# Continental Portuguese Territory Flood Social Susceptibility Index

Grosso, N., Dias, L., Costa, H., Garrett, P.

## Abstract

**Keywords:** Flood Susceptibility, Vulnerability, Susceptibility, Multicriteria Evaluation.

## 1. Introduction

The number of natural disasters as well as the number of people affected has been increasing in the last decades, showing that that societies are currently more vulnerable and exposed to these phenomena (Ge et al., 2013). Extreme climate events are responsible for 80% of the damage caused by those natural disasters worldwide, with floods affecting more than a billion people in the last decade and causing thousands of deaths every year (Vörösmarty et al., 2013). In Europe, floods, together with windstorms, are the most frequent natural disaster and their damages correspond to a third of total economic losses related to these type of phenomena(EEA et al., 2008, IPCC, 2012).

In the last decades the frequency and intensity of natural extreme events has been increasing (Ge et al., 2013) as a result of climate change induced changes in climatic patterns, which, most likely, will be aggravated in the next years (e.g. Øystein Hov et al., 2013, IPCC, 2012)

For this reason, vulnerability assessment techniques are becoming a fundamental tool in flood risk management, helping to define more effective risk reduction strategies and promoting societal disaster resilience (Birkmann, 2006). The concept of vulnerability was introduced in the 1970's in the context of social sciences and was originally oriented to the risk perception related to catastrophes (Birkmann, 2006). Currently, there are currently several definitions derived from the different scopes of application of the scientific communities behind them (Veen et al., 2009, Thywissen, 2006).

In general, vulnerability can be defined as the loss potential of assets or individuals when exposed to a natural disaster of a certain magnitude (Ionescu et al., 2009, Cutter et al., 2000, Schanze et al., 2006). This definition covers several vulnerability dimensions, namely, physical, social, economic, politic, cultural and environmental that, when aggregated with a physical component (Thywissen, 2006), form a composed vulnerability index (See e.g. Balica et al., 2012, Sebald, 2010). This scope has been expanding to include nowadays concepts such as coping capacity ad resilience (Armaş and Gavriş, 2013). The work presented here refers solely to the social component of this composed index.

Nowadays, there are still many difficulties to determine the flood loss potential due to the lack of data to estimate affected area and their associated costs, mainly at the national level. For that reason, most of the studies developed at this scale only include the main characteristics that define the societal or individual predisposition to be affected, resist, adapt or recover, when exposed to a flood (Ge et al., 2013, Armaş and Gavriş, 2013). In the opinion of the authors of this paper, this characterization, also adopted here, is better suited to define flood social susceptibility (FSS) and therefore the developed index was designated as a Social Susceptibility

under review for publishing

Index (SSI). Nevertheless the adopted methodology derives from the existing bibliography on flood vulnerability indexes.

There are usually two different methodologies to evaluate flood social vulnerability: a) the SoVI (Social Vulnerability Index) model and; b) the SeVI (Social vulnerability assessment using spatial multi-criteria analysis) model. The first was developed by Cutter et al. (2003) and uses a Principal Component Analysis (PCA) to select the most representative indicators to compose the final index without providing different variable weights. Since its formulation, this method has been widely used in the United States and more recently in Europe, becoming the standard vulnerability assessment method (Armaș and Gavriș, 2013, Ge et al., 2013). The second is based in a multicriteria analysis developed by Saaty (1980) named analytical hierarchical process (AHP). This method combines expert evaluation and statistical methods to determine the relative weight for each variable.

This main objective of this work is to develop a SSI for the Portuguese territory based on the approach initially proposed by Cutter et al. (2003) and further developed by Fekete (2010). Although there are some studies in European countries, to develop national flood vulnerability indexes, in Portugal there is only one published social vulnerability index for some municipalities, implemented by de Oliveira Mendes (2009), that includes both natural and technological risks and does not differentiate floods.

Although outside the scope of this paper, the results presented here are part of a composed flood vulnerability index for continental Portugal that also includes exposure and physical susceptibility. This index was developed in the scope of the CIRAC project (Flood Risk Mapping in Climate Change Scenarios - http://siam.fc.ul.pt/cirac/).

## 2. State of the Art

## 3. Materials and Methods

### 3.1 Study area

Continental Portugal, situated in the southwest of Europe, is part of the Iberian Peninsula and occupies an area of 89 015 km², currently divided into five NUTS II regions, 278 municipalities and 2882 4050 parishes . In 2001 the number of parishes was significantly higher (4037) and only decreased to the current number in 2013, after a national administrative reorganization process (INE, 2011) (Figure 1).

under review for publishing

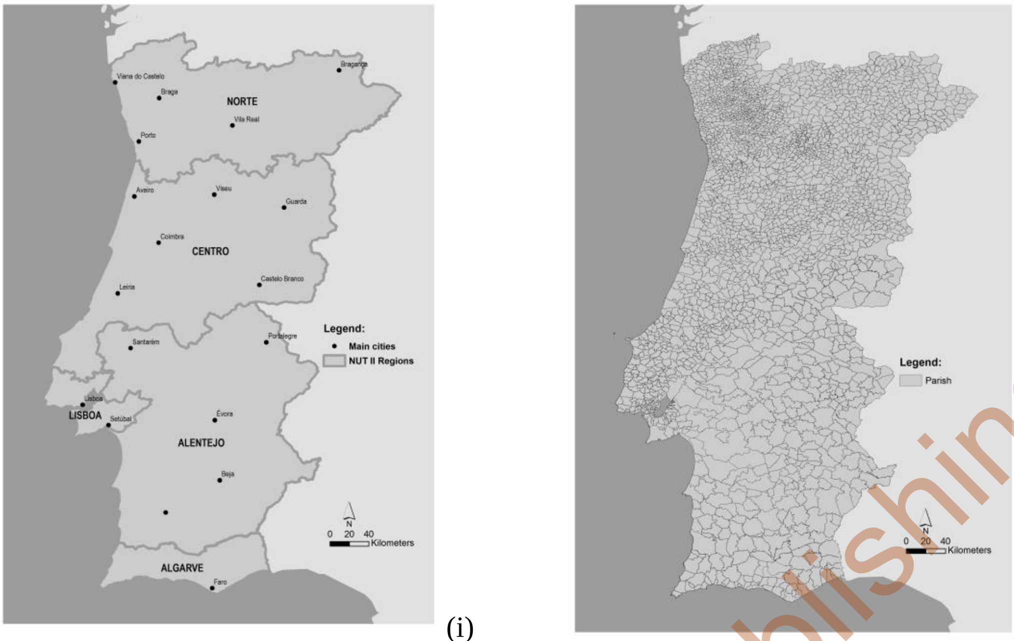



**Figure 1 - Characterization of the study area – (i) Portuguese NUTS II regions, main cities and municipalities; (ii) Portuguese Parishes.**

According to the 2011 census data (INE, 2011), its number of inhabitants increased approximately 2%, between 2001 and 2011, from 9 869 343 to 10 047 083, which represented a decrease in the growth rate, when compared to the 5% registered in the previous decade. From the 278 municipalities, 171 in 2001 and 198 in 2011 have registered a decrease in population, contributing to an unbalance in population spatial distribution (INE, 2001), with an overall movement from rural to urban municipalities. In the last decades, the migratory movements within the Portuguese territory, together with the emigration and, more recently, immigration phenomena contribute to this tendency. In fact, until the mid-1970s, there was a significant exodus from rural inland regions towards the urban coastal areas, especially in the Lisbon region, where employment opportunities were higher. At the same time, some of those rural populations also emigrated to other European countries, resulting in a decrease of the country's population. In a second phase and until the end of the 1990s, population increased due to a decrease in emigration fluxes, associated with an economic growth after Portugal joined the EU, and an influx of Portuguese, during the African decolonization process. This process also originated a smaller immigration movement to Portugal from the former colonies that has remained constant since then. In this last decade, there was a significant increase in immigration from the new Eastern countries joining the EU, which has been progressively replaced, in the last few years, by immigrants from Brazil and Asia. In parallel, the migratory movements from urban to rural areas inside Portugal continue through: a) the concentration of population along the coastline and; b) the population displacement from rural inland areas to the main cities nearby. Despite this last process the inland municipalities still register an overall population decrease.

Parallely, other demographic phenomena have intensified in Portugal. On one hand, according to the 2011 census, the double aging of the population process, characterized by a decrease in youth population and an increase in older aging groups, has continue to strengthen in the last 40

under review for publishing

years. The total dependency index, defined by ratio between the sum of the population in the 0-14 and over 65 age groups and the active population, defined by the 15-64 age group, has increased 4% in the last decade, supported solely by the 21% growth in the older population.

On the other hand, two other factors had a positive evolution in the last 10 years: education and income. Regarding the first, the percentage  of people with higher education almost double, going from approximately 6 to 12% (INE 2011), while the percentage of people with no education or only the first two cycles of basic education (between the 1st and 6th grade) completed from approximately 67% to 57%. There is also a significant regional unbalance in the evolution of the Portuguese population educational level, with higher educated people usually more concentrated in the coastal urban municipalities. As for average monthly income[1],) statistics show an increase from 729.4 euros in 2000 to 1083.8 euros in 2011. The average income spatial distribution also highlights the same coastal/inland differences shown for other indicators. Those regional differences are visible when analyzing the classifications of the Portuguese NUTS II regions regarding their eligibility to European Cohesion Funds. Under the EU convergence objective, only Lisbon is considered to be a competitiveness and employment region, while Algarve is in the phasing-out stage, and the remaining tree NUTS are still in the group of convergence regions (European Communities, 2007).

Unemployment rate is another important socioeconomical to characterize flood social vulnerability in continental Portugal. In the last 10 years, this rate rose significantly from 6.8 to 13.2%, mostly after the 2008 crisis, after 20 years of low and stable values[2]

In summary, this characterization shows a slow growing and aging country with increasingly lower birth rates, higher education and higher income. Also highlighted by these indicators is the existence of significant regional inequalities between the densely populated, higher educated and richer costal urban areas and the depopulating, lower educated, poorer inland rural regions. This snapshot of the continental Portuguese territory will surely be reflected in the social vulnerability index described in the next sections.

### 3.2 Datasets

Table 1 presents the 39 variables used initially in this study, providing information on its origin, production year, the acronym used in this study to label them as well as information on the indicator group they represent and a first evaluation of its role flood social susceptibility characterization. This evaluation is represented by: one or two minus signs in the case of variables that contribute to a high or a very high flood social susceptibility, respectively; one or two plus signs if a variable decreases it and; one minus and one plus signs, where variables can play both a positive and negative role in flood social susceptibility. Regarding the label, it should be noted that, the acronyms of the final normalized variables used in the composition of the index are equal to the ones presented in the table but with the prefix “NORM”.

under review for publishing

[1] Data retrieved for PORDATA website: http://www.pordata.pt/Portugal/Salario+medio+mensal+dos+trabalhadores+por+conta+de+outrem+remuneracao+base+e+ganho-857 in 17/2/2014

[2] Data retrieved for PORDATA website: http://www.pordata.pt/Municipios/Ambiente+de+Consulta/Tabela in 17/2/2014

**Table 1 – Variables used in this study (with the exception of the Percentage of urban area all data was obtained from Statistics Portugal)s.**

| Description | Name | Weight | Group | Year |
|---|---|---|---|---|
| Buildings with concrete structure | EBAR | ++ | Building construction typology | 2001 |
| Buildings with walls of masonry mortar | EARG | -+ | | 2001 |
| Buildings with walls of stone adobe or pug masonry | EPAT | -- | | 2001 |
| Buildings with other resistance elements (wood, metal) | EORE | -- | | 2001 |
| Exclusively residential buildings | ER | -- | Building function | 2001 |
| Mainly residential buildings | PR | -+ | | 2001 |
| Traditional families without unemployed | FCD_0 | ++ | Income | 2001 |
| Traditional families with one unemployed | FCD_1 | -+ | | 2001 |
| Employed population | IR_EP | ++ | | 2001 |
| Unemployed population seeking the 1st employment | IRD1E | - | | 2001 |
| Unemployed population seeking a new employment | IRDNE | -- | | 2001 |
| Not economically active population | IR_SAC | -+ | | 2001 |
| Foreign population with legal resident status (no UK)[3] | IMIG_VAR | - | | 2010 |
| Guaranteed minimum income[1] | RSI | -- | | 2010 |
| Percentage of social housing buildings | HAB_SOCIAL | - | | 2010 |
| Monthly net average wage[1] | GMMTCO | + | | 2009 |
| Average annual value of pensions[1] | VMAP | + | | 2010 |
| Traditional families with people with less than 15 years | FCPME15 | - | Dependent | 2001 |
| Traditional families with people with 65 or more years | FCPMA65 | -- | | 2001 |
| Families with children under 6 years old | NFF6 | - | | 2001 |
| Child dependency ratio[4] | IND_DJ | - | | 2001 |
| Aged dependency ratio[2] | IND_DI | - | | 2001 |
| Total dependency ratio[2] | IND_DT | - | | 2001 |
| Resident population between 0 and 4 years old | R0_4 | -- | Age | 2001 |
| Resident population between 5 and 9 years old | R5_9 | -- | | 2001 |
| Resident population between 10 and 13 years old | R10_13 | - | | 2001 |
| Resident population between 14 and 19 years old | R14_19 | + | | 2001 |
| Resident population between 20 and 64 years old | R20_65 | ++ | | 2001 |
| Resident population with 65 years and over | R65 | -- | | 2001 |
| Retired persons and pensioners | IR_PR | - | | 2001 |
| Residents with no qualification | IRQA_001 | -- | Education | 2001 |
| Residents with 1st Cycle of basic education | IRQA_110 | - | | 2001 |
| Residents with 2nd Cycle of basic education | IRQA_120 | + | | 2001 |
| Residents with 3rd Cycle of basic education | IRQA_130 | ++ | | 2001 |
| Residents with secondary education | IRQA_200 | ++ | | 2001 |

[3] Value given for the entire municipality and calculated for the parish by pondering the original value by the percentage of area each parish represents in the municipality

[4] Calculated from the 2001 census (Population - n / parish area -$km^2$)

under review for publishing

| | | | | |
|---|---|---|---|---|
| Residents with post-secondary education | IRQA_300 | ++ | | 2001 |
| Residents with Higher education | IRQA_400 | ++ | | 2001 |
| Population density[4] | DENS_POP | - | Urban/ Rural | 2001 |
| Percentage of urban area[5] | PERC_AURB | - | | 2007 |

The selection of indicators took into account their ability to characterize the relevant socioeconomic (e.g. age, income, dependence) and built environment characteristics (building age and typology) for flood social susceptibility assessment in the different parishes within the continental Portuguese territory.

Whenever possible, datasets of similar origin were used to assure input data homogeneity in the development of the final index. For that reason most of the selected data refer to the 2001 census. The 2011 census were not included in this study because only provisional data was available at the time. Whenever the required indicators were not available through this dataset, alternative datasets were used, available in the statistical yearbooks published by Statistics Portugal (INE, 2010a, INE, 2010b, INE, 2010c, INE, 2010d, INE, 2010e) or by other governmental sources (IGP, 2010). All the values were originally provided at parish level, except in the cases indicated in the footnotes, where calculations had to be performed to adjust to this scale. In the specific cases of the Dependency Ratios the values were calculated based on the 2001 census and refer to:

a. Youth Dependency Ratio (IND_DJ)– defined by ratio between the sum of the population in the 0-14 age groups and the active population, defined by the 15-64 age group;
b. Aged Dependency Ratio (IND_DI) – defined by ratio between the sum of the population in the over 65 age groups and the active population;
c. Total Dependency Ratio (IND_DT) – the ratio between the sum of the population in the 0-14 and over 65 age groups and the active population.

### 3.3 Methods

The methodology adopted to develop the Portuguese flood social vulnerability index was based on the work of Fekete (2010), and it is comprised of three main stages: a) pre-selecting census data variables that could better describe social vulnerability to floods in Continental Portugal (Table 1) and characterizing their role and influence; b) using a Principal Component Analysis to define the variables or group of variables that better represent the different components of flood social susceptibility; c) aggregating those variables into indicators, according to the components defined in the previous step. This aggregation takes into account the role and influence in flood social susceptibility of the variables (subtracting the sum of the negative ones from the sum of the positive variables); d) composing the final index by summing the different components. This methodology follows the SoVI model, an approach perceived as more appropriate for this study, since it provides a less subjective selection procedure of the most representative variables in large datasets.

under review for publishing

[5] Values calculated based on Land Use Map provided by the Portuguese Geographic Institute (IGP)

The variable pre-selection step consisted of an expert analysis comparing the statistical datasets available for the Portuguese territory with the most relevant factors, identified in previous studies (e.g. Vörösmarty et al., 2013, Fekete, 2010, Azar and Rain, 2007, Cutter et al., 2003), influencing flood social susceptibility: age, income, education, urban/rural background and building function/typology.

After arriving to the final set of variables, shown in Table 1, a PCA was performed, using SPSS 20, to reduce dataset dimensionality to the variables that summarize the main characteristics of flood social susceptibility (Field, 2007). In parallel, analyzing the variables with higher loadings within the main final components variables can help derive a set of indicators that define a social susceptibility profile (Fekete, 2010). Before performing the PCA, a standardization procedure was implemented to render the variable values between different parishes comparable. The standardization reference values differed, according to the different variables: a) building construction and typology variables were normalized by the total number of buildings; b) family income related datasets by the total number of families; c) employed and unemployed population variables by the total number of economically active people; d) the not economically active population by the 2001 total population; e) the foreign population variables and the number of people receiving guaranteed minimum income were divided by the 2010 total population; f) the percentage of social housing buildings by the 2010 total number of buildings; g) monthly net average wage and average annual pensions were not normalized because they already averaged values; h) all gender, age and education variables were normalized by the total number of residents and; i) the total, aged and youth dependency ratios, percentage of urban area and population density are already normalized values. All the reference values are given at the parish scale for the same year of the dataset being normalized.

After standardization, a variable correlation matrix was computed to identify cases of extreme multicollinearity, defined as the variables pairs with an absolute value of the Pearson's Correlation Coefficient R higher than 0.9. In these cases two variables have very similar behaviors and therefore the individual contribution cannot be assessed correctly within the PCA and therefore one of those variables is excluded from the analysis.

The PCA was applied using a full model approach (all variables included) in a Varimax rotation with Kaiser normalization to maximize the sum of the variances of the squared loadings of each variable across the different components, providing a higher loading in a specific component and lower on the remaining. This method provides a clearer interpretation of the correspondence between variables and components. The selection of the final set of variables was established on three criteria based on PCA outputs:

I. The overall Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO statistic) (Kaiser, 1974) should be higher than 0.5 (Hutcheson and Sofroniou, 1999). This statistic provides a general measure of the adequacy of the collected data to perform a factor analysis, based in their correlation matrices. A value higher than 0.5 is considered to be the minimum value to consider that the included variables share a significant common variance and therefore can be further reduced through factor analysis. If the KMO value is lower, individual variables should be dropped, preferentially the ones with lower communality values, a measure of how well each variable is represented in the different components;

II. The diagonal values of the anti-image correlation matrix should also be greater than 0.5. The anti-image correlation matrix contains the negative of the partial correlation

under review for publishing

coefficients between each pair of variables. The diagonal of this matrix provides the individual KMO statistics and when one its values is below the 0.5 threshold, one of the two variables involved should be excluded since this means that they are not well factored into the principal components (Feteke,2010);

III. The off-diagonal values of the anti-image correlation matrix, representing the negative of the partial correlations between variables, should be as small as possible in a good factor model (Field, 2007). A threshold value of 0.6 was established for this study (Feteke, 2010). If lower values are found one of the involved variables should be excluded.

These three criteria were applied in the order they are presented in this paper and whenever one variable was excluded, the PCA was reprocessed, since removing one variable changes the final model and it is necessary to recalculate all statistics.

After arriving to a final model, the final set of principal components was chosen based on an evaluation of the eigenvalues, a measure of the standardized variance associated with a particular factor, related to each principal component or factor. Only the components with an eigenvalue higher than 1 were included as flood social susceptibility indicators. Each variable was attributed to one of those specific components based on their highest loading value. A lower threshold loading value of 0.5 was defined to consider that a certain variable is strongly factored into a component. The final grouping of the variables into the different components and their respective signs was interpreted to identify the flood social susceptibility indicators being characterized by each component.

From the variables contained in each component/indicator, only two variables with a positive influence on flood social susceptibility and two with a negative influence were chosen to be included in the index, based on the highest loadings. To arrive to the final values per parish of each of the identified indicators, the values of the corresponding variables were aggregated by calculating the difference between the averaged sums of the variables with positive and negative influence, as can be seen in Equation 1 (adapted from Feteke, 2010):

$$Indicator = \frac{\sum Var_P}{N_P} - \frac{\sum Var_N}{N_N}$$ **Equation 1**

where $Var_P$ and $Var_N$ correspond to the values of the variables with positive and negative influence, and $N_P$ and $N_L$ to their respective number of variables. All variables were previously normalized to a 0 to 1 scale, based on their minimum and maximum values. Therefore, the final indicator values varied between -1 (indicating higher flood social susceptibility) and 1 (lower).

The final step was to aggregate the different indicators into the final flood susceptibility per parish index by summing the values of all indicators. Since all indicator values could theoretically vary from -1 to 1, the index can vary between –N (highest flood social susceptibility) to N (lowest), where N is the total number of indicators.

## 4. Results and Discussion

This results section is divided into two parts. The first focuses on the description of the main PCA results that established the set of indicators and variables introduced in the final index. The

under review for publishing

second discusses the index's capability to characterize flood social susceptibility index across the Portuguese territory and the main reasons behind its spatial distribution.

As described in the Methods section, the first variable selection step was to compute a correlation matrix based on the normalized variable values to identify cases of extreme multicollinearity ($|R| \geq 0.9$). As shown in Table 2, several age related variables pairs exhibited high correlation values. This was expected for several reasons:

1) some variables often refer to very similar age groups like, for instance:
   a) the aged dependency index (IND_DI) and the traditional families with people with 65 or more years (NORM_FCPMA65);
   b) the retired persons and pensioners (NORM_IR_PR) and the
2) one variable is included in a broader one and can be the main responsible for its variance, such as:
   a) the youth dependency index (IND_DJ) and the resident population between 5 and 9 years old (NORM_R5_9);
   b) the traditional families with people with less than 15 years (NORM_FCPME15) and the resident population between 0 and 4 years old (NORM_R0_4) and 5 and 9 years old (NORM_R5_9);
   c) the total dependency ratio (IND_DT) and the resident population over 65 years old (NORM_R65)
3) the two variables are inversely correlated, as is the case of:
   a) the resident population over 65 years and the resident between 20 and 65 years old, since areas with a higher percentage of active population, usually have a smaller percentage of residents in the older age groups (typically the parishes located around cities) and vice-versa (like the rural areas)

Since for all these cases, maintaining the two variables would not add any extra information to the final model, one of the variables was excluded (variables marked in grey in Table 2). Preference was given, in one hand, to variables with a broader scope and, on the other hand, a focus on flood susceptible age groups (such as the children and the elderly). An example is the selection of the dependency ratios and the traditional families' indicators over the different age groups of the resident population. The only exception was the exclusion of the aged dependency ratio (IND_DI), because it was already highly correlated with other broad variables such as the total dependency ratio (IND_DT) and the traditional families with people with 65 or more years (NORM_FCPMA65). By adopting this strategy it was possible to exclude a wider number of variables and maintain only the more transversal ones with useful information in flood social susceptibility. Nevertheless, it should be noted that this type of analysis is subjective and therefore open to different interpretations.

Apart from the age related variables, only three other pairs were found, all inversely correlated meaning that they are complementary variables:

a. exclusively residential buildings (NORM_ER) and mainly residential buildings (NORM_PR);
b. traditional families without unemployed (NORM_FCP0) and traditional families with one unemployed (NORM_FCP1);
c. not economically active population (NORM_IR_SAC) and employed population (NORM_IR_EP).

under review for publishing

The criteria for maintaining one variable from each pair was either a higher representativity of the variable in the Portuguese territory (a. and c.) or a higher information content regarding flood social susceptibility (b.).

**Table 2 – Variable pairs within the correlation matrix with extreme multicollinearity (|R|≥0.9). In grey are the variables excluded from the PCA. In some pairs both variables are marked as excluded because of other high correlations they exhibited with different variables.**

| Variable pairs with \|R\|≥0.9 | |
|---|---|
| IND_DI | NORM_FCPMA65 |
| IND_DI | NORM_R20_65 |
| IND_DI | NORM_IR_PR |
| IND_DI | NORM_R65 |
| IND_DJ | NORM_R5_9 |
| IND_DT | NORM_R20_65 |
| IND_DT | NORM_IR_PR |
| IND_DT | NORM_R65 |
| IND_DT | IND_DI |
| NORM_FCPMA65 | NORM_R20_65 |
| NORM_IR_PR | NORM_FCPMA65 |
| NORM_NFF6 | NORM_FCPME15 |
| NORM_R0_4 | NORM_FCPME15 |
| NORM_R0_4 | NORM_NFF6 |
| NORM_R20_65 | NORM_FCPMA65 |
| NORM_R5_9 | NORM_FCPME15 |
| NORM_R65 | NORM_FCPME15 |
| NORM_R65 | NORM_FCPMA65 |
| NORM_R65 | NORM_R20_65 |
| NORM_R65 | NORM_IR_PR |
| NORM_PR | NORM_ER |
| NORM_FCD_0 | NORM_FCD_1 |
| NORM_IR_SAC | NORM_IR_EP |

This step excluded 11 variables which meant only 28 were introduced into the PCA.

The first full model approach PCA provided an overall KMO statistic of approximately 0.7, well above the 0.5 minimum threshold referred in the Methods section. This means that the variables have some common variance and therefore the dataset can be reduced using a factor analysis method like the PCA. This value progressively increased to a final value of 0.86 as the variables with individual KMO statistics lower than 0.5 were removed in a recursive way, following the order given in Table 3. Three of removed variables refer to building typology (NORM_EORE, NORM_EPAT and NORM_EARG): This is not surprising since most of the variables in the dataset refer to socioeconomic characteristics of either individuals or families

under review for publishing

which might not correlate as well with building related variables. The remaining variables refer to income/unemployment (NORM_IRD1E, GMMTCO and NORM_IRDNE), one to education (NORM_IRQA_110) and another to building function (NORM_IRQA_110). Although any of these variables could help characterize flood social susceptibility, the decision to remove them took into consideration that other variables could provide similar information, like, for instance, in the case of building typology, the "Buildings with concrete structure" (NORM_EBAR) variable.

**Table 3 – Excluded variables due to low individual KMO values (<0.5) taken from the diagonal of the anti-image correlation matrix**

| Excluded variables (individual KMO<0.5) |
|---|
| NORM_EORE |
| NORM_EPAT |
| NORM_IRD1E |
| GMMTCO |
| NORM_IRDNE |
| NORM_IRQA_110 |
| NORM_EARG |
| NORM_ER |

Finally, as shown in Table 4, the off-diagonal values exclusion criteria also reduced the number of variables included in the final model. As in previous steps, the selection of the excluded variables within each pair took in consideration their relative territorial representativeness and their importance to characterize flood social susceptibility. For instance, the decision to keep the variable "Residents with secondary education" (NORM_IRQA_200) and exclude the variables "Residents with 3rd Cycle of basic education" (NORM_IRQA_130) and "Residents with Higher education" (NORM_IRQA_400) was based on two reasons: a) it is broader variable than NORM_IRQA_130 since it represents all stages of secondary education and; b) in the opinion of the authors, it represents a more significant cut-off education group regarding social susceptibility to floods than NORM_IRQA_400.

**Table 4 – Variable pairs with off-diagonal anti-image correlation matrix values > 0.6. In grey are the excluded variables based on this criterion.**

| Variable pairs | |
|---|---|
| IND_DJ | NORM_FCPME15 |
| IND_DT | NORM_FCPMA65 |
| PERC_AREAURB_FREG | DENS_POP |
| IND_DJ | NORM_R10_13 |
| NORM_IRQA_200 | NORM_IRQA_130 |
| NORM_IRQA_400 | NORM_IRQA_200 |

After arriving to a set of the most representative variables to include in the final model, the PCA was recalculated. From all the calculated components, three were selected to define the main flood social susceptibility indicators that will compose the SSI (Table 5). These three

under review for publishing

components were the only with eigenvalues higher than 1, explaining approximately 63% of the total dataset variability. Table 5 shows the correspondence between original variables and components based on their higher loadings. The definition of the three flood social susceptibility indicators represented by these components resulted from an interpretation of their main variables:

1. Regional conditions included most of the education variables (NORM_IRQA_001, NORM_IRQA_120, NORM_IRQA_200, NORM_IRQA_300) as well as an income variable related to average annual value of pensions (VMAP), a population density variable (DENS_POP) able to differentiate urban and rural areas and a building typology variable that identifies areas with higher or lower presence of concrete based buildings. As referred above in the description of the study area, all these variables can help to characterize the significant regional inequalities between less susceptible coastal urban areas and the more vulnerable inland regions. Furthermore, those variables, can also help distinguish, within the inland areas, some important urban areas from the remaining more rural territory. The assumption of a higher vulnerability in inland regions is mainly associated to lower education and income levels and distance;
2. Age, that includes all variables related to more susceptible age groups (the children - NORM_FCPME15 - and the elderly - NORM_FCPMA65) as well as the more resilient (active population - NORM_IR_EP)
3. Social Exclusion, defined by variables characterizing the lower income (NORM_RSI_Total, NORM_Edif_habit_Social) or possibly less integrated emigrant communities (NORM_Imigrantes_Varios).

**Table 5 – Final components and their corresponding variable loadings. The name given to each component was based on the interpretation of the flood social susceptibility characterization given by the variable group that composes it**

| Variables | Component | | |
|---|---|---|---|
| | Regional conditions | Age | Social Exclusion |
| NORM_IRQA_001 | -0.647 | | |
| NORM_IRQA_120 | | 0.835 | |
| NORM_IRQA_200 | 0.882 | | |
| NORM_IRQA_300 | 0.753 | | |
| VMAP | 0.784 | | |
| DENS_POP | 0.715 | | |
| NORM_EBAR | 0.385 | | |
| NORM_R14_19 | | 0.747 | |
| NORM_FCPME15 | | 0.925 | |
| NORM_FCPMA65 | | -0.801 | |
| NORM_IR_EP | | 0.634 | |
| NORM_Imigrantes_Varios | | | 0.800 |
| NORM_RSI_Total | | | 0.432 |
| NORM_Edif_habit_Social | | | 0.787 |

Finally, for each indicator, up to two variables with a positive influence on flood social susceptibility and two with a negative influence were selected to determine its final value. The

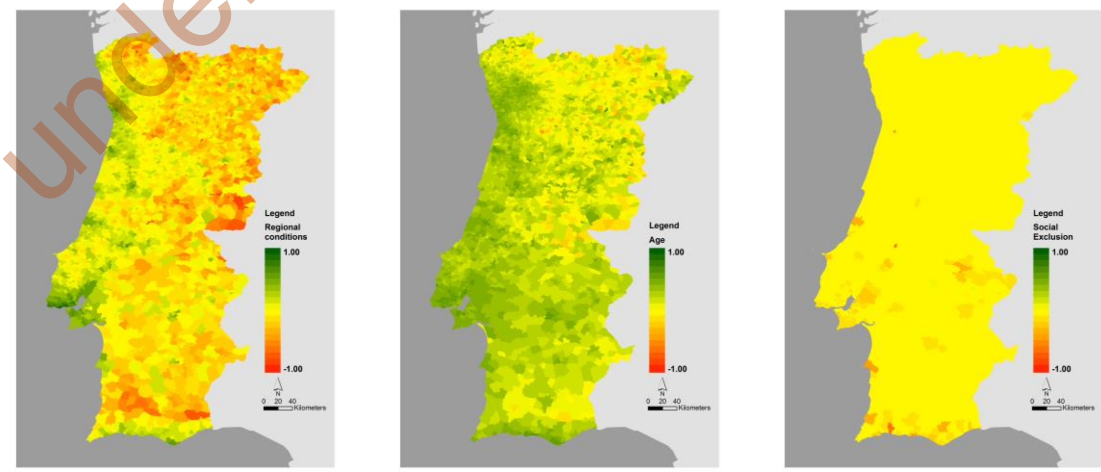
under review for publishing

selection was based on the highest loadings present in each indicator and in the interpretation of the role each variable played regarding flood social susceptibility (negative or positive influence. Table 6 shows that: a) the first indicator uses two different positive variables (higher value, lower susceptibility), residents with secondary education (NORM_IRQA_200) and average annual value of pensions (VMAP), to characterize education and income and only one negative variable (higher value, higher susceptibility) to characterize the presence of populations with lower education (residents with no qualification, NORM_IRQA_001); b) in the age indicator the selected positive variable is related to the presence of people in active age, usually less susceptible to floods and the two negative variables are related to the existence of higher susceptible age groups (children under 15 and elderly over 65 years old); c) the social exclusion indicator is composed of two negative indicators related to the presence of emigrant lower income communities, which is understandable since it is an indicator aimed at characterizing highly vulnerable populations.

**Table 6 – Final set of variables included in each indicator that composed the final flood SSI.**

| Indicators | Final Index Variables | |
|---|---|---|
| | Positive influence on FSS | Negative influence on FSS |
| **Regional conditions** | NORM_IRQA_200<br>VMAP | NORM_IRQA_001 |
| **Age** | NORM_IR_EP | NORM_FCPME15<br>NORM_FCPMA65 |
| **Social Exclusion** | | NORM_Imigrantes_Varios<br>NORM_Edif_habit_Social |

The maps with the results, per parish, of each indicator and the aggregated index are shown in Figure 2 and Figure 3. All indicators maps use a common scale of equal 0.1 intervals between -1 (higher susceptibility) and 1 (lower susceptibility). The SSI index final map also uses a 0.1 equal interval scale between -1.5 and 1.5. Although the indicators do not cover the full scale range, the definition of a common scale facilitates indicator interpretation, intercomparison and the characterization of their relative influence to the final index.




(a) (b) (c)

under review for publishing

**Figure 2 – Maps of the three flood social susceptibility indicators for the continental Portuguese territory: (a) Regional Conditions; (b) Age; (c) Social Exclusion.**

The regional conditions indicator, related to education and income variables, expresses the significant regional inequalities described in the Study Area section. The lower susceptibility values are concentrated in the Setubal-Viana do Castelo coastal axis and along Algarve's coastline (see Figure 1). Those correspond to the more developed Portuguese regions, where the population has higher education and income levels. The major inland urban centers where most of the youth population of the surrounding rural areas migrated, in search of better work conditions, also present low susceptibility values. The higher susceptibility values are associated with rural inland areas with a more fragile economy and an aging population.

This territorial dichotomy is also present in the age indicator, although the higher values are mostly focused in the Centre and North inland regions, due to a lower presence of individuals in active age and a higher incidence of elderly rural populations. In the northern part of Alentejo the aging population factor is partially absorbed by the higher presence of people in active age.

Finally the social exclusion indicator shows a more limited territorial influence, concentrated in the southern regions with a high incidence of low income and emigrant communities.

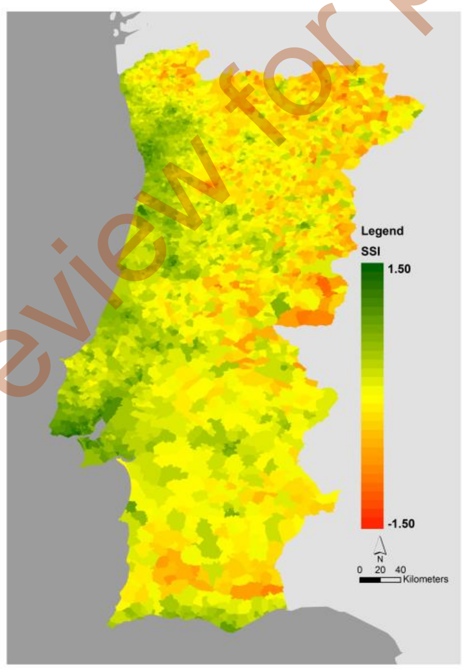



**Figure 3 – Flood Social Susceptibility Index (SSI) for the continental Portuguese Territory**

The SSI index compiles the partial information given by its indicators, highlighting, as expected, the coastal/inland differences and showing a higher ability to cope with floods in the more populated and developed coastal urban centers along the Atlantic coast. Within those areas, the metropolitan regions of Lisbon and Oporto have the lowest SSI values, mainly due to their higher per capita incomes and education and lower unemployment. Higher social

under review for publishing

susceptibility values are located in the poorer inland regions, with a focus on the north and center eastern quadrant and the northern and southern part of Alentejo.

## 5. Conclusions

The main objective of this work was to develop a flood social susceptibility index for the continental Portuguese territory based on the most representative variables able to characterize different influencing factors such as age, income, education and building typology. This goal was achieved effectively using a PCA based methodology to reduce the original set of 42 variables to eight, representing three indicators used in the final index: regional conditions, which aggregated income and education variables; age with parameters related to susceptible age groups and; social exclusion characterizing particularly susceptible very low income and emigrant communities. The PCA based technique avoided successfully most of the subjective selection processes based on expert analysis methodologies that can add bias to the final index, based on personal assumptions. Furthermore, the use of a restrict set of variables contributed to index simplicity and consequently to its transparency, as shown in the straightforward interpretation of the results given in the previous section. In general, the index correctly identified populations more socially susceptible to floods, mostly concentrated in rural inland areas with lower income and education levels, when compared with the coastal region between Viana do Castelo and Setúbal.

Nevertheless this index would benefit in the future from a validation procedure similar to the one developed by Feteke (2010). This study correlated questionnaire answers given by people affected by floods in Germany with the variables in the main PCA components to choose the variables to include in the index. The main reason not to pursue this methodology in the work presented here was the lack of systematized information on flood events in Portugal. Future integration with the results of projects like DISASTER (GIS database on hydro-geomorphologic disasters in Portugal: a tool for environmental management and emergency planning - http://riskam.ul.pt/disaster/) can improve this type of information and provide a good framework for an extensive nationwide validation of the current SSI.

under review for publishing

## 6. Bibliographic references

ARMAȘ, I. & GAVRIȘ, A. 2013. Social vulnerability assessment using spatial multi-criteria analysis (SEVI model) and the Social Vulnerability Index (SoVI model) – a case study for Bucharest, Romania. *Nat. Hazards Earth Syst. Sci.,* 13**,** 1481-1499.

AZAR, D. & RAIN, D. 2007. Identifying population vulnerable to hydrological hazards in San Juan, Puerto Rico. *GeoJournal,* 69**,** 23-43.

BALICA, S. F., WRIGHT, N. G. & MEULEN, F. 2012. A flood vulnerability index for coastal cities and its use in assessing climate change impacts. *Natural Hazards,* 64**,** 73-105.

BIRKMANN, J. 2006. *Measuring vulnerability to natural hazards - Towards disaster resilient societies*, United Nations University Press.

CUTTER, S. L., BORUFF, B. J. & SHIRLEY, W. L. 2003. Social Vulnerability to Environmental Hazards. *Social Science Quarterly,* 84**,** 242-261.

CUTTER, S. L., MITCHELL, J. T. & SCOTT, M. S. 2000. Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina. *Annals of the Association of American Geographers,* 90**,** 713-737.

EEA, WHO & JRC 2008. *Impacts of Europe's changing climate—2008 indicator-based assessment*, Copenhagen, Denmark, European Environment Agency.

FEKETE, A. 2010. *Assessment of Social Vulnerability to River Floods in Germany*, Bonn, University Institute for Environment and Human Security (UNU-EHS).

FIELD, A. 2007. *Discovering Statistics Using SPSS*, SAGE Publications.

GE, Y., DOU, W., GU, Z., QIAN, X., WANG, J., XU, W., SHI, P., MING, X., ZHOU, X. & CHEN, Y. 2013. Assessment of social vulnerability to natural hazards in the Yangtze River Delta, China. *Stochastic Environmental Research and Risk Assessment*, 27**,** 1899-1908.

HUTCHESON, G. D. & SOFRONIOU, N. 1999. *The Multivariate Social Scientist: Introductory Statistics Using Generalized Linear Models*, SAGE Publications.

IGP 2010. Carta de Uso e Ocupação do Solo de Portugal para 2007 (COS 2007) - Memória Descritiva. Lisboa.

INE 2010a. *Statistical Yearbook of Alentejo Region 2010*, Instituto Nacional de Estatística, I.P.

INE 2010b. *Statistical Yearbook of Algarve Region 2010*, Instituto Nacional de Estatística, I.P.

INE 2010c. *Statistical Yearbook of Centro Region 2010*, Instituto Nacional de Estatística, I.P.

INE 2010d. *Statistical Yearbook of Lisboa Region 2010*, Instituto Nacional de Estatística, I.P.

INE 2010e. *Statistical Yearbook of Norte Region 2010*, Instituto Nacional de Estatística, I.P.

INE 2011. *Censos 2011 – Resultados Provisórios*, Instituto Nacional de Estatística, I.P.

IONESCU, C., KLEIN, R. T., HINKEL, J., KAVI KUMAR, K. S. & KLEIN, R. 2009. Towards a Formal Framework of Vulnerability to Climate Change. *Environmental Modeling & Assessment*, 14**,** 1-16.

IPCC 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation - Special Report of the Intergovernmental Panel on Climate Change. *In:* FIELD, C. B., BARROS, V., STOCKER, T. F., DAHE, Q., DOKKEN, D. J., EBI, K. L., MASTRANDREA, M. D., MACH, K. J., PLATTNER, G.-K., ALLEN, S. K., TIGNOR, M. & MIDGLEY, P. M. (eds.) Cambridge Univeristy Press ed.: First Joint Session of Working Groups I and II.

KAISER, H. 1974. An index of factorial simplicity. *Psychometrika*, 39**,** 31-36.

ØYSTEIN HOV, ULRICH CUBASCH, ERICH FISCHER, PETER HÖPPE, T. I., NILS GUNNAR KVAMSTØ, ZBIGNIEW W. KUNDZEWICZ, DANIELA REZACOVA, DAVID RIOS, FILIPE DUARTE SANTOS, BRUNO SCHÄDLER, OTTÓ VEISZ, C. Z., RASMUS BENESTAD, J. M., M. DONAT, GREGOR C. LECKEBUSCH & ULBRICH, U. 2013. Extreme Weather Events in Europe: preparing for climate change adaptation. Banson production, Cambridge, CB1 2BH, UK ed. Norway: The Norwegian Academy of Science and Letters.

SAATY, T. L. 1980. *Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw Hill Higher Education.

SCHANZE, J., ZEMAN, E. & MARSALEK, J. 2006. *Flood Risk Management: Hazards, Vulnerability and Mitigation Measures* Springer Netherlands.

Under review for publishing

SEBALD, C. 2010. *Towards an Integrated Flood Vulnerability Index – A Flood Vulnerability Assessment.* MSc, University of Twente.

THYWISSEN, K. 2006. Core terminology of disaster reduction: A comparative glossary. *In:* BIRKMANN, J. (ed.) *Measuring Vulnerability to natural hazards: towards disaster resilient societies.* New York, Tokyo: United Nations University Press.

VEEN, A. V. D., DOPHEIDE, E., PARKER, D., TAPSELL, S., HANDMER, J., GREGG, C., BONADONNA, C. & FERRARA, F. F. 2009. Del. 1.1.3 State-of-art on vulnerability of socio-economic systems (ENSURE Project). *Enhancing resilience of communities and territories facing natural and na-tech hazards.*

VÖRÖSMARTY, C. J., BRAVO DE GUENNI, L., WOLLHEIM, W. M., PELLERIN, B., BJERKLIE, D., CARDOSO, M., D'ALMEIDA, C., GREEN, P. & COLON, L. 2013. Extreme rainfall, vulnerability and risk: a continental-scale assessment for South America. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371.

under review for publishing