Calculation of false detection rate for Nasu interferometric sky survey

T. Aoki[1,2], T. Tanaka[3], K. Niinuma[4], K. Asuma[1], S. Kida[2], T. Nakamizo[1], N. Furukawa[1], J. Ikouga[1], K. Odai[1], Y. Yamada[1], R. Hiruma[1], T. Endo[1], K. Konishi[1], T. Tsuda[1], T. Daishido[2]

ABSTRACT

This paper proposes a method for calculating the false detection/alarm rate for the Nasu interferometric sky survey. This method considers both spectral intensity and time width of a fringe in the time-frequency domain. It detects periodic waves with a given time width in a noise band and calculates the distribution of the spectral intensity of the waves. Using this distribution, we can calculate the false positive rate of fringe detection and determine the reliability of the detection results as an objective probability. The proposed method can be applied to interferometric observations such as the Nasu sky survey whose observation results depend entirely on the existence of fringes.

Subject headings: Data Analysis and Techniques

1. Introduction

Temporal Fourier analysis of time-series data, particularly called the short-time Fourier transform (STFT), is used for revealing wave features and detecting signals from background noise (e.g., Nawab & Quatieri 1987). On observation with an astronomical interferometer, the Fourier analysis enables us to detect fringes as signals and discover astronomical objects. In the Nasu sky survey that is conducted by Waseda University in Japan, the STFT is used for detecting fringes in time-series data (Kuniyoshi et al. 2006; Takefuji et al. 2007).

[1]Graduate School of Advanced Science and Engineering, Waseda University, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan. takahiro_aoki@aoni.waseda.jp

[2]School of Education, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku-ku, Tokyo 169-8050, Japan.

[3]Mitsubishi Electric Corporation, 2-7-3 Marunouchi, Chiyoda-ku, Tokyo 100-8310, Japan.

[4]Graduate School of Science and Engineering, Yamaguchi University, 1677-1 Yoshida, Yamaguchi-shi, Yamaguchi 753-8512, Japan.

However, it is yet to be proven whether the detected signal-like wave is a genuine signal. A large signal-to-noise ratio (SNR) and a small false detection rate (FDR) are essential to conclude that the detected wave is indeed a signal.

The FDR of a periodic wave is derived from the distribution of spectral intensity in the frequency domain. The spectral amplitude and the spectral power of the Gaussian noise have a Rayleigh and an exponential distribution, respectively (see e.g., Papoulis 1984 or Appendix A). The FDR of a wave with a certain spectral intensity is calculated as the upper probability of these distributions, i.e., the probability that the intensity is caused by noise. This FDR calculation method is effective only if one imposes a condition C1, according to which the spectral intensity is large enough to not be generated by noise, on signal detection. In the field of radio astronomy, condition C2, according to which the signal-channel intensity must be the maximum in the total band or in the two-dimensional grid (fringe frequency-delay plane), is additionally adopted. Then, the FDR is calculated from the distribution of the maximum value of Rayleigh-distributed random variables. This calculation method is discussed in Moran (1976) and Thompson et al. (1986).

However, these FDR calculation methods do not consider the time width of the signal appearance (hereinafter called signal duration) and therefore ignore whether the detected signal-like wave maintains a desirable wave-form. Here the term "duration" does not necessarily signify the time width in which a certain physical phenomenon continues. In a drift-scanning interferometer observation such as the Nasu sky survey, for example, a fringe appears with the specified duration that corresponds to the instrumental characteristics, i.e., the full width at half maximum of the receiving sensitivity pattern when a point source is observed. For reliable signal detection, we factor in a condition C2′, according to which a signal wave must have a duration that is greater than the specified duration. This idea implies that condition C2 should be replaced with condition C2′.

In this paper, we discuss an FDR calculation method that is useful when one imposes conditions C1 and C2′ on signal detection. This FDR calculation enables us to assess the authenticity of the detection result or obtain the SNR threshold for signal detection. In Section 2, we show the methodology by using the data that include fringes as the signal output from a basic two-element interferometer for the following application of the method to the Nasu sky survey. In Section 3, we apply the method to Gaussian noise data and derive the probability distribution of the intensity of a periodic wave generated by noise. Section 4 shows the application to the Nasu sky survey. We present the observation data and compute the FDR for each fringe having an astronomical origin. Section 5 summarizes the methodology and the results of the applications, and Appendix A provides the basics of the statistical distributions used in this study.

2. Methodology

The FDR calculation method treats the time-series data $x(t)$, which include infrequently appearing signal waves in the background noise. Suppose that the signals have the following two properties: a large intensity that satisfies condition C1, and a duration greater than the given time width that satisfies condition C2′. Figure 1 shows an example of a signal that exhibits these properties and satisfies the signal-detection conditions C1 and C2′. A wave that satisfies these conditions is regarded as a signal. However, such a wave is generated not only in the signal band but also in the noise band. In other words, even noise energizes a periodic wave (hereinafter called noise wave) with a large intensity and duration. Accordingly, while identifying the detected signal-like wave as a genuine signal, we must calculate the probability of SNR caused by noise waves. This technique calculates the probability, i.e., FDR, from the distribution of the SNR of a noise wave that has a duration greater than the specified duration $\Delta t_{\rm s}$. The strict signal-detection condition restricts the detection of signal-like waves, but the FDR of the once-detected wave is expected to be small.

In this section, we use the simulated data shown in Figure 1 for describing the methodology of the technique. The interferometric observation data include fringes as signals that satisfy conditions C1 and C2′. Although the physical dimension of data $x(t)$ is arbitrary, for the following application to the Nasu sky survey, we suppose that data $x(t)$ are the cross-correlation of two voltage outputs from two telescopes. Hence, in this case, the dimension of $x(t)$ is equivalent to that of power. We also mention the case in which data $x(t)$ refer to voltage, if necessary. Let $x(t)$ and $x[\hat{t}]$ denote the continuous and discrete data, respectively, where the time index $\hat{t}$ is defined by $\hat{t} = t/t_{\rm smp}$ with a sample interval $t_{\rm smp}$. Table 1 summarizes the analysis parameters and values that are consistently used in this paper.

2.1. Definition of SNR

First, we derive the STFT of the data $x[\hat{t}]$ by using the direct method defined as

$$X[\hat{t}, \hat{f}] = \sum_{k=0}^{n-1} x[k]\, w[k - \hat{t}]\, \mathrm{e}^{-\mathrm{i}2\pi \hat{f} k/n}, \tag{1}$$

where $w[\hat{t}]$ is a window function and n is the window width. A frequency index, i.e., a spectral channel $\hat{f}$, is an integer defined by $\hat{f} = f \times n t_{\rm smp}$ and ranges from 0 to $n/2$ if the data $x[\hat{t}]$ are real numbers. A compactly supported function, e.g., Hann window, is desirable for the window function in order to avoid the Gibbs phenomena that hinder the processing mentioned in Section 2.2. The power spectrum at a given time is defined by $|X[\hat{t}, \hat{f}]|$ because

the data $x[\hat{t}]$ refer to the power. This time-series power spectrum is called a spectrogram $S[\hat{t},\hat{f}] = |X[\hat{t},\hat{f}]|$. Now, the data $x[\hat{t}]$ are cross-correlation values, and thus, $|X[\hat{t},\hat{f}]|$ is called a cross-power spectrum. However, note that the spectrogram is defined by $S[\hat{t},\hat{f}] = |X[\hat{t},\hat{f}]|^2$ if the data $x[\hat{t}]$ refer to voltage.

Next, we define the SNR from the spectrogram $S[\hat{t},\hat{f}]$. Fixing time $\hat{t}$ reduces the spectrogram $S[\hat{t},\hat{f}]$ to the power spectrum $S[\hat{f}]$. Then, transform the power spectrum $S[\hat{f}]$ to the ratio $R[\hat{f}]$ defined as

$$R[\hat{f}] = \frac{S[\hat{f}]}{\langle S[\hat{f}_{\rm n}]\rangle}, \quad \langle S[\hat{f}_{\rm n}]\rangle = \frac{1}{n-1}\sum_{\hat{f}_{\rm n}\neq\hat{f}} S[\hat{f}_{\rm n}]. \tag{2}$$

In the signal band $\hat{f} = \hat{f}_{\rm s}$, the ratio $R[\hat{f}_{\rm s}]$ refers to the SNR that is the signal intensity normalized with the average of the noise intensity $\langle S[\hat{f}_{\rm n}]\rangle$. Equation (2) indicates that the SNR definition is expanded to the noise band. The SNR can be arbitrarily defined and will change the distribution discussed in Section 2.3, but does not affect the subsequent FDR calculation. The chronologically arrayed SNR $R[\hat{t},\hat{f}]$ refers to a modified spectrogram that is transformed from the spectrogram $S[\hat{t},\hat{f}]$.

A signal appears with a wave-form in the time-series data and with an embossment-form in the spectrogram. However, embossments also appear in the noise band in the spectrogram. Consequently, corresponding waves are observed in the time-series data. A loud noise wave exists with an SNR of 4 in the area $[\hat{t},\hat{f}] \simeq [450, 63]$ in Figure 1. This noise wave can be erroneously detected as a signal if it appears with a time width greater than the specified signal duration $\Delta\hat{t}_{\rm s}$ and consequently satisfies the signal-detection condition despite the presence of intrinsic noise. In order to search for noise waves that might have been erroneously detected, measurements of the duration of the wave or embossment are required.

2.2. Definition of duration

The duration of a wave can be defined by using the modified spectrogram $R[\hat{t},\hat{f}]$ as follows. We consider the time width between the rise and decay of an embossment as the duration in the time-series SNR $R[\hat{t}]$ at a given frequency (Figure 2). Note that an inappropriate window function impedes the measurement of the duration because of the Gibbs phenomena and the rippling curve $R[\hat{t}]$. Then, we record the peak value of the embossment as the SNR of the wave if the measured duration is greater than the specified duration $\Delta\hat{t}_{\rm s}$. This leads to the following mathematical expression for the SNR r of the

detected wave:

$$r = \text{local max}\{\ldots, R[\hat{t}],\ R[\hat{t}+1],\ R[\hat{t}+2], \ldots\} \tag{3}$$

because the variables $R[*]$ are not deterministic but stochastic.

2.3. Derivation of SNR distribution

There remain some unanswered questions such as whether a noise wave with a duration greater than $\Delta\hat{t}_{\rm s}$ exists, and if it does, how many such waves exist and what are their SNR values. To answer these questions, we count the number of noise waves detected by the processing described in Section 2.2 and derive the histogram as shown in Figure 3. The histogram shows the number of noise waves classified by the SNR r. The left side of Figure 3 represents the histogram in the signal band and shows the detection of the real signal at a bin of SNR > 7. On the other hand, the right side of the figure shows the SNR distribution of the detected noise waves, which indicates that the possibility of eruption of noise waves with SNR > 5 is low. In order to calculate this probability, we derive the probability density function of the distribution.

2.4. Calculation of false detection rate

To calculate the FDR, we estimate the probability density function of the SNR of noise waves. If the observation data $x[\hat{t}]$ originate from stationary random noise, an approximation of the probability distribution of the SNR is theoretically derived (see Section 3 and Appendix A). However, the noise in the actual observation data is both random and systematic. Moreover, the stochastic process of noise is not consistently stationary in a long-term observation. Therefore, we estimate the probability density function that is equivalent to the histogram of Figure 3 by using a non-parametric method. In this study, we use the kernel density estimation method.

Let $\{r_j;\ j = 1, 2, \ldots, N\}$ denote the SNR values of noise waves that constitute the histogram, where N is the number of detected noise waves. Note that the random variable r is a positive real number and the probability density function $p(r)$ has bounded support on $[0, \infty)$. Accordingly, we conduct an estimation based on a logarithmic conversion described below (Copas & Fryer 1980; Silverman 1986). we estimate the probability density $q(\log r)$ using a Gaussian kernel:

$$q(\log r)\,\mathrm{d}(\log r) = \frac{1}{Nh}\sum_{j=1}^{N} K\left(\frac{\log r - \log r_j}{h}\right)\mathrm{d}(\log r). \tag{4}$$

The parameter h refers to the smoothing bandwidth, and the function $K(*)$ is the Gaussian kernel, i.e., the probability density function of the standard normal distribution. We determine the bandwidth by using the method provided by Scott (1992) and the *R stats package*. Consequently, the probability density $p(r)$ is defined as

$$p(r)\,\mathrm{d}r = \frac{q(\log r)}{r}\,\mathrm{d}r. \tag{5}$$

In this example, the parameters are $N = 300$ and $h = 0.11$, and the estimated function is represented by the curve on the right side of Figure 3. The FDR is calculated by integrating the estimated function $p(r)$. The signal-like wave of Figure 1 has a duration of $\Delta\hat{t}_t exts$ and an SNR of $r_\mathrm{s} \simeq 13$. If such a wave is generated by noise, the probability is $\int_{13}^{\infty} p(r)\,\mathrm{d}r < 1.0 \times 10^{-25}$, which implies that the FDR is extremely small. Therefore, we can conclude that the detected wave is a genuine signal. Summarizing the discussion, when a signal-like wave has a duration of $\Delta\hat{t}_\mathrm{s}$ and an SNR of r_s, we can calculate the FDR as

$$\mathrm{FDR} = \int_{r_\mathrm{s}}^{\infty} p(r)\,\mathrm{d}r. \tag{6}$$

However, in this example, the FDR computed above is not highly reliable because of the small number of data points. The observation data $x[\hat{t}]$ should be dominated by noise because this FDR calculation method analyzes noise in detail. The application of the proposed FDR method to actual observation data is presented in Section 4. In the next section, we discuss the distribution of the SNR in case the data $x[\hat{t}]$ consist of stationary Gaussian noise.

3. Application to Gaussian noise

We apply this FDR calculation method to Gaussian noise data with 100,000 data points. The noise data $x[\hat{t}]$ consist of normally distributed random numbers with mean 0 and variance 1, and are generated by the *Mersenne Twister* random number generator (Matsumoto & Nishimura 1998). The distribution of the SNR of a noise wave is shown in Figure 4, and is obtained by analyzing the Gaussian noise data with the parameters given in Table 1.

This distribution depends on the analysis parameters but is approximated by the distribution of the maximum value of the independent and identically distributed Rayleigh random variables. Therefore, the probability density can be approximated by

$$p(r; m, \sigma) \simeq m\left(1 - \mathrm{e}^{-r^2/(2\sigma^2)}\right)^{m-1} \cdot \frac{r}{\sigma^2}\,\mathrm{e}^{-r^2/(2\sigma^2)}. \tag{7}$$

This is because the SNR r of a noise wave is defined by Equation (3) as the local maximum value of the embossment $R[\hat{t}]$, and the variables $R[\hat{t}]$ are dependently distributed Rayleigh

variables. This approximation is also discussed in Appendix A. Then, the parameter is estimated at $m \simeq 3.47$ and $\sigma \simeq 0.907$ by using the maximum likelihood estimation for the distribution shown in Figure 4. For confirmation, we conduct the χ^2 test of the goodness of fit between the obtained distribution and $p(r)$. The result shows the p-value to be 0.33, which is greater than the common significance level of 0.05. Therefore, the null hypothesis, i.e., the SNR distribution is described by Equation (7), is not rejected.

The probability density function $p(r)$ provides the SNR threshold necessary for signal detection. The SNR threshold value is calculated by

$$r_{\rm th} \simeq \sigma \sqrt{-2 \ln \left[1 - (1 - {\rm FDR})^{1/m}\right]} \tag{8}$$

which is derived from Equations (6) and (7). The parameter σ, which is equivalent to noise intensity, remains on the right hand side even though the left hand side refers to the SNR in Equation (8). This is because of the preservation of the parameter σ as a free parameter for the approximation of Equation (7), as mentioned in Appendix A. In this case, the threshold is $r_{\rm th} \simeq 3.7$ for ${\rm FDR} = 10^{-3}$ and $r_{\rm th} \simeq 5.0$ for ${\rm FDR} = 10^{-6}$. Using the SNR threshold of 5, we can reliably detect a signal provisionally.

4. Application to Nasu sky survey data

We apply the proposed FDR calculation method to the actual observation data of the Nasu sky survey. The Nasu sky survey is an observation project at the Nasu Observatory of the Waseda University in Japan. The Nasu Observatory is located 150 km north of Tokyo and is equipped with eight east-west arrayed 20 m radio telescopes. The main reflectors of these telescopes are spherical in shape and are fixed to the ground; the Arecibo 305 m telescope is also used. We construct basic two-element interferometers using a combination of the above-mentioned telescopes and observe the sky using a drift-scanning technique, which considers the transit of an astronomical source. The basic specifications of the Nasu Observatory are reported by Takeuchi et al. (2005), Kuniyoshi et al. (2006), Niinuma et al. (2007), and Takefuji et al. (2007).

Table 2 lists the observation parameters for the data used in this section. The observation started on 2010-06-29, as specified in the ISO 8601 date format. We sample the voltages that are outputted from two telescopes at the Nyquist rate, digitally calculate the cross-correlation of the two voltages, and acquire the cross-correlation data $x[\hat{t}]$ sampled at intervals of $t_{\rm smp} = 0.629\,15$ s. An observed fringe has a duration of $\sim$144 s, which corresponds to a beam width of $36'$ when a point source is received at the beam center. Hence, the signal duration in the observation data is $\Delta t_{\rm s} = 144$ s, and the duration time index is

$\Delta\hat{t}_{\rm s} = \Delta t_{\rm s}/t_{\rm smp} = 228$.

Figure 5 shows the modified spectrogram of the observation data that consist of 68,000 data points observed for 12 h. The downward arrows indicate the detected fringes at this single epoch observation. One of the fringes originates dominantly from galaxy J004055+331007 and has an SNR of 10.2, as indicated by arrow (6) and shown in Figure 6. However, this fringe-like wave has not yet been proven to be a true fringe having astronomical origins. In order to validate the detection result, we need to calculate the FDR for each fringe.

The proposed FDR calculation method involves the derivation of the SNR distribution from the modified spectrogram, as shown in Figure 7. The distribution in the signal band shows the detection of nine fringes having a duration index of more than 228 and an SNR greater than the threshold of 5. These fringes are indicated by arrows in Figure 5. From the right side of Figure 7, we estimate the probability density function $p(r)$ from Equation (5) for the SNR of the noise waves. Integrating $p(r)$ with Equation (6) provides FDR $= 5.9 \times 10^{-68}$ for $r_{\rm s} = 10.2$ that is the SNR of the J004055+331007 fringe. Thus, we can conclude that the detected fringe-like wave of J004055+331007 is definitely a fringe. The FDR for each of the six fringes is listed in Table 3, where these six of the nine detected fringes are identified as known astronomical objects. This table shows that the observation configuration, presented in Table 2, of the Nasu Observatory can be used to detect sources whose flux density is greater than ~ 1 Jy, with a high level of confidence at a single epoch observation.

Furthermore, we observed the same declination strip for nine epochs, every day from 2010-06-29 to 2010-07-09 but for 07-01 and 07-02. This observation result is summarized in Table 4. We detected all the sources that were expected to be observed with a flux density greater than 0.5 Jy and no sources that were concluded to be variable only from these observations. However, the comparison of the single observation with the nine observations implies that the counterparts of the fringes tagged as (1) and (2) in both Tables 3 and 4 may indicate variability. The counterpart of fringe (1), J160207+332653, was detected with a flux density of ~ 2.3 Jy, which is significantly less than the presumably received flux density of 2.9 Jy at 1.4 GHz. On the other hand, Kuehr et al. (1981) reported that the flux density of the source is 2.5 Jy, which is in the error range of our observed flux density. The source J160207+332653, also known as blazar 4C +33.38 (Pilkington & Scott 1965), may show ~ 0.5 Jy variation, which is consistent with the variability of $\sim 15\%$ at 15 GHz reported by Tremblay et al. (2010). Moreover, a counterpart of fringe (2), J190348+331920 or a nearby source, may indicate intraday or intraweek variability. To confirm the implication, other observations are required. Table 4 shows that the observation configuration can be used to detect sources whose flux density is greater than ~ 0.3 Jy at nine epoch observations.

5. Conclusion

In this paper, we proposed a method for calculating the FDR in order to verify an observation result for the Nasu sky survey. This method calculated the SNR distribution of a periodic wave that was generated by noise for the data. During the detection of a signal-like wave, we had to calculate the FDR for the wave and verify that the wave was definitely a signal. The proposed method was useful when we imposed a signal-detection condition, which stipulates that a real signal maintains a desirable wave-form and has a duration greater than the specified time width. The application of the proposed method to Gaussian noise data was discussed in Section 3. In this case, the SNR distribution of the noise wave was approximately equal to the distribution of the maximum value of Rayleigh-distributed random variables. The result indicated that we could reliably detect signals with SNR > 5. The SNR threshold of 5 allowed us to detect fringes from the Nasu sky survey data, as mentioned in Section 4. In order to statistically assess the authenticity of the detection result, we carried out the kernel density estimation for the SNR distribution and calculated the FDR. The result revealed the detection of astronomical sources with a flux density greater than ~ 1 Jy at the single epoch observation. The FDR of each source was extremely low and the confidence level was significantly high. Moreover, the proposed FDR calculation revealed the detection of sources with a flux density greater than ~ 0.3 Jy at the nine epoch observations. This method is helpful for verifying the reliability of a detection result from various perspectives for the Nasu interferometric sky survey.

This work was supported by Grant-in-Aid for JSPS Fellows 22-4016. In this research, we used data obtained from the High Energy Astrophysics Science Archive Research Center (HEASARC), provided by NASA's Goddard Space Flight Center. We also used the NASA/IPAC Extragalactic Database (NED), which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

A. Basics of statistical distributions

The derivation of Equation (7) is outlined below. The detailed calculation is explained by Papoulis (1984) for example. For simplicity, let random variables and values have the same symbols and statistical distributions as those mentioned in Table 5. Assume that a random variable x has a normal distribution $\mathrm{N}(\mu, \sigma^2)$. The set of values $\{x_j;\ j = 0, 1, \ldots, n-1\}$ corresponds to a subset of data $x[\hat{t}]$ mentioned in Section 2. The validity of the assumption is explained below in the case of the data obtained using a radio interferometer. The data x_j are a cross-correlation between two voltage outputs from the telescopes, i.e., an ensemble

average of the product of ergodic voltages. Hence, a large number of averaging data points normally distributes the data x_j according to the central limit theorem.

When random variable x_j has a normal distribution, the orthogonal transform of x_j defined as

$$X_k = \sum_{j=0}^{n-1} a_{kj} x_j, \ \sum_{l=0}^{n-1} a_{kl} a_{lj} = \delta_{kj} \tag{A1}$$

is normally distributed $\mathrm{N}(\mu \sum_{j=0}^{n-1} a_{kj}, \sigma^2)$, where δ_{kj} is the Kronecker delta. The discrete Fourier transform is the orthogonal transform of $a_{kj} \propto \exp(-\mathrm{i}2\pi kj/n)$, which reveals that the real part $\mathrm{Re}X_k$ and the imaginary part $\mathrm{Im}X_k$ of the Fourier transform are independently and identically normally distributed $\mathrm{N}(n\mu\,\delta_{k0}, \sigma^2)$. Consider the case of $k \neq 0$; the random variables $\mathrm{Re}X_k$ and $\mathrm{Im}X_k$ have a normal distribution $\mathrm{N}(0, \sigma^2)$. Now, define the amplitude spectrum as

$$A_k = |X_k| = \sqrt{(\mathrm{Re}X_k)^2 + (\mathrm{Im}X_k)^2} \ ; \tag{A2}$$

the random variable A_k has the distribution Rayleigh(σ), mentioned in Table 5, because of the transformation of the random variables. Define the power spectrum as

$$P_k = |X_k|^2 = A_k^2 \ ; \tag{A3}$$

the random variable P_k has the exponential distribution $\mathrm{Exp}(1/(2\sigma^2))$. We can calculate the FDR by using the Rayleigh or exponential distributions if condition C1 is imposed on signal detection (see Section 1).

By additionally imposing condition C2, we obtain the distribution of the maximum value of spectral intensity. When the data x_j are real numbers with n data points, the number of independent spectral channels is $m = n/2$ and the range of the integer k is $k = 0, 1, \ldots, m-1$. Hence, the probability density of the maximum value $\mathcal{A}$ of the amplitude spectrum $\{A_k\}$ is

$$p(\mathcal{A}; m, \sigma)\,\mathrm{d}\mathcal{A} = m\left(1 - \mathrm{e}^{-\mathcal{A}^2/(2\sigma^2)}\right)^{m-1} \cdot \frac{\mathcal{A}}{\sigma^2}\,\mathrm{e}^{-\mathcal{A}^2/(2\sigma^2)}\,\mathrm{d}\mathcal{A}. \tag{A4}$$

Likewise, the probability density of the maximum value $\mathcal{P}$ of the power spectrum $\{P_k\}$ is

$$p(\mathcal{P}; m, \sigma)\,\mathrm{d}\mathcal{P} = m\left(1 - \mathrm{e}^{-\mathcal{P}/(2\sigma^2)}\right)^{m-1} \cdot \frac{1}{2\sigma^2}\,\mathrm{e}^{-\mathcal{P}/(2\sigma^2)}\,\mathrm{d}\mathcal{P}. \tag{A5}$$

The parameter m is replaced by $m - 1$ without considering $k = 0$. Then, we can calculate the FDR using Equation (A4) or (A5) while imposing conditions C1 and C2 on the signal detection. The above discussion is based on a spectrum, i.e., only in the frequency domain.

The method discussed in this paper considers the time extent of a wave in the time-frequency domain and derives the distribution of a local maximum value of the spectral

intensity or SNR in this range. If the data $x[\hat{t}]$ have a normal distribution $\mathrm{N}(\mu, \sigma^2)$, the spectrogram at a certain time $S[\hat{f}] = |X[\hat{f}]|$ has the distribution Rayleigh(σ), and the mean value is $\langle S \rangle = \sigma\sqrt{\pi/2}$. Therefore, the modified spectrogram at a certain time $R[\hat{f}]$ defined by Equation (2) has the distribution Rayleigh($\sqrt{2/\pi}$). The distribution of the maximum value r of the variables R_k, i.e., $r = \max\{R_0, R_1, \ldots, R_{m-1}\}$, is represented by the probability function $p(r; m, \sqrt{2/\pi})$ in Equation (A4) if the variables R_k are independent of each other. In Section 3, the variable r is defined as the local maximum value of the embossment $R[\hat{t}]$ and is expressed by Equation (3), i.e., $r =$ local $\max\{\ldots,\ R[\hat{t}],\ R[\hat{t}+1],\ R[\hat{t}+2], \ldots\}$. Note that the variables $R[*]$ are not independent, and consequently, the distribution of r is not described by $p(r; m, \sqrt{2/\pi})$. However, the distribution is roughly approximated by $p(r; m, \sigma)$ that preserves the parameter σ as a free parameter. This is why we use Equation (7) for calculating the interim SNR threshold.

REFERENCES

Condon, J. J., Cotton, W. D., Greisen, E. W., Yin, Q. F., Perley, R. A., Taylor, G. B., & Broderick, J. J. 1998, AJ, 115, 1693

Copas, J. B., & Fryer, M. J. 1980, J. R. Statist. Soc. A, 143, 167

Kuehr, H., Witzel, A., Pauliny-Toth, I. I. K., & Nauber, U. 1981, A&AS, 45, 367

Kuniyoshi, M., et al. 2006, PASP, 118, 901

Matsumoto, M., & Nishimura, T. 1998, ACM Trans. Model. Comput. Simul., 8, 3

Moran, J. M. 1976, Methods of Experimental Physics, 12, 228

Nawab, S. H., & Quatieri, T. F. 1987, Short-time Fourier transform (Upper Saddle River, NJ, USA: Prentice-Hall, Inc.), 289–337

Niinuma, K., et al. 2007, PASP, 119, 112

Papoulis, A. 1984, Probability, random variables, and stochastic processes, McGraw-Hill series in electrical engineering (McGraw-Hill)

Pilkington, J. D. H., & Scott, P. F. 1965, MmRAS, 69, 183

Scott, D. 1992, Multivariate density estimation: theory, practice, and visualization, Wiley series in probability and mathematical statistics: Applied probability and statistics (Wiley)

Silverman, B. 1986, Density estimation for statistics and data analysis, Monographs on statistics and applied probability (Chapman and Hall)

Takefuji, K., et al. 2007, PASP, 119, 1145

Takeuchi, H., et al. 2005, PASJ, 57, 815

Thompson, A. R., Moran, J. M., & Swenson, G. W. 1986, Interferometry and synthesis in radio astronomy (Wiley)

Tremblay, S. E., Taylor, G. B., Richards, J. L., Readhead, A. C. S., Helmboldt, J. F., Romani, R. W., & Healey, S. E. 2010, ApJ, 712, 159

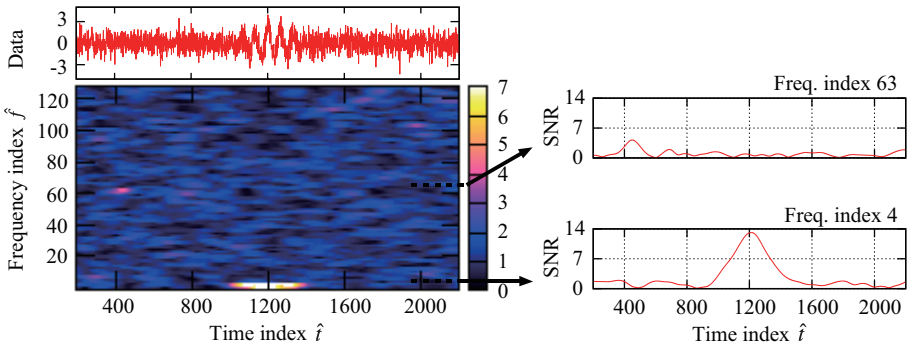

Fig. 1.— The left side shows the simulated time-series data $x[\hat{t}]$ and the modified spectrogram $R[\hat{t}, \hat{f}]$, and the right side shows the modified spectrogram at a certain frequency index $R[\hat{t}, 4]$ and $R[\hat{t}, 63]$. This FDR calculation method deals with data that include a signal wave with a duration of $\Delta\hat{t}_{\rm s}$ in the background noise. The data are created from the simulation of the Nasu sky survey, and the signal channel is a frequency index $\hat{f}_{\rm s} = 4$; the data are similar to the actual observation data, as mentioned in Section 4. The signal wave appears as an embossment in an area $[\hat{t}, \hat{f}] \simeq [1200, 4]$ on the modified spectrogram. Likewise, in an area $[\hat{t}, \hat{f}] \simeq [450, 63]$ in the noise band, a noise wave appears with SNR $R \simeq 4$. If the noise wave has the duration of the specified value $\Delta\hat{t}_{\rm s}$, this wave satisfies the signal-detection condition despite the presence of intrinsic noise. This FDR calculation method counts the number of noise waves and draws a histogram, as shown in Figure 3.

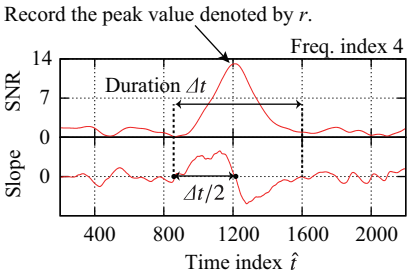

Fig. 2.— The upper curve shows the temporal variation of SNR $R[\hat{t}, 4]$ at the frequency index $\hat{f} = 4$, and the lower curve shows the slope of the curve $R[\hat{t}, 4]$. We use the slope of the local regression line of the curve $R[\hat{t}]$ to determine whether SNR $R[\hat{t}]$ increases with time $\hat{t}$. The duration is defined as twice the time width between a zero-crossing point from negative to positive and one from positive to negative. The local regression line is derived from 20 points around each data point. The 20 data points are set to approximately $\Delta\hat{t}_{\rm s}/10$, but this setting has no theoretical basis. Then, the peak value of the embossment is recorded such that the measured duration is greater than the specified signal duration $\Delta\hat{t}_{\rm s}$. This figure reveals that the signal has an SNR of approximately 13.

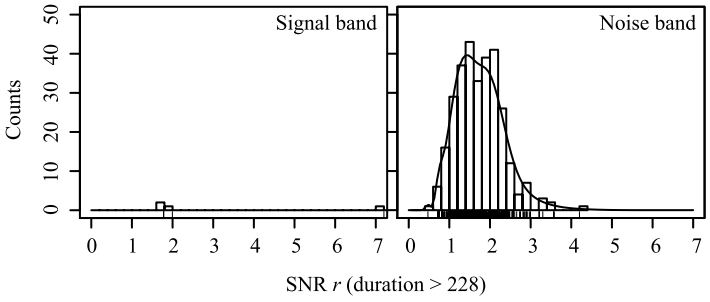

Fig. 3.— Histogram of the SNR of a wave that has a duration greater than $\Delta\hat{t}_{\rm s}$ (= 228). The data points form a rug plot on the horizontal axis. The left side of the figure shows that four waves exist in the data at the signal band $\hat{f}_{\rm s} = 4$; three of these waves have SNR < 2, and one has SNR > 7. The wave with SNR > 7 is the true signal in the area $[\hat{t}, \hat{f}] \simeq [1200, 4]$ in Figure 1. Likewise, the right side shows the SNR distribution of waves in the noise band, where the noise band is defined as a band of $\hat{f} \geq 10$ with a view to avoid leakage of the signal component. We acquire the FDR from a probability density function of the distribution estimated with a kernel function.

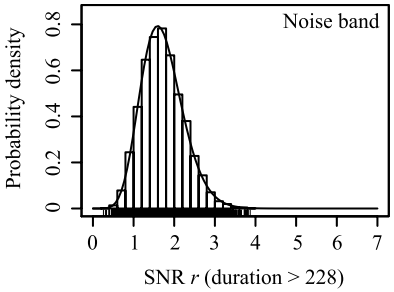

Fig. 4.— Same as Figure 3, but for the application to Gaussian noise. The shape of this distribution depends on the analysis methods and parameters given in Table 1 but is approximately expressed by Equation (7). This approximation provides the interim threshold of SNR for signal detection.

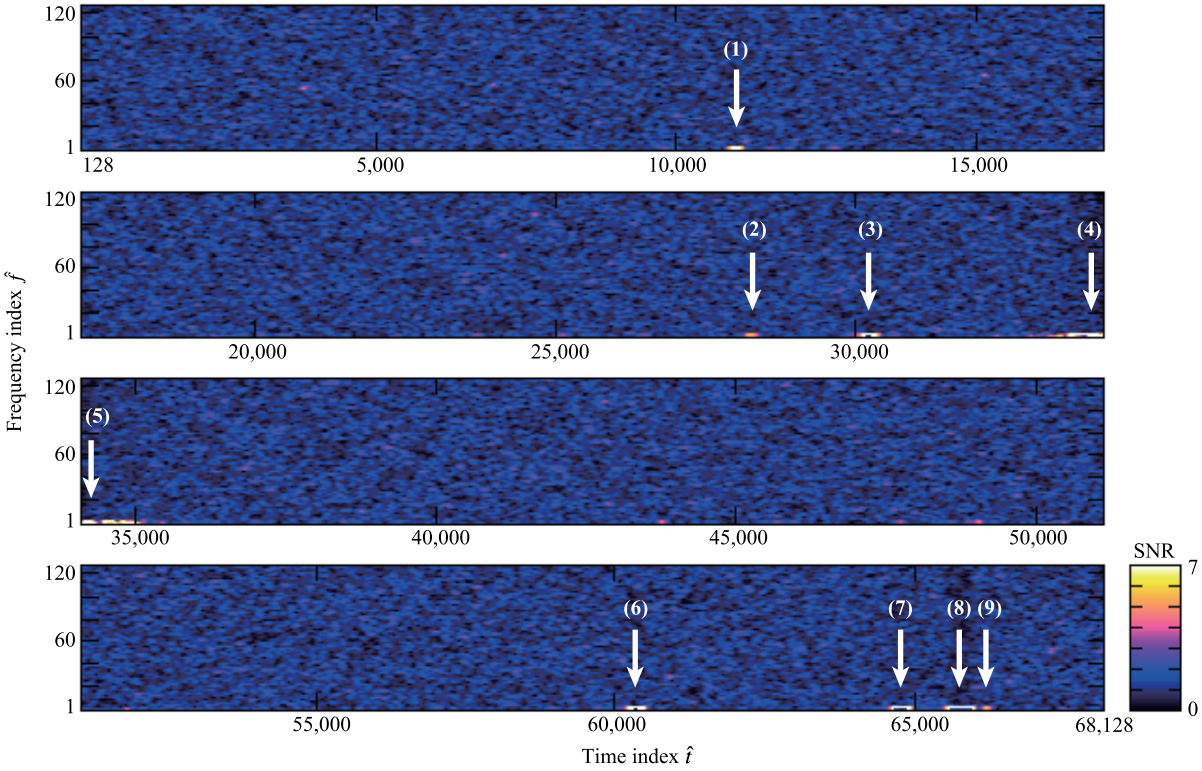

Fig. 5.— Same as Figure 1, but for the case in which the Nasu sky survey data are used. The downward arrows indicate the reliably detected fringes. The magnification of the fringe tagged as (6) is shown in Figure 6 as an example.

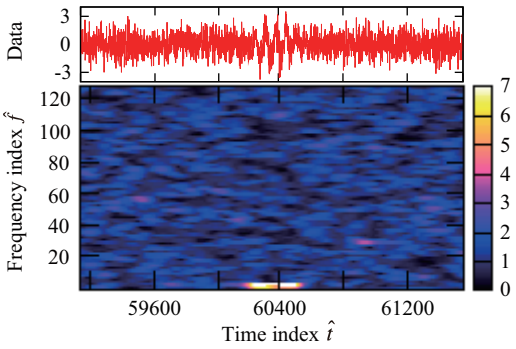

Fig. 6.— Same as Figure 1; the magnification of the fringe tagged as (6) in Figure 5. The FDR of the displayed fringe is 5.9×10^{-68}, and we can, therefore, conclude that the wave is a genuine fringe having an astronomical origin. The counterpart of this fringe is the galaxy NVSS J004055+331007 .

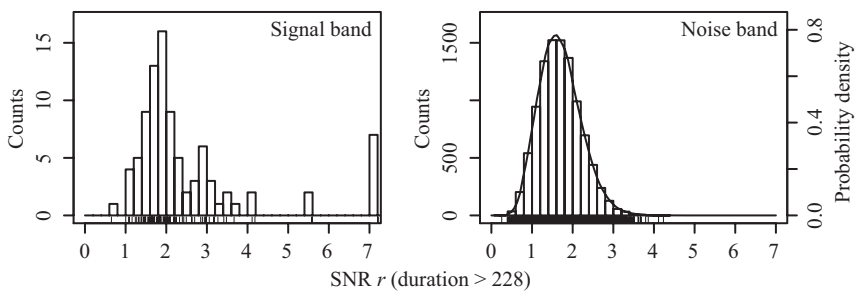

Fig. 7.— Same as Figure 3, but for the application to the actual observation data. The left figure shows the detection of nine fringes at SNR > 5.

Table 1. Analysis parameters and values.

Parameter	Value
Sample interval t_{smp}	0.629 15 s
Signal duration Δt_{s}	144 s
Signal duration index $\Delta \hat{t}_{\mathrm{s}}$	228
STFT window $w[\hat{t}]$	Hann window
Window width n	256

Table 2. Observation configuration.

Parameter	Value
Observation start	2010-06-29 10:00:00 (UT)[a]
Observation end	2010-06-29 22:00:00 (UT)[a]
Frequency	1420 ± 10 MHz
Telescope diameter	20 m
Baseline length	84 m
Integration time	0.629 15 s
System temperature	100 K
Declination	$33°30'$
Right ascension	14^h–24^h & 0^h–2^h
Resolution for Decl.	$36'$
Resolution for R.A.	$9'$

[a]The date and time representation is based on ISO 8601.

Table 3. Detected fringes at a single epoch.

Tag	Detected fringes		Counterparts		FDR		
	$\alpha_{\rm obs}$	$F_{\rm obs}$	Name	$F_{\rm exp}$	C1	C1 & C2	C1 & C2′
(1)	$16^{\rm h}02^{\rm m}02^{\rm s} \pm 14^{\rm s}$	2.3 ± 0.3 Jy	J160207+332653	2.9 Jy	10^{-22}	10^{-20}	10^{-40}
(2)	$19^{\rm h}03^{\rm m}45^{\rm s} \pm 20^{\rm s}$	1.6 ± 0.3 Jy	J190348+331920	0.8 Jy	10^{-10}	10^{-8}	10^{-10}
(3)	$19^{\rm h}24^{\rm m}12^{\rm s} \pm 11^{\rm s}$	3.1 ± 0.3 Jy	J192417+332929	3.1 Jy	10^{-42}	10^{-39}	10^{-79}
(6)	$00^{\rm h}40^{\rm m}54^{\rm s} \pm 10^{\rm s}$	3.3 ± 0.3 Jy	J004055+331007	3.2 Jy	10^{-35}	10^{-33}	10^{-67}
(7)	$01^{\rm h}26^{\rm m}43^{\rm s} \pm 8^{\rm s}$	4.4 ± 0.3 Jy	J012644+331309 J012811+332432	3.7 Jy 0.3 Jy	10^{-69}	10^{-67}	10^{-123}
(8)	$01^{\rm h}37^{\rm m}41^{\rm s} \pm 3^{\rm s}$	14.1 ± 0.4 Jy	J013741+330935	14.0 Jy	10^{-300}	10^{-300}	10^{-300}

Note. — The tags refer to the fringes numbered in Figure 5. The detected fringes or radio sources are indicated by their right ascension $\alpha_{\rm obs}$ (J2000.0), where their declination is $+33°30' \pm 36'$ because of the drift-scanning observation. The observed flux density is denoted by $F_{\rm obs}$. The calibrators of position and flux density are J013741+330935 and J192417+332929, respectively; these calibrators are catalogued in the NRAO VLA Sky Survey (NVSS) compiled by Condon et al. (1998). The possible counterparts are formally taken from the NVSS sources with a flux density greater than a noise level of 0.3 Jy and within a distance of $36'$ from the field center. The listed flux density of the counterparts, $F_{\rm exp}$, expresses an intensity that is expected to be received by a sensitivity pattern of the telescopes. The observed flux density $F_{\rm obs}$ is expected to be equal to a sum of the expected flux density $F_{\rm exp}$, if the luminosities of the counterparts are not variable. However, the sources other than the listed counterparts also increase or decrease the spatial coherence and the observed flux density. The fringes tagged as (4) and (5) in Figure 5 are not listed because they originate from sources in the galactic plane and are not resolved by the instruments of the Nasu Observatory. Furthermore, fringe (9) is eliminated because it is from the listed sources J013741+330935 and is received by the side-lobe of the telescope. The FDR based on the condition C1 is calculated by using Equation (A4), where $m = 1$ and $\sigma = \sqrt{2/\pi}$. Similarly, in the case of C1 and C2, $m = 127$ and $\sigma = \sqrt{2/\pi}$. The FDR based on C1 and C2′ is calculated with the estimated function, as described in Section 2.

Table 4. Detected fringes at nine epochs.

Tag	Detected fringes		Counterparts		FDR		
	$\alpha_{\rm obs}$	$F_{\rm obs}$	Name	$F_{\rm exp}$	C1	C1 & C2	C1 & C2′
	$15^{\rm h}04^{\rm m}18^{\rm s} \pm 24^{\rm s}$	0.3 ± 0.1 Jy	J150429+334337	0.3	10^{-7}	10^{-5}	10^{-8}
			J150508+331832	0.1			
	$15^{\rm h}50^{\rm m}50^{\rm s} \pm 16^{\rm s}$	0.6 ± 0.1 Jy	J155049+332016	0.7	10^{-17}	10^{-15}	10^{-37}
(1)	$16^{\rm h}02^{\rm m}03^{\rm s} \pm 05^{\rm s}$	2.2 ± 0.1 Jy	J160207+332653	2.9	10^{-131}	10^{-128}	10^{-218}
[a]	$16^{\rm h}18^{\rm m}32^{\rm s} \pm 31^{\rm s}$	0.4 ± 0.1 Jy	J161820+334837	0.2	10^{-10}	10^{-8}	10^{-17}
	$17^{\rm h}41^{\rm m}18^{\rm s} \pm 20^{\rm s}$	0.5 ± 0.1 Jy	J174128+333221	0.4	10^{-9}	10^{-7}	10^{-12}
			J174109+330538	0.1			
			J174111+330623	0.1			
	$18^{\rm h}12^{\rm m}43^{\rm s} \pm 35^{\rm s}$	0.3 ± 0.1 Jy	J181252+333702	0.4	10^{-8}	10^{-6}	10^{-10}
			J181322+333132	0.1			
	$18^{\rm h}15^{\rm m}55^{\rm s} \pm 15^{\rm s}$	0.7 ± 0.1 Jy	J181600+332744	0.5	10^{-19}	10^{-17}	10^{-43}
	$18^{\rm h}44^{\rm m}40^{\rm s} \pm 18^{\rm s}$	0.6 ± 0.1 Jy	J184440+331614	0.6	10^{-18}	10^{-16}	10^{-40}
(2)	$19^{\rm h}03^{\rm m}48^{\rm s} \pm 11^{\rm s}$	1.0 ± 0.1 Jy	J190348+331920	0.8	10^{-32}	10^{-29}	10^{-75}
			J190308+334040	0.1			
(3)	$19^{\rm h}24^{\rm m}12^{\rm s} \pm 04^{\rm s}$	3.1 ± 0.1 Jy	J192417+332929	3.1	10^{-232}	10^{-230}	10^{-298}
			J192345+332219	0.1			
	$21^{\rm h}38^{\rm m}16^{\rm s} \pm 24^{\rm s}$	0.4 ± 0.1 Jy	J213856+331415	0.2	10^{-8}	10^{-6}	10^{-10}
			J213814+332616	0.1			
			J213827+330115	0.1			
			J213813+330406	0.1			
	$21^{\rm h}46^{\rm m}58^{\rm s} \pm 28^{\rm s}$	0.3 ± 0.1 Jy	J214730+325902	0.2	10^{-7}	10^{-5}	10^{-8}
			J214620+333146	0.1			
			J214626+333240	0.1			
			J214626+333344	0.1			
	$22^{\rm h}06^{\rm m}32^{\rm s} \pm 20^{\rm s}$	0.4 ± 0.1 Jy	J220634+333142	0.4	10^{-9}	10^{-7}	10^{-14}
			J220719+333012	0.3			
			J220644+331555	0.3			
			J220418+331519	0.0			
	$22^{\rm h}09^{\rm m}53^{\rm s} \pm 13^{\rm s}$	0.8 ± 0.1 Jy	J220955+331408	0.6	10^{-19}	10^{-17}	10^{-43}
			J220931+334026	0.1			

Table 4—Continued

Tag	Detected fringes		Counterparts		FDR		
	$\alpha_{\rm obs}$	$F_{\rm obs}$	Name	$F_{\rm exp}$	C1	C1 & C2	C1 & C2′
			J220953+333008	0.1			
	$22^{\rm h}27^{\rm m}59^{\rm s} \pm 24^{\rm s}$	0.4 ± 0.1 Jy	J222805+333138	0.4	10^{-9}	10^{-7}	10^{-14}
	$22^{\rm h}42^{\rm m}00^{\rm s} \pm 12^{\rm s}$	0.8 ± 0.1 Jy	J224127+333713	0.4	10^{-32}	10^{-30}	10^{-76}
			J224250+332153	0.2			
	$23^{\rm h}27^{\rm m}26^{\rm s} \pm 19^{\rm s}$	0.4 ± 0.1 Jy	J232731+331805	0.3	10^{-6}	10^{-4}	10^{-6}
			J232802+332829	0.1			
			J232813+330802	0.1			
[a]	$23^{\rm h}51^{\rm m}56^{\rm s} \pm 19^{\rm s}$	0.3 ± 0.1 Jy	J235220+330403	0.5	10^{-9}	10^{-7}	10^{-13}
[a]	$23^{\rm h}54^{\rm m}17^{\rm s} \pm 20^{\rm s}$	0.5 ± 0.1 Jy	J235411+325511	0.4	10^{-11}	10^{-9}	10^{-19}
	$00^{\rm h}10^{\rm m}20^{\rm s} \pm 23^{\rm s}$	0.4 ± 0.1 Jy	J001025+333009	0.3	10^{-9}	10^{-7}	10^{-12}
			J001025+332901	0.2			
			J001003+340315	0.1			
	$00^{\rm h}26^{\rm m}07^{\rm s} \pm 25^{\rm s}$	0.3 ± 0.1 Jy	J002543+331648	0.2	10^{-6}	10^{-4}	10^{-6}
			J002757+333531	0.1			
			J002653+325840	0.1			
(6)	$00^{\rm h}40^{\rm m}54^{\rm s} \pm 04^{\rm s}$	2.8 ± 0.1 Jy	J004055+331007	3.2	10^{-174}	10^{-172}	10^{-256}
			J003938+335716	0.1			
[a]	$01^{\rm h}09^{\rm m}48^{\rm s} \pm 15^{\rm s}$	0.6 ± 0.1 Jy	J011112+330706	0.1	10^{-16}	10^{-14}	10^{-35}
(7)	$01^{\rm h}26^{\rm m}43^{\rm s} \pm 03^{\rm s}$	4.4 ± 0.1 Jy	J012644+331309	3.7	10^{-300}	10^{-300}	10^{-300}
			J012811+332432	0.3			
			J012753+330533	0.1			
(8)	$01^{\rm h}37^{\rm m}41^{\rm s} \pm 01^{\rm s}$	14.0 ± 0.2 Jy	J013741+330935	14.0	10^{-300}	10^{-300}	10^{-300}

Note. — Same as Table 3, but for the observation result of the nine epochs. The tagged fringes are also detected at a single epoch observation, as shown in Figure 5 and Table 3. The listed counterparts are the NVSS sources that are expected to be received with a flux density greater than a noise level of 0.1 Jy by the telescope sensitivity.

[a]These fringes are strongly diffused and confused, leading to errors in the position and flux density.

Table 5. Notation of distributions.

Notation	Probability density function
$\mathrm{N}(\mu, \sigma^2)$	$p(x) = (1/\sqrt{2\pi\sigma^2})\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}$
$\mathrm{Rayleigh}(\sigma)$	$p(x) = (x/\sigma^2)\,\mathrm{e}^{-x^2/(2\sigma^2)}$
$\mathrm{Exp}(1/(2\sigma^2))$	$p(x) = (1/(2\sigma^2))\,\mathrm{e}^{-x/(2\sigma^2)}$