Geostatistical method for independent sampling in kiwifruit vine to estimate yield components.

Robert Habib, D. Tisne-Agostini, M.P. Vanniere, Pascal P. Monestiez

New Zealand Journal of Crop and Horticultural Science, 1991, Vol. 19: 329–335
0114–0671/91/1904–0329 $2.50/0

Geostatistical method for independent sampling in kiwifruit vine to estimate yield components

ROBERT HABIB
INRA Station d'Agronomie
F-84143 Montfavet Cedex
France

DOMINIQUE TISNE-AGOSTINI
MARIE-PASCALE VANNIERE
INRA Station d'Agronomie San Giuliano
F-20230 San Nicolao
France

PASCAL MONESTIEZ
INRA Laboratoire de Biométrie
F-84143 Montfavet Cedex
France

Abstract Classically used in mining problems, the semi-variogram function has been tested for its application to kiwifruit (*Actinidia deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson) sampling. The fruit characteristics studied were: weight, dry matter (%), soluble solid content, acidity, and the distance between fruit was taken as the number of "active" forks to go from one fruit to another. It appeared that there was a dependency between samples taken at short distance (i.e., borne by the same lateral) whatever the fruit characteristic under consideration. The variogram function increased with increasing distances, indicating less and less correlations between samples. Fruits were essentially independent where taken from different canes separated by at least two forks. An anomalous feature in fruit weight correlations at $h = 8$ appeared to be related to the interactions between morphogenesis and pruning in kiwifruit vines, but has a quite small effect in terms of dependency between samples.

H91032
Received 27 April 1991; accepted 4 October 1991

Keywords *Actinidia deliciosa*; geostatistics; variogram function; fruit quality; sampling scheme; soluble solids content; fruit dry matter; fruit weight

INTRODUCTION

Problems of sampling in horticultural experiments have not been extensively studied. When tree characteristics are estimated through total harvest or global-tree data (e.g., total yield), there are statistical solutions of the problem of sample size estimate (Marini 1985; Millier et al. 1986). However, when the tree characteristic under consideration is an averaged value which must be estimated through samples taken in the tree, the problem of sampling in a fruit tree or in a fruit vine has not yet received a general solution. The main statistical difficulty is that the samples taken from within a given vine cannot generally be considered independent (Hopkirk et al. 1986).

This problem was recently addressed by Monestiez et al. (1989, 1990). They proposed to use geostatistics (Matheron 1965; Journel & Huijbregts 1978)) to analyse the spatial dependence between fruits growing on the same tree. This method was firstly developed in mining prospecting and is now classically used in spatial data analysis. Monestiez et al. (1989) showed that the method applied to a theoretical tree structure. Then, they applied it with success to analyse the spatial dependence of fruit weight and soluble solids content, and leaf nitrogen content in an actual peach tree, showing that samples taken at short distances were not independent. The method was later applied to apricot trees (Audergon et al. 1989).

We are interested here in fruit sampling within a kiwifruit (*Actinidia deliciosa* (A. Chev.) C.F. Liang et A.R. Ferguson)vine, when the purpose is to estimate mean values averaged within the whole vine, or within some selected parts (e.g., Grant & Ryugo 1984; Snelgar et al. 1988).

Our aim is to apply geostatistics to describe the spatial dependences between fruits sampled at different distance from each other within the vine

structure. Consequently, the distance at which fruits could be considered as statistically independent would be determined from data and used to define a simple sampling scheme.

SPATIAL ANALYSIS

The mathematical background of the geostatistical methods as applied to a tree structure is fully described in Monestiez et al. (1989). We will briefly describe here the application of the variogram function (also called semi-variogram) to analyse the spatial dependence of kiwifruit characteristics.

Within the vine structure, each pair x, y of fruits can be considered to be at a distance h, calculated as the number of active forks between the two fruits. The term "fork" is taken here as an actual fork or as a change in age within a given branch (e.g., a leader). The term "active" is used to designate a fork which lies on the path from a given sample to another one, and which matches with a shoot number change in the path. The fruit peduncle is considered as an active fork. Then, for two fruits borne on the same lateral $h = 2$. Two fruits at a Distance 4 are borne on the same cane on two different laterals, and two fruits at a Distance 6 are borne on two different canes borne on the same main shoot. Other types of distance can be used (Monestiez et al. 1989, 1990), but are less easy to measure.

The estimation of the variogram function at each h, is done from data by calculating the non-parametric estimator, $G(h)$:

$$G(h)\frac{1}{2\times N(h)}\sum_{N(h)}(Z(x)-Z(y))^2$$

where N is the total number of fruit pairs at distance h, and $Z(x)$ and $Z(y)$ are the experimental values of fruits x and y for the fruit variable Z under study (e.g., fruit acidity). Note that $G(h)$ is an estimate of half the variance of the variable difference. Consequently, if $Z(x)$ and $Z(y)$ are independent whatever x and y at distance h, $G(h)$ is an estimate of the sampled fruit variance.

Variogram functions are characteristic of the relations of dependence existing between fruits on the tree. Classically, three different parts can be defined:

(1) A discontinuity at the origin. The variogram function is by definition null at the origin. For a distance close to zero, there is often a significant value observed, usually called the "nugget effect". This can be broken down into two factors—measurement error and microscale variations. It means that fruit variability exists at very short distances and that two fruits on the same position may have quite different behaviour.

(2) The variogram function increases with increasing distances. This means that the further apart the fruits, the more independent they are.

(3) The function eventually reaches a steady value, the sill. At this distance, the fruits have no more interaction and can be considered as independent. This distance is named the range and is a important parameter for spatial description. The range is generally estimated by the beginning of the sill located on the curve. The value of the sill is close to the variance of the global population. For the highest values of the distance, random fluctuations can appear on the estimated variogram. Two possible explanations are a non-regular phenomenon at the largest distances or a low accuracy in estimation as a result of the small number of pairs at those distances.

MATERIALS AND METHODS

Four 10-year-old kiwifruit vines (cv. Hayward) labelled K1-K4, 7 × 4.5 m spaced, grown on "T-bar" trellises in north-eastern Corsica (France) were used. They were pruned in order to keep 30, 36, 49, and 24 canes per vine respectively. Vines K1 and K2 were close to the irrigation system (sprinkler) and received ample amount of water. Vines K3 and K4 were further from the irrigation system and received less but adequate water. Both pruning and irrigation treatments were done to increase between-vine variability in order to test the method.

The branching system of each vine was extensively described. All canes, laterals, and older shoots were tagged. Afterwards, some shoot variables were recorded: shoot number, bearing shoot number, and shoot age.

About 150 fruits were tagged and numbered on each vine. First, all the fruits of two randomly sampled laterals were chosen. This enabled us to estimate spatial dependences at short distance. Then, the sample was made up to 150 by random selection of fruits throughout the whole vine, to estimate spatial dependences at intermediate and large distances. During the growing season, some fruits were lost so that final sample sizes were 137, 140, 144, and 144, for vines K1–K4 respectively. At harvest, each sampled fruit was analysed for fresh weight (FW), dry matter (DM as %FW), soluble solids content

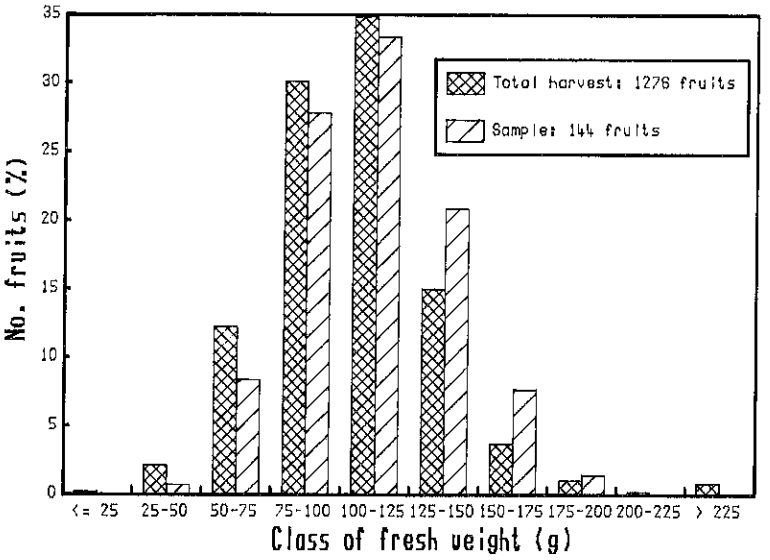

Fig. 1 Distribution of the number of fruits (%) of Vine K3 of the 1276 harvested fruits, and of the 144 sampled fruits.

(SSC) (ATAGO refractometer, expressed as °Brix), and titratable acidity (TA as ml NaOH 0.1*N*). For the vines K2, K3, K4, the fresh weight of all fruits was recorded.

RESULTS AND DISCUSSION

Data description

All vines were harvested on the same day. The total yield of vines K2, K3, K4 was respectively 80, 134, and 91 kg. The mean fruit weights were respectively 95, 105, and 107 g. Variation coefficients were c. 30%. The total number of fruits was strongly affected by the pruning treatment. For K2, K3, K4, the total number of fruits was respectively 842, 1276, and 852. Thus, the yield of vine K2 was less than the yield of vine K4 which had a similar number of fruits but less canes, and a higher mean fruit weight. The difference was probably related to difference in water supply.

Concerning quality characteristics, the largest variation between the four vines was in SSC. The mean values were respectively 8.8, 8.4, 7.2, and 6.9 °Brix for vines K1–K4. The intra-vine variation coefficients were c. 10%. Titratable acidity ranged between 14.0 and 14.3 ml NaOH (0.1*N*). The highest DM content was 18.5% for K1 and the lowest was 16.5% for K4.

Spatial analysis

The fruit weight distribution of the 150-fruit sample was compared with that of the total harvest from the vine for K3 (Fig. 1). There was good agreement except for the smallest and the largest values. The same pattern was found for vines K2 and K4. A problem was caused by the existence of a class of very large abnormal flat fruits having effects on variance estimation and therefore on estimation of the variogram function. For Vine K3, we made a computer simulation study (data not presented) testing different sample sizes by taking random samples without replacement of 25–200 fruits by Step 25 from the 1276 fruit weight file. The results indicated large fluctuations in the estimation of the variance of fruit weight until the sample size was 200. The situation was materially improved by discarding such abnormal fruit from further calculations. A simple way to do so was to exclude all fruits above 160 g, the maximal commercial size, and to limit the study to the population of commercial interest. Within this population, it can be verified that a sample size of 150 gives good results for variance estimation.

The variogram functions $G(h)$ of Vine K2 for SSC, TA, DM, and FW are shown in Fig. 2. The number of fruit pairs at each observed distance is shown in Fig. 3.

SSC, TA, and DM variograms exhibit a classical pattern. When fruits were borne on the same lateral ($h = 2$), the $G(h)$ values calculated from 168 fruit pairs (Fig. 3) were less than the sample variance (straight lines in Fig. 2) indicating a positive dependence (i.e., a positive correlation) between the paired fruits. It is to be noted that we do not take into account here the possible effect of the inflorescence

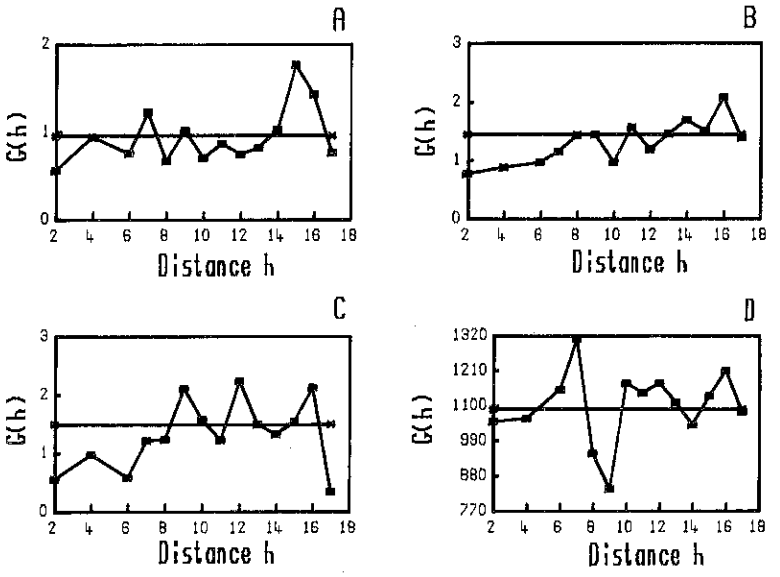

Fig. 2 Variogram functions $G(h)$ of Vine K2 estimated from the 150-fruit sample. **A**, SSC; **B**, TA; **C**, DM; **D**, FW. The *y*-axis is expressed as °Brix^2, ml^2, g^2, and (%)^2 respectively for **A–D**. Horizontal line indicates the total variance of the sample. Distance h is calculated as the number of active forks between the paired fruits.

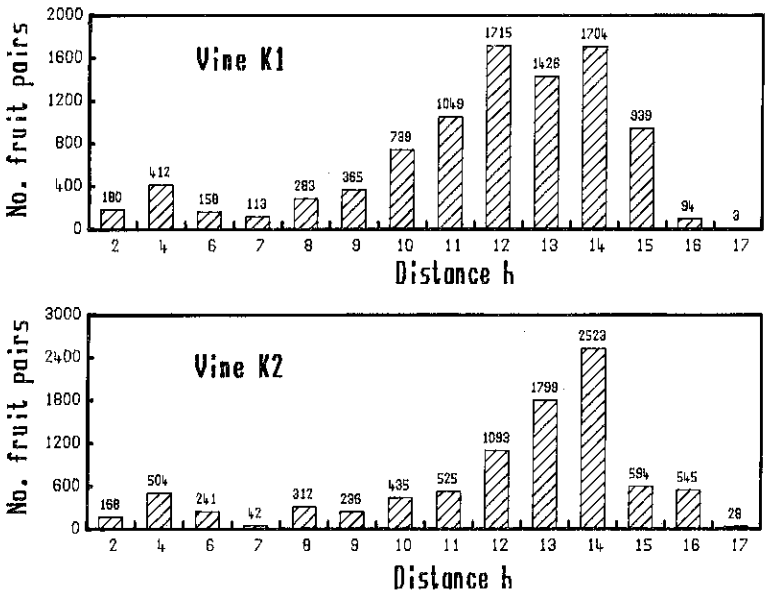

Fig. 3 Number of fruit pairs at given distances in the 150-fruit sample for Vines K1 and K2. Distance h is calculated as the number of active forks between the paired fruits.

type (i.e., single or multiple flower). In the set of paired fruits at $h = 2$, all fruits were considered equivalent, whatever the type of inflorescence. However, it has been shown elsewhere that it may have an effect on fruit characteristics (Tisné-Agostini et al. 1991). An easy refinement of the method would be to modify the distance, at $h = 2$, making subsets of fruit pairs as regards to the type of inflorescence. But it would be necessary to get larger samples at $h = 2$ to get significant subsample sizes.

Then $G(h)$ increases when h increases. The range (i.e., the distance at which the fruits can be considered independent) appears to depend on the variable under study. The range value is four (i.e., paired fruits borne on the same cane on two different laterals) for SSC, seven or eight for TA and DM. Note that in K2

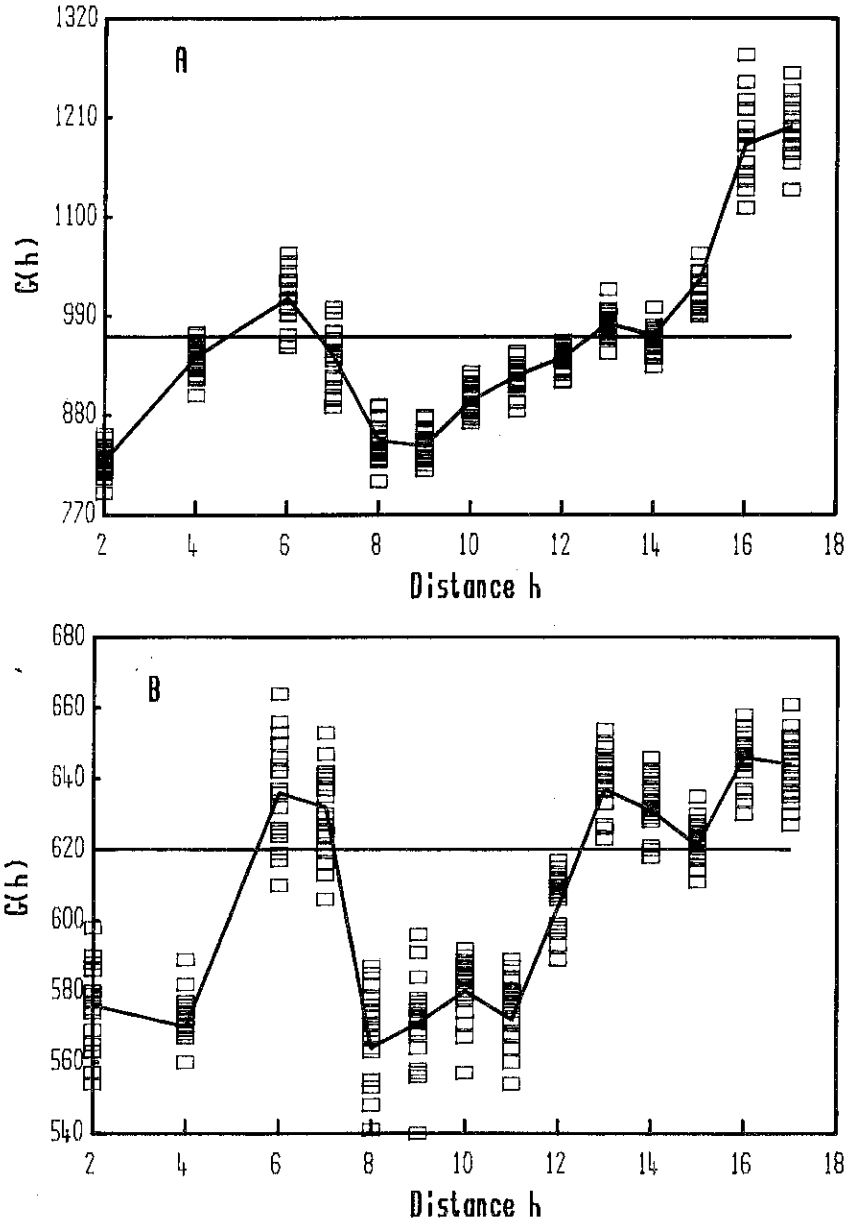

Fig. 4 Comparison between the fresh weight variogram function $G(h)$ of: A, Vine K2; and B, Vine K3—calculated from all harvested fruits (solid line)—and 20 variogram functions calculated from 150 randomly sampled fruits (rectangle). The y-axis is expressed as g^2. Horizontal line indicates the global variance of the vine. Distance h is calculated as the number of active forks between the paired fruits.

for $h = 7$ only 42 fruit pairs were found. It corresponds to a peculiar case. The comparison is of fruits borne on two different canes, the first one having a normal branching scheme, and the second one having arisen from a latent bud on a 2-year-old shoot. At higher h values, $G(h)$ exhibits fluctuations around the global variance. Thus, when paired fruits were taken on canes borne on two different main shoots separated by at least two forks ($h = 8$, corresponding to classical pruning), they can be considered independent whatever the variable under consideration. Although the variogram functions of the three other vines may be different (the value of the range may vary with vine and be shorter than $h = 8$), the same general conclusion applies and seems independent of the intensity of pruning.

The FW variogram exhibits a very different pattern. In particular, at large distances ($h = 8$ and $h = 9$) the $G(h)$ value decreases sharply indicating a strong dependence between the paired fruits at this

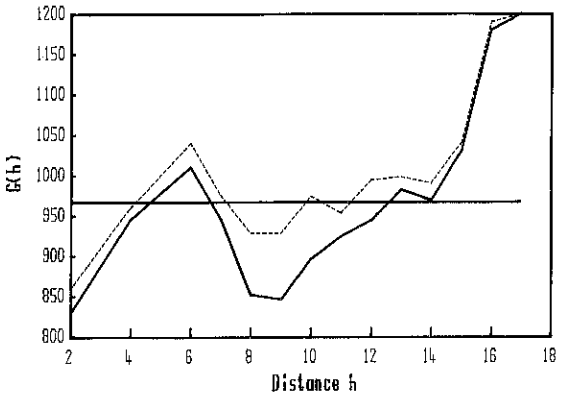

Fig. 5 Fresh weight variogram function $G(h)$ of Vine K2 calculated from all harvested fruits (solid line) and after excluding the fruits borne by canes derived from a common 2-year-older main shoot (dashed line). The y-axis is expressed as g^2. Horizontal line indicates the global variance of the vine. Distance h is calculated as the number of active forks between the paired fruits.

separation. The question arises if this pattern is related to the peculiar sample under consideration or if it is characteristic of the vine spatial structure.

This was tested by computer simulation studies for vines K2–K4. Twenty random samples of 150 fruits were made from the global file, and we calculated the new variogram functions. The results of Vine K2 are shown in Fig. 4a where the random sample values are compared to the global variogram calculated from all harvested fruits. It is clear from the simulation study that this decrease was related to the spatial structure of the vine, not inherent in fruit variability. The results of Vine K3 are shown in Fig. 4b. It may be noted that the same pattern was observed: a decrease of $G(h)$ at $h = 8$. A difference is that $G(h)$ has low values at $h = 10$ and $h = 11$. On the other hand, Vine K4 did not exhibit a similar pattern, there was no decrease in $G(h)$ at $h = 8$. This was confirmed by the simulation study (data not presented).

Comparing the structure of Vine K2 and Vine K3 to Vine K4, it appears that K2 and K3 had a typical structure at $h = 8$. Part of the paired fruits of Vine K2 and Vine K3 at $h = 8$ were borne symmetrically by canes derived from the same 2-year-old main shoot by normal pruning. For $h = 9$, the common main shoot was 3 years old, but one of the two canes came from a latent bud on a 2-year-old shoot. For $h = 10$ and $h = 11$, it was the same process. Such a structure was not found for the Vine K4 where the pruning was much more severe. When the variogram was calculated without these fruits, there was no more strong decrease of $G(h)$ at $h = 8$. This is shown in Fig. 5 for Vine K2. The same was observed for Vine K3. A possible explanation of this dependence at large distance may be the existence of a morphogenetical gradient. The canes bearing the paired fruits were derived from buds on the same basal position borne by pruned canes themselves borne by the same main shoot.

The results show that variogram functions can be satisfactorily estimated with c. 150 sampled fruits. In practice, only the network of shoots bearing the sampled fruits has to be described. The easiest way to do so is to describe the path from each lateral bearing a sampled fruit to the "trunk". This is enough to calculate the distance between any pair of sampled fruits.

Nevertheless, when one wants to verify the estimated variogram by comparing it with the variogram calculated from all borne fruits, as was done in this work, it may be interesting to test if a half vine may give a very good approximate of the global variogram function in order to save labour in measuring the fruits and describing the shoot network.

Using the fresh weight data file at harvest of Vines K2, K3, and K4, it was possible to study the effect of the fruit position on the estimation of the variogram. The fruits were separated into two sectors and variograms were calculated for each sector. North and south, or east and west sectors were compared (the vine lines are oriented north–south). Results for Vine K2 are shown in Fig. 6. Note that the scales of the y-axes are different. It is clear that a "north" or "south" subsample gives results in very good agreement with the variogram calculated from all data. The same result was found for Vines K3 and K4. The "east" subsample in Fig. 6 was quite different to the reference variogram, whereas the "west" variogram gave reasonable agreement. For Vines K3 and K4, the east and west subsamples were also less efficient in estimating the variogram function. This indicates that it is possible to estimate the variogram function from part of the vine as far as this part is representative of the vine structure.

CONCLUSION

This study has shown that geostatistic techniques can be applied to kiwifruit using "fork number" to describe between-fruit distance. The variogram function estimated from fruit samples is in good agreement with the expected pattern. The simulation study showed that a sample size of c. 150 fruits gave an

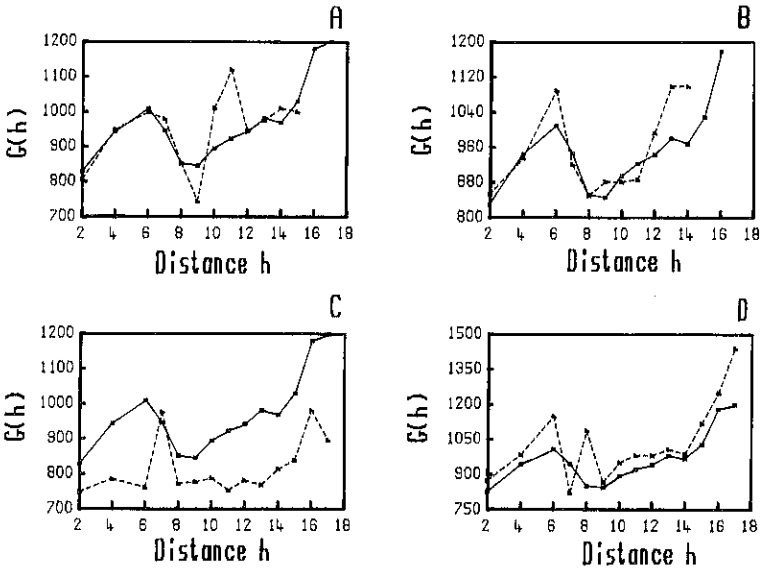

Fig. 6 Fresh weight variogram function $G(h)$ of Vine K2 calculated from all harvested fruits (full line) and from fruits taken in a geographical sector (dashed line). **A**, north; **B**, south; **C**, east; **D**, west. North variogram is to be compared to south, and east to west. The *y*-axis is expressed as g^2. Distance h is calculated as the number of active forks between the paired fruits.

accurate estimate of the actual variogram of fruit FW calculated from all harvested fruits. This means it should be possible to study the spatial dependence of fruits within the vine structure by sampling.

The method was applied here to fruit characteristics at harvest. It can also be used during the growing season when applied to variables such as fruit diameter, and we plan studies on rate of budbreak, flowering rate, and mean fruit weight. This offers prospects for studying more precisely some other important yield components.

It appears that as long as fruits are separated by at least eight forks, as when fruits are on different canes separated by at least two forks, they can be assumed to be independent, and classical statistics can then be used. This is not dependent on the number of fruits borne by the vine in the range 850–1300. This conclusion applies to SSC, TA, and DM content. It applies also to FW, with a restriction when the paired fruits were borne by canes derived from a common 2-year-older main shoot.

Finally, it was shown that when a vine is divided into sectors (e.g., north and south), a sector subsample can give a good estimation of the total variogram function, increasing research applications of this technique.

REFERENCES

Audergon, J. M.; Monestiez, P.; Habib, R. 1991: Sampling in a fruit tree: a new concept applied to an apricot tree. IX International Symposium on Apricot Culture. *Acta horticulturae 293*: 685–691.

Grant, J. A.; Ryugo, K. 1984: Influence of within-canopy shading on fruit size, shoot growth, and return bloom in kiwifruit. *Journal of the American Society of Horticultural Science 115*: 704–707.

Hopkirk, G.; Beever, D. J.; Triggs, C. M. 1986: Variation in soluble solids concentration in kiwifruit at harvest. *New Zealand journal of agricultural research 29*: 475–484.

Journel, A. G.; Huijbregts, Ch. J. 1978: Mining geostatistics. Academic Press. 600 p.

Matheron, G. 1965: Les variables régionalisées et leur estimation. Chap. XIII. Paris, Masson. 306 p.

Marini, R. P. 1985: Sample size estimates for peach tree growth and yield experiments. *Journal of the American Society of Horticultural Science 110*: 604–608.

Millier, C.; Nys C.; Ranger J. 1986: L'échantillonnage dans l'étude de la minéralo-masse forestière: l'exemple des taillis ardennais. *Annales des sciences forestières 43*: 165–178.

Monestiez, P.; Habib R.; Audergon J. M. 1989: Estimation de la covariance et du variogramme pour une fonction aléatoire à support arborescent: application à l'étude des arbres fruitiers. Pp. 39–56 *in*: Geostatistics, Vol. 1. Dordrecht, M. Armstrong *ed*. Kluwer Academic Publishers.

———1990: Geostatistics, spatial dependences in a tree: a new approach in fruit tree studies. *Acta horticulturae 276*: 257–263.

Snelgar, W. P.; Bayley, G. S.; Manson, P. J. 1988: Temperature studies on kiwifruit vines using relocatable greenhouses. *New Zealand journal of experimental agriculture 16*: 329–339.

Tisné–Agostini, D.; Vannière, M–P.; Habib, R. 1991: Influence du type d'inflorescence sur la qualité du kiwi. Pp. 59–67 *in*: 9ème Colloque sur les Recherches Fruitières Avignon INRA-CTIFL.