

Adjustment for missing plants in cassava evaluation trials

J. C. Pérez · H. Ceballos · I. C. Ramírez · J. I. Lenis ·
F. Calle · N. Morante · G. Jaramillo · M. Lentini

Received: 19 February 2009 / Accepted: 9 September 2009 / Published online: 26 September 2009
© Springer Science+Business Media B.V. 2009

Abstract Evaluations with a uniform plant stand are fundamental for efficient selections in cassava (*Manihot esculenta* Crantz) breeding. However, it is difficult to correct data of missing plants. The overall objective of this study was to analyze yield losses in agronomic trials and to develop a function that satisfactorily adjusts plot yields affected by a varying number of missing plants. Eight clones were initially evaluated in different locations for five consecutive years. For every genotype, mean plot yields decreased as the number of missing plants increased. Average losses ranged from 10.6 to 78.8% by removing from one to eight plants, respectively. Yields per plant increased significantly when more than four plants were removed due to a compensatory growth effect. Graphic analyses showed that the power function was the best explaining the relationship between fresh root yield and number of harvested plants. A model that properly adjusted yield for all varieties but one is proposed. Hopefully, this formula to adjust yields will

help to improve the quality of cassava trials worldwide.

Keywords Competition · Experimental error · Missing plants · Compensatory growth

Introduction

Uniform competition in cassava (*Manihot esculenta* Crantz), as for most crops, is fundamental for acceptable experimental errors in evaluation trials. However, it is not always feasible to achieve perfect plant stands in cassava trials and precision of results is frequently not satisfactory.

The morphological characteristics of cassava are highly variable. Plant height can vary from 1 to 4 m and plant type ranges from highly branching to non-branching erect types. Plant architecture influences the amount of planting material that a mother plant can produce. Erect, non-branching types generally produce larger amount of planting material and the harvest, storage and transport of stems is greatly facilitated (Ceballos and de la Cruz 2002; Alves 2001). The number of commercial stakes obtained from a single mother plant in a year ranges from 3 to 30, depending upon growth habit, climate, management, and soil conditions. This is considerably less than the propagation rate that can be achieved with other commercial crops (Leihner 2001).

J. C. Pérez (✉) · H. Ceballos · J. I. Lenis ·
F. Calle · N. Morante · G. Jaramillo
Cassava Breeding Project, International Center
for Tropical Agriculture (CIAT), Palmira, Colombia
e-mail: j.c.perez@cgiar.org

H. Ceballos
e-mail: h.ceballos@cgiar.org

H. Ceballos · I. C. Ramírez · M. Lentini
Universidad Nacional de Colombia, Medellín, Colombia

When roots are harvested the previous season, the stems are also collected and stored, typically under the shade of a tree (Ceballos et al. 2007; Morante et al. 2005). Stems can only be stored for 1 to 2 months, depending on the environmental conditions. Several physiological and sanitary factors affect sprouting capacity which, combined with the low multiplication rate of planting material, results in frequent and chronic problems of variation in plant densities.

The effect of missing plants on plot yield may not be noticeable when there are one or two missing plants. The compensatory growth of neighboring plants usually helps to reduce differences in total plot yield. However, as the proportion of missing plants increases, the compensatory growth of the remaining plants is not enough to correct total plot yield (Gomez and De Datta 1972; James et al. 1973; Kamidi 1995; Mead 1968).

The covariate analysis can some times adjust cassava experimental plot yields when the plans are missing only for a short time before harvest time. However, when plants are missed throughout the growing season, competition effects and compensatory growth invalidate the linear covariance adjustment. Therefore, the relationship between plot yield and plot stand is no longer linear, and linear covariance analysis may result in unreliable yield estimates and failure to reduce experimental errors. The relationship between plant density and crop yield has received previous attention (Schmidt et al. 2001; Vencovsky and Cruz 1991; Verones et al. 1995; Willey and Heath 1969). Kamidi (1995) proposed an exponential model to correct plot stands in maize reinforcing the concept that the linear covariance analysis using plot stand as the covariate is no satisfactory, especially when the plants are missing long before maturity.

The objectives of this work were to estimate yield losses due to missing plants in experimental cassava trials and propose a model that can be used to correct yields based on ideal plant stands.

Materials and methods

Field evaluation trials

A set of agronomic evaluations, with eight different varieties, were conducted during five years at four

contrasting environments in Colombia (Departments of Atlántico, Cauca, Meta and Valle del Cauca). For each variety, eight different treatments were applied by removing one, two, and up to eight ‘central’ plants of each plot, as well as a control treatment (no missing plant). The plants inside experimental plots were numbered as illustrated in Fig. 1, and were randomly removed from the treatment plot to achieve the specific treatment target 2 months after planting (before plant competition between neighboring plants started). Plots consisted of five rows of five plants, spaced 1 m apart within rows and 1 m between rows (standard plant density for cassava).

The number of experiments per variety was variable (Table 1). Some varieties were evaluated only 1 year, while others were evaluated for the 5 years this study lasted and at more than one environment. Eight cassava varieties were originally evaluated: CM 4919-1 and MTAI 8 are adapted to sub-humid environment; CM 4574-7, CM 6438-14 and CM 6740-7 are adapted to acid-soil savannas; CM 523-7, MCOL 1505 and SM 1058-13 are adapted to the mid-altitude valleys environment. These varieties differ in branching type and, therefore, competitive ability.

The design used was a randomized complete block design with three replications per experiment. Individual analyses of variances were performed for each experiment and combined for each variety within each environment and year. Graphic analyses were used to identify the best model to explain the

0	0	0	0	0
0	1	2	3	0
0	4	5	6	0
0	7	8	9	0
0	0	0	0	0

Fig. 1 Scheme illustrating the identification of each ‘central’ plant inside experimental plots for measuring the effect the missing plant in cassava evaluation trials

Table 1 Analysis of variance per experiment for fresh root plot evaluated at four contrasting environments in Colombia

Variety	Experiment	Mean squares			CV (%)
		Reps	Treatment	Error	
MTAI 8	1	4.454	95.817**	8.085	19.1
	2	1.449	51.048**	3.783	19.1
	3	10.490	209.155**	4.217	10.5
CM 4919-1	4	3.794	353.523**	12.525	15.7
CM 4574-7	5	2.850	108.800**	6.433	17.8
	6	3.201	44.504**	4.691	24.7
CM 6438-14	7	3.411	41.952**	3.981	23.2
	8	3.917	69.021**	3.287	18.9
	9	18.703*	17.077**	4.722	29.7
	10	49.901*	49.480**	9.434	47.9
	11	49.696*	45.299**	8.115	43.9
CM 6740-7	12	0.161	27.471**	3.806	25.6
	13	8.108	80.355**	6.291	29.7
	14	17.391	103.696**	4.959	16.2
	15	5.807	99.239**	6.785	22.0
	16	7.880	84.546**	7.604	23.0
	17	12.591	237.022**	18.516	21.7
CM 523-7	18	3.207	104.794**	4.285	13.9
	19	18.918	42.059**	5.896	21.3
MCOL 1505	20	3.009	43.242**	6.861	28.3
	21	20.867	92.671**	9.540	23.0
	22	0.158	24.879	16.847	40.8
CM 4574-7	23	13.963	122.649**	7.718	16.0
	24	0.858	53.652**	7.398	25.9
CM 6740-7	25	16.234*	51.679**	4.211	15.9
	26	12.339*	79.705**	3.378	14.1
	27	122.492	472.643**	39.864	17.7
SM 1058-13	28	75.371*	362.933**	18.620	16.3

* Significant at the $P < 0.05$ probability level

** Significant at the $P < 0.01$ probability level

relationship between fresh root yield and number of plants harvested. The information produced from the analysis of these trials was used to compare different models and select the best one based on its capacity to correct measured yields to the perfect plant stand.

Estimation of yield losses due to missing plants

Graphic analysis was initially performed to understand the relationship between plot yields of perfect plant stands (no missing plant) versus those for each treatment (different number of missing plants). This

analysis showed that the power function best explained the relationship between fresh root yield and number of harvested plants. This function, in all cases, presented an R^2 value above those of the others functions analyzed (exponential, logarithmic and linear). The proposed model considers a power decline of yield associated with decreasing plant stand. The adjusted plot yield was, therefore, a function of both observed plot yield and plot stand, as follows:

$$y_a = y_o \left[1 + \left(1 - \frac{N_o}{N_a} \right) \alpha \left(\frac{N_o}{N_a} \right)^{-\beta} \right]$$

where y_a is the adjusted plot yield, y_o the observed plot yield, N_a is the ideal plot stand (in this case nine plants), N_o is the number of harvested plants (in this experiment varying from one to nine depending on the assigned treatment), and α and β are unknown parameters. This model imposes the requirement that the adjusted yield should coincide with the observed yield when ideal and observed plot stands are equal. Additionally, a linear model was also fitted to the data, as follows:

$$\frac{y_a}{y_o} = \alpha + \beta \frac{N_o}{N_a}$$

where y_a , y_o , N_a , and N_o are the same parameters as indicated above.

The α and β values were estimated by a non-linear least squares iterative procedure. The SAS non-linear regression procedure based on the modified Gauss–Newton method was used to fit the proposed model. The models fit were assessed from the coefficients of determination (R^2) and magnitude of the residual values. Additionally, the invariances of the best models were tested. This test helped to define whether or not to fit with a common α and β for all cassava varieties and all environments (Boché and Lavalley 2004). Statistical tests of these hypotheses were performed on the basis of the extra sums of squares or the conditional error principle (Milliken and Debruin 1978).

Results

Analysis of variance for each experiment showed, as expected, significant differences among treatments

for fresh root yield per plot (Table 1). The coefficient of variations (CV) ranged from 10.5 (Experiment 3) to 47.9% (Experiment 10). Experiments with CV above 30% (Experiments 10, 11 and 22) and two experiments with high root-rot incidence (Experiments 8 and 9) were eliminated from further analyses. The combined analyses of variance for each variety across years within environments showed highly

significant differences among years and treatments (Table 2). The treatment-by-environment interactions were not significant, except for clone MTAI 8, which presented highly significant values. It is important to note that varieties CM 4574-7 and CM 6740-7, adapted to the acid-soil savannas (Meta), were also evaluated in the mid-altitude valleys environment (Valle del Cauca).

Table 2 Mean squares from the ANOVA for each variety combined across years within environment

Source of variation	Year/exp	Reps	Treatment	T × Y/exp	Error	CV (%)	SE
MTAI 8 (Atlántico)	600.7**	4.2	319.8**	18.1**	5.4	15.6	1.10
CM 4574-7 (Meta)	274.5**	5.7	182.9**	6.2	4.8	20.8	1.03
CM 4574-7 (Valle)		14.0	122.7**		7.7	16.0	2.27
CM 4574-7 (All)	495.3**	4.5	287.9**	10.0	5.8	19.6	0.98
CM6740-7 (Meta)	181.0**	1.2	360.3**	8.8	6.2	23.3	0.91
CM6740-7 (Valle)	56.1**	3.1	161.1**	12.5	5.4	19.2	1.01
CM6740-7 (All)	133.9**	1.2	505.8**	10.8	5.9	21.6	0.71
MCOL1505 (Valle)	138.9**	14.0	152.7**	7.3	6.7	22.5	1.22
CM523-7 (Valle)	335.0**	2.0	322.7**	19.1	11.5	19.6	1.96
SM1058-13 (Cauca)	837.8**	19.4	861.0**	59.1	31.8	18.4	3.26
CM4919-1 (Atlántico)		3.8	353.5**		12.5	15.7	2.89

** Significant at the $P < 0.01$ probability level

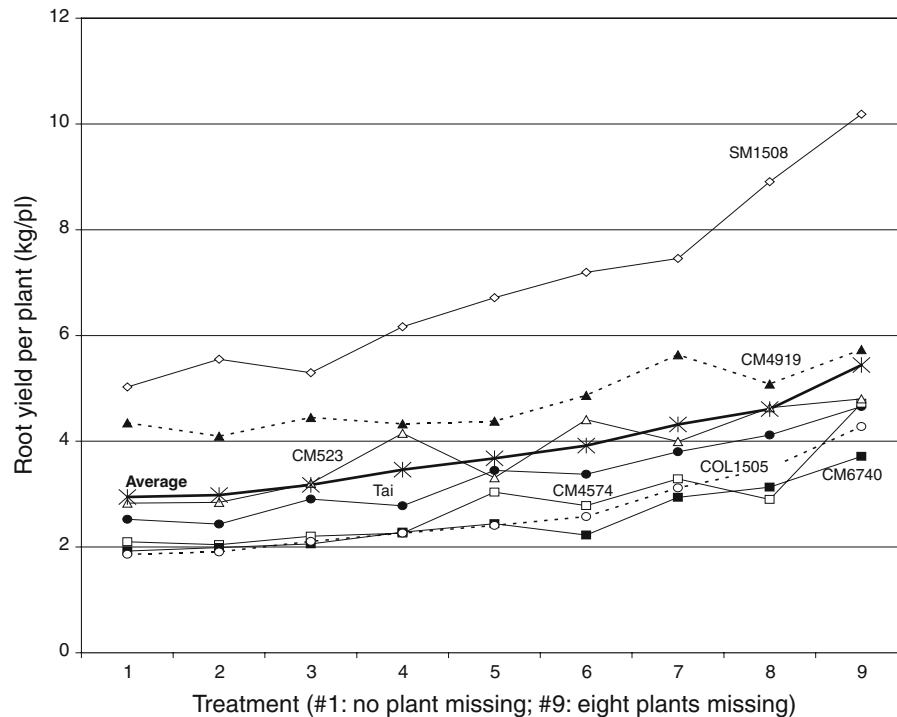


Fig. 2 Yield per plant of different varieties (across trials). The non-linear relationship becomes evident when yield per plant increases as the number of missing plants is higher than 4

Table 3 Varieties mean of observed plot yield and adjusted plot yield

Harvested Plants	MTAI 8		CM 4919-1		CM 4574-7		CM 6740-7		CM 523-7		MCO1 1505		SM 1058-13	
	y_o	y_a	y_o	y_a	y_o	y_a	y_o	y_a	y_o	y_a	y_o	y_a	y_o	y_a
9	22.7	22.7	39.1	39.1	18.9	18.9	17.3	17.3	25.5	25.5	16.8	16.8	45.2	45.2
8	19.5	21.2	32.8	35.7	16.4	17.8	15.9	17.3	22.8	24.8	15.3	16.6	44.4	48.4
7	20.3	24.4	31.1	37.3	15.4	18.5	14.4	17.3	22.4	26.9	14.7	17.7	37.1	44.4
6	16.7	22.3	26.0	34.7	13.6	18.1	13.7	18.3	24.9	33.3	13.6	18.1	37.0	49.5
5	17.2	26.2	21.9	33.3	15.2	23.1	12.2	18.5	16.6	25.1	12.0	18.3	33.6	51.0
4	13.5	24.0	19.5	34.6	11.1	19.8	8.9	15.8	17.6	31.3	10.3	18.3	28.8	51.5
3	11.4	24.8	16.9	36.7	9.9	21.4	8.8	19.2	12.0	26.0	9.4	20.3	22.4	48.6
2	8.2	23.8	10.2	29.4	5.8	16.8	6.3	18.1	9.3	26.8	7.0	20.2	17.8	51.5
1	4.7	22.2	5.7	27.3	4.7	22.5	3.7	17.7	4.8	22.9	4.3	20.4	10.2	48.5
SD ^a	5.96	1.54	10.86	3.78	4.83	2.20	4.59	0.96	7.33	3.30	4.12	1.46	11.98	2.50
SD ^b	6.79	9.69	5.39	13.45	7.58	8.13	5.49	6.48	6.88	12.93	7.29	5.40	15.48	18.77

y_o observed plot yield, y_a adjusted plot yield

^a For each variety, standard deviations for observed plot yields (left) and using the general model correction (right)

^b For each variety, standard deviations for observed plot yields corrected by the yield per plant approach (left) or the linear approach (right)

Figure 2 illustrates the relationship between yield per plant and number of missing plants. As expected, fresh root yield on a per plant basis remains relatively stable when few plants are missing. However, as the number of missing plants is ≥ 4 , the yield per plant tends to increase considerably. For all varieties the mean plot yield decreased as the number of missing plant increased (Table 3). Average yield losses by removing one up to eight plants ranged from 10.6 to 78.8%, respectively.

The R^2 values were computed from the analysis of variance routine provided on the SAS listing. The power model was associated with a largest value of R^2 (0.9438) considerably better than the linear regression model ($R^2 = 0.5973$). Convergence of the power model was achieved in fewer than four iterations. Plots of the predicted yield ratio against the corresponding observed values indicated that the power model was appropriate. The fitted curve and the actual values are shown in Fig. 3. It can be observed that variability increases as the number of missing plants increases. In other words, as the number of plants increased the reliability of the adjustment was reduced.

Analysis of residuals for all analyses indicated little evidence to disprove the hypothesis that residuals were normally distributed with a mean equal to zero. The approximate F statistics developed by

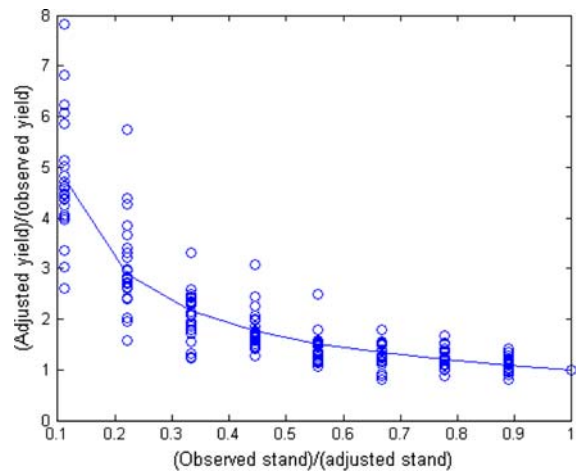


Fig. 3 Prediction curve for general values

Milliken and Debruin (1978) were used to test the significance of the extra sums of squares due to common fits. Significant differences were detected between parameters for varieties and environment ($P < 0.05$). Table 4 shows the estimated parameter values individually for each variety, environment and combined data.

The fitted curves for all varieties are depicted in Fig. 4. Invariance analysis for most varieties did not show significant differences between their models indicating similar responses. Variety CM 4919-1, on

Table 4 Estimated parameter values for varieties, environment and for combined data

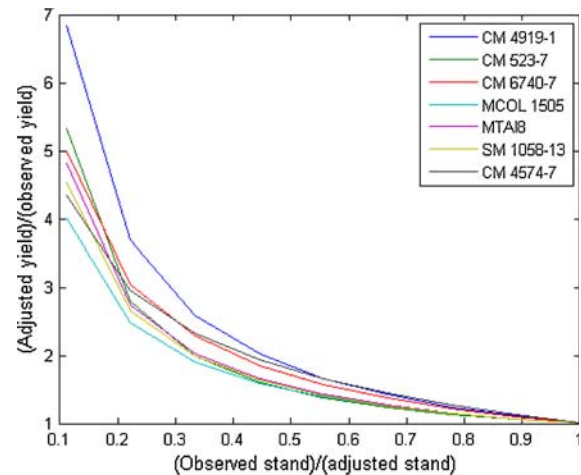
	α	β
Variety		
CM 4574-7	1.052 (0.1956)	0.579 (0.0985)
CM 4919-1	0.870 (0.0836)	0.920 (0.0477)
CM 523-7	0.462 (0.1122)	1.070 (0.1183)
CM6740-7	0.821 (0.1761)	0.774 (0.1183)
MCOL 1505	0.537 (0.1786)	0.840 (0.1673)
MTA18	0.550 (0.1293)	0.934 (0.1165)
SM 1058-13	0.548 (0.1177)	0.901 (0.1070)
Average	0.691	0.718
Environment		
Cauca	0.548 (0.1177)	0.901 (0.1070)
Atlántico	0.630 (0.1449)	0.929 (0.1114)
Meta	1.1591 (0.1747)	0.6415 (0.0787)
Valle	0.455 (0.0882)	0.948 (0.0958)
Average	0.698	0.855
Common fit (all experiments)	0.727 (0.0805)	0.805 (0.0559)

the other hand, showed highly significant differences compared with the other varieties. According to the information generated, as expected, there was a variation in the response to missing plants for different varieties or the different environments where the trials were conducted. Nonetheless, a general model across varieties and environments was evaluated resulting in estimates for $\alpha = 0.727$ and $\beta = 0.805$.

The general model was used to estimate adjusted yield to uniform full plot stands for each variety (Table 3). The analysis of variance (data not shown) indicated no significant difference after adjusted yield for all varieties, except for CM 4919-1, indicating a good fit of the proposed general model to adjust yield plot when there are missing plant in experimental plots, regardless of the environment where the trials are conducted, the varieties used or the number of missing plants.

Discussion

The results obtained in this work supported the expected effects of missing plant in field evaluations of cassava. The yield per plant increased along with the number of missing plants, mainly because the remaining plants around the missing one(s) were favored by

**Fig. 4** Prediction curve for all varieties

less competition for limiting environmental factors such as light, water and nutrients (Fig. 2). The average yield when only one plant was harvested varied from 3.7 (CM 6740-7) to 10.2 kg (SM 1058-13), indicating large variation between varieties (Table 3).

The ultimate objective of this study was to develop a model capable of adjusting total plot yields (for treatments where one or more plants were missing) as close as possible to the values observed in the perfect plant stand of the same variety. The analysis of invariance, taking into account varieties and environments, showed similar responses to different groups. However, some models had significant differences indicating that for specific varieties and environment their α and β parameters were different.

The general model across environments and varieties (based on $\alpha = 0.727$ and $\beta = 0.805$) was used to adjust total plot yields as presented in Table 3. It is recognized that, ideally, the correction for missing plants should be done individually for each variety and/or location. However, the information required to make such adjustment is usually missing beforehand and, consequently, such adjustment is seldom possible. The application of a more general model that can be applied by default in the analysis of different trials is highly desirable (Gomez and De Datta 1972), even if the precision in the adjustment is not perfect. The interest to develop a general model applicable to different cassava varieties and environmental conditions defined the nature of this study. Different set of environments with varying average yield potential and the use of varieties with contrasting plant

architectures was purposely chosen, therefore, for this study.

There are few available options to reduce the experimental errors derived from missing plants. The most obvious strategy would be to maximize the possibility of obtaining perfect plant stands. In many crops it is feasible to overplant and then reduce the number of surviving plants down to the desired plant density. However, in the case of cassava, availability of planting material is a chronic limitation because of the low multiplication rate (Ceballos et al. 2007). This is particularly the case in recurrent selection schemes (Morante et al. 2005). Therefore, the occurrence of missing plants is unavoidable and approaches to adjust yields a necessity. The simplest correction would be a linear approach based on the yield per plant estimate: (total plot yield/number of harvested plants)*ideal plant stand. As demonstrated (Fig. 2), however, this approach would tend to overestimate corrected yields when the number of missing plants is high. Another linear correction could be based on the co-variance analysis. At the bottom of Table 3 the standard deviation of the corrected plot means for these two approaches is presented. In every case the application of the general model proposed in this study resulted in considerably smaller standard deviation values, indicating that the general model is better than other available methods.

The particular performance of cultivar CM 4919-1, a widely grown variety in the sub-humid environment of Colombia's northern coast, was interesting because it failed to fit the general model proposed in this study. Plant height of CM 4919-1 was relatively low and it was the only clone that did not branch at all, showing a very distinctive, completely erect plant type. These features could explain the atypical behavior of this clone. Still the general model provided the best adjustment for total plot yields (Table 3).

Until now, cassava breeders have used two unsatisfactory approaches to overcome the problem of missing plants (re-planting a stem cutting to replace the missing plant or harvest plants in a border row). The general model proposed in this study can only be applied in trials where no such corrective measures have been used. Comparisons of coefficient of variation before and after adjusting the means would provide a fair estimate of the relative values of the method. The proposed general model is:

$$y_a = y_o \left[1 + \left(1 - \frac{N_o}{N_a} \right) 0.727 \left(\frac{N_o}{N_a} \right)^{-0.805} \right]$$

where each parameter is the same as described in the materials and method section.

References

- Alves AAC (2001) Cassava botany and physiology. In: Hillcocks RJ, Thresh JM, Bellotti AC (eds) Cassava: biology production and utilization. CABI Publishing, Wallingford, pp 67–89
- Boché S, Lavalle A (2004) Comparación de modelos no lineales, una aplicación al crecimiento de frutos de carozo. *Rev Soc Arg Gen* 8(1). <http://www.s-a-e.org.ar/revista-vol8.html>
- Ceballos H, De la Cruz GA (2002) Taxonomía y Morfología de la Planta. In: Ospina B, Ceballos H (eds) La Yuca en el Tercer Milenio. CIAT, Cali, pp 17–33
- Ceballos H, Fregene M, Pérez JC, Morante N, Calle F (2007) Cassava genetic improvement. In: Kang MS, Priyadarshan PM (eds) Breeding major food staples. Blackwell Publishing, Ames, pp 365–391
- Gomez KA, De Datta SK (1972) Missing hill in rice experimental plots. *Agron J* 64:163–164
- James WC, Lawrence CH, Smith CS (1973) Yield losses due to missing plants in potato crops. *Am Potato J* 50(10):345–352
- Kamidi RE (1995) Statistical adjustment of maize grain yield for sub-optimal plots stands. *Exp Agric* 31:299–306
- Leihner D (2001) Agronomy and cropping systems. In: Hillcocks RJ, Thresh JM, Bellotti AC (eds) Cassava: biology, production and utilization. CABI Publishing, Wallingford, pp 91–113
- Mead R (1968) Measurement of competition between individuals plants in a population. *J Ecol* 56(1):35–45
- Milliken GA, Debruin RL (1978) A procedure to test hypotheses for nonlinear models. *Commun Stat Theory Methods* 7:65–69
- Morante N, Moreno X, Pérez JC, Calle F, Lenis JI, Ortega E, Jaramillo G, Ceballos H (2005) Precision of selection in early stages of cassava genetic improvement. *J Root Crops* 31:81–92
- Schmidt ER, Cruz CD, Zanuncio JC, Pereira PRG, Ferrão RG (2001) Avaliação de métodos de correção do estande para estimar a produtividade em milho. *Pesq Agropec Bras*, Brasília 36(8):1011–1018
- Vencovsky R, Cruz CD (1991) Comparação de métodos de correção do rendimento de parcelas com estandes variados, I. Dados Simulados. *Pesq Agropec Bras*, Brasília 26(5):647–657
- Verones JA, Cruz CD, Correa LA, Scapim CA (1995) Comparação de métodos da ajuste do rendimento de parcelas com estandes variados. *Pesq Agropec Bras*, Brasília 30(2):169–174
- Willey RW, Heath SB (1969) The quantitative relationship between corn populations and yield. *Adv Agron* 21:281–321