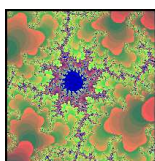


# AJUSTE DE DATOS Y MODELOS

---

## Ajuste de curvas y modelos biológicos

---



Una de las fases más importante en el proceso de modelización matemática es la de verificar si los resultados obtenidos al realizar la simulación del modelo se ajusta “lo mejor posible” a los datos experimentales. Naturalmente es muy importante definir con precisión este concepto de “buena aproximación” y si los errores cometidos al hacer la diferencia entre los valores aportados por el modelo y los datos reales están dentro de cierto rango de tolerancia. Tengamos en cuenta que al construir un modelo se ha realizado una gran simplificación de la realidad lo cual puede originar un modelo matemático cuya simulación se encuentre muy alejada de los datos obtenidos en el laboratorio.

### Ajuste de curvas a una colección de datos

Supongamos que los datos obtenidos en el laboratorio relacionados con cierto cultivo de bacteria son los que aparecen en la Tabla 1.

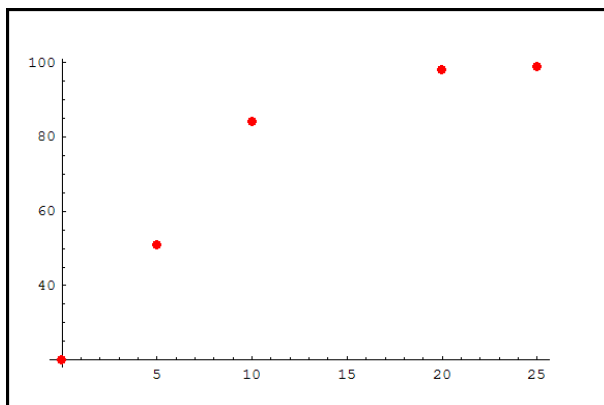
Tiempo $t$	0	5	10	15	20	25	30
Datos $d(t)$	20	51	84	-	98	99	99

**Tabla 1**

Como podemos apreciar, los datos se han obtenido de 5 en 5 horas pero con la salvedad de que desconocemos el valor correspondiente a las 15 horas. El problema

matemático de encontrar un valor aproximado del número de bacterias en ese momento se conoce con el nombre de interpolación. Por otro lado, si lo que deseamos es conocer el número de bacterias en el minuto 35 el problema recibe el nombre de extrapolación.

Para ajustar los datos a una curva lo primero que tenemos que hacer es representarlos.

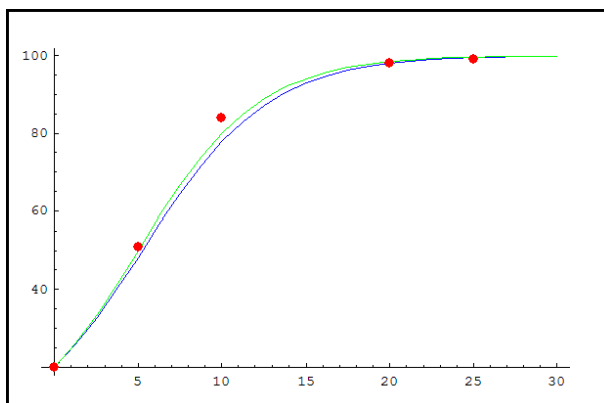


**Figura 1.-** Representación de la Tabla 1

En la gráfica se pueden apreciar dos etapas bien diferenciadas. Durante las primeras 15 horas la población de bacterias está creciendo muy rápidamente (exponencialmente) y después el crecimiento es muy lento, de tal manera que la población se estabiliza en torno a las 100 bacterias. Como sabemos, estamos en presencia del modelo logístico,

$$y(t) = \frac{k}{1 + A e^{-Bt}},$$

con una capacidad de carga  $k = 100$ .



**Figura 2**

Para encontrar los valores de los parámetros  $A$  y  $B$  utilizamos el hecho de que  $y(0) = 20$  y que  $y(20) = 98$ . Es inmediato comprobar que para estos valores,

$$y_1(t) = \frac{100}{1 + 4e^{-0.263906t}},$$

y la población a las 15 horas será aproximadamente de  $y(15) = 93$  bacterias. Por supuesto, también se pueden elegir otros valores, como por ejemplo  $y(0) = 20$  y  $y(10) = 84$ , en cuyo caso,

$$y_2(t) = \frac{100}{1 + 4e^{-0.277259t}}.$$

En la Figura 2 puede apreciarse el ajuste realizado, en color azul para el primer conjunto de valores  $y_1(t)$  y en verde en el segundo caso. La pregunta que surge de forma natural es ¿cual de estos ajustes es el mejor? Naturalmente la Figura 2 es de gran ayuda, pero lo interesante sería encontrar un parámetro que nos permita responder a la pregunta de una manera más objetiva. Para ello calculamos los errores  $e(t) = d(t) - y(t)$ . Como tenemos que trabajar con cantidades positivas, para que no se anulen mutuamente, tomamos valores absolutos y sumamos las cantidades (desviación total), o bien elevamos al cuadrado y sumamos (suma de los cuadrados del error).

Tiempo $t$	0	5	10	20	25	30
Datos $d(t)$	20	51	84	98	99	99
$y_1(t)$	20	48.33	77.78	98	99.46	99.85
$y_2(t)$	20	50	80	98.46	99.61	99.90
$e_1(t) = d(t) - y_1(t)$	0	2.67	6.22	0	-0.46	-0.85
$e_2(t) = d(t) - y_2(t)$	0	1	4	-0.46	-0.61	-0.90

**Tabla 1**

La desviación total para el primer ajuste es de 10.2 y para el segundo 6.27. La suma de los cuadrados del error en el primer caso es 46.75 y 18.3937 en el segundo. En ambos casos el segundo de los ajustes es el más adecuado. Por último, tenemos que poner de manifiesto que no siempre ocurre así. Dependiendo del criterio utilizado un ajuste puede ser mejor que el otro ya que una desviación total menor no implica que la suma de los cuadrados de los errores sea más pequeña, y recíprocamente.

### Ajuste por mínimos cuadrados

Suele ser bastante frecuente que en una primera etapa del crecimiento una población crezca según el modelo exponencial,

$$y(t) = Ae^{rt}, \quad A, r \in \mathbb{R}^+. \quad (1)$$

donde  $y(t)$  es el número de individuos de la población en el tiempo  $t$ .

Si en (1) tomamos logaritmos neperianos,

$$\ln(y(t)) = \ln(A) + rt, \quad A, r \in \mathbb{R}^+, \quad (2)$$

donde esta expresión, como puede apreciarse, es del tipo lineal, y por lo tanto, podemos realizar un ajuste lineal de los datos:  $\ln(y(t))$  utilizando el método de los mínimos cuadrados.

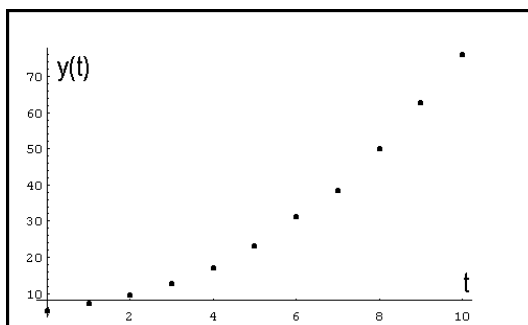
**En la tabla siguiente aparecen los datos de la población de Estados Unidos para cada década en el período 1800-1900:**

AÑO	t	y(t) (en millones)	$\ln(y(t))$
1800	0	5.3	1.666771
1810	1	7.2	1.97408
1820	2	9.6	2.26176
1830	3	12.9	2.55723
1840	4	17.1	2.83908
1850	5	23.2	3.14415
1860	6	31.4	3.44681
1870	7	38.6	3.65325
1880	8	50.2	3.91602
1890	9	62.9	4.14155
1900	10	76.2	4.33336

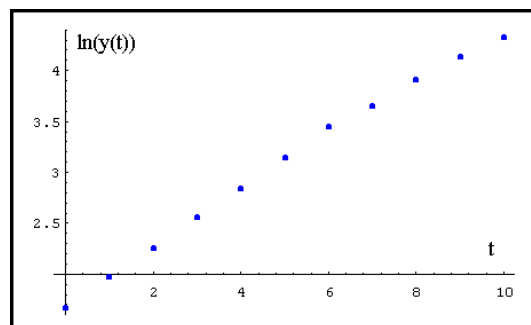
Podemos utilizar Mathematica® para la representación gráfica de estos datos.

```
A=ListPlot[{{0,5.3},{1,7.2},{2,9.6},{3,12.9},{4,17.1},{5,23.2},{6,31.4},
{7,38.6},{8,50.2},{9,62.9},{10,76.2}}]
```

```
B=ListPlot[{{0,Log[5.3]},{1,Log[7.2]},{2,Log[9.6]},{3,Log[12.9]},{4,Log[17.1]},
{5,Log[23.2]},{6,Log[31.4]},{7,Log[38.6]},{8,Log[50.2]},{9,Log[62.9]},
{10,Log[76.2]}], PlotStyle -> RGBColor[0,0,1]]
```



Representación gráfica  $(t, y(t))$



Representación gráfica  $(t, \ln(y(t)))$

- Empezaremos encontrando la recta  $y = at + b$  de ajuste de mínimos cuadrados para estos datos. Como es conocido, debemos buscar los parámetros  $a$  y  $b$  tales

que hagan mínimo la expresión,

$$\Phi(a, b) = \sum_{i=0}^{i=10} (\ln(y(t_i)) - at_i - b)^2 ,$$

lo que obliga a resolver el sistema

$$\begin{cases} \frac{\partial \Phi}{\partial a} = 0 \longrightarrow -2 \sum_{i=0}^{i=10} (\ln(y(t_i)) - at_i - b) t_i = 0 \\ \frac{\partial \Phi}{\partial b} = 0 \longrightarrow -2 \sum_{i=0}^{i=10} (\ln(y(t_i)) - at_i - b) = 0 . \end{cases}$$

Ahora bien, en lugar de resolver el sistema de dos ecuaciones con dos incógnitas anterior, es preferible hacer uso de algunos de los múltiples programas diseñados para tal fin, por ejemplo **Mathematica**<sup>®</sup>.

```
Fit[{{0,Log[5.3]},{1,Log[7.2]},{2,Log[9.6]},{3,Log[12.9]},{4,Log[17.1]},{5,Log[23.2]},{6,Log[31.4]},{7,Log[38.6]},{8,Log[50.2]},{9,Log[62.9]},{10,Log[76.2]}},{1,t},t]
```

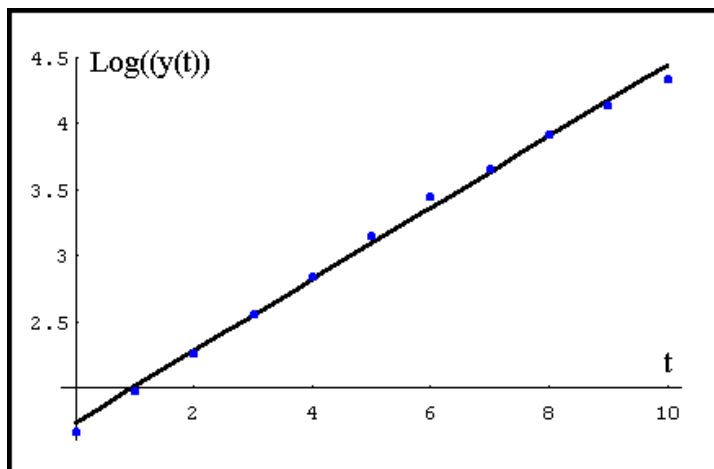
La respuesta que obtenemos es:

$$1.73224 + 0.270552t$$

Ahora representamos la recta anterior

```
ajuste=Plot[1.73224+0.270552t,{t,0,10}]
```

y finalmente superponemos los datos reales  $\ln(y(t))$  y la recta obtenida en el ajuste.



**Figura 3.-** Datos y recta de ajuste.

Para terminar encontramos los parámetros  $A$  y  $r$  de (1),

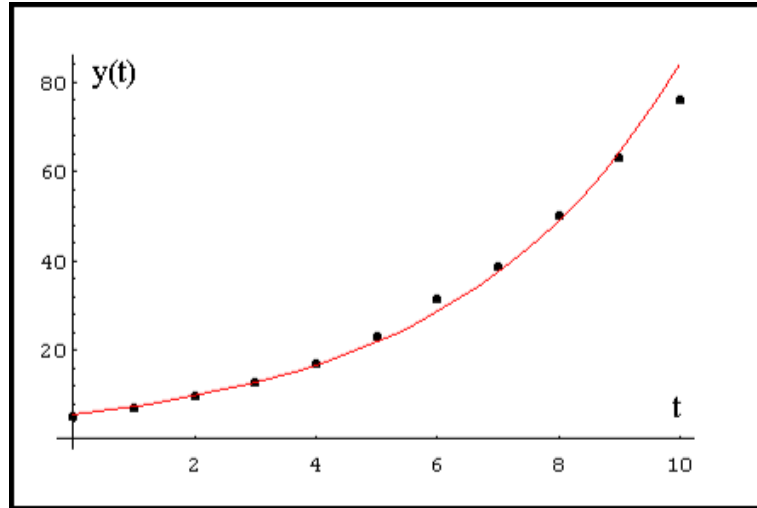
$$\ln(y(t)) = \ln(A) + r t = 1.73224 + 0.270552 t,$$

o bien,

$$A = e^{1.73224} \approx 5.65, \quad r \approx 0.27,$$

es decir

$$y(t) = Ae^{rt} = 5.65e^{0.27t}.$$



**Figura 4.-** Datos  $y(t)$  y función  $5.65 e^{0.27t}$ .

De esta manera, podemos tener una estimación de la población para el año 1910,

$$y(11) = 5.65 e^{0.27 \cdot 11} = 110.13 \text{ millones.}$$

- Un segundo método consiste en hacer uso de los datos

$$y(0) = 5.3, \quad y(5) = 23.3,$$

a fin de determinar los parámetros  $A$  y  $r$  del modelo (1). De esta manera

$$y(0) = 5.3 \longrightarrow 5.3 = Ae^0 \longrightarrow A = 5.3,$$

además

$$y(5) = 23.2 \longrightarrow 23.2 = 5.3 e^{5r} \longrightarrow r = \frac{1}{5} \ln \left( \frac{23.2}{5.3} \right) \approx 0.295289.$$

En consecuencia,

$$y(t) = 5.3 e^{0.295t}.$$

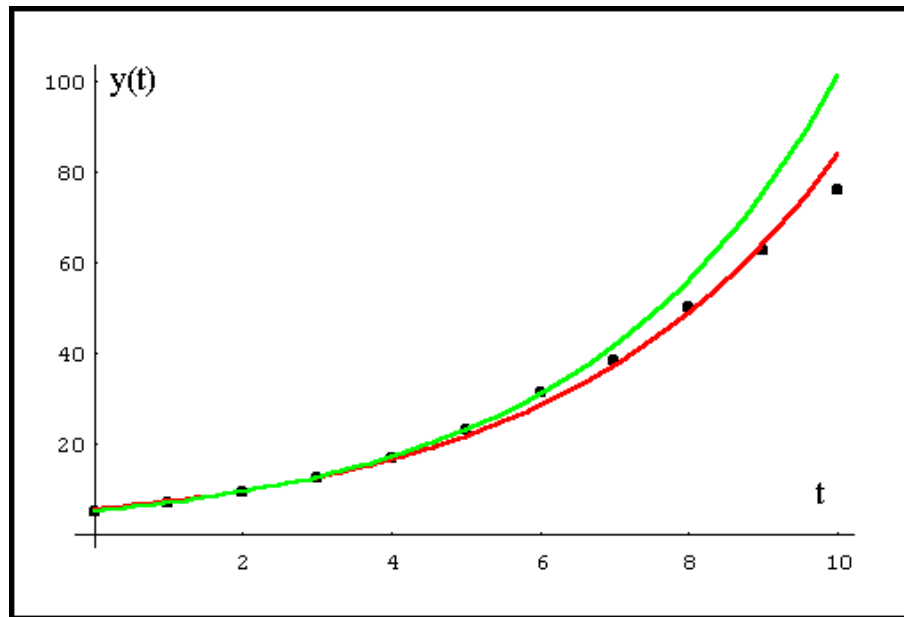


Figura 5.- Puntos =  $y(t)$ ; Rojo =  $5.65 e^{0.27 t}$ ; Verde =  $5.3 e^{0.295 t}$

**Ejercicio propuesto 1.** En la tabla siguiente se encuentran los datos de población para Estados Unidos de 1910 a 1980. Realizar un análisis similar al ejemplo anterior para estimar la población en el año 1990.

AÑO	$y(t)$ (en millones)
1910	92.2
1920	106.0
1930	123.2
1940	132.2
1950	151.3
1960	179.3
1970	203.3
1980	226.5

**Ejercicio propuesto 2.** Recientemente hay un gran debate sobre la importancia de preservar parte del terreno para mantener la biodiversidad. Muchos de los argumentos utilizados están basados en estudios realizados en islas del Caribe. En este ejercicio se utilizará la ley potencial, en la cual se supone que el número de animales  $N$  en el área  $A$  de la isla viene dado por

$$N = kA^a, \quad k, a \in \mathbb{R}^+. \quad (3)$$

Supongamos los siguientes datos:

	Área = A ( $Km^2$ )	Número = N
Redunda	1	3
Montserrat	33	10
Jamaica	4.41	38
Cuba	46.74	97

Hacer uso de la metodología utilizada en el Ejemplo 1.1 para completar la tabla:

	Área = A ( $Km^2$ )	Número = N
Saba	5	
Puerto Rico		40
Santa Cruz	80	
Española		88

---



# Bibliografía

- [1] ALLMAN E.S.; RHODES J.A. *Mathematical Models in Biology. An Introduction*. Cambridge University Press, (2004)
- [2] ARANA J. Seminario impartido en Palencia el 5 de abril de 2001.
- [3] BATSCHELET E.; BRAND L.; STEINER A. *On the Kinetics of lead in the human body*. Journal of Mathematical Biology 15-23, (1979).
- [4] BELTRAMI E. *Mathematical Models for Society and Biology*. Academic Press, San Diego, (2002).
- [5] BORRELLI R.; COLEMAN C.S. *Ecuaciones diferenciales. Una perspectiva de modelización*. Oxford University Express, Mexico, (2002).
- [6] BRITTON N.F. *Essential Mathematical Biology*. Springer-Verlag, London, (2003).
- [7] CAUGHLEY G. *Mortality patterns in mammals*. Ecology, **47**, 906-918, (1966).
- [8] CAUGHLEY G.; GUNN A. *Conservation Biology in Theory and Practice*. Blackwell Science, Massachusetts, 147-149, (1996).
- [9] CHUN-LIAN ZHU; YA JIA; QUAN LIU; XUAN ZHAN *A mesoscopic stochastic mechanism of cytosolic calcium oscillations*, Biophysical Chemistry, 125, 201-212, (2007).
- [10] CONEJERO C.M.; MARTÍNEZ FUNES J.; NAVAS J.; ESTEBAN F.J. *Aplicación de la Geometría fractal a la Ecología de Paisajes*. Departamentos de Matemática y Biología Experimental, Universidad de Jaén, (2005).
- [11] DE L HAZA I.; COBO-MOLINOS J.; GARRIDO-GARCÍA M. NAVAS J.; RUEDA P.; TORRES C.; CARÚZ A.; ESTEBAN F.J. *Fractal dimension of U373 astrocytoma cells in DMEM or RPMI cultures*. HarFA -Harmonic and Fractal Image Analysis 94-96, (2005). [http : //www.fch.vutbr.cz/lectures/imagesci/download\\_ejournal/221.Haza.pdf](http://www.fch.vutbr.cz/lectures/imagesci/download_ejournal/221.Haza.pdf)
- [12] ESTEBAN F.J.; NAVAS J.; JIMÉNEZ J.P.; QUESADA J.M.; GOÑI J.; VÉLEZ DE MENDIZÁBAL N.; VILLOSLADA P. *System Dynamics*

*as a tool in Stem Cells/Tissue Engineering*. Histology and Histopathology, Suppl. 1, 150-151, (2005).

- [13] ESTEBAN F.J.; NAVAS J.; QUESADA J.M. *Aplicaciones bioinformáticas en la docencia de Modelos Matemáticos en Biología*. II Reunión de la Red de Bioninformática de Andalucía. Baeza, ESPAÑA.
- [14] FOLLARI, R. *Interdisciplinaridad, espacio ideológico*. Simposio sobre Alternativas Universitarias, UAM-Azcapotzalco, México, (1980).
- [15] FONTDEVILA, A. M. *Introducción a la genética de poblaciones*. Síntesis, Madrid 112-117,(1999).
- [16] FORESTER, J.W. *Principles of Systems*. SMIT Press, Cambridge, M.A.(1968).
- [17] GALACH, M. *Dynamics of the tumor-immune system competition. The effect of time delay*, Int. J. Appl. Math. Comput. Sci.,Vol. 13, n<sup>o</sup> 3, 395-406, (2003).
- [18] GARCÍA CASTRO A.; GARRIDO PALOMERA E., ENDRINO M.I.; NAVAS J.; ESTEBAN F.J. *Dimensión Fractal de Imágenes MRI-T2 en la enfermedad de Alzheimer*. Departamentos de Matemáticas y Biología Experimental, Universidad de Jaén, (2005).
- [19] GÜNDÜZ *Ancient and Current Chaos Theories*, Interdisciplinary Description of Complex Systems, Vol. 4, n<sup>o</sup> 1, 1-18, (2006).
- [20] HAEFNER, J.W. *Modeling Biological Systems. Principles and Applications*. Chapman and Hall, New York, (1996).
- [21] HANNON B.; RUTH M. *Modeling Dynamic Biological Systems*. Systems Springer, New York 82-86, (1997).
- [22] HANNON B.; RUTH M. *Modeling Dynamic Biological Systems*. Systems Springer, New York 65-68, (1997).
- [23] HARTLOVE J.; SHAFFER D.; RAGAN S. *Glucose-Insuline Model*. The Maryland Virtual High School of Science and Mathematics, (2001).
- [24] HORN H.S. *The ecology of secondary succession*. Annual Review of Ecology and Systematics 5:25-37, (1974).
- [25] LEMAIRE V.; TOBIN F.L.; GRELLER L.D.; CHO C.R.; SUVA L.J. *Modeling the interactions between osteoblast and osteoclast activities in bone remodeling*, Journal of Theoretical Biology, 229, 293-309, (2004).
- [26] LI B.L. *Ecological Modelling*, 132, 33-50, (2000).

- [27] MAHAFFY J.M.; ZYSKIND J.W. *A model for the initiation of replication in Escherichia coli*. Journal Theory Biology, **140**, 453-477, (1989).
- [28] MANDELBROT B. *The Fractal Geometry of Nature*, Eds. W.H.Freeman and Company, New York, (1983).
- [29] MARTÍN-GARCÍA J. *Teoría y Ejercicios Prácticos de Dinámica de Sistemas*, Barcelona, (2003).
- [30] NOWAK M.A.; MAY R.M. *Virus Dynamics*, Oxford University Press, New York 16-26, (2000).
- [31] MARUSIC M., BAJZER Z., FREYER J.P.; VUC-PAVLOVIC, S. *Analysis of growth of multicellular tumour spheroids by mathematical models*. Cell, Prolif., **27**, 73-94, (1994).
- [32] NAVAS J.; QUESADA J.M.; GOÑI J.; VÉLEZ DE MENDIZÁBAL N.; VILLOSLADA P.; ESTEBAN F.J. *Glioma-Immune evasion: a system dynamics approach*. proceedings of the II International Conference on Computational Bioengineering, 699-710, Rodrigues et al. (Eds.), IST Press, ISBN: 972-8469-37-3, (2005).
- [33] ORTEGA H. *Un modelo logístico para el crecimiento tumoral en presencia de células asesinas*. Revista Mexicana de Ingeniería Biomédica, volumen **XX**, nº 3, 61-67, (1999).
- [34] RIOS, S. *Modelización*, Alianza Universidad, Madrid, (1995).
- [35] SERRANO C.; RODRÍGUEZ I.; MARTOS D.; NAVAS J.; ESTEBAN F.J. *Geometría Fractal de la Sustancia Blanca Cerebral durante el desarrollo de Esclerosis Múltiple*. Departamentos de Matemáticas y Biología Experimental, Universidad de Jaén, (2005).
- [36] VRIES G. DE; HILLEN T.; LEWIS M.; MÜLLER, J. SCHÖNFISH B. *A Course in Mathematical Biology. Quantitative Modeling with Mathematical and Computational Methods*. SIAM (Mathematical Modeling and Computation) (2006).
- [37] WARTON D.A.; GRAHAM W. *An approach to the teaching of host-parasite population modelling*. International Journal for Parasitology vol. 19; 451-45, (1989).
- [38] WERNER P.A.; CASWELL H. *Population growth rates and age vs. stage distribution models for teasel (Dipsacus sylvestris) Huds*. Ecology 58: 1103-1111, (1977).
- [39] YEARGERS E.K.; SHONKWILER R.W.; HEROLD J.V. *An Introduction to the Mathematics of Biology: with Computer Algebra models*, Birkhäuser, Berlin, (1996).