A New Statistical Method to Discrimination between Water Waves and Reflected and/or Refracted Waves in OBS data

[1]Tetsuo TAKANAMI and [2]Genshiro KITAGAWA

[1]Hokkaido University, Institute of Seismology and Volcanology, Kita-ku, N10,W8, Sapporo, 060-0810 Japan
[2]The Institute of Statistical Mathematics, 4-6-7 Minami-Ababu, Minatok-ku, Tokyo 106-8569

In order to explore velocity structure in the crust based on OBS (Ocean Bottom Seismogram), it is necessary to detect seismic signals reflected and/or refracted waves in the crust from the OBS data contaminated with relatively large direct water wave and its multiples. In this paper, we propose the new methods based on the time series and spatial-temporal decompositions of the data.

1. Extraction by Time Series Modeling

Based on the time series modeling (Kitagawa and Takanami, 1985; Takanami, 1991; Kitagawa and Higuchi, 1998), we consider the model

$$y_n = r_n + s_n + \varepsilon_n \tag{1}$$

where, r_n, s_n and ε_n represent the a direct wave and/or its multiples traveled in the water, the reflected and/or refracted waves generated in the submarine underground structure and the observation noise, respectively. To separate these three time series, it is assumed that both r_n and s_n are expressed by the autoregressive (AR) models

$$r_n = \sum_{i=1}^{m} a_i r_{n-i} + u_n\ , \qquad s_n = \sum_{i=1}^{l} b_i s_{n-i} + v_n\ , \tag{2}$$

where, the AR orders m, ℓ and the AR coefficients a_i and b_i are unknown and u_n, v_n and ε_n are white noise sequence with $u_n \sim N(0,\tau_1^2)$, $v_n \sim N(0,\tau_2^2)$ and $\varepsilon_n \sim N(0,\sigma^2)$, respectively.

The models in (1) and (2) can be joined together in the state space model form

$$x_n = Fx_{n-1} + Gw_n, \qquad y_n = Hx_n + \varepsilon_n \tag{3}$$

with the state vector defined by $x_n = (r_n,\cdots,r_{n-m+1},s_n,\cdots,s_{n-\ell+1})^T$.

If all the parameters m, ℓ and $\theta = (a_i, b_i, \tau_1^2, \tau_2^2, \sigma^2)^T$ are given, the state vector x_n can be estimated by the Kalman filter and the fixed interval smoother (Anderson and Moore, 1979). Although they are really unknown, they can be estimated by maximizing the log-likelihood of the model defined by

$$\ell(\theta_m) = -\frac{N}{2}\log 2\pi - \frac{1}{2}\sum_{n=1}^{N}\log r_n - \frac{1}{2}\sum_{n=1}^{N}\frac{\varepsilon_n^2}{r_n} \tag{4}$$

where $\varepsilon_n = y_n - Hx_{n|n-1}$ and $r_n = HV_{n|n-1}H^T + \sigma^2$ with $x_{n|n-1}$ and $V_{n|n-1}$ being the mean and the variance covariance matrix of the one-step-ahead predictor of the state obtained by the Kalman filter (Jones, 1980). The variance of AR model for the water, reflected and/or refracted waves, τ_1^2 and τ_2^2 are related to the amplitude of the waves and are actually time varying. To put it

concretely, the variance is almost zero before the water, reflected and/or refracted waves arrives, becomes large depending on the amplitude of the wave and then goes back to zero as signal dies out. These variance parameters play the role of signal to noise rations, and the estimation of these parameters is the key to the success of the time series decomposition.
In a self-organizing state space model (Kitagawa, 1998), the original state vector x_n is augmented with the time-varying parameters as

$$z_n = [x_n^T, \log_{10} \tau_{1n}^2, \log_{10} \tau_{2n}^2]. \tag{5}$$

We assume that the parameter changes according to the random walk model

$$\log_{10} \tau_{j,n}^2 = \log_{10} \tau_{j,n-1}^2 + \eta_{j,n}, \quad j = 1,2 \tag{6}$$

where $\eta_{j,n}$ is the Gaussian white noise with $\eta_{j,n} \sim N(0, \xi_j^2)$. The state space model for this extended state can be expressed in nonlinear state space model form. Then by applying the Monte Carlo filter / smoother (Kitagawa, 1996), we can estimate the state z_n from the observations. Since the extended state z_n contains the original state x_n and the parameters, this means that the marginal posterior distributions of the state and the parameters can be obtained simultaneously and automatically.
We applied this single channel time series decomposition to the entire 1560 channels. As far as we investigate the results channel by channel, it looks very reasonable. However, if we plot the expected series on one figure, the on-set times of the series are not so smoothly changing, suggesting the possibility of improving the estimates by incorporating the spatial structure of the signal. In this paper, we consider a method of incorporating spatial-temporal structure.

2. Decomposition by Spatial-Temporal Modeling

The real time series observed at OBS contains signals of a direct and multiples water waves, reflected, refracted waves and observation noise (Kuwano, 2000). Figure 1 shows the raw data recorded by OBS-4 as an example.

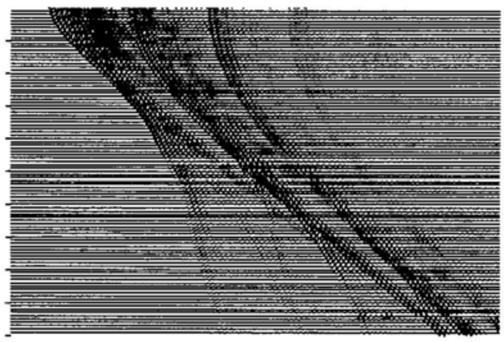

Fig.1. The 1560 air-gun traces recorded by OBS-4

Just beneath the air-gun, the direct wave (1.48km/sec) that travels through the water arrives first and dominates in the time series. However, since the velocity of the water in the solid structure is larger than it, in our present case, the reflected and the refracted waves arrive before the water waves for the offset distance larger than approximately 1.4 km and 14 km, respectively. At each OBS, 1560 time series were observed, with the location of the explosion shifted by 200m. Therefore the consecutive two series are cross-correlated and by using the structure assumed by Kuwano (2000), it is expected that we can detect the information that was difficult to obtain from a single time series. The moveout, namely the difference of the arrival times between two consecutive time series, computed for each wave type and for some offset distance, D.

Table 1. Wave type and moveout of arrival times for various offset distance.

Wave type	Offset Distance (km)				
	0	5	10	15	20
wave(0)	0.6	16.5	16.5	16.7	16.7
wave(000)	0.2	14.5	14.5	15.8	16.0
wave(00000)	0.1	8.0	12.0	14.2	15.0
wave(01)		10.5	10.0	10.1	9.9
wave(0121)	0.2	7.5	7.1	7.2	7.1
wave(012321)	0.1	3.8	3.6	3.6	3.5

The computed moveout of the waves that travel on the surface between two layers are constants independent on the offset distance D. Meanwhile, for the water waves that travel in the water, the amount of the moveout-time gradually increase with the increase of the offset distance D, and converges to approximately 17 for distance $D > 5$ km. This points out that the arrival time is approximately a linear function of the distance, D > 5km. Taking into account of this time-lag structure, and temporally ignoring the time series structure, we consider the following spatial model:

$$s_{n,j} = s_{n-k,j-1} + v_{n,j}, \qquad y_{n,j} = s_{n,j} + w_{n,j} \tag{7}$$

where k is the moveout of the water wave or reflected and refracted wave, namely the difference of the arrival times between channels $j-1$ and j. By defining the state vector by $x_{n,j} = [s_{n,j}, s_{n-k,j-1}]^T$, we obtain the state space representation, $x_{n,j} = Fx_{n-k,j-1} + Gv_{n,j}$, $y_{n,j} = Hx_{n,j} + w_{n,j}$.
Therefore, if the moveout k is given, we can easily obtain estimates of the "signal" (water, reflected, and/or refrated waves) by the Kalman filter and smoother. If one value of k dominates in one region, we can estimate it by maximizing the localized log-likelihood. However, in actual data, several different waves may appear in the same time and the same channel. To cope with this situation, we consider a mixture-log model defined by

$$s_{n,j} = \sum_{k=1}^{K} a_{n,j,k}\hat{s}_{n,j,k}, \qquad y_{n,j} = s_{n,j} + w_{n,j}, \tag{8}$$

where $\hat{s}_{n,j,k}$ is the one step ahead predictor of the $s_{n,j}$ defined by $\hat{s}_{n,j,k} = s_{n-k,j-1}$ and $\alpha_{n,j,k}$ is the mixture weight at time n and channel j. In the recursive filtering, this mixture weight can be up dated by

$$\alpha_{n,j,k} \propto \alpha_{n-k,j-1,k} \exp\{-(y_{n,j} - \hat{s}_{n,j,k})^2 / 2r_{n|n-1}\}. \tag{9}$$

Spatial-Temporal Model :

$$y_{n,j} = r_{n,j} + s_{n,j} + \varepsilon_{n,j} \tag{10}$$

where $r_{n,j}$, $s_{n,j}$, and $\varepsilon_{n,j}$ denote a direct water wave and its multiples, reflected / refracted wave and the observation noise component in channel j. Assuming that they are assumed to follow the AR models,

$$r_{n,j} = \sum_{i=1}^{m} a_{i,j} r_{n-i,j} + v_{n,j}^{r}, \qquad s_{n,j} = \sum_{i=1}^{\ell} b_{i,j} s_{n-i,j} + v_{n,j}^{s}, \tag{11}$$

respectively. On the other hand, by considering the delay structure discussed in the previous subsection, we also use the following spatial models,

$$r_{n,j} = r_{n-k,j-1} + u_{n,j}^{r}, \qquad s_{n,j} = s_{n-h,j-1} + v_{n,j}^{s}. \tag{12}$$

Here the moveout k and h are actually function of the wave type and the distance D (or equivalently the channel j).

By an approximate estimation algorithm combining the filtering and smoothing algorithms in time and space (channel), we can obtained the decomposition of the data shown in Figure 1. Figure 2 shows the results of the decomposition. The left plot shows the extracted direct and its multiple water wave. The right plot shows the extracted reflection and the refraction waves. Several waves are enhanced by this decomposition.

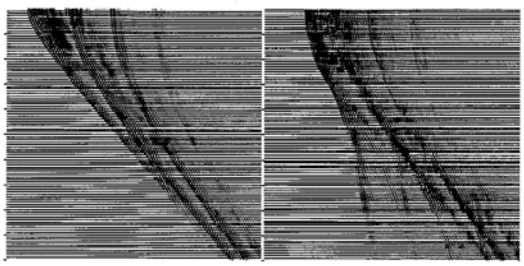

Fig.2. Extracted water waves (left) and reflected and/or refracted waves (right). Air gun 972-1071 traces, Sampling rate = 1/125 second, data length n=2000.

Conclusion

Time series model and spatial-temporal model for extracting reflection or refraction waves from OBS data are shown. In time series decomposition, a state space model based on AR representations of both direct/multiple water waves and reflection and refraction waves traveled in the crust are used. Unknown parameters of the models are estimated by the maximum likelihood method and by the self-organizing state space model. In spatial modeling the delay structure of various types of waves are considered. This spatial model is combined with the time series model and approximated method of spatial-temporal smoothing is obtained.