

Finding Climate Characteristics Associated with Primary Modes of Global Climate Variability

Hirotaka SATO

Tokyo Climate Center, Japan Meteorological Agency (TCC/JMA)

1. How to determine the thresholds for dividing data into three classes

As the first step of this exercise, you need to divide observation data into three classes (Dry/Normal/Wet for precipitation and Cold/Normal/Warm for temperature) so carefully that the occurrence probability of each class is equal. This idea is simple and easy to understand but the actual process for optimal classification is not as simple as it looks.

Fig. 1 is the example of tables relating to the classification. When you make occurrence probability figures, you might only check the “Tertile thresholds” table colored in yellow, but the detail of defining the thresholds is elaborated here.

J	K	L	M	N	O	P	Q
c_k	d_k		$n_1(k)$	$n_2(k)$	$n_3(k)$	$S_1(k)$	$S_2(k)$
Options of tertile thresholds			Cold	Normal	Warm	S1-Value	S2-Value
(1)	26.8	27.566667	21	17	19	0.0024623	0
(2)	26.7	27.566667	18	20	19	0.0006156	0.1
(3)	26.8	27.466667	21	14	22	0.0116959	0.1
(4)	26.7	27.466667	18	17	22	0.004309	0.2
			$N = n_1(k) + n_2(k) + n_3(k) = 57$				
Tertile thresholds		Probability	Cold	Normal	Warm		
	26.7	27.566667	0.3684211	0.2982456	0.3333333		

Fig. 1. An example of tables relating to determining the classification thresholds. n_1 , n_2 and n_3 are the number of years in “Cold”, “Normal” and “Warm” class, respectively. The definitions of S_1 and S_2 are given in the text.

You can find four options for tertile thresholds as indicated by (1) – (4) in Fig. 1. Hereinafter, (c_k, d_k) is the smaller and larger tertile thresholds for the option (k) ($k = 1, 2, 3, 4$), respectively. The (1)’s thresholds (c_1, d_1) are the most straightforward ones, where c_1 and d_1 are the values corresponding to 1/3 and 2/3 of whole data from the lowest, respectively (see also “Notes” at the end of this section). But in this example, if we select this option (1) as the classification thresholds, the numbers of years in each class are not equal as $(n_1(1), n_2(1), n_3(1))$ is $(21, 17, 19)$ among the total $N = 57$ years, where $(n_1(k), n_2(k), n_3(k))$ is the number of years that the observation data meet “ $\leq c_k$ ”, “ $c_k <$ and $\leq d_k$ ” and “ $d_k <$ ”, respectively, and $N = n_1(k) + n_2(k) + n_3(k)$.

Sometimes the simplest option (1) is not optimal like this example when there are several years having same observation values around a threshold. Then we consider other options (2) – (4). These thresholds are derived from (1) by subtracting Δ from (c_1, d_1) (Tab. 1), where Δ is 1 [mm] for precipitation and 0.1 [°C] for temperature.

Tab. 1. Four options for classification thresholds

Option (k)	c_k	d_k
(1)	c_1	d_1
(2)	$c_1 - \Delta$	d_1
(3)	c_1	$d_1 - \Delta$
(4)	$c_1 - \Delta$	$d_1 - \Delta$

As a metrics of the equality of classification, S_1 is calculated for these 4 options, where

$$S_1(k) = \left(\frac{1}{3} - \frac{n_1(k)}{N}\right)^2 + \left(\frac{1}{3} - \frac{n_2(k)}{N}\right)^2 + \left(\frac{1}{3} - \frac{n_3(k)}{N}\right)^2.$$

Then we select the one whose S_1 value is the least among four. This metrics S_1 is based on the idea of the least square methods. You can find out soon that S_1 has its minimum ($S_1 = 0$) if $n_1 = n_2 = n_3 = 3/N$ is possible. In this example, $S_1(2)$ is the least among the four and the option (2) is eventually selected as the classification thresholds.

Nevertheless, if the S_1 values of some options are equal, we introduce another metrics S_2 and chose the one whose S_2 value is the least among them.

$$S_2(l) = |c_1 - c_l| + |d_1 - d_l| \quad (l = 2, 3, 4).$$

Notes 1:

This MS-Excel tool calculates (c_1, d_1) with PERCENTILE function of MS Excel software. This function interpolates to determine (c_1, d_1) as appropriate (For example, if $N \neq 3M + 1$ ($M \in \mathbb{N}$), we cannot find the exact value c_1 in the data itself). That is why sometimes thresholds have ridiculously many decimal places even if you input data in 1/10 °C. See also the technical support information provided by Microsoft.

<https://support.office.com/en-US/article/PERCENTILE-function-91B43A53-543C-4708-93DE-D626DEBDDCA>

Notes 2:

It should be noted that climatological probabilities can far exceeds 33% in some cases. For example, if there are many years having 0[mm] precipitation values, the climatological probability of "Dry" is far above 33%. This is common in desert areas where it hardly rains, but you might find the situation like this for some stations in dry season. In such a case (namely, $c_1=0$ or $d_1=0$), we let $\Delta=0$ when calculating corresponding (c_l, d_l) , where $l = 2, 3, 4$.

2. Removing a long-term linear trend from temperature data

Anomalies you get from data can be interpreted as combination of various fluctuation components containing different time scales. Possibly there might be impacts from ENSO (El Niño-Southern Oscillation) and teleconnection patterns, both of which have monthly to seasonal time scale, and of course we cannot overlook an influence from global warming that cause a long-term change.

In general there are long-term trends in temperature records. If we let a temperature time series $T(t)$ (t is year) as its linear trend $(a \times t + b)$ plus the residuals $\epsilon(t)$ like

$$T(t) = (a \times t + b) + \epsilon(t),$$

for example, $(a \times t + b)$ and $\epsilon(t)$ are indicated by the blue line in Fig. 2 (a) and red dots in Fig. 2 (b), respectively. Since the aim of this exercise is to find climate characteristics associated with climate variability modes, **we should eliminate long-term linear trends from your data to minimize impacts from global warming.**

This MS-Excel tool has two sheets for temperature analysis, “Temp. (raw)” and “Temp. (detrend).” On the “Temp. (raw)” sheet, temperature data on the “climat” sheet is copied without any change. By contrast, as could be seen in Fig. 3, in the column C on the “Temp. (detrend)” sheet, you can obtain residuals from the long-term linear fit to the raw temperature data. You will also find “Liner Regression” table around the bottom-right corner in Fig. 3. This table gives you the slope a , the intercept b and the coefficient of determination, rsq , which is the square of the correlation coefficients between the detrended temperature data (column C) and year data itself (column A).

Generally, **it is recommended that you should use the detrended temperature data on the “Temp. (detrend)” sheet** even though statistical testing of the slope a and the intercept b is left unmentioned (it is beyond this exercise). The “Temp. (raw)” sheet is just for your information.

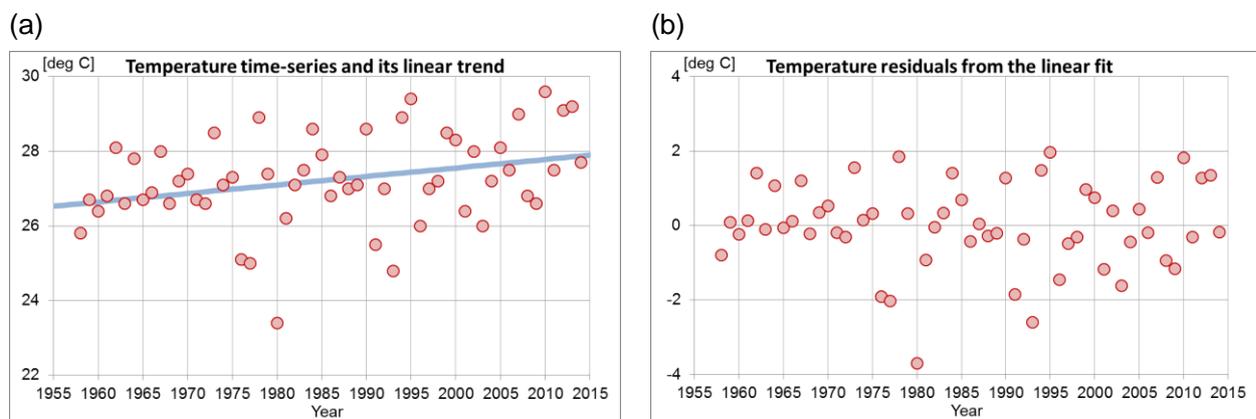


Fig.2. (a) An example of the time series (red dots) of a raw temperature data and its linear trend (blue line). (b) The residuals from the linear trend indicated in (a).

	A	B	C	D	E	F	G	H
1	Year	Month	Mean Temp. (detre	Index	Event			
2	1958	8	-0.797519661	0.2	0		Start Year	1958
3	1959	8	0.079710051	-0.1	0		End Year	2014
4	1960	8	-0.243060237	0	0		Calendar Month	8
5	1961	8	0.134169475	-0.4	0			
6	1962	8	1.411399188	-0.2	0			
7	1963	8	-0.1113711	0.9	1		Calculate and Draw !!	
8	1964	8	1.065858612	-0.8	-1		Total Year	57
9	1965	8	-0.056911676	1.3	1		Effective Year	57
10	1966	8	0.120318036	-0.1	0			
11	1967	8	1.197547749	-0.4	0		Sampling bias rate(%)	0
12	1968	8	-0.225222539	0.4	0			
13	1969	8	0.352007173	0.7	1			
14	1970	8	0.529236885	-1	-1			
15	1971	8	-0.193533402	-0.6	-1			
16	1972	8	-0.31630369	1.8	1			
17	1973	8	1.560926022	-0.9	-1		Linear Regression	
18	1974	8	0.138155734	-0.1	0		a	0.022770288
19	1975	8	0.315385446	-0.8	-1		b	-17.98670383
20	1976	8	-1.907384841	1.1	1		rsq	0.097066122

Fig. 3. An example of “Temp. (detrnd)” sheet.

3. Binomial Test (complementary explanation)

In binomial testing you assess whether an occurrence probability of an individual class is higher than the climatological probability. But you can also consider a probability that a class does **NOT APPEAR**.

An example of a table for binomial testing is given in Fig. 4 and there you can find “Not Cold” and “Not Warm” columns other than “Cold”, “Normal” and “Warm” columns. In this example, first, you will find the p-value of “Warm” class under La Niña condition is less than 0.01 and so you can say that La Niña affects so significantly that the occurrence probability of “Warm” class is higher than the climatological one (namely about 33.33...%).

At the same time, you will notice that the p-value of “Not Cold” associated with La Niña is also significant with the cell colored in yellow. This can be interpreted as follows; there are many “Not Cold” years associated with La Niña condition and this relation is also statistically significant. In other words, the occurrence probability of “Not Cold” year is significantly higher than that of climatological probability, which is the sum of “Normal” and “Warm” classes’ climatological probability (about 67% = 33.33... % + 33.33...%).

So how about the years when El Niño condition persisted? The possible answer is; the occurrence probability of “Cold” class is 38%, actually higher than the climatological probability of the class, but this is not statistically significant. However, it can be said that there are significantly less “Warm” years associated with El Niño conditions.

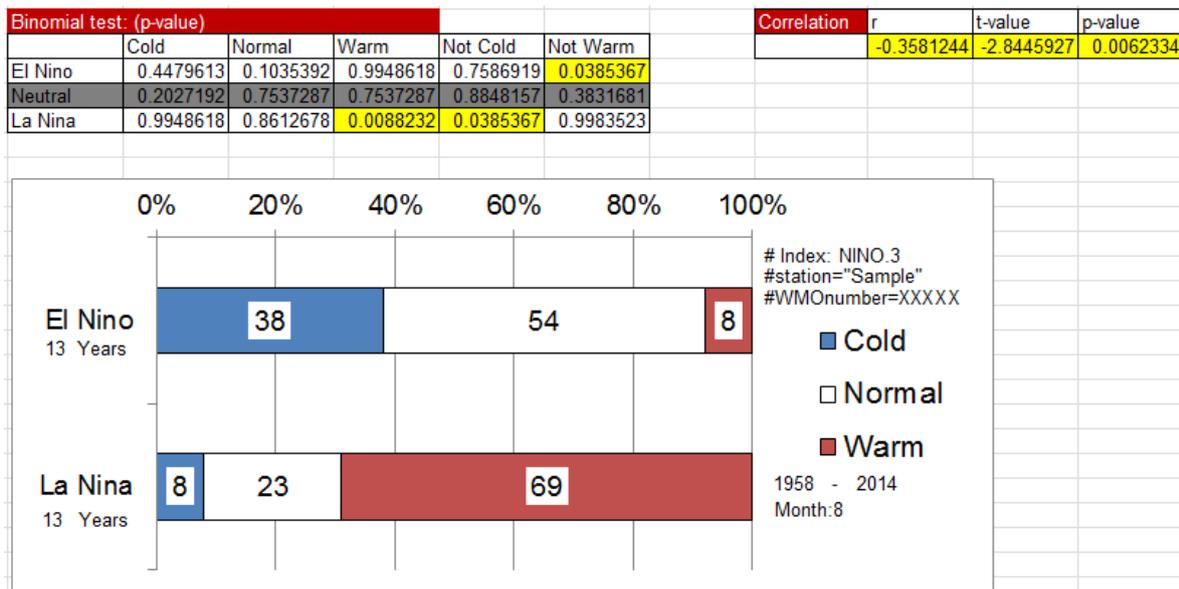


Fig. 4. An example of tables about binomial test and correlation statistics and occurrence probability figure.

4. Correlation

You can also get correlation coefficients between climate variability indices and temperature/precipitation data you input. Even if you cannot get significant characteristics through binomial testing, correlation statistics might give you interesting and helpful results.

The example of correlation statistics is also given in Fig. 4. You can find the “Correlation” table around the top-right corner of the figure. **This table gives the correlation coefficients (r), t-value and p-value. If the correlation is statistically significant with 90% significance level, these cells are colored in yellow.** The detail of t-value is not given in the text, but you can interpret it as the one of the measures that is often used for assessing statistical significance of correlation coefficients. Considering the sample size of the data you use in this exercise, roughly speaking, it would be safe to say that the correlation is significant (with 90% level) if the absolute value of t-value exceeds 1.7.

5. How to apply other climate variability indices

You can apply various climate variability indices other than NINO.3 SST index, which is for monitoring ENSO (El Niño-Southern Oscillation). On inputting indices data, we assume two types of index, “**ENSO-Related**” type and “**Teleconnection**” type. The former is ENSO-related indices, specifically, NINO.3, NINO.WEST and IOBW, and the latter is teleconnection indices (AO, NAO, PNA and others). These two types differ in the way each event is identified.

(a) ENSO-related type indices (NINO.3, NINO.WEST and IOBW)

Data Source

You can find ENSO-related indices in the directory “index” (each file name is

5rmean_XXXX.txt, where XXXX is the name of each index). These indices are five-month running mean SST deviation in each area. These SST indices and the details are also available at TCC's El Niño Monitoring webpage (See the link below).

<http://ds.data.jma.go.jp/gmd/tcc/tcc/products/elnino/index/>

*Please follow SST (sliding 30-year base period) > Nino 3/ Nino West/ IOBW > 5-month running mean SST deviation.

Input data into the MS-Excel tool

To make figures with this tool, please copy the whole data in 5rmean_XXXX.txt and paste it to the cell "A2" on "Index" sheet (Fig. 5(a)). Here you need set some parameters about the index, "Index Type", "Index Name" and "Threshold" (Tab. 1). "Index Type" and "Index Name" should be self-explanatory. **"Threshold" is needed to identify each event**, but it is fixed for this "ENSO-Related" type index. For example, **if the index value exceeds the threshold, we regard that positive SST anomaly event persisted in that calendar month. In other word, +1 is set as the event value.**

Then please click "Set" button. Index values on "Index" sheet will be arranged (Fig. 5(b)) and event values will be set on "Event index" sheet (Fig. 5(c)). As mentioned in the presentation materials, event values +1/0/-1 are correspond to warmer/neutral/cooler SST events, respectively. El Niño (La Niña) events, in which the SST in NINO.3 is above (below) normal, can be identified when the five-month running mean value of the NINO.3 SST index (5rmean_NINO3.txt) was 0.5°C or above (-0.5°C or below) for at least six consecutive months. With NINO.WEST (5rmean_NINOWEST.txt) or IOBW SST index (5rmean_IOBW.txt), warmer (cooler) SST events can be identified in the same way, but the thresholds were 0.15°C for warmer events and -0.15°C for cooler events.

(b) Teleconnection type indices (AO, EU, NAO, PNA, WP and others)

Data Source

You can find teleconnection pattern indices also in the directory "index" (each file name is YYY.txt, where YYY is the name of each teleconnection pattern).

Input data into the MS-Excel tool

You can input teleconnection indices in the similar way to ENSO-related type indices, but you have to set also "Threshold" parameter (Tab. 1). You can select the values in the pull-down box or set arbitrary values by inputting it into the box. But, **at first, it would be better to set "1" for trial and modify it as necessary** (for example, in the case of the sampling bias rate being not preferable).

If you want clear the index data, please click "Reset Index Data" button on "Index" sheet.

(a) "Index" sheet before clicking "Set" button on the same sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	V --- Copy and paste an index file's content here (A2!)							Index Type		Index Name		Threshold	
	Set		Reset Index Data				<input checked="" type="radio"/> ENSO-Related <input type="radio"/> Teleconnection		NINO.3		0.5		
2	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	
3	1949	-0.2	-0.1	0.1	0.0	-0.2	-0.3	-0.5	-0.8	-0.9	-1.0	-1.2	-1.2
4	1950	-1.2	-1.0	-0.9	-0.8	-0.6	-0.6	-0.5	-0.4	-0.4	-0.4	-0.4	-0.3
5	1951	-0.2	-0.1	0.1	0.2	0.5	0.7	0.8	1.0	1.1	0.9	0.9	0.8
6	1952	0.6	0.5	0.4	0.1	-0.1	-0.2	-0.2	-0.2	-0.2	0.0	0.1	
7	1953	0.1	0.3	0.4	0.5	0.6	0.6	0.6	0.6	0.5	0.5	0.4	
8	1954	0.3	-0.1	-0.3	-0.6	-0.7	-0.9	-0.8	-0.8	-0.7	-0.8	-0.7	-0.6
9	1955	-0.6	-0.6	-0.6	-0.7	-0.8	-0.8	-0.9	-0.9	-1.1	-1.2	-1.2	-1.0
10	1956	-0.7	-0.5	-0.3	-0.2	-0.2	-0.2	-0.3	-0.3	-0.4	-0.5	-0.5	-0.5
11	1957	-0.3	0.0	0.3	0.6	0.9	1.1	1.2	1.2	1.3	1.3	1.3	1.4
12	1958	1.3	1.1	0.8	0.6	0.4	0.3	0.2	0.2	0.2	0.2	0.1	0.1
13	1959	0.1	0.1	0.1	0.0	-0.1	-0.2	-0.1	-0.1	0.0	0.1	0.1	
14	1960	0.1	0.0	0.0	0.0	0.0	0.1	0.0	-0.1	-0.1	-0.2	-0.2	

(b) "Index" sheet after clicking "Set" button on the same sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	V --- Copy and paste an index file's content here (A2!)							Index Type		Index Name		Threshold	
	Set		Reset Index Data				<input checked="" type="radio"/> ENSO-Related <input type="radio"/> Teleconnection		NINO.3		0.5		
2	Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	APR	SEP	OCT	NOV	DEC
3	1949	-0.2	-0.1	0.1	0	-0.2	-0.3	-0.5	-0.8	-0.9	-1	-1.2	-1.2
4	1950	-1.2	-1	-0.9	-0.8	-0.6	-0.6	-0.5	-0.4	-0.4	-0.4	-0.4	-0.3
5	1951	-0.2	-0.1	0.1	0.2	0.5	0.7	0.8	1	1.1	0.9	0.9	0.8
6	1952	0.6	0.5	0.4	0.1	-0.1	-0.2	-0.2	-0.2	-0.2	-0.1	0	0.1
7	1953	0.1	0.3	0.4	0.5	0.6	0.6	0.6	0.6	0.5	0.5	0.5	0.4
8	1954	0.3	-0.1	-0.3	-0.6	-0.7	-0.9	-0.8	-0.8	-0.7	-0.8	-0.7	-0.6
9	1955	-0.6	-0.6	-0.6	-0.7	-0.8	-0.8	-0.9	-0.9	-1.1	-1.2	-1.2	-1
10	1956	-0.7	-0.5	-0.3	-0.2	-0.2	-0.2	-0.3	-0.3	-0.4	-0.5	-0.5	-0.5
11	1957	-0.3	0	0.3	0.6	0.9	1.1	1.2	1.2	1.3	1.3	1.3	1.4
12	1958	1.3	1.1	0.8	0.6	0.4	0.3	0.2	0.2	0.2	0.2	0.1	0.1
13	1959	0.1	0.1	0.1	0	0	-0.1	-0.2	-0.1	-0.1	0	0.1	0.1
14	1960	0.1	0	0	0	0	0	0.1	0	-0.1	-0.1	-0.2	-0.2
15	1961	-0.1	0	0.1	0.1	0	-0.1	-0.3	-0.4	-0.5	-0.4	-0.3	-0.2

(c) "Event Index" sheet after clicking "Set" button on the "Index" sheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Event values are based on the data on "Index" sheet.												
2	# Index: NINO.3												
3	# Year 1949 2015 5-months-running mean :: 6 months consecutive constraint												
4	# Index<=-1: La Nina -1<Index<1 Neutral 1<=Index: El Nino												
5	Year	JAN	FEB	MAR	APR	MAY	JUN	JUL	APR	SEP	OCT	NOV	DEC
6	1949	0	0	0	0	0	0	0	-1	-1	-1	-1	-1
7	1950	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0
8	1951	0	0	0	0	0	1	1	1	1	1	1	1
9	1952	1	1	0	0	0	0	0	0	0	0	0	0
10	1953	0	0	0	0	1	1	1	1	1	1	1	0
11	1954	0	0	0	-1	-1	-1	-1	-1	-1	-1	-1	-1
12	1955	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
13	1956	-1	-1	0	0	0	0	0	0	0	0	0	0
14	1957	0	0	0	1	1	1	1	1	1	1	1	1
15	1958	1	1	1	1	0	0	0	0	0	0	0	0
16	1959	0	0	0	0	0	0	0	0	0	0	0	0
17	1960	0	0	0	0	0	0	0	0	0	0	0	0
18	1961	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 5. An example of "Index" sheet and "Event Index" sheet.

Tab. 1. Parameters for climate variability indices

Index Type	Index Name	Threshold
ENSO-Related	NINO.3	0.5 (fixed)
	NINO.WEST	0.15 (fixed)
	IOBW	
Teleconnection	AO	Variable (But, it would be better to set “1” for trial at first.)
	EU	
	NAO	
	PNA	
	WP	
	(Other Indices)	