

## Metody v klimatologii

### II. Využití vícerozměrných metod, PCA, EOF

#### Metody vícerozměrné analýzy

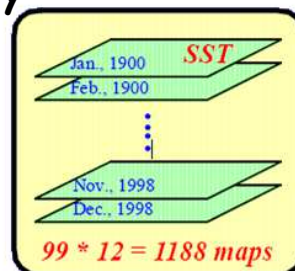
- Velké datové soubory
- Redundantní informace
- Stochastický charakter studovaných procesů
- Signál je často překryt šumovou složkou

##### Hlavní cíle

- **Redukovat množství dat** (separovat klimatický signál od přirozené variability resp. od šumu)
- **Detekovat strukturu** vztahů mezi proměnnými (klasifikovat, vytvořit typologii dat)
- **Identifikovat fyzikální procesy**, které generují klimatický signál

Vícerozměrné metody nabízí užitečný způsob, jak si zorganizovat data

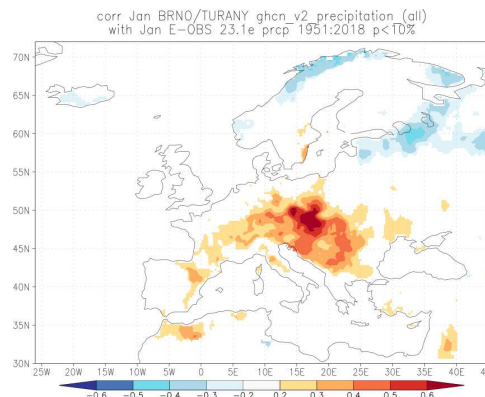
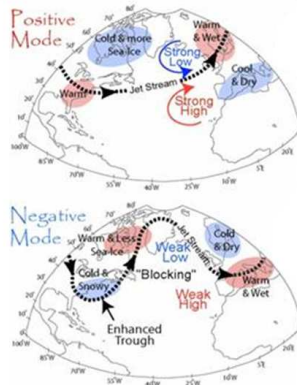
Vícerozměrné metody často dokáží prezentovat časovou a prostorovou variabilitu klimatických dat v přehledné (snáze interpretovatelné) podobě



## Metody vícerozměrné analýzy

- Vícerozměrné metody umožňují v datech **identifikovat hlavní vzory (módy) variability klimatu** (NAO, ENSO, ...)

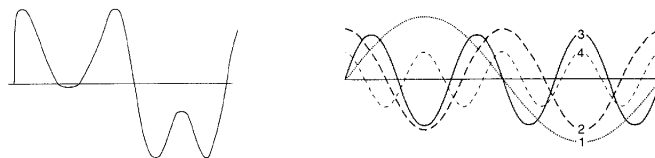
### North Atlantic Oscillation



## Metody vícerozměrné analýzy

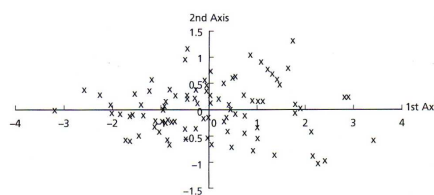
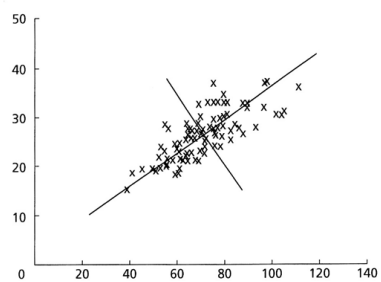
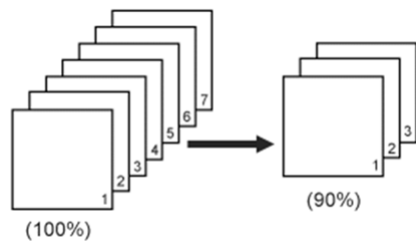
- Princip na příkladu jednorozměrných dat
- Např. časovou řadu o délce  $N$  členů lze vyjádřit jako její průměr plus  $N-1$  funkcí  $\sin$  a  $\cos$  s různou amplitudou a frekvencí
- Fourierova analýza**
- Koeficienty  $(a, b)$  značí zastoupení dané složky v původní řadě

$$y_t = a_1 \cdot \sin \frac{2\pi}{T} t + b_1 \cdot \cos \frac{2\pi}{T} t + a_2 \cdot \sin \frac{4\pi}{T} t + b_2 \cdot \cos \frac{4\pi}{T} t + a_3 \cdot \sin \frac{6\pi}{T} t + b_3 \cdot \cos \frac{6\pi}{T} t + \dots$$

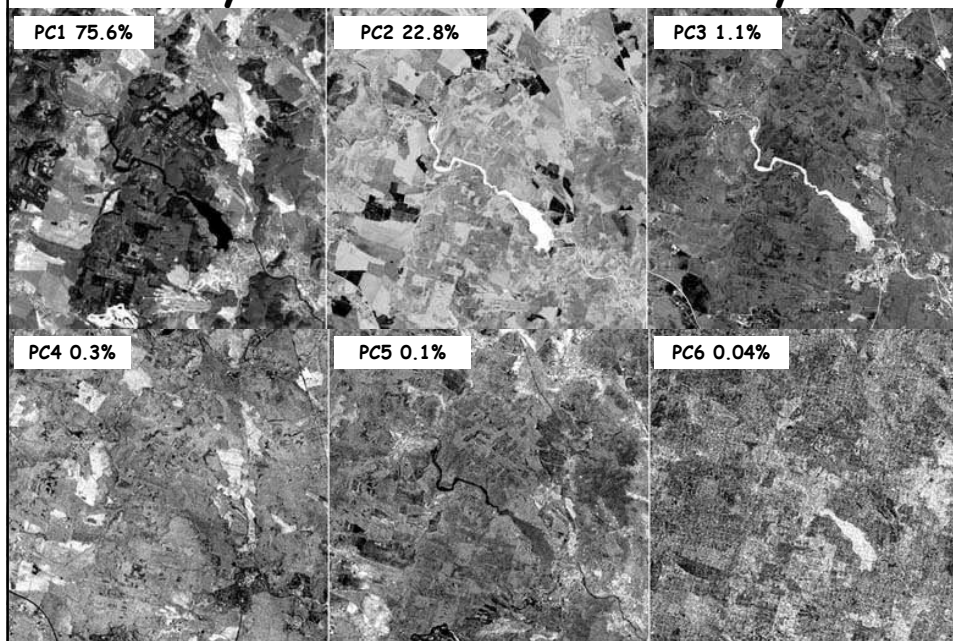


- Abychom popsali z původní řady to podstatné, nemusíme použít všech  $N-1$  funkcí.
- Nejvíce zastoupené („nejdůležitější“) složky prozradí typické chování v čase (např. cyklické chování).

## Analýza hlavních komponent Principal Component Analysis (PCA)



## Příklad využití PCA v obrazové analýze



## Analýza hlavních komponent

Nové proměnné jsou vypočteny jako lineární kombinace původních proměnných

$$PC_1 = a_1TM_1 + a_2TM_2 + a_3TM_3 + a_4TM_4 + a_5TM_5 + a_6TM_7$$

$$PC_2 = b_1TM_1 + b_2TM_2 + b_3TM_3 + b_4TM_4 + b_5TM_5 + b_6TM_7$$

...

Číslo PC	Vlastní čísla	Procenta rozptylu	Kumulov. procenta	Zátěže					
				TM 1	TM 2	TM 3	TM 4	TM 5	TM 7
1	2262,96	75,62	75,62	0,243	0,181	0,346	0,230	0,728	0,454
2	682,34	22,80	98,42	0,115	0,050	0,229	-0,936	-0,012	0,237
3	33,80	1,13	99,55	0,553	0,323	0,513	0,201	-0,531	-0,064
4	7,79	0,26	99,81	-0,264	-0,141	-0,037	0,168	-0,432	0,833
5	4,54	0,15	99,96	0,712	-0,102	-0,668	-0,034	0,000	0,186
6	1,21	0,04	100,00	-0,212	0,911	-0,343	-0,044	-0,022	0,069

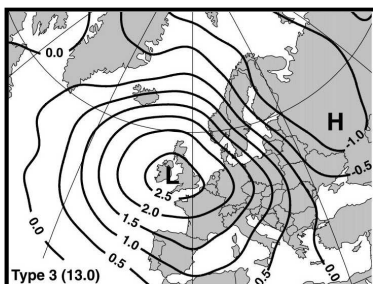
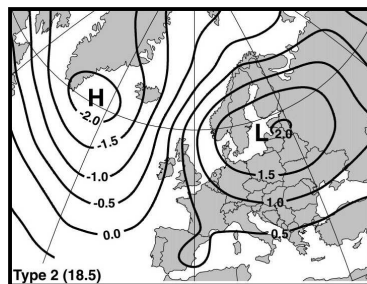
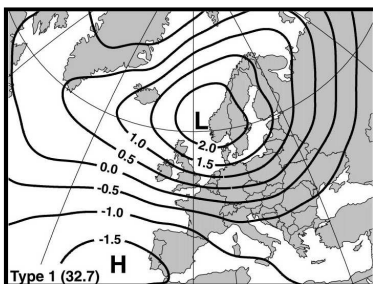
eigenvalue = vlastní číslo

eigenvector = vlastní vektor

zátěž = loading

## Příklad využití PCA v klimatologii

Jaké jsou hlavní módy cirkulace při výskytu zimních vichřic ve střední Evropě?



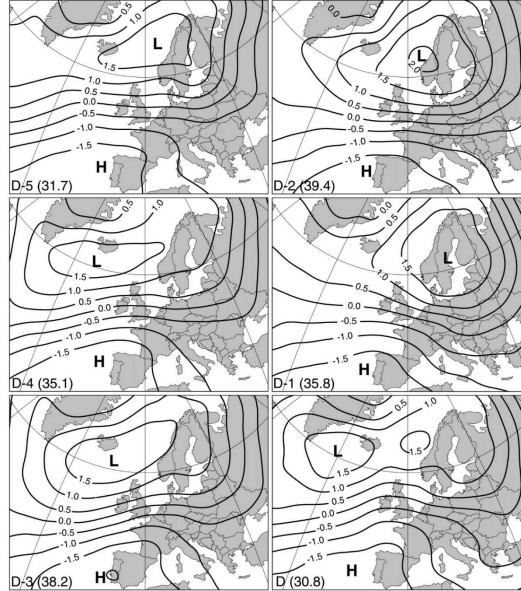
### SLP patterns of winter windstorms

- 98 cases of winter windstorms from the 20th century
- 121 grided values describing MSLP patterns during windstorms
- objective classification using principal components analysis

## Příklad využití PCA v klimatologii

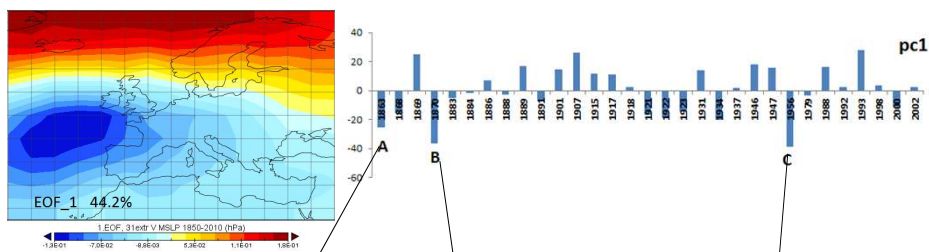
Jaký je charakter pole přízemního tlaku vzduchu v období 5 dní před výskytem povodní na Vltavě?

The first components of the SLP field on days D-5 to D calculated by the Principal Component Analysis (PCA) for 37 floods (1881-2000) of the winter synoptic type on the river Vltava in Prague (brackets - explained variance in %)

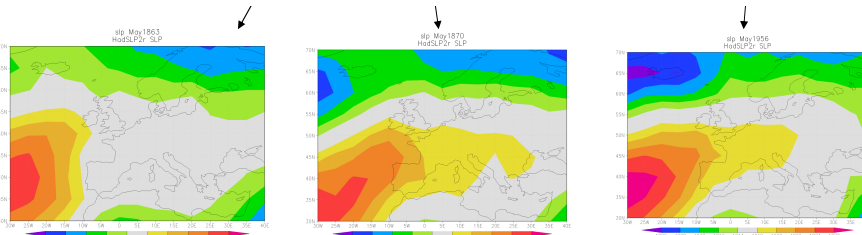


## Empirical Orthogonal Function (EOF)

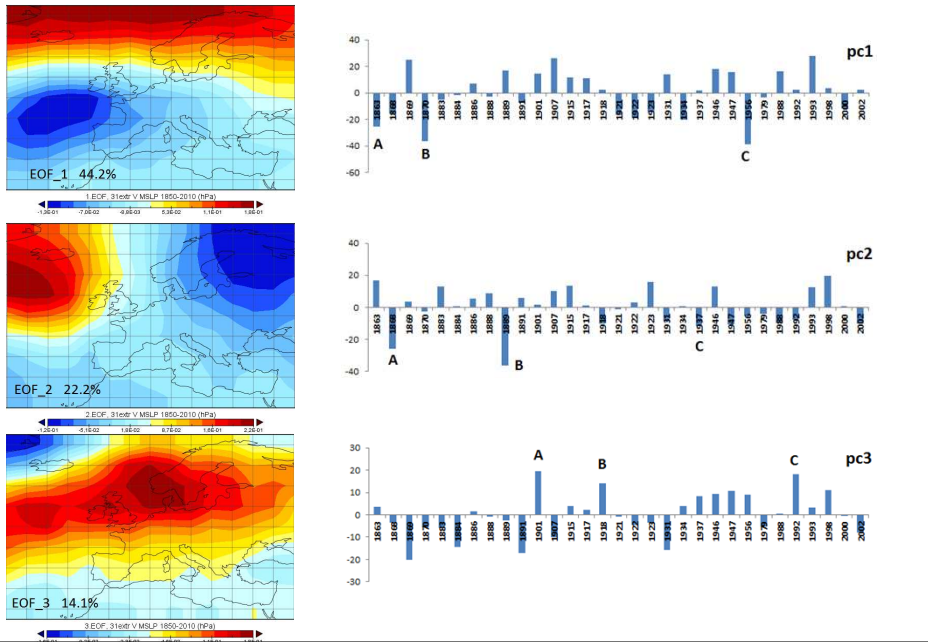
Z původní datové matice extrahuje **prostorovou složku** (EOF - mód prostorové variability) a **časovou složku** (PC - jak se EOF mění v čase).



Three the most similar monthly means according to pc1 used for interpretation of EOF\_1 mode.



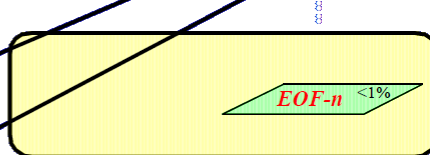
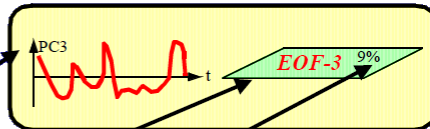
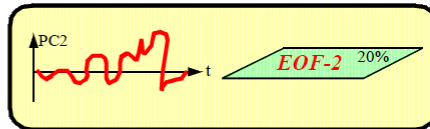
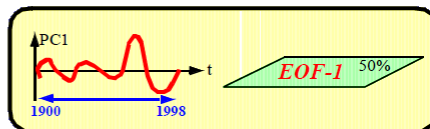
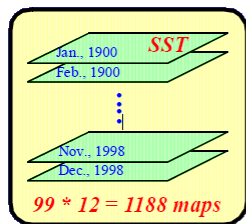
# Empirical Orthogonal Function



# Princip EOF

Původní datová matice (F) se rozkládá na časovou (PC) a prostorovou (EOF) složku.

$$F(x,y,t) = \sum_{k=1}^N PC(t) \cdot EOF(x,y)$$



hlavní komponenta

Principal Component

vlastní vektor

EOF (Eigen Vector)

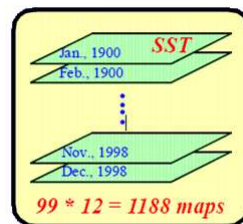
vlastní číslo

Eigen Value

## Princip EOF

- **Vstupní data:** spojité pole studovaného prvku („mapa“, např. tlak či teplota vzduchu) reprezentované množinou bodů (stanice či vrcholy pravidelné sítě) zaznamenané za určitý časový interval (časová řada)
- **EOF:** metoda jako PCA, která primárně slouží ke zmenšení rozměrnosti vstupních dat bez podstatné ztráty informace
- Poprvé prezentovaná Edwardem Lorenzem v roce 1956
- Hojně využívaná v klimatologii, oceánografii (v globálním či regionálním měřítku)
- **Cíl:** kompaktní popis časové a prostorové variability studovaného prvku pomocí tzv. ortogonálních (pravoúhlých = nekorelovaných) funkcí - tzv. statistických „módů“
- Podstatná část variability je uložena v několika prvních funkcích (módech), jejichž prostorové uspořádání i časové chování může být interpretováno fyzikálně zdůvodnitelnými jevy a procesy

## EOF - vstupní data



- Soubor  $N$  map pro časové okamžiky  $t = 1 \dots N$
- Každá mapa obsahuje měření nějaké proměnné  $\psi$  v lokalitách  $m = 1 \dots M$
- Máme  $M$  časových řad  $\psi_m(t)$ , každá má délku  $N$
- Předpokládáme, že  $N > M$  (počet časových kroků je větší než počet lokalit (gridových bodů))
- Pokud data obsahují roční chod, je nutné jej odstranit před vlastní analýzou - odečíst průměrný roční chod od hodnot původního pole  $\psi_m(t)$ .

## EOF - standardizace (normalizace) dat

Původní data  $\psi_m(t)$  upravíme tak, že od nich odečteme průměr a výsledek dělíme směrodatnou odchylkou:

$$F_m(t) = \frac{\psi_m(t) - \mu_m}{\sigma_m}$$

where  $\mu_m$  is the record mean:

$$\mu_m = \frac{1}{N} \sum_{t=1}^N \psi_m(t)$$

and  $\sigma_m$  is the record standard deviation:

$$\sigma_m = \left[ \frac{1}{N-1} \sum_{t=1}^N \psi_m^2(t) \right]^{1/2}$$

Tím zajistíme, že žádná z lokalit nedominuje svojí variabilitou a můžeme též kombinovat různé proměnné

## EOF - vstupní datová matice

Na počátku máme matici  $\mathbf{F}$  o rozměru  $M$  řádků (stanice, gridové body) a  $N$  sloupců (délka časové řady):

Time  $\longrightarrow$

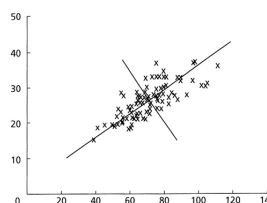
$$\mathbf{F} = \begin{bmatrix} F_1(1) & F_1(2) & \dots & F_1(N) \\ F_2(1) & F_2(2) & \dots & F_2(N) \\ \dots & \dots & \dots & \dots \\ F_M(1) & F_M(2) & \dots & F_M(N) \end{bmatrix} \quad \downarrow \text{Location}$$



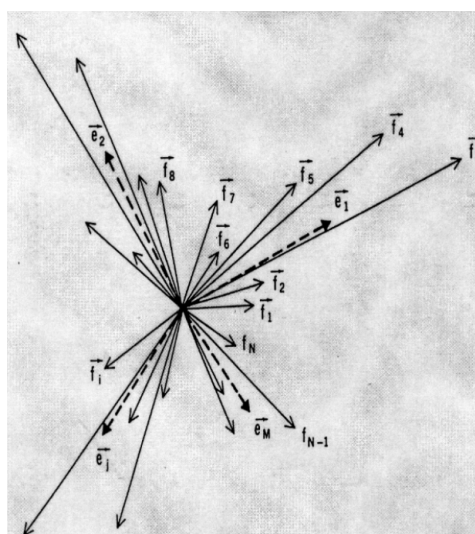
## Výpočet EOF

Dva základní přístupy, v obou případech se jedná o maticový počet (rozklad matice):

- Výpočet **matice kovariancí** a její rozklad na vlastní (charakteristická) čísla a vlastní (charakteristické) vektory
- **Singulární rozklad matice** (SVD - Singular Value Decomposition)
- **Vlastní vektory** definují směr nových faktorových os v prostoru původních proměnných a jsou navzájem ortogonální - tj. nezávislé - tedy každý nese unikátní informaci
- **Vlastní čísla** informují o variabilitě vyčerpané vytvářenými faktorovými osami. Představují rozptyl „nové“ proměnné definované v souřadném systému vlastních vektorů



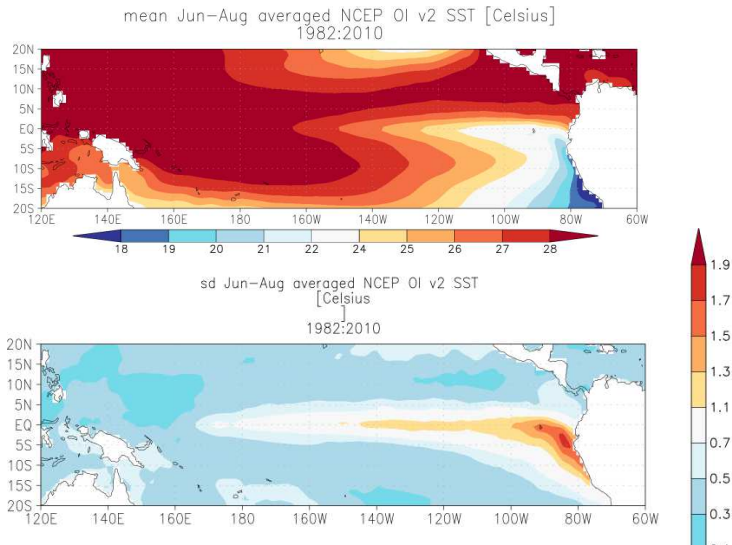
## Geometrická interpretace vlastních čísel a vlastních vektorů



Possible configuration of the data vectors  $f_n$  ( $n = 1 \dots N$  denote the time steps) and the empirical orthogonal vectors  $e_m$ ,  $m = 1 \dots M$ . (from Peixoto and Oort, 1992)

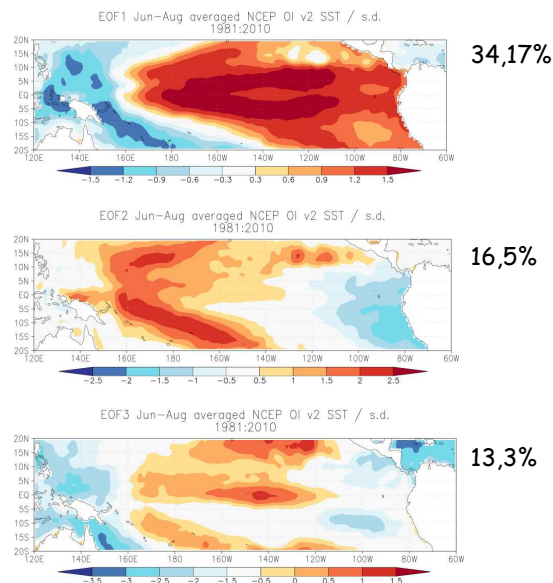
## Příklad analýzy EOF

Hlavní módy variability pole povrchové teploty oceánu (SST) v létě (JJA) v oblasti centrálního Pacifiku, vstupní data: [1982-now: 1° NOAA \("Reynolds"\) OI v2 SST](#)



## Příklad analýzy EOF

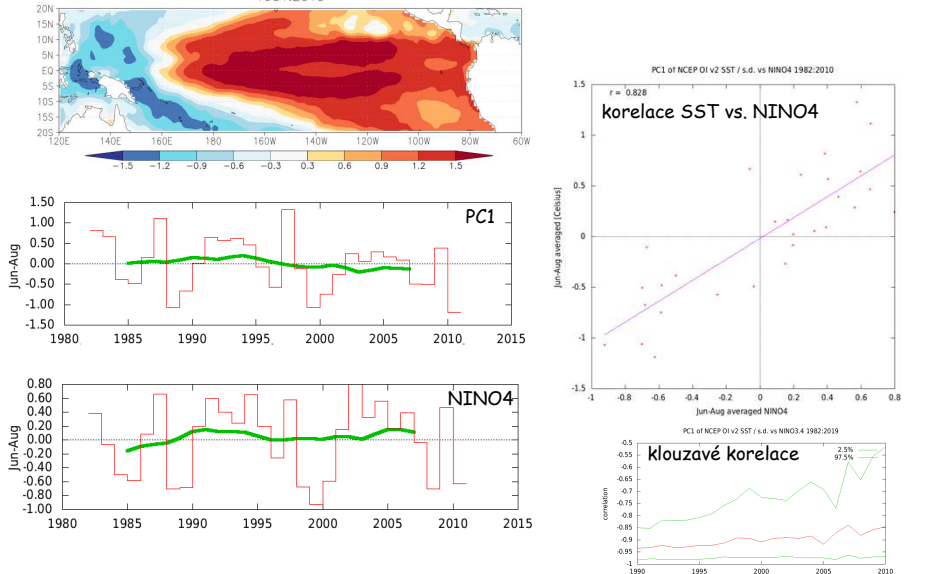
První tři hlavní módy variability pole povrchové teploty oceánu (SST)



## Příklad analýzy EOF

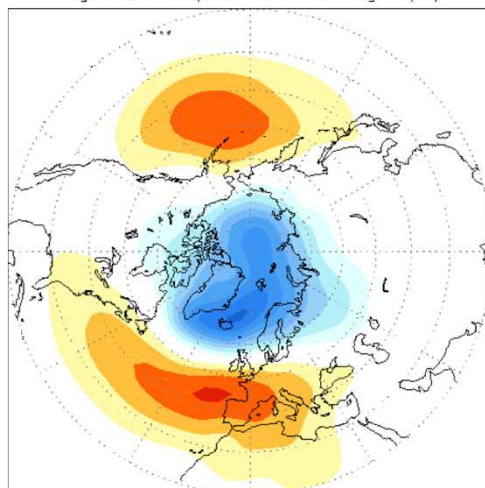
EOF1 (34.2 %), její časová složka (PC1) a korelace s NINO4

EOF1 Jun-Aug averaged NCEP OI v2 SST / s.d.  
1981:2010



## Příklad analýzy EOF

Leading EOF (19%) shown as  
regression map of 1000mb height (m)



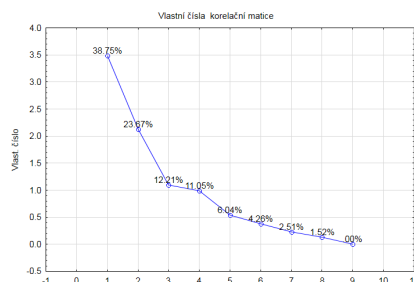
Map showing surface pressure anomaly pattern of the Arctic Oscillation in its positive phase, using the statistical technique of unrotated principal components analysis, which is another way to define typical preferred anomaly patterns. The units are height (meters) of the 1000mb pressure surface. Image credit: NOAA Climate Prediction Center.

<https://iri.columbia.edu/news/other-climate-patterns-that-impact-u-s-winter-climate/>

## Počet významných módů

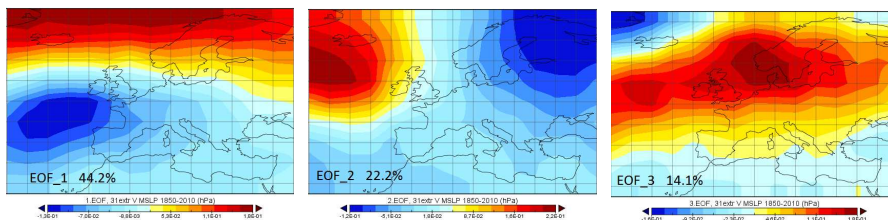
- Cíl redukce dat je důležitější než fyzikální interpretace módů
- Snažíme se především separovat signál od šumové složky
- Existuje několik metody určení počtu významných módů:
  - Tzv. sutinový graf (scree plot)
  - Vlastní čísla  $> 1$  (Guttmanovo kritérium)
  - Módy s vlastními čísly, která jsou větší než průměr vlastních čísel všech módů
  - Módy, které v součtu vysvětlují více než 70 - 90 % celkové variability

No.	Eigenvalue	Individual Cumulative		Scree Plot
		Percent	Percent	
1	3.487151	38.75	38.75	
2	2.130173	23.67	62.41	
3	1.098958	12.21	74.63	
4	0.994483	11.05	85.68	
5	0.543218	6.04	91.71	
6	0.383428	4.26	95.97	
7	0.225754	2.51	98.48	
8	0.136790	1.52	100.00	
9	0.000046	0.00	100.00	



## Rotované EOFs

- Fyzikální interpretace je důležitější než redukce (komprese) dat
- Některé módy nejsou interpretovatelné v důsledku požadavku na jejich vzájemnou ortogonalitu



- Vypočtené EOF módy se následně znovu pootočí
- Rotace nemusí být pravouhlá