TITLE: REGRESSION MODELS BY GRETL AND R STATISTICAL PACKAGES FOR DATA

ANALYSIS IN MARINE GEOLOGY

AUTHORS: Polina LEMENKOVA

PAGES: 39-59

ORIGINAL PDF URL: https://dergipark.org.tr/tr/download/article-file/741828

Research Article

REGRESSION MODELS BY GRETL AND R STATISTICAL PACKAGES FOR DATA ANALYSIS IN MARINE GEOLOGY

Polina Lemenkova [1]*

[1] ORCID ID number: 0000-0002-5759-1089. Ocean University of China, College of Marine Geo-sciences.

238 Songling Rd., 266100, Qingdao, Shandong, P. R. C. Tel.: +86-1768-554-1605.

Abstract

Gretl and R statistical libraries enables to perform data analysis using various algorithms, modules and functions. The case study of this research consists in geospatial analysis of the Mariana Trench, a hadal trench located in the Pacific Ocean. Technically, data modelling was performed using multi-functional combined approach of both Gretl and R libraries. The study aim is modelling and visualizing trends in the variations of the trench's properties: bathymetry (depths), geomorphology (steepness gradient), geology, volcanism (igneous rocks). The workflow included following statistical methods computed and visualized by Gretl and R libraries: 1) descriptive statistics; 2) box plots, normality analysis by quantile-quantile (QQ) plots; 3) local weighted polynomial regression model (loess), 4) linear regression by several methods: weighted least squares (WLS) regression, ordinary least squares (OLS) regression, maximal likelihood linear regression and heteroskedasticity regression model; 5) confidence ellipses and marginal intervals for data distribution; 6) robust estimation by Nadaraya–Watson kernel regression fit; 7) correlation analysis and matrix. The results include following conclusions. First, the slope angle gradient has a correlation with the geological settings of the trench and distribution of volcanic igneous rocks. Second, the sediment thickness varies by the tectonic plates showing unequal distribution in space. Third, there is a correlation between the slope gradient and aspect degree. Forth, geospatial analysis of the bathymetry shows that the deepest part of the trench is located in the south-west.

_____
[1] Corresponding Author
Email: pauline.lemenkova@gmail.com

INTRODUCTION

Gretl (abbreviation stands for: GNU Regression, Econometrics and Time-series Library) is a functional cross-platform software package for the econometrics and statistical data analysis developed by A. Cottrell and R. Lucchetti, written in the C programming language [1], [2]. Gretl is a part of the a GNU (GNU is Not Unix) project by the Free Software Foundation (FSF) promoting free open source software [3]. Having an intuitive GUI to operate, flexible functionality of the modules and embedded statistical functions, Gretl was chosen in this research as a useful software for statistical data analysis. R, a programming language with a powerful functionality of statistical packages [4], [5] was selected as a comparative instrument to perform advanced plotting (multi-facetted plots and scatterplots with overlaid dots) for complex data analysis and modelling through scripting.

Nowadays, data science and data analysis are important part both in industry and in academic research, because the computerization and increased amounts of information lead to the need of processing large amount of data. Therefore, in both academia and economic domain, proper effective data analysis, methods, approaches and techniques are widely discussed [6], [7], [8], [9]. Data analysis specifically in geosciences as a general Earth science domain, and in geology as its sub-branch, also received attention [10], [11], [12]. As for marine geology, there are some examples of using R [13], [14] and Python [15], [16] libraries for the data modelling and analysis. However, processing data in marine geology is mostly limited by the GIS analysis based on geoinformatics and cartography. Current research highlights the effectiveness of the geospatial data processing by means of pure statistical software, such as Gretl and R, filling the gap between the mathematical approaches in the data analysis and geospatial data processing. Presented study aims at modelling, visualizing and understanding trends in spatial variations of the properties of the ocean trench.

A case study of this research is the Mariana Trench, the deepest place of the Earth located southeast of the island of Guam, west Pacific Ocean, reaching from Japan to Palau as a 2550 km long geometric arc. It is the deepest trench in the world, with a maximum depth of ~11,000 m [17]. Mariana Trench is a geologically unique place formed as a typical example of the tectonic plate spreading in an oceanic subduction system: the Pacific tectonic plate subducts beneath the Philippine Sea plate forming Mariana Trench. Such tectonic activities results in the formation of the active volcanic arc near the Mariana Trench which consists in 9 islands and more than 60 seamounts, around 20 of which are hydrothermally active [18]. The

motion of the tectonic plates naturally affects the geomorphology of the trench. Furthermore, the interaction between the subducting plates and the underlying mantle cause gradual motion of the trench itself. Thus, the seafloor of the Mariana trench represents the site where oceanic lithosphere subducts into the mantle with many factors that influence the rate and direction of the trench migration [19]. The details of the tectonics and tectonophysics properties of the Mariana Trench subduction system are discussed in more details in various research papers (e.g. [20], [21], [22]).

Although until recent, the hadal deep-sea trenches were thought to be an almost biological deserts and understanding of their biology and ecology was very scarce due to the inaccessibility, it is now proved [23] that there are deep-ocean ecosystems even at such great depths as the Challenger Deep, the southernmost part of the Mariana Trench, existing due to the supply of nutrients and energy in form of chemical species. This becomes possible because of the constant ocean motion and waves turbulence upbringing sufficient nutrients for the organisms living in the trench [24]. Nevertheless, Mariana Trench is notable for the low temperature and low productivity comparing to other world trenches, as well as intermediate bottom temperature underlying the lowest surface productivity comparing to other trenches [25]. Being a hadal trench, Mariana Trench has higher hydrostatic pressure and relatively isolated bathymetry, while other geophysical and biogeochemical conditions, such as temperature, salinity, and dissolved oxygen are similar to those in abyssal settings [26].

As for topography of the Mariana Trench, it reflects rather complex geological situation: there are detected changes in strike of the bathymetric shape of the trench axis around 142°20E, being N 85°E to the east and N 80°E to the west of that longitude. The elongated depressions are located along the trench axis, not parallel with the trench axis. Other remarkable topographic feature of the Mariana Trench include a set of the elongated ridges and escarpments, having a shape of a half graben topographically. These geomorphic structures are accompanied by the plate bending on the outer slope, as well as soft and smooth seafloor of the western depression of the Mariana Trench [27].

2. MATERIALS AND METHODS

The research aim and specific task of this work is to analyze the correlation trends in several factors affecting the submarine ecosystem of the Mariana Trench: bathymetry (depths), geomorphology (slope steepness gradient), geology (location on four tectonic

plates: Mariana, Caroline, Pacific and Philippine Sea), volcanism in the adjacent area (igneous rocks). The data set consisting of the 25 cross-sectioning profiles with 518 observation samples in each. The workflow included following statistical methods computed and visualized in Gretl and R libraries: 1) descriptive statistics (mean, extremes, median); 2) data distribution by box plots, normality analysis by quantiles (QQ) plots; 3) local weighted polynomial regression (loess), 4) linear regression forecast evaluation by weighted least squares (WLS), ordinary least squares (OLS), limited information maximal likelihood (LML) and heteroskedasticity; 5) visualized confidence ellipses and marginal intervals for data distribution; 6) robust estimation modelling by Nadaraya-Watson fit; 7) correlation analysis: pairwise, triple and matrix plotting for all factors from the data set.
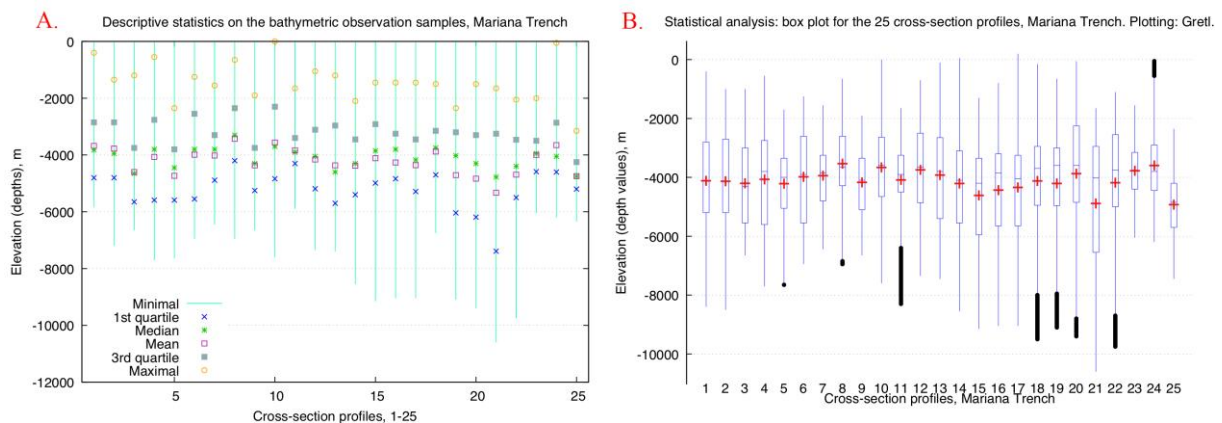

Figure 1. Descriptive statistical analysis of the bathymetry in the study area: minimal, mean, median and maximal values (A); box plots of the cross-sectional profiles (B). Plotting: Gretl

The descriptive statistical analysis (Figure 1) was computed by the 'View / Graph specified vars / Boxplots' where the initial data set (25 profiles with 518 observations in each) was imported as a .csv table with structure of the data set chosen as 'Panel' type of the data structure and Panel data organization as 'Stacked cross sections'. The multi-faceted quantiles (QQ) and regression analysis plots (Figure 3 and 4) were performed in R for visualizing the variability of the bathymetric data values (elevation depths) by 25 profiles. The statistical approach followed common existing methods [28], [29]. The quantile-quantile plot (QQ) for the 25 profiles was modeled (Figure 3) by R packages {ggplot} and {qqplotr} by function stat_qq(). The faceting for the 25 profiles combined in one plot was performed using {ggplot2} package of R by function: facet_wrap(~ profiles, labeller = label_both). The same scales were used for all panels for comparability of the sub-plots (Figure 3).
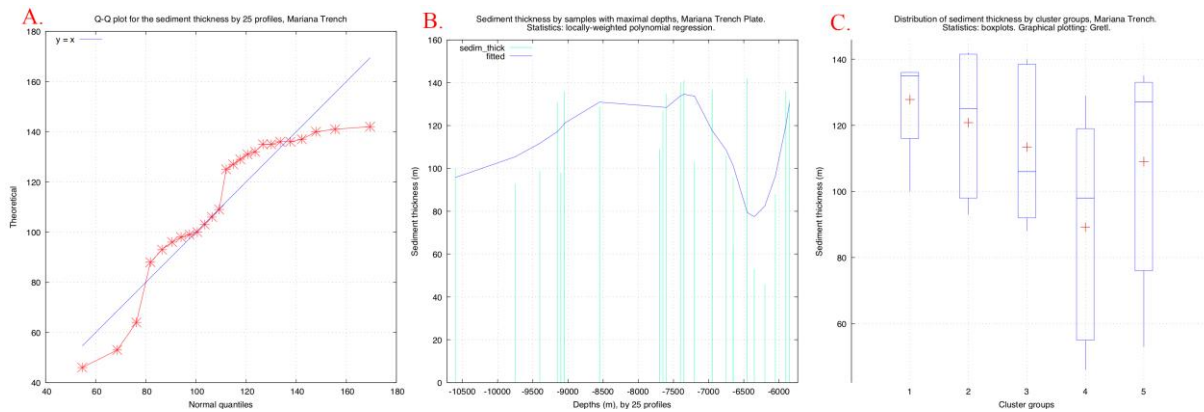
Figure 2. Statistical analysis of the sediment thickness: QQ plot (A),   locally-weighted polynomial regression (B); box plots ranked by clusters (C).

The regression analysis (Figure 4) was plotted using {ggplot} package of R by the following function: ggplot(MarDF_NEW, aes(x = observ, y = value, shape = "Observation points", color = "Observation points", size = "Observation points"). Statistical lines were added by special calls of geom_point(), geom_quantile() and geom_smooth(). The local polynomial regression was computed by fitting a smooth curve between the two variables by 25 profiles, respectively: depths values by 518 observation samples in each profile (Figure 4).
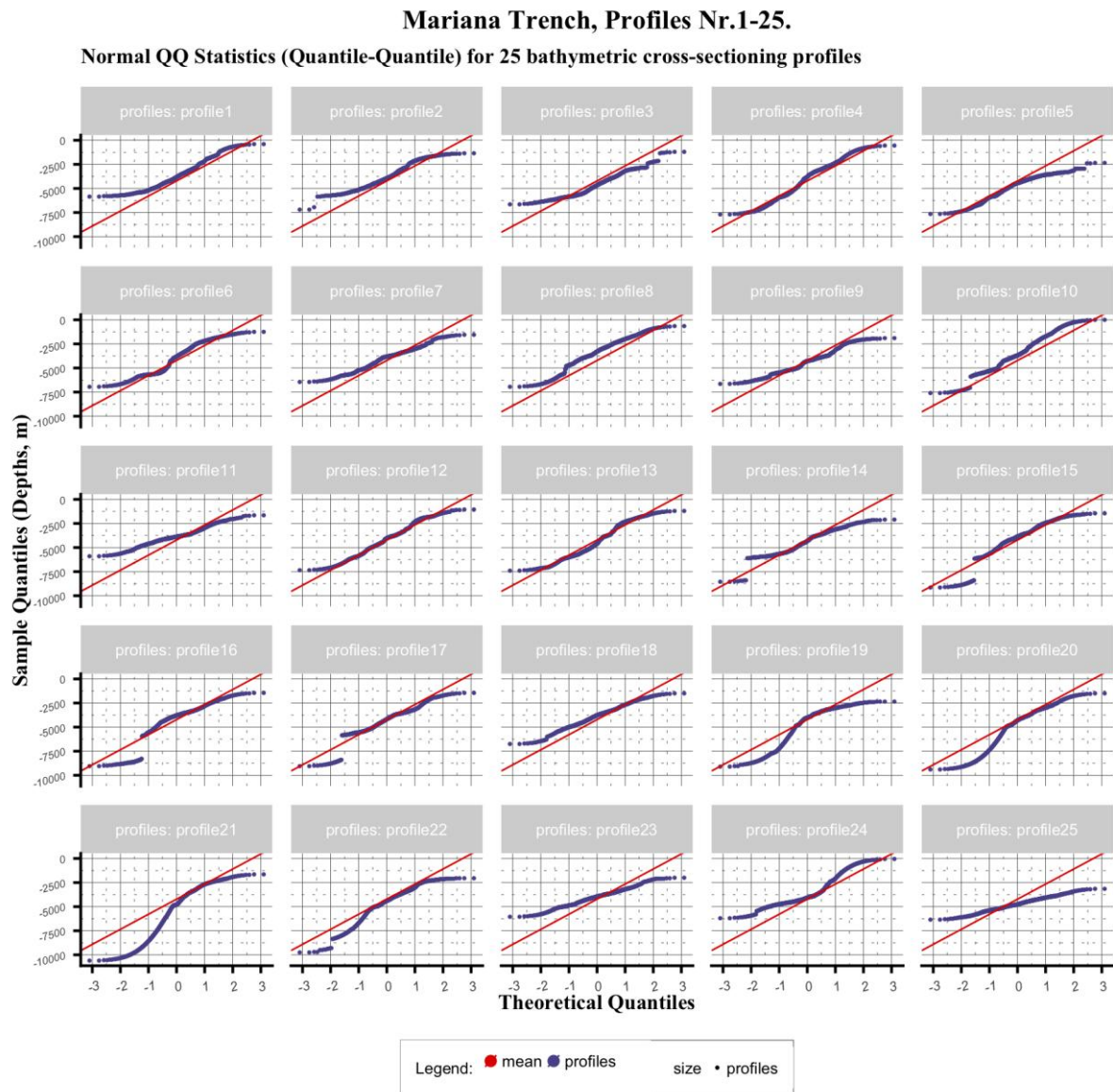
Figure 3. Statistical quantile-quantile (QQ) plot for the 25 cross-sectioning profiles

The order of the polynomial in the independent variable and the proportion of the observation samples used in each local regression were specified in bandwidth, defined in R package {ggplot} by the following command: geom_smooth(aes(x = observ, y = value, colour = "Loess method"), method = loess, se = TRUE, span = .4, size=.2, linetype = "solid", show.legend = TRUE), Figure 4.
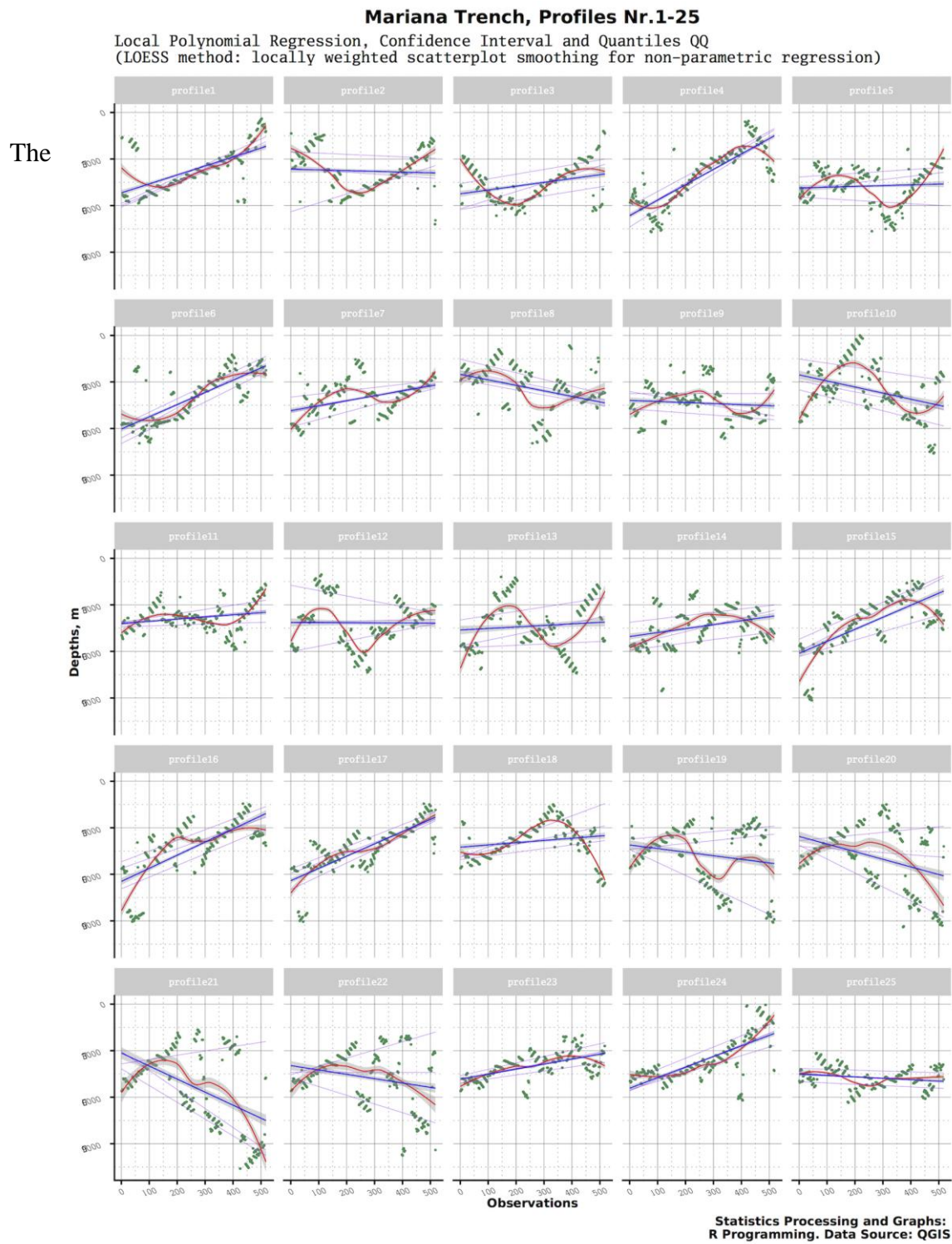
The

Figure 4. Statistical regression analysis for the 25 cross-sectioning profiles

The analysis of inter-dependancies by factors was performed using data set containing multiple variables (25 rows for values on geomorphology, geology, maximal and minimal depths, gradient slope steepness angle in tangent, etc. The table was imported as a .csv into Gretl environment with a 'Cross-sectional' type of the data structure. The method for

Ordinary Least Squares data modelling (Figure 6) followed 'Model / OLS' function in the menu. The dependent variable was selected as sediment thickness and several regressors were chosen for each sub-plot (Figure 6, A-F): tectonic plates (Mariana, Pacific, Philippine Sea), gradient angle steepness, volcanic rocks, tangent slope angle.
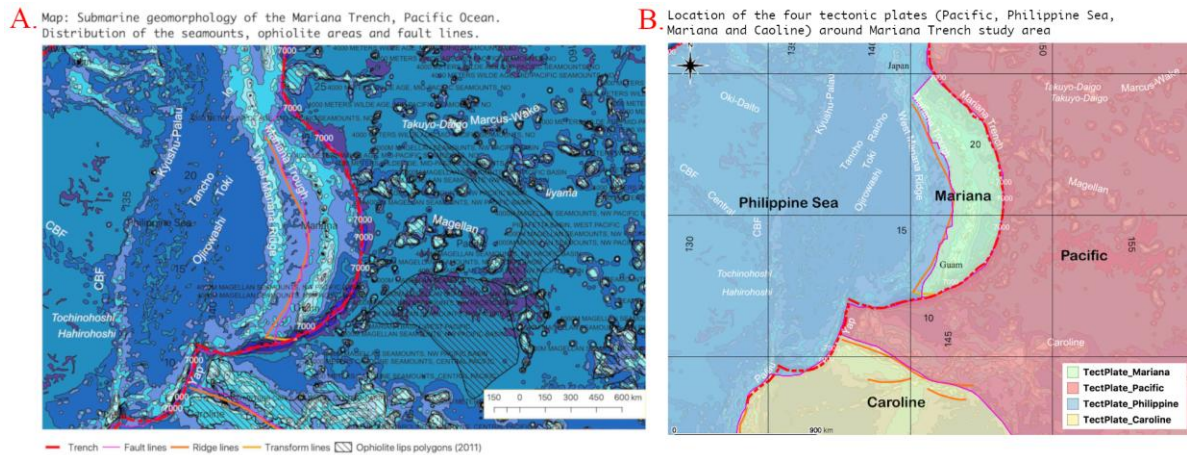

Figure 5. Study area: Mariana Trench crossing four tectonic plates

Other statistical methods were tested following existing described methodological approaches [30], [31], [32] via the command 'Models / Other Linear Models' in Gretl menu: Limited Information Maximal Likelihood (LIML), Weighted Least Squares (WLS), Heteroskedasticity-corrected, respectively (Figure 6, G-I). methodology for plotting Locally-Weighted Polynomial Regression (Loess) model [33] (Figure 7) followed 'Model / Robust Estimation / Loess' command of Gretl. Loess visualizes a series of sub-plots (Figure 7, A-D) showing predicted values of the dependent geological variable of sediment thickness for each non-missing value of the independent variable of the tectonic plates: Mariana, Pacific, Philippine Sea, Caroline. Confidence ellipses (Figure 8) illustrate the generalization of a confidence interval represented as an ellipsoid around an interval of observation samples containing 95% of the pairwise against geologic data variable (sediment thickness) (Figure 8, following from A to F): igneous volcanic areas, Caroline Plate, Pacific Plate, minimal depths, Philippine Plate, Mariana Plate.
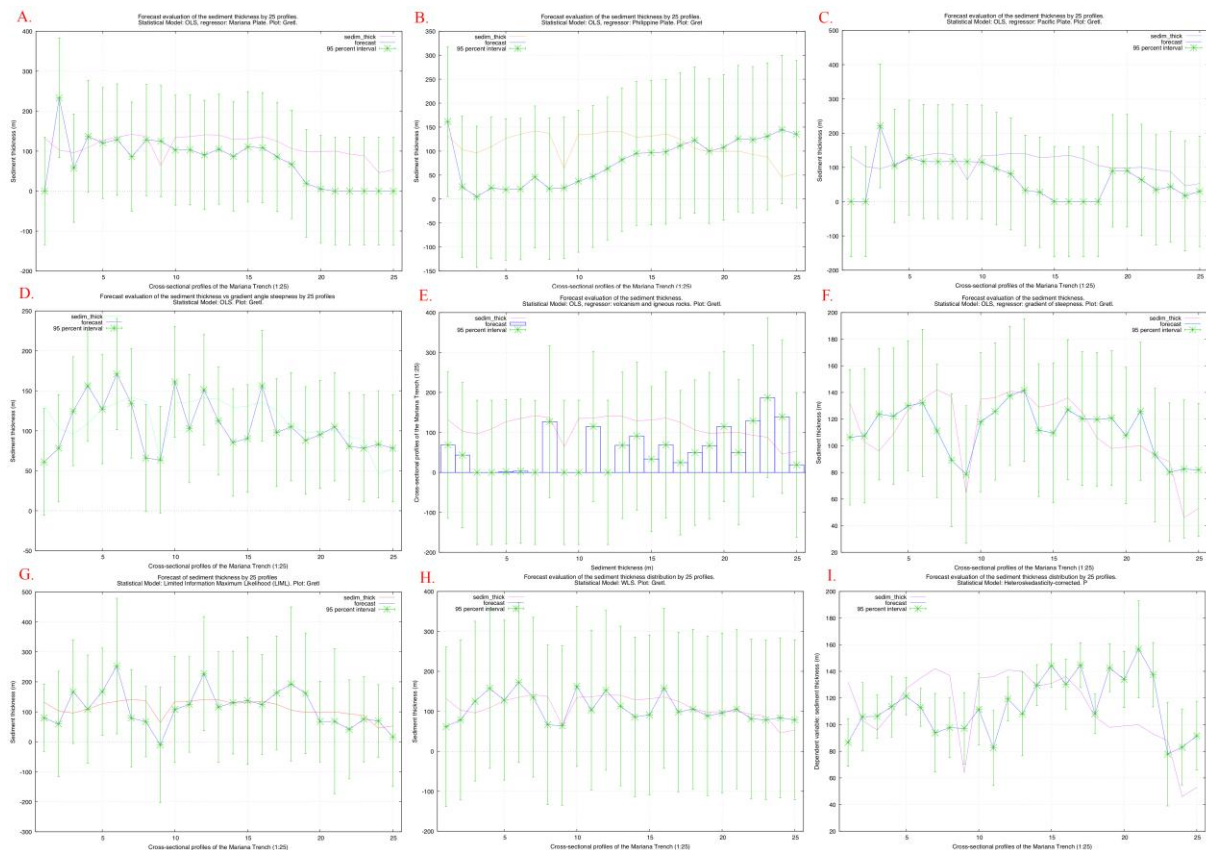
Figure 6. Forecast evaluation of the spatial variations in sediment thickness. Modelling methods: OLS (A-F); LML (G); WLS (H); Heteroskedastisity-corrected (I). Plotting: Gretl

Methodological approach in finding correlation between the variables consists in finding similarities between the selected values by existing algorithms [34], [35], [36]. The algorithms for the used approaches of OLS, WLS and LIML are provided in the existing statistical and reference literature [37], [38]. Specifically in this research the correlation matrix was computed using the 'View / Graph specified vars / Correlation Matrix' command.

3. RESULTS

3.1. Descriptive statistical analysis of the bathymetric and geological factors, Plotting: Gretl

Findings in the primary statistical analysis of the examination of bathymetric data distribution show the most prominent depth by the profiles 20, 21, 22 located in the south-western part of the Mariana Trench (Figure 1).
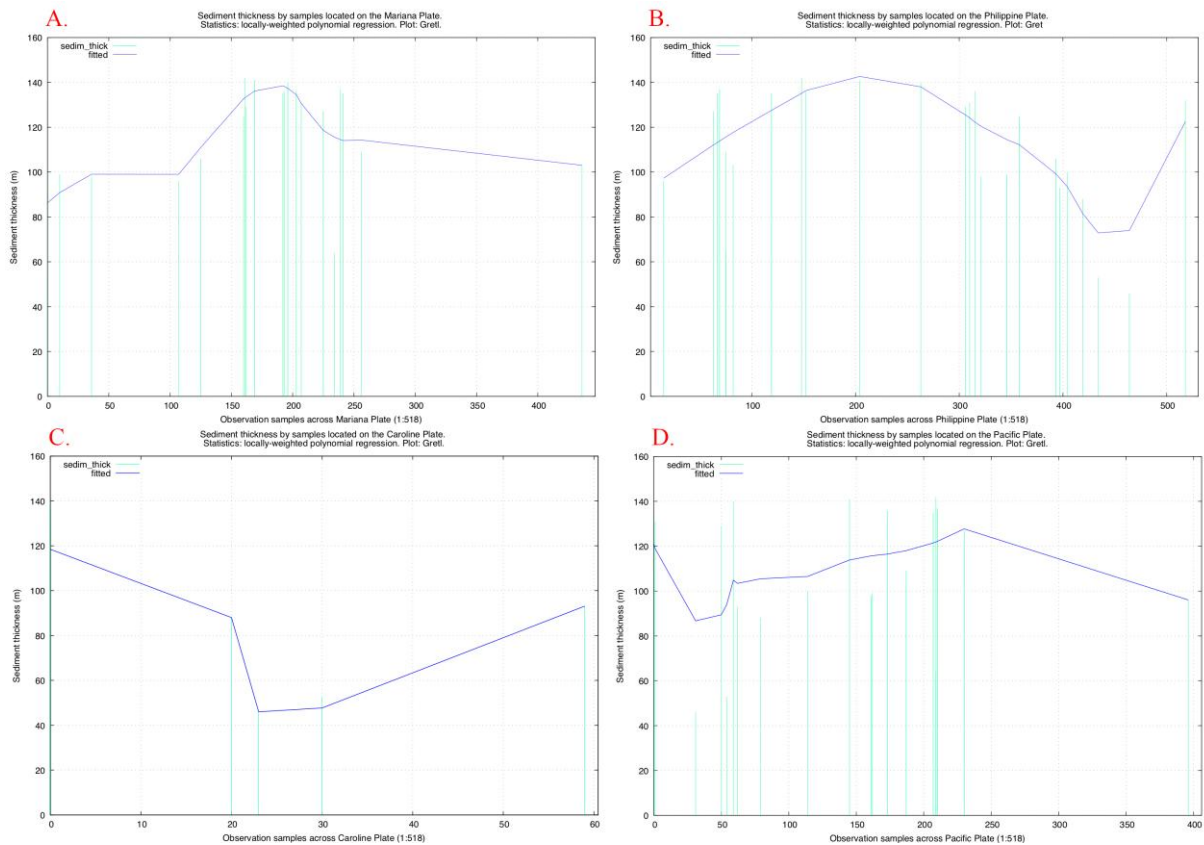
Figure 7. Spatial variations of the sediment thickness by four tectonic plates. Plotting: Gretl.

Data records on the sediment thickness have normal distribution patterns (Figure 2 A). The increased values in the sediment thickness can be noted (Figure 2) in the profile groups with the depths range -9,000 to -9,200 m (Figure 2 B) and from -7,700 to -7,000 meters with a gradual decrease thereafter. Profiles with depths ranging from -6,500 to -6,000 meters show lesser values in sediment thickness (Figure 2 B). Grouping profiles by values of the sediment thickness values (Figure 2 C) shows five distinct clusters with mean values of 138, 124, 114, 98 and 125 m, respectively (Figure 2 C). The QQ for the 25 profiles shows sample and theoretical quantiles for the bathymetric data (depths, in m) for each of the 25 profiles. The data shows normal distribution character. The Local Polynomial Regression for the 25 profiles was modeled in a facetted plot (Figure 4) using methodology explain in a previous chapter.

3.2. Evaluation of the spatial variations of the geological factors. Plotting: Gretl.

The particular spatial characteristics of the Mariana Trench consists in its form: a crescent shape crossing four tectonic plates: Mariana, Caroline, Pacific and the Philippine Sea (Figure

5). Hence, to analyze spatial variations in the attribute data distribution across the plates, a spatial analysis by four plates was performed, which is discussed in this chapter.
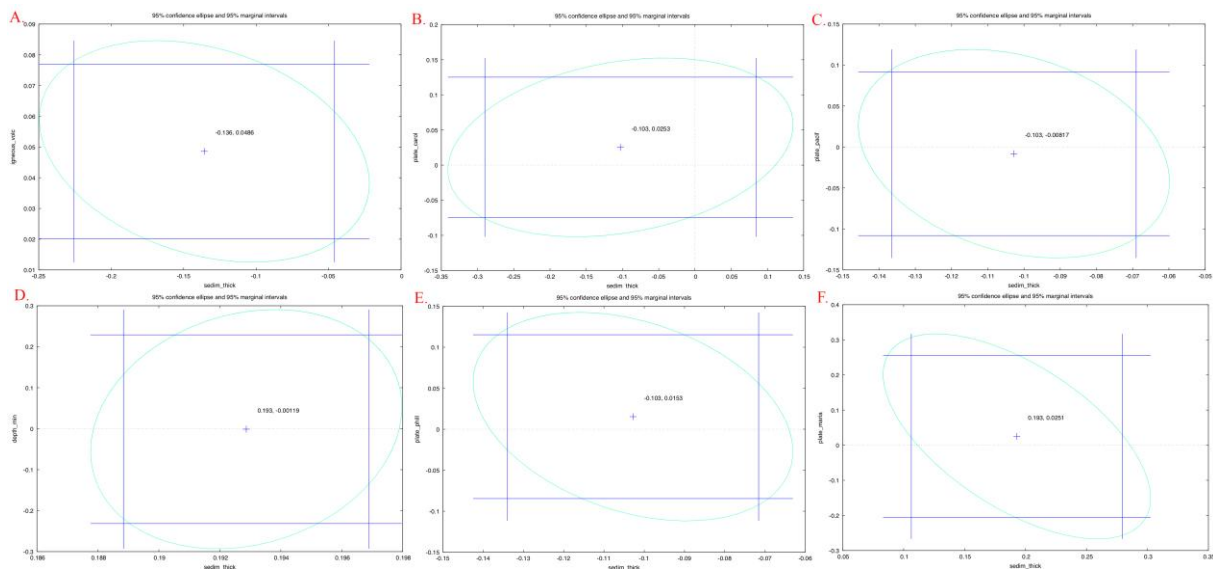

Figure 8. Confidence ellipse and marginal intervals for the data distribution. Plotting: Gretl.

Close examination of the Figure 9 shows that the geomorphic, geologic and bathymetric variables are distributed unequally by tectonic plates. There include the following tested parameters: slope angle expressed in tangent degree, sediment thickness, maximal depths and igneous volcanic rocks. The variations may be explained by the diverse geospatial conditions and geological settings in these areas (Figure 9).
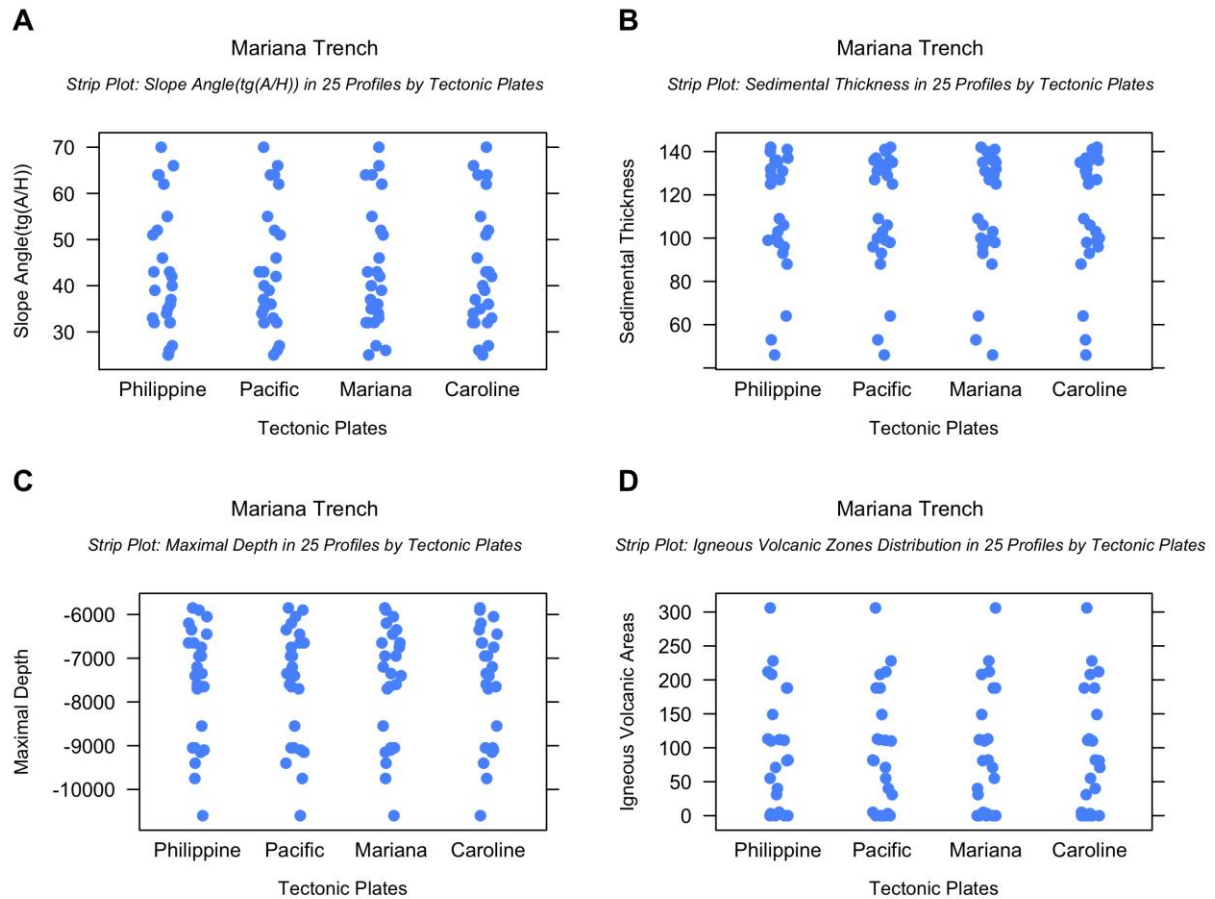
Figure 9. Variations in the geomorphology (A), geology (B), bathymetry (C), volcanism (D) visualized by strip plot. Plotting: R.

The geology of the trench (sediment thickness) has a correlation with geomorphic parameters of the slope shape: slope angle gradient, steepness tangent angle and aspect degree location of the igneous rocks (volcanism effect), Figure 10.
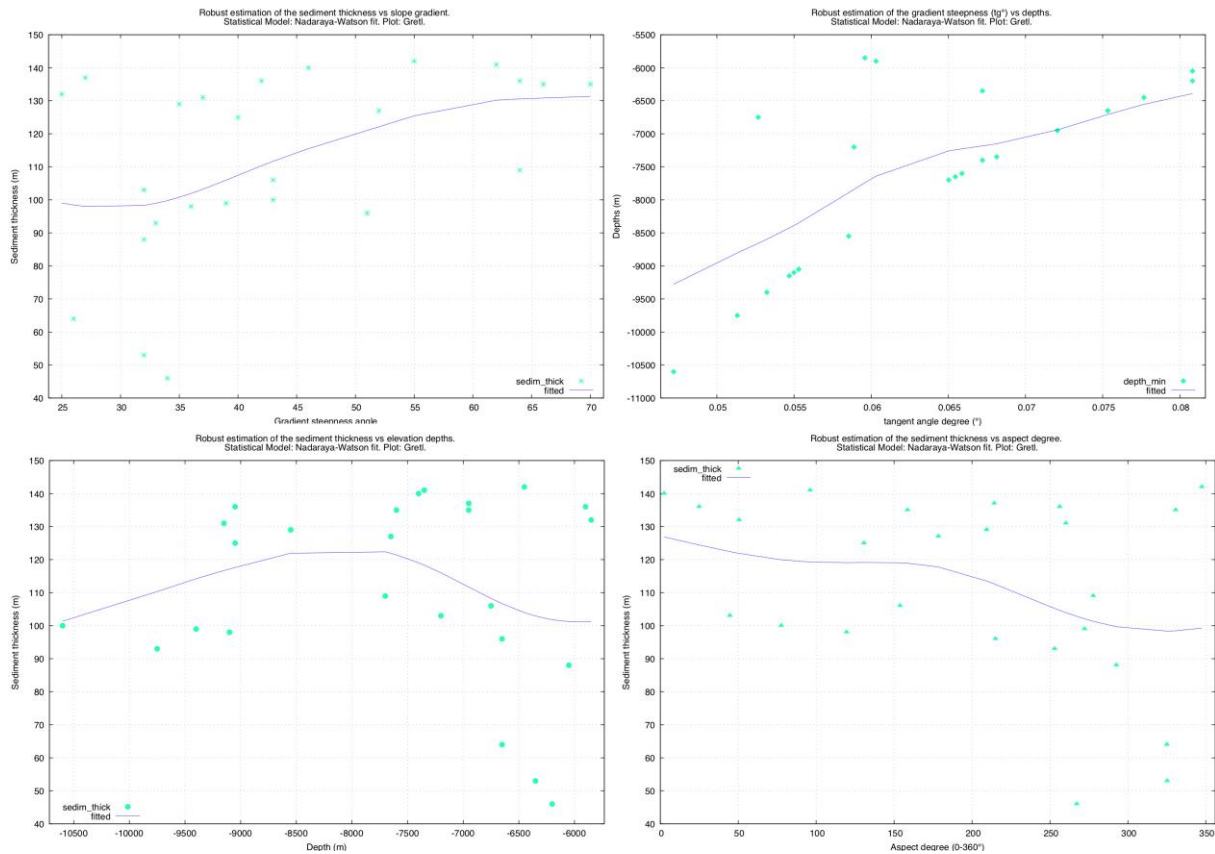
Figure 10. Robust estimation of the correlation trend. Model method: Naradaya-Watson Fit.

3.3. Correlation analysis of various factors: geomorphology, bathymetry and geography

The relevance of variations in multiple factors (Figure 13) was studied in relation to the cross-sectional profiles across the trench (Figure 12), and specifically pairwise, for slope angle gradient and sediment thickness, slope angle gradient and igneous volcanic areas, and slope angle gradient and aspect degree (Figure 12 from upper to lower sub-plots, respectively). Besides, there is correlation between the slope gradient observations and aspect degree for the profiles located in the north and north-eastern part of the trench: 1-6, 16-19, 13-14, Figure 12. The 3D visualization of the triple correlation between geomorphic and geological variables (Figure 12) shows dependancies between these factors and tectonics, examples of the Pacific and the Philippine Sea plates (Figure 12 A).
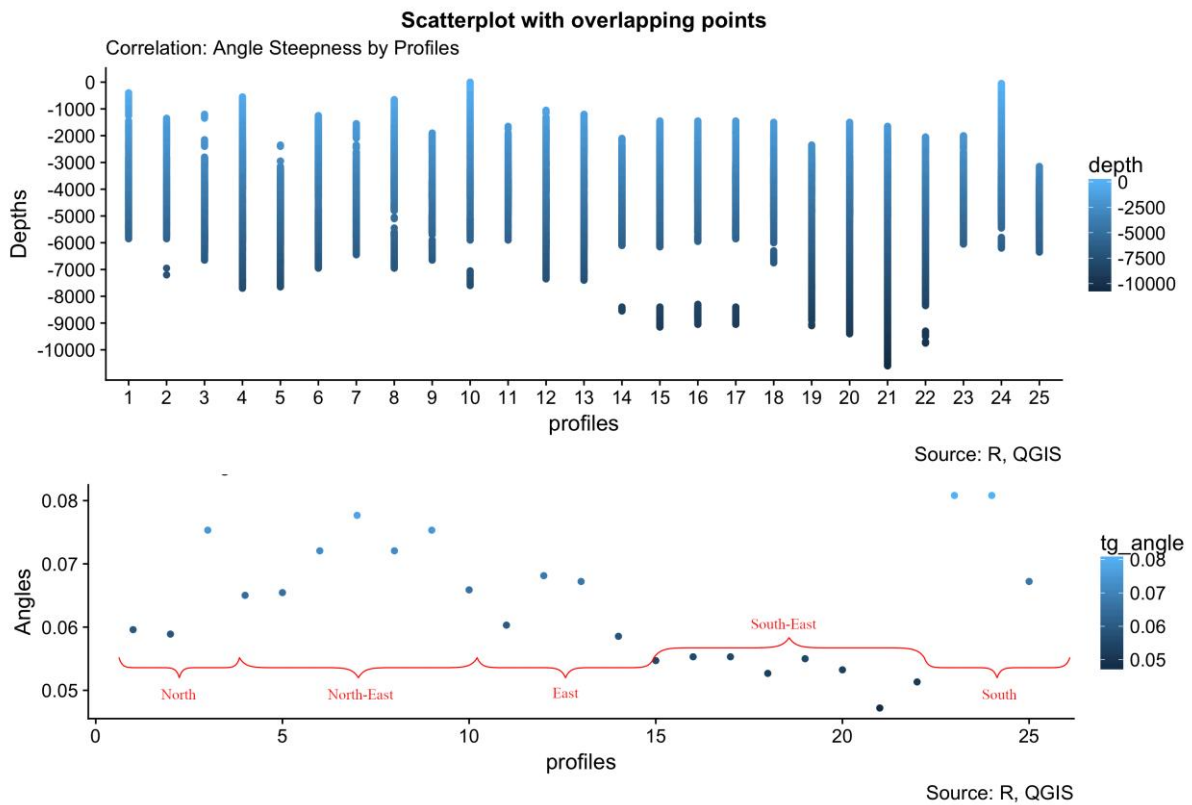
Figure 11. Scatterplot visualization of the dependancies between factors: elevation range by cross-sectioning profiles (upper plot); slope gradient and geospatial location (lower plot).

Overall, a strong relationship was not observed between tectonic plates (dependence < -0.5). However, a closer look revealed that the bathymetric values (depths, m) have a high correlation value, especially at profiles 14 to 17 and 1 to 6 (Figure 13, B).
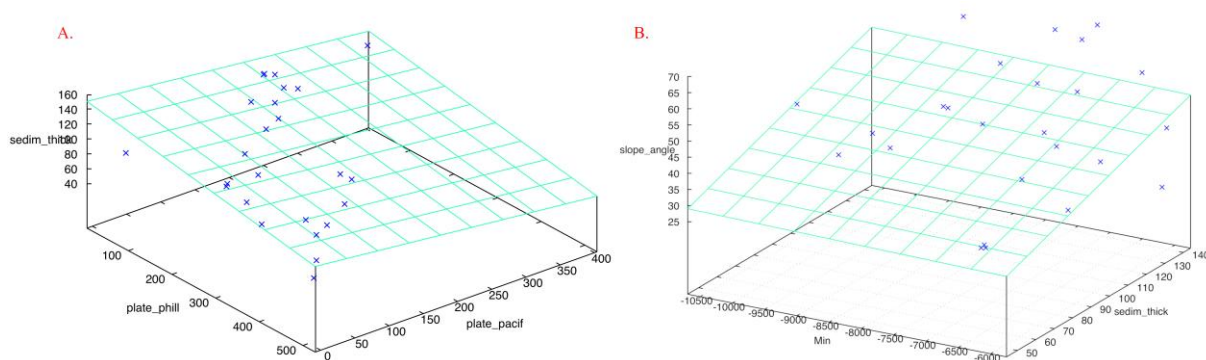

Figure 12. 3D visualization of the triple correlation between variables. Plotting: Gretl.

In this case, the geospatial effects could be the factors explaining this relationship (location in the northern and central part of the trench), Figure 11. A similar pattern, but with a higher

correlation, was observed for the slope gradient angle and sediment thickness (correlation > 0.75, Figure 13, A).
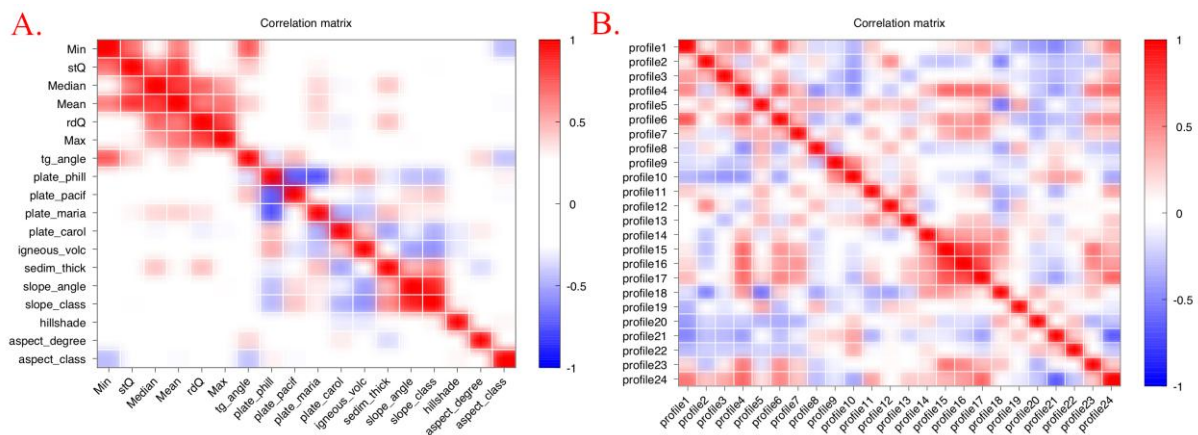

Figure 13. Correlation matrix between variables: similarities on factors (A); Correlation matrix between cross-sectioning profiles: similarities on geomorphology (B). Plot: Gretl.

4. CONCLUSION

Mariana Trench marks the location at which the Pacific plate subducts beneath the eastern edge of the Philippine Sea plate [39]. Therefore, the tectonic activities in the study area strongly affect the geomorphology of the trench. This study demonstrated quantitative and qualitative analysis of the correlation between various geospatial factors affecting Mariana Trench geomorphology by technical means of Gretl and R based data modelling. Hence, the research aim was to analyze factors affecting the variations of the Mariana Trench shape by various tectonic plates and how they correlate with each other.

Gretl based tests and data models proved to be supportive of the geospatial dependence between the factors affecting the trench geomorphology, which further reinforces geospatial data modelling interpretations made in this regard by R showing strip plots of data correlation: slope angles, maximal depths, sediment thickness and volcanism in the study area. The distributions of the depth values in the bathymetric patters across the trench show an important difference: the most prominent depth is detected in the south-western part of the Mariana Trench and decrease gradually from there until the north-eastern part of the trench. The increased values in the sediment thickness can be noted in the profile groups with the depths range -9,000 to -9,000 m and from -7,700 to -7,000 meters with a gradual decrease thereafter. Profiles with bathymetric values ranging around -6,500 meters show a contrast of

the intensity of sedimentation rates showing lesser values in the sediment thickness. A simulation model applied to the profiles grouped by the sediment thickness showed five distinct clusters showing variation of the sediment thickness according to the spatial distribution of the data samples. Overall, the bathymetric values (depths, m) have a high correlation value in the central part of the trench. Hence, the parameter of sediment thickness has a distinct correlation with geomorphic parameters of the slope shape: angle gradient, steepness and aspect, as well as distribution of the igneous rocks.

This brings up the issue of how these parameters are actually distributed in space. Gretl-simulated data modelling showed that the geomorphic, geologic and bathymetric variables are distributed unequally by the tectonic plates. Thus, the 3D visualization of the triple correlation between the geomorphic and geological variables shows dependancies between these factors and tectonics, e.g. the Pacific and the Philippine Sea plates. Virtually, the variations may be explained by the diverse geospatial conditions and geological settings in these areas. The remaining outliers in the data can be explained by the local variations in the geological parameters. The relevance of variations in multiple factors was studied in relation to the cross-sectional profiles across the trench and specifically pairwise, for the slope angle gradient with sediment thickness, igneous volcanic areas, and aspect degree, respectively. Finally, there is correlation between the slope gradient observations and aspect degree for the profiles located in the north and north-eastern part of the trench.

As demonstrated in the results, there is a distinct correlation between the volcanic activities and slope gradient angle of the trench, sediment thickness and depths in the places of the bathymetric sample points. This proves that trench motion caused by high tectonic activities in the study region directly impacts the geometry and shape geomorphology of the Mariana Trench. Furthermore, the influence of the geomorphic, geologic and bathymetric variables is assessed by the computing and measuring correlation effects, as well as visualized by graphic plotting and data modelling using Gretl and R statistical packages. The descriptive statistical analysis and correlation matrices yielded following results: sediment thickness correlate with the variations in depths and slope gradient angles, as well as changes in geographic locations moving from north southwards. Using a combination of statistical packages of R and Gretl demonstrated effective results visualized by the multifaceted plots and subplots, pairwise and triple correlation graphs, correlation matrices and other types of the graphical output.

Selecting a suitable software for statistical analysis is always important step in the research methodology. Most of all, it requires a review in choice, an argued approach comparing both pros and contras in the functionality of the particular software [40], [41], [42]. Methodologically, comparing Gretl with R, Gretl offers a more user-friendly possibilities: intuitive GUI enabling user to more smoothly perform analysis. In terms of functionality, R, comparing to Gretl, contains more packages and additional modules for diverse types of the statistical analysis that were not performed here due to the scope of this paper.

Graphical plotting is important part of the research in data analysis [43]: besides aesthetic value of the high-quality scientific plotting, effective visualization enables to highlight dependancies between variables and to put an accent on the groups variables that can be overlooked otherwise. R has powerful possibilities of graphical visualization due to such packages as {RColorBrewer}, {viridis}, {graphics}, {ggplot2}, {plotfunctions}, {ggplotgui}, {extrafont}, {fontLiberation}, {crayon}, {gridExtra}, {grid}. R enables to switch between necessary packages, adjust graphical plotting at the advanced level, create multi-plots, faceted plots, as well as enables a variety and flexibility in the statistical modelling and approaches, and many more. However, Gretl, due to the user-friendly GUI enables to focus more on the research and statistical analysis without strong skills in syntax of programming language, as required by R. To conclude, both Gretl and R are useful for the geological data analysis. Using combination of the both packages demonstrated positive results: data analysis, statistical modelling with possibilities to tune functions, select variables, options, as well as fine visualization plotting.

References

1. Cottrell, A. & Lucchetti, R. 2019. Gretl: Gnu Regression, Econometrics and Time-series Library. [Online] http://gretl.sourceforge.net/
2. Baiocchi, G. & Distaso, W. 2003. 'GRETL: Econometric software for the GNU generation', Journal of Applied Econometrics 18: 105–110.

3. Stallman, R. 1983. GNU Operating System. [Online]
https://www.gnu.org/software/software.en.html

4. R Development Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing URL: http://www.R-project.org Vienna, Austria, 2014.

5. Kleiber, C. & Zeileis, A. 2008. Applied Econometrics with R (1st ed.). Springer, New York, NY

6. Cielen, D., Meysman, A.D.B. & Ali, M. Introducing Data Science. Big Data, Machine Learning and More, Using Python Tools. Manning, Shelter Island, U.S., 2012.

7. Grus, J. 2015. Data Science from Scratch. First Principles with Python. O'Reilly.

8. Cowan, G. 1998. Statistical Data Analysis. Oxford Science Publications. Clarendon Press, Oxford, UK

9. Cuesta, H. & Kumar, S. 2016. Practical Data Analysis, 2nd Edition. A practical guide to obtaining, transforming, exploring, and analyzing data using Python, MongoDB, and Apache Spark. 360 pp. ISBN-10: 1785289713. Packt Publishing Ltd. Livery Place, Birmingham, UK.

10. Davis, J. Statistics and Data Analysis in Geology. Kansas Geological Survey John Wiley and Sons, 1990.

11. Roberts, N. M., Tikoff, B., Davis, J. R. & Stetson-Lee, T. 2018. 'The utility of statistical analysis in structural geology'. Journal of Structural Geology, 1-39. doi: 10.1016/j.jsg.2018.05.030

12. Arfken, G. B. & Weber, H.-J. Mathematical Methods for Physicists, 4th edition, Academic Press, New York, 1995.

13. Lemenkova, P. 2019. An empirical study of R applications for data analysis in marine geology. Marine Science and Technology Bulletin (2019) 8(1): 1–9. doi: 10.33714/masteb.486678

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

15. Marta-Almeida M., Ruiz-Villarreal M., Otero P., Cobas M., Peliz A., Nolasco R., Cirano M., Pereira J. 2011. OOF3: A Python engine for automating regional and coastal ocean forecasts. Environmental Modelling & Software 26, 680-682. doi: 10.1016/j.envsoft.2010.11.015

16. Lemenkova, P. 2019 Processing Oceanographic Data by Python Libraries Numpy, SciPy And Pandas. Aquatic Research 2(2), 73-9. doi: 10.3153/AR19009

17. Gardner, J.V., Armstrong, A.A., Calder, B.R., Beaudoin, J., 2014. So, how deep is the Mariana Trench? Marine Geodesy, 37, 1–13. http://dx.doi.org/10.1080/01490419.2013.837849.

18. Chadwick, W. Jr., Merle, S. G., Baker, E. T., Walker, S. L., Resing, J. A., Butterfield, D. A., Anderson, M. O., Baumberger, T., and Bobbitt, A. M. 2018. A Recent Volcanic Eruption Discovered on the Central Mariana Back-Arc Spreading Center. Frontiers in Earth Science, 6, 1-16, doi: 10.3389/feart.2018.00172

19. Faccenna C., Holt, A. F., Becker, Th. W., Lallemand, S. Royden, L. H. 2018. Dynamics of the Ryukyu/Izu-Bonin-Marianas double subduction system. Tectonophysics, 229-238, doi: 10.1016/j.tecto.2017.08.011

20. Fujioka, K. and Okino, K. and Kanamatsu, T. and Ohara, Y. 2002. Morphology and origin of the Challenger Deep in the Southern Mariana Trench. Geophysical Research Letters, 29 (10), 1372, doi: 10.1029/2001gl013595

21. Hilst R., van der and Seno, T. 1993. Effects of relative plate motion on the deep structure and penetration depth of slabs below the Izu-Bonin and Mariana island arcs. Earth and Planetary Science Letters, 120, 395-407, doi: 10.1016/0012-821X(93)90253-6

22. Hussong, D. M. and Uyeda, S. 1982. Tectonic process and the history of the Mariana Arc: A synthesis of the results of deep sea drilling project leg 60. Initial Reports of the Deep Sea Drilling Project, 60, 909-929.

23. Taira, K., Kitagawa, S., Yamashiro, T., Yanagimoto, D., 2004. Deep and bottom currents in the Challenger Deep, Mariana Trench, measured with super-deep current meters. Journal of Oceanography, 60 (6), 919–926.

24. Haren, H. van, Berndt, Ch., Klaucke, I. 2017. Ocean mixing in deep-sea trenches: New insights from the Challenger Deep, Mariana Trench. Deep-Sea Research Part I: Oceanographic Research Papers, doi: 10.1016/j.dsr.2017.09.003

25. Linley, T. D., Stewart, A. L., McMillan, P. J., Clark, M. R., Gerringer, M. E., Drazen, J. C., Fujii, T., Jamieson, A. J. 2017. Bait attending fishes of the abyssal zone and hadal boundary: Community structure, functional groups and species distribution in the Kermadec, New Hebrides and Mariana trenches. Deep-Sea Research Part I: Oceanographic Research Papers, 121, 38-53, doi: 10.1016/j.dsr.2016.12.009

26. Luo, M., Gieskes, J., Chen, L., Shi, X., Chen, D. 2017. Provenances, distribution, and accumulation of organic matter in the southern Mariana Trench rim and slope: Implication for carbon cycle and burial in hadal trenches. Marine Geology, 386, 98-106, doi: 10.1016/j.margeo.2017.02.012

27. Nakanishi, M., Hashimoto, J. 2011. A precise bathymetric map of the world's deepest seafloor, Challenger Deep in the Mariana Trench. Marine Geophysical Research, 32, 455-463, doi: 10.1007/s11001-011-9134-0.

28. Cameron, A. C. & Trivedi, P. K. 2013. Regression Analysis of Count Data, 2nd edition, Econometric Society Monograph 53, Cambridge University Press, 566 pp.

29. Bera, A. K., Jarque, C. M. & Lee, L. F. 1984. 'Testing the normality assumption in limited dependent variable models', International Economic Review 25: 563–578.

30. Bulmer, M. G. Principles of statistics. Dover Publications, New York, 1979.

31. Brownlee, K. A. Statistical theory and methodology in science and engineering. John Wiley & Sons, New York, 2nd edition, 1965.

32. Arfken, G. B. & Weber, H.-J. Mathematical Methods for Physicists, 4th edition, Academic Press, New York, 1995.

33. Cleveland, W. S. 1979. 'Robust locally weighted regression and smoothing scatterplots', Journal of the American Statistical Association 74(368): 829–836.

34. Bhargava, A., Franzini, L. & Narendranathan, W. 1982. 'Serial correlation and the fixed effects model', Review of Economic Studies 49: 533–549.

35. Sa, J. P. M. de. 2007. Applied Statistics Using SPSS, Statistics, Matlab and R. Springer, 520 pp. 978-3-540-71971-7

36. Lemenkova, P. 2018. R scripting libraries for comparative analysis of the correlation methods to identify factors affecting Mariana Trench formation. Journal of Marine Technolology and Environment, 2, 35-42. arXiv: 1812.01099

37. Everitt, B. S. 2002. The Cambridge Dictionary of Statistics. Cambridge, UK. 293 pp. Doi: 10.1016/j.geoderma.2003.11.001

38. Borradaile, G. J. 2003. Statistics of Earth Science Data. Springer, 79 pp.

39. Michibayashi, K., Tasaka, M., Ohara, Y., Ishii, T., Okamoto, A., Fryer, P. 2007. Variable microstructure of peridotite samples from the southern Mariana Trench: Evidence of a complex tectonic evolution, Tectonophysics, 444, 111-118, doi: 10.1016/j.tecto.2007.08.010

40. Yalta, A. T. & Yalta, A. Y. 2010. Should Economists Use Open Source Software for Doing Research? Computational Economics, 35 (4), 371–394. doi: 10.1007/s10614-010-9204-4

41. Altman, M. & McDonald M. P. 2001. Choosing reliable scientific software. PS: Political Science and Politics 34: 681–687

42. McCullough, B. 1998. Assessing the reliability of statistical software. The American Statistician 52: 358–366.

43. Wilkinson, L. The Grammar of Graphics. Statistics and Computing, 2nd edition. Springer-Verlag, 2005.