Models for Sustainable Population Growth on Mars

Chris Dunlap, Allen Koh, Matt May

CSE 6730 Project 2 Final Report, Spring 2016

Contents

1 Problem Statement and Approach

In the United States, the National Aeronautics and Space Administration (NASA) has announced its plan to send humans to Mars during the 2030s. This ambitious goal requires a variety of studies be conducted to effectively plan the endeavor. General habitation, food production, resource extraction, communication, spacecraft, and many other areas must be studied to determine their optimal configuration.

For our second project, we performed a simulation of population growth dynamics on Mars, with the goal of determining insight into strategies for sustainable population growth. Population growth models have been extensively studied in the literature [8], [7], [21], [6], [12] but generally only in the context of our own planet.

Often, natural populations without resource limitations exhibit exponential growth [5]. However, this type of rapid growth will likely be unsustainable under the extreme resource constraints of Mars. By considering several proposed habitation models for Mars, and modeling the effects of uncertainty, we hope to better understand the viability of these approaches, and by that develop insight into sustainable growth.

2 Related Work

2.1 Population Growth Models

Moore [25] introduces five models of human colonization. The study focuses around expansion of colonies by modeling migration patterns of the population as well as mortality and fertility rates. The five models of colonization mentioned are the matrix model, beachhead model, string of pearls, outpost model and the pulse model. The paper concludes that regardless of population size, low fertility rates and/or high mortality rates will cause colonization to fail.

There has been previous literature on modeling population growth with limited resources. The most popular model used for population growth is the matrix model. Miller et al. [23] use a Leslie matrix model to estimate the annual increase of a gray wolf population. This model takes inputs of survival and fertility rates and is modified for an environment with limited resources. A simple density-dependent matrix model is used based on a discrete time scalar logistic equation with a defined carrying capacity factor. The estimates for the projection matrix, including survival rates, fertility rates, litter size, and carrying capacity, were taken from field studies of real populations. Aikio & Pakkasmaa present additional characteristics to model population growth by linking growth and reproduction rates with an individual's biomass and the number of individuals they interact with [4]. Clark & Innis uses a model that integrates energy and protein relationships for jack rabbit population growth where the limited resource is food [9]. Food intake is controlled by energy balance and gut fill while foraging selection is used to balance nutrients. The growth and reproduction rates have energy and protein requirements while mortality rates are influenced by predation from coyotes and natural causes. Peterson et al. [27] also introduce a matrix model structure around population dynamics of trees in a forest and mentions the ease of computer simulation as a factor behind using the matrix model structure. The forest matrix model predicts population dynamics using vectors of live trees as well as growth and recruitment matrices.

2.2 Food Production

A stochastic model of population growth during the Neolithic transition focused on foragers and farmers is presented by Fedotov [13]. A two-population model is used in which foragers and farmers are modeled separately but maintain a relationship through total population density. Crop production is also modeled by a formula based on soil nutrients and production rate. The density of soil nutrients is modeled as a partial differential equation, taking into account population size and crop production per unit of time. The study discusses the change in food supply as population density increases and farm land degrades; however, there are underlying assumptions that do likely not apply to the case of colonizing Mars (such as erosion and flooding). Despite the model's objectives being different from our own, the modeling of crop production and soil nutrients appears transferable to our application with the proper tailoring.

Fedotov's study [13] suggests the use of phosphorus as the predominant indicator of nutrients in soil. As such, the relationship of phosphorus excreted by human subjects as a function of protein intake [22] can be applied to our problem to quantify the ability to reconstitute soil for farming by using human excrement.

A study by the Food and Agriculture Organization of the United Nations

(FAO) [15] further adds value to our study by providing the amount of animal and plant-based protein consumed by individuals from a multitude of countries. Given our study's focus on NASA, figures from the United States can be gleaned. Other information from the FAO [14] provides figures for crop efficiency, quantifying edible energy and protein per hectare of farming land for a selection of key crops.

Finally, a study related to hydroponics [10] presents a final useful component to the modeling of food our study, where the effectiveness of hydroponic gardening is compared to that of conventional crop growing techniques by analyzing one of the crops found in [14]. At the conclusion of the study, a multiplier is noted that could be used to approximate the amount of food the Mars colony can grow using hydroponics when compared to the amount grown by conventional means. In this light, it may be attractive for the Mars colony to use hydroponics in lieu of conventional farming techniques.

2.3 Mars-Specific Habitation Models

2.3.1 Oxygen Generation

The distance from Earth to Mars varies between roughly 58 million and 400 million km. Because of the great distance, resupply capabilities for human missions to Mars are virtually nonexistent. Due to this challenge, innovative approaches must be taken to regenerate necessary resources in-situ. To this end, the Mars Design Reference Architecture 5.0 [11] proposes the use of an "In-Situ Resource Utilization System" (ISRU) that converts Mars atmosphere into oxygen for both propellant and life support purposes. The plant operates by using electrolyzers that convert carbon dioxide into oxygen and carbon monoxide, which is then vented. A hydrogen feedstock is brought from Earth and reacted with Mars-produced oxygen to generate water. Furthermore, carbon dioxide, nitrogen, and argon that are extracted from the atmosphere of Mars can be employed as a buffer gas for crew breathing.

2.3.2 Power Generation

The lack of known resources on Mars that can be mined for power generation requires either a solar-based power source or a power source transported from Earth [18]. As noted, Mars is roughly 50% farther from the Sun on average relative to the Earth, so only around half of the solar radiation that Earth experiences actually reaches Mars. Thus, the Design Reference Architecture [11] proposes the use of a nuclear fission-based reactor. Of the power sources which can be transported from Earth, a nuclear power source is the only known option that concentrates sufficient energy in a reasonable mass and volume [18]. The reactor operates at a low temperature, allowing stainless steel (which is compatible with Mars' atmosphere) to be used for reactor components. The ISRU plant mentioned previously is a predominant consumer of power (consuming 25 kWe when operating continuously).

2.3.3 Space Agriculture and Waste Processing

The daily resource requirements of humans have been studied through computer simulation [29]. Simulations indicate that a human requires (per day) 855g of food inputs, 4577g of drinking/food preparation water, 128g of water in food, 18,000g of wash/flush water, and 804g of oxygen for food metabolism. In terms of outputs per day, they are separated into three categories: water, solids, and carbon dioxide. Humans produce 3025g of water in urine and feces, 406g of metabolic water (vapor), 1680g of perspiration water (vapor), and 18,000g of wash/flush water. 161g of solids in the form of feces, urine, and sweat solids are produced. 1092g of carbon dioxide from food metabolization is produced. Over the course of a year, this means that the average human is consuming roughly "three times his body weight in food, four times his weight in oxygen, and eight times his weight in drinking water." [24]. Thus, bioregenerative systems are essential to sustainable long-term habitation on Mars.

To enable space agriculture, hyper-thermophilic aerobic composting bacteria have been studied specifically in the context of habitation on Mars [19]. This technology can be implemented as a subsystem that oxidizes inedible biomass/wastes, converting them to fertilizer. High quality compost is an important part of a regenerative food production system, but as noted, must be implemented carefully to avoid the spread of pathogenic bacteria. A "marsh-based waste processing system" [26] has also been studied which exploits the natural ability of aquatic plant/ microbial associations to perform processing of waste. These systems metabolize, or concentrate, pollutants while generating useful biomass growth. Furthermore, aquatic plants can be used as purified water sources by condensing moisture evapotranspired from plant leaves.

In another study [20], a menu for a sustainable human diet on Mars was developed, concluding that a combination of rice, soybeans, sweet potatoes, green-yellow vegetables, silkworm pupa, and loach would fulfill human nutritional requirements.

3 Simulation Description

In population dynamics, birth and death rates are often density-dependent [16]. The birth rate eventually decreases as population size increases, and the death rate begins to increase with population size due to resource unavailability, environmental deterioration, or both. The *logistic growth model* is often used for modeling population growth that does not grow without bound. In continuous time, the logistic growth equation takes the form of the nonlinear differential equation:

$$\frac{dN}{dt} = r\left(1 - \frac{N}{K}\right)N \tag{1}$$

where r is the intrinsic growth rate and K is the carrying capacity, or the maximum population a species can sustain indefinitely. For our simulation, we instead take a discrete-time approach. In discrete time, the logistic model takes the form of the difference equation:

$$N_{t+1} = r\left(1 - \frac{N_t}{K}\right)N_t \tag{2}$$

We use this equation as a key inspiration for our model, but also incorporate multiple stochastic elements. Several simulation properties will be determined through probability distributions and a random number generator will be used to select from the distribution. For example:

- Whether a natural disaster will occur, and if so, the magnitude of the impact
- A best-case age of death by natural causes for each human
- A "gender" attribute for each human that influences the amount of resource consumption that will be necessary for the human to sustain life, as well as the human's waste production

Care will be taken to ensure that modeled stochastic elements are homogeneous stochastic processes.

We intend to model humans as consumer entities, and several types of resources such as food, air, facility, and power availability as resource entities. We take a modular approach to the simulation, in which resource entities are modeled as independent classes and interacted with through object methods. Furthermore, during each simulation run we model the availability of these resources using a stochastic procedure. Also, as David Quammen notes [28], there are four sources of uncertainty to which a population may be subject: demographic, environmental, natural catastrophes, and genetic. We model the occurrence of several of these to provide enhanced realism.

3.1 Inputs

The simulation will be a discrete, time-stepped simulation with one primary input: a stream of newly born humans. For each iteration, the current population size will be noted to determine the size of the population for the next iteration using a modified version of the difference equation given in Eq. 2. The population size will affect the production and consumption of the resource entities in the simulation. For example, food resource consumption will depend on the number of people to be fed and its production will depend on the amount of waste produced, which is also dependent on the number of people. Every timestep, the amount of other resources such as air, facility, and power, to be consumed will also be dependent on the population size. Although not an explicit input to the simulation, the probability of uncertainty events such as natural disasters occurring at any timestep will be a part of the simulation. The occurrence of a natural disaster event will be determined by a stochastic process with a given distribution. The size of the disaster will also be determined by a distribution and will determine the population size and number of resources to be lost.

3.2 Outputs

To determine whether sustainable population growth occurred, simply measuring the difference in population size per iteration will not be suitable as there may be fluctuations in size as the overall population may grow over time (oscillatory behavior). One way to determine population growth regardless of oscillatory behavior would be to target a population size (carrying capacity) and deem the simulation successful if that size is maintained within a range for a certain number of timesteps. Since our focus is on sustainability, which implies smooth, not jagged, growth dynamics, we develop a population size variation metric based on the standard deviation which is used to compare different values (parameters) of the initial population size. We deem the "most sustainable" strategies to be those which have relatively low variation in the population size after some defined number of timesteps, assuming a carrying capacity exists.

Another output could be the ratio of humans who live to their expected lifespan (natural death) against the total population size. A larger ratio would mean that a larger number of people sustained their given life expectancy and this value can be used as a metric to compare between parameters. The total numbers can be tracked through each timestep and the ratio can be calculated once the simulation completes.

There will be several "outputs" from each iteration of the simulation. As the main objective is to simulate population growth dynamics, one output will be the size of the population at a given timestep. This number must be greater than zero for the population to be sustained and will determine the population input for the next iteration. The amount of food resources will also be an output as this will determine how many people will receive food during the next iteration. Depending on the population size, waste production will also be captured as an output which will feed into the rate of food growth. The amount of oxygen (O_2) consumed was also tracked throughout the simulation. To determine this number, we employed NASA's daily human O_2 consumption estimate of 840g [17], multiplying this by the number of the people in the simulation at the current timestep. Furthermore, facility usage and power available will also be outputs and their increase or decrease will be calculated during the timestep. Power consumption data was compiled from the World Bank, which estimates annual per-capita power consumption in the United States to be 12,985 kWh [3].

3.3 Parameters

The primary parameter for our simulation is the initial population size, which we call N. We also parameterize food storage, which can be either be "off" or "on". Food storage allows surplus food from the previous iteration to be considered as part of the available food for the current iteration.

The growth of food will be determined by the amount of human waste, which is dependent on the initial number of humans. The initial population size and resources available will determine how much growth (positive or negative) will occur for each of the entities as the simulation progresses due to their inter-dependencies (consumption and production).

3.4 Driver Loop

The simulation is run through a main driver loop. This contains the logic to execute each simulation iteration. For this simulation, one iteration is equivalent to approximately one month of time. The driver loop acts as the integration point between the different entities (i.e. Population, Facility, Food, etc.) and executes the interactions in a specified order. Before the loop is executed, there are two initial steps. The simulation first initializes all entities - Population, Air, Power, Food, Facility, and Disaster (Disaster is a simple abstraction used for injecting uncertainty into the simulation). The simulation then processes the inputs by creating the Persons for the initial population size and setting the initial Food resources. After these steps are complete, the driver loop executes with the logic listed below:

1. Determine if disaster strikes
2. Determine if facility expansion should occur
3. Create new Persons (i.e., babies due to be born)
4. Remove any Persons that are set to die
5. Calculate the food needed
6. Calculate the food produced
7. Remove Persons, starting with the oldest, if there is not enough food
8. Determine number of babies to be born in nine months' time

For step one, disaster is included to introduce randomness to the simulation and model any natural disasters or accidents that cause deaths as well as affect existing resources. For this simulation, it is estimated that a disaster kills anywhere from one to 20 people and destroys existing food for 10 to 30 people. The number of people and resources to destroy as well as which people are killed are all done through stochastic processes. The uniform distribution is used.

For step two, if enough excess food is produced to feed about 10 more people and the total population is within 10 people of the personnel capacity designated by the Facility, a new "pod" begins construction. A pod is estimated to provide housing for 30 more people and add enough crop area to support those 30 additional people. It is estimated that the pod construction will complete in about three months, and therefore the additional capacity is added only after three iterations. Furthermore, only one pod can be in construction at any given time.

For step three, when new Persons are created, a random death age is assigned, which is chosen from a distribution created from data gathered by the CDC in 2007 [1]. The Person is also randomly given a gender with an equal probability of being male or female. A new Person is only added if there is enough Facility capacity to hold additional people.

For step four, any Persons who are set to die of natural causes (i.e., have reached their death age) are removed from the simulation.

For steps five through seven, the total kilocalories necessary to sustain the existing population are calculated, and then the food produced this iteration based on the previous iteration's fertilizer amount is calculated. If there is an excess of food, then everyone in the current population survives; however, if there is not enough food produced, the driver will remove people from the population, starting with the oldest Person, and continue removing people until only the number of people that can be sustained remain. The logic to start by removing the oldest of the population is that the adults will most likely feed the children first before themselves - a behavior often found in society.

For step eight, the total number of adults between the ages of 18 and 50 are calculated. The average birthrate in the US in 2014 was 13.42 per 1,000 (1.342%) [2], and considering one of the main goals of the population on Mars is to sustain itself, an assumption is made that the birthrate may be slightly higher. The birthrate per iteration is randomly selected between 1% and 2% from a uniform distribution and only the number of adults is considered in this process. The number of people to be born is then stored in the simulation and the Person objects will be created nine months later (nine iterations later).

After this, the driver starts again from step one until the maximum number of iterations defined are completed.

3.5 Assumptions & Simplifications

As there are an enormous amount of phenomena that could be modeled in a simulation like this, care must be taken to focus initially on the most salient features. Thus, many assumptions and simplifications are made to the simulation due to time constraints and to allow for feasible implementation. First, our simulation is concerned with long-term sustainable growth, and thus we assume some initial infrastructure has been put in place which provides power, crops (food), facilities for living, and oxygen. This is consistent with the literature, which suggests that prior to long-term colonization efforts, both autonomous and human-led missions to Mars will occur in which this initial infrastructure will be created.

Another assumption is that water is modeled as part of the food entity as humans require both to sustain life. There is an implicit assumption made that all the resources are consumed in the same way - but not necessarily same amount - by consumer entities. This means that every human is able to consume food and that all waste, regardless of which human it came from, has the same amount of effect as compost for food growth. Furthermore, humans consume the same amount of resources every timestep as defined by their size attribute, as well as excrete the same amount of waste every timestep. Our simulation does not take into account the energy necessary for humans to carry out tasks such as growing or harvesting more crops, building more facilities, and burying the dead.

3.5.1 Food Process Modeling

Modeling the food production cycle involves several interdependent processes, and also represents one of the core challenges with regard to supporting a sustainable population on Mars [19]. The basis of modeling food consumed by the population is found in [13], where the food production is represented using equation 3.

$$q(x,t) = \alpha \cdot \left(\frac{n(x,t)}{n_0 + n(x,t)} \right) \cdot \left(1 - e^{-\beta \cdot F(x,t)} \right) \tag{3}$$

Where:

- $q(x,t)$ is the food production rate $\left(\frac{kg_{Food}}{Person \cdot Year}\right)$
- α is the production rate coefficient $\left(\frac{kg_{Food}}{Person \cdot Year}\right)$

- $n(x,t)$ is the population density at location x and time t $\left(\frac{Person}{m^2}\right)$
- n_0 is the initial population density $\left(\frac{Person}{m^2}\right)$
- $\beta = 200$ is the nutrient parameter $\left(\frac{m^2}{kg_P}\right)$; kg_P is kilograms of Phosphorus, and the value comes from [13]
- $F(x,t)$ is the soil nutrient density $\left(\frac{kg_P}{m^2}\right)$

The intuition captured in equation 3 is the following:

- The production rate coefficient α is the nominal amount of food that can be grown.
- The fraction involving population densities $n(x,t)$ and n_0 captures the intuition of group solidarity; in the Neolithic study, the migration of human settlements was modeled, and tendencies toward group solidarity was rewarded with higher yields, whereas more dispersed people were penalized.
- The final term with the exponential captures how yields would change due to soil fertility; erosion, nutrient depletion from crop production, and nutrient reconstitution are all accounted for in the fertility computation.

Tailoring of Food Production Equation

Due to the nature of our study, equation 3 is tailored to fit our model. First, the population density term is removed entirely; because the population on Mars is assumed to be confined in one facility, the population is always together and doesn't spread out, so no group solidarity metric is needed. Also, there is no reliance upon position like there was in [13], so the dependency upon x is also removed. Equation 4 shows the resultant of these modifications:

$$q(t) = \alpha \cdot \left(1 - e^{-\beta \cdot F(t)}\right) \tag{4}$$

Next, the units of $q(t)$ and α are modified from $\frac{kg_{Food}}{Person \cdot Year}$ to $\frac{kcal}{Month}$ so that they are more relevant to the actual needs of our study's population. The time units are changed to align with the time step in our simulation.

The mass of food produced was changed to its caloric content due to the population's caloric requirements, which vary depending on age and gender as specified in [14]. These parameters are stored in the population, and thus the total caloric needs of the population are much more dynamic than only relying upon the number of individuals. A byproduct of this paradigm shift is that it is improper to compute food production on a per capita basis, so the normalization with respect to population size (i.e. $Person$ in the units of $q(t)$) is also removed.

As a result of the units change, α must be computed in a manner different from [13]. The value for α is computed from information contained in [14], [10], and the *Facility* class in equation 5 (recall the units of α are $\frac{kcal}{Month}$).

$$\alpha = \dot{E}_{crop} \cdot 10^{-5} \cdot \left(\frac{365}{12}\right) \cdot F_h \cdot F_r \cdot A_{crop} \tag{5}$$

Where:

- $\dot{E}_{crop} = 54492.7$ is the edible energy rate of crop production $\left(\frac{kcal}{ha \cdot Day}\right)$; this value is computed below
- 10^{-5} converts ha^{-1} to m^{-2}
- $\left(\frac{365}{12}\right)$ converts Day^{-1} to $Month^{-1}$
- $F_h = 3.86$ is the factor by which hydroponic farming increases production relative to farming the same area, from [10]
- $F_r = 2$ is a factor indicating the number of vertical levels to replicate the hydroponic setup on a given plot of Martian land; in this setup, we model two levels
- A_{crop} is the area of the facility dedicated to crop production, from the *Facility* class (m^2)

In order to quantify the crop's edible energy rate $\dot{E}_{crop}$, a combination of calorie-rich and protein-rich food in [14] that are still high-yielding crops is taken, such that the proportion of calorie intake and protein intake is consistent with what is found in [14]. To determine this proportion, the ratio of calorie requirements divided by protein requirements was computed for each age and gender combination. Then, the smallest ratio was chosen to ensure that enough protein is supplied for any calorie needs. The smallest ratio was found to be 14-16 year old females:

$$R_{nom} = \frac{2150kcal}{46g_{prot}} \approx 46.739\frac{kcal}{g_{prot}} = 46739\frac{kcal}{kg_{prot}} \tag{6}$$

The ratio using similar numbers found in [15] is smaller (requiring more protein), but this is an aggregated figure compared to the specificity found in [14], so it is not used in favor of the higher-granularity data. While other references indicate different dietary needs of a population on Mars [20], for purposes of out simulation's food production, a combination of sweet potatoes and wheat will be used.

To match this ratio, a combination of sweet potatoes (high calorie food) and wheat (high protein food) is found. The numbers below are from [14]:

$$R_{SP} = \frac{70kcal}{1g_{prot}} = 70\frac{kcal}{g_{prot}} = 70000\frac{kcal}{kg_{prot}} \tag{7}$$

$$R_{W} = \frac{40kcal}{1g_{prot}} = 40\frac{kcal}{g_{prot}} = 40000\frac{kcal}{kg_{prot}} \tag{8}$$

The combination was found to be 48.309% sweet potatoes, and 51.691% wheat. With this information, the value of $\dot{E}_{crop}$ may be computed as follows:

$$\dot{E}_{crop} = 0.48309 \times 70000 + 0.51691 \times 40000 = 54492.7\frac{kcal}{ha \cdot Day} \tag{9}$$

One can compute the conversion from energy supplied by the crops to their respective dry mass using figures contained in [14]:

$$K_{m|E} = \frac{0.48309 \times 22}{70000} + \frac{0.51691 \times 14}{40000} = 3.32747 \times 10^{-4}\frac{kg_{Food}}{kcal} \tag{10}$$

This concludes the tailoring of the food production equation and its associated parameters; another equation, relating to soil nutrient density, also must be tailored.

Computation of Soil Nutrient Density

Soil nutrient density is denoted by $F(x,t)$ in [13]. The equation for $F(x,t)$ is not explicitly given by [13]: instead, it is conveyed as a partial differential equation:

$$\frac{\partial F(x,t)}{\partial t} = \xi_1 - \xi_2 \cdot F(x,t) - \gamma \cdot q(x,t) \cdot n(x,t) \tag{11}$$

Where:

- $F(x,t)$ is the soil nutrient density $\left(\frac{kg_P}{m^2}\right)$; as mentioned earlier, kg_P is kilograms of phosphorus
- ξ_1 is the rate soil nutrients regenerate naturally $\left(\frac{kg_P}{m^2 \cdot Year}\right)$
- ξ_2 is the rate of nutrient depletion due to environmental reasons such as erosion, flooding, etc $(Year^{-1})$
- γ is the rate nutrients are depleted due to growing crops $\left(\frac{kg_P}{kg_{Food}}\right)$
- $q(x,t)$ is the food production rate $\left(\frac{kg_{Food}}{Person \cdot Year}\right)$
- $n(x,t)$ is the population density at location x and time t $\left(\frac{Person}{m^2}\right)$

However, several modifications, discussed earlier, were made in light of our simulation environment:

- The dependency upon x is removed, making this an ODE.
- There is no erosion or flooding, so $\xi_2 = 0$.
- The units of q are changed from $\left(\frac{kg_{Food}}{Person \cdot Year}\right)$ to be $\frac{kcal}{Month}$

In addition to the above modifications, $\frac{\partial F}{\partial t}$'s dependency upon q in the discretized time domain is modeled to use the *previous iteration's* q. The value of ξ_1 also changes as a function of time. Thus, the differential equation becomes:

$$\frac{dF(t)}{dt} = \xi_1(t) - \frac{\gamma \cdot K_{m|E} \cdot q(t - \Delta t)}{F_r \cdot A_{crop}} \tag{12}$$

Substituting equation 4 for q, we have:

$$\frac{dF(t)}{dt} = \xi_1(t) - \frac{\gamma \cdot K_{m|E} \cdot \alpha \left(1 - e^{-\beta F(t - \Delta t)}\right)}{F_r \cdot A_{crop}} \tag{13}$$

Where:

- $F(t)$ is the soil nutrient density $\left(\frac{kg_P}{m^2}\right)$

- $\xi_1(t)$ is the rate soil nutrients regenerate naturally $\left(\frac{kg_P}{m^2 \cdot Month}\right)$
- $\gamma = 5.43 \times 10^{-3}$ is the rate nutrients are depleted due to growing crops $\left(\frac{kg_P}{kg_{Food}}\right)$; the value used is from [13]
- $q(t-\Delta t)$ is the food production rate from the previous iteration $\left(\frac{kcal}{Month}\right)$
- $K_{m|E}$ is the conversion factor from calories to dry mass computed in equation 10
- F_r is a factor indicating the number of vertical levels to replicate the hydroponic setup on a given plot of Martian land; value is defined in equation 5
- A_{crop} is the area of the facility dedicated to crop production, from the *Facility* class (m^2)

The primary means of fertilization is the reconstitution of human excrement as soil fertilizer; this is used as the value of ξ_1, in lieu of static values used by [13]. The amount of phosphorus is regarded by [13] as the primary fertilizing material and conduit, so this is the only element considered when assessing the fertility of soil. Thus, the conversion from protein mass consumption of the population (split up by total protein, $m_{Prot|Tot}$, and non-animal based protein, $m_{Prot|Veg}$) to phosphorous mass excrement in human waste found in [22] is utilized:

$$m_P = 0.11 \left(m_{Prot|Tot} + m_{Prot|Veg}\right) \tag{14}$$

Given the food source of our study does not involve animal-based protein, $m_{Prot|Tot} = m_{Prot|Veg}$. Also, because this computation is performed each iteration of time, this equation is used as a mass flow rate to become as follows:

$$\dot{m}_P = 0.22 \cdot \dot{m}_{Prot|Veg} \tag{15}$$

Where:

- $\dot{m}_P$ is the mass flow rate of phosphorus produced by the population per iteration $\frac{kg_P}{Month}$
- $\dot{m}_{Prot|Veg}$ is the mass flow rate of protein ingested by the population per iteration $\frac{kg_{Prot}}{Month}$

A substitution of $\dot{m}_{Prot|Veg}$ is made using the consumption of crops by the population during the previous iteration and the ratio of calories to protein ingested from equation 6:

$$\dot{m}_P = 0.22\frac{\dot{E}_{Pop}(t - \Delta t)}{R_{nom}} \tag{16}$$

Where:

- $\dot{E}_{Pop}(t - \Delta t)$ is the energy consumption of the population during the previous iteration (kcal)
- R_{nom} is the ratio of calories eaten to mass of protein eaten $\left(\frac{kcal}{kg_{Prot}}\right)$

The phosphorous harvested from excrement is assumed to be distributed uniformly throughout the crop area, so to arrive at the value for ξ_1, the amount of phosphorous harvested is divided by the area dedicated to farming:

$$\xi_1(t) = \frac{\dot{m}_P}{A_{crop}} = 0.22\frac{\dot{E}_{Pop}(t - \Delta t)}{R_{nom} \cdot A_{crop}} \tag{17}$$

Combining equations 13 and 17, the final fertilization rate of the soil becomes:

$$\frac{dF(t)}{dt} = \frac{0.22 \cdot \dot{E}_{Pop}}{A_{crop} \cdot R_{nom}} - \frac{\gamma \cdot K_{m|E} \cdot \alpha \left(1 - e^{-\beta F(t-\Delta t)}\right)}{F_r \cdot A_{crop}} \tag{18}$$

This is the fertilization rate over the course of one time step, so the formula for $F(t)$ at the current time step may be found:

$$F(t) = F(t - \Delta t) + \Delta t \cdot \frac{dF(t)}{dt} \tag{19}$$

Finally, the initial condition $F(t_0)$ is needed; this is able to be found utilizing the assumption stated earlier in Section 3.5, where the requisite amount of crops are present at simulation time t_0. This means that the soil fertilization level is exactly enough to feed the population at t_0, and $\dot{E}_{pop}(t_0)$ is the exact amount of consumption by the population to meet each individual's caloric needs:

$$q\left(t_0\right) = \dot{E}_{Pop}(t_0) = \alpha\left(e^{-\beta \cdot F(t_0)}\right) \tag{20}$$

Solving for $F(t_0)$, the initial condition becomes:

$$F(t_0) = \frac{-1}{\beta} \cdot \ln\left(1 - \frac{\dot{E}_{Pop}(t_0)}{\alpha}\right) \tag{21}$$

Note that the energy needs of the population $\dot{E}_{Pop}(t_0)$ must be checked against the crop's maximum output α; if the population needs more than the crops can provide, then this is an invalid condition (the natural log will return a complex value due to a negative argument), and the population cannot survive or thrive.

4 Simulation Architecture

The simulation is broken down into several pieces implemented as classes in the Python language. The classes contain data dependencies as shown in Figure 1.

Consumer entities are the following classes:

- *Population*, containing instances of the *Person* class
- *Food*, which is both a consumer and resource

Resource entities are the following classes:

- *Air*
- *Facility*
- *Food*
- *Power*

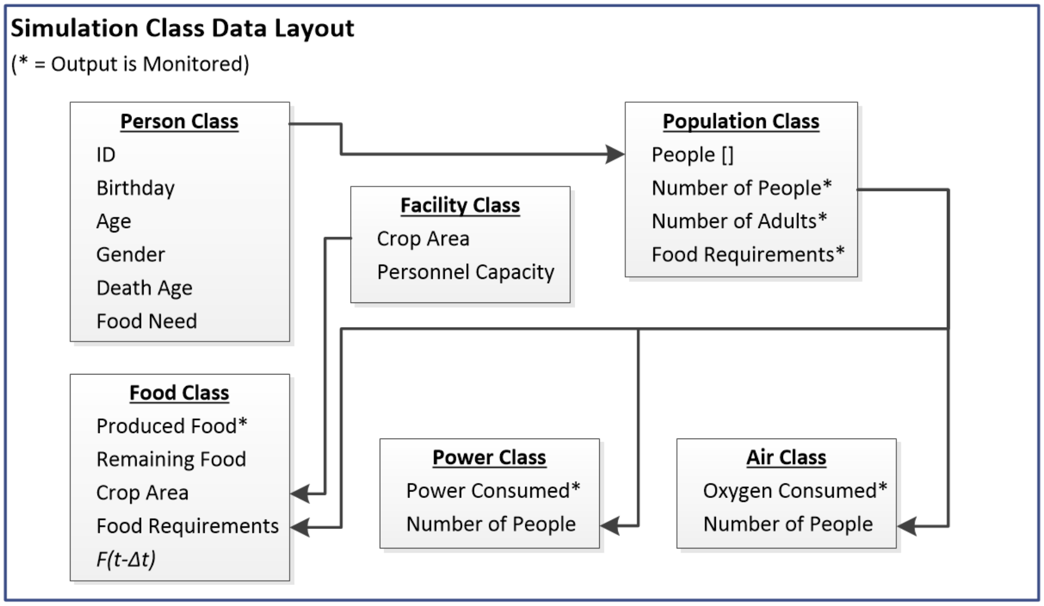

Figure 1: Simulation Data Layout.

A description of each of the aforementioned classes follows.

4.1 *Person* Class

The *Person* class is meant to encapsulate the relevant information about each person in the population. Any information that is modeled on an individual basis is stored at this level, as shown in Table 1.

Attribute	Description
ID	Unique identifier of the person.
Birthday	The person's birthday in simulation time, for the computation of their age.
Age	The person's age at the current simulation time, relative to their birthday.
Gender	The person's gender; this is used to compute their food need.
Death Age	The age at which the person is to die; this is set upon initialization.
Food Need	The amount of food (in $\frac{kcal}{month}$) required by the person per unit time for the person to survive.

Table 1: Attributes of the *Person* class.

4.2 *Facility* Class

The *Facility* class describes the facility in which the population lives. Living quarters, food production, and any other modeled aspects of where the personnel live are part of this class, shown in Table 2.

Attribute	Description
Crop Area	Area dedicated to the growing of crops for food (in m^2).
Personnel Capacity	The maximum number of people that the facility has space for.

Table 2: Attributes of the *Facility* class.

4.3 *Population* Class

The *Population* class encapsulates properties of the population as a whole, primarily the number of people present and instances of each person. The full list of attributes in the *Population* class are in Table 3.

Attribute	Description
People	Array of *Person* instances comprising the individuals of the population.
Number of People	Number of people in the population.
Number of Adults	Number of people at least 18 years of age in the population.
Food Requirements	The amount of food required by the population to survive (in $\frac{kcal}{month}$).

Table 3: Attributes of the *Population* class.

4.4 *Food* Class

The *Food* class has attributes that reflect both dependencies on other classes to compute food output, in addition to the food output itself. The attributes for the *Food* class are in Table 4.

Attribute	Description
Produced Food	The amount of food produced at a given time step.
Remaining Food	The amount of food remaining at a given time step.
Crop Area	The area dedicated to crop production (from the *Facility* class).
Food Requirements	The amount of food required by the population to survive (from the *Population* class)
$F(t - \Delta t)$	The soil fertilization level from the previous iteration.

Table 4: Attributes of the *Food* class.

The key functions belonging to the *Food* class are below. with an in-depth description following.

4.4.1 Constructor

The constructor of the *Food* class takes as input the facility (to get the area dedicated to crops in m^2) and the calorie requirements of the population for the first time step $\dot{E}_{Pop}(t_0)$. Then, the value of $F(t_0)$ is computed using equation 21. The *Remaining Food* attribute are set to the value of $\dot{E}_{Pop}(t_0)$, and $F(t_0)$ is stored into the $F(t - \Delta t)$ attribute for the next iteration. The *Produced Food* attribute is set to zero because exactly enough food was prepared for the population.

4.4.2 *update_food* Function

This function computes the amount of food produced in a given iteration using equation 4. The function takes as input the food requirements of the population, *Food Requirements*, at the current iteration to see if the food produced is sufficient for the population.

First, the value of $\xi_1(t)$ is computed using equation 17, where $\dot{E}_{Pop}(t-\Delta t)$ is the difference between the *Produced Food* and *Remaining Food* attributes; this subtraction yields the amount of food consumed by the population over the previous iteration.

Next, equation 18 is used to compute the change of soil fertility between time steps, followed by equation 19 to compute $F(t)$, the fertilization level at the current iteration.

Finally, this value of $F(t)$ is plugged into equation 4 to arrive at the crop production $q(t)$ for the current iteration; this is saved into the *Produced Food* attribute. After this, $F(t)$ is stored into the $F(t - \Delta t)$ attribute for the next iteration.

Next, the *Remaining Food* attribute is set to the difference of the produced food and the population's food requirements, but constrained to be a number greater than or equal to zero; $RemainingFood = max(ProducedFood - FoodRequirements, 0)$. In this way, the next call to *update_food* will incorporate the correct information for $\dot{E}_{Pop}(t_0)$.

The subtraction of $ProducedFood$ and $FoodRequirements$ is returned, even if the value is negative; the driver then uses this information to adjust the population accordingly based on whether or not the entire population was fed.

4.5 *Air* Class

The *Air* class monitors the air consumption of the population to provide resource utilization insight for scoping out ISRUs for the purpose of colonizing Mars.

Attribute	Description
Oxygen Consumed	Amount of air consumed by the population per time step (in kg of O_2 per month).
Number of People	Number of people in the population (from the *Population* class).

Table 5: Attributes of the *Air* class.

4.6 *Power* Class

The *Power* class monitors the power consumption of the population to provide resource utilization insight for scoping out power generators for the purpose of colonizing Mars.

Attribute	Description
Power Consumed	Amount of power consumed by the population per time step (in kWh per month).
Number of People	Number of people in the population (from the *Population* class).

Table 6: Attributes of the *Power* class.

5 Verification & Validation

5.1 Verification

Throughout development of the model, team members routinely reviewed code that they did not write; this ensured that outside perspectives and fresh sets of eyes reviewed the code. Code is very modular and compartmented to avoid a multitude of programmatic pitfalls.

Before running Monte Carlo sets using the model to attain the derived scalar output values (DSOV), individual runs were made with special scrutiny placed on intermediate simulation states returned in the command window. This facilitated rapid identification and fixing of bugs in a scalable simulation, and instilled confidence that the simulation was running as expected on the larger batch run set.

Visualization was a critical aspect of verifying the simulation model, and the development team regularly reviewed visualization outputs, in addition to carrying out real-time visual inspection of runs as they happened.

Random number generation was done using built-in statistical packages to avoid implementation errors of a home-brewed random number generator. One such generator from the statistical package underwent a goodness-of-fit test in Project 1, so it will be assumed that other distributions generated by the same statistical package will also yield satisfactory outputs.

5.2 Validation

Literature of comparable modeling efforts, in conjunction with relevant studies containing relevant findings used in our simulation, was extensively leveraged to define details regarding our model. A comprehensive description of takeaways from literature may be found in Section 2.

Since this is not a readily observable system and Mars colonization has not actually occurred, there is a deficiency of information to use for model validation with real or observed data. As such, face validation was performed on individual subcomponents of the simulation by their respective authors to ensure no erratic behavior of the model was experienced.

Again, visualization was utilized to ensure the simulation of key population parameters matched the authors' expectations based on the expected behavior of subcomponents they implemented. Any observed phenomena had reasonable explanations for occurrence.

Finally, only practical and achievable inputs were utilized to avoid subjecting the system to unrealistic input parameters, which opens the possibility of causing erratic behavior that would not realistically occur.

6 Results

6.1 Preliminary Observations

To obtain preliminary results that would help us understand the relationships between the various entities, we performed a simulation run with an initial population of $N = 50$ humans, over 1500 timesteps (with each timestep corresponding to one month). We first made the assumption that unconsumed food could be stored and used later, so that food left over from one iteration would be available as a surplus for the following iteration. We also assume there to be a single "facility" in which the population resides. We assume this facility has a max personnel capacity (carrying capacity), which can dynamically increase as certain conditions are met by way of new construction. Results from a sample run are shown in Fig. 2. The x-axis for the figure corresponds to simulation timesteps.

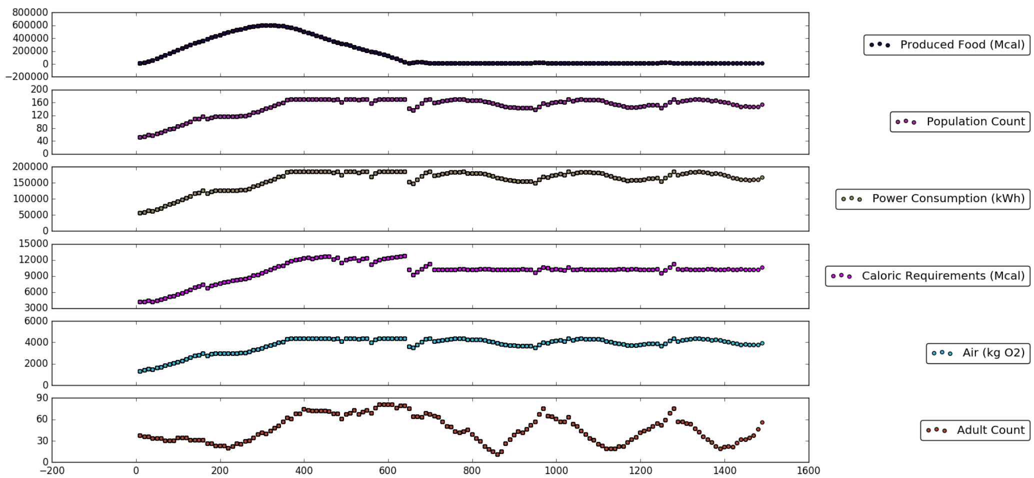

Figure 2: Results from a sample simulation run with food storage enabled ($N = 50$).

Initially, food produced by the population advances faster than the rate at which it is being consumed, creating a surplus of available food resources. However, as the population increases and children come of age, the amount of surplus decreases, and a shortage occurs, resulting in the death of some adults in the simulation. We make the assumption that, in the case of a food shortage, adults choose to prioritize giving food to children. This assumption can result in relatively few numbers of adults being left in the simulation, as seen in 2 around roughly month 850. Further work may be needed to determine how to handle food shortages with enhanced realism.

From Fig. 2, it appears that after an initial warm-up period in which a food surplus exists, the amount of produced food reaches a steady-state value after approximately 650 months (54 years). Furthermore, the population count remains roughly constant after approximately 350 months (30 years). After approximately 700 months (58 years), the adult population essentially oscillates, with food shortages leading to the death of adults in the simulation, following by a gradual coming of age of children and increase in the adult population, and so forth.

Throughout the simulation, O_2 consumption, power consumption, and caloric requirements track closely with the population count, as expected. O_2 and power consumption are simply scalar multiples of the total population count, and caloric requirements are closely related to the population size, with gender and age also being considered in the computation.

To explore the effects of eliminating our assumption that food could be stored, we removed food storage capabilities and then re-ran the simulation with the same initial conditions ($N = 50$ people). The results are shown in Fig. 3.

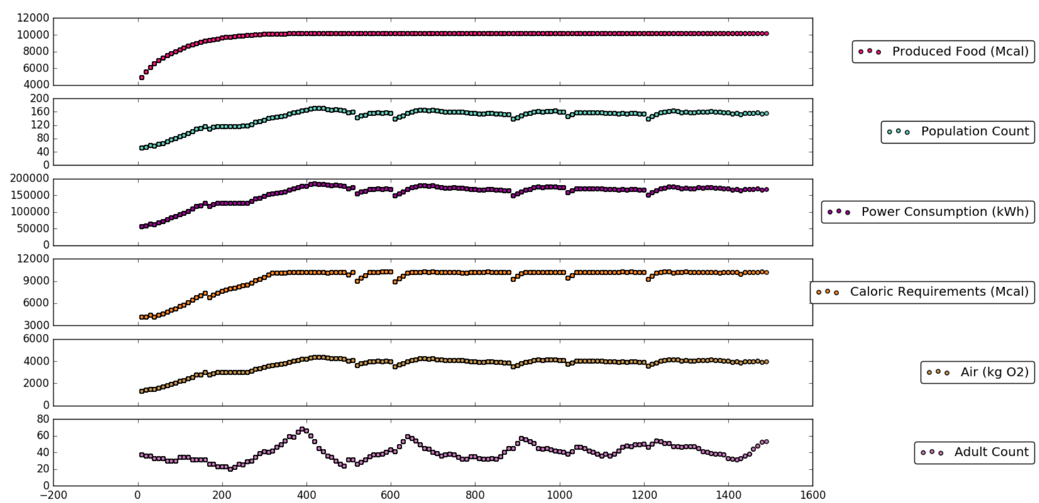

Figure 3: Results from a sample simulation run with food storage disabled ($N = 50$).

In Fig. 3, food production has an initial warm-up period, but then reaches a steady state at approximately 300 months. Over time, the population remains relatively constant, while the number of adults in the simulation fluctuates. However, the fluctuations appear to become less severe over time, suggesting a possible move to a long-term steady state. Caloric requirements, power consumption, and O_2 consumption again follow the population size closely. Due to the largely organic diet which the settlers of Mars will likely subsist on, we believe the scenario presented in Fig. 3 is likely a more realistic portrayal compared with that of Fig. 2, as long-term storage of large quantities of organic food would be fairly unrealistic.

We further explore higher values of N, the initial population size. Results from N=250, 500, and 10000 are shown in Figs. 4, 5, and 6. Interestingly, we find that as initial population size increases, the stability of the adult population generally increases (a key indicator of population stability). The population as a whole also becomes very stable, reaching a carrying capacity of the facility and remaining at that value. However, as N increases to very large values (Fig. 6), the population becomes unstable at roughly 600 months, exhibiting a jagged fall, following by a jagged increase. Furthermore, the population seems to never exceed the starting population size, instead appearing to have the starting size as an upper bound.

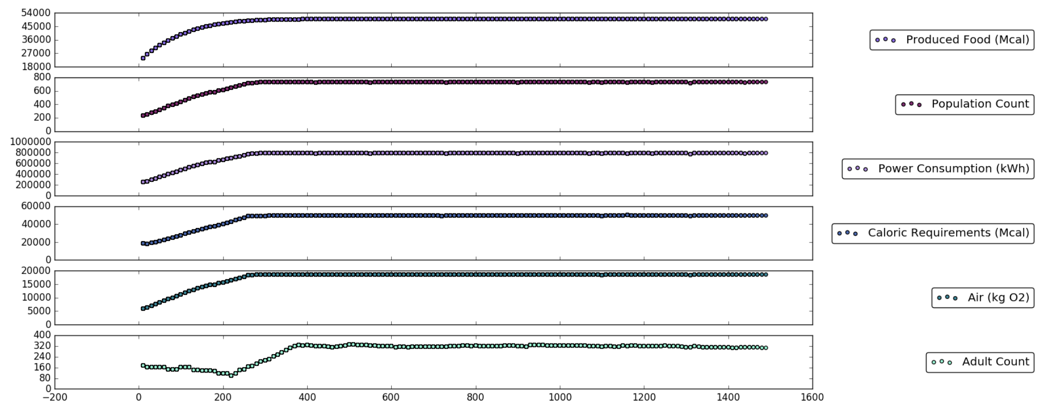

Figure 4: Results from a sample simulation run with $N = 250$.

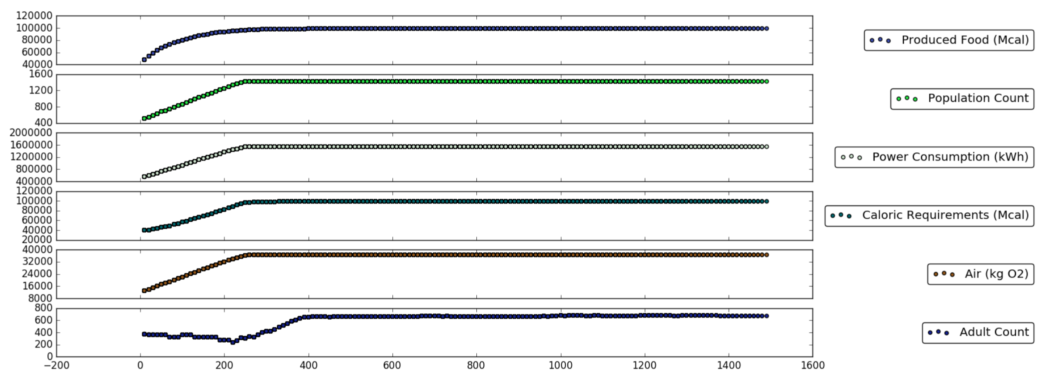

Figure 5: Results from a sample simulation run with $N = 500$.

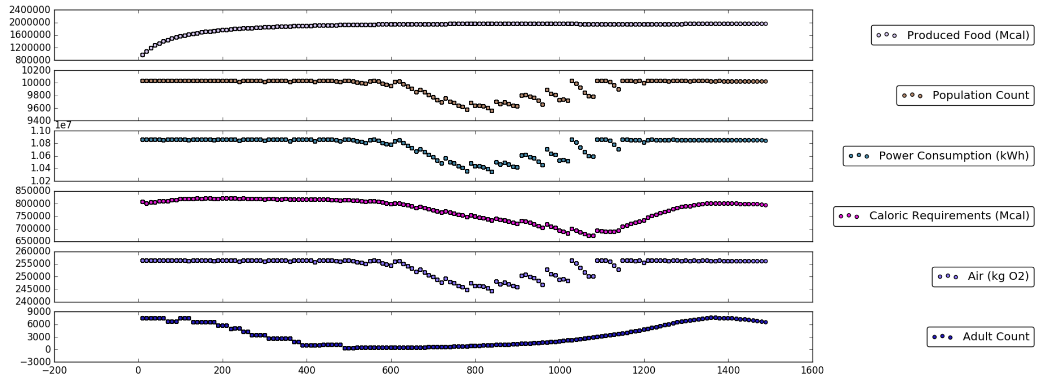

Figure 6: Results from a sample simulation run with $N = 10,000$.

6.2 Examining Population Stability

To examine the stability of varying values of N, the initial population size, we first chose several values of N which, based on our initial experimentation, would give us a variety of interesting behavior to observe. It should be noted that we define N to be the initial population size at an arbitrary time $t = 0$ at which all the necessary infrastructure is in place to support the initial population.

Next, for each N value, we performed 30 simulation runs for 1500 timesteps (125 years), with different random seeds for each run. The same sequence of random seeds was used for each batch of runs. We then examined the variation in the adult population (as this appears to be a key stability indicator) for the last 120 timesteps (10 years) using the standard deviation. From our initial experimentation, it appeared that this time period was a good candidate for examining the long-term stability, as most simulation runs seemed to either be at a steady-state position or exhibiting oscillatory behavior by this time. Our hypothesis was that for larger values of N, the stability would generally be greater (smaller variation), up to a certain point. We computed 95% confidence intervals for the stability metric for each of the values of N. Our results are presented in Fig. 7.

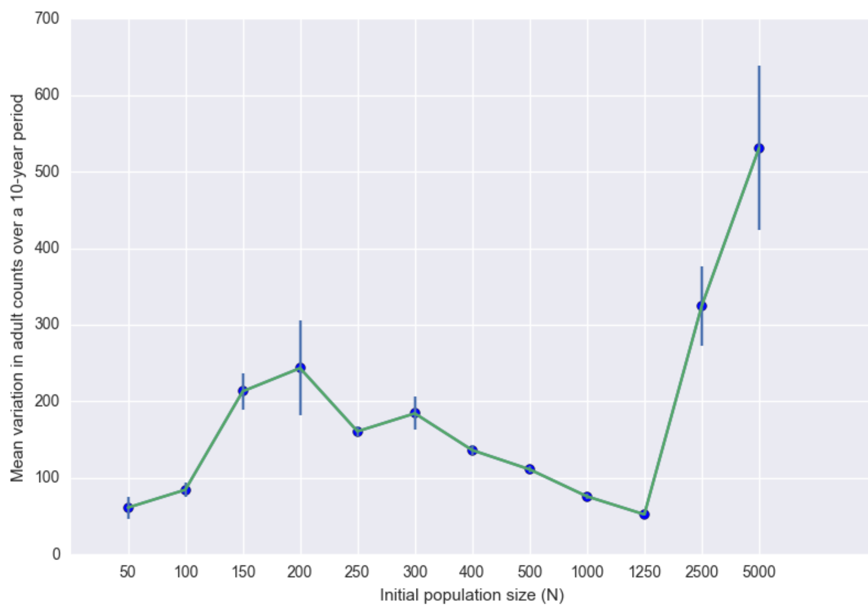

Figure 7: Variation in adult population size during a 10-year period after 115 years of habitation, as a function of initial population size.

As seen in Fig. 7, variation initially increases with N, then falls, reaching a trough at an N of 1250. Then, it again begins to climb at values of 2500 and greater. Thus, it appears that either a relatively small initial population of 50-100 individuals, or a somewhat larger population of 1000-1250 individuals will most likely achieve long-term stability. However, it should be noted that the variation is still somewhat high in a relative sense for small values of N (such as N=50). In the future, it would be helpful to compute a more advanced normalized stability metric that provides additional insight into the stability of the population. Other data sources besides the adult population could also be explored.

7 Future Work

Over the course of developing this simulation, coupled with team brainstorming meetings, there were items identified that could be incorporated into the model at a future date to either increase the fidelity of the model, or to include un-modeled aspects of the system that were omitted due to time:

- Higher-pedigree facility expansion logic, perhaps incorporating the ability for materials to be sent from Earth to Mars or the possibility of construction using available supplies.
- Incorporating a more formal food storage and spoiling model component; at the moment, if the simulation is configured to save leftover food, it never spoils and there is always enough room in the facility for its storage.
- Modeling of energy, air, and water as resources that have limits, instead of monitoring their use and assuming sufficient supply is present.
- Incorporate crop power and air requirements; this study focused on air and power requirements of the population.
- Instead of assuming fertilizer is uniformly distributed over the crop area, use it to fertilize a subset of crop area.

References

[1] Cdc: Data and statistics. http://www.cdc.gov/datastatistics/. Accessed: 2016-05-01.

[2] Country facts explorer. http://www.indexmundi.com/. Accessed: 2016-05-01.

[3] Electric power consumption (kwh per capita). http://data.worldbank.org/indicator/EG.USE.ELEC.KH.PC. Accessed: 2016-05-01.

[4] Sami Aikio and Susanna Pakkasmaa. Relatedness and competitive asymmetry–implications for growth and population dynamics. *Oikos*, 100(2):283–290, 2003.

[5] Teresa Audesirk, Gerald Audesirk, and Bruce E Byers. *Biology: life on earth*. Prentice Hall, 1996.

[6] Ester Boserup, N Makhoul, RE Munn, TN Srinivasan, JA Robinson, and C Rocha. Population and technological change: A study of long-term trends. *International Journal of Health Services*, 13(1):15–31, 1983.

[7] Hal Caswell. *Matrix population models*. Wiley Online Library, 2001.

[8] Colin Clark et al. *Population growth and land use*. London: MacMillan., 1967.

[9] William R Clark and George S Innis. Forage interactions and black-tailed jack rabbit population dynamics: a simulation model. *The Journal of Wildlife Management*, pages 1018–1035, 1982.

[10] Ricardo M. Correa, Jose Eduardo B. Pinto, Cesar Augusto Brasil Pereira Pinto, Valdemar Faquin, Erika S. Reis, Aline B. Monteiro, and Willian E. Dyer. A comparison of potato seed tuber yields in beds, pots, and hydroponic systems. *Scientia Horticulturae*, 116(1):17–20, 2008.

[11] Bret G Drake, Stephen J Hoffman, and David W Beaty. Human exploration of mars, design reference architecture 5.0. In *Aerospace Conference, 2010 IEEE*, pages 1–24. IEEE, 2010.

[12] Paul R Ehrlich, John P Holdren, et al. Impact of population growth. *Science*, 171(3977):1212–1217, 1971.

[13] Sergei Fedotov, David Moss, and Daniel Campos. Stochastic model for population migration and the growth of human settlements during the neolithic transition. *Physical Review E*, 78(2):026107, 2008.

[14] Food and Agriculture Organization of the United Nations (FAO). Roots, tubers, plantains and bananas in human nutrition. 1990.

[15] Food and Agriculture Organization of the United Nations (FAO). Food balance sheets. http://faostat.fao.org/site/368/default.aspx, 2011. Accessed: 2016-04-04.

[16] John Guckenheimer, George Oster, and Ai Ipaktchi. The dynamics of density dependent population models. *Journal of Mathematical Biology*, 4(2):101–147, 1977.

[17] Shuangsheng Guo, Weidang Ai, Yongkang Tang, Quanyong Cheng, Yunze Shen, Lifeng Qin, Jialu Ma, Jingtao Zhu, and Jin Ren. Study on o 2 generation and co 2 absorption capability of four co-cultured salad plants in an enclosed system. *Advances in Space Research*, 53(11):1551–1556, 2014.

[18] Stephen J Hoffman and David I Kaplan. Human exploration of mars: the reference mission of the nasa mars exploration study team. 1997.

[19] S Kanazawa, Y Ishikawa, K Tomita-Yokotani, H Hashimoto, Y Kitaya, M Yamashita, M Nagatomo, T Oshima, H Wada, and Space Agriculture Task Force. Space agriculture for habitation on mars with hyperthermophilic aerobic composting bacteria. *Advances in Space Research*, 41(5):696–700, 2008.

[20] Naomi Katayama, Masamichi Yamashita, Hidenori Wada, and Jun Mitsuhashi. Entomophagy as part of a space diet for habitation on mars. *The Journal of Space Technology and Science*, 21(2):2_27–2_38, 2005.

[21] Donella H Meadows, Dennis L Meadows, Jørgen Randers, et al. *Beyond the limits: global collapse or a sustainable future.* Earthscan Publications Ltd., 1992.

[22] James R. Mihelcic, Lauren M Fry, and Ryan Shaw. Global potential of phosphorus recovery from human urine and feces. *Chemosphere*, 84, 2011.

[23] David H Miller, Al L Jensen, and James H Hammill. Density dependent matrix model for gray wolf population projection. *Ecological Modelling*, 151(2):271–278, 2002.

[24] M Modell and JM Spurlock. Rationale for evaluating a closed food chain for space habitats. 1980.

[25] John H Moore. Evaluating five models of human colonization. *American Anthropologist*, 103(2):395–408, 2001.

[26] Mark Nelson, John P Allen, and William F Dempster. Biosphere 2: a prototype project for a permanent and evolving life system for mars base. *Advances in Space Research*, 12(5):211–217, 1992.

[27] Randy L Peterson, Jingjing Liang, and Tara M Barrett. Modeling population dynamics and woody biomass in alaska coastal forest. *Forest Science*, 60(2):391–401, 2014.

[28] David Quammen. The song of the dodo: island biogeography in an age of extinction. *Hutchinson, London*, 1996.

[29] John D Rummel and Tyler Volk. A modular blss simulation model. *Advances in space research*, 7(4):59–67, 1987.