Correlation-based evaluation for earthquake forecast

17 June 2015

K.Yamashina, H.Tsuruoka and T.Himeno

CSEP typical forecast test

Forecast

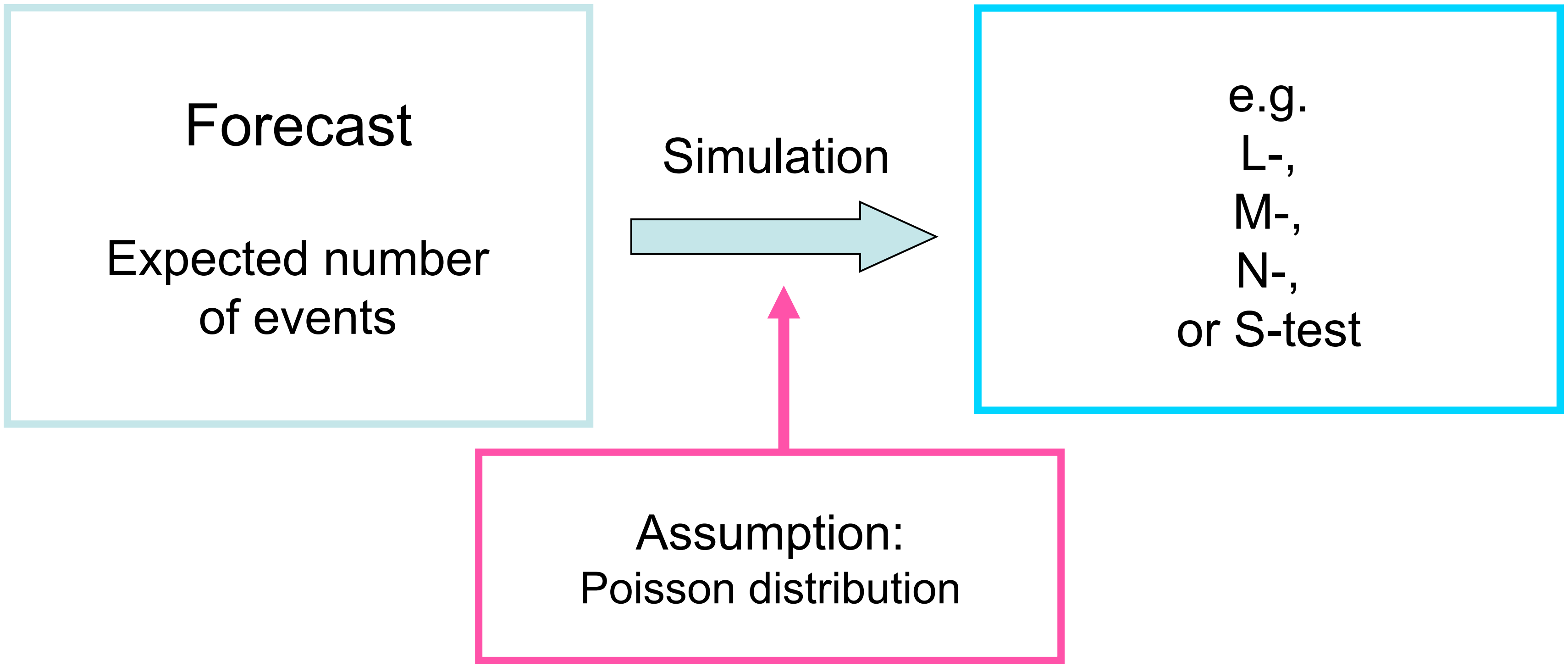
Expected number of events

Simulation →

Can results explain observations ?

In the CSEP test, results of repeated simulations are compared with observation.

CSEP typical forecast test



Here, Poisson distribution is currently assumed to carry out the simulation.

5-year frequency of earthquakes in Italy

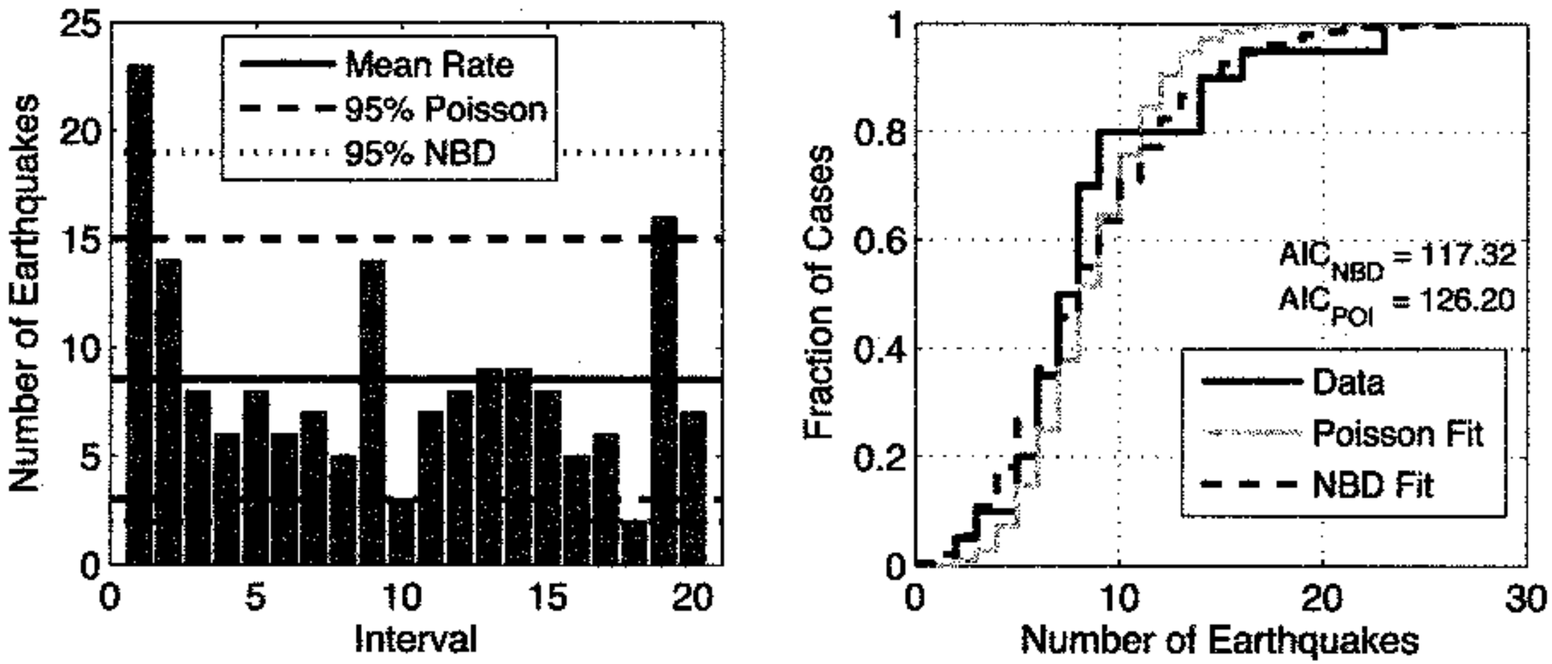


Figure 1. Left: Observed number of earthquakes in the 20 non-overlapping five-year intervals in the CPTI, from 1907 to 2006 (inclusive). Solid line, mean number of observed events; dashed lines enclose the 95% confidence interval of the Poisson distribution; dotted lines, the 95% confidence interval of the negative binomial distribution (NBD). Right: Cumulative distribution functions. Solid black line, observations; solid gray line, data fit with a Poisson distribution; dashed black line, data fit with a NBD. The Akaike Information Criterion (AIC) values of the fitted distributions are also shown.

Werner et al.(2010) pointed out that the Poisson distribution does not explain the observation. They discussed the N-test using the negative binomial distribution.

2.5-year frequency of earthquakes in California

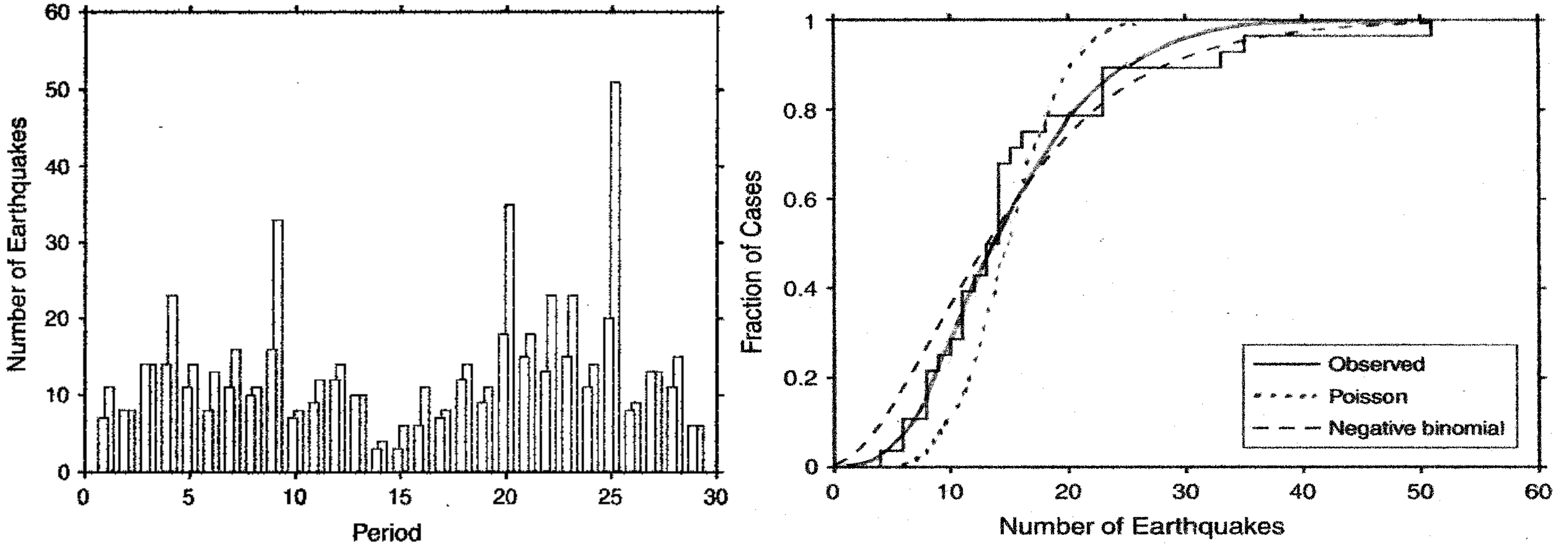


Figure 4
Earthquake rates in California from 1 January 1932 to 30 June 2004. (*left*) *Bar graph* showing the number of earthquakes in 29 non-overlapping periods of 2 years and 6 months duration. *White and gray bars* indicate the number of earthquakes in the declustered catalog, thus mainshocks only, and complete catalog, respectively. (*right*) Cumulative distribution function of the earthquakes rates in the complete catalog from the left frame. The *solid black line* indicates the observation, the *solid gray line* indicates the Poissonian distribution of rate $\lambda = 15.45$, the *dashed black line* indicates the best-fit negative binomial distribution

Another example that the negative binomial distribution was applied to the total number of earthquakes (Schorlemmer et al., 2010).

Proposed distribution functions

- Type I

$$p(x) = c\, x^{\alpha} \exp(-\beta\, x)$$

Gamma distribution

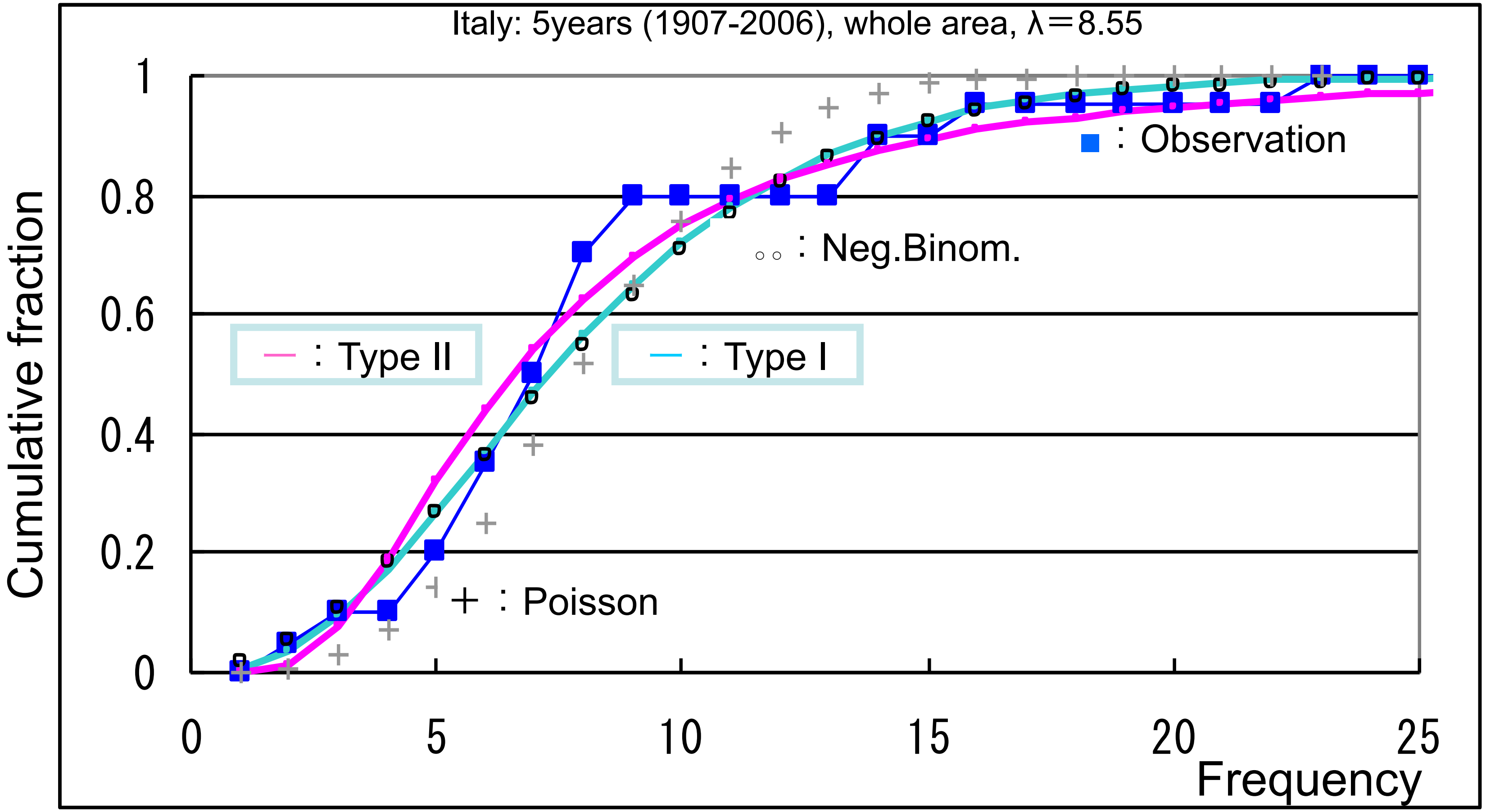
- Type II

$$p(x) = c\, x^{\alpha} \exp(-\beta / x)$$

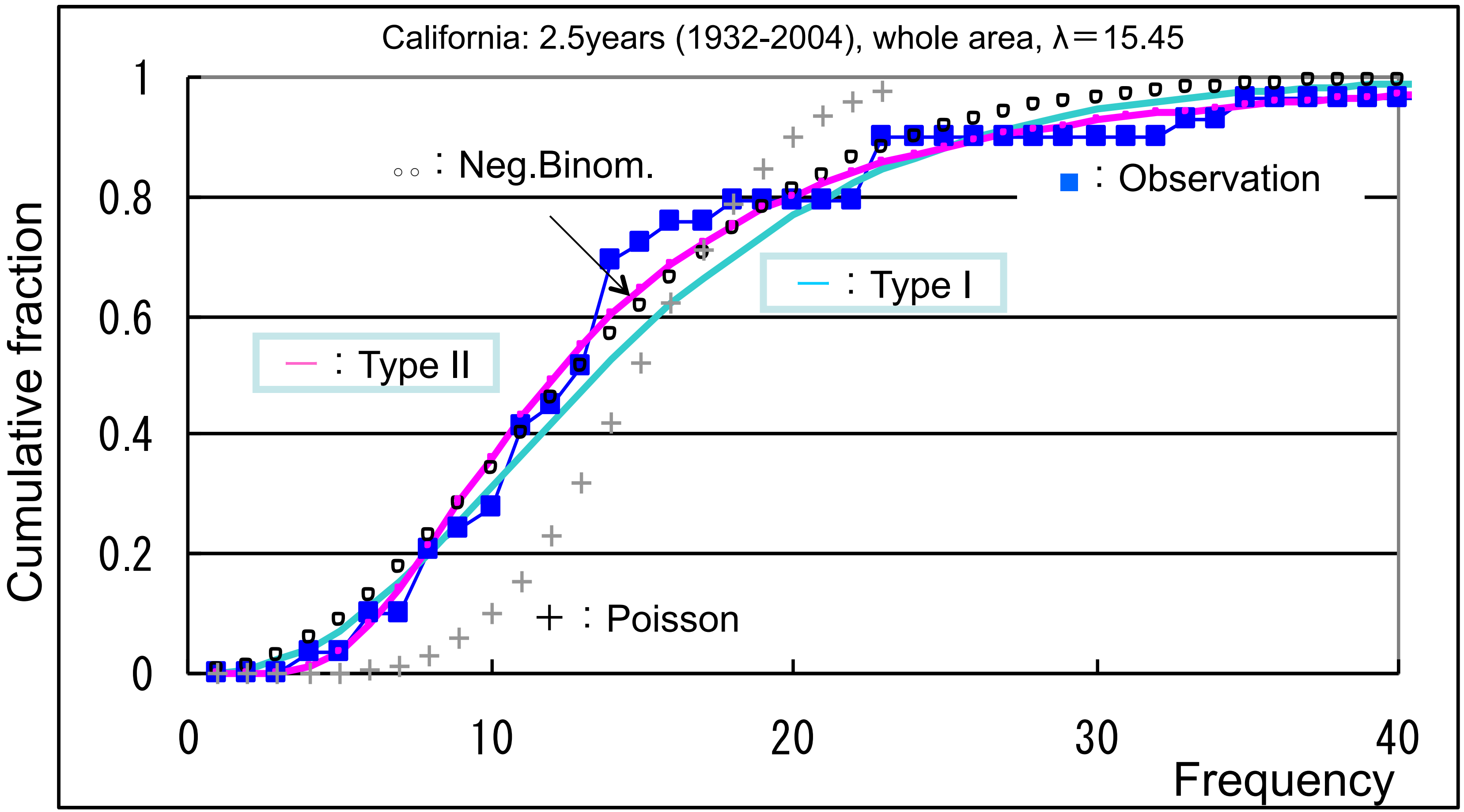
Inverse gamma distribution

1) $p(x)$ is determined only for positive integers, i.e. x =1, 2, 3, ...

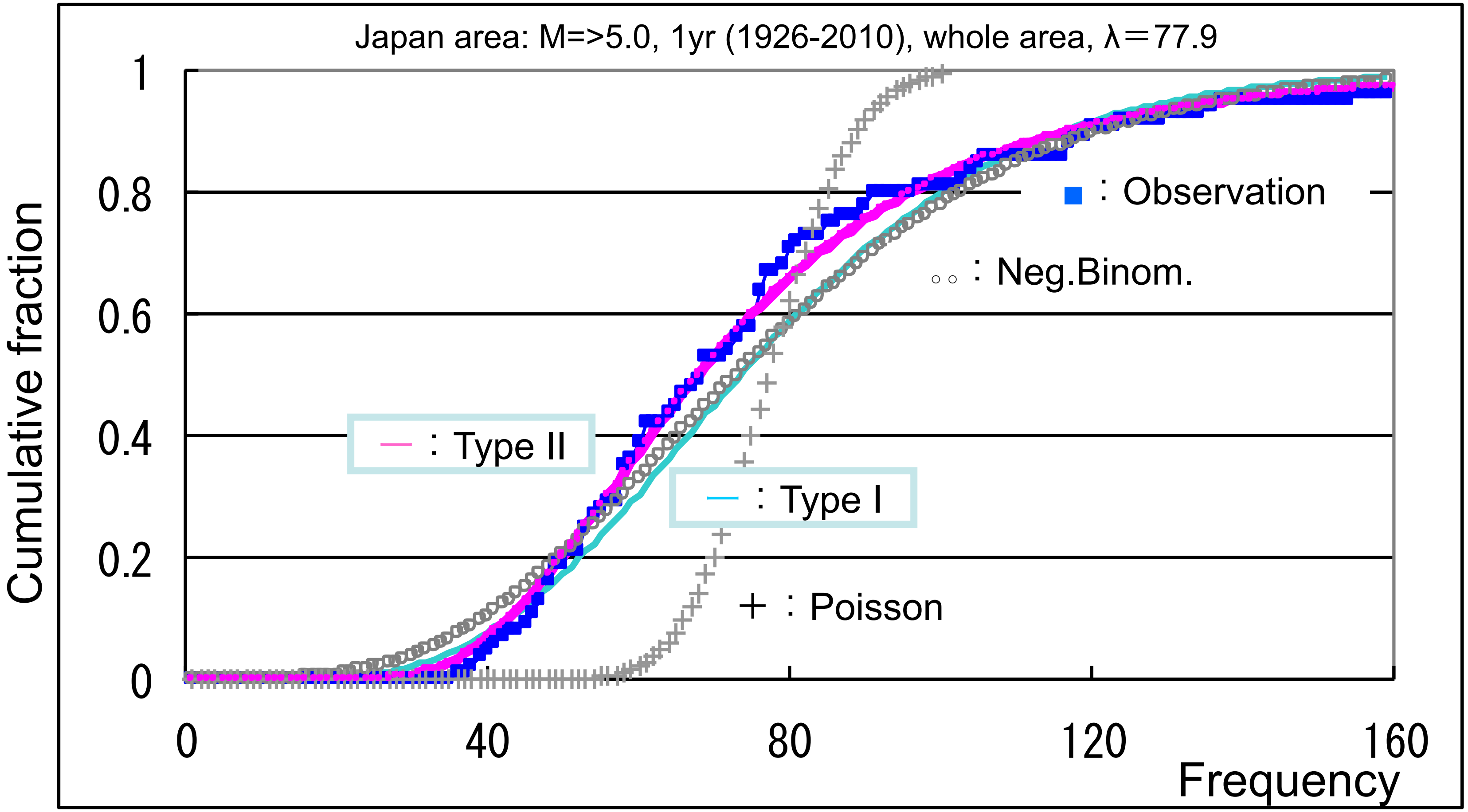
2) $p(0)$ is determined independently.



In case of Italy,
the difference among the negative binomial, Type I,
and Type II distributions is not so large with each other.



In case of California, goodness-of-fit is:
Type II > Type I > negative binomial.



In the third example obtained in Japan area, the difference appears more clearly because of the increase in the number of data. The fitness is Type II > Type I>negative binomial.

Tentative comments

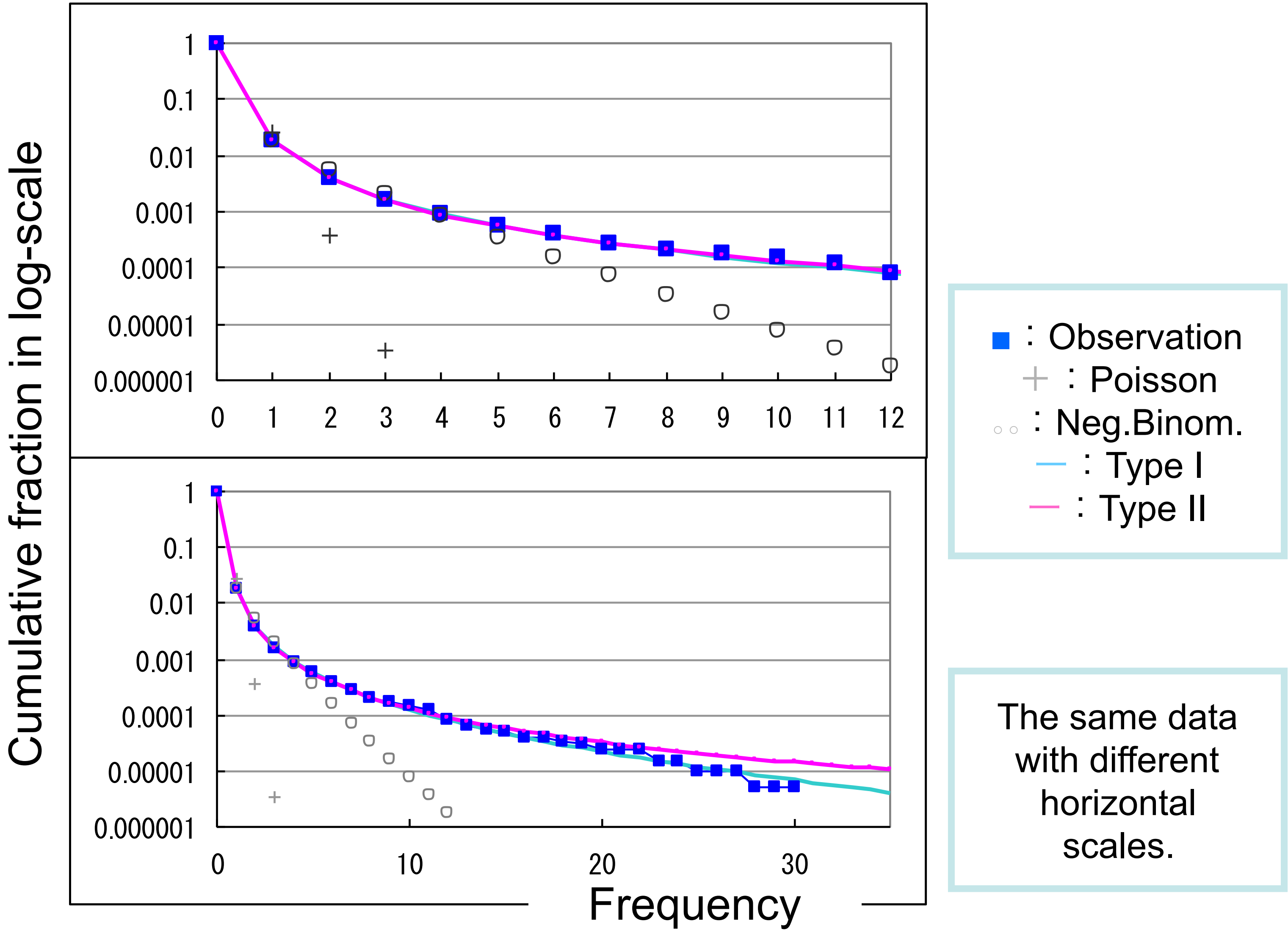
When an average λ of the frequency is much larger than 1:

$$\text{Average } \lambda : \frac{\text{total number of events}}{\text{Total number of frequency data}}$$

1) Poisson fits poorly to the observation,

2) Negative binomial and Type I are not so different from each other,

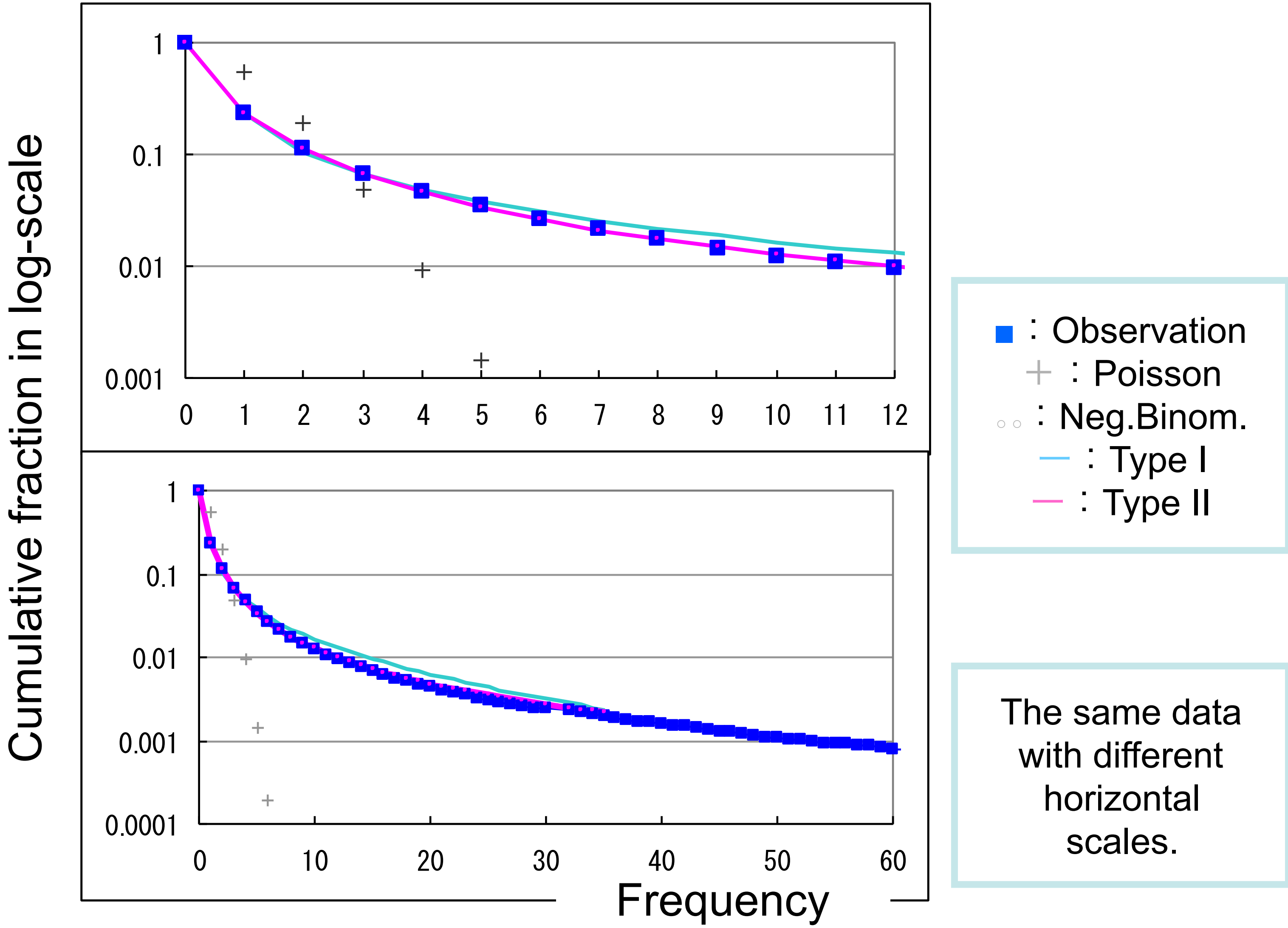
3) Type II is sometimes better than the others.



The same data with different horizontal scales.

Case study for a small average (1):

Japan area: M=>4.0, 1yr (2001-2010), 0.1x0.1 degree, λ=0.027



The same data with different horizontal scales.

Case study for a small average (2):

Japan area: M=>2.0, 1yr (2001-2010), 0.1x0.1 degree, λ=0.805

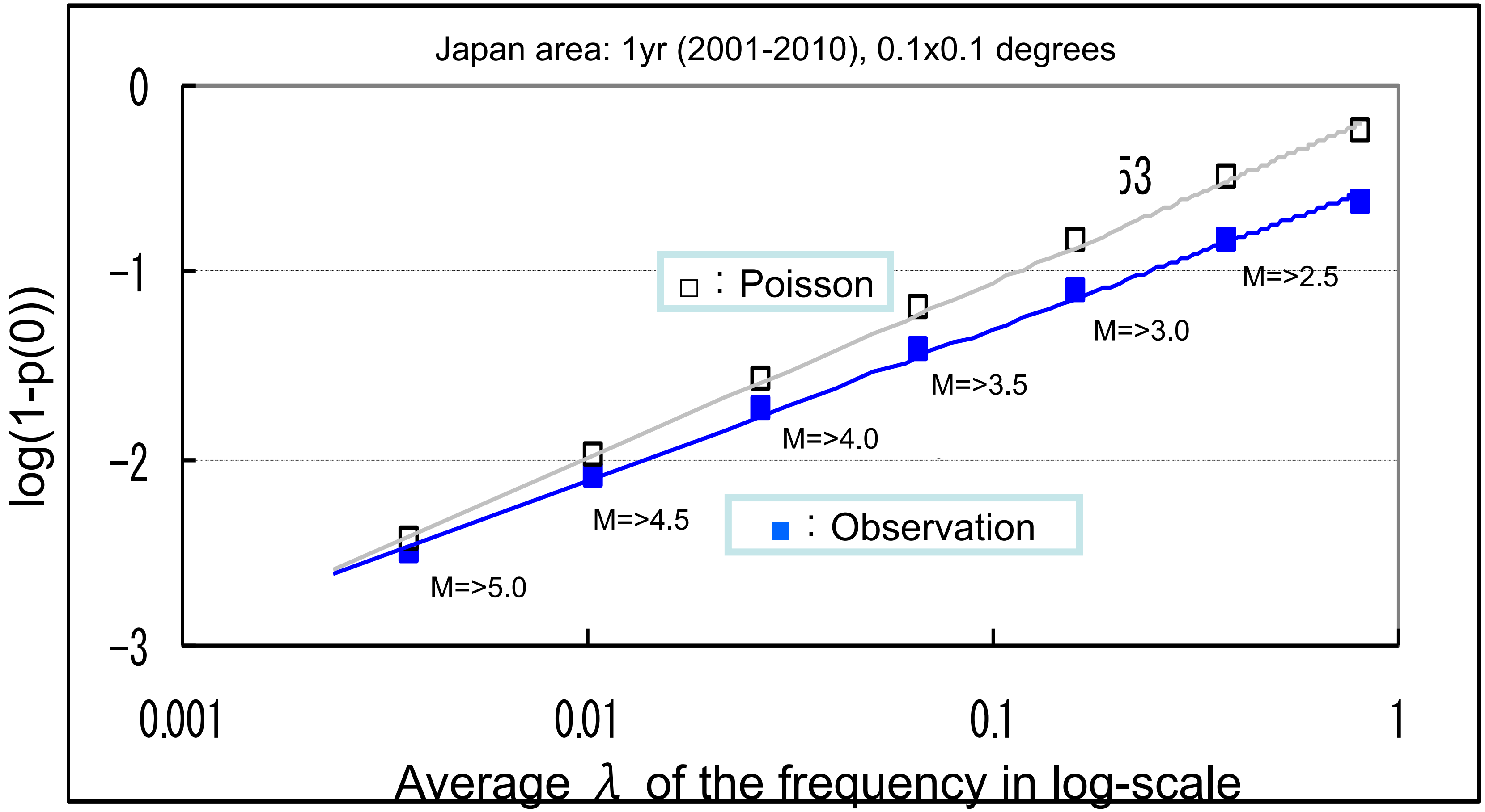
Second tentative comments

When an average λ of the frequency is less than 1:

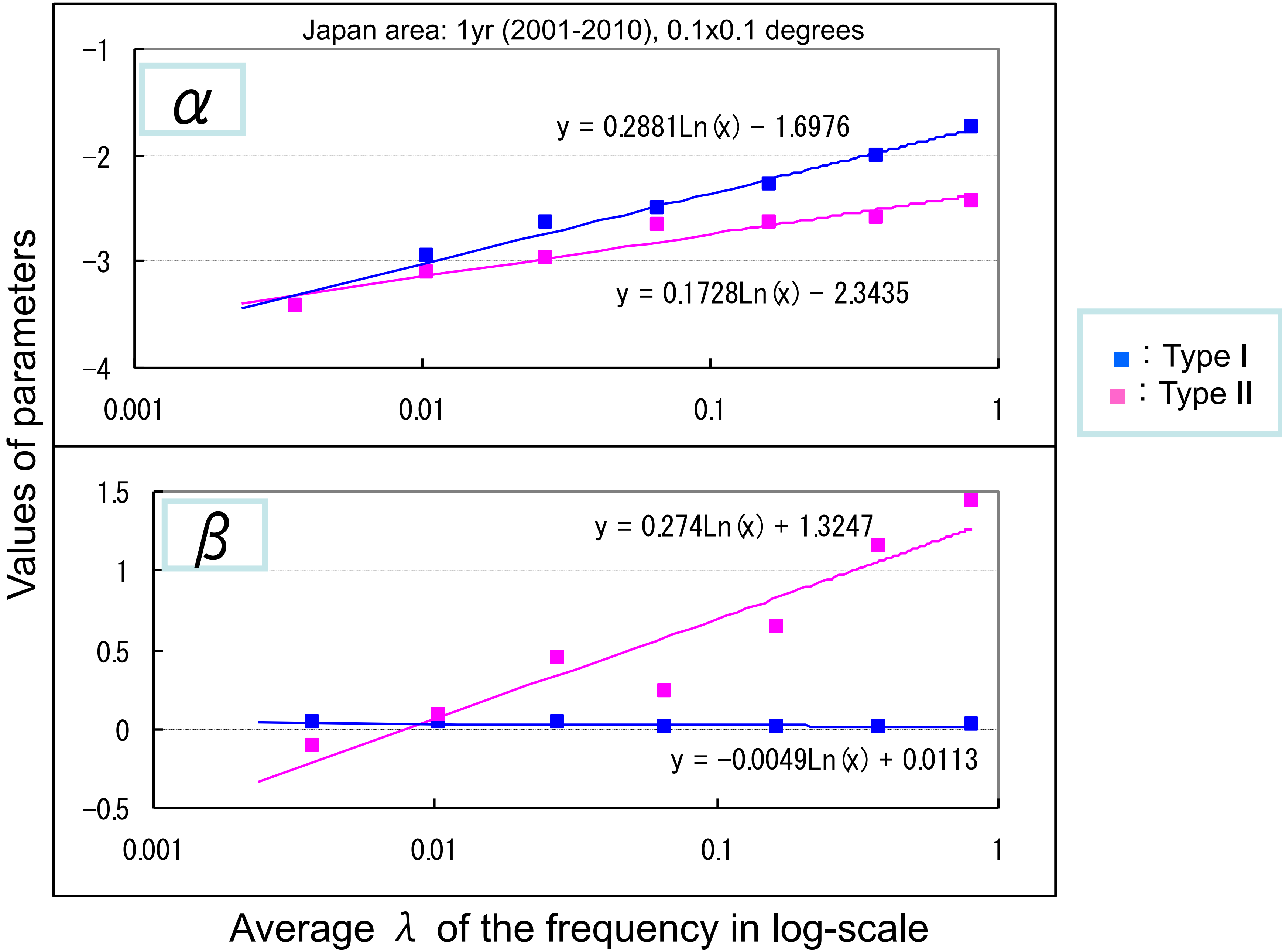
1) Not only Poisson, but also Negative binomial can not explain the observations.

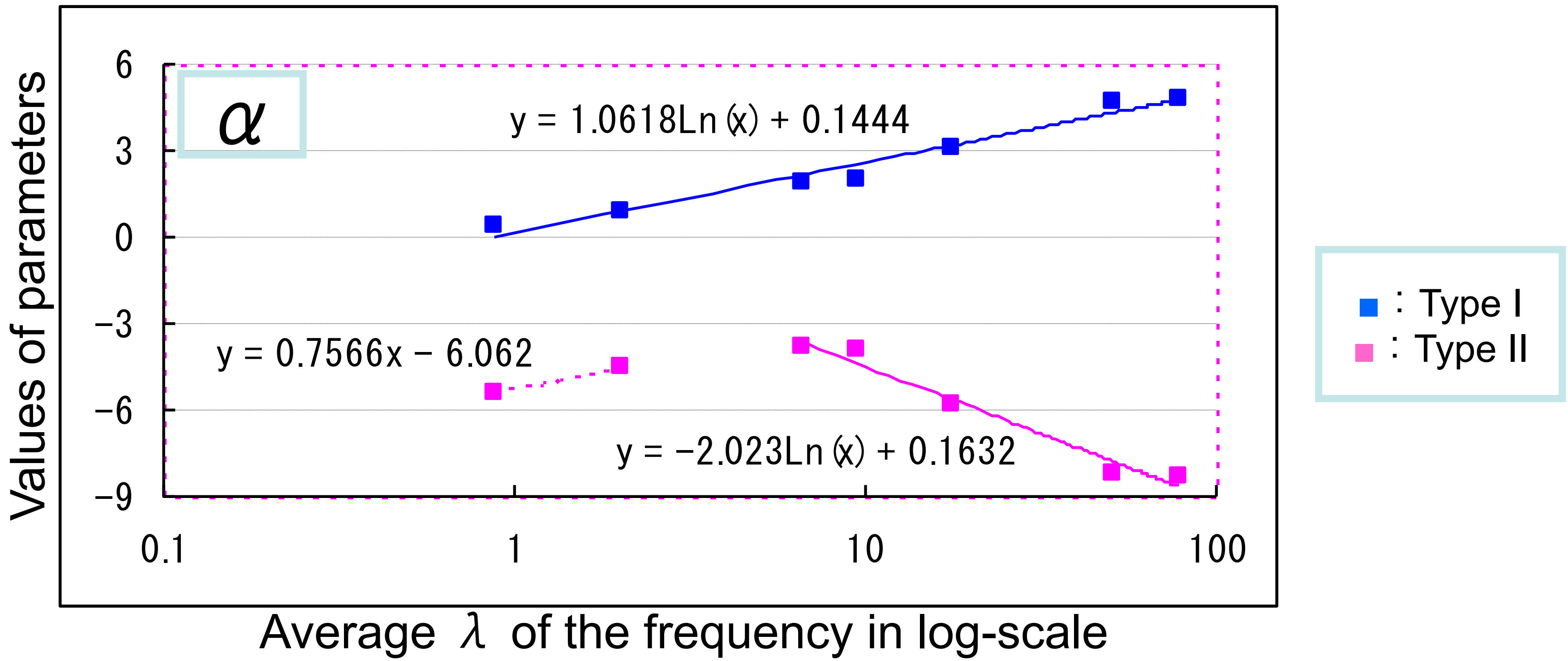
2) Types I and II explain the observations well.

3) Type II is sometimes better than Type I.



For each data set, fraction p(0) can be obtained
if the value of an average is known.


Japan area: 1yr (2001-2010), 0.1x0.1 degrees
α
y = 0.2881Ln(x) − 1.6976
y = 0.1728Ln(x) − 2.3435
−1
−2
−3
−4
0.001
0.01
0.1
1
β
y = 0.274Ln(x) + 1.3247
y = −0.0049Ln(x) + 0.0113
1.5
1
0.5
0
−0.5
0.001
0.01
0.1
1
Values of parameters
Average λ of the frequency in log-scale
: Type I
: Type II



In case of a large average, a trend reverses in Type II function.

Conclusions

- Type I $$p(x) = c\, x^{\alpha} \exp(-\beta\, x)$$ ← The second best

- Type II $$p(x) = c\, x^{\alpha} \exp(-\beta / x)$$ ← Recommendation

1) These functions explain the observed data well.

2) Parameters of these functions can be obtained, if an average of the frequency is known.

3) Referring to the inferred parameters, forecast test in the CSEP project will become more reasonable.