



# LOW DIMENSIONAL EMBEDDING OF ENVIRONMENTAL VARIABLES

**EA MAP581**

6 mars 2018

---

Flore Martin and Lorraine Roulier



## Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Principal Component Analysis - PCA</b>	<b>3</b>
2.1	Method . . . . .	3
2.2	Results . . . . .	4
<b>3</b>	<b>Kernel Principal Component Analysis</b>	<b>4</b>
3.1	Method . . . . .	4
3.2	Results . . . . .	4
<b>4</b>	<b>Multidimensional Scaling - MDS</b>	<b>4</b>
4.1	Method . . . . .	4
4.2	Results . . . . .	4
<b>5</b>	<b>Isomap</b>	<b>4</b>
5.1	Method . . . . .	4
5.2	Results . . . . .	4
<b>6</b>	<b>Comparing the different methods</b>	<b>4</b>
<b>7</b>	<b>Conclusion</b>	<b>4</b>
<b>8</b>	<b>Bibliography</b>	<b>4</b>

# 1 Introduction

Climate data amounts very quickly to a lot of unused data. In a day, we can collect temperature, pressure, wind data all over the world with satellites, even hourly. Our project was two sided. First, we familiarized with various dimension reduction techniques, then we attempted to show that the geographical position of a point on the planet - e.g. it's latitude and longitude - were embedded in the climate data one could gather on it.

Dimension reduction techniques can be divided in two classes, linear dimension reduction and non linear dimension reduction. However, in all methods, the main goal is to figure out a similarity function between vectors. Such a function will then enable to sort the dataset into classes of vectors with similar features, which would have been more intricate with the initial dataset. We used a set of datasets we found on the NASA website, that gathered various means on climate variables over 22 years at every given latitude and longitude. These variables are gathered in the table below

Temperature °C	Pressure kPa	Relative Humidity %	Wind Speed m/s	Radiation kWh/m <sup>2</sup> /day
-------------------	-----------------	------------------------	-------------------	--------------------------------------

FIGURE 1 – First lign of our dataset

## 2 Principal Component Analysis - PCA

### 2.1 Method

Principal Component Analysis detects tendencies in the data by maximizing the variance of the dataset matrix. This yields an orthonormal matrix that can be diagonalized. The largest eigenvalues point to the eigenvectors that contain the most information about the dataset.

Let  $X \in \mathbb{R}^{d \times n}$  be our dataset, PCA maximizes the following equation :

$$\|X - MM^T X\|^2,$$

subject to  $M \in \mathcal{O}^{d \times r}$  where  $r < d$

## 2.2 Results

# 3 Kernel Principal Component Analysis

## 3.1 Method

## 3.2 Results

# 4 Multidimensional Scaling - MDS

## 4.1 Method

## 4.2 Results

# 5 Isomap

## 5.1 Method

## 5.2 Results

# 6 Comparing the different methods

# 7 Conclusion

# 8 Bibliography