Lecture 24
Computer Experiments & Principal Component Analysis

STAT 8020 Statistical Methods II
November 19, 2020

Whitney Huang
Clemson University

Agenda

1. Computer Experiments
2. Multivariate Analysis
3. Principal component analysis (PCA)

What is a Computer Experiment

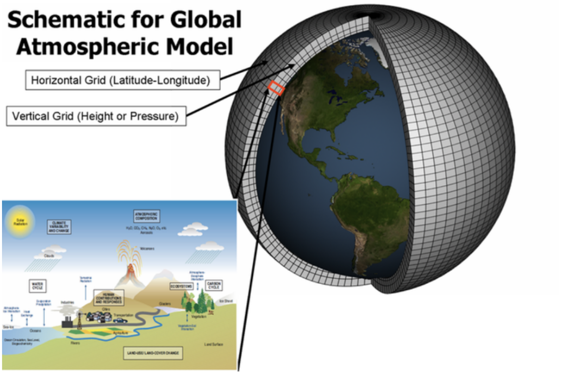
In some situations it is economically, ethically, or simply not possible to run a **physical experiment**. Instead, the following scenario might be feasible:

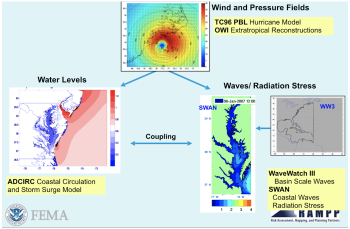
- the physical process can be described by a mathematical model (e.g., a system of differential equations)
- computer code (simulator) can be written to compute the response from the mathematical model

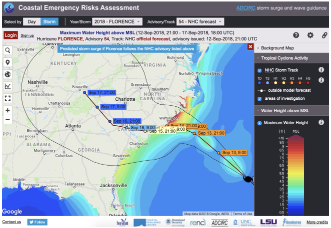
$$\boxed{\begin{array}{c}\text{Input}\\ \boldsymbol{x} \in \mathcal{X}\end{array}} \longrightarrow \boxed{\begin{array}{c}\text{Model}\\ f : \mathcal{X} \mapsto \mathcal{Y}\end{array}} \longrightarrow \boxed{\begin{array}{c}\text{Output}\\ y = f(\boldsymbol{x})\end{array}}$$

In this case, a researcher can conduct a **computer experiment** by running the computer code, which serves as a proxy for the physical process, to compute a "response" at any combination of values of the inputs

Examples of Computer Models







Computer Experiments vs. Physical Experiments

- *“Experimental results are believed by everyone, except for the person who ran the experiment”*
- *“Computational results are believed by no one, except the person who wrote the code”*

Computer Experiments vs. Physical Experiments

- *"Experimental results are believed by everyone, except for the person who ran the experiment"*
- *"Computational results are believed by no one, except the person who wrote the code"*

Replication, randomization and blocking are irreverent for a computer experiment because many **computer codes are deterministic** and **all the inputs to the code are known and can be controlled**

Design & Analysis of Computer Experiments

- **Design**:
  where to make the runs, i.e., the selection of inputs $\{\boldsymbol{x}_i\}_{i=1}^n$ where $\boldsymbol{x}_i = (x_{1,i}, x_{2,i}, \cdots x_{d,i})$

- **Analysis**:
  fit a statistical model using the model inputs-output $\{y_i, \boldsymbol{x}_i\}_{i=1}^n$ to "emulate" the simulator and to quantify the prediction uncertainty for $y(\boldsymbol{x}_{\text{new}})$, usually via a Gaussian Process Model $\mathrm{GP}\left(m\left(\cdot\right), K\left(\cdot,\cdot\right)\right)$, where

  - $m(\boldsymbol{x}) = \mathrm{E}[y(x)]$ is the mean function
  - $K(\boldsymbol{x}, \boldsymbol{x}') = \mathrm{Cov}(y(\boldsymbol{x}), y(\boldsymbol{x}'))$ is the covariance function

An Overview of Multivariate Analysis

- In many studies, observations are collected on several variables on each experimental/observational unit
- Multivariate analysis is a collection of statistical methods for analyzing these multivariate data sets
- **Common Objectives**
  - Dimensionality reduction
  - Classification
  - Grouping (Clustering)

Multivariate Data

We display a multivariate data that contains n units on p variables using a matrix

$$\boldsymbol{X} = \begin{pmatrix} X_{1,1} & X_{2,1} & \cdots & X_{p,1} \\ X_{1,2} & X_{2,2} & \cdots & X_{p,2} \\ \vdots & \cdots & \ddots & \vdots \\ X_{1,n} & X_{2,n} & \cdots & X_{p,n} \end{pmatrix}$$

Summary Statistics

- Mean Vector: $\bar{\boldsymbol{X}} = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_p)^T$
- Covariance Matrix: $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{p}$, where $\sigma_{ii} = \mathrm{Var}(X_i), \quad i = 1, \cdots, p$ and $\sigma_{ij} = \mathrm{Cov}(X_i, X_j), i \neq j$

Multivariate Data

We display a multivariate data that contains n units on p variables using a matrix

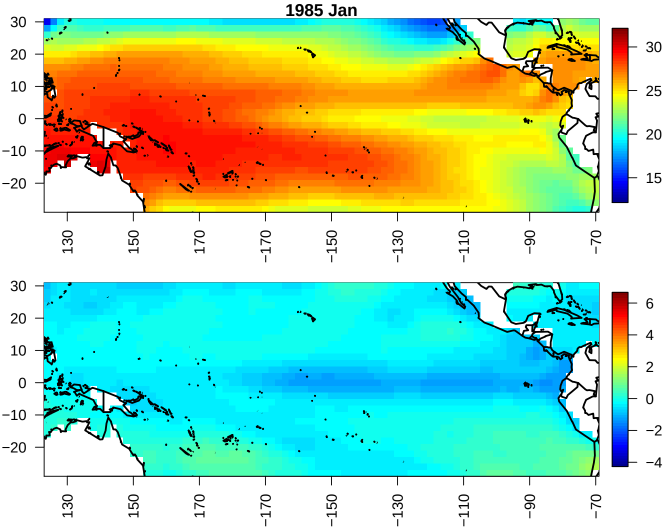
$$\boldsymbol{X} = \begin{pmatrix} X_{1,1} & X_{2,1} & \cdots & X_{p,1} \\ X_{1,2} & X_{2,2} & \cdots & X_{p,2} \\ \vdots & \cdots & \ddots & \vdots \\ X_{1,n} & X_{2,n} & \cdots & X_{p,n} \end{pmatrix}$$

Summary Statistics

- Mean Vector: $\bar{\boldsymbol{X}} = (\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_p)^T$
- Covariance Matrix: $\Sigma = \{\sigma_{ij}\}_{i,j=1}^{p}$, where $\sigma_{ii} = \mathrm{Var}(X_i), \quad i = 1, \cdots, p$ and $\sigma_{ij} = \mathrm{Cov}(X_i, X_j), i \neq j$

Next, we are going to introduce **Principal Component Analysis (PCA)**, a useful tool for conducting dimension reduction

Example: Monthly Sea Surface Temperatures



Sea Surface Temperatures and Anomalies

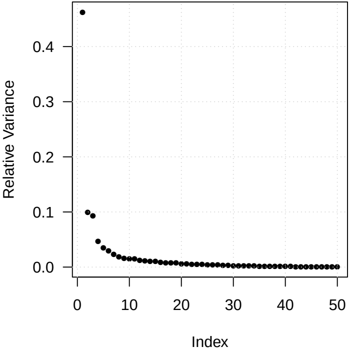
- The “data” are gridded at a $2°$ by $2°$ resolution from $124°E – 70°W$ and $30°S – 30°N$. The dimension of this SST data set is 2303 (number of grid points in space) $\times$ 552 (monthly time series from 1970 Jan. to 2015 Dec.)
- Sea-surface temperature anomalies are the temperature differences from the climatology (i.e. long-term monthly mean temperatures)
- We will demonstrate the use of Empirical Orthogonal Function (EOF) analysis to uncover the low-dimensional structure of this spatio-temporal data set

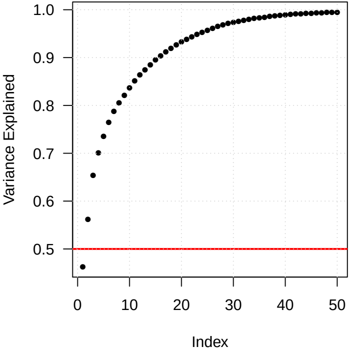
The Emipirical Orthogonal Function (EOF) Decomposition

Empirical orthogonal functions (EOFs) are the geophysicist's terminology for the eigenvectors in the eigen-decomposition of an empirical covariance matrix. In its discrete formulation, EOF analysis is simply Principal Component Analysis (PCA). EOFs are usually used

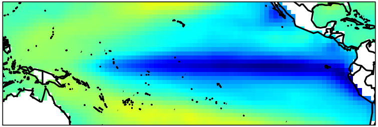
- To find principal spatial structures
- To reduce the dimension (spatially or temporally) in large spatio-temporal datasets

Screen Plot for EOFs

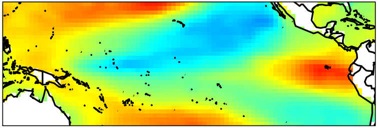




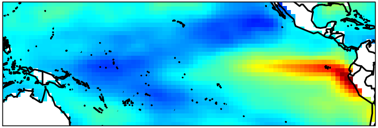
Perform EOF Decomposition and Plot the First Three Modes



EOF1: The classic ENSO pattern

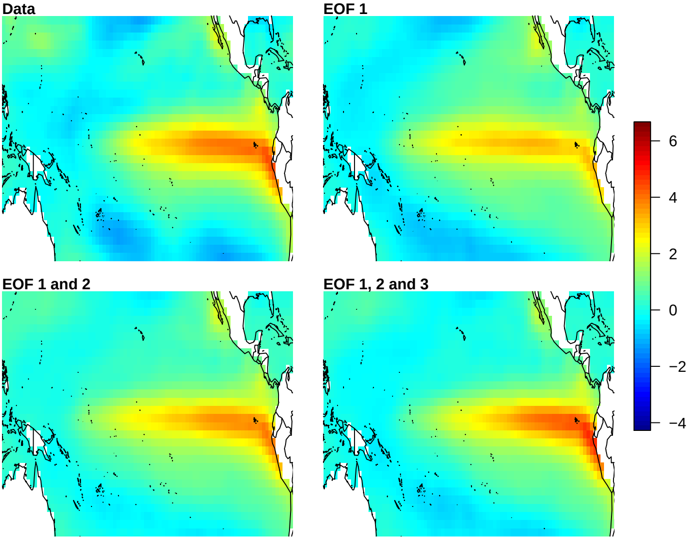


EOF2: A modulation of the center



EOF3: Messing with the coast of SA and the Northern Pacific.

1998 Jan El Niño Event



Principal Component Analysis

Given a random sample from a p-dimensional random vector $\boldsymbol{X}_i = \{X_{1,i}, X_{2,i}, \cdots, X_{p,i}\}, \quad i = 1, \cdots, n$

- Dimension reduction technique
  - Large number of variables (p)
  - Number of variables (p) may be greater than number of observations (n)
- Create new, uncorrelated variables (principal components) for the follow up analysis
  - Principal Component Regression
  - Interpretation of principal components can be difficult in some situations

Finding Principal Components

Principal Components (PC) are uncorrelated **linear combinations** $\tilde{X}_1, \tilde{X}_2, \cdots, \tilde{X}_p$ determined sequentially, as follows:

1. The first PC is the linear combination $\tilde{X}_1 = \boldsymbol{c}_1^T \boldsymbol{X} = \sum_{i=1}^{p} c_{1i} X_i$ that maximize $\mathrm{Var}(\tilde{X}_1)$ subject to $\boldsymbol{c}_1^T \boldsymbol{c}_1 = 1$
2. The second PC is the linear combination $\tilde{X}_2 = \boldsymbol{c}_2^T \boldsymbol{X} = \sum_{i=1}^{p} c_{2i} X_i$ that maximize $\mathrm{Var}(\tilde{X}_2)$ subject to $\boldsymbol{c}_2^T \boldsymbol{c}_2 = 1$ and $\boldsymbol{c}_2^T \boldsymbol{c}_1 = 0$

$$\vdots$$

3. The j_{th} PC is the linear combination $\tilde{X}_j = \boldsymbol{c}_j^T \boldsymbol{X} = \sum_{i=1}^{p} c_{ji} X_i$ that maximize $\mathrm{Var}(\tilde{X}_j)$ subject to $\boldsymbol{c}_j^T \boldsymbol{c}_j = 1$ and $\boldsymbol{c}_j^T \boldsymbol{c}_k = 0 \forall k < j$

Principal Components

- Let Σ, the covariance matrix of $\boldsymbol{X}$, have eigenvalue-eigenvector pairs $(\lambda_i, \boldsymbol{e}_i)_{i=1}^{p}$ with with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ Then, the k_{th} principal component is given by

$$\tilde{X}_k = \boldsymbol{e}_k^T \boldsymbol{X} = e_{k1}X_1 + e_{k2}X_2 + \cdots e_{kp}X_p$$

- Then,

$$\mathrm{Var}(\tilde{X}_i) = \lambda_i, \quad i = 1, \cdots, p$$

$$\mathrm{Cov}(\tilde{X}_j, \tilde{X}_k) = 0, \quad \forall j \neq k$$

PCA and Proportion of Variance Explained

- It can be shown that

$$\sum_{i=1}^{p} \mathrm{Var}(\tilde{X}_i) = \lambda_1 + \lambda_2 + \cdots + \lambda_p = \sum_{i=1}^{p} \mathrm{Var}(X_i)$$

- The proportion of the total variance associated with the k_{th} principal component is given by

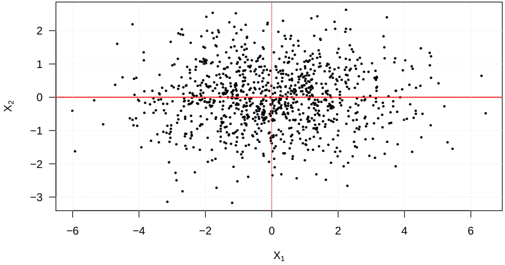
$$\frac{\lambda_k}{\lambda_1 + \lambda_2 + \cdots + \lambda_p}$$

- If a large proportion of the total population variance (say 80% or 90%) is explained by the first k PCs, then we can restrict attention to the first k PCs without much loss of information

Toy Example 1

Suppose we have $\boldsymbol{X} = (X_1, X_2)^T$ where $X_1 \sim \mathrm{N}(0, 4)$, $X_2 \sim \mathrm{N}(0, 1)$ are independent

- Total variation $= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = 5$
- X_1 axis explains 80% of total variation
- X_2 axis explains the remaining 20% of total variation



Toy Example 2

Suppose we have $\boldsymbol{X} = (X_1, X_2)^T$ where $X_1 \sim \mathrm{N}(0,4)$, $X_2 \sim \mathrm{N}(0,1)$ and $\mathrm{Cor}(X_1, X_2) = 0.8$

- Total variation
  $= \mathrm{Var}(X_1) + \mathrm{Var}(X_2) = \mathrm{Var}(\tilde{X}_1) + \mathrm{Var}(\tilde{X}_2) = 5$
- $\tilde{X}_1 = .9175X_1 + .3975X_2$ explains 93.9% of total variation
- $\tilde{X}_2 = .3975X_1 - .9176X_2$ explains the remaining 6.1% of total variation