

# KPCA over PCA to assess urban resilience to floods

Narjiss SATOUR<sup>1\*</sup>, Badreddine BENYACOUB<sup>2</sup>, Badr EL MAHRAD<sup>3,4</sup>, Ilias KACIMI<sup>1</sup>

<sup>1</sup>Geosciences, Water and Environment Laboratory Mohammed V University in Rabat, Morocco.

<sup>2</sup>National Institute of Statistics and Applied Economics, Rabat, Morocco.

<sup>3</sup>Murray Foundation, Brabners LLP, Horton House, Exchange Street, Liverpool L2 3YL, UK

<sup>4</sup>CIMA, FCT-Gambelas Campus, University of Algarve, 8005-139 Faro, Portugal

**Abstract.** Global increases in the occurrence and frequency of flood have highlighted the need for resilience approaches to deal with future floods. The principal component analysis (PCA) has been used widely to understand the resilience of the urban system to floods. Based on feature extraction and dimensionality reduction, the PCA reduces datasets to representations consisting of principal components. Kernel PCA (KPCA) is the nonlinear form of PCA, which efficiently presents a complicated data in a lower dimensional space. In this work the KPCA techniques was applied to measure and map flood resilience across a local level. Therefore, it aims to improve the performance achieved by non-linear PCA application, compared to standard PCA. Twenty-one resilience indicators were gathered, including social, economic, physical, and natural components into a composite index (Flood Resilience Index). The experimental results demonstrate the KPCA performance to get a better Flood Resilience Index, guiding q decision making to strengthen the flood resilience in our case of study of M'diq-Fnideq and martil municipalities in Northern of Morocco.

**Key words:** Floods, Resilience, KPCA, PCA, Morocco

## 1 Introduction

Achieving local and regional development goals has become a significant challenge for governments and communities worldwide [1]. As referred to the Intergovernmental Panel on Climate Change (IPCC.2012), the risks associated to global warming are going to increase during the future [2], [3]. These challenges require a concentration on strategies rising practitioners and researcher's interest in investigating how to improve urban resilience [4].

Urban resilience enhancement has gained a broad interest, as an ultimate goal of adaptation plans. In fact, several works have been dedicated to urban resilience measurement process [5], [6], [7], where composite indicators were often used as resilience metrics. Many indicators have been used to assess urban resilience level to a particular hazard at a local scale [8]. The Flood Resilience Index (FRI) is a specific index used to assess urban resilience to floods [9]. Nevertheless, conclusions have highlighted that most of challenges in resilience assessment to data availability and quality constraint. Most of the time, these data were uncertain and heterogeneous.

In fact, computing Flood Resilience Index means mapping some non-linear behaviour among different parameters (natural, social, physical, economical, institutional). This could be achieved using clustering algorithms. The classical principal component analysis (PCA) method was previously used to assess FRI [10].

The main purpose of PCA is the analysis of data to identify patterns that reflect the data. PCA is extensively applied to reduce the number of indicators highly correlated in a multivariate dataset to a smaller set of intermediate indicators [11]. In other words, PCA aims to find the axes with maximum variances along which the data is most spread. However, PCA suffers from many issues and fails to perform well in nonlinear problematics compared to other clustering technique [12]. Kernel PCA (KPCA) is able to be more successful than conventional PCA, capturing the nonlinear correlations among data [13]. Fortunately, kernel PCA allows to standardise PCA to nonlinear dimensionality reduction [14].

Taking into account the mentioned scope, in this paper, kernel PCA method was proposed to identify a better Flood Resilience Index for local scale. FRI was evaluated in three coastal municipalities in Northern Morocco (Fig.1).

To achieve a comparative analysis of alternative nonlinear method on constructing composite indicators to measure the FRI, this paper is organized as follows: Section 2 exposes a brief presentation of the study area, and data used in this work. Section 3 is devoted to the methodology applied to assess FRI using PCA, KPCA and Geographic Information Systems (GIS). Section 4 presents the results obtained, and draws the discussion to examine, and compare the spatial distribution of the Flood Resilience Index developed using PCA linear and Nonlinear PCA method (KPCA), with its own three varieties.

\* Corresponding author: [narjiss.satour@gmail.com](mailto:narjiss.satour@gmail.com)

## 2 Study area and data collection

### 2.1 Fnideq, M'diq and Martil municipalities

Over a length of 44 Km extending on the coastal edge of Tangier-Tetouan metropolitan area in Morocco, Fnideq, M'diq and Martil (FMM) municipalities are located and localized downstream three watersheds: Fnideq, Smir and Martil-Alila (Fig.1). The whole area is ranked as the one of the northern areas at risk from floods. While the frequency of floods events increased gradually over time since 1980 [12]. Few studies have operationalized an urban resilience to floods [9]. Meanwhile, historical records show frequent flood concurrency from 2000 until 2010 [15], and from 2000 to 2021 [16].

### 2.2 Data Sources and Collection

Variables section guide was used to measure resilience, is based on the principles of resilience outlined by [15] (factors shown to be linked to social, economic, infrastructural and ecological flood resilience

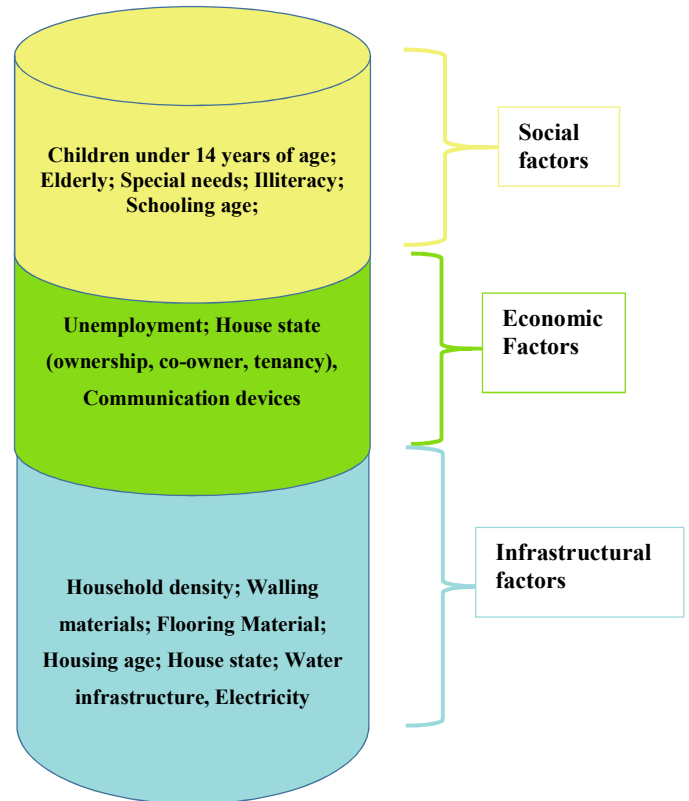


**Figure 1.** Study area delimitation

[16]).

Data were collected at the local level from the last 2014 Census of Morocco (RGPH, 2014) as well as government publications, municipal planning documents and an online RGPH 2014 database (Fig.2).

Available data were used to describe the continuous variables for the urban resilience assessment. Calculation was performed using MATLAB, while Visualizations were done using Free and Open GIS Source tools.



**Figure 2.** Flood Resilience variables used to assess FRI

## 3 Methodology: Flood Resilience Assessment

### 3.1 Linear Principal Component Analysis (PCA)

PCA, is a popular technique for dimensionality reduction and extracting structure in a higher dimensional space. The complexity of data in the original feature space can be efficiently represented in linear subspace of lower dimensionality using PCA, where the new features have the largest variance.

The principal components can be understood as new axes of the dataset that maximize the variance along those axes (the eigenvectors of the covariance matrix). Consider a dataset  $\{x_i\}$  where  $i = 1, 2, \dots, N$ , and each  $x_i$  is a  $p$ -dimensional vector. PCA method aims to project the data into a  $q$ -dimensional subspace, where  $q < p$ . Let  $S_x$  be the covariance matrix of  $\{x_i\}$ . PCA is usually achieved by solving an Eigenvalue problem. PCA diagonalizes the covariance matrix

$$S_x = \frac{1}{N} \sum_i^N (x_i - \bar{x})(x_i - \bar{x})^T \quad (1)$$

Where  $\bar{x} = \frac{1}{N} \sum_i^N x_i$ .

Thus, the Eigenvalue equation needs to be solved:

$\lambda v = S_x v$ , for Eigenvalue  $\lambda \geq 0$  and  $v \in R^p$ . Each Eigenvectors calculated  $v$  allows to determine the extracted principal components that maximize the variance of  $S_x$ .

### 3.2 Kernel Principal Component Analysis (KPCA)

#### 3.2.1 Nonlinear dimensionality reduction

The PCA are a linear combination of the extracted variables. In such cases, the data is linearly inseparable and linear combination does not satisfy the basic assumptions. For multivariate data analysis technique, a nonlinear is required if the task is to reduce the dimensionality of a dataset.

The data needs to be projected into a higher dimensional space to deal with the linearly inseparable of the coordinate system in which we describe our data. The basic idea is to transform the data to a space where it becomes linearly separable.

Assuming a nonlinear transformation  $\phi$  called mapping function, from the original  $p$ -dimensional feature space to an  $d$ -dimensional feature space, where usually  $d > p$ . Each data point  $x_i$  is projected to a point  $\phi(x_i)$  using this mapping function which can be written as  $x_i \rightarrow \phi(x_i)$ .

First, we make the assumption that the projected data have zero mean:  $\frac{1}{N} \sum_i^N \phi(x_i) = 0$ .

The covariance matrix of the projected points is  $d \times d$ , calculated by:  $\bar{C} = \frac{1}{N} \sum_i^N \phi(x_i) \phi(x_i)^T$ .

We now have to find its eigenvalues  $\lambda \geq 0$  and eigenvectors  $v$  satisfying:  $\lambda v = \bar{C} v$ .

By the same argument as above, it is possible to perform standard PCA in the new feature space. Thus, we can extract the nonlinear principal components corresponding to  $\phi$ , but this can be extremely costly and inefficient. Fortunately, we can use kernel methods to simplify the computation.

#### 3.2.2 Kernel functions

The nonlinear method use the kernel functions map the data to an often higher dimensional space. This approach need to compute dot products mapped by  $\phi$ , with a possibly high computational cost. In order to implement the kernel PCA we just need to compute dot products of the form  $\phi(x_i)^T \phi(x_j)$  using kernel representations of the form

$$k(x_i, x_j) = (\phi(x_i), \phi(x_j)) = \phi(x_i)^T \phi(x_j) \quad (2)$$

Which allow to compute the value of the dot product without having to perform the mapping function  $\phi$ . Then, the implementation of kernel PCA dimensionality reduction is based on the computation of the dot product matrix  $K$  where  $K_{ij} = (k(x_i, x_j))_{ij}$ .  $K$  is symmetric positive semi-definite. In the next, we solve the Eigenvalue equation by diagonalizing the kernel matrix  $K$ . Three commonly kernel functions are used as the kernel trick.

*Linear*

$$K(x_1, x_2) = x_1^T x_2 \quad (3)$$

*Polynomial*

$$K(x_1, x_2) = (x_1^T x_2 + 1)^p \quad (4)$$

$p$  is the order of the polynomial.

*Gaussian or Radial Basis Function (RBF)*

$$K(x_1, x_2) = \exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \quad (5)$$

$\sigma$  is the width of the kernel.

In practice the covariance matrix based on mapping function is not calculated explicitly. We can directly construct the kernel matrix using an a priori chosen kernel function  $k(x_i, x_j)$  for all occurrences of  $(\phi(x_i), \phi(x_j))$ .

### 3.3 Calculating the flood resilience index FRI using PCA and KPCA

Indicators has been mostly used as a useful tool for policymaking and public communication [5]. Nevertheless, its use to measure resilience is relatively new [19], less developed in developing countries. Among different weighting techniques used to assess resilience to floods this section presents the application of the PCA and KPCA techniques to construct the Flood Resilience Index (Flood Resilience Index). We describe also, the process to calculate the index FRI based on the extracted principal components.

To measure the index FRI, the authors propose employing the extracted principal components to construct composite indices. [11] suggest using an intermediate resilience indicator  $IRI_j$  that correspond to each principal component to FRI.

In contrast to the standard PCA, the extraction of a number of kernel principal components can exceed the input dimensionality  $p$  and depend to the chosen kernel function. In practice, the proposed method does not yield the principal component axes, but the obtained eigenvectors can be understood as projections of the data onto the principal components.

Unlike linear PCA method as proposed by [11], those eigenvectors already are the data points projected and can be employed directly to calculate FRI. In our case, the intermediate resilience indicators are given by eigenvectors of the centered kernel matrix that correspond to the largest eigenvalues. Taking into account these assumptions, 4 eigenvectors are retained to present the intermediate indicators based on the three standard kernel functions. Then, the index composite FRI can be calculated as a weighted aggregation of the intermediate resilience indicators:

$$FRI_i = \sum_{k=1}^{k=4} \alpha_k IRI_{ki}$$

Where  $FRI_i$  is the value of the composite indicator for the ward  $i$  and  $\alpha_j$  is the weight applied to the intermediate resilience indicator  $j$ . These weights are calculated as follows:

$$\alpha_j = \frac{eigvalue_j}{\sum_{j=1}^{j=6} eigvalue_j}$$

The result corresponding to the index scores shows that obtained values can be negative or positive. The normalization using min-max is used to standardize the index scores.

### 3.4 FRI external validation

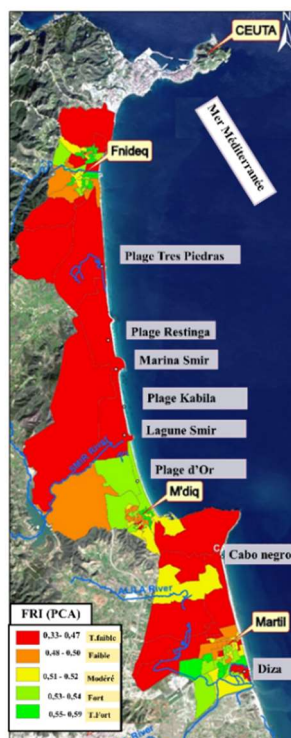
To identify a suitable methodology for a useful Flood Resilience Index development, this section undertakes the validation step. Based on risk and vulnerability works [20,21,22], phenomenon strongly related to Flood Resilience, and the last extreme weather events registered in March 2021, an external validation was performed to examine PCA or KPCA method's accuracies. The best index to recommend is supporting the purpose announcing that areas with higher vulnerability levels examined have lower resilience levels [7, 24].

## 4 Experimental results and analysis

In this section FRI results were presented assessed through two methods: PCA and the three types of KPCA (Linear; Polynomial and Gaussian). GIS were used to map the spatial distribution of FRI assessed using PCA and KPCA.

### 4.1. Mapping the FRI scores with PCA

Explaining 75% of the total variance PCA provide six principal components. Only four principal components were retained. These components correspond to the eigenvalues of PCA. The variance accounting for each of these six components are respectively 28.11, 17.13, 12.06, 7.75, 5.46 and 5.05%. Fig.3 shows how a low FRI (red color) has gained a large flood spatial distribution.



**Figure 3.** PCA application to assess Flood Resilience Index

To better understand the comparative step, and the spatial distribution of FRI for 126 urban sectors, the

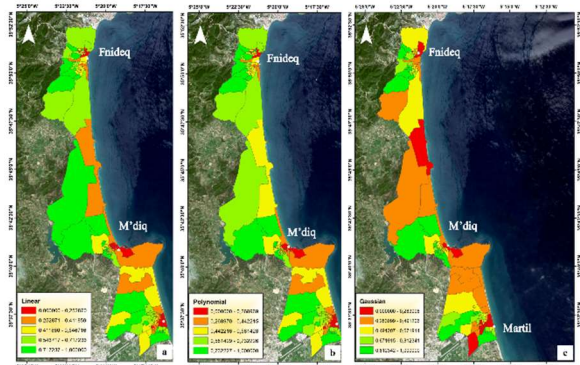
standard deviation from mean (Z-scores) were calculated to classify the sectorial resilience levels in each municipality under 5 classes:

- The sectors with a score greater than 0.55 are considered as high resilience and visualized as dark green.
- The sectors with a score between 0.53 and 0.54 are classified as relatively high resilience (light green).
- Between 0.51 and 0.52 as moderate resilience (yellow).
- Between 0.48 and 0.50 as relatively low resilience (orange).
- The sectors with a score less than 0.47 are classified as low resilience (red) (Figure.3 & 4).

### 4.2. Mapping the FRI scores with KPCA

The Kernel Principal Component Analysis (KPCA) has not been widely used in urban resilience assessment comparing to conventional PCA. KPCA was applied to feature extraction of flood urban resilience.

Thus, the three commonly kernel functions were used as the kernel trick. The experimental results shown that the calculated KPCA with Gaussian or Radial Basis Function (Fig4. c) had better performance than the linear function (Fig4. a) or the polynomial one (Fig4.b): the linear function and the polynomial, express a high resilience (Dark green) in highly vulnerable sectors having testified a huge damage during the last floods event in March 2021 (e.g South-west of Martil municipality)



**Figure 4.** KPCA applications to assess FRI: a). Linear function; b) Polynomial function, and c) Gaussian function

### 4.3. PCA and KPCA comparison

Taking into account the results of the PCA and KPCA (linear, Gaussian, polynomial), a first statistical analysis was performed checking the impact of the method on the degree of resilience explained before.

For this purpose, a pair comparison between PCA and the three-kernel type of KPCA were done using paired sample t-tests for equality of means (Tab.1).

**Table 1.** t-tests for FRI between PCA and KPCA (Gaussian, linear, polynomial)

		Mean	Std .deviation	t	df	p-value
Pair.1	KPCA Gauss - KPCA -linear	.0626	.156	4.49	125	.000
Pair 2	KPCA Gauss - KPCA -Poly	.0331	.155	2.39	125	.018
Pair 3	KPCA Gauss -PCA	.0807	.260	3.47	125	.001
Pair 4	KPCA Linear - KPCA -Poly	- .0295	.018	- 17.5	125	.000
Pair 5	KPCA Linear -PCA	.0180	.222	.911	125	.364
Pair 6	KPCA -Poly-PCA	.0475	.047	2.39	125	.018

Results show significant differences ( $p\text{-value} < 0.05$ ) between most of pairs of methods compared, unlike the pair 5 (PCA and PCA kernel linear) with ( $p\text{-value} > 0.05$ ), showing no significant differences. This can be explained by the linear transformation of data using a mapping function.

Validation step based on the overall performance interpretation of the four FRI developed, (PCA linear, KPCA (Linear, Polynomial and Gaussian)), highlight the performance of nonlinear PCA system using a Gaussian kernel to assess Flood Resilience Index. Therefore, it was the only index that supports the purpose of this study and announcing that areas with higher vulnerability levels examined have lower resilience levels.

Furthermore, Kernel KPCA is able to capture the nonlinear relationship between input variables and the output which is FRI. Thus, Gaussian kernel PCA is more successful than conventional PCA. Thus, the kernel PCA is able to produce higher results than the linear PCA approach, reducing dimensional data and using a fewer principal component (4).

## 5. CONCLUSION

In this paper, PCA and KPCA algorithms were used to evaluate Flood Resilience Index in our study area. PCA is a classical technique of dimensionality reduction, applied in many problems related to floods. The improved algorithm of PCA is KPCA analysis method. This method is performing better when it comes to a nonlinearly transformed feature space.

Three kernel functions were used. Experiments were performed based on three municipalities. Experimental results show that PCA kernel achieves dimensionality reduction and help to get better Flood Resilience Index

than conventional PCA. The spatial analysis highlights the large disparity between FRI assessed with PCA and KPCA. Further evaluation assessments methodologies will be applied to choose the most significant FRI. The FRI desegregation step will allow us to identify the main drivers of flood resilience within the study area to provide the local decision makers, the target to strength flood resilience.

## References

- Schipper, L., & Pelling, M. *Disasters*, 30(1), 19-38. (2006)
- IPCC: Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. p. 1202. (2014).
- Rus, K., Kilar, V., & Koren, D. *IJDRR* (2018).
- Cutter, S. L., Burton, C. G., & Emrich, C. T. *Journal of Homeland Security and Emergency Management*, 7(1) (2010).
- Sharifi. A and Yamagata. Y., *International Journal of Disaster Risk Reduction*, (2016).
- Hung, H. C., Yang, C. Y., Chien, C. Y., & Liu, Y. C. *Land Use Policy*, 50, 48-58. (2016)
- Asadzadeh, A., Kötter, T., Salehi, P., & Birkmann, J. *International journal of disaster risk reduction*, 25, 147-62. (2017).
- Satour, N., Raji, O., El Moçayd, N., Kacimi, I., and Kassou, N.: *Nat. Hazards Earth Syst. Sci.*, 21, 1101–1118, (2021).
- Kotzee, I., & Reyers, B. *Ecological Indicators*, 60, 45-53. (2016).
- Gómez-Limón. A. Riesgo.l. *Paper prepared for Presentation at the 107th EAAE Seminar "Modelling of Agricultural and Rural Development Policies"*. (2008).
- Das, Gopal, Manojit Chattopadhyay, and Sumeet Gupta. *International Journal of Market Research* 58.6 (2016): 815-834.
- Quan Wang. (2012). arXiv preprint arXiv:1207.3538.
- Scholkopf, Bernhard, Smola, Alexander, and Muller, Klaus-Robert. *Neural computation*, 10 (5):1299{1319, (1998).
- ABHL
- PATEAU M. *Dynamiques superficielles et gestion des risques de l'arrière-pays du complexe portuaire Tanger-Med (Rif septentrional, Maroc)*. Thèse de l'Université Paris Diderot, 282 p. (2014).
- Biggs, Reinette, et al. *Annual review of environment and resources* 37 :421-448.16 (2012).
- Cutter, Susan L., et al. *Global environmental change* 18.4: 598-606. (2008):

18. Prior, Timothy, and Jonas Haggmann. *Journal of risk research* 17.3: 281-298. (2014).
19. Niazi. S. Evaluation des impacts des changements climatiques et de l'élévation du niveau de la mer sur le littoral de Tétouan (Méditerranée occidentale du Maroc) : Vulnérabilité et adaptation, *Thèse de doctorat*. Univ. Mohamed V, Rabat, Maroc. (2007).
20. Nejari Abdelkader, *Revue AFN Maroc*, N° : 12-14. (2014).
21. Satta, A., Snoussi, M., Puddu, M., Flayou, L., & Hout, R. *Coastal and Shelf Science*, 175, 93-105. (2016).
22. Scherzer, Sabrina, Päivi Lujala, and Jan Ketil Rød. " *International Journal of Disaster Risk Reduction* 36: 101107(2019).