Canopy probabilistic reconstruction inferred from Monte Carlo point-intercept leaf sampling

Rita Giuliani, Eugenio Magnanini, Fabrizio N erozzi, Enrico Muzzi, Hervé Sinoquet

Dipartimento di Coltura Arborea, Università degli Studi di Bologna, Via Fanin 46, 40127 Bologna, Italy
ARPA, Servizio Meteorologico Regione Emilia-Romagna, Viale Silvani 22, 40122 Bologna, Italy
UMR PIAF, INRA-Université Blaise Pascal, 234 Avenue du Brézet, F-63039 Clermont-Ferrand, France

Received 13 January 2004; received in revised form 23 September 2004; accepted 23 September 2004

Abstract

The usefulness of a probabilistic technique to describe 3D foliage geometrical traits depends on the possibility of quickly and cheaply collecting a leaf data sample. The present simulation study was designed to provide the statistical information necessary to drive and control a foliage random sampling process on a real tree canopy (considered as a population of leaves) and to evaluate two corresponding statistical reconstruction outputs. A point-intercept leaf collection method was simulated on a single walnut-tree crown (7.20 m² surface area) by performing a Monte Carlo (MC) sampling from a data set obtained by digitising all 1558 leaves. Random (R) and adaptive Kernel (aK) methods were employed for foliage synthesis. Thirty canopies reconstructed from the same MC data sample (describing 250 leaves, i.e. less than 20% of crown leaf number) by both procedures indicated that the variability associated with each simulation process can be neglected, i.e. MC leaf sampling played the principle role in driving foliage reconstruction and deciding canopy geometrical traits. A crucial point of the procedure was to define a method to construct confidence envelopes for the artificial geometrical canopy parameters, given that for practical concerns in the field there is a need to sample the real canopy only once: 30 independent samples (describing 250–300 leaves) were generated by resampling with replacement (bootstrapping) one extracted leaf data sample (describing 300 leaves) and the corresponding canopies were simulated. The reconstructed canopies were characterised from a structural perspective in terms of foliage surface area, vertical leaf area density, single leaf area, and leaf angles. The synthesised canopies were evaluated from a functional perspective in terms of sunlit surface area projected orthogonal to sunbeam direction (silhouettes), and sky vault-integrated silhouette to (canopy) area ratio (STARSKY). The virtual walnut-tree foliage reproduced by electromagnetic digitising was the reference to corroborate all corresponding reconstructed canopies. The aK method appeared more accurate and precise than the corresponding R technique to represent real foliage geometrical traits. In general, the proposed aK statistical canopy

Abbreviations: Bootstrap, resampling with replacement; Contact or hit, sampled leaf; DOY, day of year; IQR, inter-quartile range (75–25th percentiles); LAD, leaf area density (m² m⁻³); LSCV, least squares cross validation technique; PDF, probability density function; Silhouette, canopy area projected orthogonal to sunbeam direction (m²); STAR, silhouette to (canopy) area ratio; STARSKY, sky vault-integrated STAR values

* Corresponding author. Tel.: +39 051 2096400; fax: +39 051 2096401.
E-mail address: rgiulian@agrsci.unibo.it (R. Giuliani).
reconstruction method appears to be promising to infer the geometrical features of a broad-leaf tree crown from a foliage sub-set, even if its applicability depends on the size of the considered tree, i.e. on the practicability of collecting a leaf data sample. © 2004 Elsevier B.V. All rights reserved.

Keywords: Adaptive Kernel methods; Bootstrap; Canopy modelling; Canopy rendering; Distribution-free statistics; *Juglans regia* L.

### 1. Introduction

Canopy geometry plays a major role in deciding foliage mass and energy exchange – whether herbaceous or shrubs or trees are considered – with the aerial environment. Thus, canopy’s geometrical traits are necessary inputs – whether the scale is single plant, stand or plant community – for eco-physiological models designed to describe processes involving radiation, photosynthesis and transpiration fluxes (Campbell and Norman, 1989; Parker, 1995).

When modelling plant canopy structure and/or functions it may be necessary to acquire both green and non-green vegetation elements (Pearcy and Yang, 1996; Valladares et al., 2002; Falster and Westoby, 2003; Weiss et al., 2004). Focusing on trees, the woody architecture is strictly involved in biophysical processes such as canopy transpiration, and when the woody fraction is a significant component of the canopy (Gower et al., 1999; Kucharik et al., 1999) it may sensibly affect foliage activities such as radiation flux distribution. Therefore, to model tree canopy structure and/or activities, both foliage geometry and woody architecture may be required as inputs. On the contrary, when the woody fraction can be considered as a negligible crown component in relation to the research aim, it can be reasonable to acquire only foliage geometrical data.

Unfortunately, in tree functional studies the parameterisation of canopy geometry represents a limitation, as much as a challenge: only a few indirect methods (remote sensing techniques, Wulder and Franklin, 2003) involving the use of a laser technique...
applications of point-quadrat methods have been employed on tree canopies and only recently enhanced and point-intercept sampling has not been widely for herbaceous cover extrapolation. By contrast, Wilson, 1959 methods, has been extensively adopted Madden, 1933 and modiﬁed point-intercept vegetation sampling (Bonham, 1989), especially using point-quadrat (Levy and Madden, 1933) and modiﬁed point-quadrat (Warren Wilson, 1959) methods, has been extensively adopted for herbaceous cover extrapolation. By contrast, (line- and) point-intercept sampling has not been widely employed on tree canopies and only recently enhanced applications of point-quadrat methods have been

exploited in orchards (Wünsche et al., 1997) and forest stands (Radke and Boltstad, 2001). In any case, the schematic approach adopted for point-intercept canopy sampling does not make it possible to produce a detailed (leaf by leaf) foliage reconstruction, and the convenience in its use for tree crown functional modelling is thus limited.

Lately, there has been much interest in point-intercept leaf techniques involving Monte Carlo methods (Hammerley and Handscomb, 1964), i.e. techniques enabling random and proﬁtable (capable of accounting for foliage spatial variability while not being too laborious) leaf sampling to infer foliage geometrical and functional traits (Kucharik et al., 1999; Ross and Ross, 1998). The encouraging results reported on fruit tree crowns (Poni et al., 1996; Succi et al., 1997) and forest stands (Reich et al., 1994; Lookingbill and Zavala, 2000; Nackaerts et al., 2000) make Monte Carlo sampling a particularly attractive approach for leaf data collection on different scales.

Given this context, the present study aimed to simulate Monte Carlo (MC) point-intercept leaf sampling on a single walnut tree (Juglans regia L.) by data extraction from a whole foliage data set (attained by leaf electromagnetic digitising, Sinoquet et al., 1997) and to perform foliage only reconstruction by random (R) and adaptive Kernel (aK) procedures. The latter process refers to a Kernel non-parametric method (Silverman, 1986) widely applied in spatial statistics (Cressie, 1991) and, while it has been lately exploited to recognise patterns that can simulate the aggregate (i.e. non-random) distribution of trees in a stand (Batista and Maguire, 1998), this is to our knowledge at least the ﬁrst time that an aK method has been used for foliage reconstruction.

Foliage syntheses by both stochastic procedures were compared with the corresponding digitised (i.e. real) canopy. Corroboration of foliage reconstruction – canopy stereological inference after Baddeley (1991) – included computation of such geometrical parameters as leaf blade dimensions and orientation angles, foliage extension area, vertical leaf area density proﬁles, and canopy fractions projected orthogonal to direct sunbeam radiation (silhouette). Also, a sky vault-integrated value of silhouette to (canopy) area ratio (STAR: Oker-Blom and Smolander, 1988; Stenberg et al., 1994; Farque et al., 2001) was computed as a canopy shape parameter.
(STARsky) associated with canopy light interception efficiency and used to corroborate the reconstructed crowns.

The originality of the present paper resides in:

1) simulation of MC sampling able to address random leaf data collection on real broad-leaf canopies;

2) possibility of inferring foliage geometry from a leaf data sample based on a few parametric statistical assumptions: in particular, an aK method relying on distribution-free statistics (Silverman, 1986);

3) Opportunity to corroborate reconstructed foliage and leaf features against the corresponding digitised (real) values.

2. Sample-based procedure for canopy simulation

It is accepted that a random data sample extracted from a population (of unknown distribution) can make it possible to estimate the parameters representative of the whole data set. A sample-based procedure for canopy simulation was thus developed under the assumption that a random sample of leaf geometrical data can provide sufficient statistical information to infer the foliage geometrical traits.

Given a foliage geometrical data set identifying all leaves of an individual tree, canopy simulation went through MC leaf data extraction and foliage synthesis by applying both random (R) and adaptive Kernel (aK) approaches. The software driving MC sampling and R synthesis was written with Mathcad 6.0 package (Mathsoft Inc., Cambridge, MA, USA) and the software driving aK reconstruction was written in Fortran 77. Canopy rendering and computations of geometrical parameters such as silhouette and STARsky were performed by VegeSTAR 3.0 software program (Adam et al., 2002).

2.1. Canopy data set

A geometrical data set acquired at PIAF-INRA, Clermont Ferrand, France, by electromagnetic digitising a single 2-year-old walnut tree (J. regia L.) canopy with 7.20 m² surface area was used for simulation (Fig. 1a). The foliage geometry had been acquired after Sinoquet et al. (1997); and Sinoquet and Rivet (1997): an electromagnetic 3D digitiser (3SPACE FASTRAK, Polhemus Inc., VT, USA) was used to identify all 1558 canopy leaves (leafletlets from a botanical perspective, because J. regia has composite leaves) considered as planar areas.

The leaf data set was organised in lines including blade length (l, cm) and maximum width (w, cm), Cartesian coordinates (x, y, z) of the blade insertion point, midrib azimuth and elevation angles (ψ and θ, respectively), and blade rolling angle around the midrib (φ). On the basis of the adopted digitising methodology, the X, Y axes were related to a horizontal flat surface and the Z axis was associated with tree height. While digitising, a random sample of 30 leaves was collected: blade l and w were measured by a ruler and leaf area (A_l, cm²) was assessed with a LI-3000 leaf area meter (LI-COR, Inc., Lincoln, NE). The collected leaf sample made it possible to identify the following allometric relationship

\[ A_l = 0.72lw \quad (R^2 = 0.98) \] (1)

that was adopted to compute the area of all digitised leaves and to choose a polygon representative of the leaf lamina necessary for canopy rendering.

A x = 1.88 m by y = 1.82 m rectangle able to contain the foliage projection, and a z = 3.00 m able to cover the vertical leaf distribution identified the smallest parallelepiped including the virtual canopy under study.

2.2. Leaf data extraction and computation of foliage surface area and leaf number

2.2.1. Monte Carlo sampling

Each digitised leaf lamina was modelled as an eight-sided flat polygon inscribed within an ellipse (l corresponded to the ellipse’s major axis, and w to the minor axis). The digitised leaf lamina angles (ψ, θ, and φ) and the spatial (x, y, z) coordinates of the lamina insertion point were assigned to each corresponding polygon so that its vertices could be spatially identified.

A random-seed instruction \( rdn \) (x) in Mathcad 6.0 was adopted to generate two sequences of pseudo-random numbers in \( 0 \leq x \leq 1.88 \) and \( 0 \leq y \leq 1.82 \) intervals, respectively; an X, Y pseudo-random point series \( (N_x) \) could then be identified in the rectangular
base area containing the foliage projection, all points having the properties of stationarity, i.e. uniformly distributed (Stoyan et al., 1987).

Leaf data extraction was simulating the measurements on a real canopy by only using a ruler and a compass. A Monte Carlo method (Hammerley and Handscomb, 1964) was applied as point-intercept technique to verify, for each X, Y monitored location, if one, more than one, or no virtual leaf polygonal area could be intercepted (contacts or hits, after Warren Wilson, 1959) on its vertical, for 0 ≤ z ≤ 3.00 m (Fig. 1b).

Each line of data acquired by leaf digitising and corresponding to each contact was automatically extracted from the whole data set, and the digitised x, y coordinates of lamina insertion were replaced by those of the associated X, Y monitored location. All selected and modified data lines were then stored separately to constitute a leaf data sample.

A fundamental step of the procedure was then the estimation of a convenient minimum number of hits to be identified to represent the canopy that passed through computation of the foliage surface area projected on the X, Y plane (A_c'), and its update while adding contacts. An unbiased estimate of A_c' on the X, Y plane was performed after Warren Wilson (1959) by applying the following formula

\[ A_c' = \left( \frac{\sum_{i=1}^{N} R_i}{N} \right) A_g \] (2)

where A_c' is the projected foliage surface area (m^2), A_g the flat monitored area (m^2), N the number of X, Y monitored points, and n the number of hits (Fig. 1c).

When negligible changes in A_c' standard deviation were observed (order ≤ 2–3% of A_c') while adding Y, Y monitored locations (Fig. 1d), that is when n/N (mean number of hits per monitored generated point) plotted against N showed almost a constant pattern (Fig. 1e), MC exploring was stopped (Fig. 1f).

Different series of points in the X, Y plane were generated and the MC technique was re-applied to identify a range of X, Y locations sufficient to sample the canopy. In general, the values of A_c' and n/N plotted against N started showing stable trends when approaching 100 monitored locations, and 100–120 monitored points detecting 270–320 digitised data lines (corresponding to 17–21% of canopy leaves) were considered as adequate for foliage reconstruction attempts.

2.2.2. Estimation of foliage extension area

A weighed mean of leaf elevation angle (⟨cos θ⟩) was estimated for the extracted canopy data sample in terms of

\[ \langle \cos \theta \rangle = \frac{\sum_{i=1}^{n} (A_i \cos \theta)_i}{\sum_{i=1}^{n} (A_i)_i} \] (3)

where A_i is the surface area of a single leaf, and n the number of detected hits. It has to be noted that A_i should be computed using the lamina inclination angle calculated from θ and ϕ after Farque et al. (2001): here, only the elevation angle (θ) has been used because the ϕ contribution appeared negligible.

The foliage surface area (A_c) was then estimated from the corresponding projected value (A_c', calculated by Eq. (2)), by the formula

\[ A_c = \frac{A_c'}{\langle \cos \theta \rangle} \] (4)

2.2.3. Estimation of artificial canopy leaf number

The mean single leaf area projected on the X, Y plane (⟨A'_i⟩) was computed by the formula

\[ \langle A'_i \rangle = \frac{1}{n} \sum_{i=1}^{n} 0.72(lw \cos \theta)_i \] (5)

where l is lamina length, w maximum width, θ the leaf elevation angle, and n the number of contacts. The number of leaves composing the synthesised canopy (L_c) was then calculated as

\[ L_c = \frac{A_c'}{\langle A'_i \rangle} \] (6)

where A_c' is the foliage area projected on the X, Y plane. Since A_c' and A'_i variations can be considered negligible in correspondence of 100–120 X, Y monitored locations, also L_c can be considered stable in that interval of points.

2.3. Canopy reconstruction methods

Foliage reconstruction comprised the estimate, on the basis of the canopy data sample, of leaf geometrical properties (such as coordinates of lamina insertion point, lamina dimensions and angles) to size and locate within the canopy volume all the virtual leaves calculated in Eq. (6) so as to assemble the foliage.
Fig. 1. Flow diagram of the whole sampling-reconstruction simulation procedure. (a) Two-year-old walnut tree photo taken at full canopy extension from an East–West horizontal perspective; (b) Corresponding rendered image obtained by digitising all leaves. The dots represent the ground random point series \( (N_r, \text{in the text}) \) used to perform Monte Carlo leaf data sampling; the dashed arrows indicate a few of the vertical lines monitored to detect the presence of artificial leaves (contacts or hits); (c) foliage surface area projected on the \( X, Y \) plane \( (A_0^c, \text{continuous line}) \) estimated via the extracted leaf data. The dotted line corresponds to the canopy surface area computed by the digitising data; (d) standard deviation (S.D.) of \( A_0^c \) in relation to the number of \( X, Y \) monitored points; (e) number of hits per \( X, Y \) point (dotted line) and corresponding S.D. (continuous line); (f) extracted leaf sample; (g) random (R) reconstruction procedure: \( N_r \) random point series in the \( X, Y \) plane is represented by open circles and \( M_r \) random point series is represented by filled circles. The dashed cylinder has the basal radius corresponding to the computed
2.3.1. Random simulation

The minimal inter-point distance (nearest-neighbour distance, after Stoyan et al., 1987) for each X, Y generated point was computed. As expected, the nearest-neighbour distance frequency (using classes of 0.030 m) showed a Poissonian distribution (data not shown) and the mean minimal inter-point distance ($\bar{d}_{\text{min}} = 0.220$ m) was picked as representative.

An X, Y random point series ($M_r$) was generated. Each x, y point part of $M_r$ was considered the centre of a circle on the X, Y plane with radius = $\bar{d}_{\text{min}}$: if at least three virtual leaves (contacts) of the canopy sample were present in the cylinder having that circle as base, one artificial leaf with lamina insertion at x, y coordinates was assumed (Fig. 1g). This process was reiterated until the x, y lamina insertion coordinates of a number of virtual leaves = $L_v$ were assigned.

For each virtual leaf, all the associated parameters of the lamina (z, l, w, and angles) were estimated from the corresponding statistics of the eight elements in the (virtual) canopy sample having the closest lamina insertion point in x, y coordinates. The number of eight-leaf data replicates resulted from a statistical requirement of having a sufficient number of observations and the need of bounding the canopy volume explored to take into account the local variability. A Gaussian algorithm for random number generation ($r = \text{rnorm}(n, m, \text{sd})$ instruction in Mathcad 6.0) was applied to assign each leaf variable value: required inputs for generating the random process were number of values to be extracted (n:one), mean (m) calculated over the eight leaf replicates, and corresponding standard deviation (S.D.).

A representative canopy synthesised by R method is displayed in Fig. 1(h).

2.3.2. Adaptive Kernel simulation

Referring to a distribution-free statistics, the Kernel method (Silverman, 1986) involves a density estimation, i.e. the construction from a discrete set of data points of a density function, called the probability density function (PDF) after Kimoto and Ghil (1993). The basic version of Kernel density estimation entails density “bumps” associated with the observed data points, i.e. a localised Kernel density function is assigned to each data point; the sum of these continuous functions yields an estimation of PDF at any point in a $d$-dimensional space. A Kernel function decides the shape of the kernels while a bandwidth parameter, which determines the radius of influence of each sampled point (Fig. 1j), controls the smoothness of the estimate of density function.

In the present study, an estimate of density function for univariate ($l$, $\psi$, $\theta$, and $\phi$) and multivariate ($x$, $y$, $z$ blade insertion point coordinates) leaf data was identified (Fig. 1ii). The application of a variable Kernel procedure (Silverman, 1986) made it possible for the smoothness of the kernels to vary from one point to one other. The fundamental mathematical steps leading to the adaptive algorithm identification are reported in Appendix 1.

The geometrical values of each of the virtual leaves ($L_v$) computed by Eq. (6) were randomly extracted from the corresponding PDF (Fig. 1j); foliage reconstruction was performed by placing all leaves in the canopy volume.

A representative canopy synthesised by the aK method is shown in Fig. 1k.

3. Statistical variability associated with the simulation procedure

The digitised foliage data set collected on the walnut tree made it possible to evaluate the variability generated by each reconstruction method, by the whole sampling-reconstruction procedure, and by one-leaf data sample.

3.1. Assessment of the canopy statistics variability generated by the reconstruction method (a) and hold by the sampling and reconstruction processes (b)

(a) Three random independent leaf data samples were extracted and each of them (1S: 250–300 contacts) was used to synthesise 30 canopies by both the R and aK methods.
(b) Thirty random independent leaf data samples were extracted (30S) and each of them was used to synthesise two canopies, one by R and one by aK.

The distribution of leaf dimensions, angles and spatial coordinates of insertion points of all (a) and (b) extracted samples and of the digitised (i.e. real) data set resulted belonging to the same population on the basis of Kolmogorov–Smirnov homogeneity test (\(P > 0.629\)).

In general, 30 reconstructions were considered the minimum threshold number making it possible to identify 25th and 75th percentiles associated with the simulated geometrical data (after Simon, 1997; Chernick, 1999; Polansky, 1999), given that the Mathcad 6.0 (to extract leaf data sample and reconstruct the canopy data set) and VegeSTAR 3.0 (to render and compute geometrical parameters) programs cannot exchange data files automatically.

For the digitised tree crown and each canopy simulated by both R and aK reconstruction methods, vertical leaf area density (LAD, \(\text{m}^2\text{m}^{-3}\)) was computed by 0.25 m high step, whereas silhouettes were simulated at hourly steps during summer days at 45°N latitude. Also, STARSKY was computed after Sonohat et al. (2002) from the integration of 42 directional STAR values corresponding to a sample of 42 directions of the sky vault weighed according to the standard overcast conditions (SOC: Moon and Spencer, 1942) radiance distribution.

A Kolmogorov–Smirnov homogeneity test was applied to assess whether simulated (R and aK) and digitised leaf data in terms of dimension, angles and lamina insertion \(x, y, z\) coordinates were part, as desirable, of the same population: for the aK reconstruction, the homogeneity was accepted for all considered parameters (\(P > 0.527\)) but, by contrast, for the R simulations the homogeneity test was rejected for leaf blade azimuth (\(P > 0.032\)) and \(z\) values (\(P > 0.041\)).

The normal distribution of structural reconstructed canopy parameters such as \(L_c, A_c, \text{LAD, silhouettes, and STARSKY}\) was assessed by the Shapiro–Wilks test for both canopy reconstruction processes: for both R and aK data normality was accepted for all variables. However, distributions of foliage and single leaf statistics simulated by both the R and aK methods and the corresponding digitised canopy data were analysed in a non-parametric way: central tendency (mean and median), dispersion (minimum, maximum and inter-quartile range, IQR), and shape (skewness and kurtosis) parameters were computed. In particular, even if 30 reconstructed canopies output geometrical parameters such as foliage surface area, silhouette and STARSKY meeting the normality statistical requirements, the number of data replicates was not considered sufficient to construct confidence intervals (after Simon, 1997, Chernick, 1999; Polansky, 1999), so only IQR ranges were calculated.

The most probable R and aK canopies were identified among the 30 corresponding reconstructed canopies by minimising the deviations between simulated \(A_c\), vertical LAD and STARSKY normalised medians and the corresponding values of the digitised tree crown.

### 3.2. Assessment of the canopy statistics variability generated by one extracted canopy data sample

One leaf data sample (1S\textsubscript{30boot}) composed by 300 lines associated with 120 monitored random points was extracted by the MC method and used to generate 30 independent sub-samples by re-sampling with replacement technique referring to a bootstrap method (Efron and Tibshirani, 1993). More precisely, to generate each sub-sample, 100 \(x, y\) points were bootstrapped and the data of the associated (250–300) contacts were automatically selected.

The distributions of leaf lamina dimensions, angles and spatial coordinates of insertion points in the 30 sub-samples and in the digitised data set belonged to the same population on the basis of the Kolmogorov–Smirnov homogeneity test (\(P > 0.483\)). Each sub-sample was used to synthesize two canopies, one by each reconstruction process.

The sub-sample generation was required to match the feasibility of monitoring only once a real canopy with the requisite to dispose of reconstructed canopy replicates for statistical analysis: again, the number of 30 sub-samples was a compromise between the need to dispose of a sufficient number of observations for percentile computations and the inability of Mathcad 6.0 and VegeSTAR 3.0 programs to share data files. Non-parametric analysis was performed on all foliage and leaf variables and IQR were computed. The most probable R and aK canopies were identified among the
30 corresponding reconstructed canopies by minimising the deviations between simulated $A_c$, vertical LAD and $\text{STAR}_{\text{SKY}}$ normalised medians and the corresponding values of the digitised tree crown.

4. Results

4.1. Assessment of the canopy statistics variability generated by the canopy reconstruction method (a) and hold by the sampling and reconstruction processes (b)

(a) The statistics summarising 30 foliage and leaf parameter distributions of synthesised canopies (30R and 30aK) obtained from the same leaf data sample (1S) indicated, in general, how the reconstruction error associated with each adopted method tended to be constant and negligible. For example, $0.02 \leq \text{IQR} \leq 0.04 \text{ m}^2$ for foliage surface area and $15 \times 10^{-3} \leq \text{IQR} \leq 23 \times 10^{-3} \text{ cm}^2$ for single leaf area were output by R reconstructed canopies. The aK reconstruction proved even more precise than the corresponding R simulations (data not reported).

(b) The statistics summarising foliage and leaf parameter distributions of the synthesised canopies (30R and 30aK) obtained from 30 leaf data samples (30S) and of the digitised tree crown are reported in Table 1. Note that median foliage surface areas of 7.06 m$^2$ and 7.53 m$^2$ were found for the R and aK reconstructed canopies, respectively, against 7.20 m$^2$ of the digitised tree crown. The median number of leaves was 1614 for all reconstructed R and aK canopies, against 1558 leaves digitised in the walnut tree. The median $\text{STAR}_{\text{SKY}}$ was 0.26, 0.19 and 0.20 for R, aK and digitised canopies, respectively.

The median number of leaves was 1614 for all reconstructed R and aK canopies, against 1558 leaves digitised in the walnut tree. The median $\text{STAR}_{\text{SKY}}$ was 0.26, 0.19 and 0.20 for R, aK and digitised canopies, respectively.

Median single leaf surface area in R, aK and digitised canopies was 43.9, 44.9 and 44.0 cm$^2$, respectively. Comparable medians of all leaf angles were found for R, aK, and the digitised canopies: the median elevation angle for R reconstruction was 20° and 22° for both aK and digitised canopies.

In the digitised tree crown, vertical LAD tended to increase from the canopy bottom up to 1.50 m height, with a maximum of about 1.75 m$^2$ m$^{-3}$; a LAD of 0.50–0.75 m$^2$ m$^{-3}$ was then computed from 1.75 to 2.75 m foliage height, i.e. almost the canopy top. Vertical LAD distribution in the R simulated canopies appeared underestimated in the lower and upper foliage fractions and overestimated in the central canopy portion compared with the corresponding distribution in the digitised tree crown. A maximum median value of 2.25 m$^2$ m$^{-3}$ at 1.50 m canopy height was observed against a corresponding value of 1.75 m$^2$ m$^{-3}$ in the real canopy (Fig. 2a). By contrast, medians of the vertical LAD profile in aK and digitised canopies almost overlapped (Fig. 2b).

Daily silhouettes of digitised and reconstructed canopies tended to show a bimodal behaviour: the results obtained at DOY 200 (19 July) are displayed as an example. For the digitised tree crown a relative maximum of 1.50 m$^2$ was computed at noon, a minimum of 1.45 m$^2$ was shown at 14:00 h, and an absolute maximum of 1.55 m$^2$ was simulated at 16:00 h. By contrast, the corresponding silhouette medians of R synthesised canopies tended to increase from morning to afternoon, with a maximum value of 1.60 m$^2$ at 16:00 h, and generally showed higher values (around 5%) compared with the digitised crown (Fig. 3a). Silhouette medians of the aK canopies tended to approach the corresponding values of the real tree crown in the morning and in the late afternoon, whereas a more appreciable decrease with respect to the digitised canopy was observed from 13:00 to 16:00 h, with a minimum of 1.35 m$^2$ at 14:00 h (Fig. 3b). The most probable R and aK artificial tree crowns identified among the 30 corresponding reconstructed canopies are shown in Fig. 1h and k, respectively.

4.2. Canopy statistics variability generated by one extracted canopy sample

The statistics summarising canopy and leaf parameter distributions of R reconstructed canopies (30R) obtained from the 30 bootstrapped sub-samples and of the digitised tree crown are reported in Table 1. The aK canopies were more precise and accurate than the corresponding R reconstruction, i.e. they more closely approached the digitised canopy and leaf geometrical traits (data not shown).

Medians of 7.51 m$^2$ for the foliage area, 1666 for leaf number, and 0.24 for $\text{STAR}_{\text{SKY}}$ were associated...
Table 1
Leaf and foliage statistical parameters for reconstructed (R = Random; aK = adaptive Kernel) and digitised canopies

<table>
<thead>
<tr>
<th>Level</th>
<th>Parameter</th>
<th>Reconstructions</th>
<th>Number of observations</th>
<th>Mean</th>
<th>Median</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Range</th>
<th>25th</th>
<th>75th</th>
<th>IQR</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canopy</td>
<td>Foliage surface area (m$^2$)</td>
<td>Digitised (*)</td>
<td>1</td>
<td>7.20</td>
<td></td>
<td></td>
<td></td>
<td>9.00</td>
<td>3.30</td>
<td>6.86</td>
<td>7.83</td>
<td>0.97</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>30</td>
<td>7.22</td>
<td>7.06</td>
<td>5.70</td>
<td>9.00</td>
<td>3.30</td>
<td>6.86</td>
<td>7.83</td>
<td>0.97</td>
<td>0.31</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>30</td>
<td>7.51</td>
<td>7.53</td>
<td>5.78</td>
<td>8.86</td>
<td>3.08</td>
<td>7.34</td>
<td>7.64</td>
<td>0.30</td>
<td>–0.57</td>
<td>–0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>30</td>
<td>7.57</td>
<td>7.51</td>
<td>6.79</td>
<td>7.92</td>
<td>1.13</td>
<td>7.50</td>
<td>7.75</td>
<td>0.25</td>
<td>–1.49</td>
<td>2.72</td>
</tr>
<tr>
<td>Leaf</td>
<td></td>
<td>Digitised (*)</td>
<td>1</td>
<td>1641</td>
<td>1614</td>
<td>1248</td>
<td>2043</td>
<td>795</td>
<td>1550</td>
<td>1760</td>
<td>210</td>
<td>0.14</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>30</td>
<td>1558</td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td>0.14</td>
<td>0.22</td>
<td>0.25</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>30</td>
<td>1641</td>
<td>1614</td>
<td>1248</td>
<td>2043</td>
<td>795</td>
<td>1550</td>
<td>1760</td>
<td>210</td>
<td>0.14</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>30</td>
<td>1558</td>
<td></td>
<td></td>
<td></td>
<td>210</td>
<td>0.14</td>
<td>0.22</td>
<td>0.25</td>
<td>0.20</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>STARSKY</td>
<td>Digitised (*)</td>
<td>1</td>
<td>40.2</td>
<td></td>
<td></td>
<td></td>
<td>60.8</td>
<td>31.7</td>
<td>0.4</td>
<td>–0.3</td>
<td>1.18</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>30</td>
<td>44.0</td>
<td>43.9</td>
<td>0.1</td>
<td>80.6</td>
<td>80.5</td>
<td>37.5</td>
<td>50.5</td>
<td>0.02</td>
<td>0.11</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>30</td>
<td>44.4</td>
<td>44.9</td>
<td>0.3</td>
<td>122.9</td>
<td>122.6</td>
<td>28.8</td>
<td>61.3</td>
<td>0.5</td>
<td>–0.2</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>30</td>
<td>44.4</td>
<td>44.9</td>
<td>0.3</td>
<td>122.9</td>
<td>122.6</td>
<td>28.8</td>
<td>61.3</td>
<td>0.5</td>
<td>–0.2</td>
<td>0.11</td>
</tr>
<tr>
<td>Leaf</td>
<td>Blade surface area (cm$^2$)</td>
<td>Digitised</td>
<td>1588</td>
<td>46.2</td>
<td>44.0</td>
<td>1.1</td>
<td>119.2</td>
<td>118.1</td>
<td>29.1</td>
<td>60.8</td>
<td>31.7</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>49230</td>
<td>44.0</td>
<td>43.9</td>
<td>0.1</td>
<td>80.6</td>
<td>80.5</td>
<td>37.5</td>
<td>50.5</td>
<td>0.02</td>
<td>0.11</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>49230</td>
<td>44.4</td>
<td>44.9</td>
<td>0.3</td>
<td>122.9</td>
<td>122.6</td>
<td>28.8</td>
<td>61.3</td>
<td>0.5</td>
<td>–0.2</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>49230</td>
<td>44.4</td>
<td>44.9</td>
<td>0.3</td>
<td>122.9</td>
<td>122.6</td>
<td>28.8</td>
<td>61.3</td>
<td>0.5</td>
<td>–0.2</td>
<td>0.11</td>
</tr>
<tr>
<td>(+)</td>
<td>Azimuth angle (°)</td>
<td>Digitised</td>
<td>1588</td>
<td>–3</td>
<td>–4</td>
<td>–180</td>
<td>180</td>
<td>360</td>
<td>–95</td>
<td>89</td>
<td>184</td>
<td>0.04</td>
<td>–1.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>49230</td>
<td>–1</td>
<td>–2</td>
<td>–179</td>
<td>180</td>
<td>359</td>
<td>–50</td>
<td>48</td>
<td>98</td>
<td>0.06</td>
<td>–0.62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>49230</td>
<td>–5</td>
<td>–5</td>
<td>–180</td>
<td>180</td>
<td>360</td>
<td>–95</td>
<td>81</td>
<td>167</td>
<td>0.05</td>
<td>–1.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>49230</td>
<td>–4</td>
<td>–4</td>
<td>–180</td>
<td>180</td>
<td>360</td>
<td>–55</td>
<td>45</td>
<td>104</td>
<td>0.13</td>
<td>–0.52</td>
</tr>
<tr>
<td>(+)</td>
<td>Elevation angle (°)</td>
<td>Digitised</td>
<td>1588</td>
<td>20</td>
<td>22</td>
<td>–70</td>
<td>79</td>
<td>149</td>
<td>10</td>
<td>33</td>
<td>23</td>
<td>–0.91</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>49230</td>
<td>19</td>
<td>20</td>
<td>–44</td>
<td>66</td>
<td>110</td>
<td>14</td>
<td>26</td>
<td>12</td>
<td>–0.40</td>
<td>1.37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>49230</td>
<td>19</td>
<td>22</td>
<td>–67</td>
<td>84</td>
<td>151</td>
<td>9</td>
<td>34</td>
<td>25</td>
<td>–0.78</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49860</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td>–32</td>
<td>75</td>
<td>107</td>
<td>15</td>
<td>28</td>
<td>13</td>
<td>–0.08</td>
<td>0.49</td>
</tr>
<tr>
<td>(+)</td>
<td>Rolling angle (°)</td>
<td>Digitised</td>
<td>1588</td>
<td>1</td>
<td>0</td>
<td>–108</td>
<td>96</td>
<td>204</td>
<td>–17</td>
<td>19</td>
<td>36</td>
<td>0.13</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30R</td>
<td>49230</td>
<td>1</td>
<td>0</td>
<td>–54</td>
<td>63</td>
<td>117</td>
<td>–8</td>
<td>9</td>
<td>17</td>
<td>0.20</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>30S → 30aK</td>
<td>49230</td>
<td>0</td>
<td>–1</td>
<td>–120</td>
<td>104</td>
<td>224</td>
<td>–20</td>
<td>19</td>
<td>39</td>
<td>0.01</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lS$_{30boot}$ → 30R</td>
<td>49230</td>
<td>–1</td>
<td>–1</td>
<td>–66</td>
<td>59</td>
<td>125</td>
<td>–10</td>
<td>8</td>
<td>18</td>
<td>–0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

(+) Angles are considered as deviations from 0. (*) = single value; IQR = Interquartile range = 75th–25th percentiles; 30S = 30 canopy data samples; 30R and 30aK = 30 random and 30 adaptive Kernel reconstructions; lS$_{30boot}$ = one canopy data sample bootstrapped 30 times.
with R reconstructed canopies against corresponding medians of 7.20 m², 1558 and 0.20 in the digitised tree crown. Median single-leaf area of R reconstructed crowns was 45.6 cm² against 44.0 cm² of the digitised canopy. Comparable medians of all lamina angles were found for R synthesised and digitised canopies: the median elevation angle was 22° for R reconstruction against the 21° for the digitised canopy.

The most probable R and aK artificial tree crowns identified among the 30 corresponding reconstructed canopies appeared similar to those reported in Fig. 1h and 1k, respectively.

5. Discussion

5.1. Monte Carlo sampling

The random leaf data sampling performed in the present study was affected by an unavoidable distortion since leaves composing a tree canopy are usually clumped and not randomly distributed and, in particular, because a walnut tree has composite leaves; then, too, the foliage area projected on the X, Y plane (A₀) was the only parameter adopted to decide the sample size. However, the collected canopy data

Fig. 2. Vertical leaf surface area density (—) computed for 30 random (a) and 30 adaptive Kernel (b) reconstructed canopies; (…) represents the 25th and 75th percentiles, and (-----) represents the digitised canopy values.

Fig. 3. Diurnal variation of silhouettes for 19 July at a latitude of 45°N (—) simulated for 30 Random (a) and 30 adaptive Kernel (b) reconstructed canopies; (…) represents the 25th and 75th percentiles, and (-----) represents the digitised canopy values.
samples made possible satisfactory probabilistic canopy reconstruction (aK artificial canopies, in particular), demonstrating that the data samples were able to represent the canopy leaf data population.

The MC sampling technique did not require any statistical assumptions on canopy data distribution, and a statistical control was applied while extracting the leaf data to optimise data sample size. Note that dispersion measures (Fig. 1c–e) were computed to decide when to stop canopy data sampling. Neither number of generated random points nor sample size was fixed a priori, and this absence of constraints differentiated the adopted approach from a foliage MC sampling performed by others (Reich et al., 1994).

In the present study, a number of 100–120 generated random points generally appeared sufficient to collect an adequate leaf data sample (involving 20% as maximum of leaves composing the walnut tree canopy). In preliminary experiments, a comparable number of monitored points was considered sufficient to sample kiwifruit canes with 10,000–15,000 leaves (Succi et al., 1997), i.e. with a foliage extension almost 10 times bigger than that of the walnut tree considered here: a cane reconstruction was attempted, but only a visual comparison of the canopy picture with the corresponding rendered image (as indirect testing of the performed leaf sampling) was possible.

The canopy data extraction carried out in the present study simulated the work necessary to collect a random leaf sample on a real broad leaf tree without any technical support other than a ruler (for leaf $l$, $w$ and $z$ measurements) and a compass (for azimuth, elevation and rolling leaf angle assessment). This is the reason why the same $x$, $y$ coordinates of a generated monitored point were assigned to all (leaf) contacts intercepted on its vertical axis: in the field it would be excessively time-consuming, in relation to the gained canopy reconstruction precision, to measure the real $x$, $y$ lamina insertion coordinates for each contact. By contrast, if an electromagnetic digitiser is available, the real spatial coordinates of the blade insertions could be automatically acquired.

Although from a theoretical perspective the suggested MC method could be applied to all tree crowns, practically the proposed sampling approach has remote application to big canopies and tree crowns with small leaf size, as those of coniferous species. Also, windy conditions compromise leaf measurements.

Note, however, that different practical solutions could be adopted for leaf sampling, depending on canopy size, shape, leaf spatial dispersion and available equipment. In the present investigation the $X$, $Y$ plane represented the horizontal field ground, but in different experimental situations it could simulate any other oriented surfaces, e.g. the vertical one.

A canopy sampling could also be performed by directly hitting the leaves, i.e. without the need to generate random points and to apply an MC point-intercept technique, to acquire leaf geometrical traits (3D coordinates of blade insertion point, blade dimensions and angles). The use of an electromagnetic digitiser to this purpose would be essential, and by digitising one leaf over five (or six or seven) the number of leaves composing the canopy could be computed.

A further application of an MC method under evaluation involves the sampling of tree branches and shoots instead of leaves: the woody geometry could be reconstructed and leaves placed on the axis on the basis of phyllotaxis rules (Sonohat et al., 2004), whereas leaf geometrical traits could be derived by experimental measurements. An MC process to sample axis (though still to be defined) could be theoretically applied even to herbaceous stands for single-stem identifications.

5.2. Canopy reconstruction procedures

In the present investigation, canopy reconstruction was achieved by two alternative stochastic procedures; one of them, based on the aK method, made no assumptions about the 3D canopy geometry under study. By contrast, two statistical assumptions were made in the R reconstruction process: the first was the adoption of a mean minimal inter-point distance among the generated random points (empirically estimated as indicated in Section 2.3.1) and the second was a Gaussian probabilistic model used to assign leaf parameters. The normal model employed can be considered as mainly responsible for the bias observed between R reconstructed and real canopies regarding leaf azimuth angles and vertical leaf area distributions. More precisely, in the R reconstructed tree crowns the azimuth leaf angle distribution showed a central tendency instead of a hyper geometric pattern exhibited by the digitised canopy (data not shown),
and the vertical distribution of leaves was centred around 1.50 m height. For vertical leaf assignment, other distributions such as Weibull, $\chi^2$ and $\beta$- could have been chosen, even if their modelling convenience seems to be debatable (Massman, 1982). By contrast, the frequency distribution patterns of all leaf geometrical parameters in the aK reconstructed canopies reflected the corresponding statistical patterns in the digitised tree crown (data not shown). This because the aK reconstruction procedure treated simultaneously the co-relations among leaf variables, whereas an R reconstruction method can process only one leaf geometrical variable at a time.

In the present study, the variability associated with the whole sampling-reconstruction process and to the reconstruction method were estimated by only 30 canopy replicates; this means that the computing of percentile-type confidence intervals for artificial leaves and foliage parameters was not possible because a higher number of simulated canopies would have been required.

Even if both statistical procedures acceptably simulated geometrical parameters such as foliage surface area ($A_c$) and single-leaf surface area ($A_l$), the aK approach output more accurate geometrical features, such as vertical LAD, silhouette and STARSKY: these traits are recognised as playing a main role in energy and mass exchange between canopy and atmosphere. Note in this connection that the higher silhouettes and STARSKY exhibited by R reconstructed canopies as compared to the corresponding aK simulated tree crowns (see Fig. 1h compared to Fig. 1k). The two statistical assumptions made in the R reconstruction process, that is the technique employed to reveal the presence of leaves in the canopy space and the normal probabilistic model chosen to place them along the vertical axis, have to be considered as mainly responsible for this bias. The possibility of including a local leaf density analysis in the R reconstruction procedure should improve accuracy and precision of the R synthesised canopies.

5.3. Canopy variability generated by one extracted canopy sample

It has been shown how foliage surface area ($A_c$), leaf number ($L_c$) and STARSKY of 30 R canopies generated from one bootstrapped canopy sample (1S$_{30\text{boot}}$ $\rightarrow$ 30R) – assuming that only one sample can be extracted on the real canopy – can include only about 1/3 of the variability associated with the whole reconstruction procedure. More precisely, IQR = 0.26 m$^2$ against IQR = 0.97 m$^2$ was found for $A_c$, IQR = 79 against IQR = 210 and IQR = 0.01 m$^2$ against IQR = 0.03 m$^2$ were found for $L_c$ and STARSKY, respectively. By contrast, the corresponding variability of leaf parameters such as single-leaf surface area ($A_l$) and angles resulted comparable to that of the whole reconstruction procedure. Canopy and leaf traits of 30 corresponding aK synthesised canopies were found to explore a smaller fraction of total variability than the 30 “bootstrapped” R canopies (data not reported). On the basis of this statistical behaviour, and taking into consideration that the reconstruction error can be neglected and that the synthesised parameters tend to exhibit normal distributions, the probability of randomly extracting a sample on the real canopy that is capable of representing the leaf population should be subordinate to a gaussian probability model.

For future investigations, a much higher number (order of thousands) of bootstrapped samples, i.e. of synthesised canopy replicates, from the only one-leaf data sample is desirable so as to define stable confidence limits for all reconstructed parameters. Because in the present investigation the limitation in the number of bootstrapped leaf data samples (i.e. reconstructed canopies) was related to the fact that Mathcad 6.0 and VegeSTAR 3.0 programs did not share files automatically, this software limitation needs to be overcome before future applications of our canopy probabilistic reconstruction method(s). Then, too, in a future study, canopy CO$_2$ and H$_2$O exchange activity should be simulated on both synthesised and reference (digitised) canopies to evaluate the reconstruction outputs from a functional perspective (Carboni, 1998).

6. Conclusions

The results of the present simulation study are promising and indicate the potential of MC canopy sampling combined with R and aK methods for foliage probabilistic reconstruction. The opportunity to
acquire the foliage geometry quickly and inexpensively makes the proposed MC sampling, especially when combined with aK canopy simulation, a convenient method to be exploited in eco-physiological studies aimed, for example, at describing and explaining the interaction between foliage and atmosphere, and/or whole-plant productivity.

Alterations of MC leaf sampling and reconstruction methods have to be conceived to make the whole procedure even more accurate and/or to extend its application opportunities to several canopy architectures. However, stochastic methods should receive increasing application in the future as a means of detecting canopy patterns using a minimum of measurements and, in particular, a minimum of human intervention.

Acknowledgements

We thank Prof. Rodolfo Rosa (Dipartimento di Statistica, University of Bologna) for his expertise on Monte Carlo methods, and Prof. Jan Goudriaan for reading the manuscript and offering advice.

Appendix A

Suppose that we have a set of observed data points assumed to be a sample from an unknown probability density function. Density estimation is the construction of an estimate of the probability density function from the observed data. A non-parametric Kernel method can be applied to identify, for univariate and multivariate sampled data, an estimate of density function from local densities (kernels) of data (Silverman, 1986).

Given this context, an adaptive Kernel method was here applied to construct a Kernel estimate of univariate (lamina ψ, or θ, or φ) and multivariate (l, w, x, y, z lamina insertion coordinates) sampled leaf data consisting of kernels associated with the observed data points and to allow the smoothness of the kernels to vary from one point to another.

The adaptive algorithms were constructed after Kimoto and Ghil (1993), with reference to Silverman (1986), by assuming the data points to be in a three-dimensional space; the symbol f is used to denote whatever density estimator is being considered.

A three-step procedure was followed:

A.1. Step 1: Find a first estimate of density function

A first estimate was identified to get a rough idea of the density; this estimate yielded a pattern of bandwidths corresponding to the various observations (choosing small smoothing parameters makes fine structure visible, while large parameters obscure the nature of the distribution), and these bandwidths were used to construct the adaptive estimator itself. Note, however, that the adaptive algorithm is relatively insensitive to pilot estimate parameters, meaning that there is no need for the pilot estimate to have any particular smoothness properties.

A local density function \( f_n(x) \) was defined by

\[
 f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta(x - X_i) \tag{A.1}
\]

where \( n \) is the number of observations (hits, in our study), \( x \) the scalar or tri-dimensional vector of leaf variables, \( X \), the vector designating the \( i \)th point, and \( \delta \) Dirac’s delta function. Dirac’s delta function has, in particular, the property to be equal to the unity in correspondence of an observed data; otherwise it is zero.

The mathematical convolution of the local density function with a Kernel function \( K(x) \), i.e. a non-negative function normalised to the unity, outputs a Kernel estimator function \( \hat{f}(x) \):

\[
 \hat{f}(x) = \int f_n(y)K(x - y) \ dy = \frac{1}{n} \sum_{i=1}^{n} K(x - X_i) \tag{A.2}
\]

The Epanechnikov Kernel (Epanechnikov, 1969) was chosen as the Kernel function \( K(x) \):

\[
 K(x) = \begin{cases} 
 1 - x^t x & \text{if } x^t x < 1 \\
 0 & \text{otherwise} 
\end{cases} \tag{A.3}
\]

where \( t \) denotes transpose.

A \( h_p \) smoothing parameter was automatically chosen (\( h_p = 0.15 \)) and an \( r \) leaf-parameter vector size \( (r = 1 \) for univariate data; \( r = 3 \) for \( x, y, z \) multivariate data) was introduced into the Kernel estimator function, so that a pilot estimate \( \hat{f}_p(x) \) could be identified as

\[
 \hat{f}_p(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_p^r} K \left( \frac{x - X_i}{h_p} \right) \tag{A.4}
\]
A.2. Step 2: Define local bandwidth factors

A local bandwidth factor \( \eta_i \) was determined as

\[
\eta_i = \left( \frac{\hat{f}_p(X_i)}{G} \right)^{-\alpha} \tag{A.5}
\]

where \( X_i \) is the vector designating the \( i \)th point, \( \alpha \) a sensitivity parameter satisfying \( 0 \leq \alpha \leq 1 \) (\( \alpha = 0.5 \), automatically set) and \( G \) the geometric mean of \( \hat{f}_p(X_i) \) determined by

\[
\log G = \frac{1}{n} \sum_{i=1}^{n} \log \hat{f}_p(X_i) \tag{A.6}
\]

A.3. Step 3: Compute the final estimate

The adaptive algorithm \( \hat{f}(x) \) could be identified as

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{(h\eta_i)^\alpha} K \left( \frac{x - X_i}{h\eta_i} \right) \tag{A.7}
\]

where \( h \) is the final smoothing parameter as determined by least-squares cross validation technique (LSCV). After Kimoto and Ghil, \( h \) was assessed by minimising the score function \( M_{\theta}(h) \) defined as

\[
M_{\theta}(h) = \int \hat{f}(x)^2 \, dx - \frac{2}{n} \sum_{i=1}^{n} \hat{f}_{i-1}(X_i) \tag{A.8}
\]

where the first right term is computed numerically and the second was the probability density estimate at \( X_i \) found by discarding the \( i \)th data point. The numerical computation of the integral was performed over a 3D grid for multivariate data (the value of the estimate of density function was computed at each node), or a regularly discretised segment for univariate data.

References


