Calculating slope gradient variations in the submarine landforms by R and Python statistical libraries

Polina Lemenkova

Ocean University of China, College of Marine Geo-sciences. Address: 238 Songling Road, Laoshan, 266100, Qingdao, Shandong Province, People's Republic of China, pauline.lemenkova@gmail.com, ORCID 0000-0002-5759-1089

A B S T R A C T

This research focuses on the analysis of the submarine geomorphology in the Mariana Trench located in west Pacific Ocean. The research question is to identify variations in the geomorphic form and bathymetry in different segments of the trench. Technically, the paper applies Python and R programming statistical libraries for geospatial modelling of the data sets. The methodological approach of the statistical data analysis by scripting libraries aimed to visualize geomorphic variations in the 25 transect profiles of the trench. Multiple factors affect submarine geomorphology causing variations in the gradient slope: geological settings (rock composition, structure, permeability, erodibility of the materials), submarine erosion, gravity flows of water streams, tectonics, sediments from the volcanic arcs, transported by transverse submarine canyons. Understanding changes in geomorphic variations is important for the correct geospatial analysis. However, modelling such a complex structure as hadal trench requires numerical computation and advanced statistical analysis of the data set. Such methods are proposed by R and Python programming languages. Current research presented usage of statistical libraries for the data processing: Matplotlib, NumPy, SciPy, Pandas, Seaborn, StatsModels by Python. The research workflow includes following steps: Partial least squares regression analysis; Ordinary Least Square (OLS); Violin plots and Bar plots for analysis of ranges of the bathymetric data; Isotonic Regression by StatsModels library; Data distribution analysis by Bokeh and Matplotlib libraries; Circular bar plots for sorting data by R; Euler-Venn diagrams for visualizing overlapping of attributes and factors by Python. As a result of the data analysis, the geomorphology of the trench slopes in 25 transecting profiles was modelled. The results achieved by the statistical data modelling show differences in the gradient slope in various segments of the trench depending on its spatial location. This shows complex geological structure of the trench. The paper contributes towards the methodological development of the data analysis in marine geology through the stepwise workflow explanations with a case study of Python and R applications

A R T I C L E I N F O

1. Introduction

The geomorphology of the ocean trenches and continental-shelf submarine canyons plays crucial role for the deep-sea ecosystems, because, along with other environmental factors, steep gradient slopes create conditions for the downward transport from the flat abyssal plains and accumulation of the organic matter and sediments along the hadal trench axis [1, 2, 3]. Therefore, understanding geomorphic slope processes is of particular interest to the geoscientists, since slopes form conditions for the ecosystem life [4] and reflect changes in lithology [5]. However, the geomorphology of the submarine slopes of the ocean trenches is often overlooked, due to the unreachable nature of the study object that can only be studied by means of the remote sensing and computer based modelling.

Existing studies on the landform geomorphology are mostly focused on the landforms structure and variations, land surface processes [6], classification of the geomorphic classes as land forms existing on the Earth in general [7, 8, 9, 10]. Nevertheless, the submarine landforms as a particular class of the geomorphic shapes on the ocean seafloor have lesser attention comparing to the land geomorphology. Some existing works report recent findings on the on hadal trench geomorphology used in the current work [11, 12, 13, 14, 15, 16, 17, 18]. The particular focus of this study lies on the calculations of the slope gradients and classification of the

hadal trench's slope steepness by multi-disciplinary approach of GIS and data analysis using programming libraries. Estimation of slopes and ranking data were performed by R language with auxiliary statistical analysis by Python libraries.

2. Methods and materials

The methodology of this work is illustrated by the workflow chart (Fig. 1). The core methods used in this work are based on the machine learning algorithms provided by Python and R programming languages and statistical libraries.
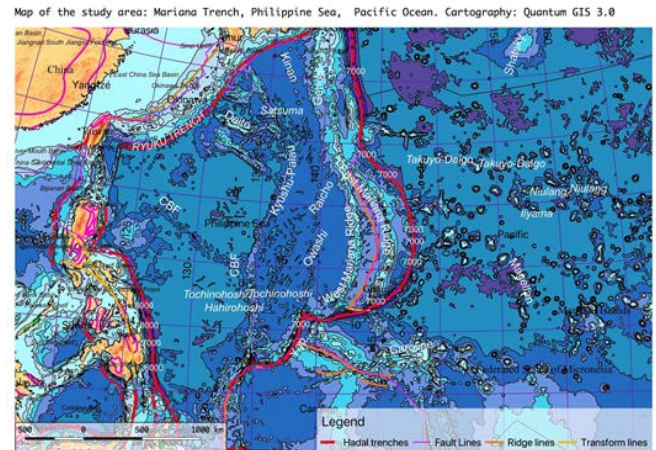

Figure 1. *The methodological workflow in 10 steps.*

The research started from the installing Python and necessary libraries. Following steps are described in the relevant subsections below.

2.1. Spatial data processing using GIS

The first step included GIS based analysis performed in QGIS where geospatial layers were stored, re-projected and visualized. These include vector .shp files of thematic data: geology, bathymetry, tectonics, sediment thickness covering study area of Mariana Trench, Pacific Ocean (Fig. 2). The next step included plotting profiles crossing the study area (Fig. 3) plotted by 'ProfileFromPoints' QGIS plugin. The data pool now consisted in 25 cross-section profiles within a 1.000km length each and a distance of 100 between each couple in the area which extend from Philippine Sea crossing Mariana Trench (Fig. 2) in a perpendicular direction. The visualization of the profiles and calculations of the deepest gradient slope angle by profiles are shown on the facetted plot of Fig. 3 for each profile, respectively. The attributes values of the the digitized layers were stored in a table (.csv) and exported to R and Python libraries for further processing.
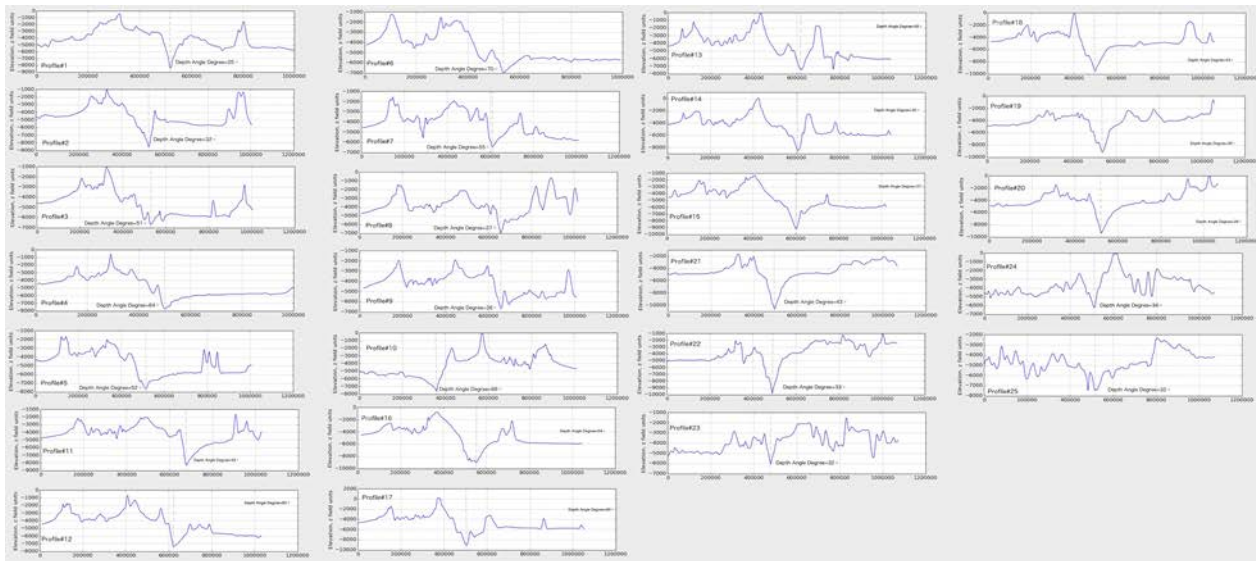

Figure 2. *Study area: Mariana Trench. Geomorphic and geological features.*

2.2. Installing tools and preparing data

The following step consists in the statistical data processing by libraries StatsModels, Matplotlib, Pandas, using existing methodologies on Python data processing [19, 20, 21, 22]. The general syntax of Python and solutions of the statistical algorithms by libraries were derived from the existing literature on data science [23, 24, 25, 26, 27, 28] and modified for the particular case study. The research is technically based on versions of Python 3.7.2 and R 3.6.1. The installation of the Python libraries and dependancies was done via the command line of the bash terminal. Because this process is not trivial, a brief explanation of the crucial installation steps is provided below. Upon installing core package of Python, the settings tools were adjusted. The package installer for Python 'pip' 19.0.3 and Xcode were installed. These steps are illustrated in the following command: $ curl https://bootstrap.pypa.io/get-pip.py -o get-pip.py $ python get-pip.py $ xcode-select –install

Then, important Python libraries were installed via the console utility 'pip'. First, NumPy, a Python's package for the scientific computing providing fast N-dimensional array processing. Second, a SciPy package version 1.1.0, depending on NumPy 1.14.5, was installed. SciPy was used for the computations. Third, Pandas 1.11.0, the package for data manipulation in tabular format was installed. Plotting packages for data visualization, Matpotlib and Seaborn, were installed via pip using commands lines, stepwise: $ pip install scipy $ pip install numpy $ python3 -m pip install pandas $ sudo python3 -m pip install matplotlib $ python3 -m pip install seaborn

Figure 3. *25 cross-sectioning profiles transecting Marian*

a Trench in perpendicular direction.

The advantages of the Matplotlib library of Python for geographic studies consists in its versatility and capability of creating a variety of plots. Seaborn package is build atop the Matplotlib. It combines the power of the Matplotlib with more intuitive approach for data visualization.

After all necessary libraries were installed, the statistical analysis was performed using previously made geographic dataset on the Mariana Trench. Importing data using table with variables was done using Pandas library, installed in the previous step. The data were stored and transformed in the .csv file format. The table was received from the QGIS project containing vector layers and attribute tables. The Pandas ability to read-in, manipulate and write data to and from csv file is a key feature enabling effective data analysis. Three data frames are data type for storing tabular 2D data in Pandas. Using read_csv() function the files were read into Pandas DataFrames.

2.3. Using Python Pandas library for logical queries for geospatial analysis in data frame

A Pandas library of Python was used for manipulations with geological data: loading from the csv table, viewing and editing its content, analysing data range and structure, selecting target rows and columns, automatization of the data processing, manipulation and inquiries, examining the descriptive statistics on data. Therefore, the start of this project initialized with the loading an existing geospatial data set into Pandas environment.

First, the csv file was imported into Pandas Data Frame using command: df = pd.read_csv("Tab-Bathy.csv"). The table for the bathymetry had 517 rows and 18 columns. Second, the data were analysed using logical queries by Pandas. Selecting

necessary columns or rows and their attributes in Pandas is relatively straightforward. Thus, using 'iloc' and 'loc' as operations for retrieving data from data frames, necessary columns were selected for modelling by logical selections. For example, X = df.iloc[: , 1:13] was done to select 12 first columns in the Bathymetry table. Alternative syntax would be (for the case of 4 columns) X = df.iloc[: , [1, 2, 3, 4]] for smaller data sets where the number of columns can be defined manually. Similarly, selecting necessary row during the data analysis was done using loc[] function, e.g. B = df.loc[12] returned the bathymetric observation nr. 13 (starting from 0). For instance, selecting the slope classes fir geomorphic analysis across the trench was done using following Python code:

```
very_steep = data['slope_class'] == 10,
extreme = data['slope_class'] == 9,
strong = data['slope_class'] == 8,
```

which returns a boolean expression of True/False for all 25 profiles.

Selecting a particular profile and corresponding geologic columns for this profile was done using following code snippet (example of profile nr. 15):

```
profile_15 = data. loc[data['profile'] == 15, ['sedim_thick', 'slope_angle', 'igneous_volc']]
```

Analysing geological data set was done through logical queries by Pandas. E.g. to find maximal depths, or to find a profile that has the thickest layer of sediments is done by the code lines:

```
# What is the maximal bathymetric depth? data['Min'].min()
    [Out 3]: 10.600 m.
# What is the maximal sediment thickness?
    data['sedim_thick'].max() [Out 3]: 142
# What is the maximal slope angle steepness?
    data['slope_angle'].max()
```

[Out 3]: 70

Other examples of the logical queries:

\# How many bathymetric data['plate_pacif'].value_counts() observations are there for each tectonic plate?

\# How many data belong to each of the slope class? data['slope_class'].value_counts()

\# Summary displaying statistics for the selected variable: Mariana Plate data['sedim_thick'].describe()

This returns a descriptive statistics on the column of sediment thickness (Name: sedim_thick, dtype: float64)

\# Splitting data into groups ba variables was done using groupby():

data.groupby(['slope_class']).groups.keys()

[Out 4]: dict_keys([6, 7, 8, 9, 10]) # that is, 5 slope classes of the geomorphic steepness.

\# Analysing how many profiles have specific slope steepness using groupby(): data.groupby('slope_class', as_index=False)[['profile']].sum() # produces Pandas DataFrame
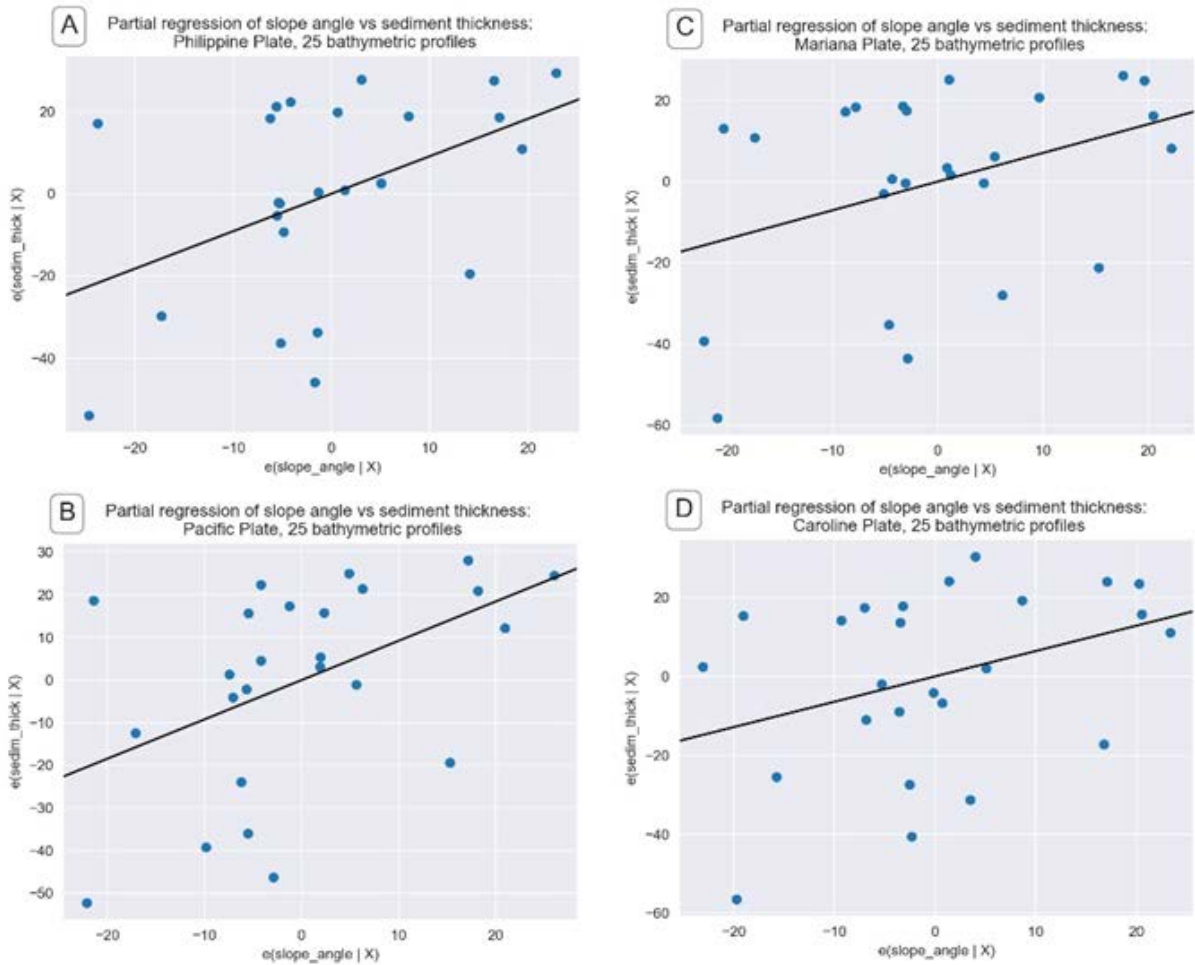
Similar logical queries are possible through Pandas environment. All scripting was done on Jupyter Notebook. The tables were examined to correctly process columns and rows. Thus, a df.shape function was used to know the number of rows and columns; df.ndim – to check up the dimension of the array; df.head() and df.tail() for analysis of the first and last rows in the data frame. The types of each column were checked using '.dtypes' function.

2.4. Partial least squares regression analysis

Partial least squares regression analysis was performed in Python library StatsModels to estimate the relationships between two factors of the data set: slope gradient angle steepness as independent variable and sediment thickness as dependent variable, by four tectonic plates. The results are presented as a facetted plot on Fig. 4. Geological materials deposited during transit downward the slope is presented by a mixture of rock and sediment derived from the abyssal plain [29]. Therefore, as can be seen from the Fig. 4, the slope gradient and sediment thickness have certain correlation of the observed and predicted values. Moreover, the dependences between the shape of the profiles varies due to the effects from the external factors and location of the segment on one of the four tectonic plates: Mariana, Caroline, Pacific and Philippine Sea.

2.5. Ordinary Least Square (OLS) by Python

The Ordinary Least Square (OLS) method of the statistical analysis aims at estimating the unknown parameters in a linear regression model [30]. It produced (Fig. 5) the best possible coefficient estimating that the model satisfies the OLS assumptions for the linear regression.

Figure 4. *Partial regression analysis for the slope gradient of the profiles by tectonic plates.*

The calculation of the OLS minimizes the sum of the squares of the differences between the observed and predicted variables in the dataset on Mariana Trench geomorphology. The predicted values are given by a plot visualizing linear partial regression analysis of a set of explanatory variables, presented as bathymetric profiles and their depths (Fig. 4). The Fig. 5 shows a screenshot of the OLS's execution and results by Python's StatsModels library.
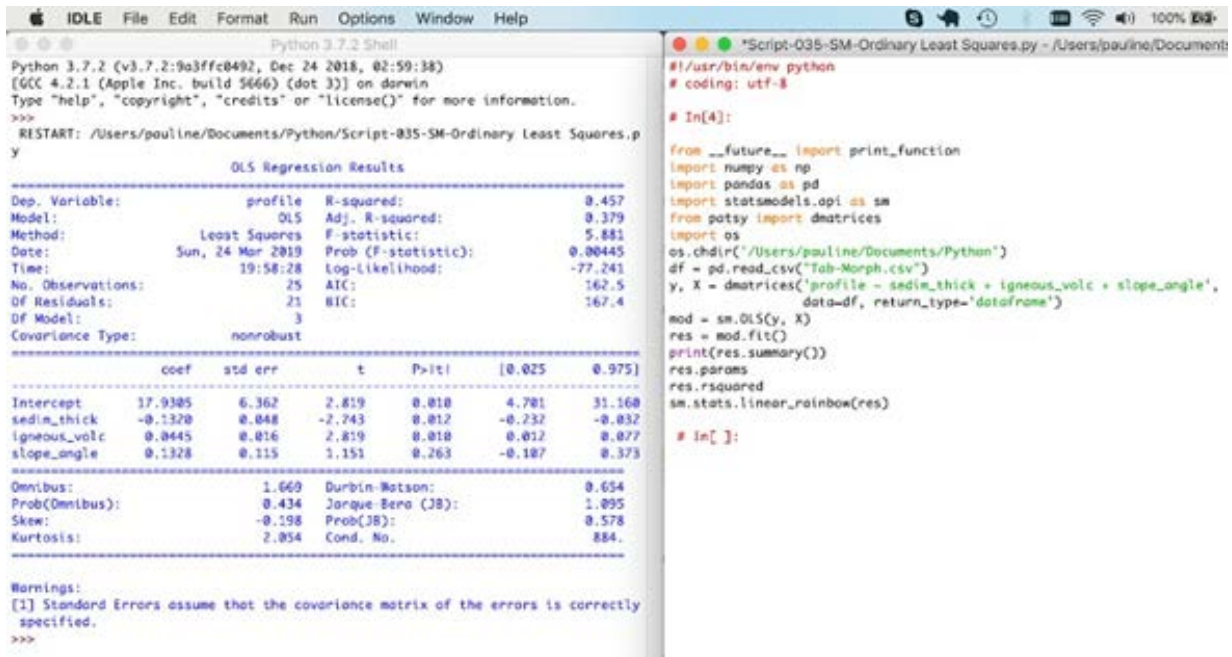
Figure 5. *Ordinary Least Square calculation by Python library StatsModel, for the Mariana Trench project on the geomorphic analysis of the profiles.*

2.6. Violin plots for descriptive statistical analysis of the bathymetry

The 'violin' plots illustrated on Fig. 6 represent statistical analysis of the bathymetric data by profiles. The name 'violin' plots is derived due to the visual similarity of the plot to the music instrument. The violin plot illustrates statistical values for each profile: markers indicating median, interquartile range box plot, rotated smoothed probability Kernel Density Plot (KDE) on each side (curved shape on each profile) of the bathymetric data at different profiles (Fig. 6).
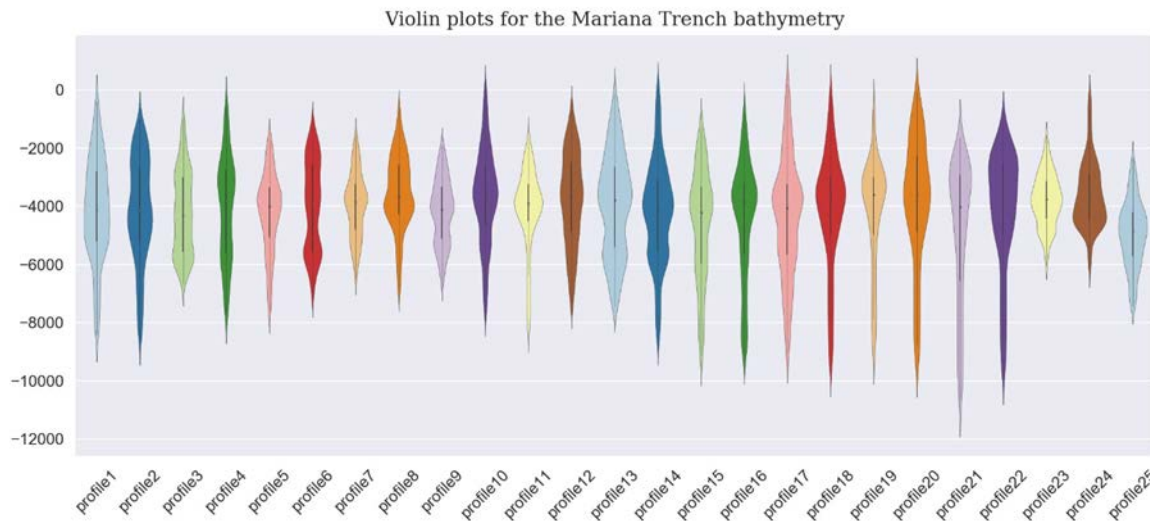

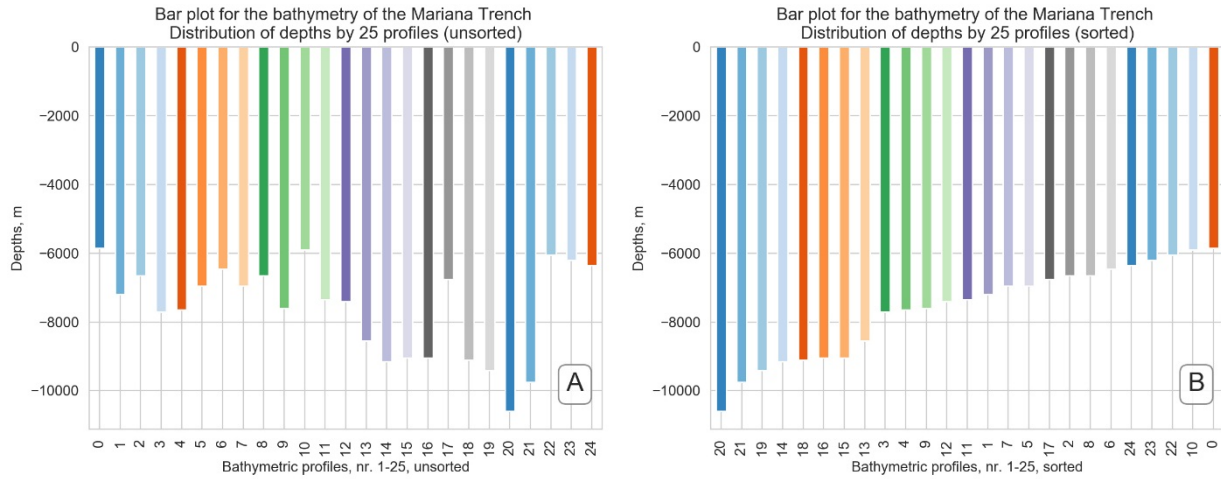**Figure 6**. *Statistics on the bathymetry by violin plots: median, KDE, means of the profiles*

2.7. Bar plots: ranges of the bathymetric depths

The data distribution are presented on Fig. 7, made by Python library Matplotlib. By profiles, they ranged from -10.600m (maximal depth in this data set) to -3150m (the shallowest point in the data set). The maximum depths in the profiles were found in profiles Nr. 20, 21, 19, 14 (Fig. 7, right, sorted bathymetry), and the minimal depths (i.e. the shallowest) consist the first group, that is profiles 24, 23, 22 and 10 (Fig.

7, right, sorted bathymetry). The deepest part of the trench is located in the south-west and central segments where the trench crosses Philippine and Pacific Plates. The moderate depths correlate with Mariana Plate and the majority of the Philippine Plate. The depths ranges detected across the profiles of the present study are also higher in the southern part of the trench than in northern one (Fig. 7). In general, the central part of the Mariana Trench has variations in depth.

However, the general trend shows the increased depths towards the southern part were the depths are higher than in the northern segments. By comparing the profiles Nr. 3, 4, 9, 12 from the Pandas array (Fig. 7, coloured green) with profiles Nr. 18, 16, 15, 13 (coloured orange on Fig. 7), it is clear that depths increase reaching their maximal values in the following group (Profiles Nr. 20, 21, 19, 14, group coloured blue, Fig. 7, right).

Figure 7. *Data distribution analysis: ranges of the maximal depths. Python library: Matplotlib*

2.8. Isotonic Regression by Python's library StatsModels

The statistical method of the isotonic regression (Fig. 8) was applied as a non-metric multidimensional scaling aimed at fitting a line to a sequence of the observations. The conditions are that the line has variation in forms, non-decreasing and non-increasing across the samples and the line lies most close to the observations pool. The concept of the isotonic regression was introduced by [31] and further developed by [32]. The principle of the computation algorithm of the isotonic regression lies in the following formula (1) of the problem of quadratic program (QP):
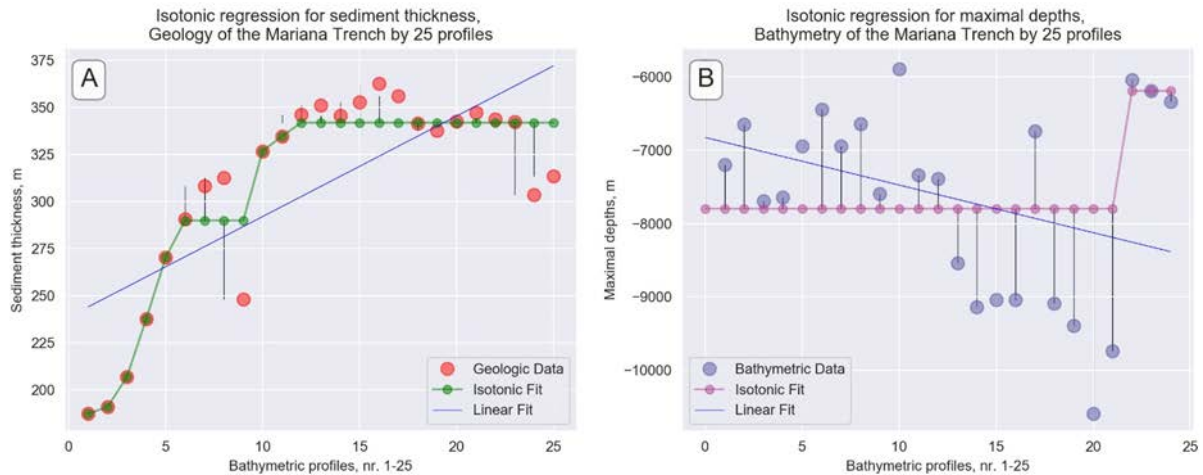
$$min\sum wi*(xi-ai)2 \; subject{:}xi\leq xj \; forall(i,j)\in E \qquad (1)$$

where, $x(i)$ and $x(j)$ are constraints; E are the edges or the is the set of pairs (i, j) for each constraint; w is a weights vector;

n is a number of observations. The isotonic estimator, g minimizes the weighted least squares-like condition (2):

$$ming\in A\sum wi*(g(xi)-f(xi))2 \qquad (2)$$

where, A is the set of all piecewise linear, non-decreasing continuous functions; f is a known function. Further references on isotonic regression and its applications are given by [33]. The presented plot of the isotonic regression (Fig. 8) shows a non-decreasing approximation of the function while minimizing the mean squared error on the bathymetric data set. The benefit of this kind of a model applied for the Mariana Trench analysis consists in the flexibility of the model: the function does not assume any form for the target function, i.e. linearity. Thus, the real data of the observed bathymetric profiles can be compared with the isotonic curves (Fig. 8). Further references are given by [34].
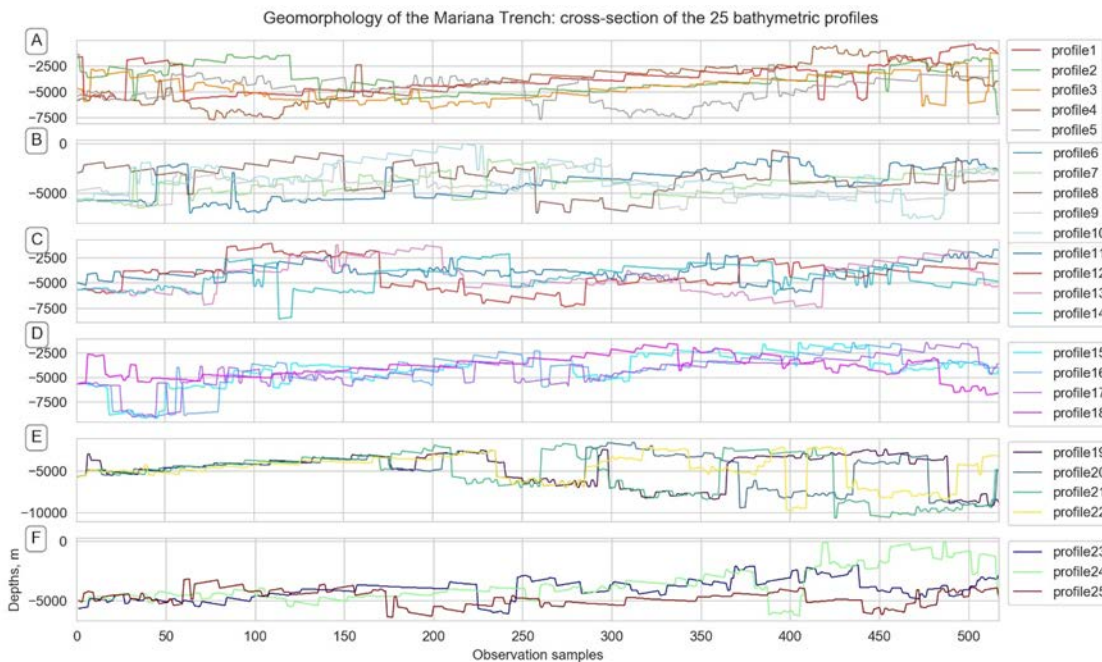
Figure 8. *Isotonic regression for the selected variables: geological data on sediment thickness layer (A) and maximal depths recorded by 25 profiles (B). Python library: StatsModels.*

2.9. Data distribution analysis by Python's libraries Bokeh and Matplotlib

Now we can have a more detailed look on the profiles' shape and outline (Fig. 9). While Fig. 7 shows maximal depths by profiles (i.e. each profile has only one dot point on the plot), Fig. 9 illustrates the cross-section profiles by 518 samples in each profile using a 'stepwise' visualization.
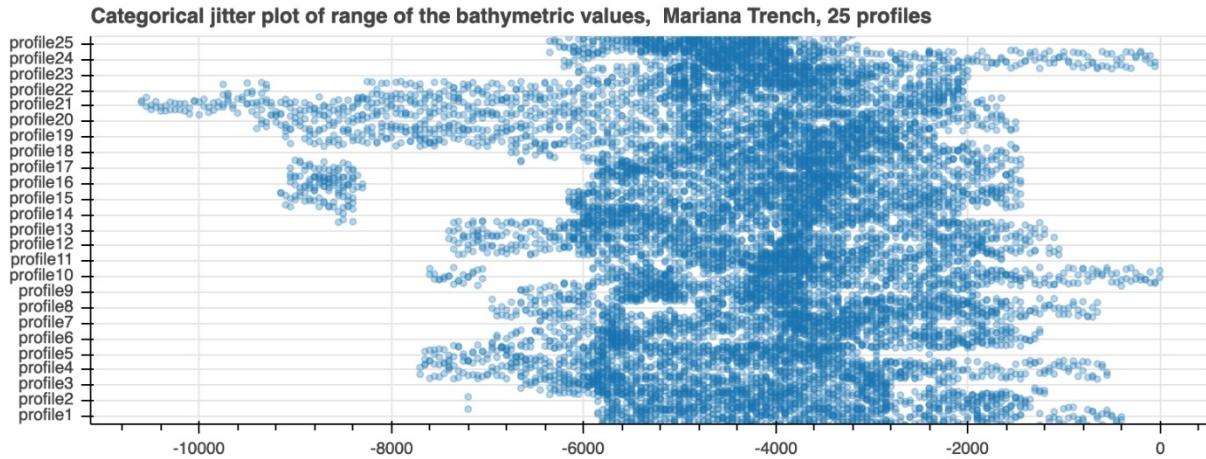

Figure 9. *Visualization of the cross-section profiles, Mariana Trench*

A jitter plot (Fig. 10) plotted using Python library Bokeh is visualizing overlapped samples by profiles. It shows concentration of the most frequent, repeated depths by profiles. Comparing to the Fig. 7 and Fig. 9 focused on the data range, Fig. 10 shows analysis of frequency.

Figure 10. *Data distribution analysis: ranges of the maximal depths. Python library: Bokeh.*
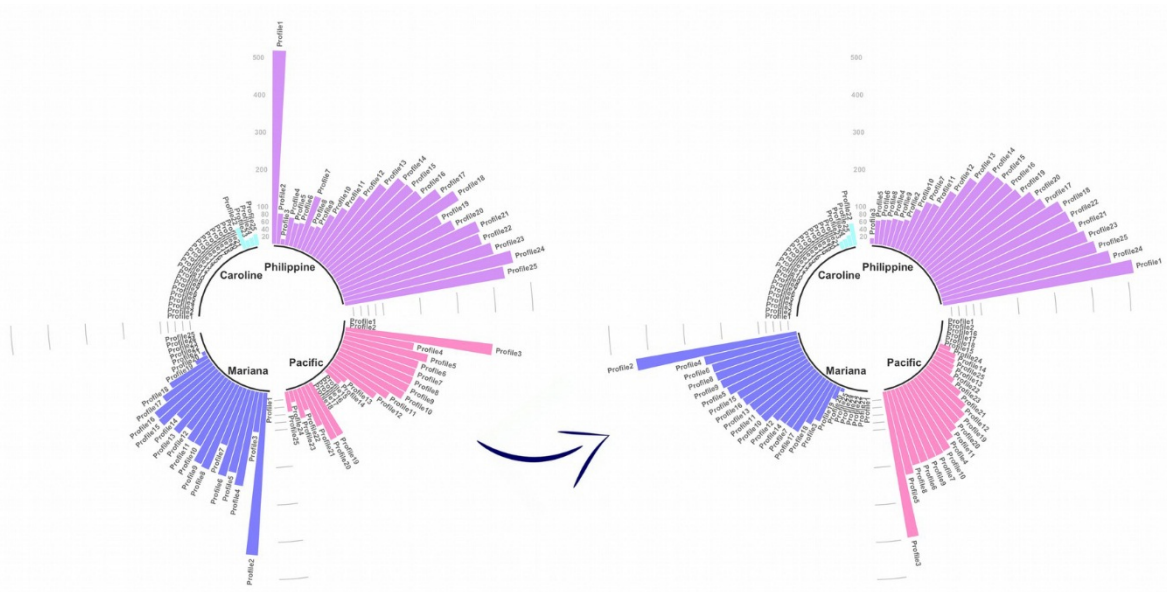
Here we can see (Fig. 10) that the most deep values are notable for the profiles 21 and 22 (upper left part of the graph), where the deepest place on the Earth is located: the Challenger Deep. In general, the graph shows distribution of the bathymetric data by the cross-section profiles.

2.10. Circular bar plots for sorting data

If we want to have more detailed insight into the bathymetry and geologic settings of the geospatial data set, we can extend other functions of spatial analysis embedded in R plotting. Hence, the variations in the sediment thickness by profiles were studied, in addition to the slope degree by four tectonic plates that Mariana Trench crosses: Mariana, Caroline, Pacific, Philippine Sea [35] .To enable multiple comparisons of the variables, existing methodology [36, 37] was used to analyse the distribution of the observation points across these plates (Fig. 11). The least points are located on the Caroline Plate, south-western part of the trench (cyan coloured lines on the Fig. 11). The most dense distribution of the samples is observed on Mariana Plate (purple colour, Fig. 11) and Philippine Sea Plates (magenta colour, Fig. 11). As for Pacific Plate, there is a absence of the samples on the central part of

the trench between the profiles Nr. 15 to 19 where the trench has a curvature (Fig. 2).

Ranking geologic data has a step in geomorphic analysis where the levels of the slope degree are determined by the geospatial impact. Therefore, geomorphic variations typically necessitate an additional sub-task of the geological analysis of the study area. For example, if a certain segment of the trench is located near the active volcanic area the sedimentation rates may increase significantly in this particular region. A {tidyverse} R package, was used to visualize a circular plot showing the comparative analysis of the bathymetric ranged by the sediment thickness. The algorithms and syntax of R coding was applied from the R manuals and available materials. Variations in the slope gradient steepness in the Mariana Trench subduction system are caused by various factors, among which sediment dispersal: longitudinal transport of sediments along the trench and transverse transport into the forearc basins. As a result of these physical processes, Mariana Trench formed as the most significant elongate depressions on the Earth seafloor with a deepest place detected in the Challenger Deep [38].

Figure 11. *Unsorted (left) and sorted (right) distribution of the profiles according to the sediment thickness, circular plot by tidyverse library, R.*

2.11. Euler-Venn diagrams for visualizing overlapping of attributes and factors
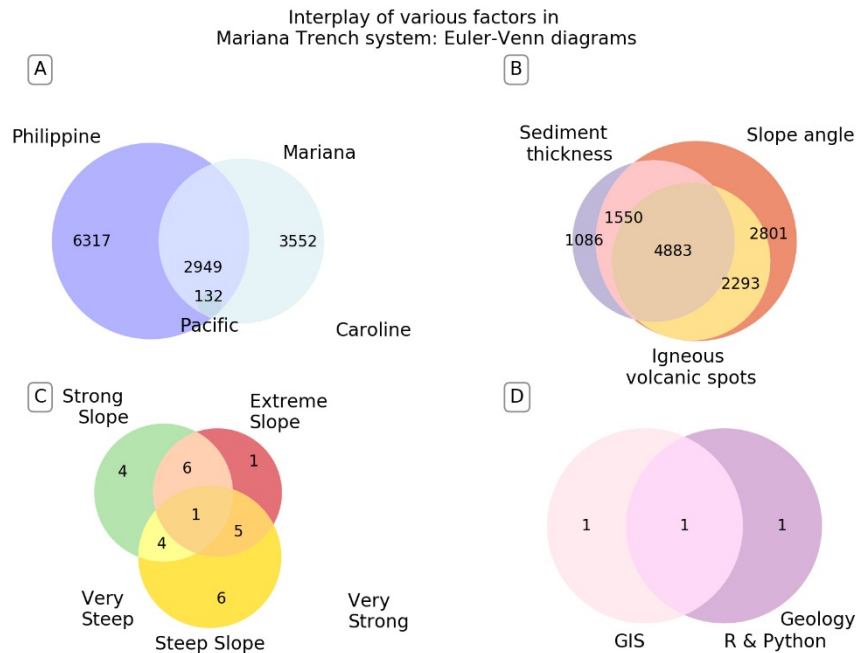
The Euler-Venn diagram shows four cases where overlapping in values or concepts can be visualized. The Fig. 12, subplot A shows cumulative total sediment thickness by tectonic plates. The subplot B shows detected igneous volcanic spots. Subplot C shows slope degree and D shows methodological overlay showing the approaches of this work: GIS data mapping and digitizing, geological analysis and Python and R programming languages.

3. Results

Python and R based data analysis presented in this research resulted in following findings. Among various shapes of the profiles (Fig. 9) the section of profiles 19, 20, 21 and 22 (Fig. 9, E) had initially uniform shape of the geomorphology that decreased abruptly after the samples 200. The unevenness of the profiles can be explained by the changes geologic conditions and transfer to the neighbour tectonic plate (from

the Philippine to the Pacific Plate, and from Mariana to the Philippine Plate). Fig. 9 and 10 illustrate comparison of the data distribution by profiles, made by Python libraries Bokeh and Matplotlib.

The groups of profiles 23, 24 and 25, (Fig. 9, F) have similar geomorphic shapes: the first part of the profiles is relatively even while the southern part is notable for the increased depths. This analyses can be drawn from Fig. 9. The profiles are coloured with the shared X axis for better visualization. The increase of depths by profiles can be influenced by multiple factors: geographic location and subducting tectonic plates, geologic settings, sedimentation, etc. The geomorphic shapes in the seafloor of the Mariana Trench subduction system vary by four tectonic plates: Mariana, Caroline, Pacific and the Philippine Sea. The Caroline Plate has the least sampling points due to the location in the south-west of the trench. The majority of the samples are located on the Mariana and the Philippine Sea, followed by the moderate number on the Pacific Plate.

Figure 12. *Modelling intersection and correlation by various factors in the project Mariana Trench by Euler-Venn diagrams. A): Bathymetric observations by plates, cumulative computing; B): Sediment thickness, slope degree and closeness of the volcanic spots; C): Subdivision of the slope steepness by classes; D): Methodological approaches.*

The following results are received on sediment thickness based on the data modelling, Fig. 11. The profiles Nr. 1, 24, 25, 23 are notable for the high sediment thickness on the Philippine Plate, profiles Nr. 2, 4, 6, 8 have the highest values for the Mariana Trench, profiles Nr. 3, 5, 8, 6 have the highest values of the sediment thickness for the Pacific Plate, etc. As can be seen from the Fig. 8, the profiles located on the Caroline plate have the least values, since this part of the trench has only a few bathymetric points from the sample pool that cross the Caroline Plate. This can be explained by the fact that sediment thickness is indirectly impacted by the variations in slope degree that creates conditions for the sediment accumulation.

The methodological goal of this research was to identify variations of the steepness in various segments of the trench. The slope steepness degrees and depths were analysed and determined using statistical analysis by libraries of Python and R programming languages. The results revealed that the most steep part of the trench is located in its south-west, north-west and central segments (profiles Nr. 3, 7, 9, 23, 24) where trench crosses Philippine and Pacific tectonic plates. This segment of the trench can be characterised by the largest local depths and sharp abrupt changes in the geomorphic shape. The highest depth gradient can be seen at the profile Nr. 21 is at the distance of 10.600 m from the seal level. The middle moderate slope degrees (Profiles Nr. 1, 11, 4, 5, 10) correlate with the areas where the trench crosses Mariana Plate. These geomorphic segments have relatively balanced bathymetric

elevations records ranging from -7.600 m (profile Nr. 11), -7.750 m (profile Nr. 4), -7.800 m (profile Nr. 5), -7.750 m (profile Nr. 10), to -6.200 m (profile Nr. 1). The modelled depths are shown in on Fig. 5 (jitter plot showing depths ranges).

Finally, the profiles with the least slope degree Nr. 21, 22, 18, 20 are mostly located in the southern segments where trench crosses Caroline Plate. Mostly, the depth values of deeper than -9600 m below the sea level prevails on the bathymetric model for the deepest depth range (Fig. 5), representing the depths of the very steep slopes of the Mariana Trench. Correlation between the geomorphic shape with geologic settings and geographic location revealed that submarine geomorphology of the Mariana Trench is generally influenced by the geospatial settings of the underlying tectonic plates. They affect its shape and cause variation of the steepness across the trench segments. The variation of the shape form is affected by the location of the trench crescent.

4. Recommendations

As a recommendation for the future similar research, it can be noted that it is necessary to select statistical methods the most suitable to the modelling type: data ranking, data distribution (e.g. box plots or 'violin' plots, linear regression or isotonic regression), correlation matrix or correlogram ellipses, etc. Differences in modelling geomorphological variation may result from the application of various methods and algorithms

or different approaches proposed by selected libraries of programming languages (R, Python or Octave), software, e.g., Gretl [39] or SPSS [40]. To achieve the optimal results, it is crucial to follow methodological principles, select correct statistical algorithms for better data visualisation.

Modelling submarine geomorphology can be done using various approaches: by the deepest values (i.e. each profile can have only one dot point), by showing the full profile line in its length, by circular plots, bar plot, dot plots, as demonstrated. The extra- or interpolation of the depths may possibly increase random errors due to the complex nature of the physical landscapes. It can be either the effects of the environmental factors over the structure of the profile (e.g. tectonic slab movement affecting the shape of the trench segment) or mathematical approaches of the selected algorithm for depths approximation. For better quality of the data set, the profiles should be taken at equal lengths and equal distance in between each two. The whole length of the trench should be covered by the profiles. In difficult segments (e.g., where trench crosses several tectonic plates), is is possible to do more dense digitizing. However, in this case it would be reasonable to enlarge the scale of this particular area as a separate fragment.

5. Discussion

Slope gradients are of particular interest for the geomorphological and bathymetric modelling. Slopes of the Mariana Trench are formed by twofold interconnected factors: primary slopes and decreasing slopes caused by the erosion and modification of the first ones. The submarine erosion of the slopes is triggered by the sediment failures and active erosion caused by gravity flows. Factors affecting slope morphology are diverse. Geological factors, such as rock composition, structure, permeability and erodibility of the slope materials (rock) that controls the detachability of the materials from the slope by a set of physical processes. Other factors include inner features of rocks and sediments structure: degree of consolidation, cohesiveness, grain size and geometry (angularity or roundness). As demonstrated in the research, a set of the slopes by 25 cross-sectioning profiles illustrates the shape formed as a results of the internal structure and external processes of the sediment transport and submarine erosion. Various geospatial factors cause formation of the slopes in Mariana Trench, and cause changes in gradient. A complex pattern of the erosional moats and depositions of the sediment drifts along the profiles cause varieties in the profile shapes (Fig. 3).

Submarine erosion has origin in sediment failures and gravity flows by the water streams, erosional processes causing formation of the submarine canyon valleys [41, 42]. However, it should be noted that submarine geomorphology and bathymetric landforms along the trench may in turn affect hydrological processes by accelerating currents. Other factors

that contribute to the formation of the slopes and slightly affect variations in their gradient: benthos, tectonics faults and plates [43], volcanic activities [44], contribution from the alluvial submarine fans, etc. Mariana Trench is a large subduction zone with depths exceeding 11 km. In these areas, cross-sections of the trench illustrate a complex association of the processes of subduction and the deposition of the deep-water sediments. The sediment that fill up the trench area are derived mainly from the volcanic arcs and basins, being transported by the transverse canyons [45].

Understanding such a complex set of processes and dynamics of the submarine geomorphology prerequisite an interdisciplinary approach. Therefore, besides the general geological knowledge and a set of data, quantitative numerical modelling of the processes is a requirement for the correct modelling. Existing papers report questions of the geomorphic analysis, bathymetric and environmental mapping [46, 47, 48]. The advantage of using machine learning approach lies is the increase precision of the results, free from the subjective errors.

Current research shown Python and R based statistical modelling of the complex geological processes that include several variables aimed at processing data sets for submarine geomorphic modelling. Prognosis of the geological processes is possible by means of R and Python based data analysis and plotting. Their libraries are based on the existing statistical and mathematical methods of data processing [49, 50, 51]. Using numerical methods by R and Python libraries, current study indicated geomorphic variations of the profiles affected by geologic and tectonic factors. The study demonstrated effective usage of the statistical methods and advantages of the multi-disciplinary implementation of the IT methods in marine geology.

6. Conclusion

Since rapid development of the programming and machine learning in the last 20 years, some papers reported using Python and R. Testing functionality of various packages, algorithms and modelling methods of these languages, these works have challenged data science and brought fresh ideas in methodology of geosciences. However, insufficient attention was done so far to the statistical data processing by Python and R specifically in marine geology. Some research examples demonstrate application of the traditional GIS methods and mapping for spatial environmental analysis [52, 53]. In marine geology, despite obvious importance of the linkage between the GIS analysis and Python and R statistical libraries, enough case studies are not yet available. This paper contributed in this field by connecting scripting based data analysis with geospatial modelling. Technical application of Python and R statistical libraries enabled successful for modelling of the geomorphic variations of the hadal trench. The study contributed towards methodological development

of the applied statistical analysis by Python and R in marine geology.

Acknowledgement

References

[1]. Jamieson A.J., Fujii T., Mayor D.J., Solan M., Priede, I. G. "Hadal trenches: the ecology of the deepest places on Earth", Trends in Ecology and Evolution, 25 (3), (2009), 190-197.
doi: 10.1016/j.tree.2009.09.009

[2]. Lemenkova P. "Factor Analysis by R Programming to Assess Variability Among Environmental Determinants of the Mariana Trench", Turkish Journal of Maritime and Marine Sciences, 4, (2018), 146–155.
doi: 10.6084/m9.figshare.7358207

[3]. Lemenkova P. "R scripting libraries for comparative analysis of the correlation methods to identify factors affecting Mariana Trench formation", Journal of Marine Technology and Environment, 2, (2018), 35-42.
doi: 10.6084/m9.figshare.7434167

[4]. Duineveld G. "Activity and composition of the benthic fauna in the Whittard Canyon and the adjacent continental slope (NE Atlantic)", Oceanologica Acta, 24, (2001), 69–83.

[5]. McCullagh P. Slopes, in: FitzGerald, B.P. (ed.), Modern Concepts in Geomorphology, (1988).

[6]. Ritter D.F., Kochel C.R., Miller J.R. Process Geomorphology (3rd Edition): Wm. C. Brown Publishers, Dubuque, IA, (2002), 544.

[7]. Chorley R.J., Schumm S.A., Sugden D.E. Geomorphology. London: Methuen and Co. Ltd. (1984).

[8]. Summerfield M.A. Global Geomorphology. New York: John Wiley and Sons, (1991).

[9]. Easterbrook D.J. Surface Processes and Landforms: Macmillan Pub. Co., (1993).

[10]. Bloom A. Geomorphology, A systematic analysis of Late Cenozoic landforms, 3rd ed. Prentice Hall, Upper Saddle River, N.J., (1998).

[11]. Harris P. T., Macmillan-Lawler M., Rupp J., Baker E.K. Geomorphology of the oceans. Marine Geology, 352, (2014), 4–24.

[12]. Wilson M.F.J., O'Connell,B., Brown,C., Guinan,J.C., Grehan A.J. "Multiscale terrain analysis of multibeam bathymetry data for habitat mapping on the continental slope", Marine Geodesy 30, (2007), 3–35.

[13]. Lemenkova P. "Hierarchical Cluster Analysis by R language for Pattern Recognition in the Bathymetric Data Frame: a Case Study of the Mariana Trench, Pacific Ocean. Virtual Simulation, Prototyping and Industrial Design". Proceedings of the 5th Int'l Sci.-Pract. Conference, 2(5), (2018), Ed. M. N. Krasnyansky, Tambov: TSTU Press, 147–152.
doi: 10.6084/m9.figshare.7531550

[14]. Pickering K.T., Hiscott R.N. Deep marine systems: processes, deposits, environments, tectonics and sedimentation. Wiley, Chichester, (2015).

[15]. Wu S., Takahashi N., Tokuyama H., Wong H.K. "Geomorphology, sedimentary processes and development of the Zenisu deep-sea channel, northern Philippine Sea", Geo-Marine Letters, 25(4), (2005), 230-240.
doi: 10.1007/s00367-005-0210-9

[16]. Lemenkova P. "An Empirical Study of R Applications for Data Analysis in Marine Geology", Marine Science and Technology Bulletin, 8(1), (2019), 1-9.
doi: 10.33714/masteb.486678

[17]. Okada H. "Anatomy of trench-slope basins: examples from the Nankai trough", Palaeogeography, Palaeoclimatology, Palaeoecology, 71 (1-2), (1989), 3-13.
doi: 10.1016/0031-0182(89)90026-6

[18]. Yu H.S. "Geological characteristics and distribution of submarine physiographic features in the Taiwan region", Marine Georesources and Geotechnology, 21 (3-4), (2003), 139–153.
doi: 10.1080/ 713773391

[19]. VanderPlas, J. Python Data Science Handbook. Essential Tools for Working with Data. O'Reilly, (2016).

[20]. McKinney W. PyData Development Team. Pandas: powerful Python data analysis toolkit Release 0.24.0. [Online] http://www.python.org [Accessed: 10 April 2019].

[21]. Lemenkova P. "Processing oceanographic data by Python libraries NumPy, SciPy and Pandas", Aquatic Research, 2, (2019), 73-91.
doi: 10.3153/AR19009

[22]. Duchesnay E., Löfstedt T. Statistics and Machine Learning in Python Release 0.2, 2019. [Online] http://www.python.org [Accessed: 10 April 2019].

[23]. R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation

for Statistical Computing. Vienna, Austria, (2014). http://www.R-project.org [Accessed: 13 March. 2019].

[24]. Lemenkova P. "K-means Clustering in R Libraries {cluster} and {factoextra} for Grouping Oceanographic Data", International Journal of Informatics and Applied Mathematics, 2(1), (2019), 1-26. doi: 10.6084/m9.figshare.9891203

[25]. Rossum, G. van. Python Programming Language, 2011. https://www.python.org/ [Accessed: 10 April 2019].

[26]. Rossetier D.G. Tutorial: An example of statistical data analysis using the R environment for statistical computing, (2017).

[27]. NumPy community. NumPy Reference. Release 1.16.1, (2019), https://www.python.org/ [Accessed: 10 April 2019].

[28]. Lemenkova P. "Testing Linear Regressions by StatsModel Library of Python for Oceanological Data Interpretation", Aquatic Sciences and Engineering, 34, (2019), 51–60. doi: 10.26650/ASE2019547010

[29]. Crozier M. Slope Evolution, in Goudie, A.S., ed., Encyclopedia of Geomorphology, Volume 2, Routledge, New York, NY, (2004).

[30]. Savage L.J. The Foundations of Statistics, 2nd revised ed., Dover, New Yor, (1972).

[31]. Kruskal J.B. "Nonmetric Multidimensional Scaling: A numerical method", Psychometrika, 29 (2), (1964), 115–129. doi:10.1007/BF02289694

[32]. Best M.J. Chakravarti N. "Active set algorithms for isotonic regression; a unifying framework", Mathematical Programming, 47 (1–3), (1990), 425–439. doi:10.1007/BF01580873

[33]. Wu W.B., Woodroofe M., Mentz G. "Isotonic regression: Another look at the changepoint problem", Biometrika, 88 (3), (2001), 793–804. doi:10.1093/biomet/88.3.793

[34]. Leeuw J. de Hornik K., Mair P. "Isotone Optimization in R: Pool-Adjacent-Violators Algorithm (PAVA) and Active Set Methods", Journal of Statistical Software, 32 (5), (2009), 1–24. doi:10.18637/jss.v032.i05.

[35]. Lemenkova P. Scatterplot Matrices of the Geomorphic Structure of the Mariana Trench at Four Tectonic Plates (Pacific, Philippine, Mariana and Caroline): a Geostatistical Analysis by R. In: Problems of Tectonics of Continents and Oceans. Proceedings of the 51st Tectonics Meeting, Ed. Degtyarev, K. E. (1) RAS Institute of Geology. Moscow: GEOS, (2019), 347–352. doi: 10.6084/m9.figshare.7699787.v1

[36]. Bretz F., Hothorn T. aWestfall P. Multiple Comparisons Using R. Taylor and Francis Group, LLC. (2011).

[37]. Lemenkova P. "Statistical Analysis of the Mariana Trench Geomorphology Using R Programming Language", Geodesy and Cartography, 45(2), (2019), 57–84. doi: 10.3846/gac.2019.3785

[38]. Gardner J.V., Armstrong A.A., Calder B.R. Beaudoin J. So. "How Deep Is the Mariana Trench?", Marine Geodesy, 37, (2014), 1-13, doi: 10.1080/01490419.2013.837849

[39]. Lemenkova P. "Regression Models by Gretl and R Statistical Packages for Data Analysis in Marine Geology", International Journal of Environmental Trends, 3(1), (2019), 39–59. doi: 10.6084/m9.figshare.8313362.v1

[40]. Lemenkova P. "Numerical Data Modelling and Classification in Marine Geology by the SPSS Statistics", International Journal of Engineering Technologies, 5(2), (2019), 90–99. doi: 10.6084/m9.figshare.8796941

[41]. Covault J.A., Fildani A., Romans B.W., McHargue T. "The natural range of submarine canyon-and-channel longitudinal profiles", Geosphere, 7, (2011), 313–332. doi:10.1130/GES00610.1

[42]. Clark M.J., Small R.J. Slopes and weathering: Cambridge University Press, Cambridge, England, (1982).

[43]. Jones O.P., Simons R.R., Jones E.J.W., Harris J.M. "Influence of seabed slope and Coriolis effects on the development of sandbanks near headlands", Journal of Geophysical Research, 111, (2006), 1–23. doi: 10.1029/2005JC002944

[44]. Dadson S., Hovius N., Pegg S., Dade W.B., Horng M.J., Chen H. "Hyperpycnal river flows from an active mountain belt", Journal of Geophysical Research Earth Surface, 110 (F4), (2005). doi: 10.1029/2004JF000244

[45]. Karig D.E., Sharman G.F. "Subduction and accretion in trenches", Geological Society of America Bulletin, 86 (3), (1975), 377–389.

[46]. Mayer L. Introduction to Quantitative Geomorphology: Prentice Hall, Englewood Cliffs, NJ, (1990).

[47]. Lemenkova P., Promper C., Glade T. Economic Assessment of Landslide Risk for the Waidhofen a.d. Ybbs Region, Alpine Foreland, Lower Austria. In E. Eberhardt, C. Froese, A. K. Turner, & S. Leroueil (Eds.), Protecting society through improved understanding, (2012), 279-285. doi: 10.6084/m9.figshare.7434230

[48]. Schenke H.W., Lemenkova P. "Zur Frage der Meeresboden-Kartographie: Die Nutzung von AutoTrace Digitizer für die Vektorisierung der Bathymetrischen Daten in der Petschora-See", Hydrographische Nachrichten, 25(81), (2008), 16–21. doi: 10.6084/m9.figshare.7435538.v2

[49]. Myers J.L. Well A.D. Research Design and Statistical Analysis. Ed. 2, Lawrence Erlbaum, U.S., (2003).

[50]. Cowan G. Statistical Data Analysis. Oxford Science Publications. Clarendon Press, Oxford, UK, (1998).

[51]. Brownlee K.A. Statistical theory and methodology in science and engineering. 2nd ed., New York: John Wiley & Sons, (1965).

[52]. Klaučo M., Gregorová B., Stankov U., Marković,V., Lemenkova P. "Determination of ecological significance based on geostatistical assessment: a case study from the Slovak Natura 2000 protected area", Central European Journal of Geosciences, 5(1), (2013), 28-42. doi: 10.2478/s13533-012-0120-0

[53]. Suetova I.A., Ushakova L.A., Lemenkova P. Geoinformation mapping of the Barents and Pechora Seas. Geography and Natural.