
	Weibull	(978.35,0.47)	2208.31	-1462.15
	Lognormal	(5.80,2.67)	334.72	-1455.18
High	Exponential	0.00036	2794.97	-1992.63
	Weibull	(866.38,0.46)	2046.32	-1777.91
	Lognormal	(5.69,2.26)	107.37	-1765.95

Table 2: Estimated parameters for each distribution and for each force of spread when NPI(1) is deployed.

Spread	Distribution	Parameters	MTTI	LIK
Low	Exponential	0.00270	369.46	-2419.22
	Weibull	(338.52,0.85)	368.29	-2410.31
	Lognormal	(5.16,1.39)	16.75	-2418.44
Medium	Exponential	0.001135	880.82	-2723.30
	Weibull	(727.24,0.81)	816.77	-2706.31
	Lognormal	(6.06,1.08)	11.58	-2646.97
High	Exponential	0.000785	1273.83	-2852.42
	Weibull	(1039.47,0.81)	1167.44	-2834.98
	Lognormal	(6.45,1.03)	11.39	-2766.91

Table 3: Estimated parameters for each distribution and for each force of spread when NPI(2) is deployed.

4 Discussion

To estimate the distribution parameters of the mentioned three distributions, maximum likelihood function in (1) is maximized based on numerical optimization methods. For exponential distribution, single parameter is estimated based on the $Nelder-Mead$ method [9]. For Weibull and Lognormal distributions, $Broyden-Fletcher-Goldfarb$ and *Shanno* (BFGS) method [10] is considered to estimate the two-dimensional parameters. The estimation results for three parametric survival distributions are implemented via *fitdistr* function in *R*.

The negative log-likelihood (LIK) is considered as the goodness-of-fit measure to evaluate and compare the performance among different distributions. A smaller value of LIK indicates a better goodness-of-fit. Among all 6 scenarios, Exponential distribution gives the largest LIK values and thus exhibits the poorest goodness-of-fit. It is mainly due to its less flexibility in representing different shapes of infection data.

As depicted in table 3, the weibull distribution outperforms other distribution with better goodness-of-fit in the low spread scenario. However, its corresponding LIK value is not far away from its counterpart of lognormal distribution. As shown in all scenarios, the estimated shape parameters of weibull distribution is less than 1. It implies a decreasing hazard rate over time. In the context of

pandemic outbreak, it indicates that the infection rate is decreasing over time and eventually the pandemic will stop. The estimation results of weibull distribution also explain the reason why exponential distribution exhibits unsatisfactory goodness-of-fit. In exponential distribution, hazard rate is constant. It indicates that the inflection rate will not change over time, which fails the capture the actual pandemic outbreak process with dynamically changing inflection rate.

Among 5 out of 6 scenarios, lognormal distribution gives the most superior performance of goodness of-fit. There are mainly two reasons. From the data fitting perspective, comparing to weibull distribution, lognormal distribution has similar capability and flexibility to represent various heavy tails and right-skewed inflection time data. From the pandemic outbreak process perspective, lognormal distribution can more closely mimic the underlying process, and thus gives satisfactory results. In addition, it is noticed that goodness-of-fit results of weibull and lognormal distributions are similar and lognormal exhibits slightly better results. The hazard rate of lognormal is in complex form and less interpretable compared to weibull distribution. Thus, weibull distribution is still valuable in providing compact and interpretable information (e.g., decreasing infection rate) while maintaining a reasonable well goodness-of-fit of data.

5 Conclusion and future work

An inconvenient for this study is that we do not have the exact time of infection but just the number of people infected daily. Having more detailed information from official sources (e.g., CDC, CDC China, WHO) could help to: i) expand the study including time of death and time of recovery, and ii) have better estimates.

Another important improvement to consider is to disaggregate the analysis (infected and deaths) according to age groups: $< 19yrs$, $20 - 64yrs$ and $65 + yrs$. This could help to the making decision process since H7N9 attacks mainly elder people ($65 + yrs$ group). and other influenza viruses can have different age-targets.

It is also known that not all infected persons are symptomatic and a significant part of the population can stay as asymptomatic (they don't show symptoms, hence they don't know that they are sick). Official reports do not capture those asymptomatic cases that may affect our estimations. We propose to develop some multipliers and incorporate them to better estimate the number of infected persons and to capture the real nature of the infection process.

Finally, as seen in figure 5, distribution of number of infected persons exhibit multi-modal shapes at different waves and none of the proposed distributions can capture the real nature of that behavior. We will address this issue in details by investigating mixture distributions and their variants in the future.

References

[1] World Health Organization, "Summary and assessment, 14 february to 16 march 2017," 2017. [Online; accessed 31-March-2017].

[2] World Health Organization, "Analysis of recent scientific information on avian influenza a(h7n9) virus," 2017. [Online; accessed 11-April-2017].

[3] World Health Organization, "Cumulative number of confirmed human cases for avian influenza a(h5n1) reported to who, 2003-2016," 2016. [Online; accessed 21-April-2017].

[4] A. Uribe-Sánchez, A. Savachkin, A. Santana, D. Prieto-Santa, and T. K. Das, "A predictive decision-aid methodology for dynamic mitigation of influenza pandemics," *OR spectrum*, vol. 33, no. 3, pp. 751–786, 2011.

[5] D. L. Martinez and T. K. Das, "Design of non-pharmaceutical intervention strategies for pandemic influenza outbreaks," *BMC public health*, vol. 14, no. 1, p. 1, 2014.

[6] T. K. Das, A. A. Savachkin, and Y. Zhu, "A large-scale simulation model of pandemic influenza outbreaks for development of dynamic mitigation strategies," *Iie Transactions*, vol. 40, no. 9, pp. 893–905, 2008.

[7] Marie Laure Delignette-Muller, Christophe Dutang, "An r package for fitting distributions," 2014. [Online; accessed 30-March-2017].

[8] Vito Ricci, "Fitting distributions with r," 2005. [Online; accessed 30-March-2017].

[9] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[10] D. F. Shanno, "Conditioning of quasi-newton methods for function minimization," *Mathematics of computation*, vol. 24, no. 111, pp. 647–656, 1970.