Annual Maximum Rainfall Data Analysis using Extreme Value Theory

For the Boston University - Keio University Workshop

August 18th, 2016

By Ayato Kashiyama

School of Fundamental Science and Technology
Graduate School of Keio University

Table of Contents

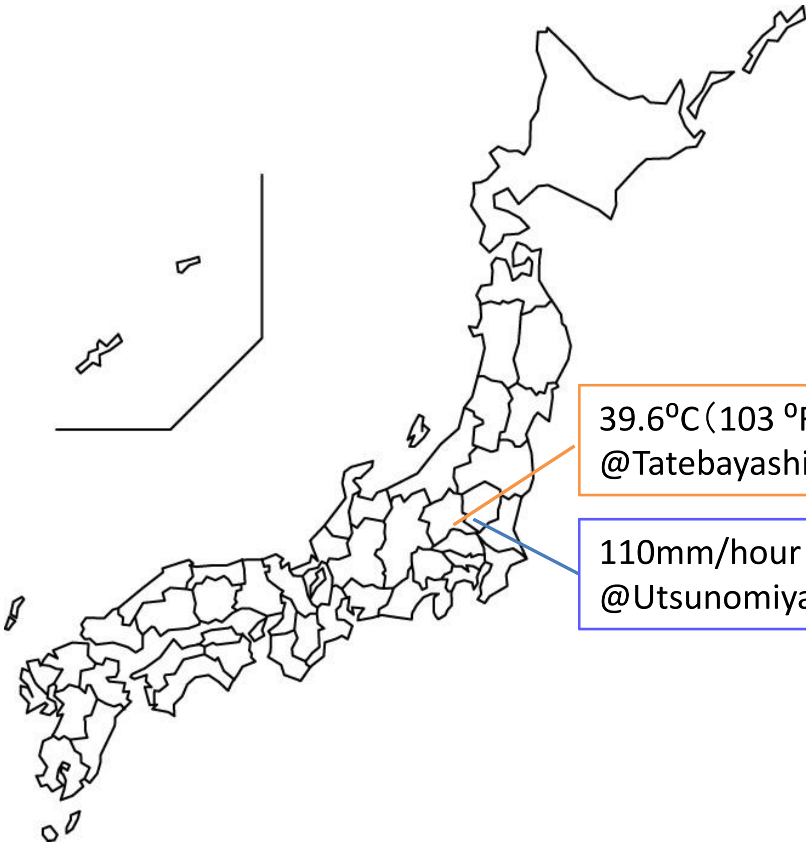
1. Introduction

- Background
  - In recent years, severe weather occurs frequently in Japan, and we often encounter the occurrence of natural disasters
  - The occurrence of natural disaster will increase in the future

Observational data and computer forecasting models suggest

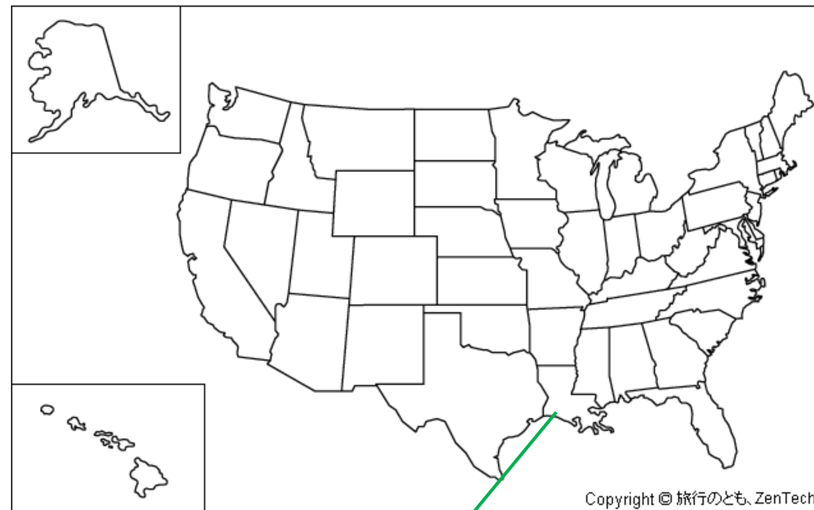
It is important to assess their possible consequences.

1. Introduction



39.6°C（103 °F）
@Tatebayashi (8/17)

110mm/hour
@Utsunomiya (8/18)



In Louisiana, severe rain occurred, and many people evacuated. (8/15-)

1. Introduction

- Background
  - In recent years, severe weather occurs frequently in Japan, and we often encounter the occurrence of natural disasters
  - The occurrence of natural disaster will increase in the future

Observational data and computer forecasting models suggest

It is important to assess their possible consequences.

1. Introduction

- Extreme Value Theory / Analysis
  - Aims at modeling maximum or minimum data
  - In meteorological data, such data corresponds when natural disaster occurs

"Large (or small) random numbers"

||

Extreme Value Data

  - Estimate the value that is more extreme than any that have been already observed (=extrapolate).

2. Extreme Value Theory and Models

- We define an extreme value statistics as

$$M_n = \max(X_1, X_2, \ldots, X_n) = \max_{1 \leq i \leq n} X_i$$

where $X_1, X_2, \ldots, X_n$ is a sequence of independent random variables having a common distribution F .

- We now consider the behavior of M_n when sample size n is large
- In extreme value modeling, we want to estimate the behavior of distribution F^n

*Theoretical distribution of M_n is exactly F^n.

2. Extreme Value Theory and Models
- Definition of Extreme Value Distribution

Definition

If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ such that

$$\Pr\left\{\frac{M_n - b_n}{a_n} \leq z\right\} = F(a_n z + b_n)^n \to G(z) \quad \text{as} \quad n \to \infty \qquad (1)$$

where $G(z)$ is a non-degenerate distribution function.
Then $G(z)$ is called *Extreme Value Distribution*.

Standardization with a_n and b_n appears natural since

$$F^n \to \begin{cases} 0 & x < x^* \\ 1 & x \geq x^* \end{cases} \qquad \text{as} \quad n \to \infty$$

where $x^* := \sup\{x : F(x) < 1\} \leq \infty$

2. Extreme Value Theory and Models
- Extreme Types Theorem

Theorem (Extreme Types Theorem)
[Fisher and Tippet (1928), Gnedenko (1943), Coles(2001)]

Extreme value distribution $G(z)$ defined by (1) is only the following non-degenerate distribution:

$$G(z) = \exp\left\{ -\left[1 + \xi \left(\frac{z-\mu}{\sigma} \right) \right]^{-1/\xi} \right\}$$

defined on the set $\{z : 1 + \xi(z-\mu)/\sigma > 0\}$
where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$

Generalized Extreme Value (GEV) Distribution

$\xi < 0$: Frechet Distribution
$\xi > 0$: Weibull Distribution
$\xi \to 0$: Gumbel Distribution

2. Extreme Value Theory and Models
- Regression Analysis

By using parameter $\boldsymbol{\beta}$ and design matrix (covariance) $\boldsymbol{X}$,
we consider the regression analysis using extreme value distributions.

$\Downarrow$

$\boxed{\text{Able to include temporal trend or other information to the model}}$

For $\mathrm{GEV}(\mu(\boldsymbol{X}), \sigma(\boldsymbol{X}), \xi(\boldsymbol{X}))$,

$$\theta(\boldsymbol{X}) = h(\boldsymbol{X}\boldsymbol{\beta})$$

where θ denotes μ, σ or ξ , h is a specified function called link function.

*Extreme value distribution doesn't belong to exponential family.

2. Extreme Value Theory and Models
- Return Period/ Return Level

- In the extreme value analysis, it is very important to quantify the occurrence of particular (rare/ catastrophic) events
- The value is obtained based on pointwise estimation of high quantile of distribution

→ Return Level

For annual maximum data z, T year return level is obtained by

$$G(z_T) = 1 - 1/T$$

When defined above, z_t is the return level associated with the return period T.

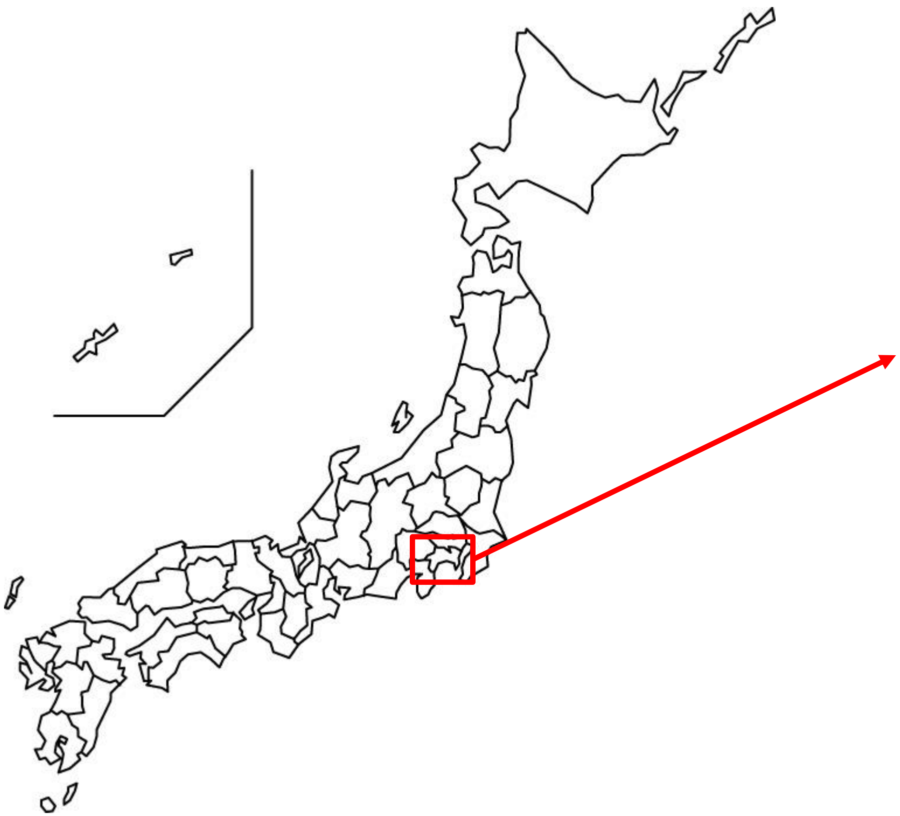
Return level means

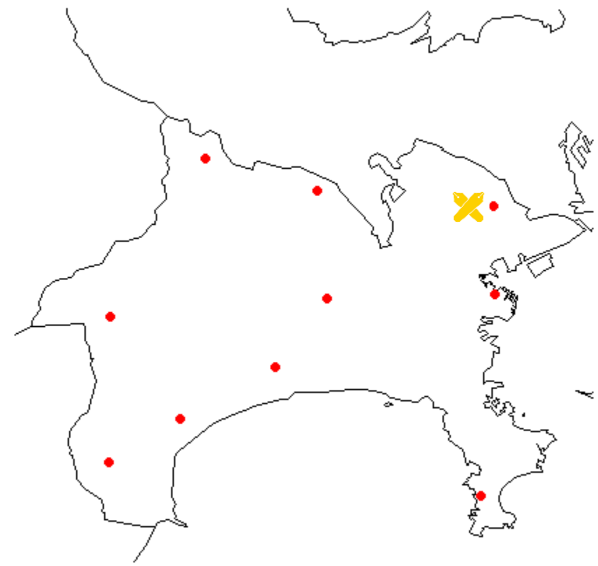
the value that is expected to be exceeded on average once in T years

(= the value expected to exceed once in 1 year with probability 1/T)

3. Annual maximum rainfall data analysis

- Data explanation





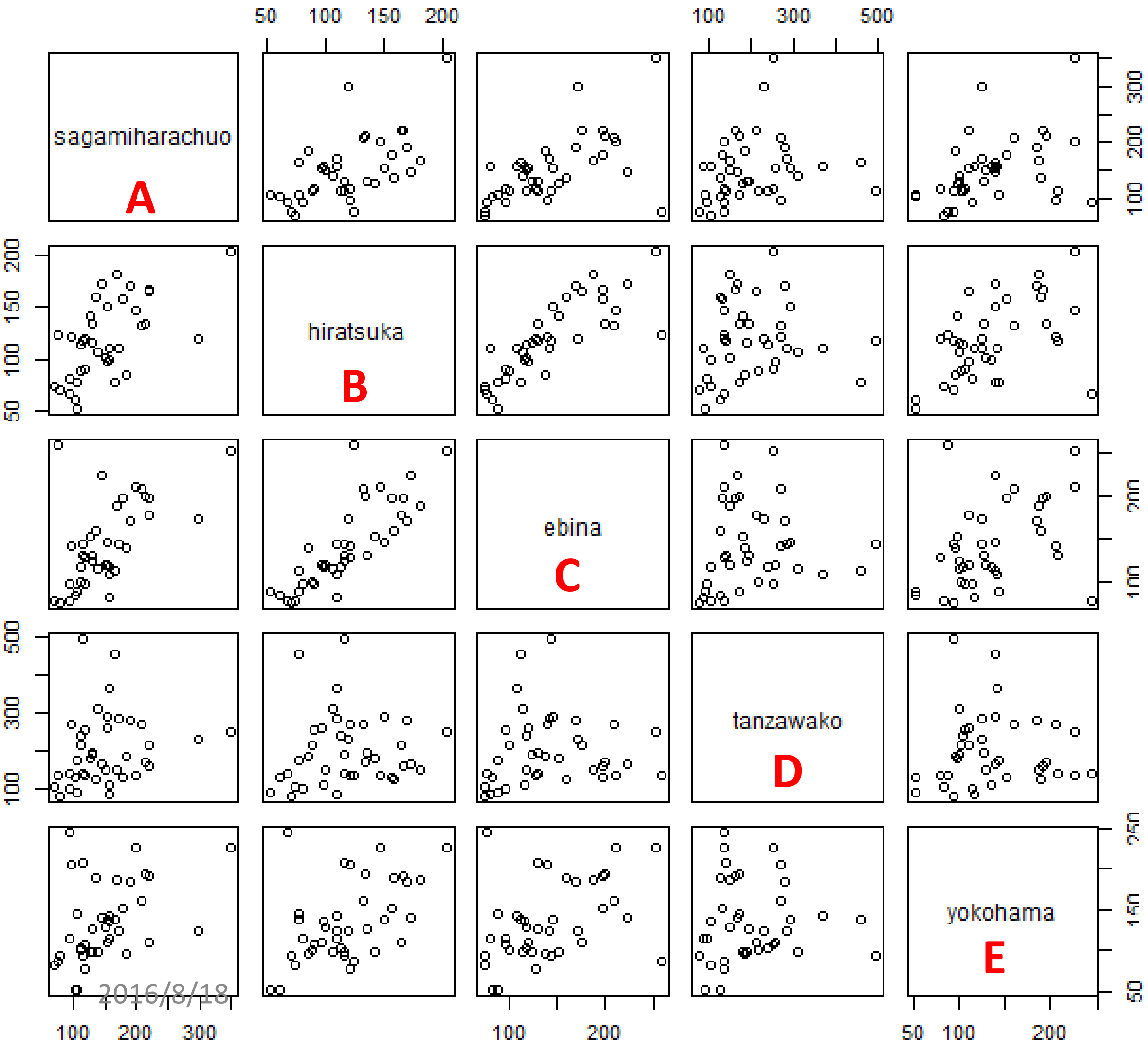
Kanagawa prefecture, Japan

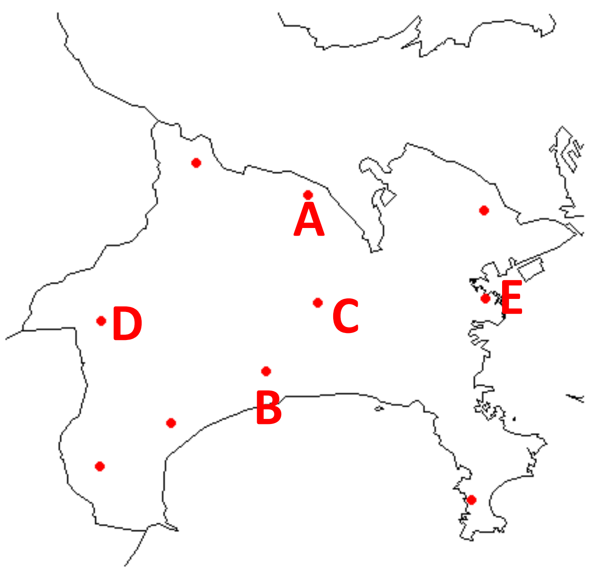
- South and east sides face to the sea
- West region is mountainous district

- 10 observation sites
- 40 years data (1976-2015) are available

3. Annual maximum rainfall data analysis

- Data explanation





3. Annual maximum rainfall data analysis
- Purpose of data analysis

- Problems of annual maximum rainfall data
  - Because of annual maximum data, the number of available data tends to be small when considering single site analysis
  - Severe rain tends to occur more frequently especially in recent years than before

Aims of our research:

- By considering the models that includes data at multiple sites, we want to compensate for the limited number of observations
- We want to construct the model that includes temporal trend

3. Annual maximum rainfall data analysis
- Analysis Procedure

Analysis Procedure

1. Estimate parameters of GEV distribution with maximum likelihood method $\mathrm{GEV}(\mu, \sigma, \xi)$

2. Consider the model that includes temporal trend in location parameter: $\mu(t) = \beta_0 + \beta_1 t$

3. Perform likelihood ratio test to judge whether β_1 should be included in the model or not

These calculations are done by using R package "ismev".

3. Annual maximum rainfall data analysis
- Model fitting/ evaluation

Single site analysis & temporal trend

Fitting the GEV distribution for each site, the results as follows.

Observation Site	Distribution	Constant Term β_0	Temporal Coefficient β_1	Scale Parameter	Hypothesis test: $\beta_1 = 0$
Ebina(C)	Gumbel	112.62	0.3142	37.92	Not Reject
Hiratsuka(B)	Gumbel	90.86	0.4982	30.41	Not Reject
Yokohama(E)	Gumbel	97.10	0.7763	37.86	Not Reject
Sagamihara(A)	Gumbel	112.69	0.6637	39.70	Not Reject
Tanzawako(D)	Gumbel	126.93	1.641	64.04	Not Reject

Hypothesis test for $\beta_1 = 0$ (= no temporal trend) is <u>not rejected</u> for all sites.

⇩

We combine all data and consider a model that includes sites as a factor.

3. Annual maximum rainfall data analysis
- Model fitting/ evaluation

Combined data model

Location / scale parameter: include sites as a factor.
Assume that observations are independent

$$
\begin{aligned}
\hat{\mu} = {} & 107.133 + 80.814(\text{hakone}) - 18.952(\text{hiratsuka}) - 7.463(\text{hiyoshi}) \\
& - 9.416(\text{miura}) + 1.073(\text{odawara}) + 6.068(\text{sagamihara}) + 13.272(\text{sagamiko}) \\
& + 39.717(\text{tanzawako}) - 7.263(\text{yokohama}) + 0.6052t
\end{aligned}
$$

$$
\begin{aligned}
\hat{\sigma} = {} & 38.853 + 42.879(\text{hakone}) - 8.412(\text{hiratsuka}) - 0.32(\text{hiyoshi}) \\
& - 4.528(\text{miura}) - 6.490(\text{odawara}) + 0.652(\text{sagamihara}) + 16.696(\text{sagamiko}) \\
& + 25.738(\text{tanzawako}) - 1.335(\text{yokohama})
\end{aligned}
$$

$$\hat{\xi} = 0(\text{Gumbel distribution})$$

3. Annual maximum rainfall data analysis
- Model fitting/ evaluation

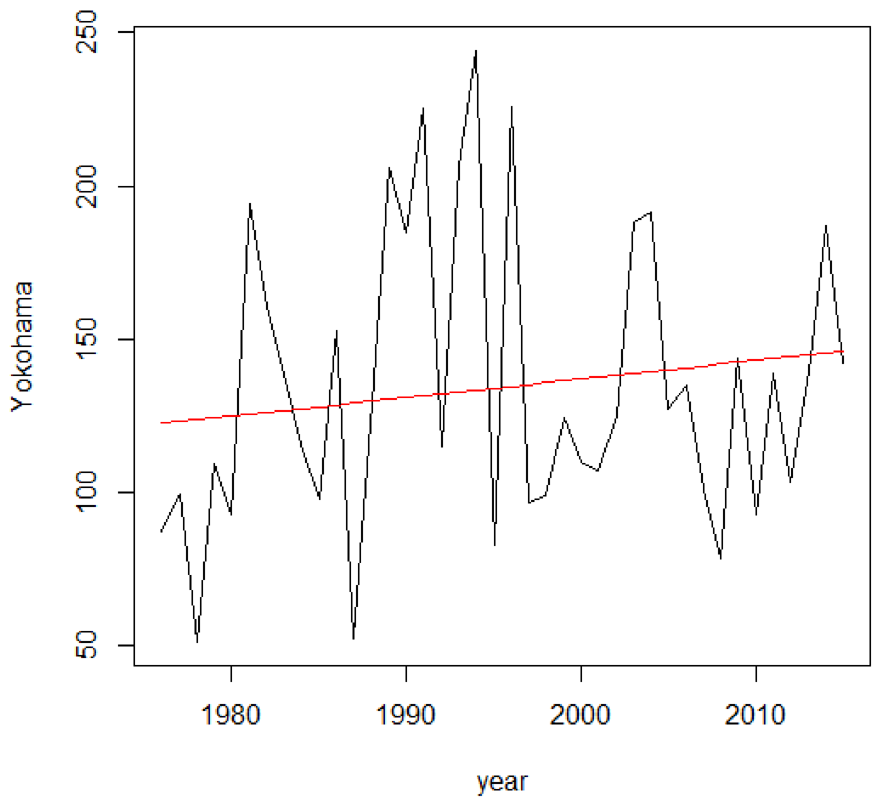
Combined data model Summary table

Observation site	Location Parameter	Std. Error	Scale Parameter	Std. Error
Ebina	107.133	7.437	38.853	5.142
Hakone	187.947	15.011	81.732	11.962
Hiratsuka	88.181	8.223	30.441	6.393
Hiyoshi	99.669	9.103	38.533	7.126
Miura	97.716	8.619	34.325	6.672
Odawara	108.205	8.419	32.364	6.473
Sagamihara	113.200	9.202	39.505	7.204
Sagamiko	120.405	11.298	55.549	8.763
Tanzawako	146.850	12.511	64.591	9.799
Yokohama	99.869	8.992	37.518	6.915
(Temporal Trend)	0.60521	0.174		

Hypothetical test for no temporal trend is rejected.

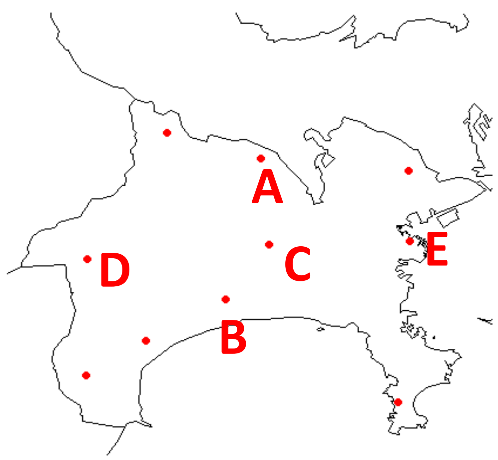
We should include temporal trend.

3. Annual maximum rainfall data analysis
- Model fitting/ evaluation



From the red line,
temporal trend (0.6 increase per year) is reasonable.

Black line: Observed value
Red line : Temporal Trend by linear fitting

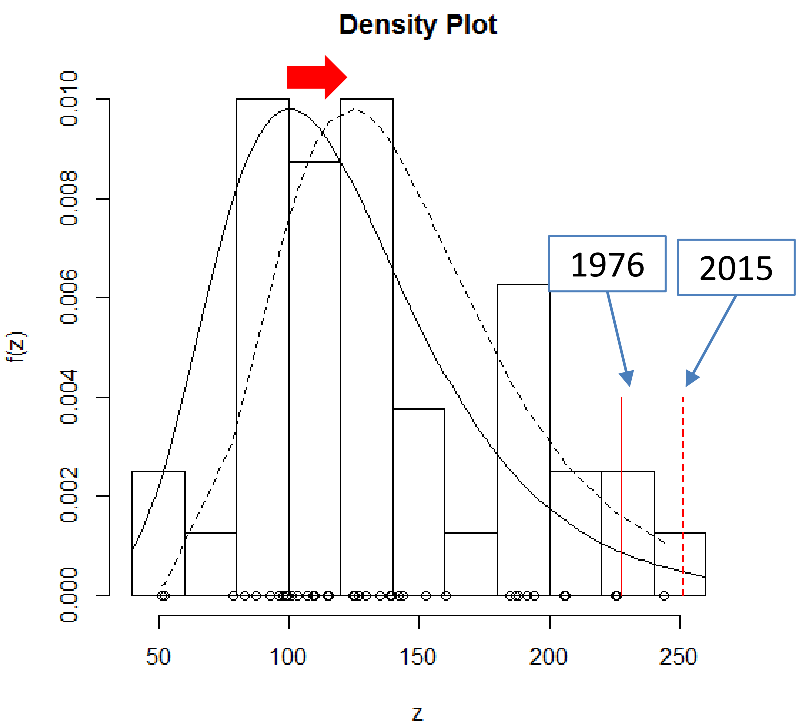


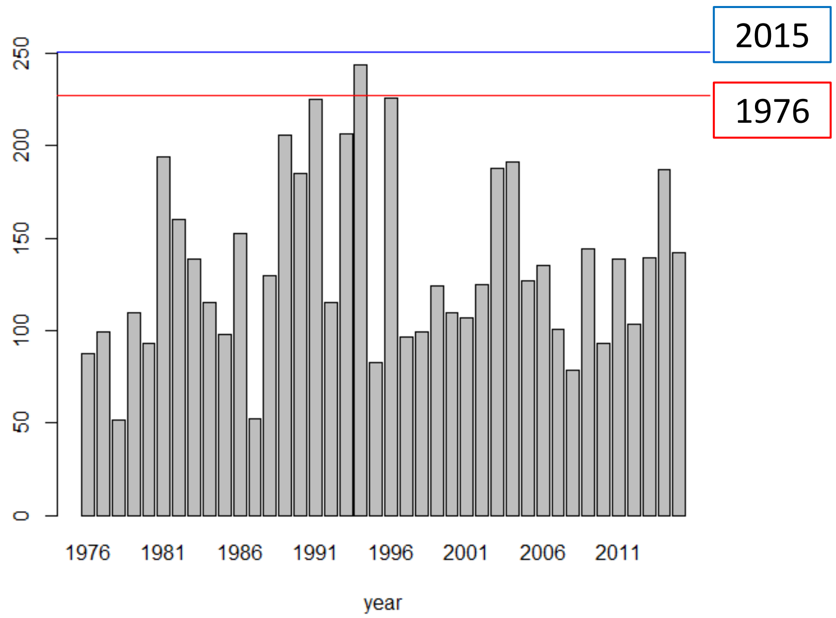
3. Annual maximum rainfall data analysis -Evaluation of return level

Based on the structural model , we consider the return level at Yokohama.

30 year return level by distribution of 1976 is 227.45 mm
30 year return level by distribution of 2015 is 251.05 mm

About 24mm increase





4. Summary & Future Work

- As a statistical modeling of extremes, I introduced extreme value theory and data analysis method
- By using the theory, I showed the result of annual maximum daily rainfall data analysis in Kanagawa prefecture

4. Summary & Future Work

From the result,

- The model of single site analysis cannot capture the temporal trend
- The model that includes data of all sites in Kanagawa can capture the temporal trend
  - This time, we consider the model for overall data that include sites as a factor and assume that observations are independent

- By considering the temporal trend, return level is increased about 24mm in 40 years

4. Summary & Future Work

- Future Work

Observations are assumed to be independent

(=there is no dependence) : Unrealistic assumption

⇩

Constructing the dependent model

e.g) Multivariate Extreme Value Distribution

Copula Model

⇩

Compare the result of independent model

Calculate the return level with dependent model

References

- Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes, Theory and Applications,* Wiley.
- Coles, S, G. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer.
- Finkenstadt. B., and Rootzen, H. (2004). *Extreme Values in Finance, Telecommunications, and the Environment,* Chapman& Hall.
- Fisher, R.A. and Tippett, L.H.C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.* **24**, 180-190.
- Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Ann.Math.* **44**,423-453.
- Leadbetter, M. R., Lindgren, G. and Rootzen, H. (1983). *Extremes and Related Properties of Random Sequences and Process*. Springer-Verlag, NY.

Thank you for your attention!