www.simiode.org
SIMIODE A SYSTEMIC INITIATIVE FOR MODELING INVESTIGATIONS & OPPORTUNITIES WITH DIFFERENTIAL EQUATIONS

STUDENT VERSION
Faking Gause

Brian Winkel
SIMIODE, Cornwall NY USA

STATEMENT

In his book, *The Struggle for Existence*[1] studied how best to combine strains of yeast to improve vodka production in the Soviet Union. Before studying yeast he studied strains of paramecium.

Gause presumed logistic growth for all his populations

$$y'(t) = by(t)\left(\frac{K - y(t)}{K}\right) \tag{1}$$

with initial conditions $y(0) = y_0$.

He studied species separately, attempting to find b and K, from observational data - we will do that here.

Often he noted his populations in units of numbers of organisms per cubic centimeter, and often he would start with just $y(0) = 2$ "critters."

He then went on to study mixed species (say Paramecium cadautum and Parameciam aurelia) or competing growth models. Gause called these equations "equations of the struggle for existence." How proletariat, how Soviet, how Stalinesque!!! Here are the equations in the manner he posed them:

$$y_1'(t) = b_1 y_1(t)\frac{K_1 - (y_1(t) + \alpha y_2(t))}{K_1} \qquad \text{and} \qquad y_2'(t) = b_2 y_2(t)\frac{K_2 - (y_2(t) + \beta y_1(t))}{K_2}\,, \tag{2}$$

where $y_1(t)$ and $y_2(t)$ are the population sizes for the respective species, often given in numbers per cubic centimeter, b_1 and b_2 are the species' growth rates,K_1 and K_2 are the species' carrying capacity. These four numbers were determined from observing and parameter estimation techniques. We will do this too, here.

Incidentally, Gause called b_1 and b_2 the "coefficients of geometric increase" and K_1 and K_2 "maximum volumes" (of the populations). The numbers α and β were called the "coefficients of the struggle for existence" by Gause. Quite often these numbers are reciprocals of each other. For example if$\alpha = 4$ that says each member of species y_2 will take up $\alpha = 4$ spaces of what is left of y_1's environment, $K_1 - y_1(t)$. $\alpha = 4$ means that $\beta = 1/\alpha = 1/4$ and this says that says each member of species y_1 will take up $\beta = 1/4$ spaces of what is left of y_2's environment, $K_2 - y_2(t)$.

Rewriting our equations as follows:

$$y_1'(t) = b_1 y_1(t) \left(\frac{K_1 - \alpha y_2(t) - y_1(t)}{K_1} \right) \qquad \text{and} \qquad y_2'(t) = b_2 y_2(t) \left(\frac{K_2 - \beta y_1(t) - y_2(t)}{K_2} \right)$$

shows the effect of α and β on the respective available space for growth in each species, i.e. $K_1 - y_1(t)$ and $K_2 - y_2(t)$. Specifically, the terms $\alpha y_2(t)$ and $\beta y_1(t)$ refer to just how much of the respective carrying capacities, K_2 and K_1, one species' numbers would have on the other with the coefficients α and β measuring that effect.

Gause sought to determine α and β from coexisting data, building on his knowledge of b_1 and b_2 and K_1 and K_2 which he determined from his studies of the populations living separately. We shall do that.

One approach Gause gave for "determining" α and β was the following:

"To find the latter let us solve this system of two equations with two unknown values in respect to α and β. We obtain:

$$\alpha = \frac{K_1 - \frac{y_1'(t) * K_1}{(b_1 * y_1(t)} - y_1(t)}{y_2(t)} \quad \text{and} \beta = (K_2 - \frac{y_2'(t) * K_2}{(b_2 * y_2(t)} - y_2(t))/y_1(t) \text{."}$$

Now when it comes to the single population equations $y'(t) = by(t)\frac{(K-y(t)}{K}$ with initial conditions $y(0) = y_0$ we can solve explicitly for $y(t)$ and then use a least squares method for estimating the parameters b and K. However, when it comes to the multiple species, competitive model

$$y_1'(t) = b_1 y_1(t) \left(\frac{K_1 - \alpha y_2(t) - y_1(t)}{K_1} \right) \qquad \text{and} \qquad y_2'(t) = b_2 y_2(t) \left(\frac{K_2 - \beta y_1(t) - y_2(t)}{K_2} \right)$$

we cannot, in fact, solve analytically for $y_1(t)$ and $y_2(t)$, so we might consider Gause's approach for determining α and β in the competing species model AFTER we determine the b and K parameters in the respective separate species modeling.

Finally, we come to your problem.

We are given the following data sets, data1, data2, and data12, in which we have population size as a function of time (in days) first in data1 y_1 vs. time - first number of each pair is time, second is population, then in data2 y_2 vs time - first number of each pair is time, second is population, and then in data12, y_1 and y_2 vs. time, - first number of each pair is time, second number is population y_1, and third number is populationy_2. This is NOT Gause's data which is quite "dirty." Rather we have constructed this pretty "clean" data set for your consideration in the hope that you can get more immediate success before modeling in the real and murky world!!!

1. Helpful information on estimating derivatives. There are several estimates for the derivative of a function. One may be more useful than another, one may encompass more information, one may be smother, etc. Consider several approaches. Do you see what each one is doing? Which one might be best to use in these analyses? Why? Draw a sketch to illustrate each - by hand sketches are in order - at some point $(x, f(x))$ on a typical function plot.

$$f'(x) \sim \frac{f(x+h)-f(x)}{h)}, f'(x) \sim \frac{f(x)-f(x-h)}{h},$$

$$f'(x) \sim \frac{\frac{f(x+h)-f(x)}{h} + \frac{f(x)-f(x-h)}{h}}{2} = \frac{f(x+h)-f(x-h)}{2h}$$

are three examples.

2. Using data set data1 for the single species where the pairs of data are (t - time - in days, population 1 size at time t) estimate the parametersb_1 and K_1 in the model

$$y_1'(t) = b_1 y_1(t) \left(\frac{K_1 - y_1(t)}{K_1} \right), \tag{3}$$

with initial conditions $y_1(0) = y_{10}$.

in two ways (a) using least squares strategy we have been employing on the solution to the ODE vs. data and (b) by estimating the slope $y_1'(t)$ at points of time in the data and fitting an appropriate function of $y_1(t)$ to match your derivative estimate, (c) and using a technique that Gause used by fitting a straight line between the variables $y_1'(t)/y_1(t)$ and $y_1(t)$ - be sure to explain the why? how? and what? of your approaches. Using parameter results from (a) and your choice of (b) or (c) solve the ODE model above and compare your solved models to the data. Which approach yields the best results?

3. Using data set data2 for the single species where the pairs of data are (t - time - in days, population 2 size at time t) estimate the parameters b_2 and K_2 in the model

$$y_2'(t) = b_2 y_2(t) \left(\frac{K_2 - y_2(t)}{K_2} \right), \tag{4}$$

with initial conditions $y_2(0) = y_{20}$, in two ways (a) using least squares strategy we have been employing on the solution to the ODE vs. data and (b) by estimating the slope $(y_2'(t))$ at points of time in the data and fitting an appropriate function of $y_2(t)$ to match your derivative estimate. Use parameter results from both (a) and (b) solve the ODE model above and compare your solved models to the data. Which approach yields the best results?

4. Using data set data12 for the multiple species where the triples of data are (t - time - in days, population 1 size at time t, population 2 size at time t) estimate the parameters α and β in the model

$$y_1'(t) = b_1 y_1(t) \left(\frac{K_1 - \alpha y_2(t) - y_1(t)}{K_1} \right) \quad \text{and} \quad y_2'(t) = b_2 y_2(t) \left(\frac{K_2 - \beta y_1(t) - y_2(t)}{K_2} \right)$$

with initial conditions $y_1(0) = y_{10}$ and$y_2(0) = y_{20}$.

using parameters b_1 and K_1 and b_2 andK_2 found in (1) and (2) above, by estimating the slopes, $y_1'(t)$ and $y_2'(t)$, at points of time in the data and using Gases approach (outlined above) for estimating α and β outlined above. Use all these estimated parameters to solve the ODE model above for multiple species and compare your solved models to the data.

5. Finally use your combined species population model to predict the future for both populations and compare to the "actual" data we offered.

REFERENCES

[1] Gause, G. F. 1971. *The Struggle for Existence.* New York: Dover Publications, Inc. First published in 1934 by The Williams & Wilkins Company.

APPENDIX - Data Sets

The data sets given below are also in an Excel spreadsheet 6-026-S-FakingGause-Excel-StudentVersion.xlsx found in the Modeling Scenario's Supporting Docs.

Data set data1

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$y_1(t)$	5	7	9	13	17	24	34	43	60	79	101	136	171	214	255	312
t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
$y_1(t)$	354	398	439	478	497	512	542	549	566	572	571	585	581	595	586	

Data set data2

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$y_1(t)$	6	8	10	13	17	21	29	37	53	63	89	114	139	173	219	273
t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
$y_1(t)$	343	448	540	581	680	816	931	1041	1036	1181	1171	1318	1368	1380	1453	

Data set data12

t	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$y_1(t)$	4	7	10	12	16	24	33	45	61	78	93	121	154	199	239	271
$y_2(t)$	5	8	11	13	18	21	30	37	49	57	75	86	109	122	144	151
t	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
$y_1(t)$	317	368	410	392	417	453	504	472	512	514	550	504	528	524	538	
$y_2(t)$	154	168	172	184	175	171	169	172	158	146	145	139	128	120	111	