



Center for **S**tatistical **E**cology and **E**nvironmental **S**tatistics

STATISTICAL SELECTION OF PERIMETER-AREA MODELS FOR PATCH MOSAICS IN MULTISCALE LANDSCAPE ANALYSIS

By L. Grossi¹, G. P. Patil², and C. Taillie²

¹Statistical Institute, Faculty of Economics
Department of Environmental Science
University of Parma, Parma, Italy

²Center for Statistical Ecology and Environmental Statistics
Department of Statistics
The Pennsylvania State University
University Park, PA 16802

To appear in *Environmental and Ecological Statistics*, 11(1), 2004:
Special Institutional Thematic Issue on Center for Statistical Ecology and Environmental Statistics

Technical Report Number 2003-0201
TECHNICAL REPORTS AND REPRINTS SERIES
February 2003



Statistical Selection of Perimeter-Area Models for Patch Mosaics in Multiscale Landscape Analysis

L. Grossi¹, G. P. Patil² and C. Taillie²

¹Statistical Institute, Faculty of Economics
Department of Environmental Science
University of Parma, Parma, Italy

²Center for Statistical Ecology and Environmental Statistics
Department of Statistics, Penn State University
University Park, PA 16802 USA

Abstract. This paper presents a statistical method for detecting distinct scales of pattern for mosaics of irregular patches, by means of perimeter-area relationships. Krummel *et al.* (1987) were the first to develop a method for detecting different scaling domains in a landscape of irregular patches, but this method requires investigator judgment and is not completely satisfying. Grossi *et al.* (2001) suggested a modification of Krummel's method in order to detect objectively the change points between different scaling domains. Their procedure is based on the selection of the best piecewise linear regression model using a set of statistical tests. Even though the change points were estimated, the null distributions used for testing purposes were those appropriate for known change points. The present paper investigates the effect that estimating the change points has on the underlying distribution theory. The procedure we suggest is based on the selection of the best piecewise linear regression model using a likelihood ratio (LR) test. Each segment of the piecewise linear model corresponds to a fractal domain. Breakpoints between different segments are unknown, so the piecewise linear models are non-linear. In this case the frequency distribution of the LR statistic cannot be approximated by a chi-squared distribution. Instead, Monte Carlo simulation is used to obtain an empirical null distribution of the LR statistic. The suggested method is applied to three patch types (CORINE biotopes) located in the Val Baganza watershed of Italy.

Keywords. CORINE biotopes, Likelihood ratio tests, Map of Italian Nature, Monte Carlo simulation, Piecewise linear regression, Scaling domains.

1 Introduction

Fractal geometry can provide a more realistic picture of the geometry of naturally occurring objects than classical Euclidean geometry. When natural “objects” like vegetation are not constrained by human activities and land manipulation, they result in highly irregular shapes, determined by iterative and diffusive growth, which can reproduce at different scales independently of size. A shift in the boundary fractal dimension, or in related fractal parameters, may indicate a substantial change in processes generating and maintaining landscape patches at different scales (Krummel *et al.*, 1987; Palmer, 1988; Sugihara and May, 1990; Milne, 1991). Determination of such change-points can provide focus and direction for proactive land management activities.

This paper presents a statistical method for detecting distinct scales of pattern for mosaics of irregular patches, by means of fractal analysis. Krummel *et al.* (1987) were the first to develop a method for detecting different scaling regions in a landscape of irregular patches, but this method requires investigator judgment and is not completely satisfying. Grossi *et al.* (2001) suggested a modification of Krummel’s method in order to detect objectively the change points between different scaling domains. The new procedure is based on the selection of the best piecewise linear regression model using a set of statistical tests. Even though the change points were estimated, the null distributions used for testing purposes were those appropriate for known change points. The present paper investigates the effect that estimating the change points has on the underlying distribution theory. The procedure we suggest is based on the selection of the best piecewise linear regression model using a likelihood ratio (LR) test. Parametrizations for continuous and for discontinuous piecewise linear models are given in the second section. In the third section the selection of models is discussed. The LR statistic is based on the minimum residual sum of squares of the piecewise linear regression models. A method for efficient computation of the minimum residual sum of squares is introduced in section four.

Each segment of the piecewise linear model corresponds to a fractal domain. Breakpoints between different segments are unknown, so the piecewise linear models are non-linear. In this case the frequency distribution of the LR statistic cannot be approximated by a chi-squared distribution. Instead, Monte Carlo simulation is used to obtain an empirical null distribution of the LR statistic. See Kim *et al.* (2000) for a nonparametric approach to fitting continuous piecewise linear regression models.

In the final sections, the suggested method is applied to three patches type (CORINE biotopes, CEC, 1991) located in the Val Baganza watershed of Italy. For illustrative purposes, we develop the procedure for perimeter-area relationships but, with straightforward modifications, it can be applied to all power relations with fractal exponents (cf. Nikora *et al.*, 1999).

2 The Models

The aim of the suggested procedure is to detect different scaling domains over each of which the fractal dimension is approximately constant. To estimate fractal dimension (D) in each domain, we use perimeter-area relation as suggested by Lovejoy (1982). Given the areas and perimeters of n patches, we can write the perimeter-area relation as follows:

$$P_i = cA_i^{D/2},$$

where P_i and A_i are the perimeter and area of the i -th patch. Applying the logarithm transform gives

$$y_i = c + \frac{D}{2} x_i, \quad i=1, 2, \dots, n, \quad (1)$$

where $y_i = \ln(P_i)$ and $x_i = \ln(A_i)$. Note that D is twice the slope of a linear regression model, under the assumption of self-similarity, i.e., when the Hurst exponent for shape is equal to 1 (Nikora *et al.*, 1999). Different scaling domains are reflected by breakpoints where parameters of model (1) change. Breakpoints can be detected comparing model (1) to more complex models. Let θ be the breakpoint for models with one breakpoint and θ_1, θ_2 ($\theta_1 < \theta_2$) the first and second breakpoint for models with two breakpoints. We consider five alternative models:

$$y = \beta_0 I_{x \leq \theta} + \beta'_0 I_{x > \theta} + \beta_1 x + \varepsilon \quad (2)$$

$$y = \beta_0 + \beta'_1(x I_{x \leq \theta} + \theta I_{x > \theta}) + \beta''_1(\theta I_{x \leq \theta} + x I_{x > \theta}) + \varepsilon \quad (3)$$

$$y = (\beta'_0 + \beta'_1 x) I_{x \leq \theta} + (\beta''_0 + \beta''_1 x) I_{x > \theta} + \varepsilon \quad (4)$$

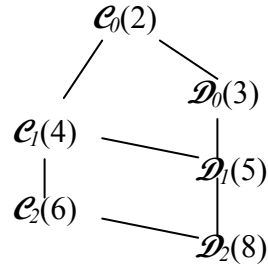
$$y = \beta_0 + \beta'_1(x I_{x \leq \theta_1} + \theta_1 I_{x > \theta_1}) + \beta''_1(\theta_1 I_{x \leq \theta_1} + x I_{\theta_1 < x \leq \theta_2} + \theta_2 I_{x > \theta_2}) + \beta'''_1(\theta_2 I_{x \leq \theta_2} + x I_{x > \theta_2}) + \varepsilon \quad (5)$$

$$y = (\beta'_0 + \beta'_1 x) I_{x \leq \theta_1} + (\beta''_0 + \beta''_1 x) I_{\theta_1 < x \leq \theta_2} + (\beta'''_0 + \beta'''_1 x) I_{x > \theta_2} + \varepsilon, \quad (6)$$

where I is an indicator variable that is equal to 1 when the subscripted condition is true and is equal to 0 otherwise. In model (2) two parallel segments describe the data: the intercept changes in going from the first to the second segment but the slope remain unchanged so that there is only one fractal domain. In models (3) and (4) two fractal domains are considered, while in models (5) and (6) there are three fractal domains. Parameters β'_1, β''_1 and β'''_1 are the fractal dimension in first, second and third domains, respectively. Model (2) is a piecewise linear regression model (Draper and Smith, 1998) with two discontinuous segments and no change of slope, and is written as \mathcal{D}_0 . Models (3) and (5) are piecewise linear regression models with, respectively, two and three continuous segments and they have one and two changes (change points) of slope, so we called them model \mathcal{E}_1 and \mathcal{E}_2 . Models (4) and (6) are piecewise linear models with, respectively, two (one change point) and three (two change points) discontinuous segments and are called, respectively, \mathcal{D}_1 and \mathcal{D}_2 .

More generally, let \mathcal{E}_r , $r = 0, 1, 2, \dots$, be the continuous piecewise linear model with r change points and $(r + 1)$ domains. The number of estimated parameters in \mathcal{E}_r is $2(r + 1)$: one intercept, r change points and $(r + 1)$ slopes. Let \mathcal{D}_r , $r = 0, 1, 2, \dots$ be the discontinuous piecewise linear model

with $(r + 1)$ domains. The number of estimated parameters in \mathcal{D}_r is 3 when $r = 0$ (two intercepts and one slope) and $3r + 2$ when $r \geq 1$: one slope and one intercept for each of $(r + 1)$ domains and r change points. Putting the number of regression parameters in parentheses, we have the following nested collection of models:



In a previous paper (Grossi *et al.*, 2001), a \mathcal{D}_2 model was selected by a procedure applied to 378 patches. The scatterplot of the data and the corresponding fitting is reported in Figure 1.

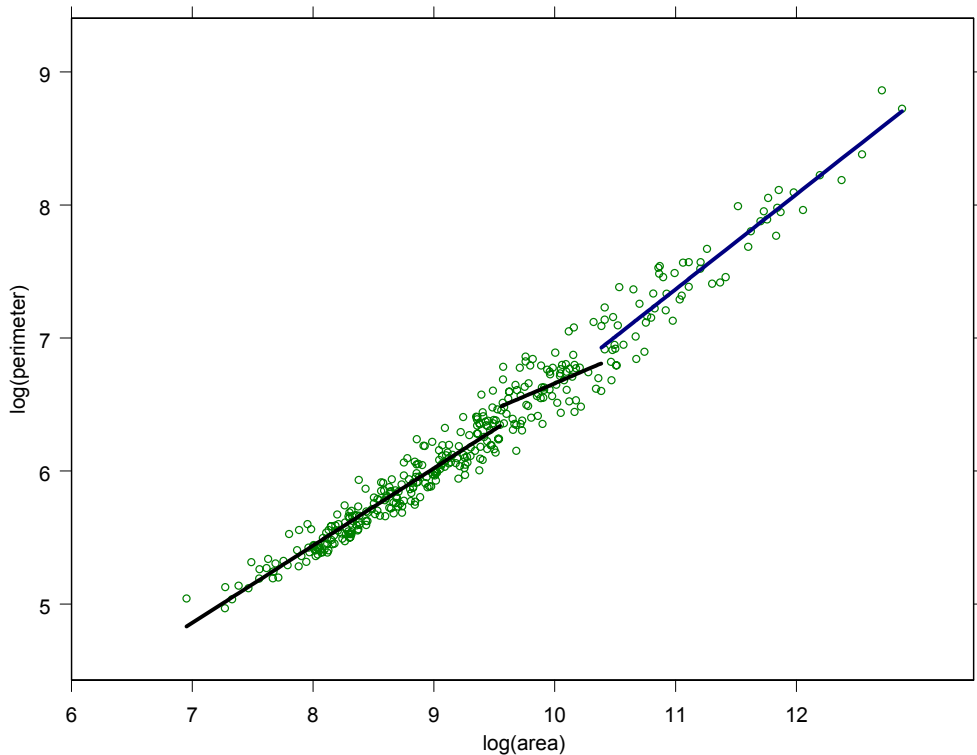


Figure 1. Scatterplot of $\log(\text{perimeter})$ vs. $\log(\text{area})$ with fitted model superimposed. Data from Grossi *et al.* (2001).

The referees questioned the use of a discontinuous piecewise linear model and asked if a continuous piecewise linear model might fit just as well. Specifically, would the use of a continuous model yield only one change point (and two scaling domains) in the above example? This paper develops the methodology for fitting the continuous piecewise linear model and for testing continuity vs. discontinuity.

Asymptotic distribution theory for maximum likelihood supposes that likelihood is a smooth function of the parameters (Gallant, 1987). In our case, the likelihood is not differentiable and, for the discontinuous piecewise linear model, is not even continuous. Accordingly, this paper examines (via simulation) the sampling distribution of the likelihood ratio test statistic under these nonstandard conditions.

3 Model Fitting and Testing

The choice among the nested models is a typical problem of variable selection. In multiple linear regression, variable selection is typically uses the F -test to measure the statistical significance of adding variables. Since the breakpoints in Models (2)–(6) are unknown parameters that have to be estimated, the corresponding regression models are not linear. But for non-linear models, the F -distribution does not necessarily apply to variable selection procedures. The problem can be studied using maximum likelihood and likelihood ratio (LR) tests.

Let $Y_i \sim N(\mu_i, \sigma^2)$, $i = 1, 2, \dots, n$ be the dependent variable of a linear regression model where the errors are Gaussian with means $\mu_i = \mu(X_i, \Phi)$; here, $\Phi = (\phi_1, \phi_2, \dots, \phi_p)'$ is a vector of unknown parameters that can vary independently of σ^2 . The maximum likelihood estimate of Φ minimizes the residual sum of squares,

$$SSE = \sum [Y_i - \mu(x_i, \hat{\Phi})]^2.$$

Let ω and Ω be two nested regression models having the same variance σ^2 , with p and $p + q$ regression parameters, respectively. The hypotheses $H_0 : \omega$ vs. $H_A : \Omega$ can be tested using the following LR test statistic:

$$\lambda = n \ln \left(\frac{SSE_{\hat{\omega}}}{SSE_{\hat{\Omega}}} \right),$$

where $SSE_{\hat{\omega}}$ and $SSE_{\hat{\Omega}}$ are the residual sum of squares of ω and Ω , respectively. The rejection region can be expressed as

$$\lambda > c_1 \tag{7}$$

or

$$F = \frac{n - p - q}{q} \left(\frac{SSE_{\hat{\omega}}}{SSE_{\hat{\Omega}}} - 1 \right) > c_2. \tag{8}$$

Note that

$$\frac{SSE_{\hat{\omega}}}{SSE_{\hat{\Omega}}} - 1 = \frac{SSE_{\hat{\omega}} - SSE_{\hat{\Omega}}}{SSE_{\hat{\Omega}}} = \frac{SSR(\hat{\Omega} | \hat{\omega})}{SSE_{\hat{\Omega}}},$$

so

$$F = \frac{MSR(\hat{\Omega} | \hat{\omega})}{MSE(\hat{\Omega})}$$

Under regularity conditions and asymptotically for large n , the critical point c_1 in (7) is obtained from a chi-squared distribution with q degree of freedom. For a linear model the critical point c_2 in (8) is obtained from an F -distribution with q and $(n-p-q)$ degree of freedom (Myers, 1986, pp. 95–111). The F -test in (8) is usually preferred, for linear models, since it is exact even for small sample sizes. However for non-linear models, the F -statistic in (8) does not have to follow an F -distribution and the chi-squared test (7) is usually preferred (Gallant, 1987, pp.47–104), unless a simulation study shows for the particular models involved that the F -distribution gives a better approximation to (8) than the chi-squared gives to (7).

The distribution theory for λ and F comes out of the likelihood. But likelihood theory is premised on the likelihood function being a smooth function of the parameters. However, the models \mathcal{D}_r , $r \geq 1$, have a likelihood function that is a discontinuous, piecewise constant, function of the breakpoints. For \mathcal{C}_r , $r \geq 1$, the likelihood is a continuous function of the breakpoints, but is not differentiable when a breakpoint equals a data value X_i . Therefore, simulation is needed to check whether χ_q^2 and/or $F(q, n-p-q)$ are good approximations to the sampling distribution of λ and F .

4 Computing the Minimum Residual Sum of Squares

Let \mathbf{Y} be the vector containing the values of the dependent variables in a regression model and \mathbf{X} the corresponding design matrix. The residual sum of squares can be expressed as

$$SSE = \mathbf{Y}'\mathbf{Y} - \mathbf{U}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{U} \quad (9)$$

where $\mathbf{U} = \mathbf{X}'\mathbf{Y}$. When the breakpoints $\theta_1, \theta_2, \dots$ are given, the corresponding models are linear, so $SSE(\theta_1, \theta_2, \dots)$ may be computed according to (9) and then be minimized with respect to the breakpoints. The resulting minimum is the value of SSE to be used in the test statistics (7) and (8). When n is large and, particularly, when simulations have to be run, we need efficient computational procedures for SSE . The procedures are different for discontinuous and continuous models.

4.1 Discontinuous Models

First Consider the \mathcal{D}_1 model in (4). Let θ be the unknown breakpoint. Let x_1, x_2, \dots, x_n and y_1, y_2, \dots, y_n the values of the X and Y variables arranged in increasing order with respect to X . The following symbols are then used:

$$\begin{aligned}
n &= \#x_i & S_x &= \sum x_i & S_{xx} &= \sum x_i^2 & S_y &= \sum y_i & S_{xy} &= \sum x_i y_i \\
n' &= \#(x_i \leq \theta) & S'_x &= \sum_{x \leq \theta} x_i & S'_{xx} &= \sum_{x \leq \theta} x_i^2 & S'_y &= \sum_{x \leq \theta} y_i & S'_{xy} &= \sum_{x \leq \theta} x_i y_i \\
n'' &= \#(x_i > \theta) & S''_x &= \sum_{x > \theta} x_i & S''_{xx} &= \sum_{x > \theta} x_i^2 & S''_y &= \sum_{x > \theta} y_i & S''_{xy} &= \sum_{x > \theta} x_i y_i
\end{aligned} \tag{10}$$

Then we can write

$$\mathbf{U} = [S'_y, S'_{xy}, S'_y, S'_{xy}] \text{ and } \mathbf{X}'\mathbf{X} = \begin{bmatrix} n' & S'_x & 0 \\ S'_x & S'_{xx} & 0 \\ 0 & n'' & S''_x \\ & S''_x & S''_{xx} \end{bmatrix}$$

Using (9), it can be shown that

$$SSE(\theta) = \mathbf{Y}'\mathbf{Y} - \left[\frac{S'_y S'_y}{n'} + \frac{\left(S'_{xy} - \frac{1}{n'} S'_x S'_y \right)^2}{S'_{xx} - \frac{1}{n'} S'_x S'_x} \right] - \left[\frac{S''_y S''_y}{n''} + \frac{\left(S''_{xy} - \frac{1}{n''} S''_x S''_y \right)^2}{S''_{xx} - \frac{1}{n''} S''_x S''_x} \right] \tag{11}$$

Now, $SSE(\theta)$ changes only when θ crosses one of the x_i values. So the minimization of (11) can be performed putting $\theta = x_3, x_4, \dots, x_{n-3}$. When θ changes from x_i to x_{i+1} , the residual sum of squares of the new model can be computed without fitting the model, but simply subtracting from the S'' quantities the values corresponding to the $(i+1)$ -th case and adding the same values to the S' quantities. The model \mathcal{D}_2 defined in (6) is handled similarly, except another set of quantities S''' is needed in (10) and another term must be added in (11) corresponding to the S''' quantities. When model \mathcal{D}_0 defined in (2) is considered the final formula for $SSE(\theta)$ is the following:

$$SSE(\theta) = \mathbf{Y}'\mathbf{Y} - \frac{S'_y S'_y}{n'} + \frac{S''_y S''_y}{n''} + \frac{1}{n' n''} \frac{\left(n' n'' S_{xy} - n' S''_x S''_y - n'' S'_x S'_y \right)^2}{\left(n' n'' S_{xx} - n' S''_x S''_x - n'' S'_x S'_x \right)}.$$

4.2 Continuous Models

First Consider the \mathcal{C}_l model. Again, let θ be the breakpoint. Given the parameterization in (3) it can be found that

$$\mathbf{U} = [S_y, S'_{xy} + \theta S''_y, \theta S'_y + S''_{xy}] \text{ and } \mathbf{X}'\mathbf{X} = \begin{bmatrix} n & S'_x + n'' \theta & S''_x + n' \theta \\ S'_{xx} + n'' \theta^2 & \theta S'_x + \theta S''_x & \\ & S''_{xx} + n' \theta^2 & \end{bmatrix}.$$

In this case, expressing $SSE(\theta)$ symbolically, in terms of the quantities in (10), is very complicated, so it is more efficient to compute it numerically from (9). Now, $SSE(\theta)$ is a continuous function of θ , which is also differentiable except when $\theta = x_i$, so it is not possible to find its minimum as in the case of the discontinuous models. In the following, we suggest one way to minimize $SSE(\theta)$ for continuous models. The first step of this procedure is the computation of $SSE(\theta)$ for $\theta = x_3, x_4, \dots, x_{n-3}$ using the quantities in (10) to find the x_i which corresponds to the minimum. Then it is reasonable to suppose that the global minimum occurs for θ in $[x_{i-1}, x_i] \cup [x_i, x_{i+1}]$. It can be proved (see Appendix) that $\frac{\partial SSE(\theta)}{\partial \theta}$ can vanish for at

most two critical values of θ which we call θ^* and θ^{**} . Further these values are the solutions of the following quadratic equation:

$$a\theta^2 + b\theta + c = 0$$

This quadratic equation is derived in the Appendix. When θ^* and θ^{**} have been estimated, one way of determining which of the critical point corresponds to the minimum is the following: compute $SSE(\theta)$ for any of the critical values in the intervals $[x_{i-1}, x_i]$ and $[x_i, x_{i+1}]$, and then compare those SSE values with the SSE values at the endpoints of the intervals. This method does not guarantee a global minimum, but it is a fast way to come near it.

Now consider model \mathcal{E}_2 . Let θ_1 and θ_2 be the breakpoints with $\theta_1 < \theta_2$. Given the parameterization in (5), the relevant matrices for computing SSE are

$$\mathbf{U} = \begin{bmatrix} S_y, & S'_{xy} + \theta_1 S''_y + \theta_1 S'''_y, & \theta_1 S'_y + S''_{xy} + \theta_2 S'''_y, & \theta_2 S'_y + \theta_2 S''_y + S'''_y \end{bmatrix}$$

$$\mathbf{X'X} = \begin{bmatrix} n & S'_x + \theta_1(n'' + n''') & \theta_1 n' + S''_x + \theta_2 n''' & \theta_2(n' + n'') + S'''_x \\ S'_{xx} + \theta_1^2(n'' + n''') & \theta_1(S'_x + S''_x) + n''' \theta_1 \theta_2 & \theta_2 S'_x + n'' \theta_1 \theta_2 + \theta_1 S'''_x \\ & \theta_1^2 n' + S''_{xx} + \theta_2^2 n''' & n' \theta_1 \theta_2 + \theta_2(S''_x + S'''_x) \\ & & \theta_2(n' + n'') + S'''_{xx} \end{bmatrix}.$$

As in the case of model \mathcal{E}_1 , $SSE(\theta_1, \theta_2)$ is more efficiently computed numerically from (9). First $SSE(\theta_1, \theta_2)$ is computed for each combination of $\theta_1 = x_3, x_4, \dots, x_{n-6}$ and $\theta_2 = x_6, x_7, \dots, x_{n-3}$, given that $\theta_1 < \theta_2$, using the quantities in (10) to find the x_i and x_j which correspond to the minimum. Next, the intervals $[x_i, x_{i+1}]$ and $[x_j, x_{j+1}]$ are each divided into four subintervals of equal length. Then $SSE(\theta_1, \theta_2)$ is compared to SSE estimated at each combination of the limits of the subintervals.

5 Landscape Analysis in the Map of the Italian Nature

The procedure has been applied to landscape pattern data produced within the Map of the Italian Nature project (Zurlini *et al.*, 1999). The data concern mosaics of contiguous, irregular patches at

a scale of 1:25000 corresponding to different habitat types in the CORINE biotopes classification (CEC, 1991). The CORINE classification, derived by satellite, airborne (photographs and hyperpectral images) and field data, permits a biophysical identification of ecosystems (sensu Tansley, 1935) as patches, based on vegetation covers, physiognomy, soil types and land-forms, and an integration of abiotic and biotic components extremely closely related, historically and evolutionary, like habitats and syntaxa (Usher, 1991). CORINE units, called habitats, are coded at different levels of a hierarchical system of mosaics of patches within patches that can go from very broad syntaxa at the landscape level down to alliances and associations. Among the 2,327 CORINE patches identified in the Val Baganza watershed, the three most frequent types of patches were here considered for analysis: *low land hay meadows* (CORINE code 38.2; 378 patches), *brachypodium grassland* (CORINE code 36.334; 131 patches), *northern apennine mesobromion* (CORINE code 34.3266; 77 patches). Other types of biotopes were not analyzed because the corresponding sample sizes are too small to perform the procedure. Original perimeter-area data are available from the first author on request.

Table 1. Residual sum of squares (SSE) and degrees of freedom (df) for the best fitting models for each biotope.

model	CORINE 38.2		CORINE 36.334		CORINE 34.3266	
	SSE	df	SSE	df	SSE	df
C_0	7.0035	376	3.0638	129	1.1634	75
C_1	6.5581	374	2.3676	127	1.137	73
C_2	6.4115	372	2.2869	125	1.1108	71
D_0	6.5591	374	2.3092	127	1.0989	73
D_1	6.5113	373	2.2822	126	1.0985	72
D_2	6.2278	370	2.0926	123	0.99994	69

Models \mathcal{C}_0 , \mathcal{C}_1 , \mathcal{C}_2 , \mathcal{D}_0 , \mathcal{D}_1 and \mathcal{D}_2 have been fitted to each biotope and estimated minimum sum of squares are reported in Table 1. In order to select the best model for each biotope, we should compare nested models computing the corresponding LR statistics. As pointed out, the LR in this case does not necessarily follow a chi-squared distribution, so a simulation study is needed.

6 Simulations

Recall that we want to test $H_0 : \omega$ vs. $H_A : \Omega$. The problem is that we don't know the exact distribution of the LR λ (7) and of the F -statistic (8), so that we run simulations. In running simulations, the data must be generated from ω . Then the test statistics λ and F can be computed using $SSE_{\hat{\omega}}$ and $SSE_{\hat{\Omega}}$ estimated from the generated data. Different null models are possible.

Case 1. When the null model ω is \mathcal{C}_0 the simulated distributions of the statistics (7) and (8) do not depend on the parameters of \mathcal{C}_0 . To see this, observe that none of the SSE values computed for any of the considered models are affected if Y_i is replaced by $Y_i - (u + vX_i)$, $i = 1, 2, \dots, n$, for

given constants u, v . Also, if Y_i is divided by σ , then all of the SSE values are divided by σ^2 . But, since (7) and (8) involve ratios of SSE , the test statistics are unchanged. Consequently, in doing simulations, we can make any convenient transformation of the form

$$Y_i^* = \frac{Y_i - (u + vX_i)}{\sigma}, \quad i = 1, 2, \dots, n$$

without affecting the null distribution. Let \mathcal{C}_0 be the null model ω . The alternative model Ω might be any of the more complex models $\mathcal{C}_1, \mathcal{C}_2, \mathcal{D}_0, \mathcal{D}_1$, and \mathcal{D}_2 . The parametric representation of ω is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$

Take $u + vX_i = \beta_0 + \beta_1 X_i$ to get

$$Y_i^* = \varepsilon_i^*, \quad \varepsilon_i^* \sim N(0, 1). \quad (12)$$

The conclusion is that the null hypothesis does not depend on the parameters of ω . The data are generated from simulation model (12). Of course, $SSE_{\hat{\omega}}$ and $SSE_{\hat{\Omega}}$ must still be computed for the fully parameterized ω and Ω .

Case 2. Suppose the null model ω is \mathcal{C}_1 . In this case, the alternative model Ω might be any of $\mathcal{C}_2, \mathcal{D}_1, \mathcal{D}_2$. The parametric representation of ω is in (3) with $\varepsilon \sim N(0, \sigma^2)$. Take $u + vx = \beta_0 + \beta'_1 x + \beta''_1 \theta$ to get the simulation model

$$y^* = \begin{cases} \varepsilon^* & x \leq \theta \\ \frac{\beta''_1 - \beta'_1}{\sigma} (x - \theta) + \varepsilon^* & x > \theta \end{cases}$$

where $\varepsilon^* \sim N(0, 1)$. Therefore, the null distribution depends on at most θ and $\frac{\beta''_1 - \beta'_1}{\sigma}$. The dependence on θ is not with respect to its numerical value, but rather in relation to the data values.

Case 3. Suppose the null model ω is \mathcal{D}_1 . The alternative model Ω can only be \mathcal{D}_2 . Parametric representation of ω is given in (4). Take $u + vx = \beta'_0 + \beta'_1 x$ to get

$$y^* = \begin{cases} \varepsilon^* & x \leq \theta \\ \frac{\beta''_0 - \beta'_0}{\sigma} + \frac{\beta''_1 - \beta'_1}{\sigma} \theta + \frac{\beta''_1 - \beta'_1}{\sigma} (x - \theta) + \varepsilon^* & x > \theta \end{cases}$$

So the null distribution may depend on θ and on $\frac{\beta_0'' - \beta_0'}{\sigma}, \frac{\beta_1'' - \beta_1'}{\sigma}$.

Case 4. When the null model ω is \mathcal{C}_2 , the alternative model Ω can only be \mathcal{D}_2 . Parametric representation of ω is given in (5). Take $u + vX_i = \beta_0 + \beta_1'X_i + \beta_1''\theta_1 + \beta_1'''\theta_2$ to get the simulation model as follows

$$Y_i^* = \begin{cases} \varepsilon_i^*, & X_i \leq \theta_1 \\ \frac{\beta_1'' - \beta_1'}{\sigma}(X_i - \theta_1) + \varepsilon_i^*, & \theta_1 < X_i \leq \theta_2 \\ \frac{\beta_1'' - \beta_1'}{\sigma}(\theta_2 - \theta_1) + \frac{\beta_1''' - \beta_1'}{\sigma}(X_i - \theta_2) + \varepsilon_i^*, & \theta_2 < X_i \end{cases}$$

where $\varepsilon_i^* \sim N(0,1)$. In this case, the null distribution depends on at most $\theta_1 < \theta_2$ and

$$\frac{\beta_1'' - \beta_1'}{\sigma}, \frac{\beta_1''' - \beta_1'}{\sigma}.$$

In running the simulations, we used as regressors the surfaces of the three CORINE biotopes previously described. The sample sizes of these data sets are 77, 131 and 378. We found that both F and λ have empirical null distributions that cannot be approximated with the nominal F and chi-squared distribution, respectively. Thus, we limited the analysis to the LR statistic. For each null model and for each sample size the 0.95 percentiles (simulated critical values) of the empirical simulated distribution of the LR statistic are reported in Table 2. There were 6000 replicates for each model Ω when the null model is \mathcal{C}_0 and 5000 in the other cases.

Table 2. Simulated critical values for LR statistic when: a) $H_0: \mathcal{C}_0$ (6000 replications), b) $H_0: \mathcal{D}_0$ (5000 replications), c) $H_0: \mathcal{C}_1$ (5000 replications) d) $H_0: \mathcal{D}_1$ (5000 replications), e) $H_0: \mathcal{C}_2$ (5000 replications). The sample sizes (number of patches) of the generated data sets are equal to the sizes of three data sets in the application. The critical values are the 0.95 percentiles of the empirical distribution.

a) $H_0: \mathcal{C}_0$

N. of patches	Simulated crit. Value for LR statistics				
	\mathcal{C}_1	\mathcal{C}_2	\mathcal{D}_0	\mathcal{D}_1	\mathcal{D}_2
378	7.42587	13.64751	11.0674	13.59459	25.47887
131	7.353848	12.88947	10.62594	13.05476	23.49382
77	7.302105	12.48273	10.42061	12.64406	22.60455

b) $H_0: \mathcal{D}_0$

N. of patches	\mathcal{D}_1	\mathcal{D}_2
378	4.62296	17.2541
131	5.2296	16.2808

c) $H_0: \mathcal{C}_1$

N. of patches	\mathcal{C}_2	\mathcal{D}_1	\mathcal{D}_2
378	9.05461	8.55445	21.1631
131	8.29954	7.06306	18.1489

d) $H_0: \mathcal{D}_1$

N. of patches	\mathcal{D}_2
378	15.4899
131	13.5358

e) $H_0: \mathcal{C}_2$

N. of patches	\mathcal{D}_2
378	15.113
131	12.9376

7 Best Fitting Models for the Val Baganza Watershed

The LR statistics and the corresponding empirical p -values are reported in Table 3. It is worth noting that, when \mathcal{C}_0 is the null model, the p -value is always less than 0.01 for sample size 131 and 378, while is always over 0.50 for sample size equal to 77 (Table 3a). This means that the null hypothesis (model \mathcal{C}_0) is always rejected for patches with sample sizes 131 or 378, while the hypothesis of a simple straight line cannot be rejected for patches with sample size of 77 (Figure 2).

Table 3. LR statistics and empirical p -values computed for alternative models when: a) $H_0: \mathcal{C}_0$, b) $H_0: \mathcal{D}_0$, c) $H_0: \mathcal{C}_1$, d) $H_0: \mathcal{D}_1$, and e) $H_0: \mathcal{C}_2$.

a) $H_0: \mathcal{C}_0$

N. of patches		\mathcal{C}_1	\mathcal{C}_2	\mathcal{D}_0	\mathcal{D}_1	\mathcal{D}_2
378	LR stat.	24.8380	33.3837	24.7804	27.5452	44.3722
	<i>Emp. p-value</i>	0.0000	0.0000	0.0002	0.0002	0.0000
131	LR stat.	33.7691	38.3121	37.0409	38.5816	49.9436
	<i>Emp. p-value</i>	0.0000	0.0000	0.0000	0.0000	0.0000
77	LR stat.	1.7674	3.5625	4.3919	4.4199	11.6583
	<i>Emp. p-value</i>	0.6098	0.8588	0.5645	0.8242	0.8457

b) $H_0: \mathcal{D}_0$

N. of patches		\mathcal{D}_1	\mathcal{D}_2
378	LR stat.	2.7650	19.5900
	<i>Emp. p-value</i>	0.1282	0.0182
131	LR stat.	1.5410	12.9030
	<i>Emp. p-value</i>	0.2806	0.1526

c) $H_0: \mathcal{C}_1$

N. of patches		\mathcal{C}_2	\mathcal{D}_1	\mathcal{D}_2
378	LR stat.	8.5460	2.7070	19.5340
	<i>Emp. p-value</i>	0.0650	0.6224	0.0946
131	LR stat.	4.5430	4.8130	16.1750
	<i>Emp. p-value</i>	0.2878	0.1586	0.0966

d) $H_0: \mathcal{D}_1$

N. of patches		\mathcal{D}_2
378	LR stat.	16.8270
	<i>Emp. p-value</i>	0.0274
131	LR stat.	11.3620
	<i>Emp. p-value</i>	0.1296

e) $H_0: \mathcal{C}_2$

N. of patches		\mathcal{D}_2
378	LR stat.	10.9890
	<i>Emp. p-value</i>	0.2484
131	LR stat.	11.6310
	<i>Emp. p-value</i>	0.0934

We can conclude that for the northern apennine mesobromion CORINE biotope (code 34.3266) the one-straight line model can not be rejected, so that all patches belong to the same fractal domain. This kind of CORINE biotope does not apparently present any statistically significant shift in boundary fractal dimension. This may indicate no substantial change in scale as regards

generating and maintaining processes of landscape patches. This kind of biotope is indeed present in naturally stressed environments due to drought, and thus already adaptive to relatively extreme environmental conditions and rather insensitive to human disturbances.

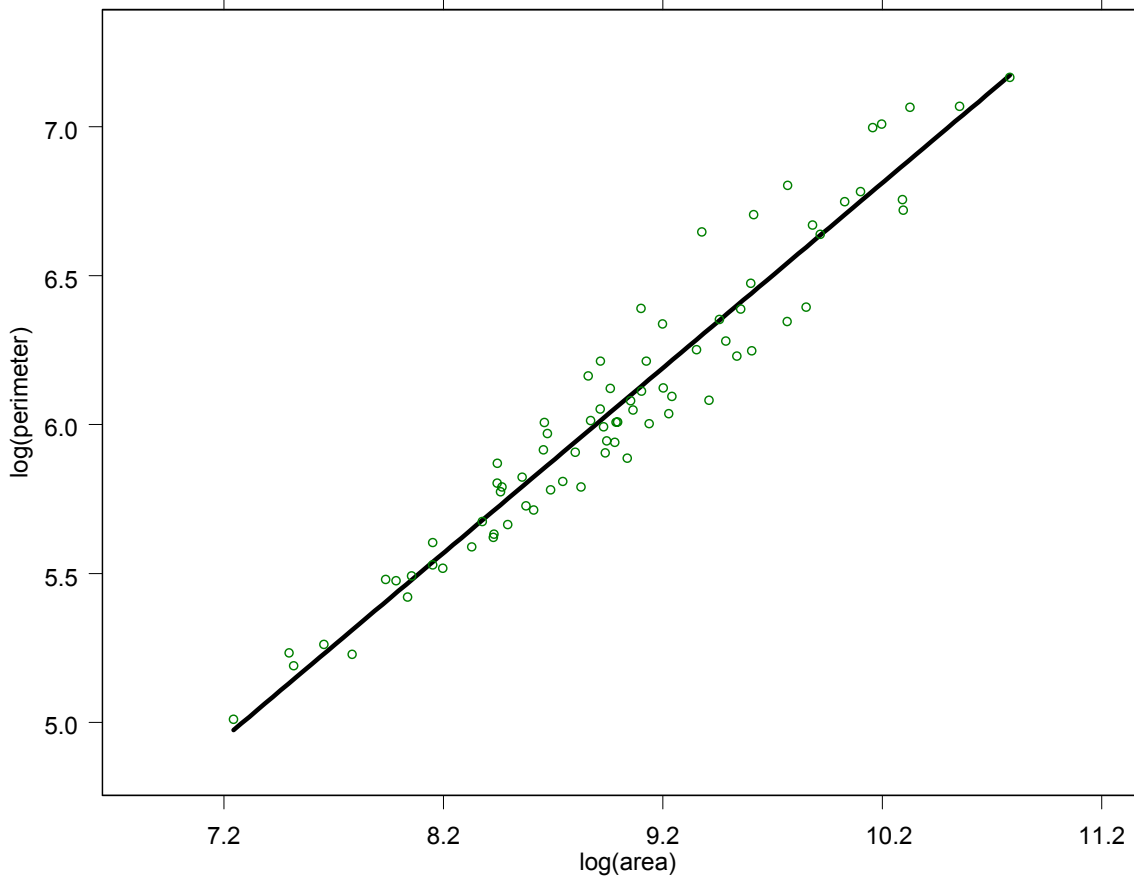


Figure 2. CORINE code 34.3266. Scatterplot of the data with superimposed fitted model \mathcal{C}_0 .

For the remaining biotopes, further tests are necessary to select the best descriptor of the data, at least among the models that we have considered. Different hypotheses are possible. The null hypothesis of \mathcal{D}_0 (Table 3b) cannot be rejected for the CORINE biotope 36.334, *brachypodium grassland*, so that no discontinuous models are warranted for this biotope. On the other hand, \mathcal{D}_0 should be rejected for the CORINE code 38.2, *low land hay meadows*, with the largest number of patches, because \mathcal{D}_2 is clearly better. Model \mathcal{C}_1 (Table 3c) cannot be rejected for either biotope at the 0.05 probability level. Nevertheless, the p -value is very near to 0.05 for model \mathcal{C}_2 in the case of low land hay meadows. There is an indication that something more complicated than \mathcal{C}_1 is needed for this larger data set. Looking at the other null hypotheses (table 3d and table 3e), we observe that for CORINE code 38.2, \mathcal{D}_2 is better than \mathcal{D}_1 , but not better than \mathcal{C}_2 . The conclusion could be that the best descriptor of the behaviour of this biotope is \mathcal{C}_2 . In the case of the CORINE code 36.334 the best model is \mathcal{C}_1 .

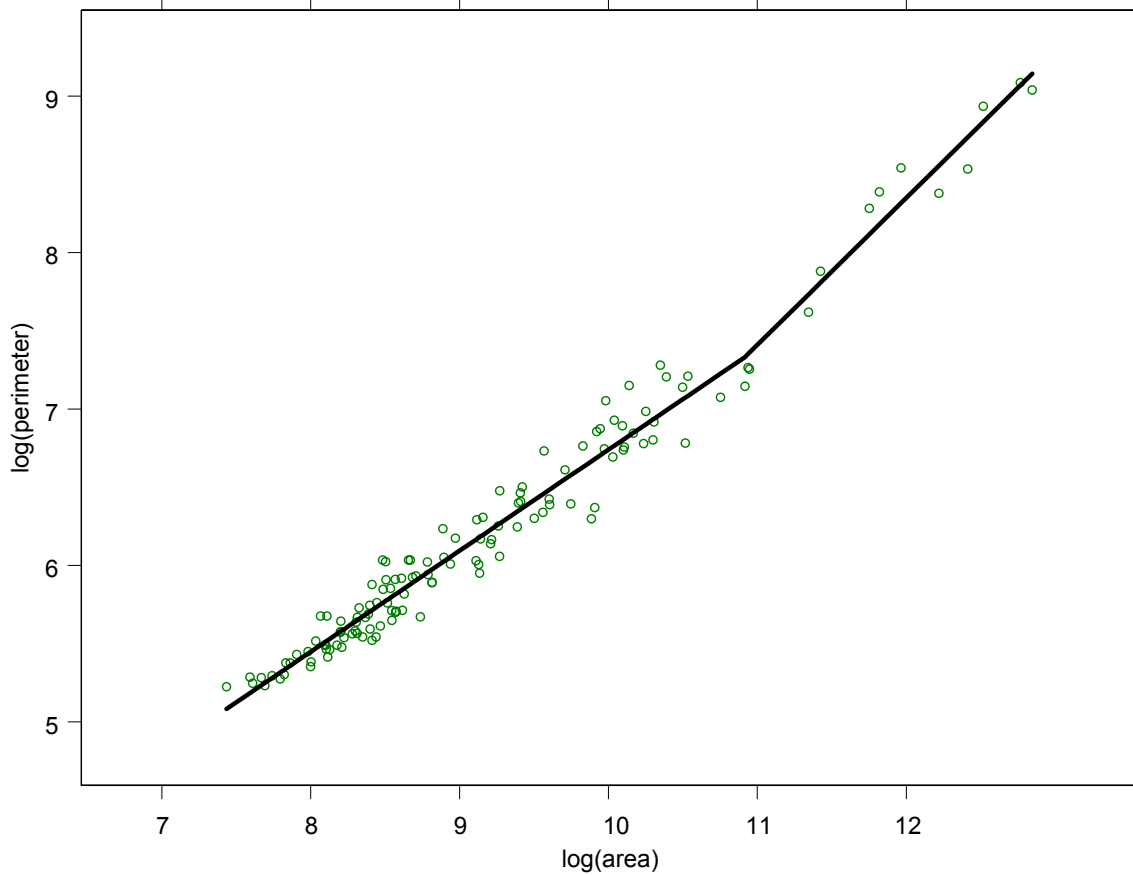


Figure 3. CORINE code 36.334. Scatterplot of the data with superimposed fitted model \mathcal{C}_1 .

Figure 3 reports the fitting of the \mathcal{C}_1 model for *brachypodium grassland* (CORINE code 36.334). There is a clear change point (at about 5 ha). At larger scales, patches assume a more complex shape, perhaps related to a lower human disturbance. In Table 4 are reported the fractal dimensions and the corresponding standard errors in each domain for the CORINE code 36.334.

Table 4. Fractal dimensions, standard errors and fractal domains for CORINE code 36.334.

	Value	st. error	domain (m ²)
D1	1.292	0.0277	0 - 55143
D2	1.879	0.079	> 55143

In Figure 4, the fitting of model \mathcal{C}_2 for *low land hay meadows* (CORINE code 38.2) is reported. Change points are at 3277 m² and 31611 m². Fractal dimension increases as area increases, so that largest patches look to be more regular and less disturbed than smaller ones. The first fractal dimension is less than one but statistically not different from one at the 0.05 level. In Table 5 are reported the fractal dimensions and the corresponding standard errors in each domain for the CORINE code 38.2.

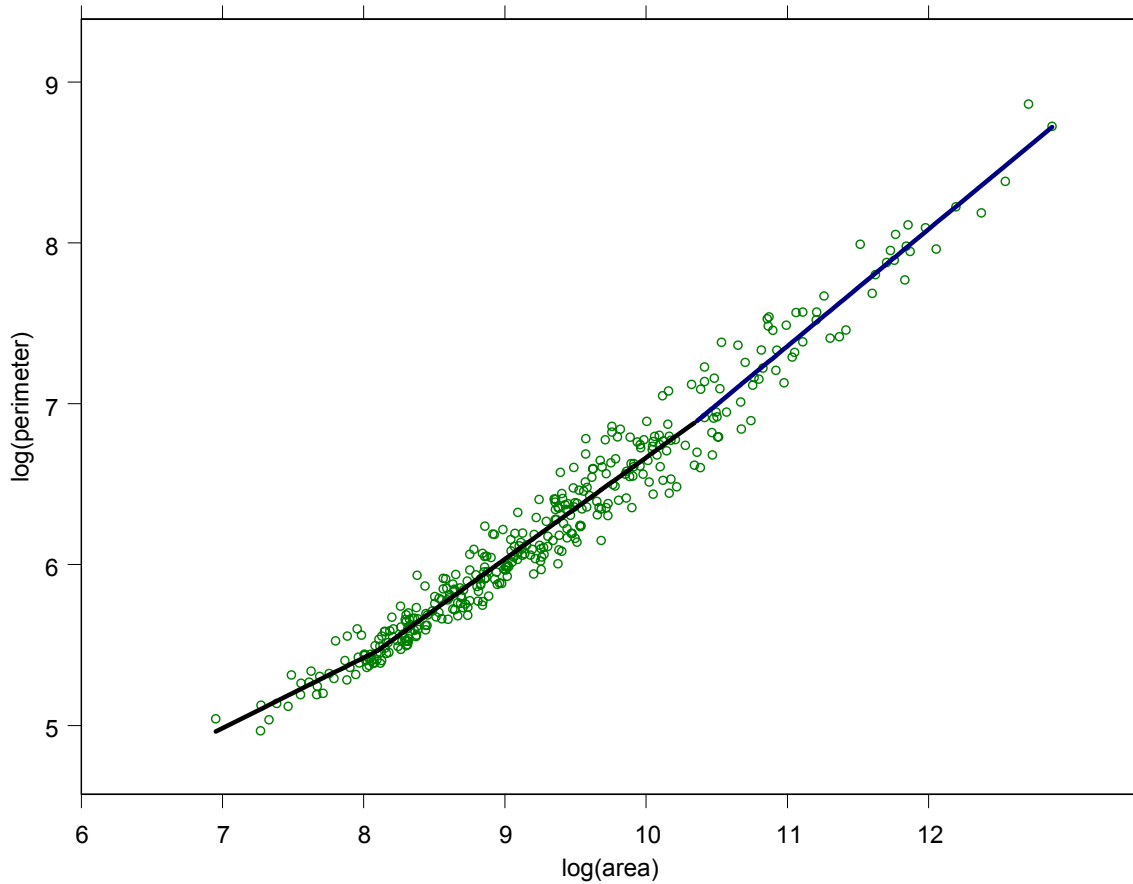


Figure 4. CORINE code 38.2. Scatterplot of the data with superimposed fitted model \mathcal{C}_2 .

Table 5. Fractal dimensions, standard errors and fractal domains for CORINE code 38.2.

	value	st. error	domain (m ²)
D1	0.8786	0.111	0 - 3277
D2	1.2622	0.0206	3277 - 31611
D3	1.4532	0.04	> 31611

8 Conclusions

We have presented a method to detect breakpoints in the fractal dimension of irregular patches. The relevance of this topic is related to the identification of landscape scaling regions with distinct dimensions connected by transition zones (Holling, 1992). According to hierarchical theory in ecology, distinct levels in ecological system should be reflected in corresponding distinct scales of patterns in space. Two distinct models have been suggested to fit perimeter-area data: continuous piecewise linear models and discontinuous piecewise linear models. In the first case segments of the regression curve are connected while in the second case segments are disconnected.

Since the change points between different segments are unknown, the classical procedure of variable selection in regression analysis based on the F -distribution can not be applied. Therefore the best piecewise model has been selected using a LR statistic based on the ratio between the minimum sums of squares of two alternative models. The suggested statistical method requires efficient computation of the minimum sum of squares because all models related to any possible breakpoint must be estimated. Some suggestions have been given in order to cope with these computational issues.

Simulation has revealed that the nominal null distribution can not be used for testing different models when breakpoints are unknown. Its use would lead to spurious significance. Approximation is worse for discontinuous models. Empirical distributions of the test statistics have been obtained by means of a simulation procedure. Simulation results are only applicable to the specific data set used in the application.

The procedure has proven to be effective in detecting landscape patterns when applied to patch mosaics of the Map of the Italian Nature Project. In this context, the procedure was helpful in discriminating among CORINE biotopes with respect to their landscape pattern. CORINE biotope types seem to behave differently according to their scaling properties resulting from the interaction between their ecological properties and human disturbance.

An open question is the eventual existence of an asymptotic null distribution for the LR statistic different from the nominal chi-squared distribution. If so, it would be desirable to describe these limiting distributions in order to avoid simulations when the method is applied to different data. One possible approach is a transformation $G(\cdot)$ so that

$$G(\lambda) = \chi_q^2. \quad (13)$$

The simplest such transformation would be $G(\lambda) = \lambda / c$ for some constant c . Assuming that G is monotone increasing, (13) implies that

$$G(\lambda_\alpha) = \chi_{q,\alpha}^2 \quad (14)$$

for all levels α . The right hand side of (14) can be computed while λ_α can be estimated from the simulations. Therefore, a plot of $\chi_{q,\alpha}^2$ versus $\hat{\lambda}_\alpha$ is an estimate of G . Additional questions are whether this plot depends on the size of the data set and on the null hypothesis.

9 Appendix

This appendix studies the behavior of $SSE(\theta)$ between any two consecutive x -values, say x_i and x_{i+1} , and shows that $\frac{\partial SSE(\theta)}{\partial \theta} = 0$ for at most two critical values of θ which we call θ^* and θ^{**} .

It is shown also that these values can be obtained as the solution of a quadratic equation.

We can write $(\mathbf{X}'\mathbf{X})^{-1}$ as $\frac{\mathbf{A}}{\det}$, where \mathbf{A} is the adjoint matrix of $\mathbf{X}'\mathbf{X}$ and $\det = \det(\mathbf{X}'\mathbf{X})$.

Then we obtain that $\mathbf{U}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{U} = \frac{\mathbf{U}'\mathbf{A}\mathbf{U}}{\det}$. Using the quantities in (10) it can be shown that $\mathbf{U}'\mathbf{A}\mathbf{U}$ and \det are both quadratic functions of θ , so we can do a polynomial division to get

$$\mathbf{U}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{U} = Q + \frac{R}{\det}, \quad (\text{A1})$$

The quotient Q in (A1) does not depend on θ and R is linear in θ . So the residual sum of squares is

$$SSE(\theta) = \mathbf{Y}'\mathbf{Y} - Q - \frac{R}{\det}. \quad (\text{A2})$$

Since $\mathbf{Y}'\mathbf{Y} - Q$ is independent of θ , the behavior of (A2) between two consecutive data points is determined by R/\det . Write

$$\frac{R}{\det} = \frac{\alpha + \beta\theta}{\gamma + \delta\theta + \varepsilon\theta^2}, \quad (\text{A3})$$

where $\alpha, \beta, \gamma, \delta, \varepsilon$ can be expressed in terms of the quantities in (10). But (A3) vanishes for exactly one value of θ , since the numerator is linear. Further, for large θ , $R/\det \approx \beta/\varepsilon\theta$, so $(R/\det) \rightarrow 0$ as $\theta \rightarrow \pm\infty$ and there are only two critical points θ^* and θ^{**} . To find the value of these critical points, put the derivative of R/\det equal to zero, that is

$$\frac{R'\det - R\det'}{\det^2} = 0$$

which is equivalent to

$$H \equiv R'\det - R\det' = 0.$$

Since R is linear and \det is quadratic in θ , H is quadratic in θ , so that

$$H = a\theta^2 + b\theta + c. \quad (\text{A4})$$

Coefficients a, b, c of (A4) might be expressed in terms of the quantities in (10), so that they can be computed any time without fitting the regression model. These expressions are not reported here because they are very long.

10 References

- CEC (Commission of the European Community) (1991). CORINE Biotopes manual, habitats of the European Community. A method to identify and describe consistently sites of major importance for nature conservation, EUR 12587/3, Bruxelles.
- Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis* (third edition). Wiley, New York.
- Gallant, A. R. (1987). *Nonlinear Statistical Models*. Wiley, New York.
- Grossi, L., Zurlini, G., and Rossi, O. (2001). Statistical detection of multiscale landscape patterns. *Environmental and Ecological Statistics*, **8**, 253–267.
- Holling, C. S. (1992). Cross-scale morphology, geometry, and dynamics of ecosystems. *Ecological Monographs*, **62**, 447–502.
- Kim, H.-J., Fay, M. P., Feuer, E. J., and Midthune, D. N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in medicine*, **19**, 335–351.
- Krummel, J. R., Gardner, R. H., Sugihara, G., O'Neill, R. V. and Coleman, P. R. (1987). Landscape patterns in a disturbed environment. *Oikos*, **48**, 321–324.
- Lovejoy, S. (1982). Area-perimeter relation for rain and cloud areas. *Science*, **216**, 185–187.
- Milne, B.T. (1991). Lessons from applying fractal models to landscape patterns. In *Quantitative Methods in Landscape Ecology*, M.G. Turner and R.H. Gardner (eds.). Springer-Verlag, Berlin. pp.199–235.
- Myers, R. H. (1986). *Classical and Modern Regression with Applications*, (second edition), PWS-KENT, Boston.
- Nikora, V. I., Pearson, C. P. and Shankar, U. (1999). Scaling properties in landscape patterns: New Zealand experience. *Landscape Ecology*, **14**, 17–33.
- Palmer, M.W. (1988). Fractal geometry: a tool for describing spatial patterns of plant communities. *Vegetatio*, **75**, 91–102.
- Sugihara, G. and May, R.M. (1990). Applications of fractals in ecology. *Trends in Ecology and Evolution*, **5**, 79–86.
- Tansley, A.G. (1935). The use and abuse of vegetational concepts and terms. *Ecology*, **16**, 284–307.
- Usher, M. B. (1991). Habitat structure and the design of nature reserves. In: *Habitat Structure - The Physical Arrangement of Objects in Space*, S.S. Bell, E.D. McCoy and H.R. Mushinsky (eds.). Chapman and Hall, London. pp. 373–391.

Zurlini, G., Amadio, V., and Rossi, O. (1999). A landscape approach to biodiversity and biological integrity planning: The Map of the Italian Nature. *Ecosystem Health* (submitted).