GeoPAT 2.0 user's manual

Note: This manual is work in progress

`http://sil.uc.edu`

August 22, 2017

Contents

1 Introduction

GeoPAT (Geospatial Pattern Analysis Toolbox) is a standalone suite of modules written in C and dedicated to analysis of large Earth Science datasets in their entirety using spatial and/or temporal patterns. Global scale, high resolution spatial datasets are available but are mostly used in small pieces for local studies. GeoPAT enables studying them in their entirety. GeoPATs core idea is to tessellate global spatial data into grid of square blocks of original cells (pixels). This transforms data from its original form (huge number of cells each having simple content) to a new form (much smaller number of supercells/blocks with complex content). Complex cell contains a pattern of original variable. GeoPAT provides means for succinct description of such patterns and for calculation of similarity between patterns. This enables spatial analysis such as clustering, segmentation, and search to be performed on the grid of complex cells (local patterns).

1.1 Basic concepts

- **Grid** is a topological structure applied to input data. Grid is a lattice of cells defined by its geographical position, spatial extent, number of cells, and cell size which defines grid resolution. Each cell in the grid contains a vector of values.
- **Pattern signatures** - compact numerical descriptors of patterns. We use histograms of patterns features as scene signatures.
- **Motifel** is an elementary unit of analysis - a square block of pixels representing local pattern. Motifel is represented by histogram of features (co-occurrence, decomposition). Distance between motifels is a distance between their histograms (for example using Jensen-Shannon Divergence).

2 GeoPAT 2.0 architecture

2.1 General description

The GeoPAT consists of twelve modules (Fig. 1). Five modules are designed for extracting pattern signatures from an original data grid, four modules for actual geoprocessing of the patterns and three utility modules.

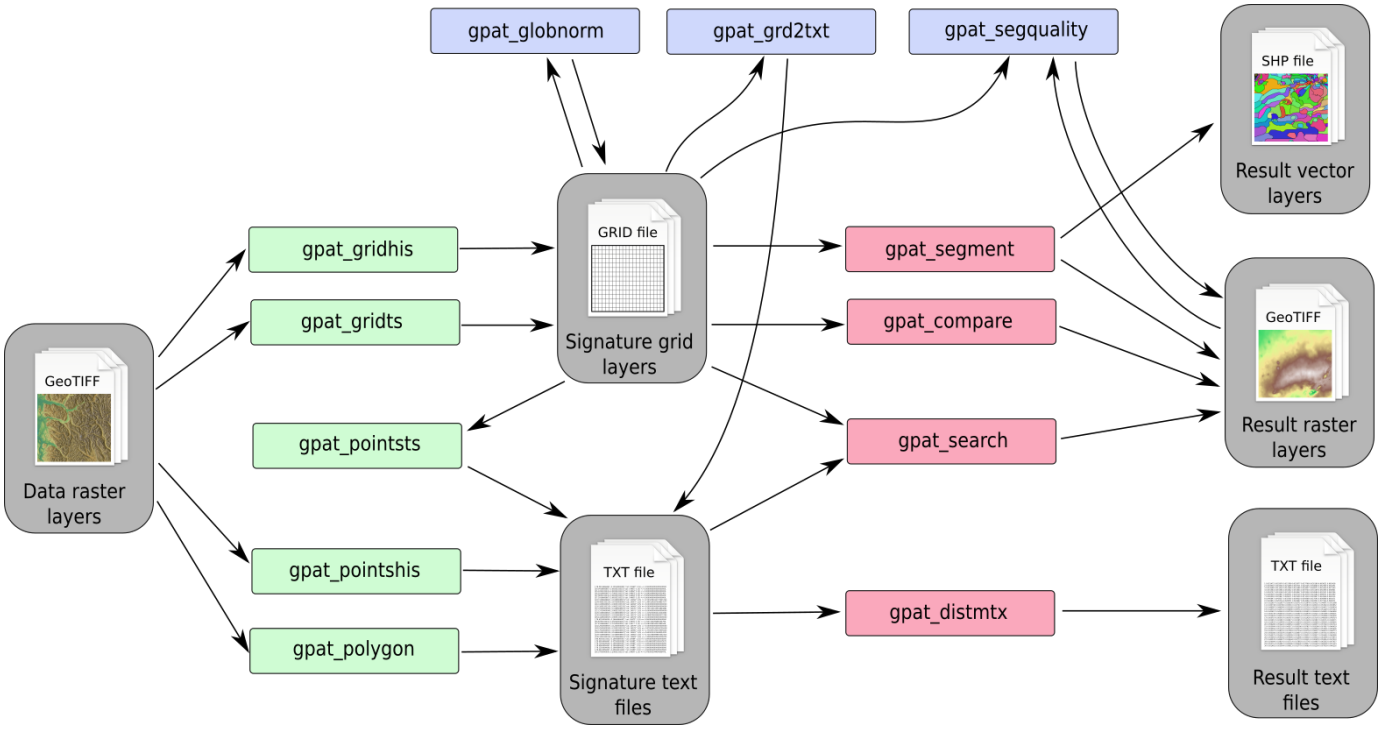

Figure 1: Outline of the GeoPAT 2.0 architecture

The role of signature extraction modules (gpat_gridhis, gpat_gridts, gpat_pointshis, gpat_pointsts, gpat_polygon) is to calculate pattern signatures for scenes defined:

- **gpat_gridhis** - by a neighborhood of each cell in a grid (spatial pattern)
- **gpat_gridts** - by a neighborhood of each cell in a grid (spatiotemporal pattern)
- **gpat_pointshis** - in the neighborhoods of selected points (spatial pattern)
- **gpat_pointsts** - in the neighborhoods of selected points (spatiotemporal pattern)
- **gpat_polygon** - over irregular polygons

The role of geoprocessing modules (gpat_search, gpat_compare, gpat_segment, gpat_distmtx) is to perform:

- **gpat_search** - searching

- **gpat_compare** - comparision
- **gpat_segment** - segmentation (based on pattern data generated by the signature extraction modules)
- **gpat_distmtx** - clustering

The role of utility modules (gpat_grd2txt, gpat_globnorm, gpat_segquality) is to:

- **gpat_grd2txt** -
- **gpat_globnorm** -
- **gpat_segquality** - calculate the quality metrics (inhomogeneity and isolation) of a segmentation

2.2 GeoPAT Modules

2.2.1 Signature building

2.2.1.1 gpat_gridhis

Creates a binary grid of signatures from a categorical raster map(s).

Usage:

```
gpat_gridhis [−lh] −i <file_name> −o <file_name> [−s <signature_name>] [−−level=<n>] [−z <n>]
    [−f <n>] [−n <normalization_name>] [−t <n>]

−i, −−input=<file_name> name of input file (GeoTIFF)
−o, −−output=<file_name> name of output file (GRID)
−s, −−signature=<name> motifel's signature (use −l to list all signatures, default: 'cooc')
−−level=<n> full decomposition level (default: 0, auto)
−z, −−size=<n> motifel size in cells (default: 150)
−f, −−shift=<n> shift of motifels (default: 100)
−n, −−normalization=<name> signature normalization method (use −l to list all methods, default: 'pdf')
−l list all signatures and normalization methods
−t <n> number of threads (default: 1)
−h, −−help print this help and exit
```

Options:

input

Defines a categorical raster layer(s) which will be used as a source for extracting pattern signature. Layers must be a categorical one. For the Cartesian product method ('prod') there can be more than one input map. Other methods use a one map only (the first one provided). In order to provide more than one input map, type multiple input options ("-i map1.tif -i map2.tif" or "–input=map1.tif –input=map2.tif").

output

Output consists of two files: one of them is a dataset containing a grid of signatures in binary form, the other one is a header text file (the .hdr extension) containing a grid topology and an information about the input data parameters. Modifying the header is strongly discouraged as tt may cause some calculations to fail. Structure of the header is as follows:

- dim – number of dimensions of each signature
- dims – size of each dimension
- type – type of data stored in the grid (integer, float, etc.)
- at0 – top left x
- at1 – w-e grid resolution
- at2 – rotation (0 if grid is north-up)
- at3 – top left y
- at4 – rotation (0 if grid is north-up)
- at5 – n-s grid resolution
- rows – number of rows in the grid
- cols – number of cols in the grid
- proj – projection in wkt style
- desc – command used to create the grid

size

Size of a motifel (local calculation window) expressed in the number of pixels. It defines the extent of a local pattern. It is the length of the side of square-shaped block of pixels (motifel). It must be at least 10 and cannot exceed the input map size.

shift

Parameter defines the shift between adjacent scenes along the grid in n-s and w-e directions. It describes the density of the output gird and defines a new topology of the grid. Formula original_resolution $\times$ shift = new_resolution explains how resolution of the original map will be reduced. If shift is set to the same value as 'size', the input map will be simply divided into a grid of non-overlapping motifels. Setting shift to a value smaller than 'size' parameter will result in grid of overlapping motifels. Shift cannot be larger than 'size', and cannot be smaller than 5.

signature

Defines method of calculating signatures of motifels. GeoPAT offers the following methods:

- prod – Cartesian product of input category lists
- cooc – spatial coocurrence of categories
- sdec – simple 2-level decomposition
- fdec – full decomposition
- lbp – histogram of local binary patterns
- lind – landscape indices vector
- linds – selected landscape indices vector

See Appendix B for the details.

level

This option is used only if 'full decomposition' (fdec) is used. It defines the highest level of decomposition. See Appendix B for the details.

normalization

Specifies normalization method used on the signatures. See Appendix B for the details.

threads (-t)

The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

This module extracts a "grid-of-scenes" (grid of pattern signatures, grid of motifels). The output is a grid of the same spatial extent as the input raster map, but with a different cell size. Each cell in a new grid has only one attribute - the signature of its pattern. Pattern is calculated over a window centered on the cell and having a user-defined size. Resolution of the output grid equals to the resolution of the input raster map multiplied by the shift parameter. A signature of pattern for each scene is stored as a numerical vector in a binary form. The module outputs a header file (.hdr) containing a topology of a grid-of-scenes and a binary file containing signatures ordered by rows.

This module uses categorical raster data in GeoTIFF format as an input. It might be more than one map for the Cartesian product signature. Raster maps must be categorical and its size must be greater than the scene size specified by the user. Size defines the size of individual scene for which the histogram is calculated. Shift shows how scenes shift along the grid in n-s and w-e directions. Shift defines the resolution of the new grid. If window size is bigger than shift (recommended), motifels will overlap. If shift and size is equal, windows will not overlap. Shift cannot be greater than size.

The output grid may be an input to one of the following GeoPAT modules: *gpat_search, gpat_compare, gpat_segment, gpat_segquality, gpat_grd2txt, and gpat_globnorm.*

2.2.1.2 gpat_gridts

Usage:

```
gpat_gridts [−nh] −i <file_name> [−i <file_name>]... −o <file_name> [−d <n>]

−i, −−input=<file_name> name of input file(s) (GeoTIFF)
−o, −−output=<file_name> name of output file (GRID)
−d, −−dimension=<n> dimension of vector that describes time series element (default: 1)
−n, −−normalize normalize each vector coordinate to [0.0, 1.0] (default: no)
−h, −−help print this help and exit
```

input

output

dimension

normalize

Description:

2.2.1.3 gpat_pointshis

Calculates numerical signatures of individual motifels within a raster map.

Usage:

```
gpat_pointshis [−lah] −i <file_name> −o <file_name> [−s <signature_name>] [−−level=<n>] [−z <n>]
    [−n <normalization_name>] [−x <double>] [−y <double>] [−d <string>] [−−xy_file=<file_name>]

−i, −−input=<file_name> name of input file (GeoTIFF)
−o, −−output=<file_name> name of output file (TXT)
−s, −−signature=<name> motifel's signature (use −l to list all signatures, default: 'cooc')
−−level=<n> full decomposition level (default: 0, auto)
−z, −−size=<n> motifel size in cells (default: 150)
−n, −−normalization=<name> signature normalization method (use −l to list all methods, default: 'pdf')
−l list all signatures and normalization methods
−x <double> x coord
−y <double> y coord
−d, −−description=<string> description of the location
−−xy_file=<file_name> name of file with coordinates (TXT)
−a, −−append append results to output file
−h, −−help print this help and exit
```

Options:

input

Defines categorical raster map(s) in the GeoTIFF format, which will be used as a source for extracting pattern signature. For the Cartesian product method ('prod') there can be more than one input map. Other methods use one map only (the first one provided). In order

to provide more than one input map, type multiple input options ("-i map1.tif -i map2.tif" or "–input=map1.tif –input=map2.tif").

output

Name of output text file containing signatures. In the output file, each line corresponds to a single motifel's signature. Each signature is preceded by its header. A header always consists of: coordinates of the midpoint of motifel (scene), name, number of dimensions, and length of each dimension. Modifying the signature file is strongly discouraged as it may cause some calculations to fail.

signature

Defines method for calculating signatures of motifels. GeoPAT offers the following methods:

- prod – Cartesian product of input category lists
- cooc – spatial coocurrence of categories
- sdec – simple 2-level decomposition
- fdec – full decomposition
- lbp – histogram of local binary patterns
- lind – landscape indices vector
- linds – selected landscape indices vector

See Appendix B for details.

level

This option is used only when 'full decomposition' (fdec) is used. It defines the highest level of decomposition. See Appendix B for the details.

size

Size of a motifel (scene) for which signature is calculated. Scene is a square centered at the coordinate pair location, while its extent is defined by size.

normalization

Specifies normalization method used on the signatures. See Appendix B for the details.

x

X coordinate of the midpoint of motifel for which signature will be calculated.

y

Y coordinate of the midpoint of motifel for which signature will be calculated.

description

Description of a motifel. It can be a name of location, description of pattern, or anything else. If not provided, the default description is "location".

xy_file

Name of a text file with list of coordinates. This option is useful for calculating signatures for multiple motifels in a single run.

append (flag)

Append mode. Useful when using the coordinate pair mode and when scenes are processed one by one (see description below). If the append flag is used and the output file already exists, signatures will be appended at the end of file instead of overwriting it.

Description:

This module extracts signatures for a scene (motifel) or collection of scenes defined over square-shaped windows. User provides coordinates of the center of each scene and the size of the scene. The module outputs a list of scene-labeled signatures. As an input the module uses categorical raster map in the GeoTIFF format (or a few maps if user calculates the Cartesian product of multiple maps), coordinates of the center of each scene and size of scenes. The coordinates must be in the input map's coordinate system. There are two ways of defining the scenes:

Definition by coordinate pair: This is the simplest mode designed for a batch or interactive processing. In this mode, scene is defined by a pair of its midpoint's coordinates (the *x* and *y* options) and scene size parameter (the *size* option). Only one scene can be processed at once. The *xy_file* option is not used in this mode. If used with the append flag (-a), signatures can be calculated one by one and added to the same output file. Additionally, the *description* option lets user name a scene. This name will be stored in the output file.

Definition by text file: In this mode, scene midpoint's coordinates are defined in a external text file in which each line contains X,Y coordinates and, optionally, name/description using the following syntax: x_coordinate,y_coordinate,scene_name. The name of a file with coordinates is provided using the *xy_file* option. Extent of a scene is defined by the *size* parameter. The *x, y* and *description* options are not used in this mode. Example of the content of coordinates file:

```
1260500, 1277638, water
1289493, 1251381, urban
1304934, 1285000, crops
1316205, 1253206, pasture
1277700, 1261936
```

If a description is not provided (like in the example above), the program will assign a default name ("location") with id representing the number of line in the coordinate file.

The output signature text file can be used as an input to the following modules: *gpat_search*, *gpat_distmtx*.

2.2.1.4 gpat_pointsts

Usage:

```
gpat_pointsts [−ah] −i <file_name> −o <file_name> [−x <double>] [−y <double>] [−d <string>]
    [−−xy_file=<file_name>]

−i, −−input=<file_name> name of input file (GRID)
−o, −−output=<file_name> name of output file (TXT)
−x <double> x coord
−y <double> y coord
−d, −−description=<string> description of the location
−−xy_file=<file_name> name of file with coordinates (TXT)
−a, −−append append results to output file
−h, −−help print this help and exit
```

input

output

x coord (-x)

y coord (-y)

description

xy_file

append

Description:

2.2.1.5 gpat_polygon

Calculates numerical signatures of irregular regions.

Usage:

```
gpat_polygon [−lh] −i <file_name> −e <file_name> −o <file_name> [−s <signature_name>] [−n
    <normalization_name>] [−t <n>]

−i, −−input=<file_name> name of input file (GeoTIFF)
−e, −−segments=<file_name> name of input file (GeoTIFF, int)
−o, −−output=<file_name> name of output file (TXT)
−s, −−signature=<name> signature method (use −l to list all methods, default: 'cooc')
−n, −−normalization=<name> signature normalization method (use −l to list all methods, default: 'pdf')
−l list all signatures and normalization methods
−t <n> number of threads (default: 1)
−h, −−help print this help and exit
```

Options:

input

Categorical raster map(s) in the GeoTIFF format, which will be used as a source for extracting pattern signature. For the Cartesian product method ('prod') there can be more than one input map. Other methods only use one map (the first one provided). In order to provide more than one input map, type multiple input options ("-i map1.tif -i map2.tif" or "–input=map1.tif –input=map2.tif").

segments

Categorical raster map in the GeoTIFF format which defines spatial division of area of interest into polygonal regions. In this map, regions should be defined as areas of unique ids. Layer must be a raster and its projection, extents and resolution should be identical to the main input map. This layer may be a result of the segmentation module *gpat_segment* or any other categorical map (e.g. map of ids of watersheds).

output

Name of output text file containing signatures. In the output file each line corresponds to a region's signature. Each signature is preceded by its header. A header always consists of: the coordinates of the midpoint of region, name, number of dimensions, and length of each dimension. Modifying the signature file is strongly discouraged as it may cause some calculations to fail.

signature

Defines method for calculating signatures of motifels. GeoPAT offers the following methods:

- prod – Cartesian product of input category lists
- cooc – spatial coocurrence of categories
- sdec – simple 2-level decomposition
- fdec – full decomposition
- lbp – histogram of local binary patterns
- lind – landscape indices vector
- linds – selected landscape indices vector

See Appendix B for the details.

normalization

Specifies normalization method used on the signatures. See Appendix B for the details.

threads (-t)

The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

Module extracts signatures for a collection of irregular regions using the same methods as in the *gpat_pointshis* and *gpat_gridhis* modules. As an input, apart from categorical raster map from which signatures are calculated, user provides a categorical raster map which defines spatial division of area of interest into polygonal regions. The module outputs a list of polygon labeled-signatures stored in the form of a text file. Output can be transferred to the p.sim.distmatrix for further processing.

The output signature text file can be used as an input to the following modules: *gpat_search*, *gpat_distmtx*.

2.2.2 Similarity measuring

2.2.2.1 gpat_search

Produces similarity maps which show similarity value between query motifels (scenes) and every motifel in the input grid.

Usage:

```
gpat_search [−dlh] −i <file_name> [−o <file_name>] −r <file_name> [−m <measure_name>]
    [−−type=Byte/....] [−p <file_name>] [−n <n>] [−t <n>]

−i, −−input=<file_name> name of input file (GRID)
−o, −−output=<file_name> name of output file (TIFF)
−r, −−reference=<file_name> reference data to calculate similarity (TXT)
−d, −−description use description of the reference histogram(s) as name of output file(s)
−m, −−measure=<measure_name> similarity measure (use −l to list all measures; default 'jsd')
−l, −−list_measures list all measures
−−type=Byte/.... output data type (default: Float64)
−p, −−palette=<file_name> name of the file with colors definition (CSV)
−n, −−no_data=<n> output NO DATA value (default: none)
−t <n> number of threads (default: 1)
−h, −−help print this help and exit
```

input

Binary file containing grid of signatures (grid-of-scenes). Grid is mandatory. User has to provide a name of grid file (without the .hdr extension). Header file is read automatically. Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Header of the grid will define topology of the output layers. Do not modify header file.

output

Raster (GeoTIFF) file or files containing the map of similarity. The number of outputs depends on the number of scenes/polygons in the input reference file.

reference

Text file containing one or more signatures of either individual motifels (scenes) or irregular polygons. This option is mandatory. Input text file can be created using the *gpat_pointshis*, *gpat_pointsts* or *gpat_polygons* module. The number of outputs depends on

the number of scenes in the input scene file. At least one scene is required.

description

measure

Defines method for calculating distance between motifels. GeoPAT offers the following measures:

- jsd – Jensen Shannon Divergence
- euc – Euclidean distance
- eucn – Normalized euclidean distance
- eucp – Normalized euclidean distance (periodic)
- wh – Wave-Hedges distance
- cos – Cosine distance
- jac – Jaccard distance
- roz – Rozicka distance
- rozp – Rozicka distance (extended)
- hass – Hassanat distance
- tsEUC – time series - euclidean distance
- tsEUCP – periodic time series - euclidean distance
- tsDTW – time series - Dynamic Time Warping distance
- tsDTWP – time series - Periodic Dynamic Time Warping distance
- tsDTWPa – time series - Synchronized Dynamic Time Warping distance

See Appendix C for the details.

list_measures

Lists all the available measures.

type

Data type of the output raster (GeoTIFF) file. Types of Byte, UInt16, Int16, UInt32, Int32, Float32, Float64, CInt16, CInt32, CFloat32 and CFloat64 are supported.

palette

no_data

Specify NO DATA value of the output raster (GeoTIFF) file.

threads (-t)

The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

2.2.2.2 gpat_compare

Usage:

```
gpat_compare [−lh] −i <file_name> −i <file_name> −o <file_name> [−−type=Byte/....] [−p
    <file_name>] [−n <n>] [−m <measure_name>] [−t <n>]

−i, −−input=<file_name> name of input files (GRID)
−o, −−output=<file_name> name of output file with similarity (TIFF)
−−type=Byte/.... output data type (default: Float64)
−p, −−palette=<file_name> name of the file with colors definition (CSV)
−n, −−no_data=<n> output NO DATA value (default: none)
−m, −−measure=<measure_name> similarity measure (use −l to list all measures; default 'jsd')
−l, −−list_measures list all measures
−t <n> number of threads (default: 1)
−h, −−help print this help and exit
```

input

Two binary files containing grid of signatures (grid-of-scenes). User has to provide names of grid files (without the .hdr extension). Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Headers of the grids will define topology of the output layers. Do not modify header files.

output

Raster (GeoTIFF) file containing the map of similarity between two input grid of signatures.

type

Data type of the output raster (GeoTIFF) file. Types of Byte, UInt16, Int16, UInt32, Int32, Float32, Float64, CInt16, CInt32, CFloat32 and CFloat64 are supported.

palette

no_data

Specify NO DATA value of the output raster (GeoTIFF) file.

measure

Defines method for calculating distance between motifels. GeoPAT offers the following measures:

- jsd – Jensen Shannon Divergence
- euc – Euclidean distance
- eucn – Normalized euclidean distance
- eucp – Normalized euclidean distance (periodic)
- wh – Wave-Hedges distance
- cos – Cosine distance
- jac – Jaccard distance
- roz – Rozicka distance
- rozp – Rozicka distance (extended)
- hass – Hassanat distance
- tsEUC – time series - euclidean distance
- tsEUCP – periodic time series - euclidean distance
- tsDTW – time series - Dynamic Time Warping distance
- tsDTWP – time series - Periodic Dynamic Time Warping distance
- tsDTWPa – time series - Synchronized Dynamic Time Warping distance

See Appendix C for the details.

list_measures
Lists all the available measures.

threads (-t)
The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

This module compares two grids-of-scenes (grid of histograms) in a scene-by-scene fashion. Both grids must be calculated using the same shift parameter and the same method. The output of gpat_compare is a raster having the same topology as the input grids. Each cell in the output file contains the value of similarity between corresponding pairs of scenes where one scene is coming from the first grid and the other scene is coming from the second grid. The module is useful for comparing data of the same area created at two different times. It is also useful for estimating the uncertainty of the results.

2.2.2.3 gpat_segment

Usage:

```
gpat_segment [−lcgaqh] −i <file_name> −o <file_name> [−v <file_name>] [−−size=<n>] [−m
    <measure_name>] [−−lthreshold=<double>] [−−uthreshold=<double>] [−−swap=<double>]
    [−−minarea=<n>] [−−maxhist=<n>] [−t <n>]

−i, −−input=<file_name> name of input files (GRID)
−o, −−output=<file_name> name of output file with segments (TIFF)
−v, −−vector=<file_name> name of output vector file with segments (SHP)
−−size=<n> output resolution modyfier (default: 1)
−m, −−measure=<name> similarity measure (use −l to list all measures; defult: jsd)
−l, −−list_measures list all measures
−−lthreshold=<double> minimum distance threshold to build areas (default: 0.1)
−−uthreshold=<double> maximum distance threshold to build areas (default: 0.3)
−−swap=<double> improve segmentation by swapping unmatched areas. −1 to skip (default: 0.001)
−−minarea=<n> minimum number of motifels in individual segment (default: 0)
−−maxhist=<n> create similarity/distance matrix for maxhist histograms; leave 0 to use all (default:
    200)
−c, −−complete use complete linkage (default is average)
−g, −−no_growing skip growing phase
−a, −−no_hierarhical skip herarhical phase
−q, −−quad quad mode (rook topology)
−t <n> number of threads (default: 1)
−h, −−help print help and exit
```

input

Binary file containing grid of signatures (grid-of-scenes). Grid is mandatory. User has to provide a name of grid file (without the .hdr extension). Header file is read automatically. Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Header of the grid will define topology of the output layers. Do not modify header file.

output

Categorical raster map in the GeoTIFF format which defines spatial division of area of interest into polygonal regions (segments). In this map, regions (segments) are defined as areas of unique ids.

vector

Vector (shapefile) map of segmentation. This data is automatically created by vectorization of the output file.

size

measure

Defines method for calculating distance between motifels. GeoPAT offers the following measures:

- jsd – Jensen Shannon Divergence

- euc – Euclidean distance
- eucn – Normalized euclidean distance
- eucp – Normalized euclidean distance (periodic)
- wh – Wave-Hedges distance
- cos – Cosine distance
- jac – Jaccard distance
- roz – Rozicka distance
- rozp – Rozicka distance (extended)
- hass – Hassanat distance
- tsEUC – time series - euclidean distance
- tsEUCP – periodic time series - euclidean distance
- tsDTW – time series - Dynamic Time Warping distance
- tsDTWP – time series - Periodic Dynamic Time Warping distance
- tsDTWPa – time series - Synchronized Dynamic Time Warping distance

See Appendix C for the details.

list_measures
Lists all the available measures.

lthreshold
Controls segment's sizes.

uthreshold
Prevents growth of inhomogeneous segments.

swap

minarea
Minimum number of cells in an individual segment not subject to removal process. The removal proecess is turned off if the value equals to 0.

maxhist

complete

no_growing

no_hierarhical

quad

threads (-t)
The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

2.2.2.4 gpat_distmtx

Usage:

```
gpat_distmtx [−lsh] −i <file_name> −o <file_name> [−m <measure_name>]

−i, −−input=<file_name> name of input file witch signatures (TXT)
−o, −−output=<file_name> name of output file (CSV) with similarity matrix
−m, −−measure=<name> similarity measure (use −l to list all measures; default 'jsd')
−l, −−list_measures list all measures
−s, −−similarity output is a similarity matrix
−h, −−help print this help and exit
```

input

output

measure
Defines method for calculating distance between motifels. GeoPAT offers the following measures:

- jsd – Jensen Shannon Divergence
- euc – Euclidean distance
- eucn – Normalized euclidean distance
- eucp – Normalized euclidean distance (periodic)
- wh – Wave-Hedges distance
- cos – Cosine distance
- jac – Jaccard distance
- roz – Rozicka distance

- rozp – Rozicka distance (extended)
- hass – Hassanat distance
- tsEUC – time series - euclidean distance
- tsEUCP – periodic time series - euclidean distance
- tsDTW – time series - Dynamic Time Warping distance
- tsDTWP – time series - Periodic Dynamic Time Warping distance
- tsDTWPa – time series - Synchronized Dynamic Time Warping distance

See Appendix C for the details.

list_measures
Lists all the available measures.

similarity
Description:

2.2.3 Tools

2.2.3.1 gpat_grd2txt

Usage:

```
gpat_grd2txt [−h] −i <file_name> −o <file_name>

−i, −−input=<file_name> name of input file (GRID)
−o, −−output=<file_name> name of output file (TXT)
−h, −−help print this help and exit
```

input
Binary file containing grid of signatures (grid-of-scenes). Grid is mandatory. User has to provide a name of grid file (without the .hdr extension). Header file is read automatically. Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Header of the grid will define topology of the output layers. Do not modify header file.

output
Description:

2.2.3.2 gpat_globnorm

Usage:

```
gpat_globnorm [−lh] −i <file_name> −o <file_name> [−m <method_name>] [−t <n>]

−i, −−input=<file_name> name of input file (GRID)
−o, −−output=<file_name> name of output file (GRID)
−m, −−method=<method_name> normalization method (use −l to list all methods, default: '01')
−l, −−list_methods list all methods
−t <n> number of threads (default: 1)
−h, −−help print this help and exit
```

input

Binary file containing grid of signatures (grid-of-scenes). Grid is mandatory. User has to provide a name of grid file (without the .hdr extension). Header file is read automatically. Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Header of the grid will define topology of the output layers. Do not modify header file.

output

Normalized binary file containing grid of signatures (grid-of-scenes) and the header text file (the .hdr extension) containing a grid topology and an information about the input data parameters.

method

Specifies normalization method used on the signatures.

- 01 – normalize coordinates to [0,1]
- N01 – normalize coordinates to N(0,1)
- ind01 – normalize coordinates to [0,1] for 72 landscape indices

See Appendix B for the details.

list_methods

Lists all the available normalization methods.

threads (-t)

The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

2.2.3.3 gpat_segquality

Usage:

```
gpat_segquality [−lcqwh] −i <file_name> −s <file name> [−g <file_name>] [−o <file_name>] [−m
    <measure_name>] [−−maxhist=<n>] [−t <n>]

−i, −−input=<file_name> name of input file (GRID)
−s, −−segments=<file name> name of input segmentation map (TIFF)
−g, −−inhomogeneity=<file_name> name of output file with segment inhomogeneity (TIF
F)
−o, −−isolation=<file_name> name of output file with segment isolation (TIFF)
−m, −−measure=<name> similarity measure (use −l to list all measures; defult: jsd)
−l, −−list_measures list all measures
−−maxhist=<n> create similarity/distance matrix for maxhist histograms; leave 0 to use all (default:
    200)
−c, −−complete use complete linkage (default is average)
−q, −−quad quad mode (rook topology)
−w, −−no_weight switch off edge−based weighting in isolation
−t <n> number of threads (default: 1)
−h, −−help print help and exit
```

input

Binary file containing grid of signatures (grid-of-scenes). Grid is mandatory. User has to provide a name of grid file (without the .hdr extension). Header file is read automatically. Grid is an output from either *gpat_gridhis* or *gpat_gridts* module. Header of the grid will define topology of the output layers. Do not modify header file.

segments

Categorical raster map in the GeoTIFF format which defines spatial division of area of interest into polygonal regions. In this map, regions should be defined as areas of unique ids. Layer must be a raster and its projection, extents and resolution should be identical to the main input map. This layer may be a result of the segmentation module gpat segment or any other categorical map (e.g. map of ids of watersheds).

inhomogeneity

The name of a raster (GeoTIFF) file containing the values of segment's inhomogeneity.

isolation

The name of a raster (GeoTIFF) file containing the values of segment's isolation.

measure

Defines method for calculating distance between motifels. GeoPAT offers the following measures:

- jsd – Jensen Shannon Divergence
- euc – Euclidean distance
- eucn – Normalized euclidean distance
- eucp – Normalized euclidean distance (periodic)
- wh – Wave-Hedges distance

- cos – Cosine distance
- jac – Jaccard distance
- roz – Rozicka distance
- rozp – Rozicka distance (extended)
- hass – Hassanat distance
- tsEUC – time series - euclidean distance
- tsEUCP – periodic time series - euclidean distance
- tsDTW – time series - Dynamic Time Warping distance
- tsDTWP – time series - Periodic Dynamic Time Warping distance
- tsDTWPa – time series - Synchronized Dynamic Time Warping distance

See Appendix C for the details.

list_measures
Lists all the available measures.

maxhist

complete

quad

no_weight

threads (-t)
The module is parallel, therefore use more than one processing thread can be used in order to speed up calculations. This option specifies how many threads will be used. The default is 1.

Description:

3 Basic workflow paths with examples

In this section, the basic GeoPAT procedures are presented. These are:

- **Search** - search for areas similar to a query
- **Change detection** - comparison of local patterns between two maps
- **Segmentation** - division of a map into regions of cohesive patterns
- **Clustering** - grouping patterns that are similar to each other

The procedures are explained using the workflow schemes and examples. All of the examples how to process categorical data are shown on a 42×69 km (1400×2300 px) part of the National Land Cover Dataset (NLCD) covering area around the city of Augusta, GA (Figure 2). This area is characterized by high diversity of land cover categories and patterns. Thus it is perfect for various pattern analyzes.

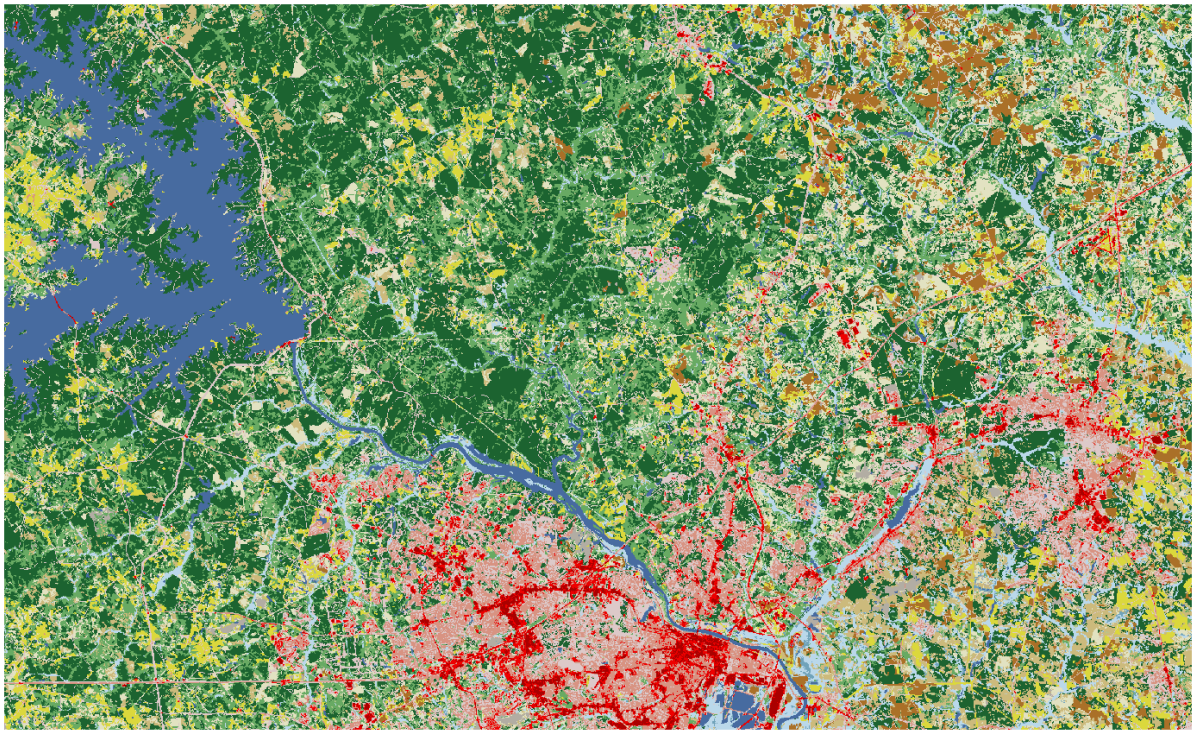

Figure 2: Part of NLCD, covering area around Augusta, GA used in the examples

3.1 Search

Search functionality enables user to produce maps of similarity. These maps show the level of similarity between a specified motifel (query) and a grid of motifels. The input is one or more GeoTIFF raster maps (depending on a data and signature type; see appendix B for

more information), and XY coordinates of one or more points in space. The result is one or more GeoTIFF raster maps which have the same extent as the grid of motifels specified by user. The number of resulting maps is the same as the number of points provided. The workflows for categorical and time series maps differ.

3.1.1 Search on categorical maps

Figure 3 presents general workflow path for producing similarity maps using a categorical raster data.

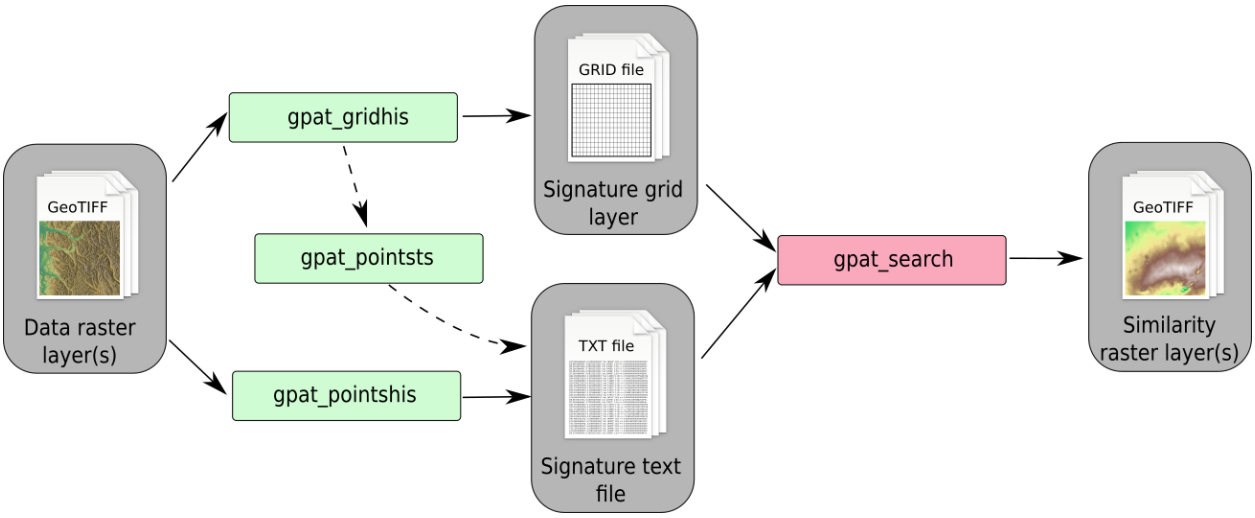

Figure 3: Workflow path for search on categorical maps

The first step is to prepare signature files for both, query motifels and grid of motifels, using two separate modules, *gpat_pointshis* and *gpat_gridhis* respectively. The second step is to use these signature files as inputs to *gpat_search* module in order to the produce similarity maps.

Example:

```
gpat_gridhis -i Augusta2011.tif -o grid -s cooc -z 50 -f 50 -n pdf
gpat_pointshis -i Augusta2011.tif -o query_signatures.txt -s cooc -z 50 -n pdf
    --xy_file=coordinates.txt
gpat_search -i grid -r query_signatures.txt
```

do not have to have the same size (but is recommended)

3.1.2 Search on time series

TODO

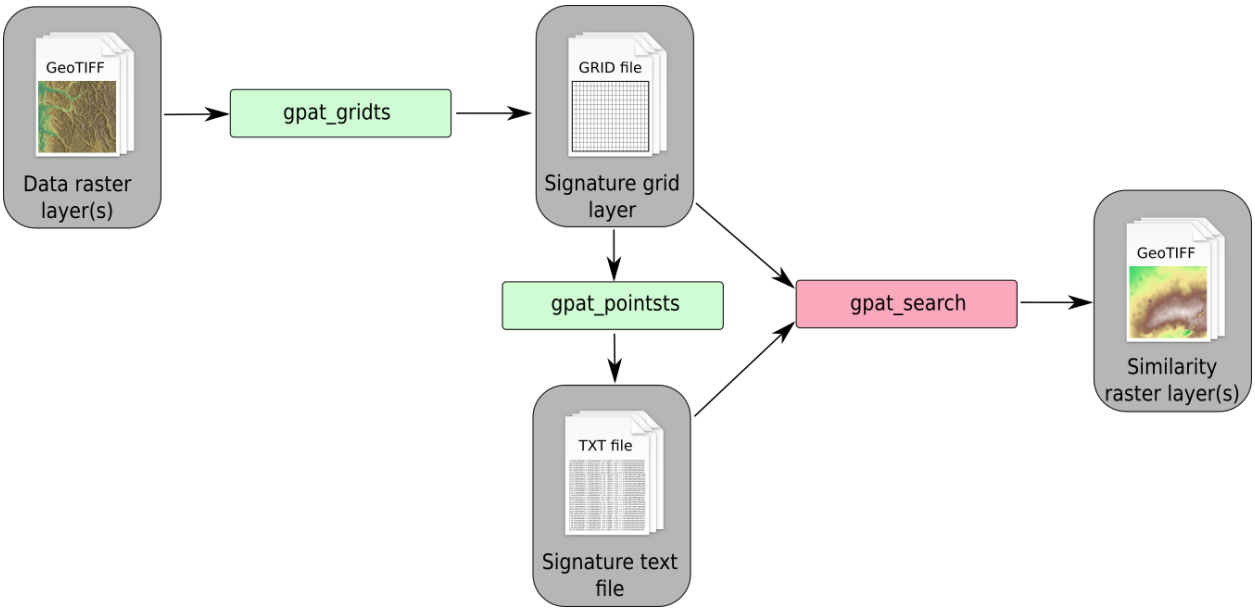

Figure 4: Workflow path for search on time series.

```
gpat_pointsts -i grid -o query_signatures.txt --xy_file=coordinates.txt
```

3.2 Change detection

TODO

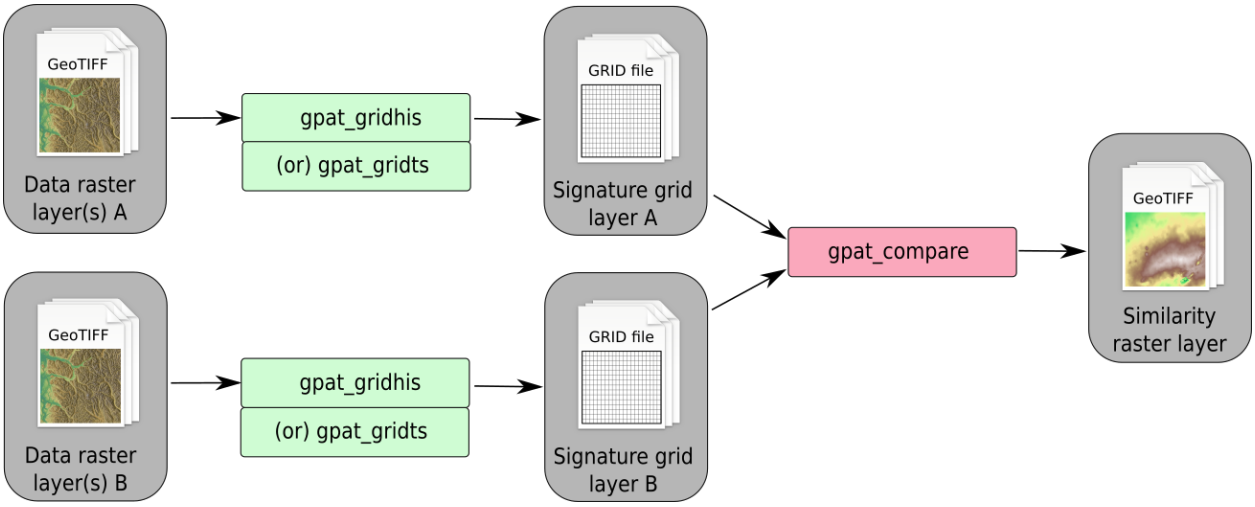

Figure 5: Workflow path for change detection.

3.3 Segmentation

TODO

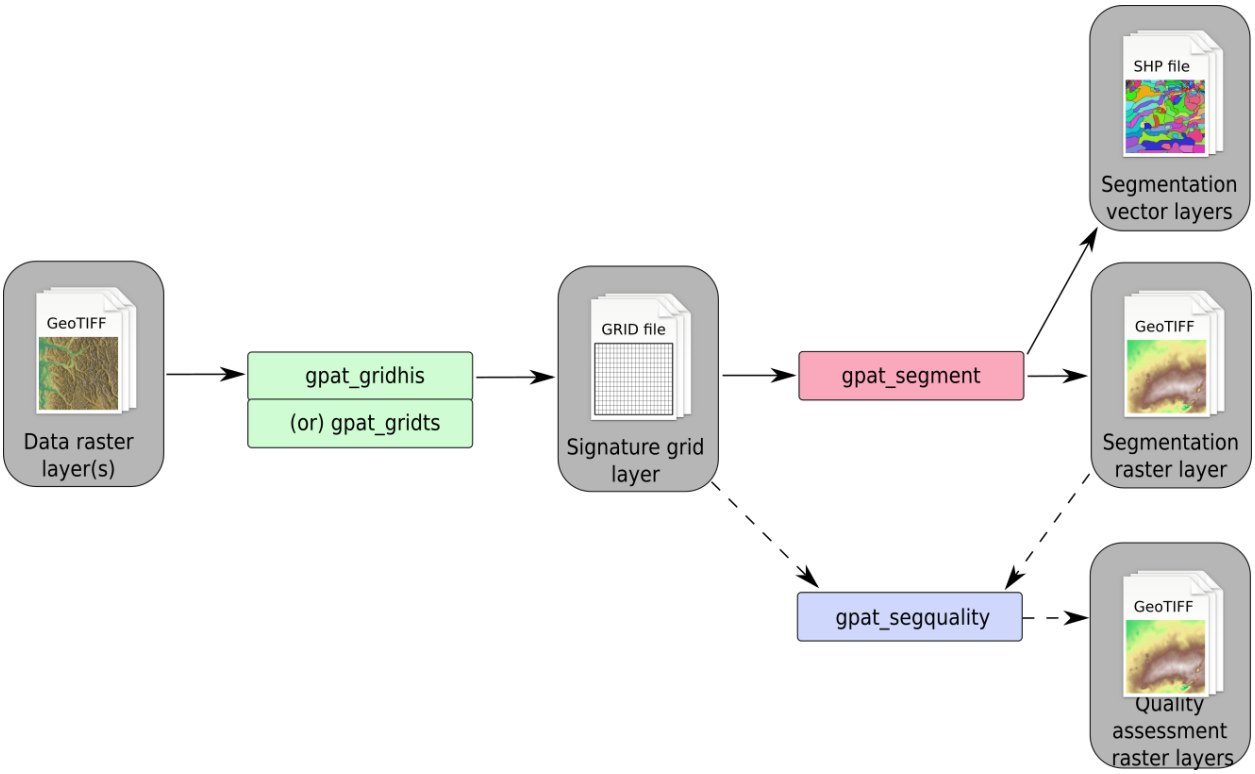

Figure 6: Workflow path for segmentation.

3.4 Clustering

TODO

3.4.1 Clustering of individual motifels

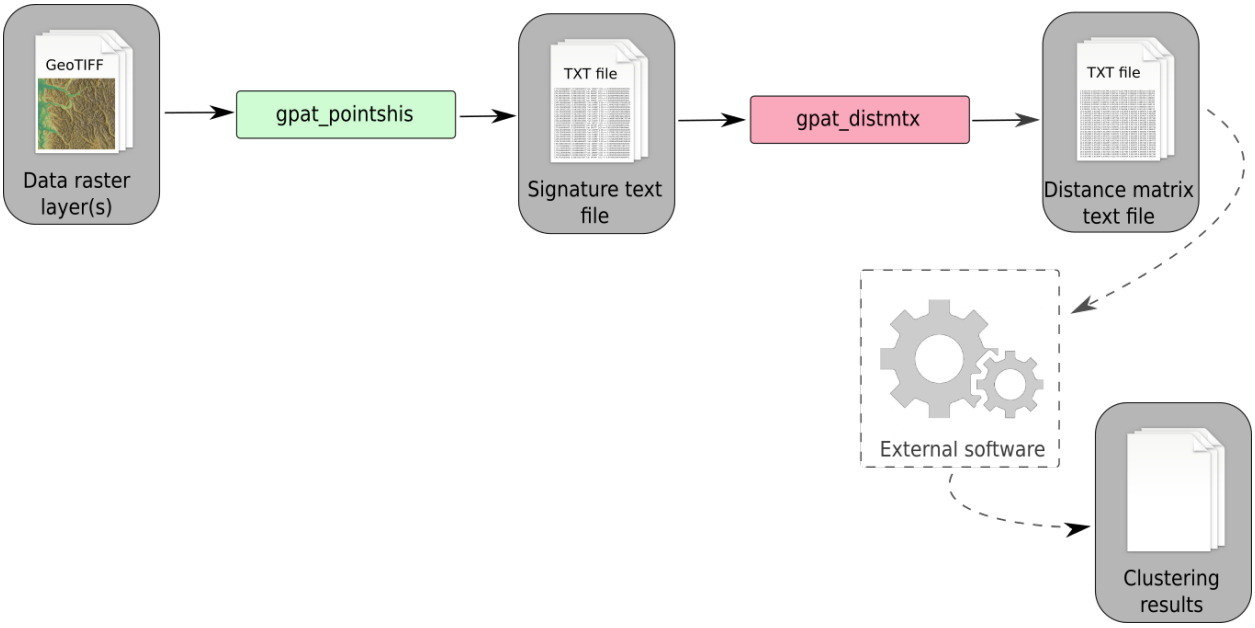

Figure 7: Workflow path for clustering of motifels.

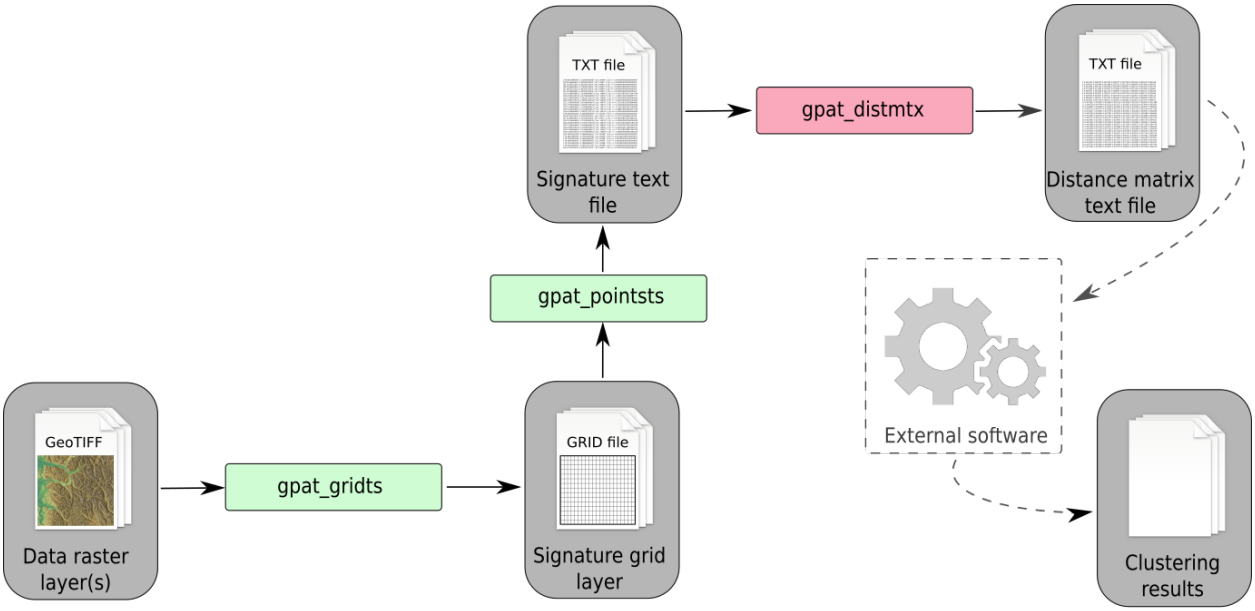

Figure 8: Workflow path for clustering of time series motifels.

3.4.2 Clustering of grid of motifels

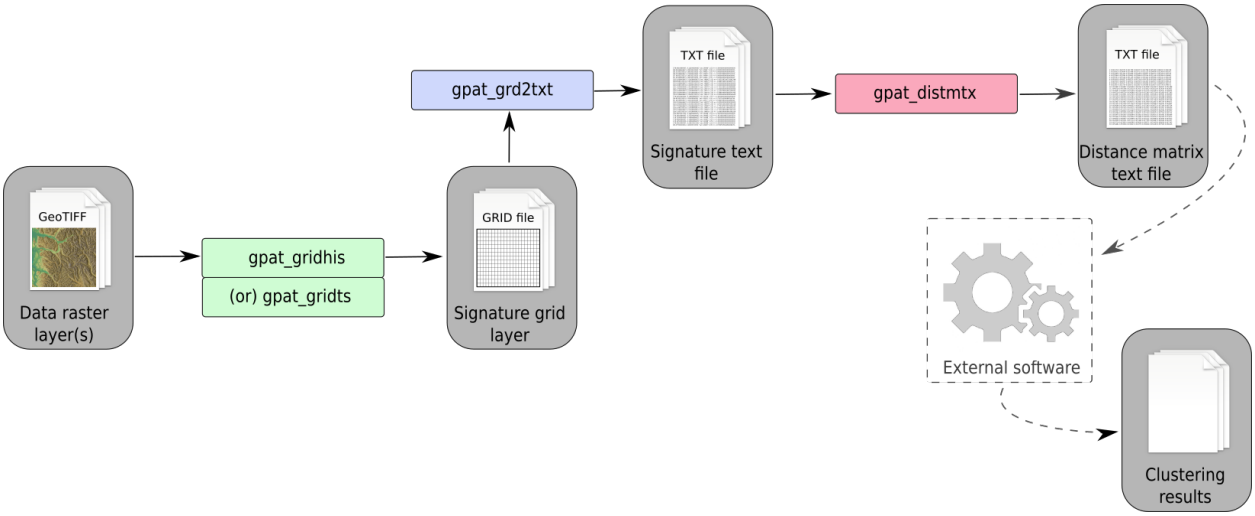

Figure 9: Workflow path for clustering of grid of motifels.

3.4.3 Clustering of segments/predefined irregular regions

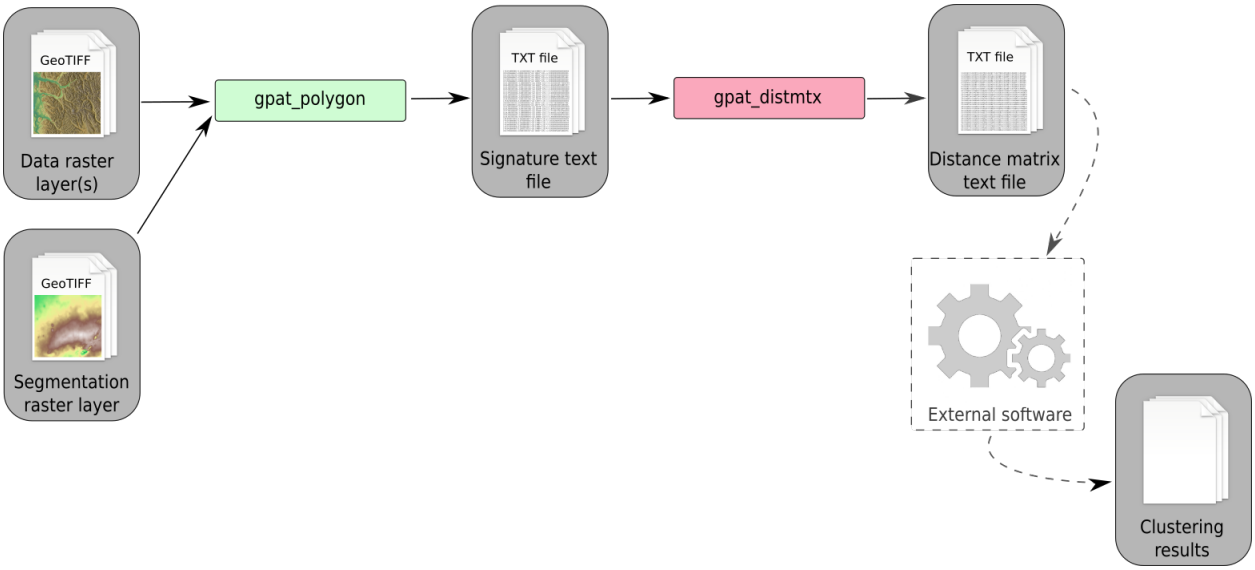

Figure 10: Workflow path for clustering of segments (regions).

Appendices

A GeoPAT 2.0 installation

A.1 System requirements

The Windows installer and Linux binaries provides all necessary components. To build GeoPAT 2.0 from the source code, the user needs to install the GDAL library and compile and install the ezGDAL and SML libraries.

A.2 Windows installer

The Windows installer works under 64 bit versions of Windows 7, 8.1, and 10. The installer provides four components:

- GDAL library package
- ezGDAL and SML libraries
- GeoPAT 2.0 software
- Microsoft Visual C 13 runtime libraries (optional)

To start the installation process user has to run GPAT20setup.exe. The setup program should be run in "Run as administrator" mode. If an antivirus software is running on the computer, user should turn it off temporary for the time of GeoPAT installation. The installer will create the directory for GDAL and GeoPAT and copy all necessary files. Optionally, GPAT20setup.exe will start the Microsoft Visual C runtime libraries installer. When installation is completed, the user can find "SIL GeoPAT 2.0" in the Windows start menu with two files - "GeoPAT console" and "Uninstall GeoPAT".

The installer for Windows x64 is available at:

```
http://sil.uc.edu/cms/data/uploads/software_data/GPAT20setup.exe
```

A.3 Fedora 25 binary installation

To install the binary version of GeoPAT 2.0, user has to copy contents of the gpat20 directory from gpat20.tar.gz file to the /usr/local directory. The GeoPAT 2.0 binaries require the GDAL package to be installed on the computer. The user can install this package using dnf package manager on a Fedora system.

```
dnf install gdal
```

The additional requirement is a proper configuration of the libraries paths. The Fedora system should look for libraries in /usr/local/lib. Therefore, it could be necessary to create the local.conf file containing the following text: "/usr/local/lib". The new file has to be placed in the /etc/ld.so.conf.d directory.

The Fedora 25 x64 binaries are available at:

```
http://sil.uc.edu/cms/data/uploads/software_data/gpat20.tar.gz
```

A.4 Building from source code

The building from source code procedure is presented using the Fedora 25 Linux distribution. This procedure could diffet between Linux distributions, therefore the user should modify it to their system.

To build GeoPAT 2.0 from the source, the user has to do the four following steps:

1. Install the developement version of GDAL
2. Build and install the ezGDAL library
3. Build and install the SML library
4. Build and install the GeoPAT 2.0 software

The dnf package manager can be used in a Fedora system to install the developement version of GDAL:

```
dnf install gdal-devel
```

To install the ezGDAL library the user has to download the ezGDAL source code and unpack it. The source code of ezGDAL is available at:

```
http://pawel.netzel.pl/data/uploads/software/libezgdal.src.tar.gz
```

Next, the user has to compile the code by calling the following command in an unpacked source code directory:

```
make
```

And install it using:

```
make install
```

By default, the library is placed in the /usr/local/lib directory and the include file is placed in /usr/local/include. The user can change the destination directory by adding the PREFIX parameter.

```
make PREFIX=/my/destination/directory
```

When PREFIX is provided, the library is placed in /my/destination/directory/lib and the include file is placed in /my/destination/directory/include.

The installation procedure of the SML library is similar. The source code of SML is available at:

```
http://pawel.netzel.pl/data/uploads/software/libsml.src.tar.gz
```

After an extraction of the source code of SML, the user should call:

```
make
make install
```

The command "make install" should be called using the sudo command or in the root user context. After finishing libraries installation procedure the user has to ensure that PREFIX/lib is on the library search path.

The last step of the installation procedure is a compilation of the GeoPAT 2.0 source code. The source code of GeoPAT 2.0 is available at:

```
http://sil.uc.edu/cms/data/uploads/software_data/gpat2.0src.tar.gz
```

GeoPAT 2.0 depends on the GDAL, SML, and ezGDAL libraries. Therefore, after the installation of the above libraries, the user has to unpack, compile and install the GeoPAT 2.0. The installation procedure is similar to the installation procedures of the ezGDAL and SML libraries.

```
make
make install
```

The "make" and "make install" commands should be run in the main GeoPAT 2.0 source code directory. PREFIX parameter works in the same way.

B Numerical signatures and normalization methods available in GeoPAT

A signature is a numerical description of a motifel

B.1 Cartesian product

B.2 Class co-occurrence histogram

B.3 Decomposition histogram

B.4 Local binary pattern histogram

B.5 Landscape indices vector

C Dissimilarity measures available in GeoPAT

C.1 Jensen Shannon Divergence

C.2 Euclidean distance

C.3 Normalized euclidean distance

C.4 Normalized euclidean distance (periodic)

C.5 Wave-Hedges distance

C.6 Cosine distance

C.7 Jaccard distance

C.8 Rozicka distance

C.9 Rozicka distance (extended)

C.10 Euclidean distance - time series

C.11 Dynamic Time Warping distance - time series

C.12 Periodic Dynamic Time Warping distance - time series

C.13 Synchoronized Dynamic Time Warping distance - time series