International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:13, No:11, 2019

The Quality Assessment of Seismic Reflection Survey Data Using Statistical Analysis: A Case Study of Fort Abbas Area, Cholistan Desert, Pakistan

U. Waqas, M. F. Ahmed, A. Mehmood, M. A. Rashid

***Abstract*—**In geophysical exploration surveys, the quality of acquired data holds significant importance before executing the data processing and interpretation phases. In this study, 2D seismic reflection survey data of Fort Abbas area, Cholistan Desert, Pakistan was taken as test case in order to assess its quality on statistical bases by using normalized root mean square error (NRMSE), Cronbach's alpha test (α) and null hypothesis tests (t-test and F-test). The analysis challenged the quality of the acquired data and highlighted the significant errors in the acquired database. It is proven that the study area is plain, tectonically least affected and rich in oil and gas reserves. However, subsurface 3D modeling and contouring by using acquired database revealed high degrees of structural complexities and intense folding. The NRMSE had highest percentage of residuals between the estimated and predicted cases. The outcomes of hypothesis testing also proved the biasness and erraticness of the acquired database. Low estimated value of alpha (α) in Cronbach's alpha test confirmed poor reliability of acquired database. A very low quality of acquired database needs excessive static correction or in some cases, reacquisition of data is also suggested which is most of the time not feasible on economic grounds. The outcomes of this study could be used to assess the quality of large databases and to further utilize as a guideline to establish database quality assessment models to make much more informed decisions in hydrocarbon exploration field.

***Keywords*—**Data quality, null hypothesis, seismic lines, seismic reflection survey.

I. Introduction

THE exploration work to discover the hydrocarbons and related derivatives is very crucial phase for a country's economy. A number of well-established geophysical techniques are available to explore and estimate subsurface oil and gas reserves. Mostly 2D and 3D seismic reflection surveys are conducted to evaluate the hydrocarbons potential of any subsurface oil and gas reservoir. The selection of the suitable location for drilling an exploratory well is highly dependent upon the quality of the acquired geophysical survey data prior to processing and interpretation. Errors in database significantly affect its quality and decision making power, which may lead to unfavorable outcomes.

Data quality can be defined as "the extent to which a set of collected facts and figures fulfill the requirements or fitness for use" [1], [2]. A number of past investigators developed different categories to assess and investigate database quality including; consistency, accuracy, information to noise ratio, availability, cohesiveness etc. [3]-[5]. Low quality input data makes output data erratic and falsified, and it is not economically favorable and increases operational cost, reduces decision-making power and decreases the potential to execute strategies [6]-[8].

Quality assessment of acquired data before processing and interpretation is a mandatory step. Karr et al. [9] focused on preliminary filtering, significance of exploratory data analysis and modification of data quality. They explained that the improper management and utilization of biased data can cause significant economical inefficiencies because data itself is nothing but its proper quality assessment makes it good or bad. High quality data acquisition practice, directly related to level of skill and experience of the data acquisition crew, performance of data recording instrument, environmental conditions and specificity of geophysical surveys.

To acquire high quality data, proper management and understanding of field parameters design, seismic line geometry and seismic reflection survey layout is utmost important. Talagapu [10] explained that seismic lines geometry by accepting general concepts for its location, direction and orientation. Dip lines are better than strike lines. New Lines might be stacked with nearby existing lines or data. Field parameters are designed first before the data acquisition. Survey time, economics, energy source, geophone type etc. are considered as designed field parameters.

High quality data reduces processing and interpretation time. Eppelbaum and Khesin [11] expressed that the selection of geophysical survey is related to ground and environmental conditions. For example, the execution of resistivity survey in an arid environment is not suitable because of very high resistance and low conductance. Sometimes the saline water is sprinkled to enhance the conductivity of the ground. Similarly, GPR (Ground Penetration Radar) survey in tropical environment is not adequate, penetration of radar waves cease in the presence of moisture.

In exploration geophysics quality of acquired database adversely affected by number of factors including subsurface geological complexities and variations, nearby ground vibrations and noises, non-specificity of geophysical survey with environmental conditions etc. The signal to noise ratio (S/N) is another parameter that adversely affects the quality of

U. Waqas is with the Department of Geological Engineering, University of Engineering and Technology, Lahore, Pakistan (corresponding author, phone: +92-322-4446487; e-mail: umerwaqas@uet.edu.pk).

M.F. Ahmed, A. Mehmood and M.A. Rashid are with the Department of Geological Engineering, University of Engineering and Technology, Lahore, Pakistan (e-mail: mfanr5@mst.edu, ahsange07@hotmail.com and awaisgeo@yahoo.com).

acquired data. Chetia and Dimri [12] showed that seismic noises such as ground rolls, guided waves, multiples and shallow refractions are consistent. In some cases, magnitude of theses noises might be stronger than seismic signals. Bigger and deeper charges are used to damp local ambient and source generated noises respectively.

This study only deals with the assessment of quality of 2D seismic reflection survey data by using different statistical tools prior to processing and interpretation stages. An erratic data needs too much static correction to make it interpretable. Statistical operations and algorithms associated with data recording devices can be used to filter out unwanted signals to improve the quality of acquired data at the spot. The outcomes of this study could be used to analyze the reliability of database acquired from geophysical surveys for planning and designing phase in site investigation, exploration and estimation of hydrocarbon reserves and for fluctuating ground water table in hydrogeological studies.

II. Materials and Methods

A. Study Area

Fort Abbas Area of Cholistan Desert, Punjab, Pakistan, is a historical city, which is located at latitude (29.19 degrees) 29° 11′ 24″ north of the Equator and longitude (72.85 degrees) 72° 50′ 59″ east of the Prime Meridian, which is situated at Indo-Pak border as shown in Fig. 1. It is the part of Punjab Platform, which has no outcrop of sedimentary rocks while dipping towards Sulaiman Depression as a huge monocline [13]. It is far away from Indian and Eurasian plate's convergence boundary that makes it tectonically least affected [14]. Various geologic formations including Khewra, Patala and Namal are encountered in this area. Khewra sandstone and Patala shale are proven reservoir and source rocks of Punjab Platform respectively [15].

Fig. 1 Location of study area on Google map

B. Data Acquisition

Oil and Gas Development Company Limited (OGDCL) conducted the 2D seismic reflection survey in 1993 to explore and estimate oil and gas reserves in study area. Several seismic lines were spread of specific length, orientation and geometry. Seismic waves were generated by explosive shot in a borehole. These waves traveled through subsurface materials, reflect, refract and scatter when encounter with a material having different densities. Geophones at the surface having specific distance from the shot point, received reflected waves, convert them into electrical pulses and record their travel time and velocity data at the recording instrument. A strong reflection of seismic waves occurs at those points where sharp density contrast between the materials was existed. To assess the quality of acquired database only 15 seismic lines were sampled.

Robinson and Coruh [16] described a method to measure the depth of reflectors and average velocity of seismic waves as follows:

- Obtain travel time of seismic waves T_1, T_2, T_3,.........,T_n recorded by each geophone located at a specific distance X_1, X_2, X_3,.......,X_n from shot point.
- Square these values to get T_1^2, T_2^2, T_3^2,.........,T_n^2 and X_1^2, X_2^2, X_3^2,.......,X_n^2 and plot the results on T^2-X^2 graph as shown in Fig. 2.
- Join the points by drawing a straight line on T^2-X^2 graph and find its slope by first order linear equation as demonstrated in Table I. Inverse of this slope gives the value of velocity.

C. Data Quality and Reliability Indices

Root mean square error (RMSE) and NRMSE were used to assess the quality of acquired dataset. The RMSE is frequently used to integrate residuals between observed and predicted data into single predictive value. The NRMSE is the dimensionless form of RMSE and resemble with coefficient of variation (C_v) as shown in (1) and (2).

$$\mathrm{RMSE} = \sqrt{\frac{\sum (X_{\mathrm{obs}} - X_{\mathrm{pre}})^2}{n}} \tag{1}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\overline{X_{\mathrm{obs}}}} \tag{2}$$

where: $X_{\mathrm{obs}}$is the experimentally determined values, X_{pre} is the predicted value and $\overline{X_{\mathrm{obs}}}$is the average experimentally determined value.

Large errors produce larger effect on the database quality and impact of errors increases by squaring the parameters of an erratic data [17]. The Cronbash's alpha test is conducted to measure reliability or internal consistency of acquired data. It evaluates the accuracy of a test that measures the variables of interest. Larger value of α conform excellent internal consistency and strong correlation among the variables of data. The value of α varies from -∞ to 1 but only positive values make sense [18]. It can be expressed as follows:

$$\alpha = \frac{N.\bar{C}}{\bar{V} + (N-1).\bar{C}} \tag{3}$$

where: N is the number of items, C¯is an average covariance and V¯is an average variance

D.Null Hypothesis Test

After assessing, the quality of acquired database by data quality index, F-test and t-test were performed to further investigate data reliability. In hypothesis, testing conditions were set as follows:

Null Hypothesis

Mean Difference = 0 (Quality of acquired data is good to give meaningful interpretation)

Alternate Hypothesis

Mean Difference ≠ 0 (Quality of acquired data is erratic and falsified and not adequate to give meaningful interpretation). If F-sig > α (α = 0.05) null hypothesis would be accepted.

If F-sig < α alternate hypothesis would be accepted

After the application of F-test, database quality is further evaluated by applying t-test as follows:

If t-stats > t-critical OR t-sig < α (α = 0.05) alternate hypothesis would be accepted and vice versa.

III. Results and Discussions

In this study, 2D seismic reflection survey was performed by spreading several seismic lines of specific orientation and geometry at Fort Abbas Area of Cholistan Desert, Pakistan. An acquisition of high quality data in seismic reflection surveys ultimately reduce processing and interpretation time and enhance the chances to explore productive well.

TABLE I
First Order Linear Equations For Selected Seismic Lines Generated By T^2-X^2 Plot

Seismic Lines	Equations
931-FAB-16A (Khewra)*	$T^2 = -0.09\ X^2 + 433552$
931-FAB-33 (Khewra)	$T^2 = -0.11\ X^2 + 397993$
931-FAB-34 (Khewra)	$T^2 = -0.52\ X^2 + 478592$
931-FAB-35 (Khewra)	$T^2 = -0.34\ X^2 + 628181$
931-FAB-37 (Khewra)	$T^2 = 0.33\ X^2 + 367052$
931-FAB-16A (Patala)*	$T^2 = -0.05\ X^2 + 195397$
931-FAB-33 (Patala)	$T^2 = -0.02\ X^2 + 149836$
931-FAB-34 (Patala)	$T^2 = 0.01\ X^2 + 153175$
931-FAB-35 (Patala)	$T^2 = -0.03\ X^2 + 162049$
931-FAB-37 (Patala)	$T^2 = -0.02\ X^2 + 135650$
931-FAB-16A (Namal)*	$T^2 = -0.07\ X^2 + 283844$
931-FAB-33 (Namal)	$T^2 = -0.04\ X^2 + 230334$
931-FAB-34 (Namal)	$T^2 = -0.09\ X^2 + 268268$
931-FAB-35 (Namal)	$T^2 = 0.31\ X^2 + 201312$
931-FAB-37 (Namal)	$T^2 = -0.03\ X^2 + 188039$

*Khewra, Patala and Namal are the geological formations

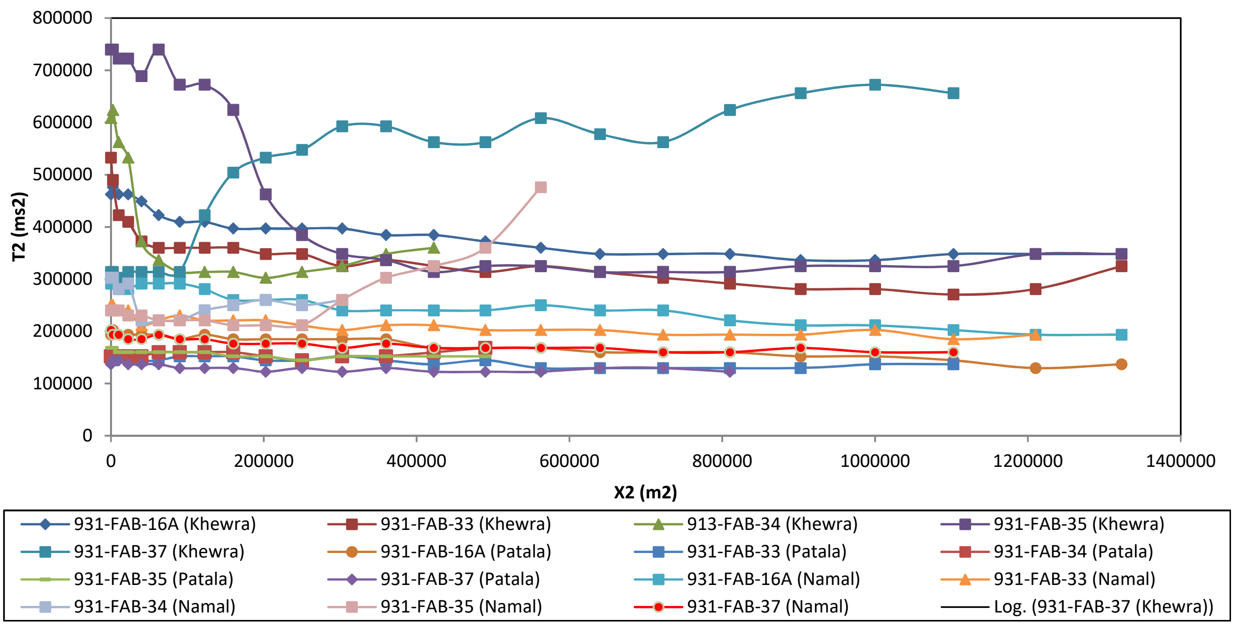

Fig. 2 T^2-X^2 plot for selected seismic lines spread over Khewra, Patala and Namal formations

A database of selected seismic lines was used to assess its reliability and quality that used to model subsurface geological conditions. At each geophone, depth of reflector was determined by simply taking the product of seismic wave velocity and travel time. Data were loaded into surfer software version 13 and by using Kriging interpolation method 3D wireframe and contour maps were generated (see Figs. 3 and 4). Both maps demonstrated that the subsurface geological strata had complexities and intense folding. However, previous seismic reflection survey results and literature review conformed that the area is flat, tectonically least affected and rich in oil and gas reserves [14], [15]. Diversity in the results clearly conform the low quality of acquired database.

TABLE II
NRMSE VALUES FOR VELOCITY DATA OF SELECTED SEISMIC LINES

Seismic Lines	NRMSE
931-FAB-16A (Khewra)*	0.57
931-FAB-33 (Khewra)	0.42
931-FAB-34 (Khewra)	0.37
931-FAB-35 (Khewra)	0.19
931-FAB-37 (Khewra)	0.20
931-FAB-16A (Patala)*	1.35
931-FAB-33 (Patala)	2.70
931-FAB-34 (Patala)	3.59
931-FAB-35 (Patala)	2.01
931-FAB-37 (Patala)	2.79
931-FAB-16A (Namal)*	0.77
931-FAB-33 (Namal)	1.48
931-FAB-34 (Namal)	0.58
931-FAB-35 (Namal)	0.08
931-FAB-37 (Namal)	1.77

*Khewra, Patala and Namal are the geological formations.

NRMSE resembles with coefficient of variation (Cv) in which standard deviation is replaced by RMSE. Low value of NRMSE indicates less variance in residuals between observed and predicted data. Quality and reliability of acquired database can easily be judged by measuring NRMSE values. In this case, NRMSE values were determined between seismic waves velocities recorded by instrument and measured by T2-X2 plots to assess the acquired database quality as shown in Table II. The seismic lines spread over Patala formation including, 931-FAB-16A, 931-FAB-33, 931-FAB-34, 931-FAB-35, and 931-FAB-37 had the highest values of NRMSE among all seismic lines as describe in Table II. The seismic lines 931-FAB-33 and 931-FAB-37 over Namal formation were found with their larger value of NRMSE (see Table II). Lowest values of NRMSE were found for the seismic line spread over Khewra formation. All seismic lines having larger NRMSE values could be used for processing and interpretation after some static correction and treatment.

To check the internal consistency or reliability of data Cronbach's alpha test was conducted. Its value varies from -∞ to 1 however, positive value makes sense to interpret reliability of data. Alpha (α) value close to 1 indicates a strong correlation and less residual variance among the variables. It is commonly observed that the value of α increases with increase in number of items of data, unlike sample having narrow range of items deflate it [19]. Past researchers made a rule for α value to interpret internal consistency of data as demonstrated in Table III [20]-[22].

TABLE III
RULE TO CHECK THE RELIABILITY OF DATA IN TERMS OF ALPHA (Α)

Cronbach's alpha	Internal Consistency
$0.9 \leq \alpha$	Excellent
$0.8 \leq \alpha < 0.9$	Good
$0.7 \leq \alpha < 0.8$	Acceptable
$0.6 \leq \alpha < 0.7$	Questionable
$0.5 \leq \alpha < 0.6$	Poor
$\alpha < 0.5$	Unacceptable

In this study, seismic waves travel time, and velocity data for all selected seismic lines was loaded into SPSS software package version 23 to conduct Cronbach's alpha test. Fig. 5 depicts that the value of α for wave velocities and travel time database was found 0.24 and 0.55 respectively. As described in Table III, alpha (α) value less than 0.5 shows unacceptable reliability of data. Therefore, reliability of wave velocity database was considered unacceptable. Similarly, α value for database of wave travel time was found in between the 0.5 and 0.6, which proved its poor internal consistency.

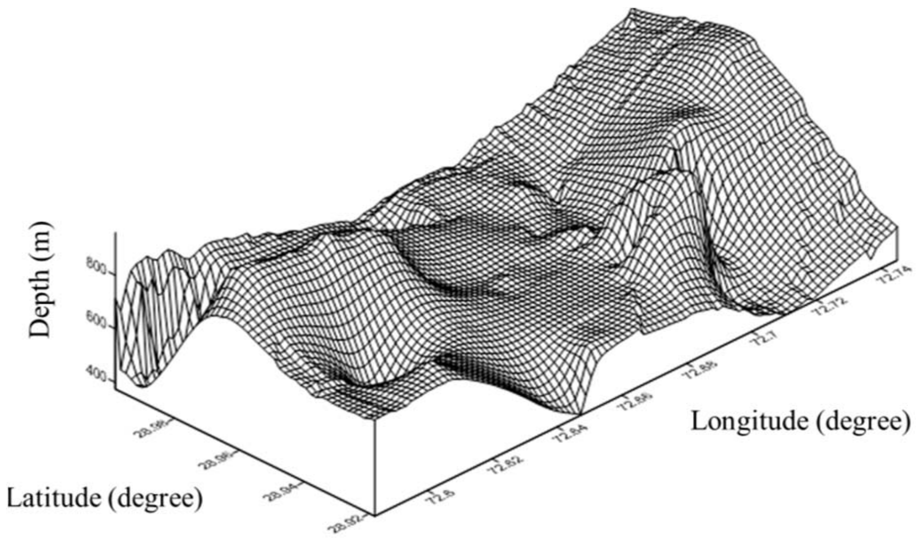

Fig. 3 A 3D-wireframe map showing intense folding in subsurface geological strata

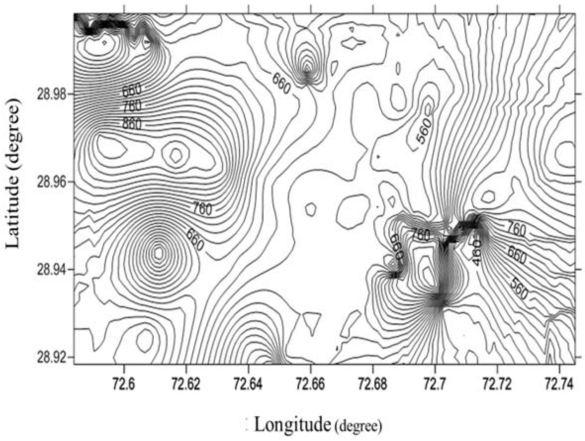

Fig. 4 A contour map showing intense undulations in subsurface geological strata

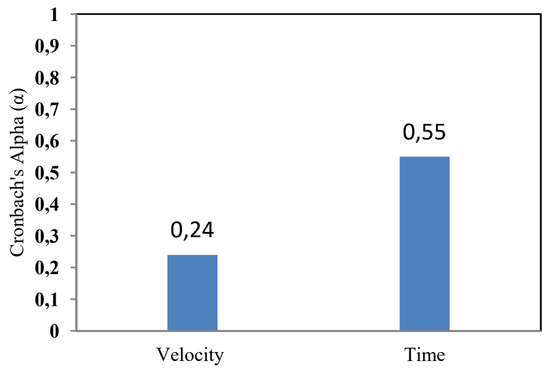

Fig. 5 Cronbach's aplha value for seismic wave velocity and travel time database

TABLE IV
T-TEST FOR THE EVALUATION OF ACQUIRED DATABASE QUALITY

Seismic lines	t-stats	t-critical	Null hypothesis
931-FAB-16A (Khewra)*	180.33	2.06	Rejected
931-FAB-33 (Khewra)	192.42	2.07	Rejected
931-FAB-34 (Khewra)	145.06	2.16	Rejected
931-FAB-35 (Khewra)	128.13	2.07	Rejected
931-FAB-37 (Khewra)	153.88	2.08	Rejected
931-FAB-16A (Patala)*	109.03	2.04	Rejected
931-FAB-33 (Patala)	173.01	2.08	Rejected
931-FAB-34 (Patala)	247.15	2.14	Rejected
931-FAB-35 (Patala)	103.42	2.14	Rejected
931-FAB-37 (Patala)	348.15	2.10	Rejected
931-FAB-16A (Namal)*	27.04	2.07	Rejected
931-FAB-33 (Namal)	169.16	2.07	Rejected
931-FAB-34 (Namal)	164.69	2.20	Rejected
931-FAB-35 (Namal)	142.07	2.13	Rejected
931-FAB-37 (Namal)	559.88	2.08	Rejected

*Khewra, Patala and Namal are the geological formations

To further investigate the quality of acquired database t-tests and F-test were conducted. The t-test uses t-distribution to find the probable values to accept or reject the null hypothesis. F-test follows f-distribution to compare known variances of two samples population under hypothesis statistics. Table IV shows that the t-test for all selected seismic lines rejected the null hypothesis (i.e. quality of acquired database is good for meaningful interpretation) which satisfied the poor quality of acquired database.

Similar kind of results also found in F-test in which alternate hypothesis (i.e. quality of acquired data was found erratic and falsified that was not adequate to provide meaningful interpretation) accepted that proved biasness in database as shown in Table V. Nearby ground, vibrations and noises adversely affect the quality of acquired database. Acquisition of low quality data, inadequate processing and misinterpretation could be the main reasons of unfavorable outcomes. A vast knowledge about subsurface lithology, structural geology, and meaningful interpretation through acquired seismic data enhances the chances to drill a productive well.

TABLE V
F-TEST FOR THE EVALUATION OF ACQUIRED DATABASE QUALITY

Seismic lines	F-sig	Sig-α	Null hypothesis
931-FAB-16A (Khewra)*	0.00	0.05	Rejected
931-FAB-33 (Khewra)	0.00	0.05	Rejected
931-FAB-34 (Khewra)	0.00	0.05	Rejected
931-FAB-35 (Khewra)	0.00	0.05	Rejected
931-FAB-37 (Khewra)	0.00	0.05	Rejected
931-FAB-16A (Patala)*	0.00	0.05	Rejected
931-FAB-33 (Patala)	0.00	0.05	Rejected
931-FAB-34 (Patala)	0.00	0.05	Rejected
931-FAB-35 (Patala)	0.00	0.05	Rejected
931-FAB-37 (Patala)	0.00	0.05	Rejected
931-FAB-16A (Namal)*	0.00	0.05	Rejected
931-FAB-33 (Namal)	0.00	0.05	Rejected
931-FAB-34 (Namal)	0.00	0.05	Rejected
931-FAB-35 (Namal)	0.00	0.05	Rejected
931-FAB-37 (Namal)	0.00	0.05	Rejected

*Khewra, Patala and Namal are the geological formations

IV. CONCLUSIONS

In exploration geophysics, the quality and reliability of acquired database should be good enough to get meaningful results before processing and interpretation stages. There are number of factors including non-specificity of geophysical survey, nearby ground vibrations, intensely complex and problematic subsurface geological strata etc. that significantly reduces acquired database quality and reliability.

The main objective of this study was to assess the quality of seismic reflection survey database of Fort Abbas Area of Pakistan by using statistical tools. The study reveals that seismic reflection survey database was biased, erratic and falsified. To investigate the subsurface structural and geological set up the reflection survey database was used to generate 3D wireframe and contour maps that were very questionable to understand the conditions of subsurface geological strata. NRMSE between seismic wave velocity data recorded at field and calculated by T^2-X^2 plot conformed high degree of residuals in database. The seismic lines spread over Patala and Namal formations including, 931-FAB-16A, 931-

FAB-33, 931-FAB-34, 931-FAB-35, 931-FAB-37 and 931-FAB-33, 931-FAB-37 had the highest values of NRMSE respectively. The Cronbach's alpha test proved unacceptable and poor reliability of seismic wave velocity and travel time database respectively. The alpha (α) value of wave velocity and travel time database was determined 0.24 and 0.55, respectively. The t-test and F-test results under null hypothesis statistics conformed poor quality and reliability of database.

Such type of studies could be used to handle and assess the quality and reliability of large database especially related to geophysical and geosciences studies. It could further be utilized to prepare database quality assessment models and algorithms with reasonable accuracy.

References

[1] New York: McGraw-Hill, 1988.

[2] R. Y. Wang and D. M. Strong, "Beyond accuracy: what data quality means to data consumers", Journal of Management Information Systems, vol. 12, no. 4, pp. 5-33, 1996

[3] M. A. Tate, "Web wisdom: how to evaluate and create information quality on the web", CRC Press, 2009.

[4] P. Katerattanakul and K. Siau, "Measuring information quality of web sites: Development of an instrument", Proceedings of the 20th International Conference on Information Systems, North Carolina: ACM, pp. 279–285, 1999.

[5] G. Shanks and B. Corbitt, "Understanding data quality: Social and cultural aspects", Proceedings of the 10th Australasian Conference on Information Systems, Wellington: MCB University Press Ltd, pp. 785–797, 1999.

[6] T. W. Mcrae, "The impact of computers on accounting", London: Wiley, 1964.

[7] T. C. Redman, "The impact of poor data quality on the typical enterprise", Communications of the ACM, vol. 41, pp. 79-82, 1998

[8] G. K. Tayi and D. P. Ballou, "Examining data quality", Communications of the ACM, vol. 41, pp. 54-57, 1998

[9] A. F. Karr, P. Ashish, Sanil and D. L. Banks, "Data quality: a statistical perspective", National Institute of Statistical Sciences, Technical Report, vol. 150, pp. 1-9, 2005

[10] K. K. Talagapu, "2D and 3D land seismic data acquisition and seismic data processing", Thesis, Department of Geophysics College of Science and Technology, Andhra University of Visakhaptanam, India, pp. 36-38, 2005.

[11] Eppelbaum, and B. Khesin, "Geophysical studies in Caucasus", Springer-Verlag Berlin Heidelberg, pp. 39, 2012

[12] B. Chetia and P. K. Dimri, "Noise problem and quality data acquisition in Himalayan foot hill area", GEOHORIZONS, vol. 23, 2005

[13] G. M. D. Sohail, "Structural & stratigraphic interpretation of migrated seismic sections of fortabbas area, Punjab platform, Pakistan", In: 5th Saint Petersburg International Conference & Exhibition–Geosciences: Making the most of the Earth's resources, Saint Peters-burg, Russia, April 2-5, 2012

[14] I. B. Qadri, "Petroleum geology of Pakistan", Pakistan Petroleum Ltd, 1995

[15] N. Aadil and G. M. D. Sohail, "Stratigraphic correlation and isopach maps of Punjab platform in middle Indus Basin, Pakistan. AAPG, Search and Discovery Article-10364, Tulsa. 2011.

[16] E. S. Robinson and C. Coruh, "Basic exploration geophysics", New York: Willey, 1988.

[17] R. Pontius, Thontteh, Olufunmilayo, Chen, and Hao, "Components of information for multiple resolution comparison between maps that share a real variable" Environmental Ecological Statistics, vol. 15, pp. 111–142, 2008.

[18] N. Ritter, "Understanding a widely misunderstood statistic: Cronbach's alpha", Paper presented at South-Western Educational Research Association (SERA) Conference: New Orleans, LA, 2010.

[19] J. M. Cortina, "What is coefficient alpha? An examination of theory and applications", Journal of Applied Psychology, vol. 78, pp. 98–104, 1993

[20] D. George and P. Mallery, "SPSS for Windows step by step: A simple guide and reference", Boston: Allyn & Bacon, 2003.

[21] P. Kline, "The handbook of psychological testing", London: Rout-ledge, 2000.

[22] R. F. DeVellis, "Scale development: Theory and applications", Los Angeles: Sage, pp. 109–110, 2012.