ORBIT - Online Repository of Birkbeck Institutional Theses

Enabling Open Access to Birkbeck's Research Degree output

The application of multivariate cluster analysis in the assessment of volcanic social vulnerability

https://eprints.bbk.ac.uk/id/eprint/40182/

Version: Full Version

Deposit Guide
Contact: email

THE APPLICATION OF MULTIVARIATE CLUSTER ANALYSIS IN THE ASSESSMENT OF VOLCANIC SOCIAL VULNERABILITY

Iain Stuart Willis

Department of Geography, Environment and Development studies
Birkbeck, University of London

DECLARATION

I, Iain Willis, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, it has been duly indicated.

Signed: Date:

ACKNOWLEDGEMENTS

THESIS OUTPUTS

PEER REVIEWED JOURNAL PUBLICATIONS

Willis, I., Gibin, M., Barros, J., & Webber, R. (2010). Applying neighbourhood classification systems to natural hazards: a case study of Mt Vesuvius. Natural Hazards, 70(1), 1-22.

PEER REVIEWED CONFERENCE PROCEEDINGS

Willis, I., Barros, J., and Gibin, M. (2010). *The application of geodemographics to social vulnerability and volcanic hazard assessment*. GISRUK, 14-16 Apr 2010, University College London. (Unpublished)

CONFERENCE PRESENTATIONS

Willis, I., Barros, J., and Gibin, M. (2013). *Living in the shadow: a neighbourhood classification approach to volcanic vulnerability in Quito*. EUGEO, 5-7 Sept 2013, Sapienza, University of Rome. (Unpublished)

Willis, I., and Fitton, J. (2015). A case study review of the multiple methodologies used to define social vulnerability classifications in hazardous environments. European Geosciences Union (EGU), 12-17 Apr, Vienna.

RELATED RESEARCH WEBSITES

Willis, I. (2010). Dispatches from a geographer. Online. Available at: http://iainwillis.wordpress.com/

Willis, I. and Florisca, A. (2011). The Cloud Client project: free geodemographic analysis. Online. Available at: http://cloud-client.co.uk/

ABSTRACT

The 20th Century was characterized by increasing human population settlement in volcanically active regions of the world. This continued growth, particularly in less developed nations, has led to an increasing exposure of households and communities more predisposed to the social and physical risks a disaster could present. This thesis proposes a new methodology for the identification, targeting and assessment of these socially vulnerable communities. Drawing from specific examples of Mount Vesuvius (Italy) and Guagua Pichincha (Ecuador), multivariate statistics are applied to population census data to characterise the frailties and assumed coping capacity of different neighbourhood types to volcanic risk. Using cluster analysis and geodemographic discriminatory techniques, results show that communities more pre-disposed to the social and economic pressures of a disaster can be identified using this method. This approach looks to enhance upon current disaster risk metrics that tend to focus on single or cumulative risk scores, rather than seeking to define the behavioural traits and attitudinal perceptions of a neighbourhood. The peripheral and often informal barrios around Quito, Ecuador are shown to be highly susceptible to volcanic social vulnerability, whilst the Campania province around Vesuvius, Italy, highlights that the greatest risk to community resilience is associated with the high density settlements along the coastal towns near the volcano. The complex nature and site-specific characteristics of volcanic hazards, as well as the cultural landscape in which a volcanic eruption takes place are found to be key determinants in all aspects of disaster reduction. Vulnerability indicators, as defined in previous studies of disaster response are often independent of each other, and in many cases, non-transferrable in different cultural settings. Similarly, vulnerability and risk perception are as much a consequence of culture and state as they are of geographical setting and the physical characteristics of a volcanic eruption. Whilst caution is advised on the application and treatment of vulnerability metrics for mitigation, examples are provided as to how a neighbourhood classification systems methodology can be practically applied for disaster risk reduction. The output of this thesis is proposed as being of direct use to disaster risk managers (DRM), civil

authorities and NGOs as an alternative tool in community outreach, exposure management, disaster mitigation and disaster preparedness plans. The contribution is also discussed in the wider context of disaster risk reduction measures, recent conceptual frameworks, and ongoing global initiatives such as the United Nations' Hyogo Framework for Action and its intended replacement, HFA2.

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Definition
DRR	Disaster Risk Reduction
GIS	Geographic Information System
HFA	The Hyogo Framework for Action (2005-2015)
IDNDR	International Decade for Natural Disaster Reduction
INEC	The Ecuador Institute of National Statistics
IPCC	Intergovernmental Panel on Climate Change
LDC	Least developed country
MDC	More developed country
OAC	Output Area Classification
PAHO	Pan American Health Organisation
PAR	Pressure and Release Model
PDC	Pyroclastic Density Current
SFDRR	Sendai Framework for Disaster Risk Reduction (2015-2030)
SoVI	Social vulnerability Index
SREX	Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation (SREX)
UNDP	United Nations Development Programme
UNISDR	United Nations Office for Disaster Risk Reduction
VEI	Volcanic Explosivity Index

1. INTRODUCTION

An estimated nine percent of the world's population lives within 100km of an active volcano (Small and Naumann 2001). In the last 100 years there have been over 96,000 fatalities from volcanic eruptions around the world (EM-DAT 2013) with economic losses during the same period exceeding 7bn USD (Annen and Wagner 2003). Despite the 1990s being declared the International Decade for Natural Disaster Reduction (IDNDR), and further progress towards disaster risk reduction (DRR) volcanic eruptions during this time were responsible for the lives of 1300 people with a further 520,000 being displaced from their home.

Figure 1-1 highlights the increasing exposure of human population to volcanic eruptions throughout the 20th century, as the number of people impacted from these events can be seen to have increased significantly despite the size and frequency of very large volcanic eruptions remaining largely constant (EM-DAT 2015). It's important to note that empirical fatality data highlights that volcanoes are not the most deadly natural phenomenon. When compared with floods, earthquakes, and storms, fatalities due to volcanoes account for less than 1% of deaths (96,000) in the period from 1900-2015 (EM-DAT 2015). While other natural hazards have been more costly to human life on a global scale over the course of the 20th century (EM-DAT 2015), the increasing settlement of human populations around volcanoes underlies a growing exposure to these violent and destructive forces. While communities at risk to the effects of climate change are of particular research interest currently, there appears little work on assessing the vulnerability of increasingly large population groups living around active volcanoes. Given the historical damage due to volcanic disasters, there appears to be little urgency in addressing this risk.

Notable disasters in the 20th century included eruptions caused by volcanoes such as Nevado Del Ruiz in Colombia (1985), Santa Maria in Guatamala (1902) and Mount Pelée on the island of Martinique (1902). Despite the commonality of destruction these volcanic eruptions left behind, the manifestation of the volcanic hazards were often very different.

The eruption of Nevado Del Ruiz occurred on November 13th 1985 following a period of increased fumarole and phreatic activity (Pierson et al 1990). The volcano forced tephra more than 30km into the atmosphere and ejected around 35 million tonnes of erupted ash and magma. The event has subsequently been characterised as having a Volcanic Explosivity Index (VEI) of 3, though the humanitarian impact of the volcano had a much more deadly legacy. The eruption melted summit glaciers causing four extremely large fast flowing mud flows to occur (more commonly known as *lahars*). With an average speed of 60km/hr, the lahars combined with existing watercourses such as the Guali River, and the resultant floods killed 23,000 people in the local towns of Armero and Chinchiná (Pierson et al 1990). Economic losses from the same event are estimated to have been in the order of 1.7bn USD (EM-DAT 2015) but the event sparked a wider debate about the importance of risk communication, evacuation measures and the role of state intervention. Despite being aware of the imminent threat the volcano posed in the weeks preceding the eruption, the Colombian government was reluctant to issue any 'false alarm'. Hence, no evacuation of the surrounding towns or villages had taken place prior to the onset of the catastrophic lahars (Voight 1990).

Such events are not in isolation. The town of St.Pierre in Martinique was completely destroyed in 1902 as super-heated pyroclastic flows from Mount Pelée engulfed the town's population, killing an estimated 30,000 people just a few minutes after the onset of the eruption. Such events remain a stark reminder of both the devastating potential of volcanoes and the urgent requirement for populations living in close proximity to such hazards to mitigate, prepare and respond to such risks.

Similarly, the loading of volcanic ash on houses surrounding a volcano can cause large scale damage and subsequent collapse. The town of Rabaul, Papua New Guinea, has provided the basis of several empirical studies assessing building vulnerability due to the repeated eruptions of the towns neighboring volcanoes Tavurvur and Vulcan (Blong 2003).

While Figure 1-1 highlights the trend of increasing human exposure to volcanic risk, it is important to be mindful that such conclusions may be overly simplistic and not without

epistemic uncertainty owing to issues in historic epidemiology data sources. The most credible disaster event database is the *Centre for Research on the Epidemiology of Disasters* (CRED), maintained by the University Catholique de Louvain, yet given the large uncertainty associated with historic data from developing nations, disaster information should be treated with some caution.

In terms of human exposure, David Chester's work (Chester et al 2000) highlights the hazardous relationship of populations and volcanoes by showing the high correlation between large cities and volcanoes that have been active in the last 10,000 years. In doing so, it demonstrates the spatial and cultural diversity of these population settlements. It draws attention to the fact that many of these cities are in developing nations that have seen rapid population rise in the last 50 years. Such rises have often been the result of increasing globalization and state-led political doctrine, leaving them extremely vulnerable to the physical and social risks a large volcanic eruption could entail (Chester et al 2000).

Such findings are not solely reserved to volcanic risk. In consideration of natural hazards more generally, and due to the occurrence of particularly high profile disasters such as the Indian Ocean Tsunami in 2004, has led governments and academics to pursue more structured approaches to crisis management.
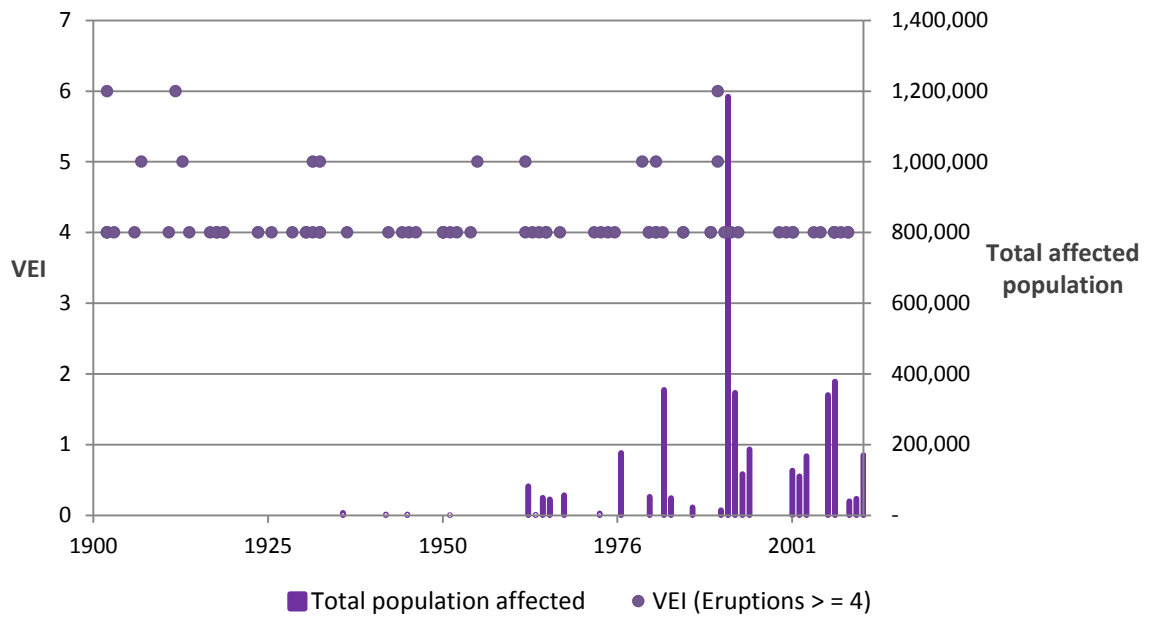
FIGURE 1-1 - A COMPARISON OF POPULATIONS IMPACTED BY VOLCANOES THROUGH THE 20TH CENTURY MAPPED AGAINST GLOBAL ERUPTIONS WITH A VOLCANIC EXPLOSIVITY INDEX (VEI) GREATER THAN 4 (CRED 2013)

One of the major initiatives to come out of disaster prevention programs in recent years has been the United Nations' Hyogo Framework for Action (HFA) which is set to run until 2015. This internationally recognised framework is focused around disaster reduction and was created to more carefully direct funds and research for the purpose of DRR. Its key aim is to make communities more resilient to natural hazards. Among the priorities and goals of the framework has been a ten-year plan (2005-2015) of action to improve existing disaster risk methodologies and management strategies. With 168 countries signed up to the United Nations International Office for Disaster Risk Reduction (UNISDR), the key priority of Hyogo was to "*reduce the underlying risk factors*" that lead to natural disasters (HFA 2005). As Hyogo came to an end in March 2015, the UNISDR is already underway coordinating priorities for its replacement, the Sendai Framework for Disaster Risk Reduction (SFDRR) that took place at the World conference on disaster risk reduction in Sendai, Japan, 2015. SFDRR is seeking to go further than HFA and place a greater focus on the minutiae, emboldening some of the previous guiding

principles of HFA whilst lowering the practical constraints that has stopped previous attempts at DRR. Four priority areas have identified in SFDRR to focus on;

- Priority 1: Understanding disaster risk.
- Priority 2: Strengthening disaster risk governance to manage disaster risk.
- Priority 3: Investing in disaster risk reduction for resilience.
- Priority 4: Enhancing disaster preparedness for effective response and to "Build Back Better" in recovery, rehabilitation and reconstruction.

The priorities are aimed at building on the success of HFA, with a set of common standards, a framework with targets, and a more legally-based instrument for disaster mitigation strategies (HFA 2013) whilst also recognising that more can be done. A key element that Hyogo sought to address and which remains a moot point has been the apparent lack of funding that governments invest in DRR. This was an issue raised at the *Rio +20*, the United Nations Conference on Sustainable Development in Brazil (UNCDS 2012). It would seem evident that this will also remain a key point throughout the lifetime of SFDRR as well.

However, during the course of the HFA framework, DRR has become of increasing importance to governments, NGOs and academics. As countries seek to identify, understand and ultimately manage their risk, there has been a large amount of research to more precisely identify the 'at risk' population groups (on both a local and state level). Numerous quantitative and qualitative methods have been developed during this time to model these processes. For example, antecedent conditions that lead to a "*disaster*" have been well chronicled by conceptual frameworks such as the *Risk-Hazard* and *Pressure and Release* models (Turner et al 2003; Blaikie et al 1994). Likewise, in the progression of quantitative methods, to which this thesis can be considered to be closely aligned, there has been a plethora of research activity on classifying natural hazard social vulnerability (Cutter et al 2003; Cutter et al 2000; Rygel 2006; Elliot and Pais 2006; Clark et al 1998; Davidson 1997; D'Ercole 1996; Dibben and Chester 1999; Dwyer et al 2004; Marti et al 2008; Spence et al 2004). Such research has varied greatly in scale, methodology, and purpose.

Initiatives such as the World Bank *Hotspots* project (Dilley 2005) or the IDEA's *Americas indexing programme* (Cardona 2005) attempted to classify vulnerability at global and continental levels. More recent applications have been national or sub-national but have tended to be location or hazard specific (Tobin and Whiteford 2002). Some have quantified and assessed vulnerability based on specific social variables, such as ethnicity (Elliot and Pais 2006) or population density (Small and Naumann 2001); whereas others have focused on classifying risk to specific natural hazards, such as volcanoes or earthquakes (Marti et al 2008, Spence et al 2004). Many of these areas of research have been successfully applied to small and very specific geographic areas or were reflective studies of recent natural disasters. However, despite these research interests taking note of the numerous factors that give rise to social vulnerability, recent critique of such offerings has highlighted an urgent need to more fully describe vulnerability within the site specific context and culture in which a population lives (Fuchs et al 2011; Fekete 2012).

The development of the Social Vulnerability Index (SoVI) recipe by Susan Cutter provided a new paradigm in quantifying social vulnerability using Principal Component Analysis (PCA) on US census variables (Cutter et al 2003; Cutter et al 2000). The method proposed a transferrable technique though it's important to note that Cutter's work has been heavily focused on classifying risk in the US and not in the world's least developed nations, where natural disasters tend to have the greatest impact. Methodologies based on highly evolved risk classification may be adept at ranking 'risk indicators' or census zones but a fundamental caveat is that it does not necessarily make them useable to DRR practitioners and those seeking to ultimately mitigate disaster risk. Facilitating NGO workers and civil authorities with the practical tools they require to undertake community outreach and risk mitigation are not always a primary consideration in academic research.

It is also worth examining the way in which many nation states react, communicate and seek to mitigate their exposure to disasters. As well as the losses and stress from a humanitarian perspective, natural disasters (including volcanoes) very often have high economic consequences that can greatly impact a country's sovereign wealth. Despite ongoing global initiatives (e.g. HFA, SFDRR) there has been a lack of traction with DRR among developing nations. Kellett and Sparks (2012) estimate that natural disasters from 2000-2009 have cost 840,000 lives and cost an estimated $891 (USD) billion. Further to this, it is currently estimated that for every $1 (USD) spent on disaster mitigation and prevention, $7 are saved from the cost of future disaster response (UNDP 2012). Therefore, there appears to be a strong business case for increased mitigation among such nations.

With this mind, there is an increasing amount of financial instruments that developing nations are turning their focus on with regards to *risk transfer.* Catastrophe pools, insured-linked securities (ILS) and catastrophe re/insurance are becoming commonplace in many of the world's developing nations. With many of these bonds and insurance schemes based on natural hazard models, it is important to note the contribution that *catastrophe modelling* has within re/insurance industry.

Insurance has become the most standard method of risk transfer regarding the capacity of an individual, organisation or state to cope with the economic loss of a natural disaster (Grossi and Kunreuther 2005). In the last 20 years, there has been a significant increase in catastrophe risk modelling. The insurance, reinsurance and capital markets now have several competing software modelers that are used to help manage natural disaster market exposure on a global basis. EQECAT, RMS and AIR Worldwide (Grossi and Kunreuther 2005) are the three main software vendors providing both the local installation of software and more recently, cloud capability (e.g. RMS one) for clients to import, analyse and model their insured portfolio to numerous natural disasters in multiple territories. Based on the model's loss output, and key

industry metrics, underwriters are then able to price an acceptable level of premium for their clients to cover the risk of these disasters.

Though the quantitative assessment of earthquake probability (Algermissen 1969) and forecasting of hurricane events by the National Oceanic Atmospheric Administration (NOAA) had already been underway (Neumann 1972), it was not until the onset of several industry defining disasters that catastrophe modelling began to gather momentum. Hurricane Hugo in 1989 caused widespread damage in South Carolina, costing an estimated $4 billion USD. Likewise, a month later in Loma Prieta, San Francisco; an earthquake causing $6 billion USD in damages occurred (Stover and Coffman 1993). Similarly, and more significant in overall loss was hurricane Andrew in 1992. Making landfall in Florida, the insured damage to property and contents reached in excess of $15.5 billion in claims (Property Claims Services 2014). This latter event was particularly damaging to the Lloyd's insurance market of London. It highlighted Lloyd's over-exposure in the US hurricane market and a lack of best practice as several syndicates were rendered insolvent. Likewise, these events proved to be something of a benchmark as primary reinsurers became aware they could not always underwrite the entire loss of these catastrophic events. Hence, reinsurance and catastrophe bond treaties became increasingly common in the following years (Grossi and Kunreuther 2005).

From a catastrophe modelling perspective, however, these events provided strong evidence that underwriting practice needed greater scientific and empirical judgment in establishing the level of exposure they were prepare to underwrite. Since then, catastrophe modelling technique, practice and validation has progressed significantly. Such is the growth in this sector, it is now common place for insurers, reinsurers, and catastrophe model software vendors to have dedicated teams of meteorologists, seismologists, statisticians and GIS specialists to help analyse a company's exposure to natural hazard risk (Grossi and Kunreuther 2005).

As insurance penetration varies greatly from country to country, the insured market to natural disasters is polarised in certain regions. For example, though a natural hazard may be extremely destructive and dangerous to human life in a particular region (e.g. flooding in Bangladesh or

earthquakes in Pakistan), the insured market may be very low. Therefore, it is notable that the leading catastrophe model software vendors, insurers, and reinsurers focus on key market territories, and perils. As Figure 1-2 highlights, the key insurance losses for natural disasters are largely based around the US, Europe and Asia. However, as also shown in Figure 1-2, recent loss trends (2009-2011) suggest that the traditional catastrophe insurance markets are changing significantly, with emerging markets also using risk transfer techniques (such as catastrophe modelling) to manage their exposure. The asian market in particular has grown rapidly during the last few years.
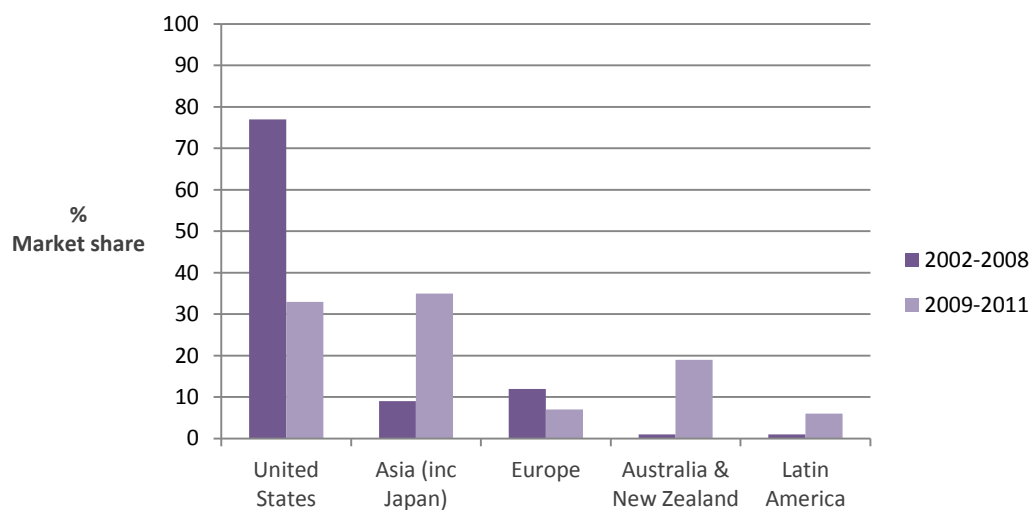
FIGURE 1-2 - NATURAL CATASTROPHE INSURED LOSSES BY REGION (SWISS RE 2014)

The key perils for the leading catastrophe model vendors are US hurricanes, US earthquakes, European Windstorm, Japan Earthquake and Asia Typhoon. There are numerous other perils insured across the world but these are considered to be the main territories.

The stimulus for the recent increase in risk modelling capability has been a mixture of technological advances in computer functionality, advances in natural hazard research, the availability of good quality exposure data and most importantly, the economic incentive to reduce the cost of recent disasters to insurance industries and governments.

This last point is clearly illustrated in Figure 1-3 by Grossi and Kunreuther (2005) as they draw attention to the rising cost of natural disasters in the US over the last 50 years. As world population growth has increasingly put pressure on natural resources, many developing nations now have many more people living in harm's way of volcanoes, flood plains, and earthquake zones. Aside from the overall rise in the economic cost of natural disasters, year-by-year, there are large seasonal changes as most natural hazards remain highly unpredictable in their occurrence.
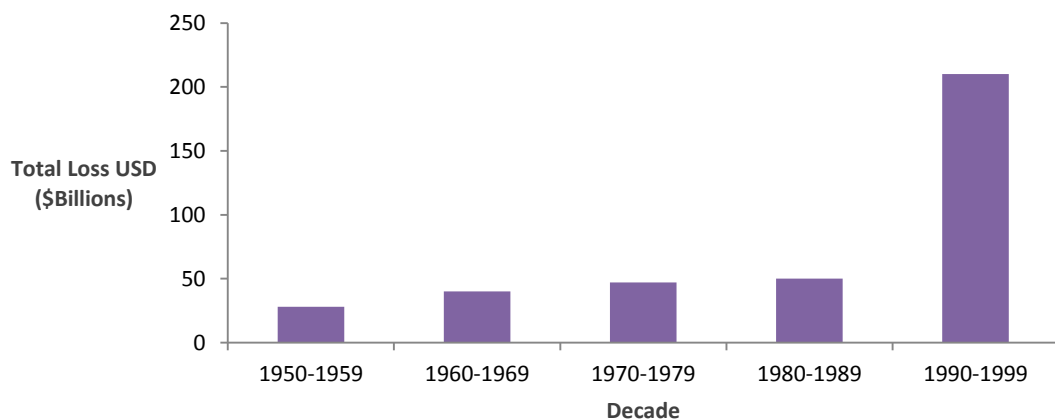
FIGURE 1-3 - ECONOMIC COST OF NATURAL DISASTERS TO THE CAPITAL MARKETS IN THE US (GROSSI AND KUNREUTHER 2005)

Despite the difficulties in predicting natural hazards, catastrophe modeling is based around the process of creating several key industry metrics that are used by the underwriter to assess risk. The most critical of these is the *exceedance probably* (EP) curve (Kozlowski and Mathewson 1995). This non-linear relationship is shown in Figure 1-4 and is based on an empirically derived hazard frequency for a given natural peril (e.g. 1/number of return period years) and the estimated loss that such an event could have on an insured market portfolio.
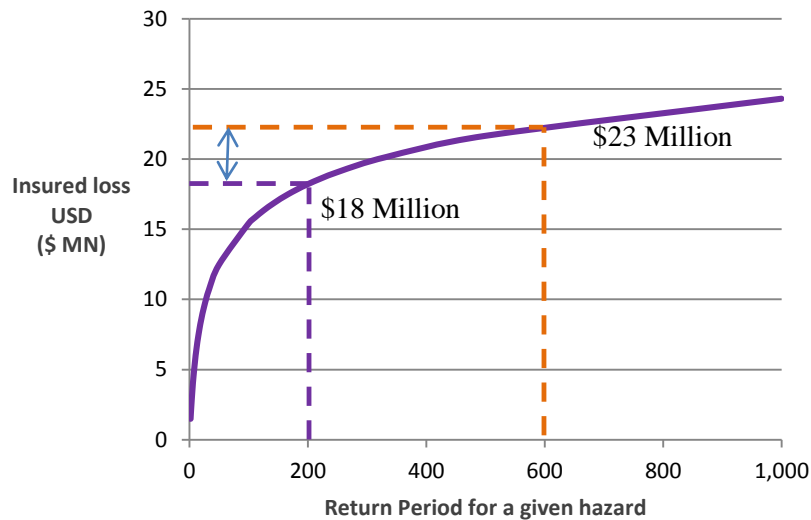
FIGURE 1-4 - EXCEEDANCE PROBABILITY CURVE – (AUTHOR'S ILLUSTRATION)

Figure 1-4 highlights how the EP curve can be used by underwriters to establish the basis of their company's risk. For example, *Company X* decides it has $18 million in capital that it will use to underwrite hurricane risk in Florida and by using the EP curve they can establish that they can write insurance coverage upto a 1 in 200 year hurricane event. Similarly, if they choose to have a subsequent reinsurance treaty that will cover their exposure for another $5 million, they may be able to increase their exposure to a 1 in 600 year event.

To provide these metrics, catastrophe models essentially run off four main components; *hazard, exposure, vulnerability* and *loss*. The *hazard* refers to the specific natural hazard, it's average periodicity and the spatial extent this risk presents; the *exposure* refers to the amount of the housing stock, the number of businesses, and the people that would be impacted by the hazard in a given area; the *vulnerability* component in these models is defined as the ratio of hazard intensity to loss, based on the engineering performance of the building and likely failure mechanisms. Based on the relative distribution of the hazard intensity and inherent vulnerability differences (e.g timber frame vs reinforced concrete) a distribution of loss is calculated. The relationship of these components can be seen in Figure 1-5.
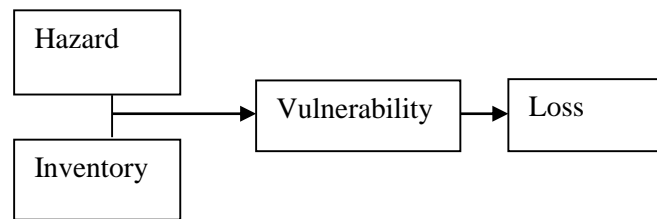
By modelling natural disasters at varying scales and then comparing these events to their annual expected frequency (e.g. the average number of category 4 hurricanes per year), an insurer is able to gauge their expected losses on an exceedance probability curve.

It should be noted that before Hurricane Hugo (1989), Hurriance Andrew (1992) and the Northridge earthquake (1994), a natural disaster had never cost more than $1 billion to the insurance industry. However, with the increased take up of insurance and population settlement changes the catastrophe insurance industry has radically changed. These events were estimated to have cost a total of over $30 billion in damages (Grossi and Kunreuther 2005).

VOLCANIC INSURANCE RISK

With specific regard to volcanoes, it is extremely difficult to estimate insured loss as the effects of volcanoes are often indirect, particularly diverse and complex to assess (Annen and Wagner 2003). Likewise, separating out the insured loss from the economic cost can be difficult given the structuring of insurance policies and diverse ways in which a volcanic eruption can affect infrastructure. The 2010 Eruption in Iceland had an economic loss of $1.7 billion USD (to the airlines) yet the full insured loss has yet to be disclosed.

Despite these diffculties Table 1-1 provides a good introduction to the scale of economic losses of previous large volcanic eruptions. It's noticable that the financial loss of volcanic eruptions is highly correlated to the economic status of the resident nation where the eruption occurred.

As Table 1-1 highlights, though Mount Pinatubo's (Philipines) economic loss was less than 10% of Mount St.Helens in the USA, nearly 800 more lives were lost during Mount Pinatubo's 1991 eruption. This provides a stark reminder that the geographic location of volcanoes is still the driving force behind many hazard risk, preparedness, and response initiatives.

TABLE 1-1 - WORLD BANK DAMAGE ESTIMATES. (EM-DAT 2007)

Year	Volcano	Country	Damage in USD $ (millions) (2007)
1973	Eldafjell	Iceland	93
1980	Mount St. Helens	United States	3327
1982	Mount Galunggung	Indonesia	306
1982	El Chichon	Mexico	224
1983	Mount Gamalama	Indonesia	275
1985	Nevado Del Ruiz	Colombia	1719
1991	Mount Pinatubo	Philippines	300
1994	Rabaul/Tavarvur	Papua New Guinea	531
1996	Grimsvotn	Iceland	21
1997	Soufriere	Montserrat	10
2001	Etna	Italy	4
2002	Stromboli	Italy	1
2006	Tungurahua	Ecuador	154

Funded by the Willis Research Network (WRN), Spence, Gunesekara, and Zuccaro (2010) researched the insured loss potential of multiple European volcanoes, identifying Vesuvius and Etna as being among the most dangerous. They estimated potential economic losses for a Sub-Plinian eruption of Vesuvius as being in the order of $24 billion (taking into account Residential loss potential only).

DISASTER RISK REDUCTION

Returning to the literature on vulnerability, recent research has helped highlight the notion that studies based on quantitative or qualitative research exclusively cannot provide a panacea for disaster risk reduction. Similarly, it remains increasingly important to recognise the

contribution that indigenous knowledge can provide to DRR. Kelman, Mercer, and Gaillard highlighted three key lessons from their case study research of Papua New Guinea and the Philippines (Kelman at al 2012):

- Requirement to understand contextualization, transferability and non-transferability of (DRR) knowledge

- Promote trust of different knowledge forms and self-help

- Do not assume community homogeneity

There is a growing acknowledgement of the importance to understand the context of a community within an 'at risk' area. Indigenous knowledge provides an invaluable source of information to best address disaster risk reduction but is often not transferrable to other perils, cultures or geography regions. For example, the indigenous community of Simeulue (Sumatra) reacted to the 26th December 2004 earthquake by climbing to higher ground, thus evacuating the coastal region and avoiding mass casualties from the devastating tsunami that ensued (Shaw et al 2008). Such knowledge and adaptive behavior was a direct consequence of an oral tradition to pass such information on to successive generations within the community. Whilst appropriate and valid in the context of Simeulue, such knowledge would not necessarily translate to protection against other perils, such as volcanoes or tropical cyclones, both of which would require different adaptive strategies.

Though not the single focus of this thesis, it is also noting the relative debate in the literature on disaster risk reduction more widely. Researchers such as Bankoff have questioned the definition of a disaster, and in particular, how developed nations are prone to systemic labelling of such events as an affliction of 'poverty' and 'disease stricken' areas of the world (Bankoff 2001). Bankoff raises the question of whether vulnerability is a separate entity, or whether it cannot be truly disentangled from 'development' and 'tropicality', two factors inextricably linked to the creation of vulnerable populations.

Likewise, another debate on disaster reduction regards the conflict that exists in communities that co-exist with tribal belief systems, orthodox religions and scientific knowledge about the nature of the hazard. Chester draws out this paradox in his research into 'Theology and disaster studies' (Chester 2005) and the way that some disasters are labelled as 'Acts of God' and others are considered a consequence of human frailty. During fieldwork undertaken for this thesis, this latter point was noted repeatedly following interviews with local Quitenos in Ecuador. The co-existence of their Catholic faith, pre-Colombian beliefs that neighbouring volcanoes were related deities and a basic understanding of plate tectonics were not considered to cause conflict.

In terms of the contribution of this thesis, it can be seen to focus on the application of social vulnerability assessment. It is proposed that 'neighbourhoods' can be segmented to define social vulnerability rather than as the classification of discrete, disconnected statistical units. This thesis is a proponent of the philosophy to not "assume community homogeneity". The inherent complexities of what constitutes one area to be more socially vulnerable to a natural hazard than another area may be understood at a community level rather than simply as a weighted social classification. Our risk perceptions, attitudinal concerns, and risk behaviour cannot be removed from the ties of our local community. As discussed further in Chapter 2, social vulnerability is very much a product of our environment. With this paradigm, it is proposed that insight can be leveraged to gain a better understanding of the attitudinal and behavioural aspects of human disaster response, resilience and risk perception.

It is worth drawing attention to the fact that *Volcanic vulnerability* is a term discussed repeatedly within this thesis. Whilst this notion is treated as a distinct term, and with sole application to volcanic risk, many of the concepts at the core of the methodology proposed in this thesis are interchangeable with social vulnerability research more generally, and could be considered transferrable to volcanic risk. Similarly, the final methodology and techniques proposed in this work could be partly transferrable to other perils and are not considered to be exclusive. However, given the site specific nature of volcanic risk and the settlement pattern of

large populations around active volcanoes, it was important to keep the notion of volcanic vulnerability as a distinct notion in developing the methodology.

A multivariate methodology is proposed here for the purposes of disaster mitigation and community outreach to neighbourhoods at risk of volcanic hazards but wider consideration is afforded to other natural catastrophes where the techniques outlined in this thesis may also have meaningful application (e.g. earthquakes, floods, windstorms).

1.1 AIMS

In developing this thesis, research was initially focused on investigating current risk classification methodologies and existing practice within the broader disaster reduction framework. As the title of this thesis alludes to, volcanic regions have provided the context for this research given large historic eruptions and the increasing exposure of population settlements in close proximity of major volcanoes. With a view to developing an alternative methodology, and to focus research, the following four key aims were identified:

1. *Investigate and review the existing capability of risk classifications in DRR more broadly.*
2. *Develop an original methodology that can be used to assess social vulnerability based on classifying neighbourhoods in volcanically active regions.*
3. *Apply and validate the methodology (outlined in aim #2) in different spatial and cultural settings.*
4. *Assess whether the methodology could be transferrable and how it could be used in practical DRR outreach and communication.*

Prior to commencement of any practical assessment or methodological work, it was first proposed to consult the literature to understand the key drivers of social vulnerability during disaster experience and research how any relationship between such indicators could be drawn with neighbourhood classification systems. On this basis, conceptual frameworks were reviewed and assessed to help define the relationship between social vulnerability indicators

and geodemography. Following this, the assessment of social vulnerability using geodemographics in volcanically active regions (or study areas) could be addressed in two phases:

- The assessment of an *existing/commercial* neighbourhood classification system.

- The assessment of a neighbourhood system created from original census output data.

By comparing model results in volcanically active study regions to anecdotal evidence of vulnerable populations during historic eruptions, the benefits of using neighbourhood classification systems for volcanic vulnerability assessment could be assessed. Likewise, the testing of both an existing and bespoke geodemographic solution would likely identify any differences in the practical constraints of applying such models to DRR.

Regarding the further aims of this research agenda, and whether neighbourhood classification systems could be successfully applied across different spatial and cultural landscapes, it was important to consider if such a vulnerability model (and its methodology) could be both transferrable and scalable. To help assess this question, two very different geographic areas would be required. With large variations in geography, culture, socioeconomics, and data availability, Italy and Ecuador were selected as appropriate regions for assessment.

1.2 DOCUMENT STRUCTURE

This thesis presents research towards a PhD on the application of geodemographics in volcanic hazard risk assessment.

The thesis is split into seven chapters. The first section begins with a narrative background to the research interest and the fundamental contribution of this research within the global context of work towards DRR. This is followed by chapter two, which provides a review of the literature, including key elements of this research such as social vulnerability, risk perception, and geodemographics. It should be noted that there is also due consideration to fringing research interests and commercial aspects of risk transfer such as catastrophe risk insurance.

The third section documents the nature and occurrence of volcanic hazards, specifically detailing their geophysical origin and the risk posed to human life. Chapter four is the methodology section, which provides an account of the steps taken towards the application of multivariate statistics in both risk assessment and quantification, and which can be considered as the primary contribution of this thesis.

Chapter four is split into two distinct phases of development, based on divergent study regions, with a commercial geodemographic used in the first part of the methodology section (*phase 1*) and the second part focused on the creation of a bespoke geodemographic from Ecuadorian census data (*phase 2*). Chapter five presents both tabular and graphical results from this research and highlights aspects of model validation. Chapter six provides a worked example of how this contribution could theoretically be applied for the purposes of both disaster mitigation and community outreach. Chapter seven discusses the results from the two study regions (Ecuador and Italy) as well as the relative strengths, applications and limitations of this work within the wider DRR framework. Consideration is given to the global context, practical implementation of such initiatives to NGOs, model transferability, and the technical concerns and considerations of applying this methodology.

1.3   RESEARCH AREAS

Research interests during this thesis have included contributions from many diverse areas, not always well-aligned in the literature. Development of this thesis can be considered to be multi-disciplinary in its origin. Though not all of these interests are mentioned in detail during the literature review, the core subject areas, and discussed at length in chapter 2 include: volcanic hazards, social vulnerability, risk perception, quantitative risk assessment, factor analysis, data reduction, DRR, risk transfer and Geographical Information Systems (GIS).

GIS is implicit to this research as it has provided much of the analytical capability as well as facilitating map production and the communication of model results. Although a necessary

mechanism binding these disparate research and methodological areas, it should be noted that GIS is only discussed within this thesis with regard to its practical use in conveying and communicating disaster risk to key audiences. However, its contribution cannot be overstated as it can be considered to be at the convergence of several other research interests.
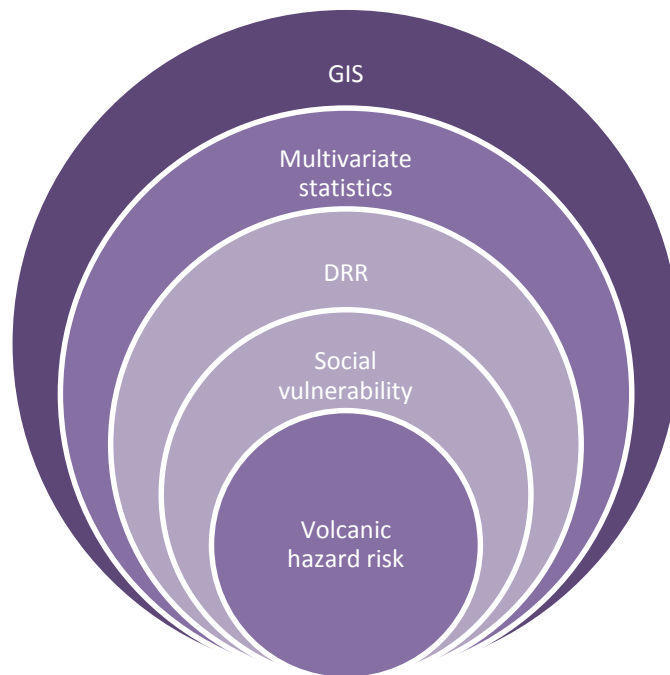
FIGURE 1-6 – STACKED VENN DIAGRAM OF THE PRINCIPLE AREAS OF THESIS RESEARCH

Figure 1-6 provides an illustration of these divergent research interests and how they overlap within the context of this study. Vulnerability is discussed here in both its universal meaning and with regard to its specific application to the 'at risk' population (social vulnerability). This is to help provide context to its use in this research and in assessment of human exposure. Likewise, risk perception is also considered within the framework to comprehend how people view natural hazards in their everyday lives and the various societal and cultural influences that impact communities. Multivariate statistics is an umbrella term to describe the various statistical tests and quantitative methods that are at the core of this thesis and ultimately provide the backbone of the methodology and study region applications. This includes a broad range of

parametric and non-parametric approaches in applied statistics and therefore it is important to note that this work was solely focused on just a few of these specialised practices. The following literature review discusses the derivation and etymology of neighbourhood classification systems, which have become of increasing commercial and practical interest in recent years. The quantitative techniques employed in their creation are considered (e.g. standardisation, cluster analysis, and propensity indices) as well as the strengths and limitations of their principle data sources (e.g. census data, telemarketing surveys).

Volcanic hazards are very wide ranging and although light discussion is given in this thesis to the long term and disparate effects of volcanic aerosols on climatic conditions and distal populations, the principle basis of research and modelling in this thesis was focused on the proximal volcanic hazards that could impact cities next to volcanoes. In this regard, the literature review considers the following four key volcanic hazards:

- Ash fall (often termed *tephra*),

- Pyroclastic flows (sometimes termed *gravity flows*)

- Lahars (destructive torrents of mud and water)

- Volcanogenic earthquakes

It is worth mentioning that significant consideration is given in this thesis to the practice of catastrophe modelling, which provides a mechanism for nation states and the insurance industry to manage the impact of disaster risk to many natural hazards. Although volcanic risk is not currently a principle risk for the insurance industry (in comparison to hazards such as earthquakes and hurricanes), notably large eruptions in recent years have had profound financial effects on the markets and are worth further discussion (e.g. the Icelandic eruption of Eyjafjallajökull in 2010). Catastrophe modelling is discussed within the framework of risk transfer and as a financial instrument that states and capital markets use to manage their

exposure, with primary insurance, reinsurance, insurance-linked securities (also known as cat

bonds) and state disaster pools.

2.   LITERATURE REVIEW

This chapter begins with reviewing the definition of a volcano and how the term itself encompasses a variety of tectonic forms. A description is then provided of the various earth processes that lead to the formation of volcanoes on the earth's surface and the current displacement of these edifices along existing plate boundaries. The next section identifies the numerous direct and indirect hazards caused by volcanoes as well as a comparison of their severity with regard to other natural hazards on a global basis. Whilst this chapter does not provide a detailed account of every possible volcanic hazard, discussion is prioritised to the direct hazards most likely to impact populations living in close quarters of volcanoes.

The chapter then explores and reviews the literature on a wide range of topics pertinent to the core contribution of this thesis. It includes a detailed discussion on the definition of vulnerability and social vulnerability. Various concepts, models, and frameworks are then reviewed that have been developed to understand the antecedent conditions that lead to vulnerability and possible mitigation strategies. During this discourse, a review is provided of the various methods to quantify social vulnerability in hazardous locations, including volcanically active regions of the world. A section is provided on the global risk transfer market as it provides valuable insight to how disaster risk is managed in developed nations, and increasingly, in several developing countries. Lastly, the contribution of multivariate statistics is discussed with regard to data reduction methods such as cluster analysis. Given the substantial contribution of this in later chapters, it is important to cover the background, development and previous applications of this technique in the literature.

2.1 VOLCANIC HAZARDS

2.1.1 DEFINITION OF A VOLCANO

The literal definition of a volcano is provided as follows; "*A mountain or hill, typically conical, having a crater or vent through which lava, rock fragments, hot vapour, and gas are or have been erupted from the earth's crust*" (OED 2013, p.1619)

Although the Oxford English Dictionary definition will help classify volcanoes that consist of a single vent and source, in reality, many volcanoes have multiple vents, or share the same source of magma. For example, the Ngauruhoe volcano is regarded by many as a secondary vent of the Tongariro volcano in New Zealand, yet they have two very distinct volcanic cones (Rothery 2007). Likewise, the term 'cone' can be misleading as many volcanoes do not have a defined conical shape. The volcanic complexes of Yellowstone in Wyoming or Kilauea in Hawaii are good examples of such divergence in characteristic shape.

The USGS provides a more technical definition of a volcano that seeks to place greater emphasis on the tectonic depth of the magma chambers that feed the complex; "*A volcano is a structure containing a vent or cluster of vents fed by magma rising directly from great depth within the earth, generally more than 30 km (18 miles) and in Hawaii more than 100 km (60 miles)*" (USGS 1999)

Likewise, while volcanic formations may consist of subterranean vents or sea mounts, such as those found along the mid-Atlantic ridge or the perpetually effusive volcanoes of the Hawaiian islands, the volcanic hazards discussed in the context of the research presented here are primarily concerned with areas where large human settlements have located around volcanoes with a long history of associated explosive and life-threatening geophysical hazards.

2.1.2   STRUCTURE OF THE EARTH

To understand the origin of volcanoes, it is first required to comprehend the compositional layering of the earth's interior. Earth's interior structure is concentrically layered with distinction between four fundamental components; the inner core, outer core, mantle and crust. The structure of these layers is shown in Figure 2-1, with the inner core at the centre of the earth. The inner core is under extreme pressure and remains in a solid state, whereas the outer core is molten. The principle elements in the composition of the earth's core are iron with a small proportion of nickel (Decker and Decker 2005). However, most of the earth's mass is derived from the mantle, the layer surrounding the core and which is less dense. The most abundant elements in the mantle are silicon and oxygen, combining with other elements such as magnesium and iron to form a number of different silicates. At over 1000 degrees in temperature, the mantle is in a solid state but can deform slowly in a plastic nature.
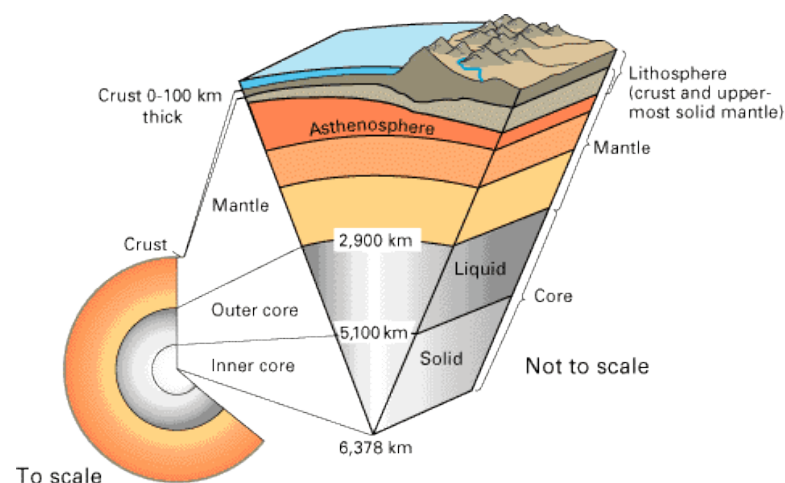
FIGURE 2-1 – EARTH'S INTERIOR (OPENHAZARDS 2015)

The crust overlays the mantle and can be regarded as a thin skin, ranging in thickness from 6-11Km over the oceans and 25-90Km over continent (Decker and Decker 2005). Given that the crust is much richer and varied in the silicates present, the rocks have a noticeably different composition to the underlying mantle and core. Oceanic crust is typically composed of basalt

whereas continental crust is highly variable, typically with a greater amount being derived from igneous and metamorphic rocks. On average, continental crust comprises of 62% silicon whereas oceanic crust is 49% (Rothery 2007).

2.1.3   MAGMA GENERATION

As well as the relative composition of a volcanoes source rocks within the earth's interior, several factors involved are in the creation of magma that provide the fundamental driving force for an eruption to occur. The melting of silicate rocks to generate magma can be considered a direct consequence of three key mechanisms; heating, decompression, and hydration (addition of water).

If the temperature of a rock rises (at a given pressure and rock composition) sufficiently beyond the solidus, the rock will begin to melt. Rock temperatures in the earth are largely controlled by the geothermal gradient and the radioactive decay within the rock. Such temperature are found to have a range of 5-10 degrees/Km around subduction zones and 30-80 degrees/Km under mid-ocean ridges and volcanic arcs (Decker and Decker 2005). As buoyant magma rises, and the rock temperature falls through adiabatic cooling, it will liquefy and form lava if erupted. Decompression melting can also occur due to a rapid decrease in pressure and the upward movement of solid mantle. It is understood that such melting is critical in the creation of oceanic crust at mid-ocean ridges as well as intraplate volcanic regions attributed to mantle plumes found in Europe, Africa, and the Pacific.

As magma cools, fractional crystallization occurs, whereby different crystals melt at different temperatures, thus resulting in a residual melt that will differ in composition to its parent magma. For example, a gabbroic magma may partially melt to form a granitic magma; these magmas thus have a very different composition and melting point.

2.1.4 OCCURRENCE AND DRIVERS OF GLOBAL VOLCANISM

Volcanoes are predominantly found at convergent or divergent tectonic plate boundaries. They can also be found where there is a mantle plume rising in the middle of a tectonic plate (e.g. Hawaiian Islands) but as Figure 2-3 shows, the greatest density of global volcanoes are found at the linear or arcuate belts of active tectonics. Lateral density variations in the mantle are understood to give rise to convective forces in the earth's interior although there is on-going debate as to the exact derivation of plate motion (Conrad and Lithgow-Bertelloni 2002). It was historically understood that mantle dynamics was the sole driver of plate tectonics but where subduction is not occurring, factors such as gravity and earth's rotation are understood to also play a key role (Conrad and Lithgow-Bertelloni 2002).

The principle mechanism of mantle dynamics revolves around two components that cause the plates to move. Firstly, basal drag provides plate motion driven by friction between the convection currents in the asthenosphere and the rigid overlying floating lithosphere. This happens concurrently to slab suction (or gravity) where convection currents exert a downward frictional pull on plates at ocean trenches. Slab suction may occur in a geodynamic setting where basal tractions continue to act on the plate as it dives into the mantle. There is also debate as to the convective forces that drive continental drift in the first instance - current uniform models highlight adiabatic processes of convection (and hence a uniform earth interior). These seem in contrast to earth measurements in recent geodynamics that suggest a more asymmetric structure (Conrad and Lithgow-Bertelloni 2002).
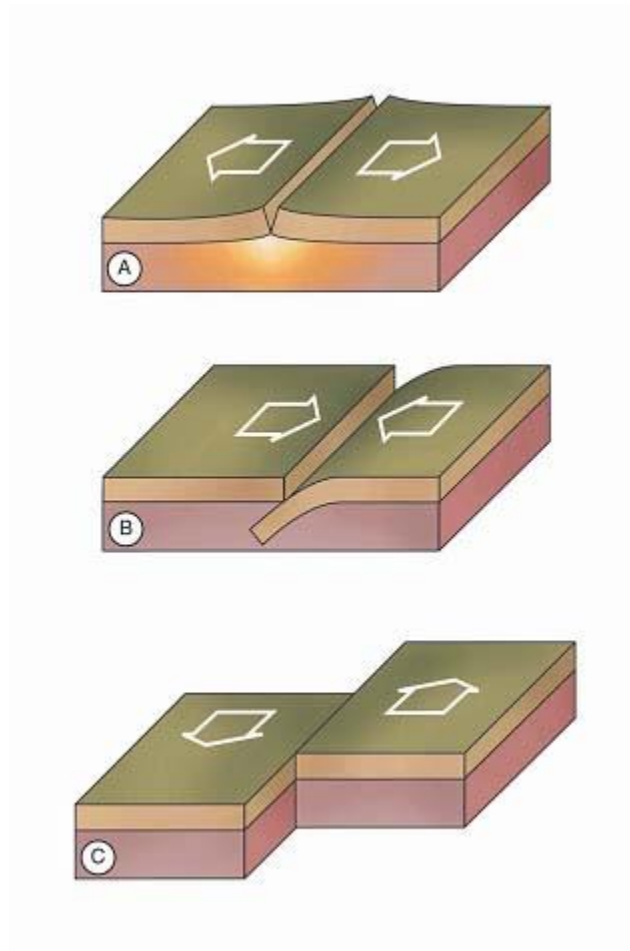
FIGURE 2-2 - PLATE TECTONIC BOUNDARY TYPES A) DIVERGENT B) CONVERGENT C) TRANSFORM (FUB 2014)

Though the origin of convective forces may be debated, the consequences of such plate movements are better understood (Rothery 2007). As lithospheric plates move across the top of the more fluid asthenosphere, their continual collision results in convergence, divergence and sometimes moving parallel to each other (transform movement). At divergent plate boundaries (such as the Mid-Atlantic ridge), the thin oceanic lithosphere becomes stretched, allowing magma to rise to the surface and form new oceanic crust. This process is shown in Figure 2-2 (a). At convergent plate boundaries, the denser oceanic plate is subducted beneath the continental plate and plunged deep into the asthenosphere, also shown in Figure 2-2 (b). Similarly, the rich silicates of the oceanic crust are released in the mantle through melting allowing for explosive and violent volcanic eruptions to occur.
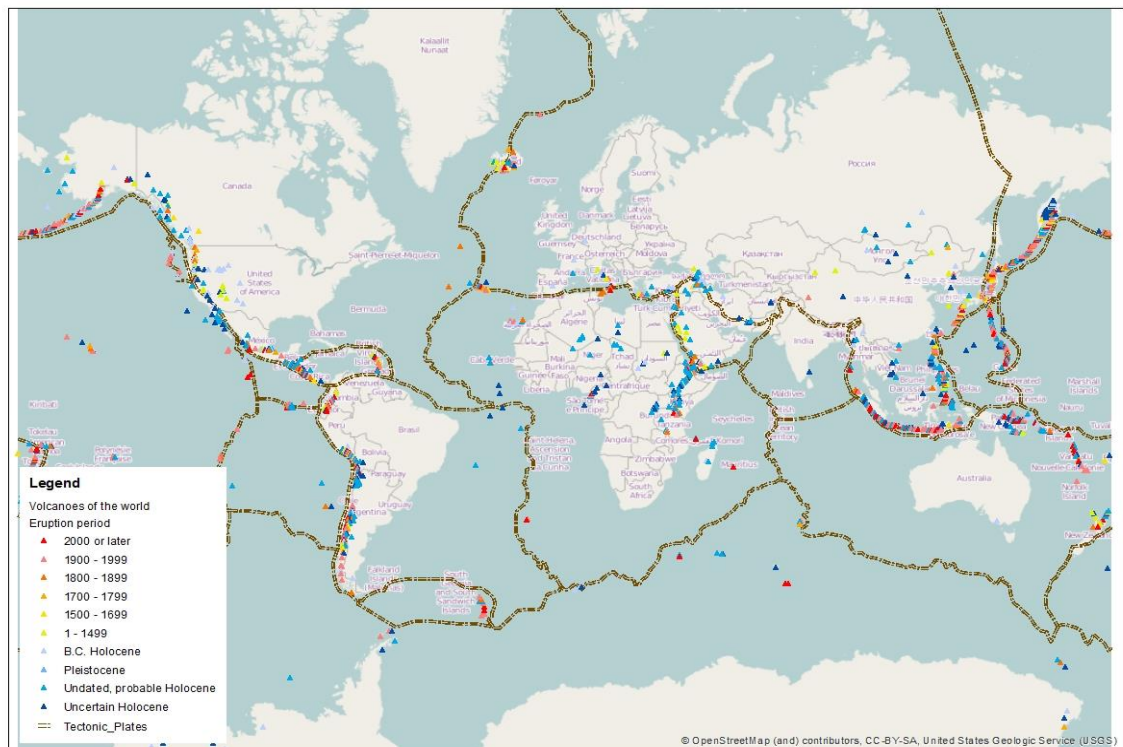
FIGURE 2-3 - GLOBAL VOLCANOES AND ERUPTION PERIOD (DURING THE HOLOCENE – 10,000 YEARS BP) (SMITHSONIAN INSTITUTE 2014)

Figure 2-3 highlights the global coverage of volcanoes that have been active in the Holocene (10,000 years to BP) and draws attention to the high correlation of these edifices with earth's tectonic plate boundaries. Volcanoes created at these destructive boundaries have two general origins:

1) Oceanic plate subducting beneath another oceanic plate

2) Oceanic plate subducting beneath a continental plate

Examples of *oceanic-oceanic* subduction include the Indonesian island arc and the Caribbean volcanoes of the Lesser Antilles. For example, the Indian-Australian plate is being subducted below the Indonesian archipelago, a process which gives rise to the volcanic island arc of Indonesia. This has led to the formation of volcanoes such as Mount Merapi, a stratovolcano very near the city of Yogyakarta. Similarly, Caribbean volcanoes formed in the Lesser Antilles are created through the subduction of the Atlantic Ocean floor beneath the Caribbean plate.

Volcanoes created through this process include islands such as Martinique, St. Lucia, and St. Vincent (Rothery 1997).

Volcanic arc settings can also be created through subduction at *oceanic-continental* plate boundaries. The most prominent example of this would be the subduction of the Nazca plate beneath South America. This gives rise to the Andean volcanoes in the North, Central, Southern, and Austral volcanic zones of South America. Likewise, similar processes are responsible for the Central American volcanoes and Cascade volcanoes of North America, albeit on a smaller scale.

The volcanoes of the Mediterranean are a result of the progressive movement of the African land mass northwards into Europe. The oceanic crust between these plates has been steadily subducting, creating the volcanoes of the Mediterranean. This has led to the formation of highly active volcanoes such as Mount Etna (Sicily) and Mount Vesuvius (Campania), the latter of which provides a study area for this thesis.

Volcanoes created at constructive boundaries or in the middle of continental plates, are primarily the result of mantle plumes and are termed 'hot spots'. Where the lithospheric crust becomes compromised, considered to be due to rising plumes of magma in the earth's asthenosphere, volcanoes can be formed where there is an appropriate level of pressure and heat for magma to upwell onto the surface and cool. A characteristic trait of these hot spot volcanoes is that the mantle plume may remain geostationary, but due to the movement of the tectonic plate above, a series of extinct volcanoes shows evidence of the historic progression of a plate's movement over long time periods (Rothery 2007). There are numerous mantle plumes across the world but the most significant volcanic examples of such origins are the Hawaiian Islands and Iceland.

Volcanoes can also be created in *rift valley* settings, such as East Africa, where the divergent continental plates have created generally more effusive volcanoes. At this location, the Africa plate is essentially being split into two new protoplates, the Nubian plate and the Somali plate.

Due to the colossal tectonic faulting, the process of rifting has created numerous active and dormant volcanoes along this boundary, including Mount Kilimanjaro, Mount Kenya, Mount Longonot, Menengai Crater, Mount Karisimbi, Mount Nyiragongo, Mount Meru, and Mount Elgon (Rothery 1997).

2.1.5   VOLCANIC HAZARDS

There are 1540 potentially active volcanoes across the world, with approximately 50 eruptions per year (Smith 2013). Though the public risk perception of volcanic hazards is greatly enhanced due to the dramatic size and explosive nature of these rare events, Figure 2-4 helps demonstrate that in comparison to other global natural disaster phenomenon, such as floods, severe storms, and droughts, it would be accurate to say that in the last 100 years, volcanoes have not been the most deadly natural disaster from a humanitarian perspective.
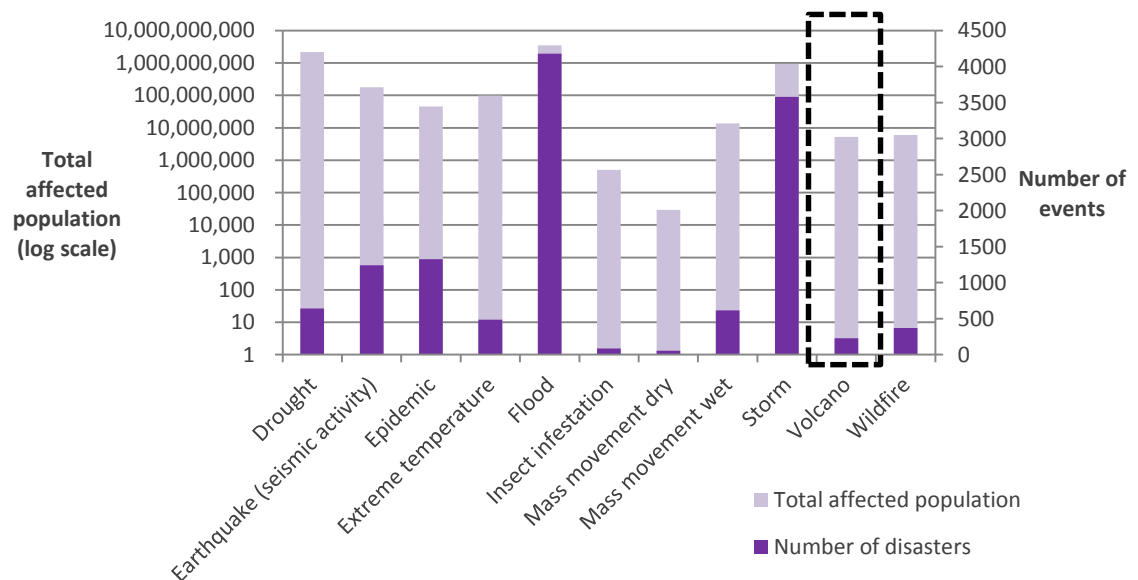
FIGURE 2-4 - GLOBAL NATURAL DISASTERS AND AFFECTED POPULATIONS (1900-2013) (EM-DAT 2013)

From 1900-2013, the average number of deaths from volcanic eruptions can be calculated to have occurred at an average rate of approximately 1300 people per year (EM-DAT 2013). However, this number is a little misleading as the range from year to year in this number is significantly higher, with a range average (X^{max}-X^{min}) of approximately 40,000. The irregular spatial and temporal pattern of the largest and most devastating volcanic eruptions is largely responsible for this deviation. The most significant loss of life from 1900-2013 was recorded in 1902, for the eruption of Mont Pelée on the Caribbean Island of Martinique. This event registered a Volcanic Explosivity Index (VEI) of 4, and killed 29,000 inhabitants in the port of Saint Pierre. This single event is responsible for 40% of volcano related deaths in the last 110 years. It is important to note that affected population figures and disaster data during this long observation period (110 years) has a great deal of uncertainty associated with it. Although the Centre for Research on the Epidemiology of Disasters (CRED) is arguably the most recognised source of such disaster information, the humanitarian impact of historic volcanoes is highly complex and may not have always been succinctly captured by such databases. Data records in developing countries, where many of the largest and most destructive volcanoes have occurred in the past, are often poor or incomplete. Therefore, calculating losses and impacted populations beyond the last 100 years should be treated with a high degree of uncertainty.

Due to the complex geological setting and geochemistry that give rise to volcanoes, there is substantial variation in the severity and nature of hazards that a volcanic eruption may produce. Whereas public risk perception is more familiar with the media portrayal and historical evidence of effusive lava being erupted from Hawaiian volcanoes or the explosive eruptions of Etna, Stromboli and Vesuvius (Italy), there appears to be less awareness of more indirect volcanic hazards. In 1986, Lake Nyos in Cameroon, a crater lake previously thought to be inactive, released a huge volume of CO_2, subsequently suffocating 1700 people living in the surrounding towns and villages (Kling et al 1987). Furthermore, the indirect effect of volcanic aerosols from an eruption on global weather pattern has profound and lasting effects on human population. The effect of volcanoes on climate change in tropical latitudes has shown to

increase the temperature gradient between the equator and pole as a result of shifting the jet stream and altering precipitation patterns (Robuck 2000). Such shifts in global rainfall may lead to drought, floods and famine in many parts of the world due to increased weather volatility.

Whilst it is important to note that there are multiple indirect volcanic risks, primary consideration during this thesis has been given to the more proximal hazards of a volcanic eruption. This is because these primary hazards are perceived as being more pertinent to the population settlements living in close vicinity of the volcano and their risk displacement is better understood. These hazards include pyroclastic density currents (PDCs), tephra (also termed ash fall), lahars, and volcanogenic earthquakes.

TEPHRA

Volcanic ash (which will sometimes be referred to as Tephra in this document) consists of the fragmented rock, minerals and glass material ejected from a volcano during an eruption. It is formed as the dissolved gases in the magma expand and are released violently into the atmosphere (Smith 2007). The reduction in pressure on the magma as it rises in the magma chamber and volcanic vent are the core driver of this explosive behaviour.

The chemical make-up of the ash is directly related to the magma's geochemistry, and therefore a consequence of the source material of the volcano. Describing tephra usually focuses on the volume of silica content (SiO_2) as this is a strong indication as to the composition of the volcano. If the silica content is <55%, the ash fall is typically considered to be basaltic (usually containing more Iron and Magnesium), whereas if the silica content is closer to 55-70%, the eruption is likely to be more explosive and related to dacite, andesite or rhyolite (Rose and Durant 2009).

Upon initial dispersal from the eruption column, tephra is driven upwards by the high velocity with which it was ejected. With further convection from air drawn into the eruption cloud, the ash cloud rises buoyantly into the atmosphere. At the point where bulk density is equal to the

surrounding atmosphere, the column cannot rise further and instead, begins to move laterally, giving the characteristic *mushroom* appearance to the eruption column. This phenomenon is illustrated in Figure 2-5 (diagram a). The top part of the cloud is often referred to as the umbrella region given its wide lateral spread.
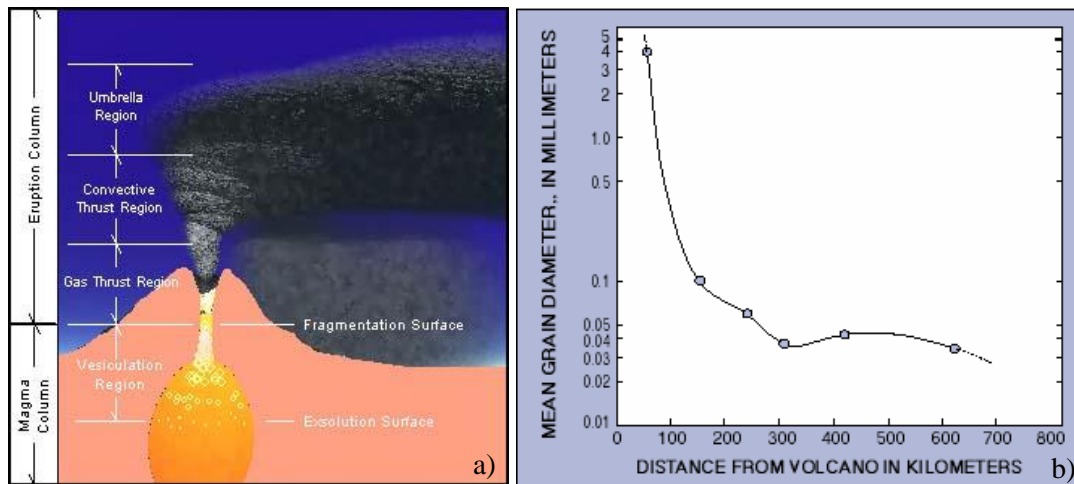
FIGURE 2-5 - A) ERUPTION COLUMN PROCESS (USGS 2013) B) GRAIN SIZE CHART (SARNA-WOJCICKI ET AL 1981)

As Figure 2-5 (diagram b) shows, based on the Mount St Helens Eruption in the US, there is a direct relationship between grain size and distance a given particle can travel away from the volcanic vent. The densest material ejected falls to the ground within a short distance of the volcano (typically fragments > 32mm in length) with the finer material being carried further away from the volcano (< 2mm). If the eruption is of significant intensity and given favourable wind conditions, the finest material can travel thousands of miles across the world, often carried by the jet streams of the upper atmosphere. If the ash cloud is impacted by tropical or extra-tropical cyclone activity, its projection and dispersal can be significantly widespread. In 1991, the strong winds of typhoon Yunya passed over the Philippines and acted to widely disperse the recently ejected tephra of Mount Pinatubo (1991).

Due to the large spatial coverage of tephra fall, as well as its potential to rapidly disperse, it can cause numerous physical risks to human populations. These risks are primarily focused around the physical damage caused to dwellings through the rapid accumulation of tephra on housing, and direct respiratory ailments caused by the inhalation of particularly fine grain sized volcanic ash.

As volcanic ash can accumulate to significant depths in the proximity of a volcano, its subsequent weight and vertical loading pressure can lead to damage or in extreme cases, complete roof collapse (Blong 2003). The 1994 eruption of the Tavurvur and Vulcan volcanoes over the town of Rabaul in Papua New Guinea covered thousands of houses in a thick blanket of ash. The thickness of loading varied from 100mm-950mm, creating loading pressures of between 2-16 KN m^{-3} (Blong 2003). Likewise, with high precipitation often occurring simultaneously to ash fall, resulted in wet ash which further increases the loading and subsequent risk to structural damage. After the Rabaul eruption, over 80% of the houses were noted to have suffered roof collapse as a consequence of the ash loading. Table 2-1 highlights the relationship of tephra loading to collapse as observed during the Rabaul eruption (Blong and McKee 1995).

TABLE 2-1 - USGS, ASH FALL THICKNESS AND OBSERVED DAMAGE (ADAPTED FROM BLONG AND MCKEE 1995)

Ash thickness[1] in cm	Estimated load[2] in kPa	Observed damage to roofs (Rabaul eruption 1994)
<10	1.5 - 2.0	Roofs and guttering mostly intact.
<20	3.0 - 4.0	Sagging or partial collapse occurred in some buildings.
<30	4.5 - 6.0	> 50% of roofs did not collapse.
50-60	7.5 - 12.0	>50% of roofs collapsed.
>60	9.0 - 12.0	It is doubtful that buildings survived without significant damage even when the roof remained relatively intact.

As well as the physical risk to human settlements through the collapse of housing, it is important to note the large scale damage to agricultural land, machinery, and business

interruption that volcanic ash fall can cause. In the short-term, and in accumulations exceeding 150mm, tephra can act to engulf and destroy arable crops. Occasionally, if hot enough, it can set the ground ablaze (Smith 2013). As well as depriving the soil of oxygen, toxicity of the ash creates a significant hazard to livestock. Volcanic ash is typically rich in fluorine, which if consumed in large enough doses by grazing animals, can lead to fatal conditions such as fluorosis. Similarly, the acidic coating of the ash can get washed into waterways or aquifers (IVHHN 2013) causing a major risk to human health if part of an active source of water for large populations.

There is also a strong link between direct ash fall and adverse effects on human health, such as respiratory ailments, eye symptoms, and skin irritation. Following the eruption of Quito's Guagua Pichincha volcano in April 2000, Estrella et al (2005) noted the substantial increase in upper and lower respiratory infections and asthma-related health problems for children living near the volcano. Similarly, there is a large amount of consensus from the literature showing evidence that children and those with existing breathing difficulties (such as asthma sufferers) are disproportionally more at risk to moderate or sustained levels of volcanic ash fall (Forbes et al 2003; Bernstein et al 1986; Vallyathan et al 1984). Typical symptoms of exposure to ash fall include nasal irritation, sore throat, heavy coughing or if there is sustained exposure over a long-term period, lung diseases such as silicosis may occur (IVHHN 2013).

Tephra can also cause major disruption to commerce, travel and industrial enterprise. As ash falls on airport runways, roads, or railways, the business interruption and financial implications are significant. The April 2010 eruption of the Eyjafjallajökull volcano in Iceland shut much of continental European airspace intermittently for several weeks. The International Air Transport Association estimated that the closures cost $200 million (USD) per day in lost revenue for the airlines, with a total of approximately 1.7 billion USD (IATA 2013) thereafter. The danger of ash fall to commercial flights concerns the airborne particles of volcanic glass that can be melted in the jet engines of airplanes and may ultimately compromise the safety and performance of the aircraft.

Unlike other volcanic hazards, the hazards of tephra fall are particularly difficult to quantify and assess because the several parameters that determine the risks are enormously changeable. For example, the prevailing wind direction and strength on the day of a volcanic eruption ultimately determine the spatial extent and volume of accumulated tephra. Similarly, the magnitude and intensity of the eruption will determine the volume of material ejected as well as the tephra grain size. To model the secondary effects of an eruption (e.g. economic disruption) and arguably, the tertiary consequences that may be ultimately caused by a large scale tephra fall (e.g. altered global precipitation patterns /global warming/droughts) remains extremely uncertain and is beyond the scope of this thesis.

PYROCLASTIC FLOWS

These super-heated, fast moving torrents of hot volcanic gas and rock are usually formed when there is a partial or total collapse of the volcanic edifice. Also referred to as *nuees ardentes* or *gravity flows*, pyroclastic density currents (PDC) are denser than the surrounding air and rush down the volcanoes flanks at unprecedented speeds, often in excess of 100km/hr. (Rothery 2007). The gas can reach temperatures in excess of 1000°C and they are a formidable hazard if human populations are impacted.

The formation of PDCs differ with each volcano but is usually the result of partial or total collapse of the eruption column. When collapse of a lava dome occurs, the resulting ash and dense rock forming the PDC is referred to as a *block and ash flow.* The term *nuees ardentes* is used if the flow is particularly hot or incandescent. With larger eruptions, such as Plinian events, PDCs may comprise of less dense material, such as ash and pumice. If this is the case, the coverage may be extensive (10-100 Km^3). The term *ignimbrites* refers to the deposits left behind from a PDC. The third type of PDC formation became widely recognised following the Mount St Helen's volcanic eruption in May 1980, USA. During this eruption, a major collapse of the side of the column occurred. The sudden exposure of crystalline magma that was previously held under extreme pressure in the magma chamber, allowed for gases to

explosively expand from the volcano. Commonly referred to as a *lateral blast*, it was triggered by the growth of a cryptodome resulting in enormous PDCs that engulfed much of the surrounding landscape in Oregon (Donnadieu and O'Merle 1998).

Unlike lava flows and tephra fall, which tend to be slowly evolving risks during an eruption, and typically allow a reasonable time for civil evacuation measures to be implemented, PDCs inherent danger is largely caused by their rapid escalation and ability to destroy almost everything in their path. The significant kinetic energy and heat of PDCs allows them to reach populated areas near volcanoes in seconds, whilst the heat and impact of the surge incinerates most vegetation, livestock and even the built environment they encounter.

Historical deaths from volcanic eruptions have often been the consequence of PDCs. If we study the period 1900-2010, and as noted previously in this chapter, the Caribbean town of St. Pierre in Martinique was the single largest loss from a volcanic eruption. On the 6th May 1902, the on-going eruption of Mount Pelee intensified, causing a partial collapse of the lava dome and creating a devastating PDC that engulfed 10km of shoreline, and more significantly, engulfed the town of St. Pierre. Figure 2-6 provides photographic evidence of the destruction witnessed in St.Pierre following the eruption. The resultant PDC killed almost everyone in the town (around 23,000 people). The temperature of the PDC surge at St. Pierre has been estimated at around 700°C (Smith 2004).

FIGURE 2-6 - THE DESTRUCTION OF ST. PIERRE, MARTINIQUE, FOLLOWING THE MAY 1902 ERUPTION OF MOUNT PELEE (HEILPRIN 1908)

Due to the heat of the ash and preceding hot air mass that may accompany a PDC, fatalities are thought to be a combination of factors such as internal and external burns, as well as asphyxiation caused by the air blast. It is the rapid onset and intensity of this hazard that has largely accounted for historic loss of life.

LAHARS

Lahars are volcanic mudflows comprising of small grain sediments that may occur in the aftermath of an eruption. They are often formed when recently deposited tephra fall interacts with water to cause fast moving deluges down the flanks and river systems of the volcano. The word is derived from Javanese (Indonesia) and it is in this region of the world where they have been a particularly prevalent hazard (Smith 2004). For example, the 1919 eruption of Kelut killed 5000 people in a lahar that was triggered after the main eruption. They are sometimes generated either by a crater lake collapsing, and hence sending a deluge of water and ash down the mountain side, or else indirectly through intense periods of rainfall interacting with recently deposited ash mobilising a mudflow.

2.2 RISK PERCEPTION

Aside from the quantitative measures of risk assessment that have been previously employed to measure social vulnerability, it is worth taking note of the literature on risk perception. As this research is focused around neighbourhood profiling, it is important to consider the attitudinal aspects of those communities living around natural hazards.

The history of volcanic risk perception can essentially be viewed as emanating from two related yet still separate strands of social science research. The first can be referred to as general risk perception and pertains to associated risks in our everyday lives. As will be demonstrated shortly, many of these risks are entered into voluntarily. The second aspect of risk perception research is associated with natural hazards, and although it is a related branch of the first area of literature discussion, it assumes our attitude to risk is a product of our environment.

The derivation of research in risk perception can be dated back to Starr (1969) and his work on how modern society accepts the nature and consequences of voluntarily entered risks (such as someone's decision to smoke or drive a car). We often have a muted perception of such risks whereas conversely, involuntary entered risks are often deemed far more serious to us (Starr 1969). This idea was subsequently developed and further enriched by Fischoff (1978) and Slovic (1980) in work focused on determining how we come to define the risks around us. They essentially addressed three main variables that we subconsciously assess in any hazardous situation: a) the dread, b) familiarity c) number of people exposed.

The dread can be summarised as the fear of an event; the familiarity defines our previous experience of the risk, and therefore what we might expect from any subsequent hazard. Lastly, the number of people exposed allows us to quantify the hazard before making our final assessment of the risk.

One of the most significant areas of research in this field was the development of the 'Cultural theoretical' approach (Douglas 1966, Thompson 1980). This approach stipulates that our risk

perception arises largely from society. The influence of our community, religion and political beliefs are tantamount to how we come to perceive a risk. The 'Cultural theoretical' approach advocates that it is not just our subjective knowledge of a hazard that defines our risk perception but our cultural heritage and social organisation. In summary, our risk perception is the very product of our social environment.

Dake and Wildavsky (1993) highlight that an individual's fears are in direct accordance with their lifestyle. What we fear (or do not fear) may actually be a subconscious decision in order to support an existing way of life. For example, though an individual may live and work on the flanks of an extremely dangerous volcano, if the land is their ancestral homeland or they have valuable animal stock to protect, their risk perception of the volcano erupting may well be mitigated by their personal circumstances. This example would suggest that the farmer has consciously or subconsciously deemed it as a risk worth taking.

Likewise, perceptions of risk can be heavily biased by our social beliefs and cultural history (Greene et al 1980). The effects of community and societal mechanisms are tantamount to our inherent perceptions of the hazards that surround us. The various stakeholders within such social networks play a vital role in how a perceived risk can be minimised or amplified. Kasperson's 'Risk Amplification theory' (1992) identifies points of contact that can act as a leverage for raising/lowering awareness of risk. These 'stations' can include the manner in which local media report a risk, pressure groups, or the effects of government intervention (Haynes et al 2008).

The focus of previous risk perception has essentially been based on the premise that risk taking behaviour is linked to risk perception, and that if we perceive something as a risk, we are much less likely to pursue that option.

The 'Perception adjustment paradigm' was created (White 1945, Kate 1976, Burton et al 1993) to describe the process by which an individual appraises, evaluates and concludes how to deal with a risk. There is some debate in the literature as to whether individuals really do decide

their risk perception in this subjective manner or whether our decisions are much more the result of organisational influence e.g. society, government, friends and our communities (Torry 1979, Susman 1983).

In recent years, risk perception has become much more focused on the minutiae as bespoke areas of research have taken precedent. One such example is volcanic risk perception.

The history of volcanic risk perception has very much been centred on hazard research. Early work has sought to define local knowledge of the possible hazards to a volcano, thus comparing these to the actual risk. Pioneers of this new field included Lachman and Bonk (1960), Murton and Shimabukuro (1974) studying the Hawaiian volcano and Kapoho and Green's (1981) work on the Mount St Helen eruption. Unsurprisingly, much of this early research was triggered by large eruptions of these latter volcanoes, both of which occurred in the US.

Until the 1990s, risk perception work on volcanically active regions relevant to the study areas of this research (Italy and Ecuador) was minimal. Essentially, the US volcanoes were the main focus. It was not until contributions by D'Ercole (1991, 1994, 1996), Peltre (1992) and Tobin and Whiteford (2002) that the countless volcanoes of Ecuador gained greater attention. Peltre's (1992) work highlighted how the length of residence and distance of an individual from the Cotopaxi volcano influenced people's perception of the volcanic hazard. Lane, Tobin and Whiteford (2003) have undertaken significant survey work in Banos, a town that has seen substantial and numerous evacuations due to its emplacement at the foot of the Tungurahua volcano. Their work has identified the strong links that exist between health problems and evacuation. The ramifications of which are often perceived as worse than the likely volcanic hazards that could affect the town. Likewise, this work was further corroborated during the eruptions of Tungurahua in May 2010, when local Banos evacuees were subjected to theft and violence in the temporary camps being used for Evacuation (PAHO 2010).

An increasingly important aspect of volcanic risk perception regards the role of the media and local stakeholders. The various scientists, politicians, and journalists involved in an ongoing

disaster are integral to how we receive, rationalise and react to a given risk or disaster. Ongoing eruptions on the Caribbean island of Montserrat since 1994 have fuelled a wealth of research regarding social policy and disaster communication outreach. Through qualitative research on Montserrat, it has been identified that existing communication channels to the public through all stages of the ongoing eruption were often found lacking. Such miscommunications often lead to feelings of frustration, mistrust, and resentment among the islands population (Haynes et al 2008; McGuire et al 2009; Donovan and Oppenheimer 2014).

In producing a handbook entitled *Communication During Volcanic Emergencies*, McGuire et al (2009) discussed the poor and untimely communication of information between monitoring scientists, emergency managers and the role of the media. Likewise, in a series of semi-structured interviews, Haynes et al (2008) found that an individual's trust and subsequent response to disaster outreach regarding the ongoing eruption was directly related to the source of the message. It was found that information relayed by family members or close friends was considered more reliable than the global media relaying the information.

Donovan and Oppenheimer (2014) reflect that such issues may be a consequence of the existing linear relationship of such policy making in volcanic disasters, whereby the public are only informed at the end of the chain and scientists are consulted only at the beginning. It is proposed that a more worthwhile model would have a matrix structure, allowing the public access to every stage of the policy process. It is argued that such transparency would greatly reduce the noted problems in volcanic disaster communication. A more unorthodox approach to reflecting on the experiences of Montserrat was taken by Donovan et al (2011) in demonstrating how local poetry and prose could be used to objectively assess the ongoing disaster and provide insight into community perception and outreach initiatives. Using an approach based on 'practical criticism' of literature, poetry and narrative were shown to be a cultural product of a disaster and that stakeholders could make use of such feedback to help reflect on the changing community perceptions of such events to direct policy making (Donovan et al 2011).

On the practical aspect of communicating volcanic risk, there are several channels available for stakeholders to disseminate such vital information. For the local authorities, the media, and NGOs to believe and take the necessary steps towards preparedness during a disaster, it is important that scientific information is both timely and conveyed through an appropriate channel (Mileti and Sorensen 1990). Potter et al (2014) provide a good overview of how volcano information is communicated to the public before, during and after a crisis. Through a combination of presentations, workshops, public lectures, website content, email, pager alerts, and SMS messages, registered users of GNS Science gain vital information throughout the course of a crisis. More recently, social media has been put to effective use with the availability of an option for online users of the GNS Science website in New Zealand to 'ask an expert' and receive answers in real-time from volcanologists (Potter et al 2014). A fundamental challenge of disaster communication however, remains in the cultural and technological availabilities that may exist between nations and even within a country. For example, smart phone penetration in the US among the entire population is 56%, compared with only 14% in Indonesia (Google 2015). Clearly, such differences have major impacts in the way civil authorities choose to alert the at risk population.

2.3   VULNERABILITY

In this section, the concept of vulnerability is explored in detail and with its relevance to the research presented here. Current understanding and definition of the term is explored under the wider concepts of DRR, coping capacity and global frameworks (e.g. UNISDR, HFA, SFDRR). This is followed by a review of the literature on *social vulnerability* whereby several conceptual models are presented that have sought to clarify the fragility of an individual during the onset of a natural disaster (Davidson 1997; Bohle 2001; Chambers and Conway 1992; Wisner et al

1994). The commonality of social vulnerability indicators are thus identified from previous research which can ultimately be used to help identify, characterise and target those 'at risk' communities in vulnerable settlements. Lastly, consideration is given to the history of research in 'risk perception' as this can be regarded as a pivotal component of vulnerability and any attempt to profile a community's risk.

2.3.1   DEFINITION OF VULNERABILITY

Vulnerability can be a very complex notion to explain. It has multiple dimensions and site-specific characteristics in different places. For example, vulnerability may involve the physical world (e.g. risk of the built environment), the social (e.g. inequality), environmental factors (e.g. natural hazards), institutional (e.g. political regimes), as well as factors of human frailty. The added difficulty is that all of these dimensions can be hard to quantify.

The Oxford English Dictionary (OED) literary definition of the term *vulnerability* is shown below. As seen, it is somewhat restricted to the notion of *risk* or physical *harm*. This is most likely because it is often interchangeably used with other terms to describe the same phenomenon. For example, terms such *risk, harm, hurt, exposed, danger* may also used.

"*exposed to the possibility of being attacked or harmed, either physically or emotionally*" (OED 2013, p.1622)

Progressing this notion, the current definition for the term as quoted in the United Nations Office for Disaster Risk Reduction (UNISDR) is provided in the quotation below and provides much more focus on the concepts of resilience and social condition. Likewise, it is interesting to note the introduction of controlling factors in their terminology. This includes mention of the role of the *community, state* and *economy* in the concept and outcome of vulnerability.

"*The characteristics and circumstances of a community, system or asset that make it susceptible to the damaging effects of a hazard - there are many aspects of vulnerability, arising from various physical, social, economic, and environmental factors. Examples may*

include poor design and construction of buildings, inadequate protection of assets, lack of public information and awareness, limited official recognition of risks and preparedness measures, and disregard for wise environmental management. Vulnerability varies significantly within a community and over time. This definition identifies vulnerability as a characteristic of the element of interest (community, system or asset) which is independent of its exposure. However, in common use the word is often used more broadly to include the element's exposure." Terminology for Disaster Risk Reduction (UNISDR 2009, p.34)

Also noted in the passage above, vulnerability is almost always discussed or referenced in the same context as resilience. This notion is now something of a standard definition and captures the role of the individual or community to cope with a disaster. Capacity and resilience are seemingly considered to be an integral ingredient of vulnerability. Further validation of this contemporary definition is found in the terminology of other key non-governmental organisations. As seen in the Red Cross quote below there appears to be a growing consensus in this definition among DRR practitioners.

"*The diminished capacity of an individual or group to anticipate, cope, resist and recover from the impact of a natural or man-made hazard. The concept is relative and dynamic. Vulnerability is most often associated with poverty, but it can also arise when people are isolated, insecure and defenceless in the face of risk, shock or stress.*" (IFRC 2013)

During the course of this research, the terms *vulnerability* and *risk* are used extensively and therefore it is important to state their definition and relationship to each other and within the context of this study. The UNISDR (2002) currently define risk with equation 1) shown below.

Risk = Hazard (H) x Vulnerability (V) / Capacity (C)   1)

Although previously used definitions have not always included the notion of capacity, such as the United Nation's Department of Humanitarian Affairs (UNDHA 1992), when it defined the equation **Risk = Hazard x Vulnerability**, this study is being undertaken within the concept that increased capacity (or resilience) can act to decrease a perceived risk. This has important consequences in the quantification of risk indices in later chapters as certain social indicators can be recognized as having an ameliorating effect on risk.

Aside from the definitions of vulnerability provided earlier in this section, it is important to recognise there are multiple, and often conflicting interpretations of this concept. Katharina Marre provides a glossary of 36 unique definitions of vulnerability in chapter 23 of *Measuring vulnerability to natural hazards* (Birkmann 2006). A summary of several of these definitions is provided in Table 2-2. It is interesting to note how the concept of vulnerability is almost always interpreted into a defined area of research, depending on the risk, whether referencing global warming (Adger et al 2001), socio-economic impacts (Susman et al 1983) or financial loss (Buckle et al 2000).

Table 2-2 – Glossary of vulnerability definitions (adapted from Birkmann 2006; 595-602)

Definition	Author/Publication
"The propensity or predisposition to be adversely affected."	IPCC 2012
"Vulnerability should be recognized as a key indicator of the seriousness of environmental problems such as global warming."	Adger et al 2001
"The degree of loss to a given element of risk or set of such elements resulting from the occurrence of a natural phenomenon of a given magnitude and expressed on a scale from 0 (no damage) to 1 (total loss) or in percent of the new replacement value in the case of damage to property."	Buckle et al 2000
"Vulnerability expresses the severity of failure in terms of its consequences. The concern is not how long the failure lasts but how costly it is."	Correira et al 1987
"Is a broad measure of the susceptibility to suffer loss or damage, The higher the vulnerability, the more exposure there is to loss and damage."	Department of Human Service 2000
"Vulnerability defines the inherent weakness in certain aspects of the urban environment which are susceptible to harm due to social, biophysical, or design characteristics."	Rashed and Weeks 2002
"The degree to which different classes in society are differentially at risk, both in terms of the probability of occurrence of an extreme event and the degree to which the community absorbs the effects of extreme physical events and helps different classes to recover."	Susman et al 1983

The notion of *volcanic vulnerability* is typically referenced in the literature with respect to the physical damage afforded by an eruption. For example, Juan Carlos Villagran de Leon defines the vulnerability to volcanoes in terms of the structural frailty that could impact a house during an eruption, depending on the build (Birkmann 2006; 431). Moreover, research into volcanic vulnerability has tended to focus on single volcanic risks rather than encapsulating the full spectrum of volcanic risks that may occur from an eruption. Recent studies are addressing this shortfall, with greater focus on assessing the social vulnerability of volcanic risks over multi-year time periods. Anna Hicks and Roger Few provide a robust account of how social vulnerability to the Soufriere Hills volcano in Montserrat is not a stationary concept, but has a transitory nature, evolving over time as hazards and population changes (Hicks and Few 2014).

2.4   SOCIAL VULNERABILITY

The concept of 'social vulnerability' appears to have existed in academic literature for over 65 years. The work of White (1945), Hewitt and Burton (1971) and Kates (1978) have helped to fundamentally change our understanding of natural disasters. Much of this early research appears to have been qualitative and it was not until the 1990s that there began to be a substantial growth in the number of researchers and disaster management professionals seeking to quantify and derive definitive methodologies for these concepts.

Drawing from the literature on natural disaster risk (Quarantelli 1978, Hewitt 1983, Wisner 2004), certain demographic and socio-economic characteristics can be recognised as increasing an individual or household's vulnerability before, during and post-disaster. The term itself alludes to partly 'social inequalities' and partly 'place inequalities' from the existing built environment (Cutter et al 2003). It includes (but is not exclusively) the socio economic status of groups of people, demographic traits, perception and attitudinal differences towards people and places, social networks, access to capital and resources, physically weak individuals, cultural beliefs, access to basic infrastructure and access to political power (Cutter et al 2003; Blaikie et al 2004). Following the work of early pioneers in disaster management, such as Westgate and

O´Keefe (1976), greater focus began to be placed on understanding how hazards became 'disasters'.

2.4.1 CONCEPTUAL FRAMEWORKS AND MODELS

To better understand and measure how disasters came about, researchers began to turn their attention to conceptual frameworks and models. Within these models they hoped to better capture some of the complex drivers of disasters. Whether at a community, local, state or global level, these conceptual models have helped provide practitioners with greater ability to measure vulnerability.

The models are markedly different in how they represent vulnerability as an input, output or a root cause of a disaster but they share many of the same fundamental mechanisms, such as *coping capacity, environment* or the role of *state intervention*. There are four principle schools of thoughts with regard to conceptual models of vulnerability; (a) political economy (b) social-ecology (c) vulnerability and disaster risk (d) climate change systems science (Birkmann 2006).While the proceeding section is not an exhaustive review of all these conceptual frameworks, it seeks to summarise the core principles of some of the most diverging concepts.

Bohle's conceptual framework (2001), also referred to as the *double structure of vulnerability*, focuses on two key themes; *exposure* and *coping* mechanisms. The interplay and relative weighting of these forces provide the key driver of vulnerability. *Exposure* is perceived as the external side, to which the individual/group may be subject to risk or shock, and *coping* is the internal side focusing on their capacity to manage the event. Bohle based much this concept on the research undertaken on the impact of intensive famine on households and communities. It is interesting to note hazard is defined as an input to vulnerability in Bohle's framework (2001).

Alternatively, Chambers and Conway's *Sustainable Livelihood framework* (1992) has a more abstract approach, putting focus on how to drive positive outcomes through sustainability. In defining key assets as human, natural, financial, social and physical capital, the element of

vulnerability is seen in this context as an input (rather than an output) to the model. Transforming processes and structures (such as laws, culture and government) are defined as the controlling mechanism to vulnerability, with the individual or group able to influence these processes. Criticism has been raised regarding the lack of awareness in this framework for positive livelihood feedback mechanism (e.g. how sustainable land use can reduce environmental hazards such as flooding or landslides).

An opposing paradigm to these latter approaches is provided by Davidson's (1997) *Disaster Risk framework*. This model is focused on input components to create a disaster, defining four principle categories; hazard, vulnerability, exposure, and capacity measures. The key difference in this theory regards its treatment of vulnerability as a discrete input to disaster (Birkmann 2006).

Sustainable development context

Socio-cultural

The focus of disaster risk reduction

AWARENESS RAISING
for change in behavior

KNOWLEDGE DEVELOPMENT
· Information
· Education & training
· Research

RISK FACTORS
Vulnerability
· Social
· Economic
· Physical
· Environmental

Hazards
· Geological
· Hydrometeorological
· Biological
· Technological
· Environmental

Vulnerability / capability analysis

Hazard analysis & monitoring

RISK IDENTIFICATION & IMPACT ASSESSMENT

POLITICAL COMMITMENT
· International, regional, national, local levels
· Institutional framework (governance)
  - policy development
  - legislation and codes
  - organizational development
· Community actions

DISASTER IMPACT

EARLY WARNING

Political

Ecosystems / Environmental

PREPAREDNESS

APPLICATION OF RISK REDUCTION MEASURES
· Environmental management
· Social and economic development practices (including poverty alleviation, livelihoods, financial mechanisms, health, agriculture, etc.)
· Physical and technical measures
  - land-use/urban planning
  - protection of critical facilities
· Networking and partnerships

EMERGENCY MANAGEMENT

RECOVERY

Economic

FIGURE 2-7 - UN-ISDR FRAMEWORK FOR DISASTER RISK REDUCTION (UNISDR 2004)

A similar structure is provided by the UN-ISDR's *framework for disaster risk reduction* (UNISDR 2002). Figure 2-7 shows how this framework also splits vulnerability into separate contributory components in a similar manner to Davidson's model (1997). It defines these vulnerability sub-categories as being social, economic, physical or environmental in origin. Although there are similarities in the model structures, the UN-ISDR framework begins to distinguish how disaster risk mitigation strategy should be focused around themes such as early warning, sustainable development, political commitment and knowledge development. Interestingly, as seen in Figure 2-7, it perceives that factors such as preparedness and emergency management are not able to be entirely controlled or mitigated.

The *Pressure and Release Model* (PAR) developed by Blaikie et al (2004) is based on the premise that disasters are a result of root causes (e.g. macro-economic), dynamic pressures (e.g. rapid population growth) and unsafe antecedent conditions (e.g. living in a dangerous location) with an opposing pressure coming from the risk of natural hazards (Blaikie et al 2004). This is illustrated in Figure 2-8, which acknowledges that disasters are a consequence of the complex interplay of these factors, often in multi-dimensional and multi-scalar environments. A key point about the PAR model is that as well as dynamic pressures and root causes creating a vulnerable environment, there are equal opportunities to reduce such stresses by many of the same instruments (e.g. state intervention, planning, and risk mitigation).
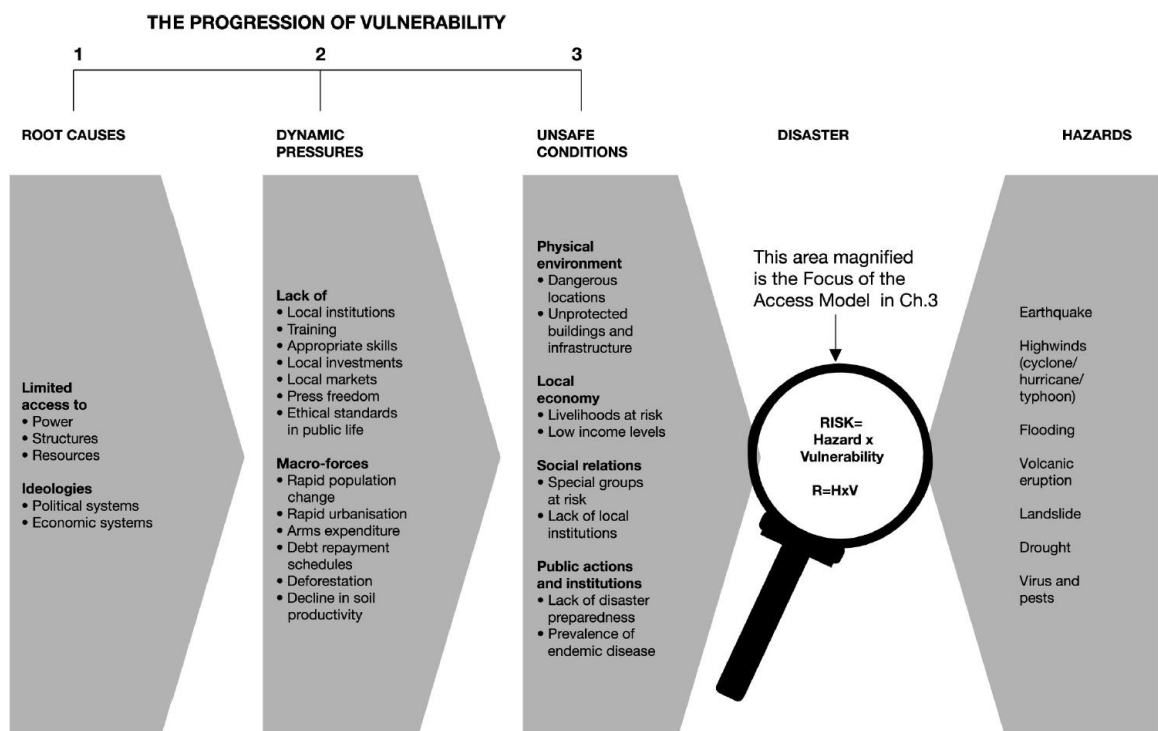
FIGURE 2-8 - THE PRESSURE AND RELEASE MODEL (PAR) (BLAIKIE ET AL 1994)

Likewise, it is important to note in the PAR model that it is essentially scaling causative factors in order of their administrative significance. For example, root causes are seen as the economic

and political systems (e.g. Bretton-Woods ideology) that subsequently drive dynamic pressures such as population growth, migration, and urbanisation. This, in turn, formulates into unsafe antecedent conditions and further exacerbates vulnerability as populations are forced to live and work in hazardous places. The parameters of the physical hazard are seen as the opposing pressure within this model. Further mention is given to the PAR model in the discussion section – where it's been used to demonstrate the vulnerable population settlements of Quito (Ecuador) and Naples (Italy).

2.4.2   QUANTIFYING VULNERABILITY

Research applications in the 21st century have begun to branch in their use and quantification of vulnerability. It is now becoming as commonplace for vulnerability indices to be used to assess natural disaster risk (such as earthquake or hurricane risk) as it is for them to be used to assess the impact of a heat wave or the long-term impacts of climate change (Cutter et al 2003; Gaither et al 2011; Shah et al 2013; Wolf and McGregor 2013). The geographic territory may vary as well as the physical risk variants, but the treatment of human exposure and capacity in such indices is often well-aligned. In using aggregated census or survey data to identify social vulnerability indicators, the context is very often the defining element. For example, Gaither et al (2011) created a vulnerability index to focus on wildfire risk, whereas De Andrade (2010) used similar indicators to formulate an index for oil spills in the Amazon. Similarly, Shah et al (2013) and Hahn et al (2009) created indices to study livelihood sustainability but with a sharper focus on climate change adaptation.

Such indices have followed in the wake of particularly seminal work on developing transferrable methodologies. Dwyer et al (2004) provided a detailed study of the quantitative identification of social vulnerability in Australia and though the concept of 'social vulnerability' was not particularly new, it was really in the late 1990s, early 2000s with the principles espoused by Susan Cutter (1996; 2000; 2003) that there began to emerge a growth in the application of quantitative methods to social vulnerability analysis.

The notion of a Social Vulnerability Index (SoVi) recipe was pioneered by Susan Cutter (2003) based on the notion that antecedent social conditions could be mathematically combined with natural hazard risk/frequency to formulate a single over-arching index. The value of the index can be attributed to a set administrative boundary (e.g. county, district, census zone) and thus thematic risk maps can be created. Much of Cutter's work has been focused on the characterisation of the US, and with particular focus on the state of South Carolina to numerous natural hazards, such as hurricanes or earthquakes (Cutter et al 2000; 2003; Schmidtlein 2011). Cutter devised a statistical and somewhat integrative approach to classifying the regions based almost exclusively on indicators provided by the census data. Similarly, in subsequent research at a national scale, this involved ranking US counties according to their social natural hazard risk. The indices have been formulated using both factor analysis and data reduction statistical techniques such as Principal Component Analysis (PCA).

Cutter (2003) identifies three distinct themes in the study of vulnerability:

1. Vulnerability as a risk or pre-existing condition.
2. Vulnerability as a social response.
3. Vulnerability as a hazard of place.

The concept of vulnerability as a pre-existing condition is concerned with the distribution of the noted hazard, whereas the second point focuses on vulnerability as a lack of response or resistance to a given event, thus impacting the severity of the risk. This second theme corresponds with what is commonly referred to as resilience or societal resistance. The third point regarding hazard of place is essentially the combination of the first two factors and is not in itself distinct. It is important to note that there are certain hazards that are unavoidable due to our spatial location. This is an important point. Despite our best efforts to engineer, prepare, understand and hopefully negate a hazard, some natural perils are unavoidable.

A limitation of Cutter's work could be that it has been heavily focused on the US, and not in the world's developing nations, where natural disaster resilience and DRR initiatives are arguably the most needed. Disasters such as hurricane Katrina (2005) and Andrew (1995) have clearly demonstrated the polarized experience of a US disaster, depending on your ethnographic or socio-economic status or gender (Elliot and Pais 2006; Morrow 1999), but the most severe loss of life from disasters has typically occurred in the most vulnerable countries of the world. The devastating earthquake in Port au Prince, Haiti (2011) and the Indian Ocean Tsunami (2004) highlighted the exposure of developing nations, with a combined death toll of nearly 500,000 people (DEC 2014). Therefore, it is also worth discussing recent initiatives aimed at measuring vulnerability on a global scale to natural hazards, with particular focus on those developed by the UN and World Bank in recent years.

In 2004 and 2005, several initiatives were launched by the UNISDR. One of the most comprehensive has been the *Disaster Risk Index* (DRI); a deductive measure of vulnerability based on hazard exposure across the world to earthquakes, hurricanes and flooding. Although consideration was given to volcanic risk, due to the very long periods of repose between large eruptions and the subsequent loss of life, this hazard was discounted from the index. The model considered the loss of life to disasters at a national level for a time period from 1980-2000, and then used a metric of mortality divided by exposed population to study vulnerbability (Peduzzi et al 2009). The model proved successful at raising awareness to the vulnerability of many developing nations across the globe but was certainly limited by its focus on a national scale, and did not provide insight at a regional or local level (Birkmann 2006).

Another initiative was created by Colombia University, in collaboration with the World Bank, in developing the *Hotspots* project in 2005. Using gridded population data for 4.1 million cells across the globe, vulnerability was assessed on the basis of the census based population in a location measured against the loss of life due to various natural hazards, including earthquakes, volcanoes, landslides, floods, drought, and cyclones (Arnold 2006). The resolution of the *Hotspots* project provided a sub-national assessment of risk across the world, though it is worth

mentioning that regions where population density was less than 5 people/Km were excluded from the model.

Some projects have been focused at a continental level, such as the *Americas Indexing Programme* (Cardona 2005). Created by the Instituto de Estudios Ambientales Nacional de Colombia and the Inter-American Development Bank (IDEA), Central and South American countries were indexed to multiple disaster indices at a national level to assess relative risk and coping capacity in the region. The metrics of assessment included a disaster deficit index (DDI), a local disaster index (LDI), a prevalent vulnerability index (PVI) and a risk management index (RMI). The DDI measured a country's financial capacity of recover from a disaster (based on various return period scenario events) whereas the LDI measured impact of disaster on a regional basis within a country. The PVI was concerned with a socio-economic based assessment of human vulnerability within a country, if a disaster were to occur. The RMI index score was based on the country's likely performance in risk management and effectiveness in coping with a natural disaster. Although the Americas Indexing Programme provided a robust picture of national security and financial coping capacity within the region it did not provide the granular level of risk assessment need to facilitate DRR practitioners at the local level (Birkmann 2006).

2.4.3   SOCIAL VULNERABILITY INDICATORS

The precise extent and causes of social vulnerability can vary significantly from place to place dependent on the specific cultural and demographic structures of a population (Blaikie et al 1994). However, historical disaster experience has demonstrated that commonalities can certainly be found across diverging spatial and temporal landscapes. Factors that can be used to define, understand and ultimately measure social vulnerability are often referred as *indicators* (Blaikie et al 2004). These indicators can be measured with varying success by the use of census or survey data. Blaike et al (2004) defined the following core indicators in social vulnerability to natural disasters:

- Income (poorest are worse affected)

- Gender (women are worst affected by disasters)

- Children (particularly infants)

- The elderly

- Minority groups

For the purposes of methodology design and the practical applications described in later chapters of this thesis, the following census variables and statistics were considered as principle factors in defining social vulnerability around volcanic regions.

AGE

Age can be a key determinant in the human experience of a disaster. For example, the elderly and the young are considered to be more difficult to move during disaster evacuation and have a higher propensity to adverse health conditions (McMaster and Johnson 1987). It has been found that the age spectrum affects people's movement out of harm's way, with the young and the elderly being far less mobile in such situations. This point was reiterated recently in the aftermath of the Japanese earthquake and subsequent tsunami (March 2011). It was found that 65% of the known fatalities were people over 60 years old (House of Japan 2011). The same pattern was noted when hurricane Andrew hit the Florida coastline (1994). The elderly were noted to be more at risk from mechanical asphyxia or cardiovascular effects as house collapse occurred (Lew & Wetli 1996).

ETHNICITY

Similarly, ethnicity plays a vital role in both differential disaster exposure and response. Pulido's work (2000) has highlighted the concept of white privilege and environmental racism in Los Angeles, where it is shown that predominantly white suburbs are in the cleaner, less vulnerable neighbourhoods through geographical processes of decentralization and

suburbanization. Conversely, many black and ethnic minority groups live in the older less desirable neighbourhoods. Such neighbourhood segregation has major implications for social vulnerability. Likewise, aside from geographical differentiation, there is strong evidence to suggest that ethnicity can be strongly aligned with the capacity to cope during and after a natural disaster. Ethnicity can have an impact on the ability to access vital resources (Elliot and Pais 2006). This may include a lack of political power or the social networks required to cope and recover from a disaster. A good example of this is provided by Elliot and Pais (2006) as they showed how hurricane Katrina gave rise to polarised social responses depending on the given community race and class. It was found that low income black home owners from New Orleans were far more vulnerable than white Americans both during and after the hurricane. They were found to lack the emotional, financial and political support they desperately needed.

INCOME

One of the most discriminating variables dividing communities can be household wealth. Less affluent households are very much more likely to struggle in terms of their financial resilience and subsequent economic recovery following the onset of a natural disaster (Burton et al 1992). Masozera et al (2007) showed this phenomenon with their case study of New Orleans and the differential exposure among income groups. It was noted that the poorest areas of the city were more adversely affected than the more affluent regions, taking longer to recover despite suffering similar levels of flood damage. Factors such as a lack of savings, insurance protection and unsecured loans have meant that disenfranchised areas find it harder to recover. Such differential patterns of destruction have been more notable in historic earthquakes, where building stock and affluence have been shown to be highly correlated. A magnitude 7.4 earthquake in Izmit, Turkey (1999) and a 7.5 event in Guatemala City (1977) were both later dubbed as 'class quakes' due to the poor areas of the respective cities being infinitely worse affected than more affluent suburbs (Blaikie et al 2004). There was shown to be an intrinsic link between building code adherence, quality, and its structural performance during an

earthquake. These factors were found also to be a reflection of affluence, with higher income families living in seismically more robust houses (Blaike et al 2004).

GENDER

Gender is not factored within the methodology described in later chapters of this thesis geodemographic classifications are based around aggregated statistical data to which gender does not provide enough level of differentiation to make it an effective method of classifying an area. However, it should be recognised that gender plays a very important role in social vulnerability as women are often noted to be more vulnerable than men during previous natural disasters (Morrow 1999). Morrow's work highlighted how disasters take place on a 'gender terrain' and that vulnerability factors such as financial and emotional stress have a disproportional effect on women more adversely than men. This was noted in the aftermath of hurricane Andrew in 1992 (Morrow 1999).

POPULATION DENSITY

Population density is also considered a significant factor during evacuation procedures (Johnson and Ziegler 1986; Chakraborty et al 2005; Dow and Cutter 2002). Johnson and Ziegler (1986) highlighted how densely populated areas are more difficult to evacuate than rural regions due to factors including traffic congestion and urban infrastructure. Chakraborty et al (2005) use population density as a key input in their social vulnerability evacuation index for Hillsborough County, Florida. Assessing the implications of evacuating densely populated areas in the event of a hurricane, population density was considered the most prominent factor to increased vulnerability due to extreme traffic congestion during evacuation. Residents wishing to drive inland, and away from the more hazardous coastal areas before and during a hurricane have been subject to very long delays; a phenomenon made worse in highly populated urban areas. Such assumptions have been corroborated from historic evidence during civil evacuation procedures, including the notable traffic and delays that occurred on the

interstate prior to hurricane Floyd (Dow and Cutter 2002) and later hurricane Katrina (Litman 2006).

2.4.4 INDICES AND RANKING PROCEDURES

As discussed previously in section 2.4.2, there has been wide spread use of social vulnerability indices in recent years, at both a government and academic level. The United Nations Development Programme (UNDP) has begun to measure the risk of death posed by particular hazards compared to the differential exposure rate. This has included analysis of three hazard types; earthquakes, cyclones, and floods (UNDP 2004). The South Pacific Commission (SOPAC) has created an index for Environmental vulnerability with 50 indicators related to human environments (SOPAC 2005).

Likewise from the literature, there are numerous examples showing the effective use of a social vulnerability index. Clark et al (1998) assessed coastal community risk to storm surges using census blocks in Massachusetts, USA. Rygel et al (2006) used Pareto ranking to formulate an index for hurricane storm surges in any developed nation. Granger et al (2001) used 31 indicators in their Cities project to identify vulnerable census districts. More relevant to natural disasters has been the development of the Earthquake Disaster Risk Index (Davidson 1997) or more lately the hurricane Disaster Risk Index (HDRI). Both are based on the premise that risk is the consequence of four factors:

- Hazard
- Exposure
- Vulnerability
- Emergency response and recovery

More specific to the concerns of this research, there have been several studies in recent years to quantify social vulnerability around volcanoes. Dibben and Chester (1999) carried out field

work on the island of Sao Miguel in the Azores. Their work helped show how a low status of risk perception among the people increased their vulnerability. Although not an index, but rather as part of the EXPLORIS project (EU) (Spence et al 2004; Zuccaro et al 2008) Robin Spence provides a thorough assessment of the likely casualty impacts if there was a forthcoming eruption of Mount Vesuvius. Aside from dynamic pyroclastic flow models, Spence used population data and housing structure to devise a vulnerability study of building stock and hence create an impact assessment. Marti et al (2008) used a similar approach in their work on building exposure and volcanic risk assessment of Icod de los Vinos, Tenerife (Canary Islands). However, it could be stated there have been limited attempts at creating a bespoke *Volcanic* Social Vulnerability Index.

Whereas the development of indices for the UNDP and SOPAC were undertaken with large aggregated datasets (at national or regional levels), Susan Cutter's development of a Social Vulnerability Index (SoVI) was based around classification of US census data at either the output area or county level (2003). The hazards-of-place model (Cutter 1996) aimed to capture a quantitative, objective measure of social vulnerability. Publishing the methodology in 2003, Cutter proposed a transferrable 'recipe' based on applying PCA to pre-defined census indicators. Preserving only highly correlated variables, an additive model ranked the counties, with the highest scoring 20% percentile being classified as having greater social vulnerability (Cutter al 2003).

2.5   MULTIVARIATE STATISTICS

The application of multivariate statistics forms the principle methodology in the construction of neighbourhood classification systems. As geodemographics is the principle basis of the models presented in this research, this section provides both a narrative background on how such systems became widely used in the commercial sector, open source initiatives in recent years and the methodological principles of those statistical techniques used in their construction, such as hierarchical and non-hierarchical cluster analysis.

2.5.1   NEIGHBOURHOOD CLASSIFICATION SYSTEMS (GEODEMOGRAPHICS)

As this thesis makes extensive use of neighbourhood classification systems (also termed geodemographics), it is important to review the literature on both the statistical methodology of these databases and their current use in commercial, academic and public sectors.

The dictionary definition of geodemographics states it is "*The study and grouping of the people in a geographical area according to socioeconomic criteria, especially for market research.*" (CED 2013).

Whilst this definition is factual, it does not seek to capture the diverse range of applications that geodemographics has beyond consumer markets. Likewise, although socio-economic criteria is an integral input into the creation of the neighbourhood clusters, it does not elicit its use in capturing a household's lifestage or behaviours. This last point is something that is a key component of geodemographics. Jan Kestle, founder and president of Environics Analytics provides a more expansive definition in his blog, '10 reasons to use geodemography'. Kestle states that "*Geodemography is a branch of market research that assigns the attributes of small areas – usually neighbourhoods – to the consumers who live within them and, based on this assignment, divides the consumer marketplace into meaningful segments that are locatable and reachable. The discipline leverages spatial and mathematical patterns in how people live and shop to help marketers make inferences about consumer behaviour.*" (Kestle 2003)

This latter description begins to touch upon the key virtue of using geodemographics that transcends its use in solely consumer insight. Aside from business interest, geodemographics is able to capture the processes by which population settlements evolve through time. Such capability has meant that geodemographics have attained widespread use in developing social policy (Longley et al 2008), understanding our communication preferences (Acorn 2011; Mosaic 2011) and directing political campaigns (Webber 2006).

The origin of geodemographics can be traced back to the work of Charles Booth (Orford et al 2002). In 1891, Booth segmented residential roads into different socio-economic categories based on the concept of neighbourhood commonality. Booth used the 1891 census survey to produce a multivariate classification of London districts by social class (Orford et al 2002). These principles of commonality were further strengthened by the Chicago school of sociology in the 1920's and 1930's, whereby research focused on how settlement patterns in urban environments were highly correlated with a household's lifestage and economic standing (Harris et al 2005).

Wyly (2001) identifies this progression as coinciding with the advent of digital computing in the 1950's and 1960's, which provided a catalyst for the origins of modern geodemographics. Factor analysis and Principle Component Analysis (PCA), previously used for the purposes of educational psychology at the Chicago School began to be used by urban geographers seeking to understand census data for households and population (Wyly 2001). One of the key advantages of using PCA in geodemographics was to help reduce the number of variables required to successfully classify neighbourhoods. Before the advent of modern computing, such multivariate analysis was extremely time consuming and thus, techniques such as PCA helped determine the most influential variables required in a geodemographic.

Proceeding this, the modern application of geodemographics (with particular focus on the UK) takes its origin from Richard Webber's work at the Centre of Environmental studies as he used such techniques to define areas of deprivation in Liverpool using the 1971 census data. A seminal paper entitled 'The utility to market research to the classification of residential neighbourhoods' began to recognise the commercial application of Webber's work (Baker et al 1997). Webber later went on to develop two leading UK commercial classifications, CACI's Acorn and Experian's Mosaic product.

Since then, the growth of client segmentation and marketing analytics appears to have been extremely rapid and increasingly more diverse in its application. Aside from the census source data, additional data sources have been used to help correlate and gauge attitudinal, lifestyle

and behavioural aspects that are not afforded from census data alone. This relationship hinges on the premise that it is not only socio-economic variance that can be shared between spatially diverse neighbourhoods but also cultural, behavioural and attitudinal differences. Supplemented by survey variables derived from sources such as consumer lifestyle surveys, telemarketing information, and national crime surveys, classification groups can be defined to describe a neighbourhood's 'average' profile. This 'insight' has become a highly sought after strategic tool for directing marketing campaigns and customer retention. There are now several competing neighbourhood classifications systems available in the UK (e.g. Acorn, Mosaic, OAC and Cameo).

In recent years, there has been a divergence as geodemographics has started to become integral to many developed nations around the world. Likewise, other changes in this market have seen the emergence of bespoke geodemographic systems aimed at segmenting emerging markets and diverse areas of public awareness. For example, in the UK there are commercial classification systems purely focused on IT literacy, such as Acorn's e-types (Acorn 2011) or Mosaic's TrueTouch categories (Mosaic 2011). Similarly, there are geodemographic systems focused solely on health concerns, such as HEALTH Acorn (Acorn 2011) which defines communities more pre-disposed to certain ailments given their spatial location and commonalities of behaviour.

As well as the commercial products, it is important to note that there has been a steady and paralleled increase in 'open' geodemographic initiatives. One such example is the UK Output Area Classification (OAC) devised by Dan Vickers for the Office of National Statistics (ONS 2006). Aside from essentially offering free marketing insight data, a series of white papers and technical information produced in conjunction with the release of the OAC has served to educate many on how these once perceived 'black box' systems are constructed.

Likewise, there is a large volume of academic research on classification construction for academic purpose. Longley et al (2008) explore how a classification system can be used to help target wider catchment of university students, particularly in less privileged areas. Similarly,

research by Petersen et al (2011) explored the application of geodemographics in targeting health campaigns.

The application, principles and methodological assumptions involved in geodemographics have not been without their critics. The very use of geodemographics in the growth of consumer insight and targeted marketing has long been questioned on an ethnical basis, with the implication that such use is tantamount to '*electronic surveillance and the erosion of privacy*' (Goss 1995). The notion that consumers can be 'pigeon holed' into specific behavioural segments on the basis of multivariate statistics makes many feel uneasy. It also questions personal identify and the concept of personal freedom. Voas and Williamson (2001) further highlighted the fundamental problem of data aggregation in neighbourhood segmentation. Their paper, *The diversity of diversity* used the 1991 census data and *superprofiles* to demonstrate how similar geodemographic segments may include inherent individual diversity that will never be captured at an aggregated level. This is particularly true of the behavioral (non-census) information of individuals in a given enumeration district. Though the areal census statistics of two districts may be identical, there is no assurance that individuals in the respective areas are alike. The concept of statistical bias is known more commonly in geography as the Modifiable Areal Unit Problem (MAUP) and regards the fact that the aggregation of discrete observations may dramatically affects the results and assumptions of an area depending on how the boundaries are defined (Openshaw 1984). Similarly, another geographical concept that is at the core of the debate on geodemographics concerns the *ecological fallacy*. Inferences about an individual based on the areal statistics (or averages) of the bounding region can be greatly misleading (Freedman 1999). For these reasons many have questioned the relative merits ethnical use of geodemographics (Crampton 1995).

Despite the controversy, geodemographics have continued to develop and there's now widespread use of classifications for marketing, planning, public health, education, politics, and academic research.

However, with specific regard to the contribution of thesis, neighbourhood classifications systems do not appear to have been used previously in the study of natural hazards and DRR. The only known publications on this topic (at the time of writing and to the current understanding of this author) regards the research undertaken in the application of geodemographics to volcanic hazard management (see Willis et al 2010). An account of the associated research output from this thesis can be found in thesis outputs section of this document. Many of the concepts and findings covered in these publications are explored in further detail within this thesis.

2.5.2   CLUSTER METHODOLOGIES

Clustering is an essential component in the construction of geodemographics and accounts for a large portion of the methodological process required in developing the thesis presented here. Therefore, in this section, the theoretical principles of diverse clustering methodologies are discussed as well as their relative application in different industries.

There are several methods of clustering data and no single definitive solution exists. Among these numerous techniques are hierarchical methods, non-hierarchical and the use of neural networks; a form of highly evolved clustering algorithm. For the purposes of this research, and given the expansive nature of this subject matter, only methods more familiar with neighbourhood classification systems are discussed in further detail. The following paragraphs provide an overview of the cluster techniques that were extensively reviewed during the course of this research.

HIERARCHICAL METHOD (WARD)

In hierarchical cluster analysis, techniques typically fall into two categories, *Agglomerative* and *Divisive*. The former applies a bottom-up approach, where each observation begins in its own cluster, slowly reducing as pairs of clusters are matched. The latter method applies a top-down approach, with all observations in one cluster, which is duly split into more numerous clusters.

One such method of Agglomertative clustering is the Ward's method. First devised by Joe H. Ward, this method was used to cluster a large number of objects, symbols or persons into smaller numbers of "mutually exclusive groups" (Ward 1963). Each cluster is based on the similarity of distance as fewer and fewer clusters are created as the process continues. Hierarchical clustering is often referred to as 'connectivity' clustering. The number of clusters is not pre-specified and provides a bottom-up hierarchical approach to clustering. The process can be summarized:

- Each object is its own cluster file to begin with

    Cluster file x = C1, C2, C3, .. , C_{n-2}, C_{n-1}, C_n

- The closest cluster to each other cluster is computed {C_i, C_j}

- These paired clusters are removed from the cluster file x.

- C_i, C_j are merged to form a new cluster.

- The process returns to the first step (without C_i and C_j) and continues.

- The method will continue until there is only one cluster remaining.

NON-HIERARCHICAL (K-MEANS)

This is a non-parametric clustering method. Therefore, no assumptions are made about the data regarding variance or distribution prior to the algorithm being run. The object is to reduce the within cluster variability. One of the major differences with the K-means method is that it allows the user to pre-define a desired number of cluster outputs. It can be regarded as a deterministic method. A series of centroid locations (seeds) are first determined by the algorithm, and thus, iterations occur whereby the clusters are defined by their relative proximity to data observations (Aldenderfer and Blashfield 1984).

The iteration is key to establishing seed locations. It is on the basis of these seeds that clusters are formed around.

K-means is a form of cluster analysis which aims to partition a given number of observations (n) into a user defined amount of clusters. Figure 2-9 is a 2D conceptual illustration of a Voronoi diagram to demonstrate how a given set of observations may be clustered based on their respective coordinates. The observation is assigned to the cluster with the nearest mean. It is an iterative process relocation algorithm with the aim of reducing the sum of squared deviations within each cluster (Aldenderfer and Blashfield 1984).

Given a set of observations (x1, x2, x3..), where each observation is a d-dimensional real vector, k-means clustering partitions the n observations into k sets (k ≤ n) so as to minimize the within-cluster sum of squared deviations.

$$\arg \frac{min}{S} \sum_{i=1}^{k} \sum_{X \in jSi} \| Xj - \mu i \|^2 \qquad \text{2)}$$

K-means has become a particularly common form of classification for multivariate data, and is very widely used in both commercial and open source geodemographics (Harris et al 2005). A large part of the reason for this is that the algorithm provides a robust method of minimising cluster variation whilst also allowing the user to set the number of clusters created. This is likely to be a significant consideration when commercial users have to then create maps, statistics and convey to lay audiences about the demographic trends of data.

Due to fundamental differences in hierarchical and non-hierarchical cluster methods, the assumptions and outcomes of these methods can vary greatly and the preferred method is not always clear. It could be stated that K-means analysis is more suited to multi-variate data with a normal distribution. This is because the algorithm seeks to iterate around seed locations that are defined by extreme data observations. Therefore, if an observation variable is particularly skewed, or a dataset has numerous outliers, the K-means analysis may define these as single-entity clusters. Thus, the process may result in very unequal sized cluster members. Conversely, in agglomerative hierarchical cluster analysis, if data has a normal distribution, it may not adequately define the smaller or niche cluster groups that K-means would define due to the

bottom-up approach of cluster membership. These differences mean that it is often necessary to perform exploratory data analysis to determine the most appropriate method for a given dataset.
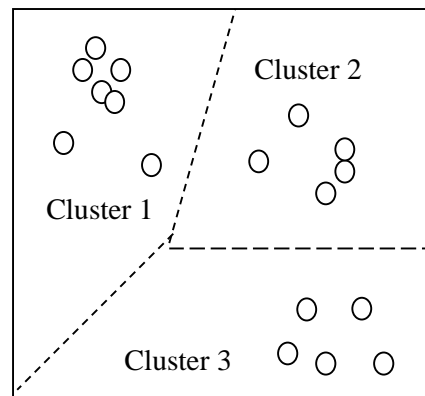
FIGURE 2-9 - A CONCEPTUAL *VORONOI* DIAGRAM FOR THE K-MEANS CLUSTERING ALGORITHM

ALTERNATIVE CLUSTER METHODS

As well as hierarchical and non-hierarchical cluster methodologies, it is worth noting that there are alternative methods to grouping census data.

The 1991 census data was split at a district level by Openshaw's 'Self-organising maps' (Openshaw and Turton 1996). This technique grouped enumeration districts into defined clusters. Likewise, it is also possible to use a 'Two-step' clustering method which accounts for categorical data as well as numerical information.

Unlike the connectivity models (hierarchical) and centroid models (k-means) discussed previously, other forms of cluster analysis include the use of distribution models. For example, the Expectation-maximization algorithm is commonly used in computer vision and image processing. In this method, a data set is modelled with a fixed number of Gaussian distributions to more closely fit the dataset. Objects are then assigned to the Gaussian distribution they most likely belong to. Although a strong clustering method, this clustering technique relies heavily

on the complex distribution models created by the user and therefore requires a thorough understanding of data parameters.

Similarly, there are density clustering models such as the DBSCAN method (density-based spatial clustering of applications with noise), which defines clusters as connected dense regions in space (Ester et al 1997). Essentially, areas of higher density are defined as separate cluster seeds from sparse areas. This method is similar to hierarchical methods, in that it measures the distance of connecting points. The algorithm begins by selecting an arbitrary point location, and depending on the relative density of points, a cluster will be started or else defined as noise. When a point is found with sufficient density of points around it, they will also be assigned to the cluster. DBSCAN is not widely used in geodemography but is one of the most popular methods of clustering, particularly in digital imagery analysis, genetic sequencing and spatial analysis.

2.6   CHAPTER SUMMARY

This chapter highlighted that the very definition of a volcano can be questioned, given that many volcanoes may share the same sources, have multiple vents, or consist of subterranean sea mounts. The global pattern of volcanoes was then discussed and their displacement both along and within some tectonic plate boundaries. The driving forces of plate tectonics are detailed as well as the ongoing research and debate as to origin of these principle forces. When comparing global epidemiology data, volcanic eruptions and their associated severity are shown to have caused fewer fatalities than more frequently observed hazards such as flood and drought. Similarly, attention was drawn to the long repose period of the largest eruptions and that the most deadly eruption of the last 100 years accounts for 40% of known fatalities from a volcano. A detailed account of the most common volcanic hazards then followed, identifying tephra, PDCs, and lahars as the main concern for populations living close to active volcanoes.

This next section considered the definition of vulnerability and how it is found to be highly subjective and often dependent on the context in which the term is being applied. Multiple conceptual frameworks have been developed to describe the relationship of vulnerability within the concerns of livelihood sustainability and disaster risk reduction. A common theme of these models is the appreciation that vulnerability is created through external forces, such as government or political doctrine as it is from local influences or the antecedent physical risk.

Likewise, reviewing the literature of social vulnerability, there are numerous factors that have been found to contribute to an individual or community being more vulnerable before, during or after a natural disaster. Experience has shown that an individual's age, gender, ethnicity, and income have been highly correlated with disaster experience, given previous research findings. Based on these *indicators,* several quantitative methods have been created striving to derive universal metrics of vulnerability. These initiatives have been aimed at different scales of geography, for different perils, and using different methodologies, yet none have aimed to define vulnerability by defined neighbourhood profiles.

Conversely, geodemographics provide a widely used method of describing commonalities in neighbourhood profiling yet have never been previously applied to describe natural hazard social vulnerability. This chapter summarised their origin and provided an overview of the various multivariate cluster analysis techniques applied in their methodology.

3. METHODOLOGY

This chapter describes how multivariate cluster analysis can be applied to the identification of socio-economic indicators of social vulnerability as well as proposing a new conceptual framework and applied methodology with which to classify such neighbourhoods in volcanically active regions of the world.

The application of geodemographics has been developed in two separate phases of work; the first using an existing commercial geodemographic (Mosaic Italy 2007) and the second phase utilises publically available census data to construct and then analyse a bespoke geodemographic database for Ecuador. These phases provide the basis of social and physical vulnerability assessment, as well as the application of third party marketing/survey data and model validation. The methodology outlined in this chapter has been developed in an iterative manner during research and was focused on addressing the following key aims, as identified in section 1.1:

1. *Investigate and review the existing capability of risk classifications in DRR more broadly.*
2. *Develop an original methodology that can be used to assess social vulnerability based on classifying neighbourhoods in volcanically active regions.*
3. *Apply and validate the methodology (aim #2) in different spatial and cultural settings.*
4. *Assess whether the methodology could be transferrable and how it could be used in practical DRR outreach and communication.*

In reviewing the existing derivation and capability of risk classifications (as defined in aim #1), a guiding model and conceptual framework was developed that would provide a foundation for the basis of the relationship between social vulnerability indicators and neighbourhood classification systems.

3.1 A FRAMEWORK TO DEFINE NEIGHBOURHOOD BASED SOCIAL VULNERABILITY

Based on current knowledge, associated literature and vulnerability models discussed earlier in this thesis, the human experience of volcanic disasters is not solely caused by the hazard but is instead a complex combination of multiple root causes. This includes influence from an individual's family, local community, government (both local and central), cultural values as well as their innate coping capacity. Taking this concept, as well as the previously defined notion that certain social vulnerability indicators can be used to identify the relative risk to a population, a new methodological framework is proposed in this chapter.

Figure 3-1 illustrates how the root cause of a disaster is not always clear given that there are numerous internal and external sources that contribute to a 'disaster'. The impact of macro forces such as globalisation, economic doctrine, and a nation state's political stability have radically changed demographic and global migration patterns in the last 50 years (Chester et al 2000). Such changes have resulted in significant waves of inter and intra-state migration (Castles and Miller 2003). This pattern is exacerbated in developing nations, where aside from international migration; there has been significant growth in internal migration, witnessed through rural-urban migration. Such movements have resulted in increasing urban populations in many cities, as well as the formation of illegal settlements, and the use of unsafe land for living (Carrion et al 2003). Based on these external root causes as well as the notions discussed earlier in section 2.4.3 (that some individuals or groups are more pre-disposed to natural disaster social vulnerability), vulnerability *indicators* can be used to further highlight community frailty.

The application of geodemographics within this framework is based on the premise that by using multiple census variables (also including those characteristics not associated with social vulnerability) and hierarchical or non-hierarchical cluster analysis, *neighbourhoods* can be identified, not just as discrete areas that share vulnerable traits. Given that geodemographics is

based around the commonality of neighbourhood types, it is able to encompass a wider remit and practical application than simply the statistical similarity of aggregated administrative areas.

For example, two neighbourhoods may be spatially very far apart but may quintessentially be very similar in their demographic, cultural and attitudinal aspects. The proof of this concept is evidenced by the successful application of geodemographics in both the commercial and public sectors. It is used to direct marketing campaigns (Harris et al 2005), provide insight on client behaviour/attitudes, and for political parties to gain leverage in electoral strategies (Webber 2006). Based on this premise, it is proposed here that natural hazard vulnerability, risk perception, and coping capacity can be identified within similar communities and targeted for DRR purposes using geodemographic techniques. In Figure 3-1, this section of the framework is referred to as the *application* component. Although specific mention is only made in the application section of the framework to disaster *mitigation* and *recovery*, this is not seen as an exhaustive list and a classification tool for DRR would also have application in disaster response. A specific application of this research is presented in a later section regarding how geodemographics can be combined with survey data to target neighbourhoods for disaster preparedness, education and community outreach.

Causes   **Indicators**   **Application**

Less state control

Age (Elderly)

Age (Young)

Women

Disability

Ethnic groups

Employment

Population density

Income

More state control

Demographic drivers: Not necessarily due to external factors but can often be aided/mitigated by intervention (e.g. gender)

Macro/External drivers: This can be due to direct/indirect macro-economic/political doctrine or cultural beliefs (e.g. globalisation, IMF intervention, WTO)

Mitigation (Planning / Education / Awareness)

Neighbourhood level
- Evacuation
- Access to resources
- Financial recovery
- Household exposure
- Physical risk of hazard

Recovery (Risk assessment /Targeted aid)

FIGURE 3-1 - CONCEPTUAL FRAMEWORK FOR IDENTIFYING NEIGHBOURHOOD-LEVEL SOCIAL VULNERABILTY TO NATURAL HAZARDS

The framework presented here takes its origins from other such conceptual models, such as the Pressure and Release model (PAR) (Blaikie et al 2004) which aim to describe the complex interdependencies of state and local governance that contribute towards a natural disaster, whilst providing a framework aimed at describing indicators that could be used for DRR application. As highlighted in Figure 3-1, there are certain social vulnerability traits that are very often symptomatic of disaster vulnerability. These *indicators* can include (but are not limited to) age, ethnicity, gender, disability, population density and income. Whilst it is acknowledged that these factors play a pivotal role in disaster vulnerability, this framework does not consider their relative weighting – which varies significantly, depending on both the disaster and the contextual setting. The framework also attempts to separate vulnerability

indicators into those that are likely to have more capacity for state control, such as income and employment and those that have less state control, such as gender and disability. For example, whilst a nation state has a direct and controlling influence on income and economic factors such as the GDP, demographic vulnerabilities (e.g. age) would appear to have less capacity for state intervention.

The final component of the framework is aimed at providing a practical measure or *application*. The model highlights that the vulnerability *indicators*, previously discussed, can be identified at a neighbourhood level. Noting vulnerability concepts at a local level, such as *access to resources, financial recovery, household exposure, and the physical risk of the hazard,* the framework presents the notion that such traits can be identified at a neighbourhood level. The model doesn't specify the methodology for how to do this, but it does propose possible DRR applications, such as post-disaster recovery, targeting aid, mitigation and education.

After creating the conceptual framework illustrated in Figure 3-1 it was then necessary to develop a methodology and apply this model in practical study areas, as this would provide a suitable procedure for assessing the relative merits, limitations and application that such a methodology could provide. If multivariate cluster analysis was to give insight into DRR in volcanically hazardous areas, it would be important to test this hypothesis in different volcanic and demographic settings.

Each volcano is tectonically unique and therefore the hazards posed by an eruption can be very different from one area to the next. It would be important to change the volcanic setting as the types of hazards associated with the 'at risk' population would warrant different vulnerabilities. Similarly, by changing the location of the study settings, the nature of the base population would change. This would mean significant differences in the geodemographic characteristics of the 'at risk' groups. Likewise, the attitudinal and cultural origins unique to a country or culture would provide significant challenges for disaster risk reduction classification.

A two phase approach was taken in applying the conceptual framework, whereby the **first phase** would be focused on using a commercially available geodemographic and then in the **second phase**, a bespoke geodemographic would be developed from first principles using freely downloadable census data and where available, supplementary survey data.

3.2   DATASETS AND STUDY REGIONS

Although volcanoes are spatially divergent across much of the world, it could be argued that several volcanic areas pose far more danger to mankind than others. Given the ultimate purpose of this research, it seemed prudent to focus the modelling on the more hazardous volcanic regions as any contribution made from this research could thus demonstrate its scope and success in a realistic geographic setting. With this in mind, it was necessary to consider what constitutes a *dangerous* volcano; a subject which is largely subjective and open to divergent interpretation.

The fundamental metric of choosing study areas was determined by selecting volcanic areas that posed *a significant and realistic threat to a large human population.* To do this it was necessary to consider both the explosive/eruptive history of the volcanic edifice as well as the prominence of a high human population settlement near the hazard. However, it remains a moot point about which volcanoes are the most dangerous and the decisions made so far in this research represent a very subjective interpretation. Factors increasing this ambiguity include the fact that eruptions are not entirely predictable and risk to human life depends as much on the size and nature of the volcanic hazard as it does on the distance of the population. For example, the eruption of the Icelandic volcano Eyjafjallajökull (April 2010) sent ash spewing into the atmosphere that caused a clear and present danger to all aircraft flying in Northern Europe for several days. This hazard had not previously been considered a major concern in Europe.

To help assess suitable study areas, David Chester's analysis (Chester et al 2001) regarding a survey of cities with increasing exposure to volcanic eruptions was consulted as it provided a thorough assessment using multivariate based risk assessment of cities lying close to dangerous volcanoes. Chester's assessment was focused on identifying settlements that were highly correlated with the following parameters; large population size, close proximity to volcanic edifice and a history of volcanic hazard(s) that have posed a risk to human life previously. In doing so, the following global cities were identified as being of major concern regarding their proximity of less than 25km from an active volcano; Quito (Ecuador), Arequipa (Peru), Kagoshima (Japan) and Naples (Italy).

It was also necessary to consider the practicalities of where the demographic data for this research would come from. To create a geodemographic and vulnerability indices it would be necessary to choose only those locations where there was an available neighbourhood classification system or else freely available census data to construct one. After researching several relevant organisations and requesting available datasets, Richard Webber from Experian plc kindly offered to loan the use of the Mosaic Italy 2007 dataset for this research. Mosaic Italy 2007 is a commercial geodemographic classification that was provided in a database format along with a series of related index tables. With direct focus on the area around Vesuvius, Italy, this dataset would form the substantive dataset for phase one of the proposed methodology.

In considering the second phase of this research, and to further evaluate the original hypothesis, it was decided that another study region was needed. After careful consideration, Ecuador in South America became a suitable candidate. There were several reasons for this; it is one of the most volcanically active regions in the world; it has the highest population density in South America and with millions of people living in harm's way of multiple active volcanoes (Chester et al 2001), it provided an opportunity to test the conceptual framework in a different geographic and cultural setting. From a research perspective, Ecuador's social and economic status made it markedly different from the first phase of research on Italy. This was an

important consideration as Ecuador would represent different challenges in creating a transferrable vulnerability methodology. Likewise, the availability of freely downloadable census data for Ecuador (at an individual level) provided a useable household and population dataset with which to construct a geodemographic database.

3.3   CONTEXT TO STUDY REGIONS

This section is intended to provide a narrative background to the study regions presented in this thesis. The physical and social risks posed by a volcano to the surrounding population cannot be considered in isolation to the geographic, social, and economic context of their surroundings. Such understanding is implicit in disaster experience and is duly represented in DRR conceptual frameworks. It is central to the core contribution of the methodology and application presented here and provides the context for discussing model results in later chapters. In phase 1 of development, Italy and the surrounding Campania study region are considered as well as the migrant and economic drivers that have meant Naples is one of the poorest cities in Italy. In progressing this research agenda to Quito, Ecuador (phase 2), the section also discusses the colonial history of the country as well as increasing impact of globalization in recent years that has ultimately shaped Ecuadorian macro-economic policy.

3.3.1   CONTEXT TO STUDY AREA: ITALY

Located at the boundary of the Eurasian and African tectonic plates, Italy's geology has led to a significant amount of seismic and volcanic activity. There are several active volcanoes in Italy but the two most iconic symbols of these colossal geohazards are Mount Vesuvius and Mount Etna. They are arguably Europe's most dangerous volcanoes and have very large populations living within close proximity of the various hazards.

Before 1861, Italy was made up of a series of separate provincial states. In what is termed the *Unification,* Italy finally became a single nation state. Between the *Unification* and the so-

called *economic miracle* of the 1950's, Italy experienced the migration of up to 25 million Italians, the largest Diaspora in modern times (Castles and Miller 2009).

In the early 20th century, most migration from Italy was taking place from the south of the country, where food shortages and low wages were very common. The recipient of this mass migration was the Americas (both North and South). This flow of migrants slowed due to the onset of the First World War as well as acts of government to restrict population flows. Emigration from many large cities was notable. Naples was particularly affected by the unification (Castles and Miller 2009). After losing all its bureaucratic power as the capital of the Campania province, there was mass emigration from the city and poverty related outbreaks such as cholera became rife.

Mass migration began again in 1921 and continued up to the Second World War as Italy had been left in economic and political turmoil. As Italy witnessed the rise of a fascist regime, an estimated 1.5 million people left the country for the United States. Significant recipients of the Diaspora included Argentina, where it is estimated that around 50% of the current population are of Italian descent and Brazil, which is estimated to have around 25 million Italian Brazilians (Castles and Miller 2009).

ECONOMICS

Following the emergence of Mussolini's fascist regime in the 1920's, the Italian economy, which had largely been laissez-faire with regard to tax regulations and trade, progressively took on a more active interest in state finance. This continued after the great depression of 1929, when banking assets were nationalised. This economic model of government intervention in wage fixing became known as corporatism.

Italy's subsequent involvement in the Second World War was crippling to their economy as they contributed to the war fund of the Axis. Following the invasion of the Allied forces in 1943, Italy's economy all but collapsed. Although Italy's economy was at its lowest point since

the turn of the 20th century, between 1950 and 1960, there was something of an economic miracle. The Italian economy was transformed from one of poverty, mass emigration and agriculture to a leading industrial nation. This miraculous change was the result of several factors including Italian entrepreneurship in the post war years together with a growing manufacturing base (Crafts and Toniolo 1996). In the period 1957-60, there was an increase in manufacturing production of 31%. Despite a decrease in financial fortune during the oil crisis of 1973, Italy's economy has been steadily growing and was considered in 2009 to be the world's 7th largest economy by the IMF and World Bank (Crafts and Toniolo 1996). However, there are still many social inequalities in Italy and there remains a clear North-South divide with regard to social mobility. Italy remains a dichotomy, with the south of the country having one of the lowest GDP's in Europe (16,294 Euros) and the North being one of the highest (32,900 Euros). Such differences reflect the agricultural nature of production in the South compared to the industrial North of the country.

Among the more deprived cities of Southern Italy is Naples. It is located between two very significant volcanic areas in Italy; Mount Vesuvius and the Phlegraean Fields. Because of its close proximity to Vesuvius, it is an integral part of the first study region. Naples is one of the oldest population settlements in the world, and despite successive years of outward migration, the population of the metropolitan area is estimated to be around 4.4 million (ISTAT 2001). This represents a significant population in harms' way of volcanic hazards.

3.3.2   CONTEXT TO STUDY AREA: QUITO

The country of Ecuador is a republic in South America. It is bordered by Peru to the South and East and by Columbia to the North. The western coast lies adjacent to the Pacific Ocean. Located on the equator, the capital city of Quito is found in the middle of the Andes, approximately 2800m above sea level. Due to numerous volcanoes found along the Pan American highway in this stretch of the Andes, it is commonly known as 'volcano alley'. With the current population estimated at 15 million, 60% Ecuadorians live in large urban settlements

such as the cities of Quito (2.5 million) and Guayaquil (3.1 million). The rapid population growth in urban areas during the 20th century has meant that human settlements are often formed by illegal barrios rather than government initiatives. Ecuadorian ethnicity is predominantly a consequence of indigenous Inca, historic migration and colonial links to Spain. Although Ecuador has officially been independent from Spain since 1830, both political and financial instability have been an ever present factor in the country's development.

POPULATION SETTLEMENT

In the period from 1800-1970, an estimated 21 million immigrants came to South America (Castles and Miller, 2009). The vast majority of these workers were Spanish, Italian and Portuguese. The largest share of movement occurred before the great depression of the 1930's resulting in international migration slowing down greatly. Despite further large-scale migration to Venezuela (due to state led incentives) in the post-war era, migration from Europe declined sharply (Castles and Miller, 2009). Instead, a new wave of intra-continental migration began to develop in South America.

Migrant flows began to occur in Ecuador and in most Latin American countries. These movements were primarily the result of economic development in the leading nations of South America (Brazil, Argentina) as a new labour force being sought. From 1935 until the mechanisation of labour in the 1960's, unregulated seasonal flows of workers from Bolivia to Argentina developed. Similarly, in the 1960's, Paraguayan and Chilean labour flows came to northern Argentina and Patagonia. Encouraged by the higher wages and prospect of regular work, South America was a hub for intra-continental migration (migration within the same continent). Added to this there were regional incentives such as MERCOSUR and GRAN (Derisbourg 2002) which actively encouraged more economic cooperation and migration between the South American nations. Unfortunately, and a recurring theme with regard to South American migration data, getting accurate figures for migrant numbers can be difficult.

Despite these large scale movements within South America, Ecuador has not been the recipient of particularly large scale migrations since its colonisation by the Spanish in 1531.

In the last century, the largest process of migration in Ecuador has been internal. The movement of rural workers to the cities has characterised the 20th century. This is also the case for most developing nations in South America as mechanisation led to a collapsing rural economy. There is also a growing movement from the overpopulated highlands of the *Sierra* to the virgin areas of the *Oriente* and coast.

As of May 1997 there were around 14,500 people who concerned the UNHCR (United Nations High Commissioner for Refugees) in Ecuador. Most of these people were from Cuba and Colombia, living and working in the capital Quito. There were also a number of rural refugees from Colombia, living in the provinces of Esmeraldes, Corchi, and Sucumbios (near Colombia). In 1999, there were 277 refugees under the UNHCR's program in Ecuador, mostly from Colombia and Peru. However, Ecuador, not unlike other Latin American countries, was beginning to experience an increasing number of asylum seekers from outside the Americas in 1999. Migrants were commonly from Africa, Asia, and the Middle East. The net migration rate for Ecuador in 1999 was 0.55 migrants per 1,000 in the population. The total number of migrants in 2000 was 82,000. In that year worker remittances were around $1,317,000, or 9.6% of GDP (Encyclopaedia of the Nations 2011).

As well as those that have sought work voluntarily through migration to neighbouring countries, there has also been a substantial amount of forced migration in Latin America. Due to the fighting in Central America in the 1980s, approximately 2 million refugees left for countries such as the US or Cuba. There have been subsequent repatriations but many have stayed in the US or Canada trying to seek asylum. It is understood that from 1984-1994, 440,000 Central Americans applied for asylum in the US (Castles and Miller, 2009).

Political instability and poor economic fortune have definitely been the key factors in much of the migration flows out of Latin America in recent years. However, emigration from the region

has a stark demographic difference to earlier movements. There has been a noticeable 'Feminization' of migration mobility in Ecuador. Young women are becoming increasingly likely to migrate and find work away from Latin America as job prospects are higher as well as there being an 'opening' up of asylum agreements between many leading nations. Other trends include the recent inter-migratory routes, whereby populations move between large and mid-sized South American cities as well as to more suburban or peripheral city locations.

However, Ecuador's greatest contribution to global migration is currently as a net exporter. It is estimated that 11% of Ecuadorians live outside of Ecuador (Castles and Miller 2009). The vast majority of these migrants go to work in Spain or the United States. In 2005, an estimated 500,000 Ecuadorians were estimated to be living in Spain. Because of the number of those living abroad, foreign remittance is the second highest income source for the Ecuadorian economy. The biggest driver of these migrations has been the poor economic performance of the country in the 20th Century. Such migrations have been taking place since the 1950s when the economy of the nation began to flounder. Ecuador's president, Rafael Correa was himself a migrant in the US where he studied economics before returning to Ecuador to lead the current populist regime.

ECONOMICS

In Ecuador's capital, Quito, there are approximately 2.5 million people living in the shadow of the Guagua Pichincha volcano. Due to global and regional economic policy, Ecuador's urban population escalated in the latter stages of the 20th century and now represents 60% of the population that live in harms' way of volcanoes. Arguably the biggest factor driving this increased social vulnerability can be traced back to the adoption of the economic principles of Raul Prebisch.

Prebisch, an Argentinean economist and former head of the United Nations Economic Commission for Latin America and the Caribbean (UNECLAC or ECLAC) believed in the Marxist principles that countries wishing to develop economically in the modern world could

not do so under the capitalist and dominant Bretton-Woods system (previously adopted by the western superpowers). Instead, he believed that development had to be attained by working on the internal dynamics and trade of a country rather than solely exporting a countries natural resources to more developed nations (otherwise known as 'dependency theory'). 'This new world model was called Import Substitutive Industrialization (ISI) and became the de facto economic principle for most South American nations from the 1930's to the 1980s.

As well as the growth of state infrastructure, the major consequence of ISI for Ecuador was a mass rural to urban migration shift as populations flocked to the budding Ecuadorian cities in search of work. An economy largely based on Cacao exports became replaced with an inward looking market. Large scale manufacturing grew and there was a collapse in the rural economy and labour markets flooded to the primate cities of Quito and Guayaquil. The consequence of these social reforms to Quito was a phenomenal rise in the metropolitan population; from 1950 to 2001, the population grew by 1.2 million. This represented a staggering increase in a very short period of time. Quito now houses approximately 12% of the Ecuadorian population.

In terms of social vulnerability, it is these population rises, coupled with a general lack of civil planning that have helped create Latin America's biggest social problem, the barrios (informal housing settlements).

Figure 3-2 illustrates a Quito community living on the side of the Pichincha volcano. These informal housing settlements, most commonly found on the Northern and Southern peripheries of Quito are particularly vulnerable to natural hazards. Not only are inhabitants among the most deprived in the province, houses are often without basic living amenities (water, sanitation, telephone). Properties are often built illegally on the dangerous, unconsolidated flanks of the Guagua Pichincha volcano. These suburban fringes of Quito lie in the path of several natural hazards, such as flash flooding, mudslides, pyroclastic flows, tephra fall and lahars.

FIGURE 3-2 - QUITO BARRIOS (CARRION ET AL 2003)

It was this socialist economic reform has been the largest single influence of population movement in Ecuador. However, to suggest the vulnerable population around Guagua Pichincha is solely the consequence of Marxist ideology is to over simplify the situation. Urban population increases in Ecuador have also been achieved by a steadily falling mortality rate as health care improved and life expectancy progressively increased during the 20th Century. Likewise, it's hard to assess the possible population influence that opposing economic theories such as *Export Industrialisation* might have brought to a largely agrarian economy.

It's important to be mindful that the last 20 years of politics in Ecuador have started to show the trademark of an opposing neo-liberal ethos. And in some ways it is this last political (and global) phenomenon that keeps the barrios, as well as much of the developing world in a vulnerable position. The forces of *globalisation* have culminated in the large wage discrepancies and an ever increasing poverty gap that appears to hinder Ecuadorian social welfare and mobility.

In 1972, Ecuador's relatively stagnant agro exporting economy was transformed by the discovery of oil in the *Oriente* region of the country. As this was coupled with the global 'oil crisis' of the 1970s, Ecuador's economy grew by 10.4% per annum during 1972-76 (Middleton 2007). The 1970s were marked as being a decade of rapid industrialisation and rural to urban migration. The oil revenues became an attractive proposition for foreign investment and foreign debt started to escalate from 1976. By 1981 nearly 70% of the export earnings in Ecuador were needed to service the enormous foreign debt that had been building. Petrol prices began to rise, taxes were increased and GDP began to fall from 1981. It was following these disastrous economic circumstances and the rise in influence of the IMF (International Monetary Fund) that Ecuador was forced to adopt structural adjustment policies in order to access loans. It was these neo-liberal policies that several Ecuadorian governments followed until the collapse of the *Sucre* currency in 1999. As globalisation increased so did Ecuador's economic instability, culminating in the freezing of state accounts in March 1999, the closure of two thirds of the financial institutions and their adoption of the US dollar as the national currency (Middleton 2007). It should be noted that since the collapse, GDP has increased and foreign debt has been steadily reducing in recent years (IMF 2014).

Evidence suggests that absolute poverty in the nation is decreasing. Ecuadorian GDP increased from 24,605 to 87,495 USD in the period between 2001-2014 (IMF 2014). However, such assumptions should be tempered to comparisons on a global basis, where Ecuador's economy has dropped from the 61st to 63rd during the same period.

3.4   SOURCE DATA

ITALY

The Mosaic Italy dataset has 223 survey variables that describe each neighbourhood profile (or cluster) of the Italian population. The neighbourhood classifications are derived from a clustering algorithm (K-means) that splits all Italian households into one of 47 types. Each of these 47 types is then aggregated into one of 12 neighbourhood profiles (Experian 2009). The 12 groups have taglines such as *Low status appartments*, *Wealthy Elite* and *Elderly Households* that describe each profile.

With each classification is provided a wealth of statistics to characterise the group in more detail. For example, the *Wealthy Elite* neighbourhood has an above average proportion of professional workers, with very high degree attainment compared to the national average. A common method of expressing variation in geodemographic variables is by using an index score. This number provides a relative scale of a variable in comparison to the mean national average (which is expressed as 100). The calculation for the index score is described in more detail in section 3.5. For example, in the Mosaic Italy dataset, the geodemographic cluster *Low status apartments* has a very low literacy index score (<50) and a very high unemployment index score (>200).

The basis of geodemographic databases is the commonality of neighbourhood trends. By using this approach, *Elderly Household* areas in Rome could be considered close enough statistically to merit the same classification as similar neighbourhoods in any other area of the country. It is on this basis that neighbourhood classifications have been so popular in marketing. It should be noted that the 12 groups breakdown further to the more detailed 47 sub-groups mentioned previously. These descriptions are discussed at more length in the results section.

Mosaic Italy 2007, like its UK counterpart, is compiled largely off the last available census survey in the country (ISTAT 2001) as well as telemarketing data with each census output region containing approximately 60 households. Drawing from the literature on DRR, each of the 223 variables is assessed according to its discriminatory ability to define a household's vulnerability to evacuation, access to resources, financial recovery and physical risk of collapse. A range of social statistical methods has been incorporated to analyse each of the variables. This includes Gini-coefficients, Pearson's correlation coefficients and the index range of variables (Leventhal 1995). These factors are then weighted and combined with geophysical risk models used to characterise a Sub-Plinian eruption of Vesuvius and formulate a vulnerability index for the surrounding 'at risk' census areas.

ECUADOR

In the second phase of research, and to create an Ecuadorian social vulnerability model it was necessary to start at first principles and construct an Ecuadorian geodemographic. Unlike Italy, Ecuador does not have a Mosaic classification or in fact, any similar neighbourhood classification system. To analyse Ecuadorian social vulnerability using the same methodology as Italy, it was necessary to construct a classification system using downloaded Ecuadorian census data.

Only by applying the same statistical and methodological approach used on Italy could there be a fair comparison made in the application of neighbourhood classification systems to volcanic social vulnerability. At the time of writing, there is currently no commercial geodemographic classification system for Ecuador.

The last Ecuadorian census took place between November and December 2010 (INEC 2011). However, small area statistics from the 2010 census are currently unavailable for download and the purposes of this research. Therefore, during the course of this thesis it was necessary to use data from the 2001 census. Clearly a more recent census would have been preferable as it gives a more accurate portrait of the contemporary Ecuadorian population. The 2001 census data for

Ecuador is freely available online and can be downloaded from the Ecuador Institute of National Statistics and Census (INEC) at their website, www.inec.gov.ec (INEC 2011).

The Ecuadorian census of population and housing (Censo de poblacion y vivienda) is managed by the INEC in Quito. The census is conducted every ten years and although preliminary data and findings are shared in summary on the website and in downloadable tables, data is not made publically available at a household level for several years.

3.5   PHASE 1: ASSESSMENT OF A COMMERCIAL GEODEMOGRAPHIC – MOSAIC ITALY

This section details the first phase of the methodology and procedures undertaken to assess the application of a commercial geodemographic to vulnerability assessment. Details are provided as to how Mosaic Italy 2007 was used to compile a social vulnerability index, as well as the approach and rationale to the discriminatory statistical testing and the weighting of variables. Results are presented in the proceeding chapter of this thesis.

Statistical techniques are discussed and highlighted that have direct application in the assessment of geodemographic variables. Statistical measures such as the index range, gini-co-efficient and Pearson's correlation of variables were applied to the Italy study region to define the overall discriminatory ability of variables in describing social vulnerability. Using the outcome of these tests and measures, appropriate weighting is then assigned to variables as a social vulnerability index is produced. Further discussion around the results and methodology presented here can then be found in later chapters.

Given that the principal output from neighbourhood classification systems is an index value (weighted comparison to the overall variable average) in defining each census/survey variable, it is important to discuss the quantitative methods available to researchers in analysing these widely used indices.

Given the widespread use and availability of *propensity indices,* it was a key concern of this methodology that the construction of a vulnerability index should be based on suitable techniques for assessing these metrics. While there is no standard practice or formula for the creation of a vulnerability metric(s), this research was keen to adopt a multivariate approach as well as using the DRR literature to inform on social vulnerability indicators.

Therefore, the first step in creating a social vulnerability index for the Mosaic Italy data was define which variables would be considered as providing a discriminatory indicator in the assessment of vulnerable populations to a volcanic disaster. This process required close consultation with disaster experience as highlighted in the literature and during the development of the conceptual model presented in Figure 3-1. On this basis, variables focused on three different social components of DRR were identified as well as variables associated with physical/buildings vulnerability:

- *Access to resources* - Indicators relating to the political and physical access/capability of a household during/post a disaster
- *Financial Recovery* - Economic indicators that are correlated with disaster experience and recovery
- *Evacuation* - Variables that could help assess the likely difficulty of evacuation during an eruption
- *Physical/Buildings* - Housing information that could provide insight into the buildings likely structural performance given volcanic hazards (tephra, PDC, volcanogenic earthquakes)

Based on these core DRR factors, Table 3-1 in this section shows 24 variables from the Mosaic Italy dataset that were identified as being contributory factors in the assessment on these social and physical risk aspects.

After identification of these variables, the next step regarded the use of quantitative methods of assessment for these discriminating neighbourhood classification system indicators. Likewise, this section proposes a formula for the creation of specific vulnerability indices to define

population groups in hazardous volcanic areas. Aside from the social risks of a disaster, physical risk maps are provided to help contextualize the likely impact of a Sub-Plinian eruption and the spatial correlation of the likely physical and social risks in the Campania province.

3.5.1 INDEX CALCULATION

In social statistics, a common method of measuring a classification system's ability to discriminate a population is by using an index. The index can be defined as a quantitative measure of a variable value against the given variable average. As seen in equation 3) below, a propensity index value is calculated by a dividing the variable result ($Variable_X$) by the given variable average $(Variable_{\bar{x}})$ and then multiplying this by 100.

$$\frac{Variable_X}{Variable_{\bar{x}}} * 100 = Index\ value_X \qquad 3)$$

If a result is exactly the same as the average, an index value of 100 will occur. However, if a variable result was half the average, it would have an index of 50. Similarly, an index of 200 would indicate that a variable was twice the average. To assess the relative merits and discriminatory ability of a classification system, propensity index tables were produced for both study areas (Ecuador and Italy).

3.5.2 INDEX RANGE

The index range of a variable is a good measure of the spread of data among a population. It gives an indication of how varied survey variables are between clusters. This was used to gain an understanding of how well defined cluster neighbourhoods were in both upper and lower tier cluster groups. The larger the index range, the more discriminating the variable (Leventhal 1995).

$$Index\ Range_X = Index_{Max\ X} - Index_{Min\ X} \qquad 4)$$

Equation 4) demonstrates how the index range is calculated, where x is the chosen variable, the range is measured by deducting the minimum index value from the maximum index value. Figure 3-3 shows the comparison of all 223 variables for both levels of clustering. The separation in the two lines show that there is greater variation in the 47 profiles than there is with just 12. This would be expected and is confirmation that increasing the number of cluster groups provides a greater amount of heterogeneity and variability within the dataset.
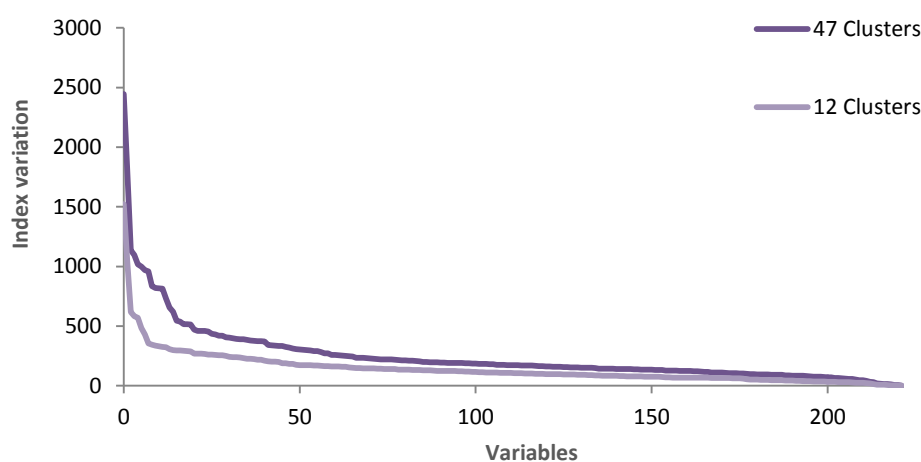
FIGURE 3-3 - INDEX RANGE VARIATIONS (MOSAIC ITALY 2007)

3.5.3   GINI CO-EFFICIENT AND LORENZ CURVES

Lorenz curves are a cumulative distribution function more commonly associated with macroeconomics. Developed by Max O. Lorenz in 1905 they were originally designed to show inequality and wealth distribution among populations (Gastwirth 1972). They can also provide a graphical representation of geodemographic discrimination as they highlight how variable data is skewed amongst a cumulative population. Equation 5) shows how the *generalised* Gini coefficient can be calculated for a population with values yi, i=1 to n, that are indexed in non-decreasing order. The *generalised* method outlined in Bellù and Liberati's (2006) work for the FAO formed the basis of the calculation of the Gini Indices in this research.

$$G = \frac{1}{n} \left(n + 1 - 2 \frac{\sum_{i=1}^{n}(n + 1 - i)yi}{\sum_{i=1}^{n} yi} \right) \qquad \text{5)}$$

(Bellù and Liberati 2006)

With application to the Mosaic Italy data, discriminatory differences can be brought out by analysing population distributions for each variable using this approach. Aside from direct calculation of the index, the Lorenz curve provides a graphical representation of the inequality of the variable within a population. The coefficient can also be calculated directly calculated from the graph. Figure 3-4 shows the skew within the population distribution for several variables (including *% black ethnicity, % asian ethnicity, % households with access to drinking water, % illiteracy*) in the census areas around the 50km radius of Mount Vesuvius, Italy. Essentially, Figure 3-4 reveals the relative differences in the socio-economic status of population groups across the study region.
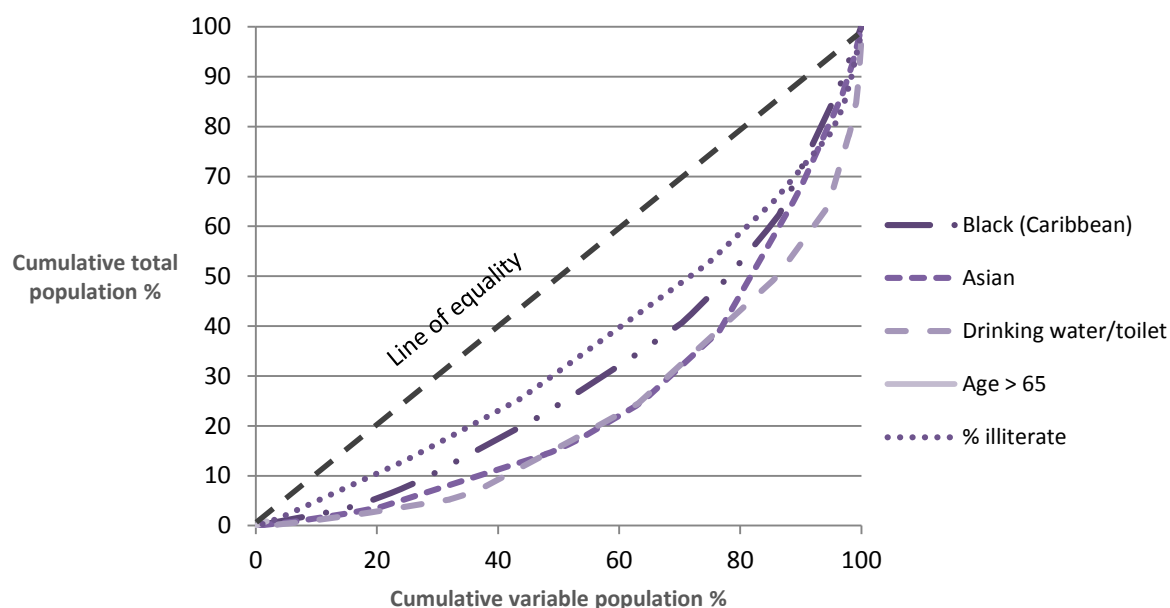
FIGURE 3-4 - LORENZ CURVES, MOSAIC ITALY INDEX VARIABLES, POPULATION (MOSAIC 2007)

The area between the hypothesised 'line of equality' and the 'actual' cumulative distribution observed is also known as the Gini-coefficient. This area provides a quantitative measurement of discrimination within a population and can be calculated from either direct measurement off a Lorenz curve or from mathematical calculation of the cumulative distribution of the variable. Gini-coefficients were calculated for 24 variables identified in the Mosaic Italy dataset as being the most pertinent indicators to DRR. The calculated values can only range from 0-1 and are independent of whether the final value is positive/negative. If the coefficient is closer to 0 than 1, the more equally distributed the variable is determined to be. If the figure is closer to 1, there is a more unequal distribution of a variable. Table 3-1 shows that the most unequal distribution of observations was found within the following variables; *Divorce, Buildings with 3-10 flats* and *Houses without drinkable water/ or a toilet*. These results show there are neighbourhoods where these factors are far more prevalent than others.

Variables	Gini Coefficient (47)
Divorced	0.30
Age <5	0.16
Age >65	0.22
Daily movement (inside Commune)	0.14
Buildings with 3-10 floors	0.29
Buildings with more than 10 floors	0.59
Population Density	0.45
Illiterate	0.49
Unemployed workforce	0.31
Retired	0.19
Rented house	0.34
Loan (2006)	0.10
% Without reinforced concrete	0.31
Buildings Built before 1919	0.45
Buildings Built between 1919-1945	0.33
Buildings with >4 floors	0.53
Multiple flat apartments	0.14
Ethnicity (Africa)	0.45
Ethnicity (Asia)	0.54
Origin Pakistan	0.32
Origin Black African	0.24
Origin Bangladesh	0.32
Origin Black Caribbean	0.31
House without water or toilet	0.72

3.5.4   CORRELATING VARIABLES

In analysing the geodemographic classifications for their use in a vulnerability index, it is necessary to test variables for their correlation and thus reduce data redundancy within a risk category. This was undertaken using a statistical software package by comparing the covariance of two variables divided by the product of their standard deviations. This calculation is known as the Pearson's product-moment coefficient. Correlation coefficients vary between -1 and 1. The closer a value is to 1 or -1, the greater the linear correlation between the variables. Table 3-2 shows the correlation between variables that would increase a household's vulnerability during disaster evacuation. Several of the initial variables in the Mosaic Italy dataset were taken out of the social vulnerability index because their correlation was very high. This would have resulted in data redundancy and effectively over representing an aspect of vulnerability within

the model. Variables with particularly high correlation in the Mosaic Italy dataset are highlighted below:

- Population density - Household density (0.978)
- Divorced – Separated (0.967)
- Over 65 - Widowed (0.970)
- Household density – Buildings with more than 10 flats (0.847)

TABLE 3-2 - EVACUATION VARIABLE CORRELATIONS, MOSAIC ITALY 2007

Variables	%Separated	% Widowed	% Divorced	% Aged < 5	%Aged> 65	%Daily movement (inside commune)	%Buildings with 3-10 flats	%Buildings with more than 10 flats	People per Household	Household Density	Population Density
% Separated		0.26	0.967	-0.17	0.18	0.559	0.418	0.616	-0.67	0.579	0.466
% Widowed	0.26		0.375	-0.669	0.97	-0.36	-0.07	0.04	-0.752	0.2	0.1
% Divorced	0.967	0.375		-0.28	0.318	0.526	0.367	0.585	-0.753	0.568	0.435
% Aged < 5	-0.17	-0.669	0.28		-0.682	0.21	0.23	-0.13	0.485	-0.13	-0.06
% Aged > 65	0.18	0.97	0.318	-0.682		-0.346	-0.14	0.01	-0.745	0.16	0.05
% Daily movement (inside commune)	0.559	-0.36	0.526	0.21	-0.346		0.407	0.715	-0.03	0.601	0.593
% Buildings with 3-10 flats	0.418	-0.07	0.367	0.23	-0.14	0.407		0.13	-0.13	0.21	0.23
% Buildings with more than 10 flats	0.616	0.04	0.585	-0.13	0.01	0.715	0.13		-0.22	0.847	0.844
People per Household	-0.67	-0.752	0.753	0.485	-0.745	-0.03	-0.13	-0.22		-0.356	-0.19
Household Density	0.579	0.2	0.568	-0.13	0.16	0.601	0.21	0.847	-0.356		0.978
Population Density	0.466	0.1	0.435	-0.06	0.05	0.593	0.23	0.844	-0.19	0.978	

Decisions had to be made regarding which of the variables to remove from the index. It was necessary to compare each variable to see their relative inter-dependencies and correlations in a vulnerability subset. This resulted in the variables *Household density, % Separated,* and *% Widowed* being removed from the index. There was no requirement to have both *% Separated* and *% Divorced* variables as this was effectively data duplication with both variables showing very strong positive correlation.

Likewise, Household density was removed from the variable list because there was already a Population density variable. Table 3-3 shows the revised evacuation variables following re-classification.

TABLE 3-3 - REVISED EVACUATION VARIABLES, MOSAIC ITALY

Variables	% Divorced	%Aged < 5	%Age> 65	% Daily movement (inside commune)	%Buildings with 3-10 flats	%Buildings with more than 10 flats	Population Density
% Divorced		-0.28	0.318	0.526	0.367	0.585	0.435
% Aged < 5	0.28		-0.682	0.21	0.23	-0.13	-0.06
% Aged > 65	0.318	0.682		-0.346	-0.14	0.01	0.05
% Daily movement (inside commune)	0.526	0.21	-0.346		0.407	0.715	0.593
% Buildings with 3-10 flats	0.367	0.23	-0.14	0.407		0.13	0.23
% Buildings with more than 10 flats	0.585	-0.13	0.01	0.715	0.13		0.844
Population Density	0.435	-0.06	0.05	0.593	0.23	0.844	

3.5.5 CREATING A SOCIAL VULNERABILITY INDEX – MOSAIC ITALY
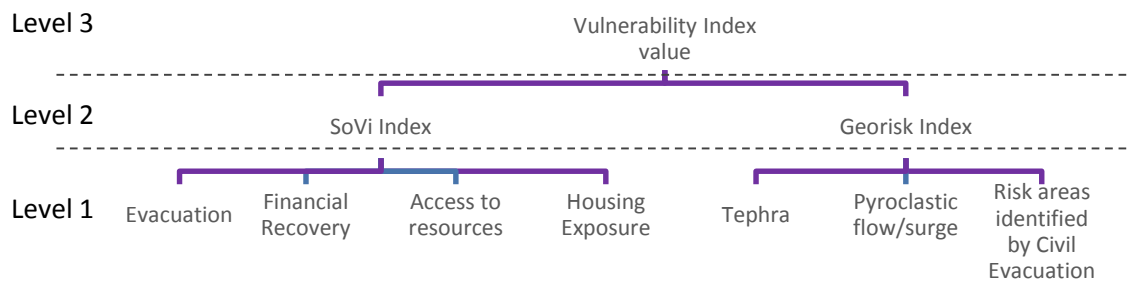
FIGURE 3-5 - PROPOSED VULNERABILITY MODEL USING MOSAIC ITALY DATA

Figure 3-5 highlights the data structure used to create a social vulnerability index and each component module. In applying this methodology to Italy, there are essentially three levels of vulnerability assigned to each census output region in the province around Vesuvius. The following narrative provides a summary of this hierarchy.

Level 1: These are the individual social vulnerability scores for each household for both social and physical risks; Evacuation, Financial recovery, Access to resources, Building exposure, Tephra fallout, Pyroclastic surges and Civil Evacuation (according to the 1995 DCP plans).

Level 2: These index scores are created as a composite of the respective social and physical risk scores. This creates a social vulnerability score and a Georisk Index for the physical risks.

Level 3: Overall vulnerability of place is calculated as an index from all physical and social variables in levels 1 and 2.

To factor in a level of weighting in this methodology, Gini-coefficients were used for each Mosaic variable. The main reason for this was to allow a level of discriminatory weighting. Therefore, those variables with Gini-coefficients closer to 0 were given less weighting in the overall vulnerability score, thus reducing their impact of the overall score.

Equations 6), 7) and 8) below describe how the index scores were calculated as a metric for each variable (x). These were then combined for all social factors to create the *Georisk Index*.

$$Weighted\ variable_X = (Gini_X) * (Mosaic\ Index_X) \qquad 6)$$

As shown in equation 6), the weighted variable is calculated by multiplying the Gini Co-efficient by the Mosaic index value for each variable.

$$Vulnerability\ score_Y = \sum Weighted\ variable_{Xn} \qquad 7)$$

Equation 7) shows that the vulnerability score for a given area of social vulnerability is then constructed from the sum of component weighted variables calculated in equation 6).

$$Total\ social\ vulnerability = \sum (Vulnerability\ Scores_{Yn}) \qquad 8)$$

The overall vulnerability score is then calculated as the sum of all aspects (e.g. *Access to resources, Financial recovery, Evacuation risk*) as shown in equation 8).

$$Social\ vulnerability\ index = \frac{Total\ vulnerability_X}{Total\ vulnerability_{\bar{X}}} * 100 \qquad 9)$$

To produce the final *social* vulnerability propensity index score it is necessary to compare each observation to the variable average as shown in equation 9).

However, to calculate an *overall* vulnerability index, it was necessary to also calculate risk ranks for the physical risk of the eruption. This included areas subject to Tephra, PDCs and the Civil Evacuation around the volcano based on previous risk assessment. For simplicity with regards to the index model, a numeric risk integer between 0-3 was assigned for each Census

area and for each hazard (where 3=very high risk and 0=very low/no risk). It should be noted that there is no numerical justification for the choice of physical risk weighting used here. The decision to have a weighting variation of 0-3 was based largely on the need to quantify the non-numeric variables such as Civil Evacuation zones, which are provided in categorical format (e.g. High risk area/Low risk area). Clearly, such choices of physical ranking remain a moot point and are discussed in more detail in section 5.1.4. These values were then multiplied by a factor of 10 and accumulated to provide an indicative *Georisk Index score*.

The final step was to calculate an overall vulnerability score on the basis of aggregating the social and physical risk scores as an index. Equation 10) shows the calculation for this metric.

$$Overall\ vulnerability\ index_X = \sum\ (Social\ index) + \sum (Physical\ index) \qquad \text{10)}$$

Indicative results and model outputs from *phase 1* are presented in the proceeding chapter of this thesis.

3.5.6   VOLCANIC HAZARD MODELLING USING GIS

In order to create a vulnerability model for the areas around Mount Vesuvius, Italy, it was necessary to make some modelling assumptions and hypothesise a likely eruption scenario. It was on the basis of these assumptions that risk areas could be identified around the volcanic edifice. GIS was then used to model the defining physical risk boundaries around the volcano and spatial analysis was used to assign demographic groups to these physical risk factors.

Volcanoes pose multiple geophysical hazards to an environment and Vesuvius is no exception. Past eruptions of Vesuvius have included several of the following hazards:

• tephra (ash fall)

• pyroclastic flows/surges (superheated ash)

• lava inundation

- lahars (mudflows)

- outputs of poisonous gases

- pyroclastic bombs

- generation of ocean tsunamis

- volcanogenic earthquakes.

Superficial deposits from previous volcanic eruptions can also be found in the surrounding geological and geomorphological strata of the Neopolitan area (Esposti Ongaro et al 2008). In general, past behaviour of the volcano has been characterised by short periods of high explosivity and longer periods of lower intensity eruptions (Cioni et al 2008). This can be defined by the Magma Discharge Rate (MDR) which has been steadily increasing for the last 3,000 years. Higher MDR rates are more characteristic of frequent but lower magnitude eruptions. Though all volcanic events are quite unique in their exact size and nature, Table 3-4 highlights how Vesuvius has exhibited several characteristic styles of eruptive behaviour and magnitude over the last 20,000 years.

TABLE 3-4 - ERUPTIONS OF MOUNT VESUVIUS (ADAPTED FROM CIONI ET AL 2008)

VEI	General characteristics	Eruption type	Eruption example	Date
3	Last explosive eruption in 1944. Eruptions are becoming more frequent but less explosive. MDR is increasing.	Open-conduit/ strombolian	1944 Eruption	1944 AD
4	Eruption columns 13-19km high, ashfall, earthquakes, lahars and pyroclastic flows.	Subplinian	1631 Eruption	1631 AD
4		Subplinian	Pollena	472 AD
5	Pyroclastic flows, ash columns >25km high. Infrequent, very large explosive events, sporadic and involved total column collapse.	Plinian	Pompeii	79 AD
4		Plinian	Pomici Verdoline	16,000 BP
5		Plinian	Pomici di Base	18,300 BP

In terms of assessing the geophysical risk to loss of life that an eruption of Vesuvius threatens, the scope of this research did not take all volcanic hazards into account. For example, lava inundation and volcanogenic earthquakes are a common phenomenon associated with a

volcanic eruption and frequently result in casualties due to house fires, building collapse, and hill slope failure. When categorising physical risk boundaries for the vulnerability model presented here, historical evidence suggests the most likely loss of life from Vesuvius would be due to pyroclastic density currents (PDC) and Tephra fall (Cioni et al 2008). The principal boundaries for the geophysical index of this study comprised of tephra loading maps (based on the *Tephra 2* software model) and a PDC map based on 3D column collapse (Esposti Ongaro et al 2008). Also included in this analysis was the use of civil evacuation maps for the areas around the volcano. This was the approximate basis for the size of the geodemographic population analysed around the Mount Vesuvius summit: a concentric area 50km from the volcanic vent.

After considering several historic eruptions of Vesuvius (Andronico 2002), it was determined the most appropriate scenario to assume for the vulnerability model was a large sub-Plinian (II) event. A magmatic eruption of this type would be in the order of 3-4 on the VEI scale. There were two main reasons for the choice of this eruption scenario. Civil protection measures (including evacuation plans) for the Neopolitan area are currently based upon the supposition of a sub-Plinian eruption, though this remains a contentious issue (Rolandi 2010). With evacuation being one of the key vulnerabilities in the geodemographic model it seemed prudent to keep scenario parameters consistent with current evacuation spatial extents. In doing so, more pragmatic comparisons can be made to existing evacuation procedures. Likewise, given the past activity of the volcano, including geomorphologic deposits (Andronico 2002), a large sub-Plinian eruption in the near future has a large probability. Although, it is important to note that eruptive history for the volcano suggests that more frequent, lower magnitude events are becoming more likely, this eruption type provides a realistic worst case scenario with which to model possible impacts.

Ash fall from a volcano can be devastating in terms of both loss of life as well as the destruction of rich agricultural areas and transportation networks. Although indirect consequences of tephra fall include flash floods of mud (lahars) and the potential hazards from the volcanic ash plume as it disrupts airplane routes, it is ash loading on houses that is regarded in this study as the greatest threat to households around Vesuvius. Using TEPHRA2 (Bonadonna et al 2005), a bespoke numerical modelling package that simulates the accumulation of sedimentation across a spatial area, isopach maps were created to quantify the distribution of tephra. TEPHRA2 is an advection-diffusion model that takes into account the grain size-dependent diffusion of ash fall in a stratified atmosphere as the volcanic plume rises and deposits erupted material. The user must define all input parameters such as particle sizes, eruption magnitude, wind characteristics and a geographical output grid. Data is then output in text file format containing discrete point locations with given accumulation volumes. Using spatial analysis in a GIS, these locations could then be interpolated into a surface feature using isopach maps. It should be noted, that TEPHRA2 is not the only available advection-diffusion model available and the analysis undertaken during this research could have made use of other freely available software packages; including ASHFALL (Hurst and Turner 1999), FALL3D (Costa et al 2006) or HAZMAP (Macedonio et al 1998). However, given that this aspect of the methodology was not the core concern of the thesis contribution, no further sensitivity testing of alternative tephra models was undertaken.

Input parameter	Quantity
Plume height (m)	27000
Eruption mass (Kg)	9.00E+12
Maximum grain size (phi units)	-5
Minimum grain size (phi units)	5
Median grain size (phi units)	1
Standard grain size (phi units)	1.5
Vent Elevation (m)	1281
Eddy constant (m2/s)	0.04
Diffusion coefficient (m2/s)	20
Fall time threshold (s)	288
Lithic density (Kg/m2)	2500
Pumice density (Kg/m2)	1000
Column steps	100
Plume ratio (of total plume height)	0.2

TABLE 3-5 – INPUT PARAMETERS, TEPHRA 2 MODELLING OF VESUVIUS ERUPTION

Input parameters used for the tephra dispersal model can be seen in Table 3-5. These were

based on the large sub-Plinian eruption as hypothesised by Macedonio et al (2008) for a likely

eruption of Vesuvius. One of the crucial parameters that any tephra model must be completely

transparent about is the subjective choice of wind direction. This input essentially dictates the

spatial orientation of the heaviest ash loading around the volcanic vent yet is also subject to

some debate. Due to most locations having a varied wind field, the requirement to pick one

direction for modelling ashfall can be understood to be a deterministic view rather than any

guarantee. Therefore, studying the wind field breakdown of the Vesuvius area for a given year,

the choice of prevailing wind on the day of the eruption was set to North North East (322°).

Over the course of a year, 18.6% of the prevailing wind emanated from this direction

(Bonadonna et al 2005) and forms the most likely direction within an given year (Macedonio et

al 2008). Once initial results were output, it was then necessary to establish those areas that

were most at risk of building collapse. Based on work by (Pareschi et al 1999), a threshold of

300-400 kg/m² was assumed to represent the demarcation of those areas at highest risk of

building collapse. After conversion of the Tephra 2 output file to a GIS format, areas within the

corresponding tephra accumulation zones were assigned an integer risk classification of

between 0-3. The integer classification corresponded with a modelled accumulation of 100, 200, 300 and 400 kg/m^2 respectively to provide a risk level. The corresponding map can be seen in Figure 4-6 (c), which highlights the tephra loading model.

PYROCLASTIC FLOW

Pyroclastic flows are one of the most deadly forces of nature and perhaps the most characteristic of large explosive volcanic eruptions. The Pompeii eruption of AD79 is still the most infamous pyroclastic event in history as thousands of Neapolitans died from asphyxiation and subsequent burial from the resulting debris flows. This was due to the superheated gravity flows of hot ash that swept down to coastal towns through systematic column collapse of Vesuvius. Unlike lava inundation, pyroclastic flows can travel at over 100km/hr and may reach proximal towns in a matter of minutes. The area designated as being at the highest risk from PDCs in this study was based on a transient 3D flow model by Esposti Ongaro et al (2008). This takes into account the topography of the land around Vesuvius to simulate total column collapse during a sub-Plinian eruption. Propagation maps of the PDCs 800 seconds after column collapse are the basis for the pyroclastic flow boundary digitised for use in this analysis. This can be seen in Figure 4-6 (b).

EVACUATION

The evacuation regions for the Georisk Index were based on the official Civil Protection plans for the area around Vesuvius (DPC 2005). This included the Blue, Red and Yellow zones that corresponded to a given level of risk. Red is deemed the highest risk area and evacuation from this region is of priority in the event of an imminent eruption. The Blue zone is the next highest risk area and the Yellow zone the area of likely tephra fall around the volcano. The boundaries of the evacuation zones are highlighted in Figure 4-6 (c).

3.6   PHASE 2: CREATING A NEIGHBOURHOOD CLASSIFICATION SYSTEM - ECUADOR

Phase 2 presents the various techniques and methods employed in creating the Ecuadorian neighbourhood classification system from downloaded census data. Whereas phase one made use of Mosaic Italy, a commercially available geodemographic system, no such database or commercial product currently exists for Ecuador. Therefore, to test the research hypothesis of this thesis in a different geographical setting, as well as document the various factors that influence a classification system, this section describes how census data was prepared, analysed, aggregated and clustered to create the Ecuadorian classification system. Likewise, discussion and justification regarding the various variable choices made in creating the classification are considered as well as the methods employed in data reduction.

The Ecuadorian census is undertaken at a *Census Area* level. Each *Census Area* has an individual *Census Code,* of which there are 3926 designated areas. The average population of a census area in the 2001 data is 3147, although the individual sizes vary greatly. The largest census area consists of 29,275 people whereas the smallest is made up of just 2 individuals. It was not possible during the course of this research to obtain delineated GIS digital mapping boundaries for all the census zones across Ecuador. Therefore, results are not displayed in a GIS at this level of geography across all territories. During the course of field work, the author visited the headquarters of the INEC in Quito (September 2010), with the department reiterating their stance that the digital census boundary data is not made publically available nor could it be made available for external research purposes. Therefore, in the current absence of the census boundary data, results shown graphically at a national scale have been calculated at census level and then re-aggregated to the next level of geography, the regional *Parroquia* (translated as Parish in English). The exception to this rule is in the metropolitan district of Quito, where census boundary details were obtained from a hard copy map courtesy of census

data research by the Andean University of Simon Bolivar (UASB 2012). This latter research displayed results for the Quito Metropolitan District. These areas have been subsequently digitised from the original maps using a GIS software package and cross-referenced to Quito census zones, allowing detailed analysis for the study region.

As mentioned, and shown in Table 3-6, the next level of geography recorded in the census is the local *Parroquia*. There are 995 Parishes in mainland Ecuador. Parishes have a mean average of 12,218 people but due to the remoteness of some areas in the amazon, there is a large amount of variation in population size. It should be noted that digital Parish boundary data is made available online in a GIS (ESRI shapefile) format from the Ecuador Institute of National Statistics and Census website (INEC 2011).
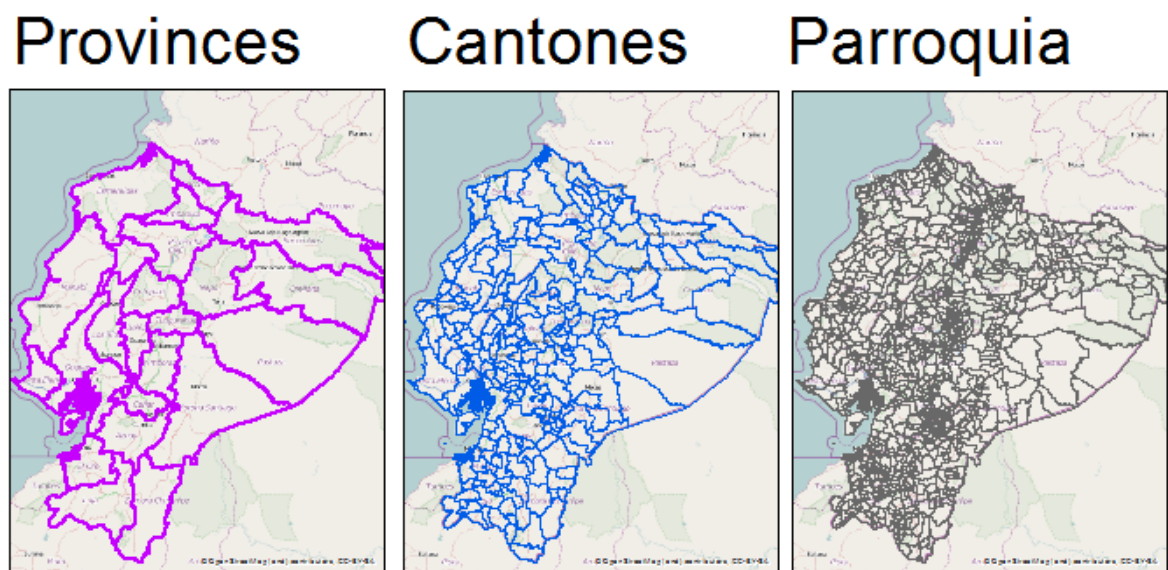
FIGURE 3-6 - MAPS OF THE VARIOUS ECUADORIAN ADMINISTRATIVE BOUNDARIES AT A COUNTRY-WIDE SCALE (INEC 2011)

Parroquia are grouped into *Cantons* (referred to as townships). These 226 *Cantons* converge to form the 24 *Provinces* of Ecuador. Table 3-6 shows the hierarchy of these geographies and Figure 3-6 highlights the nested spatial relationship of these administrative boundaries.

TABLE 3-6 - GEOGRAPHIC ADMINISTRATIVE LEVELS IN ECUADOR (INEC 2013)

Geographic level	Number of administrative divisions
Census	3926
Parroquia (Parish)	995
Canton	226
Province	24

The highest level of geography recorded in the census is that of *Province*. Although not officially defined, *Provinces* can be loosely recognised as belonging to one of three geographical territories in Ecuador; the *Costa* (coast), the *Sierra* (mountains) or the *Amazonas* (amazon basin). The topographical divide of the Andes mountain range provides much of the basis for the historic ethnographic variation within Ecuador. This theme is explored further in the proceeding results and discussion chapters.

3.6.1 REFORMATTING CENSUS DATA

In its original state, the Ecuadorian census is a mixture of text, coded categorical data and numerical information covering seven sections. The questions seen in Table 3-7 cover categories including geographic location, dwelling information, household construction and amenities, emigration details, household composition, demographics of the household and the marital status of the population.

TABLE 3-7 - A SUMMARY TABLE OF THE POPULATION AND HOUSING CENSUS, 2001, ECUADOR (INEC 2011)

Census section	Section variables
I. Geographic Location	• Information on the Province, Canton, Parish and Census zone
II. Information about the Dwelling	• Private (house, flat, shack) or collective (residence hall, hostal, hospital)
III. Information about the household.	• House occupied/unoccupied/under construction. • Construction materials (roof, walls, floors, frame) • Water services to the dwelling • Sewage service, Phone connection, Garbage removal • Number of bedrooms in the house • Number of household occupants (male/female) • Housing tenure (mortgage, rent, owned, state-owned)
IV. Information about emigrants to other countries.	• Has anyone in the household emigrated (age, gender, location)
V. Identification of the persons in the household.	• Relationship of people in the household (spouse, head of household, son, daughter, grandparent)
VI. Population data	Demographics of household members: • Age • Sex • Ethnicity • Language(s) spoken • Literacy • Educational attainment • Disability • Occupation
VII. Civil or Married status	Married, Divorced, Separated, Co-habiting, Single

The *information about the household* section includes questions about the type of housing, including whether the property is owned, rented, mortgaged or a state-owned property. This section also regards the physical attributes of the home, and whether it has fundamental amenities such as a shower, telephone and basic levels of sanitation. The *emigration* section asks only two questions, which are concerned with whether any member(s) of the family has emigrated from Ecuador prior to the census. *Household composition* is provided by focused questions on which family member(s) still live in the household, as well as their relative age, gender and status. Also shown in Table 3-7 is that the census includes information on the individual's ethnicity, employment status, occupation, and questions about their health.

To be able to accurately compute and aggregate data for the purposes of clustering, it was necessary to make sure all data was formatted and standardised into areal counts. Many variables, such as language spoken, ethnicity and disability are multiple-choice questions in the

Ecuadorian census, and therefore, there is a unique code referring to different categorical answers. This meant that a single question regarding ethnicity changed from being a single variable into six categories to reflect the multiple answers regarding ethnic background. This process is illustrated in Figure 3-7 in a screenshot taken from the SPSS software program. In this instance, category counts can be shown to be created for Mestizo (*mixed race indigenous and white*), Mulato (*mixed race white and black*), Negro (*black*), Blanco (*white*), Otro (*other*), and Indigena (*indigenous*) populations.

A notable problem with increasing the number of columns is that the file size of the census data quickly becomes very large and difficult to process. This wouldn't normally be an issue with datasets containing thousands of rows, but the statistical software package used in this work (SPSS) struggled to open, edit and compute data exceeding a million rows. In the Ecuadorian population census data, there are 12.6 million rows and thus computation became extremely slow and difficult while retaining data at anonymous individual level.
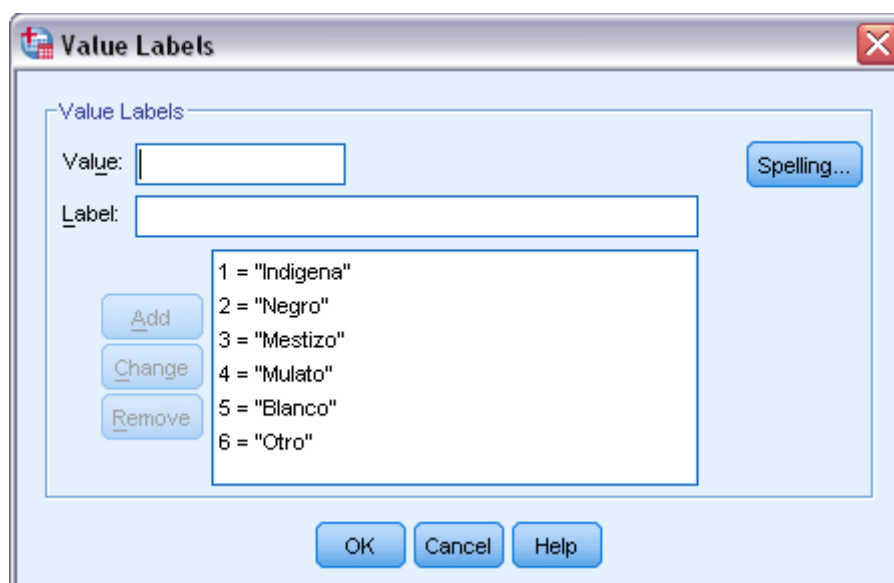
FIGURE 3-7 - CATEGORICAL CODES FOR ETHNICITY (SPSS V.16)

Because of the processing speed of the entire dataset, it was determined that it would be prudent to split the file into category types for further processing. In doing so, it became

necessary to decide which census variables would not be needed for further analysis or inclusion in the neighbourhood classification model. To guide this selection process, the following principles were followed.

1.   **Removing redundant variables** - A geodemographic profile is a rounded description/portrait of a neighbourhood, where certain site-specific variables (e.g. an individual's village, town or city of birth) are too complex to be appropriately modelled or suitably aggregated into meaningful segments. There are simply too many possible outcomes for this type of question and the amount of variables included would dominate the model.

2.   **Common geodemographic variables** - Neighbourhood classification development for Ecuador was focused on closely aligning the methodology of *phase 2* with those used to define the profiles in *phase 1 (Mosaic Italy)*. For this reason, and so that meaningful comparisons could be made to existing geodemographic solutions, the same principles of variable selection were applied to the Ecuadorian data.

TABLE 3-8 - FINAL 37 VARIABLES USED FOR THE ECUADORIAN CLASSIFICATION

Type of variable	Variables	Type of variable	Variables
Tenure	Owned (Propia)	Age	0-4
	Leased (Arrendada)		
	Mortgage (En anticresis)		5-14
	Public housing (Gratuita)		15-25
Marital status	Co-habiting couples(a)		24-45
	Single		>45
	Married	Disability	Disabled/LLTI
	Divorced	Educational attainment	Degree
	Widower		Literate
	Separated	Ethnicity	Indigena (e.g. Inca desc)
Employment	Administrative		Negro (Black)
	Agricultural/Fishing		Mestizo (Mixed race - Native)
	Artisans		Mulato (Mixed race - Black)
	Civil Service & Social work		Blanco (White)
	Defense		Other ethnic origin
	Managerial/Financial	Household composition	Single person HH
	Manufacturing		Two person HH
	Mining / Construction		HH of 3 to 5 people
	Skilled Professional		HH of greater than 6

To help advice on the selection process, Harris' book on geodemographics (Harris et al 2005) was closely consulted as well as open source initiatives such as the white papers produced by Dan Vicker's for the UK Output Area Classification (2005) and available material on both Acorn and Mosaic classifications. Vicker's (2005) used 41 variables from the UK census data as the input for a two-tiered classification. These were divided into *demographic, socio-economic, housing* and *household composition* variables. A similar approach was undertaken for Ecuador by defining 37 final variables for use in the classification system (Table 3-8). However, it's relevant to note that initially, there were 45 census variables intended for the classification, with many being subsequently removed.

For example, *child mortality* is not typically recorded in census based classification systems and therefore, to maintain continuity, it was regarded as unnecessary for inclusion here. Of course, it is very important to note that many of these variables could have significant impact upon the Ecuadorian geodemographic presented here, and could certainly be argued to be social vulnerability indicators in disaster risk analysis. This point is explored further in proceeding discussion chapters.

The next step in transforming the data was focused on aggregation. As mentioned, the lowest level of Ecuadorian 2001 census data is at person level. However, such a level is far too detailed in granularity for the purposes of the model as geodemographic classifications are built around the concept that an entire neighbourhood can be described rather than the individual. Conversely, if information is aggregated to a level of geography too large, data variance is lost as areas become too homogenous (Harris et al, 2005). For this reason it was proposed that the census area zones were used as the appropriate level for aggregation and clustering. Given that other open source geodemographics (e.g. UK Output Area Classification) have been successfully produced at Census area, and this geographic level was the most detailed available for Ecuador, it was decided to aggregate and cluster the data to this level.

There is no unique ID provided to identify a particular census area in the INEC data so it was therefore necessary to create an ID by concatenating the defining location columns such as the *Province, Canton, Parroquia and Census sector* numbers. (E.g. Province 001 + Canton 10150 + Parroquia 001+ Census sector 006 = ID 00110150001006)

The next step was to then aggregate the census data by these newly created IDs. This reduced the data from 12.6 million records to 3296 rows.

After the data had been aggregated into counts for every variable for a given census area, it was then necessary to convert these counts into percentages. Figure 3-8 demonstrates how this conversion process took place. The percentages for each variable were based on the population count for that area. However, for some variables (such as *age* or *employment*) a substantial

amount of pre-processing was involved as new categories needed to be created and then

converted into percentages for an area.

Census ID	0-4	5-14	15-24	25-44	>45
Count	34	47	145	478	104

Row	Age
100	3
101	40
102	59

Percentage	4%	6%	18%	59%
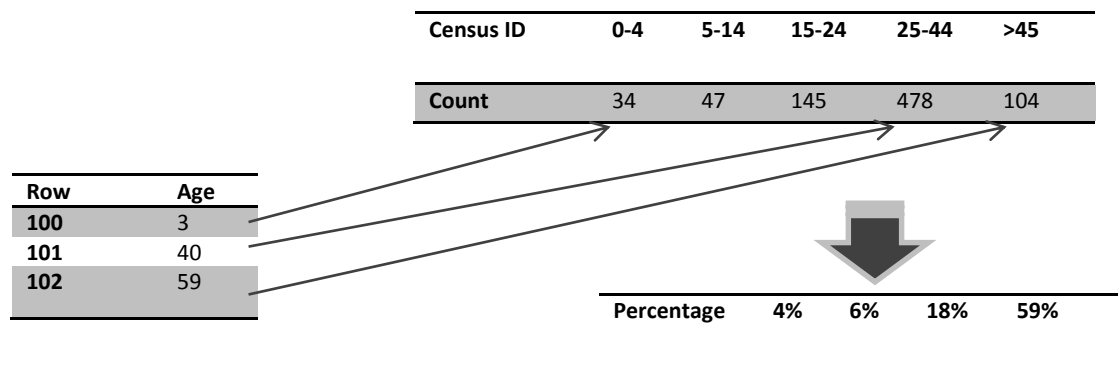
FIGURE 3-8 - CREATING NEW CENSUS CATEGORIES AND CONVERTING COUNTS TO PERCENTAGES

3.6.2   CORRELATION

The next step in the methodology was to examine the correlation matrix of the census variables.

It was important to establish the degree to which input variables were correlated (both

positively and negatively) as very high correlation could indicate a lack of heterogeneity in the

data. Similarly, the duplication of variables with very high correlation may act to systematically

over or under-weight the bias when clustering.

Table 3-9 provides a subset of the results from this analysis, and highlighting those variables

that had correlation either above 0.6 or below -0.6, where values closest to 1 and-1 are perfectly

correlated and those closer to 0 show little/no correlation.

The most highly correlated variables were *degree attainment* and people in a *skilled*

professional occupation with a correlation of 0.96. Similarly, strong correlation was also shown

between *degree attainment* and *managerial employment* (0.86) as well as people who had been

divorced (0.74). Degree attainment is a direct indicator of higher educational achievement,

which in turn would be more likely to warrant an individual going into skilled or managerial

work. The link between *divorce* and *degree attainment* is more surprising given that in countries such as the US, these variables are negatively correlated (Guardian 2009). It is not exactly clear what would cause this relationship in Ecuador but perhaps important factors indirectly linked to higher education, such as social mobility, female empowerment, or fiscal capability are involved here. For example, the high cost of divorce or existing cultural traditions may mean that it is more commonly the reserve of those that can afford it.

Strong negative correlation (-0.93) was shown between several variables including *Married* and *Co-habiting couples*. This phenomenon makes sense given that the two variables are mutually exclusive categories describing the status of a couple in a household. A similar trend was noted for *literacy* and the age grouping of *0-4*, which was an expected correlation (-0.7).

TABLE 3-9 - CORRELATION MATRIX (SHOWING ONLY STRONGLY CORRELATED VARIABLES) - ECUADORIAN CENSUS DATA

Variables	Population Density	0-4	5-14	24-45	Degree	Literate	Indigena	Blanco	Owned (Propia)	Leased (Arrendada)	Co-habiting couples	Divorced	Admin	Agriculture/ Fishing	Managerial/ Financial
24-45	0.4	-0.47	-0.78												
Literate	0.39	-0.7	-0.47	0.56	0.56										
Mestizo	0.02	-0.21	-0.07	0.07	-0.04	0.32	-0.72								
Blanco	0.31	-0.4	-0.4	0.47	0.64	0.47	-0.24								
Leased (Arrendada)	0.6	-0.36	-0.46	0.5	0.55	0.54	-0.22	0.43	-0.83						
Married	0.18	-0.11	0.11	-0.1	0.21	0.16	0.19	-0.11	-0.06	0.11	-0.93				
Divorced	0.54	-0.49	-0.44	0.47	0.74	0.57	-0.17	0.51	-0.54	0.61	-0.27				
Administrative	0.61	-0.7	-0.58	0.6	0.72	0.61	-0.2	0.51	-0.59	0.68	-0.2	0.69			
Agricultural/Fishing	-0.5	0.41	0.54	-0.63	-0.51	-0.58	0.33	-0.51	0.54	-0.65	-0.01	-0.51	-0.69		
Civil Service and Social work	0.44	-0.38	-0.45	0.51	0.44	0.51	-0.17	0.4	-0.44	0.54	-0.16	0.51	0.59	-0.63	
Managerial/Financial	0.45	-0.39	-0.42	0.41	0.86	0.47	-0.14	0.6	-0.38	0.43	-0.2	0.66	0.64	-0.42	
Skilled Professional	0.5	-0.49	-0.42	0.47	0.96	0.58	-0.16	0.58	-0.48	0.6	-0.29	0.74	0.75	-0.54	0.78

Negatively correlated
< -0.61

Positive correlation

> 0.61

Following correlation analysis and a review of the variables, it was determined that the 37 variables shown in Table 3-8 were significantly autonomous and meaningful to be used in the clustering. Most variables were largely uncorrelated, with an overall average of 0.01. Despite high positive and negative correlation in several variables described previously, the value of including these to help define the neighbourhoods outweighed the perceived risk of removing them from the clustering data.

3.6.3   PRINCIPAL COMPONENT ANALYSIS

The key aims of introducing Principal Component Analysis (PCA) into this research were two-fold; it was firstly intended that PCA could help identify and understand the underlying variance in the census information; and secondary, it was thought that PCA might help identify whether certain variables or groups of variables should be removed prior to cluster analysis. PCA provides an effective method for exploratory statistics and data reduction: it has also been used previously in the construction of geodemographic classification systems, including the creation of the 1981 UK SuperProfiles (Charlton et al 1985). Charlton, Openshaw and Wymer (Charlton et al 1985) used PCA to help inform on the relative influence of 55 cluster variables following an initial assessment of 465 census indicators. Similarly, understanding the controlling 'Principal components' of the data would be key to understanding the basis of the profiles created in clustering the Ecuadorian census data. The inter-dependencies and inherent variance within the Ecuadorian census data would help account for much of the heterogeneity between the neighbourhoods created.

After importing the standardised census data, the first step in PCA is to produce a table of communalities, as seen in Table 3-10. The extraction values inform the user of the most

prominent variables in a given dataset. (I.e. Variables that have the most influence in the dataset have higher extraction values).

In Table 3-10, among the most influential variables in the Ecuadorian census data were *degree attainment*, ethnicity types such as *Mestizo/Indigena* and whether the individual worked in a *skilled profession*. Perhaps unsurprisingly, these factors greatly influence the nature of the data much more than other factors. It could be argued that these variables are intrinsically linked to both financial achievement and cultural heritage which in turn drive the demographic trends seen in the data. Similarly, there were variables that appeared much less influential in the dataset, including *Artisan* employment, and age ranges such as *15-25*, as well as less dominant ethnicity types (e.g. *Mulato*).

TABLE 3-10 - EXTRACTION VALUES (10 HIGHEST VARIABLES) - INITIAL PCA ANALYSIS

Communalities	Extraction
Degree	0.911
Co-habiting couples	0.906
Mestizo	0.890
Skilled Professional	0.876
Married	0.872
Owned (Propia)	0.866
Indigena	0.837
Agricultural/Fishing	0.821
Leased (Arrendada)	0.812
Managerial/Financial	0.792

The PCA analysis helped identify the number of components that are driving the underlining trends in the dataset. Thus, in Table 3-11 we are able to determine that 11 components account for 70% of the cumulative variance in the Ecuadorian census data. Furthermore, the first component accounts for 26% of the overall variability. This makes *component 1* the most significant or principal factor.

TABLE 3-11 - PRINCIPAL COMPONENTS ANALYSIS COMPONENT LOADINGS

PCA Analysis	Before rotation			Varimax	Quartimax	Promax	Oblimin	Equamax
Component	Initial Eigenvalues			Rotation Sums of Squared Loadings				
	Total	% of Variance	Cumulative %	Total	Total	Total	Total	Total
1	10.733	25.555	25.555	9.539	10.429	10.434	4.613	4.462
2	4.452	10.601	36.156	3.465	3.557	3.873	3.292	3.541
3	2.594	6.176	42.332	2.678	2.669	6.171	2.572	3.083
4	2.406	5.730	48.062	2.678	2.174	4.144	3.139	2.930
5	1.892	4.504	52.566	2.378	2.067	3.165	3.291	2.836
6	1.568	3.734	56.299	1.908	1.830	3.245	3.879	2.809
7	1.468	3.495	59.795	1.558	1.540	1.930	4.332	2.404
8	1.274	3.033	62.828	1.540	1.527	1.968	4.732	2.375
9	1.061	2.525	65.353	1.342	1.320	2.340	4.426	1.885
10	1.032	2.457	67.810	1.274	1.264	1.659	3.907	1.800
11	1.012	2.409	70.219	1.133	1.114	1.840	2.985	1.368
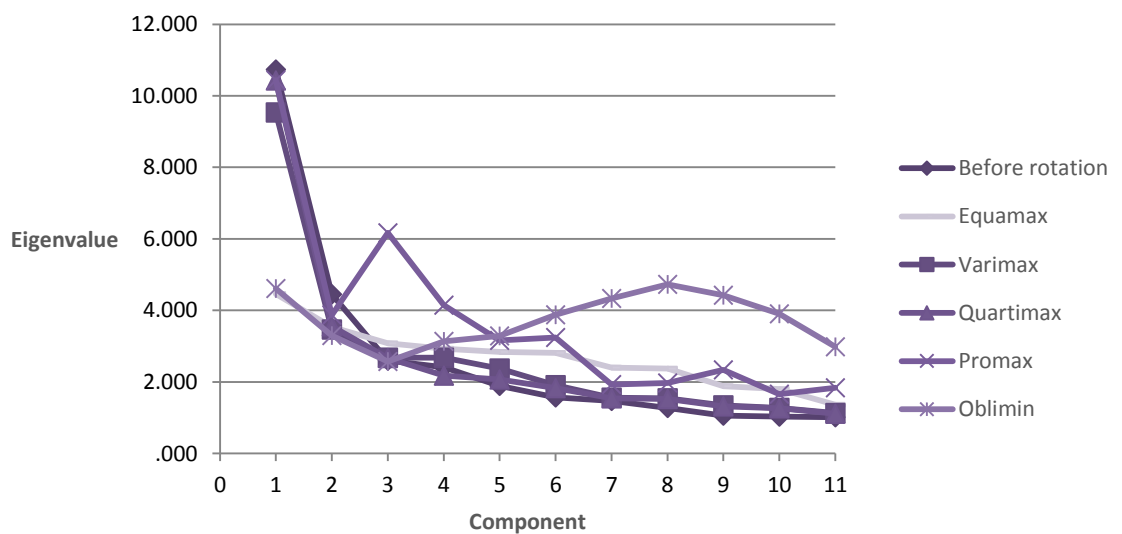
FIGURE 3-9 - SCREE PLOT OF COMPONENT LOADINGS BEFORE AND AFTER ORTHOGONAL ROTATION (PCA ANALYSIS OF ECUADORIAN CENSUS VARIABLES)

Following the initial orthogonal PCA analysis, it was decided to experiment with rotating the

components to see how this may affect the weighting factors of the Eigen values. As can be

seen in Figure 3-9, the unrotated loading on the first component is significantly higher than in the rest of the components, which decrease rapidly thereafter. This highlighted that a considerable amount of the variance could be explained using only the first component. To help reduce this loading, and thus, in seeking to define a more equal distribution of loadings, an alternative picture of variance can be gained by rotating the components in multi-dimensional space. The purpose of rotating principal components (also referred to as eigenvectors) is performed to obtain a new set of factor loadings from a given data set (MacDonald 1985). In doing so, the analyst can define 'simple structure', and thus provide a clearer perspective on the loadings (Bryant and Yarnold 1995). To do this, there are two groups of rotation that can be performed. *Orthogonal* rotation methods assume that the input variables are uncorrelated, whereas *Oblique* rotation assumes the variables are highly correlated. To further explore the Ecuadorian data, both orthogonal rotations (e.g. *varimax, quartimax,* and *equamax*) were performed as well as an oblique rotation (*oblimin*). It is important to note that many of the variables in the Ecuadorian dataset are highly correlated, and therefore less suited to orthogonal exploratory data analysis. This is also the reason that 70% of the data variance can be explained with a single component.

Unrotated PCA

Oblimin rotation

FIGURE 3-10 - COMPONENT PLOTS: TOP – UN-ROTATED: BOTTOM – ROTATED (USING OBLIMIN TECHNIQUE)

The intention of this exercise was to explore the underlying variance and correlation in the data prior to clustering. As seen in Figure 3-9 and Table 3-11 the oblique *oblimin* rotation provided a more proportionate spread of the first four Eigen loadings, and hence was less skewed on its loading of the 1st and 2nd components. Given that many of the input variables were highly correlated (positively or negatively) using an orthogonal method (which assumes no correlation) is less appropriate than use of an oblique method. Figure 3-10 provides a 3D plot of components 1, 2, and 3 in both rotated and un-rotated space.

There are several possible outcomes from the information provided using PCA. One option would have been to use the rotated component data as the basis for loading the initial census data variables prior to clustering to perhaps reduce some of the bias. By applying loading factors back into the data, it can serve to reduce the weighting of the more influential variables that had been identified using PCA. Another option is to use the correlation matrix and component variance data to reduce the data set on the basis of highly correlated variables that may skew the cluster analysis. Lastly, as was the case with the UK superprofiles (Charlton et al 1985), the PCA analysis can serve to inform on less influential variables that can be removed altogether prior to clustering. However, for the purposes of this research, it was important to try and recreate the construction methods of commercial geodemographics where possible, given that the thesis agenda was to assess their application to DRR. Therefore, following consultation with the book *Geodemographics, GIS, and Neighbourhood targeting* (Harris et al 2005), specific mention was given to the fact that Experian's Mosaic product did not use PCA to reduce the number of variables and that for the purposes of neighbourhood heterogeneity and characterisation, it advised keeping most census variables.

However, prior to cluster analysis, and due to the extreme standard deviation in census area populations, it was decided to weight input data by population size per census area, thus ensuring that areas with fewer people were given less weight in clustering than more populous areas. This was done to ensure there was less skew in cluster membership. It should be noted

that another method or complimentary approach to population weighting is to remove the outlier areas with very few residents (that may skew the clustering).

3.6.4   CLUSTER ANALYSIS OF ECUADORIAN DATA

The choice of clustering technique, hierarchy and final number of profiles was not pre-defined. It was instead decided using a range of exploratory cluster analysis tests, spatial analysis (using GIS), model validation, and perhaps most importantly, taking into account the end-user requirements of this work.

Cluster hierarchy – It is a common practice in neighbourhood classification systems to have at least two hierarchical profile levels. This is largely driven by the requirements of the commercial and public sectors. For example, UK neighbourhood classification systems such as Mosaic, Acorn, and OAC are based around at least two levels of hierarchy. The first tier usually ranges from anything between 7-10 clusters and the second tier between 21-70 cluster members (Harris et al 2005). A third and fourth tier can be used but depending on the end user requirements but this is uncommon. These profile ranges provide marketing and customer retention management (CRM) with the ability to segment and target population groups at various scales of geography and precision. For example, if you have a broad demographic customer base and you are looking to locate a new hypermarket, suitably broad cluster segments might be wholly appropriate for your campaign. Although variation within these cluster groups may be diverse, such analysis may provide the user with sufficient data to make this decision. Conversely, if the user is targeting a very defined niche demographic group (e.g. pensioners living in low rise flats within large cosmopolitan town/cities) a greater number of cluster groups will be more suitable to search and focus on a particular segment that most closely matches your target demographic (Harris et al 2005).

Applying the previous analogy to DRR, it seems appropriate that such usability would also be needed to identify, target and communicate with specifically vulnerable population groups. A

hierarchical approach would also help provide this flexibility, allowing NGOs, aid agencies and civil authorities to manage disaster risk at multiple levels.

Cluster membership – The number of clusters is proportional to the heterogeneity of the profile being described. For example, in a population of 100,000, a classification split into 1000 clusters would provide a detailed profile picture of each group (average cluster member = 1000). Similarly, increasing the number of clusters reduces the Euclidian cluster distance from the cluster centroid to each case. In contrast, using the same dataset, if a classification with just 10 cluster groups was produced (average cluster membership = 10,000) there would be a larger average cluster distance for case, resulting in less detailed information about each case and greater heterogeneity within a single cluster group.

The number of clusters can also be a consideration for display purposes. Although increasing the number of clusters provides a more vivid picture of heterogeneity, it is cumbersome and hard to visually represent such differentiation on a map. With 1000 clusters it would be incredibly time consuming and possibly redundant to use such a classification for public consumption. Thematically showing 1000 different colours on a map would make it impossible to convey and communicate risk and any underlying trends to stakeholder groups. Given that the end-use of this model is to empower DRR and facilitate decision-making/strategy, a user-friendly classification is of prime concern.

There is no single definitive solution to the number of cluster groups that should be defined for each tier of a geodemographic. However, Martin Callingham helped advice on the creation of the UK Output Area Classification (Vickers et al 2006) and suggested the use of 6, 20 and 50 cluster groups for each respective tier: *"At the highest level of aggregation, the cluster groups should be about 6 in number to enable good visualisation and these clusters should also be given descriptive names."* (Callingham 2003)

Further to Martin Callingham's quote above, descriptive or memorable names for the highest level of aggregation of cluster profiles is an important factor in geodemography. It appears to

help the end users in their task to communicating to a lay audience the attitudinal commonalities of different neighbourhood types (Harris et al 2005). Likewise, Callingham points out the requirement that good visualization is aided by reduced cluster profiles at the highest level; "*At the next level of aggregation, the number of groups should be about 20. This would be good for conceptual customer profiling…..at the next level of aggregation (beyond this), the number of groups should be about 50. This can be used for market propensity measures …..this (final) level would probably also be good for use with the current government surveys. These clusters do not need names.*" (Callingham 2003)

Following careful review of previous geodemographic databases and personal communication with Richard Webber in 2012, it was decided to have a two-tiered geodemographic with cluster membership based on experimentation of the squared Euclidian distance. A lower number of clusters would then be created by re-clustering the first group. This analysis is shown in Figure 3-13.

Data transformation – Because of the inherent variation in census data types, many variables are prone to outliers. These are data anomalies that although numerically correct, serve to distort the normal distribution of data expected in nation-wide surveys. The consequence of these outliers is that some census records cannot be grouped into other clusters. If not corrected, this can result in very uneven cluster results.

There are two options for counter acting this:

- Remove the outliers manually
- Smooth the data using a log transformation

If there are very few outliers, a manual transformation of the result can be advisable to smooth the model before clustering. This can involve changing data to match the next highest/lowest figure. Thus, it should result in more rounded clusters.

The second option is to use a log transformation. By converting all aggregated data in the census to the logarithm it is possible for the effect of outliers to be greatly reduced (Harris et al 2005). In essence, it serves to smooth the model before clustering and reduces the gaps between large and disparate variables. A log can be defined as the exponent of the power to which a base number must be raised to equal a given number (Vickers et al 2006). For example, the logarithm to the base 10 of 1000 is 3 (i.e. $1000 = 10^3$). Figure 3-11 provides an illustrative example of this process.
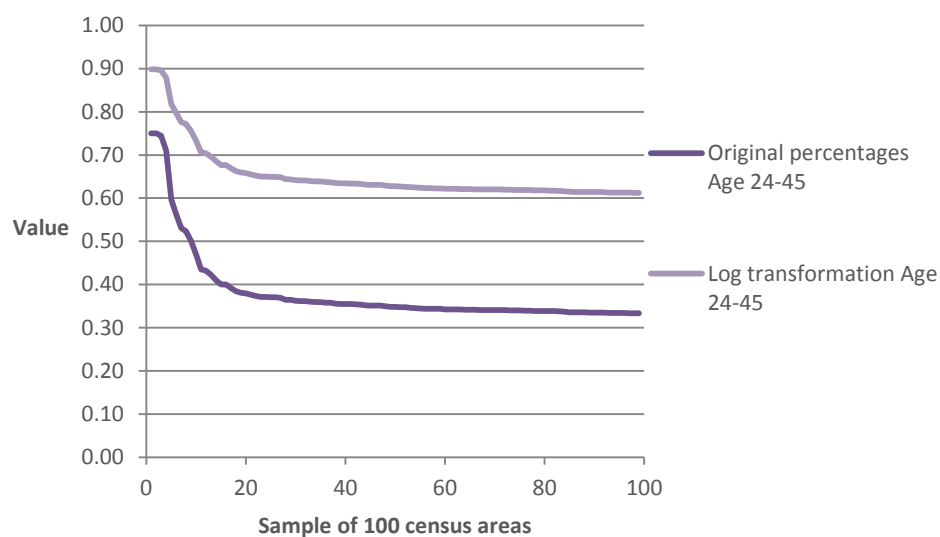
FIGURE 3-11 - EXAMPLE OF A LOG TRANSFORMATION ON ECUADOR CENSUS AGE DATA (AGE BAND 24-45)

When undertaking cluster analysis using the K-means method, users are required to select the number of iterations that the model should be ran for, with each iteration assigning cluster membership to cases. The number of iterations is largely subjective and it is often not clear how many iterations will be required to provide the optimal result. In exploratory testing, it was noted that increasing the number of iterations resulted in reducing the average cluster centroid distance to each member as the closest fit is found. However, after just a few iterations, an optimum solution is often reached, whereby further iterations did not necessarily warrant any further precision. As a rule, it was therefore decided that the model would be ran for the

maximum number of iterations available (999). Given the aggregated dataset now had nearly

4000 rows, this did not significantly impact cluster computation time, which was very rapid

(10-20 seconds).
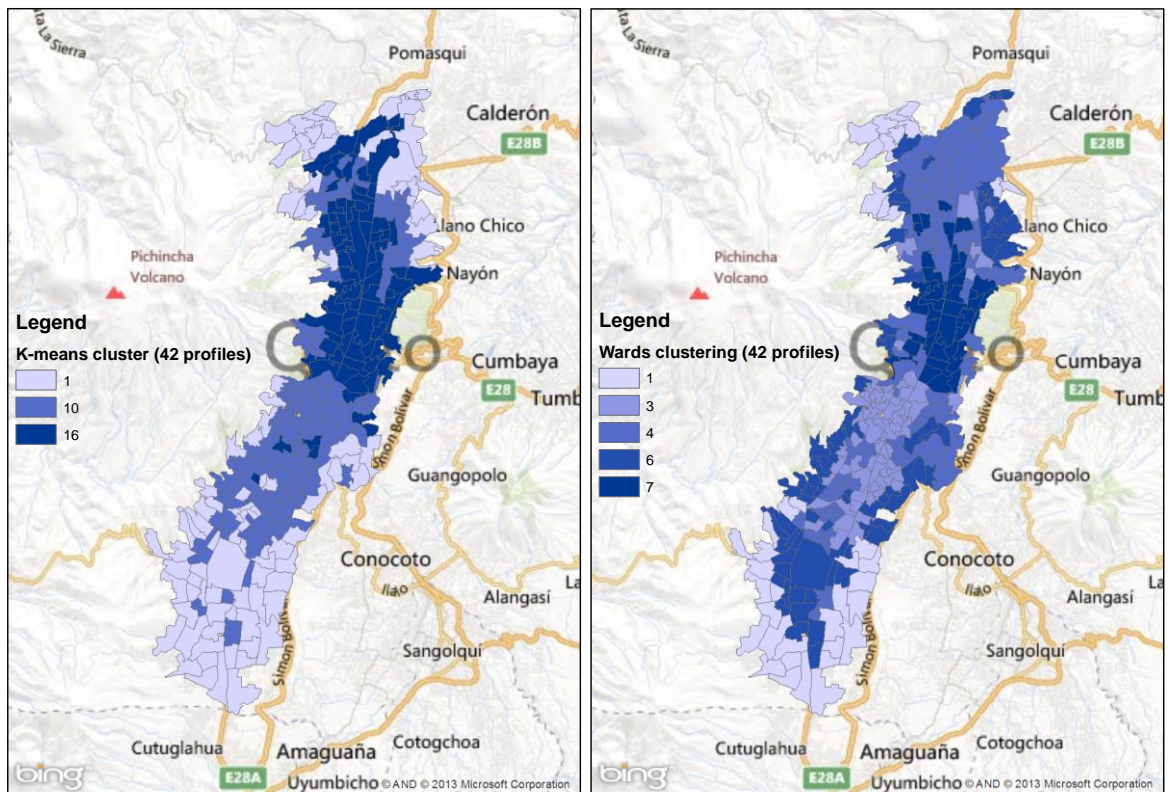
FIGURE 3-12 – A COMPARISON OF CLUSTER ANALYSIS METHODOLOGIES FOR ECUADORIAN CENSUS ZONES IN THE METROPOLITAN DISTRICT OF QUITO. LEFT: K-MEANS ALGORITHM, RIGHT: WARD'S ALGORITHM

Given that the K-means clustering and Ward's hierarchical algorithms have been widely used

in geodemographics (Jain 2010), the Ecuadorian census data was segmented applying both

these techniques to see which approach would provide more robust definition in neighbourhood

delineation. Figure 3-12 shows a comparison of census zones in Quito, using K-means and

Wards to classify the Ecuadorian population into 42 unique clusters. Both techniques show a

very strong alignment with the concentric structure of neighbourhood development in Quito.

However, as seen in Figure 3-12, the K-means approach defines the population of Quito into three groups, whereas the Wards algorithm defines five groups. Although there were broad similarities in the results of these two methods, the Wards technique was considered more favourable; largely due to the increased cluster definition it provided in urban areas and with less skewed cluster groups than the K-means approach.

Cluster distributions highlighted that the Ward's method provided a more even spread of clusters suitable for the Ecuadorian geodemographic. In contrast, the K-means cluster procedure iterates on the basis of statistical outliers as the primary seeds. Due to the statistical outliers of many Ecuadorian census areas, this latter method resulted in disproportionate cluster membership and was less suitable for the production of the classification.

A two-step approach to the creation of the cluster groups was undertaken, whereby the first step focused on creation of *Tier 2* and grouping the census zones into 42 clusters. The decision to have 42 groups was taken on the basis that this figure appeared to be the optimum number of cluster groups before significant heterogeneity would be lost from the data. Figure 3-13 shows the relationship between reducing the cluster count and the increasing Squared Euclidian Distance between clusters. This transition area is highlighted in Figure 3-13 for both clustering stages and is commonly referred to as the '*Elbow point*' – a point where the Squared Euclidian Distance increases exponentially. The second stage consisted in creating *Tier 1*. This is a top tier of eight clusters created by *re-clustering* the 42 profiles created in stage one from the Ecuadorian dataset Figure 3-13(b) highlights the second stage of clustering. Similarly, the decision to have eight cluster groups was largely driven by the Euclidian distance as well as practical applications of geodemographics discussed earlier in the chapter.
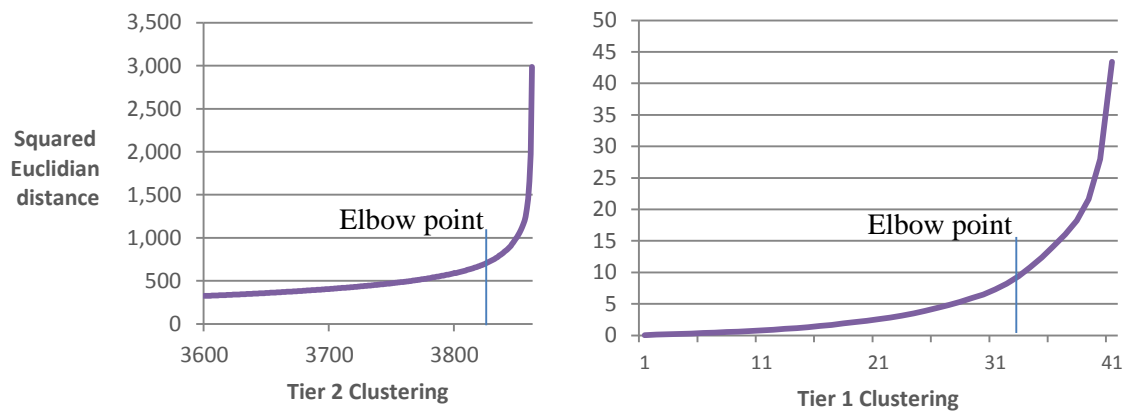
FIGURE 3-13 - CLUSTER ANALYSIS OF ECUADORIAN CENSUS AREAS USING WARDS METHOD LEFT (A) TIER 2-42 CLUSTERS RIGHT (B) TIER 1-8 CLUSTERS

Figure 3-14 shows a comparison of the propensity index ranges in the two Ecuadorian cluster tiers. It highlights that among the tier one clusters (with 42 profiles) there is greater differentiation for all Ecuadorian census variables than for tier two clusters (with 8 profiles).
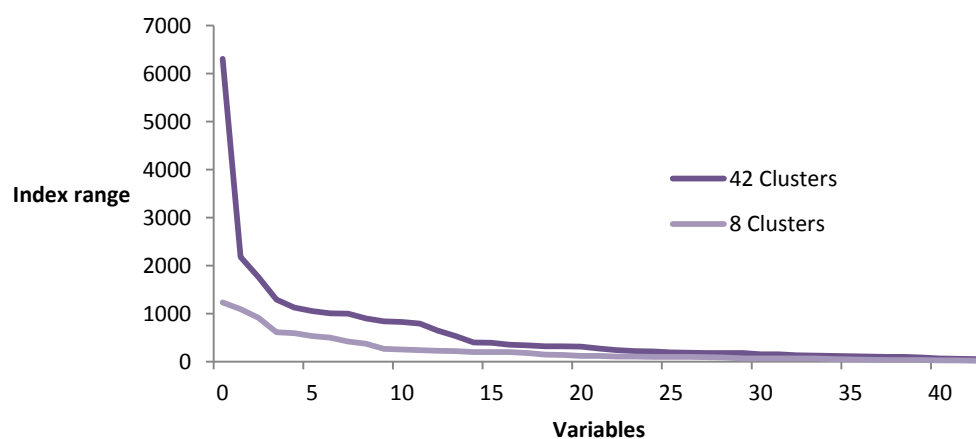
FIGURE 3-14 – INDEX RANGE VARIATIONS, TIER 1 AND TIER 2 ECUADORIAN CLUSTERS

3.7   APPLYING GEODEMOGRAPHICS TO DISASTER PREPAREDNESS AND MITIGATION

A key component regarding the application of neighbourhood classification systems to DRR is that it enables aggregated demographic profiles to describe multiple areas, rather than just discrete single statistical units. By adopting this principle, it allows statisticians, policy makers, and NGOs to make use of third party survey data and link such insights to the pre-defined clusters.

This chapter details a practical example of how readily available Ecuadorian telecommunications survey data (INEC 2013) can be linked to the geodemographic clusters detailed in chapter 4. This worked example shows how insight on personal communication devices supplements the geodemographic profiles and can be used to help direct disaster alerts to different neighbourhood profiles based on propensity indices.

Lastly, this chapter highlights the value that risk perception research can bring to DRR. During field work in 2010, quantitative and qualitative survey work was undertaken in Quito, Ecuador to help provide further information on the risk perceptions of the resident population. Findings from these workshops are presented in this chapter as they have direct relevance to the core contribution of this thesis as well as providing key contextual background to the study region.

3.7.1   COMBINING THE ECUADORIAN NEIGHBOURHOOD CLASSIFICATION WITH SURVEY DATA ON MOBILE TECHNOLOGY AND INTERNET ACCESS

In 2010, the INEC surveyed 73,686 individuals across Ecuador regarding their demographic status, their educational achievement and 21 questions concerning both their access and use of telecommunications and media. Questions included whether the individual had access to a mobile phone, a smartphone and/or the internet as well as how much they used the devices on a daily/yearly/weekly basis. The survey data is readily available to download via the INEC website in either SPSS/txt data formats.

The survey data is referenced to each Ecuadorian census zone and Parroquia. However, to help focus the application of this work to volcanic vulnerability, and to align with the previously reported analysis, only results for Quito census zones are discussed in detail. For meaningful use of this data, it was required to aggregate the survey replies for relevant questions into counts by census zones. As the INEC survey was not a census itself, it did not cover responses in all the census areas across Ecuador or the metropolitan district of Quito. Survey replies were provided for only 25% of the Quito census zones used in creation of the classification system.

The next consideration regarded applying single % values and scores for each of these questions to the previously created geodemographic profiles found in Quito. A summary from this analysis can be seen in Table 3-12.

TABLE 3-12 - QUITO CLUSTER GROUPS (TIER 1 AND 2)

Tier 1 clusters	Have enabled cell phone	Have a SMARTPHONE	Have used a computer in the last 12 month	Have used the internet (last 12 months)
6	65%	11%	51%	48%
7	78%	27%	69%	68%
Tier 2 clusters				
1	56%	9%	44%	41%
3	66%	14%	53%	50%
4	73%	17%	64%	59%
6	63%	4%	41%	39%
7	78%	27%	69%	68%
Category average	67%	13%	53%	50%

To do this it was necessary to match cluster categories (for tiers 1 and 2) to every census area covered in the survey data (including regions outside of Quito). A pivot table was then produced from this cross-referenced data and each cluster profile was assigned the mean average score for that category, and for each survey question. Table 3-12 presents the results for only those profile groups found in Quito.

As seen in Figure 3-15 and Figure 3-16, these cluster percentage scores have been mapped to the corresponding census zones in the metropolitan district of Quito. Figure 3-15 represents the

percentage of respondents in the census zone that have access to a mobile phone, whereas

Figure 3-16 shows the percentage that have used the internet in the last 12 months.
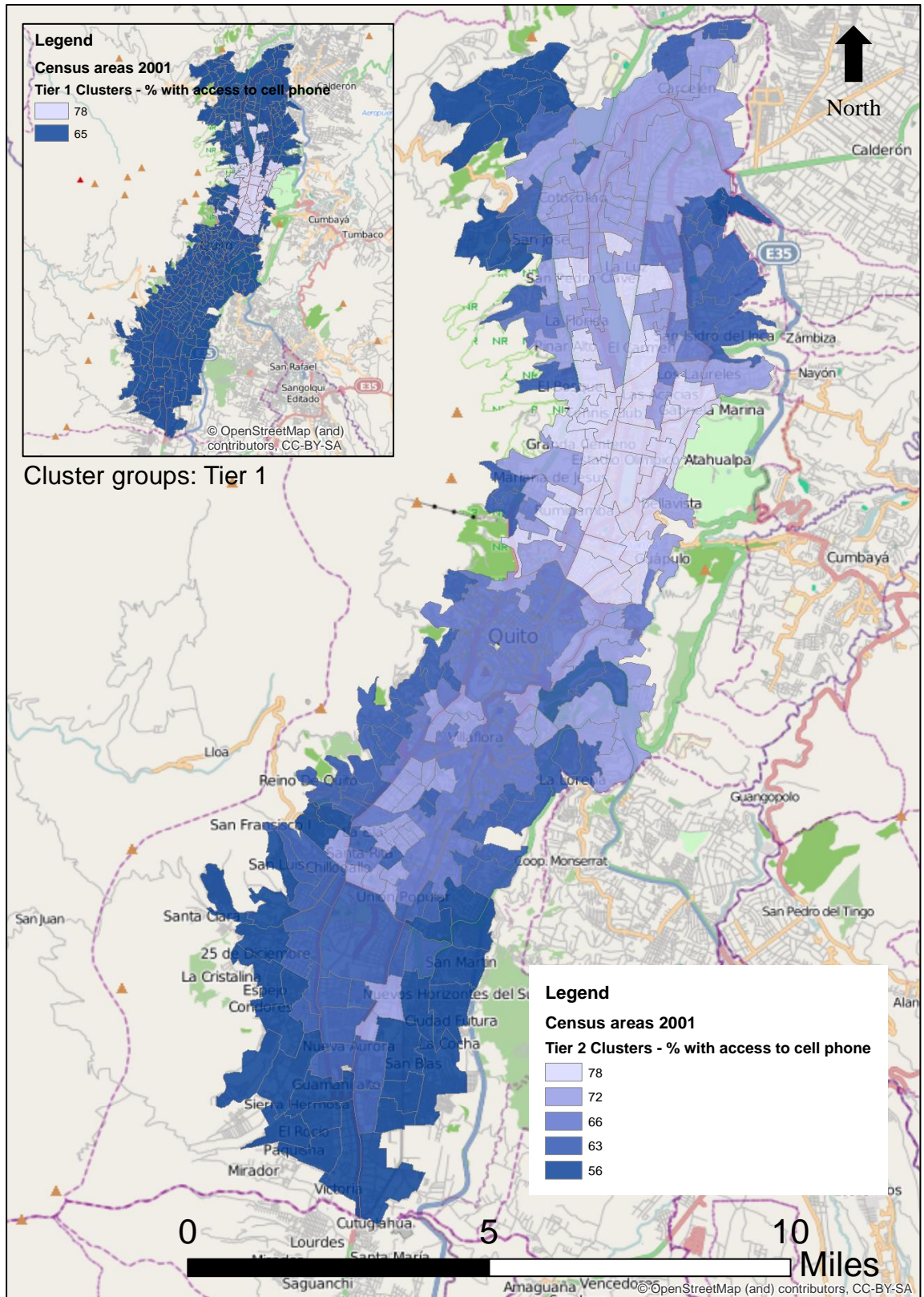
FIGURE 3-15 - TIER 2 GROUPS IN QUITO MAPPED TO MARKET SURVEY DATA (% WITH ACCESS TO A MOBILE PHONE)
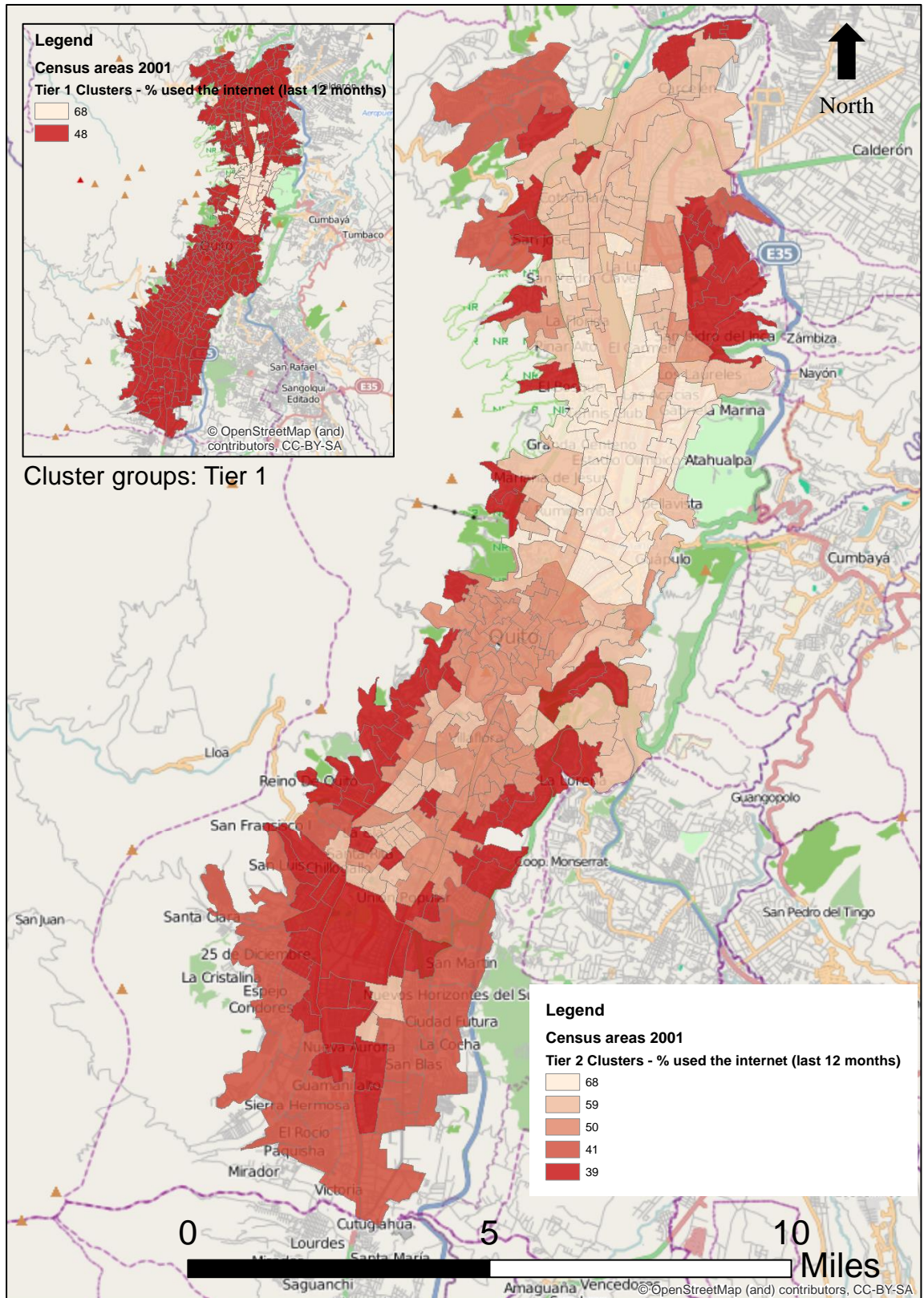
FIGURE 3-16 - TIER 2 CLUSTER GROUPS IN QUITO MAPPED TO MARKET SURVEY DATA (% USED THE INTERNET IN THE LAST 12 MONTHS)

It is interesting to note the concentric nature of the results in relation to the structure of the city. In summary both Figure 3-15 and Figure 3-16 show that people living closer to the city centre of Quito are more likely to have access to both a mobile phone and the internet. A similar pattern of results was recorded for all the questions regarding telecommunications and media, including 'SMARTPHONE ownership' and 'access to a computer'.

Using INEC data on the average household income for Quito census areas, the survey results were compared to see if there was any correlation between these variables. As Figure 3-17 shows, the survey results detailing people's access to telecommunications and media technology appear to show a positive correlation with household income in Quito. For example, mobile phone access has a correlation of 0.6 with household income.

% that have a mobile phone

$R^2 = 0.3621$

% that have a SMARTPHONE

$R^2 = 0.2856$

% that have used a computer (last 12 months)

$R^2 = 0.3745$

% that have accessed the internet (last 12 months)

$R^2 = 0.3881$
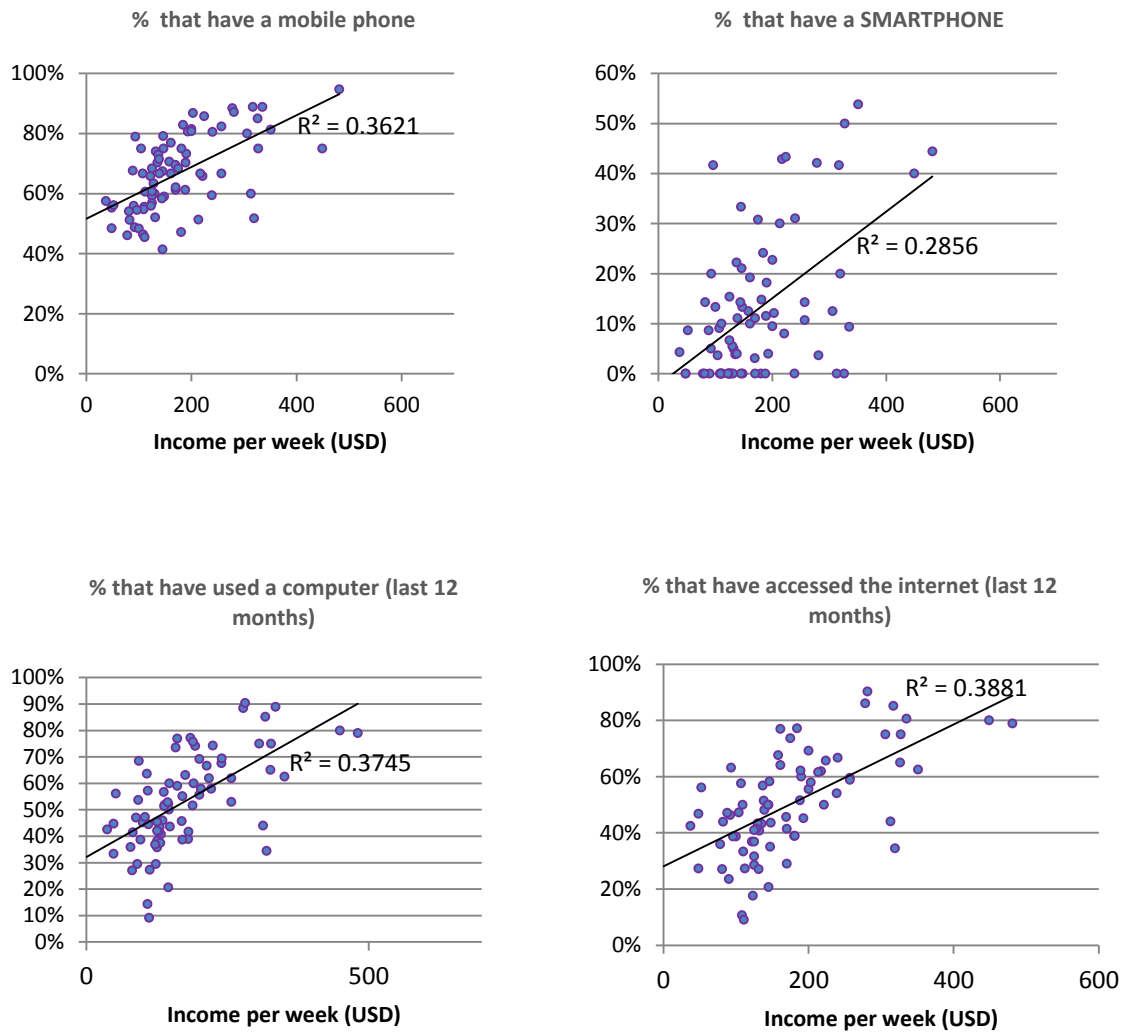
FIGURE 3-17 - CORRELATION OF MULTIMEDIA COMMUNICATION DATA AND HOUSEHOLD INCOME, QUITO. CLOCKWISE (TOP LEFT) % MOBILE PHONE, % SMARTPHONE, % COMPUTER ACCESS, % INTERNET ACCESS

It is worth noting that although the cluster groups show a strong spatial alignment, because each cluster profile is assigned a single value, the variation in actual survey results recorded for each census area is lost. For example, based on the entire survey dataset, cluster group 7 (*Affluent city living*) implies that an average of 78% of the population in this neighbourhood will have access to a mobile phone. This average is based on the survey results of all census areas with the same geodemographic classification being assigned. However, when analysing only group 7 cluster areas that are in Quito, the average from the survey data is 65%. The standard deviation for this variable in Quito is 31.3%, with a minimum recorded value of 6% and a maximum of

94.74%. This shows a high volatility and uncertainty in the survey results of this question and is also a reminder that data reduction and aggregation techniques substantially impact the underlying variance of survey data.

3.7.2 PRACTICAL CONSIDERATIONS FOR DISASTER ALERT AND COMMUNICATION OUTREACH

A key benefit of the methodology described in the previous section is that it can be used by relevant stakeholders as a practical source of insight for disaster risk management and preparedness. The penetration of mobile/smartphone/internet usage among different community types provides key information as to the relevant effectiveness of possible DRR campaigns and alert systems. Figure 3-16 demonstrates thatamong the most vulnerable communities in Quito, only an average of 39% of the population have used the internet in the last 12 months. Such information may severely hinder a civil authority measures or NGO campaigns aimed at raising awareness to the imminent threat of volcanic tephra fall given increased activity from a volcano. Conversely, Figure 3-15 highlights that an average of 56% of the population have access to a smart phone. This may infer that SMS or telephone alerts may provide a more effective communication channel. It should be noted that the example hypothesised here is greatly simplified, given that risk communication is an extremely complex process and such descriptions do not take into account the effectiveness (or lack of) such campaigns.

Warning channels can refer to multiple mediums by which hazard/risk information is communicated to the public: e.g. face-to-face contact, telephone, SMS, siren, radio, newspapers, television and the internet. The behavioural response to warning messages is very often governed by the pre-existing beliefs about the nature of the hazard that the recipient has (Lindell and Perry 2004). Likewise, the credibility of the source of a warning appears to be of critical importance as to whether an individual will comply or reject the message (Mileti and Sorensen 1987). With particular focus on this point, Haynes et al (2008) conducted semi-structured interviews with Montserrat residents following a long period of volcanic activity (1995-2005) on the island. They observed that distrust was highest with groups such as the

world media and lowest with *friends, family,* and *scientists*. Likewise, similar findings have also been noted following other volcanic eruptions, such as Mount Helens and in New Zealand (Perry and Greene 1983; Ronan et al 2000).

3.8   SURVEY OF RISK PERCEPTION IN QUITO

As well as the multivariate statistical approach to volcanic vulnerability that was presented in the previous chapters of this thesis, and forms the principal contribution of this research agenda, qualitative field work was also undertaken in Quito, Ecuador, between September-October 2010. The purpose of this work was to gain a greater understanding of the cultural and physical risk perceptions of resident Quitenos' with specific regard to the neighbouring volcano of Guagua Pichincha. This work proved extremely insightful. While the findings and research presented in this section do not directly influence the social vulnerability indices, model and geodemographic classification, the survey responses and qualitative results help profile the inherent complexity of risk perception of Quito residents as well as the challenges in mitigating and communicating disaster risk in Ecuador. Such findings can be considered integral to the conceptual model presented later as well as in the overall discussion and findings of this research.

Field work included the hosting of four focus groups and individual surveys to gain feedback on people's thoughts around the various aspects of volcanic risk perception in Quito. As well as providing a presentation on volcanic risk, group discussions typically lasted 60 minutes. The questionnaire had 25 questions, and was split into five sections; risk perception, disaster evacuation, the effect of a disaster on income, the respondent's demographic status and housing stock. As student's English ability varied significantly within and between groups, from basic levels to a very high competency, it was necessary to prepare versions in both English and Spanish – this ensured that all questions could be sufficiently communicated to survey

respondents. Copies of the surveys are provided in the *Appendices* section of this thesis. The survey was purposely a mixture of both quantitative and qualitative questions so that further research, beyond the scope of this thesis, could be carried out regarding attitudinal/cultural risk perceptions. The focus groups were conducted in two adult learning language schools in Quito; the Wall Street institute (http://www.wsi.com.ec/) and the Berlitz English School (http://www.berlitz.com.ec/).

3.8.1   FOCUS GROUP FEEDBACK

Feedback from the focus groups was based around five discussion questions. The questions were intended to facilitate group debate on key research themes and to gain a greater understanding of the real cultural and attitudinal perceptions that Ecuadorians have towards volcanic risk:

1)   How are volcanoes formed?

This question was asked to establish how much knowledge survey respondents had with regard to the geophysical processes that form volcanoes (e.g. plate tectonics, subduction/divergent boundaries).

Discussion: Awareness varied among focus group attendees and among individuals but in general terms, responses and discussion indicated a moderate-high level of knowledge regarding the geological processes that form volcanoes. References were made repeatedly to "*plates colliding*" and the association of "*earthquakes*" at the same tectonic boundaries. Similarly, many were aware that effusive lava had formed the neighbouring volcanoes.

2)   What did your ancestors say about the volcanoes of Ecuador?

Discussion of historical perceptions of volcanoes opened up numerous and somewhat revealing accounts of cultural heritage that belie many Ecuadorian families.

A young Mestizo female student recounted:

"My grandmother said the volcanoes are 'Gods of the people'......my ancestors used to give thanks to the Chimborazo (volcano), because that was the biggest volcano. They said that volcanoes are family members (and are related to each other). For example, Rocha Pichincha is the little brother of Guagua Pichincha, and Tungurahua is the cousin of Cotopaxi"

Other opinions were similar in their religious commentary:

"They thought the volcanoes were Gods. The angry gods occasionally erupted."

"They used to talk about all the damage caused when they erupted – ash and fire and floods. The ash would sometimes break the windows."

"It was said that the Virgin Mary saved the town of Banos from Tungurahua!"

"Our Inca ancestors thought that they were gods."

Such feedback was not isolated and although nearly all respondents acknowledged the scientific reasoning behind the formation of volcanic hazards, they were also conscious and to some extent still believed in the religious omen afforded by their ancestral/pre-existing Inca, pre-Colombian, or Catholic beliefs. It proved an interesting paradox in discussions.

It is also worth noting that many ancestral beliefs were focused on the positive aspects of the volcano – which in many cases seems the acknowledgement that volcanoes bring fertile soil and help sustain the micro-climate of the Sierra region.

"Volcanoes provide life to the sierra."

"Do not be scared of volcanoes. They are good for Ecuador."

3)   Do you think that there'll be a large eruption of Guagua Pichincha in the future?

Most people did not seem to believe a large eruption would occur in their lifetimes but several were quick to note, that given the small eruptions since the volcano's re-activation in1999 (when Quito was covered in a fine layer of ash), future eruptions could be likely. One mature student commented:

"No. This volcano has been quiet for many years."

Others were more cautious of ruling out a future eruption.

"Maybe, but I'm not very worried"

"Not a big eruption. The last one was hundreds of years ago."

4)   What do you think you would do if there was an eruption?

This question proved enlightening in the variation of responses. It suggested a severe lack of existing civil evacuation measures and the formal communication of government advice in this regard:

"I don't know. People are not prepared for the ash fall or eruption."

"I'd try and breathe through a mask."

"Evacuating the city would be very difficult. There'd be too much traffic. I'd have to stay here"

"I'd stay at home. The city would be too crazy to leave."

"I'd try and evacuate."

"It depends how big the eruption is. I'd stay at home if it was small"

5)   Do you feel confident in how the government/local authority would help you if there was an eruption?

Similarly, the level of confidence in existing plans appeared very low:

"The Ecuadorian government don't plan – they just react! The level of alert at the moment is low so there's not much going on. If there was an eruption they'd just improvise."

"No, there's no evacuation plan."

"No. They're useless"

"The only evacuation training was at school."

It is also worth noting that there appeared to be no urgency on the part of individuals regarding the need for a plan or lack of communication regarding current measures.

3.8.2   SURVEY RESPONSES AND FINDINGS

As seen in Figure 3-19 (c), 78% of survey respondents were either 'not worried' or only 'a little concerned' about the implications of a future eruption of the neighbouring Guagua Pichincha volcano. Further to this, respondents were asked to quantify their answer to the statement *"There will be a large volcanic eruption of Pichincha in the next 10 years". Please indicate the level to which you agree/disagree with this statement*" with a numeric value between 0-100 (0 indicating strong disagreement, 100 indicating strong agreement). The mean average response for this question was 49, a value indicating respondents were 'not sure'.

Similarly, as Figure 3-18 highlights, a correlation test of these quantitative survey responses was undertaken. Survey responses were compared to the age of the respondent to see if there was any correlation. When both variables, *Impact of an eruption on household income* and the *Likelihood of an eruption*, were compared to student's age, correlation values were very low (-0.30 and -0.36 respectively). Although such a result would indicate little/no correlation significance, it is important to reflect on the small survey number used in this field work (n=37).
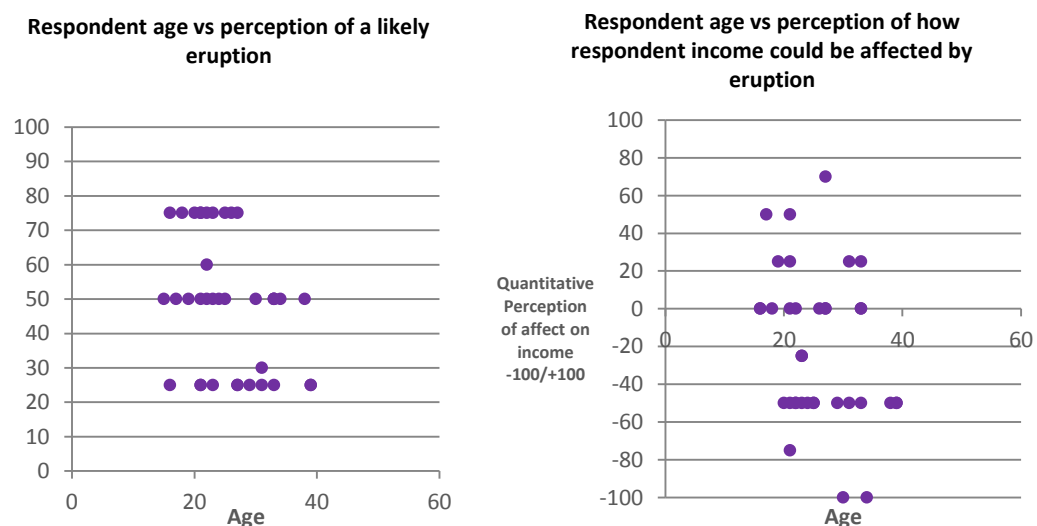
FIGURE 3-18 - CORRELATING QUANTITATIVE SURVEY VARIABLES OF QUITO RISK PERCEPTION. LEFT: RESPONDENT AGE VS INCOME, RIGHT: RESPONDENT AGE VS ERUPTION RISK PERCEPTION

The mean average age of survey respondents was 26 years, with 68% of individuals classifying themselves as single and 76% being of Mestizo ethnicity (i.e. of combined European and Native South American descent). Employment type within the survey was very mixed, with the largest proportion reporting to be in skilled work (30%) or financial/managerial roles (22%).
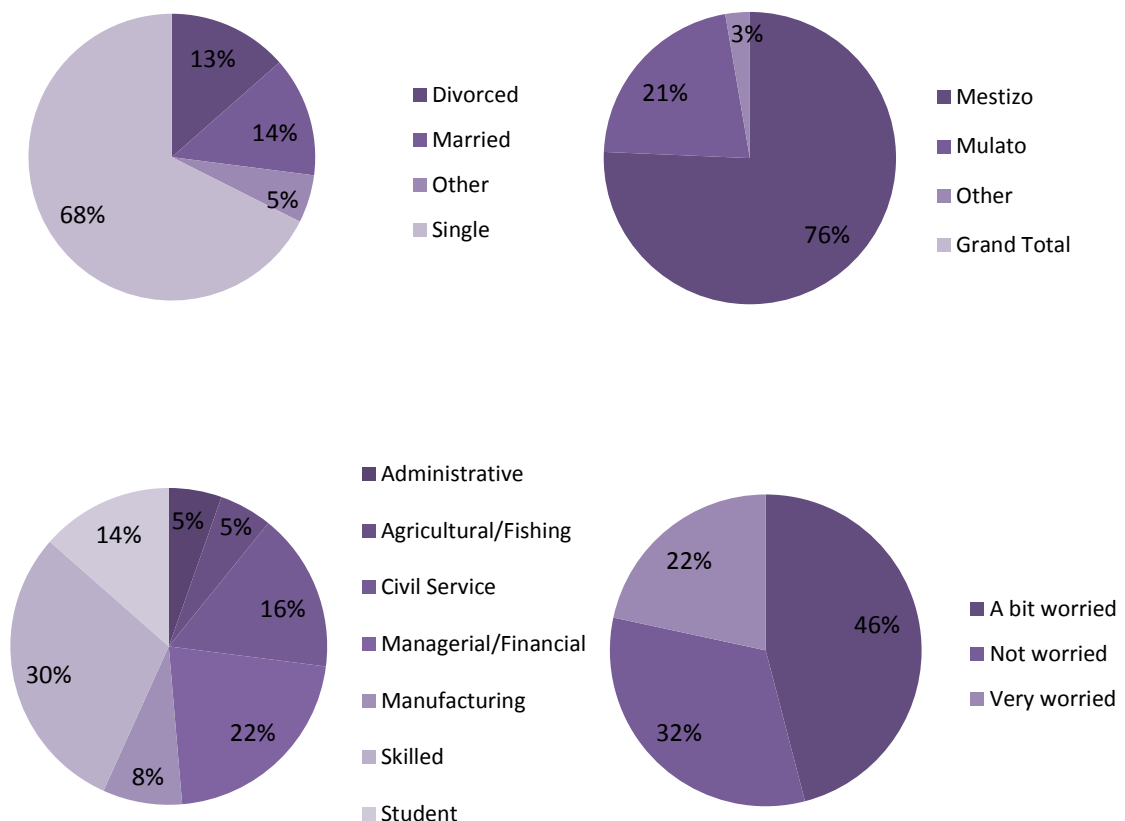
FIGURE 3-19 - SURVEY RESULTS FROM VOLCANIC RISK PRECEPTION SURVEY. CLOCKWISE FROM TOP LEFT: A) MARRITAL STATUS B) ETHNICITY C) PERCEPTION OF A FUTURE ERUPTION D) EMPLOYMENT TYPE (FIELD WORK, QUITO, 2010)

Students were also asked to provide information on the parish (Parroquia) in which they lived. As would be expected, most respondents lived in the neighbouring Parishes to the English school that they attended, such as Cotocolloa, Pomasqui and Carcelen. However, some had travelled from further suburbs around the city to attend the classes.

Further summary and the relevance of survey results is also provided in chapter 5 of this thesis.

3.9   CHAPTER SUMMARY

Building on the conceptual neighbourhood framework presented at the beginning of this
chapter, a methodology was devised based on the selection of social vulnerability traits within
communities using a commercial geodemographic classification. The weighting of the various
input variables was focused on the discriminatory characteristics of each geodemographic
propensity index. This assessment involved using multiple statistical tests, including correlation
analysis, index ranges and the Gini-coefficient of a given variable. The methodology was then
applied to Mosaic Italy 2007 data to create a series of vulnerability indices for different aspects
of DRR.

To further assess the relative merits of the previously applied methodology in a different
regional context, a bespoke geodemographic classification was then created for Ecuador using
the INEC 2001 census data. The development outlined in Section 3.6 highlights how PCA and
cluster analysis were used to create a two-tiered neighbourhood classification for Ecuador that
could then be applied to assess the social vulnerability of populations living around
volcanically active parts of the country.

The resulting neighbourhood classification, social vulnerability indices, and GIS spatial
analysis of the relevant study regions are presented in chapter 4.

In the last section of this chapter, it was highlighted how tertiary marketing or survey data can
be applied to geodemographic information to provide additional propensity indices for a given
area. This technique is based on regression of the survey results to areas with similar
geodemographic profiles and provides an effective means of enhancing the neighbourhood
profiles for a variety of uses.

In the example provided in section 3.7.1, it was shown how telecommunication data could be
inferred on the Ecuadorian geodemographic to help segment the Quito population in terms of

communication channel access. It was demonstrated that the highest proportion of those with a mobile phone were found to be in the *Affluent city living* cluster profile, located near the city centre of Quito. Conversely, it was shown that the *Peripheral barrios* profiles had the lowest percentage of those with access to the internet.

Such insights could be used by civil authorities, NGOs and practitioners to help infer on likely communication channels for disaster preparedness, mitigation and response.

Lastly, section 3.8 presented findings from qualitative and quantitative risk perception workshops held in Quito in 2010. The wide spectrum of survey results from this work helped highlight that volcanic risk is perceived as being inherently low in Quito. This was found to be a combination of ancestral attitudes to volcanic risk, the inactivity of Pichincha in the 20th and 21st century, and the competing social issues of life in Quito, such as crime and high unemployment.

4.   RESULTS

In this chapter, results are presented for the two phases of research involved in this thesis. Phase 1 results present GIS derived vulnerability index maps for the relevant social and physical risks presented by a Sub-Plinian eruption of Mount Vesuvius, Italy. Brief discussion is also given to the spatial distribution and correlation of risk with respect to the Mosaic Italy classifications used as the basis of the indices presented here. Phase 2 presents a geodemographic classification for Ecuador, with a 2-tier cluster solution and accompanying propensity index tables for all modelled variables. As well as tabular results, cluster maps are also provided to show the spatial variance of the divergent cluster groups across Ecuador. Phase 2 results also provide a case study analysis of the census based geodemographic to analyse the modern day impact of a 10th century eruption of Guagua Pichincha, the neighbouring volcano to the metropolitan city of Quito. The final section of this chapter highlights the results from a model validation exercise of the Ecuadorian geodemographic. A vulnerability index based solely on original census information is compared and contrasted to geodemographic cluster results to assess the degree of correlation in defining social vulnerability.

4.1   PHASE 1: VESUVIUS, ITALY

In this section, the results and findings from using a commercial neighbourhood classification system to measure volcanic social vulnerability around Mount Vesuvius, Italy, are presented and discussed. Thematic GIS-derived maps have been created for the various social vulnerability measures as well as an overall social and physical vulnerability index. The results are based at the census output level for the analysis territory around the volcano and help show the geodemographic drivers of vulnerability around Vesuvius. They also show the most vulnerable demographic classifications and patterns of vulnerability across the study area.

Likewise, detailed discussion is given to the basis and nature of the volcanic risks that have been modelled to hypothesise a Sub-Plinian eruption of Vesuvius. These hazards include the Tephra fall, Pyroclastic density currents (PDC) and the official evacuation zones as defined by the civil authority.

4.1.1   MOSAIC ITALY MODEL RESULTS

The following section presents the Mosaic vulnerability indices and model outputs for the census areas around Vesuvius. Results are presented using GIS derived thematic maps. As well as spatial interpretation of these results within the context of the local geography, each social and physical vulnerability variable is examined separately.
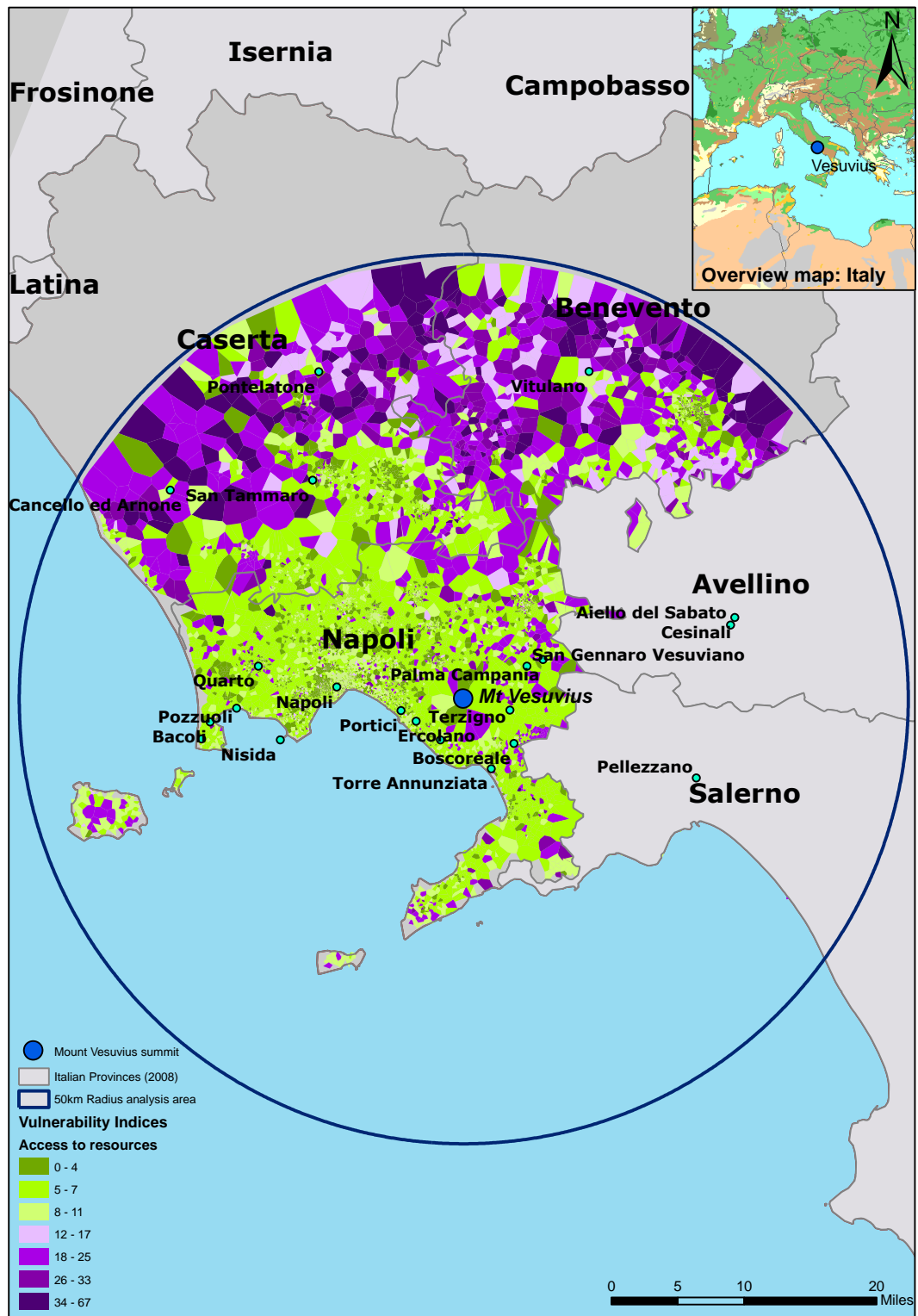
FIGURE 4-1 - ACCESS TO RESOURCES VULNERABILITY INDEX – MOSAIC ITALY 2007

Vulnerability in accessing resources during a disaster appears to increase as the location of the community becomes more rural. Figure 4-1 highlights this spatial relationship, showing the areas most at risk.

These census areas are geographically focused on the periphery of the analysis region in largely rural census areas. Conversely, the urban areas, found around the Naples conurbation (displayed here in dark green) show a lower level of vulnerability. This variation suggests urban areas have higher scores concerning those factors that reduce a household's vulnerability to access resources during the onset of a natural disaster. Given that contributing factors to this category included a household's *access to water*, many rural areas were impacted because of having a high Mosaic index score for that particular variable. In terms of the most vulnerable Mosaic groups for inadequate access to resources, the profiles over-represented in this category came from the *Large Farmhouses* Mosaic group.

Types 47 and 46 were among the highest ranking areas for this category, which correspond to *Very remote self-employed farmers* and *Large farms* in very low density areas. These neighbourhoods are largely self-employed agricultural workers with big families. Household amenities are often very limited and hence this group can be particularly vulnerable to having limited physical resources during a disaster. The lowest social risk categories were largely formed from the *Wealthy Elite* group. The census regions assigned to this group are located in the highly urban areas, predominantly in the city centre apartments and prosperous suburbs around Naples.

FIGURE 4-2 - Evacuation vulnerability index – Mosaic Italy 2007

In marked contrast to an individual or household's access to resources, the risk of evacuation to communities is found to be highest in urban areas and the least in distant rural regions. As seen in Figure 4-2, the highest ranking areas to disaster evacuation are found principally in and around the city of Naples as well as other smaller towns and conurbations scattered around the Mount Vesuvius area. This included the settlements of *Caserta, Avelino, Atripalda, Mercato San Severino, Salerno*, and *Benevento*.

The Mosaic profile of these high risk areas to evacuation are inextricably linked to the defining variables of the risk category, as defined earlier in the study. Demography, household size, the number of stories in a building and population density were considered key parameters in determining an area's difficulty in disaster evacuation. Therefore, older people living in city centre apartments are defined as the most vulnerable neighbourhood group with regards to the stresses of evacuation. The Mosaic Italy classifications *Elderly Households* and *Urban Apartments* are the main categories, with types 8 and 7 among the highest ranking within this group. The lowest scoring classifications for evacuation risk were *Large Farmhouses* and *Rural low income*. The *Elderly Households* are very often single occupancy, likely to be over 65 years old and live in older, more vulnerable housing stock.
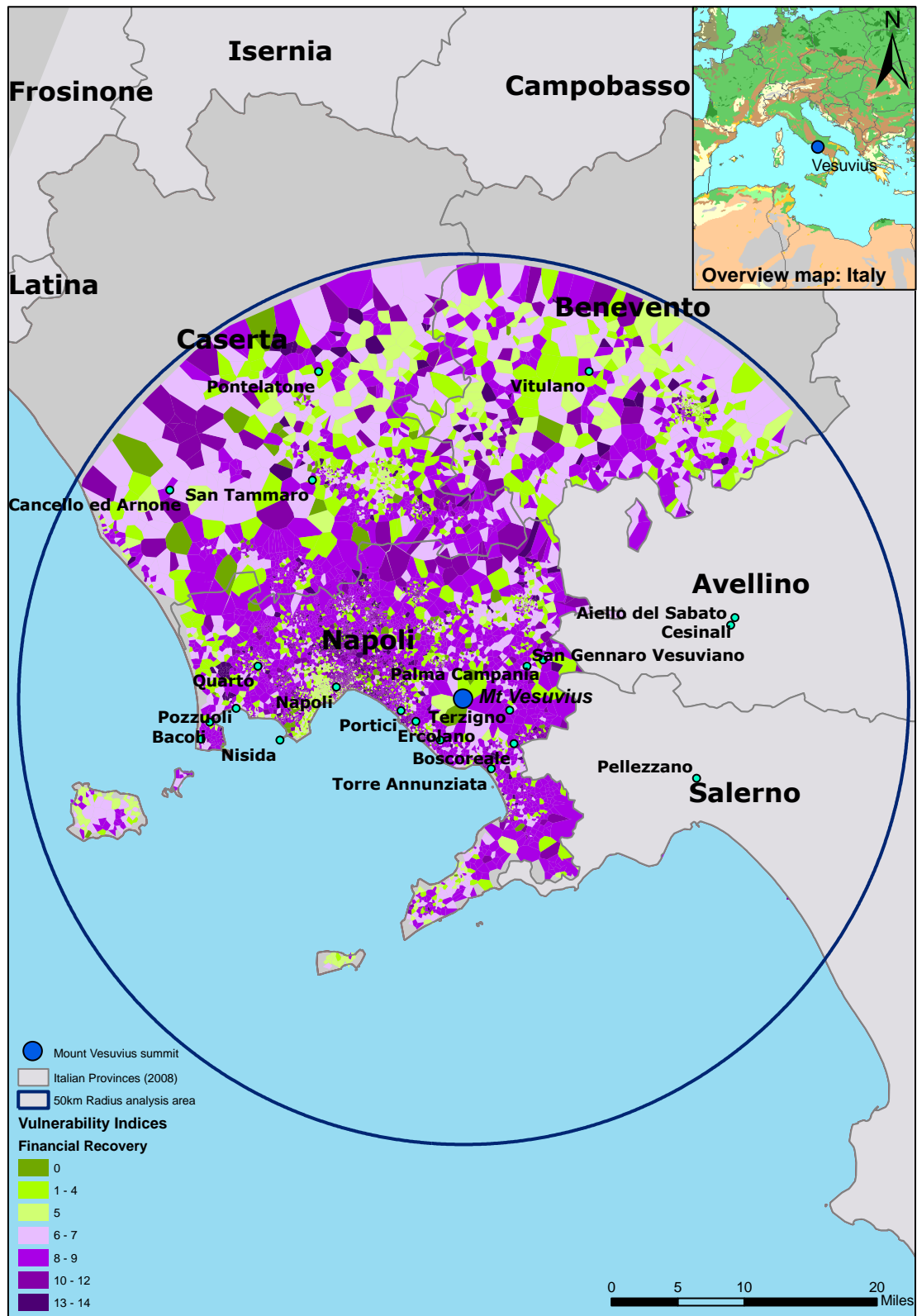
FIGURE 4-3 - Financial recovery vulnerability index – Mosaic Italy 2007

The defining characteristic of those areas most vulnerable in terms of financial recovery from an eruption would be those profiles of particularly low income and educational attainment. The spatial distribution of less affluent areas around the volcano has a mixed pattern. Areas of higher risk are found in both urban town centres as well as rural isolation. In fact, the only common denominator is low income, as other characteristics such as age and ethnicity are not factored into the risk score. Likewise, factors such as having personal credit and loans, which could be perceived as a risk regarding financial recovery from a disaster, are equally among Mosaic groups, and do not appear to have high spatial correlation within the study area.

However, the Mosaic Italy profiles that are over-exposed to the risk of financial recovery are *Low status apartments* and *Rural low income*. This explains the mixed geographic distribution of financial risk seen in Figure 4-3 as these categories have an inverse spatial pattern with regards to urban proximity. It should be noted that Low status apartments is also statistically the most common neighbourhood classification in the Naples province, making up 46% of all household types.
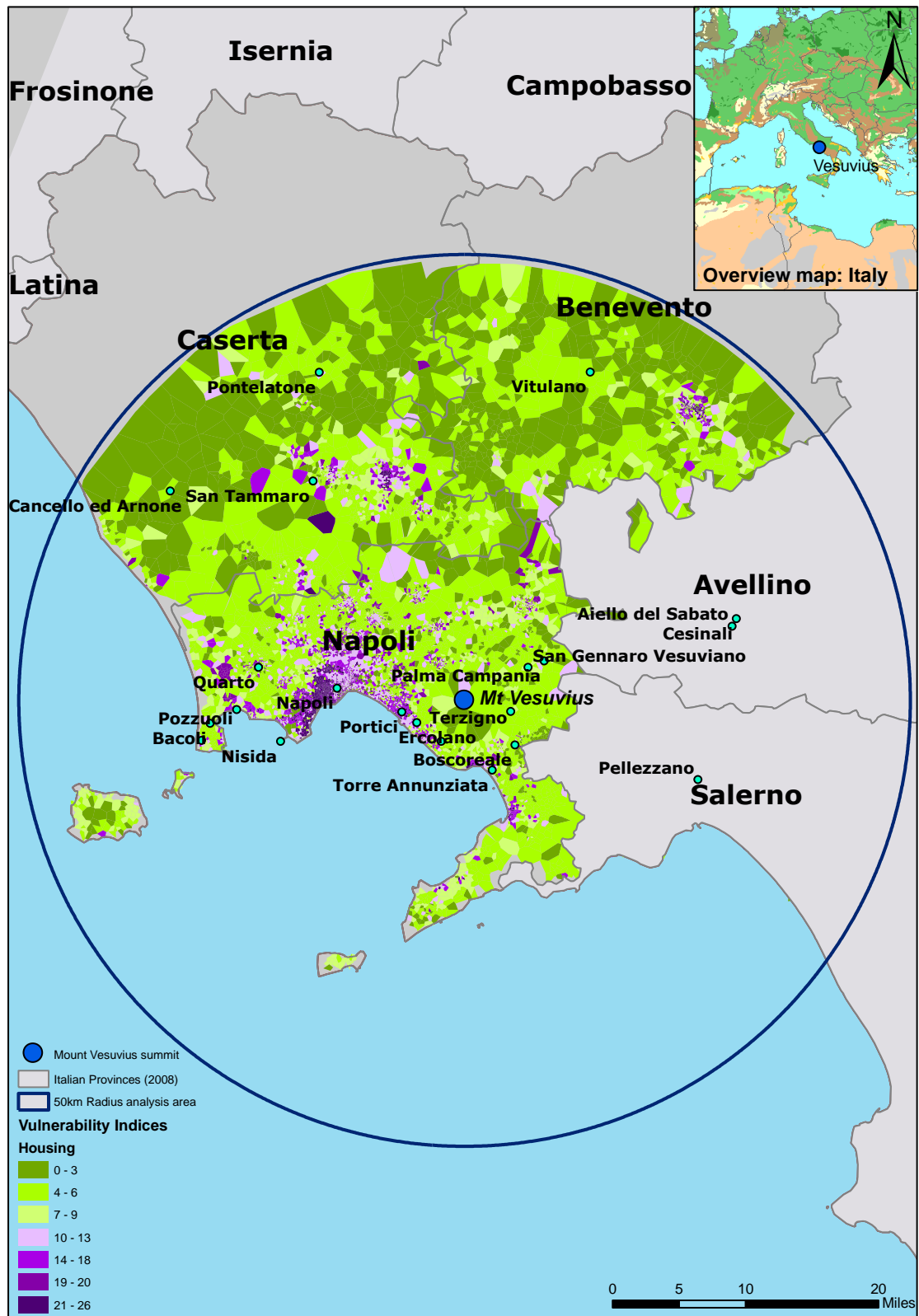
FIGURE 4-4 - Housing vulnerability index – Mosaic Italy 2007

Household vulnerability scores were highest in the principal urban areas. This is because the combination of input factors was largely concerned with the physical/structural integrity of the buildings in a given census area. Vulnerability was higher if the building was pre-1919 or high-rise, due to likely structural performance of the building with regard to hazards such as pyroclastic flows/tephra fall. Because the proportion of older building stock is higher in many wealthy urban areas, these regions were among the highest risk, as seen in Figure 4-4.
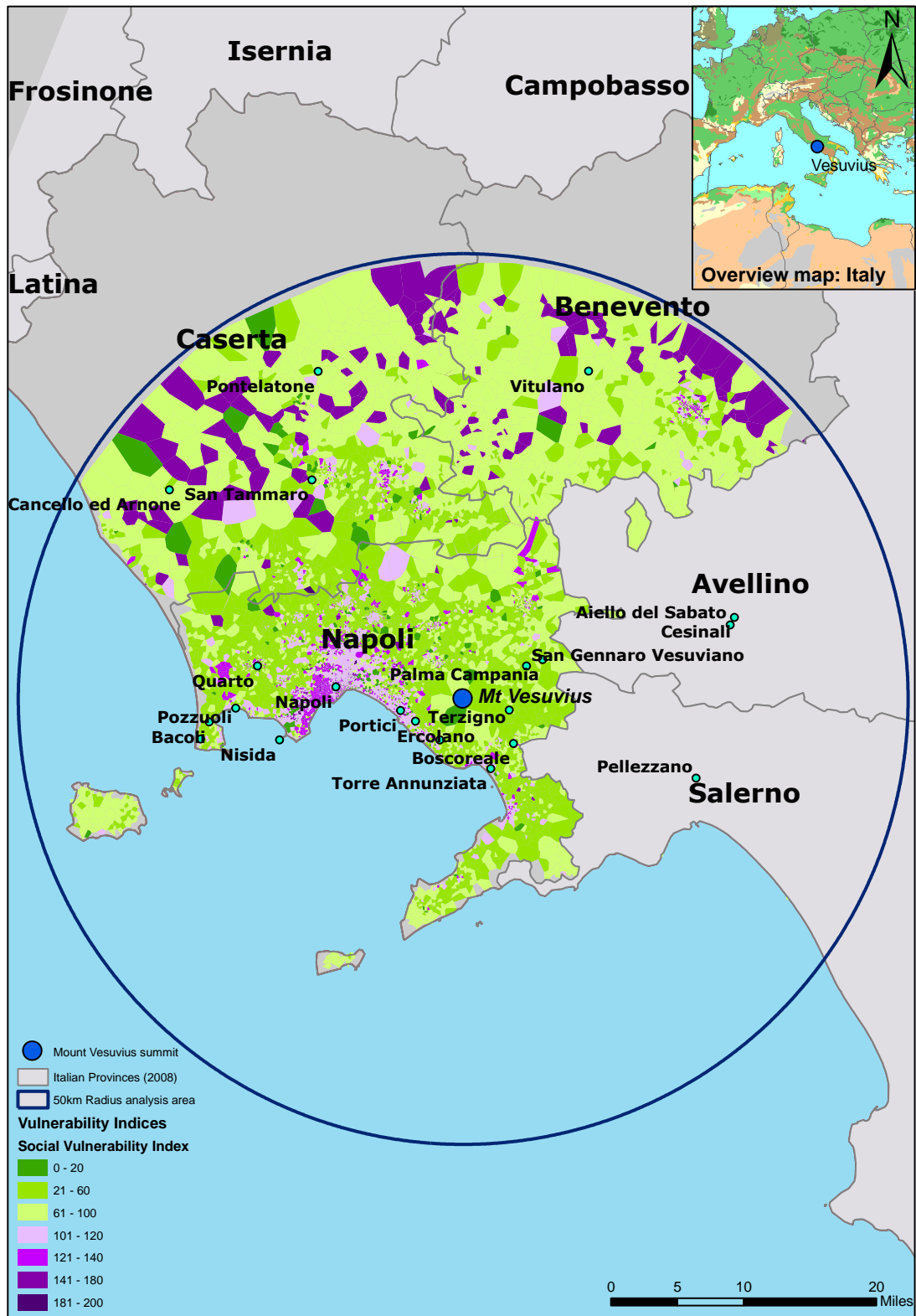
FIGURE 4-5 - Social vulnerability index – Mosaic Italy 2007

The combination of the previously mentioned vulnerabilities during an eruption (access to resource, evacuation, financial recovery, household physical risk) creates a polarised pattern of social vulnerability across the Neopolitan area when combined into a single index as shown in Figure 4-5. Pockets of highly populated urban areas are deemed equally as vulnerable as low-income rural regions. This pattern is reflecting the spatial conflict of the various geodemographic profiles making up each risk category. For example, though a lack of access to resources is defined as having a rural pattern, in the social vulnerability index urban areas are considered equally as vulnerable in other categories, such as the likely problems of such areas during evacuation. Therefore, in some instances, variables are off-setting each other in different risk categories. Add to this the mixed distribution of financial recovery and the overall geographical pattern becomes polarised, with both highly urban and highly rural areas being deemed equally at risk. However, as noted in the Mosaic Italy profile of the overall social vulnerability Index, the commonalities in the worst hit areas are *age* and *wealth*. Elderly households that are financially less secure are the main areas of risk in the social vulnerability index. This is reflected in the Mosaic sub groups of both *Elderly Households* and *Urban apartments* that constitute the highest-ranking areas.
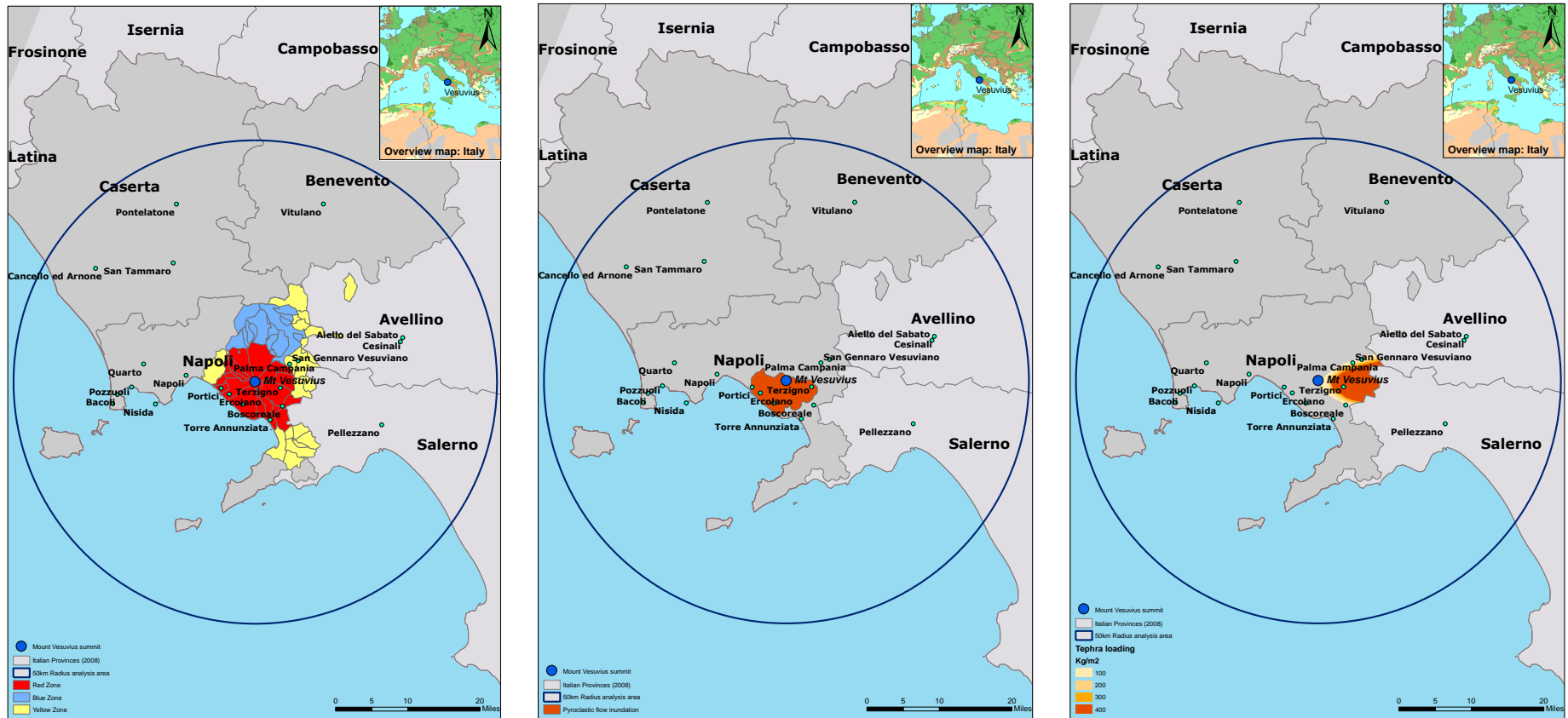
FIGURE 4-6 - (a)-(c): (a) Evacuation Zones around Mount Vesuvius, (b) Pyroclastic flow modelling (c) Tephra loading model

Figure 4-6 (a)–(c) show the physical risk maps for civil evacuation priority areas, pyroclastic flow inundation and volcanic ash fall loading. Using GIS spatial queries for each of these map layers, Figure 4-7 is the culmination of these boundaries assigned to census areas and given a composite physical risk weighting. The highest risk areas defined from this analysis were directly southeast of the volcano. In the event of a sub-Plinian eruption, with tephra fall largely to the East of the cone and pyroclastic flow south of Vesuvius, this region would be hit by a combination of two major geophysical hazards.

FIGURE 4-7 - Physical risk index (composite)

The areas within the analysis region that would likely be least impacted by these hazards are predominantly found directly North of the volcano; a region that is topographically less susceptible to column collapse and tephra fall.
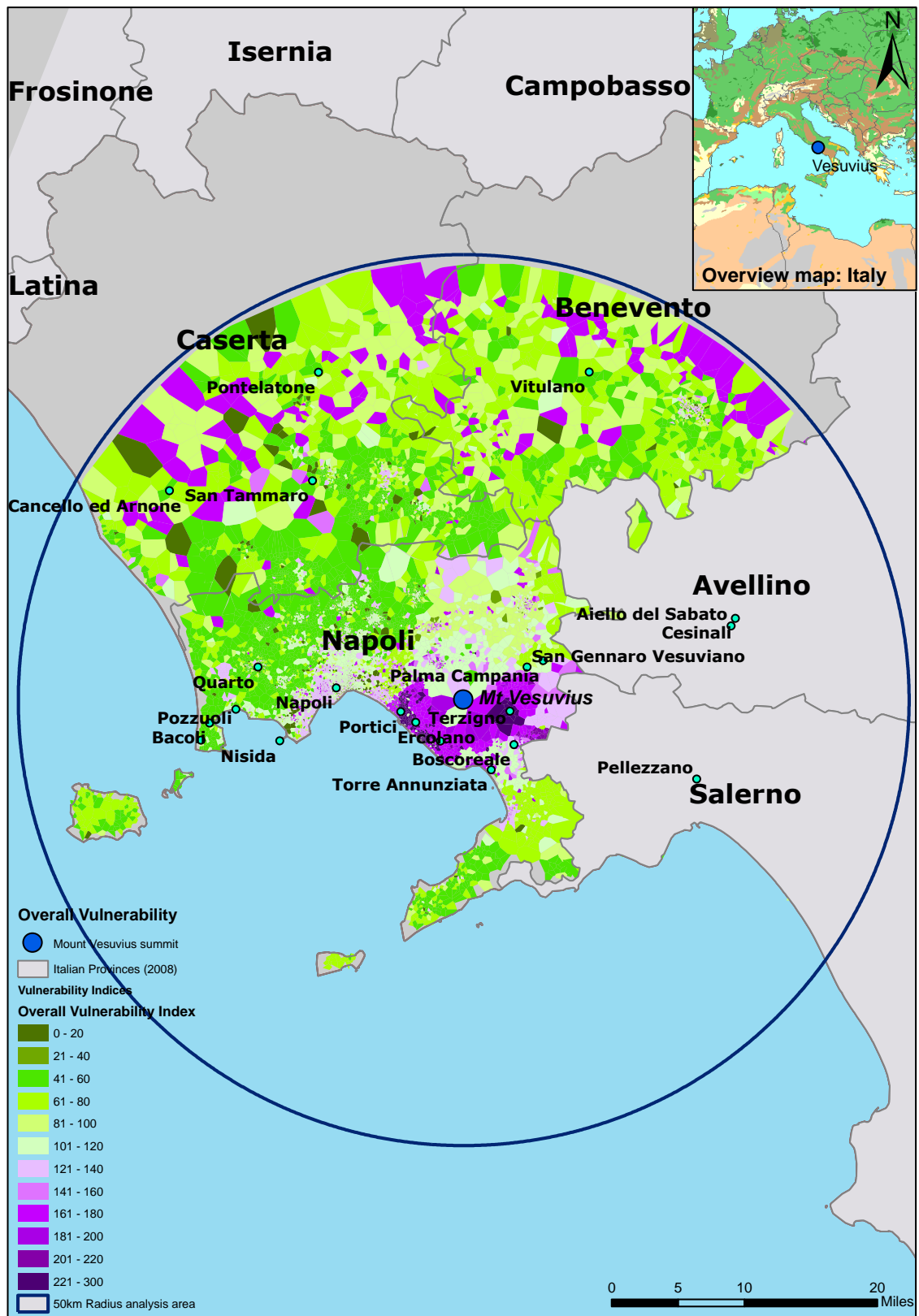
FIGURE 4-8 - OVERALL VULNERABILITY INDEX – MOSAIC ITALY 2007

Having combined the physical risks associated with the onset of a volcanic eruption and the social vulnerability of households, an overall vulnerability classification was assigned to each Census region. In Figure 4-8, the physical risks of an eruption (tephra fall, pyroclastic flow, household risk) largely dominate the overall vulnerability weighting to the analysis region. This can be seen by the geographical patterns of high index scores (140-200) immediately South of the volcano and following the estimated ash fall dispersal directly to the East of the Mount Vesuvius summit. Looking more closely at the high risk areas on the flanks of Vesuvius, the social vulnerability factors start to become more apparent. Within a 5-mile radius of the volcano we see the areas most at risk from Vesuvius. These consist of highly populated coastal towns along this stretch of the Campania region, including *Torre del Greco, Ercolano, San Giorgio a Cremano* and *Portici*.
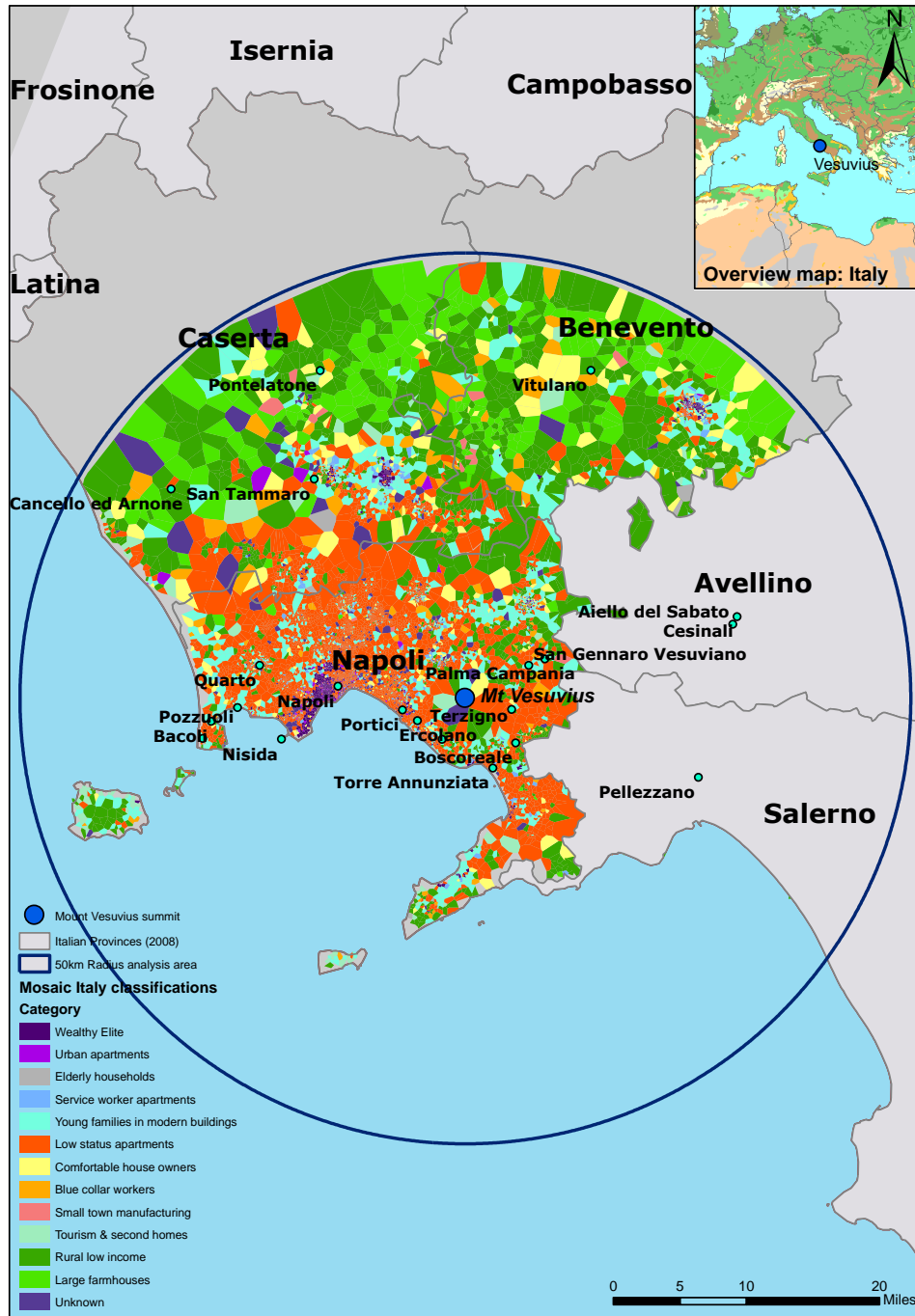
FIGURE 4-9 - Mosaic Italy 2007 geodemographic categories
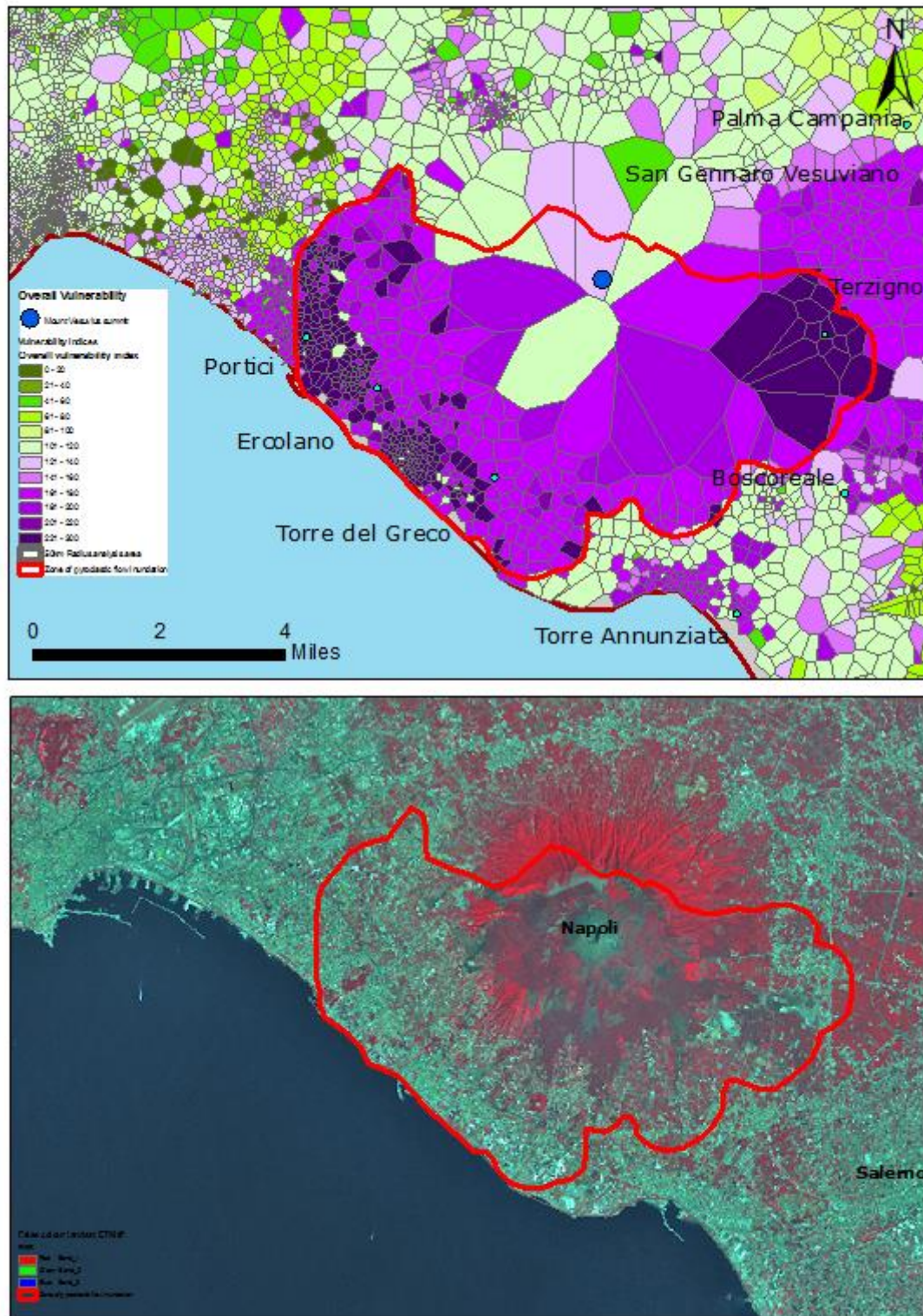
FIGURE 4-10 - Comparison TOP: Overall vulnerability index, Campania, Italy BOTTOM: False colour composite image of the urban infrastructure around Mount Vesuvius (Landsat spectral imagery analysis)

Not only are these towns exposed to substantial volcanogenic hazards, but they are extremely vulnerable to the social consequences of a volcanic eruption. They largely consist of communities with elderly, low-income households in areas of high population density. Figure 4-10 contrasts a false colour composite landsat satellite image with the overall vulnerability index defined in this study. It is interesting to note that the areas of highest risk, as defined by the dark purple census areas within the red line (pyroclastic flow limits) are also the regions of highest population density (as shown in blue-green by the composite image).

4.2   PHASE 2: QUITO, ECUADOR

Quito, Ecuador formed the second phase of this research activity and provided completely different geographical and technical challenges to the previous work on Vesuvius. The principle difference being that no commercial geodemographic classification currently exists for Ecuador, and therefore, one had to be created from first principles. In this chapter, the country-wide classification results and variable indices are presented for the 8 x *tier-1* clusters and 42 x *Tier-2* clusters created during this research. Likewise, profile descriptions are provided along with a series of GIS profile maps to define the neighbourhoods created in the analysis. To assess the viability and application of the geodemographic model to disaster risk reduction, a historic footprint of a 10th century eruption of Guagua Pichincha was recreated and discussed in terms of the population groups 'at risk'. Findings are compared to the city average and results are presented at the 2001 census boundary level using GIS maps. Likewise, in the opening section of this chapter context is provided regarding the landscape, population migration, and cultural heritage of Ecuadorians living close to volcanically active areas. Likewise, a detailed discussion is presented regarding Ecuador's economic history and current status as it helps provide the context in which social vulnerability takes place.

4.2.1 ECUADORIAN CLASSIFICATION SYSTEM MODEL RESULTS

This section provides both a cartographic and tabular representation of the Ecuador classification and vulnerability index created during this research. Model results are presented for two classification tiers using GIS derived maps at a national scale and with inset maps specifically focused on the city of Quito. Quito forms the basis of the case study described previously and results are shown at a census area level using thematic maps. Results are also provided at a national and regional scale using Parroquia centroids to provide point cluster maps and thematic canton maps. Using the same methodology as described in Phase 1 of this research, vulnerability scores have been based on geodemographic index scores relating to a household's *access to resources, evacuation, financial recovery* and an overall social vulnerability index score. GIS maps of the relative results are presented for the metropolitan districts of Quito in Figure 4-13. Likewise, tephra maps from the 1st, 10th and 17th century eruptions are shown to highlight how a contemporary eruption of Guagua Pichincha could impact highly populated parts of the modern day city. These findings are also discussed in more detail in the discussions section.

TABLE 4-1 - ECUADORIAN NEIGHBOURHOOD CLASSIFICATION, CLUSTER MEMBERSHIPS BY TIER

Tier 1 (Group)	1	2	3	4	5	6	7	8
Tier 2 (Sub-group)	17	9	19	22	5	1	7	10
	14	18	26	32	16	2	8	11
	17	20	27	41	21	3	28	13
	25	23	34		33	4		15
	37	24	35			6		36
		29	38					
		30	39					
		31	40					
		42						

Table 4-1 provides an overview of the structure of the geodemographic classification produced at census output level using the 2001 census data for Ecuador. As can be seen, there are 8

groups in Tier 1, with a further 42 sub-groups defined in Tier 2. The table also shows the structure of cluster membership with group 2 having the most sub-groups defined (a total of 8).

Further detail on the Tier 1 groups of the classification is provided in Table 4-2 with a propensity index table of all the variables used in creating the classification. Groups 7 and 8 are highlighted separately as these groups were the only profiles to be found in the Metropolitan Districts of Quito and are pertinent to analysis discussed in section 4.2.3. and the model validation section in 4.3.

Tier 1 Groups		1	2	3	4	5	6	7	8
Type of variable	Variables	Small town average	Afro Caribbean	Farming communities	Indigenous farming	Outlier group	City living	City living (affluent)	Migrant communities
	% Population	**26.48**	**11.39**	**17.97**	**4.00**	**0.05**	**15.54**	**7.39**	**17.19**
Density	Population density	65.55	14.40	8.43	11.47	106.34	623.08	225.49	87.04
Age	0-4	96.94	101.50	114.73	135.56	83.06	86.05	67.03	86.99
	5-14	95.67	107.80	108.49	114.48	60.14	85.31	82.43	91.04
	15-25	104.14	94.27	96.66	93.41	84.72	110.88	103.21	104.74
	24-45	107.82	92.07	85.88	79.90	147.85	116.12	118.68	113.21
	>45	91.92	104.44	100.39	86.45	130.65	98.25	124.89	98.99
	More than 2 children	99.11	108.21	104.15	105.34	41.91	86.83	87.27	93.39
Disability	Disabled_LLTI	100.71	116.30	99.82	109.83	180.75	69.55	64.64	92.52
Educational attainment	Degree	78.40	38.80	15.37	14.57	84.64	226.77	514.60	194.80
	literate	102.72	98.11	93.04	82.29	99.58	108.09	113.09	107.12
Ethnicity	Indigena (e.g. Inca desc)	29.74	39.32	102.63	1109.86	141.68	38.69	15.71	41.10
	Negro (Black)	102.40	209.89	58.66	8.91	66.42	36.65	31.56	46.02
	Mestizo (Mixed race - Native)	101.90	104.90	108.42	19.38	93.46	105.04	90.63	102.22
	Mulato (Mixed race - Black)	147.73	110.83	75.29	9.37	133.99	71.79	79.53	97.92
	Blanco (White)	126.21	69.91	47.07	15.07	82.93	137.18	277.83	147.74
	Other ethnic origin	144.11	76.57	50.00	46.82	1278.55	67.20	154.72	108.85
Household composition	Single person HH	85.27	113.42	107.13	84.16	95.23	90.78	93.22	101.08
	Two person HH	90.44	101.26	98.22	88.41	91.04	111.96	116.05	107.14
	HH of 3 to 5 people	106.33	95.41	89.15	78.40	54.07	117.65	114.82	109.78
	HH of greater than 6	98.97	104.10	121.22	158.11	95.17	61.83	64.46	76.39
Tenure	Owned (Propia)	100.83	112.56	113.97	126.43	20.56	65.34	92.73	70.86
	Leased (Arrendada)	104.27	57.26	25.22	22.50	18.98	244.63	148.02	223.65
	Mortgage (En anticresis)	127.72	93.37	72.06	44.30	13.24	128.89	97.22	120.50
	Public housing (Gratuita)	96.14	103.20	120.03	63.19	153.21	97.69	56.32	93.85
	For services	68.19	75.05	207.96	51.22	638.46	46.04	75.39	52.31
	Other	117.43	98.18	95.78	73.75	16.45	95.28	103.37	98.01
Marrital status	Co-habiting couples(a)	129.98	108.54	109.16	67.41	90.39	47.53	51.15	84.44
	Single	98.26	99.20	100.76	96.90	113.71	101.77	102.61	101.40
	married	86.96	97.34	98.56	121.57	83.55	120.95	114.13	102.64
	divorced	94.01	64.79	44.31	39.85	167.66	202.26	259.17	163.50
	Widower	89.23	112.49	103.48	114.29	46.45	84.91	96.71	97.20
	Separated	140.14	91.66	69.57	49.40	249.78	87.32	106.69	117.61
Employment	Administrative	115.52	53.48	27.06	35.66	131.97	223.65	264.01	165.51
	Agricultural/Fishing	40.39	124.85	209.24	208.26	44.23	7.53	7.57	27.27
	Artisans	127.27	96.75	35.78	91.39	216.10	171.94	71.14	124.67
	Civil Service & Social work	122.43	86.00	49.92	59.84	95.08	156.89	148.89	132.21
	Defence	80.46	33.78	101.54	40.38	945.84	100.89	67.42	257.46
	ManagerialFinancial	77.19	40.79	26.42	18.95	148.13	219.91	553.28	156.55

Manufacturing	134.05	91.49	47.47	59.72	296.45	153.43	95.39	114.50
Miscellaneous	134.62	74.28	26.34	30.74	168.84	172.68	164.03	166.45
Mining / Construction	140.76	88.40	42.06	55.11	414.84	143.44	83.66	125.45
NA	104.82	103.48	99.07	94.44	69.63	93.16	91.58	98.24
Skilled Professional	88.07	48.28	16.83	23.47	73.04	220.23	433.42	188.82

Table 4-3 shows the propensity indices for sub-groups 1, 6, 3, 4, and 7, all found in
Metropolitan Quito.

Propensity indices related to all 42 sub-groups can be found in accompanying thesis data as
well as Appendix 5: Propensity Index – Ecuador Classification (Sub-groups).

TABLE 4-3 - ECUADORIAN CLASSIFICATION INDEX VALUES: TIER 2 – ONLY SHOWING THOSE
CLUSTER GROUPS FOUND IN THE METROPOLITAN DISTRICT OF QUITO

Tier 2 Clusters found in Quito	1	6	3	4	7
Variables	Peripheral barrios 1	Peripheral barrios 2	Working class 1	Working class 2	City Centre
% Population	2	3.1	4.3	4.2	1.6
Population density	580.5	603.3	601.7	592.5	603.3
0-4	97.5	95.4	82.9	75.2	56.1
5-14	92.2	85.4	85.1	80.6	76.2
15-25	113.3	113.7	108.3	106.7	99.2
24-45	108.1	116.4	118.9	121.2	122.2
>45	86	86.5	99.7	112.9	145.6
More than 2 children	92.1	87.5	85	83.4	84.7
Disabled/LLTI	79.1	65.9	74.8	56.9	54.2
Degree	61.3	112.4	243.6	392	803.5
literate	103	105.7	109.1	111.7	116.6
Indigena (e.g. Inca desc)	65.5	36.6	51.3	22.5	18.7
Negro (Black)	65	51.6	31.2	25.2	21.5
Mestizo (Mixed race - Native)	105.2	108.9	103.6	102.1	83.5
Mulato (Mixed race - Black)	89.4	84.2	74.2	59.8	46.7
Blanco (White)	99.8	97.8	139.9	183.4	348.3
Other ethnic origin	62.3	56.5	71.1	74	192.8
Single person HH	64.1	77.9	125.3	91.6	154.6
Two person HH	88.7	103.9	131.1	119.4	163.6
HH of 3 to 5 people	116.6	120.7	111.9	121.9	105.6
HH of greater than 6	87.5	65.2	49.3	48.6	32.8
Owned (Propia)	86.8	60.9	44.6	72.4	76.9
Leased (Arrendada)	154	260.9	326.9	222.3	209.4

Mortgage (En anticresis)	90	103.3	131.3	166.4	114.7
Public housing (Gratuita)	123.5	105	92.7	76.2	52.2
For services	46.6	34.9	40.4	57.3	85.5
Other	116.2	115.2	87.3	81.2	72.9
Co-habiting couples(a)	56.7	61.5	53.9	36	21.9
Single	101.2	98.9	101.5	101.5	101.9
Married	121	120.1	115.1	124.8	122.7
Divorced	123.6	149	235.8	257.3	428
Widower	77.1	75.4	96.5	84.4	111.6
Separated	83.6	90.3	101.4	86.3	84.6
Administrative	133.5	191.6	245	289.6	325
Agricultural/Fishing	15.6	5.9	4.8	4.6	4.4
Artisans	219	177.4	211.3	115.5	52.1
Civil Service & Social work	159.7	178.5	150.6	156.1	147.6
Defense	57.8	97.3	88	104.6	51.3
Managerial/Financial	67.5	124.4	217.3	408.8	1003.1
Manufacturing	198.2	180.4	143.8	102.5	40.6
Miscellaneous	140	162.3	202.2	176.1	161.9
Mining / Construction	214.1	175.3	122.6	103.2	48.7
Skilled Professional	63.1	122.3	244	357.9	642.4

4.2.2   PROFILE PORTRAITS FOR CENSUS-AREA LEVEL CLASSIFICATION OF ECUADOR

It is a common practice in the production of geodemographic classifications that as well as providing a quantitative measurement of a cluster profile, a narrative description of a neighbourhood is provided. Essentially, these terse descriptions help translate the aggregated statistics of these neighbourhoods into an accessible portrait of a particular profile. Although somewhat subjective, provided below is a highlight portrait created for the eight tier 1 neighbourhood types identified during cluster analysis. Likewise, the names given to these cluster groups in tier 1 are for representative purposes only and although not describing every individual, they give an overview to help identify broad Ecuadorian neighbourhood traits.

TABLE 4-4 - ECUADORIAN CLASSIFICATION DESCRIPTIONS

Demographic profile description	Maps (Profile count/Canton)

1: Small town average

These communities are found in various areas across Ecuador and account for the greatest collective population of any profile group (approx. 27% of the total). The largest proportion of this group is located in the Sierra (mountains) and Costa (coastal lowlands) regions of the country and population density in these areas is typically a little lower than the national average (index score of 65). Most demographic and household tenure census statistics for this profile are very close to the national average although it is noticeable that this group has a very low indigenous population (index score of 29) and employment tends to be in mining or administrative work. These communities are found in small towns across the country.

2: Afro Caribbean

Making up 11% of the population, this neighbourhood type is characterised by the fact that the black population within these communities is 2 x the national average and there are strong historical and cultural roots behind such Ecuadorian neighbourhoods. Costa provinces such as Esmeraldas have historically had a large black community due to post-colonial emancipation and subsequent colonisation of the North-West coastal regions of Ecuador. Livelihoods are largely focused around agricultural/fishing (index score of 125) and most own their home outright rather than renting/mortgaging.

3: Farming communities

As the name suggests, this neighbourhood type has a strong agricultural connection. With the lowest population density in Ecuador (index score of 8) and making up 18% of all households, employment centres around agricultural work. Ethnicity comprises mainly of Mestizo, Indigenous and Mulato, with a below average representation of White families (index score of 47).

4: Indigenous farming

These communities are found predominantly in the Amazonas and Sierra regions of Ecuador. The defining characteristic of these areas is their Indigenous heritage, typically from Inca or tribal communities, such as those found in the Amazon (Index score of 1110). All other ethnic denominations are significantly below the national average in this profile. Work is usually agricultural or artisan and these neighbourhoods have the lowest degree attainment rate of all tier 1 Ecuadorian groups (Index score of 15). The demographic variables show families are more likely to have young children (Index score of 136) and family sizes (> 5 children) are above average (Index score of 158).
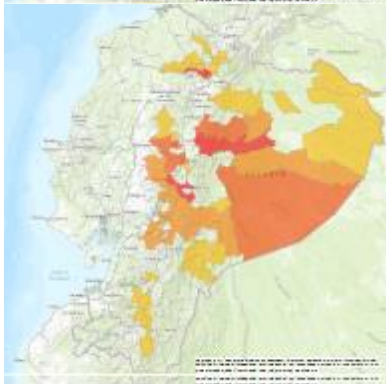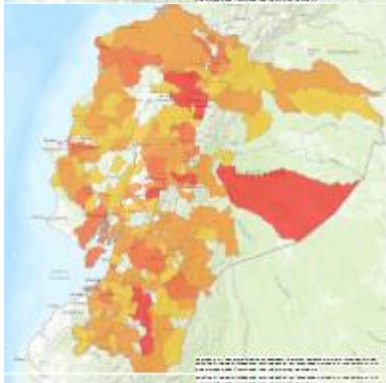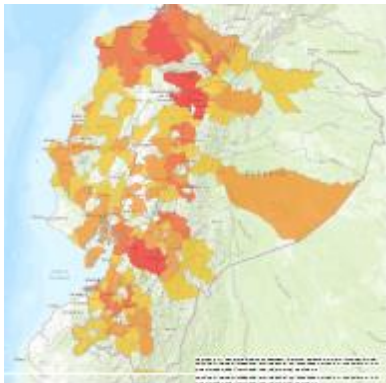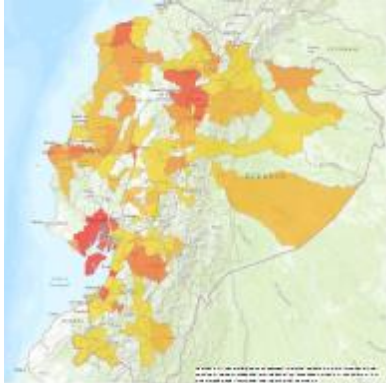
5: Outliers (oil/mining/army towns)

This profile is something of an outlier group, comprising of just 0.5% of the population. The combination of variables such as employment in mining being 4 x national average and jobs in defence being 9 x average, together with a demographic comprised of minorities from outside Ecuador (index score of 1279) implies these areas are common to migrant workers. This includes mining/oil towns and those living in military accommodation. Further evidence for this is provided by an above average level of single marital status, the high propensity of public housing (gratuita), the relatively high educational attainment, and low representation for the age group 1-15. Together with the often remote location of these census areas in places such as the Amazonas suggest a strong connection with mining/oil towns on the fringes of the Amazon basin (e.g. Coca).
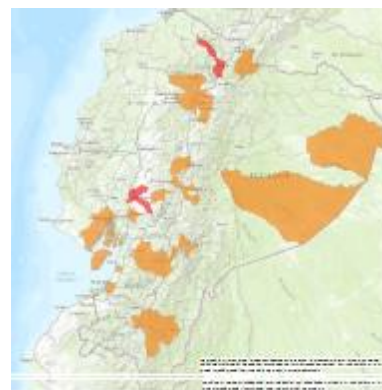
6: City living

Constituting 16% of the entire population and with a population density 6 x national average, these areas are found exclusively in the large cities such as Quito, Cuenca and Guayaquil, the principal cities of Ecuador. Ethnicity is predominantly white or mestizo, educational attainment is more than 2 x national average, and employment is high in skilled professions, artisans, as well as manufacturing and mining. Many of the tier 2 classification found in this group could be regarded as the working class suburbs of the large cities.
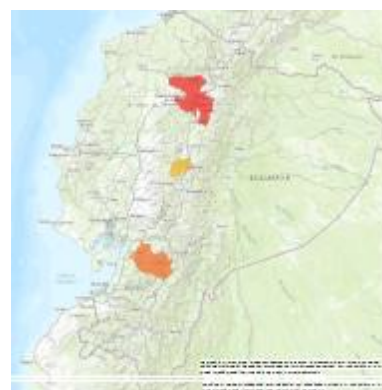
7: Affluent city living

Similar to the 'City living' profile with the key distinction of these neighbourhoods being economically defined. Located in the very heart of large cities, such as the historic quarters of Quito, these areas show a population with 5 x national average of degree attainment and people working in financial/managerial roles. Interestingly, the proportion of those classed as divorced is also highest among all profiles, being more than 2 x national average. This profile has the lowest proportion of black and indigenous people (Index scores of 32 and 16 respectively) and the largest proportion of white people (Index score of 278)
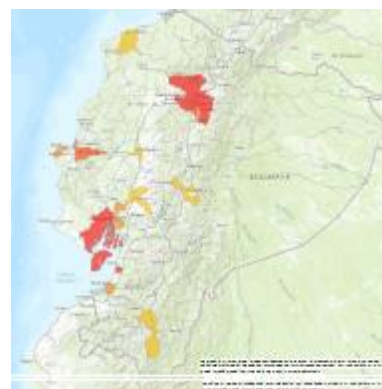
8: Migrant communities

With an urban profile but found most commonly across the Sierra, Costa and some regions of the Amazonas, these neighbourhoods could be associated with small Andean towns. Ethnicity is usually white, mestizo or from outside Ecuador with most opting to rent their home rather than buy or mortgage. Degree attainment is approximately 2 x the national average, as is those working in skilled professional employment.
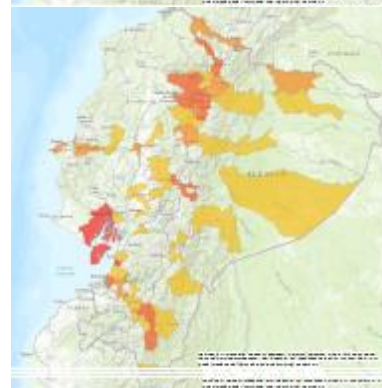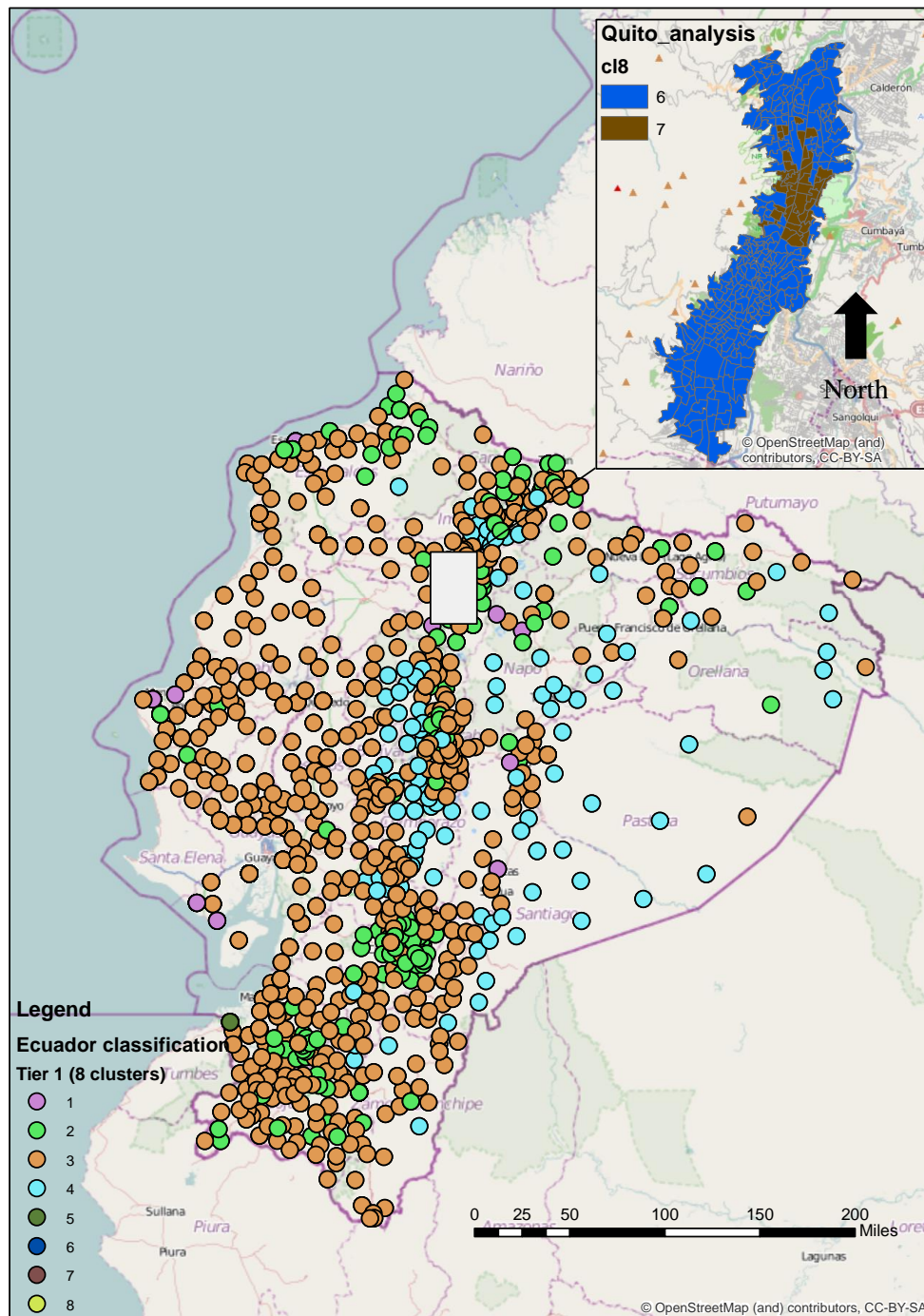
FIGURE 4-11 - Ecuador neighbourhood classification mapped to the centroid of the Parroquia (inset Quito): Tier 1: 8 clusters
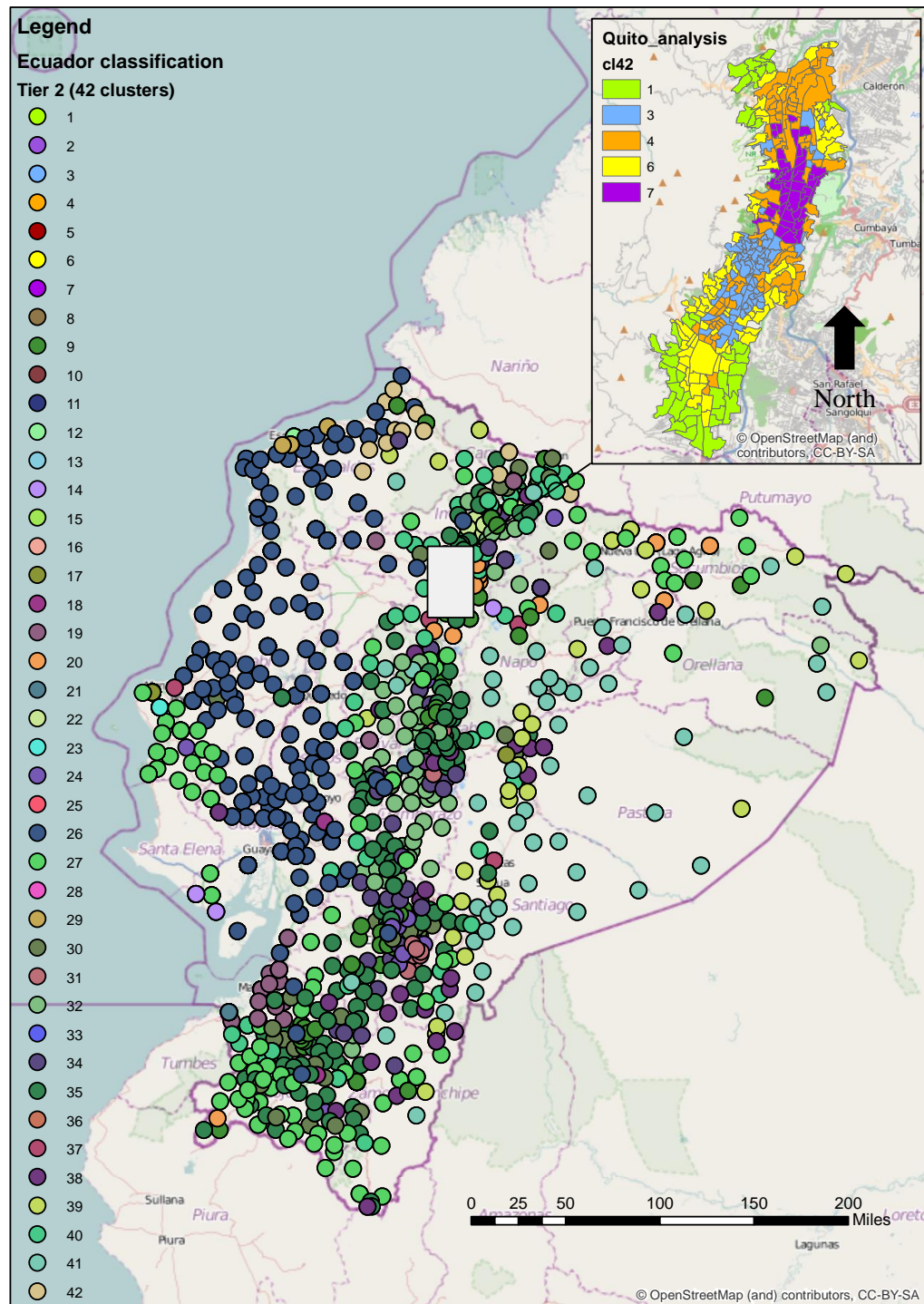
FIGURE 4-12 - ECUADOR NEIGHBOURHOOD CLASSIFICATION MAPPED TO THE CENTROID OF THE PARROQUIA (INSET QUITO): TIER 2: 42 CLUSTERS
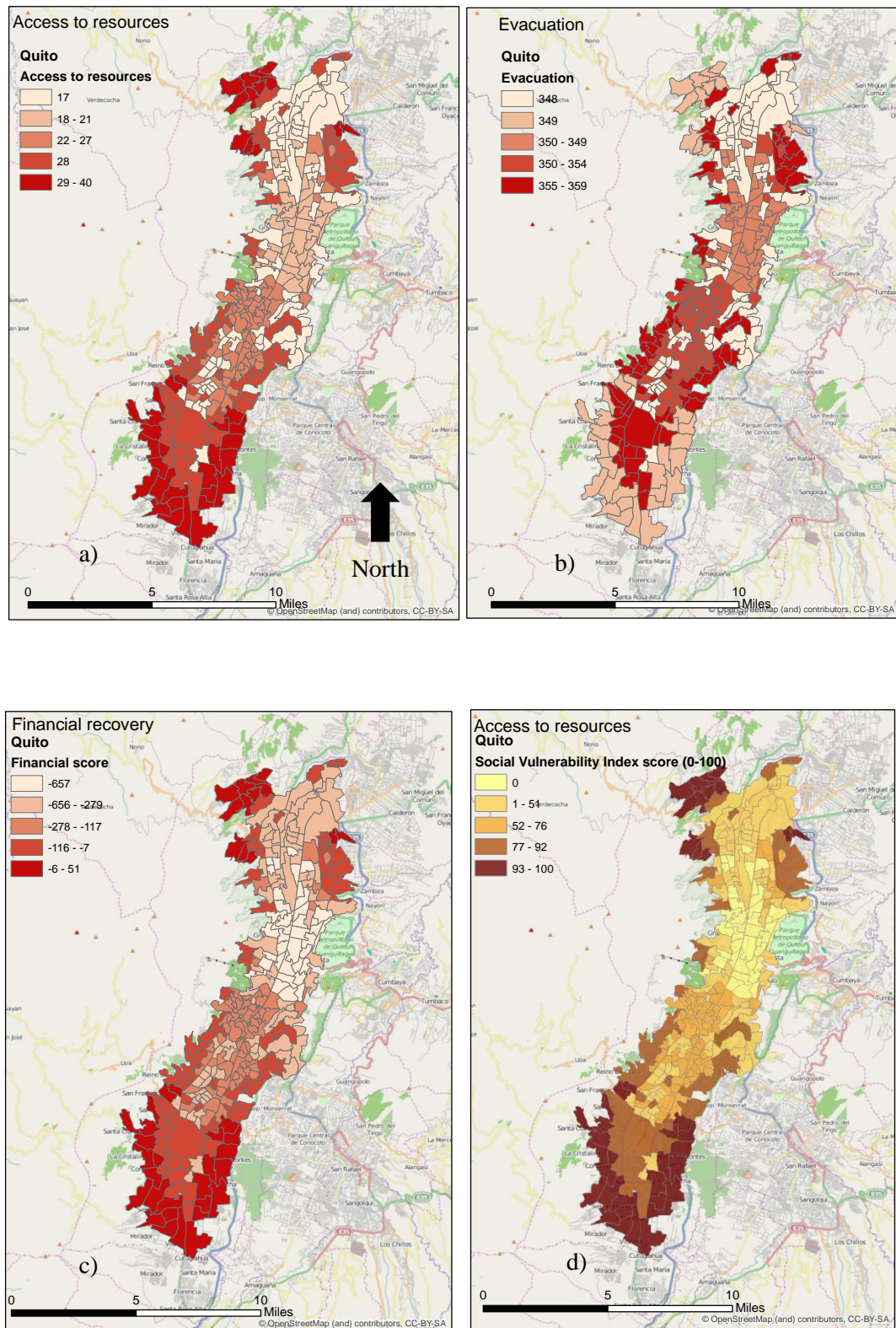
FIGURE 4-13 – QUITO SOCIAL VULNERABILITY MAPS SHOWN BY INEC 2001 CENSUS ZONES (A) ACCESS TO RESOURCES (B) EVACUATION (C) FINANCIAL RECOVERY (D) SOCIAL VULNERABILITY INDEX
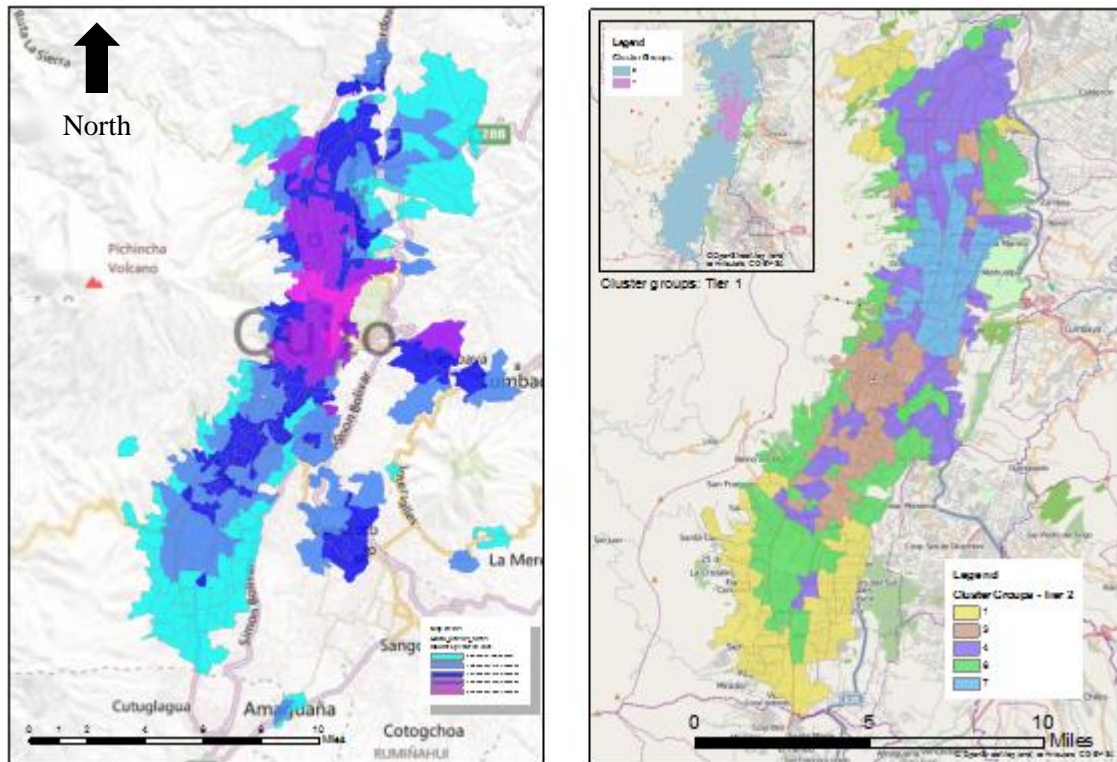
FIGURE 4-14 – LEFT: AVERAGE CONSUMPTION, USD/MONTH, 2004-2006 BY CENSUS ZONE RIGHT: ECUADOR GEODEMOGRAPHIC CLASSIFICATION BY 2001 CENSUS AREAS

4.2.3 GUAGUA PICHINCHA ERUPTION CASE STUDY

After creating the Ecuador geodemographic clusters, it was decided that it would be important to apply the classification in a realistic volcanic disaster scenario. In this section, analysis and findings results are presented for a case study analysis of a Plinian eruption of Guagua Pichincha, an active volcano very close to the capital city of Quito.

Though Guagua Pichincha is currently in a state of repose (at the time of writing), significant eruptive behavior has been chronicled over the last 2 millennia (Robin et al 2008), and provides evidence that previous volcanic events have impacted large regions of the city that are now densely populated. Given the scenario of a large Plinian eruption of Guagua Pichincha in the near future, many of the communities in harm's way of volcanic hazards (tephra, volcanogenic earthquakes and lahars) are also among the most disenfranchised socio-economic groups. Many

of these houses lack the most basic of amenities, such as drinking water or sewerage (Quito planning office 2002). With a significant portion of the population being informal housing settlements (barrios) established on the flanks of the volcano, they are very much a consequence of poor land regulation and the rapid urban migration that has characterized numerous South American countries over the last 50 years (Carrion et al 2003).

Quito volcanic hazard

The oblique subduction of the Nazca plate with respect to the South American plate at the latitudes of Ecuador occurs at a rate of 50–70 mm/ year (DeMets et al., 1990; Kellogg and Vega, 1995). This process gives rise to the North Andean Volcanic Zone (NAVZ), whereby the Ecuadorian mountain chain essentially forms an elbow between the Peruvian Cordillera and the Colombian Cordillera (Legrand et al 2002). It is understood that the mountain chain induces high stress and deformation fields thus responsible for the high seismic and volcanic activity in Ecuador, particularly when compared to Peru and Colombia.

The Pichincha volcanic complex, located within the NAVZ, consists of the older Rucu Pichincha edifice (approximately 1 Ma years old) which essentially shields the city of Quito from the younger Guagua Pichincha (Barberi et al 1992). Although the volcano has seen long periods of both quiescence and episodic phreatic activity in the 19th century, the most significant eruptive activity is largely evidenced by events in the Late Holocene phase (Robin et al 2008). This period was characterized by three large Plinian events, in the 1st, 10th and 17th centuries. The 10th century eruption was the largest of these events and is estimated as a VEI 5 event.

Previous eruptions have included sizable pyroclastic flows, lahars, volcanogenic earthquakes, and ballistic impacts from the volcano. However, for the purposes of this study, historic tephra fall maps have been used as the basis of the principal volcanic hazard impacting human population in the city. Though several volcanic hazards continue to pose a large threat to the

surrounding area, historic evidence suggests tephra fall would be likely to directly impact populated areas of the city.

Given the natural barrier afforded by the older edifice of Rucu Pichincha, pyroclastic flows from previous eruptions have historically been orientated South-West of the volcano, and more critically, away from the populated provinces of the city (Barberi et al, 1992). Earthquake swarms believed to be triggered by reactivation of the volcano during the 1999 eruption (Legrand et al 2002) affected the north of the city but due to the very small nature of the magnitude of these events, they posed only a minor threat to housing stock and human settlement. Topographic analysis using GIS indicates that lahars generated from a large eruption could reach the populated flanks of the volcano (Canuti et al 2002) but the impact of this hazard was not be included in the proceeding assessment.

Tephra loading

Based on the assumptions of building vulnerability, and in consideration of the 10th century Plinian eruption of Pichincha, the spatial footprints of two different tephra thresholds were considered here in analysing the impacted population groups. Relative ash fall thickness' of 100mm and 150mm are considered here as being the critical thresholds of vulnerability, although it's important to note that building stock and structural performance analysis with specific regard to tephra loading has not previously been undertaken in Quito.

Assumptions concerning vulnerability thresholds were based on association to structural performance during the Rabaul and Mount Pinatubo eruptions. Based on previous research by Robin et al (2008) in mapping previous tephra accumulation levels, digitized boundaries for the 10th Century eruption of Guagua Pichincha were re-created using a GIS. Based on the spatial footprint of the tephra fall of the 10th Century Plinian eruption, census zones in the metropolitan district of Quito were identified that were found to lie within the impacted accumulation areas. The extent of these spatial areas are highlighted in Figure 4-15.
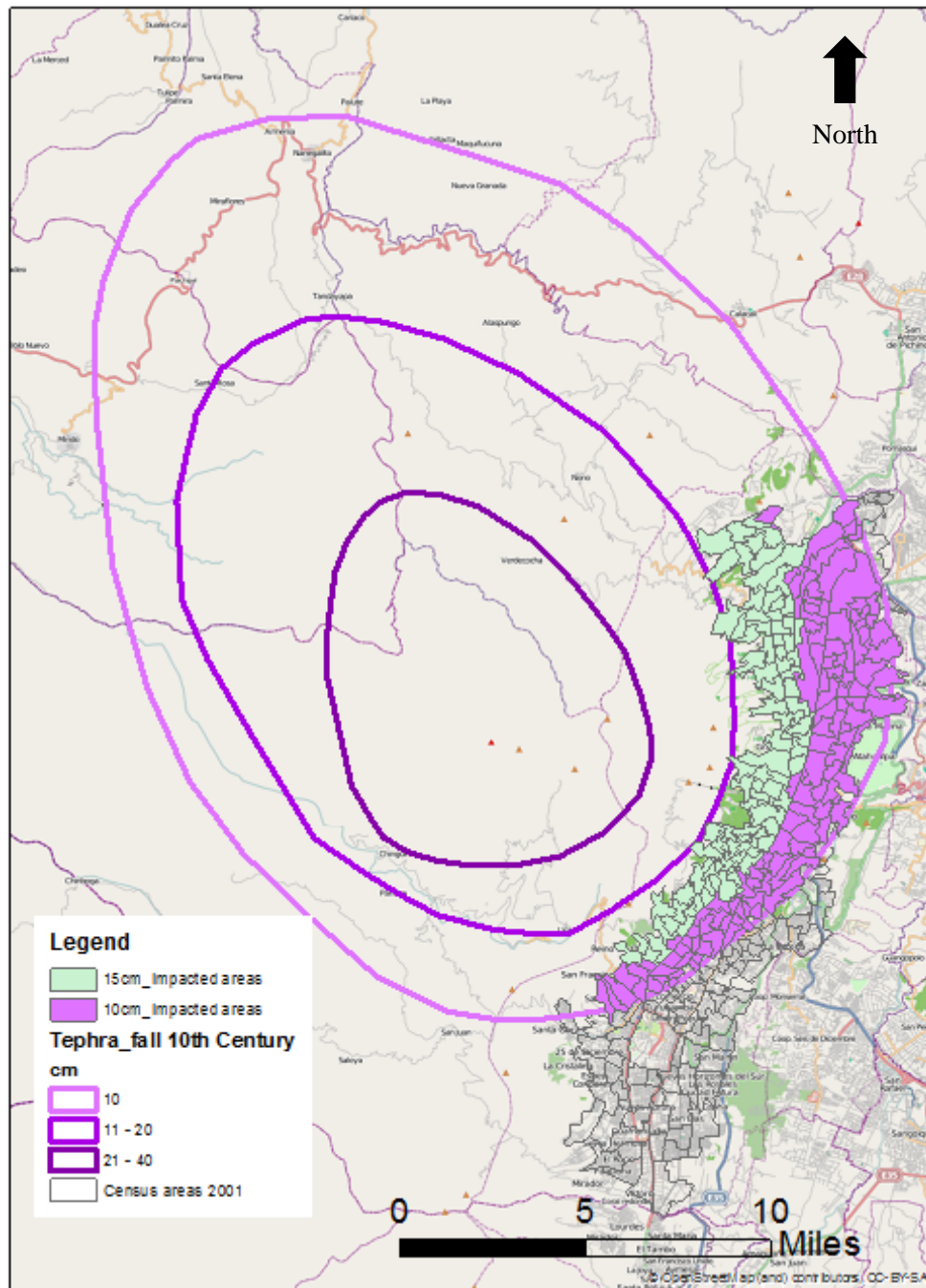
FIGURE 4-15 - DIGITISED BOUNDARY OF ACCUMULATION LEVELS FOR THE 10TH CENTURY TEPHRA FALL

After identifying affected population groups, statistical analysis was then undertaken,

comparing the demographics of the populations groups impacted at different levels of tephra

accumulation with the city average for the Metropolitan district of Quito. Of key interest during

this spatial analysis was to examine the hypothesis that less affluent, peripheral barrios may be disproportionately vulnerable during a large volcanic eruption.
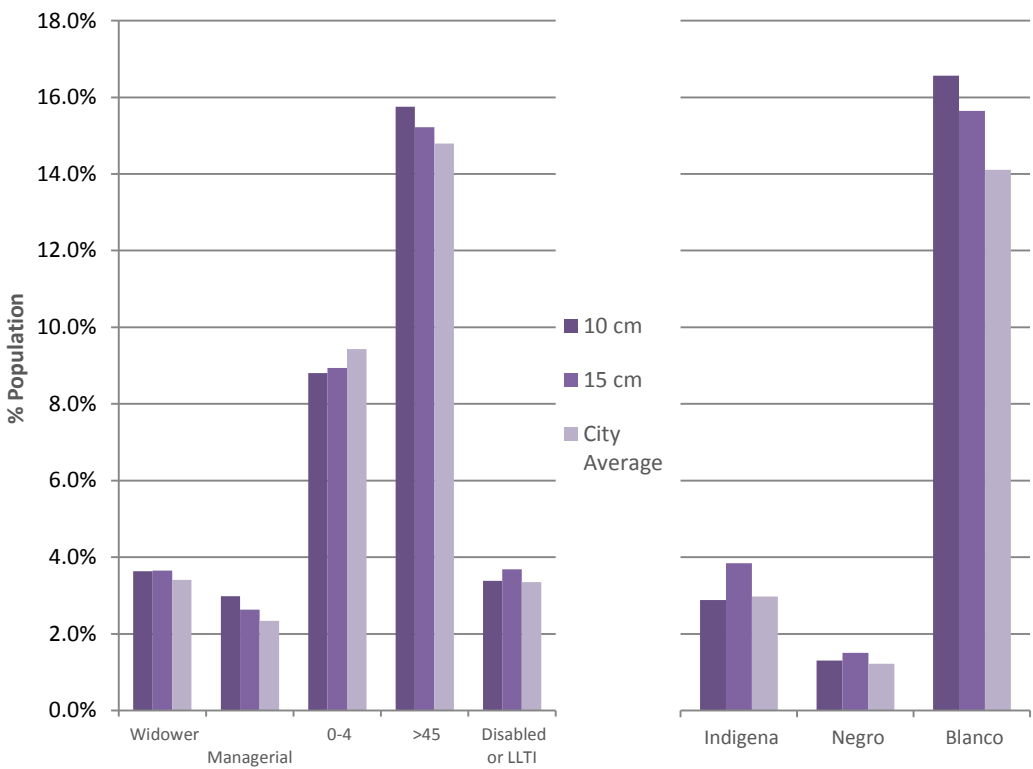
FIGURE 4-16 – COMPARISON OF VARIOUS SOCIAL VULNERABILITY INDICATORS WITHIN THE IMPACTED POPULATIONS COMPARED TO THE CITY AVERAGE (CENSUS 2001)

Figure 4-16 highlights the findings when comparing the city average to the various social vulnerability indicators in the impacted population groups. Due to the footprint of the 10th Century tephra fall disproportionately impacting the more affluent sections of Quito, there is a general trend of many social vulnerability indicators decreasing in the affected population group. For example, *Managerial employment*, typically associated with a reduced vulnerability, increases proportionately at both the 100mm and 150mm levels. Likewise, the proportion of those *aged 0-4 years* shows a proportional decrease in the tephra fall populations. Perhaps the most significant evidence of this trend regards the populations of those with *white ethnicity*.

This ethnic variable is highly correlated with educational attainment, employment type and financial resilience. As show in Figure 4-16, there is marked increase of this variable in the impacted populations.
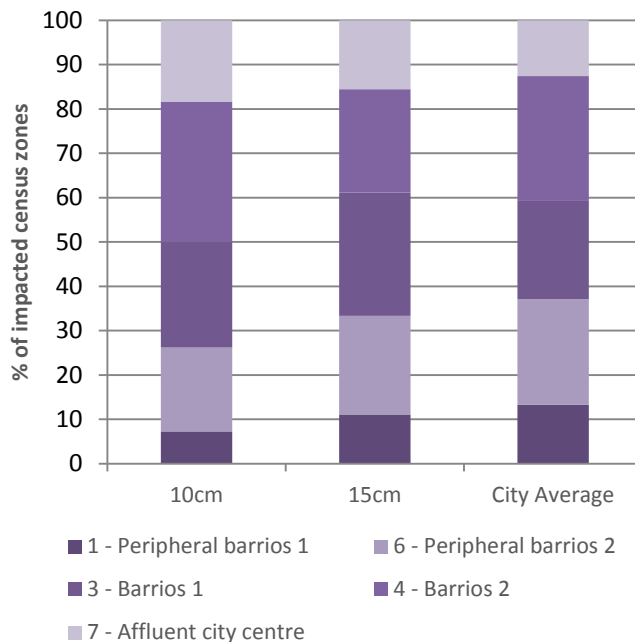
FIGURE 4-17 – COMPARISON OF THE TIER 2 GEODEMOGRAPHIC CLASSIFICATIONS AT DIFFERENT TEPHRA ACCUMULATION LEVELS

Figure 4-17 shows the geodemographic profiles found within these areas compared and contrasted at different accumulations levels of tephra (100mm and 150mm respectively) with the Quito city average. This analysis was undertaken to establish whether the tephra hazard would disproportionately affect specific neighbourhood profile groups.

Findings from this research show that the *affluent city centre* profile appears over represented in the 'at risk' population by a magnitude of 6% given a 10cm ash fall threshold. Conversely, and perhaps more significantly, Figure 4-17 shows that the *peripheral barrios* profile surrounding the city edges of Quito's metropolitan area has a 5% decrease in their proportional representation at this same threshold. There would also be a 5% increase in the highly populous conventillos suburbs of the old quarter of the city.

These inner city barrios of Quito appear to account for 57% of the population at risk of 100mm tephra threshold, with an estimated 149,400 households affected. At the more critical 15cm threshold, where structural damage and the risk of collapse to roofs has been significant in previous volcanic eruptions (Blong and McKee 1995), the total population at risk comprises of 111,700 households. This corresponds to approximately 30% of the Quito metropolitan population.

These findings imply that a tephra impact of the same orientation and magnitude as the Plinian event evidenced in the 10th Century would not disproportionately affect the peripheral barrios of Quito compared to other neighbourhood profiles. This assertion appears primarily due to the South-East orientation of the 10th Century tephra fall, and the fact that Quito suburbs are largely constrained by the steep topography of the surrounding volcanoes. Similarly, the eruption scenario avoided the large peripheral barrios conurbations to the South of the city.

Further discussion on this chapter can be found in the chapter 5 but such results are inherently reliant on the tectonic scenario footprint of the 10th Century eruption (Figure 4-15). For example, any sensitivity analysis around the key controls of tephra fall, such as prevailing wind speed, wind direction or alternative VEI scenarios would duly warrant very different results. It is recommended that probabilistic analysis around these thresholds could provide a more accurate picture of tephra vulnerability in Quito.

4.3   MODEL VALIDATION

This thesis highlights the application of geodemographics to DRR through the use of neighbourhood profiling indices and subsequent social vulnerability risk metrics. The additive model methodology presented within this framework (for both Italy and Ecuador) shows how relevant vulnerability variables can be accounted for whilst factoring in the discriminatory ability of a given variable based on the corresponding Gini-Coefficient. Given the novel application of this procedure, in developing this research it has been necessary to test and

ultimately validate the assumptions of this application and methodology with alternative methods also aimed at deriving risk metrics.

Therefore, to help assess the effectiveness and limitations of this method as a technique for identifying socially vulnerable communities to volcanic risk, it was decided that a process of model validation should be undertaken. Two divergent tests were identified:

1. *Aggregated areal unit test* – A comparative analysis of the use of vulnerability metrics based on original census area information contrasted against data derived from aggregated statistical units such as the geodemographic clusters created in this research.

2. *Alternative risk methodology test* – A comparison of the Gini-coefficient based methodology presented in this thesis compared to an alternative, widely published methodology in the literature focused on deriving single social vulnerability risk metrics at the same areal extent and using the same variables.

It should be noted that model validation was not undertaken for the Italy Mosaic data as the information required to perform such scrutiny on a comparative areal scale would not be possible. The Mosaic Italy 2007 data used in developing this thesis does not show original census data but instead pertains to the Mosaic indices produced after clustering. Therefore, it was not possible to create an independent vulnerability index using census information and then compare this with the Mosaic clusters.

However, validation tests were undertaken for the Ecuador study region as the census data was available at the output area level as well as the geodemographic clusters constructed during this research. Where *Test 1* was focused around assessment of whether a limited number of cluster groups can adequately represent social vulnerability variation at a smaller level of geography, *Test 2* required comparison to an alternative but widely published vulnerability metric methodology. For the purposes of this latter examination, it was decided to compare social vulnerability in Quito using both Susan Cutter's Principal Component Analysis SoVI methodology (1996) against the Gini-Coefficient methodology proposed earlier in this thesis

and published by Willis et al (2010). To focus the resultant application of this work to volcanic risk and the wider DRR community, the following results are presented for Quito metropolitan district census areas (2001).

Test 1 - Aggregated areal unit test

In the Quito study region, a comparison was made between the Tier 2 geodemographic clusters created from the 2001 census data and calculating a social vulnerability risk score for the same census areas based on the following variables associated with social vulnerability:

- Age 0-4 (+)
- Age >45(+)
- Low degree attainment (-)
- Ethnicity (Negro) (+)
- Households greater than 6 people (+)
- Widower (+)

Note: The (+) and (-) denote whether the variable cardinality and whether it is perceived in the literature to increase or decrease perceived social vulnerability.

Social vulnerability scores were then applied using the methodology and equations outlined in section 3.5 whereby the Gini-Coefficient of each variable acts as a multiplier before the cumulative risk score is normalized using z-scores (resulting in a value between 0-1). Each of the 304 observations represents a census area within the metropolitan district of Quito. The relative risk scores within each of the five Tier 2 cluster groups found in Quito are presented below in Figure 4-18.
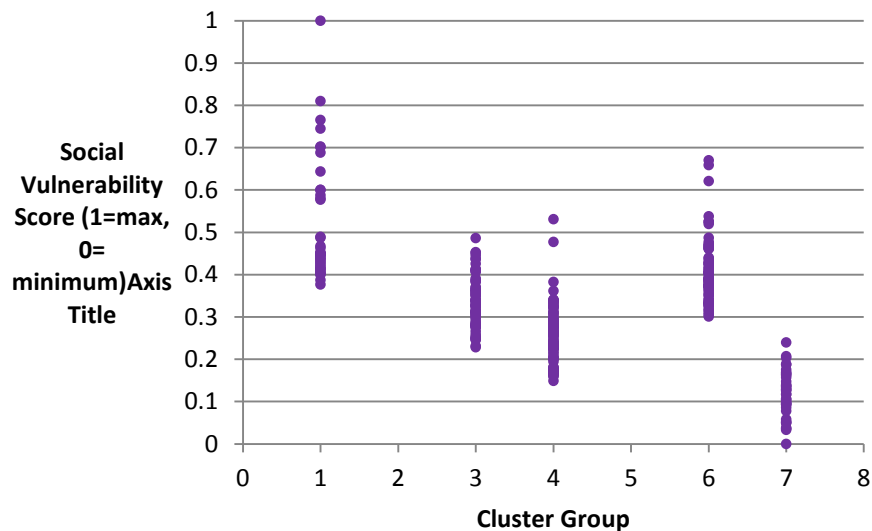
FIGURE 4-18 – COMPARISON OF CLUSTER GROUPS FOUND IN QUITO AND THEIR ASSOCIATED SOCIAL VULNERABILITY SCORES

As can be seen in Figure 4-18, where the Y axis represents social vulnerability on a scale of 0-1 (1 = maximum social vulnerability) there appears to be an interesting relationship to social vulnerability within each of the Tier 2 cluster groups found in Quito. This can be evidenced by the distribution of vulnerability scores within each cluster and in comparison between clusters. Taking the mean score of each Tier 2 cluster group only, groups 1 has an average social vulnerability score of 0.52 and group 7 had the lowest mean with a score of 0.12. This finding alone implies that social vulnerability in cluster group 7 is lower than in group 1. However, whilst some groups show a higher average score relative to other groups, this does not account for the uncertainty in the distribution of scores within a cluster. For example, Figure 4-18 highlights that the minimum score in group 4 (0.15) is lower than the maximum score in group 7, (0.24), yet on average, group 7 has a greater social vulnerability score.

This helps show the large amount of uncertainty. This is further highlighted in Figure 4-19.
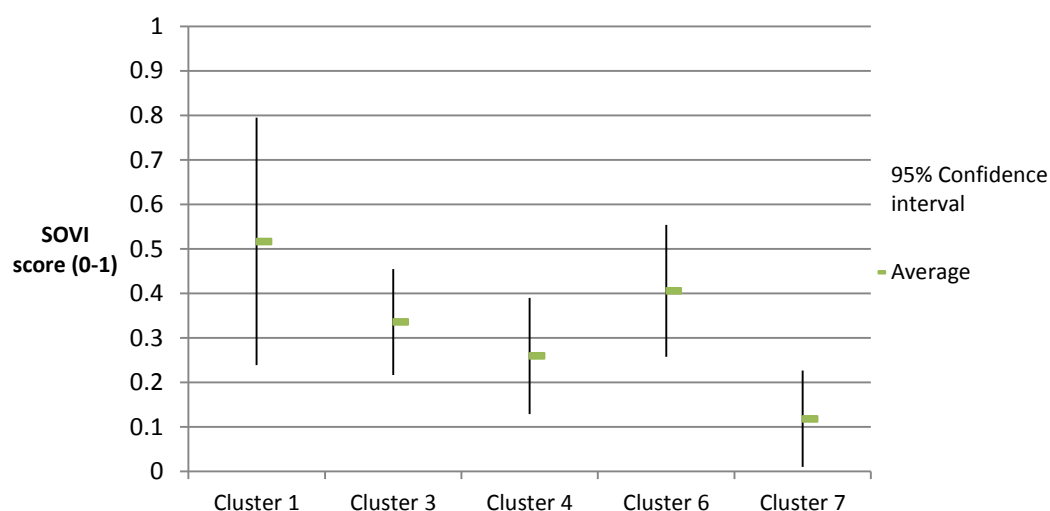
FIGURE 4-19 – VULNERABILITY UNCERTAINTY WITHIN DEFINED CLUSTERS, QUITO, ECUADOR

Figure 4-18 showed the discrete distribution of vulnerability distribution within a given classification. This can be seen more clearly in Figure 4-19, which shows the uncertainty boundaries in terms of the standard deviation for each cluster group in defining vulnerability. This is evidenced by considering the standard deviations of each cluster. Cluster 1 has the highest standard deviation (0.14) and cluster 7 the lowest (0.06). A Chi Square test was also undertaken to assess the statistical significance of the average vulnerability scores for each cluster group, compared to the expected values. With a probability value of 0.97, these findings could not be considered to be statistically significant. Although the averages for the Quito cluster groups appear sufficiently independent, there is substantial overlap between many of the cluster groups when applying a 95% confidence interval from the mean. Figure 4-19 reaffirms that although the neighbourhood groups appear to be correlated with social vulnerability scores, they do not allow for a binary assessment of vulnerability as there remains substantial variance regarding vulnerability within a single aggregated classification groups.

Test 2 - Alternative methodology test

To further explore the validity and assumptions of the Gini-Coefficient weighting methodology, it was decided to contrast this technique against an alternative and widely published social vulnerability index methodology devised by Susan Cutter (1996). Cutter's hazard-of-place conceptual model (1996) aimed an objective appraisal of vulnerability by first identifying variables that are associated with social vulnerability and then classifying them through factor analysis in an additive process. This research culminated in the publication of Cutter's SoVI methodology (2006), providing a guide to how fellow researchers and practitioners could replicate Cutter's technique.

Cutter's '*SoVI recipe*' (2006) recommends the following iterative steps:

1. Selected relevant variables associated with social vulnerability for inclusion in the model.

2. Standardise the variable input data to be between 0-1 (e.g. using z-scores, range standardisation).

3. Run PCA (using a Varimax rotation) to determine the number of relevant components and cardinality of vectors, making ensure that all variables are highly correlated.

4. Remove any variables that are not correlated with the principal components, or alternatively, inverse any negatively correlated variables so that they are positively correlated.

5. Extract the component scores for each observation using the Varimax rotation and use this data as an additional input variable.

6. Sum all the input variables and produce and index.

Using the same six variables as previously used in *Test 1*, the SoVI recipe steps outlined above were followed to produce a vulnerability index.

Prior to the PCA assessment, Table 4-5 shows the Pearson's correlation matrix produced for the six variables. As highlighted in red, there appears to be greater positive correlation between variables such as 'Age 0-4' and 'Households >6' and high negative correlation between 'Age>45' and 'Age0-4'.

TABLE 4-5 – PEARSON'S CORRELATION MATRIX, ECUADOR CENSUS VARIABLES (2001) (POSITIVE/NEGATIVE CORRELATION VALUES >0.5 ARE HIGHLIGHTED)

Variables	Age 0-4	Age >45	Degree	Ethnicity-Black	Household>6	Widower
Age0-4	N/A	-0.51	-0.49	0.15	0.52	-0.16
Age>45	-0.51	N/A	0.29	-0.10	-0.40	0.44
Degree	-0.49	0.29	N/A	-0.09	-0.47	-0.01
Ethnicity-Black	0.15	-0.10	-0.09	N/A	0.13	-0.07
Household>6	0.52	-0.40	-0.47	0.13	N/A	-0.18
Widower	-0.16	0.44	-0.01	-0.07	-0.18	N/A

In running the PCA, the underling variance and principal cardinality of the test data could be seen more clearly. Figure 4-20 shows variables in relation to the first two underlying principal components by using a biplot. This figure illustrates how variables 'Ethnicity-Black', 'Household>6' and 'Age 0-4' are closely aligned whereas variables 'Widower' and 'Age >45' are not as correlated. 'Degree' is polarised as an inverse variable to the principally correlated vulnerability vectors.
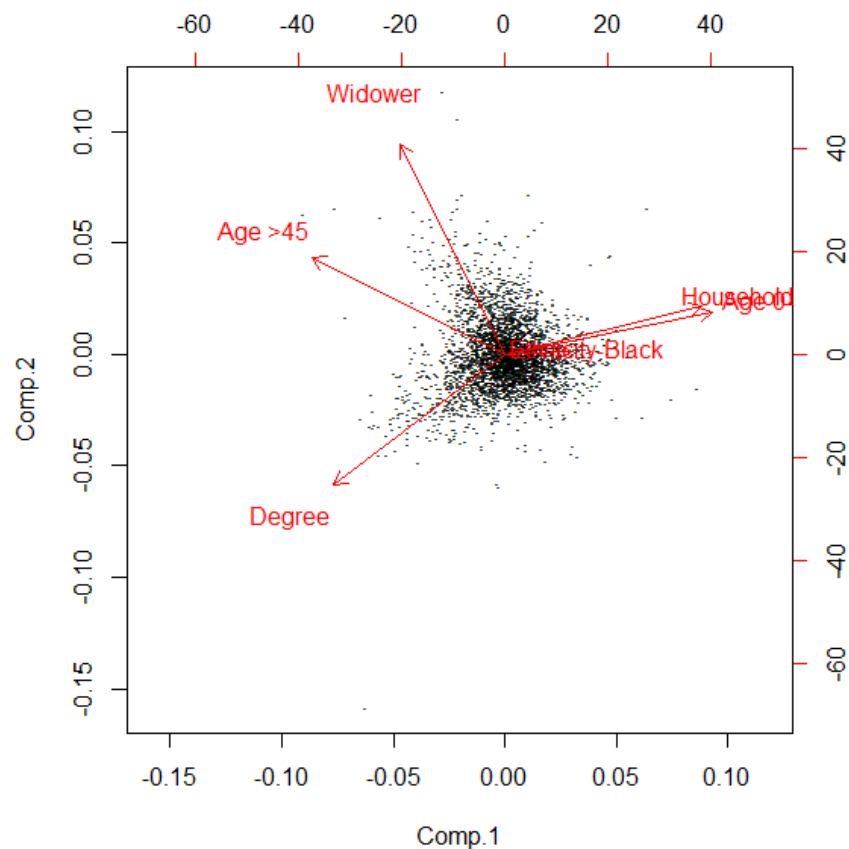
FIGURE 4-20 – BIPLOT OF PRINCIPAL COMPONENT VARIABLES (MAPPED AS VECTORS) AND CENSUS AREAS (MAPPED AS OBSERVATIONS)

On the basis of these findings and in following Cutter's methodology, that seeks to preserve only highly correlated factors, the less correlated variables 'Widower' and 'Age >45' were subsequently removed from the test.

In following direction from the *SoVI recipe*, the negatively correlated 'Degree' variable was inverted by subtracting the standardised observation score (x) from 1 (e.g. 1-x), thus creating a new input variable.

The varimax extraction scores for each census area were calculated and input as a separate variable before the additive equation shown below (11) was used to formulate the vulnerability score.

$$\text{Vulnerability score (Cutter)}_x = \text{Age}(0-4)_x + \text{Degree}_x + \text{Household} > 6_x \qquad 11)$$

$$+\text{Ethnicity} - \text{black}_x + \text{ PCA weighting}_x$$

$$\text{Vulnerability score (Gini method)}_x = \text{Age}(0-4)_x^{Gini\,x} + \text{Degree}_x^{Gini\,x} \qquad 12)$$

$$+\text{Household} > 6_x^{Gini\,x} + Ethnicity - black_x^{Gini\,x}$$

Cutter's SoVI approach was then contrasted to the Gini-Coefficient methodology devised during thesis research and shown below in equation 12). To make sure the comparison was entirely objective, the same four variables were used in the calculation, and then plotted against each observation as coordinates in Figure 4-21.
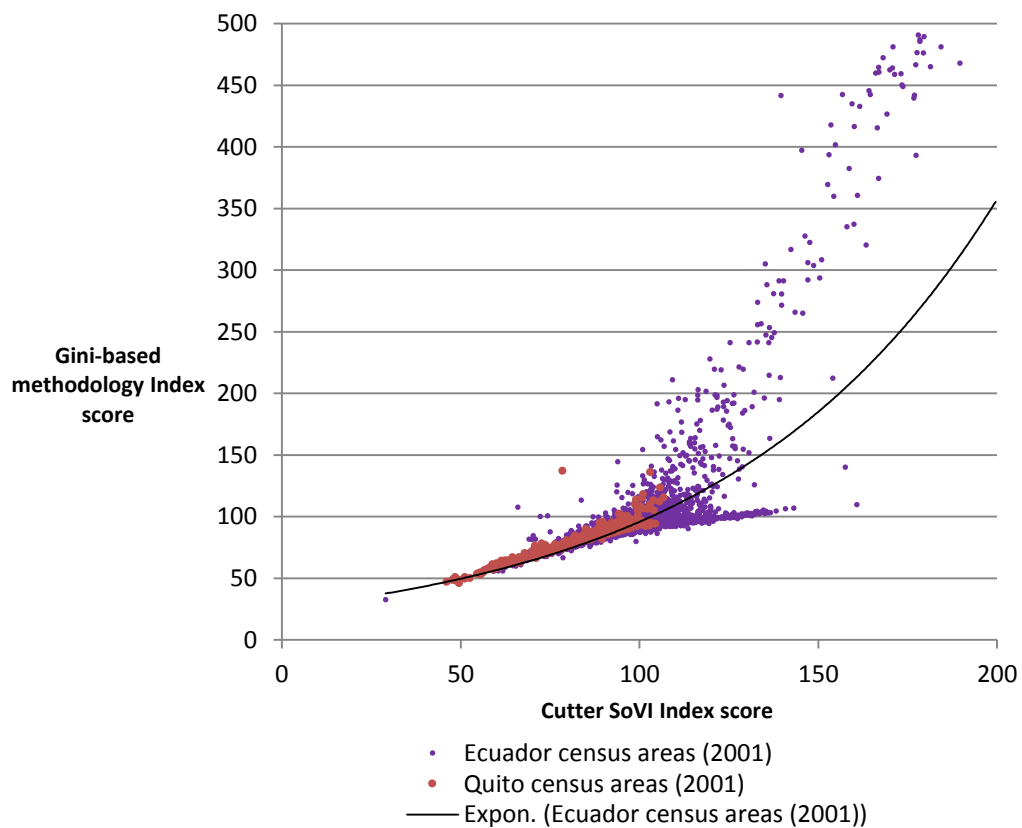
FIGURE 4-21 – ECUADOREAN CENSUS LEVEL SOCIAL VULNERABILITY INDICES MAPPED AS COORDINATES (PCA-BASED METHODOLOGY VS GINI-COEFFICIENT METHODOLOGY)

Figure 4-21 shows that the contrasting methodologies are closely correlated (0.72). In assessing social vulnerability on a national level, the chart also shows how the Gini-Coefficient methodology displays an increasingly exponential relationship above the mean average (x=100). This is evidenced by higher correlation for the low scoring census areas of Quito (0.93). Such results indicate a greater value of vulnerability is attributed to the same census area by the Gini method in the most vulnerable zones.

Conversely, Cutter's approach, without an exponential multiplier, results in preserving a more linear distribution. Census areas for Quito are highlighted in Figure 4-21 (in red), and help demonstrate that in comparison to overall social vulnerability ranking across Ecuador, Quitenos neighbourhoods are nearly all below average (<100) using both methods.
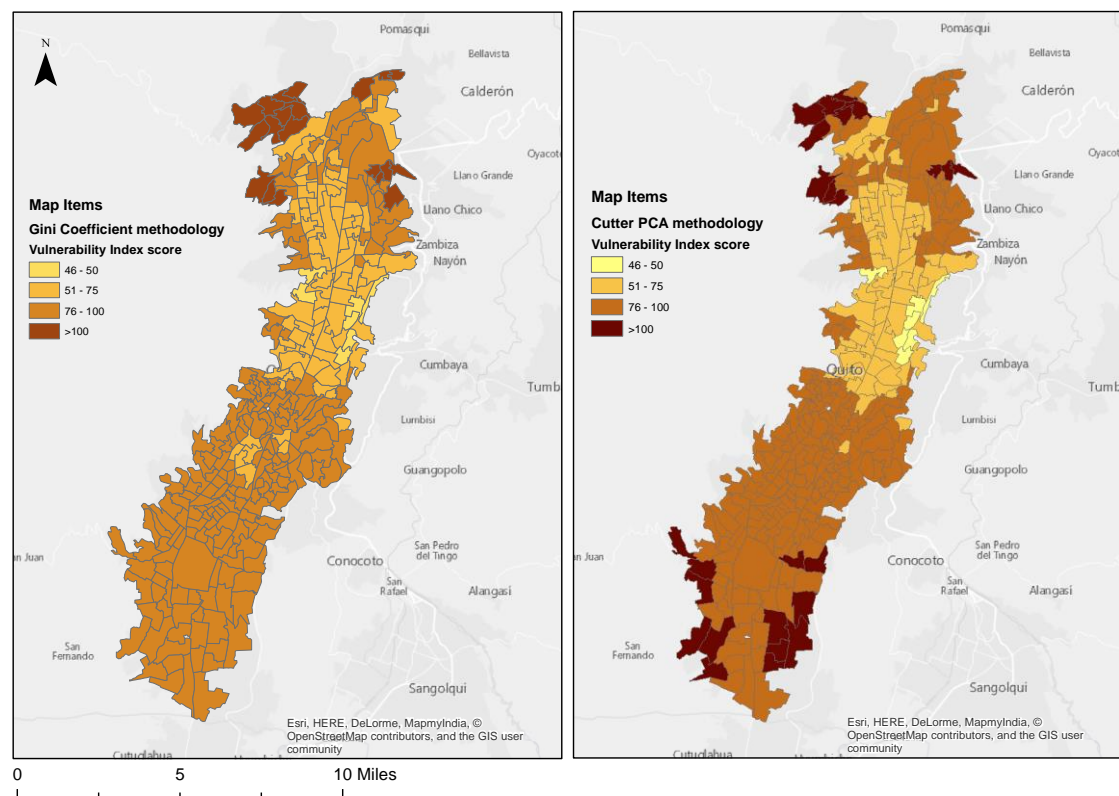
FIGURE 4-22 – A SPATIAL COMPARISON OF SOCIAL VULNERABILITY METHODOLOGIES, QUITO CENSUS ZONES, ECUADOR. LEFT: GINI COEFFICIENT (WILLIS ET 2010), RIGHT: PCA METHODOLOGY (CUTTER 1996)

Figure 4-22 shows a spatial analysis of the contrasted social vulnerability methods for Quito Metropolitan census areas. Both methods show display a similar overall relationship of vulnerability, with the most socially vulnerable areas (dark brown areas) appearing in the fringe peripheral barrios of city. Likewise, the less vulnerable areas (shown in light yellow) are found in the central, historic areas of the city.

There is also some differentiation noted in Figure 4-21. Cutter's approach results in greater variation of vulnerability within the south of Quito, show vulnerable communities on the southern periphery. In applying an index to both scores, the more extreme values of the Gini approach result in less variation within the lower scoring areas, such as those found in Quito. The result is that in comparison to Cutter's method, there appear to be less variation in areas like Quito.

For further discussion regarding the findings and assessment of the model validation tests, please refer to the *Discussion and Conclusions* chapter (5).

4.4   CHAPTER SUMMARY

This chapter began by presenting the research outputs from the application of a commercial geodemographic database (Mosaic Italy 2007) used in constructing a social vulnerability index to assess the Campanian population around Mount Vesuvius. Coupled with GIS –derived volcanic hazard maps, a methodology based on weighting index scores by the Gini-Coefficient of the variable highlighted the spatial relationship of social vulnerability and neighbourhood profiles. The constituent parts of the overall index (a household's access to resources, financial frailty, evacuation vulnerability) demonstrated a complex and often inverse relationship to each other. However, key drivers of the overall vulnerability, aside from the proximity to the

principal volcanic hazards, were seen in census areas of low employment, high population density and among households of advanced age.

In section 4.2, the two-tiers Ecuador geodemographic created from 2001 census data was presented, along with profile portrait descriptions for the Tier 1 groups. Index tables and GIS cluster maps highlight the substantial variance in Ecuadorian culture and demography within both tiers. Ethnicity, education and population density are key determinants in the outcome of the geodemographic groups.

Gini-Coefficient weighted social vulnerability scores based on the geodemographic are then presented for Quito Metropolitan districts to highlight the relative risk associated with the neighbouring Guagua Pichincha volcano. Figure 4-13 shows that overall social vulnerability appears to be heavily correlated with income, identifying the peripheral barrios of the city as scoring highest. Likewise, the assessment of social vulnerability in relation to a historic Plinian eruption of Guagua Pichincha implies that at critical tephra thresholds, more vulnerable sections of the city would not be disproportionately affected (in population) than more affluent central areas.

The final section compared the methodology presented within this thesis with both less aggregated datasets and with an alternative method of vulnerability calculation. Results are presented using GIS maps for Quito and with comparative scatter plots of the resultant indices. Indications of these results show that aggregated vulnerability indices show high variance at 95% confidence levels, although this varies significantly between cluster groups. In testing Susan Cutter's PCA approach to creating a SoVI (2006), the Ecuador data highlights that Gini-Coefficient weighting exponentially impacts the more extreme vulnerability scores. However, given the same input variables, the two contrasted approaches show a 0.72 correlation at a national level and a 0.93 correlation for Quito census areas.

5. DISCUSSION AND CONCLUSIONS

This chapter discusses the overall application and value of this contribution within the wider framework of DRR. The study region results presented in this thesis (for both Italy and Ecuador) are discussed and contrasted, as well as an open debate regarding the current limitations of this approach. Similarly, due consideration is given to how this methodology could be applied in different geographic settings and at different spatial levels across the world for the purposes of disaster risk mitigation. Lastly, ideas for the continuation of this research topic are proposed as well as how such endeavours could further contribute to global UNISDR strategic plans following the end of the Hyogo framework for disaster reduction (2015) and its successor, SFDRR.

5.1 STUDY REGIONS

Findings from the study region areas of this research imply that neighbourhood classification systems have great value in helping to identify socially vulnerable communities in hazardous settlement areas around volcanoes. The approaches outlined within this thesis also have resonance to other natural and anthropogenic hazards, given the ubiquitous nature of social vulnerability and historic patterns of human settlement. Rather than identifying risk based on single or composite scores of census or survey indicators (e.g. age, ethnicity, employment, housing, population density), multivariate clustering, both in commercially available geodemographics or academic solutions, allows for the classification of vulnerable populations at a community level.

Unlike methodologies focused on single risk or vulnerability indicators, cluster analysis is able to remove much of the statistical bias or weighting that such techniques typically place on variables associated with vulnerability. In developing a geodemographic, the methodology places more focus on the behavioural, cultural and life-status variables (such as household

composition, marital status, educational attainment) that would not otherwise be factored into DRR assessment. There are key benefits to this approach – it allows practitioners to identify rounded descriptions of neighbourhoods, which are grounded in definitive profile types and can be linked to tertiary sources of data, such as telemarketing or survey data. The result is a tool than can provide key insight into behavior traits and risk perception. It can help DRR planners and NGOs target vulnerable communities for disaster mitigation and preparedness.

5.1.1   USING A COMMERCIAL GEODEMOGRAPHIC (MOSAIC ITALY)

Results highlighted that in the areas around Mount Vesuvius, Italy, the Mosaic classifications of *Elderly households* and *Low status apartment* could be particularly vulnerable in the event of a Sub-Plinian volcanic eruption. Aside from the geographical location of these households within the highly populated areas near the coastal base of the volcano, these communities showed a greater prevalence in the social vulnerability indicators. The Mosaic group *Elderly households* comprises of a high population density (nearly twice the national average) and had a significant proportion of 65 year olds (approximately 50% higher than the national average) living in a poor economic situation. These same variables have been inextricably linked to particularly vulnerable communities during and following the onset of a natural disaster. The *Low status apartments* Mosaic category is also particularly vulnerable, given the high population density and with an illiteracy rate twice the national average. With the highest percentage of those under 5 years old of any category, evacuation of these neighbourhoods prior to an eruption could be considerably more difficult than other areas. Unemployment in this cluster group was also the highest of any Mosaic Italy group, indicating that the financial ramifications of an eruption could have a disproportionate impact on these neighbourhoods.

Perhaps the more concerning factor from a disaster risk reduction perspective is that the *Low status apartments* Mosaic group is significantly over-represented in its population within the 50km analysis territory around the Vesuvius volcano. Figure 5-1 illustrates that only 8% of Italian households belong to this cluster group on a national level, but if we consider only the

population in the volcano analysis territory, the cluster population rises to represent 65% of the total population. Such results indicate that the population around Vesuvius, and in neighbouring Naples, has an increased social vulnerability.
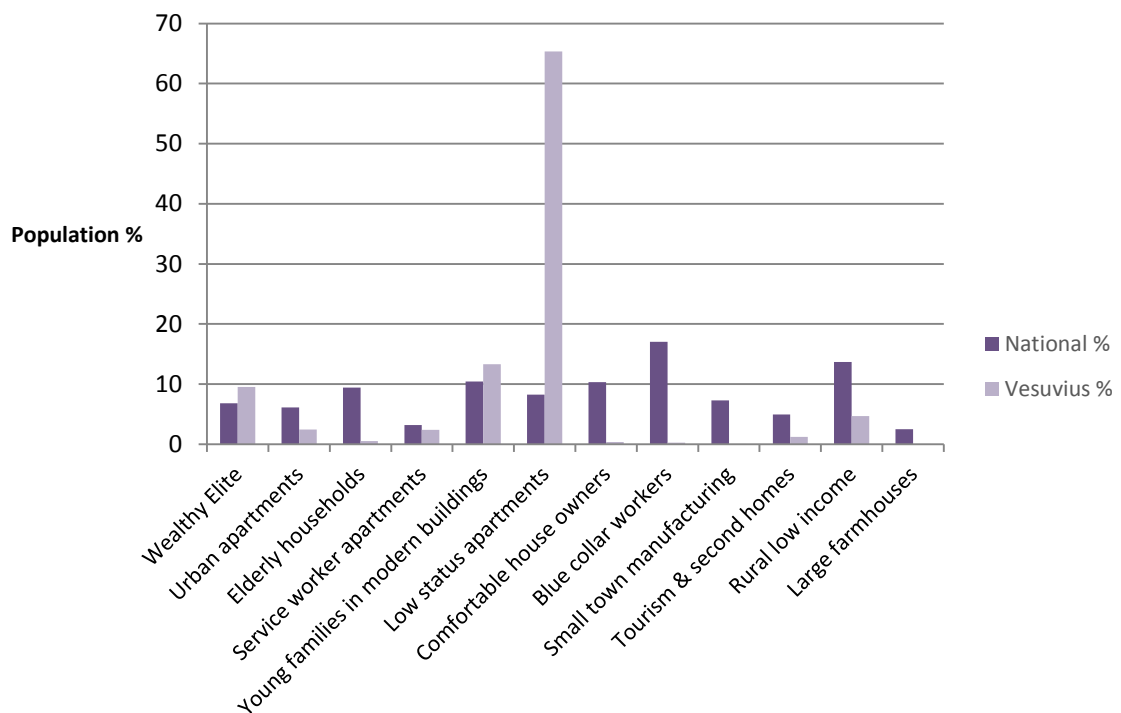
The Vesuvius study showed how the creation of social vulnerability indices based on geodemographic index scores provided a valuable method of aggregating and ranking different vulnerability indicators in a meaningful way. By analysing maps of only those vulnerable to *evacuation, financial recovery,* and *access to resources¸* it helped demonstrate the complex spatial and socio-economic interaction of these diverse risk types. For example, those vulnerable to the stress and rigors of disaster *evacuation* have a profoundly different spatial footprint to areas of *financial* vulnerability.

As Figure 4-2 highlighted earlier in this thesis, *evacuation* is highly correlated with population density and therefore, rural locations in the analysis region showed a very low score compared to the city provinces and larger towns in Campania. Although the *evacuation* index is a

composite score of multiple factors including *age and daily movement*, the most influential variables appear to have been *population density* and *building height* – these last two factors are greatly reduced in rural locations hence the low scores. In sharp contrast, *financial* recovery is linked to variables such as *literacy, housing tenure* and *unemployment,* thus resulting in higher social vulnerabilities in the rural areas of the province. This helped show that measuring social vulnerability is not a uniform concept but rather, a relative term depending on the specific attributes that are assigned.

Given this inverse relationship for different social vulnerability types, this study raises the bigger question of whether it is right to assign a single index to define social vulnerability? In this study, the overall vulnerability index implied that the highly populous areas around the Volcano were the most socially vulnerable. However, if more weighting had been given to the *financial* variables, it is possible a very different outcome would have been achieved. Further to this, the choice of the geodemographic variables used within a vulnerability classification is also paramount. Though *gender* can be a defining characteristic of social vulnerability during a natural disaster, it has very low capability for statistical discrimination in aggregated datasets. To highlight this lack of distinction within the Mosaic Italy dataset, the calculated Gini-coefficients for the variables *% Females* and *% Ethnicity-Asian* are 0.02, and 0.54 respectively. Given that a low Gini-coefficient (i.e. close to 0) shows greater equality within a distribution, this helps confirm that gender is almost perfectly equal in its distribution within the Mosaic areas, and therefore, holds little value in discriminating aggregated populations. Conversely, the *Asian-Ethnicity* Gini-coefficient shows a high discriminatory ability and is able to model the variance among different cluster groups and indices. With this mind, it could be argued that geodemographics by their very construct are incapable of capturing the full statistical significance of key variables associated with social vulnerability. With many global census datasets (such as Ecuador) available at anonymised individual level, it is conceivable to construct a vulnerability index at the most granular level (i.e. assigned to an individual). Such information could help inform disaster risk reduction research as to the distribution pattern of

social vulnerability within a community and also identify how gender and lifestyle characteristics could be accounted for. However, it is worth noting, the anonymous nature of the data would continue to limit how such information could be meaningfully disseminated or used to target/communicate disaster preparedness. Though such knowledge would help with understanding individual risk, it may not be entirely practical in terms of outreach. In recent years, there have been several government initiatives aimed at providing open data, such as the UK government's data.gov.uk. Such movements allow researchers free access to fundamental datasets on population dynamics, lifestyle, and behavior. In consideration of this thesis, if such initiatives were adopted more widely by developing nations, it would greatly help in the global assessment of social vulnerability.

Given the commercial intentions of Mosaic Italy to provide users with tools for direct marketing and geographically defined consumer insight, a key benefit of such classifications is the large amount of lifestyle survey variables associated with the dataset. The Mosaic Italy dataset is provided with 222 variables, covering a range of lifestyle and socio-demographic indicators that go far beyond the Italian census. Many of these supplementary variables are provided by tertiary data sources including Experian's credit checking service or through the acquisition of telemarketing survey data. This allows Mosaic and other commercial neighbourhood classification providers to create propensity indices for factors such as credit card ownership or the proportion of those who have taken out a personal loan. Although such information is not a direct cause of natural hazard social vulnerability, it provides a valid measure of assessing financial risk to post disaster recovery.

An interesting development in recent years has been the concept of global neighbourhood classification products. Experian created the commercial product, *Mosaic Global* on the basis that similarities could be identified in communities internationally, as well as on a regional basis. Their global product currently classifies 880 million consumers into 10 categories (Mosaic 2014). Covering 25 countries, the classification is largely made up of European nations, North America, Australia, New Zealand, Japan, Hong Kong and Singapore. The 10

categories are fairly broad but include profiles that cover affluent neighbourhoods (e.g. *Sophisticated singles, Bourgeois prosperity*), older demographics (e.g. *Low income elders*) and those employed in more routine industries (e.g. *Hard working blue collar*). Having discussed the derivation of Mosaic Global with Richard Webber, a leading expert on geodemographics and creator of the Mosaic Global product, he intimated that the offering was a natural progression based on the similarities that had been observed between different country-level classifications.

"*The same type of neighbourhoods could be identified in Australia as well as Japan, the UK, or Singapore. It was then a process of mapping the country level clusters to the most appropriate global description.*" Richard Webber, (personal communication, 2013)

The availability of a global classification system could offer valuable insight and practical application in the analysis and outreach of disaster risk reduction. The global commonality of neighbourhood types would imply a natural correlation with social vulnerability indicators, given that factors such as wealth, ethnicity, and demography are key inputs in both. Similarly, a series of global clusters linked to social vulnerability indicators could provide researchers with a transferrable and scalable tool in disaster response, mitigation or preparedness.

However, there could be major issues with the practical implementation of a single transferrable geodemographic for DRR purposes. Firstly, existing classifications such as Mosaic Global are inherently built for a commercial purpose and therefore, their creation has been targeted around leading world economies and markets. Figure 5-2 highlights Experian's full coverage in Europe, North America and Japan, yet this also illustrates that there is no coverage in South America, Africa, and large parts of Asia. There are likely to be several reasons for this: the availability of accurate census data; detailed electoral or financial data; available and reliable telemarketing information; and commercial demand for consumer insight information in Least Developed Countries (LDC).
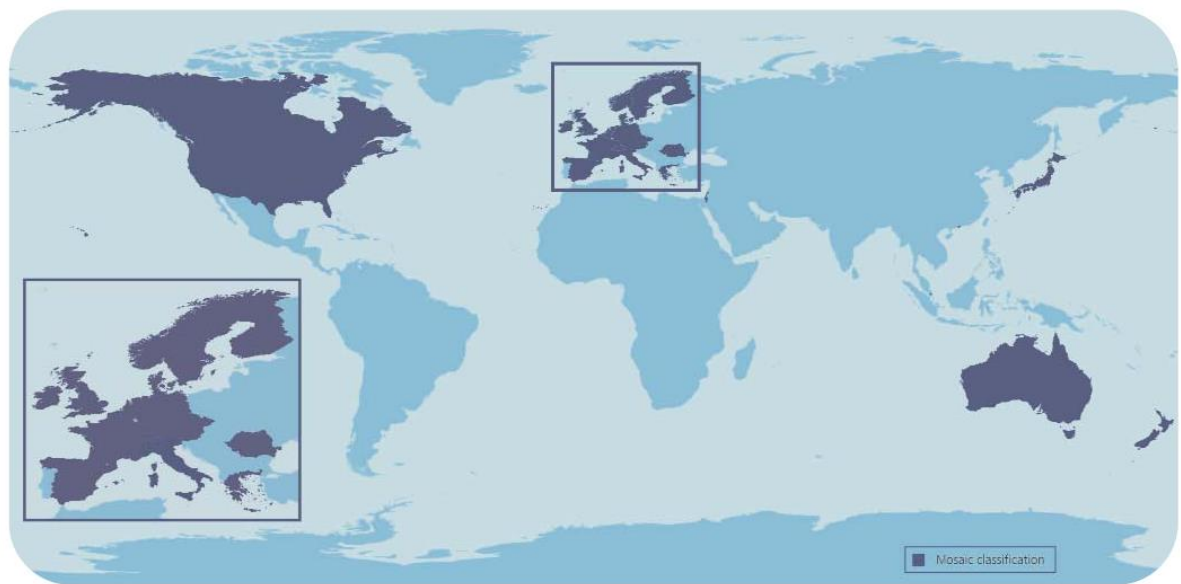
FIGURE 5-2 - MOSAIC GLOBAL COVERAGE (MOSAIC GLOBAL COUNTRIES ARE HIGHLIGHTED IN GREY) (MOSAIC 2014)

A secondary concern in creating a global vulnerability index pertains to the level of aggregation. By mapping across a countries geodemographic groups into fewer global classifications, far greater uncertainty and homogeneity is introduced. The cultural and statistical nuances of a single region are lost in the process of data reduction. For example, would the lifestyle, behaviours, and social vulnerabilities of a category described as *Hard working blue collar* be that similar for a household in the UK as to one in Japan? Likewise, would their risk perceptions and vulnerability to natural hazards be the same? Given the way society, religion and culture help influence our risk perception, it would seem unlikely that a global classification could accurately capture such vulnerability on a universal basis.

5.1.2   ERUPTION SCENARIO, GUAGUA PICHINCHA, ECUADOR

The disaster scenario of the 10th Century eruption of the Guagua Pichincha volcano, Quito, a presented in section 4.2.3 provided key insight into how the specific dynamics of volcanic hazards could impact a large urban population. Unlike many volcanoes, where the principal risk

is from pyroclastic flows, lava inundation, or lahars, the more significant (and immediate) risk of Guagua Pichincha on the neighbouring city of Quito appears to be from tephra fall.

Although the steep flanks of the Guagua Pichincha volcano are home to tens of thousands of Quitenos, often living in cramped, close quarters with a lack of basic amenities or infrastructure (Carrion et al 2003), the physical risk posed by an eruption may be more likely due to the heavy ash fall loading (given prevailing winds and proximity to the volcanic vent) than other volcanic hazards. In creating the 10th Century tephra fall scenario for Guagua Pichincha, it was interesting to note that the orientation of the city, as well as the concentric structure of its neighbourhoods imply that the proportion of areas impacted with dangerous levels of ash loading (>10cm) would not significantly over-expose any defined cluster group. A key factor in determining this was the assumed ash fall orientation of the 10th Century eruption, which did not appear to have greatly impacted the southern section of the metropolitan district of Quito. Since this region houses the largest proportion of the 'peripheral barrios' cluster group, it could be stated that many of the most socially vulnerable communities in Quito were not impacted in the analysis population.

However, such conclusions may only be referenced within the context of the 10th century eruption scenario presented here. Any sensitivity testing around this eruption scenario (e.g. changing the level of ash fall loading, prevailing wind direction, or precipitation experienced during the eruption) would almost certainly warrant a different tephra footprint and thus a different impact would be noted on demographic groups.

Similarly, this study was limited to the sole consideration of tephra fall and did not explore the possible impact of lahars and pyroclastic flows. Canuti et al (2002) highlighted that since the Holocene, numerous Pichincha lahars have impacted areas now occupied by settlers on the edge of the city limits. Their GIS-based modelling of the Rumipamba and Rumiurca streams showed that a sizeable hydrological discharge could impact areas on the populated flanks of the volcano. Likewise, although the older volcanic edifice of Rucu Pichincha protects the city from the imminent risk of pyroclastic flows, given lower magnitudes of eruption, it is not certain that

such protection would still be afforded if more catastrophic circumstances occurred. For example, if caldera collapse occurred to both the Rucu and Guagua volcanic edifices, the threat of gravity driven hazards such as pyroclastic flows would be infinitely more significant.

It is important to note that this example was focused on the Metropolitan Districts of Quito and made no attempt to define the risk posed to the areas immediately surrounding these census zones. This is an over-simplification of the likely impact of an eruption event scenario given that 24% (approximately 450,000) of the population in the Quito Canton live within the *Parroquia rurales* (rural parishes) and not in the more densely populated capital city. Many of these rural neighbourhoods are reliant on agricultural work and would be directly impacted by an eruption of Guagua Pichincha, and therefore, it is vital to consider such areas in future DRR initiatives. Likewise, this study was focused on a single volcano in Ecuador and has not covered the other numerous and arguably more active volcanoes that align the Sierra valley, such as Cotopaxi or Tungurahua. This latter volcano erupted in February 2014 (BBC 2014) and continues to cause disruption and health concerns to the neighbouring town of Banos (Tobin and Whiteford 2002).

5.1.3   CREATING A BESPOKE GEODEMOGRAPHIC (ECUADOR STUDY REGION)

By using similar statistical methodologies as those employed in creating commercial or open initiative geodemographic databases, this research has successfully demonstrated that such techniques can be applied in creating an Ecuadorian neighbourhood classification system from census data. Similarly, sensitivity testing using principle component analysis (PCA) and multivariate data reduction tests helped inform on the most appropriate method for constructing the Ecuadorian geodemographic clusters. Results highlighted that key factors in understanding the variance in the census data were related to three principle factors; *Ethnicity, Population density* and *Education*.

Ethnicity is the greatest determinant of neighbourhood types that were defined in the classification. The regional occurrence and variance in the Ecuadorian clusters was largely

217 | P a g e

derived from the ethnicity variables. The categories *Black, Mulato, White, Mestizo* and *Indigenous* defined the neighborhood's socio-economic and geographic distribution. For example, in consideration of the Tier 1 clusters, neighborhoods such as '*Indigenous farming*' have over 11 times the national average of those identified as *indigenous*. Similarly, the '*City living*'category has an over-representation of *white* ethnicity (around three times higher than the national average) whereas *black* communities are strongly correlated with the '*Afro Caribbean*' category.

Likewise, *population density* and *educational achievement* are inherently linked to the variance seen in the Ecuadorian ethnicity. It is no coincidence that the '*Indigenous farming*' cluster has the second lowest population density of any category and the lowest index in degree attainment (1/10 of the national average). Conversely, the 'City living' category, as the name suggests, is found solely around the largest urban centres of Ecuador (e.g. Guayaquil and Quito) and has a markedly different socio-economic structure. Individuals living in these areas are more than twice as likely to be divorced and up to five times as likely to work in a financial or managerial role. Such trends imply that ethnicity, population density and educational achievement are not only intrinsically linked to each other but are an integral factor in the regional and socio-economic landscape of Ecuador.

Given historical population trends in the migration and colonization of Ecuador, as well as more recent movements associated with globalization, it is perhaps not surprising to see such stark regional trends in demography. The North-West province of Ecuador, known as Esmeraldas, has over one million individuals of African descent and with cultural origins dating back to the emancipation of slaves in 1553 (Labarga 1997), it is completely distinct within Ecuador regarding its cultural identity. Similarly, those communities associated with large indigenous populations are predominantly found in the Eastern provinces of the country, and are particularly prevalent within territories of the Amazon basin. With cultural origins dating back to a pre-Colombian era, these indigenous populations often retain strong tribal traditions and heritage.

Although the Ecuador geodemographic has been created in two tier of classification (8 and 42 groups respectively), it is clear that further variance and different neighborhoods could also have been defined during this research. For example, by increasing the number of clusters created, greater heterogeneity could certainly have been defined in the census areas in Quito.

The rationale for the number of clusters was based on the optimal amount of groups that could be defined for communication and DRR outreach purposes on a national level whilst also preserving sufficient heterogeneity between groups. This is largely subjective, and the use of increased clustering in high population centres (e.g. Quito) or around significant volcanoes may have provided more precision in segmenting the population. Furthermore, the application of the *ward's* method for clustering, in preference to *k-means* has certainly influenced the outcome of clusters and neighbourhood definition.

As discussed in previous chapters, using variable distributions that are prone to outliers, as was the case with the Ecuadorian census variables, resulted in multiple instances of single entity clusters when using the *k-means* method. During the first iteration, the clustering seeds selected by the algorithm were based on these initial outliers, which despite repeated iterations, ultimately resulted in a very uneven spread of clusters. Conversely, the hierarchical approach of *Ward's* method provided a more even distribution and greater distinction in the urban population zones. However, further sensitivity testing and spatial analysis of both these cluster methods could also have been carried out in different regions of Ecuador to assess the overall impact of cluster method on social vulnerability assessment in Ecuador.

Similarly, the Ecuador geodemographic was limited by the INEC choice of census questions and the fact that there was no telemarketing or supporting survey data related to the model. For example, many vulnerability traits were not used in the construction of the Ecuador model but could have included important household amenities (e.g. household telephone, internet connection) or financial information (credit card ownership, loans). Likewise, if the methodologies proposed here were transferred and applied to other countries and at different

scales, similar issues of data availability would have an impact on the vulnerability indices produced.

5.1.4 TEPHRAVOLCANIC HAZARD MODELING AND VULNERABILITY INDICES

The research on Ecuador further raised a debatable point regarding the social vulnerability of communities to volcanic hazards. Because volcanic hazards are such a complex peril and vary so significantly depending on the exact geology, geomorphology and changeable antecedent conditions (e.g. wind direction or rainfall), identifying the significant hazard to life can be challenging. In contrast to other geophysical hazards, such as earthquakes, where metrics such as the spectral acceleration or peak ground acceleration can be directly related to the likely damage an event could cause to building stock, volcanoes are multi-peril phenomena and different locations provide very different hazards. Though not surprising, such site-specific conditions create significant challenges in providing a social vulnerability index to evaluate the risk of volcanoes. For example, assigning appropriate weights to hazards such as pyroclastic flows or lahars is a highly subjective process. The relative impact of volcanic hazards varies so significantly, depending on geophysical conditions, the magnitude of a given eruption and the vulnerability of the housing stock, that it is not possible to produce a single, transferrable weighting for each sub-peril. For example, though gradual tephra fall may lead to the collapse of a roof, and subsequent death or injury to a household, tephra fall is widely regarded as less life threatening than more dynamic hazards such as pyroclastic flows, lava inundation or lahars.

For example, given that many vulnerable communities are spatially disconnected, can it be assumed that the populations in these areas would react the same way during the onset of an eruption? It seems unlikely that a community living at the foot of the perpetually active Tungurahua volcano would have the same coping mechanisms as a community living at the base of the dormant Pichincha volcano, even though both communities may have been assigned the same geodemographic profile in the database. Local knowledge and education of a neighbouring volcano are key themes that the research has not factored into account but which

would certainly play a pivotal role prior or even during an eruption. Ancestral knowledge and insight acquired through learned experience (i.e. previous eruptions) are key elements in both risk perception and risk behaviour. Though key assumptions can be made about the attitudinal qualities of a neighbourhood, based on their socio-economic or ethnographic status, geodemographic databases cannot capture all intricacies of site-specific neighbourhood behaviour.

Because the underlying use of neighbourhood classification systems is entirely based on their statistical creation from census data, it is important to consider the methodology in creating both the clusters and the vulnerability indices. For example, by choosing different variables or weighting techniques in the creation of the indices, entirely different results could have been gained. This could certainly be an area of further research. Greater sensitivity testing around both the initial clustering and vulnerability model would be likely to warrant very different results and therefore expose the more critical components of the vulnerability model. An early finding from the case studies implied that the assessment of social vulnerability measures must be robust and appropriate as the subsequent weighting of variables determines the outcome of the model.

5.1.5   ISSUES OF DRR SOCIAL VULNERABILITY CATEGORIES

In some cases, reviewing the literature has shown that a variable can have both a positive and negative effect on social vulnerability, depending on the geographic context. A good example of this can be seen in the ethnicity variables of the Mosaic Italy 2007 dataset. Non-native ethnicities and ethnic minority areas are broadly considered in DRR literature to be more predisposed to risk during a natural disaster due to a lack of political access, financial resource or communication access. However, this is something of a blanket summary in social vulnerability studies and appears to be highly context-specific.

In terms of social statistics, areas with high alternative ethnicity can also found in the very centre of the wealthiest towns and cities. Statistically, these demographic groups often consist

of very well educated, affluent individuals who would not ordinarily be declared as socially vulnerable. Thus, the studies of Italy and Ecuador highlighted that census variables should be understood within the context of their geographic and cultural setting otherwise misleading assumptions can be made about the underlying population.

Good examples to help highlight this complexity and ambiguity were shown during the methodological development of this research. In the Mosaic Italy dataset, the *Urban apartments* cluster category displayed the highest vulnerability to the vulnerability category *access to resources* with many of the variables being associated with the census area's population of Afro-Caribbean and Asian descent. However, many of the Mosaic sub-groups within this category were associated with being among the most affluent in the Campania province (e.g. *Well educated professionals, Young singles*).

In stark contrast, this pattern of inner city ethnic diversity was not seen in the Ecuador classification, where the prominent inner city neighbourhoods were largely under-represented by those individuals of Black or Afro-Caribbean descent. Such duality in variable trends creates a key dilemma for those seeking to define composite vulnerability metrics. The relationship of ethnicity to criteria such as financial and educational achievement is extremely complex, varies widely, and depends on the cultural and economic setting.

Ecuador and Italy are fundamentally very different economies and thus have different socio-economic structures. Italy is regarded as a More Developed Country (MDC) and one of the world's ten largest economies (World Bank 2012). Its post-industrial market economy and foreign policy on immigration have led to highly cosmopolitan cities with multiple ethnicities. Ecuador's colonial history and rapid urban-rural migration have led to a more conservative city structure, with black and indigenous communities still being somewhat ostracized from more affluent areas. Likewise, such results have implications for statisticians and those seeking to define global transferrable vulnerability metrics.

5.1.6   Transferrable methodology

The indicators of social vulnerability to a given hazard are peril and location dependent. Though a peril may exhibit similar hazard behaviour in two divergent geographic settings, the cultural and socio-economic differences between the countries may be vast and varied. This may result in a very different disaster experience from one setting to the next. The Chile and Haiti earthquakes of 2010 provided a stark reminder of this phenomenon. A 7Mw earthquake in Haiti (12th January 2010) killed 230,000 and affected 3 million, particularly around the capital city of Port-au-Prince (Bilham 2010). Conversely, the Mw 8.8 earthquake in Chile on the 27th February 2010 killed 562 people and generated a tsunami that hit many coastal communities (Guha-Sapir et al 2011). This latter event, near the Chilean city of Concepcion, was 500 times stronger than the Haiti event a month before, yet caused considerably less damage and destruction to infrastructure and livelihood. Haiti and Chile are considerably different in financial terms, with economic output per head around 10 times greater in Chile (BBC 2014). Likewise, and perhaps most significantly, since 1960 Chile has administered a revised seismic building code to be among the most advanced in the world. Whereas most buildings in Chile are designed with reinforced concrete beams, Haiti's building stock was poorly prepared and thousands of masonry structures were reduced to rubble, including schools, hospitals and municipal buildings (Bilham 2010).

Another consideration is city structure. Given the marked differences in the city structure of the case studies presented in this thesis, it raises a concern for creating a transferrable methodology. Latin American cities such as Ecuador typically have informal (and often) illegal housing settlements on the outskirts of the city. These peripheral barrios can be particularly vulnerable. In contrast, European cities typically have some of their wealthiest areas located in the peripheral suburbs. Creating a single vulnerability methodology that can represent such an inverse relationship in different geographic settings presents a range of issues. For example, how will such differences affect disaster evacuation?

Although geodemographics has been used successfully to create global neighbourhood profiles (e.g. Mosaic global) on the basis of statistical similarities, such profiles are limited to

describing the mature economies of MDCs. Thus, applying the same social vulnerability indicators of ethnicity, economy, and demographics to LDCs appears to be fraught with issues. Inter and intra migration, indigenous populations, as well as the continued reliance on agrarian economy provides countries such as Ecuador with a markedly different structure to MDCs such as Italy. The research undertaken during this thesis appears to confirm the notion that social vulnerability to natural hazards should be defined on a bespoke basis. It is only by qualitative and quantitative consideration of the underlying population that vulnerability can be understood within the specific geographical and cultural setting. Such approaches appear the best way to capture the nuances of attitudinal and behavioural differences between communities, and how different neighbourhoods respond before, during and post eruption.

5.2   LIMITATIONS AND FUTURE RESEARCH CONSIDERATIONS

The following section discusses the limitations of this research agenda as well as proposing ideas for how this work can be further progressed in the future.

Though the aims and results of this research have been transparent in methodology, construction and findings, it is important to take note of the various uncertainties, both aleatory and epistemic, that underpin the use of neighbourhood classification systems for volcanic hazard vulnerability analysis. The proceeding section describes some of the methodological principles, issues of model application, and statistical uncertainty that may be further reduced, understood or accepted within the framework and wider application of this work to disaster risk DRR.

STATISTICAL TESTING AND SENSITIVITY ANALYSIS

Section 4.3 illustrates the results of model validation and sensitivity testing. Tests were undertaken to objectively assess both the Gini-weighting methodology and the application of geodemographic index scores in suitability of vulnerability assessment. Whilst the merits of this technique are discussed in the proceeding *Thesis conclusions* chapter, it would be fair to acknowledge that more scrutiny of this methodology could have been performed.

For example, is the discriminatory ability of using Gini coefficients the most valid and robust method of weighting census variable index scores? By using this variable value as a weighting ratio, it has been seen polarise extreme index values and thus, exacerbate relative risk measures. The application of this weighting may be considered unwise in circumstances where a single variable, recording a very high index score, can serve to skew the vulnerability measure in a particularly category (e.g. evacuation, access to resources, financial recovery). Likewise, the implications of this step in the methodology will have a cumulative effect in the final classification of the 'at risk' profiles. A more appropriate form of discrimination may be to weight each variable on the basis of subjective social vulnerability assessment. A factor could be applied based on *qualitative* consideration of DRR literature rather than quantitative assessments such as Gini-coefficients or PCA extraction scores (as advocated by Cutter et al 2003).

Likewise, the implications of this step in the methodology may have major consequences in the final classification of the 'at risk' profiles. Lisa Rygel argues that using a Pareto ranking of vulnerability indicators can remove much of the inherent bias of weighting indicators (Rygel et al 2006). However, perhaps the most fundamental step in defining a vulnerability index is to first ascertain the conceptual framework to which the practitioner subscribes (Rygel et al 2006). This will inherently dictate the indicators and thus, the key drivers of the vulnerability assessment. This is succinctly summed up by Alwang et al (2001), *"Practitioners from different disciplines use different meanings and concepts of vulnerability, which, in turn, have led to diverse methods of measuring vulnerability."*

This research proposes a new neighbourhood based conceptual framework, which principally takes its roots from the Pressure and Release Model (PAR) (Blaikie et al 2004). In progressing this research agenda further, it may be interesting to perform sensitivity testing around different conceptual models given the same region and available data.

PHYSICAL RISK MODELLING AND RANKING

In the Vesuvius study, the choice of ranking Civil Evacuation areas, tephra loading maps and pyroclastic flow regions with an integer of between 0-3 is without doubt an oversimplification of the gradations of volcanic hazard around Vesuvius. For example, the Tephra loading model is defined by a sliding scale so that exact volumes were calculated. However, a risk rank of 0-3 does not take into account these variations. Likewise, the Tephra and Pyroclastic flow models used in assessment of the hazardous areas do not take into account the physical impact of these perils on households. To more precisely represent the vulnerability of communities around both study regions (Vesuvius and Guagua Pichincha) housing structure type, quality of construction, and the age of construction would be key parameters in how a building and its inhabitants coped with such hazards. For example, Spence at al (2008) identified classes of roof vulnerability in Guadeloupe that would be more susceptible during a sub-Plinian eruption of La Soufriere. Spence highlighted that sheet roofs in poor condition and vaulted roofs are much more vulnerable to collapse through ash fall loading than reinforced concrete roofs and those constructed in the last 20 years.
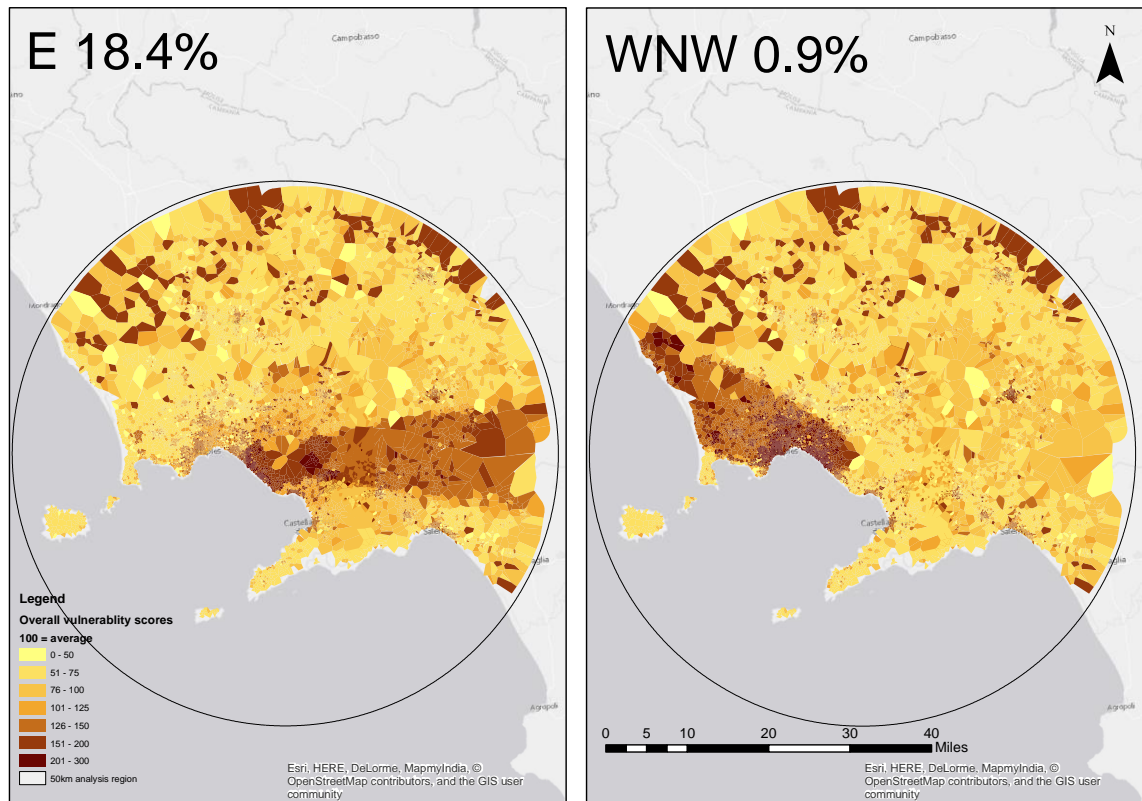
FIGURE 5-3 – SENSITIVITY TESTING OF THE PREVAILING WIND DIRECTION TO OVERALL VULNERABILITY, VESUVIUS

Figure 5-3 highlights the difference that the prevailing wind may have upon both tephra dispersal and on the overall vulnerability assessment of the census areas around Vesuvius. As shown in the image, two possibilities are presented based on NCEP reanalysis data for prevailing winds in the region (Macedonio et al 2008); on the left, an Easterly wind, which occurs 18.4% of the year, and on the right, a West-North-West (WNW) wind, which occurs just 0.9% of the time. Though it has a very low probability, a WNW wind would mean that the city of Naples and the heavily populated bay area would be particularly vulnerable if such an event occurred. Such tests highlight the sensitivity of input parameters to modelling volcanic hazard risk.

It is critical to note that the formal Italian Civil Evacuation areas used in the hazard assessment of the Vesuvius study example have been subsequently criticised in recent years as being

outdated, simplistic and inaccurate (Barberi et al 2008). The pyroclastic flow maps were derived from numerical modelling of PDCs using a 3D flow model (Esposti Ongaro et al 2008) which made necessary assumptions about the process and likely dynamics of column collapse. However, without more consideration and sensitivity testing of these model inputs, it is impossible to categorically define the likely intensity and footprint of the volcanic hazards. Like most models that are based on multiple-disciplines and probabilistic inputs, there is tremendous uncertainty in the likely impact of volcanic hazards near Vesuvius. These uncertainties have not been exhaustively addressed in this work. Perhaps the most fundamental of these remains the key question of whether there'll be a major eruption in the foreseeable future. The periodicity of repose between large eruptions can be upwards of millennia and given that volcanic systems often have the capacity for unpredictable behaviour (Sparks 2003), it is both a complex and troublesome peril to forecast. Eruptive behaviour (such as degassing, seismicity and dome growth) does not always manifest itself in a full eruption and therefore, risk communication prior to an impending eruption is increasingly difficult for NGOs, scientists and authorities as they seek to mitigate the impact of these events. An example of such complexity was found during the ongoing eruptions on the island of Montserrat (1995-2003). Public misunderstanding and mistrust of both the Montserrat Volcano Observatory (MVO) and government officials led people to pursue informal and often misleading informal communication channels (Haynes et al 2008).

DATA AGGREGATION AND THE ECOLOGICAL FALLACY

The nature of geodemographics and clustering implies that the characteristics of an aggregated area are applied to all individuals within that region. This falsely implies that everyone in a given neighbourhood has exactly the same characteristics. This is an invalid assumption, yet is one on which aggregated statistics are based. Though every individual has different characteristics and social vulnerabilities, it is the area average that is being both quantitatively

and qualitatively assessed. The heterogeneity of a census region may be significantly high, resulting in a census profile that is not accurately represented by the geodemographic. However, such variance is lost upon aggregation to higher levels of geography and no single person in the aggregated census area is likely to have a perfect match to the defined profile. With this approach essentially 'pigeon-holing' people into the characteristics of their neighbourhood, there is a clear risk this study is guilty of the 'ecological fallacy' (Robinson 1950). Using aggregated statistics for disaster risk reduction purposes, there is a danger that the minutiae of real and present social vulnerability could be missed during aggregation. Other critics of geodemographics make reference to the questionable ethics of segmenting unknowing consumers in a growing culture of 'dataveillance' (Goss 1995). Likewise, the integral use of decennial surveys such as census data in geodemographics brings into question the relative accuracy of neighbourhood descriptors that are so closely based on census data and survey information (Flowerdew 1991). Given that neighbourhoods can develop or change rapidly, it is questionable whether classification systems (commercial or otherwise) based on 10-year old census data can accurately reflect such flux and precisely represent their changing vulnerability.

TRANSFERRABLE SOCIAL VULNERABILITY INDICATORS

By using social vulnerability indicators drawn from the literature, the risk models in this research are therefore very dependent on specific areas of social science research. Although these publications detail both anecdotal and statistical evidence regarding our current understanding of what makes someone more socially vulnerable during a natural disaster, it would be untrue to say that they are entirely dependable or transferable.

For example, though social vulnerability was noted to be greater for black and ethnic minorities during the onset and post disaster recovery of Hurricane Katrina in the US (Cutter 2006), the same social/cultural vulnerabilities may not be true of an earthquake in Quito. There are countless noticeable differences. In Quito, for example, the more affluent areas of housing stock are often in the Mariscal district, but these houses are also among the oldest and

frequently built without reinforced concrete, could be considered vulnerable. Whereas, houses found in the most economically deprived areas of the city, such as the peripheral barrios, are usually built with reinforced concrete. Therefore, an assumption made for one hazard in one culture does not necessarily translate that the same principal can be applied in alternative geographic settings.

Conversely, and as discussed earlier in this chapter, geodemographic neighbourhoods have been successfully created and applied on a global scale (Mosaic Global 2014) across multiple countries. However, it is important to note, that such neighbourhood segments are extremely broad in their description and are limited to highly developed nation states only. In summary, they have identified similarities between some leading market economies but are certainly not universal.

Unfortunately, given the level of statistical bias, weighting and experience needed to model the complexities of every different community living in a volcanically active area, it would seem unrealistic at present that a universal solution could be derived. Similarly, the choice of census variables for both the Italy and Ecuador case studies demonstrated the reliance of indicators taken from the literature as well as the practical limitations of using census data.

PHYSICAL VULNERABILITIES OUTWEIGH SOCIAL VULNERABILITY

The outlays from this research show the social vulnerabilities of households across very large areas of Italy and Ecuador. If there was a large catastrophic volcanic eruption in one of these study areas, the understanding of this model might imply that a community in very close proximity to an erupting volcano would be affected to the same extent as a more distal area with the same index classification. It would be greatly misleading to assume this.

Volcanoes are, by and large, proximal in their hazards and areas closer to the pyroclastic surges, earthquakes, lahars and large scale ash fallout will almost always be impacted far greater than those regions further away. Maps of social vulnerability to volcanic hazards can

fundamentally be argued to be flawed in principle because no two areas will be physically impacted the same. Likewise, the direct impact of volcanoes is typically limited to certain spatial thresholds (with the exception of very large scale and dispersed ash fallout) and therefore it may be misleading to disseminate maps that may be interpreted otherwise. It is important to remain mindful that defining an area as having low social vulnerability does not discount a volcanic eruption's capacity to still cause significant death, financial ruin or destruction in a community.

Likewise, vulnerability maps should be taken within the context that they were created; i.e. they should be used to help understand the factors, drivers and processes that lead to social vulnerability as well as providing a spatial decision support tool for those seeking to manage and mitigate disasters (e.g. NGOs, government planners, disaster management committees).

CENSUS DATA

As well as the concerns previously mentioned it is worth taking note of some fundamental model limitations in the data used within this research. The 2001 Ecuadorian census is currently 13 years old (as of 2014) and a more recent census was conducted in 2010, subsequent to the initiation of this research agenda (December 2008). The 2010 data remains unavailable at Census output area level at the present time, but there will clearly be marked differences in Ecuadorian demography since 2001. As well as a population growth increases equating to nearly 14% (INEC 2010), there are likely to have been major social, economic and geographic changes to the population landscape during this time. Processes such as a social mobility, migration and health will have radically altered in many areas and therefore the classification presented here is not able to reflect the most recent Ecuadorian census (2010). If the 2010 data is released at census output level in the near future, and in considering further progression of this research agenda, comparative studies in the output of the 2001 and 2010 census data would provide very valuable feedback in the changing nature of social vulnerability. Economic conditions between 2001-2010 have undoubtedly changed, thus having an effect on

employment and GDP in Ecuador. As Figure 5-4 shows, annual GDP has increased following the global economic crisis in 2008/2009, largely driven by manufacturing, construction and public administration. This latter phenomenon has been the consequence of government loans from China (Economist 2013).

Note: Agriculture also includes livestock, forestry, hunting and fishing. Mining and extraction also includes refined petroleum products. Other services inclde: hospitality, communications, rental housing, business and household services, education, and health. Other includes: public utilities, finance, and private households with domestic staff.
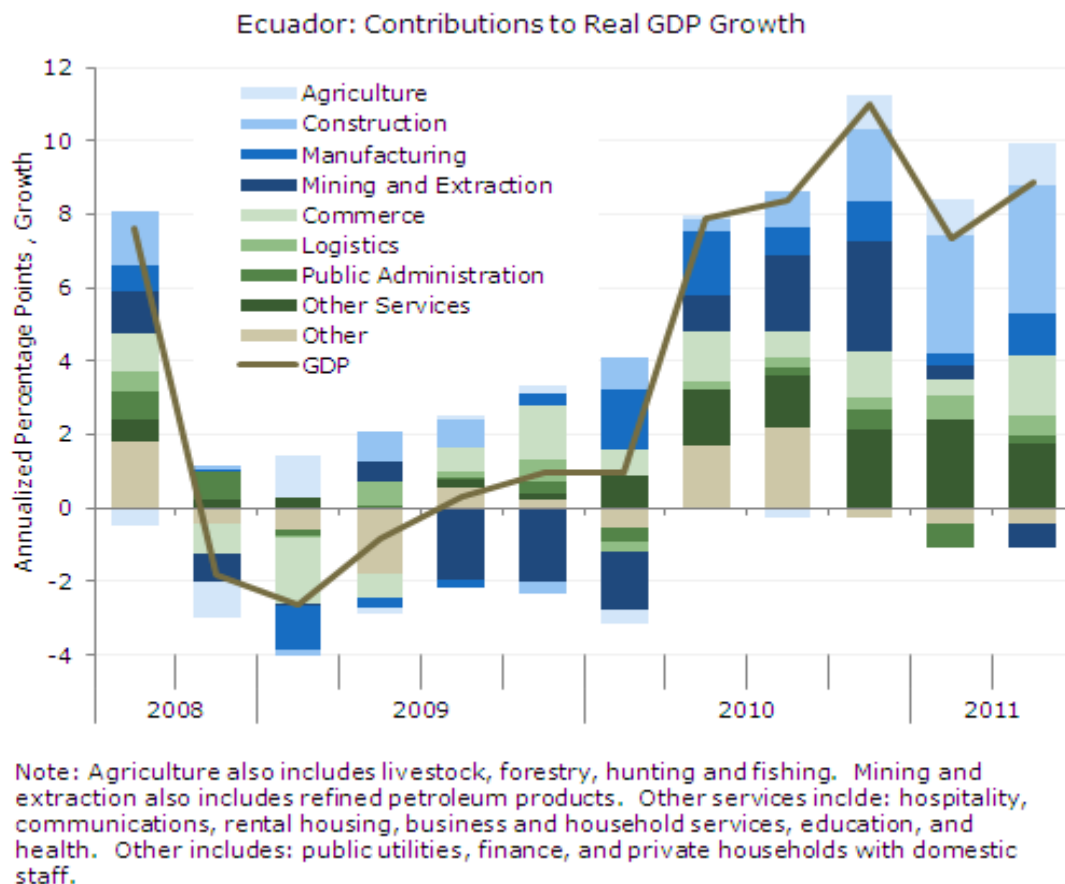
FIGURE 5-4 - ANNUAL GDP AND RELATIVE CONTRIBUTIONS, ECUADOR. (CEPR 2014)

Interestingly, in September 2013, Ecuador's president, Rafael Correa, announced the government will begin oil and gas exploration (and extraction) within the Yasuni national park, Ecuador (Economist 2013). This area is currently a UNESCO protected habitat, home to thousands of endangered species of animal as well as indigenous tribal populations, such as the Huaorani, Tagaeri and Taromenane. Similarly, and with direct relevance to social vulnerability, the ongoing demographic trend of Ecuador in recent years is highlighted in Figure 5-5, and shows an increase in both the *15-64* and *64+* age ranges, and a decreasing *0-14* age range. Such

changes will be reflected in Ecuador's neighbourhood classification and aspects of social vulnerability.
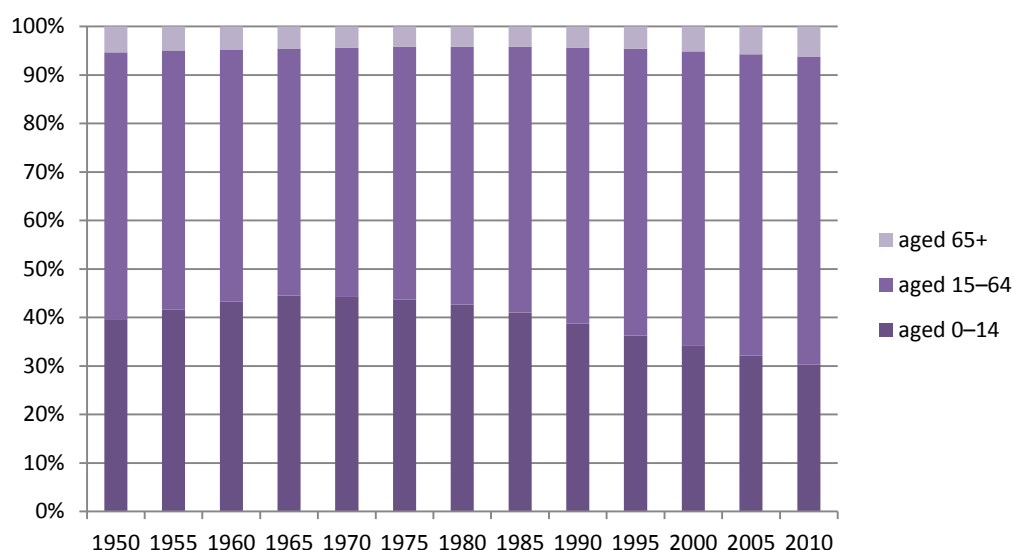
FIGURE 5-5 - ECUADOR DEMOGRAPHIC PROPORTION OF POPULATION 1950-2010 (%) (UN 2010)

SURVEY DATA – GEODEMOGRAPHICS

A factor more relevant to the Italian vulnerability model regards the use of sample marketing data. Because many of the data variables for this model were derived from telemarketing surveys, they represent only a small sample of the total population, yet assumptions are regressed across an entire country.

In doing this, large uncertainty is brought into the behavioural and attitudinal assumptions of a neighbourhood classification system. This is a fundamental component in the use of the geodemographics and uncertainty is directly related to the nature and distribution of the sample taken. It was not within the scope of this agenda to address the sampling regimen and precision but clearly, it will have an effect on the interpretation of neighbourhoods. Given the practical methodology of how telemarketing surveys are conducted, key factors influencing such

uncertainty may include the scale and sampling strategies involved. Without addressing such uncertainly, there is clearly a concern that geodemographic profiles may be associated with large amounts of volatility. Likewise, and with greater relevance to commercial geodemographics, the details of survey data used are not always extensively detailed. Further research around the sensitivities of such questionnaires would be greatly encouraged and hence, would help reduce uncertainties in the use of geodemographics for social vulnerability assessment.

5.3   THESIS CONCLUSIONS

The quantitative methodology, application of geodemography and conceptual model presented in this research represent a new contribution towards the further understanding and management of social vulnerability in volcanic regions. It provides an alternative paradigm in current DRR research methods, which have traditionally measured vulnerability by focusing on single or multiple indicators rather than seeking to profile and understand the behaviour of community types. This thesis highlights how multivariate statistics can be applied to aggregated census output areas to successfully identify and define neighbourhood profiles that exhibit greater social vulnerability. In doing so, it may enable practitioners concerned with DRR research and management (DRM) to make use of such tools for targeted decision support, planning, community outreach, sustainable development, and ultimately, disaster mitigation.

RISK INDICES

Attempts to define disaster risk and social vulnerability at various scalar levels have continued to gain greater traction through the lifetime of HFA. On a global scale, the Disaster Risk Index (DRI) was a major step in attempting to derive a universal approach to classifying risk. The indicator-based method and use of the emergency disasters data-base (EM-DAT) provided a near-global coverage reviewing relative and social vulnerability (UNEP 2004). However, it's important to note that the DRI was limited to earthquakes, tropical cyclones and floods, as well

as being at a national scale of resolution. While the UNEP note that these three perils take into account 96% natural disaster mortality rates, they do not take into account many natural disasters, including volcanic risk, wildfire, and severe convective storm.

With direct application to this research agenda, the Hotspots project (Dilley 2005) sought to measure vulnerability to earthquakes, volcanoes, landslides, floods, drought and cyclones using a GIS grid-based approach. The increased resolution of the Hotspots project (grid squares of approximately 55 Km) and its multi-hazard approach provide a valid and rapid assessment of natural hazard risk at a sub-national resolution. However, the spatial scales of these global assessments remain too broad to capture the intricacies and local impact of natural disasters. Using macro-economic or statistical measures to weight disaster risk, such as the proportion of GDP, the total population at risk, arbitrary defined population grids, does not successfully address the need of policy-makers, NGOs and Governments to mitigate these perils.

Susan Cutter's work regarding the Hazards-of-Place (HOP) model and constructing a US county-level Social Vulnerability Index provided further focus on the granular assessment of risk (Cutter and Mitchell 2000; Cutter et al 2003) and can be seen to be closely aligned in ideology with the contribution presented here. Whilst section 4.3 highlighted the strong correlation between the Gini-Coefficient methodology and Cutter's PCA-based approach (0.72), there remain some fundamental differences in the practical approach of these alternative techniques. Cutter's approach aims at retaining only highly correlated vulnerability variables, thus discounting any factors which may affect social vulnerability but which don't share the same cardinality as the principal factors. The ideology presented in this thesis does not advocate this approach as discounting factors that may impact vulnerability or coping capacity could results in misleading assessments. It is on this basis that the neighbourhood-based approach presented here adds further definition. It seeks to fully characterise the 'at risk' community and type rather than focusing on the individual zone. By doing so, it is proposed a rich geodemographic profile of a neighbourhood can be created which will implicitly help in assessing vulnerability, risk perception and risk behaviour.

Whilst there are benefits of indicator based vulnerability assessments, the research presented here also highlights that vulnerability remains a highly relative, site and hazard specific term. What constitutes a community to be more vulnerable to evacuation in one area, may be entirely different in another cultural/geographic setting or scale. Risk perception, behaviour, and disaster experience are heavily influenced by cultural, ethno-graphic and socio-economic forces. Measuring such complexity and attempting to derive and transfer methodologies in different settings is fraught with uncertainty. Ethnicity has been shown to be highly correlated with social vulnerability in previous disaster experience (Elliot and Pais 2006), yet the same ethnic classification, as defined using census data, can have inverse relationships with vulnerability in different geographic settings.

FUTURE PROGRESSION

This research does not currently propose a single transferrable DRR tool exists or could be derived. However, with further revision of the processes and methodology presented here, there is convincing evidence that the application of geodemography to define social vulnerability in hazardous areas is warranted and would benefit from further progress.

For example, continued testing of vulnerability assumptions and alternative weighting strategies within different geographic settings would be likely to reduce underlying uncertainty and offer insight into the transferrable aspects of indicator-based approaches. Similarly, the necessity and value of 'ground-truthing' quantitative statistical methods with further qualitative survey techniques cannot be underestimated. It establishes a means of validating the methods, assumptions and output of the models. Socially vulnerable communities defined in the geodemographic should be assessed qualitatively to seek consensus with the models. Likewise, understanding cultural risk perception and attitudinal differences is paramount in the interpretation of how individuals and communities will respond during and post-disaster. During the course of this thesis, field work in Ecuador was undertaken to gather qualitative feedback. This work was limited to semi-structured interviews, and concerned a very small

sample size (N=37). In continuing the progression of the research agenda, further work into the direct application and alignment of survey data and geodemographics would be greatly encouraged. This work would be centred on more fully defining the profiles of the neighbourhood.

Although commercial geodemographics remain proprietary and users are not enabled access to the underlying survey data, index values can be sufficiently scrutinized to allow for broad vulnerability studies. A notable drawback of using commercial products regards their transparency of construction. The exact clustering methods, weighting, and treatment of outliers used to define neighbourhoods is typically not chronicled in user guides or literature. Therefore, a large amount of uncertainty exists for statisticians seeking to thoroughly scrutinise such systems. Conversely, geodemographic databases created for academic purposes, such as the Ecuador classification produced during this research allow complete transparency in construction, but are limited by the constraints of academic research. This means researchers are limited in their access to commercial survey data and marketing information needed to more fully define neighbourhood profiles for assessing social vulnerability.

GLOBAL CONTEXT

In consideration of the global research into methods and models for measuring vulnerability to natural hazards, the research presented here builds directly on the principles of earlier work and DRR concepts. As the Hyogo Framework for Action 2005-2015 was superseded by SFDRR in Sendai at the 2015 World Conference on Disaster Risk Reduction (WCDRR), it is apparent that, while a great deal has been accomplished since HFA's conception in 2005, there is much more action needed by SFDRR. HFA provided a platform to raise the agenda on physical, social, economic and environmental factors that influence vulnerability, but it has not proposed to local or national governments how to achieve such resilience in the future. HFA successfully reported and summarised the context and social setting of several disasters during its lifespan but this does not equate to the ultimate target of successful vulnerability assessment and

disaster mitigation. A criticism of HFA and associated initiatives (e.g. Rio +20) was that while numerous countries signed up to the charter, it is arguable how much vulnerability assessment / preparedness has actually been undertaken since that time (Birkmann 2006).

In recent years, the narrative on DRR has increasingly widened in scope from focusing on disaster management (DRM) and sustainability to encompass the growing requirement of nation states to address the risk of future climate change in the context of sustainable livelihoods. The early Intergovernmental Panel on Climate Change (IPCC) reports do make reference to vulnerability, but only with regard to the environmental exposure, the loss of the economic potential (e.g. based on the physical area of impacted land) or where the population density of a likely impacted area was greater than 100 people / Km^2 (Tegart et al 1990). Subsequent IPCC assessments (e.g. AR2, AR3, AR4), and in particular, the special report on '*Managing the risks of extreme events and disasters to advance climate change adaptation*' have attempted to directly address climate change adaptation (CCA) within the complex interactions that underlie disaster risk:

"*Extreme and non-extreme weather or climate events affect vulnerability to future extreme events by modifying resilience, coping capacity, and adaptive capacity.*" (SREX 2012)

As the HFA was replaced in 2015 by SFDRR, it is hoped that the contribution presented here may help further progress our understanding of the future assessment, classification and mitigation of natural disaster risk. Further to this, the conceptual model and method presented here could provide a transferrable paradigm in the measurement of social risk to multiple hazards and is not limited to volcanoes. It is hoped that further progression in this research area will duly continue.

BIBLIOGRAPHY

Acorn. (2011). CACI Geodemographics. Accessed 10th May 2011, Available at: www.caci.co.uk.

Aldenderfer, M.S. and Blashfield, R.K. (1984). *Cluster Analysis*, Sage University Paper series on Quantitative Applications in the social science. Newbury Park, California publications series, 07-044.

Andronico, D. and Cioni, R. (2002). Contrasting styles of Mount Vesuvius activity in the period between the Avellino and Pompeii Plinian eruptions, and some implications for assessment of future hazards. *Bulletin of Volcanology*. 64, 372–391.

Annen, C. and Wagner, J. J. (2003). The impact of volcanic eruptions during the 1990s. *Natural hazards review*, 4 (4), 169-175.

Arnold, M. (Ed.). (2006). *Natural disaster hotspots case studies*. World Bank Publications, (6).

Baker, K., Bermingham, J., and McDonald, C. (1997). The utility to market research of the classification of residential neighbourhoods (Reprinted from Proceedings of The Market Research Society Conference 1979). *Journal of the Market Research Society*, 39 (1), 55-66.

Bankoff, G. (2001). Rendering the world unsafe: 'vulnerability' as western discourse. *Disasters*, 25(1), 19-35.

Barber, E.W., and Barber, P.T. (2004). *When they Severed Earth from Sky: How the Human Mind Shapes Myth*. Princeton University Press, Princeton. 312.

Barberi, F., Davis, M.S., Isaia, R., Nave, R., and Ricci, T. (2008). Volcanic risk perception in the Vesuvius population. *Journal of Volcanology and Geothermal Research,* 172: 244–258.

BBC. (2011). *Devastation of the Haiti Earthquake*. Accessed 20th May 2014, http://www.bbc.co.uk/news/world-latin-america-25706060.

BBC. (2014). *Tungurahua volcano news*. BBC news website. Accessed 15th May 2014, http://www.bbc.co.uk/news/world-latin-america-26019046.

Bellù, L. G., & Liberati, P. (2006). Inequality analysis: The gini index. FAO, EASYPol Module, 40.

Bernstein, R. S., Baxter, P. J., Falk, H., Ing, R., Foster, L., and Frost, F. (1986). Immediate public health concerns and actions in volcanic eruptions: lessons from the Mount St. Helens eruptions, May 18-October 18, 1980. *American Journal of public health*, 76, 25-37.

Bilham, R. (2010). Lessons from the Haiti earthquake. *Nature*, 463 (7283), 878-879.

Birkmann, J. (2006). Measuring vulnerability to promote disaster-resilient societies: Conceptual frameworks and definitions. *Measuring vulnerability to natural hazards: Towards disaster resilient societies*, 9-54.

Blaikie, P. C., Cannon, T. T., Davis, I. and Wisner, B. (1994). *At risk: natural hazards, people's vulnerability, and disaster*. Routledge.

Blong, R. (2003). Building damage in Rabaul, Papua New Guinea, 1994. *Bulletin of volcanology*, 65(1), 43-54.

Blong, R. and McKee, C. (1995). *The Rabaul eruption 1994: destruction of a town*. Natural Hazards Research Centre, Macqauarie University, Australia, 52.

Bonadonna, C., Connor, C.B., Houghton, B.F., Connor, L., Byrne, M., Laing, A., and Hincks, T. (2005). Probabilistic modeling of tephra dispersion: hazard assessment of a multi-phase eruption at Tarawera, New Zealand. *Journal of Geophysical Research*, (110).

Bryant, F. B., and Yarnold, P. R. (1995). Principal-components analysis and confirmatory factor analysis. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics. Washington, DC: *American Psychological Association*, 99-136.

Burton, I., Kates, R.W. and White, G.F. (1978). *The environment as a hazard*. New York. Oxford University Press.

Callingham, M. (2003) *Current commercial sector use of geodemographics and the implications for the ONS area classification systems*. First referenced in Vickers et al 2005 14th October 2003.

Calvache, M. (1995). *The Geological Evolution of Galeras volcanic complex*. Thesis Doctorate. Arizona State University, Tempe, AZ, 180.

Canuti, P., Casagli, N., Catani, F. and Falorni, G. (2002). Modeling of the Guagua Pichincha volcano (Ecuador) lahars. *Physics and Chemistry of the Earth*, Parts A/B/C, 27 (36), 1587-1599.

Cardona, O. D. (2005). *Indicators of Disaster Risk and Risk Management: Program for Latin America and the Caribbean: Summary Report*. Inter-American Development Bank.

Carrion, D., Vasconez, J., and Bermudez, N. (2003). *The global report on human settlements*. United Nations Human Settlements Programme. UN-HABITAT

Castles, S. and Miller M.J. (2003). *The age of migration: International population movements in the modern world*. Routledge.

CED. (2013). Collins English Dictionary - *Complete & Unabridged 10th Edition*. Accessed 6th January 2015, Available at http://dictionary.reference. com/browse/geodemographics.

Chester, D.K., Degg, M., Duncan, A.M., Guest, J.E. (2000). The increasing exposure of cities to the effects of volcanic eruptions: a global survey. *Global Environmental Change B*, Environmental Hazards. Elsevier, 2 (89-103)

Chester, D. K. (2005). Theology and disaster studies: the need for dialogue. Journal of volcanology and geothermal research, 146(4), 319-328.

Clark, G. E., Moser, S. C., Ratick, S. J., Dow, K., Meyer, W. B., Emani, S., Jin, W., Kasperson, J. X., Kasperson, R. E., and Schwarz, H. E. (1998). Assessing the Vulnerability of Coastal Communities to Extreme Storms: The Case of Revere, MA, USA, *Mitigation and Adaptation Strategies for Global Change*, 3, 59–82.

Condie, K. C. (1997). *Plate Tectonics*. Butterworth-Heinemann.

Conrad, C. P., and Lithgow-Bertelloni, C. (2002). How mantle slabs drive plate tectonics. *Science*, 298 (5591), 207-209.

Chakraborty, J., Tobin, G. A., and Montz, B. E. (2005). Population evacuation: assessing spatial variability in geophysical risk and social vulnerability to natural hazards. *Natural Hazards Review*, 6 (1), 23-33.

Chambers, R., & Conway, G. (1992). *Sustainable rural livelihoods: practical concepts for the 21st century*. Institute of Development Studies. UK.

Chouet, B. A. (1996). Long-period volcano seismicity: its source and use in eruption forecasting. *Nature*, 380 (6572), 309-316.

Cioni, R., Bertagnini ,A., Santacroce, R. and Andronico, D. (2008). Explosive activity and eruption scenarios at Somma-Vesuvius (Italy): Towards a new classification scheme. *Journal of Volcanology and Geothermal Research*, 178: 331-346.

Costa, A., Macedonio, G., and Folch, A. (2006). A three-dimensional Eulerian model for transport and deposition of volcanic ashes. *Earth and Planetary Science Letters*, 241 (3-4), 634-647.

Crafts, N., Toniolo, G. (1996). *Economic growth in Europe since 1945*. Cambridge University Press, 441.

Crampton, J. (1995). The ethics of GIS. Cartography and Geographic Information Systems, 22(1), 84-89.

Crisp, J.A. (1984). Rates of volcanic emplacement and volcanic output. *Journal of Volcanology and Geothermal Research*, 20, 177–211.

Cutter, S.L. (1996). Vulnerability to Environmental Hazards. *Progress in Human Geography* 20 (4), 529–39

Cutter, S. L., J. T. Mitchell, et al. (2000). Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina. *Annals of American Geographers* 90 (4), 713-737.

Cutter, S. L., Boruff, B. J. and Shirley, W. L. (2003), Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, 84, 242–261

Daily mail. (2010). *Icelandic volcano*. The daily mail online. Accessed 15th November 2010, http://www.dailymail.co.uk/debate/columnists/article-1267904/Iceland-volcano-eruption-The-price-pay-society-overreacts-risk.html

Dake,K. and Wildavsky, A. (1993). *Theories of risk perception: who fears what and why?* Burger, E.J. (Ed.), Risk. University of Michigan Press, Michigan, 41–58.

Davidson, R. (1997). *An urban earthquake disaster risk index*. The John A. Blume Earthquake Engineering Center, report no. 121. Stanford University, California.

D'Ercole, R. (1991). *Vulnérabilité des populations face au risque volcanique: le cas de la région du Cotopaxi (Equateur)*. Unpublished Ph.D. dissertation, Université Joseph Fourier. Grenoble I.

D'Ercole, R. (1994). Mesurer le risque: le volcan Cotopaxi et les populations proches. André, Y., Bailly, A., Ducret, B., Guérin, J.-P. (Eds.), *Enseigner les Risques Naturels: Pour une Géographie Physique Revisitée. Anthropos / GIP Reclus*, Paris, 111–150.

D'Ercole, R. (1996). Représentation cartographique des facteurs de vulnérabilité des populations exposées à une menace volcanique: application à la region du volcan Cotopaxi (Equateur). *Bulletin de l'Institut Français des Etudes Andines*, 25 (3), 479–507.

De Andrade, M. M. N., Szlafsztein, C. F., Souza-Filho, P. W. M., Araújo, A. D. R., and Gomes, M. K. T. (2010). A socioeconomic and natural vulnerability index for oil spills in an Amazonian harbor: a case study using GIS and remote sensing. *Journal of environmental management*, 91(10), 1972-1980.

DEC. (2014). *Disasters Emergency Committee*. Accessed 27th November 2014, http://www.dec.org.uk/

Decker, R.W. and Decker, B.B. (1991). *Mountains of Fire: The Nature of Volcanoes*. Cambridge University Press, New York.

Derisbourg, J. P. (2002). L'Amérique latine, entre Etats-Unis et Union européenne. *Politique étrangère*, 67(2), 415-434.

Dibben, C., Chester, D.K. (1999). Human vulnerability in volcanic environments: the case of Furnas, Sao Miguel, Azores, *Journal of Volcanology and Geothermal Research*, 92, (1-2), 133-150

Dilley, M. (2005). *Natural disaster hotspots: a global risk analysis*. World Bank Publications, (5).

Donnadieu, F., & Merle, O. (1998). Experiments on the indentation process during cryptodome intrusions: new insights into Mount St. Helens deformation. Geology, 26(1), 79-82.

Donovan, A., & Oppenheimer, C. (2014). Science, policy and place in volcanic disasters: Insights from Montserrat. *Environmental Science & Policy*, 39, 150-161.

Donovan, A., Oppenheimer, C., & Bravo, M. (2011). Rationalising a volcanic crisis through literature: Montserratian verse and the descriptive reconstruction of an island. *Journal of Volcanology and Geothermal Research*, 203 (3), 87-101.

Douglas, M. (1966). Purity and Danger: An Analysis of Concepts of Pollution and Taboo. Routledge, London.

Dow, K., & Cutter, S. L. (2002). Emerging hurricane evacuation issues: hurricane Floyd and South Carolina. *Natural hazards review*, 3 (1), 12-18.

DPC. (1995). *Pianificazione Nazionale d'Emergenza dell'Area Vesuviana*. Dipartimento della Protezione Civile, Roma.

Dumezil, G. (1966). *Archaic Roman Religion: Volume one*. John Hopkins University Press

Dwyer, A., Zoppou, C., Nielsen, O., Day, S., and Roberts, S. (2004). *Quantifying Social Vulnerability: A methodology for identifying those at risk to natural hazards*, Geoscience Australia, 2004/14.

Economist. (2013). *It's hard to be green*. Accessed 22nd of March 2014, http://www.economist.com

Elliott, J. R., & Pais, J. (2006). Race, class, and Hurricane Katrina: Social differences in human responses to disaster. *Social Science Research*, 35 (2), 295–321.

EM-DAT. (2007). *The International Disaster Database*. Accessed 21 September 2011, http://www.emdat.be/

EM-DAT. (2013). *The International Disaster Database*. Accessed 15th May 2014, http://www.emdat.be/

Encyclopedia of the Nations. (2011). Accessed 8th July 2011, www.nationsencyclopedia.com.

Erickson, J., (1988). *Volcanoes and Earthquakes*. Tab Books, Blue Ridge Summit, Pennsylvania.

Ester, M., Kriegel, H. P., and Sander, J. (1997). Spatial data mining: A database approach. *Advances in spatial databases*. Springer Berlin Heidelberg. 47-66

Estrella, B., Estrella, R., Oviedo, J., Narváez, X., Reyes, M. T., Gutiérrez, M., and Naumova, E. N. (2005). Acute respiratory diseases and carboxyhemoglobin status in school children of Quito, Ecuador. *Environmental health perspectives*, 113 (5), 607.

Esposti Ongaro, T., Neri, A., Menconi, G., de Michieli Vitturi, M., Marianelli, .P, Cavazzoni, C., and G. Erbacci, Baxter, P.J. (2008). Transient 3D numerical simulations of column collapse and pyroclastic density current scenarios at Vesuvius. *Journal of Volcanology and Geothermal Research*, 178, 378–396.

Eurostat. (2009). *European Commission website*. Accessed 15th November 2009, http://epp.eurostat.ec.europa.eu

Everitt, B.S., Landau, S., Leese, M. (2001). *Cluster analysis*. 2001. 4th edn. London. Arnold.

Experian. (2011). *Mosaic index*. Accessed 15 November 2009, http://strategies.experian.co.uk

Fekete, A. (2012). Spatial disaster vulnerability and risk assessments: challenges in their quality and acceptance. *Natural hazards*, 61(3), 1161-1178.

Fischoff, B., Slovic, P., Lichtenstein, S., Read, S. and Combs, B., (1978). How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. *Policy Sciences*, 9, 127–158.

Flowerdew, R. (1991). Classified residential area profiles and beyond. University of Lancaster, North West Regional Research Laboratory.

Forbes, L., Jarvis, D., Potts, J., and Baxter, P. J. (2003). Volcanic ash and respiratory symptoms in children on the island of Montserrat, British West Indies. *Occupational and environmental medicine*, 60 (3), 207-211.

Freedman, D. A. (1999). Ecological inference and the ecological fallacy. International Encyclopedia of the social & Behavioral sciences, 6, 4027-4030.

FUB. (2014). *Plate Boundaries diagram*, Freie Universitat of Berlin. Accessed 1st October 2014, http://www.cms.fu-berlin.de /geo/fb/elearning/geolearning/en/mountain_building/plate tectonics/index.html

Fuchs, S., Kuhlicke, C., and Meyer, V. (2011). Editorial for the special issue: vulnerability to natural hazards—the challenge of integration. *Natural Hazards*, 58 (2), 609-619.

Fytikas et al. (1984). *Tertiary to Quaternary evolution of volcanism in the Aegean region.* Geological Society, London, Special Publications, 17, 687-699

Gaither, C. J., Poudyal, N. C., Goodrick, S., Bowker, J. M., Malone, S. and Gan, J. (2011). Wildland fire risk and social vulnerability in the Southeastern United States: An exploratory spatial data analysis approach. *Forest policy and economics*, 13 (1), 24-36.

Gastwirth, Joseph L. (1972). The Estimation of the Lorenz Curve and Gini Index. *The Review of Economics and Statistics,* 54, 306–316

Geotimes. (2005). Accessed 15th November 2010, http://www.geotimes.org/july05/NN_NVEWS.html

Gondard, P. and Lopez, F. (1983). *Inventario Arquelogico Preliminar de los Andes Septentrionales del Ecuador*, MAG, PRONAREG and ORSTOM, Quito.

Google. (2015). *Our Mobile Planet*. Accessed 15th October 2015, http://think.withgoogle.com/mobileplanet/en/

Goss, J. (1995). We Know Who You Are and We Know Where You Live: The Instrumental Rationality of Geodemographic Systems. *Economic Geography*, 171-198.

Greene, M.R., Perry, R.W. and Lindell, M.K., (1981). The March 1980 eruptions of Mt. St. Helens: citizen perceptions of volcano hazard. *Disasters*, 5 (1),49–66.

Guha-Sapir, D., Vos, F., Below, R., and Ponserre, S. (2011). *Annual disaster statistical review 2010*. Centre for Research on the Epidemiology of Disasters.

Hahn, M. B., Riederer, A. M., & Foster, S. O. (2009). The Livelihood Vulnerability Index: A pragmatic approach to assessing risks from climate variability and change—A case study in Mozambique. Global Environmental Change, 19 (1), 74-88.

Hall, M.L, Mothes P.A. (2008). Quilotoa volcano - Ecuador: An overview of young dacitic volcanism in a lake-filled caldera. *Journal of Volcanology and Geothermal Research*. 176 (1), 44-55

Harris, R., Sleight, P. and Webber, R. (2005). *Geodemographics, GIS and Neighbourhood Targeting*. Chichester. Wiley.

Haynes, K., Barclay, J., & Pidgeon, N. (2008). Whose reality counts? Factors affecting the perception of volcanic risk. *Journal of Volcanology and Geothermal Research*, 172 (3), 259-272.

Heilprin, A. (1908). The eruption of Pelee: Philadelphia Geographic Society, Mt. Pelee, Martinique; A study of an active island-arc volcano, Boulder, Colorado, *Geological Society of America Memoir*, 175, 105.

Hewitt, K., Burton, I. (1971). *The Hazardousness of a place: a regional ecology of damaging events*. Toronto: University of Toronto Press.

Hewitt, K. (1983). *Interpretations of calamity from the viewpoint of human ecology*. Winchester M.A. Allen & Unwin

HFA. (2005). *Hyogo Framework for Action 2005-2015: Building the resilience of nations and communities to disasters*, United Nations Office for Disaster Risk Reduction. Accessed 20th September 2014, http://www.unisdr.org/files/1037_hyogoframeworkforactionenglish.pdf

Hicks, A., & Few, R. (2014). Trajectories of social vulnerability during the Soufrière Hills volcanic crisis. Journal of Applied Volcanology, 4(1), 1-15.

House of Japan. (2011). *Tsunami special report: Most disaster victims are elderly*. Accessed 10th October 2013, http://www.houseofjapan.com/local/most-disaster-victims-elderly

Hurst, A.W., and R. Turner, (1999). Performance of the program ASHFALL for forecasting ashfall during the 1995 and 1996 eruptions of Ruapehu volcano, New Zealand. *Journal of Geology and Geophysics*, 42 (4), 615-622.

IATA. (2013). *Volcano Crisis Costs*. Accessed 16th May 2014, http://www.iata.org/.

IFRC. (2013). *International Federation of Red Cross*. Defining vulnerability. Accessed 10th September 2013, http://www.ifrc.org/.

INEC. (2011). *Instituto Nacional de Estadisticas y Censos de Ecuador*. Ecuador. Accessed 10th June 2009, http://www.inec.gob.ec/web/guest/inicio.

ISTAT. (2001). *Istituto Nazionale di Statistica*. Italy. Accessed 9th June 2009, http://www.istat.it/english/.

Jain, A. (2010). Data Clustering: 50 years beyond K-Means. *Pattern Recognition Letters*. 31, 651-666.

Johnson, J. and Zeigler, D. (1986). Evacuation Planning for Technical Hazards: An Emerging Imperative. *Cities*, 3, 148-156.

Kasperson, R.E., (1992). The social amplification of risk: progress in developing an integrative framework. In: Krimsky, S., Golding, D. (Eds.), *Social theories of risk*. Praeger, Connecticut.

Kates, R. W., (1978). *Risk Assessment of Environmental Hazard*, ICSU/SCOPE, Wiley International, 8.

Kellett, J., & Sparks, D. (2012). *Disaster Risk Reduction. Spending where it should count*. Somerset, UK. Global Humanitarian Assistance.

Kelman, I., Mercer, J., & Gaillard, J. C. (2012). Indigenous knowledge and disaster risk reduction. *Geography*, 97, 12.

Kestle, J. (2003). *10 Reasons to use geodemography*. Blog Geodemography, Environics Analytics. Accessed 1st May 2014. environicsanalytics.ca/blog/ea-blog/2011/05/03/10-reasons-to-use-geodemography

Kling, G.W., Clark, M.A., Wagner, G.N., Compton, H.R., Humphrey, A.L., Devine, J.D., Evans, W.C., Lockwood, J.P., Tuttle, M.L., and Koenigsberg, E.J. (1987). The 1986 Lake Nyos gas disaster in Cameroon, West Africa. *Science*, 236, 169–175.

Knapp, G., (1988). *Ecologia Cultural Prehispanica del Ecuador*. Quito. Banco Central del Ecuador.

Kozlowski, R. T. and Mathewson, S. B. (1995). Measuring and Managing Catastrophe Risk. Discussion Papers on Dynamic Financial Analysis, *Casualty Actuarial Society*, Arlington, Virginia.

Labarga, C.M. (1997). Origin and evolution of the African American communities of Esmeraldas (Ecuador). Historical and genetic data. *Annals of the Museum of America*, (5), 129-138

Lachman, R., Bonk,W.J., (1960). Behavior and beliefs during the recent volcanic eruption in Kapoho, Hawaii. *Science,* 131 (3407), 1095–1096.

Leventhal B (1995). Evaluation of geodemographic classifications. *Journal of targeting, measurement and analysis for marketing*. 4, 173-183.

Lew, E.O., & Wetli, C.V. (1996). Mortality from hurricane Andrew. *Journal of Forensic Science*, 41, 449–452.

Litman, T. (2006). Lessons from Katrina and Rita: What major disasters can teach transportation planners. *Journal of Transportation Engineering*, 132 (1), 11-18.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ. Lawrence Erlbaum Associates.

McGuire, W. J., Solana, M. C., Kilburn, C. R. J., & Sanderson, D. (2009). Improving communication during volcanic crises on small, vulnerable islands. *Journal of Volcanology and Geothermal Research*, 183 (1), 63-75.

Macedonio, G., Costa, A., Folch, A.(2008). Ash fallout scenarios at Vesuvius: Numerical simulations and implications for hazard assessment. *Journal of Volcanology and Geothermal Research,* 178, 366–377

Macedonio, G., Pareschi, M.T., and Santacroce, R. (1998). A numerical simulation of the Plinian fall phase of 79 AD eruption of Vesuvius. *Journal of Geophysical Research-Solid Earth and Planets*, 93 (B12), 14817-14827

Marti, J., Spence, R., Calogero, E., Ordonez, A., Felpeto, A. and Baxter, P. (2008). Estimating building exposure and impact to volcanic hazards in Icod de los Vinos, Tenerife (Canary Islands), *Journal of Volcanology and Geothermal Research*, Evaluating Explosive Eruption Risk at European Volcanoes Contribution from the EXPLORIS Project,178 (3), 553-561

Masozera, M., Bailey, M., & Kerchner, C. (2007). Distribution of impacts of natural disasters across income groups: A case study of New Orleans. *Ecological Economics*, 63 (2–3), 299–306.

Masse, W.B., Barber, E.W., Piccardi, L., Barber, P.T. (2007). *Exploring the nature of myth and its role in science*. In: Piccardi, L., Masse, W.B. (Eds.), Myth and Geology. Geological Society of London Special Publication, 1–28.

McMaster, R.B. and Johnson, J.H., jr, (1987). Assessing community vulnerability to hazardous materials with a geographic information asystem. In Chrisman, N.R., editor, Autocarto 8 Proceedings, Falls Church, VA: *American Society for Photogrammetry and American Congress on Surveying and Mapping*, 471-80.

Menard, H.W. (1964). *Marine Geology of the Pacific*. McGraw-Hill, New York.

Middleton, A. (2007). *Globalization, Free Trade, and the Social Impact of the Decline of Informal Production: The Case of Artisans in Quito, Ecuador*, World Development 35 (11)

Mileti, D. S., & Sorensen, J. H. (1987). Natural hazards and precautionary behavior. Taking care: Understanding and encouraging self-protective behavior, 189-207.

Minc, L. (1986). Scarcity and survival: the role of oral tradition in mediating subsistence crises. *Journal of Anthropological Archaeology*, 5, 39–113.

Morrow, B. H. (1999). Identifying and Mapping Community Vulnerability, *Disasters*, 23, 1–18.

Mosaic. (2011). *Mosaic Product suite*, Experian. Accessed 10th May 2011, http://www.experian.co.uk/economics/mosaic-economics.html

Mosaic. (2014). *Mosaic Global*, Experian website. Accessed 4th December 2014, http://www.experian.co.uk/assets/business strategies/ brochures/Mosaic_Global_factsheet [1].pdf

MSNBC. (2010). *Volcanic ash grounds aircraft*. Accessed November 15th 2010, http://www.msnbc.msn.com/id/36543312

Murton, B.J., and Shimabukuro, S. (1974). Human adjustment to volcanic hazard in Puna District, Hawaii. *Natural Hazards: Local, National, Global*. Oxford University Press, New York, 151-159.

OED. (2013). *The Oxford English Dictionary*. Oxford University Press.

ONS. (2010). *Office of National Statistics*. Accessed 10th May 2011, http://www.statistics.gov.uk.

Openshaw, S. (1983). Multivariate analysis of census data: the classification of areas. A census user's handbook. Edited by David Rhind. London, Eng., Methuen, 1983. 243-263.

Openshaw, S., and Turton, I. (1996). A parallel Kohonen algorithm for the classification of large spatial datasets. *Computers & Geosciences*, 22 (9), 1019-1026.

Orford, S., Dorling, D., Mitchell, R., Shaw, M., and Smith, G. D. (2002). Life and death of the people of London: a historical GIS of Charles Booth's inquiry. *Health & place*, 8 (1), 25-35.

PAHO. (2010). *News website*, Pan American Health Organisation. Accessed October 15th 2010, www.paho.org/ecu/

Pareschi, M.T., Favalli, M and Boschi, E. (2006). Impact of the Minoan tsunami of Santorini: Simulated scenarios in the eastern Mediterranean. *Geophysical Research Letters* (33).

Pareschi, M.T., Cavarra, L., Favalli, M., Giannini, F., and Meriggi, A. (2000). GIS and Volcanic Risk Management. *Journal of Natural Hazards*. 21, 361-379.

Peduzzi, P., Dao, H., Herold, C., & Mouton, F. (2009). Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index. *Natural Hazards and Earth System Science*, 9 (4), 1149-1159.

Pelling, M. (2003). *Natural Disasters and Development in a Globalizing World: In a Globalizing World*, Routledge, London.

Pelling, M. (2003). *The Vulnerability of cities*, Earthscan Ltd, London.

Perry, R. W., & Greene, M. (1983). *Citizen response to volcanic eruptions: The case of Mt. St. Helens*. Ardent Media.

Perry, R. W., & Lindell, M. K. (2008). Volcanic risk perception and adjustment in a multi-hazard environment. *Journal of Volcanology and Geothermal Research*, 172 (3), 170-178.

Pierson, T. C., Janda, R. J., Thouret, J. C., & Borrero, C. A. (1990). Perturbation and melting of snow and ice by the 13 November 1985 eruption of Nevado del Ruiz, Colombia, and consequent mobilization, flow and deposition of lahars. *Journal of Volcanology and Geothermal Research*, 41(1), 17-66.

Polsky, Alexander Pulsipher, and Andrew Schiller. (2003). A framework for vulnerability analysis in sustainability science. PNAS, *Proceedings of the National Academy of Sciences of the United States of America*, 100 (14), 8074-8079

Potter, S. H., Jolly, G. E., Neall, V. E., Johnston, D. M., & Scott, B. J. (2014). Communicating the status of volcanic activity: revising New Zealand's volcanic alert level system. *Journal of Applied Volcanology*, 3 (1), 1-16.

Property Claims Services (2014). Accessed 4th March 2014, http://www.verisk.com/property-claim-services/.

Pulido, L. (2000). Rethinking Environmental Racism: White Privilege and Urban Development in Southern California. *Annals of the Association of American Geographers*, 90, 12–40.

Quarantelli, E.L (1978). *Disasters: theory and research*. Sage Publications, London ; Beverly Hills, 282.

Robinson, W.S. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review,* 15 (3): 351–357.

Robock, A. (2000). Volcanic eruptions and climate, *Review of Geophysics*., 38 (2), 191–219.

Rolandi, G. (2010). Volcanic hazard at Vesuvius: An analysis for the revision of the current emergency plan. *Journal of Volcanology and Geothermal Research*, 189, (3-4), 347-362

Ronan, K. R., Paton, D., Johnston, D. M., & Houghton, B. F. (2000). Managing societal uncertainty in volcanic hazards: a multidisciplinary approach. *Disaster Prevention and Management*, 9(5), 339-349.

Rose, W.I.; Durant, A.J. (2009). Fine ash content of explosive eruptions. *Journal of Volcanology and Geothermal Research*, 186 (1-2) 32–39

Rygel, L., O'Sullivan, D., and Yarnal, B. (2006). A Method for Constructing a Social Vulnerability Index. *Mitigation and Adaptation Strategies for Global Change*, 11, 741–

764.

Sarna-Wojcicki, A. M., Shipley, S., Waitt Jr, R. B., Dzurisin, D., & Wood, S. H. (1981). Areal distribution, thickness, mass, volume, and grain size of air-fall ash from the six major eruptions of 1980. *U.S. Geological Survey*, 577-600.

Shah, K. U., Dulal, H. B., Johnson, C., & Baptiste, A. (2013). Understanding livelihood vulnerability to climate change: Applying the livelihood vulnerability index in Trinidad and Tobago. *Geoforum*, 47, 125-137.

Schmidtlein, M. C., Shafer, J. M., Berry, M., and Cutter, S. L. (2011). Modeled earthquake losses and social vulnerability in Charleston, South Carolina. *Applied Geography*, 31(1), 269-281.

Schwartz, E. (2006). *A Needless toll of natural disasters*, Boston Globe.

Shaw, R., Uy, N., & Baumwoll, J. (2008). Indigenous Knowledge for Disaster Risk Reduction: good practices and lessons learned from experiences in the Asia-Pacific Region. *United Nations International Strategy for Disaster Reduction*, *Bangkok*.

Sigurdsson, H. (2000). *Introduction in Encyclopedia of Volcanoes.* Academic Press, San Diego, CA, 1–13.

Slovic, P., Fischhoff, B., Lichenstein, S., (1980). Facts and fears: understanding perceived risk. In: Schwing,R., Albers,W.A. (Eds.), *Societal Risk Assessment: How Safe is Safe Enough?* Plenum, New York, 181–214.

Small, C. and Naumann, T.R. (2001). The Global Distribution of Human Population and Recent Volcanism, *Environmental Hazards*, 3, 93-109.

Small, C., Cohen, J.E., (1999). Continental physiography, climate and the global distribution of human population. *Proceedings of the International Symposium on Digital Earth*, Beijing China, 11, 965–971.

Smith, K. (2013). Environmental hazards: assessing risk and reducing disaster. Routledge.

Sparks, R. S. J. (2003). Forecasting volcanic eruptions. Earth and Planetary Science Letters, 210 (1), 1-15.

Spence, R.J.S., Baxter, P.J., Zuccaro, G. (2004). Building vulnerability and human casualty estimation for a pyroclastic flow: a model and its application to Vesuvius, *Journal of Volcanology and Geothermal Research*, The Neapolitan Volcanoes: Vesuvius, Campi Flegrei and Ischia, 133 (1-4), 321-343.

Spence, R., Komorowski, J. C., Saito, K., Brown, A., Pomonis, A., Toyos, G., and Baxter, P. (2008). Modelling the impact of a hypothetical sub-Plinian eruption at La Soufrière of Guadeloupe (Lesser Antilles). *Journal of Volcanology and Geothermal Research*, 178 (3), 516-528.

Spence, R., Gunesekera, R., Zuccaro, G. (2010). *Insurance Risks from volcanic eruptions in Europe*. Willis Research Network. Accessed 4th September 2012, http://www.willisresearch network.com

SREX, I. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation: Special report of the intergovernmental panel on climate change*. United Nations.

Starr, C. (1969). Social benefit versus technological risk. *Science*. 165 (3899),1232-1238

Susman, P., O'Keefe, P., Wisner, B. (1983).Global disasters, a radical interpretation. In: Hewitt, K. (Ed.), Interpretation of Calamities. *The Risks and Hazards series,* 1, 263–283.

Swiss re. (2014). *Sigma Insurance research tool*. Accessed 28th November 2014, www.swissre.com/sigma/

Tegart, W. J., Sheldon, G. W., & Griffiths, D. C. (1990). *Climate change: The IPCC impacts assessment*. Australian Government Pub. Service.

Thompson, M. (1980). An outline of the cultural theory of risk. *IIASA Working Paper*, 80-177 (December 1980). Accessed 13th March 2012, http://webarchive.iiasa.ac.at/Admin/PUB/Documents/WP-80-177.pdf

Tobin, G.A., Whiteford, L.M. (2002). Community resilience and volcano hazard: the eruption of Tungurahua and the evacuation of the Faldas in Ecuador. *Disasters*, 26 (1), 28–48.

Torry,W.I. (1979).Hazards, hazes and holes: a critique of the environment as hazard and general reflections on disaster research. *Canadian Geographer*, 23 (4), 368–383.

Turner, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., and Schiller, A. (2003). A framework for vulnerability analysis in sustainability science. *Proceedings of the national academy of sciences*, 100(14), 8074-8079

UASB. (2012). *Socio-Economic maps*, Quito, Universidad Andina Simon Bolivar. Accessed 21st March 2014, http://www.uasb.edu.ec/.

UNCDS. (2012). *United Nations Conference on Disaster sustainability*. Accessed 16th May 2014, http://www.uncsd2012.org/.

UNDHA. (1992). Internationally Agreed Glossary of Basic Terms Related to Disaster Management, United Nations Department of Humanitarian Affairs, Geneva.

UNDP. (2012). Act Now, Save later. United Nations Development Programme. Accessed 3rd May 2014, http://www.undp.org/.

UNISDR. (2002). *World disasters report 2002*, United Nations International Strategy for Disaster Reduction, ISBN/ISSN:9291390828

UNISDR. (2004). *Living with risk: a global review of disaster risks reduction*, United Nations International Strategy for Disaster Reduction , ISBN/ISSN:9211010640

UNISDR. (2009). *Disaster Risk Reduction Terminology*. Accessed 10th September 2013, http://www.unisdr.org/

UNWHO. (2010). Personal Correspondence, Pan American World Health Organisation, Quito, Ecuador, November 2010.

UN Stats. (2011). *United Nations Statistics*, UN website, Accessed 17th August 2011. www.UNstats.un.org.

USGS (1999). *What is a volcano*. Accessed 4th July 2014, http://hvo.wr.usgs.gov/.

Vallyathan, V., Robinson, V., Reasor, M., Stettler, L., & Bernstein, R. (1984). Comparative in vitro cytotoxicity of volcanic ashes from Mount St. Helens, El Chichon, and Galunggung. *Journal of Toxicology and Environmental Health*, Part A Current Issues, 14 (5-6), 641-654.

Vickers, D., Rees, P. and Birkin, M. (2005) *Creating the National Classification of Census Output Areas: data, methods and results*. Working Paper 05/2. School of Geography, University of Leeds, Leeds. Accessed 15th May 2012, http://www.geog.leeds.ac.uk/wpapers/05-2.pdf.

Voight, B. (1990). The 1985 Nevado del Ruiz volcano catastrophe: anatomy and retrospection. *Journal of Volcanology and Geothermal Research*, 44(3), 349-386.

Wallace, M. and Denham, C. (1996). *The ONS Classification of Local and Health Authorities of Great Britain*. London. Stationery Office.

Webber, R.J. (1977): *An introduction to the National Classification of Wards and Parishes, Planning Research Applications Group*, London, Centre for Environmental Studies, Technical paper, 23.

Webber, R.J. (1985): The use of census-derived classifications in the marketing of consumer products in the United Kingdom, *Journal of Economic & Social Measurement,* 13, 113-124.

Webber, R.J. (1989): Using multiple data sources to build an area classification system: operational problems encountered by MOSAIC, *Journal of the Market Research Society,* 31, 103-9.

Webber, R.J. (2006). How parties used segmentation in the 2005 British general election campaign. *Direct, Data and Digital Marketing Practice*, 7(3), 239-252.

Webber, R.J. (2012). Advice on geodemographic clustering, personal communication, February, 2012.

Westgate, K., O'Keefe, P. (1976). Some Definitions of disaster. Bradford: University of Bradford Disaster Research Unit, 69

White, G.F. (1945). Human adjustment to floods. Chicago: The University of Chicago.

Whitten, T., Soeriaatmadja, R. E., Suraya A. A . (1996). *The Ecology of Java and Bali*. Hong Kong: Periplus Editions, 309–312

Willis, I., Gibin, M., Barros, J., and Webber, R. (2010). Applying neighbourhood classification systems to natural hazards: a case study of Mt Vesuvius, *Natural Hazards*, Springer, 70 (1), 1-22

Wisner, B., Blaikie, P., Cannon, T., and Davis, I. (2004) At Risk: *Natural Hazards*, People's vulnerability and disasters. London. Routledge. New York.

Wolf, T., & McGregor, G. (2013). The Development of a Heat Wave Vulnerability Index for London, United Kingdom. *Weather and Climate Extremes*, 1, 59-68.

World Bank Group (Ed.). (2012). *World Development Indicators 2012*. World Bank Publications.

World Migration Report 2010. (2010). International Organisation for Migration. Accessed 17th August 2011. http://www.iom.int/jahia/Jahia/world-migration-report-2010.

WorldStat. (2011). Accessed 17th August 2011, http://en.worldstat.info.

Wyly, E. (2001). Quantitative Geographic Analysis, Background on Principal Components and Factor Analysis, University of British Columbia, 2001. Accessed July 7th 2014. http://ibis.geog.ubc.ca/~ewyly/teaching/606_pca.pdf

APPENDIX

APPENDIX 1: INTRODUCTION TO SURVEY QUESTIONNAIRE - QUITO

Department of Geography
Birkbeck, University of London
Malet Street, Bloomsbury
London WC1E 7HX
tel: 020 7631 6000
fax: 020 7631 6270

International research project

We are currently involved in an international research project to study people's thoughts and perceptions to living near Volcanoes. Because Ecuador is a volcanically active region, we very interested in engaging with locals in Quito to understand their thoughts, feelings and possible concerns about the **Guagua Pichincha** volcano.

We would be extremely appreciative if you could spare a few minutes to answer some survey questions to help us gather data in this research topic.
If you would rather not take part in this questionnaire we fully understand and thank you kindly for your time.

This research is being coordinated by Mr Iain Willis, Birkbeck University.

For more information on this subject/topic please contact Iain at one of the following email addresses:
Iwilli02@bbk.ac.uk
iain@gis-consult.co.uk

Thank you for your time,

Iain

APPENDIX 2: RISK PERCEPTION SURVEY – QUITO (ENGLISH)

Introduction

Questionnaire - English

Perceptions

1. There will be a large volcanic eruption of Pichincha in the next 10 years? *Please indicate the level to which you agree/disagree with this statement.*

0%	25%	50%	75%	100%
(Strongly disagree)		(Not sure)		(Strongly agree)
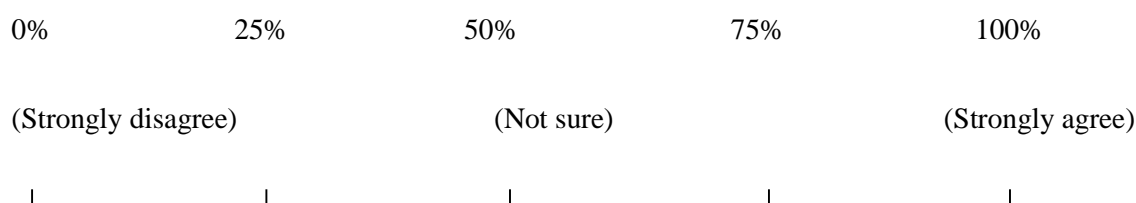
2. Are you worried about a volcanic eruption of Pichincha?

    □ = Not concerned

    □ = A bit worried

    □ = Very worried

3. I feel confident that my house would adequately protect me from ashfall and lahars if there was an eruption? *Please indicate the level to which you agree/disagree with this statement.*

0%	25%	50%	75%	100%
(Strongly disagree)		(Not sure)		(Strongly agree)
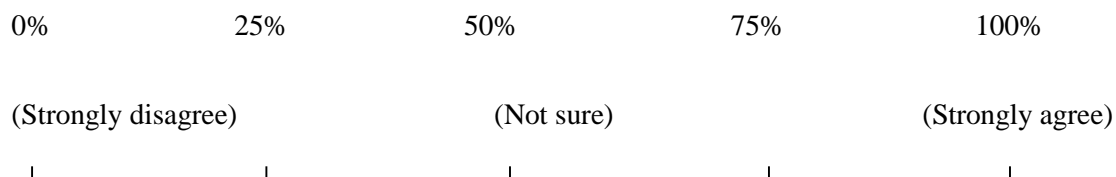
Evacuation

4. Do you have an evacuation plan if there is a volcanic eruption?

☐ = Yes

☐ = No

If 'Yes', can you briefly describe in the box below what your plan would be….

```



```

5. Have you received any official training/guidance/information about evacuation?

☐ = Yes

☐ = No

5a. If 'Yes', did you find the evacuation information useful?

☐ = Yes

☐ = No

Income

6. Would your financial income be affected by a Volcanic eruption?

☐ = Yes

☐ = No

6a. If 'Yes', by how much do you think your earnings would change?

-100%   -50% ←———   0% ———→   50%   100%

(Decrease)   (No change)   (Increase)

└─────────────┴─────────────┴─────────────┘
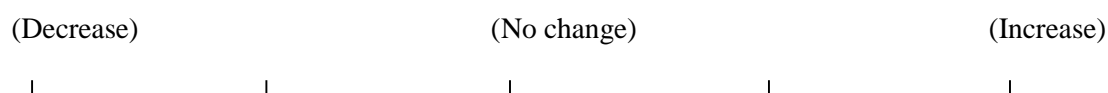
6b. Would this change in income be enough to live on?

☐ = Yes

☐ = No

Demographics

7. How old are you? _____

8. In which parish of Quito do you live? _____

9. What is your marital status? ☐ = Married ☐ = Single ☐ = Divorced ☐ = Other

10. How many people live in your house? Males = ___ Females = ___

11. What is your ethnicity? ☐ = Indigena ☐ = White ☐ = Black ☐ = Mestizo ☐ = Mulato ☐ = *Other*

12. Which category would best describe the principal household income provider's employment type? *Please circle one*

Managerial/Financial	Agricultural/Fishing
Skilled Professional	Mining/Quarrying/Construction
Administrative	Artisan
Civil Service/Social work	Manufacturing
Defence	Other

13. As of September 2010, would you consider yourself to have an illness, disability or infirmity that affects your general mobility? ☐ = Yes ☐ = No

Housing vulnerability – (completed by interviewer)

1. Vertical loading

CF   (Reinforced concrete, infilled frame)

CS   (Reinforced concrete, shear wall)

MB (Masonry, block/squared/cut stone unreinforced)

MC (Masonry, confined or reinforced reinforced)

MR (Masonry, rubble)

ST   (Steel Frame)

TI   (Timber)

2. Building Height

S   1   (single-storey)

L   2   (low-rise)

M   3,4,5   (medium-rise)

H   6+   (high-rise)

3. Age of Building

O   (Old)   Pre-1920

I   (Intermediate)   1920-1940

M (Modern)   1940-1990

R   (Recent)   Post 1990

4. Air Conditioning = **Yes/No**

5. *Building use* = **residential/mixed**

6. *Roof type* = **timber, concrete, coverings**

7. *Combustible material* = **present/not present**

8. Distance between buildings = **<0.7   0.7-1.5   >1.5**

9. Types of window frames = **metal/timber**

10. Types of shutters = **metal/louvred/roller**

11. Condition of openings = **good/poor**

12. Roof type

Roof class	Description	Typical design load range	Mean collapse load
WE (weak)	Sheet roofs, old or in poor condition. Tiled roof, old or in poor condition. Masonry vaulted roof.	Pre-design code, or no design code.	2.0 kPa
MW (medium weak)	Sheet roof on timber; average quality; average or good quality tiled roof on timber rafters or trusses. Steel or precast reinforced concrete joists and flat terrace roof.	1–2 kPa	3.0 kPa
MS (medium strong)	Flat reinforced concrete roof not all above characteristics; sloping reinforced concrete roof. Sheet roof on timber rafters or trusses, good quality and condition, designed for cyclone areas.	2–3 kPa	4.5 kPa
ST (strong)	Flat reinforced concrete roof designed for access; recent, good quality construction, younger than 20 years.	>3 kPa	7.0 kPa

End of the Survey

APPENDIX 3: RISK PERCEPTION SURVEY – QUITO (SPANISH)

Introducción

Cuestionario - Español

Percepciones

1. Habrá una gran erupción volcánica de Pichincha en los próximos 10 años? Por favor, indique el nivel al que usted está de acuerdo / en desacuerdo con esta afirmación.

0%	25%	50%	75%	100%

(Totalmente en desacuerdo)   (No estoy seguro)   (Totalmente de acuerdo)
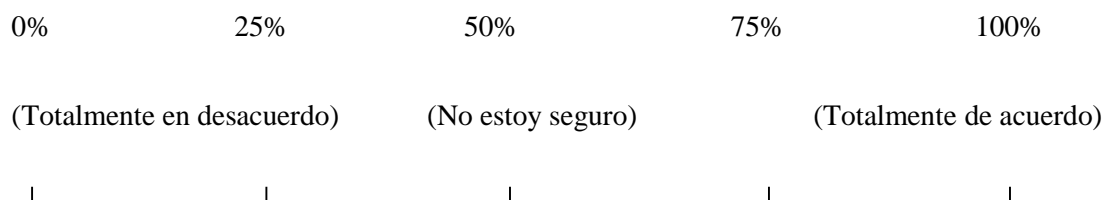
2. ¿Está preocupado por una erupción volcánica del Pichincha?

   ☐ = No es que se trate

   ☐ = Un poco preocupado

   ☐ = Muy preocupado

3. Confío en que mi casa adecuadamente me proteja de la lluvia de cenizas y lahares si hubiera una erupción? *Por favor, indique el nivel al que usted está de acuerdo / en desacuerdo con esta afirmación.*

0%	25%	50%	75%	100%

(Totalmente en desacuerdo)   (No estoy seguro)   (Totalmente de acuerdo)
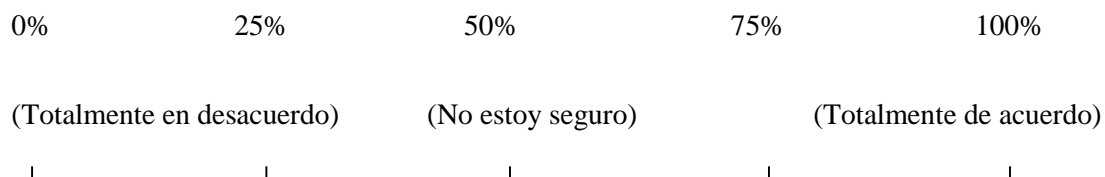
Evacuación

4. ¿Tiene un plan de evacuación si hay una erupción volcánica?

☐ = Si

☐ = No

Si "Sí", ¿puede describir brevemente en el cuadro de abajo lo que su plan sería ..

```



```

5. ¿Ha recibido una formación oficial o de orientación e información acerca de la evacuación?

☐ = Si

☐ = No

5a. Si 'Sí', ha encontrado la información de evacuación útil?

☐ = Si

☐ = No

Ingresos

6. ¿Estaría su renta financiera verse afectados por una erupción volcánica?

☐ = Si

☐ = No

6a. Si "Sí", ¿cuánto cree usted que sus ganancias cambiarías?

-100%   -50% ⟵   0%   ⟶ 50%   100%

(Disminución)	(Sin cambios)	(Aumento)
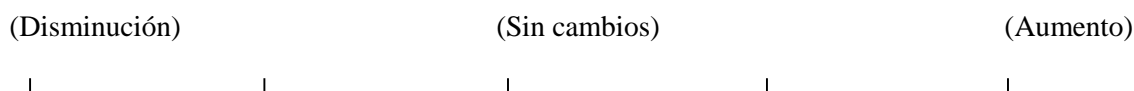
6b. ¿Cambiaría esto en el ingreso suficiente para vivir?

□ = Si

□ = No

Demografía

7. ¿Cuántos años tienes? _____

8. ¿En qué parroquia de Quito vive usted? _____

9. ¿Cuál es su estado civil? □ = Casado □ = Individual □ = Divorciado □ = Otro

10. ¿Cuántas personas viven en tu casa? Los varones = ___ Los varones = ___

11. ¿Cuál es su etnicidad? □ = Indígena □ = Blanco □ = Negro □ = Mestizo □ = Mulato □ =
Otro

12. ¿Qué categoría describe mejor tipo de la casa principal proveedor de ingresos en materia de
empleo? *Por favor, marque uno*

Dirección / Financiero	Agropecuario y Pesca
Experto Profesional	Minería/Extracción/ Construcción
Administrativo	Artesanal
Administración Pública	Producción
Defensa	Otros

13. Hasta septiembre de 2010, se considera usted tiene una enfermedad, discapacidad o enfermedades que afectan a la movilidad en general? ☐ = Si ☐ = No

Vulnerabilidad de la Vivienda – (completada por el entrevistador)

1. Vertical de carga   2. Altura del edificio

CF   (hormigón armado, marco rellenados)   **S**   1   (un solo piso)

CS   (de hormigón armado, muros de cortante)   **L**   2   (de baja altura)

MB   (Albañilería,bloque/cuadrado/corte de piedra sin refuerzo)   **M**   3,4,5   (mediana altura)

MC   (albañilería, confinada o reforzada reforzado)   **H**   6+
(rascacielos)

MR   (Albañilería, escombros)

ST   (estructura de acero)

TI   (Madera)

3. Edad de la Construcción   4. Aire Acondicionado = **Yes/No**

O   (antigua)   before1920   5. *Uso del edificio* = **residencial/mixto**

I   (Intermedio)   1920-1940   6. *Tipo de techo*= **timber, concrete, coverings**

M   (Moderno)   1940-1990   7. *El material combustible* = **presente/ausente**

R   (Recientes)   posterior 1990

8. Distancia entre edificios = **<0.7   0.7-1.5   >1.5**

9. Los tipos de marcos de ventana = **metal / madera**

10. Los tipos de persianas de metal = **metal / lamas / rodillo**

11. Estado de las aberturas **= buena / mala**

12. Tipo de techo

| Roof class	Description	Typical design load range	Mean collapse load
WE (weak)	Sheet roofs, old or in poor condition. Tiled roof, old or in poor condition. Masonry vaulted roof.	Pre-design code, or no design code.	2.0 kPa
MW (medium weak)	Sheet roof on timber; average quality; average or good quality tiled roof on timber rafters or trusses. Steel or precast reinforced concrete joists and flat terrace roof.	1–2 kPa	3.0 kPa
MS (medium strong)	Flat reinforced concrete roof not all above characteristics; sloping reinforced concrete roof. Sheet roof on timber rafters or trusses, good quality and condition, designed for cyclone areas.	2–3 kPa	4.5 kPa
ST (strong)	Flat reinforced concrete roof designed for access; recent, good quality construction, younger than 20 years.	>3 kPa	7.0 kPa

Fin de la Encuesta

Appendix 4: Census Questionnaire – INEC 2001

VI Population Census and V Dwelling Census

2001

Republic of Ecuador

INEC National Institute of Statistics and Census

National Census of Population and Dwelling 2001

The information solicited is strictly confidential.

Article 21 of the Law of Statistics establishes the Law of Statistical Secrecy.

I. Geographic Location

1. [_ _] Province

2. [_ _] County

3. [_ _] County Seat or Rural Parish

4. [_ _] Zone or Community, locality divided by blocks

5. [_ _ _] Sector; Name of disperse section

6. [_ _] Census Area

7. [_ _] Block Number

8. [_ _] Community, Population center, locality, neighborhood, etc.

9. [_ _ _] Number of the dwelling (order of the visit)

10. Dwelling address:

___ Avenue, street and number, block, department, etc.

__ Other identification Road, highway, path, etc.

II. Information about the Dwelling

1. Type of dwelling

Private

[ ] 1 House or villa

[ ] 2 Apartment

[ ] 3 Room in rental house

[ ] 4 Basic housing, zinc roof

[ ] 5 Rural house

[ ] 6 Shack

[ ] 7 Hut

[ ] 8 _____ Other (specify)

Collective

[ ] 11 Hotel, pension, residence hall, hostal

[ ] 12 Military or Police quarters

[ ] 13 Prison

[ ] 14 Hospital, clinic, etc.

[ ] 15 Convent or religious institution

[ ] 16 Other (specify)

If the dwelling is collective, continue with Chapter V (identification of the persons of the household).

2. Occupation condition of the dwelling:

Home Variables Create Extract FAQ Contact Us Login

1 [ ] Oc[ ] cupied present

[ ] 2 Occupied with persons absent (end of interview)

[ ] 3 Unoccupied (end of interview)

[ ] 4 Under construction (end of interview)

3. Predominant construction materials of the dwelling:

A. Roof or covering

[ ] 1 Reinforced concrete

[ ] 2 Asbestos or similar (example: Eternit brand of roofing sheets)

[ ] 3 Zinc

[ ] 4 Clay tiles

[ ] 5 Straw or similar

[ ] 6 Other materials

B. External walls

[ ] 1 Reinforced concrete, brick, or cement block

[ ] 2 Adobe wall or adobe brick wall

[ ] 3 Wood

[ ] 4 Cane with mud covering

[ ] 5 Cane without mud covering

[ ] 6 Other materials

C. Floors

[ ] 1 Wood or parquet

[ ] 2 Tile or vinyl

[ ] 3 Brick or cement

[ ] 4 Cane

[ ] 5 Dirt

[ ] 6 Other materials

D. Frame or structure

[ ] 1 Reinforced cement

[ ] 2 Steel

[ ] 3 Stone

[ ] 4 Wood

[ ] 5 _____ Other (specify)

4. Water service to the dwelling:

A. How does the household acquire water for the dwelling?

[ ] 1 Piped water inside the dwelling

[ ] 2 Piped water outside the dwelling but inside the building, plot, or land

[ ] 3 Piped water outside the dwelling

[ ] 4 No piped water, acquires water through other methods.

B. Where does the water provided to the household come from?

[ ] 1 From a public network

[ ] 2 From a well

[ ] 3 From a river, fall, ditch, or channel

[ ] 4 From a delivery truck

[ ] 5 _____ Other (specify) (example: rain water)

5. How does the household remove the sewage or wastewater from the dwelling?

[ ] 1 Through a public drain or sewer

[ ] 2 Black well

[ ] 3 Septic Tank

[ ] 4 _____ Other form (specify)

6. Doe the dwelling have electricity?

[ ] 1 Yes

[ ] 2 No

7. Does the dwelling have telephone service?

[ ] 1 Yes

[ ] 2 No

8. How does the household dispose of trash?

[ ] 1 Trash collecting vehicle

[ ] 2 Empty lot or ditch

[ ] 3 Burning or burying

[ ] 4 _____ Other form (specify)

9. Without counting the bathroom, how many rooms are in the dwelling?

[_ _] Number

10. Are there persons or groups that prepare their food separately, but sleep in this dwelling?

[ ] 1 Yes

[ ] 2 No

11. How many groups of persons (households) cook their food separately and sleep in this dwelling?

[_ _] Number

Summary of Household Population

Remember to complete this summary as soon as you finish the interview.

[_ _] Men

[_ _] Women

[_ _] Total

For the dwelling where you are interviewing more than one household, use a form for each household,
and repeat Chapter 1 (location).

III. Information about the household.

Number of the household that you are interviewing [_] of [_]

1. Without counting the kitchen or bathroom, how many rooms are occupied by this household?

[_ _] Number

2. In this household, how many rooms are used only for sleeping?

[_ _] Number

3. Does this household have a room, building, or space exclusively for cooking?

[ ] 1 Yes

[ ] 2 No

4. What is the principal combustible or energy used for cooking?

[ ] 1 Natural gas

[ ] 2 Electricity

[ ] 3 Gasoline

[ ] 4 Kerex (kerosene) or diesel fuel

[ ] 5 Wood or charcoal

[ ] 6 _____ Other (specify)

[ ] 7 None (they do not cook)

5. The sanitary service available to this household is:

[ ] 1 Bathroom for the exclusive use of the household

[ ] 2 Bathroom shared by various households

[ ] 3 Latrine

[ ] 4 None

6. The bathing system (shower) available to this household is:

[ ] 1 For the exclusive use of the household

[ ] 2 Shared by various households

[ ] 3 None

7. In this household, is part of the dwelling used for an economic activity?

[ ] 1 ____ Yes; what is the principal activity of the establishment? specify

[ ] 2 No

8. The dwelling occupied by this household is:

[ ] 1 Owned by occupants

[ ] 2 Rented

[ ] 3 Loan-backed habitation contract (anticresis, defined: occupants have use of property in payment
of a debt, until the debt is paid)

[ ] 4 Free

[ ] 5 In payment of services

[ ] 6 ___ Other (specify)

IV. Information about emigrants to other countries.

1. Has one or more persons who were members of this household traveled to another country since
November 1996 (during the past five years) and not returned?

[_] Yes; continue with question 2

[_] No; continue with Chapter 5 (identification of the persons in the household)

How many members of this household traveled?

[_] Number

[There are rows to complete the information about persons numbered 1 through 6].

1. Number, in order (numbers 1-6)

2. Sex

[ ] 1 Man

[ ] 2 Woman

3. Age (at the time the person left the country)

4. Year of departure

5. Motive of the trip (code)

6. Name of the destination country (code)

V. Identification of the persons in the household.

1. What are the names and last names of each of the persons that spent the night from the 24th to
the
25th of November in this household? Begin with the head of household and continue with the rest of
the
members of the family (do not forget newborns and the elderly).

Mr. Census Taker: Write the names and last names of each person in the following order:

1. Head of Household (male or female)

2. Spouse or partner

3. Son or daughter (unmarried children, married children, from eldest to youngest)

4. Son-in-law or daughter-in-law

5. Grandson or granddaughter

6. Parents or parents-in-law

7. Other relatives

8. Other non-relatives

9. Domestic servants

10. Member of the collective household

[There are lines for seven persons]

1. Number of the person

2. Names and last names

3. Relationship to the head of household

4. Sex; Man 1 [ ] Woman 2 [ ]

Total

Important

Do not forget to transcribe the names and last names of each of the persons in the initial

table of the Chapter IV interviews (population info)

questions.

: if the number of persons is more than 7, use an additional form

[One line of instructions at the end of the page is cut off.]

VI. Information about the population

Interview only for the head of household (male or female).

[_ _ _ _] Person Number

___ Names and last names

A. General Characteristics

1. Head of household (male or female) [ ]

2. Is man or woman?

[ ] 1 Man

[ ] 2 Woman

3. How many years have you completed? Write the age at the last birthday.

[_ _] Years completed

4. Does the person have any permanent physical, sensorial, or mental disability? (Incapacity)

[ ] 1 Yes

[ ] 1 To see (blindness, only shadows)

[ ] 2 To move or use his/her body (paralysis, amputations)

[ ] 3 Is deaf or uses hearing aids? (deaf, deaf/mute)

[ ] 4 Mental retardation

[ ] 5 Psychiatric Illness (craziness)

[ ] 6 Multiple (two or more of the above)

[ ] 7 Other (disfigurations, internal organs)

[ ] 2 No

[ ] 9 Don't Know

5. What is the language or dialect that he/she speaks?

[ ] 1 Only Spanish

[ ] 2 Only native language

___ What native language?

[ ] 3 Only foreign language

[ ] 4 Spanish and native language

___ What native language?

[ ] 5 Other (specify) ___

6. How does he/she consider himself/herself: Indigenous, Black (Afro-Ecuadorian), Mestizo, Mulato, White, or Other?

[ ] 1 Indigenous

___ What indigenous nationality or group does he/she belong to?

[ ] 2 Black (Afro-Ecuadorian)

[ ] 3 Mestizo

[ ] 4 Mulato

[ ] 5 White

[ ] 6 Other

7. Where was he/she born?

[ ] 1 In this rural parish or county seat

In another part of the country

___ Rural Parish or County Seat

___ County

___ Province

___ In another country (specify)

[_ _ _ _] Year of arrival in Ecuador

[ ] 9 Don't know

8. Where does he/she usually live?

[ ] 1 In this rural parish or county seat

In another part of the country

___ Rural Parish or County Seat

___ County

___ Province

___ In another country (specify)

[ ] 9 Don't know

9. How long have you lived in the place indicated in the previous question? If less than one year, write 0

[ ] 98 Always

[_ _] Number of years

[_ _] Number of months

[ ] 99 Don't know

10. Five years ago (in November 1996), in what rural parish or county seat did you usually live?

[ ] 1 In this rural parish or county seat

In another part of the country

___ Rural Parish or County Seat

___ County

___ Province

___ In another country (specify)

[ ] 9 Don't know

B. Educational Characteristics

11. Does he/she know how to read and write? If he/she can only read or only write, mark the "no" box.

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

12. Does he/she currently attend a regular educational center? (Literacy Center, Elementary, Secondary,

Basic School, Middle School, Post-high School, University, Post-graduate).

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

13. What is the highest level of education that he/she attends or attended?

[ ] 1 Literacy Center

[ ] 2 Elementary

[ ] 3 Secondary

[ ] 4 Basic School

[ ] 5 Middle School

[ ] 6 Post-high School

[ ] 7 University

[ ] 8 Post-graduate

[ ] 9 Don't know.

14. What is the last grade or highest year that he/she completed in the indicated level?

[ ] 00

[ ] 01

[ ] 02

[ ] 03

[ ] 04

[ ] 05

[ ] 06

[ ] 07

[ ] 08

[ ] 09

[ ] 10

[ ] 99 Don't know

15. Does he/she have a university degree?

Only for persons who have completed their higher education studies.

[ ] 1 Yes

___ What degree does he/she have?

[ ] 2 No

[ ] 9 Don't know.

C. Economic Characteristics

16. Is he/she affiliated with Social Security?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

Does he/she currently contribute?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

17. Is he/she member of a rural [campesina] organization?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

18. What did he/she do last week?

Read the possible answers in the indicated order: Worked (at least one hour), has a job but did not work

(because of illness, vacation, strike, etc), looked for work having worked before (unemployed), looked for

work for the first time, etc. When you receive an answer, mark the corresponding box and move on to

the next question. This question only allows one answer.

[ ] 01 Worked (at least one hour); Continue with question 20

[ ] 02 Has a job but did not work; Continue with question 20

[ ] 03 Looked for work having worked before (unemployed); Continue with question 19

[ ] 04 Looked for work for the first time; Continue with question 19

[ ] 05 Only did housework; Continue with question 19

[ ] 06 Only student; Continue with question 19

[ ] 07 Only retired; Continue with question 19

[ ] 08 Only pensioner; Continue with question 19

[ ] 09 handicapped, cannot work; Continue with question 19

[ ] 10 Other, Specify___; Continue with question 19

[ ] 99 Don't know; Continue with question 19

19. Did he/she do or help do any activity, even if it was not for payment?

For example: planted, harvested, raised animals to sell, washed, ironed, or sewed clothes for someone

else; caught fish to sell; helped serve the public in any business; sold food, artisan products, fruit,

newspapers, clothing or other articles; cared for or watched children or elderly, cured persons who were

ill, helped with childbirth for women who are not part of their household, or carried out any similar

activity.

[ ] 1 Yes Continue with question 20

[ ] 2 No

Unemployed, Continue with question 20

Women, Continue with question 24

Men, Continue with question 28

20. What was the principal activity or work that you carried out during the past week or the last time you

worked before you became unemployed?

Examples: elementary school teacher, construction laborer, agricultural day-laborer, food vendor,

laundress, install steering wheel covers, hairdresser, dressmaker, domestic servant, etc.

_____

21. How many hours did you work at this activity last week or the last week you worked before you

became unemployed?

[_ _ _] Number of hours

22. What is the main activity or main product of the business or establishment where you worked at the

above activity?

Example: Elementary education, industrial textile production, traveling sales, cattle, fishing industry,

mechanic's shop, tailoring, beauty shop, etc.

_____

For the head of household (male or female):

23. What was the position or job category of the work indicated?

Read the possible answers in order and when you receive and answer, mark the corresponding box.

[ ] 1 Owner or active partner

[ ] 2 Self-employed

Employee or salaried worker:

[ ] 3 of the municipality or Provincial Council

[ ] 4 of the State

[ ] 5 of the private sector

[ ] 6 Unpaid family worker

[ ] 9 Don't know

Only for female heads of household:

24. What is the total number of live children you have given birth to?

[ ] 98 None; Continue with question 28

[_ _] Number

[ ] 99 Don't know

25. How many of these children that were born alive are still living?

[ ] 98 None

[_ _] Number

[ ] 99 Don't know

26. On what date was the last son or daughter born alive?

Date: month [_ _] year [_ _]

27. Is the last son or daughter that was born alive still living?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know

For the head of household (male or female):

E. Civil or Married status

28. Currently, are you: in a consensual union, single, married, divorced, widowed, or separated? Mark only one box.

[ ] 1 in a consensual union

[ ] 2 single

[ ] 3 married

[ ] 4 divorced

[ ] 5 widowed

[ ] 6 separated

[ ] 9 don't know

For all persons except the heads of household:

VI. Population data

Questionnaire for the rest of the people in the household.

Person number [_ _ _] Names and last names ___

A. General Characteristics

1. What is your relationship to the head of household?

[ ] 1 Spouse or partner

[ ] 2 Son or daughter

[ ] 3 Son-in-law or daughter-in-law

[ ] 4 Grandson or granddaughter

[ ] 5 Parents or parents-in-law

[ ] 6 Other relatives

[ ] 7 Other non-relatives

[ ] 8 Domestic servants

[ ] 9 Member of a collective household

2. Is man or woman?

[ ] 1 Man

[ ] 2 Woman

3. How many years have you completed? Write the age at the last birthday. For persons less than one year old, write 00.

[_ _] Years completed

4. Does the person have any permanent physical, sensorial, or mental disability? (incapacity)

[ ] 1 Yes

[ ] 1 To See (blindness, only shadows)

[ ] 2 To move or use his/her body (paralysis, amputations)

[ ] 3 Is deaf or uses hearing aids? (deaf, deaf/mute)

[ ] 4 Mental retardation

[ ] 5 Psychiatric Illness (craziness)

[ ] 6 Multiple (two or more of the above)

[ ] 7 Other (disfigurations, internal organs)

[ ] 2 No

[ ] 9 Don't Know

5. What is the language or dialect that he/she speaks? Note to Census Taker: For children less than one

year old, leave this question blank.

[ ] 1 Only Spanish

[ ] 2 Only native language

___ What native language?

[ ] 3 Only foreign language

[ ] 4 Spanish and native language

___ What native language?

[ ] 5 ___ Other (specify)

6. How does he/she consider himself/herself: Indigenous, Black (Afro-Ecuadorian), Mestizo, Mulato,

White, or Other?

[ ] 1 Indigenous

___ What indigenous nationality or group does he/she belong to?

[ ] 2 Black (Afro-Ecuadorian)

[ ] 3 Mestizo

[ ] 4 Mulato

[ ] 5 White

[ ] 6 Other

7. Where was he/she born?

[ ] 1 In this rural parish or county seat

In another part of the country

___Rural Parish or County Seat

___ County

___ Province

In another country (specify) ___; Year of arrival in Ecuador [_ _ _ _]

[ ] 9 Don't know

8. Where does he/she usually live?

[ ] 1 In this rural parish or county seat

In another part of the country

___ Rural Parish or County Seat

___ County

___ Province

___ In another country (specify)

[ ] 9 Don't know

9. How long have you lived in the place indicated in the previous question? If less than one year, write

0

[ ] 98 Always

[_ _] Number of years

[_ _] Number of months

[ ] 99] Don't know

For children less than five years old, finish the interview here and continue with the next person.

10. Five years ago (in November 1996), in what rural parish or county seat did you usually live?

[ ] 1 In this rural parish or county seat

In another part of the country

___ Rural Parish or County Seat

___ County

___ Province

___ In another country (specify)

[ ] 9 Don't know

B. Educational Characteristics

11. Does he/she know how to read and write? If he/she can only read or only write, mark the "no" box.

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

12. Does he/she currently attend a regular educational center? (Literacy Center, Elementary, Secondary,

Basic School, Middle School, Post-high School, University, Post-graduate).

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

13. What is the highest level of education that he/she attends or attended?

[ ] 1 Literacy Center

[ ] 2 Elementary

[ ] 3 Secondary

[ ] 4 Basic School

[ ] 5 Middle School

[ ] 6 Post-high School

[ ] 7 University

[ ] 8 Post-graduate

[ ] 9 Don't know.

14. What is the last grade or highest year the he/she completed in the indicated level?

[ ] 00

[ ] 01

[ ] 02

[ ] 03

[ ] 04

[ ] 05

[ ] 06

[ ] 07

[ ] 08

[ ] 09

[ ] 10

[ ] 99 Don't know

15. Does he/she have a university degree?

Only for persons who have completed their higher education studies.

[ ] 1 Yes

___ What degree does he/she have?

[ ] 2 No

[ ] 9 Don't know.

C. Economic Characteristics

16. Is he/she affiliated with Social Security?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

Does he/she currently contribute?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

17. Is he/she member of a rural [campesina] organization?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know.

18. What did he/she do last week?

Read the possible answers in the indicated order: Worked (at least one hour), HAS A JOB BUT DID NOT WORK (Because of illness, vacation, strike, etc), LOOKED FOR WORK HAVING WORKED BEFORE (unemployed), LOOKED FOR WORK FOR THE FIRST TIME, etc. When you receive an answer, mark the corresponding box and move on to the next question. This question only allows one answer.

[ ] 01 Worked (at least one hour); Continue with question 20

[ ] 02 Has a job but did not work; Continue with question 20

[ ] 03 Looked for work having worked before (unemployed); Continue with question 19

[ ] 04 Looked for work for the first time; Continue with question 19

[ ] 05 Only did housework; Continue with question 19

[ ] 06 Only student; Continue with question 19

[ ] 07 Only retired; Continue with question 19

[ ] 08 Only pensioner; Continue with question 19

[ ] 09 handicapped, cannot work; Continue with question 19

[ ] 10 Other, Specify___; Continue with question 19

[ ] 99 Don't know; Continue with question 19

19. Did he/she do or help do any activity, even if it was not for payment?

For example: planted, harvested, raised animals to sell, washed, ironed, or sewed clothes for someone else; caught fish to sell; helped serve the public in any business; sold food, artisan products, fruit, newspapers, clothing or other articles; cared for or watched children or elderly, cured persons who were

ill, helped with childbirth for women who are not part of their household, or carried out any similar activity.

[ ] 1 Yes Continue with question 20

[ ] 2 No

Unemployed, Continue with question 20

Women, Continue with question 24

Men, Continue with question 28

20. What was the principal activity or work that you carried out during the past week or the last time you

worked before you became unemployed?

Examples: elementary school teacher, construction laborer, agricultural day-laborer, food vendor, laundress, install steering wheel covers, hairdresser, dressmaker, domestic servant, etc.

_____

21. How many hours did you work at this activity last week or the last week you worked before you

became unemployed?

[_ _ _] Number of hours

22. What is the main activity or main product of the business or establishment where you worked at the

above activity?

Example: Elementary education, industrial textile production, traveling sales, cattle, fishing industry, mechanic's shop, tailoring, beauty shop, etc.

_____

For the head of household (male or female):

23. What was the position or job category of the work indicated?

Read the possible answers in order and when you receive and answer, mark the corresponding box.

[ ] 1 Owner or active partner

[ ] 2 Self-employed

Employee or salaried worker:

[ ] 3 of the municipality or Provincial Council

[ ] 4 of the State

[ ] 5 of the private sector

[ ] 6 Unpaid family worker

[ ] 9 Don't know

Only for females over the age of twelve, except female heads of household:

24. What is the total number of live children you have given birth to?

[ ] 98 None; Continue with question 28

[_ _] Number

[ ] 99 Don't know

25. How many of these children that were born alive are still living?

[ ] 98 None

[_ _] Number

[ ] 99 Don't know

26. On what date was the last son or daughter born alive?

Date: month [_ _] year [_ _]

27. Is the last son or daughter that was born alive still living?

[ ] 1 Yes

[ ] 2 No

[ ] 9 Don't know

For all persons over twelve years old, except the head of household (male or female):

E. Civil or Married status

28. Currently, are you: in a consensual union, single, married, divorced, widowed, or separated? Mark

only one box.

[ ] 1 in a consensual union

[ ] 2 single

[ ] 3 married

[ ] 4 divorced

[ ] 5 widowed

[ ] 6 separated

[ ] 9 don't know

Appendix 5: Propensity Index – Ecuador Classification (Sub-groups)

Tier 2 Sub-groups (1-10)	1	2	3	4	5	6	7	8	9	10
% Population	2.0%	1.9%	4.3%	4.2%	0.0%	3.1%	1.6%	3.8%	2.4%	5.6%
Population density	581	825	602	593	327	603	603	131	39	57
0-4	98	86	83	75	18	95	56	76	101	99
>45	86	98	100	113	254	86	146	110	95	89
Disabled_LLTI	79	80	75	57	323	66	54	67	80	94
Degree	61	228	244	393	41	113	805	335	40	119
literate	103	109	110	113	99	106	119	112	99	104
Indigena (e.g. Inca desc)	65	16	51	22	261	37	19	19	154	49
Negro (Black)	65	13	31	25	143	52	22	40	20	48
Mestizo (Mixed race - Native)	105	108	104	102	83	109	84	100	107	107
Blanco (White)	100	143	140	184	40	98	349	189	44	107
Single person HH	64	68	125	92	4	78	155	66	97	94
HH of greater than 6	88	80	49	49	6	65	33	76	105	89
Owned (Propia)	87	75	45	72	2	61	77	97	110	77
Leased (Arrendada)	154	201	327	222	1	261	209	127	46	199
Mortgage (En anticresis)	90	133	131	166	3	103	114	109	54	117
Public housing (Gratuita)	123	111	93	76	2	105	52	65	144	101
Single	101	108	101	101	128	99	102	102	102	103
Married	121	127	115	125	45	120	123	110	127	105
Divorced	124	196	236	257	360	149	428	176	61	127
Administrative	134	200	245	290	196	192	325	245	72	119
Agricultural/Fishing	16	13	5	5	25	6	4	10	106	48
Defence	61	192	93	110	0	103	54	87	58	301

ManagerialFinancial	67	163	217	409	0	124	999	314	45	97
Mining / Construction	214	135	123	103	559	175	49	109	215	126
Skilled Professional	63	229	244	358	0	123	643	310	42	123
Tier 2 Sub-groups (11-20)	11	12	13	14	15	16	17	18	19	20
% Population	5.7%	4.9%	4.5%	4.8%	1.4%	0.0%	4.6%	3.2%	0.5%	1.4%
Population density	109	138	98	35	122	92	52	18	15	9
0-4	83	85	79	91	66	5	94	97	121	109
>45	103	99	106	102	123	91	95	101	94	92
Disabled_LLTI	94	101	89	111	95	38	108	112	62	102
Degree	232	114	232	115	315	290	72	35	35	36
literate	110	108	110	105	114	130	102	97	98	98
Indigena (e.g. Inca desc)	30	17	19	48	52	170	12	9	35	33
Negro (Black)	47	231	44	24	42	26	78	115	72	48
Mestizo (Mixed race - Native)	101	90	103	111	85	78	106	106	108	111
Blanco (White)	162	156	157	84	273	270	122	118	91	79
Single person HH	111	73	83	105	159	10	69	84	156	102
HH of greater than 6	67	97	76	84	54	225	108	105	95	114
Owned (Propia)	64	95	81	95	42	2	104	116	65	100
Leased (Arrendada)	251	132	180	117	345	1	86	55	30	97
Mortgage (En anticresis)	130	137	136	111	68	3	167	107	180	86
Public housing (Gratuita)	96	91	95	113	60	2	110	79	240	120
Single	101	101	99	102	103	74	96	93	95	101
Married	100	74	105	124	87	140	75	55	77	93
Divorced	190	116	167	115	241	275	78	64	76	70
Administrative	181	145	197	126	239	52	103	49	63	50
Agricultural/Fishing	18	18	14	58	6	50	39	101	210	141
Defence	96	143	91	84	112	2570	25	16	114	37

ManagerialFinancial	171	117	190	88	295	151	67	43	89	47
Mining / Construction	128	146	130	129	111	452	132	77	48	67
Skilled Professional	219	119	222	127	299	113	80	39	24	46
Tier 2 Sub-groups (21-30)	21	22	23	24	25	26	27	28	29	30
% Population	0.0%	0.4%	0.6%	0.8%	7.9%	9.0%	2.3%	1.9%	0.6%	1.7%
Population density	32	59	8	12	124	13	5	117	15	9
0-4	128	118	107	97	103	116	125	57	118	92
>45	99	98	91	124	81	91	95	139	84	118
Disabled_LLTI	180	104	115	160	82	74	100	68	100	130
Degree	66	14	22	50	46	11	14	641	31	48
literate	90	77	100	98	102	89	92	117	93	103
Indigena (e.g. Inca desc)	90	1132	16	41	22	15	51	6	10	30
Negro (Black)	59	9	61	8	154	139	20	23	800	136
Mestizo (Mixed race - Native)	99	18	117	116	93	109	117	77	55	112
Blanco (White)	57	14	40	63	169	55	34	400	78	54
Single person HH	163	84	59	179	87	82	93	98	108	126
HH of greater than 6	72	134	142	76	94	128	137	66	139	90
Owned (Propia)	15	119	124	114	120	111	122	96	109	113
Leased (Arrendada)	6	23	30	54	51	24	25	143	68	52
Mortgage (En anticresis)	27	71	140	87	92	114	63	58	141	89
Public housing (Gratuita)	290	118	62	95	51	132	84	41	100	114
Single	118	96	95	100	94	95	103	104	99	103
Married	87	139	126	122	66	44	113	116	32	121
Divorced	34	56	37	72	79	32	32	296	65	73
Administrative	164	28	53	62	135	29	25	254	59	53
Agricultural/Fishing	57	87	64	142	10	183	183	6	78	149
Defence	121	11	66	24	42	23	40	52	28	49

ManagerialFinancial	268	26	27	32	70	26	21	687	43	45
Mining / Construction	91	89	122	80	174	26	41	60	77	69
Skilled Professional	102	14	28	64	55	13	20	507	43	61

Tier 2 Sub-groups (31-42)	31	32	33	34	35	36	37	38	39	40	41	42
% Population	0.5%	2.6%	0.0%	1.7%	2.9%	0.0%	4.3%	0.3%	0.4%	0.8%	0.9%	0.2%
Population density	29	13	19	16	7	15	9	5	2	5	2	3
0-4	100	115	100	117	97	120	109	118	140	116	155	142
>45	108	109	83	102	120	46	84	101	74	97	67	84
Disabled_LLTI	107	131	55	103	128	73	101	94	109	101	95	89
Degree	38	12	42	10	19	114	47	10	23	15	17	14
literate	99	77	99	85	97	102	99	92	86	95	83	73
Indigena (e.g. Inca desc)	71	1081	96	444	52	358	39	242	561	64	1128	40
Negro (Black)	4	3	21	4	6	34	84	70	123	61	14	2184
Mestizo (Mixed race - Native)	115	22	104	82	117	82	103	99	60	113	18	34
Blanco (White)	52	16	66	29	48	73	120	31	69	46	15	33
Single person HH	116	123	123	102	135	87	81	113	84	119	53	126
HH of greater than 6	89	109	113	122	98	89	117	131	169	112	201	155
Owned (Propia)	120	128	72	124	120	40	90	121	113	100	127	123
Leased (Arrendada)	36	16	82	15	21	349	137	23	41	36	28	28
Mortgage (En anticresis)	58	56	15	36	54	29	143	15	65	59	32	89
Public housing (Gratuita)	114	71	84	96	119	27	111	79	88	180	47	78
Single	101	94	114	100	105	95	98	104	104	101	99	97
Married	131	140	83	134	126	132	80	124	92	112	104	43
Divorced	71	46	123	50	59	66	81	49	45	46	32	31
Administrative	45	28	22	19	22	104	74	10	34	36	43	31
Agricultural/Fishing	97	246	37	202	232	31	64	322	186	226	202	154
Defence	20	17	0	13	38	6643	141	104	1172	78	69	12

ManagerialFinancial	29	13	59	12	25	54	53	8	30	37	23	28
Mining / Construction	93	69	911	109	42	64	126	29	52	43	38	46
Skilled Professional	41	17	71	11	17	95	59	12	32	16	31	30