HydroPredict 2010, 2nd International Interdisciplinary Conference on
Predictions for Hydrology, Ecology, and Water Resources Management:
Changes and Hazards Caused by Direct Human Interventions and Climate Change
20–23 September 2010; Prague, Czech Republic

Invited Paper

Non-Parametric Frequency Analysis of Hydrological Extreme Events

2010/09/21

Kaoru TAKARA and Kenichiro KOBAYASHI
Disaster Prevention Research Institute
Kyoto University, Japan

River planning and design of flood control facilities

are based upon hydrological prediction (HydroPredict !! ☺):

- **Design rainfall / design flood**
- ***T*-year return period**
  (*T*= 5, 10, 20, 30, 50, 100, 150, 200, …)
- **Frequency analysis**
- **Probabilistic/stochastic approaches**
- **Stationary or non-stationary?**

DPRI-KU

Today's talk

Part-1: Brief review of "Parametric" method

Part-2: "Non-parametric" method for long-term extreme-value series combined with the bootstrap resampling -- New idea

Part-3: Trend analysis and How to consider Climate Change effect

- Non-parametric method with unequal occurrence probability -- New idea

Part-1: Brief review of "Parametric" method

HYDROLOGIC FREQUENCY ANALYSIS

Traditional "Parametric" Approach Using Probability Distribution Functions (PDFs)

Frequency analysis of hydrological extreme events

Hydrological variable
(Extreme event)

$$X$$

Its realizations
(Observed data)

$$x_1, x_2, \cdots, x_N$$

Issues of frequency analysis

(1) data characteristics (homogeneity, independence),

(2) sample size (effect of years of records on accuracy and appropriate estimation method),

(3) parameter estimation (selection of parameter values of distribution functions),

(4) model evaluation (selection of a distribution), and

(5) accuracy of quantile estimates (unbiasedness, estimation error).

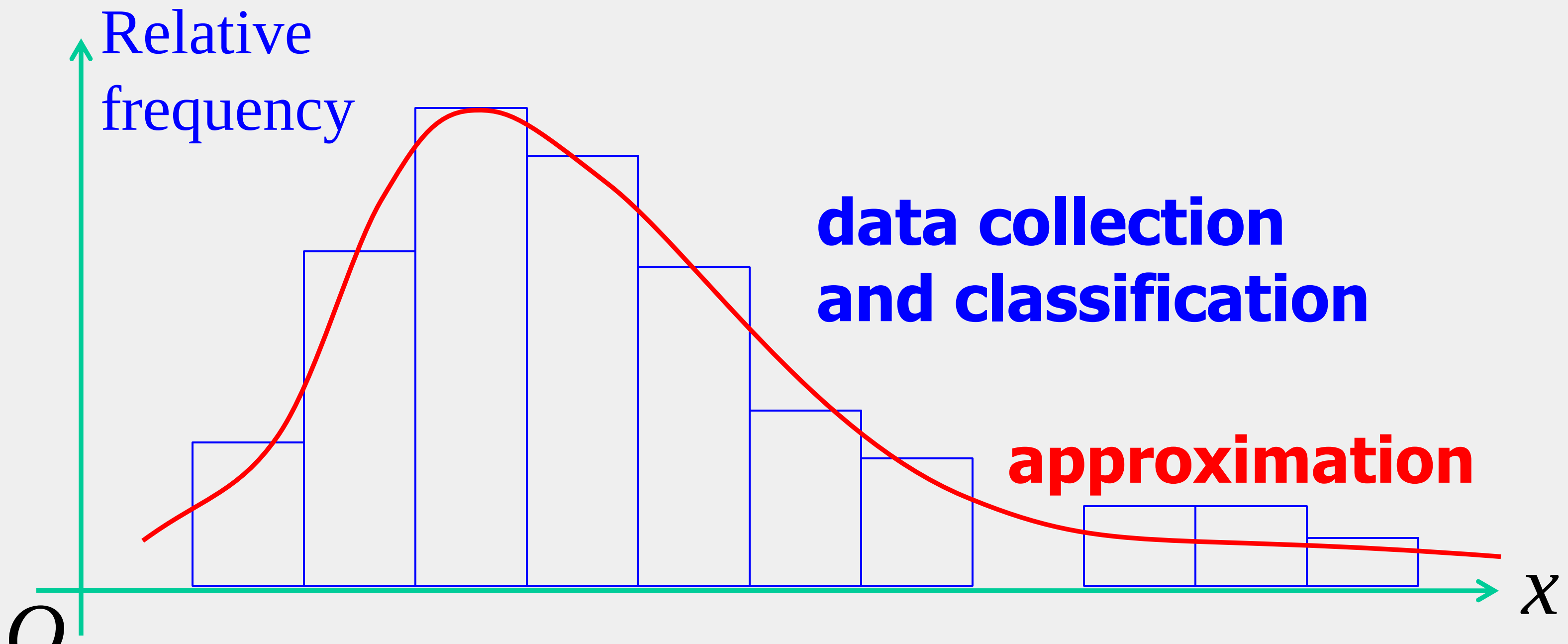
Non-exceedance probability

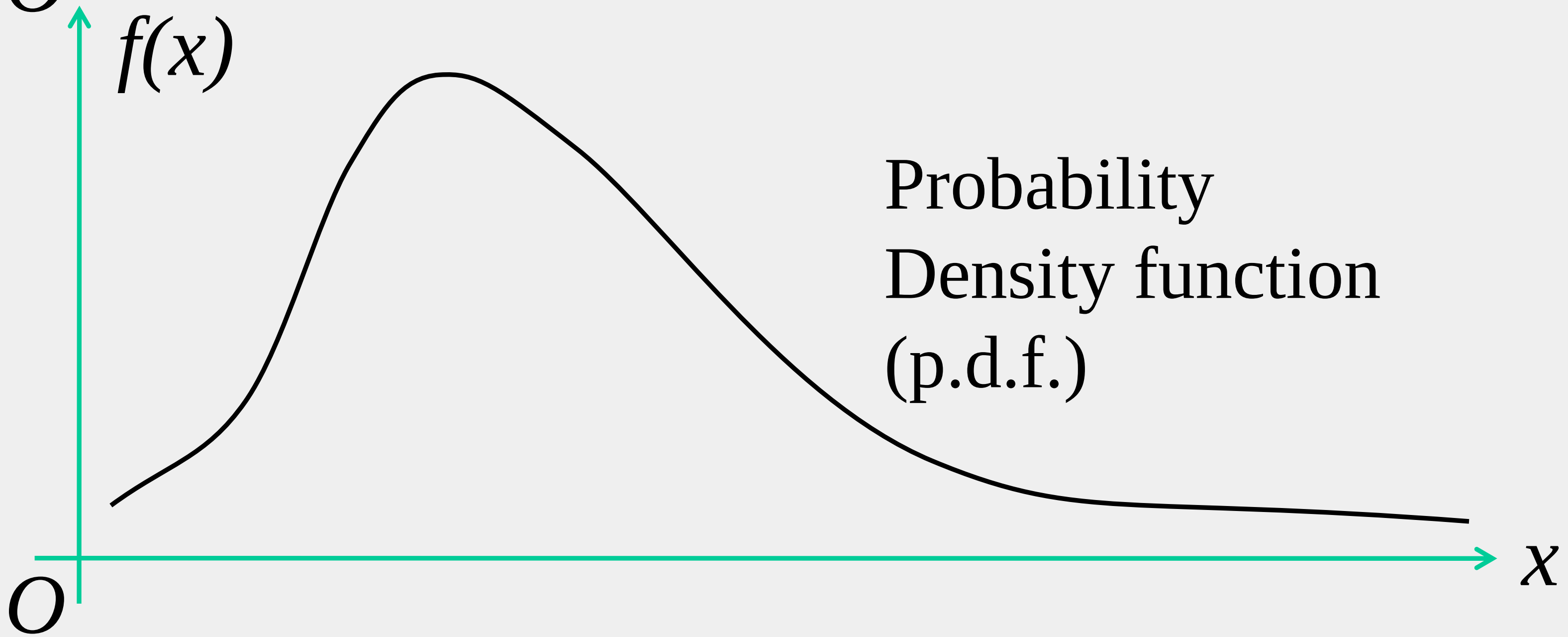
$$F_X(x) = P\left[X \leq x\right]$$

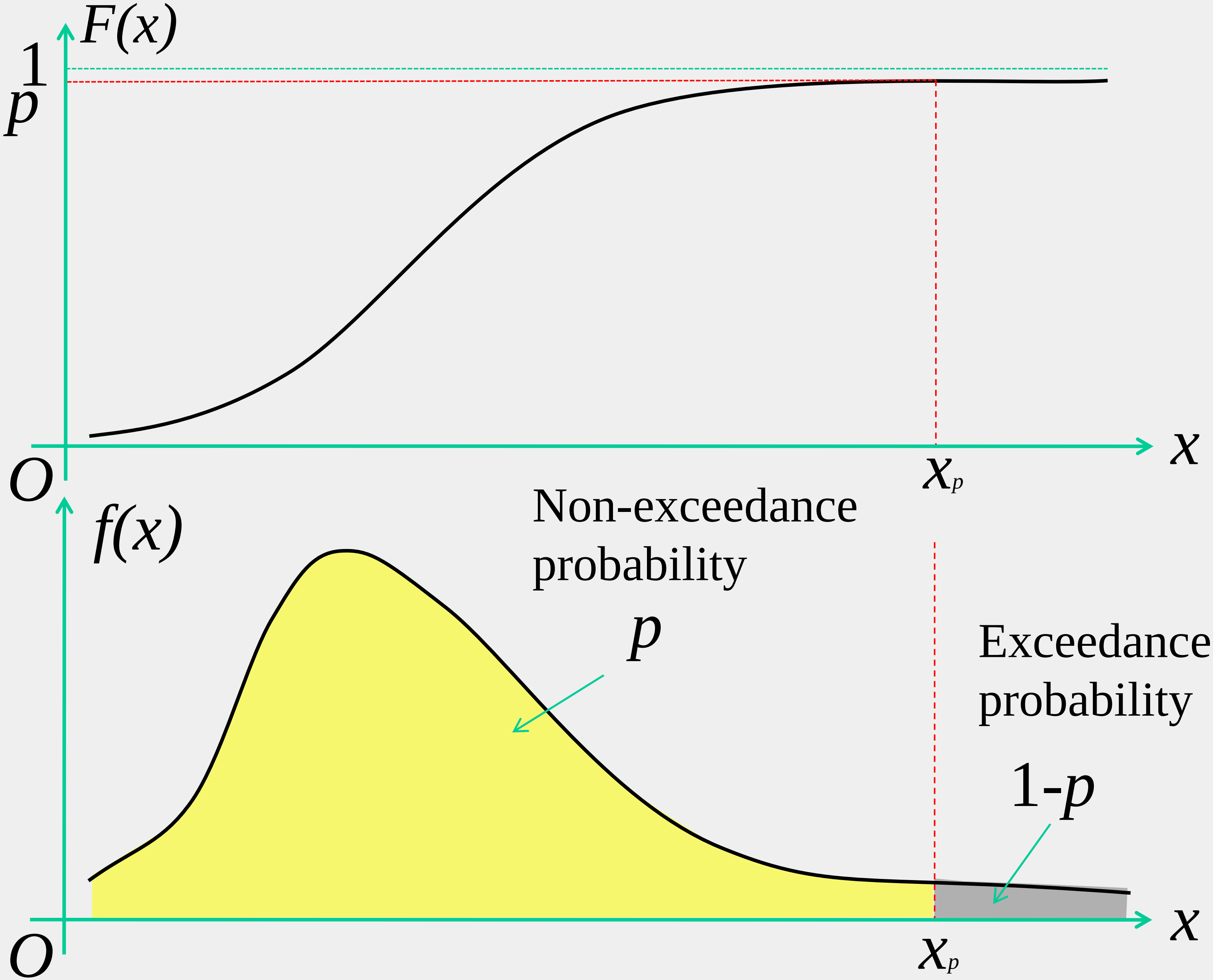
$$F(x) = \int_{-\infty}^{x} f(t)dt$$

F : cumulative probability distribution function

f : probability density function


Relative
frequency
data collection
and classification
approximation
x
O


f(x)
Probability
Density function
(p.d.f.)
x
O


F(x)
1
p
O
x
x_p
f(x)
Non-exceedance probability
p
Exceedance probability
1-p
O
x
x_p

LN(3) and GEV(3)

$$f(x) = \frac{1}{(x-c)\sigma_Y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{\ln(x-c)-\mu_Y}{\sigma_Y}\right\}^2\right]$$

$$F(x) = \exp\left[-\left(1-k\,\frac{x-c}{a}\right)^{1/k}\right]$$

Probability distribution *F*, non-exceedance probability *p* and return period *T*

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

$$p = F(x_p)$$

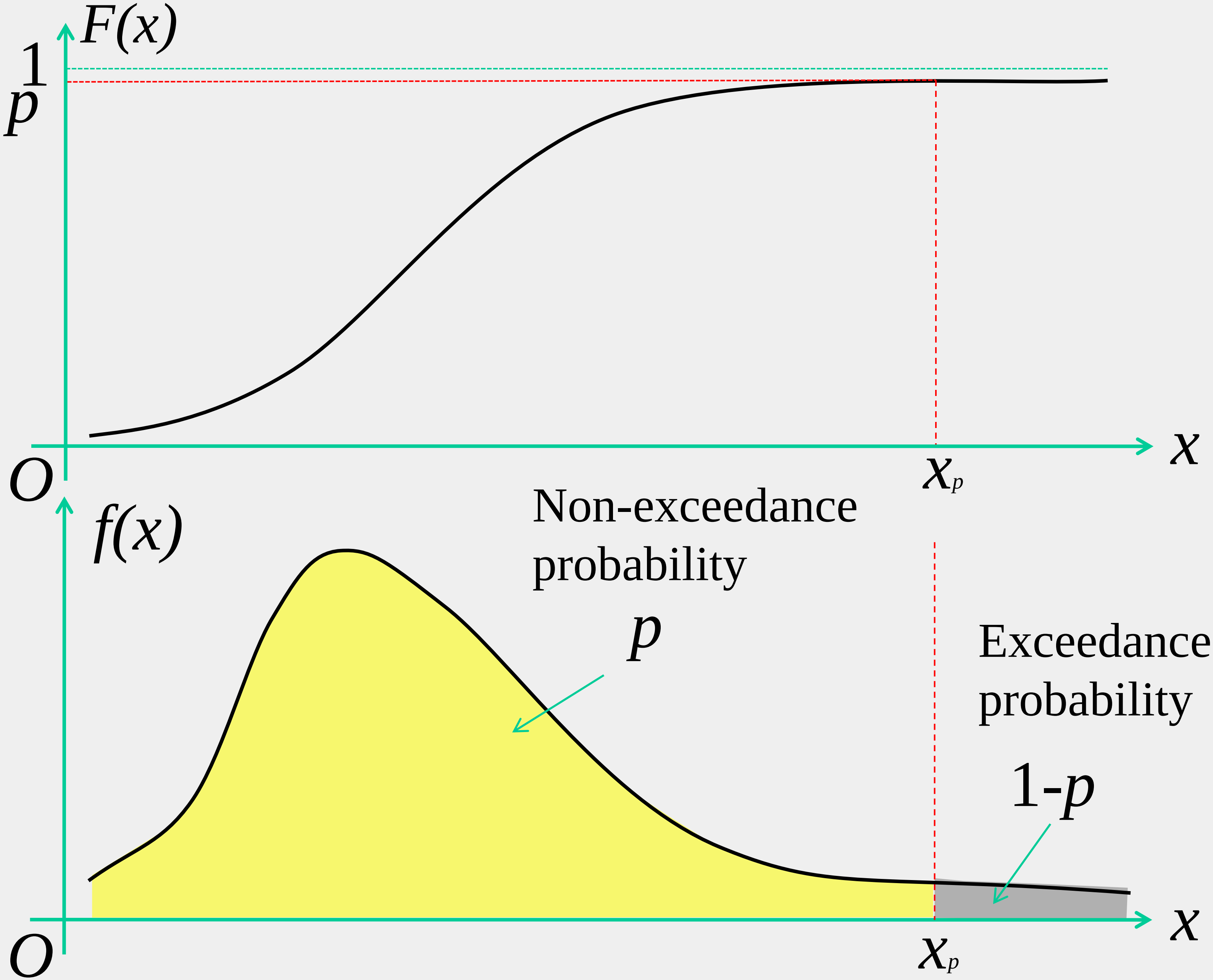
$$x_p = F^{-1}(p)$$

$$T = \frac{1}{n(1-p)}$$

For annual maximum series, n=1:

$$T = \frac{1}{1-p}$$

DPRI-KU

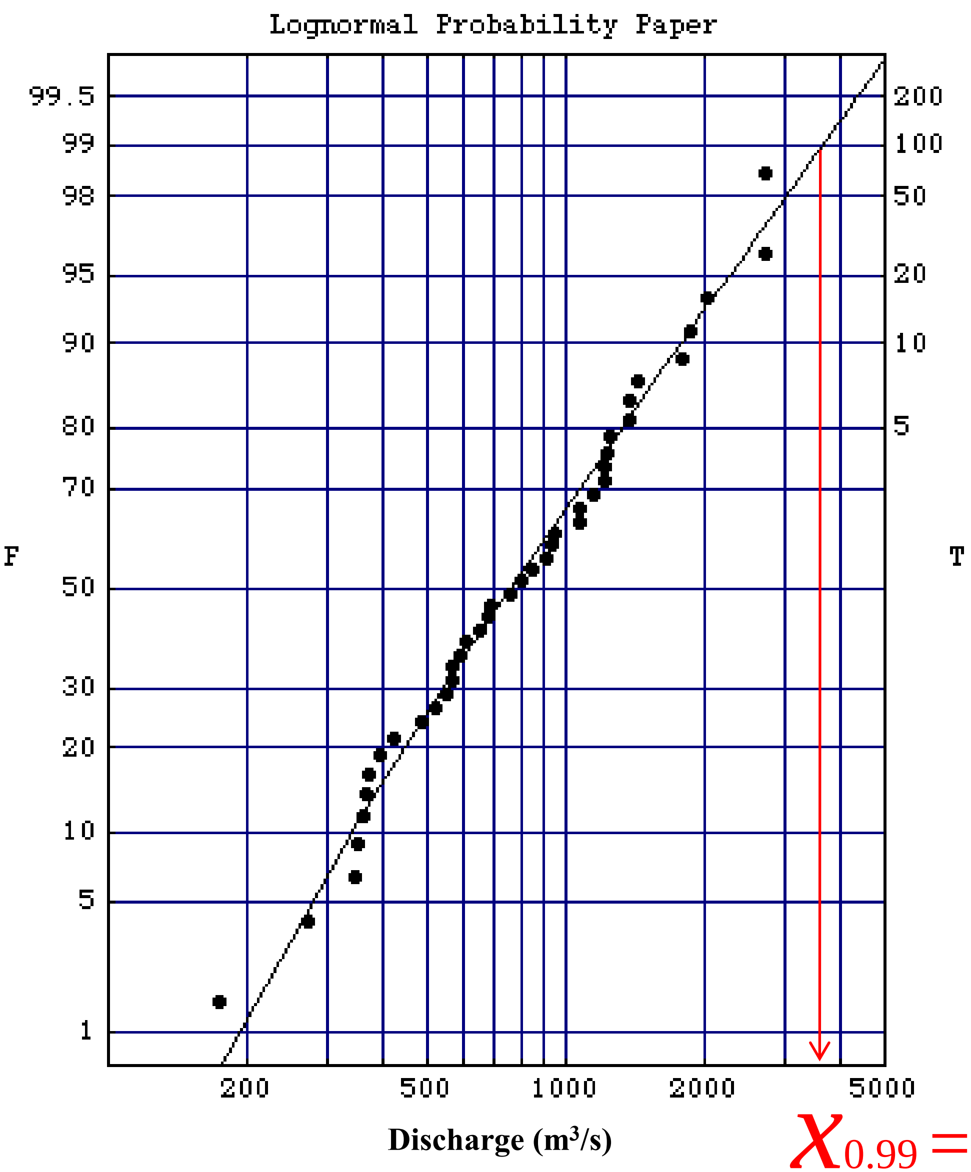


F(x)
1
p
O
x
x_p
f(x)
Non-exceedance probability
p
Exceedance probability
1-p
O
x
x_p

Non-exceedance probability *p* and return period *T*

$$T = \frac{1}{1-p}$$

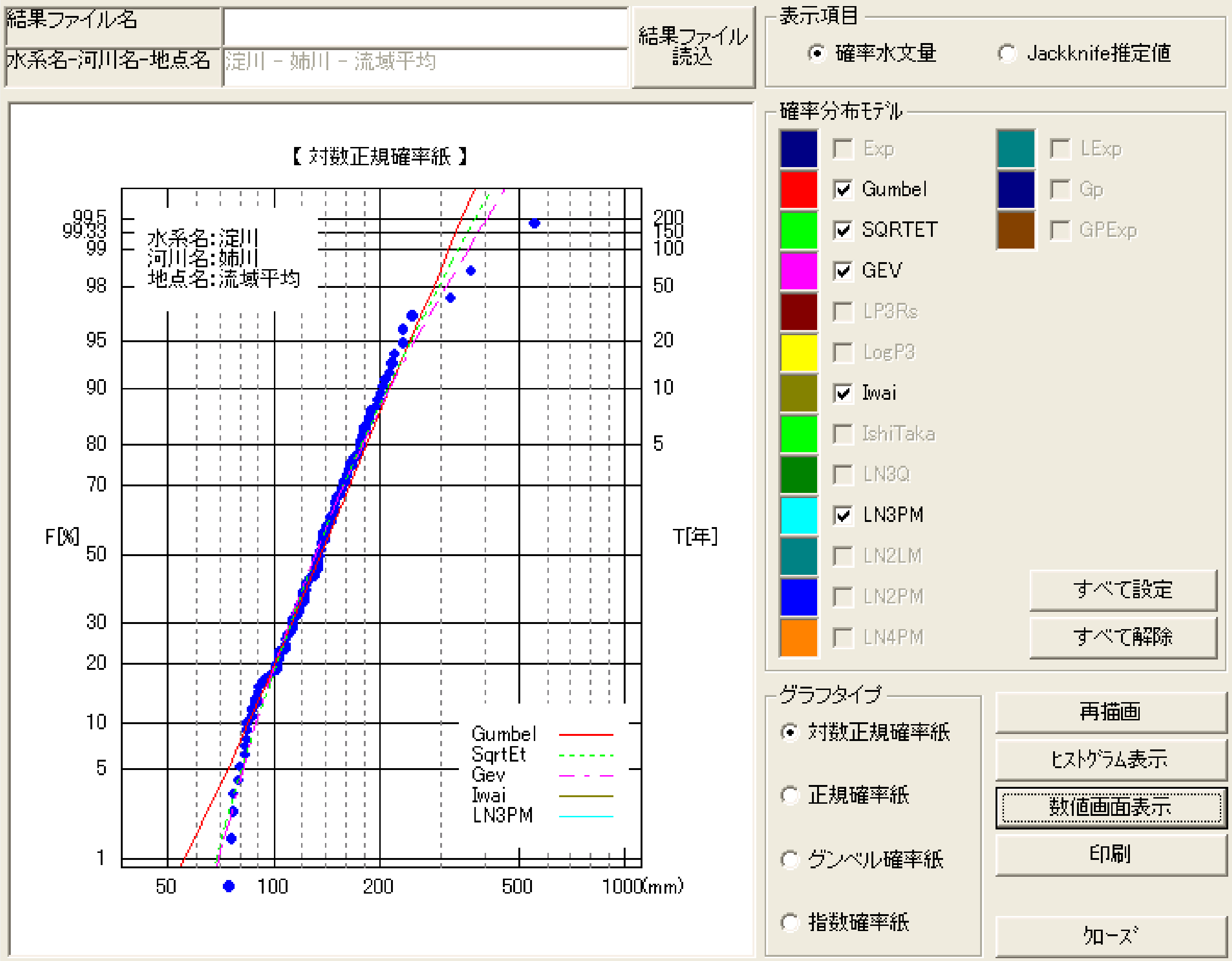
T (year)	*p*
2	0.5
10	0.9
20	0.95
50	0.98
100	0.99
200	0.995



Example of fitting discharge data to lognormal (LN) distribution

$X_{0.99}$ = 3600 m3/s

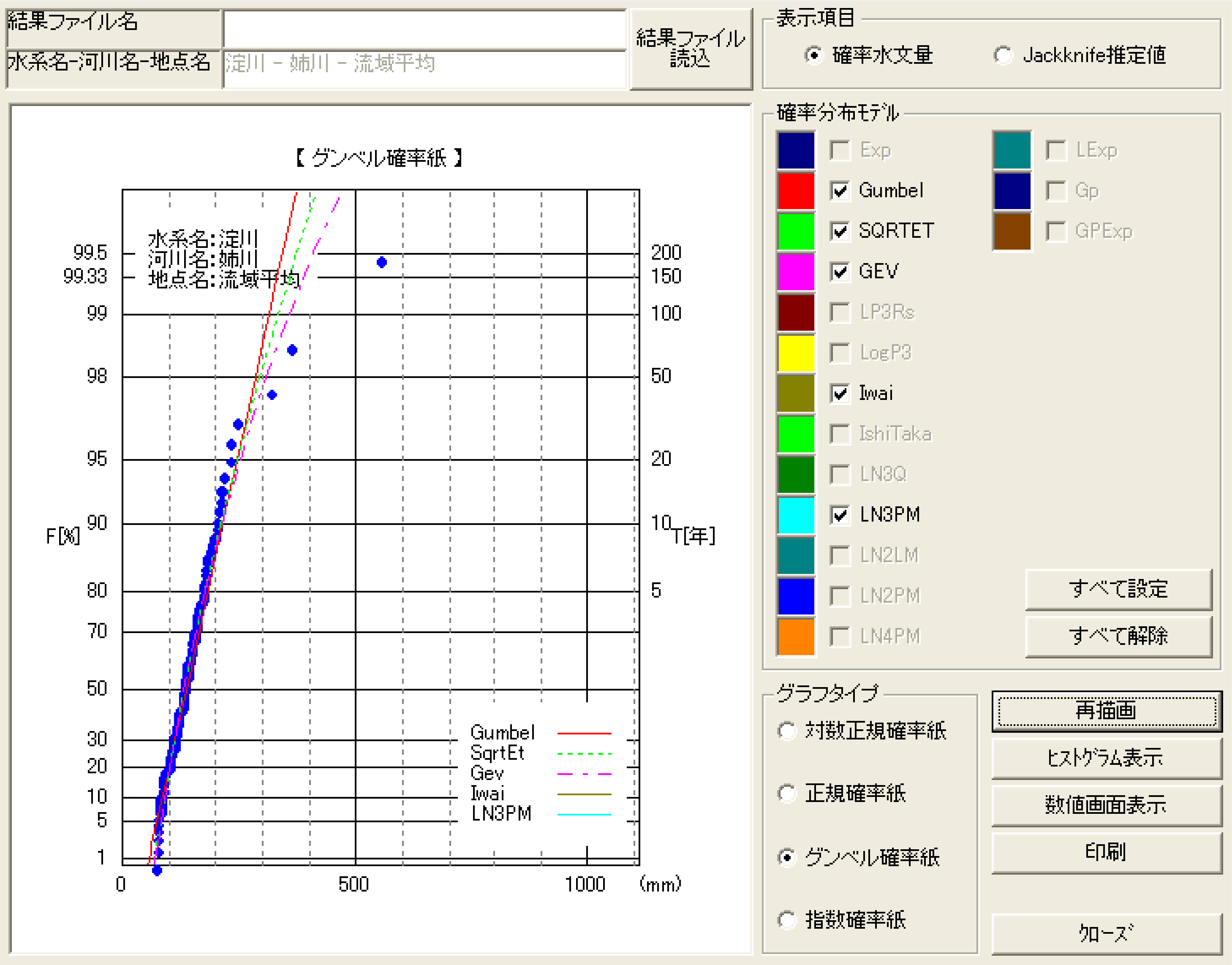
Hydrological Statistics Utility Ver. 1.5



LN Prob. paper

See JICE (Japan Institute of Construction Engineering)
http://www.jice.or.jp/sim/t1/200608150.html

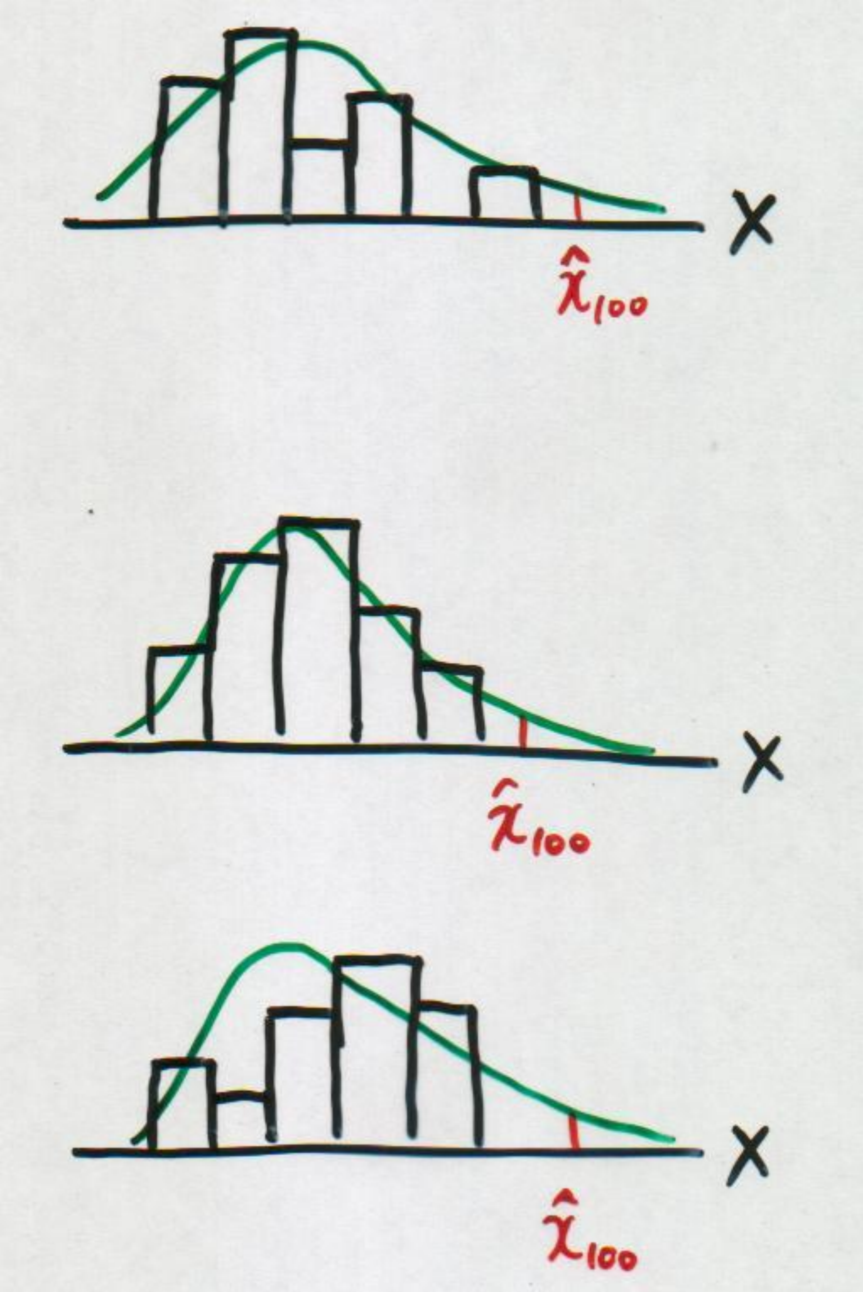
Gumbel probability paper (JICE)



http://www.jice.or.jp/sim/t1/200608150.html

T-year 2-day rainfall at the Anegawa river basin with three different distributions

T	GEV	Gumbel	SQRTET
20	252.0	245.1	245.5
30	276.2	262.5	266.6
50	308.8	284.3	294.1
100	357.1	313.7	333.1
150	387.7	330.9	356.9
200	410.7	343.0	374.2



Quantile estimates (T-year events)

Depend on:

- **Combination of data,**
- **Number of data (sample size),**
- **Probability Distribution,**

and

- **Parameter estimation method Used**

Today's talk

Part-1: Brief review of "Parametric" method

Part-2: "Non-parametric" method for long-term extreme-value series combined with the bootstrap resampling -- New idea

Part-3: Trend analysis and How to consider Climate Change effect

- Non-parametric method with unequal occurrence probability -- New idea

HYDROLOGIC FREQUENCY ANALYSIS

New “Non-Parametric” Approach

No PDFs !

New era for hydrological frequency analysis

- Sample size is getting larger.

  Extreme-value samples with N>100 years.

- How can we consider Climate Change effect ?

  No more return period?

  What is the non-stationary analysis ?

Sample size (Record length)

In 1950's it is often said that hydrological samples have only 10- to 50-year records.

- This was true.

But …

Now we are in 2010.

→We have large samples with N>100 at many locations in the world.

20 years later → much more !!

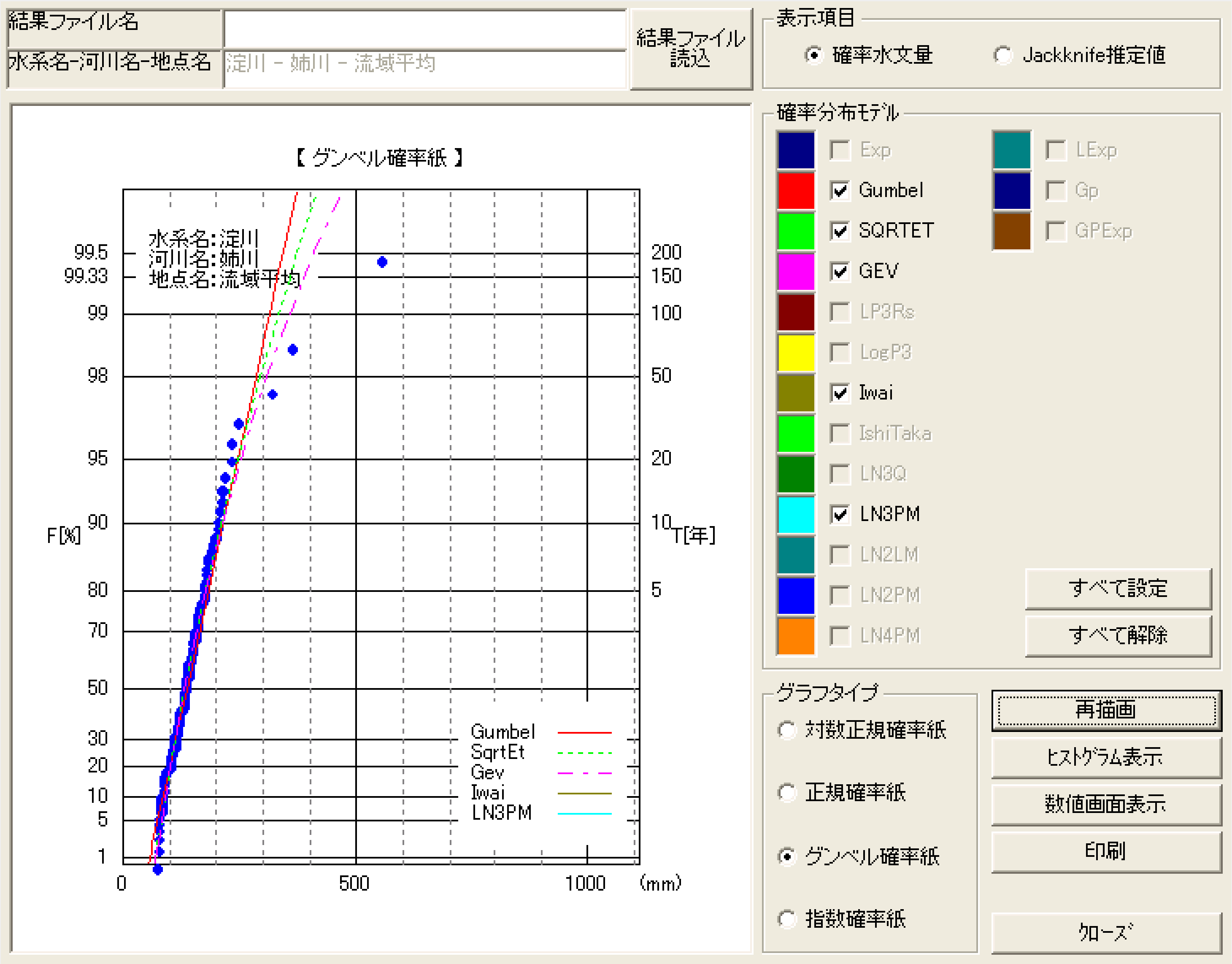
No PDFs → Non-parametric

Proposing

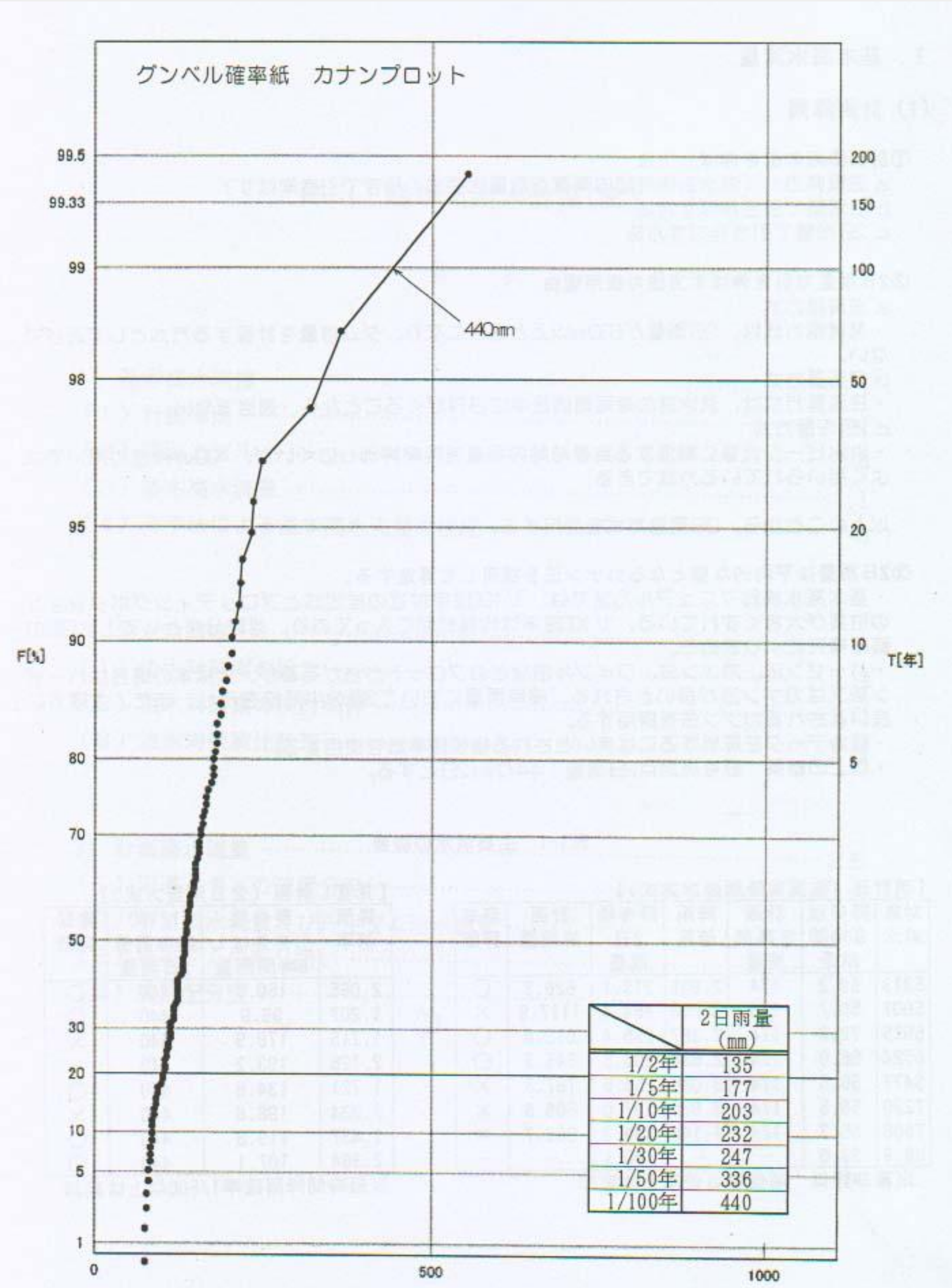
- a non-parametric method estimating 100-year events based on a sample with a size $N>100$.
- To verify the estimation accuracy with a computer intensive statistics (CIS) method: the bootstrap resampling.

DPRI-KU

Gumbel probability paper (JICE)



http://www.jice.or.jp/sim/t1/200608150.html



Plot on the Gumbel probability paper
The Ane-gawa River basin two-day rainfall

- Sample size N=108>100

- Parametric method
  - Gumbel: 313 mm
    (bad fitting)
  - GEV: 357 mm
  - Log-Pearson III: 373 mm
  - SQRT-ET-max: 333 mm

- Graphical method with the empirical distribution
  $x_{0.99}$ = 440 mm

100-year rainfall by parametric methods and graphical method with empirical distribution (yellow)

100-year daily rainfall estimated by parametric methods [*two-day rainfall only for the Ane]

PDF	Gumbel (EV1) distribution		Gumbel (EV1) distribution		GEV distribution	
Estimator	L.S.	Graph+Empirical	L-moment	SLSC	L-moment	SLSC
Ane R.*	360	440	290	0.033	298	0.026
Amano R.	330	380	250	0.032	281	0.021
Seri R.	430	480	313	0.149	401	0.039
Hikone C.	335	380	230	0.181	292	0.044
Yogo R.	260	285	195	0.035	177	0.020
Toyo R.	302	285	302	0.021	300	0.021

L.S.: Least-Squares method.

SLSC<0.03 indicates the goodness of fit is very well.

Hydrologic Frequency Analysis

If we use empirical distribution, analysis becomes simpler.

1. Data Collection (AMS or PDS)
2. Checking Data (Homogeniety and I.I.D.)
3. ~~Enumerate Candidate Models (Many models)~~
4. ~~Fitting Models to Data (Parameter Estimation~~)
5. ~~Goodness-of-Fit Evaluation (Criteria)~~
6. Stability of Quantiles (T-year events) (Resampling)
7. ~~Final Model Selection (Criterion)~~

Method of “Non-Parametric” Analysis

- Probability plot on a normal-scale paper
- Plotting position formula giving a non-exceedance probability to the i-th data

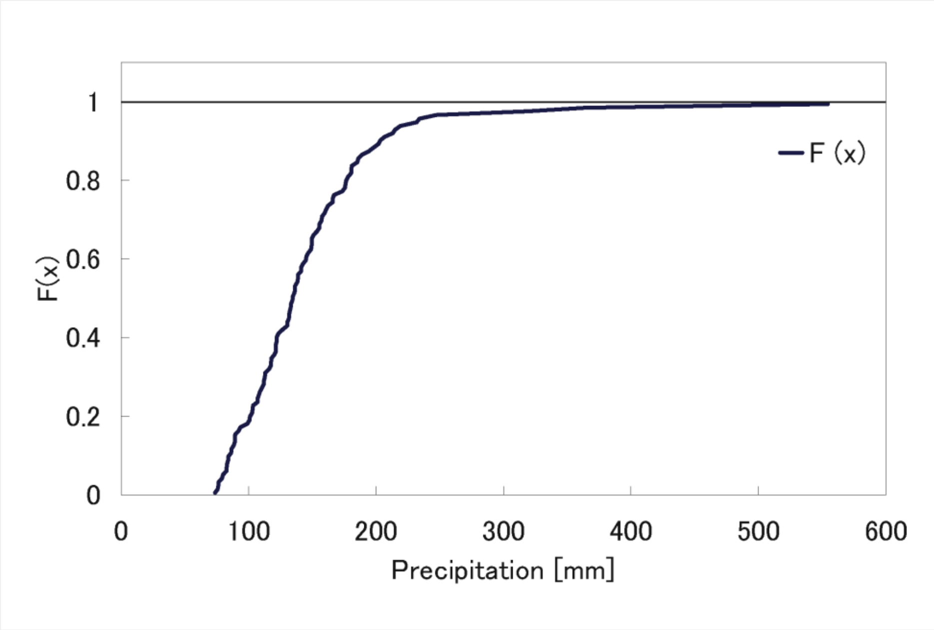
  Here, we compare the differences of

  - Weibull: $F_i = i/(N+1)$

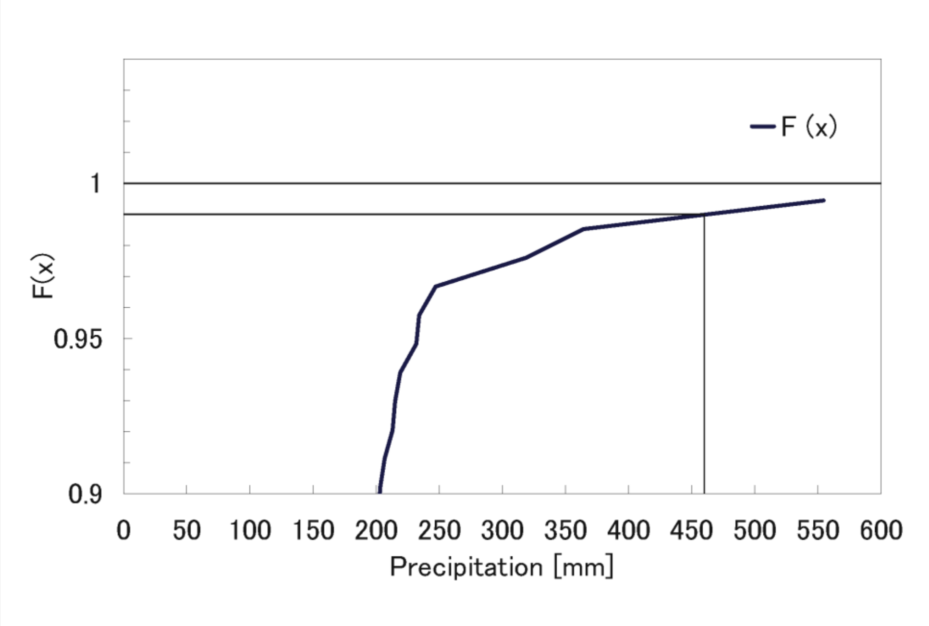
  - Cunnane: $F_i = (i-0.4)/(N+0.2)$

- Linear interpolation of the plotted points

DPRI-KU



Empirical distribution plotted on a normal-scale paper



Top: all data

Bottom: enlargement of a part of non-exceedance probability p>0.90

Plotting position formula

$$F_i = \frac{i - \alpha}{N + 1 - 2\alpha}$$

α = 0 : Weibull

α = 0.25 : Adamowski (1981)

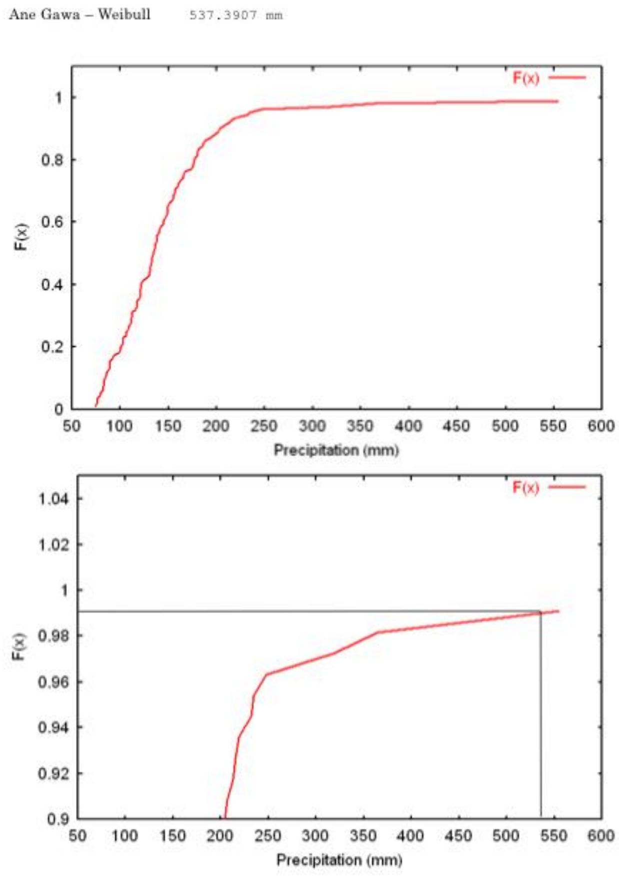
α = 0.375 : Blom

α = 0.4 : Cunnane (1978)

α = 0.44 : Gringorten

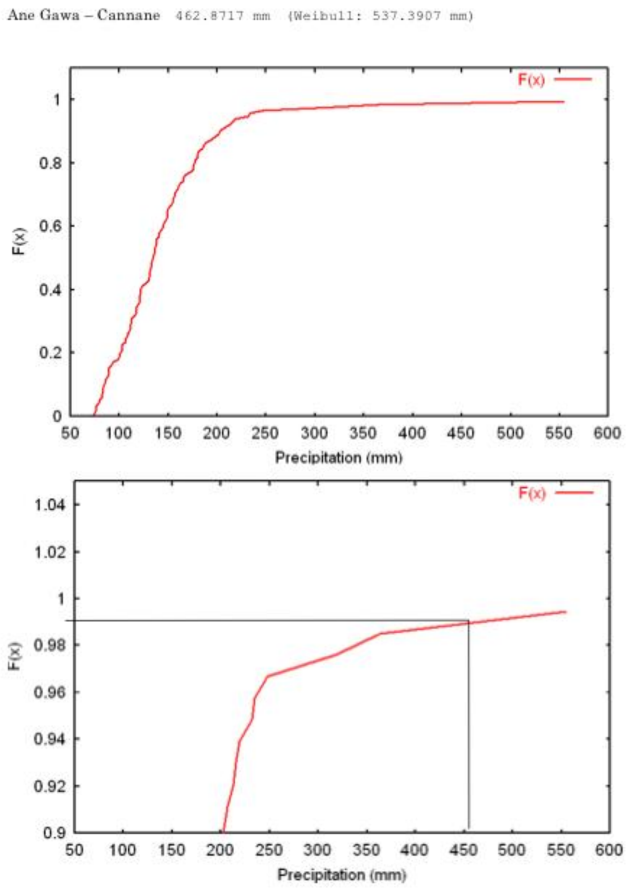
α = 0.5 : Hazen (1914)

Gives non-exceedance probability of the *i*-th order statistics with sample size: *N*



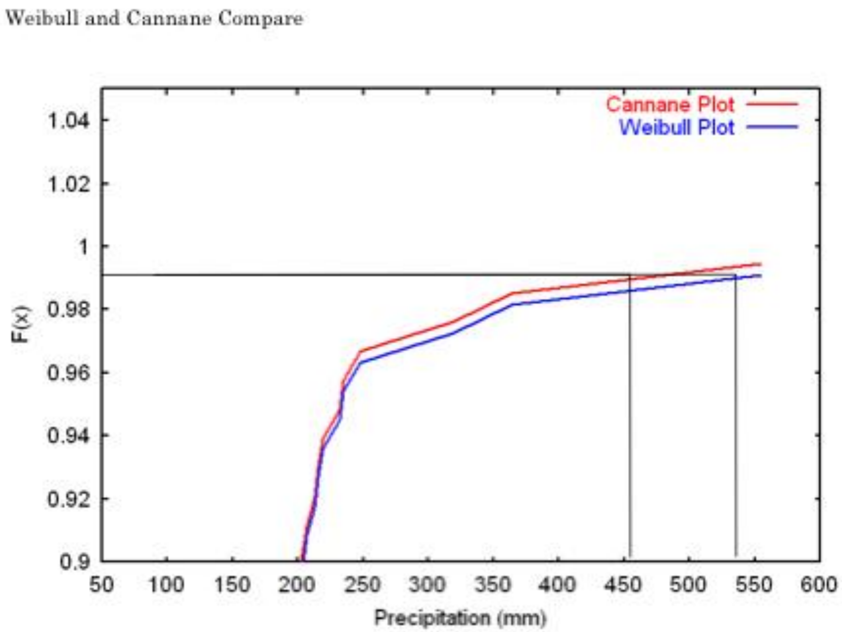
Annual maxima of 2-day precipitation in the Ane River (1886-2003, N=108)

- Weibull Plot
- $X_{0.99}$ = 537 mm



Annual maxima of 2-day precipitation in the Ane River (1886-2003, N=108)

- Cunnane Plot
- X(0.99)=462.9 mm



The Weibull plot gives larger quantile (T-year event) estimates.

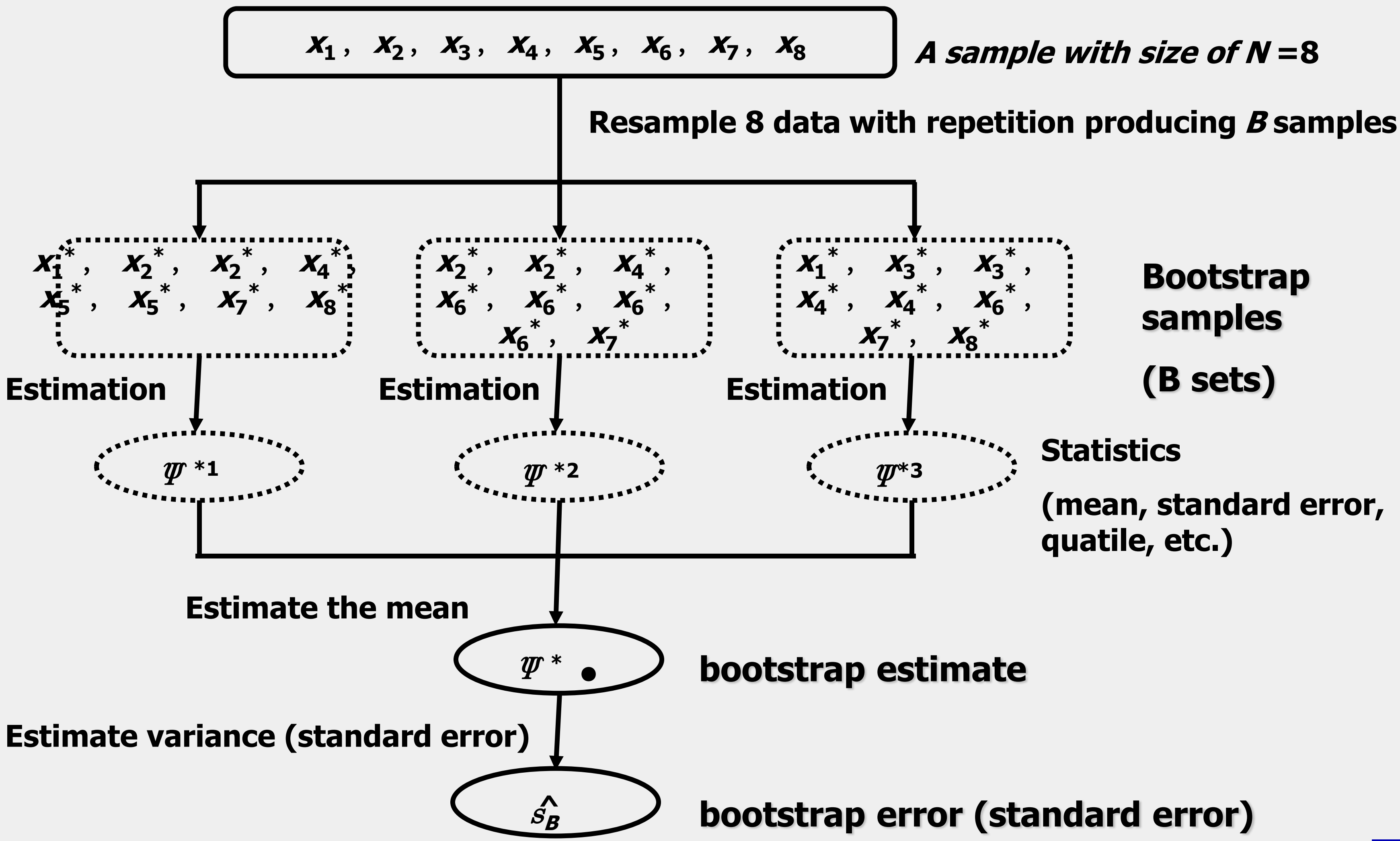
Annual maxima of 2-day precipitation in the Ane River (1886-2003, N=108)

- Cunnane Plot
  $x_{0.99}$= 463 mm
- Weibull Plot
  $x_{0.99}$= 537 mm

Bootstrap resampling

- A random resampling from the original dataset with data $\{x_1, \ldots, x_N\}$ generates a sample with the same size of N. The generated sample is described as $\{x_1{}^*, \ldots, x_N{}^*\}$. This set of data is called *bootstrap sample*.
- Repeating the operation in the previous step independently many times (B times), we can obtain a statistic G for each of the bootstrap samples.
- Using this bootstrap sample, we can obtain a statistic. The statistic obtained for the i-th bootstrap sample is written as G_i.
- The average value of G_i is the bootstrap estimate of the statistic .
- The bootstrap estimate of the variance in the statistic is $\mathrm{Var}\{G_i\}$.

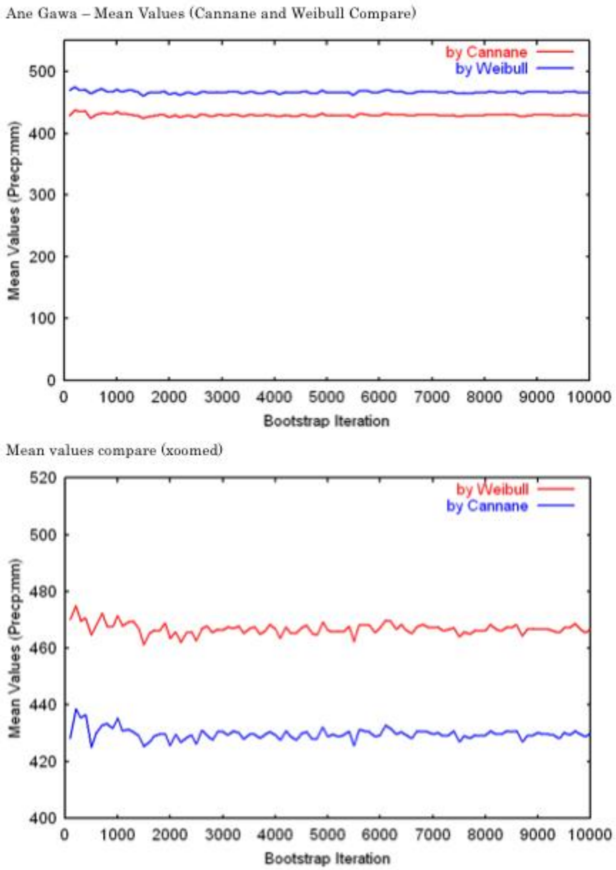
Bootstrap resampling



100-year rainfall estimated by the non-parametric method

100-year daily Precipitation (mm) estimated by the non-parametric method [*two-day rainfall only for the Ane river]

Location	*N*	(I) Weibull	(II) Cunnane	(II') Bootstrap with Cunnane
Ane River*	108	537	463	430
Amano River	107	517	413	386
Seri River	107	735	541	499
Hikone City	100	560	436	386
Yogo River	106	413	315	291
Toyo River	104	286	284	280



Result of the bootstrap resampling

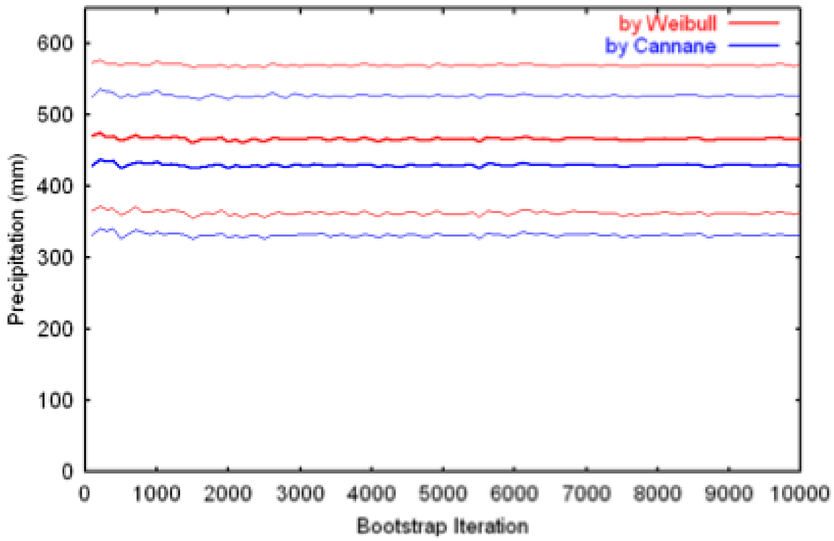
- Number of bootstrap samples:

  B=100 ~ 10000
- B should be B>2000 to obtain stable quantile estimates
- Cunnane Plot

  $x_{0.99}$ = 430 mm
- Weibull Plot

  $x_{0.99}$ = 470 mm

Ane Gawa – Mean Values and StDev (Weibull:red line, Cannane: blu line)



Result of the bootstrap resampling

Bootstrap Standard Error

- Cunnane Plot
- Weibull Plot
- $S_{X(0.99)} \sim 100$ mm

Results summary (Part-2)

- Non-parametric method (using the empirical distribution) could be useful for samples larger than the return period considered.
- Non-parametric method applied to the original sample overestimates the 100-year quantile (463 mm for the Ane River). Bias correction could be done by the bootstrap resampling (430 mm).
- Weibull's plot tends to overestimate（470 mm）. Cunnane's plot is recommended (430 mm).

DPRI-KU

How can we treat long-term extreme events?

- Parametric method using probability distribution functions – traditional method
- Non-parametric method using empirical distribution (Takara, 2006)
- Resampling methods such as the jackknife and bootstrap can be used for bias correction and estimation of accuracy of quantile (T-year event) estimates

Today's talk

Part-1: Brief review of "Parametric" method

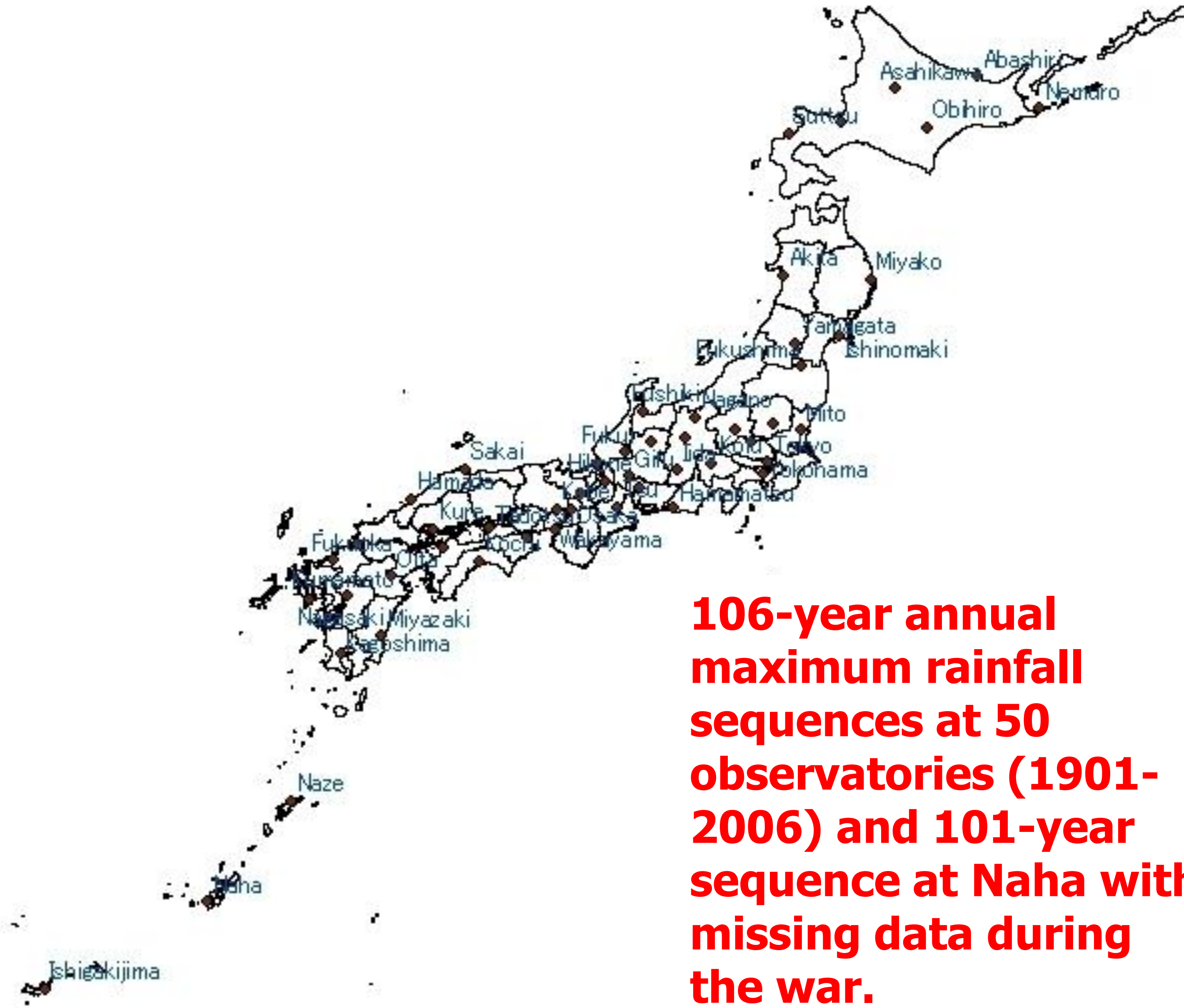
Part-2: "Non-parametric" method for long-term extreme-value series combined with the bootstrap resampling -- New idea

Part-3: Trend analysis and How to consider Climate Change effect

- Non-parametric method with unequal occurrence probability -- New idea

DPRI-KU

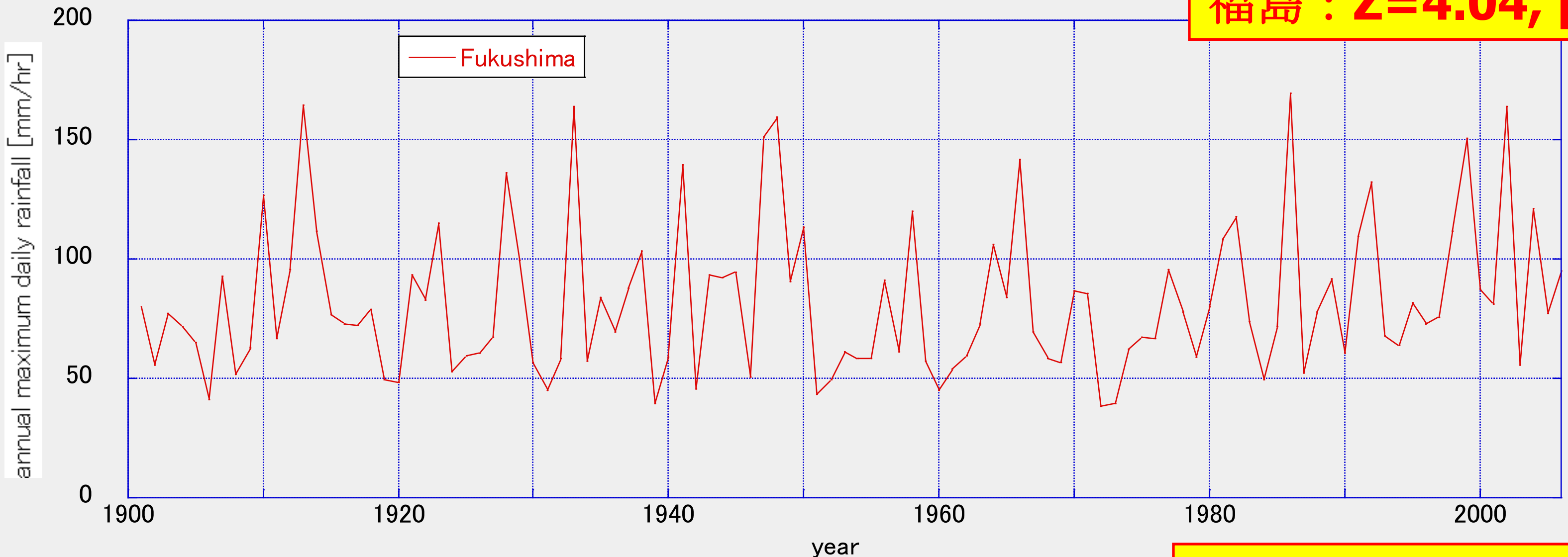
51 JMA meteorological observatories



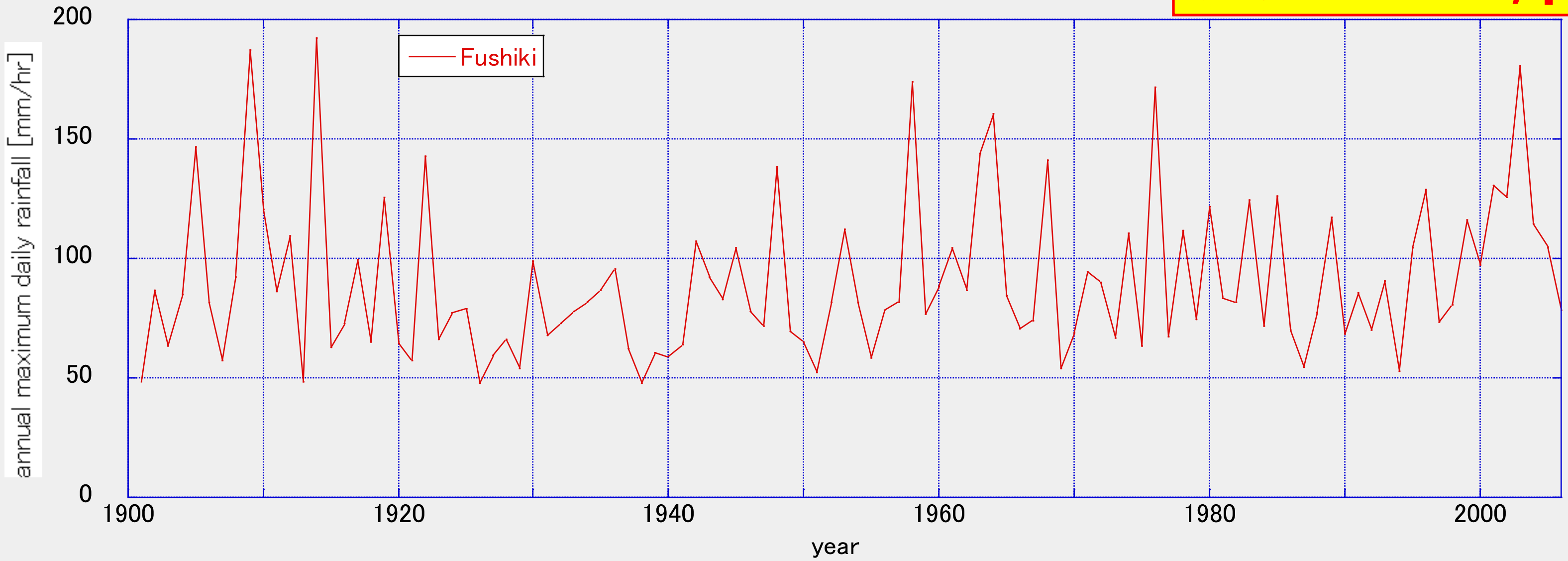
106-year annual maximum rainfall sequences at 50 observatories (1901-2006) and 101-year sequence at Naha with missing data during the war.

Locations indicating increasing trend by the Mann-Kendall Test

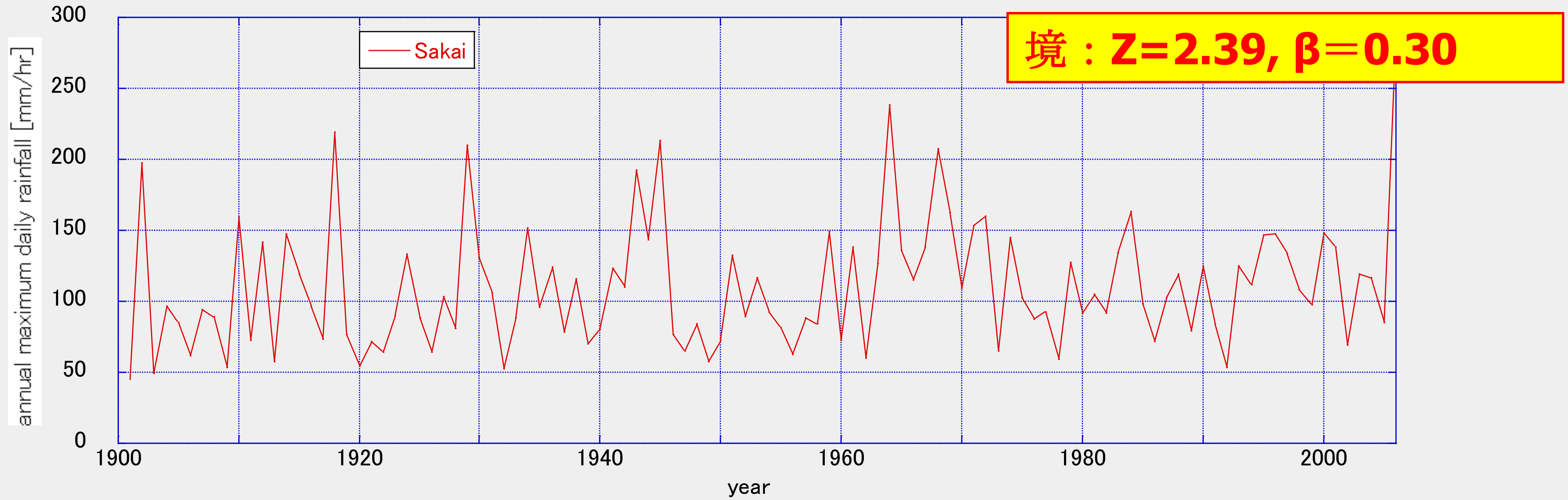
福島：Z=4.04, β＝0.50

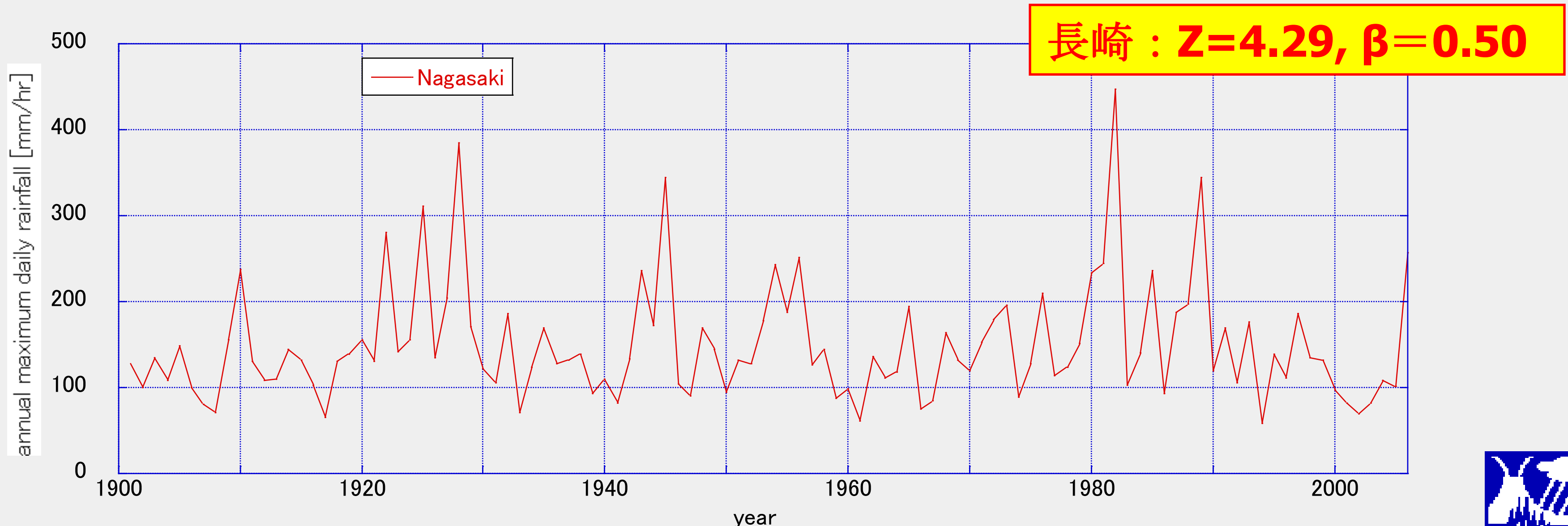


伏木：Z=2.29, β＝0.19



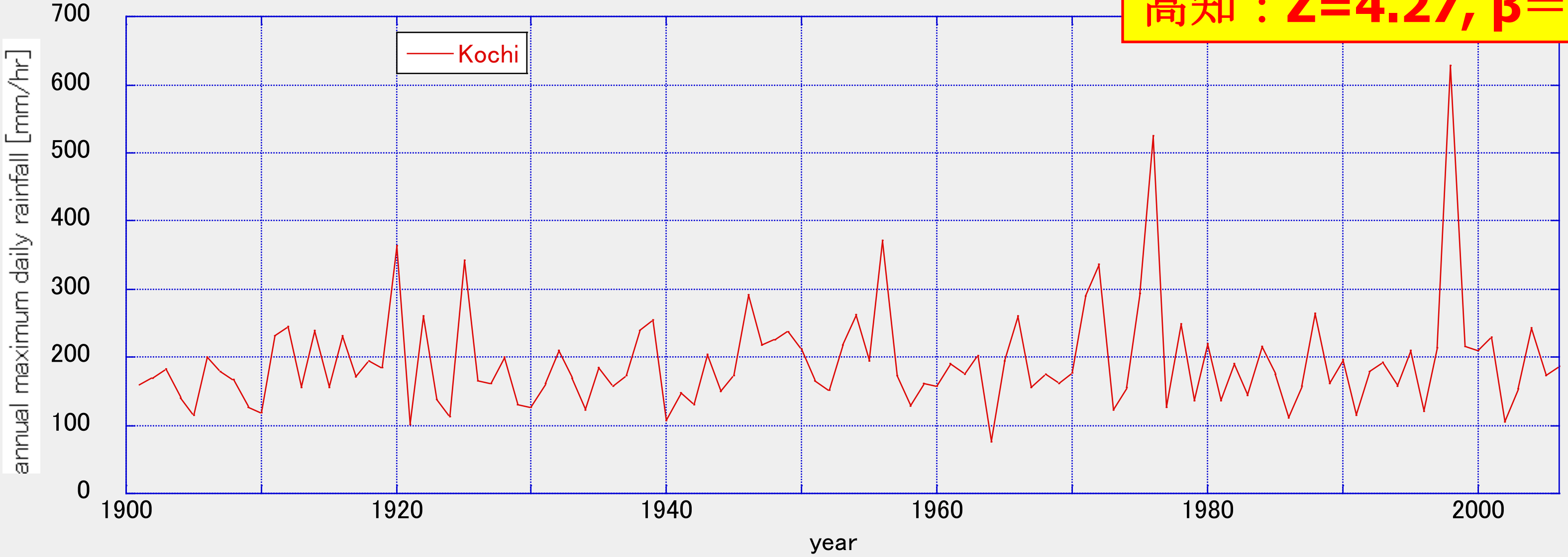
Locations indicating increasing trend by the Mann-Kendall Test



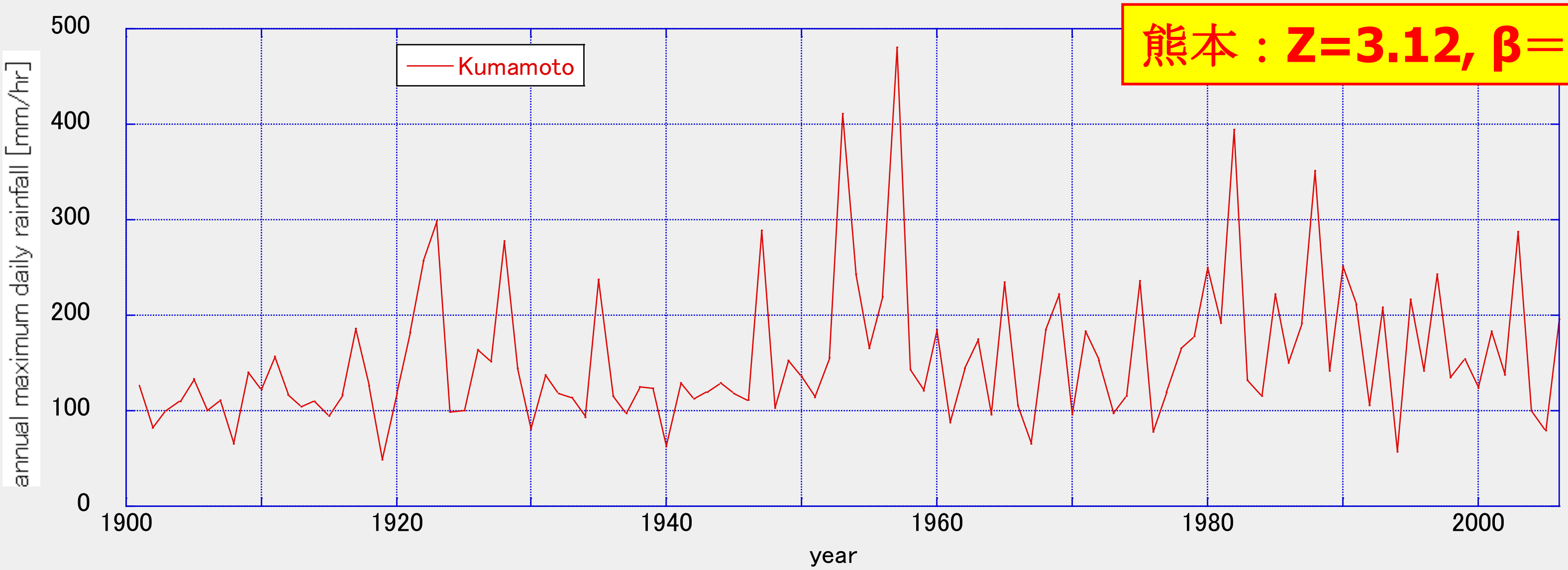


Locations indicating increasing trend by the Mann-Kendall Test

高知：Z=4.27, β＝0.50

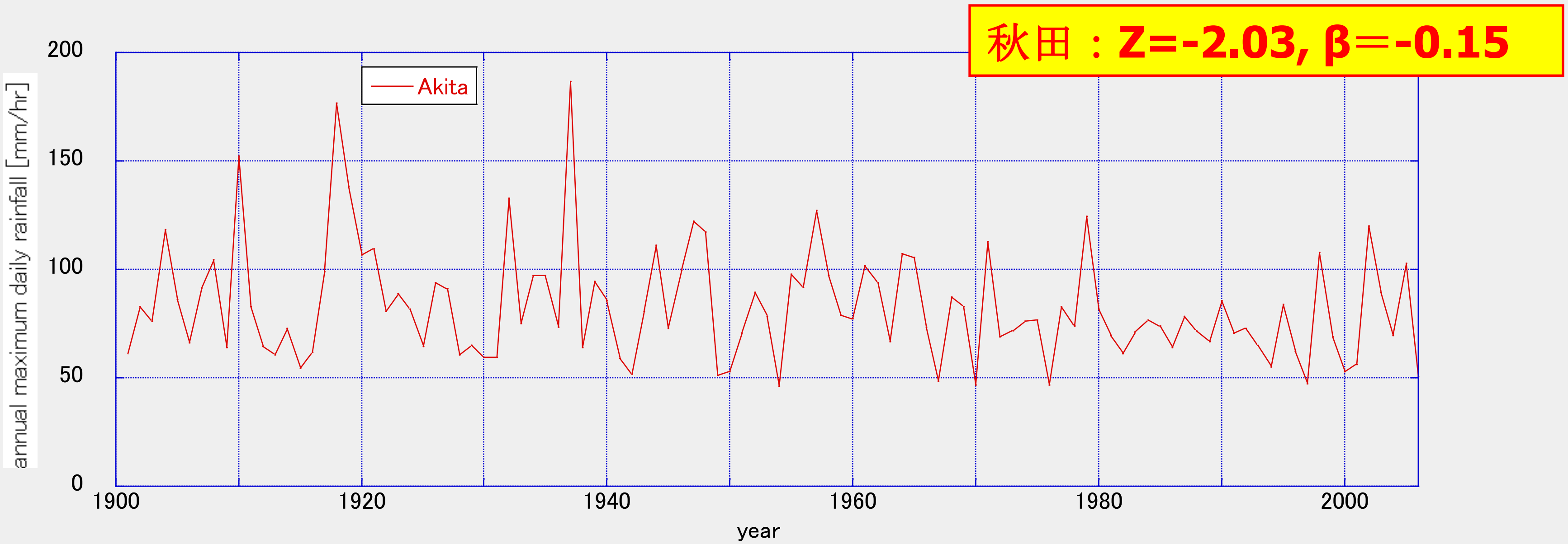


熊本：Z=3.12, β＝0.53



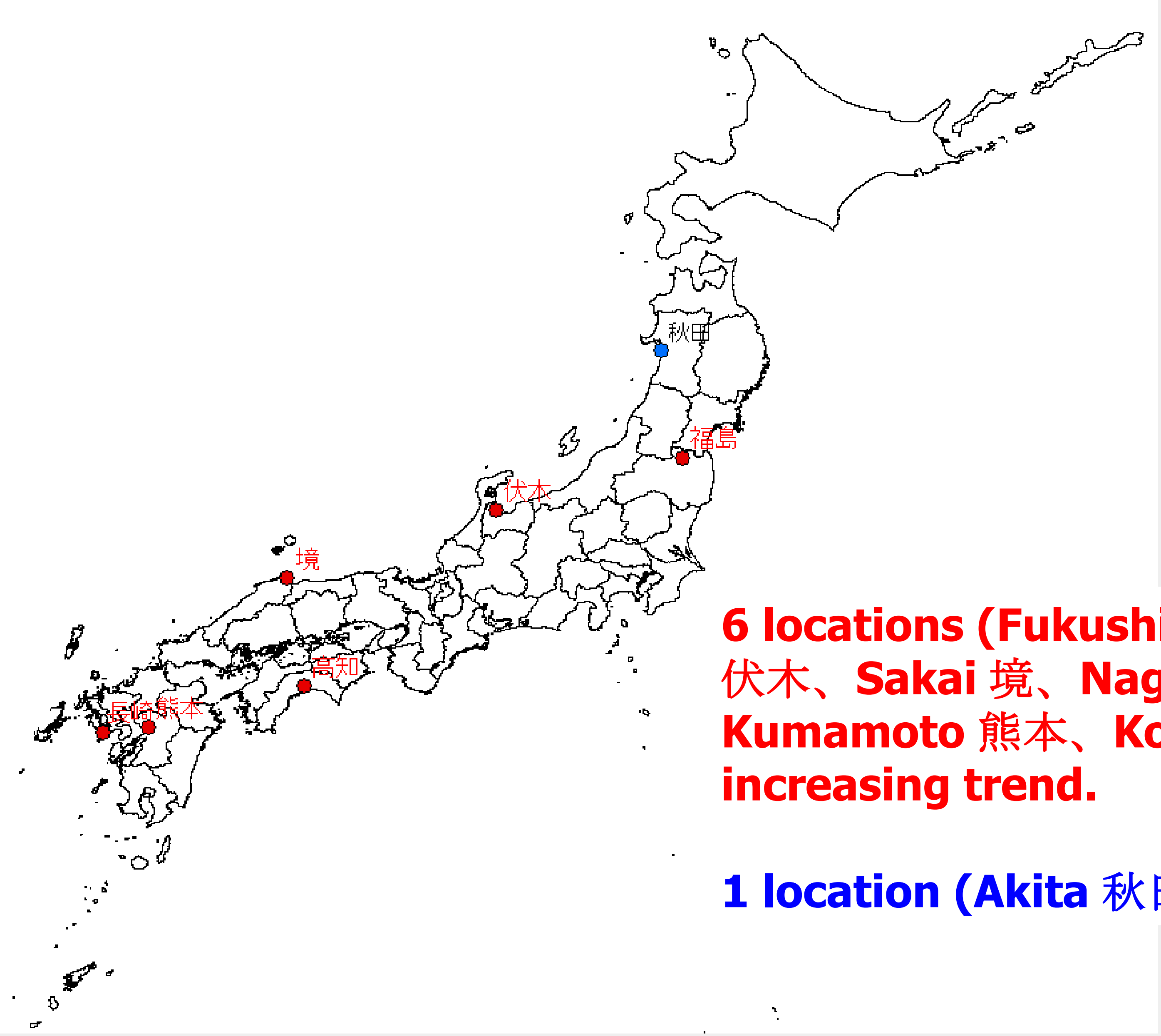
DPRI-KU

Locations indicating decreasing trend by the Mann-Kendall Test



DPRI-KU

Mann-Kendall Test Results (3)

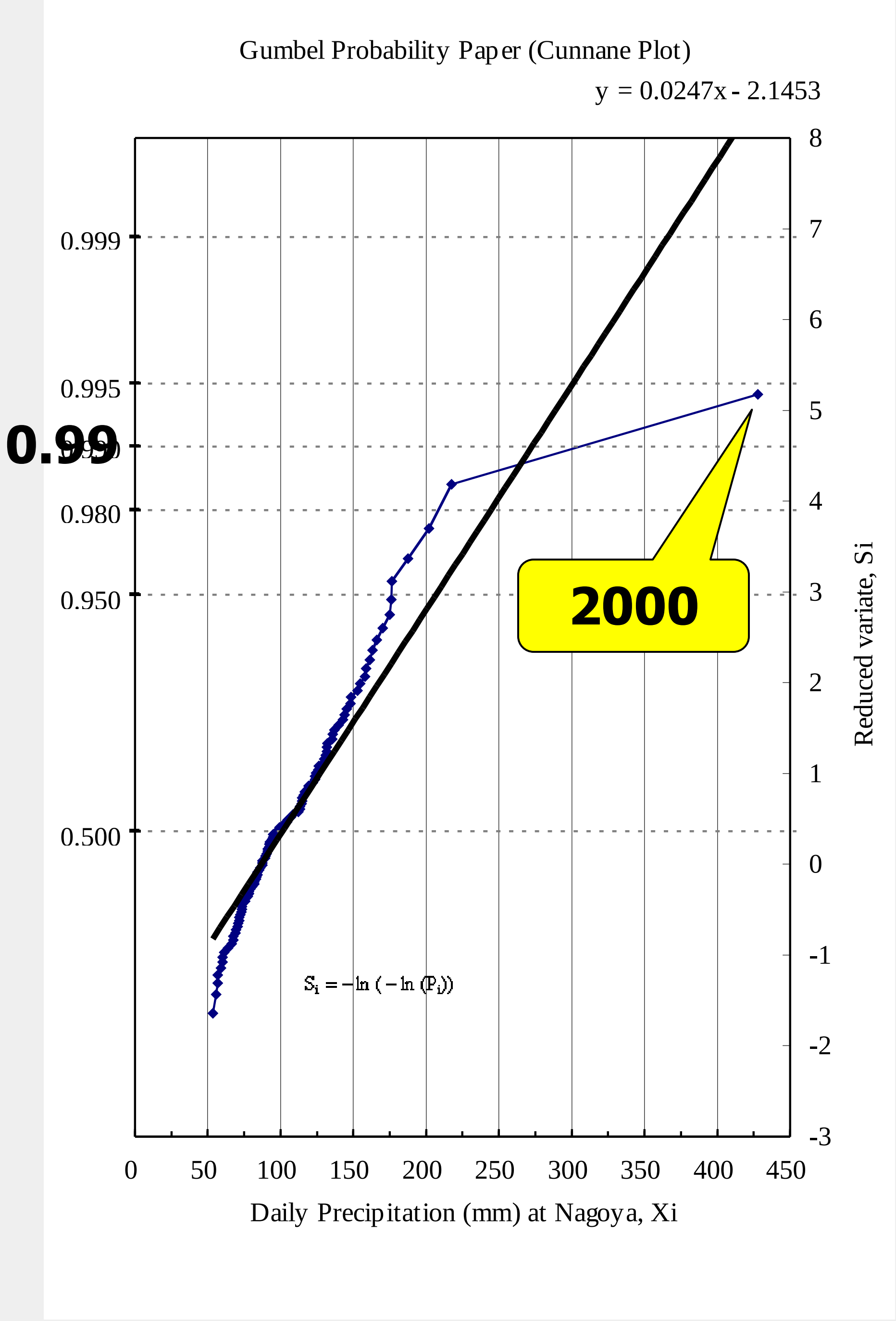


6 locations (Fukushima 福島、Fushiki 伏木、Sakai 境、Nagasaki 長崎、Kumamoto 熊本、Kochi 高知) showed increasing trend.

1 location (Akita 秋田) decreasing.

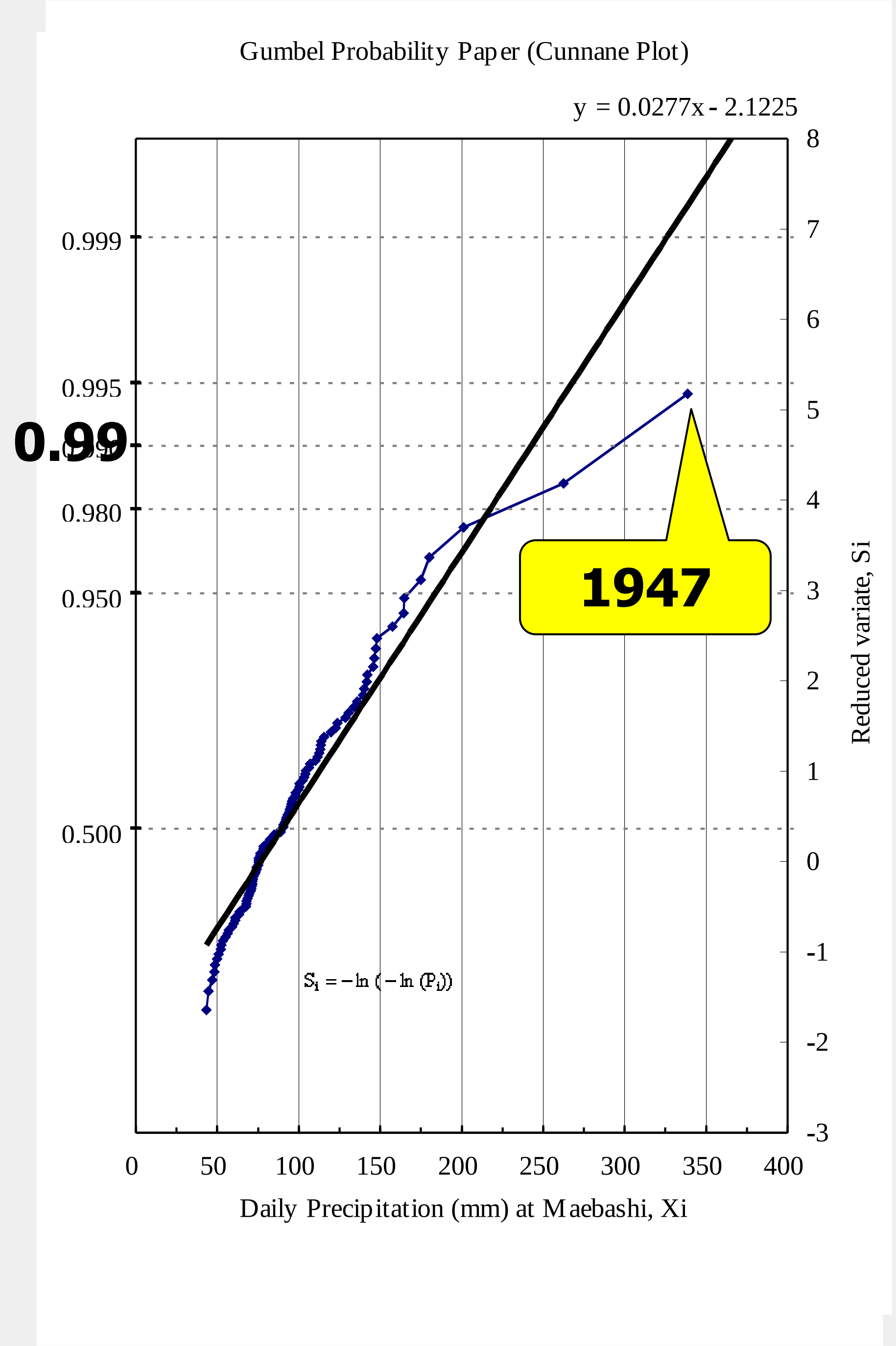
Nagoya

Non-parametric: 330.7 mm



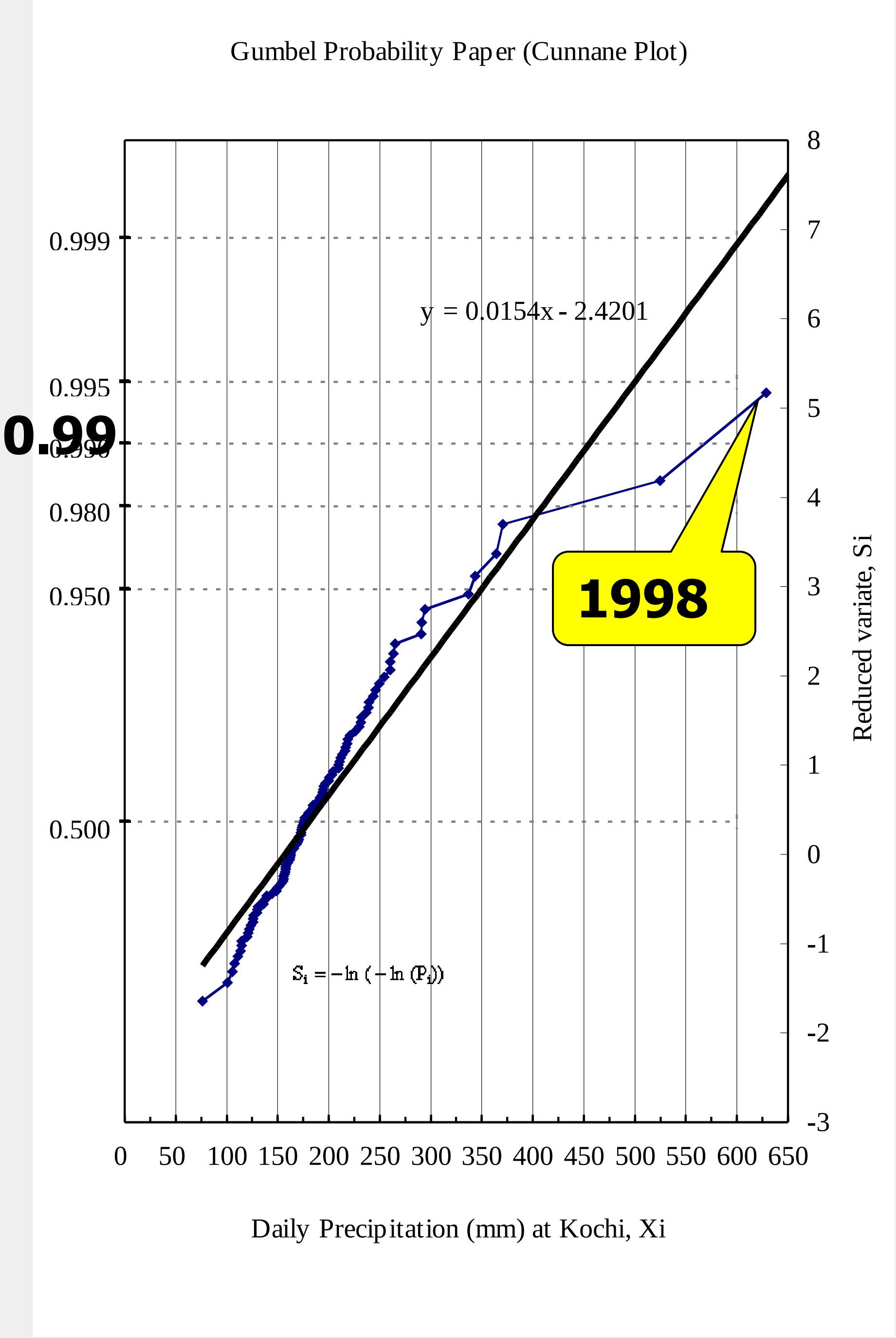
Maebashi

Non-parametric 303.4 mm



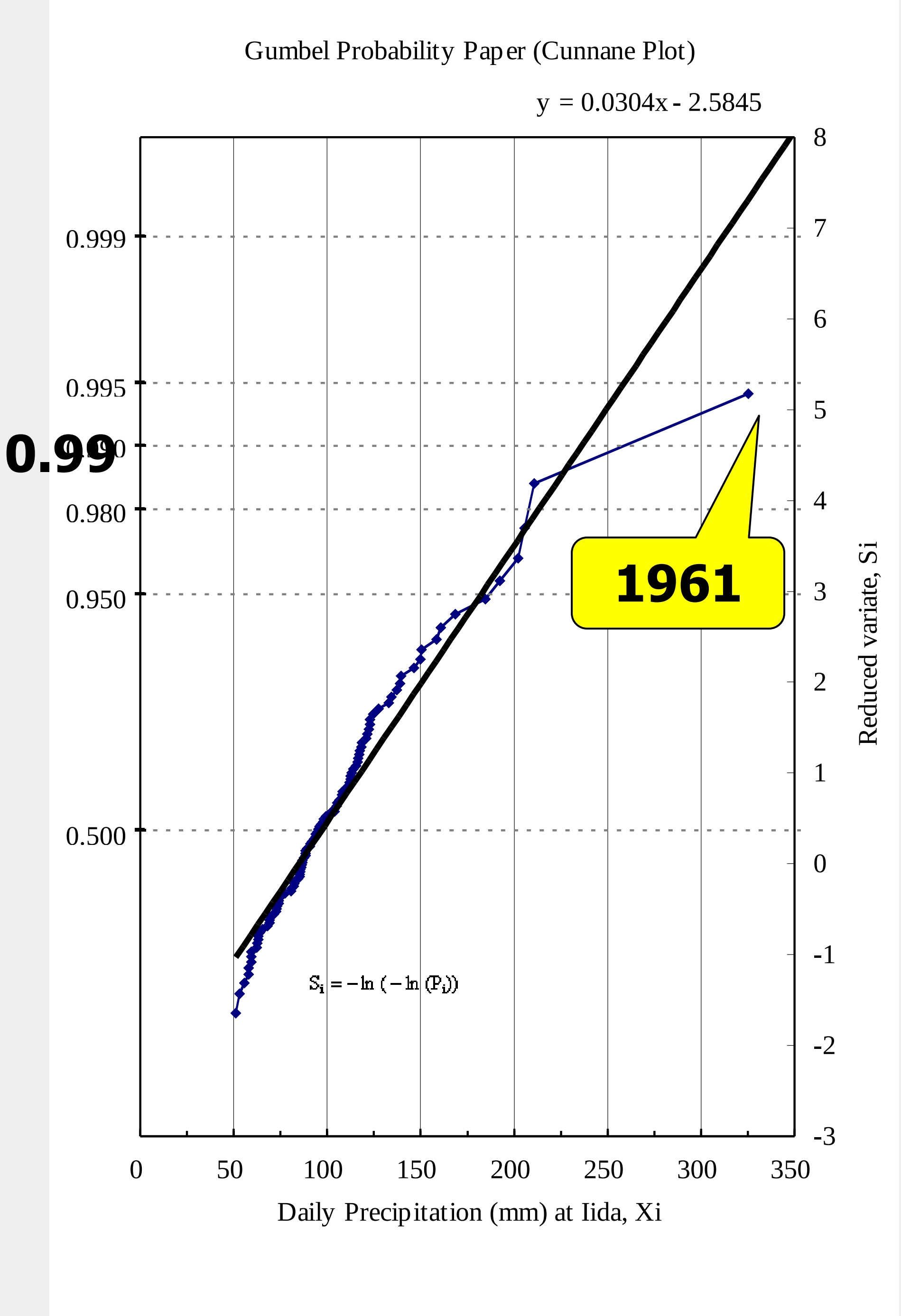
Kochi

Non-parametric: 580.5 mm



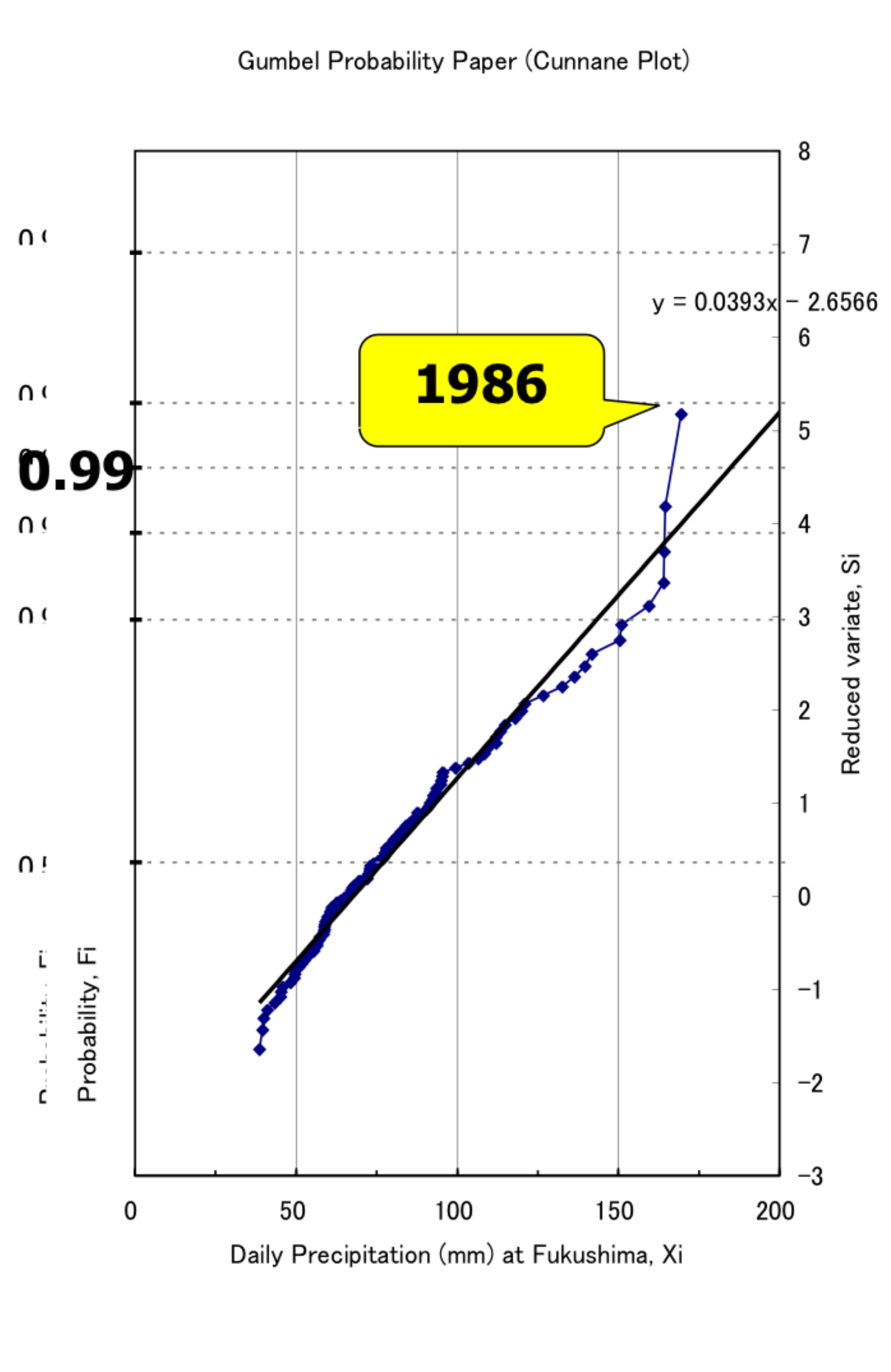
Iida

Non-parametric: 272.4 mm



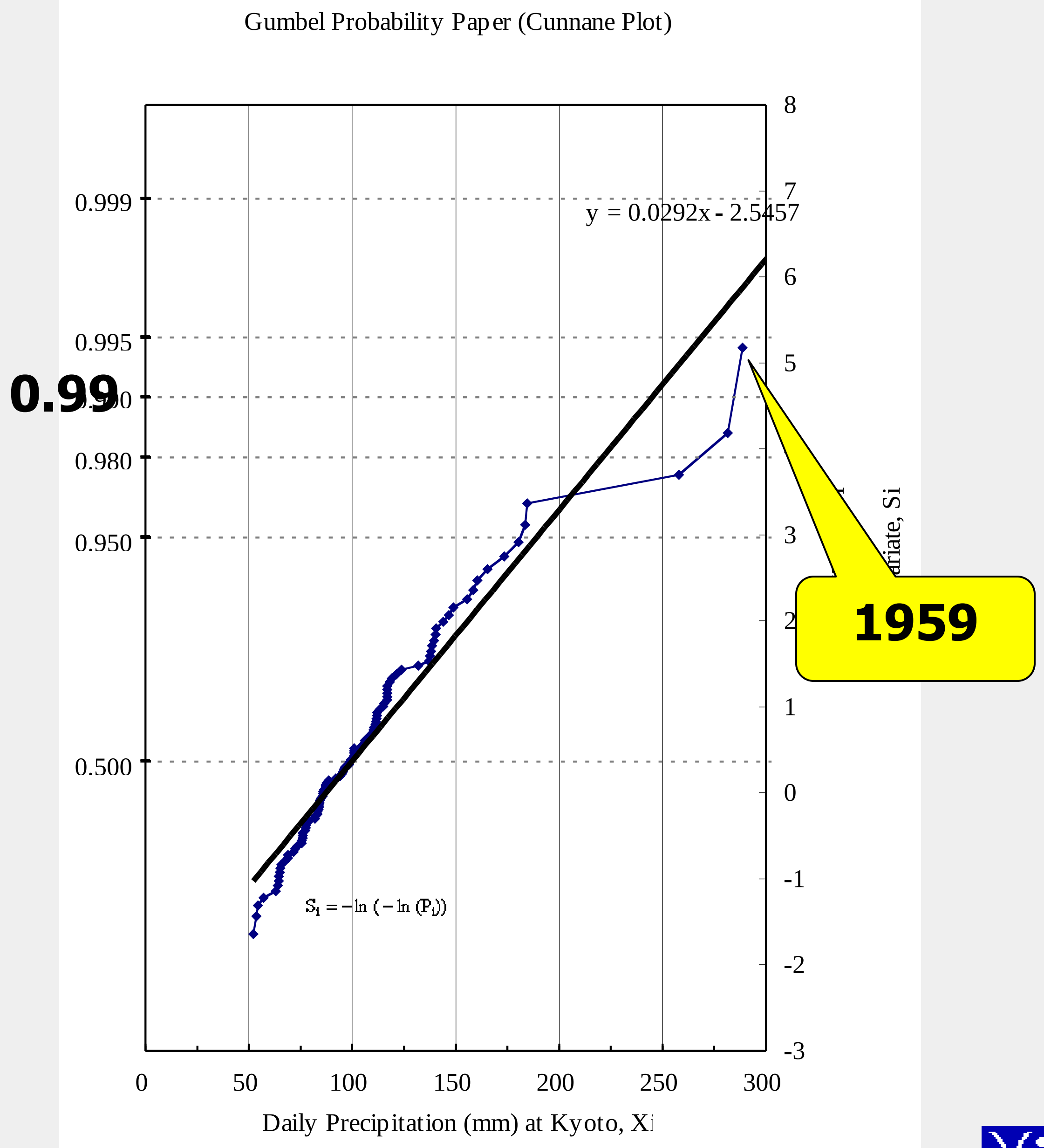
Fukushima

Non-parametric: 167.2 mm



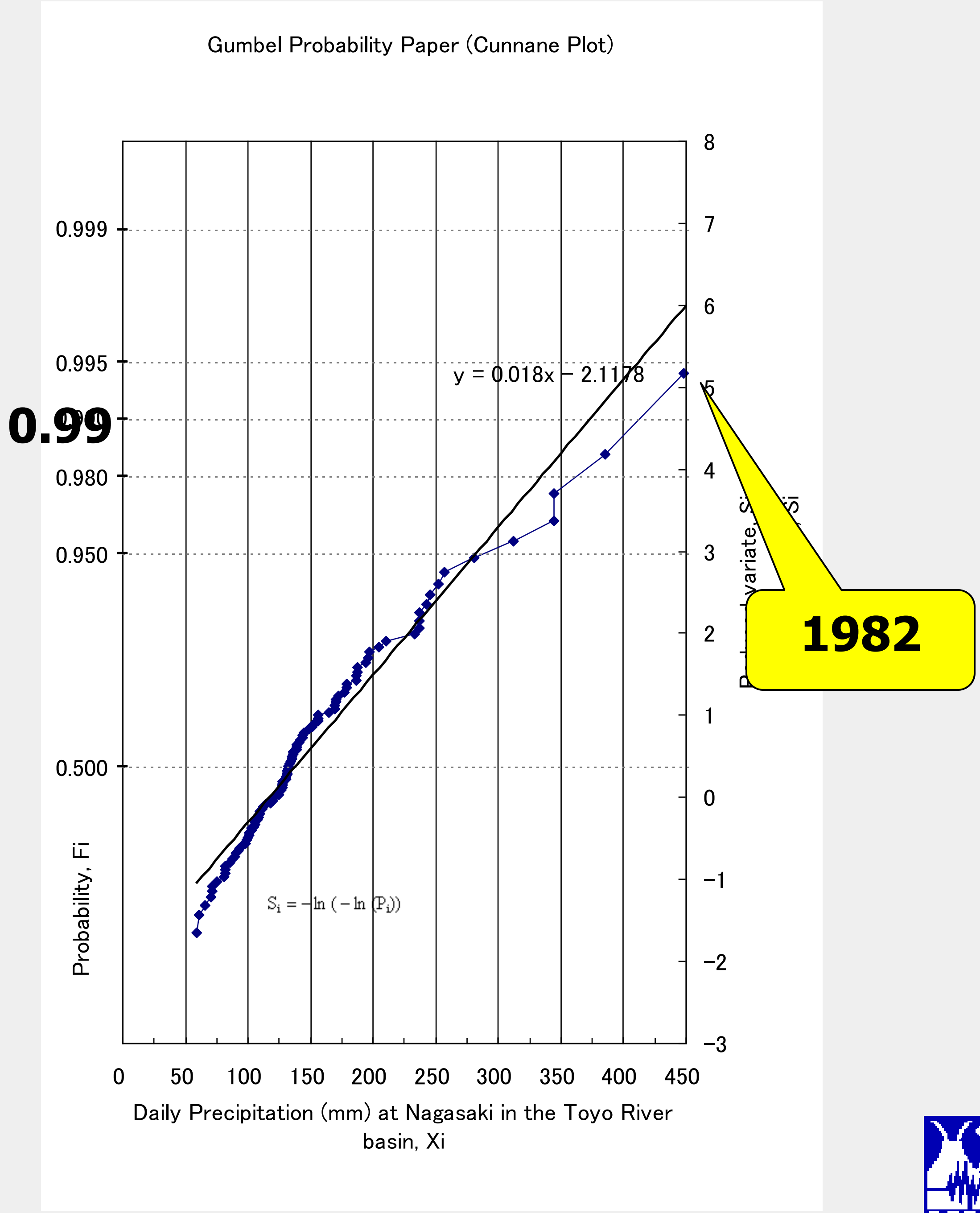
Kyoto

Non-parametric: 285.4 mm



Nagasaki

Non-parametric: 419.1 mm



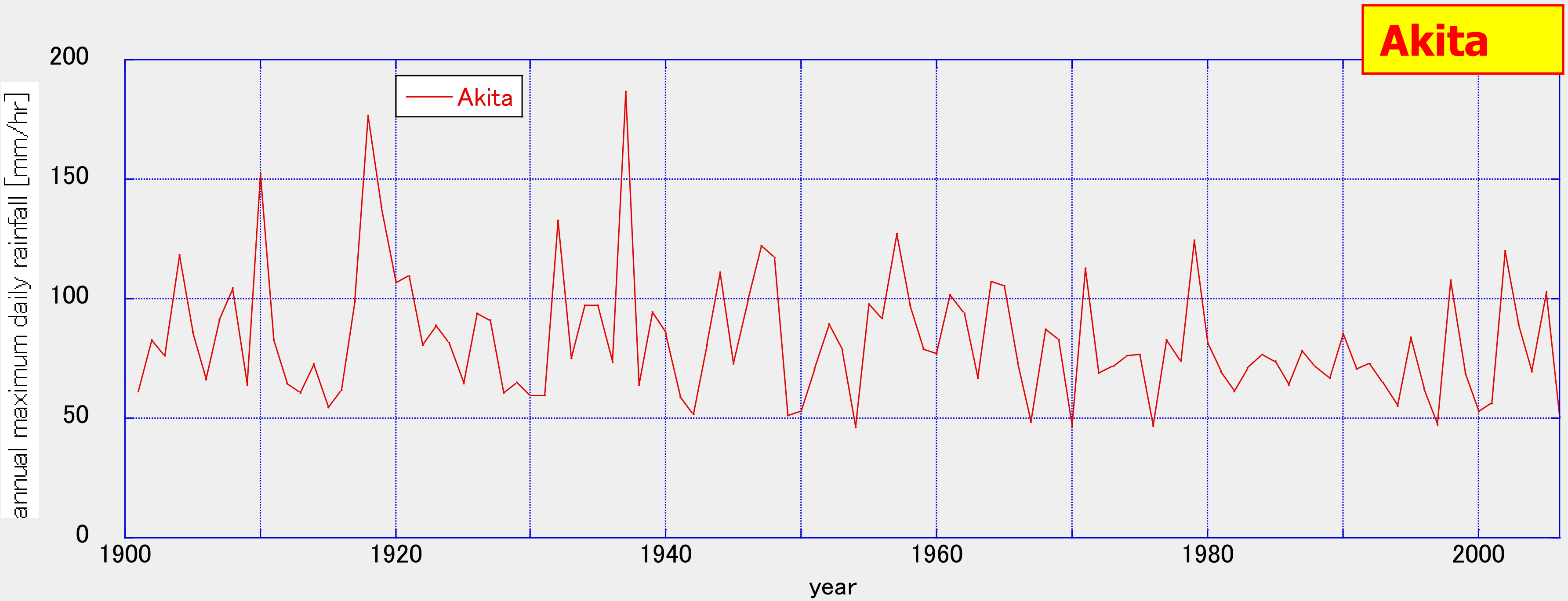
Non-stationary Non-Parametric Hydrological Frequency Analysis Incorporating Climate Change Effect

Extreme events are brought by a global climate system

A working hypothesis:
"The global climate system is changing."

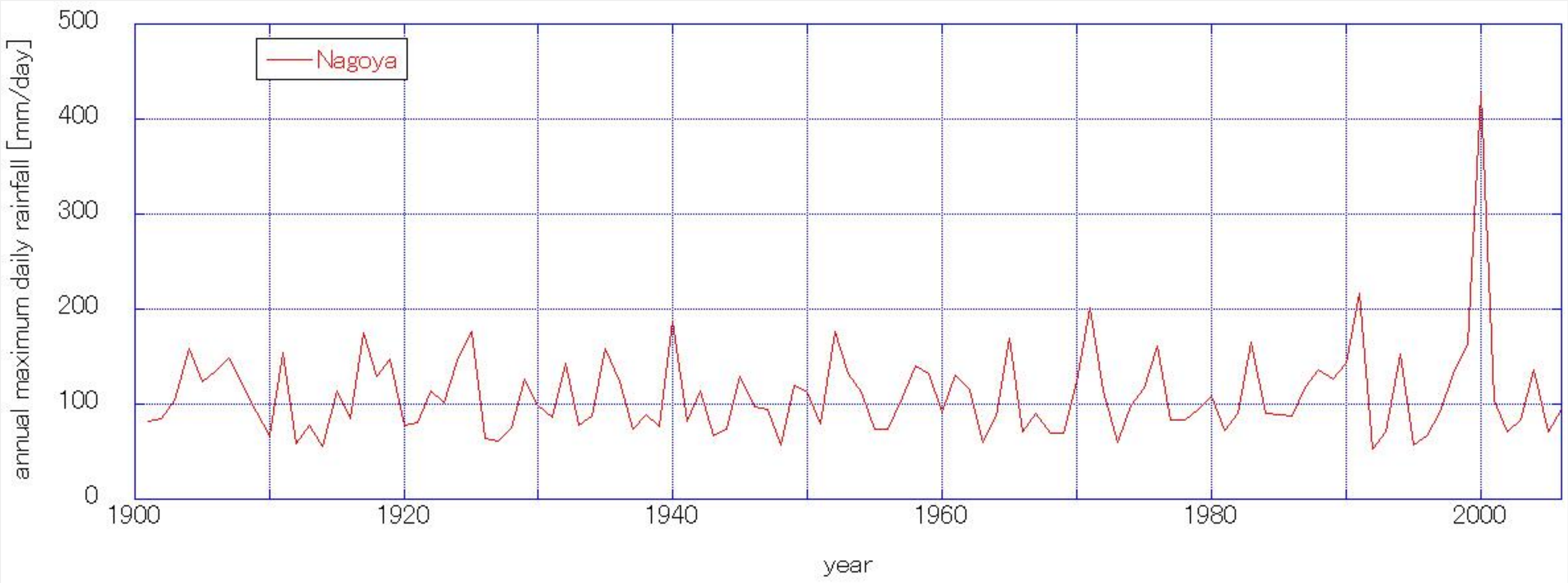
If so, how we can deal with historical extreme events?

106-year annual maximum rainfall series at Akita

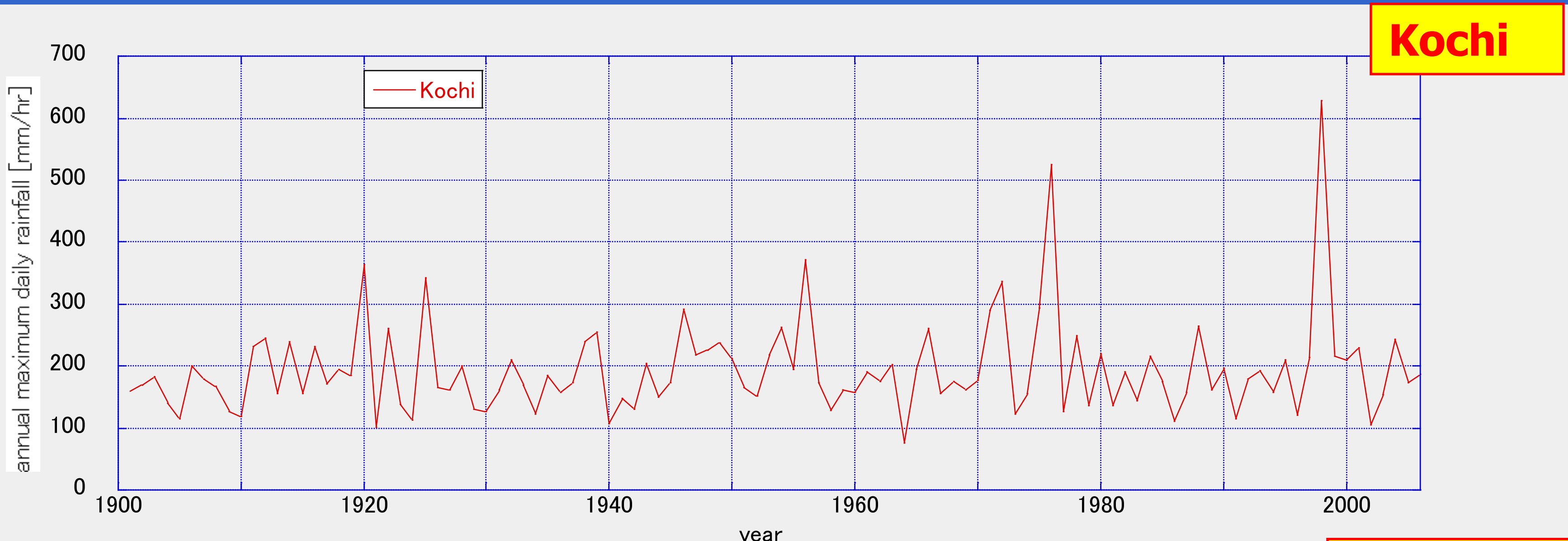


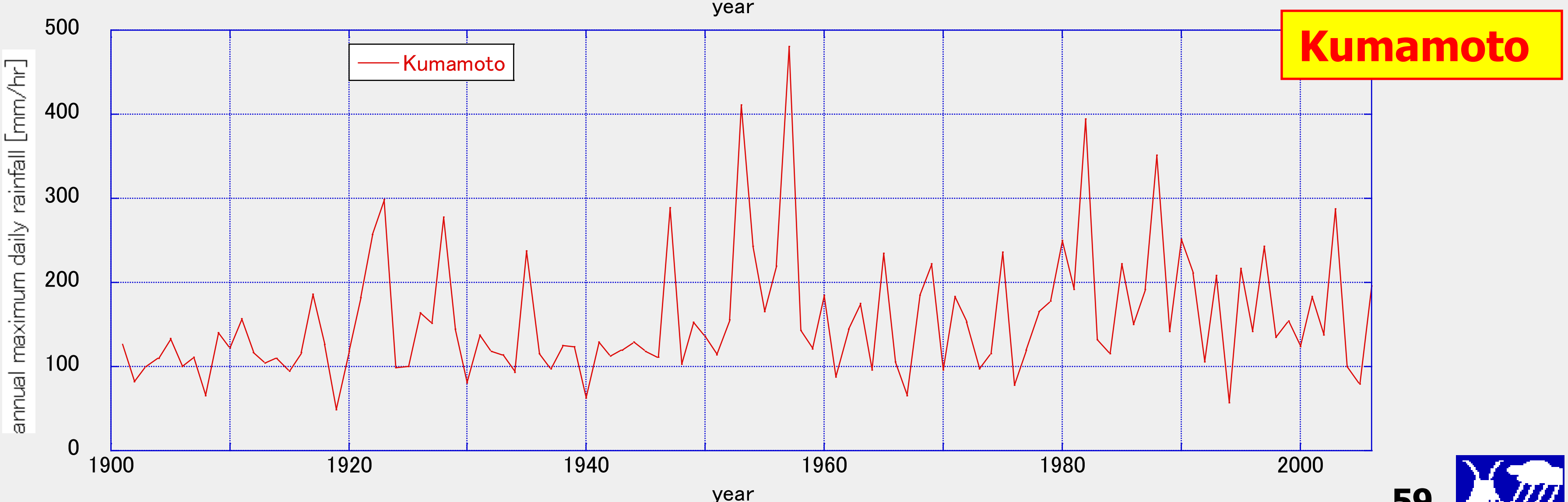
106-year annual maximum rainfall series at Nagoya

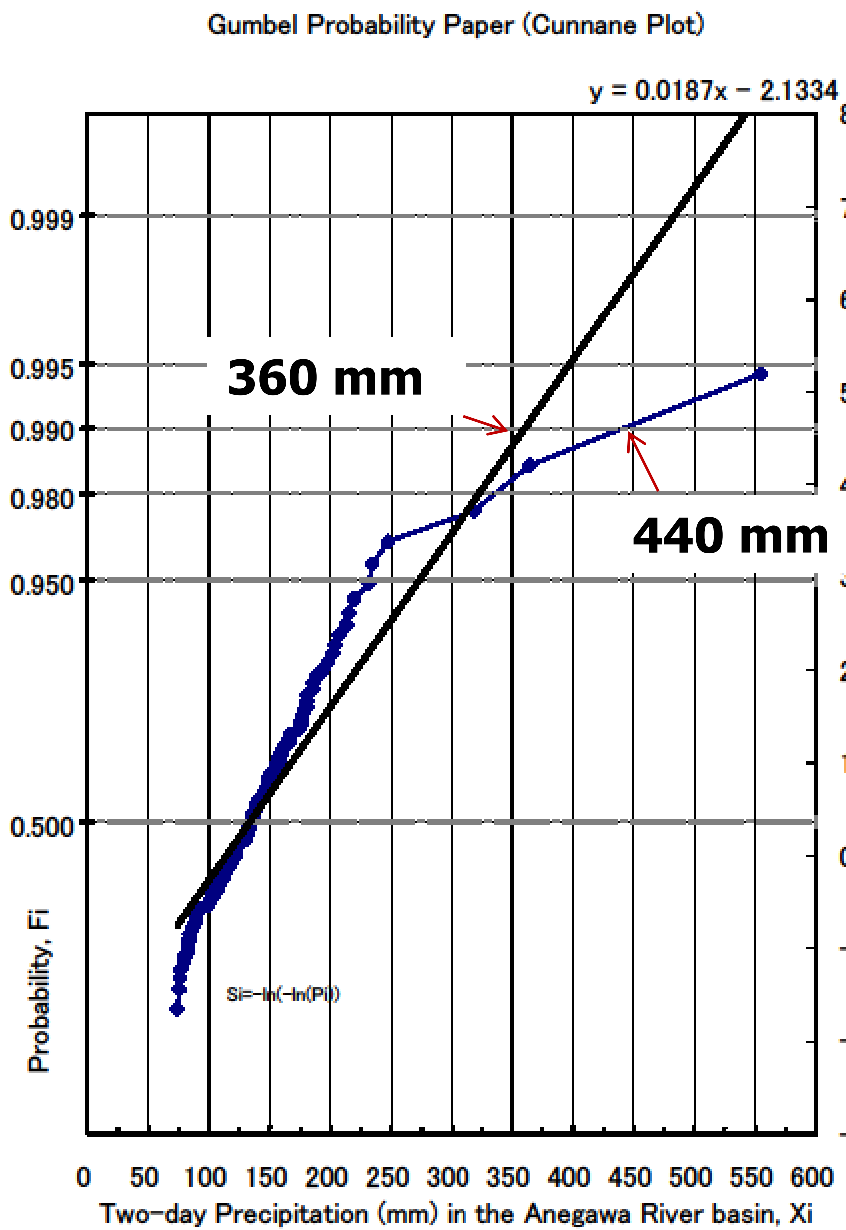
Nagoya



106-year annual maximum rainfall series at Kochi and Kumamoto







Parametric method with Gumbel probability paper

Example: a 108-year sequence of annual maximum 2-day rainfalls in the Anegawa river basin (1886-2003)

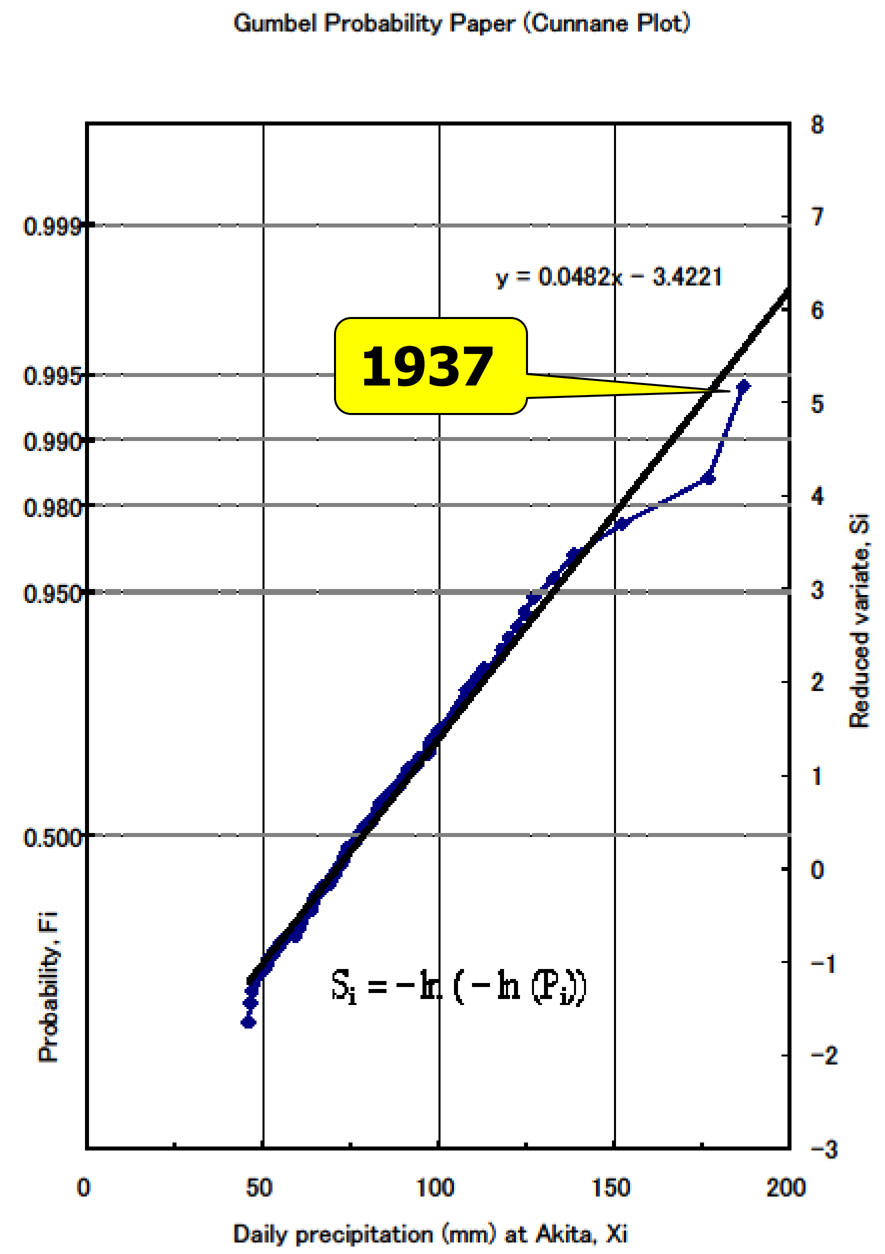
秋田 Akita

Gumbel

X100= 168 mm

Non-Parametric (one time)

X100= 182 mm



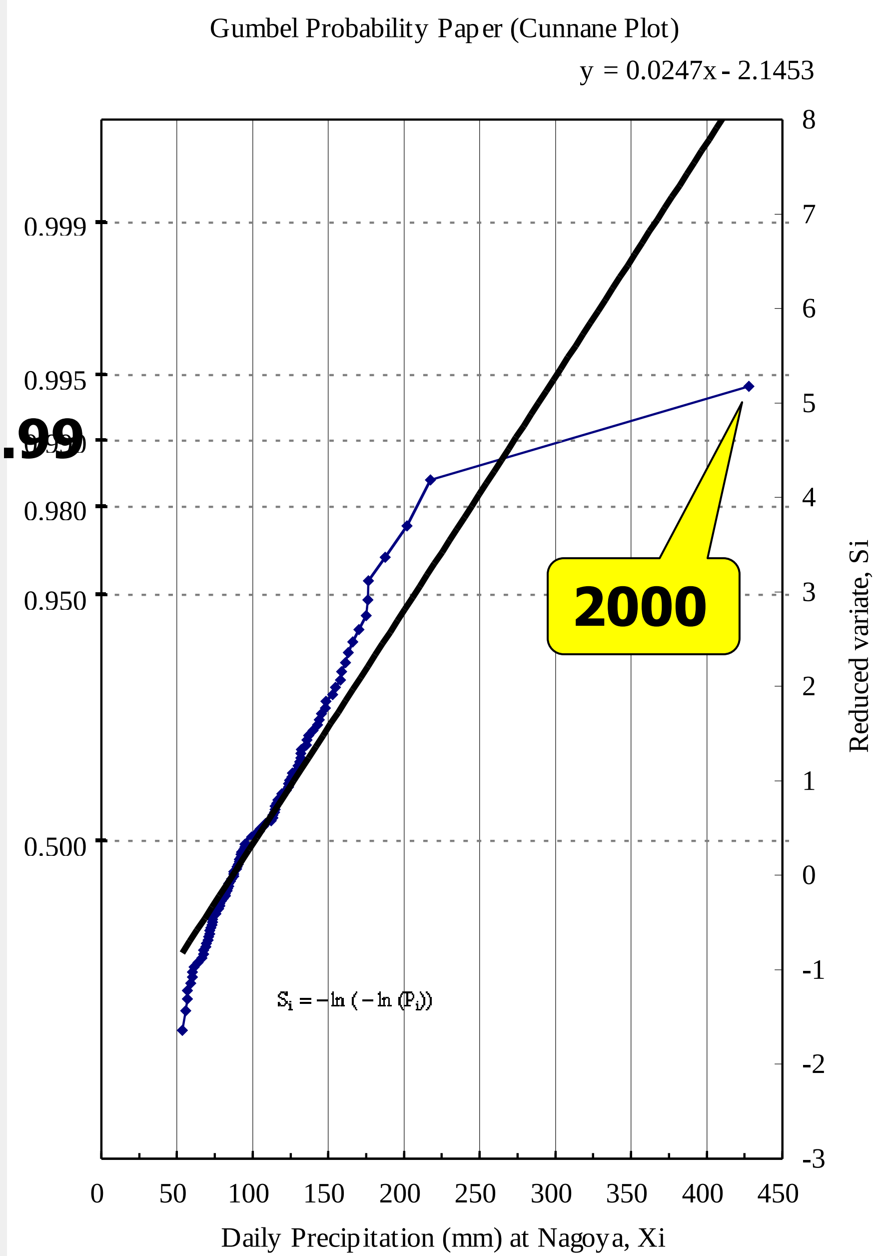
名古屋 Nagoya

Gumbel

X100= 280 mm

Non-Parametric (one time)

X100= 330.7 mm



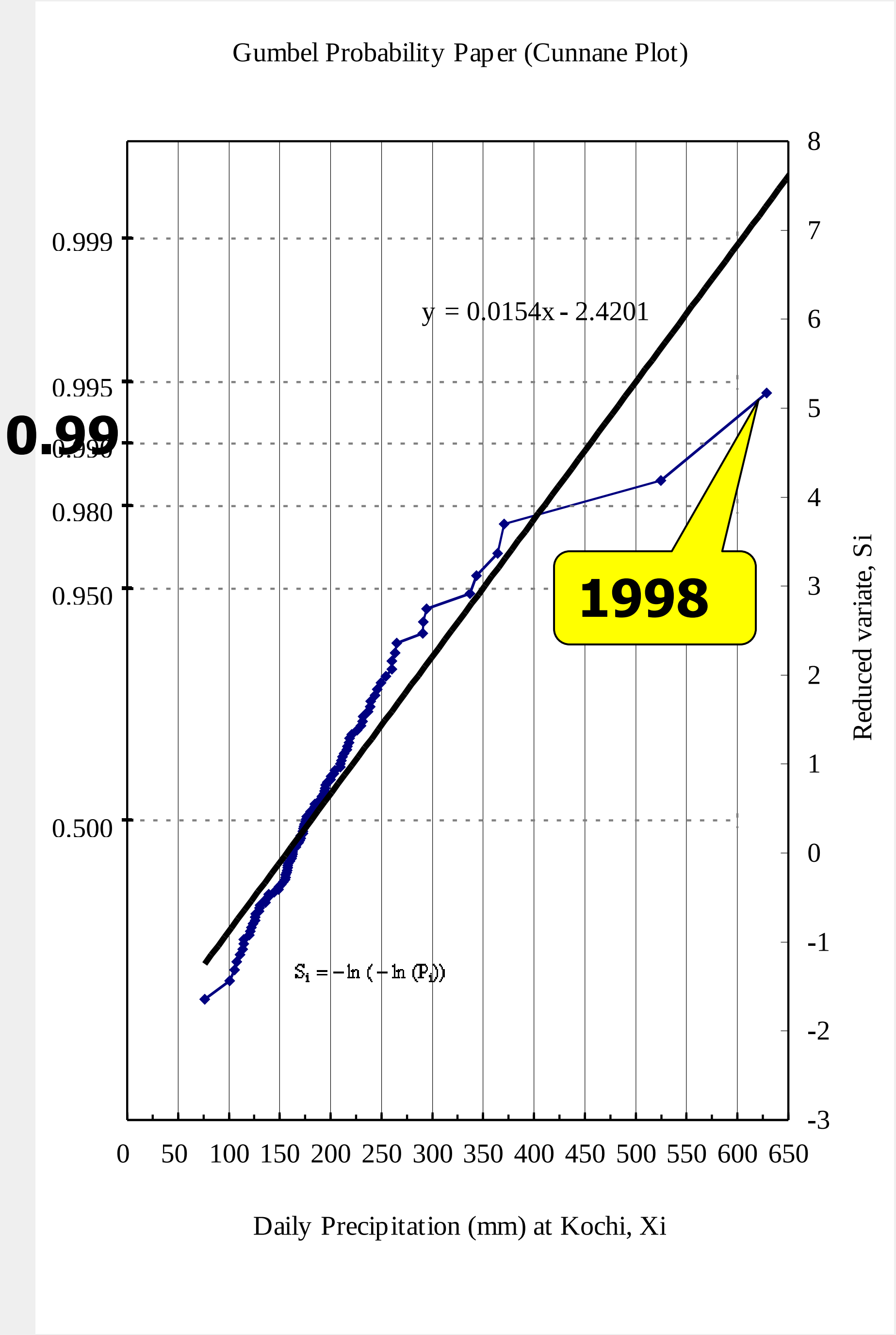
高知 Kochi

Gumbel:

X100= 445 mm

Non-Parametric (one time):

X100= 580 mm



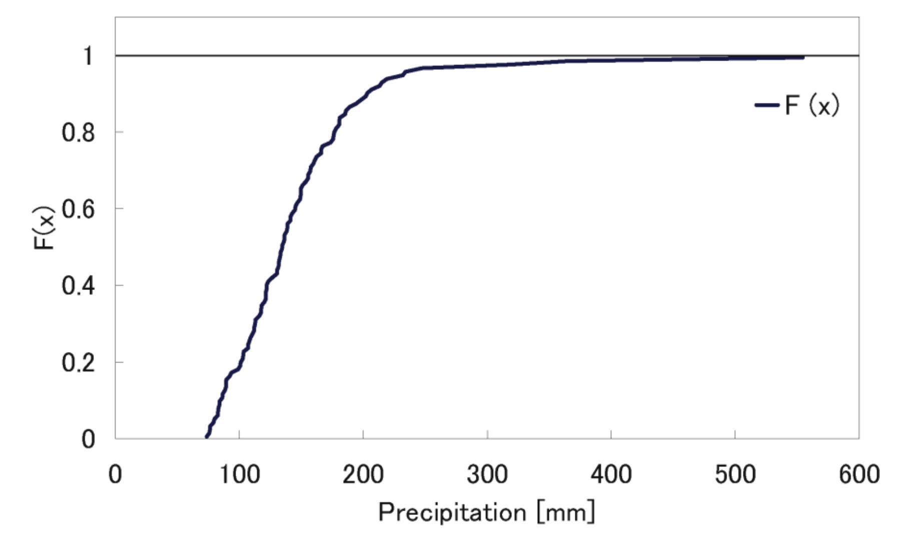
Non-parametric method proposed by Takara (2006)

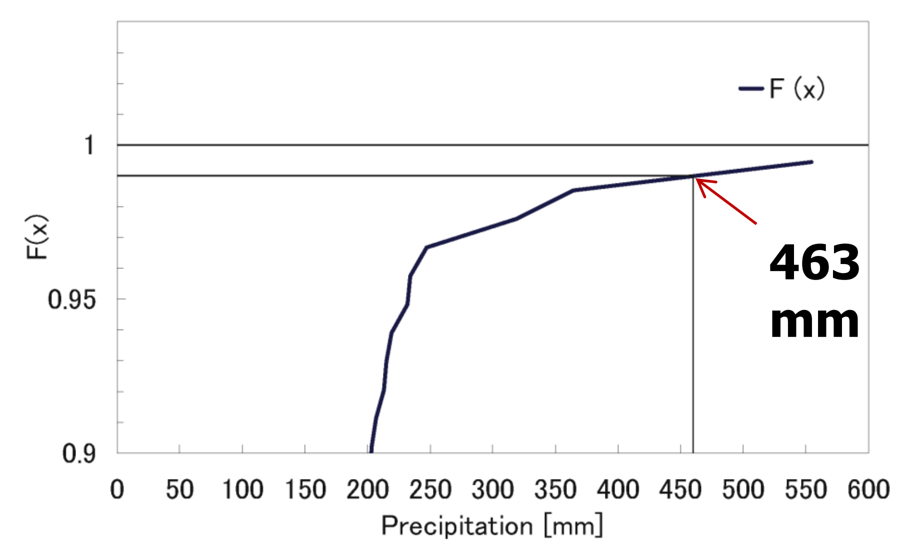
Step 1: Plot annual maximum series ($N > T$) with normal axes, using Cunnane plotting position formula: $(i-0.4)/(N+0.2)$.

Step 2: Connect all the points to obtain the empirical distribution.

Step 3: Estimate a quantile x_T with Non-exceedance probability $F(x)=1-1/T$ (T: return period) by linear-interpolation.

Step 4: Correction of the bias of quantile estimates by the Bootstrap: x_T = 430 mm





A 108-year sequence of annual maximum 2-day rainfalls in the Ane River basin

Extreme events are brought by a global climate system

A working hypothesis:
"The global climate system is changing."

If so, how we can deal with historical extreme events?

Question:
Equal probability ?

$$x_1, x_2, \cdots, x_N$$

Traditional frequency analysis has been dealing with the realizations with equal probability of 1/N.

Is this OK under global climate change situation?

DPRI-KU

Equal probability
to
Unequal probability

$$x_1, x_2, \cdots, x_N$$

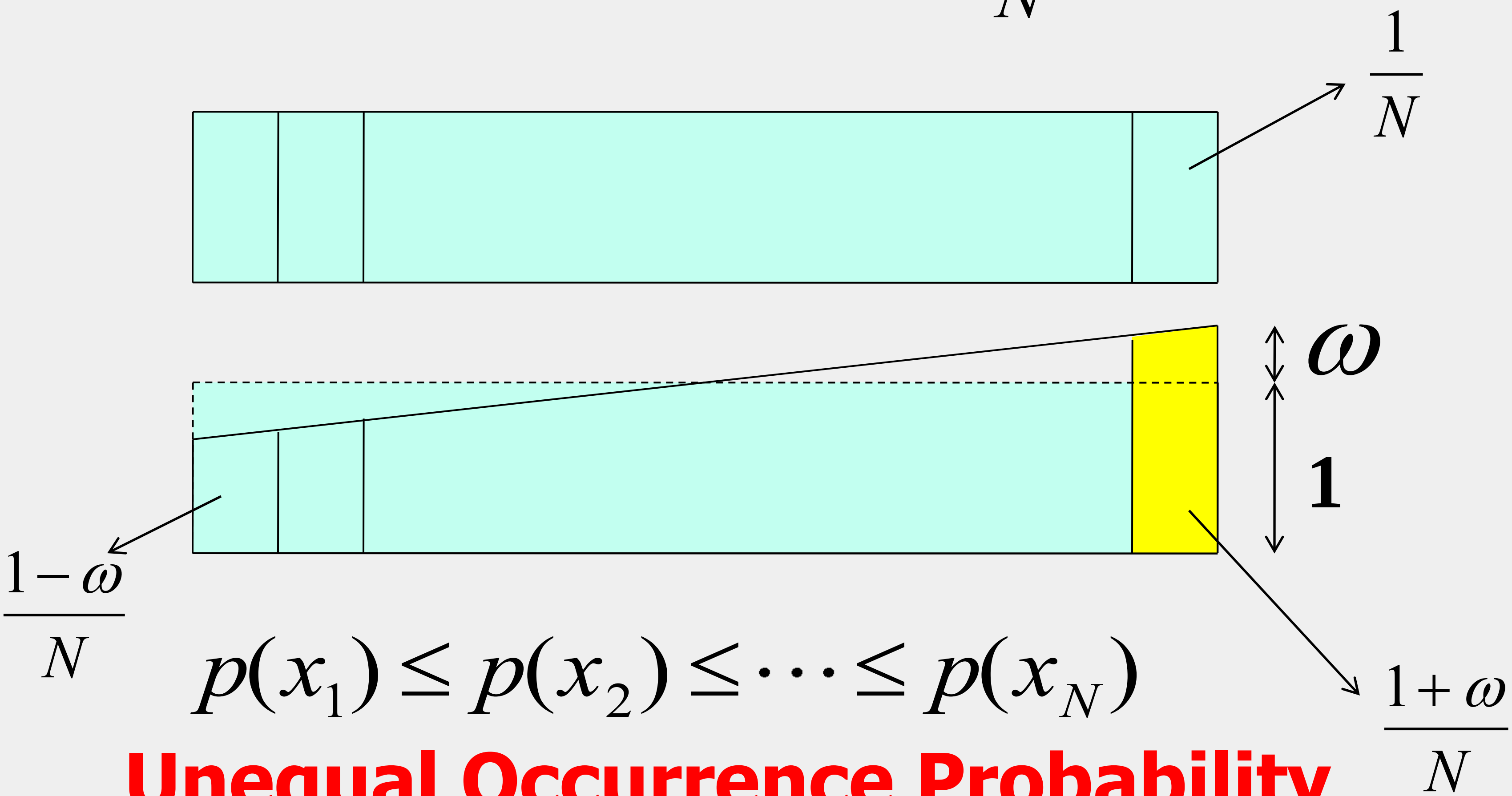
$$p(x_1) = p(x_2) = \cdots = p(x_N)$$

$$\Downarrow$$

$$p(x_1) \le p(x_2) \le \cdots \le p(x_N)$$

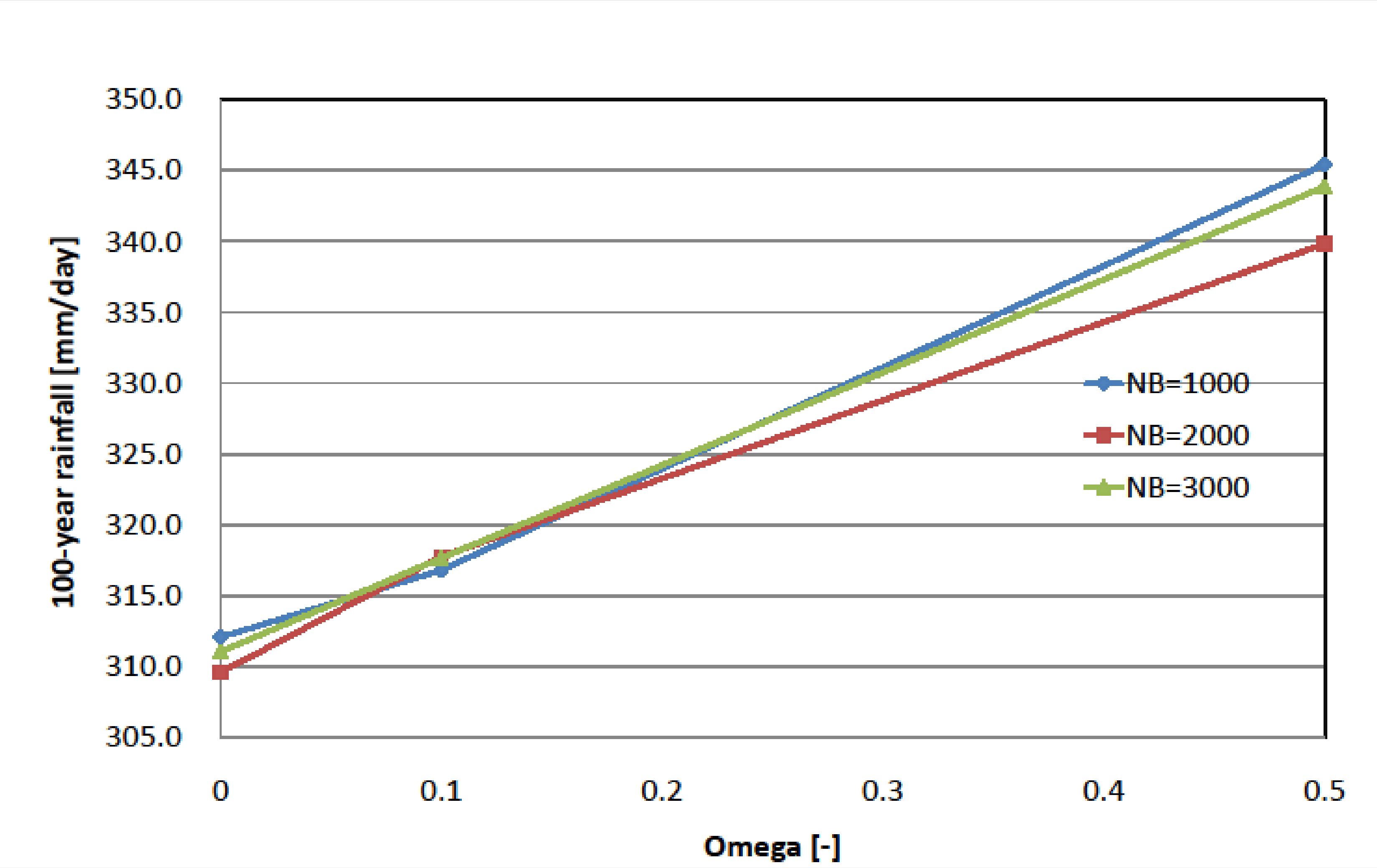
Recent climatic system affects recent events more

$$p(x_1) = p(x_2) = \cdots = p(x_N) = \frac{1}{N}$$

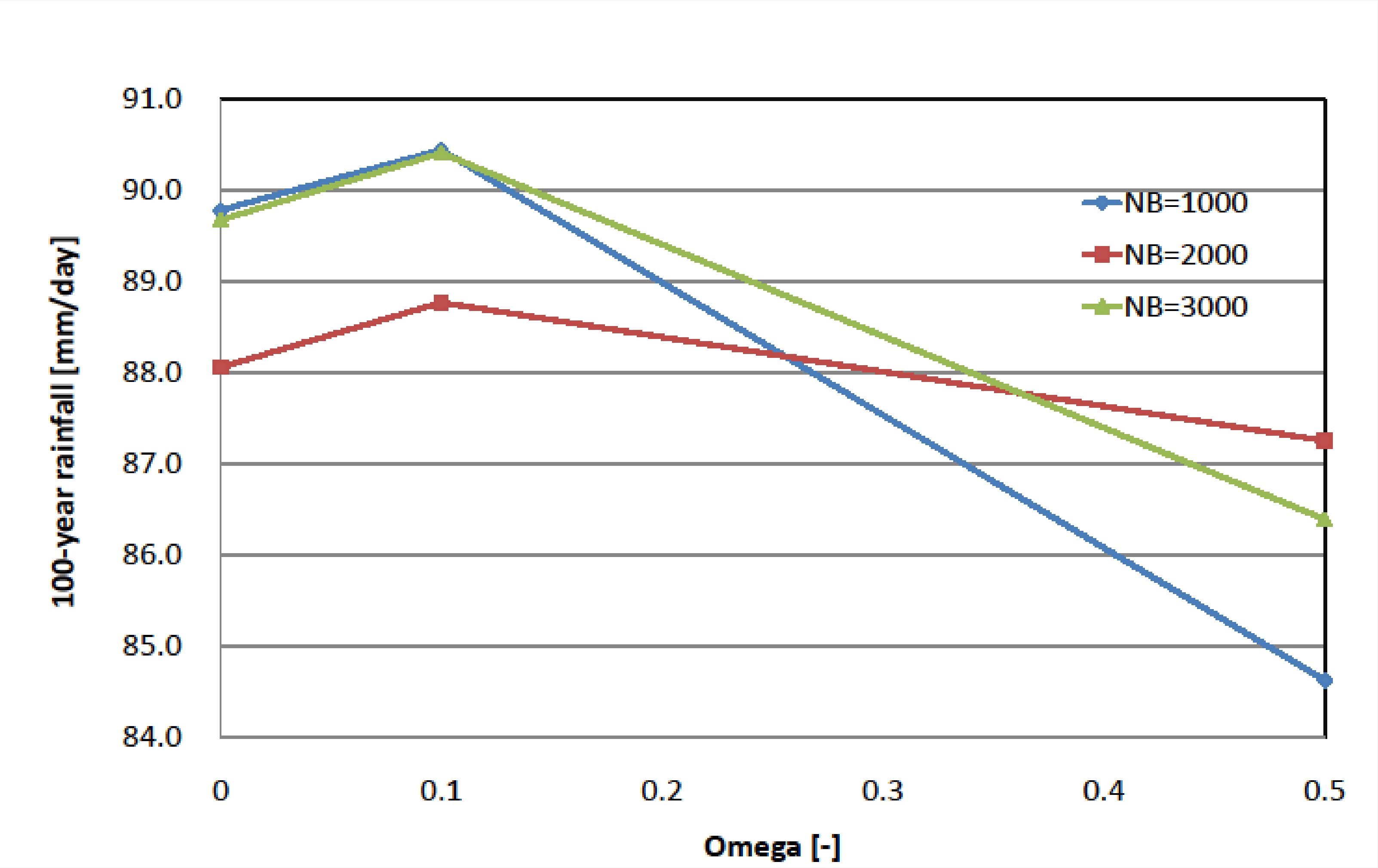


$$p(x_1) \leq p(x_2) \leq \cdots \leq p(x_N)$$

Unequal Occurrence Probability



Bootstrap estimate for different Bootstrap sample sizes NB And Omega (100-year daily rainfall at Nagoya)



Bootstrap error for different Bootstrap sample sizes NB (100-year daily rainfall at Nagoya)

Results of new "Non-Parametric" method

Method	**Gumbel**	**Non-parametric (one time)**	**Non-parametric with the Bootstrap (Bootstrap samples = 3000)**		
	Equal occurrence probability			Unequal occurrence probability	
Location			ω= 0.0	ω= 0.1	ω= 0.5
Akita (1901-2006)	168	182	172.9 (14.7)	172.3 (13.8)	166.9 (10.0)
Nagoya (1901-2006)	280	331	311.1 (89.7)	317.7 (90.4)	343.9 (86.4)
Kochi (1901-2006)	445	580	529.6 (89.9)	538.6 (87.1)	560.7 (81.2)

Results Summary

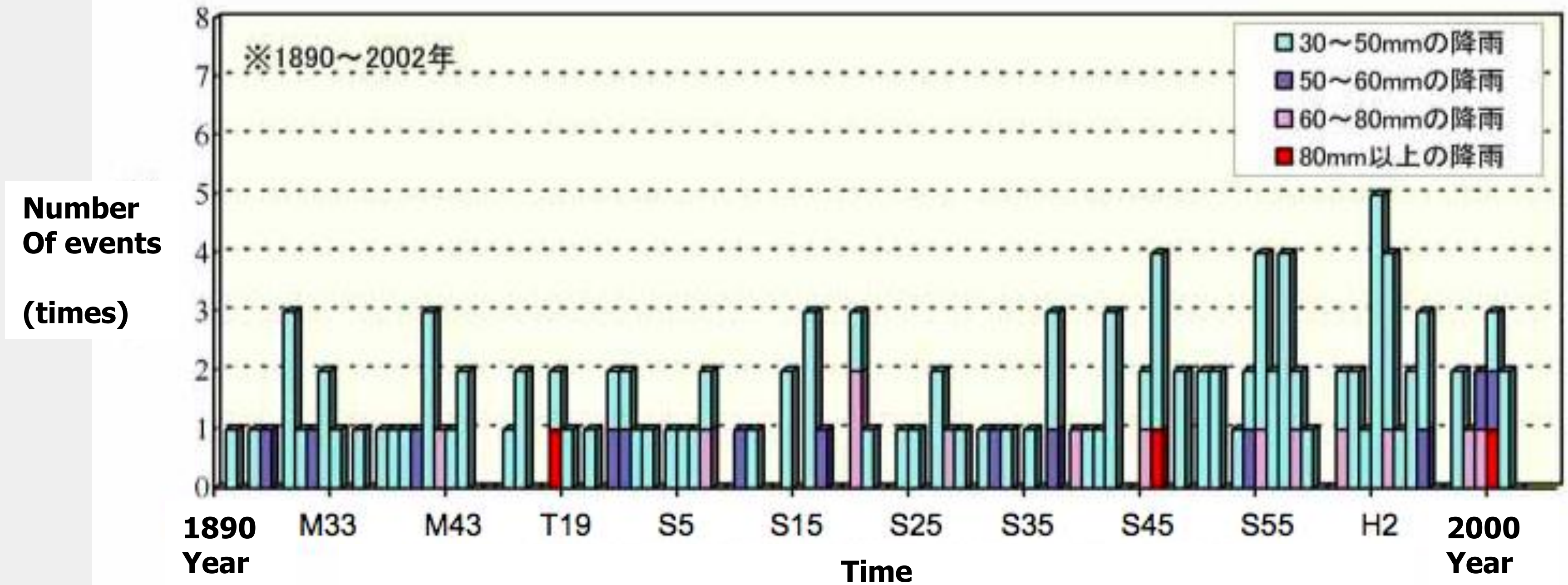
- This presentation proposed a new idea for dealing with extreme rainfall events, which are affected by the global climatic system.
- The global climatic change effect is considered as the unequal occurrence probability (UOP) concept.
- Non-parametric frequency analysis method is successfully combined with the UOP concept.
- Further discussion is necessary to understand what the results indicate.

Conclusions

- Basic theory of parametric hydrological frequency analysis was briefly reviewed.
- For samples with N>T, a non-parametric method combined with the Bootstrap resampling could give better T-year estimates than the parametric method.
- Based on a trend analysis Man-Kendall test for long-term (106 years) annual maximum rainfall series at 51 locations in Japan, 6 locations indicated increasing trend, while only one location showed decreasing trend.
- The non-parametric method is modified for adapting the climate change effect by introducing “Non-equal probability of occurrence”, which means that the recent events have more possibility to recur. The probability weighting factor OMEGA should be connected to Climate Change scenarios such as A1B and B2.

TREND ANALYSES OF HYDROLOGIC EXTREME EVENTS

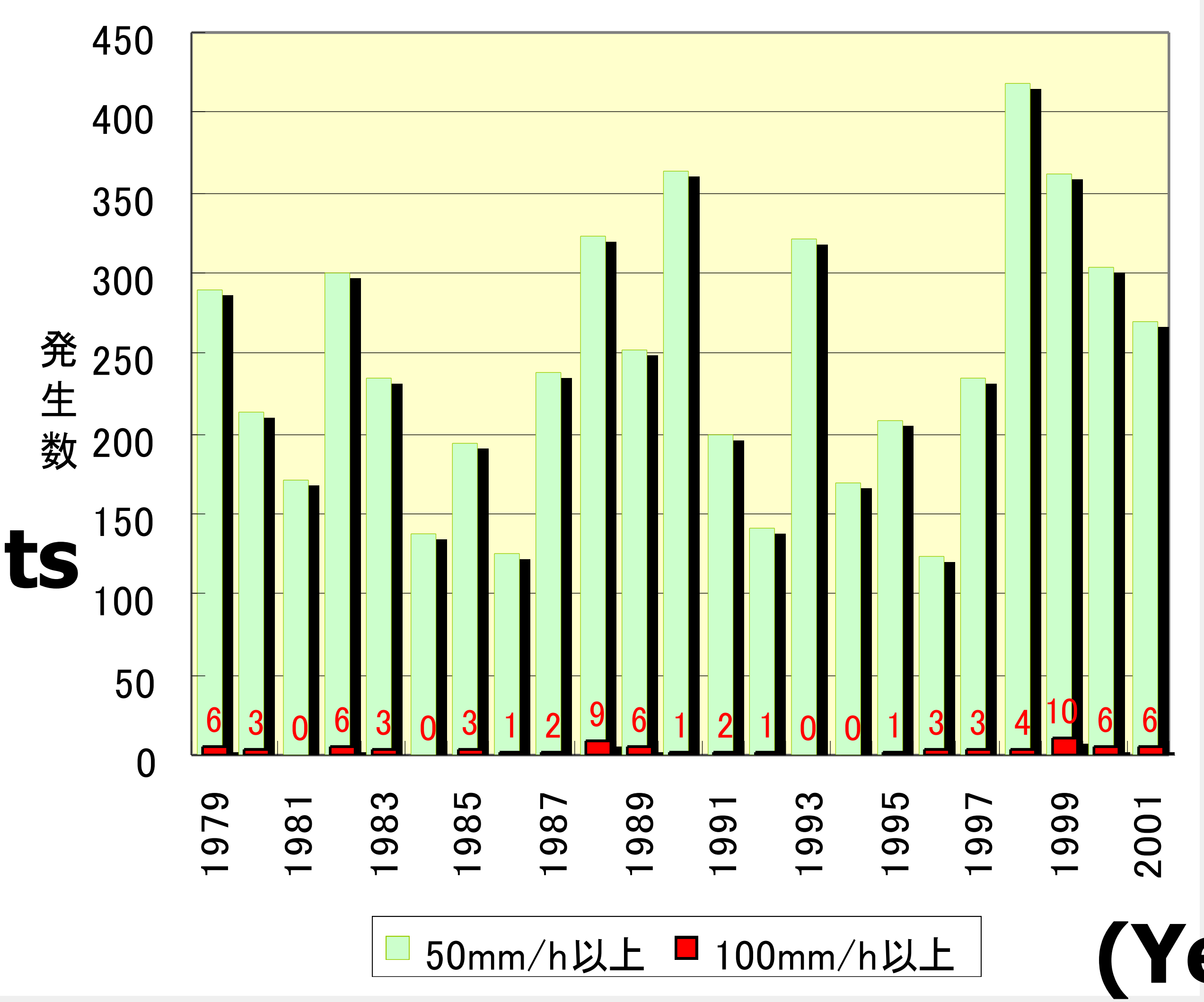
Short-term rainfall is amplified (Example of Nagoya City)



名古屋市

Number of heavy rainfalls (AMeDAS data)
More than 50 mm/h and 100 mm/h (red)

No. Of Events

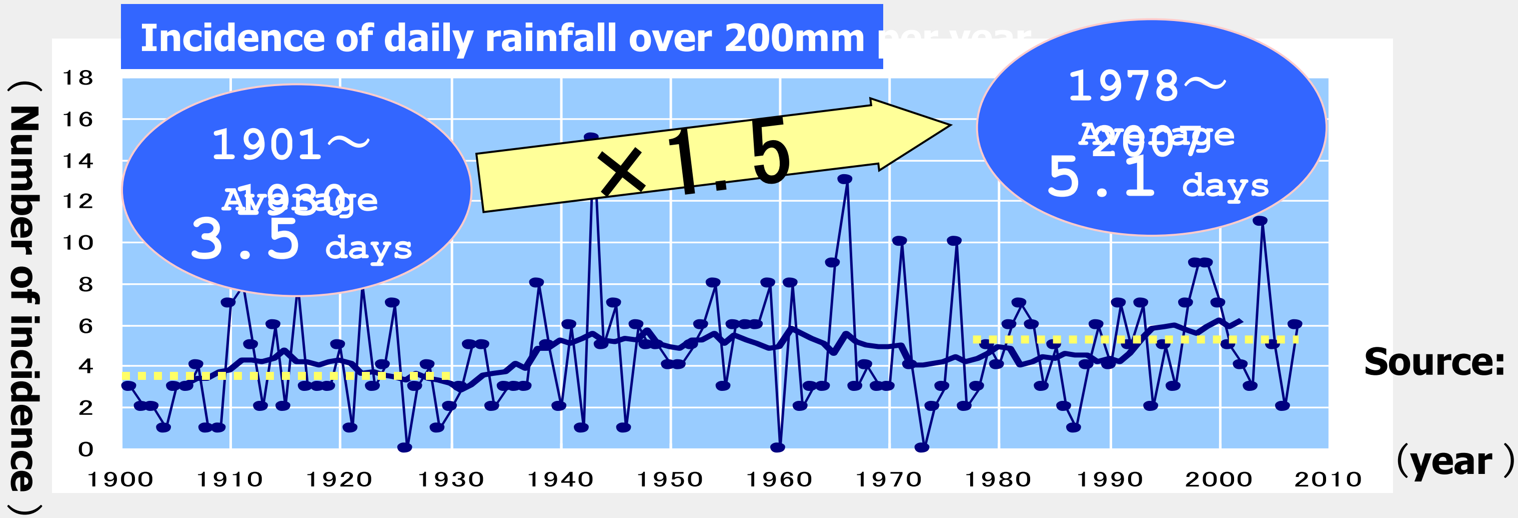


(Year)

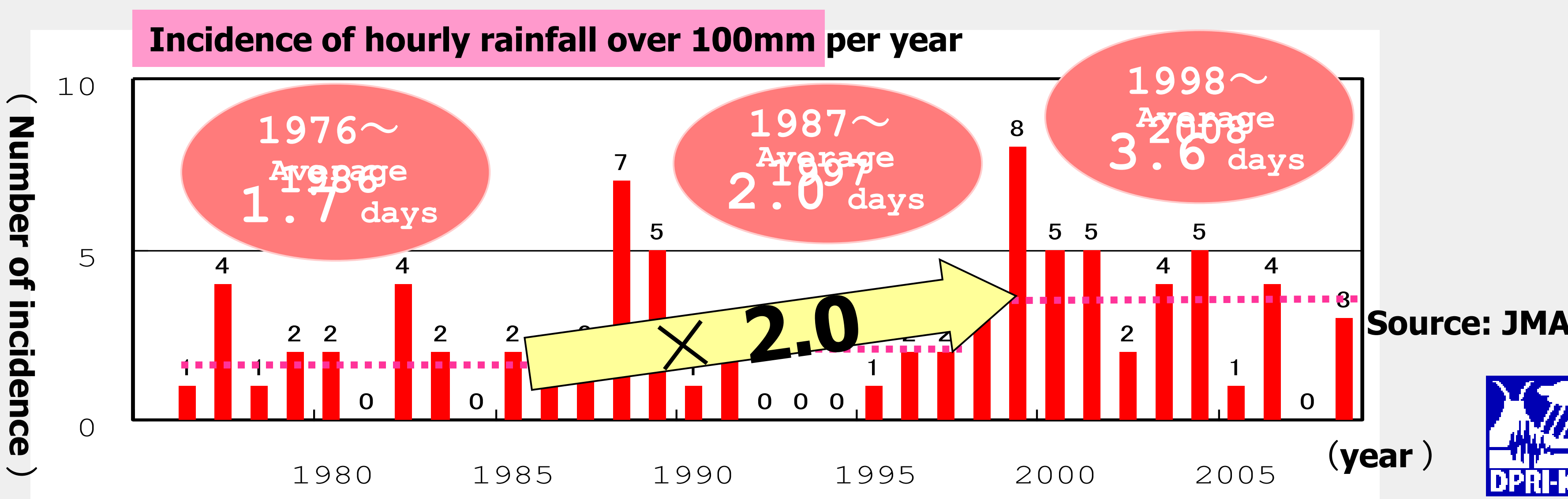
AMeDAS has 1300+ raingauges all over Japan.

Recent Change in Climate

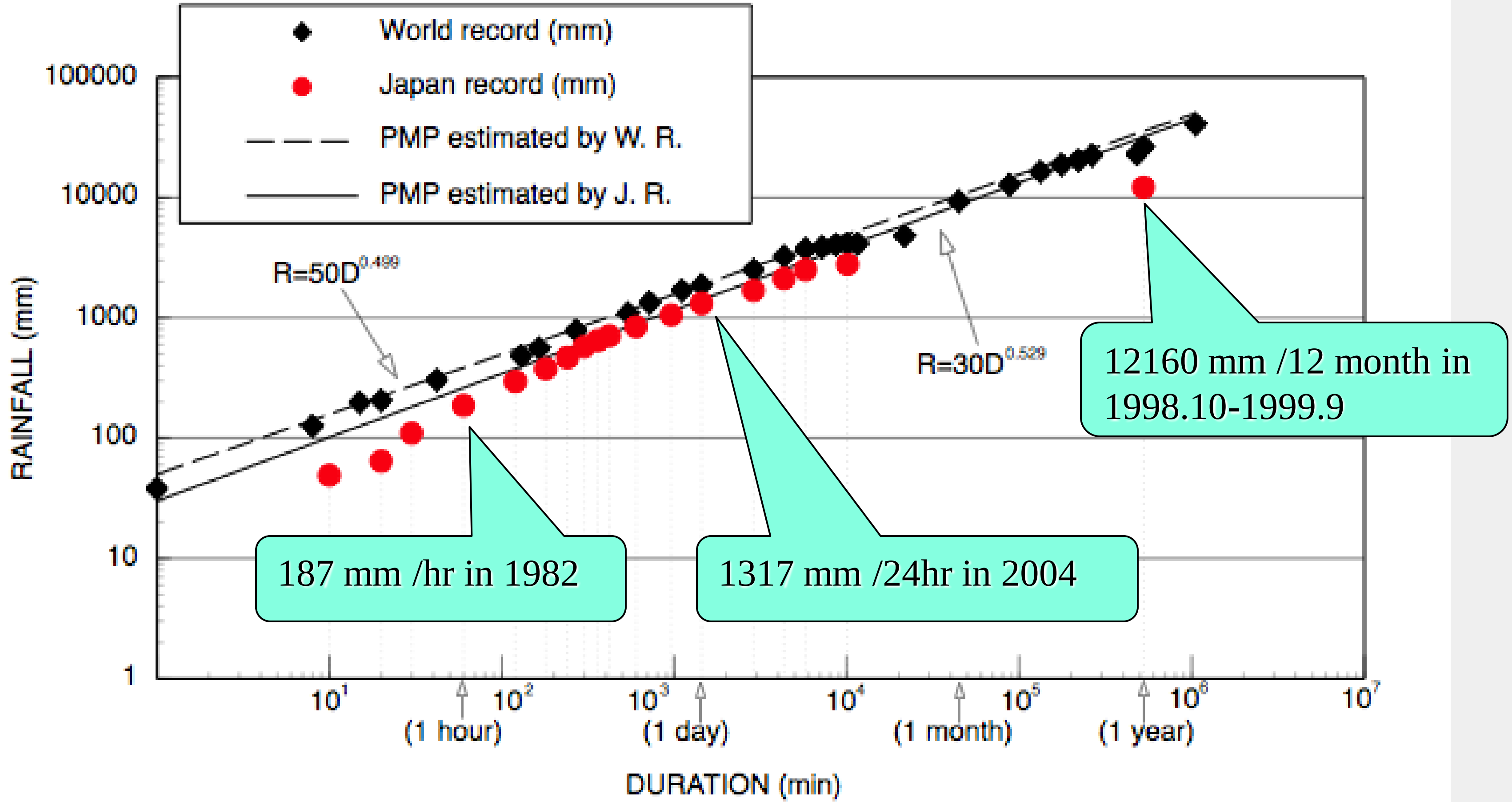
Daily rainfall over 200 mm is significantly increasing



Hourly rainfall over 100 mm is increasing



Precipitation records in the World and in Japan and the probable maximum precipitation (PMP)



Recent record-breaking precipitations in Japan are also indicated.

Impacts of GW on water-related disasters (MLIT)

Increase of CO_2 concentration in the air causes air temperature rise.

Melting of glaciers, ice caps and ice sheets

Thermal expansion of sea water

Change in evapotranspiration

Stronger typhoons

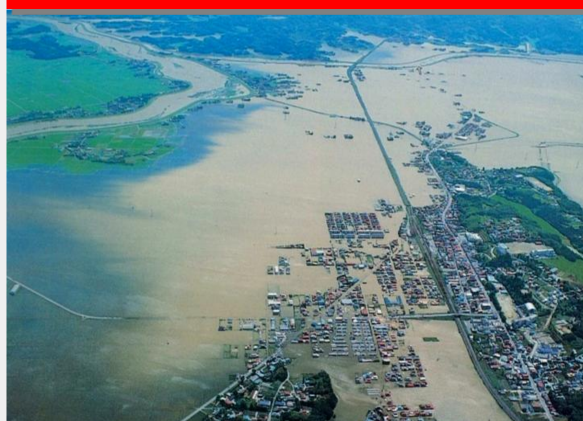
Precipitation increase

More frequent heavy rains and droughts

Increase of river flow

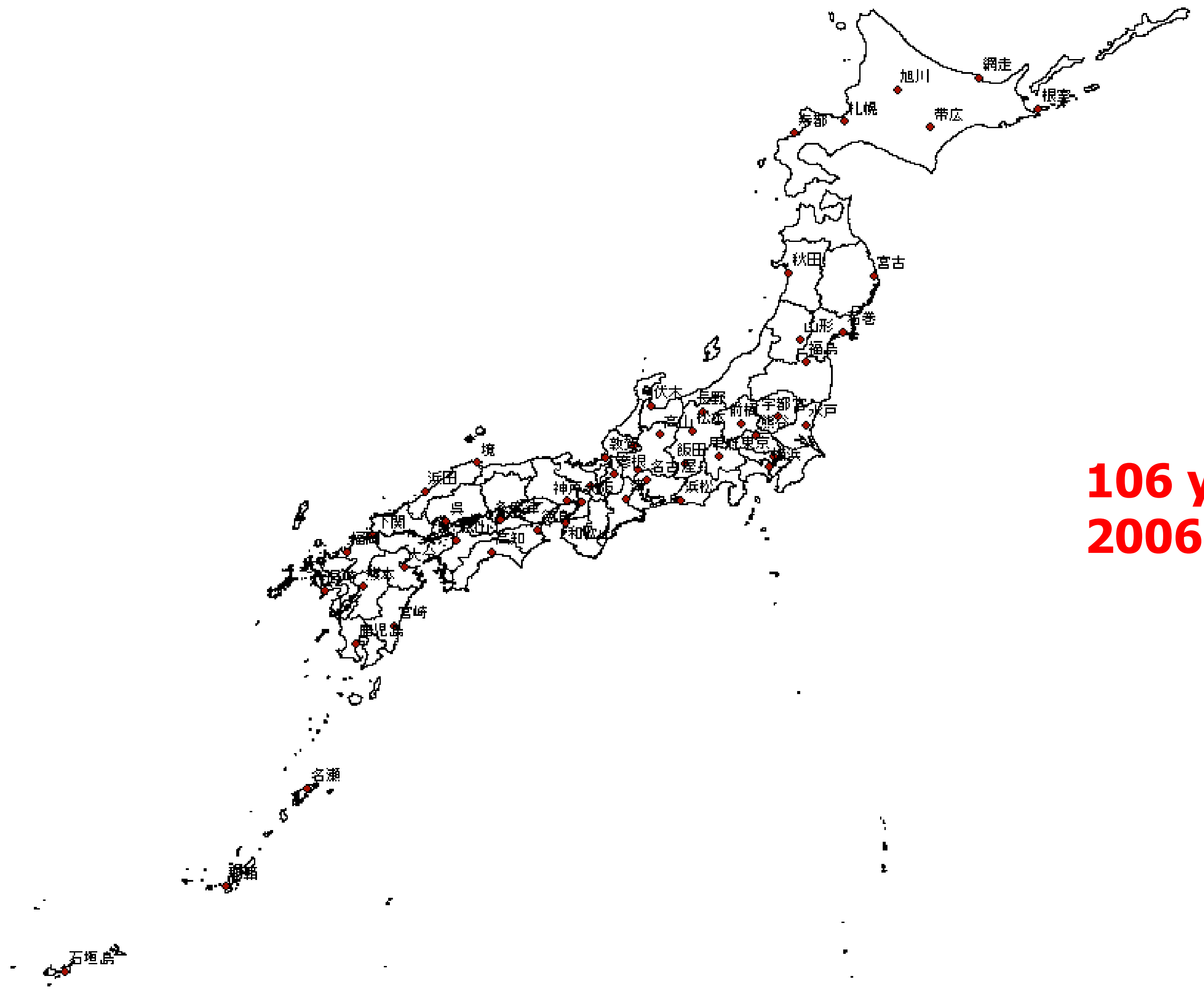
Frequent high tides and coastal erosions

Frequent floods



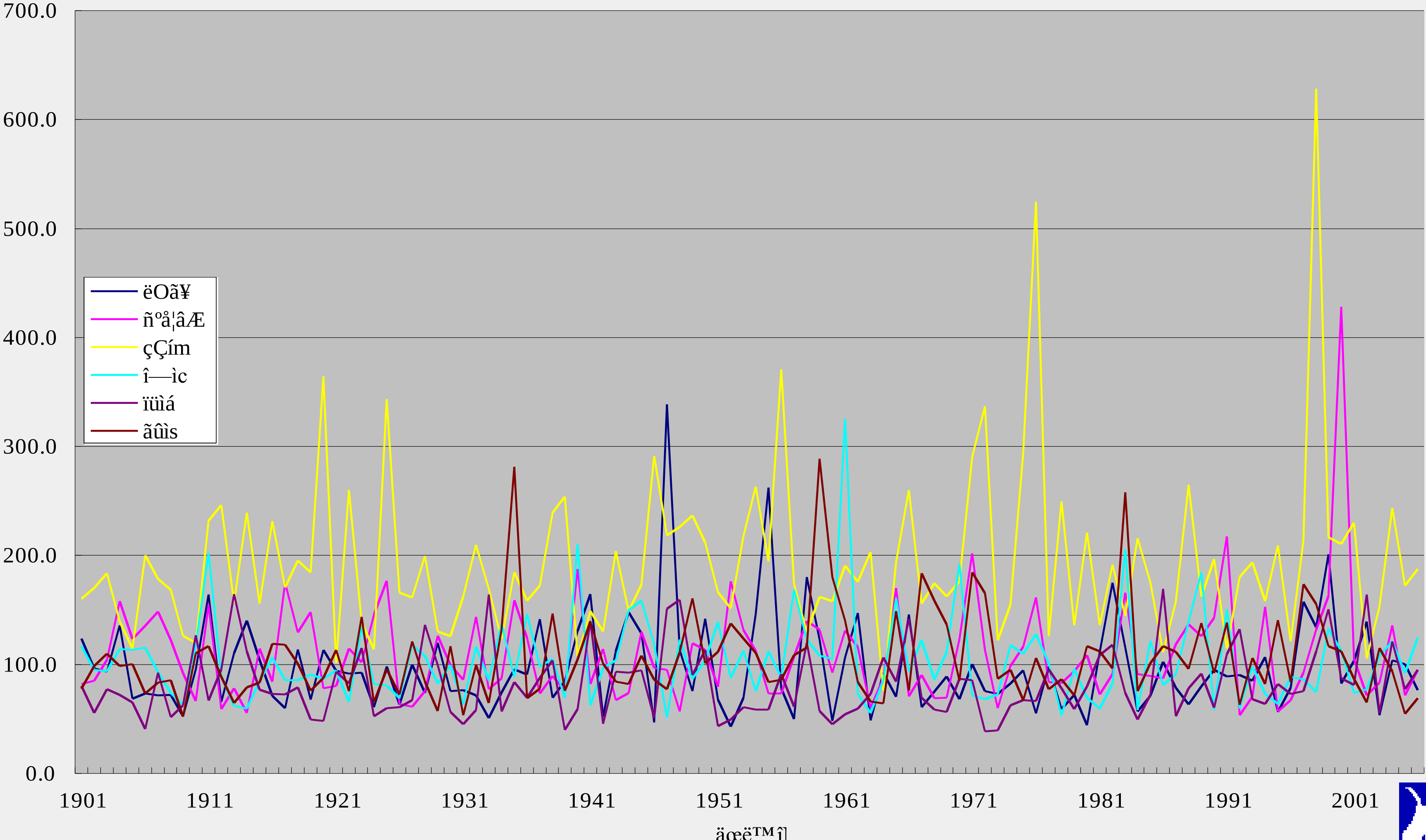
Serious debris flow

51 JMA raingauges with a long history

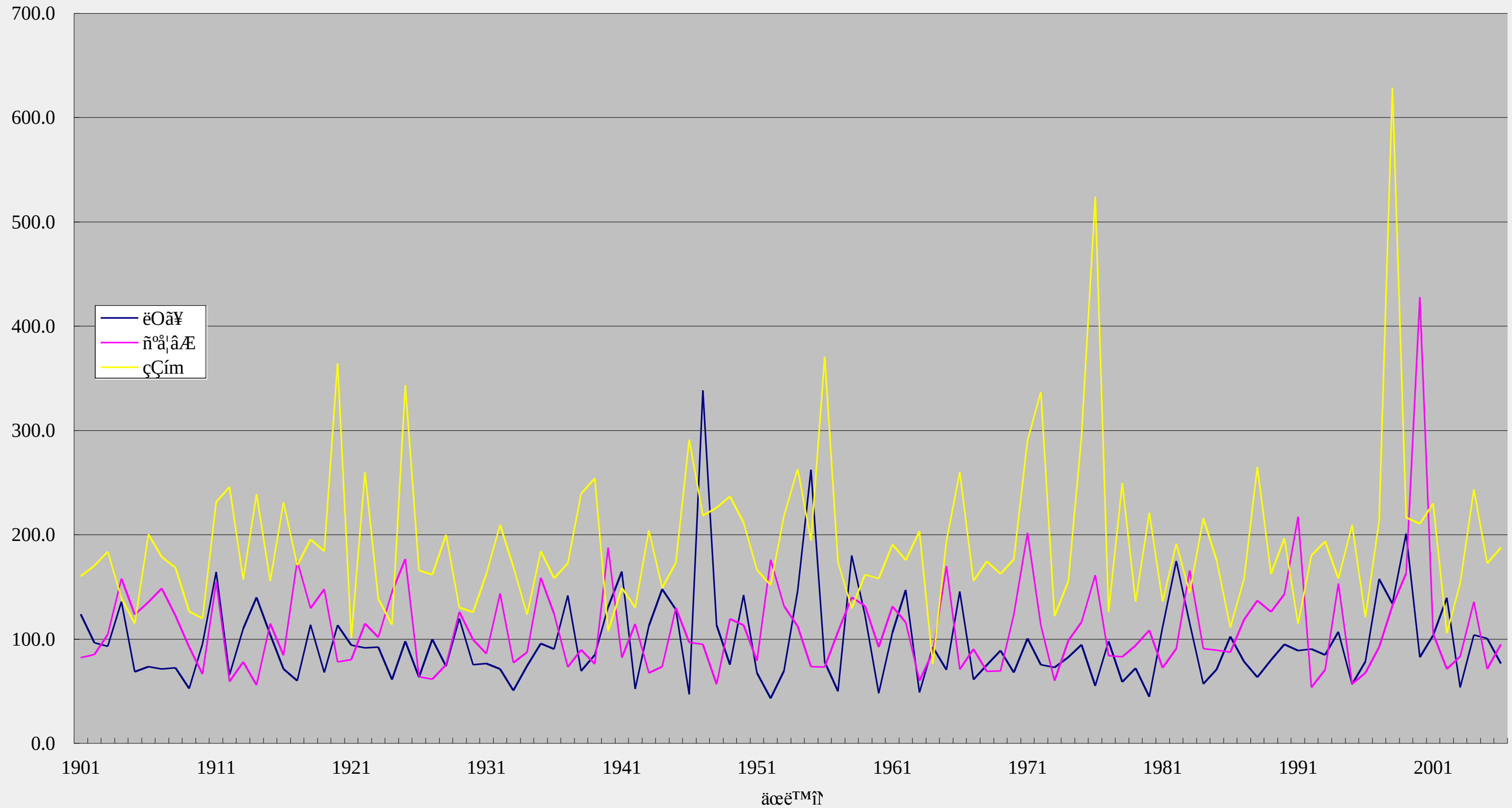


106 years (1901-2006)

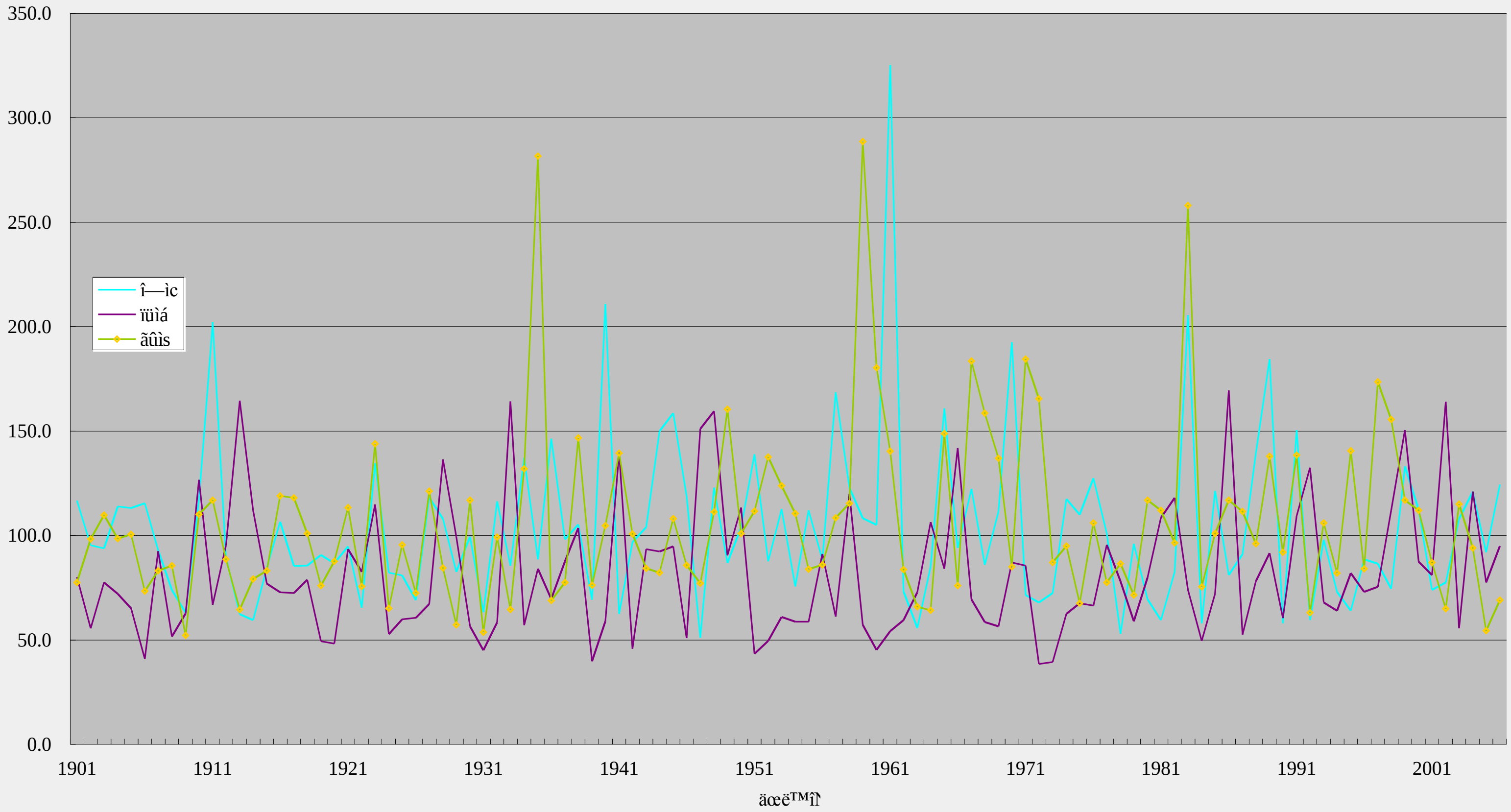
Annual Maximum series at Maebashi, Nagoya, Kochi, Iida, Fukushima and Kyoto: 1901-2006



Annual Maximum series at Maebashi (Black), Nagoya (red) and Kochi (yellow):1901-2006



Annual Maximum series at Iida (blue), Fukushima (red) and Kyoto (yellow): 1901-2006



The year of the maximum daily precipitation event

Two-decade period	# Locations	Location names
1901-1920	7	Yamagata, Takayama, Matsumoto, Hamamatsu, Shimonoseki, Tokushima, Naze
1921-1940	6	Akita, Ishinomaki, Fukui, Mito, Miyazaki, Ishigakijima
1941-1960	13	Asahikawa, Utsunomiya, Maebashi, Kumagaya, Kofu, Tokyo, Yokohama, Kyoto, Osaka, Matsuyama, Fukuoka, Kumamoto, Naha
1961-1980	7	Suttsu, Tsuruga, Gifu, Tsuruga, Gifu, Iida, Hikone, Kobe, Kure
1981-2000	14	Abashiri, Sapporo, Obihiro, Nemuro, Miyako, Fukushima, Nagoya, Hamada, Wakayama, Tadotsu, Kochi, Oita, Nagasaki, Kagoshima
2001-2006	4	Fushiki, Nagano, Tsu, Sakai

More than half (26) of the 51 locations have the historical maximum event before 1960.

Estrangement of the maximum of annual maximum series of daily precipitation

$$e = \frac{x_{\max} - \bar{x}}{s}$$

$\bar{x}$: mean of annual maxima

s : sample standard deviation

Statistics at some locations

City name	Maximum (mm)	Year of event	Estrangement
Nagoya	428.0	2000	6.67
Kochi	628.5	1998	5.59
Iida	325.3	1961	5.52
Maebashi	338.7	1947	5.50
Nagasaki	448.0	1982	4.35
Kyoto	288.6	1959	4.34
Fukushima	169.5	1986	2.57

Mann-Kendall Test (1)

Mann-Kendall Test verifies the trend in a time series statistically.

For x_1, x_2, ..., x_n (x_j , j=1, …, n), Mann-Kendall statistics S is defined as

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \operatorname{sgn}(x_j - x_k)$$

where

$$\operatorname{sgn}(x_j - x_k) = \begin{cases} 1 & x_j - x_k > 0 \\ 0 & x_j - x_k = 0 \\ -1 & x_j - x_k < 0 \end{cases}$$

Mann-Kendall Test (2)

$$E[S]=0$$

$$VAR[S]=\frac{n(n-1)(2n+5)}{18}-\sum_{j=1}^{k}\frac{t_j(t_j-1)(2t_j+5)}{18}$$

$$Z=\begin{cases}\dfrac{S-1}{\sqrt{VAR[S]}} & if\ S>0\\ 0 & if\ S=0\\ \dfrac{S+1}{\sqrt{VAR[S]}} & if\ S<0\end{cases}$$

Mann-Kendall Test

Z＞1.96 : Increasing trend
with a confidence level of 95％

Z＜-1.96 : Decreasing trend
with a confidence level of 95％

Mann-Kendall Test Results（1）

	S	Z	Beta	90% significance	95% significance
Asahikawa 旭川	135	0.37	0.03	1.64	1.96
Abashiri 網走	106	0.29	0.02	1.64	1.96
Sapporo 札幌	400	1.09	0.09	1.64	1.96
Obihiro 帯広	295	0.80	0.07	1.64	1.96
Nemuro 根室	231	0.63	0.06	1.64	1.96
Suttsu 寿都	553	1.51	0.09	1.64	1.96
Akita 秋田	-744	-2.03	-0.15	1.64	1.96
Miyako 宮古	257	0.70	0.08	1.64	1.96
Yamagata 山形	252	0.69	0.05	1.64	1.96
Ishnomaki 石巻	-305	-0.83	-0.07	1.64	1.96
Fukushima 福島	1481	4.04	0.50	1.64	1.96
Fushiki 伏木	841	2.29	0.19	1.64	1.96
Nagano 長野	203	0.55	0.03	1.64	1.96
Utsunomiya 宇都宮	359	0.98	0.11	1.64	1.96
Fukui 福井	185	0.50	0.05	1.64	1.96
Takayama 高山	-84	-0.23	0.00	1.64	1.96
Mtasumoto 松本	250	0.68	0.04	1.64	1.96
Maebashi 前橋	83	0.22	0.03	1.64	1.96
Kumagaya 熊谷	83	0.22	0.03	1.64	1.96
Mito 水戸	525	1.43	0.14	1.64	1.96
Tsuruga 敦賀	-472	-1.29	-0.12	1.64	1.96
Gifu 岐阜	163	0.44	0.05	1.64	1.96
Nagoya 名古屋	-20	-0.05	-0.01	1.64	1.96
Iida 飯田	11	0.03	0.00	1.64	1.96
Kofu 甲府	-422	-1.15	-0.12	1.64	1.96

Mann-Kendall Test Results (2)

	S	Z	Beta	90% significance	95% significance
Tsu 津	-56	-0.15	-0.03	1.64	1.96
Hamamatsu 浜松	-249	-0.68	-0.08	1.64	1.96
Tokyo 東京	126	0.34	0.05	1.64	1.96
Yokohama 横浜	332	0.90	0.15	1.64	1.96
Sakai 境	876	2.39	0.30	1.64	1.96
Hamada 浜田	444	1.21	0.12	1.64	1.96
Kyoto 京都	588	1.60	0.14	1.64	1.96
Hikone 彦根	203	0.55	0.03	1.64	1.96
Shimonoseki 下関	254	0.69	0.07	1.64	1.96
Kure 呉	550	1.50	0.17	1.64	1.96
Kobe 神戸	131	0.35	0.04	1.64	1.96
Osaka 大阪	455	1.24	0.11	1.64	1.96
Wakayama 和歌山	-259	-0.70	-0.09	1.64	1.96
Fukuoka 福岡	464	1.26	0.17	1.64	1.96
Oita 大分	166	0.45	0.07	1.64	1.96
Nagasaki 長崎	1573	4.29	0.50	1.64	1.96
Kumamoto 熊本	1144	3.12	0.53	1.64	1.96
Kagoshima 鹿児島	491	1.34	0.21	1.64	1.96
Miyazaki 宮崎	-63	-0.17	-0.03	1.64	1.96
Matsuyama 松山	195	0.53	0.05	1.64	1.96
Tadotsu 多度津	146	0.40	0.04	1.64	1.96
Kochi 高知	1564	4.27	0.50	1.64	1.96
Tokushima 徳島	554	1.51	0.24	1.64	1.96
Naze 名瀬	-172	-0.47	-0.13	1.64	1.96
Ishigakijima 石垣島	441	1.20	0.21	1.64	1.96
Naha 那覇	56	0.15	0.03	1.64	1.96