Description:

OpenCLC is an open source program that estimates the similarity of two sets of linear (vector-based) features that represent the same phenomenon. The program was primarily written to estimate similarity of two sets of linear features that represent the same set of surface-water flow networks. The program computes the Coefficient of Linear Correspondence (CLC) for the tested datasets by estimating the matching and mismatching linear features in each dataset. Subsequently, the CLC is computed as the sum of the length of the matching lines from both datasets divided by the sum of the length of all lines from both datasets. Matching and mismatching lines are estimated by classifying likely matching and mismatching pixels of a line-density difference raster dataset determined for the input sets of linear features. Additional details of the process are presented by Stanislawski, Buttenfield, Doumbouya (2015). In addition, the program summarizes the spatial distribution of matching and mismatching features by computing a CLC value for each cell within a grid of about 200 square polygonal cells covering the study area.

References:

Stanislawski, L.V., Buttenfield, B.P, and Doumbouya, A. (2015) A rapid arpporach for automated comparison of independently derived stream networks, Cartography and Geographic Information Science, 42:5, 435-448, DOI: 10.1080/15230406.2015.106086

Stanislawski, L.V., Li, T., and Shavers, E. (2018) Similarity assessment of linear hydrographic features using high performance computing. FOSS4G North America. May14-16, 2018, St. Louis, Missouri.

Setup:

The OpenCLC program has been implemented through Python 2.7. Programs have only been tested on the following Linux environment:

Linux a101 2.6.32-504.8.1.el6.x86\_64 #1 SMP Wed Jan 28 21:11:36 UTC 2015 x86\_64 x86\_64 x86\_64 GNU/Linux

Environment list includes:

Currently Loaded Modulefiles:

1) binutils/2.25

2) GCC/4.8.4

3) numactl/2.0.10-GCC-4.8.4

4) hwloc/1.10.1-GCC-4.8.4

5) OpenMPI/1.8.4-GCC-4.8.4

6) OpenBLAS/0.2.13-GCC-4.8.4-LAPACK-3.5.0

7) gompi/1.7.20

8) FFTW/3.3.4-gompi-1.7.20

9) ScaLAPACK/2.0.2-gompi-1.7.20-OpenBLAS-0.2.13-LAPACK-3.5.0

10) goolf/1.7.20

11) filegdb/1.4

12) NASM/2.11.06-goolf-1.7.20

13) libjpeg-turbo/1.4.0-goolf-1.7.20

14) PROJ/4.8.0-goolf-1.7.20

15) GEOS/3.4.2-goolf-1.7.20

16) Szip/2.1-goolf-1.7.20

17) HDF5/1.8.10-patch1-goolf-1.7.20

18) netCDF/4.2.1.1-goolf-1.7.20

19) flex/2.5.35-goolf-1.7.20

20) Bison/2.5-goolf-1.7.20

21) JasPer/1.900.1-goolf-1.7.20

22) HDF/4.2.7-patch1-goolf-1.7.20

23) zlib/1.2.8-goolf-1.7.20

24) ncurses/5.9-goolf-1.7.20

25) libreadline/6.3-goolf-1.7.20

26) Tcl/8.6.3-goolf-1.7.20

27) SQLite/3.8.11.1-goolf-1.7.20

28) spatialite/4.3.0a

29) gdal/2.0.1

30) gdal2-stack

31) bzip2/1.0.6-goolf-1.7.20

32) Tk/8.6.3-goolf-1.7.20-no-X11

33) Python/2.7.9-goolf-1.7.20

34) LineDensity2

The line density program is written in C and has been compiled for the installed Linux version mentioned above. The following python modules are imported:

import os, string, time

from osgeo import gdal

from osgeo import ogr

import glob

import sys

from osgeo import osr

import numpy

import math

import scipy

from shapely.geometry import shape, mapping

from fishnet import fishnet

from skimage.measure import label, regionprops

import fiona

import subprocess

Implementation:

The program is currently setup to test the similarity of National Hydrography Dataset (NHD) flowlines for the Forked Creek (HUC 10) watershed with drainage lines extracted from 3m resolution elevation data. These test datasets are under the ForkedCreek\_Illinois/streams/ folder. The 3-m resolution elevation model is also included under the ForkedCreek\_Illinois/data/ folder for reference. The test\_CLC\_ForkedCreek\_rect\_3m.sh shell is set up to run the OpenCLC program on the Forked Creek data. Once the environment is setup the shell script can be run from a terminal as follows:

bash tests\_CLC\_ForkedCreek\_rect\_3m.sh ForkedCreek\_Illinois

Aside from the dataset folder, the arguments for the CLC program are hardcoded into the shell script. Arguments include line shapefile 1, line shapefile 2, water polygon shapefile, boundary polygon shapefile, cell size for line-density raster dataset, and radius for line-density computations.

Resulting Data:

Files generated by the program are written to a “results” folder under the dataset folder. Each run on a dataset folder will write over the results folder, so it should be renamed to maintain multiple results from multiple runs. An example of the results for the Forked Creek data are under the results\_3m folder. Results are summarized in a text file clc\_lvs\_results\_”name of lineset file one”\_”name of lineset file two”.txt