MULTI-SCALE APPLICATION OF SPATIAL METRICS FOR QUANTIFYING FOREST SPATIAL STRUCTURE AND DIVERSITY FROM CORINE LAND COVER AND FMERS-WIFS RASTER DATA

N.C. Nielsen[a] and G.A. Blackburn[b]

[a] Center for Historie, Syddansk Universitet, Niels Bohrs vej 9, DK-6700 Esbjerg, mail: ncn@hist.sdu.dk
[b] Department of Geography, Lancaster University, e-mail: alan.blackburn@lancaster.ac.uk

ABSTRACT

In this paper, the moving-windows approach to calculation and analysis of spatial metrics is tested with particular focus on forest mapping. The influence of window size on average metrics values, agreement between values from different EO-based data sources and local variance of metrics values is analysed using standard statistical approaches. Forest Concentration Profiles, based on forest-non forest masks for moving windows is presented as an approach to characterise structure and distribution of forest over certain areas of interest.

Keywords: Forest, Spatial Metrics, Structure, Diversity, Moving windows, Scale.

1 INTRODUCTION

Spatial metrics are commonly employed for characterising landscape structure and can be used as descriptors of ecological states and processes as well as indicators of sustainable land use and forest management. Such metrics can be derived through processing of satellite images and from existing digital map data stored in Geographical Information Systems. Ideally, the metrics should be insensitive or predictable with respect to scale changes and at the same time appropriate for description of landscape diversity and structure [1]. Furthermore, the metrics used should be uncorrelated, thus ensuring that they describe different aspects of landscapes.

The methods used and results presented here stem from a study on development of criteria and indicators for sustainable forestry, based on Earth Observation data, as part of the Eurolandscape project at the Joint Research Centre, Ispra, Italy. The study area consists of Northern Italy and smaller parts of neighbouring countries. The goals of the study were threefold: to evaluate different RS-based data sources for mapping and evaluation of forest structure and diversity, to select and evaluate a suite of spatial metrics and finally to implement a moving-windows (MW) approach as envisioned by O'Neill at el [2] and demonstrated by Häusler et al [3] and evaluate the scale effects of particularly window-size on metrics values and distributions.

2 DATA

Data used include raster and vector maps, based on automatic classification procedures as well as manual interpretation of imagery.

The images were fitted geographically to a window of size 700 by 500 km, in the Corine Land Cover (CLC) projection (spherical Lambert Azimuthal), with upper left corner at (-200.000, -200.000) corresponding to app. 6.5 deg. E, 46.2 deg. N.

2.1 CORINE LAND COVER

Data from the CLC are used here in the form of raster images with a pixel size of 100m. 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 currently becoming available through the web site `http://terrestrial.eionet.eu.int/CLC2000/`). This makes CLC data useful for monitoring purposes and comparisons across Europe [4]. 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 [5]. 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 1.

2.2 WIFS-FMERS

The forest map derived 'directly' from EO data used here is based on a mosaic of WiFS images from the IRS 1-C satellite. The map was produced by VVT-Finland on contract to SAI, and the steps of the image preparation and processing are described in [6]. 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 been defined in accordance with the FIRS nomenclature system [7]. The spatial resolution of the original images is 188m pixel size, the mosaic was re-sampled to a pixel size of 200m.

2.3 ANCILLARY DATA

A Digital Elevation Model (DEM) with grid cell size 250m was available for the study area. It was re-sampled to the cell sizes of the respective images, using bi-linear interpolation. Ancillary vector data were used to extract information from the metrics images, using the statistical functionalities of the image processing software.

The data set described in [8], which has been developed at JRC, was used to delineate watersheds serving as reporting units for metrics values. A subset of watersheds were extracted for the upper Po valley and for the entire Tevere (the Tiber) catchment area, supplemented with two 4th order watersheds 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. 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.

3 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 derive metrics from data sets of raster images are presented.

Table 1 Matching CORINE and FMERS forest cover classes for the current study.

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

3.1. SELECTING A SUITE OF 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 [9], for each class and for combined forest layers
- the Square Patch index (SqP) – for forest-non forest structure
- Patches Per Unit area (PPU), both following Frohn's definition [1] and a modified, 'normalised' version that accounts for changing window sizes.

The edge pixel percentages and proportions area are used only as intermediate steps to get to the M and SqP metrics, and for development and testing purposes, though they have the potential to function as indicators in their own right. The structural indices are defined as follows :

$$M = 10^* \frac{\text{number of edges between forest and other LC type pixels}}{\sqrt{(\text{number of forest pixels})}^* \sqrt{(\text{total number of pixels})}} \quad (1)$$

the scaling factor of 10 has been applied in order to assure that typical output values will be in the interval from 1 to 10.

$$\text{PPU} = \frac{m}{(n^* \lambda)} \quad (2)$$

where m is the total number of patches (in the window), n is the total number of pixels in the area of interest (window) and λ is the scaling constant equal to the area of a pixel, dependent on the extent of the area of interest the unit of λ can be m^2, ha or km^2.

$$\text{SqP} = 1 - \frac{4^* \sqrt{A}}{P} \quad (3)$$

where A is the total area of all pixels and P is the total perimeter of all pixels belonging to the land cover class of interest in the area (window). The theoretical values for this index ranges from 0 to 1; 0 is for the case of the landscape mask element (the forest) consisting of one large square; if it is made up of more patches, the values will be > 0; the value will approach 1 when the cover type becomes more scattered over the landscape.

The diversity metrics used are Number of class types (richness), Simpson's diversity and 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 follows [10]:

$$SIDI = \sum_{i=1}^{n} P_i^2 \quad (4)$$

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, is based on information theory, expresses the 'bandwidth' needed for description of a system and thus the 'disorder' or distance from predictability of it [9]. The index defined as:

$$SHDI = -\sum_{i=1}^{n} (P_i^* \ln P_i) \quad (5)$$

The maximum value of the SHDI for a landscape with n classes is simply ln(n), and the minimum value 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 not 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 should not have taken place, and the various natural and agricultural land cover types preserved as separate classes. 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 'matrix' or background class).

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" [11]. 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 patches per km^2. In the study it was investigated whether it was possible to compensate – 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:

$$PPUN = \frac{NP - 1}{A} + \frac{1}{A_{min}} \tag{6}$$

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. 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 [12]. 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.

3.2. MOVING WINDOWS

The MW calculations were carried out using scripts for the IDL data processing software. The scripts allow modification of the window and the step size, as part of which overlap between windows is possible. The main difference between this implementation and the one used in, for instance, Fragstats for Windows and various kernel operations in image processing software, 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.

MW-operations can thus be seen as a coarsening filter and/or as a way of compressing information on landscape structure to images at lower spatial resolution. The practical issues concerning implementation of MW are discussed in more detail in chapter 4 of [13], where examples of the IDL scripts constitute Appendix 1.

3.2.1 Masking for inclusion/exclusion of cells

Values of spatial metrics are reported as average values for the parts of the entire study area that are covered by a mask (bit-map image) showing the presence of more than one percent forest cover. No further rules are applied whether to incluede a part of the image in the calculations or not. These mask-images proved useful in further calculations, not only of metrics and variance values, but also for characterisation of spatial structure over larger regions.

3.3. FOREST CONCENTRATION PROFILES

The work with image masks at different output cell sizes has 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 or region of interest. The metric is defined:

$$FC = \frac{Cover\_mask}{Cover\_landscape} - 1 \qquad (7)$$

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 RESULTS

The results of image processing and subsequent statistical analysis are presented along the lines laid out in the objectives of this paper. For brevity, the results are shown here only as plots of scalogram, variogram and correllogram types, and for illustrations of the immediate MW-outputs, the reader is referred to [13].

In Figure 1 metrics 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 metrics from the two datasets behave very much in the same way, for the shape as well as the relative position of the curves. Thus, they show similar *scaling properties*. The Matheron index and PPUN however have almost twice s high numerical values for the FMERS data. 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 become included in each window.

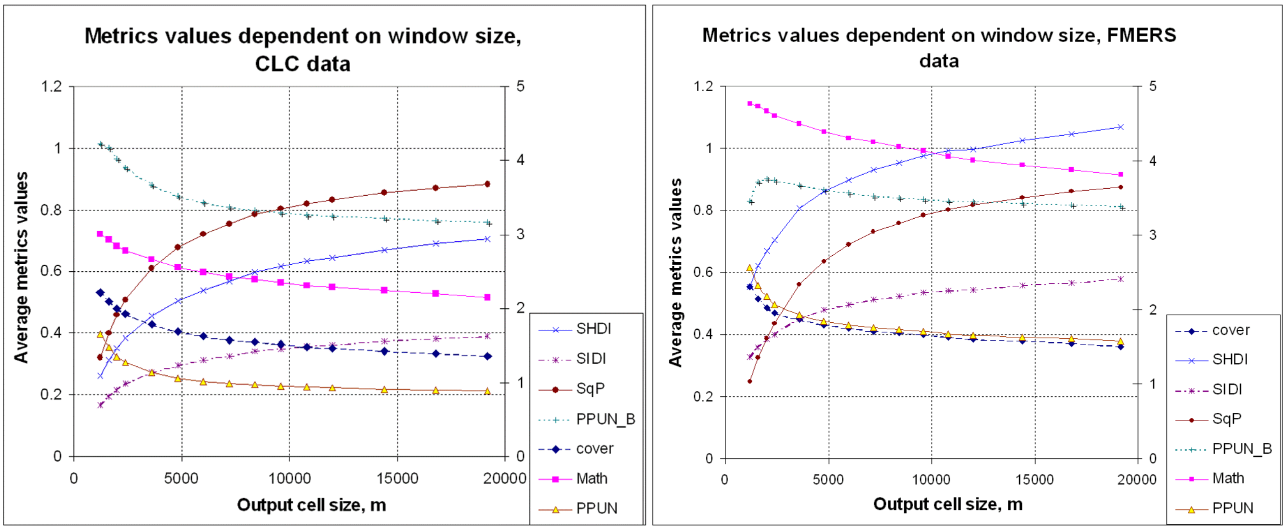

Figure 1 Metrics 'response curves' or scalograms with values plotted against window size or (sub-landscape) extent. CLC and FMERS data for the entire study area (under the forest masks). Note that M and PPUN metrics map to 2nd axis values.

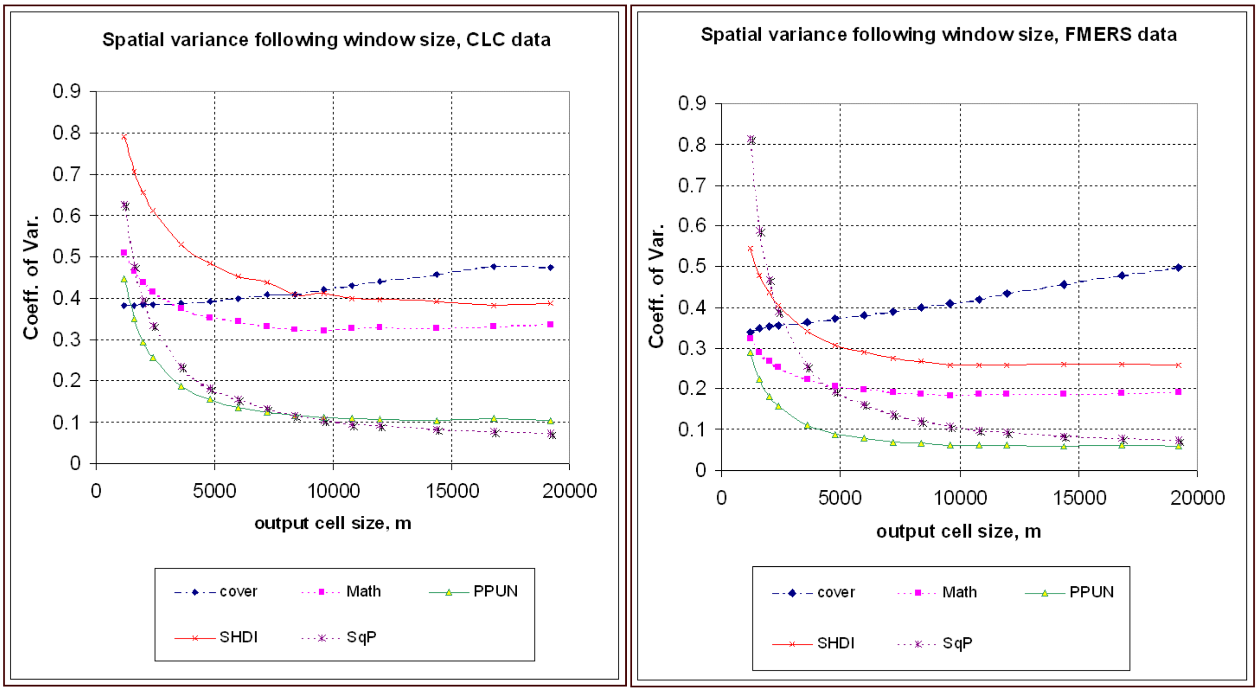

Figure 2 Local variability of CLC and FMERS data. Coefficient of variation from the suite of spatial metrics as function of the window sizes for which they are calculated.

In Figure 2, the coefficient of variation (CV) of the MW-outputs within a 3*3 pixel kernel is shown as function of window size. This has been used to assess the structure of the produced metric-maps. The CV of the Matheron metric of fragmentation increases slightly after having its minimum average values around a window size of 10 km, most clearly for the CLC data, but also visible for FMERS. The CV of cover fraction increases steadily with window size, while for all other metrics it is falling steadily.

The different metrics show quite different correlations at the same window size, as seen in Figure 3. More surprisingly they respond in different ways to the changes in window size, as expressed by the shape of the window size-correlation curves.

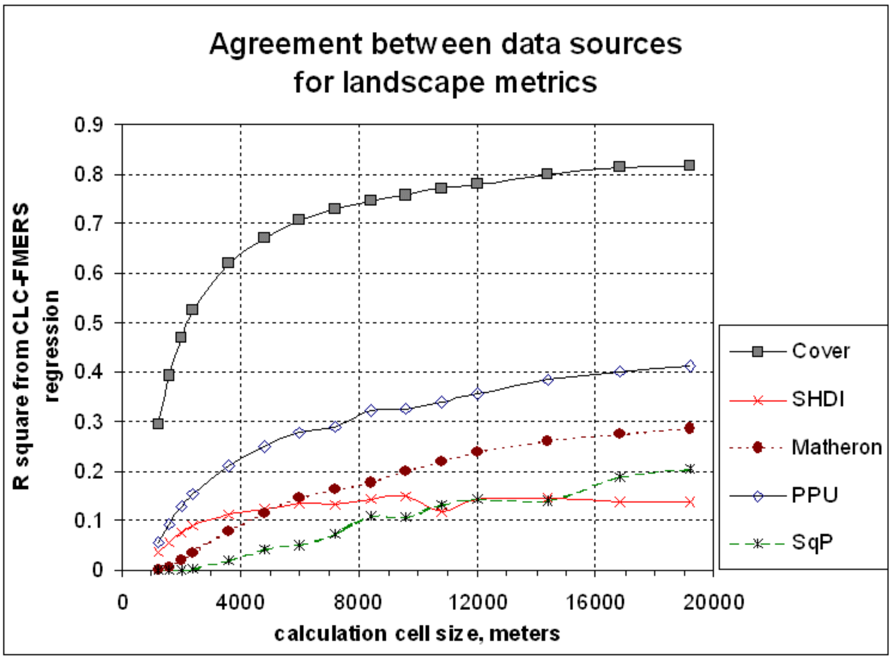

Figure 3 R-square, expressing agreement between metrics values from CLC and FMERS data, plotted against spatial extent/window size. Smallest windows are 6*6 pixels for FMERS and 12*12 pixels for CLC data, largest windows 96*96 pixels for FMERS and 192*192 pixels for CLC.

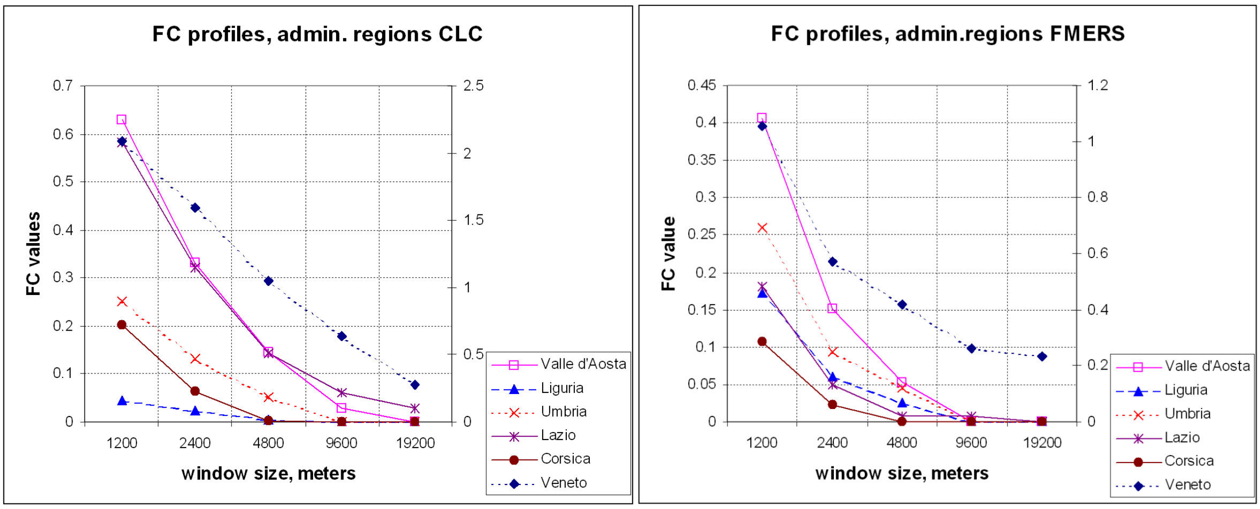

Figure 4 CLC and FMERS inputs compared for creation of FC-profiles of selected administrative (NUTS-level 2) regions. Note that for both data sets the curve for Veneto corresponds to the 2nd y-axis.

In general the increasing window size will even out differences between spatial structure as mapped in the two data sets, 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 neighboring 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.

For the shape of the FC curves shown in Figure 4, the selected administrative regions show significant differences, but there is good agreement between the two different data sources. Marked differences between the two data sets are seen for Liguria, where the forest cover in the CLC maps is so dense that hardly any non-forest cells are found (and when they are found in the FMERS map it can 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. The highest absolute values found for the Veneto region, where the forested and non-forested parts are spatially well separated, in the Dolomite mountains and the lower plains of the Po river respectively.

5 DISCUSSION AND CONCLUSIONS

Metrics from FMERS and CLC maps show good agreement 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. Differences in absolute values of the metrics at similar window sizes can be attributed to the ways in which the two data sets are derived, in particular the CLC data having a minimum mapping unit, while the FMERS map was classified on a pixel-by-pixel basis, generally producing less geographically concentrated forest and more scattered patches.

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 (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. F-C curves represent truly multi-scale summaries of the spatial structure of certain land cover types (forest) over relevant regions.

The application of M-W methods could be seen as a way of addressing the Modifiable Area Unit Problem as it appears in the use of different reference units for reporting of landscape metrics. Metrics should be normalised to facilitate comparison and cluster and Principal Component analysis. The set of methods described here provide a promising approach for assessment of structural and compositional properties of forests over large areas from medium-resolution imagery (100-200m grain size), comparison between regions and monitoring of environmental conditions, given the availability of regularly updated images or maps.

ACKNOWLEDGMENTS

The work for this thesis was made possible through a Training and Mobility grant from the commission of the European Communities, that allowed N.C. Nielsen to stay at the JRC-Ispra site from 1998 to 2001 and access the data used for the work presented here. During that period Dr. S. Folving and Dr. P. Kennedy provided valuable inspiration and discussions. Dr. B. Ejstrud at SDU in Esbjerg initiated a project on spatial analysis of cultural environment that allowed N.C. Nielsen to continue working with land cover data and spatial metrics and finish his Ph.D. thesis.

REFERENCES

1 Frohn, R.C. (1998): Remote sensing for landscape ecology - new metric indicators for monitoring, modelling and assessment of ecosystems. Lewis Publishers, CRC Press, Boca Raton, Florida.

2 O'Neill, R.V., Hunsaker, C.T., Jones, K.B., Riitters, K.H. , Wickham, J.D. , Schwartz, P.M. , Goodman, I.A. , Jackson, B.L. Baillargeon, W.S. (1997): Monitoring environmental quality at the landscape scale. BioScience vol. 47, no. 8, pp. 513-519.

3 Häusler, T., Akgöz, E., Ekstrand, S., Gallaun, H., Lagarde, I., Loefmark, M., Schardt, M., Pelz, D.R., Obergföll, P. (2000): Monitoring Changes and Indicators for Structural Diversity of Forested Areas. In: Proceedings of the IUFRO conference on "Remote Sensing and Forest Monitoring" Rogow, Poland 1-3 June 1999, eds. T. Zawila-Niedzwiecki, M. Brach. EUR 19530 EN - European Communities, Luxembourg. Pages 392-414.

4 European Environment Agency (2001): Towards agri-environmental indicators: Integrating statistical and administrative data with land cover information. EEA Topic report no. 6. EEA, Copenhagen. Available at: http://reports.eea.eu.int/topic_report_2001_06/en/Topic_6_2001.pdf (accessed 20/9 2004)

5 Bossard, M., Feranec, J., Otahel, J. (2000): CORINE land cover technical guide - Addendum 2000. Technical Report 40, European Environment Agency, Copenhagen. Available at: http://terrestrial.eionet.eu.int/CLC2000/documents/Addendum (accessed 19/9 2004)

6 Häme, T., Andersson, K., Carfagna, E., Jeanjean, H., Rapaport, P., Spence, I. (1999): Forest Monitoring in Europe with Remote Sensing (FMERS) Final report – Regional mapping (Stage 2) Executive summary. Report to: Joint Research Centre/CEO Contract number: 13105-97-07 F1ED ISP SF. Available at: http://www.vtt.fi/tte/research/tte1/tte14/proj/fmers/final/fmers_executi.htm (accessed 19/9 2004)

7 Kennedy, P., K. Kuusela, S. Folving (1997): Regionalisation and stratification of Europe's forests: the foresters perspective. FIRS project report. Available at http://www.vtt.fi/tte/research/tte1/tte14/proj/firs/firs_regions_97.pdf (accessed 19/9 2004)

8 Vogt, J.V., Puumalainen, J., Kennedy, P., Folving, S. (2004): Towards the Characterisation and Analysis of European Landscapes: Integration of Information on River Networks, Catchments and Major Forest Types. Landscape and Urban Planning vol. 67, no. 1-4. pp. 27-41.

9 Mayaux, P., Lambin E.F. (1995): Estimation of Tropical Forest Area from Coarse Spatial Resolution Data: A Two-Sted Correction Function for Proportional Errors Due to Spatial Aggregation. Remote Sensing of Environment vol. 53, pp. 1-15.

10 McGarigal, K. and Marks, B.J. (1995): FRAGSTATS: Spatial Pattern Analysis Program for Quantifying Landscape Structure. General Technical Report PNW-GTR-351. Pacific Nrothwest Research Station. August 1995. Available at http://www.fs.fed.us/pnw/pubs/gtr_351.pdf (accessed 19/9 2004).

11 O'Neill, R.V., Hunsaker, C.T., Timmins, S.P., Jackson, B.L., Jones, K.B., Riitters, K.H. Wickham J.D. (1996): Scale problems in reporting landscape pattern at the regional scale. Landscape Ecology vol. 11, no. 3, pp. 169-180.

12 Plotnick, R.E, Gardner, R.H., O'Neill R.V. (1993): Lacunarity indices as measures of landscape texture. Landscape Ecology vol. 8, no. 3, pp. 201-211.

13 Nielsen, N.C. (2004): Spatial Metrics of Structure and Diversity: Calculation from Earth Observation and map data, for use as indicators in environmental management. Unpublished Ph.D. thesis, Lancaster University (248 pp.). Available at http://www.humaniora.sdu.dk/kulturmiljoe/publikat.html