Research on the management of seismic data quality assessment methods

Li Bo^{1,a}, Liu Chengyu^{*1,b}, Zhang zhengshuai^{1,c}, Li Guoyi^{1,d}

¹Earthquake Administration of Shandong Province Jinan, China

Abstract. The quality of real-time waveform data from seismic networks not only directly affects the accuracy of seismic positioning accuracy, but also has an irreversible impact on the credibility of in-depth research findings in seismic disciplines. The article introduces an engineering perspective evaluation method for seismic data quality assessment, and uses real-time waveform data from the seismic station network in Shandong to form a visualization report on the quality of seismic quasi-real-time data, so as to determine the data quality of the seismic station network more intuitively and accurately.

1 Introduction

The digital seismic network in China has become a high-density monitoring network integrating earthquake monitoring and forecasting and scientific research after three five years of construction and renovation since the Tenth Five-Year Plan; especially for the digital network in Shandong, 117 seismic observation stations have been established. Especially for the digital station network in Shandong, 117 seismic observation stations have been established, and the China Earthquake Science and Exploration Array and the Lushan Array have set up many mobile observation seismometers in Shandong, especially after the completion of the earthquake early warning and intensity reporting system, and the reconstruction of 79 new reference stations in Shandong. The quality assessment of real-time seismic data has become the most critical part of monitoring and forecasting, and the quality of seismic data has a direct impact on the accuracy of the results of the earthquake reporting and early warning system.

Incorporated Research Institutions for Seismology (IRIS) [1] currently has hundreds of seismic data producing institutions around the world, The IRIS Data Management Center(DMS) [2] was established in 1986 in Seattle, USA, to provide data management, such as data processing, member sharing. It was also one of the first institutions to raise the issue of seismic data quality[3], but it has been a long process from raising the issue of data quality to quantitatively solving the data quality analysis

tools. At present, the National Seismic Network Data Backup Center of the Institute of Geophysics[4], China Earthquake Administration has also established a platform for data quality assessment and operational quality monitoring of seismic stations and built a seismic data quality assessment system. Xu Jiajun et al. used the observation data of Fujian seismic network to perform quality testing on the probability density function (PDF) of earthquake noise. Huang Lingzhu et al. used the real-time waveform data to detect the waveform anomalies of the station network, and the system output results can provide the basis for data quality determination for the earthquake early warning and intensity quick reporting system.

2 Waveform detection method

Seismic data quality assessment system is based on JOPENS flow service system of seismic network. Starting from different data source access methods, two storage modes of short-term data and long-term data are stored based on different data analysis methods, and then the data are interpreted. After data preprocessing, the weight distribution is carried out according to different methods, and the software interface or Web interface of data quality assessment report and visual display are generated. Data quality assessment software design ideas are shown in Figure 1 :

^a63899166@qq.com ^{b*}Corresponding author:670393809@qq.com
^czhangzs1989@126.com ^d596802488@qq.com

Figure 1 Data quality assessment software architecture diagram

In the process of data quality assessment of the seismic network, the data continuity, completeness and availability are judged by the data step rate assessment method; the time stamp difference method is used to monitor the delay of data; the calculation of the background noise level of the station base environment is realized by the root mean square value of the station base environment noise level method.

2.1 Step-rate evaluation method

In the transmission process of seismic real-time waveform, if the continuous occurrence of the same value of count is called data step, it often occurs that the system automatically fills zero (or random value) [5] when the signal is interrupted in the data recording process, or the pseudo signal in the recording process (the electrical signal generated by the pure data acquisition without seismic timing), which is also an important indicator to evaluate the continuity, availability and integrity of the recording data. The specific calculation method of data step-rate : the original waveform signal is calculated twice by difference, and then the proportion of zero value is calculated to represent the step rate of the data.

2.2 Timestamp Difference Method

Firstly, the server timing of the transmitter and receiver of the seismic data flow is consistent, and then the time stamps of the two servers are recorded[6]. One is the time stamp A when the data flow is sent, and the other is the time stamp B received by the data flow. The corresponding data delay is obtained by subtracting the two time stamps from B-A. According to the criteria, judging once per second, if the time delay is less than 2 seconds, it belongs to the normal phenomenon, if greater than or equal to 2 seconds, recording a delay. All delays are counted daily.

2.3 Noise level root mean square method

The use efficiency of seismic data is usually greatly reduced due to the presence of noise. The level of noise directly affects the quality of observation data[7]. The background noise level[8] is used to evaluate the reliability of data quality. The ground noise level of the observation environment is an important index to measure

the observation efficiency of the station, and it is also the technical basis for protecting the observation environment and site selection of the seismic station.

Power spectral density calculation process:

① Data preparation. Split channel, convert count number to velocity quantity.

② Initialization parameters. Number of data points (Nnum), number of Fourier transform points (Nfft), sampling rate (Fs), number of window length points (Wd), number of overlap points (Novelap), Hamming window length Hd. Note: $Wd \leq Nfft$; $Novelap < Nfft$; Nfft should preferably be 2^n ; Normally, you can set $Novelap = 0.5 * Wd$; $Hd = Wd$;

③ Segmentation to calculate the power spectrum. Data segmentation. Each calculated sample is segmented by length Nnum, and the segmentation process uses the length of each segment to take Wd, and then sliding overlap according to Novelap. Data windowing. In order to minimize the effect of "spectral leakage" and increase the width of the frequency peaks, the data in the Wd window is windowed, and the window length is the same as Wd, i.e., $Wd * Hd$. Fourier transform. Fourier transform of $Wd * Hd$ data to get the amplitude spectrum value FxL , the amplitude spectrum value are taken as absolute values, and then squared and divided by the Euclidean length that is to get the power spectrum of the section PxL .

④ Average power spectrum. The average power spectrum Px can be obtained by superimposing the amplitude spectrum of each segment obtained in step 3 and dividing it by the number of segments. the unit of the average power spectrum Px is converted to decibel value, that is, $Px = 10 * \log_{10}(Px)$, the unit is dB.

⑤ Calculate the power spectrum corresponding to the frequency point.

Background noise level calculation method:

① Data preparation. Splitting channels, converting digital numbers to velocity quantities.

② Calculate the power spectrum. The calculation process is the same as the steps in the power spectrum probability density function method, and finally the frequency points and the average power spectrum are obtained.

③ Convert the velocity power spectrum to acceleration power spectrum.

④ Transfer function correction.

⑤The RMS value of the noise is calculated based on the power spectral density using a 1/3 octave bandwidth.

3 Design of seismic data quality evaluation report

In the test phase of data quality evaluation report, the real-time flow data of 20 stations in Shandong Seismic Network (or the waveform data file read from the local

hard disk) are used. Some functional display (including waveform delay and continuity) can automatically read the waveform data of continuous waveform from the data acquisition of stations, and analyze the count number of data and other parameters, so as to realize the real-time and quasi-real-time monitoring and display of the main contents of the report, including the delay display of the network data, the continuity rate and the noise level of the base environment.

Figure 2 Data flow and WeChat end display chart

The principle of WeChat display is to generate the corresponding WeChat template message every day after the report is generated in the service background, and call the WeChat template sending interface to read the corresponding personnel to be sent, and send the template message to the corresponding personnel through multi-threaded message processing. After the message arrives, the receiver can click on the template message and obtain the specific data of the report by hyperlink.

4 Quality assessment report effect display

Shandong seismic network observation data quality assessment report mainly includes four parts : Data introduction, including data storage, station distribution, station latitude and longitude, station code, IP address and so on ;

On January 1, 2021, there were 20 real-time connected stations in the seismic network of Shandong Province. The total amount of data stored in the seismic network on that day was 2.01 G. The distribution of the stations is shown in Figure 3.

Figure 3 Data flow and WeChat end display chart
Data integrity, including data integrity rate, data

interruption frequency and statistical illustrations; Data availability, The specific evaluation indexes include data delay and environmental noise level; Examine the completeness of waveform data records and the extent to which statistical data records are missing. The data continuity rate of three stations in the statistical cycle was below 98%, the lowest of which was 97.2% for Tuesday's Lishan station.

Figure 4 Statistical chart of weekly data continuity rate

The results show that the ambient noise level of 7 stations is Class I, 6 stations are Class II, and 7 stations are Class III. As shown in Figure 5.

Fig. 5 Statistical chart of environmental ground noise level

Environmental ground noise level examination index description is mainly to examine whether there is an abnormality in the single-divisional noise. The calculation method is to calculate the average noise value of single station three divisions within a certain period of time (such as 24 hours, 30 days, etc.) respectively. If there is a phenomenon that the noise value of single-division direction is higher than the other two divisions, and exceeds three times the value of the other two single-division direction, it is considered that the single-division noise is abnormal.

Fig. 6 The environmental ground noise level of the station

Comprehensive conclusion, summarize the situation of quality assessment, and put forward suggestions in the quality problems of the station in the quick report and early warning.

5 Conclusion

(1) The visual quality assessment report effectively provides a basis for judging data quality. The operation and maintenance personnel of the seismic station network can see the quality assessment report to analyze the interruption rate, frequency of interruption, delay and noise in the waveform data quality of the previous day without opening the waveform, which provides effective technical support for the next step of data maintenance.

(2) At present, the seismic data quality assessment report uses 20 seismic stations for visual report display. The next step is to develop the waveform data quality scoring standard, which will include all the seismic instruments and equipment of Shandong seismic network in the quality assessment report, and divide the stations into grades according to the total data quality score. There are also some other missing quality judgments such as instrument leaning pendulum, data format and metadata, waveform straight line, zero point drift, number of sharp pulses, count value anomaly, etc. We will continue to supplement the assessment parameters and indicators to give full play to the scientific benefits of the seismic monitoring station network and improve the scientific, validity, timeliness, dynamics and operability of the seismic waveform data quality assessment.

References

1. Ahern T, Benson R, Casey R, et al. 2015. Improvements in data quality, integration and reliability: New developments at the IRIS DMC[J]. *Advances in Geosciences*, 40: 31-35.
2. Beauduin R, Lognonné P, Montagner J P, et al. 1996. The effects of the atmospheric pressure changes on seismic signals or how to improve the quality of a station[J]. *Bulletin of the Seismological Society of America*, 86(6) : 1760-1769.
3. Davis J P, Berger J. 2007. Calibration of the global seismographic network using tides[J]. *Seismological Research Letters*, 78(4) : 454-459.
4. Hutt C R, Ringler A T. 2011. Some possible causes of and corrections for STS-1 response changes in the global seismographic network[J]. *Seismological Research Letters*, 82(4) : 560-571.
5. Larson E W F, Ekstrom G. 2002. Determining surface wave arrival angle anomalies[J]. *Journal of Geophysical Research*, 107(B6) : ESE 7-1-ESE 7-15, doi: 10.1029/2000JB000048.
6. Laske G, Masters G. 1996. Constraints on global phase velocity maps from long-period polarization data[J]. *Journal of Geophysical Research*, 101(B7): 16059-16075.
7. Li Q S, Chen L, Wang L S, et al. 2010. USArray and

latest scientific discoveries[J].Acta Geologica Sinica
(in Chinese) ,84(6) : 847-853.

8. McN amara D E,Buland R P.2004.Ambient noise
levels in the continental United States [J]. Bulletin
of the Seismological Societyof America , 94(4) :
1517-1527.