*Journal of Ecology* 2001
**89**, 616–628

# Periodic spotted patterns in semi-arid vegetation explained by a propagation-inhibition model

P. COUTERON and O. LEJEUNE*

*Ecole Nationale du Génie Rural des Eaux et des Forêts/CNRS-FRE 2366, 648 rue J.F. Breton, 34033 Montpellier cédex 1, France; and *Collaborateur Scientifique FNRS, Faculté des Sciences, CP 231, Université Libre de Bruxelles, B-1050 Bruxelles, Belgium*

**Summary**

**1** Vegetation cover regularly punctuated by spots of bare soil is a frequent feature of certain semi-arid African landscapes, which are also characterized by banded vegetation patterns (i.e. tiger bush).

**2** The propagation-inhibition (PI) model suggests that a periodic pattern characterized by a dominant wavelength can theoretically establish itself through a Turing-like spatial instability depending only on a trade-off between facilitative and competitive interactions among plants. Under strictly isotropic conditions, spotted and banded patterns are distinct outcomes of a unique process, whereas anisotropy leads to a banded structure. The model predicts that spotted patterns will have a lower dominant wavelength than bands.

**3** We test some outcomes of the PI model against vegetation patterns observable in aerial photographs from West Africa. Two sites with rainfall of *c.* 500–600 mm year$^{-1}$ were studied: a 525-ha plain in north-west Burkina Faso and a 300-ha plateau in southern Niger. Digitized photographs were subjected to spectral analysis by Fourier transform in order to quantify vegetation patterns in terms of dominant wavelengths and orientations.

**4** Spotted vegetation proved highly periodic. The characteristic range of dominant wavelengths (30–50 m) was similar at two sites more than 500 km apart. The PI model suggests that spots may occur as a hexagonal lattice but there is little evidence of such patterning in the field. A dominant wavelength was far quicker to establish in simulations (*c.* $10^2$–$10^3$ years for annual grasses) than a hexagonal symmetry (*c.* $10^5$ years), and observed patterns are therefore likely to be far from the asymptotic structure.

**5** Elongated and smudged spots that locally became flexuous bands have been observed in southern Niger. This pattern that had a dominant wavelength of 50 m but lacked any dominant orientation can be interpreted as a transition from spots to bands under fairly isotropic conditions.

**6** The PI model provides a framework for further investigation of patterns in semi-arid vegetation and may be of a broader ecological application.

*Key-words*: facilitation/competition, Fourier analysis, tiger bush, Turing instability, West Africa

*Journal of Ecology* (2001) **89**, 616–628

## Introduction

Striking geometric patterns were revealed in the vegetation of extensive areas of arid or semi-arid Africa when aerial photographs became available in the early 1940s (Macfadyen 1950; Clos-Arceduc 1956). The most amazing, made up of stripes or arcs of vegetation alternating with bare ground, were given the name of *brousse tigrée* (tiger bush) by Clos-Arceduc (1956). Similar vegetation patterns have since been reported in several semi-arid regions of the African, American and Australian continents (White 1971; Cornet *et al.* 1988; Ludwig & Tongway 1995). Most suggested explanations for the origin of banded systems (White 1971; Greig-Smith 1979; Wilson & Agnew 1992) invoke water re-allocation from bare areas to vegetated bands through run-off, and

Correspondence: Pierre Couteron, ENGREF, 648 rue J.F. Breton, 34033 Montpellier cédex 1, France (tel. 33 4 67 04 71 26; fax 33 4 67 04 71 01; e-mail Couteron@engref.fr).

hence postulate that environmental anisotropy (usually a gentle slope) is necessary for establishment. As a consequence, models simulating the emergence of a periodic vegetation cover apply only to conditions marked by a slope-induced anisotropy (Mauchamp *et al*. 1994; Thiéry *et al*. 1995; Dunkerley 1997; Klausmeier 1999).

The propagation-inhibition (PI) model provides an alternative quantitative explanation of how stable periodic patterns may arise, based solely on the intrinsic dynamics of the vegetation (Lefever & Lejeune 1997). Periodicity results from an interplay between short-range facilitative interactions and long-range inhibitory interactions between plants (Lejeune *et al*. 1999). The PI model, like others, predicts that anisotropy will lead to development of a banded system (tiger bush) and will determine its orientation. However, it differs by predicting that a range of periodic patterns, such as spots of bare soil arranged on a hexagonal lattice or flexuous stripes of vegetation alternating with bare soil, with well-defined spatial wavelength, can emerge in the absence of any environmental anisotropy.

The concept of spatial phase transition, on which patterning in the PI model is based, has been recognized as crucial in determining large-scale emergent properties from numerous local interactions. It has found several applications in biological sciences (Hopfield 1982; Milne *et al*. 1996), as in generation of a sharp boundary from a smooth environmental gradient due to intra- and interspecific interactions (Yamamura 1976; Wilson & Nisbet 1997; Holt & Keit 2000). The PI model presents a pattern-generating instability (*sensu* Turing 1952; Glansdorff & Prigogine 1971) which is specific to vegetation dynamics, whereby minor variations of the vegetation cover are amplified and strongly contrasting regions of high and low phytomass develop even in strictly homogeneous and isotropic environments.

As far as we know, periodic patterns have never been explicitly reported for vegetation of arid or semi-arid environments devoid of a consistent source of anisotropy. However, vegetation punctuated by bare spots, referred to as 'termitaria-peppering' in Somalia (Macfadyen 1950) and as *brousse tachetée* (spotted bush) in West Africa (Clos-Arceduc 1956; Boudet 1972; Ambouta 1997), has been reported in some regions with tiger bush. Spotted vegetation sometimes occurs in the immediate vicinity of banded structures (e.g. plate 10 in Macfadyen 1950) but has received far less attention. Furthermore, several authors who have presented data on tiger bush did not report their observations about spotted vegetation (e.g. in Mexico, A. Cornet, personal communication). Spotted patterns may therefore be much more widely distributed than it would appear from the existing literature, and, as tiger bush, may have a world-wide distribution.

According to the PI model, banded and spotted patterns are distinct outcomes of a unique dynamic process, and its predictions can be tested against field observations of vegetation patterns. We present such a quantitative analysis of air photographs from semi-arid West Africa, with a main study site in north-west Burkina Faso and additional data from southern Niger. Although these landscapes include gentle slopes covered by tiger bush (Ambouta 1997; Couteron *et al*. 2000), we focus on areas devoid of a consistent slope. Such areas are characterized by savanna vegetation punctuated by spots of bare soil, a common physiognomy in the Gondo plain on the Burkina Faso-Mali border at 13°30′–14°30′ N (Boudet 1972; Couteron & Kokou 1997). Spotted savanna is also a dominant feature under the similar climatic conditions (L'Hôte & Mahé 1996) of the plateaux in Niger below 13° N (Clos-Arceduc 1956; Ambouta 1997).

We aim to quantify the spotted pattern in terms of dominant wavelength and orientation, and to document its relationship with tiger bush. Our main purpose has been to provide objective descriptions that can be tested against the outcomes of the PI model. Digitized air photographs provided an interesting trade-off between spatial extent of the study area and resolution of the data. Numerous methods are available for quantitative image analysis (Haralick 1979), and spectral analysis by Fourier transform (Niblack 1986; Mugglestone & Renshaw 1998) was selected as periodicity was apparent from both preliminary visual appraisal and modelling outcomes.

## Materials and methods

### BRIEF DESCRIPTION OF THE PI MODEL

As the PI model has been extensively analysed (Lefever & Lejeune 1997; Lejeune & Tlidi 1999; Lejeune *et al*. 1999; Lefever *et al*. 2000), only broad outlines are provided here. Vegetation dynamics are described in terms of a single state variable $\rho(\mathbf{s}, t)$ defined as the total phytomass density at time $t$ of all plant species present at point $\mathbf{s} \equiv (x, y)$. Local changes in the density are modelled through a generalized logistic equation expressing a growth–death balance (see Harper 1977 for the use of the logistic equation in vegetation science). The influences exerted by neighbouring vegetation on the local dynamics at $\mathbf{s}$ are averaged by means of spatial weighting functions (referred to in physics as a mean-field treatment). The formation of periodic vegetation patterns is interpreted as being the outcome of a symmetry-breaking instability analogous to the one initially described by Turing (1952) for chemical systems (Castets *et al*. 1990; Ouyang & Swinney 1991). Turing-like spatial instability is a central concept to explain morphogenesis in biological systems (Meinhardt 1982; Murray 1993). In the context of plant ecology, the PI model sustains the idea that periodic vegetation patterns result from an interplay between two main processes, namely propagation by reproduction and inhibition through competition.

Two conditions are required for spatial instability to occur: (i) vegetation must exert a positive (facilitative) influence on the process of reproduction; and (ii) competitive interactions must occur over a larger

range than facilitative ones. Periodic patterns thus result from a trade-off between facilitation and competition. The resulting wavelength is much larger than the range of either facilitative or competitive interactions between individual plants, because the spatial instability amplifies local individually based processes and leads to the emergence of a community-level pattern. As a consequence, the spectral analysis of digitized images, although useful to quantify vegetation patterns, cannot yield any direct information on the ranges of plant interactions. Indirect information may, nevertheless, be obtained through the PI model (see Lejeune *et al*. 1999 for a first attempt).

Competitive and facilitative influences co-occur in many plant communities (Callaway & Walker 1997). In arid and semi-arid environments, competition is generally exerted via the root-system, while the above-ground plant parts tend to have facilitative effects (Stuart-Hill & Tainton 1989; Vetaas 1992; Belsky 1994). For perennial grasses, bare areas between tussocks are often maintained by root competition (Aguilera & Lauenroth 1993), while for trees and shrubs, lateral roots extend beyond the limit of the crown (Belsky 1994; Breman & Kessler 1995; Martens *et al*. 1997) and extract water in inter-canopy areas (Breshears *et al*. 1997). Hence, having a larger range for below-ground competition than for above-ground facilitation is at least plausible in semi-arid plant communities and, as far as we know, the PI model is the first theoretical attempt at exploring the consequences of such a discrepancy.

The PI model, for both analytical investigation and numerical integration, considers a situation with low density and shallow density gradient. Arid regions, characterized by a low average phytomass and by vegetation patterns with a large wavelength in comparison with the size of individual plants, fulfil these assumptions. For the idealized situation of a strictly isotropic and homogeneous environment, the resulting equation is (Lejeune & Tlidi 1999; Lejeune *et al*. 1999; Lefever *et al*. 2000):

$$\frac{\partial}{\partial t}\rho(\mathbf{s},t) = \rho(\mathbf{s},t)\,[1-\mu+(\Lambda-1)\rho(\mathbf{s},t)-\rho^2(\mathbf{s},t)\,]+ \frac{1}{2}(L^2-\rho(\mathbf{s},t))\;\Delta\rho(\mathbf{s},t)-\frac{1}{8}\rho(\mathbf{s},t)\Delta^2\rho(\mathbf{s},t) \qquad \text{eqn 1}$$

where $\Delta = \dfrac{\partial^2}{\partial x^2}+\dfrac{\partial^2}{\partial y^2}$ is the two-dimensional laplacian operator. The dynamics are characterized by three dimensionless parameters: μ, the decay-to-reproduction ratio; Λ, the intensity of the facilitation on reproduction; and *L*, the facilitation to inhibition range ratio. Note that, as a basic premise, the parameter μ can be interpreted as an indirect measure of aridity. Indeed, all things being equal, a lower rate of phytomass reproduction, possibly with a higher rate of phytomass decay, may be expected when the environment becomes drier (Walter 1971; Fowler 1986; Kadmon 1995).

An increase in μ leads to a lower average density ($<\rho>$) of the vegetation and to a larger wavelength of the pattern, with evenly distributed vegetation being replaced, as its value rises, by spots of bare soil perforating the vegetation cover according to a hexagonal lattice, and then by vegetation stripes with no preferential orientation. It should be noted that the patterning mechanism is intrinsically dynamic in nature and that periodic structure is solely determined by parameter values and not by initial or boundary conditions.

A generalized version of the PI model shows that pattern formation is modified in anisotropic environments. In particular, when anisotropy influences inhibitory interactions, vegetation bands develop perpendicular to the direction of anisotropy (Lefever & Lejeune 1997; Lejeune *et al*. 1999; Lefever *et al*. 2000). However, dynamics under spatial constraints, such as on sloping ground, are beyond the scope of the present paper and the generalized version of the PI model will not be discussed further.

For simulations, equation 1 was numerically integrated using the finite difference method that considers a discrete approximation of space based on pixels whose size is determined by computational considerations (see Manneville 1990 for details). The integration domain was a square-shaped area with periodic boundary conditions (as if the area were mapped on a torus). The initial condition of every simulation was a uniform value of ρ (homogeneous cover) disturbed by a random white noise of small amplitude that expresses the inherent variability of vegetation cover. A given model simulation was conceptually related to a given life-form by setting each iteration to the average time needed by an individual plant to reach maturity (generation time), namely one rainy season for an annual grass and 30–40 years for a small tree such as *Pterocarpus lucens* (Couteron 1998). The observed patterns were chiefly made of a continuous cover of annual grasses punctuated by bare spots and disappeared from digital images after application of a threshold that captured only woody cover (Couteron 1998). All simulations were therefore run to model annuals.

## MAIN STUDY AREA

The main study area was located in the northern part of the Yatenga Province (Burkina Faso), between 13°45′ and 14°15′ N, and 2°20′ and 2°40′ W. The climate is semi-arid tropical with mean annual temperatures between 29 °C and 30 °C and a potential evapotranspiration (Penman) slightly under 2.000 mm year$^{-1}$. There is a long dry season from October to May, with a short wet season from June to September (heaviest rainfall in August). Average annual rainfall for 1951–89 was between 500 mm and 600 mm (L'Hôte & Mahé 1996).

The plain on which the study was situated had no consistent overall slope. Soil depth was 0–80 cm over ironstone and sandstone debris that has been locally re-consolidated into a discontinuous petroferric cuirass (Couteron & Kokou 1997). The texture of this topsoil

was sandy-clay to clay-silt with locally abundant ironstone gravels denoting the proximity of the underlying cuirass, and thus adverse edaphic conditions.

Vegetation belonged to the 'Sahel regional transition zone', with most woody species related to the 'Sudanian regional centre of endemism' (White 1983). According to a detailed field description (Couteron & Kokou 1997), the most abundant woody species was *Combretum micranthum* G. Don (mean adult height 2.8 m), although *Pterocarpus lucens* Lepr (5.3 m) and *Anogeissus leiocarpus* (D.C.) G. et Perr. (7.5 m) accounted for more of the woody phytomass. The overall density of woody individuals (with a height above 1.5 m) was 300 $ha^{-1}$. The herbaceous strata consisted of annual grasses and forbs. During the dry season, herders use fire to promote regrowth of perennial grasses but as such species were absent, fire was very unlikely and the woody vegetation was indeed dominated by fire-sensitive species. The area had a rather low human population (less than 10 inhabitants $km^{-2}$); none of the vegetation types under study had been cleared for crops and pastoral utilization was moderate (Couteron & Kokou 1997).

### SITE FROM NIGER

Additional data (aerial photographs from the 1950s) were obtained from southern Niger (approximately 13°06′ N, 2°07′ E) to describe some periodic patterns that were not observable in north-west Burkina Faso. Although only limited field data were available from a cursory inspection carried out in 1992 (Couteron & Kokou, unpublished report for UNESCO), a great deal of information was provided by intensive field studies conducted in neighbouring locations (White 1970; Seghieri *et al*. 1997; Couteron *et al*. 2000). Climatic conditions were similar to those of the Burkina site and the average annual rainfall for 1951–89 was also in the range 500–600 mm (L'Hôte & Mahé 1996).

In southern Niger, patterned vegetation (either spots or bands) was observable on laterite-capped plateaux with very gentle slopes ranging from zero to less than 1%. The soils were thin (usually less than 40–50 cm) over a cemented iron-pan that limited root penetration. Texture was sandy-clay with petroferric gravels (White 1970; Thiéry *et al*. 1995; Seghieri *et al*. 1997). The vegetation, as observed in the 1980s and 1990s, fitted very well with the overall description provided for the main site in Burkina Faso, except that *Combretum micranthum* showed a more pronounced dominance in southern Niger and the woody vegetation was thus denser and of lower stature (Couteron *et al*. 2000). No human settlement was observable within the area covered by aerial photographs.

### AERIAL PHOTOGRAPHS AND DIGITIZED IMAGES

The main site (Burkina Faso) was photographed on 10 October 1994 around 10.30 (universal time), i.e. with a zenith solar angle of about 29°. Pictures were taken from an elevation of 750 m using a Pentax ILX camera (50 mm focal length and 35 mm lens) in a door-mounted arrangement for unmodified aircraft. The film (Kodak Gold 100 ASA) was machine-processed into coloured prints at approximately 1 : 10.000 scale. The 15 photographs used, corresponded with a sampled area of about 525 ha. Prints were digitized into grey levels of reflectance (range 0–255) at a resolution of 300 dots per inch (DPI) through an AGFA® Studiostar scanner (each pixel side corresponded to a distance of 0.8 m in the field). Note that there is no relation between the size of pixels in simulated patterns (as in Figs 2 and 3), resulting from computational considerations, and the size of pixels in photographs that was chosen to allow enough details in displayed figures (e.g. in Figs 5, 6 and 7).

A limited set of panchromatic aerial photographs from southern Niger was acquired from IGN-France. These were taken during a high altitude flight (6300 m above ground level) on 14 November 1955, with a camera with a focal length of 125 mm. Contacts prints were at a scale of 1 : 50 000 and digitizing at a resolution of 1500 DPI gave pixels corresponding to the same field measurement (0.8 m side) as at the main site.

On all digitized images, bright pixels corresponded with bare soil, dark ones with areas dominated by woody vegetation, and intermediate grey-scale values corresponded with continuous grass cover. The above-ground phytomass of continuous grass and woody vegetation averages 1500 kg $ha^{-1}$ and 20 000 kg $ha^{-1}$, respectively (Le Houerou 1989; Couteron, unpublished data), and grey-scale values can therefore be seen as a monotonic non-decreasing function of the phytomass.

### SPECTRAL ANALYSIS OF DIGITIZED PHOTOGRAPHS

Two-dimensional spectral analysis aims to model a digital image in terms of cosine and sine functions. Detailed information on the method has been provided by Ripley (1981) and Renshaw & Ford (1984), whilst Mugglestone & Renshaw (1998) presented an application on digitized aerial photographs. The main tool is the periodogram, which is a set of values, $Ipq$ in Cartesian co-ordinates, or, as here, $Gr\theta$ in polar co-ordinates, each representing the portion of image variance $\sigma^2$ that can be accounted for by a simple cosine wave repeating itself $r$ times (wavenumber) along a travel direction of $\theta$. For the present paper we chose to display the periodogram as a square grey-tone image (with re-scaling in the range 0–255), for which the centre corresponds to $r = 0$. For the simple periodic pattern in Fig. 1(a) (straight bands), the resulting periodogram image (Fig. 1b) featured only two symmetric spikes at the points $(r, \theta) = (5, \pm 90°)$.

For a more complex image, it is convenient to carry out a separate investigation for periodicity and orientation by computing polar spectra (Renshaw & Ford 1984), i.e. (i) a 'radial spectrum' $I(r)$ (as in Fig. 1c) that

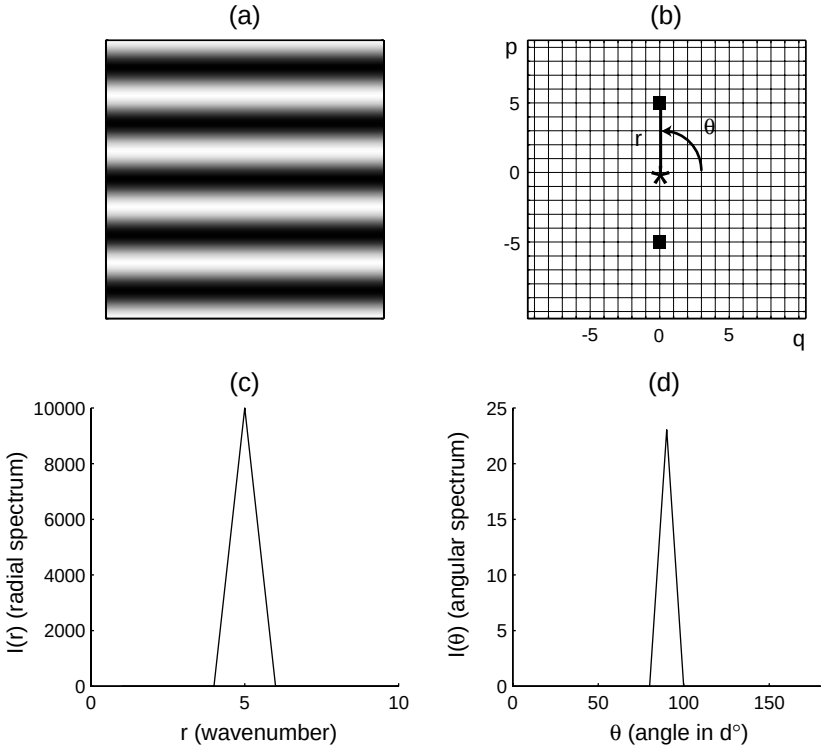


**Fig. 1** Main tools of two-dimensional spectral analysis as illustrated through a pure cosine wave (i.e. straight bands). (a) Pattern under analysis. (b) Central part of the two-dimensional periodogram with its centre denoted by a star. The Cartesian co-ordinates $p$ and $q$ are spatial frequencies from which wavenumbers, $r = \sqrt{p^2+q^2}$, and orientations, $\theta = \tan^{-1}[p/q]$, are deduced. (c) Radial spectrum. (d) Angular spectrum. Spectra are composed of dimensionless values, namely portions of the image variance.

is obtained by binning $Gr\theta$ values for each successive wavenumber and (ii) an 'angular spectrum' $I(\theta)$ (as in Fig. 1d) that is computed by binning periodogram values into angular segments, such as $-5° < \theta \leq 5$, ... , $165 \theta \leq 175°$. The radial and angular spectra are re-scaled by dividing by $\sigma^2$ and $k\sigma^2$ respectively, where $k$ is the number of periodogram values corresponding to a particular bin $\theta$. In the absence of spatial structure, values of the re-scaled angular spectrum are distributed as $\chi^2_{2k}$ with expected value one (Renshaw & Ford 1984). Hence, results significantly above one indicate dominant directions in the image. Dominant wavenumbers can be determined by visual inspection of the radial spectrum. The wavelength is computed as the image size divided by the wavenumber, and its assessment is fairly consistent for any size of image that is at least 3–5 times as large as the pattern under investigation.

### SYSTEMATIC SCREENING AND PATTERN DETECTION

Periodicity analysis was based on the computation of the radial spectrum for square non-overlapping windows, which were systematically sampled from digitized photographs. Several window sizes, ranging from 140 m to 250 m in the field, were used to ensure that the results were independent of window size.

Strong directional effects, that may correspond to the main asymptotic outcomes of the PI model, i.e. banded or hexagonal patterns, were systematically detected using a specific adaptation of a generic approach called 'template matching in Fourier space' (Niblack 1986; Anonymous 1997). The basic idea is to compare the angular spectrum observed for successive positions of a sliding window with the angular spectrum of any target pattern (i.e. a 'template'). Positions with spectra that best match the target spectrum are retained for further examination based on the whole periodogram. Two theoretical 'target spectra' were defined, with one spike (as in Fig. 1d, representing a banded pattern) or three identical spikes shifted from each other by 60° (Fig. 2a, representing a spotted vegetation). An 'optimal correlation' (denoted by $C_{op}$) was computed between the angular spectrum $I(\theta)$ obtained from each particular position of the sliding window and a given target spectrum $I_T(\theta)$. $C_{op}$ was the highest value of Pearson's coefficient of correlation that was found after having shifted one of the two spectra over the whole set of angular bins.

Pattern detection was carried out by computing the optimal correlation with each target spectrum for a sliding window of 140 m by 140 m, i.e. three times the dominant wavelength of the pattern under study (see below). The digitized image was screened in both Cartesian directions by centring the window on successive nodes of a square grid (25 m by 25 m). Optimal correlation was recorded on every node, yielding a two-dimensional array of $C_{op}$ values for each target pattern.

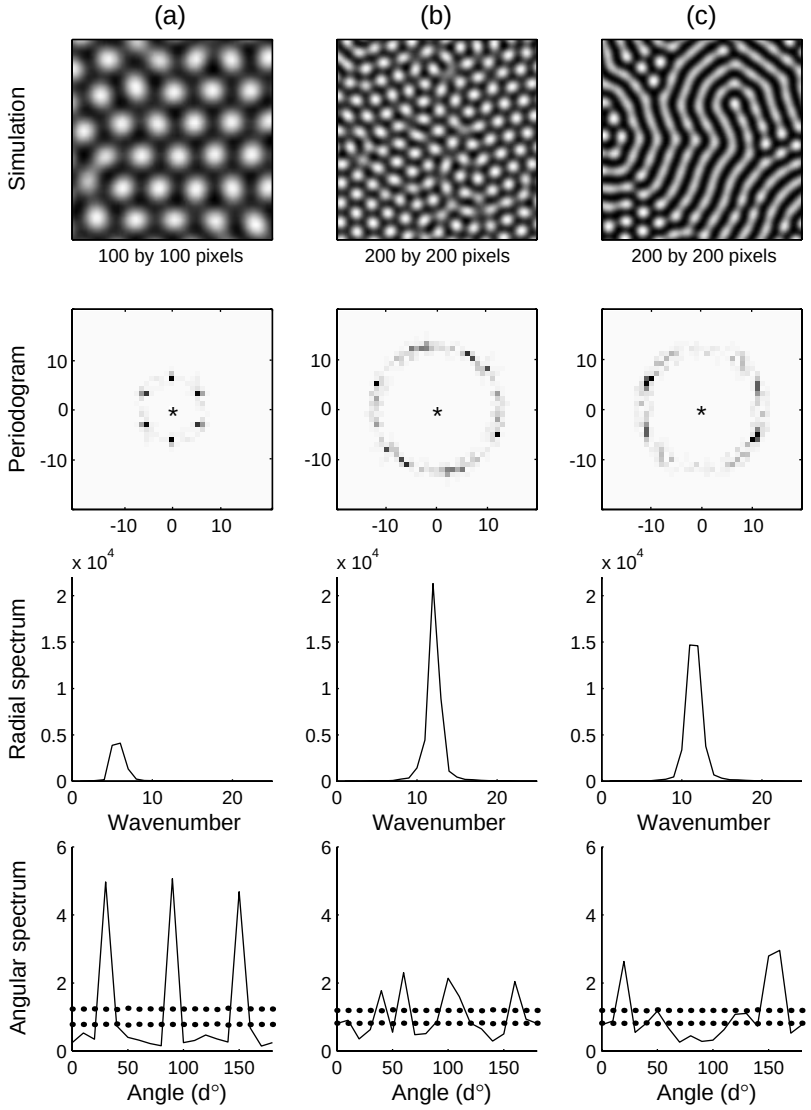


**Fig. 2** Spectral attributes for some patterns obtained with the isotropic version of the PI model for parameters $\Lambda = 1.2$ and $L = 0.2$ ($\mu = 0.98$ for spots and 1.0 for bands). (a) Spotted pattern as perceived through a small window selected for its clear hexagonal symmetry. (b) The same pattern as perceived through a larger window. (c) Banded pattern from a large window. For angular spectra, the dotted lines indicate the 5% bilateral confidence interval computed from the $\chi^2$ distribution.

## Results

### SPECTRAL CHARACTERIZATION OF GENERATED PATTERNS

The first test pattern was generated using the isotropic version of the PI model for a set of parameters ($\mu = 0.98, \Lambda = 1.2, L = 0.2$) that lead to a pattern of denuded spots within a continuous vegetation cover. The area used for simulation was sufficient for spots to occur at least 25 times along each side (i.e. $r > 25$). The windows shown in Fig. 2(a,b) have sides of 20% and 40% of the simulated area, respectively, and represent portions of the pattern obtained after 300 000 iterations (considered 'asymptotic' since further changes in the average phytomass density were negligible).

Systematic screening of the whole pattern identified the window in Fig. 2(a) as providing the highest optimal correlation ($C_{op} = 0.98$) with the angular spectrum of a typical hexagonal symmetry. Its periodogram had six dominant entries (dark dots) constituting the vertices of a hexagonal frame, and the hexagonal pattern is completely characterized by its wavelength and by three dominant orientations shifted from each other by 60°. The radial spectrum displayed a spike ($r = 5$–$6$), while the angular spectrum had three significant peaks for the dominant directions, i.e. $\theta = 30°, 90°, 150°$ (Fig. 2a). Angular spectra are plotted only for angles between 0° and 180° as results between 180° and 360° are redundant. However, the whole square periodogram is displayed since the pattern of main entries may be highly suggestive of an underlying hexagonal symmetry.

When considered through a larger window (Fig. 2b), the spot distribution was, however, less regular. For instance, the main directions were not consistent throughout the pattern, some spots had five or seven (rather than six) nearest neighbours, and neighbouring spots sometimes coalesced. Such defects, although less

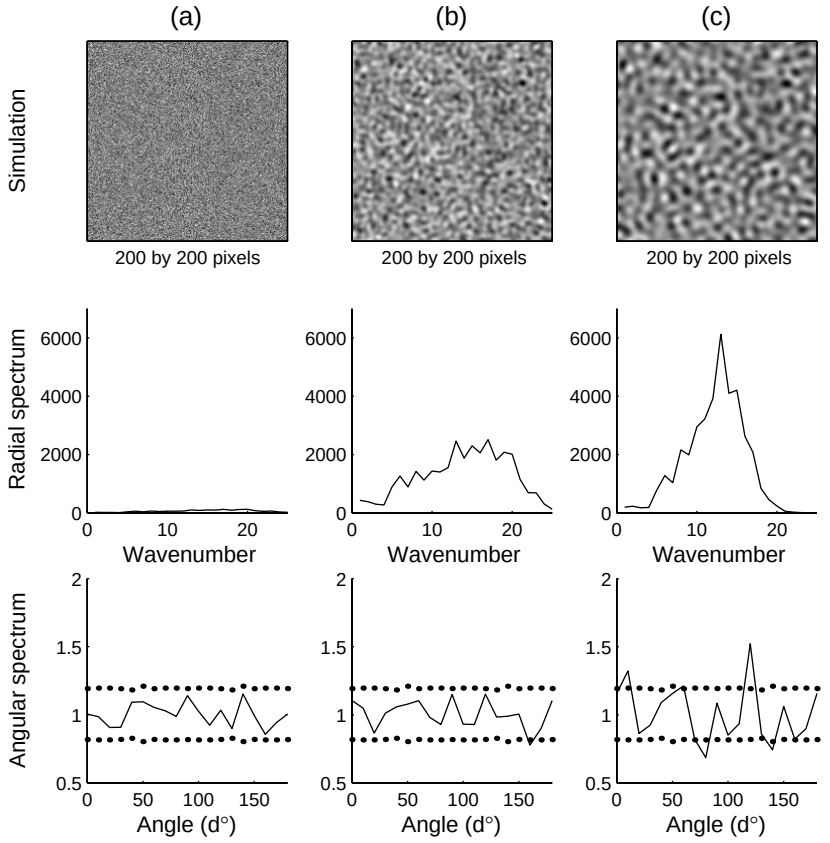


**Fig. 3** Spectral attributes for successive transient patterns from the specific simulation leading to Fig. 2(b). (a) Initial condition. (b) Pattern after 70 iterations. (c) Pattern after 1000 iterations. For angular spectra, the dotted lines indicate the 5% bilateral confidence interval computed from the $\chi^2$distribution.

perceptible from a smaller window, are classic features of periodic patterns in spatially extended systems (Ciliberto *et al*. 1990). The periodogram had numerous dominant entries forming an annulus. A dominant wavenumber, $r = 14$, was therefore still obvious from the radial spectrum, but three main directions were no longer perceptible on the angular spectrum.

The PI model was also run for a set of parameters whose asymptotic result was a banded structure ($\mu = 1$, $\Lambda = 1.2$, $L = 0.2$). Because isotropic conditions were imposed, the bands were flexuous and their orientation changed within the area used for simulation. Even for the subsample in Fig. 2(c), there were two main orientations that were reflected by two peaks in the angular spectrum instead of a single one (cf. Figure 1a,d). Furthermore, the bands broke down to give spots. The whole pattern corresponded, nevertheless, to a precise wavelength, with a radial spectrum pointing towards $r = 11$–$12$.

Additional, transient, patterns were seen as the initially aperiodic structure developed into an asymptotic pattern. A dominant periodicity emerged progressively from the virtually flat radial spectrum (Fig. 3a). It should be noted that for the specific simulation that led to Fig. 2(b), a slight dominance of wavenumbers in the range 12–20 could be seen after only 70 iterations (Fig. 3b) and that the dominance of $r = 14$ (asymptotic value) was clearly established by 1000 iterations (Fig. 3c). The hexagonal symmetry that could be identified in the asymptotic pattern (at least from small windows, Fig. 2a) took much longer to develop than the asymptotic wavelength. Indeed, only 10% of the positions of the sliding window after 100 000 iterations yielded a high optimal correlation (> 0.7) with the three-spiked angular spectrum, instead of 30% for the asymptotic pattern.

## PERIODICITY ANALYSIS

The 15 aerial photographs from Burkina Faso were systematically sampled through non-overlapping windows of 250 m by 250 m, and radial spectra were computed. The average radial spectrum (Fig. 4a) suggested dominant wavenumbers in the range $r = 5$–$8$, and thus wavelengths of 31–50 m. The analysis was repeated with a 140-m by 140-m window (Fig. 4b), giving $r = 2$–$5$ (wavelengths between 28 m and 70 m). Results from the two window sizes proved consistent, although using a smaller window automatically implies a lower spectral precision (Kumaresan 1993) and less reliability.

Inspection of individual windows nevertheless allowed us to distinguish a fine-grained pattern (Fig. 5a), with a spread of wavenumbers in the range $r = 5$–$13$ and a modest spike for $r = 8$ (31 m), from a coarse-grained pattern (Fig. 5b) with several strongly dominant wavenumbers in the range $r = 5$–$10$ (25–50 m). In both cases, the structure leading to the dominant periodicity was

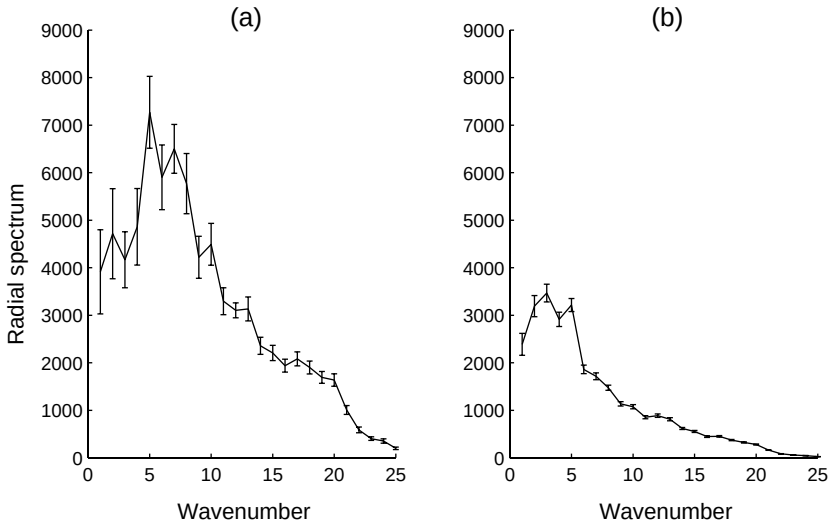


**Fig. 4** Average radial spectrum for two sizes of sampling windows in Burkina Faso. (a) Spectrum for $n = 30$ windows of 250 m by 250 m. (b) Spectrum for $n = 88$ windows of 140 m by 140 m. Error-bars indicate the standard error around the mean spectrum.

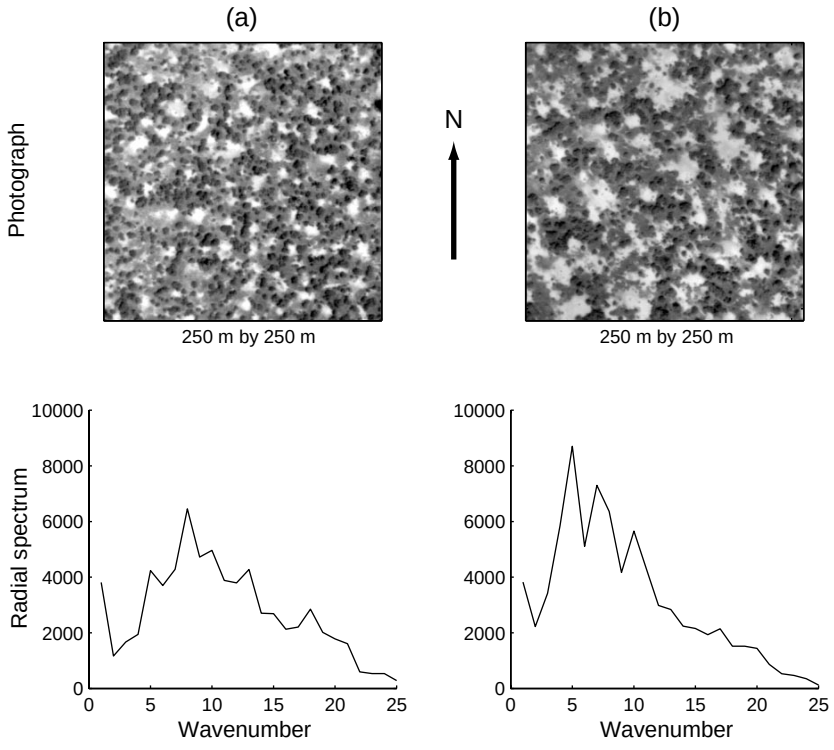


**Fig. 5** Two scales of pattern seen in individual 250 m by 250 m windows in Burkina Faso. (a) Fine-grained pattern. (b) Coarse-grained pattern.

apparent to the human eye, namely, the regular punctuation of continuous vegetation by spots of bare ground.

## DOMINANT ORIENTATIONS AND HEXAGONAL SYMMETRY

Aerial photographs from the Gondo plain, Burkina Faso, were systematically screened for a hexagonal symmetry (as in Fig. 2a) using the three-spiked angular spectrum as target. Large values ($C_{op} > 0.7$) were detected for only 3% of the positions on which the sliding window was centred (along the 25 m × 25 m grid) and not all of these yielded a convincing periodogram. Although typical hexagonal patterns proved rare, examples of both fine- and coarse-grained patterns were found (Fig. 6a, b). The selected examples were more than 12 km apart and were separated by at least one stretch of seasonally flooded woodland. They should thus be considered as independent realizations of the hexagonal pattern.

About 5–10% of the positions taken by the sliding window yielded a very high optimal correlation ($C_{op} >$ 0.8) with the 'single-spike' spectrum, suggesting the dominance of a unique orientation. Such a dominance was frequently occasioned by an alignment of spots,

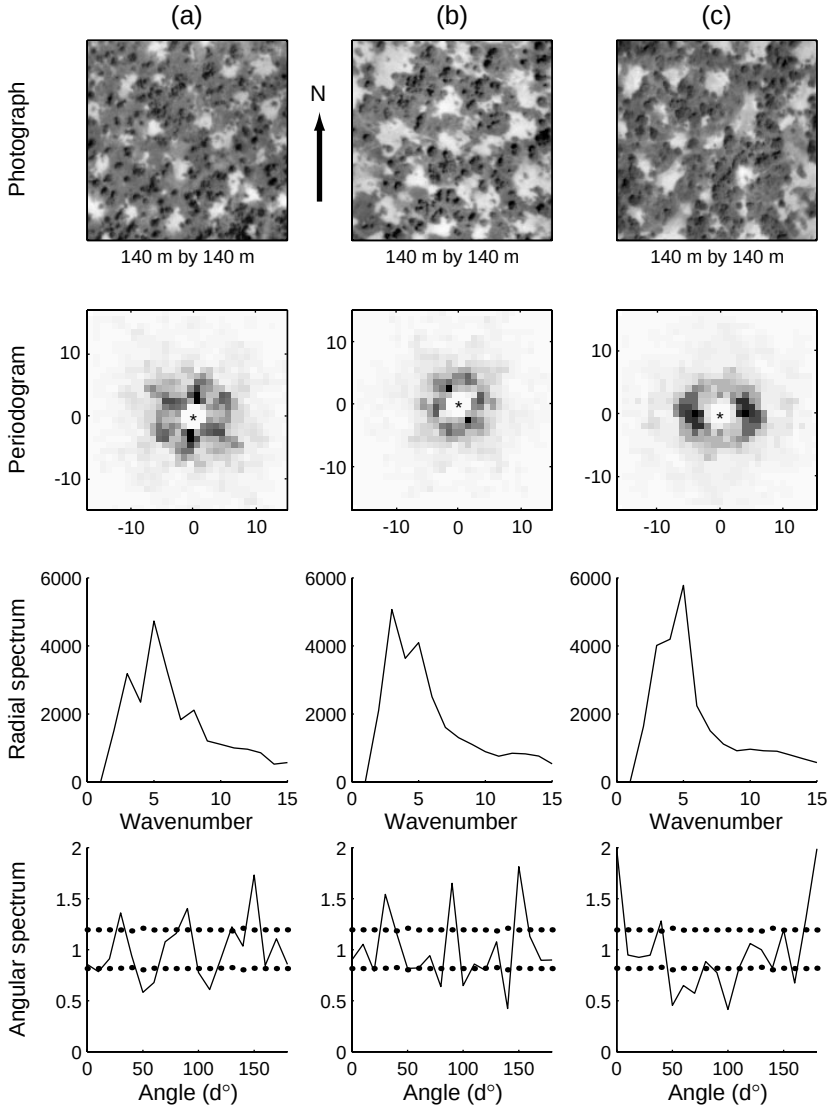


**Fig. 6** Spectral attributes for specific 140 m by 140 m windows from Burkina Faso that displayed strong directional effects. All windows are orientated to the North. (a) and (b) Hexagonal symmetry. (c) Unique privileged direction. The dotted lines associated with the angular spectrum indicate the 5% bilateral confidence interval computed from the $\chi^2$ distribution.

which might be interpreted as a trend towards bands (Fig. 6c). However, the direction of the alignment was not consistent across the Gondo plain, or even within a given aerial photograph, spots rarely coalesced and true denuded bands were never encountered. Bands were observed on nearby slopes (i.e. under anisotropic conditions) but on the Gondo plain the environment determined a spotted pattern.

## ADDITIONAL RESULTS FROM NIGER

Periodicity analysis was carried out for the southern Niger site using non-overlapping windows with the same size as in Burkina Faso, i.e. 250 m by 250 m in the field. The spotted pattern (Fig. 7a) yielded dominant wavenumbers that were strikingly consistent with the results from Burkina Faso (i.e. wavelengths ranging between 30 m and 50 m). Coarse-grained and fine-grained patterns tended to coexist within the studied area of 800 m by 800 m. Angular spectra were computed from sliding windows of 140 m by 140 m and screened for dominant directions, but no consistent preferential direction was found. Some positions (4–5%) yielded an optimal correlation with the three-spiked angular spectrum above 0.7, but few of them provided convincing examples of hexagonal symmetry.

The Niger data provided an opportunity to study the transition from spots to bands, a feature that was not observable in north-west Burkina Faso. Patterns (as in Fig. 7b) of elongated patches of bare ground, sometimes coalescing to form flexuous and ramified bands, had some similarity with the simulated pattern in Fig. 2(c). No consistent preferential direction was encountered, but a strongly dominant wavenumber was found ($r = 5$, i.e. a wavelength of 50 m). Typical banded structures (as in the upper-right part of Fig. 7c) were observed at the fringes of the plateau under study, i.e. at less than 5 km from the two other patterns. Some sharp transitions

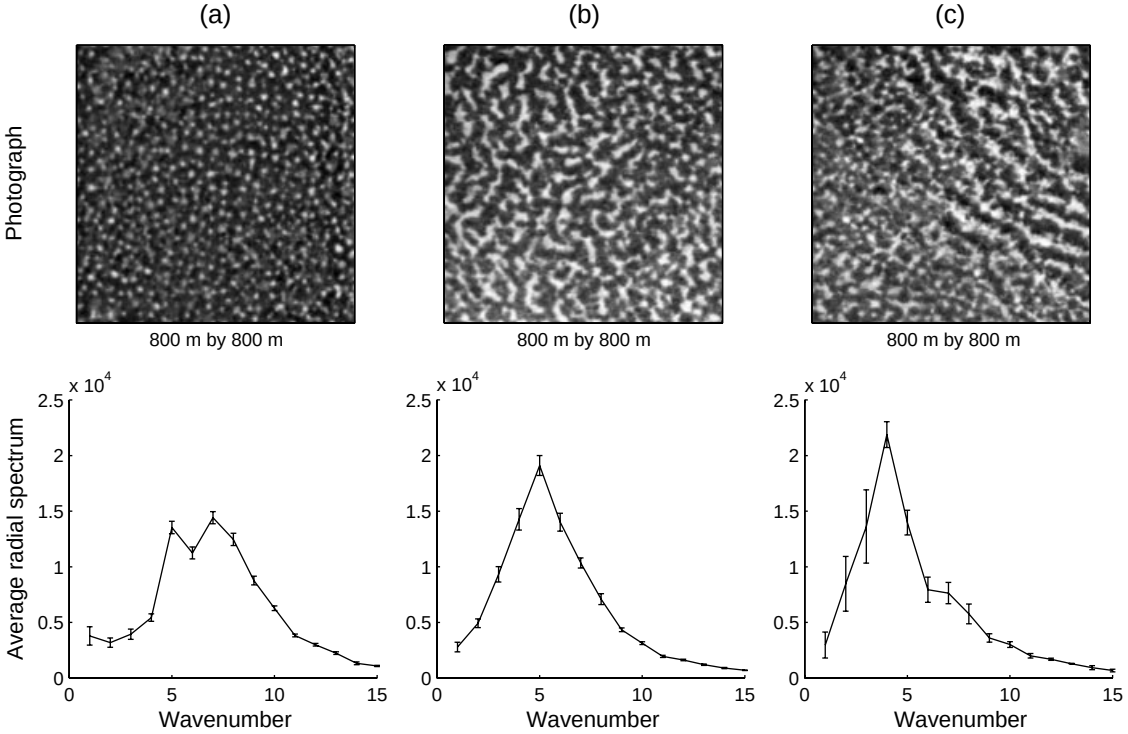


**Fig. 7** Three types of pattern observed on the same plateau in southern Niger with their respective radial spectrum (computed from 250 m by 250 m windows). Error-bars indicate the standard error on the mean spectrum. (a) Spots. (b) Elongated spots with no dominant orientation. (c) Banded system organized in relation to a local declivity near the edge of the plateau (note the sharp transition with the spotted pattern).

from spots to bands were observed (Fig. 7c), suggesting that environmental conditions change markedly towards the edge of the plateau. Banded patterns were characterized by a dominant wavelength (62.5 m, $r = 4$) that was larger than the neighbouring spotted or intermediate patterns and similar to values (50–75 m) reported for tiger bush in southern Niger (White 1970; Wu *et al*. 2000) and in Burkina Faso (Couteron *et al*. 2000). The increase in wavelength from 30 to 50 m in a spotted pattern to *c.* 50 m for an intermediate structure and 50–75 m in bands is consistent with a fundamental prediction of the PI model.

## Discussion

In north-west Burkina Faso, the vegetation of the Gondo plain can be described as a savanna, i.e. trees coexisting with grass (Belsky 1994), regularly punctuated by spots of bare soil. Most lateritic-capped plateaux in southern Niger displayed a very similar physiognomy and a strong floristic affinity. In both locations, the use of spectral analysis demonstrated that the spotted pattern was not random, but corresponded to a characteristic range of wavelengths (i.e. 30–50 m) that proved strikingly consistent between two sites more than 500 km apart. This corroborated the visual impression that spots had a specific size and were regularly distributed throughout the continuous vegetation cover. Although coalescing spots were not frequent, patterns with elongated spots, that locally became flexuous bands, were observed in southern Niger (but not in north-west Burkina Faso). Such patterns also proved highly periodic with a dominant wavelength (50 m) situated at the upper bound of the range obtained for typical spotted patterns.

Predictions of the isotropic version of the PI model were supported, such as: (i) spotted and banded patterns may be two distinct outcomes of a unique dynamic process; (ii) if so, they should be characterized by specific dominant wavelengths; and (iii) the shift from spots to bands should be accompanied by an increase in wavelength. Dominant wavelengths in the range 50–75 m for highly organized banded systems on gentle slopes in both north-western Burkina Faso and southern Niger (White 1970; Leprun 1999; Couteron *et al*. 2000) are also consistent. It should be noted that the introduction of a slight anisotropy in the PI model does not significantly modify the resulting wavelength as long as the three fundamental parameters ($\mu$, $\Lambda$, $L$) are held constant (Lejeune 1999).

On the sub-horizontal plateaux of southern Niger, the existence of clearly orientated systems of bands (as on Fig. 7c) is likely to result from a locally very gentle yet consistent slope (White 1970; Seghieri *et al*. 1997). Hence the observation of periodic structures devoid of dominant orientations (as in Fig. 7a,b) pointed towards the absence of any significant slope-induced anisotropy, and corroborated the fundamental prediction of the PI model that periodic patterns may emerge even in a truly isotropic environment.

In spite of defects, a hexagonal symmetry is easily detectable in simulated patterns through systematic screening using a sliding window that is reasonably

small with respect to the pattern (i.e. $r < 6–7$). In this case, about 30% of window positions yielded an optimal correlation above 0.7. The figures were far lower for the real-world periodic patterns that were analysed for the present study, i.e. less than 3% in Burkina Faso and 5% in Niger for, respectively, 525 ha and 300 ha sampled. Furthermore, not all windows with a high optimal correlation corresponded to convincing hexagonal patterns (but see Fig. 6).

There are at least two reasons that may explain why hexagonal symmetry may not be detectable over wide areas: (i) a high level of noise may result from edaphic heterogeneity, vegetation complexity (trees vs. grass) or climatic fluctuations; and (ii) there may not have been sufficient time to allow the emergence of a clear symmetry. For pattern formation models, the time-scale on which the wavelength appears is expected to be smaller than for symmetry ( Manneville 1990) and in our simulations, it took about 300 times as many iterations to obtain a convincing hexagonal symmetry than to get a dominant wavelength (Fig. 3). In semi-arid ecosystems, periodic patterns probably have had a limited time to evolve and settle between two drastic changes of climate. Ecologists are becoming aware that most ecosystems may be far from any equilibrium or asymptotic state (Sprugel 1991). Since an iteration (i.e. a generation time) represents at least 1 year (for annual grasses), it might be more relevant to compare observed patterns with the transient outcomes of the PI model rather than with asymptotic ones. Furthermore, real-world photographs ought to be smoothed through a relevant filter (Niblack 1986) before comparison, because all results of the PI model are smooth structures (due to the mean-field approach). Indeed, the model is designed to render an overall pattern (periodicity, orientation) and not local details.

The limited literature on this subject has interpreted bare spots as results of recurrent disturbances such as building of large termite mounds (i.e. termitaria; Macfadyen 1950), increasing aridity, or both (Clos-Arceduc 1956; Boudet 1972), which affect vegetation cover. Although relating bare spots, as perceived from air photos, to giant termitaria is tempting to anybody who has ever observed both, the bare spots are substantially larger than the denuded areas around termitaria. Hence, in Burkina Faso, the average diameter of bare spots was 13 m and 28 m for the fine-grained and coarse-grained patterns, respectively (Fig. 5), compared with mean and extreme values of 4.5 m and 10 m around mounds (Ouedraogo 1997). Furthermore, not all bare spots were circularly shaped (Fig. 5b, 7b). Macfadyen (1950) observed in Somalia circular and 'smudged' spots having respective greatest dimensions of 20 m and 40 m, and acknowledged that much smaller values were found for bare areas around mounds. Although bare spots cannot directly correspond to termitaria, such structures may play a role in the emergence of the spotted pattern via their interaction with vegetation. However, the PI model provided a theoretical demonstration that a periodic spotted pattern may be the strict outcome of local interactions between plants.

Boudet (1972) hypothesized that spotted vegetation may be an intermediate state in a regressive series driven by overgrazing and/or aridity and leading from savannas with a complete vegetation cover to tiger bush. Clos-Arceduc (1956) and Greig-Smith (1979) suggested that spots may be a first step on the way towards bands. From a geographical standpoint, savanna, spotted bush and tiger bush successively prevail along the rainfall gradient stretching from values above 800 mm $year^{-1}$ to values less than 400 mm $year^{-1}$ (Clos-Arceduc 1956; White 1970; Ambouta 1997; Leprun 1999). Consequently, an analogy between the geographical zonation and a temporal succession may be appealing. Such an analogy is not contradicted by the isotropic version of the PI model according to which homogeneous vegetation, bare spots and bands appear successively with increasing values of a modelling parameter ($\mu$) that quantifies aridity. However, although spotted vegetation appears for intermediate values of $\mu$, it is not automatically a transient state between a homogeneous cover and a banded pattern. It may also be a stable pattern that can be characterized, at least theoretically, by a hexagonal symmetry, and above all by a dominant wavelength.

The existence of a dominant wavelength in spotted vegetation, as evidenced for the first time by the present study, demonstrated that a slope-induced anisotropy is not a necessary condition for the emergence of a periodic pattern. As a consequence, there is a lack of generality in most existing models dealing with periodic semi-arid vegetation (Mauchamp *et al*. 1994; Thiéry *et al*. 1995; Dunkerley 1997; Klausmeier 1999). Indeed, the latter author had to invoke local topographic irregularities to explain the existence of a heterogeneous vegetation cover on a strictly flat terrain. The strength of the PI model is its ability to account for several kinds of periodic patterns that are observable in both isotropic and anisotropic environments.

From a more general standpoint, the PI model contributes to the growing awareness that vegetation dynamics are greatly influenced by the interplay between facilitative and competitive processes (Holmgren *et al*. 1997; Martens *et al*. 1997). More precisely, the model demonstrates that a discrepancy in the respective ranges of facilitative and competitive interactions is a necessary condition to have a large-scale periodic pattern emerging from local interactions between a great number of individual plants. Seeing spatial instability as a potential consequence of the dualism between facilitation and competition bears an ecological meaning that may extend beyond semi-arid vegetation.

## Acknowledgements

We are very grateful to Professor René Lefever (Université Libre de Bruxelles) for his fundamental contribution to the PI model, the development of which was supported by the Instituts Internationaux de Physique et de Chimie

Solvay and by the Centre for Nonlinear Phenomena and Complex Systems (ULB). The financial support of Fonds National de la Recherche Scientifique and of Fonds Emile Defay is also acknowledged. Dr Joël Chadoeuf (INRA Avignon, France) provided important suggestions on pattern detection, while Dr Moira Mugglestone (University of Leicester, UK) gave important information on two-dimensional spectral analysis. We are indebted to Dr John Ludwig (CSIRO, Australia) and Dr Bruce T. Milne (University of New Mexico, USA) who, as reviewers, provided valuable suggestions and comments, and to Dr L. Haddon for a substantial improvement of the final version.

## References

Aguilera, M.O. & Lauenroth, W.K. (1993) Seedling establishment in adult neighbourhoods – intraspecific constraints in the regeneration of the bunchgrass Bouteloua gracilis. *Journal of Ecology*, **81**, 253–261.

Ambouta, J.M.K. (1997) Définition et caractérisation des structures de végétation contractée au Sahel: cas de la brousse tigrée de l'ouest nigérien. *Fonctionnement et Gestion Des Écosystèmes Forestiers Contractés Sahéliens* (eds J.M. d'Herbès, J.M.K. Ambouta & R. Peltier), pp. 41–57. John Libbey Eurotext, Paris.

Anonymous (1997) *Matlab. Image Processing Toolbox*, 2nd edn. The Mathworks Inc. Natick, Massassuchets.

Belsky, A.J. (1994) Influences of trees on savanna productivity: tests of shade, nutrients, and tree-grass competition. *Ecology*, **75**, 922–932.

Boudet, G. (1972) Désertification de l'Afrique tropicale sèche. *Adansonia, Série*, **2** (12), 505–524.

Breman, H. & Kessler, J.-J. (1995) *Woody Plants in Agro-Ecosystems of Semi-Arid Regions*. Springer-Verlag, Berlin.

Breshears, D.D., Myers, O.B., Johnson, S.R., Meyer, C.W. & Martens, S.N. (1997) Differential use of shallow inter-canopy water by two semiarid woodland tree species: *Pinus edulis* and *Juniperus monosperma*. *Journal of Ecology*, **85**, 289–299.

Callaway, R.M. & Walker, L.R. (1997) Competition and facilitation: a synthetic approach to interactions in plant communities. *Ecology*, **78**, 1958–1965.

Castets, V., Dulos, E., Boissonade, J. & De Kepper, P. (1990) Experimental evidence of a sustained standing Turing-type nonequilibrium chemical pattern. *Physical Review Letters*, **64**, 2953–2956.

Ciliberto, S., Coullet, P., Lega, J., Pampaloni, E. & Perez-Garcia, C. (1990) Defects in roll-hexagon competition. *Physical Review Letters*, **65**, 2370–2373.

Clos-Arceduc, M. (1956) Etude sur photographies aériennes d'une formation végétale sahélienne: la brousse tigrée. *Bulletin de L'IFAN Série A*, **18**, 678–684.

Cornet, A.F., Delhoume, J.P. & Montana, C. (1988) Dynamics of striped vegetation pattern and water balance in the Chihuahuan desert. *Diversity and Pattern in Plant Communities* (eds H.J. During, M.J.A. Werger & J.H. Willems), pp. 221–231. SPB Academic, The Hague, The Netherlands.

Couteron, P. (1998) *Relations spatiales entre individus et structure d'ensemble dans des peuplements ligneux soudano-sahéliens au nord-ouest du Burkina Faso.* Doctoral thesis, Université de Toulouse, Toulouse, France.

Couteron, P. & Kokou, K. (1997) Woody vegetation spatial patterns in a semi-arid savanna of Burkina Faso, West Africa. *Plant Ecology*, **132**, 211–227.

Couteron, P., Mahamane, A., Ouedraogo, P. & Seghieri, J. (2000) Differences between banded thickets (tiger bush) at two sites in West Africa. *Journal of Vegetation Science*, **11**, 321–328.

Dunkerley, D.L. (1997) Banded vegetation: development under uniform rainfall from a simple cellular automaton model. *Plant Ecology*, **129**, 103–111.

Fowler, N. (1986) The role of competition in plant communities in arid and semiarid regions. *Annual Review of Ecology and Systematics*, **17**, 89–110.

Glansdorff, P. & Prigogine, I. (1971) *Thermodynamic Theory of Structure, Stability and Fluctuations*. Wiley – Interscience, New York.

Greig-Smith, P. (1979) Pattern in vegetation. *Journal of Ecology*, **67**, 755–779.

Haralick, R.M. (1979) Statistical and structural approaches to texture. *Proceedings of the IEEE*, **67**, 786–804.

Harper, J.L. (1977) *Population Biology of Plants*. Academic Press, New York.

Holmgren, M., Scheffer, M.S. & Huston, M.A. (1997) The interplay of facilitation and competition in plant communities. *Ecology*, **78**, 1966–1975.

Holt, R.D. & Keitt, T.H. (2000) Alternative causes for range limits: a metapopulation perspective. *Ecology Letters*, **3**, 41–47.

Hopfield, J.J. (1982) Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences (US)*, **79**, 2554–2558.

Kadmon, R. (1995) Plant competition along soil moisture gradients: a field experiment with the desert annual *Stipa capensis*. *Journal of Ecology*, **83**, 253–262.

Klausmeier, C.A. (1999) Regular and irregular patterns in semiarid vegetation. *Science*, **284**, 1826–1828.

Kumaresan, R. (1993) Spectral analysis. *Handbook for Digital Signal Processing* (eds S.K. Mitra & J.F. Kaiser), pp. 1143–1169. John Wiley & Sons, New York.

L'Hôte, Y. & Mahé, G. (1996) *Afrique de l'Ouest et Centrale. Précipitations Moyennes Annuelles (Période 1951–89)*. ORSTOM, Paris.

Le Houerou, H.N. (1989) *The Grazing Land Ecosystems of the African Sahel*. Springer-Verlag, Berlin.

Lefever, R. & Lejeune, O. (1997) On the origin of tiger bush. *Bulletin of Mathematical Biology*, **59**, 263–294.

Lefever R., Lejeune O. & Couteron P. (2000) Generic modelling of vegetation patterns. A case study of Tiger Bush in sub-Saharan Sahel. *Mathematical Models for Biological Pattern Formation* (eds P.K. Maini & H.G. Othmer), pp. 83–112. Springer-Verlag, New York.

Lejeune, O. (1999) *Une théorie champ moyen de l'organisation spatio-temporelle des écosystèmes végétaux.* Doctoral thesis, Université Libre de Bruxelles, Bruxelles, Belgium.

Lejeune, O., Couteron, P. & Lefever, R. (1999) Short range co-operativity competing with long range inhibition explains vegetation patterns. *Acta Oecologica*, **20**, 171–183.

Lejeune, O. & Tlidi, M. (1999) A model for the explanation of vegetation stripes (tiger bush). *Journal of Vegetation Science*, **10**, 201–208.

Leprun, J.C. (1999) The influence of ecological factors on tiger bush and dotted bush patterns along a gradient from Mali to northern Burkina Faso. *Catena*, **37**, 25–44.

Ludwig, J.A. & Tongway, D.J. (1995) Spatial organisation of landscapes and its function in semi-arid woodlands, Australia. *Landscape Ecology*, **10**, 51–63.

Macfadyen, W.A. (1950) Vegetation patterns in the semi-desert plains of British Somaliland. *Geographical Journal*, **116**, 199–210.

Manneville, P. (1990) *Dissipative Structures and Weak Turbulence*. Academic Press, New York.

Martens, S.N., Breshears, D.D., Meyer, C.W. & Barnes, F.J. (1997) Scales of above-ground and below-ground competition in a semi-arid woodland detected from spatial patterns. *Journal of Vegetation Science*, **8**, 655–664.

Mauchamp, A., Rambal, S. & Lepart, J. (1994) Simulating the dynamics of a vegetation mosaic: a functional spatialized model. *Ecological Modelling*, **71**, 107–130.

Meinhardt, H. (1982) *Models of Biological Pattern Formation*. Academic Press, New York.

Milne, B.T., Johnson, A.R., Keitt, T.H., Hatfield, C.A., David, J. & Hraber, P. (1996) Detection of critical densities associated with piñon-juniper woodland ecotones. *Ecology*, **77**, 805–821.

Mugglestone, M.A. & Renshaw, E. (1998) Detection of geological lineations on aerial photographs using two-dimensional spectral analysis. *Computers & Geosciences*, **24**, 771–784.

Murray, J.D. (1993) *Mathematical Biology*, 2nd edn. Springer, Berlin.

Niblack, W. (1986) *An Introduction to Digital Image Processing*. Prentice Hall, Englewood Cliffs, New Jersey.

Ouedraogo, P. (1997) *Rôle des termites dans la structure et la dynamique d'une brousse tigrée soudano-sahélienne.* Doctoral thesis, Université de Paris VI, Paris.

Ouyang, Q. & Swinney, H.L. (1991) Transition from a uniform state to hexagonal and striped Turing patterns. *Nature*, **352**, 610–611.

Renshaw, E. & Ford, E.D. (1984) The description of spatial pattern using two-dimensional spectral analysis. *Vegetatio*, **56**, 75–85.

Ripley, B.D. (1981) *Spatial Statistics*. John Wiley & Sons, New York.

Seghieri, J., Galle, S., Rajot, J.L. & Ehrmann, M. (1997) Relationships between soil moisture and growth of herbaceous plants in a natural vegetation mosaic in Niger. *Journal of Arid Environments*, **36**, 87–102.

Sprugel, D.G. (1991) Disturbance, equilibrium and environmental variability: what is natural vegetation in a changing environment? *Biological Conservation*, **58**, 1–18.

Stuart-Hill, B.C. & Tainton, N.M. (1989) The competitive interactions between *Acacia karoo* and herbaceous layer and how this is influenced by defoliation. *Journal of Applied Ecology*, **26**, 285–298.

Thiéry, J.M., d'Herbès, J.M. & Valentin, C. (1995) A model simulating the genesis of banded vegetation patterns in Niger. *Journal of Ecology*, **83**, 497–507.

Turing, A.M. (1952) The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London, Series B*, **237**, 37–72.

Vetaas, O.R. (1992) Micro-site effects of trees and shrubs in dry savannas. *Journal of Vegetation Science*, **3**, 337–344.

Walter, H. (1971) *Ecology of Tropical and Subtropical Vegetation*. Oliver & Boyd, Edinburgh.

White, L.P. (1970) *Brousse tigrée* patterns in southern Niger. *Journal of Ecology*, **58**, 549–553.

White, L.P. (1971) Vegetation stripes on sheet wash surfaces. *Journal of Ecology*, **59**, 615–622.

White, F. (1983) *The Vegetation of Africa. A Descriptive Memoir to Accompany the UNESCO/AEFTAT/UNSO Vegetation Map*. UNESCO/AETFAT/UNSO, Paris.

Wilson, J.B. & Agnew, A.D. (1992) Positive-feedback switches in plant communities. *Advances in Ecological Research*, **23**, 263–323.

Wilson, W.G. & Nisbet, R.M. (1997) Cooperation and competition along smooth environmental gradients. *Ecology*, **78**, 2004–2017.

Wu, X.B., Thurow, T.L. & Whisenant, S.G. (2000) Fragmentation and changes in hydrologic function of tiger bush landscapes, south-west Niger. *Journal of Ecology*, **88**, 490–500.

Yamamura, N. (1976) A mathematical approach to spatial distribution and temporal succession in plant communities. *Bulletin of Mathematical Biology*, **38**, 517–526.

*Received 4 July 2000*
*Revision accepted 30 January 2001*

INT. J. REMOTE SENSING, 2002, VOL. 23, NO. 17, 3407–3425

Taylor & Francis
Taylor & Francis Group

# Quantifying change in patterned semi-arid vegetation by Fourier analysis of digitized aerial photographs

PIERRE COUTERON*

Ecole Nationale du Génie Rural des Eaux et des Forêts/UMR-CNRS 5120 Botanique et Bioinformatique de l'Architecture des Plantes

(*Received 13 April 2000; in final form 24 April 2001*)

**Abstract.** Panchromatic aerial photographs from 1955 and 1985 (scale: 1:50 000) were used to quantify changes in semi-arid patterned vegetation caused by a succession of dry years in the early 1980s. The study site is located in the north-west part of Burkina Faso (West Africa), and features a plain with a savanna physiognomy and gentle slopes covered by tiger bush. Digitized data (pixel size of 3.15 m) covered a belt transect of 9 km by 1.5 km that has been divided into 315 $m^2$ square quadrats. Four reference quadrats were digitized with a pixel of 0.83 m, for comparison with high-resolution outlooks from 1994. Pattern quantification relied on spectral analysis by Fourier transform, that yielded dominant wavelengths (radial spectrum) and main orientations (angular spectrum). The vegetation in the plain displayed important changes that were related to the collapse of the herbaceous cover (and associated scattered trees), and its partial post-drought recovery. Such changes were quantified as a relative decline of small spatial wavelengths (< 40 m), followed by a bounce-back of wavelengths in the range 25–50 m. An underlying pattern of fragmented bands of woodland was also revealed with a dominant orientation and a characteristic wavelength ranging from 80 m up to more than 105 m. On the gentle slopes, dense wooded bands that constituted tiger bush (with a wavelength in the range 60–80 m) displayed only minor changes through time. Spectral analysis by Fourier transform proved to be a suitable way to monitoring patterned semi-arid vegetation for which periodicity and orientation are important characteristics. Consistent diachronic comparisons are possible even with historical photographs of varying scale and quality.

## 1. Introduction

Vegetation patterns with spatial periodicity have been reported for extensive areas experiencing arid or semi-arid climates, in Africa, Australia and America (White 1971). An example of this is the 'brousse tigrée', or tiger bush (a phrase coined by Clos-Arceduc 1956), which comprises vegetation strips alternating with bare ground suggesting the black markings of a tiger skin.

However, spotted patterns, though less frequently reported, are also encountered, sometimes near banded systems (MacFadyen 1950, Clos-Arceduc 1956, Boudet 1972). The emergence of a stable periodic structure from a homogeneous vegetation raises

*Corresponding address: ENGREF, 648 rue J.F. Breton, 34093 Montpellier cédex 5, France; e-mail: Couteron@engref.fr

*International Journal of Remote Sensing*
ISSN 0143-1161 print/ISSN 1366-5901 online © 2002 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/01431160110107699

challenging theoretical questions, which have been tackled recently by mathematical models of vegetation dynamics (Dunkerley 1997, Lefever and Lejeune 1997, Lejeune *et al.* 1999, Couteron and Lejeune 2001). However, besides theoretical considerations, quantified and objective descriptions of periodic semi-arid patterns are needed for both investigation and monitoring purposes. Until now, most available information has been based either on ground studies of small spatial extent or on visual interpretation of aerial photographs.

Visual analyses of aerial photographs have generally been restricted to an overall appraisal of a given patterned vegetation, with occasional manual measurement, i.e. along transects located on some photographs (MacFadyen 1950, Clos-Arceduc 1956, Hiernaux and Gérard 1999). Though vegetation patches often have sharp boundaries, their manual delineation is prohibitive for large areas, and as far as we know, has never been carried out. From this viewpoint, patterned vegetation of semi-arid tropics raises similar problems to those already underlined for Mediterranean shrublands by Kadmon and Harari-Kremer (1999). A more general conclusion is that visual interpretation has little relevance for highly fragmented vegetation covers, and digital image processing techniques are required, if the aim is to map and monitor extensive arid or semi-arid areas. However, as already pointed out by Kadmon and Harari-Kremer (1999), reference studies dealing with digitized air photographs have been scarce, in spite of some convincing results (Warren and Dunford 1986, Mast *et al.* 1997). The rare use of digitized photographs stems from technical difficulties, raised by geometrical deformation and reflectance levels distortion (Avery and Berlin 1992, Dymond 1992), that are not always easy to overcome. Nevertheless, for extensive areas of semi-arid Africa, aerial photographs are the only data source that offers simultaneously a metre-scale spatial resolution, a kilometre-scale extension and historical insights going back to the 1950s. Indeed, at that time, the discovery of periodic vegetation patterns was a direct outcome of the first systematic acquisition of aerial photographs (MacFadyen 1950, Clos-Arceduc 1956). Although the first author recommended a quantitative analysis of air photographs to investigate periodic vegetation, virtually nothing has been attempted since.

The aim of the present paper is to propose a specific approach that relies on historical aerial photographs to characterize and monitor semi-arid vegetation patterns with strong periodic features. This approach was designed with a special reference to the Sudano-Sahelian climatic belt in West Africa, within which patterned vegetation types are frequently encountered (Hiernaux and Gérard 1999). More precisely, the focus is on the north-west part of Burkina Faso, a region where diversified, i.e. spotted and banded, patterns are encountered (Couteron *et al.* 2001). The question of change quantification was addressed in relation to a major drought that occurred in the early 1980s (Morel 1992) with an impact on woody vegetation that was described by Couteron and Kokou (1997). Among the numerous methods for quantitative image analysis (Haralick 1979), we chose to use two-dimensional spectral analysis by Fourier transform (Niblack 1986, Mugglestone and Renshaw 1998) due to the strong presumption of periodicity that resulted from preliminary inspections of photographs. Furthermore, that the dominant wavelength of the vegetation pattern could be inversely correlated to rainfall has been suggested both by theoretical modelling (Lejeune *et al.* 1999) and field results (Hiernaux and Gérard 1999). Hence, emphasis was placed on characterizing the vegetation pattern through a spectrum of spatial wavelengths.

## 2. Material and methods

### 2.1. *Study area*

The study area is located in the northern part of the Yatenga Province (Burkina Faso), between 14°00′ and 14°15′ North latitude and 2°20′ and 2°30′ West longitude. Climate is semi-arid tropical with hot temperatures (mean annual values are 29–30°C) and a potential evapotranspiration (Penman) slightly under 2000 mm year$^{-1}$. There is a long dry season from October to May, with a short wet season from June to September. The highest monthly rainfall is usually observed in August. Average annual rainfall was 486 mm (SD = 92 mm) between 1986 and 1994 at Banh (14°05′N, 2°27′W). A broader perspective on rainfall variability can be obtained from Ouahigouya (50 km south; 13°35′N, 2°20′W) which has a continuous record from 1922. The average annual rainfall was 745 mm between 1950 and 1967, compared to only 550 mm between 1968 and 1985 (figure 1). The drought experienced during this second period was probably the worst that has occurred during this century for West Tropical Africa as a whole (Morel 1992). The early 1980s were particularly dry (less than 200 mm at Banh in 1984). A clear improvement has taken place since 1991 (figure 1).

Two main topographical conditions were encountered: (i) gentle slopes (0.5–0.8%) on old Palaeozoic sandstones (*'Grès de Firgoun'*); (ii) flat terrain made of Tertiary and Quaternary continental deposits ('Gondo plain'). On Firgoun sandstones, soils were very shallow (10–40 cm), overlying the hard bedrock. On the plain, soils were deeper (30–80 cm), yet poorly developed, overlying a more or less continuous petroferric contact, consisting of ironstone gravels and stones.

Vegetation belongs to the 'Sahel regional transition zone', with most woody species related to the 'Sudanian regional centre of endemism' (White 1983). Woody vegetation is dominated by shrubs and small trees which are often multistemmed. In the early 1990s, the Gondo plain displayed a savanna vegetation (grass and trees), punctuated by spots of bare soils, a physiognomy which was referred to as *brousse tachetée* (spotted bush) by Boudet (1972). The gentle slopes on Firgoun sandstones

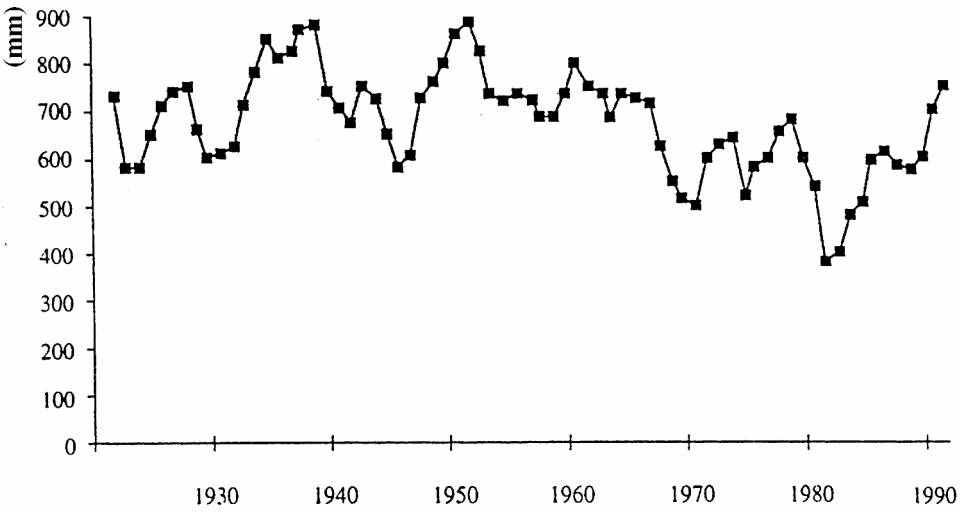


Figure 1. Rainfall series for Ouahigouya (13°35′N, 2°20′W) from 1922 to 1995 (three years moving average).

displayed dense bands of woody vegetation, roughly organized along the contour, and usually referred to as tiger bush (White 1970, Couteron *et al.* 1996, 2000). Detailed field description have been provided by Couteron and Kokou (1997) for the spotted savanna, and by Couteron (2001) for the tiger bush. In both situations, the woody stands had a basal area (BA) of 6.5 $m^2 ha^{-1}$, and were dominated by *Combretum micranthum* G. Don (30–50% of BA) and by *Pterocarpus lucens* Lepr. (30% of BA). The first species had mature individuals with average height and canopy radius of 2.8 m and 1.3 m, respectively, while the corresponding values were 5.3 m and 5.4 m for *P. lucens* (Couteron *et al.* 2000). Most woody species were without leaves from January to June. The herbaceous strata comprised annual grasses and forbs, the composition and development of which were highly dependent on rainfall abundance and distribution (Grouzis 1992, Seghieri *et al.* 1994). Standing crop phytomass ranged from 500 kg $ha^{-1}$ to 2500 kg $ha^{-1}$ (unpublished data), in relation to rainfall and local conditions. The herbaceous vegetation developed in July and senescence began in mid-September. The dead phytomass decreased progressively during the dry season. The study area had a rather low human population (less than 10 inhabitants $km^{-2}$), with no clearing for crops and no significant wood-cutting in the vegetation types under study. Pastoral utilization was moderate, with no use of fire. Woody vegetation was dominated by fire-sensitive species (Couteron and Kokou 1997).

### 2.2. *Aerial photographs and digitized images*

A first set of panchromatic aerial photographs at 1:50 000 scale was obtained from the Institut Géographique National (IGN, France). They were acquired on 11/12/1955 between 8h30 and 15h (U.T.) as part of an overall coverage referred to as AOF 1955/56-ND-30-XVI. The flight altitude was about 6500 m above ground level (a.g.l.) and the camera (referenced as A19) had a focal length of 125 mm and a film size of 190 mm. Contacts used for this paper were numbered 397, 398, 523 and 524. A second set of panchromatic printed outlooks with the same scale of 1:50 000 was acquired from the Institut Géographique du Burkina (IGB, Burkina Faso). The corresponding mission (84066-B) was conducted at the same period of the year (01/09/1985) between 11h and 12h (U.T.). The flight altitude was about 4500 m a.g.l. and the camera (SAG II 2110) had a focal length of 88 mm and a film size of 240 mm. Contacts 6687, 6689, 6703, 6704 were used. The first set of photographs depicted a vegetation that had experienced several years of favourable rainfall (pre-drought conditions; figure 1), whereas the second coverage, which was obtained after the driest rainy season on record (i.e. 1984), described a highly stressed vegetation. A final set of photographs was acquired on 10/08/94 through a low altitude flight (700–800m a.g.l.), using a PENTAX ILX camera (50 mm focus and 35 mm lens) placed in a door-mounted arrangement for unmodified tourism aircraft. The natural colour film (Kodak Gold, ASA 100) was machine processed toward printed outlooks at a scale of ca. 1:10 000. This coverage was carried out along a transect intersecting the study area in order to obtain high-resolution data depicting the post-drought vegetation.

Printed 1:50 000 photographs were digitized (grey-scale values in the range 0–255) at a resolution of 400 dots per inch (DPI) through an AGFA® Studiostar scanner (pixel side of 3.15 m in the field). The 1:10 000 colour outlooks were digitized into grey levels of reflectance, at a spatial resolution of 300 DPI (0.83 m in the field). Sampled areas corresponding to the low-altitude flight were digitized from the

1:50 000 photographs at a resolution of 1500 DPI ensuring the same pixel size. Although the main analyses were based on the low-resolution data from 1955 and 1985, useful additional interpretations came from the high-resolution images from the three dates. The scanner mechanism was checked for potential artefacts, as encountered by Mugglestone and Renshaw (1998): a plain sheet of white photographic paper was digitized, and the resulting grey-level image was submitted to spectral analysis (see below). The analysis detected neither preferential orientation nor systematic striation. Digitized data were directly submitted to quantified analyses without any prior correction of the grey-level values. Image enhancement (as histogram equalization or brightness adjustment; Niblack 1986), was used only for displaying purposes (as on figure 2 and figure 7). Reflectance corrections within and between photographs (Hall *et al.* 1991, Dymond 1992, Kadmon and Harari-Kremer 1999) were not attempted, since they do not constitute a prerequisite of the approach proposed here.

On digitized images (e.g. figure 2 and figure 7), bright pixels corresponded to bare soil, whereas dark ones contained woody vegetation, and intermediate grey-scale pixels contained continuous grass cover. Since continuous grass and woody vegetation have respective epigeneous phytomass averaging 1500 kg ha$^{-1}$ and 20 000 kg ha$^{-1}$ (Le Houerou 1989, Couteron unpublished data), grey-scale values appeared as a monotonic increasing function of phytomass.

### 2.3. *Windows extraction*

Six square 'windows', with a side of 400 pixels (1.26 km in the field), were extracted from each temporal version of the 1:50 000 coverage. They constituted a continuous belt transect 1.26 km wide and 7.56 km long, that encompassed the geological divide between the Gondo plain and the sloping area on Firgoun sandstones. (Windows w4 and w5, which were located around the divide, have their two temporal versions displayed on figure 2.) As much as possible, windows have been delineated in the central part of the photographs, by avoiding the fringes where the grey-level scale is usually distorted and where geometrical deformations may occur (Avery and Berlin 1992, Bolstad 1992). The 1985 version of the belt transect was geometrically rectified (using a second-order polynomial transformation) to match the 1955 version. Each window of the belt transect was then subdivided into 16 square 'quadrats' with a side of 100 pixels, i.e. 315 m in the field (figure 2). Quadrats were the basic items used for spectral analysis by Fourier transform (see below). Every quadrat had two temporal versions, corresponding respectively to the 1955 and 1985 aerial coverage.

### 2.4. *Spectral analysis*

Two-dimensional spectral analysis by Fourier transform is described in detail by Renshaw and Ford (1984), whilst Mugglestone and Renshaw (1998) presented an application on digitized aerial photographs. Given a square digital image consisting of an $n$ by $n$ array of pixels, let $Y_{jk}(j=1,...,n;k=1,...,n)$ be the pixel grey-scale values in the range 0–255. Spectral analysis is based on the Fourier coefficients, computed from mean-corrected observations ($X_{jk}=Y_{jk}-\bar{Y}$).

$$a_{pq}=n^{-2}\sum_{j=1}^{n}\sum_{k=1}^{n}X_{jk}\cos\{2\pi(pj/n+qk/n)\} \qquad (1)$$

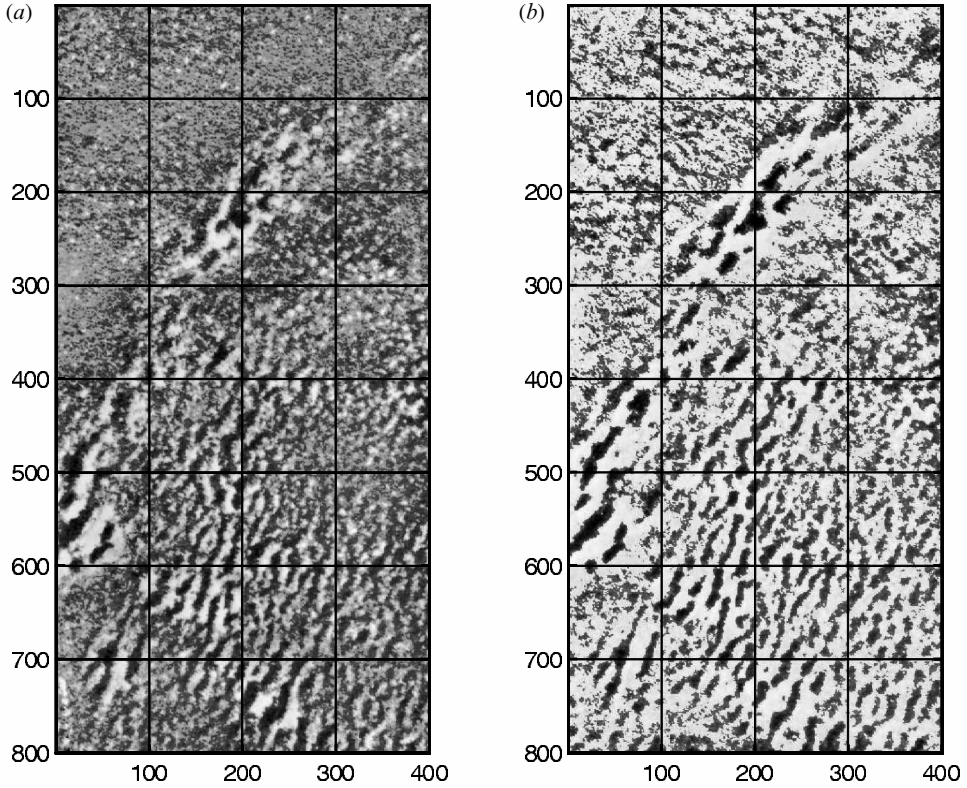


Figure 2. Intermediate stretch of the belt transect (2.52 km by 1.26 km; i.e. 800 by 400 pixels) constituted by windows w4 and w5. The grey-level scale expresses the transition from bare soil (light grey) to dense woody cover (dark grey), with intermediate levels for herbaceous vegetation. The figure is oriented to the north. (*a*) 1955 version (*b*) 1985 version.

and

$$b_{pq} = n^{-2} \sum_{j=1}^{n} \sum_{k=1}^{n} X_{jk} \sin\{2\pi(pj/n + qk/n)\} \tag{2}$$

Then the periodogram (sample spectrum) is given by $I_{pq} = n^2(a_{pq}^2 + b_{pq}^2)$. The value $I_{pq}/n^2$ represents the portion of the image variance accounted for by a cosine wave with frequency $(p,q)$ along the two Cartesian directions. Conversely, an image corresponding strictly to a simple cosine wave of the form $X_{jk} = \cos\{2\pi(p'j/n + q'k/n)\}$ would give rise to a periodogram with $I_{pq} = 0$ for all frequencies other than $(p = p', q = q')$. In that case, the travel direction of the wave, $\theta$, and the associated wavenumber, $r$, would be:

$$\theta = \tan^{-1}\{p'/q'\} \text{ and } r = \sqrt{p'^2 + q'^2} \tag{3}$$

The wavenumber $r$ expresses the number of times the pattern reproduces itself within the image. Periodogram values can be also expressed in polar coordinates as $Gr\theta$, where $r$ and $\theta$ are deduced from (3). Furthermore, the periodicity of a given pattern (quantified by $r$) and its main orientation (denoted by $\theta$) can be investigated separately, by binning $Gr\theta$ values into polar segments, as $-5^\circ < \theta \leqslant 5^\circ, ..., 165^\circ < \theta \leqslant 175^\circ$

and $0<r\leqslant 1$, $1<r\leqslant 2$, ... which have been deemed appropriate by Mugglestone and Renshaw (1998). This approach yields a 'radial spectrum' expressing the proportion of the image variance explained by successive wavenumbers, and an 'angular spectrum' measuring data variability with respect to the main directions of the geographical space (see figure 3 for an illustration through two computer-generated images). Not only do polar spectra provide a convenient summary of the two-dimensional periodogram, but they are also suitable for statistical tests (Mugglestone and Renshaw 1998): in the absence of spatial structure (figure 3(*a*)), each $2Gr\theta/\sigma^2$ value is distributed as $\chi^2_2$ ($\sigma^2$ being the variance of the image). Thus a bin containing $m$ such values has its sum distributed as $\chi^2_{2m}$ (with an expected value of $2m$). Thus, if bin $\theta$ contains $m=n_\theta$ values, the angular spectrum

$$I(\theta)=(m\sigma^2)^{-1}\sum_{bin(\theta)} Gr\theta \sim (2m)^{-1}\chi^2_{2m} \tag{4}$$

has expected value one in the absence of spatial structure. Hence, results significantly above one for specific $\theta$ values indicate directional effects at these angles (figure 3(*b*)). Similarly, the radial spectrum

$$I(r)=(m\sigma^2)^{-1}\sum_{bin(r)} Gr\theta \sim (2m)^{-1}\chi^2_{2m} \tag{5}$$

where $m=n_r$ is the number of values in bin $r$, provides information on periodicity of the pattern.

A complex structure superimposing a periodic pattern (bands) and a vertical trend is displayed on figure 3(*b*). To the human eye, the periodic pattern is far more impressive than the trend, although both have the same amplitude. The trend is denoted by a high value attached to the first wavenumber on the radial spectrum, and by a spike corresponding to a direction of 90° on the angular spectrum. The periodic pattern determines a spike for $r=5$ on the radial spectrum, and for $\theta=0^\circ$ on the angular spectrum. A trend is a very common feature for digitized images extracted from real-world aerial photographs, and the smallest wavenumbers ($r=$ 1–2) display, usually, high spectral values (Mugglestone and Renshaw 1998). The trend often results from non-stationary grey-levels, a usual feature of air photographs (Avery and Berlin 1992), or from a macro-heterogeneity at the scale of the extracted image. Since the two causes cannot be discriminated by spectral analysis, it was deemed preferable to ignore the first two wavenumbers in all quantified analyses, and to retain a reference quadrat size at least three times larger than the biggest wavelength deemed of interest, therefore quadrats of 315 m. The largest wavelength taken into account was hence 105 m. Polar spectra were thus re-scaled with reference to the variance accounted for by a limited range of wavenumbers (say, $\sigma'^2$ for $r_0-r_1$), instead of the total variance $\sigma^2(\sigma'^2=\Sigma_{r_0-r_1}Gr\theta)$. Hence, for $m$ being the number of periodogram values in the respective bins:

$$I(r/r_0-r_1)=(m\sigma'^2)^{-1}\sum_{bin(r\cap(r_0-r_1))} Gr\theta \sim (2m)^{-1}\chi^2_{2m} \tag{6}$$

and

$$I(\theta/r_0-r_1)=(m\sigma'^2)^{-1}\sum_{bin(\theta\cap(r_0-r_1))} Gr\theta \sim (2m)^{-1}\chi^2_{2m}. \tag{7}$$

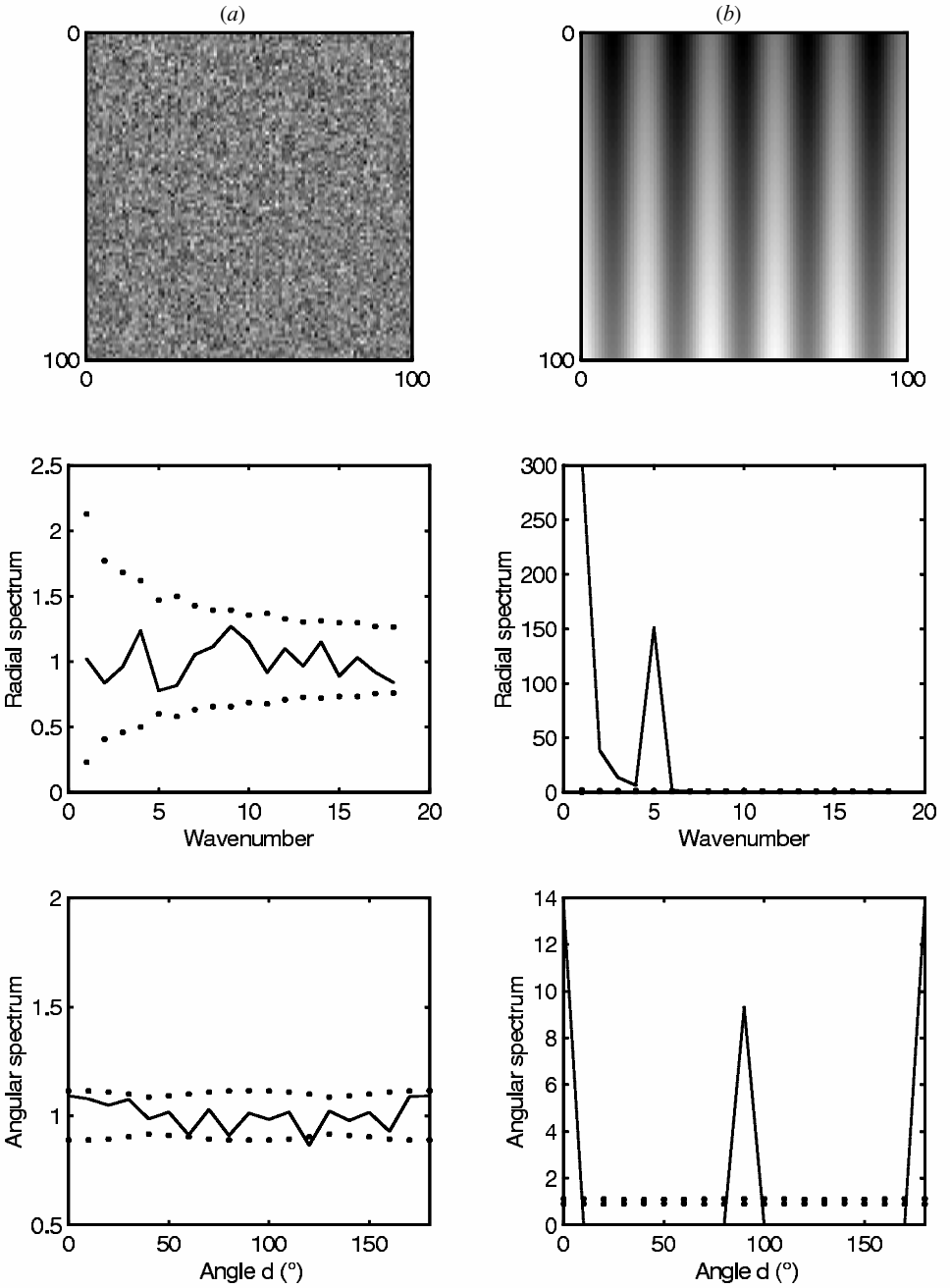


Figure 3. Examples of polar spectra from two computer-generated images (100 by 100 pixels). The solid lines denote the spectra whereas the dotted lines stand for the 5% bilateral interval around 1 (expected value in the absence of spatial structure). (*a*) No spatial structure: each pixel value was generated according to a Gaussian white noise (WN), with $\mu = 10\sigma$. (*b*) Superimposition of a cosine wave (bands) of amplitude $\beta$, a WN with $\sigma = 0.1\beta$, and a linear trend of amplitude $\beta$.

2.5. *Systematic comparison of radial spectra*

A systematic periodicity analysis was envisaged by comparing all the radial spectra obtained for the whole set of quadrats from the low-resolution images. A general data table was built, for which each row was the radial spectrum of a given quadrat, while each column contained the spectral values related to a given wavenumber. Quadrats are statistical observations that are characterized by their spectral profile, namely, the portion of their variance explained by successive wavenumbers. Conversely, wavenumbers can be seen as quantitative variables characterizing the quadrats. One way to achieve a systematic comparison of quadrats is to submit the quadrats by wavenumbers table to a Principal Components Analysis (PCA; Manly 1994) that looks for a limited number of synthetic new variables accounting for a large share of the variability between spectral profiles of quadrats. New variables are linear combinations of initial variables, i.e. wavenumbers, and are called principal components or 'axes'. They are computed from the matrix of correlations between the variables of the table (see Manly 1994 for details). Plotting quadrats against their values for principal axes ('PCA scores') enables an overall comparison between spectral profiles, since quadrats having similar profiles are characterized by similar PCA scores and *vice versa* for contrasted profiles.

The quadrats by wavenumbers table had two temporal versions (1955 and 1985) which could have been submitted separately to PCA. Another possibility, deemed preferable, was to concatenate the two versions columnwise prior to analysis. The resulting diachronic table contained the two temporal versions of the spectral profile of each quadrat. This table had twice as many rows (quadrats) as each separate temporal table, although based on the same columns, i.e. wavenumbers. PCA scores of the rows of the diachronic table can be used to compare either quadrats at a given date or temporal versions of a given quadrat. Windows were considered via the PCA scores of their constituting quadrats. At a given date, differences between distinct windows were addressed using the Wilcoxon's test with correction for multiple comparisons (Sokal and Rohlf 1995), while changes between temporal versions of a given window were tested using the sign-rank test for paired data (Sokal and Rohlf 1995).

2.6. *Search for dominant orientations*

A systematic search of potential directional effects was also carried out: for a given window, individual periodograms obtained from each quadrat were averaged. The angular spectrum of the window was then computed from the average periodogram, since a direct average of angular spectra would be meaningless. Such an approach can detect dominant orientations that are consistent throughout a window, but would be irrelevant in the case of inconsistent, though locally significant, orientations.

All computations involved in the preparation of this paper were carried out using the Matlab® software (Anon. 1997), with both pre-programmed and personal routines.

## 3. Results

3.1. *Systematic periodicity analysis at low resolution*

The first factorial plane of the PCA accounted for 76.8% of the variability of the overall table (58.5% and 18.3% for the first two axes, respectively). Subsequent axes proved of low interest since none of them yielded more than 5% of total variability.

The first axis displayed a high negative correlation with large wavenumbers ($r >= 11$; figure 4(*a*)), and a high positive correlation with small wavenumbers ($r = 3–5$). The second axis correlated strongly with intermediate wavenumbers ($r = 7–8$). Hence, in terms of quadrat scores (figures 4(*c*) to 4(*h*)), the first factorial plane expressed a progressive transition from quadrats dominated by large wavenumbers (on the left side), to quadrats dominated by small wavenumbers (on the right side). Quadrats characterized by intermediate wavenumbers had high positive scores on the second axis (on the upper side of the plane). In the factorial plane (figure 4(*b*)), windows from the Gondo plain (w1 to w3) were located on the left side, denoting a fine-grained structure, whereas windows from the sloping area (w5 and w6) were closer to the right side, due to a coarser pattern. One of them (w6) had nevertheless its 1955 version positioned on the upper side of the plot. The window w4 (displayed in the upper half of figure 2), since heterogeneous, occupied an intermediate position along the first axis (figure 4(*b*)).

At each date, window positions were compared on a pairwise basis by submitting the PCA scores of their quadrats to Wilcoxon's test. For the 1955 versions, no significant difference appeared along the second axis, with the exception of widow w6, which had higher PCA scores than windows w1 to w4 ($p < 0.0001$) and also than window w5 ($p < 0.01$). Along the first axis, significant differences were observed between windows from the Gondo plain and windows from the sloping area ($p < 0.0001$). Among the latter, one window (w1) proved different from the two others ($p < 0.05$). The composite window (w4) displayed less difference with the windows from the plain ($p > 0.05$ with w2 and w3; $p < 0.01$ with w1) than with the two windows from the slope ($p < 0.01$). This result was consistent with the visual impression drawn from figure 2(*a*) (upper half) of a greater extension of spotted fine-grained vegetation. The Wilcoxon's test, when applied to the scores of the 1985 quadrats, yielded similar results. There were two notable exceptions: (i) window w6 no longer appears as an outlier along the second axis; (ii) window w4 significantly departed from the windows from the plain ($p < 0.001$), and no longer differed from windows from the slope ($p > 0.05$). This last result was, once again, in agreement with the visual impression suggested by figure 2(*b*), since the whole window appeared, at that date, dominated by two distinct banded patterns. However, the change was revealed by an analysis solely based on periodicity, whereas the human eye is also impressed by dominant orientations.

Between 1955 and 1985, all windows experienced a significant shift toward the right side of the first axis (sign-rank test for paired data), indicating an increase in the image variability accounted for by the smallest wavenumbers. However, the change was more dramatic for windows from the plain than for windows from the slope ($p < 0.0001$ against $p < 0.01$). In the plain, all windows moved clearly rightward, the two temporal versions corresponding to distinct groups of points that did not overlap (figures 4(*c*) to 4(*e*)). Windows from the slope were obviously more heterogeneous in 1955, since the corresponding quadrats were more scattered in the factorial plane (figures 4(*g*) and 4(*h*)). Indeed, at that date, most quadrats bear a banded pattern, while some others were covered by a fine-grained and spotted vegetation (figure 2(*a*)). The latter quadrats displayed the largest shifts in factorial plane between 1955 and 1985. As a consequence, the two versions of a given window corresponded to overlapping groups of points in spite of a significant shift along both axes ($p < 0.01$). Furthermore, windows from the slope exhibited also a shift along the second axis, which was particularly significant for w6 (figure 4(*h*); $p < 0.0001$); no comparable shift was found for windows from the plain.

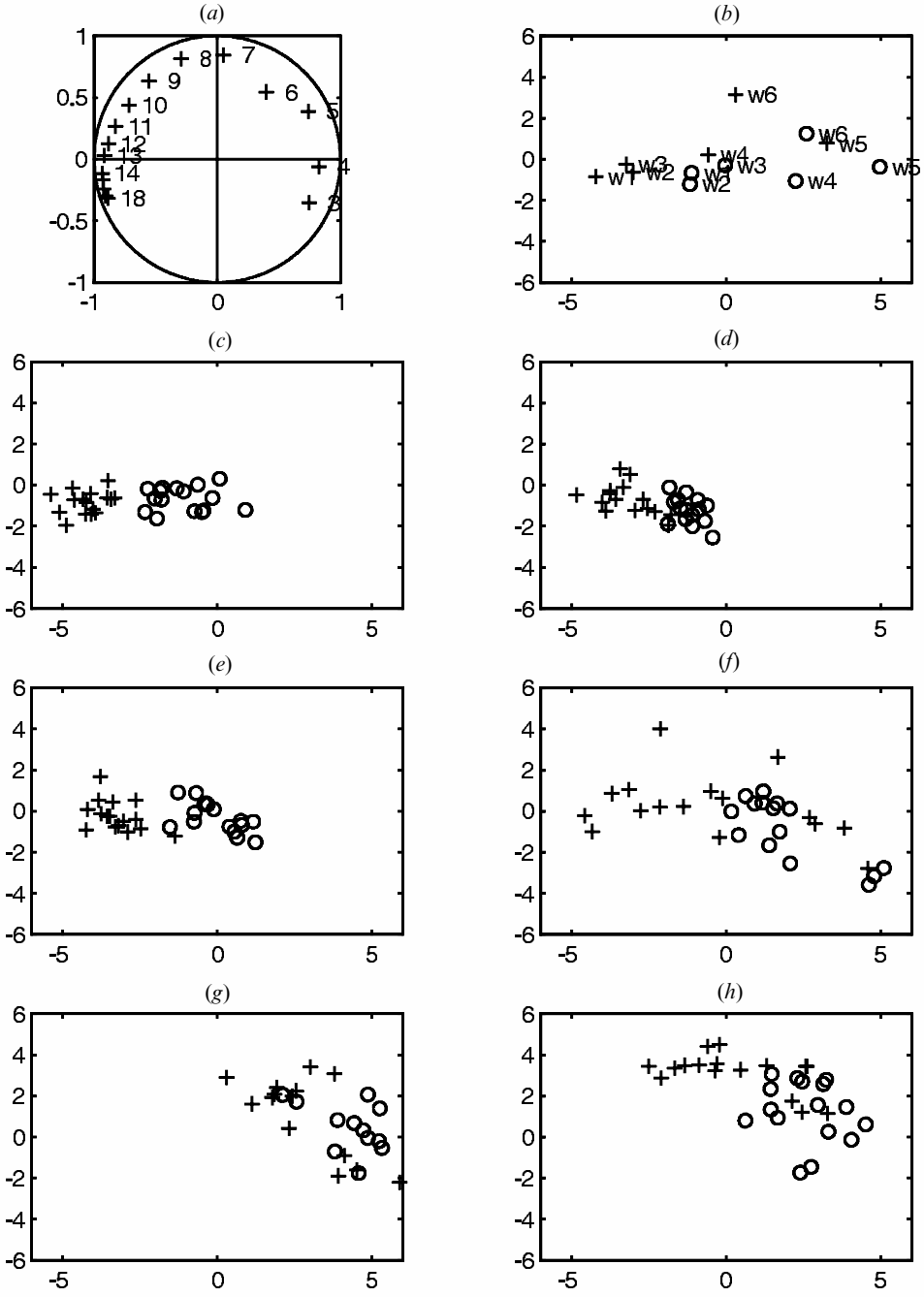


Figure 4. First factorial plane of the PCA applied to the overall table of radial spectra. (*a*) Correlation circle for wavenumbers $r = 3 \pm 18$. (*b*) Average position of the windows. (*c*), (*d*), (*e*), (*f*), (*g*), (*h*) Positions of quadrats from windows w1, w2, w3, w4, w5, w6, respectively. +: 1955 version; ○: 1985 version.

3.2. *Systematic orientation analysis*

As expected, a dominant orientation appeared clearly for the window corresponding to the tiger bush (w5) on Firgoun sandstones (figure 5(*a*)), with a spike between 150° and 170°. Most bands were, hence, perpendicular to that range of direction, corresponding roughly to the main orientation of the slope (field assessment). Furthermore, the two temporal versions of the spectrum were similar. Less expected was the existence of a consistent dominant orientation for the spotted vegetation in the Gondo plain (figure 5(*b*)). Directions in the range 60–90° proved, nevertheless, consistently significant whatever the window. This main orientation appeared on the upper-left part of figure 2(*b*) (1985 version of w4), as fragmented bands of woody vegetation (dark grey), with a lower intensity than the ones observed on Firgoun sandstones (lower half of figure 2(*b*)). The former banded system was not clearly apparent to the eye on the 1955 version of the w4 window. However, the two versions of the angular spectrum displayed only minor differences, indicating that the dominant orientation was already present in the 1955 photographs, though masked by a more impressive pattern made of bare spots.

3.3. *Mapping patterns and change*

For the whole belt transect, several maps expressing the pattern periodicity, the intensity of change, and the main orientations have been deduced from the previous results. An excerpt, homologous to figure 2 (w4 and w5), is presented in figure 6. In 1955, PCA scores underlined the main divide between the fine-grained spotted vegetation, that dominate in the northern part of the picture, and the coarser banded cover that was located down south (figure 2(*a*)). The map of principal orientations (not presented) was, moreover, fairly redundant. However, intermediate values were found on the right side on the map (figure 6(*a*)), which was perceived as dominated by an intermediate structure (figure 2(*a*)). The map of change (figure 6(*b*)) was based on the shift along the first PCA axis. It discriminated mostly the same areas, and emphasized the greater intensity of change for the spotted vegetation.

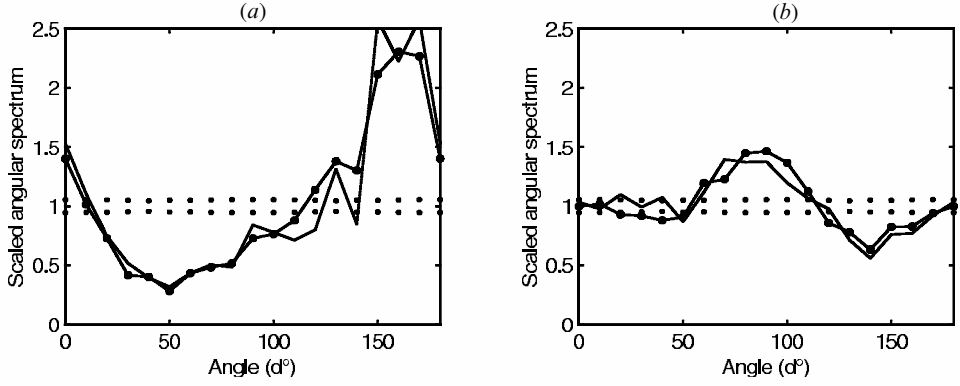


Figure 5. Angular spectra from averaged periodogram for two particular windows. (*a*) Window w5 on Firgoun sandstones. (*b*) Window w1 from the Gondo plain. The symbol '○' refers to the 1955 version, whilst the plain line refers to the 1985 version. The dotted lines denote the 5% bilateral interval around 1 (expected value in the absence of a dominant orientation).

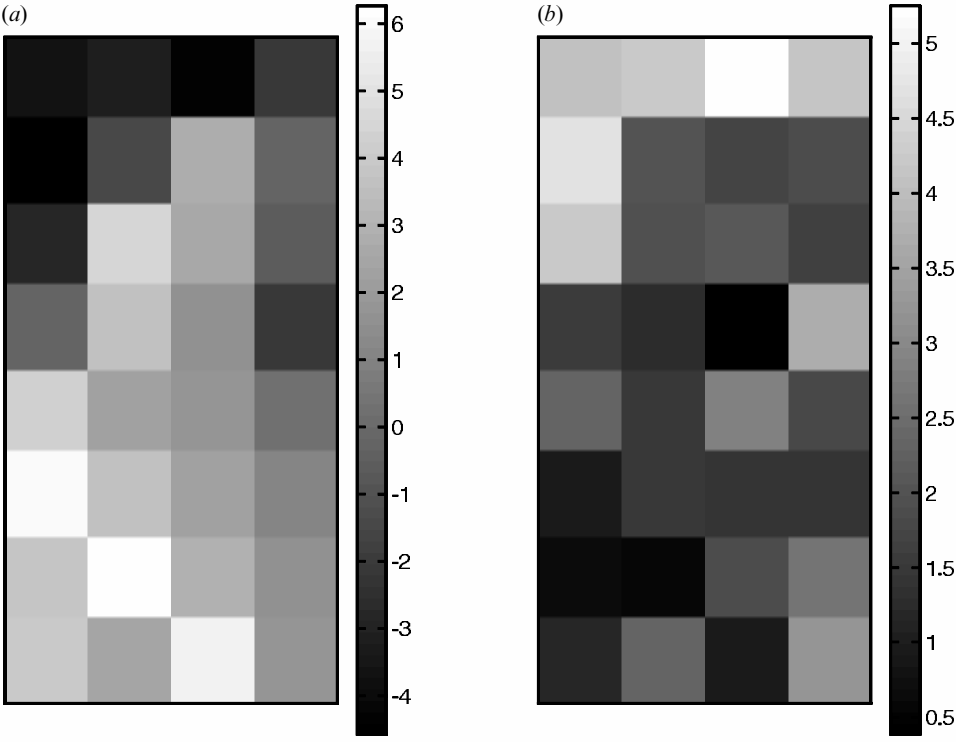


Figure 6. Maps of PCA scores along the first factorial axis for the part of the belt transect displayed on figure 2. (*a*) Scores in 1955. (*b*) Amplitude of the shift between 1955 and 1985.

3.4. *High-resolution analysis for particular quadrats*

Four particular 315 m by 315 m quadrats were considered for detailed analysis and interpretation based on high resolution digitized images (0.83 m in the field, i.e. 1500 DPI for the 1955 and 1985 photographs, and 300 DPI for the 1994 coverage). The selected quadrats corresponded to field plots of equivalent size, with a detailed mapping of woody individuals carried out in 1993–1995 (Couteron and Kokou 1997, Couteron 2001). In spite of a consistent pixel size, the actual resolution was different at the three dates, since the scanning resolution substantially exceeded the film resolution of 1955 and to a lesser extent the resolution of 1985. However, a major incidence on the following analyses was unlikely, since only wavenumbers $r < 19$ (i.e. a wavelength above 17 pixels) were taken into account.

A first quadrat was extracted from the w1 window, hence from a vegetation which had the finer grain in 1955 (figure 4(*b*)). At that date, the herbaceous cover was widespread, the punctuation of bare soil being of limited extension and density, with no obvious periodicity (figure 7(*a*)). Woody vegetation was mainly composed of small clumps of trees, with few extended thickets or groves. Hence the fine-grained texture detected on figure 4, with a rather flat radial spectrum. Just before 1985, the herbaceous cover collapsed, resulting in a more contrasted vegetation cover, with a radial spectrum expressing the overall pattern of woody vegetation, and thus clearly dominated by wavenumbers ranging from 4 to 7 (i.e. wavelengths of 45–80 m). The 1994 spectrum displayed intermediate characteristics (figure 8(*a*)), in relation to a partial recovery of the herbaceous vegetation (figure 7(*a*)). However, this conclusion

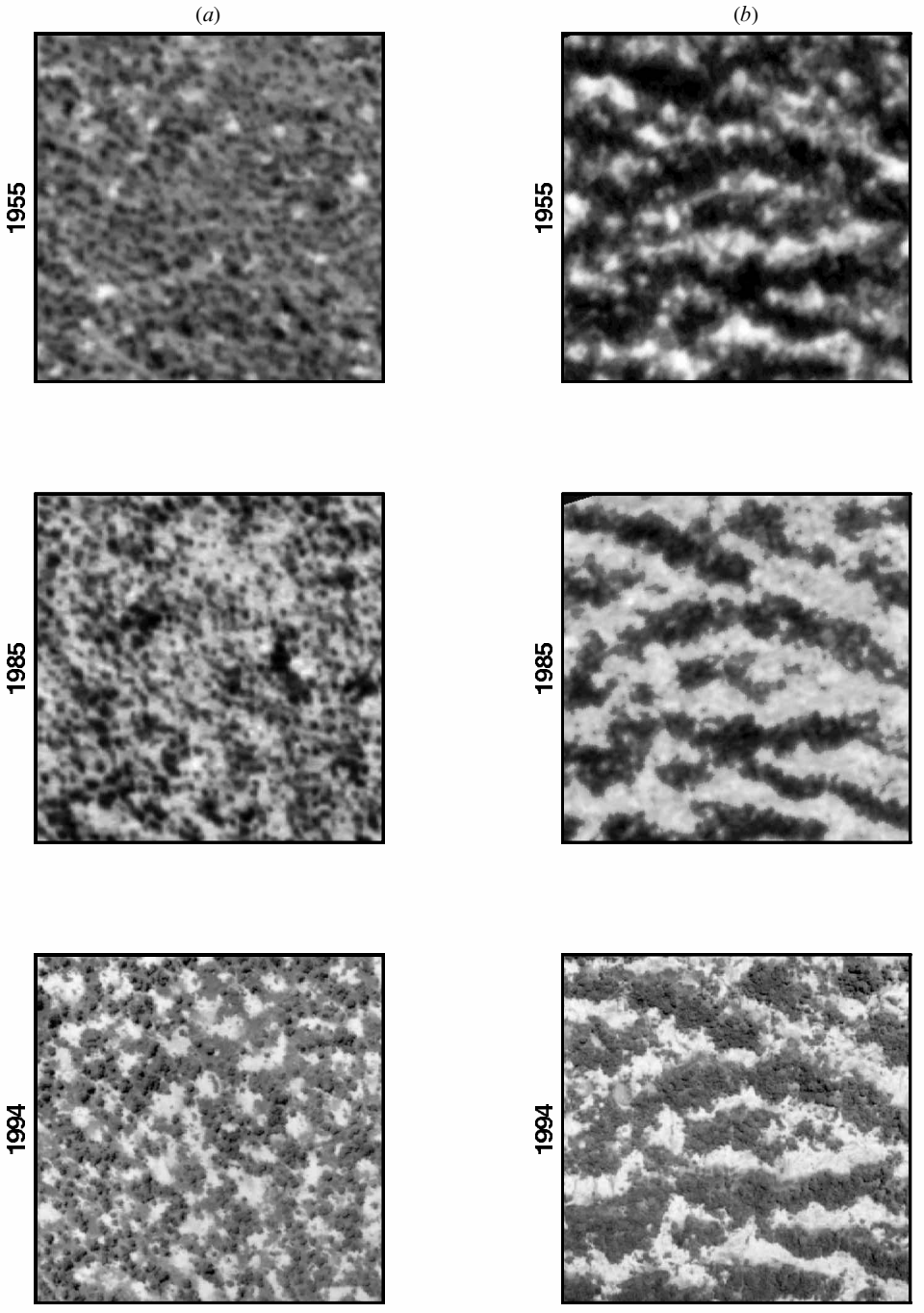


Figure 7. Three temporal versions of two specific quadrats (315 m by 315 m) digitized at high resolution. (*a*) Quadrat from the plain (oriented to the north). (*b*) Quadrat from the sloping area (oriented according to the slope).

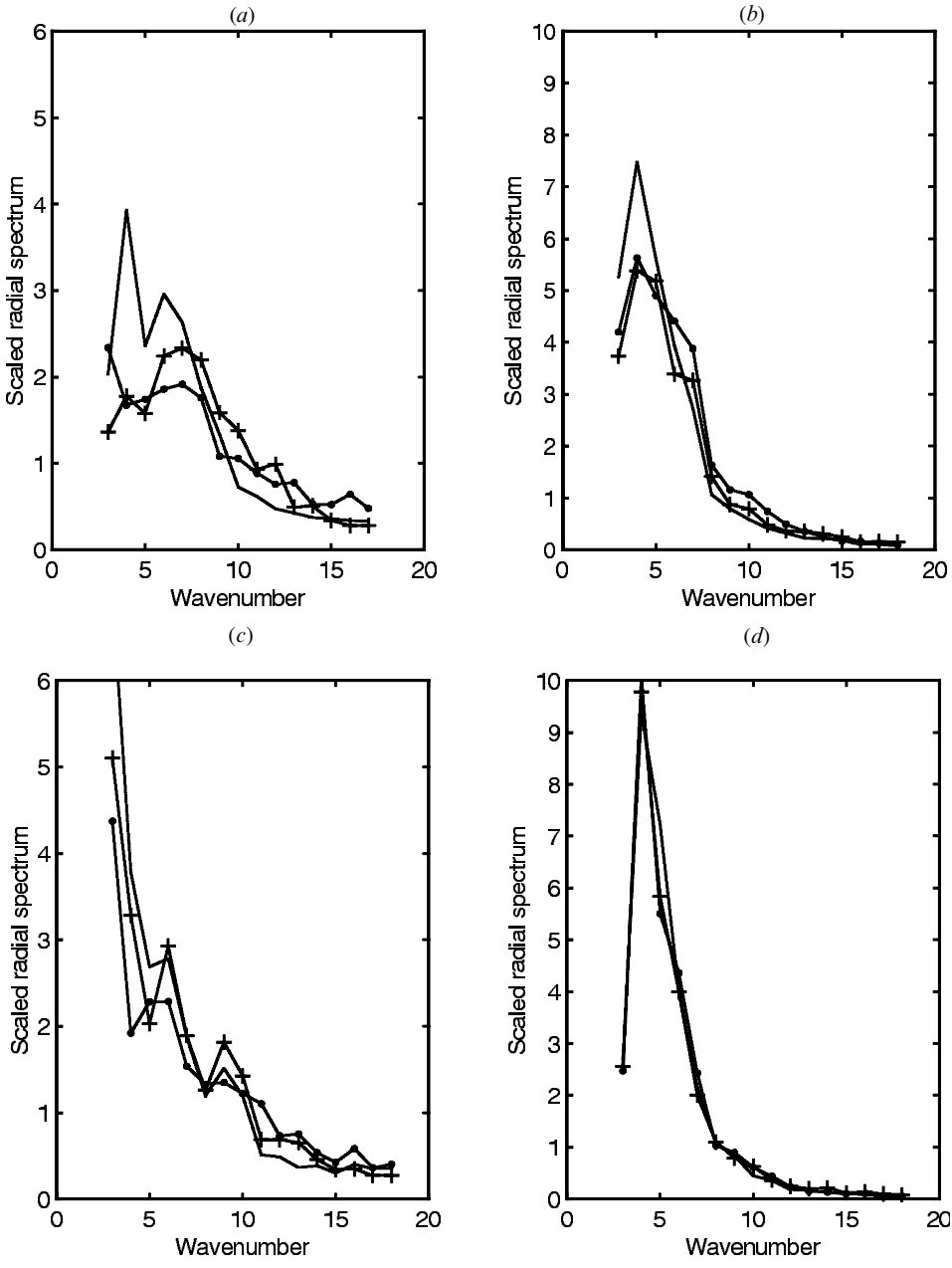


Figure 8. Diachronic radial spectra (1955, 1985 and 1994) of four quadrats analysed at high resolution. The 1955 and 1994 versions are denoted by '●' and '+', respectively while the plain line stands for 1985.

was only valid for wavenumbers $r < 8$, since for $r = 8$–12 (i.e. wavelengths of 25–40 m), the radial spectrum had larger values in 1994 than in 1955. This range of values characterized the pattern of large bare spots that has established in the post-drought decade. A similar pattern of diachronic change (figure 8(*c*)) was also identified through the radial spectra of a second quadrat (not presented on figure 7), extracted

from w2; the only difference being that the woody vegetation had a larger wavelength of at least 105 m. The radial spectrum had also higher values in 1994 than in 1955 for $r = 6$–$10$ (i.e. a wavelength of 30–50 m).

A third quadrat was extracted from w5, the structure of which appeared in 1955 as mostly banded yet locally spotted (figure 7(*b*)). At that date, the banded structure was mostly dominated by woody vegetation, with a herbaceous cover extending around the bands (on both sides) or in some inter-bands. In the latter situations, the grassy cover, when present, was regularly punctuated by spots of bare soils (e.g. on the left side of the picture). The whole pattern was dominated by wavenumbers in the range 3–7 (i.e. wavelengths of 45–105 m with a peak for 75 m; figure 8(*b*)). In 1985, wavenumbers around $r = 4$ were far more dominant, expressing a main woody structure with a wavelength around 75 m. The recovery of the herbaceous vegetation was less dramatic than in the Gondo plain, and in 1994, grass cover was mostly restricted to the upslope fringe of the wooded bands. As a consequence, no pattern of bare spots was apparent in 1994, and the related radial spectrum had lower values than its 1955 homologue (for $r > 5$). A fourth quadrat (not presented on figure 7), also extracted from w5, corresponded in 1955 to a more precise banded pattern, with a limited herbaceous cover. This quadrat experienced only minor drought-related changes, and the radial spectrum, expressing the dominance of wavelengths in the range 60–80 m, remained virtually unchanged (figure 8(*d*)).

## 4. Discussion

In 1955, the vegetation of the Gondo plain featured an extended herbaceous cover, that was punctuated by spots of bare soil of varying density (figures 2(*a*) and 7).Woody vegetation was not absent, but its overall structure was not easily perceived by the human eye. However, spectral analysis indicated a dominant orientation, corresponding to a pattern of fragmented bands that became more apparent on the 1985 photographs. At that date, the herbaceous cover had been drastically reduced by the very dry years observed in the early 1980s (figures 2(*b*) and 7; see also Serpantié *et al*. (1992) for a field account). The main woody pattern pre-existed the drought (figure 5), and was revealed by the double collapse of the herbaceous cover and its associated population of scattered trees and shrubs. In the field, a much higher mortality rate was observed for isolated woody individuals than for trees growing in dense stands (Couteron and Kokou 1997). During the 1955–1985 interval, the vegetation shifted from a superimposition of several patterns with distinct wavelengths, to a simplified physiognomy dominated by the fragmented bands of woody vegetation. Correlatively, the radial spectrum became more concentrated around the corresponding range of wavelengths (80–105 m; figure 8(*a*)), with a systematic shift in the reduced spectral space obtained from PCA (figure 4). During the post-drought decade, increasing rainfall enabled the herbaceous vegetation to bounce back, although its cover remained less extended than in 1955. Spots of bare soil were larger and more numerous than forty years before (figure 7(*a*)). They constituted a distinct periodic pattern, with a wavelength ranging from 25 to 50 m, that has been extensively considered by Couteron and Lejeune (2001) in relation to a propagation-inhibition model that can account for its emergence.

Tiger bush on Firgoun sandstones displayed less dramatic changes. This was particularly true for banded systems that already appeared as highly contrasted (dense woody cover versus bare soil) in 1955. In that case, the radial spectrum remained virtually the same through time (figure 8(*d*)). However, the tiger bush

domain on gentle slopes featured also areas the vegetation of which displayed more complex patterns in 1955, i.e. intermingled bands without a dominant orientation (as on the right side of figure 7(*b*)), and/or inter-bands that were occupied by a spotted herbaceous cover. In the field, it was established from soil pits that the occurrence of such patterns was related to more favourable substratum conditions, with a more weathered bedrock (Couteron 2001). Such situations were indicated by intermediate scores in the reduced spectral space obtained through PCA (figure 4(*d*)), and also by a notable shift between 1955 and 1985 (figures 4(*b*) and 6(*b*)). This shift was mostly determined by the collapse of the vegetation that grew in the inter-bands. Contrary to what was observed in the Gondo plain, the herbaceous vegetation did not substantially bounce back in the inter-bands during the more favourable post-drought decade.

The present study corroborates the conclusion drawn by Mugglestone and Renshaw (1998) about the relevance of two-dimensional spectral analysis for studying periodicity and orientation in digitized air photographs. Furthermore, the present paper addressed the problem of change quantification, a question that is of the utmost interest when dealing with natural vegetation. To do so, we relied mainly on a systematic comparison of radial spectra in a reduced spectral space obtained through Principal Component Analysis. Change quantification was based on factorial coordinates, and was tested through the usual non-parametric tests. Orientation analysis based on angular spectra also proved useful by revealing underlying features that were not obvious to the human eye (as for the 1955 coverage of the Gondo plain), and by demonstrating the great stability of dominant orientations.

One great advantage of the approach is its practicability without complex corrections, that are unavoidable when the analysis relies on a direct comparison of grey-levels (Dymond 1992, Hall *et al.* 1991, Kadmon and Harari-Kermer 1999). This is due to the fact that spectral analysis does not directly compare grey levels, which are inconsistent throughout an aerial coverage, but model their spatial variability in terms of sine and cosine functions. Consequently, trends and non-stationary components, whatever their origin, are mostly expressed by the smallest wavenumbers ($r = 1–2$; figure 3(*b*)), which can be easily ignored in subsequent analyses (as was done for the PCA and for the computation of polar spectra). Indeed, image reconstruction by Inverse Fourier Transform (IFT; Niblack 1986), after suppression of the smallest wavenumbers from the two-dimensional periodogram, is a convenient way of removing an overall trend and of rendering the grey scale more consistent within a digitized image. Nevertheless, image reconstruction through IFT was not necessary for the present paper. Furthermore, both synchronic and diachronic comparisons rely on the inner periodic structure of the quadrats, without assuming any consistency of the grey scale between them, and between photographs. Lastly, a highly accurate spatial matching between two temporal versions of the same window is far less critical than for a pixelwise comparison. Correlatively, geometrical distortions generally attached to air photographs should have only minor consequences on the results, especially when dealing with the gentle landscapes within which periodic vegetation is encountered (White 1971). A suitable quadrat size is nevertheless required for systematic analysis, since quadrats should be large enough to have the main pattern reproducing itself at least three or four times (smaller wavenumbers are to be discarded), yet as small as possible to increase the resolution of subsequent mapping (figure 6). It is also thinkable to base maps on moving quadrats after having computed PCA axes from non-overlapping quadrats of the same size. Thus, PCA axes would be obtained

from a limited number of statistical units (non-overlapping quadrats), and then used to compute PCA scores for a large number of non-independent moving quadrats. Thereby, spatial resolution of the maps could be enhanced.

The present approach, based upon Fourier transforms, has been designed for landscapes in which the existence of a periodic structure is likely, and its usefulness for a vegetation cover devoid of obvious periodicity is still to be established. However, patterned arid and semi-arid vegetation types dominate extensive areas on three continents, and are found in the transition zone between deserts and tropical savannas (White 1971). As a consequence, they can be thought of as particularly exposed to global change and worth monitoring on a systematic basis. Both theoretical results (Lejeune *et al.* 1999) and field evidence (White 1971, Hiernaux and Gérard 1999) suggest that the dominant wavelengths of the pattern should be greater in more arid conditions, a conclusion that is also supported by the present study. Consequently, periodicity of vegetation can be seen as a key indicator of climatic conditions experienced by extensive areas for which meteorological records are limited, if not fully absent. From this standpoint, the comprehensive air coverage that has been achieved in the 1940s and the 1950s for extensive areas of arid and semi-arid Africa represents an invaluable source of historical information. The present approach proved able to extract consistent quantified parameters from printed documents of various origin, scale and film quality. Though high-resolution data were useful for enhanced interpretations, the main results were drawn from digitized images with a spatial resolution of about 3 m. Such a modest value was purposely retained to match the pixel size that could be expected in the near future using panchromatic satellite imagery (e.g. 2.5 m with SPOT5 P). Hence, satellite data are likely to become the natural prolongation of the historical series formed of air photographs.

**Acknowledgments**

The author is particularly indebted to Bruno Henquin who greatly helped with the acquisition of the 1994 aerial photographs. Moira Mugglestone (University of Leicester, UK) provided important information on two-dimensional spectral analysis and John Dymond (Landcare Research, NZ) made valuable comments on an earlier draft of this paper.

**References**

Anonymous, 1997, *Matlab. Image Processing Toolbox* (Natick, MA: The Mathworks Inc.).

Avery, T. E., and Berlin, G. L., 1992, *Fundamentals of Remote Sensing and Airphoto Interpretation* (New York: Macmillan).

Bolstad, P. V., 1992, Geometric errors in natural resource GIS data: tilt and terrain effects in aerial photographs. *Forest Science*, **38**, 367–380.

Boudet, G., 1972, Désertification de l'Afrique tropicale sèche. *Adansonia, ser.2*, **12**, 505–524.

Clos-Arceduc, M., 1956, Etude sur photographies aériennes d'une formation végétale sahélienne: la brousse tigrée. *Bulletin de l'Institut Français d'Afrique Noire*, **T.XVIII, ser.A, 7**, 678–684.

Couteron, P., and Kokou, K., 1997, Woody vegetation spatial patterns in a semi-arid savanna of Burkina Faso, West Africa. *Plant Ecology*, **132**, 211–227.

Couteron, P., Mahamane, A., and Ouedraogo, P., 1996, Analyse de la structure de peuplements ligneux dans un fourré tigré au Nord-Yatenga (Burkina Faso) : état actuel et conséquences évolutives. *Annales des Sciences Forestières*, **53**, 867–884.

Couteron, P., Seghieri, J., Mahamane, A., and Ouedraogo, P., 2000, Differences between thickets of banded vegetation in two sites in West Africa. *Journal of Vegetation Science*, **11**, 321–328.

Couteron, P., 2001, Using spectral analysis to confront distributions of individual species with an overall periodic pattern in semi-arid vegetation. *Plant Ecology*, **156**, 229–243.

COUTERON, P., and LEJEUNE, O., 2001, Spotted patterns in semiarid vegetation explained by a propagation-inhibition model. *Journal of Ecology*, **89**, 616–628.

COUTERON, P., DESHAYES, M., and ROCHES, C., 2001, A flexible approach for woody cover assessment in semi-arid West Africa. Application in northern Burkina Faso. *International Journal of Remote Sensing*, **22**, 1029–1051.

DUNKERLEY, D. L., 1997, Banded vegetation: development under uniform rainfall from a simple cellular automaton model. *Plant Ecology*, **129**, 103–111.

DYMOND, J. R., 1992, Nonparametric modeling of radiance in hill country for digital classification of aerial photographs. *Remote Sensing of Environment*, **39**, 95–102.

GROUZIS, M., 1992, Germination des plantes annuelles sahéliennes. In *L'aridité, une contrainte au développement*, edited by E. Le Floc'h, M. Grouzis, A. Cornet and J. C. Bille (Paris: ORSTOM, Coll. Didactiques), pp. 221–224.

HALL, F. G., BOTKIN, D. B., STREBEL, D. E., WOODS, K. D., and GOETZ, S. J., 1991, Large-scale patterns of forest succession as determined by remote sensing. *Ecology*, **72**, 628–640.

HARALICK, R. M., 1979, Statistical and structural approaches to texture. *Proceedings of the IEEE*, **67**, 786–804.

HIERNAUX, P., and GÉRARD, B., 1999, The influence of vegetation pattern on the productivity, diversity and stability of vegetation: The case of 'brousse tigrée' in the Sahel. *Acta Oecologica*, **20**, 147–158.

KADMON, R., and HARARI-KREMER, R., 1999, Studying long-term vegetation dynamics using digital processing of historical aerial photographs. *Remote Sensing of Environment*, **68**, 164–176.

LE HOUEROU, H. N., 1989, *The Grazing Land Ecosystems of the African Sahel* (Berlin: Springer-Verlag).

LEFEVER, R., and LEJEUNE, O., 1997, On the origin of tiger bush. *Bulletin of Mathematical Biology*, **59**, 263–294.

LEJEUNE, O., COUTERON, P., and LEFEVER, R., 1999, Short range co-operativity competing with long range inhibition explains vegetation patterns. *Acta Oecologica*, **20**, 171–183.

MACFADYEN, W. A., 1950, Vegetation patterns in the semi-desert plains of British Somaliland. *Geographical Journal*, **116**, 199–210.

MANLY, B. F. J., 1994, *Multivariate Statistical Methods. A Primer* (London: Chapman & Hall).

MAST, N. J., STAENZ, K., and ITTEN, K. I., 1997, Tree invasion within a pine/grassland ecotone: an approach with historic aerial photography and GIS modeling. *Forest Ecology and Management*, **93**, 181–194.

MOREL, R., 1992, *Atlas agro-climatique des pays de la zone du CILSS. Notice et commentaire* (Niamey, Niger: AGRHYMET).

MUGGLESTONE, M. A., and RENSHAW, E., 1998, Detection of geological lineations on aerial photographs using two-dimensional spectral analysis. *Computers & Geosciences*, **24**, 771–784.

NIBLACK, W., 1986, *An Introduction to Digital Image Processing* (Englewood Cliffs, NJ: Prentice-Hall).

RENSHAW, E., and FORD, E. D., 1984, The description of spatial pattern using two-dimensional spectral analysis. *Vegetatio*, **56**, 75–85.

SEGHIERI, J., FLORET, C., and PONTANIER, R., 1994, Development of an herbaceous cover in a Sudano-Sahelian savanna in North Cameroon in relation to available soil water. *Vegetatio*, **114**, 175–184.

SERPANTIÉ, G., TEZENAS DU MONTCEL, L., and VALENTIN, C., 1992, La dynamique des états de surface d'un territoire agropastoral soudano-sahélien. Conséquences et propositions. In *L'aridité, une contrainte au développement*, edited by E. Le Floc'h, M. Grouzis, A. Cornet and J. C. Bille (Paris: ORSTOM), pp. 420–447.

SOKAL, R. R., and ROHLF, F. J., 1995, *Biometry: the Principles and Practice of Statistics in Biological Research* (San Francisco: Freeman).

WARREN, P. L., and DUNFORD, C., 1986, Sampling semiarid vegetation with large-scale aerial photography. *ITC Journal*, **4**, 273–279.

WHITE, F., 1983, *The Vegetation of Africa. A Descriptive Memoir to Accompany the Unesco/AEFTAT/UNSO Vegetation Map* (Paris: UNESCO/AETFAT/UNSO).

WHITE, L. P., 1970, *Brousse tigrée* patterns in southern Niger. *Journal of Ecology*, **58**, 549–553.

WHITE, L. P., 1971, Vegetation stripes on sheet wash surfaces. *Journal of Ecology*, **59**, 615–622.

*Journal of Vegetation Science 14: 163-172, 2003*




# A test for spatial relationships between neighbouring plants in plots of heterogeneous plant density

**Couteron, Pierre[1*]; Seghieri, Josiane[2] & Chadœuf, Joël[3]**

[1]*Ecole Nationale du Génie Rural des Eaux et des Forêts / UMR-CNRS 5120 (AMAP), 648 rue J.-F. Breton, B.P. 44494, F-34093 Montpellier Cédex 5, France;* [2]*Institut de Recherche pour le Développement / CESBIO (Centre d'Etudes Spatiales de la BIOsphère), 18 av. Edouard Belin bpi 2801, F-31401 Toulouse Cédex, France;* [3]*Institut National de la Recherche Agronomique, Unité de Biométrie, Domaine Saint-Paul, Site Agroparc, 84914 Avignon Cédex 9, France;*
[*]*Corresponding author; Fax +33467047101; E-mail couteron@engref.fr*

**Abstract.** Maps of plant individuals in (*x, y*) coordinates (i.e. point patterns) are currently analysed through statistical methods assuming a homogeneous distribution of points, and thus a constant density within the study area. Such an assumption is seldom met at the scale of a field plot whilst delineating less heterogeneous subplots is not always easy or pertinent. In this paper we advocate local tests carried out in quadrats partitioning the plot and having a size objectively determined via a trade-off between squared bias and variance. In each quadrat, the observed pattern of points is tested against complete spatial randomness (CSR) through a classical Monte-Carlo approach and one of the usual statistics. Local tests yield maps of *p*-values that are amenable to diversified subsequent analyses, such as computation of a variogram or comparison with covariates. Another possibility uses the frequency distribution of *p*-values to test the whole point pattern against the null hypothesis of an inhomogeneous Poisson process. The method was demonstrated by considering computer-generated inhomogeneous point patterns as well as maps of woody individuals in banded vegetation (tiger bush) in semi-arid West Africa. Local tests proved able to properly depict spatial relationships between neighbours in spite of heterogeneity/clustering at larger scales. The method is also relevant to investigate interaction between density and spatial pattern in the presence of resource gradients.

**Keywords:** Heterogeneity; Inhomogeneous Poisson process; Point process; Semi-arid vegetation; Spatial pattern; Spatial statistics; Tiger bush; West Africa.

**Abbreviations:** CSR = complete spatial randomness; IPP = inhomogeneous Poisson process.

**Nomenclature:** Hutchinson & Dalziel (1954-1972).

## Introduction

Vegetation generally displays spatial patterns, namely non-random and uneven distributions of its main characteristics, among which are densities of the constituent species (Greig-Smith 1979; Dale 1999). The patterns may either emerge from numerous, local-scale, plant-to-plant interactions (Levin 1992; Lejeune et al. 1999; Couteron & Lejeune 2001) or be imposed by larger scale environmental constraints, such as edaphic conditions. This distinction is obviously not clear-cut, because vegetation can exert a strong feedback on its environment, sometimes sharpening environmental discrepancies (Wilson & Agnew 1992; Wilson & Nisbet 1997). Whatever their nature, ecological mechanisms underlying spatial patterns generally operate at several scales and complex nested patterns are common (Lepš 1990; Levin 1992). If important insights on mechanisms were to be drawn from observed patterns, a belief that is widespread among vegetation scientists (e.g. Barot et al. 1999), then pattern analysis should be able to properly depict spatial relationships (e.g. clustering vs regularity) between neighbouring plants in spite of heterogeneity at larger scales. Such a condition, though reasonable, is not yet guaranteed by quantitative methods that aim at analysing vegetation data of the highest spatial resolution, namely maps representing plants as points.

Maps of points may be relevant for the study of spatial distributions, not only for strictly individual plants (e.g. trees), but also for clonal shoots, propagules, grass tufts etc. as long as it is meaningful to reduce the objects of study to geometrical points, and possible to map them in (*x*, *y*) coordinates. In the field of vegetation science, there have been numerous studies analysing maps of points (e.g. Duncan 1991; Skarpe 1991; Couteron & Kokou 1997). Point process theory provides a framework of analysis, via summary functions that lend themselves to statistical tests (Ripley 1977; Diggle 1983; Cressie 1993; van Lieshout & Baddeley 1997, 1999).

Such tests compare the observed pattern to those of a null hypothesis of complete spatial randomness (CSR), according to which points are independent and their distribution is ruled by a uniform density of probability (Diggle 1983: 50). For Monte-Carlo testing, CSR is simulated through a homogeneous Poisson process, which is conditional to the number of points observed in the study plot. This means that the intensity of the process, i.e. the mean number of points per unit area, is assumed to be constant throughout the study plot. The test compares the observed pattern against simulations via summary functions that are usually based on second-order characteristics (e.g. inter-point distances or number of neighbours), depicting spatial dependence between points in different regions of the plane. As these functions are generally estimated for the whole study plot, homogeneity in second order characteristics is also assumed.

Since vegetation is generally heterogeneous or patchy, the assumption of homogeneity is seldom met by field data. The consequences may be serious and second-order analyses, for instance through the classical *K*- and *L*-functions (Ripley 1977), may describe a large scale inhomogeneous distribution of points, rather than spatial relationships between neighbouring points. At worst, inhibition between neighbours may be masked by large scale heterogeneity (Lepš 1990). Such problems have been recognized by Pélissier & Goreaud (2001), who proposed to delineate regions of interest, or subplots, that would be homogeneous in terms of point density and of second order characteristics. Their approach has proved efficient in dealing with 'simple cases of heterogeneous vegetation', that lend themselves to an objective delineation of homogeneous subplots. But its applicability may be questioned with respect to less amenable situations, for instance in the presence of a smooth intensity gradient.

An approach of broader applicability has recently been proposed by Brix et al. (2001) for the analysis of inhomogeneous point patterns. The study area is split into quadrats at the scale of which the point intensity can be considered as constant. Then a local test, based on the CSR null hypothesis, is carried out for each quadrat. The final step uses the results obtained for all testable quadrats to perform a global test whose null hypothesis corresponds to an inhomogeneous Poisson process (IPP). We consider that such an approach broadens ecological insights that can be drawn from labour intensive mappings of plant individuals (or clonal shoots) and the aim of the present paper is to demonstrate its usefulness. Since methodological details as well as mathematical developments have already been provided by Brix et al. (2001), we shall restrict ourselves to a brief presentation of the method, while emphasising on illustration through both computer generated and field examples.

## Method

Three distinct steps are considered: (1) looking for an 'optimal' partition of the study area into quadrats of an adequate size; (2) computing the *p*-value of the CSR null hypothesis, for each quadrat, via a local Monte-Carlo test; (3) grouping the local *p*-values to test their frequency distribution against the uniform distribution that would be encountered if the whole pattern of points was derived from an IPP (global test).

### *Theoretical background*

The optimal partition is defined from a quadrat size that minimizes the total mean squared error (i.e. squared bias plus variance) attached to the statistic used for the global test. To carry out a test conditional to the local density, we have to assume that density is not too inhomogeneous within quadrats. Squared bias, which measures the departure from this assumption, is likely to increase in relation to quadrat size. On the other hand, variance is bound to be high when quadrat size is too small, since many quadrats may contain less than two points and be non testable against CSR. Hence the trade-off between squared bias and variance that can indicate an adequate quadrat size. Explicit estimations of squared bias and variance are made possible by assuming that the point pattern under study derives from an underlying 'Cox process', namely a point process for which the intensity function may not only vary in space, as for an IPP, but is also considered as a particular realization of a non-negative valued stochastic process, called an 'intensity process' (Diggle 1983: 58; Table 1).

From the sole statistical analysis of a point pattern, no difference between clustering and heterogeneity can be sustained, and a great deal of point processes marked by aggregation are to be considered as particular versions of a Cox process (Diggle 1983). For instance, a pattern made of clumps of young plants randomly distributed within a certain distance around parents can be modelled as a Neyman-Scott process (Diggle 1983) and, therefore, as a particular version of a Cox process. Conditional to a particular distribution of the parents, this is also an IPP. In general, plant patterns stemming from dispersion of propagules are likely to have an inhomogeneous density, whatever the nature of local plant-plant relationships. To study the latter, the IPP is a more realistic and less restrictive null model than the homogeneous Poisson process.

**Table 1.** Main characteristics of the theoretical point processes used as reference.

| | Nature of the intensity function | Stationarity of the intensity function | Stationarity of the point process |
|---|---|---|---|
| Homogeneous Poisson process | $\lambda(\mathbf{x}) = \lambda$ for any location $\mathbf{x}$ | $\lambda$ is constant (thus stationary) | Stationary |
| Inhomogeneous Poisson process | $\lambda(\mathbf{x})$ is varying with respect to $\mathbf{x}$ | $\lambda(\mathbf{x})$ is stationary | Non-stationary |
| Cox process | $\lambda(\mathbf{x})$ is a realisation of a stochastic process $\Lambda(\mathbf{x})$ | Each $\lambda(\mathbf{x})$ is stationary if $\Lambda(\mathbf{x})$ is | Non-stationary if relative to a particular realisation $\lambda(x)$ |

*Estimation of bias and variance and selection of quadrat size*

For a particular quadrat size, the intensity function is estimated as the ratio of the number of points in each quadrat divided by its area. Let *d(ij)* be this estimate for a particular quadrat *(i,j)* of the plot partition (*i* refers to rows and *j* to columns). We refer to a Cox process having a stationary intensity process, $\Lambda(\mathbf{x})$, while assuming that variations in space of the intensity function, $\lambda(\mathbf{x})$, are not 'too abrupt', so as to allow the estimation of its derivatives by finite differences. First-order derivatives, $hd_1$ and $vd_1$, are estimated according to the horizontal and vertical directions, respectively, that is:

$$hd_1(i,j)=(d(i,j+1) - d(i,j-1))/(2d(i,j)); \quad (1a)$$
$$vd_1(i,j)=(d(i+1,j)-d(i-1,j))/(2d(i,j)) \quad (1b)$$

Note that this estimation is normalized by the intensity through a division by *d(i,j)*, since the test in each quadrat will be conditional to the number of points. Then second-order derivatives, $hd_2$ and $vd_2$, are computed as:

$$hd_2(i,j)=(hd_1(i,j+1) - hd_1(i,j-1)) / 2; \quad (2a)$$
$$vd_2(i,j)=(vd_1(i+1,j) - vd_1(i-1,j)) / 2 \quad (2b)$$

A second-order derivative, $vhd_2$, is also computed in the 'diagonal' direction as:

$$vhd_2(i,j)=(vd_1(i,j+1)-vhd_1(i,j-1))/2 \text{ or} \quad (3a)$$
$$vhd_2(i,j)=(hd_1(i+1,j)-hd_1(i-1,j))/2 \quad (3b)$$

We have introduced a more concise notation by defining the vector of first-order derivatives as the transpose $\mathbf{A_{ij}}^{\mathbf{Tr}}$ of $\mathbf{A_{ij}} = [hd_1(i,j),vd_1(i,j)]$ and the matrix of second-order derivatives as:

$$\mathbf{B_{ij}} = [hd_2(i,j) \ , hvd_2(i,j)] \quad (4a)$$
$$[hvd_2(i,j), vd_2(i,j) \ ] \quad (4b)$$

Bias on the result of any local test stems from the fact that the intensity function is not strictly homogeneous in the quadrats. The less homogeneity, the greater the bias. The contribution, *bs(i,j)*, of a particular quadrat to the overall bias can be estimated using the equation provided by Brix et al. (2001: 492), which is expressed here in a more practical form as: (5)

$$bs(i,j) = \frac{n_{ij}}{24} trace\left(\mathbf{B_{ij}}\right) \pm \left( \sum_N n_{ij} \mathbf{u}^{\mathbf{Tr}} \mathbf{B_{ij}} \mathbf{u} + \frac{n_{ij}\left(n_{ij} \pm 1\right)}{2} \mathbf{u}^{\mathbf{Tr}} \left(\mathbf{A_{ij}} \mathbf{A_{ij}}^{\mathbf{Tr}}\right) \mathbf{w} \right) \Big/ N_{sim}$$

Here $N_{sim}$ is the total number of Monte-Carlo simulations of CSR (conditional to the number $n_{ij}$ of points in the quadrat) that are used to estimate *bs(i,j)*. The summation only concerns those *N* simulations that yield a value of the local test statistic (see below) exceeding the value observed in the quadrat under consideration. In each of these simulations of $n_{ij}$ points, two points $\mathbf{u}$ and $\mathbf{w}$ are randomly chosen for estimating *bs(i,j)*.

For a given quadrat size, the overall bias, *bs*, is then estimated as the mean of all *bs(i,j)* values found for the *Nq* quadrats with $n_{ij} > 1$. On the other hand, the variance of the global test (see below) is estimated as:

$$\sigma^2 = \frac{1}{12Nq} \quad (6)$$

(Brix et al. 2001: 492)

Then, the mean squared error term, $Er = bs^2 + \sigma^2$, is plotted against quadrat size or, equivalently, against the number of quadrats in the plot partition. The adequate quadrat size is the one yielding the lowest value for *Er*.

*Performing a local test for a selected quadrat size*

Once the quadrat size has been identified, spatial distribution of points within each quadrat can be tested against CSR through any one of the statistics that have been designed to investigate spatial patterns of points (Cressie 1993: 603). Three point process statistics have been considered alternatively by Brix et al. (2001), namely the mean distance to nearest neighbours, the mean distance from point to point and the mean distance from random additional points to the points of the pattern under study (referred to as $\overline{G}$, $\overline{H}$ and $\overline{F}$, respectively, in analogy with Diggle 1983). It is to be noted that the choice of a particular statistic has generally little influence on the determination of the quadrat size, although distinct statistics may not be equally powerful in detecting a particular type of departure from CSR (see below). Specifically, the $\overline{F}$ statistic proved to be of poor sensitivity for a wide set of simulated processes (Brix et al. 2001 and unpubl.). Consequently, only $\overline{G}$ and $\overline{H}$ statistics have been retained for comparison in the present paper. The local test could also have used any other point process function and, in particular, an analogue of Ripley's *K*-function, namely the edge-corrected mean distance from point to point.

The test in each quadrat is carried out via the usual Monte-Carlo approach by generating *m* realizations of point patterns under the CSR null hypothesis. The test statistic (either the mean nearest neighbour distance, $\overline{G}$, or the mean inter-point distance, $\overline{H}$) is computed for each realization and compared to the observed value. In the presence of clustered points, both $\overline{G}$ and $\overline{H}$ observed values are below their expectations under CSR since shorter distances between points are generally encountered. Conversely, inhibition between points leads to high values of both $\overline{G}$ and $\overline{H}$. The *p*-value attached to a given statistic is estimated from the relative frequency of CSR realizations exceeding the observed value. It is expected to be low in so far as inhibition is concerned and high in the presence of clustering.

*Global test*

A global test can be carried out for the whole set of quadrats partitioning a given study plot. The whole point pattern in the plot is supposed to derive from an IPP (null hypothesis) and, consequently, quadrats are expected to yield a set of *p*-values uniformly distributed on [0,1]. The goodness of fit of the observed *p*-values with the uniform distribution is tested via the usual Kolmogorov-Smirnov test (Sokal & Rohlf 1995). A departure from the null hypothesis toward local clustering (at the quadrat scale) should lead to *p*-values more abundant above 0.5 than below and vice versa in the case of local inhibition.

**Illustration from computer simulated patterns**

The approach of the local test and its usefulness are illustrated through simulations of inhomogeneous point processes featuring either clustering, inhibition or independence at a local scale. To make the results more comparable we decided to consider a family of point processes that was based on the same macrostructure, namely a single cluster of points centred in the middle of the unit square. The density of points within the cluster was chosen to decrease steadily as a Gaussian function (with a standard deviation of $R = 0.25$) of the distance from the centre (see Fig. 1a and Fig. 2a for realizations with 200 points). With such a point process, there is neither clustering nor inhibition between neighbouring points once the local density is taken into account. Nevertheless, there is a strong rejection of CSR on the basis of the *L*-function for all 100 realizations of the process. The *L*-function (Besag 1977 in Ripley 1977), which has been extensively used by vegetation scientists (e.g. Duncan 1991; Skarpe 1991; Couteron & Kokou 1997), is derived from the *K*-function (Ripley 1977) and, as such, is based on the expected number of points within successive distances, *r*, from an arbitrary point of the pattern. Under CSR, *L(r)* can be equated to *r*, and it is convenient to plot *L(r)-r* against *r*. Thus, values of *L(r)* that lay significantly above zero indicate aggregation and vice versa for values below zero. For the 100 realizations of the point process specifically considered, all the points of the *L*-curves were above the confidence envelopes (Fig. 2b), a result that was due to the inhomogeneous distribution of the points within the plot. Hence

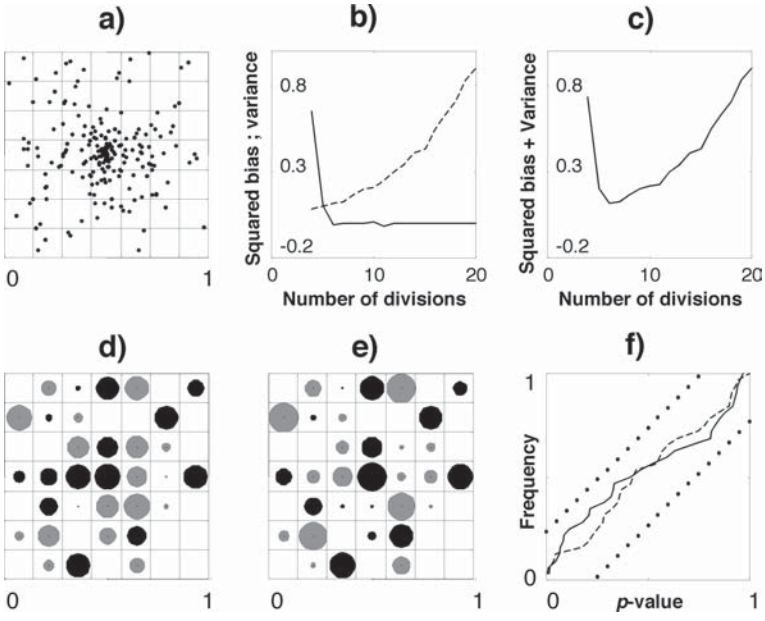


**Fig. 1.** Successive steps of the local test as illustrated from a particular realization (**a**) of an inhomogeneous Poisson process consisting of a unique central cluster of points. (**b**) Squared bias and variance as a function of the number, *n*, of divisions of the plot edge ($n \times n$ quadrats). **c**. Squared bias plus variance in relation to *n*. **d** and **e**. Maps of *p*-values for the nearest neighbour ($\overline{G}$) and the inter-point ($\overline{H}$) distance statistics, respectively (partition into $7 \times 7$ quadrats). The circle radius is proportional to the magnitude of departure from CSR (i.e. $p > 0.5$), black circles indicate local clustering (i.e. $p > 0.5$) and grey circles denote local inhibition (i.e. $|p - 0.5|$). Non-testable quadrats with less than two points are blank. **f.** Results of the global test based on the cumulative frequency distribution of quadrats *p*-values for the $\overline{G}$ and $\overline{H}$ statistics (solid and interrupted lines, respectively), with the 5% two-sided envelopes from the Kolmogorov-Smirnov distribution (dotted lines).

the *L*-function pointed towards the overall pattern but told nothing about the spatial relationship between neighbouring points within the cluster.

For the particular point pattern displayed in Fig. 1a, the squared bias dropped rapidly as a function of the number of quadrats while the variance increased steadily (Fig. 1b). Consequently, the mean squared error (i.e. squared bias + variance) was least for 6-7 divisions (36-49 quadrats) and the largest value was retained for partitioning the plot. The spatial distribution of points in each quadrat containing at least two points was then tested against CSR (conditional on local density) via the statistic (mean distance to the nearest neighbours; Fig. 1d) and the statistic (mean distance between all points; Fig. 1e). In both cases, the results were diversified with point distribution being strongly clustered in some quadrats (denoted by large black circles) while being marked by inhibition in others (large grey circles); in the remaining quadrats (small circles), departures from CSR were weak. Large departures from CSR appeared only randomly for some quadrats, and the frequency distribution of the *p*-values obtained for all testable quadrats remained within the confidence envelopes of the Kolmogorov-Smirnov distribution (global test; Fig. 1f). This was to be expected since the underlying point process implies no departure from CSR at a local scale. Furthermore, using all testable quadrats of the 100 realizations of the process, we computed the frequency distribution of *p*-values with respect to a regular partition of [0,1] into 20 classes. The frequency proved more or less the same for all classes (Fig. 2c), indicating that, at quadrat scale, all spatial configurations of points occurred a similar number of times. In fact, out of 100 realizations, the null hypothesis of local independence was rejected only once ($P<0.05$) using $\overline{G}$ and four times using $\overline{H}$ (Table 2).

We also considered IPPs that locally depart from CSR whilst still being based on a unique cluster of points (see Fig. 2d, g, j for particular realizations). Local clustering (Fig. 2d) was achieved by using the 70 points of the initial cluster as 'parent' points of a Neyman-Scott process (Diggle 1983), with a mean of three new points per parent and a Gaussian distribution of standard deviation $r_1$=0.08 around each parent. A first type of local inhibition (Fig. 2g) was simulated according to the 'hard-core' principle (Matérn 1986; Diggle 1983). All points of the initial cluster that have a 'stronger' neighbour at a distance less than a given value, e.g. *r*, are deleted (the 'strength' of a point being defined by a random value drawn from the uniform distribution). We considered, nevertheless, a non-stationary type of hard-core by allowing the radius *r* to increase proportionally to the distance from the centre of the plot (from 0 to $1/\sqrt{2}$), so as to mimic a one-sided competition for soil resource having a range that would increase in relation to the distance from the centre of a fertile patch. We also investigated another type of local thinning (Fig. 2j) for which the risk of a point being deleted is proportional to both the distance to and the 'strength' difference with its neighbours. Such a process, that belongs to the broad class of 'soft-core' processes (or 'Matérn-Stoyan models'; Stoyan 1988; Cressie 1993: 671), can be seen as a hard-core with a randomly varying radius *r*. It is fully defined by $\bar{r}$, the expected value of *r* (here we took $\bar{r} = 0.01$).

Introducing a local clustering of points did not substantially change the values taken by the *L*-function, though its variability increased (compare Fig. 2b to Fig. 2e). There was no substantial change in the *L*-function when local inhibition was considered via the non-stationary hard-core (Fig. 2h). This means that point patterns as different as in Fig. 2a, d, g cannot be properly differentiated on the sole basis of the *L*-function, which mostly reflects the macrostructure of the initial cluster of points. In particular, the first value of the *L*-function, that should depict the spatial relationships between neighbouring points separated by a distance of $< 0.005$, was always located above the upper confidence envelope, thereby pointing towards clustering. This was due to the fact that the points around the centre of the plot could be

**Table 2.** Results from 100 random simulations of four-point processes having distinct local patterns yet stemming from the same macro-structure (a unique cluster of points). $N_{psim}$ = Mean number of points per simulation; $N_{rq}$ = Retained number of quadrats; $N_{qsim}$ = Mean number of testable quadrats per simulation; $D_{K\text{-}S}$ = Mean maximal departure for Kolmogorov-Smirnov (K.-S.) test (positive sign: departure toward inhibition; negative sign: departure toward clustering); %sim percentage of simulations with rejection of local independence ($P < 0.05$; K.-S. test). $\overline{G}$ and $\overline{H}$ refer to the mean distance to nearest neighbours and to the mean inter-point distance, respectively.

| Processes | $N_{init}$ | $N_{psim}$ | $N_{rq}$ | $N_{qsim}$ | $D_{K\text{-}S}\,\overline{G}$ | $D_{K\text{-}S}\,\overline{H}$ | $\%_{sim}\,\overline{G}$ | $\%_{sim}\,\overline{H}$ |
|---|---|---|---|---|---|---|---|---|
| Local independence | 200 | 200 | $7 \times 7$ | 39.2 | – 0.1248 | – 0.1604 | 1 % | 4 % |
| Local clustering | 70 | 208 | $5 \times 5$ | 17.6 | – 0.502 | – 0.472 | 97 % | 92 % |
| Local inhibition | 300 | 191 | $10 \times 10$ | 28.6 | 0.379 | 0.244 | 97 % | 41 % |
| Local inhibition(soft-core) | 300 | 205 | $20 \times 20$ | 47.9 | 0.257 | 0.230 | 90 % | 75 % |

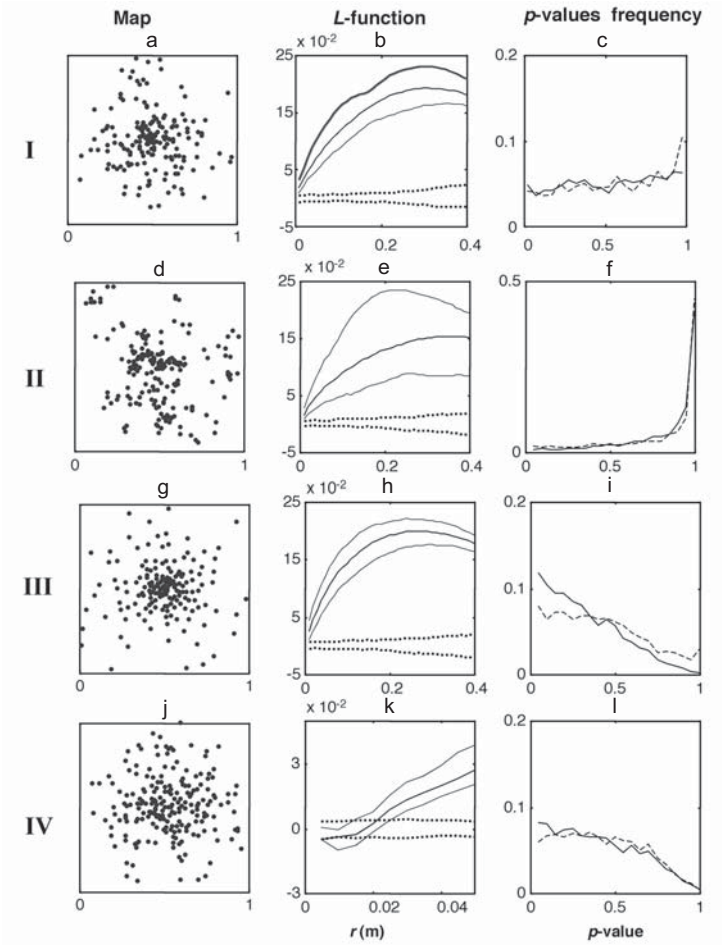


**Fig. 2.** Results for simulated patterns derived from the same macro-structure, i.e. a unique central cluster of points, as on (**a**). **I.** No local structure; **II.** Local clustering; **III.** Local inhibition (hard-core with a radius increasing in relation to the distance to the centre); **IV.** Local inhibition (soft-core with an expected radius of $r$=0.01). (**a**), (**d**),( **g**),(**j**) Examples from particular realisations. (**b**), (**e**), (**h**), (**k**) Minimum, median and maximum values (solid lines) taken by the $L$-function for 100 realizations of the point process ($L(r)$-$r$ is plotted against the distance, r); dotted lines indicate the 5% two-sided confidence interval computed from 400 Monte-Carlo simulations under CSR. (**c**), (**f**), (**i**), (**l**) Frequency distribution of $p$-values obtained for all testable quadrats and for 100 realizations, for the $\overline{G}$ (mean distance to the nearest neighbour; solid line) and $\overline{H}$ (mean inter-point distance; interrupted line) statistics.

extremely close relative to the overall density which is the reference for the computation of $K$- and $L$-functions. Only when local inhibition corresponded to the soft-core were the first values of the $L$-curve located below the upper envelope (Fig. 2k), and sometimes even under the lower envelope. However, only 50% of the realizations yielded an $L$-curve featuring at least one point under the lowest envelope and the presence of local inhibition could easily have been missed from a particular realization.

Results provided by quadrat-based local tests proved informative for all the considered kinds of point processes (Table 2). When using $\overline{G}$, the null hypothesis of an inhomogeneous Poisson process (local independence) was rejected for most simulations involving either local clustering or inhibition. The $\overline{H}$ statistic did not perform so well and encountered some problems with the non-stationary hard-core. The efficiency of the local test is due to the fact that the frequency distribution of the $p$-values yielded by all testable quadrats for 100 realizations clearly differentiate the distinct processes. Hence, in the presence of local clustering, most quadrats provided a $p$-value > 0.9, since 90% of Monte-Carlo simulations of CSR had a mean distance to the nearest neighbours (or a mean inter-point distance) above the observed values. Conversely, when we introduced local inhibition into the process, $p$-values > 0.5 were scarce, indicating that most simulations of CSR provided values of the test statistic (either $\overline{G}$ or $\overline{H}$) that were lower than the observed value.

## Application to semi-arid patterned vegetation

The first application deals with the spatial distribution of a particular species, *Pterocarpus lucens*, in a banded vegetation (tiger bush) observed in northern Burkina Faso (West Africa). The study plot of 10.24 ha (320m × 320m) was located on hard Palaeozoic sandstones with mean annual rainfall of 486 mm (1986-1994 figures; see Couteron 2001 for more details). *P. lucens* is a small tree, which had a mean height of ca. 4.3 m. Bands of dense vegetation were more or less perpendicular to the overall slope (0.8%) and were dominated by *Combretum micranthum*, a shrub with a mean height of ca. 2.9m (not mapped but counted into 10m × 10m

quadrats). This species accounted for more than 50% of the basal area of the vegetation in the plot, compared to 32% for *P. lucens*. Although ca. 53% of the 660 *P. lucens* mapped in the plot were located outside bands (defined by a density of *C. micranthum* above 500.ha$^{-1}$), the spatial distribution of *P. lucens* was not homogeneous, due to the existence of elongated areas devoid of trees (Fig. 3a). The complementary pattern of fuzzy bands and clumps determined an *L*-curve that increased steadily to a distance of ca. 25 m before declining (Fig. 3c). A peak in the *L*-curve is often considered as corresponding to the mean size of clumps (Dale 1999: 217), but such an interpretation would be questionable for the present pattern, which displayed a marked anisotropy (see Couteron 2001 for a detailed characterization). In any case, the initial part of the curve, up to 6-7 m, lay within the confidence envelopes suggesting that neighbouring *P. lucens* may be independently distributed. However, in the presence of a strong macrostructure, we know that no result at the local scale can be firmly established through the *L*-function, since the basic assumption of a homogeneous density of points is obviously violated.

Computation of squared bias and variance in relation to the number of plot divisions indicated an appropriate quadrat size of around 32 m (not presented). The local test based on such quadrats yielded *p*-values that were uniformly distributed on [0,1] for both $\overline{G}$ and $\overline{H}$ statistics (Fig. 3d), and local independence between neighbouring points was thus confirmed. A visual inspection of the map of the $\overline{G}$-based *p*-values (Fig. 3b) revealed no obvious overall spatial pattern. Indeed, correlation between *p*-values from adjacent quadrats proved weak since, in Fig. 3e, the variogram (Cressie 1993: 69) wavered between confidence envelopes obtained by randomly permuting values in quadrats (Cressie 1993: 597). Computation of the variogram illustrates, nevertheless, the fact that *p*-values at the quadrat scale convey information enabling diversified statistical treatments in addition to the global test.

A second example comes from tiger bush in Niger and is characterized by comparable climatic and edaphic conditions (see Couteron et al. 2000 for details on the 'Banizoumbou' site), though having a higher basal area (7.9m$^2$.ha$^{-1}$ against 6.6m$^2$.ha$^{-1}$ for the site in Burkina Faso). We considered a single thicket within a field plot of 0.5 ha (100m × 50m; Fig. 4a), in which all woody plants with a height >0.5m have been mapped in (*x, y*) coordinates (78% of these 903 individuals were *C. micranthum*). A substantial part of the plot was virtually devoid of woody individuals and, as suggested by Pélissier & Goreaud (2001), it appeared sensible to compute the *L*-function for the area that can be delineated as thicket (using the edge-correction proposed by Goreaud & Pélissier 1999). However, even within the thicket, the spatial distribution of woody individuals was not homogeneous (Fig. 4a) and an objective delineation proved not to be as easy as expected. In fact, the *L*-function remained marked by the inhomogeneous distribution of points with no clear decrease of the function for distances < 20m (Fig. 4d). Inhibition was, nevertheless, apparent for distances <1m, thanks to the prior delineation of the thicket, without which the first point

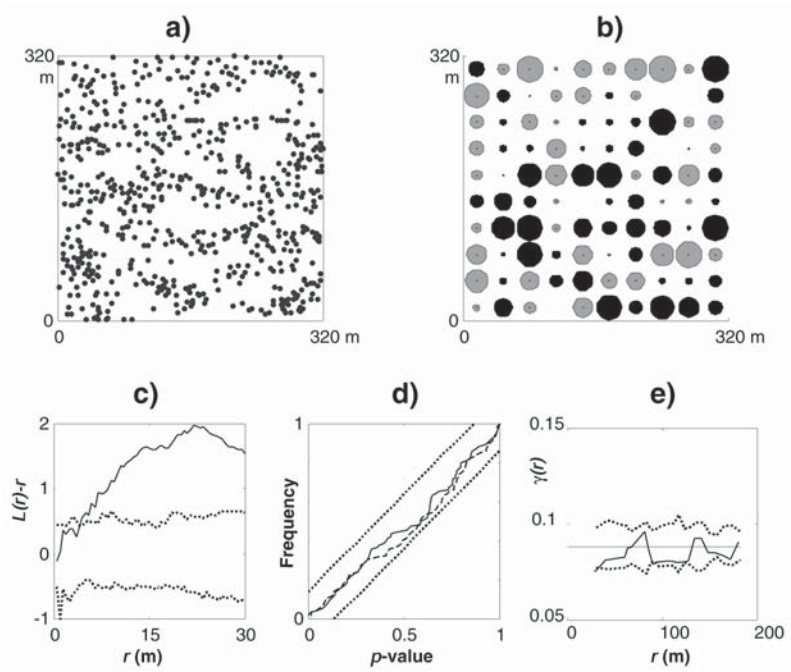


**Fig. 3.** Analysis of the spatial distribution of 660 *Pterocarpus lucens* with a height > 1.5 m in a 10.84 ha plot in NW Burkina Faso. **a.** Map of individuals (denoted by a point). **b.** Map of *p*-values ($\overline{G}$ statistic) obtained for 32m × 32m quadrats. The circle radius is proportional to the magnitude of the departure from CSR (i.e. $p > 0.5$), black circles indicate local clustering (i.e. $p > 0.5$) and grey circles denote local inhibition (i.e. $|p - 0.5|$). **c.** *L*-function ($L(r)$-$r$ is plotted against the distance, $r$ ) of the pattern (solid line) with the 5% two-sided confidence interval yielded by 400 Monte-Carlo simulations of CSR (dotted line). **d.** Frequency distribution of *p*-values obtained for all testable quadrats for the statistics $\overline{G}$ (mean distance to the nearest neighbour; solid line) and $\overline{H}$ (mean inter-point distance; dashed line); dotted lines denote the 5% two-sided confidence interval (Kolmogorov-Smirnov distribution). **e.** Variogram of the $\overline{G}$-based *p*-values with the Monte-Carlo 5% two-sided envelopes (dotted lines).

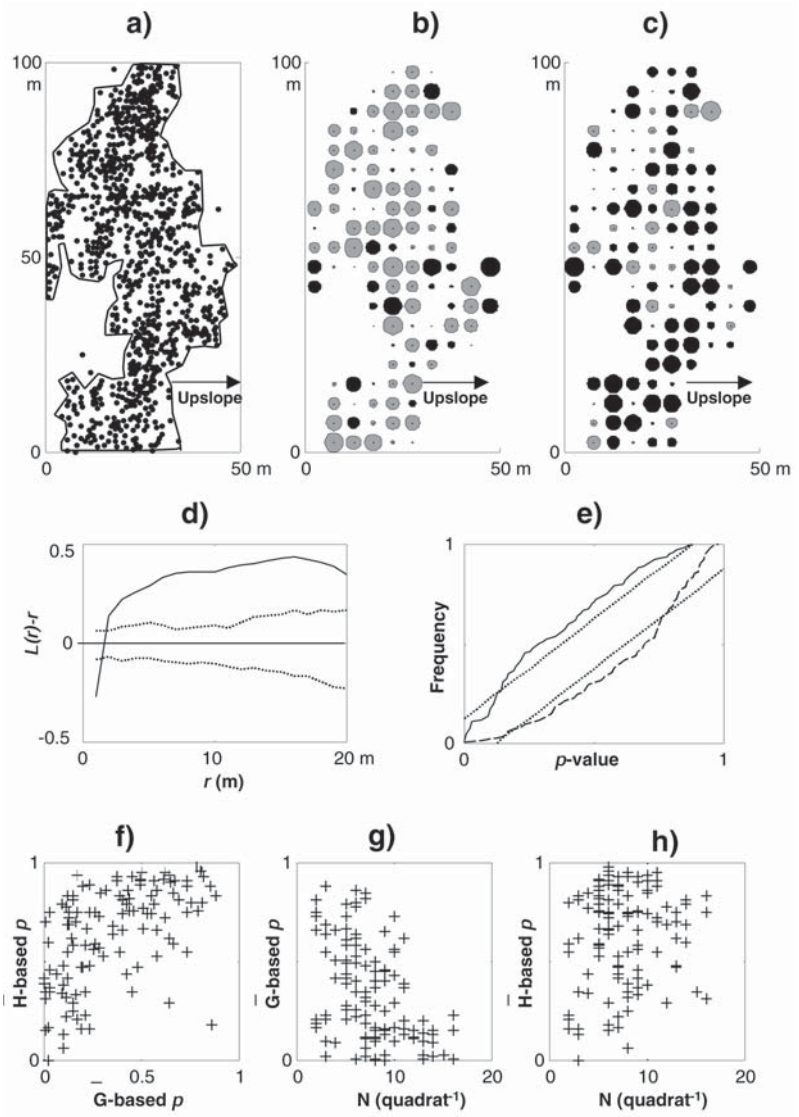


**Fig. 4.** Analysis of the spatial distribution of 903 woody individuals with a height > 0.5 m in a 0.5 ha plot in West Niger. **a.** Maps of individuals (denoted by a point) with a manual delineation of the thicket (solid line). **b.** and **c.** Maps of *p*-values obtained for 5m × 5m quadrats via the $\overline{G}$ and $\overline{H}$ statistics, respectively. The circle radius is proportional to the magnitude of the departure from CSR (i.e. $|p - 0.5|$), black circles indicate local clustering (i.e. $p > 0.5$) and grey circles denote local inhibition (i.e. $p < 0.5$). **d.** L-function for all the points within the area delineated as thicket ($L(r)$-$r$ is plotted against the distance $r$) of the pattern (solid line) with the 5% one-sided confidence interval yielded by 100 Monte-Carlo simulations of CSR (dotted line). **e.** Frequency distribution of *p*-values obtained for all testable quadrats for the $\overline{G}$ (mean distance to the nearest neighbour; solid line) and $\overline{H}$ (mean inter-point distance; dashed line) statistics; dotted lines denote the 5% one-sided confidence interval (Kolmogorov-Smirnov distribution). **f.** Scatterplot of $\overline{H}$- vs $\overline{G}$ based *p*-values. **g.** Comparison of $\overline{G}$-based *p*-values with the density of shrubs in the quadrats. **h.** As **f.** with $\overline{H}$-based *p*-values.

of the *L*-curve would have remained within the confidence envelopes (result not presented).

Computation of squared bias and variance (not presented) pointed toward a quadrat size ca. 5m for the nearest neighbour statistic ($\overline{G}$) and in the range 5-10m for the inter-point statistic ($\overline{H}$). To allow a consistent comparison we considered a single quadrat size of 5m for both statistics. The global test gave contrasting results, since local inhibition was found using $\overline{G}$ whereas $\overline{H}$ suggested local aggregation (Fig. 4e). *p*-values provided by each statistic were, nevertheless, not independent at the quadrat scale (Fig. 4f) but $\overline{H}$ gave variable results when facing local inhibition (i.e. when low *p*-values were found via $\overline{G}$) while the converse was observed in the presence of local aggregation (i.e. when $\overline{H}$ yielded high *p*-values). The relative power of each statistic thus appeared different for the two alternatives to CSR. A statistic based on distances to nearest neighbours is very efficient at detecting inhibition, while being of limited sensitivity in the presence of weak aggregation. This was made explicit by a particular feature of the pattern under study, namely the fact that the distribution of woody plants tended to be clumped in some quadrats (mostly located in the upper half of the thicket), while being marked by inhibition in others (Fig. 4b, c). The two statistics gave dual evidence of such a shift from clumping to regularity that was consistent with the existence of a lower soil water resource in the downslope half of the thicket. In fact, the amount of available water, which peaked in the middle of the thicket, decreased slowly in the direction of the upslope edge and decreased sharply in the downslope direction (Galle et al. 1999; Seghieri & Galle 1999). Another interesting result, pinpointing the potential link between resource limitation and spatial pattern, was that only regular distributions (i.e. low *p*-values) were obtained

via the $\overline{G}$ statistic for a local density above 12-15 woody individuals per quadrat (i.e. 5000-6000.ha$^{-1}$). There seems to be a threshold of density above which the stand tends towards a regular pattern. It is also noteworthy that the most clumped distributions, revealed by $\overline{H}$, corresponding to high *p*-values did not occur independently of local density, since being observed in quadrats having an intermediate density (5-10 individuals, i.e. 2000-5000.ha$^{-1}$). All these results, obtained at a local scale, suggested the existence of complex relationships between spatial pattern, density and water resource limitation. Hence additional field work is necessary, notably to compare soil water balance and spatial distribution of shrubs in 5m × 5m quadrats.

## Discussion

Spatial distributions of plants are shaped by both local scale plant to plant interactions and larger scale environmental constraints. Quantitative descriptions have, therefore, to cope with nested scales of heterogeneity, while the range of scales addressable from a particular field investigation is limited by both the spatial resolution of data ('grain') and the spatial extent of the observation window (Turner & Gardner 1991). Mapping individuals in a plot provides a very precise grain but for such an effort to be worthwhile demands that ensuing quantitative analyses allow a reliable characterization of spatial relationships between neighbours. Obviously, this is not possible when the analysis is carried out at the level of a large plot.

In some cases large and fairly homogeneous subplots can be delineated, thereby enabling simple and powerful analyses (Pélissier & Goreaud 2001). However, in many cases such as the pattern in Fig. 3, delineating homogeneous subplots would not be an easy task. Nevertheless, this approach can be seen as somewhat complementary to our method, since rapid changes of the intensity function within the plot may ease subplot delineation, while rendering questionable the way we estimated the intensity function through finite differences. Another limitation of the method is that it is unlikely to obtain useful results with a low number of points, e.g. < 40-50, within a plot.

Most methods used in the field of vegetation science for point pattern analysis rely on tests relative to a constant density of points in the plot. Getis & Franklin (1987) have proposed computing some values of the *L*-function around each point of the pattern, and to create contour maps from the results. The computation is also relative to the overall density. In the absence of a test, results may be misleading since contrasted contour maps have been obtained even for realizations of a homogeneous Poisson process (J. Chadœuf unpubl.). The 'circumcircle' method (Dale & Powell 2001) also yields local results, with a formal test, relative to the overall density. Whatever the method, testing with respect to a local density raises the difficult question of how large the area used for density estimation should be. For instance, if local density is to be estimated in circles around individual points, so as to complement the approach of Getis & Franklin (1987), how large should their radii be? There is no easy answer to this question. Furthermore, it would be difficult to go from local results to a global test, since results obtained from intersecting circles would obviously not be independent.

A regular partition of space into quadrats is a relevant basis for studying an inhomogeneous point pattern, seeing that an adequate quadrat size can be determined. Characterizing the departure from CSR via a *p*-value extends to maps of points what is referred to with digital images as a 'textural transform' (Haralick 1979), namely the production of a new image containing information on patterns within its resolution cells (quadrats). Maps of *p*-values are amenable to subsequent analyses, investigating the spatial distribution of *p*-values (e.g. computation of the variogram) or comparing such a distribution with spatialized covariates, that may relate to either environmental conditions (e.g. soil characteristics) or to vegetation (e.g. local density). Frequency distribution of *p*-values can be used, not only to test the null hypothesis of an IPP, but also to compare two maps of points (e.g. two species, or a species in two distinct plots) for which the same quadrat size is deemed acceptable.

A final element is that the approach via a local test can be based on several spatial statistics and several maps of *p*-values can be obtained from the same map of points. Since distinct statistics do not look exactly at the same aspect of a given pattern, the results may be often complementary yet sometimes redundant (see our field examples). From a broader perspective, it can be said that, due to the complexity of plant spatial distributions, a single approach cannot detect all relevant features present in a large set of field data (as underlined by Diggle 1983). There are several methods that are very efficient at characterizing large structures present in a point pattern, such as clumps or gaps (Couteron 2001; Dale & Powell 2001), but that tells us little about the relationship between neighbouring points. Ripley's *K*- or *L*-functions are often informative regarding both aspects and, in spite of some limitations (Dale 1999: 216), deserve their widespread use. However, it should be kept in mind that results obtained for small reference distances may be biased due to structures present at larger scales. Only a local approach can ensure that spatial relationships between neighbouring plants are thoroughly described, quantified and characterized.

**References**

Barot, S., Gignoux, J. & Menaut, J.-C. 1999. Demography of a savanna palm tree: predictions from comprehensive spatial pattern analysis. *Ecology* 80: 1987-2005.

Brix, A., Senoussi, R., Couteron, P. & Chadœuf, J. 2001. Assessing goodness of fit of spatially inhomogeneous Poisson processes. *Biometrika* 88: 487-497.

Couteron, P. 2001. Using spectral analysis to confront distributions of individual species with an overall periodic pattern in semi-arid vegetation. *Plant Ecol*. 156: 229-243.

Couteron, P. & Kokou, K. 1997. Woody vegetation spatial patterns in a semi-arid savanna of Burkina Faso, West Africa. *Plant Ecol*. 132: 211-227.

Couteron, P. & Lejeune, O. 2001. Periodic spotted patterns in semiarid vegetation explained by a propagation-inhibition model. *J. Ecol*. 89: 616-628.

Couteron, P., Mahamane, A., Ouedraogo, P. & Seghieri, J. 2000. Differences between banded thickets (tiger bush) at two sites in West Africa. *J. Veg. Sci*. 11: 321-328.

Cressie, N.A.C. 1993. *Statistics for spatial data*. Wiley, New York, NY.

Dale, M.R.T. 1999. *Spatial pattern analysis in plant ecology*. Cambridge University Press, Cambridge, UK.

Dale, M.R.T. & Powell, R.D. 2001. A new method for characterizing point patterns in plant ecology. *J. Veg. Sci*. 12: 597-608.

Diggle, P.J. 1983. *Statistical analysis of spatial point patterns*. Academic Press, London, UK.

Duncan, R. 1991. Competition and the coexistence of species in a mixed podocarp stand. *J. Ecol*. 79: 1073-1084.

Galle, S., Ehrmann, M. & Peugeot, C. 1999. Water balance in a banded vegetation pattern. A case study of tiger bush in western Niger. *Catena* 37: 197-216.

Getis, A. & Franklin, J. 1987. Second-order neighborhood analysis of mapped point patterns. *Ecology* 68: 473-477.

Goreaud, F. & Pélissier, R. 1999. On explicit formulas of edge effects correction for Ripley's *K*-function. *J. Veg. Sci*. 10: 433-438.

Greig-Smith, P. 1979. Pattern in vegetation. *J. Ecol*. 67: 755-779.

Haralick, R.M. 1979. Statistical and structural approaches to texture. *Proc. IEEE* 67: 786-804.

Hutchinson, J. & Dalziel, J.M. 1954-1972. *Flora of West Tropical Africa*. Crown, London, UK.

Lejeune, O., Couteron, P. & Lefever, R. 1999. Short range co-operativity competing with long range inhibition explains vegetation patterns. *Acta Oecol*. 20: 171-183.

Lepš, J. 1990. Can underlying mechanisms be deduced from observed patterns? In: Krahulec, F., Agnew, A.D.Q., Agnew, S. & Willems, J.H. (eds.) *Spatial processes in plant communities*. pp. 1-11. SPB Academic Publishing, The Hague, NL.

Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73: 1943-1967.

Matérn, B. 1986. *Spatial variation*. Springer-Verlag, New York, NY.

Pélissier, R. & Goreaud, F. 2001. A practical approach to the study of spatial structure in simple cases of heterogeneous vegetation. *J. Veg. Sci*. 12: 99-108.

Ripley, B.D. 1977. Modelling spatial patterns. *J. R. Stat. Soc. Ser. B* 39: 172-212.

Seghieri, J. & Galle, S. 1999. Run-on contribution to a Sahelian two-phases mosaic system: Soil water regime and vegetation life cycles. *Acta Oecol*. 20: 209-217.

Skarpe, C. 1991. Spatial patterns and dynamics of woody vegetation in an arid savanna. *J. Veg. Sci*. 2: 565-572.

Sokal, R.R. & Rohlf, F.J. 1995. *Biometry: the principles and practice of statistics in biological research*. Freeman, San Francisco, CA.

Stoyan, D. 1988. Thinnings of point processes and their use in the statistical analysis of a settlement pattern with deserted villages. *Statistics* 19: 45-56.

Turner, M.G. & Gardner, R.H. 1991. Quantitative methods in landscape ecology: an introduction. In: Turner, M.G. & Gardner, R.H. (eds.) *Quantitative methods in landscape ecology*. pp. 3-14. Springer Verlag, New York, NY.

van Lieshout, N.N.M. & Baddeley, A.J. 1997. A nonparametric measure of spatial interaction in point patterns. *Stat. Neerl*. 50: 344-361.

van Lieshout, N.N.M. & Baddeley, A.J. 1999. Indices of dependence between types in multivariate point patterns. *Scand. J. Stat*. 26: 511-532.

Wilson, J.B. & Agnew, A.D. 1992. Positive-feedback switches in plant communities. *Adv. Ecol. Res*. 23: 263-323.

Wilson, W.G. & Nisbet, R.M. 1997. Cooperation and competition along smooth environmental gradients. *Ecology* 78: 2004-2017.

Received 19 December 2001;
Revision received 29 May 2002;
Final version received 19 October 2002;
Accepted 19 October 2002.
Coordinating Editor: E. van der Maarel.

**FORUM**
**FORUM**
**FORUM**

FORUM is intended for new ideas or new ways of interpreting existing information. It provides a chance for suggesting hypotheses and for challenging current thinking on ecological issues. A lighter prose, designed to attract readers, will be permitted. Formal research reports, albeit short, will not be accepted, and all contributions should be concise with a relatively short list of references. A summary is not required.

# *Additive apportioning of species diversity: towards more sophisticated models and analyses*

***Pierre Couteron, ENGREF, UMR AMAP, Boulevard de la Lironde, TA40-PS2, FR-34398 Montpellier Cedex 05, France (couteron@engref.fr). – Raphaël Pélissier, IRD, UMR AMAP, Boulevard de la Lironde, TA40-PS2, FR-34398 Montpellier Cedex 05, France.***

**As a follow-up to the recent interest in simple additive models for diversity partitioning, we have used the analogy with ANOVA models to propose a consistent framework that allows both (i) apportioning the amount of regional diversity into environmentally and/or geographically-related components (ii) detecting individual species' habitat preferences using permutation tests. We addressed geographical aspects by relating diversity partition to dissimilarity measurements. An illustration based on an inventory of trees in a tropical rain forest of French Guiana is provided.**

In their recent review, Veech et al. (2002) listed some papers that explicitly referred to Allan's (1975) and Lande's (1996) additive models that apportion species diversity, assessed from a pooled set of samples, into within and among samples components (i.e. alpha vs beta diversity). The topic may be, however, of great concern for a wider community of ecologists who express a common interest in either: (i) quantifying the amount of regional diversity that may be accounted for via environmentally and/or spatially related components (Condit et al. 2002, Duivenvoorden et al. 2002); (ii) detecting distributions of individual species that are biased in favour of particular habitats or geographical areas (Clark et al. 1999, Harms et al. 2001). Although these two aspects have been mostly treated through distinct quantitative approaches, they are, nevertheless, strongly connected, and we shall aim in the sequel to demonstrate how more complete models of diversity partitioning may help to address both topics within a consistent framework. As diversity indices are nothing but particular expressions of the overall concept of variance, it is clear that the objective of diversity apportionment could greatly benefit from the considerable attention that historically has been paid to analyses of variance. Our aim here is thus to present a nested scheme of diversity partitioning, that generalises the simple additive framework proposed by Lande (1996) and Veech et al. (2002), while also allowing investigations into species/environment relationships.

## Total diversity as a function of species' variances

Let us suppose that a given region has been sampled via many sampling units (e.g. plots or traps) within which N individual organisms (e.g. trees) have been enumerated and identified with respect to a consistent level of taxonomic resolution (usually species). Let Xi(k) be a binary random variable indicating whether an arbitrary individual k belongs to the particular species i. Provided that the total number of individuals in the region is sufficiently large with respect to the number of sampled individuals, Xi has an expectation of $p_{i+}$ and a variance of $SV(i) = p_{i+}(1 - p_{i+})$, where $p_{i+}$ denotes the relative frequency (i.e. $n_{i+}/N$) of species, i.

Total diversity of the region can be expressed in a very general way as a function of the species' variances, SV(i):

$$TD = \sum_i w_i [p_{i+}(1 - p_{i+})] \quad (1)$$

where $w_i$ is a weighting function, that expresses the weight we want to give to species i in the overall quantification of regional diversity. Taking $w_i$ as constant and equal to one, whatever the species, obviously means that diversity is quantified via the Simpson–Gini index, while taking $w_i = 1/p_{i+}$ or $w_i = \log(1/p_{i+})/(1 - p_{i+})$ means equating TD with species richness (minus one) or with the Shannon index, respectively (Pélissier et al. 2003). As noted by Patil and Taillie (1982), it is clear that the sensibility to rare species of the diversity assessment is maximal when using the richness and minimal with the Simpson–Gini index. Furthermore, as long as diversity quantification relies on those three very popular indices, any partition

of species variance, SV(i), can be easily translated into a partition of total diversity via an appropriate choice of $w_i$.

## Partitioning species' variance

We shall now further suppose that the whole set of plots can be unequivocally partitioned with respect to a categorical variable, J, of which classes are denoted by the subscript j relating, for instance, to ecological factors (e.g. soil types) or to distinct ecological communities. Using a standard result (Saporta 1990 p. 68), we can express the marginal variance SV(i) as the sum of the expected value of the conditional variance and of the variance of the conditional means, namely:

$$\begin{aligned} SV(i) = Var_k(Xi) = E_j(Var_k(Xi|J=j)) \\ + Var_j(E_k(Xi|J=j)) \end{aligned} \quad (2)$$

Subscripts denote on which hierarchical level expectations and variances are computed (k = individual, j = ecological class). Eq. 2 can be further specified as (proof in Appendix):

$$\begin{aligned} SVac(i) &= Var_j(E_k(Xi|J=j)) \\ &= \sum_j p_{+j}(p_{ij}/p_{+j} - p_{i+})^2 \end{aligned} \quad (2a)$$

and

$$SVwc(i) = E_j(Var_k(Xi|J=j)) = \sum_j p_{ij}(1 - p_{ij}/p_{+j}) \quad (2b)$$

Here $p_{ij}$ denotes the relative frequency ($n_{ij}/N$) of individuals that are both belonging to species i and sampled in the ecological class j, while $p_{+j}$ stands for the relative frequency of individuals sampled in class j. We provide here a justification of the results given by Lande (1996) for diversity partitioning into within- and among-classes portions, $D_{wc}$ and $D_{ac}$ (i.e. alpha vs beta diversity), which can be directly computed from SVwc(i) and SVac(i), respectively.

Testing the value reached by SVac(i) against values found under a null hypothesis of a random distribution of species across ecological classes, would obviously be relevant to assess whether the distribution of species i is biased with respect to ecological classes. From this standpoint, SVac(i) is not unrelated to the "weighted preference index (WPI)" introduced by Clark et al. (1999), though the WPI cannot be integrated in a consistent partition of diversity. To SVac(i) values, a Monte-Carlo approach with a complete randomization of individuals (trees) across ecological classes may be thought of. But, as soon as individuals are sampled via field plots (or traps), and this is likely to be so in most studies aimed at assessing diversity, total randomization is to be questioned since individuals found in a given plot are not statistically independent from each other. As a consequence, an adequate test for SVac(i) (or for any related index) demands an explicit consideration of the underlying sampling design.

## Taking the sampling design into account

Ecological classes, indexed by j, being defined as mutually exclusive sets of plots, we shall consider a further partition of the within classes variance, SVwc(i), into among plots and within plots portions (denoted P(i) and R(i), respectively). This is done by applying, once again, the formula for variance decomposition with respect to the categorical variable, Q, that indicates the plot. Namely:

$$\begin{aligned} Var_{k\in j}(Xi|J=j) = E_{q\in j}(Var_{k\in q}((Xi|J=j)|Q=q)) \\ + Var_{q\in j}(E_{k\in q}((Xi|J=j)|Q=q)) \end{aligned} \quad (3)$$

Subsequent manipulations lead (Appendix) to an explicit expression for P(i), i.e. the portion of species variance attributable to plots, once ecological classes have been taken into account:

$$\begin{aligned} P(i) &= E_j(Var_{q\in j}(E_{k\in q}((Xi|J=j)|Q=q))) \\ &= \sum_j p_{+j} \sum_{q\in j} \frac{p_{+q}}{p_{+j}}(p_{iq}/p_{+q} - p_{ij}/p_{+j})^2 \end{aligned} \quad (3a)$$

Here $p_{iq}$ is the relative frequency ($n_{iq}/N$) of individuals belonging to species i and sampled in plot q, and $p_{+q}$ is the relative frequency ($n_{+q}/N$) of individuals in plot q.

The residual portion, R(i), which is the within plots variance is defined as (Appendix):

$$\begin{aligned} R(i) &= E_j(E_{q\in j}(Var_{k\in q}((Xi|J=j)|Q=q))) \\ &= \sum_j \sum_{q\in j} p_{iq}(1 - p_{iq}/p_{+q}) \end{aligned} \quad (3b)$$

And the complete decomposition of both species variance and total diversity can be written as:

$$SV(i) = SVac(i) + P(i) + R(i)$$

$$TD = D_{ac} + P + R$$

As for a two-level nested ANOVA (Sokal and Rohlf 1995), tests based on SVac(i) and P(i) are informative on potential biases in species distributions with respect to ecological classes and to plots within ecological classes, respectively. In analogy with nested ANOVA (Sokal and Rohlf 1995 p. 272), these values should be transformed into pseudo-F ratios taking into account the appropriate degrees of freedom (Table 1). Monte-Carlo testing of SVac(i) is to be carried out via a randomization of plots across ecological classes, while for P(i), the randomization of individuals across plots is to be carried out in each ecological class (Anderson and Ter Braak 2003). Such tests may also concern the break-down of total diversity (Eq. 1), once the choice of a weighting function ($w_i$) and, thus, of a diversity index, has been made.

Nevertheless, we are fully aware that randomization of plots across ecological classes is not a fully satisfactory null model, due to potential biases induced by spatial autocorrelation in species' abundance. Alternative null models (Roxburgh and Matsuki 1999) are available and may be useful. In this paper, we shall, however, introduce another aspect which is the direct study of spatial patterns of diversity.

## Dissimilarity between plots: investigating spatial patterns of diversity

Dissimilarity/similarity coefficients, especially Jaccard's or Sorensen's, have been widely used to quantify beta diversity (Condit et al. 2002, Ruokolainen et al. 2002), but as most other popular coefficients, they cannot be integrated within the framework of an additive partitioning of total diversity. On the other hand, considering the quantity

$$\Delta_{qq'}(i) = (p_{iq}/p_{+q} - p_{iq'}/p_{+q'})^2 \quad (4)$$

which measures the contrast between plots q and q′ relative to species i, opens up the way to various consistent apportionments of both dissimilarity and beta diversity.

First, we can note that, when considering all species, the quantity

$$D_{qq'} = \sum_j w_i \Delta_{qq'}(i) \quad (5)$$

can be a dissimilarity coefficient in the sense of Rao's (1982) very comprehensive theory about diversity/dissimilarity. We may compute $D_{qq'}$ for any couple of plots, q and q′, or for couples of plots belonging to a given ecological class.

Second, given now that any variance (i.e. a weighted sum of departures around the mean) can be expressed as a weighted sum of differences between pairs of observations (Appendix), it is possible to rewrite P(i) (defined by Eq. 3a), as:

$$P(i) = \sum_j \frac{1}{2p_{+j}} \sum_{q\in j} \sum_{q'\in j} p_{+q}p_{+q'}\Delta_{qq'}(i) \quad (6)$$

or for all species

$$P = \sum_j \frac{1}{2p_{+j}} \sum_{q\in j} \sum_{q'\in j} p_{+q}p_{+q'}D_{qq'} \quad (7)$$

These formulations of P(i) and P (Eq. 6, 7) can be used to further partition the diversity within ecological classes with respect to distances between plots, thereby providing a consistent apportionment of regional diversity into spatial (or geographical) versus environmental components (as attempted by Duivenvoorden et al. 2002 from the variability of Steinhaus' similarity coefficient).

For instance, defining a set of mutually exclusive distance classes, $H_h$, allows us to rewrite P(i) as:

$$P(i) = \sum_j \frac{1}{2p_{+j}} \sum_h \sum_{\substack{d(q,q')\in h \\ q,q'\in j}} p_{+q}p_{+q'}\Delta_{qq'}(i) \quad (8)$$

P(i) is fundamentally an interaction term between ecological and distance classes, which may be considered in several ways. One possibility is to apportion P(i) with respect to different spatial ranges, and to test whether a particular class of distance accounts for a share of species variance higher or lower than expected. The null hypothesis is generated by randomly re-allocating the floristic composition to the geographical locations of the plots (Cressie 1993 p. 597), while preserving the relationship between plots and ecological classes as to remain consistent with a hierarchical model of diversity decomposition.

In Eq. 8, contributions from successive distance classes are obviously additive though they cannot be directly compared, since the distance classes do not systematically relate to a constant number of couples of plots and of individuals. A standardisation may nevertheless be obtained by computing an average dissimilarity for each distance class, namely:

$$D(h) = \frac{\sum_{d(q,q')\in H_h} p_{+q}p_{+q'}D_{qq'}}{2\sum_{d(q,q')\in H_h} p_{+q}p_{+q'},} \quad \text{with } H_h \text{ being a distance class centred around h} \quad (9)$$

D(h) is homologous to the variogram, which is a basic tool to investigate into spatial patterns displayed by quantitative variables (Cressie 1993 p. 69). Equation 9 can thus be useful to quantify the changes in species composition with distance, either within ecological classes, landscape units or over a biogeographic region.

## Illustration from tropical rain forest data

The above principles have been applied to a data-set originating from a forest inventory covering ca 10 000 ha of tropical rain forest in French Guiana. Partitioning floristic diversity of rain forests is particularly challenging at a mesoscale of, say, $1–10^3$ km$^2$, for which results are badly missing. The inventory was based on 411 rectangular plots of 0.3 ha each, located at the nodes of a systematic sampling grid of $400 \times 500$ m. Fiftynine botanical species were recognised as being reliably identified by field tree spotters. Ecological classes were defined on the basis of a synthetic variable (12 categories) expressing both the topographical situation and

Table 1. Diversity partitioning for the 59 species observed in 411 plots in Counami Forest.

| Diversity index | Total diversity (TD) | Beta diversity | | Alpha diversity (within plots) (R) | Pseudo-F ratio* | |
|---|---|---|---|---|---|---|
| | | Among ecological classes ($D_{ac}$) | Among plots within classes (P) | | $\frac{D_{ac}/(nJ-1)}{P/(nQ-nJ)}$ | $\frac{P/(nQ-nJ)}{R/(N-nQ)}$ |
| Richness-1 | 58 | 0.3632 | 5.2059 | 52.4309 | 2.53 | 1.68 |
| Shannon | 3.1058 | 0.0354 | 0.3264 | 2.744 | 3.93 | 2.02 |
| Simpson-Gini | 0.8897 | 0.0129 | 0.1037 | 0.7731 | 4.51 | 2.27 |

* nJ, nQ and N are the number of ecological classes (12), the number of plots (411) and the total number of individuals (7189), respectively.

the water regime (e.g. flooding, seasonal soil saturation) of the plots (Couteron et al. 2003 for more details regarding the inventory and the data-set). Several R routines (R Development Core Team 2004) allowing to carry out all the computations used in this paper are freely available from the authors (http://pelissier.free.fr/Diversity.html).

Diversity partitioning for the three well-known indices is presented in Table 1. Both pseudo-F ratios tended to be higher when using Simpson–Gini than when using the richness, whilst the Shannon's index yielded intermediate values. Whatever the index, observed values of $D_{ac}$ and P appeared highly significant in the light of the Monte-Carlo tests, since neither of them was exceeded by any results of 10 000 randomizations ($P < 1.10^{-4}$).

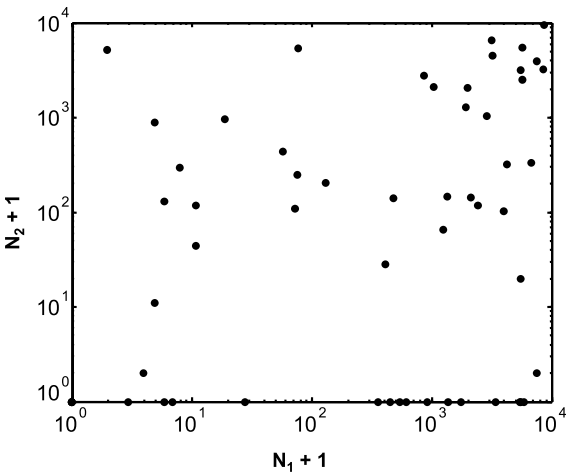


Fig. 1. Results of Monte-Carlo tests for the 59 species observed in 411 plots laid out in Counami forest (French Guiana). Each species (closed circle) is plotted with respect to the numbers $N_1$ and $N_2$ (logarithmic scale) of randomizations (among 10 000) yielding results exceeding the observed values for SVac(i) (significance of variance among ecological classes) and P(i) (significance of variance among plots within ecological classes), respectively. Note that six species are grouped together at the origin of the plot since yielding observed values of both SVac(i) and P(i) that were never exceeded by any result of the randomizations.

More analytically, we also considered results of the Monte-Carlo tests for each species (Fig. 1). By a common standard ($P < 0.01$), 23 species had a significant share of their variance attributable to topography (significant value of SVac(i)), while 28 species displayed a significant inter-plot variance within ecological classes (significant value of P(i)). Thirteen species yielded significant results for both SVac(i) and P(i). For 21 species, values provided by 10 000 randomizations never exceeded the observed value of P(i) ($P < 1.10^{-4}$), among which only six yielded a never-exceeded value for SVac(i). Distributions of some individual species displayed clearer biases with respect to space (inter-plot variation) than to ecological classes. But pseudo-F ratios, which integrate the whole set of species, pointed towards a substantial influence of ecological classes (Table 1). Hence, both standpoints appear complementary.

To investigate spatial organisation, we considered two additional variance components: the "local" variance between "neighbouring" plots and the "long-range" variance between "very distant plots". Here, we defined the local variance from distance classes less than 1 km and the long-range variance from distance classes more than 5 km. We assessed the level of statistical significance of both range-related effects by counting the number of randomization results falling above or below the observed values. At short range, we expect a similarity between two randomly chosen neighbouring plots greater than for two arbitrary plots (positive spatial dependence) and, thus, an observed variance smaller than under the null hypothesis. This should determine a very large number of randomization results exceeding the observed values. Indeed, we found eight species displaying a positive spatial dependence at a local-scale, since their observed local variances were exceeded more than 9900 times (among 10 000) by randomization results ($P < 0.01$). Conversely, we found only one species displaying a negative spatial dependence, with an observed variance being exceeded no more than 100 times. On a large scale, there were only four species having observed variances between very distant plots significantly higher than expected (macro-heterogeneity in spatial distribution), and one

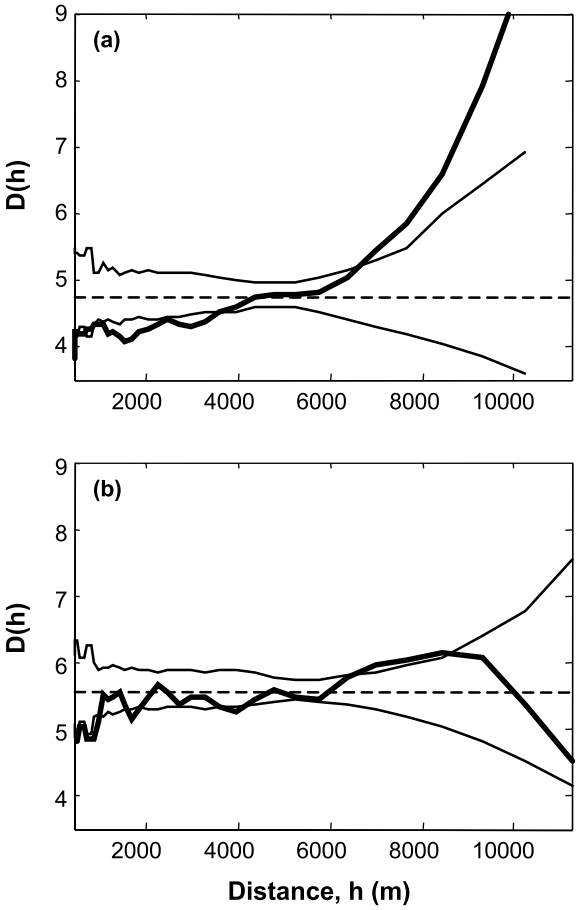


Fig. 2. Floristic average dissimilarity (based on species richness computed from a pool of 59 species), D(h), as a function of the distance between plots, h (bold line). The interrupted line expresses the mean dissimilarity for all pairs of plots, irrespective of distance, while the thin solid lines mark the envelopes of the 10% bilateral confidence interval obtained from 300 random re-allocations of floristic composition to geographical locations of plots. a) plots from uplands. b) plots from bottom-lands.

species for which the converse was observed ($P < 0.01$ for both results).

We then considered the evolution of the average plot dissimilarity, D(h), for distances ranging from 400 m to more than 12 km. We made two distinct analyses by separating plots relating to bottom-lands, namely talwegs and foot-slopes (6 ecological classes, 201 plots) from plots corresponding to uplands, i.e. slopes, plateaux and hilltops (6 ecological classes, 210 plots). On Fig. 2, we only displayed the average dissimilarity based on species richness, although versions relying on Shannon's or Simpson–Gini's indices were considered as well. Using richness, the average floristic dissimilarity has a very simple meaning, namely the average number of species not shared by two arbitrary plots, and D(h) expresses how this number varies with increasing distance, h. For upland plots (Fig. 2a), D(h) is significantly lower than expected for $h < 4$ km and significantly higher for distances above 7 km (confidence envelopes were constructed via 300 randomizations of the plots' floristic composition throughout geographic locations). In the Counami forest, there is a strong floristic macro-structure that has been explicitly mapped from ordination results (Couteron et al. 2003). Interestingly, such a structure was strongly attenuated for plots from bottom-lands, a situation in which floristic dissimilarity did not substantially depart from expectation but for very short distances ($h < 500$ m; Fig. 2b).

## Conclusion

Ecological studies searching for determinants of species diversity should benefit from analytical methods based on variance decomposition, which is a well-established statistical standard. Seeing species richness, Shannon and Simpson–Gini indexes as functions of species' variances – Eq. 1 – is relevant, not only to bridge the gap between diversity assessment and ordination methods (Pélissier et al. 2003), but also to apportion diversity using the rich and flexible framework of ANOVA, and also ANCOVA, models (Sokal and Rohlf 1995) for which permutation-based tests of statistical significance are available (Anderson and Ter Braak 2003). The link with usual methods for studying spatial variation (e.g. the variogram), is also worth noting. Such methods and models are indeed appropriate candidates to explicitly consider the particular features of different sampling designs, either simple or nested, with varying numbers of classification levels, factors and covariables,

*Acknowledgements* – We are indebted to D. Chessel (University of Lyon-1) for important insights relating to diversity quantification and also to F. Houllier (INRA) for valuable comments on a first draft of this paper.

## References

Allan, J. D. 1975. Components of diversity. – Oecologia 18: 359–367.

Anderson, M. J. and Ter Braak, C. J. F. 2003. Permutation tests for multi-factorial analysis of variance. – J. Statist. Comput. Simul. 73: 85–113.

Clark, D. B., Palmer, M. W. and Clark, D. A. 1999. Edaphic factors and the landscape-scale distributions of tropical rain forest trees. – Ecology 80: 2662–2675.

Condit, R., Pitman, N. C. A., Leigh, E. G. J. et al. 2002. Beta-diversity in tropical forest trees. – Science 295: 666–669.

Couteron, P., Pélissier, R., Mapaga, D. et al. 2003. Drawing ecological insights from a management-oriented forest inventory in French Guiana. – For. Ecol. Manage. 172: 89–108.

Cressie, N. A. C. 1993. Spatial statistics. – John Wiley & Sons Inc.

Duivenvoorden, J. F., Svenning, J.-C. and Wright, S. J. 2002. Beta diversity in tropical forests. – Science 295: 636–637.
Harms, K. E., Condit, R., Hubbell, S. P. et al. 2001. Habitat associations of trees and shrubs in a 50-ha neotropical forest plot. – J. Ecol. 89: 947–959.
Lande, R. 1996. Statistics and partitioning of species diversity, and similarity among multiple communities. – Oikos 76: 5–13.
Patil, G. P. and Taillie, C. 1982. Diversity as a concept and its measurement. – J. Am. Stat. Assoc. 77: 548–567.
Pélissier, R., Couteron, P., Dray, S. et al. 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. – Ecology 84: 242–251.
Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. – Theor. Popul. Biol. 21: 24–43.
R Development Core Team 2004. R: a language and environment for statistical computing. R Foundation for Statistical Computing (http://www.R-project.org/).
Roxburgh, S. H. and Matsuki, M. 1999. The statistical validation of null models used in spatial association analysis. – Oikos 85: 68–78.
Ruokolainen, K., H., Tuomisto, H., Chave, J. et al. 2002. Beta-diversity in tropical forests. – Science 297: 1439a.
Saporta, G. 1990. Probabilités, analyse des données et statistique. – Technip.
Sokal, R. R. and Rohlf, F. J. 1995. Biometry: the principles and practice of statistics in biological research, 3rd edn. – Freeman.
Veech, J. A., Summerville, K. S., Crist, T. O. et al. 2002. The additive partitioning of species diversity: recent revival of an old idea. – Oikos 99: 3–9.

## Appendix

**1) Proof for Eq. 2a and 2b: decomposition of species variance with respect to environmental classes**

We start from the formula for marginal variance decomposition (Eq. 2) with addition of subscripts mentioning at which level of the hierarchy expectation or variance are computed (k = individuals, j = environmental classes). For instance, $E_j(...)$ denotes an expectation computed over all environmental classes, while $Var_{k\in j}(...)$ stands for variance computation over all individuals within an environmental class.

$$SV(i) = Var_k(Xi) = E_j(Var_{k\in j}(Xi|J)) + Var_j(E_{k\in j}(Xi|J))$$

Note that J = j is omitted for simplicity.

- Variance of expectations:

Noting that:

$$Var_j(E_{k\in j}(Xi|J)) = E_j[[E_{k\in j}(Xi|J) - E_j(E_{k\in j}(Xi|J))]^2]$$

while

$$E_j(E_{k\in j}(Xi|J)) = E_k(Xi) = p_{i+}$$

and

$$E_{k\in j}(Xi|J)) = p_{ij}/p_{+j}$$

leads to

$$SVac(i) = Var_j(E_{k\in j}(Xi|J)) = \sum_j p_{+j}(p_{ij}/p_{+j} - p_{i+})^2 \text{ i.e.} \quad \text{Eq. 2a}$$

- Expectation of variances:

$$Var_{k\in j}(Xi|J) = \frac{p_{ij}}{p_{+j}}\left(1 - \frac{p_{ij}}{p_{+j}}\right) \text{ and thus:}$$

$$SVwc(i) = E_j(Var_{k\in j}(Xi|J)) = \sum_j p_{+j}\frac{p_{ij}}{p_{+j}}\left(1 - \frac{p_{ij}}{p_{+j}}\right) = \sum_j p_{ij}\left(1 - \frac{p_{ij}}{p_{+j}}\right) \text{ i.e.} \quad \text{Eq. 2b}$$

**2) Proof for Eq. 3a and 3b: decomposition with respect to plots after the partialling out of environmental classes**

We now consider the partitioning introduced in Eq. 3, namely:

$$Var_{k\in j}(Xi|J) = E_{q\in j}(Var_{k\in q}((Xi|J)|Q)) + Var_{q\in j}(E_{k\in q}((Xi|J)|Q))$$

(dropping J = j and Q = q for the sake of simplicity; subscript q means taking expectation or variance over the plots).

Environmental classes are defined as mutually exclusive sets of plots, which means $(Q = q) \subset (J = j)$ whatever the plot q considered in an arbitrary class j.

$$E_{k\in q}((Xi|J)|Q) = Prob((Xi\cap J)\cap Q)/Prob(Q) = Prob(Xi\cap Q)/Prob(Q) = E_{k\in q}(Xi|Q) = p_{iq}/p_{+q}$$

and in a similar way: $Var_{k\in q}((Xi|J)|Q) = Var_{k\in q}(Xi|Q)$

- Variance of expectations:

$$Var_{q\in j}(E_{k\in q}(Xi|Q)) = E_{q\in j}[[E_{k\in q}(Xi|Q) - E_{q\in j}(E_{k\in q}(Xi|Q))]^2] = E_{q\in j}[[E_{k\in q}(Xi|Q) - E_{k\in j}(Xi|J)]^2] = \sum_{q\in j}\frac{p_{+q}}{p_{+j}}[p_{ij}/p_{+q} - p_{ij}/p_{+j}]^2$$

Consequently:

$$P(i) = E_j(Var_{q\in j}(E_{k\in q}(Xi|Q))) = \sum_j p_{+j}\sum_{q\in j}\frac{p_{+q}}{p_{+j}}[p_{iq}/p_{+q} - p_{ij}/p_{+j}]^2 \text{ i.e.} \quad \text{Eq. 3a}$$

- Expectation of variances:

$$Var_{k\in q}(Xi|Q) = \frac{p_{iq}}{p_{+q}}\left(1 - \frac{p_{iq}}{p_{+q}}\right)$$

$$E_{q\in j}(Var_{k\in q}(Xi|Q)) = \sum_{q\in j}\frac{p_{+q}}{p_{+j}}\frac{p_{iq}}{p_{+q}}\left(1 - \frac{p_{iq}}{p_{+q}}\right)$$

$$R(i) = E_j(E_{q\in j}(Var_{k\in q}(Xi|Q)))$$

$$= \sum_j p_{+j} \sum_{q\in j} \frac{p_{iq}}{p_{+j}} \left(1 - \frac{p_{iq}}{p_{+q}}\right) \text{ i.e. Eq. 3b}$$

with $SVwc(i) = P(i) + R(i)$

**3) Proof for Eq. 6: expressing the "among-plots within-classes" variance on the basis of the expectation of the squared inter-plot difference**

Let us consider the quantity:

$$A = \frac{1}{2} E_{\substack{q\in j \\ q'\in j}} ([E_{k\in q}(Xi|Q) - E_{k\in q'}(Xi|Q')]^2)$$

$$= \frac{1}{2} \sum_{q\in j} \sum_{q'\in j} \frac{p_{+q}}{p_{+j}} \frac{p_{+q'}}{p_{+j}} (p_{iq}/p_{+q} - p_{iq'}/p_{+q'})^2$$

$$A = \frac{1}{2} E_{\substack{q\in j \\ q'\in j}} [(E_{k\in q}(Xi|Q))^2 + E_{k\in q'}(Xi|Q')^2 - 2E_{k\in q}(Xi|Q)E_{k\in q'}(Xi|Q')]$$

We can first note that:

$$E_{\substack{q\in j \\ q'\in j}} (E_{k\in q}(Xi|Q)^2) = E_{\substack{q\in j \\ q'\in j}} (E_{k\in q'}(Xi|Q')^2)$$

$$= E_{q\in j}(E_{k\in q}(Xi|Q)^2)$$

and that:

$$E_{\substack{q\in j \\ q'\in j}} (E_{k\in q}(Xi|Q)E_{k\in q'}(Xi|Q'))$$

$$= E_{q\in j}(E_{k\in q}(Xi|Q)E_{q'\in j}(E_{k\in q'}(Xi|Q')))$$

$$= E_{q\in j}(E_{k\in q}(Xi|Q)E_{k\in j}(Xi|J)) = E_{k\in j}(Xi|J)^2$$

Thus:

$$A = \frac{1}{2} [2E_{q\in j}(E_{k\in q}(Xi|Q)^2) - 2E_{k\in j}(Xi|J)^2]$$

$$= Var_{q\in j}(E_{k\in q}(Xi|Q))$$

As a consequence:

$$P(i) = E_j(Var_{q\in j}(E_{k\in q}(Xi|Q))) = E_j(A)$$

$$= \sum_j p_{+j} \frac{1}{2} \sum_{q\in j} \sum_{q'\in j} \frac{p_{+q}}{p_{+j}} \frac{p_{+q'}}{p_{+j}} \times (p_{iq}/p_{+q} - p_{iq'}/p_{+q'})^2 \text{ i.e. Eq. 6}$$

*Ecology*, 86(4), 2005, pp. 828–834


# A GENERALIZED, VARIOGRAM-BASED FRAMEWORK FOR MULTI-SCALE ORDINATION

PIERRE COUTERON[1,3] AND SÉBASTIEN OLLIER[2]

[1]*École Nationale du Génie Rural des Eaux et des Forêts/UMR botAnique et bioinforMatique de l'Architecture des Plantes (AMAP), Boulevard de la Lironde, TA40/PS2, 34398 Montpellier Cedex 05, France, and French Institute of Pondicherry, 11 Saint Louis Street, Pondicherry 605001, India*

[2]*UMR CNRS 5558, Laboratoire de Biométrie et Biologie Evolutive, Université Claude Bernard Lyon 1, 69 622 Villeurbanne Cedex, France*

STATISTICAL REPORTS

*Abstract.* Multi-scale ordination (MSO) deals with potential scale dependence in species assemblages by studying how results from multivariate ordination may be different at different spatial scales. MSO methods were initially based on two-term local covariances between species and, therefore, required sampling designs composed of adjacent quadrats. A variogram-based MSO, recently introduced by H. H. Wagner, is applicable to very diverse sampling designs and for use with principal-components analysis, correspondence analysis, and derived ''two-table'' (also called ''direct'') ordination methods, i.e., redundancy analysis and canonical correspondence analysis.

In this paper we put forward an enlarged framework for variogram-based MSO that relies on a generalized definition of inter-species covariance and on matrix expression of spatial contiguity between sampling units. This enables us to provide distance-explicit decompositions of variances and covariances (in their generalized meaning) that are consistent with many ordination methods in both their single- and two-table versions. A spatially explicit apportioning of diversity indices is proposed for some particular definitions of variance. Referring to two-table ordination methods allowed the multi-scale study of residual spatial patterns after factoring out available environmental variables. Some aspects of the approach are briefly illustrated with vegetation data from a Neotropical rain forest in French Guiana.

*Key words: canonical correspondence analysis; diversity apportioning; multi-scale ordination; multivariate geostatistics; spatial contiguity; species assemblages; tropical rainforest; variogram.*

## INTRODUCTION

Determining to what extent multi-species patterns of association may be different on different spatial scales is obviously a central issue in ecology (Levin 1992). The concern to integrate space in numerical studies of inter-species association has led to the development of a method of multi-scale ordination (MSO; Ver Hoef and Glenn-Lewin 1989). This requires data from continuous sampling designs (e.g., belt transects) since it is based on the computation of two-term local covariances between species (Greig-Smith 1983). In two recent papers, Wagner (2003, 2004) proposed a new method for MSO based on the variogram (Wackernagel 1998), thereby allowing the use of data collected by means of very diverse sampling designs. This insightful approach was proposed for two usual ordination methods, i.e., principal-components analysis (PCA) and correspondence analysis (CA). Extension to ''direct'' ordination methods (Legendre and Legendre 1998), using two data tables, such as redundancy analysis (RDA, relating to PCA) and canonical correspondence analysis (CCA, relating to CA) was also proposed by Wagner (2004).

This most recent contribution is a considerable step forward since CA is generally preferred to PCA for the study of inter-species associations. CA is a very popular and powerful method that positions species and sites along common ordination axes, by applying the same centering and weighting options to rows and columns of the site by species table. However, in spite of several attractive properties, there is no reason to consider CA as being automatically the most appropriate ordination method whatever the characteristics of the data and the aim of the study (Gimaret-Carpentier et al. 1998). Several alternatives to CA, with distinct properties, can be defined by changing weighting options for either sites or species. For instance, Pélissier et al. (2003) demonstrated that changing species weighting, i.e., placing varying degrees of emphasis on scarce species, could be used to define three ordination methods (including CA) and that each was consistent with one classical diversity index (richness, Shannon's, Simpson's). On the other hand, Dolédec et al. (2000) used a uniform weighting of sites to derive

Manuscript received 17 December 2003; revised 23 July 2004; accepted 6 September 2004; final version received 22 October 2004. Corresponding Editor: G. M. Henebry.

[3] Present address: French Institute of Pondicherry, 11 Saint Louis Street, Pondicherry 605001, India.
E-mail: Pierre.couteron@ifpindia.org

an alternative to CCA with interesting properties for the separation of species niches. Hence, it would be preferable for ecologists to become aware of the potential adaptability of both single- and two-table methods of ordination to their specific aims and to the characteristics of their data. Such adaptability should also encompass the emerging field of spatially explicit ordinations.

Our work here was triggered by the pioneering work conducted by Wagner (2003, 2004) but aims to define a broader framework for variogram-based multi-scale ordinations. We demonstrate that it is possible to partition by distance the results of very diverse ordination methods, as defined by re-scaling and weighting options for the rows and columns of the data tables. To do so, we introduce a generalized definition of covariance between species that encompasses several ordination methods while being amenable to scale-explicit decompositions. We also highlight the link with the additive decomposition of common diversity indices (Pélissier et al. 2003, Couteron and Pélissier 2004) and refer to methods of two-table ''direct'' ordination to carry out the explicit analysis of residual spatial patterns after factoring out some environmental variables. This aspect is then emphasized in a brief illustration based on vegetation data from a neotropical rainforest. In this report, we have chosen to keep mathematical developments to a minimum while providing a complete treatment in matrix form in an appendix. To accomplish the multi-scale ordinations, computer programs have been developed for Matlab or R (Ihaka and Gentleman 1996) (ade4 package, *available online*).[4]

## A Generalized Definition of Covariance

Data tables containing counts of individual organisms by sampling sites (say ''quadrats'') and taxa (usually species) are both a central and general feature of ecological studies. Let us consider such a table, based on $N$ sampled individuals, for which $f_{ai}$ is the total number of individuals belonging to species $i$ ($1 < i < S$) that were counted in quadrat $a$ ($1 < a < Q$). Let $p_{ai}$ be the corresponding relative frequency ($p_{ai} = f_{ai}/N$) while $p_{a+}$ and $p_{+i}$ are the relative frequencies for quadrat $a$ and species $i$, respectively. We have introduced our topic with explicit reference to counted individuals, though the above parameters remain meaningful as long as $f_{ai}$ is a nonnegative value (biomass measurements, semi-quantitative indices of abundance, presence/absence, . . . ) expressing the abundance of species $i$ in quadrat $a$.

Ordination methods such as correspondence analysis (CA) and various versions of Principal-components analysis (PCA; ter Braak 1983) are the usual tools employed to analyze quadrats by species tables. Central to all these methods is the application of singular values decomposition (SVD), also called ''eigenanalysis,'' to a square $S$-by-$S$ matrix, which is the usual variance–covariance matrix, $\mathbf{C}$, for the species-centered (non-standardized) PCA and which is another matrix $\mathbf{Q^2}$ in the case of CA (see Legendre and Legendre [1998:453] and Wagner [2004] for details). In $\mathbf{Q^2}$, terms on the diagonal are homologous to variances and are proportional to the portions of the total chi square of the data table (Legendre and Legendre 1998:452) that are attached to each of the $S$ species. Off-diagonal terms are homologous to the usual pairwise covariances and measure to what extent two arbitrary species may conjointly depart from expected abundance values.

We can see matrices $\mathbf{C}$ and $\mathbf{Q^2}$ as nothing more than special cases of a square matrix $\mathbf{G_T} = [g_{ij}]_{1\le i\le S, 1\le j\le S}$ based on an appropriate generalization of the notions of species variance (diagonal values) and covariance (off-diagonal values). This generalized measure of covariance is, for two arbitrary species $i$ and $j$,

$$g_{ij} = \frac{1}{2}\sum_{a=1}^{Q}\sum_{b=1}^{Q}(x_{ai} - x_{bi})\sqrt{w_i w_j}(x_{aj} - x_{bj})\delta_a\delta_b \quad (1)$$

with variance being a special case where $i = j$.

Here, $w_i$ weights the influence of species $i$, while $\delta_a$ and $\delta_b$ are the weights given to quadrats $a$ and $b$, respectively; $x_{ai}$ denotes any measure of abundance of species $i$ in quadrat $a$ that can be derived from the initial value $f_{ai}$ via re-scaling options (Table 1).

The choice of weighting options is a central yet often-overlooked question when using multivariate techniques, since weighting along with re-scaling and centering defines the nature of the distance between quadrats and, for some methods, also between species. Moreover, the choice of weighting options relates to very practical questions concerning, for instance, the influence that it seems meaningful to confer to a particular species in the definition of a multi-specific assemblage, or to a given quadrat in the investigation of an ecological gradient. Addressing such questions means that the biogeographic context must be taken into account (e.g., Are there many scarce species? How abundant are the most frequent species?) along with the sampling design (Does it give a fair estimate of species abundance in a region?) and, for two-table methods, the nature of the ecological gradients under study (Are there strong limiting factors or threshold effects?). More detailed discussions on the consequences of weighting can be found in Dolédec et al. (2000; regarding quadrats in direct gradient analysis) and in Pélissier et al. (2003; regarding species).

Combining re-scaling and weighting options opens up a wide selection of ordination methods and associated properties. Some examples, based on published methods, are presented in Table 1, but other possibilities are obviously imaginable. The presentation of classical ordination methods in terms of weighting of rows and columns was introduced by Escoufier (1987)

[4] ⟨http://pbil.univ-lyon1.fr/CRAN/⟩

TABLE 1. Definition of some ordination methods from re-scaling and weighting options.

| Re-scaling options | | Weighting options for quadrats, $\delta_a$ | Weighting options for species, $w_i$ | Corresponding ordination method | Link with diversity indices |
|---|---|---|---|---|---|
| (I) | $x_{ai} = p_{ai}$ | $\delta_a = \frac{1}{Q}$ | $w_i = 1$ | species-centered principal-components, analysis, PCA† | |
| (II) | $x_{ai} = \frac{p_{ai}}{\sqrt{V_i}}$ | $\delta_a = \frac{1}{Q}$ | $w_i = 1$ | PCA on the species correlation matrix† | $I_N = S$ (richness) |
| (III) | $x_{ai} = \frac{p_{ai}}{p_{a+}}$ | $\delta_a = \frac{1}{Q}$ | $w_i = 1$ | species-centered PCA on proportions (ter Braak 1983) | $I_N =$ Simpson–Gini |
| (IV) | $x_{ai} = \frac{p_{ai}}{p_{a+}}$ | $\delta_a = p_{a+}$ | $w_i = \frac{1}{p_{+i}}$ | correspondence analysis, CA† | $D = S - 1$ (richness − 1) |
| | | | $w_i = \log\left(\frac{1/p_{+i}}{1 - p_{+i}}\right)$ | Pélissier et al. (2003) | $D =$ Shannon |
| | | | $w_i = 1$ | nonsymmetric correspondence analysis, NSCA (Gimaret-Carpentier et al. 1998) | $D =$ Simpson–Gini |

*Notes:* The definition relates to a $Q$ quadrats × $S$ species table for which $p_{ai}$ notes the relative frequency of individuals sampled in quadrat $a$ and belonging to species $i$, while $p_{a+}$ and $p_{+i}$ are the relative frequencies of quadrats and species, respectively. $V_i$ is the variance of species $i$, i.e., $V_i = (1/Q)\Sigma_a (p_{ai} - p_{+i})^2$, and $x_{ai}$ is the value of the re-scaled table for computation of the generalized covariance (see Eq. 1 in the main text). $D$ is the total diversity ($D = \Sigma_i w_i[p_{+i}(1 - p_{+i})]$) while $I_N$ is the total ''inertia'' of the ordination method defined by $x_{ai}$, $\delta_a$, and $w_i$ (trace of the generalized variance–covariance matrix, $\mathbf{G_T}$).

† Legendre and Legendre (1998).



and was used by Sabatier et al. (1989) and Dolédec et al. (2000) for several single- and two-table methods (including CA, CCA [canonical correspondence analysis] and classical versions of PCA and RDA [redundancy analysis]). Pélissier et al. (2003) used this presentation to compare the properties of the three methods corresponding to option IV in Table 1, and to investigate their relationship with diversity measures. All these authors based their presentation of the methods on a species-centered version of the data table containing differences between individual observations, $x_{ai}$, and the $\delta_a$-weighted mean value, $\bar{x}i$, found for each species. Alternatively, in Eq. 1, we use all pairwise differences between observations to compute variance and covariance. The equivalence of the two approaches is explained in the Appendix (see Eq. A.5 to Eq. A.10).

## GENERALIZED SPATIAL COVARIANCE

From Eq. 1, the contribution made by a given pair $(a,b)$ of quadrats to the covariance between two species can be expressed as

$$g_{ij}(a, b) = \frac{1}{2}(x_{ai} - x_{bi})\sqrt{w_i w_j}(x_{aj} - x_{bj})\delta_a\delta_b. \quad (2)$$

This translates easily into a generalized version of cross-variograms ($i \neq j$) and variograms ($i = j$), namely,

$$\gamma_{\mathbf{G}_{ij}}(h) = \frac{1}{K(h)} \sum_{a,b\,|h_{ab}\approx h} g_{ij}(a, b) \quad (3)$$

where $h$ is the central value of a given distance class, and where $K(h)$ is a scaling coefficient, such as the following:

$$K(h) = \sum_{a,b\,|h_{ab}\approx h} \delta_a\delta_b. \quad (4)$$

Considering all species together leads to a generalized variogram of species composition (''generalized'' since potentially relating to several ordination methods and distance metrics):

$$\gamma_{\mathbf{G}}(h) = \sum_i \gamma_{\mathbf{G}_{ii}}(h). \quad (5)$$

Eq. 4 is a crucial point since $K(h)$ standardizes $\gamma_G(h)$ in such a manner as to equate its expected value (sill) with the total variance of the ordination method defined by weighting options. This is completely different from computing the usual experimental variogram from ordination scores, except for the special case of uniform quadrat weights (as in Wagner [2003]) where $K(h)$ is proportional to the number, $n_h$, of pairs of quadrats relating to distance class $h$. Conversely, if quadrat weights are not uniform, scaling by $K(h)$ is the only manner to ensure that, whatever the distance class, the expected value of $\gamma_G(h)$ is the total variance (''inertia'') attached to matrix $\mathbf{G_T}$ and computed from the sum of its diagonal elements. Note that such a property is not guaranteed by the manner in which Wagner (2004) defined her version (denoted as $\gamma_Q(h)$ of the CA-related variogram since the corresponding scaling remains proportional to $n_h$ despite the fact that quadrat weights are not uniform. The scaling by $K(h)$ is of particular interest if weightings of both species and quadrats are

chosen so as to relate to a diversity measurement (option IV in Table 1). In this case, the trace of $\mathbf{G_T}$ is the diversity among quadrats (Couteron and Pélissier 2004), which means that $\gamma_G(h)$ measures the average beta diversity between pairs of quadrats corresponding to distance class *h.* Equivalently,

$$\mathrm{VAR}(a, b) = \sum_i g_{ii}(a, b) \quad (6)$$

quantifies the contribution made by a given couple (*a,b*) of quadrats to beta diversity. Some classical dissimilarity indices, such as Jaccard's or Sorensen's (Legendre and Legendre 1998:256) are often used to quantify beta diversity, though these have no direct connection with either geostatiscal tools or ordination methods. Conversely, Eqs. 2 and 6 provide a family of dissimilarity indices some of which relate directly to both.

## Variograms and Cross-Variograms of Ordination Axes

Regardless of the reference ordination method chosen, a generalized variance–covariance matrix, $\mathbf{G}_h$, is computed for each distance class *h.* To ensure efficient computations by any matrix-oriented programming language, as we did with Matlab and R (Ihaka and Gentleman 1996), we introduced a matrix formulation of the method. It is based on a contiguity relationship (Thioulouse et al. 1995) consistent with the variogram, which considers two quadrats as ''neighbors at scale *h*'' if the distance between them is within the bounds of the class centered around *h* (see Appendix). Assuming that distance classes include all pairs of quadrats while being mutually exclusive, we demonstrated (see Appendix, Eq. A.13) that the matrices $\mathbf{G}_h$ sum to $\mathbf{G_T}$, whatever the initial choice of the reference ordination method by weighting options.

The eigenvectors and eigenvalues originating from the singular values decomposition (SVD) of $\mathbf{G_T}$ can be partitioned with respect to distance classes (Appendix), as a generalization of the fundamental principle introduced by Ver Hoef and Glenn-Lewin (1989). But the complete variance–covariance matrix, $\mathbf{F}_h$, between eigenvectors can also be obtained (Appendix, Eq. A.17). Considering off-diagonal elements of $\mathbf{F}_h$, namely covariances at scale *h* between eigenvectors, is a new perspective in multi-scale ordination (MSO) that can be used to investigate the potential existence of a scale-dependent covariance between distinct ordination axes. This question, though ignored by most papers devoted to MSO, is closely related to the initial concern of Noy-Meir and Anderson (1971), namely, that ordination results may substantially vary with spatial scales. This would mean, for example, that species displaying the most prominent variations of abundance may not be the same depending on the average distance between the quadrats, or that distinct species assemblages may be found for different distance classes. How can we test whether this is the case or not? One way would be to carry out an ordination for each of the $\mathbf{G}_h$ matrices and compare the results, but this is likely to be cumbersome while objective criteria for the comparison are not straightforward to define. We propose a more efficient approach by constructing cross-variograms of eigenvectors from the off-diagonal values of $\mathbf{F}_h$ matrices after appropriate scaling by $K(h)$. All these cross-variograms have an expectation of zero, since the eigenvectors of $\mathbf{G_T}$ are globally uncorrelated, but some may have significant departures from this expectation on particular scales. (Of course, only the cross-variograms for the most prominent eigenvectors are to be analyzed.) If this is the case, it is possible to know at which scales it may be worthwhile carrying out specific ordination analyses via the SVD of the corresponding $\mathbf{G}_h$ matrices.

## Taking Environmental Heterogeneity into Account

If the species-by-quadrats table is accompanied by environmental variables assessed at the quadrat scale, it is advantageous to factor out the influence of such variables prior to analyzing the residual spatial patterns of species composition. Technically, this verifies whether some basic assumptions, such as ''intrinsic'' stationarity (used to interpret the empirical variogram), or independence of residuals (assumed to fit a linear model of species–environment relationship), are met by the data (see Wagner [2004] for an extensive discussion). In terms of ecological interpretation, it is judicious to see residual spatial patterns of community composition as predominantly shaped by biotic processes, such as species dissemination or species interactions (Wagner 2004) and as potentially informative on the scale at which such processes may operate.

Any two-table ''direct'' ordination starts from the decomposition of the quadrats-by-species table, **X**, into an approximated table **A**, modeled from environmental variables by a weighted linear regression and a residual table **R**. Such a decomposition may be carried out in a manner consistent with a two-table version of any of the ordination methods mentioned in Table 1 (Sabatier et al. 1989, Pélissier et al. 2003) when the linear regression uses the quadrat weights defining the ordination. (For instance, defining table **X** along with species and quadrat weights so as to make them consistent with CA means that an ordination on **A** would be a CCA.) To study residual spatial patterns, it is possible to break down $\mathbf{G_R}$, i.e., the variance–covariance matrix computed from **R**, into additive variance–covariance matrices, $\mathbf{G_{R}}_h$, each corresponding to a certain distance class, and on which a variogram-based multi-scale analysis can be based (see Appendix).

## Brief Illustration Based on Tropical Rain Forest Data

We considered 7189 trees (diameter at breast height above 10 cm) belonging to 59 species sampled in a

lowland tropical rain forest of ~10 000 ha in French Guiana. The sampling design was based on 411 rectangular quadrats of 0.3 ha each, located at the nodes of a 400 × 500 m grid. Environmental information at the quadrat scale was expressed by a synthetic nominal variable (12 categories) primarily based on topography and soil water regime (see Couteron et al. [2003] for details). Performing correspondence analysis (CA) on the quadrats-by-species table showed two main floristic gradients corresponding to the second and third axes (CA2 and CA3). The first axis (CA1) resulted from the spurious occurrence of a scarce species (17 trees) in a particular quadrat, and this illustrates a well-known drawback of CA. Results of the nonsymmetric correspondence analysis (NSCA) were free from this problem, while the two main axes, NSCA1 and NSCA2, correlated strongly with CA3 and CA2 ($r = 0.76$ and $r = 0.84$, respectively) despite being defined from distinct species. For this data set, shifting emphasis from scarce to abundant species changed the hierarchy between the ordination axes, but the detection of two main floristic gradients proved robust with respect to species weighting. To go beyond these results established by a previous study (Couteron et al. 2003) we explicitly considered inter-quadrat distances by applying the generalized variogram-based multi-scale ordination (MSO) with CA and NSCA as reference ordination methods. First, we partitioned the total variance attached to each ordination axis (eigenvalue) among distance classes. Since diversity-related ordinations were used (Pélissier et al. 2003), it was the total among-quadrats diversity (sensu the species richness for CA or the Simpson-Gini index for NSCA) that was successively broken down with respect to main floristic gradients (eigenvalues) and distance classes.

The floristic gradient defined by CA2 and NSCA2 failed to show any obvious spatial pattern since variograms were found to waver between confidence envelopes, and this regardless of the reference ordination (Fig. 1a and b). The study of residual patterns, after factoring out the 12 environmental categories (unconstrained ordinations on the residual variance–covariance matrix, **R**) showed significant departures of the CA-based variogram for distances under 2 km. Such a change in the variogram stemming from the partialling out of the environmental variable typifies the complex interaction that can be expected between the environmental heterogeneity and spatial patterns of species assemblages. It also illustrates the advantage of studying such an interaction within a unified theoretical framework of MSO since it enabled us to express in the same unit all kinds of results derived from a particular ordination method. This renders variograms of both initial and residual patterns directly comparable—a desirable property that could not have been achieved by the computation of classical variograms from ordination scores. The other floristic gradient (defined by CA3 and NSCA1) showed a strong spatial pattern that pointed toward non-stationarity (see Wagner [2004] for a detailed definition) since both initial variograms (Fig. 1c and d) continued to rise up to 8 km without reaching a sill. The variograms of the homologous axes provided by CA and NSCA after factoring out the qualitative environmental categories appeared to be very similar. This indicated that the observed spatial patterns relating to this floristic gradient were not determined by the spatial distribution of the environmental categories.

By separately analyzing spatial patterns of distinct ordination axes we have implicitly hypothesized the absence of any scale-dependent relationship between the ordination axes or, equivalently, the stability across scales of inter-species covariances (''intrinsic'' covariances sensu Wackernagel [1998]). Such a hypothesis can be easily addressed by computing the cross-variograms between the ordination axes (from off-diagonal elements of matrices $\mathbf{F}_h$, Eq. A.18 in the Appendix). Only cross-variograms computed from the residual variance–covariance matrix (**R**) are shown (Fig. 1e and f) since homologous cross-variograms from the initial data table were very similar. No scale dependence was observed between the two ordination axes given by NSCA (Fig. 1f) and covariances between abundant species thus appeared to be stable across scales. This was not the case when the emphasis was placed on scarcer species by the use of CA since most values of the corresponding cross-variogram were outside the confidence envelopes (Fig. 1e). Indeed, by diagonalizing the pooled variance–covariance matrices for distances under 4 km vs. distances above 4 km we obtained two clearly distinct sets of species with high loadings on the ordination axes (results not shown). This result exemplified how scale dependence may be detected by analyzing cross-variograms between ordination axes, while also illustrating the influence that species weighting may have: CA results proved scale dependent though NSCA results did not.

## CONCLUDING REMARKS

In the above illustration we deliberately restricted ourselves to some particular analyses that can be obtained from the generalized variogram-based multiscale ordination (MSO), but other kinds of analyses are clearly possible. For instance, it may be of interest to compare the spatial patterns of all individual species and identify scales at which some patterns may differ from others. This can be done by analyzing, for instance by PCA, the table containing the generalized variograms of all species (prior standardization by variances of individual species so as to have all sills equal to 1 is likely to be preferable [a ''sill'' is the expected variance, i.e., the value at which a variogram is expected to level off]). It may also be of interest to study the manner in which the distribution of the eigenvalues changes with scale, by analyzing the table containing the generalized variogram of the eigenvalues. Any type of MSO can address these questions, but our unifying

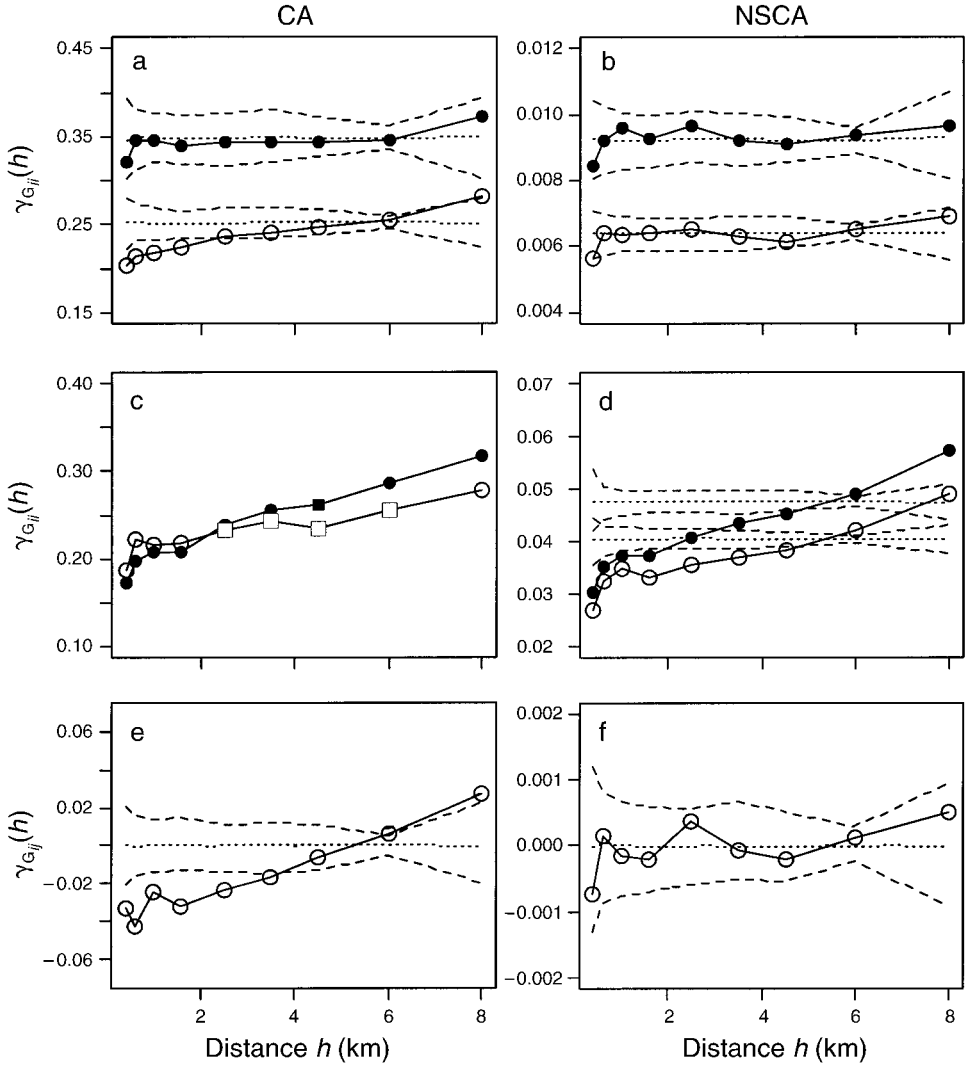


FIG. 1. Spatial patterns shown by the main ordination axes provided by the application of correspondence analysis (CA) and nonsymmetric correspondence analysis (NSCA) to vegetation data from the Counami Forest Reserve in French Guiana (7189 trees of 59 species sampled in 411 quadrats of 0.3 ha each). (a) Generalized variogram $\gamma_{G_{ii}}(h)$, for axis CA2 of the initial table (solid circles) and for the homologous axis from the residual table, after factoring out of 12 environmental categories (open circles). Dashed lines denote the 95% bilateral envelopes computed from 300 reallocations of the specific composition to geographical locations (complete randomization for variograms from the initial data table and randomization within environmental categories for variograms from the residual table). Each dotted line denotes the mean values for randomizations. (b) Same as panel (a), but for axis NSCA2. (c) Same as panel (a), but for CA3 (confidence envelopes are omitted for legibility, values within envelopes are marked by a square). (d) Same as panel (a), but for NSCA1. (e) Generalized cross-variograms, $\gamma_{G_{ii}}$, between the two main CA axes of the residual table. (f) Same as panel (e), but for NSCA.



approach also provides a choice between ordination methods while offering links with diversity measurement and apportioning. As a consequence, future users may be able to select the particular ordination method (either direct or indirect) that best suits their data and aims.

Furthermore, the presentation of the method in matrix form, with the use of contiguity matrices (Appendix), not only allows for efficient programming, but also opens up interesting methodological perspectives. Indeed, two-term local variances and covariances (TTLV/TTLC) on which the ''classical'' MSO relies (Ver Hoef and Glenn-Lewin 1989) have been also formulated using contiguity matrices (Di Bella and Jona-Lasinio 1996, Ollier et al. 2003). Such a formulation is obviously a sound basis, not only for a further generalization of the TTLV/TTLC-based MSO, as performed by us with the variogram-based MSO, but also for a thorough investigation of the respective properties of the two approaches.

ACKNOWLEDGMENTS

We are indebted to D. Chessel (University of Lyon-1) for important insights into several topics mentioned in this paper, to R. Pélissier (IRD/UMR AMAP) for his valuable comments on preliminary drafts, and to H. H. Wagner for pertinent comments on the initial version.

LITERATURE CITED

Couteron, P., and R. Pélissier. 2004. Additive apportioning of species diversity: towards more sophisticated models and analyses. Oikos **107**:215–221.

Couteron, P., R. Pélissier, D. Mapaga, J.-F. Molino, and L. Teillier. 2003. Drawing ecological insights from a management-oriented forest inventory in French Guiana. Forest Ecology and Management **172**:89–108.

Di Bella, G., and G. Jona-Lasinio. 1996. Including spatial contiguity information in the analysis of multispecific patterns. Environmental and Ecological Statistics **3**:269–280.

Dolédec, S., D. Chessel, and C. Gimaret-Carpentier. 2000. Niche separation in community analysis: a new method. Ecology **81**:2914–2927.

Escoufier, Y. 1987. The duality diagram: a means of better practical applications. Pages 139–156 *in* P. Legendre and L. Legendre, editors. Development in numerical ecology. Springer-Verlag, Berlin, Germany.

Gimaret-Carpentier, C., D. Chessel, and J.-P. Pascal. 1998. Non-symmetric correspondence analysis: an alternative for species occurrences data. Plant Ecology **138**:97–112.

Greig-Smith, P. 1983. Quantitative plant ecology. Third edition. Blackwell Science Publishing, Oxford, UK.

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics **5**:299–314.

Legendre, P., and L. Legendre. 1998. Numerical ecology. Elsevier, Amsterdam, The Netherlands.

Levin, S. 1992. The problem of pattern and scale in ecology. Ecology **73**:1943–1967.

Noy-Meir, I., and D. Anderson. 1971. Multiple pattern analysis or multiscale ordination: towards a vegetation hologram. Pages 207–232 *in* G. P. Patil, E.C. Pielou, and E.W. Water, editors. Statistical ecology: populations, ecosystems, and systems analysis. Pennsylvania State University Press, University Park, Pennsylvania, USA.

Ollier, S., D. Chessel, P. Couteron, R. Pélissier, and J. Thioulouse. 2003. Comparing and classifying one-dimensional spatial patterns: an application to laser altimeter profiles. Remote Sensing of Environment **85**:453–462.

Pélissier, R., P. Couteron, S. Dray, and D. Sabatier. 2003. Consistency between ordination techniques and diversity measurements: two strategies for species occurrence data. Ecology **84**:242–251.

Sabatier, R., J.-D. Lebreton, and D. Chessel. 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. Pages 341–352 *in* R. Coppi and S. Bolasco, editors. Multiway data analysis. Elsevier Science Publishers, Amsterdam, The Netherlands.

ter Braak, C. J. F. 1983. Principal components biplots and alpha and beta diversity. Ecology **64**:454–462.

Thioulouse, J., D. Chessel, and S. Champély. 1995. Multivariate analysis of spatial patterns: a unified approach to local and global structures. Environmental and Ecological Statistics **2**:1–14.

Ver Hoef, J. M., and D. C. Glenn-Lewin. 1989. Multiscale ordination: a method for detecting pattern at several scales. Vegetatio **82**:59–67.

Wackernagel, H. 1998. Multivariate geostatistics. Second completely revised edition. Springer-Verlag, Berlin, Germany.

Wagner, H. H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology **84**:1045–1057.

Wagner, H. H. 2004. Direct multi-scale ordination with canonical correspondence analysis. Ecology **85**:342–351.

## APPENDIX

A matrix-algebraic presentation of the concepts and computations is available in ESA's Electronic Data Archive: *Ecological Archives* E086-042-A1.

*Ecological Archives* **E086-042-A1**

http://www.esapubs.org/archive/ecol/E086/042/default.htm

**Pierre Couteron, and Sébastien Ollier. 2005. A generalized, variogram-based framework for multi-scale ordination. *Ecology* 86:828-834.**

## Appendix A. Matrix-algebraic presentation of the concepts and computations..

### General denotation

Let **X** be a table expressing a measure of the abundance $x_{ai}$ of $S$ species (columns) within $Q$ quadrats (rows). $\mathbf{x}_i$ and $\mathbf{x}_j$ are two columns of table **X**, relating to species $i$ and $j$, respectively. Let **D** be a matrix containing quadrat weights ($\delta_a$, $\sum_a \delta_a = 1$) on its main diagonal and zeros for all off-diagonal values, and let **W** be a $S$ by $S$ matrix containing the square root of species weights ($\sqrt{w_i}$) on its main diagonal and zeros outside. (In the main paper, Table 1 gives some options for abundance re-scaling and for quadrat and species weighting.)

### Contiguity relationships

Let $\mathbf{L}_h$ be a $Q$ by $Q$ matrix expressing a contiguity relationship (*sensu* Lebart 1969) between the quadrats. For our variogram-based approach, we consider quadrats $a$ and $b$ as neighbors if the distance between the two is within the bounds of the distance class centered around $h$:

$\mathbf{L}_h(a,b)=1$ if $h_{a,b} \approx h$ and $\mathbf{L}_h(a,b)=0$ otherwise. (A.1)

To introduce quadrat weights into the analysis, we define the matrix $\mathbf{M}_h$ and the vector $\mathbf{E}_h$ such that:

$\mathbf{M}_h = \mathbf{D}\mathbf{L}_h\mathbf{D}$ and $\mathbf{E}_h = \mathbf{M}_h\mathbf{1}_Q$ (A.2)

where $\mathbf{1}_Q$ is the vector containing $Q$ values equal to 1.

$\mathbf{M}_h$ contains, for each pair $(a,b)$ of neighboring quadrats at "scale" $h$, the product $\delta_a\ \delta_b$ of their weights. $\mathbf{E}_h$ features, for each quadrat $a$, the sum of the weights of its neighbors multiplied by $\delta_a$. Let $\mathbf{N}_h$ be the $Q$ by $Q$ matrix with $\mathbf{E}_h$ on its main diagonal and zeros elsewhere.

We shall assume that distance classes include all pairs of quadrats while being mutually exclusive. In such a case, the two following matrices:

$$\mathbf{M_T} = \sum_h \mathbf{M}_h \quad \text{and} \quad \mathbf{N_T} = \sum_h \mathbf{N}_h \tag{A.2b}$$

are such that $\mathbf{M_T}$ is a $Q$ by $Q$ matrix that containing zeros on the diagonal while all values off the diagonal are equal to $\delta_a\ \delta_b$; $\mathbf{N_T}$ is a $Q$ by $Q$ matrix that containing $(1\text{-}\delta_a)\delta_a$ values on the diagonal and zeros elsewhere. With $\mathbf{M_T}$ and $\mathbf{N_T}$ it is as if each quadrat has all other quadrats as neighbors. Denoting $\mathbf{I}_Q$ the $Q$ by $Q$ diagonal identity matrix, we can also write:

$$\mathbf{N_T} = \mathbf{D(I}_Q - \mathbf{D)} \quad \text{and} \quad \mathbf{M_T} = \mathbf{D(1}_Q\mathbf{1}_Q{}^t - \mathbf{I}_Q)\mathbf{D} \tag{A.3}$$

(where the exponent $'\,t\,'$ is the matrix transpose). Thus:

$$\mathbf{N_T} - \mathbf{M_T} = \mathbf{D} - \mathbf{D1}_Q\mathbf{1}_Q{}^t\mathbf{D} \tag{A.4}$$

**Equivalent expressions of the generalized variance-covariance matrix**

Let $\mathbf{G_T}$ be the generalized variance-covariance matrix, irrespective of distance classes, that can be directly computed from table $\mathbf{X}$ using weighting options for rows and columns defined by matrices $\mathbf{D}$ and $\mathbf{W}$, respectively. $\mathbf{G_T}$ contains, for each species couple $(i,j)$, the generalized covariances, $g_{ij}$ as defined by Eq. 1 and Eq. 2 in the main paper:

$$g_{ij} = \sum_{a,b} g_{ij}(a,b) \tag{A.5}$$

Usual algebraic manipulations allow us to re-write Eq. 1 and Eq. A.5 as:

$$g_{ij} = \sqrt{w_i w_j}\left(\sum_a^Q \delta_a x_{ai} x_{aj} - \bar{x}_i \bar{x}_j\right) \quad \text{(A.6)}$$

where $\bar{x}_i$ and $\bar{x}_j$ are the **D**-weighted means of $\mathbf{x}_i$ and $\mathbf{x}_j$, respectively.

($\bar{x}_i = \sum_a^Q \delta_a x_{ia}$ or, equivalently, $\bar{x}_i = \mathbf{X}_i{}^t \mathbf{D1}_Q$)

The matrix expression of $g_{ij}$ is thus:

$$g_{ij} = \mathbf{W}(\mathbf{x}_i{}^t \mathbf{D} \mathbf{x}_j - \mathbf{x}_i{}^t \mathbf{D1}_Q \mathbf{x}_j{}^t \mathbf{D1}_Q)\mathbf{W} \quad \text{(A.7)}$$

which generalizes into:

$$\mathbf{G_T} = \mathbf{W}(\mathbf{X}^t \mathbf{D} \mathbf{X} - \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}})\mathbf{W} \quad \text{(A.8)}$$

where $\overline{\mathbf{X}} = \mathbf{1}_Q \mathbf{1}_Q{}^t \mathbf{D} \mathbf{X}$ and where $\overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}} = [\bar{x}_i \bar{x}_j]$ (A.9)

Note that we may also write:

$$\mathbf{G_T} = \mathbf{W}(\mathbf{X} - \overline{\mathbf{X}})^t \mathbf{D}(\mathbf{X} - \overline{\mathbf{X}})\mathbf{W} \quad \text{(A.10)}$$

On the other hand**,** it is important to note that $\mathbf{G_T}$ can be directly computed as:

$$\mathbf{G_T} = \mathbf{W}\mathbf{X}^t(\mathbf{N_T} - \mathbf{M_T})\mathbf{X}\mathbf{W} \quad \text{(A.11)}$$

<u>Proof of Eq. A.11:</u>

$\mathbf{X}^t(\mathbf{N_T} - \mathbf{M_T})\mathbf{X} = \mathbf{X}^t(\mathbf{D} - \mathbf{D1}_Q \mathbf{1}_Q{}^t \mathbf{D})\mathbf{X}$ (using Eq. A.4)

$\mathbf{X}^t(\mathbf{D} - \mathbf{D1}_Q \mathbf{1}_Q{}^t \mathbf{D})\mathbf{X} = \mathbf{X}^t \mathbf{D} \mathbf{X} - \mathbf{X}^t \mathbf{D} \overline{\mathbf{X}}$ (using Eq. A.9)

Noting that $\mathbf{X}^t \mathbf{D} \overline{\mathbf{X}} = \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}}$, allows us to write:

$$\mathbf{X}^t(\mathbf{N_T} - \mathbf{M_T})\mathbf{X} = \mathbf{X}^t \mathbf{D} \mathbf{X} - \overline{\mathbf{X}}^t \mathbf{D} \overline{\mathbf{X}} \quad \text{(A12)}$$

**Partition of the generalized variance-covariance matrix among distance classes**

The very definition of matrices $\mathbf{N_T}$ and $\mathbf{M_T}$ (Eq. A.2b), along with Eq. A.11, enables partition of $\mathbf{G_T}$ into strictly additive components, $\mathbf{G}_h$, that relate each to a distance class:

$$\mathbf{G_T} = \sum_h \mathbf{G}_h = \sum_h \mathbf{W}\mathbf{X}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{X}\mathbf{W} \qquad \text{(A.13)}$$

$\mathbf{G}_h$ is the generalized variance-covariance matrix defined for the distance class *h* by the neighboring relationship expressed by the matrices $\mathbf{N}_h$ and $\mathbf{M}_h$. $\mathbf{G}_h$ translates easily into generalization of Wagner's variogram matrix (2003) by a division of all its values by

$$K(h) = \sum_{a,b \mid h_{ab} \approx h} \delta_a \delta_b \qquad \text{or} \quad K(h) = \mathbf{1}_Q{}^t \mathbf{M}_h \mathbf{1}_Q \qquad \text{(A.14)}$$

Equations A.2, A.13 and A.14 are used for easy programming of the method as well as efficient computations via any matrix-oriented programming environment, as we did with Matlab® and R (Ihaka and Gentleman 1996): see the freely available library "msov" on http://pbil.univ-lyon1.fr/CRAN/.)

For a particular species couple *i* and *j* we obtain:

$$g_{ij}(h) = \mathbf{W}\mathbf{x}_i{}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{x}_j\mathbf{W} \qquad \text{(A.15)}$$

Dividing by the scaling factor *K(h)* gives the value at "scale" *h* of the generalized version of either cross-variogram ($i \neq j$) or variogram ($i = j$) :

$$\gamma_{Gij}(h) = \frac{1}{K(h)} g_{ij}(h) \qquad \text{(A.16)}$$

**Multi-scale ordination**

All the ordination methods mentioned in Table 1 of the main paper are based on the singular values decomposition (svd) of the appropriate version of $\mathbf{G_T}$ to compute eigenvectors, $\mathbf{u}_f$, and associated eigenvalues, $\lambda_f$. Let $\mathbf{U}_f$ be the matrix having all the eigenvectors $\mathbf{u}_f$ as columns and let $\mathbf{\Lambda}$ be the diagonal matrix having the eigenvalues $\lambda_f$ on its diagonal. Both eigenvectors and eigenvalues of $\mathbf{G_T}$ can be partitioned by distance classes:

$$\mathbf{F}_h = \mathbf{U}_f{}^t \mathbf{G}_h \mathbf{U}_f \quad \text{and} \quad \lambda_f(h) = \mathbf{u}_f{}^t \mathbf{G}_h \mathbf{u}_f \qquad \text{(A.17)}$$

$\mathbf{F}_h$ is the variance-covariance matrix of the eigenvectors at scale *h*. Scale-dependent variogram/cross-variogram matrices of the eigenvectors are deduced by the appropriate scaling (Eq. A.16). Note also that:

$$\sum_h \mathbf{F}_h = \mathbf{U}_f{}^t \left( \sum_h \mathbf{G}_h \right) \mathbf{U}_f = \mathbf{U}_f{}^t \mathbf{G}_T \mathbf{U}_f = \mathbf{\Lambda} \qquad \text{(A.18)}$$

**Taking environmental heterogeneity into account**

Let us now suppose that a table, **Z**, containing assessments of *P* environmental variables for the *Q* quadrats, is available in addition to table of species composition. It is well established that the centered by columns table, $\mathbf{X}_C$ , may be partitioned into an approximated table,

$$\mathbf{A}_C = \mathbf{Z}(\mathbf{Z}^t\mathbf{D}\mathbf{Z})^{-1}\mathbf{Z}^t\mathbf{D}\mathbf{X}_C \qquad \text{(A.19)}$$

and a residual table, $\mathbf{R}_C = \mathbf{X}_C - \mathbf{A}\mathbf{c}$ (Sabatier et al. 1989).

In the same manner, it may also have a direct decomposition of the initial table **X**:

$$\mathbf{A} = \mathbf{Z}(\mathbf{Z}^t(\mathbf{N}_T - \mathbf{M}_T)\mathbf{Z})^{-1}\mathbf{Z}^t(\mathbf{N}_T - \mathbf{M}_T)\mathbf{X} \text{ and } \mathbf{R} = \mathbf{X} - \mathbf{A} \qquad \text{(A.20)}$$

After factoring out the environmental variables, residual spatial patterns may be studied by the multi-scale analysis of spatial covariances derived from table **R** or $\mathbf{R}_C$ . The total residual variance-covariance matrix, $\mathbf{G}_{RT}$, is computed as:

$$\mathbf{G}_{RT} = \mathbf{W}\mathbf{R}^t(\mathbf{N}_T - \mathbf{M}_T)\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_C{}^t(\mathbf{N}_T - \mathbf{M}_T)\mathbf{R}_C\mathbf{W} \qquad \text{(A.21)}$$

and is broken down with respect to distance classes:

$$\mathbf{G}_{Rh} = \mathbf{W}\mathbf{R}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}\mathbf{W} = \mathbf{W}\mathbf{R}_C{}^t(\mathbf{N}_h - \mathbf{M}_h)\mathbf{R}_C\mathbf{W} \qquad \text{(A.22)}$$

The additive partitioning of $\mathbf{G}_{RT}$ with respect to distance classes thus enables an investigation of the residual spatial patterns by a multi-scale ordination scheme analogous to that defined by Eq. A.17 and Eq. A.18.

**Literature cited in this appendix**

Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. Journal of Computational and Graphical Statistics **5**:299-314.

Lebart, L. 1969. Analyse statistique de la contiguïté. Publications de l'Institut de Statistiques de l'Université de Paris **28**: 81-112.

Sabatier, R., J.-D. Lebreton, and D. Chessel. 1989. Principal component analysis with instrumental variables as a tool for modelling composition data. Pages 341–352 *in* R. Coppi and S. Bolasco, Editors. Multiway data analysis, Elsevier Science Publishers, Amsterdam, The Netherlands.

Wagner, H.H. 2003. Spatial covariance in plant communities: integrating ordination, geostatistics, and variance testing. Ecology **84**: 1045-1057.

*Journal of Applied Ecology* 2005

METHODOLOGICAL INSIGHTS

# Predicting tropical forest stand structure parameters from Fourier transform of very high-resolution remotely sensed canopy images

PIERRE COUTERON,*† RAPHAEL PELISSIER,* ERIC A. NICOLINI* and DOMINIQUE PAGET‡

**UMR botAnique et bioinformAtique de l'Architecture des Plantes (AMAP), TA40/PS2, 34398 Montpellier Cedex 05, France; †Département d'Ecologie, Institut Français de Pondichéry, 11 St Louis Street, 605001 Pondicherry, India; and ‡151, allée des Narcisses, 74120 Megève, France*

**Summary**

**1.** Predicting stand structure parameters for tropical forests from remotely sensed data has numerous important applications, such as estimating above-ground biomass and carbon stocks and providing spatial information for forest mapping and management planning, as well as detecting potential ecological determinants of plant species distributions. As an alternative to direct measurement of physical attributes of the vegetation and individual tree crown delineation, we present a powerful holistic approach using an index of canopy texture that can be extracted from either digitized air photographs or satellite images by means of two-dimensional spectral analysis by Fourier transform.
**2.** We defined an index of canopy texture from the ordination of the Fourier spectra computed for 3545 1-ha square images of an undisturbed tropical rain forest in French Guiana. This index expressed a gradient of coarseness vs. fineness resulting from the relative importance of small, medium and large spatial frequencies in the Fourier spectra.
**3.** Based on 12 1-ha control plots, the canopy texture index showed highly significant correlations with tree density ($R^2$ = 0·80), diameter of the tree of mean basal area ($R^2$ = 0·71), distribution of trees into d.b.h. classes ($R^2$ = 0·64) and mean canopy height ($R^2$ = 0·57), which allowed us to produce reasonable predictive maps of stand structure parameters from digital aerial photographs.
**4.** *Synthesis and applications*. Two-dimensional Fourier analysis is a powerful method for obtaining quantitative characterization of canopy texture, with good predictive ability on stand structure parameters. Forest departments should use routine forest inventory operations to set up and feed regional databases, featuring both tree diameter figures and digital canopy images, with the ultimate aims of calibrating robust regression relationships and deriving predictive maps of stand structure parameters over large areas of tropical forests. Such maps would be particularly useful for forest classification and to guide field assessment of tropical forest resources and biodiversity.

*Key-words*: canopy texture, French Guiana, submetric images, two-dimensional spectral analysis

*Journal of Applied Ecology* (2005)
doi: 10.1111/j.1365-2664.2005.01097.x

## Introduction

Retrieving tropical forest stand structure parameters from remotely sensed data is of primary importance for estimating global carbon stocks in above-ground biomass (Houghton *et al.* 2001; Grace 2004) as well as for obtaining large-scale information required by regional biodiversity



Correspondence: Pierre Couteron, Département d'Ecologie, Institut Français de Pondichéry, 11 St Louis Street, 605001 Pondicherry, India (e-mail pierre.couteron@ifpindia.org).

studies and forest type classification and mapping (Tuomisto *et al*. 1995). Indeed, stand structure parameters such as tree density, basal area and canopy height not only allow predictions of forest biomass (Chave *et al*. 2003) but may also provide spatial information on potential determinants of plant species' distributions (Couteron *et al*. 2003), such as the gap-phase regeneration stages (Riéra, Pélissier & Houllier 1998) and the variations in substratum conditions and soil fertility (Ashton & Hall 1992; Paget 1999).

Thanks to the various kinds of satellite and airborne remotely sensed data that are available, for example Landsat™ (Lu *et al*. 2004), SPOT (De Wasseige & Defourny 2002), Laser Vegetation Imaging Sensor (Drake *et al*. 2002) and Synthetic Aperture Radar JERS-1 (Santos *et al*. 2002), predicting stand structure parameters over large areas with a spatial resolution of a dozen to several dozen metres may be achievable. It is, however, surprising that fine spatial resolution techniques capable of detecting small-scale variation in physical signals have only been of limited applicability for the study of tropical forest structure, and therefore of limited use for both ecological research and operational management/conservation (Read 2003). In fact, small-footprint laser altimeters have proven to be of limited efficiency in dense tropical forests because of inconsistent ground returns (Nelson, Oderwald & Gregoire 1997; Dubayah & Drake 2000; Drake *et al*. 2003; but see Clark, Clark & Roberts 2004). Recent metric and submetric resolution optical data, such as IKONOS and QuickBird panchromatic images, have mostly been used in tropical forests for visual tree crown delineation (Asner *et al*. 2002; Read *et al*. 2003; Clark *et al*. 2004a, 2004b), although Asner & Warner (2003) used an automatic quantification of shadowing while Hurtt *et al*. (2003) reported preliminary investigations on automatic crown delineation. However, tropical foresters have more than half a century of tradition of delineating and mapping forest types by visual interpretation of aerial photographs (see Table 1 in Polidori *et al*. 2004), a practice that could easily be extended to modern very-high resolution (VHR) satellite imagery. Furthermore, aerial photographs, when digitized at metric spatial resolution, have enabled numerical extraction of textural information in temperate forest (Sommerfeld, Lundquist & Smith 2000) and semi-arid vegetation (Couteron & Lejeune 2001; Couteron 2002). Such techniques could thus also be applied to satellite images. The present study aimed to demonstrate that texture analysis of digitized aerial photographs by two-dimensional Fourier transform (Mugglestone & Renshaw 1998), as adapted by Couteron (2002), is a valuable approach to characterizing tropical rain forest canopies and obtaining reasonable predictions of tropical forests stand structure parameters.



## Materials and methods

### STUDY SITE

The study was conducted as part of the DIME 'Diversité Multi-échelles (multiscale diversity)' project, within an area of 65 km$^2$ of undisturbed lowland evergreen rain forest, located at about 10 km east of the Petit-Saut reservoir dam in French Guiana (5°00′N, 52°55′W). A small river (Crique Plomb) running along a geological boundary divides the area into a hilly landscape (hill-tops from 80 to 210 m a.s.l.) lying on sedimentary rocks (pelite) to the north, and the Montagne Plomb volcano-sedimentary massif that reaches its highest point at 332 m a.s.l. to the south (Delor *et al*. 2003). The varied geomorphology of the site results in contrasted textural aspects of the forest canopy, which are apparent on aerial photographs.

### DIGITIZED AERIAL PHOTOGRAPHS

A set of black and white aerial photographs at 1 : 25000 scale (numbers 379–383, 397–401 and 456–460 of coverage 1992-GUF-91/250) were obtained from the Institut Géographique National (IGN, Saint-Mandé, France). Each image was digitized into 256 grey levels with a Nikon® Scantouch 210 flatbed scanner (Nikon Corp., Tokyo, Japan) using a resolution of 600 dots per inch (d.p.i.), corresponding to a pixel size of 1 m in the field. On images, the fully sunlit crowns of canopy trees appeared in white or light grey, while shadowed inter-crown gaps were dark-grey or black. A monotonic relationship between grey-level scale and canopy height can thus be assumed as long as there is no substantial relief-induced shadowing. Contact numbers 399 and 401 were separately submitted for spectral analysis (see below) without any prior correction. In each photograph we only analysed the central part, i.e. a $4{\cdot}8 \times 4$-km block (1920 ha) in which convex deformation as a result of the camera lens is of minor importance (Avery & Berlin 1992). For field navigation, a copy of the digitized images was assembled, georeferenced and superimposed on a digitized 1 : 50 000 topographical map (sheets Kourou S-O NB-22-VIII-1a and Haut Kourou NO NB-22-II-3c; IGN) using ArcGIS (ArcGIS™ Version 8.3; ESRI Inc., Redlands, CA).

### SPECTRAL ANALYSIS BY FOURIER TRANSFORM

Only broad outlines of two-dimensional spectral analysis by Fourier transform are provided here, as detailed presentations (Ripley 1981) as well as applications to digital images (Mugglestone & Renshaw 1998; Couteron 2002) are already available. For simplicity, only square images were considered although the method also applies to rectangular images. A digital image is defined here as an $n$ by $n$ array of grey-scale values in the range 0–255 expressing the panchromatic radiance of each pixel. Spectral analysis aims at modelling such data as a weighted sum of cosine and sine waveforms of varying travelling direction and spatial frequency. For a particular geographical direction, the wavenumber, $p$, quantifies spatial frequency and corresponds to the number of times a waveform repeats itself within the image. The

two-dimensional Fourier periodogram features the decomposition of the total image variance according to all possible integer pairs ($p$, $q$), of wavenumbers along the two Cartesian geographical directions (with $1 \leq p \leq n/2$ and $1 \leq q \leq n/2$). When expressed in polar form (Mugglestone & Renshaw 1998), periodogram values, $Ir\theta$, are portions of image variance accounted for by a waveform having spatial frequency $r$ and travelling direction $\theta$, with $r = \sqrt{p^2 + q^2}$ and $\theta = \tan^{-1}(p/q)$.

For each spatial frequency, summing values on all possible travelling directions yields an azimuthally cumulated 'radial' spectrum, $I(r)$, that provides a convenient way to quantify coarseness-related textural properties by studying the decomposition of variance among spatial frequencies. Images with a coarse texture will yield a radial spectrum that is skewed towards small wavenumbers, whilst fine-textured images are expected to produce more balanced spectra (for a schematic illustration see fig. 3 in Couteron 2002). An image in which each pixel takes a random value independent of the value taken by any other pixel will have the finest possible texture and a virtually flat spectrum.

### ORDINATION OF RADIAL SPECTRA

We carried out a systematic textural analysis that started by partitioning the canopy photographs into square windows of 1 ha, at which scale radial spectra were computed. A general table was built in which each row was the radial spectrum of a given window, while each column contained $I(r)$ values, i.e. the portions of the grey-level variance explained by a given spatial frequency or wavenumber, $r$. This table of spectra was submitted to a principal component analysis (PCA; Manly 1994), which means that windows were considered as statistical observations characterized by their spectral profiles, i.e. the way in which the grey-level variance was broken down in relation to spatial frequencies. Conversely, spatial frequencies were seen as quantitative variables that were to be linearly combined to yield principal components. We used standardized PCA, so that principal components were defined from the eigenvector analysis of the correlation matrix between spatial frequencies. Note that our method of textural analysis, which fully relies on Fourier spectra, is totally distinct from the approach of Sommerfeld, Lundquist & Smith (2000), who only use the Fourier transform to filter an image before identifying trees by applying a binary threshold.

### GROUND CONTROL DATA

Twelve 1-ha control plots were laid out to serve as ground-truth for the textural analysis. We used a Magellan SporTrak Color hand-held GPS unit (Magellan Systems Corp., San Dimas, CA) to locate the plots in the field, link them to the aerial photographs, and extract the corresponding digital images using GIS. Each plot was a $100 \times 100$-m square in which we measured diameter at breast height (d.b.h.), or above the buttresses if present, of all the trees greater than 10 cm d.b.h. We also sampled 49 canopy trees (i.e. trees with estimated Dawkin's (1958) crown index of 4 or 5) for total height measurement using a calibrated optical telemeter (Birnbaum 2001). These canopy trees were selected systematically and regardless of their size as the closest to the nodes of a $15 \times 15$-m grid covering the entire plot. We computed from these data five simple structural characteristics of the forest stands (Table 1): density ($D$), basal area ($G$), diameter of the tree of mean basal area ($Dg$), mean canopy height ($Hm$) and the standard deviation of mean canopy height as an index of canopy roughness ($R$). All measured trees were also classified into eight d.b.h. classes of 10 cm width, plus an additional class in which all the trees ≥ 90 cm d.b.h. were pooled. The table crossing plots with d.b.h. classes was submitted to correspondence analysis (CA; Manly 1994) to summarize diameter distributions. The first CA axis that was prominent (73% of the variance of the table) ranked d.b.h. classes in their natural order and expressed a gradient from the smallest class (10–20 cm) to the largest (≥ 50 cm). Plot scores on this axis were

**Table 1.** Stand structure parameters measured for 12 1-ha ground-truth plots in the study area of the DIME project near the Petit-Saut reservoir dam, French Guiana. *Dg*, the diameter of the tree of mean basal area; *R*, standard deviation of *Hm*; *CA1*, plot scores along axis 1 of the correspondence analysis carried out on diameter distributions (see text)

| Plots | Density *D* (trees ha$^{-1}$) | Basal area *G* (m$^2$ ha$^{-1}$) | Mean tree *Dg* diameter (cm) | Mean canopy *Hm* tree height (m) | Canopy *R* roughness | Diametric *CA1* structure |
|---|---|---|---|---|---|---|
| 1 | 455 | 37·45 | 32·3 | 30·4 | 8·28 | 0·236 |
| 2 | 458 | 28·41 | 28·1 | 27·5 | 7·45 | 0·020 |
| 3 | 481 | 38·57 | 31·9 | 26·0 | 7·37 | 0·134 |
| 4 | 746 | 37·08 | 25·2 | 26·8 | 6·36 | −0·098 |
| 5 | 802 | 39·44 | 25·0 | 26·3 | 5·02 | −0·107 |
| 6 | 548 | 39·26 | 30·2 | 28·1 | 6·48 | 0·156 |
| 7 | 989 | 32·96 | 20·6 | 21·1 | 5·28 | −0·284 |
| 8 | 461 | 42·38 | 34·2 | 29·9 | 8·44 | 0·337 |
| 9 | 461 | 36·55 | 31·8 | 27·5 | 9·25 | 0·184 |
| 10 | 547 | 37·16 | 33·2 | 29·0 | 5·57 | 0·058 |
| 11 | 476 | 35·14 | 30·6 | 29·5 | 6·66 | 0·168 |
| 12 | 861 | 33·03 | 22·1 | 24·4 | 4·56 | −0·208 |

then used as a latent variable summarizing the diameter distributions (*CA1* in Table 1). Subsequent CA axes proved difficult to interpret and were not considered further.

## Results

### TEXTURAL ANALYSIS

A preliminary PCA analysis dealing with all the 3840 windows of 1 ha highlighted that a limited number of windows (about 7·5% of the total) was strongly differentiated from the remainder. These windows, which displayed large features (at the scale of 1 ha) such as forest roads or stream valleys and so could not be adequately assessed for forest canopy, were excluded from subsequent analyses. A second PCA was run on the 3545 remaining windows, yielding a first factorial plane (Fig. 1, top) that reflected a progressive transition from spectra dominated by the first wavenumber ($r = 1$, positive part of PCA axis 2) to spectra characterized by the relative importance of small ($r = 2$–$3$, negative part of axis 1), intermediate ($r = 4$–$8$, negative part of axis 2) and large wavenumbers ($r > 8$, positive part of axis 1). Accordingly, we found windows displaying coarse-grained canopy aspects (Fig. 1b) on the negative extremity of axis 1, while fine-grained canopy textures were found on the positive extremity (Fig. 1f). Windows of intermediate texture were located all around the axes origin, although the best illustrations of these textural types (Fig. 1c,d) were found around the negative extremity of axis 2. The textural properties of the canopy frequently resulted from size and spatial distributions of gaps and aggregates of crowns rather than from size distribution of individual crowns. This was particularly the case for the coarser textural types (Fig. 1b).

All windows having high positive scores along axis 2 displayed heterogeneity mostly as a result of the influence of relief on sun illumination; for example, on Fig. 1a–g about one half of the window appeared sunlit and the other half in shadow. Most of those windows encompassed a prominent relief feature such as a ridge or a pronounced valley. Some of those windows displayed either a coarse-grained or a complex canopy texture (e.g. Fig. 1a) while others had a fine-grained canopy (Fig. 1g). In this latter case, the spectrum was characterized by the simultaneous dominance of first and large wavenumbers.

### RELATIONSHIPS BETWEEN CANOPY TEXTURE AND STAND STRUCTURE

We used coordinates of the plots on the textural ordination axis 1 as an index of canopy texture (PCA1), while explanatory power regarding stand structure parameters of Table 1 was tested using ordinary linear regressions (Fig. 2). It is noteworthy that while PCA1 was a good predictor of the mean stem density ($R^2 = 0{\cdot}80$, $P < 0{\cdot}0001$), the regression against mean basal area displayed almost no slope. This was because of the limited variability of basal area across the reference plots. Good predictions were also obtained for stand structure parameters such as the diameter of the tree of mean basal area ($R^2 = 0{\cdot}71$, $P < 0{\cdot}001$) and plot coordinates along axis 1 of the correspondence analysis summarizing the size (d.b.h.) distribution of trees ($R^2 = 0{\cdot}64$, $P < 0{\cdot}01$). The relationship with mean canopy height was weaker but statistically significant ($R^2 = 0{\cdot}57$, $P < 0{\cdot}01$), as with standard deviation of canopy heights ($R^2 = 0{\cdot}55$, $P < 0{\cdot}01$), a proxy for stand canopy roughness.

### PREDICTIVE MAPS OF STAND STRUCTURE PARAMETERS

To complement the textural analysis, we submitted the table of the radial spectra to *k*-means clustering (Manly 1994) using the Euclidean distance after prior standardization by column standard deviations in order to remain consistent with results of standardized PCA. In this example, clustering using five classes yielded the most meaningful results. Textural classes T2 (coarse-grained canopy) to T5 (fine-grained) were ordinated along PCA axis 1 of the textural analysis (Fig. 1). Class T1 differentiated from other coarse-grained classes along PCA axis 2, on the basis of image macro-heterogeneity as a result of relief-induced illumination discrepancy. Class T5 encompassed all fine-grained canopy windows whatever the level of illumination heterogeneity.

Although allocating canopy windows to an arbitrary number of classes provides less objective information than ordination scores, it is very convenient for mapping. For instance, Fig. 3 presents a map of the canopy texture based on the five classes for contact number 399 while Table 2 shows predictions of the most significant mean stand structure parameters of these classes, as inferred from the above regressions (Sokal & Rohlf 1995). Figure 3 reveals a clear spatial partition between:

**Table 2.** Predictions of mean stand structure parameters (with standard error) for the five classes of canopy texture (see Fig. 2 and text). Study site of the DIME project, near the Petit-Saut reservoir dam, French Guiana

| Textural classes | Predicted density $\hat{D}$ (trees ha$^{-1}$) | Predicted mean tree diameter $\hat{Dg}$ (cm) | Predicted mean canopy tree height, $\hat{Hm}$ (m) |
|---|---|---|---|
| T1 | 453·2 (35·2) | 32·2 (1·0) | 29·0 (0·7) |
| T2 | 382·0 (43·7) | 33·8 (1·3) | 29·8 (0·9) |
| T3 | 482·2 (32·3) | 31·6 (0·9) | 28·7 (0·7) |
| T4 | 583·2 (25·9) | 29·3 (0·7) | 27·5 (0·5) |
| T5 | 738·2 (32·9) | 25·8 (0·9) | 25·7 (0·7) |

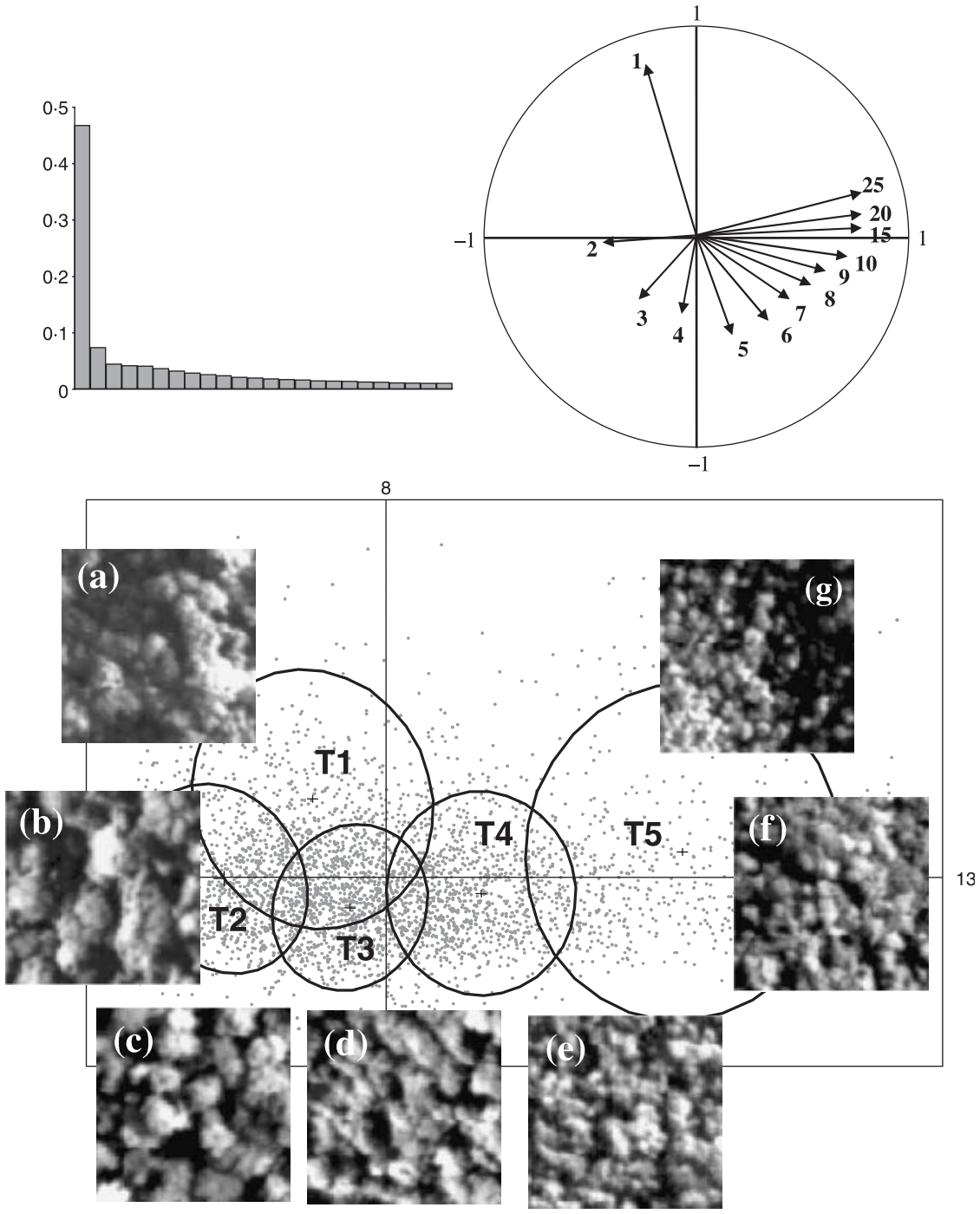


**Fig. 1.** Results of the multidimensional comparison of spectral profiles for 3545 forest canopy windows of 1 ha in the study area of the DIME project near the Petit-Saut reservoir dam, French Guiana. The five textural classes (T1–T5) identified by *k*-means clustering (see text) are plotted against the two main axes yielded by the PCA of the spectra table. The envelopes of classes delineate twice the standard deviation of PCA scores of the constituting windows. Windows (a) to (g) have been automatically selected as the most illustrative (largest distance from axes origin) with regards to successive angular directions in the PCA plane and, thus, to the relative importance of the corresponding ranges of wavenumbers. Top-left inset, distribution of relative eigenvalues. Top-right inset, correlations of wavenumbers (some of the wavenumbers above 10 are omitted for legibility) with PCA axes 1 and 2; each radial direction in the PCA plane corresponds to the relative importance in the spectrum of particular ranges of wavenumbers.



(i) coarse- to intermediate-grained canopies (T2–3; red and green squares), which were observed on volcano sedimentary materials on the slopes of Montagne Plomb in the southern part of the study site, corresponding to tall stands ($\hat{Hm}$ *c.* 29–30 m) with a low density ($\hat{D}$ *c.* 400–500 trees ha$^{-1}$) of large trees ($\hat{Dg}$ *c.* 32–34 cm d.b.h.); and (ii) intermediate- to fine-grained canopies (T4–5; yellow and light-blue squares) on pelitic formations in the northern hilly part of the site, corresponding to low stands ($\hat{Hm}$ *c.* 26–28 m) with a high density ($\hat{D}$ *c.* 600–700

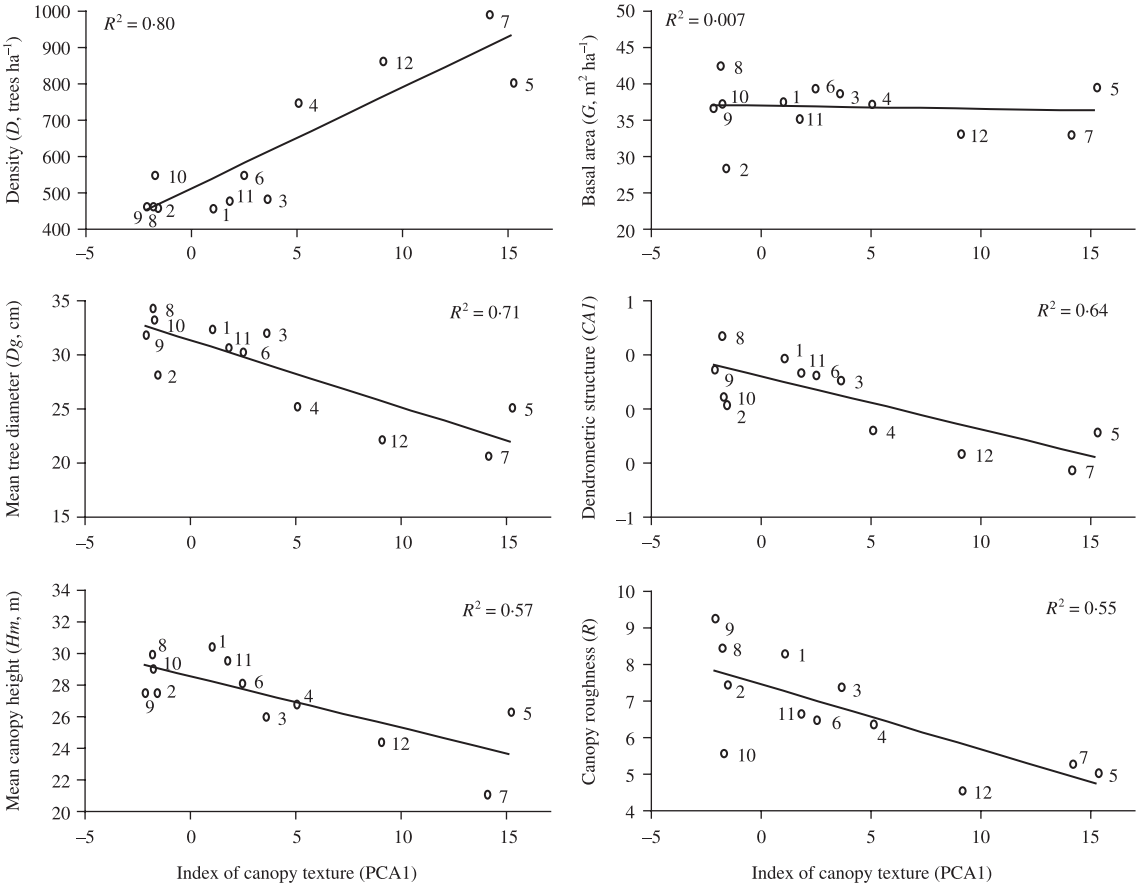


**Fig. 2.** Stand structure parameters (see Table 1) as a function of scores of the ground-truth plots along PCA axis 1 used as an index of canopy texture (see Fig. 1). Study site of the DIME project near the Petit-Saut reservoir dam, French Guiana.

trees ha$^{-1}$) of small trees ($\hat{D}g$ *c*. 26–29 cm d.b.h.). In this northern part, fine-grained windows were intermixed with coarse-grained windows (T1; dark-blue squares) containing sharp relief features such as ridges and deep valleys (Fig. 1a–g). These features, while fairly well detected by the textural analysis (Fig. 3), were infrequent on the smooth slopes of Montagne Plomb but frequent in the hilly northern part. Furthermore, it was on steep slopes in this northern part, where unfavourable substratum conditions (virtually no soil on the alterite) determine dense stands of small trees, that frequent occurrence of the finest-grained windows (Fig. 1g) was observed.

## Discussion

Two-dimensional Fourier analysis is a powerful method for obtaining both quantitative characterization of canopy texture and reasonable estimations of stand structure parameters from VHR optical images. This conclusion, drawn from the use of digitized aerial photographs, is likely to hold regarding submetric pixel satellite images, which are easier to use with respect to georeferencing and assembling into large maps. But the availability of such images at affordable cost is still problematic if the aim is to document large areas of tropical rain forest. In addition, frequent cloud coverage limits the acquisition of satellite scenes of good quality. Visual interpretation is not only time consuming but also of limited spatiotemporal consistency, and this is probably why it has not been widely applied in the tropics even though extensive air coverage has been available for as long as in the temperate zone. Indeed, only an automatic or a semi-automatic approach can deal with the subtle and spatially intricate variations that characterize tropical evergreen forest stands. Most examples of visual interpretation quoted by Polidori *et al*. (2004) dealt with contrasted and spatially segregated forest types, for example deciduous vs. evergreen.

Automatic recognition and delineation of individual tree crowns has been attempted in homogeneous, even-aged tree stands from VHR optical data at the price of complex algorithms and heavy computation (Gougeon 1995; Pouliot *et al*. 2002). Automatic delineation is also under investigation for heterogeneous natural stands (Hurtt *et al*. 2003). There are, however, good reasons why automatic delineation of individual crowns should be perceived as a difficult task in tropical stands, because of the frequent crown merging between adjacent trees. Asner *et al*. (2002), who performed a thorough manual delineation of tree crowns, acknowledged situations for which deciding between single crowns vs. merged crowns is virtually impossible and reported a serious positive bias on crown size estimation from VHR satellite imagery. Hence, broad-scale applicability of individual tree

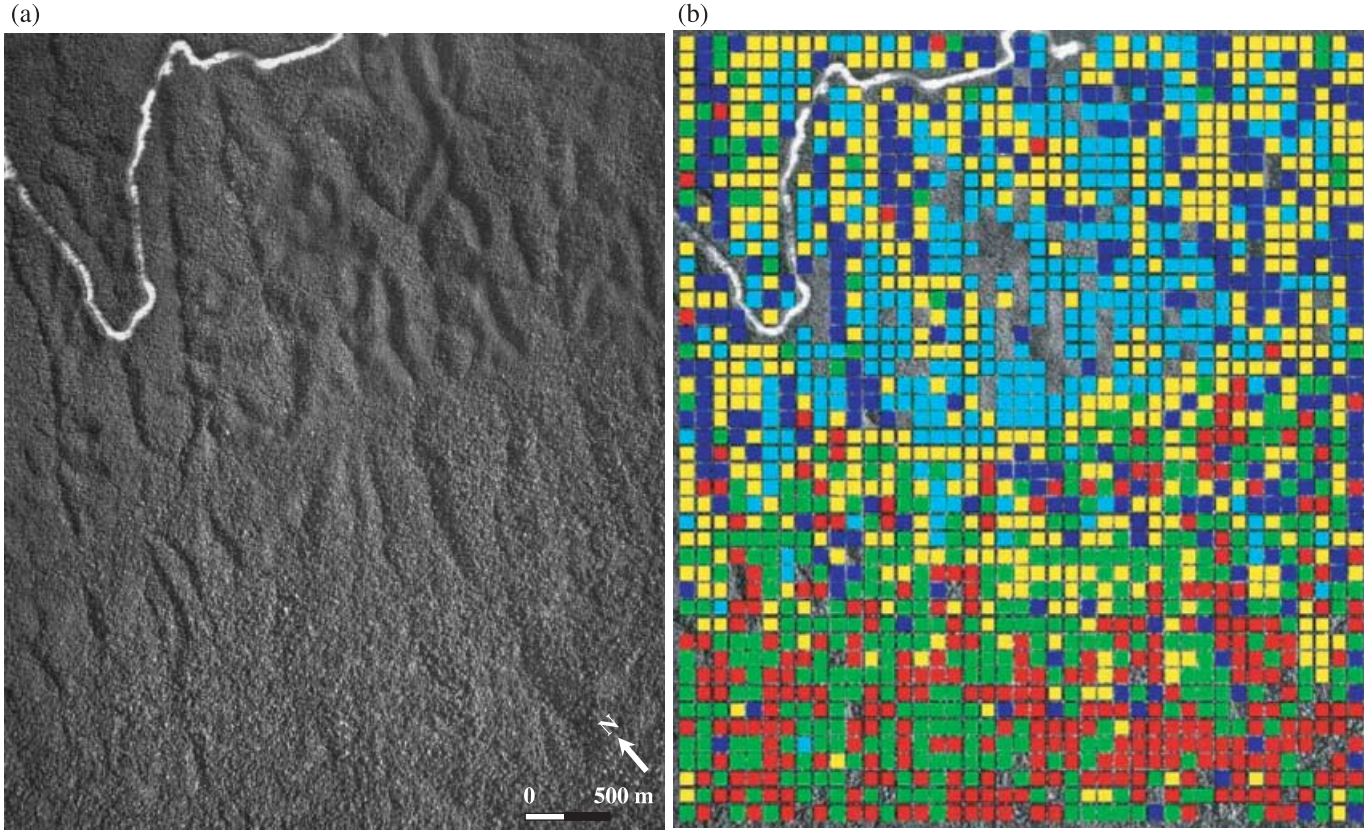


**Fig. 3.** (a) Central part of contact image number 399 (coverage 1992-GUF-91/250; IGN, France) in the study area of the DIME project, near the Petit-Saut reservoir dam, French Guiana. (b) The same area with coloured squares corresponding to 1-ha forest canopy windows coded in five classes of canopy texture (see Fig. 1, Table 2 and text): dark-blue, T1 (coarse-grained texture for windows marked by relief-induced illumination discrepancy); red, T2 (coarse-grained canopy texture); green, T3 (coarse- to intermediate-grained canopy texture); yellow, T4 (intermediate- to fine-grained canopy texture); light-blue, T5 (fine-grained canopy texture).

approaches to tropical stand characterization cannot be considered in the near future. On the other hand, our more holistic approach, which considers canopy texture as a whole, presents good prospects of immediate large-scale applicability. From this point of view, it represents a promising alternative to direct measurement of physical attributes of the vegetation, such as canopy tree height estimates from LIDAR (light detection and ranging), which require complex algorithms to separate ground returns from overlying vegetation returns, a process remaining subject to increased random errors in dense, multilayered evergreen forests (Clark, Clark & Roberts 2004).

However, as for every regression-based parameter retrieval technique, there is a need for both calibration and validation, because values of the canopy texture index are relative to a particular set of images and also because the ecological relationship between canopy texture and stand parameters has remained up to now poorly investigated. The need to measure additional field plots to recalibrate regression coefficients, for each new study area and each new set of images, may, at first sight, appear a serious hindrance to a wide-scale implementation of the method. However, it is not unrealistic to think that large data sets featuring both tree diameter data and digitized canopy images for many 1-ha plots could be efficiently set up by forest departments as part of their routine field operations of forest monitoring and inventory. From such databases, textural indices and regression coefficients with broad regional validity could be easily derived to yield automatic estimates of forest stand structure parameters over vast areas of tropical forests. This would be particularly useful to guide preparation of field prospects for consistent resource inventory, forest classification and management planning.



## Acknowledgements

This research has been carried out in the framework of the DIME project supported by the French Ministère de l'écologie et du développement durable (MEDD) through grant CV02000074. We thank C. François, J.-F. Molino, M.-F. Prévost, D. Sabatier and J.-L. Smock along with the students of the ENGREF 'FTH-2002' training stay for their contribution to the collection of field data, and F. Lokonadinpoullé for assistance with GIS. We also thank Ch. Proisy, D. Lo Seen, D.B. Clark and one anonymous referee for comments on an earlier draft of the paper.

## References

Ashton, P.S. & Hall, P. (1992) Comparisons of structure among mixed dipterocarp forests of north-western Borneo. *Journal of Ecology*, **80**, 459–481.

Asner, G.P. & Warner, A.S. (2003) Canopy shadow in IKONOS satellite observations of tropical forests and savannas. *Remote Sensing of Environment*, **87**, 521–533.

Asner, G.P., Palace, M., Keller, M., Pereira, R. Jr, Silva, J.N.M. & Zweede, J.C. (2002) Estimating canopy structure in an Amazon forest from laser range finder and IKONOS satellite observations. *Biotropica*, **34**, 483–492.

Avery, T.E. & Berlin, G.L. (1992) *Fundamentals of Remote Sensing and Airphoto Interpretation*. Macmillan, New York, NY.

Birnbaum, P. (2001) Canopy surface topography in a French Guiana forest and the folded forest theory. *Plant Ecology*, **153**, 293–300.

Chave, J., Condit, R., Lao, S., Caspersen, J.P., Foster, R.B. & Hubbell, S.P. (2003) Spatial and temporal variation of biomass in a tropical forest: results from a large census plot in Panama. *Journal of Ecology*, **91**, 240–252.

Clark, D.B., Castro, C.S., Alvarado, A.L.D. & Read, J.M. (2004a) Quantifying mortality of tropical rain forest trees using high-spatial-resolution satellite data. *Ecology Letters*, **7**, 52–59.

Clark, D.B., Read, J.M., Clark, M.L., Cruz, A.M., Dotti, M.F. & Clark, D.A. (2004b) Application of 1-m and 4-m resolution satellite data to ecological studies of tropical rain forests. *Ecological Applications*, **14**, 61–74.

Clark, M.L., Clark, D.B. & Roberts, D.A. (2004) Small-footprint lidar estimation of sub-canopy elevation and tree height in a tropical rain forest landscape. *Remote Sensing of Environment*, **91**, 68–89.

Couteron, P. (2002) Quantifying change in patterned semi-arid vegetation by Fourier analysis of digitized aerial photographs. *International Journal of Remote Sensing*, **23**, 3407–3425.

Couteron, P. & Lejeune, O. (2001) Periodic spotted patterns in semiarid vegetation explained by a propagation-inhibition model. *Journal of Ecology*, **89**, 616–628.

Couteron, P., Pélissier, R., Mapaga, D., Molino, J.-F. & Teillier, L. (2003) Drawing ecological insights from a management-oriented forest inventory in French Guiana. *Forest Ecology and Management*, **172**, 89–108.

Dawkins, H.C. (1958) The management of natural tropical high-forests with special reference to Uganda. *Imperial Forestry Institute Paper*, **34**, 1–155.

Delor, C., Lahondère, D., Egal, E., Lafon, J.-M., Cocherie, A., Guerrot, C., Rossi, P., Truffert, C., Théveniaut, H., Phillips, D. & de Avelar, V.G. (2003) Transamazonian Crustal Growth and Reworking as Revealed by the 1: 500 000-Scale Geological Map of French Guiana, 2nd edn. *Géologie de la France*, **2-3-4**, 5–57.

De Wasseige, C. & Defourny, P. (2002) Retrieval of tropical forest structure characteristics from bi-directional reflectance of SPOT images. *Remote Sensing of Environment*, **83**, 362–375.

Drake, J.B., Dubayah, R.O., Clark, D.B., Knox, R.G., Blair, J.B., Hofton, M.A., Chazdon, R.L., Weishampel, J.F. & Prince, S. (2002) Estimation of tropical forest structural characteristics using large-footprint lidar. *Remote Sensing of Environment*, **79**, 305–319.

Drake, J.B., Knox, R.G., Dubayah, R.O., Clark, D.B., Condit, R., Blair, B. & Hofton, M. (2003) Above-ground biomass estimation in closed canopy neotropical forests using lidar remote sensing: factors affecting the generality of relationships. *Global Ecology and Biogeography*, **12**, 147–159.

Dubayah, R.O. & Drake, J.B. (2000) Lidar remote sensing for forestry. *Journal of Forestry*, **98**, 44–46.

Gougeon, F.A. (1995) A crown-following approach to the automatic delineation of individual tree crowns in high spatial resolution aerial images. *Canadian Journal of Remote Sensing*, **21**, 274–284.

Grace, J. (2004) Understanding and managing the global carbon cycle. *Journal of Ecology*, **92**, 189–202.

Houghton, R.A., Lawrence, K.T., Hackler, J.L. & Brown, S. (2001) The spatial distribution of forest biomass in the Brazilian Amazon: a comparison of estimates. *Global Change Biology*, **7**, 731–746.

Hurtt, G., Xiao, X., Keller, M., Palace, M., Asner, G.P., Braswell, R., Brondizio, E.S., Cardoso, M., Carvalho, C. Jr, Fearon, M.G., Guild, L., Hagen, S., Hetrick, S., Moore, B.I.I.I., Nobre, C., Read, J.M.ST., Schloss, A., Vourlitis, G. & Wickel, A.J. (2003) IKONOS imagery for the large scale biosphere–atmosphere experiment in Amazonia (LBA). *Remote Sensing of Environment*, **88**, 111–127.

Lu, D., Mausel, P., Brondizio, E. & Moran, E. (2004) Relationships between forest stand parameters and Landsat TM spectral responses in the Brazilian Amazon Basin. *Forest Ecology and Management*, **198**, 149–167.

Manly, B.F.J. (1994) *Multivariate Statistical Methods: A Primer*. Chapman & Hall, London, UK.

Mugglestone, M.A. & Renshaw, E. (1998) Detection of geological lineations on aerial photographs using two-dimensional spectral analysis. *Computers and Geosciences*, **24**, 771–784.

Nelson, R., Oderwald, R. & Gregoire, T.G. (1997) Separating ground and airborne laser sampling phases to estimate tropical forest basal area, volume and biomass. *Remote Sensing of Environment*, **60**, 311–326.

Paget, D. (1999) *Etude de la diversité spatiale des écosystèmes forestiers guyanais: réflexion méthodologique et application*. Thèse de Doctorat. ENGREF, Nancy.

Polidori, L., Couteron, P., Gond, V., Proisy, C. & Trichon, V. (2004) Télédétection et caractérisation des paysages amazoniens. *Revue Forestière Française*, **LV**, 101–117.

Pouliot, D.A., King, D.J., Bell, F.W. & Pitt, D.G. (2002) Automated tree crown detection and delineation in high resolution digital camera imagery of coniferous forest regeneration. *Remote Sensing of Environment*, **82**, 322–334.

Read, J.M. (2003) Spatial analyses of logging impacts in Amazonia using remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, **69**, 275–282.

Read, J.M., Clark, D.B., Venticinque, E.M. & Moreiras, M.P. (2003) Application of merged 1-m and 4-m resolution satellite data to research and management in tropical forests. *Journal of Applied Ecology*, **40**, 592–600.

Riéra, B., Pélissier, R. & Houllier, F. (1998) Caractérisation d'une mosaïque forestière et de sa dynamique en forêt tropicale humide sempervirente. *Biotropica*, **30**, 251–260.

Ripley, B.D. (1981) *Spatial Statistics*. Wiley, New York, NY.

Santos, J.R., Pardi Lacruz, M.S., Aroujo, L.S. & Keil, M. (2002) Savanna and tropical rainforest biomass estimation and spatialisation using JERS-1 data. *International Journal of Remote Sensing*, **23**, 1217–1229.

Sokal, R.R. & Rohlf, F.J. (1995) *Biometry: The Principles and Practice of Statistics in Biological Research*. Freeman, San Francisco, CA.

Sommerfeld, R.A., Lundquist, J.E. & Smith, J. (2000) Characterizing the canopy gap structure of a disturbed forest using the Fourier transform. *Forest Ecology and Management*, **128**, 101–108.

Tuomisto, H., Ruokolainen, K., Kalliola, R., Linna, A., Danjoy, W. & Rodriguez, Z. (1995) Dissecting Amazonian biodiversity. *Science*, **269**, 63–66.

*Received 14 February 2005; final copy received 23 July 2005*
*Editor: Phil Hulme*