

## Part II Singular Hydrologic Amount and Rejection Test

### 1. Introduction

It is needless to say that the statistical treatment of the hydrologic data should be rational in the sense of stochastics to be significant as a useful tool for the hydraulic works. Usually, the very large or small data which seem to be singular in an ordinary sense are to be contained in a family of hydrologic observations. In hydrologic frequency analysis, the rejection test of such data is essential in the sense of stochastics, while evaluation of the singular value defined later is important in the sense of engineering. It seems, however, that the course of treatment for such problems is not always successfully planned, in the field of hydrologic statistics. In this part, a theoretical and practical method of treatment for such problems is developed<sup>10,20</sup>.

### 2. Definition of singular hydrologic amount

In a family of hydrologic data of size  $N$ , the very large variate  $x_{\varepsilon u}$  or the very small variate  $x_{\varepsilon l}$  is often contained, whose population probability of exceedance  $1-F(x_{\varepsilon u})$  or of nonexceedance  $F(x_{\varepsilon l})$  is very small.

According to the help of the sample theory, the probability  $\varepsilon \equiv P_N(x \geq x_{\varepsilon u})$  or  $P_N(x \leq x_{\varepsilon l})$  with which such a variate is contained in a sample of size  $N$  is greater than the population probability  $1-F(x_{\varepsilon u})$  or  $F(x_{\varepsilon l})$ , under the condition of finite sample size  $N$ .

$$\left. \begin{aligned} \varepsilon &\equiv P_N(x \geq x_{\varepsilon u}) \geq 1 - F(x_{\varepsilon u}) \\ \varepsilon &\equiv P_N(x \leq x_{\varepsilon l}) \geq F(x_{\varepsilon l}) \end{aligned} \right\} \quad (2.1)$$

Or, for a fixed value  $x_\varepsilon$ , the smaller the sample size  $N$  is, the larger the probability  $\varepsilon$  becomes in spite of the fixed population probability.

In hydrologic frequency analysis, the above facts should always be considered so far as the analysis is based on a sample of small size. And it is desired in the hydraulic planning and design to utilize the reasonable information obtained by the frequency analysis of the hydrologic data as far as possible.

That is, as the expected value for the desired return period  $T$ , the

value of  $x_{\varepsilon u}$  or  $x_{\varepsilon l}$  for  $\varepsilon=1/T$  rather than that for  $1-F(x)=1/T$  or  $F(x)=1/T$  should be considered for discussion. In this paper, the former estimate is named the expected singular hydrologic amount in a stochastic sense, or simply the (expected) singular value, and the latter estimate is called the expected hydrologic amount in an ordinary sense, or simply the (expected) ordinary value. Moreover, the probability  $\varepsilon$  is simply called the singular level for the singular value  $x_\varepsilon$ .

To find the theoretical relation between the singular value  $x_\varepsilon$  and its singular level  $\varepsilon$ , the sample theory should, strictly speaking, be extended on extremes, but it seems to be difficult since a satisfactory sample theory has not yet been developed. Therefore, in this paper, the theory of rejection limit in the sense of Thompson, which is obtained as a special case of the two-sample theory on normals, will be utilized as a convenient approach. This idea is all but analogous in an essential sense to that adopted by Ogawara<sup>21)</sup>, by which he studied the stochastic limit for the maximum possible amount of precipitation.

Now, let  $\bar{\eta}$  and  $S_\eta^2$  be the sample values of mean and variance, respectively, in a sample of size  $N$  from a normal population  $N(\mu, \sigma^2)$ , defined as

$$\left. \begin{aligned} \bar{\eta} &= \frac{1}{N} \sum_i \eta_i \\ S_\eta^2 &= \frac{1}{N} \sum_i (\eta_i - \bar{\eta})^2 \end{aligned} \right\} \quad (2.2)$$

If  $\eta_\varepsilon$  is another sample of size 1 which is obtained from the same population independent of the above-state sample, the following equations are satisfactory clearly,

$$\begin{aligned} E(\eta_\varepsilon - \bar{\eta}) &= 0 \\ E(\eta_\varepsilon - \bar{\eta})^2 &= \frac{N}{N+1} \sigma^2 \end{aligned}$$

The statistics

$$\chi_1^2 = \frac{NS_\eta^2}{\sigma^2} \quad \text{and} \quad \chi_2^2 = \frac{N}{N+1} \frac{(\eta_\varepsilon - \bar{\eta})^2}{\sigma^2}$$

are independent of each other and they follow the  $\chi^2$ -distribution of freedom  $N-1$  and 1, respectively, Therefore, the statistics

$$F = \frac{N-1}{N+1} \frac{(\eta_\varepsilon - \bar{\eta})^2}{S_\eta^2} \quad (2.3)$$

must follow the  $F$ -distribution of freedom 1 and  $N-1$ .

Next, let  $F_{N-1}^{1-2\mathcal{E}}$  be value of  $F$  for a given singular level  $\mathcal{E}$  which corresponds to the level of significance of one tail in the case of the rejection test on a normal sample. Then Eq. (2.3) becomes,

$$\left. \begin{array}{l} \eta_{\varepsilon u} \\ \eta_{\varepsilon l} \end{array} \right\} = \bar{\eta} \pm S_{\eta} \sqrt{\frac{N+1}{N-1} F'_{N-1}(2\mathcal{E})} \quad (2.4)$$

where  $\eta_{\varepsilon u}$  and  $\eta_{\varepsilon l}$  are the values corresponding to the singular level  $\mathcal{E}$ , provided on the normals.

Generally, the following relation is satisfactory for the upper tail.

$$\begin{aligned} \Phi(\eta_{\varepsilon u}) &= \frac{1}{\sqrt{2\pi} S_{\eta}} \int_{\eta_{\varepsilon u}}^{\infty} \exp\left\{-\frac{(\eta - \bar{\eta})^2}{2S_{\eta}^2}\right\} d\eta \leq \Phi_0(\eta_{\varepsilon u}) = \\ &= \frac{1}{\sqrt{2\pi} \sigma} \int_{\eta_{\varepsilon u}}^{\infty} \exp\left\{-\frac{(\eta - m)^2}{2\sigma^2}\right\} dy \leq \mathcal{E} \end{aligned} \quad (2.5)$$

And the similar relation to it for the lower tail can be presented. It can be easily considered that there is a simple relation between the singular level  $\mathcal{E}$  and the value of  $\Phi(\eta_{\varepsilon u})$  or  $\Phi_0(\eta_{\varepsilon u})$ , which will be defined as a function of sample size  $N$  if a relation as Eq. (2.4) can be made available.

The above discussions provided for the normals can be easily extended to the extremes.

### 3. Estimation of singular hydrologic amount based on extremes

If all parameters included in the asymptotes for the largest value distribution which are expressed as

$$\left. \begin{array}{l} F(x) = \exp(-e^{-y}), \\ \text{for the first asymptote ; } y = a(x-u), \\ \text{for the second asymptote ; } y = a \log(x+b)/(u+b), \end{array} \right\} \quad (2.6)$$

are known, the expected hydrologic amount can be obtained by the following equations,

$$\left. \begin{array}{l} \text{for the first asymptote ; } x = u + (1/a)y \\ \text{for the second asymptote ; } \log(x+b) = \log(u+b) + (1/a)y \end{array} \right\} \quad (2.7)$$

And as is well known, in estimating the amount for the desired return period  $T$  in an ordinary sense, the value of  $y$  defined by the following equation must be adopted in Eq. (2.7)

$$y = -\lg\{\lg T/(T-1)\} \quad (2.8)$$

Table 2.1 Reduced extremes  $y$  for given return period.

$T$	$1-F(x)$	$y$
2000	0.00050	7.60065
1000	100	6.90725
500	200	6.21361
400	250	5.99021
300	0.00333	5.70212
250	400	5.51946
200	500	5.29581
150	667	5.00730
100	0.01000	4.60015
80	1250	4.37574
60	1667	4.08596
50	2000	3.90194
40	0.02500	3.67625
30	3333	3.38429
25	4000	3.19853
20	5000	2.97020
10	0.10000	2.25037
8	12500	2.01342
7	14286	1.86982
6	16667	1.70199
5	0.20000	1.49994
4	25000	1.24590
3	33333	0.90273
2	50000	0.36651
	63212	0

These values are tabulated in Table 2.1.

Now, let  $y$  in Eq. (2.7) be denoted as  $y_\varepsilon$ , which should be especially used for evaluation of the expected singular hydrologic amount for the desired return period  $T=1/\varepsilon$  in a stochastic sense as discussed in the preceding section. Then, the value of  $y_\varepsilon$  can be obtained from Eqs. (2.4) and (2.5) by the transformation of variables as follows :

$$\left. \begin{array}{l} \text{For upper tail ; } \exp(-e^{-y_\varepsilon u}) = F(y_\varepsilon u) \equiv 1 - \Phi(\eta_{\varepsilon u}) \\ \text{For lower tail ; } \exp(-e^{-y_\varepsilon l}) = F(y_\varepsilon l) \equiv \Phi(\eta_{\varepsilon l}) \end{array} \right\} \quad (2.9)$$

Table 2.2 Reduced singular extremes  $y_\epsilon$  for given upper singular levels  $\epsilon$ .

$N-1$	upper singular level $\epsilon$							
	25%	12.5%	5%	2.5%	1.25%	0.5%	0.25%	0.05%
20	1.314	2.179	3.328	4.237	5.200	6.557	7.666	10.5
22	308	163	294	182	113	425	483	10.2
24	303	150	265	137	047	319	342	9.94
26	298	140	241	098	4.991	230	224	72
28	294	130	220	067	943	155	124	56
30	291	122	203	039	903	091	039	40
32	1.289	2.116	3.188	4.015	4.867	6.038	6.969	9.27
34	286	110	175	3.994	836	5.991	904	16
36	284	104	163	976	808	950	849	08
38	282	099	153	959	785	913	799	01
40	280	095	143	944	763	880	755	8.94
42	1.278	2.091	3.135	3.931	4.742	5.850	6.715	8.86
44	277	087	127	919	726	824	680	80
46	275	084	120	908	711	801	650	74
48	274	081	113	898	695	778	620	68
50	273	079	108	889	682	756	593	63
52	1.272	2.076	3.103	3.879	4.670	5.740	6.568	8.58
54	271	073	098	873	658	721	544	54
56	270	071	093	865	647	706	523	50
58	269	069	088	859	638	690	503	46
60	268	066	084	852	628	676	486	43
65	1.267	2.063	3.076	3.839	4.610	5.645	6.443	8.36
70	265	060	068	827	592	620	410	30
75	264	057	061	817	577	598	382	25
80	263	054	056	807	565	579	359	20

Table 2.3 Reduced singular extremes  $y_\varepsilon$  for given lower singular levels  $\varepsilon$ .

$N-1$	lower singular level $\varepsilon$							
	25%	12.5%	5%	2.5%	1.25%	0.5%	0.25%	0.05%
20	-0.3688	-0.8042	-1.208	-1.446	-1.649	-1.881	-2.036	-2.350
22	647	-0.7971	197	433	632	860	012	320
24	615	921	189	422	619	844	-1.993	298
26	589	875	182	413	608	830	977	277
28	567	835	176	405	599	817	964	257
30	547	801	171	398	590	807	952	240
32	-0.3529	-0.7772	-1.166	-1.392	-1.583	-1.798	-1.941	-2.227
34	513	746	162	387	577	791	932	214
36	499	722	158	382	571	784	924	204
38	487	701	155	378	566	778	917	195
40	476	682	152	375	562	772	910	187
42	-0.3466	-0.7665	-1.149	-1.371	-1.557	-1.767	-1.905	-2.178
44	457	649	147	369	554	762	899	171
46	449	633	145	366	551	758	895	165
48	441	621	143	363	548	754	890	159
50	434	610	141	361	545	751	886	153
52	-0.3428	-0.7598	-1.139	-1.358	-1.542	-1.747	-1.882	-2.148
54	422	587	138	357	540	744	879	144
56	416	577	136	355	537	741	876	139
58	410	568	135	353	535	739	873	136
60	406	560	134	351	533	736	870	132
65	-0.3396	-0.7542	-1.131	-1.348	-1.529	-1.731	-1.863	-2.124
70	387	527	129	345	525	727	858	117
75	379	514	127	342	522	723	854	111
80	371	502	125	340	519	719	850	105

$$\text{where } \Phi(\eta_{\varepsilon t}) = \frac{1}{\sqrt{2\pi} S_{\eta}} \int_{-\infty}^{\eta_{\varepsilon t}} \exp \left\{ -\frac{(\eta - \bar{\eta})^2}{2S_{\eta}^2} \right\} d\eta$$

These values,  $y_{\varepsilon u}$  and  $y_{\varepsilon t}$ , for various singular levels,  $\varepsilon$ , are tabulated as a function of sample size  $N$  in Tables 2.2 and 2.3 for the practical facilities. It may be easily seen by the help of the values in Tables 2.1 and 2.2 that the value of  $y_{\varepsilon u}$  for  $T=1/\varepsilon$  approaches to the value of  $y$  for  $T=1/1-F(x)$  with increase of the sample size  $N$ .

It is needless to say that the expected singular hydrologic amount for the desired return period  $T=1/\varepsilon$  is obtained from Eq. (2.7) by using  $y_{\varepsilon}$  instead of  $y$ , that is,

$$\left. \begin{array}{l} \text{for the first asymptote ; } \quad x_{\varepsilon} = u + (1/a)y_{\varepsilon} \\ \text{for the second asymptote ; } \quad \log(x_{\varepsilon} + b) = \log(u + b) + (1/a)y_{\varepsilon} \end{array} \right\} \quad (2.10)$$

The necessity for such an idea or the usefulness of such a proposal as discussed above in detail, may be easily understood even from several examples of application presented by Fig. 2.1 which is shown with the proposal in the next section, and by Fig. 1.8 which has already been shown in Part I.

#### 4. Method of rejection test

In the frequency analysis of observed hydrologic data, the decision of adoption or rejection of the singular variate contained in a family of the data must be objectively made from a stochastic viewpoint, because it is the duty of the hydrologic statistics to offer the most likely information to hydraulic works.

Since the rejection test proposed in this section is based on the idea of binomial distributin, it is not necessarily new as an idea, but it may be recognized to be useful as a method.

Generally, the probability of an event that  $r$  variates which are not smaller (larger) than  $x_{\varepsilon}$  for the upper (lower) singular level  $\varepsilon$  are at least contained in a family of observed data of size  $N$  is given by

$$\begin{aligned} & P\{np(x \geq x_{\varepsilon u}) \geq r \text{ or } np(x \leq x_{\varepsilon t}) \geq r\} \\ & = 1 - \sum_{j=0}^{r-1} \frac{\Gamma(N+1)}{\Gamma(j)\Gamma(N-j+1)} (1-\varepsilon)^{N-j} \varepsilon^j \end{aligned} \quad (2.11)$$

If this probability is smaller than a certain level  $\beta_0$  of significance, the danger of rejecting such variates under the hypothesis that the event is very rare may be said to be 100  $\beta_0$  % at most.

Up to now, if the singular variates in a family of observed data of small size  $N$  are two or more in number in one tail, it is not desirable to treat them simply in the usual sense of statistics, because a certain physical reason may exist in this event. Therefore, the singular variate to be made the object of rejection test should be either a set of two extreme data in both tails or an extreme datum in one tail. This concept of rejection test seems to be inevitable from the viewpoint that the treatment or the evaluation of singular value itself is to be of importance in the hydrologic statistics.

Thereupon, putting  $r=1$ , Eq. (2.11) becomes,

$$\beta = 1 - (1 - \varepsilon)^N \quad (2.12)$$

Table 2.4 Singular levels  $\varepsilon_0$  for given levels of significance,  $\beta_0$ .

$N \backslash \beta_0$	10%	5%	1%
18	0.584%	0.285%	0.056%
20	525	256	049
22	478	233	046
24	438	214	042
26	404	197	039
28	376	183	036
30	351	171	034
32	0.329	0.160	0.032
34	309	151	030
36	292	142	028
38	276	135	027
40	263	128	025
42	0.251	0.122	0.024
44	239	117	023
46	229	111	022
48	219	107	021
50	210	103	020
55	0.191	0.093	0.018
60	175	085	017
65	162	079	016
70	150	073	014
75	140	068	013
80	132	064	013



Table 2.5 Examples of rejection test.

Station	Period of Observation	Sample Size $N$	Observed Value in Question $x_e'$ (mm./day)	Equation for Rejection Test	Rejection Test			Equation for Estimates
					$\varepsilon$ (%)	$\beta$ (%)	Decision	
Wakayama	1911—1955	45	394	$x_e = 96 + 29.7y_e$	$< 0.05$	$\leq 5$	Rej.	$x_e = 96 + 29.7y_e$
Ueno	1901—1953	53	287	$\log(x_e + 513) = \log 590 + 0.020y_e$	0.27	$> 10$	Adop.	$\log(x_e + 67) = \log 144 + 0.0783y_e$
Kyoto	1911—1950	40	282	$x_e = 85 + 21.4y_e$	0.055	$\leq 5$	Rej.	$x_e = 85 + 21.4y_e$
Mineyama	1911—1955	44	308	$\log(x_e + 635) = \log 704 + 0.0162y_e$	0.13	$> 5$	Adop.	$\log(x_e - 5) = \log 660 + 0.1303y_e$
Maizuru	1911—1955	42	324	$\log x = \log 81 + 0.0964y_e$	0.35	$> 10$	Adop.	$\log(x_e - 29) = \log 54 + 0.1433y_e$

If  $\beta$  is smaller than a given significant level  $\beta_0$ , the variate for the singular level  $x_\varepsilon$  may be rejected. This is represented by,

$$\varepsilon < \varepsilon_0 = 1 - (1 - \beta_0)^{1/N} \quad (2.31)$$

That is, Eq. (2.13) indicates that when the singular level  $\varepsilon$  of variate is smaller than  $\varepsilon_0$ , such a variate is rejected at the  $\beta_0$  level of significance. The relations between  $\varepsilon_0$  and  $N$  for several values of  $\beta_0$  are shown in Table 2.4 for practical facilities.

In practice, the singular level of variate in question may be evaluated from all of the other data in which such a variate is not included, and as the value of the level of significance  $\beta_0$ , 5% may be usually adopted.

Several examples of application of those proposed approaches in hydrologic frequency analysis are shown in Fig. 2.1, basing on the data in Table 2.5.

## 5. Conclusion

In this part, several discussions on the singular data included in a actual sample are made from a stochastic viewpoint. After definition of the singular value, a practical method of evaluation of such a value was successfully developed by help of the theory of stochastic limit on normals, and also a course of the rejection test for such a data was defined practically by the use of the binomial distribution.

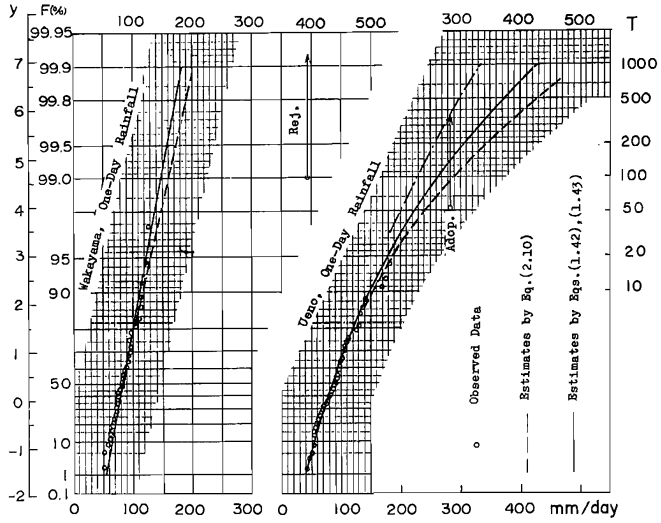


Fig. 2.1-(1) Examples of rejection test and of evaluation of singular value, (1).

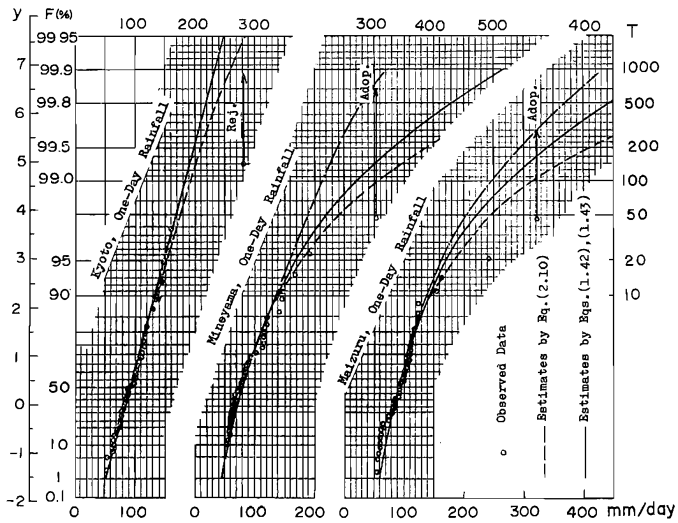


Fig. 2.2-(2) Examples of rejection test and of evaluation of singular value, (2).

The author believes that these treatments of the singular data are strongly desirable, and that the results obtained in this part must provide a useful tool in the field of hydrologic frequency analysis, although these studies were performed in 1959.

## Acknowledgment

The author is greatly indebted to Professors Tojiro Ishihara, Shigehisa Iwai and Toshio Sawada, Kyoto University, for their kind guidance throughout the course of this study, and to Professors Yasuo Ishihara and Hiroshi Nishihara, Kyoto University, for their kind suggestions and useful discussions for this study, and also to Professor Masami Ogawara, Tokyo Women's University, for his useful suggestion for the study in Part II. Moreover, the author wishes to thank Mr. Y. Sumida for his assistance in the calculation.

## References

- 1) Gumbel, E. J. The Return Period of Flood Flows. *Ann. Math. Stat.*, 12-2, 1941, pp. 163-190.
- 2) Gumbel, E. J. Statistical Theory of Droughts. *Proc. Am. Soc. Civ. Eng.* 80-439, 1954.
- 3) Gumbel, E. J. The Statistical Forecast of Floods. Ohio Water Resources Board, Bull. 15, 1949.
- 4) Thom, H. C. S. : Frequency of Maximum Wind Speeds. *Proc. Am. Soc. Civ. Eng.* 80-539, 1954.
- 5) Lieblein, J. : A New Method of Analyzing Extreme-value Data. *Nat. Adv. Comm. Aero. Tech. Note* 3053, 1954.
- 6) Jenkinson, A. F. : The Frequency Distribution of the Annual Maximum (or Minimum) values of Meteorological Elements. *Quart. Jour. Roy. Meteor. Soc.*, 81, 1955, pp. 158-171.
- 7) Kadoya, M. The Extreme-value Law and Some of its Solutions for Engineering Problems. *Jour. Agr. Eng. Soc.* 23-6, 1956, pp. 350-357 (in Japanese), and its brief report, *Proc. 6-th Nat. Cong. Appl. Mech.*, 1956, pp. 507-509.
- 8) Kadoya, M. Order Statistical Consideration on Probable Floods. *Jour. Agr. Eng. Soc.* 21-3, 1953, pp. 188-195. (in Japanese)
- 9) Kadoya, M. : Order Statistical Consideration on Probable Rainfall. *Jour. Agr. Eng. Soc.* 22-6, 1955, pp. 483-492. (in Japanese)
- 10) Kadoya, M. Statistical Method on Hydrologic Amount for Irrigation and Drainage Planning. Doctor Thesis, Kyoto Univ., 1960. (in Japanese)
- 11) Kadoya, M. Practical Method of Extreme Value Distribution for Engineering Application. *Trans. Agr. Eng. Soc. Jap.* 3, 1962, pp. 17-22. (in Japanese)
- 12) Kadoya, M. On the Plotting Positions of Hydrologic Data. *Disa. Prev. Res. Inst. Kyoto Univ. Ann.* 3, 1959, pp. 74-86. (in Japanese)
- 13) Gumbel, E. J. *Statistics of Extremes*, Columbia Univ. Press, 1958.
- 14) Fisher, R. A. and Tippett, L. H. C. Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Proc. Camb. Phil. Soc.* 24, 1928, pp. 180-190.
- 15) Gumbel, E. J. : Ref. 2) and Ref. 13), pp. 282-284.
- 16) For Examples Ref. 2), 3), 8), 9), 13) and 17) and the others.

- 17) Chow, Ven Te : The Log-probability Law and its Engineering Applications. Proc. Am. Soc. Civ. Eng. 80-536, 1954.
- 18) Thomas, H. A. · Frequency of Minor Floods. Jour. Boston Soc. Civ. Eng., 25, 1948, pp. 425-442.
- 19) Gumbel, E. J. · Simplified Plotting of Statistical Observations. Trans. Am. Geoph. Union, 26-1, 1945, pp. 69-
- 20) Kadoya, M. · Probable Singular Hydrologic Amount and its Rejection Test. Trans. Agr. Eng. Soc. Jap., 3, 1962, pp. 23-27. (in Japanese)
- 21) Ogawara, M. and Collaborators : Stochastic Limit for Maximum Possible Amount of Precipitation. Paper in Meteo. and Geoph. 5, 1955, pp. 8-20.