

논문

불규칙한 빠짐을 포함한 탄성파 탐사 자료의 머신러닝을 이용한 트레이스 기반 내삽

이재우¹ · 박지호¹ · 설순지¹ · 윤대웅² · 변중무^{1*}

¹한양대학교 자원환경공학과

²전남대학교 에너지자원공학과

Trace-based Interpolation Using Machine Learning for Irregularly Missing Seismic Data

Zeun Yeeh¹, Jiho Park¹, Soon Jee Seol¹, Daeung Yoon², and Joongmoo Byun^{1*}

¹Department of Earth Resources and Environmental Engineering, Hanyang University

²Department of Energy and Resources Engineering, Chonnam National University

요약

최근에 활발히 적용되고 있는 머신러닝 기반 탄성파 내삽 기법들은 대부분 모음 자료를 2차원 영상화 하여 빠짐을 채우는 방법으로 하는 훈련(training)-추론(inference) 전략에 기초하므로 완벽히 채워진 다수의 모음자료가 훈련을 위해 필요하게 된다. 이 연구는 이와는 달리 트레이스 기반 내삽을 수행하는 내삽 기술의 훈련-추론 전략을 기본으로, 불규칙한 빠짐이 있는 현장자료만을 이용하여 훈련-추론을 모두 수행할 수 있는 머신러닝을 이용한 트레이스 기반 불규칙한 빠짐의 내삽 기술을 제시하였다. 이 연구에서는 불규칙한 빠짐이 있는 자료를 훈련과 추론에 체계적으로 사용하는 최대 연속빠짐 간격에 따라 정해지는 네트워크를 구성하는 방법 및 훈련하는 방법을 기술하였다. 또한, 서호주 Exmouth Sub-basin 지역의 Vincent 유전에서 얻어진 시간 참반사 보정된 탄성파 자료에 개발된 방법을 적용한 후, 예측 결과를 전통적인 내삽 방법의 결과와 비교 및 분석하였다. 신호대잡음비나 구조유사성과 같은 정량적인 지표를 통해 두 방법 모두 내삽 성능이 높은 것을 확인하였으며, 모든 주파수 대역에서도 골고루 좋은 결과를 보임을 확인하였다.

주요어

트레이스 내삽, 불규칙한 빠짐이 있는 탄성파 자료, 머신러닝

ABSTRACT

Recently, machine learning (ML) techniques have been actively applied for seismic trace interpolation. However, because most research is based on training-inference strategies that treat missing trace gather data as a 2D image with a blank area, a sufficient number of fully sampled data are required for training. This study proposes trace interpolation using ML, which uses only irregularly sampled field data, both in training and inference, by modifying the training-inference strategies of trace-based interpolation techniques. In this study, we describe a method for constructing networks that vary depending on the maximum number of consecutive gaps in seismic field data and the training method. To verify the applicability of the proposed method to field data, we applied our method to time-migrated seismic data acquired from the Vincent oilfield in the Exmouth Sub-basin area of Western Australia and compared the results with those of the conventional trace interpolation method. Both methods showed high interpolation performance, as confirmed by quantitative indicators, and the interpolation performance was uniformly good at all frequencies.

KEYWORDS

trace interpolation, irregularly missing seismic data, machine learning

서론

지하자원 탐사나 지하 지질구조 파악에 있어 탄성파 탐사자

료는 중요한 역할을 수행한다. 그러나, 탄성파 탐사자료는 종종 불규칙한 빠짐이나 자료의 손실로 인해 왜곡될 수 있으며, 이로 인해 자료 해석의 정확성이 떨어질 수 있다. 이러한 문제

Received: 12 April 2023; Revised: 27 April 2023; Accepted: 1 May 2023

*Corresponding author

E-mail: jbyun@hanyang.ac.kr

Address: 222, Wangsimni-ro, Seongdong-gu, Seoul, Republic of Korea

©2023, Korean Society of Earth and Exploration Geophysicists

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

를 해결하기 위해 트레이스 내삽(trace interpolation) 기술이 활용되어 왔다. 트레이스 내삽 기술은 트레이스(trace)가 얻어지지 못한 특정한 위치에서의 트레이스를 주변의 다른 트레이스들을 이용하여 복원하는 기술이다. 이 기술을 사용하면 자료처리를 위해 필요한 특정 형태의 배열을 갖는 자료를 얻어낼 수 있어서 현장에서 발생하는 자료 취득에 관련된 시간과 비용도 줄일 수 있다.

트레이스 내삽 기술은 입력 자료의 불규칙도에 따라 3가지로 분류할 수 있다(Trad, 2009; Trad, 2014). 첫 번째 수준은 규칙적인 입력 자료의 간격을 줄이는 기술로 공간 알리아스 효과(aliasing)를 극복하는 등의 효과를 기대할 수 있다(Spitz, 1991). 두 번째 수준은 불규칙하게 빠짐을 가지는 자료를 내삽하는 기술이다. 이 경우 먼저 비닝(bin) 등의 방법을 통해 일정간격을 갖는 그리드 위에 올려놓은 뒤 이 규칙적인 자료들 중에 채워지지 못한 불규칙적인 빠짐을 채우는 기술로 대표적으로 MWNI (Minimum Weighted Norm Interpolation; Liu and Sacchi, 2004; Choi *et al.*, 2017; Yeeh *et al.*, 2020b)와 POCS (Projection Onto Convex Sets; Abma and Kabir, 2006; Kim *et al.*, 2015)가 있다. 마지막으로 실제 자료의 취득 위치를 입력 자료에 반영하여 내삽을 수행하는 가장 높은 수준의 불규칙도를 반영하는 방법이 있으며 대표적으로 ALFT (Antileakage Fourier Transform; Xu *et al.*, 2010)와 Matching pursuit (Choi *et al.*, 2016; Vassallo *et al.*, 2010) 방법이 있다.

머신러닝 기반의 트레이스 내삽 기술은 최근에 큰 관심을 받고 있으며, 특히 딥러닝(Deep learning) 알고리즘을 사용한 방법들이 많은 연구가 이루어지고 있다(Jia and Ma, 2017; Oliveira *et al.*, 2018; Park *et al.*, 2022; Park *et al.*, 2021; Wang *et al.*, 2020a; Yeeh *et al.*, 2020a; Yoon *et al.*, 2021). 영상처리 기술에서의 초고해상도화(super resolution; Dong *et al.*, 2016) 기술의 발전과 함께, 손상된 영상과 손상되지 않은 영상사이의 관계를 이용하는 접근방법으로 GAN (Generative Adversarial Networks; Goodfellow *et al.*, 2020)이나 오토인코더(Autoencoder; Bourlard and Kamp, 1988) 계열의 네트워크를 이용하는 영상 기반의 내삽 연구가 활발히 진행되고 있다. 이들 방법은 높은 정확도 성능을 보이며 두 번째 수준까지의 불규칙도를 가지는 입력 자료 모두에 대한 내삽을 수행하는데 효과적이다(Dou *et al.*, 2022; Fang *et al.*, 2023; Pan *et al.*, 2020; Wang *et al.*, 2020b).

그러나 이러한 네트워크를 훈련시키기 위해서는 빠짐이 포함되지 않은 자료가 필요하며, 다양한 많은 자료를 사용하여 훈련시켜야 일반화(generalization) 성능을 지닌다는 단점이 있다. 이로 인해 훈련-추론 전략을 세우기 까다롭고, 실제 현장 자료를 훈련자료로 사용할 경우에는 반드시 목표하는 배열로 제대로 얻어진 현장자료가 다수 필요하다는 문제가 존재한다. 따라서 머신러닝 기반의 트레이스 내삽 기술의 훈련-추론 전략을 개선하여 이러한 문제를 해결할 필요가 있다. 반면에

Yoon *et al.* (2021)의 트레이스 기반의 머신러닝을 이용한 내삽 기술은 적은 수의 입력 및 출력 트레이스를 사용하여 빠짐이 포함된 탄성과 단면에서 훈련을 시킬 수 있다. 이로 인해 훈련-추론 전략을 세우기가 더 쉽다. 하지만 이 연구는 앞서 설명한 세가지 불규칙도 중 첫 번째 수준의 불규칙도를 다루는 데 그쳤다.

이 연구에서는 Yoon *et al.* (2021)이 제안한 트레이스 기반 내삽 기술의 훈련-추론 전략을 개선한 두 번째 수준의 불규칙도를 처리하는 데 효과적인 방법을 제안한다. 불규칙한 빠짐 중 연속된 빠짐의 트레이스의 복원을 위하여 연속된 빠짐 안에서의 여러가지 상대적인 위치에 대해서 서로 다른 네트워크를 사용하였다. 이 서로 다른 네트워크의 학습에는 모든 자료가 채워진 별도의 빠짐이 없는 모음(gather)이 아닌 내삽 해야 할 대상인 불규칙한 빠짐이 있는 탄성과 현장자료가 사용되었다. 최종적으로 훈련된 네트워크를 이용해 불규칙적인 빠짐을 복원하는 방법을 제안하였다.

제안된 방법을 이용한 수치실험을 서호주 Exmouth Sub-basin 지역의 Vincent 유전에서 얻어진 시간 참반사 보정된(time migrated) 탄성과 자료를 이용하여 진행하였고, 실험 결과를 전통적인 트레이스 내삽 방법과 비교하였다. 개발된 자료의 정량적인 평가를 위하여 신호대잡음비(PSNR; Peak Signal-to-Noise Ratio; Hore and Ziou, 2010)와 구조유사성(SSIM; Structural Similarity Index Measure; Wang *et al.*, 2004)과 같은 정량적인 평가 지표들을 사용한 비교하였다.

네트워크 훈련-추론 전략

일반적으로 GAN이나 오토인코더 등 2차원 단면을 입력 자료로 사용하는 영상 처리 방법 기반의 딥 러닝 트레이스 내삽 방법 들에서는 손상된 모음과 완성된 모음 사이의 관계를 학습한다. 이 관계를 학습하기 위해 완성된 모음을 라벨(label)로 사용하며 해당 라벨에 빠짐을 추가하여 손상시킨 모음을 훈련 자료로 사용한다. 이 과정에서 라벨을 구성하기 위한 완성된 모음을 확보하는 것이 필수적이다. 우리는 이러한 완성된 모음이 필수적이지 않은 트레이스 기반 접근방법(Yoon *et al.*, 2021)을 사용하여 불규칙적인 빠짐이 포함되어 있는 관측된 자료만을 이용하여 자료를 훈련시키는 접근방법을 제안한다.

현장 탄성과 단면이 Fig. 1과 같이 총 N_p 개의 트레이스 중 일부가 불규칙적인 빠짐을 가지고 있을 때 탄성과 단면은 크기가 $N_p \times N_i$ 인 행렬 \mathbf{d} 로 식 (1)과 같이 표현이 가능하다.

$$\mathbf{d} = \begin{bmatrix} d_{1,1} & d_{1,2} & \cdots & d_{1,N_i} \\ \vdots & \vdots & \ddots & \vdots \\ d_{N_p,1} & d_{N_p,2} & \cdots & d_{N_p,N_i} \end{bmatrix}, \quad (1)$$

여기서 N_p 는 각 모음에 기록된 트레이스의 수($|S_{\text{obs}}| + |S_{\text{gap}}|$),

N_i 는 각 트레이스에 시간 방향으로 기록된 샘플의 수를 뜻한다. S_{obs} 와 S_{gap} 은 각각 관측된 트레이스의 위치와 빠짐의 위치를 원소로 가지는 집합이며, $|X|$ 는 집합 X 의 원소의 개수를 뜻한다.

그리고 아래와 같이 내삽이 필요한 i 위치의 트레이스에서 왼쪽으로 n 만큼 오른쪽으로 m 만큼 떨어진 트레이스를 이용하여 내삽을 수행한다면 일반적인 입력 자료, X 와 라벨 자료, Y 를 행렬로 아래와 같이 표현할 수 있다.

$$X_{i,(n,m)} = \begin{bmatrix} d_{i-n,1} & d_{i-n,2} & \cdots & d_{i-n,N_i} \\ d_{i+m,1} & d_{i+m,2} & \cdots & d_{i+m,N_i} \end{bmatrix}, \quad (2)$$

$$Y_i = [d_{i,1} \ d_{i,2} \ \cdots \ d_{i,N_i}], \quad (3)$$

여기서 X 는 $[2 \times N_i]$ 크기의 두 입력 트레이스를 갖는 행렬, Y 는 $[1 \times N_i]$ 크기의 한 출력 트레이스를 갖는 행벡터를 의미한다.

여기서 거리가 i 로부터 좌우로 각각 n, m 떨어진 두 입력 트레이스만을 입력자료로 사용하여 i 번째 자료를 내삽하기 때문에 훈련에 입력으로 사용되는 두 트레이스 사이에 라벨로 사용하는 i 번째 트레이스 이외에 빠짐이 있는 경우에도 훈련 자료로 사용할 수 있다. 따라서 큰 간격의 내삽을 수행하기 위한 네트워크가 필요할 때도 효율적으로 훈련자료를 구성할 수 있다. 이 때 훈련단계에서는 예측과는 달리 라벨로 현장자료가 필요하므로 $\{1-n, i, i+m\} \subset S_{\text{obs}}$ 이어야 하고, 예측 단계에서는 $\{1-n, i+m\} \subset S_{\text{obs}}, \{i\} \subset S_{\text{gap}}$ 를 만족하게 된다. 이렇게

구성한 입력 자료와 라벨 자료의 쌍을 이용하여 내삽 네트워크 $R_{(n,m)}$ 을 훈련시킬 수 있다. 이 때 훈련시키기 위한 목적 함수는 아래와 같이 식 (2)와 식 (3)에서 정의한 $X_{i,(n,m)}, Y_i$ 을 이용하여 나타낼 수 있으며 제곱 평균 오차를 사용하였다.

$$\text{loss}_{(n,m)} = |R_{(n,m)}(X_{i,(n,m)}) - Y_i|^2 \quad (4)$$

우리는 이 접근방법을 구현하기 위한 네트워크, $R_{(n,m)}$ 으로 Yoon *et al.* (2021)이 사용한 skip connection이 포함된 DBiLSTM (deep bidirectional LSTM; deep bidirectional long short-term memory)을 변형하여 사용하였다(Fig. 2).

여기서 규칙적인 손상과는 달리 불규칙적인 손상을 가지고 있는 자료는 여러 가지 서로 다른 크기의 연속된 빠짐을 가지고 있을 수 있다. 이를 채우기 위하여 존재하는 모든 크기의 연속된 빠짐에 대해서 모든 위치의 트레이스를 내삽 할 수 있는 네트워크를 각각 학습시켜야 한다. 예를 들어 Fig. 1과 같이 가장 큰 연속된 빠진 트레이스의 크기가 3이라면, $(n, m) = (1, 1), (1, 2), (2, 1), (1, 3), (2, 2), (3, 1)$ 에 대하여 총 6가지의 서로 다른 네트워크, $R_{(n,m)}$ 가 필요하다. 여기서 i 번째 트레이스를 채우기 위하여 i 에서 좌측으로 n 번째의 트레이스와 우측으로 m 번째의 트레이스를 $n:m (n,m)$ 으로 내분하는 위치의 트레이스를 $R_{(n,m)}$ 을 이용하여 내삽하는 대신에 좌우 반전된 입력자료인 i 를 중심으로 우측의 m 번째의 트레이스와 좌측의 n 번째 트레이스를 순서대로 입력 자료로 사용한 뒤 이를 $m:n(m,n)$ 으로 내분하는 위치의 트레이스를 대칭되는 네트워크인 $R_{(m,n)}$ 으로 내삽할 수 있다. 이를 통해 같은 입력자료와 출력자료를 이용하여 대칭되는 서로 다른 네트워크 $R_{(n,m)}$ 과

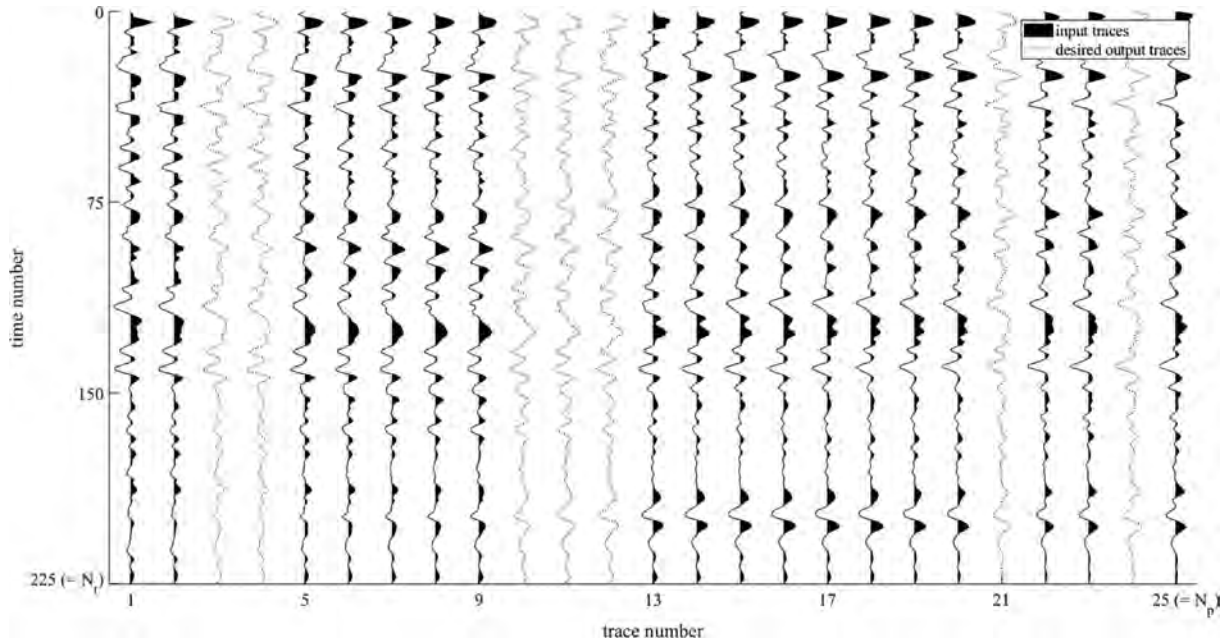


Fig. 1. The black variable area wiggle plot represents observations used as input for trace interpolation, while the wiggle plot with black dashed lines illustrates the desired output gaps in the trace interpolation process.

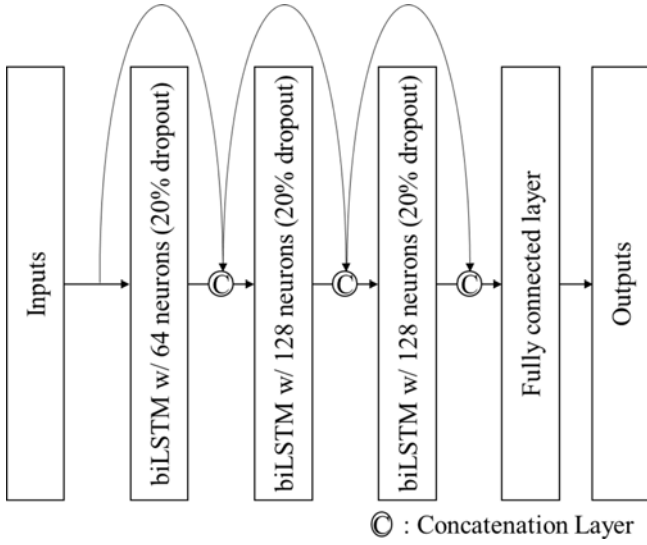


Fig. 2. The architecture of deep bidirectional long short-term memory (DBiLSTM) with skip connection network utilized for performing the interpolation process.

$R_{(m,n)}$ 을 학습시킬 수 있음을 알 수 있다. 따라서 m 과 n 사이의 대칭성 $R_{(n,m)}(\mathbf{X}_{i,(n,m)}) = R_{(m,n)}(\mathbf{X}_{i,(-m,-n)})$ 이 존재한다고 말할 수 있다. 여기서 $-$ 기호는 네트워크 입력 시 우측의 자료를 좌측으로 또는 좌측의 자료를 우측으로 바꾸는 것을 의미한다. 이를 고려할 경우 $n = m$ 인 2가지 네트워크를 제외하고 중복성을 제외하면 총 4개의 네트워크로 필요한 모든 빠짐을 채울 수 있다. 또한, $n = m$ 인 경우에도 훈련 자료 증강 (augmentation) 및 예측 정확도 향상 목적으로 좌우 바뀐 입력 자료를 학습에 사용할 수 있다. 일반적으로 가장 큰 연속된 빠짐의 길이가 gap_{max} 일 때 필요한 네트워크 수의 최대값 (number of networks)은 식 (5)와 같다.

$$(\text{number of networks}) = \text{floor}\left(\frac{gap_{max}+1}{2}\right) \times \text{ceiling}\left(\frac{gap_{max}+1}{2}\right), \quad (5)$$

여기서 $\text{floor}(x)$ 는 x 보다 크지 않은 최대 정수를 나타내고,

$\text{ceiling}(x)$ 는 x 보다 작지 않은 최소 정수를 나타낸다.

예를 들어 훈련 단계에서 Fig. 1의 자료를 사용하여 $R_{(1,2)}$ 를 훈련시키기 위하여 $\{1-n, i, i+m\} \subset S_{\text{obs}}$ 을 만족하는 $n = 1, m = 2, i \in \{6, 7, 14, \dots, 23\}$ 를 사용할 수 있다. 또한 상술한 대칭성을 이용하면 $n = -2, m = -1, i \in \{7, 8, 15, \dots, 22\}$ 도 사용하여 두 위치의 빠짐을 채우는 네트워크를 하나로 합칠 수 있다. 반면에 $R_{(2,2)}$ 를 훈련시키기 위하여 $n = 2, m = 2, i \in \{7, 15, 16, \dots, 20\}$ 을 활용할 수 있다. 그리고 데이터 증강을 위하여 $n = -2, m = -2, i \in \{7, 15, 16, \dots, 20\}$ 도 훈련 자료로 사용할 수 있다. 그리고 예측 단계에서는 Fig. 1의 빠짐 중 가장 왼쪽의 연속된 2개의 빠짐 중 왼쪽 빠짐인 3번 트레이스의 경우 $\mathbf{Y}_3 = R_{(1,2)}(\mathbf{X}_{3,(1,2)})$ 로 내삽 할 수 있다. 반면에 오른쪽 빠짐인 4번 트레이스의 경우는 $\mathbf{Y}_4 = R_{(2,1)}(\mathbf{X}_{4,(2,1)})$ 를 이용하여 내삽하거나, 상술한 대칭성을 활용하여 $\mathbf{Y}_4 = R_{(1,2)}(\mathbf{X}_{4,(-1,-2)})$ 를 이용하여 예측할 수 있다. 앞서 기술한 대로 훈련 네트워크의 수를 줄여 효율적인 연산을 하기 위하여 $R_{(2,1)}$ 은 훈련시키지 않았으므로 4번 트레이스의 경우는 $\mathbf{Y}_4 = R_{(1,2)}(\mathbf{X}_{4,(-1,-2)})$ 을 이용하여 예측한다. 이런 방법으로 연속된 빠짐을 내삽 할 수 있다. 11번 트레이스의 경우는 $\mathbf{Y}_{11} = R_{(2,2)}(\mathbf{X}_{11,(2,2)})$ 와 입력자료의 위치를 바꾸어 넣어주는 $\mathbf{Y}_{11} = R_{(2,2)}(\mathbf{X}_{11,(-2,-2)})$ 를 모두 사용할 수 있다. 정확도 향상을 위해 두 결과를 모두 구한 뒤, 두 결과의 평균을 예측에 사용하였다. 이를 고려하여 Fig. 1에 표시되어 있는 모든 빠짐을 내삽 하는 네트워크를 훈련시키기 위한 매개변수(m, n, i)들과 모든 빠짐을 내삽 하기 위한 매개변수들을 Table 1에 기술하였다.

이렇게 모든 종류의 빠짐을 내삽 하기 위하여 빠짐을 연속된 빠짐의 개수로 분류해야 하는데 이러한 분류는 S_{gap} 을 이용하여 수행할 수 있으며, 이 분류를 수행하는 알고리즘의 의사 코드(pseudo code)를 부록에 기술하였다.

현장자료 적용사례

제안한 알고리즘의 검증을 위하여 자료는 서호주 Exmouth Sub-basin 지역의 Vincent 유전에서 얻어진 시간 참반사 보정된 탄성과 자료를 사용하였다. 실험을 위하여 공간 방향으로

Table 1. Parameters required for training and prediction of the networks utilized to fill the gaps in Figure 1.

Network	Training	Prediction
$R_{(1,1)}$	$n = 1, m = 1, i \in \{6, 7, 8, 14, 15, 16, 17, 18, 19\}$	$n = 1, m = 1, i \in \{21, 24\}$
	$n = -1, m = -1, i \in \{6, 7, 8, 14, 15, 16, 17, 18, 19\}$	$n = -1, m = -1, i \in \{21, 24\}$
$R_{(1,2)}$	$n = 1, m = 2, i \in \{6, 7, 14, 15, 16, 17, 18, 20, 23\}$	$n = 1, m = 2, i = 3$
	$n = -2, m = -1, i \in \{7, 8, 15, 16, 17, 18, 19, 22\}$	$n = -2, m = -1, i = 4$
$R_{(1,3)}$	$n = 1, m = 3, i \in \{2, 6, 14, 15, 16, 17, 19, 20\}$	$n = 1, m = 3, i = 10$
	$n = -3, m = -1, i \in \{5, 8, 16, 17, 18, 19, 22\}$	$n = -3, m = -1, i = 12$
$R_{(2,2)}$	$n = 2, m = 2, i \in \{7, 15, 16, 17, 18, 20\}$	$n = 2, m = 2, i = 11$
	$n = -2, m = -2, i \in \{7, 15, 16, 17, 18, 20\}$	$n = -2, m = -2, i = 11$

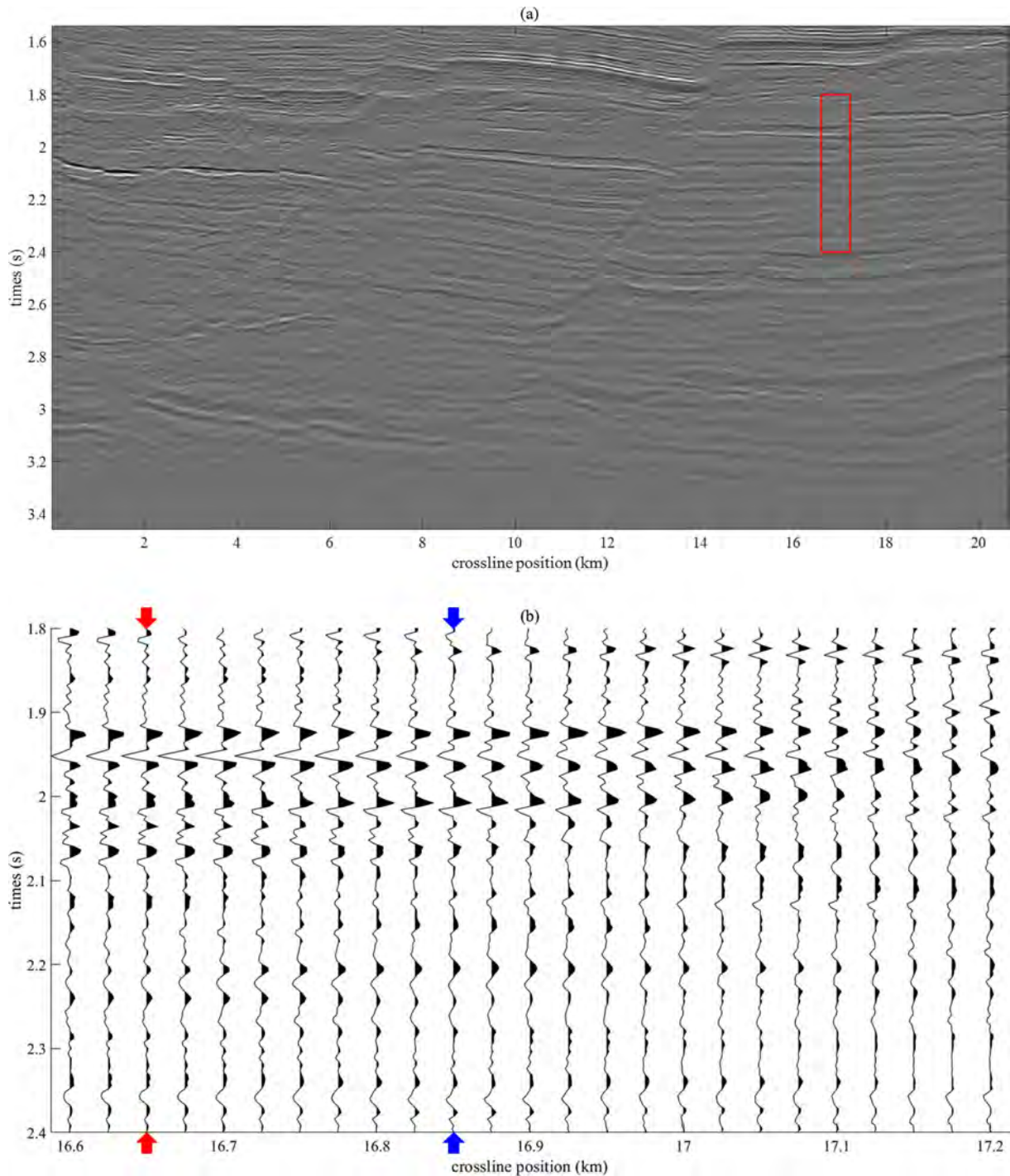


Fig. 3. (a) The selected time migrated section of field data for the Vincent oilfield in the Exmouth Sub-basin of West Australia used for the numerical experiment. (b) The enlarged data from the red solid box area in (a) displayed as a variable area wiggle plot. The red and blue arrows in (b) indicate the location of the selected traces for continuous wavelet transform.

25 m간격으로 총 길이 20.625 km의 825개의 트레이스를 선정하고, 시간 방향으로 1.540 초부터 3.456 초까지 480개 자료를 포함하도록 선정(Fig. 3)하였다. 불규칙적인 빠짐을 내삽하는 성능을 실험하기 위하여 22.5%인 186개의 트레이스를 불규칙적으로 손상시켜 최대 3개의 연속된 빠짐이 불규칙하게

존재하는 관측자료를 모사하였다(Fig. 4a). 그리고 이렇게 불규칙한 빠짐이 포함된 자료에서 불연속 구간이 포함되어 있으며 여러 가지 크기의 빠짐을 볼 수 있는 구간을 확대하여 Fig. 4b에 위글 트레이스(wiggle trace)으로 도시하였다.

먼저 최대 3개의 연속된 빠짐을 내삽 하기 위하여 필요한 4

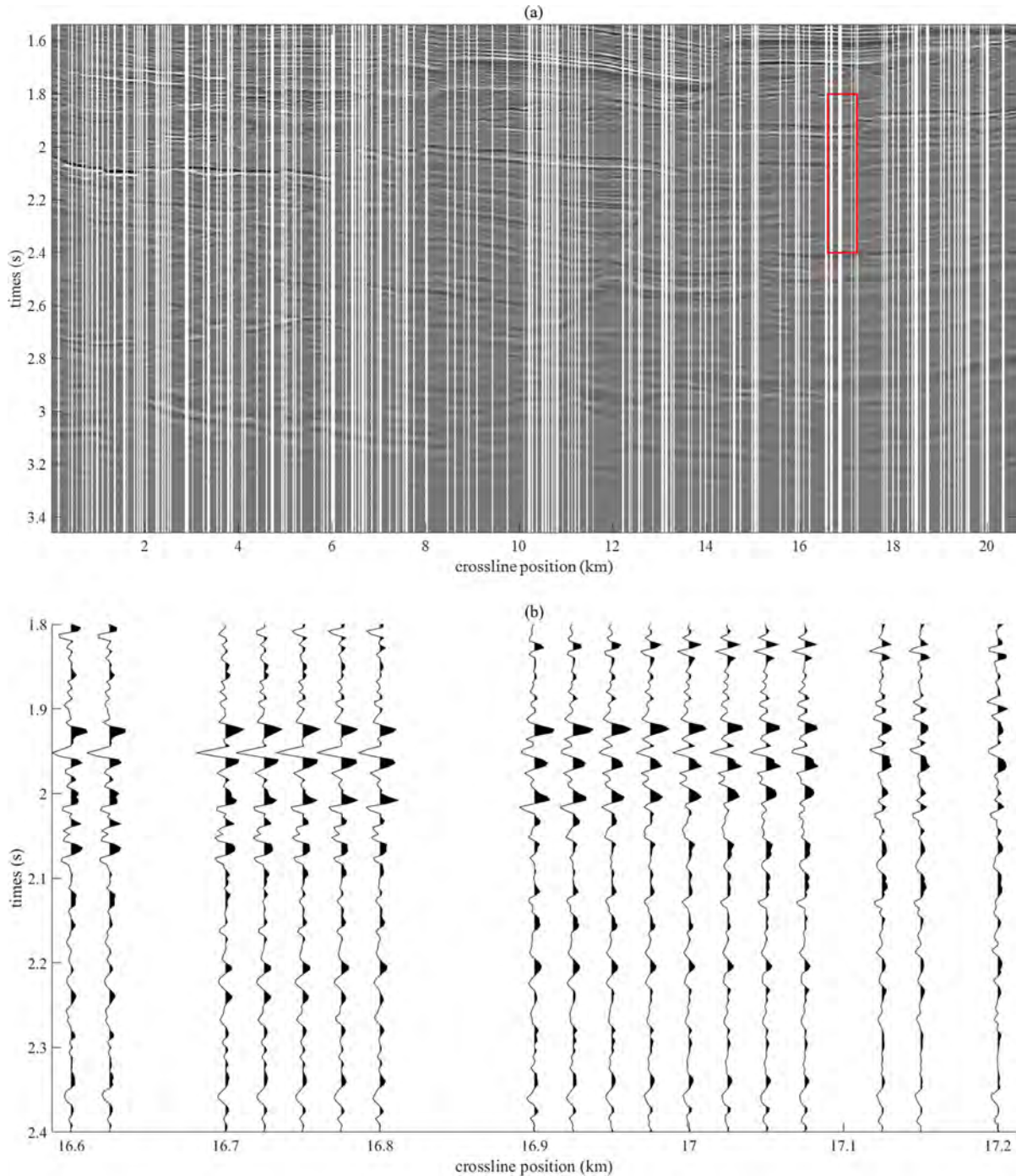


Fig. 4. (a) A simulated section containing gaps, created by introducing irregular damages to the complete data displayed in Figure 3, to validate the interpolation performance. (b) The data from the red solid box area in (a) displayed as a variable area wiggle plot.

개의 네트워크, $R_{(1,1)}$, $R_{(1,2)}$, $R_{(1,3)}$, $R_{(2,2)}$ 를 훈련시키기 위하여 훈련 자료와 라벨 자료의 쌍을 각각 구성하였다. 이 때 사용한 훈련 라벨 쌍은 상술한 방법을 이용하여 대칭성을 고려하여 구성하면 각각 764, 761, 743, 756개로 큰 차이가 나지 않는 것을 확인할 수 있다.

이렇게 구성된 훈련자료를 이용하여 90%의 자료는 네트워

크의 훈련에, 나머지 10%의 자료는 훈련 과정의 확인(validation)에 사용하였다. Fig. 2와 같이 뉴런 개수가 각각 64, 128, 128개인 3층 구조의 네트워크를 훈련시켰으며, 과적합을 피하기 위하여 각 층에는 20%의 드랍아웃(drop out)을 적용하여 네트워크를 만들었다. 그 밖에 배치(batch)크기는 16을 사용하였다. 또한 학습율(learning rate) 0.001인 Adam optimizer

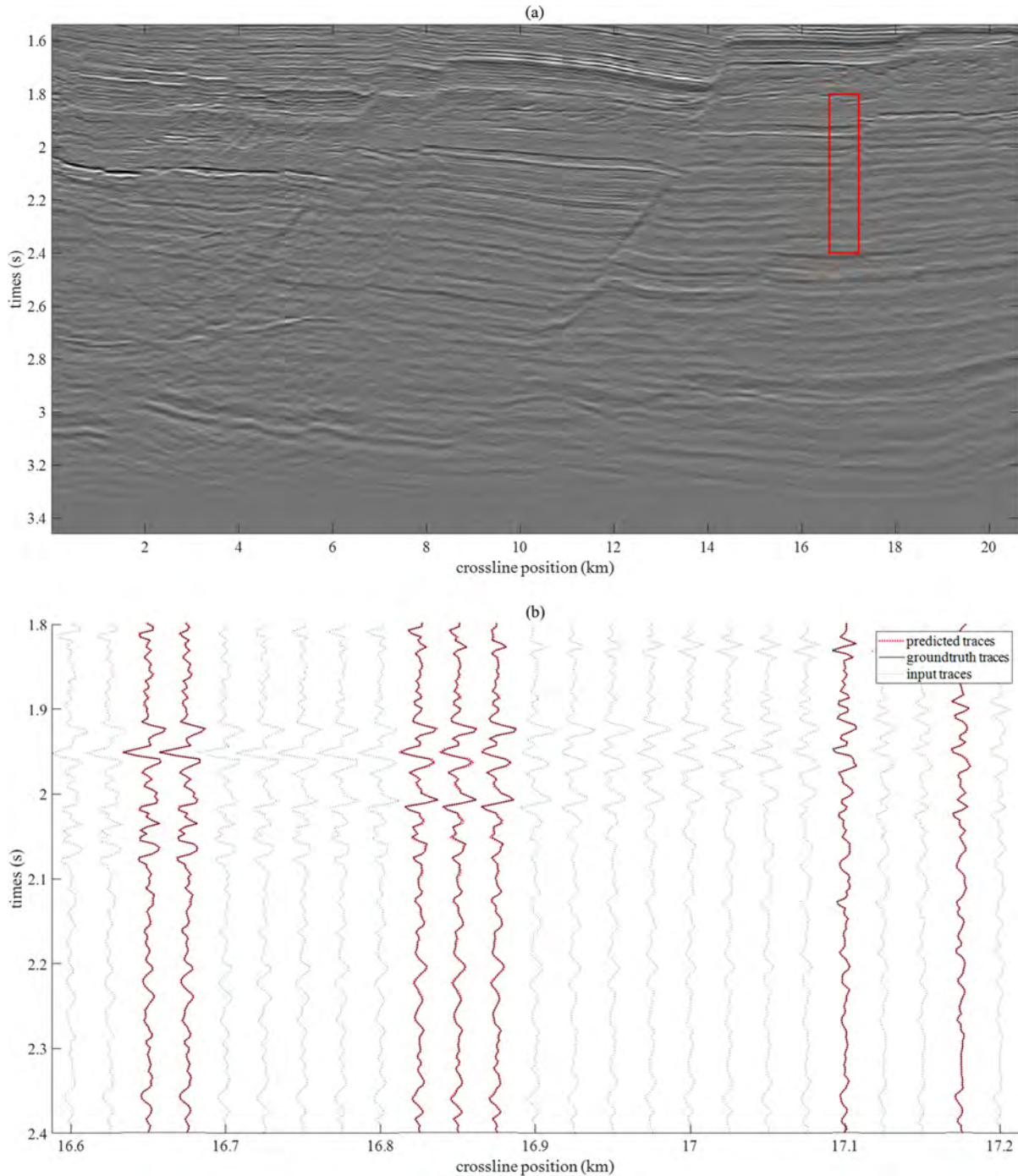


Fig. 5. (a) A section interpolated using the proposed method. (b) An enlarged view of the red solid box area in (a). The red dashed and black solid lines represent the prediction result of the proposed method and the ground truth, respectively. The black dashed line indicates the neighboring data used as input.

을 사용하여 네트워크를 훈련시켰다.

이렇게 불규칙적인 빠짐이 포함된 모음의 자료만을 가지고 총 4개의 내삽 네트워크를 훈련시킬 수 있었으며, 이렇게 훈련된 4개의 네트워크에 빠짐의 종류에 따라 적절한 네트워크를 사용하여 불규칙한 빠짐을 모두 내삽 하였고 그 결과를 Fig. 5에 도시하였다.

모델 제약조건이 적용된 MWNI방법과의 성능 비교

이 연구에서 개발되어 적용된 방법의 내삽 성능 평가를 위하여 지금까지 내삽연구에서 가장 많이 사용되고 있는 전통적인 내삽 방법인 모델 제약조건이 적용된 MWNI (Yeoh *et al.*, 2020b)를 이용한 내삽 결과(Fig. 6a)와 비교하였다. Fig. 5a와

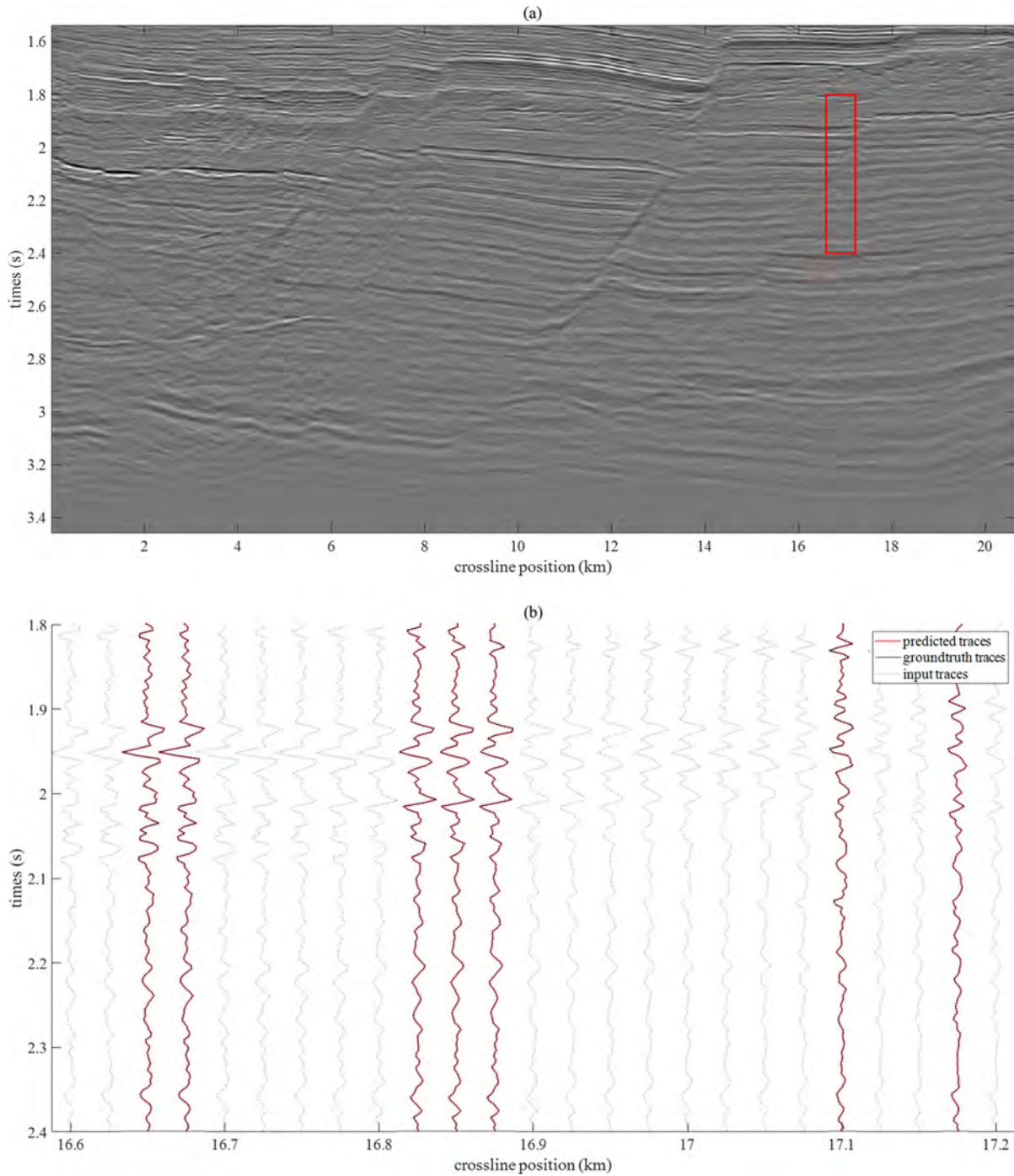


Fig. 6. (a) A section interpolated using the MWNI method. (b) An enlarged view of the red solid box area in (a). The red dashed and black solid lines represent the result of the MWNI interpolation method and the ground truth, respectively. The black dashed line indicates the neighboring data used as input.

Fig. 6a에서 볼 수 있는 것처럼 같은 손상 자료에 대한 채움 결과인 두 결과 모두 실제값(groundtruth; Fig. 3a)과 거의 유사한 모습을 보여주었다. 특히 좀 더 결과를 정확히 확인해보기 위해 1, 2, 3개의 연속된 빠짐이 모두 존재하는 영역에 대해 확대하여 비교한 Fig. 5b와 Fig 6b에 빨간색 점선으로 도시된 내삽 결과가 검정색 실선으로 표시된 실제값과 거의 겹쳐지는

것을 확인할 수 있다.

조금 더 정량적인 비교를 위하여 전체 단면에 대한 오차를 계산해 보았다. Fig. 7a와 Fig. 8a에 보이는 결과와 같이 두 결과 모두 전영역에 대해 눈에 띄는 큰 오차를 보이는 부분이 없이 0에 가까운 오차를 보이는 것으로 확인되었다. 그러나, 좀 더 자세한 비교를 위해 확대한 Fig. 7b와 Fig. 8b에서 보이는

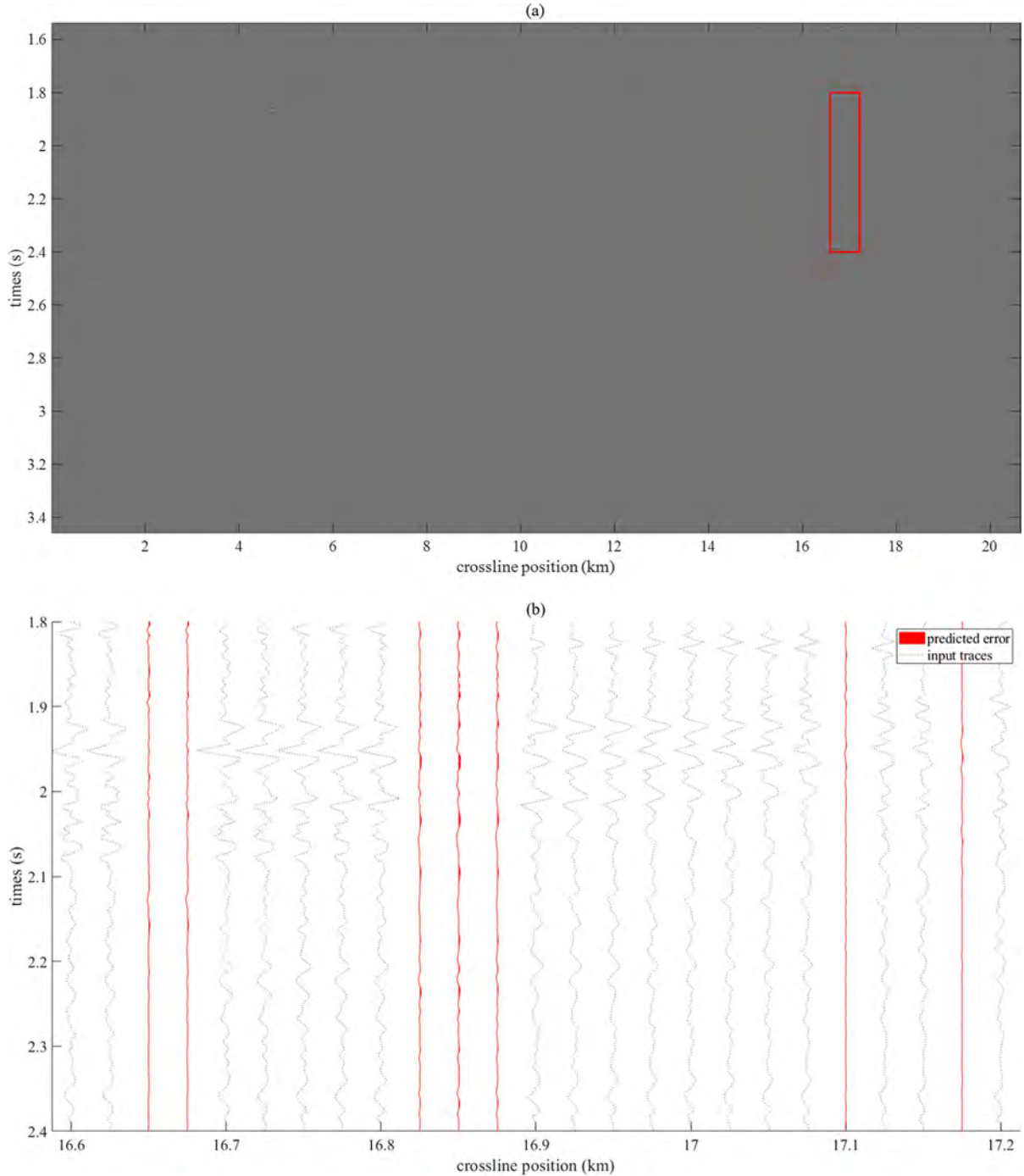


Fig. 7. (a) The difference between the proposed method and the ground truth. (b) An enlarged view of the red solid box area in (a). The red variable area wiggle plot represents the difference between the proposed method and the ground truth, and the black dashed line indicates the neighboring data used as input.

것처럼 전영역의 자료를 전부 동시에 연산에 사용하는 모델 제약조건이 적용된 MWNI방법이 내삽 할 트레이스 근처만의 자료를 이용하여 내삽을 수행하는 제안된 방법에 비해 더 정확하게 내삽을 수행했음을 알 수 있다. 전체 자료에 대해 좀 더 정량적인 비교를 위해서 Table 2에 다음식들로 표현되는 신호대잡음비 (Hore and Ziou, 2010)와 신호들의 유사도를 나타

내는 구조유사성(Wang *et al.*, 2004)을 계산하여 비교하였다.

$$\text{PSNR} = 10 \cdot \log_{10} \frac{\text{MAX}_I^2}{\text{MSE}} \quad (6)$$

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (7)$$

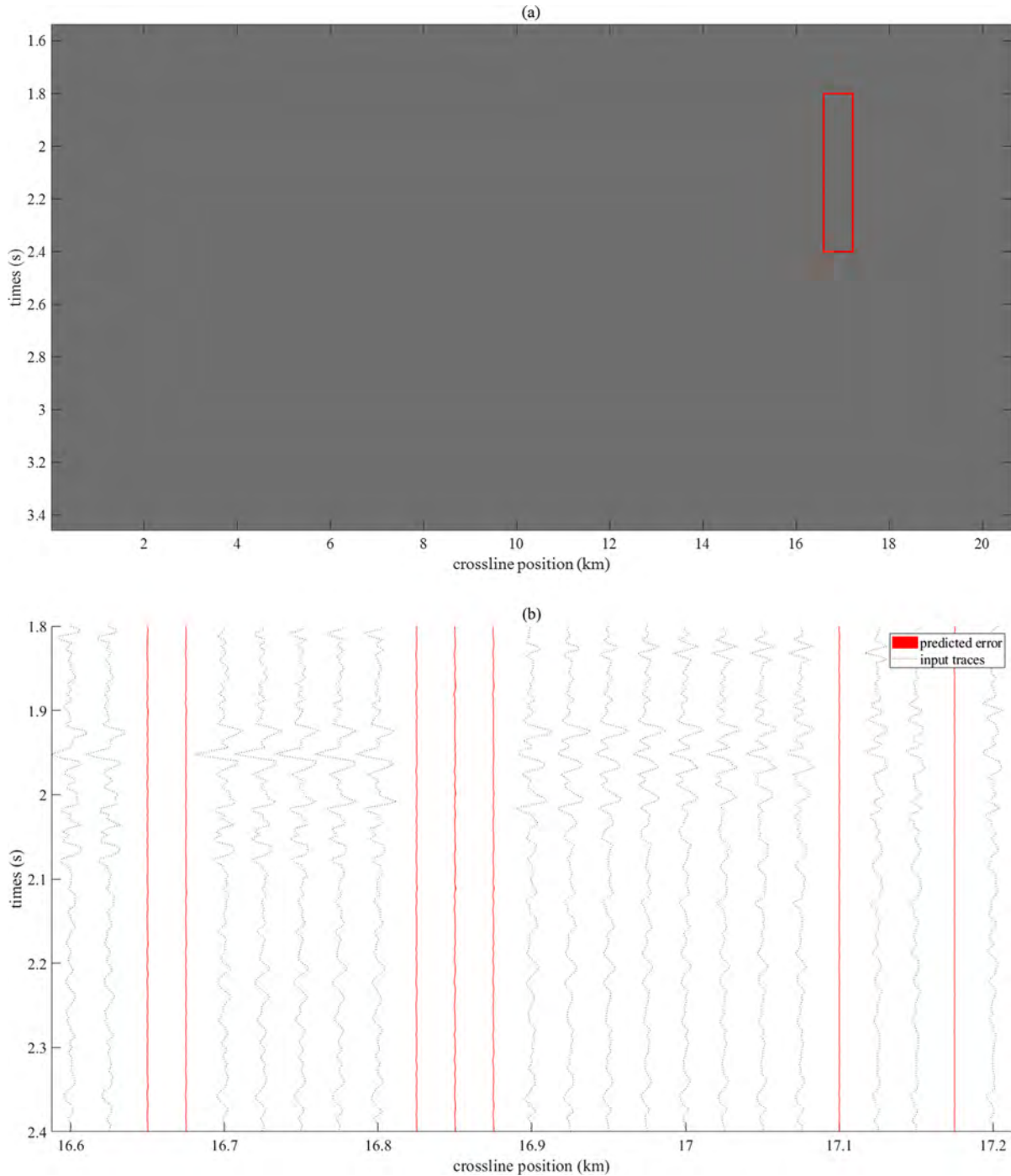


Fig. 8. (a) The difference between the MWNI method and the ground truth. (b) An enlarged view of the red solid box area in (a). The red variable area wiggle plot represents the difference between the conventional method and the ground truth, and the black dashed line indicates the neighboring data used as input.

(6)번식에서 MAX_i 은 영상에서 나타날 수 있는 최대 픽셀 값이며, MSE는 원본 영상과 복원된 영상 간의 평균 제곱 오차를 나타낸다. 이를 통해 두 영상 간의 품질을 측정할 수 있다. 한편, (7)번식에서 X, Y는 각각 원본 영상과 복원된 영상을, μ_x, μ_y 는 각각 영상 X와 Y의 평균을 나타낸다. 또한 σ_x, σ_y 는 각각 영상의 표준편차를, σ_{xy} 는 영상 X와 Y의 공분산을

나타낸다. 마지막으로, C_1, C_2 는 작은 값의 안정성을 보장하기 위한 상수이다.

Table 2에 나타난 것처럼 신호대잡음비 및 구조유사성의 기술한 정량적인 비교 결과에서 두 방법 모두 아주 높은 성능을 보여줬지만, 모델 제약조건이 적용된 MWNI 방법에서 조금 더 높은 성능을 확인하였다. 이는 모델 제약조건이 적용된

Table 2. Quantitative comparison of prediction performance between the proposed method and the model-constrained MWNI method using structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR).

		$R_{(1,1)}$	$R_{(1,2)}$	$R_{(1,3)}$	$R_{(2,2)}$	all
Proposed method	PSNR	44.7200	42.2381	36.2750	32.7271	44.0778
	SSIM	0.9655	0.9171	0.8992	0.8350	0.9843
model-constrained MWNI	PSNR	51.4314	49.0058	42.5830	38.5142	50.6061
	SSIM	0.9595	0.9334	0.9196	0.8896	0.9724

MWNI 방법이 전영역의 자료를 전부 동시에 고려한 연산을 하기 때문이다. 하지만 트레이스의 개수가 적은 경우에는 사용이 제한되며, 푸리에 변환(Fourier transform)이나 역행렬을 구하는 연산등에서 필요한 시간이 출력 자료의 크기가 커질 경

우 연산량이 푸리에 변환 연산에서 로그-선형적으로(log-linear), 역행렬 연산에서 기하급수적으로 늘어나게 된다. 하지만 제안된 방법은 두 개의 입력 트레이스만 이용하여 연산을 수행하기 때문에 이러한 제약이 비교적 적으며, 한 자료의 예

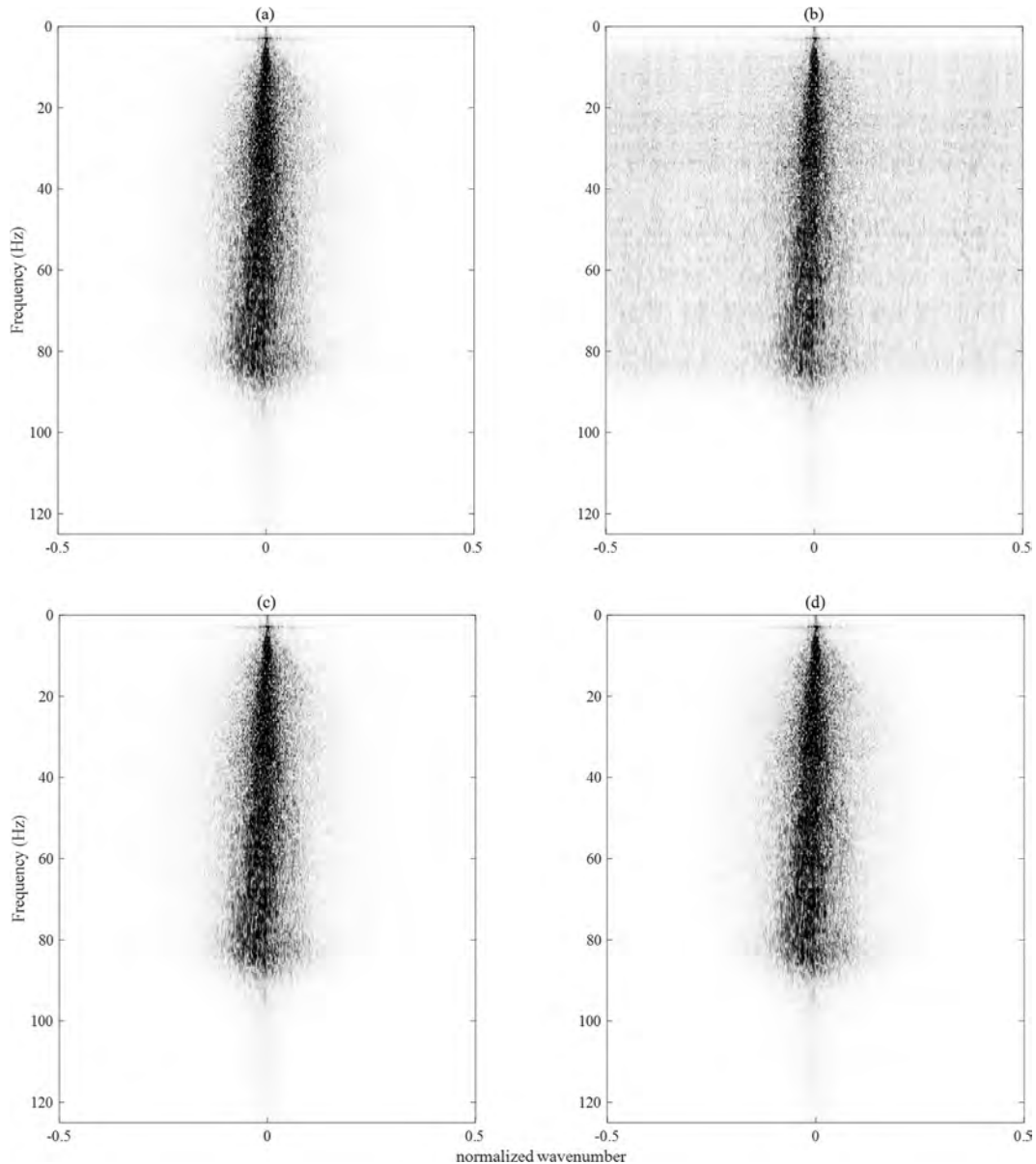


Fig. 9. (a)-(d) The f - k spectra for the complete (Fig. 3a), observation (Fig. 4a), and interpolated sections using the proposed method (Fig. 5a), and interpolated section using the MWNI method (Fig. 6a), respectively.

선형적으로 늘어나게 되어 연산량 증가가 크지 않다. 다만, 자료의 훈련 시간이 입력 자료의 크기에 따라 증가하긴 하지만 모든 자료의 훈련이 반드시 필요하지는 않고 훈련에 충분한 자료의 크기를 선택적으로 적용할 수 있기 때문에 자료의 크기가 커질 수록 연산량 측면에서 큰 이점이 있다. 자료의 크기가 작은 이 실험에서는 Intel(R) Xeon(R) CPU E5-2630 v3 프로세서와 GTX 1080 Ti 그래픽 프로세서를 사용하여 네트워크의 훈련 시간은 3시간 정도 걸렸지만 예측 및 추론 시간은 10초 미만 소요되었다. 모델 제약조건이 적용된 MWNI의 경

우는 10분 정도 소요되었다.

내삽 결과의 성능을 확인해 볼 수 있는 다른 방법으로 주파수-과수(f-k) 영역에서의 변화를 확인해 볼 수 있다. Fig. 9에 빠짐이 포함되지 않은 실제 탄성과 단면(Fig. 3a), 불규칙한 빠짐이 존재하는 탄성과 단면(Fig. 4a), 이 연구에서 제안한 네트워크에 의해 복원된 자료를 포함한 단면(Fig. 5a) 및 모델 제약조건이 적용된 MWNI로 복원된 자료(Fig. 6a)를 포함한 단면의 주파수-과수 스펙트럼을 나타내었다. 빠짐이 포함되지 않은 실제값의 단면의 주파수-과수 스펙트럼(Fig. 9a)은 주된 에

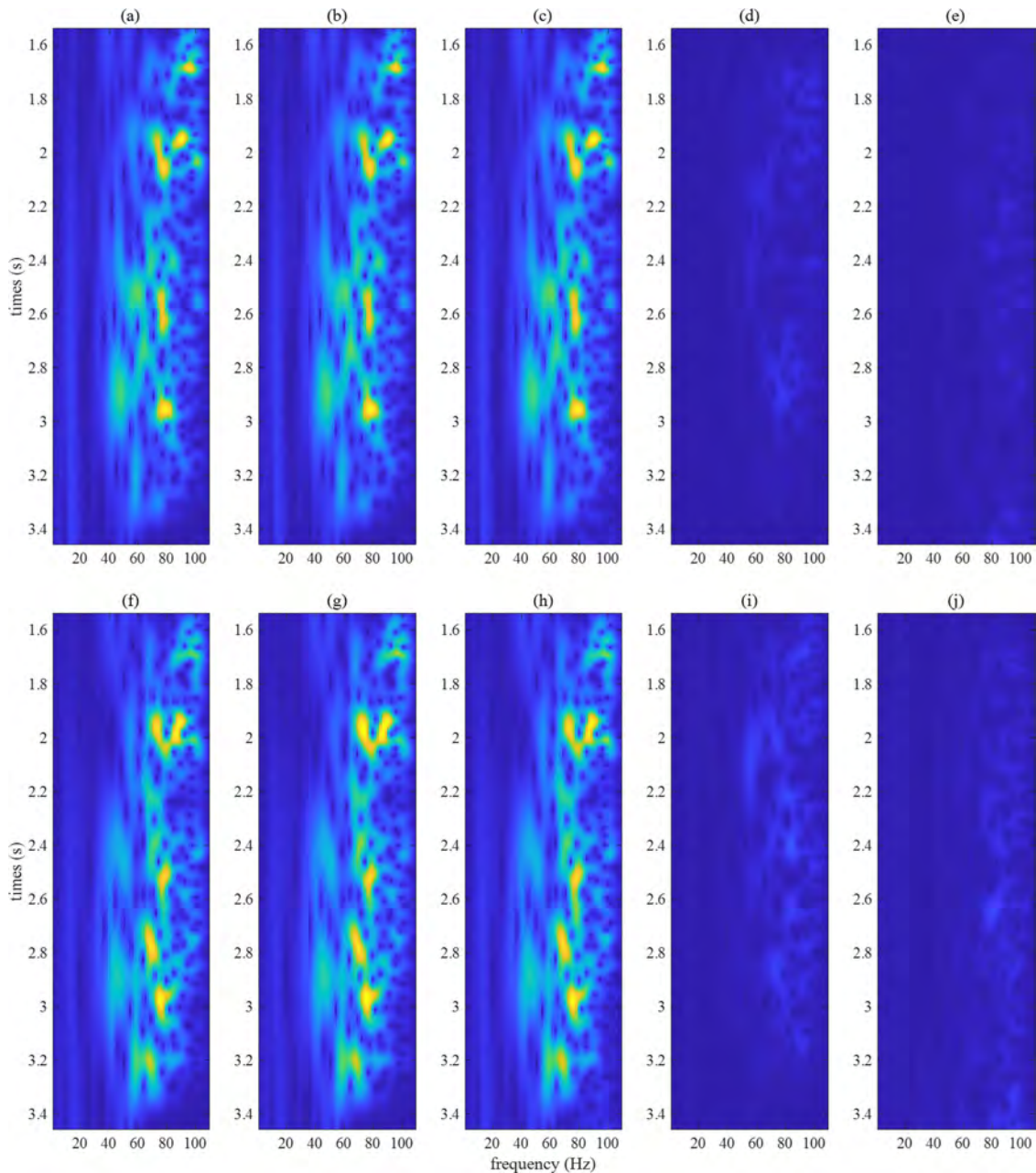


Fig. 10. Scalograms using continuous wavelet transform (CWT) at $XL = 16.65$ (a-e) and 16.85 (f-j) locations in Fig. 3b. (a, f) Ground truth. (b, g) Prediction using the proposed method. (c, h) Prediction using the MWNI method. (d, i) Difference between ground truth and prediction using the proposed method. (e, j) Difference between ground truth and prediction using the MWNI method.

너지 이외의 분산된 진폭의 누출이 없는 반면, 불규칙적인 빠짐을 포함하는 단면의 주파수-파수 스펙트럼(Fig. 9b)는 주된 에너지 주변에 진폭의 누출이 분산되어 나타나는 것을 확인할 수 있다. 이러한 누출은 제안된 방법의 결과(Fig. 9c)와 모델 제약조건이 적용된 MWNI 방법의 결과(Fig. 9d)에서 구한 두 주파수-파수 스펙트럼 모두에서 성공적으로 제거되었으며, 두 방법 모두 2차원 스펙트럼을 안정적으로 복구한 것을 확인했다.

마지막으로 트레이스에 존재하는 여러 요소파(wavelet)들이 포함하는 주파수 성분들에 따른 복원 결과의 차이가 존재하는지 확인 해보기 위하여 제안된 방법과 모델 제약조건이 적용된 MWNI 방법에 의해 복원된 자료의 XL = 16.65 km (Fig. 3b의 빨간색 화살표), 16.85 km (Fig. 3b의 파란색 화살표) 두 위치에서 1차원 트레이스에 대해 각각 연속 요소파 변환(Continuous Wavelet Transform; Lilly and Olhede, 2012)을 수행한 스케일로그래프(scalogram)을 구해 Fig. 10에 나타내었다. 스케일로그래프는 단주기 요소파 변환으로 특정 시간대에 존재하는 파형이 가지는 주파수 성분을 분석한 결과를 보여준다. 그림에서 볼 수 있듯이 두 지점에서 얻은 스케일로그래프 모두에서 모든 주파수에서 고른 성능을 확인할 수 있었다. 특히 이 분석을 위해 선정한 두 지점은 각각 $R_{(1,2)}$, $R_{(2,2)}$ 을 이용하여 예측된 결과로 비교적 성능이 떨어졌던 네트워크에서도 특정 주파수나 특정 시간대의 성능이 크게 떨어지지 않는 안정된 스펙트럼을 모든 시간대에서 만들어 내는 것을 확인하였다.

결론 및 토의

이 연구에서는 중점의 위치의 트레이스를 예측할 수 있도록 개발된 트레이스 기반 접근방법(Yoon *et al.*, 2021)이 적용된 기계학습 기반의 트레이스 내삽 방법을 중점의 위치가 아닌 위치의 트레이스를 예측할 수 있도록 확장하였다. 제안된 방법은 대부분의 머신러닝 기반 내삽방법들과는 달리 네트워크의 학습에 모든 자료가 채워진 별도의 빠짐이 없는 모음자료가 아닌 내삽 해야 할 대상인 불규칙한 빠짐이 있는 탄성파 현장 자료만이 필요하다는 장점을 가지며, 서로 다른 네트워크를 사용하여 연속된 빠짐도 채울 수 있게 개발되었다. 확장된 방법의 성능을 확인해 보기 위하여 서호주 Exmouth Sub-basin 지역의 Vincent 유전에서 얻은 시간 참반사 보정된 탄성파 자료를 사용하여 수치실험을 수행하였다. 수치실험에서는 원 자료에 연속적인 빠짐을 포함하는 불규칙적인 빠짐을 만든 후 이를 내삽 하여 원 자료와 비교하였다. 또한 전통적인 내삽 방법인 모델 제약조건이 적용된 MWNI와 비교하여 제안된 방법의 성능을 검증하였다.

현장자료를 활용한 수치실험 결과 두 방법 모두 높은 성능을 보인다는 것을 확인할 수 있었다. 보다 정량적인 비교를 위하여, 신호대잡음비 및 구조유사성 값으로 비교를 한 결과도 전체 자료를 동시에 연산에 사용하는 모델 제약조건이 적용된

MWNI 방법이 조금 더 높은 성능을 나타내나 두 방법 모두에서 수치상으로도 성공적인 내삽 결과를 보여준다는 것을 알 수 있었다. 또한 주파수-파수 스펙트럼 및 연속 요소파 변환을 통해 주파수 성분에 따른 내삽 성능 차이를 확인한 결과, 모든 주파수에서 고른 성능을 보임도 확인할 수 있었다. 하지만 입력으로 사용하는 모음에 트레이스의 수가 제한적일 때는 모델 제약조건이 적용된 MWNI 방법은 적용이 어려워 본 연구에서 제안된 방법이 더 효과적일 것이라 예상된다.

이 연구에서 제안한 방법은 규칙적인 격자 구조의 탄성파 자료에 불규칙한 빠짐이 있는 경우에도 효과적인 내삽을 제공한다. 그러나 더 높은 수준의 트레이스 정규화를 위해 격자 위에 있지 않는 관계를 학습하는 방법이 추가로 필요하다. 또한, 연속된 빠짐의 크기가 커질수록 네트워크의 크기가 기하급수적으로 커지는 문제가 있다. 일반적으로 관심영역 구간에 연속된 빠짐의 크기가 커지면 주변의 소수의 자료만으로 내삽을 수행하기는 어려워져서 머신러닝을 이용하는 내삽의 경우도 모델 제약조건이 적용된 MWNI와 같이 전체자료를 활용하는 아이디어를 사용할 수밖에 없다. 그러나 이 경우를 제외하면, 향후 어느 정도의 연속된 빠짐이나 별도의 자료처리 없이 실제 자료의 측정 위치를 활용할 수 있는 세번째 불규칙도를 하나의 네트워크로 해결할 수 있는 방향으로 이 연구를 발전시킬 필요가 있다.

감사의 글

이 논문(또는 특허 등)은(는) 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1A2C2014315).

References

- Abma, R., and Kabir, N., 2006, 3D interpolation of irregular data with a POCS algorithm, *GEOPHYSICS*, **71(6)**, E91-E97, doi: <https://doi.org/10.1190/1.2356088>
- Bourlard, H., and Kamp, Y., 1988, Auto-association by multilayer perceptrons and singular value decomposition, *Biol. Cybern.*, **59(4-5)**, 291-294, doi: <https://doi.org/10.1007/BF00332918>
- Choi, J., Byun, J., Seol, S. J., and Kim, Y., 2016, Wavelet-based multicomponent matching pursuit trace interpolation, *Geophys. J. Int.*, **206(3)**, 1831-1846, doi: <https://doi.org/10.1093/gji/ggw246>
- Choi, J., Song, Y., Choi, J., Byun, J., Seol, S. J., Kim, K., and Lee, J., 2017, Trace Interpolation using Model-constrained Minimum Weighted Norm Interpolation, *Geophys. Geophys. Explor.*, **20(2)**, 78-87, doi: <https://doi.org/10.7582/GGE.2017.20.2.078> (In Korean with English abstract)
- Dong, C., Loy, C. C., He, K., and Tang, X., 2016, Image Super-

- Resolution Using Deep Convolutional Networks, *IEEE Trans. Pattern Anal. Mach. Intell.*, **38(2)**, 295-307, doi: <https://doi.org/10.1109/TPAMI.2015.2439281>
- Dou, Y., Li, K., Duan, H., Li, T., Dong, L., and Huang, Z., 2022, MDA GAN: Adversarial-Learning-based 3-D Seismic Data Interpolation and Reconstruction for Complex Missing, *IEEE Trans. Geosci. Remote Sens.*, **61**, 1-14, doi: <https://doi.org/10.1109/TGRS.2023.3249476>
- Fang, W., Fu, L., Wu, M., Yue, J., and Li, H., 2023, Irregularly sampled seismic data interpolation with self-supervised learning, *GEOPHYSICS*, **88(3)**, 1-43, doi: <https://doi.org/10.1190/geo2022-0586.1>
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., 2020, Generative adversarial networks, *Commun. ACM*, **63(11)**, 139-144, doi: <https://doi.org/10.1145/3422622>
- Hore, A., and Ziou, D., 2010, Image Quality Metrics: PSNR vs. SSIM, *2010 20th International Conference on Pattern Recognition*, IEEE, 2366-2369, doi: <https://doi.org/10.1109/ICPR.2010.579>
- Jia, Y., and Ma, J., 2017, What can machine learning do for seismic data processing? An interpolation application, *GEOPHYSICS*, **82(3)**, V163-V177, doi: <https://doi.org/10.1190/geo2016-0300.1>
- Kim, B., Jeong, S., and Byun, J., 2015, Trace interpolation for irregularly sampled seismic data using curvelet-transform-based projection onto convex sets algorithm in the frequency-wavenumber domain, *J. Appl. Geophys.*, **118**, 1-14, doi: <https://doi.org/10.1016/j.jappgeo.2015.04.007>
- Lilly, J. M., and Olhede, S. C., 2012, Generalized Morse Wavelets as a Superfamily of Analytic Wavelets, *IEEE Trans. Signal Process.*, **60(11)**, 6036-6041, doi: <https://doi.org/10.1109/TSP.2012.2210890>
- Liu, B., and Sacchi, M. D., 2004, Minimum weighted norm interpolation of seismic records, *GEOPHYSICS*, **69(6)**, 1560-1568, doi: <https://doi.org/10.1190/1.1836829>
- Oliveira, D. A. B., Ferreira, R. S., Silva, R., and Brazil, E. V., 2018, Interpolating Seismic Data With Conditional Generative Adversarial Networks, *IEEE Geosci. Remote Sens. Lett.*, **15(12)**, 1952-1956, doi: <https://doi.org/10.1109/LGRS.2018.2866199>
- Pan, S., Chen, K., Chen, J., Qin, Z., Cui, Q., and Li, J., 2020, A partial convolution-based deep-learning network for seismic data regularization, *Comput. Geosci.*, **145**, 104609, doi: <https://doi.org/10.1016/J.CAGEO.2020.104609>
- Park, J., Yeoh, Z., Seol, S. J., and Byun, J., 2022, Seismic Data Interpolation Using Attention-Based Deep Learning, *83rd EAGE Annual Conference & Exhibition*, European Association of Geoscientists & Engineers, **2**, 1-5, doi: <https://doi.org/10.3997/2214-4609.202210245>
- Park, J., Choi, J., Seol, S. J., Byun, J., and Kim, Y., 2021, A method for adequate selection of training data sets to reconstruct seismic data using a convolutional U-Net, *GEOPHYSICS*, **86(5)**, V375-V388, doi: <https://doi.org/10.1190/geo2019-0708.1>
- Spitz, S., 1991, Seismic trace interpolation in the F-X domain, *GEOPHYSICS*, **56(6)**, 785-794, doi: <https://doi.org/10.1190/1.1443096>
- Trad, D., 2009, Five-dimensional interpolation: Recovering from acquisition constraints, *GEOPHYSICS*, **74(6)**, V123-V132, doi: <https://doi.org/10.1190/1.3245216>
- Trad, D., 2014, Five-dimensional interpolation: New directions and challenges, *CSEG Rec.*, **39(3)**, 22-29, <https://csegrecorder.com/articles/view/five-dimensional-interpolation-new-directions-and-challenges>
- Vassallo, M., Özbek, A., Özdemir, K., and Eggenberger, K., 2010, Crossline wavefield reconstruction from multicomponent streamer data: Part 1 — Multichannel interpolation by matching pursuit (MIMAP) using pressure and its crossline gradient, *GEOPHYSICS*, **75(6)**, WB53-WB67, doi: <https://doi.org/10.1190/1.3496958>
- Wang, B., Zhang, N., Lu, W., Geng, J., and Huang, X., 2020a, Intelligent Missing Shots' Reconstruction Using the Spatial Reciprocity of Green's Function Based on Deep Learning, *IEEE Trans. Geosci. Remote Sens.*, **58(3)**, 1587-1597, doi: <https://doi.org/10.1109/TGRS.2019.2947085>
- Wang, Y., Wang, B., Tu, N., and Geng, J., 2020b, Seismic trace interpolation for irregularly spatial sampled data using convolutional autoencoder, *GEOPHYSICS*, **85(2)**, V119-V130, doi: <https://doi.org/10.1190/geo2018-0699.1>
- Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., 2004, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, **13(4)**, 600-612, doi: <https://doi.org/10.1109/tip.2003.819861>
- Xu, S., Zhang, Y., and Lambaré, G., 2010, Antileakage Fourier transform for seismic data regularization in higher dimensions, *GEOPHYSICS*, **75(6)**, WB113-WB120, doi: <https://doi.org/10.1190/1.3507248>
- Yeoh, Z., Byun, J., and Yoon, D., 2020a, Crossline interpolation with the traces-to-trace approach using machine learning, *SEG Technical Program Expanded Abstracts 2020*, Society of Exploration Geophysicists, **2020-October**, 1656-1660, doi: <https://doi.org/10.1190/segam2020-3428348.1>
- Yeoh, Z., Song, Y., Byun, J., Seol, S. J., and Kim, K. Y., 2020b, Regularization of multidimensional sparse seismic data using Delaunay tessellation, *J. Appl. Geophys.*, **174**, 103877, doi: <https://doi.org/10.1016/J.JAPPGEO.2019.103877>
- Yoon, D., Yeoh, Z., and Byun, J., 2021, Seismic Data Reconstruction Using Deep Bidirectional Long Short-Term Memory with Skip Connections, *IEEE Geosci. Remote Sens. Lett.*, **18(7)**, 1298-1302, doi: <https://doi.org/10.1109/LGRS.2020.2993847>

부록: 빠짐을 크기별로 분류하는 알고리즘의 의사 코드

1. Initialize `index_list_step` as `index_list`
2. Initialize `gap_count` as 1
3. Initialize `max_gap_count` as 0
4. Initialize `gap_index_groups` as an empty dictionary
5. While `index_list_step` is not empty
6. Initialize `index_list_previous_step` as an empty list
7. Initialize `index_list_next_step` as an empty list
- 8.
9. For `i` in range from 0 to length of `index_list_step` - 1
10. If `index_list_step[i + 1] == index_list_step[i] + 1`
11. Append `index_list_step[i]` to `index_list_previous_step`
12. Else
13. Append `index_list_step[i]` to `index_list_next_step`
- 14.
15. Store `index_list_previous_step` in `gap_index_groups` with key `gap_count`
16. Increase `gap_count` by 1
17. Set `index_list_step` to `index_list_next_step`
18. Set `max_gap_count` as the maximum key in `gap_index_groups`
19. Return `max_gap_count`, `gap_index_groups`