Closer to Cayuga's Waters: An Evaluation System of the Invasive *Hydrilla* Species

By Zichen Wang, Mo Lyu, and Franklin Deng
Department of Mathematics and
Department of Computer Science, College of Arts and Sciences

Abstract

In recent years, an invasive plant species, *Hydrilla verticillate*, has been identified to pose significant threats to the local ecology of Cayuga Lake. This paper aims to provide a practical, quantitative, and real time evaluation system that helps monitor and predict the population dynamics of the plant. The system includes two main parts: sampling and simulation. In Section 3, we discuss in detail our stochastic Monte Carlo simulation based on the enhanced *Invasive Plant Population Dynamic* (IPPD) model. Human activities, such as boating, are identified to greatly accelerate the spread of *Hydrilla*. In Section 4, we employ statistical methods to promote confidence in the sample results, which in turn guarantees more accurate simulation results. We provide an adaptive algorithm to calculate the minimum sample size in real time. Other than the practical utility of our study, there is also the theoretical importance in that various approaches discussed in this paper, such as the IPPD model and the adaptive algorithm, can serve as the basis for future work on this subject.

1. Introduction

To the north of Ithaca, Cayuga Lake stands as a landmark of the state and contributes to the mild local climate and supreme natural beauty. However, in recent years, an invasive plant species, Hydrilla verticillate, has been identified to pose significant threats to the local ecology of Cayuga Lake. Hydrilla is an aggressive submersed perennial plant. As an invasive species, it out-competes native plants and creates a monoculture that disrupts the balance of the local ecosystem. Hydrilla's potato-like "tubers" that grow in the bed of the lake make herbicides largely ineffective in the long term (New York Invasive Species Information, 2019). Despite various efforts to eradicate the species, new locations of Hydrilla have been continuously spotted, and early detection of these new locations remains at present the optimal way to control their spread. Given limited time and human resources, there is thus an urgent need of a scientific evaluation system that monitors new Hydrilla spots in real-time and predicts their future developments.

We come to our evaluation system by essentially answering the following questions:

- What factors contribute to the growth and spread of the plant?
- What would the natural spread pattern of the plant be like?
- How can we accurately and timely monitor the current distribution of the plant?
- How do human activities, such as boating and chemical treatment, influence the plant?

2. Assumptions and Justifications

To answer the first question from above, we need a closer examination of the habits of Hydrilla. This will also lead to some important assumptions and simplifications of our model used to predict the future developments of the plant. Hydrilla is a perennial species with a normal life span of 7 to 12 years. Considering that the Cayuga Lake Hydrilla Management Plan is made for a 5-year interval (Cayuga Lake Hydrilla Task Force, 2021), we set our model's simulation time length to 5 years and assume that all Hydrilla will not die within this time period. Hydrilla propagates primarily by stem fragments, although turions and subterranean tubers also play an important role. The main means of the introduction of Hydrilla is as castaway fragments on recreational boats and trailers and in their live wells. Once the stem pieces get carried away to a new place, they grew to root their tubers in the substrate to establish new colonies. It is thus reasonable to use the number of tubers to represent the overall density of Hydrilla. As Hydrilla grows rapidly and caps at 6000 tubers per square meters (Missouli Stream Team, 2020), we further simplify our calculation by substituting the conventional logistic population growth model with exponential growth model that caps at 6000 tubers per square meter. This also coincides well with Audl's observation that invasive plant species have an exponential growth rate for a relatively long period (see Section 3). Figure 1 shows that when the growth rate is high, exponential growth with cap is a good approximation of the logistic growth.

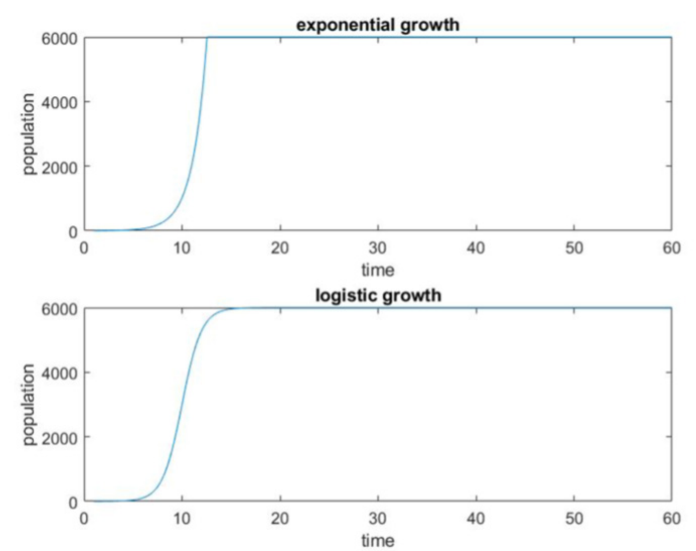

Figure 1: Comparison Between Two Population Growth Models

We also assume that the probability of spread fluctuates around a constant. That is, the probability of 35 spread at a certain area is independent of the population density at or around that area (Blackburn and Tueller, 1970). Finally, empirical evidence suggests that Hydrilla cannot grow in water regions deeper than 25 feet (7.62 meters) (New York Invasive Species Information, 2019), so we limit our consideration to water regions of 0-25 feet deep.

3. Simulation

In this section, we build a stochastic simulation model to predict the future spread of Hydrilla. We start by considering the simple setup where no human activity is present and identify plant growth and spread as two main factors of our simulation model. Modifying historical models, we propose our Invasive Plant Population Dynamics (IPPD) model as a better fit for Hydrilla simulation. Geographical information and human activity are then supplemented to finish the simulation.

3.1 Auld's Model

In 1980, Auld and Coote proposed the following classic model on invasive plant growth and spread in the natural environment (Auld & Coote, 1980).

$$P_n = P_{n-1}(1+r)(1-s) \quad (1)$$

Here P_n denotes the density of Hydrilla at time n. r is the population growth rate and s is the spread rate. According to Boughey in 1973, rather than the commonly used logistic function for population growth, empirical evidence suggests that invading plant species have a constant exponential increase rate for a relatively long period. Hence here $(1 + r)$ indicates exponential growth. Auld further assumes that the fraction of spread s is a constant, given that the dispersing fraction from a location is relatively small in relation to the annual increment of the population. Thus $(1 - s)$ denotes the remaining population that does not spread. As stated in Section 2, we will keep these two important assumptions in the following modeling.

3.2 The IPPD Model

One major problem with Audl's model is that the already established plant and the newborn have the same probability of dispersing. Since Hydrilla has the distinctive feature that its tuber roots in the lake's bed, which makes it relatively immobile, not all population can spread. Instead, only the newly grown Hydrilla, whose root yet shallow and tuber not matured, should be able to spread. This leads us to revise the time-step equation as

$$\mathbb{E}[P_n(i,j)] = P_{n-1}(i,j)(1+r(1-s)) + \mathbb{E}[I_n(i,j)] \quad (2)$$

Here $P_n(i, j)$ denotes the Hydrilla population (number of tubers) at position (i, j) at time n. $I_n(i, j)$ is the spread increment at location (i, j) at time n, calculated by the increase in tubers via dispersing from other locations. Expected values are used to reflect the stochastic nature of our model.

Considering the rather mediocre mobility of Hydrilla, we adhere to Audl's model to set $s = 0.05$ and the range of spread to be 3 units within the current location. Although such a s might seem arbitrary, later in sensitivity analysis we will show that the value of s would not effect the qualitative result of the simulation. The range of spread comes as a result of our choice of the bivariate normal distribution (see Section 3.2 for more detail). Since Hydrilla commonly grow to length over 9 meters and grow as fast as 0.3 meters a day (Schuyler County Soil and Water Conservation District), we make the assumption that every time a Hydrilla grows another 9 meters in length, it roots a new tuber in the lake's bed. In other words, a new Hydrilla is grown. Moreover, our time unit is set to a month and $0.3m/day \times 30day/month = 9m/month$, so the number of Hydrilla doubles in a month. Thus we have the formula for the growth rate:

$$r = \frac{9m}{0.3m \times 30day} = 1 \quad (3)$$

Another flaw of Audl's model is that for simplicity reasons, Audl assumes equal possibility to spread from a certain location to any place nearby within a certain range. This contradicts the simple intuition that more population should disperse to locations closer to the source point. Hence, we keep the spread radius the same as 3 units nearby but change the probability distribution of the dispersal from a uniform distribution to a bivariate normal distribution. According to our setup, any place further than 3 units, by calculating the double integral of the normal distribution over the region, has probability near zero and can be omitted. The spread increment $I(i, j)$ can be now calculated by the double integral of the population density times the probability following normal distribution $N(0, I)$. Here $f(x, y)$ denotes the probability density function of the normal distribution, which reflects the probability that Hydrilla spread from location (i+x, j+y) to (i, j)

$$\mathbb{E}[I(i,j)] = \iint f(x,y)P(i+x, j+y)rs\ dxdy \quad (4)$$

Whether a newborn Hydrilla spread or not can be seen as a Bernoulli trial with probability s of success, so the total number of spread $P_{n-1}rs$ follows a binomial distribution. As the number of tubers grows rapidly to a relatively large number, and according to the Central Limit Theorem, we can use normal distribution $N(P_{n-1}(i, j)rs, P_{n-1}(i, j)rs(1 - s))$ to approximate the experiment result. In actual coding we introduce an intermediate variable $E_n(i, j)$ as the Hydrilla population dispersing away from (i, j) at time n, generated by a random variable following the normal distribution. At program run time, we iterate all valid (i, j) in the map and calculate $E(i, j)$ following the Bernoulli process. We then divide the dispersal by the bivariate normal distribution and increment nearby units by the according amount. In other words, $I(i, j)$ is not directly calculated during the simulation but comes as a sum of separate increments from nearby units. Theoretically we should obtain the same simulation result. To focus more on the change of a single unit within each iteration of time, we can express $P_n(i, j)$ as

$$P_n(i,j) = P_{n-1}(i,j)(1+r) - E_n(i,j) + I_n(i,j) \quad (5)$$

We verify this approach by trying on a simple 200×200 map, with the initial Hydrilla at the center (100,100). The side length of a single unit is set as 44m in accordance to later simulations. After simulation of five years, the density of Hydrilla can be shown in the figure below:

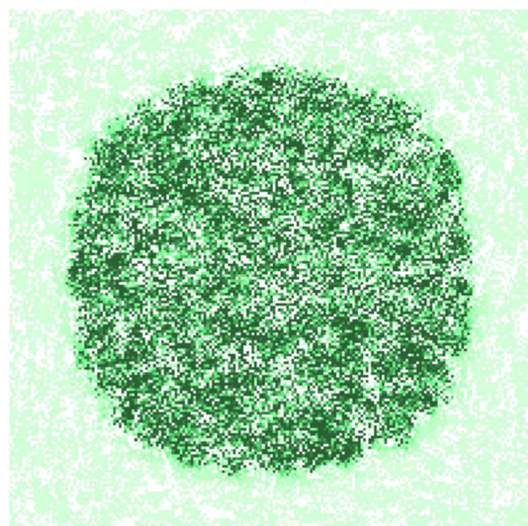

Figure 2: Single Source Spread Simulation

The darkest green signifies a density at 6000 tubers per square meters (the maximum density), while the lightest green represents a density of 0-42, as 42 is commonly used as a benchmark for high Hydrilla density (CMCM, 2021). In other words, a light green grid indicates that Hydrilla is at present in that grid, but only in a low density. After the region is left untreated for five years, Hydrilla has already spread across the entire region, and a large region around the center has a high level of density. Another interesting result is that instead of a graduate change of color across the darkest-lightest boundary (we have in fact 10 different shades of greens in our program!), there is a clear, dramatic distinction between the dark green and the light green. That said, even though Hydrilla quickly spreads to cover the entire region, the area with high concentration grows at a slower rate. But once Hydrilla establishes itself at a new location, the area quickly grows into high concentration, which is explainable by the exponential setup of the model.

3.3 Running the IPPD Model over Cayuga Lake

The map of Cayuga Lake we use captures a region of $51.04km \times 20.77km$ and is represented as a 1160×472 matrix during coding. Each matrix unit thus represents a square with side length

$$\frac{51.04km \times 1000m/km}{1160} = 44m \quad (6)$$

Image processing technique based on OpenCV is employed to produce the gray-scale map that facilitates model coding. Canny edge detection algorithm is executed to produce the shoreline and the 25 feet water depth contour line. Since Hydrilla roots at most 25 feet deep water, we limit our consideration to the white region as seen in the following figures. According to the latest obtainable data from October 2020 (Hydrilla Community Conference, 2020), 36 spots of Hydrilla were found at the southern end of the lake and 3 spots found close to the town of Aurora. We set this as our model's initial state and obtain Figure 3 (see Appendix for high resolution Figure 3.3).

As can be seen in Figure 3, within one year of time dark green that signifies high concentrations of Hydrilla appears near initial spots. The initial small populations of Hydrilla start to establish themselves. Further away, more areas of light green cover nearby regions. This means that Hydrilla is at present at these places but has not established itself yet. By the end of the second year (Figure 3.2), the original light green areas have partially turned into dark green. A much larger area is now susceptible to high concentrations of Hydrilla. After five year's simulation, the Hydrilla continues to invade the rest of the east and south coasts. While the established areas turn darker in green, the spread speed of Hydrilla, to our surprise, dramatically slows down. In other words, the areas susceptible of high concentrations of Hydrilla after five years differs little from the areas susceptible of high concentrations of Hydrilla after two years. Such results also attest the advantage of our model over Audl's model. By introducing geographical information into our IPPD model, we are able to not only predict the global developments, but also make local analysis of Hydrilla that are more sensitive to nearby geography, such as coasts and water depth. This is best seen in our analysis of how the west coast would be free of invasion. Tracking only the total number of Hydrilla, on the other hand, would lose all this information. As we will seen by the end of this section, this might also led to inaccurate results.

In comparison with the spread simulation in the plane (Figure 2), where the Hydrilla quickly spreads all over the map and continue to enlarge established areas, there seems to be an upper bound of the area that Hydrilla could spread. This gives the conclusion that certain geographical features of Cayuga Lake limits the spread of Hydrilla. One possible explanation is that the vast area with water depth more than 25 feet makes the growable region in long narrow strap shape, which in turn hampers the spread of the Hydrilla. Still, this is an unacceptably large area and certain actions need to be taken.

Next, we add boating to our model and verify it as a major contribution to the spread of Hydrilla. The magenta asterisks in the figures represent the marina locations (Tompkins County Planning and Sustainability Department, 2020). For simplicity reasons, we combine nearby marinas as one and give different weights to them according to their combined size (number of marinas, number of boats, etc.). This results in five locations to be considered: Beacon Bay (Cayuga), Frontenac Harbor (Union Springs), Don's (Genoa), Finger Lake Marine Svcs (Lansing), and City Harbor (Ithaca), with weight 3, 2, 1, 1, 3 at each location. We also make the assumption that there is one trip on the Cayuga Lake every day, whose departure and arrival location are randomly chosen from the marinas. In the real world, there are obviously much more boat trips, but for the purpose of this paper this simplification is sufficient to show how boating facilitates Hydrilla spread. Let m_i denote the weight of marina i, then the probability that on a given day, the trip is from marina a to marina b is:

$$p_{ab} = \frac{m_a m_b}{\sum_{i \neq j} m_i m_j} \quad (7)$$

Figure 3.1: After 1 Year.

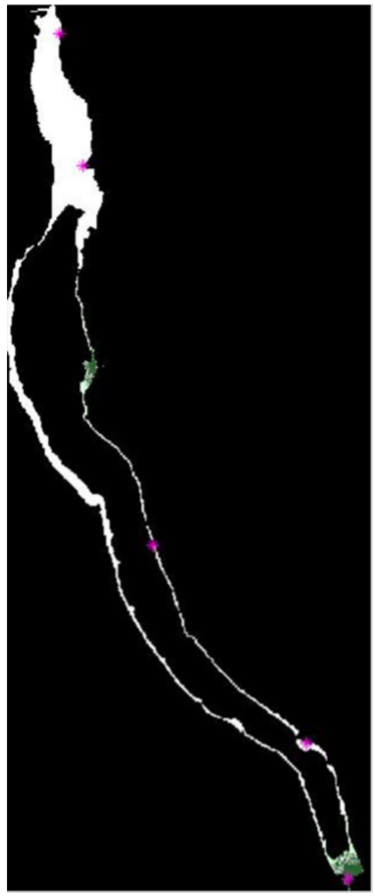

Figure 3.2: After 2 Years.

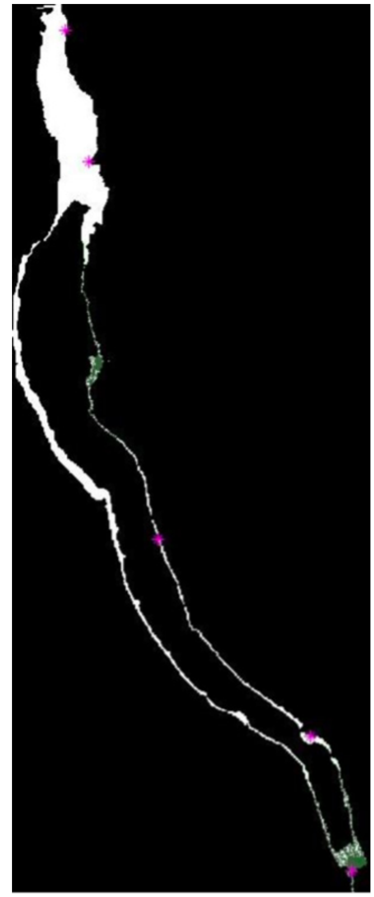

Figure 3.3: After 5 Years (No Boats).

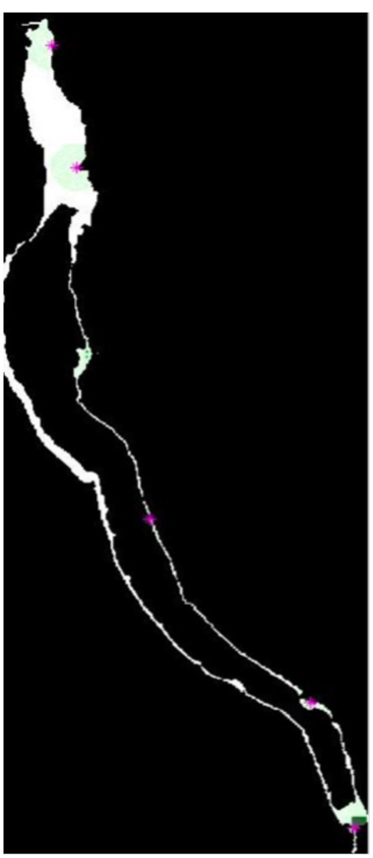

Figure 4.1: After 1 Years.

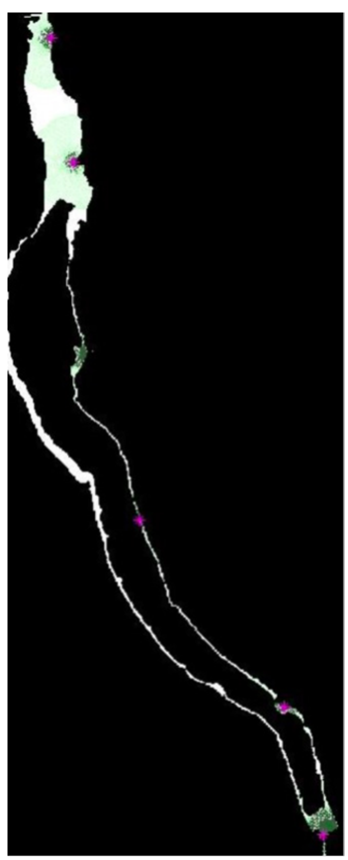

Figure 4.2: After 2 Years.

Figure 4.3: After 5 Years (with Boats).

Therefore, we have the following matrix that represents each p_{ab} probability:

	Cayuga	Union Springs	Genoa	Lansing	Ithaca
Cayuga	0	$\frac{3}{38}$	$\frac{3}{76}$	$\frac{3}{76}$	$\frac{9}{76}$
Union Springs	$\frac{3}{38}$	0	$\frac{1}{38}$	$\frac{1}{38}$	$\frac{3}{38}$
Genoa	$\frac{3}{76}$	$\frac{1}{38}$	0	$\frac{1}{76}$	$\frac{3}{76}$
Lansing	$\frac{3}{76}$	$\frac{1}{38}$	$\frac{1}{76}$	0	$\frac{3}{76}$
Ithaca	$\frac{9}{76}$	$\frac{3}{38}$	$\frac{3}{76}$	$\frac{3}{76}$	0

Figure 5: Probablity Matrix.

With the boating activity added to our model, we obtain our simulation results in Figure 4 (see Appendix for high resolution figure). Clearly a much larger area now has high concentrations of Hydrilla, including the entire north, east, and south part of the lake. The only part free of Hydrilla is the west coast of the lake. This makes sense as there is no marina on the west. In reality, however, boats might pass or temporarily stop along the west coast, which could lead to Hydrilla spread there. Figure 4.1 and Figure 4.2 help us to see how boating facilitate the spread of Hydrilla. Figure 4.1 shows that after one year of time new spots of Hydrilla showcase in all marinas, although in a small amount. This is no longer the case by the end of the second year. The area with presence of Hydrilla has significantly grown larger and they have established themselves around the marinas. From there Hydrilla continues to spread across and essentially covers the entire north of Cayuga lake. We conclude that boating significantly conduces Hydrilla spread by introducing the plant to new locations where it could not reach in natural settings.

Apart from the direct results from the figures, a quantitative analysis of the simulation also help us better support our conclusions. We keep track of the number of units that reach the maximum density of 6000 tubers per square meters, which reflects the total area that has the highest density. This combines the growth rate of Hydrilla with the global spread speed and serve as a good indicator of the overall situation. As can be seen in Figure 5, with boating added to our model, the Hydrilla continuously grow and spread at an exponential speed. In comparison, the Hydrilla reach the maximum capacity much slower when the boat is not present. An initial thought would be that the total number of Hydrilla should grow at an exponential speed regardless of the spread, but in reality Hydrilla would grow to the maximum local capacity and thus slow down the growth rate of the total population. Another interesting observation is that around 27 months of simulation, we see a stark decrease in the increase rate of the maximum density area in the case without boat. This well supports our conclusion that "the areas susceptible of high concentrations of Hydrilla after five years differs little from the areas susceptible of high concentrations of Hydrilla after two years." From another point of view, we verify the existence of an upper limit of Hydrilla spread in the natural setting.

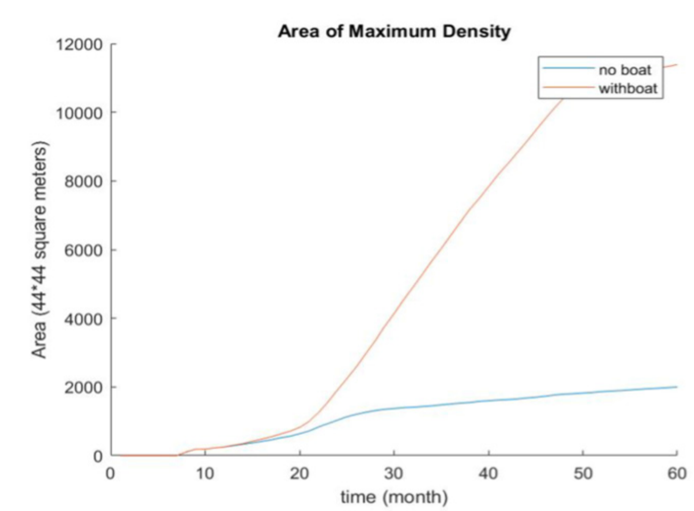

Figure 6: Area of Maximum Density.

4. Sampling

We have obtained the simulation model that predicts the future development of the plant. The question now is how do we determine the initial state of our model? The above simulations use available data in 2020, and clearly, the accuracy of such initial states would be crucial to good simulations. *Quantitative survey* provides a means to obtain accurate data for this purpose (Madsen, 1993). The current survey method at Cayuga Lake monitors previously treated sites by taking small samples of the sediment, called cores, and counting the number of tubers found in each core (CMCM 2021). The *sampling design* of the survey, such as the number of tubers to take, and the subsequent *statistical inference* of the survey results would then help generalize from local information of the samples to the overall situation of Cayuga Lake (Madsen, 1993).

4.1 Spencer's Logistic Regression

In 1994, Spencer D.F. published an important paper on subterranean propagules of submersed aquatic plants, including our most concerned *Hydrilla verticillata*. Spencer collected empirical data at Belle Haven Marina and environs, Potomac River, and Virginia, as well as historical records from Sutton and Portier (1985), Anderson and Dechoretz (1982), and Harlan, Davis, and Pesacreta (1985). The data summed up to 379 sample means and associated standard error from a total of 4942 cores (i.e. sample units). Using PROC REG in SAS (SAS Institute, 1988), Spencer claimed that the frequency distribution of Hydrilla follows not a normal distribution but a log normal distribution. Take log(s^2) and log($\bar{x}$) (all logs are in base 10) there exhibits a clear linear relationship between the two

$$\log s^2 = 1.7039 + 1.2668 \log \bar{x} \quad (8)$$

Here $\bar{x}$ is the average number of tubers per square meters of the samples and s is the sample standard deviation.

4.2 The Baseline Approach

The current sampling design at Cayuga Lake consists of taking 30 cores at each site, with each core in about 0.0187 square meters (CMCM, 2021). Suppose each core contains x_i ($1 \leq i \leq 30$) tubers, then the 200 sample mean per square meters would be

$$\bar{x} = \frac{\sum x_i / 0.0187}{30} = \frac{\sum x_i}{0.561} \quad (9)$$

The standard deviation would thus be

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{30 - 1}} \quad (10)$$

Note that the frequency distribution of Hydrilla follows not a normal distribution, thus we cannot simply add or subtract multiples of standard deviations over the sample mean to get a confident range of the actual number of tubers. Then question naturally arises as to how confident we are in using our sample mean to reflect the real situation? A simple and intuitive validation is to see whether our sample mean fits the logistic regression. If the standard deviation calculated by the samples differs only a little from the standard deviation calculated by the logistic regression, we are confident in our sample mean. Otherwise, we need to increase the sample number until the standard deviations match. In practice, however, it is often too troublesome to calculate the standard deviation from samples, especially when it is constantly changing. An alternative approach often used in the industry (Madsen, 1993; Spencer, 1994) assumes a fixed linear relationship between the standard error and the sample mean. Often, it is assumed that $SE = 0.2\bar{x}$. Plug in $SE = s/\sqrt{N}$, where N is number of samples (cores), and we have

$$s_1 = 0.2\bar{x}\sqrt{N} \quad (11)$$

Rewrite the logistic regression we have

$$s_2 = 10^{0.085195} \cdot \bar{x}^{0.6334} \approx 1.216\bar{x}^{0.6334} \quad (12)$$

Starting with $N = 30$, equation (11) becomes $s = 1.095\bar{x}$ and monotonically increases as N increases. Meanwhile $\bar{x}$ would stabilize around a constant according to the Law of Large Numbers. Thus s_1 and s_2 will eventually meet and the sampling process would terminate.

Finally, there is one last adjustment in the current approach. Given the sensitive nature of the issue, we want to be more conservative in our conclusion. Hence, when the sample data does not fit in the logistic regression, if $\bar{x}$ is above the curve (given same s has higher $\bar{x}$), we are more prone to believe this outlier actual exists and we should take immediate action to control the high concentration of Hydrilla. On the other hand, if $\bar{x}$ is below the curve, we are more suspicious of our sample results and should take more samples to fit the curve. In practice, 42 tubers per square meter is often considered the benchmark for high concentration of Hydrilla. Thus during our sampling process, if $\bar{x} \geq 42$ at any time, we cease the sampling and believe immediate action should be taken.

4.3 An Adaptive Approach

To be even more efficient, we can calculate the minimum sample size N instead of the standard deviation. The idea is to assume that the theoretical standard deviation equals the sample standard deviation to get a formula of N. Combining equation (11)(12) we have

$$N = \frac{1}{0.04} \cdot 10^{1.7039} \cdot \bar{x}^{-0.7332} \quad (13)$$

If the calculated sample number is greater than the actual sample number, by some simple mathematics we know this implies that $s1 < s2$, and we need more samples to fit the curve. Instead of comparing standard deviation, we can now directly calculate the minimum sample size. Note that as $\bar{x}$ changes as more samples are collected, according to equation (13), the minimum number of samples N also changes. Thus we come up with the following adaptive algorithm—

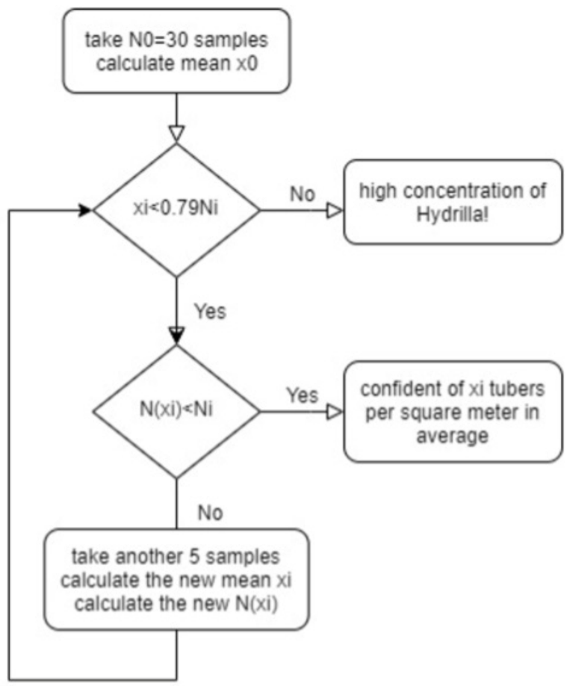

Figure 7: An Adaptive Algorithm for Sampling

Here x_i denotes the average number of tubers of all samples at iteration i, N_i denotes the performed total number of sampling at iteration i, and $N(x_i)$ denotes the minimum number of sampling to be confident of x_i. 0.79 is calculated from 0.0187 * 42, following the same procedures at the beginning of section 5.1. If the sample average is greater or equal to the high concentration thresh hold, we exit the program and claim the area has a high concentration of Hydrilla. If the total sample number is greater or equal to the minimum sample size, we also exit the program and claim that there expect to be x_i tubers per square meter. If neither of these conditions is satisfied, however, we take another 5 samples and recalculate the sample mean and minimum sample size. Notice however that as the sample average goes to 0, the minimum sample size will go up to infinity. Thus in practice it is recommended to set also a upper bound of the sample size as 100.

5. Sensitivity Test

5.1 Simulation

In our current model we set $s = 0.05$ based on the fact that Hydrilla's tubers are relatively immobile and in Auld's model the spread rate is set to 0.05 for mediocre mobility plants. A more direct statistics of Hydrilla spread would be much better but is hard to obtain. This leads us to double check the spread rate.

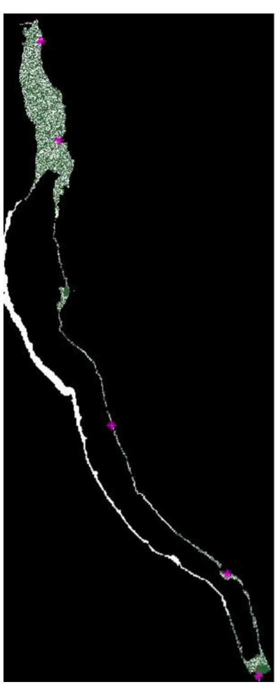

Figure 8: $s = 0.01$

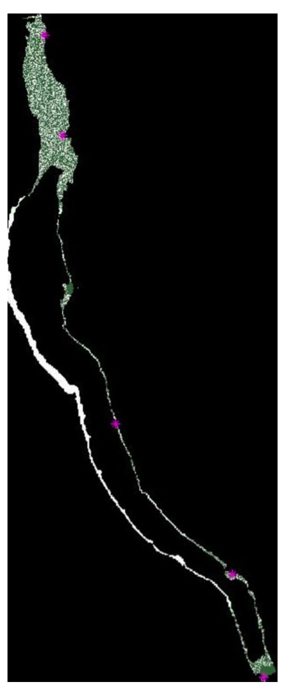

Figure 9: $s = 0.05$

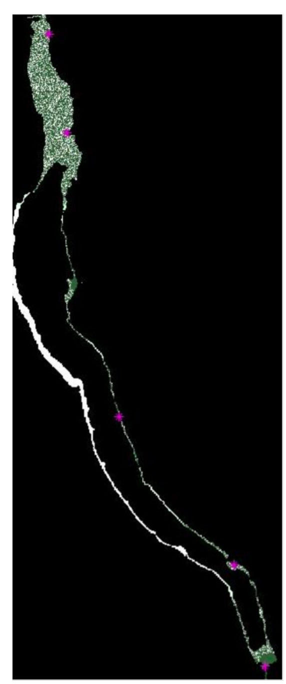

Figure 10: $s = 0.1$

We test our model with different spread rate: $s = 0.1$, $s = 0.05$ and $s = 0.01$ while keeping other parameters fixed. A closer look reveals that the dark green regions become sparse as s decreases, which means that less areas have a high density of Hydrilla and more areas have zero density with a lower spread rate. As spread rate increases, Hydrilla has a higher probability to spread from one area to others, and more areas would be affected. On the other hand, all of these three spread rates display the similar distribution pattern of the plant and only differs in the subtle density of the dark green. Thus choosing 0.05 as a middle ground value would not lead to severe inaccuracy in terms of the qualitative simulation result.

5.2 Sampling

Another important parameter that we need to test is the standard error of the sample mean. In either the baseline approach, the adaptive algorithm, or historical works, the standard error is set to be an arbitrary value of 0.2. Here we test how sensitive the minimum sample size is in our adaptive algorithm. We vary the standard error from 5% to 50% of $\bar{x}$ and plot the number of samples needed versus this proportion in Figure 10. The conclusion from the figure is that if we are willing to tolerate higher variability in our samples, we can pull fewer cores to infer the overall picture.

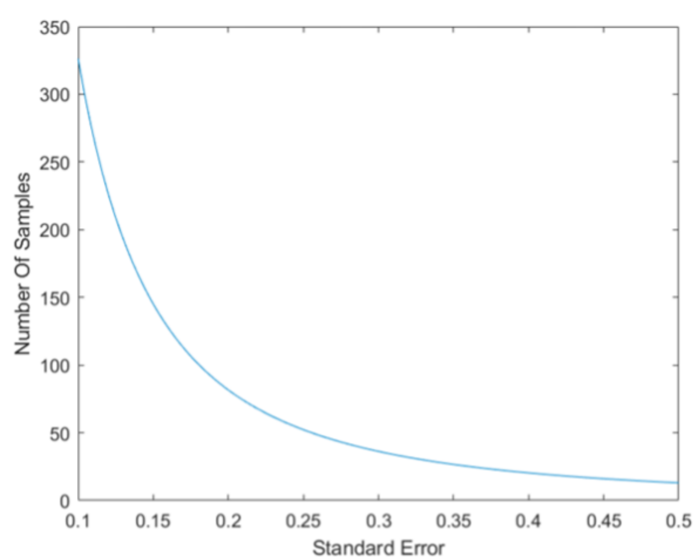

Figure 11: Minimum Sample Size in terms of SE

Often time in practice we want a rough estimate of the minimum sample size in terms of the current sample mean, and we want our rough estimate to not fluctuate too much so that the adaptive algorithm terminates faster. This leads us to vary the sample mean and see how the minimum sample size changes (Figure 11).

From Figure 11, the required number of samples decreases while the sample mean increases. Note that the minimum sample size varies greatly when the sample mean is how, which justifies why we should be more conservative of our conclusion when the sample mean is low.

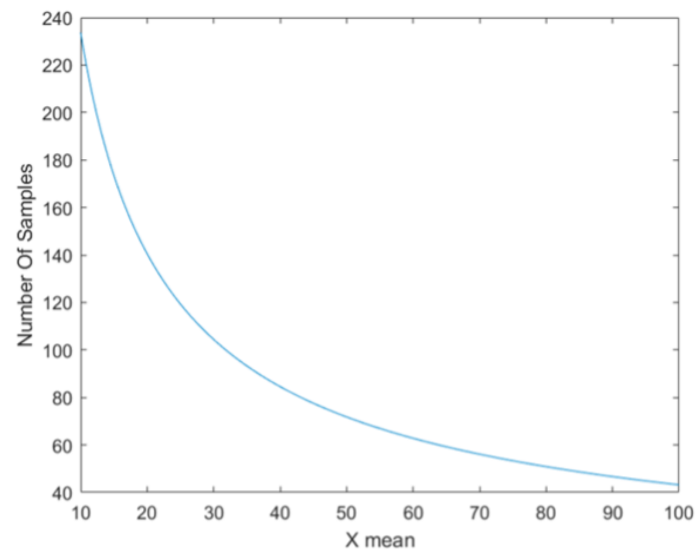

Figure 12: Minimum Sample Size in terms of Sample Mean

6. Discussions

Due to the distinctive nature of simulation of sampling, the simulation results and analysis and the reasoning of the statistical inference are discussed in detail separately in Section 3, 4. This section mainly elaborates on some of the possible future works after our study.

Refine Model So far, our model incorporates the natural growth of the plant, the dispersal of the plant, and boating as the main factor of the human spread of the species. One potential future work is to extend the current model to incorporate more human activities, such as cleaning the plant using fluridone. Environmental factors such as temperature, sunlight, wind, water level, etc., might also significantly influence the growth and spread of the plant. Also, in this paper, only the ten marinas alongside the Cayuga Lake are considered for boat activities for simplicity. One can include more details for boating activities by also considering boat ramp launch site and paddlecraft launch site.

Other Species Another direction is to modify the parameters to model other plant species with similar behaviors. For the spread simulation only the growth rate

r and spread rate s follow directly from the features of Hydrilla. While in the sampling assessment, one can run linear regression on different species to get different equations between σ and $\bar{x}$, and the assessment method as well as the adaptive algorithm should still work.

Update Data For the sampling assessment, this paper directly uses Spencer's empirical results of Hydrilla in 1994. More accurate and relevant data can be obtained from direct surveys and historical data for Cayuga Lake protection. One might also seek more up-to-date data and reproduce Spencer's regression results.

Practical Concerns Our analysis mainly focuses on the quantitative and theoretical aspects of the simulation/samples. In practice, other than how many samples to take, there is also the concern of how to take those samples. Should the sample points be selected at regular intervals or completely random? What utilities better collect the samples? Discussion on these practical questions also has theoretical importance. For example, the actual error rate of our sample mean might depend on the method of sampling as mentioned above.

7. Concluding Remarks

Overall, this paper aims to provide a practical, quantitative, and real-time evaluation system that helps monitor and predict the population dynamics of the plant. The system includes two main parts: sampling and simulation. In Section 3, we discuss in detail our stochastic Monte Carlo simulation based on the enhanced *Invasive Plant Population Dynamic* (IPPD) model. Human activities, such as boating, are identified to greatly accelerate the spread of Hydrilla. In Section 4, we employ statistical methods to promote confidence in the sample results, which in turn guarantees more accurate simulation results. We provide an adaptive algorithm to calculate the minimum sample size in real time. Altogether, the authors reach the following conclusions: (1) if left unattended, a small population of Hydrilla could quickly establish itself and expand to nearby areas; (2) in natural settings Hydrilla could only cover limited areas of Cayuga Lake due to the geographical features of the lake; (3) boating significantly conduces Hydrilla spread by bringing Hydrilla to new locations that it could not reach in natural settings; (4) generally speaking, low sample mean indicates not low density of Hydrilla, but that more samples are needed.

References

[1] Auld, B. A., & Coote, B. G. (1980). A Model of a Spreading Plant Population. Oikos, 34(3), 287. https://doi.org/10.2307/3544288

[2] Blackburn, W. H., & Tueller, P. T. (1970). Pinyon and Juniper Invasion in Black Sagebrush Communities in East-Central Nevada. Ecology, 51(5), 841–848. https://doi.org/10.2307/1933976

[3] Pennsylvania Sea Grant (2017). Hydrilla. https://seagrant.psu.edu/sites/default/files/Hydrilla%202017.pdf

[4] Schuyler County Soil and Water Conservation District. Hydrilla.
http://www.schuylerswcd.com/Hydrilla.html

[5] Cayuga Lake Hydrilla Task Force (2021). Cayuga Lake Hydrilla Management Plan 2021–2026. Cornell Co operative Extension. https://s3.amazonaws.com/assets.cce.cornell.edu/attachments/50128/Cayuga Lake Hydrilla Ma nagement Plan 2021 March 10 2021.pdf?1617048676

[6] Boughey (1973). Ecology of Populations. 2nd Ed.

[7] WRC Education Committee of Tompkins County & Tompkins County Planning and Sustain ability Department (2020). Clean boating combo map - Tompkins County NY. Tompkins County NY. https://www2.tompkinscountyny.gov/files2/planning/committees/WRC/documents/clean-boating combo map brochure-2020-web.pdf

[8] Fall 2020 Community Conference Hydrilla focus 11.4.2020 (2020). Cayuga Lake Watershed Network. https://www.youtube.com/watch?v=Qf3Xj21gslw&t=2473s

[9] Madsen & Bloomfield (1993). Aquatic Vegetation Quantification Symposium: An Overview. Lake and Reservoir Management, 7(2), 137–140. https://doi.

org/10.1080/07438149309354265

[10] Madsen, J. D. (1993). Biomass Techniques for Monitoring and Assessing Control of Aquatic Vegetation. Lake and Reservoir Management, 7(2), 141–154. https://doi.org/10.1080/07438149309354266

[11] Spencer, D. F., Ksander, G. G., & Whitehand, L. C. (1994). Estimating the abundance of subterranean propagules of submersed aquatic plants. Freshwater Biology, 31(2), 191–200. https://doi.org/10.1111/j.1365-2427.1994.tb00853.x

[12] Hydrilla – New York Invasive Species Information (2019). New York Invasive Species Information. Retrieved November 15, 2021, from http://nyis.info/invasive species/Hydrilla/

[13] Downing, J. A., & Anderson, M. R. (1985). Estimating the Standing Biomass of Aquatic Macrophytes. Canadian Journal of Fisheries and Aquatic Sciences, 42(12), 1860–1869. https://doi.org/10.1139/f85-234

[14] Missouri Stream Team. Invasive Species Alert-Hydilla. http://mostreamteam.org/assets/factsheet26.pdf

[15] Sutton D.L. & Portier K.M. (1985). Density of tubers and turions of Hydrilla in south Florida. Journal of Aquatic Plant Management, 23, 64-67.

[16] Cornell MCM (2021). Closer to Cayuga's Waters. https://e.math.cornell.edu/sites/mcm/ CMCM 2021 Cayuga problem.pdf

Appendix High Resolution Figures

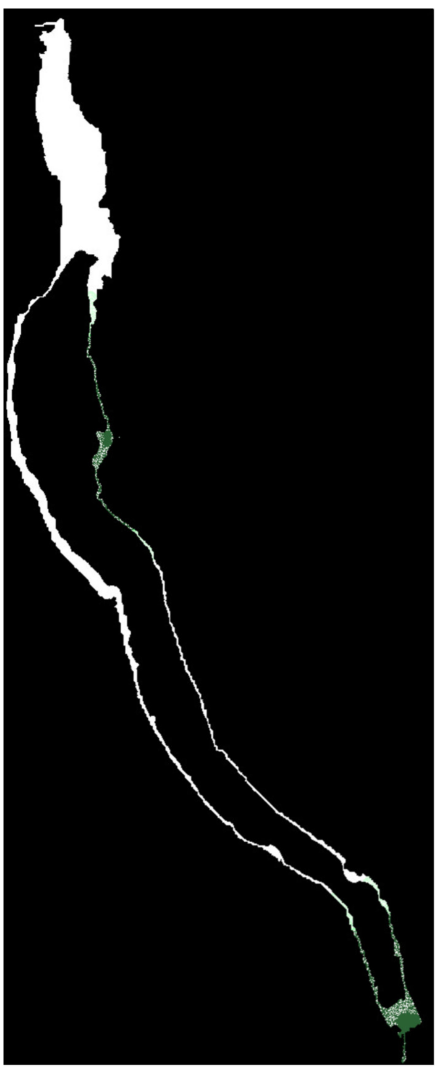

Figure 3: After 5 Years (No Boats).

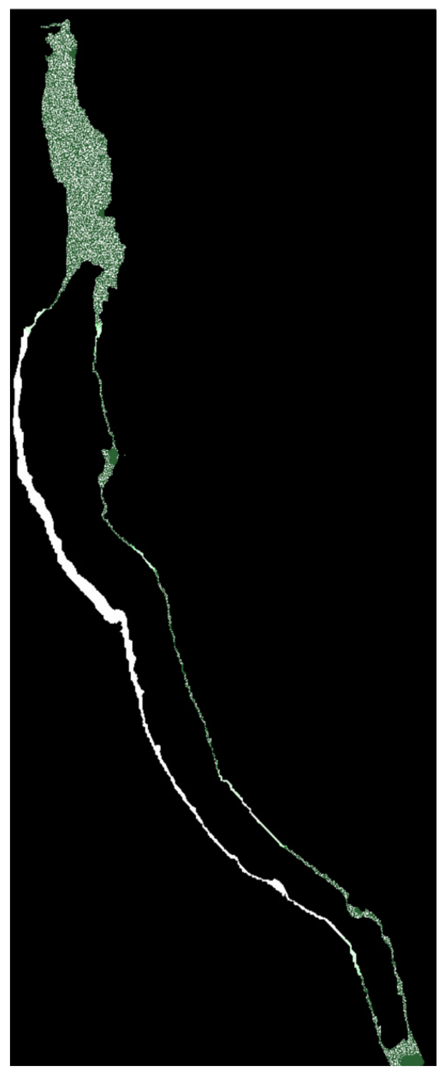

Figure 4: After 5 Years (with Boats).