MDPI

Article

Bayesian Learning-Based Clustered-Sparse Channel Estimation for Time-Varying Underwater Acoustic OFDM Communication [†]

Shuaijun Wang [1] **, Mingliu Liu** [1] **and Deshi Li** [1,2,*]

1 Electronic Information School, Wuhan University, Wuhan 430072, China; wangsj@whu.edu.cn (S.W.); liumingliu@whu.edu.cn (M.L.)
2 Collaborative Innovation Center of Geospatial Technology, 129 Luoyu Road, Wuhan 430072, China
* Correspondence: dsli@whu.edu.cn
† This paper is an extended version of our paper published in Wang, S.; Li, D.; Liu, M.; Huang, W.; Chen, H.; Cen, Y. Clustered-Sparse Bayesian Learning for Channel Estimation in Underwater Acoustic OFDM Systems. In Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 21–23 October 2020; pp. 546–551.

Abstract: Orthogonal frequency division multiplexing (OFDM) has been widely adopted in underwater acoustic (UWA) communication due to its good anti-multipath performance and high spectral efficiency. For UWA-OFDM systems, channel state information (CSI) is essential for channel equalization and adaptive transmission, which can significantly affect the reliability and throughput. However, the time-varying UWA channel is difficult to estimate because of excessive delay spread and complex noise distribution. To this end, a novel Bayesian learning-based channel estimation architecture is proposed for UWA-OFDM systems. A clustered-sparse channel distribution model and a noise-resistant channel measurement model are constructed, and the model hyperparameters are iteratively optimized to obtain accurate Bayesian channel estimation. Accordingly, to obtain the clustered-sparse distribution, a partition-based clustered-sparse Bayesian learning (PB-CSBL) algorithm was designed. In order to lessen the effect of strong colored noise, a noise-corrected clustered-sparse channel estimation (NC-CSCE) algorithm was proposed to improve the estimation accuracy. Numerical simulations and lake trials are conducted to verify the effectiveness of the algorithms. Results show that the proposed algorithms achieve higher channel estimation accuracy and lower bit error rate (BER).

Keywords: clustered-sparse channel estimation; Bayesian learning; underwater acoustic communication; orthogonal frequency division multiplexing

1. Introduction

Underwater acoustic (UWA) sensor networks can be applied to a variety of underwater activities, including environment monitoring, resource exploration and equipment navigation. Reliable high-rate communication is significant for massive data transmission between sensor nodes [1]. However, the UWA channel is widely known as one of the most difficult communication media, and high-rate communication with high reliability is quite challenging. Since the sound attenuation rapidly increases with frequency, the bandwidth available for communication is extremely limited. Due to the reflection and refraction of sound in the ocean, the multipath spread in a medium-range shallow water channel can extend over tens or even hundreds of milliseconds [2], resulting in severe frequency-selective signal distortion. Due to the dynamic environment and moving transceivers in ocean, the UWA channel is time-varying. In particular, the normalized frequency offset factor induced by Doppler can be on the order of 10^{-3} compared with 10^{-7} for high-speed mobile radio channels [2].

Due to high spectral efficiency and good anti-multipath performance, orthogonal frequency division multiplexing (OFDM) and some OFDM-based modulation techniques are

widely adopted for UWA communications [3–6]. In UWA-OFDM systems, accurate channel state information (CSI) is vital for channel equalization and adaptive transmission [7,8], which can significantly affect the reliability and throughput. However, it is difficult to estimate the time-varying channel impulse responses (CIRs) with excessive delay spread. To overcome this problem, some structural information such as sparsity can be exploited to improve channel estimation performance. The paths of the UWA channel tend to be clustered-sparse, which carry comprehensive structural information [9–13]. As shown in Figure 1, there are many weak paths centering around the eigen-paths associated with the medium refractions and surface/bottom bounces, which form the clustered-sparse structure. As different clusters are usually derived from different transmission routes, the channel coherence time usually varies across clusters. Thus, the path variation for different clusters can reflect the channel variation characteristics [14], and can be used for predicting the time-varying UWA channel [15]. However, the fading process of the UWA channel may randomly vary [2], which makes it difficult to accurately obtain the path variation for different clusters between adjacent OFDM blocks. To obtain the clustered-sparse structure of the UWA channel, we proposed a clustered-sparse Bayesian channel estimation method in [16], which is the preliminary version of this work (This work extends the conference paper [16]) by (i) adding a Bayesian learning-based channel estimation architecture; (ii) adding a noise-corrected clustered-sparse channel estimation algorithm to lessen the effect of strong colored noise; (iii) giving the detailed derivation and performance analysis of the proposed algorithms; and (iv) presenting more complete performance results related to modulation order, model hyperparameters and noise distribution).
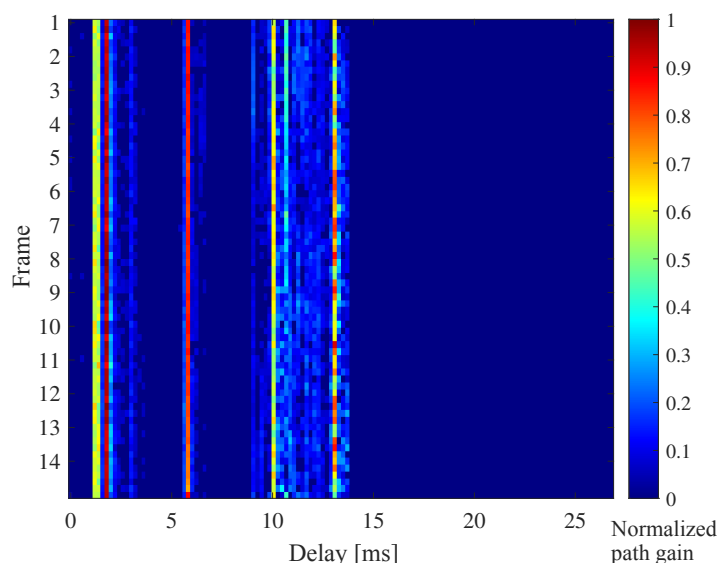

Figure 1. Time-varying multipath channel with clustered-sparse structure estimated in a lake communication trial.

In addition to the time-varying CIRs with excessive delay spread, the complex distribution of noise in the UWA channel also makes the channel estimation difficult. Generally, there are two kinds of channel noise, including the ambient noise and site-specific noise. The former type is presented in the background of the quiet deep sea, and exhibits a colored characteristic when approximated by Gaussian distribution, while for the site-specific noise, significant non-Gaussian components should be carefully considered [2]. Some experiments have confirmed that the noise is non-uniform across different locations, times and frequencies [17]. As a popular noise model in most UWA-OFDM systems, additive white Gaussian noise (AWGN) is straightforward for modeling and solving the noise distribution of the UWA channel, but it deviates from the actual underwater noise. Obviously, the deviation of the noise model means the estimation error of the noise distribution, which will lead to the degradation of channel estimation and signal detection performance.

To tackle the challenge of estimating the time-varying UWA channel with excessive delay spread and complex noise distribution, we adopted Bayesian learning to obtain the multipath and noise distribution characteristics of the UWA channel, based on which Bayesian channel estimation is derived to improve the performance of pilot-based channel estimation. The main contributions of this work are summarized as follows:

- To estimate the time-varying multipath channel with colored noise, we propose a novel Bayesian learning-based channel estimation architecture for UWA-OFDM systems. Specifically, a clustered-sparse channel distribution model is constructed to characterize the delay power spectrum and temporal correlation of each cluster in the multipath channel, and a noise-resistant channel measurement model is constructed to reduce the noise disturbance. By learning the model hyperparameters, the Bayesian channel estimation based on the two models can be iteratively optimized.
- To obtain the clustered-sparse distribution, we propose a partition-based clustered-sparse Bayesian learning (PB-CSBL) algorithm. Through the cluster partition, different clusters can learn different channel correlation coefficients, and thus the inter-cluster interference of the multipath channel can be suppressed.
- To lessen the effect of strong colored noise, we propose a noise-corrected clustered-sparse channel estimation (NC-CSCE) algorithm. Based on the iterative symbol decision and noise correction, the more accurate hyperparameters of the models can be obtained, which can improve the accuracy of the Bayesian channel estimation.

The rest of this paper is organized as follows. Section 2 discusses the related works. In Section 3, the system architecture and channel models are constructed. Then, the Bayesian learning-based clustered-sparse channel estimation method is designed in Section 4. In Section 5, the noise-corrected clustered-sparse channel estimation method is presented. The evaluation and result analysis are shown in Section 6. In Section 7, conclusions are given.

2. Related Works

The UWA channel is severely band-limited, and exhibits time-varying excessive delay spread. To this end, the sparse structure of the UWA channel can be exploited to improve the estimation performance, as most channel energy of the UWA channel is concentrated on a few paths [18,19]. The paths of the UWA channel also tend to be clustered-sparse [20], which has been frequently observed based on field experiments at different locations [9–13]. A detailed investigation on the variation of the significant paths in each cluster is presented in [21]. Based on the observation of channel sparsity, compressed sensing (CS)-based sparse channel estimation methods [22–24] have been studied, such as the greedy matching pursuit (MP) and orthogonal matching pursuit (OMP), which were adopted in [18,25], aiming to estimate the UWA channel impulse response (CIR) and the channel delay-Doppler-spread function, respectively. In order to reduce the convergence error of the OMP algorithm, the compressed sampling matching pursuit (CoSaMP) algorithm is proposed in [26]. However, CoSaMP-based channel estimators require knowledge of the channel sparsity level, which is difficult to obtain accurately in advance in practical applications. The joint sparsity among adjacent OFDM blocks was exploited by the simultaneous OMP (SOMP) [27] for higher estimation accuracy, but only stable path delays were considered and the temporal correlation of path gains was ignored. Typical convex optimization algorithms including basis pursuit (BP) and basis pursuit denoising (BPDN) tackle the NP-hard l_0 norm problem by transforming the l_0 norm to a convex l_1 norm. The BP algorithm was applied to estimate the path delays and Doppler scale factors of the UWA channel in [25], and the BPDN algorithm was proposed to estimate the slowly time-varying UWA channels in [28]. However, the global minimum of the cost function in BP or BPDN is not necessarily consistent with the sparsest solution. An empirical channel variation model, in which different clusters vary independently, was proposed in [14], and demonstrated that adaptation with multiple clusters outperforms that with one cluster for UWA channel estimation.

In recent years, Bayesian learning (BL) has been used to improve the performance of channel estimation. Compared with the CS-based methods that provide point estimation, Bayesian techniques can give the interval estimation of channel response and have a desirable property of preventing structural errors [29]. By using BL-based methods, the high accuracy of channel estimation was achieved in OFDM systems [30–33]. The temporal multiple sparse Bayesian learning (TMSBL)-based estimation method was further proposed in [32] to estimate the sparse channels by taking advantage of the channel coherence between consecutive OFDM blocks, and it can achieve better performance in strong correlated channels and maintains robustness in weak temporal correlated channels. Due to the high computational complexity of BL, some accelerated algorithms were proposed, such as the approximate message passing (AMP)-based method in [34]. More recently, some researchers utilized deep learning (DL)-based channel estimators for OFDM systems [35–37], which showed better channel estimation or equalization performance for nonlinear channel distortion. However, how to go about improving the generalization performance of pre-training DL models and reduce the complexity of training and inferring remain challenging.

3. System Architecture and Channel Models

In this section, the system architecture of Bayesian learning-based channel estimation is presented first. Then, the noise-resistant channel measurement model and the clustered-sparse channel distribution model are constructed.

3.1. Bayesian Learning-Based Channel Estimation Architecture

A typical UWA-OFDM communication scenario was shown at the top of Figure 2. The receiver can obtain the transmitted OFDM signal through dynamic multipaths, and the received signal may be affected by random colored noise. At the bottom of Figure 2, the system architecture of Bayesian learning-based channel estimation is given. To characterize the delay power spectrum and temporal correlation of each cluster in the UWA channel, the clustered-sparse channel distribution model is constructed and the model hyperparameters can be optimized by the iterative partition-based cluster evolution. Demodulated signal and pilot symbols were utilized to measure the instantaneous channel response, and the iterative noise measurement and data detection can improve the anti-noise performance of the channel measurement model. Based on the channel distribution model and the channel measurement model, the Bayesian channel estimation can be obtained.

Important notations used in following sections are given by:

- Uppercase and lowercase bold symbols are reserved for matrices and vectors, respectively. Particularly, $\mathbf{I_M}$ denotes the identity matrix with size $M \times M$. When the dimension is evident from the context, for simplicity, we only used $\mathbf{I}$;
- $\mathbf{A}^T$ and $\mathbf{A}^H$ represent the transpose and conjugate transpose of $\mathbf{A}$, respectively;
- $\mathbf{A} \otimes \mathbf{B}$ denotes the Kronecker product of the two matrices $\mathbf{A}$ and $\mathbf{B}$;
- $vec(\mathbf{A})$ denotes the vectorization of $\mathbf{A}$ formed by stacking its columns into a single column vector;
- $Tr(\mathbf{A})$ denotes the trace of $\mathbf{A}$;
- $Toeplitz(a_1, a_2, \ldots, a_N)$ denotes a Toeplitz matrix taking $a_1, \ldots, a_N$ as first row;
- $diag(\mathbf{a})$ or $diag(a_1, a_2, \ldots, a_N)$ denotes a diagonal matrix with principal diagonal elements being $a_1, a_2, \ldots, a_N$ in turn; $diag(\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_N)$ denotes a block diagonal matrix with principal diagonal blocks being the square matrices $\mathbf{A}_1, \ldots, \mathbf{A}_N$ in turn;
- If $\mathbf{A}$ is a square matrix, $Prdiag(\mathbf{A})$ denotes a diagonal matrix with principal diagonal elements being the principal diagonal elements of $\mathbf{A}$ in turn.
- If some terms in a cost function do not contribute to the subsequent optimization of the parameters, $\propto$ is used to indicate that these terms have been dropped.
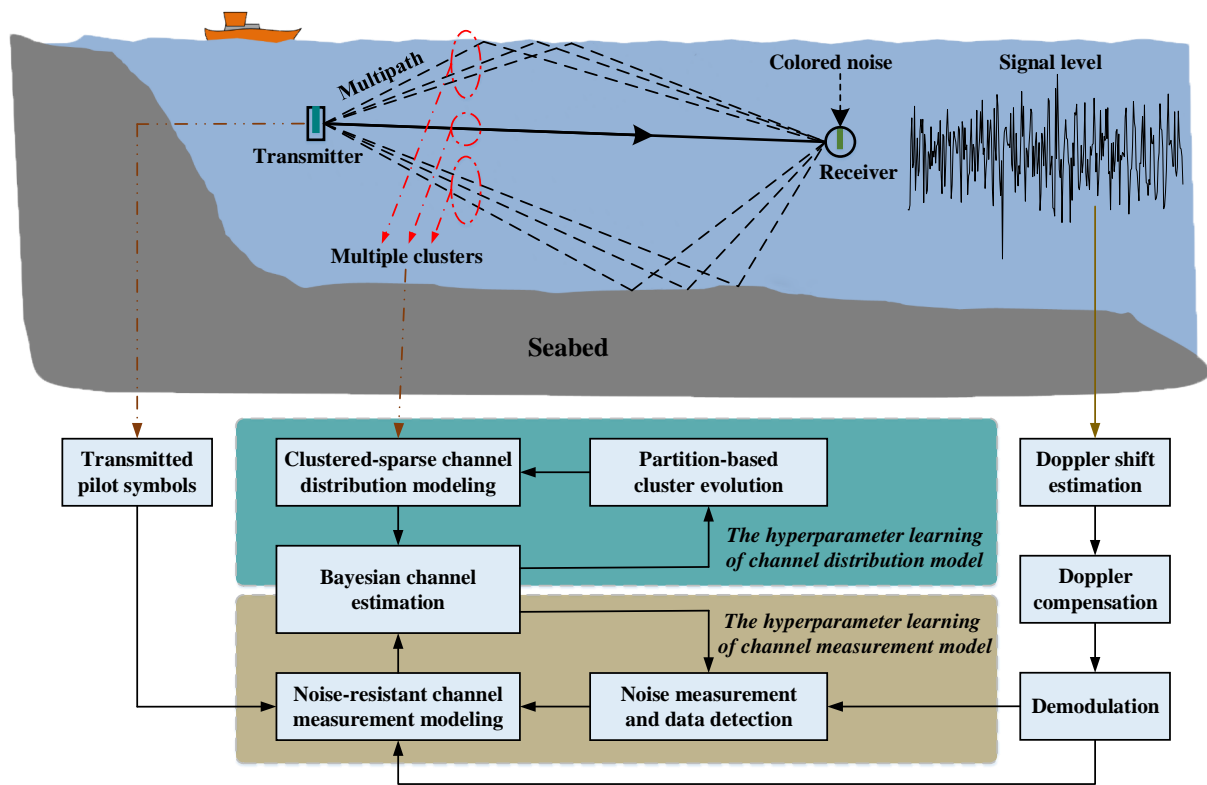
Figure 2. Architecture of Bayesian learning-based channel estimation in UWA-OFDM communication.

3.2. Noise-Resistant Channel Measurement Model

Considering the OFDM system with cyclic prefix (CP) and K subcarriers, the non-overlapping sets of K_d data subcarriers $\mathbf{x_d}$, K_p pilot subcarriers $\mathbf{x_p}$ and K_n null subcarriers $\mathbf{x_n}$ satisfy $K = K_d + K_p + K_n$. Let T denotes the OFDM symbol period without CP, T_{cp} denotes the length of CP and $T_s = T + T_{cp}$ is the whole OFDM symbol period. The kth subcarrier is at the frequency:

$$f_k = f_c + k/T, \qquad k = -K/2, \ldots, K/2 - 1, \tag{1}$$

where f_c is the carrier frequency. Assuming that there are M OFDM blocks in one frame, a passband waveform of the mth OFDM block at time $t' \in [-T_{cp} + mT_s, T + mT_s]$ is then given by

$$\widetilde{x}^{(m)}(t) = 2Re\{\frac{1}{\sqrt{K}} \sum_{k=-K/2}^{K/2-1} x_k^{(m)} e^{j2\pi f_k t} q(t)\}, \tag{2}$$

where $m = 1, 2, \ldots, M$, $t = t' - mT_s$, $x_k^{(m)}$ is defined as a transmitted symbol on the kth subcarrier at the mth OFDM block and q(t) is the pulse shaping filter:

$$q(t) = \begin{cases} 1, & t \in [-T_{cp}, T], \\ 0, & otherwise. \end{cases} \tag{3}$$

The UWA channel is a typical sparse multipath channel. For horizontal shallow water multi-path transmission with a range much greater than depth, each path has almost the same Doppler factor [38]. Thus, we assume that all the paths have the equal Doppler scale factor *a*, and the Doppler effect can be almost eliminated through effective Doppler estimation and compensation. Furthermore, the path delay remains stable across several consecutive OFDM blocks, and the path gains and Doppler scale factors are constant during

one OFDM block, but vary from block to block. The time-varying channel response during the mth OFDM block can be expressed as

$$h^{(m)}(\tau, t) = \sum_{l=0}^{L-1} A_l^{(m)} \delta\left[\tau - \left(\tau_l - a^{(m)}t\right)\right], \tag{4}$$

where L is the path number of the multipath channel, which may be distributed in more than one cluster. A_l and τ_l denote the gain and delay of the lth path.

Through the time-varying channel, the received passband signal during the mth OFDM block can be expressed as

$$\widetilde{y}^{(m)}(t) = \sum_{l=0}^{L-1} A_l^{(m)} \widetilde{x}^{(m)}\left[\left(1 + a^{(m)}\right)t - \tau_l\right] + \widetilde{w}^{(m)}(t), \tag{5}$$

where $\widetilde{w}(t)$ is the additive noise. After synchronization, a popular two-step approach [39] can be adopted to mitigate the Doppler effect, which takes a coarse Doppler estimation and resampling as the first step and then performs the fine Doppler shift compensation using the null subcarriers. Moreover, CP can also be used for Doppler shift estimation [40]. Performing CP-OFDM demodulation, the output $K \times 1$ vector of the mth OFDM block can be expressed as

$$\mathbf{y}^{(m)} = \mathbf{X}^{(m)}\mathbf{F}\mathbf{h}^{(m)} + \mathbf{w}^{(m)}, \tag{6}$$

where $\mathbf{X}^{(m)}$ is an $K \times K$ diagonal matrix whose diagonal entries are the K transmitted symbols, $\mathbf{F}$ is an $K \times L$ discrete Fourier transform (DFT) matrix and $\mathbf{w}^{(m)}$ is an $K \times 1$ additive noise vector. $\mathbf{h}^{(m)}$ is the discretized CIR vector as $\mathbf{h}^{(m)} = [h_1^{(m)}, h_2^{(m)}, \cdots, h_L^{(m)}]$.

For M consecutive OFDM blocks, the transmission model of all $K \times M$ symbols can be expressed as

$$\mathbf{Y} = \left[\boldsymbol{\Phi}^{(1)}\mathbf{h}^{(1)}, \boldsymbol{\Phi}^{(2)}\mathbf{h}^{(2)}, \cdots, \boldsymbol{\Phi}^{(M)}\mathbf{h}^{(M)}\right] + \mathbf{W}, \tag{7}$$

where $\mathbf{Y} = [\mathbf{y}^{(1)}, \mathbf{y}^{(2)}, \cdots, \mathbf{y}^{(M)}]$, $\mathbf{W} = [\mathbf{w}^{(1)}, \mathbf{w}^{(2)}, \cdots, \mathbf{w}^{(M)}]$, and $\mathbf{H} = [\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \cdots, \mathbf{h}^{(M)}]$. $\boldsymbol{\Phi}^{(\mathbf{m})}$ is defined as $\boldsymbol{\Phi}^{(\mathbf{m})} \triangleq \mathbf{X}^{(m)}\mathbf{F}$. Considering the comb-type pilot arrangement, the transmission model of $K_p \times M$ pilot symbols can be expressed as

$$\mathbf{Y_P} = \boldsymbol{\Phi_P}\mathbf{H} + \mathbf{W_P}, \tag{8}$$

where $\mathbf{Y_P} = [\mathbf{y_p}^{(1)}, \mathbf{y_p}^{(2)}, \cdots, \mathbf{y_p}^{(M)}]$ and $\mathbf{W_P} = [\mathbf{w_p}^{(1)}, \mathbf{w_p}^{(2)}, \cdots, \mathbf{w_p}^{(M)}]$ are the submatrixs of $\mathbf{Y}$ and $\mathbf{W}$ at the location of the pilot subcarriers, respectively. $\boldsymbol{\Phi_P} \triangleq \mathbf{X_P}\mathbf{F_P}$ is the known dictionary matrix where $\mathbf{X_P}$ is the $K_p \times K_p$ diagonal matrix with the known pilots along its diagonal and $\mathbf{F_P}$ is the $K_p \times L$ DFT matrix.

By vectorizing $\mathbf{H}$ as $\mathbf{h} = vec(\mathbf{H}^T)$, a pilot-based channel measurement model can be constructed according to (8), as

$$\mathbf{y_p} = \mathbf{U_P}\mathbf{h} + \mathbf{w_p}, \tag{9}$$

where $\mathbf{y_p} = vec(\mathbf{Y_P}^T)$, $\mathbf{w_p} = vec(\mathbf{W_P}^T)$ and $\mathbf{U_P} = \boldsymbol{\Phi_P} \otimes \mathbf{I}_M$. AWGN can be adopted to approximate the additive noise $\mathbf{w_p}$. However, the attenuation and noise of the UWA channel may vary over the signal bandwidth. When a small number of pilot symbols are used to estimate the UWA channel with strong colored noise, the accuracy of the pilot-based channel measurement will decrease severely. To improve the performance of channel measurement, a noise-resistant channel measurement model is constructed according to (7) and it can be expressed as

$$\mathbf{y} = \mathbf{U}\mathbf{h} + \mathbf{w}, \tag{10}$$

where $\mathbf{y} = vec(\mathbf{Y}^T)$, $\mathbf{w} = vec(\mathbf{W}^T)$ and $\mathbf{U}$ is the $KM \times LM$ matrix as

$$\mathbf{U} = \begin{bmatrix} diag\{\boldsymbol{\Phi}^{(1)}[1,1], \cdots, \boldsymbol{\Phi}^{(M)}[1,1]\} & \cdots & diag\{\boldsymbol{\Phi}^{(1)}[1,L], \cdots, \boldsymbol{\Phi}^{(M)}[1,L]\} \\ \vdots & \ddots & \vdots \\ diag\{\boldsymbol{\Phi}^{(1)}[K,1], \cdots, \boldsymbol{\Phi}^{(M)}[K,1]\} & \cdots & diag\{\boldsymbol{\Phi}^{(1)}[K,L], \cdots, \boldsymbol{\Phi}^{(M)}[K,L]\} \end{bmatrix}, \quad (11)$$

where $\boldsymbol{\Phi}^{(m)}[i,j]$ denotes the entry in the ith row and the jth column of $\boldsymbol{\Phi}^{(m)}$. All subcarriers are used for channel measurement in (10), as the unknown data symbols can be obtained by data detection or decoding. Assuming that the power distribution over all subcarriers is non-uniform, $\boldsymbol{\lambda} = [\lambda_{-K/2}, \lambda_{-K/2+1}, \cdots, \lambda_{K/2-1}]$, $\mathbf{w}$ in (10) satisfies:

$$\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Lambda}), \quad (12)$$

where $\boldsymbol{\Lambda} = diag(\boldsymbol{\lambda}) \otimes \mathbf{I}_M$.

3.3. Clustered-Sparse Channel Distribution Model

According to the measurement equation in (9) or (10), the channel vector $\mathbf{h}$ can be estimated. However, the UWA channel usually has an excessive multipath delay spread. With so many undetermined channel taps, the prior clustered-sparse structure of the UWA channel can be exploited to obtain the sparse solution of the channel vector. First, we assume all that the L paths (i.e., L rows) in $\mathbf{H}$ are mutually independent, which can form C clusters. There are L_d paths in the dth cluster and these clusters do not overlap, each of which occupies dense path delays. With the same time correlation characteristics, the dth cluster satisfies a first-order auto-regressive (AR) model, given by

$$\mathbf{H}_d[i, j+1] = \beta_d \mathbf{H}_d[i,j] + \sqrt{1 - |\beta_d|^2} \mathbf{V}_d[i,j], \quad (13)$$

where β_d is the temporal correlation coefficient of the dth cluster. $\mathbf{H}_d$ is a $L_d \times M$ submatrix of $\mathbf{H}$, $i = 1, 2, \ldots, L_d$ and $j = 1, 2, \ldots, M-1$ are the row and column indices of $\mathbf{H}_d$, respectively. $\mathbf{V}_d$ is the model noise matrix with complex Gaussian distribution and assumed that $vec(\mathbf{V}_d^T) \sim \mathcal{CN}(0, \boldsymbol{\Gamma}_d \otimes \mathbf{I}_M)$, where:

$$\boldsymbol{\Gamma}_d = diag(\gamma_{d,1}, \gamma_{d,2}, \ldots, \gamma_{d,L_d}) \quad (14)$$

is a positive semi-definite diagonal real-valued matrix. Although a higher-order AR model may express better model approximation performance, it would lead to higher complexity. More importantly, a higher-order AR model means a higher overfitting risk, especially when there are only a few OFDM blocks in one frame. Hence, a first-order AR process was adopted to model the temporal correlation characteristic for each cluster of CIRs.

By vectorizing $\mathbf{H_d}$ as $\mathbf{h_d} = vec(\mathbf{H}_d^T)$, the parameteric Bayesian prior of $\mathbf{h}_d$ can be written by

$$p(\mathbf{h}_d; \beta_d, \gamma_d) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Gamma}_d \otimes \mathbf{B}_d), \quad (15)$$

where $\gamma_d = [\gamma_{d,1}, \gamma_{d,2}, \cdots, \gamma_{d,L_d}]$ controls the sparsity of the L_d paths in the dth cluster. When $Tr(\boldsymbol{\Gamma}_d) \to 0$, the associated $\mathbf{h}_d \to \mathbf{0}$. In other words, $\boldsymbol{\Gamma}_d$ reflects the probability density of the dth cluster. $\mathbf{B}_d$ is a Toeplitz matrix with the following form:

$$\mathbf{B}_d = Toeplitz(1, \beta_d, \ldots, \beta_d^{M-1}), \quad (16)$$

and it reflects the variation characteristics of the dth cluster. $\boldsymbol{\Gamma}_d$ and $\mathbf{B}_d$ jointly determine the cluster distribution of the dth cluster. Accordingly, the prior probability density function of $\mathbf{h}$ can be written as

$$p(\mathbf{h}; \gamma_1, \gamma_2, \ldots, \gamma_C, \beta_1, \beta_2, \ldots, \beta_C) = \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma}_0), \quad (17)$$

where $\boldsymbol{\Sigma}_0 = diag(\boldsymbol{\Gamma}_1 \otimes \mathbf{B}_1, \boldsymbol{\Gamma}_2 \otimes \mathbf{B}_2, \ldots, \boldsymbol{\Gamma}_C \otimes \mathbf{B}_C)$.

4. Bayesian Learning-Based Clustered-Sparse Channel Estimation

In this section, the parametric form of Bayesian channel estimation is given first, and followed by the cluster distribution learning. Then, the partition-based clustered-sparse Bayesian learning algorithm is presented to obtain the clustered-sparse distribution and Bayesian channel estimate. At last, the complexity and performance analysis of this algorithm is given.

4.1. Bayesian Channel Estimation

To simplify the implementation complexity of Bayesian channel estimation, only known pilot symbols are used for channel measurement. As pilot symbols are sparsely distributed, it is difficult to estimate the non-uniform power spectrum at the location of pilot subcarriers. Therefore, the additive noise in (9) is approximated as white Gaussian noise with noise variance λ in this section. According to (9) and (17), using the Bayes rule, the posterior density of $\mathbf{h}$ can be obtained as

$$p(\mathbf{h}|\mathbf{y_P}; \lambda, \gamma_1, \gamma_2 \ldots, \gamma_C, \beta_1, \beta_2, \ldots \beta_C) = \mathcal{CN}(\boldsymbol{\mu_h}, \boldsymbol{\Sigma_h}), \tag{18}$$

which is the complex Gaussian distribution with the posterior mean and covariance matrix:

$$\boldsymbol{\mu_h} = \lambda^{-1}\boldsymbol{\Sigma_h}\mathbf{U_P^H}\mathbf{y_P} = \boldsymbol{\Sigma}_0\mathbf{U_P^H}\left(\lambda\mathbf{I} + \mathbf{U_P}\boldsymbol{\Sigma}_0\mathbf{U_P^H}\right)^{-1}\mathbf{y_P}, \tag{19}$$

$$\boldsymbol{\Sigma_h} = \left(\boldsymbol{\Sigma}_0^{-1} + \lambda^{-1}\mathbf{U_P^H}\mathbf{U_P}\right)^{-1} = \boldsymbol{\Sigma}_0 - \boldsymbol{\Sigma}_0\mathbf{U_P^H}\left(\lambda\mathbf{I} + \mathbf{U_P}\boldsymbol{\Sigma}_0\mathbf{U_P^H}\right)^{-1}\mathbf{U_P}\boldsymbol{\Sigma}_0. \tag{20}$$

Obviously, the maximum posterior (MAP) estimates $\widehat{\mathbf{h}}_{MAP} \triangleq \boldsymbol{\mu_h}$ and the covariance matrix $\boldsymbol{\Sigma_h}$ are related to the hyperparameters $[\lambda, \gamma_1, \gamma_2, \cdots, \gamma_C, \beta_1, \beta_2, \cdots, \beta_C]$. Using $K_n \times M$ null subcarriers $\mathbf{Y_N} = [\mathbf{y}_n^{(1)}, \cdots, \mathbf{y}_n^{(M)}]$, λ can be estimated as

$$\lambda = \frac{1}{K}E[\| \mathbf{y}^{(m)} \|_2^2] \approx \frac{1}{MK_n}\sum_{j=1}^{M}\sum_{i=1}^{K_n}|\mathbf{Y_N}[i, j]|^2, \tag{21}$$

where i and j are the row and column indexes of $\mathbf{Y_N}$, respectively. The remaining hyperparameters $[\gamma_1, \gamma_2, \cdots, \gamma_C, \beta_1, \beta_2, \cdots, \beta_C]$ are related to the cluster distribution information.

4.2. Cluster Distribution Learning

The hyperparameters $\boldsymbol{\theta} \triangleq [\gamma_1, \gamma_2, \cdots, \gamma_C, \beta_1, \beta_2, \cdots, \beta_C]$ can be estimated by maximizing the marginal likelihood function $p(\mathbf{y_P}; \boldsymbol{\theta})$. This is equivalent to minimizing $-log\, p(\mathbf{y_P}; \boldsymbol{\theta})$, giving the effective cost function:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbf{y_P^H}\boldsymbol{\Sigma}_{\mathbf{y_P}}^{-1}(\boldsymbol{\theta})\mathbf{y_P} + log|\boldsymbol{\Sigma_y}(\boldsymbol{\theta})|, \tag{22}$$

where $\boldsymbol{\Sigma}_{\mathbf{y_P}}(\boldsymbol{\theta}) = \lambda\mathbf{I} + \mathbf{U_P}\boldsymbol{\Sigma}_0(\boldsymbol{\theta})\mathbf{U_P^H}$. The above problem cannot be solved in closed form, and the expectation maximization (EM) method [41] can be employed to solve it iteratively. For unknown values of hyperparameters governing the prior density, $\mathbf{h}$ is considered as the nuisance variable and $\boldsymbol{\theta}$ is estimated by maximizing:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(old)}) = E_{\mathbf{h}|\mathbf{y_P};\boldsymbol{\theta}^{(old)}}[log\, p(\mathbf{y_P}, \mathbf{h}; \boldsymbol{\theta})] = E_{\mathbf{h}|\mathbf{y_P};\boldsymbol{\theta}^{(old)}}[log\, p(\mathbf{y_P}|\mathbf{h}; \lambda)] + E_{\mathbf{h}|\mathbf{y_P};\boldsymbol{\theta}^{(old)}}[log\, p(\mathbf{h}; \boldsymbol{\theta})], \tag{23}$$

where $\boldsymbol{\theta}^{(old)}$ denotes the estimated hyperparameters in the previous iteration. Ignoring

the first term unrelated to θ, the Q Function (23) can be simplified to:

$$Q(\beta_1, \beta_2, \ldots, \beta_C, \gamma_1, \gamma_2, \ldots, \gamma_C) = E_{\mathbf{h}|\mathbf{y_p}; \theta^{(old)}} \{ log\, p[\mathbf{h}; \mathbf{B}_1, \mathbf{B}_2, \cdots, \mathbf{B}_C, \mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \cdots, \mathbf{\Gamma}_C] \}$$

$$\propto -\frac{1}{2} \sum_{d=1}^{C} \left[log\left(|\mathbf{\Gamma}_d|^M |\mathbf{B}_d|^{L_d} \right) + \mathbf{h}_d^H (\mathbf{\Gamma}_d^{-1} \otimes \mathbf{B}_d^{-1}) \mathbf{h}_d \right] , \tag{24}$$

$$\propto -\frac{1}{2} \sum_{d=1}^{C} \left\{ M log\left(|\mathbf{\Gamma}_d| \right) + L_d log\left(|\mathbf{B}_d| \right) + Tr\left[\left(\mathbf{\Gamma}_d^{-1} \otimes \mathbf{B}_d^{-1} \right) \left(\mathbf{\Sigma}_{\mathbf{h}}^d + \boldsymbol{\mu}_{\mathbf{h}}^d (\boldsymbol{\mu}_{\mathbf{h}}^d)^H \right) \right] \right\}$$

where $\mathbf{h}_d \in \mathbb{C}^{ML_d \times 1}$, $\boldsymbol{\mu}_{\mathbf{h}}^d \in \mathbb{C}^{ML_d \times 1}$ and $\mathbf{\Sigma}_{\mathbf{h}}^d \in \mathbb{C}^{ML_d \times ML_d}$ denote the corresponding dth cluster in $\mathbf{h} \in \mathbb{C}^{ML \times 1}$, $\boldsymbol{\mu}_{\mathbf{h}} \in \mathbb{C}^{ML \times 1}$ and $\mathbf{\Sigma}_{\mathbf{h}} \in \mathbb{C}^{ML \times ML}$, respectively. $\boldsymbol{\mu}_{\mathbf{h}}$ and $\mathbf{\Sigma}_{\mathbf{h}}$ are evaluated according to (19) and (20), given the estimated hyperparameters $\theta^{(old)}$ and the noise variance λ. As shown in (14) and (16), $\mathbf{\Gamma}_d$ and $\mathbf{B}_d$ are determined by the hyperparameters γ_d and β_d, respectively.

The partial derivative of (24) with respect to $\gamma_{d,i}$ ($d = 1, 2, \ldots, C; i = 1, 2 \ldots, L_d$) is given by

$$\frac{\partial Q}{\partial \gamma_{d,i}} = -\frac{M}{2\gamma_{d,i}} + \frac{1}{2\gamma_{d,i}^2} Tr\left[\mathbf{B}_d^{-1} \left(\mathbf{\Sigma}_{\mathbf{h}}^{d,i} + \boldsymbol{\mu}_{\mathbf{h}}^{d,i} \left(\boldsymbol{\mu}_{\mathbf{h}}^{d,i} \right)^H \right) \right], \tag{25}$$

where $\boldsymbol{\mu}_{\mathbf{h}}^{d,i} \in \mathbb{C}^{M \times 1}$ denotes the corresponding ith path in $\boldsymbol{\mu}_{\mathbf{h}}^d$, and $\mathbf{\Sigma}_{\mathbf{h}}^{d,i} \in \mathbb{C}^{M \times M}$ is the corresponding principal diagonal block for the ith path in $\mathbf{\Sigma}_{\mathbf{h}}^d$. So the learning rule for $\gamma_{d,i}$ can be given by

$$\gamma_{d,i} = \frac{1}{M} Tr\left[\mathbf{B}_d^{-1} \left(\mathbf{\Sigma}_{\mathbf{h}}^{d,i} + \boldsymbol{\mu}_{\mathbf{h}}^{d,i} \left(\boldsymbol{\mu}_{\mathbf{h}}^{d,i} \right)^H \right) \right]. \tag{26}$$

The gradient of (24) over $\mathbf{B}_d$ ($d = 1, 2, \ldots, C$) is:

$$\frac{\partial Q}{\partial \mathbf{B}_d} = -\frac{L_d}{2} \mathbf{B}_d^{-1} + \frac{1}{2} \sum_{i=1}^{L_d} \frac{1}{\gamma_{d,i}} \mathbf{B}_d^{-1} \left(\mathbf{\Sigma}_{\mathbf{h}}^{d,i} + \boldsymbol{\mu}_{\mathbf{h}}^{d,i} \left(\boldsymbol{\mu}_{\mathbf{h}}^{d,i} \right)^H \right) \mathbf{B}_d^{-1}. \tag{27}$$

Thus, the learning rule for $\mathbf{B}_d$ can be derived as

$$\mathbf{B}_d = \frac{1}{L_d} \sum_{i=1}^{L_d} \frac{\mathbf{\Sigma}_{\mathbf{h}}^{d,i} + \boldsymbol{\mu}_{\mathbf{h}}^{d,i} \left(\boldsymbol{\mu}_{\mathbf{h}}^{d,i} \right)^H}{\gamma_{d,i}}. \tag{28}$$

here, we constrain $\mathbf{B}_d$ to have the form as shown in (16), and thus β_d can be empirically calculated as $\beta_d \triangleq \frac{\alpha_1}{\alpha_0}$ where α_0 (resp. α_1) is the average of the elements along the main diagonal (resp. the main sub-diagonal) of the matrix $\mathbf{B}_d$ in (28).

4.3. Partition-Based Clustered-Sparse Bayesian Learning Algorithm

When the cluster partition information is known, the hyperparameters θ, the posterior mean $\boldsymbol{\mu}_{\mathbf{h}}$ and the posterior covariance $\mathbf{\Sigma}_{\mathbf{h}}$ can be iteratively obtained through the EM method. In other words, cluster partition and cluster distribution jointly determine the clustered-sparse channel distribution. However, it is usually difficult to obtain the cluster partition in advance. As shown in Figure 1, several significant clusters with high energy in CIRs are scattered over a long time spread, and usually do not overlap each other. As shown in (28), the boundary information of these clusters is important for calculating $\mathbf{B}_d$, because different clusters have different coherence times (i.e., different AR coefficients). In particular, some weak clusters that can be ignored will cause the hyperparameter estimation error of significant clusters when these clusters are not distinguished. Therefore, the PB-CSBL algorithm is proposed to update cluster partition by pruning the paths with very small energy (i.e., $\gamma_{d,i} \to 0$). As shown in Algorithm 1, the PB-CSBL algorithm mainly includes three sub-parts: Channel Estimation, Cluster Partition, Cluster Evolution, and they will loop until the iteration stop condition is reached. In the sub-part of Cluster Partition,

only the dense paths with high energy are retained to form clusters and the paths with negligible energy are pruned. For a cluster with some pruned paths, it can further split into more clusters according to the positions of the pruned paths. To reduce the impact of an inaccurate clustering structure at the beginning of the iteration, the maximum possible number of discrete paths in one cluster is given. In the sub-part of Cluster Evolution, the derived cluster distribution learning rules are included. A good initial estimation of the unknown hyperparameters is significant for achieving the global maximum instead of a local maximum. For practical implementation, it was found that an initial estimation given by $\mathbf{\Gamma}_d = \mathbf{I}_{L_d}$ and $\mathbf{B}_d = \mathbf{I}_M$ was sufficient for the proposed PB-CSBL algorithm.

Algorithm 1: Partition-Based Clustered-Sparse Bayesian Learning Algorithm

---

Input: the received pilot signal $\mathbf{y_p}$; the dictionary matrix $\mathbf{U_P}$; the noise variance λ; the length of discrete paths L; the maximum number of iterations r_{max}; the maximum number of discrete paths in one cluster LC_{max}; the threshold for prunning small hyperparameters γ_{th}; the threshold to stop the whole algorithm ϵ.

Initialize: the list of path power $\gamma_{list} = \mathbf{1}_L$; the number of clusters $C = 1$; the list of cluster structure **CLS** : $\mathbf{\Gamma}_1 = \mathbf{I}_L$, $\mathbf{B}_1 = \mathbf{I}_M$ and the delay range of the 1st cluster $R_d = [0, L-1]$; the iteration counter $r = 0$.

Channel Estimation:

1: $\mathbf{\Sigma_0} \leftarrow diag(\mathbf{\Gamma}_1 \otimes \mathbf{B}_1, \mathbf{\Gamma}_2 \otimes \mathbf{B}_2, \ldots, \mathbf{\Gamma}_C \otimes \mathbf{B}_C)$.

2: $\mathbf{\Sigma_h} = \mathbf{\Sigma_0} - \mathbf{\Sigma_0} \mathbf{U_P}^H (\lambda \mathbf{I} + \mathbf{U_P} \mathbf{\Sigma_0} \mathbf{U_P}^H)^{-1} \mathbf{U_P} \mathbf{\Sigma_0}$.

3: $\boldsymbol{\mu_h} \leftarrow \lambda^{-1} \mathbf{\Sigma_h} \mathbf{U_P}^H \mathbf{y_p}$.

Cluster Evolution:

4: **for** $d = 1, 2, \ldots, C$ **do**

5: 　**for** $i = 1, 2, \ldots, L_d$ **do**

6: 　　$\gamma_{d,i} \leftarrow \frac{1}{M} Tr\left[\mathbf{B}_d^{-1}\left(\mathbf{\Sigma_h}^{d,i} + \boldsymbol{\mu_h}^{d,i}\left(\boldsymbol{\mu_h}^{d,i}\right)^H\right)\right]$, and update γ_{list} with $\gamma_{d,i}$.

7: 　**end for**

8: 　$L'_d \leftarrow min(L_d, LC_{max})$.

9: 　$\mathbf{B}_d \leftarrow \frac{1}{L'_d} \sum_{i=1}^{L'_d} \frac{\mathbf{\Sigma_h}^{d,i} + \boldsymbol{\mu_h}^{d,i}\left(\boldsymbol{\mu_h}^{d,i}\right)^H}{\gamma_{d,i}}$ using the L'_d most significant continuous paths.

10: 　$\beta_d \leftarrow \frac{\alpha_1}{\alpha_0}$, where α_1 and α_0 can be obtained through $\mathbf{B}_d$.

11: **end for**

Cluster Partition:

12: $\mathbf{p}^{(0)} \leftarrow \text{FindIndex}(0 < \gamma_{list} < \gamma_{th})$.

13: **if** $\mathbf{p}^{(0)}$ is not empty **then**

14: 　$\gamma_{list}^{(old)} \leftarrow \gamma_{list}$ and $\gamma_{list}[\mathbf{p}^{(0)}] \leftarrow \mathbf{0}$.

15: 　$\mathbf{CLS} \leftarrow \text{Split}(\mathbf{CLS}, \mathbf{p}^{(0)})$, where **CLS** is splitted into C clusters according to $\mathbf{p}^{(0)}$.

16: 　$\mathbf{p}^{(1)} \leftarrow \text{FindIndex}(\gamma_{list} \geq \gamma_{th})$ and $\mathbf{U_P} \leftarrow \mathbf{U_P}[:, \mathbf{p}^{(1)} \otimes (1:M)]$.

17: **end if**

18: **for** $d = 1, 2, \ldots, C$ **do**

19: 　Update R_d, $\mathbf{\Gamma}_d$ and $\mathbf{B}_d$ in **CLS** according to $\mathbf{p}^{(0)}$, γ_{list} and β_d, respectively.

20: **end for**

Check stopping conditions:

21: $r \leftarrow r + 1$.

22: **return** the sub-part of **Channel Estimation** until $r \geq r_{max}$ or $\|\gamma_{list} - \gamma_{list}^{(old)}\|_2^2 < \epsilon$.

Output: the pruned channel $\boldsymbol{\mu_h}$ with the covariance matrix $\mathbf{\Sigma_h}$ and the cluster list **CLS**.

4.4. Complexity and Performance Analysis

Considering the excessive delay spread of the UWA channel and the purpose of saving pilot overhead, we assumed $K_d \geq L \geq K_p \geq M$. The time complexity of the proposed PB-CSBL algorithm is compared with the least squares (LS) [39], OMP [25], SOMP [27] and TMSBL [32]. In order to avoid involving matrix inversion, pilot subcarriers are equispaced in the LS method. For fair comparison, the residual Doppler expansion is ignored when evaluating the time complexity of the OMP method. The comparison of time complexity is shown in Table 1, where $\overline{N}$ denotes the average iteration number. It can be seen that the time complexity of the PB-CSBL algorithm is higher than that of the LS, OMP and SOMP. This is due to the fact that the temporal correlation of CIRs is ignored in the LS and OMP. Although the stability of path delays is exploited in the SOMP, the correlation of path gains is discarded. Therefore, LS, OMP and SOMP may exhibit worse channel estimation performance compared to the PB-CSBL algorithm. As for the TMSBL with $\mathcal{O}(LK_p^2)$ per iteration, it has lower time complexity than the proposed PB-CSBL with $\mathcal{O}(LK_p^2 M^3)$ per iteration. This is mainly because it adopts the measure in [41], assuming that there is no temporal correlation among consecutive CIRs or no additive noise in received signal, but this may bring some performance loss, especially for slow-varying or high-noise channels. Therefore, when M is small, we do not adopt the simplification in [41]. The computational complexity of the PB-CSBL algorithm can be further reduced by implementing a first-order algorithm in E-step [34], but the implementation details are omitted here.

Table 1. The comparison of time complexity.

LS	OMP	SOMP	TMSBL	PB-CSBL
$\mathcal{O}(LK_pM)$	$\mathcal{O}(L^2K_pM)$	$\mathcal{O}(L^2K_pM)$	$\mathcal{O}(LK_p^2\overline{N})$	$\mathcal{O}(LK_p^2M^3\overline{N})$

According to (20), the mean square error (MSE) bound of the sparse vector $\mathbf{h}$ can be expressed as

$$E\left[\|\mathbf{h} - \widehat{\mathbf{h}}\|_2^2\right] \geq Tr(\mathbf{\Sigma_h}) = Tr\left[\mathbf{\Sigma}_0 - \mathbf{\Sigma}_0\mathbf{U}_\mathbf{P}^H\left(\lambda\mathbf{I} + \mathbf{U_P}\mathbf{\Sigma}_0\mathbf{U}_\mathbf{P}^H\right)^{-1}\mathbf{U_P}\mathbf{\Sigma}_0\right], \qquad (29)$$

where $\widehat{\mathbf{h}}$ is the estimate of $\mathbf{h}$, and its estimation accuracy is related to the hyperparameter vector $\boldsymbol{\theta}$ obtained by maximizing the marginal likelihood function. Thus, the lower bound of MSE is reached (i.e., equality holds) only when the perfect hyperparameters are obtained.

Assume that all L paths are in one cluster (i.e., $C = 1$) and consider two special cases of $\boldsymbol{\theta}$:

(1) The temporal correlation coefficient $\beta = 0$.

In this case, there is no temporal correlation among CIRs. It is assumed that each transmitted symbol is normalized to unit power. Then, the MSE bound of $\mathbf{h}^{(m)}$ can be expressed by

$$E\left[\|\mathbf{h}^{(m)} - \widehat{\mathbf{h}}^{(m)}\|_2^2\right] \geq Tr\left[(\mathbf{\Gamma}^{-1} + \lambda^{-1}\mathbf{\Phi}_\mathbf{P}^H\mathbf{\Phi_P})^{-1}\right] = \sum_{l=1}^{L}\left(\frac{K_p}{\lambda} + \frac{1}{\gamma_l}\right)^{-1}, \qquad (30)$$

where $\mathbf{\Gamma} = diag(\boldsymbol{\gamma})$, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_L]$ and γ_l controls the variance of the lth channel coefficient.

(2) The temporal correlation coefficient $\beta = 1$

In this case, the CIRs are time-invariant across M OFDM blocks. Thus, the additive noise variance λ can be reduced to λ/M, and the MSE bound of $\mathbf{h}^{(m)}$ can be expressed by

$$E\left[\|\mathbf{h}^{(m)} - \widehat{\mathbf{h}}^{(m)}\|_2^2\right] \geq \sum_{l=1}^{L}\left(\frac{MK_p}{\lambda} + \frac{1}{\gamma_l}\right)^{-1}. \qquad (31)$$

As shown in the case of the high temporal correlation (i.e., $\beta = 1$), the anti-noise performance of channel estimation can be improved. As for the low temporal correlation among CIRs (i.e., $\beta = 0$), the anti-noise performance is not significantly improved. However, with more reference information, the better hyperparameter vector γ can be obtained. Therefore, when multiple OFDM blocks are utilized simultaneously, higher channel estimation accuracy can be achieved.

5. Noise-Corrected Clustered-Sparse Channel Estimation

In this section, the parametric form of noise-resistant Bayesian channel estimation is given first. Then, the data detection and noise measurement are derived. Then, the noise-corrected clustered-sparse channel estimation algorithm is presented. At last, the complexity and performance analysis of this algorithm is given.

5.1. Noise-Resistant Bayesian Channel Estimation

As the number of pilots K_p decreases, the precision of channel estimation using the PB-CSBL algorithm will rapidly deteriorate. This is primarily because the algorithm does not utilize all the information available. Specifically, although both pilot and data symbols are transmitted, only the $K_p \times M$ matrix $\mathbf{Y_P}$ corresponding to the pilot subcarriers in (8) is used for estimating the CIRs. For the matrix $\mathbf{Y}$ in (7), the remaining $K_d \times M$ observations are typically not used as they contain the unknown data symbols. If the unknown data symbols are perfectly obtained, using the Bayes rule, the posterior mean of CIRs based on the prior probability in (17) and the channel measurement equation in (10) can be written by

$$\overline{\mu}_{\mathbf{h}} = \Sigma_0 \mathbf{U}^H \left(\Lambda + \mathbf{U}\Sigma_0 \mathbf{U}^H \right)^{-1} \mathbf{y}, \tag{32}$$

with the posterior covariance matrix:

$$\overline{\Sigma}_{\mathbf{h}} = \Sigma_0 - \Sigma_0 \mathbf{U}^H \left(\Lambda + \mathbf{U}\Sigma_0 \mathbf{U}^H \right)^{-1} \mathbf{U}\Sigma_0. \tag{33}$$

Moreover, AWGN is assumed to simplify implementation complexity in the PB-CSBL algorithm, as the noise power distribution $\lambda = [\lambda_{-K/2}, \lambda_{-K/2+1}, \cdots, \lambda_{K/2-1}]$ depends on the observation of all subcarriers. However, in a complex UWA environment, the noise power may vary significantly over the signal bandwidth. Using decided data symbols and corrected colored noise, the high estimation accuracy can be achieved by the MAP estimate in (32).

5.2. Data Detection and Noise Measurement

To obtain the MAP estimate in (32), the hyperparameters related to the cluster distribution, $\theta \triangleq [\gamma_1, \gamma_2, \cdots, \gamma_C, \beta_1, \beta_2, \cdots, \beta_C]$, can be iteratively updated, as shown in the proposed PB-CSBL algorithm, however, the detection of data signal $\mathbf{x_d}^{(m)}$ and the noise distribution λ are still unresolved. Using the notations $\mathbf{x_d} \triangleq \left[(\mathbf{x_d}^{(1)})^T, (\mathbf{x_d}^{(2)})^T, \cdots, (\mathbf{x_d}^{(M)})^T \right]$, $\varphi \triangleq [\lambda, \mathbf{x_d}]$, and $\zeta \triangleq [\theta, \varphi]$, the unknown hyperparameters set ζ can be estimated by maximizing the marginal likelihood function $p(\mathbf{y}; \zeta)$. This is equivalent to minimizing $-log\, p(\mathbf{y}; \zeta)$, giving the effective cost function:

$$\mathcal{L}(\zeta) = \mathbf{y}^H \Sigma_{\mathbf{y}}^{-1}(\zeta)\mathbf{y} + log|\Sigma_{\mathbf{y}}(\zeta)|, \tag{34}$$

where $\Sigma_{\mathbf{y}}(\zeta) = \Lambda(\lambda) + \mathbf{U}(\mathbf{x_d})\Sigma_0(\theta)\mathbf{U}^H(\mathbf{x_d})$. Similar to the cluster distribution learning, the expectation maximization (EM) method [41] can be employed to solve (34) iteratively. While $\mathbf{h}$ is considered as the nuisance variable, ζ can be estimated by maximizing:

$$Q(\zeta|\zeta^{(old)}) = E_{\mathbf{h}|\mathbf{y};\zeta^{(old)}}[log\, p(\mathbf{y}, \mathbf{h}; \zeta)] = E_{\mathbf{h}|\mathbf{y};\zeta^{(old)}}[log\, p(\mathbf{y}|\mathbf{h}; \varphi)] + E_{\mathbf{h}|\mathbf{y};\zeta^{(old)}}[log\, p(\mathbf{h}; \theta)], \tag{35}$$

where $\zeta^{(old)}$ denotes the estimated hyperparameters in the previous iteration. As shown in (35), the joint maximization simplifies into two independent maximizations over $\boldsymbol{\theta}$ and $\boldsymbol{\varphi}$. On optimizing the latter function in (35) with respect to $\boldsymbol{\theta}$, one obtains the learning rules about $\boldsymbol{\theta}$ as (26) and (28). Maximizing the former over $\mathbf{x}_{\mathbf{d}}^{(m)}$ yields the minimum mean squared error (MMSE) equalizer:

$$\widehat{\mathbf{x}}_{\mathbf{d}}^{(m)} = \left[diag(\boldsymbol{\lambda}_d)\mathbf{C}_{\mathbf{X}}^{-1} + Prdiag\left(\mathbf{F}_{\mathbf{D}}\Xi_{\mathbf{h}}^{(m)}\mathbf{F}_{\mathbf{D}}^{H} \right) \right]^{-1} diag\left[\left(\overline{\boldsymbol{\mu}}_{\mathbf{h}}^{(m)} \right)^{H} \mathbf{F}_{\mathbf{D}}^{H} \right] \mathbf{y}_{\mathbf{d}}^{(m)}, \qquad (36)$$

where $\mathbf{y}_{\mathbf{d}}^{(m)}$ and $\boldsymbol{\lambda}_d$ are the subvectors of $\mathbf{y}^{(m)}$ and $\boldsymbol{\lambda}$ at the location of data subcarriers, respectively. $\mathbf{C}_{\mathbf{X}}$ is the covariance matrix of $\mathbf{x}_{\mathbf{d}}^{(m)}$, $\mathbf{F}_{\mathbf{D}}$ is the $K_d \times L$ DFT matrix at the location of data subcarriers. $\Xi_{\mathbf{h}}^{(m)}$ is defined as $\Xi_{\mathbf{h}}^{(m)} \triangleq \overline{\boldsymbol{\mu}}_{\mathbf{h}}^{(m)}\left(\overline{\boldsymbol{\mu}}_{\mathbf{h}}^{(m)}\right)^{H} + \overline{\boldsymbol{\Sigma}}_{\mathbf{h}}^{(m)}$, where $\overline{\boldsymbol{\mu}}_{\mathbf{h}}^{(m)} \in \mathbb{C}^{L \times 1}$ and $\overline{\boldsymbol{\Sigma}}_{\mathbf{h}}^{(m)} \in \mathbb{C}^{L \times L}$ denote the posterior mean and covariance matrix of the mth OFDM block, respectively. The decisions on the data symbols, $\mathbf{x}_{\mathbf{d}}^{(m)}$, are obtained by mapping the soft decisions, $\widehat{\mathbf{x}}_{\mathbf{d}}^{(m)}$, to the closest point on the constellation, meaning that:

$$\mathbf{x}_{\mathbf{d}}^{(m)} = Quantization\left(\widehat{\mathbf{x}}_{\mathbf{d}}^{(m)} \right). \qquad (37)$$

Although the noise distribution is non-uniform, the noise powers of adjacent subcarriers are usually close (especially when the number of subcarriers is large). Then, we divide all subcarriers into non-overlapping Q groups, so that K/Q subcarriers can be grouped into the same one. Thus, the average noise power of the qth group can be calculated as

$$\lambda_g^{(q)} = \frac{Q}{KM}Tr\left[\mathbf{y}^{(q)}(\mathbf{y}^{(q)})^{H} - \mathbf{U}^{(q)}\Xi_{\mathbf{h}}(\mathbf{U}^{(q)})^{H} \right], \qquad (38)$$

where $\Xi_{\mathbf{h}}$ is defined as $\Xi_{\mathbf{h}} \triangleq \overline{\boldsymbol{\mu}}_{\mathbf{h}}(\overline{\boldsymbol{\mu}}_{\mathbf{h}})^{H} + \overline{\boldsymbol{\Sigma}}_{\mathbf{h}}$. $\mathbf{U} \in \mathbb{C}^{KM \times LM}$ can be calculated according to (11), as unknown data symbols are estimated by (36) and (37). $\mathbf{U}^{(q)} \in \mathbb{C}^{\frac{KM}{Q} \times LM}$ is the submatrix of $\mathbf{U}$ and denotes the qth group. $\mathbf{y}^{(q)}$ denotes the qth group of $\mathbf{y}$. After $\lambda_g^{(1)}, \lambda_g^{(2)}, \ldots, \lambda_g^{(Q)}$ are calculated according to (38), the non-uniform noise power distribution across all subcarriers, $\boldsymbol{\lambda} = [\lambda_{-K/2}, \lambda_{-K/2+1}, \cdots, \lambda_{K/2-1}]$, can be obtained by upsampling the Q average noise powers.

5.3. Noise-Corrected Clustered-Sparse Channel Estimation Algorithm

Based on the noise-resistant Bayesian channel estimation, noise measurement and data detection, we can develop the PB-CSBL algorithm into the NC-CSCE Algorithm. As shown in Algorithm 2, the sub-parts of Cluster Partition and Cluster Evolution have the similar process to the PB-CSBL algorithm. With the aid of Data Detection and Noise Measurement, the estimation accuracy in Channel Estimation is improved. Conversely, the improvement of channel estimation also ensures the high accuracy of data detection and noise measurement results. To achieve the global maximum instead of a local maximum, the NC-CSCE algorithm requires a good initial estimate with respect to the unknown cluster distribution, data signal and noise distribution. However, it is obviously difficult to initialize so many hyperparameters at the same time. Therefore, the PB-CSBL algorithm is utilized to initialize $\overline{\boldsymbol{\mu}}_{\mathbf{h}}$ and $\overline{\boldsymbol{\Sigma}}_{\mathbf{h}}$ instead of initializing these hyperparameters. Once $\overline{\boldsymbol{\mu}}_{\mathbf{h}}$ and $\overline{\boldsymbol{\Sigma}}_{\mathbf{h}}$ are initialized, these hyperparameters can be calculated as shown in Algorithm 2.

5.4. Complexity and Performance Analysis

The NC-CSCE algorithm utilizes all available subcarriers to achieve better channel estimation performance. Therefore, compared with the proposed PB-CSBL algorithm with $\mathcal{O}(LK_p^2M^3)$ per iteration, the NC-CSCE algorithm has higher time complexity, i.e., $\mathcal{O}((K_d + K_p)^3M^3)$ per iteration. Similar to the PB-CSBL algorithm, the complexity of the

NC-CSCE algorithm can be further simplified by using the techniques in [34,41], however, these simplifications may lead to the performance degradation of channel estimation because of the errors introduced by the approximations. The details of the simplifications are omitted here.

Algorithm 2: Noise-Corrected Clustered-Sparse Channel Estimation Algorithm

---

Input: the received signal $\mathbf{y}$; the noise variance λ; the length of discreted paths L; the maximum number of iterations r_{max}; the maximun number of discreted paths in one cluster LC_{max}; the threshold for prunning small hyperparameters γ_{th}; the threshold to stop the whole algorithm ϵ.

Initialize: The PB-CSBL algorithm is utilized to obtain the posterior mean $\overline{\mu}_{\mathbf{h}} \leftarrow \mu_{\mathbf{h}}$ and covariance matrix $\overline{\Sigma}_{\mathbf{h}} \leftarrow \Sigma_{\mathbf{h}}$; the noise vector $\lambda = \lambda \mathbf{I}_K$; the list of path power $\gamma_{list} = \mathbf{1}_L$; the iteration counter $r = 0$.

Cluster Evolution:

1: **for** $d = 1, 2, \ldots, C$ **do**
2:　　**for** $i = 1, 2, \ldots, L_d$ **do**
3:　　　$\gamma_{d,i} \leftarrow \frac{1}{M} Tr\left[\mathbf{B}_d^{-1} \left(\overline{\Sigma}_{\mathbf{h}}^{d,i} + \overline{\mu}_{\mathbf{h}}^{d,i} \left(\overline{\mu}_{\mathbf{h}}^{d,i} \right)^H \right) \right]$, and update γ_{list} with $\gamma_{d,i}$.
4:　　**end for**
5:　　$L_d' \leftarrow min(L_d, LC_{max})$.
6:　　$\mathbf{B}_d \leftarrow \frac{1}{L_d'} \sum\limits_{i=1}^{L_d'} \frac{\overline{\Sigma}_{\mathbf{h}}^{d,i} + \overline{\mu}_{\mathbf{h}}^{d,i} \left(\overline{\mu}_{\mathbf{h}}^{d,i} \right)^H}{\gamma_{d,i}}$ using the L_d' most significant continuous paths.
7:　　$\beta_d \leftarrow \frac{\alpha_1}{\alpha_0}$, where α_1 and α_0 can be obtained through $\mathbf{B}_d$.
8: **end for**

Cluster Partition:

9: $\mathbf{p}^{(0)} \leftarrow \text{FindIndex}(0 < \gamma_{list} < \gamma_{th})$.
10: **if** $\mathbf{p}^{(0)}$ is not empty **then**
11:　　$\gamma_{list}^{(old)} \leftarrow \gamma_{list}$ and $\gamma_{list}[\mathbf{p}^{(0)}] \leftarrow \mathbf{0}$.
12:　　$\mathbf{CLS} \leftarrow \text{Split}(\mathbf{CLS}, \mathbf{p}^{(0)})$, where $\mathbf{CLS}$ is splitted into C clusters according to $\mathbf{p}^{(0)}$.
13: **end if**
14: **for** $d = 1, 2, \ldots, C$ **do**
15:　　Update R_d, Γ_d and $\mathbf{B}_d$ in $\mathbf{CLS}$ according to $\mathbf{p}^{(0)}$, γ_{list} and β_d, respectively.
16: **end for**

Data Detection:

17: Refer to (36) and (37).

Noise Measurement:

18: Update the noise vector λ according to (38).

Channel Estimation:

19: $\Sigma_0 \leftarrow diag(\Gamma_1 \otimes \mathbf{B}_1, \Gamma_2 \otimes \mathbf{B}_2, \ldots, \Gamma_C \otimes \mathbf{B}_C)$.
20: Obtain $\mathbf{U}$ according to (11).
21: $\mathbf{p}^{(1)} \leftarrow \text{FindIndex}(\gamma_{list} \geq \gamma_{th})$ and $\mathbf{U} \leftarrow \mathbf{U}[:, \mathbf{p}^{(1)} \otimes (1:M)]$.
22: Refer to (32) and (33).

Check stopping conditions:

23: $r \leftarrow r + 1$.
24: **return** the sub-part of **Cluster Evolution** until $r \geq r_{max}$ or $\| \gamma_{list} - \gamma_{list}^{(old)} \|_2^2 < \epsilon$.

Output: the pruned channel $\overline{\mu}_{\mathbf{h}}$ with the covariance matrix $\overline{\Sigma}_{\mathbf{h}}$, the cluster list $\mathbf{CLS}$, the noise vector λ, and the transmitted data symbols $\mathbf{x}_{\mathbf{d}}^{(m)}$ for $m = 1, 2, \ldots, M$.

Based on the noise-resistant Bayesian channel estimation in the NC-CSCE algorithm, the MSE bound of the sparse vector $\mathbf{h}$ can be expressed as

$$E\left[\| \mathbf{h} - \widehat{\mathbf{h}} \|_2^2 \right] \geq Tr(\overline{\Sigma}_{\mathbf{h}}) = Tr\left[\Sigma_0 - \Sigma_0 \mathbf{U}^H \left(\Lambda + \mathbf{U} \Sigma_0 \mathbf{U}^H \right)^{-1} \mathbf{U} \Sigma_0 \right], \tag{39}$$

where $\widehat{\overline{\mathbf{h}}}$ is the estimate of $\mathbf{h}$. The MSE bound is related to the hyperparameters of the cluster distribution, data symbols and noise distribution. Only when the perfect hyperparameters are obtained is the lower bound of MSE reached. Similarly to analyzing the PB-CSBL algorithm, we assume that all L paths are in one cluster (i.e., $C = 1$) and consider the two special cases, namely the temporal correlation coefficient $\beta = 0$ and $\beta = 1$. For ease of analysis, the noise power keeps constant $\overline{\lambda}$ over all subcarriers. Furthermore, each transmitted symbol is normalized to unit power. Then, the MSE bounds of $\mathbf{h}^{(m)}$ can be evaluated by

$$E\left[\|\mathbf{h}^{(m)} - \widehat{\overline{\mathbf{h}}}^{(m)}\|_2^2\right] \geq \sum_{l=1}^{L}\left(\frac{K_p + K_d}{\overline{\lambda}} + \frac{1}{\gamma_l}\right)^{-1}, \tag{40}$$

$$E\left[\|\mathbf{h}^{(m)} - \widehat{\overline{\mathbf{h}}}^{(m)}\|_2^2\right] \geq \sum_{l=1}^{L}\left(\frac{M(K_p + K_d)}{\overline{\lambda}} + \frac{1}{\gamma_l}\right)^{-1}, \tag{41}$$

respectively. In addition to what has been discussed when analyzing the PB-CSBL algorithm, Equations (40) and (41) also prove that the NC-CSCE algorithm utilizing pilot symbols and unknown data symbols can achieve a smaller MSE compared with the PB-CSBL algorithm.

6. Evaluation and Result Analysis

In this section, to evaluate the performance of the proposed algorithms, numerical simulations and a lake trial were conducted. The parameter settings of CP-OFDM in the simulations and lake trial are shown in Table 2. The bandwidth, carrier frequency and sampling frequency of the lake trial are closely related to the transducers used in the trial, and these values are set smaller in the simulation system to increase the simulation speed. Because of the high cost of the lake trial, the CP length, the number of pilot subcarriers and the number of null subcarriers are designed to be larger than the simulation system to ensure the availability of trial data. In the simulation system (the specifications of our computer are as follows: Intel(R) Core(TM) i5-6400 CPU 2.7 GHz (4 cores), 16 GB RAM, 1 TB memory with Windows and MATLAB installed), it is assumed that the UWA channel has 12 randomly generated paths, and that these paths remain fixed within one OFDM frame. The maximum path delay is less than CP length, and the amplitudes of paths are Rayleigh distributed with the average power decreasing exponentially with delay. The Doppler scale factor $a \in (0, 10^{-3})$ is randomly chosen. Moreover, there is no channel code module in the simulation system.

Table 2. UWA CP-OFDM settings.

Parameters	Notations	Values of the Simulations	Values of the Lake Trial
Bandwidth	B	1.6 kHz	5 kHz
Carrier frequency	f_c	2.5 kHz	20 kHz
Sampling frequency	f_s	12.5 kHz	100 kHz
Number of subcarriers	K	256	256
Number of data subcarriers	K_d	218	193
Number of pilot subcarriers	K_p	14	32
Number of null subcarriers	K_n	24	31
Symbol duration without CP	T	160 ms	51.2 ms
CP length	T_{cp}	10 ms	25.6 ms
Blocks in one frame	M	4	5

6.1. Simulation Results

Firstly, the proposed PB-CSBL and NC-CSCE algorithms were compared with the LS, OMP, SOMP and TMSBL for channel estimation. All paths are distributed in three non-overlapping clusters, and these clusters are set with three different temporal correlation

coefficients ranges over multiple OFDM blocks, which are $[0.1, 0.3]$, $[0.5, 0.7]$, $[0.8, 1.0]$. The additive noise is assumed to be white Gaussian noise, and thus all subcarriers are divided into the same group (i.e., $Q = 1$) for the proposed NC-CSCE algorithm. The transmitted data bits are mapped to QPSK or 16-QAM constellation symbols. The performance metrics used in the numerical simulations are the MSE of CIRs and BER of data bits, which are verified in 500 simulations for each value of SNR.

Figure 3a,b show the MSE and BER performance of different channel estimation methods, respectively. LS performs worst, because the channel sparsity is not considered for this underdetermined system. Based on the joint estimation of multiple OFDM blocks, the SOMP makes full use of the strong temporal correlation of path delays and achieves better performance than the OMP. The delay and correlation of path gains are analyzed in the TMSBL, which makes it better than the SOMP. However, the proposed PB-CSBL performs even better, as it can learn different temporal correlation coefficients and reduce the interference among clusters. The proposed NC-CSCE algorithm utilizes all received pilot and data symbols for the channel estimation, which achieves the lowest BER close to the perfect CSI. When 16-QAM mapping is adopted, Figure 4 proves that the proposed algorithms still maintain good estimation performance. However, in the low SNR region from 0 to 10 dB, the NC-CSCE algorithm does not perform well. This is mainly because high-order modulation at low SNR is prone to symbol decision errors, which may be exacerbated in the iterative process of the NC-CSCE algorithm.
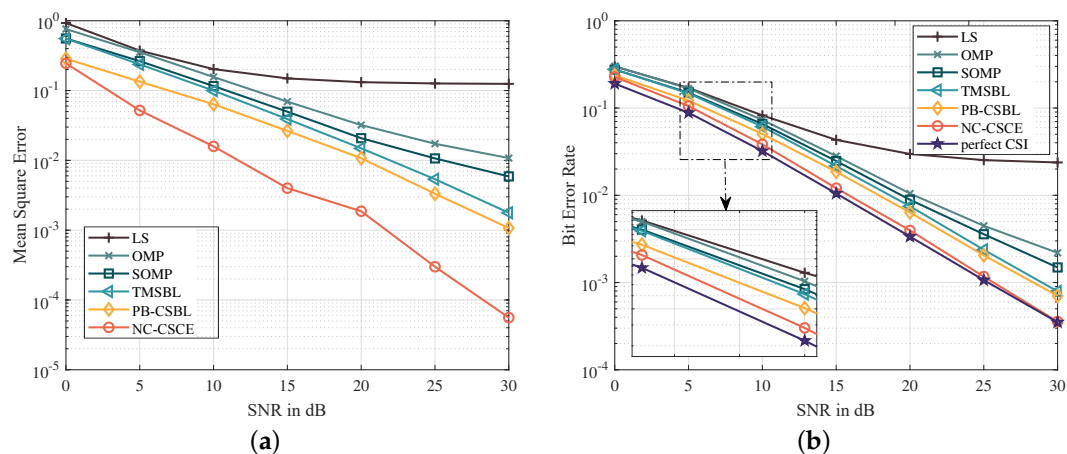

Figure 3. The comparison of MSE and BER performance based on QPSK constellation mapping: (**a**) the comparison of MSE performance; and (**b**) the comparison of BER performance.
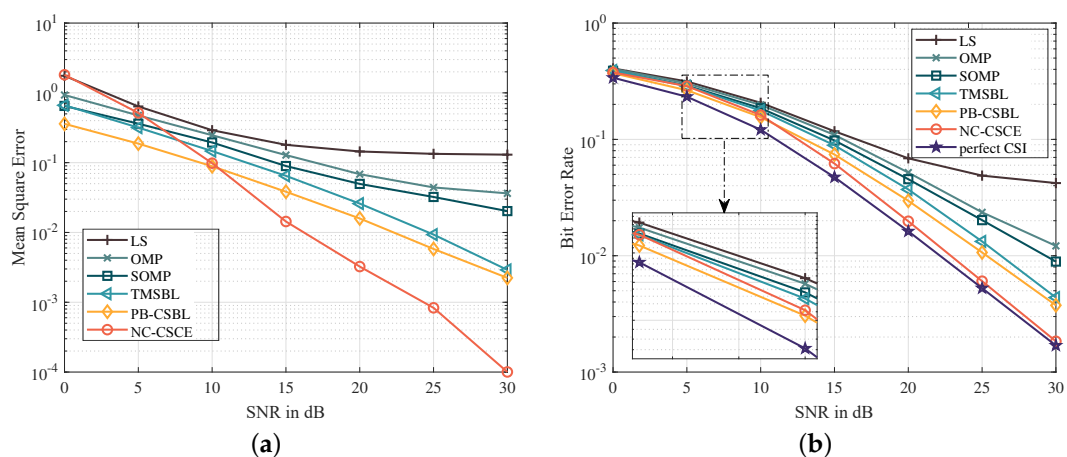

Figure 4. The comparison of MSE and BER performance based on 16-QAM constellation mapping: (**a**) the comparison of MSE performance; and (**b**) the comparison of BER performance.

Then, to verify the estimation performance of the proposed algorithms under extreme temporal correlation coefficients (i.e., $\beta = 0$ and $\beta = 1$), the proposed algorithms are compared with their theoretical lower bounds. Multipaths are assumed to be sparse and share the same correlation coefficient. Then, under QPSK mapping and white Gaussian noise, the mean square error performance is presented in Figure 5. When $\beta = 0$, the MSE of the proposed algorithms is close to the theoretical lower bound in a high SNR region. A significant gap exists in the low SNR region, especially for the NC-CSCE algorithm. This is caused by imperfect hyperparameter estimation, as the lower bound is only reached when the perfect hyperparameters are obtained. When $\beta = 1$, the gap is more obvious in the whole SNR region because of the hyperparameter estimation error. However, the MSE performance with $\beta = 1$ is better than the MSE performance with $\beta = 0$ in a low SNR region. In short, a larger temporal correlation coefficient means stronger resistance to additive noise, however, the hyperparameter estimation error, such as the error of γ, is often larger than that of small temporal correlation coefficients in the high SNR region. Figure 6a,b show the MSE of the γ estimated by the proposed PB-CSBL and NC-CSCE algorithms, respectively. The MSE is evaluated under different SNRs and temporal correlation coefficients. As can be seen, a larger temporal correlation coefficient achieves a smaller MSE in the low SNR region, while a smaller temporal correlation coefficient reduces the MSE value in the high SNR region.
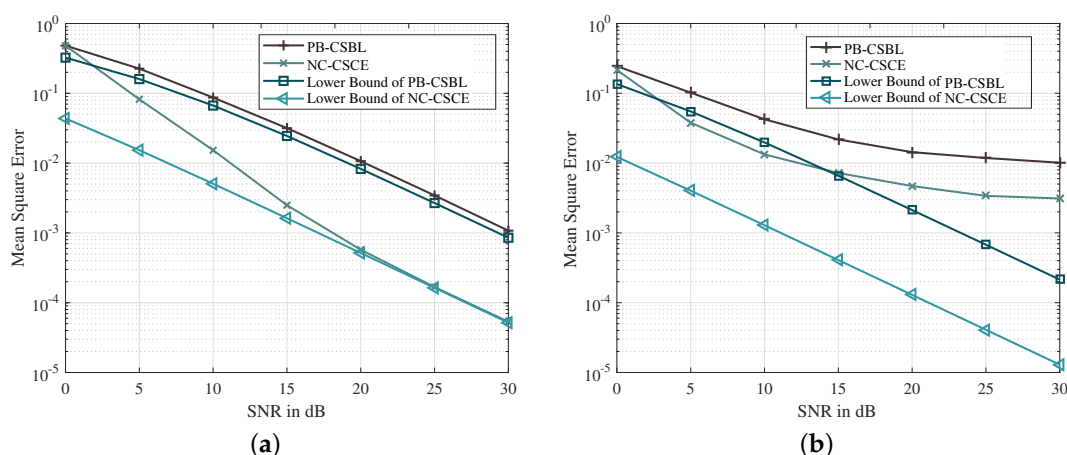

(a)   (b)

Figure 5. The MSE performance of the proposed algorithms and their theoretical lower bounds: (**a**) the temporal correlation coefficient $\beta = 0$; and (**b**) the temporal correlation coefficient $\beta = 1$.

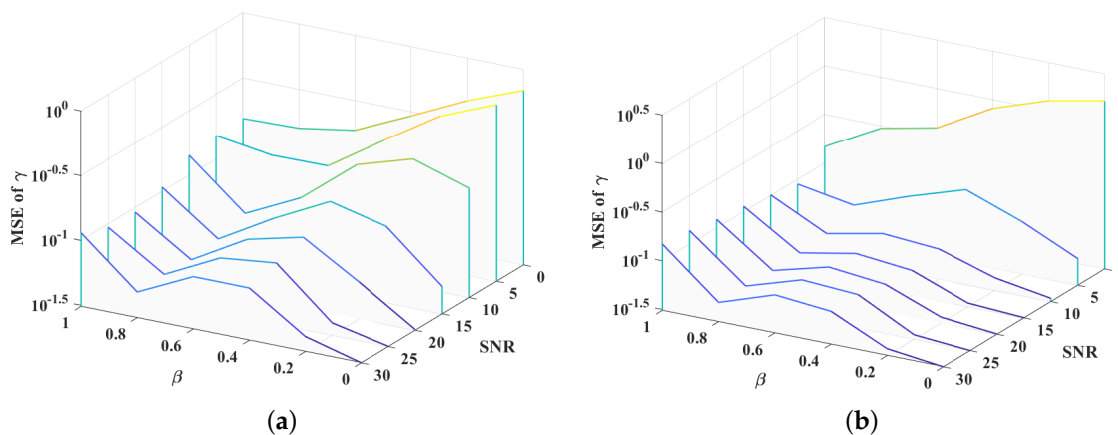(a)   (b)

Figure 6. The MSE of the estimated hyperparameter γ by the proposed PB-CSBL and NC-CSCE algorithms under different SNRs and temporal correlation coefficients: (**a**) the PB-CSBL algorithm; and (**b**) the NC-CSCE algorithm.

Finally, to evaluate the estimation performance of the proposed NC-CSCE algorithm under colored noise, we compared the NC-CSCE algorithm with the TMSBL and PB-

CSBL algorithms. The group number of subcarriers in NC-CSCE is adjusted to evaluate the channel estimation performance. In the simulation, all paths are distributed in three non-overlapping clusters. The clusters are set with three different temporal correlation coefficients ranges over multiple OFDM blocks, which are $[0.1, 0.3]$, $[0.5, 0.7]$, $[0.8, 1.0]$. All subcarriers are divided into eight groups and each group is set with a different SNR. As shown in Figure 7b, the reference SNR ranges are from 5 to 20 dB. With QPSK mapping, the MSE and BER performance curves are shown in Figure 7a, and the results prove that the proposed NC-CSCE algorithm is superior to the TMSBL and PB-CSBL algorithms. The NC-CSCE-8 denotes the NC-CSCE algorithm with eight groups, and it performs better than the NC-CSCE algorithm with other groups. Therefore, the setting of group number is critical for the NC-CSCE algorithm. In one instance, the actual distribution of colored noise can hardly be reflected with a limited group number; yet in another, a small symbol number will result in low measurement accuracy in one group. Figure 7b also shows the statistical results of the estimated SNRs based on 500 simulations. Similarly, NC-CSCE-8 matches the reference best among all compared methods.
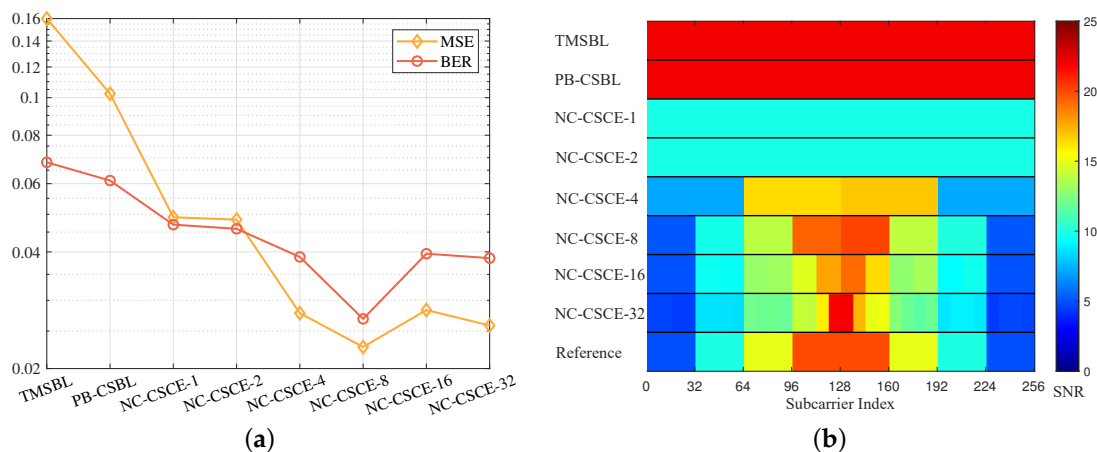

Figure 7. The performance comparison of different algorithms under colored noise: (**a**) the MSE and BER performance; and (**b**) the estimated SNRs over all subcarriers.

6.2. Lake Experimental Results

To further verify the proposed algorithms, a real lake experiment was performed at Huating Lake, Anhui Province, China, in May 2018. Huating Lake covers an area of about 70 square kilometers, with an average water depth of about 17 m. We hung the transmitter and receiver on the side of a passenger ship and a speedboat, respectively. The passenger ship has a semi-open roof and is about 10 m long, while the speedboat is about half the size of the ship. During the experiment, the transmitter depth was about 10 m and the receiver depth was about 13 m. Affected by wind and current, their relative speed was about 0.5 m/s, and the distance between them was more than 4.5 km. The parameter settings of the experiment are in Table 2. In addition, the linear frequency modulation (LFM) signal was added at the beginning and end of each frame for time synchronization and Doppler estimation. The transmitted data bits were encoded by a 1/2 rate convolutional code, and the encoded bits were mapped by QPSK constellation. As showed in Figure 1, using the proposed NC-CSCE algorithm, we obtained the estimated CIRs for a 14 frame history with five consecutive OFDM blocks in one frame. The CIRs are estimated after Doppler compensation, and the total delay spread is around 15 ms. From the estimated CIRs, we notice that the paths are distributed in multiple clusters, and the channel shows a clustered-sparse structure.

Figure 8 shows the BER performance after convolutional decoding. The BERs of the proposed two algorithms are still the lowest for most frames, followed by the TMSBL and the SOMP, while the OMP has the worst BER performance. Obviously, the performance gap

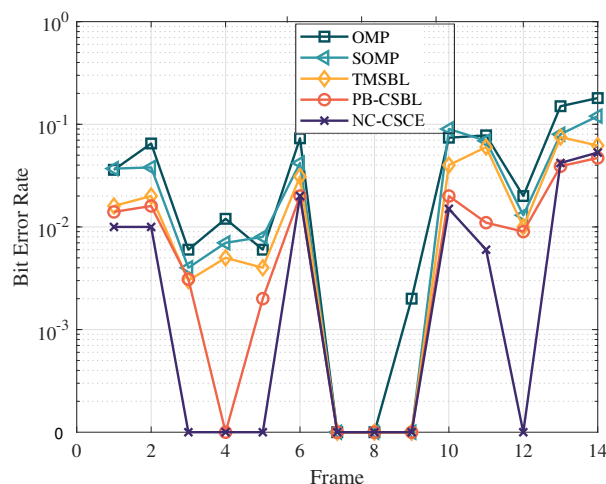between proposed algorithms and other methods further verify the superior performance of the proposed algorithms.

Figure 8. The comparison of BER performance in the lake experiment.

7. Conclusions

In this paper, we propose a novel Bayesian learning-based channel estimation architecture to estimate the time-varying multipath channel with colored noise for UWA-OFDM systems. Specifically, a clustered-sparse channel distribution model was constructed to characterize the multipath distribution, and a noise-resistant channel measurement model is constructed to reduce the noise disturbance. To obtain the clustered-sparse distribution, we propose the partition-based clustered-sparse Bayesian learning algorithm. To lessen the effect of colored noise, we proposed a noise-corrected clustered-sparse channel estimation algorithm to improve the estimation performance. Taking advantage of the iterative clustered-sparse distribution learning, symbol decision and noise correction, the accuracy of Bayesian channel estimation can be improved. Experiments proved the effectiveness of the proposed algorithms for channel estimation.

In future work, channel decoding can be incorporated into the Bayesian learning-based channel estimation architecture to explore the possibility of reducing symbol decision errors. In the noise-resistant channel measurement model, the channel gains are approximately constant within one OFDM block, which may be too optimistic in some rapidly time-varying UWA channels. Therefore, some channel estimation models that deal with doubly selective fading channels, such as the popular basis expansion model, can be combined with the clustered-sparse Bayesian learning to improve the estimation performance. At the same time, channel feedback and reconstruction will be analyzed to realize adaptive modulation and coding in UWA-OFDM communication.

Author Contributions: Conceptualization, S.W. and D.L.; data curation, S.W.; formal analysis, S.W. and M.L.; funding acquisition, D.L.; investigation, S.W.; methodology, S.W. and M.L.; project administration, D.L.; resources, D.L.; software, S.W.; validation, S.W. and D.L.; visualization, S.W.; writing—original draft, S.W.; writing—review and editing, M.L. and D.L. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Climent, S.; Sanchez, A.; Capella, J.; Meratnia, N.; Serrano Martín, J. Underwater Acoustic Wireless Sensor Networks: Advances and Future Trends in Physical, MAC and Routing Layers. *Sensors* **2014**, *14*, 795–833. [CrossRef]
2. Stojanovic, M.; Preisig, J. Underwater acoustic communication channels: Propagation models and statistical characterization. *IEEE Commun. Mag.* **2009**, *47*, 84–89. [CrossRef]
3. Coatelan, S.; Glavieux, A. Design and test of a multicarrier transmission system on the shallow water acoustic channel. In Proceedings of the OCEANS'94, Brest, France, 13–16 September 1994 ; Volume 3, pp. III/472–III/477. [CrossRef]
4. Zhou, S.; Wang, Z. *OFDM for Underwater Acoustic Communications*, 1st ed.; Wiley Publishing: Hoboken, NJ, USA, 2014.
5. Wen, M.; Cheng, X.; Yang, L.; Li, Y.; Cheng, X.; Ji, F. Index modulated OFDM for underwater acoustic communications. *IEEE Commun. Mag.* **2016**, *54*, 132–137. [CrossRef]
6. Basar, E.; Aygolu, U.; Panayirci, E.; Poor, H.V. Orthogonal Frequency Division Multiplexing With Index Modulation. *IEEE Trans. Signal Process.* **2013**, *61*, 5536–5549. [CrossRef]
7. Wan, L.; Zhou, H.; Xu, X.; Huang, Y.; Zhou, S.; Shi, Z.; Cui, J. Adaptive Modulation and Coding for Underwater Acoustic OFDM. *IEEE J. Ocean. Eng.* **2015**, *40*, 327–336. [CrossRef]
8. Zhang, Y.; Huang, Y.; Wan, L.; Zhou, H.; Zhou, S.; Shen, X.; Wang, H. Adaptive OFDMA for downlink underwater acoustic communications. In Proceedings of the 2014 Oceans-St. John's, St. John's, NL, Canada, 14–19 September 2014; pp. 1–5. [CrossRef]
9. Badiey, M.; Yongke, M.; Simmen, J.A.; Forsythe, S.E. Signal variability in shallow-water sound channels. *IEEE J. Ocean. Eng.* **2000**, *25*, 492–500. [CrossRef]
10. Preisig, J.C.; Deane, G.B. Surface wave focusing and acoustic communications in the surf zone. *J. Acoust. Soc. Am.* **2004**, *116*, 2067–2080. [CrossRef]
11. Song, A.; Mohsen, B.; Song, H.C.; Hodgkiss, W.S.; Porter, M.B. Impact of ocean variability on coherent underwater acoustic communications during the Kauai experiment (KauaiEx). *J. Acoust. Soc. Am.* **2008**, *123*, 856–865 [CrossRef]
12. Rouseff, D.; Badiey, M.; Song, A. Effect of reflected and refracted signals on coherent underwater acoustic communication: Results from the Kauai experiment (KauaiEx 2003). *J. Acoust. Soc. Am.* **2009**, *126*, 2359–2366. [CrossRef]
13. Song, A.; Badiey, M.; Newhall, A.E.; Lynch, J.F.; Deferrari, H.A.; Katsnelson, B.G. Passive Time Reversal Acoustic Communications Through Shallow-Water Internal Waves. *IEEE J. Ocean. Eng.* **2010**, *35*, 756–765. [CrossRef]
14. Wang, Z.; Zhou, S.; Preisig, J.C.; Pattipati, K.R.; Willett, P. Clustered Adaptation for Estimation of Time-Varying Underwater Acoustic Channels. *IEEE Trans. Signal Process.* **2012**, *60*, 3079–3091. [CrossRef]
15. Qiao, G.; Song, Q.; Ma, L.; Sun, Z.; Zhang, J. Channel prediction based temporal multiple sparse bayesian learning for channel estimation in fast time-varying underwater acoustic OFDM communications. *Signal Process.* **2020**, *175*, 107668. [CrossRef]
16. Wang, S.; Li, D.; Liu, M.; Huang, W.; Chen, H.; Cen, Y. Clustered-Sparse Bayesian Learning for Channel Estimation in Underwater Acoustic OFDM Systems. In Proceedings of the 2020 International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China, 21–23 October 2020; pp. 546–551. [CrossRef]
17. Enguix, I.F.; Egea, M.S.; González, A.G.; Arenas, D. Acoustic Characterization of Impulsive Underwater Noise Present in Port Facilities: Practical Case in the Port of Cartagena. *Proceedings* **2018**, *4*, 57. [CrossRef]
18. Li, W.; Preisig, J.C. Estimation of Rapidly Time-Varying Sparse Channels. *IEEE J. Ocean. Eng.* **2007**, *32*, 927–939. [CrossRef]
19. Stojanovic, M. OFDM for underwater acoustic communications: Adaptive synchronization and sparse channel estimation. In Proceedings of the 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, Las Vegas, NV, USA, 31 March–4 April 2008; pp. 5288–5291.
20. Geng, X.; Zielinski, A. An eigenpath underwater acoustic communication channel model. In Proceedings of the 'Challenges of Our Changing Global Environment', OCEANS '95 MTS/IEEE, San Diego, CA, USA, 9–12 October 1995.
21. Yang, T.C. Characteristics of underwater acoustic communication channels in shallow water. In Proceedings of the OCEANS 2011 IEEE–Spain, Santander, Spain, 6–9 June 2011.
22. Rani, M.; Dhok, S.B.; Deshmukh, R.B. A Systematic Review of Compressive Sensing: Concepts, Implementations and Applications. *IEEE Access* **2018**, *6*, 4875–4894. [CrossRef]
23. Khan, M.R.; Das, B.; Pati, B.B. Channel estimation strategies for underwater acoustic (UWA) communication: An overview. *J. Frankl. Inst.* **2020**, *357*, 7229–7265. [CrossRef]
24. Uwaechia, A.N.; Mahyuddin, N.M. A Review on Sparse Channel Estimation in OFDM System Using Compressed Sensing. *IETE Tech. Rev.* **2017**, *34*, 514–531. [CrossRef]
25. Berger, C.R.; Zhou, S.; Preisig, J.C.; Willett, P. Sparse Channel Estimation for Multicarrier Underwater Acoustic Communication: From Subspace Methods to Compressed Sensing. *IEEE Trans. Signal Process.* **2010**, *58*, 1708–1721. [CrossRef]
26. Needell, D.; Tropp, J.A. CoSaMP Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 301–321. [CrossRef]
27. Zhou, Y.H.; Tong, F.; Zhang, G.Q. Distributed compressed sensing estimation of underwater acoustic OFDM channel. *Appl. Acoust.* **2017**, *117*, 160–166. [CrossRef]
28. Yin, Y.; Liu, S.; Qiao, G.; Yang, Y. OFDM Demodulation Using Virtual Time Reversal Processing in Underwater Acoustic Communications. *J. Comput. Acoust.* **2015**, *23*, 1540011. [CrossRef]
29. Wipf, D.P.; Rao, B.D. Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.* **2004**, *52*, 2153–2164. [CrossRef]

30. Prasad, R.; Murthy, C.R. Bayesian Learning for Joint Sparse OFDM Channel Estimation and Data Detection. In Proceedings of the Global Telecommunications Conference (GLOBECOM 2010), Miami, FL, USA, 6–10 December 2010.

31. Prasad, R.; Murthy, C.R.; Rao, B.D. *Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems Using Sparse Bayesian Learning*; IEEE Press: Piscataway, NJ, USA, 2014.

32. Qiao, G.; Song, Q.; Ma, L.; Liu, S.; Sun, Z.; Gan, S. Sparse Bayesian Learning For Channel Estimation In Time-varying Underwater Acoustic OFDM Communication. *IEEE Access* **2018**, 56675–56684. [CrossRef]

33. Bajwa, W.U.; Haupt, J.; Sayeed, A.M.; Nowak, R. Compressed Channel Sensing: A New Approach to Estimating Sparse Multipath Channels. *Proc. IEEE* **2010**, *98*, 1058–1076. [CrossRef]

34. Liu, S.; Wu, H.; Huang, Y.; Yang, Y.; Jia, J. Accelerated Structure-Aware Sparse Bayesian Learning for Three-Dimensional Electrical Impedance Tomography. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5033–5041. [CrossRef]

35. Ye, H.; Li, G.Y.; Juang, B. Power of Deep Learning for Channel Estimation and Signal Detection in OFDM Systems. *IEEE Wirel. Commun. Lett.* **2018**, *7*, 114–117. [CrossRef]

36. Jiang, R.; Wang, X.; Cao, S.; Zhao, J.; Li, X. Deep Neural Networks for Channel Estimation in Underwater Acoustic OFDM Systems. *IEEE Access* **2019**, *7*, 23579–23594. [CrossRef]

37. Zhang, Y.; Li, J.; Zakharov, Y.; Li, X.; Li, J. Deep learning based underwater acoustic OFDM communications. *Appl. Acoust.* **2019**, *154*, 53–58. [CrossRef]

38. Stojanovic, M. Low Complexity OFDM Detector for Underwater Acoustic Channels. In Proceedings of the OCEANS 2006, Boston, MA, USA, 18–21 September 2006; pp. 1–6.

39. Li, B.; Zhou, S.; Stojanovic, M.; Freitag, L.; Willett, P. Multicarrier Communication Over Underwater Acoustic Channels With Nonuniform Doppler Shifts. *IEEE J. Ocean. Eng.* **2008**, *33*, 198–209.

40. Kim, B.-C.; Lu, I.T. Parameter study of OFDM underwater communications system. In Proceedings of the OCEANS 2000 MTS/IEEE Conference and Exhibition, (Cat. No.00CH37158), Providence, RI, USA, 11–14 September 2000; Volume 2, pp. 1251–1255. [CrossRef]

41. Zhang, Z.; Rao, B.D. Sparse Signal Recovery with Temporally Correlated Source Vectors Using Sparse Bayesian Learning. *IEEE J. Sel. Top. Signal Process.* **2011**, *5*, 912–926. [CrossRef]