Article

Automatic Horizon Picking Using Multiple Seismic Attributes and Markov Decision Process

Chengliang Wu [1,2] , **Bo Feng [1,2,*]**, **Xiaonan Song [2]**, **Huazhong Wang [1,2]**, **Rongwei Xu [2]** and **Shen Sheng [2]**

[1] State Key Laboratory of Marine Geology, School of Ocean and Earth Science, Tongji University, Shanghai 200092, China
[2] Wave Phenomena and Intelligent Inversion Imaging Group (WPI), School of Ocean and Earth Science, Tongji University, Shanghai 200092, China
* Correspondence: 16666012@tongji.edu.cn

Abstract: Picking the reflection horizon is an important step in velocity inversion and seismic interpretation. Manual picking is time-consuming and no longer suitable for current large-scale seismic data processing. Automatic algorithms using different seismic attributes such as instantaneous phase or dip attributes have been proposed. However, the computed attributes are usually inaccurate near discontinuities. The waveforms in the horizontal direction often change dramatically, which makes it difficult to track a horizon using the similarity of attributes. In this paper, we propose a novel method for automatic horizon picking using multiple seismic attributes and the Markov decision process (MDP). For the design of the MDP model, the decision time and state are defined as the horizontal and vertical spatial position on a seismic image, respectively. The reward function is defined in multi-dimensional feature attribute space. Multiple attributes can highlight different aspects of a seismic image and therefore overcome the limitations of the single-attribute MDP through the cross-constraint of multiple attributes. The optimal decision is made by searching the largest state value function in the reward function space. By considering cumulative reward, the lateral continuity of a seismic image can be effectively considered, and the impacts of abnormal waveform changes or bad traces in local areas for automatic horizon picking can be effectively avoided. An effective implementation scheme is designed for picking multiple reflection horizons. The proposed method has been successfully tested on both synthetic and field data.

Keywords: seismic image; horizon picking; multiple seismic attributes; Markov decision process

1. Introduction

In exploration seismology, the reflection horizons in imaging volumes play an important role in velocity inversion and seismic interpretation. For velocity inversion, the reasonable application of the reflection horizon can effectively improve the convergence and obtain a geologically plausible velocity model. For seismic interpretation, the main reflection horizon can determine the relationship between geological layers and improve the reliability and accuracy of seismic interpretation. Therefore, automatically and efficiently picking the horizons from a complex seismic image is very important.

There are two main methods for tracking horizons. The first method is to identify and extract the horizon position using various signal or image processing methods based on various seismic attributes. For example, we can use the amplitude attribute [1], phase attribute [2–6], waveform attribute [7–9], and slope or dip attribute [10–17] to track the horizon. Furthermore, multiple attributes can be combined to constrain the picking results [18–23], such as using the instantaneous phase and dip attributes [18,22] and using the envelope, phase attributes, and texture analysis [23]. In addition, there are also various tracking algorithms for horizon picking. Admasu and Tönnie [24] used the correlation of horizons across faults in 3D seismic data, and Zou et al. [9] tracked the horizon based on the waveform similarity. Bondár [25] used the edge detection algorithm to detect seismic

horizons. Aurnhammer and Tonnies [26] used a genetic algorithm for automated horizon correlation across faults in seismic images. Many global methods [27–29] or the flattening method [12–14,30–34] are also widely used for automated horizon picking.

The second method is mainly based on the Machine Learning (ML) algorithm to intelligently track the horizon; it mainly includes supervised learning methods with artificial neural networks [35–39], convolutional neural networks [40–45], and unsupervised learning methods with self-organizing networks [46] and convolutional auto-encoding networks [47].

In the first method, conventional horizon-picking methods depend heavily on the seismic attribute. When the attribute changes in the horizontal direction are relatively dramatic, it is difficult to track the correct horizon position. With the help of big data and ML methods, we can learn the structural information of the seismic image according to the data's statistical characteristics. However, supervised learning methods need a lot of training data and labels, and the results obtained using unsupervised network algorithms are often unreliable.

If the problem of automatic horizon picking is regarded as an optimal decision-making process, a kind of ML method named reinforcement learning, the core algorithm of which is based on the Markov decision process (MDP), can be used to replace manual picking by imitating human behavior. The MDP based on the Markov chain can perceive the state change in the system environment and simulate humans to make a decision. Wald [48] adopted the MDP to solve the sequential analysis problem. Shapley [49] introduced a Markov decision process when studying random games. Bellman [50] put forward the discrete-time MDP (Bellman equation) for dynamic programming, which established a complete methodology of MDP. With the improvement of the MDP theory and method [51–53], the MDP method has been widely used in sequential decision making [54–56], dynamic programming [53,57], and reinforcement learning [58,59]. In the exploration seismology, Ma et al. [60] introduced reinforcement learning for selecting the first arrival. Luo et al. [61] designed a Markov decision process for first break picking.

In this paper, we formulate the horizon picking of seismic images as a Markov decision process (MDP) with multiple seismic attributes. The paper is organized as follows: Firstly, this paper briefly reviews the basic concepts and algorithms of the MDP. Then, the finite-stage MDP is designed for automatic horizon picking. For the MDP, the decision time is defined as the horizontal spatial position, and the state is defined as the vertical spatial position on the seismic image. The reward function is defined in multi-dimensional feature attribute space including the waveform attribute, instantaneous phase attribute, envelope attribute, and extremum attribute. When the seed point is determined, we can perceive the state change and take some actions to obtain reward value, where the selected action depends on the dip attribute. Therefore, we can make the optimal decision for horizon picking through constant trial and error and interaction with the environment. Moreover, for picking multiple reflection horizons, we design an effective implementation scheme using multiple seismic attributes and MDP. Finally, synthetic and field data are used to verify the effectiveness of the proposed method.

2. Review of the Markov Decision Process

2.1. 5-Tuple $<s, a, \mu, p, r>$

The Markov decision process (MDP) mainly includes the following elements: state (s), action (a), policy (μ), transition probability (p), and reward function (r).

–*State*: The state s_i^t is defined as one of the system characteristics at the decision-making time t. If a state does not change with time, the state becomes s_i, called a stable state. State space $\mathbf{S}$ is defined as a finite set $\mathbf{S} = \{s_1, s_2, \ldots, s_M\}$, where M is the size of state space.

–*Action*: The action a is the response of a system to the state s. For a certain state s_i, the set of all possible actions can be expressed as $\mathbf{A}_{s_i} = \{a_1(s_i), a_2(s_i), \ldots, a_n(s_i)\}$. The action space of a system is defined as the union composed of action sets under all states $\mathbf{A} = U_{s_i \in s} = \{\mathbf{A}_{s_1}, \mathbf{A}_{s_i}, \ldots, \mathbf{A}_{s_M}\} = \{a_1, a_2, \ldots, a_N\}$, where N is the size of action space.

–Policy: The mapping process from state to action is called policy $\mu : s \rightarrow a(s)$. The policy depends on the whole decision process, and the sequence formed by the mapping of state to action at all times is called a policy π:

$$\pi = \left(\mu_{i,k}^0, \mu_{i,k}^1, \mu_{i,k}^2, \ldots, \mu_{i,k}^{t-1}, \mu_{i,k}^t, \ldots, \mu_{i,k}^{T-1}, \mu_{i,k}^T \right), \tag{1}$$

where $\mu_{i,k}^t = \mu\left(s_i^t, a_k^t \right)$ is the probability of mapping state s_i^t to action a_k^t. The policy space is defined as a finite set $\Pi = \{\pi_1, \pi_2, \ldots, \pi_L\}$, where L is the size of the policy space.

–Transition probability: When performing an action $a_k^t \in \mathbf{A}$ from one state $s_i^t \in \mathbf{S}$ at the time t, the system makes a transition to a new state $s_j^{t+1} \in \mathbf{S}$, where the transition is based on a probability distribution $p\left(s_j^{t+1} \middle| s_i^t, a_k^t \right) : s \times a \times s \rightarrow [0, 1]$. The transition probability of the system from the state s_i^t to the state s_j^{t+1} is:

$$p\left(s_j^{t+1} \middle| s_i^t \right) = \sum_{a_k^t} \mu\left(s_i^t, a_k^t \right) p\left(s_j^{t+1} \middle| s_i^t, a_k^t \right). \tag{2}$$

According to Markov property, the state of a system at the time $t+1$ only depends on the state at the time t and is independent of the previous state:

$$p\left(s^{t+1} \middle| s^1, s^2, \ldots, s^t \right) = p\left(s^{t+1} \middle| s^t \right). \tag{3}$$

–Reward Function: The reward function $r\left(s_i^t, a_k^t \right)$ is defined as $R : s \times a \rightarrow \mathbb{R}$,

$$r\left(s_i^t, a_k^t \right) = \sum_{s_j^{t+1}} p\left(s_j^{t+1} \middle| s_i^t, a_k^t \right) r\left(s_j^{t+1} \middle| s_i^t, a_k^t \right), \tag{4}$$

where $r\left(s_j^{t+1} \middle| s_i^t, a_k^t \right)$ is the reward function when the system executes action a_k^t under the state s_i^t at the time t, then transfers to the state s_j^{t+1}. The reward function is only related to the current state and action and has nothing to do with the previous state and action.

2.2. The Markov Decision Process

By integrating the above elements, we can define the MDP model, which can be viewed in Figure 1. This process can be completely described as: at the time t, the system is in the state s_i, and the decision-maker executes action a_k and generates reward value $r\left(s_i^t, a_k^t \right)$ and then transfers to the next state s_j^{t+1} with probability function $p\left(s_j^{t+1} \middle| s_i^t, a_k^t \right)$. The model forms the complete MDP time series at all times, which can be expressed as:

$$\tau = \left(s^0, a^0, r^1; \ldots; s^{t-1}, a^{t-1}, r^t; s^t, a^t, r^{t+1}; \ldots; s^T, a^T, r^{T+1} \right). \tag{5}$$

Taking a decision (with an action) according to policy $\mu_{i,k}^t$ from the state s_i^t at the time t, the system can obtain a reward sequence at the subsequent time,

$$\left\{ R^{t+1}, R^{t+2}, \ldots, R^T \right\}, \tag{6}$$

where $R^{t+1} = r\left(s_i^t, a_k^t \right)$ is the immediate reward function. The overall expectation from the immediate reward function R^{t+1} to the future reward function sequence R^T with the policy π is defined as value function $v_\pi\left(s^t \right)$:
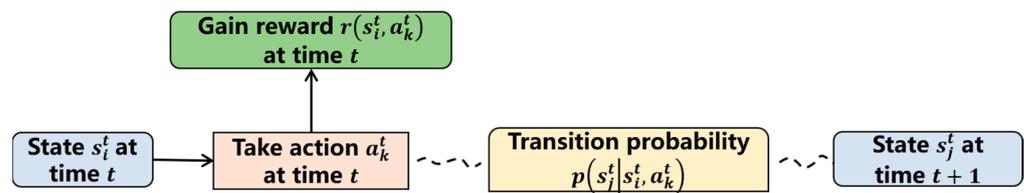
Figure 1. The MDP model.

$$v_\pi\left(s^t\right) = E_\pi\left[\sum_{k=0}^{T} \gamma^k R^{t+k+1}|s^t\right], \quad \forall s^t \in \mathbf{S} \quad 0 < \gamma < 1, \tag{7}$$

where E_π is the expectation and γ is the discount factor, and $0 < \gamma < 1$, which means that the further the reward is from the decision-making time, the smaller the impact on the value function. Based on the Markov property, Formula (7) can be written as:

$$v_\pi\left(s^t\right) = \sum_{a^t \in \mathbf{A}} \mu\left(s^t, a^t\right)\left(r\left(s^t, a^t\right) + \gamma \sum_{s^{t+1} \in \mathbf{S}} p\left(s^{t+1}\middle|s^t, a^t\right)v_\pi\left(s^{t+1}\right)\right). \tag{8}$$

This equation is the so-called Bellman equation [50] or the state transition equation. Based on the Bellman equation, the state value function at the current time can be calculated by the state value function of the next decision time.

The value function is used to evaluate how "good" a decision-maker is in a given state [59]. The goal of the MDP model is to find the best policy that can maximize the value function. Therefore, an optimal policy π_* is defined as $v_{\pi_*}\left(s^t\right) \geq v_\pi\left(s^t\right)$ for all $s^t \in \mathbf{S}$ and all policies $\pi \in \mathbf{\Pi}$. The optimal solution $v_*\left(s^t\right) = \max_\pi v_\pi\left(s^t\right)$ satisfies the following Bellman optimality equation:

$$v_*\left(s^t\right) = \max_{a^t \in \mathbf{A}}\left\{r\left(s^t, a^t\right) + \gamma \sum_{s^{t+1} \in \mathbf{S}} p_\pi\left(s^{t+1}\middle|s^t\right)v_*\left(s^{t+1}\right)\right\}. \tag{9}$$

Once the MDP model is determined, an appropriate and effective solution method can be selected to obtain the optimal policy, for example, the value iteration method [50], the policy iteration method [51], the Q learning algorithm [58], or the SARSA [62].

3. Formulation of Horizon Picking with MDP

For the problem of automatic horizon picking on seismic images, we can regard the seed point of one reflection horizon as the initial state and take some action policies to track the horizon. Based on the reward function of the current state (vertical position) and action, we can find an optimal policy to transfer to a new state at the next decision time (horizontal position). The state and transition probability distribution only depend on the current position and have nothing to do with the state and action of the previous position. Therefore, we regard the automatic horizon picking on the seismic image as an intelligent decision-making problem in a random environment which can be modeled as a discrete MDP with finite stages.

The horizon-picking model with finite-stage MDP is defined as follows:

(1) Decision time

The decision time $t \in \mathbf{T} = \{1, 2, \ldots, nx\}$ is defined as the horizontal spatial position where the vertical and horizontal grid number is nz and nx, respectively, for a 2D seismic image.

According to the position of the seed point, the decision-making time can be divided into three categories:

(a) The starting time is at the left endpoint $t_0 = 1$, and the ending time is at the right endpoint $t_n = nx$. The decision time sequence is $\mathbf{T} = \{1, 2, \ldots, nx - 1, nx\}$.

(b) The starting time is at the right endpoint $t_0 = nx$, and the ending time is at the left endpoint $t_n = 1$. The decision time sequence is $\mathbf{T} = \{nx, nx - 1, \ldots, 2, 1\}$.

(c) The starting time is at the middle position $t_0 = i$. The decision time is divided into two parts: one is to track to the right endpoint $t_n = nx$(consistent with mode a); the second is to track to the left endpoint $t_n = 1$(consistent with mode b).

(2) State

The state is defined as the vertical spatial position on a seismic image. The set of states is represented as $\mathbf{S} = [1, \ldots, nz]$. A seismic image often contains multiple horizons. When we pick one of them, the rest are irrelevant. The farther away an image is from the current axis in the vertical direction, the less influence it has on the decision. Moreover, we reduce the state space to $\mathbf{S} = \left[z_0^t - dz, \ldots, z_0^t + dz\right]$, where dz is the width. z_0^t is the middle position of the state space, which is determined by the conventional picking method. Particularly, z_0^0 is the seed point. The reduction in state space can not only weaken the impact of another horizon, but also improve computational efficiency.

(3) Action

The action is defined as a movement from a lateral position x_i to a lateral position x_{i+1}, $a(s) : x_i \rightarrow x_{i+1}$.

(4) Policy

We adopt the stochastic policy to realize the mapping from state to action under the constraint of the dip attribute

$$\mu^t(s, a) = e^{-\frac{(a - a_0)^2}{\sigma^2}}, \quad a_0 = s + p\frac{dz}{dx}, \tag{10}$$

where p is the dip at the position of the state s, and dx and dz are the sampling intervals in the lateral and vertical directions, respectively. σ is the variance in the Gaussian probability distribution. Figure 2 shows the mapping probability diagram from state to action. Following the dip direction, the probability of action is high. Staying away from the dip direction, the probability of action slowly becomes smaller. When the state position changes, the corresponding action probability density function also changes.
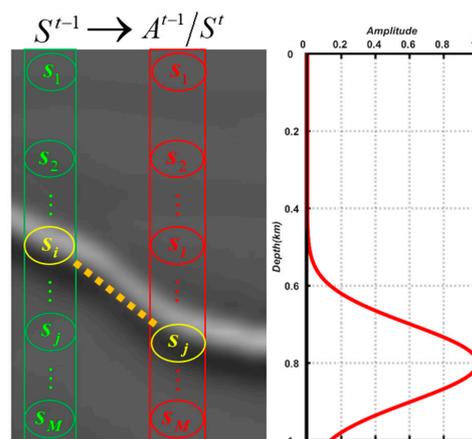

Figure 2. The diagram of the mapping process of the state to action. The background profile in the left figure is a local seismic image. The right figure shows the probability distribution followed by the relationship between state and action where the position of the current state at the time $t - 1$ is located at s_i, and the state will move to next time t. The yellow dashed line is the direction corresponding to the dip angle calculated at the position s_i. The probability at s_j is maximum. Away from s_j, the probability is getting smaller.

(5) Transition probability

For state $s_i^t \in \mathbf{S}$ and action $a_k^t \in \mathbf{A}$ at time t, the state $s_j^{t+1} \in \mathbf{S}$ at time $t+1$ is relatively independent. The transition probability is defined as

$$p\left(s_j^{t+1} \middle| s_i^t, a_k^t\right) = 1. \tag{11}$$

Therefore, the transition probability of the system from the state s_i^t to the state s_j^{t+1} is:

$$p\left(s_j^{t+1} \middle| s_i^t\right) = \sum_{a_k^t} \mu\left(s_i^t, a_k^t\right). \tag{12}$$

(6) Reward function

In the MDP, the optimal state is found according to the value function. However, the value function depends on the reward function. The reward function is the most important part of the MDP, which implicitly defines the learning objective [59]. In this paper, we use multiple seismic attributes of the seismic image to define the reward function.

Based on Formula (4) and (11), the relationship between the reward function $r\left(s_i^t, s_j^{t+1}\right)$ and $r^t\left(s_j^{t+1} \middle| s_i^t, a_k^t\right)$ is as follows:

$$r\left(s_i^t, s_j^{t+1}\right) = r\left(s_j^{t+1} \middle| s_i^t, a_k^t\right). \tag{13}$$

Therefore, the reward function is defined as

$$r\left(s_i^t, s_j^{t+1}\right) = \sum_{f \in F} \alpha_f r_f\left(s_i^t, s_j^{t+1}\right) \quad s.t. \quad \sum_{f \in F} \alpha_f = 1, \tag{14}$$

$$r_f\left(s_i^t, s_j^{t+1}\right) = R_f \times F(\Delta f), \tag{15}$$

where $\Delta f = \|f(s_i) - f(s_j)\|^2$ is the distance between the state s_i and s_j for the attribute f. The operator $F(\bullet)$ is a measure of attribute difference. R_f is the maximum reward value, $r_f\left(s_i^t, s_j^{t+1}\right)$ is the reward value, and α_f is the weight coefficient of the attribute f. The weighting coefficient is determined by the Fuzzy C-Means (FCM) clustering method, where the result is obtained by using the conventional horizon-picking method with a single attribute.

In this paper, the following attributes are selected to calculate the reward function:

(a) The waveform similarity for $r_{waveform}\left(s_i^t, s_j^{t+1}\right)$:

$$r_{waveform}\left(s_i^t, s_j^{t+1}\right) = R_{waveform} \times \overline{\gamma}\left(s_i^t, s_j^{t+1}\right)$$
$$\begin{cases} \overline{\gamma} = \frac{\sum_{k=1}^n \gamma_k}{n} \\ \gamma_k = \frac{\langle s^t(z), s^{t+1}(z) \rangle}{\sqrt{\langle s^t(z), s^t(z) \rangle}\sqrt{\langle s^{t+1}(z), s^{t+1}(z) \rangle}} \\ \langle s^t(z), s^{t+1}(z) \rangle = \int_{-\Delta z - k}^{\Delta z + k} s^t(z) s^{t+1}(z) dz \end{cases} \tag{16}$$

where Δz is the size of the local space window, and γ_k is the similarity coefficient.

(b) The extremum attribute for $r_{et}\left(s_i^t, s_j^{t+1}\right)$:

$$r_{et}\left(s_i^t, s_j^{t+1}\right) = \begin{cases} 1, & if\ f\left(s_i^t\right)\ is\ \text{extremum} \\ 0, & else \end{cases} \tag{17}$$

where the attribute f includes the waveform attribute, instantaneous phase attribute, and envelope attribute.

The instantaneous attributes of the seismic image can be obtained by applying a Hilbert transform to generate complex signals. For a 2-D seismic image $I(z, x)$, the $I(z, x)$ is treated like a real part of the complex signals, and the imaginary part $\widetilde{I}(z, x)$ can be obtained after the Hilbert transformation. The complex signals $\hat{I}(z, x)$ are:

$$\hat{I}(z, x) = I(z, x) + i\widetilde{I}(z, x). \tag{18}$$

Then, the envelope $e(z, x)$ and the instantaneous phase $\theta(z, x)$ can be obtained by

$$e(z, x) = \sqrt{I^2(z, x) + \widetilde{I}^2(z, x)}. \tag{19}$$

$$\theta(z, x) = arctan\left(\frac{\widetilde{I}(z, x)}{I(z, x)}\right). \tag{20}$$

However, the range of the instantaneous phase solved by Equation (20) has discontinuity. Based on the relationship between the seismic image and its transformation result:

$$\begin{cases} I(z, x) = e(z, x)cos(\theta(z, x)) \\ \widetilde{I}(z, x) = e(z, x)sin(\theta(z, x)) \end{cases}. \tag{21}$$

The instantaneous phase can also be represented as

$$cos(\theta(z, x)) = \frac{I(z, x)}{e(z, x)}. \tag{22}$$

Furthermore, we use the least squares inverse division [63] to avoid the instability caused by zero division:

$$cos(\theta(z, x)) = \arg min\|I(z, x) - cos(\theta(z, x))e(z, x)\|^2. \tag{23}$$

Finally, the instantaneous phase is calculated as

$$cos(\theta(z, x)) = \frac{I(z, x)e(z, x)}{e(z, x)e(z, x) + \varepsilon}, \tag{24}$$

where ε is a threshold coefficient.

Then, we use the structure tensor method to calculate the dip attribute. Considering the waveform oscillation of seismic image, the structure tensor is obtained using the envelope attribute:

$$G(z, x) = \nabla e(z, x) \bullet \nabla e^t(z, x) = \begin{pmatrix} e_x^2(z, x) & e_x(z, x)e_z(z, x) \\ e_x(z, x)e_z(z, x) & e_z^2(z, x) \end{pmatrix}, \tag{25}$$

where ∇e represents the gradient of the envelope, and e_x and e_z are the directional derivatives in the x direction and z direction, respectively.

The eigenvalues and eigenvectors of the structure tensor matrix are obtained by eigenvalue decomposition of the structure tensor matrix:

$$\begin{cases} \lambda_1(z, x) = \frac{1}{2}\left(e_x^2(z, x) + e_z^2(z, x)\right) + \frac{1}{2}\sqrt{\left(e_x^2(z, x) - e_z^2(z, x)\right)^2 + 4(e_x(z, x)e_z(z, x))^2} \\ \lambda_2(z, x) = \frac{1}{2}\left(e_x^2(z, x) + e_z^2(z, x)\right) - \frac{1}{2}\sqrt{\left(e_x^2(z, x) - e_z^2(z, x)\right)^2 + 4(e_x(z, x)e_z(z, x))^2} \end{cases} \tag{26}$$

$$\begin{cases} v_1(z, x) = \begin{bmatrix} 2e_x(z, x)e_z(z, x) \\ e_z^2(z, x) - e_x^2(z, x) + \sqrt{\left(e_x^2(z, x) - e_z^2(z, x)\right)^2 + 4(e_x(z, x)e_z(z, x))^2} \end{bmatrix} \\ v_2(z, x) = \begin{bmatrix} -e_z^2(z, x) - e_x^2(z, x) - \sqrt{\left(e_x^2(z, x) - e_z^2(z, x)\right)^2 + 4(e_x(z, x)e_z(z, x))^2} \\ 2e_x(z, x)e_z(z, x) \end{bmatrix} \end{cases}, \tag{27}$$

where λ_1 and λ_2 are the eigenvalues, and v_1 and v_2 are the corresponding eigenvectors, respectively. λ_1 represents the tensor energy of the first eigenvector, which is consistent with the direction of the gradient. The orientations of the eigenvector v_1 and v_2 are:

$$\begin{cases} \varphi_1(z,x) = arctan\left(\frac{\lambda_1(z,x) - e_z^2(z,x)}{e_z(z,x)e_x(z,x)}\right) \\ \varphi_2(z,x) = arctan\left(\frac{e_z(z,x)e_x(z,x)}{\lambda_2(z,x) - e_x^2(z,x)}\right) \end{cases}, \tag{28}$$

where φ_1 is the gradient direction.

(7) Discount factor

Considering the impact on the subsequent limited-time steps, the discount factor is defined as

$$\gamma(k) = \begin{cases} e^{-\frac{k^2}{\sigma^2}}, & if \ k < N_t \\ 0, & else \end{cases}, \tag{29}$$

where N_t is the maximum time step, and $k = t - t_c$, t_c is the current time. Figure 3 shows the discount factor for $N_t = 10$. As the time step gets further away, the discount factor gets smaller, and the subsequent value function becomes less influential on the decision.
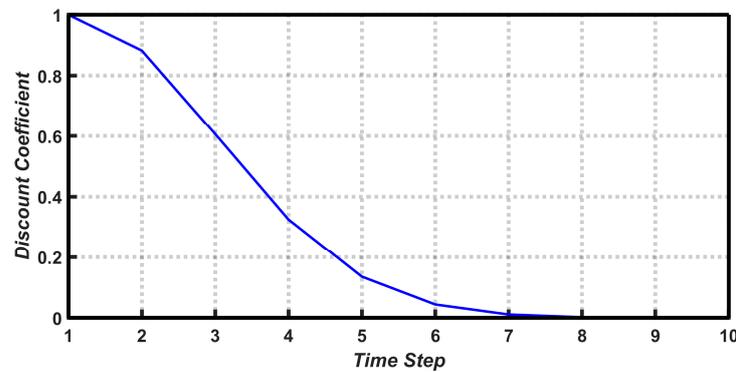
Figure 3. The discount factor for the MDP model.

4. An Effective Scheme for Horizon Picking

We designed an effective implementation scheme for automatically picking multiple horizons using multiple seismic attributes and MDP, shown in Figure 4. The scheme mainly includes five parts as follows:

(1) Preprocessing the seismic image.

The seismic image contains not only multiple horizons but also various coherent and incoherent noises. The proper preprocessing method can highlight wavelet characteristics and suppress some noises. We use the structural smoothing filtering method to process the seismic image.

(2) Determining the position of the initial seed point

We proposed two methods to obtain the seed point to adapt the different applications. One method is to rely on manual interpretation, and the second method is to use edge detection and clustering to obtain seed points automatically. For the automated method, firstly, we use an edge detection method to calculate edge points with different seismic attributes, and then cluster these edge points together to obtain the corresponding center point location. Last, the initial seed points are determined by the constraints of horizon interval and maximum horizon number. The second method is mainly aimed at the process of automatically determining seed points, such as extracting structural information in inversion iteration. However, for complex imaging results, we prefer manual interpretation.
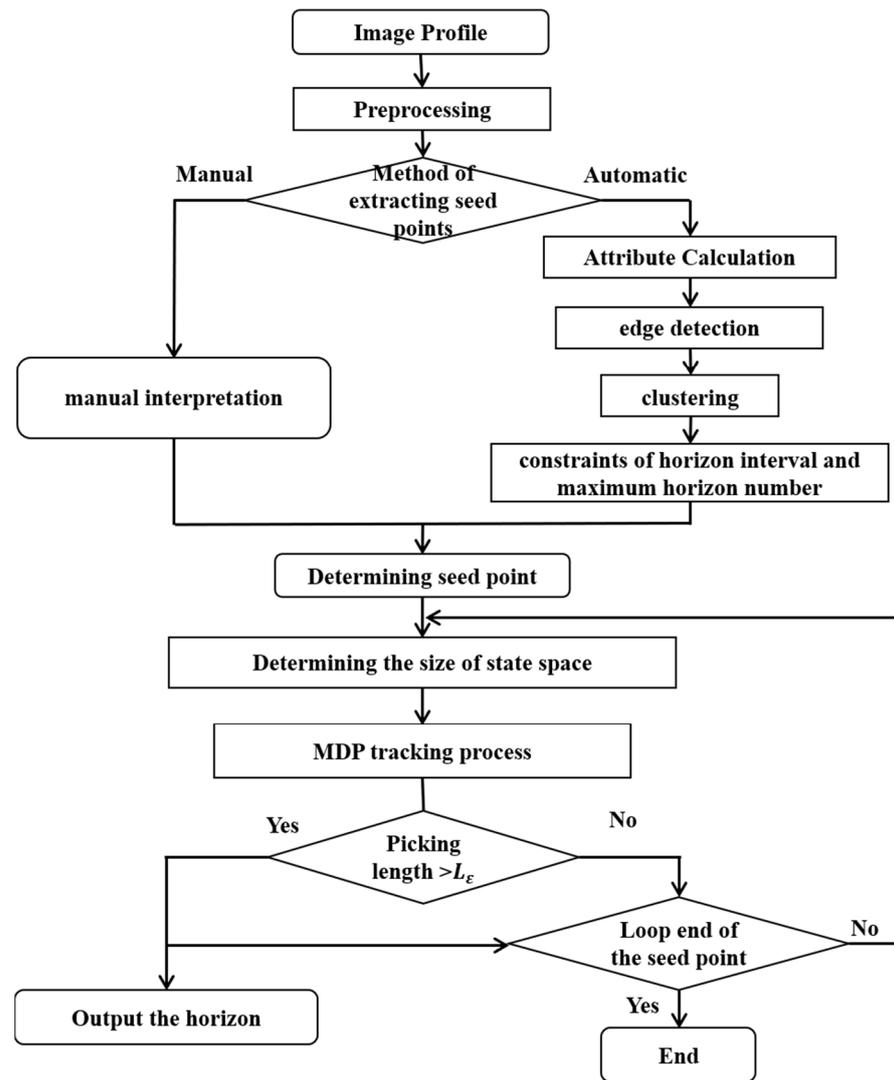
Figure 4. The flowcharts for the horizon picking of the proposed method.

(3) Determining the size of state space

According to the position of seed points z_0^0, we use the conventional method (waveform, phase, and extremum attributes) to pick up the horizon. Then, the z_0^t location is determined by superimposing these conventional results to form the initial state.

$$F_i^c = \begin{cases} F_i^c & if \ (F_i^c - z_0^0) < dz, \ \ i = \{w, p, e\} \\ z_0^0 & else \end{cases},$$
$$z_0^t = \left(F_w^c + F_p^c + F_e^c\right)/3 \tag{30}$$

where F_w^c, F_p^c, F_e^c are the results of the waveform, phase, and extremum attributes, respectively.

(4) MDP tracking process

After defining the MDP model with multiple seismic attributes, the optimal state (horizon position) can be obtained by the state value function. When the following conditions are met during MDP, the tracking will stop:

(a) When the decision time reaches the end point (the boundary of the profile), it automatically turns to the termination state and records the length of the horizon.

(b) When the long-term value function is less than the given threshold (it means that this may encounter a fault), the system will turn to the termination state and record the length of the horizon.

(5) Output the tracking result

If the length of the horizon is less than the given threshold length, the horizon will be discarded, and the MDP tracking process will be started with the new seed point. On the contrary, we will output the horizon as the final result.

5. Numerical Examples

In this section, we illustrate the effectiveness of the proposed automatic horizon-picking method in both synthetic and field data examples.

5.1. A Local Layered Model

In the first example, we use the migration result of a synthetic model to demonstrate the validity of the proposed method. Figure 5a shows the local seismic image under the condition of inaccurate migration velocity. A large number of migration artifacts appear in the profile, and the waveforms change rapidly along the horizons. The attributes of the envelope, instantaneous phase, and dip are shown in Figure 5b–d. As can be seen, the different attributes can highlight the partial characteristics of the seismic image. For example, when the envelope attribute occurs discontinuously (as shown in the black box in Figure 5b), the phase attribute has good continuity (as shown in the black box in Figure 5c). Multiple attributes can be used to supplement each other's defects. However, not all attributes may be good. At this time, if we take a conventional method such as single-trace-picking-based similarity, the reflection horizon cannot be traced accurately. Although all attributes are not good in local areas, they have certain similarities in the horizontal neighborhood of the area. This means that we can use the attributes of a neighborhood of the area to determine the horizon using a multi-step prediction method such as the MDP method. Figure 5e shows the results of the conventional single trace cross-correlation picking method and the proposed method using the multiple attributes with MDP. The sharp changes in the waveform in the horizontal direction and the attribute differences in the region lead to the inability of the conventional picking method to correctly pursue the horizon position. The Markov decision process method based on multi-attribute and multi-trace rewards can effectively avoid local noise interference and successfully track the correct reflection horizon.

5.2. The North Sea F3 Field Data

The second example illustrates the application of the proposed method to a real 3-D seismic survey F3 acquired over the North Sea. The extracted slice is shown in Figure 6a, and the instantaneous phase attribute is shown in Figure 6d. We observe that there is an obvious reflection horizon shown as the red arrow, and there is also a horizon with weak amplitude and less continuity, within the depth range of 0.8 km~0.9 km and the horizon position 0~1 km. Figure 6b,e show the troughs of the seismic data and the instantaneous phase attribute. Figure 6c,f show the peaks of the seismic data and the instantaneous phase attribute. Whatever a trough or a peak is shown the seismic data and the instantaneous phase attribute, the horizon shows poor continuity, which is marked by circles shown in Figure 6b,c,e,f. We cannot pick the correct reflection horizon according to any of these attributes. Figure 6g,h show the tracked horizon (red curve) using our proposed method on the image and waveform of seismic data, respectively. The blue cross symbols in Figure 6g,h are the seed points obtained via manual interpretation. Note that the waveform is similar and easy to track within the left portion and the right portion of the seismic slice. The waveform changes rapidly within the middle portion of the seismic slice, which makes it difficult to obtain the correct horizon position via the conventional method. However, the tracked horizon using the proposed method exactly follows the seismic reflection event.
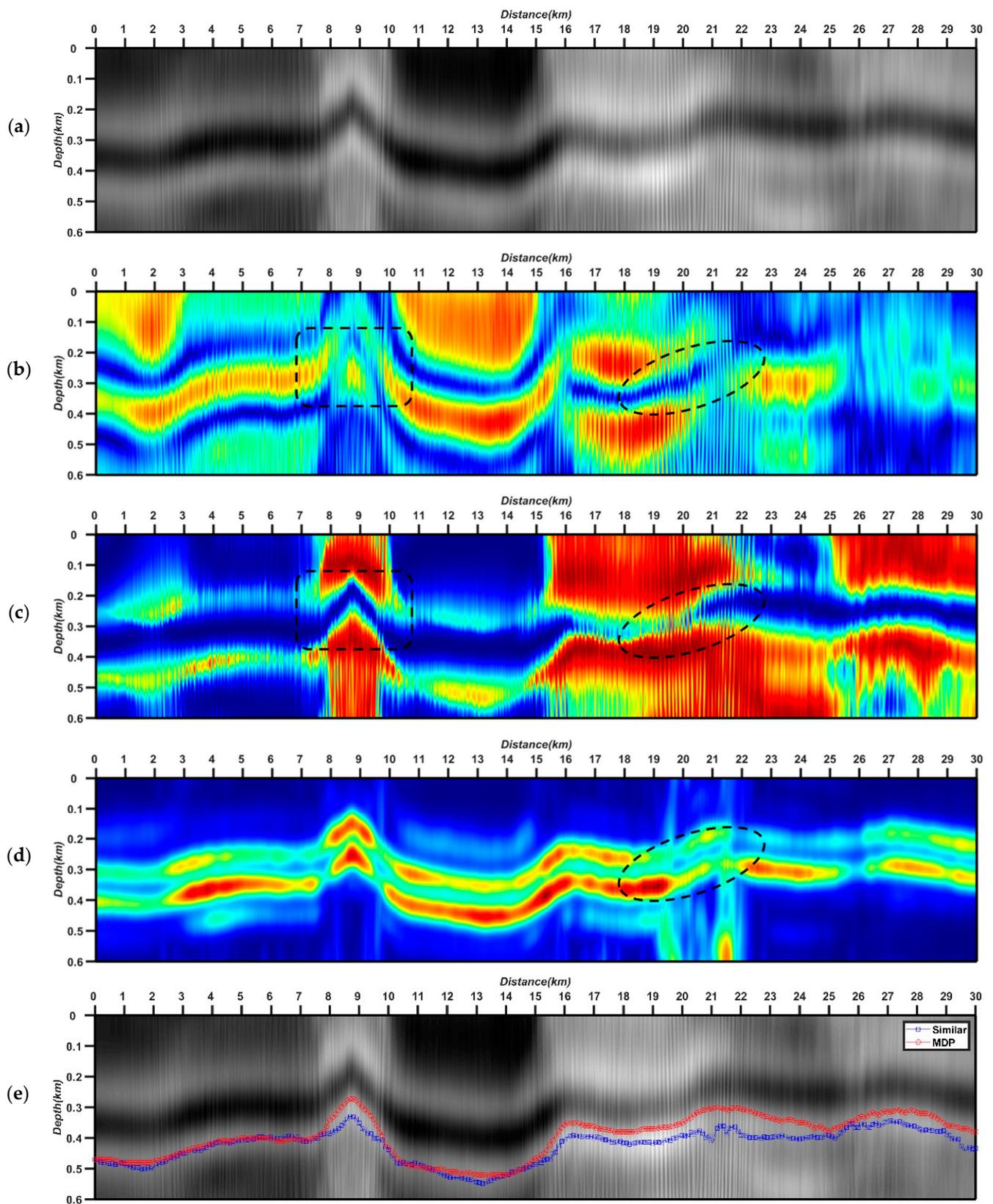
Figure 5. (**a**) Local migration imaging results, (**b**) envelope, (**c**) instantaneous phase, (**d**) dip, and (**e**) tracked horizon (the blue curve is the result via conventional waveform similar method, and the red curve is the result via the proposed method) overlaid with seismic image.
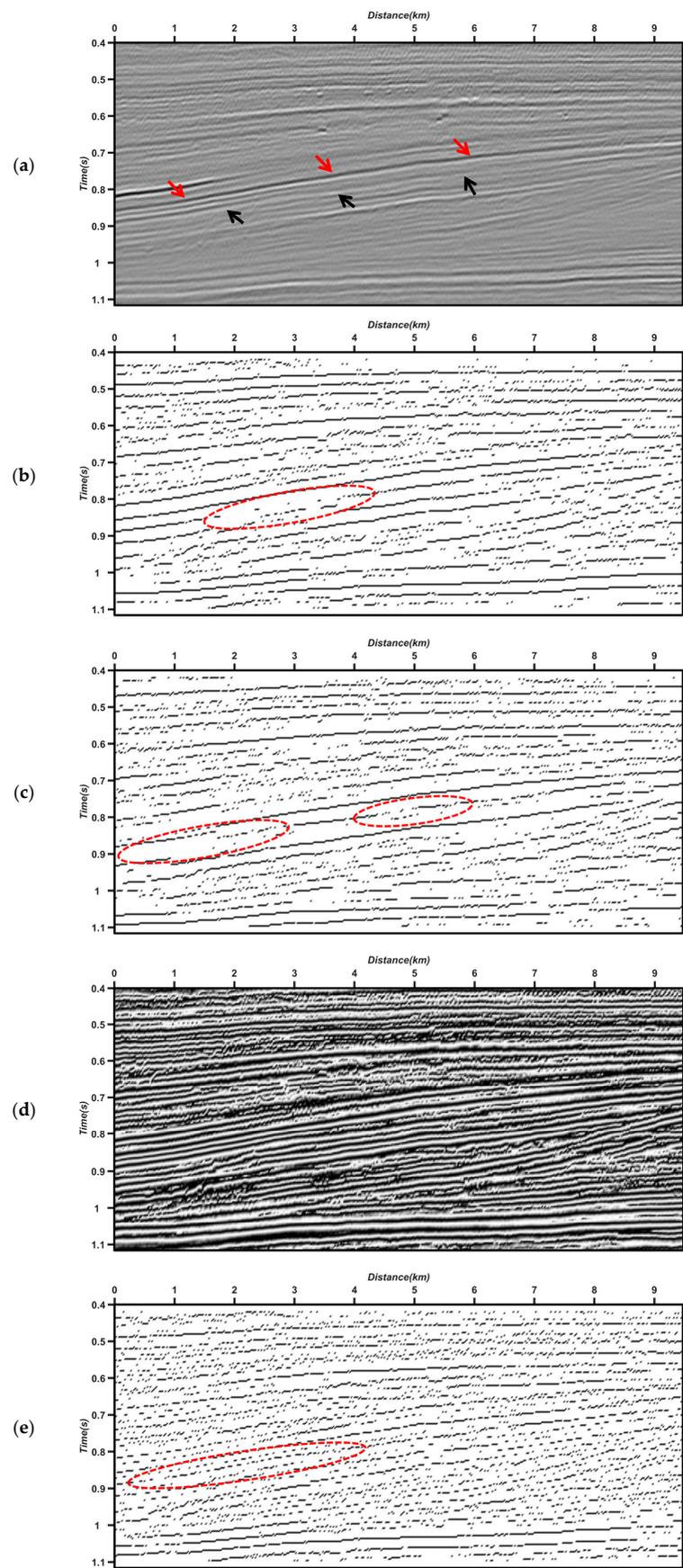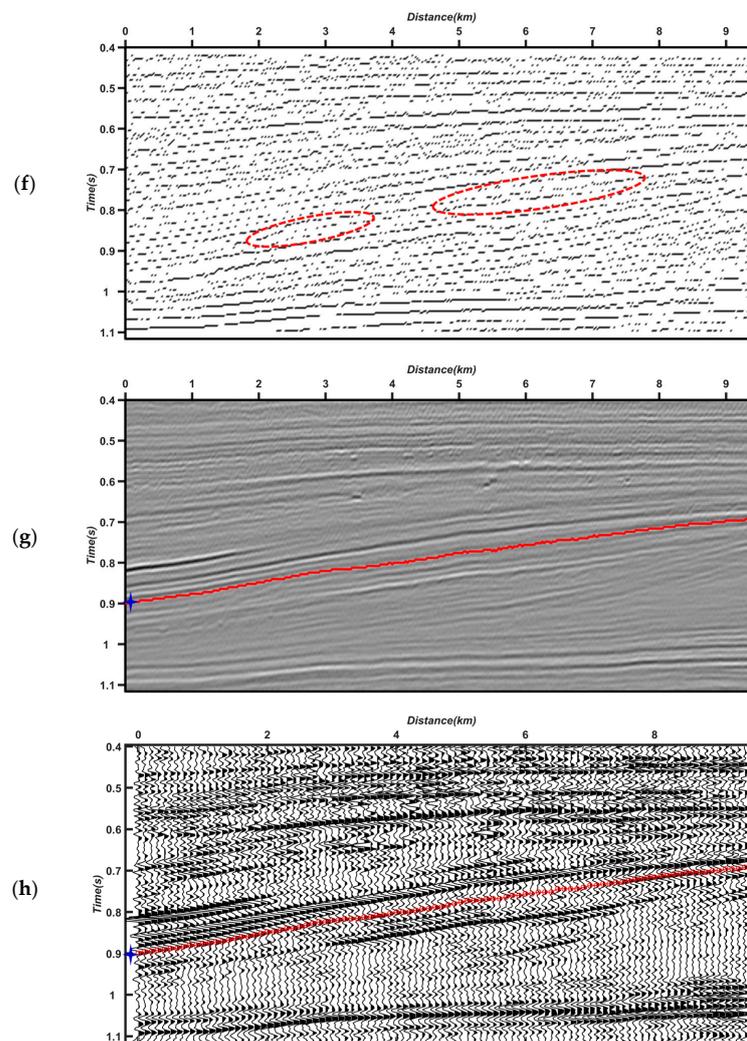
Figure 6. *Cont.*

Figure 6. The result for an extracted inline slice of F3 seismic data. (**a**) Seismic data, (**b**) troughs of seismic data, (**c**) peaks of seismic data, (**d**) instantaneous phase attribute, (**e**) troughs of instantaneous phase, (**f**) peaks of instantaneous phase, and tracked horizon (red curve) overlaid with (**g**) image and (**h**) waveform of seismic data.

5.3. The Overthrust Model for Picking Multiple Horizons

The third example illustrates the application to the overthrust model for picking multiple horizons. Figure 7a shows the migration imaging result with the background velocity model. In this example, we use the automated method to obtain seed points. The maximum number of seed points is set to eight, the minimum interval of the horizon is 0.1 km, and the minimum length is 1/10 of the number of lateral points. After the edge detection and clustering with the constraints of horizon interval and maximum horizon number, eight seed points are automatically obtained, which are marked by stars shown in Figure 7b. The tracked horizons with red curves using the proposed method overlaid on the seismic image are shown in Figure 7b. Because of the inaccurate migration velocity, the reflection horizon in the migration seismic image is not focused and discontinuous, shown as the black arrows, which makes it difficult to pick the correct reflection horizon information by using the conventional method. The reflection structure extracted using the proposed method is consistent with the real structure, which shows the effectiveness of this method. Therefore, the desired reflection horizons can be selected effectively and applied to the process of velocity modeling and the seismic interpretation process.
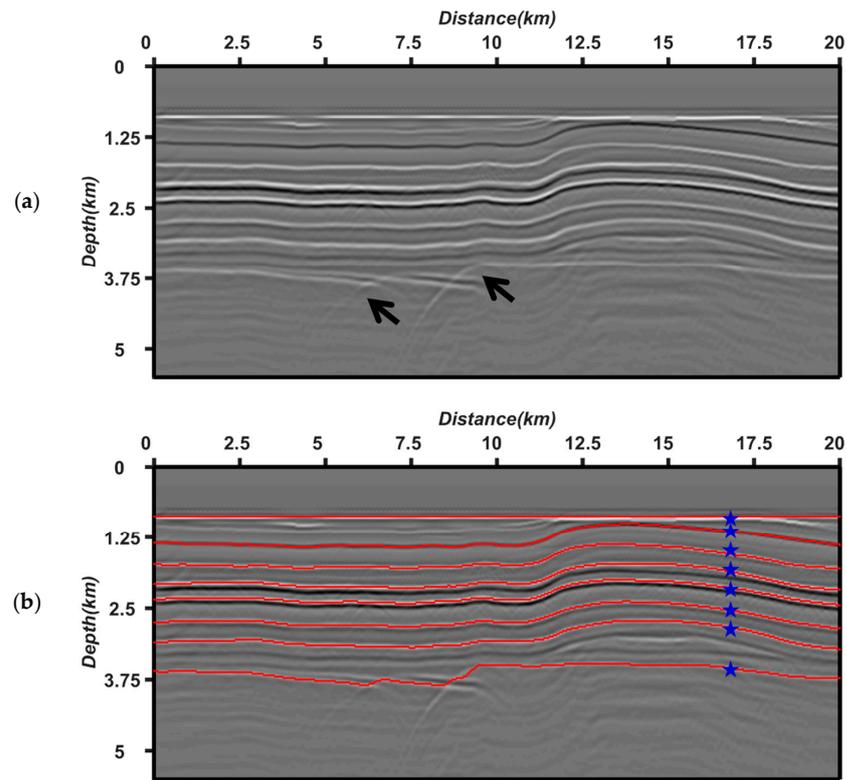
Figure 7. The overthrust model result. (**a**) Seismic image and (**b**) tracked horizons (red curve) overlaid with the seismic image.

5.4. The Field Data for Picking Multiple Horizons

We further test the effectiveness of the proposed method by using the migration image from a field dataset acquired on land. The tracked horizons with red circles using the proposed method overlaid on the local migration image are shown in Figure 8. The horizons have a small interval, and multiple waveforms are stacked together. Therefore, we cannot extract the complete wavelet waveform. This is very common in sandstone thin interbed reservoirs and is exacerbated when the result is included in internal multiples. Given the seed points, we can trace the right horizons from the superimposed waveform. The position of the extracted horizons is consistent with the initial state positions.
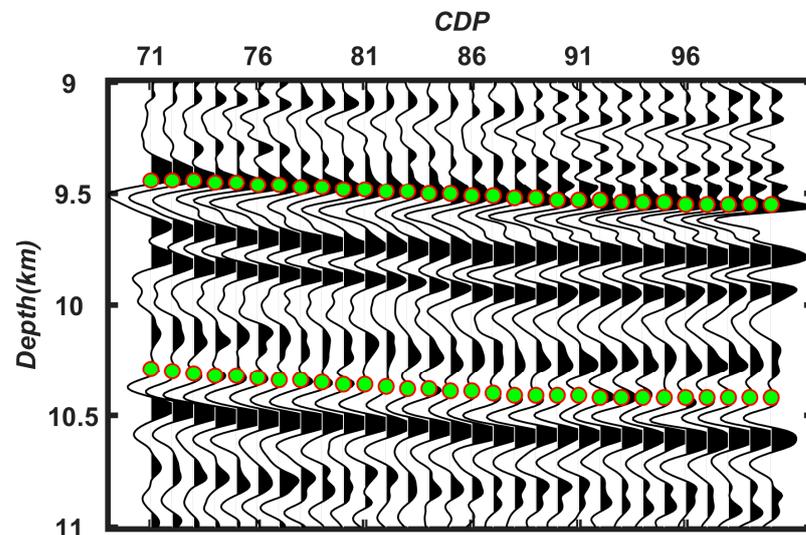

Figure 8. The extracted horizons for land data.

Finally, we apply the proposed method to field offshore data. Figure 9a shows the migration imaging result in the background velocity model from the original observed seismic data. Figure 9b shows the tracked horizons with a red curve overlaid on the migration image. The blue stars in Figure 9b are the positions of the seed points obtained via manual interpretation. As can be seen, the picking result of the proposed method is still very good, even when dealing with an actual complex seismic image.
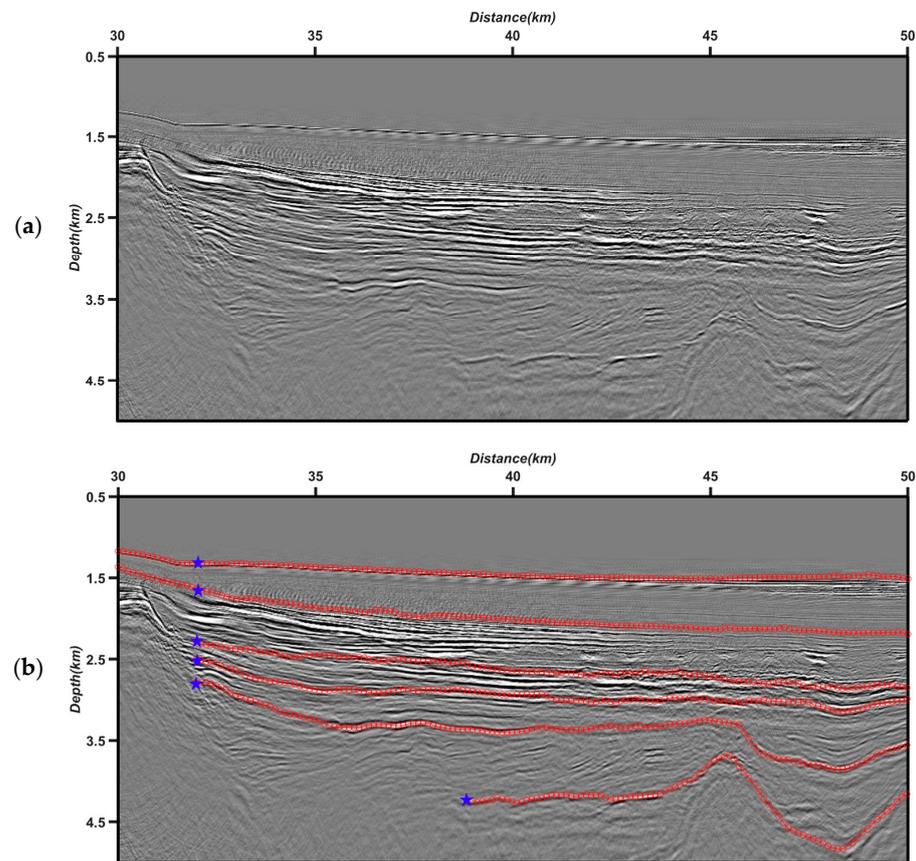

Figure 9. The extracted horizons for real offshore data. (**a**) Seismic image and (**b**) tracked horizons (red curve) overlaid with the seismic image.

6. Discussion

An appropriate data preprocessing method can effectively improve the accuracy of horizon picking. A seismic image contains not only multiple horizons but also various coherent and incoherent noises. For example, due to inaccurate migration velocity, serious migration artifacts will interfere with the continuity of the waveform. The migration operator is not amplitude-preserving, resulting in a large amplitude difference in the transverse direction of horizons. Therefore, various methods (such as a amplitude-preserving migration algorithm, edge-preserving smoothing, and a denoising method) can be applied to improve the imaging quality. For the problem that the imaging result is not focused due to an inaccurate migration velocity model, we can extract the common-image gathers (CIGs), then the flattening and denoising operators can be applied to the CIGs, and the focused imaging result can be obtained by optimal stacking [64].

The imaging results cannot be directly mapped to horizons. It is necessary to extract some attributes of a seismic image and highlight the difference between the horizon and noise, so as to detect and extract the horizon. The purpose of attribute extraction is to highlight the feature information and eliminate redundant and irrelevant information in a seismic image. However, the waveform in the horizontal direction changes dramatically, and one single attribute often cannot trace to the correct horizon. Multiple attributes can

highlight different aspects of seismic image properties. It is necessary to use multiple attributes to make up for the shortcomings of a single attribute. Therefore, the reward function is defined according to the similarity of multiple attributes for adjacent traces. The more attributes there are, the more information is available. However, it should be noted that the calculation of attributes must be accurate. Only correct attributes can contribute to picking. It is better to leave a deficiency uncovered than to have it covered without discretion. Attributes should be as independent as possible and highlight the structural features and similarities of the horizon.

Regarding horizon picking with the MDP, the key is to determine the state space, action space, and reward function. Considering the relationship between state and action at different times, and the similarity and continuity between the near horizons and inline or crossline slice, the association search constraint can be applied with the MDP. Through association search, we can transfer learning to the adjacent horizons or slices. Different MDP-solving methods can also be considered. Various kinds of reinforcement learning based on the MDP algorithm have been developed rapidly in recent years. For example, the Deep Q Network algorithm [65] can effectively search large-scale state space.

Moreover, when picking a horizon with faults, the MDP model needs to be optimized. The presence of a fault causes the low similarity of seismic attributes and low reward value. One solution is to place seed points on both sides of the fault. The position of the seed points will affect the picking result. The results are not the same for different positions of the seed points. The numbers and positions of seed points are also key factors. The method of obtaining seed points requires flexible selection through practical problems. The role of seed points is not only to provide the initial state but also to effectively select and extract the corresponding horizons.

In addition, for 2-D cases, we can regard the single horizon-picking problem as finding one optimal curve under the constraint of multiple seismic attributes. In the case of 3-D problems, the optimal curve will be replaced by the optimal surface. Tracing the optimal surface is a more complex problem. In this paper, we just explain how to track the horizon for a 2-D imaging section; extension to a 3-D problem with the MDP model is the next endeavor.

7. Conclusions

We propose a horizon-picking method based on multiple seismic attributes and the Markov decision process. We regard tracing reflection events on the seismic image as an intelligent decision problem. With the interaction process between the picking target and the environment, we can perceive the change of state and take some action to obtain reward value, where the reward function is defined by multiple seismic attributes. Multiple attributes can highlight different aspects of a seismic image and can make up for the shortcomings of a single attribute through the cross-constraint of multiple attributes. Then, the reward function will affect the state of the environment. Therefore, we can obtain the optimal picking horizon via the MDP model. Numerical examples on both synthetic and field data show that, by considering the optimal MDP with multiple attributes, we can avoid the influence caused by waveform changes in a single trace or a local area and pick up the desired reflection horizon robustly. The result of horizon picking can be used not only in velocity modeling but also in seismic interpretation.

Author Contributions: Conceptualization, C.W. and H.W.; methodology, C.W., B.F. and H.W.; writing—original draft and formal analysis, C.W. and X.S.; writing—review and editing, B.F., X.S., R.X. and S.S. All authors have read and agreed to the published version of the manuscript.

References

1. Howard, R.E. Method for Attribute Tracking in Seismic Data. U.S. Patent No. 5,1991,056-066, 8 October 1991.
2. Zeng, H.; Backus, M.M.; Barrow, K.T.; Tyler, N. Stratal slicing—Part 1: Realistic 3D seismic model. *Geophysics* **1998**, *63*, 502–513. [CrossRef]
3. Stark, T.J. Unwrapping instantaneous phase to generate a relative geologic time volume. In *73rd Annual International Meeting, SEG, Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2003; pp. 1707–1710. [CrossRef]
4. Wu, X.; Zhong, G. Generating a relative geologic time volume by 3D graph-cut phase unwrapping method with horizon and unconformity constraints. *Geophysics* **2012**, *77*, O21–O34. [CrossRef]
5. Dossi, M.; Forte, E.; Pipan, M. Automated reflection picking and polarity assessment through attribute analysis: Theory and application to synthetic and real ground-penetrating radar data. *Geophysics* **2015**, *80*, H23–H35. [CrossRef]
6. Forte, E.; Dossi, M.; Pipan, M.; Ben, A.D. Automated phase attribute-based picking applied to reflection seismic. *Geophysics* **2016**, *81*, V141–V150. [CrossRef]
7. Marfurt, K.; Kirlin, R.; Farmer, S.; Bahorich, M. 3-D seismic attributes using a semblance-based coherency algorithm. *Geophysics* **1998**, *63*, 1150–1165. [CrossRef]
8. Lacaze, F.; Valding, T. Autopicking Velocity by Path-Integral Optimization and Surface Fairing. In *79th Annual International Meeting, SEG Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2009; pp. 2592–2596. [CrossRef]
9. Zou, G.; Ji, R.; Liu, W.; Zhang, J. Horizon auto-tracking technology based on the waveform similarity for the coalmining area seismic data. In Proceedings of the 2013 2nd International Symposium on Instrumentation and Measurement, Sensor Network and Automation (IMSNA), Toronto, ON, Canada, 23–24 December 2013; pp. 236–242.
10. Gersztenkorn, A.; Marfurt, K.J. Eigenstructure-based coherence computations as an aid to 3-D structural and stratigraphic mapping. *Geophysics* **1999**, *64*, 1468–1479. [CrossRef]
11. Bakker, P. Image Structure Analysis for Seismic Interpretation. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2002.
12. Lomask, J.; Guitton, A.; Fomel, S.; Claerbout, J.; Valenciano, A.A. Flattening without picking. *Geophysics* **2006**, *71*, P13–P20. [CrossRef]
13. Lomask, J.; Guitton, A. Volumetric flattening: An interpretation tool. *Lead. Edge* **2007**, *26*, 888–897. [CrossRef]
14. Parks, D. Seismic Image Flattening as a Linear Inverse Problem. Master's Thesis, Colorado School of Mines, Golden, CO, USA, 2010.
15. Wu, X. Directional structure-tensor-based coherence to detect seismic faults and channels. *Geophysics* **2017**, *82*, A13–A17. [CrossRef]
16. Wu, X.; Fomel, S. Least-squares horizons with local slopes and multigrid correlations. *Geophysics* **2018**, *83*, IM29–IM40. [CrossRef]
17. Di, H.; Gao, D.; AlRegib, G. 3D dip vector-guided auto-tracking for weak seismic reflections: A new tool for shale reservoir visualization and interpretation. *Interpretation* **2018**, *6*, SN47–SN56. [CrossRef]
18. Lou, Y.; Zhang, B. Automatic horizon picking using multiple seismic attributes. In *88th Annual International Meeting, SEG Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2018; pp. 1683–1688.
19. Gogia, R.; Singh, R.; de Groot, P.; Gupta, H.; Srirangarajan, S.; Phirani, J.; Ranu, S. Tracking 3D seismic horizons with a new hybrid tracking algorithm. *Interpretation* **2020**, *8*, SQ39–SQ45. [CrossRef]
20. Li, S.; Zhao, J.; Zhang, H.; Bi, Z.; Qu, S. A Novel Horizon Picking Method on Sub-Bottom Profiler Sonar Images. *Remote Sens.* **2020**, *12*, 3322. [CrossRef]
21. Lou, Y.; Zhang, B.; Lin, T.; Cao, D. Seismic horizon picking by integrating reflector dip and instantaneous phase attributes. *Geophysics* **2020**, *85*, O37–O45. [CrossRef]
22. Zhang, B.; Qi, J.; Lou, Y.; Fang, H.; Cao, D. Generating seismic horizon using multiple seismic attributes. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 979–983. [CrossRef]
23. Zhao, J.; Li, S.; Zhao, X.; Feng, J. A Comprehensive Horizon-Picking Method on Subbottom Profiles by Combining Envelope, Phase Attributes, and Texture Analysis. *Earth Space Sci.* **2020**, *7*, e2019EA000680. [CrossRef]
24. Admasu, F.; Tönnies, K.D. Model-based Approach to Automatic 3D Seismic Horizon Correlation across Faults. In Proceedings of the Simulation und Visualisierung 2004 (SimVis), Magdeburg, Germany, 4–5 March 2004; pp. 239–250.
25. Bondár, I. Seismic horizon detection using image processing algorithms. *Geophys. Prospect.* **1992**, *40*, 785–800. [CrossRef]
26. Aurnhammer, M.; Tonnies, K.D. A genetic algorithm for automated horizon correlation across faults in seismic images. *IEEE Trans. Evol. Comput.* **2005**, *9*, 201–210. [CrossRef]

27. Pauget, F.; Lacaze, S.; Valding, T. A global approach in seismic interpretation based on cost function minimization. In *79th Annual International Meeting, SEG Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2009; pp. 2592–2596.

28. Hoyes, J.; Cheret, T. A review of global interpretation methods for automated 3D horizon picking. *Lead. Edge* **2011**, *30*, 38–47. [CrossRef]

29. Hong, Z.; Su, M.; Qian, F.; Han, Q. Global seismic horizon interpretation based on data mining—A new tool for seismic geomorphologic study. *Interpretation* **2020**, *8*, T131–T140. [CrossRef]

30. Wu, X.; Hale, D. Automatically interpreting all faults, unconformities, and horizons from 3D seismic images. *Interpretation* **2016**, *4*, T227–T237. [CrossRef]

31. Jin, S.; Chen, S.; Wei, J.; Li, X. Automatic seismic event tracking using a dynamic time warping algorithm. *J. Geophys. Eng.* **2017**, *14*, 1138–1149. [CrossRef]

32. Bugge, A.J.; Erik, L.J.; Evensen, A.K.; Faleide, J.I.; Clark, S. Automatic extraction of dislocated horizons from 3D seismic data using nonlocal trace matching. *Geophysics* **2019**, *84*, IM77–IM86. [CrossRef]

33. Lou, Y.; Zhang, B.; Fang, H.; Cao, D.; Wang, K.; Huo, Z. Simulating the procedure of manual seismic horizon picking. *Geophysics* **2021**, *86*, O1–O12. [CrossRef]

34. Yan, S.; Wu, X. Seismic horizon extraction with dynamic programming. *Geophysics* **2021**, *86*, 51–62. [CrossRef]

35. Harrigan, E.; Kroh, J.R.; Sandham, W.A.; Durrani, T.S. Seismic horizon picking using an artificial neural network. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, CA, USA, 23–26 March 1992; pp. 105–108.

36. Huang, K. Hopfield neural network for seismic horizon picking. In *75th Annual International Meeting, SEG, Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 1997; pp. 562–565. [CrossRef]

37. Huang, K. Neural network for seismic horizon picking. In *IEEE International Joint Conference on Neural Networks Proceedings*; IEEE World Congress on Computational Intelligence: Anchorage, AK, USA, 1998; Volume 3, pp. 1840–1844.

38. Leggett, M.; Sandham, W.A.; Durrani, T.S. *Automated 3-D Horizon Tracking and Seismic Classification Using Artificial Neural Networks*; Springer: Cham, The Netherlands, 2003; pp. 31–44.

39. Peters, B.; Granek, J.; Haber, E. Multi-resolution neural networks for tracking seismic horizons from few training images. *Interpretation* **2019**, *7*, SE201–SE213. [CrossRef]

40. Waldeland, A.U.; Jensen, A.C.; Gelius, L.-J.; Solberg, A.H.S. Convolutional neural networks for automated seismic interpretation. *Lead. Edge* **2018**, *37*, 529–537. [CrossRef]

41. Shi, Y.; Wu, X.; Fomel, S. SaltSeg: Automatic 3D salt segmentation using a deep convolutional neural network. *Interpretation* **2019**, *7*, SE113–SE122. [CrossRef]

42. Wu, H.; Zhang, B.; Lin, T.; Cao, D.; Lou, Y. Semiautomated seismic horizon interpretation using the encoder-decoder convolutional neural network. *Geophysics* **2019**, *84*, B403–B417. [CrossRef]

43. Yang, L.; Sun, Z. Seismic horizon tracking using a deep convolutional neural network. *J. Pet. Sci. Eng.* **2020**, *187*, 106709. [CrossRef]

44. Tschannen, V.; Delescluse, M.; Ettrich, N.; Keuper, J. Extracting horizon surfaces from 3D seismic data using deep learning. *Geophysics* **2020**, *85*, N17–N26. [CrossRef]

45. Zhang, K.; Lin, N.; Zhang, D.; Zhang, J.; Yang, J.; Tian, G. Automatic tracking for seismic horizons using convolution feature analysis and optimization algorithm. *J. Pet. Sci. Eng.* **2021**, *208*, 109441. [CrossRef]

46. Huang, K.; Chang, W.; Lin, C. Self-organizing neural network for picking seismic horizons. In *60th Annual International Meeting, SEG Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 1990; pp. 313–316.

47. Shi, Y.; Wu, X.; Fomel, S. Waveform embedding: Automatic horizon picking with unsupervised deep learning. *Geophysics* **2020**, *85*, WA67–WA76. [CrossRef]

48. Wald, A. Foundations of a general theory of sequential decision functions. Econometrica. *J. Econom. Soc.* **1947**, *15*, 279–313. [CrossRef]

49. Shapley, L.S. Stochastic games. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 1095–1100. [CrossRef]

50. Bellman, R.A. Markovian decision process. *J. Math. Mech.* **1957**, *6*, 679–684. [CrossRef]

51. Howard, R.A. *Dynamic Programming and Markov Processes*; John Wiley: Hoboken, NJ, USA, 1960.

52. Blackwell, D. Discrete dynamic programming. *Ann. Math. Stat.* **1962**, *33*, 719–726. [CrossRef]

53. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2014.

54. Barto, A.G.; Sutton, R.S.; Watkins, C. *Learning and Sequential Decision Making*; University of Massachusetts: Amherst, MA, USA, 1989.

55. Littman, M.L. *Algorithms for Sequential Decision-Making*; Brown University: Rhode Island, RI, USA, 1996.

56. Read, P.; Lermit, J. Bio-energy with carbon storage (BECS): A sequential decision approach to the threat of abrupt climate change. *Energy* **2005**, *30*, 2654–2671. [CrossRef]

57. Bertsekas, D. *Dynamic Programming and Optimal Control: Volume I*; Athena Scientific: Belmont, MA, USA, 2012.

58. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]

59. Sutton, R.S.; Barto, A.G. *Reinforcement Learning*; MIT Press: Cambridge, MA, USA, 1998.

60. Ma, Y.; Fei, T.; Luo, Y. A new insight into automatic first-arrival picking based on reinforcement learning. In *81st EAGE Conference and Exhibition, Expanded Abstracts*; European Association of Geoscientists and Engineers: Houten, The Netherlands, 2019; pp. 1–5.

61. Luo, F.; Feng, B.; Wang, H. Automatic first-arrival picking method via intelligent Markov optimal decision processes. *J. Geophys. Eng.* **2021**, *18*, 406–417. [CrossRef]

62. Sutton, R.S. Generalization in reinforcement learning: Successful examples using sparse coarse coding. In *Advances in Neural Information Processing Systems*; MIT Press: Denver, CO, USA, 1996; pp. 1038–1044.

63. Wu, C.; Wang, H.; Feng, B.; Sheng, S. RTM angle gathers based on the Combining Local and Global (CLG) optical flow method and wavefield decomposition method. *Chin. J. Geophys.* **2021**, *64*, 1375–1388. (In Chinese) [CrossRef]

64. Wu, C.; Wang, T.; Wang, H.; Luo, F.; Xu, P. Improving the quality of Common-Image Gathers using DTW and Local Similarity. In *89th SEG Annual International Meeting, Expanded Abstracts*; Society of Exploration Geophysicists: Tulsa, OK, USA, 2019; pp. 4336–4340.

65. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]