SORT 37 (1) January-June 2013, 95-106

Modelling "calçots" (*Allium cepa* L.) growth by the Gompertz function

Joan Simó, Marçal Plans, Francesc Casañas and Jose Sabaté*

Abstract

"Calçots" are the second-year resprouts of the "Ceba Blanca Tardana de Lleida" landrace of onions. The evolution of three "calçots" populations has been modeled to help farmers to plan the optimal time to harvest. Four different models that essentially differ in the type of distribution of the fitting Gompertz function parameters (lag time, maximum growth rate and the maximum attainable number of commercial size "calçots") have been tested. The model that considers a multinomial distribution of the fitting parameters showed the best agreement with the experimental data.

MSC: MSC2000 classification of the American Mathematical Society: 92D25

Keywords: "Calçots", onion, crop modeling, Gompertz, simulation.

1. Introduction

"Calçots" are the second-year resprouts of the "Ceba Blanca Tardana de Lleida" landrace of onions. In the production of "calçots", all the resprouts from an onion are harvested at the same time, when ≥ 50 % reach commercial size (1.7 cm–2.5 cm in diameter and 20 cm in length, as specified in the regulations for the "Calçot de Valls" (Protected Geographical Indication). Each onion yields between 1 and 20 "calçots", but their thickness is negatively correlated with the number of "calçots" per onion, so in the most productive onions many "calçots" never fulfill the commercial requirements for size. Production lasts from mid-November to the end of April, and a more or less constant release of marketable product is needed during this period. Farmers exploit genetic variability in earliness, using combinations of genotypes and/or sowing dates to adjust the production to consumer demand, but these combinations are haphazard and

* Dept. d'Enginyeria Agroalimentària i Biotecnologia, UPC, ESAB, Campus del Baix Llobregat, Carrer Esteve Terradas 8, 08860 Castelldefels, Barcelona. joan.simo@upc.edu
Received: March 2012
Accepted: October 2012

inefficient. Thus, it would be interesting to develop a methodology that would enable farmers to predict crop evolution and help them plan the optimal time to harvest.

The data recorded in our experimental crops suggest that the evolution of the number of commercial "calçots" through the growing season can be described by a sigmoidal pattern: an initial period where no commercial "calçots" are observed is followed by a second stage of rapid expansion and a final phase where the number of "calçots" asymptotically tends to a maximum value. Sigmoidal curves have been widely reported in biology, particularly in the growth of microorganisms under specific physical and chemical conditions (Rodriguez-Gonzalez et al., 2011; Zwietering et al., 1990), in the microorganisms inactivation (Gil et al., 2011), in the seasonal growth of fish (Singhi, 2011), and in the growth and development of field crops (Barker et al., 2010; Tei, Aikman and Scaife, 1996).

One of the most popular models to explain sigmoidal curves is based on the Gompertz function, which can be expressed in several forms, such as the following one, which involves three biologically meaningful parameters.

$$N = N_m e^{-e^{\frac{\mu_m e}{N_m}((\lambda - t)+1)}}, \tag{1}$$

where N and t are the measured number of individuals and time, respectively. N_m is the maximum N that can be reached at infinite time, μ_m is the maximum growth rate, and λ is the lag time.

The Gompertz model for microorganisms' growth has been used to predict the safety and shelf life of foods (Gil et al., 2011; Rodriguez-Gonzalez et al., 2011; Zwietering et al., 1990). The Gompertz function has been fitted to lettuce growth, although onions and red beets have expolinear growth (Tei et al., 1996). The Gompertz model has also been fitted to herbage mass and herbage accumulation (Barker et al., 2010); growth of tobacco leaves, stems, and roots (Ismail, Khamis and Jaafar, 2003); total biomass, leaf area index, number of plants per meter, and productivity of sugarcane (Simoes, Rocha and Lamparelli, 2005); and dry matter production and cob weight in maize cultivars (Ramachandra Prasad, Krishnamurthy and Kailasam, 1992).

We aim to i) use parameters based on the Gompertz function to discriminate among different populations and ii) use the same parameters to model the growth of "calçots" to enable the evolution of the crop to be predicted and the optimal harvest time to be planned.

2. Materials and methods

In a single location, we cultivated 100 onions from three populations ($P1$, $P2$, and $P3$) corresponding to three different genotypes. We scored the number of commercial "calçots" (N) in each viable plant every two weeks over a seven-month period.

First, the three Gompertz function parameters ($\lambda_{i,j}$, $\mu_{i,j}$ and $N_{m,i,j}$) were estimated for any plant i of the population j, by nonlinear least squares using the Gauss-Newton

algorithm. The goodness of fit was tested for each plant through the R^2, the root mean square error (RMSE) and the residuals distribution. All these data were used to compare the behaviour of the three populations and to develop a simulation process aimed at predicting the evolution of the three crops.

2.1. Comparison of the three populations

Multivariate ANOVA of the Gompertz parameters was performed following $\mathbf{Y} = \mathbf{XB} + \mathbf{E}$ where $\mathbf{Y}$ is the parameters matrix distributed as a $N_p \sim (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\mathbf{X}$ is the design matrix, $\mathbf{B}$ is the unknown parameters matrix (μ_{ip}, α_{ip}), and $\mathbf{E}$ is the error matrix. The distribution of error matrix was supposed $N_p \sim (0, \boldsymbol{\Sigma}\varepsilon)$, Wilks' statistic was used to test the significance of MANOVA. Comparisons of several multivariate means were analyzed using simultaneous confidence intervals (*CI*) with Bonferroni correction (Johnson and Witchern, 2007; Chung et al., 2009).

2.2. Simulations

Simulations were performed independently for each genotype. Viable plants of each genotype (n_j) were randomly split in 2 groups: the "calibration" and the "validation" groups, composed of n_c, j and $n_{v,j}$ plants respectively, being $n_{c,j} \approx 2n_{v,j}$ and $n_j = n_{c,j} + n_{v,j}$. Gompertz parameters achieved from the calibration set were used to generate a simulated set of $n_{v,j}$ plants.

The simulated set was generated according to four different models that essentially differ in the type of distribution of the fitting parameters of the Gompertz function. Model 1 only considers the average value of λ, μ_m and N_m. Then all the simulated plants evolved with the same rate for this model. Model 2 takes the average value of λ, μ_m and N_m and their variance-covariance matrix into account considering a normal multivariate distribution (Ripley, 1987) following the Equation (2):

$$f_{\mathbf{x}}(x_1, \ldots, x_k) = \frac{1}{(2\pi)^{\frac{k}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \Big(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \Big), \tag{2}$$

where k is the numbers of random variables, $\boldsymbol{\Sigma}$ is the Variance-Covariance matrix between variables and $\boldsymbol{\mu}$ is the mean vector of these variables. Model 3 is similar to Model 2 but uses the transformed parameters obtained from the Box-Cox method (Box and Cox, 1964; Ripley, 1987). Model 4 considers a univariate Weibull distribution of λ, μ_m and N_m independently (Johnson et al., 1994). The probability density function of a random variable x is described by Equation (3):

$$f(x; k, \lambda) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} e^{-(\frac{x}{\lambda})^k} & \text{if } x \geq 0, \\ 0 & \text{if } x > 0, \end{cases} \tag{3}$$

where k is the shape parameter and λ is the scale parameter. Then, the shape and scale parameters were obtained by maximum likelihood estimation and were subsequently used to generate the simulated set (Johnson et al., 1994).

For each model and each genotype, the simulation was repeated 100 times from the first step (i.e., from the random selection of the calibration and validation sets). The suitability of the simulations for predicting the evolution of the crop was evaluated by comparing each simulated set with its corresponding validation set. Comparisons were carried out in three ways: first, by applying the chi-square test on the total number of commercial "calçots" of any population, N_t along the 14 scoring dates, second with a parametric model survival analysis in which the target success was the time (t_X) when a plant produces a given fraction (X) of "calçots" meeting the commercial specifications, and third using a one-way ANOVA performed on the maximum number of commercial "calçots", N_m in which scores at the latest time (count 14) were taken for the N_m of validation plants.

All calculations were done with the R-program (`www.R-project.org`), using packages `agricolae` (Mendiburu, 2010), `doBy` (Højsgaard and Halekoh, 2011), `fitdistrplus` (Delignette-Muller et al., 2010), `CAR` (Fox and Weisberg, 2011), `MASS` (Venables and Ripley, 2002) and `survival` (Therneau and Lumley, 2011).

3. Results

The evolution of commercial "calçots" number for a typical plant of any population is shown in Figure 1. The experimental vs. fitted values of $N_{i,j}$ for each plant with the modified Gompertz equation yielded average R^2 values of 0.901, 0.915, and 0.906, with their corresponding standard deviations of 0.078, 0.051, and 0.066 for the populations $P1$, $P2$ and $P3$, respectively. The average values of RMSE were 0.41, 0.39 and 0.22, which can be considered low compared with the number of "calçots" expected to

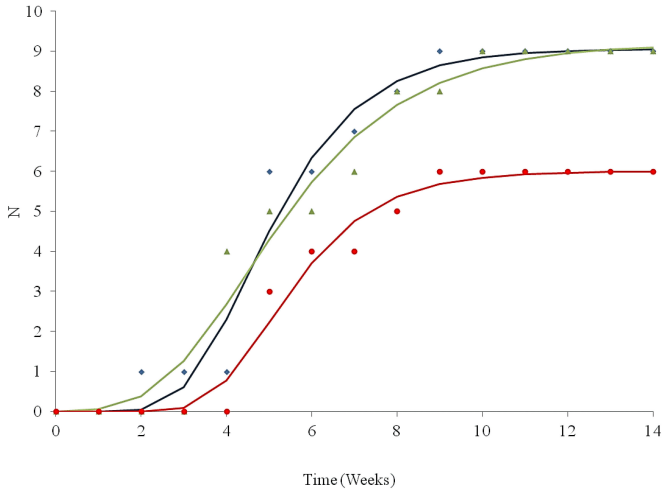

Figure 1: *Evolution of commercial "calçots" number for some typical plants:* ◇ *P1,* △ *P2 and* ○ *P3.*

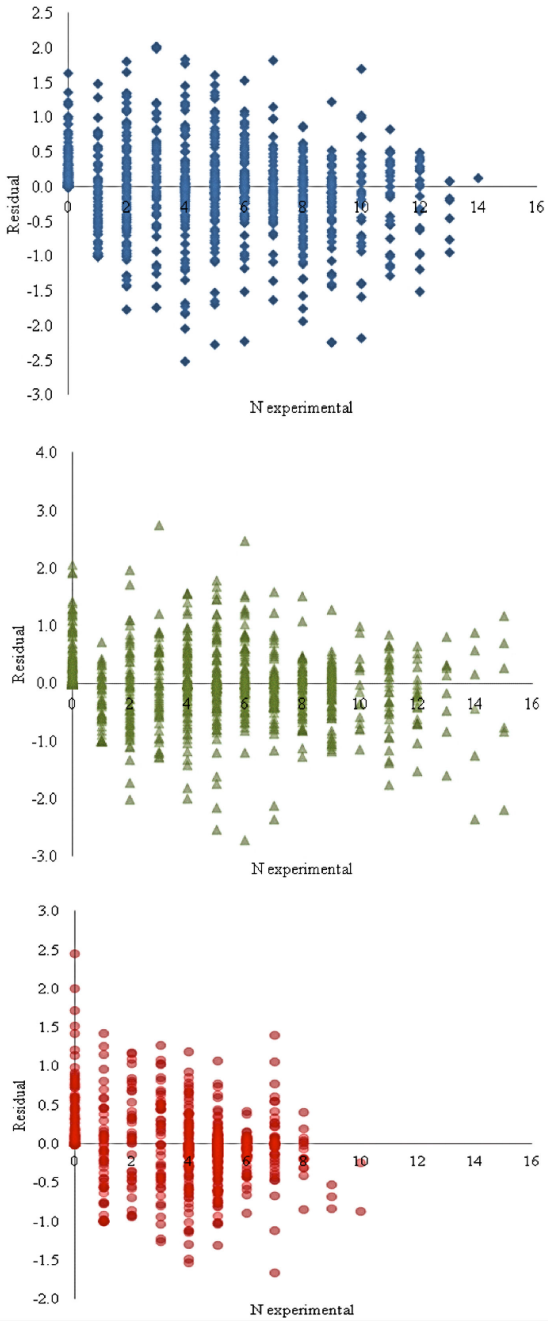

Figure 2: *Residuals vs experimental number of commercial "calçots". a)* ◇ *P1, b)* △ *P2 and c)* ○ *P3.*

harvest, 6 to 9. The residuals for of all the plants of any population, plotted in Figure 2, are nearly symmetrically distributed around the X axis for all the values of N. The only exception to this trend is observed for $N = 0$, as the predicted value of N should always

be positive according to Equation (1). All these data together support the suitability of the fitting.

3.1. Comparison of the three populations

First of all, the Shapiro-Wilks test shows that the set of estimates values of λ, μ_m and N_m of the individual plants for each population do not mainly follow a normal distribution, so they were transformed using the Box-Cox method and checked again (Box and Cox, 1964; Royston, 1982). Although conversion by the Box-Cox method resulted in transformed λ and μ_m values with a normal distribution, the transformed N_m ($P = 0.033$) still does not pass the test (Table 1). The Box-Cox parameter values for λ, μ_m and N_m were 0.7, -0.2, and -0.4, respectively. The p-values using Wilks tests were < 0.001 for both raw and transformed parameters. MANOVA with Bonferroni simultaneous *CI* test for multiple means comparisons (Table 2) indicates that the λ_2 mean value is smaller than λ_1 and λ_3 and μ_{m3} is smaller than μ_{m1} and μ_{m2} for both raw and transformed data. When the raw parameters are compared, the N_{m2} mean value is larger than N_{m1} and N_{m3}; however, when the transformed parameters are compared the N_m mean values of the three populations show significant differences.

3.2. Simulations

Mean parameters values and their coefficients of variation for simulation and variation sets can be seen in Table 3. Although a rigorous comparison was not performed, it can be seen that, generally, mean parameter values for the simulation sets are close to those corresponding to the validation sets. Further, the coefficients of variation indicate that fitting parameters are more scattered for simulation sets than for validation samples.

Table 1: *P-values for Shapiro-Wilks test of normality.*

	Raw parameters	Transformed parameters
λ	6.4×10^{-3}	0.215
μ_m	6.89×10^{-16}	0.146
N_m	2.75×10^{-13}	0.033

Table 2: *Comparison of parameters mean values with simultaneous CI for treatments' difference using Bonferroni correction in multiple comparison.*

	λ		μ_m		N_m	
	Raw	Transformed	Raw	Transformed	Raw	Transformed
$P1$	6.25 a	6.09 a	2.78 a	2.07 a	7.89 b	7.19 b
$P2$	4.84 b	4.66 b	2.51 a	1.93 a	9.40 a	8.25 a
$P3$	5.99 a	5.75 a	2.21 b	1.74 b	6.91 b	6.40 c

Table 3: *Mean values and coefficients of variation (CV) for simulation and validation sets.*

Set	Factor	Population	Model 1		Model 2		Model 3		Model 4	
			Mean	CV (%)	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)
Simulation	λ	$P1$	6.51	9.37	6.48	8.90	6.43	8.58	6.42	9.92
		$P2$	5.13	9.17	5.17	8.71	4.69	10.56	4.78	10.17
		$P3$	6.21	7.22	6.28	8.48	5.90	10.87	5.91	11.37
	μ_m	$P1$	3.20	14.75	3.16	13.99	2.89	20.56	2.90	14.06
		$P2$	2.69	10.56	2.69	10.17	2.89	19.27	2.60	13.18
		$P3$	2.37	12.53	2.46	12.93	2.25	16.54	2.24	14.19
	N_m	$P1$	8.48	7.92	8.44	6.93	6.94	5.57	8.42	7.77
		$P2$	10.19	7.14	9.99	7.76	7.35	5.68	10.32	8.00
		$P3$	7.47	6.25	7.52	6.27	6.53	5.83	7.53	7.04
Validation	λ	$P1$	6.25	7.28	6.21	8.23	6.18	6.14	6.18	7.06
		$P2$	4.83	6.40	4.80	6.88	4.88	7.38	4.80	7.51
		$P3$	5.96	8.33	5.98	6.94	6.05	7.20	5.95	7.75
	μ_m	$P1$	2.80	13.76	2.76	13.33	2.79	13.97	2.72	13.76
		$P2$	2.55	10.97	2.49	10.66	2.55	11.12	2.48	11.48
		$P3$	2.24	11.79	2.21	11.64	2.20	12.82	2.21	10.68
	N_m	$P1$	7.87	6.79	7.89	5.69	7.97	6.73	7.95	5.89
		$P2$	9.38	6.55	9.52	6.60	9.37	6.75	9.39	7.68
		$P3$	6.94	6.34	6.84	5.39	6.86	6.01	6.90	6.15

That can be understood because fitting parameters of validation set are estimated in one step. Nevertheless, the fitting parameters of simulations sets are estimated after two previous processes, the parameters estimation of calibration samples and the generation of a validation set with a limited number of samples. Additionally, for simulation sets, the parameters of Model 3 tend to be slightly lower than those of other models. Although the mean parameter values of the validation sets will not be used in future computations, comparing them to the corresponding simulation sets gives a first rough view of the goodness of the simulation.

Table 4: *Percentage of simulations that pass the chi-square test without reaching significance.*

	Model 1	Model 2	Model 3	Model 4
$P1$	0	28	32	22
$P2$	0	29	23	16
$P3$	0	27	27	18
All	0	5	10	10

First, the chi-square test was performed on the total number of commercial "calçots" in a given population. The percentage of simulations that accomplish the chi-square without significance ($p \leq 0.01$) is very low: about 20% for Model 4 and about 30% for Models 2 and 3 (Table 4). No simulations of Model 1 pass the test, which indicates that a

given distribution of λ, μ_m and N_m, and not only their mean values, has to be considered in order to make reasonable predictions of the evolution of the crop.

To analyze the reason for the failures of the chi-square test, the residuals between the predicted and experimental values of N vs. time were plotted (Figure 3). As expected, Model 1 gave the highest residual values throughout the growing season. Models 2 and 3 show a similar tendency for any population and two smaller peaks are observed at counts 4 and 9. Model 4 follows a particular trend for any population, peaks are observed at count 4 for population 1 but peaks at counts 4 and 8 appear for population 3.

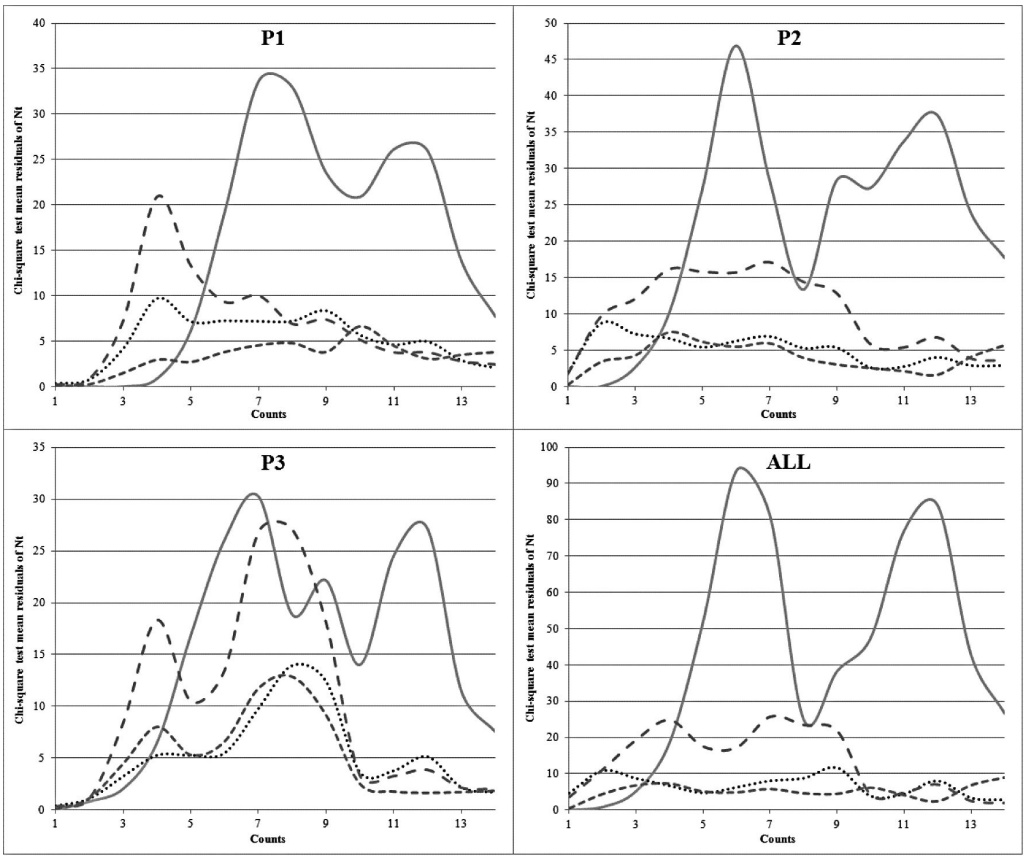

Figure 3: *Chi-square test mean residuals of N_t between validation and simulation sets for 100 simulations.*

As a second method to check the suitability of the simulations, a survival analysis was performed at the four times when 25%, 50%, 75%, and 90% of the "calçots" of one plant achieved the commercial size. In agreement with the results of the chi-square test, Model 1 fails for the three populations at all times. The other models behave differently depending on the percentage considered. When 25% of the "calçots" of a plant attained the commercial size (t_{25}), the number of successful simulations was visibly lower than

those corresponding to t_{50}, t_{75}, and t_{90}, for which more than 90% of the simulations were successful (Table 5).

Table 5: *Percentage of simulations that fulfill the survival analysis without reaching significance ($p \leq 0.01$).*

	t_{25}				t_{50}				t_{75}				t_{90}			
	*P*1	*P*2	*P*3	ALL	*P*1	*P*2	*P*3	ALL	*P*1	*P*2	*P*3	ALL	*P*1	*P*2	*P*3	ALL
Model 1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Model 2	54	86	69	92	90	94	96	98	100	95	98	100	83	97	94	98
Model 3	73	86	87	95	99	94	97	100	99	97	97	99	99	98	100	100
Model 4	70	84	74	96	94	93	97	98	96	98	98	100	76	93	93	99

The third way to evaluate the simulation was a one-way ANOVA performed on the maximum number of commercial "calçots", N_m. Again, the results showed that model 1 did not work (Table 6). Models 2 and 4 lead to more than 90% of simulations with the ANOVA test non-significant, whereas the suitability of Model 3 is clearly lower and varies greatly depending on the population.

Table 6: *Percentage of simulations that fulfill the ANOVA for N_m.*

	Model 1	Model 2	Model 3	Model 4
*P*1	4	99	32	100
*P*2	55	100	86	98
*P*3	17	94	65	99
All	55	100	93	97

A global comparison of the different models can be seen in Figure 4, where the total number of commercial "calçots" of any population, N, is represented. Points for the hundred simulations of any count are included. As stated above, the points of Model 1 are farthest from the target line, where experimental and calculated values of N would match. Models 2 and 4 tend to overestimate the production of "calçots". For the last stages of the culture, when N_t approaches its maximum value, the points of Model 3 move away from the bisector, in agreement with the ANOVA test for N_m. Globally, the best predictions were achieved by Model 2.

4. Conclusions

The modified Gompertz function allowed us to compare several populations with different genotypes throughout the growing season instead of making the comparison at peak times when scoring is carried out.

The multinomial distribution of fitting parameters of the Gompertz function used in Model 2 was the best distribution to model the growth of "calçots", predict the evolution of the crop, and decide the optimal harvest time.

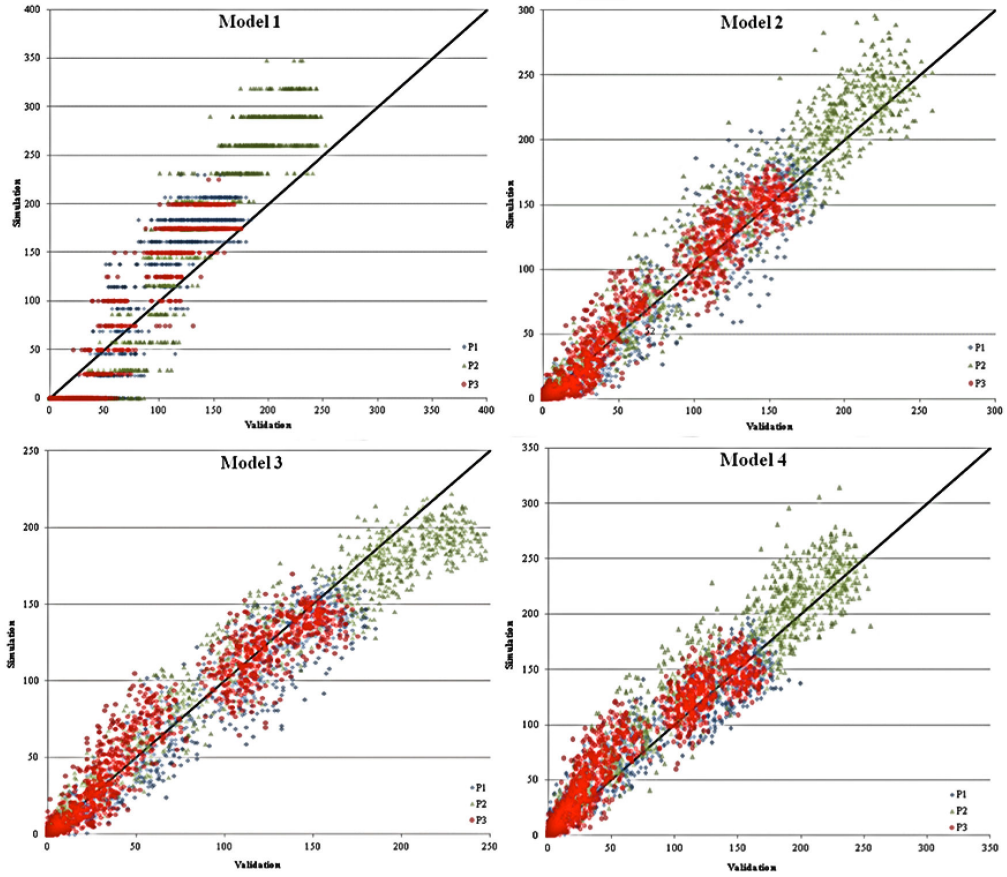

Figure 4: *Simulation vs. validation of total number of commercial "calçots". Symbols:* ◇ *P1,* △ *P2 and* ○ *P3.*

Model 1 yielded the worst results for all the tests used. Thus, models that consider a given distribution of the Gompertz fitting parameters (λ, μ_m and N_m) are much more suitable to explain the growth of "calçots" than those that consider only mean values.

Studies of the effects of the environment and genotypes on crop growth are needed to understand the different behaviour of each population so that better models can be constructed for the entire growing season.

References

Barker, D. J., Ferraro, F. P., Nave, R. L., Sulc, R. M., Lopes, F. and Albrecht, K. A. (2010). Analysis of herbage mass and herbage accumulation rate using Gompertz equations. *Agronomy Journal*, 102, 849–857.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26, 211–252.

Chung, C., Du, H., Su, Y. and Zhang, X. (2009). *The R project for Comparisons of Several Multivariate Means*. In http://www.divms.uiowa.edu/ kcowles/s166_2009/ ChungS166projectreport.pdf (September 2012).

Delignette-Muller, M. L., Pouillot, R., Denis J.-B. and Dutang C. (2010). fitdistrplus: help to fit of a parametric distribution to non-censored or censored data. http://www.R-project.org

Fox, J., and Weisberg, S. (2011). *An R Companion to Applied Regression*. *SAGE Publications*, Thousand Oaks, Calif.

Gil, M. M., Miller, F. A., Brandao, T. R. S. and Silva, C. L. M. (2011). On the use of the Gompertz model to predict microbial thermal inactivation under isothermal and non-isothermal conditions. *Food Engineering Reviews*, 3, 17–25.

Højsgaard, S. and Halekoh, U. (2011). doBy: groupwise summary statistics, general linear contrasts, LSMEANS (least-squares-means), and other utilities. http://www.R-project.org

Ismail, Z., Khamis, A. and Jaafar, Y. (2003). Fitting nonlinear Gompertz curve to tobacco growth data. *Pakistan Journal of Agronomy*, 2, 223–236.

Johnson, R. A. and Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis (6th Ed.)*. Englewood Cliffs, New Jersey: Prentice-Hall. Chapter 6.

Johnson, N. L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*, volume 1, chapter 21. Wiley, New York.

Mendiburu, F. (2010). agricolæ: statistical procedures for agricultural research. http://www.R-project.org

Royston, P. (1982). An extension of Shapiro and Wilk's W test for normality to large samples. *Applied Statistics*, 31, 115–124.

Ramachandra Prasad, T. V., Krishnamurthy, K. and Kailasam, C. (1992). Functional crop and Cob growth models of maize (*Zea mays L.*) cultivars. *Journal of Agronomy and Crop Science*, 168, 208–212.

Ripley, B. D. (1987). *Stochastic Simulation.* Wiley. Page 98.

Rodriguez-Gonzalez, O., Walkling-Ribeiro, M., Jayaram, S. and Griffiths, M. W. (2011). Cross-protective effects of temperature, pH, and osmotic and starvation stresses in *Escherichia coli* O157:H7 subjected to pulsed electric fields in milk. *International Dairy Journal*, 21, 953–962.

Simoes, M. d. S., Rocha, J. V. and Lamparelli, R. A. C. (2005). Growth indices and productivity in sugarcane. *Scientia Agricola*, 62, 23–30.

Singhi, N. O., Paul, A. K., Singh, N. G., Singh, P. and Alam, W. (2011). Modeling seasonal growth of fish using modified Gompertz model with sine wave function. *Indian Journal of Animal Sciences*, 81(6), 648–650.

Tei, F., Aikman, D. P. and Scaife, A. (1996). Growth of lettuce, onion and red beet. 2. Growth modelling. *Annals of Botany*, 78, 645–652.

Therneau, T. and Lumley, T. (2011). survival: survival analysis, including penalised likelihood. http://www.R-project.org

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S.* Springer, New York.

Zwietering, M. H., Jongenburger, I., Rombouts, F. M. and Vantriet, K. (1990). Modeling of the bacterial-growth curve. *Applied and Environmental Microbiology*, 56, 1875–1881.