Simulation of the hippo population in Columbia as an invasive species

Mika Mauro Rother
Dipartimento di Informatica
Università degli Studi di Milano
Milano, Italia
mikamauro.rother@studenti.unimi.it

***Abstract*—As an invasive species the Hippopotamus amphibius lives in Colombia along the Magdalena river. Compared to other Hippopotamus around the world the Colombians actually grow in their population size. My approach is now to simulate the population growth over the next years by considering an amount of important factors, which can be human-influenced or environmental. Using a logistic growth function, I try to predict how the amount of Hippopotamus amphibius increases. The key point in this simulation is to figure out the carrying capacity, so the maximum number of possible living species in a specific habitat.**

***Index Terms*—simulation, hippopotamus, population growth, agent based, carrying capacity**

I. Introduction

What do a drug empire and the Hippopotamus amphibius (in the following just *hippos*) have in common? Both were possessions of Pablo Escobar, but only the hippo population is still increasing. (There are still a lot of drugs, but not such a large empire than to Escobar times). So this is how the story begins: In 1981 Pablo Escobar brought four hippos to Hacienda Nápoles, Colombia for his private joy. But after he died in 1993, these hippos were still living in Colombia and over the years they became more and more. In Fig. 1, Subalusky et al. provide an overview of hippos along the Magdalena river (for more see [1]).

Although this snapshot is quite a few years old, it shows that the number of hippos is increasing. In [2], Castelblanco-Martínez et al. estimate that the number of hippos in the river could be between 65 and 80 or even higher. It is therefore an interesting topic, and worth an approach to predict the population growth of hippos as an invasive species over the next years.

So how large can such a population grow? There are already a lot of scientific work which try to predict this outcome (e.g., [1], [2], [5]), but as all predictions about a few years in the future you never know what could happen. Which influences can be dangerous for the growth of a species you cannot include in your predictions, because they are not always known yet. It is then even more important to make precise assumptions about the things we already know (with high percentage).

The main target of my approach will be that I try to simulate the carrying capacity K of the population growth model, and then apply it to a logistic growth function in order to

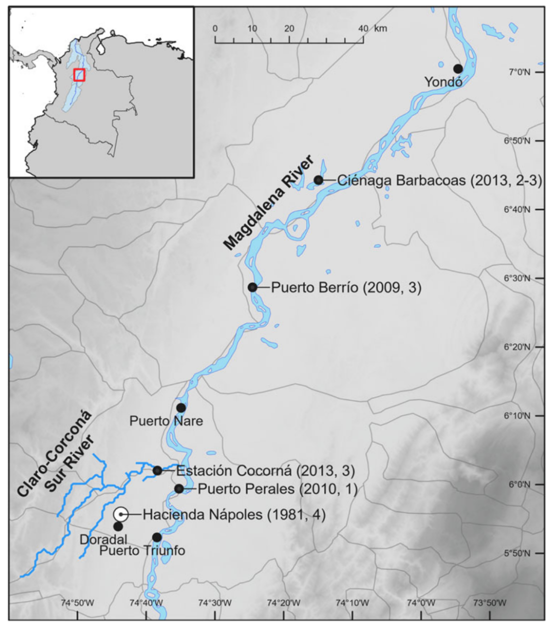

Fig. 1. Range of the hippopotamus in the Magdalena River. (Figure taken from [1])

see how the population could develop in the next 100-150 years. To complete this model already estimated growth rates r are used that were considered by Subalusky et al. in [1] for their exponential growth functions. One will see that a lot of simulations are needed to get a reasonable value for K and a trustworthy model.

II. Model

When it comes to the point of simulating a population growth, there are two different approaches, i.e., two types of population growth models:

1) Biotic potential (exponential growth), so a growth without any limits. This a more theoretical approach, since there must be always a limit in the size of a population.
2) Environmental resistence (logistic growth), which is growth that is limited through competition.

As I try to get a relative precise prediction of the hippo population growth, it is clear that I will focus myself on the second approach, the logistic growth. Given some parameters as the growth rate r, the initial population P_0, and the carrying capacity K, the population growth over time t is given by

$$P(t) = \frac{KP_0 e^{rt}}{K + P_0(e^{rt} - 1)} = \frac{K}{1 + \left(\frac{K - P_0}{P_0}\right) e^{-rt}} \tag{1}$$

where $\lim_{t\to\infty} P(t) = K$. So the carrying capacity K is the maximum number of population that can be reached.

Obviously there could be many other approaches that could predict the hippo population in Colombia better, but for my case I now assume that Eq. (1) is a suitable model for us. The next step is now to fix the three parameters r, P_0, K in order to plot the population growth $P(t)$. In [1], Subalusky et al. estimating different growth rates r between 5% and 11%. According to this range, I will apply them in my experiments later. For the initial population the amount of hippos that are living right now in Colombia are used. Since the number is not known exactly, I will choose a number somewhere around 65-100, which are reasonable values according to [1] and [2].

The tricky part is K. To find out the maximum number of hippos that are able to live in Colombia together there are many paramters to be taken into account, because population growth can be determined by density-dependent or density-independent factors.

- Density-dependent environmental factors are influenced by the relative size of a population, e.g., predators, availability of resources, etc.
- Density-independent environmental factors are not influenced by the relative size of a population, e.g., phenomena (natural disasters), weather, etc.

You can see that there are a large bunch of factors that influences the possible number of hippos alive. For reasonable results a simulation is needed.

For this I use an agent based simulation, where the hippos meme the agents. I also have to make some assumptions about the key system parameters, since not all of them can be precisely determined. In addition also simplifications have to be made. That means that there must be cut out a lot of parameters for the simulation that could influence the real system to make it still feasible.

Now I would like to give an overview about the system, influences and parameters that are used to build my model. Of course there must be an initial population of hippos to start with, this is my first *Key System Parameter* (KSP). Since my goal is to find the carrying capacity K, the initial number has to be independent of the system. So the outcome of K must be uncorrelated to it. E.g., an outcome of $K \approx 2000$ should be the result of both initial values $x_1 = 10$ or $x_2 = 100$.

Then as all other animals aswell, also hippos are born naturally and even die naturally. In [4], the San Diego Zoo Wildlife Alliance found out that under good conditions female hippos can produce one calf each year. In the wild their sexual maturity starts between 7 and 15 years. Since they die between 30 and 40 years of age, they are able to get children for around 60% of their lifetime, also in their higher ages.

Besides their natural reproduction and death rates, other factors can reduce the population. From the large amount of possible factors I use the density-dependent ones *Human Influence* and *Resources*, and the density-independent one *Phenomenon*. Let's discuss first the latter one. Phenomenon is meant as a general term which includes things like natural disasters, weather and climate change, so things that are not influenced by the hippo population itself. Since weather and environmental conditions are getting worse, and will be even more worse in the future, the probability for phenomenon should be choosen sowehow random, but it should be higher the more you look in the future.

The human influence and the resources parameters depend on the number of hippos. The more hippos there are, the more danger exists for the Colombians. In [1], Subalusky et al. report already about an approach by the Colombian government to hunt and kill some hippos. These approaches will happen more often the more hippos live along the Magdalena river, since also the human population increases. This has also an influence on the amount of resources. The more you look into the future, the less resources are available. So the probabilty that there are enough resources for the hippos should decrease over time, but is also correlated to the number of hippos. The more hippos, the more food and water is needed, which makes an influence too.

III. Implementation

The implementation of the approach is completely written in Python. For my goal it is sufficient to use simple Python code combined with some helpful modules like *numpy* or *random*. In addition for the visualization of the results the modul *matplotlib* is used to plot the results (for more see Section V).

In the following I will provide some code snippets with additional explanation to get a well understanding of the idea of the program and the whole simulation approach. For even more detailed insights, you can find the complete code in [3].

Let's have a look on the following five functions:

```
def get_num_births(num_hippos, birth_rate_mean,
     birth_rate_dev)
def get_num_deaths(num_hippos, death_rate_mean,
     death_rate_dev)
def get_num_human_deaths(num_hippos,
    human_influence_mean, human_influence_dev)
def get_num_resource_deaths(num_hippos,
    resources_mean, resources_dev)
def get_num_phenomenon_deaths(num_hippos,
    phenomenon_mean, phenomenon_dev)
```

Listing 1. "Get functions to find out the population growth"

The first function is used to calculate the number of hippos that are born in one year using the size of the hippo population of the year before, the average birth rate, and the standard deviation for the birth rate. On the other hand, the other four functions calculate the amount of hippos that die each year in a similar way. In every one of these five functions a random number within the normalvariate (of mean and standard deviation for the specific case) is used in order to get a result.

In addition some other functions are needed that tell us if there is for example human influence in a specific year. Therefore the following three functions will help us.

```
def is_there_human_influence(num_hippos,
    current_year)
def enough_resources(num_hippos, current_year)
def is_there_phenomenon(current_year)
```

Listing 2. "Boolean type functions"

As already written, the first function answers the question if there will be human influence in a specific year. I decided that in the first two years the humans are not able to do that, since it will need some time until they are able to regulate the number of hippos by themselves. After that, humans will try to reduce the number of hippos when there are 400 or more hippos alive, in this case the function returns *True*. The second function returns *False* when there are not enough resources for the hippos. This will be the case if there are too much hippos or when you are going further into the future, since the amount of resources on planet earth decreases. The third function in this listing answers the question if there will be some phenomenon in one specific year. Since this is something that does not depend on the hippos itself, the function is written more random, but with a higher probability of returning *True* the more you look into the future.

But the central function of my program is of course SIMULATE(NUM_YEARS). Using the functions described before as auxiliary functions, and the classes KSP and Key Performance Indicators (KPI) as parameters/arguments for the functions, SIMULATE(NUM_YEARS) returns an array with NUM_YEARS entries, where each entry describes the amount of hippos that are alive in a specific year.

One disclaimer: My time scale is years! Of course you cannot determine what happens first, a hippo gets born, or killed, or a large amount of animals got killed by a natural disaster. So I decided to use the same number of hippos at the beginning of each iteration for every function.

The following algorithm then describes every iteration inside the simualtion function (one iteration is equal to one year):

1) Start with a specific number of hippos that is saved in the population size class member of the KPI class (or initialized with an initial value before the first iteration).
2) Calculate the number of births, deaths, hippos killed by humans, hippos died because of the lack of resources, or deaths through to phenomenon, using for every function the number of hippos as base that was determined in Step 1).
3) Update the population size by adding the births and substracting all deaths from the current number.
4) Append the new population size to the history of the population growth.

So what we have got so far? The simulation function returns just one array with one example of how the population size could develop. Neither is that the final goal, or is representative as a scientific result. To get this, also a function that yields to a number K is needed, which must be representative! For this the function GET_CARRYING_CAPACITY(NUM, YEARS) is used, which has two arguments. The first should result in a number of simulations to perform. Inside the function there are two for loops, an outer, and an inner one. Each of the loops will perform NUM (let's use N from here on) iterations, so in total $N \times N$ simulations. In the inner loop the maximum value of the array that the simulation function returns is then taken. This is the largest amount of hippos in this specific run. Then this maximum value is added to an array. Since this happens in each iteration, there is an array of size N with all maximum values of N simulations. Then the *mean* of this array is calculated, which yields to one average maximum value after N inner loops. In addition there are also N outer loops, so such an average maximum value is created N times. Then you add each one of them to an array, which is the result after all simulations have been done. The last step is then to calculate again the *mean* value of this array. After making a cast to get an integer instead of a real number, a reasonable value K for the carrying capacity is received.

Last but not least obviously the logistic growth function must be implemented to finish the model. Here I easily implement Eq. (1).

IV. Key System Parameters (KSP) and Key Performance Indicators (KPI)

As written before the key point to make my approach work is to choose the proper key parameters. Let's start with the simplest one. The initial population, that must be uncorrelated to the outcome of K. So I just choose 100 for it, since it's also the estimated value of the current amount of hippos (see Section II). All KSP are then implemented into one class, which is further defined as follows

```
class KSP:
    initialPopulation = 100
    birthRate = 0.3
    deathRate = 0.03
    birthRateDev = 0.03
    deathRateDev = 0.01
    humanInfluenceRate = 0.2
    humanInfluenceDev = 0.025
    resourcesRate = 0.05
    resourcesDev = 0.01
    phenomenonRate = 0.1
    phenomenonDev = 0.05
```

Listing 3. "Key System Parameters"

After the initial population one have the average birth rate and its standard deviation. As written in Section II female hippos can have children in around 60% of their lifetime. The assumption was made that the distribution between male and female hippos is 50/50. As one hippo can get around one calf each year, the mean birth rate is fixed to 0.3 (30%), with a standard deviation of 0.03 (3%). Since the birth of new hippos is the only thing that increases the amount of them, it has the biggest influence on the KPI parameter POPULATIONSIZE. The evaluation and measuring of the new born hippos is performed by the KPI parameter NEWBIRTHS.

Compared to the birth parameter, there are four different ways how the population could decrease. The most influence in my approach have the humans. If the KPI parameter HUMANINFLUENCE is true, then they will try to decrease the amount of hippos by 20%, with a standard deviation of 2.5%. With 10% and a standard deviation of 5% the next biggest danger for the hippos are phenomena, followed by an average death rate of 5% with 1% standard deviation through the lack of resources. The reasons for these choices are that a phenomenon like a natural disaster will probably kill more hippos at once than the lack of resources, but is on the other hand less probable. The death rate for the natural death is the lowest one of all of them (with only 3%, and standard deviation 1%), but is therefore guaranteed every year.

I already mentioned until now some KPI's. The remaining ones are listed up here:

```python
class KPI:
    populationSize = 0
    newBirths = 0
    newDeaths = 0
    humanInfluence = False
    humanDeaths = 0
    resources = True
    resourceDeaths = 0
    phenomenon = False
    phenomenonDeaths = 0
```

Listing 4. "Key Performance Indicators"

To summarize this Section quickly, NEWBIRTHS is influenced by the KSP's that giving information about birth, all "*death*" KPI's are influenced by their respective KSP parameters. And the most important KPI, the POPULATIONSIZE is influenced by all indicators and parameters, since it is a sum of all births, deaths and the current population size.

V. EXPERIMENTAL ANALYSIS

Now the model is defined, the implementation is also clear and the most important parameters were worked out, so the experiments can start.

The first part of the experiments is to look if and how the simulation function works. For that let's have a look on the Figures 2-4.

Here five simulations were performed three times. In Fig. 2 for 100 years, in Fig. 3 for 150 years, and in Fig. 4 for 200 years time span. The first thing that is obvious by watching these figures is that the simulation function works, since you get a high variation in the number of hippos in each simulation run. And these are just 15 random examples. If you would perform them again you would get three completely different looking figures. The only thing that all simulations have in common is that their curves raise until around 20 years similarly. The reason for that is simple: Since there is no lack of resources in the first 20 years and the probability of phenomena is also really low, there are not so many influences that result in death hippos (just humans and natural deaths).

But after the first 20 years the behavior of each simulation is widely spreaded. This shows that my approach is not a good one if you would like to use it directly for a population

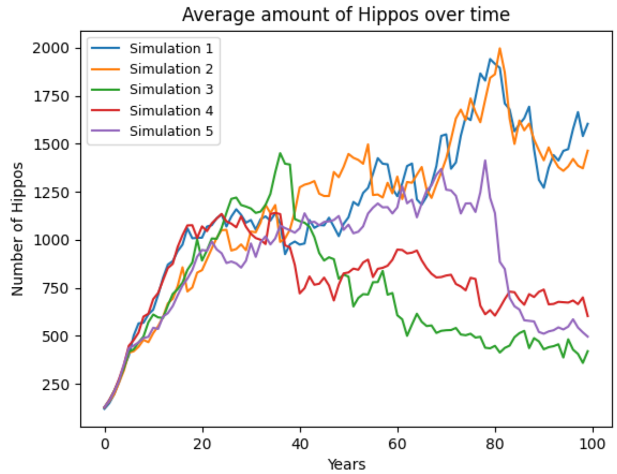

Fig. 2. Five runs of the simulation function for 100 years.

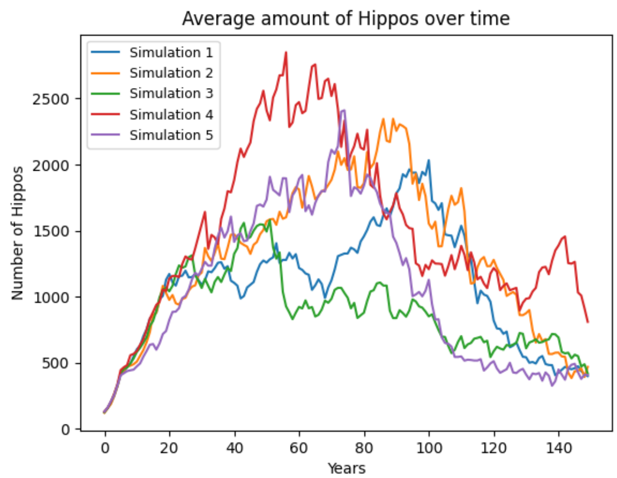

Fig. 3. Five runs of the simulation function for 150 years.

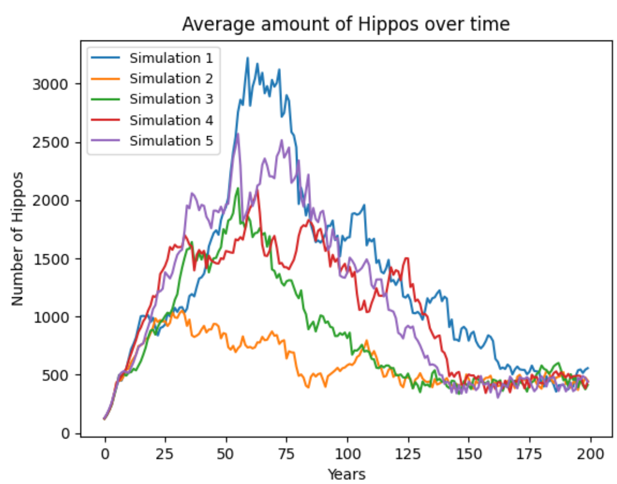

Fig. 4. Five runs of the simulation function for 200 years.

growth model. It would also be hard to estimate growth rates from these plots, because there are so many jumps from high to low and backwards. This is also one reason why I will use already estimated growth rates for my logistic growth function. However the key interest is the carrying capacity. It has been already explained how to get a reasonable value for it in Section III, so it will be not repeated here. You can extract from each of the Figures 2-4 the maximum value for the number of hippos, which is the carrying capacity K for each run.

Then I decided to create logistic curves once for 100 and once for 150 years, since after around 100-150 years there is a strong decrease in the number of hippos as you can see in Fig. 4. The reason for this is that natural phenomena are now more probable and that there is always a lack of resources through to our parameter choice. So it has been shown that the simulation approach works and can now be used to get reasonable values for K, getting a prediction for the population growth for the next 100-150 years.

In [1], Subalusky et al. have provided four different growth rates which they consider possible, $r = 0.05$, $r = 0.07$, $r = 0.08$ and $r = 0.11$. Of course this choices makes every rate between 0.05 and 0.11 reasonable too, but I will use exactly these values for the experiments. So now the growth rates are fixed, the remaining two paramters that have to be choosen for the logistic growth function are the carrying capacity and the initial population. The latter one will be, as already explained in Section IV, initialised with 100. And the carrying capacity will be calculated with the program.

Figures 5-8 show the logistic curves plotted with respective to their carrying capacity for all four different growth rates over the time span of 100 years. The first thing that you can see when you compare all these figures with each other is that the values of K are in a really similar range, which proves that they are reasonable! In Fig. 5 you further see that with a growth rate of $r = 0.05$ the population doesn't converge to the carrying capacity after only 100 years.

Otherwise for Figures 6-8, here one can see the convergence to K for the population growth, even when it is close for $r = 0.07$ in Fig. 6. Obviously, the higher the growth rate, the earlier the convergence towards K begins.

Through the choice of a logistic growth function there is a turning point t_ξ in which the acceleration of the population growth is 0. Which means until that point the population growth increases with an positive acceleration and after that point the acceleration is negative, although the population size is still increaing afterwards, but slower than before. Since it is interesting where that point is, let's find that out.

The first approach to find t_ξ is to derive Eq. (1) two times and set the result equal to 0. Then transform the equation according to t which is then t_ξ and you have the result. For this you need the second derivative with respect to t, which is

$$\frac{d^2}{dt^2}P(t) = \frac{KP_0(P_0 - K)r^2e^{rt}(P(e^{rt}+1) - K)}{(P(e^{rt}-1)+K)^3} \tag{2}$$

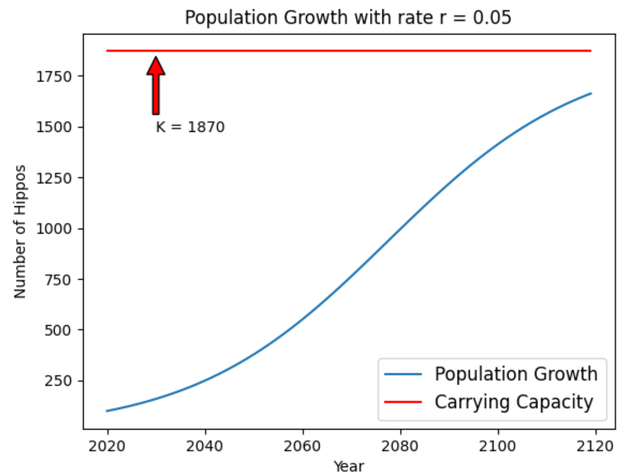

Fig. 5. Population growth, with $r = 0.05$ over the next 100 years.

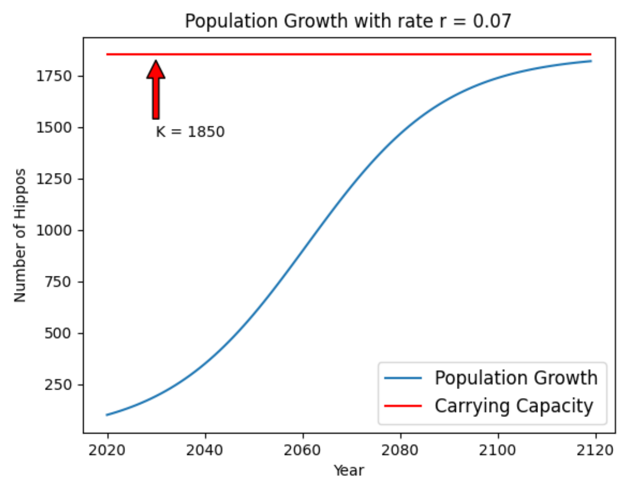

Fig. 6. Population growth, with $r = 0.07$ over the next 100 years.

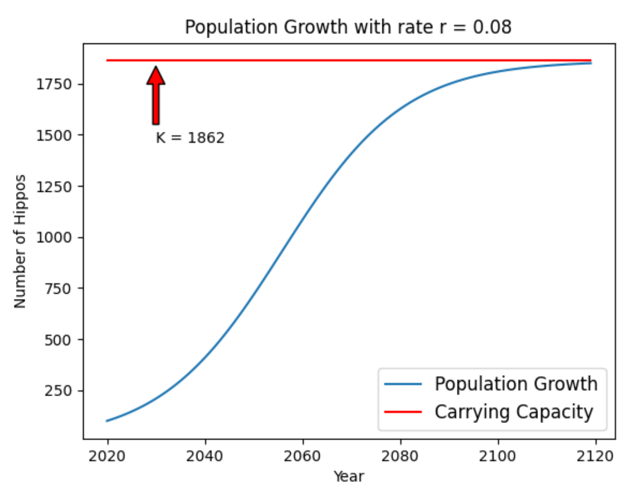

Fig. 7. Population growth, with $r = 0.08$ over the next 100 years.

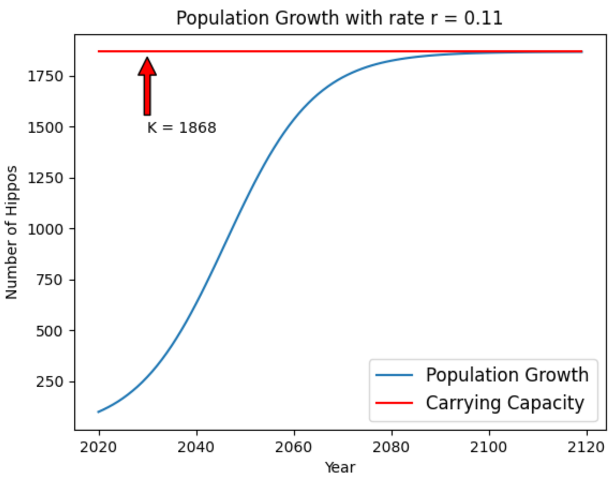

Fig. 8. Population growth, with $r = 0.11$ over the next 100 years.

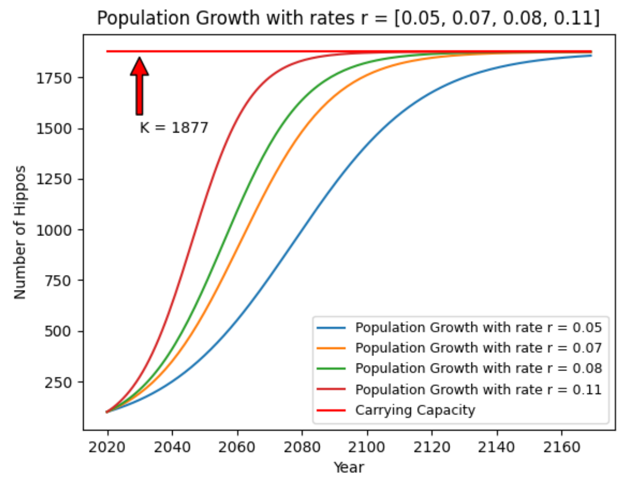

Fig. 9. Population growth, with all four growth rates over the next 150 years.

To transform Eq. (2) according to t and find the result would take some time, so a different approach was chosen instead, to take the value from the experiments. Of course the result is not completely exact, but with a closer look on the Figures 5-8 one can figure out t_ξ.

For $r = 0.05$ you get $t_\xi \approx 2075$, for $r = 0.07$ you get $t_\xi \approx 2062$, for $r = 0.08$ you get $t_\xi \approx 2055$ and for $r = 0.11$ you get $t_\xi \approx 2047$. In [1], Subalusky et al. consider that the growth rates of $r = 0.07$ and $r = 0.08$ are the most probable. Since they have chosen to plot the exponential growth instead of the logistic one, I can compare my results only until t_ξ, because the results of my approach until that point have a similar form like the exponential one. For growth rates between 7-8% a year, Subalusky et al. estimate around 400-800 hippos in the Magdalena river by 2050. With looking on the curves of the Figures 5-8 you would estimate an amount of around 800-900 hippos by 2050, so a bit more than Subalusky et al.

In [5], Shurin et al. also use a logistic curve for their approach. They expecting a full-carrying capacity of 1418 ± 144 individuals. So around 450 less than in my approach, since the value for K of that one is somewhere around 1840-1890.

In the final experiment I also want to plot all logistic growth models for the four considered growth rates in one image, which is given by Fig. 9.

Here also a time span of 150 years was used instead of only 100, so you can see also the convergence of the model with $r = 0.05$.

VI. Conclusion

The Experiments have shown above all that my model is plausible and leads to possible results. Maybe I have overestimated the population growth and the carrying capacity if I compare the results to them of Subalusky et al. and Shurin et al. (see [1], [5]). A reason for this could be that the estimated birth rate is too high. After all, female hippos can only have one calf per year under good conditions, as explained in [4]. But when the environment changes to the bad, which decreases the availability of resources and makes deadly phenomena more probable, the conditions to get children are not good anymore.

With the approach to use a logistic model for the population growth and a simualtion for the carrying capacity I found a reasonable prediction of how the hippos in Colombia could evolve. Of course these are all just predictions and mathematical models, nobody can know what will happen in the next few years, let alone in the next 100-150 years. Maybe next year the government just removes hippos completely from the Magdalena river, and then every model is useless. But it's still fascinating that you can make these approaches about an invasive species that started with only four species, I wonder if Pablo Escobar would have thought that?

References

[1] A. L. Subalusky, E. P. Anderson, G. Jiménez, D. M. Post, D. E. Lopez, S. García-R., L. J. Nova León, J. F. Reátiga Parrish, A. Rojas, S. Solari, and L. F. Jiménez-Segura, "Potential ecological and socio-economic effects of a novel megaherbivore introduction: the hippopotamus in Colombia," Oryx, vol. 55, no. 1, pp. 105–113, 2021.

[2] N. Castelblanco-Martínez, R. Moreno-Arias, J. Velasco, J. Moreno-Bernal, S. Restrepo, E. Noguera-Urbano, M. Baptiste, L.M. García Loaiza, G. Jimenez, "A hippo in the room: Predicting the persistence and dispersion of an invasive mega-vertebrate in Colombia, South America.", Biological Conservation, 253, 108923, 10.1016/j.biocon.2020.108923, 2021.

[3] M. Rother, https://github.com/suomika/HippoPopulation, "HippoPopulation: Simulation of the hippo population in Columbia as an invasive species", 2022.

[4] San Diego Zoo Wildlife Alliance, https://ielc.libguides.com/sdzg/factsheets/hippopotamus, "Hippopotamus (Hippopotamus amphibius) & Pygmy Hippopotamus (Choerpsis liberiensis) Fact Sheet: Reproduction & Development", 2021

[5] J. Shurin, N. Aranguren-Riaño, D. Negro, D. Lopez, N. Jones, O. Laverde, A. Neu, A. Pedroza Ramos. "Ecosystem effects of the world's largest invasive animal", Ecology, 101, 10.1002/ecy.2991, 2020.