The impact of data spatial resolution on flood vulnerability assessment

D. Morrison a, L. Beevers a, G. Wright a and M. D. Stewart b

a School of Energy, Geosciences, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK; b Kaya Consulting Limited, Edinburgh, UK

ABSTRACT

Index-based approaches are a popular method for assessing societal vulnerability to flooding, many of which differ in terms of indicator selection, underlying social data, spatial scale and aggregation methods. They are typically assessed at geographically broad spatial scales to provide a spatial picture of vulnerability for policy and decision-makers. However, aggregation of vulnerability at broad scales also potentially masks the true vulnerability of an area as the underlying data is not spatially refined. This research expands on a previous indicator approach, the Social Flood Vulnerability Index by using geodemographics to facilitate household and postcode level vulnerability assessment to explore the impact of spatial aggregation on vulnerability at national and local levels in Scotland. The results suggest that applying geodemographics to an existing approach increases spatial heterogeneity and has the potential to be adopted as a new dataset to guide indicator selection in future.

ARTICLE HISTORY

Received 23 May 2020
Accepted 29 March 2021

KEYWORDS

Geodemographics; social vulnerability index; flooding; flood risk management

1. Introduction

As well as being an area of active academic research, vulnerability measures are increasingly used when assessing the benefits of flood management measures, to move away from traditional cost–benefit analyses that focus on physical damages to property and commerce. Methods of estimating vulnerability to flooding are typically categorised as indicator methods (Balica et al., 2013; Lindley et al., 2011; Lindley & O’Neill, 2013; Mwale et al., 2015), vulnerability curves (Penning-Rowsell et al., 2005), disaster loss data methods (Dilley et al., 2005) and modelling approaches (Beevers et al., 2016). Indicator methods are amongst the most widely used in Flood Risk Management (Connor & Hiroki, 2005), and they consider a range of social factors considered to be indicative of vulnerability to flooding (Rufat et al., 2015); such as demographic characteristics (e.g. age, ethnicity, gender), socio-economic status (e.g. income, wealth, education, occupation), land tenure (e.g. homeowners/renters) and health (e.g. long term-illnesses, disabilities).
Despite their popularity, indicator methods are limited by the spatial resolution of input data and the infrequent gathering of this data, which results in issues related to the standardisation, weighting and aggregation methods used in the calculation of the indicators (Nasiri et al., 2016). Moreover, there is a distinct lack of research surrounding high-spatial resolution index outputs (Anderson et al., 2019). In the U.K., the key social dataset is Census data, which is limited in terms of the frequency (updated every 10 years), the variables considered and the spatial resolution of the data, which is aggregated to ‘output areas’. These are the smallest geographical reporting unit in the data and typically contain 40–125 households. These limitations are common to other national Census datasets; for example, Canada’s smallest reporting units account for 400–700 people (Statistics Canada, 2016). It is notable that this data reflects specific geographical areas unrelated to hazard exposure. Therefore, the only vulnerability of the community in that area can be assessed, without reference to the boundary of exposure to an assessed natural hazard. The problem of spatial resolution is not unfamiliar in indicator-based vulnerability assessments; Werritty et al. (2007) highlight that the ability to accurately profile socio-demographic flood vulnerability is dependent on successfully matching spatial units to areas at risk of flooding.

The Social Flood Vulnerability Index (SFVI) (Tapsell et al., 2002) and the Flood Disadvantage framework (see Lindley et al., 2011; Lindley & O’Neill, 2013; Scottish Government, 2015) are established indicator sets for measuring the flood vulnerability. However, both are limited by the base social data, with the former measuring vulnerability at Census output areas in England, and the latter at data zone level in Scotland (a larger spatial area than output areas covering approx. 500–1000 households). Whilst this level of aggregation is useful at a strategic level, it can inherently constrain local-scale risk assessments as it assumes homogeneous levels of vulnerability over a broad area which can mask the existing vulnerability at the micro-scale (Alexander et al., 2011; The Scottish Government, 2015).

Datasets that are more granular than Census are available, and neighbourhood classification systems (geodemographics) have been recently used in marketing applications, as they enable diverse classification of consumer marketing (e.g. ACORN, Experian Mosaic). They are more current and spatially discretised (postcode and household level) than Census datasets. One of these datasets, Experian Mosaic (hereafter called Mosaic) was used for indicator-based vulnerability assessments by Tomlinson et al. (2011) to demonstrate a range of vulnerabilities to the urban heat island effect at the neighbourhood scale. Willis et al. (2014) used Mosaic Italy data to determine the most vulnerable neighbourhood types to volcano eruption in Naples, whilst Fitton et al. (2018) used Mosaic data at postcode level to conduct a national-scale assessment of coastal erosion risk in Scotland.

The ultimate aim of the research reported herein is to explore the viability and added value of high-resolution datasets in flood vulnerability assessments. The research will explore the benefits and disbenefits afforded by using different underlying data for assessment of vulnerability at different spatial scales to explore flood vulnerability at a variety of scales across Scotland (U.K.). By using a well-known vulnerability index compatible with Census data (SFVI), the study assesses the potential for Mosaic to be used as an alternative data source by analysing and comparing the impact of data aggregation on final vulnerability scores. If the Mosaic can be shown to be successful in flood vulnerability
assessment, it offers the potential for further work that can develop alternative flood vulnerability indicators that are suited for such high-resolution datasets.

2. Data and methods

The assessment calculates the SFVI using Census and Mosaic data at the national scale (Scotland) and in detail for two Scottish communities, one a small rural town impacted by river (fluvial) flooding and the other an urban area impacted by surface water (pluvial) flooding. For both national and local analysis, the population is separated into those that are both at-risk and not at-risk of flooding (whole population) and those that are (exposed population). Mosaic data is considered at household and postcode levels to allow a direct comparison with the Census output area data. The exposed population analysis considers both fluvial and surface water flooding.

2.1. Input data

The two main sources of data are household and postcode level Mosaic 2019 data provided by Experian, and Census 2011 data for output areas across Scotland (see Scottish Government, 2011). Mosaic is a derived data product that classifies households and neighbourhoods into 57 household typologies and 14 lifestyle groups, using a total of 332 data elements from a range of sources, including electoral roll, council tax and lifestyle surveys conducted by Experian. Data is clustered using in-house k-means methods which provide a statistical representation of consumer behaviour and characteristics across households and postcodes (Experian, 2019). The U.K. Census is similar in that it aims to define how the population live (through household surveys) based on socio-demographic characteristics to aid with policy-making, planning and running of services and allocation of public funds (Scottish Government, 2015); there is an overlap between both datasets, as 28% of Mosaic information is sourced from 2011 Census data. For example, both gather information related to education, population and households, transport, health, housing, ethnicity, identity, language and religion. However, both datasets diverge in terms of spatial discretisation and characterisation. Census information is gathered by household, aggregated and published at output area level, whereas Mosaic data is available at higher resolution (postcode and household). Moreover, a more detailed picture of the population is delivered by Mosaic as households are characterised into unique social typologies. Another key difference is that Mosaic is updated bi-annually, whereas Census is updated once every 10 years, with the next Census release expected in 2021 (Experian, 2019; NRS, 2018).

The indicators used for measuring vulnerability are those defined in the SFVI method (Tapsell et al., 2002). Whilst more comprehensive indicator sets are available, the SFVI method is employed as it is a relatively simple and well-recognised approach, which has had previous applications in Scotland (SEPA, 2011). The SFVI uses inputs that can be directly extracted from Census data and has like-for-like inputs within the Mosaic variables. Therefore, it offers an opportunity to explicitly measure the impact of spatial aggregation without the use of more complicated indicators or further analysis of the correlation between variables in both datasets.
2.1.1 Uncertainty in the underlying data

This study uses National ‘hard edge’ hazard maps v1.3 provided by SEPA through a data-sharing agreement. The National flood hazard maps v1.3 are derived by SEPA using the most up-to-date modelling techniques and use a consistent approach to produce mapping outputs across Scotland. These cover a range of flooding scenarios coastal, fluvial and surface water flooding (SEPA, 2020). This data enables the overlaying of spatial units (e.g. Mosaic households) with flood extents to identify the exposed population for three return periods (1 in 10, 1 in 200 and 1 in 1000 year) for fluvial and surface water flooding. It is noteworthy that although this study utilises hazard maps derived from flood modelling carried out by SEPA, the methods described in Section 2.3 are independent from the base datasets used here to represent flood hazards. Therefore, any robust hazard data that represents flood extent can be used to calculate underlying vulnerability with Experian Mosaic data.

Whilst the SEPA maps are the most readily available National dataset, and are based on the best available data at the time of development, it must be acknowledged there are inherent uncertainties in the complex flood modelling process that have an impact on corresponding mapping outputs. Due to the broad-scale application of these maps at the National level, the precision at which individual properties can be allocated to hard edge flood extents must be interpreted with caution. Additional uncertainty in both demographic input datasets (Mosaic and Census) must also be acknowledged. It is highly likely that demographics in the population will change more frequently than the flood maps are updated by SEPA, therefore whilst the accuracy of overlaying Mosaic properties with SEPA’s flood extents must be considered, so must the associated demographic characteristics present within Mosaic (2019) data.

In light of the above, it must be emphasised that the scope of this study is to explore the viability of higher resolution datasets in calculating flood vulnerability.

2.2. Case study areas

The main analysis was undertaken at a national scale, considering the whole of Scotland. The available Mosaic data for Scotland provided by Experian equates to 1,048,575 households and 70,162 postcodes; accounting for 23,272 of Scotland’s 46,351 Census output areas. Following the national assessment, two more detailed analyses were undertaken at the community scale. Two Scottish communities were selected on the basis that both sites contrasted in terms of the primary flood hazard (fluvial vs. surface water) and represent both an urban and rural conurbation.

The first-community level assessment was undertaken for Inverurie (Aberdeen), a small rural town that has suffered a long history of flooding from the River Don. The most recent (and severe) was in 2015/2016 during Storm Frank, where a total of 130 homes and 16 businesses were flooded and residents of 38 homes were evacuated (see Philip et al., 2020). The costs arising to Aberdeenshire Council from Storm Frank were estimated to be £8.3m (Aberdeenshire Council, 2019). The second-community level assessment was undertaken for Shettleston, an urban community in Glasgow which experienced flash floods from intense rainfall in 2002, resulting in several million pounds damages and the evacuation of around 200 homes (BBC, 2002a 2002b).
2.3. Method overview

National and local level vulnerability assessments were carried out for the whole population using Census and Mosaic data, as per the original SFVI method. The population at risk from flooding were then identified in ArcGIS Pro v2.6 using SEPA’s national flood hazard maps for fluvial and surface water flooding (detailed steps are provided in Supplementary Material A2). These maps account for the three return periods, and any output area that lay within the flood extents was considered exposed to flooding.

The same approach was undertaken using Mosaic data, however, discretised additionally (i) at the postcode and (ii) household level. The population exposed to flooding were determined by identifying the household locations (point data) that sat within the SEPA flood extent (polygon data). All results were then aggregated back to Census output areas, due to privacy concerns related to the use of household-level and postcode-level data.

The above methodology resulted in five output pathways, each of which represents a different degree of aggregation used to calculate vulnerability scores, hence allowing the impact of spatial resolution to be investigated. Whilst data privacy issues mean that all analysis pathways were aggregated to the output area level to produce final outputs, each can be distinguished by the level of aggregation adopted at the beginning of the analysis. Figure illustrates each analysis pathway and the process used to calculate vulnerability scores, namely:

- Census_OA uses Census 2011 data and calculates vulnerability at output area;
- Mosaic_PC and Mosaic_HH use Mosaic data and calculates vulnerability at household and postcode level;
- Exp_Mosaic_HH and Exp_Census_OA uses Mosaic data as above, but accounts for exposed properties, which are identified in accordance with SEPA’s national flood hazard maps.

Finally, the paired-sample Wilcoxon test was used to determine whether the analysis pathways are statistically significant from each other. This non-parametric test is used to compare related (paired) samples that describe the same population (McDonald, 2014; Wilcoxon, 1945).

2.3.1. Output area analysis with Census data

Vulnerability is measured at the output area level using Census 2011 data and the SFVI method from Tapsell et al. (2002). The SFVI consists of seven indicators (Table 1) that represent those financially deprived and with poor health conditions.

Percentages for each indicator were calculated and then data were transformed as per the methods outlined in Table 1 to minimise skewness and kurtosis. Transformed data were then standardised as Z-scores and vulnerability scores for each output area were calculated by summing Z-scores using the following equation (Tapsell et al., 2002):

\[
\text{Vuln. Score}_{OA} = (U + O + NCO + NHO) \times 0.25 + (LTS + SP + E)
\]

where \( U = \text{unemployment}, \ O = \text{overcrowding}, \ NCO = \text{non-car ownership}, \ NHO = \text{non-home ownership}, \ LTS = \text{long-term sick}, \ SP = \text{single parents} \) and \( E = \text{elderly} \).
To facilitate comparison of vulnerability outputs across all analysis pathways (Figure 1), outputs are normalised as Z-scores. This enables categorisation of vulnerability into five bands: ‘very low’ (0–0.2); ‘low’ (0.2–0.4); ‘medium’ (0.4–0.6); ‘high’ (0.6–0.8) and ‘very

Table 1. SFVI Indicators used to calculate vulnerability with transformation methods to minimise skew and kurtosis.

| Indicator            | Description                                                                 | Transformation method |
|----------------------|-----------------------------------------------------------------------------|-----------------------|
| Unemployment         | Unemployed residents 16+ as a percentage of all economically active residents 16+ | Log natural (x+1)     |
| Overcrowding         | Households with more than one person per room as a percentage of all households. | Log natural (x+1)     |
| Non-car ownership    | Households with no car as a percentage of all households.                   | Square root           |
| Non-home ownership   | Households not owning their own home as a percentage of all households.     | Square root           |
| The long-term sick   | Residents suffering from limiting long-term illness as a percentage of all residents. | Log natural (x+1)     |
| Single parents       | Lone parents as a proportion of all residents                               | Square root           |
| The elderly          | Residents aged 75+ as a percentage of all residents                         | Log natural (x+1)     |

Figure 1. Flow chart outlining the processes in which vulnerability is calculated at each level of aggregation.
“High” (0.8–1.0). This categorisation remains consistent for analyses using Census data and Mosaic data.

2.3.2. Postcode and household analysis with Mosaic data

The Mosaic dataset defines the population as one of 57 different typologies. These typologies are described using variables as statistically representative means for each typology (e.g. on average, 0.63% of Mosaic group A01 are Aged 75+). The vulnerability classifications for each typology and how their distributions are provided in Supplementary Material A1. SFVI values for each typology are also provided and accompanied by additional information on key contrasts between selected typologies.

Data for each SFVI indicator was extracted from Mosaic data and then vulnerability scores for each respective typology are a product of calculating each of the SFVI variables as per Equation (1). These unique typology scores are normalised as before and averaged across the output area (Figure 2). The method remains the same as the original SFVI in principle, only vulnerability scores are resolved at household or postcode level first before being averaged and resolved to output area. This process is also followed for the exposed population analysis. For example, an output area of 20 households in total, the vulnerability is determined by averaging across all households. However, where only 4 are exposed to flooding, the vulnerability score in the exposed analysis would average across these 4 only.

![Figure 2](Image)

**Figure 2.** Vulnerability calculation at the three levels of spatial aggregation. The method for example showing Mosaic typology, ‘Mature Workers’ (K45) is the same for postcode level. (See Experian, 2017 for full details on each Mosaic typology).
3. Results

3.1. National analysis

3.1.1. Vulnerability of the whole population

Figure 3 compares the distributions and boxplots of vulnerability at the national level for the whole population (not just those exposed to flood hazards) at each level of spatial aggregation. The vulnerability of the population using Census 2011 data produces results with a near-normal distribution. Conversely, Mosaic data (both at household and postcode level) produce non-normal distributions. Given that Census data is resolved to output area level (accounting for up to 125 households), it can be expected that detail on the population is lost due to aggregation, thus resulting in a normal distribution of vulnerability.

![Figure 3. National level vulnerability distributions and boxplots comparing three different levels of aggregation using Census (Output Area) and Mosaic (Postcode; Household) data. Vertical dashed lines represent vulnerability classification intervals.](image-url)
These distributions suggest that Mosaic provides a more realistic representation of vulnerability than Census data because the final output area vulnerability score is an aggregation of various typologies and their diverse circumstances (i.e. SFVI indicators). A Wilcoxon signed-rank test comparing each level of aggregation shows that there is a statistically significant difference between each, as all report $p < .001$ (Table 2). The magnitude of effect sizes comparing each level of aggregation are noteworthy; output area and postcode produce a small effect size ($r = 0.198$), household and postcode are even smaller ($r = 0.097$), whereas output area and household report a moderate magnitude ($r = 0.397$). Such a small effect size between household and postcode is unsurprising as they both stem from Mosaic and the difference in resolution is not as significant as household and output area, hence the large effect size between these two levels of aggregation. They are both from different datasets and represent the largest gap in resolution. However, if we examine the data point count between household (1,048,575) and postcode (70,162), it could be argued that the difference in vulnerability distributions is not as stark as would be expected.

3.1.2. Vulnerability of the exposed population

This section compares the vulnerability of populations exposed to flooding. Figure 4 illustrates boxplots and vulnerability distributions of each exposed population group. The vulnerability distributions of both Census and Mosaic data echo that of the whole population as the former is near normal and the latter is right-skewed, showing an emphasis towards less vulnerable groups. This suggests that at the national level, exposed output areas are not necessarily any more or less vulnerable than the wider population. Furthermore, examining the vulnerability of output areas using Census data show that the difference between return periods (both fluvial and surface water) is virtually indistinguishable when examining the medians (Table 3; for all return periods the median is in the range of 0.45–0.48). Similarly, there is little variation in medians using Mosaic data. The primary exception in Figure 4 is the Mosaic data boxplots for those exposed to fluvial flooding; the presence of outliers in the upper whiskers of each return period potentially suggests that highly vulnerable Mosaic typologies are clustered within fluvial floodplains, leading to an irregularly high vulnerability score for these output areas. Furthermore, whilst the median shows little variation, the extremes of the Mosaic distribution suggest that the majority exposed to flooding are less vulnerable (fluvial and surface water), whereas Census, in comparison, underestimates the most and least vulnerable in its distribution.

### Table 2

| National | Census_OA | 0.495 | OA_HH | 0.397 | OA count | 23,272 | HH/PC count | – |
|----------|-----------|--------|-------|--------|-----------|--------|-------------|---|
| OA_HH    | 0.333     | ***    | 0.198 | 23,272 | 70,162    |        |             |   |
| Mosaic_PC| 0.378     | ***    | 0.097 | 23,272 | 1,048,575 |        |             |   |

***$p \leq .001$.**
Insights such as this favour the argument for using higher-resolution datasets in vulnerability assessments. However, it is also worth noting that this clustering effect cannot be deduced for populations exposed to surface water flooding using Mosaic. Although, this may say something about the distinction between both sources of flooding. Surface water flooding is ‘spotty’ in nature meaning that exposed Mosaic typologies will not be clustered like those situated near river banks. Therefore, when aggregated to output area they yield similar results to Census data. With regard to statistical significance, comparison of exposed population groups between Census and Mosaic (10 yr; Census against Mosaic, etc.) produces $p < .001$ for all cases in the Wilcoxon signed-rank test (Table 3). Furthermore, effect size magnitudes are all reported as moderate except for 10 and 200 yr fluvial events.

Figure 4. National level: vulnerability distributions and boxplots comparing two different levels of aggregation using Census (Output Area) and Mosaic (Household) data for exposed population. Vertical dashed lines represent vulnerability classification intervals. Counts for each return period are stacked on top of each other.
3.2. Local analysis

3.2.1. Vulnerability of the whole population

Figure 5 illustrates the vulnerability using Census and Mosaic data for the whole population of two Scottish communities, Inverurie (rural area: Aberdeenshire) and Shettleston (urban area: Glasgow). In all cases, irrespective of the underlying data used, or the level of aggregation, Shettleston is visibly more vulnerable than Inverurie as there are significant spikes in vulnerability at approximately 0.8 (highly vulnerable) which represent around 21% of output areas in Shettleston. In contrast, however, there is an absence of output areas that are classed as ‘very high’ in Inverurie. This observation is consistent across both Census and Mosaic datasets at each level of aggregation. The difference between vulnerability distributions for Census and Mosaic data are similar in Inverurie to those seen with the national-level data, with the Mosaic data being skewed to the left compared to the Census data that is more normally distributed.

However, in Shettleston, both data sets show distributions favouring higher vulnerability. Contrary to the output area, household and postcode level results in Inverurie exhibit a similar trend of vulnerability as that shown at the national level; the majority of the households are classed as having ‘low’ and ‘very low’ vulnerability. Conversely, however, ‘high’ is the dominant classification of household vulnerability in Shettleston. In terms of median vulnerability (Table 4) differences between the postcode and the household level are small, with one exception. In Inverurie, 6% of the population are classed as having a ‘very high’ vulnerability at postcode level compared to 0% at household level. Given that both household and postcode levels adopt the same underlying social data, this difference in vulnerability potentially highlights the impact of aggregation as the final vulnerability score for an output area is aggregated across 374 unique postcodes compared to 5719 households.

As before, the Wilcoxon signed-rank test reports that there are statistically significant differences between each level of aggregation as $p < .001$ in both communities. Moreover, all effect sizes can be reported as moderate, with one exception; a small effect size of 0.281 when comparing output area and postcode in Shettleston where median values are 0.67 and 0.614, respectively. The differences in vulnerability between output area and postcode vulnerability are therefore trivial, however, the latter is resolved at higher
resolution and holds the advantage of advanced insights into the population through Mosaic typologies.

3.2.2. Vulnerability of the exposed population

Lastly, we examine the vulnerability of the exposed populations at the local level. Taking the size of each local community into account and the associated number of exposed output areas for each return period (Tables 5 and 6), it should be noted that sample sizes are considerably smaller than previous sections. Therefore, our Wilcoxon signed-rank test outputs follow no emergent pattern in terms of statistical significance as sample size has a clear impact on the statistical power of each scenario, therefore the results of which must be interpreted with caution.

Figures 6 and 7 show the vulnerability of exposed populations in Inverurie and Shettleston, respectively. There are differences between Census and Mosaic for Inverurie as
### Table 4. Local level: summary statistics comparing two different levels of aggregation using Census (OA = Output Area) and Mosaic (PC = Postcode; HH = Household) data for whole population of Inverurie and Shettleston.

|          | Median | p   | r   | OA count | HH/PC count |
|----------|--------|-----|-----|----------|-------------|
| **Inverurie** |        |     |     |          |             |
| Census_OA  | 0.411  |     |     | 113      |             |
| OA_HH      |        | ***| 0.469|          |             |
| Mosaic_PC  | 0.238  |     |     | 113      | 374         |
| PC_OA      |        | ***| 0.345|          |             |
| Mosaic_HH  | 0.272  |     |     | 113      | 5719        |
| HH_PC      |        | ***| 0.335|          |             |
| **Shettleston** |      |     |     |          |             |
| Census_OA  | 0.67   |     |     | 275      |             |
| OA_HH      |        | ***| 0.412|          |             |
| Mosaic_PC  | 0.568  |     |     | 275      | 736         |
| PC_OA      |        | ***| 0.281|          |             |
| Mosaic_HH  | 0.613  |     |     | 275      | 11,816      |
| HH_PC      |        | ***| 0.416|          |             |

***p ≤ .001.

### Table 5. Inverurie: summary statistics comparing two different levels of aggregation using Census (OA = Output Area) and Mosaic (HH = Household) data for populations exposed to fluvial and pluvial flooding (F = fluvial; P = pluvial).

|          | Median | p   | r   | OA count | HH count |
|----------|--------|-----|-----|----------|----------|
| **Inverurie** |        |     |     |          |          |
| F_1 in 10 | Exp_Cen_OA | 0.275| 0.625ns| 0.365| 4        |
|           | Exp_Mos_HH | 0.212|     |         | 4        |
| F_1 in 200| Exp_Cen_OA | 0.386| 0.30ns| 0.375| 9        |
|           | Exp_Mos_HH | 0.266|     |         | 9        |
| F_1 in 1000| Exp_Cen_OA | 0.299| 0.32ns| 0.322| 11       |
|           | Exp_Mos_HH | 0.266|     |         | 11       |
| P_1 in 10 | Exp_Cen_OA | 0.387| 0.015*| 0.894| 7        |
|           | Exp_Mos_HH | 0.237|     |         | 7        |
| P_1 in 200| Exp_Cen_OA | 0.387| ***| 0.885| 11       |
|           | Exp_Mos_HH | 0.237|     |         | 11       |
| P_1 in 1000| Exp_Cen_OA | 0.397| ***| 0.744| 18       |
|           | Exp_Mos_HH | 0.22 |     |         | 18       |

ns = p > .05.

*p ≤ .05.

***p ≤ .001.

### Table 6. Shettleston: summary statistics comparing two different levels of aggregation using Census (OA = Output Area) and Mosaic (HH = Household) data for populations exposed to fluvial and pluvial flooding (F = fluvial; P = pluvial).

|          | Median | p   | r   | OA count | HH count |
|----------|--------|-----|-----|----------|----------|
| **Shettleston** |        |     |     |          |          |
| F_1 in 10 | Exp_Cen_OA | 0.659| 0.688ns| 0.214| 6        |
|           | Exp_Mos_HH | 0.623|     |         | 6        |
| F_1 in 200| Exp_Cen_OA | 0.671| 0.003**| 0.433| 44       |
|           | Exp_Mos_HH | 0.564|     |         | 44       |
| F_1 in 1000| Exp_Cen_OA | 0.671| 0.004**| 0.407| 48       |
|           | Exp_Mos_HH | 0.56 |     |         | 48       |
| P_1 in 10 | Exp_Cen_OA | 0.671| 0.232ns| 0.403| 10       |
|           | Exp_Mos_HH | 0.562|     |         | 10       |
| P_1 in 200| Exp_Cen_OA | 0.668| 0.087*| 0.295| 34       |
|           | Exp_Mos_HH | 0.596|     |         | 34       |
| P_1 in 1000| Exp_Cen_OA | 0.668| 0.004**| 0.448| 40       |
|           | Exp_Mos_HH | 0.557|     |         | 40       |

ns = p > .05.

*p ≤ .05.

**p ≤ .01.
Mosaic shows a skew towards less vulnerable groups compared to Census. In Shettleston, both datasets are similarly skewed, however, towards higher vulnerability scores. Interestingly, both Figures 6 and 7 show that Mosaic represents vulnerability at both extremes (0–0.2 and 0.8–1.0) therefore suggesting that the least and most vulnerable are not captured by Census.

Whilst not as distinguishable at the national level, determining whether or not those that are exposed to flooding are more or less vulnerable than the entire population is more apparent at the local level. Mosaic household data (whole population; Figure 5) showed that vulnerability of Inverurie ranges from 0 to 0.8 and represented vulnerability classes from ‘very low’ – ‘high’. Moreover, Shettleston ranges from 0 to 1 and represented all five classes (‘very low’ – ‘very high’). However, in both fluvial and surface water scenarios (Figure 6), Inverurie shows an absence of vulnerability between approximately 0.5 and 0.8.
('high’-'very high’) using Mosaic. Similarly, there is an absence of surface water areas with scores of approximately 0.3 ('low') in Shettleston (Figure 7). In contrast, however, fluvial vulnerability generally represents the distribution of the whole population in Shettleston.

The aforementioned differences in vulnerability between the exposed population and the whole population highlight interesting contrasts between both communities; Inverurie’s exposed population (fluvial and surface water) could be considered less vulnerable than the whole population as there are no areas that are classed as ‘high’. This means that in Inverurie, there are Mosaic typologies that are highly vulnerable as per the SFVI indicators, however, (fortunately) they are not exposed to flooding. The opposite is true for Shettleston; there is an absence of some ‘low’ vulnerability areas for populations exposed to surface water flooding only, suggesting that a segment of the population are more vulnerable to surface water than fluvial flooding. However, whilst this is the case,
vulnerability is distributed similarly to the whole population, meaning that the exposed population are equally as vulnerable as the whole community.

4. Discussion

4.1. Comparing the characteristics of Mosaic typologies

Mosaic typologies provide a more refined way to examine flood vulnerability as they use a plethora of variables to characterise members of society which can be considered in vulnerability assessments. This highlights the advantage of refined information afforded by Mosaic that Census cannot replicate; without societal characterisation in the form of typologies, such high differences between variables would alter the vulnerability of an entire output area using Census. The ability to distinguish between social typologies in this manner affords a more heterogenous approach to vulnerability characterisation. Whilst out with the scope of this study, in-depth comparison of Mosaic typologies in exposed areas would prove to be a valuable process in flood management. Similar processes are already commonplace; for example, Tapsell et al. (2002) compared two areas in the U.K. that had flood alleviation schemes implemented, similarly, Scottish Government (2015) consider the exposure of residential properties to flooding, however, express this as a proportion of residents within the floodplain only. Both studies recognise the value of focusing on the exposed populations, however, the former does so at a broad level and the latter (whilst considerate of individual properties) does not use social characteristics of these properties that geodemographics can afford.

4.2. Comparison of differences between Census and Mosaic

Overall, there are clear differences between vulnerabilities calculated at different spatial scales using different datasets. Calculating vulnerability at Census output area generally produces normal distributions at the national level for both the exposed and wider population. In contrast, the equivalent analysis pathway using Mosaic shows that at household and postcode level (exposed whole population), the distributions are not normal and are right-skewed, thus classifying the population mostly as ‘very low’ and ‘low’ vulnerability. The vulnerability of the population therefore tends to be lower than if the underlying data used is less spatially discretised initially; whilst there is an averaging process involved to resolve at output area, the SFVI method employed using Mosaic maintains a more realistic representation of national vulnerability than Census, as detail is captured at higher resolutions first.

This highlights the advantage of using data sources that have a higher resolution from the outset for calculating population vulnerability, as it considers an array of socially different typologies, whereas the aggregation involved in producing Census data does not consider this. In contrast to a single vulnerability score calculated for an output area from the outset using the aggregated Census data, the use of household-level data provides a more accurate description of the ‘collective’ vulnerability in an area as the individual vulnerabilities of households are taken into account. It should also be highlighted that whilst Census data as a whole is updated every 10 years, and Mosaic biannually, the underlying SFVI variable data used in Mosaic is extracted from Census
2011 data. Therefore, whilst survey periods are technically the same, the vulnerability profiles produced by Mosaic are strikingly different from Census data. Clearly, the issue of spatial resolution and its effect on vulnerability output remains.

4.3. Additional insights into flood vulnerability afforded by high-resolution data

The SFVI method using Census data calculates flood vulnerability of entire geographical areas, irrespective of whether households are exposed to flooding or not. The availability of spatially refined data enables the identification of households that intersect the floodplain, therefore exposure quantification is more accurate. In turn, this produces a more true representation of vulnerability, as non-exposed households are not considered. Comparative analysis of national-level fluvial and surface water vulnerability produced interesting results with regards to vulnerability after aggregation; the ‘spotty’ nature of surface water flooding is likely to result in exposed households being dispersed across an output area. For example, extreme socio-economic contrasts between exposed typologies in terms of vulnerability could be expected for surface water, whereas typologies exposed to fluvial flooding will be clustered in the same part of an output area (i.e. by the river). This means that final vulnerability scores for an output area will vary dramatically between surface water and fluvial flooding when aggregating, as the former is averaging across typologies of different classifications, whereas the latter is averaging typologies with similar levels of vulnerability. Further analysis of this phenomena will be considered as future work.

Whilst the advantage of examining the exposed population’s vulnerability using household data may not be apparent at the national level, disparities in vulnerability results are highlighted across both fluvial and surface water return periods at the local level. The exposed populations of Inverurie and Shettleston contrast each other, with lower vulnerability classifications in the former and generally a more equal representation of all vulnerability classifications in the latter. Contrary to the national level exposed analysis, the local level exposed analysis found differences in vulnerability between the exposed and wider; Shettleston were equally as vulnerable as the wider population, whereas in contrast, Inverurie’s exposed population were in fact less vulnerable than the wider population. Given that there is no significant distinction between surface water and fluvial exposed populations for the aforementioned contrasts, this is, therefore, a key finding, as it points towards markedly different socio-economic profiles between the two sites and the circumstances of those situated in areas at risk of flooding. Whilst in-depth analysis of socio-economic characteristics of both areas is out of the scope of this study, it is worth highlighting that these contrasts between exposed and the whole populations would not have emerged using broad-scale data.

4.4. Limitations of Mosaic and future research direction

Determining an area’s vulnerability to flooding using unique household level data is more advantageous than broad-scale aggregated social data, however, it is not without limitations. Firstly, it must be acknowledged that a degree of vulnerability will also be masked using Mosaic data due to the averaging of individual typologies. Moreover, there is a further averaging effect that is inherent within the raw Mosaic data as
vulnerability indicators are provided as statistically representative means as opposed to the absolute value that Census data provides. Therefore (spatial resolution aside), the certainty surrounding an output areas’ vulnerability score could be considered higher using Census, thus presenting a trade-off between enhanced spatial representation and more uncertainty in classification of vulnerable groups. Regarding accessibility, there are limitations surrounding data privacy due to the wealth of sensitive personal information available in the Mosaic dataset at the household level. However, postcode and household level vulnerability results do not diverge substantially enough to claim that one level of aggregation is superior to the other, therefore, postcode level analysis provides scope to act as a middle-ground between Census data at the output area and Mosaic data at the household level, as it is aggregated enough to comply with data privacy requirements whilst also maintaining an advantage over Census data due to a higher resolution. However, this would limit analysis of exposed populations as postcodes are typically represented as polygons (aggregating parts of streets for example) as opposed to household point data. This would therefore include or omit groups who may be not exposed at all, which in turn would bias vulnerability results. Despite this, however, the polygons representing postcode data would still be an improvement on the previous work that accounts for exposure (Scottish Government, 2015), therefore future research should seek to combine more comprehensive vulnerability indices that are more diversified in terms of socio-economic characteristics with geodemographic. Datasets such as Mosaic are more spatially discretised and also represent social groups with a wealth of information.

5. Conclusion

A fundamental limitation of index-based approaches to assess vulnerability is the spatial scale of the underlying data used. Typically, these cover broad geographical areas. Whilst this is useful for studies at the regional or national level as an initial assessment of vulnerability it is vital to capture vulnerability at a granular level as possible. This paper demonstrates the first application of high-resolution data to calculate flood vulnerability at the household level and provides a clearer picture of vulnerability from the perspective of individual entities. Mosaic produces clear differences in vulnerability categorisation when compared to traditional methods using Census data. At the national level, Scotland is found to be less vulnerable when analysis uses Mosaic data at the postcode and household level. Whilst the Mosaic approach is not without limitations, it highlights issues with broad-scale approaches as it was found that using Census data underpredicts the presence of the most vulnerable people.

The exposed population analysis using Mosaic data at the national level reveals that areas are categorised as less vulnerable overall compared to Census data. However, local-level analysis using Mosaic data highlights that exposed populations may differ in vulnerability between communities. This indicates that the vulnerability categorisation of individual Mosaic typologies is useful, as the presence of typologies who are individually more vulnerable than others are likely to significantly alter the overall vulnerability rating of at-risk areas. The source of flooding also impacts vulnerability as exposure to fluvial flooding is a function of spatial proximity to the river, whereas surface water
exposure is more localised as it is based on microtopography and corresponding rainfall in that area; this is reflected in our analysis as fluvial and surface water vulnerability distributions differ. This requires further investigation as it suggests that social profiles of those located adjacent to the river vary.

However, whilst this study utilises higher resolution demographic data as a means of improving the discretisation of flood vulnerability outputs, uncertainty can be expected in the analysis of the exposed population due to inherent uncertainty present within the flood maps. Therefore, the conclusions that can be drawn from this part of the analysis should be interpreted with caution until uncertainty can be accounted for more comprehensively. Future work would benefit from the use of probabilistic flood maps, as opposed to deterministic ‘hard edge’ maps. A dichotomous flood extent boundary runs the risk of under/over-estimating the allocation of properties that lie within flood extents. This uncertainty has a cascading effect on the vulnerability characterisation when aggregating from property level to output area. Therefore, a probabilistic approach to the exposure analysis would mitigate uncertainties associated with fluvial and surface water flood mapping by providing exposed properties with confidence intervals that can be factored into vulnerability calculation.

This study sought to examine the impact of spatial resolution on vulnerability as opposed to the indicators themselves. However, the availability of increasingly large complex datasets provides the potential to further the discussion around indicator selection. For example, the indicator selection for SFVI was assigned using the best available data at the time (i.e. Census) which inherently limited the spatial resolution of the analysis. However, even the most known indicator set (SoVI; Cutter, 1996; Cutter et al., 2003) is still applied at broader scales, the most common scale for vulnerability analyses remains at the administrative boundary level (e.g. output area) (Fekete, 2009, 2019) and the primary focus in the literature surrounding indicator methods seems focused solely around the validity of indicators (Fekete, 2009). The indicators adopted in this study were selected as they facilitate like-for-like vulnerability analysis between datasets, ensuring the focus was on a spatial scale as opposed to indicators. However, myriad objective indicator sets in the literature make it increasingly difficult to assess vulnerability objectively. This therefore presents researchers and practitioners with a choice of continuing to produce divergent sets of indicators that may still suffer from aggregation or move towards utilising datasets such as Mosaic and fitting vulnerability to the comprehensive set of variables included. This approach would both potentially overcome some of the effects of aggregation by permitting analysis at the household level and would also enhance transparency in vulnerability assessments as indices are based on a more universal, communicable and marketable dataset.

Our analysis exposes some of the issues surrounding the discretisation of the underlying data used. Geodemographics presents an opportunity to both address the issue of spatial scale whilst also providing a rich new dataset from which new or existing vulnerability indicators can be based on. Furthermore, many of Mosaic’s variables depend on the participation of consumers through lifestyle surveys. Adaptation of this approach to the context of floods or other natural hazards can potentially improve the external validity of indicators as validation does not solely rely on experts, but also relies on the population
in question (i.e. those who are deemed to be vulnerable), which in turn will increase the confidence in outputs (Emrich, 2005).

There is an opportunity for a larger research effort to collaborate with commercial organisations such as Experian in creating a geodemographic database tailored specifically to flooding or other natural hazards (Mosaic’s original purpose is for strategic marketing). However, the use of rich datasets containing vast amounts of sensitive information would require more due diligence from data processors and controllers. This may make collaboration difficult among organisations when looking at vulnerability at the household level. Interestingly though, household and postcode vulnerability were found to be almost indistinguishable, which presents a potential workaround to the issue of compromising sensitive data when sharing sensitive outputs. However, this again presents another trade-off in terms of spatial accuracy. The indicators used in this study only account for approximately 0.01% of variables available in Mosaic, therefore despite limitations, it is imperative that the most accurate, rich and representative picture of social vulnerability to flooding is created by using all available data at our disposal (whilst being conscious of data protection issues).

Acknowledgement
This study was undertaken as part of the EPSRC EP/N030419/1 project ‘Water Resilient Cities: Climate Uncertainty and Urban Vulnerability to Hydro-hazards’. We are grateful to the Scottish Environmental Protection Agency for the provision of the National Hazard Maps (available at http://map.sepa.org.uk/floodmap/map.htm) and Experian Ltd. for Mosaic (Scotland).

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
This work was supported by Engineering and Physical Sciences Research Council (EPSRC) [Grant Number EP/N030419/1].

ORCID
D. Morrison http://orcid.org/0000-0003-4612-081X
L. Beevers http://orcid.org/0000-0002-1597-273X
G. Wright http://orcid.org/0000-0003-3241-1456
M. D. Stewart http://orcid.org/0000-0003-3754-4995

References
Aberdeenshire Council. (2019). Local climate impact profile (LCILP). http://publications.aberdeenshire.gov.uk/dataset/ca4d686c-f8a8-4390-af0f-8088d2b536bb/resource/dbb94611-c5f0-492f-9ed1-a762fb0813bc/download/lclip-2019-high-res.pdf
Alexander, M., Viavattene, C., Faulkner, H., & Priest, S. (2011). A GIS-based flood risk assessment tool: supporting Flood Incident Management at the local scale. Flood Hazard Research Centre, Middlesex University.

Anderson, C. C., Hagenlocher, M., Renaud, F. G., Sebesvari, Z., Cutter, S. L., & Emrich, C. T. (2019). Comparing index-based vulnerability assessments in the Mississippi delta: Implications of contrasting theories, indicators, and aggregation methodologies. International Journal of Disaster Risk Reduction, 39(March), 101128. https://doi.org/10.1016/j.ijdrr.2019.101128

Balica, S. F., Beevers, L., Popescu, I., & Wright, N. G. (2013). Parametric and physically based modelling techniques for flood risk and vulnerability assessment: A comparison. Journal of Environmental Modelling & Software, 41(3), 81–92. https://doi: 10.1016/j.envsoft.2012.11.002.

BBC. (2002a, 31 July). Homes evacuated after flooding. BBC news. Scotland. http://news.bbc.co.uk/1/hi/scotland/2161637.stm

BBC. (2002b, 1 August). Floods cost into millions. BBC news. Scotland. http://news.bbc.co.uk/1/hi/scotland/2166701.stm

Beevers, L., Walker, G., & Strathie, A. (2016). A systems approach to flood vulnerability. Civil Engineering and Environmental Systems, 33(3), 199–213. http://doi.org/10.1080/10286608.2016.1202931

Connor, R. F., & Hiroki, K. (2005). Development of a method for assessing flood vulnerability. Water Science Technology, 51(5), 61–67. https://doi.org/10.2166/wst.2005.0109

Cutter, S. L. (1996). Vulnerability to environmental hazards. Progress in Human Geography, 20(4), 529–539. https://doi:10.1177/030913259602000407

Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. Social Science Quarterly, 84(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Dilley, M., Chen, R. S., Deichmann, U., Lerner-Lam, A. L., Arnold, M., Agwe, J., Buys, P., Kjekstad, O., Lyon, B., & Yetman, G. (2005). Natural disaster hotspots: A global risk analysis. The World Bank, Hazard Management Unit.

Emrich, C. T. (2005). Social vulnerability in US metropolitan areas: Improvements in hazard vulnerability assessment [Doctoral dissertation]. University of South Carolina. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.621.6113&rep=rep1&type=pdf

Eurostat. (2011). EU legislation on the 2011 population and housing censuses: Explanatory notes. Experian. (2017). Mosaic Scotland: Data profile, 7–8. https://www.experianintact.com/content/uk/documents/productSheets/MosaicConsumerScotland.pdf

Experian. (2019). Mosaic Scotland: The consumer classification for Scotland. https://www.experian.co.uk/assets/marketing-services/brochures/mosaic-brochure-scotland.pdf https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-RA-11-006

Fekete, A. (2009). Validation of a Social Vulnerability Index in context to river-floods in Germany. Natural Hazards and Earth System Science, 9(2), 393–403. https://doi.org/10.5194/nhess-9-393-2009

Fekete, A. (2019). Social vulnerability (re-)assessment in context to natural hazards: Review of the usefulness of the spatial indicator approach and investigations of validation demands. International Journal of Disaster Risk Science, 10(2), 220–232. https://doi.org/10.1007/s13753-019-0213-1

Fitton, J. M., Hansom, J. D., & Rennie, A. F. (2018). A method for modelling coastal erosion risk: The example of Scotland. Natural Hazards. Springer Netherlands, 91(3), 931–961. https://doi.org/10.1007/s11069-017-3164-0

Lindley, S., & O'Neill, J. (2013). Flood disadvantage in Scotland: Mapping the potential losses in well-being. Scottish Government Social Research.

Lindley, S., O'Neill, J., Lawson, N., Christian, R., & O'Neill, M. (2011). Climate change, justice and vulnerability. Joseph Rowntree Foundation. http://www.jrf.org.uk/sites/files/jrf/climate-change-social-vulnerability-full.pdf

McDonald, J. H. (2014). Handbook of biological statistics – paired t-test. Sparky House. http://www.biostathandbook.com/pairedttest.html

Mwale, F. D., Adeloye, A. N., & Beevers, L. (2015). Quantifying vulnerability of rural communities to flooding in SSA: A contemporary disaster management perspective applied to the lower Shire
valley, Malawi. *International Journal of Disaster Risk Reduction*, 12, 172–187. https://doi.org/10.1016/j.ijdrr.2015.01.003

Nasiri, H., Mohd Yusof, M. J., & Mohammad Ali, T. A. (2016). An overview to flood vulnerability assessment methods. *Sustainable Water Resources Management*, 2(3), 331–336. https://doi.org/10.1007/s40899-016-0051-x

National Record of Scotland. (2018). Plans for Scotland’s census 2021. https://www.scotlandscensus.gov.uk/documents/census2021/Plans_for_Scotlands_Census_2021.pdf

Penning-Rowsell, E. C., Johnson, C., Tunstall, S. M., et al. (2005). The benefits of flood and coastal risk management: a manual of assessment techniques. *Flood Hazard Research Centre*, 89. https://doi.org/10.1596/978-0-8213-8050-5

Philip, L., Dowds, G., & Currie, M. (2020). Long-term impacts of flooding following the winter 2015/16 flooding in North East Scotland: Comprehensive report (CRW2016_02). Scotland’s Centre of Expertise for Waters (CREW).

Rufat, S., Tate, E., Burton, C. G., & Maroof, A. S. (2015). Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction*, 14, 470–486. https://doi.org/10.1016/j.ijdrr.2015.09.013

Scottish Environmental Protection Agency. (2011). National flood risk assessment. https://www.sepa.org.uk/media/99914/nfra_method_v2.pdf

Scottish Environmental Protection Agency. (2020). How were the maps developed? https://www.sepa.org.uk/media/163436/how-were-the-maps-developed.pdf

Scottish Government. (2011). *Scotland’s census 2011 general report*. https://www.scotlandscensus.gov.uk/documents/censusresults/Scotland’s_Census_2011_General_Report.pdf

Scottish Government. (2015). *Mapping flood disadvantage in Scotland 2015: Methodology report*. 102.

Statistics Canada. (2016). Census dictionary: Dissemination area. https://www12.statcan.gc.ca/census-recensement/2016/ref/dict/geo021-eng.cfm

Tapsell, S., Penning-Rowsell, E. C., Tunstall, S. M., & Wilson, T. (2002). Vulnerability to flooding: Health and social dimensions. *Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences*, 360(1979), 1511–1525. https://doi.org/10.1098/rsta.2002.1013

Tomlinson, C. J., Chapman, L., Thorne, J. E., & Baker, C. J. (2011). Including the urban heat island in spatial heat health risk assessment strategies: A case study for Birmingham, UK. *International Journal of Health Geographics*, 10(1), 42. https://doi.org/10.1186/1476-072X-10-42

Werritty, A., Houston, D., Ball, T., Tavendale, A., & Black, A. (2007). Exploring the social impacts of flood risk and flooding in Scotland. *Scottish Executive for Social Research*, Available at: https://www.webarchive.org.uk/wayback/archive/20150219020106.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83. hdl:10338.dmlcz/135688. JSTOR 3001968. https://doi.org/10.2307/3001968

Willis, I., Gibin, M., Barros, J., Webber, R., 2014. Applying neighbourhood classification systems to natural hazards: A case study of Mt Vesuvius. *Natural Hazards* 70(1), 1–22. doi:10.1007/s11069-010-9648-9