GEOMORPHOMETRY 2021
PERUGIA, ITALY
PARTNERS SPONSORS
SEPT 13 - 17 2021

# Burial mound detection using geomorphometry and statistical methods: pixels versus objects

Mihai NICULIȚĂ
Department of Geography,
Faculty of Geography and Geology,
Alexandru Ioan Cuza University of Iași
mihai.niculita@uaic.ro

# Introduction

- Burial mound detection on high resolution data is a practical aspect of geomorphometry that has recently been in the focus of various researchers.
- Considering two of the best rated approaches in the literature (Guyot, 2019 and Niculiță, 2020), one based on a pixel approach and the other based on an object approach, a comparison was performed for the Niculiță (2020) study area in North-Eastern Romania.
- The results of the object based method were presented in 2020, while the results of the method adapted for pixels are presented here.
- Romanian burial mounds are of kurgan type, which appear from the middle Yenisei river to the middle Danube river, and they were constructed starting with the Eneolithic period (6th-5th millennia BP) and the Bronze Age to the Iron Age (3rd-2d centuries BP) by various cultures: Sredny Stog, Usatovo, Jamnaya, Catacomb, Srubnaya and Scythians.

# Study area

0.5 0 0.5 1 1.5 2 km

## Legend

**Mound types**

- landfill (8)
- error (1)
- house (1)
- landslide (2)
- meander cut-off island (1)
- mound (4)
- reservoir (1)
- electric pole base (5)
- tumuli (68)
- vegetation (7)

- settlement
- forest
- anthropic lakes
- rivers
- embankments
- dams

0.5 0 0.5 1 1.5 2 km

## Legend

**Mound types**

- landfill (1)
- vegetation (1)
- landslide (1)
- tumuli (29)

- settlement
- forest
- anthropic lakes
- rivers
- embankments
- dams

**Mound types**

- landfill (8)
- error (1)
- house (1)
- landslide (2)
- meander cut-off island (1)
- mound (4)
- reservoir (1)
- electric pole base (5)
- tumuli (68)
- vegetation (7)

Training = 98 from which 69 BM
Validation = 32 from which 29 BM

a
45
50
b
45
50
34
c
54
22
d
73
e
13
f
69
60
39
73
g
58
64
72
59
71
52
17
40
18
x3
NW
SE
BM
50
x7.5
N
S
45
x3
NW
SE
34
x10
54
22
S
N
electric pole base
x3
N
S
13
x5
W
E
73
x10
39
63
38
N
S
2D
5 m
mound
segment
peak
selected peak
seed
x10
S
60
x10
W
E
69

h
45
i
56
j
34
k
64
72
l
64
72
52
m
40
17
n
1
o
94
p
97
q
10
r
5
x5
64
BM
S
59
72
NW
SE
x5
52
W
E
x10
17
40
mound
2D
5 m
mound
segment
peak
selected peak
seed
x5
97
94
SW
NE
vegetation
x3
1
landslide
x10
10
landfill
x10
5
landfill

T – training BM [68]

V – validation BM [29]

nBM – mound of other type [33]

dem STDDEV
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
NBM
BM
dem RANGE
0 2.5 5 7.5 10 12.5 15
NBM
BM
dem VARIA
0 2.5 5 7.5 10 12.5 15 17.5 20
NBM
BM
ioc Q90
-40 -20 0 20 40 60 80
NBM
BM
ioc MAX
-50 -25 0 25 50 75
NBM
BM
ioc Q45
-60 -40 -20 0 20 40 60
NBM
BM
DmaxsqrtA
1.4 1.6 1.8 2.0 2.2 2.4
NBM
BM
DmaxDir
25 50 75 100 125 150 175
NBM
BM
Fmax90
20 30 40 50 60 70 80
NBM
BM
Local convexity
20 30 40 50 60 70 80 80
NBM
BM
Negative openness
1.40 1.45 1.50 1.55
NBM
BM
Maximum curvature
0.000 0.0025 0.005 0.0075 0.010 0.0125 0.015
NBM
BM

# Methodology – object based

Compound analitical shading [radians]
0.48
0.44
0.40
0.36
0.32
0.28
0.24
0.20
0.16
0.12
0.08
0.04
0.00
a
b
Convexity - cells [%]
80
72
64
56
48
40
32
24
16
c
Convexity - resampling [%]
d
Elevation [m a.s.l.]
160
159
158
157
156
155
154
153
152
151
150
e
Index of convergence
32
24
16
8
0
-8
-16
-24
-32
f
5 m
5 m
mound boundary
segment boundary
segment seed
peak
selected peak
3D ex. 4.5
2D
3D
N

# Local convexity (Iwahashi and Kamiya, 1995; Iwahashi and Pike, 2006)

- Percentage of convex-upward cells within a certain radius of pixels.
- It will include not only the peak as a maximum positive local relief but also the surrounding convex areas.
- This type of convexity is independent of relief magnitude.
- This parameter is ideal for the case of burial mounds, in terms of identifying the convexity of the burial mounds that is surrounded by flat or even convex areas.
- Nonetheless, by coupling the peak selection with local convexity gives the power of the method for selection the segments that are burial mound candidates.

START
Peaks
Local convexity
Local convexity segments
Burial mound
Selected peaks
Local convexity seeds
Local convexity seeds
Burial mound segments
Random Forest
Predicted burial mound segment
condition
prediction
TN
10540
2
FN
FP
42
25
TP
Sensitivity: 0.93
False positive rate:0.004

**Table S1. The list of geomorphometrical variables and their computation settings in SAGA GIS.**

| No | Code | Name | Description |
|---|---|---|---|
| 1 | A | Area | Polygon area |
| 2 | P | Perimeter | Polygon perimeter |
| 3 | P.A | Interior edge ratio | P/A |
| 4 | P.sqrt.A. | | P/sqrt(A) |
| 5 | Depqc | Equivalent projected circle diameter | 2*sqrt(A/π) |
| 6 | Sphericity | Sphericity | The ratio of the perimeter of the equivalent circle to the real perimter |
| 7 | Shape.Index | Shape index | Inverse of Sphericity |
| 8 | Dmax | Maximum diameter | Maximum distance between two polygon part's vertices |
| 9 | DmaxDir | Direction of maximum diameter | |
| 10 | Dmax.A | Dmax/A | |
| 11 | Dmax.sqrt.A | Dmax/sqrt(A) | |
| 12 | Dgyros | Diameter of gyration | Twice the maximum vertex distance to its polygon part's centroid |
| 13 | Fmax | Maximum Feret diameter | |
| 14 | FmaxDir | Direction of the maximum Ferret diameter | |
| 15 | Fmin | Minimum Feret diameter | |
| 16 | FminDir | Direction of the minimum Feret diameter | |
| 17 | Fmean | Mean Feret diameter | |
| 18 | Fmax90 | Feret diameter measured at an angle of 90° to that of the Fmax direction | |
| 19 | Fmin90 | Feret diameter measured at an angle of 90° to that of the Fmin direction | |
| 20 | Fvol | Diameter of a sphere having the same volume as the cylinder constructed by Fmin as the cylinder diameter and Fmax as its length | |
| 21 | dem | Elevation | |
| 22 | ioc | Index of convergence | |
| 23 | conv_r | Local convexity | |
| 24 | nego | Negative openness | |
| 26 | proc | Profile curvature | |
| 27 | plac | Plan curvature | |
| 28 | logc | Longitudinal curvature | |
| 29 | croc | Cross-sectional curvature | |
| 30 | minc | Minimum curvature | |
| 31 | maxc | Maximum curvature | |
| 32 | rare | Real surface area | |
| 33 | wind | Wind exposition index | |
| 34 | tpi | Topographic position index | |
| 35 | vld | Valley depth | |
| 36 | mpi | Morphometric protection index | |
| 37 | tri | Terrain ruggedness index | |
| 38 | vrm | Vector ruggedness measure | |
| 39 | txt | Terrain surface texture | |
| 40 | clo | Local curvature | |
| 41 | cup | Upslope curvature | |
| 42 | clu | Local upslope curvature | |
| 43 | cdo | Downslope curvature | |
| 44 | cdl | Local downslope curvature | |
| 45 | flo | Flow accumulation | |
| 46 | fpl | Flow path length | |
| 47 | spl | Slope length | |
| 48 | cbl | Cell balance | Ratio between flow input and output |
| 49 | twi | Topographic wetness index | SAGA implementation of TWI using a modified catchment area, that is more realistic, compared to standard TWi |
| 51 | dhratio | Diameter-height ratio | Dmax/dem RANGE |
| 52 | compactness | Compactness | (Sqrt(4*(A/π)))/P |
| 53 | formfactor | Form factor | (4*π*A)/(P/2) |
| 54 | roundness | Roundness | (4*A)/(π*Fmax) |
| 55 | elongation | Elongation | Fmax/Fmin |

Table S2. The list of geomorphometrical variables and their computation settings in SAGA GIS.

| No | Code | Name |
|---|---|---|
| 1 | A | Area |
| 2 | P | Perimeter |
| 3 | P.A | Interior edge ratio |
| 4 | P.sqrt.A. | |
| 5 | Depqc | Equivalent projected circle diameter |
| 6 | Sphericity | Sphericity |
| 7 | Shape.Index | Shape index |
| 8 | Dmax | Maximum diameter |
| 9 | DmaxDir | Direction of maximum diameter Dmax/A |
| 10 | Dmax.A | |
| 11 | Dmax.sqrt.A | |
| 12 | Dgyros | Diameter of gyration |
| 13 | Fmax | Maximum Feret diameter |
| 14 | FmaxDir | Direction of the maximum Ferret diameter |
| 15 | Fmin | Minimum Feret diameter |
| 16 | FminDir | Direction of the minimum Feret diameter |
| 17 | Fmean | Mean Feret diameter |
| 18 | Fmax90 | Feret diameter measured at an angle of 90° to that of the Fmax direction |
| 19 | Fmin90 | Feret diameter measured at an angle of 90° to that of the Fmin direction |
| 20 | Fvol | Diameter of a sphere having the same volume as the cylinder constructed by Fmin as the cylinder diameter and Fmax as its length |
| 21 | dem MIN | Minimum elevation |
| 22 | dem MAX | Maximum elevation |
| 23 | dem RANGE | Range of elevation |
| 24 | dem SUM | Sum of elevation |
| 25 | dem MEAN | Mean elevation |
| 26 | dem VARIAN | Elevation variance |
| 27 | dem STDDEV | Standard deviation of elevation |
| 28 | dem Q05 | Elevation percentiles, multiples of 5 |
| 29 | dem Q10 | |
| 30 | dem Q15 | |
| 31 | dem Q20 | |
| 32 | dem Q25 | |
| 33 | dem Q30 | |
| 34 | dem Q35 | |
| 35 | dem Q40 | |
| 36 | dem Q45 | |
| 37 | dem Q50 | |
| 38 | dem Q55 | |
| 39 | dem Q60 | |
| 40 | dem Q65 | |
| 41 | dem Q70 | |
| 42 | dem Q75 | |
| 43 | dem Q80 | |
| 44 | dem Q85 | |
| 45 | dem Q90 | |
| 46 | dem Q95 | |
| 47 | ioc MIN | Minimum index of convergence |
| 48 | ioc MAX | Maximum index of convergence |
| 49 | ioc RANGE | Range of index of convergence |
| 51 | ioc SUM | Sum of index of convergence |
| 52 | ioc MEAN | Mean index of convergence |
| 53 | ioc VARIAN | Index of convergence variance |
| 54 | ioc STDDEV | Standard deviation of index of convergence |
| 55 | ioc Q05 | Indexs of confergence percentiles, multiples of 5 |
| 56 | ioc Q10 | |
| 57 | ioc Q15 | |
| 58 | ioc Q20 | |
| 59 | ioc Q25 | |
| 60 | ioc Q30 | |
| 61 | ioc Q35 | |
| 62 | ioc Q40 | |
| 63 | ioc Q45 | |
| 64 | ioc Q50 | |
| 65 | ioc Q55 | |
| 66 | ioc Q60 | |
| 67 | ioc Q65 | |
| 68 | ioc Q70 | |
| 69 | ioc Q75 | |
| 70 | ioc Q80 | |
| 71 | ioc Q85 | |
| 72 | ioc Q90 | |
| 73 | ioc Q95 | |

# Random Forest – really a black box?

https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/

https://www.mql5.com/en/articles/3856

A decision tree consists of three components: decision nodes, leaf nodes, and a root node. A decision tree algorithm divides a training dataset into branches, which further segregate into other branches. This sequence continues until a leaf node is attained. The leaf node cannot be segregated further. The nodes in the decision tree represent attributes that are used for predicting the outcome.

Plot of the CV error of 5 replicates for 10-folds rfcv with all the variables.

The biggest the number the variables the better, although at some point the results might slightly degrade.
<u>Tuning is needed!</u>

**ATTENTION! Out of bag error is not necessarily meaningful in classification problems, where the confusion matrix is.**

The plot of OOB error showing the mtry and node size tuning with 100 ntree.

# The plot of the minimal tree depth and its mean

Investigating the minimal depth of the trees for each variable and its mean, it can be seen which are the variables that are the most predictive (the smallest mean value of minimal depth) in terms of tree topologies. From the 20 variables plotted we can see that elevation standard deviation, 90th quartile of ioc and elevation range have the longest trees and are always used to split trees at the root. The RF model constructed 100 trees and no limit to the maximum number of terminal nodes trees in a tree was set up, in the figure legend it is visible that trees were split until a depth of 28.

**The plot of accuracy decrease versus the number of nodes and the times a root for all the variables.**

Variable importance is given by the difference in the OBB error between the situation when only the variable is perturbed and when it is not, for that tree. We can see plotted all the variables used, as a relation between accuracy decrease (if that variable is perturbed) and the number of node splits for that variable, the size of the circle indicating the number of times that the variable is a root. The blue-colored circles correspond to the first ten most important variables. We can see again here variables, which are also predictive and have a presence as the number of nodes and times a root. There is a clear relationship between the accuracy decrease and number of nodes (as the number of nodes decreases the accuracy is not so decreased), but there are also not so important variables (by accuracy decrease, mainly index of convergence percentiles), with a small number of nodes but used many times as a root. This situation shows that using these variables in the model and not removing them, although will increase the OBB error, will allow the model to increase the true positives and decrease the false positives.

# Index of convergence (Koethe and Lehmeier, 1996)

- Index of convergence/divergence is computed in regard to overland flow. By its meaning it is similar to plan or horizontal curvature, but gives much smoother results and is better emphasizing the peaks and pits. The calculation uses the aspects of surrounding cells, i.e. it looks to which degree surrounding cells point to the center cell. The result is given as percentages, negative values correspond to convergent, positive to divergent flow conditions. Minus 100 would be like a peak of a cone (a), plus 100 a pit (c), and 0 an even slope (b).

## Mean minimal depth for the 30 most frequent interactions.

The standard deviation of altitude had the biggest occurrence, and the unconditional mean minimal depth of it in the forest was almost equal with its mean minimal depth across maximal subtrees. Beside altitude standard deviation, range and variance, which are measures of local peak amplitude, index of convergence measures (maximum and percentiles) define convex features that have local prominence. Shape descriptors appeared to be used for node splitting and interacting with the previous variables at a certain point, the following being used: the ratio between the diameter and square root of area (a measure of similarity to a square, since square root of area, is the side length), Feret diameter measured at an angle of 90° to that of the maximum Feret diameter and shape index.

Prediction plots (**left**) and partial dependence plots (**right**) for the most three important variables.

**The burial mound watershed delineation based on both methods of local convexity computation: (a) (2D) and (b) (3D) for cell count method, and (c) (2D) and (d) (3D) for the cell count interpolation method.**

The "goodie" of the object based method is that it can even delineate the burial mounds. Although this was not necessarily targeted, the results are impressive. The results are limited by the use of local convexity and the degree of smoothing of the burial mounds due to erosion. Ploughing even created a concave feature around the convex part.

Once the concave parts are found the method could be extend to map also the concave features.

The accuracy statistics of the watershed segmentation for the northern study area.

# The pixel based approach – better confusion matrix?

| RF parameters | OBB error | TP | TN | FN | FP | SNS | FPR |
|---|---|---|---|---|---|---|---|
| 1000 segments from which 75% burial mounds, 100 ntree, 5 mtry, 1 nodesize* | 3.1 | 25 | 10536 | 2 | 46 | 0.93 | 0.004 |
| 17858 pixels from which 8929 burial mounds (50%), 400 ntree, 3 mtry, 1 nodesize | 0.51 | 4829 | 3452293 | 878 | 171896 | 0.97 | 0.047 |

Training
27.15°E
27.20°E
47.30°N
47.25°N
Bahlui
Legend
0.5 0 0.5 1 1.5 2 km
ID from Niculita (2020)
burial mound (68)
TP
TN
FN
FP
anthropic lakes
rivers

Valdiation
27.15°E
27.20°E
47.20°N
47.15°N
Valea Oii
Bahluiet
Albești
Goești
Iazul între Văi
Sinești
Gâmboasa
Legend
0.5 0 0.5 1 1.5 2 km
ID from Niculita (2020)
burial mound (29)
TP
TN
FN
FP
anthropic lakes
rivers

The RF parameters has impact mainly on the FP pixels.
Although the confusion matrix numeric variables might “look” better, actually the spatial representation is showing that wide areas are mapped as BM.

18
19-20
22-23
14
TP
FN
TN
FP

- Flow path length becomes the most important since burial mounds are local maxima.
- Also valley depth and flow accumulation, after which the altitude and the slope.
- In the pixel based approach the curvature indices do not perform very well, although index of convergence and some curvatures are present in the first 20 variables.

- The first 4 important variables (flow path length, valley depth, flow length and altitude) are producing more nodes.
- The majority of the other variables produce about to a half than the first.
- The conclusion is that in the pixel based approach the depth of the trees is not big and the nodes are more frequent because the feature space is not well defined.
- Some few variables are defining roots, and there is no variation that can explain the different between BM and FP.

Mean minimal depth for 30 most frequent interactions
mean_min_depth
0
1
2
3
4
vld:fpl5
flo:fpl5
fpl5:vld
fpl5:dem
vld:slope
fpl5:cdo
slope:fpl5
flo:twi
fpl5:wind
fpl5:fpl5
vld:dem
flo:dem
flo:vld
vld:ioc
fpl5:cbl
fpl5:minc
vld:vld
vld:maxc
fpl5:mpi
fpl5:slope
flo:nego
fpl5:rare
fpl5:tri
vld:cup
slope:dem
slope:flo
vld:flo
fpl5:ioc
fpl5:nego
vld:twi
interaction
occurrences
345
340
335
330
minimum
unconditional

Prediction of the forest for different values of fpl5 and vld

Distribution of pixels in the training set

Distribution of pixels in the valdiation set

# Conclusions

- I have shown that the object-based approach performs better than the pixel-based approach, both in terms of confusion matrix measures and in practical validation.
- Attention is needed to the overall measures of confusion matrix like accuracy, which might be misleading: a big number of false positives can render the results unusable in practice, especially in imbalanced cases.
- The explanation of the superiority of the object-based approach is given by its power of feature space reduction, spatial aggregation and the usability of shape descriptors in the RF model. In this way the RF algorithm is able to find the geomorphometric signature of the burial-mound segments.
- Geomorphometric expert opinion is needed when landform classification is targeted.
- AI & ML algorithms are able to learn from even slight differences given by certain geomorphometric variables; a wide range of geomorphometric variables need to be investigated in this context through tuning.

# Thank you for your attention!

- Niculiță (2020) Geomorphometric Methods for Burial Mound Recognition and Extraction from High-Resolution LiDAR DEMs, Sensors, https://www.mdpi.com/1424-8220/20/4/1192/htm