

# 気象情報解析特論第7回

主成分分析を用いた気象データの分析 (EOF解析)

神山 翼, @t\_kohyama,  
[tsubasa@is.ocha.ac.jp](mailto:tsubasa@is.ocha.ac.jp),

理3-703

今日は、lat-lon格子の気象データから  
卓越する変動を見つける方法を学びます

主成分分析を用いた気象データの分析 (EOF解析)

EOF解析を使うと、データから重要な線型変動を取り出して  
客観的にインデックスを定義できる

主成分分析の知識に加えて  
大気海洋データならではの計算の工夫が必要

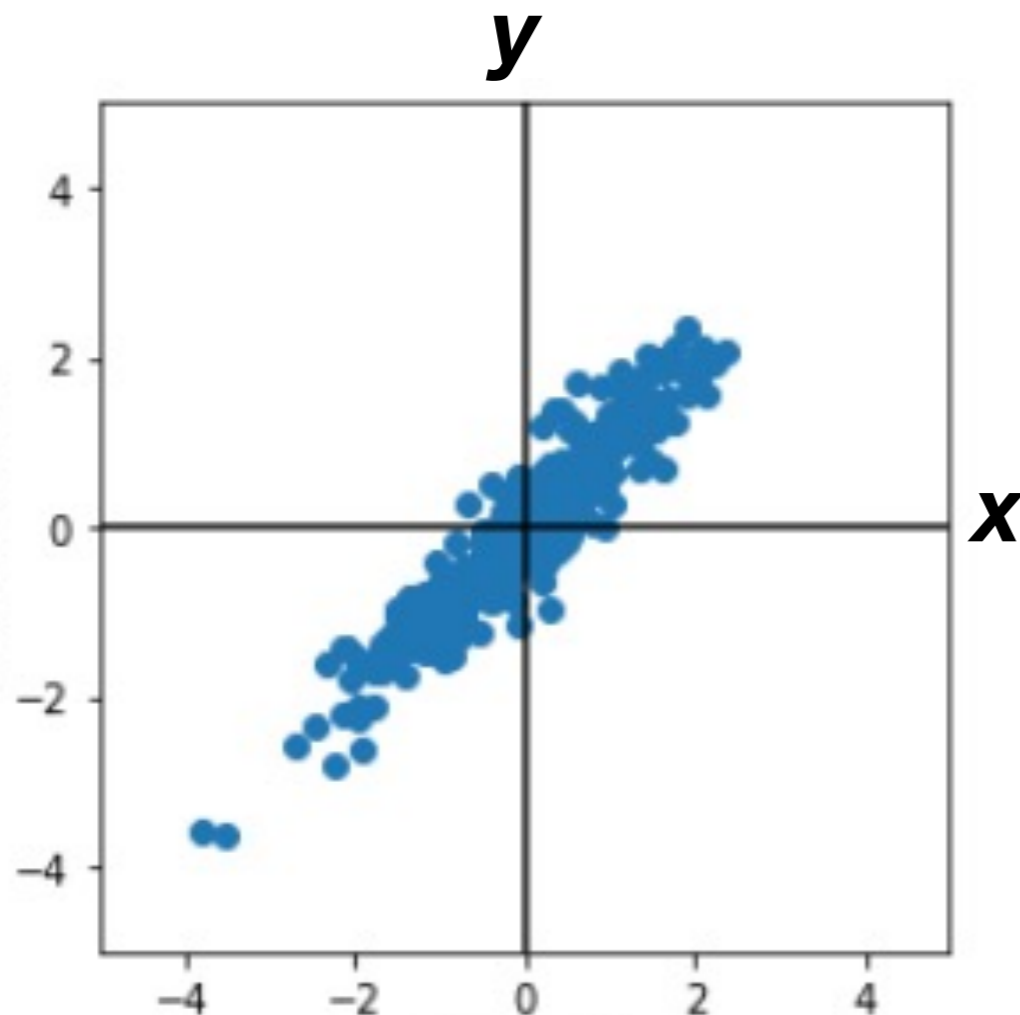
卓越変動の空間構造 (EOF) と時間変動 (PC) を  
取り出した上で、その変動の寄与率を算出する

# 主成分分析の復習

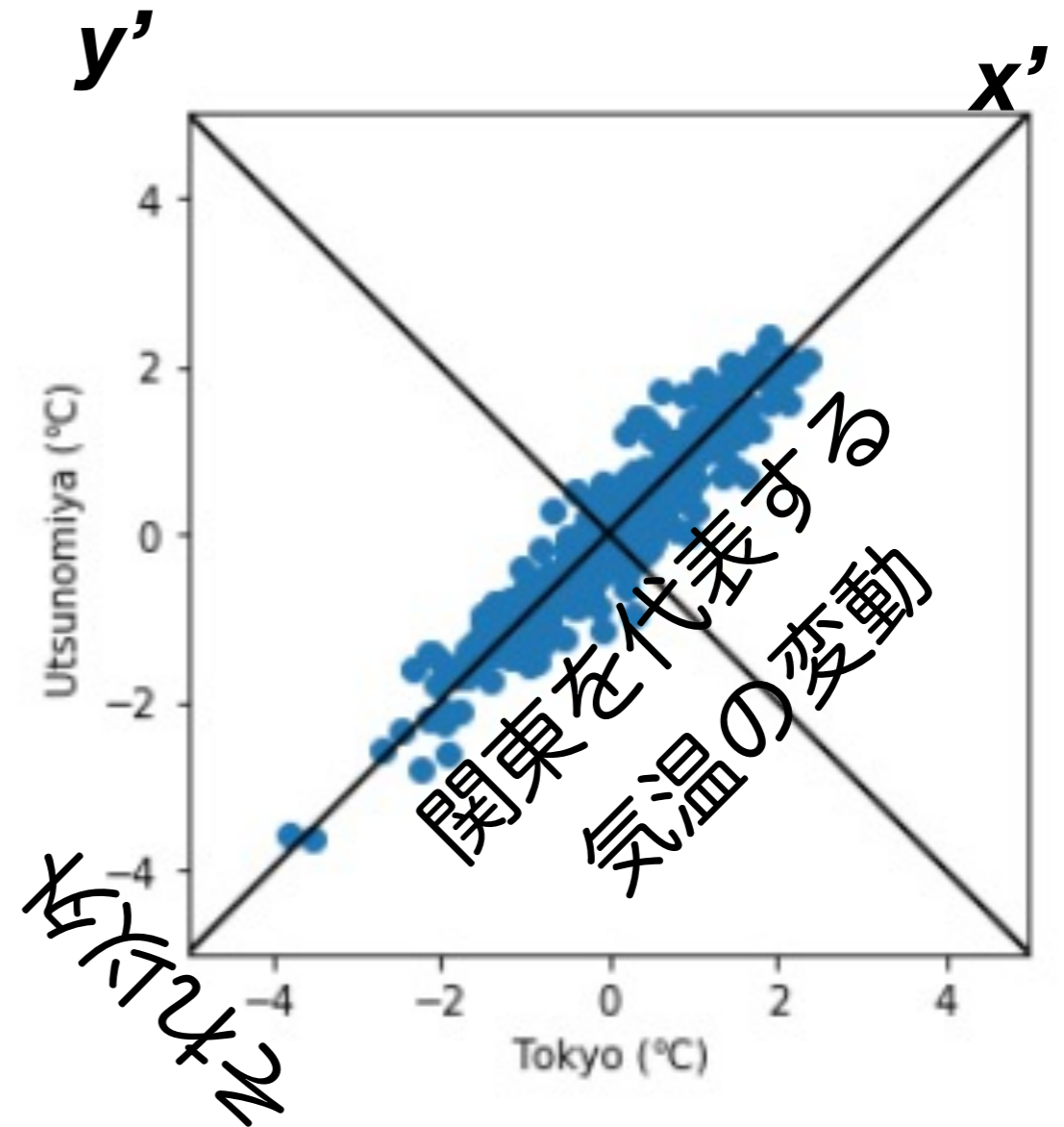
# 主成分分析の流れ（環境情報論第11, 12回の復習）

座標軸を最大分散方向とその直交方向  
になるように回転させる分析

宇都宮の気温偏差

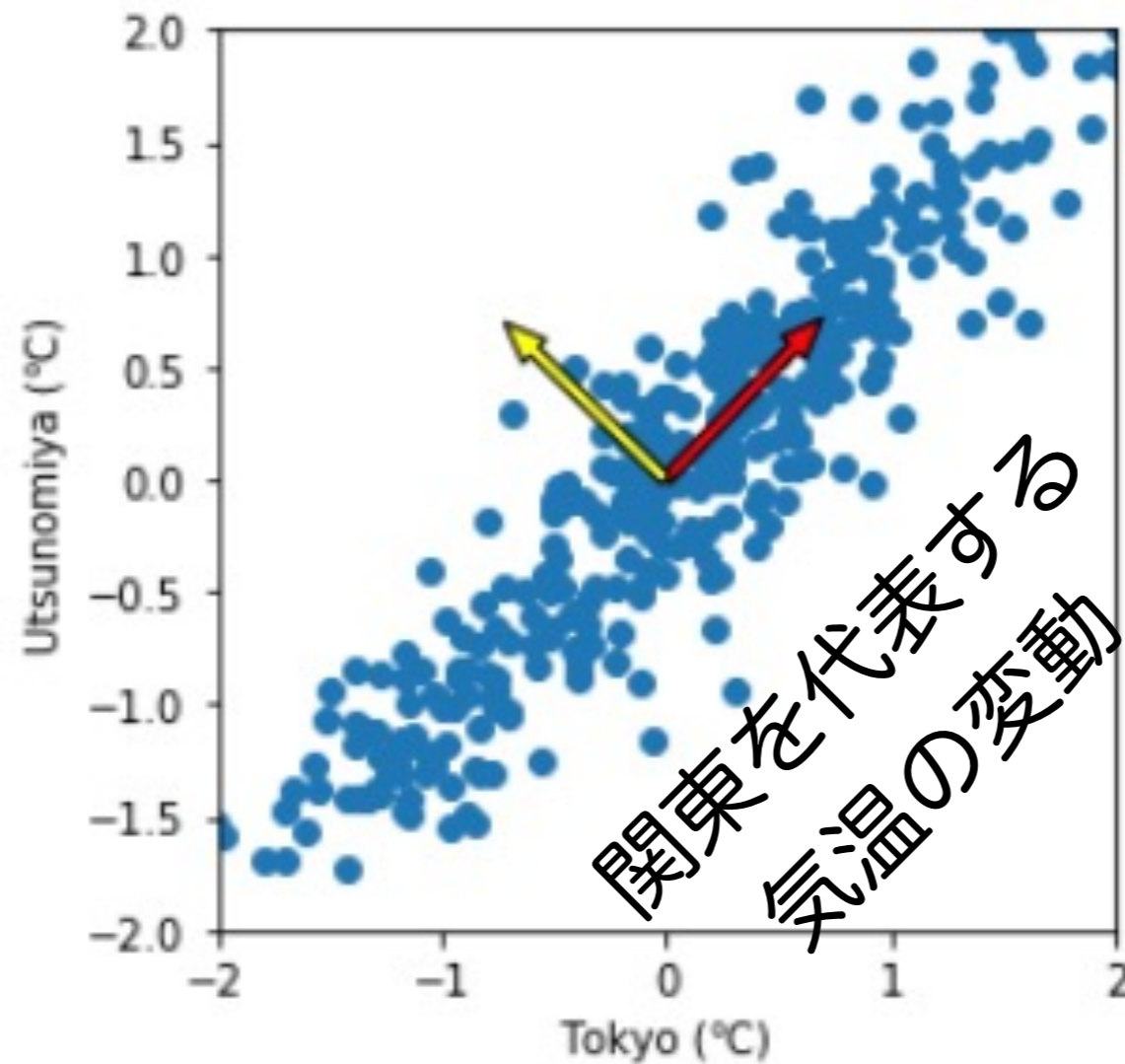


東京の気温偏差



# 主成分分析の流れ（前回の復習）

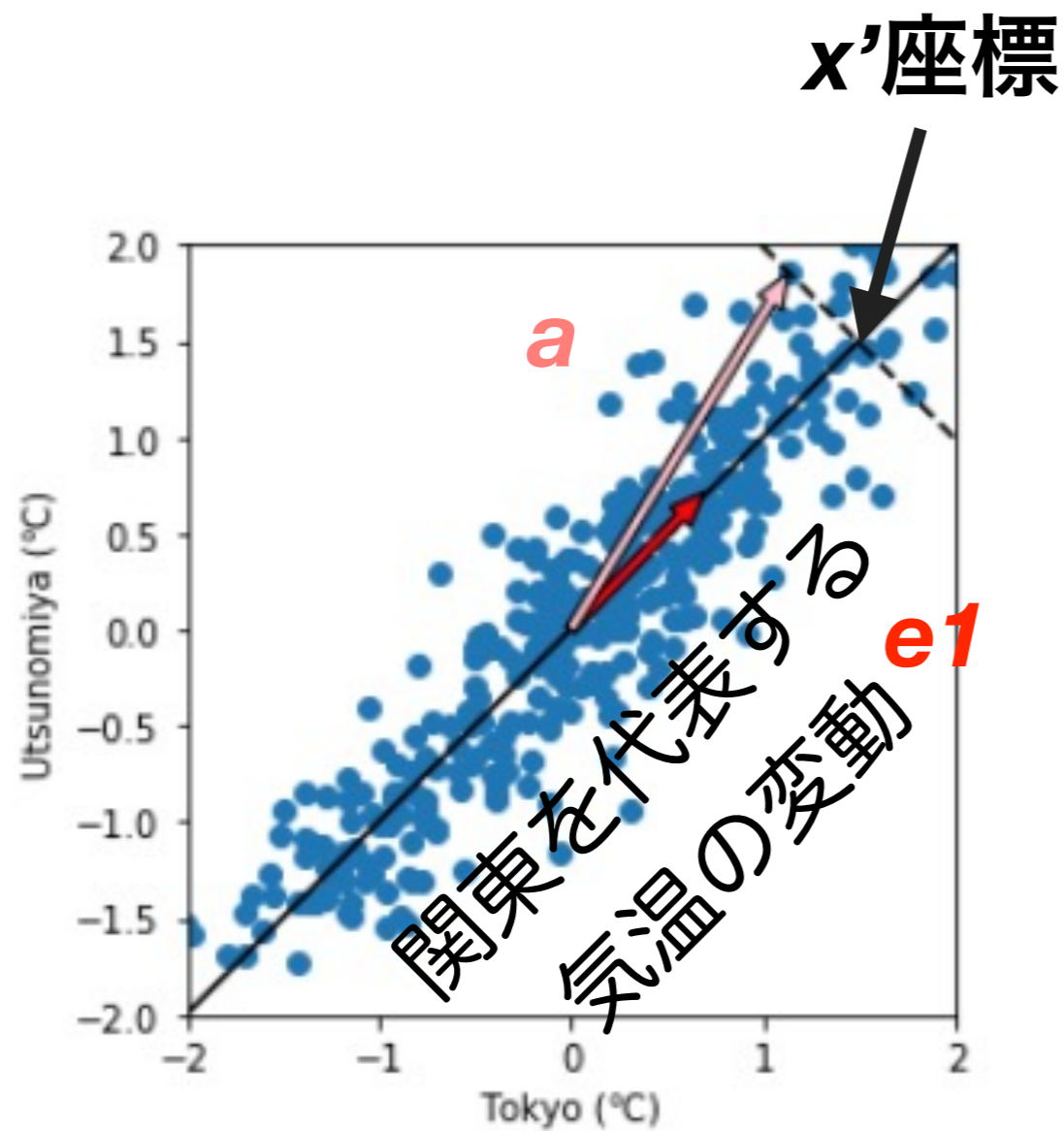
共分散行列の固有値固有ベクトルを計算し、  
分散最大方向を抽出



分散最大方向の単位ベクトル  $e_1$ （赤）と  
その直交方向の単位ベクトル  $e_2$ （黄色）

# データ点 $a$ の $x'$ 座標を求める方法

基底ベクトルとの内積を取れば良い

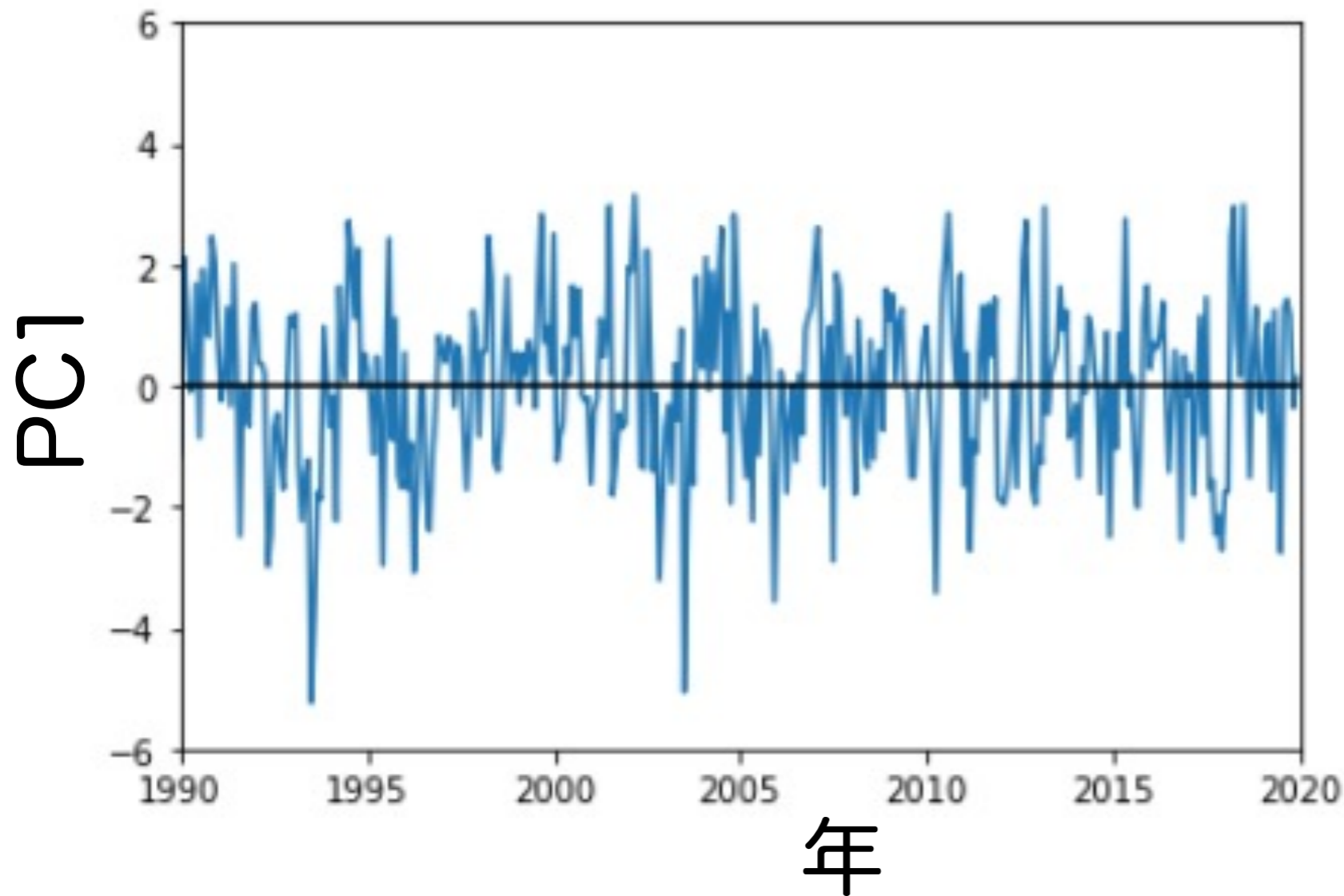


$$x' = |\vec{a}| \cos \theta = \vec{e}_1 \cdot \vec{a}$$

これを基底  $e_1$  に  
**射影する**という

# 第一主成分の時系列 = PC1

全ての月について、x'座標を計算して並べた時系列



「関東代表」の  
時系列

PC1が正の時は  
東京も宇都宮も  
暖かい

※ PC = Principal Component (主成分)

# まとめ（あとでゆっくくり資料読んで）

## I. データ行列の定義

$$X := \begin{pmatrix} \vec{x}_1 \\ \vec{x}_2 \end{pmatrix} = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \end{pmatrix}$$

## II. 共分散行列の計算

$$C := \begin{pmatrix} \text{COV}(\vec{x}_1, \vec{x}_1) & \text{COV}(\vec{x}_1, \vec{x}_2) \\ \text{COV}(\vec{x}_2, \vec{x}_1) & \text{COV}(\vec{x}_2, \vec{x}_2) \end{pmatrix} = XX^T/N$$

## III. 共分散行列の固有値固有ベクトルを求める

$$CE = E\Lambda$$



# まとめ（あとでゆっくり資料読んで）

IV. 固有値の大きい固有ベクトルから順に，第1主成分，第2主成分，...の方向とする（このとき，第1主成分は分散最大方向となり，第2主成分以降は「固有ベクトルが直交する」という制約の中で分散の大きい順となる。）

$$E = \begin{pmatrix} \vec{e}_1 & \vec{e}_2 \end{pmatrix}$$
$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \lambda_1 \geq \lambda_2 > 0$$

V. 固有ベクトルを並べた行列に，元の行列を射影すると，PC1時系列，PC2時系列，...を得る（このとき，それぞれのPC時系列は無相関であり，固有値は各PC時系列の分散を与える。主成分分析前後で，分散の和は保存される。）

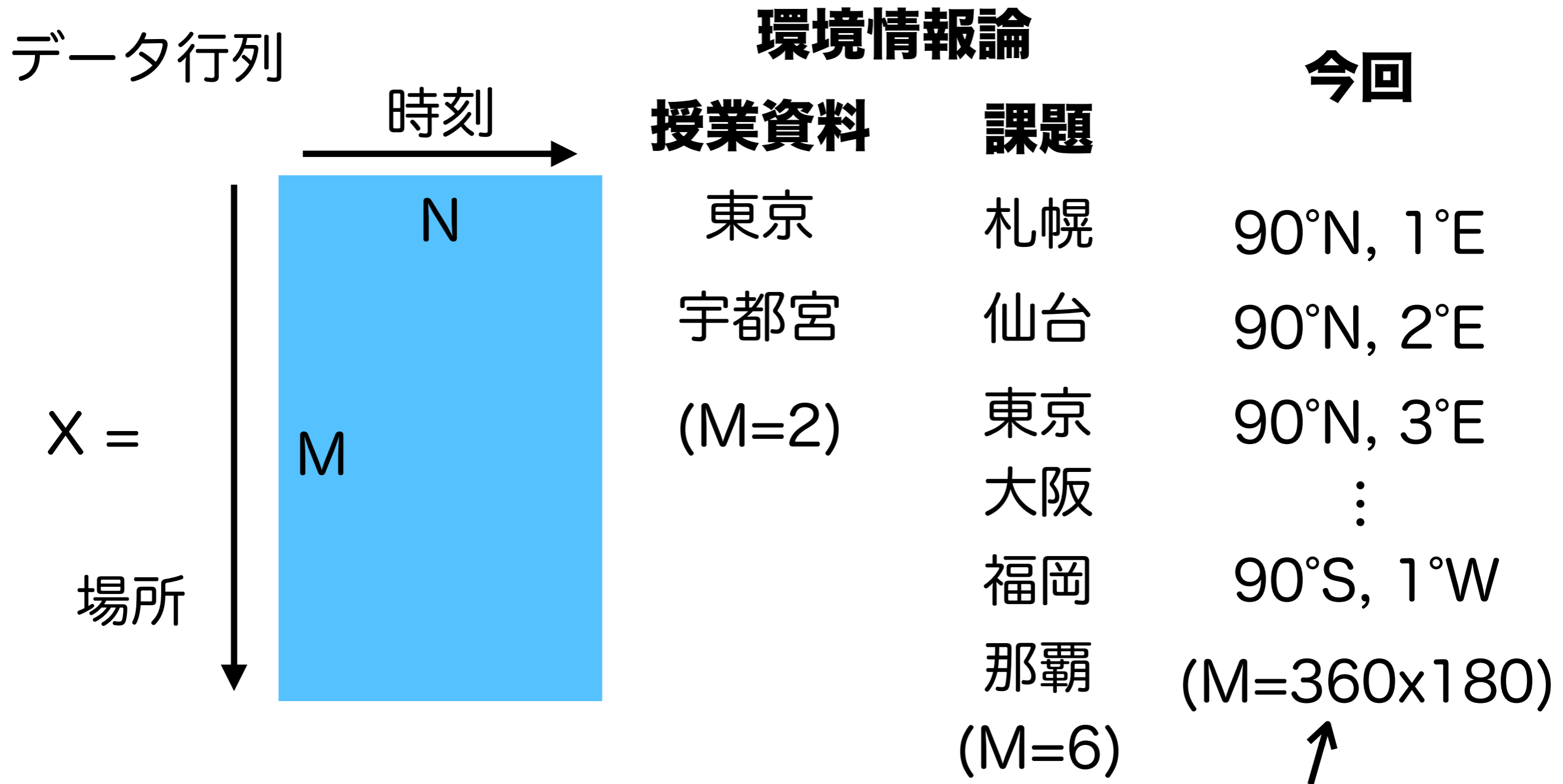
$$Z = E^T X$$
$$(\Leftrightarrow \vec{PC1} = \vec{e}_1^T X, \vec{PC2} = \vec{e}_2^T X)$$

ただし

$$Z := \begin{pmatrix} \vec{PC1} \\ \vec{PC2} \end{pmatrix}$$

これをlat-lon格子データに応用する

# 今回：2次元気象場における主成分分析 インデックスの定義などに頻繁に用いられる

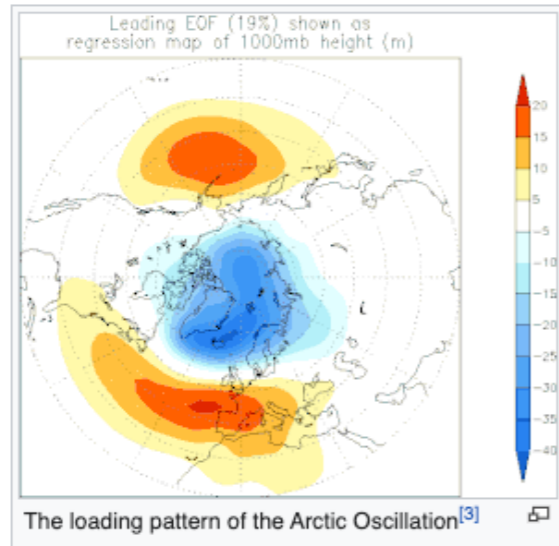


※共分散行列を $\cos \phi$  ( $\phi$ は緯度) で重み付け・

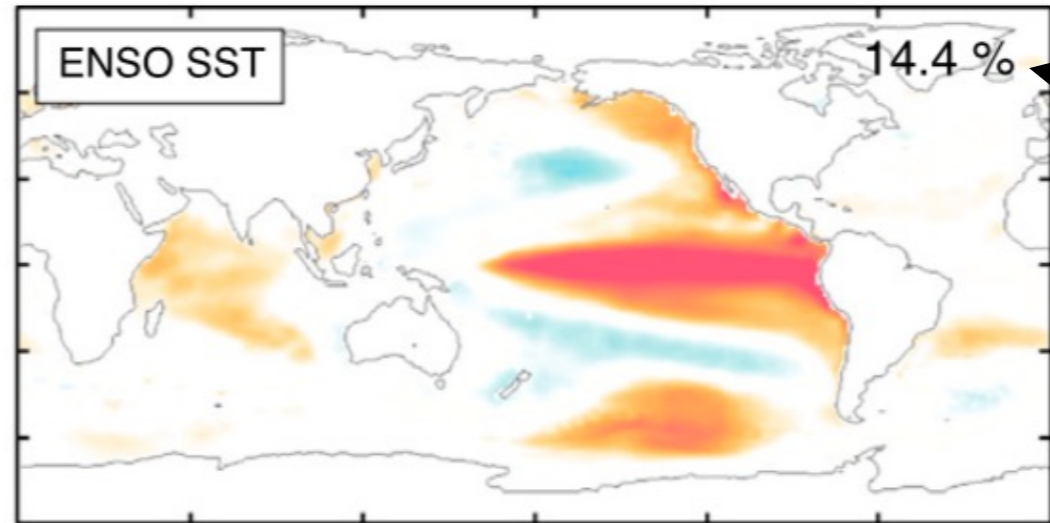
NaNのデータを除外する必要がある

# 今回：2次元気象場における主成分分析

## 固有ベクトル

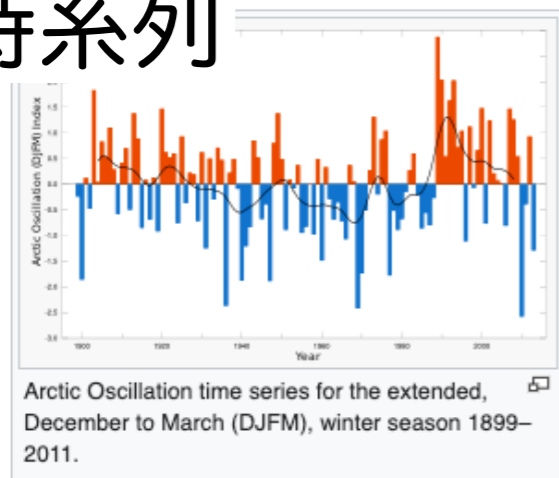


## 固有ベクトル



寄与率

## PC1時系列



## 世界の海面水温の第一主成分

=エルニーニョ南方振動のインデックス

主成分分析を行うと  
こんな風に客観的に  
卓越変動を取り出せる

## 北半球における

地表気圧の第一主成分

= 北極振動のインデックス

lat-lon格子データ特有のテクニック

大気海洋のlat-lonデータは  
大抵は空間方向の方が次元が大きい

データ行列の行と列を入れ替えて  
主成分分析するとよい

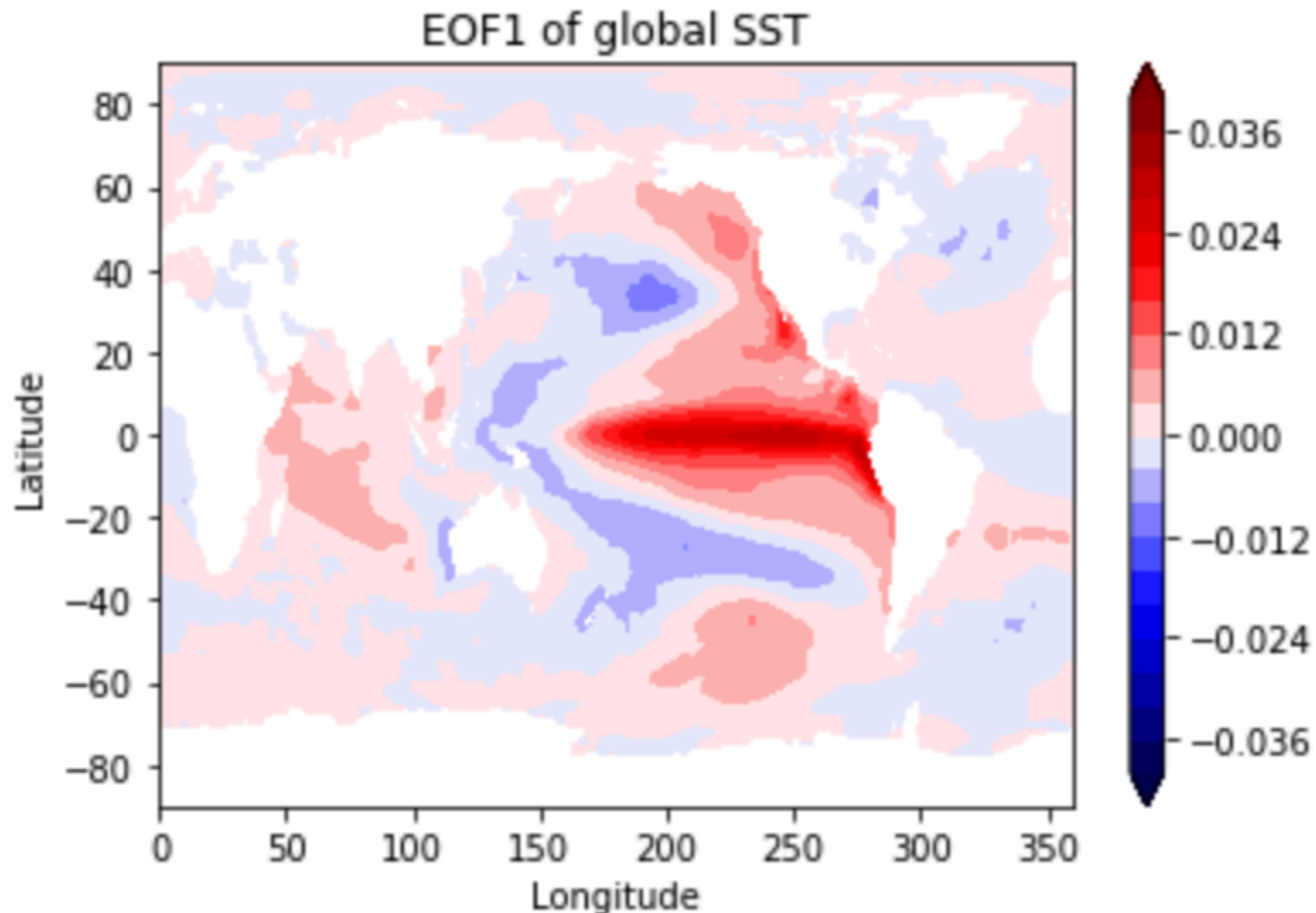
固有ベクトルがPC時系列を与え、  
固有ベクトルに元データを射影した  
係数行列が空間構造 (=EOFという)  
を与える ← 第12回と逆

# 海面水温の第一主成分

# 全球海面水温のEOF1はENSO

客観的にENSOを取り出すことができた！

(Niño3.4指数などは、太平洋の変動を見たいために主観的に決めたインデックスだった)



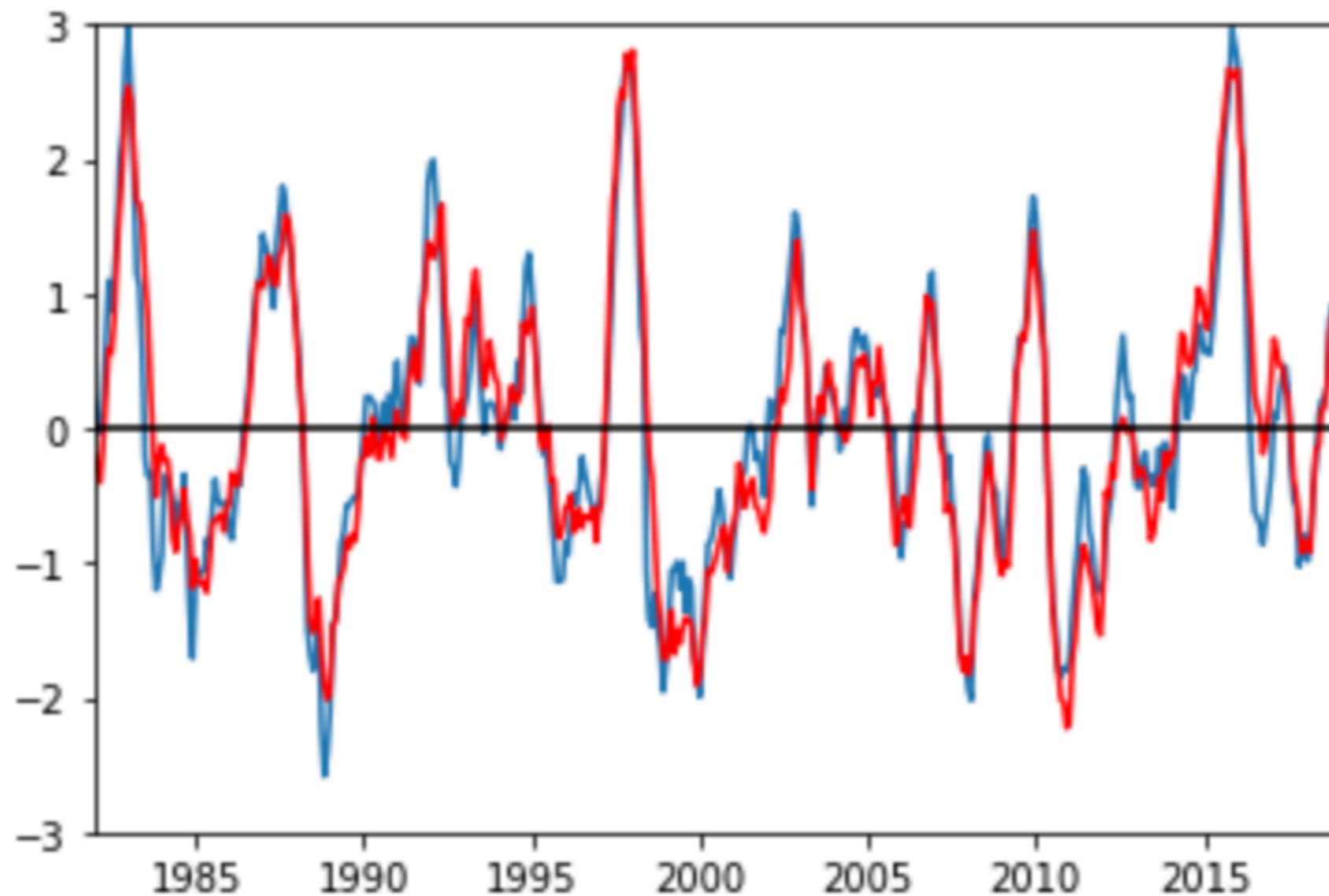


# 全球海面水温のPC1はほぼNiño3.4指数

Niño3.4インデックスは

全世界のSSTの第一主成分と一致するからこそ

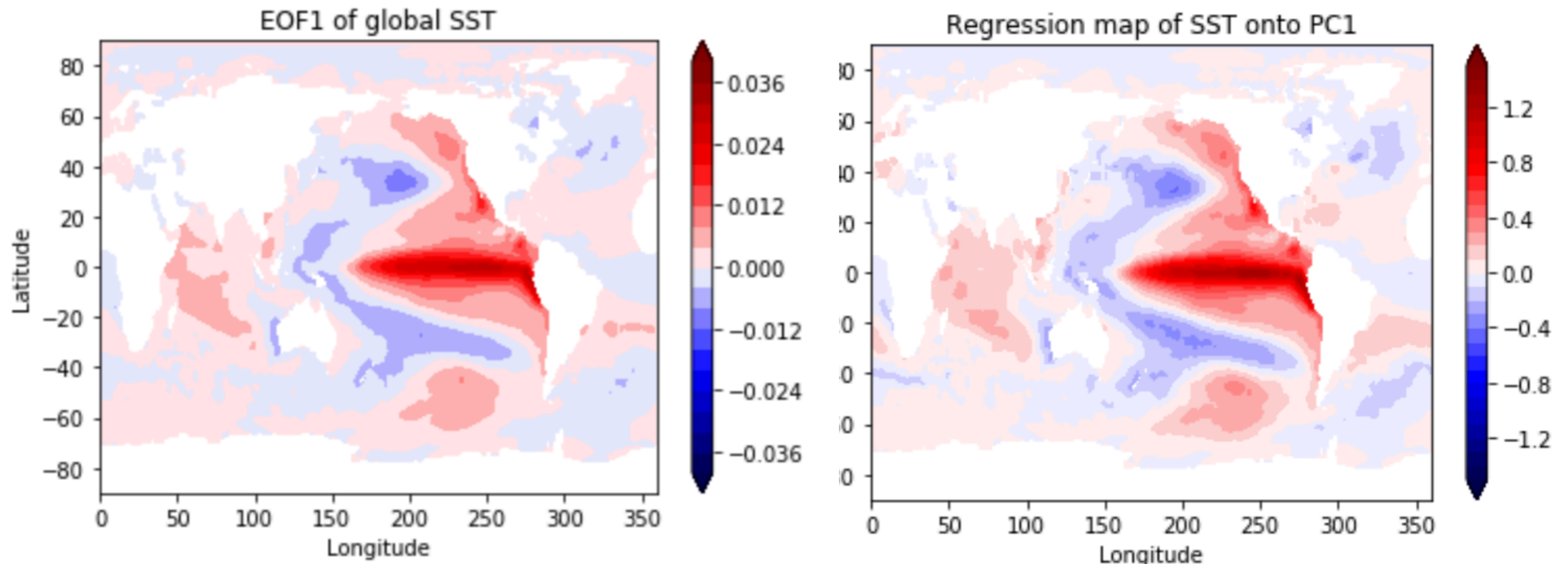
客観的で良いインデックスの定義だと言える



# 振幅の情報を得るために 正規化されたPC1への回帰図を書く

空間パターンはEOF1と同一

振幅はPC1が1標準偏差と等しいときに  
実現される大きさ



全球海面水温のPC1の寄与率は16%

思い出そう：

寄与率 = 全分散に対するEOF1の説明する分散の比

```
ev1 = L[0, 0]/np.trace(L)
```

```
ev1
```

```
0.16438340262378257
```

**世界のSST経年変動（=数年程度の時間スケールの変動）のうち約1/5弱をENSOだけで説明**

まとめ

# EOF解析を行うことによって得られる 第1主成分に関する情報は以下の3つ

- PC1時系列
- EOF1の空間パターン  
(実際はPC1時系列への回帰図として示す)
- EOF1の寄与率

第二主成分以降も同様です  
少しずつ慣れていってください

今日は、lat-lon格子の気象データから  
卓越する変動を見つける方法を学びます

主成分分析を用いた気象データの分析 (EOF解析)

EOF解析を使うと、データから重要な線型変動を取り出して  
客観的にインデックスを定義できる

主成分分析の知識に加えて  
大気海洋データならではの計算の工夫が必要

卓越変動の空間構造 (EOF) と時間変動 (PC) を  
取り出した上で、その変動の寄与率を算出する

本日の導入パートは以上です。  
何でも良いので渡した紙に  
授業に関係のあるコメントを  
してください（出席代わり）。

コメント拾いが終わったら、  
早速今日のプログラミングに進みましょう。