

PROJECT 5 - PROBLEM 1.24 in the lecture notes (p.189)
(in groups of 2 or 3 students) (grade-weight: 20%)

Principal Component Analysis (PCA)

This is a small *research* project in which you have to formulate a *research question* and a *hypothesis*, based on *theory*, which tentatively answers the research question, and *test this hypothesis*, using a statistical technique called Principal Component Analysis (PCA). PCA is explained in **Box 1.13**. For inspiration, read sections 1.27-1.30 (PART F).

The following eleven theoretical concepts, or topics, come up for discussion.

1. Theory and hypothesis
2. Teleconnection involving the atmosphere, the ocean and the cryosphere
3. Northern and Southern Annular Modes (NAM & SAM)
4. El Niño – Southern Oscillation (ENSO)
5. Sudden Stratospheric Warming (SSW)
6. Quasi-Biennial Oscillation (QBO)
7. Standard error of correlation
8. Correlation matrix
9. Physical interpretation of the eigenvalues and eigenvectors of a correlation matrix
10. Principal Components or Empirical Orthogonal Functions
11. Projecting data on the first Principal Component

PROBLEM 1.24. Principal Component Analysis (PCA) (PROJECT 5, in groups of 2 or 3 students)

Read **sections 1.27 to 1.30** and **Box 1.13 (PCA-analysis)** for inspiration. Formulate an interesting hypothesis on the teleconnection (<https://www.ncdc.noaa.gov/teleconnections/>) of three or more variables. An example of a “teleconnection” between three remote variables (temperature at De Bilt in the Netherlands, temperature at Goose in Canada and the NAM-index) is discussed in **Box 1.13**. Other examples of variables, which are “teleconnected”, are the strength of the subtropical jet and the latitude of the surface westerlies (**figure 1.64**), the latitude of the subtropical jet and the latitude of the baroclinic zone in the troposphere (**figure 1.68**), the meridional isentropic mass flux at 50°N in the Middleworld and the maximum baroclinicity at 30°N in the Middleworld (in January) (**figure 1.70**), the meridional isentropic mass flux at 50°N in the Middleworld and the NAM-index (**figure 1.74**), sea surface temperature in the equatorial east Pacific and the strength of the Trade winds in the central equatorial Pacific (**figure 1.78**), Polar cap mean pressure at 350 K and polar cap mean potential vorticity at 600 K (**figure 1.84**) and, finally, the zonal mean temperature, zonal mean zonal velocity and zonal mean vertical velocity at the equator (**figure 1.88**). These teleconnections are all manifestations of large-scale modes of variability, such as the Northern Annular Mode (NAM) (**section 1.27**), El Niño-Southern Oscillation (ENSO) (**section 1.28**), Sudden Stratospheric Warmings (SSW's) (**section 1.29**) and the Quasi-Biennial Oscillation (QBO) (**section 1.30**).

Data can be found on the internet, e.g. at <https://www.ncdc.noaa.gov> (a very complete American source of climate data), <https://cds.climate.copernicus.eu/cdsapp#!/search?type=dataset> (ERA-5 reanalysis data, among other), https://psl.noaa.gov/data/gridded/data.20thC_ReanV3.html (NOAA/CIRES/DOE 20th Century Reanalysis (V3)), <http://apps.ecmwf.int/datasets/> (ERA-Interim reanalysis data), <http://climexp.knmi.nl> (site giving links to many data-sites), <http://eca.knmi.nl> (Daily observations at many stations in especially Europe), <http://www.knmi.nl/klimatologie/> (Dutch weather observations), <https://projects.knmi.nl/klimatologie/uurgegevens/selectie.cgi> (hourly observation, the Netherlands) and <https://www.knmi.nl/nederland-nu/klimatologie/uurgegevens/Noordzee> (hourly observations over the North sea).

The structure of **project 5** is as follows.

(a) First formulate and explain your **hypothesis** and hand in the hypothesis plus the explanation on or before **a prefixed date**. You must also indicate what data you are going to use to verify the hypothesis (step b).

(b) Collect the necessary **data** (time series).

(c) Perform the **principal component analysis of the time series** (explained in **Box 1.13**).

(d) **Present** your results orally (in **12 minutes** in January 2021), in which you show or discuss the following topics.

(e) The **cross-correlation matrix**, and the associated **eigenvectors** and **eigenvalues**.

(f) The physical interpretation of the eigenvector(s) with the largest eigenvalue(s), i.e the **Principle components** .

(g) **Discuss** the result in the light of the hypothesis.

(h) **Hand in** the presentation as a **pdf-file**.