4 Comparison of Corine Land Cover and FMERS-WiFS raster images for description of forest structure and diversity over large areas

4.1 Introduction:

In the previous chapter, focus was on forest structure, and it was demonstrated how it is possible to use spatial metrics from medium-resolution satellite images to predict the values of the same metrics when derived from high-resolution images. The study area was in Umbria in central Italy, and GIS-data from the same geographical window were used to analyse the effects of scaling i.e. changing pixel size on the value of the metrics. An important finding was that in order to quantify and compare the *distribution of* spatial properties over landscapes, subsets of the particular landscapes can be analysed, and results represented in geo-referenced map or table form. For the analyses, binary images were used – allowing calculation of only structural parameters, whereas in this chapter thematic maps with a number of forest classes are used – making it possible to calculate metrics of diversity and patch numbers.

The purpose of the analyses carried out here is to evaluate the use of land cover data in raster format for mapping of forest structure and composition over large areas[10], with intended use in monitoring of ecological conditions and forest resource management. For such larger areas, i.e. at national to continental scale, a need for methods to assess landscape structure and, as part hereof, diversity has been identified, in order to supplement traditional forest area and production statistics (Haines-Young and Chopping 1996, McCormick and Folving 1998, Häusler *et al* 2000, Riitters *et al* 2000, Weber and Hall 2001). The spatial metrics were here extracted by application of a moving-windows (M-W) method to data originating from high- to medium-resolution satellite imagery. As part of the study, some software tools were developed, that take categorical maps (in raster image form) as input and output quantitative

[10] The total extent of the area studied here being 350,000 km^2, corresponding to areas for estimation of *epsilon* and *delta* diversities, ref. Figure 2.2.

information on such landscape parameters as fragmentation and diversity. The information is contained in raster images format (through the rows and columns of a landscape metrics matrix), which can be subjected to further statistical analysis for the entire image, selected regions or strata. At the same time, 'window size profiles' or scalograms were used to describe the scaling effects on the calculation of the chosen spatial metrics.

4.1.1 Large area forest mapping and M-W analyses

M-W methods are an obvious choice for extraction of large map-like sets/tables of spatial metrics from raster-format land-cover maps, as they allow comparison of spatial metrics for various landscapes (O'Neill *et al* 1996, Schumaker 1996, Saura and Millan-Martinez 2001), However, the interpretation of the outputs is not always straightforward (as discussed in detail by McGarigal and Marks 1995, O'Neill *et al* 1996, Haines-Young and Chopping 1996, EU-DG AGRI and others 2000, Remmel and Csilag 2002). In this chapter, the challenges that accompany selection of the central parameter window size will be illustrated and discussed. The task is, expressed in landscape ecological terms (Forman and Godron 1986, McGarigal and Marks 1995), to find the relevant extent of the sub-landscapes for which the different spatial metrics should be derived and used. This is primarily done by modifying the size of the 'moving window'. The MW approach with optional overlap is illustrated in Figure 4.8, below. As stated above, the outputs from MW-analysis themselves can be used as maps illustrating e.g. forest structure.

Some research and pilot projects have already been carried out, in which land cover maps and MW techniques are used to assess forest and landscape structure, even at continental to global level. In a report produced for the EU's general directorate for Agriculture, with the title "From Land Cover to Landscape Diversity in the European Union", a group of researchers investigated the use of CORINE land cover (CLC) data for assessment of landscape diversity with the use of M-W and per-region methods (EU, DG AGRI and others 2000). The methodology has later on been used for development of "Agri-environmental indicators" at EU level (EU-DG AGRI and others 2000, Gallego 2002). As a contribution to these studies, Eiden *et al* (2000) assessed different types of reference units for appropriate retrieval of landscape metrics, including administrative regions, German "Naturräumliche Einheiten" (landscape units) and French "Region Agricole" (agricultural regions) as well as a simple M-W approach with window sizes at 20, 40, 60 and 80km and 50% overlap between each window step to extract values of Shannon's Diversity index . They concluded that it was possible to delimit "hard core" zones of diversity or homogeneity of the European territory. At 20km window size, it was possible to identify region specific properties of the structural indicators, while at 80km window size, regional differences were smoothed out and only the strongest features remained. To produce a clearer image, the M-W results were re-sampled to a 2km grid, using bilinear interpolation.

At a global level, Riitters *et al* (2000) used 1-km resolution land-cover maps for analysis of forest fragmentation world wide[11], and extracted spatial information for windows ranging from 9 x 9 pixels, termed "small" scale to 243 x 243 pixels, termed "large" scale. The information on pixel numbers and adjacency was then used to characterize the fragmentation around each forested pixel. The result of the analysis was reported as a kind of thematic map, with pixels assigned to a certain 'fragmentation class'. This approach is rather subjective, although the output maps are illustrative and provide a useful overview of the selected structural parameters. It is worth noting, that as window sizes increased, forest areas shifted from being characterised as *interior, perforated* and *undetermined* into the types *edge, transitional* and *patch*. Furthermore, more fragmentation was detected as the number of included forest types increased, especially in areas where savannah is dominant. This is one among many examples of the influence on metrics values from the definition of forest in the mapping phase.

[11] This study is also presented in section 2.3.4 about measurement of fragmentation.

The European Environment Agency (EEA) has conducted a study of how forest in Europe is fragmented by transportation networks (EEA 2000)[12]. In this study CLC data were used, aggregated to 1*1 km grids – thus the forest patches were defined at a different scale from the original data. The results were clear: fragmentation measured as 'average size of non-fragmented land parcels' was highest (i.e. smallest parcel sizes found) in highly urbanised countries like Belgium and Luxemburg and lowest (largest parcel sizes) in the sparsely populated countries Finland and Sweden, which have large areas of continuous forest. This last study adopted a more traditional GIS-method, in reporting the results a country level – which makes sense as the desired output is indicators for the included countries. O'Neill *et al* (1996) analysed landscape patterns in the South-eastern USA using classified NOAA AVHRR images and metrics calculated for hexagons of 640 km^2 each. They also used compositional (Dominance) and shape (Shape Complexity) metrics and found that in order to get meaningful results, the grain should be 2 to 5 times smaller than the features of interest (i.e. forest or landscape patches); meanwhile the sample area or window must be 2 to 5 times larger than the patches in order to get representative metric values.

Medium resolution forest maps covering all or most of Europe have been constructed independently in at least two instances. During the FMERS project, the Technical Research Centre of Finland (VTT) led a consortium, which produced forest maps at a resolution of 200m for large parts of the continent, based on data from the satellites/sensors of the types Spot, Landsat, IRS-WiFS, Resurs MSU-SK, and ERS SAR. The purpose of this study was mostly method development (Häme *et al* 1999). Later on, another project concerned with creation of a pan-European forest map, also based on WiFS data was carried out by the Munich-based company GAF, on a similar contract to SAI. This project has demonstrated the

[12] The indicator fact sheet is available at http://themes.eea.eu.int/Sectors_and_activities/transport/indicators/consequences/fragmentation/TERM_2002_06_EUAC_Fragmentation.pdf Accessed 12/8 2003.

feasibility of creating a coherent and reliable forest map, which covers all of Europe, and the resulting map is available for later analysis[13] (GAF 2001).

The existence of the above-mentioned data sets, methodologies and results together provide the potential for analysis and mapping of forest structure in Europe, based on spatial metrics and land cover data. However, a need for methods to assess the robustness and flexibility/transferability of the various proposed metrics still exists. In this chapter methods for comparison of metrics derived in multiple matching geographical windows are proposed, and their use demonstrated on a data set consisting of two forest maps in raster image format.

4.2 Objectives

The main objective of this chapter is to compare the spatial metrics that result from applying similar methods of calculation to land-cover data sets available at different thematic and spatial resolutions. The goals are

(i) Development of new spatial metrics, particularly suited for description of forest structure and diversity over large areas and/or recommendations for the use of existing ones.

(ii) To find the optimal window size for display and reporting of landscape spatial metrics.

(iii) To test the robustness of the metrics through their use with two different data sources that provide forest maps of the same area.

(iv) Furthermore, the aim is to examine and compare scaling effects as expressed by window size on the values of various spatial metrics. This will be done through comparison of the values of the different spatial metrics, as well as the variability and autocorrelation against window size for each metric, and calculating the correlation coefficients for these relations.

[13] On request to the JRC which managed the project on behalf of the EU commission.

(v) Also of interest is the 'internal' correlations between values of different metrics (from the same input image) at a fixed window size, and comparison of these 'patterns of correlation' at different window sizes

(vi) To find out how well one land cover data set can substitute the other for mapping of structural features. This will be assessed and shown through correlations between values from the two different input types at similar window sizes (representing identical geographical areas).

(vii) Finally, catchment/watershed information and regional/administrative borders as vector GIS data are used for reporting and summarising metrics values, thus addressing the MAUP, which is an issue of concern in Remote Sensing and GIS, especially in relation to (the use of) spatial metrics. Though the metrics values are known to vary with window size, their relative values in different, separate regions might co-vary with window size, to yield the same order or ranking of the regions. This property is also expected for the two different data sets at similar window sizes.

Throughout this chapter, different types of scalograms will be used as tools to describe landscape structure and to compare maps and landscapes. At the end of the chapter, the MAUP addressed through different regionalisation approaches. It can however be argued that the use of M-Ws itself is an attempt to overcome the MAUP (Marceau *et al* 1994, Marceau 1999a and b, Marceau and Hay 1999a).

4.3 Data

In this section, the test area for this study is briefly presented, then the different data types used are described as well as the approaches to convert them to compatible forest maps.

4.3.1 Study area

In order to address the objectives stated above, forest maps of the study area were extracted from CLC and FMERS data respectively. The area investigated in this study is shown in Figure 4.1.

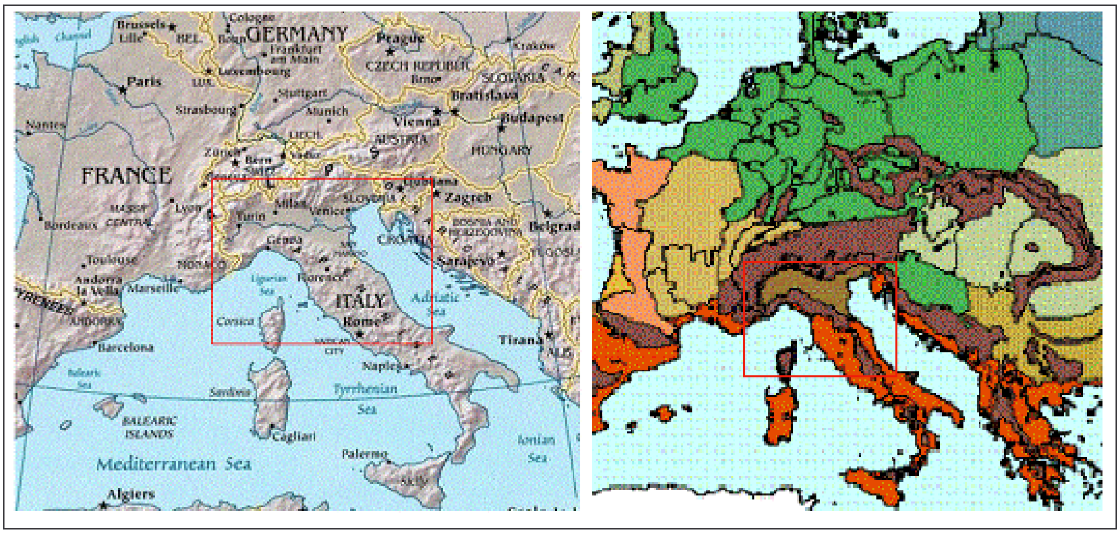

Figure 4.1 The selected subset, as shown by the red rectangle, covering Northern Italy and small parts of France, Switzerland and Slovenia. Dominant natural features are the Alpine and Apennine mountain chains and the Po river valley. The spatial extent of the subset is 500 by 700 km. To the left location on a political map with relief, to the right forest strata from the FIRS project (Kennedy *et al* 1997)[14]. Forest strata included are Mediterranean region (orange), the warm/moderate temperate region (light brown) and the Alpine and Apennine orobioms (elevational communities and associations - dark brown).

This area contains a variety of different landscape- and forest types, and includes the area in Umbria that was covered in the previous chapter. Other criteria for the selection of this test area was the presence of different forest and landscape types, and the availability of good quality forest maps. The image files are of size up to 5000*7000 pixels (at 100m cell size), large enough to produce statistically significant results even when the number of output cells decreases following the use of larger window sizes. The types of forest diversity under investigation are thus epsilon diversity (broad region) in the inventory domain and delta (between landscapes) diversity in the differentiation domain, as defined in section 2.1.4.

4.3.2 Raster data

4.3.2.1 WiFS – FMERS data

The forest map derived 'directly' from EO data used here is based on a mosaic of WiFS images from the IRS 1-C satellite, these are the same images that were used in chapter 3 – the project is introduced in section 4.1.1. The map was produced by VVT-Finland on contract to

[14] The (sub)project web site is at http://www.vtt.fi/tte/research/tte1/tte14/proj/firs/found1.html, accessed 25/4 2004

SAI, and the steps of the image preparation and processing are described in Häme *et al* (1999), for spectral properties etc. refer section 3.2. The aim of that study was in particular to develop a fast, reliable and cost-efficient method for mapping and monitoring of forest at the continental level. The 'demonstration' forest map, that was created, has the following classes, defined in accordance with the FIRS nomenclature system (Kennedy *et al* 1997):

1. **Coniferous**
2. **Broadleaved Deciduous**
3. **Broadleaved Evergreen**
4. **Mixed forest**
5. **Other Wooded Land Coniferous**
6. **Other Wooded Land Broadleaved**
7. **Other Land.**

The resolution of the original images is 188m pixel size, the mosaic was re-sampled to a pixel size of 200m. The resulting, simplified forest map is shown in Figure 4.2.

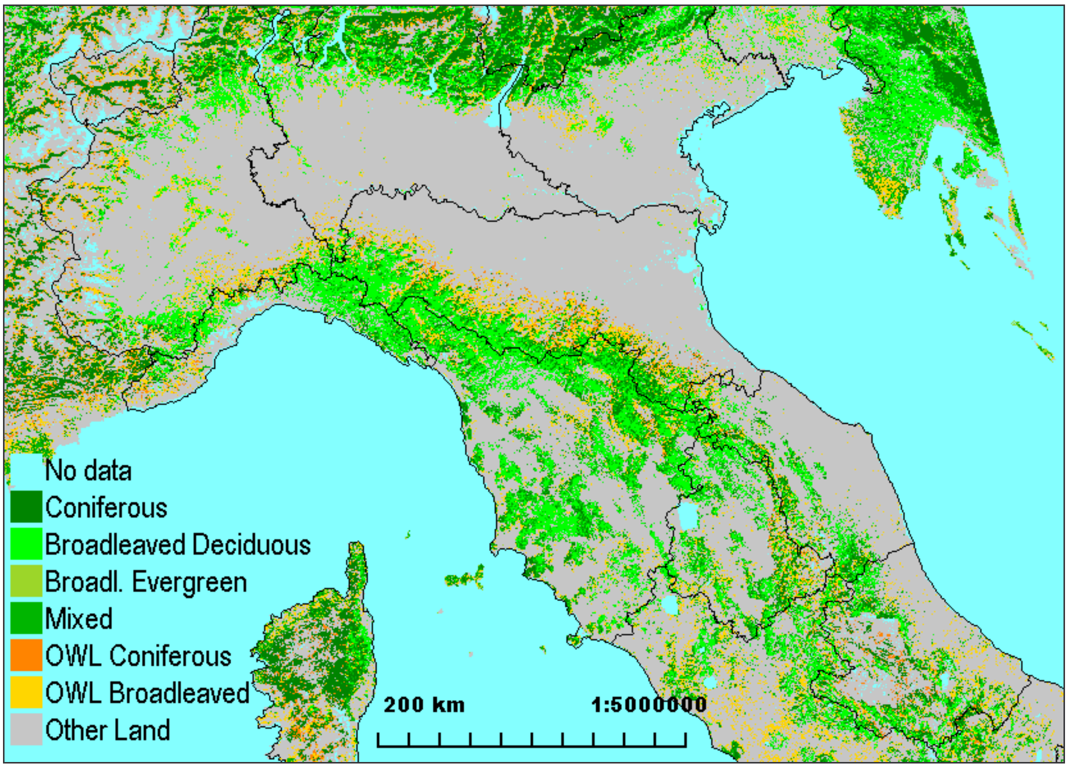

Figure 4.2 FMERS forest map for area of interest, vector layer showing NUTS-level 2 regions.

4.3.2.2 CORINE land cover data

Data from the CLC (described in section 2.3.2) are used here in the form of raster images with a pixel size of 100m. The data were extracted from the CLC database in February 2001

(Liberta 2001). The information in the database is based on Landsat TM and SPOT HRV imagery, which has been digitised manually, with a minimum patch (polygon) size of 25 ha. CLC data are interesting because they are regularly updated and standardised between the individual countries and producers (with next updated version, termed CLC2000 expected early 2004[15]). This makes CLC data useful for monitoring purposes and comparisons across Europe (EU, DG AGRI and others 2000). The three 'pure' forest classes from CLC were included in the present analysis, along with the classes Agro-forest areas, Sclerophyllous Vegetation and Transition woodland-scrub. The agro-forest class was included as forest, since it is defined as Annual crops or grazing land under the wooded cover of forestry species (Bossard *et al* 2000). This land-cover class includes areas of forest trees mixed with fruit and olive trees. The CLC image data were then re-classified to provide a forest map similar to the WiFS, though direct comparison is complicated by different nomenclatures, as seen from Table 4.1, below. Figure 4.3 shows an example of how the data are aggregated to a forest map and Figure 4.4 shows the resulting CLC-based forest map for the study area.

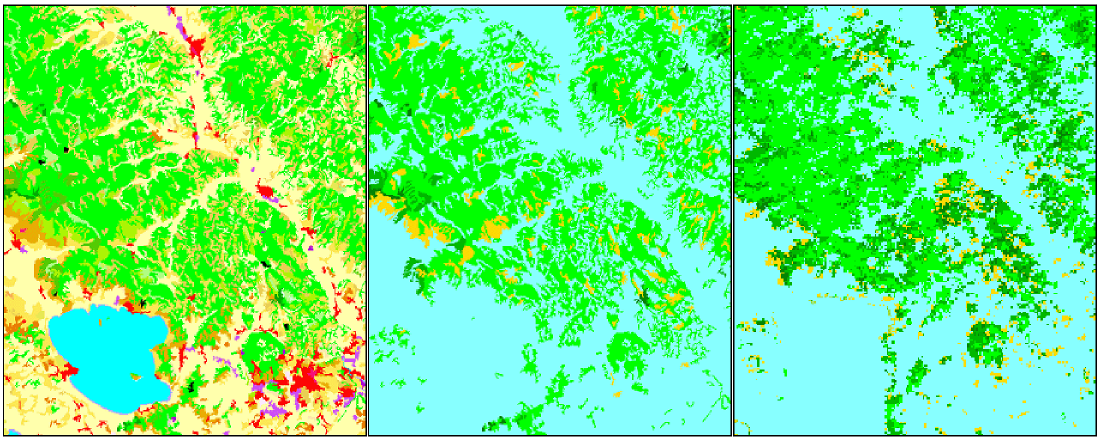

Figure 4.3 Subsets of CLC and FMERS maps located in Umbria and Toscana, Area extent 42*50 km, with the Trasimeno lake and regional capital Perugia at the bottom. From left to right: Original CLC map with all possible land cover classes, map with only the forest classes (both pixel size 100m) and FMERS map of same area (pixel size 200m).

[15] Regular updates on mapping and availability status are provided at http://terrestrial.eionet.eu.int/CLC2000

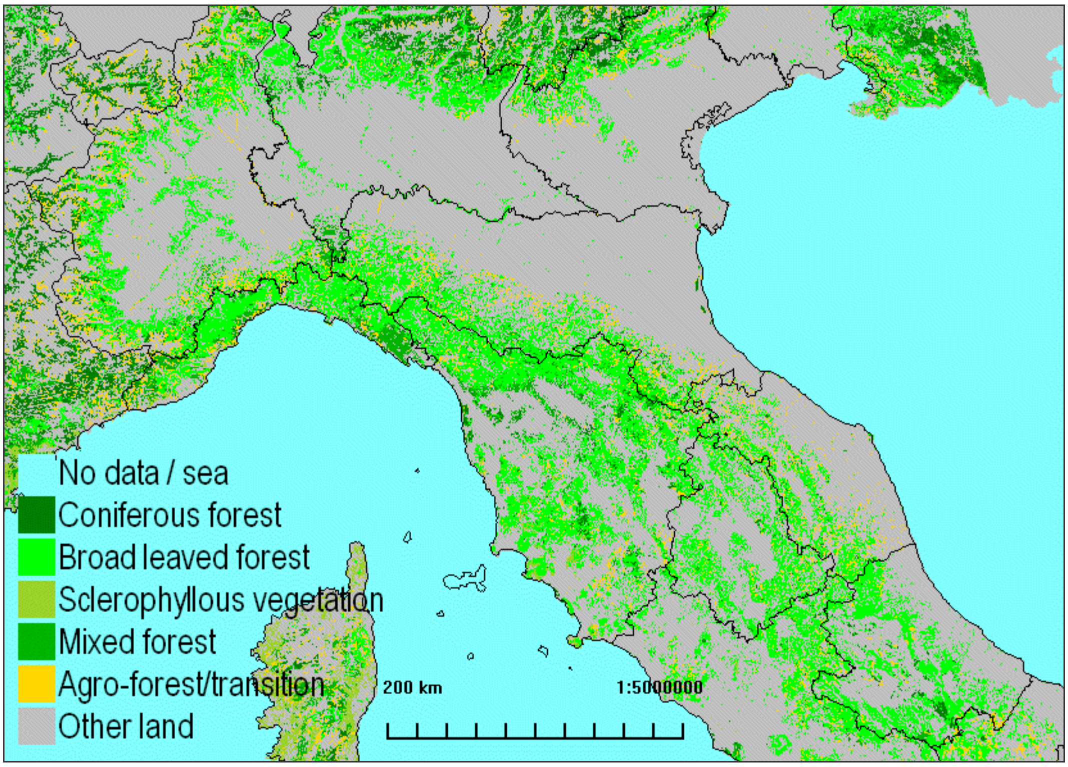

Figure 4.4 CLC image for the area of interest, after re-classification to forest map.

CORINE		FMERS	
LC class:	**Description:**	**Number**	**Description**
0	Not inventoried	0	No data
2.4.4	Agro-forest areas	6	OWL Broadleaved
3.1.1	Broad Leaved Forest	2	Broad Leaved Deciduous
3.1.2	Coniferous Forest	1	Coniferous
3.1.3	Mixed Forest	4	Mixed
3.2.3	Sclerophyllous Vegetaion	3	Broadleaved Evergreen
3.2.4	Transition woodland-scrub	6	OWL Broadleaved
	Not defined	5	OWL Coniferous

Table 4.1 Matching CORINE and FMERS forest cover classes for the current study.

4.3.2.3 Comparison of derived forest maps

The definition of forest can have a large influence on the types and degree of fragmentation detected in any survey (Riitters *et al* 2000). It is thus no trivial task to select and possibly re-classify the themes that define forest, for studies of forest structure like this, where it is a central task to compare forest maps derived from satellite imagery with land cover maps made for other purposes[16].

The two data sets are both satellite based and have more or less the same *thematic* resolution. It is however worth noticing that a partly manually delineated land cover databases like the CORINE have a very low *temporal* resolution, compared to maps based on spectral classification algorithms, which can now be updated more or less automatically. The land surface covered by the selection is approximately 195,150 km^2, of which 34.4% is forest according to the CLC classification and 37% according to the FMERS classification. The distribution of the separate classes is shown in Table 4.2.

CLC				FMERS			
LC type	pixels	percentage of forest areas	percentage of land area	LC type	pixels	percentage of forest areas	percentage of land area
0	28314165	N/A	N/A	0	6944905	N/A	N/A
1	848669	12.69%	4.35%	1	482593	26.74%	9.89%
2	3792008	56.72%	19.43%	2	574189	31.81%	11.77%
3	274834	4.11%	1.41%	3	7266	0.40%	0.15%
4	911814	13.64%	4.67%	4	296940	16.45%	6.09%
5	0	0.00%	0.00%	5	87400	4.84%	1.79%
6	858510	12.84%	4.40%	6	356707	19.76%	7.31%
TOTAL FOREST	**6685835**	**100.00%**	**34.26%**		**1805095**	**100.00%**	**37.00%**

Table 4.2 Distribution of land cover classes in the two data sets.

A direct comparison of the two data sets is done using a "confusion matrix" at per-pixel level for similar pixel sizes and calculating the Kappa statistics (Congalton and Green 1999, p 45 ff). In order to compare the input data from CLC and FMERS pixel-to-pixel, the CLC image was degraded to 200m pixel size, by assigning the dominant land cover type in a 2*2 pixel window to the pixel in the output window. Table 4.3 shows the resulting co-occurrence-matrix, on which the Kappa statistics is based.

[16] As in this case the CLC database that has been made for environmental assessment and management in general.

	0	1	2	3	4	6	**Total**
0	6420421	67101	297683	25606	53365	80729	**6944905**
1	186382	84144	104595	15653	61153	30666	**482593**
2	198056	20476	285787	2267	43671	23932	**574189**
3	4350	332	756	1230	340	258	**7266**
4	119442	13337	114982	10132	24782	14265	**296940**
5	47552	10360	14137	2435	5611	7305	**87400**
6	219627	14461	79456	7255	16000	19908	**356707**
Total	**7195830**	**210211**	**897396**	**64578**	**204922**	**177063**	**8750000**

Table 4.3 Co-occurrence of pixel values in FMERS and CORINE land cover maps. The CORINE data were re-sampled to 200m pixel size. CORINE data are in columns and FMERS data in rows.

The Kappa coefficient was calculated using IDRISI and used as accuracy measure. It assumes acceptable values for categories 1 and 2, coniferous and broadleaved evergreen, which are also the most common forest types in the images. These land cover types also have the lowest error coefficients. The overall Kappa for this comparison is as low as 0.095. It must however be noted that when comparing forest-non-forest maps from the two image types, as illustrated in Figure 4.5, below, the overall Kappa increases to 0.5218. This, along with visual inspection of the maps, clearly shows that a pixel-to-pixel comparison is not possible or meaningless, and instead we have to test whether the spatial metrics at different cell sizes are appropriate.

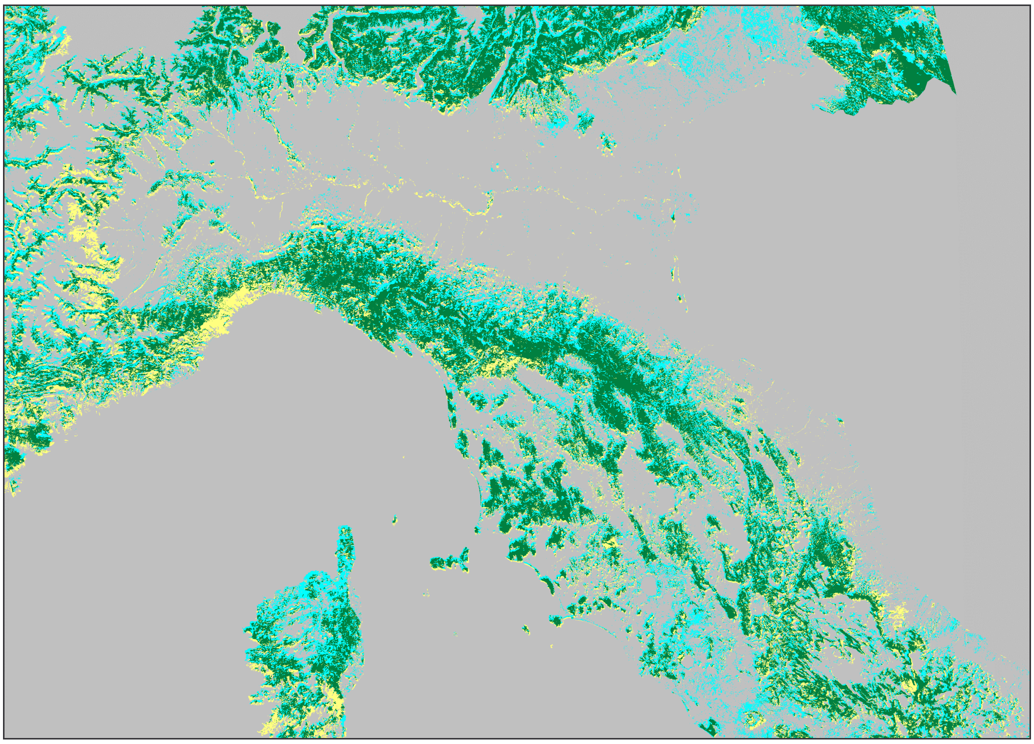

Figure 4.5 Cross-tabulated image from CORINE and FMERS forest masks: background pixels are grey, pixels only in the FMERS map are blue, pixels only in the CORINE map are yellow and pixels in both forest maps are green.

The preparation of the forest-masks for parameter extraction provided an interesting insight in the structure of the CLC and FMERS data, as illustrated in Figure 4.7, where the CLC data maps more coherent riparian forest, a feature that is typically hard to separate and eventually is 'lost' in solely spectral classifications like the one performed in the production of the FMERS map.

4.3.2.4 Digital elevation model

The digital elevation model (DEM) used here is based on the data set that was assembled and used for development of a pan-European database of rivers, lakes and catchments (Vogt *et al* 2002). The current DEM is an 8-bit version of the file that was used for deriving the river-network for Italy in the initial part of the project, this means that the altitude resolution is 20m and the grid cell size is 250m. The DEM is shown with a typical colour legend in Figure 4.6 below. For use with the different output maps of spatial metrics, the DEM was re-sampled to cell sizes to the images, using the image-rectification routine of WinChips (bi-linear interpolation), with resulting average elevation values.

4.3.3 Vector data

Ancillary vector data were used to extract information from the metrics images, using the statistical functionalities of WinChips. This was done in order to summarise and evaluate the evenly distributed (gridded) metrics values. The GIS data used include the watersheds from level 1 to 6 for Italy from the project described above, their shape and extent is shown in Figure 4.6 below. A subset of catchments were extracted for the upper Po valley and for the entire Tevere (the Tiber) catchment, supplemented with two 4th order catchments in Toscana. A set of polygon layers with the NUTS (Nomenclature of Territorial Units for Statistics) administrative regions were also used, they were made available from Eurostat[17], in the Corine projection. From this database, the Italian regions ('regioni' = NUTS-level 2) were extracted and used for derivation of average metrics values within these. The CLC dataset with 100m pixels, together with the NUTS-coverage were used to make a base-map showing land surfaces and excluding only open sea. This base-map has been re-sampled to various pixel sizes, and these derived maps have been used as background image for illustrative purposes throughout the project.

[17] Description and interactive maps at: http://europa.eu.int/comm/eurostat/ramon/nuts/home_regions_en.html (accessed 28/12 2003)

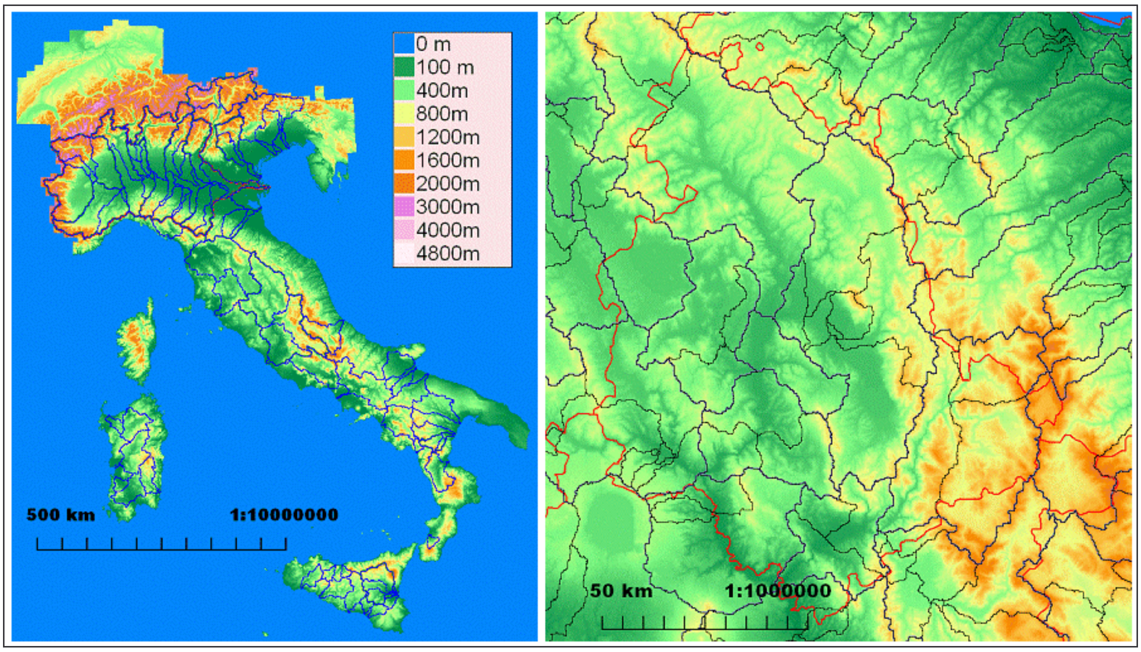

Figure 4.6 Digital Elevation Model of Italy. To the left full extent with 4st to 6rd order catchments – there is just one 6th order catchment: the Po river basin with tributaries. To the right a subset with the Umbria region (borders as red lines), overlaid by 2nd and 3rd order catchments, extent 140*150 km.

4.4 Methods

In this section, first the intended outputs in terms of spatial metrics are listed and discussed, then the practical image processing and statistical approaches for how to derived them from the input data set are presented.

4.4.1 Selection and definition of spatial metrics

The metrics selected for this study are the same as in previous chapter, supplemented by metrics of cover proportion and diversity. The types of structural metrics calculated are:

- cover (percentage), total forest and for each class
- percentage of edge pixels, of total number of pixels in window
- a simple edge index: proportion of edge pixels to number of pixels in actual class
- the Matheron (M) index, for each class and for combined forest layers
- the Square Patch index (SqP) – for forest-non forest
- Patches Per Unit area (PPU), both following Frohn's definition and a modified, 'normalised' version that accounts for changing window sizes.

These last three are described in section 2.3.4. The edge pixel percentages and proportions area used here only as intermediate steps to get to the M and SqP metrics and for development and testing purposes, though they have the potential to be used as indicators in their own right.

The diversity metrics used are:

- Number of class types (richness)
- Simpson's diversity
- Shannon's diversity (Entropy)

The richness metric is the simplest possible measure of diversity, and has the advantage of being easily understood and easily implemented. Simpson's diversity SIDI, which expresses the degree to which one or more classes *dominate,* is defined as follow (McGarigal and Marks 1995):

$$SIDI = \sum_{i=1}^{n} P_i^2$$

Where P_i expresses the proportion of the entire landscape occupied by class i, the different values of P_i should sum to 1 for each landscape/subset. In this study 1-SIDI is used for reporting the metrics values, in order to have the highest values for the smallest amount of dominance, i.e. for the landscapes with largest *evenness* between classes. Then we have <u>maximum</u> value of $SIDI_{max}$ for P1=P2=….Pn=1/n., and
$SIDI_{max}$ = 1-1/n .

Shannon's diversity index, also known as the Shannon-Weaver or Shannon-Wiener information index (Whittaker 1972), is based on information theory, expresses the 'bandwidth' needed for description of a system and thus the 'disorder' or distance from predictability of it (McGarigal and Marks 1995). The index defined as:

$$SHDI = -\sum_{i=1}^{n} \left(P_i * \ln P_i \right)$$

The maximum value of the SHDI for a landscape with n classes is simply ln(n), and the minimum values is 0 for the case when the landscape contains only one patch type (no diversity). These two diversity metrics are very commonly used in the ecological literature, and thus it is found to be of interest to look closer into their behaviour with changing pixel- and window size.

In this study, it was chosen <u>not</u> to include the pixels that represent background in the diversity calculations, since the phenomenon under study is the structure of the forests and the diversity of the forest types. Including background pixels would give a measure of *landscape diversity* rather than *forest diversity,* and then it could be argued that the aggregation (see section 2.3.3.3) should not have taken place, and the various natural and agricultural land cover types preserved as separate classes. This issue is addressed in the following chapter, when CLC and high-resolution land use data are used, compared and discussed in more detail. Thus, as part of the preparation of the images, they were processed so that only the forest classes of interest were preserved, and any other class set to zero (i.e. constituting the 'background class'), as seen in Figure 4.2 and Figure 4.4.

Concerning the structural metric Patches per unit area (PPU), based on the count of number of patches in the window, there is a problem of bias towards higher values for small window sizes, since if any part of a larger coherent forest is present in the window, one patch will already be counted there. In other words, the sampling method acts like a "cookie cutter" (O'Neill *et al* 1996, p. 174). For instance, if 10*10 km of continuous forest cover is analysed with 1*1 km windows it will result in 100 output cells with 1 patch per km^2, and from a 10*10 km window, the result will be one output cell with 0.01 patch per km^2. The present study investigate whether it is possible to remove – completely or partly - this effect of window size, especially for densely forested areas (where a low number of separate patches can be expected). This is done with PPU-Normalised (PPUN), defined as:

$$PPUN = \frac{NP - 1}{A} + \frac{1}{A_{min}}$$

Where A_{min} is the area of the smallest window used in the current analysis. The last part of the expression is included in order to avoid having the values of PPUN approach zero for large windows, thus PPUN will be one for the case of just one patch present at all sizes, with values approaching one for larger window sizes with more patches present – as examplified in Table 4.4. After inspection of the results from the first tentative runs of the patch-counting script, it was chosen also to include the number of 'background patches' as a spatial metric, for the reason that a patch of non-forest surrounded by forest is an expression of fragmentation and perforation of the forest cover in the area/window of interest. It is similar to but much simpler than metrics of lacunarity (Plotnick *et al* 1993). The PPUN_B value, as it will hereafter be called, is easily derived, as the patch counting script anyway will deliver the number of patches in the window of analysis for each land-cover class in the input image. It is calculated in the same way as the PPUN metric.

Area	No. Of patches	PPU	PPUN	No. Of patches	PPU	PPUN
1	1	1	1	5	5	5
10	1	0.1	1	5	0.5	1.4
100	1	0.01	1	5	0.05	1.04
1000	1	0.001	1	5	0.005	1.004
1	2	2	2	10	10	10
10	2	0.2	1.1	10	1	1.9
100	2	0.02	1.01	10	0.1	1.09
1000	2	0.002	1.001	10	0.01	1.009

Table 4.4 Theoretical values of PPU and PPUN for varying window sizes and number of patches.

For the regression analysis performed in order to find the agreement between the different metrics, the 'original' patch count metrics are used, i.e. the NP values from the M-W results. This is possible due to the nature of the transformations from NP to PPUN and PPUN_B, and because the regressions only take place for one window size at a time, and as such not are affected by the transformations.

When average values of spatial metrics are reported from the different output (image) files, only those output cell which represent forest cover of one per cent or more are included, the others are masked out. When values for the two different data sources are compared, the criterion for inclusion is that one of the results should represent a window with a forest cover of one per cent or more. In practical terms, this is done through constructing of a binary forest cover map, using the arithmetic functionality of WinChips. Such non-forest cells are typically found in river basins with intense agricultural activity and to a lesser extent in mountain areas above the tree line. This means that *three types of forest mask* are applied: one for each of the map types and one for analyses where they are combined or compared – in this case the "OR" image from the right hand side of Figure 4.7 below. A consequence of this masking approach is that the average forest cover values reported for entire images and selected regions will be higher than the actual forest cover as percentage of the entire land area, since output cells with no or very little forest are excluded.

The preparation of the forest-masks for parameter extraction provided an interesting insight in the structure of the CLC and FMERS data, as illustrated in Figure 4.7, where the CLC data has more and coherent riparian forest, a feature that is typically hard to separate and eventually lost in solely spectral classifications like for the FMERS map.

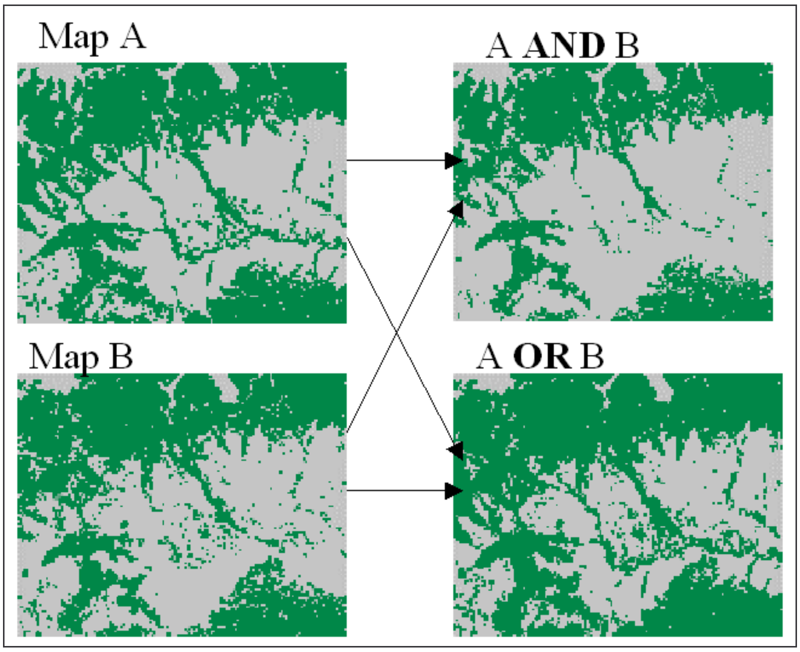

Figure 4.7 An example of how maps of forest presence are combined for masking in extraction of statistical parameters. The subset used here is the upper part of the Po river basin, for a cell size of 1200m. Map A is based on CLC and Map B on FMERS data. Note that the agreement between these two data sets improves as the cells become larger (and there are fewer cells with no forest), see for instance Table 4.19, page 154, row 'Cover'.

4.4.2 Implementation of Moving Windows and analysis of outputs

The 'moving window' calculations were carried out using IDL scripts (Research Systems Inc. 1999), that allow modification of the window and the step size, as part of which overlap between windows is possible. The principles of M-W analyses as implemented here and the basic terms referred to throughout the text are illustrated in Figure 4.8 below. The main difference between this implementation and the one used in for instance Fragstats for Windows is that here, the user can define not only the extent (size) of the window, but also the step and thus the output cell size which determines the grain size of the output image. These window sizes and steps are implemented as parameters of for-next loops that operate on image-matrices in the various IDL-scripts used here (Appendix 1).

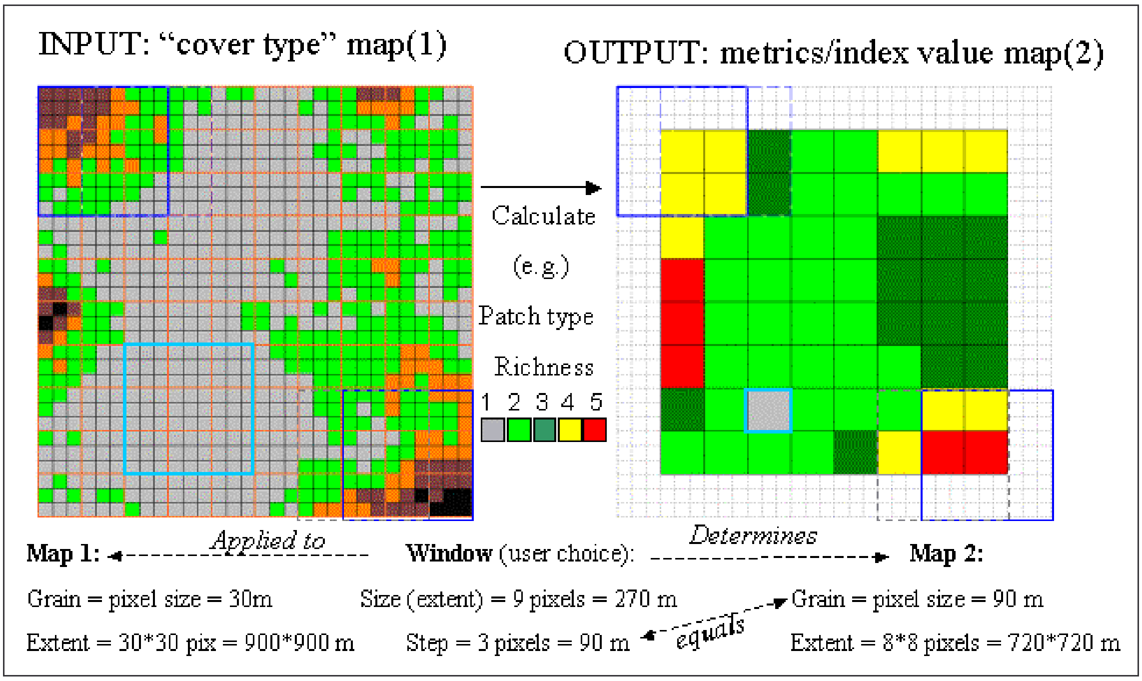

Figure 4.8 Moving windows concepts with and without overlap.

As part of the processing chain, which is only partly automated, simple spatial statistics such as cover proportion and number of land-cover classes are calculated. This information can also be used for inspection of the input data and visualisation of basic landscape properties (see for instance Figure 4.10 on page 134). The diversity metrics are based on histograms of pixel values collected for each window, the fragmentation metrics are based on per-window counts of edges, both for each land-cover class and between forest and non-forest pixels. As an aid for the further presentation of the types of calculations and files used in this study, a sketch of the way from input data to the various types of results has been made, Figure 4.9.

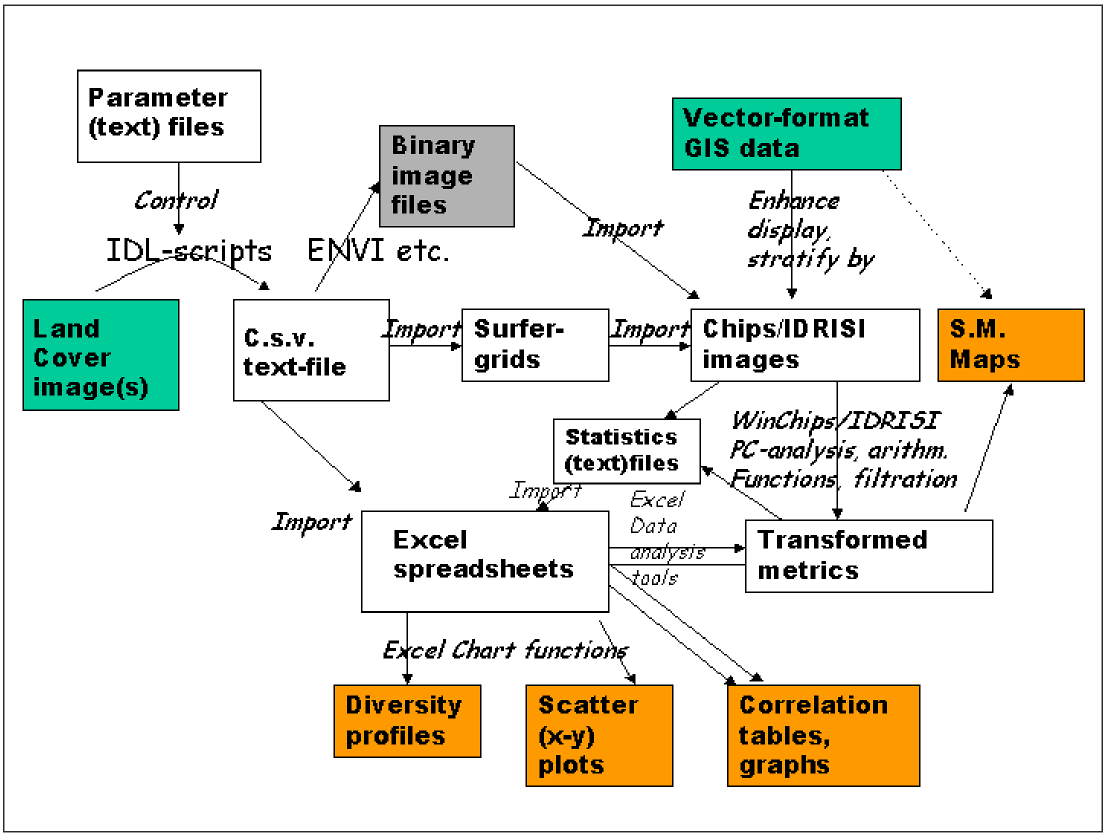

Figure 4.9 Simplified flowchart showing how the results presented below are derived. The boxes represent final or temporal data (to be) stored as files.

The output images are easily geo-referenced. The coordinates for the upper left corner of the output images depend on the parameter for the MW-calculations, in the following way:

Pixel size = step

$UL\_X_{out} = UL\_X_{in} + ((size\text{-}step)/2)$

$UL\_Y_{out} = UL\_Y_{in} - ((size\text{-}step)/2)$

The number of pixels in the output image is determined by the equations:

$Outcolumns = INTEGER\left(\dfrac{incolumns - wsize + wstep}{wstep}\right)$ for the columns (X-size) and

$Outrows = INTEGER\left(\dfrac{inrows - wsize + wstep}{wstep}\right)$ for the rows (Y-size).

These numbers are needed for correct import and geo-referencing of the resulting maps, so they for instance can be used with vector data in a GIS. In this implementation of the method, this is achieved by importing the outputs (text files) into the image processing software and assigning the correct pixel sizes and edge coordinates. All image data have 'UTM-style' lower-left coordinate systems, in the present case the CORINE-projection is used, since the CLC maps serve as reference for the entire dataset.

IDL Scripts used include (scripts are listed in Appendix 1):

a) Calculation of Cover, Diversity and Fragmentations metrics

b) Counting of patches, where each land cover class is processed separately.

c) Degradation of images, either

   - binary (forest-non-forest maps), possibly with variable threshold values, in order to keep the same cover percentage as in the input image

   - aggregation with possible weighting for land cover classes of different interest/ "value"[18]

For each dataset spatial metrics were calculated for window sizes ranging from 1200m to 19200m, corresponding to windows of 6*6 to 96*96 (9216) pixels for the FMERS map and of 12*12 to 192*192 (36864) pixels for the CLC based forest map. Further on, the two data types are compared at window level, i.e. between output cells representing the groups of pixels, that cover the same part of the forested landscape. This is done by finding the correlation coefficient for the two variables or M_{CLC} and M_{FMERS} (or in standard terms y_i and y_j) representing the spatial metrics from the two data sources. The number of observations n is the number of windows/output cells where forest is present – with the criterion that at least one of the land cover images should have a forest cover of one percent.

For a comparison of the M-W results for the entire maps with results from previously defined regions that form subsets of the test area, vector data were used to extract metrics values for catchments as well as administrative region (through the creation of WinChips statistics files, see Figure 4.9). The spatial metrics values are reported at regional level (highest level of

[18] Note that simple averaging of pixel values, as applied to photos or satellite images will not work on categorical maps.

Italian administrative regions) and for the catchments of highest orders, i.e. of largest spatial extent

4.4.2.1 Tests for significance of results

As test for the significance of the correlations, a simple 'rule of thumb' is used, namely that for large values of *n*, the minimum (absolute) value needed to attain significance is defined as (from Rogerson 2001, p. 94):

$$r_{crit} = 2/\sqrt{n}$$

.. when α=0.05. For this type of analysis a "combined forest mask" is used, where the criterion for a pixel to be included is that forest cover is > 1% in either the CLC *or* the FMERS forest map – based on the cover value calculated at the given window size. These combined, all inclusive forest masks are also used for extraction of (average) metrics values for administrative regions and watersheds.

As an alternative to the pixel-to-pixel approach described above, and in order to test whether the two different data sources give the same *general picture* of regional forest structure, the areas (admin. regions and catchments) are ranked according to the average values of each metric and compared using Spearman's Rank Correlation Coefficient (Rogerson 2001, pp 94-95). The results from these tests contribute to understanding which metrics are sufficiently robust to be used with different image sources and over large areas. The ranking approach also helps illustrate in which geographical areas or zones agreement of metrics values are found, and in which areas the differences are – and whether these 'problematic' areas are similar or different for metrics assessed in this study.

4.4.3 Local variance and autocorrelation

The concepts of variability and autocorrelation are of interest because they describe not only the structure (clustered or scattered landscape elements and derived spatial metrics) but also the information content in the output ‘maps’. For the current study focusing on mapping of spatial structure and diversity, it is assumed that higher local variability means that more information is present in the outputs (refer section 2.3.3). This information is potentially used for display of landscape properties and ultimately prediction of biological diversity. The spatial variances of the resulting ‘spatial metric maps’ are calculated in two ways: local standard deviation and autocorrelation expressed through Moran’s I.

4.4.3.1 Local variance approach

The approach to find the local variance follows the methodology described by Woodcock and Strahler (1987), for assessment of characteristic scales in remotely sensed images, insofar as the extraction of spatial metrics can be seen as applying a low-pass filter, in the same way that remote sensing imagery is degraded to lower resolutions, ref. Wu *et al* (2000).

Here, the local standard deviation (stdv.) of the metrics values is found under a mask defined by (percentage of forest cover => 1), with edge pixels excluded, as these are set to zero values during calculation of variance (as during filter operations in general).

The steps in the creation of variance statistics at each extent are:

- Create mask from pixels with cover > = 1% AND not along edges
- Calculate stdv. of pixel values in 3*3 window around each pixels
- Calculate mean and max. value of stdv. from under the mask
- Calculate coefficient of variance, based on average and st.dev.values for each metric and data source, this gives a nicely normalised expression of the local variance of the metric.

The results are reported in table and graphical form.

4.4.3.2 Autocorrelation approach

The Moran's I (MI) measure of spatial autocorrelation is derived using Idrisi (Eastman 1997), where it is implemented as a statistical function. It is defined as follows:

$$MI = \frac{n\sum_{i}^{n}\sum_{j}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{(n\sum_{i}^{n}\sum_{j}^{n} w_{ij})\sum_{i}^{n}(y_i - \bar{y})^2}$$

where n are the number of regions/pixels/windows, W_{ij} is a measure of proximity and y_i and y_j are the metrics from the different data sets. Similar to a correlation coefficient, MI assumes values from –1 to 1, where values near 1 indicate a strong spatial pattern (high values near each other, low values near each other), values around 0 indicate no particular pattern (random distribution) and values near –1 indicate a case where high values are located near low values (this is rarely seen and geographical data normally never have values of MI below 0). MI can also be seen as a simple measure of self-similarity or the potential of using cell values to predict the value of neighbouring cells in raster images (Costanza and Maxwell 1994).

4.4.4 Masking and Forest Concentration

The work with image masks at different output cell sizes have led to proposing a new spatial metric particularly for use with MW methods: a measure of forest concentration (FC) or landscape concentration. It stems from the observation of characteristic values in selected regions of the forest cover percentage for respectively the entire landscape and under the 'forest presence' mask. When the value under the forest mask is high relatively to the entire landscape it means that the forest is concentrated in a limited number of output cells, whereas when the two values are nearly similar the forest cover must be spread out over the image/region of interest. The metric is defined:

$$FC = \frac{Cover\_mask}{Cover\_landscape} - 1$$

The theoretical values range from 0 when the two cover metrics are similar (the forest presence mask covers the entire region) to near infinite, depending on the size of the output cells relative to the output image. For the same input image the values of FC will decrease with increasing output cell size, as the chance of finding windows with no forest will decrease, but also the shape of the resulting FC-profiles might provide additional information on the structure of forest (or other element of interest) in the region. To derive a FC-profile, MW analysis with a number of different window sizes is required.

4.5 Results

The results of image processing and subsequent statistical analysis are presented along the lines laid out in the objectives of this chapter. This section thus begins with a presentation of and some comments on the values of the metrics *per se* and in relation to window sizes, then the spatial structure of the output 'maps' are looked at, followed by examination of the regressions between pairs of metrics from the two data sources for a range of window sizes. After that values of metrics from different spatial units are compared (administrative regions vs. river catchments) and finally, the visual appearance of the metrics (maps), interpretation and applications for statistical reporting and use as indicators are discussed.

Figure 4.10, below shows an example of an immediate result of the application of M-Ws to the two data sets, where the resulting text files have been imported and visualised as grids using the Surfer software (Keckler 1997) , following the flow outlined in Figure 4.9. Already this visualisation of a relatively simple metric, the number of land cover classes gives the impression of not only where forest is found but also where environmental conditions allow several different forest types to be found within a limited geographical area, in this case within squares of 4.8*4.8 kilometres. The apparent broad agreement between the outputs for the two different map data sources is tested statistically in section 4.5.4.

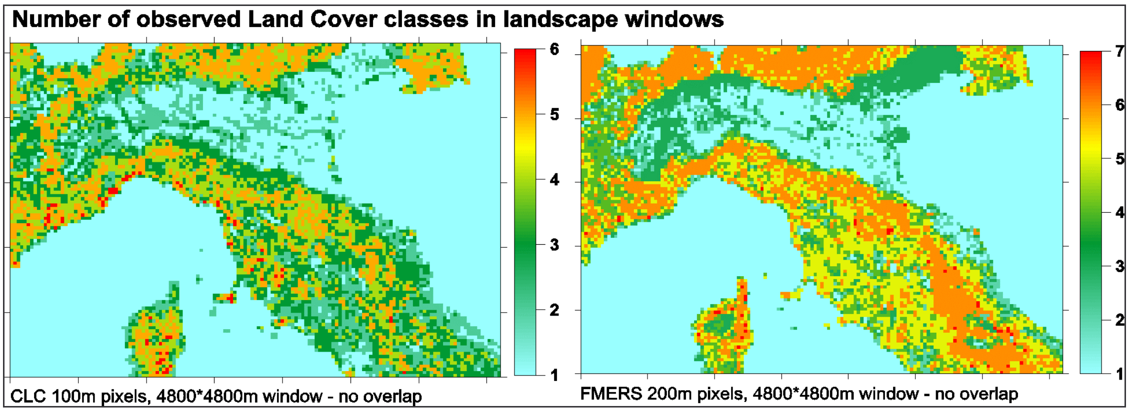

Figure 4.10 Land cover "richness", i.e. count of different land cover types present within windows of 23km^2 , figure created in Surfer for windows, using text file outputs from IDL script processing of input images.

4.5.1 Response of metrics to window size

For each output map of the specific spatial metric for each of the datasets the average value was calculated – though only for cells/pixels with a forest cover fraction >= 1%. In Figure 4.11, these values are plotted against the size of the moving window. In order to make the metrics of forest cover fit in the graph, they have been divided by 100, resulting in fraction values between 0 and 1. The percentage values of forest cover in the windows are listed in Table 4.5.

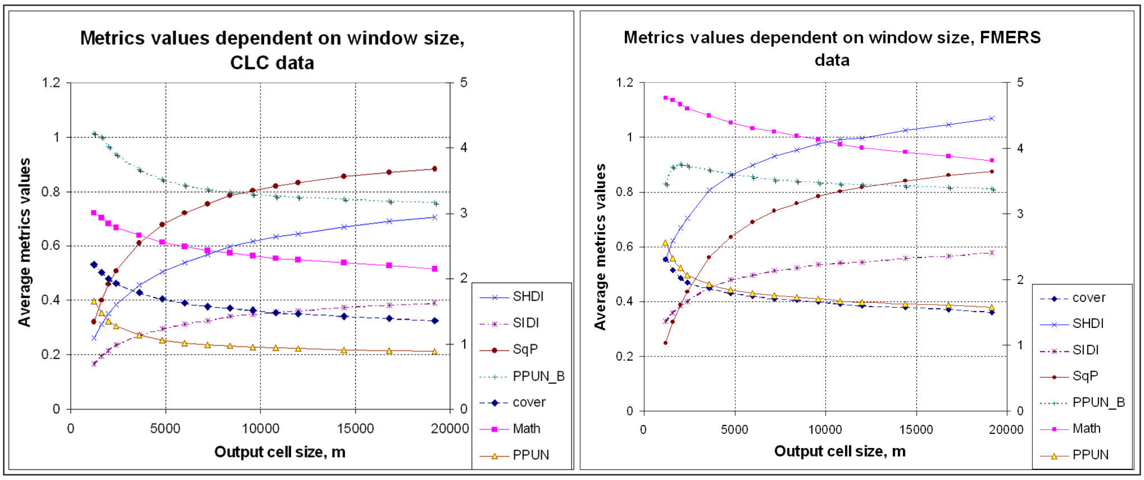

Figure 4.11 Metrics 'response curves' or scalograms with values plotted against window size or (sub-landscape) extent. CLC and FMERS data for the entire study area (under the forest masks). Note that M and PPUN metrics map to 2nd axis values.

When these graphic outputs are compared, it is obvious that the metrics behave very much in the same way for the two datasets, for the shape as well as the relative position of the curves. Thus, they show similar *scaling properties*. The almost complete overlap of the PPUN and cover curves for the FMERS data is accidental, but clearly shows the relation between these two metrics. It is noteworthy though that for the CLC data, the PPUN values are markedly lower – but not the PPUN_B values. As expected, the value of the diversity metrics increase with window size, as more land cover classes get included in each window.

The most noteworthy differences are observed for the SqP metric, that starts out at a lower level for the FMERS data and grows more rapidly than for the CLC data with increasing window size. This is probably due to the fact that the small window sizes correspond to very few pixels, where the probability of finding 'blocks' of forest is much higher, while on the other hand large windows will include a mixture of forest and non-forest. The same phenomenon is reflected in the decrease of the average forest cover with window size. Note that due to the definition of the metric, high values of SqP (approaching 1) indicate forest that is more scattered/fragmented across the landscape, i.e. distributed on a number of patches. The higher values for the SqP metric from CLC relative to the values from FMERS data is in agreement with the observation in section 3.3.1 where synthesised images were analysed, and SqP found to decrease with increasing pixel size (for a fixed size of the spatial window, thus representing the same "ground truth" = forest structure across the scalogram). Figure 4.12 shows that the numeric values of SqP is more closely related to the size of the geographical window than to the number of pixels included in the calculations. This is a reassuring result, and speaks in favour of using this metric as an indicator of forest structure, given that a correction for window size can be applied.

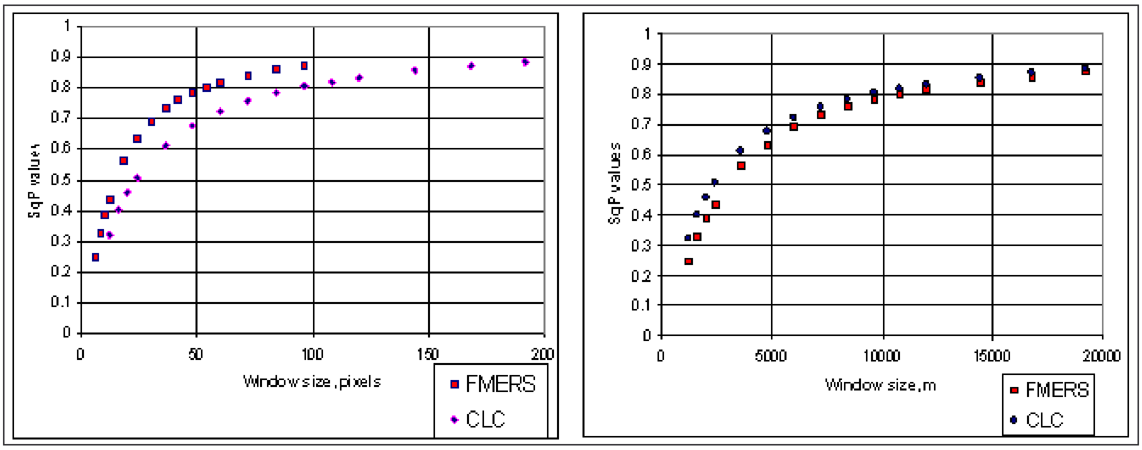

Figure 4.12 Average values of the SqP metric for the two data types plotted against window size in pixels resp. meters

4.5.1.1 Patch counting and the PPU metric

Values of patch count metric values are influenced by the size of the windows, due to the effects of "cutting off" of patches that are partly within the window, as seen from the plots of average PPU versus window size. Thus, the smaller the window, the greater the number of separate patches, which are parts of larger patches, with centre outside or on the edge of the window. This effect will also influence the values of calculated average patch sizes (a metric that only makes sense for entire landscapes or vary large windows). Another notable effect is that as the window size increases, more non-forest area is included, as seen from Table 4.5 (last column). This is due to the non-random (i.e. clumped) distribution of forest across the landscape. It is hard to separate these two effects, and caution must be taken when metrics based on the number of patches in a given area are used, especially at small (< 30-40 pixels) window sizes. In Figure 4.13 and Figure 4.14 the values of PPUN and PPUN_B are plotted against window size and forest cover fraction respectively.

cell size, m	side factor	area factor	PPUN_CLC	PPUN_FMERS	Mean_cover_percent CLC image
1200	1	1	2.57	1.66	47.33
1600	1.333	1.778	2.33	1.48	45.14
2000	1.667	2.778	2.18	1.35	43.34
2400	2	4	2.08	1.27	42.10
3600	3	9	1.94	1.13	40.23
4800	4	16	1.85	1.06	38.77
6000	5	25	1.80	1.02	37.79
7200	6	36	1.76	0.99	37.01
8400	7	49	1.73	0.97	35.90
9600	8	64	1.71	0.95	36.07
10800	9	81	1.68	0.94	35.27
12000	10	100	1.66	0.93	34.99
14400	12	144	1.63	0.91	34.10
16800	14	196	1.61	0.90	33.36
19200	16	256	1.58	0.89	32.80

Table 4.5 Patch count values from different window sizes, the unit of the PPU metric is no. of patches per square km. The shaded rows of the table indicate values calculated using WinChips, the remaining ones are calculated in Excel.

The agreement in the shape of the PPU-curves between the two data sources seen in Figure 4.13 indicates that their behaviour is an inherent effect of the way in which the metrics are calculated – as much as of the spatial distribution of forest on the Italian peninsula! Here it would be very relevant to test on data from a neutral model, as done by Saura and Martinez-Millán (2001). These authors also described the sensitivity of spatial metrics values to window size, and found that for their artificial data, measures of patch density increased with window size. Such tests were carried out early in this project, but with no conclusive results, and have since been determined to be outside the scope of this study.

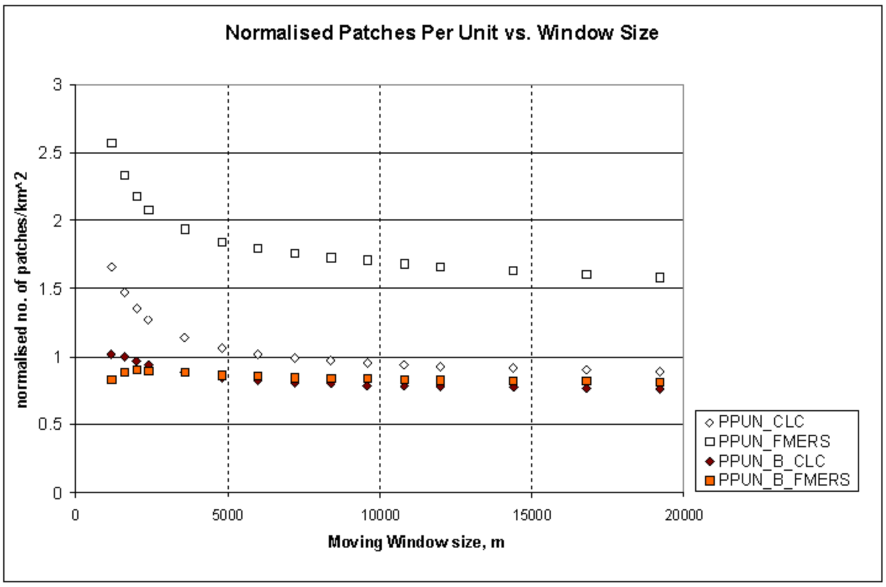

Figure 4.13 Average patch density plotted against window size, CLC and FMERS. Note the different shapes of curves for forest respectively background patch densities.

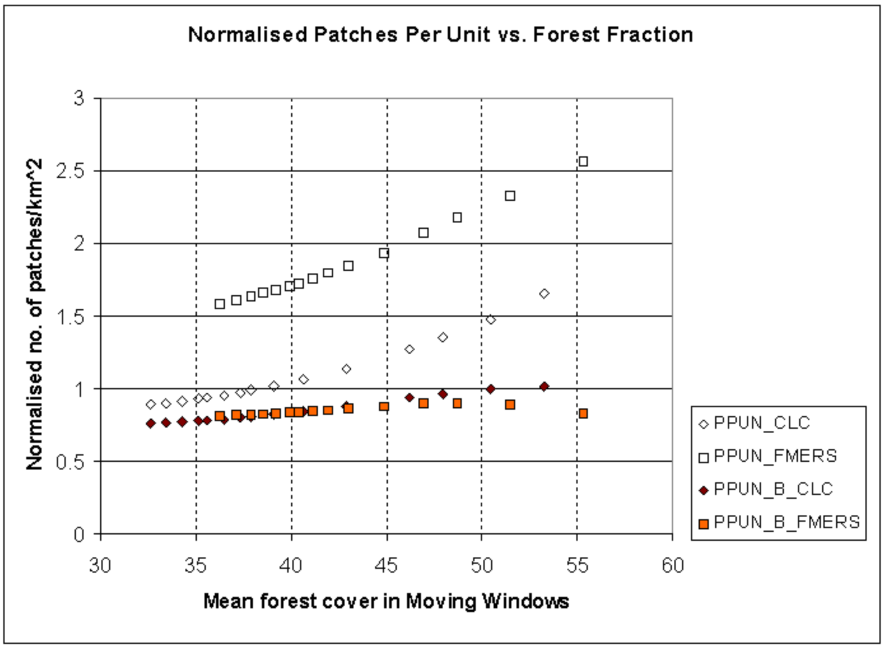

Figure 4.14 Average patch density plotted against the average forest cover, for CLC and FMERS data in the respectively included windows/output cells. Note that the right hand side of the curve, with largest forest cover values represents the smallest window sizes (compare Table 4.5).

By visual inspection of the maps produced and comparison with the input data, it appeared that the number of patches in the "background" class, i.e. all non-forest pixels was a good indicator of one aspect of forest fragmentation, namely the perforation or lacunarity of the forest landscape. Maps of the number of "background patches" at window sizes ranging from 1200 and 4800m, from the CLC and FMERS data is shown in Figure 4.15.

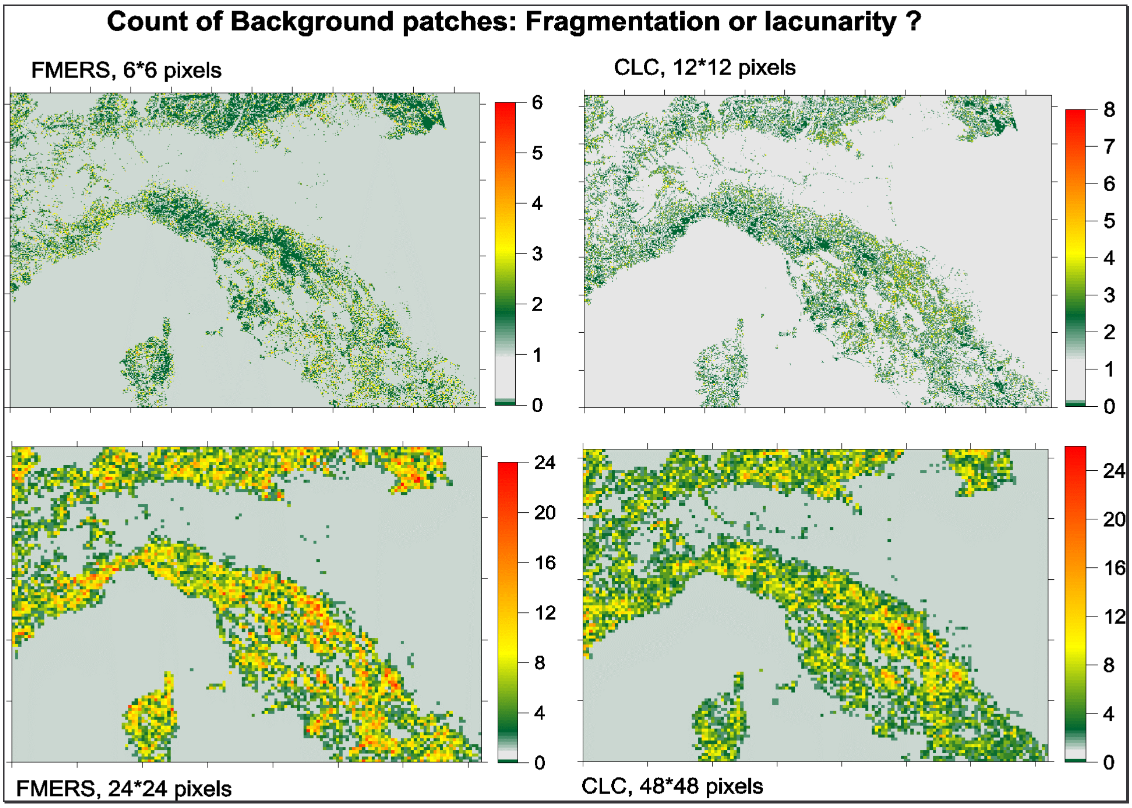

Figure 4.15 Background patch count applied as possible fragmentation metric. Derived maps show the number of separate patches belonging to the "background" class in FMERS and CLC images respectively. These maps are outputs from the Surfer software, where the text-files from the IDL calculations are converted to grids.

4.5.2 Variability and autocorrelation of the metrics

The local standard deviation was calculated for the full series of metrics images, using a WinChips filtering function (Hansen 2000) and the results extracted as a statistics file. In Figure 4.16 and Figure 4.17 the local variation of two metrics: forest proportion and Shannon's diversity are plotted against the size of the moving windows. The first observation from the figures is that the variation behaves in a similar way for the two different data types. In both cases the variability in cover fraction initially falls with increasing window size, then stabilises or increases, indicating that for CLC data there is a characteristic size of forest areas between 15 and 20 km where a slight maximum is observed, for FMERS data there is possibly a maximum above 20 km. For the CLC data there is a slight increase in the variability of the SHDI diversity metric, which is not found for the FMERS data. This is in contrast to the increase in the absolute values of this metrics seen in Figure 4.11. Similar behaviour is seen for the SqP metric, where the standard deviation decreases in spite of an increase in the absolute values of this metric with window size. For both data types, the variance of the PPUN metric decreases in the same way as the absolute values. The difference of the response curves for forest cover and diversity metrics indicate that these properties have different spatial domains/characteristic distances. This is theoretically possible, and could for instance owe to changes of composition within forested areas following altitude variations – this possibility is evaluated in the chapter 5.

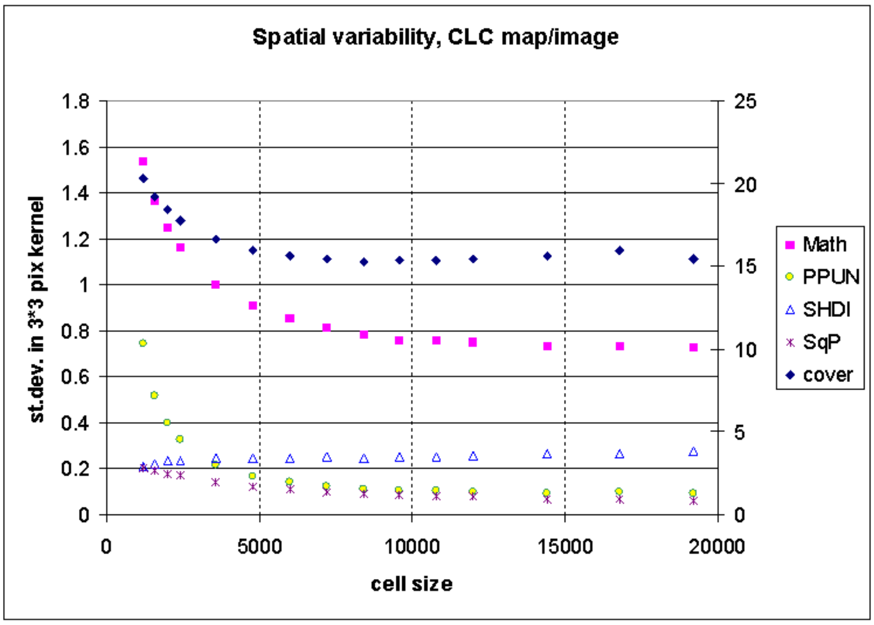

Figure 4.16 Standard deviation of the values in output cells for CLC data, calculated in 3*3 cell windows and averaged over the non-empty parts of the image. Note that the 'cover' (percentage) values map to the 2nd y-axis.

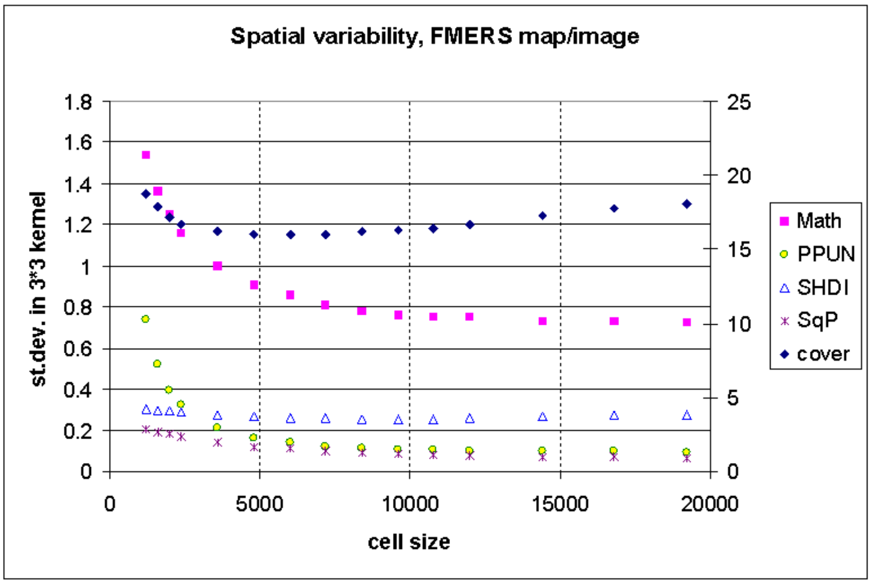

Figure 4.17 Standard deviation of the values in output cells, for FMERS data the curves of both forest cover and SHDI show a distinct minimum. Note that the cover values map to the 2nd y.axis.

When the coefficient of variance is calculated for each metric and displayed as function of the window size, it becomes clear that the different metrics show different responses to change of scale, see Figure 4.18 and Figure 4.19, below. The 'peak' in the variance of the forest cover for the CLC data is still visible, and also the Matheron metric of fragmentation increases after having its minimum average values around a window size of 10 km, most clearly for the CLC data but also visible for FMERS. All other metrics have steadily decreasing coefficient of variation.

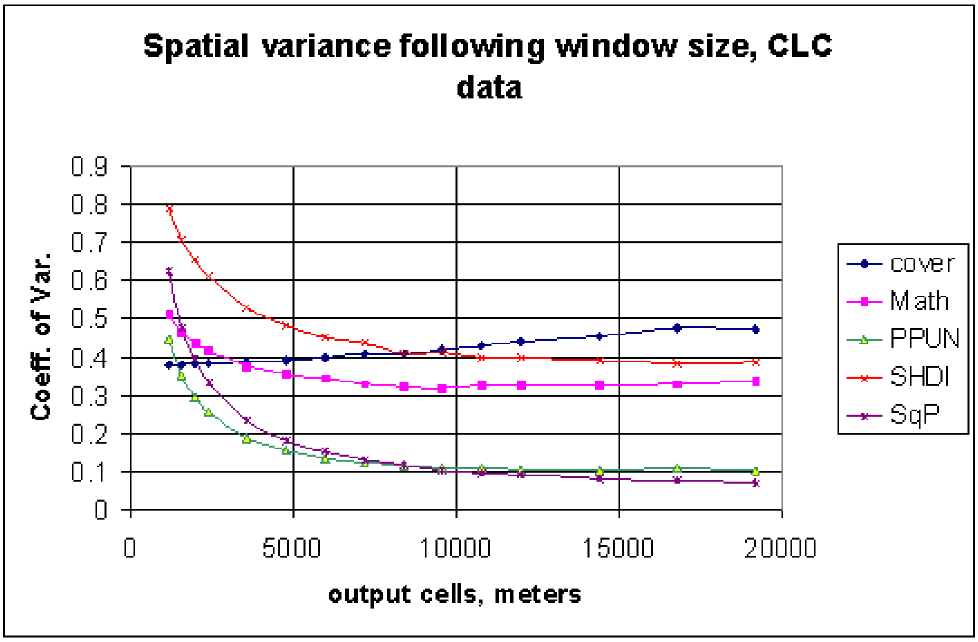

Figure 4.18 Local variability of CLC data. Coefficient of variation from the suite of spatial metrics as function of the window sizes for which they are calculated.

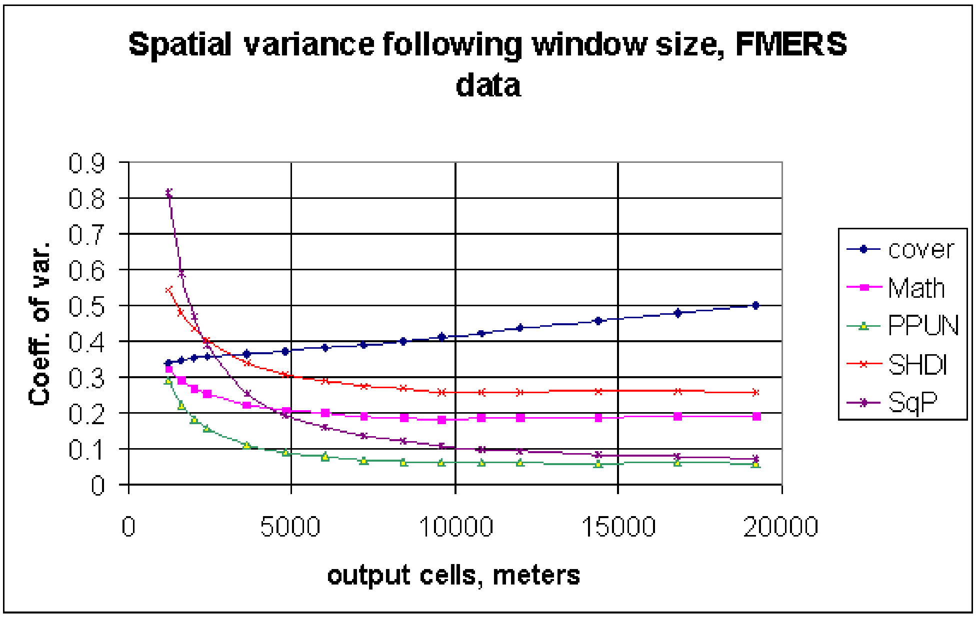

Figure 4.19 Local variability of FMERS forest map data. Calculated as described above.

An alternative way of describing spatial variability is through the Moran's I metric of autocorrelation, as shown in Figure 4.20 and Figure 4.21, below:

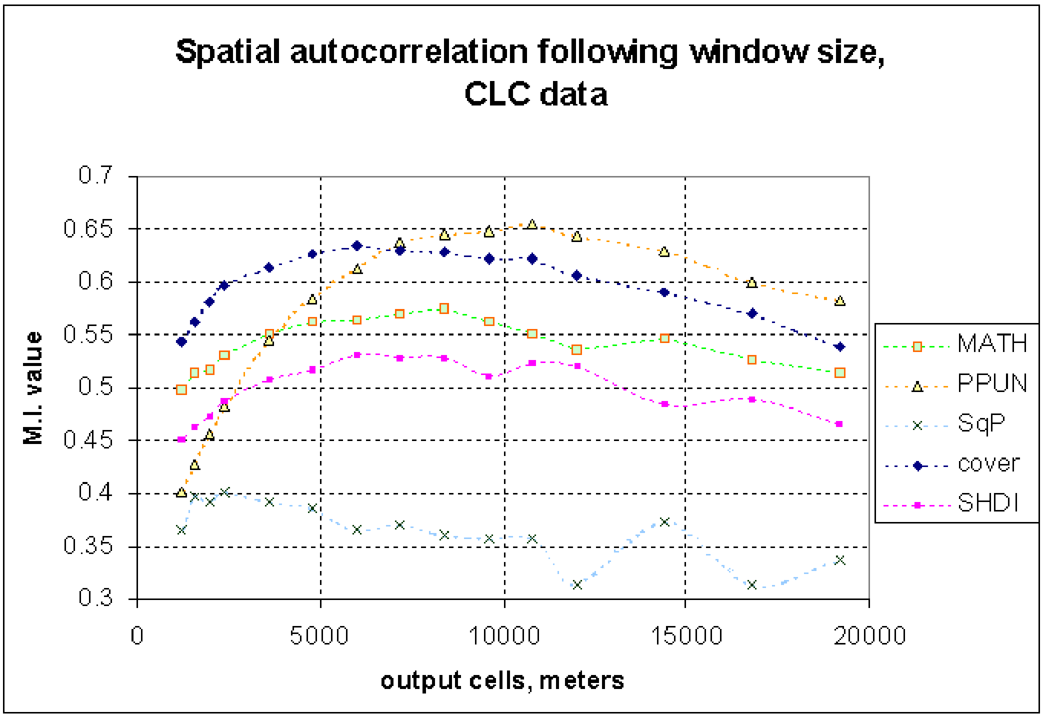

Figure 4.20 Local variability of spatial metrics from CLC data, expressed with Moran's I as function of the cells for which they are calculated.

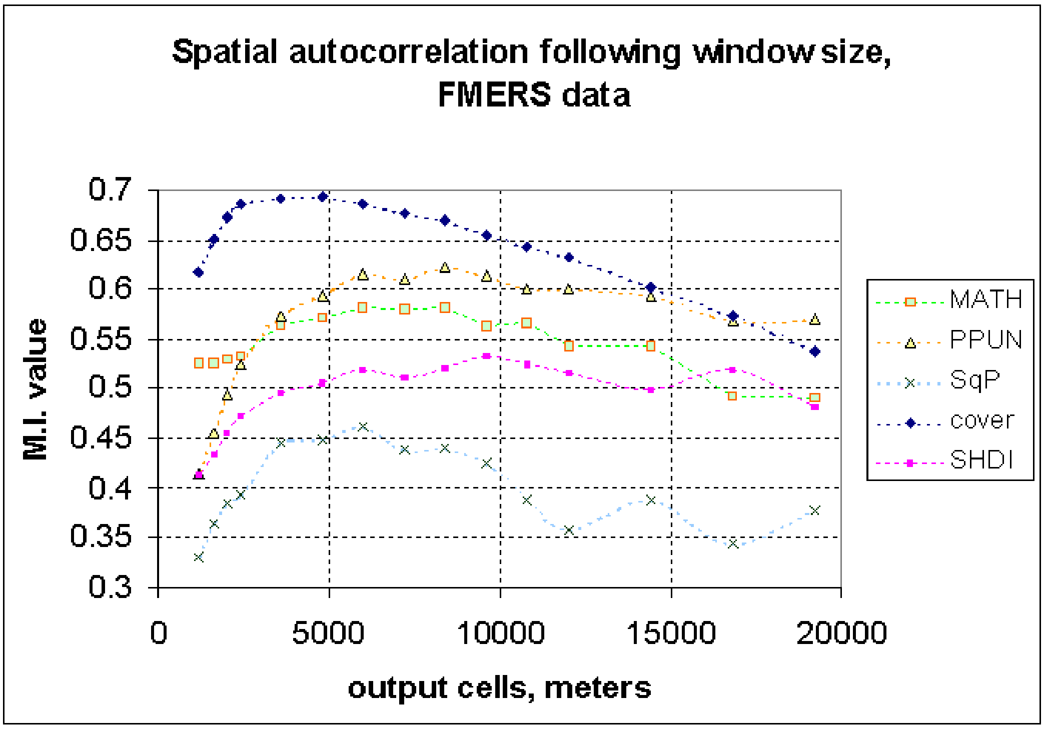

Figure 4.21 Local variability of spatial metrics from FMERS forest map expressed with Moran's I.

For this method of measuring local variance, all the metrics show distinct peaks. The shape of the curves are quite similar for the two data types, but as for variance measures, the position of the peaks differ.

In principle, low values of MI should correspond to high values of variance, but the behaviour of the values as expressed in Figure 4.20 and Figure 4.21 differ from what is observed for the standard deviation and coefficient of variance values in Figure 4.16 to Figure 4.19. The troughs on the graphs represent window sizes with relatively lower spatial autocorrelation, and thus the highest information content on landscape structure. This indicate that SqP and M should be reported and/or mapped with window size 12 km and SHDI at 14.4 km. On the other hand, the distinct peaks of MI values for the cover metric indicate that window sizes around 6 km for CLC data and 4 to 5 km for FMERS data should be avoided when maps of forest cover are made – keeping in mind that the purpose of such maps is to highlight differences between areas. The fact that the Matheron index for description of fragmentation peaks at larger window sizes than the cover fraction metric could mean that the spatial structure of the forest area is a property that characterises different regions, and is more or less independent of the actual forest cover. This assumption can be confirmed by inspection of the correlation between values of cover and M, as done in the next section.

4.5.3 Relationships between different metrics

The values of the different spatial metrics are far from independent of each other, as shown in a number of studies (for instance Riitters *et al* 1995, Hargis *et al* 1998, Gallego *et al* 2000). The way in which the correlation coefficients vary with window size is a scaling property of the metric as well as of the data. In this study with a fixed study area and increasing window sizes, the number of samples i.e. output cells or windows will decrease as the size and number of 'input-pixels' for each window increases. The number of windows with 'forest presence' has been counted, and the critical values of the correlation coefficient r to attain significance are listed in Table 4.6, below. Note that for large sample sizes, even small values of r are significant.

cell size, meters	Number of observations	r_crit	cell size, meters	Number of observations	r_crit
1200	86431	0.007	**8400**	2726	0.038
1600	60113	0.008	**9600**	2088	0.044
2000	40152	0.010	**10800**	1681	0.049
2400	28644	0.012	**12000**	1364	0.054
3600	13271	0.017	**14400**	968	0.064
4800	7773	0.023	**16800**	724	0.074
6000	5095	0.028	**19200**	574	0.083
7200	3606	0.033			

Table 4.6 Critical R values for varying number of observations with α=0.05, calculated following the formula given in section 4.4.2.1.

Table 4.7 and Table 4.8 below represent the correlations between the different spatial metrics at the smallest window size in this study, namely 1200*1200 m as defined by 6*6 pixels of the FMERS map and 12*12 pixels of the CLC map. The area of this geographical window is 1.44 km^2 or 144 hectares. The number of output pixels, representing windows included (covered by the forest mask), which is also the number of observations, is 86,431, out of a total 242,528 pixels/windows in this largest or most detailed output image.

CLC_1200m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1						
SHDI	0.437	1					
SIDI	0.421	0.993	1				
Math	-0.191	0.005	0.013	1			
SqP	-0.641	-0.264	-0.255	0.081	1		
NP	0.481	0.741	0.72	0.312	-0.339	1	
NP_back	0.044	0.027	0.032	0.476	0.153	0.106	1

Table 4.7 Correlation coefficients between metrics, CLC image with 12*12 pixels window.

FMERS_1200m	**Cover**	**SHDI**	**SIDI**	**Math**	**SqP**	**NP**	**NP_back**
Cover	1						
SHDI	0.513	1					
SIDI	0.48	0.989	1				
Math	-0.367	0.037	0.072	1			
SqP	-0.555	-0.264	-0.256	0.096	1		
NP	0.575	0.871	0.837	0.09	-0.313	1	
NP_back	-0.046	0.114	0.114	0.388	0.219	0.118	1

Table 4.8 Correlation coefficients between metrics, FMERS image with 6*6 pixels window.

According to the coefficients given in Table 4.6, all correlations in Table 4.7 and Table 4.8 are significant, as their absolute value is greater than 0.007. For both data types, the highest correlation is found between the two metrics of diversity, which is not surprising given their definitions. Therefore, for the further analysis in this chapter only SHDI will be used – in order to avoid redundancy. SHDI expressing dominance however correlates better with the cover proportion than SIDI expressing evenness. This was expected, since densely forested areas tend to be dominated by one forest type. There is a strong positive correlation between the values of NP (patches per window, shown earlier to be proportional to PPUN), the cover fraction and SHDI. The reason for the correlation between metrics of diversity and patch density is probably that, when more than one land cover type is present in the window, more than one patch is counted – there are at least as many separate patches as land cover types within each window. The correlations are stronger for the FMERS data than for the CLC data, probably due to the fact that the land cover types are more evenly distributed in the FMERS map (see Table 4.2, page 117). In both data sets, the metrics of forest structure M and SqP show strong negative correlations with the forest cover fraction. This is probably because at this small window size there are many pixels representing 100% forest cover, which by definition give zero values of M and SqP. At this window size M and SqP values are only weakly correlated, indicating that they describe different aspects of landscape structure, at least at small window sizes. The count of background patches, NP_back are, for both data types highly correlated with the Matheron index. This confirms that NP_back (or the transformed version PPUN_B) has potential for use as an indicator of one important aspect of forest fragmentation. On the other hand it is worth noting that, while M correlates quite well with NP for the CLC data, the correlation is weak for the FMERS data. Finally, it is seen that for both data types the SqP metric is negatively correlated to the NP metric but positively correlated to the NP_back metric. This may result from the fact that SqP approaches zero as the forest cover approaches 100%, and the possibility of finding background patches is reduced.

The maximum number of forest patches at this window size is 19 for the CLC data and 37 for the FMERS data. This is a somewhat counterintuitive finding, as there are four times as many pixels in the CLC windows for similar resolutions, but it must be attributed to the way in which the data are prepared, namely the pixel-by-pixel classification of the FMERS data and the area-delineation for the CLC data, compare Figure 4.3 on page 115.

When the window side length is doubled to 2400m and the size is quadrupled to 5.76 km^2, the general pattern of correlations remain the same, as shown by the coefficient values in Table 4.9 and Table 4.10. The cover proportion becomes more positively correlated with the diversity metrics, and less negatively with the fragmentation metrics. At the same time, the diversity metrics are less correlated with the patch count metrics, a trend that continues for increasingly larger windows.

CLC_2400m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1						
SHDI	0.472	1					
SIDI	0.443	0.991	1				
Math	-0.080	0.115	0.129	1			
SqP	-0.481	-0.196	-0.185	0.211	1		
NP	0.512	0.740	0.718	0.382	-0.148	1	
NP_back	0.444	0.237	0.228	0.359	0.071	0.322	1

Table 4.9 Correlation coefficients between metrics, CLC image with 24*24 pixels window.

FMERS_2400m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1						
SHDI	0.509	1					
SIDI	0.460	0.986	1				
Math	-0.323	0.131	0.168	1			
SqP	-0.389	-0.077	-0.065	0.401	1		
NP	0.622	0.829	0.784	0.131	-0.066	1	
NP_back	0.467	0.327	0.299	0.120	0.112	0.421	1

Table 4.10 Correlation coefficients between metrics, FMERS image with 12*12 pixels window

Figure 4.22, below plots the relation between two pairs of metrics: forest cover percentage – SqP metric (cover vs. fragmentation) and PPUN – SHDI (patchiness vs. diversity) for both types of input data. The window size of 2400 m represent the smallest window size for which it was possible to use the graphic functionality of Excel (initial number of windows/output cells is 291*208=60,528, and Excel in the version used can handle a maximum of 65,536 rows). The negative correlation coefficient for SqP and cover seen in the tables above point to a general pattern of more square forest patches with higher forest cover, while the positive correlation coefficient for SHDI and PPUN (or NP = number of patches) reflects an increasing land cover diversity with more separate patches – or vice versa.

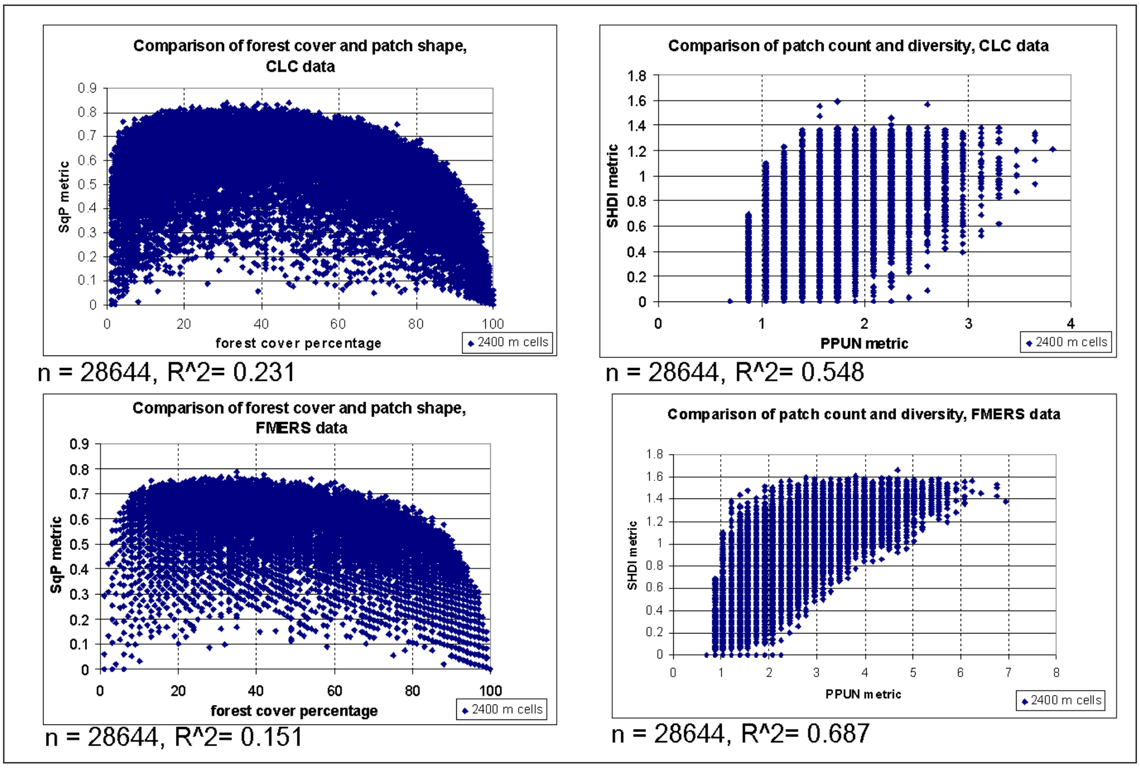

Figure 4.22 Plots of different metrics values from the same data source, here CLC and FMERS data with metrics calculated for 2400*2400 m windows. Only output cells with forest cover >= 1% are used.

The relations between metrics for the window size of 4800*4800 m corresponding to 23.02 km^2 are reported in Table 4.11 and Table 4.12. For the CLC data, the M metric is observed NOT to be significantly correlated with the cover proportion, the value of –0.003 represents a turning point, in the sense that for larger window sizes, the correlation coefficient is (significantly) positive.

CLC_4800m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1.000						
SHDI	0.468	1.000					
SIDI	0.418	0.988	1.000				
Math	-0.003	0.162	0.174	1.000			
SqP	-0.183	-0.025	-0.022	0.413	1.000		
NP	0.545	0.694	0.666	0.449	0.124	1.000	
NP_back	0.685	0.327	0.290	0.279	0.133	0.469	1.000

Table 4.11 Correlation coefficients between metrics, CLC image with 48*48 pixels window.

FMERS_4800m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1.000						
SHDI	0.482	1.000					
SIDI	0.418	0.981	1.000				
Math	-0.233	0.190	0.223	1.000			
SqP	-0.256	0.038	0.043	0.570	1.000		
NP	0.683	0.765	0.707	0.196	0.079	1.000	
NP_back	0.731	0.384	0.332	-0.009	0.048	0.585	1.000

Table 4.12 Correlation coefficients between metrics, FMERS image with 24*24 pixels window

For the window size of 9600*9600 m corresponding to 92.16 km^2, the relations are collected in Table 4.13 and Table 4.14 below. For both data types, both fragmentation metrics have now become clearly positively correlated with the diversity metrics. For the CLC data, M and SqP have positive correlations with cover proportion, while for the FMERS data, SqP is at the turning point with the value of –0.001, an r-value which is not a significant correlation to the cover proportion.

CLC_9600m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
SHDI	0.431	1.000					
SIDI	0.363	0.986	1.000				
Math	0.102	0.164	0.167	1.000			
SqP	0.165	0.175	0.163	0.621	1.000		
NP	0.590	0.604	0.575	0.522	0.374	1.000	
NP_back	0.806	0.323	0.267	0.280	0.281	0.567	1.000

Table 4.13 Correlation coefficients between metrics, CLC image with 96*96 pixels window.

FMERS_9600m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
SHDI	0.472	1.000					
SIDI	0.401	0.976	1.000				
Math	-0.119	0.230	0.253	1.000			
SqP	-0.001	0.262	0.264	0.684	1.000		
NP	0.743	0.707	0.644	0.275	0.267	1.000	
NP_back	0.850	0.402	0.337	-0.012	0.118	0.685	1.000

Table 4.14 Correlation coefficients between metrics, FMERS image with 48*48 pixels window

At the largest window size used, 19.2*19.2 km corresponding to a window area of 368.64 km^2, all correlation coefficients are positive and significant. Table 4.15 and Table 4.16 show that the correlation between cover fraction and number of background patches, which for both data types had low absolute values for small window sizes, has now grown to yield high values. This must be attributed to the fact, that for large window sizes, there are no windows which are completely covered by forest (for 19.2*19.2 km windows the maximum values are around 90% for both data types), and thus the effect that densely forested areas include a number of background patches here and there become dominant. Due to the nature of the two data sets, this effect is most apparent for the FMERS data, which have a more scattered appearance and no minimum area condition for mapping of patches – opposed to the CLC where the minimum area is 25 ha (corresponding to 6 ¼ FMERS pixels).

CLC_19200m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1.000						
SHDI	0.396	1.000					
SIDI	0.326	0.984	1.000				
Math	0.267	0.131	0.119	1.000			
SqP	0.394	0.272	0.247	0.728	1.000		
NP	0.645	0.480	0.453	0.630	0.505	1.000	
NP_back	0.872	0.318	0.256	0.373	0.397	0.657	1.000

Table 4.15 Correlation coefficients between metrics, CLC image with 192*192 pixels window.

FMERS_19200m	Cover	SHDI	SIDI	Math	SqP	NP	NP_back
Cover	1.000						
SHDI	0.441	1.000					
SIDI	0.358	0.968	1.000				
Math	0.059	0.242	0.247	1.000			
SqP	0.216	0.329	0.317	0.786	1.000		
NP	0.795	0.639	0.567	0.408	0.443	1.000	
NP_back	0.911	0.407	0.332	0.109	0.246	0.765	1.000

Table 4.16 Correlation coefficients between metrics, FMERS image with 96*96 pixels window.

The plots in Figure 4.23 below show the nature of the relations between different metrics for 19.2*19.2km windows, the largest extent examined here. These relations have been expressed here through the values of correlation coefficients – although the reality can be more complex than the linear relationships that are normally assumed. For instance, the shape of the curves for the M-SqP relations indicate a form of power-law relation between these two metrics.

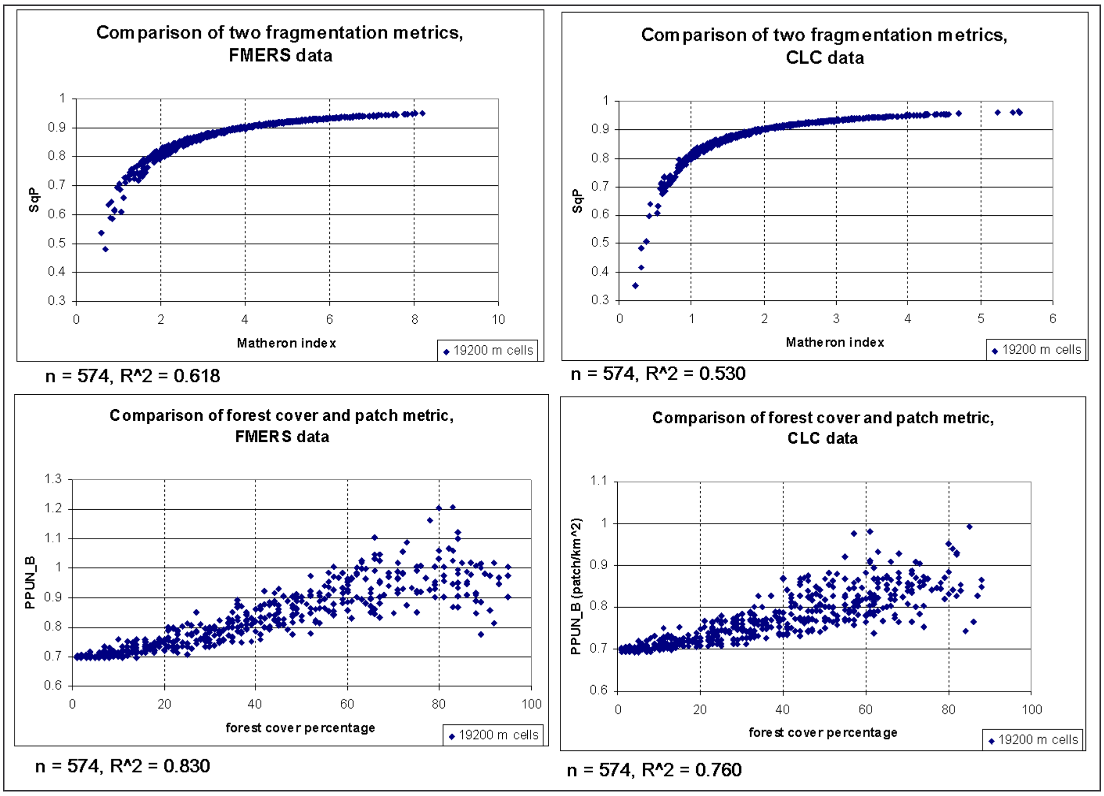

Figure 4.23 Plots of different metrics values from the same data source, here CLC (left) and FMERS (right) data with metrics calculated for 19200*19200 m windows. Only output cells with forest cover >= 1% are used.

Due to such relationships, when these metrics are used as indicators, we should not expect them to describe completely different aspects of landscape structure, but rather different interpretations of the relationship between forest area, edge length and total (landscape) area. Table 4.17 and Table 4.18 below summarise the correlations between cover fraction and the other metrics for the range of window sizes examined in this study. Correlation coefficients are observed to increase with window size for all the fragmentation and patch-count metrics and diversity metric to decrease slightly.

CLC	correllation between metric and cover%				
window size	**SHDI**	Math	SqP	NP	NP_back
1200	0.437	-0.191	-0.641	0.481	0.044
2400	0.472	-0.080	-0.481	0.512	0.444
4800	0.468	-0.003	-0.183	0.545	0.685
9600	0.431	0.102	0.165	0.590	0.806
19200	0.396	0.267	0.394	0.645	0.872

Table 4.17 Summary of correlation coefficients between cover proportion and metrics values at increasing window sizes for CORINE land cover data.

The difference between the CLC and the FMERS data is notable for the 'fragmentation metrics' M and SqP, where for the CLC data the correlations become positive for window sizes between 4800 and 9600m, while for the FMERS data they do so above 9600m. For both data types the SqP values become more highly correlated with forest cover at large window sizes.

FMERS	correllation between metric and cover%				
Window size	**SHDI**	Math	SqP	NP	NP_back
1200	0.513	-0.367	-0.555	0.575	-0.046
2400	0.509	-0.323	-0.389	0.622	0.467
4800	0.482	-0.233	-0.256	0.683	0.731
9600	0.472	-0.119	-0.001	0.743	0.850
19200	0.441	0.059	0.216	0.765	0.911

Table 4.18 Summary of correlation coefficients between cover proportion and metrics values at increasing window sizes for FMERS forest map.

4.5.4 Relationships between metrics derived from the two different data types

The degree of correlation between the values of the same spatial metric derived from two different data sets informs us about the degree to which the (metric) values from one data set can be used to predict and eventually substitute the values derived from the other (Costanza and Maxwell 1994). Examination of this degree of predictability provides information on the nature and usefulness of the (image) data sets as well as on the behaviour of the chosen metrics, however not distinguishing effects due to the 'nature' of the metrics from effects owing to the 'nature' of the data, as discussed by for instance Turner *et al* (1989) and Saura (2002).

Table 4.19 A and B summarise the agreements found between the *same metrics*, from the two *different image sources* with *different resolutions*, at varying window sizes. The R-square values are plotted against the window size in Figure 4.24.

comparing CLC-FMERS **A**		Window size, meters						
		1200	**1600**	**2000**	**2400**	**3600**	**4800**	**6000**
Cover	Multiple R	0.543	0.626	0.684	0.724	0.787	0.819	0.840
	R Square	0.295	0.392	0.468	0.524	0.619	0.670	0.705
SHDI	Multiple R	0.192	0.237	0.274	0.301	0.336	0.352	0.366
	R Square	0.037	0.056	0.075	0.090	0.113	0.124	0.134
Math	Multiple R	0.009	0.077	0.137	0.187	0.280	0.340	0.382
	R Square	0.000	0.006	0.019	0.035	0.078	0.116	0.146
PPU	Multiple R	0.237	0.305	0.360	0.394	0.459	0.499	0.527
	R Square	0.056	0.093	0.130	0.155	0.211	0.249	0.277
SqP	Multiple R	-0.019	0.014	0.027	0.045	0.144	0.204	0.227
	R Square	0.000	0.000	0.001	0.002	0.021	0.042	0.052

comparing CLC-FMERS **B**		Window size, meters							
		7200	**8400**	**9600**	**10800**	**12000**	**14400**	**16800**	**19200**
Cover	Multiple R	0.854	0.864	0.870	0.879	0.884	0.894	0.902	0.903
	R Square	0.730	0.747	0.757	0.772	0.781	0.799	0.814	0.816
SHDI	Multiple R	0.364	0.379	0.385	0.345	0.376	0.381	0.373	0.371
	R Square	0.132	0.144	0.148	0.119	0.142	0.145	0.139	0.138
Math	Multiple R	0.405	0.421	0.447	0.467	0.489	0.510	0.524	0.534
	R Square	0.164	0.177	0.200	0.218	0.239	0.260	0.274	0.285
PPU	Multiple R	0.538	0.567	0.571	0.582	0.598	0.620	0.633	0.642
	R Square	0.289	0.321	0.326	0.339	0.357	0.384	0.401	0.412
SqP	Multiple R	0.268	0.332	0.325	0.364	0.378	0.376	0.433	0.453
	R Square	0.072	0.110	0.106	0.132	0.143	0.141	0.188	0.205

Table 4.19 A and B Agreement between metric values from different image sources at varying window size. The values for output cell sizes < 2400m are calculated using the statistical functions of WinChips, for larger (and fewer) windows using the analysis module of Excel.

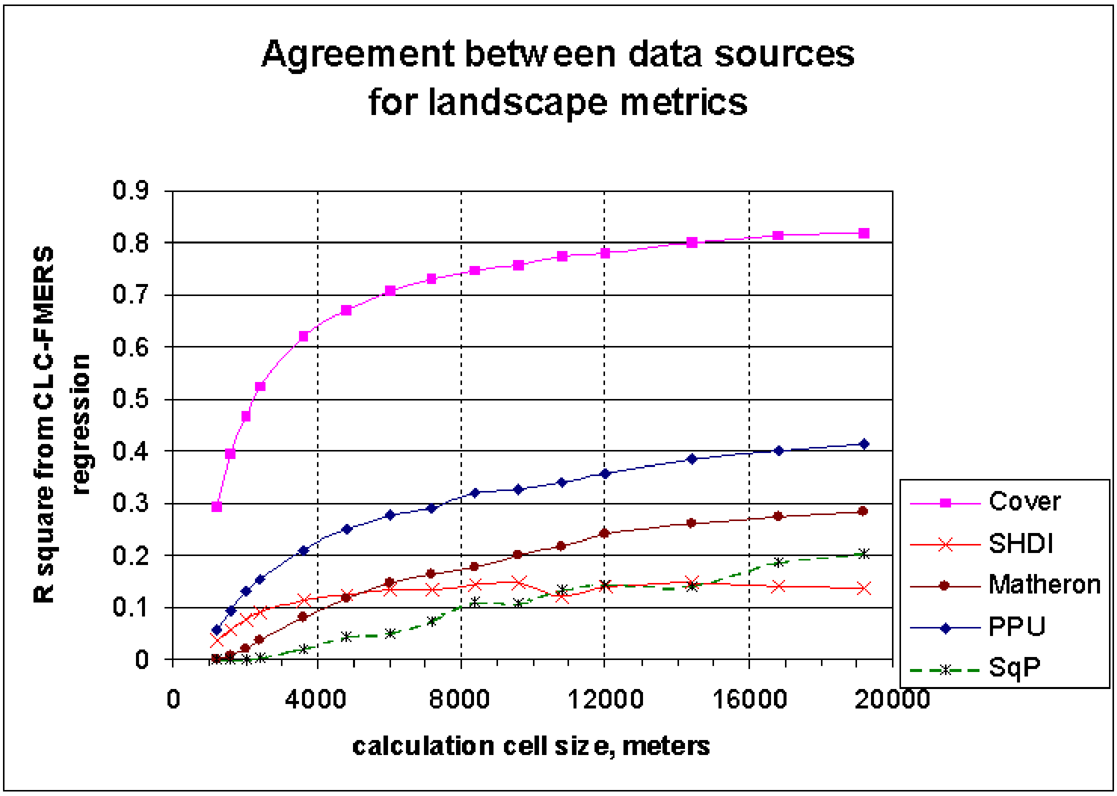

Figure 4.24 R-square, expressing agreement between metrics values from CLC and FMERS data, plotted against spatial extent/window size. Smallest windows are 6*6 pixels for FMERS and 12*12 pixels for CLC data, largest windows 96*96 pixels for FMERS and 192*192 pixels for CLC.

As shown in Chapter 3, the different metrics show quite different correlations at the same window size. More surprisingly they respond in different ways to the changes in window size, as expressed by the shape of the window size-correlation curves. In general, the increasing window size will even out differences between spatial structure as mapped in the two data sets, leading to higher correlations, most notably and understandable for the forest cover fraction, which also has the highest correlation coefficients at all window sizes. This is partly due to the elimination of possible errors in the geo-referencing of the datasets (how well the two ‘maps’ fit each other), a common problem for large-area data in grid format. The “dip” on the curve for the correlation of the SHDI–value at 10.8 km window size is not easily explained, as it has been computed in the same way as its neighbouring values and checked more than once. Perhaps the lower correlation of the SHDI diversity values at this window size reflects a change in spatial domain from landscape to regional level (following the size of characteristic landscape structuring elements like the width of valleys). Also the response curve for the SqP metric behaves in an irregular, step-wise fashion. The shape of the cover-curve suggest that the response of R^2 to window size follows a power-law or logarithmic relation, and that is confirmed by plotting these values against window size on a logarithmic scale as shown in Figure 4.25.

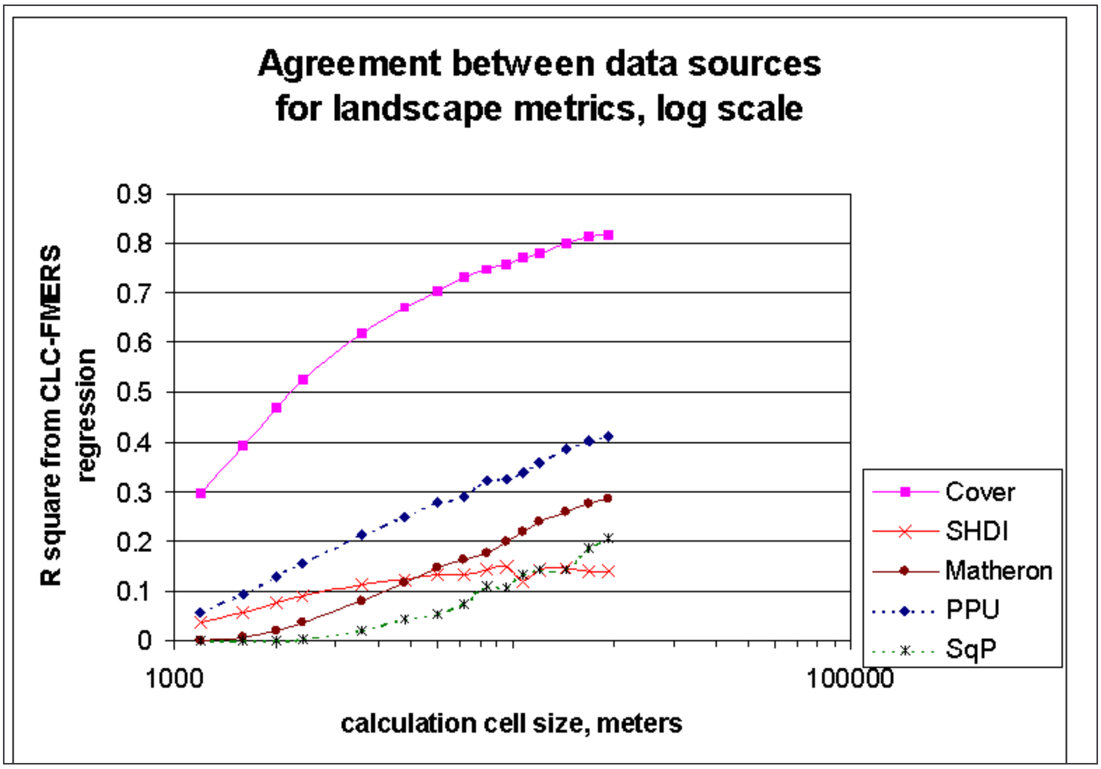

Figure 4.25 R-square-plot similar to Figure 4.24, but with window size values transformed logarithmically.

The correlation for the patch count metric PPU/PPUN (count of forest patches per unit area) improves steadily with window size, also in a log-linear fashion. This is an interesting and quite promising result, since the degree of patchiness and thus number of patches is amongst the largest differences between the CLC and the FMERS data sets (see the difference of the absolute (average) values of the metrics listed in Table 4.5). Though not shown above, correlations between the count of background patches in the two different data types were also derived for the window sizes described here in detail, and are reported in Table 4.20. The correlation of background patches-count values follows the pattern of correlation (of NP_back) with forest cover fraction seen in Table 4.17 and Table 4.18. When the metric of background patches correlate well for large windows, it is in agreement with the high correlation of forest cover-fraction values between the two data sets for large windows.

	Inter-correlation CLC-FMERS images	
window size	**NP**	**NP_back**
1200	0.237	0.093
2400	0.394	0.272
4800	0.499	0.518
9600	0.571	0.681
19200	0.642	0.791

Table 4.20 Agreement between the two data sources on the number of ”background patches”, as expressed through the correlation coefficient R, improves drastically with increasing size of output cells (and thus the number of input pixels).

4.5.5 Comparisons of metrics values with different regionalisation approaches

The use of watersheds or catchments (the term used here) is becoming increasingly popular for environmental assessment in general and for reporting of spatial metrics in particular. Intuitively it seems reasonable to use these naturally delineated, functional regions as the basis for reporting of environmental parameters, especially when these are related to water quality or sediment load. Recently, there has been a number of studies on the use of spatial metrics at watershed level (Tinker *et al* 1998, Patil *et al* 2000, Jones *et al* 2001, Cifaldi *et al* 2003). Vogt *et al* (2003) used satellite based forest maps in combination with catchment and elevation data[19] to describe forest-water interactions such as the fraction of rivers running through forest. Administrative regions, on the other hand, have the advantage of already forming a hierarchy of levels from nation state to parish, farm or forest plot for which GIS data are readily available.

A central question that can possibly be answered with the MW-approach is whether catchments are more homogeneous than the administrative regions within the study area. This is relevant because watersheds/catchments have been proposed as natural reporting units for landscape properties and environmental indicators (Apan *et al* 2000, Paracchini *et al* 2000, Patil *et al* 2000, Vogt *et al* 2003). In this section the question is addressed through extraction of spatial metrics values for selected NUTS-regions and for selected 4th to 6th order catchments. Thus, the MAUP is treated through data analysis on overlapping but different regions. Also the coefficient of variance is calculated for the administrative regions and the catchments for both data types, and for a number of window sizes – since it can be hypothesised that if a more homogeneous forest structure is found within the catchments, the variation of the metrics values that characterise structure will be smaller within the region (in practice/GIS-implementation the polygon used to extract statistical parameters).

Another dimension is the comparison of the two different data sources. When the same set of results is derived from both data sources, in terms of output cell size and metrics, the agreement between them can be investigated at the level of catchment or region. Thus, regression between CLC and FMERS metrics was performed separately within the geographical areas of interest. Finally the averaged values per region were compared. Given the limited number of regions and the problem with regression of such averaged values, the rank-size correlation was applied, in order to test whether the metrics were sufficiently robust to point out areas with high/low diversity, fragmentation etc. even with different input data.

[19] The catchment and elevation data used in this thesis are based on the ones used in Vogt *et al*'s study, which is carried out at the JRC. The current version of the database is available through the web site http://agrienv.jrc.it; follow the link Activities - Catchments, and data can be requested and downloaded.

The administrative regions used are illustrated in Figure 4.26, and the catchments with numbering are shown in Figure 4.27.

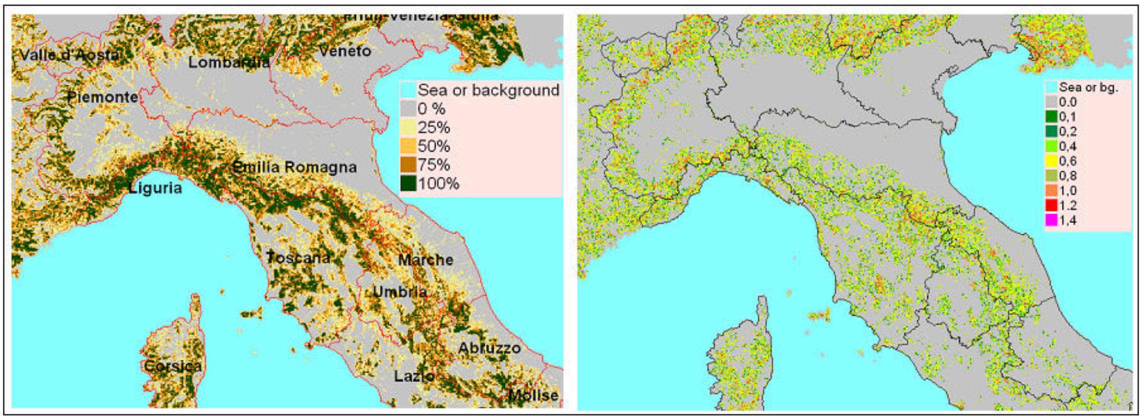

Figure 4.26 Forest cover and SHDI in 1200*1200 m cells from CLC forest map. To the left, forest cover overlaid with Italian regions (NUTS-2 level). To the right map with values of Shannon's diversity index (SHDI), created in the same IDL batch-run.

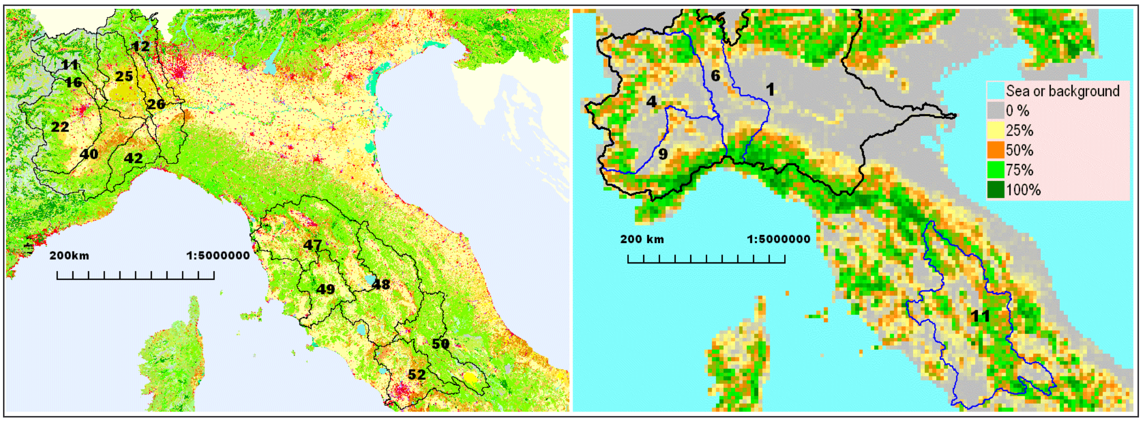

Figure 4.27 CLC data with high-order catchment polygons. To the left, the original CLC data overlaid with the 4th order catchments used in this study. To the right, forest cover fraction from the CLC-based forest maps for 4.8*4.8 km windows, overlaid with 5th and 6th order catchments.

Note that Corsica is included in the administrative theme, even though the island is a French region. In the text and tables, the catchments are named by their order, followed by an underscore and the code that functions as unique identifier, so possible names are i.e. 5_01.

Statistical properties of the MW-outputs were extracted per administrative region and catchment for a subset of spatial resolutions, namely 1200, 2400, 4800, 9600 and 19200m output cells. This is deemed sufficient to describe scale effects on the metrics values, though the entire set of metric images as used in the previous sections were available. Not all of the extracted values are shown here in table form, but scaling profiles are used to illustrate their properties for the selected geographical units. Note that the way in which the graphs are constructed result in a ‘logarithmic’ appearance, as the window size is doubled for each step.

Figure 4.28 below presents examples of how spatial metrics are derived with the M-W method and reported either as raster maps with pixels corresponding to the output cells or as vector maps with metrics values assigned to regions (the ones used for delineating the parts of the image from where statistical information is extracted). Note that in the figure, image 2 is derived from image 1, and that image 3,4 and 5 subsequently represent different way of describing the MW-outputs in image 2.

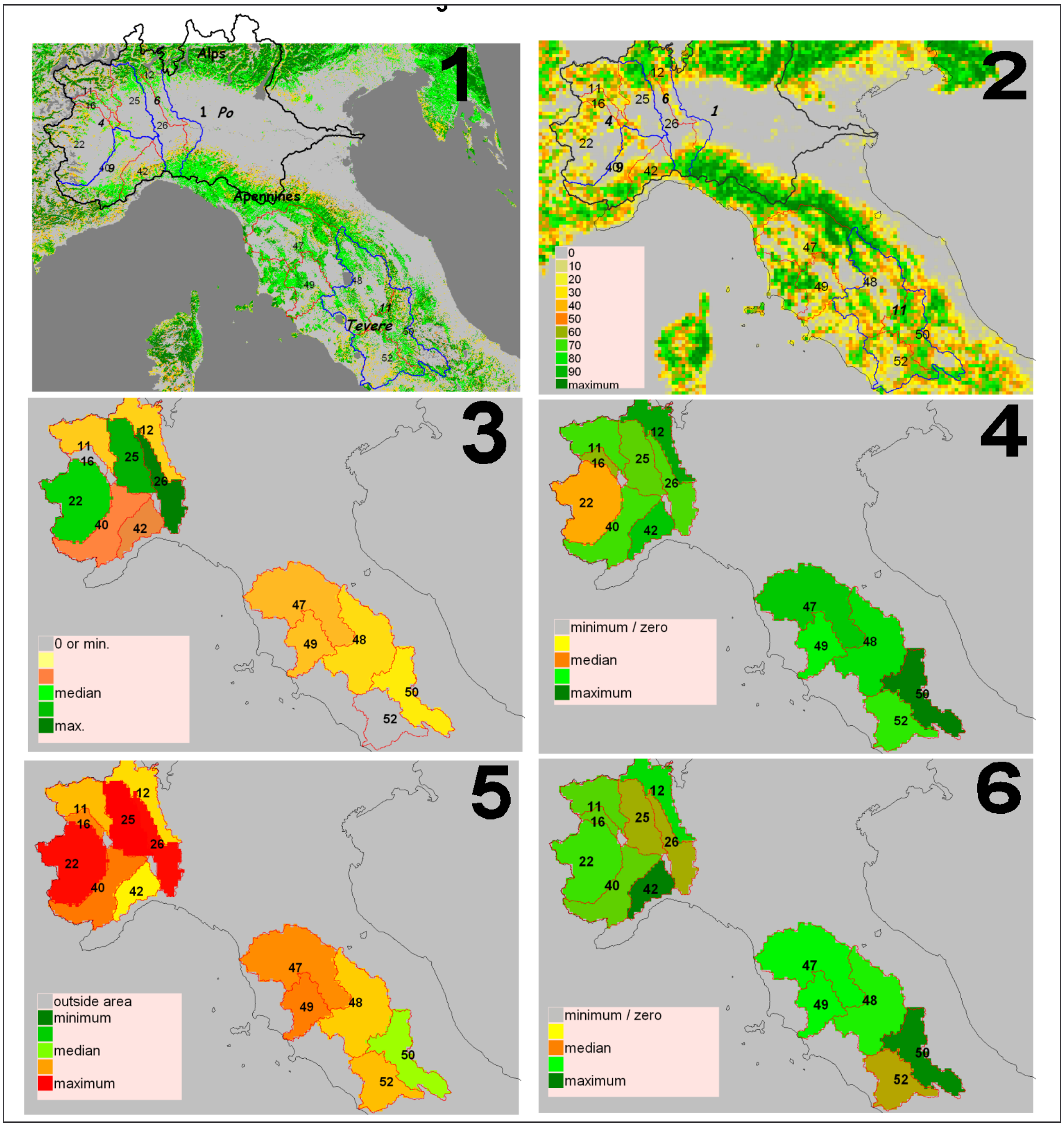

Figure 4.28 Examples of landscape metrics values reported at catchment level, in this case relatively simple forest cover information.

In Figure 4.28, Image 1 shows the catchments of 4^{th} to 6^{th} order that are used here on a background of the FMERS forest map that is used as input to the MW-calculations. Image 2 shows M-W output at window size 4800m, for each output cell measured forest cover percentage. These values are used in the following derived images. Image 3 shows the FC metric values, ranging from 0 in the lower Tevere to 0.425 on the upper Po plain. Image 4 shows the cover percentage (under the forest mask) per catchment, and image 5 shows the coefficient of variation within the catchment of the cover percentage values. Finally, image 6 shows the cover percentage values from the CLC data also at 4800m cell size, and is thus directly comparable with image 4. The rank correlation between these two particular outputs is found to be significant at 5% probability level (Table 4.34, below). Due to the nature of the image data (floating point) and a wish to use the full range of colours of the look-up-tables relative values are shown in the image legends.

4.5.5.1 Metrics values within catchments

The statistics for the various output cell sizes were collected in spreadsheet files – one for catchments and one for administrative regions, making it possible to report and summarise the metrics values. Examples for the catchments delineation are shown in Table 4.21 and Table 4.22 below. It appears here that according to CLC, the highest values of diversity metrics and lowest values of fragmentation metrics are found at relatively high altitudes in the Po catchments in the northern part of the area. The lowest diversity and highest fragmentation is then in the catchments that contribute to the Tevere. Catchment 4_48 (region 11 in the tables below) is the upper catchment of that river, an area that more or less coincides with the Umbria administrative region. The highest FC value is found for 4_26 (7 in the tables) that is situated across the Po plain on the upper to middle part of the rivers longitudinal extent, east of the confluence with the Ticino river at Pavia, while the lowest FC value is found for catchment 4_49 (10) in Toscana, with a mixture of agricultural plains and forested hills.

CLC	4800m	pixels	cover	PPUN	PPUNB	SHDI	SIDI	Math.	SqP	Elevation	FC
6th ord	6_01 Po (1)	2168	39.87	1.039	0.829	0.507	0.299	2.301	0.655	788.9	0.306
5th ord	5_04Po (1)	622	36.31	1.04	0.814	0.576	0.344	2.157	0.645	1122.4	0.172
	5_06 Po (2)	286	37.31	1.023	0.813	0.452	0.265	2.33	0.656	516.4	0.206
	5_09 Po (3)	331	42.25	1.103	0.835	0.516	0.308	2.691	0.69	679.4	0.085
	5_11Teve (4)	704	42.21	1.074	0.881	0.343	0.191	3.123	0.734	592.1	0.095
4th ord	4_42 Po (1)	111	54.81	1.146	0.868	0.58	0.341	2.408	0.654	468.6	0.036
	4_40 Po (2)	213	36.31	1.088	0.82	0.487	0.293	2.876	0.714	794.6	0.084
	4_22 Po (3)	286	38.28	1.032	0.817	0.565	0.335	2.047	0.63	1082.7	0.154
	4_16 Po (4)	35	36.43	1.101	0.847	0.728	0.436	2.266	0.663	1258.5	0.057
	4_11 Po (5)	148	36.88	1.051	0.804	0.665	0.398	2.082	0.648	1730.5	0.101
	4_25 Po (6)	146	32.71	1.04	0.807	0.497	0.297	2.357	0.661	608.0	0.308
	4_26 Po (7)	114	32.44	0.976	0.797	0.357	0.212	2.324	0.636	360.7	0.325
	4_12 Po (8)	145	45.11	1.089	0.842	0.568	0.331	2.305	0.67	684.0	0.069
	4_47 Tosc (9)	377	42.18	1.077	0.86	0.547	0.319	3.039	0.723	342.4	0.085
	4_49 Tosc (10)	154	42.01	1.068	0.862	0.469	0.274	3.102	0.727	341.3	0.019
	4_48 Teve (11)	335	39.86	1.104	0.901	0.331	0.182	3.665	0.771	462.6	0.042
	4_50 Teve (12)	213	53.76	1.092	0.904	0.407	0.223	2.536	0.708	934.7	0.047
	4_52 Teve (13)	156	31.49	0.984	0.805	0.284	0.166	2.759	0.688	402.4	0.276
	avg. 5th order	*413*	*38.63*	*1.055*	*0.821*	*0.515*	*0.306*	*2.393*	*0.664*	*772.7*	*0.154*
	avg. 4th order	*149.75*	*39.12*	*1.065*	*0.825*	*0.556*	*0.330*	*2.333*	*0.660*	*873.5*	*0.142*

Table 4.21 Summary at catchment level of spatial metrics from the CLC map, with medium window size 4800m. The Highest metrics values are highlighted in yellow, lowest values in blue. Average elevation from the terrain model is included as a supplementary description of the area. Note that this value is an average for the forested windows in the area only.

Figure 4.29 below shows the scaling profiles of the SHDI and Matheron indices respectively, for six 4th order catchments with pronounced differences in the shapes of the curves. The continuous increase and fall of the values are expected from previous results (section 4.5.1), so what is interesting are the edges on the curves. The sharp increase of the SHDI values for catchment 4_26 from 9600 to 19200m window size reflects that a characteristic forest (patch) size has been exceeded and additional forest classes are included in each instance of the window, this is especially clear for the southern part of the catchment, with hills to the north of the Apennines. On the other hand, the SHDI values for catchment 4_22 increase only little when the window side length is doubled from 9600 to 19200m, because only few of the larger windows include more forest classes (implying that the characteristic landscape or forest size in the area, in terms of side length, is not larger then 10 km). For both CLC and FMERS based metrics, the average values for the 13 fourth order and the 4 fifth order catchment areas are almost the same, and that the fifth order values show less variation, since they represent average values taken over larger areas. As the catchment areas become larger, the metrics values approach the averages for the entire study area that are shown in Figure 4.11.

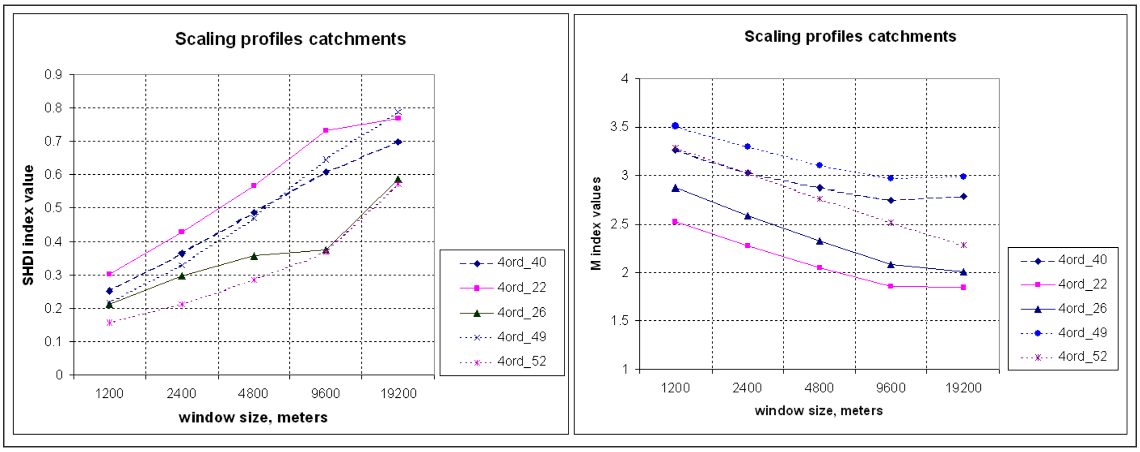

Figure 4.29 SHDI and Matheron metrics, extracted from CLC data to catchment areas, for a range of output cell sizes.

For the FMERS data, reported in Table 4.22 there is a clear difference between the catchments in the northern and southern part of the study area, as expected from the input forest maps (compare Figure 4.2 and Figure 4.27). The 4th order catchments of Tevere have high diversity and fragmentation values, including the patch count metrics. The lowest fragmentation metric values are found in the 4th order catchments of Po that include a substantial part of the plain where agriculture is dominant – and the map indicates little or no forest presence. Catchment 4_22 is shown as having surprisingly little forest cover, but this is partly due to problems with clouds in the input images, as often in mountains. This effect also contributes to observed low forest cover for the administrative region of Piemonte, and it is obviously a source of error in the calculations (where pixels marked as cloud, snow etc. should preferably not be counted in).

FMERS	4800m	pixels	cover	PPUN	PPUNB	SHDI	SIDI	Math.	SqP	Elevation	FC
6th ord	6_01 Po (1)	2027	42.99	1.684	0.851	0.799	0.446	3.913	0.608	851.7	0.397
5th ord	5_04Po (1)	598	29.32	1.507	0.803	0.745	0.431	4.273	0.636	1198.2	0.219
	5_06 Po (2)	283	40.22	1.629	0.852	0.758	0.417	3.948	0.622	531.6	0.219
	5_09 Po (3)	315	38.67	1.853	0.89	0.805	0.454	5.198	0.695	706.8	0.140
	5_11Teve (4)	757	43.66	2.129	0.915	0.966	0.527	5.404	0.716	561.0	0.018
4th ord	4_42 Po (1)	103	44.76	1.924	0.943	0.851	0.454	5.562	0.716	494.1	0.116
	4_40 Po (2)	209	35.89	1.823	0.866	0.784	0.455	5.038	0.686	807.0	0.105
	4_22 Po (3)	257	21.97	1.337	0.771	0.736	0.433	4.209	0.622	1137.2	0.284
	4_16 Po (4)	37	31.87	1.659	0.802	0.834	0.45	4.482	0.663	1347.6	0.000
	4_11 Po (5)	156	35.85	1.687	0.821	0.835	0.474	4.566	0.672	1794.0	0.045
	4_25 Po (6)	142	34.42	1.588	0.833	0.653	0.379	3.968	0.612	670.9	0.345
	4_26 Po (7)	106	35.30	1.332	0.845	0.537	0.295	3.979	0.627	382.2	0.425
	4_12 Po (8)	149	48.26	1.928	0.879	0.949	0.52	3.823	0.618	683.9	0.040
	4_47 Tosc (9)	386	44.61	1.968	0.869	0.929	0.518	4.509	0.633	334.3	0.060
	4_49 Tosc (10)	149	40.52	1.75	0.853	0.823	0.473	4.706	0.655	347.9	0.054
	4_48 Teve (11)	339	42.43	2.016	0.905	0.955	0.532	5.354	0.71	459.9	0.030
	4_50 Teve (12)	219	51.79	2.595	0.951	1.188	0.625	4.651	0.688	924.8	0.018
	4_52 Teve (13)	199	36.83	1.81	0.892	0.741	0.409	6.317	0.756	333.1	0.000
	avg. 5th order		*36.07*	*1.663*	*0.848*	*0.769*	*0.434*	*4.473*	*0.651*	*812.2*	*0.193*
	avg. 4th order		*36.04*	*1.660*	*0.845*	*0.772*	*0.433*	*4.453*	*0.652*	*914.6*	*0.170*

Table 4.22 Summary at catchment level of spatial metrics from the FMERS map, with medium window size 4800m as example. The highest metrics values are highlighted in yellow, lowest values in blue. The reason that the average elevation values are not the same as for the CLC data, is that different inclusion/forest presence masks are used.

The graphs in Figure 4.30 show the same general pattern in the selected catchments as observed for the CLC data, although for the FMERS data used here catchment 4_52, lower Tevere including the Rome metropolitan area, stands out with high values of M at all window sizes, indicating high fragmentation. For the CLC data, this area does not stand out in the same way, so the profile partly reflects the tendency of the FMERS mapping to place many small forest patches of type OWL broadleaved in areas where CLC show no forest. Catchment 4_40, reaching from the summit of the Maritine Alps to the Po valley east of Torino, has a profile of SHDI value similar to catchment 4_22 with CLC data. Also the SHDI diversity metric for this catchment reaches a maximum when the sub-landscapes get sufficiently large to include all possible forest classes. Catchment 4_26 has constantly low values for both metrics, because the two FMERS classes broadleaved and mixed forest dominate in the area, and because the forest patches are relatively coherent – in fact having the highest FC value of the catchments for this data type. Catchment 4_22 has a steeper M-value curve, with higher values at small window sizes, this must more small-scale fragmentation, i.e. more open forest or fringed edges, a structure typically found on mountain slopes.

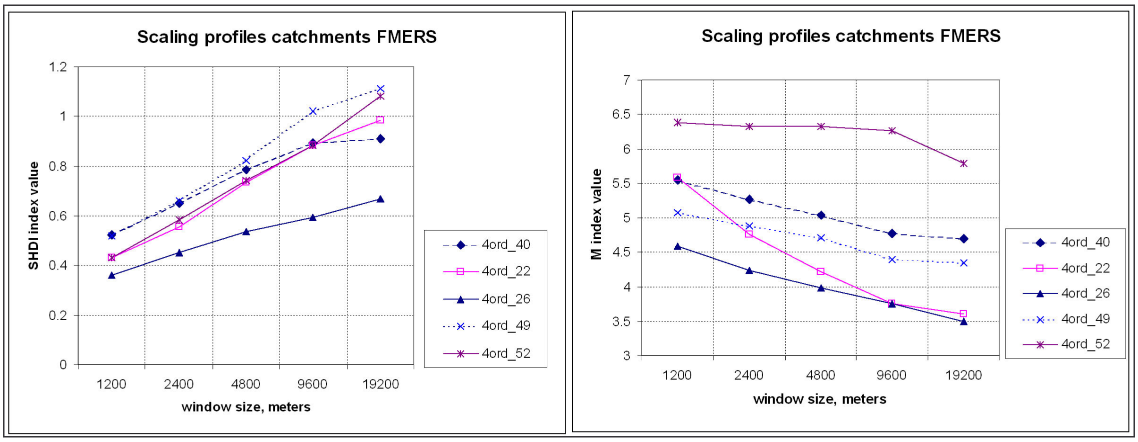

Figure 4.30 SHDI and Matheron metrics extracted from FMERS data to catchment areas, for a range of output cell sizes.

The hierarchical nature of the catchment delineations at different orders allows comparison of metrics for catchments at lower levels with those of higher levels. In general, the values at higher orders are close to the average of those at lower orders, that together constitute the catchment, as can be seen from the values in Table 4.21 and Table 4.22, and more clearly from Table 4.23, where the intention has been to make a table structure that reflects landscape structure. Matheron index values are used as examples, since fragmentation is indeed a phenomenon that manifests itself in different ways at different spatial levels. All the 5th order catchments have small areas in the lower parts that are not amongst the 4th order catchments used here, but the effect of that is assumed negligible.

CLC				Math.	FMERS			
6^{th}	5^{th}	4^{th}	4^{th} level number	4800m	6th	5^{th}	4^{th}	4^{th} level number
2.301	2.691	2.408	4_42 Po (1)		3.913	5.198	5.562	4_42 Po (1)
		2.876	4_40 Po (2)				5.038	4_40 Po (2)
	2.157	2.047	4_22 Po (3)			4.273	4.209	4_22 Po (3)
		2.266	4_16 Po (4)				4.482	4_16 Po (4)
		2.082	4_11 Po (5)				4.566	4_11 Po (5)
		2.357	4_25 Po (6)				3.968	4_25 Po (6)
	2.33	2.324	4_26 Po (7)			3.948	3.979	4_26 Po (7)
		2.305	4_12 Po (8)				3.823	4_12 Po (8)

Table 4.23 The hierarchical approach illustrated. Average Matheron index values from windows with extent 4800m, extracted for selected catchments in the upper Po valley plus the entire river basin (6^{th} order). The 5^{th} order catchments are from the top: 5_09, 5_04 and 5_06.

As expected, and shown in a previous section, the FMERS data yield higher values of diversity as well as fragmentation type metrics relative to the CLC data in all catchments. The ordering or ranking of the areas according to M value however differ significantly, as discussed below.

4.5.5.2 Metrics values within administrative regions

Administrative regions have the advantage of being known beforehand by the people who should use spatial metrics as environmental indicators. Areas like Piemonte and Toscana and are also well known for certain landscape characteristics such as mush or dense forest or large open areas with views over rolling hills. The observed metrics values for these regions are shown in Table 4.24 and Table 4.25 below for CLC and FMERS maps respectively, and scale profiles for selected areas and metrics are shown in Figure 4.31 and Figure 4.32. Two small regions almost coincide with catchments: Valle d'Aosta with 4_11 (which include a bit of the plains around Ivera to the SE of the valley) and Umbria with 4_48[20].

The regionalisation results mark Liguria, situated between the Northern coast of the Mediterranean and the Apennines, as a partiular area with dense forest cover and low

[20] Mountains can provide natural borders, and Umbria has been a stable geographical unit for thousands of years.

fragmentation. Also Valle d'Aosta has low fragmentation and high diversity, but this might be an artefact of the re-classification since the CLC class "transitional woodland-shrub" has been aggregated into OWL broadleaved (see Table 4.1 on page 116), though in this area it constitutes the zone around the tree line, where it can be questioned if it constitute a separate type of forest rather than less dense deciduous forest. Corsica, another region with large differences in elevation within short distances, has similar high diversity values. High fragmentation is found in middle Italy, with highest values for the Marche region, where the forest structure can be interpreted as rather perforated with low cover but high PPUN_B value.

CLC 4800m	nr_pix	Cover	PPUN	PPUNB	SHDI	SIDI	Math.	SqP	Elevation	FC
Veneto	327	33.59	1.012	0.842	0.465	0.27	2.594	0.661	444.8	1.046
Lombardia	612	39.55	0.952	0.83	0.419	0.245	2.085	0.632	627.1	0.511
Piemonte	931	37.75	1.086	0.824	0.573	0.339	2.423	0.67	838.6	0.147
Valle d'Aosta	123	38.55	1.069	0.802	0.712	0.43	1.921	0.631	1976.6	0.146
Emilia Romagna	589	36.08	1.1	0.838	0.45	0.269	2.903	0.702	473.8	0.630
Liguria	236	70.96	1.059	0.884	0.608	0.342	1.731	0.598	559.2	0.004
Toscana	945	48.41	1.063	0.868	0.519	0.299	2.756	0.696	385.0	0.044
Marche	385	29.06	1.331	0.875	0.511	0.317	3.948	0.78	441.1	0.096
Umbria	348	42.18	1.114	0.921	0.312	0.169	3.595	0.769	517.5	0.052
Abruzzo	423	37.85	0.921	0.803	0.254	0.151	2.397	0.664	866.3	0.116
Lazio	494	33.74	0.996	0.812	0.33	0.19	2.654	0.694	494.6	0.144
Corsica	328	43.81	1.022	0.842	0.678	0.388	2.292	0.682	635.8	0.003
average value		*40.96*	*1.060*	*0.845*	*0.486*	*0.284*	*2.608*	*0.682*	*688.3*	*0.245*

Table 4.24 Summary at administrative region level of spatial metrics from the CLC map, window size 4800m used as example. Highest metrics values are highlighted in yellow, lowest values in blue.

FMERS 4800m	nr_pix	Cover	PPUN	*4.5.5.2.1.1* *PPUNB*	SHDI	SIDI	Math.	SqP	Elevation	FC
Veneto	472	29.97	1.614	0.818	0.669	0.406	4.97	0.65	318.3	0.417
Lombardia	574	45.99	1.759	0.85	0.827	0.46	3.386	0.566	665.9	0.612
Piemonte	882	34.12	1.634	0.837	0.767	0.438	4.459	0.65	882.4	0.211
Valle d'Aosta	134	36.84	1.697	0.823	0.842	0.475	4.28	0.653	2058.9	0.052
Emilia Romagna	552	48.22	1.755	0.876	0.847	0.46	3.753	0.588	499.8	0.739
Liguria	231	50.99	1.836	0.92	0.87	0.465	4.77	0.67	561.2	0.026
Toscana	953	50.38	1.904	0.879	0.903	0.504	4.12	0.614	381.9	0.036
Marche	326	32.59	1.95	0.857	0.956	0.525	4.912	0.665	491.0	0.294
Umbria	350	43.17	2.18	0.908	1.053	0.577	5.139	0.703	514.5	0.046
Abruzzo	455	35.84	2.049	0.846	0.933	0.5	5.218	0.698	808.7	0.037
Lazio	561	38.31	1.982	0.888	0.871	0.474	5.979	0.742	448.9	0.007
Corsica	329	56.83	1.873	0.93	0.86	0.477	3.993	0.631	641.0	0.000
Average value		*41.94*	*1.853*	*0.869*	*0.867*	*0.480*	*4.582*	*0.653*	*689.4*	*0.206*

Table 4.25 Summary at administrative region level of spatial metrics from the FMERS map, window size 4800m used as example. Highest metrics values are highlighted in yellow, lowest values in blue.

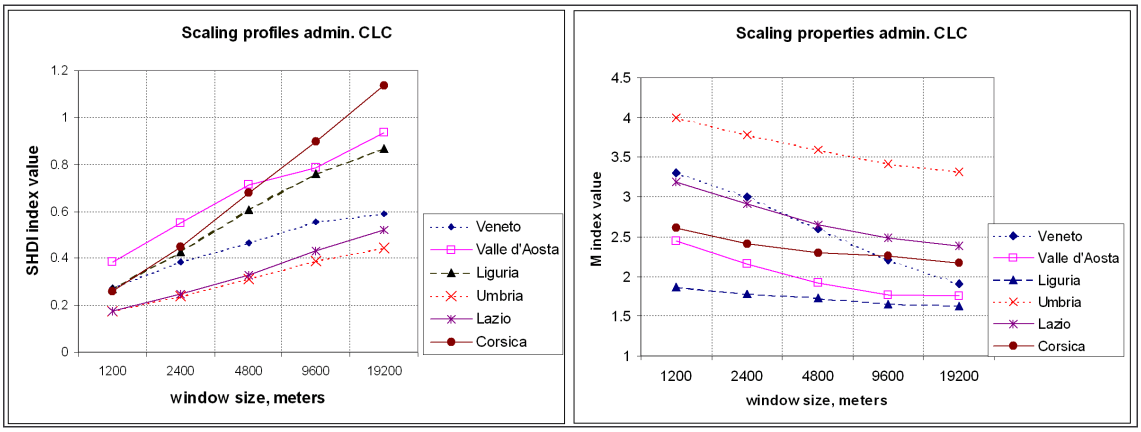

Figure 4.31 SHDI and Matheron metrics from CLC data, selected administrative regions, for a range of window sizes.

The higher contrasts in the landscapes of Corsica and Valle d'Aosta is also reflected in the shape of the scale-diversity curves, in Figure 4.31, right side. On the contrary, the Lazio and Umbria regions have low and slowly increasing diversity values. Liguria maintains low fragmentation values even at small window sizes while for Veneto they decrease rapidly with increasing window size, Figure 4.31, left side. This corresponds well with the high FC value found for this region from the CLC forest map.

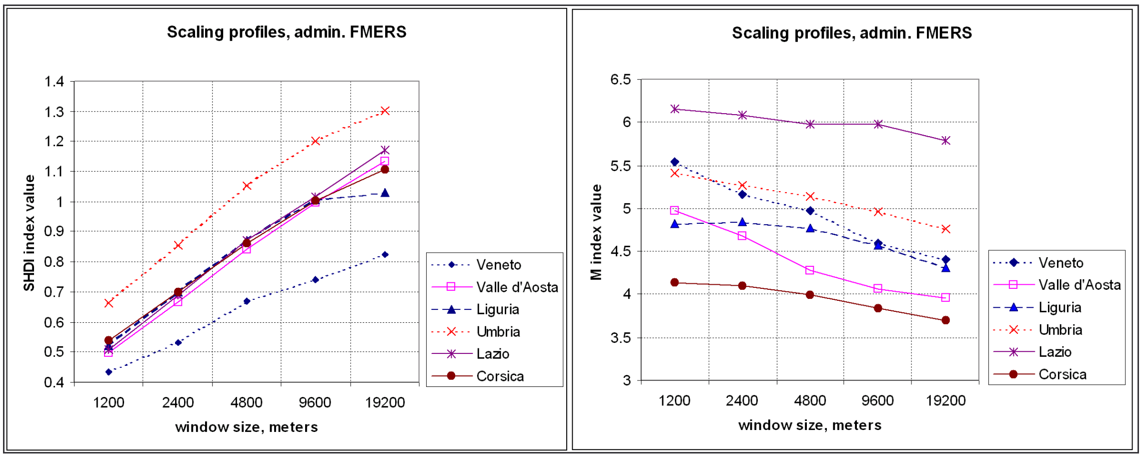

Figure 4.32 SHDI and Matheron metrics from FMERS data extracted to regions, for an interval of window sizes.

In the FMERS data, Veneto is marked by low forest cover and low diversity *within the windows of calculation*, i.e. the forest types are concentrated in specific geographical areas and not much interspersed, i.e. the (class) richness is low throughout (this is the case for both map types and all extents). High values of the fragmentation indicating metrics are found in middle Italy, most in Lazio and Abruzzo, now along with high diversity values, with Umbria having highest SHDI and SIDI values. The combination of high diversity and high fragmentation indicates a complex interspersion of forest and other land cover types.

The biggest difference between the metric values from CLC and FMERS is observed for Umbria, which has constantly high diversity values for the FMERS data, along with fragmentation values less than for neighbouring region Lazio. Inspection of statistics for the input data show that Umbria actually is a site of strong disagreement between the CLC and the FMERS classifications.

The forest cover proportions for Veneto and Umbria from CLC and FMERS are listed in Table 4.26, in order to examplify the effects of classification disagreements at region level (and to illustrate how diversity metrics are calculated). Although the forest percentage is almost the same from the two data sources in Umbria, the diversity values are at opposite ends of the scale. For Veneto there is better agreement, but again the FMERS data give a higher estimate of the forest diversity in the region.

Region	CLC	% of tot.	% of land	% forest	FMERS	% of tot.	% of land	% forest
Veneto	No data	0.01			No data	5.01		
Veneto	Coniferous	2.13	2.13	12.85	Coniferous	5.76	6.07	26.86
Veneto	Broadleaved Deciduous	9.72	9.72	58.62	Broadleaved Decid.	9.32	9.81	43.44
Veneto	Broadl. Evergreen	0.00	0.00	0.00	Broadl. Evergreen	0.00	0.00	0.00
Veneto	Mixed	2.24	2.24	13.49	Mixed	1.90	2.00	8.87
Veneto	OWL Coniferous	0.00	0.00	0.00	OWL Coniferous	0.19	0.20	0.88
Veneto	OWL Broadleaved	2.49	2.49	15.03	OWL Broadleaved	4.28	4.51	19.95
Veneto	Other Land	83.42	83.42		Other Land	73.54	77.42	
Veneto	total	100	100	100	Total	100	100	100
Veneto	land_map	1.00	**SHDI_forest**	**1.13**	land_map	0.95	**SHDI_forest**	**1.29**
Veneto	% forest	0.17	**SHDI_land**	**0.87**	% forest	0.23	**SHDI_land**	**0.94**
Umbria	No data	0.00			No data	2.67		
Umbria	Coniferous	0.60	0.60	1.50	Coniferous	5.55	5.70	13.45
Umbria	Broadleaved Deciduous	35.08	35.08	87.88	Broadleaved Decid.	13.79	14.17	33.42
Umbria	Broadl. Evergreen	0.02	0.02	0.05	Broadl. Evergreen	0.14	0.14	0.34
Umbria	Mixed	0.86	0.86	2.15	Mixed	11.89	12.22	28.82
Umbria	OWL Coniferous	0.00	0.00	0.00	OWL Coniferous	1.35	1.39	3.27
Umbria	OWL Broadleaved	3.36	3.36	8.43	OWL Broadleaved	8.54	8.78	20.70
Umbria	Other Land	60.09	60.09		Other Land	56.06	57.60	
Umbria	total	100	100	100	Total	100	100	100
Umbria	land_map	1.00	**SHDI_forest**	**0.47**	land_map	0.97	**SHDI_forest**	**1.45**
Umbria	% forest	0.40	**SHDI_land**	**0.41**	% forest	0.42	**SHDI_land**	**1.18**

Table 4.26 A comparison of forest proportion values and derived diversity metrics from the input data for two administrative regions.

4.5.5.3 Forest Concentration profiles

For the previously used metrics, the values at higher orders of regions and catchments are averages of the values for lower order areas – as a consequence of the way they are derived from the M-W outputs. This is not the case for FC values, where it is possible to have higher values at higher orders, due to the integrative nature of this metric (i.e. the files from the masking process are used indirectly). The inclusion of areas with little or no forest cover, typically in the lower parts of the catchments can give higher contrast between forested and non-forested cells and thus higher FC values. This effect is actually seen in Table 4.27 and Table 4.28, where values are reported for the smallest window size, 1200m and an intermediate window size, 4800m. Furthermore, there is a remarkably good agreement between the values extracted from the two image types, which initially shows the FC metric as a potentially useful description of landscape structure.

CLC				FC	FMERS			
6th	5th	4th	4th level number	1200m	6th	5th	4th	4th level number
0.760	0.343	0.192	4_42 Po (1)		0.866	0.452	0.320	4_42 Po (1)
		0.368	4_40 Po (2)				0.453	4_40 Po (2)
	0.673	0.626	4_22 Po (3)			0.851	1.257	4_22 Po (3)
		0.515	4_16 Po (4)				0.507	4_16 Po (4)
		0.530	4_11 Po (5)				0.342	4_11 Po (5)
		0.962	4_25 Po (6)				0.984	4_25 Po (6)
	0.668	0.998	4_26 Po (7)			0.686	1.079	4_26 Po (7)
		0.333	4_12 Po (8)				0.330	4_12 Po (8)

Table 4.27 FC values for catchments in Northern Italy for window size 1200m. Highest contrasts forest-non forest areas are found for the highest orders of catchments.

CLC				FC	FMERS			
6th	5th	4th	4th level number	4800m	6th	5th	4th	4th level number
0.306	0.085	0.036	4_42 Po (1)		0.397	0.140	0.116	4_42 Po (1)
		0.084	4_40 Po (2)				0.105	4_40 Po (2)
	0.172	0.154	4_22 Po (3)			0.219	0.284	4_22 Po (3)
		0.057	4_16 Po (4)				0.000	4_16 Po (4)
		0.101	4_11 Po (5)				0.045	4_11 Po (5)
		0.308	4_25 Po (6)				0.345	4_25 Po (6)
	0.206	0.325	4_26 Po (7)			0.219	0.425	4_26 Po (7)
		0.069	4_12 Po (8)				0.040	4_12 Po (8)

Table 4.28 FC values for same catchments as above, but with window size 4800m . The larger window/mask cells used, give lower metric values, again with highest values for highest orders of catchments.

The visual appearance of FC profiles for different types of catchments are shown in Figure 4.33 and Figure 4.34 below. Only values for window size up to 9600m are used, because most catchments have zero FC values at 19200m, and many have so few cells that calculations become statistically uncertain. The catchments (contributing to Po) in the northern part of the area generally have higher FC values, but they also decrease more rapidly with window size. The crossing of the curves for 4_22 and 4_26 from FMERS data indicates that catchment 4_22 has forest patches scattered across the landscape with typical distances between 1.2 and 2.4 km (the steepest part of the curve), while catchment 4_26 further down the valley has larger and more compact forest patches – or larger areas where no forest is found.

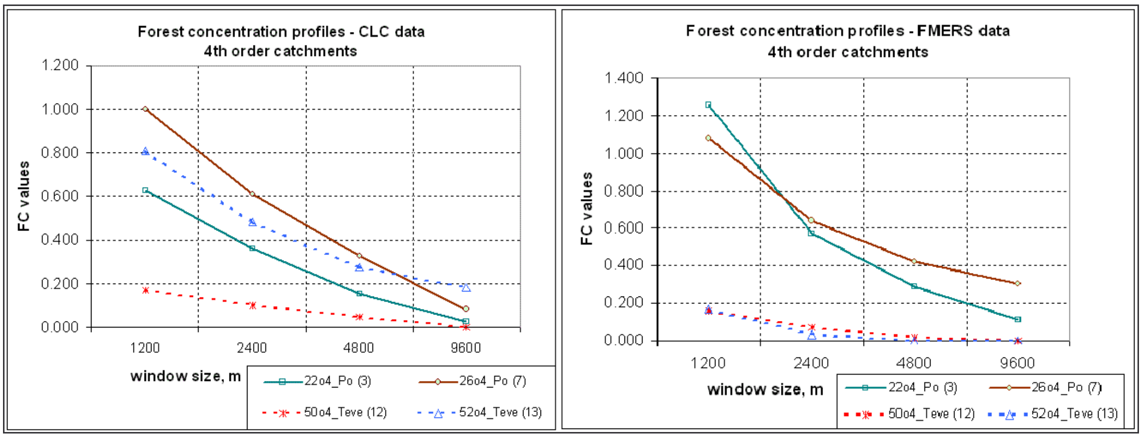

Figure 4.33 CLC and FMERS inputs compared for creation of FC-profiles of selected catchments in northern and middle Italy.

The selected administrative regions also show differences for the shape of the FC curves in Figure 4.34, but there is good agreement between the two different data sources. There are marked differences for Liguria, where the forest cover in the CLC maps is so dense that hardly any non-forest cells are found (when they *are* found in the FMERS map it can however be due to cloud cover), and for Lazio, where the CLC map has larger non-forest areas and thus higher FC values at small window sizes.

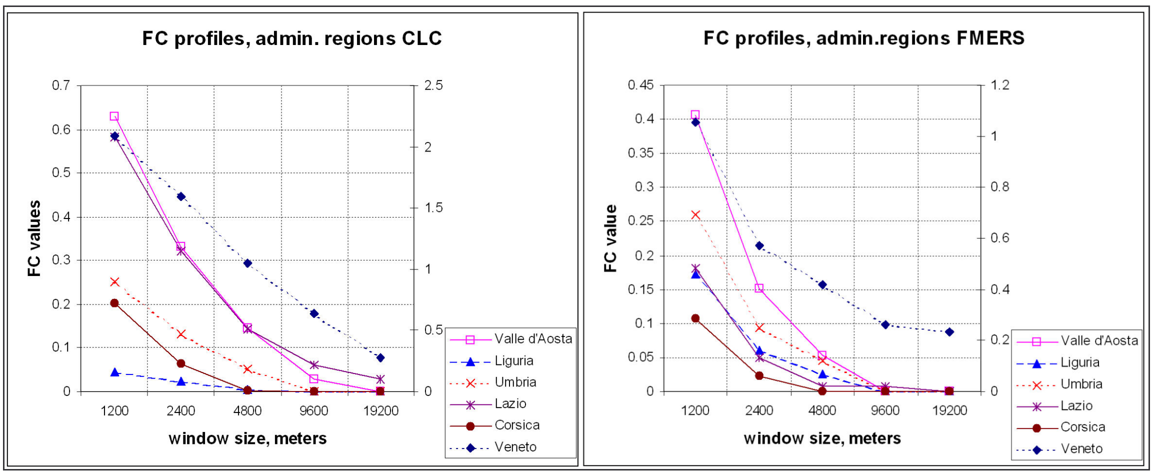

Figure 4.34 CLC and FMERS inputs compared for creation of FC-profiles of selected administrative (NUTS-level 2) regions. Note that for both data sets the curve for Veneto corresponds to the 2nd y-axis.

Generally, it seems that CLC data yield FC-curves of more different shapes and placement, thus making it easier to characterise and distinguish between regions. Again it is the more scattered nature of the FMERS data that is reflected in spatial metric values.

4.5.5.4 Regressions between metrics derived from different data sources within selected areas

The calculations made here are basically repetitions of what was done in section 4.5.4 where correlations between values from the two different input types were made for a forest mask of the entire study area, with results summarised in Table 4.19, Figure 4.24 and Figure 4.25. However, here the regressions are performed for subsets of the study area. The subsets are defined in two different ways, namely delineation by terrain and by following man-made borders. The SHDI diversity metric and the Matheron index of fragmentation are used as examples, and for comparison with the profiles that illustrate how metrics values vary with window size.

Catchment – SHDI	1200 m	2400 m	4800 m	9600 m	19200 m		Admin. - SHDI	1200 m	2400 m	4800 m	9600 m	19200 m
4_42 Po (1)	0.184	0.059	0.256	0.33	0.096		Veneto	0.291	0.449	0.563	0.622	0.675
4_40 Po (2)	0.254	0.219	0.39	0.522	0.58		Lombardia	0.219	0.383	0.476	0.519	0.538
4_22 Po (3)	0.147	0.077	0.257	0.427	0.552		Piemonte	0.255	0.418	0.505	0.622	0.682
4_16 Po (4)	0.338	0.114	0.283	0.826	N/A		Valle d'A.	0.245	0.446	0.464	0.343	0.442
4_11 Po (5)	0.268	0.192	0.278	0.614	0.874		Emilia R.	0.266	0.432	0.538	0.549	0.549
4_25 Po (6)	0.369	0.263	0.486	0.713	0.791		Liguria	0.094	0.143	0.283	0.485	0.709
4_26 Po (7)	0.267	0.19	0.368	0.583	0.744		Toscana	0.221	0.337	0.314	0.325	0.432
4_12 Po (8)	0.212	0.356	0.531	0.311	0.367		Marche	0.34	0.537	0.611	0.547	0.665
4_47 Tosc (9)	0.24	0.344	0.539	0.19	0.217		Umbria	0.203	0.263	0.228	0.203	0.293
4_49 Tosc (10)	0.263	0.489	0.644	0.234	0.254		Abruzzo	0.067	0.096	0.114	0.15	0.222
4_48 Teve (11)	0.206	0.411	0.643	0.069	0.103		Lazio	0.287	0.417	0.435	0.378	0.069
4_50 Teve (12)	0.108	0.291	0.472	0.046	-0.429		Corsica	-0.02	-0.069	-0.146	-0.219	-0.278
4_52 Teve (13)	0.326	0.104	0.163	0.614	0.209		average	*0.206*	*0.321*	*0.365*	*0.377*	*0.417*
Average	*0.245*	*0.239*	*0.408*	*0.421*	*0.363*		st.dev.	*0.106*	*0.179*	*0.219*	*0.243*	*0.298*
st.dev.	*0.075*	*0.135*	*0.156*	*0.247*	*0.369*		coeff.var.	*0.513*	*0.557*	*0.6*	*0.645*	*0.714*
coeff.var.	*0.305*	*0.563*	*0.383*	*0.587*	*1.015*							

Table 4.29 Correlation coefficients for agreement between CLC and FMERS based values of the SHDI diversity index at different output cell (window) sizes for selected geographic areas. Highest metrics values are highlighted in yellow, lowest values in blue.

Table 4.29 shows that the SHDI values have large differences between the different areas and correlations values somewhat fluctuating with respect to window size, especially for the catchment regions. This is contrary to what is observed for the entire study area. The average values for the administrative regions (representing a larger part of the maps than the catchments), however have values similar to the multiple R values at the same window sizes in Table 4.19. The regions with highest forest concentration (FC values) and lowest fragmentations or dense forest cover seem to have the best agreement between CLC and FMERS data. A notable exception is the Corsica region, where the negative correlation coefficients indicate strong disagreement between the data sources as to where the most diverse forest areas are found. The fluctuations can be attributed to random effects, such as the influence of where the windows happen to be placed in the landscape. The higher correlation coefficients for large windows do not necessarily mean that they are more reliable, this is because, with a small number of samples or output cells, confidence intervals are correspondingly narrower. Thus the potential for establishing relations or predictions of metrics values from one data type to another based on smaller areas remains doubtful. It also remains to be examined whether strata such as botanical or climatic zones or based on terrain/altitude give better agreements.

Catchment - Math.	1200m	2400m	4800m	9600m	19200 m
4_42 Po (1)	-0.039	0.261	0.181	0.34	0.046
4_40 Po (2)	0.006	0.451	0.509	0.486	0.542
4_22 Po (3)	-0.112	0.28	0.377	0.311	0.272
4_16 Po (4)	-0.132	0.544	0.498	0.509	N/A
4_11 Po (5)	-0.009	0.449	0.466	0.348	0.595
4_25 Po (6)	0.014	0.52	0.643	0.661	0.858
4_26 Po (7)	-0.092	0.504	0.577	0.57	0.554
4_12 Po (8)	0.144	0.345	0.407	0.726	0.91
4_47 Tosc (9)	0.117	0.376	0.285	0.68	0.8
4_49 Tosc (10)	0.199	0.316	0.305	0.816	0.948
4_48 Teve (11)	0.128	0.251	0.207	0.744	0.742
4_50 Teve (12)	0.121	0.182	0.145	0.542	0.407
4_52 Teve (13)	-0.052	0.465	0.429	0.275	0.539
Average	*0.023*	*0.380*	*0.387*	*0.539*	*0.601*
st.dev.	*0.109*	*0.117*	*0.155*	*0.180*	*0.271*
coeff.var.	*4.817*	*0.308*	*0.400*	*0.335*	*0.450*

Admin. - Math.	1200m	2400m	4800m	9600m	19200m
Veneto	0.101	0.015	0.091	0.179	0.281
Lombardia	-0.083	0.182	0.367	0.54	0.717
Piemonte	-0.083	0.227	0.441	0.579	0.629
Valle d'A.	0.183	0.129	0.207	0.324	0.461
Emilia R.	0.043	0.315	0.548	0.709	0.814
Liguria	-0.029	0.205	0.31	0.417	0.438
Toscana	0.031	0.402	0.581	0.704	0.747
Marche	-0.1	0.276	0.552	0.727	0.795
Umbria	-0.058	0.454	0.71	0.791	0.858
Abruzzo	0.019	0.324	0.51	0.611	0.502
Lazio	0.208	0.134	0.225	0.247	0.363
Corsica	-0.01	0.113	0.248	0.339	0.488
average	*0.019*	*0.231*	*0.399*	*0.514*	*0.591*
St.dev.	*0.102*	*0.128*	*0.187*	*0.207*	*0.193*
coeff.var.	*5.505*	*0.553*	*0.468*	*0.402*	*0.327*

Table 4.30 Correlation coefficients for agreement between values of the Matheron index, based on CLC and FMERS data, at different output cell (window) sizes for selected geographic areas. Highest metrics values are highlighted in yellow, lowest values in blue.

Table 4.30 shows that, on average M values have higher correlations for the regions used here than for the entire study area (compare Table 4.19). As expected and following the large differences in the structure and composition of the data sets, as described in the above sections, there are marked differences between the regions, and no clear pattern of zones with high correlations emerge. Surprisingly, the Corsica region has positive correlation values for this forest structure metric, so the problem of agreement lies more with composition than with extent and texture of forest across the landscape. See also, for comparison Table 4.26 with description of forest composition for Veneto and Umbria.

4.5.5.5 Test for variability

Table 4.31 and Table 4.32 report the average of the coefficient of variation for each of the spatial metrics *within* administrative and catchment regions respectively. The purpose of comparing the values is to examine whether one of the delineation approaches produces more homogenous regions in terms of metric values.

catchmts.	CLC_COV	CLC_PPUN	CLC_PPUNB	CLC_SHDI	CLC_M	CLC_SqP	CLC_Alt
1200	0.625	0.582	0.629	1.291	0.618	0.604	0.603
2400	0.680	0.363	0.277	0.995	0.500	0.356	0.653
4800	0.696	0.234	0.152	0.757	0.422	0.193	0.682
9600	0.670	0.157	0.095	0.572	0.353	0.085	0.728
Admin.							
1200	0.616	0.567	0.634	1.353	0.627	0.608	0.640
2400	0.663	0.345	0.284	1.016	0.504	0.357	0.696
4800	0.690	0.225	0.164	0.783	0.426	0.196	0.760
9600	0.694	0.157	0.106	0.608	0.370	0.097	0.799

Table 4.31 Mean values of coefficients of variation for selected metrics from the CLC data and elevation from DTM, average values from the 13 4th level catchments and the 12 regions.

catchmts.	FM_cover	FM_PPUN	FM_PPUNB	FM_SHDI	FM_M	FM_SqP	FM_alti
1200	0.673	0.625	0.612	0.847	0.598	0.693	0.661
2400	0.743	0.463	0.297	0.640	0.468	0.417	0.663
4800	0.725	0.372	0.200	0.478	0.375	0.194	0.690
9600	0.674	0.312	0.141	0.360	0.303	0.078	0.719
Admin.							
1200	0.648	0.624	0.650	0.797	0.653	0.730	0.682
2400	0.737	0.488	0.301	0.624	0.521	0.453	0.744
4800	0.727	0.401	0.206	0.454	0.431	0.222	0.783
9600	0.703	0.350	0.156	0.335	0.366	0.102	0.795

Table 4.32 Mean values of coefficients of variation for selected metrics from the FMERS data and elevation from DTM, average values from the 13 4th level catchments and the 12 regions.

When comparisons are made between values of metrics from the same data source and at the same window size (within each table), no clear differences or trends emerge. Thus, it can not be concluded that one or the other regionalisation approach produces more homogenous regions with smaller internal variance of the metrics values. The decreasing values of SqP variance with increasing window size can be attributed to the nature of the metric (more separate patches in larger windows give values closer to 1) and not to an actual smaller difference in forest structure between the windows. Note however the differences in variability of the patch count metrics, where FMERS maps have the highest values and of the SHDI metric, where CLC maps have the higher values. This is also seen from Figure 4.18 and Figure 4.19, though the variance values there are calculated only for each output cell and its immediate neighbours.

4.5.5.6 Test for agreement - CLC-FMERS

This final sub-section examines the results derived at regional level, comparing the relative values per region to examine whether they give the same general image of the study area (i.e. will thematic maps of a given spatial metric look the same, when derived from CLC and FMERS data?).

When the 12 administrative regions are compared, the critical value of **observed t** (Spearman's rank transformed to t-distribution values, assuming a two–sided distribution) is 2.201 for the rank correlation at 5% confidence interval and 1.796 at 10%, corresponding to coefficients of +/- 0.6354 and +/-0.5185 respectively. When the 13 catchment regions are compared, the critical value of observed t is 2.179 at 5% confidence and 1.728 at 10%. The values in Table 4.33 and Table 4.34 below are the rank correlations, with indications of possible significance. Note that some of the correlations are negative. Though not significant, these values indicate strong disagreement between the CLC and the FMERS data. It is no surprise that this is seen for the SHDI diversity metric as calculated on admin. regions, where the CLC data generally give highest values in the northern regions, and FMERS data give highest values in middle Italy. In this test the administrative regions have 12 instances of significant agreement, hereof one at 10% confidence level, the catchments have 16, hereof five of them at 10% confidence level, so it seems that with this approach, catchments are more effective for mapping of spatial metrics.

Italian admin. regions	window size *n=12*				
Metrics	**1200**	**2400**	**4800**	**9600**	**19200**
Cover	0.510	0.727**	0.797**	0.734**	0.720**
PPUN	-0.119	-0.077	-0.021	0.112	0.224
PPUN_B	0.136	0.168	0.549*	0.703**	0.797**
Math.	0.364	0.273	0.259	0.224	0.280
SHDI	-0.517	-0.517	-0.385	-0.329	-0.378
SIDI	-0.105	0.035	0.108	-0.017	-0.332
SqP	0.140	0.070	0.080	0.262	0.367
FC	0.811**	0.755**	0.804**	0.776**	0.781**

Table 4.33 Spearman's rank correlation coefficients for agreement between spatial metrics from CLC and FMERS forest maps, extracted for the 12 northernmost administrative regions in Italy. ** indicate significance at 5% probability level, * at 10% level, assuming a two-sided Student's t-distribution.

Italian catchments	window size *n=13*				
Metrics	**1200**	**2400**	**4800**	**9600**	**19200**
Cover	0.549*	0.738**	0.761**	0.846**	0.755**
PPUN	0.234	0.475	0.703**	0.569*	0.529*
PPUN_B	-0.092	0.443	0.635**	0.620**	0.643**
Math.	-0.069	0.275	0.623**	0.503*	0.595*
SHDI	-0.086	0.003	0.132	0.140	0.063
SIDI	-0.092	-0.169	-0.006	0.114	-0.066
SqP	0.253	0.220	0.349	0.463	0.169
FC	0.658**	0.435	0.413	0.615**	

Table 4.34 Spearman's rank correlation coefficients for agreement between spatial metrics from CLC and FMERS forest maps, extracted for 13 selected 4th level catchments in northern and middle Italy.

The difference between the two regionalisation approaches is especially pronounced for the patch count metrics, where the catchments show good agreement for the PPUN values at larger window sizes, but not so for the admin. regions. For catchments the Matheron index value show agreement at window sizes of 4800m and above, for admin. regions neither M nor SqP show significant agreement, still M seems to be the better choice for an indicator of forest fragmentation.

The results here, along with the analysis for variability indicate that for "thematic" mapping of spatial metrics, the smallest window sizes should be avoided, if the resulting pattern should be compared with metrics from other data sources. For the catchments, the fragmentation metrics of PPUN_B and the Matheron index have higher rank correlation at 4800m window size, than at 2400 or 9600m, thus a window size of around 5km seems appropriate for mapping of forest structure. In terms of pixels that is 50*50=2500 at 100m resolution or 25*25=625 at 200m resolution.

In general, the metrics are seen to behave very differently in the different regions, administrative as well as catchments. Local circumstances rather than general scaling properties dominate, and for the diversity metrics a north-south gradient of values is visible.

4.6 Discussion of results from application of Moving-Windows

In this section, the findings from the previous section are summarised, following the structure of the results section. It is here intended to interpret the results and put them into a broader context. Then the methods used are evaluated.

4.6.1 Evaluation of results

1) Responses to window size

The examination of the metrics' response to window size show a similar behaviour for the two data sets, even though the structural metrics Matheron index and PPU have markedly different numerical values, i.e. higher values for FMERS data. Also the compositional metrics SHDI and SIDI have higher values for FMERS, confirming that (according to this map) forest patches are smaller, more scattered and the classes more interspersed. With one exception (PPUN_B which initially increased for the FMERS data) the metrics values increased or decreased steadily with window size. The diversity metrics and the SqP metric constantly increase with window size, the other metrics constantly fall. Patch count metrics are known to vary with window size, but the normalisation proposed here seem to restrain that. A remarkably good agreement was found between the forest cover-background patches curves for the two data sets. Also the SqP metric vary with window size, an effect that is so far not accounted for, but quantification of the influence of extent (working with controlled/artificial landscape maps) could prove useful.

The changes in metrics value, variability and correlations with extent is in line with the observations made by Riitters *et al* (2000), of the changing fragmentation related characteristics with increasing window sizes. The relatively rapid changes in metrics values and correlations at small window sizes point to the relevance of the observation by O'Neill *et al* (1996), that the window/extent must be at least 2 to 5 times larger than the (forest) patches in order to give representative values.

2) Variability and autocorrelation

Regarding standard deviation for an output cell and its eight nearest neighbours (3*3 window), examination of variability and autocorrelation of the metrics show better agreement between the st.dev. values from CLC and FMERS data, than for the metrics values *per se*, in terms of response to changing extent (Figure 4.16 and Figure 4.17 are very similar, compared to the response curves in Figure 4.11). For the cover metric, window sizes with low standard deviation correspond roughly to sizes with high autocorrelation as expressed with Moran's I. The latter however show more distinct peaks and troughs, allowing recommendations for making maps of forest structure, and will surely provide more characteristic profiles of forest structure in separate and different study areas. The large area of study makes it hard to distinguish any characteristic forest/landscapes from the local variability values, as it was otherwise intended, for selection of appropriate window sizes for M-W based maps of forest structural metrics. Identification of such characteristic scales will probably require studies by region or stratum and using higher resolution data as well.

3) Relationships between different metrics from one data source

Calculation of the correlations between the different metrics for each data type and (geographic) window size provides interesting insight into the behaviour of the metrics, as well as of the scale of structure and processes in the landscape, and the similarities and differences between the two data sets. Given the large number of observations (output cells) even for large window sizes, almost all correlations are significant. The development of correlations between the metrics of cover and fragmentation (Table 4.17 and Table 4.18) show that the same combination of metrics cannot necessarily be used to describe an area at different resolutions or window sizes. The two diversity metrics SHDI and SIDI are so highly correlated that very little extra information is provided by reporting both. If a group of metrics to represent landscape properties should be selected, it could for instance be, for window size 4800m: cover, SIDI and the Matheron index. They represent forest fraction, composition and structure and are only weakly correlated (Table 4.11 and Table 4.12).

4) Correlations between similar metrics from different data sources

The correlations between the values from the two different image types generally increase with window size. This is to a smaller extent due to gradual elimination of possible bias from a geographic co-registration of the images that is not sufficiently precise[21]. The increase also reflects a gradual softening of the MW-output images, as small areas with special structure (in one of the image types) become integrated with their surroundings. Across scales, the cover metric shows the best agreement between the two map types, followed by the patch count metrics and the Matheron index. The diversity metrics and SqP show low correlations even at large window sizes, the former reflecting large-area differences in (classification of) forest composition that make it hard to substitute on map type with the other, the latter showing that the Matheron index is to be preferred for comparisons of forest fragmentation etc. between data sources.

5) Comparison of regionalisation approaches

Extraction of metric values for subsets of the study area in the form of catchments and administrative regions proved interesting and illustrative. At all the window sizes used,

---

[21] Here the image were co-registered to the Corine projection using the definitions from the image processing software (ENVI) – perhaps for large areas and different data sets as in this exercise, GCP-collection and pixel-to-pixel comparison is needed.

average metrics values clearly varied. The set of regions (13 4th order catchments and 12 administrative regions) was not large enough to identify north-south or altitude controlled trends, but it was possible to explain extreme values with properties of the maps and the geographic reality behind them. The region approach allowed calculation of the new forest concentration (FC) metrics, which turned out to be a good descriptor of the general forest structure, but it could also apply to other land cover, vegetation or habitat types or even urban classes or population concentration. The metric is well suited for graphic reporting in the form of FC-profiles. A hierarchical structure for reporting the initial metrics and the FC values in table form seems useful. Regressions between the two data sources was performed within the regions and results for the SHDI and Matheron metrics presented. As expected low correlation values were found for small windows, and higher but varying values for larger windows (with fewer pixels to supply values). The M index on the average showed a higher correlation coefficient than for similar window sizes for the entire area, especially within the catchments. Calculations of variability within the regions showed little differences, and the pronounced differences found were related to data type rather than to region type. Thus, the recommendations given by amongst others Apan *et al* (2000), Paracchini *et al* (2000), Vogt *et al* (2003) for use of catchments/watersheds or (more locally) headwaters as reference units for landscape metrics could not be confirmed in this study.

4.6.2 Evaluation of methods

Concerning the methods used in this chapter, the use of special IDL-scripts to carry out the M-W calculation proved practical, as it has been possible to modify the scripts after initial calculations, for instance to exclude background pixels from calculations of forest diversity and to output also the number of background patches. The process of getting from input images to final statistic was however quite tedious, as illustrated in Figure 4.9 on page 128. Work is ongoing to make scripts that output the M-W results as binary map files in Idrisi raster format – this will also save disk space, as the current comma separated text format can result in very large files for small window sizes[22]. The creation of thematic maps from Corine Land Cover was simple and straightforward, the considerations mostly being on which classes to include and how to label them (Table 4.1).

Implementing and using the M-W approach has provided many useful results and insights, and highlighted some general considerations and problems related to the calculation of spatial or landscape metrics. For instance whether or not to include background pixels in metrics calculations (typically for the ‘total number of pixels in window’ parameter) or how to handle non-forest land. In this implementation no distinction was made between background and ‘other land’, and that partly explains the decrease in forest cover percentage with increasing window sizes (as water/sea became included in total window area). The definition and use of the PPUN and PPUN_B metrics for characterising patch density has proved feasible and these metrics are used along with the structure and compositional metrics in description of the total landscape structure. The results, as expressed in the appearance of the output maps and the extent-variance curves confirm the observations by Eiden *et al* (2000) that results vary strongly with window size and that too large windows smooth out potentially useful information.

The creation of scaling profiles or *scalograms* for metrics following window size has proved a useful tool for the understanding of scale (or in this chapter rather *extent*) effects on spatial metrics values. Also calculation and graphical illustration of variance and autocorrelation of the M-W outputs has helped in understanding the effects associated with this approach. Woodcock and Strahler (1987) proposed that graphs of local variance in images as a function of spatial resolution may be used to measure spatial structure in images. Here the objective was to measure spatial structure of maps of spatial metrics, and the results were not as distinct

[22] Output as Idrisi images is possible in the latest version of Fragstats for Windows, where M-W has been implemented, although with step fixed at one pixel, which results in long calculation time and large output files. The software can be downloaded for free from: http://www.umass.edu/landeco/research/fragstats/fragstats.html (accessed 15/11 2003).

as in the examples used (ibid, figure 2 and 4), and for this type of data, graphs of extent versus autocorrelation seem to be more useful. Plots of the coefficient of variance, a normalised value, against window size seems to be more informative than plots of standard deviation against window size, as the former approach produces more distinct response curves (compare Figure 4.16 with Figure 4.18 and Figure 4.17 with Figure 4.19). The values of Moran's I were calculated in Idrisi, and it might save some time, files and space if it becomes integrated in the IDL-scripts that provide the metrics values.

Regression between images at the pixel level, in order to test agreement between calculation approaches (here: choice of input data) turned out to be simple and fast, with most of the work load lying in the preparation of spreadsheet files for creation of the graphic representations. The regressions were preformed using forest (presence) masks, following the "OR" approach, in order to make sure that all possible forest cells/windows were included in the calculations, even if some of them have zero values. It was assumed that use of the "AND" approach would be too restrictive, though it would be interesting to compare results derived with these two approaches.

The extraction of spatial metrics values for administrative regions and catchments was straightforward, following standard GIS and image processing techniques. This was also done for the creation of forest concentration (FC) profiles and hierarchical tables for reporting the values at different levels, in this case hydrological, but it could also have been administrative levels. The combination of metrics calculation within M-Ws and reporting of average and variance values for physical or administrative regions makes it possible to eliminate the influence of region size, which would for instance make patch count and richness metrics less useful. The agreement between the data sources within the study area at region level was examined using rank correlation, which proved useful in distinguishing metrics and window sizes suitable for comparison (in the form of thematic maps).

In this chapter, some interesting results for the structure of the forest landscape of the study area have been found, especially with regard to indicators for reporting at regional level, and most of the methods that were introduced and proposed in relation to M-W analysis of landscape structure have proven feasible.

4.7 Conclusions – implications for forest monitoring

The CLC dataset appears to be a useful base for a forest map at 100m pixel size, with distinct forest patches and a realistic distribution of forest types following terrain and climatic gradients. The FMERS dataset give a somewhat different picture for the sub-continental forest map at 200m pixel size, but the results here show good agreement with the CLC map for basic spatial properties such as forest cover and concentration and reasonably good agreements for structural properties such as Matheron fragmentation index and the PPUN metric.

Working with two different data sources, a suite of spatial metrics and a number of different window sizes has made it clear that, there are no obvious ‘best’ choices of metrics and window sizes for summarising and illustration forest structure and diversity. The selections must depend on the properties of the input data (particularly spatial and thematic resolution) as well as the purpose of the M-W analysis (analytical, illustrative or auxiliary to further image processing). Then inspection of the extent-variance curves and of the correlations between metrics values can help the user to choose the metrics images with the highest information content and least redundancy.

The application of M-W methods could be seen as a way of addressing the MAUP as it appears in the use of different reference units for reporting of landscape metrics. At least for production of maps of the metrics, the potentially distorting effects of region size and shape are avoided. The grainy or edgy appearance of the outputs at large window sizes could be avoided, if the results were smoothed following the approach described by Eiden *et al* (2000) or produced with a software similar to Fragstats for Windows, where the step of the window is equal to input pixel size.

In summary, the set of methods described here provide an approach for assessment of structural and compositional properties of forests over large areas from medium-resolution satellite imagery (100-200m grain size), comparison between regions and monitoring of environmental conditions, given the availability of regularly updated images or maps. In the following chapter, a thematically detailed data set on land use-land cover delivered in vector format will be compared with satellite land cover maps at a higher spatial resolution than used here, namely 25 metres. These satellite data based maps will represent the 'monitoring' approach, in contrast to the 'mapping' approach of the Danish Area Information System and the Corine Land Cover database.
