



HAL
open science

Methodology for choosing a model for wheat kernel growth

N. Robert, S. Huet, Christelle Hennequet, A. Bouvier

► **To cite this version:**

N. Robert, S. Huet, Christelle Hennequet, A. Bouvier. Methodology for choosing a model for wheat kernel growth. *Agronomie*, 1999, 19 (5), pp.405-417. <hal-02694006>

HAL Id: hal-02694006

<https://hal.inrae.fr/hal-02694006v1>

Submitted on 1 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Methodology for choosing a model for wheat kernel growth

Nathalie Robert^{a*}, Sylvie Huet^b, Christelle Hennequet^b, Annie Bouvier^b

^a Unité d'amélioration des plantes, Inra, domaine de Crouelle, 234, avenue du Brézat, 63039 Clermont-Ferrand cedex 2, France

^b Unité de biométrie, Inra, domaine de Vilvert 78352 Jouy-en-Josas cedex, France

(Received 5 December 1998, accepted 30 March 1999)

Abstract – A methodology for choosing a model describing the wheat kernel growth of 16 wheat cultivars, grown in nine environments, is presented. Indeed, it was a preliminary and essential step before comparing the cultivars for their rates and durations of grain filling. Four current growth functions, i.e. logistic with three parameters assuming that the lower asymptote equals 0, logistic with four parameters estimating the lower asymptote, Weibull and Gompertz functions, were compared. In a first step the parameters of the curve were estimated assuming that the variance of the observed kernel weight was constant. Examining the graphs of absolute values of standardized residuals against predicted values of kernel weight highlighted that the variance of errors in the regression model was not constant and suggested modelling the variance using a power function. In a second step, modelling of the variance was added to the model. The models were compared using the likelihood ratio tests, the graphs of residuals, Akaike's criterion and the biological meaning of the estimated final kernel weight. Significant likelihood ratio tests indicated that, for all functions except Weibull, the assumption of homogeneous variances had to be rejected; thus, it was necessary to model the variance. Comparisons of the four functions using Akaike's criterion led to keeping the logistic function with four parameters and modelling of the variance. Comparing the estimates of the final kernel weight (95 % of the upper asymptote) obtained with this model with observed kernel weights revealed that some of the estimates were not realistic from a biological point of view. Finally, we chose to model the kernel growth using the logistic curve with three parameters for modelling the growth curve and the power function for modelling the heterogeneity of variance. In addition, a modification of the sampling protocol is also presented. (© Inra/Elsevier, Paris.)

growth function / non linear model / variance modelling / wheat kernel

Résumé – Méthodologie pour choisir un modèle de croissance du grain de blé. Une méthodologie pour choisir un modèle décrivant la croissance du grain de 16 variétés de blé, cultivées dans 9 milieux, est présentée. En effet, c'est une étape préliminaire et indispensable pour comparer les variétés pour leurs vitesses et durées de remplissage du grain.

Communicated by Max Rives (Villeneuve-lès-Avignon, France)

* Correspondence and reprints
robert@valmont.clermont.inra.fr

Quatre fonctions de croissance classiques, i. e. la logistique à trois paramètres supposant l'asymptote inférieure égale à 0, la logistique à quatre paramètres estimant l'asymptote inférieure, les fonctions de Weibull et Gompertz, ont été comparées. Lors d'une première étape, les paramètres de la courbe sont estimés avec l'hypothèse d'une variance des poids de grain observés constante. L'examen des graphiques des valeurs absolues des résidus centrés réduits contre les valeurs prédites des poids de grain indique que la variance des erreurs n'est pas constante et suggère de modéliser la variance à l'aide d'une fonction puissance. Lors d'une seconde étape, la modélisation de la variance est introduite dans le modèle. Les modèles sont comparés à l'aide du test des rapports de vraisemblance, des graphiques des résidus, du critère d'Akaike et de la signification biologique des estimations du poids de grain final. Les tests de rapport de vraisemblance significatifs indiquent que pour toutes les fonctions excepté la fonction Weibull, l'hypothèse d'homogénéité des variances doit être rejetée; il est donc nécessaire de modéliser la variance. Les comparaisons des quatre fonctions avec le critère d'Akaike conduit à garder la fonction logistique à quatre paramètres et modélisation de la variance. La comparaison des estimations du poids final du grain (95% de l'asymptote supérieure) obtenues avec ce modèle aux valeurs observées de poids du grain révèlent que certaines estimations ne sont pas réalistes d'un point de vue biologique. Finalement, une courbe logistique à trois paramètres pour modéliser la croissance du grain, avec une fonction puissance pour modéliser l'hétérogénéité des variances est retenue. Une modification du protocole de prélèvements est aussi présentée. (© Inra/Elsevier, Paris.)

fonction de croissance / modèle non linéaire / modélisation de la variance / grain de blé

1. Introduction

The protein content of a grain is a trait involved in the end-use quality of wheat grain [11, 20]. It is calculated as the ratio of the total quantity of protein in the grain to the total dry matter quantity in the grain and results from two filling phenomena: dry matter and nitrogen filling. In order to characterize the effect of genetic and environmental variations on the protein content, the genetic, environment and genetic \times environment interaction effects for these filling phenomena have to be studied. Each of these fillings corresponds to a growth which can be described by a curve and characterized by a rate and a duration. We must then ask which growth function to choose, as the calculation of the filling rate and duration depends on this. We present a process for choosing a function describing the growth of a kernel and corresponding to dry matter accumulation.

A method often used to calculate the filling rate as a function of time, expressed in days or growing degree-days after anthesis, consists in fitting a linear regression to a part of the data observed, the slope of the line corresponding to the filling rate [9, 12, 15, 23, 26, 27]. The duration can be estimat-

ed from the day of anthesis to the day corresponding to the intersection of the regression line with the horizontal line fitted through the last samplings and considered to describe the maximal kernel weight. It can be also calculated as the final kernel weight divided by the filling rate [9, 23, 27]. It then corresponds to the time interval defined by the intersection of the abscissae axis with the regression line and by the intersection of this line with the horizontal line describing the final kernel weight [23, 29]. The time interval corresponding to the lag period [23] is then omitted in the calculation. A visual examination on the curve has also been proposed [26]. The difficulty of this method is to define the linear period for fitting the regression [18]. In order to limit the subjective choice of the linear period, regressions can be fitted successively and the interval giving the best coefficient of determination can be kept [15, 23].

Fitting a quadratic [4, 18] or cubic polynomial [1, 10] is a way of circumventing this difficulty. For some studies, either a quadratic or a cubic polynomial was fitted to each curve [2, 16, 17]. In all the studies mentioned, the duration was defined as the time when the first derivative of the polynomial is equal to 0; the maximal weight predicted by the function is reached at that time. In contrast, the

rate used to characterize curves varied according to the authors. It was generally calculated as the maximal kernel weight predicted by the function chosen, divided by the duration estimated and was called the mean rate [4, 10, 16, 17]. The instantaneous rate is equal to the first derivative of the function; the rate calculated at the time for which the second derivative of the polynomial is equal to 0, called the maximum rate, was then used to characterize the curves [15]. However, for Motzo et al. [16], the maximum rate was estimated by the slope of the linear regression providing the best fit, among all the linear regressions calculated. In Wardlaw and Moncur [28], a cubic polynomial was only used to describe the curve and the rate and duration were calculated by fitting a linear regression.

The logistic curve with a lower asymptote equal to 0 was also used to model kernel growth [5, 8, 9, 15]. Such a non linear function is fitted to all the values observed. The duration was defined as the time required to reach 95 % of the upper asymptote value and the maximum rate was calculated at the time for which the second derivative of the logistic function was equal to 0. In addition to the maximal rate, a mean rate was calculated as the sum of instantaneous rates estimated on a time interval

divided by the duration of this interval [5], for it was considered as more representative of grain filling as it took into account the total duration.

Other non linear functions, such as Weibull and Gompertz functions, are classically used to describe growth phenomena [19, 22]. In this paper, we compare the logistic, Weibull and Gompertz models in order to describe kernel growth and to estimate rate and duration of grain filling and final kernel weight. The final model was chosen using statistical criteria and biological references. The variation of the variance of the observations versus time was taken into account in the model.

2. Materials and methods

2.1. Field experiments

Sixteen bread wheat cultivars, Ami, Apollo, Arminda, Baroudeur, Camp-Rémy, Forby, Qualital, Récital, Renan, Rossini, Sidéral, Soissons, Talent, Thésée, Trémie, Viking, all registered and cultivated in France, were grown at Clermont-Ferrand in a randomized complete block design with two blocks, in nine environments. Each environment corresponded to the combination of one date of sowing, one nitrogen fertilization practice and one year (*table I*). Plots consisted of

Table I. Description and codes of the nine environments.

Environment	Date of sowing	Nitrogen fertilization
Environment 1: E1	19 October 1994	none
Environment 2: E2	19 October 1994	100 kg.ha ⁻¹ (at tillering) 90 kg.ha ⁻¹ (at heading)
Environment 3: E3	5 December 1994	none
Environment 4: E4	5 December 1994	100 kg.ha ⁻¹ (at tillering) 90 kg.ha ⁻¹ (at heading)
Environment 5: E5	3 November 1995	60 kg.ha ⁻¹ (at tillering)
Environment 6: E6	3 November 1995	60 kg.ha ⁻¹ (at tillering) 60 kg.ha ⁻¹ (at early shooting)
Environment 7: E7	3 November 1995	60 kg.ha ⁻¹ (at tillering) 60 kg.ha ⁻¹ (at early shooting) 50 kg.ha ⁻¹ (at early booting)
Environment 8: E8	21 December 1995	60 kg.ha ⁻¹ (at early shooting)
Environment 9: E9	21 December 1995	60 kg.ha ⁻¹ (at early shooting) 50 kg.ha ⁻¹ (at early booting)

Table II. Functions, maximum rate, duration and upper asymptote for the four growth functions.

Function	$P = f(t)$	Maximum rate	Duration	Upper asymptote
Logistic 3 (L3)	$f(t) = \frac{a}{1 + e^{(b-ct)}}$	$\frac{ac}{4}$	$\frac{1}{c} \left(b - \ln \frac{1}{19} \right)$	a
Logistic 4 (L4)	$f(t) = d + \frac{a}{1 + e^{(b-ct)}}$	$\frac{ac}{4}$	$\frac{1}{c} \left(b - \ln \left(\frac{0.05(a+d)}{0.95a - 0.05d} \right) \right)$	a + d
Weibull (W)	$f(t) = a \left(1 - e^{-bt^c} \right)$	$abc \left(\frac{c-1}{bc} \right)^{1-\frac{1}{c}} e^{-\left(1-\frac{1}{c}\right)}$	$\left(\frac{1}{b} \ln \frac{1}{0.05} \right)^{\frac{1}{c}}$	a
Gompertz (G)	$f(t) = a + be^{-e^{(d-ct)}}$	$\frac{bc}{e}$	$\frac{d - \ln \left(\ln \frac{b}{0.95b - 0.05a} \right)}{c}$	a + b

six rows spaced 20 cm apart, plot sizes were 4.5 m² in 1995 and 7 m² in 1996. On each plot, 100 stems were collected at anthesis and 10 days after anthesis, then 50 spikes were collected every 5 days till maturity. Spikes were oven-dried at 80 °C for 48 h. The mean dry weight of one kernel was obtained after weighing 200 kernels, except for the sampling at anthesis. Eight to 11 samplings were available per curve. As the study of grain filling was destructive, a second identical trial was planted in the same field to study yield, yield components and quality. Mean kernel weights at maturity were assessed in this trial.

2.2. Statistical methods

Four growth functions were compared:

- the logistic function with three parameters (L3), with a null lower asymptote;
- the logistic function with four parameters (L4), with a non null lower asymptote;
- the Weibull function (W);
- the Gompertz function (G).

Each growth function was characterized by its maximum rate R_m , equal to the first derivative of the function at the time which sets the second derivative to 0, and by its duration D , the time required to reach 95 % of the upper asymptote value. We defined the estimate of the final kernel weight as 95 % of the upper asymptote

value. Equations of the functions and calculations of the upper asymptote, rate and duration are presented in *table II*. Time was expressed in cumulative growing degree-days after anthesis. The sum of temperatures for a day was calculated as $T = (T_{min} + T_{max})/2 - T_b$, where T_{min} and T_{max} are the minimum and maximum temperatures of the day and T_b the base temperature, chosen to be equal to 0 [21].

The statistical model considered was a non linear model: the relationship between P , the kernel weight and t , the sum of temperatures (*figure 1*) was the sum of a systematic component $f(t, \theta)$ and a random error ϵ . The function, f , or growth function, depended on unknown parameters θ to be estimated. In each environment, for each curve corresponding to the combination of one variety by one block (16 cultivars \times 2 blocks) we wrote:

$$P_i = f(t_i, \theta) + \epsilon_i$$

where i was the sample indice, which varied from 1 to I , and I was the total number of samples per curve, and varied from 8 to 11.

In a first step, the observations were assumed to have a constant variance $\text{var}(\epsilon_i) = \sigma^2$, as usually found in the literature. In each environment, 32 vectors of parameters θ were estimated plus one parameter σ^2 . The models considered were noted L3, L4, W and G according to the growth function chosen. The graphical examination of the residuals led us to reject the assumption of variance homogeneity. In a second step, we took into account the lack of homogeneity, assuming that the

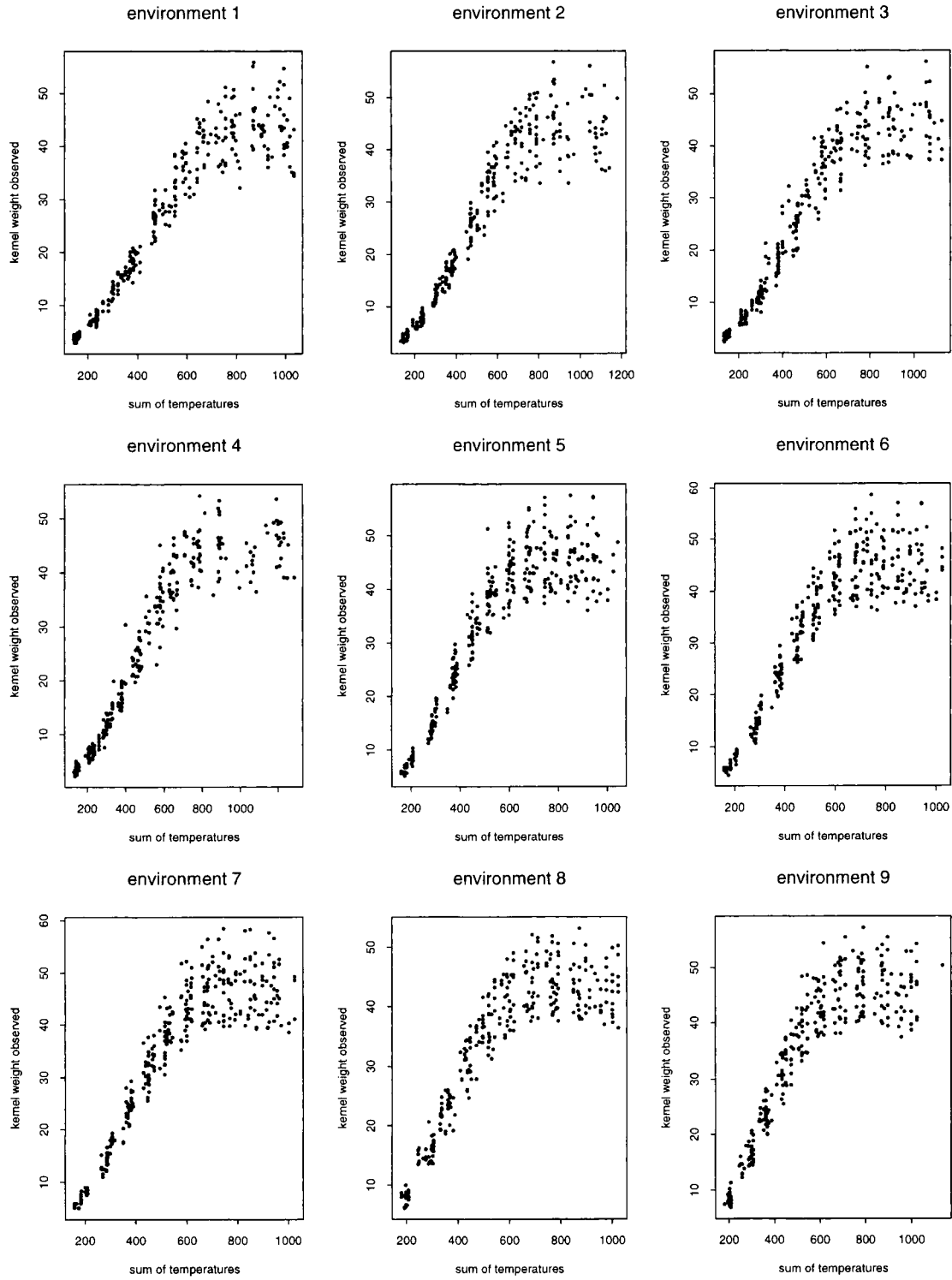


Figure 1. Kernel weight observed, P, versus sum of temperatures in each of the nine environments.

Table III. Likelihood ratio test for the hypothesis of variance homogeneity $H_0: \text{var}(\epsilon_i) = \sigma^2$ against the alternative A: $\text{var}(\epsilon_i) = \sigma^2 f(t_i, \theta)^\tau$, estimated value of τ and its standard error (in brackets) for model L3V.

Environment	L3V/L3	τ	L4V/L4	WV/W	GV/G
1	90.42 *	1.29 (0.0256)	157.64*	14.77*	21.30*
2	170.36*	1.78 (0.0257)	226.27*	3.84*	NCV
3	66.69*	1.14 (0.0261)	97.35*	1.70 ns	84.72*
4	133.27*	1.64 (0.0265)	219.02*	3.25 ns	258.51*
5	89.42*	2.36 (0.023)	NCV	3.25 ns	258.51*
6	166.41*	3.37 (0.0231)	NCV	0 ns	NCV
7	125.36*	2.83 (0.0229)	NCV	0.68 ns	NCV
8	72.85*	1.75 (0.0234)	137.13 *	11.24*	NCV
9	92.51*	2.009 (0.0233)	NCV	36.24	NCV

L3: logistic model with three parameters, L4: logistic with four parameters, W: Weibull model, G: Gompertz model.

L3V, L4V, WV, GV: same models with modelling of the variance; L4V2: logistic model with four parameters, modelling of the variance and parameter $\tau = 2$. NCV: the convergence of the estimating algorithm was not reached. No estimation available.

$\text{khi}_{0.95, 1df}^2 = 3.84$; ns: not significant; * significant at the 5 % level.

variance of an observation was proportional to a power of the kernel growth: $\text{var}(\epsilon_i) = \sigma^2 f(t_i, \theta)^\tau$, where τ and σ^2 were the parameters to be estimated. They were assumed identical for the 32 data sets belonging to the same environment. In this case, 32 vectors of parameters θ were estimated plus the two parameters σ^2 and τ . Modelling the variance for each growth function introduced four new models noted L3V, L4V, WV and GV. Once the growth function was chosen, the model assuming a constant error variance, L3, for example, was nested in model L3V. When the parameter τ was nul, L3 and L3V were identical. We compared L3 and L3V by testing the hypothesis $H_0: \tau = 0$ against the alternative A: $\tau \neq 0$. The parameters θ , σ^2 , τ were estimated by the maximum likelihood method, under the assumption of a Gaussian distribution of the errors. The tests were based on the statistics of the likelihood ratio test. The confidence intervals at 95 % of the functions of parameters R_m and D were calculated from the quantiles of the Gaussian law [13]. The comparison of the eight models considered was based on Akaike's criterion [25] which formalizes the principle of parsimony, making a compromise between likelihood, which has to be high, and the number of parameters. The higher the number of parameters, the better the fit since the discrepancy between the observed and fitted values is low, and thus the higher the likelihood. In contrast, the higher the number of parameters, the larger the variance of the estimator. The model which maximized Akaike's criterion was kept.

All calculations were made with the nls2 library [3] of SPLUS software [24].

3. Results

For each of the growth models L3, L4, W and G, the parameters were estimated for each environment and the graphs of the residuals were examined to detect possible deviations from the model. The graphs diagnosed variance heterogeneity revealed by the higher scatter of the standardized residuals for high fitted values (*figure 2A*). This result was coherent with the phenomenon studied and simply expressed that the variance of a growth phenomenon varies during the time course of the phenomenon and increases with the growth progress [6, 22]. The variance was then modelled with a power function defining models L3V, L4V, WV and GV. For each set of 32 curves, a likelihood ratio test for testing the hypothesis of variance homogeneity was performed.

3.1. Logistic models

The likelihood ratio test of the hypothesis that the model was L3 against the alternative that the

model was L3V (*table III*, column 1) led to the rejection of model L3. The examination of the graphs of standardized residuals versus fitted values of the growth function for model L3V (*figure 2B*) confirmed the improvement brought by modelling the variance: the distribution of the points was more uniform for model L3V than for model L3 (*figure 2B*) and the line joining the points after smoothing was quasi-horizontal. The graphs did not suggest any model misspecification. Comparison between L4 and L4V models was more critical as the numerical procedure to estimate the parameters for model L4V did not converge in four environments out of nine. If we restricted ourselves to the environments for which convergence was reached, results of the likelihood ratio test for model L4 against model L4V as well as the examination of residuals (*figure 2*) led to the rejection of model L4 with the same arguments as for model L3. All the environments for which convergence was not reached corresponded to the year 1996. Examining the sums of temperature highlighted that, in these environments, higher temperatures after the anthesis of all cultivars delayed the beginning of observations. Consequently, the lag period was removed from some curves, particularly those of late cultivars, as illustrated by cultivars Apollo, Forby, Arminda and Viking in *figure 3*. In consequence, estimation of a lower asymptote was not always possible.

To overcome these convergence problems, we strengthened the constraints on parameter τ . Model L4V assumed that the τ value was common to the 32 curves belonging to the same environment, but that τ could vary according to the environment. We defined a new model, L4V2, for which we imposed a fixed value of τ , common to the nine environments, as a new constraint. This parameter value was set as equal to 2, by referring to the τ values obtained in model L3V (*table III*, column 2). L4V2 can be neither compared to L4 nor to L3V with a test that can only compare embedded models, but Akaike's criterion can be used. Comparing Akaike's criterion for models L3V, L4V and L4V2 (*table IV*) led us to prefer L4V or L4V2 except for environments 3, 6 and 7, for which a value of 2 for τ may not have been the most relevant. At this

stage of the analysis, it was necessary to remember the aim of modelling. We wanted to estimate the rate and duration of grain filling as well as the final kernel weight. The definitions of these quantities as functions of the model parameters are given in *table II*. As experimental values were available for kernel weight, the biological meaning of the estimates of kernel weight could be assessed. The approximate values were compared to the kernel weight values observed on plots of the second experiment designed to study yield and quality. The estimates of the parameters differed from one model to the other. The values of the kernel weight estimates were not credible for some cultivars (*table V*). *Figure 4* showed that model L3V generally provided estimates of kernel weight closer to the observed values than model L4V. A thorough examination of the fit obtained for each variety showed that the lack of observations at early filling led to the estimation of the lower asymptote by a strongly negative value. As a consequence the maximal rate was under-estimated and the upper asymptote over-estimated. Model L3V was thus preferred among the different logistic models studied.

3.2. Weibull model

The likelihood ratio test (*table III*) and the graphs of residuals (*figure 2A* and *B*) showed that taking into account the variance heterogeneity did not improve the fit when the growth function was the Weibull function. The graph of standardized residuals showed a systematic under-estimation of the growth at the beginning of grain filling (*figure 2B*), which was not surprising, since the Weibull function we proposed assumed that the kernel weight was null when t was null, i.e. at anthesis. At this stage, the weight of the ovary, although low, cannot be considered as null. We generalized the Weibull function by considering the function defined by $f(t) = a(1 - e^{-b(t+d)^c})$, which is null when t is equal to $-d$. The algorithm maximizing the likelihood did not converge: the value of the parameter d which contributed to the solution of the algorithm seemed to tend to infinity. A Weibull function cannot thus be kept.

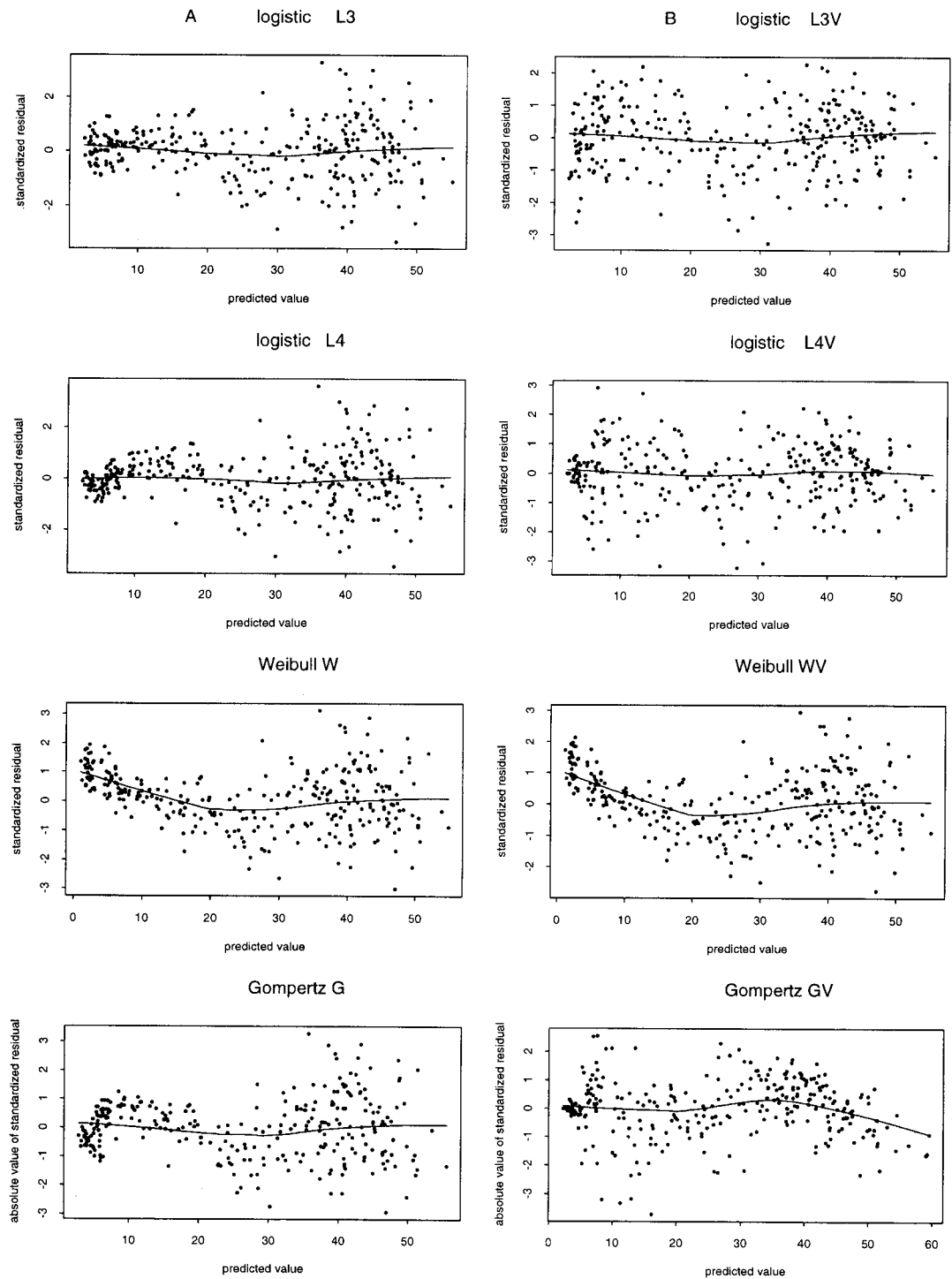


Figure 2. Graphs of standardized residuals against predicted values of kernel weight. A) Standardized residuals with constant variance. B) Standardized residuals with heterogeneous variances. A curve joining the points after smoothing was superimposed on each graph.

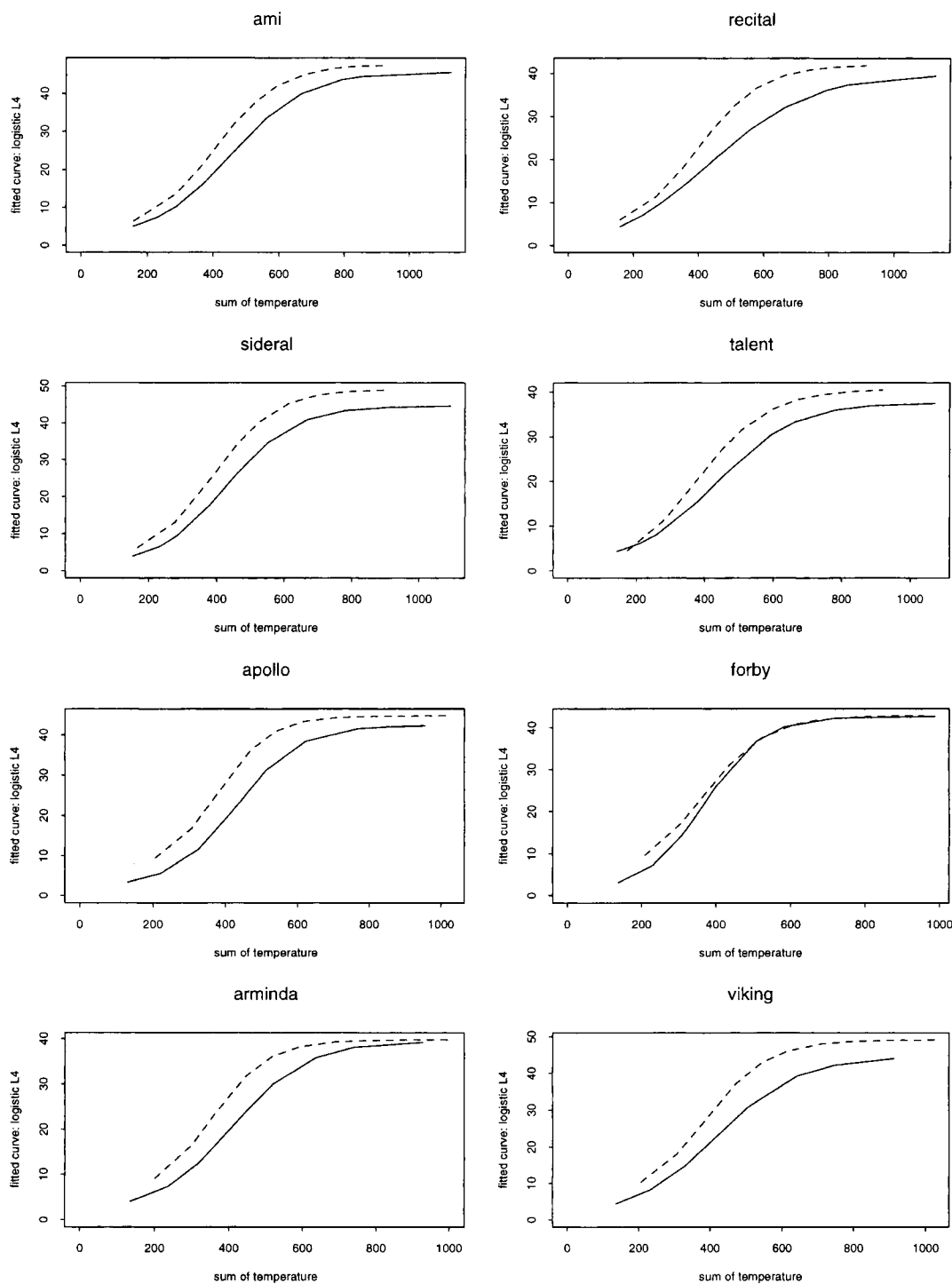


Figure 3. Fitted curves by a logistic model with four parameters (L4) in environment 3 (—, year 1995) and environment 6 (---, year 1996) for four early cultivars (Ami, Recital, Sideral, Talent) and four late cultivars (Apollo, Forby, Arminda, Viking).

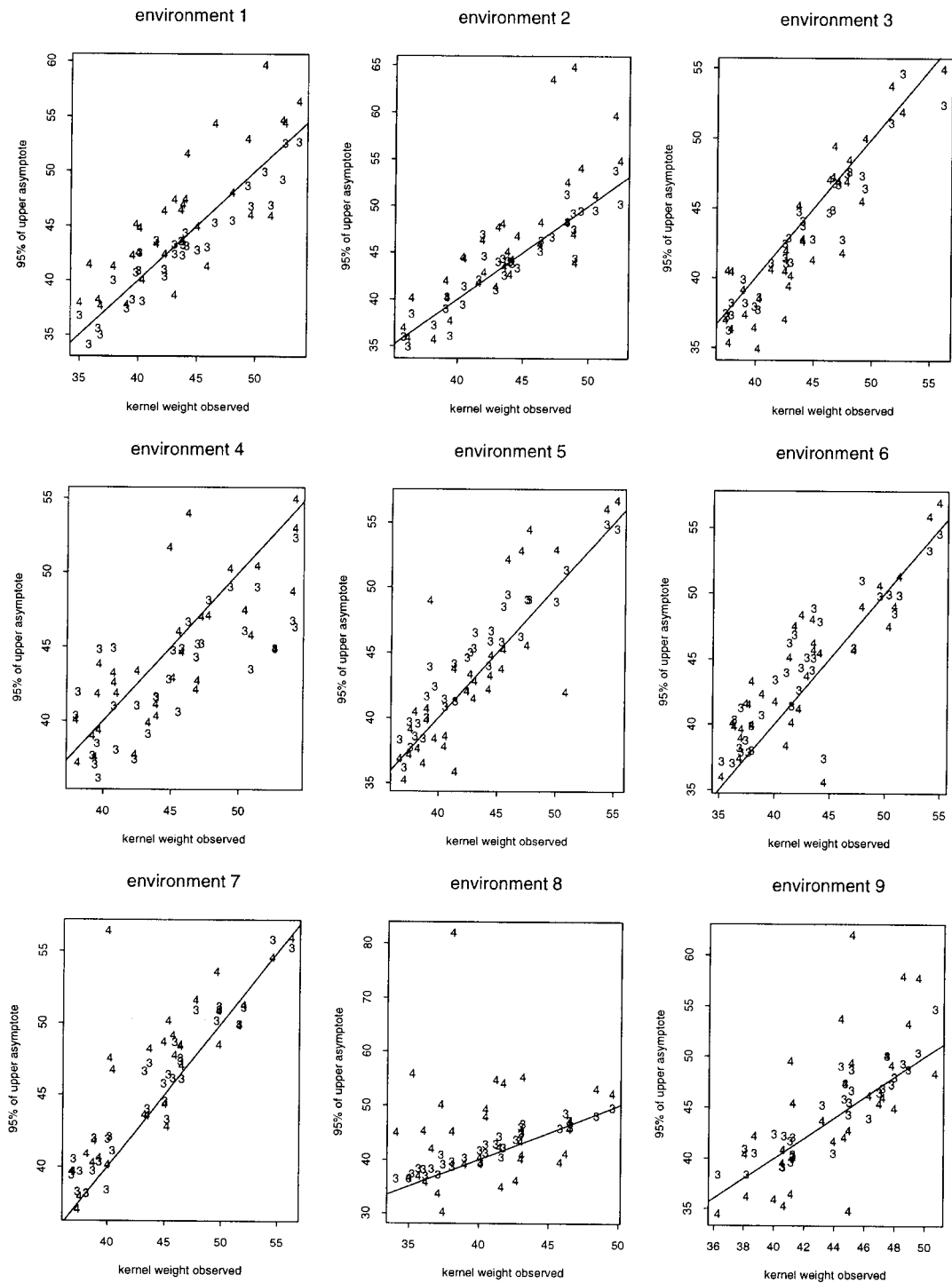


Figure 4. Estimates of the final kernel weight (95 % of upper asymptote) for model L3V (code 3) and model L4V (code 4, environments 1, 2, 3, 4, 8) or L4V2 (code 4, environments 5, 6, 7, 9) against kernel weight observed on a second experiment. The line corresponds to the equation $y = x$.

Table IV. Values of Akaike's criterion for the nine models considered.

Environment	L3	L3V	L4	L4V	L4V2	W	WV	G	GV
1	-581.21	-536.00	-602.21	-524.40	-524.69	-612.00	-605.61	-650.37	-545.72
2	-606.27	-522.10	-623.03	-510.89	-514.48	-646.10	-645.17	-672.69	NCV
3	-580.72	-548.37	-600.45	-552.78	-554.70	-629.57	-629.72	-648.23	-606.87
4	-588.12	-522.49	-602.91	-494.40	-497.16	-622.30	-621.67	-653.93	-525.67
5	-556.76	-513.05	-566.77	NCV	-492.13	-572.97	-572.21	-609.84	NCV
6	-567.64	-485.44	-586.27	NCV	-510.05	-577.36	-578.36	-629.63	NCV
7	-535.71	-474.03	-555.96	NCV	-493.46	-541.61	-542.27	-603.30	NCV
8	-558.82	-523.39	-573.14	-505.58	-519.97	-555.55	-550.93	-607.07	NCV
9	-570.71	-525.46	-585.79	NCV	-509.58	-565.72	-548.60	-620.19	NCV

L3: logistic model with three parameters, L4: logistic with four parameters, W: Weibull model, G: Gompertz model.

L3V, L4V, WV, GV: same models with modelling of the variance; L4V2: logistic model with four parameters, modelling of the variance and parameter $\tau = 2$.

NCV: the convergence of the estimating algorithm was not reached. No estimation available.

Table V. Values of the rate ($\mu\text{g}/^\circ\text{Cd}$), duration ($^\circ\text{Cd}$) and estimates of the final kernel weight (mg) and their 95 % confidence intervals (in brackets) and kernel weight observed (mg) for some variety \times environment \times block combinations.

	L3V	L4V	L4V2	Kernel weight observed
Rate	110.0 [102.6, 116.3]	93.2 [84.6, 101.7]	95.0 [86.1, 103.7]	
Duration	808 [761, 854]	959 [825, 1 092]	934 [814, 1 052]	
Estimate of kernel weight	50.0 [47.3, 52.6]	59.7 [51.3, 68.0]	58.1 [50.7, 65.4]	50.8
Rate	108.4 [101.6, 115.2]	98.4 [88.0, 108.9]	99.6 [89.6, 109.7]	
Duration	833.3 [783, 885]	925 [796, 1 053]	913 [801, 1 025]	
Estimate of kernel weight	53.8 [50.9, 56.7]	59.7 [51.6, 67.7]	59.0 [52.0, 66.0]	52.0
Rate	94.5 [88.2, 100.9]	87.0 [84.3, 89.7]	88.9 [82.7, 95.0]	
Duration	662 [628, 697]	820 [700, 940]	726 [647, 806]	
Estimate of kernel weight	37.4 [36.2, 38.6]	55.8 [41.8, 70.1]	44.3 [36.3, 48.3]	35.2
Rate	98.4 [92.1, 104.7]	89.7 [84.0, 95.5]	92.5 [85.1, 99.9]	
Duration	664 [632, 693]	744 [669, 818]	712 [644, 780]	
Estimate of kernel weight	38.18 [36.96, 39.40]	45.4 [39.5, 51.2]	42.8 [36.9, 48.3]	36.0
Rate	100.6 [93.1, 108.0]	95.7 [89.1, 102.2]	97.9 [90.0, 105.8]	
Duration	642 [606, 679]	692 [616, 769]	668 [600, 736]	
Estimate of kernel weight	39.7 [38.4, 41.0]	45.3 [38.4, 52.3]	42.8 [35.8, 49.7]	38.1
Rate	101.9 [93.6, 110.1]	118.0 [82.7, 153.2]	100.7 [93.4, 108.0]	
Duration	600 [561, 638]	777 [600, 954]	661 [569, 752]	
Estimate of kernel weight	38.7 [37.4, 40.0]	82.0 [25.3, 138.7]	49.5 [34.6, 64.3]	38.1

3.2. Gompertz model

The examination of the standardized residuals for the Gompertz model indicated an over-estimation at early filling followed by an under-estimation (figure 2B). After studying the absolute stan-

darized residuals, we chose to model the variance using the power function previously defined (figure 2A). However, the convergence was only reached for GV in three environments (table III). Similarly to model L4V, we set the value of τ to 2, making the convergence problem disappear. The

likelihood ratio test and the graph of absolute standardized residuals suggested that model GV was better than model G (*table III* and *figure 2*). However, the graph of the standardized residuals (*figure 1B*) showed an over-estimation at the end of grain filling which led us to finally reject this model.

4. Discussion

Taking into account both Akaike's criterion and the graphs of residuals showed that a logistic model was well adapted for fitting the data and thus characterizing the growth curves for their rate and duration. The relationship between the kernel weight estimates and the kernel weights observed in the second experiment planted to study yield led us to choose the logistic function with a null lower asymptote and a modelling of variance (L3V). Studies which compare models depicting grain growth are few. Jones et al. [14] compared, in rice, the negative exponential, the logistic and Gompertz functions and the cubic polynomial. The latter model was retained but the method for the choice was not described. Loss et al. [15], after comparison of the linear regression, the cubic polynomial and the logistic function with three parameters, considered the logistic as more useful owing to smaller confidence intervals for duration and to a better fit of the curve to observed values (smaller residuals). A characteristic of our study compared to the previous one was a definition of the rate and duration similar for the three growth functions; the presence of an upper asymptote in each model allowed a definition of duration common to the three curves.

None of the studies using the logistic with three parameters [5, 7, 8, 15] proposed modelling the variance errors. However, this modelling appeared necessary for well fitting the data sets when using the logistic and Gompertz functions. Thus, the possibility of modelling the variance of observations must be introduced into every process to model growth.

The introduction of two additional parameters, the lower asymptote and the parameter τ , com-

pared with the lack of observations at the beginning of grain filling, explained the problems of convergence of the estimating algorithm of the parameters. Indeed, the estimators of parameters maximized a criterion, the likelihood, for which the contribution of each observation depended on the inverse of the variance of that observation. For a model in which the variance of an observation increased with its expectation, the observations made at the beginning of grain filling contributed more to the estimation of the parameters than those made at the end of filling. This may explain why the environments in which the lag period was not observed presented problems of convergence during the estimating procedure for model L4V. This problem of no convergence for the environments of the year 1996, due to higher temperatures in early grain filling than in 1995, led us to reconsider the protocol for sampling. Indeed, in order to simplify the task of the experimenter, sampling dates are usually determined with numbers of days after anthesis, even if time is expressed as cumulative degree-days. Our results indicated that, to correctly describe growth using a thermal scale for the time, it is better to directly define sampling dates with sums of temperatures rather than with days; an alternative is to make the first sampling early after anthesis in order to integrate the lag phase. In the case of the 1996 environments, the first samplings started around 200 degree-days after anthesis, a sum of temperatures which seemed too high. Taking into account the sums of temperature observed in 1995, it seems preferable to start sampling spikes at 150 degree-days at the latest.

5. Conclusion

Comparing several functions classically used to describe growth phenomenon, using several statistical criteria, led us to keep the logistic model with three parameters and modelling of the variance to characterize dry matter filling in wheat kernel. The variations of the rate and duration of dry matter and nitrogen grain fillings and their relationship with the variations of the protein content will be presented in a subsequent paper.

Acknowledgements: We are grateful to P. Bérard for carrying out field trials, J. Gourdon and A. Chatain for assessing kernel weight.

References

- [1] Bauer A., Frank A.B., Black A.L., Estimation of spring wheat grain dry matter assimilation from air temperature, *Agron. J.* 77 (1985) 743–752.
- [2] Bertholdsson N.-O., Stoy V., Accumulation of biomass and nitrogen during plant growth in highly diverging genotypes of winter wheat, *J. Agron. Crop Sci.* 175 (1995) 167–182.
- [3] Bouvier A., Huet S., nls2. non-linear regression by s-plus functions, *Comput. Stat. Data Anal.* 18 (1994) 187–190.
- [4] Bruckner P.L., Froberg R.C., Rate and duration of grain fill in spring wheat, *Crop Sci.* 27 (1987) 451–455.
- [5] Campbell C.A., Cutforth H.W., Selles F., Depauw R.M., Clarke J.M., Dynamics of dry matter, N and P accumulation in the developing kernels of four spring wheat cultivars for irrigation and dryland, *Can. J. Plant Sci.* 70 (1990) 1043–1056.
- [6] Carroll R.J., Ruppert D., Transformation and Weighting in Regression. Monographs on Statistics and Applied Probability, Chapman and Hall, London, 1988.
- [7] Darroch B.A., Baker R.J., Grain filling in three spring wheat genotypes: statistical analysis, *Crop Sci.* 30 (1990) 525–529.
- [8] Darroch B.A., Baker R.J., Two measures of grain filling in spring wheat, *Crop Sci.* 35 (1995) 164–168.
- [9] Daynard T.B., Tanner J.W., Duncan W.G., Duration of the grain filling period and its relation to grain yield in corn, *Zea mays* L., *Crop Sci.* 11 (1971) 45–47.
- [10] Gebeyehou G., Knott D.R., Baker R.J., Rate and duration of grain filling in durum wheat cultivars, *Crop Sci.* 22 (1982) 337–340.
- [11] Hehn E.R., Barmore A., Breeding wheat for quality, *Adv. Agron.* (1965) 85–114.
- [12] Housley T.L., Kirleis A.W., Ohm H.W., Patterson F.L., Dry matter accumulation in soft red winter wheat seeds, *Crop Sci.* 22 (1982) 290–294.
- [13] Huet S., Bouvier A., Gruet M.-A., Jolivet E., Statistical Tools for Nonlinear Regression. A Practical Guide with S-PLUS Examples, Springer Series in Statistics, Springer-Verlag, New York, Inc., 1996.
- [14] Jones D.B., Peterson M.L., Geng S., Association between grain filling rate and duration and yield components in rice, *Crop Sci.* 19 (1979) 641–644.
- [15] Loss S.P., Kirby E.J.M., Siddique K.H.M., Perry M.W., Grain growth and development of old and modern Australian wheats, *Field Crops Res.* 21 (1989) 131–146.
- [16] Motzo R., Giunta F., Deidda M., Relationships between grain-filling parameters, fertility, earliness and grain protein of durum wheat in a Mediterranean environment, *Field Crops Res.* 47 (1996) 129–142.
- [17] Mou B., Kronstad W.E., Duration and rate of grain filling in selected winter wheat populations: I. inheritance, *Crop Sci.* 34 (1994) 833–837.
- [18] Nass, H.G., Reiser R., Grain filling period and grain yield relationship in spring wheat, *Can. J. Plant Sci.* 55 (1975) 673–678.
- [19] Ratkowsky D.A., Handbook of Non Linear Regression, Dekker, 1989.
- [20] Rousset M., Triboï E., Branlard G., Godon B., Influence du génotype et du milieu sur les tests d'appréciation de la valeur d'utilisation du blé tendre (*Triticum aestivum* em. Thell.), *Agronomie* 5 (1985) 653–663.
- [21] Russelle M.P., Wilhemm W.W., Olson R.A., Power J.F., Growth analysis based on degree days, *Crop Sci.* 24 (1984) 28–32.
- [22] Seber G.A.F., Wild C.J., Non Linear Regression, Wiley, New York, 1989.
- [23] Sofield I, Evans L.T., Cook M.G., Wardlaw I.F., Factors influencing the rate and duration of grain filling in wheat, *Aust. J. Plant Physiol.* 4 (1977) 785–797.
- [24] Statistical Sciences, SPLUS programmer's manual, version 3.2, Seattle: StatSci, a division of MathSoft, Inc., 1993.
- [25] Stone M., An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion, *JRSS* 29 (1977) 44–47.
- [26] Stone P.J., Savin R., Wardlaw I.F., Nicolas M.E., The influence of recovery temperature on the effects of a brief heat shock on wheat. I. Grain growth, *Aust. J. Plant Physiol.* 22 (1995) 945–954.
- [27] Triboï E., Modèle d'élaboration du poids du grain chez le blé tendre (*Triticum aestivum* em Thell), *Agronomie* 10 (1990) 191–200.
- [28] Wardlaw I.F., Moncur L., The response of wheat to high temperature following anthesis. I. The rate and duration of kernel filling, *Aust. J. Plant Physiol.* 22 (1995) 391–397.