

1. 基礎統計量

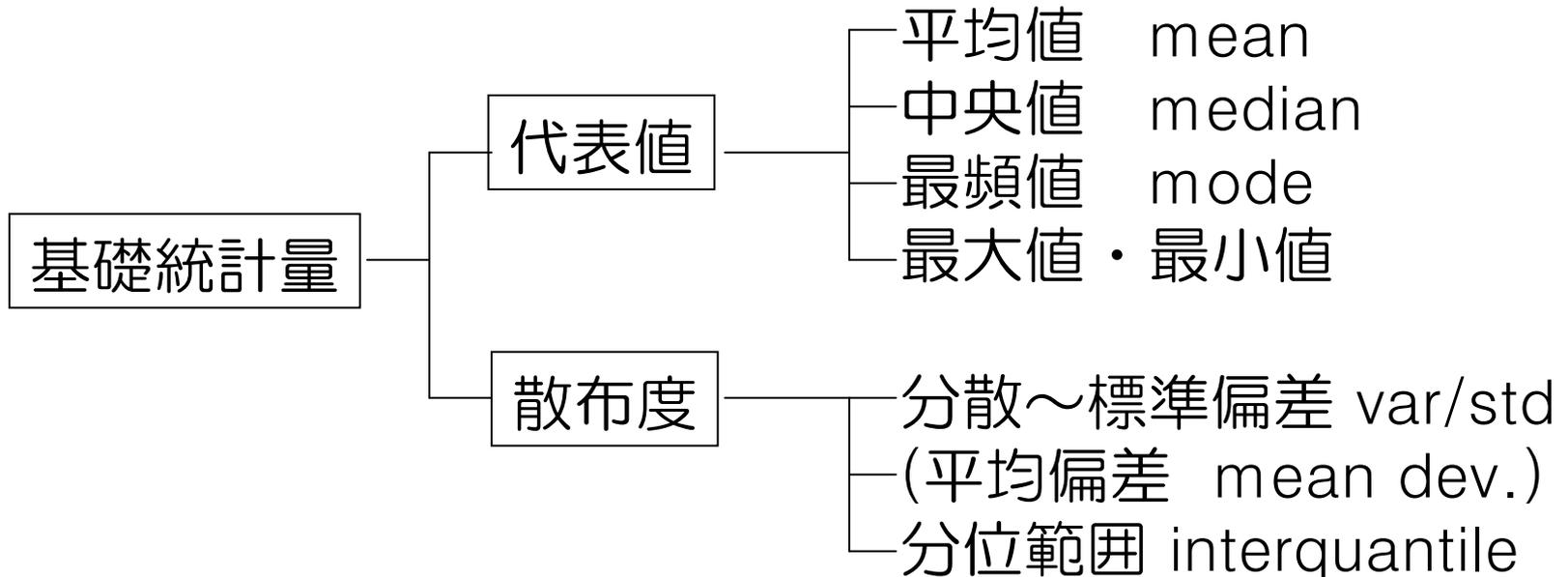
- ・ 1.0 資料の代表値と散布度
- ・ 1.1 代表値
 - 平均・最頻値・中央値
- ・ 1.2 散布度
 - 偏差
 - 分散・標準偏差
- ・ 1.3 高次モーメント
 - 歪度・尖度

キーワード
平均と偏差

1. 0 基礎統計量

- 物理的なデータを縮約するには、静的な（代表値：平均）成分と動的な（散布度：変動・偏差）成分とに分けて考察すると都合が良い。

変数分布 = 代表値 + 散布度

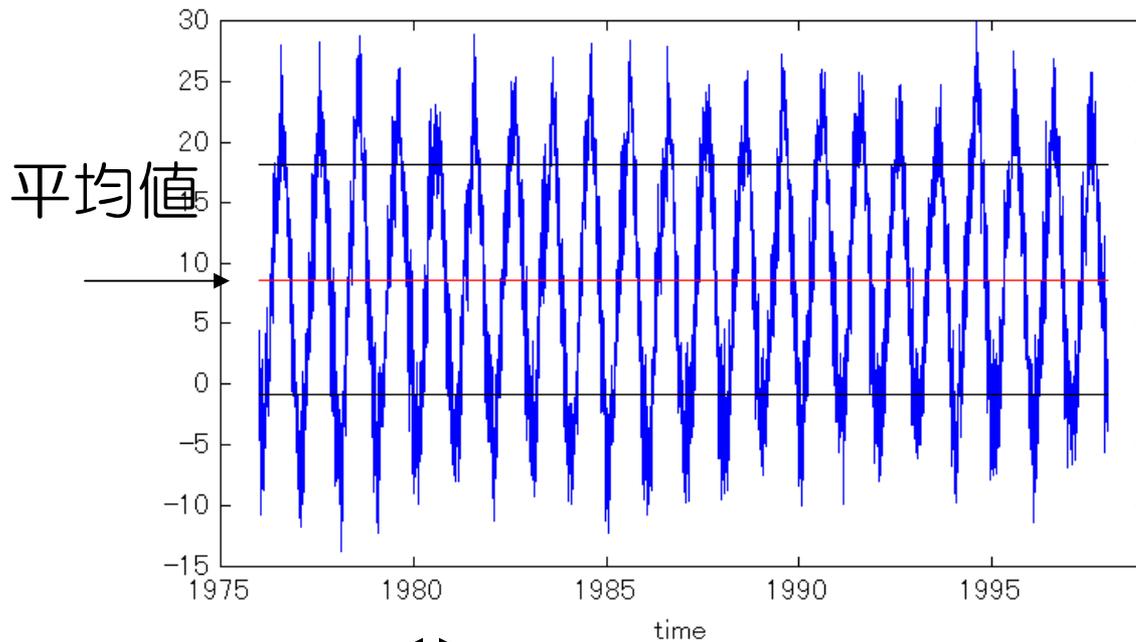
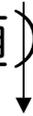


1. 1 代表値

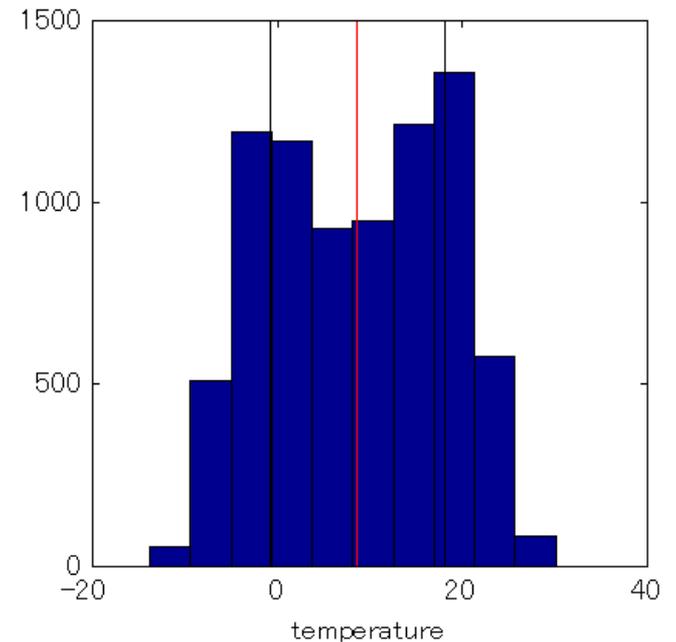
札幌の日平均気温の時系列（1976-1997）

平均

モード（最頻値）やメジアン（中心値）は？



1年

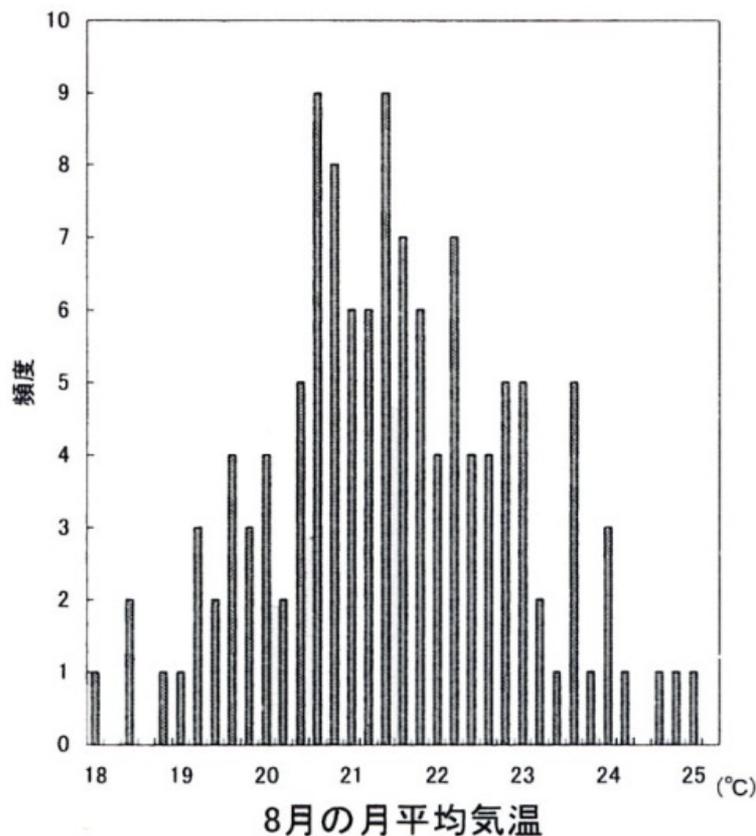


度数分布

Histogram

最頻値は、度数分布の作り方に依存

札幌における8月の月平均気温の
ヒストグラム(1877-2000年)



札幌における8月の降水量の
ヒストグラム(1877-2000年)

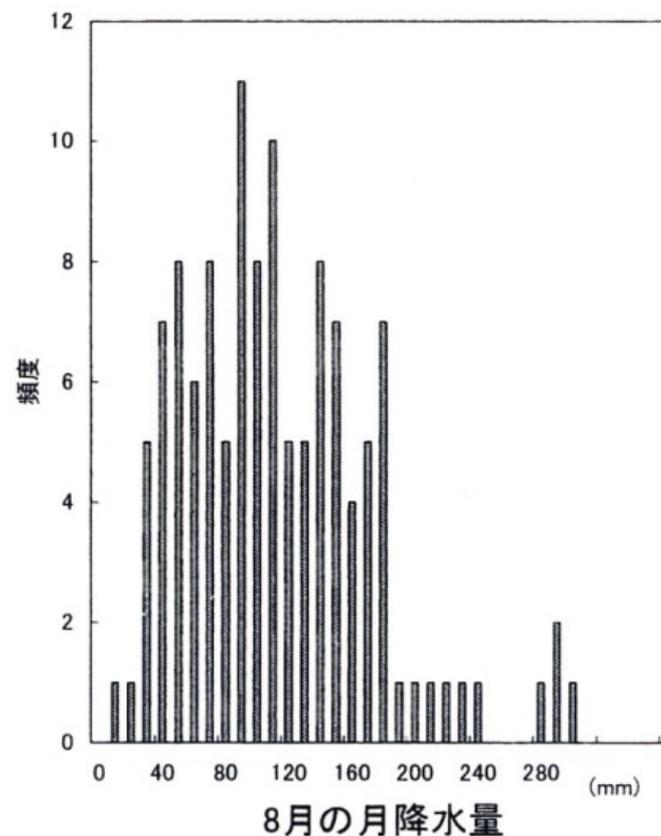


図 1: 札幌における 8 月の月平均気温 (左, 単位 °C) と月降水量 (右, 単位 mm/month) のヒストグラム.

データの期間は 1877~2000 年である. 気温の分布はほぼ左右均等に分布しているのに対し, 降水量は右の裾が長くなる分布をしている.

1. 1 代表値

百分位 percentile
四分位 quartile

中央値 (Median) 分位 (quantile)

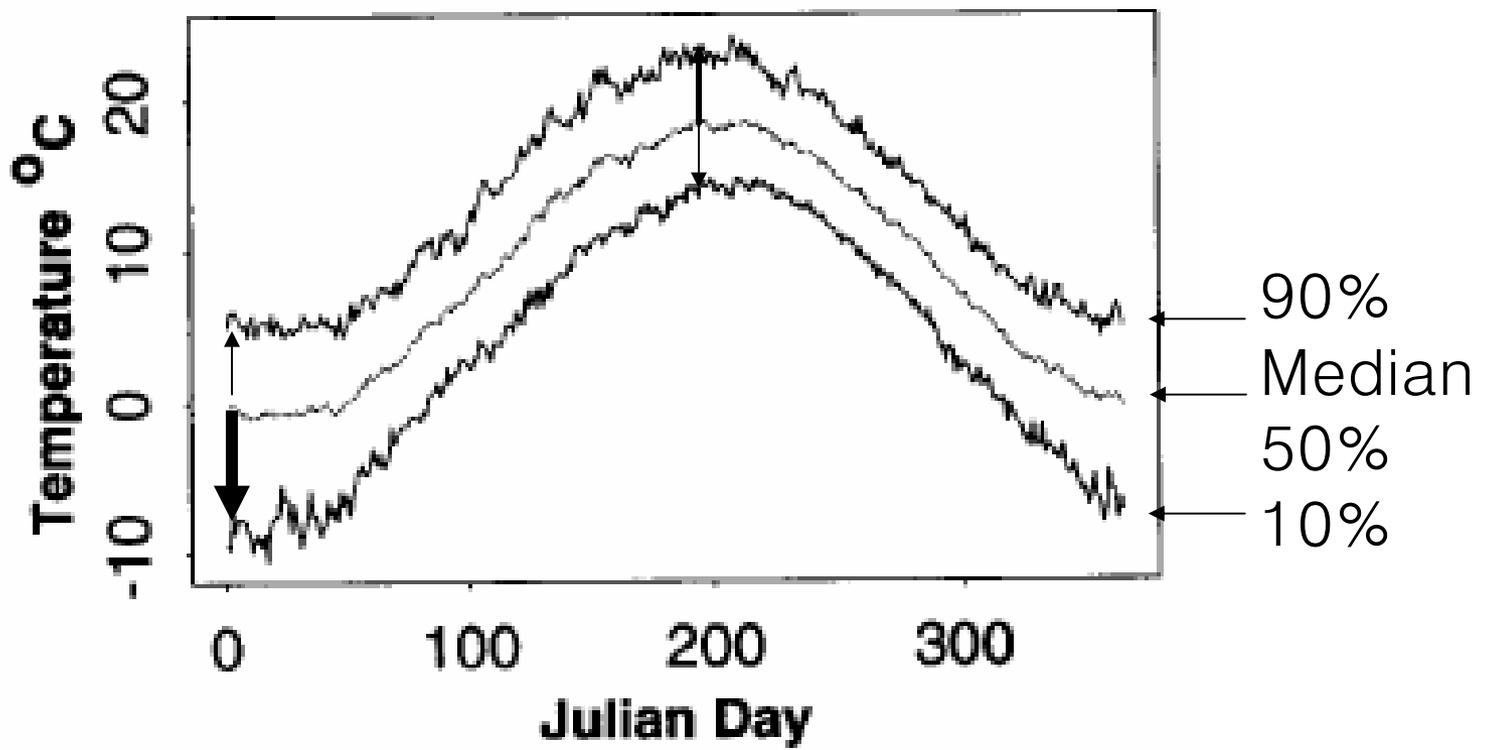


Figure 2.1: The 10th, 50th, and 90th quantiles of daily mean temperature at Potsdam, Germany (1983-94).

季節によって
分布傾向が違う
正規分布からはずれている
→外れ値(Outliers)の処理に有効

von Storch & Zwiers (1991)

メジアン・フィルター

original



added noise



average



median



3×3の領域の場合

3, 3, 3, 4, 4, 5, 5, 5, 10

5	3	4
3	10	5
3	4	5

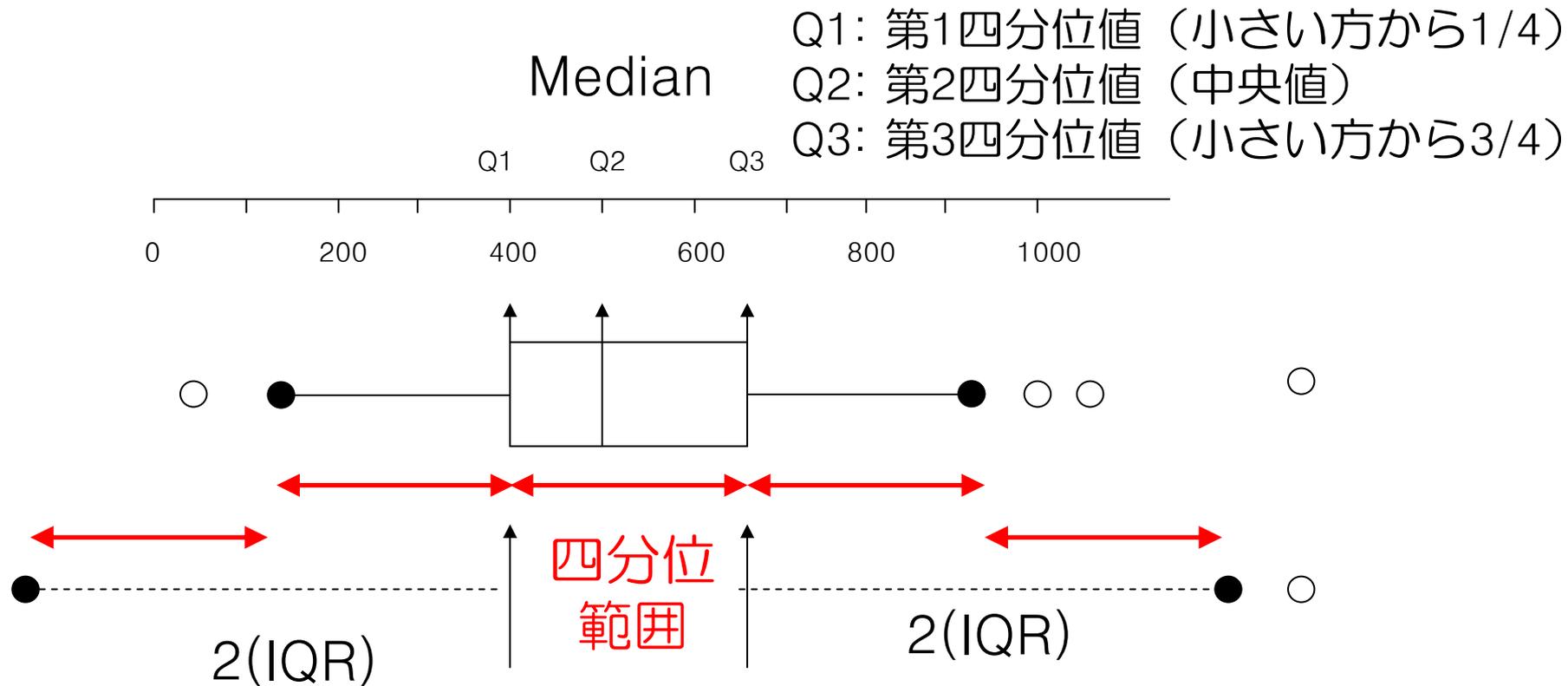


5	3	4
3	4	5
3	4	5

領域の全ての値の
平均値を中心に挿入

インパルス性の雑音を取り除く
のに有効

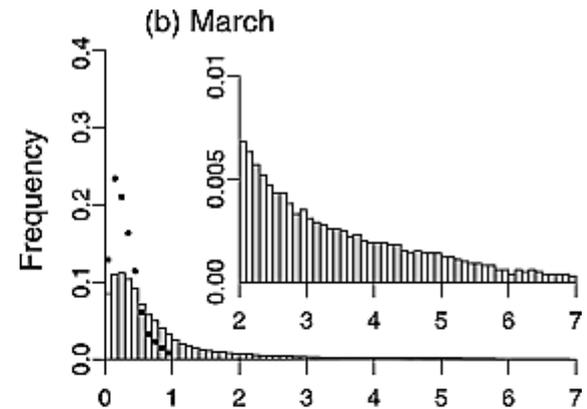
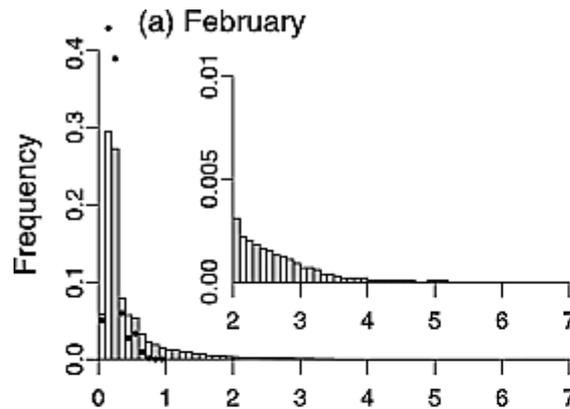
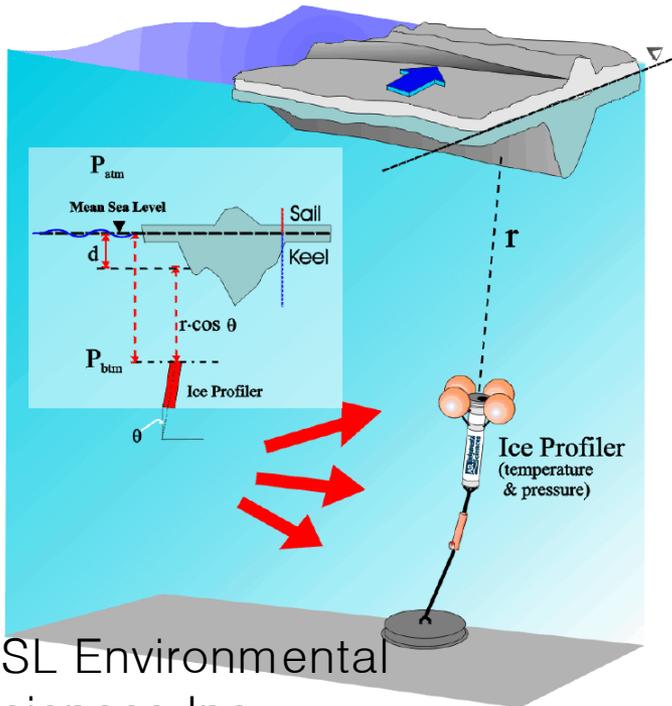
四分位範囲フィルター Interquartile range filter



外れ値 or 異常値の除去に有効

百分位 (percentile) の利用例

海水の厚さの頻度分布 (Fukamachi et al., 2003, 2006)

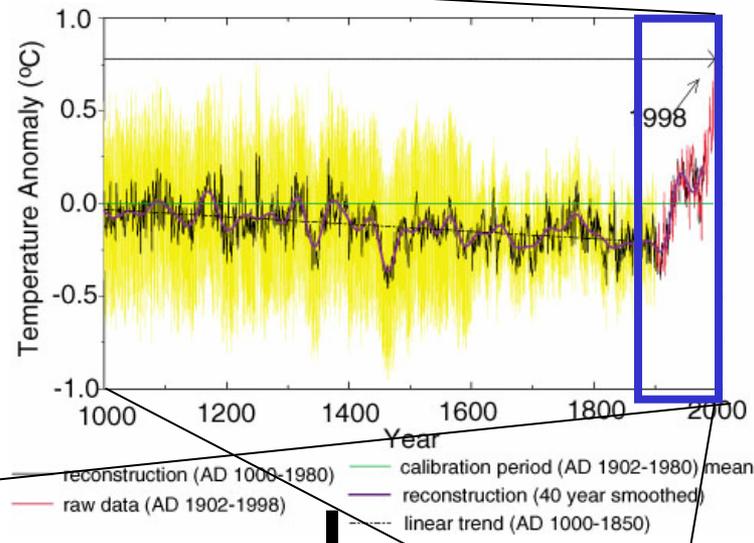
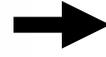
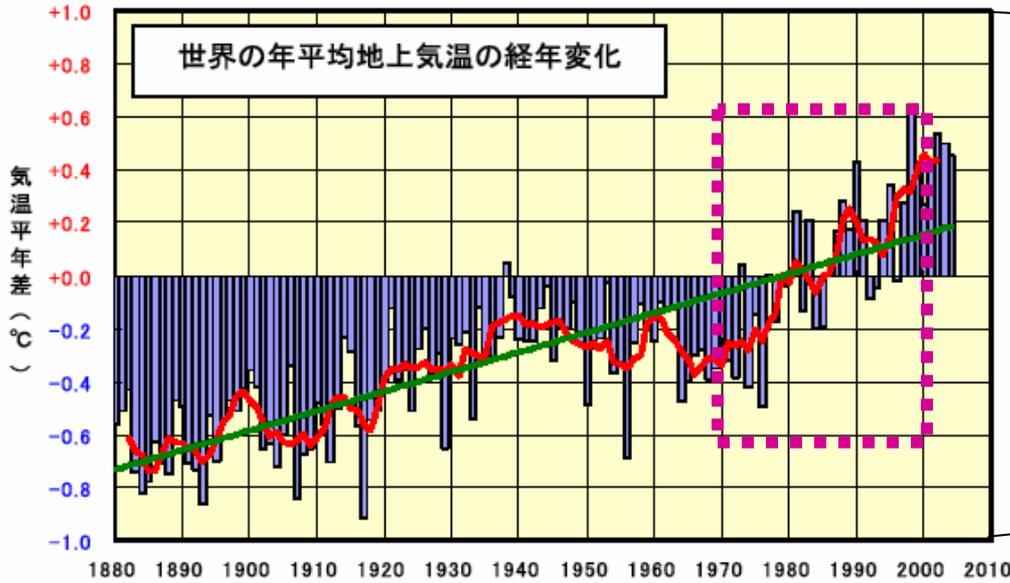


Draft, m

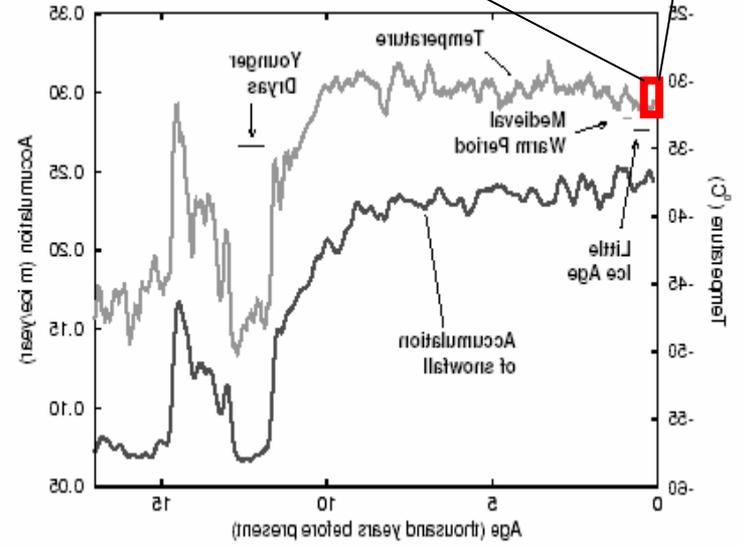
Period (Subsection)	Length, km	Mean	S. D.	90%	99%	Max.
1999						
9–28 Feb (1–7)	275	0.37 (0.20)	0.44	0.79	2.32	7.28
28 Feb–6 Mar (8)	170	— (0.00)	—	—	—	—
6–27 Mar (9–19)	442	0.77 (0.26)	1.10	1.81	5.54	17.04
9 Feb–27 Mar (All)	887	0.60 (0.24)	0.90	1.35	4.79	17.04

ASL Environmental Sciences Inc.

平均の変動



より長いスケールでは
平均も変動する

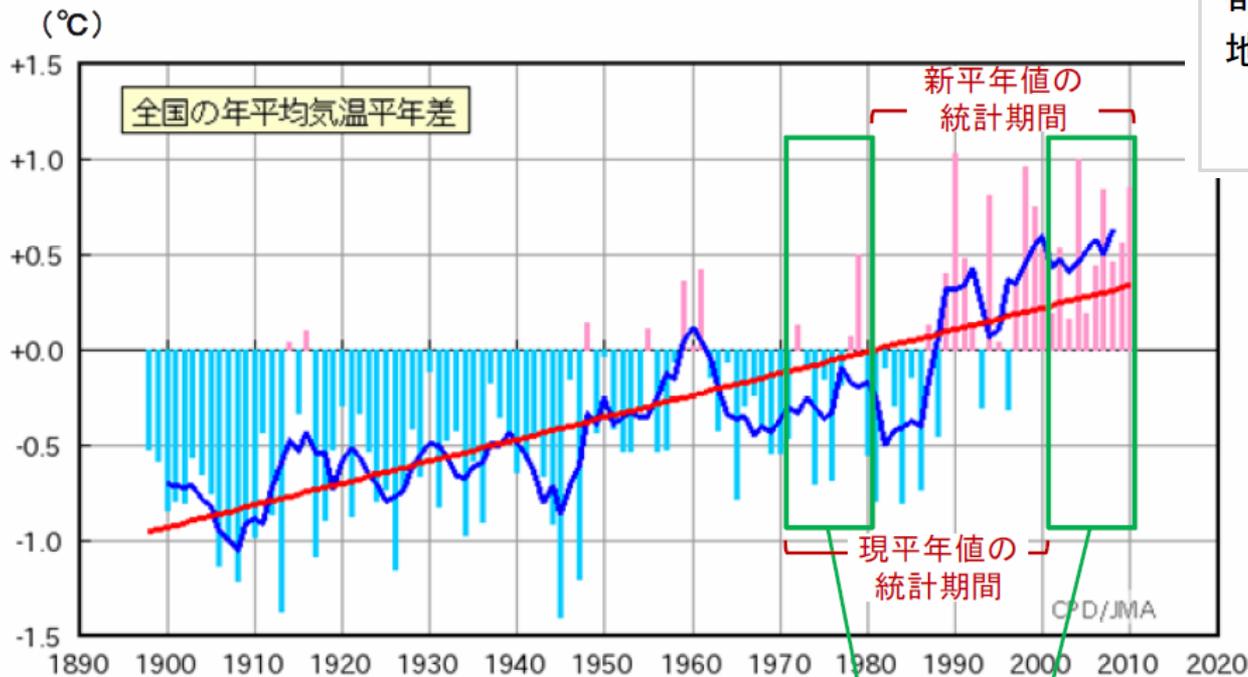


気象の平年値 (Climatological normal)

気象の平年値は30年平均 (WMOが1935年に制定)
2011年5月に、1971-2000年のものから1981-2010年のものに更新

図2 日本の年平均気温の現平年値 (1971-2000年) からの差の経年変化 (1898~2010年)
都市化の影響が比較的少ない17地点の平均気温の平年差。

年平均気温の現平年値(1971-2000年)からの差



この差が平年値の差となって現れた

1.2 散布度

なぜ必要か？

→ 平均だけではデータの特徴を捉えるのに
不十分

サンプル	1	2	3	4	5
Aグループの成績	30点	40点	50点	60点	70点
Bグループの成績	50点	50点	50点	50点	50点

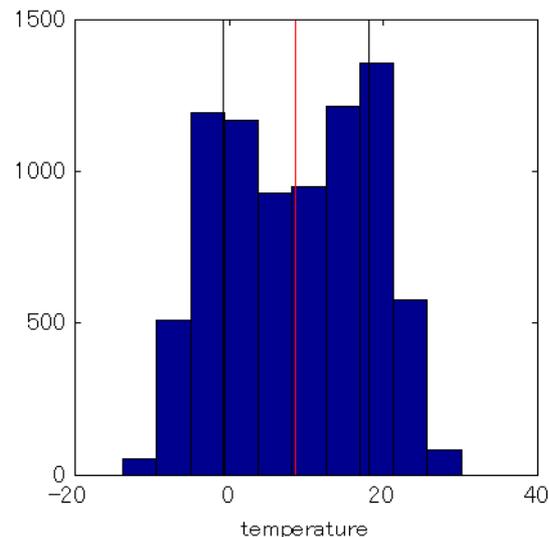
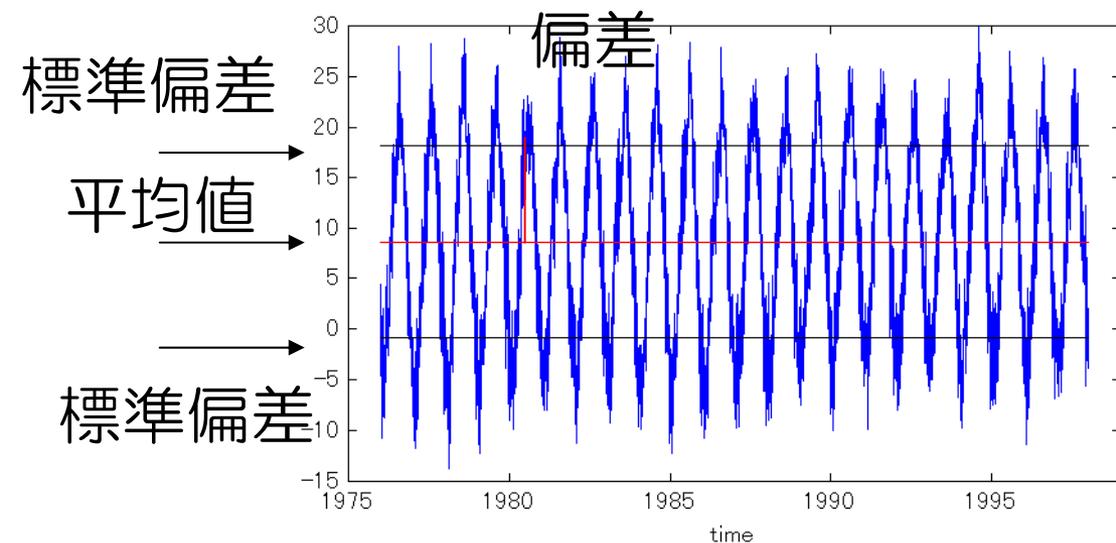
石村 (1989)

1.2 散布度

散布度 — 分散・標準偏差
variance · standard deviation

札幌の日平均気温の時系列 (1976-1997)

平均 ↓ ↓ 標準偏差



平均:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$

不偏分散 (標本分散):

$$s^2 = \frac{1}{\underline{N-1}} \sum_{i=1}^N (x_i - \bar{x})^2 = \frac{1}{\underline{N-1}} \left\{ \sum_{i=1}^N (x_i)^2 - N\bar{x}^2 \right\}$$

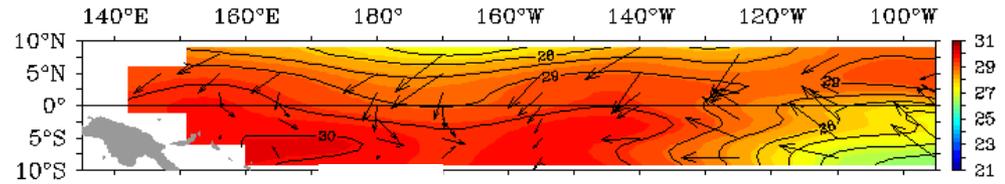
平均を計算する際には、 N 個のデータ全てが独立だが、分散を計算する際には、 $N-1$ 個のデータのみが独立。何故なら、 $x_1 - \bar{x}, \dots, x_{N-1} - \bar{x}$ までを知れば、次の $x_N - \bar{x}$ を知ることが出来るから。

1. 2 散布度

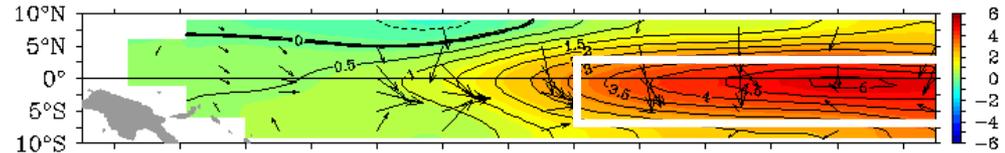
偏差

全体値 = 平均 + 偏差
 mean anomaly

TAO Monthly Mean SST (°C) and Winds (m s⁻¹)



December 1997 Total

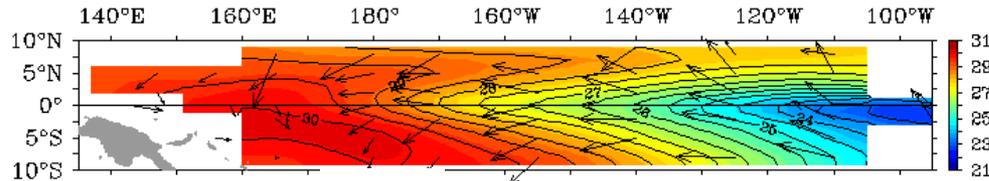


December 1997 Anomalies

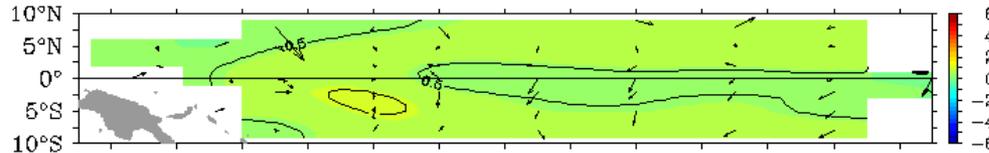
El Nino Conditions

PMEL/NOAA

TAO Monthly Mean SST (°C) and Winds (m s⁻¹)



December 1993 Total



December 1993 Anomalies

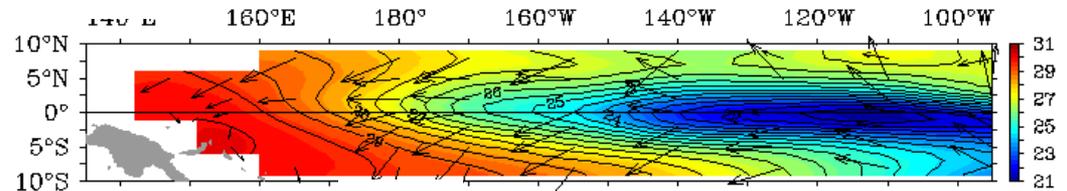
Normal Conditions

通常の状態

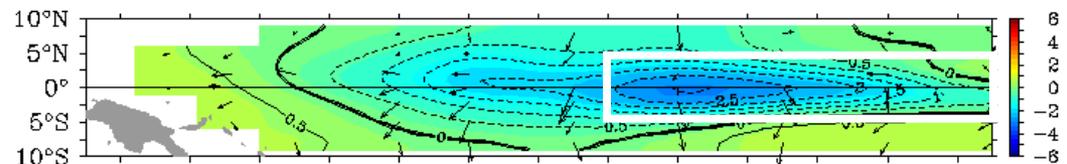
TAO Project Office/PMEL/NOAA

東風が弱い ↑ 通常より暖水
 ↓ 通常より冷水
 東風が強い

TAO Monthly Mean SST (°C) and Winds (m s⁻¹)



December 1998 Total

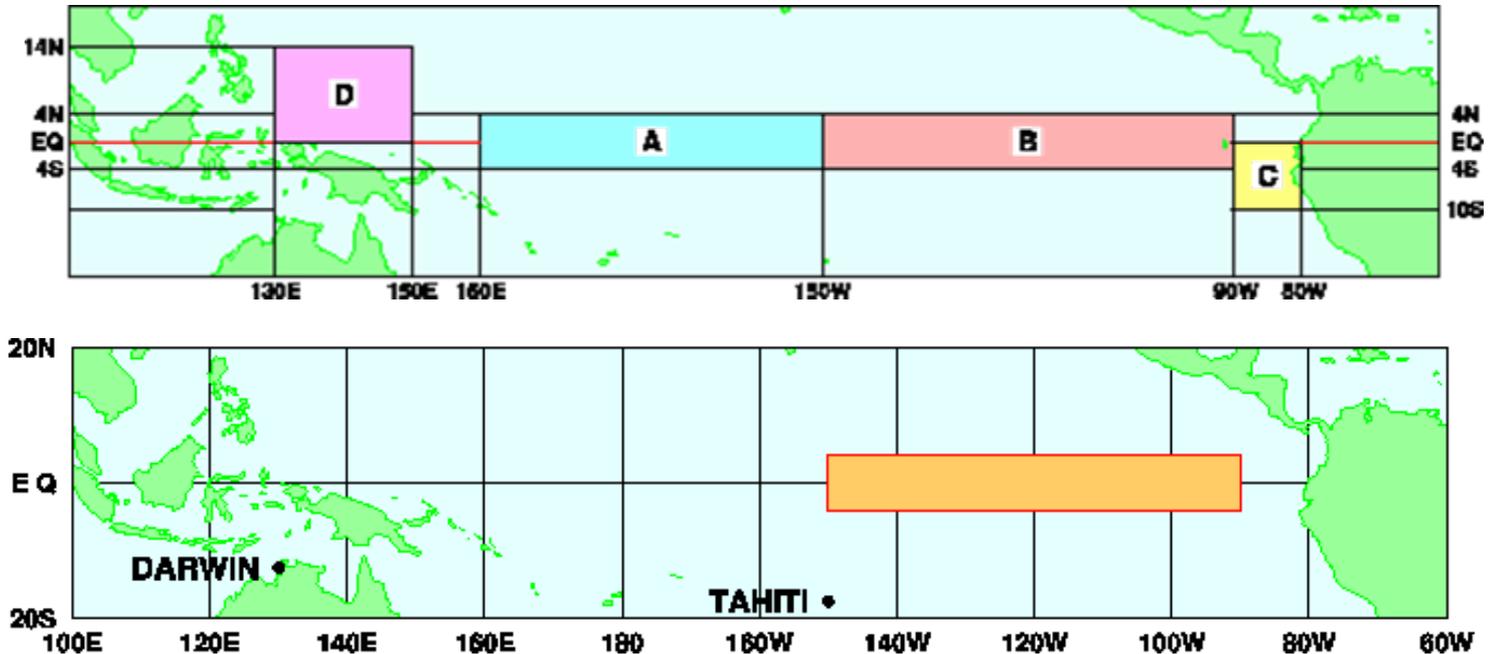


December 1998 Anomalies

La Nina Conditions

TAO Project Office/PMEL/NOAA

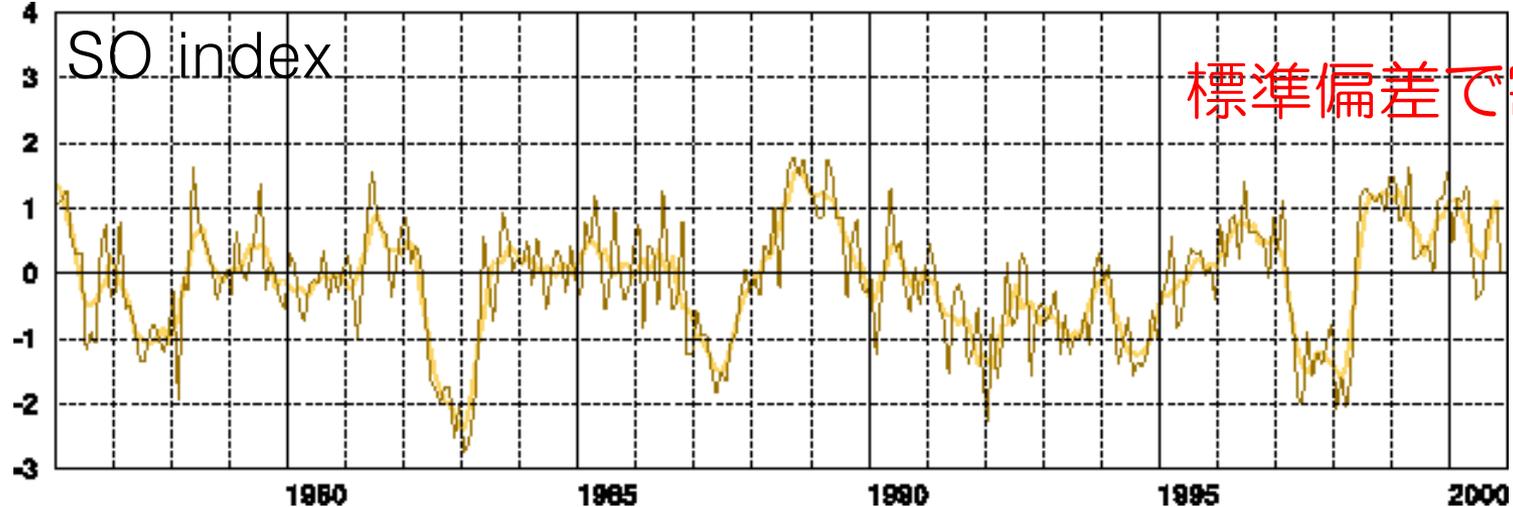
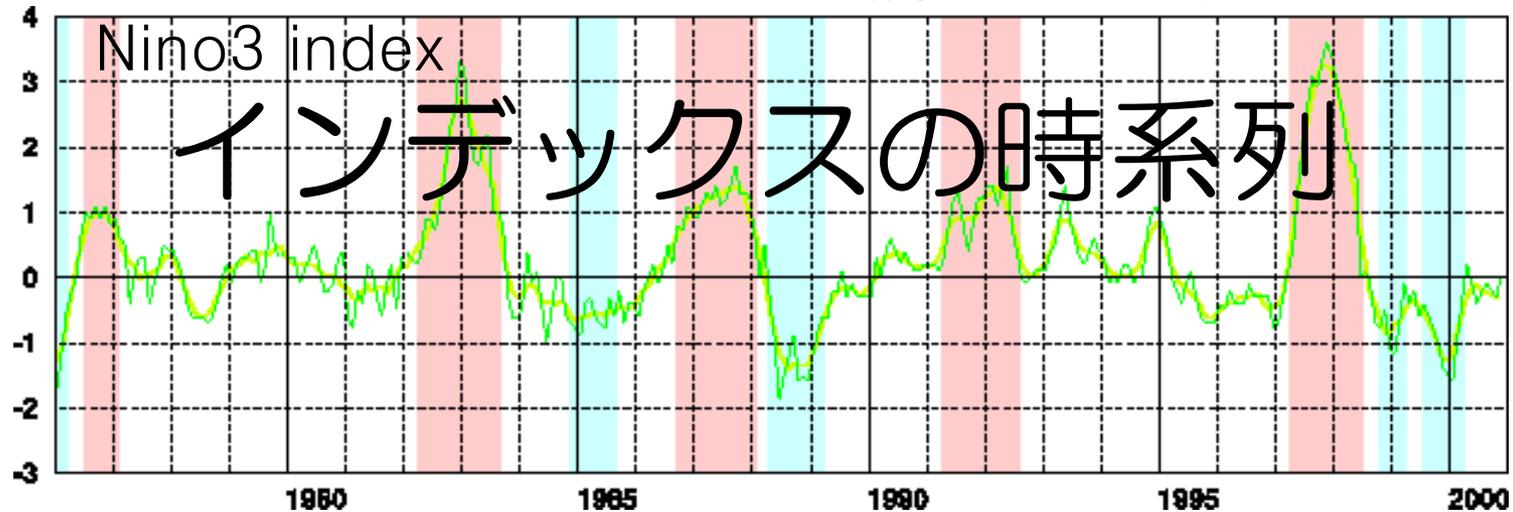
現象を代表するために 領域平均値をとる



- エルニーニョ監視速報でいうところの「エルニーニョ監視海域」は「B海域」にあたる。

<http://www.data.kishou.go.jp/climate/elnino/index/regionmap.html>

偏差をインデックスとする



エルニーニョ監視海域の海面水温と南方振動指数の推移(1976年1月～2000年12月)。

(1) エルニーニョ監視海域の海面水温の基準値との差(°C)。折線は月平均値、滑らかな太線は5か月移動平均値を示し、正の値は基準値より高いことを示す。赤はエルニーニョ現象の発生期間、青はラニーニャ現象の発生期間。

(2) 南方振動指数。正規化。太線は5か月移動平均値を示す。

<http://www.data.kishou.go.jp/climate/elnino/faq/qa/sstsoi.html>

(タヒチの海面気圧) - (ダーウィンの海面気圧)

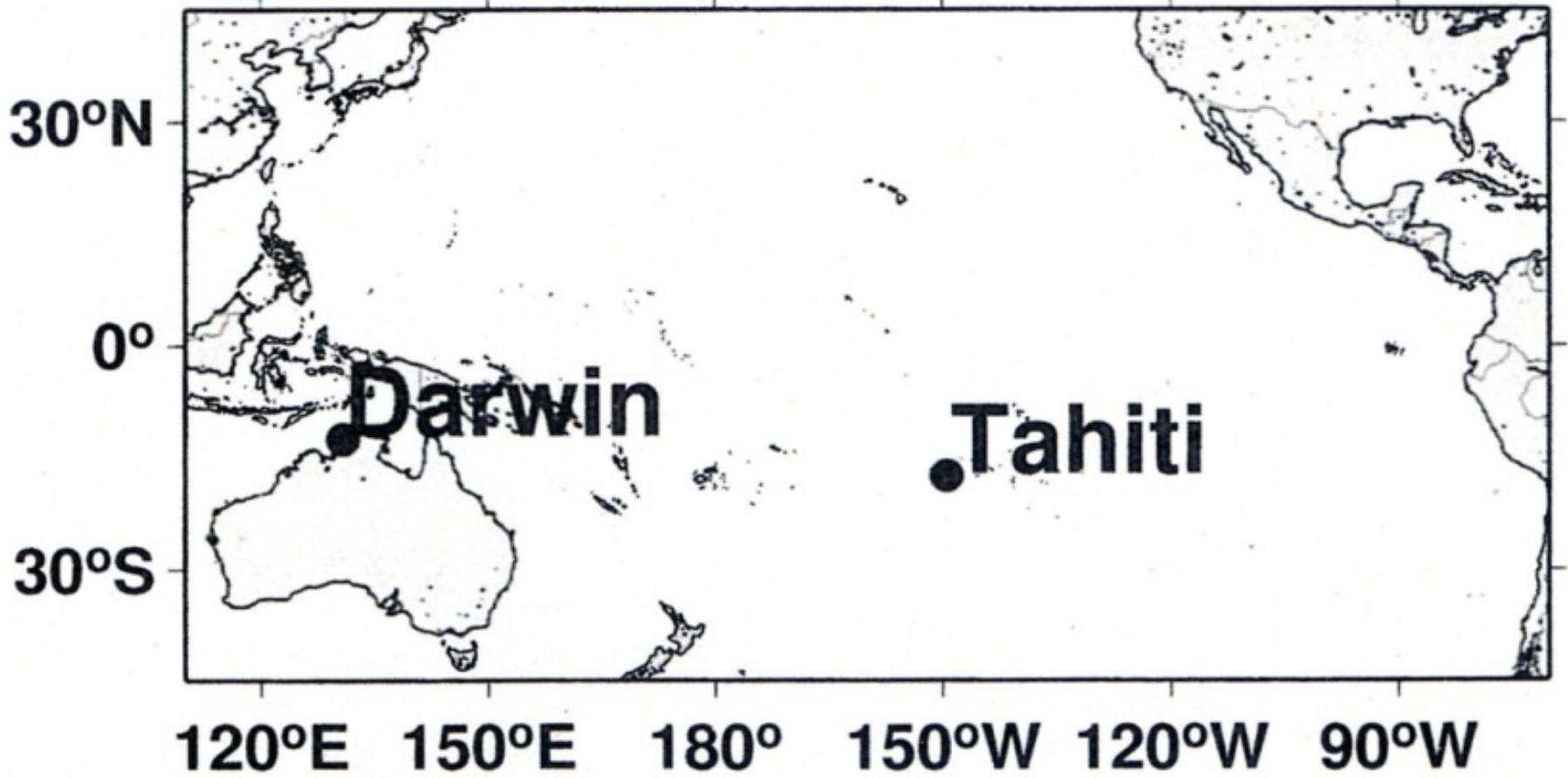
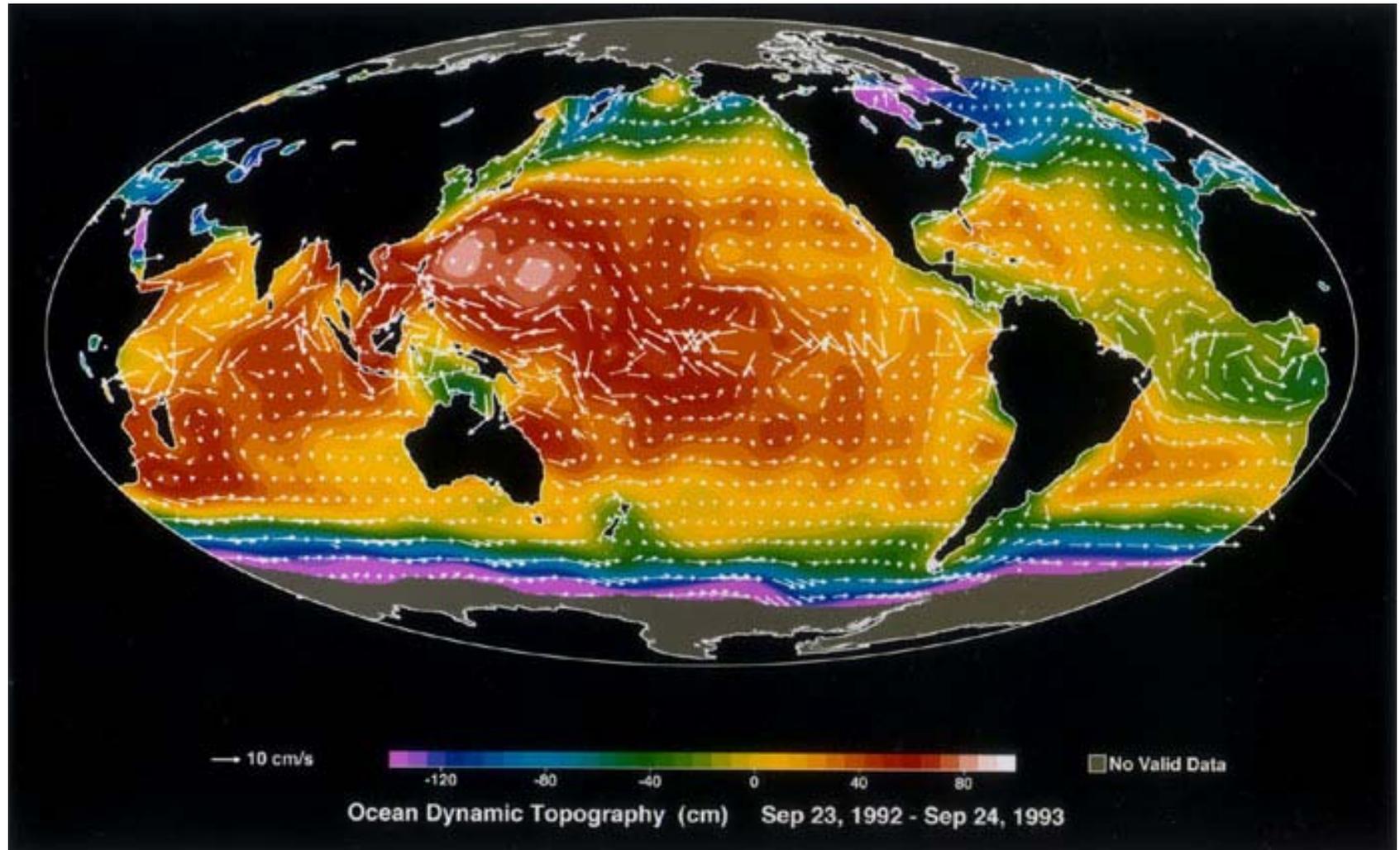


図 6: タヒチとダーウィンの位置.

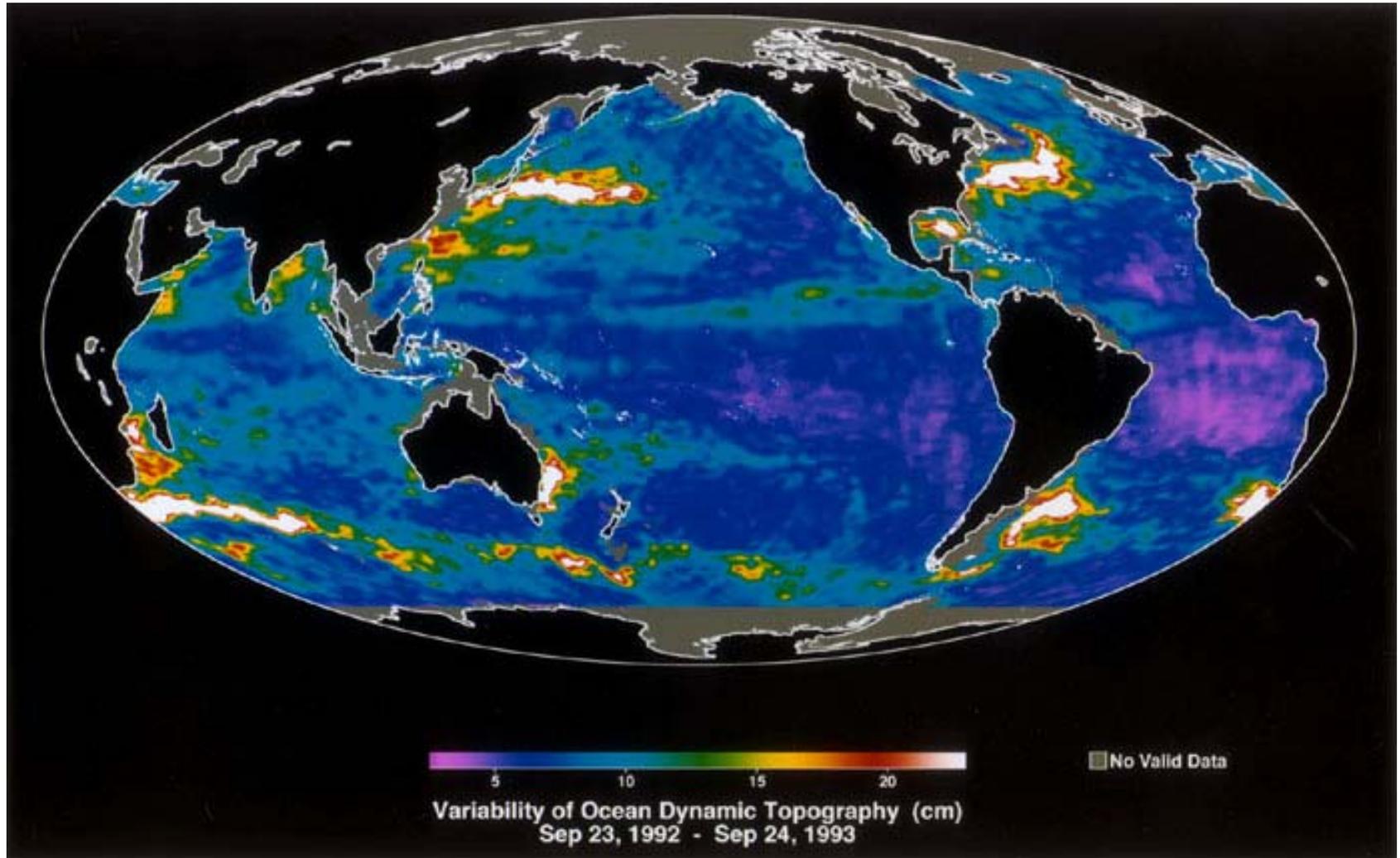
松山・谷本 (2005)

海面高度の平均値



海面の高さの分布は海流の分布を示している ⇔ 地衡流

海面高度変動の標準偏差



標準偏差が大きいところでは**変動が大きい**

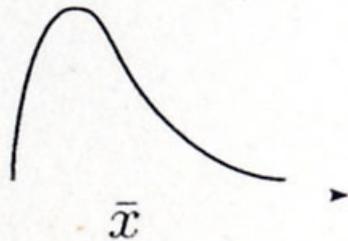
ここでの海面高度とは潮汐成分や気圧成分を除いたもの

1. 3 高次モーメント

歪度 (skewness)

最頻値 < 中央値
< 平均値

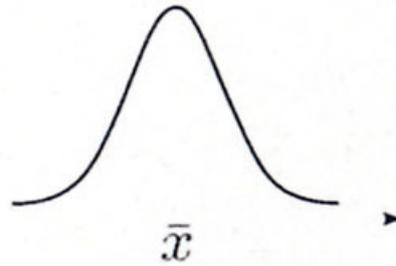
$$a_3 > 0$$



右にスノが長い

最頻値 ~ 中央値
~ 平均値

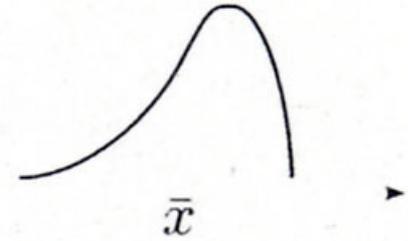
$$a_3 \approx 0$$



左右対称

平均値 < 中央値
< 最頻値

$$a_3 < 0$$



左にスノが長い

図 1.2 : 歪度 (a_3) が $a_3 > 0$, $a_3 \approx 0$, $a_3 < 0$ の場合の頻度分布の概念図。

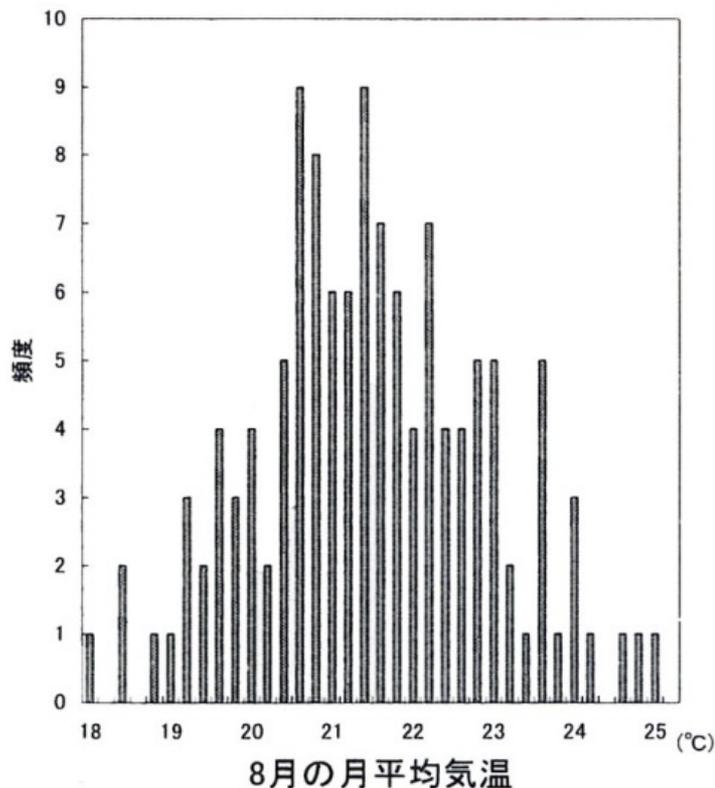
伊藤・見延 (2010)

$$a_3 = \frac{1}{(N-1)s^3} \sum_{i=1}^N (x_i - \bar{x})^3$$

Skewness: $a_3 \sim 0$

$a_3 > 0$

札幌における8月の月平均気温の
ヒストグラム(1877-2000年)



札幌における8月の降水量の
ヒストグラム(1877-2000年)

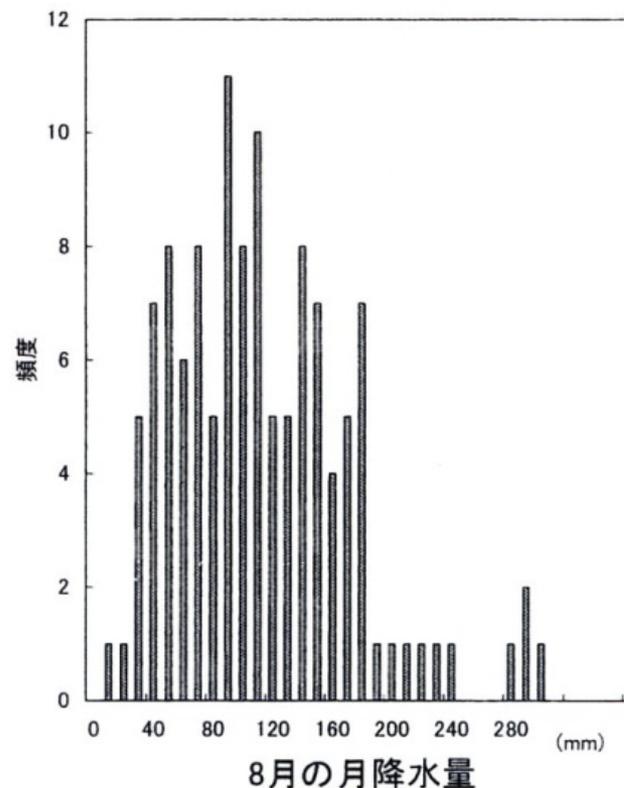


図 1: 札幌における 8 月の月平均気温 (左, 単位 °C) と月降水量 (右, 単位 mm/month) のヒストグラム.

データの期間は 1877~2000 年である. 気温の分布はほぼ左右均等に分布しているのに対し, 降水量は右の裾が長くなる分布をしている.

松山・谷本 (2005)

1. 3 高次モーメント

尖度 (kurtosis)

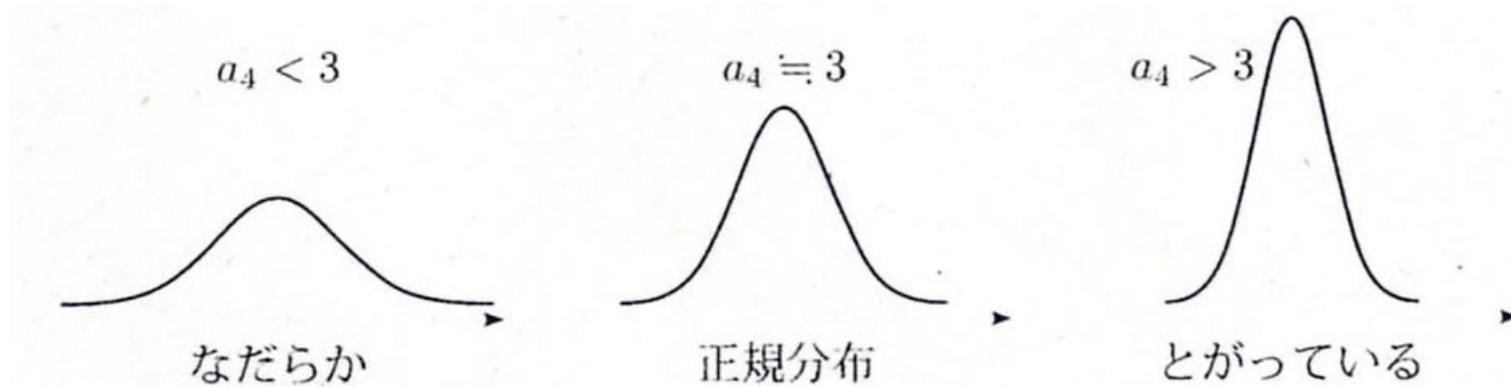


図 1.3：尖度 (a_4) が $a_4 < 3$, $a_4 \approx 3$, $a_4 > 3$ の場合の頻度分布の概念図。

伊藤・見延 (2010)

$$a_4 = \frac{1}{(N-1)s^4} \sum_{i=1}^N (x_i - \bar{x})^4$$

歪度 Skewness

500-hPa geopotential height

Sea-level pressure

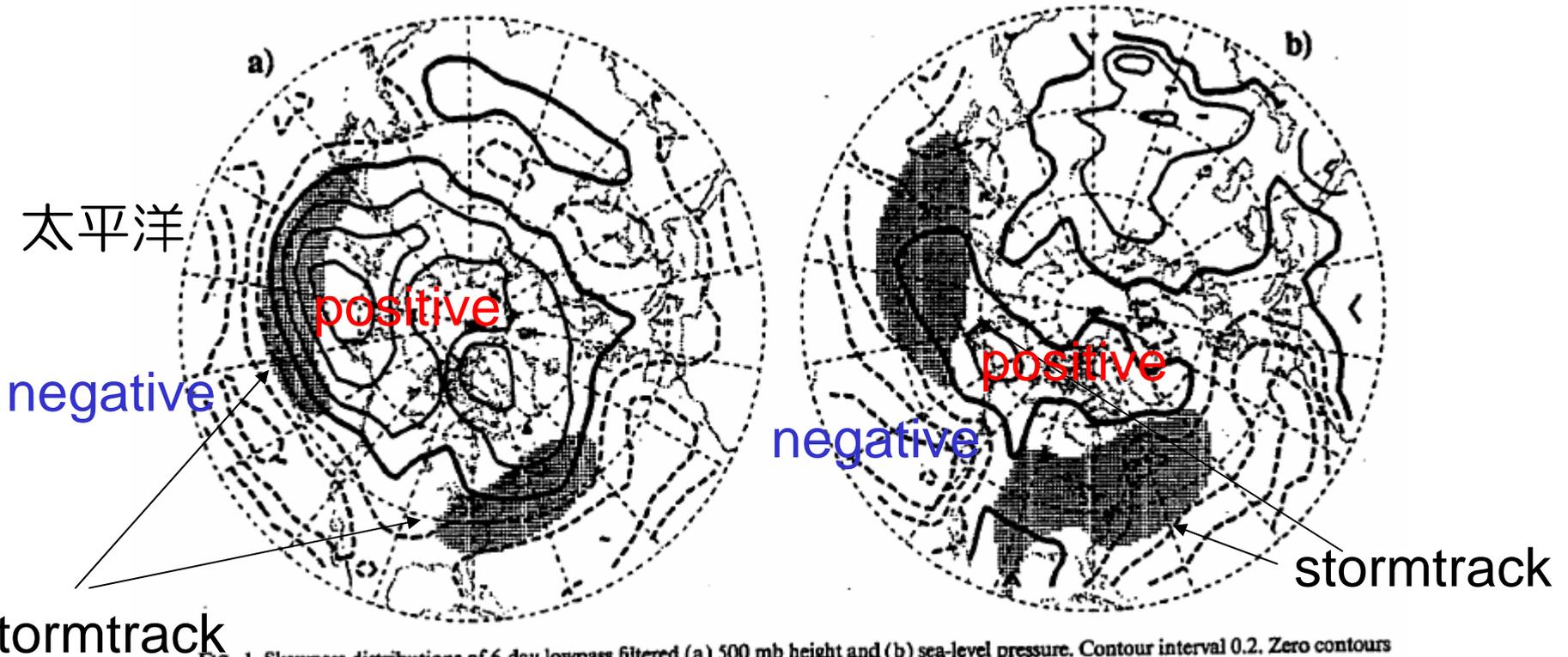


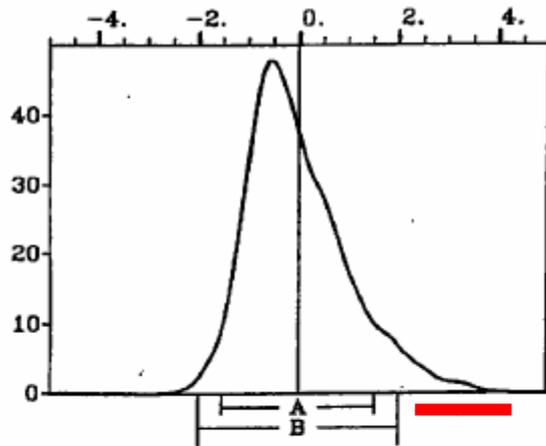
FIG. 1. Skewness distributions of 6-day lowpass filtered (a) 500 mb height and (b) sea-level pressure. Contour interval 0.2. Zero contours are thickened, and negative contours are dashed. Shaded areas indicate the climatological-mean stormtracks, defined as the regions in which the mean squared amplitude of the fluctuations in 2.5–6 day bandpass-filtered fields adjusted by the factor $[\sin^2 45^\circ / \sin^2(\text{lat})]$ exceeds 4000 (m^2) for (a) and 30 (mb^2) for (b).

ストームトラックの高緯度側ではskewnessは正で、
低緯度側では負になっている。

Positive skewness

高緯度
58N

$a_3 > 0$



Negative skewness

低緯度
23N

$a_3 < 0$

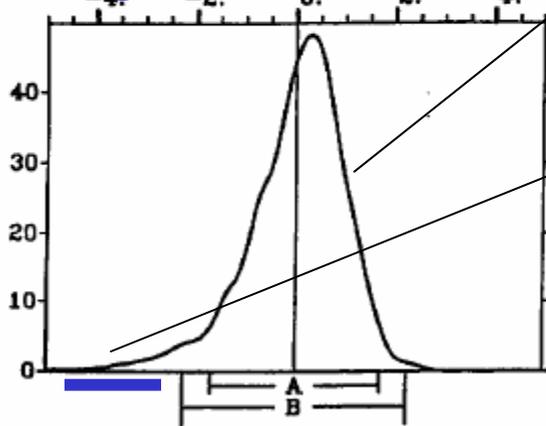


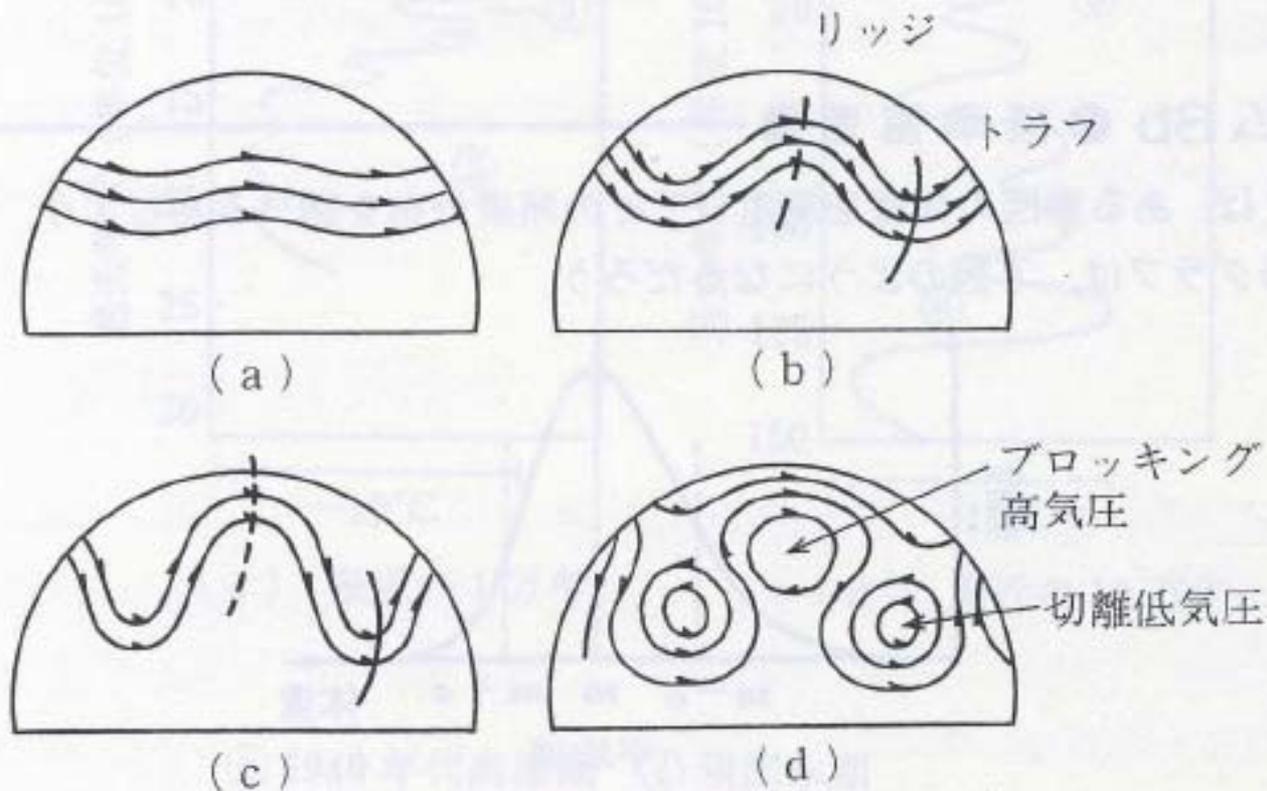
TABLE 1b. As in Table 1a but for the gridpoint at [23°N, 176°E].

Amplitude (in σ unit)	Number of samples		Contribution to skewness (a_3^*)	Cumulative skewness
	Pos	Neg		
0.00-0.25	611	566	0.00	0.00
0.25-0.50	648	399	0.00	0.00
0.50-0.75	576	356	0.01	0.01
0.75-1.00	399	285	0.02	0.03
1.00-1.25	330	285	0.04	0.07
1.25-1.50	186	174	0.01	0.08
1.50-1.75	111	157	0.02	0.10
1.75-2.00	37	90	-0.03	0.07
2.00-2.25	18	61	-0.08	-0.01
2.25-2.50	12	61	-0.09	-0.10
2.50-2.75	6	45	-0.10	-0.20
2.75-3.00	1	34	-0.08	-0.28
3.00-3.25	0	19	-0.11	-0.39
3.25-3.50	0	13	-0.09	-0.48
3.50-3.75	0	14	-0.12	-0.61
3.75-4.00	0	3	-0.03	-0.64
4.00-4.25	0	7	-0.09	-0.73
4.25-4.50	0	2	-0.03	-0.76
4.50-4.75	0	4	-0.08	-0.84
4.75-5.00	0	0	0.00	-0.84
5.00-5.25	0	2	-0.05	-0.89
5.25-5.50	0	3	-0.09	-0.98

全体で負

高緯度のブロッキング高気圧
と低緯度の切離低気圧を示し
ている

FIG. 4. Probability density distributions ($\times 100$) of 6-day lowpass filtered anomalies of 500 mb height at (a) [58°N, 176°W] and (b) [23°N, 176°E]. Height anomalies (abscissa) are normalized by their standard deviation. Density distribution is based on a biweight kernel estimator (bin and window widths are 0.1 and 0.4 standard deviations, respectively). Amplitude ranges "A" and "B" are referred to in the text.



北半球の対流圏中～上層の西風の流れ (a) がしだいに蛇行し (b), その振幅がさらに拡大し (c), ついにはブロッキング高気圧と切離低気圧が形成される (d).

図 6・7 偏西風波動の増幅とブロッキング高気圧, 切離低気圧の発達過程
 (Barry R. G. and Chorley R. J.: Atmosphere, Weather and Climate, Methuen, 1982)

まとめ

- 代表値には平均・中央値・最頻値がある。
- 値は平均と偏差とに分けられる。
- 長い時間スケールでは、平均も変化することがある。
- 分散・標準偏差はデータのばらつき具合を示すもので、変動の強さの指標となる。
- データの偏りを見るには高次のモーメントが有効である。