Some statistical consideration of azimuth and inclination angles determination based on walk - away VSP data in Python

Mateusz Zaręba[1,3,*], *Tomasz* Danek[1], *Michal* Stefaniuk[2]

[1]AGH University of Science and Technology, Department of Geoinformatics and Applied Computer Science, Krakow, Poland
[2]AGH University of Science and Technology, Department of Fossil Fuels, Krakow, Poland
[3]GK Processing, R&D Department of Seismic Signal Processing Center, Balice, Poland

Abstract. It is a common knowledge that proper inclinations and azimuth angle determination is a critical step in processing and interpretation of walk-away VSP data. Additionally, an in-depth analysis of the uncertainty of these interpreted values requires the introduction of measurement errors. In this contribution, we present a statistical analysis of obtained polarization angles from three-component, multi-depth level, walk-away VSP using Python 3 programming language. Our analysis is presented in the context of different processing sequences and correlation with local features of the geological medium. We show that the obtained values of polarization angles and their errors can be strongly affected by processing sequence and - when done correctly - can give addition inside into features of analysis medium. Moreover, in some cases, even a presence of saturation can be express by polarization angles variations. Additionally, we examined the impact of well-casing on interpretational values of polarization angles.

Keywords: statistical analysis, seismic anisotropy, exploratory data analysis, VSP seismic method, data science, machine learning

1 Introduction

Obtaining full anisotropy tensor using P-wave only from the walk-away VSP (Vertical Seismic Profiling) survey is not a trivial task, especially when the acquisition was gathered in challenging conditions [1]. The methodology for P-wave only inversion for local anisotropy estimation has been proposed by Grechka and Mateeva [2]. Statistical analysis including EDA (Exploratory Data Analysis) of VSP data and obtained inclinations and azimuths are crucial steps and are not less important than proper processing of data. The purpose of this step was to find the patterns in data and prepare it for cluster analysis to decide which samples correspond to particular layers differing in acoustic properties. Additionally, it was important to find the correlation between geological layers (based on core analysis) and calculated polarization angles. The difficulty of P-wave only inversion method is into finding minimum of target function using optimization algorithms. On its input there have to

---

[*] Corresponding author: mattzareba@gmail.com

be included the inclinations for various offsets with errors of its estimations, velocities obtained from VSP, and velocities of formations from other logs. The analysis presented in this paper allowed for proper preparation of data for the P-wave only inversion. Beside difficulty of method itself, the acquisition was performed in difficult terrain after heavy rains and half of the measurements were carried out in uncased part of the well. Due to heavy rains, ground conditions between each sweep were varying significantly [1]. The main aim of this research was to find the quantitive information about which processing scheme is the best for obtaining reliable polarization angle using PCA (Principal Component Analysis) method for components rotation. Additionally, the patterns in the data were studied for further data clustering using unsupervised machine learning.

2 Walkaway VSP acquisition and region characterization

Data were gathered in Wysin village located in Northern Poland. It is a part of the Baltic-Central Russian Depression Zone tectonic unit called the Peri-Baltic Syneclise. There were three wells drilled – one vertical (Wysin 1) and two horizontal ones (Wysin 2H and Wysin 3H) – see Fig. 1. There were 480 shot points along 12200 meters profile (parallel to horizontal wells). Wysin 1 well is placed on the center of 3D surface seismic survey. There were 96-receiver 3C BSR Array System (Oyo Geospace Company) with spacing between receivers 15 m. This system allows for simultaneous recording of zone from 2400 to 3825 m MDGL (measured depth from ground level). It allowed for analysis of 11 lithological complexes of Silurian rocks (determined from well-logs) which corresponds to 4 macro-complexes separated on the base of velocity analysis for depth migration.
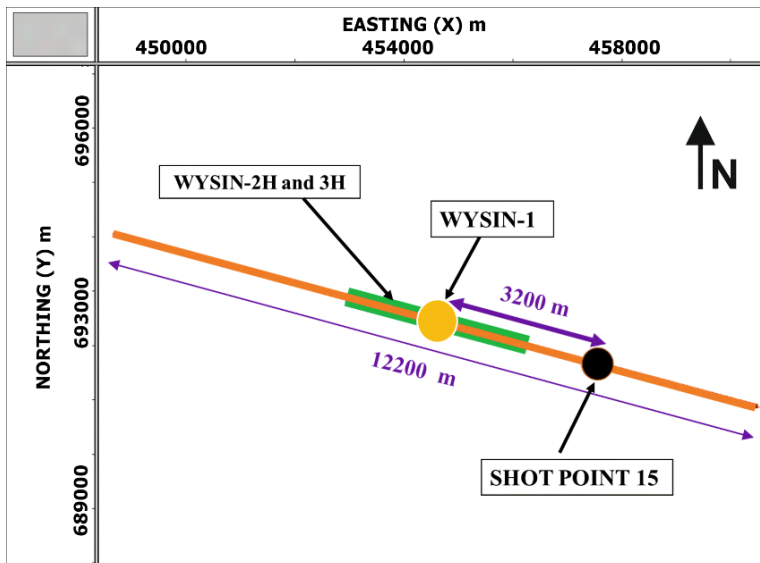
Fig. 1. Scheme of walkaway VSP acquisition [1].

The whole recording system was placed below high-attenuating Zechstein formation [1]. In the case of litology, they can be classified as anhydrite, halite, shale, and dolomite. There were up to 15 sweeps on each of 480 shot points performed in frequency range from 6 to 140 Hz.

3 Processing and rotations

In this section, we introduce basic information about seismic processing of presented data in this paper just to cover the minimum to understand the aim of presented statistical analysis. The wide description of each procedure presented in this paper: pre-processing, rotation methods, and matching filter creation (including graphical illustrations) can be found in [1].

There were 4 different options of processing tested to choose which one is the best for polarization angles determination: Option 1 - raw data rotation, then vertical stacking for each shot independently, next vertical stacking with noise removal after it. Option 2 – vertical stacked data rotation, then noise removal. Option 3 – de-noising then rotation for each shot separately, then vertical stacking. Option 4 - rotation is done for each shot after noise attenuation and vertical stacking of un-rotated data. Then average error of inclination estimation is performed for whole receivers range (R_1-R_N) in a group of five (N=5) and was calculated according to equation 1:

$$\sigma(R) = \sqrt{\frac{\sum_i^N (R_i - \overline{R})^2}{N}} \tag{1}$$

Errors for the whole measured depth range for every shot point (SP) were estimated according to equation 2:

$$\sigma(SP) = \frac{\sum_i^{N_E} \sigma(R_i)}{N_E} \tag{2}$$

where N_E – number of calculated angles with, $\sigma(R_i) < 15^0$ for azimuths and $\sigma(R_i) < 5^0$ for inclinations, and $\sigma(SP)$ – azimuth or inclination average error for each SP.

After choosing Option 4 as the best in case of polarization angles determination, a signal matching filter (SM) was designed (based on similarities between average sweep and each sweep on each shot point). The values of errors for inclinations and azimuths according to equation 2 are shown in Fig. 2 (for Shot Point 15 - offset 3200 m).
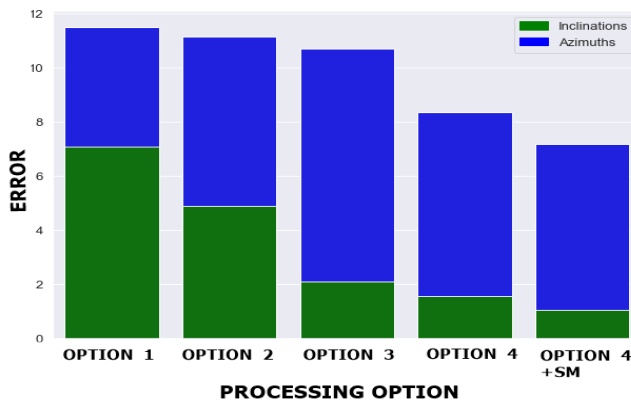

Fig. 2. Errors of polarization angles determination for Shot Point 15.

4 Statistical analysis and EDA of walkaway VSP data

The statistical analysis of polarization angles distribution and patter analysis was performed using Python programming language [3]. It is efficient and easy to use high-level programming language. There are libraries for SEG-Y files read (Segpy), framework for Seismology (ObsPy) and the possibility of large-scale processing of Adaptable Seismic Data Format (ASDF). It is also easy to import data from Excel format using Pandas package and from SQL database. It allows for multifactorial, simultaneous analysis of data from different formats (including seismic ones) and efficient visualization. It also gives opportunity to perform machine learning using all data together. In this paper the Seaborn Python library (based on matplotlib) [4] was used for visualizations (Fig. 2 to Fig. 9)

The pattern analysis and comparison of descriptive statistics between each processing option, input data, and previously option 4 with signal matching filter (SM) applied were done. For this purpose, the box plots and swarm plots were calculated (Fig. 3). Box plot allows for similar analysis like histogram or kernel density function but in simplified version. The center of a box is median, the top and bottom represent 3rd quartile and 1st quartile. Whiskers represent the minimum and maximum. The bee swarm plot shows each sample and helps to understand the structure of data. It helps to gain inside into distribution of data. It is easy to see that inclinations obtained from option 1 are three-modal, where one mode is placed far from median and close to minimum. For Option 2, the distribution is similar. However, the average value of inclination is lower, and median is closer to 1st quartile which is opposite to option 1. Option 3 and 4 have similar average inclination value, however values are more cumulated around median than in case of option 1 and 2. The most consistent distribution with no outliner values was obtained for option 4, which has a lower average estimation error. Applying matching filter for data in option 4 (Option_4 with SM in Fig. 3) has the narrowest interquartile range and a slightly lower average inclination value.
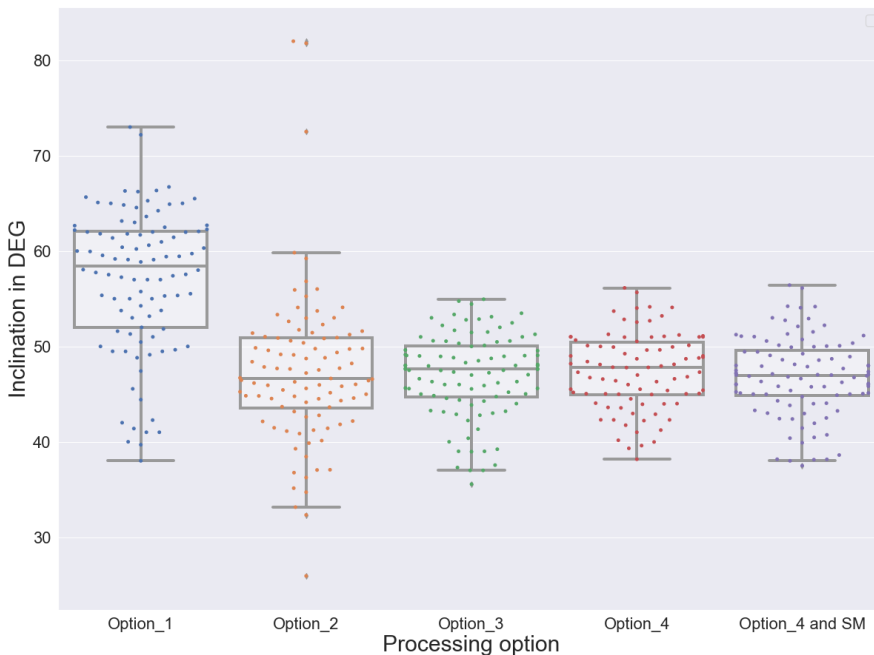
Fig. 3. Box plot and swarm plot for inclinations of different processing schemes.

Then the kernel density estimation using the Gaussian kernel method (KDE) has been calculated. The results are shown in Fig. 4. The blue graphs show the KDE of input data, option1 – 4, and option 4 with SM inclinations according to depth, and the light-brown graphs show the same data according to Option 4 with SM. It is clearly visible that data has three or four main concentrations. They can be correlated with 4 main velocities complexes shown in [1]. The worst option of processing (in case of polarization angles determination) is option 2 which makes KDE flat. The second analysis light-brown KDEs was done to answer which processing scheme results are the most similar for those obtained using option 4 with SM. It can be noticed that the answer is option 3 (Pearson correlation between results = 0.78) and 4 (Pearson correlation between results = 0.95), however for Option 3 the center of kernel is placed up-right from center and is asymmetric. The similarity is bigger for lower inclination values.

Not only proper processing of seismic data is critical, but also the elimination of bad data from input. In this case, we understand bad data as values anomalously higher or lower than surrounding values. It can be caused by wrong anchorage of receiver in well, signal transmission problem, measurement device errors or other random problems during acquisition in extreme conditions of deep well. In Fig. 5. histogram for inclinations before (blue) and after (orange) bad values elimination, with KDE plot and fitted Gauss distribution line is shown. After this process the distribution is more concentrated about central value, additionally, the modes of data distribution are stronger visible. The distributions for different Shot Points are similar in shape. However for Shot Point nr 5 and 14 the distribution is slightly different.

The last step was to plot the bee swarm plot before (Fig. 6) and after (Fig. 7) bad values removal. It can be seen that after this procedure the outline values are removed and then, the distribution analysis can be done with higher accuracy. There is an unexpected rapid change between Shot Point nr 5 and Shot Point nr 6. We expected linear trend in obtained swarms. Similar behavior can be noticed in Shot Point nr 15. Thus, the data can be divided into three different groups according to data distribution – first: from Shot Point 1 to 5, second from 8 to 14, and third from 15 to 19. These three main group data will be used for separate calculation of P-wave only inversion in the future.

5 Conclusion

Presently, the detailed investigation of seismic anisotropy for shale gas exploration is extremely important [1]. The use of new tools, which are not included in seismic processing software gives a chance to make the correct decision for a data processing scheme. The power of data analysis and statical examination with proper visualization of data distribution should not be underestimated when a detailed investigation is performed. This step is even more important when application of machine learning algorithms is planned. P-wave only inversion is not a trivial problem to solve, especially when the signal is recorded in presence of high attenuating rocks above receivers and in uncased well. The use of Python for statistical analysis and visualization allowed for fast and accurate examination of data from different data types.
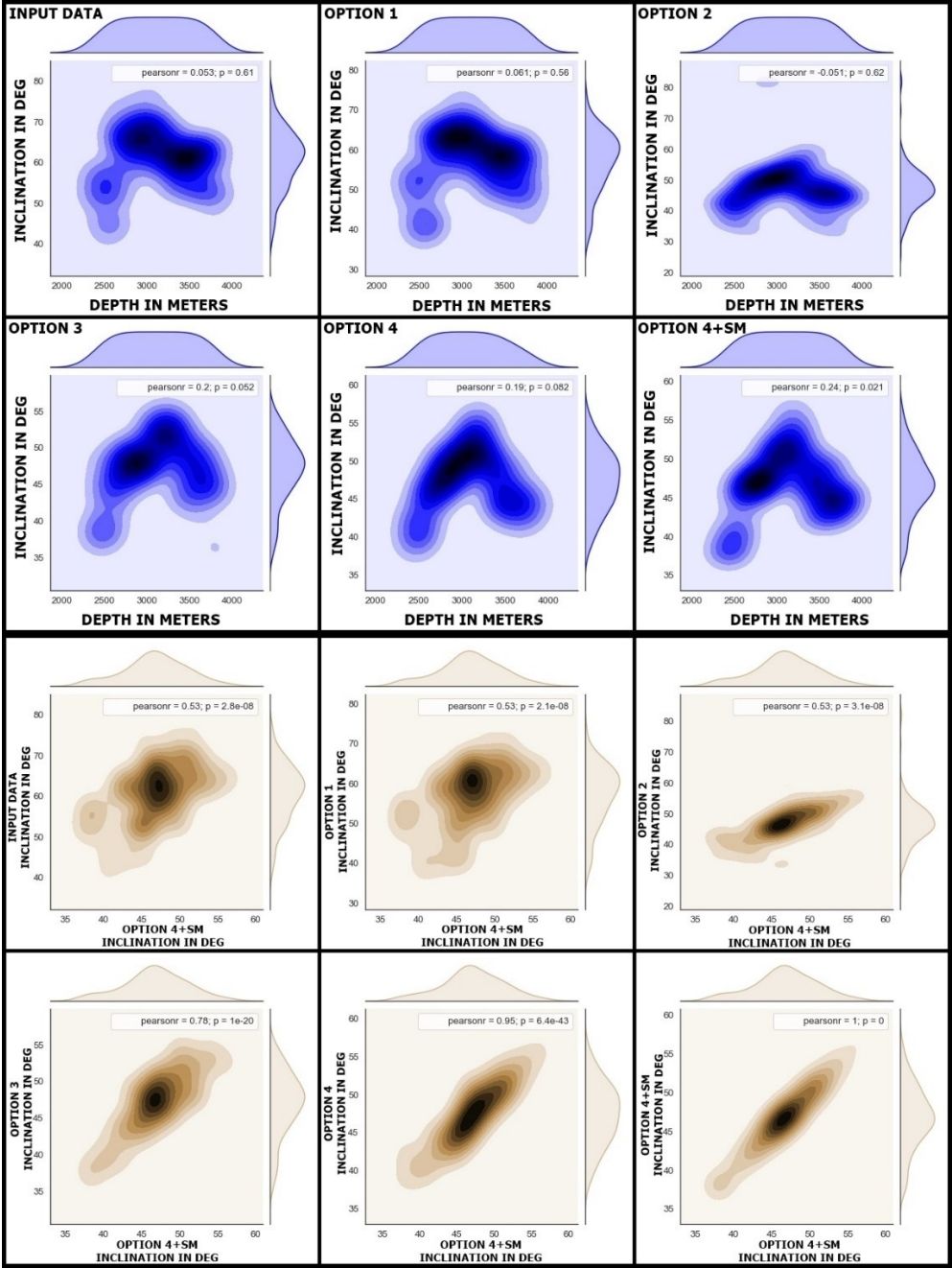
Fig. 4. Kernel Density Estimation (KDE) for different processing options.
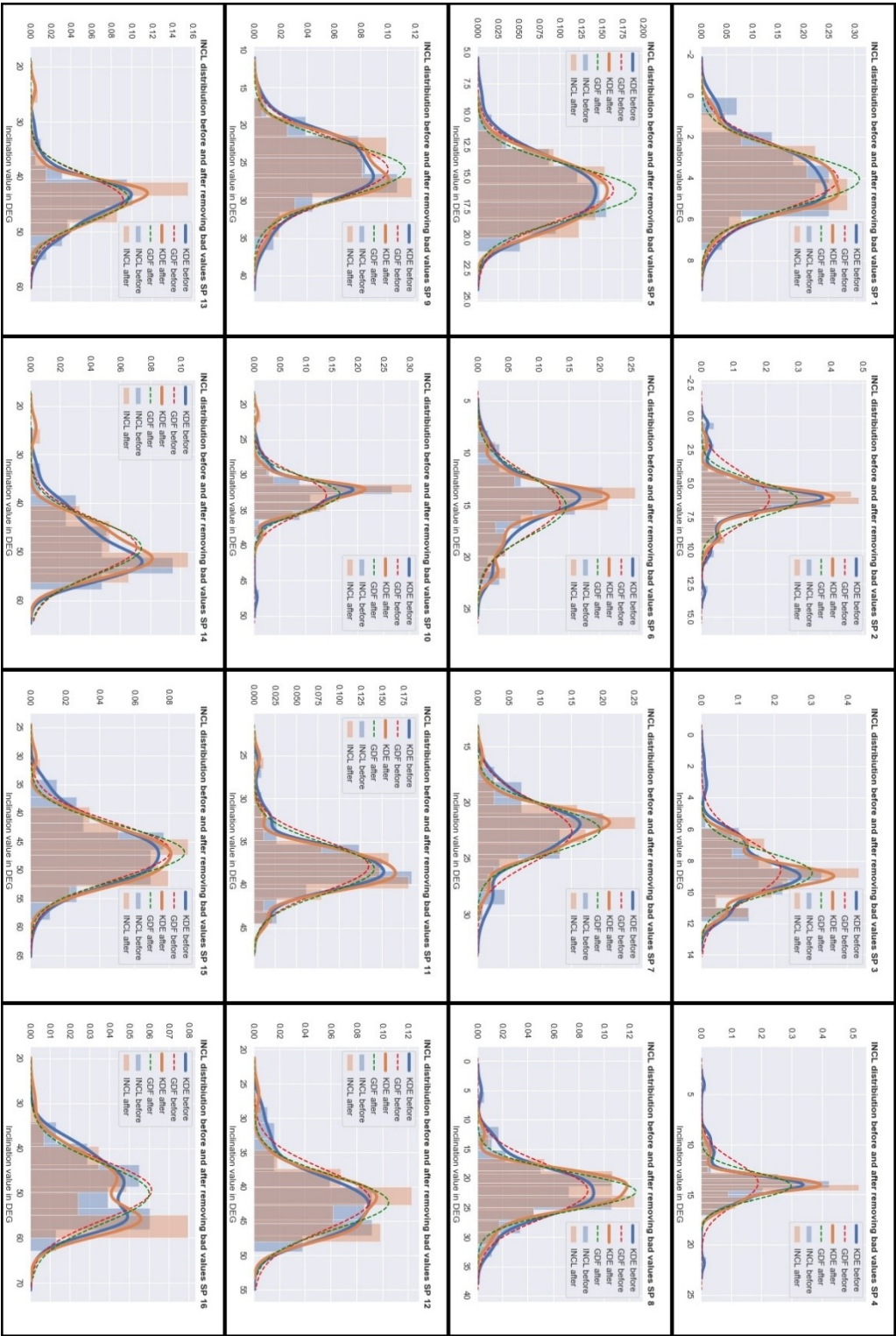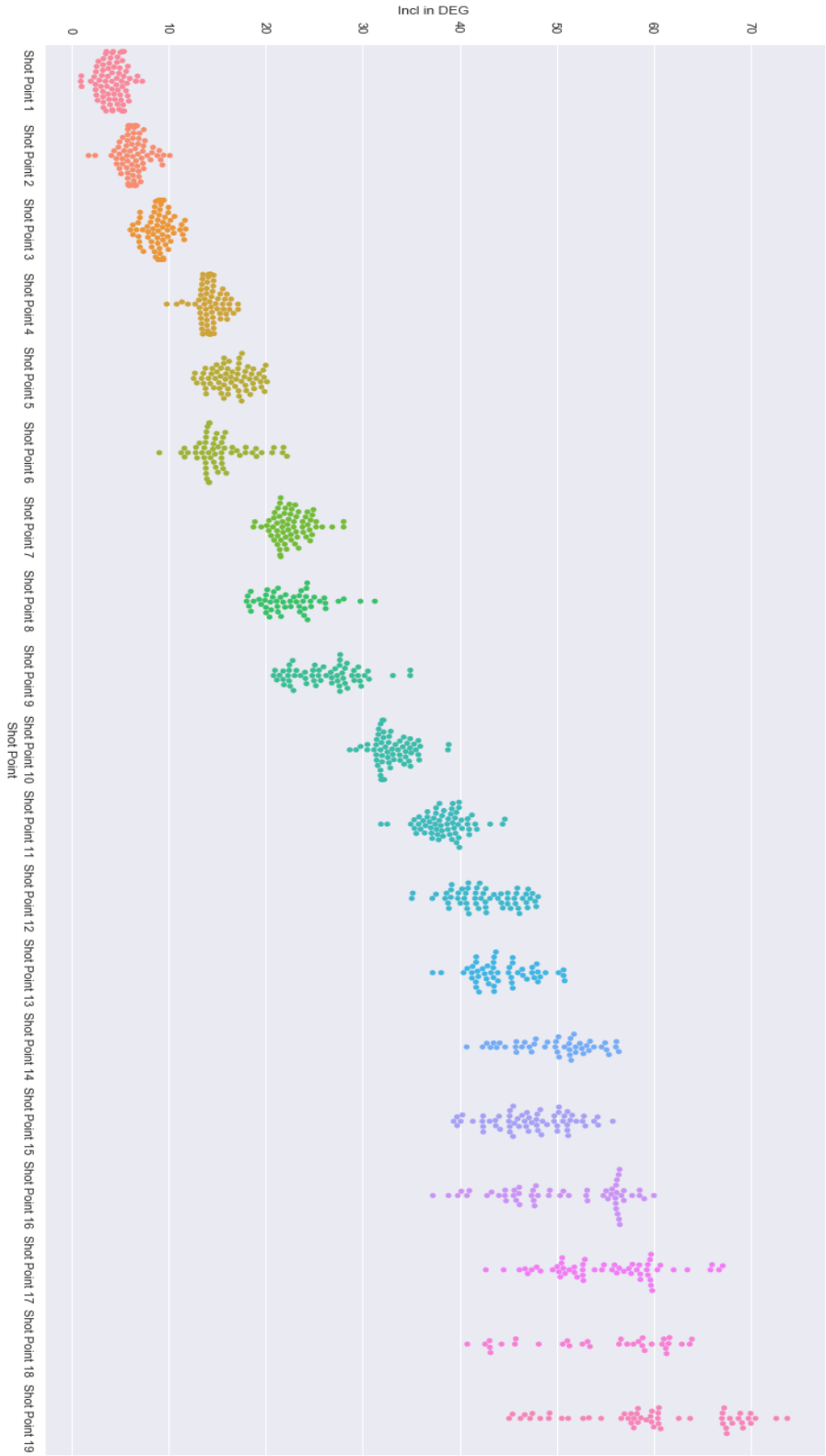
Fig. 5. The statistical distribution of inclinations for different shot points.

Fig. 6. Data distribution for different shot points before bad values elimination.

Fig. 7. Data distribution for different shot points after bad values elimination.

References

1. M. Zaręba, T. Danek, ACTA GEOPHYS, (2018) doi: 10.1007/s11600-018-0200-8

2. V. Grechka, A. Mateeva, GEOPHYSICS, **72**, 4, (2007) doi: 10.1190/1.2742970

3. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C. R Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python, ArXiv (2019)

4. J. D. Hunter, COMPUT SCI ENG, **9**, 3, (2007) doi: 10.1109/MCSE.2007.55