

# Segmentation of Model Localization Sub-areas by Getis Statistics

Minna Rätty and Annika Kangas

---

**Rätty, M. & Kangas, A.** 2010. Segmentation of model localization sub-areas by Getis statistics. *Silva Fennica* 44(2): 303–317.

Models for large areas (global models) are often biased in smaller sub-areas, even when the model is unbiased for the whole area. Localization of the global model removes the local bias, but the problem is to find homogenous sub-areas in which to localize the function. In this study, we used the eCognition Professional 4.0 (later versions called Definiens Pro) segmentation process to segment the study area into homogeneous sub-areas with respect to residuals of the global model of the form height and/or local Getis statistics calculated for the residuals, i.e.,  $G_i^*$ -indices. The segmentation resulted in four different rasters: 1) residuals of the global model, 2) the local  $G_i^*$ -index, and 3) residuals and the local  $G_i^*$ -index weighted by the inverse of the variance, and 4) without weighting. The global model was then localized (re-fitted) for these sub-areas. The number of resulting sub-areas varied from 4 to 366. On average, the root mean squared errors (RMSEs) were 3.6% lower after localization than the global model RMSEs in sub-areas before localization. However, the localization actually increased the RMSE in some sub-areas, indicating the sub-area were not appropriate for local fitting. For 56% of the sub-areas, coordinates and distance from coastline were not statistically significant variables, in other words these areas were spatially homogenous. To compare the segmentations, we calculated an aggregate standard error of the RMSEs of the single sub-areas in the segmentation. The segmentations in which the local index was present had slightly lower standard errors than segmentations based on residuals.

**Keywords** eCognition, form height, Getis statistics, image segmentation, local indicators of spatial association

**Addresses** University of Helsinki, Department of Forest Sciences, P.O. Box 27 (Latokartanonkaari 7), FI-00014 University of Helsinki, Finland **E-mail** minna.s.ratty@helsinki.fi

**Received** 17 August 2009 **Revised** 29 March 2010 **Accepted** 31 March 2010

**Available at** <http://www.metla.fi/silvafennica/full/sf44/sf442303.pdf>

---

## 1 Introduction

Growing stock properties such as height, volume, and form height are costly, time-consuming, and sometimes impossible to measure in field surveying. The height of only a fraction of all the trees is normally measured, and the volume and the form height are estimated using models. Such a model can be either parametric or non-parametric, and, in addition, can be either a local or a global one. "Global" refers to the whole of the area under consideration as a single unit, no matter how large that unit may be. A model is called global if the same model is fitted over the entire area. A local model, however, has at least one parameter which is estimated only for one particular sub-area.

The estimates of the global model may be unbiased over the whole area, but for one or more specific sub-areas they could be badly biased (e.g. Korhonen 1993). For example, there may be unexplained spatial variation in the dependent variable due to one or more variables that for some reason or other cannot be included in the model. The question is how to figure out sub-areas in which this unexplained variation is homogeneous, i.e. the most appropriate size and location of sub-areas for localizing the global model. There are several alternatives for how the localization could be carried out. In this study, we first segment the entire study area and then re-fit the original regression model individually to each sub-area. We do not add new variables to the model; only remove statistically insignificant variables from the original form height model.

Delineation of proper localization sub-areas can be problematic if localization is carried out by local fitting of the models only at certain points or small areas. There are an infinite number of different solutions to this problem. Selecting the optimal ones is, in principle, possible, but not feasible in practice. However, the sizes of the sub-areas should be large enough, to allow enough observations for local fitting of the model, and the sub-areas should not be too large, as larger sub-areas become increasingly heterogeneous. If the sub-areas are fully homogeneous, it is enough to adjust the level of predictions in each sub-area with a simple correction coefficient. The alternative is to use methods which do not need segmentation into reasonably homogene-

ous regions. These can be variance-based, e.g. kriging (Cressie 1991, Köhl and Gertner 1997), mixed models including some local-level random effect/s (Kangas and Korhonen 1995), or non-parametric such as knn imputation methods (Sironen et al. 2008) with respect to the form height in the study area. Compared to these methods, localization of a parametric model by local fitting or a local correction coefficient is simple to implement, once the segmentation of the global area to homogeneous sub-areas is carried out.

The local indicators of spatial association (LISA) measure the degree of clustering of similar values in the neighbourhood of the calculation points, called pivots. The first LISAs were introduced in the 1990s by Anselin (1995) (local Moran's  $I_i$  and local Geary's  $c_i$ ) and Getis and Ord (1992, Ord and Getis 1995) ( $G_i$  and  $G_i^*$ ). Both Geary's  $c_i$  and Moran's  $I_i$  detect departures from spatial randomness but are calculated differently and so yield different numerical results.  $G_i$  and  $G_i^*$  are commonly called "hot spot analysis" and indicate where there are spatial clusters of variable values with either high or low z-scores within the whole dataset. These indices have recently been used in many scientific disciplines, for example epidemiology and regional and urban sciences. In forestry, indices have been used to characterize within-stand structure (Wells and Getis 1999), competition between single trees (Shi and Zhang 2003, Wulder et al. 2007), for forest damage detection (Weishampel et al. 2007) and model analysis (Zhang et al. 2005, Wulder et al. 2007).

In an earlier study (Räty and Kangas 2007), we tested the suitability of LISAs such as Moran's  $I_i$ , Geary's  $c_i$ , and  $G_i$  and  $G_i^*$  for delineation of homogenous sub-areas from residuals of the global model. The Moran's  $I_i$  and Geary's  $c_i$  could show only a couple of statistically significant clusters which were small in size whereas both Getis statistics,  $G_i$  and  $G_i^*$ , did indicate several clusters dividing the study area into both positive and negative sub-areas (see Räty and Kangas 2007, Fig. 4). In the second study (Räty and Kangas 2008), we divided the whole study area using a recursive partitioning algorithm (following classification and regression trees (CART) introduced by Breiman et al. 1984) using the  $G_i^*$ -index and/or residuals and spatial coordinates

as dividers. The idea in CART is to find the most profitable divisions among the given variables (spatial coordinates), which minimizes the deviance in the dependent variable (the  $G_i^*$ -index and/or residuals). The result of the dividing process can be presented as a tree, but since the dividers were coordinates the division could also be drawn over a map. Although CART was successful, the results from this method did not differ from division of the study area into the eleven forestry administrative centres (see Räty and Kangas 2008, Fig. 2a), which were used as a baseline for the assessment of the method. The sub-areas that CART delineated were rectangular in shape, where they should have been more flexible and variable in both size and geometry.

In this study, we segment the study area using eCognition 4.0 Professional® (also known as Definiens Pro), which is designed to cluster and segment satellite and aerial images but accept all kinds of raster data layers as well (eCognition... 2009). The software uses general object-oriented image analysis, in which the initial segmentation is based on primary features, such as digital value (Benz et al. 2004). An object is created by progressively merging pixels using a bottom up region-merging technique which is controlled by a heterogeneity variable. Heterogeneity is a combination of colour and shape variables and is limited by a scale parameter, a threshold value for the optimization process in region-merging, acting for instance as a stop criterion. The initial segmentation routine of the software, which we use, has been employed in segmenting the data in various papers (e.g., Guindon et al. 2004, Bock et al. 2005, Yu et al. 2006, Johansen et al. 2007, Van Coillie et al. 2007, Straatsma and Baptist 2008, Triepke et al. 2008, Mustonen et al. 2008). The decision variables in the present study are the same as in the CART-study, the residuals and the  $G_i^*$ -index with combination to the spatial position.

The aim of this study is to evaluate 1) the localization sub-areas formed by the multiresolution segmentation, and 2) to compare these segmentations with segmentations formed by CART. The latter segmentations have been presented in detail in our previous study and here we introduce the results briefly (Räty and Kangas 2008, Table 8). Having the segmentations done we also wanted

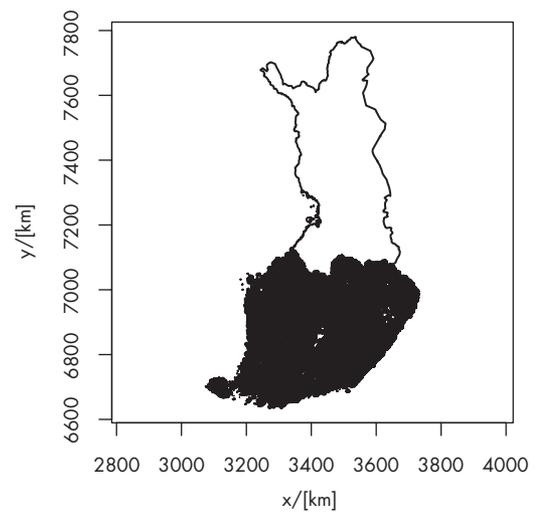
3) to test whether the properties of the local data, such as the range and the variation of the variables in each sub-area, contributed to the success of localization. We compared the values for model outputs and the residuals when they are localized, with the same values when they are not localized.

## 2 Materials

Our data is sample tree (*Pinus sylvestris*) data from the ninth National Forest Inventory of Finland (NFI9) (Fig. 1, Table 1a). The trees were selected from a plot by angle gauge (BAF=2). We modelled a measure proportional to the form height ( $f$ ), hereafter simply called form height (Eq. 1).

$$f = \frac{v}{d^2} = \beta_0 + \beta_1 d + \beta_2 d^2 + \beta_3 \ln(\text{BA}) + \alpha_1 XC + \alpha_2 XC^2 + \alpha_3 XC \cdot YC + \alpha_4 YC + \alpha_5 YC^2 + \alpha_6 \text{RDIST} + \varepsilon \quad (1)$$

where diameter at breast height ( $d$ ), and volume ( $v$ ) are from the tree data and the total basal area (BA) from the plot data. In Eq. 1 are three geo-



**Fig. 1.** The study area in Southern Finland.

**Table 1a.** Basic properties of the Scots pine (*Pinus sylvestris*) sample trees: diameter at breast height ( $d$ ), diameter at 6 metres height ( $d_6$ ), height ( $h$ ), volume ( $v$ ), basal area (BA), measure proportional to the form height ( $f$ ), and actual form height (fh).

	Min	1stQ	Med	Mean	3rdQ	Max
$d$ [cm]	5	14	19	20.55	27	67
$d_6$ [cm]	*	11	16	16.4	22	51
$h$ [dm]	20	105	146	150	193	352
$v$ [dm <sup>3</sup> ]	3.7	80	204	331	479	2882
BA [m <sup>2</sup> /ha]	1	16	21	21.15	27	60
$f$ [dm <sup>3</sup> /cm <sup>2</sup> ]	0.148	0.413	0.543	0.553	0.677	1.305
fh [m]	1.9	5.3	6.9	7	8.6	16.6

\* 2561 of the sample trees were less than six metres high.

**Table 1b.** The frequency distributions of diameter and stand basal area for the Scots pine (*Pinus sylvestris*) sample trees data.

	5–10	10–15	15–20	20–25	25–30	30–35	35–40	40–45	45–50	50–55	55–60	60–65	65–70
$d$ [cm]	1952	3575	4145	3419	2681	1924	987	369	90	24	7	1	1
BA [m <sup>2</sup> /ha]	950	2184	5116	4750	2601	1990	494	289	61	9	11	-	-

graphical measures: XC- and YC-coordinates (Eq. 2), and the distance from the coastline, RDIST (Eq. 3) calculated as:

$$\begin{cases} YC = \frac{(Y - 6620)}{1000} \\ XC = \frac{(X - 60)}{1000} \end{cases} \quad (2)$$

$$\begin{cases} RDIST = \frac{1}{(DIST + 0.05)}, & DIST \leq 20 \\ RDIST = 0, & DIST > 20 \end{cases} \quad (3)$$

where  $X$  and  $Y$  are the geographic coordinates in kilometres, and  $DIST$  is the straight-line Euclidian distance from the coastline in kilometres, which are all in Kartastokoordinaattijärjestelmä (KKJ), a national Gauss-Krüger map projection system used in Finland.

The main interest is in the residuals of this global model and later in the residuals of the

fitted local models. A residual is the difference between the true value ( $f$ ) and the fitted value ( $\hat{f}$ ) (Eq. 4):

$$\epsilon_i = f_i - \hat{f}_i \quad (4)$$

In earlier studies, we have used LISAs in the localization of regression models (Räty and Kangas 2007, 2008). Here we used the same local  $G_i^*$ -index (Eq. 5) (Getis and Ord 1992, Getis and Ord 1996),

$$G_i^* = \frac{\sum_j w_{ij} \epsilon_j}{\sum_j \epsilon_j} \quad (5)$$

where  $\epsilon_j$  is the residual of the global model (Eq. 4),  $w_{ij}$  are binary weights, and the neighbourhood is a 20-km circular (see Räty and Kangas 2007 for the neighbourhood selection), i.e., observation  $j$  has unit weight if it is less than 20 km distant from pivot  $i$ ; otherwise it is zero (see Anselin 1995). In the  $G_i^*$ -index, pivot  $i$  also belongs to neighbour-

hood  $j$ . We used the Z-standardized version of the index,  $ZG_i^*$ , hereafter  $G_i^*$ -index or simply  $G^*$  in tables (Eq. 6),

$$ZG_i^* = \frac{G_i^* - E[G_i^*]}{SD[G_i^*]} \tag{6}$$

where  $E[G_i^*]$  is the expected value, which is the mean of the  $G_i^*$ -indices, and  $SD[G_i^*]$  is the standard deviation.

### 3 Methods

In preliminary preparations, we interpolated the point data (single trees) of residuals and  $G_i^*$ -indices into grids which hereafter we call (data) “layers” with ArcMap® inverse distance weighting (IDW) where the missing pixel values were calculated as a weighted sum of the neighbours and the weights were the inverses of the distances to the power of two between five nearest observations and the pixel (Longley et al. 2005). We then passed these layers to eCognition Professional 4.0® (Definiens Imaging) for multiresolution segmentations (Benz et al. 2004, eCognition... 2009). In addition, we formed two additional layers as combinations of the former two, namely residuals and the local  $G_i^*$ -index with and without weighting by the inverse of the variance.

We produced segmentations for each of the four layers (the abbreviation is in parentheses) by changing the number of sub-areas indirectly via the scale parameter, whose values are shown under the name of the each individual segmentation after the abbreviation in Table 2. This procedure led into several segmentations. The number of segmentations of the layer is mentioned after the content of the segmentation in the following list:

- 1) Residuals of the global model (RES), six segmentations
- 2)  $G_i^*$ -index ( $G^*$ ), 11 segmentations
- 3) Residuals and  $G_i^*$ -index in layers (Res; $G^*$ ), seven segmentations
- 4) Weighted residuals and  $G_i^*$ -index: here the layers were weighted by the inverse of the layer variances (Weight), six segmentations

**Table 2.** For all sub-areas made: the number of sub-areas (DIV) in them, the name of the segmentation (SEG), and descriptive statistics of the number of trees in sub-areas (Observations).

Div	Seg	Observations <sub>i</sub>			
		Min	Med	Mean	Max
366	G*10	10	43	52	219
265	G*13	10	62	72	360
177	G*17	14	89	108	567
120	G*22	14	137	159	567
84	G*29	17	191	228	990
62	G*38	17	263	309	1156
35	G*50	17	404	548	2462
25	G*65	51	670	767	2462
17	G*87	189	817	1128	2886
6	G*103	299	1124	3196	12380
4	G*130	299	2357	4794	14160
150	Weight10	11	99	127	990
89	Weight14	11	191	215	990
54	Weight20	12	308	354	1626
24	Weight27	24	510	798	3797
9	Weight36	431	2336	2129	6008
5	Weight56	431	2453	3833	10900
140	Res10	11	110	136	552
73	Res14	20	191	262	2129
39	Res20	80	401	491	2218
22	Res27	80	535	870	3105
15	Res35	152	1393	1277	3105
10	Res47	719	1842	1915	3529
291	Res;G*10	10	52	65	253
192	Res;G*13	11	82	99	643
120	Res;G*18	11	135	159	643
78	Res;G*26	11	173	245	812
43	Res;G*35	11	319	445	1774
23	Res;G*48	72	673	833	2389
16	Res;G*60	162	662	1198	3328

Increasing the scale parameter decreased the number of sub-areas. The parameter values were not pre-determined since the effect of the scale parameter was not consistent. We used the scale parameter analysis in eCognition to determine the values individually for each of the segmentations. All other decision variables were at their defaults: the shape factor had the value of 0.1 and smoothness/compactness 0.5. The first refers to the weight given to the shape of the sub-area delineated versus that given to the digital value in the layer. Value 0.1 here means that the delinea-

tion depended mainly on the digital value, which was appropriate for our purposes, since our homogeneity information was in the layer and we had no interest in limiting the shapes of the sub-areas. The second parameter was for the smoothness of the borderline versus compactness of the objects delineated. Its value of 0.5 was a compromise; we did not favour either of these two parameters.

We allowed the program to do its initial segmentation of the study area without any limitation, and then localized the model (Eq. 1) into the resulting sub-areas. We rejected sub-areas with less than ten observations from the localization since the regression function used had nine variables and a constant, so that with ten observations the original function had no degree of freedom at all. This elimination removed altogether 128 sub-areas from a total of 2682 sub-areas from 14 segmentations for a total of 30 segmentations. However, this limit could have been higher since 30 or more observations are usually recommended for fitting regression. However setting the limit up to 30 would have raised the number of rejected sub-areas up to 472, or 18% of all sub-areas. This is why we kept the limit to ten observations.

The re-fitting of the global form height model (Eq. 1) was carried out with the stepping function with the penalty criteria (stepAIC), which can step both backward and forward to the localized model in the process of the variable selection. This function is available in the statistical software R (Venables and Ripley 2002). Each variable in a localized regression model had to reach the  $p=0.05$  level (statistically significant). The success of the localization was measured by the root mean squared error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n \varepsilon_i^2}{df}} = \sqrt{\text{bias}^2 + \sigma^2} \quad (7)$$

where  $\varepsilon_i$  is the residual of the global model (Eq. 4), and  $df$  is the degrees of freedom. The RMSE can be expressed also as a square root of the sum of the squared bias and variance,  $\sigma^2$  (Eq. 7). This RMSE was calculated for both the situation before localization and after it in the sub-areas. The bias was zero for the localized functions since the model was fitted to the local data available. The bias is the mean (Eq. 8) of residuals (Eq. 4)

$$\text{bias} = \frac{\sum_{i=1}^n \varepsilon_i}{n} \quad (8)$$

We made two comparisons with the localized model  $RMSE_{i,loc}$ :

- 1) comparison with the global model  $RMSE$  ( $RMSE_{glo}=0.1027 \text{ dm}^3/\text{cm}^2$ ) (Eq. 9), and

$$\begin{aligned} \Delta RMSE &= \frac{RMSE_{i,loc} - RMSE_{glo}}{RMSE_{glo}} \cdot 100\% \\ &= \frac{RMSE_{i,loc} - 0.1027}{0.1027} \cdot 100\% \end{aligned} \quad (9)$$

- 2) comparison with the global model  $RMSE_{i,glo}$  in sub-area  $i$  (Eq. 10).

$$\Delta RMSE_i = \frac{RMSE_{i,loc} - RMSE_{i,glo}}{RMSE_{i,glo}} \cdot 100\% \quad (10)$$

where  $loc$  is the localized function and  $glo$  is the global function for a sub-area  $i$ . The first comparison above shows the changes in respect to the global level and the latter the actual change in that sub-area under localization in percentages.

One way to assess the success of localization is to look at which variables were included in the localized function ( $X$ -, and  $Y$ -coordinates, and distance from coastline,  $RDIST$ ). If spatial variables were included in the localized function, the segmentation failed on the criterion of spatial homogeneity, i.e., the sub-area had spatial variation left.

Besides the segmentation algorithm, the success of localization may also depend on the properties of the local data used for local fitting. The accuracy of the regression model depends in many ways on the design matrix  $X$  (see Draper and Smith 1981). Here we assume that the goodness of the data for model fitting into a sub-area can be measured by the balance of the data, the range, and variance of the variables in the sub-areas. By balance, we mean that the distribution of independent variables ( $d$ ,  $BA$ ) is uniform (Table 1a). To test balance, we calculate Shannon's entropy index:

$$H = -\sum_i p_i \ln p_i \quad (11)$$

**Table 3.** The classification of the sub-areas into three classes: negative, neutral, and positive, according to the signs of the mean of  $G_i^*$ -index (Mean) and its standard deviation (SD) in the sub-area.

Class	Mean	Mean-SD	Mean+SD
Negative	–	–	–
Positive	+	+	+
Neutral	–	–	+
Neutral	+	–	+

where  $p$  is the proportion of the number of trees in a particular category, e.g. diameter class or basal area class, to the total number of the trees in a sub-area. For the whole dataset Shannon’s entropy index is 1.96 for the diameter and 1.91 for the basal area (Table 1b). The entropy for uniformly distributed data, which has exactly the same number of trees and is divided into as many classes as the present data, would be 2.56 and 2.69 respectively. We compare entropy (Eq. 11) with the local RMSEs by calculating correlations between them.

To study the effect of balance further, we divided the sub-areas into classes according to the  $G_i^*$ -index values (Table 3). If  $\bar{G}_i^* + \sigma < 0$ , the area was classified as negative. If  $\bar{G}_i^* - \sigma > 0$ , the sub-area was classified as positive. The sub-areas fitting neither criterion were classified as neutral. We calculated correlations between the localized RMSEs of the class and the Shannon’s entropy index for the diameter and properties of the variables in the sub-area.

To compare different segmentations across the whole area, we calculated an aggregate estimate of the standard error for the area as a weighted sum of squared RMSEs for all points within the area, i.e., mean squared errors (MSEs),  $se_{aggr}$ :

$$se_{aggr} = \sqrt{\frac{1}{\left(\sum_{i=1}^m n_i^2\right)} \cdot \sum_{i=1}^m n_i^2 MSE_i} \tag{12}$$

where  $m$  is the number of sub-areas in a segmentation and  $n_i$  is the number of observations in sub-area  $i$ .

## 4 Results

For the localized models the RMSEs had a median and mean 3–4% lower than the global RMSE of 0.1027 dm<sup>3</sup>/cm<sup>2</sup> (Eq. 1) (Table 4). The minimum RMSEs were 39–42% of global model values whereas the maximum RMSE was one and a half times the global model value ( $\Delta RMSE$ ). The minimum, mean and median RMSEs all increased when the number of sub-areas decreased. For the maximum RMSEs, the effect was the reverse, the RMSE decreasing as the number of sub-areas decreased. The localized RMSEs ( $\Delta RMSE_i$ ) were also 3–4% lower (mean and median) than the RMSE before localization in the sub-areas, but the variation in these values was 9.4% (–8.5 to 0.9%) when the range for the  $\Delta RMSE$  was 14.7% (–10.9 to 3.8%). In other words, in small sub-areas both the greatest benefits and the drawbacks of localization are the most probable.

Attaining spatial homogeneity in sub-areas was one target. The percentage of localized functions without spatial variables in the sub-areas varied from 0 to 75. At first, the percentage increased as the number of sub-areas increased, rising above 50% after 40 sub-areas and reaching its maximum of 60%, between 100 and 200 sub-areas. Finally, the percentage lacking spatial variables saturated at the level of 55% (Fig. 2a). The number of variables in a localized function decreased as the number of sub-areas increased (Fig. 2b), but the change was slight. The localized function had three variables after 50 sub-areas and two variables after 90 sub-areas. The number of spatial variables in all localized functions dropped from a maximum of three to zero after 50 sub-areas (Fig. 2c). We have studied here the median values for the variables and thus there may be some localized functions at any number of sub-areas which have some spatial variables left (Fig. 2d). Therefore it is not possible to create fully homogenous sub-areas with the data layers and segmentation criteria used.

Our hypothesis was that the uniformity of the distribution of variables (balance) has an impact on the localization and, as a measure of this balance, we calculated Shannon’s entropy index for diameter and basal area distributions in the sub-areas. For these two indices and for the ranges and variances of the diameter and basal area, we calculated the correlations to the localized

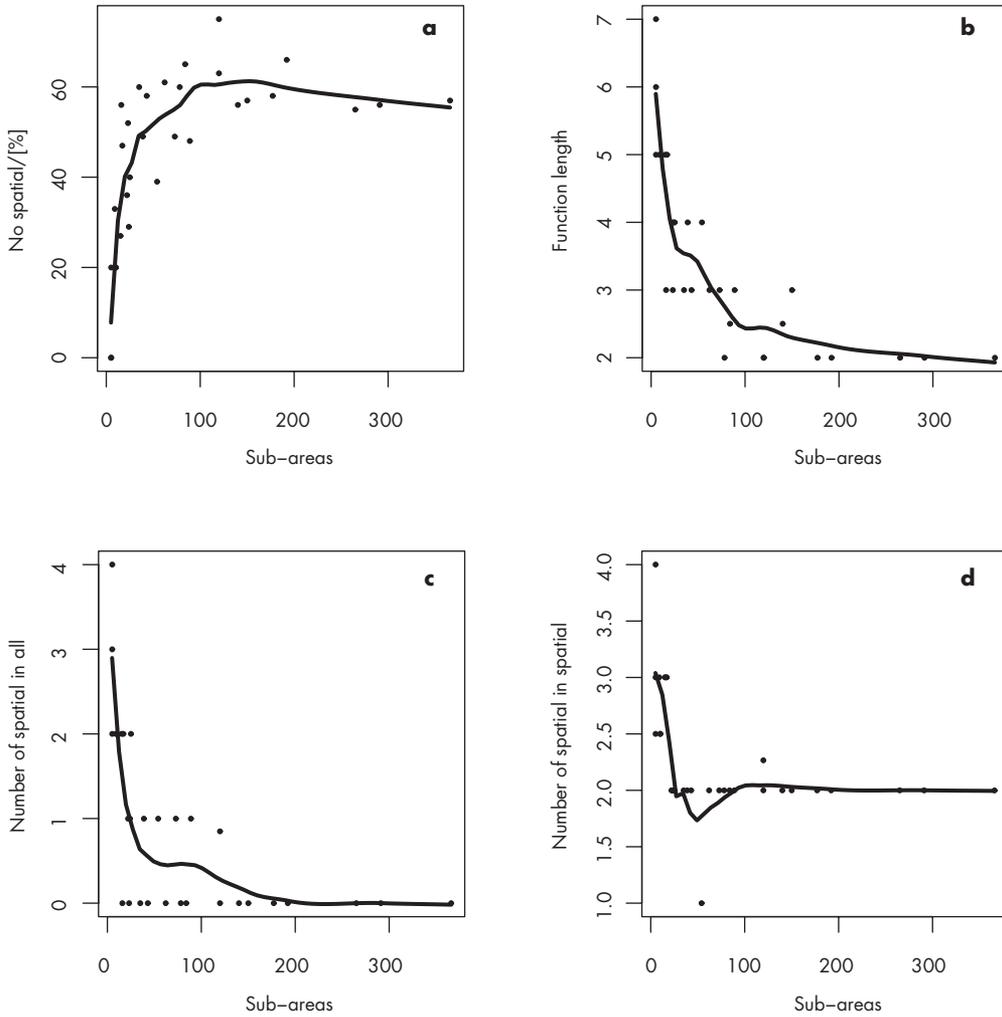
**Table 4.** The ratio (in percent) of localization RMSE to the global model RMSE and to the global model RMSE in the same sub-area.

Seg	RMSE				$\Delta$ RMSE/[%]				$\Delta$ RMSE <sub>i</sub> /[%]			
	Min	Med	Mean	Max	Min	Med	Mean	Max	Min	Med	Mean	Max
Res10	0.0308	0.0996	0.0997	0.1712	-70	-3.0	-3.0	67	-74	-3.6	-3.3	97
Res14	0.0569	0.1008	0.1008	0.1366	-45	-1.8	-1.8	33	-34	-3.3	-4.4	13
Res20	0.0730	0.0992	0.0994	0.1260	-29	-3.4	-3.3	23	-17	-2.9	-3.4	10
Res27	0.0789	0.0954	0.0968	0.1139	-23	-7.1	-5.8	11	-17	-2.7	-4.4	10
Res35	0.0822	0.0952	0.0979	0.1139	-20	-7.3	-4.7	11	-17	-1.7	-4.2	0
Res47	0.0837	0.0988	0.0990	0.1139	-19	-3.8	-3.6	11	-5	-1.6	-1.8	0
G*10	0.0141	0.0915	0.0942	0.2563	-86	-10.9	-8.3	150	-89	-7.3	-7.7	201
G*13	0.0176	0.0953	0.0964	0.2504	-83	-7.2	-6.1	144	-77	-5.6	-5.4	201
G*17	0.0598	0.0975	0.0998	0.2504	-42	-5.1	-2.8	144	-53	-4.2	-3.6	96
G*22	0.0617	0.0990	0.1011	0.2504	-40	-3.6	-1.5	144	-53	-3.0	-2.1	96
G*29	0.0626	0.0987	0.0999	0.1533	-39	-3.9	-2.7	49	-53	-3.0	-2.2	96
G*38	0.0626	0.0982	0.1003	0.1533	-39	-4.4	-2.3	49	-53	-3.2	-1.8	96
G*50	0.0790	0.0992	0.1013	0.1533	-23	-3.4	-1.4	49	-30	-2.0	0.9	96
G*65	0.0790	0.0986	0.0988	0.1237	-23	-4.0	-3.8	20	-14	-2.6	-1.8	12
G*87	0.0794	0.1013	0.1004	0.1237	-23	-1.4	-2.2	20	-14	-1.8	-1.7	12
G*103	0.0794	0.1002	0.0942	0.1032	-23	-2.5	-8.3	0	-14	-3.3	-5.6	0
G*130	0.0794	0.1002	0.0954	0.1020	-23	-2.5	-7.1	-1	-14	-6.0	-6.7	0
Res;G*10	0.0057	0.0937	0.0937	0.2359	-94	-8.8	-8.8	130	-95	-7.9	-8.5	64
Res;G*13	0.0463	0.0980	0.0994	0.2359	-55	-4.6	-3.2	130	-58	-4.1	-4.2	55
Res;G*18	0.0523	0.1020	0.1006	0.1483	-49	-0.7	-2.0	44	-48	-1.9	-2.6	54
Res;G*26	0.0523	0.0986	0.0981	0.1435	-49	-3.9	-4.4	40	-48	-3.0	-4.4	18
Res;G*35	0.0523	0.0983	0.1000	0.1435	-49	-4.3	-2.7	40	-28	-2.1	-2.3	12
Res;G*48	0.0797	0.1010	0.1030	0.1435	-22	-1.6	0.3	40	-14	-1.2	-0.3	12
Res;G*60	0.0797	0.0990	0.1007	0.1236	-22	-3.6	-2.0	20	-14	-1.7	-0.1	12
Weight10	0.0232	0.0966	0.0963	0.1516	-77	-6.0	-6.3	48	-76	-4.8	-7.3	22
Weight14	0.0232	0.0969	0.0968	0.1357	-77	-5.7	-5.7	32	-76	-3.7	-6.6	22
Weight20	0.0232	0.0990	0.0982	0.1327	-77	-3.6	-4.4	29	-76	-3.2	-6.1	11
Weight27	0.0595	0.0981	0.0977	0.1310	-42	-4.4	-4.9	28	-56	-2.5	-5.1	4
Weight36	0.0905	0.1031	0.1034	0.1163	-12	0.4	0.7	13	-6	-1.9	-1.8	2
Weight56	0.0952	0.1066	0.1062	0.1163	-7	3.8	3.4	13	-2	-1.1	-1.3	-1

model  $RMSE_{i,loc}$  in sub-areas (Table 5a). For the diameter, both Shannon’s entropy index and the variance had from medium to large positive correlation with  $RMSE_{i,loc}$ , i.e., the variance and the entropy index were large for large RMSEs. One would expect that both correlations would be negative.

In the all three classes of sub-area classification (Table 3), the entropy index showed medium strength positive correlation with localized  $RMSE_{i,loc}$  (Table 5b). The evenness in distribution of the diameters has therefore no positive impact on the localization result in the studied case. Two

correlations were large and positive: that of the localized  $RMSE_{i,loc}$  with global  $RMSE_{i,glo}$  and that of the change in localization  $\Delta RMSE_i$  (Eq. 9). If we look at the localized models and their variables with this same classification into positive, negative or neutral, we can see differences between the models (Table 6). In neutral sub-areas, where the mean  $G_i^*$ -index did not differ significantly from zero, the localized models had more variables in general and the occurrence percentages for all variables in the models were higher (except for  $YC^2$ ) than for the models in the two other classes. The largest differences



**Fig. 2.** The relationship between a) percentages of localized functions which have no spatial variables, b) number of variables in localized function, c) number of spatial variables in all localized functions, and d) number of spatial variables in those localized functions versus number of sub-areas in segmentations. The line in both figures is the locally fitted trend line.

were in the occurrences of the basal area (BA) and diameter ( $d$  and  $d^2$ ) variables, which were considerably higher for neutral sub-areas than in the two other classes, i.e., in neutral sub-areas the differences in form height were explained by tree and stand variables more often than in positive or negative sub-areas. In other words, there was more heterogeneity in form height in neutral classes than in the other two classes. This is a logical result since a sub-area belonging to the

neutral class can be by definition either a cluster of near zero values or mixture of miscellaneous  $G_i^*$ -indices having their mean near zero.

Evaluating the segmentations could be done via aggregate standard errors. The minimum was  $se_{aggr}=0.0967 \text{ dm}^3/\text{cm}^2$  when the residual layer was segmented using the scale parameter value 14 (Res14) (Table 7); i.e., the residual layer was split into 73 sub-areas. There were altogether 14 segmentations that had a  $se_{aggr}$  below the 0.1

**Table 5a.** Correlations between RMSE of localized function and following six measures: Shannon’s entropy index (H) for diameter ( $d$ ) and basal area (BA), ranges (Range) and variances (Var) of diameter and basal area in the sub-areas.

Seg	H		Range		Var	
	$d$	BA	$d$	BA	$d$	BA
Res10	0.38	0.38	0.31	0.19	0.26	0.03
Res14	0.43	0.19	0.18	0.22	0.40	0.12
Res20	0.65	0.32	0.33	0.22	0.59	0.05
Res27	0.64	0.51	0.45	0.36	0.58	-0.04
Res35	0.73	0.44	0.42	0.46	0.69	0.00
Res47	0.90	0.67	0.44	0.62	0.90	-0.09
G*10	0.31	0.19	0.15	0.14	0.17	0.12
G*13	0.40	0.11	0.18	0.17	0.25	0.14
G*17	0.40	0.05	0.15	0.06	0.33	0.06
G*22	0.35	-0.07	0.05	0.02	0.28	0.03
G*29	0.39	0.07	0.16	0.18	0.31	-0.10
G*38	0.31	0.10	0.12	0.19	0.16	-0.12
G*50	0.29	-0.23	-0.11	-0.04	0.30	-0.18
G*65	0.76	0.36	0.22	0.35	0.67	0.02
G*87	0.70	0.52	0.14	0.37	0.59	0.14
G*103	0.78	0.62	0.46	0.74	0.64	0.06
G*130	0.73	0.53	0.70	0.97	0.66	-0.12
Res;G*10	0.42	0.19	0.20	0.18	0.14	0.04
Res;G*13	0.35	0.16	0.15	0.12	0.19	0.09
Res;G*18	0.55	0.30	0.33	0.16	0.37	0.00
Res;G*26	0.54	0.38	0.35	0.39	0.33	0.04
Res;G*35	0.66	0.33	0.30	0.22	0.62	-0.09
Res;G*48	0.72	0.44	0.10	-0.10	0.70	-0.34
Res;G*60	0.78	0.47	0.18	0.17	0.73	-0.26
Weight10	0.59	0.34	0.47	0.31	0.36	0.06
Weight14	0.71	0.46	0.62	0.40	0.46	0.02
Weight20	0.68	0.36	0.51	0.48	0.44	-0.21
Weight27	0.50	0.41	0.25	0.72	0.13	0.12
Weight36	0.85	0.31	0.28	-0.14	0.87	-0.50
Weight56	0.87	-0.32	-0.27	-0.48	0.87	-0.05

**Table 5b.** Sub-areas were classified into three classes: negative, neutral, and positive according to the  $G_i^*$ . Here is the number and total area they represent reported. In the second half of the table are correlations (Cor) between RMSE of localized function ( $RMSE_{i,loc}$ ), and Shannon’s index for diameter ( $H(d)$ ), situation before localization ( $RMSE_{i,glo}$ ), and change in RMSE in the localization ( $\Delta RMSE_i$ ) for the classes.

Area class	Number	Area [km <sup>2</sup> ]	Cor	H( $d$ )	$RMSE_{i,glo}$	$\Delta RMSE_i$
Negative	997	1662	$RMSE_{i,loc}$	0.28	0.42	0.71
Neutral	574	4752		0.59	0.70	0.58
Positive	983	1538		0.48	0.58	0.73

dm<sup>3</sup>/cm<sup>2</sup> level. All four layers based on residual and/or  $G_i^*$ -index alone or together were present in that group and segmentations had from nine to 265 sub-areas. Four residual layer segmentations are among five with the highest aggregated errors, however. The maximum  $se_{agg}$ , 0.1031 dm<sup>3</sup>/cm<sup>2</sup>, was reached with a residual layer split into 22 sub-areas by scale parameter value 27 (RES27). Among the ten highest aggregate errors for segmentations, there were segmentations from all four layers having from six to 62 sub-areas.

### 5 Discussion

We have segmented the study area for smaller localization sub-areas by different methods. By

**Table 6.** Variables in the regression models after localization when the data is divided into three classes: negative, neutral and positive, according to the  $G_i^*$ -index values and deviation in the sub-areas. The number in the Table is the percentage of the localized models where the variable occurs.

$G_i^*$	$d$	$d^2$	BA	XC	YC	XC <sup>2</sup>	YC <sup>2</sup>	XCYC	RDIST
Negative	90.9	34.0	40.0	28.4	29.7	13.9	14.6	11.2	4.9
Neutral	96.7	58.4	59.8	32.6	32.8	16.9	12.4	16.9	10.8
Positive	91.9	40.9	47.5	24.7	26.6	12.5	13.7	9.6	3.5
All	92.6	42.6	47.3	27.9	29.2	14.1	13.8	11.5	5.7

$d$  = Diameter at breast height  
 BA = Total basal area  
 XC = X-coordinate  
 YC = Y-coordinate  
 RDIST = Euclidian distance from coastline

**Table 7.** The aggregate standard errors for the segmentations.

Seg	Div	se <sub>aggr</sub>
Res10	140	0.1002
Res14	73	0.0967
Res20	39	0.0988
Res27	22	0.1029
Res35	15	0.1023
Res47	10	0.1025
G*10	366	0.1009
G*13	265	0.0998
G*17	177	0.0983
G*22	120	0.1016
G*29	84	0.0986
G*38	62	0.0991
G*50	35	0.0986
G*65	25	0.0994
G*87	17	0.0983
G*103	6	0.1006
G*130	4	0.1010
Res;G*10	291	0.1014
Res;G*13	192	0.1025
Res;G*18	120	0.1016
Res;G*26	78	0.0997
Res;G*35	43	0.1018
Res;G*48	23	0.1019
Res;G*60	16	0.0981
Weight10	150	0.1004
Weight14	89	0.0988
Weight20	54	0.1004
Weight27	24	0.0997
Weight36	9	0.0998
Weight56	5	0.1018

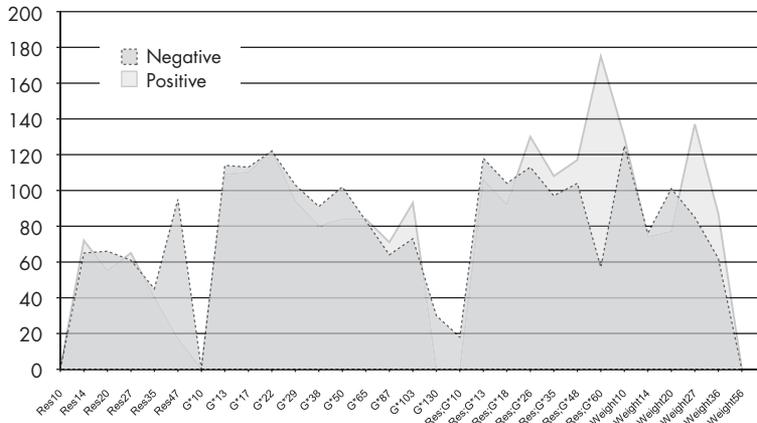
changing the variable layers on which the segmentation was based and (by changing region merging criteria) the number of the sub-areas, we ended up having 30 different segmentations for the study area with 4 to 366 sub-areas (Table 2). As in earlier research (Räty and Kangas 2008), the  $G_i^*$ -index turned out to be more efficient as the area divider than the residuals of the global model. The changes in RMSEs for the  $G_i^*$ -index segmentations were larger, i.e., the localized RMSEs were smaller (Table 4), as were the aggregate standard errors in general (Table 7). This implies that a variable, in this case the local  $G_i^*$ -index, which averages the residuals over a larger neighbourhood, reveals the trends hidden in the residuals.

**Table 8.** The RMSEs for segmentations with recursive partitioning, CART (Räty and Kangas 2008).

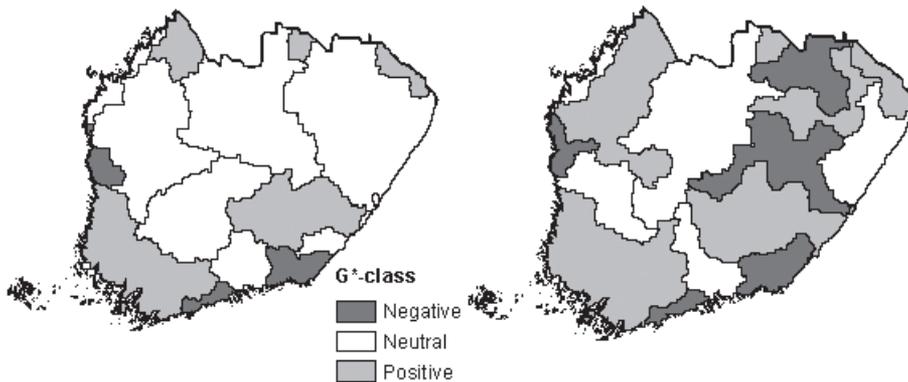
CART	Div	RMSE <sub>i</sub>				se <sub>aggr</sub>
		Min	Med	Mean	Max	
4	127	0.0652	0.1014	0.0995	0.1414	0.0992
5	50	0.0775	0.102	0.1006	0.1292	0.0980
4P1	38	0.0775	0.1001	0.0994	0.1227	0.0997
5P1	24	0.0775	0.0987	0.0977	0.1227	0.1024
3	23	0.0541	0.0997	0.0983	0.1503	0.0980
5P2	20	0.0778	0.1001	0.0995	0.1227	0.1012
2	16	0.0562	0.0988	0.1017	0.1275	0.0974
4P2	15	0.0786	0.0985	0.0992	0.1145	0.1019
5P3	12	0.0786	0.0995	0.1001	0.1128	0.1017
1	11	0.0936	0.1074	0.1049	0.1270	0.0989

A known deficiency of LISAs is that they cannot identify clusters of variable values near zero or mean value of the variables over the entire layer (Tiefelsdorf and Boots 1997). In this study the mean is zero for residuals since the fitted regression model has no bias. For localization area delineation, this is not a good feature because an index value near zero can mean either a homogenous or heterogeneous surrounding. The classification of sub-areas according to  $G_i^*$ -index into three classes, negative, neutral, and positive (Table 3), showed differences in the RMSEs and the changes in RMSEs in the localization (Table 5b). The “neutral” indices are those with values around 0, which might be expected to illustrate the difficulties of LISAs. Firstly, there were about twice as many both negative and positive sub-areas as neutral ones, and the neutral sub-areas were three times larger than the others in all segmentations (Fig. 3). This is partly due to the number of sub-areas in the segmentations. When the number decreased, the sub-areas became larger, lost homogeneity and were classified as neutral. There will thus be segmentations where all sub-areas area classified as neutral (Weight36, Weight56, RES47), or where one of the three classes is missing (G\*103, G\*130) or is small in size compared to the others (RES35) (Fig. 4). If these sub-areas are excluded from the calculations, the mean area for neutral sub-areas drops to less than twice that of the others.

Secondly, looking at the sub-areas with a



**Fig. 3.** The proportion of positive (dash line) and negative (solid line) sub-areas mean areas to neutral ones as percentages.



**Fig. 4.** Two sub-areas, RES;C\*60 (16 sub-areas, left) and G\*87 (17 sub-areas, right), drawn over map. The sub-areas are classified according to  $G_i^*$ -index in negative, neutral, and positive classes.

$G_i^*$ -index near zero reveals that the correlations between the values for Shannon’s entropy index performed on tree diameter as an underlying variable and the localized  $RMSE_{i,loc}$ , and between localized  $RMSE_{i,loc}$  and original  $RMSE_{i,glo}$  are higher in sub-areas classified as either positive or negative  $G_i^*$ -index (Table 5b), than for the sub-areas in the neutral group. The first correlation suggests that the balance in data (i.e., even distribution of diameters) does not lead to lower RMSEs in the models localized to sub-areas. The second implies that the localized  $RMSE_{i,loc}$  is similar to the original global  $RMSE_{i,glo}$  in that

an area with an originally high RMSE stays high in respect to the other localized RMSEs, whereas an originally low RMSE is low in respect to the other localized RMSEs. This is a logical result; the  $G_i^*$ -index near zero by definition means that there is nothing that could be fixed by re-fitting the model containing exactly the same variables as the original one. The correlation between the Shannon’s entropy index and the number of distribution classes in the local data was large and positive. The disadvantage of Shannon’s entropy index is that the natural logarithm of zero is undefined and therefore the index cannot be defined

for empty distribution classes and we had to omit the empty classes from the calculations. The original data here had 13 diameter classes (Table 1b), but for the single sub-areas there were from 4 to 12 classes; i.e., none of the sub-areas had objects in all available classes. All Shannon's entropy indices would have been undefined had they been calculated without our constraint. Therefore Shannon's entropy index is not appropriate to measure balance; the actual number of classes used in the calculations should somehow be included in the index.

Thirdly, the RMSEs alone do not necessarily show the relative importance of the various variables in the data following the localization. Earlier in this study we draw percentages of the models which have no spatial variables and the length of the localized models (Fig. 2a, b). Tree and stand variables were included more often than spatial variables in the localized models. The closer look at the variables in the localized models (Table 6) showed that the heterogeneity in form height in the neutral sub-areas was larger than in positive or negative sub-areas. When the segmentations with mainly neutral sub-areas were removed, the remaining neutral class became similar to the other two classes with respect to form height.

Wulder and Boots (1998) suggested that Getis statistics could be used for image segmentation. Mallinis et al. (2008) used LISA measures and classification and regression trees (CART) in the classification phase after the initial segmentation with eCognition. We have tested these methods separately (Räty and Kangas 2008) (Tables 4, 7, 8). With fewer than 30 sub-areas, the RMSEs were at the same level for both methods, but where more than 30 sub-areas were produced; the maximum RMSEs of CART tended to be higher than those for segmentations of the same size arrived at in the current research. At the same time, the minimum and median RMSEs for the current research's sub-area were lower than for the CART-sub-area. On the other hand, there was no difference between methods in the aggregate standard errors, where the methods were equivalent. In comparison to CART, sub-areas created by our current division procedure are more variable in shape and size. For an example, see Fig. 4, which compares the sub-areas created

by the current research's method, using different input layers, but with the same number of sub-areas in each).

In three of the four segmentations the layer is interpolated from a point data of the local indicators which are calculated from a 20-km neighbourhood. One data point in a dataset actually represents properties of its surroundings and the indicators next to each other are correlated if their neighbourhoods overlap. We use the indicator only for segmentation purposes to show the changes in the values in the layer and do not use them in any calculations. Therefore the above mentioned two properties are not crucial for the work, however their existence is good to realize. The indicators were calculated only for the variable of interest, i.e. the form height of a tree or actually for the residuals of the globally fitted form height model, since we wanted to find homogeneous sub-areas for it. The other variables related to the model are not taken into account in any steps of the segmentation. Their values and properties are used for the model localization and studied afterwards. The method is not meant to be able to form unchangeable sub-areas over time. The sub-areas segmented are unarguably valid only for this data, variable, and model. However, as the differences between CART and the current study were quite small, most probably the same sub-areas are applicable for many different tree-level variables. The method, on the contrary, is adoptable for other datasets, variables, and models. It would also be possible to use a weighted combination of RMSEs of several different models as a reference point, and consider several models at the same time.

To conclude, our purpose was to use the eCognition Professional 4.0 software to segment our study area using various inputs and segmentation parameters, and to document and analyze the effects of this segmentation on the localized form height regression models. The minimum local RMSEs are, after all, promisingly low, implying that the method did identify uniform sub-areas. The maximum local RMSEs are high, however. The methods for improving the estimates are a subject for future study. This could be as simple as, for example, adding variables already available in the inventory data (e.g., site index, site fertility, and temperature sum), or digital terrain

models to the segmentation method and/or the regression model. Judging the segmentations we found that a segmentation containing a layer of  $G_i^*$ -index had a lower aggregate standard error than the segmentation based only on the residuals. When we classified the sub-areas into three classes according to the mean  $G_i^*$ -index value and its variance in the sub-area, we discovered that those sub-areas whose mean  $G_i^*$ -index differed from zero i.e. which were classified as either positive or negative, differed from neutral ones, whose mean  $G_i^*$ -index were near zero. Since neutral sub-areas were more heterogeneous than the negative and positive sub-areas, a possible application could be to use the method presented in this paper to delineate positive and negative sub-areas and then to localize the models only to these sub-areas.

## References

- Anselin, L. 1995. Local indicators of spatial association – LISA. *Geographical Analysis* 27: 93–115.
- Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I. & Heynen, M. 2004. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information. *ISPRS Journal of Photogrammetry & Remote Sensing* 58: 239–258.
- Bock, M., Rossner, G., Wissen, M., Remm, K., Langanke, T., Lang, S., Klug, H., Blaschke, T. & Vrscaj, B. 2005. Spatial indicators for nature conservation from European to local scale. *Ecological Indicators* 5: 322–338.
- Breiman, L., Friedman, J.H., Olshen, R.D. & Stone, C.J. 1984. *Classification and regression trees*. Wadsworth & Brooks, Belmont (CA). 358 p.
- Cressie, N. 1991. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., London. 900 p.
- Draper, N.R. & Smith, H. 1981. *Applied regression analysis*. 2nd ed. Wiley series in probability and mathematical statistics. John Wiley & Sons, Inc., New York (NY). 709 p.
- eCognition 4.0 Professional Users guide. 2004. 2009. [Internet site]. University of Northern British Columbia (UNBC). Available at: <http://www.gis.unbc.ca/help/software/ecognition4/userguide.pdf>. [Cited 20 July 2009].
- Getis, A. & Ord, J.K. 1992. The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24: 189–206.
- & Ord, J.K. 1996. Local spatial statistics: an overview. In: Longley, P.A. & Batty, M. (eds.). *Spatial analysis: modelling in a GIS environment*. GeoInformation International, Cambridge. p. 261–278.
- Guindon, B., Zhang, Y. & Dillabaugh, C. 2004. Landsat urban mapping based on a combined spectral-spatial methodology. *Remote Sensing of Environment* 92(2): 218–232.
- Johansen, K., Coops, N.C., Gergel, S.E. & Stange, Y. 2007. Application of high spatial resolution satellite imagery for riparian and forest ecosystem classification. *Remote Sensing of Environment* 110: 29–44.
- Kangas, A. & Korhonen, K.T. 1995. Generalizing sample tree information with semiparametric and parametric models. *Silva Fennica* 29(2): 151–158.
- Köhl, M. & Gertner, G. 1997. Geostatistics in evaluating forest damage surveys: considerations on methods for describing spatial distributions. *Forest Ecology and Management* 95: 131–140.
- Korhonen, K.T. 1993. Mixed estimation in calibration of volume functions of Scots pine. *Silva Fennica* 27(4): 269–276.
- Longley, P.A., Goodchild, M.F., Maguire, D.J. & Rhind, D.W. 2005. *Geographic information systems and science*. 2nd ed. Wiley, Chichester, Hoboken (NJ). 517 p.
- Mallinis, G., Koutsias, N., Tsakiri-Strati, M. & Karteris, M. 2008. Object-based classification using Quickbird imagery for delineating forest vegetation polygons in a Mediterranean test site. *ISPRS Journal of Photogrammetry & Remote Sensing* 63: 237–250.
- Mustonen, J., Packalén, P. & Kangas, A. 2008. Automatic segmentation of forest stands using canopy height model and aerial photograph. *Scandinavian Journal of Forest Research* 23: 534–545.
- Ord, J.K. & Getis, A. 1995. Local spatial autocorrelation statistics – distributional issues and an application. *Geographical Analysis* 27(4): 286–306.
- Räty, M. & Kangas, A. 2007. Localizing general models based on local indices of spatial association. *European Journal of Forest Research* 129: 279–289.
- & Kangas, A. 2008. Localizing global models with classification and regression trees (CART). *Scandinavian Journal of Forest Research* 23: 419–430.

- Shi, H. & Zhang, L. 2003. Local analysis of tree competition and growth. *Forest Science* 49(6): 938–955.
- Sironen, S., Kangas, A., Maltamo, M. & Kalliovirta, J. 2008. Localization of growth estimates using non-parametric imputation methods. *Forest Ecology and Management* 256: 674–684.
- Straatsma, M.W. & Baptist, M. 2008. Floodplain roughness parameterization using airborne laser scanning and spectral remote sensing. *Remote Sensing of Environment* 112(3): 1062–1080.
- Tiefelsdorf, M. & Boots, B. 1997. A note on the extremities of local Moran's I and their impact on global Moran's I. *Geographical Analysis* 29(3): 248–257.
- Triepke, F.J., Brewer, C.K., Leavell, D.M. & Novak, S.J. 2008. Mapping forest alliances and associations using fuzzy systems and nearest neighbor classifiers. *Remote Sensing of Environment* 112(3): 1037–1050.
- Van Coillie, F.M.B., Verbeke, L.P.C. & De Wulf, R.R. 2007. Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. *Remote Sensing of Environment* 110: 476–487.
- Venables, W. N. & Ripley, B. D. 2002. *Modern applied statistics with S*. 4th ed. Springer, New York. 495 p.
- Weishampel, J.F., Drake, J. B., Cooper, A., Blair, J.B. & Hofton, M. 2007. Forest canopy recovery from the 1938 hurricane and subsequent salvage damage measured with airborne LiDAR. *Remote Sensing of Environment* 190: 142–153.
- Wells, M.L. & Getis, A. 1999. The spatial characteristics of stand structure in *Pinus torreyana*. *Plant Ecology* 143: 153–170.
- Wulder, M. & Boots, B. 1998. Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistics. *International Journal of Remote Sensing* 19(11): 2223–2231.
- , White, J.C., Coops, N.C., Nelson, T. & Boots, B. 2007. Using local spatial autocorrelation to compare outputs from a forest growth model. *Ecological Modelling* 209: 264–276.
- Yu, Q., Gong, P., Clinton, N., Biging, G., Kelly, M. & Schirokauer, D. 2006. Object-based detailed vegetation classification with airborne high spatial resolution remote sensing imagery. *Photogrammetric Engineering and Remote Sensing* 72(7): 799–811.
- Zhang, L., Gove, J.H. & Heath, L.S. 2005. Spatial residual analysis of six modeling techniques. *Ecological Modelling* 186: 154–177.

*Total of 33 references*