Oral presentation draft

**Begin：**

Good morning, everyone. I’m glad to make this presentation. My name is Shougui Cai, from Zhejiang University, China, and the co-author is Professor Wen Xu. The title of our paper is “Matched-field source localization using sparsely-coded neural network and data-model mixed training”.

Source localization is a basic problem in underwater acoustics. Many solving approaches have been developed, and the matched-field processing (MFP) is one of the mostly-studied. However, MFP is sensitive to the mismatch problem and performs well only when the knowledge of ocean environment is accurate.

Machine learning learns directly from the observation and can be designed to learn a generic model suitable for different scenarios. In our paper, source localization is viewed as a machine learning problem and a matched-field source localization model is learned by training a sparsely-coded feed-forward neural network (SCFNN) with mixed environment models and data.

**Contents：**

My presentation will go on in this order. First, i’ll introduce how to establish an SCFNN for source localization prediction and how to learn the model parameter from data. Then, i’ll train and test a prediction model on the experimental data, and confirm that the SCFNN works well on source localization. And in part III, the inﬂuence of SSP mismatch on the SCFNN classifier is investigated by simulations, together with performance comparison with conventional MFP, in part IV, data-model mixed training strategy is used to improve the model robustness, the end is a summary.

**SCFNN model：**

Same as Niu did in his paper (JASA 2017), we also assumed that there is a deterministic relationship between source range and sample-covariance matrix (SCM) and approximated this relationship by the FNN.

The feed-forward neural network (FNN), also known as multi-layer perceptron, is constructed using a feedforward directed acyclic architecture. The outputs are formed through a series of functional transformations of the weighted inputs. In the FNN, the outputs are deterministic functions of the inputs.

Here, a three layer model (input layer L1, hidden layer L2 and output layer L3) is used to construct the FNN. The input layer L1 is comprised of D input variables x. The output y is discrete corresponding to the estimated source range.

Only the real and imaginary parts of the complex valued entries of diagonal and upper triangular matrix in C(f) are used as input

To make the processing independent of the complex source spectra, the received array pressure is transformed to a normalized sample covariance matrix.

The output of FNN is the source range distribution prediction.

**Network structure:**

The FNN used in our paper is a simple three layer model (input layer L1, hidden layer L2 and output layer L3).

The input layer L1 is comprised of D input variables and the hidden layer L2 is a linear combination with input layer and then transformed using an activation function f(·). The [ReLu (Rectified Linear Units)](https://www.baidu.com/link?url=uhTsmz7w6DRma97E_HW-dvsWjzC3tMeQApEEewo9StEWFDDSVEYzLrgaG9b_BaQCq5ejxlDAZiB-BVMg6H5WcK&wd=&eqid=828f381700012dc60000000459f97789) function was chosen as the intermediate activation function for this study, see fig(b).

Similarly, Neurons between the hidden layer and the output layer are mapped by a linear function,

then the output of the model is normalized by *softmax* function, which is a common choice for multi-class classification task.

**Input data preprocessing：**

The complex pressure at frequency f obtained by taking the DFT of the input pressure data at L sensors is denoted by p(f) = [p1(f), · · · , pL(f)]T . The sound pressure is modeled as p(f) = S(f)g(f, r) + n, where n is the noise, S(f) is the source term, and g is the Green’s function. To reduce the effect of the source amplitude |S(f)|, this complex pressure is normalized. The normalized sample covariance matrices (SCMs) are averaged over Ns snapshots to form the conjugate symmetric matrix.

Finally, the matrix C(f) are vectorized to form the real-valued input x of size L × L.

Preprocessing the data according to there equations ensures that the Green’s function is used for localization.

**Source range mapping：**

In the classification problem, a set of source ranges is discretized into K bins, r1, ..., rK, of equal width ∆r.

**Training criterion：**

The weight matrix W and V are the parameters to be learned. Obviously, a learning criterion is needed. In our case, the parametric model defines a distribution of target location, and we can simply use the principle of maximum likelihood to determine the parameters in this model.

The first term in *E(w)* is the cross-entropy, equivalently the negative log-likelihood, between the true/desired distribution and the model predict distribution.

The second and third term are the sparsity constraints on neural networks, In practical applications, we not only want the network to be sparse, but also want the model parameters to be sparse,  
the latter making the model more interpretable. In this paper, we use L1-norm to promote sparse neurons activations, and constrain the L2-norm of each row of the weight matrix V to prevent any one hidden unit from having very large weights. Thus, the neural network is sparsely coded, we named it as SCFNN.

As the maximum likelihood criterion is consistent, the model is capable of representing the training data distribution.

By minimizing this criteria, we can learn the model weights from training data and get a Source localization prediction model.

**Definition of model accuracy：**

The model accuracy is defined the percentage of accurately predicted samples.

In this equation, rk, r = 1,…,K are the source range classes and tn is the label of data xn. The upper case N is the number of test numbers.

**Simulation environment：**

The proposed SCFNN is trained and tested on the widely studied SWell96-Ex test, conducted in a shallow water waveguide environment with depth of 216 m.

During the experiment, two moving sound sources are deployed in field, including a deep source (J-15) and a shallow source (J-13). In all of our discussions, the shallow sound source is used, which was towed about 9 m in depth and emitted with 9 frequencies between 109 Hz and 385 Hz. The number of vertical array elements is 21, and other specifc deployment parameters are shown in Fig. (a).

As Fig.(b) shows, the source ship (R/V Sproul) started its track south of all of the arrays and proceeded northward at a speed of 5 knots (2.5 m/s).

**Performance comparison：**

In this part, the proposed model is tested on SWellEx-96 experimental data, and compared with two methods denoted as Bartlett and MCE, Bartlett use the measured pressure to match with a replica field computed by a numerical propagation model, while, MCE matches the covariance. Note that, there are two kinds of replica-field used in the Bartlett processor, one is simulated by Kraken (noted as Bartlett 2), the other is the measurement data (noted as Bartlett 1), same as the training data used in SCFNN.

As we can see, whether using single frequency or multi-frequencies, the accuracy of SCFNN is always better than the Bartlett, and not worse than direct data match (noted as MCE),.

This is more obvious when it comes to the comparison of absolute mean error. It can be said that, the learned SCFNN works well on source localization and performs better than the Bartlett processor.

**SSP mismatch on performance：**

In the MFP method, the model accuracy is heavily aﬀected by the mismatch problem. Here two diﬀerent degrees of error (a large one and a light one) in the knowledge of the sound-speed profile are chosen to investigate how the SSP mismatch influence the model performance.

The optimized one is the best SSP model for real environment of SWellEx-96 experiment, while, i906 has significant change in shape from the optimized, while the change in the i905 is slight. The i906\* is slightly changed from i906, for the sake of testing.

The performance curves for SCFNN, Bartlett and MCE vs SNR are plotted by 1000 time Monte Carlo simulations. Results show that when the change in SSP is relatively slight, SCFNN positions best, followed by MCE and Bartlett worst; In the Fig., the legend ‘FNN, i905’ means that, the corresponding method is FNN and the test data is from i905 environment, rests are similar. The snapshot number here is 10.

When the change is relatively large (with shape varying), the accuracy order is unchanged, but the absolute mean error of SCFNN becomes larger than MCE. This is maybe caused by the noisy training data.

In conclusion, SCFNN is also sensitive to SSP mismatch, but still performs better than Bartlett and the performance of SCFNN is close to the MCE method.

**Data-model mixed training：**

Another strategy used in our paper is training the neural network with mixed data-model, in order to increase the processing robustness. As neural networks are strong enough to learn regular pattern over a set of changing scenarios, when training the network, we can use the examples gathered from diﬀerent mismatch scenarios to make the network be robust to mismatch. In our case, the received acoustic pressure computed by various environment models, are combined with the actual data as the training set.

As mentioned above, the SCFNN is also sensitive to SSP mismatch. When the environment SSP has a big change in shape, the classifier trained by single data set performs poorly. For example, the model, trained by data set corresponding to ssp-optimized, performs poorly on ssp-i906\*, and the accuracy drops about 40%, compared with the performance on ssp-i905.

In this section, by combining the data collected from ssp-i906 and ssp-optimized as training set, the robustness of the classifier increases significantly; as Fig. shows, the re-trained classifier predicts accurately on ssp-i906\*, just as well as on ssp-i905. Although the accuracy for i905 has a little glissade compared with data training only case, the performance for i906 is improved.

By mixed data-model training, the SCFNN classifier works well on two entirely diﬀerent SSPs. Note that, in Fig. 7, the legend ‘i905, combined’ means the model is trained by mixed data, and then tested on ssp-i905. The rest legends are similar.

**Summary & Future work：**

In our paper, we propose a method that can help reduce the mismatch problem in matched-field source localization, by using a sparsely-coded feed-forward neural network (SCFNN), combined with data-model mixed training. The proposal is examined on SWellEx-96 experiment. To be specific, we firstly train and test a prediction model on the experimental data, and confirm that the SCFNN works well on source localization. Then, the inﬂuence of SSP mismatch on the SCFNN classifier is investigated by simulations. Finally, we train the SCFNN with mixed environment model data. It can be seen that the model robustness is significantly improved and the trained classifier performs well on varying SSP environments.

The discussions on applying machine learning methods to overcome mismatch problem in underwater source localization are preliminary and only a fine-tuned FNN is used in our paper. Machine learning has potential advantages in unstable underwater acoustic environment and thus deserves more eﬀorts.

Our future work may include the following orientations.

Firstly, as the localization error at low SNR is still huge, we want to integrate adversarial learning into our model, and make the characteristic parameters can be enhanced, when the noise is high.

Second, we will do some mathematical analysis and try to explain how does the model robustness been explained.

Thank you for your listening.