PyGEOS Documentation

Release 1.0

Eric Daniels

September 07, 2016

CONTENTS

PyGEOS is a Python Package designed to support common processes in the geosciences. These tasks include interfacing with open source Python packages, GSLIB, Paraview, and Geosoft. The goal is to simplify common tasks that require batch processing or functionality not available in commercial software.

Dependencies: PyGEOS has been developed to run on Python 3.4 and includes a number of functions from pre-existing packages. To avoid issue be sure to install Pandas, NumPy, PyGSLIB, and Matplotlib

Contents:

CHAPTER

ONE

PA MODULE

pygeos.PA.**zscore**(*inputDF*)

Return a DataFrame of Zscore values

Parameters **inputDF** – Pandas Dataframe of values to be returned as zscore

Returns:

CHAPTER

TWO

IO MODULE

module for import/export functions

NOTE: many functions require the Geosoft GX plug-in Currently commented out b/c I am no longer working with geosoft regularly.

pygeos.io.**read_GSLIB**(*FilePath*, *griddef*)

Read GSLIB gridded file with grid definition in header.

Parameters

- **FilePath** – Location of .dat GSLIB file
- **griddef** – pyGSLIB gs.GridDef object

Returns

tuple containing:

griddef: pyGSLIB griddef object with grid definition read from header
data: Pandas DataFrame with Grid(s)

Return type (tuple)

pygeos.io.**read_LAS**(*FilePath*)

Read LAS files to pandas Dataframe and dictionary of tops

Parameters **Filepath** – location of LAS file

Returns

tuple containing:

TopsDict: dictionary containing tops
WellData: Pandas DataFrame with log data

NOTE: Not all LAS files are the same, this function was designed using Petra exported files. LAS files from other software may yield and error.

Return type (tuple)

pygeos.io.**read_g2d**(*FilePath*)

Read Pangeos g2d file to PyGSLIB GridDef object

Parameters **Filepath** – location of .g2d file

Returns PyGSLIB griddef object

Return type (gs.GridDef)

pygeos.io.**read_gxgrid**(*GridPath*, *name*)
Read geosoft grid to pandas dataframe

Parameters

- **GridPath** – Filepath of Geosoft grid
- **name** – Name for column header in pandas dataframe

Returns Pandas DataFrame

pygeos.io.**read_ply**(*filepath*)
Read polygon from .ply file

Parameters **filepath** – filepath for polygon file

Returns

tuple containing:

ply: pandas DataFrame containing XY coordinates for verticies
patch: matplotlib patch for use in plotting

Return type (tuple)

pygeos.io.**to_gxgrid**(*df*, *name*, *griddef*)
Convert Pandas DataFrame to Geosoft grid

Parameters

- **df** – Pandas DataFrame to be Gridded
- **name** – variable name
- **griddef** – pygslib GridDef object to set grid definition

Returns

Geosoft grid (.grd)

Note: this function does not set the projection

Return type (GXgrd)

CHAPTER

THREE

PLOT MODULE

pygeos.plot.**corr_plot**(*df*, *outfl=None*, *color='dimgray'*)

Generate a matrix plot for examining pairwise relationships between datatypes

Parameters

- **df** – Pandas DataFrame with data
- **outfl** – file path for saved plot (optional)
- **color** – color for points on scatter plots

pygeos.plot.**fancy_xplot**(*GridData*, *WellData*, *var1*, *var2*, *outfl*, *show=True*, *figsize=(15, 15)*)

Cross plotting function used to portray gridded data and extracted scattered data on the same plot Useful to show relationship between grids and data actually used in training for PA methods

Parameters

- **GridData** – Pandas DataFrame of gridded data
- **WellData** – Scattered well data extracted from grid locations (same variable names as GridData)
- **var2** (*var1,*) – variables for crossplotting
- **outfl** – filepath for saving figure
- **show** – True/False to show figure
- **figsize** – set figure size (default: 15x15)

pygeos.plot.**log_plot**(*Data*, *columns*, *topsdict*, *top*, *bottom*, *cmap=<matplotlib.colors.LinearSegmentedColormap object at 0x0000000006BD60F0>*, *cutoff=None*, *outfl=None*, *show=True*, *returnfig=False*)

Plot well log information from LAS file

Parameters

- **Data** – Pandas DataFrame with log data
- **columns** – columns within DataFrame to plot
- **topsdict** – a dictionary of tops from which top and bottom will be selected
- **top** – top formation for plot
- **bottom** – bottom formation for plot
- **cmap** – colormap (default is jet)
- **cutoff** – cutoff value (optional)
- **outfl** – filepath for saved plot (optional)

- **show** – True/False to show plot (default: True)
- **returnfig** – True/False to return figure as object, useful for color mapping / legend

pygeos.plot.**probplt**(*dataSeries*, *outfl=None*, *color='black'*)

CHAPTER

FOUR

UTILS MODULE

pygeos.utils.**gap**(*data*, *refs=None*, *nrefs=20*, *ks=range(1, 11)*)

Compute the Gap statistic for an nxm dataset in data. Either give a precomputed set of reference distributions in refs as an (n,m,k) scipy array, or state the number k of reference distributions in nrefs for automatic generation with a uniformed distribution within the bounding box of data. Give the list of k-values for which you want to compute the statistic in ks.

pygeos.utils.**getcoords**(*GridDef*)

Generate Pandas DataFrame of XYZ coordinates based on grid definition

Parameters **GridDef** – pyGSLIB Grid Definition object

Returns Pandas DataFrame with the columns X, Y and Z

Return type coords

pygeos.utils.**getidx**(*points*, *griddef*)

Return the index of a given point in a grid Best used with df.apply() function in pandas

Parameters

- **points** – Pandas DataFrame including columns labeled X and Y
- **griddef** – pyGSLIB Grid Definition object

Returns index value for location in grid

Return type (int)

pygeos.utils.**gridExtract**(*ptxy*, *griddef*, *gridDF*)

Uses getidx to extract gridded data based on point locations

Parameters

- **ptxy** – Pandas DataFrame containing X Y columns of coordinates
- **griddef** – pyGSLIB Grid Definition object
- **gridDF** – Pandas Dataframe containing grid(s)

Returns Values extracted from grid(s) at XY locations

Return type (Pandas DataFrame)

CHAPTER

FIVE

INDICES AND TABLES

- genindex
- modindex
- search

p