A comparison of social vulnerability indices specific to flooding in Ecuador: principal component analysis (PCA) and expert knowledge

Agathe Bucherie a,*, Carolynne Hultquist b, Susana Adamo b, Colleen Neely c, Fernanda Ayala d, Juan Bazo e, f, Andrew Kruczkiewicz a, e, g

a International Research Institute for Climate and Society (IRI), Columbia Climate School, Columbia University, New York, USA
b Center for International Earth Science Information Network (CIESIN), Columbia Climate School, Columbia University, New York, USA
c School of International and Public Affairs (SIPA), Columbia University, New York, USA
d Disaster Risk Management Department, Ecuadorian Red Cross (CRE), Quito, Ecuador
e Red Cross Red Crescent Climate Centre, The Hague, Netherlands
f Universidad Tecnológica del Perú, Lima, 15046, Perú
g Faculty of Geo-Information Science and Earth Observation, University of Twente, 7514, Enschede, AE, the Netherlands

ABSTRACT

Social vulnerability is a key component of the risk equation alongside the context of the hazard and exposure. Increasingly, social vulnerability indices are used to better understand and predict the consequences of disasters, and support the development of improved disaster management policies. Humanitarian organisations particularly strive to capture social vulnerability in their decision processes relative to prioritisation of actions before disasters occur. This research supports the Ecuadorian Red Cross in generating a flood-specific social vulnerability index to inform flash flood early action at the Parroquia level in Ecuador. This paper compares the results from the two most common approaches used to create composite indices, one using the weighting of variables from disaster experts’ judgments (referred to as Expert method) and the other using PCA analysis, with one or more components. While all outcomes reveal similar trends in areas where most indicators suggest the lowest (urban areas) or highest (the Amazon and northwest coastal regions) social vulnerability, the research shows that the choice of the method matters for assessing the social vulnerability in the rest of the country where there are less pronounced vulnerability signals. In those areas, PCA-driven indices suggest higher relative vulnerability levels than Expert outcomes. Further, in the Andes particularly, the PCA outcomes result in wider distribution than the Expert outcomes, and therefore more heterogeneity in the vulnerability assessment. While divergence in outcomes suggests particular attention with the use of composite indexes for decision making, our results provide support to understand the sensitivity in flood-specific social vulnerability outcomes spatially. To go further we emphasise the importance of using historical flood impact data to evaluate the contribution of each variable in the final social vulnerability scores.

* Corresponding author.
E-mail address: agathe@iri.columbia.edu (A. Bucherie).

https://doi.org/10.1016/j.ijdrr.2022.102897
Received 5 October 2021; Received in revised form 7 February 2022; Accepted 6 March 2022
Available online 15 March 2022
2212-4209/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).
1. Introduction

Vulnerability products are of increasing interest for humanitarian practitioners, to prioritise where and when to take anticipatory actions, as well as informing decisions on the type of action to take [1,2]. This study is developed particularly to support the Ecuadorian Red Cross (ERC) in building a social vulnerability index. Combined with hazard and exposure information, the social vulnerability index will be used as an input for impact-based forecasts for flash-floods, and to inform flash flood early action protocol at the administrative level 3 (Parroquia) in Ecuador. Indeed, vulnerability indices have been developed in Ecuador using different socio-economic indicators and for specific contexts such as climate change [3] and earthquakes [4]. However, there is a gap in addressing social vulnerability specific to the context of flooding.

In this study, our definition of “social vulnerability” incorporates socio-economic variables characterising the pre-event individual, household and community vulnerability, as well as variables describing how the vulnerable groups can cope with disasters, during and after an event. These two components (referred as “vulnerability” and “cooping capacity”) are defined by ERC according to the INFORM standard practises used in the humanitarian sector [5]. We worked with disaster practitioners to identify a broad selection of variables that are relevant to describe relative social and economic vulnerability [6] in the context of flooding in Ecuador, in order to build a social vulnerability index informing vulnerability to flooding. However, different input selection and construction methods exist to generate spatial composite social vulnerability indexes, resulting in different outcomes [7]. In addition, studies providing guidance on the validation and evaluation of the methods used and indices outcomes are still rare [8,9]. Therefore, our research aims to quantify the differences between flood-specific social vulnerability outputs from the two of the most common approaches (Expert weighting of variables and PCA). As recent studies propose the use of disaster impact data to validate and calibrate vulnerability index [10–13], we discuss here if and how validation with impact data could help disaster practitioners evaluating flood vulnerability indices in Ecuador.

1.1. Social vulnerability within disaster risk management

Vulnerability, referring to the propensity of predisposition for any exposed element to be adversely affected [14,15], is a key determinant of risk alongside the context of the hazard and exposure [16]. Social vulnerability - the focus of this paper - can be conceptualised as encompassing all social, political, cultural, economic, and institutional characteristics of a place and its population, resulting in different ways of preparing, experiencing and recovering from hazards [6,16,17]. Cutter et al. (2003:243) suggest that social vulnerability is the result of social inequalities (social factors that influence or shape the susceptibility of various groups to harm and their ability to respond) and place inequalities (characteristics of communities and the built environment - urbanisation level, economic performance, etc. - that contribute to the social vulnerability of places). Social vulnerability can take form at different dimensions and scale and levels of analysis (from individual to institutional, and from local neighbourhood to a whole country). Various factors influence the social vulnerability of the population, which may differ amongst individuals and social groups exposed to the same hazard, resulting potentially in different outcomes [18]. When aggregated, those characteristics describe the vulnerability of places [18]. Therefore, analysing how those factors contribute to hazard-specific social vulnerability and how it changes in time and space is essential to better understand and predict the relative impacts of disasters on exposed populations [19–22]. Doing so can lead to improved disaster risk reduction practices aiming at reducing social vulnerability, through the development of policies for long term mitigation and adaptation planning [23,24]. Social vulnerability assessments are also used by humanitarian organisations to support disaster management decisions, before, during and after disasters, and increasingly for pre-disaster anticipatory action [25]. Vulnerability assessments, such as Vulnerability and Capacity Assessment [26] are often household surveys with qualitative outcomes. They are typically conducted at local levels without comprehensive coverage which may present challenges for aggregation at sub-national levels.

1.2. Composite social vulnerability indices

While social vulnerability cannot be directly observed or measured, composite vulnerability indices are able to quantitatively estimate relative vulnerability from available proxy variables characteristics. However, a composite social vulnerability index strongly depends on the factors considered, and the indicators1 used [7], which also differ according to the context of the place or region [6].

A common way of creating an index involves using expert knowledge to select and weight the indicators to form an index [27–29]. For instance, the Social Vulnerability Index (SVI) produced by the U.S. Centres for Disease Control and Prevention (CDC) uses best judgement for the variable selection [30]. The Climate Vulnerability Index (CVI), developed in the context of Peru, explores different weighting schemes of variables based on participatory consultation and expert opinion [31]. Another example is the Environment Performance Index (EPI), developed using best judgement informed by data and trend analysis to weight the contributions of each category of indicator [32].

Alternatively, one can use a statistical method to determine the contribution of each variable. The most widely used data-driven method to quantitatively measure social vulnerability to hazards applies principal component analysis (PCA) [33–35]. An example of a PCA driven social vulnerability index is the SoVI method [6], first used in the U.S. and now applied elsewhere to spatially compare relative social vulnerability using an analytic factor approach [36,37].

While social vulnerability indices are common tools for risk analysis, how to address the quality and goodness of those indexes remains understudied [8,27,38]. However, recent research have examined the potential to validate vulnerability indexes with external data [10–13]. Decision makers may need to use their own judgement based on how the index was constructed and if it is intended to

---

1 Here we use the terms ‘indicator’ and ‘variable’ interchangeably.
capture local variations in vulnerability.

1.3. The choice of social vulnerability variables

Creation of a vulnerability index necessarily involves a subjective pre-selection of vulnerability indicators and categories based on judgement. This initial pre-selection of variables can lead to very different outcomes even within the same disaster context.

Social vulnerability indices were first generated in contexts specific to certain hazards (e.g., flooding in this study) as the pre-existing factors interacting with an exposure to generate increased risk [39,40]. Social vulnerability was therefore used to characterise the underlying conditions within the population that make some people more susceptible to experience an increased impact from a hazard [6]. More recently, integrating indicators characterising how the population can cope with a disaster [41] have been core to vulnerability assessments [42], to characterise how social vulnerability plays out in the aftermath of disasters. In addition to describing the pre-event vulnerability conditions, coping capacity refers to the resources and capabilities of individuals or groups to face adverse conditions, and how they can potentially manage hazard-related losses, during and shortly after being impacted [14,43,44].

The social vulnerability variables to include in risk analysis and their inter-connections are strongly dependent on the disaster context. While some variables such as wealth, education, age, health and settlement status have been easily identified as an increased vulnerability factor in all disaster contexts [45], other variables related to the physical vulnerability of settlements, such as house structure types, are more hazard-specific [46].

2. Studied area

In continental Ecuador, the climate and hazard occurrence are controlled by the presence of the Andes mountains (Fig. 1), dividing the country into three major geographical units; the Coastal, Andes and Amazon regions [47]. Ecuador is prone to a variety of disasters, including floods, tsunamis, earthquakes, landslides, and volcanic eruptions. Floods, the most damaging of all disasters, occur throughout Ecuador, while geophysical hazards are restricted to specific areas. However, the flood risk differs spatially and temporally [48] related to seasonal variation in precipitation and the physiography within the country. Ecuador’s population is exposed to different types of floods, including recurrent riverine and pluvial flood types, such as large-scale and long-duration floods in low elevated floodplain areas (Guayaquil and Los Rios) and flash floods in the Andes [49]. The Coastal and the Andes regions are characterised by two distinct - wet and dry-seasons, while the Amazonian rainforest area receives rainfall all year round [50,51].

In addition to being exposed to various natural hazards, Ecuador’s population of 17.54 million - amongst which 64% is urban (Word Bank) - presents a wide range of socio-economic vulnerability contexts across the country [3]. This results in significant differences in economic development levels, disaster risk perception, and coping capacities [52], and therefore vulnerability to disasters [53]. While Ecuador is in the high human development category according to the HDI classification, one in four households still lives below the poverty line, with the unemployment rate rising to 35% due to the COVID-19 crisis [54]. Climate shocks are known to have a significant impact on development and the persistence of poverty in Ecuador [55]. Composed of diverse ethnic groups as well as wide social disparities, Ecuador presents one of the highest ethnic and cultural fractionalisations in Latin America [56], 7 out of every 100 Ecuadorians self-identify as indigenous, representing the 14 indigenous nationalities existing in the country [57]. With a ratio of household per capita income between urban and rural households of 1.69, inequalities are particularly observed between the richest urban population and the rural, indigenous and poor populations, more vulnerable to shocks [56,58]. The Gini coefficient reached 0.472 in June 2018 at a national level, but subnational results reveal greater inequalities in urban than rural areas [59]. Additional statistics of poverty and inequality measures in Ecuador are presented in Table 1.

3. Materials and methods

Two methods are developed to create a social vulnerability index specific to flooding in Ecuador, at the Parroquia level. This administrative level was selected as it corresponds to the highest spatial resolution available from census data (admin level 3). In Ecuador, the median area of a Parroquia is about 120 km\(^2\) for a population ranging from less than 100 to 2.5 million inhabitants. First, we selected and compiled variables characterising social vulnerability to floods for the 1032 Parroquias of Ecuador, excluding the Galapagos Islands due to data availability reasons. Second, the variables are integrated into composite vulnerability indices using 1) principal component analysis approach and 2) an expert judgement weighting of the variables. Finally, we compare the results from the two approaches to the composite social vulnerability indices, as well as the results from the contribution of each individual variable.

3.1. Input data selection process

In Ecuador, socio-economic data are available from census survey results of the National Institute of Statistics and Census (INEC),\(^2\) conducted every 10 years. The most recent full census data (2010), as well as more recent modelled updates, are freely available at the three main administrative levels: Provincias, Cantón and Parroquias. Additional data such as poverty related indexes and population projections are jointly published by the World Bank and INEC.\(^3\) In addition, specific indexes are generated by the Ecuadorian

---

\(^2\) http://redatam.inec.gob.ec/.

\(^3\) https://data.humdata.org/dataset/poverty-and-population.
We extracted some of these indicators, aggregated at the Parroquia level, directly from the Red Cross data preparedness dashboard. In order to select relevant socio-economic indicators to characterise social vulnerability to flood in Ecuador amongst a wide range of variables, the following four-step process was used:

1. Identify potential categories representing social vulnerability to flood, specific to the Ecuador context, from literature review and consultation with disaster practitioners.
2. Define appropriate indicators or proxy variables representative of the selected flood vulnerability categories, from data availability.
3. Co-produce spatial analysis of individual indicators with representatives from the ERC to confirm the soundness of indicator selection, and better understand the indicators variation in space, directions, and dependencies.
4. Refine indicator selection after correlation analysis to explore multidimensional relationships and co-dependencies, combined with expert knowledge.

As a result of this four-step process, fifteen indicators within five categories, presented in Table 1, are selected to capture relative social vulnerability to flooding in Ecuador. Following the INFORM risk composite index definition [5], the variables and categories are organised in two components of the social vulnerability to flooding: the “vulnerability component” characterising the pre-event vulnerability to flooding and the “coping capacity component” addressing how the population can overcome adverse situations in the aftermath of a flood [14]. All indicators are corrected to prevent no-data values and checked for outlier data, and reordered so that

---

4 https://www.salud.gob.ec/.
5 https://www.agricultura.gob.ec/.
6 https://dashboard.510.global/.
Table 1
Socio-economic indicators gathered at administrative level 3 (Parroquia) of Ecuador.

<table>
<thead>
<tr>
<th>Categories</th>
<th>Indicators</th>
<th>Indicator’s description</th>
<th>Source and Date</th>
<th>National statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>VULNERABILITY COMPONENT</td>
<td>Vulnerable groups</td>
<td>Disability: Percentage of population with a permanent disability for more than a year.</td>
<td>INEC Censos [57]</td>
<td>5.63%</td>
</tr>
<tr>
<td>Socio-economic</td>
<td>Poverty Incidence</td>
<td>Estimated income-based poverty incidence based on ECV 2014 and CPV 2010.</td>
<td>World Bank and INEC (2016)</td>
<td>32.7% of population below poverty line</td>
</tr>
<tr>
<td></td>
<td>Gini Index</td>
<td>Estimated Gini Index (statistical index of inequality over income and wealth distribution) from 0 to 1.</td>
<td>World Bank and INEC (2016)</td>
<td>Gini Index of 0.45</td>
</tr>
<tr>
<td></td>
<td>Agricultural labor share</td>
<td>Percentage of population working in agriculture, animal farming, forestry and fishery.</td>
<td>INEC Censos [57]</td>
<td>27.8%</td>
</tr>
<tr>
<td>Health</td>
<td>Vector-borne disease</td>
<td>Vector-borne disease incidence (including malaria vivax, chikungunya, dengue and leishmaniasis) per 10000 persons, during the period 2013–2016.</td>
<td>Ministerio de Salud Pública (2016)</td>
<td>53 incidences per 10000 persons</td>
</tr>
<tr>
<td></td>
<td>Water-borne disease</td>
<td>Water-borne disease incidence (including diarrhoea and leptospirosis) per 10000 persons during the period 2013–2016.</td>
<td>Ministerio de Salud Pública (2016)</td>
<td>18 incidences per 10000 persons</td>
</tr>
<tr>
<td></td>
<td>Social security affiliation</td>
<td>Fraction of the population having no public health insurance (through social security affiliation).</td>
<td>INEC Censos [57]</td>
<td>59%</td>
</tr>
<tr>
<td>Education</td>
<td>Education level</td>
<td>Fraction of population with low education level (education &lt; primary school).</td>
<td>INEC Censos [57]</td>
<td>6.2%</td>
</tr>
<tr>
<td>COPING COMPONENT</td>
<td>Infrastructure</td>
<td>Sanitation: Percentage of households with no access to piped sanitation.</td>
<td>INEC Censos [57]</td>
<td>48%</td>
</tr>
<tr>
<td></td>
<td>Drinking water access</td>
<td>Percentage of households that are not connected to the public water network.</td>
<td>INEC Censos [57]</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>Power access</td>
<td>Percentage of households with no access to public electrical network.</td>
<td>INEC Censos [57]</td>
<td>7%</td>
</tr>
<tr>
<td></td>
<td>Road travel time</td>
<td>Mean road travel time to nearest economic centres, storage infrastructure or agricultural facilities.</td>
<td>Ministerio de agricultura y Ganadería (2015)</td>
<td>60% live &gt;15min away from economic centres</td>
</tr>
<tr>
<td></td>
<td>Wall structure</td>
<td>Percentage of households with strong walls. “Strong” is defined as: concrete, bricks, adobe.</td>
<td>INEC Censos [57]</td>
<td>22%</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>Mobile access: Percentage of households with no mobile phone access.</td>
<td>INEC Censos [57]</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td>Internet access</td>
<td>Percentage of households with no internet access.</td>
<td>INEC Censos [57]</td>
<td>87%</td>
</tr>
</tbody>
</table>
the indicators’ positive directions relate to a possible increase in vulnerability.

3.1.1. Vulnerability component

The vulnerability component includes economic, social and physical variables describing the predisposition of individuals and communities to be affected or destabilised by a hazard. The variables are organised into categories related to: vulnerable groups, socio-economic, health, and education status (see Table 1).

The vulnerable groups category refers to the fraction of the population likely to suffer more and/or to need more humanitarian assistance in case of a disaster. Disability status can be an indicator of vulnerability [62] as disaster preparedness efforts often lack adequate provisions to aid those with special needs [63]. The Ecuadorian Red Cross considers disability status as a vulnerability to flooding.

The socio-economic category includes indicators measuring deprivation, inequality, and the fraction of agricultural labor within the population. The relationship between disaster and poverty is one of the most important aspects to understand vulnerability, in particular for floods [64]. To measure deprivation, we used the poverty incidence, an income-based poverty indicator created by INEC and the World Bank, estimated from the Population and Housing Census 2010 (Censo de Población y Vivienda, CPV10) and the Living Conditions Survey 2014 (Encuesta de Condiciones de Vida, ECV). In addition, a measure of inequality is included by the use of the Gini Index which was developed by the World Bank to analyse the distribution of income amongst the population. High inequality exists in Ecuador, especially in urban environments, related with economic conditions and flood impacts [56,65]. Finally, we integrate an indicator related to the agricultural dimension to represent the disaster vulnerability associated with the livelihoods of rural communities [66]. Indeed, the population depending on agriculture often suffers more from floods [67], which was also observed during a focus group discussion in rural Ecuador [68].

The health category includes indicators related to the occurrence of infectious diseases, such as vector-borne and water-borne diseases, that impact populations in relation to flooding [69,70]. These indicators are used as proxies to indicate areas of Ecuador having pre-conditions for contamination which suggest a potential increased vulnerability to post-flood infectious disease outbreak [71]. In addition, access to social protection is an important factor for disaster risk reduction assessment [72,73]. In Ecuador, a lack of public social security affiliation and reduced health access are considered proxies for poverty by disaster practitioners, as they are related to underlying conditions of vulnerability.

Finally, within the education category, we selected a variable estimating the fraction of population with an education level below primary school, to address the increased vulnerability related to the ability to understand flood risk information, early warning signs, and therefore respond to a flood emergency [74].

3.1.2. Coping capacity component

The coping capacity component refers to the resources and infrastructure available to the population to be able to cope with and mitigate the impact of disasters. It includes characteristics describing physical infrastructure and communication systems [5]. In rural Ecuador, the lack of access to certain basic services is one of the main concerns for the population [68], as it decreases the coping capacity to handle floods.

The infrastructure category includes a number of indicators such as: sanitation and drinking water infrastructure, connectivity to a public electric grid, proximity to economic centres, and structural resilience of homes. In Ecuador, appropriate sanitation and drinking water infrastructure do not exist everywhere or are inadequately constructed to withstand disasters [75], leading to an increased risk of disease and illness [76,77]. The risk is especially acute in the event of flooding as flood waters can cause surface and groundwater drinking supplies to contaminate with faeces, causing contamination of drinking water supplies [76,78]. In addition, two indicators related to the connectivity to a public electric grid and proximity to economic centres are included as proxies for remote or isolated communities [79], such as those in the Amazon region of Ecuador and parts of Coastal Ecuador [80]. According to local disaster practitioners in Ecuador, remoteness is known to impact a community’s ability to cope with extreme flood events due to reduced access to strategic infrastructures, essential relief and recovery needs, and constraints on seeking aid in the impacted area [81]. However, it is important to note that community remoteness and its impact on coping capacity can be nuanced and should be taken as context specific [79,82]. Finally, disaster practitioners indicated the proportion of households with strong walls, including those made of concrete, bricks, and adobe, as an indication of structural vulnerability in Ecuador [83–85]. Where houses’ walls are built with non-structurally robust or resilient materials, there is an increased risk of damage due to flood waters and corresponding reduced ability to cope, post-flood event [86]. When entire villages or communities are decimated, reduced coping capacity may include limited financial resources to rebuild and recover, loss of community gains and the potential for limited building material availability post-disaster [86]. In addition, the absence of safe housing post-flood can exacerbate other vulnerabilities present in the wake of a flood disaster, further reducing coping capacity.

The final category within the coping capacity component, communication, is represented by two indicators: mobile phone access and internet access. While mobile phone and internet access both play an important role in disaster preparedness in terms of disseminating and receiving warning notifications, mobile phone and internet access also contribute to the coping capacity, post-flood disaster. For example, where emergency assistance is available post disaster, mobile phones and the internet can be useful in helping to dispatch aid to those in need [87]. Additionally, where financial resources are required to aid in recovery, mobile phones can provide a mean for individuals to connect with family and friends for assistance [88].
3.2. Social vulnerability index using a PCA approach

We perform the principal component analysis (PCA) method using the Python Scikit-learn PCA tool\footnote{Using Python sklearn.decomposition.PCA.} to calculate a composite vulnerability index based on the selected indicators related to flooding in Ecuador. PCA is commonly used to reduce the dimensions of data by transforming the variables that are most correlated into separate dimensions with the first component explaining the most variance \cite{89,90}.

We use the indicators listed in Table 1 and prepare them for the PCA. The indicators are normalised,\footnote{Using Python sklearn.preprocessing.MinMaxScaler scaling.} transformed when necessary for the data to have a normal distribution \cite{91}, and then standardised\footnote{Using Python sklearn.preprocessing.StandardScaler scaling.} to obtain a mean of 0 and a deviation of 1 \cite{92}. No rotation of the data was performed. After the preparation of the variables, the PCA is performed on a chosen number of components ($n$). The results from the explained variances and eigenvalues associated with each component are presented in Table 2. From these results we explored two approaches to compute a PCA vulnerability index, at the Parroquia spatial unit $k$.

$$PCA\ _{vulnerability\ _{index}}^{k} = \sum_{i=1}^{n} \eta_{i}^{2} \ast PC_{ki}$$

(1)

The first approach uses the first principal component alone ($n = 1$), accounting for 36% of the variance explained, to compute a final social vulnerability outcome map from the normalised first principal component transformed matrix ($PC_{k1}$). The second approach integrates all components with an eigenvalue close to or higher than 1 \cite{93,94} - in our case the 5 first components ($n = 5$) - into a PCA vulnerability index. The calculation is based on the aggregation of the normalised transformed matrix ($PC_{ki}$) of each principal component $i$, weighed by the corresponding percentage of variance explained ($\eta_{i}^{2}$), for individual Parroquia $k$ \cite{21,22,95}, as described in Equation (1). This method presents an alternative to the equal weighting of components proposed by Cutter et al. \cite{6}; and accounts for the relative importance of each principal component into the final composite index. The result is then normalised to a final PCA vulnerability index using a min-max scalar function \cite{92,95–97}.

3.3. Social vulnerability index using expert judgement

Through frequent and structured working sessions with ERC, we co-developed an understanding of the vulnerability of populations to flooding based on expert knowledge. The co-authors held a workshop with the participation of 13 Ecuadorian disaster practitioners representing sectors of health and community development, the disaster risk management team in charge of response and reduction lines, zonal coordinators of territory, specialists of information management and monitoring of adverse events, in addition to officials from the National Institute for Meteorology and Hydrology of Ecuador (INAMHI). Through the workshop, the participants developed their own weights of variables important for understanding social vulnerability to flooding. Guided by the INFORM index best practices for the humanitarian sector, a hierarchy is applied to the variables through the two Vulnerability and Coping capacity components. The process of assigning weights from practitioners, hereafter referred to expert judgement on the weighting of variables, is developed as follows.

1. The weights are first estimated individually by each practitioner. Each indicator is given a score from 0 to 3 points according to their perceived importance within the two components Vulnerability and Coping capacity. The same approach is followed to assign a score from 0 to 3 to the two Vulnerability and Coping capacity components. We use statistical violin plots to analyse the variance of the individual inputs from practitioners’ weightings of variables as a quality control of the results.

\begin{table}
\centering
\caption{PCA percentage of variance explained and Eigenvalues for each component.}
\begin{tabular}{|c|c|c|c|}
\hline
Number of component ($n$) & \% variance explain ($\eta_{i}^{2}$) & cum \% explain & Eigenvalue \\
\hline
1 & 0.361 & 0.361 & 5.423 \\
2 & 0.134 & 0.495 & 2.010 \\
3 & 0.083 & 0.578 & 1.251 \\
4 & 0.071 & 0.650 & 1.071 \\
5 & 0.066 & 0.716 & 0.989 \\
6 & 0.057 & 0.772 & 0.850 \\
7 & 0.046 & 0.818 & 0.690 \\
8 & 0.034 & 0.852 & 0.514 \\
9 & 0.032 & 0.884 & 0.473 \\
10 & 0.028 & 0.912 & 0.424 \\
11 & 0.021 & 0.934 & 0.321 \\
12 & 0.018 & 0.952 & 0.274 \\
13 & 0.017 & 0.969 & 0.259 \\
14 & 0.016 & 0.985 & 0.235 \\
15 & 0.015 & 1.000 & 0.229 \\
\hline
\end{tabular}
\end{table}
2. The individual scores of all participants are then averaged into a collective level for each vulnerability variable, and for the two components. The two components’ averaged scores are first transformed into loads in percent. The final weights of each vulnerability variable are calculated per component, using the fraction of the averaged scores of each variable over the total component score, weighted by the components’ loads.

We then integrate the outcomes of the collective variable weighting into a composite social vulnerability index, hereafter named Expert vulnerability index. To do so, the 15 normalised input variables are aggregated following the most commonly used weighted arithmetic averaged scheme (Backer et al., 2017), presented in Equation (2).

\[ \text{Expert vulnerability index}_{ki} = \sum_{j=1}^{d} (w_j \times x_{kj}) \]  

Where \( d \) represents the total number of aggregated variables \( j \) (in our case \( d = 15 \)), \( w_j \) is the nominal weight outcome assigned to the \( j \)-th variable, and \( x_{kj} \) is the normalised variable value of individual Parroquia \( k \).

3.4. Comparison between PCA and expert vulnerability indices

We calculate, visualise and map the similarities and differences between the results of the PCA and Expert vulnerability indices,
looking at the differences in output distribution, in the resulting attributed contribution of each individual variable and the spatial differences between the composite vulnerability index outputs. To do so, we ranked the loading factors from the first principal component analysis, as well as the resulting collective weights from the Expert Method.

3.5. Correlating indicators and indexes with impact data

We use the historical flood occurrence and impact dataset built in Ecuador [98, 99] from DesInventar and SNGRE (Servicio Nacional de Gestión de Riesgos y Emergencias de Ecuador) - the Ecuadorian Service for Risk and Emergency Management - data sources. It contains 3284 flood events for the period 2010–2019, with information about the estimated number of people affected (both directly and indirectly). The average number of affected people per flood event is calculated for each of the 519 Parroquia where impact data is available. Results are correlated with the different outcomes from the social vulnerability indexes generated for this study, as well as with the results from each social vulnerability variable.

Table 3

<table>
<thead>
<tr>
<th>Indicators</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pov_incidence</td>
<td>−0.352</td>
<td>0.098</td>
<td>0.071</td>
<td>−0.172</td>
<td>0.123</td>
</tr>
<tr>
<td>decreased_sanitation</td>
<td>−0.338</td>
<td>−0.014</td>
<td>−0.204</td>
<td>0.242</td>
<td>−0.229</td>
</tr>
<tr>
<td>Low_Drinking_water_access</td>
<td>−0.335</td>
<td>0.101</td>
<td>−0.089</td>
<td>0.162</td>
<td>−0.310</td>
</tr>
<tr>
<td>Low_Power_access</td>
<td>−0.334</td>
<td>−0.218</td>
<td>0.088</td>
<td>−0.003</td>
<td>0.085</td>
</tr>
<tr>
<td>Agricultural_labor</td>
<td>−0.319</td>
<td>−0.051</td>
<td>0.027</td>
<td>−0.244</td>
<td>0.237</td>
</tr>
<tr>
<td>Low_Mobile_access</td>
<td>−0.318</td>
<td>−0.266</td>
<td>0.020</td>
<td>0.038</td>
<td>0.029</td>
</tr>
<tr>
<td>Road_travel_time</td>
<td>−0.286</td>
<td>0.394</td>
<td>0.047</td>
<td>−0.123</td>
<td>−0.038</td>
</tr>
<tr>
<td>weak_wall</td>
<td>−0.277</td>
<td>−0.154</td>
<td>−0.282</td>
<td>−0.057</td>
<td>0.140</td>
</tr>
<tr>
<td>Low_SocialSecurity</td>
<td>−0.258</td>
<td>−0.025</td>
<td>0.513</td>
<td>−0.159</td>
<td>0.043</td>
</tr>
<tr>
<td>Disability</td>
<td>−0.254</td>
<td>0.414</td>
<td>0.216</td>
<td>0.052</td>
<td>−0.221</td>
</tr>
<tr>
<td>Waterborne</td>
<td>−0.137</td>
<td>0.150</td>
<td>−0.600</td>
<td>−0.148</td>
<td>0.411</td>
</tr>
<tr>
<td>Vectorborne</td>
<td>−0.069</td>
<td>−0.494</td>
<td>0.282</td>
<td>0.099</td>
<td>0.051</td>
</tr>
<tr>
<td>Gini_Index</td>
<td>−0.030</td>
<td>0.290</td>
<td>0.001</td>
<td>0.578</td>
<td>−0.042</td>
</tr>
<tr>
<td>Low_SocialSecurity</td>
<td>−0.027</td>
<td>0.069</td>
<td>0.299</td>
<td>0.538</td>
<td>0.686</td>
</tr>
<tr>
<td>Road_travel_time</td>
<td>0.148</td>
<td>0.383</td>
<td>0.247</td>
<td>−0.351</td>
<td>0.238</td>
</tr>
</tbody>
</table>
4. Results

4.1. Initial results and exploratory analysis

4.1.1. Indicator mapping

Nine selected indicators are presented in Fig. 2 (all 15 indicators can be found in supplementary materials). Indicator mapping reveals different vulnerability contexts. While some indicators such as the poverty incidence or the fraction of weak walls clearly highlight the vulnerability of the population in the Amazon or the northwest coastal areas, other indicators, such as the share of agricultural labor do not indicate significant regional gradients, and are characterised by more local variability in their distribution across the country.

4.1.2. Indicator correlation analysis

The correlation coefficient matrix of the 15 selected indicators is presented in Fig. 3. The selection is the result from elimination of indicators presenting too high co-dependencies (e.g., analphabesism was highly correlated to education rate) or not relevant to the Ecuador context according to disaster practitioners (e.g., gender or birth rate). It reveals that most indicators have a positive correlation with each other. Most of the negative correlations are associated with the Gini Index, presenting negative correlations with 10 of the 15 variables. On the contrary, the poverty incidence has the most occurrences of correlation to all other variables, showing correlation coefficients above 0.5 with 10 of the 15 variables. The highest correlations to poverty incidence are observed with the reduced access to mobile network (0.61), and power access (0.6). Interestingly, the highest observed correlation coefficients are not linked to poverty, but are the results of the correlation between i) the decreased sanitation condition with the reduced access to drinking water (0.73); ii) the reduced access to power networks with the increased weakness of walls (0.69); and iii) the increased share of agricultural labor with the reduced access to internet (0.69).

4.1.3. Outcomes of PCA-driven method

Table 3 presents the loadings factors of the 15 socio-economic variables, for each of the 5 first principal components, as a result from the PCA. We note the importance of the poverty incidence indicator, showing the highest absolute load within the first principal component. All first principal component loads have the same direction (highly negative values reflecting the indicators with increased contribution to the vulnerability) except for the Gini Index. This confirms that increased inequality (higher Gini Index values) does not seem to correlate with all the other social vulnerability variables. Indeed, in Ecuador, Parroquias with higher inequality are the most urbanised, which have the lowest average poverty incidence and the best infrastructure.

The two vulnerability index outputs, generated using the first principal component only PCA(n=1) and the first five principal components PCA(n=5) are presented in the two maps of Fig. 4. In both output maps, the Amazon region presents the highest social vulnerability to flooding, whereas the Major urban areas, Guayaquil and Quito, are attributed the lowest social vulnerability. Yet, the output of PCA(n=1) presents a more contrasted spatial pattern, while the output of PCA(n=5) reveals more local variability between Parroquias. Indeed, PCA(n=1) results show high vulnerability clustered in the Amazon and in the northwest coast, and lower vulnerability in the rest of the country than from PCA(n=5).

4.1.4. Outcomes of expert weighting method

The results from the expert weighting of each variable are presented in Fig. 5. It illustrates the data distribution of the individual score assigned by the 13 participants to each variable, with violin plots (including estimated Kernel density plots and the inner box plots), as well as the collective level represented by the mean values as red dots. The participants’ answer distributions are generally

![Fig. 4.](image-url) Flood-specific composite social vulnerability outputs from the PCA analysis, using the first (left) and the first five (right) components, at the Parroquia level of Ecuador.
wide and skewed toward higher score values in the assignment of individual and collective scores, confirming the importance of considering all selected vulnerability variables in the construction of a flood-specific social vulnerability index. Interestingly, the three variables ranked by experts as the most important for social vulnerability to flooding - poverty, weak walls and road travel time - are the ones showing the narrowest distribution within the practitioners’ weighting answers, and therefore the highest confidence in the results. On the contrary, as a result of the averaging method from individual to collective results, the variables with the wider individual distribution of answers are more likely to have a lower collective rank.

The mean scores from each variable are transformed into final weight percentages within each component, using the “vulnerability” component loads of 53.85% and the “lack of coping capacity load” of 46.15%, also estimated from individual practitioner results. The results of final collective variables ranking are presented in Table 4. The output from the social vulnerability index built from the collective results from expert weighting is presented in Fig. 6.

4.2. Comparing PCA-driven with humanitarian practitioners socio-economic vulnerability outputs

4.2.1. Comparing the resulting ranking of the indicators from each method

The ranking of the contribution of each variable is presented in Table 4, calculated from the first principal component analysis factor loadings, as well as from the disaster practitioner experts’ collective weights. The comparison of the ranking results with the two approaches is illustrated in a separated column. The ranking difference represents the difference between the variables’ ranking with PCA(n=1) and Expert methods. Negative values indicate variables for which the contribution to social vulnerability was attributed higher by the Expert than by the PCA approach.

The analysis of the attribution of variable’s ranks from the two approaches reveals quite some differences. Other than the poverty incidence variable, revealed to be the most important indicator for vulnerability assessment by both methods, we observe major differences in the variable ranking contribution to the composite index. For instance, while the PCA approach attributes very low

Table 4

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Ranking PCA(n=1)</th>
<th>Ranking Expert</th>
<th>Ranking difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pov_incidence</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>decreased_sanitation</td>
<td>2</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>Low_Drinking_water_access</td>
<td>3</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Low_Internet_access</td>
<td>4</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>Low_Mobile_access</td>
<td>5</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>Agricultural_labor</td>
<td>6</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Low_Power_access</td>
<td>7</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>LowEd_rate</td>
<td>8</td>
<td>12</td>
<td>4</td>
</tr>
<tr>
<td>Road_travel_time</td>
<td>9</td>
<td>3</td>
<td>-6</td>
</tr>
<tr>
<td>weak_wall</td>
<td>10</td>
<td>2</td>
<td>-8</td>
</tr>
<tr>
<td>Low_SocialSecurity</td>
<td>11</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Disability</td>
<td>12</td>
<td>4</td>
<td>-8</td>
</tr>
<tr>
<td>Vectorborne</td>
<td>13</td>
<td>9</td>
<td>-4</td>
</tr>
<tr>
<td>Waterborne</td>
<td>14</td>
<td>5</td>
<td>-9</td>
</tr>
<tr>
<td>Gini_Index</td>
<td>15</td>
<td>6</td>
<td>-9</td>
</tr>
</tbody>
</table>
contribution to variables such as the disability, and vector/water-borne disease variables, due to their lower dependency to the other variables, these variables are perceived more important by the expert judgement of the practitioners. On the contrary, the variables related to communication (internet and mobile access) are ranked high, in the 4th and 5th positions, by the PCA-driven method while they are considered as less critical by the disaster practitioners.

4.2.2. Comparing outputs of vulnerability indices

The social vulnerability indices output results derived from PCA (Fig. 4) and expert judgement (Fig. 6) show similar trends toward higher vulnerability in the Amazon region, and lower vulnerability in the most urbanised regions. However, significant divergences are revealed by the comparison of the social vulnerability outputs. The distribution of the social vulnerability outputs estimated by experts versus the two PCA methods are presented as density plots in Fig. 7 for Ecuador as a whole as well as for the Coastal, Andes, and Amazon sub-regions.

The comparison between the vulnerability indices output for each Parroquia is shown as density scatter plots in Fig. 8. The percentage difference of the Expert judgement index compared to the PCA indices for each Parroquia (Equation (3)) is used to represent the magnitude of the difference between the indices. The difference results are illustrated for each Parroquia as coloured values from blue to red in Figs. 8 and 9. Blue colours indicate areas where the estimation of the social vulnerability using Expert index is lower than using the PCA indexes. Red colours indicate areas where the Expert vulnerability index is higher than the PCA-driven indexes.

\[
\text{Percentage difference} = \frac{(\text{PCA index} - \text{Expert index}) \times 100}{(\text{PCA index} + \text{Expert index}) / 2}
\]

We observe significant differences between the different social vulnerability outcomes. First, the Expert vulnerability index values tend to systematically be lower than the PCA-driven vulnerability indices. This is especially true when comparing the Expert with PCA (n=5) outcomes. The difference between the Expert and PCA(n=5) social vulnerability results show a wider dispersion than the difference between Expert and PCA(n=1) results. The PCA(n=5) outcomes reveal an increased number of Parroquia with significant differences with the Expert results, including some Parroquia showing opposite social vulnerability index values. These are important results to highlight specific areas with increased index sensitivity to the construction method and therefore potential higher uncertainty on the social vulnerability assessment.

In addition, we observe a trend in how the social vulnerability outputs differ spatially. In the Amazon and the northwest coastal regions, as well as the urbanised areas, regions for which most indicators tend respectively toward high and low relative social vulnerability, the PCA methods assign lower vulnerability than the Expert judgement (red colours in Figs. 8 and 9). On the contrary, in the rest of Ecuador with more averaged social vulnerability estimation, particularly in the Andes, both PCA methods assign higher vulnerability than the Expert judgement.

These trends are confirmed by the comparison of the output distribution for each sub-region (Fig. 7). Further, these results indicate that, in the Andes, both PCA outcomes result in wider distribution than the Expert results, leading to more local heterogeneity in the social vulnerability results in this area. On the contrary, in the Coastal region, the PCA outcomes distribution are narrower than the Expert outcomes.
4.3. Correlating indicators and indexes with impact data

Fig. 10 presents the average number of affected people per flood event calculated for the 519 Parroquia where the impact data is available.

Results from the correlation between impacts and the different outcomes from the social vulnerability indexes generated for this study are presented in Table 5. It reveals that, even though the correlations between the social vulnerability indices and historical flood impacts at Parroquia level are small and similar from one method to the other, the values are slightly higher for the two methods using expert weighting and PCA (n=1) than for the PCA (n=5).

Fig. 11 presents the correlation analysis between the flood impact and the results from each social vulnerability variable. It shows that the individual variables correlating the most with the average number of affected people per flood event do not include the poverty incidence. The four variables most correlated to flood impacts belong to the “coping capacity” component, suggesting a strong link between post disaster social vulnerability and impacts. The variable with the highest correlation with flood impacts is the fraction of houses with weak walls, highlighting the importance of the structural vulnerability. While the structure of the walls is ranked by experts second in the social vulnerability assessment to flooding, the PCA approach does not capture the importance of this variable. On the contrary, the two other variables showing highest correlation coefficients with flood impacts - the reduced access to drinking water and decreased sanitation - are addressed as the most contributing to the social vulnerability to flooding in Ecuador by the PCA approach, but not by the experts. Finally, the variables with negative correlations with impact data - Water-borne, Disability and Gini Index - are confirmed with the lowest contribution these have in the PCA-driven social vulnerability index.
5. Discussion

This research contributes to the study of composite social vulnerability indices related to natural hazards, that can be used as part of decision-making processes by providing a context on socio-economic conditions to inform actions [100]. Specifically for this study, the social vulnerability index is intended to be combined with hazard and exposure information to inform flash flood anticipatory actions and prioritisation at the administrative level 3 (Parroquia) in Ecuador. As underlying vulnerability components and context vary considerably within a country [101] and there is local spatial variability in the factors that drive social vulnerability [102], sub-national details are generally critical for disaster practitioners ([103]. Multi-vulnerability analysis is key for flood risk management [104], however, there is no universally accepted method in the variable selection and the generation of social vulnerability indices [28], and the outputs are not “calibrable” to observed data. Inspired by the SoVI® concepts, most vulnerability indices used in developing countries are based on variable-specific weighting from experts or PCA techniques without an evaluation of the outcomes [34,36,105].

This study reveals divergence among experts on variable weightings, and divergence in results from different construction approaches, which might question the use of vulnerability indexes for decision making without undertaking validation and sensitivity analysis. However, we argue that our results help with the quantification and understanding of the reasons for spatial differences in flood-specific social vulnerability [20]. In addition, as the challenge to evaluate the goodness and the fitness for use of these indexes remains [8], we implement a way to use historical flood impact data to help evaluate vulnerability indices, and review the relevance and contribution of each indicator.

Fig. 8. Comparison of the results from the social vulnerability indices calculated by experts’ judgement versus PCA approaches, using the first (left) and the five first components (right) methods, for each Parroquia of Ecuador. The coloured scale corresponds to the percentage difference between the two resulting social vulnerability outcomes, calculated for each Parroquia.

Fig. 9. Difference maps in percentage comparing the Expert with the two PCA vulnerability indices results at Parroquia level.
Our research provides information that will likely be valuable to disaster practitioners interested in risk reduction and anticipatory action, by highlighting the most challenging or conflicting areas for the social vulnerability assessment. Below we highlight six key considerations for developing flood-specific social vulnerability indices for disaster management.

5.1. Sensitivities in method selection

Deductive methods using PCA and hierarchical approaches involving expert knowledge for indicators weighting are the most commonly used approaches for the generation of composite social vulnerability indices [38]. However, both methods present some limitations. The weightings from expert judgement are often perceived as a weak form of a deductive argument that should only be used for indicator selection [106]. While disagreement amongst experts on the appropriate weighting is common as disaster practitioners’ judgement involves subjectivity [104,107], they however have knowledge on what the characteristics of social vulnerability are in a country-specific or disaster-specific context. On the other hand, the PCA is mainly aiming at explaining the data relationship,
and generating a social vulnerability dimension of what we expect represents social vulnerability. While there are no completely good or bad methods to generate composite vulnerability indices [27], our results show that the choice of methods matters in the final outputs and can considerably influence local disaster management decisions. This is especially true when using multiple components in the PCA analysis, resulting in more extreme differences between the outputs. For instance, the Parroquia of El Santiago in the Amazon is assigned a high social vulnerability index of 0.85 by Expert and a low index of 0.18 by the PCA(n=5) method.

The differences in social vulnerability outputs show a spatial pattern, and are not randomly distributed. Social vulnerability indexes using PCA show systematically higher vulnerability in the highlands and in the southwest coastal area than the social vulnerability index calculated using expert knowledge. On the contrary, the Amazon, northwest coast and urban areas will tend to get higher prioritisation based on Expert rather than PCA-driven social vulnerability results. As a result, depending on how the social vulnerability index is used by disaster practitioners, using for instance Expert instead of PCA method could lead to lower or higher prioritisation actions in these regions.

5.2. The choice and contribution of indicators

As important as the choice of methods, which indicator to select and how to classify or group them matter for the composite output. For instance, the outcomes of broad social vulnerability indices are more variable as the selection of indicators is not constrained by their relevance to a hazard or location. It is possible to group and classify variables according to multiple schemes influencing the weighting and composite outcomes. In this study, we opt for a simple classification with two components, which is commonly adopted by humanitarian practitioners, in which we consider specific indicators related to the momentum of disaster; pre disaster with the “vulnerability” component, and “during-post” disaster vulnerability with the “coping capacity” component. This suggests that, potentially, one could use the results of each component to characterise and contrast the pre and post disaster social vulnerability.

Further, our results show that, even with the same indicator selection and classification, different methods attribute very different contributions for each variable, leading to different outcomes. In our study, only the poverty incidence is attributed the same contribution, and the variables responsible for the main deviation in the results are disability, the vector and water-borne disease incidence, as well as the variables related to the communication categories. This can provide information to disaster practitioners about the sensitivity related to specific variables and, depending on the regional context, support the decision on the final contribution of these variables to the vulnerability assessment.

Finally, the concern of the influence of the poverty incidence variable on the composite social vulnerability results should be raised. There is ample agreement that poor people are disproportionately affected by climate (e.g., floods, drought) and other environmental disasters due to exposure, proportionally larger economic losses (as a fraction of their wealth) and lack of coping capacity, and poverty is both a driver and a result of environmental impacts [108,109]. However, as the poverty incidence variable shows significant correlation with most of the variables, and the maximum contribution to the composite social vulnerability indices, it could strongly influence the outcomes. To test this, we removed the poverty incidence indicator from the input variables and observed no significant changes in the outcomes of the Expert and PCA(n=1) composite social vulnerability indices. However, this led to important reduction of the social vulnerability estimation in the Amazon when using PCA with multiple components.

5.3. Evaluation with impact data

Research to evaluate [10,11,13] and calibrate [12] social vulnerability indexes using a disaster outcome matrix such as those based on historical flood impact data have been conducted in data-rich countries, where high quality and high spatial resolution hazard and impact data are available. While comprehensive impact data is not available for most of the world, Ecuador fortunately has extensive historical disaster records.

In the absence of the hazard component, here we compare how the different vulnerability indexes as well as the individual vulnerability variables are generated, and correlate those with historical flood impact data aggregated at the Parroquia level of Ecuador. In future work, we aim to quantify, for the same hazard magnitude, the influence of the socio-economic factors on the resulting impacts.

We understand the limitations of correlating social vulnerability indicators with averaged flood impacts occurring in the past in each Parroquia, without taking into account the different flood event magnitude, as well as the number of affected populations. We argue that the correlation between the average number of people affected per flood and the population density of each Parroquia is low enough (R = −0.076) to compare how each social vulnerability indicator and composite indices contribute to the flood impacts.

Vulnerability index outcomes can vary depending on methods used, therefore we suggest that results from correlation with historical flood impact data can help practitioners differentiate which method would better benefit impact-based forecast applications. While we observed an overall low correlation between the constructed indexes and historical flood impacts, the correlation between impact data and individual variables can help to better understand the relative importance of each variable. The contribution of some important variables (e.g., wall structure) were undervalued with the PCA approach, revealing the importance of involving expert judgement in the variable weighting process.

5.4. Methodological considerations

Spatial considerations are key for social vulnerability assessments. We recognize that, as the concept of non-stationarity captures, what makes people vulnerable differs spatially. Therefore, composite social vulnerability indices could be generated at different spatial units, leading for instance to different indicator contributions and outputs for each sub-region of Ecuador. Indeed, what spatial unit to consider for the social vulnerability assessment matters. While aggregated spatial units from the census may provide the most accurate data characterising the population, it represents the average values and does not capture heterogeneity within the census units. Urban
areas have inequalities, as demonstrated by the Gini Index, that are not captured unless higher resolution layers are used to indicate the most vulnerable within the spatial unit. While our social vulnerability outputs reflect the vulnerability of each Parroquia in conformity with the aggregated census data, higher resolution local social vulnerability layers could give additional insight [110,111] while also introducing potential errors that need to be assessed before being used for decision making.

Further, we acknowledge the following limitations of our current approach. First, some of the results are specific to the context of flooding in Ecuador, and might not be directly applicable to other countries and datasets. In addition, we worked within the context of the in-country data availability and therefore realise that not all important social vulnerability factors could be included. Finally, we limited our analysis to the comparison of only two methods to build composite indices, and acknowledge that inclusion of other statistical approaches could complement our analysis.

5.5. Implication for disaster practitioners

With this comparison study, we aim to provide guidance for disaster practitioners that can be applied in most developing countries that collect and produce social data at a sub-national level. Based on our results, we recommend disaster practitioners to consider using vulnerability indexes with particular attention, to avoid the possible risks of inadequate allocation of resources before, during and after disasters due to overreliance on indexes. Composite indices can be falsely perceived as highly robust due to their quantitative nature [112], leading to discount qualitative evidence when assessing vulnerability. A critical analysis of the choice of methods and social vulnerability methods and outcomes used for decision making can help prevent potential differences in data-driven decisions linked to the use of one product to another.

- We show the value of generating composite social vulnerability index using more than one method. Doing so, disaster practitioners could compare the results and highlight the indicators and areas showing the highest sensitivity within the social vulnerability assessment.
- We do not recommend starting with more than one component in the PCA analysis, or only with the objective to highlight administrative units with more complex social vulnerability contexts, to be addressed with local knowledge.
- However, we suggest that adding a simple correlation with historical impact data can help to better understand the relative importance in the contribution of each indicator. The integration of all results could be combined into a ‘PCA and Impact’ driven expert weighting approach, for instance by analysing the results from the PCA and impact correlations with disaster experts, and integrating the observation into the exercise of variable weighting.
- Finally, to account for the different skewness of output distributions, it is recommended to define actions based on the use of quantile-based instead of value-based thresholds of social vulnerability indices.

In the case of intending to develop a sustainable service driven by the index, governance structures should be considered so that each the development, implementation and review (and potential update and/or revision) of the index can be conducted in a structured and transparent approach.

6. Conclusion

This study demonstrates the need for more significant reflection on index construction as vulnerability assessments are increasingly incorporated into the decision-making process. It is especially important to increase awareness of the limitations (and opportunities) of indexes by the decision makers seeking to use them. We particularly recommend disaster practitioners to use composite vulnerability indexes with particular attention without undertaking validation and sensitivity analysis, to avoid the possible risks of inadequate allocation of resources.

While different approaches in the social vulnerability index construction lead to similar spatial trends toward higher social vulnerability in the Amazon and lower vulnerability in the urban areas, major differences are observed at the Parroquia level, potentially leading to different decision and prioritisation actions in the case of flood events. Indeed, with the exception of the poverty variable, the two methods (PCA and Expert) attribute very different contributions to each individual variable, which are responsible for the major differences in the social vulnerability outcomes. In addition, the PCA approaches tend to generate outcomes with skewed distributions toward higher social vulnerability values, as the complexity increases in the number of PCA principal component dimensions included in the composite index.

This study reveals that capturing social vulnerability is more challenging in areas where indices result in values closer to the mean of the distribution. This is the case of Ecuador Highlands and southwest coast, where the results from the Expert index are lower and more homogeneously distributed than the results from the PCA-driven indices. In those areas, the PCA outcomes express more variability and potentially better capture some of the combinations of factors leading to relative changes in local social vulnerability. Further, the PCA(n=5) results reveal an increased number of Parroquía with significant differences with the Expert results, including some Parroquia showing social vulnerability index values on the opposite end of the spectrum. These results allow users to highlight areas with higher sensitivities in the social vulnerability assessment outcomes.

Even though it is not possible to conclusively define one approach being superior than the other, in our case the results from the correlation with historical flood impact data show that PCA(n=5) social vulnerability outputs have less correlation with impact data than the results from PCA(n=1) and Expert method. Further work is needed to explore correlations between impact, hazard and vulnerability. However, the correlation between historical flood impact data and individual variables already helps us to better understand the relative importance of each variable. As an example, the contribution of some important variables (e.g. wall structure) to flood impact is not honoured by the PCA approach, revealing the importance of involving expert judgement in the variable weighting.
process.

Data availability

The datasets resulting from the current flood-specific social vulnerability analysis of Ecuador, as well as the historical flood impact data are published on Zenodo.

The social vulnerability input variables matrix and the social vulnerability indices outcomes of each Parroquia in Ecuador are openly available. The reference to the most updated version of the dataset is stated hereafter.


The historical dataset of flood events and impact built for Ecuador is not openly available. It has restricted access because it contains data that are the property of the Ecuadorian Service for Risk and Emergency Management (SNGRE, Servicio Nacional de Gestión de Riesgos y Emergencias de Ecuador). The reference to the most updated version of the dataset is stated hereafter.


Supplementary material has been provided.

Funding

This research was funded by the National Aeronautics and Space Administration, grant number 80NSSC18K0342.

Author contribution


Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank the Cruz Roja Ecuatoriana (Ecuadorian Red Cross) for the close collaboration in this research, supporting the NASA GEO project “Towards A Global Flood & Flash Flood Early Warning Early Action System Driven by NASA Earth Observations and Hydrologic Models.”

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijdrr.2022.102897.

References