Optimizing Crop Selection Using Genetic Algorithms

Graham Gordon[1]
Interstellar Lab, Hollywood, California, USA 90028

Jeremy Jauzion[2] and David Kas[3]
Interstellar Lab, Paris, France 75013

Maarten Smits[4] and Yixuan Yang[5]
Interstellar Lab, Paris, France 75013

Interstellar Lab develops bioregenerative habitats and agriculture systems, which are naturally centered around crop cultivation. These systems are very sensitive to which plants are grown, since it determines land area, water consumption, and lighting power requirements. Hence, we developed an innovative crop selection algorithm (CSA) to address these challenges. This CSA is designed for the following purposes: (1) to fully cover the nutritional requirements of the crew, (2) to ensure the dietary diversity of the daily intake, (3) to minimize the water and land usage, and (4) to schedule crop cultivation and ensure optimal use of available growing area. To tackle this, we developed a Genetic Algorithm (GA), known to perform well on this type of tasks. A population of crop selections and schedules is generated and evaluated through an objective function to select the ones that perform well. The top performers reproduce in the next generation, and mutations are applied until an optimal solution is found. Interstellar Lab believes our approach is the first step towards a bioregenerative system that puts crop selection and scheduling in the center of the BioPods that comprise our Experimental BIOregenerative Station (EBIOS).

Nomenclature

CSA	=	crop selection algorithm
GA	=	genetic algorithm
EBIOS	=	experimental bioregenerative station
CEA	=	controlled environmental agriculture
ECLSS	=	environmental control and life support system
CGR	=	compound growth rate
PPFD	=	photosynthetic photon flux density
DLH	=	daily light hours
P_g	=	gross photosynthesis
R	=	respiration
CQY	=	canopy quantum yield
CUE	=	carbon use efficiency
ET	=	evapotranspiration
LAI	=	leaf area index
R_n	=	net radiation at crop surface
IR_n	=	net irrigation requirement

[1] Lead Systems Engineer, 6516 Leland Way, Hollywood, CA 90028
[2] Data Engineer, 1 Rue Fulton, Paris, 75013, France
[3] Lead AI Engineer, 1 Rue Fulton, Paris, 75013, France
[4] Plant Production Scientist, 1 Rue Fulton, Paris, 75013, France
[5] Plant Production Scientist, 1 Rue Fulton, Paris, 75013, France

I. Introduction

INTERSELLAR Lab is developing closed-loop sustainable living systems on Earth and in space. The company focuses on creating space-grade agriculture systems and habitat modules for humans. On Earth, we provide solutions regarding the current climate crisis and the need of new methods for crop cultivation. In Space, we develop closed-loop life support systems for lunar missions in the context of Artemis NASA Program and in the future for Mars exploration. The Experimental BIOregenerative Station (EBIOS) consists of a central connector, a habitat, and several BioPods for agriculture, water treatment and waste management. We are also developing the crop cultivation BioPod as a standalone product. This Controlled Environment Agriculture (CEA) system integrates closed-loop water treatment and air revitalization systems with Greenhouse and Vertical Farm cultivation technologies.

To size and design these regenerative systems, the metabolism of plants and animals alike must be characterized. For the crew, these values have been well understood since the days of the first astronauts. Crew dietary needs drive the selection of crops in the agricultural system, while oxygen generation and water purification are secondary benefits of cohabitation with plants. The plant growth model and selection algorithms discussed in this paper provide a novel approach to designing the diet of space colonists to minimize water and land usage while maximizing the diversity of foods and nutrients for the crew.

Crop diversity is a specifically interesting trade, as it comes with advantages in psychological aspects like palatability but requires significant operational complexity in terms of multiple crop systems, disease and pest management, or compromises between optimal growing conditions for various species. Currently no constraint is placed on the number of species to be grown, because we want to conduct experiments with as broad a range of crops as possible early on. If experimental results indicate some species will not be feasible, they will be removed from the database, and a limit can easily be set on the total number of species. Additionally, the crop scheduling function allows species with similar light and temperature needs to be grouped together for better yield.

The crop growth model aims to simulate plant biochemistry with different environmental inputs to predict crop growth rate (CGR) and total yield. The CEA crop growth model contains various sub-models, including photosynthetic energy cascade, crop evapotranspiration, and irrigation models. These models are constructed in MATLAB because of its easy operation and responsiveness. Since the models regarding soil, nutrients, pests, and diseases are component specific, these sub-models will be constructed after the specific soil structure is designed and potential pests and diseases are evaluated.

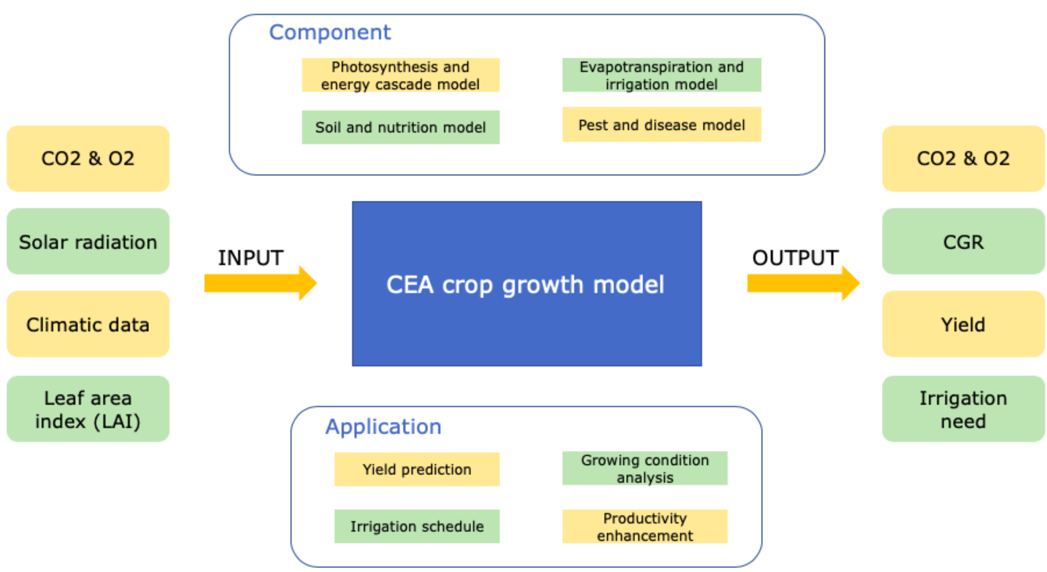

Figure 1: A schematic overview of the CEA crop growth model[0].

II. Energy Cascade Yield Prediction

The implementation of a crop growth model based on existing energy cascade models[1] serves multiple purposes. To minimise the area needed to fulfil the nutritional requirements of the crew, accurate yield estimations are necessary. To design carbon dioxide scrubbers and humidifiers, minute by minute CO_2, O_2, and H_2O flux estimations are required. To quantify the amount of biological waste that needs to be processed, precise biomass predictions are needed. The energy cascade model was originally developed for wheat. Subsequent approaches to energy cascade crop modelling have been published for only a select few crops like bean, lettuce, peanut, potato, rice, soybean, sweet potato, and tomato. Considering the sheer diversity of crops necessary to maximize dietary diversity, most crops will not have a dedicated pre-existing model approach available.

To fill in these knowledge gaps, we built a plant database which serves as a plant pool for the algorithm itself. In this database, each variety of plant is matched with micro and macro nutrient content, growth characteristics (leaf area index, days till harvest, harvest index, etc), and growth conditions (temperature, humidity, etc). Crop-specific variables were obtained from literature, but custom-built gas-exchange chambers will be used to acquire a more robust dataset. An excerpt of this plant database is shown in the table. Parameters for which a maximum value is shown, start at zero when the plant emerges from the soil and increase readily until canopy closure.

Table 1: Excerpt of plant database.

Variable	Max. LAI	Canopy cover time	Max. fAPPF	CQY	PPFD	Max. P_g	CUE	Max. R	DLH	k	Max. CGR
Unit	*N/A*	*d*	*N/A*	*N/A*	*µmol/m²/s*	*µmol/m²/s*	*N/A*	*µmol/m²/s*	*hour/day*	*kg*s/µmol/h*	*kg/m²/d*
Calculation			$1\text{-}e^{-0.78 \cdot LAI}$			*fAPPF · CQY · PPFD*		$(1\text{-}CUE) \cdot P_g$			$(DLH \cdot P_g - 24 \cdot R) \cdot k$
Hot chili pepper	3.3	71	0.93	0.078	434	31.4	0.68	10.0	20	$8.87 \cdot 10^{-5}$	0.034
Tomato	2.9	14	0.89	0.078	434	30.2	0.68	9.7	20	$8.93 \cdot 10^{-5}$	0.033
Eggplant	5.0	47	0.98	0.078	434	33.2	0.68	10.6	20	$9.14 \cdot 10^{-5}$	0.037
Cucumber	3.6	24	0.94	0.078	463	33.9	0.68	10.9	18	$8.67 \cdot 10^{-5}$	0.030

The energy cascade model consists of three distinct stages. At each stage, a fraction of the energy is lost. First light, or photosynthetic photon flux density (PPFD) is absorbed by the plant canopy. A portion of all available light is lost when it scatters off the Greenhouse dome or hits the ground instead of leaf elements. A fraction of absorbed PPFD is used as energy during photosynthesis to fix carbon into sugar. This carbon is derived from atmospheric CO_2. Most of the sucrose is converted into biomass, but a portion is burned for energy during respiration which releases CO_2 back into the atmosphere. Respiration is assumed to occur at a constant rate during both light and dark periods. Hourly Crop Growth Rate (CGR) during the total growing time is used to calculate the accumulated biomass at the time of harvest. Shown below are example growth curves for various species, where biomass increases between harvest cycles.

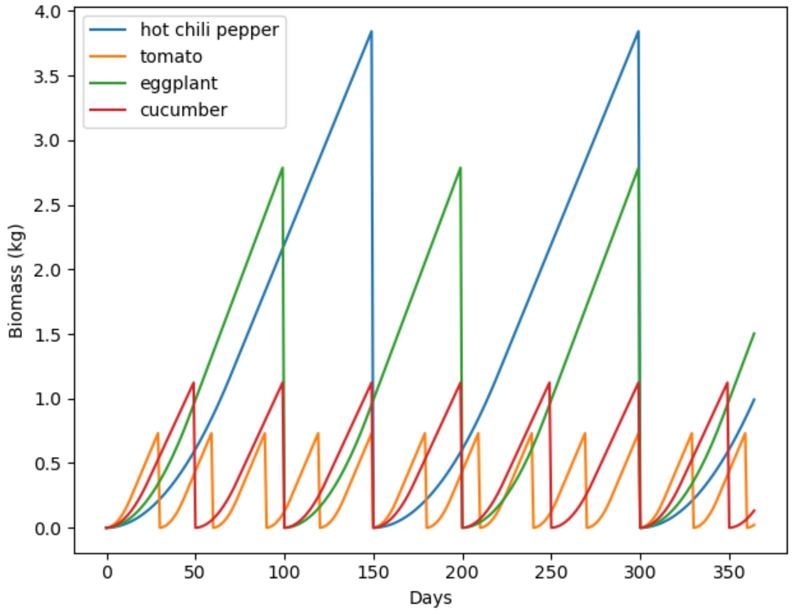

Figure 2: Model predictions of the accumulated dry biomass over time.

Crop Growth Rate is calculated based on the number of daylight hours (DLH), gross Photosynthesis (P_g), Respiration (R), and a crop-specific constant for unit conversion (k).

$$CGR = (DLH \cdot P_g - 24 \cdot R) \cdot k \qquad (1)$$

Gross Photosynthesis is calculated based on the fraction of absorbed light (fAPPF), Canopy Quantum Yield (CQY, number of moles of carbon carboxylated for each mole of absorbed light), and the incoming light per m^2 ground area (PPFD). It is important to keep in mind that CQY is strongly dependent on atmospheric conditions.

$$P_g = fAPPF \cdot CQY \cdot PPFD \qquad (2)$$

Respiration is calculated based on Carbon Use Efficiency (CUE, number of moles of carbon fixed into biomass for each mole that is carboxylated). Since CQY and CUE are generally similar among crops that use the same C3 carbon fixation cycle for photosynthesis, common values of CQY = 0.078 and CUE = 0.68 were assumed for species where specific data was unavailable.

$$R = (1 - CUE) \cdot PPFD \qquad (3)$$

The cumulative CGR of each crop until harvest determines the total dry biomass. A fraction of the total biomass is edible. There is a sharp distinction in the approach for predicting annual and biennial crop yields, versus perennial crop yields. For annual and biennial crop yields, a crop-specific harvest index is used to determine the percentage of total biomass that is edible. After harvest, the entire plant must regrow in roughly the same ratio. For perennial crops, the plant survives beyond the harvest point, which makes harvest indices a less valuable tool. Instead, it is more productive to calculate how resources are allocated to fruit, leaves, wood, and root components during the plant's life cycle. This is called dry-matter partitioning. These fruit partitioning calculations are based on Hester's apple orchard model[3].

fF, the fraction of dry matter partitioned to fruits, is calculated based on organ specific constants α, β, and γ, and Fruit Load (FL, the number of fruits per m^2 of leaf area).

$$fF = \alpha + (\beta \cdot FL)/(\gamma + FL) \qquad (4)$$

The fruit partitioning calculation predicts how much energy the tree expends to create botanical fruits, but in some crops, such as cashews, only part of the botanical fruit is consumed. In these cases, a distinction must be made between total fruit portioning, and edible fruit partitioning.

In the Vertical Farm, the properties and lighting schedules of the LED's determine the PPFD. PPFD values for plants grown in the Greenhouse are calculated from daily light integral (DLI) data throughout the year for any location on Earth from SunTracker Technologies' DLI Calculator*. PPFD per day can then be calculated via the daily light hours obtained from Time and Date AS**. The fraction of light that enters the Greenhouse is calculated based on material properties of the transparent CEA cover.

The fraction of light that hits a leaf element is based on a crop-specific Leaf Area Index (LAI). This results in a fraction of Absorbed Photosynthetic Photon Flux (fAPPF) as shown in the graph below. While LAI differs by up to 2x as shown earlier in Table 1, the logarithmic formula for fAPPF results in a difference of only 10%. Perennial plants have already reached their maximum LAI, so absorbed PPFD is solely dependent on incoming light. For annual and biennial plants, LAI starts at zero when the plant emerges from the soil, and so light absorption increases linearly until it approaches its maximum absorption at canopy closure. This monthly data is interpolated to obtain minute by minute data.

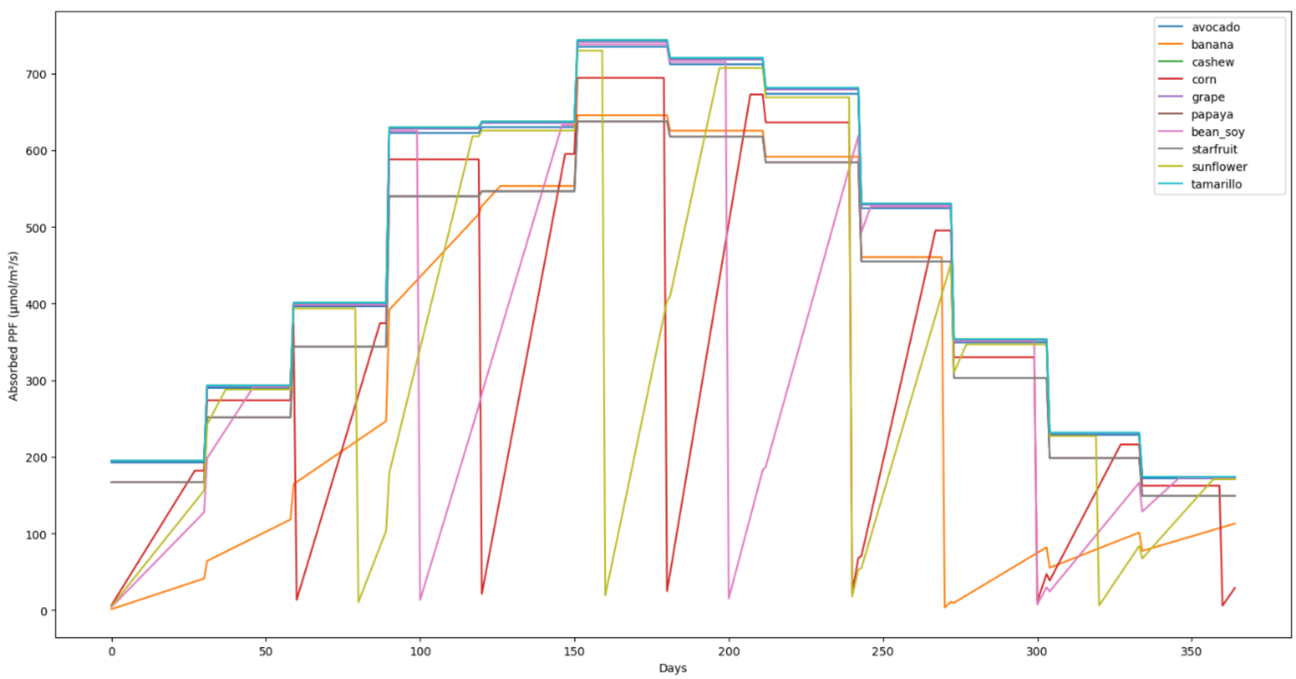

Figure 3: Model predictions for the Absorbed Photosynthetic Photon Flux m^{-2} s^{-1} over a year in Paris.

* https://dli.suntrackertech.com

** https://www.timeanddate.com/sun/france/paris

III. Evapotranspiration and Irrigation

Besides the energy cascade model, another key component in the CEA crop growth model is the evapotranspiration model and irrigation models. These two models are implemented to give a reference to the precise irrigation management in EBIOS. This section will briefly introduce some fundamental details of these models.

Water is one of the most important resources for both humans and crops. For an environmental control and life support system (ECLSS) like that of EBIOS, water circulation is one of the essential elements. In the crop cultivation system, an accurate estimation of water consumption and water loss is significant in determining the crop growth as well as for informing appropriate irrigation management. The irrigation system consequently influences crop nutrient availability and soil salinity, which are two factors that impact crop growth[5]. An appropriate model is needed to accurately estimate the water demand for the agriculture system. Models that calculate plant evapotranspiration (ET) are generally used to predict the water loss and water consumption in both field and indoor crop production systems. Plant evapotranspiration consists of an evaporation and a transpiration process. Evaporation refers to a water loss process that occurs from the soil and wet vegetation surface, while transpiration is a process in which water escapes plant tissue to the atmosphere via stomata. Meteorological factors such as solar radiation, wind speed, air humidity and air temperature have domination over these two processes[8]. Thus, the total water loss from a plant can be found as the combination of these two processes, so called evapotranspiration (ET).

Understanding ET is crucial to determine crop water requirements, and it can subsequently give a reference for the irrigation management strategy. To achieve this, a method to calculate ET inside different domes of the EBIOS CEA system must be defined. Many models have been developed by researchers in the past decades. The Penman-Monteith model is the base of all the newly conducted models for greenhouse ET[9]. Fazlil-Ilahil carried out a literature-based study in his 2009 MSc thesis which summarized existing ET models for greenhouse climate, so-called microclimate[10]. In addition, this thesis illustrated suitable models for different types of greenhouse: Low technology, medium technology, and high technology greenhouses. For the crop cultivation system in EBIOS, the recommended models for high technology greenhouses will be used as our reference. The Stanghellini model is described as one of the most accurate models (r2 = 0.96, RMSE = 0.006). Moreover, after considering the diversity of crop species in EBIOS, the Stanghellini model was chosen to be the most suitable model to calculate the reference ET (ET_0).

The Stanghellini model was first developed in 1987[10]. A well-developed tomato plant grown in a single glass, Venlo-type greenhouse with inner heating system was used as a model plant to estimate the ET_0. This model also applied leaf area index (LAI) to determine the energy exchange from different layers of leaves or canopy of the crops.

$$ET_0 = 2LAI\,\frac{1}{\lambda}\left[\frac{s(R_n-G)+K_t[VPD\rho C_p]}{s+\gamma\left[1+{}^{r_c}/_{r_a}\right]}\right] \tag{5}$$

$$R_n = \frac{0.07R_{ns}-252\rho C_p(T-T_o)}{r_R} \tag{6}$$

$$R_{ns} = 0.77R_s \tag{7}$$

$$r_R = \frac{\rho C_p}{4\sigma(T=273.15)^3} \tag{8}$$

where
ET_0: Reference evapotranspiration (mm/day)
R_n: Net radiation at the crop surface (MJ/m2/day)
G: Soil heat flux density (MJ/m2/day)
K_t: Unit conversion factor equal to 3600 (s/h)
VPD: Daily or hourly vapour pressure deficit (kPa)
ρ: Mean atmosphere density (kg/m3)
C_p: Specific heat of the air (MJ/kg/°C)
r_R: Radiative resistance (s/m)
r_c: Canopy resistance (s/m)
r_a: Aerodynamic resistance (s/m)

λ: Latent heat of vaporization (MJ/kg)
s: Slope of the saturation vapour pressure curve (kPa/℃)
γ: Psychrometric constant (kPa/℃)
R_{ns}: Net short-wave radiation (MJ/m2/day)
R_s: Ground level solar radiation (MJ/m2/day)
T: Hourly or daily mean air temperature (℃)
T_o: Leaf temperature (℃)
σ: Stefan-Boltzman constant (MJ/m2/K4/day)
LAI: Leaf area index (m2/m2)

After calculating the ET_0, the species-specific evapotranspiration (ET_c) is calculated. The ET_c refers to the crop evapotranspiration under standard condition. This condition is described as a growing condition with no disease or pest, well-fertilized, water-sufficient, and fully produced under given climatic conditions. Under such condition, the ET_c can be described as:

$$ET_c = ET_0 \times K_c \qquad (9)$$

where K_c: Crop coefficient

As a result, the daily evapotranspiration values of different crops are calculated. The result of the evapotranspiration is subsequently followed by the irrigation model implementation.

The calculation of irrigation is performed after defining the evapotranspiration model. The irrigation requirement (IR) refers to the amount of water that must be provided via irrigation to maintain optimal growth. Since there are no natural rainfall in our system, all the crop water requirement (CWR) is dependent on artificial irrigation, in which case, the IR should be at least equal to CWR[8]. An optimal water supply towards crops will prevent crops suffering from water stress, consequently, ensure the optimal yield. To achieve this, a model will be needed to illustrate different variables influencing IR and CWR.

The net irrigation requirement (IR_n) is a parameter that does not include the water loss from the process of applying water, while gross irrigation requirement (IR_g) is a combination of IR_n and the loss[8]. Net irrigation requirement (IR_n) is equal to:

$$IR_n = ET_c - (Pe + Ge + Wb) + LR_{mm} \qquad (10)$$

where
ET_c: Crop evapotranspiration (mm)
Pe: Effective dependable rainfall (mm)
Ge: Groundwater contribution from water table (mm)
Wb: Water previously stored in the soil (mm)
LR_{mm}: Leaching requirement (mm)

Hence, a comprehensive water management model is constructed by combining the evapotranspiration and irrigation model. The output of the model will be used later for the crop selection algorithm thereafter.

IV. Nutritional Requirements

The crop selection algorithm's main goal is to fulfill daily nutritional requirements for every crew member. Based on different levels of activity, and on different metabolisms, an average requirement was found so that no one would lack nor have an excess of any nutrient.

One important parameter that determined the boundaries for macronutrients is the number of calories needed. As explained on the FAO website[11], the Total Energy Expenditure (TEE, i.e., the amount of energy needed) is the product of Basic Metabolic Rate (BMR) and Physical Activity Level (PAL). NASA's HIDH tables[12] 50th percentile metabolic rate was used in our activity estimate. Given a recommended range[13] of what proportion of energy should come from the different macronutrients (fat, carbohydrate, protein), a minimum and maximum intake are computed for each of them. According to several sources[14-22], similar boundaries are computed for vitamins (A, B1, B2, C, E, K …) and some minerals (Ca, Mg, Na …). For example, the recommended value for protein intake is 0.8 gram per kilogram of bodyweight[23] but it had to be increased in accordance with the physical activity level. For several nutrients, information was lacking on what would be the effects of consuming too much or not enough of them. In these cases, either the constraint was removed, or a limit was set at a level where no adverse effect was observed.

More information is needed for the algorithm to determine if the food consumed will provide the whole crew with a steady intake for all the nutrients. If a crew member needs a certain income of energy every day, some nutrients can be absorbed over longer periods of time. So instead of checking if everyone gets more than the lower limit for each nutrient every day, a convolution is used over a certain time window specific to each nutrient. On the other hand, upper limits for nutrient intake are always checked over the smallest period to prevent from any overdose.

A particular attention was also paid to the variety of the food. For example, it is important to forbid the algorithm from selecting a menu of 10 kilograms per day per person. As an analogy, the mass of food consumed by a crewman during the 2nd stage of BIOS 3 experiment was almost 4 kg per day[24]. A daily limit on the intake of each crop was also set to avoid relying too much on one crop and to avoid having to eat 4 kg of the same food daily.

Table 2: Crew nutritional requirements derived from literature.

Daily nutritional requirement	Minimum intake	Maximum intake
Energy (kcal)	2800	-
Protein (g)	95	-
Carbohydrate (g)	350	525
Fat (g)	47	109
Thiamin (mg)	1.15	-
Riboflavin (mg)	1.15	-
Niacin (mg)	15	35
Vitamin B6 (mg)	1.3	100
Folate (µg)	800	2000
Choline (mg)	400	-
Vitamin C (mg)	45	-
Vitamin E (mg)	15	-
Vitamin K (µg)	60	-
Calcium (mg)	700	2500
Phosphorus (mg)	550	-
Magnesium (mg)	240	-
Iron (mg)	22	-
Zinc (mg)	12	50
Selenium (µg)	30	400
Potassium (mg)	2500	6000
Sodium (mg)	1300	2300
Vitamin A (µg)	550	4945
Omega-3 (mg)	250	6000
Omega-6 (mg)	2500	30000

V. Genetic Algorithm

To address all the challenges in fulfilling the dietary needs of a crew of 4 astronauts, we developed a crop scheduling algorithm that uses a genetic approach for selecting which crops to plant, when to plant them, and how to arrange the different species. The problem of scheduling is a combinatorial optimization problem that is highly complex. The challenge is an unconventional version of factory job scheduling, where the minimization process is not based on duration of work but rather underlying dietary outcomes resulting from the processes of harvesting, storing, and eating food to sustain the astronauts. Furthermore, the choice of the crops to seed at different times also depends on non-dietary constraints, for example minimizing daily water use, maximizing food diversity, minimizing waste of food, and minimizing the required grow.

Considering the unique problem and variety of constraints, we opted for a hand-made genetic optimization approach where we mutate and evolve a population of crop schedules with a fitness score selection criteria. Genetic algorithms often lack optimal performance relative to other forms of artificial intelligence but are more often used because of their simplicity and modularity. By comparison, neural networks are more commonly used in reinforcement learning, supervised or unsupervised, when the inputs are more variable and broader in range. For example, image recognition or sensor data for control systems are better suited for a neural system rather than a genetic algorithm which needs to reproduce for every data point. Since the crop data and nutritional requirements are relatively static, however, the genetic algorithm is well-suited to progressively optimize a solution. In this section we describe the main parts of the genetic algorithm, namely the individual schedule, the fitness function score, the mutation and reproduction strategy, and the pseudo-code of the genetic algorithm.

A. Individual Schedule and Initial Population

An individual in the population of the GA is described as a 2-dimensional array. One dimension (columns) represents the number of weeks we are scheduling on, for example 52 weeks rotation or 1 year. The other dimension (rows) represents growing spaces. A growing space is a subdivision of the growing area, which can be Greenhouse area, Vertical Farm area, herb garden area or mushroom area. The size of the growing space and the growing area are hyperparameters in the algorithm that was found by grid search for subdivisions of the total growing area in our crop cultivation systems. The values inside the array represent different crops, each of which occupies a space for as many weeks as the calculated harvest time from the energy cascade model. There are different subsets of growing spaces for the different growing areas.

There are 2 subtleties in the description of an individual. The first one is that we want to have a fixed scheduling period in the agriculture system, so that all crops are harvested at the end of the mission. Because of the different harvest cycle durations, we space the time between harvesting and seeding differently to have each crop complete harvesting at the end of the scheduling period. The other trick is about allowing the algorithm to schedule nothing on a growing space, for water irrigation or dietary purposes. Therefore, we have introduced an empty crop in the crop index possibilities.

At the start of the GA, we initialized a population of individual schedules. For each schedule, we randomly select crops and assign them to growing spaces until no more crops in the database can fit on the scheduling period. If the sum of harvest cycles is smaller than the scheduling period, we space them equally to fulfill the whole growing period, as described above. We then apply a random time shift to allow better alignment of the schedule in the individual search space. It means that the schedule starts with crops already in growing process in most of the subspaces.

B. Fitness Function and Selection

The score of an individual in the population is a weighted sum of different fitness scores. For scoring an individual, we create different representations of the initial individual schedule: a food schedule, a diet schedule, an irrigation schedule, and an excess schedule. Each schedule will allow us to calculate a subcomponent of the fitness score. The final fitness score is a weighted sum of these sub-components.

A food schedule is a new 2-dimensional array representing crops in one dimension (rows) and weeks in the other dimension (columns). The week we harvest a specific crop in the sub-space of a specific area, we divide the yield calculated by the energy cascade model by the vacuum storage time specific to this crop in our database and add it to the corresponding weekly range in the line of the crop. A value in the food schedule thus represents the total amount of yield available to the crew on a specific week.

This food schedule is used to threshold the maximum weekly intake of a certain crop. If the amount of food grown exceeds the maximum intake for a certain crop, we consider that the excess is waste, which will go in the excess schedule and will not add to the dietary score. The values in the excess schedule represent the total amount of food greater than the maximum authorized intake during a certain week. We take the maximum of these values as a component of our fitness score to minimize waste and enforce diversity in crops.

Out of the food schedule, we create the diet schedule which instead of having the crops in one dimension, has all the macronutrients and micronutrients. The values in the diet schedule represent the total amount of a specific nutrient that is available at a specific week. We use this schedule to calculate the dietary subcomponent of the fitness score. Each week, we calculate for each nutrient if it is in the recommended range of values. We attribute quantitatively a value of 1 if the quantity is respected and a value decreasing to 0 as it moves further outside of the range. The subcomponent of the fitness score is then the minimum of the whole diet score array, to ensure that dietary constraint is respected on a weekly basis.

By using the evapotranspiration model, we also create an irrigation schedule, which is a one-dimensional array with size equal to the number of weeks we schedule for. From the initial individual schedule, we calculate for each week the irrigation requirement of each crop. We use the maximum of this array as a subcomponent of the fitness score, then invert the value since we want to minimize water demand. We ensure that we have stability or variance reduction in the irrigation schedule or system.

C. Selection, Reproduction and Mutation

Once we calculated a weighted average of scores for everyone in a population of a certain generation, we normalized all the scores to sum up to 1 with a SoftMax function. We then select based on this multinomial distribution a pool of high-performing individuals that will reproduce, delete the rest of the population, then perform reproduction to restore the original population size.

Each new child is created by randomly selecting a schedule for each sub-space from one of its parents. This reproduction strategy is like selecting genes alleles in biology, alleles corresponding to a crop-schedule on the sub-space of a sub-area.

Once we created a child, we apply with a small probability 2 types of mutation:

- Crop schedule (allele) reset mutation: We delete the crop-schedule on the sub-space and create a new random one, the same way as we created the initial population. We select and place randomly a certain number of crops with their growth harvest time on the sub-space. In short, we replace the line in the 2-dimensional array representation of the child schedule.
- Crop schedule (allele) rotation mutation: We shift or pivot the schedule by a random number of weeks in the same way as we created an individual in the initial population. Values after the pivot are copied at the beginning of the schedule and inversely values before the pivot are copied at the end.

D. Pseudo-Algorithm

Hyper-parameters: population size, reproduction pool size, number of generations, size of growing areas, number of sub-areas (growing spaces), fitness score subcomponent weights, reset mutation rate, rotation mutation rate, crop data-frame with cultivation areas, growth harvest times, vacuum conservation time.

Step 0: Initialize population with individuals:

- For each growing space (rows), randomly place crops of size equal to their growth harvest time, side to side, until the sum of the harvest times exceeds the schedule plan. Add spaces between each crop to evenly fill the schedule.

- Randomly apply a rotation or time shift to the schedule.

For any number of generations:
- Step 1: Score everyone in the population with a fitness score.
    - Create fridge schedule, diet schedule, irrigation schedule and excess schedule.
    - Calculate dietary, irrigation and excess sub-components of the fitness scores from schedules.
    - Fitness schedule is the weighted sum of the sub-components.
    - If any individual is the best schedule over all generations, save it.
- Step 2: Select randomly a reproduction pool by sampling the normalized multinomial distribution from fitness scores.
- Step 3: Create children equal to population size minus reproduction pool size plus one:
    - Randomly select one growing space schedule from parent A or parent B (line)
    - With small probability, apply a random reset mutation.
    - With small probability, apply a random rotation mutation.
- Step 4: Re-inject the best schedule in new population.

VI. Results

Overall, the genetic approach to crop selection and scheduling provides a basis for constantly improving solutions to providing food for astronauts. When applied to the full nutritional demand for a crew size of 4, an area of 278 m^2 is needed for all crops, along with an irrigation requirement of 33 kL per year. It is worth noting that these values are smaller than conventional agriculture[25] by 1-2 orders of magnitude (10-100x), since the CEA architecture allows for faster harvest cycles, higher base productivity, and recapture of evaporated water from the air. For example, the Biosphere 2 experiment needed roughly 6x more growing area[26] since it relied on the sun as the only energy input for soil agriculture. Furthermore, the CO2 is elevated to 1500ppm which improves quantum yield, and lighting profiles are set to maximize yield of each species. The closest analytical model is from Zabel's 2019 dissertation[27], which uses a similar hybrid life support system, with physiochemical and bioregenerative functions, but is not specifically optimized to reduce the growing area required. The only data showing a more efficient use of growing area was China's Lunar Palace 1, which is likely because it provided only 55% of the diet rather than 100%, allowing them to focus on the species with highest caloric yield[28].

Table 3: Comparison of growing areas required to support crew in various studies and experiments.

Source	Growing Area (m^2)	Crew Size	Calorie Demand (kcal/d)	Crop System Efficiency
Lunar Palace 1[28] (Experiment)	69	3	2900[29] * 55%	69.3 kcal/m^2/d
Interstellar Lab Crop Selector	278	4	2800	40.3 kcal/m^2/d
Zabel 2019[27] (Analysis)	385.5	6	1900	29.6 kcal/m^2/d
Biosphere 2[26] (Experiment)	2000	8	2200 * 80%	7.0 kcal/m^2/d
Elementa Vegan[25] (USDA)	1300	1	2150	1.65 kcal/m^2/d

Among 103 crops selected, the 10 most heavily utilized have at least 120 kg harvested per year. These top contributors include potato, squash, sweet potato, cucumber, fava bean, pumpkin, soybean, tomato, and watermelon. This is indicative of the large diversity between fruits, vegetables, legumes, and tuber crops. As discussed in the introduction, this quantity of crop species may be difficult to manage for small crews, but as the group size scales up to dozens or hundreds, the high level of diversity is advantageous for maintaining physical and psychological health. As a work schedule is developed for crew operations, the topic of diversity will need to be revisited to determine an appropriate number of species for various crew sizes.

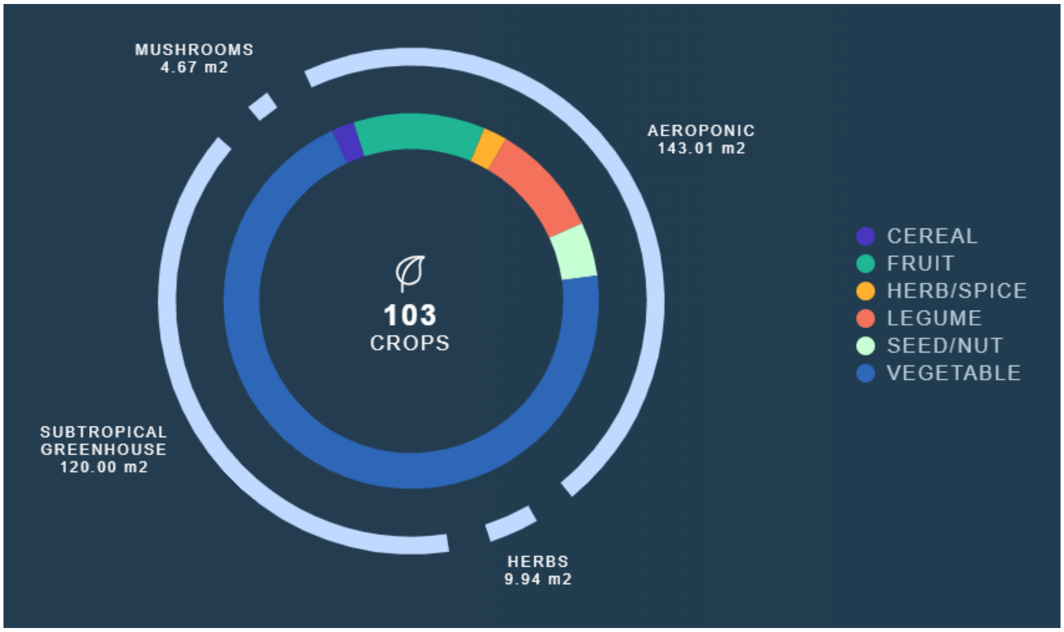

Figure 4: Example crop area allocations for a crew size of 4.

As mentioned previously, this algorithm will be used as a tool to design systems for future astronauts, from sizing greenhouses and vertical farms to predicting the exchange of biomass, water, oxygen, and carbon dioxide in the life support systems. This crops selection algorithm has been deployed to a server, where it can be run for a variety of crew sizes, specific nutritional requirements, and food allergies. To access a demonstration version of this software, please visit the link below:

https://crop-selector.interstellarlab.earth/

References

Introduction

[0]Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., ... & Ritchie, J. T. (2003). The DSSAT cropping system model. *European journal of agronomy*, 18(3-4), 235-265

Energy Cascade

[1]Jones, H., & Cavazzoni, J. (2000). Top-level crop models for advanced life support analysis (No. 2000-01-2261). *SAE Technical Paper*.

[2]Boote, K. (2020). Advances in crop modelling for a sustainable agriculture. Burleigh Dodds Science Publishing Limited.

[3]Hester, S. M., & Cacho, O. (2003). Modelling apple orchard systems. *Agricultural systems*, 77(2), 137-154.

[4]Volk, T., Bugbee, B., & Wheeler, R. M. (1995). An approach to crop modeling with the energy cascade. *Life Support & Biosphere Science*, 1(3-4), 119-127.

Evapotranspiration and Irrigation

[5]Blanco, F.F. & Folegatti, M.V. 2004. Evaluation of Evaporation-measuring Equipments for Estimating Evapotranspiration within a Greenhouse. *Revista Brasileira de Engenharia Agricola e Ambiental* v.8(n.2/3):p184-188.

[6]Jones, H., & Cavazzoni, J. (2000). Top-level crop models for advanced life support analysis (No. 2000-01-2261). *SAE Technical Paper*.

[7]Pamungkas, A., Hatou, K., & Morimoto, T. (2014). Evapotranspiration Model Analysis of Crop Water Use in Plant Factory System. *Environmental Control In Biology*, *52*(3), 183-188. doi: 10.2525/ecb.52.183

[8]Savva, A. P., & Frenken, K. (2002). *Crop water requirements and irrigation scheduling* (p. 132). Harare: FAO Sub-Regional Office for East and Southern Africa.

[9]Takakura T., Kubota C., Sase S., Hayashi M., Ishii M., Takayama K., Nishina H., Kurata K., Giacomelli G. A. 2009. Measurement of evapotranspiration rate in a single-span greenhouse using the energy-balance equation. *Biosystems Engineering* 102(3):298-304

[10]Fazlil-Ilahil, W. F. (2009). Evapotranspiration models in greenhouse. *Wageningen University*.

Nutritional Requirements

[11]FAO, Energy Requirements of Adults http://www.fao.org/3/y5686e/y5686e07.htm#bm07

[12]NASA, Human Integration Design Handbook (HIDH) 2014

[13]NCBI Bookshelf, Dietary Reference Intakes (DRIs): Acceptable Macronutrient Distibution Ranges, Food and Nutrition Board, National Academies https://www.ncbi.nlm.nih.gov/books/NBK545442/table/appJ_tab5/?report=objectonly

[14]FAO, Thiamin, Riboflavin, Niacin, Vitamin B6, Pantothenic Acid, Biotin http://www.fao.org/3/Y2809E/y2809e09.htm#bm9

[15]European Food Safety Authority (2006). Tolerable upper intake levels for vitamins and minerals

[16]Carver, J. (2006). Conditionally essential nutrients: Choline, inositol, taurine, arginine, glutamine and nucleotides. In P. Thureen (Author) & W. Hay (Ed.), Neonatal Nutrition and Metabolism (pp. 299-311). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511544712.020

[17]K.U. Ingold, A.C. Webb, Donna Witter, G.W. Burton, T.A. Metcalfe, D.P.R. Muller (1987). Vitamin E remains the major lipid-soluble, chain-breaking antioxidant in human plasma even in individuals suffering severe vitamin E deficiency, Archives of Biochemistry and Biophysics, Volume 259, Issue 1, 1987, Pages 224-225, ISSN 0003-9861

[18]EFSA NDA Panel (EFSA Panel on Dietetic Products, Nutrition and Allergies), 2016. Scientific opinion on Dietary Reference Values for choline. *EFSA Journal* 2016; 14(8):4484, 70 pp. doi:10.2903/j.efsa.2016.4484

[19]Agarwal A, Shaharyar A, Kumar A, Bhat MS, Mishra M. Scurvy in pediatric age group - A disease often forgotten? *J Clin Orthop Trauma*. 2015;6(2):101-107. doi:10.1016/j.jcot.2014.12.003

[20]Eiji Takeda, Hironori Yamamoto, Hisami Yamanaka-Okumura, Yutaka Taketani, Dietary phosphorus in bone health and quality of life, Nutrition Reviews, Volume 70, Issue 6, 1 June 2012, Pages 311–321, https://doi.org/10.1111/j.1753-4887.2012.00473.x

[21]WHO. Guideline: Potassium intake for adults and children. Geneva, World Health Organization (WHO), 2012.

[22]Niels Graudal, Gesche Jürgens (2018). Conflicting Evidence on Health Effects Associated with Salt Reduction Calls for a Redesign of the Salt Dietary Guidelines, Progress in Cardiovascular Diseases, Volume 61, Issue 1, 2018, Pages 20-26, ISSN 0033-0620

[23]S. Bilsborough, N. Mann (2006). A Review of Issues of Dietary Protein Intake in Humans, International Journal of Sport Nutrition and Exercise Metabolism, 2006, 16, 129-152

[24]V.S. Kovalev, N.S. Manukovsky, A.A. Tikhomirov (2019). Computing-feasibility study of NASA nutrition requirements as applied to a bioregenerative life support system

Results

[25]Peters, Picardy et al (2016). Carrying capacity of U.S. agricultural land: Ten diet scenarios. *Elementa: Science of the Anthropocene, 2016, 4: 000116*

[26]Silverstone and Nelson (1993). Food production and Nutrition in Biosphere 2: Results from the First Mission. *Advances in Space Research Vol 18*

[27]Zabel (2019). An investigation of the dynamic behavior of a hybrid life support system and an experiment on plant cultivation with a urine-derived nutrient solution. *Technical University Dresden*

[28]Wheeler (2017). Agriculture for Space: People and Places Paving the Way. *Open Agriculture, 2017, 2:14-32*

[29]Fu, Guo, Liu (2018). An Optimized 4-day Diet Meal Plan for 'Lunar Palace 1'. *Journal of the Science of Food and Agriculture*